INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

The Formal Theory of Birth-and-Death Processes, Lattice Path Combinatorics, and Continued Fractions

Philippe Flajolet, Fabrice Guillemin

N° 3667
Avril 1999

THÈME 2

Rapport de recherche

ISSN 0249-6399

The Formal Theory of Birth-and-Death Processes, Lattice Path Combinatorics, and Continued Fractions

Philippe Flajolet, Fabrice Guillemin

Thème 2 — Génie logiciel
et calcul symbolique
Projet Algo

Rapport de recherche n ° 3667 — Avril 1999 — 25 pages

Abstract: Classic works of Karlin-McGregor and Jones-Magnus have established a general correspondence between continuous-time birth-and-death processes and continued fractions of the Stieltjes-Jacobi type together with their associated orthogonal polynomials. This fundamental correspondence is revisited here in the light of the basic relation between weighted lattice paths and continued fractions otherwise known from combinatorial theory. Given that trajectories of the embedded Markov chain of a birth-and-death process are lattice paths, Laplace transforms of a number of transient characteristics can be obtained systematically in terms of a fundamental continued fraction and its family of convergent polynomials. Applications include the analysis of evolutions in a strip, upcrossing and downcrossing times under flooring and ceiling conditions, as well as time, area, or number of transitions while a geometric condition is satisfied.

Key-words: Lattice path combinatorics, continued fractions, orthogonal polynomials, birth-and-death process, first passage time, excursions, transient characteristics.

(Résumé : tsvp)

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
Téléphone : (33) 01 39 63 55 11 – Télécopie : (33) 01 39 63 53 30

Théorie formelle des processus de naissance et de mort, combinatoire des chemins, et fractions continues

Résumé : Les travaux classiques de Karlin-McGregor et Jones-Magnus ont établi une correspondance générale entre les processus de naissnace et de mort en temps continu et les fractions continues de type Stieltjes-Jacobi ainsi que les polynômes orthogonaux associés. Cette correspondance fondamentale est rééxaminée ici à la lumière de la relation connue en analyse combinatoire qui lie chemins valués et fractions continues, Étant donné que les trajectoires de la chaîne de Markov incluse sont des chemins, les transformées de Laplace d'un grand nombre de caractéristique transitoires peuvent être obtenues systématiquement. Les applications incluent l'analyse des évolutions dans une bande, des traversées de bande, ainsi que le temps, l'aire, ou le nombre de transitions sous conditions géométriques.

Mots-clé : Combinatoire des chemins, fractions continues, polynômes orthogonaux, processus de naissance et de mort, premiers passages, excursions, caracteristiques transitoires

Version of April 19, 1999

THE FORMAL THEORY OF BIRTH-AND-DEATH PROCESSES, LATTICE PATH COMBINATORICS, AND CONTINUED FRACTIONS

PHILIPPE FLAJOLET AND FABRICE GUILLEMIN

ABSTRACT. Classic works of Karlin-McGregor and Jones-Magnus have established a general correspondence between continuous-time birth-and-death processes and continued fractions of the Stieltjes-Jacobi type together with their associated orthogonal polynomials. This fundamental correspondence is revisited here in the light of the basic relation between weighted lattice paths and continued fractions otherwise known from combinatorial theory. Given that trajectories of the embedded Markov chain of a birth-and-death process are lattice paths, Laplace transforms of a number of transient characteristics can be obtained systematically in terms of a fundamental continued fraction and its family of convergent polynomials. Applications include the analysis of evolutions in a strip, upcrossing and downcrossing times under flooring and ceiling conditions, as well as time, area, or number of transitions while a geometric condition is satisfied.

Keywords: Lattice path combinatorics, continued fractions, orthogonal polynomials, birth-and-death process, first passage time, excursions, transient characteristics.

INTRODUCTION

Many systems occurring in diverse fields of applications, like physics, modelling of telecommunications networks, queueing theory, population growth, or analysis of dynamic data structures, are described by means of transient characteristics associated with a birth-and-death process. In favourable situations, explicit forms for probability distribution functions or at least for Laplace transforms of the main characteristics of interest can be obtained. This is often achieved via methods that combine general properties, like the Chapman-Kolmogorov equations or the strong Markov property, together with *ad hoc* calculations that reflect probabilistic decompositions of sorts.

Various aspects of the connection between birth-and-death processes, continued fractions and orthogonal polynomial systems have been investigated early in the literature. The first outstanding results date back to the papers by Karlin and McGregor in the 1950's [28, 29, 30], where it was shown that the Chapman-Kolmogorov equation satisfied by the transition probabilities of a birth-and-death process can be solved via the introduction of an orthogonal polynomial system and a spectral measure. This was notably used in the study of transience and recurrence in a purely analytic way. (The results of Karlin and McGregor are thus of great interest to the community dealing with the properties of orthogonal polynomials; see for instance [24].) Proceeding along different lines, Jones and Magnus developed around 1977 a direct treatment of some Laplace transforms associated to birth-and-death processes by means of continued fractions [26, 27].

In this paper, we adopt a radically different point of view. We first note that transient characteristics of a birth-and-death process are frequently defined by means of geometric conditions, for instance the time spent by the process in a set of states or the state of the process at an arbitrary instant, starting from given initial conditions. The trajectories of the process are lattice path, and an adequate morphism provides a representation of the Laplace transforms of probabilities of the corresponding transient characteristics (Section 1). On the other hand, the combinatorial theory due to Flajolet *at al* [10, 11, 12, 13] leads to a formal representation of geometric classes of lattice paths in terms of continued fractions and their associated orthogonal polynomials (Section 2). Accordingly, the results of the paper lead directly to a representation of the Laplace transform of a transient characteristic in terms of continued fractions (Section 3) and othogonal polynomials (Section 4). This approach not only allows us to recover in a simple way a number of results often obtained previously by special-purpose methods, but also to derive new results somewhat difficult to obtain by straight probabilistic arguments.

As is well-known, continued fractions prove useful, either symbolically or numerically. Indeed, symbolically, a large number of continued fraction expansions of formal power series or analytic functions are known and closed form expressions in terms of classical special functions can often be attained by reverse-engineering and table lookup in classical treatises like Wall's [36] or Chihara's [7]. This is especially useful for processes whose birth-and-death rates are given by "regular laws", typically $M/M/1$ or $M/M/\infty$ queues, population models, etc; see Section 5. Numerically, it is to be noted that continued fractions are also instrumental in the evaluation of underlying probability distributions, although this topic is out of the scope of the paper. Continued fraction based analyses of specials processes are to be found in [18, 19, 35] for the distribution of transient characteristics in an $M/M/\infty$ system and in [20] for the tail behaviour of the cumulative waiting time in a busy period of an $M/M/1$ queue; see also the arguments developed by Abate and Whitt in [1, 2] in combination with their Laplace transform inversion algorithm specified in [3].

The main objective of this paper as regards birth-and-death processes is to separate clearly the formal apparatus from the analytic-probabilistic machinery, and neatly delineate parameters that are amenable to a systematic treatment by means of continued fractions and orthogonal polynomials. Section 1 describes the basic connection between lattice path theory and birth-and-death processes. The Laplace transform of a transient characteristic specified by a geometric condition is shown to derive throughby an appropriate morphism from the combinatorial generating function of a set of lattice paths. In Section 2, several results from the literature on lattice path combinatorics are gathered and cast in a unified and self-contained framework. The theory exposed in the first two sections is applied next in Sections 3, 4 where several analyses, some old and some new, are obtained in this way for transient characteristics of a general birth-and-death process. We finally outline a specialization of this theory to three of the classical models ($M/M/1$, $M/M/\infty$, population growth) in Section 5.

1. Birth-and-death processes

Consider a (continuous-time) *birth-and-death* process (BDP) $\{\Lambda_t\}$, taking values in $\mathbb{N}$ and with transition rate $q_{i,j}$ from state i to state j defined by

$$q_{0,0} = -\lambda_0,\ q_{0,1} = \lambda_0,$$
$$q_{i,i-1} = \mu_i,\ q_{i,i} = -(\lambda_i + \mu_i),\ q_{i,i+1} = \lambda_i,\ i \geq 1,$$

with $q_{i,j} = 0$ otherwise, where the constants λ_i for $i \geq 0$ and μ_j for $j \geq 1$ are strictly positive. Let $\{Y_n\}$ denote the (discrete time) *Markov chain* embedded at jump instants. The transition matrix $\{p_{i,j}\}$ of $\{Y_n\}$ is given by

$$p_{0,1} = 1;\ p_{i,i-1} = \frac{\mu_i}{\lambda_i + \mu_i},\ p_{i,i+1} = \frac{\lambda_i}{\lambda_i + \mu_i},\ i \geq 1; \tag{1.1}$$

with $p_{i,j} = 0$ otherwise. It proves occasionally convenient to consider an extension of the process with a death rate μ_0 from state 0 and absorption at -1; the extended process gives back the basic process upon setting $\mu_0 = 0$.

In order to simplify the discussion, we assume throughout the paper that the birth-and-death process is ergodic. This implies in particular [5]

$$\sum_{n=0}^{\infty} \pi_n < \infty \text{ and } \sum_{n=0}^{\infty} \frac{1}{\lambda_n \pi_n} = \infty, \tag{C$_1$}$$

where the fundamental quantities π_n are defined by

$$\pi_0 = 1;\ \pi_n = \frac{\lambda_0 \ldots \lambda_{n-1}}{\mu_1 \ldots \mu_n},\ n \geq 1. \tag{1.2}$$

The π_n are known to express the long term behaviour of the system [31], the stationary probability of state j being

$$\mathbf{p}_j = \frac{\pi_j}{\sum \pi_k}. \tag{1.3}$$

Trajectories of the Markov chain $\{Y_n\}$ are sequences of nonnegative integers that can be alternatively described as *lattice paths* defined as follows.

Definition 1 (Lattice path). A (lattice) *path* $\upsilon = (U_0, U_1, \ldots, U_n)$ is a sequence of points in the lattice $\mathbb{N} \times \mathbb{N}$ such that if $U_j = (x_j, y_j)$, then $x_j = j$ and $|y_{j+1} - y_j| \le 1$. An edge $\langle U_j, U_{j+1}\rangle$ is called an *ascent* ($\underline{a}$) if $y_{j+1} - y_j = +1$, a *descent* ($\underline{b}$) if $y_{j+1} - y_j = -1$, and a *level step* ($\underline{c}$) if $y_{j+1} - y_j = 0$.
The quantity n is the *length* of the path, $o(\upsilon) := y_0$ is the *initial altitude*, $h(\upsilon) := y_n$ is the *final altitude*. The extremal quantities $\sup\{\upsilon\} := \max_j y_j$ and $\inf\{\upsilon\} := \min_j y_j$ are called the *height* and *depth* of the path.

It is assumed that paths are normalized by the condition $x_0 = 0$, the variable x_n denoting the (discrete) "time" parameter. With this normalization, a path of length n is encoded by a word with a, b, c representing ascents, descents, and level steps, respectively. What we call the *standard encoding* is such a word in which each step a, b, c is (redundantly) subscripted by the value of the y-coordinate of its associated point. For instance, $w = a_0a_1a_2b_3c_2c_2a_2b_3b_2b_1a_0c_1$ encodes a path that connects the initial state $U_0 = (0, 0)$ to the state $U_{12} = (12, 1)$. We freely identify a path υ defined as a sequence of points and its standard encoding w. With this convention, a sample path of the Markov chain $\{Y_n\}$ is in turn identified with a word without level steps. Inclusion of level steps in Definition 1 proves occasionally useful in some parts of the discussion: see (2.28), (3.11), and (3.20).

Let H be the set of trajectories ("histories" in the terminology of analysis of algorithms) of the BDP. Given a geometric condition (Q) that is expressed solely in terms of geometric constraints on trajectories, it is then possible to associate to it a "language" $H[Q]$ that comprises the collection of all path encodings satisfying the condition Q. This language can be viewed either as a set or as a formal sum,

$$H[Q] = \sum_{\{w \mid Q\}} w,$$

in which case it becomes the generating function in infinitely many indeterminates of the corresponding condition.

The major thread of this paper is the connection between formal languages expressing properties of trajectories and Laplace transforms. The Laplace transform of a function $g(x)$ is taken here as

$$\widetilde{g}(s) = \int_0^\infty e^{-sx} g(x)\, dx,$$

and the Laplace transform $\phi_X(s) = \mathbb{E}[e^{-sX}]$ of a random variable X is the transform of its density function if it exists. In particular, the transform of an exponential variable of rate a with probability density function ae^{-ax} is $a/(s + a)$. In the context of this paper, a *morphism* χ is a linear application from the set of words to an algebra of functions satisfying $\chi(u \cdot v) = \chi(u)\chi(v)$.

Proposition 1 (Basic Connection). *Consider a process with transition rates $\{\lambda_j\}$ and $\{\mu_j\}$, and the probabilities*

$$\mathbb{P}_Q(\tau) = \Pr\{\{\Lambda_t\} \text{ satisfies } Q \text{ at time } \tau\}. \tag{1.4}$$

Let χ_s be the "probabilistic morphism", defined by

$$\chi_s(a_j) = \frac{\lambda_j}{s + \lambda_j + \mu_j},\ \chi_s(b_j) = \frac{\mu_j}{s + \lambda_j + \mu_j},\ \chi_s(c_j) = 0. \tag{1.5}$$

Then, the Laplace transform of the function $\tau \mapsto \mathbb{P}_Q(t)$ is given for $s \ge 0$ by

$$\widetilde{\mathbb{P}}_Q(s) = \sum_{w \in H[Q]} \left(\chi_s(w) \cdot \frac{1}{s + \lambda_{h(w)} + \mu_{h(w)}} \right), \tag{1.6}$$

where $h(w)$ is the final altitude of the path encoding w.

Proof. The proof closely parallels the elementary construction of continuous-time Markov chains; see for instance [31, Ch. 4] or [34, Sec. 2.6]. We start by conditioning upon a particular trajectoty of the embedded chain. Given a lattice path $w = w_1 w_2 \cdots w_n \in H$, the probability that the process satisfies the condition Q at time t, having followed the path $w \in H$,

$$\mathbb{P}_{Q,w}(t) := \Pr\{\{\Lambda_u\} \text{ satisfies } Q \text{ at time } t \ / \ w\}$$

satisfies

$$\mathbb{P}_{Q,w}(t) = \Pr\left\{S_{e_0} + \cdots + S_{e_{n-1}} \le t,\ S_{e_0} + \cdots + S_{e_{n-1}} + S_{e_n} > t\right\}, \tag{1.7}$$

where S_{e_j} is the random variable that represents the sojourn time at the state e_j determined by $w_1 \cdots w_j$. It is elementary that the Laplace transform of a sum of independent random variables is a product, $\phi_{X+Y}(s) = \phi_X(s)\phi_Y(s)$, and that the transform of the probability of the event $\Pr\{X \le t < X + Y\}$ is $\phi_X(s)(1 - \phi_Y(s))/s$. Also, the sojourn time at some state i is exponential with parameter $\lambda_i + \mu_i$, so that its Laplace transform is $(\lambda_i + \mu_i)/(s + \lambda_i + \mu_i)$. Thus, owing to the strong Markov property satisfied by the BDP $\{\Lambda_t\}$, the transform of the probability of (1.7) is

$$\widetilde{\mathbb{P}}_{Q,w}(s) = \left(\prod_{j=0}^{n-1} \frac{\lambda_{e_j} + \mu_{e_j}}{s + \lambda_{e_j} + \mu_{e_j}}\right) \cdot \frac{1}{s + \lambda_{e_n} + \mu_{e_n}}. \tag{1.8}$$

The next observation (the strong Markov property still) is that the probability of a path w is the product of the individual transition probabilities, that is,

$$\alpha_w = p_{e_0,e_1} p_{e_1,e_2} \cdots p_{e_{n-1},e_n}, \tag{1.9}$$

where $p_{i,j}$ is defined by (1.1). Equivalently, $\alpha_w = \zeta_{e_0} \cdots \zeta_{e_{n-1}}$, where $\zeta_j = \lambda_j/(\lambda_j + \mu_j)$ or $\zeta_j = \mu_j/(\lambda_j + \mu_j)$ depending on whether the letter w_j is a or b. There results that the Laplace transform of the function $\mathbb{P}_Q$ is a sum over all lattice paths in $H[Q]$,

$$\widetilde{\mathbb{P}}_Q(s) = \sum_{w \in H[Q]} \alpha_w \widetilde{\mathbb{P}}_{Q,w}(s),$$

where the quantities $\widetilde{\mathbb{P}}_{Q,w}(s)$ and α_w are given by (1.8) and (1.9). The statement of the proposition follows. ☐

It is often the case that paths satisfying Q all end at the same altitude f; in that event, equation (1.6) further simplifies to

$$\widetilde{\mathbb{P}}_Q(s) = \chi_s\left(H[Q]\right) \frac{1}{s + \lambda_f + \mu_f},.$$

The next statement requires a proper definition of stopping conditions. Given two words u and v, we say that v is a strict prefix of u if $u = vw$ for some nonempty w. A language is prefix-free iff it does not contain two words u, v with u a strict prefix of v. If the set of paths $H[Q]$ associated to some condition Q is prefix-free, then it is said to be a stopping set, and accordingly Q is said to be a stopping condition. The stopping time $\theta \equiv \theta_Q$ associated with Q is then the random variable that represents the time at which the process first has a trajectory that belongs to $H[Q]$. (The prefix-free constraint on $H[Q]$ precisely formalizes the fact the problem is well-posed: no trajectory satisfying Q can extend another trajectory satisfying Q and having fewer steps.) Note that a stopping time may be a defective random variable if Q is not hit with probability 1, in which case we take $\theta = +\infty$.

Proposition 2 (Time connection). *Let Q be a stopping condition, with θ the stopping time and $H[Q]$ the corresponding set of paths. The Laplace transform $\widetilde{\theta}(s)$ of the random variable θ is given by*

$$\widetilde{\theta}(s) := \mathbb{E}\left[e^{-s\theta}\right] = \chi_s(H[Q]) \tag{1.10}$$

where χ_s is the probabilistic morphism defined by eq. (1.5).

Proof. The proof only differs marginally from that of Proposition 1. Since Q is a stopping condition, the probability $\mathbb{P}'_{Q,w}(t)$ of the process satisfying Q before time t, having followed path w, obeys a simplified form of (1.7),

$$\Pr\left\{S_{e_0}+\cdots+S_{e_{n-1}}+S_{e_n}<t\right\},$$

The fact that $H[Q]$ is a stopping set grants that the events corresponding to different w's are disjoint, and the proof concludes by the same computation as before. ☐

If the stopping time θ is defective, then, by general principles, the probability of θ being finite is

$$\Pr\{\theta<+\infty\}=\chi_0(H[Q]), \tag{1.11}$$

where χ_0 is the morphism χ_s instantiated at $s=0$. Accordingly, the Laplace transform of θ conditioned upon $\theta<+\infty$ is

$$\mathbb{E}\left[e^{-s\theta}\mid\theta<+\infty\right]=\frac{\chi_s(H[Q])}{\chi_0(H[Q])}. \tag{1.12}$$

The procedure can be generalized to derive the Laplace transform of a number of random variables associated with a birth-and-death process. An instance is the total area $\mathcal{A}_c$ swept under the BDP above a given threshold c until the stopping condition Q is satisfied (at the random time θ)

$$\mathcal{A}_c=\int_0^\theta(\Lambda_t-c)^+\,dt. \tag{1.13}$$

This parameter has provided one of the motivations for the present study. Its practical importance devolves from the fact that it measures the volume of traffic in queueing systems and related telecommunication contexts; see [19, 20, 21, 35].

Proposition 3 (Area connection)**.** *Let Q be a stopping condition. The joint Laplace transform of the (time, area)–pair $(\theta,\mathcal{A}_c)$ defined by $\mathcal{G}(s,y):=\mathbb{E}\left[e^{-s\theta-y\mathcal{A}_c}\right]$, is given by*

$$\mathcal{G}(s,y)=\chi_s^\circ(H[Q]), \tag{1.14}$$

where χ_s° is the enriched "area morphism" defined by

$$\chi_s^\circ(a_j)=\frac{\lambda_j}{s+(j-c)^+y+\lambda_j+\mu_j},\ \chi_s^\circ(b_j)=\frac{\mu_j}{s+(j-c)^+y+\lambda_j+\mu_j},\ \chi_s^\circ(c_j)=0. \tag{1.15}$$

Proof. Without loss of generality, consider the case $c=0$. The property that the stopping time occurs before time t and the area is less than v, following the path w, is the event

$$\{\theta<t,\mathcal{A}_0<v\ /\ w\}=\{S_{e_0}+S_{e_1}+\cdots+S_{e_n}<t,\ e_0S_{e_0}+e_1S_{e_1}+\cdots+e_nS_{e_n}<v\}. \tag{1.16}$$

Let X_j be independent exponential variables with rates κ_j and consider the two variables $U=\sum X_j$, $V=\sum\delta_jX_j$, for some system of weights δ_j. It suffices to observe the relation

$$\mathbb{E}[e^{-sU-yV}]=\prod_j\frac{\kappa_j}{s+\kappa_j+\delta_jy},$$

and the proof concludes like before, on the basis of (1.16). ☐

Example. Consider the condition E that represents the event "*no death has occurred*" and the closely related stopping condition F expressing "*the first death has already occurred*". Then $H[E]$ and $H[F]$ are

$$H[E]=1+a_0+a_0a_1+a_0a_1a_2+\cdots,\qquad H[F]=b_0+a_0b_1+a_0a_1b_2+\cdots,$$

where we must take $b_0=0$ and accordingly $\mu_0=0$. Then, Propositions 1 and 2 give (with $\prod_{j=0}^{-1}=1$)

$$\widetilde{\mathbb{P}}_E(s)=\sum_{n=0}^\infty\frac{\prod_{j=0}^{n-1}\lambda_j}{\prod_{j=0}^n(s+\lambda_j+\mu_j)},\qquad\widetilde{\theta}_F(s)=\sum_{n=0}^\infty\frac{\left(\prod_{j=0}^{n-1}\lambda_j\right)\mu_n}{\prod_{j=0}^n(s+\lambda_j+\mu_j)}.$$

Under ergodicity, we have $\widetilde{\theta}_F(0) = 1$, meaning that the stopping time θ_F is nondefective. In addition, the area $\mathcal{A}_0$ is taken into account by the joint Laplace transform,

$$\mathcal{G}(s,y) = \sum_{n=0}^{\infty} \frac{\left(\prod_{j=0}^{n-1} \lambda_j\right) \mu_n}{\prod_{j=0}^{n}(s+\lambda_j+\mu_j+jy)}.$$

Our purpose in this paper is to apply this formal procedure to less obvious transient characteristics of the birth-and-death process. This requires first setting up expressions for the multivariate generating functions expressing corresponding conditions.

2. Combinatorics of lattice paths and continued fraction representations

It is known that the formal theory of continued fraction expansions for power series is to a large extent equivalent to the combinatorial theory of weighted lattice paths. The goal of this section is to establish a representation for various sets of paths constrained by height and depth conditions, as arises in connection with transient characteristics of a BDP. For this purpose, we recast in a unifying framework and extend basic results established in [10, 11, 12, 13, 17] Proofs are given in this paper because they lead to results of independent interest. All the computations eventually depend only on a simple set of combinatorial decompositions (Section 2.1) and on basic properties of linear fractional transformations represented by 2×2 matrices (Sections 2.2), resulting in a wide of set of expressions (Section 2.3).

2.1. **Lattice path decompositions.** Let H denote the collection of all path encodings. A general subclass of paths of interest is defined by flooring (m), ceiling (h), as well as fixing initial (k) and final (l) altitudes

$$H_{k,l}^{[\geq m,<h]} \quad = \quad \{w \in H : o(w)=k,\ h(w)=l,\ \inf\{w\} \geq m, \text{ and } \sup\{w\} < h\}.$$

We also need the specializations,

$$H_{k,l}^{[<h]} = H_{k,l}^{[\geq 0,<h]}, \quad H_{k,l}^{[\geq m]} = H_{k,l}^{[\geq m,<\infty]}, \quad H_{k,l} = H_{k,l}^{[\geq 0,<\infty]}.$$

Consider now the symbols a_j, b_j, c_j as formal indeterminates, so that words become monomials. The *generating function* of a collection $\mathcal{C}$ of paths is defined as the formal sum

$$C = \sum_{w \in \mathcal{C}} w,$$

a formal power series in infinitely many indeterminates. (No technical difficulty arises if one operates with the gradation on the set of variables defined by $\deg a_j = \deg b_j = \deg c_j = 1$.) In the same way as words are assimilated to monomials, it proves convenient to use consistently the same notation for sets and for the associated generating functions.

Basic combinatorial principles [16] state that (disjoint) set union, (unambiguous) concatenation of sets of words, and the (unambiguous) formation of arbitrary sequences,

$$C = A \cup B, \qquad C = A \cdot B, \qquad C = \mathtt{Seq}(A) := \bigcup_{n \geq 0} A^n,$$

translate into generating functions as

$$C = A + B, \qquad C = A \times B, \qquad C = \frac{1}{1-A}.$$

There, the last identity derives from the fact that $(1-f)^{-1} = 1 + f + f^2 + \cdots$ generates symbolically all the sequences with components f.

Three obvious *combinatorial decompositions* of paths then suffice to derive all the basic formulæ.

Arch decomposition: An excursion from and to level 0 consists of a sequence of "arches", each made of either a c_0 or a $a_0 H_{1,1}^{[\geq 1]} b_1$, so that

$$H_{0,0} = (1 - c_0 - a_0 H_{1,1}^{[\geq 1]} b_1)^{-1}.$$

which relativizes to height $< h$: in general, one has the recursion

$$\left\{ \begin{array}{lcl} H_{j,j}^{[\geq j,<h]} & = & \left(1 - c_j - a_j H_{j+1,j+1}^{[\geq j+1,<h]} b_{j+1}\right)^{-1}, \\ H_{h-1,h-1}^{[\geq h-1;<h]} & = & (1-c_{h-1})^{-1}. \end{array} \right. \tag{2.1}$$

Last passages decompositions: Recording the times at which each level $0, \ldots, k$ is last traversed gives

$$H_{0,k} = H_{0,0}^{[\geq 0]} a_0 H_{1,1}^{[\geq 1]} a_1 \cdots a_{k-1} H_{k,k}^{[\geq k]} \tag{2.2}$$

First passage decomposition: The quantities $H_{k,l}$ with $k \leq l$ are implicitly determined by the first passage through k in a path connecting level 0 to l; a dual decomposition holds when $k \geq l$, so that

$$\left\{ \begin{array}{lcll} H_{0,l} & = & H_{0,k-1}^{[<k]} a_{k-1} H_{k,l} & (k \leq l), \\ H_{k,0} & = & H_{k,l} b_l H_{l-1,0}^{[<l]} & (l \leq k). \end{array} \right. \tag{2.3}$$

2.2. **Continued fraction and convergent polynomials.** The basic results of earlier studies express the generating functions of three basic types of paths in terms of a fundamental continued fraction and its associated convergent polynomials. They involve the "numerator" and "denominator" polynomials, denoted by P_h and Q_h that each satisfy the second order (or "tree-term") recurrence

$$Y_{h+1} = (1-c_h)Y_h - a_{h-1}b_h Y_{h-1}, \ h \geq 1, \tag{2.4}$$

and the initial conditions $(P_{-1}, Q_{-1}) = (1, 0)$ and $(P_0, Q_0) = (0, 1)$, with the convention $a_{-1}b_0 = 1$.

Theorem 1 (Path continued fractions [10]). *(i) The set $H_{0,0}$ of all basic excursions is represented by the fundamental continued fraction:*

$$H_{0,0} = \cfrac{1}{1 - c_0 - \cfrac{a_0 b_1}{1 - c_1 - \cfrac{a_1 b_2}{1 - c_2 - \cfrac{a_2 b_3}{\ddots}}}}. \tag{2.5}$$

(ii) The set of ceiled excursions $H_{0,0}^{[<k]}$ is given by convergents of the fundamental fraction:

$$H_{0,0}^{[<h]} = \cfrac{1}{1 - c_0 - \cfrac{a_0 b_1}{1 - c_1 - \cfrac{a_1 b_2}{1 - c_2 - \cfrac{a_2 b_3}{\cfrac{\ddots}{1 - c_{h-1}}}}}} \tag{2.6}$$

$$= \frac{P_h}{Q_h}. \tag{2.7}$$

(iii) The set of floored excursions is given by the truncation of the fundamental fraction:

$$H_{h,h}^{[\geq h]} = \cfrac{1}{1 - c_h - \cfrac{a_h b_{h+1}}{1 - c_{h+1} - \cfrac{a_{h+1} b_{h+2}}{1 - c_{h+2} - \cfrac{a_{h+2} b_{h+3}}{\ddots}}}} \tag{2.8}$$

$$= \frac{1}{a_{h-1}b_h} \frac{Q_h H_{0,0} - P_h}{Q_{h-1} H_{0,0} - P_{h-1}}, \tag{2.9}$$

Proof. Repeated use of the arch decomposition (2.1) provides a form of $H_{0,0}^{[<h]}$ with nested quasi-inverses $(1-f)^{-1}$ that is the finite fraction representation (2.6). The continued fraction representation for basic paths (namely $H_{0,0}$) is then obtained by letting $h \to \infty$ in (2.6). Finally, the continued fraction form (2.8) for ceiled excursions is nothing but the fundamental form (2.5), when the indices are shifted. The three continued fraction expressions (2.5), (2.6), (2.8) are thence established.

Finding explicit expressions for the fractions $H_{0,0}^{[<h]}$ and $H_{h,h}^{[\geq h]}$ next requires determining the polynomials that appear in the convergents of the basic fraction (2.5). By definition, the convergent polynomials P_h and Q_h are the numerator and denominator of the fraction $H_{0,0}^{[<h]}$.

For the computation of $H_{0,0}^{[<h]}$ and P_h, Q_h, one classically introduces the linear fractional transformations

$$g_j(y) = \frac{1}{1 - c_j - a_j b_{j+1} y},$$

so that

$$H_{0,0}^{[<h]} = g_0 \circ g_1 \circ g_2 \circ \cdots \circ g_{h-1}(0) \text{ and } H_{0,0} = g_0 \circ g_1 \circ g_2 \circ \cdots\ ,. \tag{2.10}$$

Now, linear fractional transformations are representable by 2×2-matrices

$$\frac{ay+b}{cy+d} \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{2.11}$$

in such a way that composition corresponds to matrix product. By induction on the compositions that build up $H_{0,0}^{[<h]}$, there follows the equality

$$g_0 \circ g_1 \circ g_2 \circ \cdots \circ g_{h-1}(y) = \frac{P_h - P_{h-1} a_{h-1} b_h y}{Q_h - Q_{h-1} a_{h-1} b_h y}, \tag{2.12}$$

where P_h and Q_h satisfy the recurrence (2.4). Setting $y = 0$ in(2.12) proves (2.7).

Finally, $H_{h,h}^{[\geq h]}$ is determined implicitly as the root y of the equation $g_0 \circ \cdots \circ g_{h-1}(y) = H_{0,0}$, an equation that, when solved using (2.12), yields (2.9). □

The proof of Theorem 1 relies on the decomposition of trajectories into arches that are themselves built of a matching ascent-descent pair and another trajectory one level up, with the pattern repeating itself forever. The continued fraction expressions directly reflect this decomposition. Then, the (2×2) matrix representations that are classically associated to continued fraction algebra open the way to a systematic calculus. In the sequel, much use is made of this line of reasoning.

An immediate corollary is a representation for the set $H_{m,m}^{[\geq m,<h]}$ of upward excursions that is simply obtained by replacing $H_{0,0}$ with P_h/Q_h in relation (2.9).

Corollary 1 (Upward excursions). *The set $H_{m,m}^{[\geq m,<h]}$ is represented as*

$$H_{m,m}^{[\geq m,<h]} = \frac{1}{a_{m-1} b_m} \frac{Q_m P_h - P_m Q_h}{Q_{m-1} P_h - P_{m-1} Q_h}. \tag{2.13}$$

2.3. **Sets of paths satisfying height and depth conditions.** The main result of this section is Theorem 2 that is contained in [17] and is closely related to earlier combinatorial studies like [10, 12]. It provides a formula for "transitions", that is, paths of general type $H_{k,l}$, in the presence of general flooring and ceiling constraints. We build up the proof in stages as it allows us to derive intermediate results of independent interest. These intermediate results are themselves corollaries of Theorem 1. To alleviate the notation, we use the abbreviation:

$$A_m = a_0 a_1 \cdots a_{m-1} \text{ and } B_m = b_1 b_2 \cdots b_m. \tag{2.14}$$

Corollary 2 (Crossings [10]). *The set $H_{0,h-1}^{[<h]}$, $h \geq 1$, of upcrossings from state 0 to $h-1$ that stay below level h is represented as*

$$H_{0,h-1}^{[<h]} = \frac{A_{h-1}}{Q_h}. \tag{2.15}$$

	Type	*Spec.*	*Formula*	*Ref.*
1.	Excursion	$H_{0,0}$	$\dfrac{1}{1-c_0-\dfrac{a_0b_1}{1-c_1-\cdots}}$	(2.5)
2.	Ceiled excursions	$H^{[<h]}_{0,0}$	$\dfrac{P_h}{Q_h}$	(2.7)
3.	Floored excursions	$H^{[\geq h]}_{h,h}$	$\dfrac{1}{a_{h-1}b_h}\dfrac{Q_hH_{0,0}-P_h}{Q_{h-1}H_{0,0}-P_{h-1}}$	(2.9)
4.	Transitions from 0	$H_{0,l}$	$\dfrac{1}{B_l}\left(Q_lH_{0,0}-P_l\right)$	(2.17)
5.	Transitions to 0	$H_{k,0}$	$\dfrac{1}{A_k}(Q_kH_{0,0}-P_k)$	(2.19)
6.	Upcrossings from 0	$H^{[<h]}_{0,h-1}$	$\dfrac{A_{h-1}}{Q_h}$	(2.15)
7.	Downcrossings to 0	$H^{[<h]}_{h-1,0}$	$\dfrac{B_{h-1}}{Q_h}$	(2.16)
8.	Transitions ($k \leq l$)	$H_{k,l}$	$\dfrac{1}{A_kB_l}Q_k\left(Q_lH_{0,0}-P_l\right)$	(2.20)
9.	Transitions ($k \geq l$)	$H_{k,l}$	$\dfrac{1}{A_kB_l}Q_l\left(Q_kH_{0,0}-P_k\right)$	(2.20)
10.	Upward excursions	$H^{[\geq m,<h]}_{m,m}$	$\dfrac{1}{a_{m-1}b_m}\dfrac{D_{m,h}}{D_{m-1,h}}$	(2.13)
11.	Downward excursions	$H^{[<l+1]}_{l,l}$	$\dfrac{Q_l}{Q_{l+1}}$	(2.27)
12.	Transitions in strip ($k \leq l$)	$H^{[\geq m,<h]}_{k,l}$	$\dfrac{1}{A_kB_l}\dfrac{D_{m-1,k}D_{l,h}}{D_{m-1,h}}$	(2.22)
13.	Transitions in strip ($l \leq k$)	$H^{[\geq m,<h]}_{k,l}$	$\dfrac{1}{A_kB_l}\dfrac{D_{m-1,l}D_{k,h}}{D_{m-1,h}}$	(2.22)

Table 1. Generating functions associated to some major path conditions. The basic continued fraction is $H_{0,0}$ in Entry 1, with convergent polynomials P_h, Q_h. Abbreviations used are: $A_m = a_0\cdots a_{m-1}$, $B_m = b_1\cdots b_m$, and $D_{i,j} = Q_iP_j - P_iQ_j$.

The set of downcrossings from state $h-1$ to 0 that stay below level h is given by

$$H^{[<h]}_{h-1,0} = \frac{B_{h-1}}{Q_h}. \tag{2.16}$$

Proof. Consider first the case of upcrossings from state 0 to state l in (2.15). The expression for ceiled excursions, when plugged inside (2.2), yields a product that telescopes, so that

$$H_{0,l} = \frac{1}{B_l}\left(Q_lH_{0,0}-P_l\right), \tag{2.17}$$

On the other hand, equations (2.10) and (2.12) together with multiplicativity of the determinant imply the classical "determinant identity",

$$P_hQ_{h-1} - Q_hP_{h-1} = (a_0a_1\cdots a_{h-2})(b_1b_2\cdots b_{h-1}). \tag{2.18}$$

The representation of $H^{[<h]}_{0,h-1}$ is obtained next by replacing $H_{0,0}$ with $H^{[<h]}_{0,0} = P_h/Q_h$ in (2.17), then setting $l = h-1$; the expression simplifies because of the determinant identity (2.18), yielding (2.15).

For downcrossings, the time-reversal transformation gives directly

$$H_{k,0} = \frac{B_k}{A_k}H_{0,k} = \frac{1}{A_k}(Q_kH_{0,0}-P_k). \tag{2.19}$$

and the proof concludes like in the case of upcrossings. ∎

Corollary 3 (Transitions [12]). *The set $H_{k,l}$ is representable by means of the fundamental set $H_{0,0}$ and the polynomials P_h and Q_h as:*

$$H_{k,l} = \begin{cases} \dfrac{1}{A_k B_l} Q_k \left(Q_l H_{0,0} - P_l\right) & (k \le l), \\ \dfrac{1}{A_k B_l} Q_l \left(Q_k H_{0,0} - P_k\right) & (k \ge l). \end{cases} \tag{2.20}$$

Proof. The first-passage decomposition (2.3) combined with (2.15) and (2.17) yields the assertion. □

We can now state:

Theorem 2 (Path transitions [17]). *Assume that $0 \le m \le k, l \le h$ and define the determinant-like quantities*

$$D_{i,j} = Q_i P_j - P_i Q_j. \tag{2.21}$$

Then, the transition sets $H_{k,l}^{[\ge m,<h]}$ are given by

$$H_{k,l}^{[\ge m,<h]} = \begin{cases} \dfrac{1}{A_k B_l} \dfrac{D_{m-1,k} D_{l,h}}{D_{m-1,h}} & (k \le l), \\ \dfrac{1}{A_k B_l} \dfrac{D_{m-1,l} D_{k,h}}{D_{m-1,h}} & (k \ge l). \end{cases} \tag{2.22}$$

Proof. It suffices to treat the representative case when $k \le l$. The starting point is relation (2.20) that relativizes to bounded height ($< h$) upon replacing $H_{0,0}$ with P_h/Q_h:

$$H_{k,l}^{[<h]} = \frac{1}{A_k B_l} \frac{Q_k}{Q_h} \left(Q_l P_h - P_l Q_h\right). \tag{2.23}$$

Let ξ be the operation that shifts the indices of variables: $\xi(x_j) = x_{j+1}$, for $x \in \{a, b, c\}$. Shifting indices of variables by m is equivalent to moving the reflecting barrier from altitude 0 to altitude m. Then, the transitions in a strip are

$$H_{k,l}^{[\ge m,<h]} = \xi^m \left(H_{k-m,l-m}^{[<h-m]}\right), \tag{2.24}$$

where the right hand side is determined by (2.23) and shifting. The problem is thus reduced to expressing the shifts of the basic family of polynomials P_h, Q_h.

The polynomials $P_h^{(m)} := \xi^m(P_h)$ and $Q_h^{(m)} := \xi^m(Q_h)$ are usually called the *associated polynomials*[7]. The connection,

$$Q_h^{(m)} = \frac{1}{A_{m-1} B_{m-1}} \left(Q_{m-1} P_{m+h} - P_{m-1} Q_{m+h}\right), \tag{2.25}$$

$$P_h^{(m)} = \frac{1}{A_m B_m} \left(Q_m P_{m+h} - P_m Q_{m+h}\right), \tag{2.26}$$

is readily established. Indeed, the associated polynomials satisfy a recurrence of second order that is, up to a shift of indices, equivalent to the recurrence (2.4) of the basic polynomials P_h, Q_h which form a basis of the linear space of all solutions. (Naturally, one has $\xi(Q_h) = P_{h+1}$.) The connection formulæ (2.25) applied to (2.23) and (2.24) then yield the statement in the case when $k \le l$. The other case $k \ge l$ relies on the dual decomposition, $H_{k,0} = H_{k,l} b_l H_{l-1,0}^{[<l]}$. □

Theorem 2 is a turnkey formula for a number of results. It gives back the earlier results provided one sets wherever appropriate: $P_\infty = H_{0,0}$, $Q_\infty = 1$. An instance of a new result is the special case $k = l$, $h = l + 1$, $m = 0$, which gives a reflex of Corollary 1.

Corollary 4 (Downward excursions). *The set $H_{l,l}^{[<l+1]}$ can be represented by*

$$H_{l,l}^{[<l+1]} = \frac{Q_l}{Q_{l+1}}. \tag{2.27}$$

The different results of this section are summarized in Table 1.

Objects	*Weights* $(\alpha_j, \beta_j \gamma_j)$	*Moments*	*Orth. pol.*
Simple paths	$1, 1, 0$	Catalan #	Chebyshev
Permutations	$j+1, j, 2j+1$	Factorial #	Laguerre
Alternating perm.	$j+1, j, 0$	Secant #	Meixner
Set partition	$j+1, 1, j+1$	Bell #	Poisson-Charlier
Nonoverlap. set part.	$j+1, 1, 1$	Bessel #	Lommel

TABLE 2. Some special families of combinatorial objects together with corresponding weights, moments, and orthogonal polynomials.

2.4. **Relations to enumerative combinatorics.** We digress momentarily from our primary motivation and briefly comment on consequences of Theorems 1, 2 as regards the enumeration of classical combinatorial structures. The theory of lattice path enumerations and continued fractions was intially motivated by the need to count "path diagrams" (equivalently, weighted lattice paths) [10], notably in the context of the analysis of dynamic data structures in computer science [12, 13]. In this framework, a system of multiplicative weights $\alpha_j, \beta_j, \gamma_j$ is associated with the steps a_j, b_j, c_j. In the simplest version, each weight is an integer that represents a number of "possibilities" for the corresponding step type. A system of weighted lattice paths has counting generating functions given by an easy specialization of the corresponding multivariate expressions we have just developed. The "combinatorial morphism" is then

$$a_j \mapsto \alpha_j z, \quad b_j \mapsto \beta_j z, \quad c_j \mapsto \gamma_j z, \tag{2.28}$$

where z marks the length of paths. Under the substitution (2.28), the coefficient $[z^n]H_{0,0}$ is thus the number of weighted paths, and the continued fractions of Table 1 all admit a direct interpretation.

Upon application of the combinatorial morphism (2.28), multivariate generating functions become functions of a single variable z. Under these circumstances, a further relation with orthogonal polynomials develops and we briefly summarize some basic aspects for completeness. Define the linear functional $\mathbf{M}$ on the space of polynomials $\mathbb{C}[x]$ by

$$\mathbf{M}[x^n] = H_{0,0,n} \qquad \text{where} \quad H_{0,0,n} = [z^n]H_{0,0}.$$

A simple rewriting of the coefficient convolution shows that for a polynomial $f(z)$, there holds

$$[z^n]\left(H_{0,0}f(z)\right) = \mathbf{M}[x^n \overline{f}(x)] \qquad \text{where } \overline{f}(x) = x^{\deg(f)}f(x^{-1}),$$

is the reciprocal of f. As an application, one finds immediately rephrasings of Entries 8 and 9 of Table 1 in the form

$$[z^n]H_{k,l} = \frac{1}{(\alpha_0 \cdots \alpha_{k-1})(\beta_1 \cdots \beta_l)}\mathbf{M}\left[x^n \overline{Q}_k(x)\overline{Q}_l(x)\right]. \tag{2.29}$$

This implies in particular that $\mathbf{M}\left[z^n \overline{Q}_k(z)\right] = 0$ for $n < k$ (consider $[z^n]H_{0,k}$ that must be 0 if $n < k$), which means that the reciprocal polynomials $\overline{Q}_k(z)$ constitute an orthogonal system with respect to $\mathbf{M}$. (See for instance [36, Ch. XI] for aspects of orthogonality, [15] for general algebraic combinatorics aspects, and [10, 12] for implications in lattice path enumerations.) The relation (2.29) constitutes a combinatorial analogues of Karlin and McGregor's formula in Theorem 5 below, and the analogy between formulæ originally developed independently in quite different contexts *is* striking.

Bijective combinatorics, continued fractions, and orthogonality relations have been exploited in [10, 12]. As a result, it was found that many classical combinatorial structures are related to classical families of continued fractions and orthogonal polynomials, a fact that has numerous enumerative consequences; see also [16] for an account. A concise summary is offered in Table 2. This table should be compared with Table 4 relative to special birth-and-death processes, and the analogies are striking. (However in the discrete case of paths, the area parameter leads to so-called q-analogues [11] that are of a form different from the basic polynomials.)

3. BIRTH-AND-DEATH PROCESSES AND CONTINUED FRACTION REPRESENTATIONS

In this and the next section, we exploit the combinatorial relations just obtained. The computation of Laplace transforms reduces to a purely mechanically application of the probabilistic morphism χ or its variants according to the principles set forth in Propositions 1, 2, 3.

3.1. **Basic probabilities.** Consider first the basic probability

$$\mathbb{P}_{0,0}(t) := \Pr\{\Lambda_t = 0 \ / \ \Lambda_0 = 0\}$$

that the process issued from 0 is in state 0 at time t. By Proposition 1, the Laplace transform $\widetilde{\mathbb{P}}_{0,0}(s)$ of the function $\mathbb{P}_{0,0}(t)$ equals the quantity $\chi_s(H_{0,0})/(s+\lambda_0)$. Introduce the shorthand notation,

$$\nu_j := \lambda_j + \mu_j,$$

with, as before, $\mu_0 = 0$. Then, one has

$$\widetilde{\mathbb{P}}_{0,0}(s) = \chi_s(H_{0,0})\frac{1}{s+\nu_0} \quad = \quad \cfrac{1}{1 - \cfrac{\frac{\lambda_0\mu_1}{(s+\nu_0)(s+\nu_1)}}{1 - \cfrac{\frac{\lambda_1\mu_2}{(s+\nu_1)(s+\nu_2)}}{\ddots}}} \cdot \frac{1}{(s+\nu_0)}$$

$$(3.1) \qquad = \cfrac{1}{s+\nu_0 - \cfrac{\lambda_0\mu_1}{s+\nu_1 - \cfrac{\lambda_1\mu_2}{\ddots}}},$$

as results from a simple equivalence-preserving transformation. The fraction (3.1) is of a type known as *Jacobi type*, a J-fraction for short.

We consider next the collection of probabilities,

$$\mathbb{P}_{k,l}^{[\geq m, < h]}(t) := \Pr\{\Lambda_t = l \ / \ \Lambda_0 = k, \ \sup_{u\leq t}\{\Lambda_u\} < h, \ \inf_{u\leq t}\{\Lambda_u\} \geq m\},$$

and the specializations

$$\mathbb{P}_{k,l}^{[<h]} = \mathbb{P}_{k,l}^{[\geq 0, <h]}, \quad \mathbb{P}_{k,l}^{[\geq m]} = \mathbb{P}_{k,l}^{[\geq m, <\infty]}, \quad \mathbb{P}_{k,l} = \mathbb{P}_{k,l}^{[\geq 0, <\infty]}.$$

Once more, the convergents of the continued fraction (3.1) play an important role in expressing various sorts of boundary conditions. It is customary to scale the convergent polynomials [28, 29] and express formulæ in terms of polynomials $\mathcal{P}_h(s), \mathcal{Q}_h(s)$, called "birth-and-death" polynomials, that are defined by the recurrence,

$$(3.2) \qquad \lambda_h\mathcal{Y}_{h+1}(s) + (s - \lambda_h - \mu_h)\mathcal{Y}_h(s) + \mu_h\mathcal{Y}_{h-1}(s) = 0,$$

together with the initial conditions $\mathcal{P}_0(s) = 0$, $\mathcal{P}_1(s) = 1/\lambda_0$ and $\mathcal{Q}_{-1}(s) = 0$, $\mathcal{Q}_0(s) = 1$. (The convergent polynomials of the continued fraction $\widetilde{\mathbb{P}}_{0,0}(s)$ are $\lambda_0\cdots\lambda_{h-1}\mathcal{P}_h(-s)$ and $\lambda_0\cdots\lambda_{h-1}\mathcal{Q}_h(-s)$.) We can now state:

Theorem 3 (Birth-and-death continued fractions). (*i*) *The Laplace transform* $\widetilde{\mathbb{P}}_{0,0}(s)$ *is formally represented as*

$$(3.3) \qquad \widetilde{\mathbb{P}}_{0,0}(s) = \cfrac{1}{s+\lambda_0 - \cfrac{\lambda_0\mu_1}{s+\lambda_1+\mu_1 - \cfrac{\lambda_1\mu_2}{s+\lambda_2+\mu_2 - \cfrac{\lambda_2\mu_3}{\ddots}}}}.$$

(ii) The Laplace transform $\widetilde{\mathbb{P}}^{[<h]}_{0,0}(s)$ *of ceiled excursions is given by*

$$\widetilde{\mathbb{P}}^{[<h]}_{0,0}(s) = \cfrac{1}{s+\lambda_0 - \cfrac{\lambda_0\mu_1}{s+\lambda_1+\mu_1 - \cfrac{\lambda_1\mu_2}{\cfrac{\ddots}{s+\lambda_{h-1}+\mu_{h-1}}}}} = \frac{\mathcal{P}_h(-s)}{\mathcal{Q}_h(-s)}. \tag{3.4}$$

(iii) The Laplace transform $\widetilde{\mathbb{P}}^{[\geq h]}_{h,h}(s)$ *of floored excursions is given by*

$$\widetilde{\mathbb{P}}^{[\geq h]}_{h,h}(s) = \cfrac{1}{s+\lambda_h - \cfrac{\lambda_h\mu_{h+1}}{s+\lambda_{h+1}+\mu_{h+1} - \cfrac{\lambda_{h+1}\mu_{h+2}}{\ddots}}} \tag{3.5}$$

$$= \frac{1}{\mu_h}\frac{\mathcal{Q}_h(-s)\widetilde{\mathbb{P}}_{0,0}(s) - \mathcal{P}_h(-s)}{\mathcal{Q}_{h-1}(-s)\widetilde{\mathbb{P}}_{0,0}(s) - \mathcal{P}_{h-1}(-s)}. \tag{3.6}$$

The representation (3.3) for $\widetilde{\mathbb{P}}_{0,0}(s)$ has been used for instance in [6] for studying the behaviour of asymptotically proportional birth-and-death processes.

Proof. Given Theorem 1 and the computation of (3.1), the proof of these assertions only necessitates the determination of the morphic images

$$\chi_s(P_h) = \frac{\lambda_0\ldots\lambda_{h-1}}{\prod_{l=1}^{h-1}(s+\lambda_l+\mu_l)}\mathcal{P}_h(-s), \tag{3.7}$$

$$\chi_s(Q_h) = \frac{\lambda_0\ldots\lambda_{h-1}}{\prod_{l=0}^{h-1}(s+\lambda_l+\mu_l)}\mathcal{Q}_h(-s), \tag{3.8}$$

obtained by comparing the recurrences (2.4) and (3.2). □

The same mechanism immediately gives the counterpart of Theorem 2.

Theorem 4 (Birth-and-death transitions). *Assume that* $1 \leq m \leq k, l \leq h$, *and let* π_n *be given by* (1.2). *Define the determinant-like quantities*

$$\mathcal{D}_{i,j}(s) = \mathcal{Q}_i(s)\mathcal{P}_j(s) - \mathcal{P}_i(s)\mathcal{Q}_j(s). \tag{3.9}$$

Then, the transition probabilities $\widetilde{\mathbb{P}}^{[\geq m,<h]}_{k,l}$ *have Laplace transform given by*

$$\widetilde{\mathbb{P}}^{[\geq m,<h]}_{k,l}(s) = \begin{cases} \pi_l \dfrac{\mathcal{D}_{m-1,k}(-s)\mathcal{D}_{l,h}(-s)}{\mathcal{D}_{m-1,h}(-s)} & (k \leq l) \\ \pi_l \dfrac{\mathcal{D}_{m-1,l}(-s)\mathcal{D}_{k,h}(-s)}{\mathcal{D}_{m-1,h}(-s)} & (l \leq k). \end{cases} \tag{3.10}$$

The dictionary provided by equations (3.1), (3.7), (3.8) yields automatically the other results summarized in Table 3.

We observe, but we don't make use of the fact, that the continued fraction representing $\widetilde{\mathbb{P}}_{0,0}(s)$ can be alternatively obtained as a translation of the combinatorial fraction expression $H_{0,0}$, making use of level steps:

$$s\widetilde{\mathbb{P}}_{0,0}(s) = \hat{\chi}_s(H_{0,0}) \quad \text{where} \quad \hat{\chi}_s(a_j) = -\frac{\lambda_j}{s}, \hat{\chi}_s(b_j) = -\frac{\mu_j}{s}, \hat{\chi}_s(a_j) = -\frac{\nu_j}{s}. \tag{3.11}$$

	Type	*Spec.*	*Laplace transform*
1.	Excursion	$\mathbb{P}_{0,0}$	$\widetilde{\mathbb{P}}_{0,0}(s) = \cfrac{1}{s+\lambda_0} - \cfrac{\lambda_0\mu_1}{s+\lambda_1+\mu_1} - \cdots$
2.	Ceiled excursion	$\mathbb{P}_{0,0}^{[<h]}$	$\dfrac{\mathcal{P}_h(-s)}{\mathcal{Q}_h(-s)}$
3.	Floored excursion	$\mathbb{P}_{h,h}^{[\geq h]}$	$\dfrac{1}{\mu_h}\dfrac{\mathcal{Q}_h(-s)\widetilde{\mathbb{P}}_{0,0}(s)-\mathcal{P}_h(-s)}{\mathcal{Q}_{h-1}(-s)\widetilde{\mathbb{P}}_{0,0}(s)-\mathcal{P}_{h-1}(-s)}$
4.	Transitions from 0	$\mathbb{P}_{0,l}$	$\pi_l\left(\mathcal{Q}_l(-s)\widetilde{\mathbb{P}}_{0,0}(s)-\mathcal{P}_l(-s)\right)$
5.	Transitions to 0	$\mathbb{P}_{k,0}$	$(\mathcal{Q}_k(-s)\widetilde{\mathbb{P}}_{0,0}(s)-\mathcal{P}_k(-s))$
6.	Upcrossings from 0	$\mathbb{P}_{0,h-1}^{[<h]}$	$\dfrac{1}{\lambda_{h-1}}\dfrac{1}{\mathcal{Q}_h(-s)}$
7.	Downcrossings to 0	$\mathbb{P}_{h-1,0}^{[<h]}$	$\dfrac{1}{\mu_h\pi_h}\dfrac{1}{\mathcal{Q}_h(-s)}$
8.	Transitions ($k \leq l$)	$\mathbb{P}_{k,l}$	$\pi_l\mathcal{Q}_k(-s)\left(\mathcal{Q}_l(-s)\widetilde{\mathbb{P}}_{0,0}(s)-\mathcal{P}_l(-s)\right)$
9.	Transitions ($k \geq l$)	$\mathbb{P}_{k,l}$	$\pi_l\mathcal{Q}_l(-s)\left(\mathcal{Q}_k(-s)\widetilde{\mathbb{P}}_{0,0}(s)-\mathcal{P}_k(-s)\right)$
10.	Upward excursions	$\mathbb{P}_{m,m}^{[\geq m,<h]}$	$\dfrac{1}{\mu_m}\dfrac{\mathcal{D}_{m,h}(-s)}{\mathcal{D}_{m-1,h}(-s)}$
11.	Downward excursions	$\mathbb{P}_{l,l}^{[<l+1]}$	$\dfrac{1}{\lambda_l}\dfrac{\mathcal{Q}_l(-s)}{\mathcal{Q}_{l+1}(-s)}$
12.	Transitions in strip ($k \leq l$)	$\mathbb{P}_{k,l}^{[\geq m,<h]}$	$\pi_l\dfrac{\mathcal{D}_{m-1,k}(-s)\mathcal{D}_{l,h}(-s)}{\mathcal{D}_{m-1,h}(-s)}$
13.	Transitions in strip ($l \leq k$)	$\mathbb{P}_{k,l}^{[\geq m,<h]}$	$\pi_l\dfrac{\mathcal{D}_{m-1,l}(-s)\mathcal{D}_{k,h}(-s)}{\mathcal{D}_{m-1,h}(-s)}$

TABLE 3. Laplace transforms of probabilities associated to some major geometric conditions of a birth-and-death process. The basic continued fraction is $\widetilde{\mathbb{P}}_{0,0}(s)$ in Entry 1, with scaled convergent polynomials $\mathcal{P}(-s), \mathcal{Q}(-s)$. The abbreviations used are: $\nu_j = \lambda_j + \mu_j$, $\pi_n = \lambda_0 \cdots \lambda_{n-1}/\mu_1 \cdots \mu_n$, and $\mathcal{D}_{i,j}(s) = \mathcal{Q}_i(s)\mathcal{P}_j(s) - \mathcal{Q}_j(s)\mathcal{P}_i(s)$.

3.2. **Stopping times.** Given the closeness between results provided by Propositions 1 and 2, the very same procedure can be used to derive Laplace transforms of a number of transient characteristics, including a variety of stopping times defined by natural geometric conditions.

As expressed by Proposition 2, $\chi_s(H[Q])$ is the Laplace transform of the stopping time θ associated with condition Q and set of paths $H[Q]$. Under the ergodicity condition (C_1) of the Introduction, certain stopping times are deficient, meaning that their total probability mass is strictly less than 1, while others, like the probability of returning to the same state, are not. The deficiency of a stopping time θ associated to a condition Q is measured by $\chi_0(H[Q])$ (the result of substituting $s = 0$ in χ_s), which represents the probability that $\theta < \infty$, or equivalently, that Q is eventually satisfied. A number of Laplace transforms and of corresponding deficiency probabilities then result from a mechanical translation of Table 3.

We first establish a few values of characteristic quantities at $s = 0$. Ergodicity implies that the process spends a nonzero fraction of its time at state 0; therefore, $\widetilde{\mathbb{P}}_{0,0}(0) = +\infty$. Define next the *scaling function*,

$$\sigma(0) = 0, \quad \sigma(m) := \sum_{n=0}^{n-1} \frac{1}{\lambda_n \pi_n}, \tag{3.12}$$

where the π_n have been defined in (1.2) and are related to the stationary probabilities in (1.3). It is easily checked by the defining recurrence that, for $h \geq 0$,

$$\mathcal{P}_h(0) = \sigma(h), \qquad \mathcal{Q}_h(0) = 1,$$

All conditions expressible in terms of the elementary conditions listed in Tables 1 and 3 must therefore have probabilities expressible in terms of the quantities λ, μ, usually by way of π and σ.

We develop below a representative sample of applications.

Excursion periods: Consider, for a given altitude $h \geq 1$, the excursions of the birth and death process $\{\Lambda_t\}$ above the level $h-1$. Let θ_h denotes the time of such an excursion period,

$$\theta_h = \inf\{t > 0 : \Lambda_t = h-1 \mid \Lambda_0 = h\}.$$

(The particular case θ_1 is known as the duration of the busy period in the context of queueing theory.) The corresponding condition is specified by $(H_{h,h}^{[\geq h]} b_h)$, which by (3.6) and Entry 3 of Table 3 leads to

$$\widetilde{\theta_h}(s) = \chi_s(H_{h,h}^{[\geq h]} b_h) = \frac{\mathcal{Q}_h(-s)\widetilde{\mathbb{P}}_{0,0}(s) - \mathcal{P}_h(-s)}{\mathcal{Q}_{h-1}(-s)\widetilde{\mathbb{P}}_{0,0} - \mathcal{P}_{h-1}(-s)}. \tag{3.13}$$

In this case, one has $\chi_0(H_{h,h}^{[\geq h]} b_h) = 1$, which is consistent with ergodicity. The result (3.13) was established by different methods in [21] and is of special interest for dealing with Laplace transform inversion issues [18, 19].

Hitting times: Consider the Laplace transforms of first hitting times of a state l starting from state k and knowing that the process remains bounded from below or from above. We thus introduce

$$\begin{aligned} \tau_{k,l}^{[\geq m]} &= \inf\{t > 0 : \Lambda_t = l \mid \Lambda_0 = k;\ \inf_{s\in[0,t]} \Lambda_s \geq m\} \\ \tau_{k,l}^{[<h]} &= \inf\{t > 0 : \Lambda_t = l \mid \Lambda_0 = k;\ \sup_{s\in[0,t]} \Lambda_s < h, \}. \end{aligned}$$

The defining conditions are respectively

$$\tau_{k,l}^{[\geq m]} \ :\ \left(H_{k,l-1}^{[\geq m,<l]} a_{l-1}\right), \qquad \tau_{k,l}^{[<h]} \ :\ \left(H_{k,l+1}^{[\geq l+1,<h]} b_{l+1}\right).$$

There result the Laplace transforms,

$$\widetilde{\tau}_{k,l}^{[\geq m]}(s) = \frac{\mathcal{D}_{m-1,k}(-s)}{\mathcal{D}_{m-1,l}(-s)}, \qquad \widetilde{\tau}_{k,l}^{[<h]}(s) = \frac{\mathcal{D}_{k,h}(-s)}{\mathcal{D}_{l,h}(-s)}, \tag{3.14}$$

and the corresponding probabilities of the events $\{\tau_{k,l}^{[\geq m]} < \infty\}$, $\{\tau_{k,l}^{[<h]} < \infty\}$

$$\widetilde{\tau}_{k,l}^{[\geq m]}(0) = \frac{[\sigma(k) - \sigma(m-1)]}{[\sigma(l) - \sigma(m-1)]}, \qquad \widetilde{\tau}_{k,l}^{[<h]}(0) = \frac{[\sigma(h) - \sigma(k)]}{[\sigma(h) - \sigma(l)]}. \tag{3.15}$$

In the particular case when $m = 0$ and when $k \leq l$, the formula (3.14) simplifies to

$$\widetilde{\tau}_{k,l}^{[\geq 0]}(s) = \frac{\mathcal{Q}_k(-s)}{\mathcal{Q}_l(-s)}, \tag{3.16}$$

and the random variable is no longer defective as the condition $\Lambda_t \geq 0$ becomes a vacuous one.

Exit times: Let $1 \leq m < k < h$ and denote by $\varepsilon_k^{[\geq m,<h]}$ the exit time from the strip $[m+1, h-1]$, starting from state k. We find similarly

$$\widetilde{\varepsilon}_k^{[\geq m,<h]}(s) = \frac{\mathcal{D}_{m,k}(-s) - \mathcal{D}_{h,k}(-s)}{\mathcal{D}_{m,h}(-s)}. \tag{3.17}$$

These examples furnish, amongst other things, various probabilistic interpretations of σ. For instance, define $T_x = \inf\{t > 0 : \Lambda_t = x\}$; from previous equations, σ is seen to satisfy the relation

$$\Pr\{T_h < T_m \mid \Lambda_0 = k\} = \frac{\sigma(k) - \sigma(m)}{\sigma(h) - \sigma(m)}.$$

3.3. **Area and level crossings.** Proposition 3 opens the possibility of recording the area $\mathcal{A}_c$ above some level c. In queueing theory terms, the case $c = 0$ measures the cost in total waiting time of all customers till a condition is met; the case of general $c > 0$ measures the cost incurred during heavy traffic periods. The morphism χ_s° of (1.15) then applies. For instance, the joint distribution of $(\theta, \mathcal{A}_h)$ above level $h-1$ is obtained as $\chi_s^\circ(H_{h,h}^{[\geq h]} b_h)$ and one gets:

Corollary 5. *The joint Laplace transform of* $(\theta_h, \mathcal{A}_h)$ *is*

$$(3.18)\quad \mathcal{G}_h(s,y) \;:=\; \mathbb{E}\left[e^{-s\theta_h - y\mathcal{A}_h}\right] \;=\; \cfrac{\mu_h}{s+y+\lambda_h+\mu_h - \cfrac{\lambda_h\mu_{h+1}}{s+2y+\lambda_{h+1}+\mu_{h+1} - \cfrac{\lambda_{h+1}\mu_{h+2}}{\ddots}}},$$

This result generalizes the formula obtained by Preater [35] in the specific case when the BDP under consideration is the occupation process of an $M/M/\infty$ queue. Such continued fraction representations are variants of the continued fraction expressing $\widetilde{\mathbb{P}}_{h,h}^{[\geq h]}$ and they derive by a substitution in the denominator of the original continued fractions,

$$\lambda_h + \mu_h \;\mapsto\; \lambda_h + \mu_h + (h+1-c)^+ y.$$

This gives rise to a new collection of modified polynomials $\mathcal{P}_h^\circ(s;y)$, $\mathcal{Q}_h^\circ(s;y)$ of the "area type" that are obtained by

$$(3.19)\qquad \lambda_h \mathcal{Y}_{h+1}^\circ(s) + (s - \lambda_h - \mu_h - (h+1-c)^+ y)\mathcal{Y}_h^\circ(s) + \mu_h \mathcal{Y}_{h-1}^\circ(s) = 0,$$

and that can be identified in some important cases; see [19, 20] and Section 5. Formulæ entirely parallel to those derived for the standard probabilistic morphism (Theorems 3, 4 and Table 3) can then be easily developed.

The morphism χ_s is susceptible to enrichment in various other ways. For instance, it is possible to keep track simultaneously of random variables recording the total number of births (N_{a_j}) or deaths (N_{b_j}) at level j. Let u_j and v_j mark respectively N_{a_j} and $N_{b,j}$. The version of the probabilistic morphism to be applied is then

$$\mathbb{E}\left[e^{-s\theta_Q} u_0^{N_{a_0}} u_1^{N_{a_1}} v_1^{N_{b_1}} \cdots\right] = \chi_s'(H[Q]),$$

where

$$\chi_s'(a_j) = \frac{\lambda_j u_j}{s+\lambda_j+\mu_j},\; \chi_s'(b_j) = \frac{\mu_j v_j}{s+\lambda_j+\mu_j}.$$

(This rule can be otherwise freely combined with an area enrichment.) Various specializations are then of interest: for instance, the number of times a transition b_k occurs in the course of an excursion is obtained by setting $u_j = v_j = 1$ if $j \neq k$ together with $u_k = 1$, $v_k = v$; ascents are marked by $u_j = u$ and $v_j = 1$, etc. Simple algebra in the style of (2.12) then yields the distribution of the corresponding number of transitions.

3.4. **Discrete-time analogues.** Until now, we have refrained from discussing the case of inhomogeneous random walks on the integer half-line—the discrete time analogue of birth-and-death processes. A continued fraction theory exists but it is in a way more "shallow". The expressions obtained are generating functions ("z-transforms") in a single variable z that records the discrete time parameter.

Let p_j, q_j, r_j be the probabilities of transitions $\langle j, j-1\rangle$, $\langle j, j+1\rangle$, and $\langle j, j\rangle$ respectively. The "walk generating function" $\mathbb{W}_{0,0}$ is defined by the fact that its coefficient $[z^n]\mathbb{W}_{0,0}(z)$ equals the probability of returning to 0 from state 0 in n steps. (See Godsil's book [15] for a nice introduction to walk generating functions.) The connection with the lattice path theory of Section 2 is expressed by the "random walk morphism",

$$(3.20)\qquad \eta_z(a_j) = p_j z, \quad \eta_z(b_j) = q_j z, \quad \eta_z(c_j) = r_j z,$$

which is of the same form as the combinatorial morphism (2.28), safe that the weights are now arbitrary positive real numbers satisfying $p_j + q_j + r_j = 1$. Consequently, one has

$$\mathbb{W}_{0,0}(z) = \cfrac{1}{1 - r_0 z - \cfrac{p_0 q_1 z^2}{1 - r_1 z - \cfrac{p_1 q_2 z^2}{\ddots}}}. \tag{3.21}$$

The developments of Section 2 apply *verbatim* once the trivial change of notations (3.20) has been performed. Basic aspects of this connection between continued fractions and inhomogeneous walks on the integer half-line seem to have been first recognized by I. J. Good in 1958; see [22]. They have been rediscovered and published many times since then.

4. Analytic properties and orthogonality

We again digress a bit from our main thread in this section: we illustrate briefly the way the continued fraction approach may be used as entry point for deriving many orthogonal representations of the type first discovered by Karlin and McGregor. Chapter 12 of Henrici's [23] book contains a lucid exposition of the convergence theory of algebraic continued fractions while the paper by Bordes and Roehner [6] presents an insightful discussion of the role of continued fraction theory in the analysis of birth-and-death processes.

The continued fraction expressing $\widetilde{\mathbb{P}}_{0,0}(s)$ in (3.3) is, up to notational details a continued fraction of the Jacobi type, known as a J-fraction. As it has been recognized for a long time, this continued fraction also admits an "expanded" form obtained via simple equivalence preserving transformations, namely

$$\widetilde{\mathbb{P}}_{0,0}(s) = \cfrac{1}{s + \cfrac{\lambda_0}{1 + \cfrac{\mu_1}{s + \cfrac{\lambda_2}{\ddots}}}}. \tag{4.1}$$

The coefficients in (4.1) being positive, this continued fraction is a Stieltjes fraction (S fraction). The classical theory developed by Stieltjes then implies, because of the ergodicity condition (C_1), that the odd and even approximants of the fraction in (4.1) each converge to analytic functions in the complex plane split along $(-\infty, 0)$. As a consequence, the J-fraction expressing $\widetilde{\mathbb{P}}_{0,0}(s)$ in (3.3), being the even part of the S-fraction of (4.1), itself defines an analytic function of s in the split plane.

We next turn to analytic properties of the $\mathcal{P}, \mathcal{Q}$ polynomials that are the central figures in Karlin and McGregor's thorough treatment [28, 30]. Our presentation here is only meant to point to alternative integral representations for some of the quantities of interest. The (linear) *moment functional* $\mathfrak{M}$ is known to play an especially important role. It is first defined formally over the linear space of polynomials $\mathbb{C}[x]$ by

$$\mathfrak{M}[x^n] = M_n, \tag{4.2}$$

where the coefficients M_n, called the *moments*, are given by the asymptotic expansion at infinity

$$\widetilde{\mathbb{P}}_{0,0}(s) \sim \sum_{n\geq 0}^{\infty} (-1)^n \frac{M_n}{s^{n+1}} \qquad (s \to +\infty). \tag{4.3}$$

Furthermore, Stieltjes' theory grants that the moment functional always admits a representation as a Laplace-Stieltjes integral

$$\mathfrak{M}[f] = \int_0^{\infty} f(x)\, d\psi(x), \tag{4.4}$$

for some *measure* $\psi(x)$ over the positive real line that may *a priori* have discrete, continuous, or mixed spectrum. Then, the continued fraction $\widetilde{\mathbb{P}}_{0,0}(s)$ is a Stieltjes transform, that is,

$$\widetilde{\mathbb{P}}_{0,0}(s) = \int_0^\infty \frac{d\psi(x)}{x+s}, \; s \in \mathbb{C} \setminus (-\infty, 0), \tag{4.5}$$

and the moments deserve their name:

$$M_n = \int_0^\infty x^n \, d\psi(x).$$

(Under the ergodicity condition (C_1), the spectral measure $\psi(x)$ is the unique solution of the Stieltjes moment problem and also the unique measure satisfying (4.5).)

A fundamental property is that the polynomials $\{\mathcal{Q}_h(s)\}$ form an *orthogonal polynomial system* (OPS) [23] with respect to the spectral measure ψ,

$$\int_0^\infty \mathcal{Q}_i(x)\mathcal{Q}_j(x) d\psi(x) = \frac{1}{\pi_i}\delta_{i,j}, \tag{4.6}$$

where $\delta_{i,j}$ is the Kronecker symbol.

As we now explain, the formulæ derived directly from lattice path combinatorics can be used to recover alternative integral representations including some of Karlin and McGregor's results in [28, 29]. (Our derivation has perhaps the merit of not being dependent upon general theorems from spectral analysis.) We consider the transition probability $\mathbb{P}_{k,l}(t) = \Pr\{\Lambda_t = l \;/\; \Lambda_0 = k\}$ for which the classical result of Karlin and McGegor provides an expression by means of the OPS $\{\mathcal{Q}_h(s)\}$ and its associated spectral (Stieltjes) measure ψ.

Theorem 5 (Karlin and McGegor [28]). *The transition probability $\mathbb{P}_{k,l}(t)$, for $k, l \geq 0$, is given by*

$$\mathbb{P}_{k,l}(t) = \pi_l \int_0^\infty e^{-tx} \mathcal{Q}_k(x)\mathcal{Q}_l(x) d\psi(x). \tag{4.7}$$

Proof (sketch). We briefly indicate the principles on which the proof is built. Details can be supplied by mimicking closely the derivations of [23, §12.9] (see especially Theorem 12.9h).

It suffices to treat the case when $k \leq l$. Let $[s^{-m}]f(s)$ denote the coefficient of s^{-m} in the expansion of some function $f(s)$ at infinity. Theorem 4 and Entry 8 of Table 3 provide an expression for $\widetilde{\mathbb{P}}_{k,l}(s)$ that implies

$$(-1)^n[s^{-n-1}]\frac{1}{\pi_l}\widetilde{\mathbb{P}}_{k,l}(s) = (-1)^n[s^{-n-1}]\mathcal{Q}_k(-s)\mathcal{Q}_l(-s)\widetilde{\mathbb{P}}_{0,0}(s) \tag{4.8}$$

$$= \int_0^\infty \mathcal{Q}_k(x)\mathcal{Q}_l(x) x^n d\psi(x), \tag{4.9}$$

where (4.8) takes care of the disappearance of terms with positive powers of s, and (4.9) is a rephrasing of convolution formulæ in terms of orthogonality. Now (4.9) means that

$$\widetilde{\mathbb{P}}_{k,l}(s) = \pi_l \int_0^\infty \frac{\mathcal{Q}_l(x)\mathcal{Q}_k(x)}{s+x} d\psi(x).$$

A Stieltjes transform is a double Laplace transform, so that Laplace inversion then gives the statement. ☐

We can finally obtain a counterpart of Theorem 4 phrased in terms analogous to Karlin and McGregor's theorem. Define the spectral measure $\psi_m^h(x)$ that is discrete, finite, and such that a Stieltjes transform representation holds,

$$\widetilde{\mathbb{P}}_{m,m}^{[\geq m, <h-m]}(s) = \int_0^\infty \frac{d\psi_m^h(x)}{s+x}. \tag{4.10}$$

The representation can be easily obtained from a partial fraction decomposition of $\widetilde{\mathbb{P}}_{m,m}^{[\geq m, <h-m]}(s)$ itself expressible in terms of the $\mathcal{P}, \mathcal{Q}$ polynomials. Then, by the same arguments as before, one finds:

Process	*Continued fr.*	*Polynomials*	$\psi(x)$	*Moments*
$M/M/1$	Quadratic	Chebyshev	Quadratic (cont.)	Catalan #
— Area	${}_0F_1$ quotient	Lommel	Bessel (discr.)	Bessel #
$M/M/\infty$	${}_1F_1$-hypergeom.	Poisson-Charlier	Poisson (discr.)	Bell-Stirling #
— Area	${}_1F_1$ quotient	mod. Poisson-Ch.	—	—
Population				
— prop.	${}_1F_1$ quotient	Laguerre	Exp. (cont.)	Factorial #
— non-prop.	${}_2F_1$ quotient	Meixner	(discr.)	—
— Area	—	mod. Meixner	—	—

TABLE 4. Some classical processes and a broad indication of the types of associated special functions.

Theorem 6. *The transition probability function* $\mathbb{P}_{k,l}^{[\geq m,<h]}(t)$ *is given by*

$$\mathbb{P}_{k,l}^{[\geq m,<h]}(t) = \mu_m^2 \pi_m \pi_l \int_0^\infty e^{-tx} \mathcal{D}_{m-1,k}(x) \mathcal{D}_{m-1,l}(x) d\psi_m^h(x). \tag{4.11}$$

Orthogonality relations prove useful for special processes where the measure ψ can be made explicit.

5. SPECIAL PROCESSES

With each system of birth-and-death rates $\{\lambda_j, \mu_j\}$ there are associated a basic continued fraction $\widetilde{\mathbb{P}}_{0,0}(s)$, a family of orthogonal polynomials $\mathcal{Q}_k(s)$, an orthogonality measure $\psi(x)$, and a moment sequence M_n. General identities relate these objects; see [4, 23]. "Special" processes are defined by systems of birth-and-death rates $\{\lambda_j, \mu_j\}$ that obey a regular pattern meant to model a physical phenomenon that has itself some strong form of regularity. Identification of the relevant class of objects may then often be achieved by table look up: see especially Chihara's book [7] for classical orthogonal polynomials or Wall's book [36] for continued fractions. A rich set of special function identities then usually supplements the general identities valid for an arbitrary birth-and-death process. In this section, we simply make explicit the classes of functions involved for three major special processes arising from queueing theory and population growth models. See Table 4 for a summary and refer to Table 2 or [10, 11, 12, 13, 14] for the corresponding combinatorial analogues. The point made here is that continued fractions may adequately serve as an entry to the analysis of special processes.

5.1. **The $M/M/1$ queue.** Consider the single server queue $M/M/1$ with arrival rate $\rho < 1$ and unit service rate [8, 32] whose parameters are

$$\lambda_k = \rho, \quad \mu_k = 1.$$

The fundamental continued fraction is then

$$\widetilde{\mathbb{P}}_{0,0}(s) = \cfrac{1}{\rho + s - \cfrac{\rho}{1 + \rho + s - \cfrac{\rho}{1 + \rho + s - \cfrac{\rho}{\ddots}}}}.$$

The continued fraction has constant coefficients reflecting the uniform structure of the $M/M/1$ process. From the continued fraction representation, we see that $\widetilde{\mathbb{P}}_{0,0}(s) = (s + \rho - X)^{-1}$ where X is a solution to the equation $X = 1/(1 + \rho + s - \rho X)$. (The quantity X is in fact $\widetilde{\theta}_1(s)$, the Laplace transform of the duration of a busy period of the $M/M/1$ queue.) Introducing the two conjugate quantities,

$$V(s) = \frac{1+\rho+s-\sqrt{(1+\rho+s)^2-4\rho}}{2\rho} \tag{5.1}$$

$$\overline{V}(s) = \frac{1+\rho+s+\sqrt{(1+\rho+s)^2-4\rho}}{2\rho}, \tag{5.2}$$

we find

$$\widetilde{\mathbb{P}}_{0,0}(s) = \frac{1-\rho-s+\sqrt{(1+\rho+s)^2-4\rho}}{2s} = \frac{V(s)}{1-V(s)}. \tag{5.3}$$

The $\mathcal{Q}_k$ polynomials satisfy a linear recurrence with constant coefficients, so that

$$\mathcal{Q}_k(-s) = \frac{V(s)^{k+1} - \overline{V}(s)^{k+1}}{V(s) - \overline{V}(s)},$$

and one has $\mathcal{P}_k = \mathcal{Q}_{k-1}$. Up to an affine transformation on the argument, these are Chebyshev polynomials with orthogonality measure

$$d\psi(x) = \frac{1}{2\pi\rho}\sqrt{4\rho-(x-1-\rho)^2}\,\mathbb{1}_{[(1-\sqrt{\rho})^2,(1+\sqrt{\rho})^2]}dx.$$

Moments M_n are related to the Catalan numbers arising in combinatorial theory.

The area polynomials appear to be modified Lommel polynomials[1], an observation that led to the distributional analysis of area during a busy period [20].

5.2. $M/M/\infty$ **queue.** An infinite server queue $M/M/\infty$ with mean arrival rate u and unit service rate is defined by

$$\lambda_k = u, \quad \mu_k = k.$$

The fundamental continued fraction is

$$\widetilde{\mathbb{P}}_{0,0}(s) = \cfrac{1}{u+s-\cfrac{1u}{1+u+s-\cfrac{2u}{2+u+s-\cfrac{3u}{\ddots}}}}.$$

This is recognizable as a confluent case of Gauss's continued fraction expansion for the quotient of two contiguous hypergeometric functions [36]. Thus,

$$\widetilde{\mathbb{P}}_{0,0}(s) = \frac{1}{s}\Phi(1, 1+s; -u), \tag{5.4}$$

where $\Phi(\alpha, \gamma; z)$ is the Kummer function (equivalently, a confluent hypergeometric function ${}_1F_1$) defined by

$$\Phi(\alpha,\gamma;z) = \frac{\Gamma(\gamma)}{\Gamma(\alpha)}\sum_{k=0}^{+\infty}\frac{\Gamma(\alpha+k)}{\Gamma(\gamma+k)}\frac{z^k}{k!},$$

where Γ is the Euler function. (A comprehensive treatment of this special function is given in [33, Chap.9].) The polynomials $\{\mathcal{Q}_n\}$ are identified by their three-term recurrence [7, 9] or by their generating function that satisfies a solvable differential equation. They are the well-known Charlier polynomials,

$$\mathcal{Q}_n(x) = \sum_{k=0}^{n}\binom{x}{k}\frac{(-u)^{n-k}}{(n-k)!}. \tag{5.5}$$

[1] The authors cannot resist inserting a personal note at this point. Our surprise at discovering the concomitant occurrence of Lommel polynomials in the seemingly unrelated contexts of area under the $M/M/1$ process [20] and of the enumeration of non-overlapping set partitions [14] led us to the present work, originally meant to elucidate the phenomenon.

The corresponding spectral measure ψ is the Poisson measure of intensity u on $\mathbb{N}$,

$$\psi = \sum_{k=0}^{\infty} e^{-u} \frac{u^k}{k!} \boldsymbol{\delta}_k.$$

The moments are closely related to Stirling polynomials of the second kind and to Bell numbers that enumerate set partitions [10].

Due to the very special nature of the process, the area polynomials turn out to reduce to the Poisson-Charlier polynomials, up to a shift of the argument; see [19] where the corresponding formulæ are worked out in full detail.

5.3. **Population growth.** Population growth under constant birth-and-death rates but with possible migration may be modelled by the process

$$\lambda_n = a(n+c+\alpha+1),\ \mu_n = n+c,\ n \geq 0,\ \mu_0 = 0.$$

The quantity a measures the tendency of the population to grow or decay and the process is said to be (asymptotically) proportional when $a=1$; the quantities c, α measure rates of immigration and emigration, and the process will be called "chauvinistic" when $c=0$. The fundamental continued fraction is

$$\widetilde{\mathbb{P}}_{0,0}(s) = \cfrac{1}{a(c+\alpha+1)+s-\cfrac{a(c+\alpha+1)(c+1)}{a(c+\alpha+2)+(c+1)+s-\cfrac{a(c+\alpha+2)(c+2)}{\ddots}}}.$$

This continued fraction is a quotient of two contiguous ${}_2F_1$ hypergeometric functions, by virtue of Gauss's classical expansion [36].

In the simplest case of a chauvinistic and proportional process ($c=0, a=1$), the orthogonal polynomials $Q_n(x)$ are the Laguerre polynomials. For a nonproportional but chauvinistic process ($c=0,\ a \neq 1$), they are the Meixner polynomials. In the case of a process with $c>0$, the polynomials become associated Laguerre ($a=1$) or associated Meixner ($a \neq 1$) polynomials. See [25] for a thorough discussion.

In the proportional case $a=1$, the continued fraction $\widetilde{\mathbb{P}}_{0,0}(s)$ is, for $\alpha > -1$,

$$\widetilde{\mathbb{P}}_{0,0}(s) = \frac{\Psi(c+1, 1-\alpha; s)}{\Psi(c, -\alpha; s)}, \tag{5.6}$$

where $\Psi(\alpha, \gamma; z)$ denotes the Tricomi function [33], again a function of the ${}_1F_1$ hypergeometric category. The spectral measure ψ is given by

$$d\psi(x) = \frac{x^{\alpha} e^{-x}}{\Gamma(c+1)\Gamma(c+1+\alpha)} \left|\Psi(c, -\alpha; x e^{-i\pi})\right|^{-2} dx. \tag{5.7}$$

In the nonproportional case $a \neq 1$, one has

$$\widetilde{\mathbb{P}}_{0,0}(s) \quad = \quad \frac{1}{s} \frac{{}_2F_1(c+1, -\alpha-s; 1+c-s; a)}{{}_2F_1(c, -\alpha-s; 1+c-s; a)}. \tag{5.8}$$

where ${}_2F_1(\alpha, \beta; \gamma; z)$ denote the usual hypergeometric function. The spectrum of the measure ψ becomes discrete and it consists of the zeros of the equation:

$$\frac{\Gamma(1-x)}{\Gamma(-x)\Gamma(c+1-x)}\ {}_2F_1(c, -\alpha-x; 1+c-x; a) = 0. \tag{5.9}$$

Loosely speaking, the moments are related to the enumeration of permutations as well as to factorial and Eulerian numbers.

Due to the leeway allowed by the two parameters of the model, the area polynomials remain in a class similar to the one of the basic polynomials.

5.4. **Other models.** The three classes of processes previously introduced each have birth-and-death rates λ_k, μ_k that are at most linear functions in the level k. The corresponding fractions are then closely related to Gauss's continued fraction expansion of the quotient of two contiguous hypergeometric functions [36], with suitable confluences. Accordingly, the orthogonal polynomials relate to the Meixner classification of special orthogonal polynomials [7], an interesting parallel with the combinatorial cases of Section 2.4.

The approach we have exposed clearly applies to finite populations, that is, processes with only a finite number of states, as this is equivalent to imposing a ceiling constraint on an infinite process. It also applies to "bimodal" systems, a typical instance of which is the $M/M/m$ queue: the rates are those of the $M/M/\infty$ queue below a certain threshold level m, after which they become of the $M/M/1$ type. In such cases, two regimes are visible in the continued fraction representation. Given the linear fractional transformation and matrix expressions available (see Eq. (2.12) typically), mixed representations are obtained that involve the two families of polynomials corresponding to the two regimes.

6. Conclusions

We have developed a formal calculus of basic events described by trajectories of birth-and-death processes on which geometric conditions are imposed. The corresponding algebra is that of chains of linear fractional transformations ("*Kettenbrüchen*"), itself expressible by finite and infinite products of 2×2 matrices. Consequently, all basic events have Laplace transforms that are expressible rationally in terms of the fundamental continued fraction of the process and of its associated polynomials. This calculus is of a mechanical nature and it may be used to treat easily fairly complex transient characteristics of processes. A byproduct is formulæ based on integral representations and orthogonality, in the style of Karlin and McGregor's studies. As an additional benefit, the formal approach induces a simple calculus of stopping probabilities and of inhomogeneous random walks on the line, while being susceptible to a number of extensions to parameters like area or level crossings easily taken into account by suitable morphisms.

Acknowledgements. The authors are grateful to Dr. Philippe Robert for a careful scrutiny of the paper. The work of the first named author has been supported in part by the Alcom-IT Project (#20244) of the European Union.

References

[1] J. Abate and W. Whitt. Computing Laplace transforms for numerical inversion via continued fractions. *INFORMS Journal on Computing*, 7(1):36–43, 1995. To appear.
[2] J. Abate and W. Whitt. Laplace transforms of density probability density functions with series representation. AT&T Labs, 1998.
[3] J. Abate and W. Whitt. Numerical inversion of Laplace transforms of probability distributions. *ORSA Journal on Computing*, 7(1):36–43, 1995.
[4] R. Askey and M.E.H. Ismail. *Recurrence relations, continued fractions, and orthogonal polynomials.* Memoirs of the American Mathematical Society, Volume 49, Number 300, 1984.
[5] S. Asmussen. *Applied probability and queues.* Wiley, N.Y., 1987.
[6] G. Bordes and B. Roehner. Applications of Stieltjes theory for S-fractions to birth and death processes. *Adv. Appl. Prob.*, 15:507–530, 1983.
[7] T. S. Chihara. *An Introduction to Orthogonal Polynomials.* Gordon and Breach, New York, 1978.
[8] J. W. Cohen. *The Single Sever Queue.* North Holland, 1982.
[9] A. Erdélyi. *Higher Transcendental Functions.* 3 volumes. McGraw Hill, NY, 1953.
[10] P. Flajolet. Combinatorial aspects of continued fractions. Discrete Mathematics, 32:125-161, 1980.
[11] P. Flajolet. On congruences and continued fractions for some classical combinatorial quantities. *Discrete Mathematics*, 41:145-153, 1982.
[12] P. Flajolet, J. Françon, and J. Vuillemin. Sequence of operations analysis for dynamic data structures. *Journal of Algorithms*, 1:111-141, 1980.
[13] P. Flajolet, C. Puech, and J. Vuillemin. The analysis of simple list structure. Information Sciences, 38:121-146, 1986.
[14] P. Flajolet and R. Schott. Non–overlapping partitions, continued fractions, Bessel functions and a divergent series. *European J. Combinatorics*, 11:421-432, 1990.
[15] C. D. Godsil. *Algebraic Combinatorics*, Chapman and Hall, 1993.
[16] I. P. Goulden and D. M. Jackson. *Combinatorial Enumeration.* John Wiley, 1983.

[17] I. P. Goulden and D. M. Jackson. Path Generating functions and continued fractions. *J. of Combinatorial Theory*, Series A, 41:1–10, 1986.

[18] F. Guillemin and D. Pinchon. Continued fraction analysis of the duration of an excursion in an $M/M/\infty$ system. *J. Appl. Prob.*, 35:165–183, 1998.

[19] F. Guillemin and D. Pinchon. On a random variable associated with excursion in an $M/M/\infty$ system. *Queueing Systems*. 1999.

[20] F. Guillemin and D. Pinchon. On the area swept under the occupation process of an $M/M/1$ queue in a busy period. *Queueing Systems*(20):383-398, 1998.

[21] F. Guillemin and D. Pinchon. Excursions of birth an death processes, orthogonal polynomials, and continued fractions. *J. Appl. Prob.* 1999.

[22] I. J. Good. Random Motion and Analytic Continued Fractions. *Proceedings of the Cambridge Philosophical Society*, 54:43-47, 1958.

[23] P. Henrici. *Applied and Computational Complex Analysis.* Vol. 2 Wiley, New York, 1977.

[24] M.E.H. Ismail, J. Letessier, D. Masson, and G. Valent. Birth and death processes and orthogonal polynomials. In P. Nevai, editor, *Orthogonal polynomials*, pages 229–255. Kluwer Academic Publishers, 1990.

[25] M.E.H. Ismail, J. Letessier, and G. Valent. Linear birth and death models and associated Laguerre and Meixner polynomials. *Journal of Approximation Theory*, 55, pp. 337-348, 1988.

[26] W. B. Jones and A. Magnus. Application of Stieltjes fractions to birth-death processes. In E. B. Saff and R. S. Varga, editors, *Padé and Rational Approximation*, pages173–179, Academic Press , 1977.

[27] W. B. Jones and W. J. Thron. *Continued Fractions: Analytic Theory and Applications.* In *Encyclopedia of Mathematics and its Applications*, vol. 11, Addison–Wesley, 1990.

[28] S. Karlin and J. Mc Gregor. The differential equation of birth and death processes, and the Stieltjes moment problem. *Trans. Amer. Math. Soc.*, 85:489–546, 1957.

[29] S. Karlin and J. Mc Gregor. The classification of birth and death processes. *Trans. Amer. Math. Soc.*, 86:366–401, 1957.

[30] S. Karlin and J. Mc Gregor. Linear growth birth and death processes. *J. Math. Mech.*, 7:643–662, 1958.

[31] S. Karlin and H. Taylor. *A First Course in Stochastic Processes.* Academic Press, 1975.

[32] L. Kleinrock. *Queueing systems, Vol.I: Theory.* J. Wiley, 1975.

[33] N. Lebedev. *Special functions and their applications.* Prentice Hall, 1965.

[34] J. R. Norris. *Markov Chains.* Cambridge University Press, 1997.

[35] J. Preater. $M/M/\infty$ transience revisited. *J. Appl. Prob.*, 34:1061-1067, 1998.

[36] H.S. Wall. *Analytic Theory of Continued Fractions.* Chelsea, New York, 1967.

P. Flajolet, Algorithms Project, INRIA, F-78150 Rocquencourt (France)
E-mail address: `Philippe.Flajolet@inria.fr`

F. Guillemin, France Telecom, BD.CNET, DAC/ATM, 2, Avenue Pierre Marzin, 22300 Lannion (France)
E-mail address: `Fabrice.Guillemin@cnet.francetelecom.fr`

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105,
78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399