FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Tobiáš Hudec

Absorption cascades of one-dimensional diffusions

Department of Probability and Mathematical Statistics

Supervisor of the master thesis:	Jan Swart, Dr.
Study programme:	Mathematics
Study branch:	Probability, Mathematical Statistics and Econometrics

Prague 2016

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In .... date .... signature of the author

I would like to thank Dr. Jan Swart for his guidance and help in writing this thesis.

Title: Absorption cascades of one-dimensional diffusions

Author: Tobiáš Hudec

Department: Department of Probability and Mathematical Statistics

Supervisor: Jan Swart, Dr., Institute of Information Theory and Automation of the ASCR

Abstract: It is known that the time until a birth and death process reaches certain level is distributed as a sum of independent exponential random variables. Diaconis, Miclo and Swart gave a probabilistic proof of this fact by coupling the birth and death process with a pure birth process such that the two processes reach the given level at the same time. We apply their techniques to find a one-dimensional diffusion and a pure birth process whose transition probabilities are related by an intertwining relation. From this we prove that the time to absorption of the diffusion has the same distribution as the time to explosion of the pure birth process, although we do not manage to couple them such that the two times are a. s. equal. This gives us a probabilistic proof of the known fact that the time to absorption of the diffusion is distributed as a sum of independent exponential random variables. We also find a coupling of a similar diffusion with the same pure birth process, which is now stopped at an arbitrary level. This allows us to interpret the diffusion as being initially reluctant to get absorbed, but later getting more and more compelled to get absorbed.

Keywords: intertwining, Markov process, diffusion

Contents

Introduction

It is known that the time until a birth and death process X_t started at the origin reaches certain level is distributed as a sum of independent exponential variables whose parameters are given by certain eigenvalues. Diaconis and Miclo [2] and Swart [10] have given a probabilistic proof of this fact by finding a pure birth process Y_t which reaches the given level at the same time as X_t. The goal of this thesis is to generalize this technique to the case that X_t is a diffusion. Since the general case is too difficult, we will mostly restrict ourselves to a special type of diffusion called the Wright-Fisher diffusion.

The technique that Diaconis, Miclo and Swart employ is called intertwining of Markov processes. For a given transition semigroup P_t of a birth and death process X_t on $\{0, \ldots, n\}$ absorbed at n, they find a transition semigroup Q_t of a pure birth process Y_t on $\{0, \ldots, n\}$ and a probability kernel K which satisfy

$$P_t K = K Q_t \tag{1}$$

and

$$K(x, n) = \mathbf{1}_{[x=n]},\ x \in \{0, \ldots, n\}\,. \tag{2}$$

The algebraic relation (1) is called intertwining, which gives the name to the intertwining of Markov processes. From (1) they derive that if both processes start at zero, then

$$\mathrm{P}\left(Y_t = y\right) = \mathrm{E}\left[K\left(X_t, y\right)\right]$$

for all $y = 0, \ldots, n$. Given (2), this means that

$$\mathrm{P}\left(Y_t = n\right) = \mathrm{P}\left(X_t = n\right),$$

and from this they conclude that the time until Y_t reaches n has the same distribution as the time until X_t reaches n. But since Y_t is a pure birth process, the time until it reaches n starting from zero is distributed as the sum of independent exponential variables, whose rates are the birth rates of Y_t.

Until now, we have only been dealing with the marginal distributions of X_t and Y_t. However, Fill [4] proved that whenever we have relation (1), the Markov processes X_t and Y_t can be coupled (i. e. defined on the same probability space) such that

$$\mathrm{P}\left(Y_t = y | X_u, 0 \le u \le t\right) = K\left(X_t, y\right). \tag{3}$$

Using this, Diaconis, Miclo and Swart proved that X_t and Y_t can be coupled such that the times of absorption of X_t and Y_t are a. s. the same.

In this thesis, we derive analogue results for a diffusion. As already mentioned, we mostly deal with the Wright-Fisher diffusion (denoted here as X_t), which is a

Markov process with state space $[-1,1]$ absorbed at its end points.[1] We find an explosive pure-birth process Y_t on $\bar{\mathbb{N}}=\mathbb{N}\cup\{\infty\}$ and a probability kernel K from $[-1,1]$ to $\mathbb{N}\cup\{\infty\}$ satisfying the intertwining relation (1) and

$$K(x,\infty)=\mathbf{1}_{[x=\pm 1]}.$$

Just as in discrete state-space, we show that if both processes start at zero, then

$$\mathrm{P}(Y_t=y)=\mathrm{E}[K(X_t,y)]$$

for all $y\in\mathbb{N}\cup\{\infty\}$, hence

$$\mathrm{P}(Y_t=\infty)=\mathrm{P}(X_t=\pm 1).$$

We conclude that the time to absorption of the Wright-Fisher diffusion starting from zero has the same distribution as the time to explosion of the pure birth process starting from zero, which is the sum of independent exponential variables whose intensities are the birth rates of Y_t.[2]

In order to find a coupling satisfying (3), we need to restrict ourselves further, as we are unable to find it for the Wright-Fisher diffusion and the pure-birth process on $\mathbb{N}\cup\{\infty\}$. However, for any $n\geq 1$, we manage to find a probability kernel K_n from $[0,1]$ to $\{0,\ldots,n\}$ and a Markov process $\left(\bar{X}_t,Y_t\right)$ such that $\bar{X}_t$ is the Wright-Fisher diffusion with reflection at zero, Y_t is the pure birth process considered before, but stopped at n, and (3) holds (with K replaced by K_n). Unfortunately, $K_n(x,n)\neq\mathbf{1}_{[x=\pm 1]}$, so $\bar{X}_t$ is not absorbed at the same time as Y_t reaches n. To get a coupling such that the time of absorption of $\bar{X}_t$ is a. s. the same as the time of explosion of Y_t, one would have to prove an analogue result for $n=\infty$. Present result find step in this direction.

Although we find the coupling only for finite n, and $\bar{X}_t$ is not absorbed at the same time as Y_t reaches n, this coupling nevertheless gives us some intuitive insight about the behavior of $\bar{X}_t$. Specifically, we find that when $Y_t=y$, $\bar{X}_t$ behaves like the Wright-Fisher diffusion with reflection at zero and with drift which depends on y. We find that this drift pushes $\bar{X}_t$ toward certain point, which we denote x_y. We also find that x_y is increasing in y, $x_0=0$ and $x_n=1$. We therefore interpret $\bar{X}_t$ as being initially reluctant to get absorbed (as Y_t is small so x_y is close to zero), but being more and more compelled to get absorbed as Y_t rises (because x_y approaches 1).

The structure of the thesis is as follows. In Chapter 1 we present the theory of Markov processes. In Chapter 2 we review the literature on the intertwining of Markov processes with discrete state-spaces. Finally, in Chapter 3 we generalize some of the results presented in Chapter 2 to the case that one of the processes is a diffusion while the other is a pure birth process.

[1] Usually, the Wright-Fisher diffusion is defined such that its state-space is $[0,1]$, but a simple linear transformation gives us a process with state-space $[-1,1]$.

[2] Just as with the birth and death processes, the distribution of the time to absorption of the Wright-Fisher diffusion has long been known, but our proof is new.

Chapter 1

Markov processes

In this chapter, we present basic results in the theory of Markov processes in continuous time. Since this theory is closely related to functional analysis, we start by stating a few functional-analytic definitions.

1.1 Preliminaries from functional analysis

Definition 1.1. Let L_1 and L_2 be Banach spaces. A function $T : L_1 \to L_2$ is called a linear operator from L_1 to L_2 if

$$T\left(\alpha f + \beta g\right) = \alpha T f + \beta T g$$

for all $\alpha, \beta \in \mathbb{R}$ and $f, g \in L_1$. If $L_1 = L_2$ we speak of a linear operator on L_1, and if $L_2 = \mathbb{R}$, we speak of a linear functional.

We say that the operator T is bounded if

$$\|T\| := \sup_{\|f\| \leq 1} \|Tf\| < \infty.$$

It is well known that a linear operator is bounded if and only if it is continuous. If $\|T\| \leq 1$, T is called a contraction. If on L_1 and L_2 we have defined partial order, we say that a linear operator $T : L_1 \to L_2$ is positive if $Tf \geq 0$ whenever $f \geq 0$, or equivalently $Tf \geq Tg$ whenever $f \geq g$.

Definition 1.2. Let L_1 and L_2 be Banach spaces and $G : \mathcal{D}(G) \to L_2$ be a linear operator, where $\mathcal{D}(G)$ is a linear subspace of L_1. We define the graph of G as $\{(f, Gf)\,;\, f \in \mathcal{D}(G)\}$. If the graph of G is a closed set in $L_1 \times L_2$ (with respect to the product topology), we say that G is a closed operator. If G is a (not necessarily closed) linear operator and the closure of the graph of G is the graph of a linear operator, which we denote $\bar{G}$, then we say that G is closable and $\bar{G}$ is its closure.

Let $(T(t))_{t\geq 0}$ be a family of continuous linear operators on a Banach space L. We say that $T(t)$ is a semigroup if $T(0) = I$, the identity operator, and $T(s+t) = T(s)T(t)$ for all $s, t \geq 0$. We say that the semigroup is strongly continuous if additionally $\lim_{t\to 0} T(t)f = f$ for all $f \in L$.

Now we define derivatives and integrals of Banach space-valued functions.

Definition 1.3. Let $u : [a, b] \to L$ for some $[a, b] \subset \mathbb{R}$ and a Banach space L. We define the derivative of u at $t_0 \in [a, b]$ as

$$\frac{\mathrm{d}}{\mathrm{d}t} u\left(t_0\right) = \lim_{t \to t_0} \frac{u(t) - u\left(t_0\right)}{t - t_0}$$

provided the limit exists.

A tagged partition of the interval $[a, b]$ is a sequence t_i, $a = t_1 < t_2 < \cdots < t_n = b$, together with a sequence s_i, $t_i \leq s_i \leq t_{i+1}$, $i = 1, \ldots, n-1$. The norm of the partition is $\max_{i=1,\ldots,n-1}\left(t_{i+1} - t_i\right)$. Suppose that for every sequence t^j of tagged partitions of the interval $[a, b]$ whose norm approaches zero the limit

$$\lim_{j \to \infty} \sum_{k=1}^{n^j - 1} u\left(s_i\right)\left(t_{i+1} - t_i\right)$$

exists and suppose that this limit does not depend on the choice of the tagged partition. Then we say that u is Riemann integrable over $[a, b]$ and write

$$\int_a^b u(t)\mathrm{d}t = \lim_{j \to \infty} \sum_{k=1}^{n^j - 1} u\left(s_i\right)\left(t_{i+1} - t_i\right).$$

It can be proved that whenever u is continuous, u is integrable. Furthermore, if u has a continuous derivative on $[a, b]$, then

$$\int_a^b \frac{\mathrm{d}}{\mathrm{d}t} u(t)\mathrm{d}t = u(b) - u(a)$$

(see Ethier and Kurtz [3], Lemma 1.4 in Chapter 1).

Next we introduce some notation.

Notation 1.4. Let $(E, \mathcal{E})$ be a measurable space. By $\mathrm{b}\mathcal{E}$ we denote the space of bounded measurable functions $f : E \to \mathbb{R}$. Furthermore, if E is a metric space (in which case we take $\mathcal{E}$ to be its Borel σ-algebra), we denote by $\mathcal{C}(E)$ the space of all continuous functions from E to $\mathbb{R}$. If E is a locally compact metric space, we denote by $\mathcal{C}_{\mathrm{K}}(E)$ the space of continuous functions with compact support and by $\mathcal{C}_0(E)$ the space of continuous functions vanishing at infinity, that is,

$$\mathcal{C}_{\mathrm{K}}(E) = \left\{f \in \mathcal{C}(E);\, \exists K \subseteq E \,\text{compact such that}\, f = 0 \,\text{on}\, E \backslash K\right\},$$

$$\mathcal{C}_0(E) = \left\{f \in \mathcal{C}(E);\, \forall \epsilon > 0\, \exists K \subseteq E \,\text{compact such that}\, |f| < \epsilon \,\text{on}\, E \backslash K\right\}.$$

Note that if E is itself compact, then $\mathcal{C}_{\mathrm{K}}(E) = \mathcal{C}_0(E) = \mathcal{C}(E)$. Finally, if $E \subseteq \mathbb{R}^d$ is an open set, we denote by $\mathcal{C}^k(E)$ the space of k-times continuously differentiable functions from E to $\mathbb{R}$, and by $\mathcal{C}^k\left(\bar{E}\right)$ the space of k-times continuously differentiable functions $f : E \to \mathbb{R}$ such that f and all its derivatives up to kth order can be extended to continuous functions on $\bar{E}$.

Lemma 1.5. *If $(E, \mathcal{E})$ is a measurable space and we equip $\mathrm{b}\mathcal{E}$ with the supremum norm, it is a Banach spaces. If E is a locally compact metric space and we equip $\mathcal{C}_0(E)$ with the supremum norm, it is also a Banach space.*

Proof. We only need to prove completeness, as the other properties from the definition of Banach space are trivial. Let $f_n \in \mathrm{b}\mathcal{E}$ be a Cauchy sequence. Then $f_n(x)$ is Cauchy in $\mathbb{R}$ for all $x \in E$, and therefore has a limit, which we denote $f(x)$. Since f is a pointwise limit of measurable functions, it is measurable. Let us fix $\epsilon > 0$. By the Cauchy property of f_n, there exists n_0 such that for all $n, m \geq n_0$ we have $\|f_n - f_m\| < \epsilon/2$. For every $x \in E$, we can find $m_x \geq n_0$ such that $|f_{m_x}(x) - f(x)| < \epsilon/2$. Then we have that for every $n \geq n_0$ and every $x \in E$, $|f_n(x) - f(x)| \leq |f_{m_x}(x) - f(x)| + |f_n(x) - f_{m_x}(x)| < \epsilon$. This proves that f_n converges to f in supremum norm, and it follows that f is bounded.

Let now f_n be a Cauchy sequence in $\mathcal{C}_0(E)$. Since $\mathcal{C}_0(E) \subseteq \mathrm{b}\mathcal{E}$, where $\mathcal{E}$ is the Borel sigma-algebra of E, there is $f \in \mathrm{b}\mathcal{E}$ such that $f_n \to f$ in the supremum norm. But convergence in the supremum norm preserves continuity, so $f \in \mathcal{C}(E)$. Let us fix $\epsilon > 0$. There is n such that $\|f_n - f\| < \epsilon/2$. Since f_n is in $\mathcal{C}_0(E)$, there is a compact set $K \subseteq E$ such that $|f_n| < \epsilon/2$ on $E \backslash K$. Hence, $|f| < \epsilon$ on $E \backslash K$, so f is also in $\mathcal{C}_0(E)$. □

1.2 Transition probabilities and semigroups

Throughout this chapter we will assume that $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathrm{P})$ is a filtered probability space.

Definition 1.6. Let $(E, \mathcal{E})$ be a measurable space and $X = (X_t)_{t \geq 0}$ be an E-valued $(\mathcal{F}_t)$-adapted random process. We say that X_t is $(\mathcal{F}_t)$-Markov process if

$$\mathrm{P}(X_t \in A | \mathcal{F}_s) = \mathrm{P}(X_t \in A | X_s) \tag{1.1}$$

a. s. for all $A \in \mathcal{E}$ and $t \geq s \geq 0$. If $\mathcal{F}_t = \sigma(X_s, 0 \leq s \leq t)$, we say that X_t is a Markov process. Observe that every $(\mathcal{F}_t)$-Markov process is a Markov process.

Next we define transition kernels.

Definition 1.7. Let $(U, \mathcal{A})$ and $(V, \mathcal{B})$ be measurable spaces. We say that function $K : U \times \mathcal{B} \to [0, 1]$ is a probability kernel (also called stochastic kernel or Markov kernel) from $(U, \mathcal{A})$ to $(V, \mathcal{B})$ if

1. for all $x \in U$ the function $K(x, \cdot)$ is a probability measure on $(V, \mathcal{B})$ and

2. for all $B \in \mathcal{B}$ the function $K(\cdot, B)$ is $\mathcal{A}$-measurable.

If $(U, \mathcal{A}) = (V, \mathcal{B})$, we say that K is a probability kernel on $(U, \mathcal{A})$.

Remark 1.8. If V is countable and $\mathcal{B} = 2^V$, then K is uniquely determined by $K(x, \{y\}), x \in U, y \in V$. To simplify the notation, we will write $K(x, y)$ instead of $K(x, \{y\})$. Then we can say that $K : U \times V \to [0, 1]$ is a probability kernel if and only if $\sum_{y \in U} K(x, y) = 1$ for all $x \in U$ and $K(\cdot, y)$ is measurable for all $y \in V$.

The Riesz representation theorem (Rudin [8], Theorem 6.19) says that there is a correspondence between measures and linear functionals. There is similar correspondence between probability kernels and linear operators, as described by the following two lemmas.

Lemma 1.9. *Let K be a probability kernel from $(U, \mathcal{A})$ to $(V, \mathcal{B})$ for some measurable spaces $(U, \mathcal{A})$ and $(V, \mathcal{B})$. For a bounded measurable function f from $(V, \mathcal{B})$ to $\mathbb{R}$ and for $x \in U$ define*

$$T_K f(x) = \int f(y) K(x, \mathrm{d}y).$$

Then T_K is a positive continuous linear operator from $\mathrm{b}\mathcal{B}$ to $\mathrm{b}\mathcal{A}$.

Proof. Measurability of $T_K f$ can be proved by approximating f by simple functions and noting that $K(\cdot, B)$ is measurable for all $B \in \mathcal{B}$. Linearity and positivity follow from the properties of the integral. Finally, $|T_K f(x)| \leq \int |f(y)| K(x, \mathrm{d}y) \leq \|f\|$, which proves that $T_K f$ is a bounded function and T_K is a bounded, hence continuous (by linearity), operator. □

There is also a converse to the previous lemma.

Lemma 1.10. *Let V be a locally compact separable metric space and $(U, \mathcal{A})$ a measurable space. Let T be a positive continuous linear operator from $\mathcal{C}_0(V)$ to $\mathrm{b}\mathcal{A}$ such that*

$$\sup_{f \in \mathcal{C}_0(V), \|f\| \leq 1} Tf(x) = 1$$

for all $x \in U$. Then there exists a probability kernel K_T from $(U, \mathcal{A})$ to $(V, \mathcal{B}(V))$ such that

$$Tf(x) = \int f(y) K_T(x, \mathrm{d}y) \tag{1.2}$$

for all $f \in \mathcal{C}_0(V)$ and $x \in U$.

Proof. (For the proof of a similar statement, see Kallenberg [5], Proposition 17.14) Let us fix $x \in U$. The functional

$$T_x : f \mapsto Tf(x)$$

is a positive linear functional on $\mathcal{C}_0(V)$. By the Riesz representation theorem (Rudin [8], Theorem 6.19), there exists a measure $K_T(x, \cdot)$ on $(V, \mathcal{B})$ such that (1.2) holds for all $f \in \mathcal{C}_0(V)$. Moreover,

$$\begin{aligned} K_T(x, V) &= \sup_{f \in \mathcal{C}_0(V), \|f\| \leq 1} Tf(x) \\ &= 1, \end{aligned}$$

so $K_T(x, \cdot)$ is in fact a probability measure. For $f \in \mathcal{C}_0(V)$, the function

$$x \mapsto \int f(y) K_T(x, \mathrm{d}y)$$

is measurable by assumption. Let G be an open set in V. Since V is a locally compact separable metric space, it is known that there exist $f_n \in \mathcal{C}_{\mathrm{K}}(V)$ such that $f_n(y) \nearrow \mathbf{1}_G(y)$ for all $y \in V$.[1] The monotone convergence theorem now proves that

$$x \mapsto \int \mathbf{1}_G(y) K_T(x, \mathrm{d}y) = K_T(x, G)$$

[1] If V is compact, we can define $f_n(y) = \min(1, nd(y, V \backslash G))$ where $d(y, V\backslash G) = \inf_{z \in V \backslash G} d(y, z)$ is the distance of y and $V \backslash G$, and $d(y, z)$ is the metric on V. However, if V is not compact, then this definition would not necessarily give us compactly supported functions. The construction of f_n in the non-compact case is more complicated and can be found at the beginning of Chapter 2 of Seidler [9].

is also measurable. If we now define

$$\mathcal{D} = \{A \in \mathcal{B}(V);\, x \mapsto K_T(x, A)\,\text{is measurable}\}\,,$$

it is easy to see that $\mathcal{D}$ is a monotone class. Since we have proved that all open sets are in $\mathcal{D}$, the monotone class theorem (Theorem 1.1 in Kallenberg [5]) now shows that $\mathcal{D} = \mathcal{B}(V)$. □

In the light of the previous lemmas we will not make a distinction between probability kernels and the corresponding linear operators. For example, we will write Kf instead of $T_K f$ where K is a probability kernel, and we will say that a linear operator satisfying the requirements of Lemma 1.10 is a probability kernel. Similarly, we will not make a distinction between measures and their corresponding linear functionals. This allows us to define composition of a measure and a kernel by operator composition. That is, if μ is a measure and K is a kernel, then by μK we denote the composition of the linear functional associated with μ with the linear operator associated with K. This composition is itself a linear functional and can be represented by a measure. It is easy to see that the measure is given by

$$(\mu K)\,(A) = \int K(x, A)\mu(\mathrm{d}x).$$

Analogously, it is possible to define composition of two kernels.

Lemma 1.10 shows that a probability kernel is uniquely determined by its restriction to continuous functions vanishing at infinity. Sometimes it will be useful to use probability kernels that map continuous functions vanishing at infinity to continuous functions vanishing at infinity.

Definition 1.11. Let U and V be locally compact metric spaces and let $\mathcal{A}$ and $\mathcal{B}$ be their Borel σ-algebras. We say that a probability kernel K from $(U, \mathcal{A})$ to $(V, \mathcal{B})$ is continuous if it maps $\mathcal{C}_0(V)$ to $\mathcal{C}_0(U)$.

Lemma 1.12. *Let U and V be compact metric spaces and let D be a dense subspace of $\mathcal{C}(V)$ containing 1. Let $K : D \to \mathcal{C}(U)$ be a positive continuous linear operator such that $K1 = 1$. Then K can be uniquely extended to a continuous probability kernel from U to V.*

Proof. Let f be in $\mathcal{C}(V)$. There exist $f_n \in D$ such that $f_n \to f$. Since the sequence f_n is Cauchy in $\mathcal{C}(V)$,

$$\|Kf_n - Kf_m\| \le \|K\|\,\|f_n - f_m\| \to 0$$

as $n, m \to \infty$, so Kf_n is also Cauchy in $\mathcal{C}(U)$. We may therefore define

$$\hat{K}f = \lim_{n\to\infty} Kf_n.$$

It is easy to see that this definition does not depend on the choice of the approximating functions f_n (if there are two such sequences f_n and g_n combine them into a new sequence h_n to show that Kh_n has a limit, and the limits of Kf_n and Kg_n must therefore be equal, since they are subsequences of Kh_n). It can also be seen that $\hat{K}f = Kf$ for $f \in D$ (just take $f_n = f$), so $\hat{K}$ is indeed an extension of K. Since K is linear, so must be $\hat{K}$, and

$$\begin{aligned}\left\|\hat{K}\right\| &= \sup_{f\in\mathcal{C}(V)}\left\|\hat{K}f\right\| \\ &= \sup_{f\in D}\|Kf\| \\ &= \|K\|\end{aligned}$$

where we have used that D is dense in $\mathcal{C}(V)$. Hence $\hat{K}$ is a continuous linear operator from $\mathcal{C}(V)$ to $\mathcal{C}(U)$. To see that $\hat{K}$ is positive, observe that if $f \in \mathcal{C}(V)$ is greater than or equal to $\epsilon > 0$ and $f_n \in D$ is its approximating sequence, then $f_n \geq 0$ for sufficiently large n. Then

$$\begin{aligned}\hat{K}f &= \lim_{n\to\infty} Kf_n \\ &\geq 0\end{aligned}$$

where we have used the positivity of K. If now $f \in \mathcal{C}(V)$ is nonnegative,

$$\hat{K}(f+\epsilon) \geq 0.$$

Taking the limit $\epsilon \to 0$ and noting that $\hat{K}$ is continuous, we get that $\hat{K}$ is positive. Finally

$$\begin{aligned}\sup_{f\in\mathcal{C}(V),\|f\|\leq 1}\hat{K}f(x) &= K1(x) \\ &= 1\end{aligned}$$

for all $x \in V$, where we have used positivity of K. Lemma 1.10 now proves that $\hat{K}$ is a continuous probability kernel. □

Let X_t be an $(\mathcal{F}_t)$-Markov process and suppose that the conditional probability on the right hand-side of (1.1) has a regular version. That is, assume that for all $t \geq s \geq 0$ there exist probability kernels $P_{s,t}$ such that

$$\mathrm{P}\left(X_t \in A|\mathcal{F}_s\right) = P_{s,t}\left(X_s, A\right)$$

a. s. for all A in $\mathcal{E}$. Observe that then we have

$$\mathrm{E}\left(f\left(X_t\right)|\mathcal{F}_s\right) = P_{s,t}f\left(X_s\right)$$

a. s. for all measurable f for which $P_{s,t}f$ is well defined, i. e. for $\mathrm{P}\circ X_s^{-1}$-almost every x, $\left(P_{s,t}f^+\right)(x)$ and $\left(P_{s,t}f^-\right)(x)$ are not both ∞ (in particular, this holds true for all $f \in \mathrm{b}\mathcal{E}$). For all $t \geq u \geq s \geq 0$ and for all $A \in \mathcal{E}$ we also have

$$\begin{aligned}P_{s,t}\left(X_s, A\right) &= \mathrm{P}\left(X_t \in A|\mathcal{F}_s\right) \\ &= \mathrm{E}\left(\mathrm{P}\left(X_t \in A|\mathcal{F}_u\right)|\mathcal{F}_s\right) \\ &= \mathrm{E}\left(P_{u,t}\left(X_u, A\right)|\mathcal{F}_s\right) \\ &= P_{s,u}P_{u,t}\left(X_s, A\right)\end{aligned}$$

a. s., which implies

$$P_{s,t}f(x) = P_{s,u}P_{u,t}f(x) \tag{1.3}$$

for all $f \in \mathrm{b}\mathcal{E}$ and $\mathrm{P}\circ X_s^{-1}$-almost every x, where the exceptional set may depend on f. Equation (1.3) is called the Chapman-Kolmogorov equality. This finding motivates the following definition, where we require that (1.3) holds for every x, not only for almost every x.

Definition 1.13. Let $(P_{s,t})_{t\geq s\geq 0}$ be a family of probability kernels on a measurable space $(E,\mathcal{E})$. We call this family a *transition probability* if for all $0\leq s\leq u\leq t$ the probability kernels satisfy

$$P_{s,t}=P_{s,u}P_{u,t}$$

and

$$P_{t,t}=I,$$

where I is the identity kernel.[2] If X_t is an $(\mathcal{F}_t)$-Markov process and $P_{s,t}$ is a transition probability such that

$$\mathrm{P}\left(X_t\in A|\mathcal{F}_s\right)=P_{s,t}\left(X_s,A\right)$$

a. s. for all $0\leq s\leq t$ and $A\in\mathcal{E}$, then we say that $P_{s,t}$ is associated with X_t.

If $P_{s,t}$ is associated with X_t and we denote by $\pi_t^X=\mathrm{P}\left(X_t\in\cdot\right)$ the distribution of X at time t, then for $0\leq s\leq t$ and $f\in\mathrm{b}\mathcal{E}$,

$$\begin{aligned}\pi_t^X(t)f &= \mathrm{E}f\left(X_t\right)\\ &= \mathrm{E}\left[\mathrm{E}\left[f\left(X_t\right)|\mathcal{F}_s\right]\right]\\ &= \mathrm{E}\left[P_{s,t}f\left(X_s\right)\right]\\ &= \pi_s^X P_{s,t}f,\end{aligned}$$

that is, $\pi_t^X=\pi_s^X P_{s,t}$. Specifically, $\pi_t^X=\pi_0^X P_{0,t}$, so the initial distribution and the transition probability determine the one-dimensional distributions of the Markov process. In a similar way, it can be proved that the initial distribution and the transition probability determine all finite-dimensional distributions. Then, whenever we are given an initial distribution and a transition probability, we can calculate finite-dimensional distributions and use the Kolmogorov extension theorem (provided we work on a Polish space) to construct an associated Markov process. This discussion leads us to the following Proposition.

Proposition 1.14. *Let π_0 be a probability measure and $(P_{s,t})_{t\geq s\geq 0}$ a transition probability on a Polish space $(E,\mathcal{E})$. Then there exists a filtered probability space $(\Omega,\mathcal{F},(\mathcal{F}_t),\mathrm{P})$ and an $(\mathcal{F}_t)$-Markov process X_t on $(E,\mathcal{E})$ satisfying*

$$\mathrm{P}\left(X_t\in A|\mathcal{F}_s\right)=P_{s,t}\left(X_s,A\right)\tag{1.4}$$

a. s. and

$$\mathrm{P}\left(X_0\in A\right)=\pi_0(A)\tag{1.5}$$

for all $t\geq s\geq 0$ and $A\in\mathcal{E}$. Moreover, if there is another filtered probability space and a random process satisfying (1.4) and (1.5), then the finite-dimensional distributions of the two processes agree.

Proof. Proposition 7.2 in Kallenberg [5] shows that transition probability and initial distribution determine finite-dimensional distributions. Theorem 7.4 in [5] in turn proves the existence. □

[2] Some authors do not require that $P_{t,t}=I$, but we will not deal with such cases.

Unfortunately, the sample paths of the constructed Markov process might be quite irregular. We would like to construct a Markov process with RCLL sample paths (right continuous having left limits). For this, further restrictions on the transition probability will be required.

In the following we will restrict our attention to time-homogeneous transition probabilities. A transition probability is called time-homogeneous if

$$P_{s,t} = P_{0,t-s}$$

for all $t \geq s \geq 0$. To simplify the notation, we will write P_t instead of $P_{0,t}$. We will also assume that E is a locally compact separable metric space and $\mathcal{E}$ is its Borel σ-algebra.

Definition 1.15. Let E be a locally compact separable metric space and let $(P_t)_{t\geq 0}$ be a family of linear operators which map $\mathcal{C}_0(E)$ into itself. We say that P_t is a Feller semigroup if it is a conservative, strongly continuous, positive semigroup, that is, for all $s, t \geq 0$,

1. $P_0 f = f$ for all $f \in \mathcal{C}_0(E)$.
2. (semigroup property) $P_{s+t} = P_s P_t$ on $\mathcal{C}_0(E)$.
3. (strong continuity) $\lim_{t\to 0+} P_t f = f$ for all $f \in \mathcal{C}_0(E)$.
4. (positivity) $P_t f \geq 0$ for every nonnegative $f \in \mathcal{C}_0(E)$.
5. (conservativity) For all $x \in E$, $\sup_{f\leq 1} P_t f(x) = 1$.

Observe that positivity and conservativity imply $\|P_t\| = 1$, hence P_t is a contraction. Note also that if E is compact, then conservativity is equivalent to $P_t 1 = 1$.

The crucial restrictions in Definition 1.15 are strong continuity and that P_t are continuous probability kernels (see Lemma 1.10 and Definition 1.11). The interpretation of the latter is that for $f \in \mathcal{C}_0(E)$,

$$P_t f(x) = \mathrm{E}\left[f\left(X_t\right)|X_0 = x\right]$$

is continuous in x and vanishing at infinity, which is a form of continuous dependence on the initial condition. With these restrictions on the transition probability, we are able to construct an RCLL process.

Theorem 1.16. *Let P_t be a Feller semigroup and π_0 be a probability distribution on a locally compact, separable metric space E. Then there exists a Markov process associated with P_t whose initial distribution is π_0 and whose sample paths are RCLL.*

Proof. See Kallenberg [5], Theorem 17.15. □

Before we proceed further, we demonstrate here an important technique of "completing" a semigroup with missing probability. Sometimes we will find a strongly continuous, positive, contraction semigroup P_t which is not conservative.

Then, the associated transition probability will be substochastic. To make P_t into a Feller semigroup we will use the following construction. Let

$$E^\Delta = E \cup \{\Delta\} \tag{1.6}$$

where Δ is a point at infinity if E is not compact (i. e. E^Δ is the one-point compactification of E) and an isolated point otherwise. Observe that E^Δ is compact, so $\mathcal{C}_0\left(E^\Delta\right) = \mathcal{C}\left(E^\Delta\right)$. For f in $\mathcal{C}_0(E)$ and $x \in E^\Delta$ define

$$f^\Delta(x) = \begin{cases} f(x), & x \in E, \\ 0, & x = \Delta, \end{cases}$$

and observe that f^Δ is in $\mathcal{C}\left(E^\Delta\right)$. Observe also that

$$\begin{aligned} \mathcal{C}\left(E^\Delta\right) &= \left\{a + f^\Delta;\, a \in \mathbb{R}, f \in \mathcal{C}_0(E)\right\}, \\ \mathcal{C}_0(E) &= \left\{f \restriction_E - f(\Delta);\, f \in \mathcal{C}\left(E^\Delta\right)\right\}, \end{aligned}$$

where by $f \restriction_E$ we mean the restriction of f to E, and define

$$P_t^\Delta f = f(\Delta) + \left(P_t\left(f \restriction_E - f(\Delta)\right)\right)^\Delta \tag{1.7}$$

for $f \in \mathcal{C}\left(E^\Delta\right)$. We will prove that P_t^Δ is a Feller semigroup on $\mathcal{C}\left(E^\Delta\right)$ and Δ is an absorbing point.

Definition 1.17. Let P_t be a transition probability on a metric space E. We say that $x \in E$ is an absorbing point if $P_t(x, \cdot) = \delta_x$.

The meaning of the absorbing point is that if a Markov process ever reaches this point, it stays there forever. It is easy to see that x is an absorbing point if and only if $P_t f(x) = f(x)$ for all $f \in \mathcal{C}_0(E)$, since a measure is uniquely determined by how it integrates continuous functions.

Lemma 1.18. *Let P_t be a strongly continuous, positive, contraction semigroup on $\mathcal{C}_0(E)$ for a locally compact separable metric space E. Let E^Δ and P_t^Δ be defined by (1.6) and (1.7). Then P_t^Δ is a Feller semigroup on $\mathcal{C}\left(E^\Delta\right)$ and Δ is an absorbing point.*

Proof. (See also Lemma 2.3 in Chapter 4 of Ethier and Kurtz [3].) First, P_t^Δ maps $\mathcal{C}\left(E^\Delta\right)$ into itself, since P_t maps $\mathcal{C}_0(E)$ into itself. Next, P_t^Δ is a semigroup, since

$$\begin{aligned} P_0^\Delta f &= f(\Delta) + \left(P_0\left(f \restriction_E - f(\Delta)\right)\right)^\Delta \\ &= f(\Delta) + f - f(\Delta) \\ &= f \end{aligned}$$

and

$$\begin{aligned} P_s^\Delta P_t^\Delta f &= P_t^\Delta f(\Delta) + P_s\left(\left(P_t^\Delta f\right) \restriction_E - P_t^\Delta f(\Delta)\right)^\Delta \\ &= f(\Delta) + P_s P_t\left(f \restriction_E - f(\Delta)\right)^\Delta \\ &= P_{s+t}^\Delta f \end{aligned}$$

for $s, t \geq 0$ and $f \in \mathcal{C}\left(E^{\Delta}\right)$. Next, for $f \in \mathcal{C}\left(E^{\Delta}\right)$ we have

$$\begin{aligned} \left\|P_t^{\Delta} f - f\right\| &= \left\|\left(P_t\left(f \restriction_E - f(\Delta)\right)\right)^{\Delta} - \left(f - f(\Delta)\right)\right\| \\ &= \left\|P_t\left(f \restriction_E - f(\Delta)\right) - \left(f \restriction_E - f(\Delta)\right)\right\|, \end{aligned}$$

which converges to 0 as $t \to 0$ by the strong continuity of P_t, hence P_t^{Δ} is also strongly continuous. To prove positivity, let $0 \leq f \in \mathcal{C}\left(E^{\Delta}\right)$ and write

$$P_t^{\Delta} f = f(\Delta) + \left(P_t\left(f \restriction_E - f(\Delta)\right)^{+}\right)^{\Delta} - \left(P_t\left(f \restriction_E - f(\Delta)\right)^{-}\right).$$

Now $\left(P_t\left(f \restriction_E - f(\Delta)\right)^{+}\right) \geq 0$ by the positivity of P_t. Also $\left\|\left(f \restriction_E - f(\Delta)\right)^{-}\right\| \leq f(\Delta)$ since f is positive, so $P_t\left(f \restriction_E - f(\Delta)\right)^{-} \leq f(\Delta)$ by the contraction property of P_t. Hence, $P_t^{\Delta} f \geq 0$. Finally,

$$\begin{aligned} P_t^{\Delta} 1 &= 1 + P_t 0 \\ &= 1, \end{aligned}$$

so P_t^{Δ} is conservative. We have therefore proved that P_t^{Δ} is a Feller semigroup. Moreover, $P_t^{\Delta} f(\Delta) = f(\Delta)$ for all $f \in \mathcal{C}\left(E^{\Delta}\right)$, so Δ is absorbing. □

In this section we have found a way to characterize a Markov process in terms of its transition probability. Unfortunately, the transition probability cannot be specified explicitly by a formula except for a few special cases. For this reason, we will define generators which uniquely determine transition probabilities and in many practical cases can be given by formulas.

1.3 Generators

In this section, we will describe how to characterize Feller semigroups in terms of their generators. Since part of the theory does not depend on specific assumptions about Feller semigroups, we will start this section by presenting results about strongly continuous, contraction semigroups, and later specialize them to the case of Feller semigroups.

Definition 1.19. Let L be a Banach space. We say that a family of linear operators $(P_t)_{t \geq 0}$ on L is a strongly continuous contraction semigroup if

1. $P_0 = I$,
2. (semigroup property) $P_{s+t} = P_s P_t$ on L for all $s, t \geq 0$,
3. (strong continuity) $\lim_{t \to 0+} P_t f = f$ for all f in L
4. (contractiveness) $\|P_t\| \leq 1$ for all $t \geq 0$.

Let P_t be a strongly continuous contraction semigroup on a Banach space L. The infinitesimal generator of P_t is defined as

$$G f = \lim_{t \to 0+} \frac{1}{t}\left(P_t f - f\right).$$

The domain of G, denoted by $\mathcal{D}(G)$, is taken to be the subspace of L such that the limit exists.

Theorem 1.20. *(Hille-Yosida) A linear operator G on L is the generator of a strongly continuous contraction semigroup on L if and only if*

1. *$\mathcal{D}(G)$ is dense in L,*

2. *G is dissipative, that is, $\|\lambda f - Gf\| \geq \lambda \|f\|$ for every $f \in \mathcal{D}(G)$ and $\lambda > 0$,*

3. *$\mathcal{R}(\lambda - G) = L$ for some $\lambda > 0$.*

Proof. See Ethier and Kurtz [3], Theorem 1.2.6. □

Corollary 1.21. *Let G be the generator of a strongly continuous contraction semigroup on a Banach space L. Then G is closed (see Definition 1.2). Moreover, $\mathcal{D}(G) = L$ if and only if G is bounded (in particular if L is finite-dimensional).*

Proof. Since G is dissipative and $\mathcal{R}(\lambda - G) = L$ is closed, Lemma 2.2 in Chapter 1 of Ethier and Kurtz [3] shows that G is closed.

Suppose now that G is bounded and let f be in L. Then there are $f_n \in \mathcal{D}(G)$ such that $f_n \to f$. Since f_n is Cauchy,

$$\begin{aligned} \|Gf_n - Gf_m\| &\leq \|G\| \, \|f_n - f_m\| \\ &\to 0 \end{aligned}$$

as $n, m \to \infty$. Therefore, Gf_n is Cauchy and has a limit. Since G is closed, f must be in $\mathcal{D}(G)$. But f was arbitrary, so $\mathcal{D}(G) = L$.

On the other hand, if $\mathcal{D}(G) = L$, then the closed graph theorem shows that G is continuous, hence bounded. □

The domain of a generator might be difficult to describe. For this reason, we often define a generator only on a "nice" set and then we take the closure.

Theorem 1.22. *(Hille-Yosida) A linear operator G on L is closable and its closure (see Definition 1.2) is the generator of a strongly continuous contraction semigroup on L if and only if*

1. *$\mathcal{D}(G)$ is dense in L.*

2. *G is dissipative.*

3. *$\mathcal{R}(\lambda - G)$ is dense in L for some $\lambda > 0$.*

Proof. See Ethier and Kurtz [3], Theorem 1.2.12. □

The preceding theorem motivates the following definition.

Definition 1.23. Let G be the generator of a strongly continuous contraction semigroup on a Banach space L, and let D be a subspace of $\mathcal{D}(G)$. We say that D is a core of G if the closure of $G \restriction_D$ is G.

It should be noted that not every dense subspace of $\mathcal{D}(G)$ is a core of G. For a counterexample, see Liggett [6], Remark 3.57. However, there is a simple sufficient condition for a dense subspace of $\mathcal{D}(G)$ to be a core.

Proposition 1.24. *(Invariance and cores) Let G be the generator of a strongly continuous contraction semigroup P_t on a Banach space L and let $D \subseteq \mathcal{D}(G)$ be a dense subspace of L such that $P_t : D \to D$ for all $t \geq 0$. Then D is a core for G.*

Proof. See Ethier and Kurtz [3], Proposition 3.3 in Chapter 1. □

Proposition 1.25. *Let G and H be generators of strongly continuous semigroups on a Banach space L. Then we have the following:*

1. *If G is an extension of H (that is, $\mathcal{D}(G) \supseteq \mathcal{D}(H)$ and $Gf = Hf$ for all $f \in \mathcal{D}(H)$), then $G = H$.*
2. *If D is a core of H such that $D \subseteq \mathcal{D}(G)$ and $Gf = Hf$ for all $f \in D$, then $G = H$.*

Proof. Ethier and Kurtz [3], Theorem 4.1 in Chapter 1, shows that if H is a generator and G is a linear and dissipative extension of H, then $G = H$. Since G is linear and dissipative by Theorem 1.20, the first assertion is proved. For the proof of the second assertion let f be in $\mathcal{D}(H)$. Then there are $f_n \in D$ such that $f_n \to f$ and $Gf_n = Hf_n \to Hf$. Since G is closed, f is in $\mathcal{D}(G)$ and $Gf = Hf$. Hence, G is an extension of H and the second assertion follows from the first. □

If $L = \mathcal{C}_0(E)$ for some locally compact metric space, then we have the following positive maximum principle, which we can use in the Hille-Yosida theorem instead of dissipativity. In addition, it will imply positivity of the associated semigroup.

Definition 1.26. Let E be a locally compact metric space and let G be a linear operator defined on a subspace of $\mathcal{C}_0(E)$. We say that G satisfies the *positive maximum principle* if $\forall f \in \mathcal{D}(G), \forall x_0 \in E$ such that $\sup_x f(x) = f(x_0) \geq 0$ we have $Gf(x_0) \leq 0$.

Theorem 1.27. *(Hille-Yosida for continuous functions) Let E be a locally compact metric space. A linear operator G on $\mathcal{C}_0(E)$ is closable and its closure is the generator of a strongly continuous, positive, contraction semigroup on $\mathcal{C}_0(E)$ if and only if*

1. *$\mathcal{D}(G)$ is dense in $\mathcal{C}_0(E)$.*
2. *G satisfies the positive maximum principle.*
3. *$\mathcal{R}(\lambda - G)$ is dense in $\mathcal{C}_0(E)$ for some $\lambda > 0$.*

Proof. See Ethier and Kurtz [3], Theorem 4.2.2. □

The Hille-Yosida theorem gives us a necessary and sufficient condition for an operator to be the generator of a semigroup. However, it does not tell us how to calculate the associated semigroup. For that purpose, the next proposition might be useful.

Proposition 1.28. *(Kolmogorov backward equation) Let P_t be a strongly continuous semigroup on a Banach space L and let G be its generator. Let $u : [0, \infty) \to L$ and assume that $f = u(0) \in \mathcal{D}(G)$. Then the following are equivalent*

1. *u is a continuous function such that $u(t) \in \mathcal{D}(G)$ for all $t \geq 0$ and $Gu : (0, \infty) \to L$ is continuous. Moreover,*

$$\frac{\mathrm{d}}{\mathrm{d}t} u(t) = Gu(t)$$

for all $t > 0$, or equivalently

$$u(t) = u(\epsilon) + \int_{\epsilon}^{t} Gu(s)\, \mathrm{d}s \tag{1.8}$$

for all $t > \epsilon > 0$.

2. *$u(t) = P_t f$ for all $t \geq 0$.*

Proof. (See also Ethier and Kurtz [3], Proposition 3.4 in Chapter 1) Let us first prove that $1 \Rightarrow 2$. Define $v : [0, \infty) \to L$ by

$$v(t) = P_t f.$$

By strong continuity and the semigroup property, v is continuous. By Proposition 1.5 in Chapter 1 of [3], $v(t) \in \mathcal{D}(G)$ for all $t \geq 0$, $Gv(t) = P_t G f$ (which implies that $Gv(t)$ is continuous) and v satisfies (1.8). Hence, by Proposition 2.10 in Chapter 1 of [3], $\|u(t) - v(t)\| \leq \|u(0) - v(0)\| = 0$ for all $t \geq 0$. Therefore, $u = v$.

Now we will prove that $2 \Rightarrow 1$. Clearly, $u(t)$ is continuous by the strong continuity and the semigroup property of P_t. Proposition 1.5 in Chapter 1 of [3] shows that $u(t) \in \mathcal{D}(G)$ and

$$\frac{\mathrm{d}}{\mathrm{d}t} u(t) = Gu(t) = P_t G f$$

for all $t \geq 0$. It follows from the strong continuity of P_t that $Gu(t)$ is continuous. $\square$

Corollary 1.29. *Let G be the generator of a Feller semigroup on a separable, locally compact metric space E. Then $x \in E$ is an absorbing point if and only if $Gf(x) = 0$ for all $f \in \mathcal{D}(G)$.*

Proof. Assume that x is an absorbing point and let $f \in \mathcal{D}(G)$. Then $P_t f(x) = f(x)$ and

$$\begin{aligned} Gf(x) &= \lim_{t \to 0} \frac{1}{t} (P_t f(x) - f(x)) \\ &= 0. \end{aligned}$$

On the other hand, suppose that $Gf(x) = 0$ for all $f \in \mathcal{D}(G)$. Let f be in $\mathcal{D}(G)$. Then by Proposition 1.28 $P_t f$ is in $\mathcal{D}(G)$ and

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} P_t f(x) &= G P_t f(x) \\ &= 0, \end{aligned}$$

hence $t \mapsto P_t f(x)$ is constant, so $P_t f(x) = f(x)$. Since $\mathcal{D}(G)$ is dense in $\mathcal{C}_0(E)$ and P_t is a continuous operator, $P_t f(x) = f(x)$ for all $f \in \mathcal{C}_0(E)$, hence x is absorbing. $\square$

Corollary 1.30. *Let L_1 be a closed subspace of a Banach space L_2 and let P_t and Q_t be strongly continuous semigroups on L_1 and L_2 with generators G and H. Suppose that $\mathcal{D}(G) \subseteq \mathcal{D}(H)$ and $G = H$ on $\mathcal{D}(G)$. Then $P_t = Q_t$ on L_1 for all $t \geq 0$. In particular, strongly continuous semigroups are uniquely determined by their generators.*

Proof. Let f be in $\mathcal{D}(G)$. Then by applying implication $2 \Rightarrow 1$ of Proposition 1.28 on $u(t) = P_t f$ we get that $P_t f$ and $\frac{\mathrm{d}}{\mathrm{d}t} P_t f$ are continuous, $P_t f$ is in $\mathcal{D}(G)$ and

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} P_t f &= G P_t f \\ &= H P_t f \end{aligned}$$

for all $t \geq 0$, where we have used that $G = H$ on $\mathcal{D}(G)$. By implication $1 \Rightarrow 2$ of the same Proposition, $P_t f = Q_t f$. Since $\mathcal{D}(G)$ is dense in L_1, $P_t f = Q_t f$ for all $f \in L_1$. □

Recall that a strongly continuous, positive, contraction semigroup P_t on $\mathcal{C}(E)$ for a compact metric space E is conservative if and only if $P_t 1 = 1$ (see Definition 1.15).

Corollary 1.31. *Let P_t be a strongly continuous, positive, contraction semigroup on $\mathcal{C}(E)$, where E is a compact metric space, and let G be its generator. Then P_t is conservative (hence Feller) if and only if 1 is in $\mathcal{D}(G)$ and $G1 = 0$.*

Proof. Assume that P_t is conservative. Then

$$\frac{P_t 1 - 1}{t} = 0,$$

hence $1 \in \mathcal{D}(G)$ and $G1 = 0$.

Assume now that 1 is in $\mathcal{D}(G)$ and $G1 = 0$ and define $u(t) = 1$. By Proposition 1.28, $P_t 1 = u(t) = 1$. □

Now we return to Lemma 1.18.

Lemma 1.32. *Let P_t be a positive, strongly continuous, contraction semigroup on $\mathcal{C}_0(E)$ for some locally compact metric space and let G be its generator. Define E^Δ and P_t^Δ as in Lemma 1.18 and let G^Δ be the generator of P_t^Δ. Then*

$$\mathcal{D}\left(G^\Delta\right) = \left\{a + f^\Delta;\ a \in \mathbb{R}, f \in \mathcal{D}(G)\right\},$$

or equivalently

$$\mathcal{D}(G) = \left\{f \restriction_E - f(\Delta);\ f \in \mathcal{D}\left(G^\Delta\right)\right\},$$

and

$$G^\Delta f = \left(G\left(f \restriction_E - f(\Delta)\right)\right)^\Delta \tag{1.9}$$

for $f \in \mathcal{D}\left(G^\Delta\right)$. Moreover, if $D \subseteq \mathcal{D}(G)$ is a core of G, then

$$D^\Delta = \left\{a + f^\Delta;\ a \in \mathbb{R}, f \in D\right\}$$

is a core of G^Δ.

Proof. Let f be in $\mathcal{C}\left(E^{\Delta}\right)$. Observe that

$$\frac{P_t^{\Delta} f - f}{t} = \left(\frac{P_t\left(f \restriction_E - f(\Delta)\right) - \left(f \restriction_E - f(\Delta)\right)}{t}\right)^{\Delta}$$

and this expression converges as $t \to 0$ if and only if $f \restriction_E - f(\Delta)$ is in $\mathcal{D}(G)$ in which case the limit is given by (1.9).

To prove the statement about the cores, let f be in $\mathcal{D}\left(G^{\Delta}\right)$. We need to find $f_n \in D^{\Delta}$ such that $f_n \to f$ and $G^{\Delta} f_n \to G^{\Delta} f$. But we know that $f \restriction_E - f(\Delta)$ is in $\mathcal{D}(G)$, so there are $\tilde{f}_n \in D$ such that $\tilde{f}_n \to f \restriction_E - f(\Delta)$ and $G\tilde{f}_n \to G\left(f \restriction_E - f(\Delta)\right)$. It is now easy to see that $f_n = \tilde{f}_n^{\Delta} + f(\Delta)$ satisfy the requirements. $\square$

1.4 Diffusions

In this section we describe diffusions. A diffusion is a Feller process with continuous sample paths. First, we give a sufficient condition for the generator of a Feller semigroup to be associated with a diffusion.

Theorem 1.33. *Let G be the generator of a Feller semigroup on $\mathcal{C}_0(E)$ where E is a locally compact separable metric space. Suppose that for each $x_0 \in E$ and $\epsilon > 0$ there exists $f \in \mathcal{D}(G)$ such that $f(x_0) = \|f\|$, $\sup_{x \in E \setminus B(x_0, \epsilon)} f(x) < \|f\|$ and $Gf(x_0) = 0$. Then almost all sample paths of the process associated with G are continuous.*

Proof. See Ethier and Kurtz [3], Proposition 2.9 and Remark 2.10 in Chapter 4. $\square$

Now we give an example of a diffusion process. Define $\mathcal{C}^2[-1,1]$ as the space of continuous functions on $[-1,1]$ whose restrictions to $(-1,1)$ are twice continuously differentiable and whose second derivative has finite limits as $x \to \pm 1$ and therefore can be regarded as a continuous function on $[-1,1]$. Let us define an operator on $\mathcal{C}^2[-1,1]$ by

$$G^{WF} = \left(1 - x^2\right)\frac{\partial^2}{\partial x^2}. \tag{1.10}$$

We will now prove that G^{WF} generates a diffusion process. The process is called the Wright-Fisher diffusion (see Liggett [6], Example 3.48).[3]

Lemma 1.34. *The operator G^{WF} defined by (1.10) is closable and its closure is the generator of a Feller semigroup. Moreover, the associated Markov process is a diffusion.*

Proof. (See also Liggett [6], Example 3.48 and Theorem 3.49) First, we verify the conditions of the Hille-Yosida theorem for continuous functions (Theorem 1.27) to prove that G^{WF} is closable and its closure generates a strongly continuous, positive, contraction semigroup.

[3] Note that the Wright-Fisher diffusion is usually defined on $[0,1]$, but a linear transformation of space yields our definition.

1. $\mathcal{D}\left(G^{WF}\right)$ contains the set of all polynomials on $[-1,1]$, which is dense in $\mathcal{C}[-1,1]$ by the Stone-Weierstrass theorem.

2. Let $f \in \mathcal{C}^2[-1,1]$ and $x_0 \in [-1,1]$ be such that $\sup_{x\in[-1,1]} f(x) = f(x_0)$. If $x_0 \in (-1,1)$ then $\frac{\partial^2 f}{\partial x^2}(x_0) \leq 0$. If $x_0 \in \{-1,1\}$ then $1 - x_0^2 = 0$. In both cases, $G^{WF} f(x_0) \leq 0$.

3. We show that $\mathcal{R}(\lambda - G)$ contains the set of all polynomials. Let

$$g(x) = \sum_{k=0}^{n} b_k x^k$$

be given. We will try to find a polynomial of the form

$$f(x) = \sum_{k=0}^{n} a_k x^k$$

such that $\left(\lambda - G^{WF}\right) f = g$. We have

$$\begin{aligned}
(\lambda - G^{WF}) f(x) &= \left[\lambda - \left(1 - x^2\right)\frac{\partial^2}{\partial x^2}\right] \sum_{k=0}^{n} a_k x^k \\
&= \sum_{k=0}^{n} a_k \left[\lambda x^k - \left(1 - x^2\right) k(k-1) x^{k-2}\right] \\
&= \sum_{k=0}^{n} a_k \left[\lambda + k(k-1)\right] x^k \\
&\quad - \sum_{m=0}^{n-2} a_{m+2}(m+2)(m+1) x^m \\
&= \sum_{k=0}^{n} \Big\{ a_k \left[\lambda + k(k-1)\right] \\
&\quad - \mathbf{1}_{[k\leq n-2]} a_{k+2}(k+2)(k+1) \Big\} x^k
\end{aligned}$$

Therefore, coefficients of f must satisfy

$$\begin{gathered}
\left(\lambda + k(k-1)\right) a_k - (k+2)(k+1) a_{k+2} = b_k,\ k = 0, \ldots, n-2, \\
\left(\lambda + (n-1)(n-2)\right) a_{n-1} = b_{n-1}, \\
\left(\lambda + n(n-1)\right) a_n = b_n.
\end{gathered}$$

These equations can be solved recursively, starting from a_n working down to a_0. Then f satisfies $(\lambda - G^{WF}) f = g$, which means $g \in \mathcal{R}(\lambda - G)$.

Since $G_{WF} 1 = (1 - x^2) \frac{\partial^2}{\partial x^2} 1 = 0$, by Corollary 1.31, the associated semigroup is conservative. We have therefore shown that G^{WF} generates a Feller semigroup. To prove that the associated Markov process is a diffusion, we verify the conditions of Theorem 1.33.

Let $x_0 \in [-1,1]$ and $\epsilon > 0$ be given and define f as

$$f(x) = 16 - (x - x_0)^4.$$

Clearly $f \geq 0$ and f attains its unique maximum at x_0, so

$$\sup_{x\in[-1,1]\setminus(x_0-\epsilon,x_0+\epsilon)} f(x) < f(x_0) = \|f\|.$$

Moreover,

$$\frac{\partial^2}{\partial x^2} f(x_0) = 0,$$

so $G^{WF} f(x_0) = 0$. □

In general, it is difficult to determine the domain of a generator of a Feller semigroup. However, in one dimension the situation is easier.

Lemma 1.35. $\mathcal{D}(\bar{G}^{WF}) = \{f \in \mathcal{C}[-1,1] \cap \mathcal{C}^2(-1,1);\ \lim_{x\to\pm 1} (1-x^2) f''(x) = 0\}$.

Proof. Use Theorem 8.1.1 in [3]. □

For future reference, we also give the following useful representation of the Wright-Fisher semigroup for polynomials.

Lemma 1.36. *Let P_t^{WF} be the semigroup associated with G^{WF}. If $p \in \mathcal{C}[-1,1]$ is a polynomial,*

$$p(x) = \sum_{k=0}^{n} a_k x^k,$$

then there exist $b_k \in \mathcal{C}[0,\infty)$, $k = 0,\ldots,n$ such that $b_k(0) = a_k$ and

$$P_t^{WF} p(x) = \sum_{k=0}^{n} b_k(t) x^k.$$

Moreover, if $p(\cdot)$ is an even polynomial, then so is $P_t^{WF} p(\cdot)$ for all $t \geq 0$.

Proof. We will find $b_k(t)$ such that $u(t) := \sum_{k=0}^{n} b_k(t) x^k$ satisfies the Kolmogorov backward equation (Proposition 1.28). First observe that[4]

$$\frac{\mathrm{d}}{\mathrm{d}t} u(t) = \sum_{k=0}^{n} b_k'(t) x^k$$

since

$$\begin{aligned}
\limsup_{s\to 0} \sup_{x\in[-1,1]} &\left| \frac{u(t+s,x) - u(t,x)}{s} - \sum_{k=0}^{n} b_k'(t) x^k \right| \\
&= \limsup_{s\to 0} \sup_{x\in[-1,1]} \left| \sum_{k=0}^{n} \left(\frac{b_k(t+s) - b_k(t)}{s} - b_k'(t) \right) x^k \right| \\
&\leq \lim_{s\to 0} \sum_{k=0}^{n} \left| \frac{b_k(t+s) - b_k(t)}{s} - b_k'(t) \right| \\
&= 0.
\end{aligned}$$

[4]The derivative here is interpreted in the sense of Definition 1.3, where the Banach space is $\mathcal{C}[-1,1]$, not simply as the point derivative.

By the Kolmogorov backward equation, we need $\frac{\mathrm{d}}{\mathrm{d}t}u(t)$ to be equal to

$$\begin{aligned} G^{WF}u(t) &= \sum_{k=0}^{n} k(k-1)b_k(t)\left(1-x^2\right)x^{k-2} \\ &= \sum_{k=0}^{n-2}(k+2)(k+1)b_{k+2}(t)x^k - \sum_{k=0}^{n} k(k-1)b_k(t)x^k. \end{aligned}$$

That is

$$b_k'(t) = \mathbf{1}_{[k\leq n-2]}(k+2)(k+1)b_{k+2}(t) - k(k-1)b_k(t). \tag{1.11}$$

From the Kolmogorov backward equation we also have the initial condition

$$u(0) = p,$$

that is

$$b_k(0) = a_k$$

for all k. This is a system of linear differential equations, hence it has a unique solution.

Finally note that the right-hand side of (1.11) only uses b_k and b_{k+2}, so if $a_k = 0$ for all odd k, then $b_k(t) = 0$ for all odd k, so $u(t)$ is an even polynomial. □

Corollary 1.37. *The set of polynomials on* $[-1,1]$ *is a core for the Wright-Fisher diffusion.*

Proof. Follows from the previous Lemma and Proposition 1.24. □

We describe one more diffusion process which we call the Wright-Fisher diffusion with reflection at zero. Let $f \in \mathcal{C}^2[0,1]$ be such that $\frac{\partial}{\partial x}f(0) = 0$ and define

$$G^{WF,r}f(x) = \left(1-x^2\right)\frac{\partial^2}{\partial x^2}f(x), \tag{1.12}$$

$x \in [0,1]$. Note that this formula is the same as that of the generator of ordinary Wright-Fisher diffusion, but the domain is different.

Lemma 1.38. $G^{WF,r}$ *is closable and its closure is the generator of a diffusion process.*

Proof. This proof is analogous to that of Lemma 1.34. We start by verifying the conditions of Theorem 1.27.

1. $\mathcal{D}\left(G^{WF,r}\right)$ contains even polynomials and they are dense in $\mathcal{C}[0,1]$ by the Stone-Weierstrass theorem.

2. Let f be in $\mathcal{D}\left(G^{WF,r}\right)$ and $x_0 \in [0,1]$ be such that $\sup_{x\in[0,1]} f(x) = f(x_0)$. If $x_0 \in (0,1)$ then $\frac{\partial^2 f}{\partial x^2}(x_0) \leq 0$. If $x_0 = 0$, then $\frac{\partial^2 f}{\partial x^2}(x_0) \leq 0$, since $\frac{\partial}{\partial x}f(x_0) = 0$. If $x_0 = 1$ then $1 - x_0^2 = 0$. In all cases, $G^{WF,r}f(x_0) \leq 0$.

3. We claim that $\mathcal{R}\left(\lambda - G^{WF,r}\right)$ contains the set of all even polynomials. In other words, for every even polynomial g there exists $f \in \mathcal{D}\left(G^{WF,r}\right)$ such that $\left(\lambda - G^{WF,r}\right)f = g$. The calculations are the same as in the proof of Lemma 1.34, we just need to point out that the equations there guarantee that whenever g is even, then so is f.

Thus, we have proved that $G^{WF,r}$ is closable and its closure generates a strongly continuous, positive contraction semigroup. Since $G^{WF,r}1 = 0$, the semigroup is conservative, hence Feller. We conclude the proof by verifying the conditions of Theorem 1.33.

Let $x_0 \in [0,1]$ and $\epsilon > 0$ be given. Define

$$f(x) = 1 - \left(x^2 - x_0^2\right)^4.$$

Then $f \geq 0$ on $[0,1]$ and it attains its unique maximum at x_0. Hence

$$\sup_{x \in [0,1] \setminus (x_0 - \epsilon, x_0 + \epsilon)} f(x) < f(x_0) = \|f\|.$$

Moreover,

$$\frac{\partial^2}{\partial x^2} f(x_0) = 0,$$

so $G^{WF,r} f(x_0) = 0$. □

Remark 1.39. In analogy to Lemma 1.36, we can prove that if p is an even polynomial, then so is $P_t^{WF,r} p$. Moreover, since the Kolmogorov backward equations are the same, $P_t^{WF,r} p(x) = P_t^{WF} p(x)$ for $x \in [0,1]$. Then, just as in the Corollary 1.37, we have that the set of even polynomials on $[0,1]$ is a core for the Wright-Fisher diffusion with reflection at zero.

We say that a Markov process whose generator is $G^{WF,r}$ is the Wright-Fisher diffusion with reflection at zero. The next lemma justifies this name.

Lemma 1.40. *Let X_t be the Wright-Fisher diffusion. Then $|X_t|$ is the Wright-Fisher diffusion with reflection at zero.*

Proof. (Adapted from Liggett [6], Example 3.55, where similar techniques are used for Brownian motion with reflection at zero.) For $f : [0,1] \to \mathbb{R}$ define the even extension by

$$f_e(x) = \begin{cases} f(x), & x \in [0,1], \\ f(-x), & x \in [-1,0). \end{cases}$$

First we will prove that

$$\left(P_t^{WF,r} f\right)_e = P_t^{WF} f_e \tag{1.13}$$

for all $f \in \mathcal{C}[0,1]$. If f is an even polynomial, then (1.13) follows from Remark 1.39. The general case now follows from the fact that even polynomials are dense in $\mathcal{C}[0,1]$.

Let now f be in $\mathcal{C}[0,1]$. Then for $s, t \geq 0$,

$$\begin{aligned}
&\mathrm{E}\left[f\left(|X_{s+t}|\right) \mid |X_u|, 0 \leq u \leq s\right] \\
&\qquad = \mathrm{E}\left[\mathrm{E}\left[f_e\left(X_{s+t}\right) \mid X_u, 0 \leq u \leq s\right] \mid |X_u|, 0 \leq u \leq s\right] \\
&\qquad = \mathrm{E}\left[P_t^{WF} f_e\left(X_s\right) \mid |X_u|, 0 \leq u \leq s\right] \\
&\qquad = \mathrm{E}\left[\left(P_t^{WF,r} f\right)_e\left(X_s\right) \mid |X_u|, 0 \leq u \leq s\right] \\
&\qquad = \left(P_t^{WF,r} f\right)\left(|X_s|\right)
\end{aligned}$$

a. s. □

1.5 Birth and death process

In this section, we define birth and death processes.

Definition 1.41. We say that a linear operator G on $\mathbb{R}^{\{0,\dots,n\}}$ is the generator of a birth and death process on $\{0,\dots,n\}$ if there exist non-negative numbers $b_0,\dots,b_{n-1}$ and $d_1,\dots,d_n$ such that

$$Gf(k)=\begin{cases} b_0 f(1)-b_0 f(0), & k=0,\\ b_k f(k+1)-(b_k+d_k)\,f(k)+d_k f(k-1), & 0<k<n,\\ d_k f(k-1)-d_k f(k), & k=n.\end{cases}$$

Similarly, we say that a linear operator G on $\mathbb{R}^{\mathbb{N}}$ is the generator of a birth and death process on $\mathbb{N}$ if there exist non-negative numbers $b_0,b_1,\dots$ and $d_1,d_2,\dots$ such that

$$Gf(k)=\begin{cases} b_0 f(1)-b_0 f(0), & k=0\\ b_k f(k+1)-(b_k+d_k)\,f(k)+d_k f(k-1), & k>0\end{cases}$$

The numbers $b_0,b_1,\dots$ are called birth rates and $d_1,d_2,\dots$ are called death rates. If all death rates are zero, we speak of a pure birth process.

Observe that the generator of a birth and death process satisfies the positive maximum principle. Indeed, if f attains its maximum at k for $0<k<n$, then $f(k)$ dominates both $f(k+1)$ and $f(k-1)$, hence $(b_k+d_k)\,f(k)\geq b_k f(k+1)+d_k f(k-1)$. On the other hand, if f attains its maximum at $k=0$, then $b_0 f(0)\geq b_0 f(1)$. Similarly, if f is maximized at $k=n$, then $d_k f(n)\geq d_k f(n-1)$. In all cases, $Gf(k)\leq 0$.

Let G be the generator of a birth and death process on $\{0,\dots,n\}$. Noting that $\mathcal{C}\left(\{0,\dots,n\}\right)=\mathbb{R}^{\{0,\dots,n\}}$ we can prove that G is the generator of a Feller semigroup. Indeed, its domain is dense (since it is the entire space $\mathbb{R}^{\{0,\dots,n\}}$), it satisfies the positive maximum principle as noted above, $(\lambda-G)^{-1}=\frac{1}{\lambda}\left(I-\frac{G}{\lambda}\right)^{-1}=\frac{1}{\lambda}\sum_{k=0}^{\infty}\left(\frac{G}{\lambda}\right)^{k}$ for $\lambda>\|G\|$, so $\mathcal{R}\left(\lambda-G\right)=\mathbb{R}^{\{0,\dots,n\}}$ for such λ, and $G1=0$.

When G is the generator of a birth and death process on $\mathbb{N}$, the situation is more complicated. We note that when G is unbounded, it does not necessarily map $\mathcal{C}_0(\mathbb{N})$ into itself. For simplicity, we will only deal with generators of pure birth processes on $\mathbb{N}$, as that is all we will need.

Proposition 1.42. *Let G be the generator of a pure birth process on $\mathbb{N}$. Define $D=\{f\in\mathcal{C}_0\left(\mathbb{N}\right);\, Gf\in\mathcal{C}_0\left(\mathbb{N}\right)\}$. Then the restriction of G to D is the generator of a strongly continuous positive contraction semigroup on $\mathcal{C}_0\left(\mathbb{N}\right)$. Moreover, $\mathcal{C}_{\mathrm{K}}\left(\mathbb{N}\right)$ is a core of this generator.*

Proof. First note that whenever f is in $\mathcal{C}_{\mathrm{K}}(\mathbb{N})$ then so is Gf, so $\mathcal{C}_{\mathrm{K}}\left(\mathbb{N}\right)\subset D$. Moreover, $\mathcal{C}_{\mathrm{K}}\left(\mathbb{N}\right)$ is dense in $\mathcal{C}_0\left(\mathbb{N}\right)$.

Let $\lambda>0$ and $g\in\mathcal{C}_{\mathrm{K}}(\mathbb{N})$ be fixed. We will find $f\in\mathcal{C}_{\mathrm{K}}(\mathbb{N})$ such that

$$(\lambda-G)\,f=g. \tag{1.14}$$

This will prove that the range of $(\lambda-G)$ contains $\mathcal{C}_{\mathrm{K}}(\mathbb{N})$ and is therefore dense in $\mathcal{C}_0(\mathbb{N})$.

Let n be such that $g(k) = 0$ for all $k \geq n$. By definition of G, $f \in \mathbb{R}^{\mathbb{N}}$ satisfies (1.14) if and only if

$$f(k+1) \quad = \quad \left(1 + \frac{\lambda}{b_k}\right) f(k) - \frac{g(k)}{b_k}, \tag{1.15}$$

By iterating 1.15 we get

$$f(n) = \prod_{l=0}^{n-1} \left(1 + \frac{\lambda}{b_l}\right) f(0) - \sum_{k=0}^{n-1} \prod_{l=k+1}^{n-1} \left(1 + \frac{\lambda}{b_l}\right) \frac{g(k)}{b_k}.$$

Now we wish to set $f(0)$ such that $f(n) = 0$. This is accomplished by setting

$$f(0) = \sum_{k=0}^{n-1} \prod_{l=0}^{k} \left(1 + \frac{\lambda}{b_l}\right)^{-1} \frac{g(k)}{b_k}.$$

If we now use (1.15) to define f recursively for all $k > 0$, then f solves (1.14). Moreover, $f(n) = 0$, hence $f(k) = 0$ for all $k > n$, so $f \in \mathcal{C}_{\mathrm{K}}(\mathbb{N})$.

Recall that we have shown above that G satisfies the positive maximum principle. Theorem 1.27 now proves that both $G \restriction_{\mathcal{C}_{\mathrm{K}}(\mathbb{N})}$ and $G \restriction_D$ are closable and their closures are the generators of strongly continuous positive contraction semigroups. Corollary 1.30 shows that these semigroups agree, so $\mathcal{C}_{\mathrm{K}}(\mathbb{N})$ is indeed a core of the closure of $G \restriction_D$. To complete the proof, it suffices to prove that $G \restriction_D$ is closed. Let $f_n \in D$ and $f, g \in \mathcal{C}_0(\mathbb{N})$ such that $f_n \to f$ and $Gf_n \to g$ in $\mathcal{C}_0(\mathbb{N})$. Then $f_n \to f$ pointwise, so $Gf_n \to Gf$ pointwise. Therefore $Gf = g$ and $f \in D$ by the definition of D. □

Let G be a generator of a pure birth process on $\mathbb{N}$ and let P_t be its corresponding semigroup. The kernel associated with P_t may be substochastic. The "missing probability" is the probability of explosion. To make P_t into a Feller semigroup we make use of Lemmas 1.18 and 1.32. Let $\bar{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$ be the one-point compactification of $\mathbb{N}$ and for f in $\mathcal{C}\left(\bar{\mathbb{N}}\right)$ define

$$P_t^{\Delta} f = \left(f(\infty) + P_t\left(f \restriction_{\mathbb{N}} - f(\infty)\right)\right)^{\Delta}.$$

Lemma 1.18 shows that P_t is a Feller semigroup and ∞ is an absorption state (see also Lemma 2.3 and Proposition 2.4 in chapter 4 of [3]). Moreover, the generator of P_t is given by (1.9) and its core are functions such that

$$f(k) = f(n);\ k \geq n$$

for sufficiently large n.

We have constructed pure birth processes from their generators using the Hille-Yosida theorem. However, there is also an alternative construction. Let $b_0, b_1, \ldots$ be positive numbers and let $T_0, T_1, \ldots$ be independent exponential random variables with parameters $b_0, b_1, \ldots$ Define

$$\tau_n = \sum_{k=0}^{n} T_k,\ n = 0, 1, \ldots$$

and

$$X_t = \sum_{k=0}^{\infty} \mathbf{1}_{[\tau_k \le t]}.$$

Then X_t is a Feller process and its generator is the generator of a pure birth process with birth rates $b_0, b_1, \ldots$ in the sense of Definition 1.41. Similarly,

$$X_t^{(n)} = \sum_{k=0}^{n-1} \mathbf{1}_{[\tau_k \le t]}$$

is a pure birth process with birth rates $b_0, \ldots b_{n-1}$. Observe that τ_{n-1} is the time it takes the process $X_t^{(n)}$ to reach its absorbing state n, and

$$\tau_\infty = \sum_{k=0}^{\infty} T_k$$

is the time it takes process X_t to reach its absorbing state ∞. For the details of the construction and proofs, see Chapter 2 of Liggett [6]. We note that similar construction using exponential waiting times is possible for more general Markov processes with countable state spaces than just pure birth processes. When the state spaces are finite, the two constructions (i. e. by the Hille-Yosida theorem and by exponential waiting times) are equivalent. But there are Markov processes with countably infinite state spaces that can be constructed by exponential waiting times, but they are not Feller processes, so they cannot be constructed using the Hille-Yosida theorem. Hence for infinite state spaces the two constructions are not equivalent (however, as mentioned above, the two constructions are equivalent for pure birth processes even on countably infinite state spaces).

Chapter 2

Intertwining of Markov processes with discrete state-space

In this chapter, we review the literature on intertwining of Markov processes with discrete state-space. We show results in both discrete and continuous time, as the former is often the basis for the latter. The basic idea is to find two generators G and H and a probability kernel K such that

$$GK = KH. \tag{2.1}$$

From this, we deduce an analogous relation for the associated semigroups, that is

$$P_t K = K Q_t \tag{2.2}$$

for all $t \geq 0$. From this relation, we may deduce distributional properties of the associated Markov processes. Algebraic relations of the type (2.1) and (2.2) are called intertwining relations, which gives the name to the intertwining of Markov processes.

2.1 Discrete time

Intertwining of Markov processes in discrete time and space was studied by Diaconis and Fill [1]. They consider the following situation. Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be countable spaces, P and Q be probability kernels on $\mathcal{S}_1$ and $\mathcal{S}_2$ (interpreted as transition probabilities of Markov processes X_n and Y_n), and K be a probability kernel from $\mathcal{S}_1$ to $\mathcal{S}_2$. Assume that the discrete-time analogue of (2.1) holds, that is,

$$PK = KQ. \tag{2.3}$$

Iterating this equation, we get

$$P^n K = K Q^n, \tag{2.4}$$

which is the discrete-time analogue of (2.2). If π_0^X and π_0^Y are initial distributions that satisfy

$$\pi_0^Y = \pi_0^X K,$$

then the time n distributions satisfy an analogous relationship, namely,

$$\begin{aligned}\pi_n^Y &= \pi_0^Y Q^n \\ &= \pi_0^X K Q^n \\ &= \pi_0^X P^n K \\ &= \pi_n^X K.\end{aligned}$$

This means that

$$\mathrm{P}\left(Y_n = y\right) = \mathrm{E}\left[K\left(X_n, y\right)\right]. \tag{2.5}$$

However, it is possible to prove more. Diaconis and Fill show that it is possible to couple X_n and Y_n in such a way that the conditional distribution of Y_n, given the history of X_n, is given by K, as is shown in the next Theorem.

Theorem 2.1. *Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be countable sets with discrete topology, $\left(\pi_0^X, P\right)$ a distribution and a probability kernel on $\mathcal{S}_1$, and $\left(\pi_0^Y, Q\right)$ a distribution and a probability kernel on $\mathcal{S}_2$, and let K be a probability kernel from $\mathcal{S}_1$ to $\mathcal{S}_2$. Then there exists a Markov chain (X_n, Y_n) with margins $X_n \sim \left(\pi_0^X, P\right)$ and $Y_n \sim \left(\pi_0^Y, Q\right)$ such that*

$$\mathrm{P}\left(Y_n = y_n | X_n = x_n, \ldots, X_0 = x_0\right) = K\left(x_n, y_n\right)$$

for every $n \geq 1$, $y_n \in \mathcal{S}_2$ and almost every $x_0, \ldots, x_n \in \mathcal{S}_1$ w. r. t. law of $(X_0, \ldots, X_n)$, if and only if

$$\begin{aligned}\pi_0^Y &= \pi_0^X K, \\ PK &= KQ.\end{aligned}$$

Remark 2.2. By $X_n \sim \left(\pi_0^X, P\right)$ we mean that, on its own, X_n is a Markov process with the initial distribution π_0^X and the transition probability P. Expression $Y_n \sim \left(\pi_0^Y, Q\right)$ is interpreted analogously.

Proof. See Diaconis and Fill [1], Theorem 2.17. The basic idea for the proof of sufficiency is to define a distribution and a probability kernel on

$$\mathcal{S} = \left\{(x, y) \in \mathcal{S}_1 \times \mathcal{S}_2;\ K\left(x, y\right) > 0\right\}$$

by

$$\boldsymbol{\pi}_0^{(X,Y)}\left(x, y\right) = \pi_0^X(x) K(x, y) \tag{2.6}$$

and

$$\boldsymbol{P}^{(X,Y)}\left(\left(x_1, y_1\right), \left(x_2, y_2\right)\right) = \begin{cases} \frac{Q(y_1,y_2)P(x_1,x_2)K(x_2,y_2)}{\Delta(x_1,y_2)}, & \text{if } \Delta\left(x_1, y_2\right) > 0, \\ 0, & \text{otherwise,}\end{cases} \tag{2.7}$$

where $\Delta = PK = KQ$. To see that $\boldsymbol{\pi}_0^{(X,Y)}$ is indeed a distribution on $\mathcal{S}$, observe that

$$\begin{aligned}\sum_{(x,y)\in\mathcal{S}} \boldsymbol{\pi}_0^{(X,Y)}(x, y) &= \sum_{x\in\mathcal{S}_1} \pi_0^X(x) \sum_{y\in\mathcal{S}_2, K(x,y)>0} K(x, y) \\ &= \sum_{x\in\mathcal{S}_1} \pi_0^X(x) \\ &= 1.\end{aligned}$$

To see that $\boldsymbol{P}^{(X,Y)}$ is a probability kernel on $\mathcal{S}$, let $(x_1, y_1) \in \mathcal{S}$ and $y_2 \in \mathcal{S}_2$ and note that

$$\sum_{x_2 \in \mathcal{S}_1, K(x_2,y_2)>0} \boldsymbol{P}^{(X,Y)}\left((x_1, y_1), (x_2, y_2)\right) = Q\left(y_1, y_2\right) \mathbf{1}_{[\Delta(x_1,y_2)>0]}, \tag{2.8}$$

where we used the fact that

$$\Delta\left(x_1, y_2\right) = \sum_{x_2 \in \mathcal{S}_1} P\left(x_1, x_2\right) K\left(x_2, y_2\right).$$

However, we also have that

$$\begin{aligned}\Delta\left(x_1, y_2\right) &= \sum_{y \in \mathcal{S}_2} K\left(x_1, y\right) Q\left(y, y_2\right) \\ &\geq K\left(x_1, y_1\right) Q\left(y_1, y_2\right).\end{aligned}$$

But (x_1, y_1) is in $\mathcal{S}$, so $K\left(x_1, y_1\right) > 0$. Hence, if $\Delta\left(x_1, y_2\right) = 0$, then $Q\left(y_1, y_2\right) = 0$, which proves that the indicator on the right-hand side of (2.8) can be dropped. Then we get

$$\begin{aligned}\sum_{(x_2,y_2) \in \mathcal{S}} \boldsymbol{P}^{(X,Y)}\left((x_1, y_1), (x_2, y_2)\right) &= \sum_{y_2 \in \mathcal{S}_2} \sum_{x_2 \in \mathcal{S}_1, K(x_2,y_2)>0} \boldsymbol{P}^{(X,Y)}\left((x_1, y_1), (x_2, y_2)\right) \\ &= \sum_{y_2 \in \mathcal{S}_2} Q\left(y_1, y_2\right) \\ &= 1,\end{aligned}$$

which shows that $\boldsymbol{P}^{(X,Y)}$ is indeed a probability kernel. Then Diaconis and Fill prove that the Markov process with the initial distribution $\boldsymbol{\pi}_0^{(X,Y)}$ and the transition probability $\boldsymbol{P}^{(X,Y)}$ has the desired properties (compare this with the proof of Remark 3.17 below). □

Remark 2.3. It is worthwhile to point out that the coupled Markov process is not necessarily unique. That is, there might be other choices for the coupled transition probability than (2.7) such that the coupled process has the desired properties, see Remark 2.23 in Diaconis and Fill [1].

Remark 2.4. Note that the intertwining relationship $PK = KQ$ is not symmetric. Even if K is invertible, K^{-1} is generally not a probability kernel. To express this asymmetry, we say that Y_n is an averaged Markov process on X_n, following the terminology used in Swart [10]. This terminology can be motivated by (2.5), which says that the distribution of Y_n is the average (expectation) of some function of X_n.

The construction used in the proof of Theorem 2.1 created the bivariate Markov chain in one step. However, Diaconis and Fill also give the following sample path construction. In this construction, they first construct the Markov chain Y_n and then for each sample path of Y_n they construct the Markov chain X_n using knowledge of the sample path of Y_n and independent randomness. Specifically, let Y_n be the Markov chain with the initial distribution π^Y and transition probability Q. Define X_0 as the random variable whose conditional distribution is

$$\mathrm{P}\left(X_0 = x_0 | Y_0 = y_0\right) = \frac{\pi_0^X\left(x_0\right) K\left(x_0, y_0\right)}{\pi_0^Y\left(y_0\right)}. \tag{2.9}$$

Assuming that $X_0, \dots, X_{n-1}$ have already been constructed, construct X_n as

$$\mathrm{P}\left(X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_0 = x_0, Y_n = y_n, \dots Y_0 = y_0\right) = \frac{P\left(x_{n-1}, x_n\right) K\left(x_n, y_n\right)}{\Delta\left(x_{n-1}, y_n\right)}. \quad (2.10)$$

They then show that (X_n, Y_n) constructed in this way is a Markov chain with the initial distribution $\boldsymbol{\pi}_0^{(X,Y)}$ of (2.6) and transition probability $\boldsymbol{P}^{(X,Y)}$ of (2.7). Hence, it is a coupled Markov chain in the sense of Theorem 2.1.

2.2 Continuous time

Fill [4] extends the results of Diaconis and Fill [1] to Markov processes with continuous time. However, he still only considers the case of the discrete state-space. Note that Fill uses formalism of Markov chains constructed by exponential waiting times, and not the formalism of Feller processes (see the discussion at the end of Chapter 1). In this theses we are interested mostly in Feller processes (indeed, entire Chapter 3 is concerned with Feller processes). Nevertheless, we use Fill's results only as an inspiration, and all our proofs in Chapter 3 are formally independent from Fill's results presented in this section. Hence, it does not matter that he uses different formalism than we do.

In discrete time, the equivalence of (2.3) and (2.4) was trivial, but in continuous time, the proof is more complicated.

Proposition 2.5. *Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be countable spaces and let K be a probability kernel from $\mathcal{S}_1$ to $\mathcal{S}_2$. Let G and H be generators on $\mathcal{S}_1$ and $\mathcal{S}_2$ and let P_t and Q_t be their associated semigroups. Then*

$$P_t K = K Q_t \quad (2.11)$$

for all $t \geq 0$ if and only if

$$GK = KH. \quad (2.12)$$

Proof. See Lemma 3 and Proposition 2 in Fill [4]. The idea of proving (2.11)⇒(2.12) is to differentiate (2.11). The proof of (2.12)⇒(2.11) uses the Kolmogorov backward equation. □

If we assume (2.11) and that the initial distributions are properly intertwined,

$$\pi_0^Y = \pi_0^X K,$$

then, just as in the discrete time, we can prove that time t distributions are intertwined:

$$\begin{aligned} \pi_t^Y &= \pi_0^Y Q_t \\ &= \pi_0^X K Q_t \\ &= \pi_0^X P_t K \\ &= \pi_t^X K. \end{aligned} \quad (2.13)$$

Again, it is possible find a coupling such that the conditional distribution of Y_t given the history of X_t is given by K.

Theorem 2.6. *Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be finite spaces and let K be a probability kernel from $\mathcal{S}_1$ to $\mathcal{S}_2$. Let π_0^X and P_t be a probability distribution and a semigroup on $\mathcal{S}_1$ and analogously let π_0^Y and Q_t be a probability distribution and a semigroup on $\mathcal{S}_2$. Then there exists a Markov chain (X_t, Y_t) with margins $X_t \sim \left(\pi_0^X, P_t\right)$ and $Y_t \sim \left(\pi_0^Y, Q_t\right)$ satisfying*

$$\mathrm{P}\left(Y_t = y | X_u, 0 \leq u \leq t\right) = K\left(X_t, y\right) \tag{2.14}$$

a. s. for all $t \geq 0$, if and only if

$$\pi_0^Y = \pi_0^X K$$

and

$$P_t K = K Q_t$$

for all $t \geq 0$.

Proof. See Theorem 2 in Fill [4]. We give here only a sketch of the proof of sufficiency. Define a state-space for the bivariate process as

$$\mathcal{S} = \left\{(x, y) \in \mathcal{S}_1 \times \mathcal{S}_2;\ K(x, y) > 0\right\}.$$

Also define a distribution on $\mathcal{S}$ by

$$\boldsymbol{\pi}_0^{(X,Y)}(x, y) = \pi_0^X(x) K(x, y).$$

In analogy with (2.7), for $t > 0$ define a probability kernel on $\mathcal{S}$ by

$$\boldsymbol{P}^{(t)}\left(\left(x_1, y_1\right), \left(x_2, y_2\right)\right) = \begin{cases} \frac{Q_t(y_1,y_2) P_t(x_1,x_2) K(x_2,y_2)}{\Delta^{(t)}(x_1,y_2)}, & \text{if } \Delta^{(t)}(x_1, y_2) > 0, \\ 0, & \text{otherwise,} \end{cases} \tag{2.15}$$

where $\Delta^{(t)} = P_t K$. It turns out, however, that $\boldsymbol{P}^{(t)}$ does not satisfy the Chapman-Kolmogorov equations and hence cannot be used to construct the bivariate process directly. However, Fill proves that there exists a generator $\boldsymbol{G}$ on $\mathcal{S}$ (for which he gives an explicit formula) such that

$$\frac{\boldsymbol{P}^{(t)} - I}{t} \to \boldsymbol{G} \tag{2.16}$$

as $t \to 0$. It is then shown that

$$\lim_{h \to 0} \left(\boldsymbol{P}^{(h)}\right)^{\lfloor t/h \rfloor} = \boldsymbol{P}_t, \tag{2.17}$$

where $\boldsymbol{P}_t$ is the semigroup generated by $\boldsymbol{G}$. Fill then uses (2.15) and (2.17) to prove that the bivariate Markov process associated with $\boldsymbol{G}$ has the desired properties. $\square$

Remark 2.7. Contrary to the discrete-time case, it is not clear if there is more than one $\boldsymbol{G}$ such that the bivariate process generated by $\boldsymbol{G}$ has the desired properties.

Remark 2.8. Note that Theorem 2.6 assumes that the state-spaces are finite. Indeed, Fill is unable to generalize the algebraic proof sketched above to the case that the state-spaces are infinite, because he is unable to prove that (2.17) holds in the infinite case.

Nevertheless, Fill finds a way to prove Theorem 2.6 even for the countably infinite case, although he does not present the entire proof. When the state-spaces are infinite, he makes some simplifying assumptions that are satisfied in all cases he is interested in. Specifically, he assumes that P_t and Q_t define non-explosive Markov chains, there exists exactly one absorbing state $\infty \in \mathcal{S}_1$ such that $K(\infty, \cdot)$ is a stationary distribution for Q_t, and for all $x \in \mathcal{S}_1 \setminus \{\infty\}$ the sets

$$\{y \in \mathcal{S}_2;\ K(x,y) > 0\}$$

are finite. Under these assumptions he observes that the explicit formula which he found for $\boldsymbol{G}$ in the proof for the finite state-spaces continues to define a generator even when the state spaces are infinite. Then he proves that

$$\Lambda \boldsymbol{G} = G\Lambda$$

and

$$\Lambda \boldsymbol{P}_t = P_t\Lambda$$

where Λ is a kernel from $\mathcal{S}_1$ to $\boldsymbol{\mathcal{S}}$ defined by

$$\Lambda\left(\tilde{x}, (x,y)\right) = K(x,y)\mathbf{1}_{[x=\tilde{x}]}$$

(see the proof of Proposition 4 in Fill [4]). He uses this to prove that if $(X_t, Y_t) \sim \left(\boldsymbol{\pi}_0^{(X,Y)}, \boldsymbol{G}\right)$ then $X_t \sim \left(\pi_0^X, P_t\right)$ and (2.14) holds (compare this to the proofs of Remark 3.17 and Theorem 3.9 below). He is unable to prove algebraically that

$$(X_t, Y_t) \sim \left(\boldsymbol{\pi}_0^{(X,Y)}, \boldsymbol{G}\right) \Rightarrow Y_t \sim \left(\pi_0^Y, Q_t\right). \tag{2.18}$$

However, he finds a continuous-time analogy of the sample path construction (2.9) and (2.10). Specifically, for a given Markov process $Y_t \sim \left(\pi_0^Y, Q_t\right)$ he constructs a process X_t such that $(X_t, Y_t) \sim \left(\boldsymbol{\pi}_0^{(X,Y)}, \boldsymbol{G}\right)$. Since the distribution of a Markov process is uniquely determined by its initial distribution and transition probability (or equivalently the generator), this proves (2.18).

Diaconis and Miclo [2] use these results to give a probabilistic representation of the time to absorption for a birth and death process. Specifically, let G be a generator of a birth and death process on $\{0, \dots, n\}$ for some $n \in \mathbb{N}$. Suppose that all birth rates are positive and n is an absorbing state. Diaconis and Miclo construct a generator G^+ of a pure birth process on $\{0, \dots, n\}$ and a probability kernel K^+ such that $G^+K^+ = K^+G$ (i. e. the birth and death process is averaged on the pure birth process). The birth rates of G^+ are the negative eigenvalues of G in descending order. Moreover, the probability kernel is constructed such that

$$K^+\left(x, \{0, \dots, x\}\right) = 1 \tag{2.19}$$

for all $x \in \{0, \dots, n\}$ and

$$K^+(n,n) = 1. \tag{2.20}$$

From Theorem 2.6 they conclude that there exists a bivariate Markov chain (X^+, X) whose margins evolve according to G^+ and G. From the properties of K^+, they conclude that $X_t^+ \geq X_t$ and that X_t^+ and X_t are absorbed at the same time. That is, the time to absorption of a birth and death process can be represented as the time to absorption of a pure birth process, which is equal to the sum of independent exponentially distributed random variables, whose intensities are equal to the birth rates of the pure birth process.[1]It turns out that the distribution intertwining (2.13) is sufficient to derive the distribution of the time to absorption, and the more general Theorem 2.6 is not needed.

Swart [10] extends this result by constructing another generator G^- of a pure birth process and another kernel K^- such that $GK^- = K^-G^-$ (i. e. the pure birth process is averaged on the birth and death process). The birth rates of G^- are again the negative eigenvalues of G, now in ascending order. The probability kernel K^- again satisfies (2.19) and (2.20). By applying Theorem 2.6 twice, he constructs Markov process $\left(X_t^-, X_t, X_t^+\right)$ such that

$$X_t^- \leq X_t \leq X_t^+$$

and all three processes are absorbed at the same time.

[1]The distribution of the time to absorption for birth and death process was known much earlier, see references in [2]. However, the proof given by Diaconis and Miclo is the first probabilistic proof of this fact. Earlier proofs derived the result via analytical methods without probabilistic interpretation.

Chapter 3

Intertwining of diffusions

In this chapter we generalize the results of Chapter 2 to the case when one of the processes is a diffusion and the other is a pure birth process on either $\{0, \ldots, n\}$ or $\mathbb{N}$. Since the general case is too difficult, we restrict ourselves to the case that the diffusion is either the Wright-Fisher diffusion or the Wright-Fisher diffusion with reflection at zero. This restriction is motivated by the fact that the generator and the semigroup of the Wright-Fisher diffusion (with or without reflection) maps polynomials to polynomials (see Lemma 1.36), which simplifies things quite a bit.

In the first part, we describe generator and semigroup intertwining, which allows us to find the distribution of the time to absorption of the Wright-Fisher diffusion. In the second part, we construct a coupling in the spirit of Theorem 2.6. We construct the coupling only for the pure-birth process on $\{0, \ldots, n\}$ and not on $\mathbb{N}$. Moreover, at one point, one of our proofs fails to work for the Wright-Fisher diffusion without reflection, so we are only able to construct the coupling for the Wright-Fisher diffusion with reflection at zero. This is our main motivation for dealing with the Wright-Fisher diffusion with reflection at zero in the first place.

3.1 Generator and semigroup intertwining

As the first step, we generalize Proposition 2.5.

Theorem 3.1. *Let L_1, L_2 be Banach spaces. Let P_t and Q_t be strongly continuous contraction semigroups defined on L_1, L_2 and let G and H be their generators. Let $K : L_2 \to L_1$ be a continuous linear operator. Then the following are equivalent:*

1. *For all $t \geq 0$,*
$$P_t K = K Q_t \tag{3.1}$$
on L_2,

2. *K maps $\mathcal{D}(H)$ into $\mathcal{D}(G)$ and*
$$GK = KH \tag{3.2}$$
on $\mathcal{D}(H)$,

3. *There exists a core D of H such that K maps D into $\mathcal{D}(G)$ and (3.2) holds on D.*

Proof. To prove (1) $\Rightarrow$ (2), fix $f \in \mathcal{D}(H)$. Then $\frac{1}{t}(Q_t f - f)$ converges to Hf, so $\frac{1}{t}(KQ_t f - Kf)$ converges to KHf. By (3.1), $\frac{1}{t}(P_t Kf - Kf)$ is also convergent, so Kf is in $\mathcal{D}(G)$ and $GKf = KHf$.

In order to prove (2) $\Rightarrow$ (1), fix $f \in \mathcal{D}(H)$ and define $u(t) = KQ_t f$. Since $Q_t f$ is in $\mathcal{D}(H)$, $u(t) \in \mathcal{D}(G)$ for all $t \geq 0$. By the continuity of K

$$\frac{\mathrm{d}}{\mathrm{d}t}u(t) = K\frac{\mathrm{d}}{\mathrm{d}t}Q_t f = KHQ_t f = GKQ_t f = Gu(t).$$

Since

$$\frac{\mathrm{d}}{\mathrm{d}t}u(t) = KQ_t Hf,$$

$Gu(t) = \frac{\mathrm{d}}{\mathrm{d}t}u(t)$ is continuous. By the Kolmogorov backward equation (Proposition 1.28), $u(t) = P_t u(0) = P_t Kf$ which proves that (3.1) holds on $\mathcal{D}(H)$. Since all operators involved in (3.1) are continuous, the assertion now follows from the density of $\mathcal{D}(H)$ in L_2.

The implication (2) $\Rightarrow$ (3) is trivial by taking $D = \mathcal{D}(H)$. To prove the converse, let f be in $\mathcal{D}(H)$. Then there exist $f_n \in D$ such that $f_n \to f$ and $Hf_n \to Hf$. Since K is continuous, $Kf_n \to Kf$ and $GKf_n = KHf_n \to KHf$, where we have used (3.2) for f_n. Since G is a closed operator, Kf is in $\mathcal{D}(G)$ and $GKf = KHf$. □

Corollary 3.2. *Under the assumptions of Theorem 3.1, assume further that $L_i = \mathcal{C}_0(\mathcal{S}_i)$ where $\mathcal{S}_i$ are locally compact metric spaces interpreted as state-spaces for random processes. Let X_t and Y_t be processes associated with P_t and Q_t with initial distributions π_0^X and π_0^Y. If the initial distributions satisfy the intertwining relationship*

$$\pi_0^Y = \pi_0^X K,$$

then the time t distributions satisfy

$$\pi_t^Y = \pi_t^X K \tag{3.3}$$

or

$$\mathrm{P}(Y_t \in A) = \mathrm{E}(K(X_t, A)) \tag{3.4}$$

for all $t \geq 0$ and $A \in \mathcal{B}(\mathcal{S}_2)$.

Proof. Write

$$\begin{aligned}
\pi_t^Y &= \pi_0^Y Q_t \\
&= \pi_0^X KQ_t \\
&= \pi_0^X P_t K \\
&= \pi_t^X K.
\end{aligned}$$

□

Now we describe an analogy of the work of Swart [10]. More precisely the birth and death process X_t of Section 2.2 is replaced by the Wright-Fisher diffusion (with or without reflection) and we construct a pure birth process Y_t which is an averaged process on X_t, analogous to the process X_t^- of Swart [10]. Moreover, we would like to construct Y_t such that it is absorbed or explodes at the same time as X_t is absorbed.

Let $n \in \mathbb{N}$ be given and define a probability kernel from $[-1,1]$ to $\{0,\dots,n\}$ by

$$K_n(x,k) = \begin{cases} (1-x^2)\,x^{2k}, & \text{if } 0 \le k < n, \\ x^{2n}, & \text{if } k = n. \end{cases} \tag{3.5}$$

Define the generator of a pure birth process on $\{0,\dots,n\}$ by

$$H_n f(k) = \begin{cases} \lambda_k \left(f(k+1) - f(k)\right) & 0 \le k < n, \\ 0, & k = n, \end{cases} \tag{3.6}$$

where

$$\lambda_k = (2k+1)\,(2k+2)\,,\ 0 \le k < n.$$

Lemma 3.3. *Let $K = K_n$ of (3.5), $H = H_n$ of (3.6) and let $G = G^{WF}$ of (1.10) be the generator of the Wright-Fisher diffusion. Then K maps $\mathcal{C}(\{0,\dots,n\})$ to $\mathcal{C}^2[-1,1]$ and $GK = KH$ on $\mathcal{C}(\{0,\dots,n\})$.*

Proof. Since $K(\cdot,k) \in \mathcal{C}^2[-1,1]$ for $0 \le k \le n$, it follows from linearity that the range of K must also be in $\mathcal{C}^2[-1,1]$. Let x be in $[-1,1]$. For $0 \le k < n$ we have

$$\begin{aligned} GK\mathbf{1}_{\{k\}}(x) &= \left(G\left(\left(1-x^2\right)x^{2k}\right)\right)(x) \\ &= \left(1-x^2\right)\frac{\partial^2}{\partial x^2}\left(x^{2k} - x^{2k+2}\right) \\ &= \left(1-x^2\right)\left(2k(2k-1)x^{2k-2} - (2k+2)(2k+1)x^{2k}\right) \end{aligned}$$

and

$$\begin{aligned} KH\mathbf{1}_{\{k\}}(x) &= K\left(\lambda_{k-1}\mathbf{1}_{\{k-1\}} - \lambda_k\mathbf{1}_{\{k\}}\right)(x) \\ &= \left(1-x^2\right)\left(2k(2k-1)x^{2k-2} - (2k+2)(2k+1)x^{2k}\right). \end{aligned}$$

For $k = n$ we have

$$\begin{aligned} GK\mathbf{1}_{\{n\}}(x) &= G\left(x^{2n}\right)(x) \\ &= \left(1-x^2\right)\frac{\partial^2}{\partial x^2}x^{2n} \\ &= 2n(2n-1)\left(1-x^2\right)x^{2n-2} \end{aligned}$$

and

$$\begin{aligned} KH\mathbf{1}_{\{n\}}(x) &= K\left(\lambda_{n-1}\mathbf{1}_{\{n-1\}}\right)(x) \\ &= 2n(2n-1)\left(1-x^2\right)x^{2n-2}. \end{aligned}$$

The assertion of the lemma now follows from linearity. □

Remark 3.4. Observe that K_n, when restricted to $[0,1]\times\{0,\dots,n\}$ is a probability kernel from $[0,1]$ to $\{0,\dots,n\}$. A simple modification of the proof of Lemma 3.3 shows that it also holds if we replace the generator of the Wright-Fisher diffusion by the generator of the Wright-Fisher diffusion with reflection at zero (defined by (1.12)) and if we interpret K_n as a probability kernel form $[0,1]$ to $\{0,\ldots,n\}$. To see this, note that $\frac{\partial}{\partial x}K_n(0,k)=0$, so K_n maps $\mathcal{C}\{0,\ldots,n\}$ into $\mathcal{D}\left(G^{WF,r}\right)$. Since the formulas defining G^{WF} and $G^{WF,r}$ are the same, the proof showing that $G^{WF}K_n = K_nH_n$ also works for $G^{WF,r}$. In the remainder of this section, we will formulate all our assertions in terms of the Wright-Fisher diffusion (without reflection), but it is easy to modify their proofs to show that they also hold for the Wright-Fisher diffusion with reflection at zero.

For the Wright-Fisher diffusion, we have found a pure birth process on $\{0,\ldots,n\}$ and a probability kernel that are properly intertwined in the sense of equation (3.2). We would like it if the time to absorption of the diffusion had the same distribution as the time to absorption of the pure birth process. This would be guaranteed by (3.4) if $K_n(x,n)$ were equal to $\mathbf{1}_{[x\in\{-1,1\}]}$. Unfortunately, this is not the case, so we take the limit as n approaches infinity. Specifically, define the generator of a pure birth process on $\mathbb{N}$ by

$$H_\infty f(k)=\lambda_k\left(f(k+1)-f(k)\right)\ 0\leq k<\infty \tag{3.7}$$

and use construction of Lemmas 1.18 and 1.32 to extend it to a generator H_∞^Δ on $\bar{\mathbb{N}}$. Define a probability kernel from $[-1,1]$ to $\bar{\mathbb{N}}$ by

$$K_\infty(x,k)=\begin{cases}\left(1-x^2\right)x^{2k}, & 0\leq k<\infty,\\ \mathbf{1}_{[x\in\{-1,1\}]}, & k=\infty.\end{cases} \tag{3.8}$$

Lemma 3.5. *K_∞ is a continuous probability kernel from $[-1,1]$ to $\bar{\mathbb{N}}$.*

Proof. Since $K_\infty(x,k)\geq 0$ and $\sum_{k=0}^\infty K_\infty(x,k)+K_\infty(x,\infty)=1$, $K(x,\cdot)$ is a probability measure for every $x\in[-1,1]$. $K_\infty(\cdot,k)$ is clearly measurable (for $0\leq k<\infty$ it is continuous and for $k=\infty$ it is an indicator of a closed set). Hence K_∞ is a probability kernel from $[-1,1]$ to $\bar{\mathbb{N}}$. To finish the proof, it suffices to show that for every $f\in\mathcal{C}\left(\bar{\mathbb{N}}\right)$, $K_\infty f$ is in $\mathcal{C}[-1,1]$.

Let f be in $\mathcal{C}\left(\bar{\mathbb{N}}\right)$. For $x\in(-1,1)$ we have

$$\begin{aligned}K_\infty f(x) &= \sum_{k=0}^\infty\left(1-x^2\right)x^{2k}f(k) \\ &= \sum_{k=0}^\infty f(k)x^{2k}-\sum_{k=0}^\infty f(k)x^{2k+2}.\end{aligned} \tag{3.9}$$

Since $K_\infty f$ can be expressed as a difference of two power series, it is continuous within the radius of convergence, which is at least one, because f is bounded. It is thus sufficient to prove that $K_\infty f$ is continuous in ±1. Fix $\epsilon>0$ and let n be such that $|f(k)-f(\infty)|<\epsilon$ for all $k>n$. For x in $(-1,1)$ we have

$$\begin{aligned}|K_\infty f(x)-f(\infty)| &\leq \sum_{k=0}^\infty\left(1-x^2\right)x^{2k}|f(k)-f(\infty)| \\ &\leq 2\|f\|\left(1-x^{2n}\right)+\sum_{k=n+1}^\infty\left(1-x^2\right)x^{2k}\epsilon \\ &\leq 2\|f\|\left(1-x^{2n}\right)+\epsilon.\end{aligned}$$

Passing to the limit we get

$$\limsup_{x\to\pm 1} |K_\infty f(x) - f(\infty)| \le \epsilon.$$

Since ϵ was arbitrary, $K_\infty f$ is continuous. $\square$

Just as in the case of K_n, by restricting its domain, we may interpret K_∞ as a continuous probability kernel from $[0,1]$ to $\bar{\mathbb{N}}$.

Theorem 3.6. *Let*

$$D = \left\{ f \in \mathcal{C}\left(\bar{\mathbb{N}}\right); \exists n \,\mathrm{s.\,t.}\, \forall k \ge n : f(k) = f(n) \right\},$$

let $H = H^{\Delta}_{\infty}$, $K = K_\infty$ and let $G = G^{WF}$ of (1.10) be the generator of the Wright-Fisher diffusion. Then D is a core of H and K maps D into $\mathcal{D}(G)$. Moreover, $GK = KH$ on D.

Proof. The fact that D is a core follows from Proposition 1.42 and the discussion following it. Fix f in D and let n be such that $f(k) = f(n)$ for all $k > n$. Then for $x \in (-1,1)$,

$$\begin{aligned} Kf(x) &= \left(1 - x^2\right) \sum_{k=0}^{n-1} x^{2k} f(k) + x^{2n} f(n) \\ &= f(0) + \sum_{k=1}^{n} x^{2k} \left(f(k) - f(k-1)\right). \end{aligned} \tag{3.10}$$

As x approaches ± 1, $Kf(x)$ approaches $f(n)$. Now

$$K_\infty f(\pm 1) = f(\infty) = f(n).$$

Hence (3.10) holds also for $x = \pm 1$, and therefore $Kf(x)$ is in $\mathcal{C}^\infty[-1,1] \subseteq \mathcal{D}(G)$. Also, for $x \in [-1,1]$,

$$GKf(x) = \left(1 - x^2\right) \sum_{k=0}^{n-1} \lambda_k x^{2k} \left(f(k+1) - f(k)\right).$$

Now for $k < n$

$$Hf(k) = \lambda_k \left(f(k+1) - f(k)\right),$$

and for $k \ge n$,

$$Hf(k) = 0,$$

hence for $x \in [-1,1]$,

$$KHf(x) = \left(1 - x^2\right) \sum_{k=0}^{n-1} \lambda_k x^{2k} \left(f(k+1) - f(k)\right).$$

$\square$

We summarize our findings in the following Theorem.

Theorem 3.7. *Let $\bar{G}^{WF}$ be the generator of the Wright-Fisher diffusion on $[-1,1]$ with semigroup P_t^{WF} (recall equation (1.10) and Lemma 1.34) and let K_n be the probability kernel from $[-1,1]$ to $\{0,\dots,n\}$ defined by (3.5). Let H_n be the generator of a pure-birth process in $\{0,\dots,n\}$ defined in (3.6) and denote by $Q_t^{(n)}$ its associated semigroup. Then*

$$P_t^{WF} K_n = K_n Q_t^{(n)} \tag{3.11}$$

on $\mathcal{C}(\{0,\dots,n\})$ for all $t \geq 0$. Moreover, let K_∞ be the probability kernel from $[-1,1]$ to $\bar{\mathbb{N}}$ defined by 3.8 and let H_∞^Δ be the generator of a pure-birth process on $\bar{\mathbb{N}}$ constructed from H_∞ of (3.7) by the means of Lemmas 1.18 and 1.32. Denote by $Q_t^{(\infty)}$ the semigroup associated with H_∞^Δ. Then

$$P_t^{WF} K_\infty = K_\infty Q_t^{(\infty)} \tag{3.12}$$

on $\mathcal{C}\left(\bar{\mathbb{N}}\right)$ for all $t \geq 0$.

Proof. Lemma 3.3 shows that K_n maps $\mathcal{C}(\{0,\dots,n\})$ into $\mathcal{D}\left(\bar{G}^{WF}\right)$ and that

$$\bar{G}^{WF} K_n = K_n H_n$$

on $\mathcal{C}(\{0,\dots,n\})$. Equality (3.11) now follows from Theorem 3.1. Theorem 3.6 shows that there exists a core D of H_∞^Δ such that K_∞ maps D into $\mathcal{D}\left(\bar{G}^{WF}\right)$ and

$$\bar{G}^{WF} K_\infty = K_\infty H_\infty^\Delta$$

on D. Equality (3.12) follows again from Theorem 3.1. □

We can use the intertwining relation that we have just found to describe the distribution of the time to absorption for the Wright-Fisher diffusion.

Theorem 3.8. *Let P_t^{WF} and $Q_t^{(\infty)}$ be as in Theorem 3.7. Let X_t be the Markov process with continuous sample paths associated with P_t^{WF}. Define τ as the time to absorption*

$$\tau = \inf\{t \geq 0; X_t = \pm 1\}.$$

Let Y_t be the Markov process with right-continuous sample paths associated with $Q_t^{(\infty)}$. Define σ as the time to explosion

$$\sigma = \inf\{t \geq 0; Y_t = \infty\}.$$

Assume that the initial distributions are properly intertwined:

$$\pi_0^Y = \pi_0^X K_\infty. \tag{3.13}$$

Then

$$\tau \stackrel{d}{=} \sigma.$$

In particular, if X_t starts from zero, then

$$\tau \stackrel{d}{=} \sum_{k=0}^{\infty} T_k \tag{3.14}$$

where $T_0, T_1, \dots$ are independent random variables such that T_k is exponentially distributed with intensity λ_k.

Remark. The distribution of τ given by (3.14) is an old result, see the work cited in [2].

Proof. From Theorem 3.7 and Corollary 3.2 we have that

$$\mathrm{P}\left(Y_t = k\right) = \mathrm{E}\left(K_\infty\left(X_t, k\right)\right).$$

Now we can calculate the distribution of σ in terms of X_t:

$$\begin{aligned} \mathrm{P}\left(\sigma \le t\right) &= \mathrm{P}\left(Y_t = \infty\right) \\ &= \mathrm{E}\left(K_\infty\left(X_t, \infty\right)\right) \\ &= \mathrm{P}\left(X_t = \pm 1\right) \\ &= \mathrm{P}\left(\tau \le t\right). \end{aligned}$$

We conclude that $\sigma \overset{d}{=} \tau$. If X_t starts from zero, then the appropriate initial distribution of Y_t so that (3.13) is satisfied, is given by

$$\begin{aligned} \pi_0^X K_\infty &= \delta_0 K_\infty \\ &= K_\infty\left(0, \cdot\right) \\ &= \delta_0 \end{aligned}$$

and it is well known that the distribution of the time to explosion of a pure birth process starting from zero is given by (3.14). □

Note that we have found the intertwining such that the pure birth process is averaged on the diffusion (see Remark 2.4). This is an analogue to the result in Swart [10] discussed at the end of Chapter 2. It is perhaps also possible to find an intertwining such that the diffusion is averaged on the pure birth process (an analogy of the result in Diaconis and Miclo [2], also discussed at the end of Chapter 2), but we do not pursue it here.

3.2 Coupled process

Let P_t and Q_t be Feller semigroups, G and H their generators and X_t and Y_t the associated Markov processes. Additionally, let K be a probability kernel. In Corollary 3.2 we have shown that if $P_t K = K Q_t$ for all $t \ge 0$ and the initial distributions satisfy

$$\pi_0^Y = \pi_0^X K,$$

then

$$\pi_t^Y = \pi_t^X K. \tag{3.15}$$

However, as described in Chapter 2, in discrete state-space it is possible to do more. It is possible to construct a coupled Markov process (X_t, Y_t) such that its margins evolve according to P_t and Q_t and its distribution satisfies (2.14) (see Theorem 2.6). Here, we generalize this result to the case where X_t is a diffusion. We shall devote the rest of this section to proving the following result.

Theorem 3.9. *There exists a generator* $\boldsymbol{G}$ *of a Markov process* (X_t, Y_t) *with the state-space* $[0,1] \times \{0,\dots,n\}$ *with the following properties. For* $f \in \mathcal{C}(\{0,\dots,n\})$ *and* $s,t \geq 0$,

$$\mathrm{E}\left[f\left(Y_{s+t}\right) \mid \left(X_u, Y_u\right), 0 \leq u \leq s\right] = Q_t f\left(Y_s\right), \tag{3.16}$$

where Q_t *is the semigroup generated by* $H = H_n$ *of (3.6). Hence* Y_t *on its own is a Markov process with generator* H*.*

If additionally the initial distribution $\boldsymbol{\pi}_0^{(X,Y)}$ *satisfies*

$$\boldsymbol{\pi}_0^{(X,Y)}\left(A \times \{y\}\right) = \int_A K(x,y) \pi_0^X(\mathrm{d}x)$$

where $K = K_n$ *is the probability kernel defined by (3.5) and* π_0^X *is an arbitrary probability distribution on* $[0,1]$*, then*

$$\mathrm{P}\left(Y_t = k | X_s, 0 \leq s \leq t\right) = K\left(X_t, k\right) \tag{3.17}$$

a. s. for all $t \geq 0$ *and* $k = 0,\dots,n$*, and* X_t *on its own is the Wright-Fisher diffusion with reflection at zero (generated by* $G^{WF,r}$ *defined by (1.12)) with the initial distribution* π_0^X*.*

Note that Theorem 3.9 is stated for the Wright-Fisher diffusion with reflection at zero. Although we believe that an analogous result also holds for the Wright-Fisher diffusion without reflection at zero, we are unable to prove it. The reason for the complications with the diffusion without reflection is that $K(0,k)$ is zero for $k > 0$. Hence, if (3.17) is to hold, X_t cannot be zero after Y_t departs from zero. Later while proving Theorem 3.9 we will find that when $Y_t = k$, X_t behaves like the Wright-Fisher diffusion with an additional drift which depends on k. This drift is infinite at zero when $k > 0$, in accordance with our observation that X_t cannot be zero in this case. Intuitively this means that if we were to prove Theorem 3.9 for the Wright-Fisher diffusion without reflection at zero, the dynamics of X_t would be as if it was only one diffusion on the entire interval $[-1,1]$ when $Y_t = 0$, but when $Y_t > 0$, the dynamics of X_t would change, and it would behave as two independent diffusions on $[-1,0]$ and $[0,1]$ (since the infinite drift at zero would not permit X_t to cross zero). It would then become difficult to link the two independent diffusions for $Y_t > 0$ with the single diffusion on the entire interval for $Y_t = 0$. This intuitive interpretation of why our proof fails for the diffusion without reflection at zero is stated more precisely in Remark 3.27.

Also note that Theorem 3.9 is stated only for the pure birth process with finite state-space. This simplifies the situation, because $[0,1] \times \{0,\dots,n\}$ is compact, while $[0,1] \times \mathbb{N}$ is not. Moreover, the compactification that we would find natural, $[0,1] \times (\mathbb{N} \cup \{\infty\})$ may not be appropriate, because $K_\infty(x,\infty) > 0$ only if $x = 1$, which could lead to problems because by (3.17), $[0,1) \times \{\infty\}$ would be unreachable. Moreover, whatever compactification we would choose, the topology would be more complicated than in the finite case. Since $\{0,\dots,n\}$ has discrete topology, $f : [0,1] \times \{0,\dots,n\} \to \mathbb{R}$ is continuous if and only if $f(\cdot,k)$ is continuous for all $k = 0,\dots,n$. However, if f were a mapping from some compactification of $[0,1] \times \mathbb{N}$ to $\mathbb{R}$, we would additionally need restrictions on the behavior of limits as $k \to \infty$ in order to make sure that f is continuous. We deal only with the finite case to avoid these complications.

We shall prove Theorem 3.9 by imitating the proofs used for the discrete state-space. In Theorem 3.10 we find an analogy to the probability kernel $\boldsymbol{P}^{(t)}$ from the proof of Theorem 2.6. Then in Lemma 3.18 we find the operator

$$\boldsymbol{G} = \lim_{t \to 0} \frac{1}{t} \left(\boldsymbol{P}^{(t)} - I \right)$$

and in Theorem 3.20 we prove that $\boldsymbol{G}$ generates a Feller semigroup. Finally, we apply Theorem 3.28, which was proved by Rogers and Pitman [7]. This theorem states that under certain conditions, a function of a Markov process is a Markov process, and it also states the relation between the original and the transformed Markov process. Theorem 2.6 and Lemma 3.18 work also for the diffusion without reflection, but the proof of Theorem 3.20 fails for this case as explained by Remark 3.27.

We start by reviewing the probability kernel $\boldsymbol{P}^{(t)}$ of (2.15). Since the state-space was discrete in Chapter 2, we were able to define the probability kernel in (2.15) as a function from $\mathcal{S}_1 \times \mathcal{S}_2 \times \mathcal{S}_1 \times \mathcal{S}_2$ to $[0,1]$ (where $\mathcal{S}_1$ and $\mathcal{S}_2$ are discrete state-spaces, recall the setting of Theorem 2.6). In order to find an analogous kernel for the continuous state-space, we need to reformulate (2.15) as an operator on functions on $\mathcal{S}_1 \times \mathcal{S}_2$. For a function $f : \mathcal{S}_1 \times \mathcal{S}_2 \to \mathbb{R}$, we multiply both sides of (2.15) by $f(x_2, y_2)$ and then sum over (x_2, y_2) to get

$$\begin{aligned} &\boldsymbol{P}^{(t)} f(x_1, y_1) \\ &\quad = \sum_{y_2 \in \mathcal{S}_2} Q_t(y_1, y_2) \mathbf{1}_{\left[\Delta^{(t)}(x_1, y_2) \neq 0\right]} \frac{\sum_{x_1 \in \mathcal{S}_1} P_t(x_1, x_2) f(x_2, y_2) K(x_2, y_2)}{\Delta^{(t)}(x_1, y_2)} \end{aligned} \tag{3.18}$$

where

$$\begin{aligned} \Delta^{(t)}(x_1, y_2) &= \sum_{x_2 \in \mathcal{S}_1} P_t(x_1, x_2) K(x_2, y_2) \\ &= (P_t K(\cdot, y_2))(x_1) \\ &= (P_t K)(x_1, y_2). \end{aligned}$$

In this expression, $P_t K$ may be viewed as the composition of the two kernels. Alternatively, K may be viewed not as a kernel, but simply as a function of two variables. Then $P_t K$ may be viewed as the application of the kernel P_t to this function. Note that although P_t operates on functions of only one variable, we may extended it to operate on functions of two variables by fixing the redundant variable.[1] Similarly, the numerator of (3.18) can be written as

$$(P_t(fK))(x_1, y_2) = (P_t f(\cdot, y_2) K(\cdot, y_2))(x_1),$$

where by fK we mean the pointwise product of the two functions (here K must be viewed as a function of two variables and not as a kernel), and $P_t(fK)$ denotes the application of the semigroup on the product, with the same caveat that the redundant variable is fixed. Then we can rewrite (3.18) as

$$\boldsymbol{P}^{(t)} f = Q_t \mathbf{1}_{[P_t K \neq 0]} \frac{P_t f K}{P_t K}. \tag{3.19}$$

[1] Probabilistically, this means that we extend the kernel P_t to the kernel on $\mathcal{S}_1 \times \mathcal{S}_2$ by defining $P_t((x_1, y_1), (x_2, y_2)) = P_t(x_1, x_2) \mathbf{1}_{[y_1 = y_2]}$. For the sake of readability, we will use the same symbol P_t both for the kernel on $\mathcal{S}_1$ and for the kernel on $\mathcal{S}_1 \times \mathcal{S}_2$.

Here, $\mathbf{1}_{[P_tK\neq 0]}\frac{P_tfK}{P_tK}$ denotes the pointwise division of P_tfK and P_tK, which is then multiplied by the indicator $\mathbf{1}_{[P_tK\neq 0]}$. Q_t is then applied to the resulting function. Although this is a function of two variables and Q_t operates on functions of only one variable, we interpret (3.19) by fixing the redundant variable. Note that (3.19) is not a suitable definition for the case that state-spaces are continuous, because we need to make sure that the resulting function is continuous, but the indicator $\mathbf{1}_{[P_tK\neq 0]}$ can introduce discontinuity. To get around this problem, we shall prove that $\frac{P_tfK}{P_tK}$ can be extended to a continuous function, and then define $\boldsymbol{P}^{(t)}$ by

$$\boldsymbol{P}^{(t)}f = Q_t\frac{P_tfK}{P_tK}. \tag{3.20}$$

Theorem 3.10. *Let $P_t = P_t^{WF,r}$ be the Wright-Fisher semigroup with reflection at zero and let $K = K_n$ be as defined by (3.5). Let f be in $\mathcal{C}\left([0,1]\times\{0,\ldots,n\}\right)$ and $t>0$. Then the function*

$$(x,k)\mapsto\frac{\left(P_tf\left(\cdot,k\right)K\left(\cdot,k\right)\right)(x)}{\left(P_tK\left(\cdot,k\right)\right)(x)},\ (x,k)\in D, \tag{3.21}$$

where

$$D=\left\{(x,k)\in[0,1]\times\{0,\ldots,n\}\,;\,\left(P_tK\left(\cdot,k\right)\right)(x)>0\right\},$$

can be uniquely extended to a continuous function on $[0,1]\times\{0,\ldots,n\}$, which we denote by

$$\frac{P_tfK}{P_tK}(x,k).$$

Moreover, the operator on $\mathcal{C}\left([0,1]\times\{0,\ldots,n\}\right)$ defined by

$$f\mapsto\frac{P_tfK}{P_tK}$$

is a continuous probability kernel.

Before we can prove Theorem 3.10, we need several lemmas. In order to prove the existence of the continuous extension of (3.21) it is enough to prove the existence of the continuous extension for every fixed k, since we assume the discrete topology on $\{0,\ldots,n\}$. Therefore, the following lemmas assume that k is fixed. To simplify notation, they also assume that f is in $\mathcal{C}[0,1]$, so we can write $f(\cdot)$ instead of $f(\cdot,k)$.

First, let us show that whenever the denominator of (3.21) is zero, then so is the numerator.

Lemma 3.11. *Under the assumptions of Theorem 3.10, let $0\leq k\leq n$, $t>0$ and $f\in\mathcal{C}[0,1]$. Then*

$$\left|\left(P_tf\left(\cdot\right)K\left(\cdot,k\right)\right)(x)\right|\leq\|f\|\left|\left(P_tK\left(\cdot,k\right)\right)(x)\right|$$

for all $x\in[0,1]$ and therefore

$$\left\|P_tf\left(\cdot\right)K\left(\cdot,k\right)\right\|\leq\|f\|\left\|P_tK\left(\cdot,k\right)\right\|.$$

Proof. Since P_t is a positive operator and $K(\cdot,k) \geq 0$,

$$(P_t f(\cdot) K(\cdot,k))(x) \leq \|f\| (P_t K(\cdot,k))(x)$$

and

$$(P_t f(\cdot) K(\cdot,k))(x) \geq -\|f\| (P_t K(\cdot,k))(x),$$

hence

$$|(P_t f(\cdot) K(\cdot,k))(x)| \leq \|f\| |(P_t K(\cdot,k))(x)|.$$

□

Now we will find all $(x,k) \in [0,1] \times \{0,\ldots,n\}$ such that the denominator of (3.21) is zero.

Lemma 3.12. *Under the assumptions of Theorem 3.10 let $0 \leq k < n$, $t > 0$ and $x \in [0,1]$. Then*

$$(P_t K(\cdot,k))(x) = 0$$

if and only if $x = 1$. Moreover

$$(P_t K(\cdot,n))(x) > 0$$

for all $x \in [0,1]$.

Proof. Since $K(\cdot,n) > 0$, it follows that $(P_t K(\cdot,n))(x) > 0$ for all x. Moreover, since $Gf(1) = 0$ for all $f \in \mathcal{D}(G)$, 1 is an absorbing point by Corollary 1.29, hence $(P_t K(\cdot,k))(1) = K(1,k) = 0$ for $0 \leq k < n$. Finally, to prove that $P_t K(x,k)$ is positive for $0 \leq k < n$ and $x \in [0,1)$, it suffices to prove that

$$P_t(x,(0,1)) > 0,\ x \in [0,1), \tag{3.22}$$

since $K(\cdot,k)$ is positive on $(0,1)$ for $0 \leq k < n$. Define

$$\begin{aligned} f(x) &= \left(1-x^2\right)x^2 \\ &= x^2 - x^4. \end{aligned}$$

If we prove $P_t f(x) > 0$ on $[0,1)$, then (3.22) will be proved.

By solving the Kolmogorov backward equations as in the proof of Lemma 1.36, we can show that for $x \in [0,1)$ and $t > 0$,

$$\begin{aligned} P_t f(x) &= \frac{1}{5}e^{-2t}\left(1-x^2\right) - \frac{1}{5}e^{-12t}\left(1-6x^2+5x^4\right) \\ &\geq \frac{1}{5}\left(1-x^2\right)\left(e^{-2t}-e^{-12t}\right) \\ &> 0, \end{aligned}$$

where we have used that $1-6x^2+5x^4 \leq 1-x^2$ since $x^4 \leq x^2$ for $x \in [0,1)$. □

Since the denominator of (3.21) can be zero only if $k < n$, it is sufficient to consider only this case. The next two lemmas show us the structure of the denominator of (3.21).

Lemma 3.13. *Under the assumptions of Theorem 3.10, let $0 \le k < n$ and $t > 0$. Then*

$$\frac{\partial}{\partial x}\left(P_t K\left(\cdot,k\right)\right)(x)\bigg|_{x=1} < 0.$$

If additionally $k < n-1$, then

$$\lim_{x\to 1}\frac{\left(P_t K\left(\cdot,k+1\right)\right)(x)}{\left(P_t K\left(\cdot,k\right)\right)(x)} > 0.$$

Proof. Let $0 \le k < n$ and $t > 0$. Let Q_t be the semigroup whose generator is $H = H_n$ defined by (3.6). Observe that for $x \in [0,1]$

$$\begin{aligned}
\left(P_t K\left(\cdot,k\right)\right)(x) &= P_t K \mathbf{1}_{\{k\}}(x) \\
&= K Q_t \mathbf{1}_{\{k\}}(x) \\
&= \sum_{l=0}^{n} K(x,l) Q_t(l,k) \\
&= \sum_{l=0}^{k}\left(x^{2l} - x^{2l+2}\right) Q_t(l,k)
\end{aligned}$$

where we have used that $Q_t(l,k) = 0$ for $l > k$. Differentiating with respect to x we get

$$\frac{\partial}{\partial x}\left(P_t K\left(\cdot,k\right)\right)(x)\bigg|_{x=1} = \sum_{l=0}^{k} -2Q_t(l,k).$$

Since $Q_t(l,k) > 0$ for $l \le k$, the first assertion is proved.

Suppose now that $0 \le k < n-1$ and $t > 0$. Since $\left(P_t K\left(\cdot,k\right)\right)(1)$ and $\left(P_t K\left(\cdot,k+1\right)\right)(1)$ are zero by Lemma 3.12, we can use l'Hospital's rule to conclude that

$$\begin{aligned}
\lim_{x\to 1}\frac{\left(P_t K\left(\cdot,k+1\right)\right)(x)}{\left(P_t K\left(\cdot,k\right)\right)(x)} &= \lim_{x\to 1}\frac{\frac{\partial}{\partial x}\left(P_t K\left(\cdot,k+1\right)\right)(x)}{\frac{\partial}{\partial x}\left(P_t K\left(\cdot,k\right)\right)(x)} \\
&> 0.
\end{aligned}$$

□

Lemma 3.14. *Under the assumptions of Theorem 3.10, let $0 \le k < n$ and $t > 0$. Then*

$$p_k(x) = \left(P_t K\left(\cdot,k\right)\right)(x),\ x \in [0,1]$$

is a polynomial whose only root in $[0,1]$ is 1 and it is a simple root. That is, there exists a polynomial u_k such that $u_k > 0$ on $[0,1]$ and

$$p_k(x) = (1-x)\, u_k(x),\ x \in [0,1]. \tag{3.23}$$

Proof. Lemma 1.36 shows that p_k is a polynomial. Since by Lemma 3.12 $p_k(1) = 0$ and $p_k > 0$ on $[0,1)$, the only root in $[0,1]$ is 1. That is, there exists a polynomial u_k such that $u_k > 0$ on $[0,1)$ and (3.23) holds. We will prove that $u_k(1) > 0$ by induction in k.

For $k = 0$, it can be verified by solving the Kolmogorov backward equations as in Lemma 1.36 that $\left(P_t K\left(\cdot,0\right)\right)(x) = e^{-2t}\left(1-x^2\right)$, hence $u_0(x) = e^{-2t}(1+x)$, so $u_0(1) > 0$.

Let us now assume that we have proved that $u_k(1) > 0$ and we wish to prove that $u_{k+1}(1) > 0$ for $0 \leq k < n-1$. Observe that

$$\frac{u_{k+1}(1)}{u_k(1)} = \lim_{x \to 1} \frac{p_{k+1}(x)}{p_k(x)}$$

which is positive Lemma 3.13. □

Given that the root of $(P_t K(\cdot, k))(x)$ is simple, it is be possible to cancel it, as the next lemma shows.

Lemma 3.15. *Under the assumptions of Theorem 3.10, let $0 \leq k < n$, $t > 0$ and let p be a polynomial. Then the function*

$$x \mapsto \frac{(P_t p(\cdot) K(\cdot, k))(x)}{(P_t K(\cdot, k))(x)},\ x \in [0, 1)$$

can be extended to a continuous function on $[0, 1]$.

Proof. Lemma 1.36 shows that both $(P_t p(\cdot) K(\cdot, k))(x)$ and $(P_t K(\cdot, k))(x)$ are polynomials in x. Lemma 3.14 shows that

$$(P_t K(\cdot, k))(x) = (1 - x)\, u(x)$$

where u is a polynomial such that $u > 0$ on $[0, 1]$. Lemma 3.11 then shows that 1 is a root of $(P_t p(\cdot) K(\cdot, k))(x)$, so

$$(P_t p(\cdot) K(\cdot, k))(x) = (1 - x)\, v(x)$$

where v is a polynomial. Then for $x \in [0, 1)$ we have

$$\frac{(P_t p(\cdot) K(\cdot, k))(x)}{(P_t K(\cdot, k))(x)} = \frac{v(x)}{u(x)}$$

which is a continuous function on $[0, 1]$. □

Now we extend this result to arbitrary continuous functions.

Lemma 3.16. *Under the assumptions of Theorem 3.10, let $0 \leq k < n$ and $t > 0$. Then for every f in $\mathcal{C}[0, 1]$, the function*

$$x \mapsto \frac{(P_t f(\cdot) K(\cdot, k))(x)}{(P_t K(\cdot, k))(x)},\ x \in [0, 1)$$

can be uniquely extended to a continuous function on $[0, 1]$. Moreover, the linear operator on $\mathcal{C}[0, 1]$ given by

$$f \mapsto \frac{P_t f(\cdot) K(\cdot, k)}{P_t K(\cdot, k)}$$

is a continuous probability kernel on $[0, 1]$.

Proof. For a polynomial $p \in \mathcal{C}[0,1]$, denote by μp the unique continuous extension of

$$\frac{P_t p(\cdot) K(\cdot, k)}{P_t K(\cdot, k)}$$

whose existence is guaranteed by Lemma 3.15. Clearly, μp is linear in p.

For a polynomial $p \geq 0$ we have that $(P_t p(\cdot) K(\cdot, k)) \geq 0$ and $(P_t K(\cdot, k)) \geq 0$ by the positivity of P_t and the fact that $K(\cdot, k) \geq 0$. Hence for $x \in [0, 1)$, $\mu p(x) \geq 0$. Since μp is a continuous function, $\mu p \geq 0$, so μ is a positive operator.

Furthermore, by Lemma 3.11 $|\mu p(x)| \leq \|p\|$ for $x \in [0, 1)$. Using continuity of μp, $\|\mu p\| \leq \|p\|$, so μ is bounded, and therefore continuous. Finally $(\mu 1)(x) = 1$ for $x \in [0, 1)$, and using continuity of $\mu 1$, we get that $\mu 1 = 1$.

By Lemma 1.12, μ can be uniquely extended to a continuous probability kernel $\hat{\mu}$ on $[0, 1]$. Let f be in $\mathcal{C}[0,1]$ and let p_m be polynomials such that $p_m \to f$. Then $\mu p_m \to \hat{\mu} f$. For $x \in [0, 1)$ we have that

$$\begin{aligned} \hat{\mu} f(x) &= \lim_{m\to\infty} \frac{(P_t p_m(\cdot) K(\cdot, k))(x)}{(P_t K(\cdot, k))(x)} \\ &= \frac{(P_t f(\cdot) K(\cdot, k))(x)}{(P_t K(\cdot, k))(x)}. \end{aligned}$$

$\hat{\mu} f$ is therefore a continuous extension of

$$\frac{(P_t f(\cdot) K(\cdot, k))}{P_t K(\cdot, k)}$$

as required. □

Now we can prove Theorem 3.10.

Proof of Theorem 3.10. Lemma 3.16 shows that for every $0 \leq k < n$ there exists a continuous probability kernel μ_k on $[0, 1]$ such that for $f \in \mathcal{C}[0,1]$, $\mu_k f$ is a continuous extension of

$$\frac{(P_t f(\cdot) K(\cdot, k))}{P_t K(\cdot, k)}.$$

Lemma 3.12 shows that for $f \in \mathcal{C}[0,1]$

$$\frac{(P_t f(\cdot) K(\cdot, n))}{P_t K(\cdot, n)}$$

is a continuous function and it is easy to see that

$$\mu_n f = \frac{(P_t f(\cdot) K(\cdot, n))}{P_t K(\cdot, n)}, \; f \in \mathcal{C}[0,1]$$

defines a continuous probability kernel.

Let us now define

$$\mu f(x, k) = (\mu_k f(\cdot, k))(x)$$

where f is in $\mathcal{C}([0,1] \times \{0, \ldots, n\})$ and (x, k) is in $[0, 1] \times \{0, \ldots, n\}$. Since μ_k are continuous probability kernels, μf is continuous. By construction, μf is a continuous extension of

$$\frac{P_t f K}{P_t K}.$$

Moreover, since μ_k are continuous probability kernels, μ is a positive continuous linear operator such that $\mu 1 = 1$. Hence μ is a continuous probability kernel by Lemma 1.12. $\square$

Having proved Theorem 3.10, we can now use (3.20) to define $\boldsymbol{P}^{(t)}$, where $\frac{P_t f K}{P_t K}$ is defined in Theorem 3.10. Note that since $\boldsymbol{P}^{(t)}$ is a composition of the probability kernels Q_t and $\frac{P_t f K}{P_t K}$, it is itself a probability kernel.

Remark 3.17. Since for a fixed $t > 0$, $\boldsymbol{P}^{(t)}$ is a probability kernel, it can be used to define a discrete-time Markov chain (X_n, Y_n). It is easy to see that Y_n marginally is a Markov chain with transition kernel Q_t. This follows from the fact that if $f \in \mathcal{C}\left([0,1] \times \{0,\dots,n\}\right)$ does not depend on its first argument, then $\left(P_t f(k) K\left(\cdot, k\right)\right)(x) = f(k)\left(P_t K\left(\cdot, k\right)\right)(x)$, hence $\boldsymbol{P}^{(t)} f = Q_t f$. Moreover, if the initial distribution satisfies

$$\boldsymbol{\pi}_0^{(X,Y)}\left(A \times \{k\}\right) = \int_A K(x,k)\pi_0^X\left(\mathrm{d}x\right), \tag{3.24}$$

then also X_n marginally is a Markov chain with transition kernel P_t and moreover

$$\mathrm{P}\left(Y_n = y | X_i, 0 \leq i \leq n\right) = K\left(X_n, y\right). \tag{3.25}$$

This result is similar to Theorem 2.1, except that now the state-space of X_n is not countable.

Proof. (This proof is inspired by the proof of Proposition 4 in Fill [4].) To prove the claims about X_n, define

$$\Lambda\left(x, A \times \{k\}\right) = \delta_x(A) K(x,k)$$

where x is in $[0,1]$, A is a Borel subset of $[0,1]$ and k is in $\{0,\dots,n\}$. Observe that Λ is a probability kernel from $[0,1]$ to $[0,1] \times \{0,\dots,n\}$, $\boldsymbol{\pi}_0^{(X,Y)} = \pi_0^X \Lambda$ and

$$\begin{aligned}\Lambda f(x) &= \sum_{k=0}^n f(x,k)K(x,k)\\ &= \left(K f\left(x,\cdot\right)\right)(x)\end{aligned}$$

where $x \in [0,1]$ and f is in $\mathcal{C}\left([0,1] \times \{0,\dots,n\}\right)$. Observe also that

$$\begin{aligned}\Lambda Q_t f(x) &= \left(\left(K Q_t\right) f\left(x,\cdot\right)\right)(x)\\ &= \left(\left(P_t K\right) f\left(x,\cdot\right)\right)(x)\\ &= \sum_{k=0}^n \left(P_t K\left(\cdot,k\right)\right)(x) f(x,k),\end{aligned}$$

hence

$$\begin{aligned}\Lambda \boldsymbol{P}^{(t)} f(x) &= \Lambda Q_t \frac{P_t f K}{P_t K}(x)\\ &= \sum_{k=0}^n \left(P_t K\left(\cdot,k\right)\right)(x) \times \frac{P_t f K}{P_t K}(x,k).\end{aligned}$$

Now note that for $(x,k) \in [0,1] \times \{0,\ldots,n\}$ such that $(P_t K(\cdot,k))(x) > 0$, we have

$$\frac{P_t f K}{P_t K}(x,k) = \frac{(P_t f(\cdot,k) K(\cdot,k))(x)}{(P_t K(\cdot,k))(x)},$$

hence

$$(P_t K(\cdot,k))(x) \times \frac{P_t f K}{P_t K}(x,k) = (P_t f(\cdot,k) K(\cdot,k))(x).$$

Since all expressions in the last equality are continuous in (x,k) and since the set of (x,k) such that $(P_t K(\cdot,k))(x) > 0$ is dense, the last equality must hold for all $(x,k) \in [0,1] \times \{0,\ldots,n\}$. Therefore we get

$$\begin{aligned} \Lambda \boldsymbol{P}^{(t)} f(x) &= \sum_{k=0}^{n} (P_t f(\cdot,k) K(\cdot,k))(x) \\ &= (P_t \Lambda f)(x). \end{aligned} \tag{3.26}$$

Now we can prove (3.25). First observe that (3.25) is equivalent to

$$\mathrm{E}\left(f(X_n,Y_n) \,|\, X_i, 0 \le i \le n\right) = \Lambda f(X_n) \tag{3.27}$$

for all f in $\mathcal{C}([0,1] \times \{0,\ldots,n\})$. We shall prove (3.27) by induction in n. For $n = 0$, it follows directly from (3.24). Let us now assume that (3.25) holds for some $n \ge 0$ and prove it for $n+1$. Observe that

$$\begin{aligned} &\mathrm{E}\left[f(X_{n+1},Y_{n+1}) \,|\, X_i, 0 \le i \le n\right] = \\ &\qquad = \mathrm{E}\left[E\left[f(X_{n+1},Y_{n+1}) \,|\, (X_i,Y_i), 0 \le i \le n\right] |\, X_i, 0 \le i \le n\right] \\ &\qquad = \mathrm{E}\left[\boldsymbol{P}^{(t)} f(X_n,Y_n) \,|\, X_i, 0 \le i \le n\right] \\ &\qquad = \Lambda \boldsymbol{P}^{(t)} f(X_n) \\ &\qquad = P_t \Lambda f(X_n) \end{aligned} \tag{3.28}$$

where we have used the induction hypothesis and (3.26). Note that if f does not depend on its second argument, then from (3.28) we get

$$\begin{aligned} \mathrm{E}\left[f(X_{n+1}) \,|\, X_i, 0 \le i \le n\right] &= P_t \Lambda f(X_n) \\ &= P_t f(X_n), \end{aligned}$$

which proves that marginally, X_n is a Markov chain with transition kernel P_t. From (3.28) we then get

$$\mathrm{E}\left[f(X_{n+1},Y_{n+1}) \,|\, X_i, 0 \le i \le n\right] = \mathrm{E}\left[\Lambda f(X_{n+1}) \,|\, X_i, 0 \le i \le n\right].$$

Hence for g_i in $\mathcal{C}[0,1]$,

$$\begin{aligned} &\mathrm{E}\left[f(X_{n+1},Y_{n+1}) \prod_{i=0}^{n+1} g_i(X_i)\right] = \\ &\qquad = \mathrm{E}\left[\mathrm{E}\left[f(X_{n+1},Y_{n+1}) g_{n+1}(X_{n+1}) \,|\, X_i, 0 \le i \le n\right] \prod_{i=0}^{n} g_i(X_i)\right] \\ &\qquad = \mathrm{E}\left[\mathrm{E}\left[(\Lambda f g_{n+1})(X_{n+1}) \,|\, X_i, 0 \le i \le n\right] \prod_{i=0}^{n} g_i(X_i)\right] \\ &\qquad = \mathrm{E}\left[\Lambda f(X_{n+1}) \prod_{i=0}^{n+1} g_i(X_i)\right], \end{aligned}$$

where in the last equality we have used that

$$\begin{aligned}
(\Lambda f g_{n+1})(x) &= \sum_{k=0}^{n} K(x,k) f(x,k) g_{n+1}(x) \\
&= g_{n+1}(x) \sum_{k=0}^{n} K(x,k) f(x,k) \\
&= g_{n+1}(x) (\Lambda f)(x).
\end{aligned}$$

This proves (3.27) for $n+1$. □

Now that we have defined an approximation of the coupled semigroup, we can try to construct the generator by a limit procedure similar to the one used by Fill [4] (see (2.16)).

Lemma 3.18. *Let $G = \bar{G}^{WF,r}$ be the Wright-Fisher generator with reflection at zero defined by (1.12), let $H = H_n$ be the generator of the pure birth process defined by (3.6) and let $K = K_n$ be the probability kernel defined by (3.5). Let P_t and Q_t be the semigroups associated with G and H and let $\boldsymbol{P}^{(t)}$ be given by (3.20). Define*

$$D = \{ f \in \mathcal{C}([0,1] \times \{0,\dots,n\}) \,;\, f(\cdot,k)\, K(\cdot,k) \in \mathcal{D}(G) \,\forall k \in \{0,\dots,n\} \}.$$

Then for all $f \in D$ and $(x,k) \in [0,1] \times \{0,\dots,n\}$ such that $K(x,k) > 0$ we have

$$\frac{1}{t}\left(\boldsymbol{P}^{(t)} - 1\right) f(x,k) \to \boldsymbol{G} f(x,k)$$

pointwise as $t \to 0$, where

$$\begin{aligned}
\boldsymbol{G} f(x,k) &= \sum_{l \in \{0,\dots,n\}, K(x,l) > 0} H(k,l) f(x,l) \\
&\quad + \sum_{l \in \{0,\dots,n\}, K(x,l)=0, GK(x,l)>0} H(k,l) \frac{(Gf(\cdot,l)\, K(\cdot,l))(x)}{(GK(\cdot,l))(x)} \\
&\quad + \frac{(Gf(\cdot,k)\, K(\cdot,k))(x) - f(x,k)\,(GK(\cdot,k))(x)}{K(x,k)}.
\end{aligned} \tag{3.29}$$

Proof. Fix $f \in D$ and $(x,k) \in [0,1] \times \{0,\dots,n\}$ such that $K(x,k) > 0$. We have

$$\begin{aligned}
&\frac{1}{t}\left(\left(\boldsymbol{P}^{(t)} - 1\right) f\right)(x,k) = \\
&\qquad = \frac{1}{t}\left[Q_t \frac{P_t f K}{P_t K}(x,k) - f(x,k) \right] \\
&\qquad = \frac{1}{t}\left[(Q_t - 1) \frac{P_t f K}{P_t K}(x,k) \right. \\
&\qquad\quad \left. + \frac{(P_t f(\cdot,k)\, K(\cdot,k))(x) - f(x,k)\,(P_t K(\cdot,k))(x)}{(P_t K(\cdot,k))(x)} \right] \qquad (3.30) \\
&\qquad = \frac{1}{t}\left[(Q_t - 1) \frac{P_t f K}{P_t K}(x,k) + \frac{((P_t - 1) f(\cdot,k)\, K(\cdot,k))(x)}{(P_t K(\cdot,k))(x)} \right. \\
&\qquad\quad \left. - f(x,k) \frac{(P_t K(\cdot,k))(x) - K(x,k)}{(P_t K(\cdot,k))(x)} \right].
\end{aligned}$$

The first term in (3.30) equals

$$\begin{aligned}
\sum_{l=0}^{n} \frac{1}{t}\left(Q_t(k,l) - \mathbf{1}_{[k=l]}\right)\frac{P_t f K}{P_t K}(x,l) = \\
&= \sum_{l\in\{0,\dots,n\},K(x,l)>0} \frac{1}{t}\left(Q_t(k,l) - \mathbf{1}_{[k=l]}\right)\frac{P_t f K}{P_t K}(x,l) \\
&\quad + \sum_{l\in\{0,\dots,n\},K(x,l)=0,(GK)(x,l)>0} \frac{1}{t}\left(Q_t(k,l) - \mathbf{1}_{[k=l]}\right)\frac{P_t f K}{P_t K}(x,l) \qquad (3.31)\\
&\quad + \sum_{l\in\{0,\dots,n\},K(x,l)=0,(GK)(x,l)\le 0} \frac{1}{t}\left(Q_t(k,l) - \mathbf{1}_{[k=l]}\right)\frac{P_t f K}{P_t K}(x,l).
\end{aligned}$$

Note that $\frac{1}{t}\left(Q_t(k,l) - \mathbf{1}_{[k=l]}\right)$ converges to $H(k,l)$. To find the limit of the first sum in (3.31), observe that $(P_tK)(x,l)$ converges to $K(x,l)$ and $(P_tfK)(x,l)$ converges to $f(x,l)K(x,l)$. Thus, if $K(x,l)>0$, then $(P_tK)(x,l)>0$ for sufficiently small t and

$$\begin{aligned}
\frac{P_t f K}{P_t K}(x,l) &= \frac{\left(P_t f(\cdot,l)\,K(\cdot,l)\right)(x)}{\left(P_t K(\cdot,l)\right)(x)} \\
&\to \frac{f(x,l)K(x,l)}{K(x,l)} \\
&= f(x,l).
\end{aligned}$$

Hence,

$$\begin{aligned}
\lim_{t\to 0} \sum_{l\in\{0,\dots,n\},K(x,l)>0} \frac{1}{t}\left(Q_t(k,l) - \mathbf{1}_{[k=l]}\right)\frac{P_t f K}{P_t K}(x,l) \\
= \sum_{l\in\{0,\dots,n\},K(x,l)>0} H(k,l)f(x,l).
\end{aligned}$$

Now assume that $K(x,l)=0$ and $\left(GK(\cdot,l)\right)(x)>0$. Then we have

$$\left.\frac{d}{dt}\left(P_t K(\cdot,l)\right)(x)\right|_{t=0} = \left(GK(\cdot,l)\right)(x) > 0,$$

so again $\left(P_tK(\cdot,l)\right)(x) > 0$ for all sufficiently small $t>0$. Now we can use l'Hospital rule to get

$$\begin{aligned}
\lim_{t\to 0}\frac{\left(P_t f(\cdot,l)\,K(\cdot,l)\right)(x)}{\left(P_t K(\cdot,l)\right)(x)} &= \lim_{t\to 0}\frac{\left(P_t G f(\cdot,l)\,K(\cdot,l)\right)(x)}{\left(P_t G K(\cdot,l)\right)(x)} \\
&= \frac{G\left(f(\cdot,l)\,K(\cdot,l)\right)(x)}{\left(GK(\cdot,l)\right)(x)},
\end{aligned}$$

therefore

$$\begin{aligned}
\lim_{t\to 0} \sum_{l\in\{0,\dots,n\},K(x,l)=0,(GK)(x,l)>0} \frac{1}{t}\left(Q_t(k,l) - \mathbf{1}_{[k=l]}\right)\frac{P_t f K}{P_t K}(x,l) \\
= \sum_{l\in\{0,\dots,n\},K(x,l)=0,(GK)(x,l)>0} H(k,l)\frac{G\left(f(\cdot,l)\,K(\cdot,l)\right)(x)}{\left(GK(\cdot,l)\right)(x)}.
\end{aligned}$$

Finally, let $K(x,l)=0$ and $(GK(\cdot,l))(x)\le 0$. Observe that

$$\begin{aligned}
(GK(\cdot,l))(x) &= \left(GK\mathbf{1}_{\{l\}}\right)(x) \\
&= KH\mathbf{1}_{\{l\}}(x) \\
&= \sum_{u=0}^{n} K(x,u)H(u,l) \\
&= \mathbf{1}_{[l>0]}\lambda_{l-1}K(x,l-1)-\lambda_l K(x,l) \\
&= \mathbf{1}_{[l>0]}\lambda_{l-1}K(x,l-1) \\
&\ge 0 \qquad (3.32)
\end{aligned}$$

where we have used Lemma 3.3, the definition of H (3.6) and the fact that $K(x,l)=0$. Note that since $K(x,k)>0$ and $K(x,l)=0$, $k\neq l$. Also note that if $k=l-1$, the inequality in (3.32) would be strict, which would be in contradiction with our assumption that $(GK(\cdot,l))(x)\le 0$. Hence, k is neither l nor $l-1$, which by definition of H (3.6) means that $H(k,l)=0$. Since we also have

$$\left|\frac{P_t fK}{P_t K}(x,l)\right| \le \|f\|,$$

we conclude that

$$\lim_{t\to 0}\sum_{l\in\{0,\dots,n\},K(x,l)=0,(GK)(x,l)\le 0}\frac{1}{t}\left(Q_t(k,l)-\mathbf{1}_{[k=l]}\right)\frac{P_t fK}{P_t K}(x,l)=0.$$

To sum up, we have shown that

$$\begin{aligned}
&\lim_{t\to 0}\frac{1}{t}\left(Q_t-\mathbf{1}_{[k=l]}\right)\frac{P_t fK}{P_t K}(x,k) \\
&\qquad = \sum_{l\in\{0,\dots,n\},K(x,l)>0} H(k,l)f(x,l) \\
&\qquad\quad + \sum_{l\in\{0,\dots,n\},K(x,l)=0,GK(x,l)>0} H(k,l)\frac{(Gf(\cdot,l)K(\cdot,l))(x)}{(GK(\cdot,l))(x)}.
\end{aligned} \tag{3.33}$$

To conclude the proof, we note that

$$\lim_{t\to 0}\frac{1}{t}\frac{((P_t-1)f(\cdot,k)K(\cdot,k))(x)}{(P_tK(\cdot,k))(x)}=\frac{(Gf(\cdot,k)K(\cdot,k))(x)}{K(x,k)}$$

and

$$\lim_{t\to 0}\frac{1}{t}f(x,k)\frac{(P_tK(\cdot,k))(x)-K(x,k)}{(P_tK(\cdot,k))(x)}=f(x,k)\frac{(GK(\cdot,k))(x)}{K(x,k)}.$$

□

Now we observe that the expression (3.29) can be slightly simplified.

Lemma 3.19. *Let $\boldsymbol{G}$ be defined as in (3.29). Then*

$$\boldsymbol{G}f(x,k)=(Hf(x,\cdot))(k)+\frac{(Gf(\cdot,k)K(\cdot,k))(x)-f(x,k)(GK(\cdot,k))(x)}{K(x,k)}$$

where f is in D and $(x,k)\in[0,1]\times\{0,\dots,n\}$ such that $K(x,k)>0$.

Proof. We have to show that for $(x,k) \in [0,1] \times \{0,\dots,n\}$ such that $K(x,k) > 0$,

$$\sum_{l\in\{0,\dots,n\},K(x,l)>0} H(k,l)f(x,l) + \sum_{l\in\{0,\dots,n\},K(x,l)=0,GK(x,l)>0} H(k,l)\frac{\left(Gf\left(\cdot,l\right)K\left(\cdot,l\right)\right)(x)}{\left(GK\left(\cdot,l\right)\right)(x)} = \left(Hf(x,\cdot)\right)(k).$$

It suffices to show that if $x \in [0,1]$ and $k,l \in \{0,\dots,n\}$, such that $K(x,k) > 0$, $K(x,l) = 0$ and $H(k,l) \neq 0$, then $\left(GK\left(\cdot,l\right)\right)(x) > 0$ and

$$\frac{\left(Gf\left(\cdot,l\right)K\left(\cdot,l\right)\right)(x)}{\left(GK\left(\cdot,l\right)\right)(x)} = f(x,l). \tag{3.34}$$

Observe that $K(x,l)$ can be zero only if $x \in \{0,1\}$, so it suffices to consider only such x.

Let us first consider the case $x = 0$. Then $K(0,k) > 0$ if and only if $k = 0$. Now $H(0,l) \neq 0$ if and only if $l \in \{0,1\}$. But for $l = 0$, $K(0,l) > 0$, so we only need to consider the case $l = 1$. It is easy to see that

$$\begin{aligned}\left(Gf\left(\cdot,1\right)K\left(\cdot,1\right)\right)(0) &= \left(1-x^2\right)\frac{\partial^2}{\partial x^2}f(x,1)\left(x^2-x^4\right)\Big|_{x=0} \\ &= 2f(0,1)\end{aligned}$$

and similarly $\left(GK\left(\cdot,1\right)\right)(0) = 2$, hence (3.34) is satisfied.

Let now $x = 1$. Then $K(1,k) > 0$ if and only if $k = n$, but then $H(k,l) = H(n,l) = 0$ for all l. □

Now that we have found the candidate generator by limiting procedure, we shall prove that it indeed is the generator of a Feller semigroup.

Theorem 3.20. *Let $G = \bar{G}^{WF,r}$ as in (1.12) be the generator of the Wright-Fisher diffusion with reflection at zero, $H = H_n$ the generator of the pure birth process defined by (3.6) and $K = K_n$ the probability kernel defined by (3.5), restricted to $[0,1] \times \{0,\dots,n\}$. Define*

$$\mathcal{D}\left(\boldsymbol{G}\right) = \left\{f \in \mathcal{C}\left([0,1]\times\{0,\dots,n\}\right);\, f\left(\cdot,k\right) \text{ is an even polynomial } \forall k \in \{0,\dots,n\}\right\}$$

For $(x,k) \in [0,1] \times \{0,\dots,n\}$ such that $K(x,k) > 0$ and $f \in \mathcal{D}\left(\boldsymbol{G}\right)$ define

$$\boldsymbol{G}f(x,k) = \left(Hf\left(x,\cdot\right)\right)(k) + \frac{\left(Gf\left(\cdot,k\right)K\left(\cdot,k\right)\right)(x) - f(x,k)\left(GK\left(\cdot,k\right)\right)(x)}{K(x,k)}. \tag{3.35}$$

Then $\boldsymbol{G}f$ can be extended to a continuous function on $[0,1] \times \{0,\dots,n\}$. Moreover, $\boldsymbol{G}$ is a closable operator and its closure is the generator of a Feller semigroup on $\mathcal{C}\left([0,1]\times\{0,\dots,n\}\right)$.

Remark 3.21. Observe that under the assumptions of Theorem 3.20, both $K\left(\cdot,k\right)$ and $f\left(\cdot,k\right)K\left(\cdot,k\right)$ are even polynomials, and therefore are in $\mathcal{D}\left(G\right)$. Hence, the expression (3.35) is well defined.

We will prove Theorem 3.20 in several lemmas.

Lemma 3.22. *$\mathcal{D}\left(\boldsymbol{G}\right)$ is dense in $\mathcal{C}\left([0,1]\times\{0,\dots,n\}\right)$.*

Proof. It suffices to prove that even polynomials are dense in $\mathcal{C}[0,1]$. This follows from the Stone-Weierstrass theorem. $\square$

Lemma 3.23. *Let f be in $\mathcal{D}(\boldsymbol{G})$. For $(x,k) \in [0,1] \times \{0,\dots,n\}$ such that $K(x,k) > 0$ we have*

$$\begin{aligned}\boldsymbol{G}f(x,k) &= \mathbf{1}_{[k\neq n]}\lambda_k\left(f(x,k+1)-f(x,k)\right)\\ &\quad+\left[\frac{4k}{x}-2(2k+2\mathbf{1}_{[k\neq n]})x\right]\frac{\partial}{\partial x}f(x,k)\\ &\quad+\left(1-x^2\right)\frac{\partial^2}{\partial x^2}f(x,k).\end{aligned} \tag{3.36}$$

Moreover, $\boldsymbol{G}f$ can be extended to a continuous function on $[0,1] \times \{0,\dots,n\}$.

Proof. For $f \in \mathcal{D}(\boldsymbol{G})$ and $(x,k) \in [0,1] \times \{0,\dots,n\}$ such that $K(x,k) > 0$ we have

$$\begin{aligned}\left(Gf(\cdot,k)\,K(\cdot k)\right)(x) &= \left(1-x^2\right)\frac{\partial^2}{\partial x^2}\left(f(x,k)K(x,k)\right)\\ &= \left(1-x^2\right)\left(\frac{\partial^2}{\partial x^2}f(x,k)\right)K(x,k)\\ &\quad+2\left(1-x^2\right)\frac{\partial}{\partial x}f(x,k)\frac{\partial}{\partial x}K(x,k)\\ &\quad+\left(1-x^2\right)f(x,k)\frac{\partial^2}{\partial x^2}K(x,k),\end{aligned}$$

hence

$$\begin{aligned}&\frac{\left(Gf(\cdot,k)\,K(\cdot,k)\right)(x)-f(x,k)\left(GK(\cdot,k)\right)(x)}{K(x,k)}\\ &\qquad=\left(1-x^2\right)\frac{\partial^2}{\partial x^2}f(x,k)+2\left(1-x^2\right)\frac{\frac{\partial}{\partial x}K(x,k)}{K(x,k)}\frac{\partial}{\partial x}f(x,k).\end{aligned}$$

For $0 \le k < n$ we have

$$\begin{aligned}2\left(1-x^2\right)\frac{\frac{\partial}{\partial x}K(x,k)}{K(x,k)} &= 2\frac{2kx^{2k-1}-(2k+2)x^{2k+1}}{x^{2k}}\\ &= \frac{4k}{x}-2(2k+2)x,\end{aligned}$$

and for $k = n$ we get

$$\begin{aligned}2\left(1-x^2\right)\frac{\frac{\partial}{\partial x}K(x,n)}{K(x,n)} &= 2\left(1-x^2\right)\frac{2nx^{2n-1}}{x^{2n}}\\ &= \frac{4n}{x}-4nx.\end{aligned}$$

This shows that

$$\begin{aligned}&\frac{\left(Gf(\cdot,k)\,K(\cdot,k)\right)(x)-f(x,k)\left(GK(\cdot,k)\right)(x)}{K(x,k)}\\ &\qquad=\left[\frac{4k}{x}-2(2k+2\mathbf{1}_{[k\neq n]})x\right]\frac{\partial}{\partial x}f(x,k)+\left(1-x^2\right)\frac{\partial^2}{\partial x^2}f(x,k)\end{aligned}$$

and that concludes the proof of (3.36). To prove that $\boldsymbol{G}f$ has a continuous extension, observe that

$$\lim_{x\to 1} \boldsymbol{G}f(x,k)$$

trivially always exists. To prove that

$$\lim_{x\to 0} \boldsymbol{G}f(x,k)$$

exists, it suffices to prove that

$$\lim_{x\to 0} \frac{4k}{x}\frac{\partial}{\partial x} f(x,k)$$

exists. For $k = 0$ this is trivial and for $k = 1, \ldots, n$ it follows from the fact that $f(\cdot, k)$ is an even polynomial. $\square$

Remark 3.24. For the interpretation of (3.36) notice that

$$\mathbf{1}_{[k\neq n]}\lambda_k\left(f(x,k+1) - f(x,k)\right) = \left(Hf(x,\cdot)\right)(k)$$

corresponds to the pure birth process,

$$\left(1-x^2\right)\frac{\partial^2}{\partial x^2} f(x,k) = \left(Gf(\cdot,k)\right)(x)$$

corresponds to the Wright-Fisher diffusion (with reflection), and

$$\left[\frac{4k}{x} - 2(2k + 2\mathbf{1}_{[k\neq n]})x\right]\frac{\partial}{\partial x} f(x,k)$$

is an extra drift for the diffusion. Once we have proved that $\boldsymbol{G}$ generates a Markov process, denoted by (X_t, Y_t), we can interpret Y_t as the pure birth process with birth rates λ_k and X_t as the Wright-Fisher diffusion with reflection at zero and with extra drift which depends on the state k of Y_t. For fixed k, the drift is positive for $x < x_k$ and negative for $x > x_k$, where

$$x_k = \sqrt{\frac{4k}{4k + 4\mathbf{1}_{[k\neq n]}}}.$$

Hence, we can interpret X_t as being pushed towards x_k when $Y_t = k$. Noting that x_k is increasing in k and $x_n = 1$, we can think of X_t as being initially reluctant to get absorbed, but getting more and more compelled to get absorbed as Y_t rises.

Lemma 3.25. *$\boldsymbol{G}$ satisfies the positive maximum principle.*

Proof. Let f be in $\mathcal{D}(\boldsymbol{G})$ and (x,k) be in $[0,1]\times\{0,\ldots,n\}$ such that

$$\sup_{(y,l)\in[0,1]\times\{0,\ldots,n\}} f(y,l) = f(x,k) \geq 0.$$

Then $\mathbf{1}_{[k<n]}\lambda_k\left(f(x,k+1) - f(x,k)\right)$ must be non-positive. If x is in $(0,1)$, then $\frac{\partial f}{\partial x}(x,k) = 0$ and $\frac{\partial^2 f}{\partial x^2}(x,k) \leq 0$, so $\boldsymbol{G}f(x,k) \leq 0$. If $x = 0$, then the second order

term of the polynomial $x \mapsto f(x,k)$, $x \in [0,1]$ must be non-positive, since it is an even polynomial. Hence

$$\lim_{x\to 0} \frac{1}{x}\frac{\partial}{\partial x} f(x,k) \leq 0$$

and

$$\lim_{x\to 0} \frac{\partial^2}{\partial x^2} f(x,k) \leq 0,$$

so $\boldsymbol{G}(0,k) \leq 0$.

Finally , if $x=1$, then $\frac{\partial}{\partial x} f(x,k) \geq 0$ and $\frac{4k}{x} - 2(2k+2\mathbf{1}_{[k\neq n]})x \leq 0$. Moreover, $(1-x^2)\frac{\partial^2}{\partial x^2} f(x,k) = 0$, so $\boldsymbol{G}f(x,k) \leq 0$. □

Lemma 3.26. *Let $\lambda > 0$. Then $\mathcal{R}(\lambda - \boldsymbol{G}) \supseteq \mathcal{D}(\boldsymbol{G})$, hence $\mathcal{R}(\lambda - \boldsymbol{G})$ is dense in $\mathcal{C}([0,1] \times \{0,\ldots,n\})$.*

Proof. Let g be in $\mathcal{D}(\boldsymbol{G})$. Then there exist $M \in \mathbb{N}$ and $b_{k,m} \in \mathbb{R}$, with $k = 0,\ldots,n$ and $m = 0,\ldots,M$ such that

$$g(x,k) = \sum_{m=0}^{M} b_{k,m} x^m,\ x \in [0,1]\,. \tag{3.37}$$

Moreover, $b_{k,m} = 0$ whenever m is odd. We will try to find $a_{k,m}$ such that

$$f(x,k) = \sum_{m=0}^{M} a_{k,m} x^m,\ x \in [0,1] \tag{3.38}$$

satisfies $(\lambda - \boldsymbol{G}) f = g$. First observe that $(\lambda - \boldsymbol{G}) f(\cdot,k)$ is also a polynomial for all $k = 0,\ldots,n$. If we define $\lambda_n = 0$ and $a_{k,m} = 0$ if either $k > n$ or $m > M$, then from (3.36) we get that the coefficient of x^m in $(\lambda - \boldsymbol{G}) f(\cdot,k)$ is

$$\begin{aligned}\lambda a_{k,m} - \lambda_k a_{k+1,m} + \lambda_k a_{k,m} - 4k(m+2)a_{k,m+2} + 2\left(2k + 2\mathbf{1}_{[k\neq n]}\right) m a_{k,m} \\ - (m+2)(m+1)\, a_{k,m+2} + m(m-1)a_{k,m}.\end{aligned} \tag{3.39}$$

$(\lambda - \boldsymbol{G}) f = g$ holds if and only if the expression in (3.39) equals $b_{k,m}$, which is equivalent to

$$a_{k,m} = \frac{b_{k,m} + \lambda_k a_{k+1,m} + 4k(m+2)a_{k,m+2} + (m+2)(m+1)\, a_{k,m+2}}{\lambda + \lambda_k + 2\left(2k + 2\mathbf{1}_{[k\neq n]}\right) m + m(m-1)}. \tag{3.40}$$

Note that the denominator of (3.40) is non-zero, because $\lambda > 0$ and the other terms are non-negative. (3.40) can be solved recursively, starting by $a_{n,M}$ proceeding to $a_{n,0}$, then $a_{n-1,M}$ to $a_{n-1,0}$ and so on. Observe that (3.40) also guarantees that $a_{k,m} = 0$ whenever m is odd, hence f is in $\mathcal{D}(\boldsymbol{G})$. □

The previous lemmas together with the Hille-Yosida Theorem 1.27 and the simple observation that $\boldsymbol{G}1 = 0$ prove Theorem 3.20.

Remark 3.27. Lemma 3.26 is the only point where our proof fails for the Wright-Fisher diffusion without reflection at zero. Originally, we tried to prove Theorem 3.20 for the Wright-Fisher diffusion without reflection. We defined $\boldsymbol{G}$ by (3.35), where (x,k) was now in $[-1,1] \times \{0,\ldots,n\}$ such that $K(x,k) > 0$

and $G = \bar{G}^{WF}$ was taken to be the generator of the Wright-Fisher diffusion (defined by (1.10)). Note that at this point we could not take $\mathcal{D}(\boldsymbol{G})$ to be functions such that $f(\cdot, k)$ are even polynomials, because even polynomials are not dense in $\mathcal{C}[-1,1]$. But if we allowed all polynomials, we could not prove that $\boldsymbol{G}f$ can be extended to a continuous function on $[-1,1] \times \{0, \ldots, n\}$ because of the term $\frac{4k}{x}\frac{\partial}{\partial x} f(x,k)$ in (3.36). The deeper reason for these problems is that $K(0,k) = 0$ for $k > 0$. The interpretation is that after the pure birth process departs from zero, the diffusion is no longer allowed to cross 0. Moreover, for $k > 0$, the drift of the diffusion points away from zero and approaches infinity as x approaches zero (see (3.36)). This means that the semigroup $\boldsymbol{P}_t f(x,k) = \mathrm{E}\left[f\left(X_t, Y_t\right) | X_0 = x, Y_0 = k\right]$ may not be expected to be continuous at $(0,k)$ for $k > 0$ – if X_0 is a small positive number, the drift will push it away from zero, so $\boldsymbol{P}_t f(x,k)$ will depend on values of $f\left(\tilde{x}, \tilde{k}\right)$, $\tilde{x} > 0$, and if X_0 is a negative number close to zero, the drift will push it further to the negative numbers, so $\boldsymbol{P}_t f(x,k)$ will depend on values of $f\left(\tilde{x}, \tilde{k}\right)$, $\tilde{x} < 0$.

To overcome this problem, we defined the state-space of the coupled process to be

$$\begin{aligned} \boldsymbol{S} &= [-1,1] \times \{0\} \\ &\cup \left(\left[-1, 0^-\right] \cup \left[0^+, 1\right]\right) \times \{1, \ldots, n\}. \end{aligned} \tag{3.41}$$

Here, we interpret 0^- and 0^+ to be two distinct points. Now we could define $\mathcal{D}(\boldsymbol{G})$ to be the set of functions on $\boldsymbol{S}$ such that

$$x \mapsto f(x, 0)$$

is a polynomial and

$$\begin{aligned} x &\mapsto f(x,k),\ x \in \left[-1, 0^-\right] \\ x &\mapsto f(x,k),\ x \in \left[0^+, 1\right] \end{aligned}$$

are even polynomials for all $k = 1, \ldots, n$. Note that when $k > 0$, the polynomials for $x > 0$ and $x < 0$ are allowed to have different coefficients. This allows us to prove that $\mathcal{D}(\boldsymbol{G})$ is dense in $\mathcal{C}(\boldsymbol{S})$ by the Stone-Weierstrass theorem. Moreover, since all polynomials except for $k = 0$ are even, $\boldsymbol{G}f$ can be extended to a continuous function on $\mathcal{C}(\boldsymbol{S})$ as in Lemma 3.23 (the fact that $f(x,0)$ is not necessarily even does not matter, since the term $\frac{4k}{x}\frac{\partial}{\partial x} f(x,k)$ vanishes for $k = 0$).

In this setting, however, the proof of Lemma 3.26 fails. The reason is that now equation (3.37) has to be rewritten to reflect the fact that the coefficients of $g(k,x)$ may differ for positive and negative x when $k > 0$. We may write it as

$$\begin{aligned} g(k,x) &= \sum_{m=0}^{M} b^+_{k,m} x^m,\ x \in \left[0^+, 1\right], \\ g(k,x) &= \sum_{m=0}^{M} b^-_{k,m} x^m,\ x \in \left[-1, 0^-\right], \end{aligned}$$

with the additional requirement that $b^-_{0,m} = b^+_{0,m}$ for all $m = 0, \ldots, M$. Equation (3.38) has to be modified in a similar way. The effect of this is that we also get two versions of equation (3.40), one for the "plus" coefficients and the other for the "minus" coefficients. Since we solve this equation backward, starting at $k = n$ and proceeding toward $k = 0$, we cannot guarantee that $a_{0,m}^{-} = a_{0,m}^{+}$ even if we know that $b_{0,m}^{-} = b_{0,m}^{+}$. We cannot even prove that the $a_{0,0}^{+}$ and $a_{0,0}^{-}$ would agree, meaning that f could be discontinuous at zero.

An obvious idea how to get around this problem is to split the interval $[-1, 1]$ into two even for $k = 0$. However, $K(0, 0) > 0$, so the drift at $(x, k) = (0, 0)$ is finite. The diffusion can therefore reach this point, and we need to specify what happens here. If we say that the diffusion should reflect (which corresponds to requiring that $f(x, 0), x > 0$ and $f(x, 0), x < 0$ are even polynomials; a natural requirement since we require the same thing for $k > 0$), then we get the Wright-Fisher diffusion with reflection at zero. Now the parts where $x > 0$ and $x < 0$ are symmetric and independent of each other, so we need to only deal with one of them. If we choose only the part where $x > 0$, we get the present formulation of Theorem 3.20.

To summarize, it seems likely that if we define $\boldsymbol{G}$ by (3.36) and make a suitable choice of $\mathcal{D}(\boldsymbol{G})$, then a version of Theorem 3.20 holds also for the process without reflection, where the state-space of the process generated by $\boldsymbol{G}$ should be defined by (3.41). However, it seems impossible to use polynomials to prove that the closure of $\boldsymbol{G}$ defined by (3.36) generates a Feller process on $\boldsymbol{\mathcal{S}}$. Using other techniques, such as stochastic differential equations, would lead too far for the present thesis.

In order to show that the generator $\boldsymbol{G}$ satisfies the requirements of Theorem 3.9, we will use the following theorem due to Rogers and Pitman [7].

Theorem 3.28. *Let $(\boldsymbol{S}, \boldsymbol{\mathscr{S}})$ and $(S, \mathscr{S})$ be measurable spaces and let $\phi : \boldsymbol{S} \to S$ be a measurable transformation. Let Λ be a probability kernel from S to $\boldsymbol{S}$ and define a probability kernel from $\boldsymbol{S}$ to S by*

$$\Phi f = f \circ \phi.$$

Let X be a continuous-time Markov process with state space $(\boldsymbol{S}, \boldsymbol{\mathscr{S}})$, transition semigroup $\boldsymbol{P}_t$ and initial distribution $\boldsymbol{\pi}_0 = \pi_0 \Lambda$, for some distribution π on S. Suppose further:

1. *$\Lambda\Phi = I$, the identity kernel on S,*
2. *for each $t \geq 0$ the probability kernel $P_t = \Lambda \boldsymbol{P}_t \Phi$ from S to S satisfies*

$$\Lambda \boldsymbol{P}_t = P_t \Lambda. \tag{3.42}$$

Then P_t is a transition semigroup on S, $\phi \circ X$ is Markov with the initial distribution π_0 and

$$\mathrm{P}\left(X_t \in A | \phi \circ X_s, 0 \leq s \leq t\right) = \Lambda\left(\phi \circ X_t, A\right)$$

a. s. for all $t \geq 0$ and $A \in \boldsymbol{\mathscr{S}}$.

Proof. Rogers and Pittman ([7], Theorem 2) proved this for the case that $\pi_0 = \delta_y$ for some $y \in S$. The general case follows by integration with respect to π_0. □

Now we are able to prove Theorem 3.9.

Proof of Theorem 3.9. We shall prove that the generator $\boldsymbol{G}$ defined by (3.35) satisfies the requirements of the Theorem. Let f be in $\mathcal{C}(\{0,\ldots,n\})$. We may view $\mathcal{C}(\{0,\ldots,n\})$ as a subspace of $\mathcal{C}([0,1]\times\{0,\ldots,n\})$ using the natural embedding

$$\Psi : f \mapsto f \circ \psi$$

where

$$\psi(x,k) = k.$$

Since

$$\mathrm{E}\left[f\left(Y_{s+t}\right) \mid \left(X_u, Y_u\right), 0 \le u \le s\right] = \boldsymbol{P}_t f\left(X_s, Y_s\right)$$

where $\boldsymbol{P}_t$ is the semigroup corresponding to $\boldsymbol{G}$, (3.16) will be proved if we prove that $\boldsymbol{P}_t f(x,k)$ does not depend on x and equals $Q_t f(k)$ for all $f \in \mathcal{C}(\{0,\ldots,n\})$ and $t \ge 0$. Let f be in $\mathcal{D}(H) = \mathcal{C}(\{0,\ldots,n\})$. Then f is in $\mathcal{D}(\boldsymbol{G})$ and $\boldsymbol{G}f = Hf$ by (3.35), so $Q_t = \boldsymbol{P}_t$ on $\mathcal{C}(\{0,\ldots,n\})$ by Corollary 1.30.

In order to prove the claims about X_t, we will use Theorem 3.28. In the present setting, $\boldsymbol{S} = [0,1]\times\{0,\ldots,n\}$ and $S = [0,1]$. Define a function $\phi : [0,1]\times\{0,\ldots,n\} \to [0,1]$ by $\phi(x,k) = x$ and a probability kernel Φ from $[0,1]\times\{0,\ldots,n\}$ to $[0,1]$ as in Theorem 3.28. Define also a probability kernel from $[0,1]$ to $[0,1]\times\{0,\ldots,n\}$ by

$$\Lambda\left(x, A\times\{y\}\right) = \delta_x(A)K(x,y),$$

where $x\in[0,1]$, $y\in\{0,\ldots,n\}$ and A is a Borel subset of $[0,1]$. In other words,

$$\Lambda f(x) = \sum_{k=0}^{n} K(x,k) f(x,k)$$

for $f \in \mathcal{C}([0,1]\times\{0,\ldots,n\})$ and $x\in[0,1]$. Observe that

$$\begin{aligned}\pi_0^X \Lambda\left(A\times\{y\}\right) &= \int \delta_x(A)K(x,y)\pi_0^X(\mathrm{d}x)\\ &= \boldsymbol{\pi}_0^{(X,Y)}\left(A\times\{y\}\right).\end{aligned}$$

Also observe that for $f\in\mathcal{C}[0,1]$ we have

$$\Lambda\Phi f(x) = \sum_{k=0}^{n} K(x,k) f(x) = f(x),$$

hence

$$\Lambda\Phi = I. \tag{3.43}$$

Let us now prove that

$$\Lambda\boldsymbol{P}_t = P_t\Lambda \tag{3.44}$$

where $P_t = P_t^{WF,r}$ is the semigroup of the Wright-Fisher diffusion with reflection at zero. By Theorem 3.1, it suffices to prove that

$$\Lambda\boldsymbol{G}f = G\Lambda f$$

for all $f \in \mathcal{C}([0,1] \times \{0,\dots,n\})$ such that $f(\cdot,k)$ is an even polynomial for all $k = 0,\dots,n$. Let f be such a function and $(x,k) \in [0,1] \times \{0,\dots,n\}$ such that $K(x,k) > 0$. By (3.35),

$$\begin{aligned}&K(x,k)(\boldsymbol{G}f)(x,k)\\&\quad = K(x,k)(Hf(\cdot,x))(k) + G(f(\cdot,k)K(\cdot,k))(x) - f(x,k)(GK(\cdot,k))(x).\end{aligned}$$

Since both sides of the equality are continuous in (x,k) and the set where $K(x,k) > 0$ is dense, the equality must hold for all $(x,k) \in [0,1] \times \{0,\dots,n\}$. Hence

$$\Lambda \boldsymbol{G} f(x) = \sum_{k=0}^{n} K(x,k)(Hf(x,\cdot))(k) + G\Lambda f(x) - \sum_{k=0}^{n} f(x,k)(GK(\cdot,k))(x).$$

The first term can be rewritten as

$$\sum_{k=0}^{n} K(x,k) \sum_{l=0}^{n} H(k,l) f(x,l).$$

The last term can be written as

$$\begin{aligned}\sum_{k=0}^{n} f(x,k)\left(GK\mathbf{1}_{\{k\}}\right)(x) &= \sum_{k=0}^{n} f(x,k)\left(KH\mathbf{1}_{\{k\}}\right)(x)\\ &= \sum_{k=0}^{n} f(x,k) \sum_{l=0}^{n} K(x,l)H(l,k)\end{aligned}$$

where in the first equality we have used Lemma 3.3. Therefore, $\Lambda \boldsymbol{G} f = G\Lambda f$.

Finally, from (3.43) and (3.44) we get that $P_t = \Lambda \boldsymbol{P}_t \Phi$. Thus, we have verified all requirements of Theorem 3.28. It follows that X_t is the Wright-Fisher diffusion with reflection at zero with the initial distribution π_0^X and

$$\begin{aligned}\mathrm{P}(Y_t = k | X_s, 0 \le s \le t) &= \Lambda(X_t, [0,1] \times \{k\})\\ &= K(X_t, k).\end{aligned}$$

□

Remark 3.29. Observe that in the proof of Theorem 3.9 we have shown that

$$\Lambda \boldsymbol{P}_t = P_t \Lambda$$

and then applied Theorem 3.28. Meanwhile, in the proof of Remark 3.17 we have derived an analogous relation

$$\Lambda \boldsymbol{P}^{(t)} = P_t \Lambda$$

and then proved (3.25) by induction. We could have proved Theorem 3.9 by an inductive argument similar to that of Remark 3.17 without using Theorem 3.28. To see this, note that (3.17) is equivalent to saying that for all $m \in \mathbb{N}$ and for all $0 \le t_1 < \cdots < t_m$,

$$\mathrm{P}(Y_{t_m} = k | X_{t_m}, \dots, X_{t_1}) = K(X_{t_m}, k) \text{ a.s., } k = 0,\dots,n. \tag{3.45}$$

Moreover, saying that X_t on its own is the Wright-Fisher diffusion with reflection at zero is equivalent to saying that for all $m \in \mathbb{N}$ and for all $0 \le t_1 < \cdots < t_m$,

$$\mathrm{P}\left(X_{t_m} \in A | X_{t_{m-1}}, \dots, X_{t_1}\right) = P^{WF,r}_{t_m - t_{m-1}}\left(X_{t_{m-1}}, A\right) \text{ a.s., } A \in \mathcal{B}([0,1]), \tag{3.46}$$

where $P_t^{WF,r}$ is the semigroup of the Wright-Fisher diffusion with reflection at zero. We could have proved (3.45) and (3.46) by induction in m.

Conclusion

In this thesis we have generalized several results that were previously proved only for Markov processes with discrete state-space. In particular, we have shown that for arbitrary Feller semigroups, the intertwining of their generators of the form

$$GK = KH$$

is equivalent to the intertwining of the semigroups

$$P_t K = K Q_t, \tag{3.47}$$

see Theorem 3.1. We have used this result to find an intertwining between the Wright-Fisher diffusion and a pure birth process on $\mathbb{N} \cup \{\infty\}$. This intertwining relation allowed us to find a probabilistic proof of the fact that the time to absorption of the Wright-Fisher diffusion is distributed as a sum of independent exponential random variables (although the distribution of the time to absorption was already known, our proof of this fact is new).

A simple consequence of the intertwining relation (3.47) is that if the initial distributions satisfy

$$\pi_0^Y = \pi_0^X K,$$

then also the time t distributions satisfy

$$\pi_t^Y = \pi_t^X K,$$

in other words

$$\mathrm{P}(Y_t = k) = \mathrm{E}\left[K\left(X_t, k\right)\right].$$

As in the discrete setting [4], one can try to generalize this by finding a coupling of the processes X_t and Y_t which satisfies

$$\mathrm{P}\left(Y_t = k | X_u, 0 \leq u \leq t\right) = K\left(X_t, k\right).$$

We have tried to find such a coupling for the Wright-Fisher diffusion and a pure-birth process on $\mathbb{N} \cup \{\infty\}$, but due to technical difficulties, we were only able to find it for the Wright-Fisher diffusion with reflection at zero and for a pure-birth process on $\{0, \ldots, n\}$. In the process of constructing the coupling, we noticed that when $Y_t = k$, X_t behaves like the Wright-Fisher diffusion with reflection at zero and with additional drift which depends on k. This drift pushes X_t toward certain point, and this point increases in k and reaches 1 when $k = n$. This can be interpreted as X_t being initially reluctant to get absorbed, but getting more compelled to get absorbed as Y_t rises.

We have constructed the coupled process directly from the generator. However, as described in Chapter 2, Diaconis and Fill [1] and Fill [4] showed that in the discrete setting, there is also an alternative, sample path construction. We believe that this sample path construction can be extended to construct a coupling for the Wright-Fisher diffusion (with or without reflection). To outline the construction (for concreteness, we demonstrate the construction for the diffusion without reflection), let $X_t^{(0,x_0)}$ be the Markov process on $[-1,1]$ generated by

$$G^{(0,x_0)}f(x) = -4x\frac{\partial}{\partial x}f(x) + \left(1-x^2\right)\frac{\partial^2}{\partial x^2}f(x),\ f \in \mathcal{C}^2[-1,1]$$

and started at x_0, where $x_0 \in [-1,1]$.[2] Similarly, for each $x_0 \in (0,1]$ and $k = 1,\dots,n$, let us have Markov process $X_t^{(k,x_0)}$ with state-space $[0,1]$ started at x_0 and generated by

$$\begin{aligned} G^{(k,x_0)}f(x) &= \left[\frac{4k}{x} - 2(2k + 2\mathbf{1}_{[k\neq n]})x\right]\frac{\partial}{\partial x}f(x) + \left(1-x^2\right)\frac{\partial^2}{\partial x^2}f(x), \\ & \qquad f \in \mathcal{C}^2[0,1]\,\text{s.t.}\,\frac{\partial}{\partial x}f(x) = 0, \end{aligned}$$

and for each $x_0 \in [-1,0)$ and $k = 1,\dots,n$, let us have Markov process $X_t^{(k,x_0)}$ with state-space $[-1,0]$ started at x_0 and generated by

$$\begin{aligned} G^{(k,x_0)}f(x) &= \left[\frac{4k}{x} - 2(2k + 2\mathbf{1}_{[k\neq n]})x\right]\frac{\partial}{\partial x}f(x) + \left(1-x^2\right)\frac{\partial^2}{\partial x^2}f(x), \\ & \qquad f \in \mathcal{C}^2[-1,0]\,\text{s.t.}\,\frac{\partial}{\partial x}f(x) = 0. \end{aligned}$$

Let Y_t be the pure birth process on $\{0,\dots,n\}$ generated by H_n of (3.6) and started at zero. Assume that Y_t and all $X_t^{(k,x_0)}$ are mutually independent. Let τ_k, $k = 1,\dots,n$ be the times when Y_t jumps from $k-1$ to k, and define $\tau_{n+1} = \infty$. Now define X_t as

$$X_t = \begin{cases} X_t^{(0,0)}, & t \le \tau_1, \\ X_{t-\tau_k}^{\left(k,X_{\tau_k}\right)}, & \tau_k < t \le \tau_{k+1}, k = 1,\dots,n. \end{cases}$$

Note that X_t is not well defined for $t > \tau_k$ in the event that $X_{\tau_k} = 0$, but since this event has zero probability, we can define X_t for $t > \tau_k$ arbitrarily (e. g. defining $X_t = 0$ for $t > \tau_k$ if $X_{\tau_k} = 0$) without affecting its distribution. In view of Remark 3.24, we conjecture that (X_t, Y_t) is a coupled process in the sense of Theorem 3.9. We also conjecture that this construction could be used to construct a coupling of the Wright-Fisher diffusion (with or without reflection at zero) with the pure birth process on $\mathbb{N} \cup \{\infty\}$.

[2]For the motivation for this and the following generators, see Lemma 3.23.

Bibliography

[1] Persi Diaconis and James Allen Fill. Strong stationary times via a new form of duality. *The Annals of Probability*, 18(4):1483–1522, 1990.

[2] Persi Diaconis and Laurent Miclo. On times to quasi-stationarity for birth and death processes. *Journal of Theoretical Probability*, 22(3):558–586, 2009.

[3] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes: Characterization and Convergence*. John Wiley & Sons, 2009.

[4] James Allen Fill. Strong stationary duality for continuous-time markov chains. Part I: Theory. *Journal of Theoretical Probability*, 5(1):45–70, 1992.

[5] Olav Kallenberg. *Foundations of Modern Probability*. Springer Science & Business Media, 1997.

[6] Thomas Milton Liggett. *Continuous Time Markov Processes: An Introduction*. American Mathematical Soc., 2010.

[7] L. C. G. Rogers and J. W. Pitman. Markov functions. *The Annals of Probability*, 9(4):573–582, 1981.

[8] Walter Rudin. *Real and Complex Analysis (3rd ed.)*. McGraw-Hill Inc, 1986.

[9] Jan Seidler. Markovské procesy. Available online `http://simu0292.utia.cas.cz/seidler/teaching.html`.

[10] Jan M. Swart. Intertwining of birth-and-death processes. *Kybernetika*, 47(1):1–14, 2011.