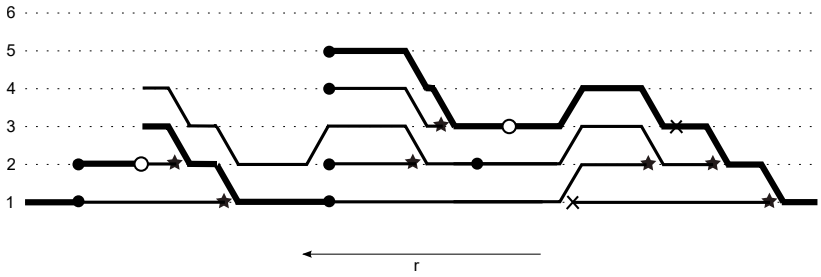
Ancestral lines under mutation and selection



Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich 12 (Informatik und Mathematik)
der Johann Wolfgang Goethe -Universität
in Frankfurt am Main

GOETHE
UNIVERSITÄT
FRANKFURT AM MAIN

von
Ute Lenz
aus Frankfurt am Main

Frankfurt 2016
(D 30)

vom Fachbereich 12 (Informatik und Mathematik)
der Johann Wolfgang Goethe - Universität
als Dissertation angenommen.

Dekan:	Prof. Dr. Uwe Brinkschulte
Gutachter:	Prof. Dr. Anton Wakolbinger (Universität Frankfurt)
	Prof. Dr. Ellen Baake (Universität Bielefeld)
Datum der Disputation:	24.10.2016

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisor Anton Wakolbinger for the continuous support, for sharing his great experience and knowledge with me, and for always taking time to answer questions of any kind. Attending his lectures was always a great pleasure. Every week there was at least one '*Botschaft für's Leben*', e.g. '*wo Tauben sind, fliegen Tauben hin*'.

Secondly, I am very grateful to Ellen Baake for the fruitful collaboration between Bielefeld and Frankfurt, many enlightening discussions (also via e-mail), and for reviewing this thesis.

I thank my co-author (of Chapter 3) Sandra Kluth, and Sebastian Hummel and Fernando Cordero from Bielefeld University for fruitful discussions.

For the purpose of Chapters 3 and 4, I am grateful to Martin Möhle for valuable hints. Concerning the content of Chapter 3, it is my pleasure to thank Tom Kurtz and Peter Pfaffelhuber for stimulating and fruitful discussion. I would also like to thank Jay Taylor and an anonymous referee for their help in improving the first version of [LKBW15].

I am very grateful to all the members of the working group 'Stochastik' in Frankfurt for their (mathematical) support. I am especially indepted to Götz Kersting for always having an open door and for giving many instructive lectures on weak convergence, martingale problems and stochastic analysis. For the good mood in the office, always having an open ear for my problems, and lots of advice (not only on the mathematical side) I am very grateful to Iulia, Cornelia, Andrea and Christina. Thank you very much!
I also enjoyed very much the 'Blockseminar Stochstik' in Riezlern, the 'Sommerfest Stochastik', and other social events. Thank you, Ralph Neininger, Gaby Schneider, Hermann Dinges, Brooks Ferebee, and Gertraud Wakolbinger!
I would also like to thank Anna Weiglhofer for her technical support.

Having a great lunch or coffee break with discussions on mathematical problems, exchanging news on FB 12 or just talking about the weekend activities with friendly colleagues was essential for writing this thesis. I am very thankful to Sven, Björn, Kai, Lukas, Mareike, Thorsten and Thorsten.

I am tremendously thankful to my family and friends for their enduring support. Most of all I thank Christian for his unconditional help and advice in all situations, and for carefully reading a pre-version of this thesis.

I gratefully acknowledge the funding received from Deutsche Forschungsgemeinschaft (SPP Priority Programme 1590: Probabilistic Structures in Evolution, grant no. WA967/4-1).

This thesis was written with the document preparation system LaTeX.
Figures 2.5, and 3.15 are created with the programming language *Maple*.
Figures 5.3, 5.4, and 5.5 are produced with the software environment *R*.
Figures 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 2.1, 2.2, 2.3, 2.4, 2.6, 2.7, 2.8, 2.9, 2.10, 2.11, 2.12, 2.13, 2.14, 2.15, 2.16, 2.17, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 3.10, 3.11, 3.12, 3.13, 3.14, 3.16, 4.1, 4.2, 4.3, 5.1, 5.2, 5.6, 5.7, 5.8 are created with the vector graphics editor *Inkscape*.

Contents

0 Deutsche Zusammenfassung: Anzestrale Linien unter dem Einfluss von Mutation und Selektion

Die Populationsgenetik beschäftigt sich mit dem Einfluss von zufälliger Reproduktion, Mutation, Selektion, Rekombination und Migration auf die genetische Struktur einer Population (vergleiche z.B. [Dur08, Ewe04]). In dieser Arbeit wird das Zusammenspiel von zufälliger Reproduktion, Selektion und Zweiwegmutation untersucht.

Dazu betrachten wir eine haploide Population[1] der Größe $N \in \mathbb{N}$ im Moran-Modell, in der jedes Individuum einen von zwei Typen aus $S := \{0,1\}$ trägt. Die Mutationsrate pro Individuum pro Generation nach Typ 0 sei $u_N v_0$ und nach Typ 1 $u_N v_1$ mit $u_N, v_1, v_0 \geq 0$, $v_0 + v_1 = 1$. Sei 0 der selektiv bevorzugte ('gute') Typ mit Selektionsvorteil s_N und 1 der benachteiligte ('schlechte') Typ. Dann ist die Dynamik des Modells die folgende (siehe Abbildung 0.1): Jedes Individuum bekommt unabhängig von

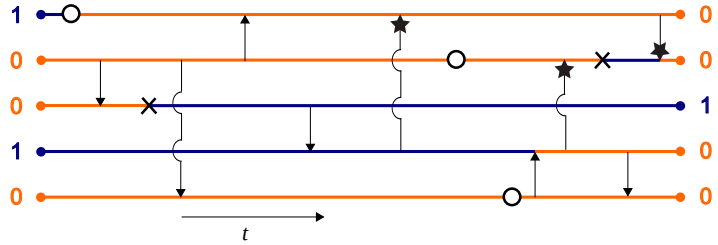


Abbildung 0.1: *Das Moran-Modell mit Zweiwegmutation und Selektion. Mutationen nach Typ* 0 *werden durch Kringel dargestellt, Mutationen nach Typ* 1 *durch Kreuzchen. Neutrale Reproduktionsereignisse sind durch Pfeile mit keilförmigem Kopf dargestellt. Selektive Reproduktionspfeile mit sternförmigem Kopf dürfen nur von Individuen des Typs* 0 *benutzt werden. Im Bild ist* $N = 5$, $K^5 = 3$ *auf der linken Seite und* $K^5 = 4$ *auf der rechten Seite.*

allen anderen Individuen mit Rate $1/2$ ein Kind desselben Typs. Dieses ersetzt ein rein zufällig aus der Population gewähltes Individuum (dargestellt durch *Pfeile* mit keilförmigem Kopf). Zusätzlich zu dieser neutralen Komponente der Reproduktion gibt es eine selektive Komponente der Reproduktion mit Rate s_N pro Individuum (dargestellt durch *Sternchenpfeile*, die nur dann genutzt werden dürfen, wenn das Individuum am Pfeilschaft Typ 0 trägt[2]). Außerdem mutiert jedes Individuum unabhängig mit Rate $u_N v_0$ nach Typ 0 (dargestellt durch *Kringel*) und mit Rate $u_N v_1$ nach Typ 1 (dargestellt

[1] Haploid heißt, dass jedes Individuum nur eine Kopie jedes Chromosoms besitzt (z.B. könnte man sagen eine Mutter) hat.
[2] Diese Arbeit beschäftigt sich mit *fertility* Selektion.

durch *Kreuzchen*). Sei K_t^N die Anzahl der Individuen vom Typ 0 zur Zeit $t \geq 0$, $K_0^N = k_0$, für ein $k_0 \in \{0,1,\ldots,N\}$, und $X_t^N := K_t^N/N$ der Typ-0-Anteil in der Population. Dann ist (K^N) ein Geburts- und Todesprozess mit Geburtenrate λ_k^N und Todesrate μ_k^N, gegeben durch

$$\begin{aligned} \lambda_k^N &= \frac{1}{2N}k(N-k) + k(N-k)\cdot\frac{s_N}{N} \;+\; (N-k)\cdot u_N v_0, \\ \mu_k^N &= \frac{1}{2N}k(N-k) \;+\; k\cdot u_N v_1. \end{aligned} \tag{0.1}$$

Wir nehmen an, dass $k_0/N \to x_0 \in [0,1]$, $Nu_N \to \theta \geq 0$ und $Ns_N \to \sigma \geq 0$ für $N \to \infty$. Dann konvergiert der Prozess $(X_{Nt}^N)_{t\geq 0}$ in Verteilung gegen eine Wright-Fisher-Diffusion mit Mutation und Selektion $(X_t)_{t\geq 0}$ mit Startwert $X_0 = x_0$ und Generator G_X, gegeben durch

$$G_X g(x) = \frac{1}{2}x(1-x)g''(x) + [(1-x)\theta v_0 - x\theta v_1 + \sigma x(1-x)]g'(x), \quad \text{für } g \in \mathscr{C}^2[0,1]. \tag{0.2}$$

Die zugehörige Gleichgewichtsverteilung wird als Wright-Verteilung bezeichnet. Sie hat die Dichte

$$w(x) = c_w \cdot (1-x)^{2\theta v_1 - 1} x^{2\theta v_0 - 1} \cdot \exp\{2\sigma x\}, \tag{0.3}$$

mit der Normierungskonstanten $c_w = \left[\int_0^1 (1-x)^{2\theta v_1 - 1} x^{2\theta v_0 - 1} \cdot \exp\{2\sigma x\} dx\right]^{-1}$ (siehe auch [Dur08, Kapitel 7.2] oder [KHB13, Sektion 2]).

In einer Population, die sich gemäß der eben beschriebenen Dynamik entwickelt, gibt es zu jedem Zeitpunkt τ genau ein Individuum, dessen Nachkommen ab einem bestimmten zukünftigen Zeitpunkt $s > \tau$ die gesamte Population ausmachen werden. Wir nennen dieses Individuum den *gemeinsamen Vorfahren (common ancestor)* zum Zeitpunkt τ, da alle Individuen zu allen Zeitpunkten $t > s$ von ihm abstammen (vgl. Abbildung 0.2). Sei R_τ dessen Typ zum Zeitpunkt τ. In Abbildung 0.1 ist der gemeinsame Vorfahre auf der linken Seite des Bildes das zweite Individuum von oben, also vom Typ 0. Die Linie, die durch alle gemeinsamen Vorfahren durch die Zeit geht, auf der sich also zu jedem

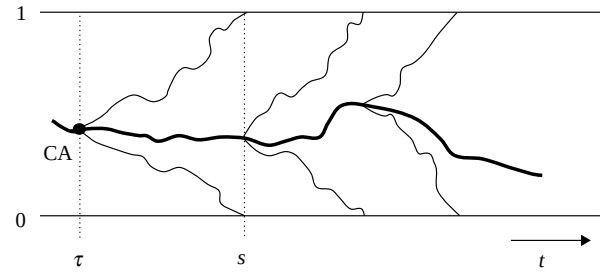
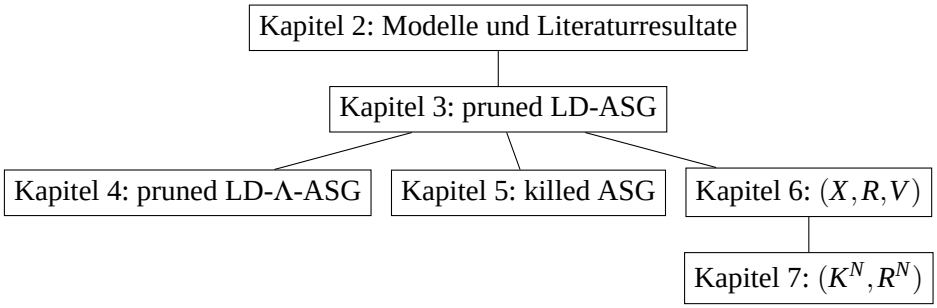


Abbildung 0.2: *Der gemeinsame Vorfahre zur Zeit $t = \tau$ (CA) ist das Individuum, dessen Nachkommen ab dem Zeitpunkt $t = s$ die gesamte Population ausmachen. Die Linie durch alle gemeinsamen Vorfahren durch die Zeit ist die unsterbliche Linie (fette Linie im Bild).*

Zeitpunkt genau das Individuum befindet, dessen Nachkommen zu einem zukünftigen Zeitpunkt in der Population fixiert sein werden, nennen wir *unsterbliche Linie (immortal line)* bzw. (der Notation

von Fearnhead [Fea02] folgend) *reelle Linie (real line)*. Da die anzestrale Linie jedes Individuums nach fast sicher endlicher Zeit in die unsterbliche Linie hinein verschmilzt, kann man den gemeinsamen Vorfahren zum Zeitpunkt τ auch bestimmen, indem man ein beliebiges Individuum zur Zeit $t = +\infty$ aus der Population auswählt und dessen anzestrale Linie aus der Zukunft bis zur Zeit $t = \tau$ zurück verfolgt.
O.B.d.A. sei $\tau = 0$ und der Prozess X zum Zeitpunkt 0 im Gleichgewicht. Dann definieren wir die Wahrscheinlichkeit, dass die unsterbliche Linie zum Zeitpunkt 0 Typ 0 hat, gegeben $X_0 = x$, durch $h(x) := \mathbb{P}(R_0 = 0 \mid X_0 = x)$. Eine Darstellung von $h(x)$ wurde bereits von Fearnhead [Fea02] und Taylor [Tay07] gefunden und dort vorwiegend mit Mitteln der Analysis bewiesen. In dieser Arbeit entwickeln wir ein neues Teilchenbild, den *pruned LD-ASG*, der für sich selbst genommen interessant ist und eine neue probabilistische Interpretation der Darstellung von $h(x)$ liefert.

Die Struktur dieser Arbeit lässt sich als Diagramm mit kurzen Stichworten folgendermaßen darstellen. Dabei baut jedes Kapitel auf allen Kapiteln auf, die sich jeweils in direkter Linie weiter oben im Baumdiagramm befinden.



Kapitel 1 ist eine Einleitung mit Eingliederung dieser Arbeit in den historischen Kontext und Kapitel 2 enthält eine Einführung in die Hauptmodelle, die in dieser Arbeit verwendet werden. Außerdem werden einige Resultate aus der Literatur kurz vorgestellt. In Kapitel 3 dieser Arbeit entwickeln wir ein neues Teilchenbild zur Bestimmung von $h(x)$, das wir in Kapitel 4 von klassischen Wright-Fisher-Diffusionen auf Λ-Wright-Fisher-Diffusionen erweitern. Die sogenannte Siegmund Dualität ist in diesem Fall ein wichtiges Hilfsmittel.
In Kapitel 5 entwickeln wir einen Algorithmus zur perfekten Simulation der Typen einer Stichprobe von $m \in \mathbb{N}$ Individuen, die aus einer Wright-Fisher-Population mit Mutation und Selektion im Gleichgewicht gezogen werden.
Eine Verbindung zwischen Ideen von Taylor [Tay07], der den gemeinsamen Prozess (X,R) untersucht hat, und einem von Fearnhead [Fea02] betrachteten Prozess (R,V), der die Entwicklung des Typs R der unsterblichen Linie in einer Umgebung von V sogenannten virtuellen Linien beschreibt, stellen wir in Kapitel 6 her. Dort bestimmen wir die gemeinsame Dynamik des Tripels (X,R,V). In Kapitel 7 kommen wir zurück zum diskreten Bild mit endlicher Populationsgröße N und schlagen dort eine Brücke zu Resultaten von Kluth, Hustedt und Baake [KHB13].

Die Resultate von Kapitel 3 basieren auf gemeinsamer Arbeit mit Sandra Kluth, Ellen Baake und Anton Wakolbinger und sind publiziert in [LKBW15]. Kapitel 4 basiert auf gemeinsamer Arbeit mit Ellen Baake und Anton Wakolbinger und ist zur Veröffentlichung eingereicht [BLW16]. Die

Inhalte der Kapitel 5, 6 und 7 sind nach gemeinsamen Diskussionen mit Ellen Baake und/oder Anton Wakolbinger entstanden.

0.1 Ein probabilistischer Zugang zur Bestimmung des Typs des gemeinsamen Vorfahren

Um die Verteilung des Typs des gemeinsamen Vorfahren einer Wright-Fisher-Population mit Mutation und Selektion zum Zeitpunkt τ, ohne Beschränkung der Allgemeinheit $\tau = 0$, im Gleichgewicht zu bestimmen, kann man sich alle Individuen der Population zu einem in der Zukunft liegenden Zeitpunkt $s > 0$ anschauen und deren Ahnenlinien zurück bis zum Zeitpunkt 0 verfolgen. Wenn s (in Abhängigkeit von der Realisierung) groß genug gewählt ist, dann haben alle Individuen einen eindeutigen gemeinsamen Vorfahren zum Zeitpunkt 0.
Der gemeinsame Prozess der (potentiellen) Ahnenlinien einer Stichprobe unter Selektion im Teilchenbild wurde erstmals von Krone und Neuhauser erforscht [KN97, NK97]. Der sogenannte *Anzestrale Selektionsgraph (ancestral selection graph)*, kurz *ASG*, ist ein Graph mit Verzweigungen und Verschmelzungen von Linien. Ein Reproduktionsereignis zwischen zwei Linien im Moran-Modell vorwärts in der Zeit (Pfeil mit keilförmiger Spitze) führt zu einem Verschmelzungsereignis im ASG rückwärts in der Zeit[3]. Ein selektiver Sternchenpfeil wird genau dann benutzt, wenn das am Pfeilschaft sitzende Individuum vom Typ 0 ist. Kennt man dessen Typ jedoch nicht, so kommen zwei potentielle Vorfahren des Individuums an der Pfeilspitze als Mutter in Frage und es kommt zu einer Verzweigung im ASG. Die Linie des potentiellen Vorfahren an der Pfeilspitze wird *continuing line* genannt, die Linie, von der der Pfeil stammt, heißt *incoming line*.
Wenn s groß genug gewählt ist, dann muss man sich zur Bestimmung des Typs des gemeinsamen Vorfahren zur Zeit 0 nicht die Ahnenlinien aller Individuen zur Zeit s in der Population anschauen. Da bis zur Zeit 0 zurück alle Ahnenlinien in die unsterbliche Linie verschmolzen sein werden, reicht es, sich eine beliebige Linie zur Zeit s zu ziehen und dessen ASG zu betrachten. (In Abbildung 0.3 ist der ASG der zweitobersten Linie am rechten Rand von Abbildung 0.1 gezeigt.) Wir betrachten daher immer den mit nur einer Linie zu einer sehr großen Zeit s gestarteten ASG.
Mutationen werden im ASG folgendermaßen modelliert: Für jede Linie gibt es einen unabhängigen

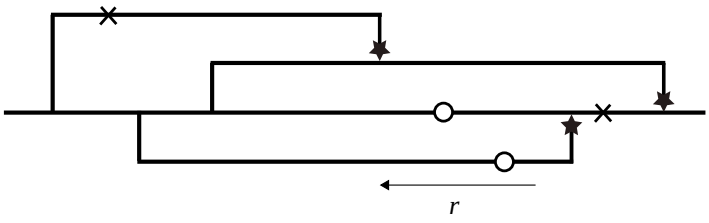


Abbildung 0.3: *ASG mit Mutationen auf den Linien. Es sind alle potentiellen Vorfahren des zweitobersten Individuums aus Abbildung 0.1 gezeigt.*

[3] In dieser Arbeit indizieren wir Rückwärtszeit stets mit r und Vorwärtszeit mit t. In den Abbildungen läuft die Rückwärtszeit immer von rechts nach links und Vorwärtszeit von links nach rechts.

Poissonschen Punktprozess von Typ-0-Mutationen (Kringeln) mit Rate $\theta\nu_0$ und Typ-1-Mutationen (Kreuzchen) mit Rate $\theta\nu_1$.

Eine Realisierung der Typen der Linien im ASG zur Zeit 0, gegeben $X_0 = x$, kann man durch ein Bernoulliexperiment mit Erfolgsparameter x bekommen. Kennt man dann die Typen aller Linien im ASG zur Zeit 0, so kann man den Typ des gemeinsamen Vorfahren zur Zeit 0 (und den Typ des Individuums zur Zeit s) bestimmen. Dies ist im Fall ohne Mutationen einfach (vgl. [Man09, Theorem 2.1]): Der Typ des gemeinsamen Vorfahren zur Zeit 0 ist genau dann 0, wenn mindestens eine der Linien im ASG zum Zeitpunkt 0 Typ 0 trägt. Denn Typ 0 setzt sich immer durch: Sobald entweder die continuing line oder die incoming line vom Typ 0 ist, ist auch der Nachkomme eines selektiven Ereignisses vom Typ 0.
Im Fall mit Mutationen ist die Bestimmung des gemeinsamen Vorfahren zur Zeit 0 aber recht komplex. Man muss zuerst die Typen aller potentiellen Vorfahren zur Zeit 0 kennen (bzw. mit einem Bernoulli(x)-Experiment simulieren indem man die potentiellen Vorfahren gemäß den Erfolgen und Misserfolgen des Experiments mit den Typen 0 oder 1 *einfärbt*). Diese Typen werden dann vorwärts in der Zeit entlang der Linien des ASG unter Respektierung der Mutationen bis zu dem Individuum zur Zeit s transportiert. Dann wird der wahre Vorfahre zu jedem selektiven Ereignis rückwärts in der Zeit bis zur Zeit 0 bestimmt und somit die unsterbliche Linie zurück bis zur Zeit 0 verfolgt.

Diese recht komplizierte Prozedur wird in Kapitel 3 dieser Arbeit durch unser neues Teilchenbild, den *pruned LD-ASG*, erheblich vereinfacht. Dazu starten wir mit dem ASG und führen eine Ordnung der Linien (inspiriert von Donnelly und Kurtz [DK99a]) ein: Bei einem selektiven Verzweigungsereignis zeichnen wir die incoming line immer direkt unter die continuing line (vgl. Abbildung 0.4). Bei einem Verschmelzungsereignis führen wir immer die untere der beiden beteiligten Linien weiter, d.h. es verschmilzt stets die obere der beiden beteiligten Linien in die untere hinein. Um zu jedem

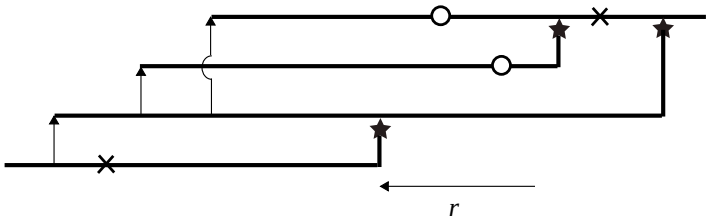


Abbildung 0.4: *Geordnete Version des in Abbildung 0.3 dargestellten ASG.*

Zeitpunkt eine Nummerierung der Linien zu bekommen, platzieren wir die Linien auf Level $1, 2, \ldots$; und zwar so, dass zu jedem Zeitpunkt die unterste Linie auf Level 1 ist, die zweitunterste auf Level 2, u.s.w. Somit gibt also zu jedem Zeitpunkt die Nummer des höchsten besetzten Levels die Anzahl der vorhandenen Linien im ASG an. Dies führt dazu, dass bei einem selektiven Verzweigungsereignis auf Level $k \in \mathbb{N}$ die incoming line auf Level k platziert wird, die continuing line auf Level $k+1$ und alle Linien auf den Levels $\ell > k$ um ein Level nach oben auf $\ell + 1$ geschoben werden. Bei einem Verschmelzungsereignis der Linien auf Level k_1 und Level $k_2 > k_1$ verschmilzt die Linie auf Level k_2 in die Linie auf Level k_1. Um den freien Platz aufzufüllen werden alle Linien auf den Levels $\ell > k_2$ um ein Level nach unten auf $\ell - 1$ geschoben (vgl. Abbildung 0.5). Den auf diese Art und Weise

entstandenen Graphen nennen wir *lookdown ASG* oder kurz *LD-ASG*, in Anlehnung an das *lookdown Modell* von Donnelly und Kurtz [DK99a].

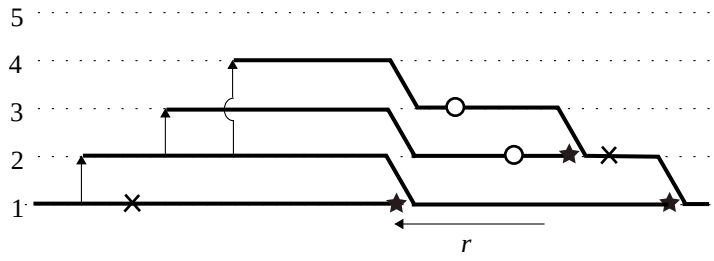


Abbildung 0.5: *LD-ASG zu Abbildung 0.4. Die Levels sind von unten nach oben durchnummeriert.*

Im Fall ohne Mutationen führt die spezielle Ordnung der Linien bereits dazu, dass wir die unsterbliche Linie im lookdown ASG zur Zeit 0 direkt angeben können, wenn wir die Typen der Linien zur Zeit 0 kennen. Dazu sei die *immune Linie* (*immune line*) zu jeder gegeben Zeit die Linie, die die unsterbliche Linie zu dieser Zeit ist, falls alle Linien zu dieser Zeit vom Typ 1 sind. Bei jedem auf ihrem aktuellen Level stattfindenden selektiven Ereignis rutscht sie zusammen mit der continuing line dieses Ereignisses um ein Level nach oben. Ansonsten verhält sie sich so wie jede andere Linie auch.

Proposition 0.1 *Für fast alle Realisierungen des LD-ASG im Gleichgewicht im Fall ohne Mutationen ist das Level der unsterblichen Linie zur Zeit* 0 *entweder das niedrigste Typ*-0-*Level zur Zeit* 0 *oder, falls alle Linien zur Zeit* 0 *mit Typ* 1 *gefärbt sind, das Level der immmunen Linie zur Zeit* 0*.*

Der Beweis von Proposition 0.1 wird induktiv von Ereignis zu Ereignis (Verzweigung oder Verschmelzung) geführt. Dabei nutzt man aus, dass die incoming line (die immer direkt unterhalb der continuing line liegt) bei jedem selektiven Verzweigungsereignis genau dann anzestral ist, wenn sie vom Typ 0 ist.

Im Fall mit Mutationen muss der LD-ASG weiter modifiziert werden. Jede Mutation liefert nämlich neue Informationen über die Typen der Linien. Daher fallen manche Linien als potentielle Vorfahren weg und können aus dem Graphen gelöscht werden. Die Pruning-Prozedur, die vom LD-ASG zum *pruned LD-ASG* führt, ist die folgende (vgl. Abbildung 0.6): Wird eine Linie, die nicht die immune

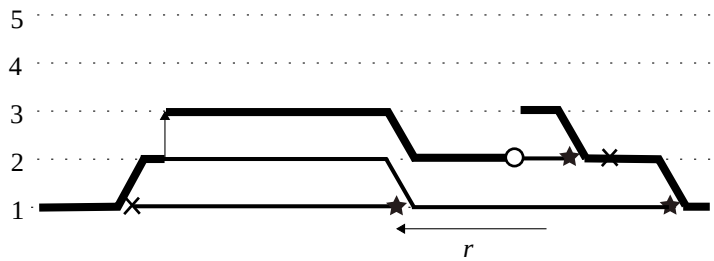


Abbildung 0.6: *Pruned LD-ASG abgeleitet aus Abbildung 0.5. Die immune Linie ist fett markiert.*

Linie ist, von einer Typ-1-Mutation (Kreuzchen) getroffen, so wird diese Linie abgeschnitten. Alle anderen Linien auf höheren Levels rutschen um ein Level nach unten um die frei gewordene Lücke wieder aufzufüllen. Wenn eine Typ-1-Mutation auf der immunen Linie passiert, dann wird diese nicht abgeschnitten (daher auch der Name der Linie). Damit allerdings die Ordnung der Linien respektiert wird, rutschen alle Linien, die auf höheren Levels als die immune Linie sitzen, um ein Level nach unten. Die immune Linie selber wird dann auf das vor der Umordnung höchste besetzte Level platziert. Wird eine Linie von einer Typ-0-Mutation (Kringel) getroffen, so werden alle Linien auf höheren Levels abgeschnitten. Die Linie auf dem Level des Kringels setzt die immune Linie fort.
Durch das Abschneiden von Linien nach Mutationen überträgt sich die Aussage von Proposition 0.1 dann auch auf den Fall mit Mutationen.

Theorem 0.2 *Für fast alle Realisierungen des pruned LD-ASG im Gleichgewicht ist das Level der unsterblichen Linie zur Zeit* 0 *entweder das niedrigste Typ*-0*-Level zur Zeit* 0 *oder, falls alle Linien zur Zeit* 0 *mit Typ* 1 *gefärbt sind, das Level der immmunen Linie zur Zeit* 0*. Insbesondere ist die unsterbliche Linie zur Zeit* 0 *genau dann und nur dann vom Typ* 1*, wenn alle Linien zur Zeit* 0 *vom Typ* 1 *sind.*

Für den Beweis von Theorem 0.2 nutzt man einerseits aus, dass eine incoming line, die durch eine Typ-1-Mutation schlecht gefärbt wurde, nicht anzestral sein kann. Andererseits bewirkt jede incoming line, die durch eine Typ-0-Mutation gut gefärbt wurde, dass die zugehörige continuing line kein potentieller Vorfahre mehr sein kann.

Sei L der Linienzählprozess des pruned LD-ASG. Dann sind Übergangsraten von L gegeben durch

$$\begin{aligned} &q_L(n,n-1) = \frac{1}{2}n(n-1) + (n-1)\theta\nu_1 + \theta\nu_0, \\ &q_L(n,n+1) = n\sigma, \\ &q_L(n,n-\ell) = \theta\nu_0, \qquad\qquad 2 \le \ell < n, \quad n \in \mathbb{N}. \end{aligned} \tag{0.4}$$

Mithilfe von Theorem 0.2 ist es nun möglich eine Interpretation der Wahrscheinlichkeit $h(x)$ und gleichzeitig das Haupttheorem von Kapitel 3 anzugeben.

Theorem 0.3 *Für die Wahrscheinlichkeit, dass der gemeinsame Vorfahre zur Zeit* 0 *vom Typ* 0 *ist, gegeben die Frequenz der Typ*-0*-Individuen zur Zeit* 0 *ist* $X_0 = x$*, erhält man*

$$h(x) = \sum_{n\ge 1} x(1-x)^{n-1}\mathbb{P}(L_0 \ge n), \tag{0.5}$$

wobei L_0 *die Anzahl der Linien eines stationären pruned LD-ASG zur Zeit* 0 *ist.*
Die Tailwahrscheinlichkeiten $a_n := \mathbb{P}(L_0 > n)$ *sind eindeutig bestimmt durch die Rekursion*

$$(n+1+2\theta\nu_1)a_{n+1} - (n+1+2\sigma+2\theta)a_n + 2\sigma a_{n-1} = 0, \quad n \ge 1, \tag{0.6}$$

mit den Randbedingungen

$$a_0 = 1, \qquad \lim_{n\to\infty} a_n = 0. \tag{0.7}$$

Eine Reihendarstellung von h, $h(x) = \sum_{n\geq 1} x(1-x)^{n-1}a_n$, wobei die Koeffizienten (a_n) - die wir '*Fearnhead Koeffizienten*' nennen - durch die Rekursion (0.6) gegeben sind, kann man bereits in [Fea02] und [Tay07] finden. Die probabilistische Interpretation der a_n als Tailwahrscheinlichkeiten und der pruned-LD-ASG selbst sind neue Resultate dieser Arbeit (bzw. der Veröffentlichung [LKBW15]).

0.2 Der Typ des gemeinsamen Vorfahren einer Λ-Wright-Fisher-Diffusion

In Kapitel 4 dieser Arbeit erweitern wir die Resultate des Kapitels 3 (und somit auch die Resultate von Fearnhead [Fea02] und Taylor [Tay07]) auf Λ-Wright-Fisher-Diffusionen mit Mutation und Selektion.

Wir betrachten Populationen dessen Typ-0-Frequenz sich gemäß des folgenden Generators entwickelt (vgl. [EGT10, Gri14]),

$$\begin{aligned} G_X g(x) = \int_{(0,1]} \big[x(g(x+z(1-x))-g(x)) + (1-x)(g(x-zx)-g(x))\big] \frac{\Lambda(dz)}{z^2} \\ + \Lambda(\{0\}) \cdot \frac{1}{2} x(1-x) g''(x) \; + \; \big[\sigma x(1-x) - \theta\nu_1 x + \theta\nu_0(1-x)\big] g'(x), \end{aligned} \tag{0.8}$$

$g \in \mathscr{C}^2$, wobei das *Reproduktionsmaß* Λ ein Wahrscheinlichkeitsmaß auf $[0,1]$ ist. Der Spezialfall $\Lambda = \delta_0$ ist der in Kapitel 3 (vgl. Sektion 0.1) behandelte Fall.

Sei weiterhin $h(x) = \mathbb{P}(R_0 = 0 \mid X_0 = x)$ die Wahrscheinlichkeit, dass die unsterbliche Linie in einer stationären Situation zur Zeit 0 Typ 0 hat, gegeben $X_0 = x$. Man beachte nur, dass X jetzt eine Λ-Wright-Fisher-Diffusion (mit Generator (0.8)) ist. Dann gilt unter der Bedingung $0 \leq \sigma < -\int_0^1 \log(1-x)\frac{\Lambda(dx)}{x^2}$ die folgende Aussage.

Theorem 0.4 *Die Wahrscheinlichkeit $h(x)$ hat die Reihendarstellung*

$$h(x) = \sum_{n\geq 0} x(1-x)^n a_n, \tag{0.9}$$

wobei die Koeffizienten a_n in (0.9) *monoton fallend sind und die eindeutige Lösung des folgenden Systems von Gleichungen,*

$$\begin{aligned} \left[\sum_{n+1<c\leq\infty} \frac{1}{n}\binom{c-1}{c-n}\lambda_{c,c-n}\right](a_n - a_{c-1}) + (\sigma+\theta)a_n = \sigma a_{n-1} + \theta\nu_1 a_{n+1}, \quad n \geq 1, \\ a_0 = 1, \quad a_\infty := \lim_{n\to\infty} a_n = 0, \end{aligned} \tag{0.10}$$

mit der Konvention

$$\binom{\infty-1}{\infty-d+1} := \begin{cases} 0 & \text{falls } d = 1 \\ 1 & \text{falls } d \geq 2 \end{cases} \qquad \text{und} \quad \lambda_{\infty,\infty} := \Lambda(\{1\}). \tag{0.11}$$

Der Beweis von Theorem 0.4 führt über eine Erweiterung des pruned LD-ASG auf den Λ-Fall, den *pruned lookdown Λ-ASG*. Dieser verhält sich bei Mutationen und selektiven Ereignissen genau so wie der pruned LD-ASG. Zusätzlich zu den Koaleszenzereignissen von zwei Linien gibt es Verschmelzungsereignisse beliebiger Größe. Diese werden wie im von Pitman [Pit99], Sagitov [Sag99] und Donnelly und Kurtz [DK99b] eingeführten Λ-*Koaleszenten* modelliert. Die Rate mit der jedes beliebige aber feste Tupel von j aus b Linien in eine verschmilzt ist gegeben durch

$$\lambda_{b,j} := \int_0^1 z^j (1-z)^{b-j} z^{-2} \Lambda(dz), \quad j \leq b. \tag{0.12}$$

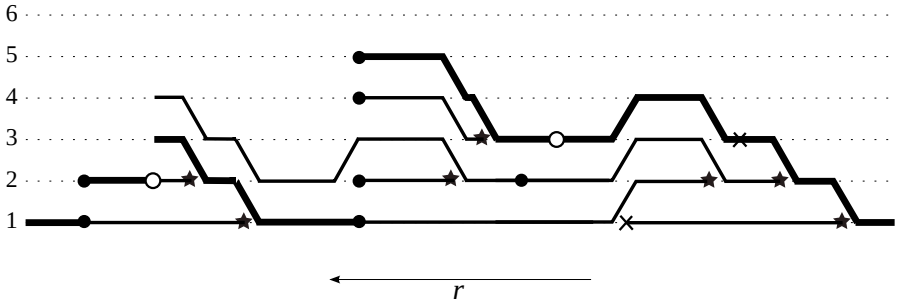


Abbildung 0.7: *Ausschnitt einer Realisierung eines pruned LD-Λ-ASG. Die immune Linie ist fett markiert.*

Ein Ausschnitt einer Realisierung eines pruned LD-Λ-ASG ist in Abbildung 0.7 gezeigt. Bei einem Verschmelzungsereignis werden alle teilnehmenden Linien mit einem fetten Punkt markiert. Alle punktierten Linien verschmelzen dann in die Linie mit Punkt, die das niedrigste Level besetzt.

Sei $(L_r)_{r\geq 0}$ der Linienzählprozess des pruned LD-Λ-ASG, der sich zur Zeit 0 im Gleichgewicht befinde, und $a_n := \mathbb{P}(L_0 \geq n)$, $n \geq 0$. Der Generator G_L von L ist dann gegeben durch

$$\begin{aligned} G_L g(\ell) = &\sum_{c=1}^{\ell-1} \binom{\ell}{\ell-c+1} \lambda_{\ell,\ell-c+1} [g(c) - g(\ell)] + \ell\sigma [g(\ell+1) - g(\ell)] \\ &+ (\ell-1)\theta\nu_1 [g(\ell-1) - g(\ell)] + \sum_{k=1}^{\ell-1} \theta\nu_0 [g(\ell-k) - g(\ell)]. \end{aligned} \tag{0.13}$$

Außerdem überträgt sich Theorem 0.2 und somit auch Theorem 0.3 auf den Λ-Fall.

Corollary 0.5 *Für die Wahrscheinlichkeit, dass die unsterbliche Linie im stationären pruned LD-Λ-ASG zur Zeit* 0 *vom Typ* 0 *ist, gegeben die Frequenz der Typ-*0*-Individuen zur Zeit* 0 *ist* $X_0 = x$*, erhält man*

$$h(x) = \sum_{n\geq 0} x(1-x)^n a_n. \tag{0.14}$$

Um ein Gleichungssystem für die Tailwahrscheinlichkeiten (a_n) zu bekommen, konstruieren wir einen Prozess D, der in *Siegmund-Dualität* zum Prozess L steht, d.h. $\mathbb{P}_\ell(L_u \geq d) = \mathbb{P}_d(D_u \leq \ell)$ für

alle $u \geq 0$, $\ell, d \in \mathbb{N}$ (vgl. [Sie76] oder Sektion 4.1 in der Übersichtsarbeit von Jansen und Kurt über Dualität [JK14]).
Denn dann können wir die Tailwahrscheinlichkeiten von L über Treffwahrscheinlichkeiten des dualen Prozesses D ausrechnen. Genauer gilt (siehe auch [CR84, Thm. 1]) $a_n = \mathbb{P}_{n+1}(\exists t \geq 0 : D_t = 1)$ für alle $n \geq 0$.

Einer Idee von Clifford und Sudbury folgend [CS85] beschreiben wir erst den Prozess L über sogenannte *Flights* und konstruieren dann den dualen Prozess D über die zugehörigen *dualen Flights*. Ein Flight f ist eine ordnungserhaltende Abbildung (also $f(k) \leq f(\ell)$ für alle $k \leq \ell$), die eine Menge in sich selber abbildet (hier $f : \mathbb{N} \cup \{\infty\} \to \mathbb{N} \cup \{\infty\}$). Wir setzen zusätzlich $f(\infty) = \infty$ und nehmen an, dass ∞ für den Prozess L unerreichbar ist.
Ein Flight, der zur Zeit r auftaucht, induziert einen Übergang nach $L_r = f(\ell)$, gegeben $L_{r-} = \ell$ (und zwar für alle $\ell \in \mathbb{N}$). Grafisch stellen wir einen Flight als eine Menge simultaner Pfeile dar (siehe Abbildung 0.8).

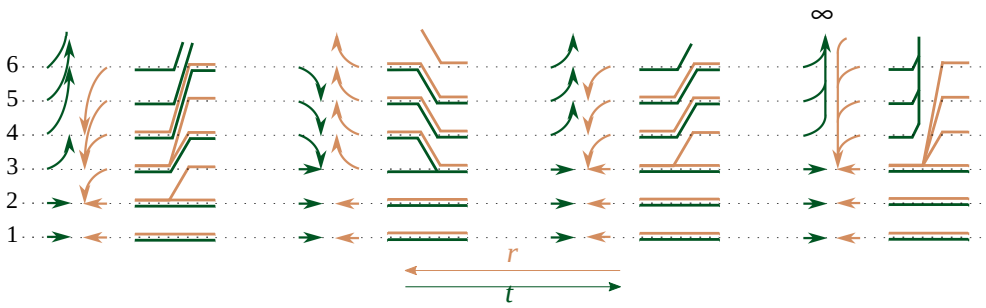


Abbildung 0.8: *Grafische Darstellung der vier Arten von Flights, die zur Beschreibung von L notwendig sind (hellbraune Pfeile), und deren duale Flights (dunkelgrüne Pfeile) zusammen mit den zugehörigen Pfadstücken von L (hellbraun) und D (dunkelgrün). Die Pfade von L folgen den Pfeilen in Richtung der Rückwärtszeit r (von rechts nach links) und von D in Richtung der Vorwärtszeit t (von links nach rechts).*

Die Dynamik des Prozesses L können wir nun durch Poissonsche Punktprozesse von vier Typen von Flights beschreiben, die Verschmelzungen, selektive Verzweigungen, Typ-1-Mutationen und Typ-0-Mutationen repräsentieren. Die Realisierung eines Pfades des Prozesses L können wir auf zweistufige Weise bekommen. Zuerst generieren wir eine Realisierung der Poissonschen Punktprozesse der Flights. Gegeben diese Realisierung lesen wir dann den L-Pfad direkt von den Flights ab: Wir fangen mit einem vorgegebenen Startlevel an und gehen dann in Rückwärtszeitrichtung induktiv von Flight zu Flight. Bei jedem Flight folgen wir jeweils dem Pfeil, der vom aktuellen Level des L-Pfades ausgeht.

Zu jedem Flight f definieren wir für $d \in \mathbb{N}$ den *dualen Flight* $\widehat{f}$ durch $\widehat{f}(d) = \min(f^{-1}(\{d, d+1, \ldots\}))$, mit der Konvention $\min(\varnothing) = \infty$ (vgl. auch Abbildung 0.8).

Zu einem aus einer Realisierung Φ der Poissonschen Punktprozesse der Flights abgelesenen L-Pfad definieren wir einen D-Pfad, der auf folgende Weise ebenfalls von Φ abgelesen werden kann: Zu der Realisierung Φ definieren wir eine duale Realisierung $\widehat{\Phi}$. Diese geht aus Φ durch die Abbildung

$(r,f) \mapsto (-r,\widehat{f}) =: (t,\widehat{f})$ hervor. Der D-Pfad wird nun von $\widehat{\Phi}$ abgelesen, indem man zur Zeit $t=0$ bei einem vorgegebenen Level startet und dann induktiv in Vorwärtszeitrichtung von dualem Flight zu dualem Flight geht und bei jedem dualen Flight jeweils dem Pfeil folgt, der vom aktuellen Level des D-Pfades ausgeht.

In Kapitel 4 zeigen wir dann, dass der auf diese Art und Weise definierte D-Prozess in (pfadweiser) Siegmund-Dualität zum L-Prozess steht. Den Generator G_D von D bekommt man dann über die Form der dualen Flights und die Übergangsraten von L,

$$\begin{aligned} G_D g(d) = &\sum_{d<c\leq\infty} \binom{c-1}{c-d+1} \lambda_{c,c-d+1}\left[g(c)-g(d)\right] + (d-1)\sigma\left[g(d-1)-g(d)\right] \\ &+ (d-1)\theta\nu_1\left[g(d+1)-g(d)\right] + (d-1)\theta\nu_0\left[g(\infty)-g(d)\right] , \\ &d \in \mathbb{N}, g : \mathbb{N}\cup\{\infty\} \to \mathbb{R}. \end{aligned} \tag{0.15}$$

Eine 'Zerlegung nach dem ersten Schritt' des Ereignisses $\{\exists t \geq 0 : D_t = 1\}$, gegeben $D_0 = n+1$, mit den Raten (0.15) von D führt dann schließlich zum Gleichungssystem (0.10).

0.3 Das Simulieren von Stichproben mit dem killed ASG

Kapitel 5 dieser Arbeit steht in keinem direkten Zusammenhang zu Kapitel 4. Es kann daher auch direkt nach Kapitel 3 gelesen werden[4].
In Kapitel 5 betrachten wir eine klassische Wright-Fisher-Diffusion $(X_t)_{t\geq 0}$ mit Mutation und Selektion, dessen Generator G_X gegeben ist durch

$$G_X g(x) = \frac{1}{2}x(1-x)g''(x) + [(1-x)\theta\nu_0 - x\theta\nu_1 + \sigma x(1-x)]g'(x), \quad g \in \mathscr{C}^2[0,1]. \tag{0.16}$$

Dabei verstehen wir X_0 wieder als die Frequenz der Typ-0-Individuen in einer Population zur Zeit 0 und nehmen an, dass die Verteilung von X_0 im Gleichgewicht ist. In diesem Kapitel sind wir interessiert an $\mathbb{E}\left[\binom{m}{\ell} X_0^\ell (1-X_0)^{m-\ell}\right]$, in Worten also an der Wahrscheinlichkeit, dass genau ℓ Individuen einer Stichprobe der Größe m aus der Population zur Zeit 0 vom Typ 0 sind.

Eine Rekursion dieser Wahrscheinlichkeit wurde bereits von Neuhauser und Krone über den ASG hergeleitet [KN97, Theorem 5.2]. Wir entwickeln eine neue Rekursion über eine modifizierte Version des ASG, den *killed ASG*, die auch gleichzeitig ein einfach implementierbarer Simulationsalgorithmus ist. Im Spezialfall $m=1$ entspricht unser Algorithmus einem bereits bekannten Resultat von Shiga [Shi88, Theorem 4.1], [SU86, Lemma 2.1].

Theorem 0.6 *Sei $m \in \mathbb{N}$ die Größe einer Stichprobe von Individuen, die aus einer stationär verteilten Wright-Fisher-Population mit Selektion und Mutation gezogen wird. Die Wahrscheinlichkeit, dass es in der Stichprobe genau ℓ Individuen vom Typ 0 gibt, $0 \leq \ell \leq m$,*

$$d_{m,\ell}\Big(\big((\{1\},\{2\},\ldots,\{m\}),0\big)\Big) := \mathbb{E}\left[\binom{m}{\ell} X_0^\ell (1-X_0)^{m-\ell}\right],$$

[4] Auch die Inhalte des Kapitels 3 sind nicht zwingend zum Verständnis von Kapitel 5 notwendig. Wir betrachten zwar den ASG in seiner geordneten Variante, wie sie in Kapitel 3 beschrieben wird, man kann hier jedoch jederzeit auch an den 'ungeordneten' ASG von Neuhauser und Krone [KN97, NK97] denken.

erfüllt das folgende System von Gleichungen,

$$
\begin{aligned}
&d_{m,\ell}\big((b_1,\dots,b_k),j\big)\\
&\quad= \frac{2\sigma}{k(k-1)+2k\theta+2k\sigma}\sum_{i=1}^{k} d_{m,\ell}\big((b_1,\dots,b_{i-1},b_i,b_i,b_{i+1},\dots,b_k),j\big)\\
&\qquad+\frac{2}{k(k-1)+2k\theta+2k\sigma}\sum_{i,p=1,i<p}^{k} d_{m,\ell}\big((b_1,\dots,b_{i-1},b_i\cup b_p,b_{i+1},\dots,b_{p-1},b_{p+1},\dots,b_k),j\big)\\
&\qquad+\frac{2\theta\nu_1}{k(k-1)+2k\theta+2k\sigma}\sum_{i=1}^{k} d_{m,\ell}\big((b_1,\dots,b_{i-1},b_{i+1},\dots,b_k),j\big)\\
&\qquad+\frac{2\theta\nu_0}{k(k-1)+2k\theta+2k\sigma}\sum_{i=1}^{k} d_{m,\ell}\big((b_1\setminus b_i,\dots,b_{i-1}\setminus b_i,b_{i+1}\setminus b_i,\dots,b_k\setminus b_i),j+\#b_i\big),\\
&\qquad\qquad\qquad\qquad j\in\{0,1,\dots,m\},\ b_\nu\subseteq\{1,\dots,m\}\ \forall\,\nu=1,\dots,k,\ k\geq 0,
\end{aligned}
$$

$$
d_{m,\ell}\big(\varnothing,\ell\big)=1,\quad d_{m,\ell}\big(\varnothing,j\big)=0 \text{ für alle } j\neq\ell. \tag{0.17}
$$

Den Beweis von Theorem 0.6 führt über die Konstruktion unseres '*killed ASG*'. Da der 'normale' ASG von Krone und Neuhauser [KN97, NK97] (im Gleichgewicht) alle potentiellen Vorfahren von allen m Individuen in der Stichprobe enthält, lohnt es sich mit ihm bzw. seiner geordneten Variante (vgl. Kapitel 3) anzufangen. Verfolgt man die potentiellen Ahnenlinien rückwärts in die Zeit, so trifft man nach f.s. endlicher Zeit bei jeder Linie auf eine erste Mutation. Diese Färbungen der Linien jeweils mit Typ 0 oder 1 wird dann wiederum vorwärts in der Zeit unter Respektierung der Verzweigungs- und Verschmelzungsereignisse entlang des Graphen bis zur Zeit 0 und somit bis zu den Individuen in der Stichprobe propagiert. Da die Färbung der Ahnenlinien jenseits der ersten Mutation keinen Einfluss mehr auf die Typen der m Individuen zur Zeit 0 hat, kann man sie bei dieser ersten Mutation abschneiden. Da auf diese Art und Weise Linien 'gekillt' werden, heißt der entstandene Graph *killed ASG*: Eine Ahnenlinie im ASG wird nach einer Typ-1-Mutation abgeschnitten, da sie nicht mehr dazu beiträgt, dass irgendein Individuum der Stichprobe zur Zeit 0 mit Typ 0 gefärbt wird. Eine Typ-0-Mutation hingegen sorgt sogar dafür, dass alle Individuen der Stichprobe, die die nach 0 mutierte Linie als potentiellen Vorfahren haben, zur Zeit 0 Typ 0 bekommen. Man kann also nicht nur die mutierte Linie selber abschneiden, sondern auch alle Linien, die potentielle Vorfahren der betroffenen Individuen der Stichprobe zur Zeit 0 sind (vgl. Abbildung 0.9). Um zu jedem Zeitpunkt $r\geq 0$ Buch darüber führen zu können, welche Linie zur Zeit r potentieller Vorfahre von welchen Individuen der Stichprobe zur Zeit 0 ist, versehen wir jede Linie mit einem zusätzlichen Label. Das Label einer Linie zur Zeit r enthält die Nummern aller Nachkommen in der Stichprobe zur Zeit 0. Außerdem definieren wir noch den Zählprozess $(J_r^m)_{r\geq 0}$. Die Zahl J_r^m gibt die Anzahl der Individuen aus der m-Stichprobe an, bei denen zur Zeit r bereits klar ist, dass sie vom Typ 0 sind. Sei τ der Zeitpunkt, zu dem die letzte noch vorhandene Linie durch eine Mutation gekillt wird. Dann ist J_τ^m die Anzahl der Individuen der m-Stichprobe zur Zeit 0, die Typ 0 tragen.

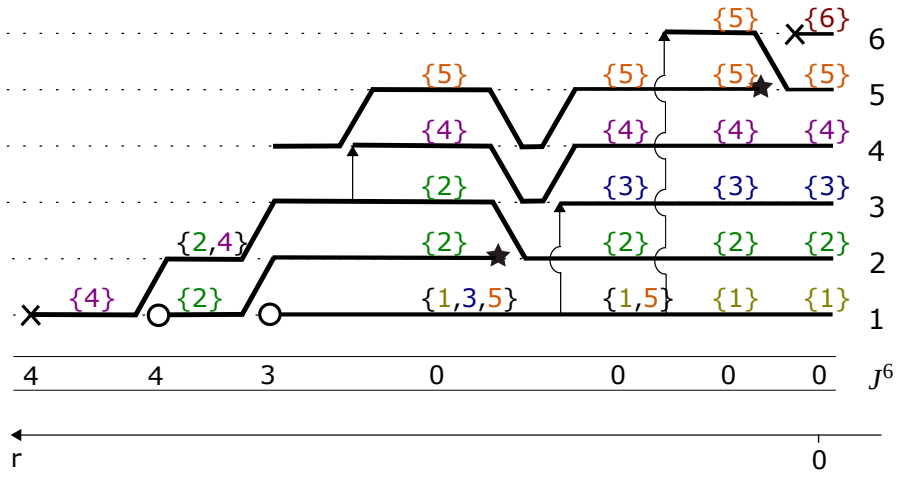


Abbildung 0.9: *Realisierung eines killed ASG, der zur Zeit* 0 *mit einer Stichprobe der Größe* $m = 6$ *startet. Die Startlabel auf der rechten Seite des Bildes sind* $\{1\},\{2\},\ldots,\{6\}$ *und es ist* $J_0^6 = 0$*. Bei einem durch einen Stern symbolisierten selektiven Ereignis verzweigt die Linie in sich selbst und einen Klon mit identischem Label. Bei einem Koaleszenzereignis trägt die resultierende Linie als Label die Vereinigung der Labels beider beteiligten Linien. Wird eine Linie von einer Typ-*1*-Mutation (Kreuzchen) getroffen, so wird sie gelöscht. Jede Typ-*0*-Mutation bewirkt die Erhöhung von* J^6 *um die Mächtigkeit ihres Labels. Gleichzeitig werden die in diesem Label enthaltenen Zahlen aus allen Labels gelöscht und Linien mit leeren Labels werden komplett weggeschnitten. In diesem Bild weden die Individuen* $1,3,5,2$ *der Stichprobe mit Typ* 0 *gefärbt. Daher ist* J^6 *auf der rechten Seite des Bildes gerade* 4*.*

Startet man einen killed ASG zur Zeit 0 mit m Linien, so führt eine ‘Zerlegung nach dem 1. Schritt’ des Ereignisses $\{J_\tau^m = \ell\}$ mit den Raten des ASG auf das System (0.17) von Gleichungen.

In Kapitel 5 dieser Arbeit zeigen wir dann noch Simulationsergebnisse (Abbildungen 5.3, 5.4 und 5.5) und zeigen eine Verbindung zum sogenannten *decision tree* im neu erschienenen Buch von Dawson und Greven [DG14] auf.

0.4 Der Typ des gemeinsamen Vorfahren und die virtuellen Linien in einer Wright-Fisher-Umgebung

In Kapitel 6 schlagen wir eine Brücke zwischen den Herangehensweisen von Fearnhead [Fea02] und Taylor [Tay07] mithilfe des in Kapitel 3 diskutierten pruned LD-ASG. Die Resultate der Kapitel 4 und 5 werden hierfür nicht benötigt.

Um die Verteilung des Typs des gemeinsamen Vorfahren in einer (standard-) Wright-Fisher-Population mit Mutation und Selektion im Gleichgewicht zur Zeit 0, gegeben $X_0 = x$, zu bekommen, untersucht Fearnhead [Fea02] einen Markovschen Prozess $(R_r, V_r)_{r\in\mathbb{R}}$, wobei R_r der Typ der unsterblichen Linie (der reellen Linie) zur Zeit r ist und V_r die Anzahl sogenannter virtueller Linien zur Zeit r. Eine virtuelle Linie ist bei Fearnhead stets vom Typ 1 und verhält sich (rückwärts in der Zeit) ansonsten

genau so wie eine Linie im gefärbten ASG: Sie kann mit anderen virtuellen Linien oder der reellen Linie, wenn diese gerade Typ 1 hat, verschmelzen, in sich selbst und eine neue Linie vom Typ 1 verzweigen, und von Typ 1 weg-mutieren. Passiert allerdings solch eine Mutation, so wird die Linie einfach abgeschnitten - denn virtuelle Linien dürfen ja nur vom Typ 1 sein. Fearnhead bekommt dann die Wahrscheinlichkeit $h(x)$, indem er die Gleichgewichtsverteilung des Prozesses (R,V) berechnet und über die Verteilung von V integriert: $h(x) = \sum_{v\geq 0} \mathbb{P}(R_0 = 0, V_0 = v)$.

Taylor hingegen betrachtet den Prozess $(X_r, R_r)_{r\in\mathbb{R}}$ des Typs R der reellen Linie in seiner Wright-Fisher Umgebung; X_r ist die Frequenz der Typ-0-Individuen zur Zeit r, mit Dynamik gegeben durch den Generator G_X in (0.2). Er bestimmt den Generator des gemeinsamen Prozesses (X,R), stellt $h(x)$ als bestimmte Treffwahrscheinlichkeit dar, löst ein Randwertproblem und bekommt so eine Lösung für $h(x)$.

In Kapitel 6 kombinieren wir beide Ansätze, indem wir die Dynamik des Tripelprozesses (X,R,V) rückwärts und auch vorwärts in der Zeit untersuchen. Durch dieses feinere Bild bekommt man einen weiteren Einblick in die Dynamik des Typs der unsterblichen Linie.
Wir bestimmen zuerst die Dynamik von (X,R,V) *rückwärts in der Zeit* und bekommen folgendes Resultat, welches wir dann auch intuitiv begründen.

Lemma 0.7 *Der Generator $\widehat{G}_{(X,R,V)}$ des Prozesses $(X_r,V_r,R_r)_{r\in\mathbb{R}}$ rückwärts in der Zeit ist gegeben durch*

$$\begin{aligned}\widehat{G}_{(X,R,V)}g(x,0,v) &= G_X g(x,0,v) + \frac{1-x}{x}\theta\nu_0\left[g(x,1,v)-g(x,0,v)\right]\\ &\quad + (v+1)\,\sigma(1-x)\left[g(x,0,v+1)-g(x,0,v)\right]\\ &\quad + \left[v\theta\nu_1 + \frac{1}{2}v(v-1)\right]\frac{1}{1-x}\left[g(x,0,v-1)-g(x,0,v)\right]\end{aligned} \tag{0.18}$$

$$\begin{aligned}\widehat{G}_{(X,R,V)}g(x,1,v) &= G_X g(x,1,v) + \frac{x}{1-x}\theta\nu_1\left[g(x,0,v)-g(x,1,v)\right]\\ &\quad + (v+1)\sigma(1-x)\left[g(x,1,v+1)-g(x,1,v)\right]\\ &\quad + \left[v\theta\nu_1 + \frac{1}{2}v(v+1)\right]\frac{1}{1-x}\left[g(x,1,v-1)-g(x,1,v)\right]\end{aligned} \tag{0.19}$$

für alle $g \in \mathscr{C}^2(0,1)\times\{0,1\}\times\mathbb{N}_0$.

Dieser Generator hat eine stationäre Verteilung, die man durch Disintegration nach der stationären Verteilung (siehe (0.3)) des Prozesses X folgendermaßen schreiben kann.

Theorem 0.8 *Die Dichte φ der stationären Verteilung des Prozesses (X,R,V) ist gegeben durch*

$$\begin{aligned}\varphi(x,0,v) &= w(x)\cdot a_v x(1-x)^v,\\ \varphi(x,1,v) &= w(x)\cdot(a_v - a_{v+1})(1-x)^{v+1},\end{aligned} \tag{0.20}$$

für $x \in [0,1]$, $v \in \mathbb{N}_0$, wobei die Koeffizienten $(a_v)_{v\in\mathbb{N}_0}$ bestimmt sind durch

$$\begin{aligned} &\left[\frac{1}{2}(v+1)+\sigma+\theta\right] a_v = \left[\frac{1}{2}(v+1)+\theta v_1\right] a_{v+1} + \sigma a_{v-1}, \quad v \geq 1, \\ &\textit{mit} \quad a_0 = 1 \ \textit{und} \ \lim_{v\to\infty} a_v = 0. \end{aligned} \tag{0.21}$$

(Die Koeffizienten (a_v) erweisen sich somit als die 'Fearnhead Koeffizienten' aus Theorem 0.3.)

Die Dynamik von (X,R,V) *vorwärts in der Zeit* bekommt man, indem man den Generator $\widehat{G}_{(X,R,V)}$ in der Zeit bezüglich der stationären Verteilung ψ des Prozesses (X,R,V) umkehrt. Der Vorwärtsgenerator $G_{(X,R,V)}$ muss also die Gleichung

$$\int g_1 (G_{(X,R,V)} g_2) d\psi = \int (\widehat{G}_{(X,R,V)} g_1) g_2 d\psi \tag{0.22}$$

für alle Testfunktionen $g_1, g_2 \in \mathscr{C}^2[0,1] \times \mathbb{N}_0 \times \{0,1\}$ erfüllen. Auf analytischem Weg erhält man das folgende Resultat.

Theorem 0.9 *Der Generator $G_{(X,R,V)}$ des Prozesses $(X_t, V_t, R_t)_{t\in\mathbb{R}}$ vorwärts in der Zeit ist gegeben durch*

$$\begin{aligned} G_{(X,R,V)} g(x,0,v) \;=\; & G_X g(x,0,v) + (1-x-vx) g'(x,0,v) \\ & + \theta v_1 \frac{a_v - a_{v+1}}{a_v} [g(x,1,v) - g(x,0,v)] \\ & + v\sigma \frac{a_{v-1}}{a_v} [g(x,0,v-1) - g(x,0,v)] \\ & + \left[(v+1)\theta v_1 + \frac{1}{2} v(v+1)\right] \frac{a_{v+1}}{a_v} [g(x,0,v+1) - g(x,0,v)] \end{aligned} \tag{0.23}$$

$$\begin{aligned} G_{(X,R,V)} g(x,1,v) \;=\; & G_X g(x,1,v) - x(v+1) g'(x,1,v) \\ & + \theta v_0 \frac{a_v}{a_v - a_{v+1}} [g(x,0,v) - g(x,1,v)] \\ & + v\sigma \frac{a_{v-1} - a_v}{a_v - a_{v+1}} [g(x,1,v-1) - g(x,1,v)] \\ & + \left[(v+1)\theta v_1 + \frac{1}{2}(v+1)(v+2)\right] \frac{a_{v+1} - a_{v+2}}{a_v - a_{v+1}} [g(x,1,v+1) - g(x,1,v)] \end{aligned} \tag{0.24}$$

für alle $g \in \mathscr{C}^2[0,1] \times \{0,1\} \times \mathbb{N}_0$.

Die Gleichungen (0.23) und (0.24) lassen sich auch anschaulich mit Hilfe des pruned LD-ASG erklären. Dazu betrachtet man für jede einzelne Übergangsrate, die eine Änderung von R oder V bewirkt, den Startzustand und den Endzustand des Prozesses. Diese Zustände bettet man dann jeweils in einen pruned LD-ASG ein. Das Ereignis $\{(R,V) = (1,v)\}$ impliziert $\{L = v+1\}$ und aus $\{(R,V) = (0,v)\}$ folgt $\{L > v\}$, wobei L der Linienzählprozess des pruned LD-ASG ist. Diese Überlegungen zusammen mit den Übergangsraten von L führen dann auf eine anschauliche Interpretation der Übergangsraten von (X,R,V).

Am Ende von Kapitel 6 beschäftigen wir uns dann noch mit der Dynamik eines Viererprozesses (L,X,R,V). Dessen stationäre Verteilung bekommt man wieder, indem man (R,V) in den pruned LD-ASG einbettet.

Lemma 0.10 *Die Dichte ϖ der stationären Verteilung des Viererprozesses (L,X,V,R) ist gegeben durch*

$$\varpi(\ell,x,r,v) = w(x)\left[\mathbf{I}_{\{r=0,v<\ell\}}\rho_\ell \cdot x(1-x)^v + \mathbf{I}_{\{r=1,v+1=\ell\}}\rho_\ell \cdot (1-x)^{v+1}\right], \tag{0.25}$$

wobei ρ die diskrete Dichte der stationären Verteilung von L ist, $\rho_\ell = \mathbb{P}(L_0 = \ell) = a_{\ell-1} - a_\ell$, $\ell \in \mathbb{N}$.

Es zeigt sich allerdings, dass der Viererprozess (L,X,R,V) kein Markovprozess ist.

0.5 Der Typ des gemeinsamen Vorfahren in einer Population endlicher Größe

Kapitel 7 dieser Arbeit beschäftigt sich mit einer Population endlicher Größe $N \in \mathbb{N}$, die einem Moran-Modell mit Zweiweg-Mutation und Selektion folgt, wie zu Beginn dieses Kapitels beschrieben. Die Anzahl K^N der Individuen vom Typ 0 ist dann ein Geburts- und Todesprozess mit den Raten (0.1).
Kapitel 7 baut auf den Kapiteln 2 und 3 auf und ist unabhängig von den Kapiteln 4 und 5. Es ist empfehlenswert vorher Kapitel 6 zu lesen, da es besonders in Abschnitt 7.3 Verweise auf Kapitel 6 gibt.

Um eine Brücke zwischen den Arbeiten von Taylor [Tay07] und Kluth, Hustedt und Baake [KHB13] (und auch zu Kapitel 6) zu schlagen, bestimmen wir die Übergangsraten (vorwärts und auch rückwärts in der Zeit) eines Doppelprozesses $(K_r^N, R_r^N)_{r\in\mathbb{R}}$, wobei R^N der Typ der unsterblichen Linien im diskreten Modell ist. Die Raten bekommt man direkt über die Raten der Pfeile (neutrale Reproduktion), Sternchenpfeile (selektive Reproduktion), Kreuzchen (Typ-1-Mutationen) und Kringel (Typ-0-Mutationen) im Moran-Modell. Es ergeben sich folgende Generatoren.

Lemma 0.11 *Der Generator $\widehat{G}_{(K^N,R^N)}$ des Prozesses (K^N,R^N) rückwärts in der Zeit ist gegeben durch*

$$\begin{aligned}
\widehat{G}_{(K^N,R^N)}g(k,0) = {} & \mu_{k+1}^N \frac{w^N(k+1)}{w^N(k)}\left[g(k+1,0) - g(k,0)\right] \\
& + \frac{1}{k}u_N v_0 (N-k+1)\frac{w^N(k-1)}{w^N(k)}\left[g(k-1,1) - g(k,0)\right] \\
& + \left[\lambda_{k-1}^N - \frac{1}{k}u_N v_0(N-k+1)\right]\frac{w^N(k-1)}{w^N(k)}\left[g(k-1,0) - g(k,0)\right],
\end{aligned}$$

$$\begin{aligned}
\widehat{G}_{(K^N,R^N)}g(k,1) = {} & \frac{1}{N-k}u_N v_1 (k+1)\frac{w^N(k+1)}{w^N(k)}\left[g(k+1,0) - g(k,1)\right] \\
& + \left[\mu_{k+1}^N - \frac{1}{N-k}u_N v_1(k+1)\right]\frac{w^N(k+1)}{w^N(k)}\left[g(k+1,1) - g(k,1)\right]
\end{aligned}$$

$$+\lambda_{k-1}^N \frac{w^N(k-1)}{w^N(k)}\left[g(k-1,1)-g(k,1)\right],$$

mit der Testfunktion $g:\mathbb{N}_0\times\{0,1\}\to\mathbb{R}$ *und* $k\in\{0,1,\ldots,N\}$.

Den Vorwärtsgenerator $G_{(K^N,R^N)}$ bekommt man wieder durch Zeitumkehr des Rückwärtsgenerators $\widehat{G}_{(K^N,R^N)}$ über die stationäre Verteilung mit der Dichte

$$\pi^N(k,r)=w^N(k)\left[\mathbf{1}_{\{r=0\}}h_k^N+\mathbf{1}_{\{r=1\}}(1-h_k^N)\right]. \tag{0.26}$$

Lemma 0.12 *Der Generator* $G_{(K^N,R^N)}$ *des Prozesses* (K^N,R^N) *vorwärts in der Zeit ist gegeben durch*

$$\begin{aligned}
G_{(K^N,R^N)}g(k,0) = {} & \left[\lambda_k^N-\frac{1}{k+1}u_N v_0(N-k)\right]\frac{h_{k+1}^N}{h_k^N}\left[g(k+1,0)-g(k,0)\right]\\
&+\frac{1}{N-k+1}u_N v_1 k\frac{1-h_{k-1}^N}{h_k^N}\left[g(k-1,1)-g(k,0)\right]\\
&+\mu_k^N\frac{h_{k-1}^N}{h_k^N}\left[g(k-1,0)-g(k,0)\right],
\end{aligned}$$

$$\begin{aligned}
G_{(K^N,R^N)}g(k,1) = {} & \frac{1}{k+1}u_N v_0(N-k)\frac{h_{k+1}^N}{1-h_k^N}\left[g(k+1,0)-g(k,1)\right],\\
&+\lambda_k^N\frac{1-h_{k+1}^N}{1-h_k^N}\left[g(k+1,1)-g(k,1)\right]\\
&+\left[\mu_k^N-\frac{1}{N-k+1}u_N v_1 k\right]\frac{1-h_{k-1}^N}{1-h_k^N}\left[g(k-1,1)-g(k,1)\right],
\end{aligned}$$

mit der Testfunktion g *und* $k\in\{0,1,\ldots,N\}$.

Dabei ist h_k^N, $k\in\{0,1,\ldots,N\}$, das diskrete Analogon zu $h(x)$, also die Wahrscheinlichkeit, dass der gemeinsame Vorfahre in einer Gleichgewichtssituation zu einer beliebigen aber festen Zeit vom Typ 0 ist, gegeben die Anzahl der Typ-0-Individuen zu dieser Zeit ist k. Wir zeigen dann, dass $(h_k^N)_{k=0,1,\ldots N}$ die Lösung des Gleichungssystems

$$\begin{aligned}
&h_k^N\left(\lambda_k^N+\mu_k^N\right)\\
&\quad= h_{k+1}^N\lambda_k^N+h_{k-1}^N\mu_k^N+\left(1-h_{k-1}^N\right)u_N v_1\frac{k}{N-k+1}-h_{k+1}^N u_N v_0\frac{N-k}{k+1},
\end{aligned} \tag{0.27}$$

mit den Randwerten $h_0^N=0$, $h_N^N=1$ ist. Dies wurde auch von Kluth, Hustedt und Baake bewiesen [KHB13, (21)], allerdings mit anderen Methoden.

Die Verbindung zu Taylor [Tay07] bauen wir auf, indem wir den Diffusionslimes der Generatoren $\widehat{G}_{(K^N,R^N)}$ und $G_{(K^N,R^N)}$ berechnen. Wir erhalten durch Übergang zum Grenzwert gerade die Generatoren $\widehat{G}_{(X,R)}$ und $G_{(X,R)}$ [Tay07, (4) und (15)].

Im letzten Abschnitt von Kapitel 7 skizzieren wir die Bestimmung der Dynamik eines diskreten Tripelprozesses (K^N, R^N, V^N). Dabei beschreibt V^N das diskrete Analogon zu V in Kapitel 6, nämlich die Anzahl virtueller Linien im Modell mit Gesamtpopulationsgröße N. Auf diese Weise lassen sich durch Übergang zum Diffusionslimes alternative Beweise für Lemma 0.7 und Theorem 0.9 erhalten.

1 Introduction

Population genetics plays an important role in biological and mathematical research. It provides a basis for understanding evolutionary processes and features both experimental and theoretical components. This thesis deals with the theoretical side. As the processes of evolution are very complex, it is almost impossible to model them mathematically in all details. Nevertheless, mathematical models are often helpful to understand (at least qualitatively) the interplay between different evolutionary factors. In addition, the models are building elements of fundamental research.

Today, mathematical population genetics research deals, amongst other topics, with the influence of random neutral reproduction, selective reproduction, recombination, migration, and mutation onto the genetic structure of a population; see e.g. [Dur08] or [Ewe04]. In this thesis, we concentrate on the interplay of random neutral reproduction, selective reproduction, and mutation. We consider a population with two types; the deleterious type 1 and the beneficial (selectively advantaged) type 0.

A basic model for analysing the influence of random reproduction in a population is the classical *Wright-Fisher model*, introduced in the 1920's and named after Sewall Wright and Ronald Fisher. It deals with a haploid population of finite and fixed size N where time is measured in terms of discrete generations. Reproduction is modelled in the following way: In each generation, each child chooses its single parent independently and uniformly at random among all N individuals in the previous generation. This model was generalised by Chris Cannings in the 1960's. In the *Cannings model*, the offspring vector per generation is not necessarily multinomially (as in the Wright-Fisher model) but still exchangeably distributed.
In 1958, Pat Moran provided a new model that goes with a continuous time scale [Mor58]. In the *Moran model*, each individual reproduces independently at a constant rate. Its offspring then replaces a randomly chosen individual from the population. Selection and mutation can be added in a natural way to all these models (see e.g. [Dur08, Ewe04]). In the Moran model, for example, Selection is visualised either by additional reproduction events only of individuals of beneficial type (fertility selection) or by additional death events only of individuals of deleterious type (viability selection).
When considering large populations, it may be convenient to pass to the (diffusion) limit $N \to \infty$. Then, with a convenient rescaling of time, the proportion of type 0 individuals converges in distribution to a *Wright-Fisher diffusion* [Dur08, Chapter 7.2]. There even exists a model with which one may (in the absence of selection) gain this convergence in the 'strong' sense (realisation by realisation). The *lookdown model*, introduced by Stephen Donnelly and Tom Kurtz, arises from the Moran model basically by a rearrangement of the lines [DK99b, DK99a].

So far, all of the above models describe the evolution of populations from the present into the future (or from the past to the present). Thus, they can be classified as *forward in time* models. When interested in genealogies, i.e. in the evolution of ancestral lines, one has to look *backward in time*.
In the absence of selection, the genealogy of a (yet untyped) random sample of individuals is modelled

by a *Kingman coalescent*, introduced by John Kingman in 1982 [Kin82]: Each pair of ancestral lines coalesces independently at a constant rate (say at rate 1). The types of the individuals in the sample are then determined by adding a type to the most recent common ancestor. Optionally, mutations may also be added to the lines (modelled by independent Poisson point processes). When selection comes into play, the Kingman coalescent has to be extended by additional branching events of ancestral lines. This branching and coalescing graph, the *ancestral selection graph (ASG)*, was introduced in 1997 by Stephen Krone and Claudia Neuhauser [KN97, NK97]. The ASG of a (yet untyped) sample contains all potential ancestors of this sample, it consists of so-called *virtual* and *real* branches. The true ancestors, the real branches, can only be resolved when adding types to the graph (see also [BB08]).

At each time, there exists a unique individual whose progeny will eventually take over in the population. Since eventually, at some later time, all individuals in the population are offspring of children of this single individual, it is called the *common ancestor*. In this thesis, we are especially interested in the stationary probability $h(x)$ that the common ancestor is of the beneficial type 0 at a given time, given the frequency of type-0 individuals at that time is x. In the case with selection, this probability is larger than in the case without selection. The intuition behind this fact is that individuals of the beneficial type have higher reproduction rates and therefore are more successful in the long run.
The common ancestor type distribution was first investigated in 2002 by Paul Fearnhead [Fea02]. His approach uses a pruned version of the ASG, the *common ancestor process* (R,V). $R \in \{0,1\}$ is the type of the single real line (thus the common ancestor) and $V \in \mathbb{N}_0$ a number of virtual lines that is needed in order for the process to be Markovian. In 2007, Jesse Taylor was able to gain the common ancestor type distribution using a different approach [Tay07]. By starting from the *structured coalescent* [BES04], he derived and solved a boundary value problem for (X,R), where R is again the type of the common ancestor and X the type-0 frequency in the entire population. Then, in 2013, the common ancestor type distribution was investigated by Sandra Kluth, Thiemo Hustedt, and Ellen Baake via a discrete approach [KHB13]. They worked all the way through in the discrete Moran model with population size N and passed to the diffusion limit $N \to \infty$ only in the very end.

In Chapters 3, 6, and 7, we build a bridge between these three approaches.
Chapter 3 gives a probabilistic approach to the common ancestor type distribution. We develop the *pruned lookdown ancestral selection graph (pruned LD-ASG)*, a particle picture that lies behind a recursion that is part of the formula determining $h(x)$. This way, we are able to give a probabilistic meaning to Taylor's and Fearnhead's results on the probability $h(x)$. Inspired by Donnelly's and Kurtz's idea to order the lines [DK99a], the pruned LD-ASG arises from the ASG by a rearrangement of the lines and additional pruning of lines upon mutation. Although our ordering is a different one, we borrow the name 'lookdown' from [DK99a]. In the classical ASG in the case with mutations, it is rather involved to find the line of the common ancestor. Namely one has to assign types to all ancestors, let the types propagate forward in time along the lines of the ASG with respecting the mutations, and resolve afterwards each branching event backward in time. In the pruned LD-ASG, the line of the common ancestor can be found quite easily. It is just, if such a line exists, the type-0 line that occupies the lowest level. If all lines are of type 1, it is the so-called *immune line*. Therefore, we believe that our pruned LD-ASG is interesting in its own right.
Chapter 6 links Fearnhead's common ancestor process (R,V) and Taylor's process (X,R). Namely, we define the triple process (X,R,V) and describe the evolution of the type-0 frequency X together

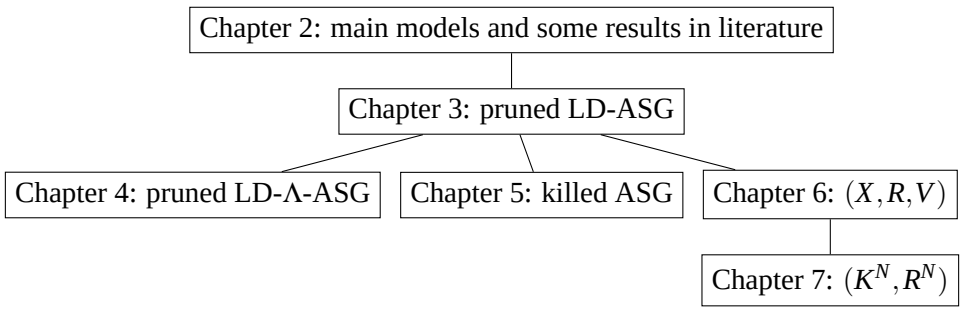
with the type R of the immortal line and the number V of Fearnhead's virtual lines. We determine the backward and forward in time generators of (X,R,V) and its stationary distribution.
Then, in Chapter 7, we start with a discrete approach that links [Tay07] and [KHB13]. We investigate the process (K^N,R^N), a discrete version of (X,R) for a population of finite size N. We recapitulate results by Kluth, Hustedt, and Baake [KHB13] on the probability h_k^N that the common ancestor is of beneficial type, given the number of type-0 individuals at that time is k. By taking the limit $N\to\infty$ for the backward and forward in time generators of the process (K^N,R^N) we regain Taylor's [Tay07] generators of the process (X,R).

While the pruned LD-ASG is introduced in Chapter 3 for classical Wright-Fisher populations only, we extend it to the case with heavy tailed offspring in Chapter 4.
In a classical Wright-Fisher population, at a reproduction event, only an infinitesimal size of the population is replaced by offspring of the single individual that reproduces. But it is also convenient to consider reproduction events where the offspring in a single reproduction event replaces a macroscopic fraction of the population. We are then in the more general setting of Λ-*Wright-Fisher processes*. They belong to the larger class of Λ-Fleming-Viot processes. While *Fleming-Viot processes* (named after Wendell H. Fleming and Michel Viot [FV79], see also [EK93]) allow for each individual to inherit a type that is chosen out of a a continuum of types (e.g. the type space can be the unit interval $[0,1]$), we again stick to the Wright-Fisher case in Chapter 4 and only allow for two types 0 and 1. The symbol 'Λ' in 'Λ-Wright-Fisher process' indicates that we are dealing with a general reproduction measure Λ that also models the case when a single individual produces offspring that replaces a macroscopic fraction of the population at a single birth event.
Backward in time, in the case without selection and mutation, the ancestral processes are so-called Λ-*coalescents*. They were introduced independently by Jim Pitman [Pit99], Serik Sagitov [Sag99], and Peter Donnelly and Tom Kurtz [DK99b] (see also [Ber09] for an introductory review). In comparison to Kingman's coalescent, they also include multiple mergers. In the case with selection, the situation becomes rather involved: For example, one has to include branching events to the coalescent and again deal with an ancestral graph of potential parents. Although not much is known on the Λ-coalescent with selection, some progress has been made recently (see e.g. [DK99a, EGT10, DGP12, Fou13, Gri14, BP15]).
We enlarge the collection of results. With the help of the the *pruned LD-*Λ*-ASG*, our extension of the pruned LD-ASG to the Λ-case, we determine the type distribution ($h(x)$, $x\in[0,1]$) of the common ancestor of a Λ-Wright-Fisher processes. In order to determine the stationary distribution of the line counting process of the pruned LD-Λ-ASG, we use Siegmund duality techniques (see e.g. [Sie76, CR84], or [JK14] for a survey the notion(s) of duality for Markov processes).
To this end, in the footsteps of Peter Clifford and Aidan Sudbury [CS85], we exploit the Siegmund dual process of the pruned LD-ASG. As it turns out, the latter process can be seen as a generalisation of the fixation line. The latter was introduced by Peter Pfaffelhuber and Anton Wakolbinger [PW06] for Kingman coalescents, and generalised to Λ-coalescents by Olivier Hénard [Hén15].

We restrict again to the classical Wright-Fisher case in Chapter 5. But we do not consider the immortal line any more; we investigate sampling probabilities for samples of size $m\in\mathbb{N}$. Namely, we think of sampling m individuals uniformly at random out of a stationary Wright-Fisher population with selection and mutation and ask for the probability of obtaining ℓ individuals of type 0 and $m-\ell$

of type 1. While this probability was computed already by Neuhauser and Krone [KN97, Theorem 5.2] in the form of a recursion that is rather difficult to handle, we develop a simulation algorithm via the so-called *killed ASG*. We also show simulation results. In addition, we show a link to Don Dawson's and Andreas Greven's *descision tree*, that is part of their recently published book [DG14] on Fleming-Viot processes.

The structure of this thesis is shortly visualised in following diagram. Each chapter is based on all chapters that are in direct line above this chapter in the tree diagram. For example, it is recommended to read Chapters 2, 3, and 6 before reading Chapter 7.



The thesis is organised as follows. In Chapter 2, we introduce the main models and the quantities that we are interested in. We review the Wright-Fisher model, the Moran model, the ancestral selection graph, the lookdown model, (Λ-) Wright-Fisher diffusions and the Λ-coalescent. We also present results on the common ancestor type distribution by Fearnhead [Fea02], Taylor [Tay07], and Kluth, Hustedt, and Baake [KHB13]. In Chapter 3, we introduce the pruned lookdown ancestral selection graph and develop a probabilistic meaning for the common ancestor type distribution in a Wright-Fisher population with two-way mutation and selection. This chapter is already published in [LKBW15]. We further extend these results to the case with heavy-tailed offspring distributions in Chapter 4. To gain the common ancestor type distribution for a Λ-Wright-Fisher process with mutation and selection, we exploit Siegmund duality. Chapter 4 is submitted for publication [BLW16]. Chapter 5 then deals with the probability of taking a sample of size $m \in \mathbb{N}$ with $\ell \leq m$ individuals of type 0. A simulation algorithm for this sampling probabilities is gained via a killed version of the ancestral selection graph. We build a bridge between the approaches on the common ancestor type distribution by Fearnhead [Fea02] and Taylor [Tay07] in Chapter 6. By using a discrete approach in Chapter 7, we extend this bridge towards Kluth, Hustedt, and Baake [KHB13].

2 Main models and some mathematical background

In *population genetics*, inter alia, the evolution of the frequency of alleles within a population is analysed. In this thesis, we will consider a haploid population[5] where each individual inherits one of two possible types. We will then analyse the evolution of the frequencies of the two alleles under genetic drift, mutation and selection.
Our main processes will be the *Wright-Fisher diffusion*, describing the frequency evolution *forward* in time, and a modification of the *ancestral selection graph*, describing the genealogy of a sample *backward* in time.

In this chapter we will introduce our main models and quantities, and review some results available in the recent literature. Some models will be reviewed again when they are needed in the different chapters.

2.1 Main models

2.1.1 Discrete Wright-Fisher model and Wright-Fisher diffusion

Let us consider a haploid population of fixed size $N \in \mathbb{N}$ with discrete generations. In generation 0, each individual gets one of two types, either 0 or 1. Then the Wright-Fisher model evolves (forward in time) from generation to generation as follows (compare also [Eth11, Chapter 2.1], or [Dur08, Chapter 1.2]). Independently in each generation and independently of the other individuals, each individual picks a uniformly chosen mother (with replacement) and copies her type. Then, for all $g \in \mathbb{N}$, the number of children in generation $g+1$ of an arbitrary but fixed individual in generation g is binomially distributed with parameters N and $1/N$.
The number of type-0 individuals in generation g is a Markov chain as well as a (bounded) martingale. As such, it converges to one of the two boundary points 0 and N for $g \to \infty$ a.s. Eventually one of the two types dies out and the other type fixates (i.e. takes over the entire population) a.s.
When considering large populations, it is convenient to renormalise the number of type-0 individuals and speed up time. This way, we think of N generations per time interval of unit length and let $\check{X}_t^{(N)}$, $t \in \{0, 1/N, 2/N, \ldots\}$ be the proportion of type-0 individuals in generation $g = Nt$. It is then easily calculated that

$$\mathbb{E}\left[\check{X}_{t+\Delta t}^{(N)} \,\middle|\, \check{X}_t^{(N)}\right] = \check{X}_t^{(N)} \quad \text{and} \quad \mathbb{E}\left[\left(\check{X}_{t+\Delta t}^{(N)} - \check{X}_t^{(N)}\right)^2 \,\middle|\, \check{X}_t^{(N)}\right] = \check{X}_t^{(N)}\left(1 - \check{X}_t^{(N)}\right) \cdot \Delta t \quad \text{a.s.}, \tag{2.1}$$

[5] Haploid individuals have only one parent. Diploid individuals have two copies of chromosomes from two different parents.

with $\Delta t = \frac{1}{N}$.

By taking the limit $N \to \infty$ in the discrete Wright-Fisher model, one arrives at a diffusion process with diffusion coefficient $x(1-x)$. Namely, when the initial proportion $\check{X}_0^{(N)}$ of type-0 individuals converges to some $x_0 \in [0,1]$, then $(\check{X}_t^{(N)})_{t\in\{0,1/N,\dots\}}$ converges to $(\check{X}_t)_{t\geq 0}$ as $N \to \infty$ where the Markov process $(\check{X}_t)$ has initial state $\check{X}_0 = x_0$ and generator

$$G_{\check{X}} g(x) = \frac{1}{2} x(1-x) g''(x) \tag{2.2}$$

with test function g in $\mathscr{C}^2[0,1]$, the set of all continuous and twice differentiable functions on $[0,1]$ (see e.g. [Dur08, Chapter 7.2]). The *Wright-Fisher diffusion* $(\check{X}_t)$ describes the evolution of the type-0 individuals in the population (in the absence of mutation and selection).

In fact, there is not only one model but a class of discrete models that converge to the Wright-Fisher diffusion. However, for the purpose of this thesis one may think of the Wright-Fisher diffusion as approximating either as a large population evolving according to the discrete Wright-Fisher mechanism with N generations per time unit or of a large population in a Moran model, which is described in the next paragraph.

2.1.2 Moran model

The Moran model again deals with a haploid population of fixed size $N \in \mathbb{N}$ but now in continuous time [Mor58] (compare also [Dur08, Chapter 1.5]). Therefore, generations may overlap but we think of an average of one generation per time interval of length one. Thus, each individual reproduces independently at rate $1/2$ and its single offspring replaces one uniformly chosen individual in the population (possibly its own parent).

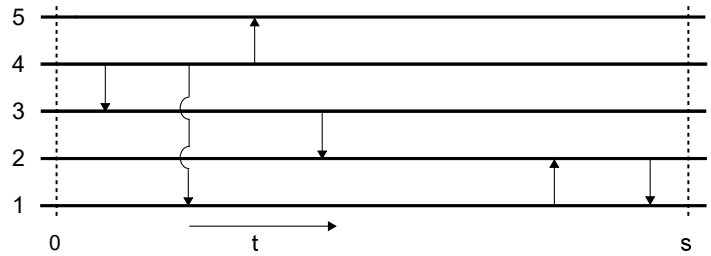


Figure 2.1: *Moran model with population size $N = 5$. The individuals are placed on levels and reproduction events are indicated by arrows. (Forward) time t runs from left ($t = 0$) to right ($t = s$).*

Graphically, this process can be realised by placing the N individuals on levels labelled with the numbers $1,\dots,N$. For example, given the proportion of type-0 individuals at time $t = 0$ is $X_0^N = k_0/N$, k_0 uniformly chosen levels at time $t = 0$ are coloured with type 0. The population then evolves forward

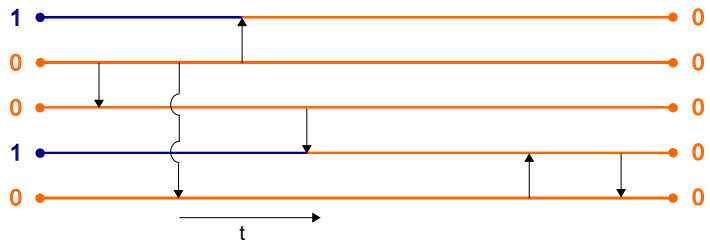


Figure 2.2: *Transportation of types in the Moran model. Types* 0 *and* 1 *are marked orange and blue, respectively.*

in time. A reproduction event is indicated by an arrow pointing from the location of the parent to the location of the child (compare also Figs. 2.1 and 2.2). The individual at the tip of the arrow dies and is replaced by a copy of the individual at the tail of the arrow. Each such arrow pointing from individual i to $j \neq i$, $i, j \in \mathbb{N}$, appears independently at rate $1/(2N)$. This way, the lines in the Moran model are exchangeable at all times $t \geq 0$, and type 0 eventually again either dies out or fixates in the population. In Fig. 2.2 type 1 dies out and type 0 fixates.
Then the birth rate at which the proportion of type-0 individuals is increased from $\frac{k}{N}$ to $\frac{k+1}{N}$ is given by $\frac{1}{2N} \cdot k \cdot (N-k)$ and the death rate at which it is decreased to $\frac{k-1}{N}$ is also given by $\frac{1}{2N} \cdot (N-k) \cdot k$.
The 'plain' Moran model describes the evolution of type frequencies in a population with reproduction events only. But it can be extended by including mutation and selection. To do so, we continue with the Moran model with finite population size N and take the diffusion limit $N \to \infty$ afterwards.

Here, we consider parent independent two-way *mutation*. Each individuals mutates independently to type 0 at rate $u_N v_0$ and to type 1 at rate $u_N v_1$, with $u_N \in [0, \infty)$, and $v_0, v_1 \in [0,1]$. This way, the overall mutation rate per individual is u_N and the probabilities for a mutation to be to type 0 is v_0 and to type 1 is v_1, $v_0 + v_1 = 1$. Silent mutations from type 0 to 0 and from type 1 to 1 are included. Graphically, mutations are visualised by independent Poisson point processes of 'circles' and 'crosses' at rates $u_N v_0$ and $u_N v_1$ per line (see Fig. 2.3). In the case of a strictly positive mutation rate to type i, $i \in \{0,1\}$, type i a.s. cannot die out in the population any more.

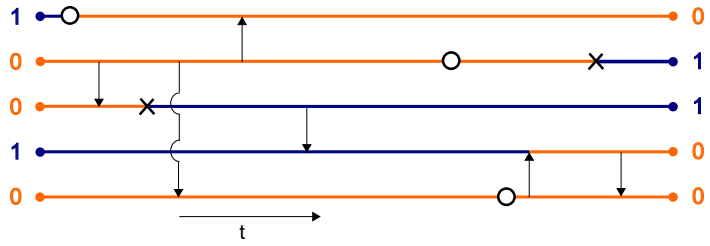


Figure 2.3: *Moran model with mutation. Mutations to type* 0 *are indicated by circles, mutations to type* 1 *by crosses.*

When speaking of selection in this thesis we always deal with *genic selection* (which is also named

directional selection). In addition, we focus on *fertility* selection only[6]. This is modelled in the following way. Let 0 be the beneficial and 1 the deleterious type such that type 0 reproduces faster than type 1. As the reproduction rate of a type-1 individual is only the neutral rate $1/2$, an individual of type 0 reproduces at rate $1/2+s_N$. This is visualised by including selective arrows into the model. A selective arrow from individual i to j, $i \neq j$, $i, j \in \mathbb{N}$ comes at rate s_N/N. In comparison to neutral ones, they have star-shaped heads (compare Fig. 2.4) and they can only be used by propagating type 0 from its tail to its tip. If the individuals at the tail of the selective arrow inherits type 1, no birth event happens and the arrow is just ignored.

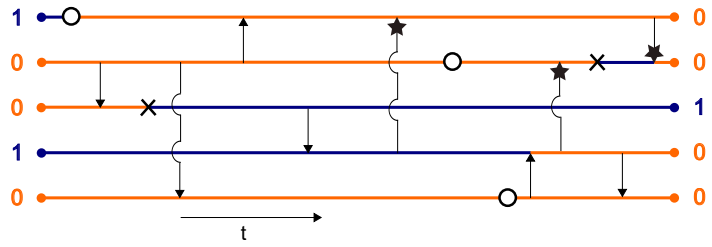


Figure 2.4: *Moran model with mutation and selection. Selective reproduction events are marked by arrows with star shaped heads. They can only be used by individuals of type* 0*.*

Mutation and selection preserves the exchangeability property of the N lines at all times. The number of type-0 individuals in the Moran model with mutation and selection is a birth and death process with birth rate λ_k^N and death rate μ_k^N, $k \in \{1,2,\ldots,N\}$,

$$\begin{aligned} \lambda_k^N &= \frac{1}{2N}k(N-k) + k(N-k)\cdot\frac{s_N}{N} \;+\; (N-k)\cdot u_N v_0, \\ \mu_k^N &= \frac{1}{2N}k(N-k) \;+\; k\cdot u_N v_1. \end{aligned} \tag{2.3}$$

Let X_t^N be the proportion of type-0 individuals at time $t \in [0,\infty]$ in the Moran model with mutation and selection with population size N.

To analyse large populations, we want to take the diffusion limit X of $X^N := (X_t^N)_{t\geq 0}$. Therefore, time is speeded up by a factor N such that we again have an average of N generations per time interval of unit length. The corresponding time scale is denoted by 'evolutionary time'. Neutral and selective reproduction arrows can then be seen at rates $1/2$ and s_N per ordered pair of lines, respectively.
To take the limit $N \to \infty$, we assume that the proportion of type-0 individuals at time 0, X_0^N, converges to $x_0 \in [0,1]$. In addition, let

$$Nu_N \to \nu \quad \text{and} \quad Ns_N \to \sigma, \quad \text{as } N \to \infty.$$

[6]Speaking of *fertility selection* means that individuals of beneficial type have a higher reproduction rate than those of deleterious type. In return, *viability selection* means that individuals of deleterious type die at a higher rate than those of beneficial type.

Then $(X_t^N)_{t\geq 0}$ converges to a Wright-Fisher diffusion with mutation and selection, $X := (X_t)_{t\geq 0}$, with generator G_X given by

$$G_X g(x) = \frac{1}{2}x(1-x)g''(x) + [(1-x)\theta\nu_0 - x\theta\nu_1 + \sigma x(1-x)]g'(x), \quad g \in \mathscr{C}^2[0,1], \tag{2.4}$$

compare also [Dur08, Chapter 7.2], or [KHB13, Section 2][7]. There exists a stationary probability measure for the generator G_X, Wright's density, which is given by [Dur08, (7.28)]

$$w(x) = c_w \cdot (1-x)^{2\theta\nu_1 - 1} x^{2\theta\nu_0 - 1} \cdot \exp\{2\sigma x\}, \tag{2.5}$$

with the normalising constant $c_w = \left[\int_0^1 (1-x)^{2\theta\nu_1-1} x^{2\theta\nu_0-1} \cdot \exp\{2\sigma x\} dx\right]^{-1}$. In other words, the frequency of type-0 individuals in a stationary population evolving according to the Wright-Fisher generator with mutation and selection (2.4) has density $w(x)$ (examples are shown in Fig. 2.5).

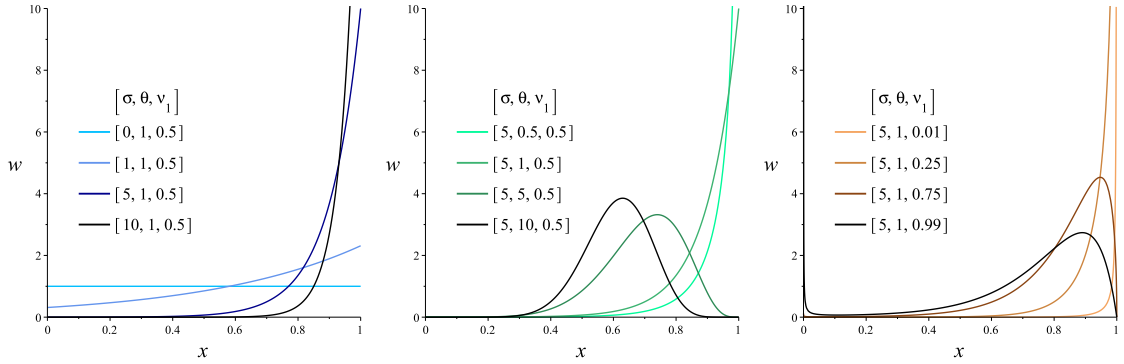


Figure 2.5: *Stationary density w of a Wright-Fisher process with mutation and selection given by* (2.5) *for different combinations of the parameters* σ, θ, *and* ν.

2.1.3 Neutral Genealogies: Kingman's Coalescent

So far, the evolution of type frequencies was modelled *forward* in time. But we also want to analyse genealogies in the Moran model (and in particular in its diffusion limit with infinite population size). Genealogical processes evolve *backward* in time. To differentiate between the two, we use throughout this thesis the variable t for forward and r for backward time. When illustrating two time scales in one single picture, we insinuate $r = -t$.

[7] Note that there is a difference of a factor $1/2$ in the scaling of the diffusion term in (2.4) and [Tay07, BLW16] in comparison to [KN97, Fea02, KHB13, KB13, LKBW15]. This is because the four last mentioned papers use the diffusion part of the Wright-Fisher generator without the factor 1/2. This corresponds to a pair coalescence rate of 2, while in this thesis we assume pair coalescence rate 1 in the diffusion limit.

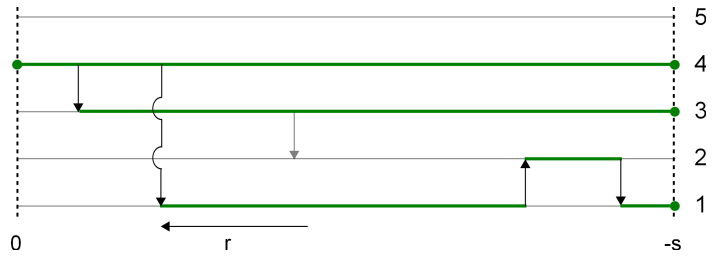


Figure 2.6: *Kingman's coalescent. The corresponding Moran model is shown in Fig. 2.1). (Backward) time r runs from right to left. The genealogy of the individuals on levels* 1, 3, *and* 4 *between times $r = -s$ and $r = 0$ is shown (green lines).*

Let us start with the neutral Moran model (i.e. the Moran model without selection) and without mutation. We want to consider the genealogy of a sample of n individuals (see Fig. 2.6) on the evolutionary time scale. This genealogical process was first investigated by Kingman in 1982 [Kin82]. It evolves backward in time and is driven by the reproduction arrows. As the rates of Poisson point processes are the same forward and backward in time, also backward in time each arrow pointing from individual i to j (and also from j to i), $i, j \in \mathbb{N}$, $i \neq j$, appears at rate $1/2$. Thus, backward in time, each unordered pair of individuals (i, j) independently coalesces at rate 1 into a common parent. In fact, the genealogical process of a sample of n individuals, denoted by *Kingman coalescent*, is a pure coalescing process. The Kingman-n-coalescent starts with n lines (or individuals or blocks) and eventually gets absorbed in only one remaining line. Its line counting process K^0 has the generator G_{K^0},

$$G_{K^0}g(n) = \binom{n}{2}\big[g(n-1) - g(n)\big], \quad g : \mathbb{N} \to \mathbb{R}, \tag{2.6}$$

and the expected time till all n lines have merged into one common ancestor is $2[1 - \frac{1}{n}]$, thus finite.

In the case without mutations, the type of each ancestor is directly handed over to all its offspring. Therefore, to determine the distribution of the types of a sample of n randomly chosen individuals taken at time $t = s$ from a population evolving according to a Wright-Fisher diffusion one may also start a Kingman-n-coalescent at time $r = -s$ with $K^0_{r=-s} = n$ individuals, let it evolve backward in time, and then colour its random remaining $K^0_{r=0}$ lines at time $r = t = 0$ with types 0 and 1 according to a Bernoulli experiment with success probability $\check{X}_0 = x_0$, the frequency of type-0 individuals at time 0.
Indeed, these heuristics are formalised in the well-known moment duality between the Wright-Fisher diffusion and the Kingman coalescent

$$\mathbb{E}\left[\check{X}_s^n \mid \check{X}_0 = x_0\right] = \mathbb{E}\left[x_0^{K_0^0} \mid K^0_{-s} = n\right], \tag{2.7}$$

or equivalently

$$\mathbb{E}\left[(1-\check{X}_s)^n \mid \check{X}_0 = x_0\right] = \mathbb{E}\left[(1-x_0)^{K_0^0} \mid K^0_{-s} = n\right]. \tag{2.8}$$

All n individuals are of type 0 (or 1) if and only if all ancestors are of type 0 (1).

It is then straight forward to add mutations to the Kingman coalescent. In the Moran model, mutations come at rates $\theta\nu_0$ and $\theta\nu_1$ independently per line. Thus, backward in time, these Poisson point processes run independently on the lines of the Kingman coalescent.
Since there are mutations on the ancestral lines, the type of one individual in a sample and the type of its ancestor may differ. Therefore the moment duality (2.8) becomes more difficult. To determine the type of an individual in a Wright-Fisher population at time $t = s$, given the type of its ancestor at time $t = 0$, one has to take a closer look at the whole ancestral line between times s and 0. If the line is not affected by any mutation, the type of the individual in the sample is exactly the type of the ancestor. If there is at least one mutation, then the type of the individual in the sample is determined by the most recent mutation before time s.
More formally, let $K^{0,cut}$ be a modified version of the Kingman coalescent. To be more precise, in addition to the coalescing dynamics, a pruning procedure is added: after each deleterious mutation, the affected line is deleted from the graph. Then $K^{0,cut}$ has generator

$$G_{K^{0,cut}}g(n) = \left[\binom{n}{2} + n\theta\nu_1\right]\big[g(n-1) - g(n)\big], \quad g : \mathbb{N} \to \mathbb{R}, \tag{2.9}$$

and the duality Equation (2.8) can be adapted to the case with mutations (compare [Shi88, Theorem 4.1] in the case without mutation, without migration, and with only one colony),

$$\mathbb{E}\left[(1-X_s)^n \mid X_0 = x_0\right] = \mathbb{E}\left[(1-x_0)^{K_0^{0,cut}} \cdot \exp\left\{-\theta\nu_0 \int_{-s}^{0} K_u^{0,cut}\, du\right\} \middle| K_{-s}^{0,cut} = n\right], \tag{2.10}$$

where the dynamics of X are determined by the generator (2.4) with $\sigma = 0$ and $\theta, \nu_0, \nu_1 \geq 0$.
Indeed, the left-hand side of (2.10) is the probability that n randomly chosen individuals at time $t = s$ in a Wright-Fisher population are of type 1, given the frequency of type-0 individuals at time 0 is x_0. On the right-hand side of (2.10), the term $\mathbb{E}\left[(1-x_0)^{K_0^{0,cut}} | K_{-s}^{0,cut} = n\right]$ is the probability that all remaining (and therefore not deleted) $K_0^{0,cut}$ ancestors at backward time $r = 0$ of a Kingman-n-coalescent started at time $r = -s$ are of type 0. The remaining ancestral lines are those lines that are not already assigned type 1 by a deleterious mutation between times $r = -s$ and $r = 0$. In addition, the remaining ancestors can only push their deleterious types through to the n sampled individuals if no mutation to type 0 can be found anywhere on the pruned Kingman graph. Given the sum of the lengths of all lines in the pruned graph is ℓ, this has probability $\exp\{-\theta\nu_0\ell\}$, which completes the explanation of the right-hand side of (2.10).

When selection is included, the backward in time picture which contains all ancestors of a sample becomes rather involved. It is briefly reviewed in the following section.

2.1.4 Genealogies with selection: The ancestral selection graph

In the Moran model, forward in time, selection is represented by selective arrows which can only be used for reproduction events of type-0 individuals. Thus, in the untyped case, at each selective event, the child at the tip of the arrow (also named *descendant*) has two potential parents. The line at the

tail of the arrow is denoted *incoming branch* and the line hit by the tip of the arrow is the *continuing branch*. When types are assigned to the lines, the 'pecking' order is the following: If the incoming branch is of type 0, it is the parental one. If it is of type 1, the continuing branch is the true parent of the selective event (see also Fig. 2.7). Branches that are true parents are denoted *real* branches, and potential ancestral branches that are not parental are called *virtual* branches [KN97].

Figure 2.7: *Selective reproduction events. At each of the four pictures the incoming branch is drawn below the continuing branch. Real branches (true parents and descendants) are marked bold, slim lines are virtual lines.*

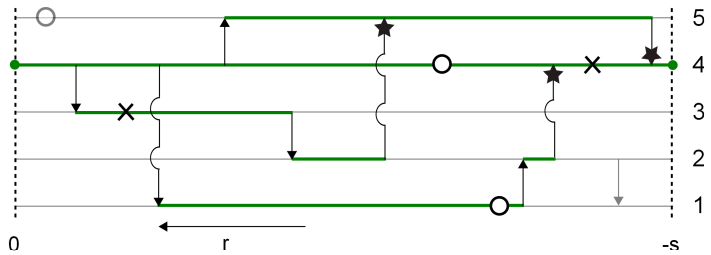


Figure 2.8: *Ancestral selection graph. The corresponding Moran model is shown in Fig. 2.4. All potential ancestors of the single individual on level* 1 *between times* $r = -s$ *and* $r = 0$ *are shown (green lines).*

When keeping track of all (potential) ancestors of a sample of size n in a Wright-Fisher population with selection, the Kingman coalescent has to be extended by the selective events (compare Fig. 2.8). Thus, backward in time, in addition to the coalescing structure of the neutral case, branching events come into play. Note that, in the discrete model with population size N, on the genealogical time scale, each line is hit by a star-shaped arrow that indicates a potential selective event at rate $(N-1)\cdot s_N/N$. In the diffusion limit, this rate converges to σ on the evolutionary time scale. But, in a sample of size n, on the genealogical time scale in the finite-size model, selective arrows are interchanged between any two individuals in the sample at rate $n(n-1)\cdot s_N/N$. On the evolutionary time scale, the diffusion limit of this rate is just 0. Thus, a.s. selective arrows are not interchanged between individuals among the sample but only hit sampled individuals from the outside. Quantitatively, each line in the ancestral graph branches independently at rate σ. This way, a branching and coalescing graph arises. This graph is denoted *ancestral selection graph* or *ASG* and was introduced by Neuhauser and Krone in 1997 [KN97, NK97]. Its line counting process K has the generator

$$G_K\, g(n) = \binom{n}{2}\big[g(n-1)-g(n)\big] + n\sigma\big[g(n+1)-g(n)\big], \quad g:\mathbb{N}\to\mathbb{R}. \tag{2.11}$$

As the coalescence rate of K is quadratic and the branching rate is only linear, the process a.s. does not explode and we have $K_r \in \mathbb{N}$ for all $r \geq 0$. In addition, for $\sigma > 0$, the ASG is recurrent. It has a stationary distribution, the Poisson distribution on $\mathbb{N}$, i.e. $\overline{g}(n) := \mathbb{P}(K_r = n) = \sigma^n/[n!(exp(\sigma)-1)]$, $n \geq 1$ (compare [PP13]).

Without mutations, the duality result (2.8) for the Kingman coalescent holds true also for the ASG: $\mathbb{E}\left[(1-X_s)^n \mid X_0 = x_0\right] = \mathbb{E}\left[(1-x_0)^{K_0} \mid K_{-s} = n\right]$ (see e.g. [Man09, Theorem 2.1]). At a branching event, the descendant is of type 1 if and only if incoming and continuing branch are both of type 1. Note that (2.7) is not valid in the ASG, since the descendant also has type 0 if only one of the two branches is of type 0 and one of type 1.

Mutations are added to the ASG in the same way as to the Kingman coalescent: They come at rates $\theta\nu_0$ and $\theta\nu_1$ independently on the lines. A modified duality for the ASG with mutations may again be gained in the same way as for the Kingman coalescent: delete each line after a deleterious mutation to get a pruned version K^{cut} of the ASG,

$$G_{K^{cut}}g(n) = \left[\binom{n}{2} + n\theta\nu_1\right]\left[g(n-1)-g(n)\right] + n\sigma\left[g(n+1)-g(n)\right], \quad g : \mathbb{N} \to \mathbb{R}. \tag{2.12}$$

Then again all n individuals in the sample are of type 1 if all remaining ancestors after the pruning are of type 1 and no mutation to type 0 appears on the pruned graph (compare [Man09, Eq. (2.23)] or [Shi88, Theorem 4.1] specialised to the case with only one colony).

$$\mathbb{E}\left[(1-X_s)^n \mid X_0 = x_0\right] = \mathbb{E}\left[(1-x_0)^{K_0^{cut}} \cdot \exp\left\{-\theta\nu_0 \int_{-s}^{0} K_u^{cut}\,du\right\} \middle| K_{-s}^{cut} = n\right]. \tag{2.13}$$

The connection between the Wright-Fisher diffusion (with mutation and selection) and the modified process K^{cut} of the ASG is a useful tool to analyse types of individuals of a sample. However, Equation (2.13) only allows for the probability that all individuals are of type 1.
Given the frequency $X_0 = x_0$ of type 0 in a Wright-Fisher population with mutation and selection at time $t = 0$, the types of n randomly chosen individuals at time $t = s$ can be gained in distribution using the ASG with mutations in the following way. First, start with $K_{-s} = n$ lines and generate their potential ancestry from time $r = -s$ up to $r = 0$ backward in time using the ASG-dynamics (2.11). Then, given a realisation of the ASG, add mutations to the lines using independent Poisson point processes. Add types to all K_0 potential ancestors according to a Bernoulli experiment with success probability x_0. Then propagate the types forward in time from time $t = 0$ ($r = 0$) to time $t = s$ ($r = -s$).

In an equilibrium situation, let $p(\ell, m-\ell)$ be the probability that a sample of m individuals in a Wright-Fisher population carries ℓ individuals of type 0 and $m-\ell$ individuals of type 1, $m \geq 1$, $\ell \leq m$,

$$p(\ell, m-\ell) = \frac{m!}{\ell!(m-\ell)!}\mathbb{E}\left[X^\ell(1-X)^{m-\ell}\right] \tag{2.14}$$

Krone and Neuhauser have shown [KN97, Thm. 5.2] that the sampling probabilities $p(m, m-\ell)$ then satisfy the recursion

$$p(\ell, m-\ell) = \frac{2\theta}{m+2\theta+2\sigma-1}\left[\frac{\ell+1}{m}p(\ell+1, m-\ell-1) + \frac{m-\ell+1}{m}p(\ell-1, m-\ell+1)\right]$$

$$+\frac{m-1}{m+2\theta+2\sigma-1}\left[\frac{\ell-1}{m-1}p(\ell-1,m-\ell)+\frac{m-\ell-1}{m-1}p(\ell,m-\ell-1)\right]$$
$$+\frac{2\sigma}{m+2\theta+2\sigma-1}\left[\frac{\ell(\ell+1)}{m(m+1)}p(\ell+1,m-\ell)+\frac{(m-\ell+1)(m-\ell)}{m(m+1)}p(\ell,m-\ell+1)\right.$$
$$\left.+2\frac{\ell(m-\ell+1)}{m(m+1)}p(\ell,m-\ell+1)\right], \tag{2.15}$$

with $p(k,j) := 0$ for $k<0$ or $j<0$. The probabilistic proof[8] of (2.15) uses the structure of the ASG [KN97, page 230] by using a 'first step decomposition' with respect to the most recent event back in time. Let us briefly review the intuition behind the first term in (2.15). Given there are m lines in the ASG at time $t=0$, the (forward in time) rates for mutation, branching, and coalescence are $m\theta$, $m\sigma$, and $m(m-1)/2$. Thus, the probability that the first event back in time before time $t=0$ is a mutation is given by $2\theta/(m+2\theta+2\sigma-1)$. If the first event back in time was a mutation to type 1 at time $t=-\tau$, there must have been $\ell+1$ individuals of type 0 and $m-\ell-1$ of type 1 at time $(-\tau)-$, which by stationarity has the probability $p(\ell+1,m-\ell-1)$. In addition, one of the type-0 individuals has to be chosen for the mutation at probability $(\ell+1)/m$. Taking factors together yields the first term in (2.15). The other terms can be derived by drawing analogous thoughts.

In Chapter 5, we give an alternative representation of sampling probabilities together with some simulation results.

When determining the types of a sample with the help of the ASG, one is in need of the shape of the whole graph in the time interval of length s together with the types. This may get really involved for large s. In Chapter 3, a refined construction of the ASG (the pruned LD-ASG) is presented. It allows for determining the type distribution of the ancestor of one individual or of a sample of arbitrary size for large times s without knowing the shape of the whole graph. Since our construction of the LD-ASG is inspired by an ordering of the lines as it is done in the so-called lookdown model, some aspects of the latter model are explained in the following section.

2.1.5 Lookdown model

The *lookdown model* was introduced by Donnelly and Kurtz [DK99b, DK99a]. Similar to the Moran model, it also is a particle picture which can be used for describing the evolution of type frequencies in a Wright-Fisher population[9]. Let us start with a population of fixed finite size $N \in \mathbb{N}$ in the absence of selection. Again, the individuals are identified with levels $1,\dots,N$. But now we start in the evolutionary time scale right from the beginning (not in the genealogical time scale as in Section 2.1.2). A reproduction event is modelled as follows: Two levels are selected uniformly at random without replacement; each (unordered) pair of levels independently at rate 1. Then the individual at the lower of these two levels ℓ_{low} gives birth to a child which is placed at the larger level ℓ_{high}.

[8] Since the scalings of the pair coalescence rate in [KN97] and this thesis differ by a factor 2, this factor has to be taken into account again when looking-up the proof of recursion (2.15) in [KN97, Thm. 5.2].

[9] Note that the model by Donnelly and Kurtz is much more general, but not all aspects are needed for the purpose of this thesis.

In comparison to the Moran model (without selection), reproduction arrows can only point in one direction in the lookdown model (without selection): upwards with tail at ℓ_{low} and tip at ℓ_{high}. The higher level 'looks down' at the lower level and adopts its type. In order to keep the population size constant, one individual has to die in the lookdown model as well. This is not the individual at level ℓ_{high} but the individual at level N. At the same time, all individuals at levels $k \geq \ell_{high}$ are shifted one level upwards to $k+1$ (compare Fig. 2.9). The Poisson point processes of reproduction arrows from i to j come at rate 1, independently for each ordered pair $i, j \in \mathbb{N}$, $i < j$. Therefore, a line on level ℓ is pushed to level $\ell+1$ at rate $\ell(\ell-1)/2$.

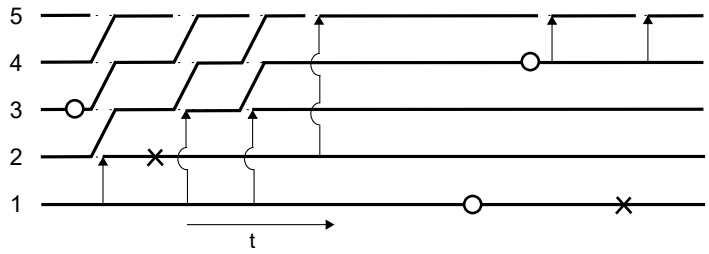


Figure 2.9: *Lookdown model with population size $N = 5$. Reproduction arrows are always pointing from lower to higher levels.*

Mutations are modelled the same way as in the Moran model: Mutations to types 0 and 1 come as independent Poisson point processes at rates $\theta\nu_0$ and $\theta\nu_1$ independently per level.

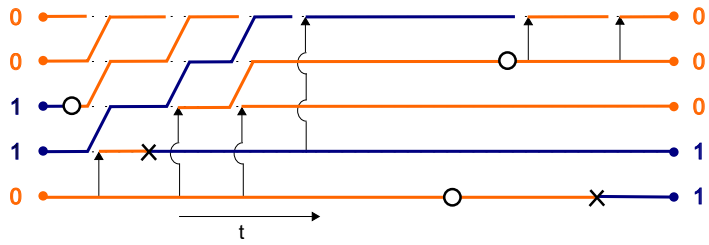


Figure 2.10: *Transportation of types in the lookdown model. Since the lines in this picture can be gained by a rearrangement of the lines in the corresponding Moran model (Fig. 2.3), also the genealogies agree.*

Donnelly and Kurtz [DK99b, Thm. 1.1] proved that if the types are assigned to the N lines at time 0 in an exchangeable way, then the lines stay exchangeable for all times $t \geq 0$. In addition, the proportion of type-0 individuals in the population in the lookdown model has the same distribution as the type-0 proportion in the Moran model, given the initial frequencies agree. Indeed, intuitively, the realisation of the lookdown model can be gained from a realisation of the Moran model by reordering the lines (compare Figs. 2.3 and 2.10). The individual which dies first is placed on level N, the individuals

with the second shortest lifetime on level $N-1, \ldots$, the individual whose descendants have the longest lifetime (and therefore live forever) on level 1.

Since the line on level 1 in the lookdown model without selection is never hit by any arrow, it lives forever. As it is the only line that never dies and whose offspring fixates in the population at some future time, it is named *immortal line* or *line of the common ancestor*.
Indeed, all lines at higher levels are affected at least by arrows originating from level 1. In fact, the higher the occupied level of a line, the faster this line is pushed to even higher levels until it dies when pushed above level N. The distribution of the time until a line on level ℓ dies at level N is given by $\mathscr{E}(\ell)+\mathscr{E}(\ell+1)+\ldots+\mathscr{E}(N)$, where $\mathscr{E}(j)$ is an exponential random variable with parameter j. This time is finite a.s. for all $\ell \in \{2,3,\ldots,N\}$ (and for all $2 \leq N \leq \infty$) with expectation $2(1/(\ell-1)-1/N)$.

The immortal line also exists in the Moran model. But it is not located on one distinguished level and its location changes throughout times, always depending on the future. The advantage in the lookdown model (with mutations) that the immortal line is always located on level 1 allows for an easy study of the type of the common ancestor of a future population:
Let, for some fixed time s, $h(x)$ be the probability that the immortal line is of type 0, given the proportion of type-0 individuals at time s is x. Then $h(x)=x$, independent of the time s and the population size N. Indeed, since the lines in the lookdown model are exchangeable at all times, each line (including the immortal line) has the same probability of being of type 0.
In an equilibrium situation, the probability that the immortal line is of type 0 is just $\frac{\theta\nu_0}{\theta\nu_0+\theta\nu_1}=\nu_0$, the probability that the most recent mutation on the ancestral line was to type 0.

Another advantage of the lookdown model in comparison to the Moran model (in the case without selection) is its consistency when adding individuals. The ordering of the lines insures that the genealogy of the first n individuals is not affected by individual $n+1$. All arrows with tips at most at level n have tails below level n. Therefore, when increasing the population size from N to $N+1$, the genealogy of the first N individuals does agree in the pictures with N and with $N+1$ lines not only in distribution but realisation by realisation.
To arrive again at a diffusion limit, one may take the limit $N \to \infty$ in the lookdown model with mutation (this is done e.g. in [DK99b]). One then arrives again at a Wright-Fisher diffusion with mutation with diffusion coefficient $x(1-x)$ and drift coefficient $(1-x)\theta\nu_0 - x\theta\nu_1$. But now this is a 'strong' limit (realisation by realisation) whereas it was only a weak limit (limit in distribution) in the Moran model.

Adding *selection* to the lookdown model was done first by Donnelly and Kurtz [DK99a] in 1999. Whereas they focus on viability selection, we again concentrate on fertility selection (as done in Section 2.1.2). In the viability selection case, type-1 individuals die at a higher rate than type-0 individuals. Thus, in addition to the neutral reproduction arrows, selective death indicators at rate σ per level are added. These indicators can only be used by individuals of type-1 and are ignored by individuals of type 0. At each death event, the free space is filled by a clone of a randomly chosen individual before the event (maybe a copy of the dying individual itself). Fertility selection means that type-0 individuals reproduce at a higher rate than type-1 individuals. In comparison to neutral reproduction arrows, the added selective arrows in the lookdown model may point in both possible directions: upwards or downwards. Therefore, (fertility) selection can be modelled exactly as in

the Moran model. In addition to the reproduction arrows, independently per level, a Poisson point process of selective arrows originates from each level at rate σ. The tip of a selective arrow points to a uniformly chosen level. An example is shown in Fig. 2.11.

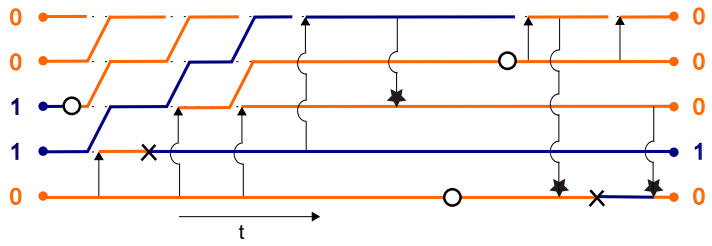


Figure 2.11: *Lookdown model with selection and mutation. Reproduction arrows always point upwards but selective arrows may point in both vertical directions. This picture emerges from Fig. 2.4 by a rearrangement of the lines.*

Donnelly and Kurtz [DK99a] show (for viability as well as fertility selection) that exchangeability is preserved and the type frequency process in the infinite-population limit is a Wright-Fisher diffusion with mutation and selection; its generator is given by (2.4). In addition, the genealogy of each sample of size n performs an ancestral selection graph started with n lines.

A slightly different lookdown model with selection was introduced by Bah, Sow, and Pardoux [BPS12] in 2012. In their model, which deals with viability selection, each individual of type 1 dies at rate σ. But it is not that level at which a clone of a randomly chosen individual is born. Instead, when an individual at level ℓ dies due to a selective event, all individuals at levels $k > \ell$ are shifted one level downwards to levels $k-1$. The empty space at level N is filled by a child that is a copy of a randomly chosen individual.

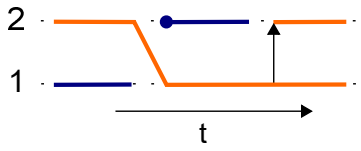


Figure 2.12: *Lookdown model with (viability) selection with population size $N = 2$ in the version by Bah, Sow, and Pardoux [BPS12]. The type-1 line at level 1 dies. The line at level 2 is shifted down to 1 and a copy of a randomly chosen individual (here of the type-1 individual that was available at level 1 just before the death event) is placed on level N. This selective event is followed by a birth event at which it is the newly born line at level N that has to die a.s.*

Bah, Sow, and Pardoux show that their version of a lookdown model with infinite population size preserves the exchangeability of the lines at all times. In addition, the proportion of type-0 individuals in the finite population model with population size N converges in probability to a type frequency

process which is again a Wright-Fisher diffusion with selection. However, in comparison to the finite population version of the lookdown model by Donnelly and Kurtz [DK99a], the version by Bah et al. [BPS12] (or [BP15] in the special case $\Lambda((0,1]) = 0$) in the case of finite population size is not a representation of the Moran model. It is not Markovian any more but has some predictable elements that yield a bias at level N. Indeed, at neutral reproduction events, the line that dies is always located at level N. At selective death events, the free space that is filled with a new child is also located at level N. Therefore, if there was a selective death event forward in time leading to the birth of the line on level N, this newly born line (and not a randomly chosen one) is for sure the line to die if this selective event is followed by a neutral reproduction event. A minimal example is shown in Fig. 2.12: The line that dies is the line that was born in the last step a.s. In a Moran model, both possible lines (the newly born one or the older line) would die with probability $1/2$.
Although the model by Bah, Sow and Pardoux is not a representation of the Moran model, we will use some similar ideas when defining our *lookdown ASG* in Chapter 3.

A different approach of ordering the lines in the Moran model with (fertility) selection (and without mutation) was introduced by Kluth and Baake [KB13]. In their *labelled Moran model* every individual is assigned a label (in addition to its level). Reproduction events are distinguished as either 'neutral' or 'selective' ones. In contrast to the rules in the lookdown model of Donnelly and Kurtz, neutral reproduction arrows can point in all directions but selective reproduction arrows are only allowed from lower to higher labels. Although this approach lacks important properties of the lookdown model (e.g. exchangeability of the lines), it gives some insight into the number and nature of selective events that yield fixation of one of two possible types.

2.1.6 Λ-Wright-Fisher diffusion and Λ-coalescent

So far, forward in time, we only considered binary reproduction events in the Moran and lookdown model. At each reproduction event, the mother has only one single offspring. This assumption will be made through Chapters 3, 5, and 6. But as we will also analyse birth events with more than one child in Chapter 4, let us briefly give some insight into the Moran model with reproduction events of arbitrary size and the genealogical process backward in time here.

In the Moran model, a birth event with $j-1$ children, $j \geq 2$, is modelled by a 'multi-arrow' with one tail (emanating from the level of the parent) but $j-1$ tips pointing to the locations of the newly born children. In order to make space for the children, the individuals located at the levels of the tips of the arrows die (compare Fig. 2.13).

Let us model the rates of the reproduction events according to a probability measure on $[0,1]$, denoted by Λ. Let $\lambda_{b,j}$,

$$\lambda_{b,j} := \int_{0,1} z^j (1-z)^{b-j} z^{-2} \Lambda(dz), \quad j \leq b, \tag{2.16}$$

be the rate at which a 'multi-arrow' is interchanged between j fixed lines among a collection of b lines in total. Consequently, in a population of size N, the Poisson point process of arrows with tail at level ℓ and tips at levels $\ell_1, \ell_2, \ldots, \ell_{j-1}$ comes independently at rate $\lambda_{N,j}/j$ for each arbitrary but fixed and pairwise disjoint tuple $(\ell, \ell_1, \ldots, \ell_{j-1}) \in \{1, \ldots, N\}^j$. The lines at levels $\ell, \ell_1, \ldots, \ell_{j-1}$ merge into

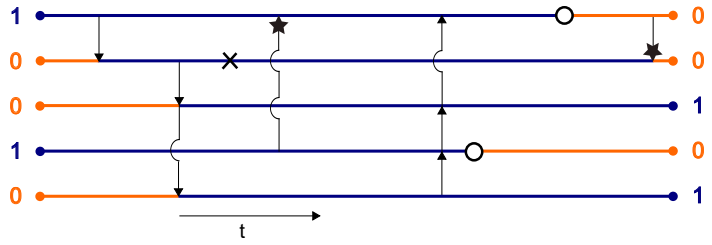


Figure 2.13: *Moran Model with mutation and selection with birth events of up to three children.*

one line at rate $\lambda_{N,j}$ and the probability that the parent is located at level ℓ is $1/j$.
Selection and mutation is modelled the same way as already described in Section 2.1.2.

Ordering again the lines in the Moran model by persistence results in the lookdown model, we get the following: At a reproduction event with j involved lines, the parent is always the individual that is located at the lowest participating level. It places its $j-1$ children at all other affected levels (compare also [DK99a, DK99b]). That way, the location of the parent of each reproduction event is determined by the locations of all participating levels. For clarity, when drawing a realisation of a lookdown model with birth events of arbitrary size, we omit the reproduction arrows and only draw bullets[10] at all affected levels (as shown in Fig. 2.14). In order to make space for the newly born children, lines again do not die but are shifted upwards. Mutation and selection is modelled as already described in Section 2.1.5.

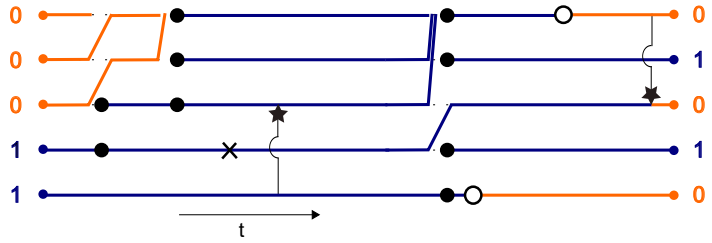


Figure 2.14: *Lookdown model with mutation and selection. Neutral reproduction events are modelled by bullets. The parent of each birth event is always the individual at the lowest level with a bullet. This picture emerges from Fig. 2.13 by a rearrangement of the lines.*

The frequency X of type 0 individuals in a population of infinite size with birth events according to the reproduction measure Λ, is then a so-called Λ-*Wright-Fisher diffusion* with mutation and selection

[10]The drawing of bullets instead of arrows at reproduction events is also motivated by the Poissonian construction of the Λ-coalescent (reviewed e.g. in [Ber09]).

with generator given by

$$G_X g(x) = \int_{(0,1]} \big[x(g(x+z(1-x))-g(x)) + (1-x)(g(x-zx)-g(x))\big] \frac{\Lambda(dz)}{z^2} + \Lambda(\{0\}) \cdot \frac{1}{2} x(1-x) g''(x) \; + \; \big[\sigma x(1-x) - \theta \nu_1 x + \theta \nu_0 (1-x)\big] g'(x), \tag{2.17}$$

$g \in \mathscr{C}^2[0,1]$ (see for example [Gri14]). As this generator is further investigated in Chapter 4, we omit the details here and continue with the Λ-Wright-Fisher process *without* mutation and selection (i.e. $\theta = \sigma = 0$).

By the analogy with the Kingman coalescent being the moment dual to the Wright-Fisher process with binary reproduction events only (compare (2.7) and (2.8) in Section 2.1.3), the Λ-Wright-Fisher process has a moment dual as well: The Λ-*coalescent* was introduced in 1999 in three independent papers by Pitman [Pit99], Sagitov [Sag99], and Donnelly and Kurtz [DK99b]. N. Berestycki worked out a general review [Ber09].
The Λ-n-coalescent evolves backward in time and gives the genealogy of a sample of n individuals in a Λ-Wright-Fisher population (see Fig. 2.15, top, for a realisation in the lookdown model). By definition, the Markovian family of Λ-n-coalescents takes values in the set of partitions of $\{1,\dots,n\}$. It starts with the set of all singletons $\{\{1\},\dots,\{n\}\}$ and finally gets absorbed in the trivial partition $\{1,\dots n\}$, when all blocks (or ancestral lines of individuals) have coalesced. The blocks (or lines) are required to be exchangeable at all times. The consistency property holds, that is the Λ-n-coalescent restricted to the first m lines is a Λ-m-coalescent for all $m \in \{1,\dots,n\}$. In addition, the rates at which any arbitrary but fixed tuple of j out of b blocks (or lines) merges into one is given by $\lambda_{b,j}$ as defined in (2.16).

When the measure Λ takes certain values, some special classes of coalescents appear. For example, in the case $\Lambda = \delta_0$, the Dirac mass at 0, the Kingman coalescent (described in Section 2.1.3) arises. We will refer to this case as *Kingman case*. When $\Lambda = \delta_1$, we have the so-called *star-shaped-coalescent*. This coalescent is remindfull of a star because at rate 1 all lines coalesce at once into one single line. For $\Lambda =$unif$[0,1]$, the uniform measure, the Λ-coalescent is a *Bolthausen-Sznitman coalescent* [BS98]. In the case when Λ is a Beta$(2-\alpha,\alpha)$ distribution, the class of *Beta-coalescents* arises. A review on all just described types of coalescents can be found in [Ber09].

The intuition behind (2.16), concentrated on $(0,1]$, becomes transparent when considering the Poissonian construction of the Λ-coalescent (e.g. in the lookdown representation shown in Fig. 2.15): In the first stage, generate a random configuration of points on $\mathbb{R} \times (0,1]$ according to a Poisson point process with intensity measure $dr \cdot \Lambda(dz)/z^2$. Given the point configuration, in the second stage, perform for each point (r,z) a Bernoulli experiment at time r with success probability z on the levels $1,2,\dots$. Then, given the successful levels are $\ell_1,\ell_2,\ell_3,\dots$, the lines at levels $\ell_2,\ell_3,\dots$ merge into the line at level ℓ_1.
Given the parameter of the Bernoulli experiment is z, the probability that exactly j fixed levels out of b are successful is $z^j(1-z)^{b-j}$. Integrating over all possible success probabilities z gives (2.16).

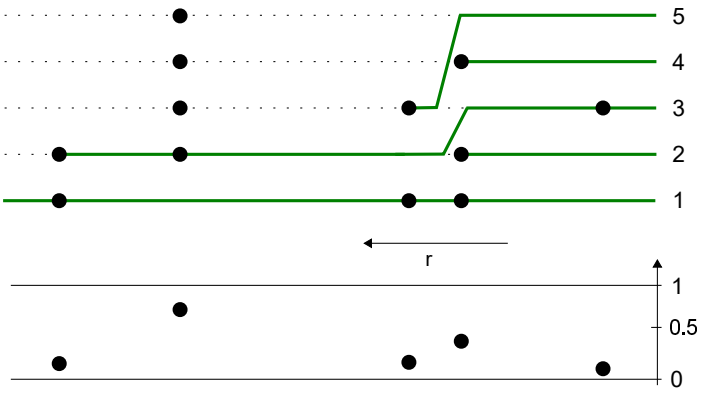


Figure 2.15: *Poissonian construction of the Λ-coalescent. Bottom: The Poisson point process of coalescing times and coalescing probabilities per level. Top: The bullets are the successes of the Bernoulli experiment, the green lines are the lines of the corresponding Λ-5-coalescent.*

The generator of the line counting process K of the Λ-coalescent is given by

$$G_K g(b) = \sum_{c=1}^{b-1} \binom{b}{b-c+1} \lambda_{b,b-c+1} \left[g(c) - g(b) \right], \quad g : \mathbb{N} \to \mathbb{R}, \tag{2.18}$$

because there are $\binom{b}{b-c+1}$ different tuples of $b-c+1$ out of b lines, each of them coalesces at rate $\lambda_{b,b-c+1}$ into one single line, and the remaining number of lines after such a coalescence is c.

Again, like in the case with only binary birth events, when selection and mutation is added to the Λ-Wright-Fisher diffusion, it becomes rather involved to find the true genealogy of a sample. The Λ-coalescent with mutation and selection is investigated in Chapter 4.

All results in this thesis require the assumption that a.s. there are no simultaneous multiple birth events, i.e. at each point in time, there is at most one birth event with a single mother.
When thinking of a scenario with simultaneous multiple birth events forward in time, the genealogy of a sample backward in time has to perform simultaneous multiple mergers. Although this is not a subject here, let us briefly note that one then arrives at the class of so-called Ξ-coalescents, a generalisation of Λ-coalescents. These coalescents with simultaneous multiple mergers were introduced by Schweinsberg in 2000 [Sch00].

2.2 The common ancestor type distribution

In this section we come back to the classical Wright-Fisher diffusion. We present the main quantity of this thesis, namely the type distribution of the common ancestor (or immortal line). As already outlined for the Wright-Fisher diffusion without selection on page 16, at a given time s, the probability that the common ancestor is of type 0 is denoted by $h(x)$, given the type-0 frequency at time s is x. We also review results by Kluth, Hustedt and Baake [KHB13], Fearnhead [Fea02], and Taylor [Tay07].

2.2.1 The common ancestor's type in a population of finite size N

In a population with particle representation given by the Moran model or lookdown model with selection and mutation with population size $N \in \mathbb{N}$, at any time $t \geq 0$ there a.s. exists a unique individual whose progeny will take over the whole population at some time $s > t$. All individuals at time s will then have this single individual as their ancestor at time t. Therefore, this individual is denoted by *common ancestor* at time t. For example, the common ancestor in the Moran model in Fig. 2.4 (at the time that corresponds to the left-hand side of the picture) is the type-0 individual located at level 4, and in the lookdown model in Figs. 2.11 and 2.14 (on the left-hand side of both pictures) it is located at level 1. Since such a common ancestor exists a.s. for all times $t > 0$, the line of the common ancestor throughout times is also denoted by *immortal line*.

In a population of size N, let h_k^N be the probability that the common ancestor has type 0 at time t, given the proportion of type-0 individuals at time t is k/N. As already explained on page 16, in the case without selection ($s_N = 0$), we have $h_k^N = k/N$, due to exchangeability of the lines.
When selection is present, the type of the immortal line has a tendency towards the beneficial type such that $h_k^N > k/N$. This scenario was analysed by Kluth, Hustedt and Baake [KHB13] in 2013. Using 'first-step analysis' methods, they derived a system of equations for h_k^N [KHB13, Equation (21)],

$$\begin{aligned}(\lambda_k^N + \mu_k^N)h_k^N &= \lambda_k^N h_{k+1}^N + \mu_k^N h_{k-1}^N + k u_N \nu_1 \frac{1-h_{k-1}^N}{N-(k-1)} - (N-k)u_N\nu_0 \frac{h_{k+1}^N}{k+1}, \quad 0<k<N, \\ h_0^N &= 0, \quad h_N^N = 1.\end{aligned} \tag{2.19}$$

Substituting $\psi_k^N = h_k^N - k/N$, $0 \leq k \leq N$, one can speak of ψ^N as the additional absorption probability of type 0 individuals due to selection. Using then the ansatz

$$\psi_{N-k}^N = (N-k)\sum_{i=1}^{k} a_i^N \frac{k(k-1)\dots(k-i+1)}{N(N-1)\dots(N-i)} \tag{2.20}$$

gives a recursion for the coefficients a_n^N, $1 \leq n \leq N-1$, compare [KHB13, Theorem 2],

$$\begin{aligned}&a_0^N = 1, \quad a_1^N = N\psi_{N-1}, \\ &(N-n)\left[\left(\frac{n}{N}+u_N\nu_1\right)a_n^N - \left(\frac{n}{N}+\frac{N-(n-1)}{N}s_N + u_N\right)a_{n-1}^N + \frac{N-(n-1)}{N}s_N a_{n-2}^N\right] = 0, \quad n \geq 2.\end{aligned} \tag{2.21}$$

Kluth, Hustedt and Baake also show that for $N \to \infty$ and $k/N \to x$, $x \in [0,1]$, the equations in their discrete setting converge to the equations given in the continuous setting and analysed by Fearnhead and Taylor. Namely $h_k^N = \psi_k^N + k/N$ (determined by (2.20)) converges to $h(x)$ (given by Equation (2.48)) and the system of Equations (2.19) converges the the recursion of Fearnhead's coefficients (2.33). The continuous setting with infinite popolation size is presented in the next sections.

2.2.2 The ASG as a backbone for a population of infinite size

In an (infinite size) Wright-Fisher population with selection and mutation, the ASG (started with n lines) contains all information about the genealogy of a sample of n individuals. For $\sigma < \infty$, as the

coalescence rate of the ASG is quadratic but the branching rate only linear, $n = \infty$ is an entrance boundary. We can start the ASG at time $r = -s$ with all individuals in the population (infinitely many lines). Then, for all $\varepsilon > 0$, there are only finitely many lines left at time $-s + \varepsilon$. The ASG is said to *come down from infinity*. In addition, the time length T_{UA} until all lines have coalesced into one

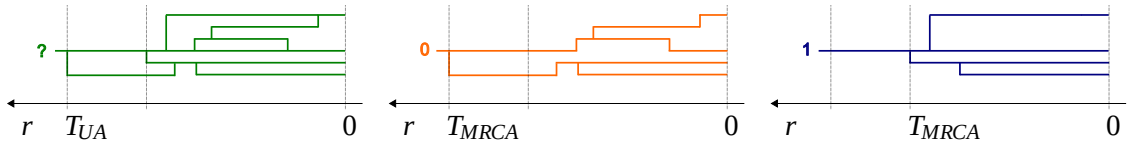


Figure 2.16: *The time to the ultimate ancestor and the most recent common ancestor of a sample of* 4 *individuals taken at time* $r = 0$ *(for simplicity in the case without mutations). Left: ASG (containing all potential ancestors). The time* T_{UA} *is the first time at which all lines have merged into one single line. The types of all individuals are unknown. Middle: True genealogy in the case that the type of the ultimate ancestor is* 0*. The time* T_{MRCA} *to the most recent common ancestor in this example coincides with* T_{UA}*. Right: True genealogy in the case when the type of the ultimate ancestor is* 1*. In this example we have* $T_{MRCA} < T_{UA}$*.*

single line (the *ultimate ancestor*) for the first time, satisfies $T_{UA} < \infty$ a.s. (compare [KN97, Thm. 3.2]). Note that a common ancestor of the whole population is the single line at time $-s + T_{UA}$ (see also Fig. 2.16 in the case $s = 0$). The time $-s + T_{MRCA}$ till the *most recent* common ancestor of the population (sampled at time $r = -s$) is bounded by $-s + T_{UA}$ from above. In addition, for all times $r > -s + T_{MRCA}$ the (unique) line of the common ancestor (the immortal line) is embedded in the ASG started with the single line at time $-s + T_{UA}$. Therefore, letting $s \to \infty$, the immortal line exists at all times and is included in the ASG started with one single line at time $r = -\infty$. At all times $r \in \mathbb{R}$, the number of lines in this ASG at time r is distributed according to the stationary distribution. Thus, we term the ASG that is started with one single line at time $r = -\infty$ *equilibrium ASG*. In some sense, we can think of the equilibrium ASG as the backbone of the population; eventually, the ancestral line of any individual sampled at any time $r < \infty$ coalesces into one line of this ASG a.s. (compare Fig. 2.17).
Since the immortal line is included in the equilibrium ASG at all times, the distribution of the type of the common ancestor of a stationary Wright-Fisher process can be determined by analysing the equilibrium ASG instead of regarding the whole population. Such an approach was carried out by Fearnhead in 2002 [Fea02].

2.2.3 Fearnhead's approach to the common ancestor type distribution

Fearnhead's analysis of the type of the immortal line in a stationary Wright-Fisher population starts with the ASG of Krone and Neuhauser [KN97]. In detail, it starts with the equilibrium ASG. As Fearnhead wants to determine the distribution of the type of the common ancestor, he does not need to construct the ancestral graph of a sample of arbitrary size but it suffices to choose sample size 1 (as explained in the last preceding paragraph). The ASG of Neuhauser and Krone is an untyped branching and coalescing graph: in the first step, each branching event remains unresolved. Without

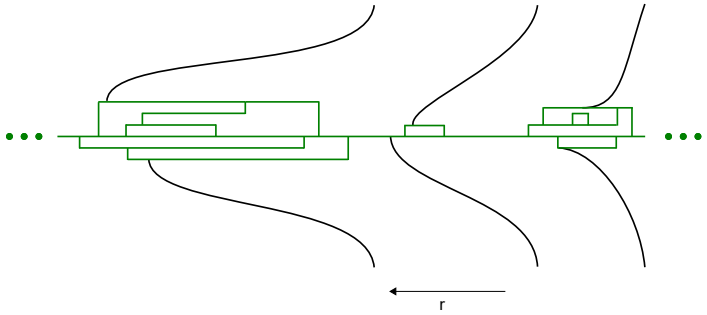


Figure 2.17: *The ASG started with one single line at time $r = -\infty$ is a backbone of the population (green lines). The ancestral lines of all individuals eventually coalesce into this ASG (some ancestral lines are drawn in black).*

types, it is not clear whether the incoming or continuing line is the true parent to the branching event. To determine the real branch (i.e. the true ancestral line), types are only assigned to a given realisation of the ASG. In his graph, instead of using an a priori untyped situation, Fearnhead keeps track of the types of all lines at all times. This also amounts to keeping track of which branches are virtual (i.e. those potential ancestors, that turn out to not be ancestral at all) and which branch is the single real branch (the immortal line). The state space is then $\{(r;n_0,n_1) \mid r \in \{0,1\}, n_0,n_1 \in \mathbb{N}_0\}$, where r is the type of the real line, and n_0 and n_1 are the numbers of virtual lines of types 0 and 1, respectively.
Remember from Equation (2.14) that $p(n_0,n_1)$ was the stationary probability of drawing a sample with n_j individuals of type j, $j \in \{0,1\}$, in a Wright-Fisher population and let $p(j \mid (n_0,n_1)) := p(n_0+\delta_{j0}, n_1+\delta_{j1})/p(n_0,n_1)$ be the conditional probability of drawing an individual of type j, given already a sample of configuration (n_0,n_1). When the current state is $(i;n_0,n_1)$, the transition rates of the typed version of the ASG are then given by[11]

(a) coalescence of two branches of type j;

transititon to state $(i;n_0-\delta_{j0}, n_1-\delta_{j1})$:

$$\frac{(n_j+\delta_{ij})(n_j+\delta_{ij}-1)}{2} \cdot \frac{p(n_0+\delta_{i0}-\delta_{j0}, n_1+\delta_{i1}-\delta_{j1})}{p(n_0+\delta_{i0}, n_1+\delta_{i1})} \tag{2.22}$$

(b) mutation of the real branch to type $k \neq i$;

transititon to state $(k;n_0,n_1)$:

$$\theta \nu_i \cdot \frac{p(n_0+\delta_{k0}, n_1+\delta_{k1})}{p(n_0+\delta_{i0}, n_1+\delta_{i1})} \tag{2.23}$$

(c) mutation of a virtual branch from type j to $k \neq j$;

transition to state $(i;n_0+\delta_{k0}-\delta_{j0}, n_1+\delta_{k1}-\delta_{j1})$:

$$n_j \theta \nu_j \cdot \frac{p(n_0+\delta_{i0}-\delta_{j0}+\delta_{k0}, n_1+\delta_{i1}-\delta_{j1}+\delta_{k1})}{p(n_0+\delta_{i0}, n_1+\delta_{i1})} \tag{2.24}$$

[11] Note that there are some small typos in the rates of [Fea02, p. 42/43].

(d) selective event, the incoming branch being of type 1 and the continuing branch of type k;

transition to state $(i; n_0, n_1 + 1)$:

$$(n_k + \delta_{ik})\sigma \cdot \frac{p(n_0 + \delta_{i0}, n_1 + 1 + \delta_{i1})}{p(n_0 + \delta_{i0}, n_1 + \delta_{i1})} \tag{2.25}$$

(e) selective event, the incoming branch being of type 0 and the continuing of type k;

transition to state $(i; n_0 + \delta_{k0}, n_1 + \delta_{k1})$:

$$(n_0 + \delta_{i0})\sigma \cdot \frac{p(n_0 + \delta_{i0} + \delta_{k0}, n_1 + \delta_{i1} + \delta_{k1})}{p(n_0 + \delta_{i0}, n_1 + \delta_{i1})} \cdot p\big(k \mid (n_0 + \delta_{i0}, n_1 + \delta_{i1})\big) \tag{2.26}$$

For a formal derivation of these rates, Fearnhead refers to [SD00, Thm. 1]. Intuitively, the rates are again evident from a 'first step decomposition' according to the first event back in time (compare the paragraph following Equation (2.15) on page 14).
Theoretically, to find the common ancestor type distribution, one may now calculate the stationary distribution with respect to these rates and then integrate over the types of the virtual lines. But as the rates are rather involved, it is not clear how that should work in practice.

Fearnhead's idea is then to simplify the typed ASG and arrive at a Markovian process which he calls *common ancestor process* (*CAP*). The CAP arises from the ASG by deleting some virtual lines that do not influence the evolution of the immortal line. In detail, all virtual lines of type 0 are deleted from the graph such that only virtual lines of type 1 remain in the ASG. For example, a virtual line that mutates to type 1 is deleted directly after the mutation and branching events are only allowed to result in an additional line of type 1. This results in the state space $\{(r,v) \mid r \in \{0,1\}, v \in \mathbb{N}_0\}$, where r is the type of the real line and v the number of virtual lines, all of type 1.
If the current state is (r,v), the transition rates of the CAP are given by

(i) mutation of the real branch to type $k \neq r$;

transititon to state (k,v):

$$\theta\nu_r \cdot \frac{p(\delta_{k0}, v + \delta_{k1})}{p(\delta_{r0}, v + \delta_{r1})} \tag{2.27}$$

(ii) coalescence of two branches of type 0 or mutation of a virtual branch to type 0;

transititon to state $(r, v-1)$:

$$\left[\frac{(v + \delta_{r1})(v + \delta_{r1} - 1)}{2} + v\theta\nu_1\right] \cdot \frac{p(\delta_{r0}, v + \delta_{r1} - 1)}{p(\delta_{r0}, v + \delta_{r1})} \tag{2.28}$$

(iii) selective event, the additional branch being of type 1;

transition to state $(r; v+1)$:

$$(v+1)\sigma \cdot \frac{p(\delta_{r0}, v + 1 + \delta_{r1})}{p(\delta_{r0}, v + \delta_{r1})} \tag{2.29}$$

Rate (i) coincides with rate (b) and rate (iii) with rate (d) of the ASG, rate (ii) arises from rates (a) and (c). Since branching events with an additional line of type 0 are forbidden in the CAP, rate (e) is not present among the rates of the CAP.

Fearnhead shows that indeed the CAP (R,V), where R is the type of the real line and V the number of virtual lines, is a Markov process. In addition, he shows that branches of type 0 (except for the real line) in the typed ASG contain no further information on the history of the line of the common ancestor. The ancestry of the immortal line (real line) in the CAP coincides with the ancestry of the immortal line in the ASG [Fea02, Thm. 1].

The CAP describes the distribution of the type process R on the immortal line, it is a Markov process that includes less lines than the ASG, and is therefore an object much simpler to analyse than the typed ASG. With the CAP, Fearnhead succeeds to specify the distribution $\phi(r,v) := \mathbb{P}\big((R,V) = (r,v)\big)$ of the type of the common ancestor together with the number of virtual lines in a stationary situation [Fea02, Theorem 3].
In detail[12], for $\theta > 0$ and $0 < \nu_1 < 1$, the probability weights ϕ of the unique stationary common ancestor type distribution together with the number of virtual lines is given by

$$\phi(r,v) = \left(a_v \mathbf{1}_{\{r=0\}} + (a_v - a_{v+1})\mathbf{1}_{\{r=1\}} \right) \cdot \frac{(v+1)!}{(v+\delta_{r1})!} p(\delta_{r0}, v+\delta_{r1}) \tag{2.30}$$

with the coefficients (a_n) given by

$$a_0 = 1, \quad a_n = \prod_{j=1}^{n} \lambda_i, \quad n = 1, \ldots, \tag{2.31}$$

and $\lambda_i = \lim_{k\to\infty} \lambda_i^{(k)}$ specified by $\lambda_{k+1}^{(k)} = 0$ and the recursion

$$\lambda_{i-1}^{(k)} = \frac{2\sigma}{i + 2\theta + 2\sigma - (i + 2\theta\nu_1)\lambda_i^{(k)}}, \quad i \geq 2. \tag{2.32}$$

For the proof, Fearnhead states that the common ancestor process is irreducible and recurrent and that $\big(\phi(r,v)\big)$ defines a probability distribution. Then, in the main part of the proof, he checks that the stationarity condition is true,

$$\sum_{(r,v)} \phi(r,v) q\big((r,v),(r',v')\big) = 0, \quad \text{for all } (r',v') \in \{0,1\} \times \mathbb{N}_0,$$

where $q\big((r,v),(r',v')\big)$ is the 'Q-matrix' with the rates $(i) - (iii)$ given by Equations (2.27), (2.28), and (2.29).
Although this calculation works quite straight forward, a probabilistic interpretation of the coefficients (λ_i) or (a_n) does not become clear from the proof.

Remark 2.1 (Fearnhead's recursion) *Note that the definition of the coefficients (a_n) in Equations* (2.31) *and* (2.32) *is equivalent to defining them via*

$$\left(\frac{n+1}{2} + \theta\nu_1\right) a_{n+1} - \left(\frac{n+1}{2} + \sigma + \theta\right) a_n + \sigma a_{n-1} = 0, \quad n \geq 1, \tag{2.33}$$

with the constraints $a_0 = 1, \lim_{n\to\infty} \frac{a_{n+1}}{a_n} = 0$.

[12] Typographically, the mutation probability 'ν_1' to type 1 and the variable 'v' that stands for the number of virtual lines may not be easy to distinguish on the first sight. But note that only the mutation probability, ν_1 or ν_0, always goes with an index.

The coefficients (a_n) are called *Fearnhead's coefficients*. Relation (2.33) is denoted by *Fearnhead's recursion*.

Writing out (2.30) for the two cases $r = 0$ and $r = 1$ (and using the definition of p in Equation (2.14)) yields the following representation of ϕ,

$$\begin{aligned} \phi(0,\nu) &= a_\nu \cdot \mathbb{E}\left[X(1-X)^\nu\right], \\ \phi(1,\nu) &= (a_\nu - a_{\nu+1}) \cdot \mathbb{E}\left[(1-X)^{\nu+1}\right]. \end{aligned} \tag{2.34}$$

The stationary distribution of the type of the common ancestor, which we denote by I, can then be gained by summing over the number of virtual lines.

$$\begin{aligned} \mathbb{P}(I=0) &= \sum_{\nu\geq 0} \phi(0,\nu) = \sum_{\nu\geq 0} a_\nu \cdot \int_0^1 x(1-x)^\nu \cdot (1-x)^{2\theta\nu_1-1} x^{2\theta\nu_0-1} \exp\{2\sigma x\}\, dx, \\ \mathbb{P}(I=1) &= \sum_{\nu\geq 0} \phi(1,\nu) = \sum_{\nu\geq 0} (a_\nu - a_{\nu+1}) \cdot \int_0^1 (1-x)^{\nu+1} \cdot (1-x)^{2\theta\nu_1-1} x^{2\theta\nu_0-1} \exp\{2\sigma x\}\, dx \end{aligned} \tag{2.35}$$

In [Fea02, Remark 3], Fearnhead gives a simulation algorithm for a stationary realisation of (R,V) according to 2.34. The '*Fearnhead simulator*' (as we call it here) works as follows:

1. Start with $n = 0$ virtual lines.
2. Take one individual at random from the stationary population.
3. If it is of type 0, call it the common ancestor. We have $(r,\nu) = (0,n)$ and are done.
   If it is of type 1,
   - call it the common ancestor with probability $1 - \frac{a_{n+1}}{a_n}$, we have $(r,\nu) = (1,n)$ and are done.
   - call it a virtual line with probability $\frac{a_{n+1}}{a_n}$, take $n = n+1$ and continue with step 2.

2.2.4 Taylor's approach to the common ancestor type distribution

In 2007, J. Taylor published a new approach to the distribution of the common ancestor's type in a stationary Wright-Fisher population [Tay07]. While Fearnhead's ansatz uses a pruned version of the ASG with the immortal line evolving in the environment of some virtual lines, Taylor's concept starts with a structured coalescent of the immortal line in the random environment of the frequency of type-0 individuals.

Remember that the frequency process X of type-0 individuals in a Wright-Fisher population with selection and mutation has drift coefficient $(1-x)\theta\nu_0 - x\theta\nu_1 + \sigma x(1-x)$ and diffusion coefficient $x(1-x)$. The (forward in time) generator G_X is given by (2.4) and its stationary probability density $w(x)$ by (2.5). But genealogies evolve backward in time. One important property of the Wright-Fisher diffusion is its invariance under time reversal (compare [BES04, Section 3]). Thus, the backward in time generator[13] $\hat{G}_X$ coincides with the forward in time generator G_X.

[13] Throughout, generators or rates indexed with $\hat{}$ are always defined backward in time. The symbol $\hat{}$ is used to distinguish them from the forward in time quantities.

In the random environment of the backward in time Wright-Fisher diffusion with mutation and selection one may consider the dynamics of typed ancestral lines. In a sample (n_0, n_1) of size $n_0 + n_1$, let n_0 and n_1 be the number of lines of types 0 and 1, respectively. backward in time, the line counting process of the sample performs a *structured coalescent* in the random environment X with generator given by (compare [Tay07, eq. (3)])

$$\begin{aligned}\hat{G}_{(X,N_0,N_1)}\, g(x,n_0,n_1) \;&= G_X\, g(x,n_0,n_1)\\ &\quad + \binom{n_0}{2}\frac{1}{x}\,\big[g(x,n_0-1,n_1)-g(x,n_0,n_1)\big]\\ &\quad + \binom{n_1}{2}\frac{1}{1-x}\,\big[g(x,n_0,n_1-1)-g(x,n_0,n_1)\big]\\ &\quad + n_0\theta\nu_0\frac{1-x}{x}\,\big[g(x,n_0-1,n_1+1)-g(x,n_0,n_1)\big]\\ &\quad + n_1\theta\nu_1\frac{x}{1-x}\,\big[g(x,n_0+1,n_1-1)-g(x,n_0,n_1)\big],\end{aligned} \tag{2.36}$$

with the test function $g \in \mathscr{C}^2(0,1)\times\mathbb{N}_0\times\mathbb{N}_0$ and the convention $\binom{1}{2} = \binom{0}{2} = 0$. The first part of (2.36) is the usual Wright-Fisher generator with mutation and selection that acts on the type frequency process. The parts with binomial factors describe the coalescence of two lines of type 0 or two lines of type 1, at type-0 frequency x. Lines of different types are not allowed to coalesce. The last two terms in (2.36) represent mutations of a line from type 0 to 1 or from 1 to 0. Since we are working in the typed case, the parent of an individual at a selective event can be determined directly (it is the incoming branch if and only if the incoming branch is of type 0). Thus, branching events do not appear here.
We will take a closer look at the shape of the rates of $\hat{G}_{(X,N_0,N_1)}$ in Chapter 6 when analysing the triple process (X,R,V).

For considering the line of the common ancestor, a sample of size 1 (either $(n_0,n_1) = (0,1)$ or $(n_0,n_1) = (1,0)$) suffices. Note that the evolution of the immortal line coincides with the evolution of the ancestral line of a randomly sampled individual. This is due to the fact (compare also Taylor's explanation [Tay07, p. 817]) that eventually the line of any randomly sampled individual coalesces into the immortal line.
Taylor expands the process $(X_r)_{r\in\mathbb{R}}$, the frequency of type 0 individuals (evolving backward in time), by adding the type R_r of the immortal line (real line) at backward time r. This yields the backward in time generator ([Tay07, eq. (4)])

$$\begin{aligned}\hat{G}_{(X,R)}\, g(x,0) \;&= G_X\, g(x,0) + \theta\nu_0\frac{1-x}{x}\,\big[g(x,1)-g(x,0)\big]\\ \hat{G}_{(X,R)}\, g(x,1) \;&= G_X\, g(x,1) + \theta\nu_1\frac{x}{1-x}\,\big[g(x,0)-g(x,1)\big], \qquad g\in\mathscr{C}^2[0,1]\times\{0,1\}.\end{aligned} \tag{2.37}$$

Therefore, the process (X,R) backward in time is a jump diffusion process: the type frequency process X performs a Wright-Fisher diffusion and the type process R on the real line is a jump process on $\{0,1\}$ with rates depending on the current state of X.

Let $(\pi(x,r))_{x\in[0,1],r\in\{0,1\}}$ be the stationary distribution of (X_r,R_r). Due to the shape of the generator $\hat{G}_{(X,R)}$, it is possible to write the density in terms of its factors,

$$\pi(x,r) = w(x)\cdot\Big[h(x)\mathbf{1}_{\{r=0\}} + (1-h(x))\mathbf{1}_{\{r=1\}}\Big], \tag{2.38}$$

with $w(x)$ being the equilibrium distribution (2.5) of X and $h(x)$ again defined as

$$h(x) := \mathbb{P}\left(\text{immortal line is of type 0 at time } s \mid X_s = x\right). \tag{2.39}$$

In order to gain a representation of $h(x)$, Taylor investigates the stationarity condition. Since π is the equilibrium distribution for $\hat{G}_{(X,R)}$ it is true that

$$0 = \sum_{r\in\{0,1\}}\int_{x\in[0,1]} \hat{G}_{(X,R)}\, g(x,r)\pi(dx,r) \quad \forall g\in\mathscr{C}^2[0,1]\times\{0,1\}. \tag{2.40}$$

Using some algebra and reordering the terms, Taylor proves that Equation (2.40) is equivalent to the boundary value problem [Tay07, eq. (9)]

$$G_X\, h(x) - h(x)\left[\theta\nu_0\frac{1-x}{x} + \theta\nu_1\frac{x}{1-x}\right] = -\theta\nu_1\frac{x}{1-x}\,, \quad h(0)=h(1)=0\,. \tag{2.41}$$

Taylor shows that the solution to (2.41) is unique. Before reviewing this solution, let us mention that Taylor also calculates the generator $G_{(X,R)}$ of the process (X,R) *forward* in time [Tay07, Section 2.2]. This can be done by reversing time with respect to the stationary distribution π,

$$\int g_1\cdot(G_{(X,R)}\, g_2)\, d\pi = \int(\hat{G}_{(X,R)}\, g_1)\cdot g_2\, d\pi \qquad \text{for all test functions } g_1, g_2. \tag{2.42}$$

This results in the following representation of the generator $G_{(X,R)}$ of the process (X,R) forward in time,

$$G_{(X,R)}\, g(x,0) \;=\; G_X\, g(x,0) + x(1-x)\frac{h'(x)}{h(x)}g'(x,0) \;+\; \frac{x(1-h(x))}{(1-x)h(x)}\theta\nu_1\left[g(x,1)-g(x,0)\right] \tag{2.43}$$

$$G_{(X,R)}\, g(x,1) \;=\; G_X\, g(x,1) - x(1-x)\frac{h'(x)}{1-h(x)}g'(x,1) \;+\; \frac{(1-x)h(x)}{x(1-h(x))}\theta\nu_0\left[g(x,0)-g(x,1)\right] \tag{2.44}$$

for all $g\in\mathscr{C}^2[0,1]\times\{0,1\}$. We will take a closer look at the rates of this generator when investigating the triple process (X,R,V) in Section 6.4.

To solve the boundary value problem (2.41), Taylor tries the ansatz of using a power series in σ for $h(x)$. Indeed, the calculations then yield the solution

$$h(x) = x + 2\sigma\int_0^x(\tilde{x}-q)\exp\big(2\sigma(q-x)\big)\left(\frac{q}{x}\right)^{2\theta\nu_0}\left(\frac{1-q}{1-x}\right)^{2\theta\nu_1}dq, \tag{2.45}$$

where $\tilde{x}$ is the conditional expectation of drawing the third individual in a sample of type 0 from a stationary population given the first two individuals in the sample are of different types,

$$\tilde{x} = \frac{\int_0^1 q^2(1-q)w(q)dq}{\int_0^1 q(1-q)w(q)dq} = \frac{\mathbb{E}\left[X^2(1-X)\right]}{\mathbb{E}\left[X(1-X)\right]}, \tag{2.46}$$

with the Wright density w given by (2.5).

The stationary distribution of the type I of the common ancestor, can then be gained by integrating over the frequency of type-0 individuals,

$$\begin{aligned} \mathbb{P}(I=0) &= \int_0^1 h(x)w(x)dx, \\ \mathbb{P}(I=1) &= \int_0^1 \big(1-h(x)\big)w(x)dx. \end{aligned} \tag{2.47}$$

Comparing his results for $h(x)$ with Fearnhead's aproach, Taylor shows [Tay07, Lemma 4.1] that the boundary value problem (2.41) is also solved by the function

$$h(x) = x + \sum_{n\geq 1} x(1-x)^n \cdot a_n, \tag{2.48}$$

where the coefficients a_n are Fearnhead's coefficients determined by Fearnhead's recursion (2.33). Since the solution to the boundary value problem is unique, the right-hand side of (2.45) equals the right-hand side of (2.48) and we have

$$x\sum_{n\geq 1}(1-x)^n \cdot a_n = 2\sigma \int_0^x (\tilde{x}-q)\exp\big(2\sigma(q-x)\big)\left(\frac{q}{x}\right)^{2\theta\nu_0}\left(\frac{1-q}{1-x}\right)^{2\theta\nu_1} dq. \tag{2.49}$$

With the help of rewriting a_n,

$$a_n = \frac{(-1)^n}{n!}\left.\frac{d}{dx}\right|_{x=1}\left(\sum_{n\geq 1} a_n(1-x)^n\right),$$

and applying some algebra, Taylor develops from (2.49) a representation for a_1,

$$a_1 = \frac{2\sigma}{1+2\theta\nu_1}(1-\tilde{x}). \tag{2.50}$$

Taking the results of Fearnhead and Taylor together, for each triple of parameters (σ, θ, ν_0), Fearnhead's coefficients a_n can be derived recursively via Fearnhead's recursion (2.33) using the starting values $a_0 = 1$ and a_1 as in (2.50). Once the coefficients are calculated, the stationary type distribution of the common ancestor can be determined either with Fearnhead's approach via virtual lines (Equation (2.35)) or with Taylor's approach via the frequency of type-0 individuals (Equations (2.45) and (2.47)).

So far, the results by Fearnhead and Taylor are only valid for Wright-Fisher diffusions with mutation and selection *with binary birth events*. It is a priori not clear how to extend them to Λ-Wright-Fisher diffusions with reproduction measure $\Lambda \neq \delta_0$. To some extend, this is due to the fact that so far there is no explicit probabilistic interpretation of Fearnhead's coefficients a_n in terms of a particle picture. Since the representation of $h(x)$ in Equation (2.48) has a somehow geometric structure and we have $a_n \in [0,1]$ for all $n \geq 0$, the curiosity of finding a probabilistic meaning for $(a_n)_{n\geq 0}$ strongly motivated us for discovering the results given in Chapter 4 of this thesis.

A slightly different step towards the interpretation of Fearnhead's coefficients is done by Kluth, Hustedt, and Baake in terms of taking limits in the finite population size case. In [KHB13, Theorem 3] they show that recursion (2.21) for the coefficients (a_n^N) in the discrete setting with population size N converges to Fearnhead's recursion in the limit $N \to \infty$. As reviewed in Section 2.2.1, the coefficients (a_n^N), $1 \leq n \leq N$, in the finite population size case appear as factors in Equation (2.20). They somehow quantify the additional probability for the common ancestor to be of the beneficial type due to selection $(\sigma > 0)$ in comparison with the neutral case $(\sigma = 0)$. They also show that their representation $h_k^N = \psi_k^N$ with ψ given by (2.20) of the probability that the immortal line is of type 0, given the proportion of type-0 individuals in the size-N population is k/N converges to $h(x)$, as $N \to \infty$ and $k/N \to x$.

3 Looking down in the ancestral selection graph: A probabilistic approach to the common ancestor type distribution

The results presented in this chapter are based on joint work with Sandra Kluth, Ellen Baake, and Anton Wakolbinger. They are published in [LKBW15][14].

In this chapter, we deal with a (two-type) Wright-Fisher diffusion with directional selection and two-way mutation. In detail, we investigate $h(x)$, the probability that among all individuals of today's population, the individual whose progeny will eventually take over in the population is of the beneficial type, given today's frequency of the beneficial type is x. We develop a construction that allows a transparent derivation of Fearnhead's coefficients (a_n), the series coefficients of $h(x)$, and gives them a probabilistic meaning. This construction, the *pruned lookdown ancestral selection graph* (p-LD-ASG) contains elements of both the ancestral selection graph and the lookdown construction and includes pruning of certain lines upon mutation.

3.1 Introduction

The understanding of ancestral processes under selection and mutation is among the fundamental challenges in population genetics. Two central concepts are the ancestral selection graph (ASG) and the lookdown (LD) construction. The ancestral selection graph ([KN97]; [NK97]; see also [SU86] for an analogous construction in a diffusion model with spatial structure) describes the set of lines that are potential ancestors of a sample of individuals taken from a present population. In contrast, the lookdown construction [DK99b, DK99a] is an integrated representation that makes all individual lines in a population explicit, together with the genealogies of arbitrary samples. See [Eth11, Chapter 5] for an excellent overview of the area.
Both the ASG and the LD are important theoretical concepts as well as valuable tools in applications. Interest is usually directed towards the genealogy of a sample, backward in time until the most recent common ancestor (MRCA). However, the ancestral line that continues beyond the MRCA into the distant past is of considerable interest on its own, not least because it links the genealogy (of a sample from a population) to the longer time scale of phylogenetic trees. The extended time horizon then shifts attention to the asymptotic properties of the ancestral process. The stationary type distribution on the ancestral line may differ substantially from the stationary type distribution in the population. This mirrors the fact that the ancestral line consists of those individuals that are successful in the long run; thus, its type distribution is expected to be biased towards the favourable types.

[14] Remember that the scaling of this thesis (including this chapter) is chosen such that the pair coalescence rate is 1 whereas it is 2 in [LKBW15].

When looking at the evolution of the system in (forward) time $[0,\infty)$, one may ask for properties of the so-called *immortal line*, which is the line of descent of those individuals whose offspring eventually takes over the entire population. In other words, the immortal line restricted to any time interval $[0,t]$ is the common ancestral line of the population back from the far future. It then makes sense to consider the type of the immortal line at time 0. To be specific, let us consider a Wright-Fisher diffusion with two types of which one is more and one is less fit. The *common ancestor type (CAT) distribution* at time 0, conditional on the type frequencies $(x,1-x)$, then has weights $(h(x),1-h(x))$, where $h(x)$ is the probability that the population ultimately consists of offspring of an individual of the beneficial type, when starting with a frequency x of beneficial individuals at time 0.

The quantity $h(x)$ can also be understood as the limiting probability (as $s\to\infty$) that the ancestor at time 0 of an individual sampled from the population at the future time s is of the beneficial type, given that the frequency of the beneficial type at time 0 is x. Equivalently, $h(x)$ is the limiting probability (as $s\to\infty$) that the ancestor at the past time $-s$ of an individual sampled from the population at time 0 is of the beneficial type, given that the frequency of the beneficial type at time $-s$ was x.

Fearnhead [Fea02] computed the common ancestor type distribution for time-stationary type frequencies, representing it in the form $\int_0^1 (h(x),1-h(x))w(dx)$ (where w is Wright's equilibrium distribution) and calculating a recursion for the coefficients of a series representation of $h(x)$. Later, $h(x)$ has been represented in terms of a boundary value problem [Tay07, KHB13], see also Section 3.7.

In the case without mutations (in which $h(x)$ coincides with the classical fixation probability of the beneficial type starting from frequency x), Mano [Man09] and Pokalyuk and Pfaffelhuber [PP13] have represented $h(x)$ in terms of the *equilibrium ASG*, making use of a time reversal argument (see Section 3.2.2). However, the generalisation to the case with mutation is anything but obvious. One purpose of this article is to solve this problem. A key ingredient will be a combination of the ASG with elements of the lookdown construction, which also seems of interest in its own right.

The chapter is organised as follows. In Section 2, we start by briefly recapitulating the ASG (starting from the Moran model for definiteness). We then recall the Fearnhead-Taylor representation of $h(x)$ and give its explanation in terms of the equilibrium ASG in the case without mutations, inspired by [PP13]. In Section 3, we prepare the scene by ordering the lines of the ASG in a specific way; in Section 4, we then represent the ordered ASG in terms of a fixed arrangement of levels, akin to a lookdown construction. In Section 5, a pruning procedure is described that reduces the number of lines upon mutation. The stationary number of lines in the resulting pruned LD-ASG will provide the desired connection to the (conditional) common ancestor type distribution. Namely, the tail probabilities of the number of lines appear as the coefficients in the series representation. In Section 6, the graphical approach will directly reveal various monotonicity properties of the tail probabilities as functions of the model parameters, which translate into monotonicity properties of the common ancestor type distribution. Section 7 is an add-on, which makes the connection to Taylor's boundary value problem for $h(x)$ explicit; Section 8 contains some concluding remarks.

3.2 Concepts and models

3.2.1 The Moran model and its diffusion limit

Let us consider a haploid population of fixed size $N \in \mathbb{N}$ in which each individual is characterised by a type $i \in S := \{0,1\}$. An individual of type i may, at any instant in continuous time, do either of two things: it may reproduce, which happens at rate $1/2$ if $i = 1$ and at rate $1/2 + s_N$, $s_N \geq 0$, if $i = 0$; or it may mutate to type j at rate $u_N v_j$, $u_N \geq 0$, $0 \leq v_j \leq 1$, $v_0 + v_1 = 1$. If an individual reproduces, its single offspring inherits the parent's type and replaces a randomly chosen individual, maybe its own parent. Concerning mutations, u_N is the total mutation rate and v_j the probability of a mutation to type j. Note that the possibility of silent mutations from type j to type j is included.

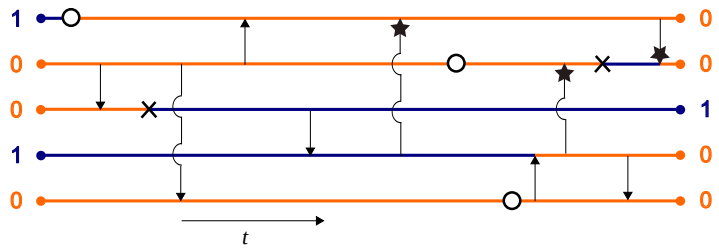


Figure 3.1: *The Moran model with two-way mutation and selection. The types are indicated for the initial population (left) and the final one (right). Crosses represent mutations to type* 1*, circles mutations to type* 0*. Selective events are depicted as arrows with star-shaped heads.*

The Moran model has a well-known graphical illustration as an interacting particle system (cf. Fig. 3.1). The individuals are represented by horizontal line pieces, with forward time running from left to right in the figure. Arrows indicate reproduction events with the parent at its tail and the offspring at its head. For later use, we decompose reproduction events into neutral and selective ones. Neutral arrows appear at rate $1/(2N)$, selective arrows (those with a star-shaped arrowhead in Fig. 3.1) at rate s_N/N per ordered pair of lines, irrespective of their types. The rates specified above are obtained by the convention that neutral arrows may be used by all individuals, whereas selective arrows may only be used by type-0 individuals and are ignored otherwise. Mutations to type 0 are marked by circles, mutations to type 1 by crosses.

The usual diffusion rescaling in population genetics is applied, i.e. rates are rescaled such that $\lim_{N\to\infty} N s_N = \sigma$ and $\lim_{N\to\infty} N u_N = \theta$, $0 \leq \sigma, \theta < \infty$, and time is sped up by a factor of N. Let X_t be the frequency of type-0 individuals at time t in this diffusion limit. Then, the process $(X_t)_{t\in\mathbb{R}}$ is a Wright-Fisher diffusion which is characterised by the drift coefficient $a(x) = (1-x)\theta v_0 - x\theta v_1 + x(1-x)\sigma$ and the diffusion coefficient $b(x) = x(1-x)$. The stationary density w is given by $w(x) = C(1-x)^{2\theta v_1 - 1} x^{2\theta v_0 - 1} \exp(2\sigma x)$, where C is a normalising constant (cf. [Dur08, Chapters 7, 8] or [Ewe04, Chapters 4, 5]).

3.2.2 The ancestral selection graph

The *ancestral selection graph* was introduced by Krone and Neuhauser [KN97, NK97] to construct samples from a present population, together with their ancestries, in the diffusion limit of the Moran model with mutation and selection. The basic idea is to understand selective arrows as unresolved reproduction events backward in time: the descendant has two *potential ancestors*, the *incoming branch* (at the tail) and the *continuing branch* (at the tip), see also Fig. 3.2. The incoming branch is the ancestor if it is of type 0, otherwise the continuing one is ancestral. For a hands-on exposition, see [Wak09, Chapter 7.1].

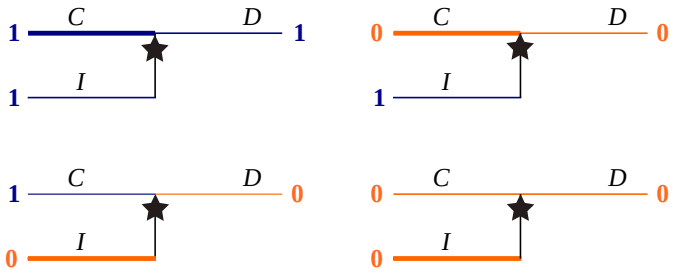


Figure 3.2: *Incoming branch (I), continuing branch (C), and descendant (D). The ancestor is marked bold.*

The ASG is constructed by starting from the (as yet untyped) sample and tracing back the lines of all potential ancestors. In the finite graphical representation, a neutral arrow that joins two potential ancestral lines appears at rate $1/N$ per currently extant pair of potential ancestral lines, then giving rise to a *coalescence event*, i.e. the two lines merge into a single one. In the same finite setting, a selective arrow that emanates from outside the current set of n potential ancestral lines and hits this set appears at rate $n(N-n)s_N/N$. This gives rise to a *branching event*, i.e., viewed backward in time, the line that is hit by the selective arrow splits into an incoming and continuing branch as described above. Thus, in the diffusion limit, since $N-n \sim N$ as $N \to \infty$, the process $(K_r)_{r\in\mathbb{R}}$, where K_r is the number of lines in the ASG at time $r=-t$, evolves backward in time with rates

$$q_K(n,n-1) = \frac{1}{2}n(n-1), \quad q_K(n,n+1) = n\sigma, \qquad n \in \mathbb{N}. \tag{3.1}$$

At a coalescence event a randomly chosen pair of lines coalesces, while at a branching event a randomly chosen line splits into two.
The (reversible) equilibrium distribution of the dynamics (3.1) turns out to be the Poisson(2σ)-distribution conditioned to $\{1,2,\dots\}$, i.e.

$$\mathbb{P}(K_r = n) = \frac{(2\sigma)^n}{n!(\exp(2\sigma)-1)}, \quad n \in \mathbb{N}. \tag{3.2}$$

We may construct the *equilibrium ASG* as in [PP13] in two stages: first take a random path $(K_r)_{-\infty<r<\infty}$, and then fill in the branching and coalescence events, with a random choice of one of the K_r lines at each upward jump, and of one of the $\binom{K_r}{2}$ pairs at each downward jump of (K_r).

Mutation events (at rates $\theta\nu_0$ and $\theta\nu_1$) are superposed on the lines of the (equilibrium) ASG by Poisson processes with rates $\theta\nu_0$ and $\theta\nu_1$. Given the frequency x of the beneficial type at time 0, one then assigns types to the lines of the ASG in the (forward) time interval $[0,\infty)$ by first drawing the types of the lines at time 0 independently and identically distributed (i.i.d.) from the weights $(x,1-x)$, and propagates the types forward in time, respecting the mutation events. In this way, the (backward in time) branching events may now be resolved into the true parent and a fictitious parent.
Note that there are various ways to illustrate the same realisation of the ASG graphically. See, for instance, Fig. 3.3, with backward time r running from right to left. The left and right panels of Fig. 3.3 represent the same realisation of the ASG, but differ in the ordering of the lines.

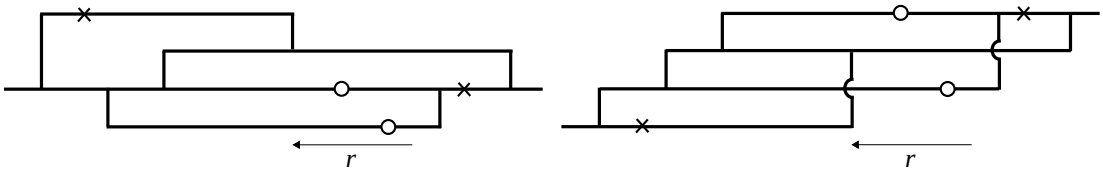


Figure 3.3: *Different representations of the same ASG realisation with superimposed mutation. All potential ancestors of the line next to the top in Fig. 3.1 are shown (before resolution into true and fictitious parents).*

3.2.3 The common ancestor

In the population, at any time t, there almost surely exists a unique individual that is, at some time $s > t$, ancestral to the whole population; cf. Fig. 3.4. The descendants of this individual become fixed, and we call it the *common ancestor at time t*. The lineage of these distinguished individuals over time defines the so-called *ancestral* (or *immortal*) *line*.

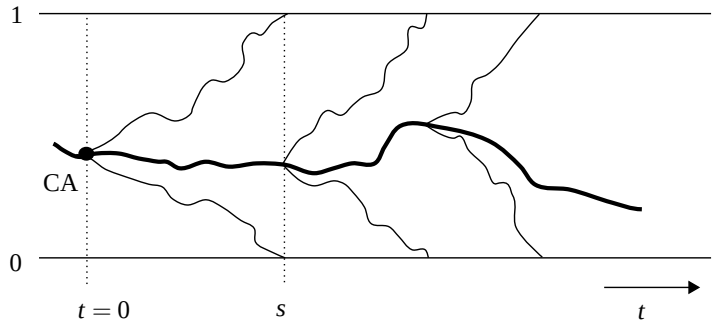


Figure 3.4: *The common ancestor at time $t = 0$ (CA) is the individual whose progeny will eventually fix in the population (at time s).*

Looking at the population at time t, say $t = 0$, we are interested in $h(x)$, the probability that the common ancestor is of type 0, given $X_0 = x$. Equivalently, one may understand $h(x)$ as the probability that the offspring of all type-0 individuals (regardless of the offspring's types) will ultimately be

ancestral to the entire population, if $X_0 = x$. The probability $h(x)$ does depend on the type-0 frequency x at that time but not on the time itself. According to previous results by Fearnhead [Fea02] and Taylor [Tay07], it reads

$$h(x) = \sum_{n\geq 0} a_n x(1-x)^n, \tag{3.3}$$

where the coefficients a_n, $n \geq 0$, are characterised by the recursion

$$(n+1+2\theta\nu_1)a_{n+1} - (n+1+2\sigma+2\theta)a_n + 2\sigma a_{n-1} = 0, \quad n \geq 1 \tag{3.4}$$

under the constraints

$$a_0 = 1, \qquad \lim_{n\to\infty} \frac{a_{n+1}}{a_n} = 0. \tag{3.5}$$

Also, it is shown in [Fea02] that (3.4) and (3.5) imply

$$1 = a_0 \geq a_1 \geq \cdots, \qquad \lim_{n\to\infty} a_n = 0. \tag{3.6}$$

These results were reviewed by Baake and Bialowons in 2008 [BB08], and re-obtained with the help of a *descendant process* (forward in time) by Kluth, Hustedt, and Baake in 2013 [KHB13].
Eq. (3.3) quantifies the bias towards type 0 on the immortal line. The $n = 0$ term on the right-hand side of (3.3) is $a_0 x = x$, which coincides with the fixation probability in the neutral case ($\sigma = 0$). Indeed, for $\sigma = 0$, we have $a_0 = 1$, $a_i = 0$ for $i \geq 1$ (this is easily seen to satisfy (3.3) and (3.4)). For $\sigma > 0$, however, all a_i are positive (again by inspection of (3.3) and (3.4)), and the terms for $n \geq 1$ in (3) quantify the long-term advantage of the favourable type.
In order to get a handle on the representation (3.3) and the recursion (3.4) in terms of the equilibrium ASG, one observes that the type of the common ancestor at time $t = 0$ may be recovered in the following way. In the equilibrium ASG marked with the mutation events (as described in Section 3.2.3), assign i.i.d types to the lines at time 0 and propagate them forward in time, respecting the mutation events. The immortal line is then encoded in the realisation of the marked ASG.
The event of fixation of the beneficial type is easily described in the case without mutations. First, recall that, as stated in Section 3.2.3, the number K_0 of lines in the equilibrium ASG at time 0 is Poisson distributed with parameter 2σ, conditioned to be positive. Next, observe that, with probability 1, the equilibrium ASG has bottlenecks, i.e. times at which it consists of a single line. Let t_0 be the smallest among all the non-negative times at which there is a bottleneck (see Fig. 3.5). This way, the unique individual is identified that is the true ancestor of the single individual at forward time t_0 and, at the same time, of the entire equilibrium ASG at any later time (and ultimately of the entire population).
As observed by Mano [Man09] and Pokalyuk and Pfaffelhuber [PP13], type 0 becomes fixed if and only if the single line at time t_0 carries type 0, and this, in turn, happens if and only if at least one ancestral line at time $t = 0$ is of type 0. The latter probability is $1-(1-x)^{K_0}$, given that the frequency of type-0 individuals is x at this time. Therefore, with the help of (3.2), the fixation probability can be obtained as

$$h(x) = \mathbb{E}\big(1-(1-x)^{K_0}\big) = \frac{1}{\exp(2\sigma)-1} \sum_{n\geq 1} \big(1-(1-x)^n\big) \frac{(2\sigma)^n}{n!} = \frac{1-\exp(-2\sigma x)}{1-\exp(-2\sigma)}, \tag{3.7}$$

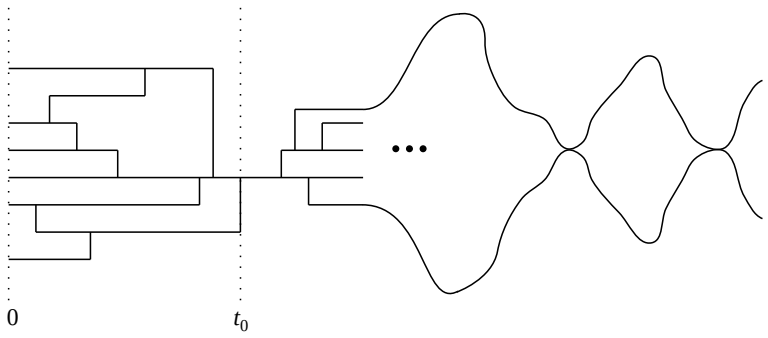


Figure 3.5: *A realisation of the equilibrium ASG, with its first bottleneck (after time* 0*) at time* t_0.

which coincides with the classical result of Kimura [Kim62]. Putting $\gamma_n := \mathbb{P}(K_0 > n)$, $n \geq 0$, the left-hand side of (3.7) may also be expressed as

$$h(x) = \sum_{n\geq 1} \big(1-(1-x)^n\big)\big[\mathbb{P}(K_0 \geq n) - \mathbb{P}(K_0 \geq n+1)\big] = \sum_{n\geq 0} \gamma_n x(1-x)^n,$$

which is the representation (3.3). (Indeed, one checks readily that the tail probabilities γ_n satisfy the recursion (3.4) in the case $\theta = 0$.) The elegance of this approach lies is the fact that one does not need to know the full representation of the ASG, in particular one does not need to distinguish between incoming and continuing branches. As soon as mutations are included, however, keeping track of the hierarchy of the branches becomes a challenge. We thus aim at an alternative representation of the ASG that allows for an orderly bookkeeping leading to a generalisation of the idea above, and yields a graphical interpretation of (3.3)-(3.5). This will be achieved in the next three sections.

3.3 The ordered ASG

In the previous section, we have reminded ourselves that one may represent the same realisation of one ASG in different ways. In the following, we propose a construction, which we call the *ordered ASG*, and which is obtained backward in time from a given realisation of the ASG as follows (compare Fig. 3.6).

- *Coalescence:* Each coalescence event is represented by a (neutral) arrow pointing from the lower participating line to the upper one. The (single) parental line continues back in time from the lower branch.
- *Branching:* A selective arrow with star-shaped head is pointed towards the splitting line at a branching event. The incoming branch is always placed directly beneath the continuing branch at the tail of the arrow; in particular, there are no lines between incoming and continuing branch at the time of the branching event.
- *Mutation:* Mutations, symbolised here by circles and crosses, occur along the lines as in the original ASG.

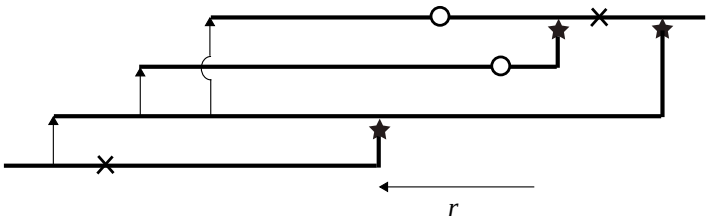


Figure 3.6: *The ordered ASG corresponding to Fig. 3.3 or directly deduced from Fig. 3.1.*

The ordered ASG corresponding to both representations in Fig. 3.3 is shown in Fig. 3.6.

3.4 The lookdown ASG

To each point in the ordered ASG, let us introduce two coordinates: its *time* and its *level*, the latter being an element of $\{1,2,\ldots\}$ which coincides with the number of lines in the ASG. Since this is in close analogy to ideas known from the lookdown processes by Donnelly and Kurtz [DK99b, DK99a], we call this construction the *lookdown ASG (LD-ASG)*. It can be obtained backward in time from a given realisation of the ordered ASG, or it may as well be constructed in distribution via Poissonian elements representing coalescence, branching, and mutation. The two possibilities are described in Sections 3.4.1 and 3.4.2, respectively.

3.4.1 Construction from a given realisation of the ordered ASG

Backward in time, the realisation of the LD-ASG corresponding to a given realisation of the ordered ASG is obtained in the following way. Start with all n individuals (respectively lines) that are present in the (ordered) ASG and place them at levels 1 to n by adopting their vertical order from the ordered ASG. Then let the following events happen (backward in time):

- *Coalescence:* Coalescence events between levels i and $j > i$ are treated the same way as in the ordered ASG: The remaining branch continues at level i. In addition, all lines at levels $k > j$ are shifted one level downwards to $k-1$ (cf. Fig. 3.7, left).

- *Branching:* A selective arrow with star-shaped head in the ordered ASG is translated into a star at the level i of the branching line. The incoming branch emanates out of the star at the same level and all lines at levels $k \geq i$ are pushed one level upwards to $k+1$. In particular, the continuing branch is shifted to level $i+1$ (cf. Fig. 3.7, right).

- *Mutation:* Mutations (symbolised again as circles and crosses) are taken from the ordered ASG.

Fig. 3.8 gives a realisation that corresponds to the realisation of the ordered ASG in Fig. 3.6. Note that we obviously have a bijection between realisations of the ordered ASG and the LD-ASG and that the LD-ASG is just a neat arrangement of the ordered ASG.

Figure 3.7: *Coalescence (left) and branching event (right) in the LD-ASG.*

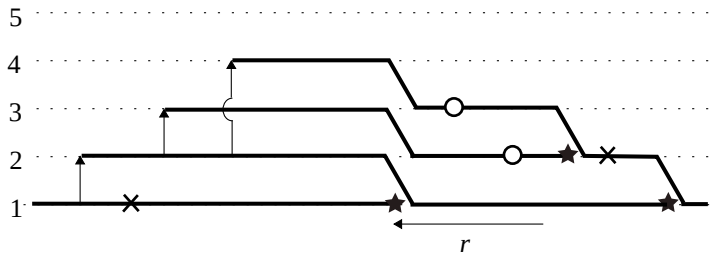


Figure 3.8: *LD-ASG corresponding to Fig. 3.6. Levels are numbered from bottom to top.*

3.4.2 Construction from elements of Poisson point processes

The LD-ASG may, in distribution, as well be constructed backward in time via the elements 'arrows', 'stars', 'circles', and 'crosses' arising as representations of independent Poisson point processes:

- *Coalescence:* For each ordered pair of levels (i, j), where $i < j$ and level j is occupied by a line, *arrows* from level i to j emerge independently according to a Poisson point process $\Gamma^{\uparrow}_{ij}$ at rate 1. An arrow from i to j is understood as a coalescence of the lines at levels i and j to a single line on level i. In addition, all lines at levels $k > j$ are shifted one level downwards to $k-1$ (cf. Fig. 3.7, left).

- *Branching:* On each occupied level i *stars* appear according to independent Poisson point processes Γ^{*}_{i} at rate σ. A star at level i indicates a branching event, where a new line, namely the incoming branch, is inserted at level i and all lines at levels $k \geq i$ are pushed one level upwards to $k+1$. In particular, the continuing branch is shifted to level $i+1$ (cf. Fig. 3.7, right).

- *Mutation:* Mutations to type 0 and type 1, i.e. circles and crosses, occur via independent Poisson point processes Γ°_{i} at rate $\theta\nu_0$ and $\Gamma^{\times}_{i}$ at rate $\theta\nu_1$, respectively, on each occupied level i.

The independent superposition of these Poisson point processes and their effects on the lines characterises the LD-ASG. Recall that $(K_r)_{r\in\mathbb{R}}$ is the line counting process of the (ordered) ASG and thus K_r is also the highest occupied level of the LD-ASG at time r. It evolves backward in time with transition rates given by (3.1).
Note that, although we will ultimately rely on the ASG in equilibrium only, neither the ordering of the ASG nor the LD-ASG construction are restricted to the equilibrium situation. The equilibrium comes back in when we search for the immortal line, which will be done next.

3.4.3 The immortal line in the LD-ASG in the case without mutations

We consider a realisation $\mathscr{G}$ of the equilibrium LD-ASG, write K_r for its highest occupied level at (backward) time r, and again write

$$t_0 = t_0(\mathscr{G}) := -\sup\{r \leq 0 : K_r = 1\} \tag{3.8}$$

for the smallest (forward) time at which $\mathscr{G}$ has a 'bottleneck', see Fig. 3.9.

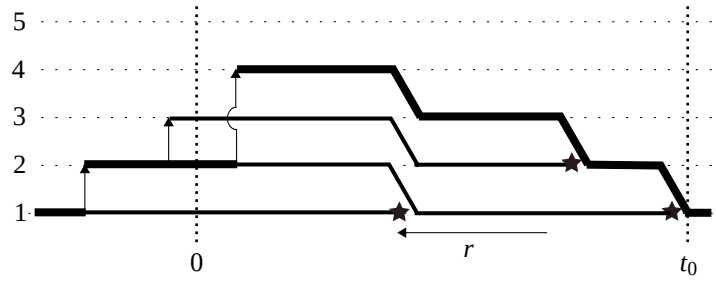


Figure 3.9: *LD-ASG (without mutations) corresponding to Fig. 3.6. The immune line is marked bold.*

The level of the immortal line at time 0 does not only depend on $\mathscr{G}$, but also on the types $I^1, \ldots, I^{K_0} \in \{0,1\}$ that are assigned to the levels $1, 2, \ldots, K_0$ at time 0. We now define a distinguished line which we call the *immune line*. The reason for this naming will become clear in the next section: the immune line will be exempt from the pruning.

Definition 3.1 *At any given time, the* immune line *is the line that will be immortal if all lines at that time are of type 1.*

The following is immediate from the construction of the LD-ASG: back from each bottleneck of $\mathscr{G}$, the immune line goes up one level at each branching event that happens at a level smaller or equal to its current level, and follows the coalescence events in a lookdown manner, see the bold line in the right panel of Fig. 3.10. In particular, the immune line follows the continuing branch whenever it is hit by a branching event at its current level.
The next proposition is illustrated by Fig. 3.10.

Proposition 3.2 *In the absence of mutations, for almost every realisation $\mathscr{G}$ of the equilibrium LD-ASG with types assigned at time* 0*, the level of the immortal line at time* 0 *is either the lowest type-*0 *level at time* 0 *or, if all lines at time* 0 *are of type* 1*, it is the level of the immune line at time* 0*.*

Proof. We proceed by induction along the Poissonian elements "branching" and "coalescence" described in Sec. 3.4.2, backward from t_0, the first time after time 0 at which the number of lines in $\mathscr{G}$ is one. Let $t_k < 0 < t_{k-1} < \cdots < t_0$ be the times at which these elements occur (note that for almost every realisation $\mathscr{G}$ the number k is finite, and all the t_j are distinct) and choose times $0 =: s_k < \cdots < s_0$ with $t_j < s_j < t_{j-1}$, $1 \leq j \leq k$. We claim that for all $j = 1, \ldots, k$, when assigning types at time s_j, the

level of the immortal line at time s_j is either the lowest type-0 level at time s_j or, if all lines at time s_j are of type 1, it is the level of the immune line at time s_j.
The assertion is obvious for $j = 1$, since by assumption no event has happened between s_1 and t_0. Now consider the induction step from j to $j+1$. If all lines at time s_{j+1} are assigned type 1, then the level of the immortal line at time s_{j+1} is by definition that of the immune line at time s_{j+1}. Now assume that at least one line at time s_{j+1} is assigned type 0. If the event at time t_j was a coalescence (such as the leftmost event in Fig. 3.10), then by the induction assumption the level of the immortal line at time s_{j+1} is the lowest type-0 level at time s_{j+1} (this is because the types are propagated along the lines, in particular along the line at the tail of the arrow). If, on the other hand, the event at time

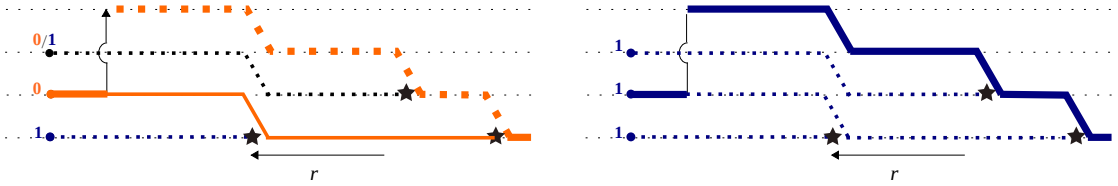


Figure 3.10: *LD-ASG with types. The level of the immortal line (solid) starting out from time* 0 *depends on the type assignment at time* 0*. The immune line is marked bold (and is the immortal line in the right picture).*

t_j was a branching, then again by the induction assumption and by the "pecking order" illustrated in Fig. 3.11, the level of the immortal line at time s_{j+1} is the lowest type-0 level at time s_{j+1}. This completes the proof of the proposition. □

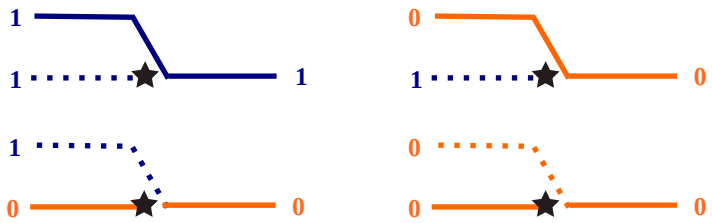


Figure 3.11: *Branching event in the LD-ASG. The four possible combinations of types are shown (in analogy with Fig. 3.2). The parental branch (bold line) is the incoming one (upper two diagrams) if it is of type* 0*, and the continuing one (lower two diagrams) if the incoming branch is of type* 1*.*

Proposition 3.2 specifies the immortal line in the case of selection only. The aim in the next section is to establish an analogous statement in the case of selection *and* mutation.

3.5 The pruned equilibrium LD-ASG and the CAT distribution

We now consider the equilibrium LD-ASG marked with the mutation events. Working backward from the bottleneck time t_0 (see Eq. (3.8) and Figs. 3.8 and 3.9), we see that the mutation events that

occur along the lines may eliminate some of them as candidates for being the immortal line. Cutting away certain branches that carry no information has been used, explicitly or implicitly, in various investigations of the ASG (e.g. by Slade [Sla00], Fearnhead [Fea02], Athreya and Swart [AS05], and Etheridge and Griffiths [EG09]), but our construction requires a specific *pruning procedure* which we now describe.

3.5.1 Pruning the LD-ASG

In addition to the events *branching* and *coalescence* we now have the *deleterious mutations* and the *beneficial mutations*. As in the proof of Proposition 3.2, let $t_k < 0 < t_{k-1} < \cdots < t_0$ be the times at which all these events occur, and choose times $0 =: s_k < \cdots < s_0$ with $t_j < s_j < t_{j-1}$, $1 \leq j \leq k$. Assume branching and coalescence events (but no mutation events) happen at the times $t_1, \ldots, t_{i-1}$, and a mutation event happens at time t_i. Recall from Definition 1 that at any given time the immune line is the line that will be immortal if all lines at that time are of type 1; but now the rule for the immune line must be adapted due to the impact of mutations. First consider the case in which our first mutation is deleterious (symbolised by a cross). Since there is no mutation between times t_i and t_0, Proposition 3.2 applies (with time 0 replaced by time s_{i+1}), showing that the line that is hit by the deleterious mutation at time t_i cannot be the immortal one unless it is the immune line. In our search for the true ancestor of the line that goes back from time t_0 we can therefore *erase* the line segment to the left of t_i, unless the line in question is the immune one; all lines above the one that is erased slide down one level to fill the space, see Fig. 3.12 (left). If the immune line is hit by a deleterious mutation at t_i, it is the immortal line at time s_{i+1} if and only if all the other lines at time s_{i+1} are of type 1. In order to tie in with our picture that the level of the immortal line at any time is the lowest level that is assigned type 0 (given there is at least one lineage at this time that is assigned type 0), we *relocate* our mutated immune line to the currently highest level of the LD-ASG at time t_i, whereby the levels of all the other lines that were above the immune line at time s_i are shifted down by 1 (compare Fig. 3.13).
Next consider the case in which the mutation occurring at time t_i is beneficial (symbolised by a circle) and happens at level ℓ, say. Then, again appealing to Proposition 3.2, we see that none of the lines that occupy levels $> \ell$ at time t_i can be parental to the single line that exists at time t_0. We can therefore erase all these lines from the list of candidates for the ancestors. We indicate this by inserting a barrier of infinite height above the circle, see Fig. 3.12 (right). If all lines at time s_{i+1} are assigned type 1, then the line on level ℓ becomes the only one that carries type 0 at time s_i and therefore is immortal. Thus, the immune line is relocated to level ℓ at time t_i.

Figure 3.12: *Pruning procedure in the LD-ASG due to deleterious (left) and beneficial (right) mutations that appear on lines that are not immune.*

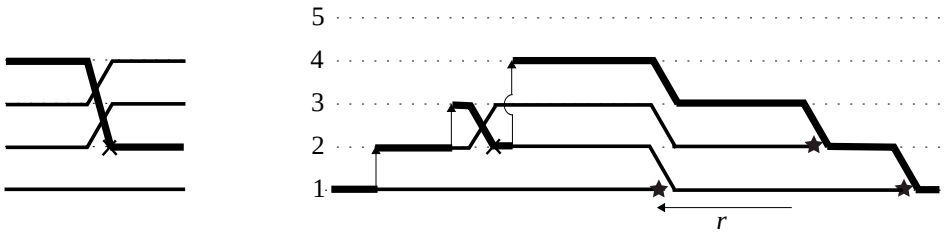


Figure 3.13: *Relocation procedure in the LD-ASG due to deleterious mutations on the immune line (bold).*

Proceeding to the next mutation event on the remaining lines back from time t_i (which happens at time t_m at level ℓ', say), we can iterate this procedure: if the mutation is deleterious, the line is killed unless it is the immune one. If the immune line is hit by a deleterious mutation, then it is relocated to the currently highest level of the LD-ASG. If the mutation is beneficial, all the lines at higher levels are killed, with the line starting back from t_m at level ℓ' being declared the new immune line.
Having worked back to $t = 0$, we arrive at the *pruned LD-ASG* between times 0 and t_0. Note that except for the immune line, due to the pruning procedure there are no mutations on any line of the *pruned LD-ASG*. In other words, each line present at time 0 is either the immune line, or has no mu-

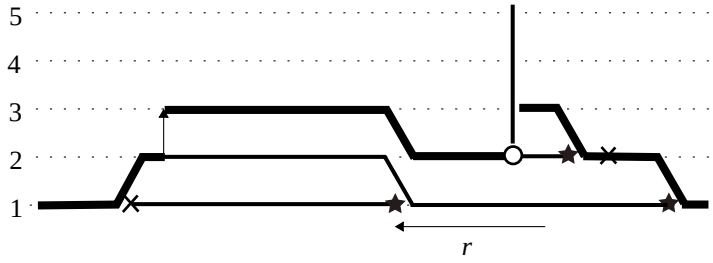


Figure 3.14: *Pruned LD-ASG derived from Fig. 3.8. The immune line is marked bold.*

tations on it between times 0 and ζ, where ζ is the time when that line was incoming to a branching event with the immune line. Note also that beneficial mutations can only be present on the current top level of the pruned LD-ASG.

As in Section 3.4.2 we can construct the pruned LD-ASG (together with the level of the immune line) backward in time in a Markovian way, using the Poisson processes $\Gamma_{ij}^{\uparrow}$ and Γ_i^* (for all occupied levels i and j, cf. Fig. 3.7), and $\Gamma_i^{\times}$ and Γ_i°, where the pruning procedure is applied as described above (cf. Fig. 3.12). Fig. 3.14 gives a realisation that corresponds to the realisation of the LD-ASG in Fig. 3.8.

3.5.2 The line-counting process of the pruned LD-ASG

The construction of the previous subsection shows that the process $(L_r)_{r\in\mathbb{R}}$, where L_r is the *level of the top line* (which coincides with the number of lines) at the backward time $r = -t$, evolves with

transition rates

$$\begin{aligned} q_L(n,n-1) &= \frac{1}{2}n(n-1)+(n-1)\theta\nu_1+\theta\nu_0, \\ q_L(n,n+1) &= n\sigma, \\ q_L(n,n-\ell) &= \theta\nu_0, \qquad\qquad 2\le\ell<n, \quad n\in\mathbb{N}. \end{aligned} \tag{3.9}$$

In words, when the top level is currently n, it decreases by one when either a coalescence event happens between any pair of lines (rate $n(n-1)/2$), or one of the $n-1$ lines that are not immune experiences a deleterious mutation (rate $(n-1)\theta\nu_1$), or line $n-1$ experiences a beneficial mutation (rate $\theta\nu_0$). The top level increases by one when one of the lines branches (rate $n\sigma$). It decreases by ℓ, $2\le\ell<n$, when level $n-\ell$ experiences a beneficial mutation (rate $\theta\nu_0$) .

Remark 3.3 *The process L is stochastically dominated by the process K (the highest level of the unpruned LD-ASG). In fact, using the above-described pruning procedure in a time-stationary picture between all the bottlenecks of the equilibrium ASG line counting process $K=(K_r)_{r\in\mathbb{R}}$, we obtain that $L_r\le K_r$ for all $r\in\mathbb{R}$.*

The stochastic dynamics induced by (3.9) thus has a unique equilibrium distribution, which we denote by ρ. In the following, let $L=(L_r)_{r\in\mathbb{R}}$ be the time-stationary process with jump rates (3.9). We then have

$$\rho_n=\mathbb{P}(L_0=n), \quad n\in\mathbb{N}, \tag{3.10}$$

and $\rho=(\rho_n)$ is the probability vector obeying

$$\rho Q=0, \tag{3.11}$$

with Q being the generator matrix determined by the jump rates (3.9).

3.5.3 The type of the immortal line in the pruned LD-ASG

We will now show that the type of the immortal line at time 0 is determined by the type configuration assigned at time 0 in a way quite similar to the case without mutations.

Theorem 3.4 *For almost every realisation $\mathcal{G}$ of the pruned equilibrium LD-ASG with types assigned at time 0, the level of the immortal line at time 0 is either the lowest type-0 level at time 0 or, if all lines at time 0 are of type 1, it is the level of the immune line at time 0. In particular, the immortal line is of type 1 at time 0 if and only if all lines in $\mathcal{G}$ at time 0 are assigned the type 1.*

Proof. We proceed by induction along the Poissonian elements described in Sec. 3.5.1 backward from t_0, the first time after time 0 at which the number of lines in $\mathcal{G}$ is one. As in Sec. 3.5.1, let $t_k<0<t_{k-1}<\cdots<t_0$ be the times at which these elements occur, and choose times $0=:s_k<\cdots<s_0$ with $t_j<s_j<t_{j-1}$, $1\le j\le k$. We will prove that for all $j=1,\dots,k$ the level of the immortal line at time s_j is either the lowest type-0 level at time 0 or, if all lines at time s_j are of type 1, it is the level of the immune line at time s_j.
If the event that occurs at t_j is a branching or a coalescence, then the induction step from j to $j+1$ is precisely as in the proof of Proposition 3.2. If the event at time t_j is a deleterious mutation, then we

distinguish two cases. In the first case, assume that this deleterious mutation happens on the immune line. Then, according to the rule prescribed in Sec. 3.5.1, the immune line is relocated to the top level of $\mathscr{G}$ at time t_j (compare Fig. 3.13). If there is a line that is assigned type 0 at time s_{j+1}, then the immortal line at time s_{j+1} is found at the lowest type-0 level. (In particular, if the uppermost line at time s_{j+1} is assigned type 0, then it is the immortal line if and only if all other lines at time s_{j+1} are assigned type 1. This is due to the deleterious mutation at time t_j and the relocation to the top level.) In the second case assume that the deleterious mutation happens on a line different from the immune one. Then the number of lines at time s_{j+1} is one less than the number of lines at time s_j; more specifically, all the lines from time s_{j+1} can be found in the same order also at time s_j, and in addition at time s_j there is one line carrying type 1 which cannot be ancestral since it is not the immune line. Thus the induction assumption from time s_j carries over to give the required assertion for time s_{j+1}. Finally, assume that the event at time t_j is a beneficial mutation. In this case, if at time s_{j+1} type 0 is assigned to a level ℓ that is occupied by one of the lines that remain after the pruning at time t_j, and if all the levels below ℓ are assigned type 1, then we can infer from the induction assumption that the line at level ℓ at time s_{j+1} is the immortal one. On the other hand, if all the lines that remain at time s_{j+1} are assigned type 1, then, because of the beneficial mutation at time t_j, the top line at time s_{j+1} is the immortal one, and due to our relocation rule this is also the immune line at time s_{j+1}. Thus, the induction step is completed, and the theorem is proved. □

3.5.4 The CAT distribution via the pruned equilibrium LD-ASG

With the help of Theorem 3.4, it is now possible to provide an interpretation of the probability $h(x)$ that the common ancestor is of type 0, given that the frequency of the beneficial type at time 0 is x.

Theorem 3.5 *Given the frequency of the beneficial type at time* 0 *is x, the probability that the common ancestor at time* 0 *is of beneficial type is*

$$h(x) = \sum_{n\geq 1} x(1-x)^{n-1}\mathbb{P}(L_0 \geq n), \tag{3.12}$$

where L_0 is the number of lines at time 0 *in the time-stationary pruned LD-ASG, see formula* (3.10).

Proof. Let $I^k \in \{0,1\}$ be the type that is assigned to the individual at level $k \in \{1,\dots,L_0\}$ in the pruned equilibrium LD-ASG at time 0. According to Theorem 3.4, the event that the common ancestor at time 0 is of type 0 equals the event that at least one of the I^k, $k \in \{1,\dots,L_0\}$, is 0. Conditional on the initial frequency of the beneficial type being x, these types are assigned in an i.i.d. manner with $\mathbb{P}(I^k = 0) = x$. The quantity $h(x)$ thus is the probability that at least one of a random number of i.i.d. trials is a success, where the success probability is x in a random number of trials L_0 (which is independent of the Bernoulli sequence with parameter x). A decomposition of $h(x)$ according to the first level which is occupied by type 0 yields

$$\begin{aligned} h(x) &= \sum_{n\geq 1} \mathbb{P}(I^n = 0,\ I^k = 1\ \forall k < n,\ L_0 \geq n) \\ &= \sum_{n\geq 1} \mathbb{P}(I^n = 0,\ I^k = 1\ \forall k < n)\ \mathbb{P}(L_0 \geq n). \end{aligned} \tag{3.13}$$

The right hand side of (3.13) equals that of (3.12), which completes the proof of the theorem. □

To compare (3.12) with (3.3), we rewrite its right-hand side as $\sum_{n\geq 0} x(1-x)^n \mathbb{P}(L_0 \geq n+1)$. It is then clear from the comparison with (3.3) that the tail probabilities $\alpha_n := \mathbb{P}(L_0 > n), n \geq 0$, agree with Fearnhead's coefficients a_n. They must therefore obey the recursion (3.4). The proof of the following proposition gives a direct argument for this.

Proposition 3.6 *The tail probabilities $\alpha_n = \mathbb{P}(L_0 > n)$, $n \geq 0$, obey the recursion* (3.4).

Proof. Let $\rho = (\rho_n)$ be the probability vector determined by (3.11). We then have

$$\alpha_n = \sum_{i>n} \rho_i, \qquad n \in \mathbb{N}_0 \,. \tag{3.14}$$

For $n \geq 2$, the n^{th} entry of the vector ρQ is

$$\begin{aligned}(\rho Q)_n = \rho_{n-1} q_L(n-1,n) + \rho_{n+1} q_L(n+1,n) + \sum_{j\geq n+2} \rho_j q_L(j,n) \\ - \rho_n \left[q_L(n,n-1) + q_L(n,n+1) + \sum_{\ell=0}^{n-2} q_L(n,\ell) \right].\end{aligned}$$

Thus, plugging in the jump rates (3.9), Eq. (3.11) is equivalent to

$$\begin{aligned}0 = \rho_{n-1}(n-1)\sigma + \rho_{n+1}\left[\frac{1}{2}n(n+1) + n\theta\nu_1\right] + \theta\nu_0 \sum_{j\geq n+1} \rho_j \\ - \rho_n \left[\frac{1}{2}n(n-1) + (n-1)\theta + n\sigma\right], \quad n \geq 2.\end{aligned}$$

Writing this in terms of the tail probabilities (3.14), rearranging terms, and shifting the index, we obtain

$$\begin{aligned}0 = n\{-\alpha_n[n+1+2\theta+2\sigma] + \alpha_{n+1}[n+1+2\theta\nu_1] + \alpha_{n-1}2\sigma\} \\ + (n+1)\{\alpha_{n+1}[n+2+2\theta+2\sigma] - \alpha_{n+2}[n+2+2\theta\nu_1] - \alpha_n 2\sigma\}, \quad n \geq 1,\end{aligned}$$

which we abbreviate by

$$n(\alpha F)_n = (n+1)(\alpha F)_{n+1}$$

with the (tridiagonal) matrix F that appears in the recursion (3.4). In view of these equalities, the proposition is proved if we can show that $\lim_{n\to\infty} n(\alpha F)_n = 0$, or, even better, that

$$\lim_{n\to\infty} n^2 \alpha_n = 0. \tag{3.15}$$

To see (3.15), recall that as stated in Remark 3.3, L_0 is stochastically dominated by the number K_0 of lines in the equilibrium ASG, which has distribution (3.2). In particular, L_0 has a finite third moment. From this, (3.15) is immediate since for any non-negative integer-valued random variable ξ one has $\mathbb{E}[\xi(\xi-1)(\xi-2)] = 3\sum_{n=0}^{\infty} n(n-1)\mathbb{P}(\xi > n)$. Thus, Proposition 3.6 is proved. □

The proof of Proposition 3.6 allows us to conclude the following

Corollary 3.7 *The solution* (a_n) *of* (3.4) *and* (3.5) *is also characterised by* (3.4) *together with the constraint* (3.6)*.*

Proof. The constraint (3.6) ensures that $r_n := a_{n-1} - a_n$, $n \geq 1$, r_n are probability weights on $\mathbb{N}$. Starting from (3.4) and working back in the proof of Proposition 3.6 we arrive at Eq. (3.11), which (by irreducibility and recurrence of Q) has a unique solution among all the probability vectors on $\mathbb{N}$. This shows that the solution of the recursion (3.4) is unique also under the constraint (3.6). □

It is worth noting that property (3.5) must hold for the tail probabilities $\alpha_n = \mathbb{P}(L_0 > n),\ n \geq 0$, as well since they agree with the a_n. This translates into an assertion on the asymptotics of the hazard function of L_0:

$$\lim_{n\to\infty} \mathbb{P}(L_0 = n+1 | L_0 > n) = 1 - \lim_{n\to\infty} \mathbb{P}(L_0 > n+1 | L_0 > n) = 1 - \lim_{n\to\infty} \frac{\alpha_{n+1}}{\alpha_n} = 1.$$

3.6 Monotonicities in the model parameters

The (conditional) probability $h(x)$ that the immortal line at time 0 carries the beneficial type does not only depend on the frequency x of this type but also on three parameters: selection coefficient σ, mutation rate θ, and mutation probability ν_1 to the deleterious type. As shown in Fig. 3.15, some monotonicity properties apply. Since $h(x) = \sum_{n\geq 0} \alpha_n x(1-x)^n$ depends on the tail probabilities (α_n) monotonically, an increase of α_n for all $n \in \mathbb{N}$ yields an increase of $h(x)$ as well. Let us now explain how the dependence of the tail probabilities on the three parameters can be understood in terms of the pruned equilibrium LD-ASG.
To this end, we consider the tail probabilities as functions of the parameters, i.e., $\alpha_n = \alpha_n(\sigma, \theta, \nu_1)$.

- If $\sigma_1 > \sigma_2$, then $\alpha_n(\sigma_1, \theta, \nu_1) > \alpha_n(\sigma_2, \theta, \nu_1)$. This is due to the fact that higher selection coefficients result in higher intensities of the Poisson point process Γ^* of stars (compare Section 3.4.2). Since each star indicates the birth of a line in the pruned LD-ASG, in distribution more lines are born, which increases the tail probabilities of the top level L_0.

- For $\theta_1 > \theta_2$, one observes $\alpha_n(\sigma, \theta_1, \nu_1) < \alpha_n(\sigma, \theta_2, \nu_1)$. This is because each mutation results in deleting lines from the pruned LD-ASG (unless it is a deleterious mutation on the immune line or a beneficial mutation on the top line), and a higher mutation rate results in more lines being cut away (in distribution). This decreases the tail probabilities for L_0.

- For $\nu_{1,1} > \nu_{1,2}$, one has $\alpha_n(\sigma, \theta, \nu_{1,1}) > \alpha_n(\sigma, \theta, \nu_{1,2})$. The reason is that increasing ν_1 (at constant θ) means replacing each circle in a realisation of the Poisson point processes Γ° by a cross (with a given probability), which thus adds to $\Gamma^\times$. Since the pruning procedure can cut away more than one line at each circle but at most one line at each cross, we get, in distribution, more lines at higher ν_1, which explains the increased tail probabilities.

To summarise: For fixed x, the quantity $h(x)$, as a function of one of the parameters σ, θ, and ν_1 (with the other two parameters being fixed), is strictly increasing in σ, strictly decreasing in θ and strictly increasing in ν_1. We will comment on the third of these monotonicity relations in Sec. 3.7.2.

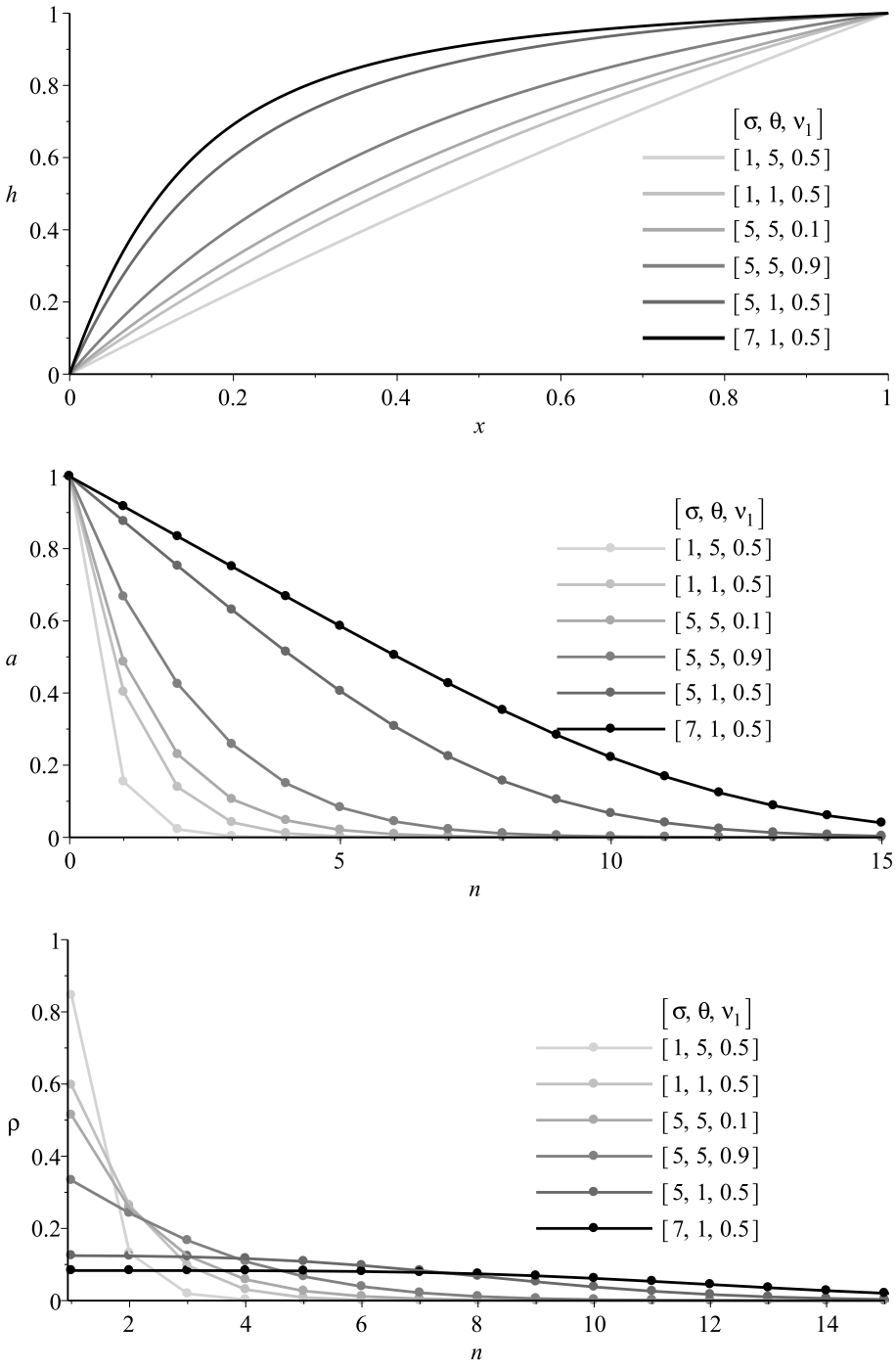


Figure 3.15: *Probability $h(x)$ that the immortal line at time 0 carries the beneficial type, given the frequency of this type is x (top), tail probabilities $\alpha_n = \mathbb{P}(L_0 > n)$, $n \geq 0$, of the stationary distribution of the highest occupied level in the pruned LD-ASG (middle), and probability weights $\rho_n = \mathbb{P}(L_0 = n)$ (bottom), $n \geq 1$. Results are shown for different combinations of selection coefficient σ, mutation rate θ, and mutation probability ν_1 to type 1.*

An illustration of the probability weights (ρ_n) of L_0 (i.e., $\rho_n = a_{n-1} - a_n, n \geq 1$) for various parameter combinations is also included in Fig. 3.15 (bottom).

3.7 Taylor's representation of the CAT distribution via a boundary value problem

Taylor [Tay07] shows by analytic methods (see also [KHB13]) that the (conditional) common ancestor type probabilities $h(x)$ arise as the solution of the boundary value problem

$$\widetilde{A}h(x) = 0, \quad 0 < x < 1 \tag{3.16}$$

$$\lim_{x\to 0} h(x) = 0, \quad \lim_{x\to 1} h(x) = 1, \tag{3.17}$$

where, for $\phi \in C^2([0,1],\mathbb{R})$,

$$\widetilde{A}\phi(x) = A\phi(x) + \theta\nu_0 \frac{1-x}{x}\big(\phi(0)-\phi(x)\big) + \theta\nu_1 \frac{x}{1-x}\big(\phi(1)-\phi(x)\big), \tag{3.18}$$

and A is the generator of the Wright-Fisher diffusion

$$A\phi(x) := \left(\frac{1}{2}x(1-x)\frac{d^2}{dx^2} + (\theta\nu_0(1-x) - \theta\nu_1 x + \sigma x(1-x))\frac{d}{dx}\right)\phi, \tag{3.19}$$

see [Tay07, Proposition 2.4]. Together with his Proposition 2.5, Taylor then suggests the following interpretation of (3.16): Given the frequency of the beneficial type at time 0 is x, sample two lineages at time 0, one of the beneficial type and one of the unfavourable type, and trace them back into the past. He writes: "By comparing this generator with that of the structured coalescent for a sample of size 2 ... [with two different alleles], it is evident that the type of the common ancestor has the same distribution as the type of the sampled lineage which is of the more ancient mutant origin." While Taylor here proposes to take the type frequency path observed from time 0 back into the past as a background process for the structured coalescent, his idea leads to a direct derivation and interpretation of (3.16) after a time reversal and a time shift (see Fig. 3.16). We take the chance to briefly explain this derivation here, as an add-on to [Tay07] and to the approach developed in the previous sections. For this we start from the illustration in the right part of Fig. 3.16.

3.7.1 A representation of $h(x)$ as a hitting probability

Let us fix $x \in (0,1)$ and consider the following two-stage experiment: In the first stage, generate a random Wright-Fisher path $X = (X_t)_{t\geq 0}$ started in $X_0 = x$ with generator (3.19). In the second stage, given X, we consider the ancestral lineage of an individual sampled at random from the population at a late time $s > 0$. For $0 \leq u \leq s$, let J^s_{s-u} be the type of that lineage at time $s-u$, i.e. u units of time back from the time of sampling. In particular, $J^s_s \in \{0,1\}$ is the type at time 0. We abbreviate $b(x) := (1-x)/x$. Then, conditioned on the path $X = (X_t)_{t\geq 0}$ of the frequency of the beneficial type, the dynamics of $J^s := (J^s_u)_{0\leq u\leq s}$ arises when restricting the structured coalescent investigated by Barton, Etheridge and Strum [BES04] and Taylor [Tay07] to a single ancestral lineage: It is a $\{0,1\}$-valued

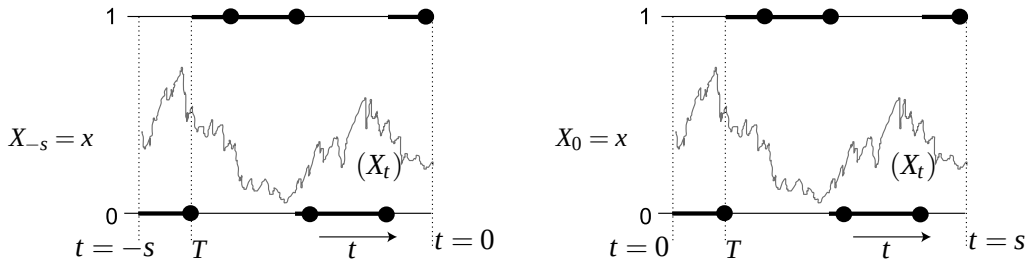


Figure 3.16: *Taylor's interpretation of* (3.16) *after a time-reversal (between $t = -s$ and $t = 0$) (*left*) and a time shift by s (*right*), see Sec. 3.7.1. The type frequency path X of the beneficial type figures as a background process. In this realisation, the type of the common ancestor is* 0 *because the (backward in time) jump at time T is to type* 0.

jump chain with time-inhomogeneous backward-in-time jump rates $\lambda_t^{0,X} := \theta\nu_0 b(X_t)$ from 0 to 1 and $\lambda_t^{1,X} := \theta\nu_1/b(X_t)$ from 1 to 0 at time $t = s-u \in [0,s]$, compare Fig. 3.16, right part. Conditioned on X, the processes J^s, $s > 0$, can be coupled, i.e. constructed on the same probability space, by using two independent Poisson point processes $\Pi^{0,X}$ and $\Pi^{1,X}$ on $\mathbb{R}_+$ with time-inhomogeneous intensities $\lambda^{0,X}$ and $\lambda^{1,X}$. This coupling works as follows: backward in time, each of the processes J^s jumps to 1 at any point $\tau_0 < s$ of $\Pi^{0,X}$ (or remains in 1 if it was already there), and jumps to 0 at any point $\tau_1 < s$ of $\Pi^{1,X}$ (or remains in 1 if it was already there). Let us note that such a coupling would not be possible if one considers the time intervals $[-s,0]$ as in Fig. 3.16, right part, since then the distribution of (an initial piece of) X would vary with s. Thus, while the interpretation that goes along with the left part of Fig. 3.16 is more appealing from a biological point of view, the (mathematically equivalent) picture after the translation to the time interval $[0,\infty)$ (Fig. 3.16, right part) makes the analysis more convenient.

In the above-described coupling, the common ancestor at time 0 is of type 0 if and only if $\lim_{s\to\infty} J_s^s = 0$, which happens if and only if the point in the union of $\Pi^{0,X}$ and $\Pi^{1,X}$ that is closest to 0 belongs to $\Pi^{1,X}$. As a matter of fact, such a closest point to 0 exists: since $X_0 = x \in (0,1)$ and since the rates $\lambda^{0,X}$ and $\lambda^{1,X}$ are bounded as long as X is bounded away from the points $\{0,1\}$, there is a minimal point T_0 in $\Pi^{0,X}$ and a minimal point T_1 in $\Pi^{1,X}$. Let $T := \min\{T_0,T_1\}$. We have thus derived the representation

$$h(x) = \mathbb{E}_x[\mathbb{P}(T = T_1 \mid X)]. \tag{3.20}$$

Now consider a jump-diffusion process $\widetilde{X}$ with generator $\widetilde{A}$ that starts in x, and let $\tilde{T}$ be the time of its first jump to the boundary. We then claim that

$$\Big((X_t)_{0\le t<T}, \mathbb{I}_{\{T=T_1\}}\Big) \stackrel{d}{=} \Big((\widetilde{X}_t)_{0\le t<\tilde{T}}, \widetilde{X}_{\tilde{T}}\Big), \tag{3.21}$$

where $\mathbb{I}$ is the indicator function.

To see this equality in law, recall that, given X, points of $\Pi^{0,X}$ arrive at rate $\lambda^{0,X}$, while a jump of $\widetilde{X}$ to the boundary point 0 occurs at rate $\lambda^{0,\widetilde{X}}$, and that, given X, points of $\Pi^{0,X}$ arrive at rate $\lambda^{0,X}$, while a jump of $\widetilde{X}$ to the boundary point 1 occurs at rate $\lambda^{1,\widetilde{X}}$.

In view of (3.21), the representation (3.20) translates into

$$h(x) = \mathbb{P}_x(\widetilde{X}_{\widetilde{T}} = 1). \tag{3.22}$$

This shows that h is a hitting probability of $\widetilde{X}$, and thus satisfies the Dirichlet problem (3.16). The boundary conditions (3.17) are explained by the fact that the jump rates $\theta\nu_0 b(x)$ and $\theta\nu_1 b(x)^{-1}$ converge to ∞ as x converges to 0 and 1, respectively.
Thus the "forward picture" of Fig. 3.16 leads to the same characterisation of h (in terms of a hitting probability of a jump-diffusion process) as Taylor's above-mentioned "backward picture". The reason for this is the time-reversibility of the one-dimensional Wright-Fisher diffusion. This symmetry breaks down when the allele frequency dynamics are not invariant under time reversal (e.g., with multiple alleles and parent-dependent mutation). Still, a two-stage construction along the lines of Fig.3.16 might, in connection with a suitable "coupling from the future", lead to a related (but then more complicated) characterisation of the multitype analogue of $h(x)$.

3.7.2 Discussion of monotonicities in the parameter ν_1

We have proved in Sec. 3.6 that $h(x)$ is monotonically increasing in ν_1 (for every fixed x). At first sight, this may seem paradoxical: how can it be that an increase in the mutation rate towards the disadvantageous type increases the probability that the common ancestor is of the beneficial type? The representation (3.20) resolves at least part of this paradox:
For fixed X, an increase of ν_1 yields an increase of $\lambda^{1,X}$ and a decrease of $\lambda^{0,X}$. This results in an enhancement of $\mathbb{P}(T = T_1 \mid X)$. In other words, given X, the intensity of mutations "back to the beneficial type" increases as ν_1 increases.
Since, under $\mathbb{P}_x$, X_t (for $t > 0$) has a tendency to become smaller as ν_1 increases, the monotonicity of (3.20) for fixed X is not quite sufficient to prove the monotonicity of $h(x)$. Still, the explanation invoking the intensity of mutations "back to the beneficial type" gives some intuition why $h(x)$ should be increasing in ν_1 - a result which we have derived in Sec. 3.6 via the line counting process of the pruned equilibrium ASG.
Let us now turn to the common ancestor type distribution. To this end, we make the dependence of the stationary density on X (mentioned at the end of Sec. 3.2.1) explicit and denote it by $w_{\nu_1}(x)$, $0 < x < 1$. Consider $g_{\nu_1} := \int_{[0,1]} h_{\nu_1}(x) w_{\nu_1}(x) dx$, that is the probability that, in the equilibrium of X, the common ancestor's type is beneficial. We now have two opposing monotonicities: On the one hand, $h_{\nu_1}(x)$ increases with both ν_1 and x; on the other hand, larger x get lower weight under $w_{\nu_1}(x)dx$ as ν_1 increases. The monotonicity of g_{ν_1} is therefore not obvious. As noticed by Jay Taylor, it is plausible to conjecture that g_{ν_1} should be decreasing with ν_1, which then would be an instance of Simpson's paradox.

3.8 Conclusion

The aim of this contribution was to find a transparent graphical method to identify the common ancestor in a model with selection and mutation, and in this way to obtain the type distribution on the immortal line at some initial time, given the type frequencies in the population at that time. This ancestral distribution is biased towards the favourable type. This bias, which is quantified in

the series representation (3.3), reflects its increased long-term offspring expectation (relative to the neutral case). A closely related phenomenon is well known from multi-type branching processes and deterministic mutation-selection models (see [BG07] and references therein).

Our construction relies on the following key ingredients. We start from the *equilibrium* ASG (without types), and from the insight that the immortal line is the one that is ancestral to the first bottleneck of this ASG. Identifying this ancestral line had previously appeared to be difficult, since it requires keeping track of the hierarchy of (incoming and continuing) branches, which quickly may become confusing. We overcome this problem by *ordering* the lines, in this way introducing a lookdown version of the ASG and a neat arrangement of the lines according to their hierarchy. Next, we *mark* the lines of the equilibrium ASG by the *mutation events* and, working backward in time, apply a *pruning* procedure, which cuts away those branches that cannot be ancestral. Finally, we *assign types* at time 0 to the lines of the resulting pruned LD-ASG by drawing the types of its lines from the initial frequency and thus determine the type of the immortal line at time 0.

This equilibrium lookdown ASG is the principal (and new) tool in our analysis: backward in time, the top level in the pruned ASG performs a Markov chain whose equilibrium distribution can be computed, and the tail probabilities of this equilibrium distribution are shown to obey the Fearnhead-Taylor recursion. This provides the link to the simulation algorithm described by Fearnhead [Fea02] for the common ancestor type distribution in the stationary case. More precisely, our Theorems 3.4 and 3.5 together connect the LD-ASG to the simulation algorithm and thus provide a *probabilistic* derivation for it. At the same time, they imply a generalisation to an arbitrary rather than a stationary initial type distribution. Furthermore, Theorem 3.5 sheds new light on the series representation for the conditional CAT distribution, whose coefficients now emerge as the tail probabilities of the number of lines in the pruned LD-ASG. As a nice by-product, the graphical approach directly reveals various monotonicity properties of the tail probabilities depending on the model parameters, which translate into monotonicity properties of the common ancestor type distribution.

We believe that the pruned equilibrium (lookdown) ASG has potential for the graphical analysis of type distributions and genealogies also beyond the applications considered in the present chapter. Let us also emphasise that, unlike Fearnhead's original approach which builds on the stationary type process (and unlike other pruning procedures that work in a *stationary typed* situation), we start out from the *untyped* lookdown ASG which then is marked and pruned, with the assignment of types at the fixed (initial) time being delayed until the very last step of the construction. This is essential to be able to assign the types i.i.d. with a given frequency, and in this way to arrive at the desired probabilistic derivation of the conditional common ancestor type distribution.

4 The common ancestor type distribution of a Λ-Wright-Fisher process with selection and mutation

The results presented in this chapter are based on joint work with Ellen Baake and Anton Wakolbinger. They are submitted for publication and can be found on arXiv [BLW16].

In this chapter, we extend the results from the previous chapter [LKBW15] to the case of heavy-tailed offspring, directed by a reproduction measure Λ. We obtain a representation of the type distribution of the ancestor in a two-type Λ-Wright-Fisher population with mutation and selection, conditional on the overall type frequency in the old population. This representation is in terms of the equilibrium tail probabilities of the line-counting process L of the graph. We identify a strong pathwise Siegmund dual of L, and characterise the equilibrium tail probabilities of L in terms of hitting probabilities of the dual process.

4.1 Introduction

We consider a Wright-Fisher process with two-way mutation and selection. This is a classical model of mathematical population genetics, which describes the evolution, forward in time, of the type composition of a population with two types. Individuals reproduce and change type, and the reproduction rate depends on the type (the beneficial type reproduces faster than the less favourable one).
In a previous paper [LKBW15], we have presented a graphical construction, termed the *pruned look-down ancestral selection graph (p-LD-ASG)*, which allows us to identify the common ancestor of a population in the distant past, and to represent its type distribution. This construction keeps track of the collection of all *potential* ancestral lines of an individual. As the name suggests, the p-LD-ASG combines elements of the *ancestral selection graph (ASG)* of Krone and Neuhauser [KN97] and the *lookdown construction* of Donnelly and Kurtz [DK99a], which here leads to a hierarchy that encodes who is the *true* ancestor once the types have been assigned to the lines. In addition, a pruning procedure is applied to reduce the graph.
A key quantity is the process L, which counts the number of potential ancestors at any given time. The *ancestral* type distribution is expressed in terms of the stationary distribution of L together with the *overall* type distribution in the past population. The two distributions may be substantially different. This mirrors the fact that the true ancestor is an individual that is successful in the long run; thus, its type distribution is biased towards the favourable type. Explicitly, the ancestral type distribution is represented as a series in terms of the frequency of the beneficial type in the past, where the coefficients are the tail probabilities of the stationary distribution of L and are known in terms of a recursion.

The results obtained so far referred to Wright-Fisher processes. These arise as scaling limits of processes in which an individual that reproduces has a single offspring that replaces a randomly chosen individual (thus keeping population size constant); in the ancestral process, this corresponds to a coalescence event of a pair of individuals. Here we will consider a natural generalisation, the so-called Λ-Wright-Fisher processes. These include reproduction events where a fraction $z > 0$ of the population is replaced by the offspring of a single individual; this leads to *multiple merger events* in the ancestral process.

The Λ-Wright-Fisher processes belong to the larger class of Λ-Fleming-Viot processes (which also include multi-(and infinite-)type generalisations). These, together with their ancestral processes, the so-called Λ-coalescents, have become objects of intensive research in the past two decades. Although less is known for the case with selection, progress has been made in this direction as well (see for example [BP15, DGP12, DK99a, EGT10, Gri14]).

Besides deriving our main result on the *common ancestor type distribution* of a Λ-Wright-Fisher process (stated in Sec. 4.2), the purpose of our paper is twofold: First, we will extend the p-LD-ASG to include multiple-merger events; this will lead to the *p-LD-Λ-ASG*. Second, in the footsteps of Clifford and Sudbury [CS85], we will construct a *Siegmund dual* of the line-counting process L of the p-LD-Λ-ASG. In line with a classical relation between entrance laws of a monotone process and exit laws of its Siegmund dual (discovered by Cox and Rösler [CR84]), the tail probabilities of L at equilibrium correspond to hitting probabilities of the Siegmund dual. This Siegmund dual is a new element of the analysis: In [LKBW15], the recursions for the tail probabilities were obtained from the generator of L, in a somewhat technical manner. The duality provides a more conceptual approach, which is interesting in its own right, and yields the recursion in an elegant way, even in the more involved case including multiple mergers. It will also turn out that the Siegmund dual of L is a natural generalisation (to the case with selection) of the so-called *fixation line* (or fixation curve), introduced by Pfaffelhuber and Wakolbinger [PW06] for Kingman coalescents and investigated by Hénard [Hén15] for Λ-coalescents.

The paper is organised as follows. In Section 4.2, we recapitulate the Λ-Wright-Fisher model with mutation and selection, and the corresponding ancestral process, the Λ-ASG; we also provide a preview of our main results. In Section 4.3, we extend the p-LD-ASG to the case with multiple mergers. Section 4.4 is devoted to the Siegmund dual. The dynamics of this dual process is identified via a pathwise construction and thus yields a strong duality. Once the dual is identified, it leads to the tail probabilities of L with little effort.

4.2 Model and main result

We will consider a population consisting of individuals each of which is either of deleterious type (denoted by 1) or of beneficial type (denoted by 0). The population evolves according to random reproduction, two-way mutation, and fertility selection (that is, the beneficial type reproduces at a higher rate), with constant population size over the generations. The parameters of the model are

- the *reproduction measure* Λ, which is a probability measure on $[0,1]$, and whose meaning will be explained along with that of the generator G_X below Eq. (4.1),

- *the selective advantage* σ (a non-negative constant that quantifies the reproductive advantage of the beneficial type and is scaled with population size),
- the *mutation rates* $\theta\nu_0$ and $\theta\nu_1$, where θ, ν_0, and ν_1 are non-negative constants with $\nu_0+\nu_1 = 1$. Thus, ν_i, $i \in \{0,1\}$, is the probability that the type is i after a mutation event; note that this includes *silent events*, where the type remains unchanged.

We will work in a scaling limit in which the population size is infinite and time is scaled such that the rate at which a fixed pair of individuals takes part in a reproduction event is 1. The process $X := (X_t)_{t\in\mathbb{R}}$ describing the type-0 frequency in the population then has the generator (cf. [EGT10, Gri14])

$$\begin{aligned} G_X g(x) = \int_{(0,1]} \big[x(g(x+z(1-x))-g(x)) + (1-x)(g(x-zx)-g(x))\big] \frac{\Lambda(dz)}{z^2} \\ + \Lambda(\{0\}) \cdot \frac{1}{2} x(1-x) g''(x) \; + \; \big[\sigma x(1-x) - \theta\nu_1 x + \theta\nu_0(1-x)\big] g'(x). \end{aligned} \tag{4.1}$$

The first and second terms of this generator describe the neutral part of the reproduction. In the case $\Lambda = \delta_0$ (to which we refer as the *Kingman case*), the first term vanishes and X is a Wright-Fisher diffusion with selection and mutation. Concerning the part of Λ concentrated on $(0,1]$, the measure $dt\,\Lambda(dz)/z^2$ figures as intensity measure of a Poisson process, where a point (t,z), $t \in \mathbb{R}, z \in (0,1]$, means that at time t a fraction z of the total population is replaced by the offspring of a randomly chosen individual. Consequently, if the fraction of type-0 individuals is x at time $t-$, then at time t the frequency of type-0 individuals in the population is $x+z(1-x)$ with probability x and $x(1-z)$ with probability $1-x$. The third term of generator (4.1) describes the systematic (logistic) increase of the frequency x due to selection, and the type flow due to mutation.
In the absence of both selection and mutation (i.e. when $\sigma = \theta = 0$), the moment dual of the Λ-Wright-Fisher process is the line-(or block-)counting process of the Λ-*coalescent*. The latter was introduced independently by Pitman [Pit99], Sagitov [Sag99], and Donnelly and Kurtz [DK99b], see [Ber09] for an introductory review.
The rate at which any given tuple of j out of b blocks merges into one is

$$\lambda_{b,j} := \int_0^1 z^j (1-z)^{b-j} z^{-2} \Lambda(dz), \quad j \le b. \tag{4.2}$$

Thus the transition rate of the line-counting process from state b to state $c < b$ is given by $\binom{b}{b-c+1}\lambda_{b,b-c+1}$. Note that $\Lambda = \delta_0$ corresponds to Kingman's coalescent; here, $\lambda_{b,j} = \delta_{2,j}$ for all $b \ge 2$. The measure Λ is said to have the property CDI if the Λ-coalescent *comes down from infinity*, i.e. ∞ is an entrance boundary for its line-counting process.
When *selection* is present (i.e. $\sigma > 0$), an additional component of the dynamics of the genealogy must be taken into account. In this case, in addition to the (multiple) coalescences just described, the lines (or blocks) may also undergo a binary branching at rate σ per line. The resulting branching-coalescing system of lines is a straightforward generalisation of the ancestral selection graph (ASG) of Krone and Neuhauser [KN97] to the multiple-merger case; we will call it the Λ-*ASG*. The Λ-ASG belonging to a sample of n individuals taken from the population at time $\bar{t} = 0$ describes all potential ancestors of this sample at times $t < 0$. Throughout we use the variables t and r for *forward* and *backward* time, respectively.

We denote the line-counting process of the Λ-ASG by $K = (K_r)_{r\geq 0}$. It takes values in $\mathbb{N}$ and its generator is

$$G_K g(b) = \sum_{c=1}^{b-1} \binom{b}{b-c+1} \lambda_{b,b-c+1} [g(c) - g(b)] + b\sigma [g(b+1) - g(b)]. \tag{4.3}$$

The process K is the moment dual of the Λ-Wright-Fisher process with selection coefficient σ and mutation rate $\theta = 0$, in the sense that

$$\mathbb{E}[(1-X_t)^n \,|X_0 = x] = \mathbb{E}[(1-x)^{K_t} \,|\, K_0 = n], \tag{4.4}$$

see e.g. [EGT10, Thm. 4.1].
Throughout we will work under the

Assumption 4.1 $0 \leq \sigma < \sigma^* := -\int_0^1 \log(1-x)\frac{\Lambda(dx)}{x^2}$.

Combining results of [Fou13] and [Gri14], one infers that Assumption 4.1 is equivalent to the positive recurrence of the process K on $\mathbb{N}$. Indeed, it is proved in [Gri14, Theorem 3] (for the case $\sigma^* < \infty$) and [Fou13, Theorem 1.1] (for the case $\sigma^* = \infty$) that Assumption 4.1 is *equivalent* to $\mathbb{P}[X_\infty = 1 \mid X_0 = x] < 1$ for all $x < 1$, where X_∞ denotes the a.s. limit of X_t as $t \to \infty$. Combined with the moment duality (4.4), this readily implies that Assumption 4.1 is equivalent to the positive recurrence of K on $\mathbb{N}$ if $\sigma > 0$.
A direct proof that Assumption 4.1 *implies* the positive recurrence of K on $\mathbb{N}$ in the case $\sigma > 0$ is provided by [Fou13, Lemma 2.4]. (Note in this context that K is clearly non-explosive because it is dominated by a pure birth process with birth rate $b\sigma$, $b \in \mathbb{N}$; this makes the first assumption in [Fou13, Lemma 2.4] superfluous).
For $\sigma = 0$, the process K, when started in $b \in \mathbb{N}$, is eventually absorbed in 1. This complements the previous argument in showing that under Assumption 4.1 the process K has a unique equilibrium distribution and a corresponding time-stationary version indexed by $r \in \mathbb{R}$. Similarly, there exists a time-stationary version of the Λ-ASG, which we call the *equilibrium Λ-ASG*, and which will be a principal object in our analysis.

Remark 4.2 *It is proved in [HM12] that* $\sigma^* = \lim_{k\to\infty} \frac{\log k}{\mathbb{E}_k[T_1]}$, *where* T_1 *is the first time at which the line-counting process of the* Λ*-coalescent hits* 1. *In particular, if the measure* Λ *has the property CDI, then* $\sigma^* = \infty$ *and hence Assumption 4.1 is satisfied for all* $\sigma \geq 0$.

Mutations can be superimposed as independent point processes on the lines of the Λ-ASG: On each line, independent Poisson point processes of mutations to type 0 ('beneficial mutation events') come at rate $\theta\nu_0$ and to type 1 ('deleterious mutation events') at rate $\theta\nu_1$.
For $\underline{t} < \bar{t}$ and for a given frequency x of type-0 individuals in the population at time $\underline{t}$, the Λ-ASG may be used to determine the types in a sample $\mathscr{S}$ taken at time $\bar{t}$, together with its ancestry between times $\underline{t}$ and $\bar{t}$, by the following generalisation of the procedure in [KN97]. Each line of the Λ-ASG at time $\underline{t}$ is assigned type 0 with probability x and type 1 with probability $1-x$, in an iid fashion. Let the types then evolve forward in time along the lines: after each beneficial or deleterious mutation, the line takes type 0 or 1, respectively. At each neutral reproduction event (which is a coalescence

event backward in time), the descendant lines inherit the type of the parent. This is also true for the (potential) selective reproduction events (the branching events backward in time), but here one first has to decide which of the two lines is parental. The rule is that the *incoming branch* (the line that issues the potential reproduction event) is parental if it is of type 0; otherwise, the *continuing branch* (the target line on which the potential offspring is placed) is parental. When all selective events have been resolved this way, the lines that are not parental are removed, and one is left with the *true genealogy* of the sample $\mathscr{S}$.

Because of the positive recurrence (and the assumed time-stationarity) of the line-counting process $(K_r)_{-\infty<r<\infty}$, there exists a.s. a sequence of positive (random) times $t_1 < t_2 < \dots$ such that $t_n \to \infty$ and $K_{t_n} = 1$ for all n. Thus, for a given assignment of types to the lines of the stationary Λ-ASG $\mathscr{A}$ at time 0, and for all $n \in \mathbb{N}$, removing the non-parental lines leaves exactly one true ancestral line, between the times $\underline{t} = 0$ and $\bar{t} = t_n$, of the single individual in $\mathscr{A}$ at time t_n. The resulting line between times $\underline{t} = 0$ and $\bar{t} = \infty$ is called the *immortal line* or *line of the common ancestor* in the stationary Λ-ASG.

Our main result is a characterisation of its type distribution at time 0, conditional on the type frequency in the population at that time. For the following definition, let I_t be the type of the immortal line in the stationary Λ-ASG at time t.

Definition 4.3 (Common ancestor type distribution) *In the regime of Assumption 4.1, and for $x \in [0,1]$, let $h(x) := \mathbb{P}(I_0 = 0 \mid X_0 = x)$ be the probability that the immortal line in a stationary Λ-ASG with two-way mutations carries type 0 at time 0, given the type-0 frequency in the population at time 0 is x.*

By shifting the time interval $[0,t]$ back to $[-t,0]$, it becomes clear that $h(x)$ is also the limiting probability (as $t \to \infty$) that the ancestor at the past time $-t$ of the population at time 0 is of the beneficial type, given that the frequency of the beneficial type at time $-t$ was x.

Theorem 4.4 *The probability $h(x)$ has the series representation*

$$h(x) = \sum_{n\geq 0} x(1-x)^n a_n, \tag{4.5}$$

where the coefficients a_n in (4.5) *are monotone decreasing, and the unique solution to the system of equations*

$$\begin{aligned} \sum_{n+1<c\leq\infty} \left[\frac{1}{n}\binom{c-1}{c-n}\lambda_{c,c-n}\right](a_n - a_{c-1}) + (\sigma+\theta)a_n &= \sigma a_{n-1} + \theta\nu_1 a_{n+1}, \quad n \geq 1, \\ a_0 = 1, \quad a_\infty := \lim_{n\to\infty} a_n &= 0, \end{aligned} \tag{4.6}$$

with the convention

$$\binom{\infty-1}{\infty-d+1} := \begin{cases} 0 & \text{if } d = 1 \\ 1 & \text{if } d \geq 2 \end{cases}, \text{ and } \lambda_{\infty,\infty} := \Lambda(\{1\}). \tag{4.7}$$

Let us discuss some special cases. In the *neutral case*, we clearly have $a_0 = 1$ and $a_n = 0$ for $n > 0$, so $h(x) = x$, which is the neutral fixation probability. For $\sigma > 0$, we have $a_n > 0$ for all n, so $h(x) > x$

due to the higher-order terms in the series (4.5). In the *Kingman case*, the system of equations (4.6) simplifies to

$$\left[\frac{1}{n}\binom{n+1}{2}+\sigma+\theta\right]a_n=\frac{1}{n}\binom{n+1}{2}a_{n+1}+\sigma a_{n-1}+\theta\nu_1 a_{n+1},\quad n\geq 1, \tag{4.8}$$

and we immediately obtain

Corollary 4.5 (Fearnhead's recursion) *In the Kingman case, the coefficients in* (4.5) *satisfy the recursion*

$$\left[\frac{1}{2}(n+1)+\sigma+\theta\right]a_n=\left[\frac{1}{2}(n+1)+\theta\nu_1\right]a_{n+1}+\sigma a_{n-1},\quad n\geq 1, \tag{4.9}$$

with $a_0=1$ *and* $\lim_{n\to\infty}a_n=0$.

The case $\Lambda(dz)=dz$, $0\leq z\leq 1$, leads to the so called *Bolthausen-Sznitman coalescent*. Although the latter does not have the property CDI, we still have $\sigma^*=\infty$. In this case one has the identity $\frac{1}{n}\binom{c-1}{c-n}\lambda_{c,c-n}=\frac{1}{(c-n-1)(c-n)}$ (cf. [Ber09] Sec. 6.1), and the system (4.6) simplifies to

$$[1+\sigma+\theta]\,a_n=\sigma a_{n-1}+\theta\nu_1 a_{n+1}+\sum_{j=1}^{\infty}\frac{1}{j(j+1)}a_{n+j},\quad n\geq 1, \tag{4.10}$$

with $a_0=1$ and $\lim_{n\to\infty}a_n=0$.
Recursion (4.9) appears in [Fea02] in connection with a time-stationary Wright-Fisher diffusion (with selection and mutation).[15] In [Tay07], the representation (4.5) together with (4.9) was derived by analytic methods. In [LKBW15], again for the Kingman case, we gave a new, more probabilistic proof, interpreting the coefficients a_n as *equilibrium tail probabilities* of the *line-counting process of the pruned lookdown ASG* (see Sec. 4.3). In the present paper we give a twofold extension: (i) we include the case of multiple mergers, and (ii) we use a strong Siegmund duality (and thus a fully probabilistic method) in order to derive the recursion (4.6).
An analogue of the quantity $h(x)$ can also be defined for a Moran model with finite population size N: for $k\in\{0,1,\ldots,N\}$, let h_k^N be the probability that the individual whose offspring will take over the whole population at some later time is of type 0 at time 0, given the number of type-0 individuals in the population at time 0 is k. In [KHB13] it is shown (for the Kingman case) that h_k^N converges to $h(x)$ as $N\to\infty$ and $k/N\to x$. Here, we work in the infinite-population limit right away, in order to carve out some important features of the underlying mathematical structure.

4.3 The pruned lookdown-Λ-ancestral selection graph

In the previous section, we have outlined the construction of the equilibrium Λ-ASG and layed out how the immortal line within it may be identified: Types are assigned at time 0, and the evolution is

[15] Note that there is a difference of a factor $1/2$ in the scaling of (4.9) in comparison to [Fea02, LKBW15, Tay07]. This is because these papers use the diffusion part of the Wright-Fisher generator (see (4.1)) without the factor 1/2. This corresponds to a pair coalescence rate of 2 in the Kingman case, while in the present paper we assume pair coalescence rate 1 throughout.

then followed forward in time. In practice, however, this procedure is entangled due to the nested case distinctions required to identify the parental branch (incoming or continuing, depending on the type). In the Kingman case, we have solved this problem by *ordering* the lines, and by *pruning* certain lines upon mutation [LKBW15]. The *ordering* is achieved by arranging the coalescence events in a lookdown manner, and by inserting the incoming branch below the continuing branch at every selection event. The *pruning* takes care of the fact that the mutations convey information on the types of lines; this entails that some lines in the ASG can never be ancestral, no matter which types are assigned at time 0, and can thus be deleted from the set of potential ancestors. By construction, this removal does not affect the immortal line.
More precisely, consider a realisation $\mathscr{A}$ of the ordered equilibrium ASG, decorated with the mutation events. The corresponding lookdown version is obtained by placing the lines on consecutive levels, starting at level 1. We now proceed from $r = 0$ in the direction of increasing r. When a beneficial mutation event is encountered, we delete all lines above it. When a deleterious mutation event occurs, we erase the line that carries it; the lines above the affected line slide down to fill the space. One of the lines, called the *immune line*, is distinguished in that it is *not* killed by mutations; rather, it is relocated to the top. Let us anticipate that this is the line that is immortal if all lines at time 0 are assigned type 1. For illustrations and more details about the pruning procedure, see [LKBW15].
The resulting *pruned lookdown ASG* can also be generated in one step, backward in time, in a Markovian manner. In what follows, we review this construction and extend it to the pruned lookdown Λ-ASG.
At each time r, the pruned lookdown Λ-ASG $\mathscr{G}$ consists of a finite number L_r of *lines*, i.e. the process $(L_r)_{r\in\mathbb{R}}$ takes values in the positive integers and L_r is the number of lines in $\mathscr{G}$ at time r. The lines are numbered by the integers $1,\dots,L_r$, to which we refer as *levels*. The evolution of the lines as r increases is determined by a point configuration on $\mathbb{R}\times\big(\mathscr{P}(\mathbb{N})\cup(\mathbb{N}\times\{*,\times,\circ\})\big)$, where $\mathscr{P}(\mathbb{N})$ is the set of subsets of $\mathbb{N}$ and $\mathscr{P}(\mathbb{N})$ is equipped with the σ-algebra generated by $\eta\mapsto\mathbf{1}_\eta(i)$, $i\in\mathbb{N}$, $\eta\in\mathscr{P}(\mathbb{N})$. Each of the points (r,τ) stands for a *transition element* τ occurring at time r, that is, a *merger*, a *selective branching*, a *deleterious mutation*, or a *beneficial mutation* at time r. The level of the immune line at time r is denoted by M_r; its precise meaning will emerge from Proposition 4.6.
Let us now detail the transition elements and their effects on $\mathscr{G}$ (see Figs. 4.1 and 4.2):

- A *merger* at time r is a pair (r,η), where η is a subset of $\mathbb{N}$. If $|\{1,\dots,L_{r-}\}\cap\eta|\le 1$, then $\mathscr{G}$ is not affected. If, however, $\{1,\dots,L_{r-}\}\cap\eta=\{j_1,\dots,j_\kappa\}$ with $j_1<\dots<j_\kappa$ and $\kappa\ge 2$, then the lines at levels $j_2,\dots,j_\kappa$ merge into the line at level j_1. The remaining lines in $\mathscr{G}$ are relocated to 'fill the gaps' while retaining their original order; this renders $L_r=L_{r-}-\kappa+1$. The immune line simply follows the line on level M_{r-}.

- A *selective branching* at time r is a triple $(r,i,*)$, with $i\in\mathbb{N}$. If $L_{r-}<i$, then $\mathscr{G}$ is not affected. If $L_{r-}\ge i$, then a new line, namely the incoming branch, is inserted at level i and all lines at levels $k\ge i$ (including the immune line if $M_{r-}\ge i$) are pushed one level upward to $k+1$, resulting in $L_r=L_{r-}+1$. In particular, the continuing branch is shifted from level i to $i+1$.

- A *deleterious mutation* at time r is a triple $(r,i,\times)$, with $i\in\mathbb{N}$. If $L_{r-}<i$, then $\mathscr{G}$ is not affected. If $L_{r-}\ge i$ and $i\ne M_{r-}$, then the line at level i is killed, and the remaining lines in $\mathscr{G}$ (including the immune line) are relocated to 'fill the gaps' (again in an order-preserving way), rendering

$L_r = L_{r-} - 1$. If, however, $i = M_{r-}$, then the line affected by the mutation is not killed but relocated to the currently highest level, i.e. $M_r = L_{r-}$. All lines above i are shifted one level down, so that the gaps are filled, and in this case $L_r = L_{r-}$.

- A *beneficial mutation* at time r is a triple $(r, i, \circ)$, with $i \in \mathbb{N}$. If $L_{r-} < i$, then $\mathscr{G}$ is not affected. If $L_{r-} \geq i$, then all the lines at levels $> i$ are killed, rendering $L_r = i$, and the immune line is relocated to $M_r = i$.

Proposition 4.6 *Assume that for some $r_0 < 0$ we have $L_{r_0} = 1$, and assume there are finitely many transition elements that affect $\mathscr{G}$ between times r_0 and 0. Consider an arbitrary assignment of types to the L_0 lines at time $r = 0$. Then the level of the immortal line at time 0 is either the lowest type-0 level at time 0 or, if all lines at time 0 are of type 1, it is the level M_0 of the immune line at time 0. In particular, the immortal line is of type 1 at time 0 if and only if all lines in $\mathscr{G}$ at time 0 are assigned type 1.*

Proof. In the absence of multiple mergers (i.e. if all mergers have exactly two elements), this is Theorem 4 in [LKBW15]. In its proof, the induction step for binary mergers directly carries over to multiple mergers. □

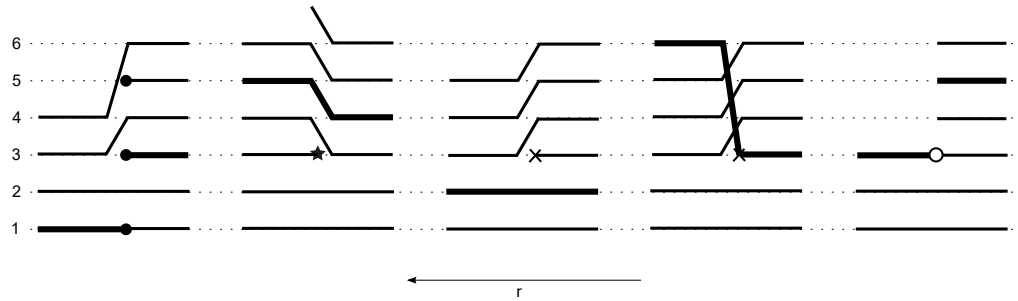


Figure 4.1: *Transitions of the pruned lookdown Λ-ASG. Since the graph evolves 'into the past', time r runs from right to left in the figure. The value of L is 6 before the jump; the immune line is marked in bold. From left to right: A 'merger' of the lines on levels 1, 3, and 5 (indicated by bullets); a 'star' at level 3; a 'cross' at level 3, outside the immune line; a 'cross' on the immune line at level 3; a 'circle' at level 3.*

Taking together the above description of $\mathscr{G}$ and the rates defining the Λ-ASG (Sec. 2), we can now summarise and formalise the law of $\mathscr{G}$ as follows. The transition elements arrive via independent Poisson processes: For each $i \in \mathbb{N}$, the 'stars', 'crosses', and 'circles' at level i come as Poisson processes with intensities σ, $\theta\nu_1$ and $\theta\nu_0$, respectively. For each 2-element subset η of $\mathbb{N}$, the 'η-mergers' come as a Poisson process with intensity $\Lambda(\{0\})$. In addition, we have a Poisson process with intensity measure $\mathbf{1}_{\{z>0\}}\frac{1}{z^2}\Lambda(dz)\,dr$, where each z generates a random subset $H(z) := \{i : V_i = 1\} \subset \mathbb{N}$, with $(V_i)_{i\in\mathbb{N}}$ being a Bernoulli(z)-sequence, and the point (r,z) gives rise to the merger $(r,H(z))$. All these Poisson processes are independent. The points (r,τ) constitute a Poisson configuration Ψ, whose intensity measure we denote by $\mu\otimes\rho$, where μ is Lebesgue measure on $\mathbb{R}$. With the transition rules described above, this induces Markovian jump rates upon L_r and (L_r, M_r). With the help of (4.2), it

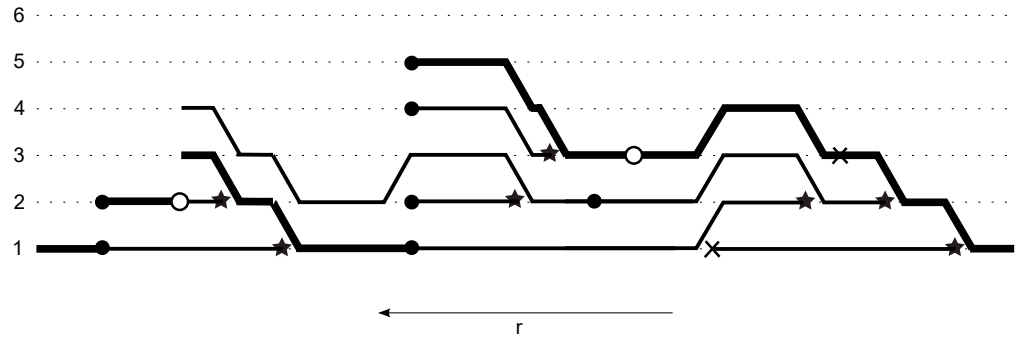


Figure 4.2: *A cut-out of a realisation of the pruned lookdown Λ-ASG. The immune line is marked in bold.*

is easily checked that the generator G_L of L is given by

$$\begin{aligned} G_L g(\ell) = &\sum_{c=1}^{\ell-1} \binom{\ell}{\ell-c+1} \lambda_{\ell,\ell-c+1} [g(c) - g(\ell)] + \ell\sigma [g(\ell+1) - g(\ell)] \\ &+ (\ell-1)\theta\nu_1 [g(\ell-1) - g(\ell)] + \sum_{k=1}^{\ell-1} \theta\nu_0 [g(\ell-k) - g(\ell)]. \end{aligned} \tag{4.11}$$

Due to Assumption 4.1 and Remark 4.2b), and because L is stochastically dominated by K, the process L obeys

$$\mathbb{E}_\ell[T_1] < \infty, \qquad \ell \in \mathbb{N}. \tag{4.12}$$

Thus L has a time-stationary version $\widetilde{L}$ (which is $\widetilde{L} \equiv 1$ if $\sigma = 0$), and likewise the pruned lookdown Λ-ASG has an equilibrium version as well. We now set $L_{\text{eq}} := \widetilde{L}_0$ and denote the tail probabilities of L_{eq} by

$$\alpha_n := \mathbb{P}(L_{\text{eq}} > n), \quad n \in \mathbb{N}_0. \tag{4.13}$$

Because of (4.12), for almost all realisations of $\widetilde{L}$, there exists an $r_0 < 0$ such that $\widetilde{L}_{r_0} = 1$. Hence, arguing as in [LKBW15, proof of Theorem 5], we conclude from Proposition 4.6 the following

Corollary 4.7 *Given the frequency of the beneficial type at time* 0 *is x, the probability that the immortal line in the equilibrium p-LD-Λ-ASG at time* 0 *is of beneficial type is*

$$h(x) = \sum_{n\geq 0} x(1-x)^n \alpha_n. \tag{4.14}$$

In order to further evaluate the representation (4.14), we need information about the equilibrium tail probabilities α_n. This is achieved in the following sections via a process D which is a Siegmund dual for L.

4.4 An application of Siegmund duality

A central point in our proof of Theorem 4.4 will be that the equilibrium tail probabilities of L can be expressed as certain hitting probabilities of a process D which is a so-called Siegmund dual of

L. The relationship between the transition semigroups of L and D is given by formula (4.15) below. Intuitively, the process D may be seen as going into the opposite time direction as L. In a suitable representation via stochastic flows, which turns out to be available for monotone processes, (4.15) means that the paths of D remain 'just above' those of L, see Sec. 4.4.2 below.

4.4.1 Tail probabilities and hitting probabilities

It is clear that L is stochastically monotone, that is, $\mathbb{P}_n(L_r \geq i) \geq \mathbb{P}_m(L_r \geq i)$ for $n \geq m$ and for all $i \in S$ (where the subscript refers to the initial value of the process). It is well known [Sie76] that such a process has a Siegmund dual, that is, there exists a process D such that

$$\mathbb{P}_\ell(L_u \geq d) = \mathbb{P}_d(D_u \leq \ell) \tag{4.15}$$

for all $u \geq 0$, $\ell, d \in \mathbb{N}$.

Lemma 4.8 *The tail probabilities of the stationary distribution of L are hitting probabilities of the dual process D. To be specific,*

$$\alpha_n = \mathbb{P}_{n+1}(\exists t \geq 0 : D_t = 1) \quad \forall n \geq 0. \tag{4.16}$$

Proof. This is a special case of [CR84, Thm. 1] for entrance and exit laws. In our case the entrance law is the equilibrium distribution of L, the exit law is a harmonic function (in terms of hitting probabilities), and the proof reduces to the following elementary argument. Namely, evaluating the duality condition (4.15) for $\ell = 1$ and $d = n+1$, $n \geq 0$, gives

$$\mathbb{P}_1(L_u \geq n+1) = \mathbb{P}_{n+1}(D_u = 1) \quad \text{for all } u \geq 0,\ n \geq 0. \tag{4.17}$$

Taking the limit $u \to \infty$, the left-hand side converges to $\mathbb{P}(L_{\text{eq}} > n) = \alpha_n$ by positive recurrence and irreducibility. Setting $\ell = d = 1$ in (4.15), we see that 1 is an absorbing state for D. Hence we have for the right-hand side of (4.17)

$$\lim_{u\to\infty} \mathbb{P}_{n+1}(D_u = 1) = \mathbb{P}_{n+1}(\exists t \geq 0 : D_t = 1) \quad \forall n \geq 0,$$

and the lemma is proven. □

Next we want to show that the (shifted) hitting probabilities

$$\alpha_n = \mathbb{P}_{n+1}(\exists t \geq 0 : D_t = 1), \quad n \geq 0, \tag{4.18}$$

satisfy the system of equations (4.6). More precisely, (4.6) will emerge as a first-step decomposition of the hitting probabilities. For this purpose, we first have to identify the jump rates of D. This can be done via a generator approach that translates the jump rates of the process L (which appear in (4.11)) into their dual jump rates, see, for instance, formula (12) in [CS85] or in [Sie76]. For the jump rates coming from the mergers this is somewhat technical, see the calculations in the appendix in [Hén15]. Inspired by [CS85] we will therefore take a 'strong pathwise approach' that consists in decomposing the dynamics of L into so-called *flights*, which can be 'dualised' one by one. While Clifford and Sudbury, starting from the generator of a monotone process, in [CS85, Thm 1] construct a special Poisson process of flights for which they form the duals ([CS85, Thm 2]), in our situation the Poisson process of flights is naturally given (being induced by the transition elements for $\mathscr{G}$ defined in Sec. 4.3, see Sec. 4.4.3 below). Consequently, we will show in Proposition 4.10 that the approach of [CS85, Thm 2] works also when starting from a more general Poisson process of flights.

4.4.2 Flights and their duals

In [CS85], Clifford and Sudbury introduced a graphical representation that allows us to construct a monotone homogeneous Markov process $\mathscr{L}$ together with its Siegmund dual $\mathscr{D}$ on one and the same probability space. The method requires that the state space S of the processes $\mathscr{L}$ and $\mathscr{D}$ is (totally) ordered. We restrict ourselves to the case $S := \mathbb{N} \cup \{\infty\}$, which is the relevant one in our context (and which is prominent in [CS85] as well).

The basic building blocks of Clifford and Sudbury's construction are so-called *flights*. A flight f is a mapping from S into itself that is order-preserving, so $f(k) \leq f(\ell)$ for all $k < \ell$ with $k, \ell \in S$; let us add that each flight leaves state ∞ invariant, so $f(\infty) = \infty$. By the construction described below, a flight f that appears at time r will induce the transition to $\mathscr{L}_r = f(\ell)$, given $\mathscr{L}_{r-} = \ell$. This way, transitions from different initial states will be coupled on the same probability space. A flight f is graphically represented as a set of simultaneous arrows pointing from ℓ to $f(\ell)$, for all $\ell \in S$, so that the process simply follows the arrows. Examples are shown in Fig. 4.3.

We denote the set of all flights by $\mathscr{F}$, and consider a Poisson process Φ on $\mathbb{R} \times \mathscr{F}$ whose intensity measure is of the form $\mu \otimes \gamma$, where μ is again Lebesgue measure on $\mathbb{R}$, and the measure γ has the property

$$\gamma(\{f \in \mathscr{F} : f(\ell) \neq \ell\}) < \infty, \qquad \ell \in \mathbb{N}. \tag{4.19}$$

Property (4.19) implies that with probability 1, for all $\ell \in \mathbb{N}$ and $r \in \mathbb{R}$, among all the points (s, f) in Φ with $s > r$ and $f(\ell) \neq \ell$, there is one whose s is minimal. We denote this time by $\nu(r, \ell)$. For $r \in \mathbb{R}$ and $\ell \in \mathbb{N}$, we define inductively a sequence $(s_0, \ell_0), (s_1, \ell_1), \ldots$ with $r =: s_0 < s_1 < \cdots$, $\ell =: \ell_0, \ell_1, \ell_2, \ldots \in S$, by setting $s_i := \nu(s_{i-1}, \ell_{i-1})$, $\ell_i := f(\ell_{i-1})$, with $(s_i, f) \in \Phi$. (Note this procedure will terminate if $\ell_i = \infty$ for some $i \in \mathbb{N}$.)

With the notation just introduced, Φ induces a semi-group (a *flow*) of mappings, indexed by $r < s \in \mathbb{R}$, and defined by

$$F_{r,s}(\ell) := \begin{cases} \ell_i & \text{if } s_i \leq s < s_{i+1}, \\ \infty & \text{if } \lim_{i \to \infty} s_i \leq s \end{cases} \tag{4.20}$$

for $\ell \in \mathbb{N}$, with $F_{r,s}(\infty) := \infty$.

Assuming property (4.19), we say that Φ *represents* the process $\mathscr{L}$ if for all $s>0$ the distribution of $F_{0,s}(\ell)$ is a version of the conditional distribution of $\mathscr{L}_s$ given $\{\mathscr{L}_0=\ell\}$, $\ell\in\mathbb{N}$. Equivalently, for all $r\in\mathbb{R}$ and $u>0$,

$$\mathbb{P}_\ell(\mathscr{L}_u\in(.))=\mathbb{P}(F_{r,r+u}(\ell)\in(.)). \tag{4.21}$$

We now describe, in the footsteps of Clifford and Sudbury [CS85], the construction of a strong pathwise Siegmund dual $\mathscr{D}$, based on the same realisation of the flights as for the original process $\mathscr{L}$. Def. 4.9 *a*) formalises the statement at the beginning of Sec. 4.4 that the paths of D remain 'just above' those of L, see also Fig. 4.3 for an illustration.

Definition 4.9 (Dual flights) *a) For a flight $f:S\to S$, its* dual flight $\widehat{f}$ *is defined by*

$$\widehat{f}(d)=\min(f^{-1}(\{d,d+1,\ldots\})),\quad d\in S, \tag{4.22}$$

with the convention $\min(\varnothing)=\infty$.
b) For a Poisson process Φ on $\mathbb{R}\times\mathscr{F}$, we define $\widehat{\Phi}$ as the result of Φ under the mapping $(r,f)\mapsto(-r,\widehat{f})=:(t,\widehat{f})$. Moreover, under the assumption

$$\gamma(\{f\in\mathscr{F}:\widehat{f}(d)\neq d\})<\infty,\quad d\in\mathbb{N}, \tag{4.23}$$

we define $\widehat{F}$ in terms of $\widehat{\Phi}$ in the same way as F was defined in terms of Φ by (4.20).

It is clear that $\widehat{f}$ is order preserving. Since f is monotone increasing by assumption, we have $\max(f^{-1}(\{1,\ldots,d-1\}))\le\min(f^{-1}(\{d,d+1,\ldots\}))$. As $f^{-1}(\{1,\ldots,d-1\})\cap f^{-1}(\{d,d+1,\ldots\})=\varnothing$ and $f^{-1}(\{1,\ldots,d-1\})\cup f^{-1}(\{d,d+1,\ldots\})=S$, we see that (4.22) is equivalent to

$$\widehat{f}(d)=\max(f^{-1}(\{1,\ldots,d-1\}))+1,\quad d\in S, \tag{4.24}$$

with the convention $\max(\varnothing)=0$. Note further that (4.23) is implied by (4.19) together with

$$\gamma(\{f\in\mathscr{F}:\exists k>\ell\text{ s.t. }f(k)\le\ell\})<\infty,\qquad \ell\in\mathbb{N}. \tag{4.25}$$

The following proposition is an adaptation of [CS85, Theorem 2] to our setting. Compare also [JK14, Section 4.1].

Proposition 4.10 *Assume* (4.19) *and* (4.25)*, and assume that ∞ is unattainable for the process $\mathscr{L}$ represented by the Poisson process Φ with intensity measure $\mu\otimes\gamma$. Then the following strong pathwise duality relation is valid: For all $s>0$; $\ell,d\in\mathbb{N}$,*

$$\mathbf{1}_{\{F_{-s,0}(\ell)\ge d\}}=\mathbf{1}_{\{\widehat{F}_{0,s}(d)\le\ell\}},\qquad \textit{almost surely.} \tag{4.26}$$

Proof. Let $Y:=(Y_r)_{r\in[-s,0]}:=(F_{-s,r}(\ell))_{r\in[-s,0]}$, and $\widehat{Y}:=(\widehat{Y}_t)_{t\in[0,s]}:=(\widehat{F}_{0,t}(d))_{t\in[0,s]}$, for given ℓ, d, and s. Due to (4.19) and the assumption that ∞ is unattainable, Y has a.s only finitely many jumps; let us denote the jump times by $-r_1,\ldots,-r_n$. We write $\widehat{J}$ for the union of $\{r_1,\ldots,r_n\}$ and the set of jump times of $\widehat{Y}$. Because of (4.23), $\widehat{J}$ has a smallest element, a second-smallest element, and so on. We denote these elements by $u_1<u_2<\ldots$, and show that

$$Y_0\ge\widehat{Y}_0\text{ if and only if }Y_{(-u_i)-}\ge\widehat{Y}_{u_i},\qquad i=1,2,\ldots \tag{4.27}$$

Proceeding by induction, for (4.27) it is sufficient to show

$$\mathbf{1}_{\{f(j)\geq k\}} = \mathbf{1}_{\{\widehat{f}(k)\leq j\}} \tag{4.28}$$

for all flights f, and $j,k \in \mathbb{N}$. Let $f \in \mathscr{F}$. On the one hand, $f(j) \geq k$ yields

$$\widehat{f}(k) \leq \widehat{f}\big(f(j)\big) = \min\big(f^{-1}(\{f(j), f(j)+1, \ldots\})\big) = \min\big(f^{-1}(f(j))\big) \leq j,$$

where we have used order preservation of $\widehat{f}$ and f as well as (4.22). On the other hand, $f(j) < k$ is equivalent to $f(j)+1 \leq k$. By order preservation and (4.24), this entails

$$\widehat{f}(k) \geq \widehat{f}\big(f(j+1)\big) = \max\big(f^{-1}(\{1, \ldots, f(j)\})\big) + 1 = \max\big(f^{-1}(f(j))\big) + 1 \geq j+1 > j.$$

We have thus shown (4.28), and hence also (4.27).
If (u_i) has no accumulation point, then it has a maximal element, say u_m. Choosing $i = m$ in the r.h.s. of (4.27) yields (4.26) (since $u_m \neq s$ with probability 1). If (u_i) has an accumulation point, say τ, then, because of (4.23), we have $\lim_{t\uparrow\tau} \widehat{Y}_t = \lim_{i\to\infty} \widehat{Y}_{u_i} = \infty$. Because Y remains bounded by assumption, this together with (4.27) enforces that $Y_0 < \widehat{Y}_0$. This means that the l.h.s. of (4.26) takes the value 0. However, this is the case also for the r.h.s of (4.26), since $\infty = \widehat{Y}_\tau = \widehat{Y}_s > \ell$. □

In view of (4.21) we immediately obtain the following

Corollary 4.11 *In the situation of Proposition 4.10, let $\mathscr{D}$ be a process represented by $\widehat{\Phi}$. Then $\mathscr{L}$ and $\mathscr{D}$ satisfy the duality relation* (4.15)*, with L and D replaced by $\mathscr{L}$ and $\mathscr{D}$.*

4.4.3 A Siegmund dual for the process L

Let us now turn to our case where $\mathscr{L} = L$. With each of the transition elements η, $(i,*)$, $(i,\times)$, $(i,\circ)$ introduced in Sec. 4.3 we associate a flight defined as follows ($\ell \in S, i \in \mathbb{N}$):

$$\begin{aligned} &f_\eta(\ell) = \ell - |\{1,....,\ell\} \cap \tilde{\eta}|, \qquad \text{where } \tilde{\eta} := \eta \setminus \{\min(\eta)\}, \\ &f_{i,*}(\ell) = \begin{cases} \ell, & \ell < i, \\ \ell+1, & \ell \geq i, \end{cases} \qquad f_{i,\times}(\ell) = \begin{cases} \ell, & \ell \leq i, \\ \ell-1, & \ell > i, \end{cases} \qquad f_{i,\circ}(\ell) = \begin{cases} \ell, & \ell \leq i, \\ i, & \ell > i, \end{cases} \end{aligned} \tag{4.29}$$

compare also Fig. 4.3. The flights are indeed order preserving. The structure of f_η, $f_{i,*}$, and $f_{i,\circ}$ is clearly inherited from that of the corresponding transition elements. The flights $f_{i,\times}(\ell)$ forget about the position (but not about the existence) of the immune line within the p-LD-Λ-ASG. Indeed, recall that the downward jump rate of L due to deleterious mutations is $(\ell-1)\theta\nu_1$; this reflects the fact that crosses arrive at rate $\theta\nu_1$ per line, but are ignored on the immune line, no matter where it is located. This is taken into account in the definition of the flight $f_{i,\times}$ by setting $f_{\ell,\times}(\ell) = \ell$.
Let us now start from the Poisson configuration Ψ (of points (r,τ) with intensity measure $\mu \otimes \rho$), as described in Sec. 4.3. Let γ be the image of the measure ρ under the mapping $\tau \mapsto f_\tau$, where f_τ is the flight belonging to the transition element τ as defined in (4.29). The measure γ has property (4.19). To see this we write $\gamma = \gamma_m + \gamma_* + \gamma_\times + \gamma_\circ$, where the 4 summands describe the intensity measures of the flights stemming from the mergers, the selective branchings, the deleterious mutations and the

beneficial mutations. It is straightforward that $\gamma_*, \gamma_\times$ and $\gamma_\circ$ obey (4.19). To see that also γ_m obeys (4.19), note that for $\ell \in \mathbb{N}$

$$\gamma_m(\{f \in \mathscr{F} : f(\ell) \neq \ell\}) = \rho(\{\eta : |\eta \cap \{1,\dots,\ell\}| \geq 2\}) \leq \binom{\ell}{2}, \tag{4.30}$$

since $\{\eta : |\eta \cap \{1,\dots,\ell\}| \geq 2\} \subset \bigcup_{1\leq i<j\leq \ell}\{\eta : \{i,j\} \subset \eta\}$ and because for all $i < j \in \mathbb{N}$

$$\rho(\{\eta : \{i,j\} \subset \eta\}) = \int_{(0,1]} z^2 \frac{1}{z^2}\Lambda(dz) + \Lambda(\{0\}) = 1. \tag{4.31}$$

Writing Φ for the Poisson point process with intensity measure $\mu \otimes \gamma$, it is now clear that Φ represents the process L in the sense of (4.20) and (4.21), because the jump rates match those appearing in the generator (4.11).

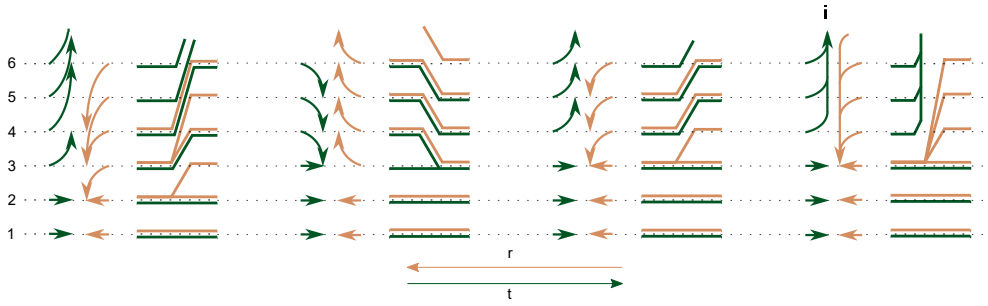


Figure 4.3: *Graphical representation of the four types of flights defined in* (4.29) *(light brown arrows) and their dual flights as defined in* (4.34) *(dark green arrows), together with the resulting paths of L (light brown) and D (dark green). The flights displayed are* f_η *(with* $\eta \cap \{1,\dots,6\} = \{1,3,5\}$*),* $f_{3,*}$*,* $f_{3,\times}$*,* $f_{3,\circ}$*; and* $\widehat{f}_\eta$*,* $\widehat{f}_{3,*}$*,* $\widehat{f}_{3,\times}$*,* $\widehat{f}_{3,\circ}$*. The flight* $\widehat{f}_{3,\circ}$ *maps all states* $d > 3$ *to the absorbing state* ∞*. The paths of L and D follow the arrows in the direction of backward and forward time, respectively.*

Let us now check that γ also satisfies assumption (4.25). It is straightforward that $\gamma_*, \gamma_\times$ and $\gamma_\circ$ obey (4.25). To see that also γ_m obeys (4.25), we note that for $k \geq 2\ell+2$ and $\eta \subset \mathbb{N}$ the inequality $f_\eta(k) \leq \ell$ implies that $|\eta \cap \{1,\dots,\ell+1\}| \geq 1$ and $|\eta \cap \{\ell+2,\dots,2\ell+2\}| \geq 1$. Let H_ℓ denote the set of all $\eta \in \mathscr{P}(\mathbb{N})$ having the latter property. Then we have for all $\ell \in \mathbb{N}$ the estimate

$$\begin{aligned}
&\gamma_m(\{f \in \mathscr{F} : \exists k > \ell \text{ s.t. } f(k) \leq \ell\}) \\
&\leq \sum_{k=\ell+1}^{2\ell+1} \gamma_m(\{f \in \mathscr{F} : f(k) \neq k\}) + \gamma_m(\{f \in \mathscr{F} : \exists k \geq 2\ell+2 \text{ s.t. } f(k) \leq \ell\}) \\
&\leq \sum_{k=\ell+1}^{2\ell+1} \binom{k}{2} + \rho(H_\ell) < \infty,
\end{aligned}$$

because of (4.30) and (4.31), since $H_\ell \subset \bigcup_{1\leq i\leq \ell+1<j\leq 2\ell+2}\{\eta : \{i,j\} \subset \eta\}$.
Following Definition 4.9, we can now consider a process D represented by $\widehat{\Phi}$. According to Corollary 4.11, L and D then obey the duality relation (4.15). It remains to read off the jump rates of D from the intensities of the (dual) flights.

Lemma 4.12 *The generator G_D of the process D is given by*

$$\begin{aligned} G_D g(d) = &\sum_{d<c\le\infty} \binom{c-1}{c-d+1} \lambda_{c,c-d+1} [g(c)-g(d)] + (d-1)\sigma [g(d-1)-g(d)] \\ &+ (d-1)\theta\nu_1 [g(d+1)-g(d)] + (d-1)\theta\nu_0 [g(\infty)-g(d)] \ , \quad d \in \mathbb{N}, g : S \to \mathbb{R}, \end{aligned} \tag{4.32}$$

where we again use the convention (4.7)*.*

Proof. We claim that the flights that are dual to those in (4.29) are of the form

$$\widehat{f}_\eta(d) = \min\{\ell : |\{1,\dots,\ell\} \cap (\mathbb{N} \setminus \tilde{\eta})| = d\}, \qquad \text{again with } \tilde{\eta} = \eta \setminus \{\min(\eta)\}, \tag{4.33}$$

$$\widehat{f}_{i,*}(d) = \begin{cases} d, & d \le i, \\ d-1, & d > i \end{cases} \qquad \widehat{f}_{i,\times}(d) = \begin{cases} d, & d \le i, \\ d+1, & d > i \end{cases} \qquad \widehat{f}_{i,\circ}(d) = \begin{cases} d, & d \le i, \\ \infty, & d > i, \end{cases} \tag{4.34}$$

$d \in S$, $i \in \mathbb{N}$ (see Fig. 4.3).
The expressions in (4.34) are obvious consequences of (4.22) and (4.29). To verify (4.33), we first note that, due to Definition 4.9, we have $\widehat{f}_\eta(d) = \min\big(f_\eta^{-1}(d)\big)$, since f_η is surjective and monotone increasing. Consequently, in the case $|\{1,\dots,d\} \cap \eta| \le 1$ we have $\widehat{f}_\eta(d) = d$, whereas otherwise we have $\widehat{f}_\eta(d) = \min\{\ell : |\{1,....,\ell\} \cap \tilde{\eta}| = \ell - d\} > d$, both in accordance with (4.33).
Let us now consider the contribution of the various types of flights to G_D. For $c \ne d \in \mathbb{N}$ we have to compute $\gamma(\{f : \widehat{f}(d) = c\})$. It is clear that the contributions from γ_*, $\gamma_\times$ and $\gamma_\circ$ yield the last 3 summands in (4.32). For the contribution coming from γ_m, we have for $d < c < \infty$

$$\gamma_m(\{f : \widehat{f}(d) = c\}) = \rho(\{\eta : c \notin \eta, |\{1,\dots,c-1\} \cap \eta| = c-d+1\}). \tag{4.35}$$

The contribution from the Kingman mergers to the right-hand side of (4.35) is $\Lambda(\{0\})\binom{c-1}{2}$ if $c = d+1$, and 0 otherwise. For $z > 0$, the probability that a z-merger does not affect level c but does affect $c-d+1$ out of the levels $1,\dots,c-1$ is $\binom{c-1}{c-d+1} z^{c-d+1}(1-z)^{d-1}$. Integrating this with respect to $\frac{1}{z^2}\Lambda(dz)$ and adding the Kingman component shows that the right-hand side of (4.35) equals $\binom{c-1}{c-d+1}\lambda_{c,c-d+1}$. These are the jump rates from d to $c < \infty$ that appear in the first sum on the r.h.s. of (4.32). It remains to take into account the jump rate of D from d to ∞. For this we note that $f_{\mathbb{N}}(\ell) = 1$, $\ell = 1,2,\dots$, and consequently $\widehat{f}(d) = 1$ if $d = 1$ and $\widehat{f}(d) = \infty$ if $d \ge 2$. These flights appear at rate $\Lambda(\{1\})$, and thus for $d \ge 2$ add the term $(g(\infty) - g(d))\Lambda(\{1\})$ to the generator. □

Remark 4.13 *In the case without selection and mutation (that is, $\sigma = \theta = 0$), our process D shifted by one, that is, $D-1$, is equal to the so-called* fixation line *in [Hén15]. In this case one has no pruning, and the line-counting process K has generator* (4.3) *(with $\sigma = 0$). The (Siegmund) duality between K and D is stated in [Hén15, Lemma 2.4]. For a corresponding statement on the still more general class of exchangeable coalescents see [GM, Thm 2.3].*

We now come to the
Proof of Theorem 4.4. Consider the tail probabilities $\alpha_n = \mathbb{P}(L_{eq} > n)$, $n \in \mathbb{N}_0$, as defined in (4.15). Lemma 4.15 allows us to write them as hitting probabilities of D. Specifically, with

$$\omega(n) := \mathbb{P}_n(\exists t \ge 0 : D_t = 1),$$

we have $\omega(n) = \alpha_{n-1}$. The hitting probabilities $\omega(n)$, $2 \le n < \infty$, constitute a G_D-harmonic function, that is,

$$G_D\omega(n) = 0, \quad n \ge 2. \tag{4.36}$$

It is this relation that is equivalent to the system (4.6). Indeed, (4.36) translates into the system

$$\begin{aligned}\left[\sum_{n+1<c\le\infty}\binom{c-1}{c-n}\lambda_{c,c-n} + n\sigma + n\theta\nu_1 + n\theta\nu_0\right]\alpha_n \\ = \sum_{n+1<c\le\infty}\binom{c-1}{c-n}\lambda_{c,c-n}\alpha_{c-1} + n\sigma\alpha_{n-1} + n\theta\nu_1\alpha_{n+1}, \quad n \ge 1,\end{aligned} \tag{4.37}$$

again using the convention (4.7). Being tail probabilities, the α_n, $n \ge 0$, are monotone, with $\alpha_0 = 1$, and $\alpha_\infty := \lim_{j\to\infty}\alpha_j = 0$. Together with these boundary conditions, Eq. (4.37) divided by n gives the system (4.6) with a_n replaced by α_n.

To prove uniqueness, let (α_n) be as above, (a_n) be a solution of (4.6), and put $b_n := a_{n-1} - \alpha_{n-1}$. Then we have the boundary conditions $b_1 = 0$ and $b_n \to 0$ for $n \to \infty$. In addition, since both $(\alpha_{n-1})_{2\le n<\infty}$ and $(a_{n-1})_{2\le n<\infty}$ are G_D-harmonic, $(b_n)_{2\le n<\infty}$ is G_D-harmonic as well. Let $T(k) := \min\{t \ge 0 : D_t \in \{1,k,k+1,\ldots\}\}$. Note that $T(k)$ is finite a.s. for every $k > 1$. Since, given $D_0 = \ell$, $(b_{D_t})_{t\ge0}$ is a bounded martingale, due to the optional stopping theorem we have $b_\ell = \mathbb{E}[b_{D_{T(k)}} \mid D_0 = \ell]$ for all $k > 1$. Because $b_{D_{T(k)}} \to 0$ as $k \to \infty$, by dominated convergence this implies $b_\ell = 0$ for all ℓ, and hence the desired uniqueness. □

5 The killed ASG and a representation of sampling probabilities

In this chapter we again deal with a classical two-type Wright-Fisher population with mutation and selection in the notation of Chapter 3. The generator G_X of the frequency X of beneficial individuals (of type 0) is then given by (2.4) and its stationary distribution w is determined by (2.5).

While we have determined the type distribution of the common ancestor in the previous two chapters, in this chapter we investigate the type configurations of samples of size $m \in \mathbb{N}$ taken uniformly at random out of a Wright-Fisher population.
We derive a representation of sampling probabilities. In detail, we are interested in the probability of choosing exactly $\ell \in \{0,\dots,m\}$ individuals of the beneficial type 0 in a sample of size $m \in \mathbb{N}$ drawn from a stationary population.
As already described in Chapter 2 on page 13, a recursion for this quantity is given by [KN97, Theorem 5.2] ((2.15) in this thesis) in terms of the ancestral selection graph (ASG).
Since working with this equation may become rather involved, we here derive new a recursion that yields an easily implementable simulation algorithm, and also show some simulation results. This is done by using a rather detailed model (with labelled lines on levels) which we denote *killed ASG*. It arises by starting with the (lookdown) ASG, adding a label process on the lines that keeps track of the ancestors to each individual in the sample, and cutting away lines via a pruning procedure. This pruning procedure is interlaced with the label process and results for every realisation of the (lookdown) ASG in deleting at least as many lines as it is done by the pruning procedure of the pruned LD-ASG. Thus, the killed ASG can be embedded in the pruned LD-ASG.

5.1 The killed ASG

Let us consider a Wright-Fisher process with two-way mutation and selection. We denote the frequency of type-0 individuals again by $X := (X_r)_{r\in\mathbb{R}}$. Let the process be in a stationary situation at time $r = 0$. At time 0, we choose $m \in \mathbb{N}$ different individuals uniformly at random. Let $I_j \in \{0,1\}$ be the type of individual j, $j = 1,\dots,m$. We are interested in the distribution of the random variable $\sum_{j=1}^m I_j$. In detail, we want to gain a representation of the probability weights $\mathbb{E}\left[\binom{m}{\ell} X_0^\ell (1-X_0)^{m-\ell}\right] = \mathbb{P}\left(m - \sum_{j=1}^m I_j = \ell\right)$.
In the footsteps of Shiga [Shi88], to catch the idea and since it is interesting in its own right, we start this section with the killed ASG with sample size $m = 1$.

5.1.1 Sample size one

Let us sample one single individual at random out of a Wright-Fisher population in equilibrium and denote its type by I_1. In this section we want to compute the distribution of the type of this single individual. Namely, we are interested in

$$\mathbb{P}(I_1 = 0) = \mathbb{E}(X_0). \tag{5.1}$$

Let w be the Wright density with mutation and selection, given by (2.5). Then the expectation (5.1) may naturally be computed via solving the integral

$$\mathbb{E}(X_0) = \int_0^1 x\, w(dx). \tag{5.2}$$

But this cannot be done in an easy way analytically. Thus, here, we want to give a way to get $\mathbb{E}(X_0)$ via solving a recursion. As a by-product, the moments $\mathbb{E}\left[(1-X_0)^k\right]$ drop out of the calculation as well. Let us first state the result, then introduce the killed ASG with sample size one (that is the particle picture behind [Shi88, Theorem 4.1], compare also (2.13) in Section 2.1.4), and give the proof afterwards.

Theorem 5.1 *The moments of the distribution of* $(1-X_0)$, $d(k) := \mathbb{E}\left[(1-X_0)^k\right]$, *satisfy the recursion*

$$d(k) = \frac{2\sigma}{k-1+2\theta+2\sigma} d(k+1) + \frac{k-1+2\theta\nu_1}{k-1+2\theta+2\sigma} d(k-1), \quad k \geq 1, \tag{5.3}$$

with boundary conditions

$$d(0) = 1, \quad \lim_{k\to\infty} d(k) = 0.$$

Note that the recursion (5.3) has a shape similar to the recursion of Fearnhead's coefficients (2.33).

The killed ASG of one single individual

The sample to start with is one randomly chosen individual from the stationary population at time $r = 0$. We want to determine its type by following each of its potential ancestors back to the most recent mutation.
The collection of all potential ancestors of this individual in an untyped situation is given by the ASG started at time 0 with one single line and constructed backward in time. The genealogy is then typed by beneficial or deleterious mutations which come as independent Poisson point processes at rates $\theta\nu_0$ and $\theta\nu_1$ along the lines of the ASG.

The type of the single individual at time $r = 0$ is 0 if and only if there is at least one line among all potential ancestors on which the most recent (considered forward in time) mutation is to type 0. The distribution of the sampled individual may be determined in the following way.

First, generate a realisation of a lookdown ASG (as explained in Section 3.4), started with one single (yet untyped) individual at time $r = 0$, and with mutations on the lines. Then, given this realisation, let r_1 be the time of the first mutation after time 0, and inductively r_{i+1} be the time of the first mutation after time r_i, $i \geq 1$. The *killed ASG of one single individual* can then be constructed from this given realisation of the lookdown ASG in the following way (compare also Figure 5.1).

1. Start with $i = 1$ and go to step 2.

2. If the mutation at time r_i is a beneficial mutation (a 'circle'), we are done. Since then there is one ancestor carrying the beneficial type, the individual at time 0 has type 0 as well. We can delete all remaining potential ancestors, stop the procedure, and conclude $I_1 = 0$.
   If the mutation at time r_i is a deleterious mutation (a 'cross'), then the first mutation backward in time (which is the most recent mutation forward in time) on the affected branch is to type 1. Since this branch does not contribute to the event that the sampled individual at time $r = 0$ is of type 0 any more, it may not be considered further and can be erased from the graph. Go to step 3.

3. If no more branches are left after the pruning, we are done. Then all potential ancestors of the sample of size one inherit type 1. Therefore, the single sampled individual has to be of type 1 as well, $I_1 = 1$.
   If there is at least one branch left, set $i = i + 1$ and continue with step 2.

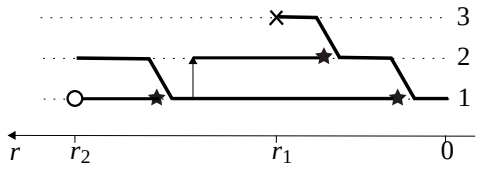


Figure 5.1: *Killed ASG of one single individual. The first mutation after time* 0 *is a deleterious mutation at time* r_1*. The mutation at time* r_2 *is beneficial. It assigns type* 0 *to the single individual at time* 0 *and kills all remaining lines.*

The killed ASG of one single individual can reach its absorbing state (i.e. number of lines is equal to zero) via two different possibilities. Either a beneficial mutation hits the graph. Then the sampled individual holds type 0. Or all lines are erased due to deleterious mutations before a beneficial mutation appears. In this second case the sampled individual inherits type 1.

Let (K_r) be the line counting process of the killed ASG with sample size 1 and $J_r \in \{0,1\}$; $J_r := 0$ if $K_r > 0$ and if there is no beneficial mutation between times r and 0. Otherwise, if $K_r = 0$, $J_r := 0$ if $I_1 = 1$, and $J_r := 1$ if $I_1 = 0$[16]. Then, due to the construction of the killed ASG with sample size one, the generator $G_{(K,J)}$ of the joint process (K,J) is given by

$$\begin{aligned} G_{(K,J)} f(k,j) = k\sigma\left[f(k-1,0) - f(k,0)\right] + \left[\frac{1}{2}k(k-1) + \mathbf{1}_{\{k>1\}} k\theta\nu_1\right]\left[f(k-1,0) - f(k,0)\right] \\ + \mathbf{1}_{\{k=1\}}\theta\nu_1\left[f(0,0) - f(1,0)\right] + k\theta\nu_0\left[f(0,1) - f(k,0)\right]. \end{aligned} \tag{5.4}$$

With the help of the killed ASG of one single individual, we can now state the proof of Theorem 5.1.

Proof of Theorem 5.1. Let $d(k) := \mathbb{E}\left[(1-X_0)^k\right]$ be the probability that k randomly chosen individuals from a stationary population at time $r = 0$ are all of type 1. The types of these k individuals can

[16] J counts the number of type-0 individuals in the sample (also in the general case treated in Section 5.1.2).

be determined by constructing an ASG with k lines backward in time, marked with beneficial and deleterious mutations. Then, to read off whether all individuals are of type 1, check if the most recent mutation back from time 0 on each line is deleterious. Thus, we are exactly in the setting of the killed ASG of one single individual but now started with k lines at time $r = 0$ and with dynamics given by (5.4).
All lines are of type 1 if and only if the killed ASG gets absorbed in state $(K,J) = (0,0)$ and not in state $(K,J) = (0,1)$. Therefore, we may define the event

$$E := \{\text{The killed ASG gets absorbed in state } (0,0) \text{ and not in } (0,1)\} \tag{5.5}$$

and get

$$d(k) = \mathbb{P}\big(E \mid (K,J) = (k,0)\big).$$

We have $d(0) = 1$ and $\lim_{k\to\infty} d(k) = 0$ since $X_0 \neq 0$ a.s. A 'first step decomposition' of the event E by using the rates given by (5.4) yields

$$\left(\frac{1}{2}k(k-1) + k\sigma + k\theta\right) d(k) = k\sigma d(k+1) + \left(\frac{1}{2}k(k-1) + k\theta\nu_1\right) d(k-1), \quad k \geq 1, \tag{5.6}$$

which is equivalent to (5.3) and the proof is complete. □

5.1.2 Sample size $m \geq 1$

If we have a sample of size larger than one, there may be some individuals of type 1 and some of type 0 in the sample. To identify the types of all individuals in the sample, we again start with the (yet untyped) ASG. But in contrast to the killed ASG with sample size one, it is not convenient here to kill the complete graph at a single beneficial mutation. Potential ancestors of the individuals that are not affected by the mutation should not be deleted.
We introduce an additional label to each ancestor to keep track of the relationship between its descendants and the individuals in the sample. These labels then determine all branches that should be deleted at a beneficial mutation. In detail, the label at any arbitrary but fixed ancestral line consists of a subset of $\{1,2,\ldots,m\}$. This subset contains the numbers of all individuals in the sample that are offspring of this potential ancestor. At a beneficial mutation, the numbers that included in the label of the line that is affected by the mutation are erased from all labels. Lines with empty labels are deleted from the graph. Deleterious mutations are treated in the same way as for the killed ASG with sample size 1: the affected branch (together with its label) is pruned. We denote the resulting labelled and pruned version of the (lookdown) ASG the *killed ASG* (for an example see Fig. 5.2). Its transition rates are given and explained in detail in the proceeding section.

The killed ASG

The state space $\mathscr{S}^m \times [0,\infty]$ of the killed ASG started with sample size m at time 0 consists of a time coordinate $r \in [0,\infty]$ and a collection of tuples whose entries are subsets of $\{1,2,\ldots,m\}$ (the labels) together with a natural number (the number of individuals in the sample that are already identified to inherit type 0),

$$\mathscr{S}^m := \varnothing \cup \{(b_1,b_2,\ldots,b_k) \mid b_i \subseteq \{0,1,\ldots,m\}\ \forall i = 1,\ldots,k\ ,\ k = 1,2,\ldots\} \otimes \{0,1,\ldots,m\}.$$

Let $(B_r^m)_{r\geq 0}$ be the configuration of labelled lines of the killed ASG with sample size m and let $(J_r^m)_{r\geq 0}$ be the random variable which counts the number of type-0 individuals in the sample, these are the first and second component of $\mathscr{S}^m$, $(B_r^m, J_r^m) \in \mathscr{S}^m$ for all $r \geq 0$.
The starting configuration of the killed ASG is $(B_0^m, J_0^m) = \big((\{1\},\{2\},\ldots,\{m\}),0\big)$, and the set $\varnothing \times \{0,1,\ldots,m\} \subset \mathscr{S}^m$ is absorbing. The killed ASG can be constructed from a given realisation of the ASG with mutations in the following way.

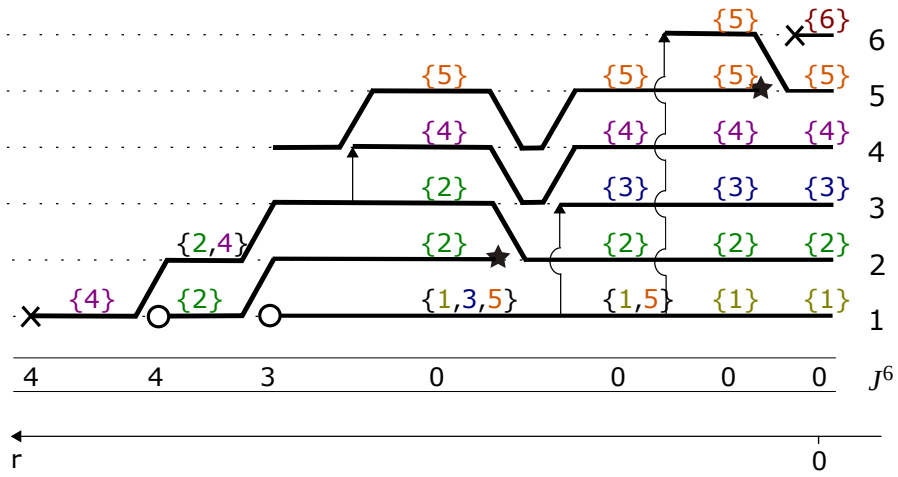


Figure 5.2: *Realisation of the killed ASG started with $m = 6$ lines. The starting configuration on the right hand side of the picture is $\big((\{1\},\{2\},\ldots,\{6\}),0\big)$. At a branching event (symbolised by star), the affected line branches into itself and a clone with identical label. Arrows indicate coalescence events of lines. The new label is the union of the labels of both lines. Each line hit by a deleterious mutation (symbolised by a cross) is deleted. Each beneficial mutation (symbolised by a circle), yields the removal from all labels of all numbers included in the label of the line affected by the mutation. Lines with empty lables are deleted. In addition, the variable J^m is increased by the cardinality of the label of the mutating line. Here, since type 0 is assigned to the labels $1,3,5,2$ due to beneficial mutations, the absorbing state (on the left hand side of the picture) is $(\varnothing,4)$. Labels 4 and 6 are coloured with type 1 due to deleterious mutations.*

In the first step, generate a realisation of a lookdown ASG (compare Section 3.4) with mutations started with m individuals at time $r_0 := 0$ (an example is shown in Fig. 3.14). The elements of this realisation are coalescence events, selective branching events, deleterious mutations, and beneficial mutations. Let us arrange them according to their appearance in time and let the times be $0 < r_1 < r_2 < \ldots$.
We read off the corresponding realisation of the killed ASG from the given realisation of the lookdown ASG by starting at time $r_0 = 0$ with $B_0^m = (\{1\},\{2\},\ldots,\{m\})$, thus assigning labels $\{1\},\{2\},\ldots\{m\}$ to the lines on levels $1,2,\ldots,m$. In addition, we set $J_0^m = 0$.
We then proceed inductively from time r_{i-1} to time r_i, $i \leq \tau$, until the absorbing state $(B_\tau^m, J_\tau^m) = (\varnothing, v)$, $v \in \mathbb{N}_0$, is reached at time r_τ for some $\tau > 0$. In the example shown in Fig. 5.2 the process gets absorbed in state $(\varnothing,4)$. Labels always stick to their lines such that if a line changes its level, the

corresponding label moves to the same new level together with its line. The variable J^m only changes its value at beneficial mutations. If the element at time r_i does affect a line of the lookdown ASG that is not included in the killed ASG any more at time r_i, nothing happens. Otherwise:

- If the element at time r_i is a **selective branching** event of the line at level j, $j \geq 1$, the labels of the lines on levels j and $j+1$ at time r_i are both a copy of the label of the line at level j at time r_i-. The lines at levels $k > j$ together with their labels are shifted one level upwards to make space for the newly born line.

- If the element at time r_i is a **coalescence** of the lines at levels j and k, $1 \leq j < k$, the label on the remaining line at level j at time r_i is the union of the labels of the lines at levels j and k at time r_i-. All lines at levels $\ell > k$ together with their labels are shifted down to level $\ell - 1$ to fill the free space.

- If the element at time r_i is a **deleterious mutation** on the line at level j, proceed exactly as in the pruned LD-ASG and delete the label at level j together with the line at level j (but note that the line is also deleted if it is the immune line). All lines at levels $k > j$ together with their labels are shifted down to level $k-1$ to fill the free space.

- If the element at time r_i is a **beneficial mutation** on the line at level j, $j \geq 1$, the new value of J^m, $J^m_{r_i}$, is gained by adding $J^m_{r_i-}$ and the cardinality of the label at level j. In addition, all numbers that are contained in the label on level j are deleted from the labels of all remaining lines. This yields at least one line with an empty label (the line on level j). All lines with empty labels are deleted from the graph and all remaining lines together with their labels are shifted down in an order preserving way such that the free spaces are filled again.

The transition rates of the killed ASG are directly passed on from the rates of the (lookdown) ASG with mutations. For $i, p \in \{1, \ldots, k\}$ we have

a) coalescence of the lines with labels b_i and b_p, $i < p$:

$$q\Big(\big((b_1,\ldots,b_k), j\big),\ \big((b_1,\ldots,b_{i-1}, b_i \cup b_p, b_{i+1},\ldots,b_{p-1},b_{p+1},\ldots,b_k), j\big) \Big) = 1,$$

b) branching of a line with label b_i :

$$q\Big(\big((b_1,\ldots,b_k), j\big),\ \big((b_1,\ldots,b_{i-1}, b_i, b_i, b_{i+1},\ldots,b_k), j\big) \Big) = \sigma,$$

c) killing of line with label b_i due to a deleterious mutation :

$$q\Big(\big((b_1,\ldots,b_k), j\big),\ \big((b_1,\ldots,b_{i-1}, b_{i+1},\ldots,b_k), j\big) \Big) = \theta\nu_1,$$

d) colouring with type 0 of all sampled individuals with ancestor on the line with label b_i due to a beneficial mutation :

$$q\Big(\big((b_1,\ldots,b_k), j\big),\ \big((b_1 \setminus b_i,\ldots,b_{i-1} \setminus b_i, b_{i+1} \setminus b_i,\ldots,b_k \setminus b_i), j + \#b_i\big) \Big) = \theta\nu_0. \tag{5.7}$$

Now, the killed ASG is an helpful tool to gain the distribution of the number of type-0 individuals in a sample of size m (drawn from a stationary Wright-Fisher population with mutation and selection).

Lemma 5.2 *Let $\tau := \min\{r \geq 0 : B_r^m = \varnothing\}$ be the absorption time of a killed ASG $(B_r^m, J_r^m)_{r\geq 0}$, started with m lines labelled with $\{1\},\dots\{m\}$ at time $r = 0$. Then*

$$\mathbb{P}(J_\tau^m = \ell) = \mathbb{E}\left[\binom{m}{\ell} X_0^\ell (1-X_0)^{m-\ell}\right], \quad 0 \leq \ell \leq m, \tag{5.8}$$

where X_0 is the frequency of the beneficial type in a sationary Wright-Fisher diffusion with selection and mutation at time $r = 0$.

Proof. It is well known (compare [KN97], for example) that a system of equations for the sampling probabilities
$\mathbb{E}\left[\binom{m}{\ell} X_0^\ell (1-X_0)^{m-\ell}\right]$, $0 \leq \ell \leq m$, can be gained in terms of an ASG with mutations, started with m lines.
In a first step, one starts with an untyped sample of size m at time $r = 0$. The genealogy of this sample forms an ASG, evolving in backward time $r \geq 0$. Given this ASG, one assigns mutations to the lines independently according to Poisson point processes with rates $\theta\nu_0$ and $\theta\nu_1$. Then, for almost every realisation of the ASG with mutations, there exists a time $s \in \mathbb{R}^+$ (chosen large enough) such that each potential ancestral line inherits at least one mutation between times $r = 0$ and $r = s$. On each potential ancestral line, the mutation that is closest to time 0 determines the type that is transported forward in time on this line from the time of the mutation towards time 0. To determine the types of the m individuals at time 0, one may therefore cut away all branches after their first mutation. On the remaining (now typed) graph, each selective event can be resolved: The true parent to each selective event (and therefore also the child) is of type 0 if and only if at least one of its two potential parents inherits type 0. Inductively, by resolving the true parent to all selective events in the remaining graph, one gets that each individual in the sample of size m is of type 0 if and only if the closest mutation to time 0 is to type 0 on at least one of its potential ancestral lines.
In the killed ASG, the label on an arbitrary but fixed line marks all individuals in the sample of size m at time $r = 0$ that are offspring of this corresponding ancestral line. Due to construction, the variable J_r^m gives the number of individuals in the sample that have at least one ancestral line assigned type 0 due to having the closest mutation to time 0 to the beneficial type. As all potential ancestors to the sample at time 0 are marked by mutations at time τ, J_τ^m gives the number of type-0 individuals in the sample. Thus, $\mathbb{P}(J_\tau^m = \ell)$ is the probability that ℓ individuals are of type 0 among a sample of size m taken from a stationary Wright-Fisher population. □

Lemma 5.2 yields the following representation of a random type configuration in a sample of size m.

Theorem 5.3 (Representation of sampling probabilities) *Let $m \in \mathbb{N}$ be the size of a random sample of individuals out of a stationary Wright-Fisher population with selection and mutation. The probability that there are exactly ℓ individuals of type 0 in this sample, $0 \leq \ell \leq m$,*

$$d_{m,\ell}\Big(\big((\{1\},\{2\},\dots,\{m\}),0\big)\Big) := \mathbb{E}\left[\binom{m}{\ell} X_0^\ell (1-X_0)^{m-\ell}\right],$$

can be determined by solving the following system of equations,

$$d_{m,\ell}\big((b_1,\dots,b_k),j\big) = \frac{2\sigma}{k(k-1)+2k\theta+2k\sigma} \sum_{i=1}^{k} d_{m,\ell}\big((b_1,\dots,b_{i-1},b_i,b_i,b_{i+1},\dots,b_k),j\big)$$

$$+\frac{2}{k(k-1)+2k\theta+2k\sigma}\sum_{i,p=1,i<p}^{k} d_{m,\ell}\big((b_1,\dots,b_{i-1},b_i\cup b_p,b_{i+1},\dots,b_{p-1},b_{p+1},\dots,b_k),j\big)$$
$$+\frac{2\theta\nu_1}{k(k-1)+2k\theta+2k\sigma}\sum_{i=1}^{k} d_{m,\ell}\big((b_1,\dots,b_{i-1},b_{i+1},\dots,b_k),j\big)$$
$$+\frac{2\theta\nu_0}{k(k-1)+2k\theta+2k\sigma}\sum_{i=1}^{k} d_{m,\ell}\big((b_1\setminus b_i,\dots,b_{i-1}\setminus b_i,b_{i+1}\setminus b_i,\dots,b_k\setminus b_i),j+\#b_i\big),$$
$$j\in\{0,1,\dots,m\},\ b_\nu\subseteq\{1,\dots,m\}\ \forall\,\nu=1,\dots,k,\ k\geq 0,$$

$$d_{m,\ell}\big(\varnothing,\ell\big)=1,\quad d_{m,\ell}\big(\varnothing,j\big)=0 \text{ for all } j\neq\ell. \tag{5.9}$$

Proof. Let $d_{m,\ell}\big((b_1,\dots,b_k),j\big)$ be the probability that the killed ASG started in state $\big((b_1,\dots,b_k),j\big)$ gets absorbed in state $(\varnothing,\ell)$,

$$d_{m,\ell}\big((b_1,\dots,b_k),j\big):=\mathbb{P}\big(\exists r\geq 0:(B_r^m,J_r^m)=(\varnothing,\ell)\mid(B_0^m,J_0^m)=((b_1,\dots,b_k),j)\big), \tag{5.10}$$

$m\in\mathbb{N}$, $j,\ell\in\{0,1,\dots,m\}$, $b_\nu\subseteq\{1,\dots,m\}$ for all $\nu=1,\dots,k$, $k\geq 0$.
Then we have $\mathbb{E}\left[\binom{m}{\ell}X_0^\ell(1-X_0)^{m-\ell}\right]=d_{m,\ell}\Big(\big((\{1\},\{2\},\dots,\{m\}),0\big)\Big)$. Since the state $\big(\varnothing,i\big)$ is absorbing for every $i\in\mathbb{N}_0$, $d_{m,\ell}\big(\varnothing,\ell\big)=1$, and $d_{m,\ell}\big(\varnothing,j\big)=0$ for all $j\neq\ell$. A 'first step decomposition' of $d_{m,\ell}$ according to the rates (5.7) yields (5.9). □

Lemma 5.2 may also be used in terms of simulating sampling probabilities for a sample of size m with the help of the killed ASG. The results of such a simulation algorithm are shown in Section 5.2. The algorithm gets along with rates of the killed ASG only and is not in need of a realisation of the frequency X_0 of type 0 individuals in the whole population. To get a realisation of the number of type-0 individuals in a sample of size m, start a killed ASG with configuration $(B_0^m,J_0^m)=\big((\{1\},\{2\},\dots,\{m\}),0\big)$ and let the algorithm evolve according to the rates $a)-d)$ of the killed ASG, given on page 76. When arrived at its final (absorbing) state ($B_\tau^m=\varnothing$ for some $\tau\geq 0$), the algorithm outputs a perfect sample of J_τ^m from the stationary distribution.

Remark 5.4 *The killed ASG inherits properties from the ASG, e.g. the two following ones.*

(i) *The killed ASG may be started with m lines ($m\in\mathbb{N}$), but one may also use a sample of infinitely many lines at time 0. The quadratic death but only linear birth rates ensure that the killed ASG comes down from infinity. Therefore, there are only finitely many lines left at time ε for all $\varepsilon>0$.*

(ii) *One may run the killed ASG for a given time horizon $r\geq 0$. If not all lines are killed until time r, the types of all (potential) ancestors (that are not typed until time r) can be identified by adding types to the remaining lines at time r: type 0 with probability X_r and 1 with probability $1-X_r$ independently per line.*

5.2 Simulation results

In this section we present some simulation results for sample sizes $m = 3, 5, 20$ and for various combinations of the parameters θ, ν_1, and σ. The algorithm encoded in Theorem 5.3 was implemented in the programming language R. In Figs. 5.3, 5.4, and 5.5, simulations of the number of type-0 individuals in a sample drawn from a stationary Wright-Fisher population with mutation and selection are displayed.

The green diagrams on the top of Figures 5.3, 5.4, and 5.5 show that high mutation rates (together with $\nu_1 = 0.5$) shift the simulated densities towards shapes that look similar to Binomial distributions. This observation may be explained intuitively in the following way. For large mutation rates, the role of the selective advantage becomes very small as each line changes its type frequently. Thus, the type of each line is mostly determined by its most recent mutation. Since this mutation is to type 0 with probability 0.5 (almost) independently for each sampled individual, the simulated distribution is similar to a Binomial distribution with parameters 3, 5, or 20 and 0.5.
The brown diagrams in the middle of the Figures indicate that a high mutation probability to type 1 shifts the simulated density to to the left: More individuals in the population are then of type 1 such that it becomes less probable to sample individuals of type 0.
The blue diagrams at the bottom of the three Figures show that a large selection coefficient results in a shift of the simulated density towards more individuals of type 0. This is evident because a big selective advantage increases the number of type-0 individuals in the population.

In the next section, we close this chapter with a glimpse on the so-called *decision tree* by Dawson and Greven [DG14, Chapter 5]. It is also a tool that can be used to determine the types of individuals of a sample. The decision tree is a collection of all potential histories of this sample (with an a priori unknown type configuration). One can think of it as being some kind of stencil: Some potential type configurations of the sample do not fit into the decision tree and can therefore be removed from the collection of all possible type configurations of the sample.

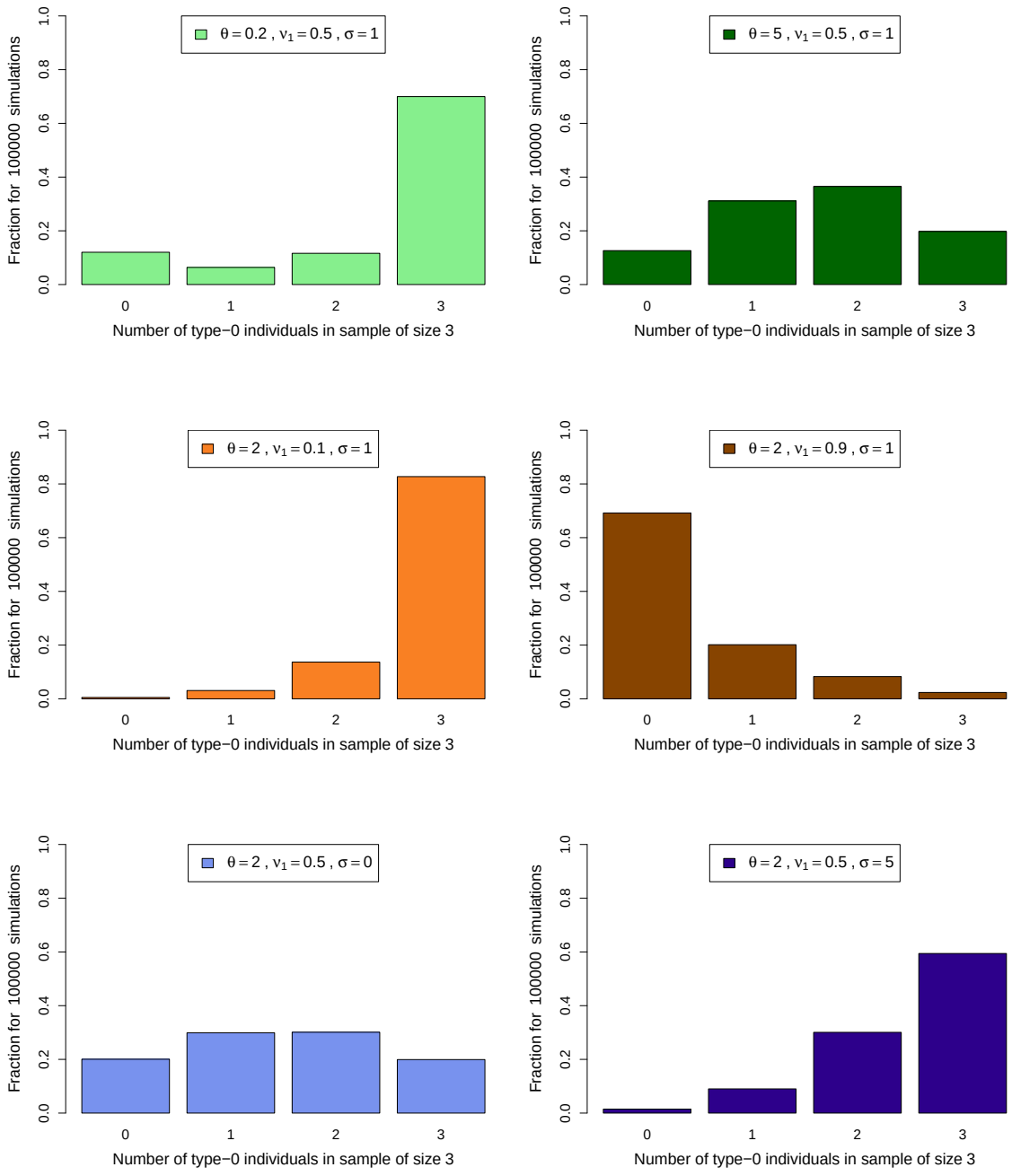


Figure 5.3: *Simulations for sample size* 3.

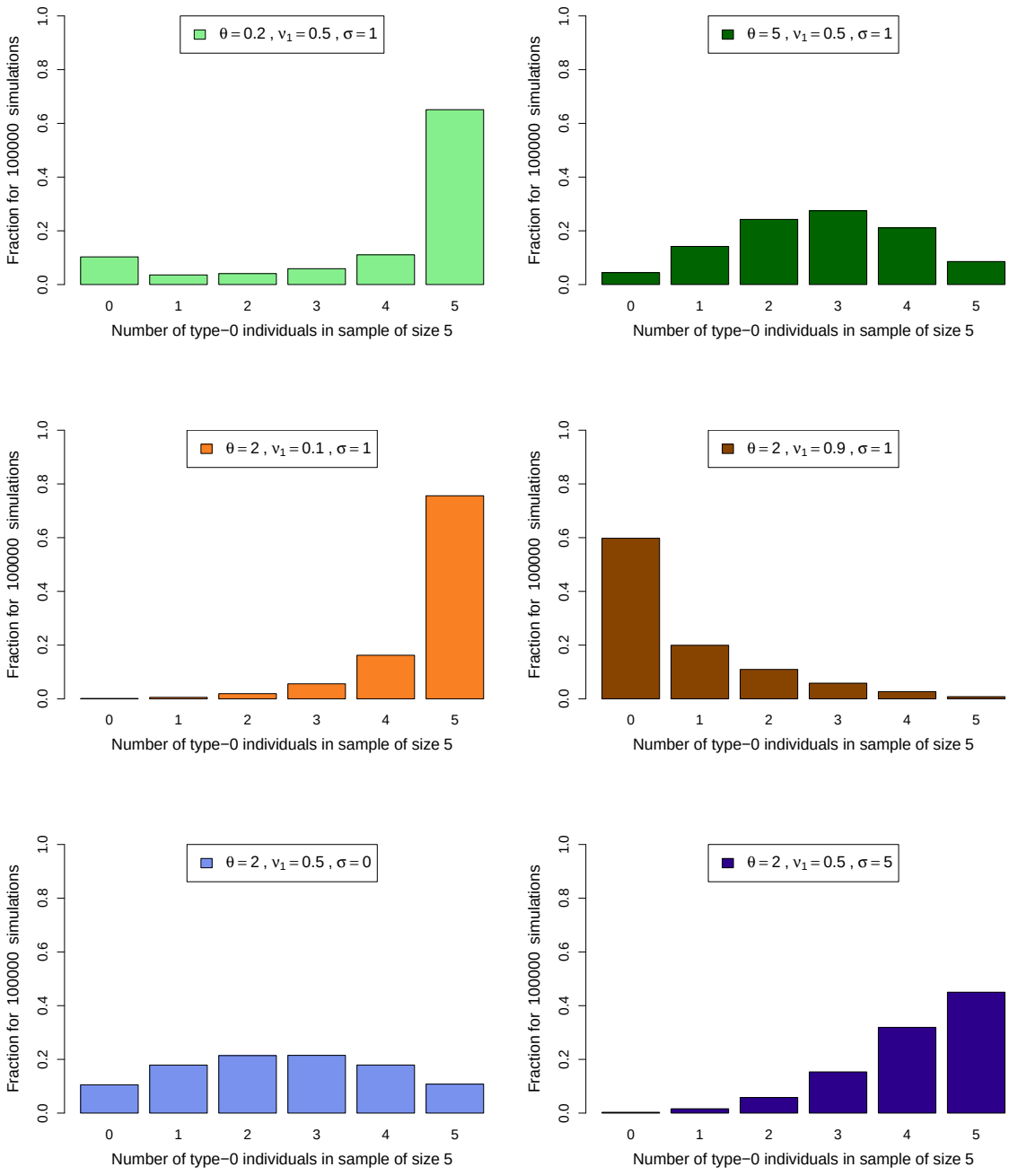


Figure 5.4: *Simulations for sample size* 5.

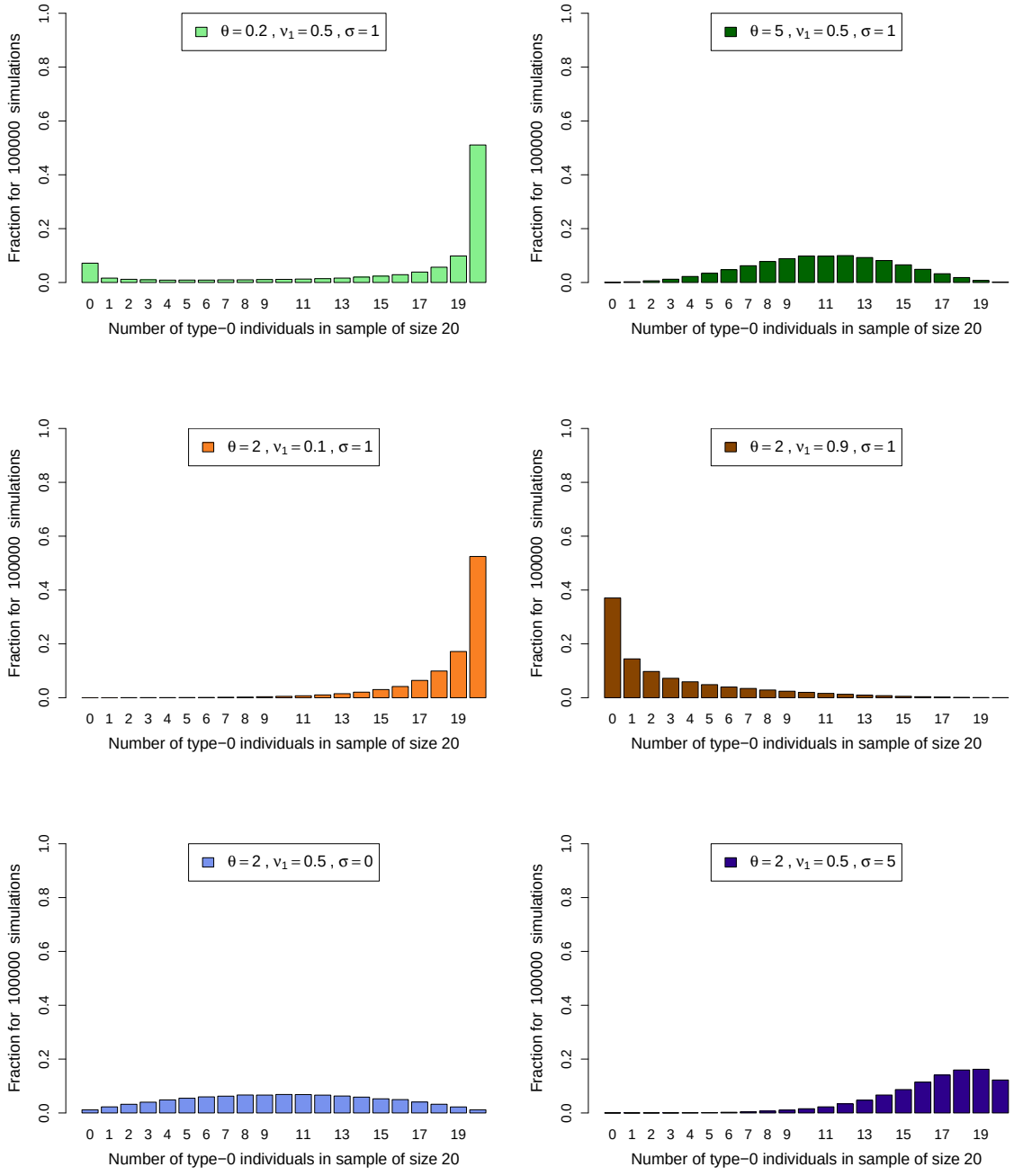


Figure 5.5: *Simulations for sample size* 20.

5.3 Dawson's and Greven's decision tree

Recently, Dawson and Greven published a monograph on spatial Fleming-Viot models with selection and mutation [DG14]. In Chapter 5, they introduce duality as a basic tool for analysing Fleming-Viot processes (and therefore also the special case of Wright-Fisher diffusions) with selection, mutation, and migration. This way, they also consider samples of individuals drawn from a (stationary) population and analyse their ancestry. The sampling probabilities can then be gained by summing over all potential histories weighted with their probabilities. In Section 5.5.4, Dawson and Greven explain how to get these potential histories in terms of a marked graph, denoted as *decision tree* (cf. [DG14, picture on page 141]).

The decision tree may (like the killed ASG) be used for investigating the distribution of the types of individuals in a sample. Therefore, we review the construction of a decision tree in our special case of a Wright-Fisher diffusion (with two types) with mutation and selection. As this gets very voluminous for large samples, we consider only sample size 1 here. This fits well into the setting of [DG14, picture on page 141] and is also comparable with Section 5.1.1 of this chapter.

To this end, we construct the decision tree for the following example (illustrated in Fig. 5.6): A single individual is sampled at time $r = 0$. In order to determine its type, take a look at all its potential ancestors. In the example, the (lookdown) ASG of this individual consists of a selective branching event at time s_1 on level 1, a selective branching event at time s_2 on level 2, a deleterious mutation on level 3 at time s_3, a coalescence between levels 1 and 2 at time s_4, a selective branching event on level 1 at time s_5, and a beneficial mutation at level 1 at time s_6.

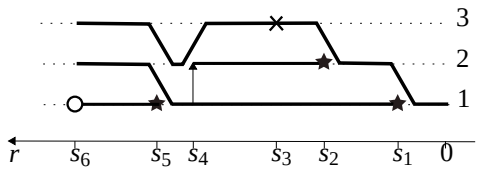
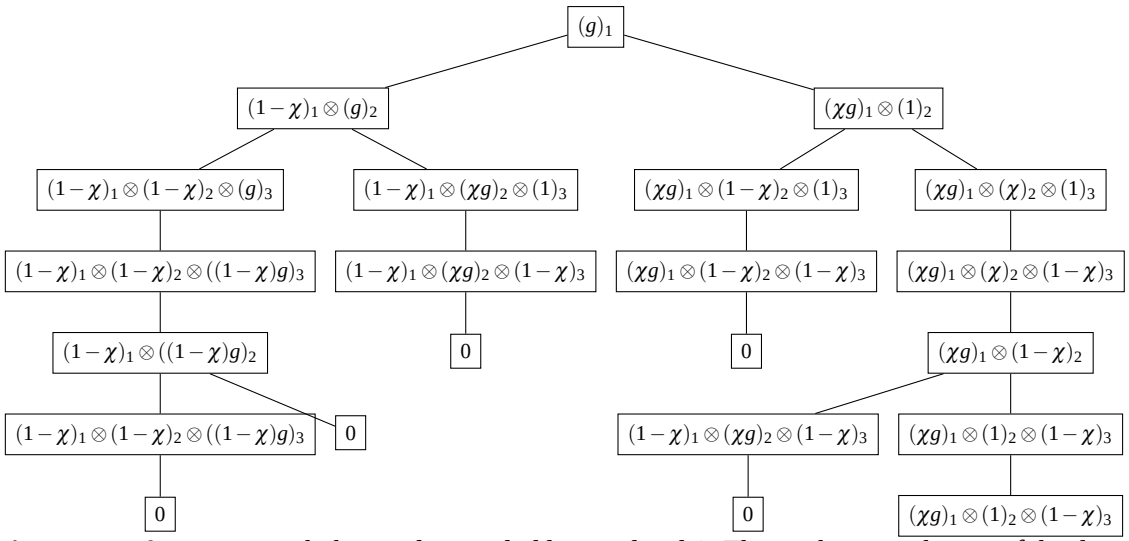


Figure 5.6: *(Lookdown) ASG of a single individual sampled at time* 0 *(the corresponding killed ASG is shown in Fig. 5.1).*

For a test function $g : \{0,1\} \to \{0,1\}$, let $(g)_\ell := g(i_\ell)$, with i_ℓ being the type of the line at level ℓ. Especially, let $(\chi)_\ell := \mathbf{1}_{\{i_\ell=0\}}$. The decision tree for the just described example looks as follows. Time r goes down from top to bottom. Let us identify the first row with $r = 0$, the second row with $r = s_1$, ..., and the seventh row with $r = s_6$. We adopt notation from [DG14]: The variable χ always stands for a line of type 0, and $1 - \chi$ for a line of type 1.
We first show the diagram and then explain its entries and the notation.



At time $r = 0$, we start with the single sampled line on level 1. This is shown at the top of the diagram, $(g)_1$. The index always indicates the level, which in this case is 1.
Then this line branches into two lines (at time $r = s_1$): The incoming branch on level 1 and the continuing branch on level 2. There are two potential parents now: If the incoming branch is of type 1 (which is indicated in the diagram by $(1-\chi)_1$), the parental branch is the continuing one. Therefore, the function g is now assigned to the continuing branch on level 2, $(g)_2$. Thus, this case coincides with $(1-\chi)_1 \otimes (g)_2$ (first entry in the second row, compare also Fig. 5.7, left). In the second case, if the incoming branch is of type 0 (indicated by χ), it is the parental one. So, we have the component $(\chi g)_1$. The type of the continuing branch is not specified; it may be either 0 or 1, thus we have the component $(\chi + (1-\chi))_2 = (1)_2$. Together, this case gives the second entry in the second row: $(\chi g)_1 \otimes (1)_2$ (compare also Fig. 5.7, right).

Figure 5.7: *Selective branching event of the line on level* 1. *Left: The incoming branch is of type* 1*, the true parent is the continuing line on level* 2. *Right: The incoming branch is of type* 0 *and thus the true parent. The labels on the branches are the entries of Dawson's and Greven's decision tree.*

The branching event at time s_1 is followed by another branching event a time s_2 at level 2 which gives the third row in the decision tree.
Then a deleterious mutation at time s_3 on level 3 assigns type 1 to the line on level 3. This is indicated by the additional factor $(1-\chi)$ in the components $(\cdot)_3$ in the fourth row of the decision tree.
Then, at time s_4, the lines on levels 1 and 2 coalesce. But this can only happen if both lines are of the same type (compare Fig. 5.8). If both lines are of type 1 (as in the first entry in fourth row), they coalesce into one line of type 1 and the line on level 3 moves to level one. This gives the first entry

in the fifth row[17]. If both lines are of type 0 (as in the last entry in fourth row), they coalesce into one line of type 0 yielding the last entry in the fifth row. Otherwise, if the two lines carry different types, they cannot coalesce. The second and third entry in the fifth row are therefore just 0.

Figure 5.8: *Coalescence of the lines at levels* 1 *and* 2 *is only possible if both lines inherit the same type. Two lines of type* 1 *(indicated by* $(1-\chi)$*) or of type* 0 *(indicated by* (χ)*) coalesce.*

At time s_5 there is again a branching event at level 1 which gives the sixth row. The second entry is 0 because a line of type 1 is not allowed to branch into an incoming branch of type 0.
Then a mutation at time s_6 assigns type 0 to the line on level 1. As this is only compatible with the right branch in the tree, this branch gives the true ancestry. As we have $(\chi g)_1$, the true parent at time s_6 of the individual sampled at time 0 is located at level 1 and inherits type 0.

This way, the decision tree gives a possibility to determine types of sampled individuals. In comparison to the killed ASG, mutations are treated differently. For example, a line (that is not the immune one) is pruned from the killed ASG at a deleterious mutation. But in the decision tree it is not deleted in all cases.

[17] Note that we are working in the framework of the lookdown ASG. Therefore, we indicate the lines by their levels. At a coalescence event, free levels are filled again by lines at higher levels. Dawson and Greven do not place the lines on levels but number the lines. Therefore, their first entry in the fifth row would keep the index 3, $(1-\chi)_1 \otimes ((1-\chi)g)_3$.

6 The evolution of the common ancestor process together with its Wright-Fisher background

In the setting of Chapter 3, we come back to a classical (two-type) Wright-Fisher diffusion with two-way mutation and directional selection. Fearnhead [Fea02] described the evolution of the type process $(R_r)_{r\in\mathbb{R}}$ on the line of the common ancestor together with a process $(V_r)_{r\in\mathbb{R}}$ of so-called 'virtual lines'.
In this chapter, we analyse the joint evolution of $(R_r)_{r\in\mathbb{R}}$, $(V_r)_{r\in\mathbb{R}}$, and the joint Wright-Fisher background $(X_r)_{r\in\mathbb{R}}$. Using time-reversal in equilibrium, we derive the backward and forward in time generators of the triple process (X,R,V) (taking values in $[0,1]\times\{0,1\}\times\mathbb{N}_0$) and explain all rates heuristically with the help of the pruned lookdown ASG (introduced in Chapter 3). This approach also allows for new insights into the probability $h(x)=\mathbb{P}(R_0=0\mid X_0=x)$. In addition, it ties together results by Fearnhead [Fea02] (who analysed the process (R,V)) and Taylor [Tay07] (who investigated the process (X,R)).

6.1 Introduction

Let us deal with the same setting as in the previous chapter: We consider a classical two-type Wright-Fisher population with two-way mutation and selection (in the notation of Chapter 3).

Remember that $\theta\nu_1$ and $\theta\nu_0$ are the mutation rates per line to types 0 and 1, and that type 0 is the beneficial type with selective advantage σ. X_t is the frequency of type 0 individuals in the population at time $t\in\mathbb{R}$ and $X=(X_t)$ evolves according to the generator G_X, given by (2.4),

$$G_X g(x)=\frac{1}{2}x(1-x)g''(x)+[(1-x)\theta\nu_0-x\theta\nu_1+\sigma x(1-x)]g'(x),\quad g\in\mathscr{C}^2[0,1], \tag{6.1}$$

The stationary probability measure w for this generator is given by (2.5),

$$w(x)=c_w\cdot(1-x)^{2\theta\nu_1-1}x^{2\theta\nu_0-1}\cdot\exp\{2\sigma x\},\quad x\in[0,1]. \tag{6.2}$$

with $c_w=\left[\int_0^1(1-x)^{2\theta\nu_1-1}x^{2\theta\nu_0-1}\cdot\exp\{2\sigma x\}dx\right]^{-1}$.

In this chapter, we are not only interested in the stationary distribution $\big(h(x),1-h(x)\big)$ of the type of the common ancestor but also in the evolution backward and forward in time. In detail, we want to take a closer look at the evolution of the common ancestor process (R,V) (introduced by Fearnhead [Fea02], compare also Chapter 2, especially Section 2.2.3) together with its random Wright-Fisher environment X.

This gives the triple process (X,R,V) which may also arise by incorporating the number V of virtual lines into the process (X,R) that was introduced by Taylor [Tay07].

With the help of the *pruned LD-ASG* (compare Chapter 3), we build a bridge between the two approaches by Fearnhead [Fea02] (reviewed already in Section 2.2.3) and Taylor [Tay07] (reviewed in Section 2.2.4).

Recall that Fearnhead defines the *common ancestor process* (CAP), a modification of the ancestral selection graph ([KN97]), which includes the type R of the common ancestor (or *real line*) and a number V of so-called virtual lines. The process $(R_r,V_r)_{r\in\mathbb{R}}$ is Markovian and its stationary distribution can be calculated from the generator matrix (given by the rates (2.27), (2.28), and (2.29)).
Taylor considers the type R of the immortal line together with the frequency process X. He calculates the stationary distribution of the Markovian process $(X_r,R_r)_{r\in\mathbb{R}}$ by solving a boundary value problem.
In this chapter, we tie together all three variables. We will analyse the evolution of the triple process $(X_r,R_r,V_r)_{r\in\mathbb{R}}$ backward and forward in time and identify its stationary distribution. Via an analytic proof, we will also see that this stationary distribution can also be gained with the help of the pruned LD-ASG directly and intuitively without many calculations.

6.2 The backward in time generator of (X,R,V)

(R_r,V_r) is just a typed sample of individuals (or lines) for all times $r\geq 0$. Thus, backward in time, the triple process (X,R,V) is a structured coalescent process in a random Wright-Fisher environment.
Such a structured coalescent was introduced and analysed by Barton, Etheridge and Sturm [BES04], compare also [Tay07, Section 2].
Here, in the case $(R_0,V_0)=(0,v)$, the structured coalescent starts with v lines of type 1 and one line of type 0. In the case $(R_0,V_0)=(1,v)$ the initial state consists of $v+1$ lines of type 1. When there is a mutation to type 0 on any line (except for the real line), this virtual line is deleted. Due to selection, each branching event may result in an additional virtual line that is required to be of type 1 (compare also Section 2.2.3). The generator of the triple process (X,R,V) is therefore given by adding the mutation and selection dynamics of the virtual lines to the generator of the structured coalescent (given by (2.36) in this thesis, Lemma 3.1 in [BES04], or (3) in [Tay07]).

Lemma 6.1 *The generator $\widehat{G}_{(X,R,V)}$ of the process $(X_r,V_r,R_r)_{r\in\mathbb{R}}$ backward in time is given by*

$$\begin{aligned}\widehat{G}_{(X,R,V)}g(x,0,v) &= G_X g(x,0,v)+\frac{1-x}{x}\theta\nu_0\left[g(x,1,v)-g(x,0,v)\right]\\ &\quad+(v+1)\,\sigma(1-x)\left[g(x,0,v+1)-g(x,0,v)\right]\\ &\quad+\left[v\theta\nu_1+\frac{1}{2}v(v-1)\right]\frac{1}{1-x}\left[g(x,0,v-1)-g(x,0,v)\right]\end{aligned}\tag{6.3}$$

$$\begin{aligned}\widehat{G}_{(X,R,V)}g(x,1,v) &= G_X g(x,1,v)+\frac{x}{1-x}\theta\nu_1\left[g(x,0,v)-g(x,1,v)\right]\\ &\quad+(v+1)\sigma(1-x)\left[g(x,1,v+1)-g(x,1,v)\right]\\ &\quad+\left[v\theta\nu_1+\frac{1}{2}v(v+1)\right]\frac{1}{1-x}\left[g(x,1,v-1)-g(x,1,v)\right]\end{aligned}\tag{6.4}$$

for all $g \in \mathscr{C}^2(0,1) \times \{0,1\} \times \mathbb{N}_0$.

The (analytic) proof of Lemma 6.1 can be carried out analogously to the proof of Lemma (3.1) in [BES04] as a diffusion limit of the discrete model.

Instead of reviewing the analytic calculations here, we verify the rates by giving heuristic and intuitive arguments.

Interpretation of the generator $\widehat{G}_{(X,R,V)}$

The rates of the generator $\widehat{G}_{(X,R,V)}$ can be interpreted[18] in terms of the possible transitions of (R,V), given X. Let us decompose $\widehat{G}_{(X,R,V)}$ into its different summands and look at them separately:

$$\begin{aligned}\widehat{G}_{(X,R,V)}g(x,r,v) \;=\; & \underbrace{G_X\, g(x,r,v)}_{\textcircled{1.}} + \underbrace{\widehat{M}_R^x\left[g(x,\bar{r},v)-g(x,r,v)\right]}_{\textcircled{2.}} \\ & + \underbrace{\widehat{B}_V^x\left[g(x,r,v+1)-g(x,r,v)\right]}_{\textcircled{3.}} + \underbrace{\widehat{C}_V^{x,r}\left[g(x,r,v-1)-g(x,r,v)\right]}_{\textcircled{4.}}\end{aligned} \tag{6.5}$$

with $\bar{r} = \begin{cases} 0, & \text{if } r=1 \\ 1, & \text{if } r=0 \end{cases}$.

1. The first part of the generator describes the dynamics of the **Wright-Fisher diffusion** process X. Due to time reversibility, we have $\widehat{G}_X = G_X$.

2. The second part is the generator of the **mutation process on the real line**, given $X = x$,

$$\widehat{M}_R^x = \frac{1-x}{x}\theta\nu_0\mathbf{1}_{\{r=0\}} + \frac{x}{1-x}\theta\nu_1\mathbf{1}_{\{r=1\}}\,. \tag{6.6}$$

The factors $(1-x)/x$ and $x/(1-x)$ in front of the forward in time mutation rates $\theta\nu_0$ and $\theta\nu_1$ are time change factors of a structured coalescent (they also appear e.g. in [KDH88, (6) and (7)]) and can be understood intuitively in the following manner.
The forward in time mutation rate per line to type 0 is $\theta\nu_0$ (silent mutations from type 0 to type 0 included). Mutation events from type 1 to type 0 (forward in time) happen when the line hit by the mutation carries type 1. As, for each line, the probability to inherit type 1 is $1-x$ (when the frequency of type 0 in the population at the time of the mutation is x), the overall rate at which a mutation from type 1 to type 0 happens is $(1-x)\theta\nu_0$ per line. Thus, backward in time the overall mutation rate from type 0 to type 1 per line should be $(1-x)\theta\nu_0$ as well. But as such a mutation can only affect a line of type 0 (at probability x), the backward in time mutation rate to type 1 of each type-0 line has to equal $\theta\nu_0\,(1-x)/x$.
In detail, the quantity $\lim_{t\to 0}\mathbb{P}(R_0=1,R_t=0 \mid X_0=x)/t$ can be calculated by starting with one line of type 1 and letting it mutate *forward in time* to type 0. This yields $\lim_{t\to 0}\mathbb{P}(R_0=1,R_t=0 \mid X_0=x)/t = (1-x)\theta\nu_0$. Or it can be evaluated by starting with one line of type 0 and letting it mutate *backward in time* to type 1, which gives $\lim_{r\to 0}\mathbb{P}(R_0=0,R_r=1 \mid X_0=x)/r = x\cdot\widehat{m}_R^x(0,1)$.

[18] Remember that the index '$\widehat{\ }$' indicates the direction *backward in time*.

Thus, the backward in time mutation rate $\widehat{m}_R^x(0,1)$ on the real line from type 0 to type 1 is given by $\theta\nu_0\ (1-x)/x$. Equivalently, we have $\widehat{m}_R^x(1,0) = \theta\nu_1\ x/(1-x)$.

3. The third part of the generator describes an increase of the number of virtual lines due to **selection**, given $X = x$,

$$\widehat{B}_V^x = (\nu+1)\sigma(1-x)\ . \tag{6.7}$$

A branching event backward in time is the analogue to a selective reproduction event forward in time (compare e.g. [KN97], [Fea02]). The forward in time rate of a selective reproduction event concerning one particular particle is given by σ. Since we have ν virtual lines and the real line, the total rate at which selective events arrive forward in time is $(\nu+1)\sigma$. Backward in time a branching means an appearance of one additional line. Here, the branching can only happen if this line is a virtual one and has type 1, which has probability $1-x$. Therefore the overall branching rate is $(\nu+1)\sigma(1-x)$.

4. The fourth part describes a decrease of the number of virtual lines due to **coalescence or mutation**, given $X = x$,

$$\widehat{C}_V^{x,r} = \left[\nu\theta\nu_1 + \frac{1}{2}\nu(\nu-1)\mathbf{1}_{\{r=0\}} + \frac{1}{2}\nu(\nu+1)\mathbf{1}_{\{r=1\}}\right]\frac{1}{1-x}\ . \tag{6.8}$$

Forwards in time, a mutation to type 1 (from type 0 or a silent mutation from type 1) per line arrives at rate $\theta\nu_1$. Thus, the overall mutation rate to type 1 of ν lines forward in time is $\nu\theta\nu_1$. Therefore, the backward in time mutation rate from type 1 (to type 0 or 1) must be equal to $\nu\theta\nu_1/(1-x)$, because the probability for one particle to be of type 1 is $1-x$ and $(1-x)\cdot\nu\theta\nu_1/(1-x) = \nu\theta\nu_1$.
The rate for a branching event forward in time is just the number of (unordered) pairs (between which the "reproductive arrow" can be interchanged). Backward in time we have $\nu(\nu-1)/2$ of these pairs of type-1 particles if the real line is of type 0 and $\nu(\nu+1)/2$ if the real line is of type 1. Since the vanishing line is of type 1, this rate has to be multiplied again by $1/(1-x)$. We have the total backward in time coalescence rate $\left[(1/2)\nu(\nu-1)\mathbf{1}_{\{r=0\}} + (1/2)(\nu+1)\nu\mathbf{1}_{\{r=1\}}\right]/(1-x)$.

Remark 6.2 *Backward in time, the process (X,R,V) can be decomposed in the following way:*

1. *In the first step, generate a random environment $X = (X_r)_{r\in\mathbb{R}}$ with respect to the Wright-Fisher generator G_X.*
2. *Then, in the second step, given X, generate the common ancestor process $(R_r,V_r)_{r\in\mathbb{R}}$ in this random environment with respect to the conditioned generator $\widehat{M}_R^x + \widehat{B}_V^x + \widehat{C}_V^{x,r}$.*

6.3 Stationary distribution of the triple process (X,R,V)

Let $\varphi(x,r,\nu)$ be the density of the equilibrium distribution (with respect to the Lebesgue measure on $[0,1]$ times counting measure on $\{0,1\}\times\mathbb{N}_0$) of the triple process (X,R,V). Following Taylor's ideas of using a product ansatz for the stationary distribution ((2.38) in this thesis), we consider a disintegration for φ,

$$\varphi(x,n,r) = w(x)\cdot\phi^x(n,r), \tag{6.9}$$

with the stationary densitiy w of the process X, given by (6.2) and $\phi^x(r,v) = \mathbb{P}\big((R_s,V_s) = (r,v) \mid X_s = x\big)$. Taking a closer look at (2.34) or [Fea02, Remark 3], one can make the educated guess

$$\begin{aligned} \phi^x(0,v) &= a_v x(1-x)^v, \\ \phi^x(1,v) &= (a_v - a_{v+1})(1-x)^{v+1}, \end{aligned} \tag{6.10}$$

with Fearnhead's coefficients $(a_v)_{v\in\mathbb{N}_0}$ determined by Remark 2.1,

$$\begin{aligned} &\left[\frac{1}{2}(v+1)+\sigma+\theta\right] a_v = \left[\frac{1}{2}(v+1)+\theta v_1\right] a_{v+1} + \sigma a_{v-1}, \quad v \geq 1, \\ &\text{with} \quad a_0 = 1 \ \text{ and } \ \lim_{v\to\infty} a_v = 0. \end{aligned} \tag{6.11}$$

Indeed, we have the following theorem.

Theorem 6.3 *The density φ of the stationary distribution of the process (X,R,V) is given by*

$$\begin{aligned} \varphi(x,0,v) &= w(x)\cdot a_v x(1-x)^v, \\ \varphi(x,1,v) &= w(x)\cdot (a_v - a_{v+1})(1-x)^{v+1}, \end{aligned} \tag{6.12}$$

for $x \in [0,1]$, $v \in \mathbb{N}_0$, and with Fearnhead's coefficients (a_v) defined by (6.11).

Let us first explain (6.12) intuitively before giving an analytical proof.

The intuition behind this theorem becomes clear when considering the following facts. The common ancestor process $(R_r,V_r)_{r\in\mathbb{R}}$ can be embedded in the pruned LD-ASG with the line counting process $(L_r)_{r\in\mathbb{R}}$ (introduced in Chapter 3). Let us assume that we are in an equilibrium situation at time 0. Then a realisation of (X_0,R_0,V_0) can be gained by first generating a realisation of X_0 and a realisation of L_0 independently. Then, in the footsteps of the *Fearnhead simulator* (see Page 27), given $X_0 = x$ and $L_0 = \ell$, we can get the realisation of R_0 and V_0 via a Bernoulli experiment. In detail, we toss ℓ times a coin with success probability x. If there is at least one success among the ℓ tosses, we set V_0 equal to the number of unsuccessful tosses before the first success, and $R_0 = 0$. If all ℓ tosses were unsuccessful, we set $V_0 = \ell - 1$ and $R_0 = 1$.
Let C_0 be the time of the first success in a coin tossing experiment with success probability X_0 and remember from Section 3 that Fearnhead's coefficients are tail probabilities of L, $a_v = \mathbb{P}(L_0 > v)$. Then we have $\mathbb{P}((R_0,V_0) = (0,v) \mid X_0 = x) = \mathbb{P}(C_0 = v+1, L_0 > v \mid X_0 = x)$ and $\mathbb{P}((R_0,V_0) = (1,v) \mid X_0 = x) = \mathbb{P}(C_0 > v+1, L_0 = v+1 \mid X_0 = x)$. Therefore, we get

$$\begin{aligned} \mathbb{P}((X_0,R_0,V_0) = (x,0,v)) &= w(x)\cdot a_v \mathbb{P}(C_0 = v+1 \mid X_0 = x), \\ \mathbb{P}((X_0,R_0,V_0) = (x,1,v)) &= w(x)\cdot (a_v - a_{v+1})\mathbb{P}(C_0 > v+1 \mid X_0 = x), \end{aligned} \tag{6.13}$$

which is exactly (6.12).

Proof of Theorem 6.3. φ is the density of the stationary distribution of the triple process (X,R,V) if and only if $\int_{[0,1]} dx \sum_{r\in\{0,1\}} \sum_{v\geq 0} \big[\widehat{G}_{(X,R,V)} g(x,r,v)\big] \varphi(x,r,v) = 0$ for all g in the domain of $\widehat{G}_{(X,R,V)}$ (compare [EK86, Theorem 9.17, Chapter 4]). For this, via integration by parts, it suffices to show

$$\widehat{G}^*_{(X,R,V)} \varphi(x,r,v) = 0 \tag{6.14}$$

for all $x \in (0,1)$, $r \in \{0,1\}$, $v \in \mathbb{N}_0$, where $\widehat{G}^*_{(X,R,V)}$ is the formal adjoint operator of $\widehat{G}_{(X,R,V)}$. To

simplify this condition we decompose $\widehat{G}_{(X,R,V)}$ into two parts:

$$\widehat{G}_{(X,R,V)} = G_x + \widehat{F} \tag{6.15}$$

with the Wright-Fisher generator G_X given by (6.1) and the jump generator $\widehat{F}$,

$$\begin{aligned}\widehat{F}g(x,r,v) \;=\; & \left[\frac{1-x}{x}\theta\nu_0\mathbf{1}_{\{r=0\}} + \frac{x}{1-x}\theta\nu_1\mathbf{1}_{\{r=1\}}\right][g(x,\bar{r},v) - g(x,r,v)] \\ & + (v+1)\sigma(1-x)\,[g(x,r,v+1) - g(x,r,v)] \\ & + \left[v\theta\nu_1 + \frac{1}{2}v(v-1)\mathbf{1}_{\{r=0\}} + \frac{1}{2}v(v+1)\mathbf{1}_{\{r=1\}}\right]\frac{1}{1-x}\,[g(x,r,v-1) - g(x,r,v)]\end{aligned} \tag{6.16}$$

for all $g \in \mathscr{C}^2(0,1)\times\{0,1\}\times\mathbb{N}_0$; $\bar{r}=1$ if $r=0$ and $\bar{r}=0$ if $r=1$. Note that $\widehat{F}$ depends on x but we omit the index here for simplicity.
The formal adjoint operator $\widehat{G}^*_{(X,R,V)}$ of $\widehat{G}_{(X,R,V)}$ can therefore be written as

$$\widehat{G}^*_{(X,R,V)} = {G_X}^* + \widehat{F}^*. \tag{6.17}$$

Thus, (6.14) is satisfied if

$$({G_X}^* + \widehat{F}^*)\varphi = 0\,. \tag{6.18}$$

As w is stationary density, it follows by the same argument as in (6.14) that ${G_X}^*[w(x)g(x)] = w(x)\cdot G_X g(x) \quad \forall g \in \mathscr{C}^2[0,1]$, which again simplifies (6.18) into

$$(G_X + \widehat{F}^*)\phi^x = 0\,. \tag{6.19}$$

Note that $\widehat{F}$ is a generator matrix (also known as a 'Q-matrix'). Thus, the formal adjoint operator $\widehat{F}^*$ of $\widehat{F}$ is just given by the transposition of $\widehat{F}$.

Now, here we only show $(G_X + \widehat{F}^*)\phi^x(0,v) = 0$. The proof of $(G_X + \widehat{F}^*)\phi^x(1,v) = 0$ works completely analogously.

To calculate $G_X\phi^x(0,v)$, we first need the derivatives of $\phi^x(0,v)$, expressed in terms of $\phi^x(0,v)$,

$$\begin{aligned}\frac{d}{dx}\phi^x(0,v) \;&=\; \left(\frac{1}{x} - \frac{v}{1-x}\right)\phi^x(0,v), \\ \frac{d^2}{dx^2}\phi^x(0,v) \;&=\; \frac{v(x-2+vx)}{x(1-x)^2}\phi^x(0,v)\,.\end{aligned}$$

We have

$$G_X\phi^x(0,v) = \frac{1}{2}x(1-x)\frac{d^2}{dx^2}\phi^x(0,v) + [\theta\nu_0(1-x) - \theta\nu_1 x + \sigma x(1-x)]\frac{d}{dx}\phi^x(0,v),$$

which is equal to

$$G_X\phi^x(0,v) = \phi^x(0,v)\left[(v+1)\sigma(1-x) + \left[\frac{1}{2}v(v-1) + v\theta\nu_1\right]\frac{1}{1-x} + \theta\nu_0\frac{1}{x} - \frac{1}{2}(v+1)v - \theta(v+1) - v\sigma\right]. \tag{6.20}$$

To express $\widehat{F}^*\phi^x(0,v)$ in terms of $\phi^x(0,v)$, the following equalities are helpful,

$$\phi^x(0,v+1) \;=\; \frac{a_{v+1}}{a_v}(1-x)\phi^x(0,v),$$

$$\phi^x(0,v-1) \;=\; \frac{a_{v-1}}{a_v}\frac{1}{1-x}\phi^x(0,v),$$

$$\phi^x(1,v) \;=\; \frac{1-x}{x}\left(1-\frac{a_{v+1}}{a_v}\right)\phi^x(0,v)\;.$$

Thus,

$$\widehat{F}^*\phi^x(0,v) = \theta\nu_1\frac{x}{1-x}\phi^x(1,v) + \left[(v+1)\theta\nu_1+\frac{1}{2}v(v+1)\right]\frac{1}{1-x}\phi^x(0,v+1) + v\sigma(1-x)\phi^x(0,v-1)$$
$$-\left[\theta\nu_0\frac{1-x}{x}+\left[v\theta\nu_1+\frac{1}{2}v(v-1)\right]\frac{1}{1-x}+(v+1)\sigma(1-x)\right]\phi^x(0,v),$$

which reads in terms of factors of $\phi^x(0,v)$

$$\widehat{F}^*\phi^x(0,v) = \phi^x(0,v)\Bigg\{\theta + v\theta\nu_1\frac{a_{v+1}}{a_v}+\frac{1}{2}v(v+1)\frac{a_{v+1}}{a_v}+v\sigma\frac{a_{v-1}}{a_v}$$
$$-\left[\theta\nu_0\frac{1}{x}+\left[v\theta\nu_1+\frac{1}{2}v(v-1)\right]\frac{1}{1-x}+(v+1)\sigma(1-x)\right]\Bigg\}. \tag{6.21}$$

Taking (6.20) and (6.21) together, we see that all terms depending on x in the sum of both generator parts cancel each other,

$$(G_X+\widehat{F}^*)\phi^x(0,v) = \;\phi^x(0,v)\left[v\theta\nu_1\frac{a_{v+1}}{a_v}+\frac{1}{2}v(v+1)\frac{a_{v+1}}{a_v}+v\sigma\frac{a_{v-1}}{a_v}-\frac{1}{2}(v+1)v-v\theta-v\sigma\right].$$

Therefore, $\widehat{G}^*_{(X,R,V)}\varphi(x,0,v)=0$ if and only if

$$0 = v\theta\nu_1\frac{a_{v+1}}{a_v}+v(v+1)\frac{a_{v+1}}{a_v}+v\sigma\frac{a_{v-1}}{a_v}-(v+1)v-v\theta-v\sigma \tag{6.22}$$

which is equivalent to Fearnhead's recursion (6.11). □

6.4 The forward in time generator of (X,R,V)

In this section we reverse time and discuss the evolution of the triple process (X,R,V) forward in time. We first state the theorem that gives the generator. Then we discuss the heuristics and intuition and we end this section with proving the theorem.

Theorem 6.4 *The generator $G_{(X,R,V)}$ of the process $(X_t,V_t,R_t)_{t\in\mathbb{R}}$ forward in time is given by*

$$\begin{aligned} G_{(X,R,V)}g(x,0,v) \;=\;& G_X g(x,0,v)+(1-x-vx)g'(x,0,v)\\ &+\theta\nu_1\frac{a_v-a_{v+1}}{a_v}\left[g(x,1,v)-g(x,0,v)\right]\\ &+v\sigma\frac{a_{v-1}}{a_v}\left[g(x,0,v-1)-g(x,0,v)\right]\\ &+\left[(v+1)\theta\nu_1+\frac{1}{2}v(v+1)\right]\frac{a_{v+1}}{a_v}\left[g(x,0,v+1)-g(x,0,v)\right] \end{aligned} \tag{6.23}$$

$$\begin{aligned} G_{(X,R,V)}g(x,1,v) &= G_X g(x,1,v) - x(v+1)g'(x,1,v) \\ &\quad + \theta\nu_0 \frac{a_v}{a_v - a_{v+1}} \left[g(x,0,v) - g(x,1,v)\right] \\ &\quad + \nu\sigma \frac{a_{v-1} - a_v}{a_v - a_{v+1}} \left[g(x,1,v-1) - g(x,1,v)\right] \\ &\quad + \left[(v+1)\theta\nu_1 + \frac{1}{2}(v+1)(v+2)\right] \frac{a_{v+1} - a_{v+2}}{a_v - a_{v+1}} \left[g(x,1,v+1) - g(x,1,v)\right] \end{aligned} \tag{6.24}$$

for all $g \in \mathscr{C}^2[0,1] \times \{0,1\} \times \mathbb{N}_0$.

Lemma 6.5 *$G_{(X,R,V)}$ can be decomposed in the following way*

$$G_{(X,R,V)}g = G_X g + x(1-x)\frac{\phi^{x\prime}}{\phi^x} g' + Fg, \tag{6.25}$$

with the generator matrix F, defined by

$$\begin{aligned} Fg(x,r,v) &= \left[\theta\nu_1 \frac{a_v - a_{v+1}}{a_v} \mathbf{1}_{\{r=0\}} + \theta\nu_0 \frac{a_v}{a_v - a_{v+1}} \mathbf{1}_{\{r=1\}}\right] \left[g(x,\bar{r},v) - g(x,r,v)\right] \\ &\quad + \left[\nu\sigma \frac{a_{v-1}}{a_v} \mathbf{1}_{\{r=0\}} + \nu\sigma \frac{a_{v-1} - a_v}{a_v - a_{v+1}} \mathbf{1}_{\{r=1\}}\right] \left[g(x,r,v-1) - g(x,r,v)\right] \\ &\quad + \Bigg[\left((v+1)\theta\nu_1 + \frac{1}{2}\nu(v+1)\right) \frac{a_{v+1}}{a_v} \mathbf{1}_{\{r=0\}} \\ &\qquad + \left((v+1)\theta\nu_1 + \frac{1}{2}(v+1)(v+2)\right) \frac{a_{v+1} - a_{v+2}}{a_v - a_{v+1}} \mathbf{1}_{\{r=1\}}\Bigg] \left[g(x,r,v+1) - g(x,r,v)\right] \end{aligned} \tag{6.26}$$

for all $g \in \mathscr{C}^2[0,1] \times \mathbb{N}_0 \times \{0,1\}$ *with* $\bar{r} = \begin{cases} 0, & \text{if } r = 1, \\ 1, & \text{if } r = 0\,. \end{cases}$

Note that Lemma 6.5 is equivalent to Theorem 6.4 because we have

$$x(1-x)\frac{\phi^x(v,0)'}{\phi^x(v,0)} = 1 - x - \nu x\,, \quad x(1-x)\frac{\phi^x(v,1)'}{\phi^x(v,1)} = -x(v+1). \tag{6.27}$$

Let us first explain the rates of the generator $G_{(X,R,V)}$ before proving Theorem 6.4.
Note that the rates of $G_{(X,R,V)}$ that indicate a change of (R,V) do not depend on X. Therefore, we have the following remark.

Remark 6.6 *Forward in time, the process (X,R,V) can be decomposed in the following way:*

1. *In the first step, generate the common ancestor process $(R_r, V_r)_{r\in\mathbb{R}}$ with respect to the generator F.*
2. *Then, in the second step, given (R,V), generate the conditioned Wright-Fisher diffusion with selection and mutation $(X_r)_{r\in\mathbb{R}}$ in the (R,V)-environment with respect to the conditioned generator $G_x + x(1-x)\frac{\phi^{x\prime}}{\phi^x}\frac{d}{dx}$.*

Interpretation of the generator $G_{(X,R,V)}$

For understanding the generator $G_{(X,R,V)}$ intuitively, it is essential to point out the following two facts connecting the common ancestor process and the pruned LD-ASG (that were already discussed on Page 91): The event $\{(R,V)=(1,v)\}$ yields $\{L=v+1\}$ and $\{(R,V)=(0,v)\}$ results in $\{L>v\}$.

The forward in time generator $G_{(X,R,V)}$ of the process (X,V,R) can be interpreted in terms of its different transition rates. To this end, let us decompose $G_{(X,R,V)}$ into its different summands:

$$\begin{aligned} G_{(X,R,V)}g(x,r,v) \;=\; & \underbrace{G_X g(x,r,v)}_{\textcircled{1.}} + \underbrace{A_X^{r,v} g(x,r,v)}_{\textcircled{2.}} + \underbrace{M_R^v\left[g(x,\bar{r},v)-g(x,r,v)\right]}_{\textcircled{3.}} \\ & + \underbrace{B_V^r\left[g(x,r,v-1)-g(x,r,v)\right]}_{\textcircled{4.}} + \underbrace{C_V^r\left[g(x,r,v+1)-g(x,r,v)\right]}_{\textcircled{5.}} \end{aligned} \tag{6.28}$$

1. The first part of the generator describes just the unconditioned dynamics of the **Wright-Fisher diffusion** X with selection and mutation.

2. The second part gives an **additional drift term** for X, given $(R,V)=(r,v)$,

$$A_X^{r,v} = x(1-x)\frac{\phi^{x\prime}}{\phi^x}\frac{d}{dx}\,. \tag{6.29}$$

This term already appears in the forward in time generator of the process (X,R) (compare [Tay07, Equation (15)]) and Taylor explains that it "reflects the fact that because the common ancestor contributes more offspring to the population than an individual chosen at random, the population has a tendency to evolve towards the type of the common ancestor" [Tay07, p. 824].
The random environment (R,V) applies a drift on the frequency X of type-0 individuals. As shown in [HP86, Theorem2.1] (compare also [FW86] for the case of a Brownian motion), this additional drift coefficient is given by a product of the diffusion coefficient of the Wright-Fisher diffusion times the gradient of the logarithm of the stationary distribution of (R,V). This is equal to $x(1-x)\nabla\log(\phi^x) = x(1-x)\frac{\phi^{x\prime}}{\phi^x}$.

3. The third part is the generator of the **mutation process along the real line**, given $(R,V)=(r,v)$,

$$M_R^v = \theta\nu_1\frac{a_v-a_{v+1}}{a_v}\mathbf{1}_{\{r=0\}} + \theta\nu_0\frac{a_v}{a_v-a_{v+1}}\mathbf{1}_{\{r=1\}}\,. \tag{6.30}$$

Let us consider the case $r=0$. The additional factor $(a_v-a_{v+1})/a_v$ to the mutation rate $\theta\nu_0$ per line is comparable to the structured coalescent factor $(1-x)/x$ of the backward in time generator $\widehat{G}_{(X,R,V)}$. The explanation of this factor is therefore analogous to the explanation of item 2 on Page 89:
We have two possibilities to calculate $\lim_{t\to 0}\mathbb{P}\big((R_0,V_0)=(0,v),(R_t,V_t)=(1,v)\mid X_0=x\big)/t$. On the one hand, backward in time, we can start the pruned LD-ASG with $v+1$ lines, colour them all with type 1, and let the real line mutate to type 0. This results in

$$\lim_{r\to 0}\frac{1}{r}\mathbb{P}\big((R_0,V_0)=(1,v),(R_r,V_r)=(0,v)\mid X_0=x\big) = (a_v-a_{v+1})\cdot(1-x)^{v+1}\cdot\theta\nu_1\frac{x}{1-x}.$$

On the other hand, forward in time, we can start with the pruned LD-ASG having more than v lines, colour the first v lines with type 1 and line $v+1$ with type 0, and then let the real line mutate to type

1 at rate $m_R^v(0,1)$. Then

$$\lim_{t\to 0}\frac{1}{t}\mathbb{P}\big((R_0,V_0)=(0,v),(R_t,V_t)=(1,v)\mid X_0=x\big)=a_v\cdot x(1-x)^v\cdot m_R^v(0,1)$$

which results in $m_R^v(0,1)=\theta v_1(a_v-a_{v+1})/a_v$. The explanation of the rate in the case $r=1$ follows the same line of argument.

4. The fourth part is the generator describing a decrease of the number of virtual lines due to **selective events** given $X=x$,

$$B_V^r=v\sigma\frac{a_{v-1}}{a_v}\mathbf{1}_{\{r=0\}}+v\sigma\frac{a_{v-1}-a_v}{a_v-a_{v+1}}\mathbf{1}_{\{r=1\}}\,. \tag{6.31}$$

Let us again consider the case $r=0$. (The case $r=1$ then works analogously). Backward in time, we can start the pruned LD-ASG with at least v lines, colour the first $v-1$ with the deleterious type and line v with the beneficial one. Then, as a branching event happens at rate $v\sigma(1-x)$,

$$\lim_{r\to 0}\frac{1}{r}\mathbb{P}\big((R_0,V_0)=(0,v-1),(R_r,V_r)=(0,v)\mid X_0=x\big)=a_{v-1}\cdot x(1-x)^{v-1}\cdot v\sigma(1-x).$$

Forward in time, the pruned LD-ASG can be started with more than v lines, v of them at type 1 and line $v+1$ at type 0. Then

$$\lim_{t\to 0}\frac{1}{t}\mathbb{P}\big((R_0,V_0)=(0,v),(R_t,V_t)=(0,v-1)\mid X_0=x\big)=a_v\cdot x(1-x)^v\cdot b_V^0(v,v-1),$$

and therefore $b_V^0(v,v-1)=v\sigma a_{v-1}/a_v$.

5. The fifth part describes an increase of the number of virtual particles due to **neutral reproduction or mutation**, given $X=x$,

$$C_V^r=\left[(v+1)\theta v_1+\frac{1}{2}v(v+1)\right]\frac{a_{v+1}}{a_v}\mathbf{1}_{\{r=0\}}+\left[(v+1)\theta v_1+\frac{1}{2}(v+1)(v+2)\right]\frac{a_{v+1}-a_{v+2}}{a_v-a_{v+1}}\mathbf{1}_{\{r=1\}}\,. \tag{6.32}$$

We consider again $r=0$. Backward in time, when starting the pruned LD-ASG with at least $v+2$ lines (at least $v+1$ of type 1 and one of type 0) and considering coalescence between two virtual lines or the removal of one line due to a deleterious mutation,

$$\begin{aligned}\lim_{r\to 0}\frac{1}{r}\mathbb{P}\big((R_0,V_0)&=(0,v+1),(R_r,V_r)=(0,v),\mid X_0=x\big)\\&=a_{v+1}\cdot x(1-x)^{v+1}\cdot\left(\frac{1}{2}v(v+1)+(v+1)\theta v_1\right)\frac{1}{1-x}.\end{aligned}$$

Forwards in time, when starting with the pruned LD-ASG with more than v lines (at least v of type 1 and one of type 0), we have

$$\lim_{t\to 0}\frac{1}{t}\mathbb{P}\big((R_0,V_0)=(0,v),(R_t,V_t)=(0,v+1)\mid X_0=x\big)=a_v\cdot x(1-x)^v\cdot c_V^0(v,v+1).$$

This yields $c_V^0(v,v+1)=\big(\frac{1}{2}v(v+1)+(v+1)\theta v_1\big)\,a_{v+1}/a_v$.

Proof of Theorem 6.4. To prove Theorem 6.4, we have to reverse in time the backward in time generator $\widehat{G}_{(X,R,V)}$ with respect to its equilibrium distribution; we are searching for the time reversal of the generator $\widehat{G}_{(X,R,V)}$ given by Theorem 6.1. Since Theorem 6.4 is equivalent to Lemma 6.5, we

here prove the decomposition of $G_{(X,R,V)}$ given by (6.25),

$$G_{(X,R,V)} \;=\; G_X + x(1-x)\frac{\phi^{x\prime}}{\phi^x}\frac{d}{dx} + F\ . \tag{6.33}$$

The first term G_X of the backward in time generator $\widehat{G}_{(X,R,V)}$ describes the evolution of X according to the Wright-Fisher dynamics. As the Wright-Fisher diffusion is time reversible with respect to w, one part of the forward in time generator $G_{(X,R,V)}$ is required to be again G_X. But since the evolution of R and V influences the dynamics of X, two additional summands arise. We will see that they are given by $x(1-x)\frac{\phi^{x\prime}}{\phi^x}\frac{d}{dx}$ (which is the second part of the generator in the decomposition above) and $\frac{G_X\phi^x}{\phi^x}$. The rates in the last part of the generator, the Q-matrix F, are the time reversal of the backward in time rates given by the Q-matrix $\widehat{F}$. But because a Q-matrix has to be *conservative* (i.e. has row sum zero), the diagonal elements in F are not just the diagonal elements in $\widehat{F}$. It turns out that the compensating summands are just $-\frac{G\phi^x}{\phi^x}$ which cancel out with the additional summands of the time reversal of the conditioned generator G_X, given R,V.
We now prove these statements properly.

Let H be a forward in time generator with stationary distribution ψ. Then the backward in time generator $\widehat{H}$ has to fulfil (compare [Nag64, (3.7)] or [Nel58])

$$\int g_1(Hg_2)d\psi = \int (\widehat{H}g_1)g_2 d\psi \qquad \text{for all test functions } g_1,g_2 \text{ in the domain of } H. \tag{6.34}$$

In our case, ψ is the stationary distribution of the process (X,R,V), we have $H=\widehat{G}_{(X,R,V)}$, and the generator we want to calculate is $\widehat{H}=\widehat{\widehat{G}}_{(X,R,V)}=G_{(X,R,V)}$. Our test functions (whose linear span is dense in $\mathscr{C}^2[0,1]\times\{0,1\}\times\mathbb{N}_0$) are

$$g_i : (x,r,v)\mapsto h_i(x)\cdot \mathbf{I}_{\{r,v\}},\quad x\in[0,1],\ r\in\{0,1\},\ v\in\mathbb{N}_0,\ \ i=1,2. \tag{6.35}$$

Let $\widehat{q}^x_{(R,V)}\big((j,l),(k,m)\big)$ be the backward in time and $q^x_{(R,V)}\big((k,m),(j,l)\big)$ the forward in time transition rate of (R,V) from state (j,l) to state (k,m) and from state (k,m) to state (j,l), given X, thus the element $((j,l),(k,m))$ of the Q-matrix $\widehat{F}$, (6.16), and the element $((k,m),(j,l))$ of the Q-matrix F (that we want to prove to be equal to (6.26)), respectively.

For the test functions $g_1(x,r,v) = h_1(x)\cdot\mathbf{I}_{\{(j,l)\}}(r,v)$, $g_2(x,r,v)=h_2(x)\cdot\mathbf{I}_{\{(k,m)\}}(r,v)$, $h_1(x),h_2(x)\in\mathscr{C}^2[0,1]$, we can distinguish between the two cases $(j,l)\neq(k,m)$ and $(j,l)=(k,m)$.

Let us start with the case $(j,l)\neq(k,m)$. By using the decomposition (6.15) of $\widehat{G}_{(X,R,V)}$, we get

$$\begin{aligned}
\int g_1(\widehat{G}_{(X,R,V)}\,g_2)d\psi &= \int_0^1 \sum_{(r,v)} h_1(x)\cdot\mathbf{I}_{\{(j,l)\}}(r,v)\Big(\widehat{G}_{(X,R,V)}h_2(x)\cdot\mathbf{I}_{\{(k,m)\}}\Big)(r,v)w(x)\phi^x(r,v)dx\\
&= \int_0^1 h_1(x)h_2(x)\cdot\Big(\widehat{F}\cdot\mathbf{I}_{\{(k,m)\}}\Big)(j,l)w(x)\phi^x(j,l)dx\\
&\quad + \int_0^1 \sum_{(r,v)} h_1(x)\cdot\mathbf{I}_{\{(j,l)\}}(r,v)\cdot\mathbf{I}_{\{(k,m)\}}(r,v)\Big(G_Xh_2(x)\Big)w(x)\phi^x(r,v)dx,
\end{aligned}$$

and since we are dealing with the case $(j,l)\neq(k,m)$, the second term on the right-hand side is equal

to zero. Rewriting $\widehat{F}$ in terms of its rates $\widehat{q}^{x}_{(R,V)}$ yields

$$\int g_1(\widehat{G}_{(X,R,V)}\, g_2)d\psi = \int_0^1 h_1(x)h_2(x)\cdot\Big(\widehat{q}^{x}_{(R,V)}\big((j,l),(k,m)\big)\frac{\phi^x(j,l)}{\phi^x(k,m)}\Big)w(x)\phi^x(k,m)dx+0.$$

On the other hand, by using the decomposition ansatz of $G_{(X,R,V)}$ into a generator matrix F of (R,V) and some generator that works on X , and keeping in mind that all summands containing operators that leave R and V invariant vanish due to $(j,l)\neq(k,m)$, we have

$$\begin{aligned}\int (G_{(X,R,V)}g_1)g_2 d\psi &= \int_0^1 \sum_{(r,v)}\Big(G_{(X,R,V)}h_1(x)\cdot\mathbf{I}_{\{(j,l)\}}\Big)(r,v)h_2(x)\cdot\mathbf{I}_{\{(k,m)\}}(r,v)w(x)\phi^x(r,v)dx\\ &= \int_0^1 h_1(x)h_2(x)\cdot\Big(q^{x}_{(R,V)}\big((k,m),(j,l)\big)\Big)w(x)\phi^x(k,m)dx+0.\end{aligned}$$

Thus, we require

$$q^{x}_{(R,V)}\big((k,m),(j,l)\big)=\widehat{q}^{x}_{(R,V)}\big((j,l),(k,m)\big)\frac{\phi^x(j,l)}{\phi^x(k,m)},\quad (j,l)\neq(k,m). \tag{6.36}$$

The part of the forward in time generator $G_{(X,R,V)}$ only concerning changes in (V,R), given X, can therefore be calculated easily by just plugging the backward in time rates $\widehat{q}^{x}_{(R,V)}$ (see Lemma 6.1) and the weights ϕ^x (see (6.10)) into (6.36). We arrive exactly at the generator (6.26).

Let us now continue with the case $(j,l)=(k,m)$: To get the drift and diffusion part concerning X, given (R,V), of the forward in time generator, we now choose $g_1(x,r,v)=h_1(x)\cdot\mathbf{I}_{\{(j,l)\}}(r,v)$, $g_2(x,r,v)=h_2(x)\cdot\mathbf{I}_{\{(j,l)\}}(r,v)$ for $j\in\{0,1\}$, $l\in\mathbb{N}_0$ and $h_1(x),h_2(x)\in\mathscr{C}^2[0,1]$. With the decomposition (6.15) of $\widehat{G}_{(X,R,V)}$ we have

$$\int g_1(\widehat{G}_{(X,R,V)}\, g_2)d\psi = \int g_1(G_X g_2)d\psi+\int g_1(\widehat{F}g_2)d\psi. \tag{6.37}$$

For the first term on the right-hand side of (6.37), let $b(x):=[(1-x)\theta\nu_0+x\theta\nu_1+\sigma x(1-x)]$. Then, using the generator G_X (6.1), we have

$$\int g_1(G_X g_2)d\psi = \int_0^1 w(x)\phi^x(r,v)h_1(x)\Big(\frac{1}{2}x(1-x)\frac{d^2}{dx^2}h_2(x)+b(x)\frac{d}{dx}h_2(x)\Big)dx,$$

and integration by parts gives

$$\int g_1(G_X g_2)d\psi = \int_0^1\Big[\frac{d^2}{dx^2}\Big(\frac{1}{2}x(1-x)w(x)h_1(x)\phi^x(r,v)\Big)-\frac{d}{dx}\Big(b(x)w(x)h_1(x)\phi^x(r,v)\Big)\Big]h_2(x)dx.$$

An application of the product rule and rearrangement of the summands yields

$$\begin{aligned}\int g_1(G_X g_2)d\psi = \int_0^1\Big[&\Big(\phi^x(r,v)w(x)G_X h_1(x)\Big)+2\frac{d}{dx}\Big(\phi^x(r,v)\Big)\frac{1}{2}x(1-x)w(x)\frac{d}{dx}\Big(h_1(x)\Big)\\ &+2\frac{d}{dx}\Big(\phi^x(r,v)\Big)b(x)w(x)h_1(x)+\frac{1}{2}x(1-x)w(x)h_1(x)\frac{d^2}{dx^2}\Big(\phi^x(r,v)\Big)\\ &-b(x)w(x)h_1(x)\frac{d}{dx}\Big(\phi^x(r,v)\Big)\Big]h_2(x)dx\end{aligned}$$

$$
\begin{aligned}
&= \int_0^1 \left[\left(\phi^x(r,v) w(x) G_X h_1(x) \right) + \phi^x(r,v) w(x) \cdot x(1-x) \frac{\frac{d}{dx}\left(\phi^x(r,v)\right)}{\phi^x(r,v)} \frac{d}{dx}\left(h_1(x)\right) \right. \\
&\qquad \left. + \left(\phi^x(r,v) w(x) \frac{G_X \phi^x(r,v)}{\phi^x(r,v)} h_1(x) \right) \right] h_2(x) dx \\
&= \int_0^1 \left[G_X h_1(x) + x(1-x) \frac{\frac{d}{dx}\left(\phi^x(r,v)\right)}{\phi^x(r,v)} \frac{d}{dx}\left(h_1(x)\right) \right. \\
&\qquad \left. + \left(\frac{G_X \phi^x(r,v)}{\phi^x(r,v)} h_1(x) \right) \right] h_2(x) \phi^x(r,v) w(x) dx.
\end{aligned}
$$

Thus, we have

$$
\int g_1 (G_X g_2) d\psi = \int \left(\left[G_X + x(1-x) \frac{\phi^{x\prime}}{\phi^x} \frac{d}{dx} + \frac{G_X \phi^x}{\phi^x} \right] g_1 \right) \cdot g_2 d\psi. \tag{6.38}
$$

The second term on the right-hand side of (6.37) is equal to

$$
\begin{aligned}
\int g_1 (\widehat{F} g_2) d\psi &= \int_0^1 \sum_{(r,v)} h_1(x) \cdot \mathbf{I}_{\{(j,l)\}}(r,v) \left(\widehat{F} h_2(x) \cdot \mathbf{I}_{\{(j,l)\}} \right)(r,v) w(x) \phi^x(r,v) dx \\
&= \int_0^1 \left(\widehat{F} \mathbf{I}_{\{(j,l)\}}(j,l) \cdot h_1(x) \right) h_2(x) \cdot w(x) \phi^x(r,v) dx.
\end{aligned} \tag{6.39}
$$

Further, using (6.16) and (6.20), we have

$$
\begin{aligned}
\left(\widehat{F} \mathbf{I}_{\{0,v\}} \right)(0,v) + \frac{G_X \phi^x(0,v)}{\phi^x(0,v)} &= -\frac{1-x}{x} \theta v_0 - (v+1)\sigma(1-x) - \left[v\theta v_1 + \frac{1}{2} v(v-1) \right] \frac{1}{1-x} \\
&\quad + (v+1)\sigma(1-x) + \left[\frac{1}{2} v(v-1) + v\theta v_1 \right] \frac{1}{1-x} + \theta v_0 \frac{1}{x} \\
&\quad - \frac{1}{2}(v+1)v - \theta(v+1) - v\sigma.
\end{aligned}
$$

Summing up terms and using Fearnhead's recursion (6.11) then gives

$$
\begin{aligned}
\left(\widehat{F} \mathbf{I}_{\{0,v\}} \right)(0,v) + \frac{G_X \phi^x(0,v)}{\phi^x(0,v)} &= -\frac{1}{2}(v+1)v - v\theta - \theta v_1 - v\sigma \\
&= -\left\{ \left[\frac{1}{2} v(v+1) + (v+1)\theta v_1 \right] \frac{a_{v+1}}{a_v} + v\sigma \frac{a_{v-1}}{a_v} + \theta v_1 \frac{a_v - a_{v+1}}{a_v} \right\} \\
&= \left(F \mathbf{I}_{\{0,v\}} \right)(0,v).
\end{aligned} \tag{6.40}
$$

Analogue, we also have

$$
\left(\widehat{F} \mathbf{I}_{\{1,v\}} \right)(1,v) + \frac{G_X \phi^x(1,v)}{\phi^x(1,v)} = \left(F \mathbf{I}_{\{1,v\}} \right)(1,v). \tag{6.41}
$$

Therefore, (6.37), (6.38), (6.39), and (6.40)/(6.41) together yield

$$
\int g_1 (\widehat{G}_{(X,R,V)} g_2) d\psi = \int \left(\left[G_X + x(1-x) \frac{\phi^{x\prime}}{\phi^x} \frac{d}{dx} + F \right] g_1 \right) g_2 d\psi,
$$

which completes the proof. □

6.5 The quadruple process (L,X,R,V)

So far, we have analysed the Markovian process (X,R,V) and interpreted its dynamics with the help of the stationary distribution of the line counting process L of the pruned LD-ASG. Now, we want to take a closer look at the evolution of all four components jointly together.
Recall from (3.9) in Chapter 3 the generator G_L of the line counting process (or the top level) L of the pruned LD-ASG,

$$\begin{aligned} G_L g(\ell) = &[\ell(\ell-1)+(\ell-1)\theta\nu_1]\cdot[g(\ell-1)-g(\ell)] \\ &+\ell\sigma\cdot[g(\ell+1)-g(\ell)] \;+\; \sum_{k=1}^{\ell-1}\theta\nu_0\cdot[g(\ell-k)-g(\ell)]\,. \end{aligned} \tag{6.42}$$

Recall from Section 3.5 the stationary tail probabilities of the line counting process of the pruned LD-ASG, $(a_\ell)_{\ell\in\mathbb{N}}$, $a_{\ell-1}=\mathbb{P}(L\geq\ell)$, and the stationary probability weights (ρ_ℓ), $\rho_\ell=\mathbb{P}(L=\ell)=a_{\ell-1}-a_\ell$.

Before discussing the (not any more Markovian) dynamics of the quadruple process (L,X,R,V) we identify its stationary distribution ϖ in the following section.

6.5.1 The stationary distribution of (L,X,R,V)

Throughout this section we assume all four processes L,X,R,V to be together in equilibrium at time 0. To simulate a sample from the stationary distribution of (L,X,R,V), first generate a realisation X_0 of the frequency of type-0 individuals in the population at time 0 according to Wright's density (6.2) and an independent realisation of L according to the probability weights ρ. Then, in the second step, given $X_0=x$ and $L_0=\ell$, the number of virtual lines V_0 and the type of the immortal line R_0 can be simulated according to Remark 3 in [Fea02] (Fearnhead's simulator, explained on Page 27 and already used on Page 91 in this chapter) as follows:
Toss ℓ times a coin with success probability x. If there is no success at all among the ℓ tosses, set $R_0=1$, $V_0=\ell-1$. Otherwise, $R_0=0$ and V_0 is the number of unsuccessful tries prior to the first success.

We have v virtual lines and the immortal line has type 0 if and only if the top level of the pruned LD-ASG is at least $v+1$, the first v lines in the pruned LD-ASG are assigned type 1, and the line on level $v+1$ is of type 0. This yields

$$\mathbb{P}((R_0,V_0)=(0,v)\mid L_0=\ell,X_0=x)=\mathbf{I}_{\{v<\ell\}}x(1-x)^v\,. \tag{6.43}$$

The type of the immortal line is 1 and the number of virtual lines equal to v if and only if there are exactly $v+1$ lines in the pruned LD-ASG and all lines are assigned type 1,

$$\mathbb{P}((R_0,V_0)=(1,v)\mid L_0=\ell,X_0=x)=\mathbf{I}_{\{v+1=\ell\}}(1-x)^{v+1}\,. \tag{6.44}$$

This yields the following lemma.

Lemma 6.7 *The density ϖ of the stationary distribution of the quadruple process (L,X,V,R) is given by*

$$\varpi(\ell,x,r,v) = w(x)\left[\mathbf{I}_{\{r=0,v<\ell\}}\rho_\ell \cdot x(1-x)^v + \mathbf{I}_{\{r=1,v+1=\ell\}}\rho_\ell \cdot (1-x)^{v+1}\right]. \tag{6.45}$$

Proof. Lemma 6.7 is implied by

$$\varpi(\ell,x,v,r) = \mathbb{P}((R_0,V_0)=(r,v) \mid L_0=\ell, X_0=x)\cdot\mathbb{P}(L_0=\ell)\cdot\mathbb{P}(X_0=x).$$

□

6.5.2 The dynamics of the process (L,X,V,R)

In Sections 6.2 and 6.4 we have analysed the backward and forward in time generators of the Markovian triple process (X,R,V).
But incorporating the line counting process L of the pruned LD-ASG into this triple leads us away from the class of Markov processes. Namely, the process (L,X,V,R) *is not Markovian* any more.

Indeed, the following (backward in time) counterexample shows that the quadruple process $(L_r,X_r,V_r,R_r)_{r\in\mathbb{R}}$ is not Markovian. Let us choose the initial state $(L_0,X_0,R_0,V_0)=(v+1,X_0,0,v)$. Let $\tau_1 := \min\{r\geq 0 : (L_r,R_r,V_r)\neq(L_0,R_0,V_0)\}$ and $(L_{\tau_1},X_{\tau_1},R_{\tau_1},V_{\tau_1})=(v+2,X_{\tau_1},0,v)$. Thus, at time τ_1 there is a branching event to an additional line that is included in the pruned LD-ASG but not included as a virtual line in the common ancestor process. Thus, since the line that branches is a virtual line of the common ancestor process but the newly born line is not included, this newly born line is required to be of type 0. This yields additional information (that is not coded in the state $(v+2,X_{\tau_1},0,v)$) which changes the transition rates. For example, each line of unknown type may be hit by a 'cross' backward in time with positive probability. But when a line is known to be of type 0, such a backward mutation has probability 0.
Let $\tau_2 := \min\{r\geq\tau_1 : (L_r,R_r,V_r)\neq(L_{\tau_1},R_{\tau_1},V_{\tau_1})\}$. Then, the probability for the transition at time τ_2 from state $(L_{\tau_1},X_{\tau_1},R_{\tau_1},V_{\tau_1})=(v+2,X_{\tau_1},0,v)$ to be to state $(L_{\tau_2},X_{\tau_2},R_{\tau_2},V_{\tau_2})=(v+1,X_{\tau_2},0,v)$ due to a 'cross' is strictly positive. But given $(L_0,X_0,R_0,V_0)=(v+1,X_0,0,v)$, this probability is equal to 0.

The counterexample indicates that a reason for the process (L,X,R,V) being not Markovian is the lack of knowledge of the types of those lines in the LD-ASG that are between levels V_r+1 and L_r. At each time $r\in\mathbb{R}$, the state (L_r,X_r,R_r,V_r) of the process does not contain the information on the types of L_r-V_r-1 many lines.
In fact, in order to again gain a Markov process, one has to include the information on the types of all lines in the pruned LD-ASG. Although we will not discuss it here, let us note that this would result in a rather complicated object similar to the *structured coalescent* of Barton, Etheridge, and Sturm [BES04] (reviewed briefly in Section 2.2.4).

7 The type of the common ancestor and the proportion of beneficial individuals in a population of finite size

In the previous chapter we have analysed the triple process (X,R,V) in the diffusion limit. In this chapter, we want to concentrate on Taylor's process (X,R) [Tay07] and come back to the case of a population of finite size $N \in \mathbb{N}$; the model was introduced in Section 2.1.2. We denote the number of type-0 individuals at time t by K_t^N and the proportion of type-0 individuals by $X_t^N = K_t^N/N$. Let R_t^N be the type of the immortal line in the finite population size model at time t.
The type of the common ancestor R^N together with the proportion X^N of type-0 individuals, $(X^N,R^N) = (X_t^N,R_t^N,)_{t\in\mathbb{R}}$, gives new insight into the underlying particle picture. In addition, by taking the diffusion limit $N \to \infty$, we get back the process (X,R) and therefore also alternative proofs for (2.37), (2.43) and (2.44).
Our approach recapitulates some results of Kluth, Hustedt, and Baake [KHB13], and by taking the diffusion limit, we also regain some of Taylor's results [Tay07].

In the following section we revisit the model. Then we analyse the backward and forward in time dynamics of the process (K^N,R^N), the discrete analogue of Taylor's process (X,R). In the last part of this chapter, we give some future perspectives concerning a discrete triple process (K^N,R^N,V^N), where V^N is the number of Fearnhead's virtual lines [Fea02] in the finite population size model.
Since many proofs consist of straightforward but lengthy calculations, they are shifted to the appendix.

7.1 Introduction

Let us consider a population of fixed size $N \in \mathbb{N}$. Each individual is marked either with the beneficial type 0 or the deleterious type 1. We define K_t^N as the number of individuals of type 0 at (forward) time[19] $t \geq 0$ and $X_t^N := K_t^N/N$ their fraction in the population. Each individual of type 1 reproduces independently at rate $1/2$ and its offspring replaces a randomly chosen individual of the population. In addition to these neutral reproduction events, individuals of type 0 can also reproduce in terms of selective events. The reproduction rate of each individual of type 0 is $1/2+s_N$. In addition, each individual mutates independently at rate $u_N v_0$ to type 0 and at rate $u_N v_1$ to type 1, $u_N, v_0, v_1 \geq 0$, $v_0 + v_1 = 1$.
The process (K_t^N) is a birth and death process with birth rate λ_k^N and death rate μ_k^N, $k \in \{0,1,2,\ldots,N\}$,

[19] As in the previous chapter, we denote *forward time* by 't' and the *backward time* by 'r'.

already given in (2.3),

$$\begin{aligned}\lambda_k^N &= \frac{1}{2N}k(N-k) + k(N-k)\cdot\frac{s_N}{N} \;+\; (N-k)\cdot u_N\nu_0,\\ \mu_k^N &= \frac{1}{2N}k(N-k) \;+\; k\cdot u_N\nu_1.\end{aligned} \tag{7.1}$$

Let $(w^N(k))_{k=0,1,\ldots,N}$ be the stationary probability vector of the rates (7.1) of the process K^N. Then an easy calculation (just check that (7.3) holds) shows that w^N is given by

$$w^N(k) = c_W^N \cdot \prod_{i=1}^{k} \frac{\lambda_{k-1}^N}{\mu_k^N}\,, \quad 0 \le k \le N. \tag{7.2}$$

with the normalising constant c_W^N. Due to time reversibility, the backward and forward in time rates of K^N agree,

$$w^N(k)\lambda_k^N = w^N(k+1)\mu_{k+1}^N, \quad k = 0,1,\ldots,N-1. \tag{7.3}$$

As we are also interested of the limit processes, we assume

$$N\cdot s_N \to \sigma, \quad N\cdot u_N \to \theta, \quad \text{and} \quad X_0^N \to x_0 \text{ for some } x_0 \in [0,1] \qquad \text{as } N\to\infty. \tag{7.4}$$

Then it is well-known (see e.g. [Dur08, Chapter 7.2]) that the diffusion limit $(X_t)_{t\ge0} := \lim_{N\to\infty}(X_{Nt}^N)_{t\ge0}$ of the proportion process of type-0 individuals exists and is a Wright-Fisher diffusion (with generator G_X) with drift coefficient $(1-x)\theta\nu_0 + x\theta\nu_1 + \sigma x(1-x)$ and diffusion coefficient $x(1-x)$. The proof is also given in Section 7.4.1 of the appendix (Theorem 7.6).

In the following section we set up a discrete analogue for Taylor's process (X,R) [Tay07] by considering a tuple $(X_t^N, R_t^N)_{t\ge0}$, where R_t^N is the type of the common ancestor in our discrete setting. In fact, we consider the tuple $(K_t^N, R_t^N)_{t\ge0}$. But there is no big difference because K^N and X^N only differ by a factor N.
While the discrete approach by Kluth, Hustedt, and Baake [KHB13] characterises the distribution of R_t^N by taking all individuals of type 0 at time t and calculating their fixation probability, we are more interested in the evolving picture that deals with the dynamics of (K^N, R^N) here. Thus, we are able to regain results of Kluth, Hustedt, and Baake with a different approach.

7.2 The process (K^N, R^N)

In this section we present the generators (forward and backward in time) of the discrete process (K^N, R^N), and compare them with results by Kluth, Hustedt, and Baake [KHB13] (some of them are briefly reviewed in Section 2.2.1), and with Taylor's boundary value problem [Tay07, Equation (9)] by passing to the limit.

7.2.1 The backward in time generator of (K^N, R^N)

Forward in time rates of K^N

To add the evolution of the type R^N of the immortal line to the evolution of the number K^N of type-0 individuals, we have to decompose the transition rates λ^N and μ^N with respect to their two 'ingredients' *reproduction* and *mutation*. This gives an enlargement of the state space in the footsteps of the Markov mapping theorem (compare [Kel82]).
Let us denote the forward in time rate of K^N from state k to state ℓ, $k,\ell = 0,1,\ldots,N$, $k \neq \ell$, by $q_{K^N}(k,\ell)$. Then we have

- $q_{K^N}(k,k+1;\text{ due to mutation}) \;=\; u_N \nu_0 (N-k)$,
- $q_{K^N}(k,k+1;\text{ due to reproduction}) \;=\; k\frac{(N-k)}{N}\left(\frac{1}{2}+s_N\right)$,
- $q_{K^N}(k,k-1;\text{ due to mutation}) \;=\; u_N \nu_1 k$,
- $q_{K^N}(k,k-1;\text{ due to reproduction}) \;=\; (N-k)\frac{k}{2N}$.

Backward in time rates of K^N

Let $\widehat{q}_{K^N}(k,\ell)$ be the backward in time rate[20] of the process K^N from state k to state ℓ, $k,\ell = 0,1,\ldots,N$, $k \neq \ell$.
Again enlarge the state space and also decompose the backward in time rates into their two sources: reproduction and mutation. To do so, and as it will become more convenient later, we rewrite the rates using (7.3) and we get

- $\widehat{q}_{K^N}(k,k+1;\text{ due to mutation}) \;=\; u_N \nu_1 (k+1)\frac{w^N(k+1)}{w^N(k)}$,
- $\widehat{q}_{K^N}(k,k+1;\text{ due to reproduction}) \;=\; \left[\mu^N_{k+1} - u_N \nu_1 (k+1)\right]\frac{w^N(k+1)}{w^N(k)}$,
- $\widehat{q}_{K^N}(k,k-1;\text{ due to mutation}) \;=\; u_N \nu_0 (N-k+1)\frac{w^N(k-1)}{w^N(k)}$,
- $\widehat{q}_{K^N}(k,k-1;\text{ due to reproduction}) \;=\; \left[\lambda^N_{k-1} - u_N \nu_0 (N-k+1)\right]\frac{w^N(k-1)}{w^N(k)}$.

Backward in time rates of (K^N, R^N)

This decomposition of the rates of the process K^N enables us to get the backward in time generator of the process (K^N, R^N), where $K^N \in \{0,1,...,N\}$ is again the number of particles of type 0 and $R^N \in \{0,1\}$ the type of the immortal line. Let $q_{(K^N,R^N)}$ and $\widehat{q}_{(K^N,R^N)}$ be its forward and backward in time transition rates, respectively.

Backward in time, each individual cannot just die. It has an ancestral line that continues through times. But it may mutate or coalesce with individuals of the same type. As each ancestral line

[20] Remember that the index '$\widehat{\ }$' indicates the direction *backward in time*.

eventually coalesces into the line of the common ancestor, backward in time, the dynamics of the immortal line coincides with the dynamics of a randomly chosen line.

The change of the type on the immortal line from 0 to 1 is coupled with the decrease of K^N by 1 due to a mutation. Conditioned on the transition of K^N from k to $k-1$ due to a mutation, the probability that the real line was affected by this mutation (from 0 to 1) is just $\frac{1}{k}$ (a uniformly chosen individual of type 0). Thus, for this transition, we get the rate

$$\begin{aligned}
&\widehat{q}_{(K^N,R^N)}((k,0),(k-1,1)) \\
&\quad = \mathbb{P}(\text{real line mutates} \mid \text{one of the k lines mutates}) \cdot \widehat{q}_{K^N}(k,k-1;\text{due to mutation}) \\
&\quad = \frac{1}{k} u_N v_0 (N-k+1) \frac{w^N(k-1)}{w^N(k)}.
\end{aligned}$$

Then, the probability that the immortal line does not change its type from 0 to 1 at a transition of K^N from k to $k-1$due to a mutation is just $k-1/k$. Therefore we get the rate

$$\begin{aligned}
&\widehat{q}_{(K^N,R^N)}((k,0),(k-1,0)) \\
&\quad = \mathbb{P}(\text{real line does not mutate} \mid \text{one of the k lines mutates}) \cdot \widehat{q}_{K^N}(k,k-1;\text{due to mutation}) \\
&\qquad + \widehat{q}_{K^N}(k,k-1;\text{due to reproduction}) \\
&\quad = \frac{k-1}{k} u_N v_0 (N-k+1) \frac{w^N(k-1)}{w^N(k)} + \left[\lambda^N_{k-1} - u_N v_0 (N-k+1)\right] \frac{w^N(k-1)}{w^N(k)} \\
&\quad = \left[\lambda^N_{k-1} - \frac{1}{k} u_N v_0 (N-k+1)\right] \frac{w^N(k-1)}{w^N(k)}.
\end{aligned}$$

The probability that the immortal line mutates from 0 to 1 given there is a transition from k to $k+1$, is just zero. Thus, we have $\widehat{q}_{(K^N,R^N)}((k,0),(k+1,1)) = 0$ and

$$\begin{aligned}
&\widehat{q}_{(K^N,R^N)}((k,0),(k+1,0)) \\
&\quad = \widehat{q}_{K^N}(k,k+1;\text{due to mutation}) + \widehat{q}_{K^N}(k,k+1;\text{due to reproduction}) \\
&\quad = \mu_{k+1} \frac{w^N(k+1)}{w^N(k)}.
\end{aligned}$$

Similar arguments lead to the transition rates out of state $(k,1)$ and we get the following lemma.

Lemma 7.1 *The (backward in time) generator $\widehat{G}_{(K^N,R^N)}$ of the process (K^N,R^N) is given by*

$$\begin{aligned}
\widehat{G}_{(K^N,R^N)}g(k,0) = \; & \mu^N_{k+1}\frac{w^N(k+1)}{w^N(k)}\left[g(k+1,0)-g(k,0)\right] \\
& + \frac{1}{k}u_N v_0(N-k+1)\frac{w^N(k-1)}{w^N(k)}\left[g(k-1,1)-g(k,0)\right] \\
& + \left[\lambda^N_{k-1}-\frac{1}{k}u_N v_0(N-k+1)\right]\frac{w^N(k-1)}{w^N(k)}\left[g(k-1,0)-g(k,0)\right], \\
\widehat{G}_{(K^N,R^N)}g(k,1) = \; & \frac{1}{N-k}u_N v_1(k+1)\frac{w^N(k+1)}{w^N(k)}\left[g(k+1,0)-g(k,1)\right] \\
& + \left[\mu^N_{k+1}-\frac{1}{N-k}u_N v_1(k+1)\right]\frac{w^N(k+1)}{w^N(k)}\left[g(k+1,1)-g(k,1)\right] \\
& + \lambda^N_{k-1}\frac{w^N(k-1)}{w^N(k)}\left[g(k-1,1)-g(k,1)\right],
\end{aligned} \tag{7.5}$$

with the test function $g:\mathbb{N}_0\times\{0,1\}\to\mathbb{R}$, and $k\in\{0,1,\dots,N\}$.

The stationary distribution of the process (K^N,R^N)

For $u_N, v_0, v_1, s_N > 0$, the process (K^N,R^N) is irreducible and recurrent and therefore has a stationary distribution π^N. Let us assume that the process is at stationarity at time 0. Then $\pi^N(k,r) = \mathbb{P}(K^N_0 = k, R^N_0 = r)$.
In this section we derive a characterisation of π^N from Lemma 7.1. This way, we regain results by Kluth, Hustedt, and Baake [KHB13] (but with a different approach).

Motivated by the structure of the stationary distribution π of (X,R) ((2.38) in Section 2.2.4), we also use a product ansatz for the stationary distribution in the discrete setting,

$$\pi^N(k,r) = w^N(k)\left[\mathbf{1}_{\{r=0\}}h^N_k + \mathbf{1}_{\{r=1\}}(1-h^N_k)\right]. \tag{7.6}$$

From this equation it is clear that h^N_k can be interpreted as the conditioned probability that the immortal line is of type 0 at time 0, given the number of type 0 individuals at time 0 is k.
Therefore, the natural boundary conditions are

$$h^N_0 = 0, \quad h^N_N = 1, \tag{7.7}$$

and (analogue to the continuous case) the probabilities h^N_k are quantified by a recursion.

Lemma 7.2 *The stationary distribution π^N of the process (K^N,R^N) is given by* (7.6)*, where $(h^N_k)_{k=0,1,\dots N}$ is determined by the recursion*

$$\begin{aligned}
& h^N_k\left(\lambda^N_k+\mu^N_k\right) \\
& \quad = h^N_{k+1}\lambda^N_k + h^N_{k-1}\mu^N_k + \left(1-h^N_{k-1}\right)u_N v_1\frac{k}{N-k+1} - h^N_{k+1}u_N v_0\frac{N-k}{k+1},
\end{aligned} \tag{7.8}$$

with boundary conditions $h^N_0 = 0$, $h^N_N = 1$.

Note that the statement of this lemma was already formulated and proven (via a different approach) by Kluth, Hustedt and Baake: (7.2) is exactly [KHB13, (21)].

Proof of Lemma 7.2. Since π^N is the equilibrium distribution for the generator matrix $\widehat{G}_{(K^N,R^N)}$ with the rates $\widehat{q}_{(K^N,R^N)}\big((\tilde{k},\tilde{r}),(k,r)\big)$, $(\tilde{k},\tilde{r}),(k,r) \in \{0,1,\ldots,N\}\times\{0,1\}$, the following equation (together with the boundary conditions) determines π^N and hence also h^N,

$$\sum_{(\tilde{k},\tilde{r})} \pi^N(\tilde{k},\tilde{r})\widehat{q}_{(K^N,R^N)}\big((\tilde{k},\tilde{r}),(k,r)\big) = 0 \qquad \forall\, k \in \{0,1,\ldots,N\},\ r \in \{0,1\}. \tag{7.9}$$

Plugging in the rates and dividing by w^N_k, we have

$$\begin{aligned} 0 = {} & h^N_{k+1}\left(\lambda^N_k - u_N v_0 \frac{N-k}{k+1}\right) + h^N_{k-1}\mu^N_k \\ & + \left(1-h^N_{k-1}\right) u_N v_1 \frac{k}{N-k+1} - \frac{h^N_k}{w^N(k)}\left(\lambda^N_{k-1} w^N(k-1) + \mu^N_{k+1} w^N(k+1)\right). \end{aligned} \tag{7.10}$$

By (7.3), (7.10) is equivalent to (7.8) and the proof is complete. □

7.2.2 The forward in time generator of (K^N,R^N)

Forward in time rates of (K^N,R^N)

By using the appropriate analogue of (7.3) for π^N instead of w^N, we can now calculate the forward in time rates $q_{(K^N,R^N)}\big((\cdot,\cdot),(\cdot,\cdot)\big)$ of the process $(K^N_r,R^N_r)_{r\in\mathbb{R}}$,

$$q_{(K^N,R^N)}\big((k,r),(\tilde{k},\tilde{r})\big) = \widehat{q}_{(K^N,R^N)}\big((\tilde{k},\tilde{r}),(k,r)\big)\frac{\pi^N(\tilde{k},\tilde{r})}{\pi^N(k,r)}, \tag{7.11}$$

$k \neq \tilde{k}$, $k,\tilde{k} \in \{0,1,\ldots N\}$, $r,\tilde{r} \in \{0,1\}$. We get the following lemma.

Lemma 7.3 *The (forward in time) generator $G_{(K^N,R^N)}$ of the process (K^N,R^N) is given by*

$$\begin{aligned} G_{(K^N,R^N)}g(k,0) = {} & \left[\lambda^N_k - \frac{1}{k+1}u_N v_0(N-k)\right]\frac{h^N_{k+1}}{h^N_k}\left[g(k+1,0)-g(k,0)\right] \\ & + \frac{1}{N-k+1}u_N v_1 k \frac{1-h^N_{k-1}}{h^N_k}\left[g(k-1,1)-g(k,0)\right] \\ & + \mu^N_k \frac{h^N_{k-1}}{h^N_k}\left[g(k-1,0)-g(k,0)\right], \\ G_{(K^N,R^N)}g(k,1) = {} & \frac{1}{k+1}u_N v_0(N-k)\frac{h^N_{k+1}}{1-h^N_k}\left[g(k+1,0)-g(k,1)\right], \\ & + \lambda^N_k \frac{1-h^N_{k+1}}{1-h^N_k}\left[g(k+1,1)-g(k,1)\right] \\ & + \left[\mu^N_k - \frac{1}{N-k+1}u_N v_1 k\right]\frac{1-h^N_{k-1}}{1-h^N_k}\left[g(k-1,1)-g(k,1)\right], \end{aligned} \tag{7.12}$$

with the test function g, and $k \in \{0,1,\ldots,N\}$.

7.2.3 Diffusion limit

In this section, we let the population size $N \to \infty$ and calculate the diffusion limit of the backward and forward in time generators of (K^N, R^N) to compare them with Taylor's results. The discrete process $(K^N/N, R)$ then converges to the continuous process (X, R) under the Assumptions (7.4) and with $h = \lim_{N\to\infty} h^N$ (for the proper convergence of the probability h^N in the discrete setting to the probability h in the continuous setting see [KHB13]).

We gain the dynamics of the diffusion limit of the process $(K^N/N, R^N)$ by letting $N \to \infty$ for the (backward and forward in time) generators $\widehat{G}_{(K^N,R^N)}$, given by (7.5), and $G_{(K^N,R^N)}$, given by (7.12). We get the following results.

Theorem 7.4 *Under Assumptions* (7.4)*, the backward in time generator* $\widehat{G}_{(X,R)}$ *of the process* $(X_r, R_r)_{r\geq 0} := \lim_{N\to\infty}(K^N_{Nr}/N, R^N_{Nr})_{r\geq 0}$ *is determined by*

$$\begin{aligned} \widehat{G}_{(X,R)}g(x,0) &= G_X g(x,0) + \frac{1-x}{x}\theta\nu_0\left[g(x,1) - g(x,0)\right], \\ \widehat{G}_{(X,R)}g(x,1) &= G_X g(x,1) + \frac{x}{1-x}\theta\nu_1\left[g(x,0) - g(x,1)\right], \end{aligned} \tag{7.13}$$

for all $g \in \mathscr{C}^2[0,1]\times\{0,1\}$.

Theorem 7.5 *Under Assumptions* (7.4)*, the generator* $G_{(X,R)}$ *of the process* $(X_t, R_t)_{t\geq 0} := \lim_{N\to\infty}(K^N_{Nt}/N, R^N_{Nt})_{t\geq 0}$ *forward in time is determined by*

$$\begin{aligned} G_{(X,R)}g(x,0) &= G_X g(x,0) + x(1-x)\frac{h'(x)}{h(x)}g'(x,0) + \frac{x(1-h(x))}{(1-x)h(x)}\theta\nu_1\left[g(x,1) - g(x,0)\right], \\ G_{(X,R)}g(x,1) &= G_X g(x,1) - x(1-x)\frac{h'(x)}{1-h(x)}g'(x,1) + \frac{(1-x)h(x)}{x(1-h(x))}\theta\nu_0\left[g(x,0) - g(x,1)\right], \end{aligned} \tag{7.14}$$

for all $g \in \mathscr{C}^2[0,1]\times\{0,1\}$.

Note that Theorem 7.4 is exactly a result by Taylor [Tay07, eq. (4)]. Theorem 7.5 can again be found in Taylor's work: It is equivalent to [Tay07, eq. (15)].

The proofs of Theorems 7.4 and 7.5 are given in Sections 7.4.2 and 7.4.3 in the appendix.

7.3 Perspective: The process (K^N, R^N, V^N)

In the previous part of this chapter we investigated the process (K^N, R^N) of the number K^N of individuals of type 0 in a population with total size N together with the type R^N of the immortal line. By passing to the limit $N \to \infty$ we recapitulated the forward and backward in time generator of the continuous process (X, R). Our results in the discrete model are similar to those in [KHB13] and the limiting continuous generators are identical to those derived by Taylor [Tay07].

In this last part of this chapter we want to give an outlook onto some ideas of extending the double process (K^N, R^N) by adding Fearnhead's number of virtual lines V^N [Fea02] (but in the discrete setting). We do not give rigorous proofs but heuristic ideas.

To derive the (backward in time) transition rates of the process (K^N, R^N, V^N), we have to take into account that virtual lines are always of type 1. The transition dynamics of the virtual lines together with the immortal line are (due to Fearnhead [Fea02]) determined by mutations on the real line, disappearance of virtual lines due to coalescence or mutation, and appearance of virtual lines due to selective reproduction. Therefore, we need to decompose the rates of K^N not only into 'mutation' and 'selection' (as done in Section 7.2.1), but into the *three* classes 'mutation', 'neutral reproduction', and 'selective reproduction'.

Forward in time the decomposed rates are

- $q_{K^N}(k,k+1;\text{ due to mutation}) = u_N \nu_0 (N-k)$,
- $q_{K^N}(k,k+1;\text{ due to neutral reproduction}) = \frac{k(N-k)}{2N}$,
- $q_{K^N}(k,k+1;\text{ due to selective reproduction}) = s_N k \frac{(N-k)}{N}$,
- $q_{K^N}(k,k-1;\text{ due to mutation}) = u_N \nu_1 k$,
- $q_{K^N}(k,k-1;\text{ due to neutral reproduction}) = \frac{k(N-k)}{2N}$,

and backward in time we have

- $\widehat{q}_{K^N}(k,k-1;\text{ due to mutation}) = u_N \nu_0 (N-k+1) \frac{w^N(k-1)}{w^N(k)}$,
- $\widehat{q}_{K^N}(k,k-1;\text{ due to neutral reproduction}) = \frac{(k-1)(N-k+1)}{2N} \frac{w^N(k-1)}{w^N(k)}$,
- $\widehat{q}_{K^N}(k,k-1;\text{ due to selective reproduction}) = s_N \frac{(k-1)(N-k+1)}{N} \frac{w^N(k-1)}{w^N(k)}$,
- $\widehat{q}_{K^N}(k,k+1;\text{ due to mutation}) = u_N \nu_1 (k+1) \frac{w^N(k+1)}{w^N(k)}$,
- $\widehat{q}_{K^N}(k,k+1;\text{ due to neutral reproduction}) = \frac{(k+1)(N-k-1)}{2N} \frac{w^N(k+1)}{w^N(k)}$.

Note that the number of Fearnhead's virtual lines may increase or decrease in the discrete model while K^N does not change at all. This can be understood in terms of the Moran model:
The process of mutations to type 1 is usually modelled by a Poisson point process of crosses at rate $u_N \nu_1$ per line. At each cross, the affected line changes its type to type 1 if it was of type 0. If it was already of type 1, nothing happens and we have a silent mutation. But backward in time a virtual line (of type 1) is removed if there is a cross (even if this cross indicates a silent mutation). Thus, backward in time, silent mutations from 1 to 1 cause a transition from v to $v-1$ virtual lines but no transition of the number K^N of particles of type 0.
Similarly, there can be a coalescence or branching of virtual particles which leaves K^N invariant. This is represented in the Moran model e.g. by neutral arrows between two individuals of type 1.
Therefore, to find the corresponding rates of the process (K^N, R^N, V^N) we have to take into account the following (forward in time) transitions as well.

- $q_{K^N}(k,k;\text{ mutation from 1 to 1}) \;=\; u_N v_1(N-k)$,
- $q_{K^N}(k,k;\text{ neutral reproduction between 1 and 1}) \;=\; \frac{(N-k)(N-k-1)}{2N}$,
- $q_{K^N}(k,k;\text{ selective reproduction between 1 and 1}) \;=\; s_N(N-k)\frac{(N-k-1)}{N}$.

Since the proportion of type 0 individuals does not change, there is no factor needed to time reverse the rates ($w^N(k)/w^N(k) = 1$). The corresponding rates backward in time therefore are

- $\widehat{q}_{K^N}(k,k;\text{ mutation from 1 to 1}) \;=\; u_N v_1(N-k)$
- $\widehat{q}_{K^N}(k,k;\text{ reproduction between 1 and 1}) \;=\; \frac{(N-k)(N-k-1)}{2N}$
- $\widehat{q}_{K^N}(k,k;\text{ selective reproduction between 1 and 1}) \;=\; s_N(N-k)\frac{(N-k-1)}{N}$.

7.3.1 Ideas on the backward and forward in time rates of (K^N, R^N, V^N)

We denote backward and forward in time rates of (K^N, R^N, V^N) by $\widehat{q}_{(K^N,R^N,V^N)}\big((\cdot,\cdot),(\cdot,\cdot)\big)$ and $q_{(K^N,V^N,R^N)}\big((\cdot,\cdot),(\cdot,\cdot)\big)$, respectively.
The ideas leading to the backward in time rates $\widehat{q}_{(K^N,R^N,V^N)}$ are similar to those in Section 7.2.1: Simultaneously with the change of the number K^N of individuals of type 0, the type R^N of the immortal line or the number V^N of virtual lines may change with a certain probability. In addition, the number of virtual lines may change while K^N stays constant.

The rate of the transition of the immortal line from type 0 to type 1 together with a decrease of K^N by one is already part of the generator $\widehat{G}_{(K^N,R^N)}$ and therefore given by $\widehat{q}_{(K^N,R^N,V^N)}((k,0,v),(k-1,1,v)) = (1/k)u_N v_0(N-k+1)w^N(k-1)/w^N(k)$.
A transition of the number of virtual lines from v to $v+1$ may happen together with a decrease of K^N from k to $k-1$ due to a selective reproduction event at which the immortal line (of type 0) branches into an additional line of type 1 (and this additional line replaces another line of type 0 in the population). The rate for this event is

$$\begin{aligned}
&\widehat{q}_{(K^N,R^N,V^N)}((k,0,v),(k-1,0,v+1))\\
&\quad= \mathbb{P}(\text{immortal line branches} \mid \text{selective reproduction with transition } k\to k-1)\\
&\qquad\cdot \widehat{q}_{K^N}(k,k-1;\text{due to selective reproduction})\\
&\quad= \frac{1}{k}s_N\frac{(k-1)(N-k+1)}{N}\frac{w^N(k-1)}{w^N(k)}.
\end{aligned}$$

But a transition of the number of virtual lines may also happen while K^N stays constant. This happens at a selective reproduction event between two lines of type 1 at which the line 'hit by the tip of the selective arrow' is a virtual line and the line 'at the tail' is not virtual,

$$\begin{aligned}
&\widehat{q}_{(K^N,R^N,V^N)}((k,0,v),(k,0,v+1))\\
&\quad= \mathbb{P}(\text{virtual line is at the tip of the selective arrow and line at tail is not virtual} \mid\\
&\qquad\text{selective reproduction with transition } k\to k)\\
&\qquad\cdot \widehat{q}_{K^N}(k,k;\text{selective reproduction between 1 and 1})
\end{aligned}$$

$$= \frac{v(N-k-v)}{(N-k)(N-k-1)} s_N \frac{(N-k)(N-k-1)}{N} = \frac{v(N-k-v)}{N} s_N.$$

A transition of the number of virtual lines from v to $v-1$ may happen together with an increase of K^N from k to $k+1$ due to neutral reproduction between two virtual lines or mutation of a virtual line,

$$\begin{aligned}
&\widehat{q}_{(K^N,R^N,V^N)}((k,0,v),(k+1,0,v-1)) \\
&\quad = \mathbb{P}(\text{reproduction between two virtual lines} \mid \text{neutral reproduction with transition } k \to k+1) \\
&\qquad \cdot \widehat{q}_{K^N}(k,k+1;\text{due to neutral reproduction}) \\
&\quad + \mathbb{P}(\text{virtual line mutates} \mid \text{mutation with transition } k \to k+1) \\
&\qquad \cdot \widehat{q}_{K^N}(k,k+1;\text{due to mutation}) \\
&\quad = \frac{v(v-1)}{(N-k)(N-k-1)} \frac{(k+1)(N-k-1)}{2N} \frac{w^N(k+1)}{w^N(k)} + \frac{v}{N-k} u_N \nu_1 (k+1) \frac{w^N(k+1)}{w^N(k)} \\
&\quad = \left[\frac{v(v-1)(k+1)}{2N(N-k)} + \frac{v(k+1)}{N-k} u_N \nu_1 \right] \frac{w^N(k+1)}{w^N(k)}.
\end{aligned}$$

A transition of the number of virtual lines from v to $v-1$ can also happen while K^N stays constant due to a neutral reproduction between two virtual individuals of type 1,

$$\begin{aligned}
&\widehat{q}_{(K^N,R^N,V^N)}((k,0,v),(k,0,v-1)) \\
&\quad = \mathbb{P}(\text{participating lines are both virtual} \mid \text{neutral reproduction between individuals of type 1}) \\
&\qquad \cdot \widehat{q}_{K^N}(k,k;\text{neutral reproduction between 1 and 1}) \\
&\quad = \frac{v(v-1)}{(N-k)(N-k-1)} \frac{(N-k)(N-k-1)}{2N} = \frac{v(v-1)}{2N}.
\end{aligned}$$

Then the remaining rate for the transition of K^N from k to $k-1$ while R^N and V^N stay constant is given by

$$\begin{aligned}
&\widehat{q}_{(K^N,R^N,V^N)}((k,0,v),(k-1,0,v)) \\
&\quad = \widehat{q}_{K^N}(k,k-1) - \widehat{q}_{(K^N,R^N,V^N)}((k,0,v),(k-1,1,v)) - \widehat{q}_{(K^N,R^N,V^N)}((k,0,v),(k-1,0,v+1)) \\
&\quad = \left[\lambda_{k-1} - (1/k) u_N \nu_0 (N-k+1) - s_N \frac{(k-1)(N-k+1)}{N} \right] \frac{w^N(k-1)}{w^N(k)}.
\end{aligned}$$

The remaining rate for the transition of K^N from k to $k+1$ while R^N and V^N stay constant is

$$\begin{aligned}
&\widehat{q}_{(K^N,R^N,V^N)}((k,0,v),(k+1,0,v)) \\
&\quad = \widehat{q}_{K^N}(k,k+1) - \widehat{q}_{(K^N,R^N,V^N)}((k,0,v),(k+1,0,v-1)) \\
&\quad = \left[\mu_{k+1} - \frac{v(v-1)(k+1)}{2N(N-k)} - \frac{v(k+1)}{N-k} u_N \nu_1 \right] \frac{w^N(k+1)}{w^N(k)}.
\end{aligned}$$

Similar arguments give the transition rates of (K^N,R^N,V^N) with initial state $(k,1,v)$.

Let φ^N be the stationary distribution of (K^N,R^N,V^N). Then, comparing with the shape of the stationary distribution φ of the continuous process (X,R,V), (6.12), and with the stationary distribution of

(K^N, R^N), (7.6), we suppose that φ^N should be given by

$$\varphi^N(k,r,v) = w^N(k) \cdot \phi_k^N(r,v), \tag{7.15}$$

where the coefficients $\phi_k^N(r,v)$ should be somehow comparable to the summands in [KHB13, (40)].

Then one may determine the forward in time generator of the process (K^N, R^N, V^N): The rates $q_{(K^N,V^N,R^N)}$ of the forward in time process (K^N, R^N, V^N) can be calculated via the appropriate analogue of (7.11),

$$q_{(K^N,R^N,V^N)}\big((k,r,v),(\tilde{k},\tilde{r},\tilde{v})\big) = \widehat{q}_{(K^N,R^N,V^N)}\big((\tilde{k},\tilde{r},\tilde{v}),(k,r,v)\big)\frac{\varphi^N(\tilde{k},\tilde{r},\tilde{v})}{\varphi^N(k,r,v)}, \tag{7.16}$$

$k \neq \tilde{k}$, $k, \tilde{k} \in \{0,1,\dots N\}$, $r, \tilde{r} \in \{0,1\}$, $v, \tilde{v} \in \{0,1,\dots N-k\}$.

Now, assuming that the backward and forward in time generators of the discrete process (K^N, R^N, V^N) are given, one may proceed by taking the diffusion limit $N \to \infty$. This way, one should arrive at the backward and forward in time generators $\widehat{G}_{(X,R,V)}$ and $G_{(X,R,V)}$ of the process (X,R,V).
Therefore, the discrete approach should yield an additional proof for Lemma 6.1 and Theorem 6.4.

7.4 Appendix

7.4.1 Convergence of the discrete type frequency process K^N to a Wright-Fisher diffusion

Theorem 7.6 *The generator G_X of the process $(X_t)_{t\geq 0}$ (backward and forward) in time, defined as the limiting object of the generator of the process $(X^N_{Nt})_{t\geq 0}$ with rates* (7.1) *is determined by*

$$G_X g(x) \;=\; \frac{1}{2}x(1-x)g''(x) + [\theta v_0(1-x) - \theta v_1 x + x(1-x)\sigma]\, g'(x)\,, \tag{7.17}$$

for all $g \in \mathscr{C}^2[0,1]$.

Proof. Due to reversibility it is true that the forward and backward in time generators of K^N (and therefore also X^N) agree (it is $\hat{q}_{K^N}(k,k+1) = q_{K^N}(k,k+1)$, $\hat{q}_{K^N}(k,k-1) = q_{K^N}(k,k-1)$). Thus, we have to take into account only one direction in time, e.g. the calculation of the generator forward in time. Let $g \in \mathscr{C}^2[0,1]$. Then

$$\begin{aligned} G_X g(x) &= \lim_{N\to\infty} N \sum_{\ell} q_{K^N}(k,\ell)\left[g\left(\frac{\ell}{N}\right) - g\left(\frac{k}{N}\right)\right] \\ &\qquad \Big(\downarrow \text{ plugging in the rates } q_{K^N}(k,k+1) = \lambda_k^N \text{ and } q_{K^N}(k,k-1) = \mu_k^N \\ &\qquad\quad \text{of the birth and death process } K^N\Big) \\ &= \lim_{N\to\infty} N\Bigg\{\lambda^N_{Nx_N}\left[g\left(x_N + \frac{1}{N}, 0\right) - g(x_N,0)\right] \\ &\qquad\qquad + \mu^N_{Nx_N}\left[g\left(x_N - \frac{1}{N}, 0\right) - g(x_N,0)\right]\Bigg\} \end{aligned}$$

$$\Big(\downarrow \text{plugging in the definitions (7.1) of } \lambda^N \text{ and } \mu^N\Big)$$

$$\begin{aligned}
&= \lim_{N\to\infty} N\Bigg\{ \left[Nx_N \frac{N(1-x_N)}{N}\left(\frac{1}{2}+s_N\right) + u_N v_0 N(1-x_N)\right]\left[g\left(x_N+\frac{1}{N},0\right)-g\left(x_N,0\right)\right] \\
&\qquad + \left[\frac{1}{2}N(1-x_N)\frac{Nx_N}{N} + u_N v_1 N x_N\right]\left[g\left(x_N-\frac{1}{N},0\right)-g\left(x_N,0\right)\right]\Bigg\}
\end{aligned}$$

$$\Big(\downarrow \text{ordering terms}\Big)$$

$$\begin{aligned}
&= \lim_{N\to\infty}\Bigg\{ \frac{1}{2}x_N(1-x_N)\cdot N^2\left[g\left(x_N+\frac{1}{N},0\right)-2g\left(x_N,0\right)+g\left(x_N-\frac{1}{N},0\right)\right] \\
&\qquad + \left[x_N(1-x_N)Ns_N + Nu_N v_0(1-x_N)\right]\cdot N\left[g\left(x_N+\frac{1}{N},0\right)-g\left(x_N,0\right)\right] \\
&\qquad + Nu_N v_1 x_N\cdot N\left[g\left(x_N-\frac{1}{N},0\right)-g\left(x_N,0\right)\right]\Bigg\}
\end{aligned}$$

$$\Big(\downarrow \text{passing to the limit}\Big)$$

$$= \frac{1}{2}x(1-x)g''(x) + [x(1-x)\sigma + \theta v_0(1-x)]g'(x) - \theta v_1 x g'(x),$$

which is (7.17). □

7.4.2 Proof of Theorem 7.4

Theorem 7.4 *Under Assumptions* (7.4)*, the backward in time generator* $\widehat{G}_{(X,R)}$ *of the process* $(X_r,R_r)_{r\geq 0} := \lim_{N\to\infty}(K^N_{Nr}/N, R^N_{Nr})_{r\geq 0}$ *is determined by*

$$\widehat{G}_{(X,R)}g(x,0) \;=\; G_X g(x,0) + \frac{1-x}{x}\theta v_0\left[g(x,1)-g(x,0)\right], \tag{7.18}$$

$$\widehat{G}_{(X,R)}g(x,1) \;=\; G_X g(x,1) + \frac{x}{1-x}\theta v_1\left[g(x,0)-g(x,1)\right], \tag{7.19}$$

for all $g\in\mathscr{C}^2[0,1]\times\{0,1\}$.

Proof. The process $(K^N_{Nr}/N, R^N_{Nr})$ takes values in $\{0,1/N,2/N,\ldots,1\}\times\{0,1\}$, which is a subset of $[0,1]\times\{0,1\}$. Since $[0,1]\times\{0,1\}$ is compact, the convergence of the backward in time generator of the process $(K^N_{Nr}/N, R^N_{Nr})$ to some limit object $\widehat{G}_{(X,R)}$ (together with the convergence of the initial values that holds due to Assumptions (7.4)) implies the convergence of the backward in time process $(K^N_{Nr}/N, R^N_{Nr})_{r\geq 0}$ to some backward in time process $(X_r,R_r)_{r\geq 0}$ (cf. [Ker13, Satz 2.8]).

To calculate the generator $\widehat{G}_{(X,R)}$ of the continuous process (X,R) backward in time we use the rates of Section 7.2.1 and pass to the limit.

$$\begin{aligned}
&\widehat{G}_{(X,R)}g(x,0) \\
&\quad = \lim_{N\to\infty} N\sum_{(\ell,s)}\widehat{q}_{(X,R)}\big((k,0),(\ell,s)\big)\left[g\left(\frac{\ell}{N},s\right)-g\left(\frac{k}{N},0\right)\right]
\end{aligned}$$

$$\Big(\downarrow \text{plugging in the rates } \widehat{q}_{(X,R)}(\cdot,\cdot)\ (7.5)\Big)$$

$$\begin{aligned}
&= \lim_{N\to\infty} N\Bigg\{ \mu^N_{Nx_N+1}\frac{w^N(Nx_N+1)}{w^N(Nx_N)}\left[g\left(x_N+\frac{1}{N},0\right)-g(x_N,0)\right] \\
&\qquad + \frac{1}{Nx_N}u_N v_0(N-Nx_N+1)\frac{w^N(Nx_N-1)}{w^N(Nx_N)}\left[g\left(x_N-\frac{1}{N},1\right)-g(x_N,0)\right] \\
&\qquad + \left[\lambda^N_{Nx_N-1}-\frac{1}{Nx_N}u_N v_0(N-Nx_N+1)\right]\frac{w^N(Nx_N-1)}{w^N(Nx_N)} \\
&\qquad\quad \left[g\left(x_N-\frac{1}{N},0\right)-g(x_N,0)\right]\Bigg\}
\end{aligned}$$

$$\Big(\downarrow \text{applying } (7.3)\Big)$$

$$\begin{aligned}
&= \lim_{N\to\infty} N\Bigg\{ \mu^N_{Nx_N+1}\frac{\lambda^N_{Nx_N}}{\mu^N_{Nx_N+1}}\left[g\left(x_N+\frac{1}{N},0\right)-g(x_N,0)\right] \\
&\qquad + \frac{1}{Nx_N}u_N v_0(N-Nx_N+1)\frac{\mu^N_{Nx_N}}{\lambda^N_{Nx_N-1}}\left[g\left(x_N-\frac{1}{N},1\right)-g(x_N,0)\right] \\
&\qquad + \left[\lambda^N_{Nx_N-1}-\frac{1}{Nx_N}u_N v_0(N-Nx_N+1)\right]\frac{\mu^N_{Nx_N}}{\lambda^N_{Nx_N-1}} \\
&\qquad\quad \left[g\left(x_N-\frac{1}{N},0\right)-g(x_N,0)\right]\Bigg\}
\end{aligned}$$

$$\Big(\downarrow \text{transforming expressions}\Big)$$

$$\begin{aligned}
&= \lim_{N\to\infty} \Bigg\{ N\lambda^N_{Nx_N}\left[g\left(x_N+\frac{1}{N},0\right)-g(x_N,0)\right] \\
&\qquad + \frac{1-x_N+\frac{1}{N}}{x_N}Nu_N v_0\frac{\mu^N_{Nx_N}}{\lambda^N_{Nx_N-1}}\left[g\left(x_N-\frac{1}{N},1\right)-g(x_N,0)\right] \\
&\qquad + N\mu^N_{Nx_N}\left[g\left(x_N-\frac{1}{N},0\right)-g(x_N,0)\right]\Bigg\} \\
&\qquad - \frac{1-x_N+\frac{1}{N}}{x_N}Nu_N v_0\frac{\mu^N_{Nx_N}}{\lambda^N_{Nx_N-1}}\left[g\left(x_N-\frac{1}{N},0\right)-g(x_N,0)\right]\Bigg\}
\end{aligned}$$

$$\Big(\downarrow \text{rearranging terms}\Big)$$

$$\begin{aligned}
&= \lim_{N\to\infty} \Bigg\{ N\lambda^N_{Nx_N}\left[g\left(x_N+\frac{1}{N},0\right)-g(x_N,0)\right] \\
&\qquad + N\mu^N_{Nx_N}\left[g\left(x_N-\frac{1}{N},0\right)-g(x_N,0)\right]\Bigg\}
\end{aligned}$$

$$
\begin{aligned}
&+ \frac{1-x_N+\frac{1}{N}}{x_N} N u_N v_0 \frac{\mu^N_{Nx_N}}{\lambda^N_{Nx_N-1}} \left[g\left(x_N - \frac{1}{N}, 1\right) - g(x_N, 0) \right] \\
&- \frac{1-x_N+\frac{1}{N}}{x_N} N u_N v_0 \frac{\mu^N_{Nx_N}}{\lambda^N_{Nx_N-1}} \left[g\left(x_N - \frac{1}{N}, 0\right) - g(x_N, 0) \right] \Bigg\}
\end{aligned}
$$

$\left(\downarrow \text{passing to the limit}\right)$

$$
\begin{aligned}
&= G_X g(x,0) + \frac{1-x}{x} \theta v_0 \cdot 1 \cdot [g(x,1) - g(x,0)] - \frac{1-x}{x} \theta v_0 \cdot 1 \cdot 0 \\
&= G_X g(x,0) + \frac{1-x}{x} \theta v_0 [g(x,1) - g(x,0)],
\end{aligned}
$$

thus (7.18) is proven.

$\widehat{G}_{(X,R)} g(x,1)$

$$
\begin{aligned}
&= \lim_{N\to\infty} N \sum_{(\ell,s)} \widehat{q}_{(X,R)}\big((k,1),(\ell,s)\big) \left[g\left(\frac{\ell}{N}, s\right) - g\left(\frac{k}{N}, 1\right) \right] \\
&\quad \left(\downarrow \text{plugging in the rates } \widehat{q}_{(X,R)}(\cdot,\cdot) \text{ (7.5)}\right) \\
&= \lim_{N\to\infty} N \Bigg\{ \frac{1}{N - Nx_N} u_N v_1 (Nx_N + 1) \frac{w^N(Nx_N+1)}{w^N(Nx_N)} \left[g\left(x_N + \frac{1}{N}, 0\right) - g(x_N, 1) \right] \\
&\qquad + \left[\mu^N_{Nx_N+1} - \frac{1}{N - Nx_N} u_N v_1 (Nx_N + 1) \right] \frac{w^N(Nx_N+1)}{w^N(Nx_N)} \\
&\qquad\quad \left[g\left(x_N + \frac{1}{N}, 1\right) - g(x_N, 1) \right] \\
&\qquad + \lambda^N_{Nx_N-1} \frac{w^N(Nx_N-1)}{w^N(Nx_N)} \left[g\left(x_N - \frac{1}{N}, 1\right) - g(x_N, 1) \right] \Bigg\} \\
&\quad \left(\downarrow \text{applying (7.3)}\right) \\
&= \lim_{N\to\infty} N \Bigg\{ \frac{1}{N - Nx_N} u_N v_1 (Nx_N + 1) \frac{\lambda^N_{Nx_N}}{\mu^N_{Nx_N+1}} \left[g\left(x_N + \frac{1}{N}, 0\right) - g(x_N, 1) \right] \\
&\qquad + \left[\mu^N_{Nx_N+1} - \frac{1}{N - Nx_N} u_N v_1 (Nx_N + 1) \right] \frac{\lambda^N_{Nx_N}}{\mu^N_{Nx_N+1}} \\
&\qquad\quad \left[g\left(x_N + \frac{1}{N}, 1\right) - g(x_N, 1) \right] \\
&\qquad + \lambda^N_{Nx_N-1} \frac{\mu^N_{Nx_N}}{\lambda^N_{Nx_N-1}} \left[g\left(x_N - \frac{1}{N}, 1\right) - g(x_N, 1) \right] \Bigg\} \\
&\quad \left(\downarrow \text{rearranging terms}\right) \\
&= \lim_{N\to\infty} \Bigg\{ \frac{x_N + \frac{1}{N}}{1 - x_N} N u_N v_1 \frac{\lambda^N_{Nx_N}}{\mu^N_{Nx_N+1}} \left[g\left(x_N + \frac{1}{N}, 0\right) - g(x_N, 1) \right]
\end{aligned}
$$

$$
\begin{aligned}
&+\ \lambda^N_{Nx_N}\cdot N\left[g\left(x_N+\frac{1}{N},1\right)-g\left(x_N,1\right)\right]\\
&-\frac{x_N+\frac{1}{N}}{1-x_N}Nu_N\nu_1\frac{\lambda^N_{Nx_N}}{\mu^N_{Nx_N+1}}\cdot\left[g\left(x_N+\frac{1}{N},1\right)-g\left(x_N,1\right)\right]\\
&+\ \mu^N_{Nx_N}\cdot N\left[g\left(x_N-\frac{1}{N},1\right)-g\left(x_N,1\right)\right]\Bigg\}\\
&\left(\downarrow \text{passing to the limit}\right)\\
&=G_Xg\left(x,1\right)+\frac{x}{1-x}\theta\nu_1\left[g\left(x,0\right)-g\left(x,1\right)\right],
\end{aligned}
$$

so (7.19) is true as well and the proof is complete. □

7.4.3 Proof of Theorem 7.5

Similar to the case backward in time there is are analogue results for the generator $G_{(X,R)}$ forward in time.

Theorem 7.5 *Under Assumptions* (7.4)*, the generator* $G_{(X,R)}$ *of the process* $(X_t,R_t)_{t\geq 0}:=\lim_{N\to\infty}(K^N_{Nt}/N,R^N_{Nt})_{t\geq 0}$ *forward in time is determined by*

$$G_{(X,R)}g\left(x,0\right)\ =\ G_Xg\left(x,0\right)+x(1-x)\frac{h'(x)}{h(x)}g'(x,0)\ +\ \frac{x(1-h(x))}{(1-x)h(x)}\theta\nu_1\left[g\left(x,1\right)-g\left(x,0\right)\right],\quad(7.20)$$

$$G_{(X,R)}g\left(x,1\right)\ =\ G_Xg\left(x,1\right)-x(1-x)\frac{h'(x)}{1-h(x)}g'\left(x,1\right)\ +\ \frac{(1-x)h(x)}{x(1-h(x))}\theta\nu_0\left[g\left(x,0\right)-g\left(x,1\right)\right],\quad(7.21)$$

for all $g\in\mathscr{C}^2\left[0,1\right]\times\{0,1\}$.

Proof. The same compactness argument as in the beginning of the proof of Theorem 7.4 yields convergence of the forward in time process $(K^N_{Nt}/N,R^N_{Nt})_{t\geq 0}$ to some forward in time process $(X_t,R_t)_{t\geq 0}$.

To calculate the generator $G_{(X,R)}$ of the continuous process (X,R) forward in time, we use the rates of Section 7.2.2 and again pass to the diffusion limit.
We start with two small calculations which will determine some limits in the main calculations. The first identity is needed for the proof of (7.20),

$$
\begin{aligned}
&N^2\Bigg\{g\left(x_N+\frac{1}{N},0\right)h^N(Nx_N+1)-g\left(x_N,0\right)h^N(Nx_N+1)-g\left(x_N,0\right)h^N(Nx_N-1)\\
&\quad+g\left(x_N-\frac{1}{N},0\right)h^N(Nx_N-1)\Bigg\}\\
&\left(\downarrow \text{rearranging terms}\right)\\
&=N^2\Bigg\{h^N(Nx_N+1)\left[g\left(x_N+\frac{1}{N},0\right)-g\left(x_N,0\right)\right]-h^N(Nx_N-1)\left[g\left(x_N,0\right)-g\left(x_N-\frac{1}{N},0\right)\right]\Bigg\}\\
&\left(\downarrow \text{representation in terms of integrals}\right)
\end{aligned}
$$

$$= N^2\left\{ h^N(Nx_N+1)\int_{x_N}^{x_N+\frac{1}{N}} g'(\xi,0)\mathrm{d}\xi - h^N(Nx_N-1)\int_{x_N}^{x_N+\frac{1}{N}} g'(\xi-\frac{1}{N},0)\mathrm{d}\xi \right\}$$

$\left(\downarrow \text{rearranging terms}\right)$

$$= N^2\Bigg\{ \left[h^N(Nx_N+1)-h^N(Nx_N-1)\right]\int_{x_N}^{x_N+\frac{1}{N}} g'(\xi,0)\mathrm{d}\xi + h^N(Nx_N-1)\left[\int_{x_N}^{x_N+\frac{1}{N}} g'(\xi,0)\mathrm{d}\xi - \int_{x_N}^{x_N+\frac{1}{N}} g'(\xi-\frac{1}{N})\mathrm{d}\xi\right]\Bigg\}$$

$\left(\downarrow \text{representation in terms of integrals again}\right)$

$$= N^2\left\{ \int_{x_N-\frac{1}{N}}^{x_N+\frac{1}{N}} h'(\zeta)\mathrm{d}\zeta \int_{x_N}^{x_N+\frac{1}{N}} g'(\xi,0)\mathrm{d}\xi + h^N(Nx_N-1)\int_{x_N}^{x_N+\frac{1}{N}}\int_{\xi-\frac{1}{N}}^{\xi} g''(\eta,0)\mathrm{d}\eta\mathrm{d}\xi\right\}$$

$\left(\downarrow \text{passing to the limit}\right)$

$$\xrightarrow{N\to\infty} 2h'(x)g'(x,0)+h(x)g''(x,0). \tag{7.22}$$

The following calculation is a building block for the proof of (7.21),

$$N^2\Bigg\{ g\left(x_N+\frac{1}{N},1\right)\left(1-h^N(Nx_N+1)\right) - g\left(x_N,1\right)\left(1-h^N(Nx_N+1)\right) - g\left(x_N,1\right)\left(1-h^N(Nx_N-1)\right) + g\left(x_N-\frac{1}{N},1\right)\left(1-h^N(Nx_N-1)\right)\Bigg\}$$

$\left(\downarrow \text{rearranging terms}\right)$

$$= N^2\Bigg\{ \left(1-h^N(Nx_N+1)\right)\left[g\left(x_N+\frac{1}{N},1\right)-g\left(x_N,1\right)\right] - \left(1-h^N(Nx_N-1)\right)\left[g\left(x_N,1\right)-g\left(x_N-\frac{1}{N},1\right)\right]\Bigg\}$$

$\left(\downarrow \text{rearranging terms further}\right)$

$$= N^2\Bigg\{ \left[g\left(x_N+\frac{1}{N},1\right)-2g\left(x_N,1\right)+g\left(x_N-\frac{1}{N},1\right)\right] - h^N(Nx_N+1)\left[g\left(x_N+\frac{1}{N},1\right)-g\left(x_N,1\right)\right] + h^N(Nx_N-1)\left[g\left(x_N,1\right)-g\left(x_N-\frac{1}{N},1\right)\right]\Bigg\}$$

$\left(\downarrow \text{passing to the limit (compare previous calculation)}\right)$

$$\xrightarrow{N\to\infty} g''(x,1)-\left[2h'(x)g'(x,1)+h(x)g''(x,1)\right] \quad = \quad (1-h(x))g''(x,1)-2h'(x)g'(x,1)\,. \tag{7.23}$$

Now we prove the two assumptions of the theorem.
The generator operating on $g(x,0)$ is

$G_{(X,R)}g(x,0)$

$$= \lim_{N\to\infty} N \sum_{(\ell,s)} q_{(X,R)}\big((k,0),(\ell,s)\big)\left[g\left(\frac{\ell}{N},s\right) - g\left(\frac{k}{N},0\right)\right]$$

$\Big(\downarrow$ plugging in the rates (7.12)$\Big)$

$$\begin{aligned}
= \lim_{N\to\infty} N \Bigg\{ & \left[\lambda^N_{Nx_N} - \frac{1}{Nx_N+1} u_N v_0 (N - Nx_N)\right] \frac{h^N(Nx_N+1)}{h^N(Nx_N)} \\
& \qquad \left[g\left(x_N + \frac{1}{N}, 0\right) - g(x_N, 0)\right] \\
& + \frac{1}{N - Nx_N + 1} u_N v_1 N x_N \frac{1 - h^N(Nx_N - 1)}{h^N(Nx_N)} \left[g\left(x_N - \frac{1}{N}, 1\right) - g(x_N, 0)\right] \\
& + \mu^N_{Nx_N} \frac{h^N(Nx_N - 1)}{h^N(Nx_N)} \left[g\left(x_N - \frac{1}{N}, 0\right) - g(x_N, 0)\right] \Bigg\}
\end{aligned}$$

$\Big(\downarrow$ plugging in the definitions (7.1) of λ^N and $\mu^N\Big)$

$$\begin{aligned}
= \lim_{N\to\infty} N \Bigg\{ & \left[N x_N \frac{N - Nx_N}{N}\left(\frac{1}{2} + s_N\right) + u_N v_0 (N - Nx_N) - \frac{1}{Nx_N+1} u_N v_0 (N - Nx_N)\right] \\
& \qquad \frac{h^N(Nx_N+1)}{h^N(Nx_N)} \left[g\left(x_N + \frac{1}{N}, 0\right) - g(x_N, 0)\right] \\
& + \frac{1}{N - Nx_N + 1} u_N v_1 N x_N \frac{1 - h^N(Nx_N - 1)}{h^N(Nx_N)} \left[g\left(x_N - \frac{1}{N}, 1\right) - g(x_N, 0)\right] \\
& + \left[\frac{1}{2}(N - Nx_N)\frac{Nx_N}{N} + u_N v_1 N x_N\right] \frac{h^N(Nx_N - 1)}{h^N(Nx_N)} \left[g\left(x_N - \frac{1}{N}, 0\right) - g(x_N, 0)\right] \Bigg\}
\end{aligned}$$

$\Big(\downarrow$ rearranging terms and summarising lower order terms as $o(1)\Big)$

$$\begin{aligned}
= \lim_{N\to\infty} \Bigg\{ & N^2 \cdot \frac{1}{2} x_N (1 - x_N) \frac{1}{h^N(Nx_N)} \Bigg[g\left(x_N + \frac{1}{N}, 0\right) h^N(Nx_N + 1) \\
& \qquad - g(x_N, 0)\, h^N(Nx_N + 1) - g(x_N, 0)\, h^N(Nx_N - 1) \\
& \qquad + g\left(x_N - \frac{1}{N}, 0\right) h^N(Nx_N - 1)\Bigg] \\
& + \left[N \cdot x_N(1 - x_N) N s_N + N \cdot N u_N v_0 (1 - x_N)\right] \frac{h^N(Nx_N + 1)}{h^N(Nx_N)} \\
& \qquad \left[g\left(x_N + \frac{1}{N}, 0\right) - g(x_N, 0)\right] \\
& + \frac{x_N}{1 - x_N} N u_N v_1 x_N \frac{1 - h^N(Nx_N - 1)}{h^N(Nx_N)} \left[g\left(x_N - \frac{1}{N}, 1\right) - g(x_N, 0)\right] \\
& + N \cdot N u_N v_1 x_N \frac{h^N(Nx_N - 1)}{h^N(Nx_N)} \left[g\left(x_N - \frac{1}{N}, 0\right) - g(x_N, 0)\right] + o(1) \Bigg\}
\end{aligned}$$

$$
\begin{aligned}
&\left(\downarrow \text{passing to the limit by applying (7.22)}\right)\\
&= \frac{1}{2}x(1-x)g''(x,0) + x(1-x)\frac{h'(x)}{h(x)}g'(x,0) + [x(1-x)\sigma + \theta\nu_0(1-x)]\,g'(x,0)\\
&\qquad + \frac{x(1-h(x))}{(1-x)h(x)}\theta\nu_1\,[g(x,1)-g(x,0)] - \theta\nu_1 x g'(x,0)\\
&\left(\downarrow \text{rearranging terms}\right)\\
&= \frac{1}{2}x(1-x)g''(x,0) + [\theta\nu_0(1-x) - \theta\nu_1 x + x(1-x)\sigma]\,g'(x,0)\\
&\qquad + x(1-x)\frac{h'(x)}{h(x)}g'(x,0) + \frac{x(1-h(x))}{(1-x)h(x)}\theta\nu_1\,[g(x,1)-g(x,0)]\,,
\end{aligned}
$$

and the generator operating on $g(x,1)$ is

$$
\begin{aligned}
&G_{(X,R)}g(x,1)\\
&= \lim_{N\to\infty} N \sum_{(\ell,s)} q_{(X,R)}\big((k,1),(\ell,s)\big)\left[g\left(\frac{\ell}{N},s\right) - g\left(\frac{k}{N},1\right)\right]\\
&\left(\downarrow \text{plugging in the rates (7.12)}\right)\\
&= \lim_{N\to\infty} N\Bigg\{ \frac{1}{Nx_N+1}u_N\nu_0(N-Nx_N)\frac{h^N(Nx_N+1)}{1-h^N(Nx_N)}\left[g\left(x_N+\frac{1}{N},0\right) - g(x_N,1)\right]\\
&\qquad + \lambda^N_{Nx_N}\frac{1-h^N(Nx_N+1)}{1-h^N(Nx_N)}\left[g\left(x_N+\frac{1}{N},1\right) - g(x_N,1)\right]\\
&\qquad + \left[\mu^N_{Nx_N} - \frac{1}{N-Nx_N+1}u_N\nu_1 Nx_N\right]\frac{1-h^N(Nx_N-1)}{1-h^N(Nx_N)}\\
&\qquad\quad \left[g\left(x_N-\frac{1}{N},1\right) - g(x_N,1)\right]\Bigg\}\\
&\left(\downarrow \text{plugging in the definitions (7.1) of } \lambda^N \text{ and } \mu^N\right)\\
&= \lim_{N\to\infty} N\Bigg\{ \frac{1}{Nx_N+1}u_N\nu_0(N-Nx_N)\frac{h^N(Nx_N+1)}{1-h^N(Nx_N)}\left[g\left(x_N+\frac{1}{N},0\right) - g(x_N,1)\right]\\
&\qquad + \left[Nx_N\frac{N-Nx_N}{N}\left(\frac{1}{2}+s_N\right) + u_N\nu_0(N-Nx_N)\right]\\
&\qquad\quad \frac{1-h^N(Nx_N+1)}{1-h^N(Nx_N)}\left[g\left(x_N+\frac{1}{N},1\right) - g(x_N,1)\right]\\
&\qquad + \left[\frac{1}{2}(N-Nx_N)\frac{Nx_N}{N} + u_N\nu_1 Nx_N - \frac{1}{N-Nx_N+1}u_N\nu_1 Nx_N\right]\\
&\qquad\quad \frac{1-h^N(Nx_N-1)}{1-h^N(Nx_N)}\left[g\left(x_N-\frac{1}{N},1\right) - g(x_N,1)\right]\Bigg\}
\end{aligned}
$$

$$\Big(\downarrow \text{rearranging terms and summarising lower order terms as } o(1)\Big)$$

$$\begin{aligned}
&= \lim_{N\to\infty} \Bigg\{ \frac{1-x_N}{x_N} N u_N v_0 \frac{h^N(Nx_N+1)}{1-h^N(Nx_N)} \left[g\left(x_N+\frac{1}{N},0\right) - g\left(x_N,1\right) \right] \\
&\qquad + \frac{1}{2} N^2 \cdot x_N(1-x_N) \\
&\qquad\qquad \left[g\left(x_N+\frac{1}{N},1\right) \frac{1-h^N(Nx_N+1)}{1-h^N(Nx_N)} - g\left(x_N,1\right) \frac{1-h^N(Nx_N+1)}{1-h^N(Nx_N)} \right. \\
&\qquad\qquad\qquad \left. - g\left(x_N,1\right) \frac{1-h^N(Nx_N-1)}{1-h^N(Nx_N)} + g\left(x_N-\frac{1}{N},1\right) \frac{1-h^N(Nx_N-1)}{1-h^N(Nx_N)} \right] \\
&\qquad + \left[N \cdot x_N(1-x_N) N s_N + N \cdot N u_N v_0 (1-x_N) \right] \frac{1-h^N(Nx_N-1)}{1-h^N(Nx_N)} \\
&\qquad\quad \left[g\left(x_N+\frac{1}{N},1\right) - g\left(x_N,1\right) \right] \\
&\qquad + N \cdot N u_N v_1 x_N \frac{1-h^N(Nx_N-1)}{1-h^N(Nx_N)} \left[g\left(x_N-\frac{1}{N},1\right) - g\left(x_N,1\right) \right] + o(1) \Bigg\}
\end{aligned}$$

$$\Big(\downarrow \text{passing to the limit by applying } (7.23)\Big)$$

$$\begin{aligned}
&= \frac{(1-x)h(x)}{x(1-h(x))} \theta v_0 \left[g(x,0) - g(x,1) \right] + \frac{1}{2} x(1-x) g''(x,1) - x(1-x) \frac{h'(x)}{1-h(x)} g'(x,1) \\
&\qquad + \left[x(1-x)\sigma + \theta v_0 (1-x) \right] g'(x,0) - \theta v_1 x g'(x,0)
\end{aligned}$$

$$\Big(\downarrow \text{rearranging terms}\Big)$$

$$\begin{aligned}
&= \frac{1}{2} x(1-x) g''(x,1) + \left[\theta v_0 (1-x) - \theta v_1 x + x(1-x)\sigma \right] g'(x,1) \\
&\qquad - x(1-x) \frac{h'(x)}{1-h(x)} g'(x,1) + \frac{(1-x)h(x)}{x(1-h(x))} \theta v_0 \left[g(x,0) - g(x,1) \right].
\end{aligned}$$

Thus, the proof is complete. □

List of Figures

Bibliography

[AS05] S. R. Arthreya and J. M. Swart. Branching-coalescing particle systems. *Prob. Theory Relat. Fields*, 131:376–414, 2005.

[BB08] E. Baake and R. Bialowons. Ancestral processes with selection: Branching and Moran models. *Stochastic Models in Biological Sciences (R. Bürger, C. Maes, and J. Miękisz, eds.)*, 80:33–52, 2008.

[Ber09] N. Berestycki. Recent progress in coalescent theory. *Ensaios Mat.*, 16, 2009.

[BES04] N. H. Barton, A. M. Etheridge, and A. K. Sturm. Coalescence in a random background. *Ann. Appl. Probab.*, 14:754–785, 2004.

[BG07] E. Baake and H.-O. Georgii. Mutation, selection, and ancestry in branching models: a variational approach. *J. Math. Biol.*, 54:257–303, 2007.

[BLW16] E. Baake, U. Lenz, and A. Wakolbinger. The common ancestor type distribution of a Λ-Wright-Fisher process with selection and mutation. *Electron. Commun. Probab.*, 21, 2016.

[BP15] B. Bah and E. Pardoux. The Λ-lookdown model with selection. *Stoch. Proc. Appl.*, 125:1089–1126, 2015.

[BPS12] B. Bah, E. Pardoux, and A. B. Sow. A look-down model with selection. *Stoch. Anal. Related Topics: In Honour of Ali Süleyman Üstünel, Paris, June 2010*, pages 1–28, 2012.

[BS98] E. Bolthausen and A.-S. Sznitman. On ruelle's probability cascades and an abstract cavity method. *Commun. Math. Phys.*, 197:247–276, 1998.

[CR84] J. T. Cox and U. Rösler. A duality relation for entrance and exit laws for markov processes. *Stoch. Proc. Appl.*, 16:141 – 156, 1984.

[CS85] P. Clifford and A. Sudbury. A sample path proof of the duality for stochastically monotone markov processes. *Ann. Probab.*, 13:558–565, 1985.

[DG14] D. A. Dawson and A. Greven. *Spatial Fleming-Viot models with selection and mutation*, volume 2092 of *Lecture Notes in Mathematics*. Springer, 2014.

[DGP12] A. Depperschmidt, A. Greven, and P. Pfaffelhuber. Tree-valued Fleming-Viot dynamics with mutation and selection. *Ann. Appl. Probab.*, 22:2560–2615, 2012.

[DK99a] P. Donnelly and T. G. Kurtz. Genealogical processes for Fleming-Viot models with selection and recombination. *Ann. Appl. Probab.*, 9:1091–1148, 1999.

[DK99b] P. Donnelly and T. G. Kurtz. Particle representations for measure-valued population models. *Ann. Probab.*, 27:166–205, 1999.

[Dur08] R. Durrett. *Probability models for DNA sequence evolution*. Probability and its Applications. Springer, New York, second edition, 2008.

[EG09] A. M. Etheridge and R. C. Griffiths. A coalescent dual process in a Moran model with genic selection. *Theor. Popul. Biol.*, 75:320 – 330, 2009.

[EGT10] A. M. Etheridge, R. C. Griffiths, and J. E. Taylor. A coalescent dual process in a Moran model with genic selection, and the lambda coalescent limit. *Theor. Popul. Biol.*, 78:77 – 92, 2010.

[EK86] S. N. Ethier and T. G. Kurtz. *Markov processes*. Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986.

[EK93] S. N. Ethier and T. G. Kurtz. Fleming-Viot processes in population genetics. *SIAM J. Control Optim.*, 31:345–386, 1993.

[Eth11] A. Etheridge. *Some Mathematical Models from Population Genetics: École D'Été de Probabilités de Saint-Flour XXXIX-2009*, volume 2012. Springer, 2011.

[Ewe04] W. J. Ewens. *Mathematical Population Genetics 1: I. Theoretical Introduction*, volume 27. Springer, 2 edition, 2004.

[Fea02] P. Fearnhead. The common ancestor at a nonneutral locus. *J. Appl. Probab.*, 39:38–54, 2002.

[Fou13] C. Foucart. The impact of selection in the λ-Wright-Fisher model. *Electron. Commun. Probab.*, 18:1–10, 2013.

[FV79] W. Fleming and M. Viot. Some Measure-Valued Markov Processes in Population Genetics Theory. *Indiana Univ. Math. J.*, 28:817–843, 1979.

[FW86] H. Föllmer and A. Wakolbinger. Time reversal of infinite-dimensional diffusions. *Stoch. Proc. Appl.*, 22:59–77, 1986.

[GM] F. Gaiser and M. Möhle. On the block counting process and the fixation line of exchangeable coalescents. preprint available at arXiv:1603.09077 [math.PR].

[Gri14] R. C. Griffiths. The Λ-Fleming-Viot process and a connection with Wright-Fisher diffusion. *Adv. in Appl. Probab.*, 46:1009–1035, 2014.

[Hén15] O. Hénard. The fixation line in the Λ-coalescent. *Ann. Appl. Probab.*, 25:3007–3032, 2015.

[HM12] P. Herriger and M. Möhle. Conditions for exchangeable coalescents to come down from infinity. *ALEA Lat. Am. J. Probab. Math. Stat.*, 9:637–665, 2012.

[HP86] U. G. Haussmann and E. Pardoux. Time reversal of diffusions. *Ann. Probab.*, 14:1188–1205, 1986.

[JK14] S. Jansen and N. Kurt. On the notion(s) of duality for markov processes. *Probab. Surveys*, 11:59–120, 2014.

[KB13] S. Kluth and E. Baake. The Moran model with selection: Fixation probabilities, ancestral lines, and an alternative particle representation. *Theor. Popul. Biol.*, 90:104 – 112, 2013.

[KDH88] N. L. Kaplan, T. Darden, and R. R. Hudson. The coalescent process in models with selection. *Genetics*, 120:819–829, 1988.

[Kel82] F. P. Kelly. Markovian functions of a markov chain. *Sankhya*, 44:372–379, 1982.

[Ker13] G. Kersting. Schwache Konvergenz und Martingalprobleme. Skript, 2013.

[KHB13] S. Kluth, T. Hustedt, and E. Baake. The common ancestor process revisited. *Bull. Math. Biol.*, 75:2003–2027, 2013.

[Kim62] M. Kimura. On the probability of fixation of mutant genes in a population. *Genetics*, 47:713, 1962.

[Kin82] J. F. C. Kingman. The coalescent. *Stoch. Proc. Appl.*, 13:235–248, 1982.

[KN97] S. M. Krone and C. Neuhauser. Ancestral processes with selection. *Theor. Popul. Biol.*, 51:210 – 237, 1997.

[LKBW15] U. Lenz, S. Kluth, E. Baake, and A. Wakolbinger. Looking down in the ancestral selection graph: A probabilistic approach to the common ancestor type distribution. *Theor. Popul. Biol.*, 103:27 – 37, 2015.

[Man09] S. Mano. Duality, ancestral and diffusion processes in models with selection. *Theor. Popul. Biol.*, 75:164 – 175, 2009.

[Mor58] P. A. P. Moran. Random processes in genetics. *Math. Proc. Cambridge*, 54:60–71, 1958.

[Nag64] M. Nagasawa. Time reversions of markov processes. *Nagoya Math. J.*, 24:177–204, 1964.

[Nel58] E. Nelson. The adjoint Markoff process. *Duke Math. J.*, 25:671–690, 1958.

[NK97] C. Neuhauser and S. M. Krone. The genealogy of samples in models with selection. *Genetics*, 145:519–534, 1997.

[Pit99] J. Pitman. Coalescents with multiple collisions. *Ann. Probab.*, 27:1870–1902, 1999.

[PP13] C. Pokalyuk and P. Pfaffelhuber. The ancestral selection graph under strong directional selection. *Theor. Popul. Biol.*, 87:25 – 33, 2013.

[PW06] P. Pfaffelhuber and A. Wakolbinger. The process of most recent common ancestors in an evolving coalescent. *Stoch. Proc. Appl.*, 116:1836–1859, 2006.

[Sag99] S. Sagitov. The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.*, 36:1116–1125, 1999.

[Sch00] J. Schweinsberg. Coalescents with simultaneous multiple collisions. *Electron. J. Probab.*, 5:1–50, 2000.

[SD00] M. Stephens and P. Donnelly. Inference in molecular population genetics. *J. Roy. Stat. Soc. B*, 62:605–635, 2000.

[Shi88] T. Shiga. Stepping stone models in population genetics and population dynamics. *Stoch. Proc. Phys. Engin.*, 42:345–355, 1988.

[Sie76] D. Siegmund. The equivalence of absorbing and reflecting barrier problems for stochastically monotone markov processes. *Ann. Probab.*, 4:914–924, 1976.

[Sla00] P. F. Slade. Simulation of selected genealogies. *Theor. Popul. Biol.*, 57:35 – 49, 2000.

[SU86] T. Shiga and K. Uchiyama. Stationary states and their stability of the stepping stone model involving mutation and selection. *Prob. Theory Rel. Fields*, 73:87–117, 1986.

[Tay07] J. E. Taylor. The common ancestor process for a Wright-Fisher diffusion. *Electron. J. Probab.*, 12:808–847, 2007.

[Wak09] J. Wakeley. *Coalescent theory: an introduction*. Greenwood Village, Colorado, 2009.