INTERTWINING OF THE WRIGHT–FISHER DIFFUSION

Tobiáš Hudec

It is known that the time until a birth and death process reaches a 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. Their coupling is of a special type called intertwining of Markov processes. We apply this technique to couple the Wright–Fisher diffusion with reflection at 0 and a pure birth process. We show that in our coupling the time of absorption of the diffusion is a. s. equal to the time of explosion of the pure birth process. The coupling also 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 of Markov processes, Wright–Fisher diffusion, pure birth process, time of absorption, coupling

Classification: 60J60, 60J35, 60J27

1. INTRODUCTION AND THE MAIN RESULT

1.1. Introduction

It is known that the time until a birth and death process X_t started at the origin reaches a certain level is distributed as a sum of independent exponential variables whose parameters are the negatives of the non-zero eigenvalues of the generator of the process stopped at the given level (see Karlin and McGregor [7]). Diaconis and Miclo [2] and Swart [12] gave 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 technique that Diaconis, Miclo and Swart employ is called intertwining of Markov processes. This technique was developed by Rogers and Pitman [11], Diaconis and Fill [1], and Fill [4]. It allows them to add structure to the process X_t such that it is initially reluctant to be absorbed, but after each exponential time (which corresponds to jump times of Y_t) it changes its behavior to be more and more compelled to be absorbed. Since one-dimensional diffusions can be obtained as limits of birth and death processes, it is interesting to investigate whether this technique can be extended to the case that X_t is a diffusion.

Recently, Fill and Lyzinski [5] have studied intertwining relations between two one-dimensional diffusion processes. In this paper, we consider something different, namely an intertwining between a diffusion process and a pure birth process. For technical

reasons, in this paper we restrict ourself to one particular diffusion, namely the Wright–Fisher diffusion with reflection at 0, which has state-space $[0, 1]$ and is absorbed at 1. The generator and the semigroup of this diffusion have the nice property that they map polynomials to polynomials of the same order, which simplifies our proofs. We need that the diffusion is reflected at 0 for technical reasons; without it, one of our proofs would not work (see Remark 8). We construct an explosive pure birth process Y_t such that X_t is absorbed at the same time as Y_t explodes.

The idea of Diaconis, Miclo and Swart can be summarized as follows. For a given transition semigroup P_t of a birth and death process X_t on $\{0, \ldots, n\}$ absorbed at n, Swart finds a transition semigroup Q_t of a pure birth process Y_t on $\{0, \ldots, n\}$ and a probability kernel K which satisfies

$$P_t K = K Q_t \qquad (t \geq 0). \tag{1}$$

The algebraic relation (1) is called intertwining, which gives the name to the intertwining of Markov processes. Swart builds on an earlier work of Diaconis and Miclo, who found an intertwining of the form $K P_t = Q_t K$. However, we focus on Swart's construction, because our work on the Wright–Fisher diffusion is closer in spirit to his. He uses a result proved by Fill [4] which says that if X_t and Y_t are Markov processes with finite state-spaces related by (1), then the two processes can be coupled (i. e. defined on the same probability space) such that[1]

$$\mathrm{P}\left(Y_t = y | X_u, 0 \leq u \leq t\right) = K\left(X_t, y\right) \qquad \text{a.s.} \quad (t \geq 0). \tag{2}$$

Using (2) and the fact that his kernel satisfies

$$K(x, n) = \mathbf{1}_{[x=n]} := \begin{cases} 1, & \text{if } x = n, \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

Swart proves 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 paper, we derive analogue results for the case that X_t is the Wright–Fisher diffusion with reflection at zero. We find an explosive pure-birth process Y_t on the one-point compactification $\bar{\mathbb{N}} := \mathbb{N} \cup \{\infty\} := \{0, 1, \ldots, \infty\}$ of $\mathbb{N}$ and a probability kernel K from $[0, 1]$ to $\bar{\mathbb{N}}$ satisfying the intertwining relation (1) and

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

We couple the two processes such that they satisfy (2) which allows us to conclude that the time of absorption of X_t is a. s. equal to the time of explosion of Y_t. But since the time of explosion of the pure birth process is the sum of independent exponential variables whose intensities are the birth rates of Y_t, this gives us a new proof of the distribution of the time to absorption of X_t.[2]

---

[1]Under certain conditions, Fill showed that this coupling can be extended to countably infinite state-spaces. Fill built on earlier work of Diaconis and Fill [1], where an analogous result is proved for processes with discrete time.

[2]Just as with the birth and death processes, the distribution of the time of absorption of the Wright–Fisher diffusion has long been known (see e. g. Kent [8]), but our proof is new.

1.2. Intertwining of the Wright–Fisher diffusion

Define

$$\mathcal{D}(G) = \left\{ f \in \mathcal{C}^2[0,1]; \ \frac{\partial}{\partial x} f(0) = 0 \right\}$$

and

$$Gf(x) = \left(1 - x^2\right) \frac{\partial^2}{\partial x^2} f(x), \qquad f \in \mathcal{D}(G), x \in [0,1]. \tag{4}$$

In the appendix we show that G is closable and its closure generates a Feller semigroup, which we denote P_t. We also show that the associated Markov process, which we call the Wright–Fisher diffusion with reflection at zero, has continuous sample paths. Note that the generator of the Wright–Fisher diffusion is usually defined as

$$\frac{1}{2} x(1-x) \frac{\partial^2}{\partial x^2} f(x), \quad f \in \mathcal{C}^2[0,1],\ x \in [0,1] \tag{5}$$

(see e.g. Liggett [9, Example 3.48]). However, if $\tilde{X}_t$ is generated by (5), then $X_t = \left| 2\tilde{X}_{2t} - 1 \right|$ is generated by (4).

Define H as the generator of an explosive pure birth process on $\bar{\mathbb{N}}$ which jumps from y to $y+1$ with the rate

$$\lambda_y = (2y + 1)(2y + 2),\ y \in \mathbb{N}.$$

That is, define H as an operator from $\mathbb{R}^{\bar{\mathbb{N}}}$ to $\mathbb{R}^{\bar{\mathbb{N}}}$ by

$$
\begin{aligned}
Hf(y) &= \lambda_y \left(f(y+1) - f(y) \right), \quad y \in \mathbb{N}, \\
Hf(\infty) &= 0,
\end{aligned}
$$

where f is in $\mathbb{R}^{\bar{\mathbb{N}}}$. It is shown in the Appendix that the restriction of H to a suitable domain is the generator of a Feller semigroup on $\mathcal{C}\left(\bar{\mathbb{N}}\right)$, which we denote by Q_t. Define a probability kernel from $[0,1]$ to $\bar{\mathbb{N}}$ by

$$K(x,y) = \begin{cases} \left(1 - x^2\right) x^{2y}, & \text{if } 0 \le y < \infty, \\ \mathbf{1}_{[x=1]}, & \text{if } y = \infty. \end{cases} \tag{6}$$

It can be shown that K maps $\mathcal{C}\left(\bar{\mathbb{N}}\right)$ into $\mathcal{C}[0,1]$. We claim that there is an intertwining relation:

Theorem 1. We have

$$P_t K = K Q_t, \qquad (t \ge 0). \tag{7}$$

Using (7), we are able to couple the two processes in the spirit of [2, 4, 12]. We define the state-space $\boldsymbol{S}$ of the coupled process as the one-point compactification of $[0,1] \times \mathbb{N}$, where we denote the point at infinity by $(1, \infty)$. Using this notation, we can think of $\boldsymbol{S}$ as a subset of $[0,1] \times \bar{\mathbb{N}}$, but keep in mind that the topology of $\boldsymbol{S}$ is not the one induced by $[0,1] \times \bar{\mathbb{N}}$. Note that if the coupled process is to satisfy an analogue of (2), then it would be natural to construct it on the space

$$\left\{ (x,y) \in [0,1] \times \bar{\mathbb{N}};\ K(x,y) > 0 \right\},$$

as was done for Markov processes with discrete state-spaces [1, 4]. However, this space is not compact, so if we want to use the theory of Feller semigroups we must compactify it (either implicitly or explicitly). It turns out that $\boldsymbol{S}$ is the right compactification.

An analogous result to the following theorem was proved by Fill [4, Theorem 2] for processes with discrete state-spaces and by Diaconis and Fill [1, Theorem 2.33] for processes with discrete time and space.

Theorem 2. There exists a Feller process (X_t, Y_t) on $\boldsymbol{S}$ such that

$$\mathrm{E}\left(f\left(Y_{s+t}\right)|X_u, Y_u, 0 \leq u \leq s\right) = \left(Q_t f\right)\left(Y_s\right) \text{ a.s.} \tag{8}$$

for all $f \in \mathcal{C}\left(\bar{\mathbb{N}}\right)$ and $s, t \geq 0$. Hence, Y_t on its own is a pure birth process on $\bar{\mathbb{N}}$ with birth rates λ_y. If the initial distribution satisfies

$$\pi_0^{(X,Y)}(A \times \{y\}) = \int_A K(x, y) \pi_0^X(\mathrm{d}y), \tag{9}$$

where π_0^X is an arbitrary probability measure on $[0, 1]$, then X_t on its own is the Wright–Fisher diffusion with reflection at zero with initial distribution π_0^X and we have

$$\mathrm{P}\left(Y_t = y | X_s, 0 \leq s \leq t\right) = K\left(X_t, y\right) \quad \text{a.s.} \tag{10}$$

for all $y \in \bar{\mathbb{N}}$ and $t \geq 0$.

Note that if both X_t and Y_t start from zero, then (9) is satisfied. Using Theorem 2 we can prove that the time of absorption of the diffusion is a.s. equal to the time of explosion of the pure birth process. Indeed, from (10) we have

$$\mathrm{P}\left(X_t \in A, Y_t \in B\right) = \mathrm{E}\left(\mathbf{1}_A\left(X_t\right) K\left(X_t, B\right)\right).$$

But since $K(x, \cdot)$ is concentrated on $\mathbb{N}$ for $x < 1$ and on $\{\infty\}$ for $x = 1$, we have

$$\mathrm{P}\left(X_t < 1, Y_t = \infty\right) = \mathrm{P}\left(X_t = 1, Y_t < \infty\right) = 0.$$

1.3. Discussion

In addition to proving the a.s. equality of the time of absorption and the time of explosion of the two processes, Theorem 2 shows more about the underlying structure. By inspection of the formula for the coupled generator (17) below we see that conditionally on $Y_t = y \in \mathbb{N}$ we can interpret X_t as the Wright–Fisher diffusion with reflection at zero and with additional drift, which at the point $X_t = x$ equals

$$\frac{4y}{x} - 4(y + 1)x.$$

It can be shown that the scale function $u(x)$ and the speed measure $m(\mathrm{d}x)$ of this diffusion satisfy

$$
\begin{aligned}
u'(x) &= \frac{1}{x^{4y}\left(1 - x^2\right)^2}, \\
m(\mathrm{d}x) &= x^{4y}\left(1 - x^2\right) \mathrm{d}x.
\end{aligned}
$$

Hence, by Mandl [10, pp. 24–25], both boundaries are entrance for $y > 0$ and 0 is a regular boundary while 1 is an entrance boundary for $y = 0$. In particular, the coupled process lives on the set $\big\{(x, y) \in [0, 1] \times \bar{\mathbb{N}}; \, K(x, y) > 0\big\}$ as one might expect.

Moreover, we can see that there is an equilibrium point

$$x_y = \sqrt{\frac{y}{y+1}}$$

such that the drift is positive when $x < x_y$ and negative when $x > x_y$. We can interpret this as, conditionally on $Y_t = y$, X_t is pushed toward the equilibrium point x_y. Obviously, x_y is monotonous in y and goes from 0 to 1 as y goes from 0 to ∞. Thus, we can think of X_t as being initially reluctant to be absorbed, but later getting more and more compelled to get absorbed.

In our paper we construct Markov processes from generators using the Hille-Yosida theorem. We could also construct them as solutions to martingale problems or stochastic differential equations. However, we chose the Hille-Yosida theorem for its simplicity. Theorem 1 extends results of Diaconis and Miclo [2] and Swart [12], who proved similar theorems for birth and death processes. Theorem 2 extends results of Fill [4], who proved similar result for Markov processes with continuous time and discrete state-space, and of Diaconis and Fill [1], who proved it for the case of discrete time and space. It remains an open problem whether results like Theorems 1 and 2 hold for other diffusions than the modified version of the Wright–Fisher diffusion we consider in our paper. It seems that our proof of Theorem 1 does not exploit any peculiarity of the Wright–Fisher diffusion and we believe it could be extended to other types of diffusions as well. On the other hand, our proof of Theorem 2 depends strongly on the fact that the generator of the Wright–Fisher diffusion maps polynomials to polynomials of the same order, and it seems that entirely different proof techniques would be required for other diffusions.

2. PROOFS

2.1. Intertwining

To prove Theorem 1 we need to show that there is an intertwining between semigroups P_t and Q_t. The following theorem says that we can show this by proving that there is an intertwining between the generators. An analogous result for Markov processes with discrete state-spaces was proved by Fill [4, Lemma 3]. Although we use the following theorem only when P_t and Q_t are Feller semigroups and K is a probability kernel, we are able to prove it more generally.

Theorem 3. 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 $t \geq 0$,

$$P_t K = K Q_t \tag{11}$$

on L_2.

2. K maps $\mathcal{D}(H)$ into $\mathcal{D}(G)$ and

$$GK = KH \tag{12}$$

on $\mathcal{D}(H)$.

3. There exists a core D of H (i.e. D is a dense subspace of $\mathcal{D}(H)$ such that the closure of the restriction of H to D is H) such that K maps D into $\mathcal{D}(G)$ and (12) holds on D.

P r o o f. To prove (1) $\Rightarrow$ (2), fix $f \in \mathcal{D}(H)$. Then $\frac{1}{t}(Q_t f - f)$ converges to Hf, so by the continuity of K, $\frac{1}{t}(KQ_t f - Kf)$ converges to KHf. By (11), $\frac{1}{t}(P_t Kf - Kf)$ 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 $\mathcal{D}(H)$ by [3, Proposition 1.1.5], $u(t) \in \mathcal{D}(G)$ 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 as a function of t. By Proposition 1.3.4 in Ethier and Kurtz [3], $u(t) = P_t u(0) = P_t Kf$ which proves that (11) holds on $\mathcal{D}(H)$. Since all operators involved in (11) 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 (12) for f_n. Since G is a closed operator, Kf is in $\mathcal{D}(G)$ and $GKf = KHf$. $\qquad\square$

Now let us apply the preceding result to the operators G, H and K introduced in Theorem 1. Theorem 3 shows that it suffices to prove the equality (12) on a core of H. In the Appendix it is shown that

$$D_H = \left\{ f \in \mathcal{C}\left(\bar{\mathbb{N}}\right); \exists y_0 \in \mathbb{N} \,\mathrm{s.t.}\, \forall y > y_0 \, f(y) = f(\infty) \right\} \tag{13}$$

is a core of H. The following theorem verifies condition 3 of Theorem 3.

Theorem 4. K maps D_H into $\mathcal{D}(G)$ and

$$GK = KH$$

on D_H.

P r o o f. Fix f in D_H and let $y_0 \in \mathbb{N}$ be such that $f(y) = f(\infty)$ for all $y \geq y_0$. Then for $x \in [0, 1)$,

$$
\begin{aligned}
Kf(x) &= \left(1 - x^2\right) \sum_{y=0}^{y_0 - 1} x^{2y} f(y) + x^{2y_0} f(y_0) \\
&= f(0) + \sum_{y=1}^{y_0} x^{2y} \left(f(y) - f(y-1)\right).
\end{aligned} \tag{14}
$$

As x approaches 1, $Kf(x)$ approaches $f(y_0)$. Now

$$Kf(1) = f(\infty) = f(y_0).$$

Hence (14) holds also for $x = 1$, and therefore Kf is in $\mathcal{C}^\infty[0,1] \subseteq \mathcal{C}^2[0,1]$. Moreover

$$\frac{\partial}{\partial x} Kf(x) = \sum_{y=1}^{y_0} 2y x^{2y-1} \left(f(y) - f(y-1) \right),$$

hence

$$\frac{\partial}{\partial x} Kf(0) = 0.$$

We have shown that Kf is in $\mathcal{D}(G)$.

From (14) we have that for $x \in [0,1]$,

$$
\begin{aligned}
GKf(x) &= \left(1 - x^2\right) \sum_{y=1}^{y_0} 2y(2y-1) x^{2y-2} \left(f(y+1) - f(y) \right) \\
&= \left(1 - x^2\right) \sum_{y=0}^{y_0-1} \lambda_y x^{2y} \left(f(y+1) - f(y) \right).
\end{aligned}
$$

Now for $y < y_0$

$$Hf(y) = \lambda_y \left(f(y+1) - f(y) \right),$$

and for $y \geq y_0$,

$$Hf(y) = 0,$$

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

$$KHf(x) = \left(1 - x^2\right) \sum_{y=0}^{y_0-1} \lambda_y x^{2y} \left(f(y+1) - f(y) \right).$$

$\square$

P r o o f. [Proof of Theorem 1.]
We use Theorem 3. In the present context, $L_1 = \mathcal{C}[0,1]$ and $L_2 = \mathcal{C}\left(\bar{\mathbb{N}}\right)$. Thus we need to show that K maps $\mathcal{C}\left(\bar{\mathbb{N}}\right)$ to $\mathcal{C}[0,1]$. This is equivalent to saying that the measures $K(x,\cdot)$ are continuous in x with respect to the weak convergence. But this is easy to prove, since $K(x,\cdot)$ is geometric distribution with success parameter $1 - x^2$ if $x < 1$, and it is the degenerate distribution δ_∞ if $x = 1$. Theorem 4 verifies condition 3 of Theorem 3. Theorem 3 shows that this is equivalent to condition 1, and this is what we had to prove. $\square$

2.2. Coupling

In order to find a coupling of Theorem 2, recall that $\boldsymbol{S}$ is the one-point compactification of $[0,1] \times \mathbb{N}$, where $(1,\infty)$ denotes the point at infinity. It is easy to see that $f : \boldsymbol{S} \to \mathbb{R}$ is continuous if and only if $f(\cdot, y)$ is continuous for all $y \in \mathbb{N}$ and $f(x,y) \to f(1,\infty)$ as $y \to \infty$, uniformly in x. It is also easy to see that

$$\{f \in \mathcal{C}(\boldsymbol{S}); f(x,y) = f(1,\infty) \text{ for all } x \in [0,1] \text{ and } y > y_0 \text{ for some } y_0 \in \mathbb{N}\}$$

is dense in $\mathcal{C}(\boldsymbol{S})$. Since even polynomials are dense in $\mathcal{C}[0,1]$ by the Stone–Weierstrass theorem, it follows that

$$
\begin{aligned}
\mathcal{D}(\boldsymbol{G}) = \{f \in \mathcal{C}(\boldsymbol{S}); \exists y_0 \in \mathbb{N} \, \text{s.t.} \, f(x,y) = f(1,\infty) \text{ for all } x \in [0,1] \text{ and } y > y_0, \\
f(\cdot, y) \text{ is an even polynomial for all } y \le y_0\}
\end{aligned}
\tag{15}
$$

is dense in $\mathcal{C}(\boldsymbol{S})$.

We now define an operator $\boldsymbol{G}$ with domain $\mathcal{D}(\boldsymbol{G})$ which we later prove generates a Feller process satisfying Theorem 2. For the motivation of this definition, see section 2.3. For $f \in \mathcal{D}(\boldsymbol{G})$ and $(x,y) \in (0,1) \times \mathbb{N}$ define

$$\boldsymbol{G}f(x,y) = (Hf(x,\cdot))(y) + \frac{G\left(f(\cdot,y) K(\cdot,y)\right)(x) - f(x,y)\left(GK(\cdot,y)\right)(x)}{K(x,y)}. \tag{16}$$

Here $G\left(f(\cdot,y) K(\cdot,y)\right)$ denotes the application of the operator G to the product of $f(\cdot,y)$ and $K(\cdot,y)$, and $(GK(\cdot,y))$ is the application of the operator G to $K(\cdot,y)$. In both cases, y is held fixed, so $f(\cdot,y)$ and $K(\cdot,y)$ are viewed as functions of x only. $(Hf(x,\cdot))$ is interpreted similarly, but here x is held fixed. Note that $f(\cdot,y)K(\cdot,y)$ and $K(\cdot,y)$ are even polynomials, hence they are in $\mathcal{D}(G)$. Moreover, $f(x,\cdot)$ is in D_H of (13), which is a core of H as shown in the Appendix, hence $f(x,\cdot)$ is in $\mathcal{D}(H)$. Finally, $K(x,y) > 0$ since x is in $(0,1)$ and $y < \infty$. Therefore, all the expressions in (16) are well defined. After plugging in the definitions of H and G, we can get an explicit formula for $\boldsymbol{G}$.

Lemma 5. Let f be in $\mathcal{D}(\boldsymbol{G})$ and $(x,y) \in (0,1) \times \mathbb{N}$. Let $\boldsymbol{G}f$ be defined by (16). Then

$$\boldsymbol{G}f(x,y) = \lambda_y\left(f(x,y+1) - f(x,y)\right) + \left(1 - x^2\right)\frac{\partial^2}{\partial x^2}f(x,y) + 4\left[\frac{y}{x} - (y+1)x\right]\frac{\partial}{\partial x}f(x,y). \tag{17}$$

P r o o f. Observe that for $(x,y) \in (0,1) \times \mathbb{N}$,

$$
\begin{aligned}
G\left(f(\cdot,y) K(\cdot,y)\right)(x) \;=\; & \left(1 - x^2\right)\left(\frac{\partial^2}{\partial x^2}f(x,y)\right) K(x,y) \\
& + 2\left(1 - x^2\right)\frac{\partial}{\partial x}f(x,y)\frac{\partial}{\partial x}K(x,y) \\
& + \left(1 - x^2\right)f(x,y)\frac{\partial^2}{\partial x^2}K(x,y).
\end{aligned}
$$

Hence

$$\frac{G\left(f\left(\cdot,y\right)K\left(\cdot,y\right)\right)(x) - f(x,y)\left(GK\left(\cdot,y\right)\right)(x)}{K(x,k)}$$

$$= \left(1 - x^2\right)\frac{\partial^2}{\partial x^2}f(x,y) + 2\left(1 - x^2\right)\frac{\frac{\partial}{\partial x}K(x,y)}{K(x,y)}\frac{\partial}{\partial x}f(x,y).$$

Noting that

$$2\left(1 - x^2\right)\frac{\frac{\partial}{\partial x}K(x,y)}{K(x,y)} = 2\frac{2yx^{2y-1} - (2y+2)x^{2y+1}}{x^{2y}}$$

$$= \frac{4y}{x} - 4(y+1)x,$$

we get

$$\frac{G\left(f\left(\cdot,y\right)K\left(\cdot,y\right)\right)(x) - f(x,y)\left(GK\left(\cdot,y\right)\right)(x)}{K(x,k)}$$

$$= \left(1 - x^2\right)\frac{\partial^2}{\partial x^2}f(x,y) + 4\left[\frac{y}{x} - (y+1)x\right]\frac{\partial}{\partial x}f(x,y). \quad (18)$$

Plugging (18) and the definition of H into (16), we get (17). $\qquad\square$

Formula (17) is well defined even for $x = 1$. Moreover, since $f(x,y)$ is an even polynomial in x, it follows that

$$\lim_{x\downarrow 0}\frac{y}{x}\frac{\partial}{\partial x}f(x,y)$$

exists, hence we can define $\boldsymbol{G}f$ on $[0,1]\times\mathbb{N}$ by taking the limit. Observe that for $y > y_0$ (where y_0 is as in (15)), $\boldsymbol{G}f(x,y) = 0$. Therefore, if we define $\boldsymbol{G}f(1,\infty) = 0$, then $\boldsymbol{G}f$ is in $\mathcal{D}\left(\boldsymbol{G}\right) \subseteq \mathcal{C}(\boldsymbol{S})$ and we can view $\boldsymbol{G} : \mathcal{D}(\boldsymbol{G}) \to \mathcal{C}(\boldsymbol{S})$ as a linear operator.

Theorem 6. Operator $\boldsymbol{G}$ is closable and its closure generates a Feller semigroup.

In order to prove Theorem 6, we use the following corollary to the Hille-Yosida theorem.

Proposition 7. Let E be a compact metric space, $\mathcal{D}(G)$ a subspace of $\mathcal{C}(E)$, and $G : \mathcal{D}(G) \to \mathcal{C}(E)$ a linear operator. Suppose that 1 is in $\mathcal{D}(G)$ and $G1 = 0$, G satisfies the positive maximum principle, and there exist a sequence $(L_n)_{n\in\mathbb{N}}$ of finite-dimensional subspaces of $\mathcal{D}(G)$ such that $\bigcup_{n\in\mathbb{N}} L_n$ is dense in $\mathcal{C}(E)$ and $G : L_n \to L_n$. Then G is closable and its closure generates a Feller semigroup.

P r o o f . Lemma 4.2.1 in Ethier and Kurtz [3] shows that G is dissipative, and Proposition 1.3.5 in [3] then proves that G is closable and its closure generates a strongly continuous contraction semigroup. Finally, the fact that $G1 = 0$ proves that G is conservative. $\qquad\square$

P r o o f. [Proof of Theorem 6]

Let us first prove that $\boldsymbol{G}$ satisfies the positive maximum principle. Let $f \in \mathcal{D}(\boldsymbol{G})$ and $(x_0, y_0) \in \boldsymbol{S}$ be such that

$$\sup_{(x,y)\in\boldsymbol{S}} f(x,y) = f(x_0, y_0) \geq 0.$$

First if $y_0 = \infty$, then $\boldsymbol{G}f(x_0, y_0) = 0$ by definition. Second, let us assume that $(x_0, y_0) \in [0,1] \times \mathbb{N}$. Then we have $f(x_0, y_0 + 1) - f(x_0, y_0) \leq 0$. If $x_0 \in (0,1)$, then $\frac{\partial}{\partial x} f(x_0, y_0) = 0$ and $\frac{\partial^2}{\partial x^2} f(x_0, y_0) \leq 0$, hence $\boldsymbol{G}f(x_0, y_0) \leq 0$. If $x_0 = 1$, then

$$4 \left[\frac{y_0}{x_0} - (y_0 + 1)x_0 \right] \frac{\partial}{\partial x} f(x_0, y_0) \leq 0$$

and

$$\left(1 - x_0^2\right) \frac{\partial^2}{\partial x^2} f(x_0, y_0) = 0,$$

so $\boldsymbol{G}f(x_0, y_0) \leq 0$. And if $x_0 = 0$, then the second-order term of the polynomial $f(\cdot, y_0)$ must be non-positive, for otherwise $(0, y_0)$ could not be a point of maximum. Hence

$$\lim_{x\downarrow 0} \frac{y_0}{x} \frac{\partial}{\partial x} f(x, y_0) \leq 0$$

and

$$\left(1 - x_0^2\right) \frac{\partial^2}{\partial x^2} f(x_0, y_0) \leq 0,$$

so we again get that $\boldsymbol{G}f(x_0, y_0) \leq 0$. We have shown that $\boldsymbol{G}$ satisfies the positive maximum principle.

Define

$$
\begin{aligned}
L_n \quad = \quad & \{f \in \mathcal{C}(\boldsymbol{S}); \ f(\cdot, y) \text{ is an even polynomial of degree at most } 2n \text{ for all } y \leq n, \\
& f(\cdot, y) = f(1, \infty) \text{ for all } y > n\}.
\end{aligned}
$$

It is easy to see that $\boldsymbol{G} : L_n \to L_n$ and $\bigcup_{n\in\mathbb{N}} L_n = \mathcal{D}(\boldsymbol{G})$ is dense in $\mathcal{C}[0,1]$. Finally, $\boldsymbol{G}\mathbf{1} = 0$, so by Proposition 7, $\boldsymbol{G}$ is closable and its closure generates a Feller semigroup. $\qquad \square$

Remark 8. The proof of Theorem 6 is the only place where our argument fails for the Wright–Fisher diffusion without reflection at zero. Indeed, we could take P_t to be the semigroup of the diffusion on the whole interval $[-1, 1]$, that is, P_t would be generated by

$$Gf(x) = \left(1 - x^2\right) \frac{\partial^2}{\partial x^2} f(x), \qquad f \in \mathcal{C}^2[-1, 1], x \in [-1, 1]. \tag{19}$$

We could now extend kernel K to be from $[-1, 1]$ to $\bar{\mathbb{N}}$ using the same formula (6). Our proof of Theorem 1 would still work. We could define $\boldsymbol{G}$ by (16) where G would now be defined by (19). But now we could not take $\mathcal{D}(\boldsymbol{G})$ to be functions such that $f(\cdot, y)$ are even polynomials, because they are not dense in $\mathcal{C}[-1, 1]$. But if we allowed

all polynomials, then for $y > 0$ we could not extend $(\boldsymbol{G}f)(\cdot, y)$ to a continuous function on $[-1, 1]$ because of the term $\frac{y}{x}\frac{\partial}{\partial x}f(x, y)$ in (17). The deeper reason for this problem is that $K(0, y) = 0$ for $y > 0$, so if the process (X_t, Y_t) satisfies (10), then after Y_t departs from zero, X_t is no longer allowed to cross zero, so the behavior of the diffusion on $[-1, 0]$ and $[0, 1]$ are independent. To overcome this problem, we could define

$$\boldsymbol{S} = [-1, 1] \times \{0\} \cup [-1, 0^-] \times \{1, 2, \dots\} \cup [0^+, 1] \times \{1, 2 \dots\} \cup \{-1, 1\} \times \{\infty\},$$

where we think of 0^- and 0^+ as two different points. Now we could take $\mathcal{D}(\boldsymbol{G})$ to be functions such that $f(\cdot, 0)$ is a polynomial, $f(\cdot, y)$ is an even polynomial with possibly different coefficients on $[-1, 0^-]$ and on $[0^+, 1]$ and from some y_0, $f(x, y)$ equals either $f(-1, \infty)$ or $f(1, \infty)$ depending on whether x is in $[-1, 0^-]$ or $[0^+, 1]$. This set is dense in $\mathcal{C}(\boldsymbol{S})$. Then, however, $(Gf)(\cdot, 0)$ could be discontinuous at $x = 0$ because of the term $\lambda_y (f(x, y+1) - f(x, y))$ in (17). To get around this problem, we decided to work with the Wright–Fisher diffusion with reflection at zero.

In order to prove Theorem 2, we need the following theorem due to Rogers and Pittman [11].

Theorem 9. Let $(\boldsymbol{S}, \mathcal{S})$ and $(S, \mathcal{S})$ 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 $\boldsymbol{X}_t$ be a continuous-time Markov process with state space $(\boldsymbol{S}, \mathcal{S})$, transition semi-group $\boldsymbol{P}_t$ and initial distribution $\boldsymbol{\pi}_0 = \pi_0 \Lambda$, for some distribution π_0 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

$$\Lambda\boldsymbol{P}_t = P_t\Lambda. \tag{20}$$

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

$$\mathrm{P}(\boldsymbol{X}_t \in A | \phi \circ \boldsymbol{X}_s, 0 \leq s \leq t) = \Lambda(\phi \circ \boldsymbol{X}_t, A)$$

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

P r o o f. Rogers and Pittman [11, 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.   □

P r o o f. [Proof of Theorem 2.]

It is a standard result that there exists a process $\boldsymbol{X} = (X, Y)$ generated by $\boldsymbol{G}$ and it is intuitively clear from the form of $\boldsymbol{G}$ that Y_t on its own is generated by H, but here we give a short formal proof. Note that for $f \in \mathcal{C}\left(\bar{\mathbb{N}}\right)$ and $s, t \geq 0$,

$$\mathrm{E}\left(f\left(Y_{s+t}\right) | X_u, Y_u, 0 \leq u \leq s\right) = \left(\boldsymbol{P}_t \Psi f\right)\left(X_s, Y_s\right),$$

where $\boldsymbol{P}_t$ is the semigroup generated by $\boldsymbol{G}$ and Ψ is a kernel given by $\Psi f = f \circ \psi$ where $\psi(x, y) = y$. Hence, in order to prove (8) we need to show that

$$\boldsymbol{P}_t \Psi f = \Psi Q_t f \tag{21}$$

for all $f \in \mathcal{C}\left(\bar{\mathbb{N}}\right)$. By Theorem 3 it suffices to prove that there exists a core D_H of H such that Ψ maps D_H into $\mathcal{D}(\boldsymbol{G})$ and

$$\boldsymbol{G}\Psi = \Psi H \tag{22}$$

on D_H. It is shown in Lemma 12 that

$$D_H = \left\{f \in \mathcal{C}\left(\bar{\mathbb{N}}\right) ; \exists y_0 \text{ s.t. } f(y) = f(\infty) \text{ for all } y > y_0\right\}$$

is a core of H and it is easy to see that for $f \in D_H$, Ψf is in $\mathcal{D}(\boldsymbol{G})$. Moreover, for $f \in D_H$ we have $(Gf(y)K(\cdot, y))(x) = f(y)(GK(\cdot, y))(x)$, so from (16) it is easy to see that (22) holds.

In order to prove the claims about X_t, we will use Theorem 9. In the present setting, $S = [0, 1]$ and $\phi(x, y) = x$. Define a probability kernel from $[0, 1]$ to $\boldsymbol{S}$ by

$$\begin{aligned}
\Lambda(x, A \times \{y\}) &= \delta_x(A) K(x, y) \\
\Lambda(x, (1, \infty)) &= K(x, \infty)
\end{aligned}$$

where x is in $[0, 1]$, A is in $\mathcal{B}[0, 1]$ and $y \in \mathbb{N}$. In other words,

$$\Lambda f(x) = \sum_{0 \leq y \leq \infty} K(x, y) f(x, y) \tag{23}$$

for $f \in \mathcal{C}(\boldsymbol{S})$ and $x \in [0, 1]$. Observe that $\pi_0^X \Lambda = \boldsymbol{\pi}_0^{(X,Y)}$. Also observe that for $f \in \mathcal{C}[0, 1]$ and $x \in [0, 1]$ we have

$$\Lambda \Phi f(x) = \sum_{0 \leq y \leq \infty} K(x, y) f(x) = f(x),$$

hence

$$\Lambda \Phi = I. \tag{24}$$

Let us now prove that

$$\Lambda \boldsymbol{P}_t = P_t \Lambda. \tag{25}$$

By Theorem 3, it suffices to prove that Λ maps $\mathcal{D}(\boldsymbol{G})$ into $\mathcal{C}^2[0, 1]$ and

$$\Lambda \boldsymbol{G} = G\Lambda \tag{26}$$

on $\mathcal{D}(\boldsymbol{G})$. Let f be in $\mathcal{D}(\boldsymbol{G})$. Then there is y_0 such that $f(x, y) = f(1, \infty)$ for $y > y_0$. Since we know that $\Lambda\boldsymbol{G}1 = G\Lambda1 = 0$, we may without loss of generality assume that $f(\infty) = 0$. Then

$$\Lambda f(x) = \sum_{y=0}^{y_0} K(x, y) f(x, y),$$

which is a polynomial, hence in $\mathcal{C}^2[0, 1]$.

By (16) we have for $x \in (0, 1)$ and $y \le y_0$ that

$$K(x, y)\,(\boldsymbol{G}f)\,(x, y) = K(x, y)\,(Hf\,(\cdot, x))\,(y) + G\,(f\,(\cdot, y)\,K\,(\cdot, y))\,(x)$$
$$-f(x, y)\,(GK\,(\cdot, y))\,(x). \quad (27)$$

Since both sides of the equality are continuous in x, the equality also holds for all $x \in [0, 1]$. Using (23), (27) and noting that $(\boldsymbol{G}f)\,(x, y) = 0$ for $y > y_0$ by (16) and $\boldsymbol{G}f(1, \infty) = 0$ by definition, we get

$$\Lambda\boldsymbol{G}f(x) = \sum_{y=0}^{y_0} K(x, y)\,(Hf\,(x, \cdot))\,(y) + \sum_{y=0}^{y_0} G\,(f\,(\cdot, y)\,K\,(\cdot, y))\,(x)$$
$$-\sum_{y=0}^{y_0} f(x, y)\,(GK\,(\cdot, y))\,(x). \quad (28)$$

The second term is just $G\Lambda f$, since $f(x, y) = 0$ for $y > y_0$. The first term can be rewritten as

$$\sum_{y=0}^{y_0} K(x, y) \sum_{z=0}^{y_0} H(y, z) f(x, z),$$

where we have again used that $f(x, z) = 0$ for $z > y_0$. The last term can be written as

$$\sum_{y=0}^{y_0} f(x, y)\,\left(GK\mathbf{1}_{\{y\}}\right)(x) \;=\; \sum_{y=0}^{y_0} f(x, y)\,\left(KH\mathbf{1}_{\{y\}}\right)(x)$$
$$=\; \sum_{y=0}^{y_0} f(x, y) \sum_{z=0}^{y_0} K(x, z) H(z, y)$$

where in the first equality we have used Theorem 4 and in the second equality we have used that H is an upper triangular matrix. Therefore, $\Lambda\boldsymbol{G}f = G\Lambda f$.

Finally, from (24) and (25) we get that $P_t = \Lambda\boldsymbol{P}_t\Phi$. Thus, we have verified all requirements of Theorem 9. It follows that X_t is the Wright–Fisher diffusion with reflection at zero with the initial distribution π_0^X and

$$\mathrm{P}\,(Y_t = y | X_s, 0 \le s \le t) \;=\; \mathrm{P}\,(\boldsymbol{X}_t \in [0, 1] \times \{y\} | X_s, 0 \le s \le t)$$
$$=\; \Lambda\,(X_t, [0, 1] \times \{y\})$$
$$=\; K\,(X_t, y)\ a.s.$$

for $y \in \bar{\mathbb{N}}$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

2.3. Derivation of the generator for the coupled process

In this section we show how formula (16) for the generator of the coupled process can be derived. Strictly speaking, this derivation is not necessary since (16) can be taken as a definition (and we therefore choose to make this derivation informal for the sake of brevity). However, we believe that this derivation can provide insight into the problem. We use the technique of Diaconis and Fill [1] who derived an analogous result for Markov processes with discrete space and time. Fill [4] then extended the result to Markov processes with continuous time and discrete space. In Fill's setting, P_t and Q_t are transition semigroups of Markov processes with discrete state-spaces $\mathcal{S}_1$ and $\mathcal{S}_2$. For a fixed $t > 0$, he defines a probability kernel on

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

by

$$\boldsymbol{P}^{(t)} f = Q_t \mathbf{1}_{[P_t K \neq 0]} \frac{P_t f K}{P_t K}, \tag{29}$$

where f is a function on $\mathcal{S}$. In (29), $P_t f K$ denotes the application of semigroup P_t to the product of functions f and K (here, K is not viewed as a kernel but simply as a function of x and y). Although P_t normally acts on functions of x, we make it act on functions of x and y by fixing y. Similarly, $P_t K$ denotes the application of P_t to K (this can alternatively be interpreted as the composition of the two kernels). Then we take the pointwise division of $P_t f K$ and $P_t K$, which we define to be zero if the denominator is zero. We then get $\boldsymbol{P}^{(t)} f$ as the application of Q_t to this function. This time, x is considered fixed when we apply Q_t.

The probability kernel $\boldsymbol{P}^{(t)}$ could be used to construct an intertwining of discrete-time processes where time is measures in steps of t. However, as a function of t, it turns out that $\boldsymbol{P}^{(t)}$ does not satisfy the Chapman-Kolmogorov equations and hence cannot be used to construct the coupled continuous-time process directly. Nevertheless, for small t, one can expect $\boldsymbol{P}^{(t)}$ to be "almost right". Indeed, Fill proves that there exists a generator $\boldsymbol{G}$ on $\mathcal{S}$ (for which he gives an explicit formula) such that

$$\frac{\boldsymbol{P}^{(t)} - I}{t} \to \boldsymbol{G} \tag{30}$$

as $t \downarrow 0$. He then shows the bivariate Markov process associated with $\boldsymbol{G}$ (with suitable initial distribution) has the desired properties, i.e. its margins on their own are processes with transition semigroups P_t and Q_t and satisfy (2).

Now we return back to our setting where P_t is the semigroup of the Wright–Fisher diffusion with reflection at zero and Q_t is the semigroup of an explosive pure birth process. Note that (29) is not a suitable definition in this case, since P_t operates on continuous functions, but the indicator in (29) can introduce discontinuity. To get around this problem, it can be proved that $\frac{P_t f K}{P_t K}$ can be extended to a continuous function (Hudec [6, Theorem 3.10] proves this for a slightly different kernel K, but his proof can easily be adapted to our setting). Now we can define $\boldsymbol{P}^{(t)}$ by

$$\boldsymbol{P}^{(t)} f = Q_t \frac{P_t f K}{P_t K}. \tag{31}$$

Observe that $P_t f K \to f K$, $P_t K \to K$ and $\frac{P_t f K}{P_t K} \to \frac{f K}{K} = f$ as $t \downarrow 0$ (provided we choose (x, y) such that $K(x, y) > 0$). Hence,

$$\frac{1}{t}\left(\boldsymbol{P}^{(t)} f - f\right) = \frac{1}{t}\left(Q_t - 1\right) \frac{P_t f K}{P_t K} + \frac{\frac{1}{t}\left((P_t - 1) f K\right)}{P_t K} - f \frac{\frac{1}{t}\left((P_t - 1) K\right)}{P_t K}$$

is expected to converge to

$$H f + \frac{G f K}{K} - f \frac{G K}{K}.$$

A. WRIGHT–FISHER DIFFUSION AND AN EXPLOSIVE PURE BIRTH PROCESS

Recall that G is defined by

$$G f(x) = \left(1 - x^2\right) \frac{\partial^2}{\partial x^2} f(x)$$

where x is in $[0, 1]$ and f is in $\mathcal{C}^2[0, 1]$ such that $\frac{\partial}{\partial x} f(0) = 0$.

Theorem 10. Operator G is closable and its closure generates a Feller semigroup. Moreover, the associated Markov process has continuous sample paths.

P r o o f . In order to prove that G is closable and generates a Feller semigroup, we will use Proposition 7. It is obvious that $G1 = 0$. Moreover, if we define L_n as the set of all even polynomials of order at most $2n$, then $\bigcup_{n \in \mathbb{N}} L_n$ is dense in $\mathcal{C}[0, 1]$ by the Stone–Weierstrass theorem and G maps L_n into L_n. Finally we prove that G satisfies the positive maximum principle. Let f be in $\mathcal{D}(G)$ 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, $G f(x_0) \leq 0$.

To prove that almost all sample paths are continuous, it suffices to show that for each $x_0 \in [0, 1]$ and $\epsilon > 0$ there exists $f \in \mathcal{D}(G)$ such that

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

(Ethier and Kurtz [3, Proposition 4.2.9 and Remark 4.2.10]). 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 $Gf(x_0) = 0$. □

Recall that $H : \mathbb{R}^{\bar{\mathbb{N}}} \to \mathbb{R}^{\bar{\mathbb{N}}}$ is an operator defined by

$$
\begin{aligned}
Hf(y) &= \lambda_y \left(f(y+1) - f(y) \right), \quad y \in \mathbb{N}, \\
Hf(\infty) &= 0,
\end{aligned}
$$

where

$$
\lambda_y = (2y+1)(2y+2), \quad y \in \mathbb{N}.
$$

Proposition 11. Define $D = \left\{ f \in \mathcal{C}\left(\bar{\mathbb{N}}\right); Hf \in \mathcal{C}\left(\bar{\mathbb{N}}\right) \right\}$. Then the restriction of H to D is the generator of a Feller semigroup on $\mathcal{C}\left(\bar{\mathbb{N}}\right)$.

P r o o f. We verify the conditions of Proposition 7. First note that $H1 = 0 \in \mathcal{C}\left(\bar{\mathbb{N}}\right)$, hence 1 is in D. Second, if we define L_n as the set of all functions $f \in \mathcal{C}\left(\bar{\mathbb{N}}\right)$ such that $f(y) = f(\infty)$ for all $y > n$, then $\bigcup_{n \in \mathbb{N}} L_n$ is dense in $\mathcal{C}\left(\bar{\mathbb{N}}\right)$ and H maps L_n to L_n. Finally, we verify the positive maximum principle. Let f be in D and y in $\bar{\mathbb{N}}$ such that

$$
\sup_{y \in \bar{\mathbb{N}}} f(y) = f(y_0).
$$

If $y_0 < \infty$, then

$$
Hf(y_0) = \lambda_{y_0} \left(f(y_0 + 1) - f(y_0) \right) \le 0,
$$

and if $y_0 = \infty$, then $Hf(y_0) = 0$, so H satisfies the positive maximum principle. □

Lemma 12. The set

$$
D_H = \left\{ f \in \mathcal{C}\left(\bar{\mathbb{N}}\right); \exists y_0 \text{ s.t. } f(y) = f(\infty) \, \forall y > y_0 \right\}
$$

is a core of H.

P r o o f. It is easy to see that D_H is dense in $\mathcal{C}\left(\bar{\mathbb{N}}\right)$. Moreover, since the process associated with H can only jump upward, the semigroup maps D_H into itself. The statement of the lemma now follows from Proposition 1.3.3 in Ethier and Kurtz [3]. □

REFERENCES

[1] P. Diaconis and J. A. Fill: Strong stationary times via a new form of duality. Ann. Probab. *18* (1990), 4, 1483–1522. DOI:10.1214/aop/1176990628

[2] P. Diaconis and L. Miclo: On times to quasi-stationarity for birth and death processes. J. Theoret. Probab. *22* (2009), 3, 558–586. DOI:10.1007/s10959-009-0234-6

[3] S. N. Ethier and T. G. Kurtz: Markov Processes: Characterization and Convergence. John Wiley and Sons, 1986. DOI:10.1002/9780470316658

[4] J. A. Fill: Strong stationary duality for continuous-time markov chains. Part I: Theory. J. Theoret. Probab. *5* (1992), 1, 45–70. DOI:10.1007/bf01046778

[5] J. A. Fill and V. Lyzinski: Strong stationary duality for diffusion processes. J. Theoret. Probab. *29* (2016), 4, 1298–1338. DOI:10.1007/s10959-015-0612-1

[6] T. Hudec: Absorption Cascades of One-dimensional Diffusions. Master's Thesis, Charles University in Prague, 2016. Available at `https://is.cuni.cz/webapps/zzp/detail/92318?lang=en`

[7] S. Karlin and J. McGregor: Coincidence properties of birth and death processes. Pacific J. Math. *9* (1959), 4, 1109–1140. DOI:10.2140/pjm.1959.9.1109

[8] J. T. Kent: The spectral decomposition of a diffusion hitting time. Ann. Probab. *10* (1082), 1, 207–219. DOI:10.1214/aop/1176993924

[9] T. M. Liggett: Continuous Time Markov Processes: An Introduction. American Mathematical Soc., 2010. DOI:10.1090/gsm/113

[10] P. Mandl: Analytical Treatment of One-dimensional Markov Processes. Springer, 1968.

[11] L. C. G. Rogers and J. W. Pitman: Markov functions. Ann. Probab. *9* (1981), 4, 573–582. DOI:10.1214/aop/1176994363

[12] J. M. Swart: Intertwining of birth-and-death processes. Kybernetika *47* (2011), 1, 1–14.

Tobiáš Hudec, Institute of Information Theory and Automation, The Czech Academy of Sciences, Pod Vodárenskou věží 4, 182 08 Praha 8. Czech Republic.
    e-mail: tobias.hudec@gmail.com