Probability in the Engineering and Informational Sciences, **24**, 2010, 129–144.
doi:10.1017/S0269964809990179

ZIPF AND LERCH LIMIT OF BIRTH AND DEATH PROCESSES

B. KLAR, P. R. PARTHASARATHY, AND N. HENZE
Institut für Stochastik
Karlsruhe Institute of Technology
76133 Karlsruhe, Germany
E-mail: bernhard.klar@kit.edu

Birth and death processes are useful in a wide range of disciplines from computer networks and telecommunications to chemical kinetics and epidemiology. Data from many different areas such as linguistics, music, or warfare fit Zipf's law surprisingly well. The Lerch distribution generalizes Zipf's law and is applicable in survival and dispersal processes. In this article we construct a birth and death process that converges to the Lerch distribution in the limit as time becomes large, and we investigate the speed of convergence. This is achieved by employing continued fractions. Numerical illustrations are presented through tables and graphs.

1. INTRODUCTION

Many types of data studied in physical and social sciences can be adequately described by the Zipf distribution. Zipf's law states that if we rank a collection of subjects in nondecreasing order according to their size, the product of a power of the rank and of the size of each object is constant throughout the collection. Consider a sequence of data values, ordered as $x_{(1)} \geq x_{(2)} \geq x_{(3)} \geq \cdots \geq x_{(n)}$. We can think of r as the rank and $x_{(r)}$ as the size of the rth data value in the ordered set. The relation $rx_{(r)} =$ constant seems to hold for various kinds of objects, including cities in the United States by population, books by number of pages, words in an essay by their frequency of occurrence, and the biological genera by number of species. The rank-size relation is known as Zipf's law and its graph is a rectangular hyperbola.

Zipf's law was later generalized to $r^q x_{(r)} =$ constant, $q > 0$, leading to a discrete probability distribution

$$P(X = r) = \zeta(1 + q)\, r^{-(1+q)}, \qquad r = 1, 2, 3, \ldots,$$

 0269-9648/10 $25.00

where

$$\zeta(s) = \sum_{r=1}^{\infty} \frac{1}{r^s}, \qquad s > 1$$

is the zeta function. Data from many different areas such as linguistics, music, and warfare fit such a law surprisingly well [17]. This empirical law is quite relevant to many economic fields: to model the firm's size (measured either by the number of employees or by the receipts account), the distribution of wealth invested by individual investors in financial markets, population sizes of the cities, the proportion of genera with exactly r species, and the number of customers across a wide area network, see [13]. Hill [7] presented a theoretical derivation of the rank-frequency form of Zipf's law based on a Bose–Einstein form of the classical occupancy model, with the additional feature that the number of cells is random.

Generalizing the Zipf distribution, we have the Lerch distribution with

$$P(X = r) = \frac{cz^r}{(\nu + r)^s}, \qquad r = 0, 1, 2, \ldots,$$

where $1/c = \sum_{r=0}^{\infty} z^r/(\nu + r)^s$ and ν is a positive parameter.

Motivation for using the Lerch distribution includes the following considerations. Kulasekera and Tonkyn [8] and Aksenov and Savageau [1] have demonstrated that the Lerch distribution is useful to model survival processes because its hazard function can be constant, monotonically decreasing, or monotonically increasing depending on the value of one parameter, and dispersal processes because its variance can be greater, equal, or less than the mean.

In this article we explain how birth and death processes can be constructed with limiting Lerch distribution.

2. BIRTH AND DEATH PROCESSES

Birth and death processes (BDPs) are widely used in a variety of fields, including queues, inventories, reliability, communication, production management, neutron propagation, optics, chemical reactions, epidemics, population dynamics, and many other domains of application. They are useful in investigating phenomena that are essentially concerned with a flow of events in time, especially those exhibiting such highly variable characteristics as birth, death, transformation, evolution, arrival, and departure. What makes BDPs so useful is that standard methods of analysis are available for determining numerous important quantities such as transient and stationary distributions and first-passage times.

Connections among BDPs, continued fractions (CFs), and orthogonal polynomials (OPs) are well known (Jones and Thron [19], Lorentzen and Waadeland [9], Parthasarathy and Lenin [14]). CF approximations occupy a conspicuous place in the mathematical literature owing to their interesting convergence properties as well as their connections with many branches of mathematics such as number theory, special functions, differential equations, and moment problems. Because count of their algorithmic nature, they are used in numerical analysis, computer science, automata, and electronic communication, among other fields. Their importance has grown further with the advent of fast-computing facilities. Bowman and Shenton [4] have discussed statistical applications of CFs.

3. CONTINUED FRACTIONS AND BIRTH AND DEATH PROCESSES

Let $\{X(t)\}$ be a BDP with state-dependent birth and death rates λ_n and μ_n, respectively [2]. Then the probabilities

$$P(X(t) = n | X(0) = m) = P_{mn}(t)$$

satisfy the forward Kolmogorov equations

$$\begin{aligned} P'_{m0}(t) &= -\lambda_0 P_{m0}(t) + \mu_1 P_{m1}(t), \\ P'_{mn}(t) &= \lambda_{n-1} P_{m,n-1}(t) - (\lambda_n + \mu_n) P_{mn}(t) + \mu_{n+1} P_{m,n+1}(t), \qquad n = 1, 2, 3, \ldots . \end{aligned} \tag{1}$$

Let us define the coefficients

$$\pi_0 = 1, \qquad \pi_n = \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n}, \qquad n = 1, 2, 3, \ldots .$$

The stationary distribution exists when $\sum_{n=1}^{\infty} \pi_n < \infty$, and its probabilities are given by

$$P_n = \frac{\pi_n}{\sum_{j=0}^{\infty} \pi_j}, \qquad n = 0, 1, 2, \ldots .$$

Writing $f_{mn}(z) = \int_0^{\infty} e^{-zt} P_{mn}(t)\, dt$, $\mathrm{Re}(z) > 0$, for the Laplace transform of $P_{mn}(\cdot)$, (1) yields

$$\begin{aligned} f_{00}(z) &= \frac{1}{z + \lambda_0 + (f_{01}(z)/f_{00}(z))}, \\ \frac{f_{0,n}(z)}{f_{0,n-1}(z)} &= \frac{-\lambda_{n-1}\mu_n}{z + \lambda_n + \mu_n} + \frac{f_{0,n+1}(z)}{f_{0,n(z)}}, \qquad n = 1, 2, 3, \ldots . \end{aligned}$$

These equations result in the CF expansion

$$f_{00}(z) = \frac{1 |}{| z + \lambda_0} - \frac{\lambda_0 \mu_1 |}{| z + \lambda_1 + \mu_1} - \frac{\lambda_1 \mu_2 |}{| z + \lambda_2 + \mu_2} \ldots,$$

which, in turn, leads to $P_{00}(t))$ by inverting the corresponding Laplace transform $f_{00}(z)$. Then $P_{0n}(t)$ can be recursively obtained from

$$\mu_n P_{0n}(t) = (P_{00}(t) + \cdots + P_{0,n-1}(t))' + \lambda_{n-1} P_{0,n-1}(t).$$

https://doi.org/10.1017/S0269964809990179 Published online by Cambridge University Press

4. DISCRETE DISTRIBUTIONS

Consider a discrete distribution of a random variable X on nonnegative integers; that is,

$$P(X = r) = p_r, \qquad r = 0, 1, 2, \ldots.$$

Let $\mu_1, \mu_2, \mu_3, \ldots > 0$ be given real numbers. Define $\lambda_{i-1} = (p_i/p_{i-1})\mu_i$, $i = 1, 2, 3, \ldots$. Then every BDP with birth parameters λ_{i-1} and death parameters μ_i, $i = 1, 2, 3, \ldots$, will have the same discrete distribution as its stationary distribution. We will present several examples assuming $P(X(0) = 1) = 1$. The first one is simple.

Example 1: Let $p_0 + p_1 = 1$ and $\mu_1 > 0$. Take $\lambda_1 = \mu_1 p_1/p_0$. Then the transition probability $P_{0,0}(t)$ is given by

$$P_{0,0}(t) = p_0 + (1 - p_0)e^{-(\mu_1/p_0)t}.$$

Example 2: Let $p_0 + p_1 + p_2 = 1$ and $\mu_1, \mu_2 > 0$. Take $\lambda_i = (p_{i+1}/p_i)\mu_{i+1}$, $i = 0, 1$. If $a_1 = ((p_1 + p_0) + 1)\mu_1$, $a_2 = ((p_2 + p_1) + 1)\mu_2$, and

$$\alpha_1, \alpha_2 = \frac{-(a_1 + a_2) \pm \sqrt{[}(a_1 - a_2)^2 + 4(p_2/p_1)\mu_1\mu_2]}{2},$$

then

$$P_{0,0}(t) = C_0 + C_1 e^{\alpha_1 t} + C_2 e^{\alpha_2 t},$$

where $C_0 = \mu_1\mu_2/\alpha_1\alpha_2$, $C_1 = [\alpha_1^2 + \alpha_1(\lambda_1 + \mu_1 + \mu_2) + \mu_1\mu_2]/[\alpha_1(\alpha_1 - \alpha_2)]$, and $C_2 = [\alpha_2^2 + \alpha_2(\lambda_1 + \mu_1 + \mu_2) + \mu_1\mu_2]/[\alpha_2(\alpha_2 - \alpha_1)]$.

Example 3: The probability functions of the Poisson, negative binomial, and binomial distribution satisfy *Panjer's recurrence relation* $p_k = (a + b/k)p_{k-1}$, $k = 1, 2, 3, \ldots$, where $a < 1$ and b are some constants.

(a) The case $a = 0$ and $b > 0$ leads to a Poisson distribution with parameter b.

(b) If $0 < a < 1$ and $a + b > 0$, we obtain a negative binomial distribution with parameters $1 + b/a$ and a.

(c) The case $a < 0$, $b = -a(n + 1)$, where n is a positive integer, leads to a binomial distribution with parameters $-1 - b/a$ and $a/(a - 1)$.

There are no other distributions that will satisfy this recurrence relation. These distributions have important applications in insurance [20].

In our situation, a BDP can be constructed with death rates μ_k, $k = 1, 2, 3, \ldots$, and birth rates $(a + b/k)\lambda_k$, $k = 0, 1, 2, \ldots$, which will have Poisson, binomial, or negative binomial distributions for their stationary distribution.

Example 4: Let $\{p_n\}$ follow a geometric distribution (i.e., $p_n = \rho^n(1 - \rho)$, $n = 0, 1, 2, \ldots$ and $\mu_1, \mu_2, \mu_3, \ldots > 0$. Taking $\lambda_{n-1} = \rho\mu_n$, $n = 1, 2, 3, \ldots$, the BDP with birth and death parameters λ_{n-1} and $\mu_n, n = 1, 2, 3, \ldots$, respectively, will have this geometric distribution as its stationary distribution. In particular, for an $M/M/1$ queue, we have $\mu_n = \mu,\ n = 1, 2, 3, \ldots$.

Mandelbaum, Hylnka, and Brill [10] have related the stationary distribution of BDPs to nonhomogeneous geometric random variables.

5. LERCH LIMIT

The Lerch distribution is a discrete distribution with

$$P(X = r) = \frac{cz^r}{(\nu + r)^s}, \qquad r = 0, 1, 2, \ldots, z, \nu > 0, s > 1,$$

where the normalization constant is the Lerch's transcendent

$$\frac{1}{c} = \Phi(z, s, \nu) = \sum_{r=0}^{\infty} \frac{z^r}{(\nu + r)^s}.$$

This distribution has mean $(\Phi(z, s-1, \nu)/\Phi(z, s, \nu)) - \nu$ when $s > 2$ and variance

$$(\nu + \mu)^2 + [\Phi(z, s-2, \nu) - 2(\nu + \mu)\Phi(z, s-1, \nu)]/\Phi(z, s, \nu) \quad \text{when } s > 3.$$

Moments and estimators for this distribution were derived by Zörnig and Altmann [18], and the structural properties, reliability properties, and statistical inference were investigated by Gupta, Gupta, Ong, and Srivatsava [6].

Let $\mu_1, \mu_2, \mu_3, \ldots > 0$ be arbitrary and take

$$\lambda_{n-1} = \left[\frac{\nu + n - 1}{\nu + n}\right]^s z\mu_n, \quad n = 1, 2, 3, \ldots .$$

Birth and death processes with these parameters will have a Lerch distribution for their stationary distribution.

For one particular set, we discuss the method to obtain time-dependent probabilities. Then, assuming $P(X(0) = 0) = 1$, the Laplace transform $f_{0,0}(z)$ of $P_{0,0}(t)$ is given by the CF

$$f_{0,0}(z) = \frac{1}{|z + \lambda_0} - \frac{\lambda_0\mu_1}{|z + \lambda_1 + \mu_1} - \frac{\lambda_1\mu_2}{|z + \lambda_2 + \mu_2} \cdots .$$

We thus have

$$f_{00}(z) \approx \frac{A_{n-1}(z)}{B_n(z)}, \tag{2}$$

https://doi.org/10.1017/S0269964809990179 Published online by Cambridge University Press

where $\approx$ signifies "approximately" and $B_n(z)$ can be written in tridiagonal determinant form as follows:

$$B_n(z) = \begin{vmatrix} z+\lambda_0 & 1 & & & \\ \lambda_0\mu_1 & z+\lambda_1+\mu_1 & 1 & & \\ & \lambda_1\mu_2 & z+\lambda_2+\mu_2 & 1 & \\ & & & \ddots & \\ & & & \lambda_{n-2}\mu_{n-1} & z+\lambda_{n-1}+\mu_{n-1} \end{vmatrix}_{n\times n}. \tag{3}$$

$A_{n-1}(z)$ is obtained from $B_n(z)$ by deleting the first row and first column.

We observe that $B_n(z)$ is the determinant of a quasisymmetric matrix, which can be transformed into a real symmetric matrix by a similarity transformation. The latter is a real symmetric diagonal-dominant positive-definite tridiagonal matrix with nonzero subdiagonal elements and, therefore, the eigenvalues are real and distinct. We denote these eigenvalues by $s_1^n, s_2^n, \ldots, s_n^n$. Similarly, the roots of $A_{n-1}(z)$ are negative, real, and distinct, and we denote them by $z_1^n, z_2^n, \ldots, z_{n-1}^n$. Using partial fractions, (2) can be expressed as

$$f_{0,0}(s) \approx \sum_{j=1}^{n} \frac{\prod_{r=1}^{n-1}\left(z_r^{(n)} - s_j^{(n)}\right)}{\left(z+s_j^{(n)}\right)\prod_{r=1,r\neq j}^{n}\left(s_r^{(n)} - s_j^{(n)}\right)}.$$

On inverting, we get

$$P_{0,0}(t) \approx \sum_{j=1}^{n} \frac{\prod_{r=1}^{n-1}\left(z_r^{(n)} - s_j^{(n)}\right)}{\prod_{r=1,r\neq j}^{n}\left(s_r^{(n)} - s_j^{(n)}\right)} e^{-s_j^{(n)}t}. \tag{4}$$

Murphy and O'Donohoe [12] have discussed the method of finding $P_r(t)$ with initial number m in the system. Specifically,

$$P_r(t) \approx \sum_{j=1}^{k} H_j^{(r)} e^{-s_j^{(k)}t}, \qquad r = 0, 1, 2, \ldots, \tag{5}$$

where

$$H_j^{(r)} := \frac{B_r\left(-s_j^{(k)}\right)B_m\left(-s_j^{(k)}\right)A_{k-1}\left(-s_j^{(k)}\right)}{\left(\prod_{i=0}^{m-1}\lambda_i\right)\left(\prod_{i=1}^{r}\mu_i\right)B_k'\left(-s_j^{(k)}\right)}, \tag{6}$$

$$B_k'\left(-s_j^{(k)}\right) = \prod_{i=1,i\neq j}^{k}\left(s_i^{(k)} - s_j^{(k)}\right),$$

and

$$k = \begin{cases} m+n+1 & \text{for } r \leq m \\ r+n+1 & \text{for } r \geq m. \end{cases}$$

https://doi.org/10.1017/S0269964809990179 Published online by Cambridge University Press

The error due to the truncation of the CF is given by (Murphy and O'Donohoe [12])

$$T_n(P_r(t)) = \frac{\left(\prod_{i=0}^{m+n-1}\lambda_i\right)\left(\prod_{i=1}^{m+n}\mu_i\right)t^{2n+m-r}}{\left(\prod_{i=0}^{m-1}\lambda_i\right)\left(\prod_{i=1}^{r}\mu_i\right)(2n+m-r)!}\left\{(1+t\rho_{r,m+n})^{-2n-m+r+1}+O(t^2)\right\}$$

for $r \leq m$ and

$$T_n(P_r(t)) = \frac{\left(\prod_{i=0}^{r+n-1}\lambda_i\right)\left(\prod_{i=1}^{r+n}\mu_i\right)t^{2n+r-m}}{\left(\prod_{i=0}^{m-1}\lambda_i\right)\left(\prod_{i=1}^{r}\mu_i\right)(2n+r-m)!}\left\{(1+t\rho_{r,r+n})^{-2n-r+m+1}+O(t^2)\right\}$$

for $r \geq m$, where

$$\rho_{r,m+n} = \frac{\sigma_{m+n}+\sigma_{m+n+1}-\sigma_m-\sigma_r}{(2n+m-r+1)(2n+m-r-1)}$$

and

$$\sigma_n = \lambda_0 + \sum_{r=1}^{n-1}(\lambda_r+\mu_r).$$

Given a value of n (the nth convergent) and a sufficiently small error ϵ, the above error formulas can be used to estimate a range of t for which this error is not exceeded.

6. NUMERICAL WORK

In this section we examine some numerical examples with exponent $s = 4$ having Lerch limit

$$P(X=r) = \frac{cz^r}{(\nu+r)^4} \quad (r = 0, 1, 2, \dots)$$

with $\nu = 1$ and $\nu = 4$. We choose $\mu_k = 1/z$ $(k = 1, 2, 3, \dots)$ and, hence,

$$\lambda_{k-1} = \left(\frac{\nu+k-1}{\nu+k}\right)^s, \qquad k = 1, 2, 3, \dots .$$

The computation of P_0 in (4) requires careful consideration. Computing the numerator and denominator separately and dividing subsequently leads to a irreversible loss of precision. The numerically most stable way would be to order the $n-1$ terms in numerator and denominator in decreasing order of magnitude, dividing the corresponding terms, and multiplying.

However, in our experience, it suffices to take the ordered roots $z_r^{(n)}$ and $s_r^{(n)}$; then $z_r^{(n)} - s_j^{(n)}$ and $s_r^{(n)} - s_j^{(n)}$ are of comparable magnitude, and dividing both terms and multiplying the $n-1$ factors leads to a stable procedure. Since numerical routines for eigenvalue computation typically deliver ordered eigenvalues, no additional ordering is necessary.

https://doi.org/10.1017/S0269964809990179 Published online by Cambridge University Press

Note that these remarks also apply for the computation of P_r in (5) since $H_j^{(r)}$ in (6) contains the same products as above (with n replaced by $k = n + 1 + \max\{m, r\}$).

In the following examples, we assume that the initial number in the system is $m = 3$, and we draw graphs for $P_r(t), r = 0, 1, 2, 3$ and the mean function $EX(t)$.

As the first example, we take $z = 1/16$; hence, $\mu_k = 1/z = 16$. For $\nu = 1$ and 4, the stationary probabilities are the following:

$$\nu = 1: \qquad p_0 = 0.996, \qquad p_1 = 3.89 \times 10^{-3}, \qquad p_2 = 4.80 \times 10^{-5},$$
$$p_3 = 9.50 \times 10^{-7};$$
$$\nu = 4: \qquad p_0 = 0.974, \qquad p_1 = 2.49 \times 10^{-2}, \qquad p_2 = 7.52 \times 10^{-4},$$
$$p_3 = 2.54 \times 10^{-5}.$$

Figure 1 shows the transient probabilities and the mean for $\nu = 1$. For $\nu = 4$, the plot looks very similar and is omitted.

As the second example, we take $z = 1/2$, $\mu_k = 2$. For $\nu = 1$ and 4, the stationary probabilities are the following

$$\nu = 1: \qquad p_0 = 0.966, \qquad p_1 = 3.02 \times 10^{-2}, \qquad p_2 = 2.98 \times 10^{-3},$$
$$p_3 = 4.72 \times 10^{-4};$$
$$\nu = 4: \qquad p_0 = 0.785, \qquad p_1 = 0.161, \qquad p_2 = 0.0388,$$
$$p_3 = 0.0105.$$

Figures 2 and 3 show the transient behavior for $\nu = 1$ and $\nu = 4$, respectively.

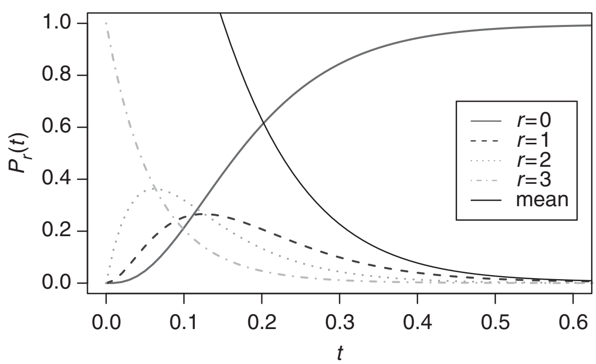

FIGURE 1. Transient behavior for $z = 1/16, \mu = 16, s = 4, m = 3$, and $\nu = 1$.

https://doi.org/10.1017/S0269964809990179 Published online by Cambridge University Press

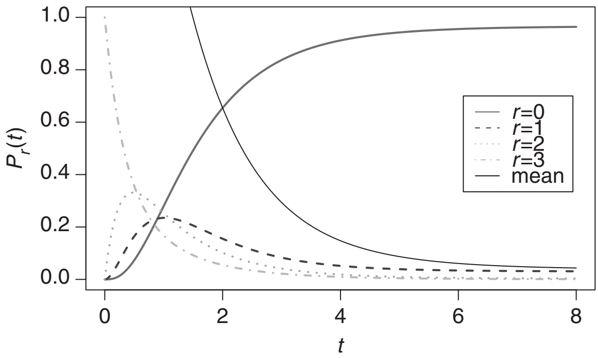

FIGURE 2. Transient behavior for $z = 1/2, \mu = 2, s = 4, m = 3$, and $\nu = 1$.

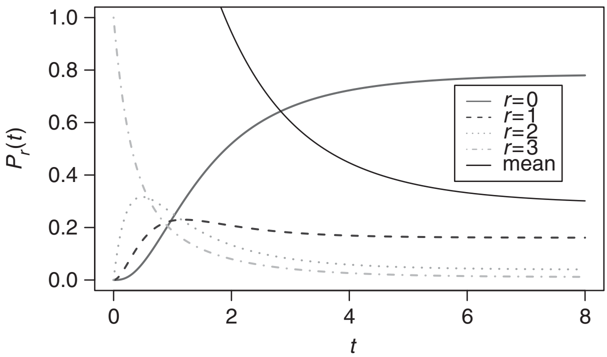

FIGURE 3. Transient behavior for $z = 1/2, \mu = 2, s = 4, m = 3$, and $\nu = 4$.

In the last example, we use $z = \mu_k = 1$. For $\nu = 1$ and 4, the stationary probabilities are the following:

$$\nu = 1: \qquad p_0 = 0.924, \qquad p_1 = 5.77 \times 10^{-2}, \qquad p_2 = 1.14 \times 10^{-2},$$
$$p_3 = 3.61 \times 10^{-3};$$
$$\nu = 4: \qquad p_0 = 0.523, \qquad p_1 = 0.214, \qquad p_2 = 0.103, \qquad p_3 = 0.0557.$$

Graphs of the transient behavior for $\nu = 1$ and $\nu = 4$ are shown in Figures 4 and 5, respectively.

To illustrate the speed of convergence of the numerical method in Section 5, consider the case $\nu = 1$. The computation of P_0 and P_1, accurate to six significant digits, needs $k = 8$ for $z = 1/16$ and $k = 16$ for $z = 1/2$. For $z = 0.9$ and $z = 1$, we have to use approximately $k = 45$ and $k = 60$. Hence, the speed of convergence decreases with increasing value of z; but in all cases, $k = 100$ is enough to get highly reliable results.

https://doi.org/10.1017/S0269964809990179 Published online by Cambridge University Press

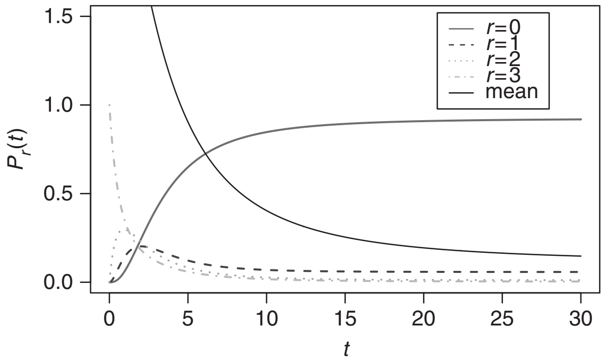

FIGURE 4. Transient behavior for $z = 1, \mu = 1, s = 4, m = 3$, and $\nu = 1$.

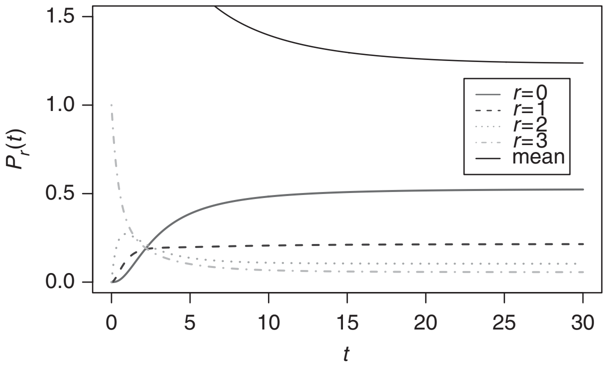

FIGURE 5. Transient behavior for $z = 1, \mu = 1, s = 4, m = 3$, and $\nu = 4$.

6.1. Accelerating the Speed of Convergence for the Zipf Limit

The last section showed that the computation is possible with high accuracy even for the case $\nu = z = 1$, which leads to the Zipf limit law:

$$P(X = r) = \frac{1}{\zeta(s)} \frac{1}{(1+r)^s}, \qquad r = 0, 1, 2, \dots . \tag{7}$$

However, in addition to the parameter of the Zipf distribution, the speed of convergence depends crucially on the birth and death rates.

If we put $\lambda_0 = 1$ and $\lambda_k = \mu_k = (k+1)^s$, $k = 1, 2, \dots$, we obtain, again, the Zipf law (7) as a limiting distribution, but the development of the process over time and the convergence behavior of the numerical procedure are completely different.

TABLE 1. Speed of Convergence for $P_0(t)$ with Unmodified and Modified Rates

	k	$t = 0.5$	$t = 1$	$t = 2$	$t = 5$
Unmodified rates	20	0.9117042	0.8905969	0.8499887	0.7389401
	50	0.9192899	0.9106700	0.8939266	0.8455217
	200	0.9229279	0.9206318	0.9163837	0.9037576
	1000	0.9238799	0.9232780	0.9224247	0.9198709
	5000	0.9240693	0.9238067	0.9236356	0.9231237
Modification 1	20	0.9235242	0.9222317	0.9199904	0.9133002
	50	0.9240222	0.9236713	0.9233226	0.9222785
	200	0.9241107	0.9239223	0.9239004	0.9238360
	1000	0.9241164	0.9239382	0.9239369	0.9239343
	5000	0.9241166	0.9239388	0.9239383	0.9239382
Modification 2	10	0.9241167	0.9239389	0.9239384	0.9239384
	20	0.9241167	0.9239389	0.9239384	0.9239384
	50	0.9241167	0.9239388	0.9239384	0.9239384
	100	0.9241167	0.9239388	0.9239384	0.9239384

As an example, we consider the case $s = 4$ and $m = 0$. The upper part of Table 1 shows the approximation of $P_0(t)$ for $t = 0.5, 1, 2$, and 5. We see very slow convergence, becoming worse as time increases. Comparing $P_0(2)$ with the exact value $1/\zeta(4) = 0.9239384$ shows that even for $k = 5000$, we obtain only three correct digits.

However, we have found that a simple modification of the highest birth and death rates, which appear in $B_n(z)$ in (3), dramatically increases the speed of convergence. This modification consists in an enlargement of μ_{k-1} and a downsizing of λ_{k-1}. Specifically, we used the following:

$$\text{Modification 1:} \qquad \lambda_{k-1}^{\text{mod}} = \lambda_{k-1}/(k-1), \qquad \mu_{k-1}^{\text{mod}} = (k-1)\mu_{k-1};$$

$$\text{Modification 2:} \qquad \lambda_{k-1}^{\text{mod}} = \lambda_{k-1}/(k-1)^2, \quad \mu_{k-1}^{\text{mod}} = (k-1)^2\mu_{k-1}.$$

The results can be found in the middle and lower parts of Table 1. For modification 1, we have four correct digits of $P_0(2)$ for $k = 200$, a considerable improvement. Even more astonishing, modification 2 delivers the correct value (up to seven digits) even for $k = 10$. We found the same behavior for other values of m, r, and s. A theoretical explanation of this phenomenon is an open question.

6.2. Computing the Transient Distribution via the Matrix Exponential

In this subsection we compare the numerical methods described in Section 5 with the direct solution of the forward Kolmogorov equations in (1). The first k equations can be written as

$$\mathbf{P}'(t) = \mathbf{QP}(t), \tag{8}$$

https://doi.org/10.1017/S0269964809990179 Published online by Cambridge University Press

where

$$\mathbf{P}(t) = \begin{pmatrix} P_{m0}(t) \\ P_{m1}(t) \\ \vdots \\ P_{m,k-1}(t) \end{pmatrix},$$

$$\mathbf{Q} = \begin{pmatrix} -\lambda_0 & \mu_1 & & & \\ \lambda_0 & -\lambda_1 - \mu_1 & \mu_2 & & \\ & \ddots & \ddots & \ddots & \\ & & & \lambda_{k-2} & -\lambda_{k-1} - \mu_{k-1} \end{pmatrix}_{k\times k}.$$

The differential equation (8) has the solution

$$\mathbf{P}(t) = e^{\mathbf{Q}t}\,\mathbf{P}(0).$$

There are various ways to compute the matrix exponential $e^{\mathbf{Q}t}$ (see, e.g., Moler and Van Loan [11]); we used the function `expm` in the R library `Matrix` [3], which uses Ward's diagonal Pade approximation with three-step preconditioning; this method is among the procedures recommended by Moler and Van Loan [11].

The results with the same birth and death rates and parameter values as in the example in Section 6.1 are shown in Table 2; thus, they can be compared directly with the results in Table 1.

For the unmodified birth and death rates, the entries in the upper part of Table 2 are in very good agreement with the corresponding entries in Table 1. The two modifications of the highest birth and death rates again lead to a significant increase in the speed of convergence; however, the effect in not as pronounced as in Section 6.1.

Table 2. Speed of Convergence for $P_0(t)$ with Unmodified and Modified Rates by Computing the Matrix Exponential

	k	$t = 0.5$	$t = 1$	$t = 2$	$t = 5$
$\mathbf{Q}$ with unmodified rates	20	0.9117042	0.8905969	0.8499887	0.7389400
	50	0.9192899	0.9106700	0.8939266	0.8455217
	200	0.9229279	0.9206318	0.9163836	0.9037575
	1000	0.9238798	0.9232778	0.9224244	0.9198700
$\mathbf{Q}$ with modification 1	20	0.9235242	0.9222317	0.9199904	0.9133002
	50	0.9240222	0.9236713	0.9233226	0.9222784
	200	0.9241107	0.9239222	0.9239002	0.9238355
$\mathbf{Q}$ with modification 2	20	0.9241519	0.9239721	0.9239651	0.9239454
	50	0.9241189	0.9239410	0.9239404	0.9239398
	100	0.9241169	0.9239390	0.9239384	0.9239380
$\tilde{\mathbf{Q}}$ with unmodified rates	20	0.9241486	0.9239718	0.9239714	0.9239714
	50	0.9241188	0.9239410	0.9239406	0.9239406
	100	0.9241169	0.9239391	0.9239386	0.9239386
	200	0.9241167	0.9239388	0.9239384	0.9239383

The main disadvantage of the direct computation of the matrix exponential consists of the large time and memory requirements: The computation for $k = 1000$ needed about 30 min compared to 5 s for the method in Section 6.1. Further, $k = 5000$ did not work due to memory overflow.

As an alternative approach, one can truncate the state space to $\{0, 1, \ldots, k-1\}$. Then the forward Kolmogorov equations of this finite BDP can be written as in (1) for $j = 0, \ldots, k-2$, but for $k-1$, we obtain

$$P'_{m,k-1}(t) = \lambda_{k-2}P_{m,k-2}(t) - \mu_{k-1}P_{m,k-1}(t).$$

Therefore, $\mathbf{P}'(t) = \tilde{\mathbf{Q}}\mathbf{P}(t)$, where $\tilde{\mathbf{Q}}$ is the generator of the truncated process; that is,

$$\tilde{\mathbf{Q}} = \begin{pmatrix} -\lambda_0 & \mu_1 & & & \\ \lambda_0 & -\lambda_1 - \mu_1 & \mu_2 & & \\ & \ddots & \ddots & \ddots & \\ & & & \lambda_{k-2} & -\mu_{k-1} \end{pmatrix}_{k\times k}.$$

The differential equation now has the solution $\mathbf{P}(t) = e^{\tilde{\mathbf{Q}}t}\mathbf{P}(0)$. For large k, one can expect the same solution as earlier; however, the convergence is much faster with $\mathbf{Q}$, as the last part of Table 2 clearly shows.

An unreckoned bottom line from this section follows that the choice of the highest birth and death rates is decisive for the quality of the solution when treating an infinite BDP numerically.

7. SPEED OF CONVERGENCE TO STATIONARITY

Consider a BDP $\{X(t)\}$ with birth rates $\lambda_0, \lambda_1, \lambda_2, \ldots$ and death rates $\mu_0, \mu_1, \mu_2, \ldots$. We assume $X(0) = 0$, $\sum_{j=0}^{\infty} \pi_j < \infty$, and $\sum_{j=0}^{\infty} (\lambda_j \pi_j)^{-1} = \infty$, where

$$\pi_0 = 1 \quad \text{and} \quad \pi_j = \frac{\lambda_0\lambda_1 \cdots \lambda_{j-1}}{\mu_0\mu_1 \cdots \mu_j}, \qquad j \geq 1,$$

as in Section 3. Then $\{X(t)\}$ is uniquely determined by its birth and death rates, $P_j = \pi_j / \sum_{k=0}^{\infty} \pi_k$, where $P_j = \lim_{t\to\infty} P_{0j}(t)$ and $P_{0j}(t) = P(X(t) = j | X(0) = 0)$. Further, $\lim_{t\to\infty} \mathbb{E}X(t) = \mathbb{E}X$, with X denoting a random variable with distribution $\{P_j, j \geq 0\}$.

The speed of convergence to stationarity of the BDP is usually characterized by

$$\alpha^\star = \sup\{\alpha \geq 0 : P_j - P_{0j}(t) = O(e^{-\alpha t}) \text{ as } t \to \infty \text{ for all } j \geq 0\}$$

or its reciprocal $1/\alpha^\star$, the relaxation time. However, there exists no general expression for $\alpha^\star$ in terms of the birth and death rates. As an alternative measure, Stadje and

https://doi.org/10.1017/S0269964809990179 Published online by Cambridge University Press

Parthasarathy [16] and Coolen-Schrijner and Van Doorn [5] considered the quantities $I = \int_0^\infty [\mathbb{E}X - \mathbb{E}X(t)]\,dt$ and the normalized value

$$m = \int_0^\infty [1 - \mathbb{E}X(t)/\mathbb{E}X]\,dt = I/\mathbb{E}X.$$

Putting

$$T = \sum_{k=0}^{\infty} \frac{r_k^2}{\lambda_k P_k}, \qquad \tau_0 = 0, \qquad \tau_j = P_j \sum_{k=0}^{j-1} \frac{r_k}{\lambda_k P_k}, \quad j = 1, 2, \ldots,$$

where $r_k = \sum_{l=k+1}^{\infty} P_l$ for $k \geq 0$, Coolen-Schrijner and Van Doorn [5] proved that

$$I_j = \int_0^\infty [P_{0j}(t) - P_j]\,dt = TP_j - \tau_j, \qquad j = 0, 1, 2, \ldots\,,$$

and $I = \sum_{j=0}^{\infty} j\tau_j - T \cdot \mathbb{E}X$.

We computed these quantities for the examples in Section 6; parts of the results are given in Table 3. The upper part of Table 3 shows I_0, I_1, I_2, T, I, and m for $\mu_k = 1/z$ and

$$\lambda_{k-1} = \left(\frac{\nu + k - 1}{\nu + k}\right)^s, k \geq 1.$$

We refer to this choice of birth and death rates as in case 1. The middle and lower part of Table 3 show the same quantities for $\mu_k = (\nu + k)^s/z$, $\lambda_{k-1} = (\nu + k - 1)^s, k \geq 1$ (case 2) and $\mu_k = (\nu + k)^s$, $\lambda_{k-1} = z(\nu + k - 1)^s, k \geq 1$ (case 3), respectively. We always chose $s = 4$ and the same values of z and ν for the three sets of rates.

Although the limiting distribution is the same in all three cases, Table 3 shows that the speed of convergence to stationarity is quite different. Looking at the normalized

Table 3. Speed of Convergence to Stationarity for Different Birth and Death Rates

z	ν	I_0	I_1	I_2	T	I	m
1/16	1	2.5e-04	−2.5e-04	−6.1e-06	2.5e-04	2.6e-04	6.5e-02
1/16	4	1.7e-03	−1.6e-03	−9.6e-05	1.7e-03	1.8e-03	6.7e-02
1/2	1	0.021	−0.016	−0.0034	0.022	0.027	0.71
1/2	4	0.15	−0.076	−0.045	0.20	0.29	0.99
1/16	1	1.6e-05	−1.5e-05	−2.3e-07	1.6e-05	1.6e-05	4.0e-03
1/16	4	2.6e-06	−2.5e-06	−1.1e-07	2.7e-06	2.7e-06	1.0e-04
1/2	1	1.2e-03	−1.0e-03	−1.2e-04	1.2e-03	1.4e-03	3.5e-02
1/2	4	2.1e-04	−1.3e-04	−5.2e-05	2.6e-04	3.2e-04	1.1e-03
1/16	1	2.5e-04	−2.5e-04	−3.6e-06	2.5e-04	2.5e-04	6.3e-02
1/16	4	4.2e-05	−4.0e-05	−1.8e-06	4.3e-05	4.4e-05	1.7e-03
1/2	1	2.3e-03	−2.0e-03	−2.5e-04	2.4e-03	2.7e-03	7.1e-02
1/2	4	4.1e-04	−2.6e-04	−1.0e-04	5.3e-04	6.5e-04	2.2e-03

value m in the last column, we see that the convergence in the first case is slower than in the other two cases.

As a general rule, large birth and death rates lead to fast convergence to stationarity. To be specific, we consider the choice $z = 1/2$ and $\nu = 4$. Here, the birth and death rates in case 1 are much smaller than in case 3, which, in turn, are smaller than in case 2. Accordingly, convergence is much slower in the first than in the other two cases, and slower in the third than in the second case.

Next, let $z = 1/16$ and $\nu = 1$. Here, the rates for cases 1 and 3 are smaller than for case 2; hence, convergence is slower. Between case 1 and 3, things are mixed, which leads to a similar speed of convergence.

Remark: All computation in this and the preceding section are done using the statistical software R [15].

8. CONCLUSIONS

In view of the importance of Lerch's distribution in several disciplines, it is of interest to construct a function of time that converges to the Lerch distribution as time becomes large. This is achieved by recourse to BDPs. We applied a numerical procedure using CFs to obtain time-dependent probabilities. There, the truncation of the number of terms plays an important role. A modified numerical procedure with altered highest birth and death rates leads to a much faster rate of convergence and thereby a considerable reduction in computation time. As alternative method, we considered the direct computation of the matrix exponential; again, the highest rates play a crucial role. Finally, we applied a recently proposed criterion to describe the speed of convergence to stationarity, which highly depends on the parameter of the process.

References

1. Aksenov, S.V. & Savageau, M.A. (2008). Some properties of the Lerch family of discrete distributions. http://arXiv.org/abs/math/0504485v1.
2. Anderson, W.J. (1991). *Continuous-time Markov chains: An applications-oriented approach*. New York:Springer.
3. Bates, D. & Maechler, M. (2008). Matrix: Sparse and dense matrix classes and methods. R package version 0.999375-16. http://www.R-project.org
4. Bowman, K.O. & Shenton, L.R. *Continued fractions in statistical applications.* Statistics: Textbooks and Monographs, vol. 103. New York:Marcel Dekker.
5. Coolen-Schrijner, P. & Van Doorn, E.A. (2001). On the convergence to stationarity of birth-death processes. *Journal of Applied Probability* 38: 696–706.
6. Gupta, P.L., Gupta, R.C., Ong, S., & Srivatsava, H.M. (2008). A class of Hurwitz–Lerch zeta distributions and their applications in reliability. *Applied Mathematics and Computation* 196: 521–531.
7. Hill, B.M. (1974). The rank-frequency form of Zipf's law. *Journal of The American Statistical Association* 69: 1017–1026.
8. Kulasekera, K.B. & Tonkyn, D.W. (1992). A new discrete distribution, with applications to survival, dispersal and dispersion. *Communications in Statistics: Simulation Computation* 21(2): 499–518.

https://doi.org/10.1017/S0269964809990179 Published online by Cambridge University Press

9. Lorentzen, L. & Waadeland, H. (1992). *Continued fractions with applications.* Studies in Computational Mathematics, Vol. 3. Amsterdam: North-Holland.
10. Mandelbaum, M., Hylnka, M., & Brill, H.P. (2007). Nonhomogeneous geometric distributions with relations to birth and death processes. *TOP* 15: 281–296.
11. Moler, C. & Van Loan, C. (2003). Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review* 45: 3–49.
12. Murphy, J.A. & O'Donohoe, M.R. (1975). Some properties of continued fractions with applications in Markov processes. *Journal of Institute of Mathematics and Its Application* 16: 57–71.
13. Naldi, M. (2003). Concentration indices and Zipf's law. *Economic Letters* 78: 329–334.
14. Parthasarathy, P.R. & Lenin, R.B. (2004). *Birth and death process (BDP) models with applications—queueing, communication systems, chemical models, biological models: The state-of-the-art with a time-dependent perspective.* American Series in Mathematical and Management Sciences 51. Syracuse, NY: American Science Press.
15. R Development Core Team (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Available from http://www.R-project.org.
16. Stadje, W. & Parthasarathy, P.R. (1999). On the convergence to stationarity of the many-server Poisson queue. *Journal of Applied Probability* 36: 546–557.
17. Zipf, G.K. (1949). *Human behaviour and the principle of the least effort*. Reading, MA:Addison Wesley.
18. Zörnig, P. & Altmann, G. (1995). Unified representations of Zipf's distributions. *Computational Statistics & Data Analysis* 19: 461–473.
19. Jones, W.B. & Thron, W.J. (1980). *Continued fractions. Analytic theory and applications. Encyclopedia of Mathematics and its Applications, 11*. Reading MA: Addison-Wesley Publishing Co.
20. Rolski, T., Schmidli, H., Schmidt, V. & Teugels, J. (1999). *Stochastic processes for insurance and finance*. *Wiley Series in Probability and Statistics*. New York, NY: John Wiley & Sons, Ltd.