

Lecture 4b: Continuous-Time Birth and Death Processes

Continuous-time Markov chains are stochastic processes whose time is continuous, $t \in [0, \infty)$, but the random variables are discrete. Prominent examples of continuous-time Markov processes are Poisson and death and birth processes. These processes play a fundamental role in the theory and applications that embrace queueing and inventory models, population growth, engineering systems, etc [3]. Here we discuss the birth and death process with its invariants.

1 Preliminaries

1.1 Probability Generating Functions

The probability generating function is a concept used for distributions with non-negative integers.

Consider a discrete random variable X taking nonnegative values. Let f denote the probability mass function (p.m.f.) of X , defined as

$$f_k = \text{Prob}\{X = k\} = P_k, \quad k = 0, 1, 2, \dots$$

where $\sum_{k=0}^{\infty} P_k = 1$.

The probability generating function (p.g.f.) of X , denoted \mathbb{P}_X , is defined as

$$\mathbb{P}_X(z) = E[z^X] = \sum_{k=0}^{\infty} P_k z^k, \quad (1)$$

for some $z \in \mathbb{R}$, or, inversely,

$$P_k(t) = \frac{1}{k!} \left. \frac{\partial^k \mathbb{P}_X(z)}{\partial z^k} \right|_{z=0}. \quad (2)$$

Note that

- The series converges absolutely for $|z| \leq 1$:

$$\begin{aligned}\mathbb{P}_X(1) &= \sum_{k=0}^{\infty} P_k(1)^k \\ &= \sum_{k=0}^{\infty} P_k \\ &= 1.\end{aligned}$$

Thus, $\mathbb{P}_X(z)$ is well defined for $|z| \leq 1$.

- Also, the p.g.f. generates the probabilities associated with the distribution

$$\mathbb{P}_X(0) = P_0, \quad \mathbb{P}'_X(0) = P_1, \quad \mathbb{P}''_X(0) = 2!P_2,$$

which are resulted from, by taking

$$\mathbb{P}_X(z) = P_0 + P_1 z + P_2 z^2 + P_3 z^3 + \dots$$

whence at $z = 0$, then $P_0 = \mathbb{P}_X(0)$. And

$$\frac{\mathbb{P}_X(z)}{\partial z} = P_1 + 2P_2 z + 3P_3 z^2 + \dots$$

whence at $z = 0$, then $P_1 = \mathbb{P}'_X(0)$. And so forth.

- In general, the n th derivative of the p.g.f. of X is

$$\mathbb{P}_X^{(n)}(0) = n!P_n.$$

The p.g.f. can be used to calculate the mean and variance of a random variable X :

- (1) The first derivative of \mathbb{P}_X is

$$\mathbb{P}'_X(z) = \frac{d\mathbb{P}_X(z)}{dz} = \sum_{k=1}^{\infty} k P_k z^{k-1}, \quad \text{for } -1 < z < 1.$$

By letting z approach one from the left, $z \rightarrow 1^-$, we obtain the mean of X

$$\mathbb{P}'_X(1) = \sum_{k=1}^{\infty} k P_k = E[X] = \mu_X.$$

(2) The second derivative of \mathbb{P}_X is

$$\mathbb{P}_X''(z) = \frac{d^2\mathbb{P}_X(z)}{dz^2} = \sum_{k=1}^{\infty} k(k-1)P_k z^{k-2},$$

so that as $z \rightarrow 1^-$,

$$\mathbb{P}_X''(1) = \sum_{k=1}^{\infty} k(k-1)P_k = E[X(X-1)] = E[X^2 - X].$$

If the mean is finite, then the variance of X is

$$\begin{aligned} \mathbb{P}_X''(1) + \mathbb{P}_X'(1) - (\mathbb{P}_X'(1))^2 &= E[X^2 - X] + E[X] - (E[X])^2 \\ &= E[X^2] - E[X] + E[X] - (E[X])^2 \\ &= E[X^2] - (E[X])^2 \\ &= \text{Var}(X). \end{aligned}$$

2 Birth and Death Processes

Consider a large class of infinite state space, continuous-time Markov chain whose states are $\{0, 1, 2, \dots\}$. Transitions of state may go only from n to $n+1$ or from n to $n-1$. We can view the state of the system as the size of a population that can increase by a “birth” ($n \rightarrow n+1$) or decrease by a “death” ($n \rightarrow n-1$). Suppose that whenever there are n people in the system, then: (i) new individuals enter the system at an exponential rate λ_n , and (ii) individuals leave the system at an exponential rate μ_n . That is, whenever there are n people in the system, the time until the next arrival is exponentially distributed with mean $1/\lambda_n$ and is independent of the time until the next departure which is also exponentially distributed but with mean $1/\mu_n$. The parameters $\{\lambda_n\}_{n=0}^{\infty}$ and $\{\mu_n\}_{n=1}^{\infty}$ are called the arrival (or birth) and departure (or death) rates, respectively. If the population is 0, there are no deaths and $\mu_0 = 0$ [5, 7].

2.1 General Birth and Death Process

The continuous-time birth and death Markov chain $\{X(t) : t \in [0, \infty)\}$ may have either a finite $\{0, 1, 2, \dots, N\}$ or infinite $\{0, 1, 2, \dots\}$ state space. Assume that its transition probabilities $P_{ij}(t)$ are stationary, *i.e.*,

$$P_{ij}(t) = \text{Prob}\{X(t+s) = j | X(s) = i\}, \quad \text{for all } \Delta t \geq 0.$$

In addition, assume the infinitesimal transition probabilities for this process are [1]

$$\begin{aligned}
P_{i,i+j}(\Delta t) &= \text{Prob}\{X(t + \Delta t) - X(t) = j \mid X(t) = i\} \\
&= \begin{cases} \lambda_i \Delta t + o(\Delta t), & j = 1 \\ \mu_i \Delta t + o(\Delta t), & j = -1 \\ 1 - (\lambda_i + \mu_i) \Delta t + o(\Delta t), & j = 0 \\ o(\Delta t), & j \neq -1, 0, 1. \end{cases} \tag{3}
\end{aligned}$$

for Δt sufficiently small, $\mu_0 = 0$, $\lambda_0 > 0$, and $\lambda_i > 0$, $\mu_i > 0$ for $i = 1, 2, \dots$. It is often the case that $\lambda_0 = 0$, except when there is immigration. The initial conditions are

$$P_{ij}(0) = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases}$$

In a small time interval Δt , at most one change in state can occur, either a birth, $i \rightarrow i + 1$ or a death, $i \rightarrow i - 1$ [1].

In the same way as for the Poisson process, we define $P_i(t) = \text{Prob}\{X(t) = i\}$ and assume $X(0) = 0$. If $\Delta t > 0$, $i \geq 1$, by invoking the law of total probability and the Markov property, we obtain from the Chapman-Kolmogorov equations that,

$$\begin{aligned}
P_{ij}(t + \Delta t) &= \sum_{k=0}^{\infty} P_{ik}(t) P_{kj}(\Delta t) \\
&= \sum_{k=0}^{\infty} P_{ik}(t) \cdot \text{Prob}\{X(t + \Delta t) = j \mid X(t) = k\} \\
&= \sum_{k=0}^{\infty} P_{ik}(t) \cdot \text{Prob}\{X(t + \Delta t) - X(t) = j - k \mid X(t) = k\} \\
&= \sum_{k=0}^{k=j} P_{ik}(t) \cdot \text{Prob}\{X(t + \Delta t) - X(t) = j - k \mid X(t) = k\}.
\end{aligned}$$

Now, for $k = j$, for the right hand side we have

$$\begin{aligned}
&= P_{ij}(t) \cdot \text{Prob}\{X(t + \Delta t) - X(t) = j - j \mid X(t) = j\} \\
&= P_{ij}(t) \cdot \text{Prob}\{X(t + \Delta t) - X(t) = 0 \mid X(t) = j\} \\
&= P_{ij}(t) \cdot [1 - (\lambda_j + \mu_j) \Delta t + o(\Delta t)],
\end{aligned}$$

for $k = j - 1$,

$$\begin{aligned} &= P_{i,j-1}(t) \cdot \text{Prob}\{X(t + \Delta t) - X(t) = j - (j - 1) \mid X(t) = j - 1\} \\ &= P_{i,j-1}(t) \cdot \text{Prob}\{X(t + \Delta t) - X(t) = 1 \mid X(t) = j - 1\} \\ &= P_{i,j-1}(t) \cdot [\lambda_{j-1}\Delta t + o(\Delta t)], \end{aligned}$$

for $k = j + 1$,

$$\begin{aligned} &= P_{i,j+1}(t) \cdot \text{Prob}\{X(t + \Delta t) - X(t) = j - (j + 1) \mid X(t) = j + 1\} \\ &= P_{i,j+1}(t) \cdot \text{Prob}\{X(t + \Delta t) - X(t) = -1 \mid X(t) = j + 1\} \\ &= P_{i,j+1}(t) \cdot [\mu_{j+1}\Delta t + o(\Delta t)], \end{aligned}$$

whereas for $k \neq j, j - 1, j + 1$, that is for $k \leq j - 2$,

$$\begin{aligned} &= P_{ik}(t) \cdot \text{Prob}\{X(t + \Delta t) - X(t) \geq 2 \mid X(t) = k\} \\ &= P_{ik}(t) \cdot o(\Delta t), \end{aligned}$$

and $k \geq j + 2$,

$$\begin{aligned} &= P_{ik}(t) \cdot \text{Prob}\{X(t + \Delta t) - X(t) \leq -2 \mid X(t) = k\} \\ &= P_{ik}(t) \cdot o(\Delta t). \end{aligned}$$

Collecting all together,

$$\begin{aligned} P_{ij}(t + \Delta t) &= P_{ij}(t) \cdot [1 - (\lambda_j + \mu_j)\Delta t + o(\Delta t)] + P_{i,j-1}(t) \cdot [\lambda_{j-1}\Delta t + o(\Delta t)] + \\ &\quad P_{i,j+1}(t) \cdot [\mu_{j+1}\Delta t + o(\Delta t)] + P_{ik}(t) \cdot o(\Delta t) \\ &= P_{i,j-1}(t)\lambda_{j-1}\Delta t + P_{i,j+1}(t)\mu_{j+1}\Delta t + P_{ij}(t)[1 - (\lambda_j + \mu_j)\Delta t] + o(\Delta t), \end{aligned}$$

which holds for all i and j in the state space with the exception of $j = 0$ and $j = N$ (if the population size is finite).

If $j = 0$, then (and due to $\mu_0 = 0$)

$$P_{i0}(t + \Delta t) = P_{i1}(t)\mu_1\Delta t + P_{i0}(t)[1 - \lambda_0\Delta t] + o(\Delta t).$$

If $j = N$, which is the maximum population size, then

$$P_{iN}(t + \Delta t) = P_{i,N-1}(t)\lambda_{N-1}\Delta t + P_{iN}(t)[1 - \mu_N]\Delta t + o(\Delta t),$$

where $\lambda_N = 0$ and $P_{kN} = 0$ for $k > N$. Subtracting $P_{ij}(t)$, $P_{i0}(t)$, and $P_{iN}(t)$ from the preceding three equations, respectively, dividing by Δt , and taking the limit as $\Delta t \rightarrow 0$, yields the forward Kolmogorov differential equations for the general birth and death process,

$$\begin{aligned}\frac{dP_{ij}(t)}{dt} &= \lambda_{j-1}P_{i,j-1}(t) - (\lambda_j + \mu_j)P_{ij}(t) + \mu_{j+1}P_{i,j+1}(t) \\ \frac{dP_{i0}(t)}{dt} &= -\lambda_0P_{i0}(t) + \mu_1P_{i1}(t), \quad \text{for } j = 0 \\ \frac{dP_{iN}(t)}{dt} &= \lambda_{N-1}P_{i,N-1}(t) - \mu_NP_{iN}(t), \quad \text{for } j = N.\end{aligned}\tag{4}$$

The forward Kolmogorov differential equations can be written in matrix notation,

$$\begin{aligned}\frac{d\mathbf{P}(t)}{dt} &= \mathbf{P}(t)\mathbf{Q} \\ \begin{bmatrix} dP_{i0}/dt \\ dP_{i1}/dt \\ dP_{i2}/dt \\ dP_{i3}/dt \\ \vdots \end{bmatrix} &= [P_{i0} \ P_{i1} \ P_{i2} \ P_{i3} \ \dots] \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \dots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & \dots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & \dots \\ 0 & 0 & \mu_3 & -(\lambda_3 + \mu_3) & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix},\end{aligned}$$

where the generator matrix \mathbf{Q} for the infinite state space

$$\mathbf{Q} = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \dots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & \dots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & \dots \\ 0 & 0 & \mu_3 & -(\lambda_3 + \mu_3) & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix},\tag{5}$$

and for the finite state space

$$\mathbf{Q} = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \mu_{N-1} & -(\lambda_{N-1} + \mu_{N-1}) & \lambda_{N-1} \\ 0 & 0 & 0 & \mu_N & -\mu_N \end{bmatrix}. \quad (6)$$

Similarly, to obtain the backward Kolmogorov differential equations, we start from the Chapman-Kolmogorov equations,

$$\begin{aligned} P_{ij}(\Delta t + t) &= \sum_{k=0}^{\infty} P_{ik}(\Delta t) P_{kj}(t) \\ &= \underbrace{P_{i,i-1}(\Delta t) P_{i-1,j}(t)}_{k=i-1} + \underbrace{P_{ii}(\Delta t) P_{ij}(t)}_{k=i} + \underbrace{P_{i,i+1}(\Delta t) P_{i+1,j}(t)}_{k=i+1} + \sum_k P_{ik}(\Delta t) P_{kj}(t), \end{aligned}$$

where the last summation is over all $k \neq i-1, i+1, i$.

For $k = i$,

$$P_{ii}(\Delta t) P_{ij}(t) = P_{ij}(t) \cdot [1 - (\lambda_i + \mu_i)\Delta t + o(\Delta t)],$$

for $k = i-1$,

$$P_{i,i-1}(\Delta t) P_{i-1,j}(t) = P_{i-1,j}(t) \cdot [\mu_i \Delta t + o(\Delta t)],$$

for $k = i+1$,

$$P_{i,i+1}(\Delta t) P_{i+1,j}(t) = P_{i+1,j}(t) \cdot [\lambda_i \Delta t + o(\Delta t)],$$

and for $k \neq j, j-1, j+1$, we have

$$\sum_k P_{ik}(\Delta t) P_{kj}(t) = P_{kj}(t) \cdot o(\Delta t).$$

Collecting all together,

$$P_{ij}(t + \Delta t) = [1 - (\lambda_i + \mu_i)\Delta t] P_{ij}(t) + \mu_i \Delta t P_{i-1,j}(t) + \lambda_i \Delta t P_{i+1,j}(t) + o(\Delta t),$$

or, rearranging,

$$P_{ij}(t + \Delta t) - P_{ij}(t) = \mu_i \Delta t P_{i-1,j}(t) - (\lambda_i + \mu_i) \Delta t P_{ij}(t) + \lambda_i \Delta t P_{i+1,j}(t) + o(\Delta t).$$

Dividing by Δt and taking the limit as $\Delta t \rightarrow 0$, we obtain the backward Kolmogorov differential equations

$$\begin{aligned} \frac{dP_{ij}(t)}{dt} &= \mu_i P_{i-1,j}(t) - (\lambda_i + \mu_i) P_{ij}(t) + \lambda_i P_{i+1,j}(t), \\ \frac{dP_{0j}(t)}{dt} &= -\lambda_0 P_{0j}(t) + \lambda_0 P_{1j}(t), & \text{for } i = 0 \\ \frac{dP_{Nj}(t)}{dt} &= \mu_N P_{N-1,j}(t) - \mu_N P_{Nj}(t), & \text{for } i = N. \end{aligned} \quad (7)$$

where $\mu_0 = 0$ and $\lambda_N = 0$, and $j \geq 0$, $i \geq 0$.

The backward Kolmogorov differential equations can be written in matrix notation,

$$\begin{aligned} \frac{d\mathbf{P}(t)}{dt} &= \mathbf{Q}\mathbf{P}(t) \\ \begin{bmatrix} dP_{0j}/dt \\ dP_{1j}/dt \\ dP_{2j}/dt \\ \vdots \\ dP_{Nj}/dt \end{bmatrix} &= \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & \cdots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \mu_N & -\mu_N \end{bmatrix} \begin{bmatrix} P_{0j} \\ P_{1j} \\ P_{2j} \\ \vdots \\ P_{Nj} \end{bmatrix}. \end{aligned}$$

2.2 Stationary Probability Distribution (The Limiting Behavior of Birth and Death Processes)

For a general birth and death process that has no absorbing states, the limits

$$\lim_{t \rightarrow \infty} P_{ij}(t) = \pi_j \geq 0 \quad (8)$$

exist and are independent of the initial state i . The limits form the limiting distribution (or limiting probability) of the process, or also known as a stationary probability distribution. For

a more detailed discussion on this topic, see subsection 1.7 of Lecture 4a (Continuous-Time Markov Chain Models). To determine if a limiting distribution exists and what its values are, we rewrite the forward Kolmogorov differential equations in Eq. (4),

$$\begin{aligned}\frac{dP_{i0}(t)}{dt} &= -\lambda_0 P_{i0}(t) + \mu_1 P_{i1}(t), \\ \frac{dP_{ij}(t)}{dt} &= \lambda_{j-1} P_{i,j-1}(t) - (\lambda_j + \mu_j) P_{ij}(t) + \mu_{j+1} P_{i,j+1}(t), \quad j \geq 1,\end{aligned}$$

with initial conditions $P_{ij}(0) = \delta_{ij}$. If we take the limit as $t \rightarrow \infty$ to these equations, the limit of the right hand sides exists according to Eq. (8). The limit on the left hand side, the derivatives $P'_{ij}(t)$, also exists. Since the probabilities are converging to a constant, the limit of these derivatives must be zero [9]. These equations can be solved recursively. For the first equation, or when $j = 0$:

$$\begin{aligned}\lim_{t \rightarrow \infty} \frac{dP_{i0}(t)}{dt} &= -\lambda_0 \lim_{t \rightarrow \infty} P_{i0}(t) + \mu_1 \lim_{t \rightarrow \infty} P_{i1}(t) \\ 0 &= -\lambda_0 \pi_0 + \mu_1 \pi_1,\end{aligned}$$

or

$$\pi_1 = \frac{\lambda_0}{\mu_1} \pi_0.$$

From the second equation,

$$\begin{aligned}\lim_{t \rightarrow \infty} \frac{dP_{ij}(t)}{dt} &= \lambda_{j-1} \lim_{t \rightarrow \infty} P_{i,j-1}(t) - (\lambda_j + \mu_j) \lim_{t \rightarrow \infty} P_{ij}(t) + \mu_{j+1} \lim_{t \rightarrow \infty} P_{i,j+1}(t) \\ 0 &= \lambda_{j-1} \pi_{j-1} - (\lambda_j + \mu_j) \pi_j + \mu_{j+1} \pi_{j+1},\end{aligned}$$

when $j = 1$:

$$\pi_2 = \frac{\lambda_1}{\mu_2} \pi_1 = \frac{\lambda_0 \lambda_1}{\mu_1 \mu_2} \pi_0,$$

when $j = 2$:

$$\pi_3 = \frac{\lambda_2}{\mu_3} \pi_2 = \frac{\lambda_0 \lambda_1 \lambda_2}{\mu_1 \mu_2 \mu_3} \pi_0,$$

and so on. Thus,

$$\pi_{i+1} = \frac{\lambda_i}{\mu_{i+1}} \pi_i. \tag{9}$$

Or, applying the induction hypothesis, the stationary probability distribution equals

$$\pi_i = \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i} \pi_0, \quad i = 1, 2, 3, \dots \quad (10)$$

Then

$$\begin{aligned} \mu_{i+1} \pi_{i+1} &= (\lambda_i + \mu_i) \pi_i - \lambda_{i-1} \pi_{i-1} \\ &= \left(\frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1} (\lambda_i + \mu_i)}{\mu_1 \mu_2 \cdots \mu_i} - \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_{i-1}} \right) \pi_0 \\ &= \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_{i-1}} \left(\frac{\lambda_i + \mu_i}{\mu_i} - 1 \right) \pi_0, \end{aligned}$$

or,

$$\pi_{i+1} = \frac{\lambda_0 \lambda_1 \cdots \lambda_i}{\mu_1 \mu_2 \cdots \mu_{i+1}} \pi_0. \quad (11)$$

Thus, if the state space is infinite, $\{0, 1, 2, \dots\}$, a unique positive stationary probability distribution π for a general birth and death process exists, with

$$\mu_i > 0 \quad \text{and} \quad \lambda_{i-1} > 0 \quad \text{for } i = 1, 2, 3, \dots$$

Since $\sum_{i=0}^{\infty} \pi_i = 1$, which we expand

$$\begin{aligned} \sum_{i=0}^{\infty} \pi_i &= 1 \\ \pi_0 + \sum_{i=1}^{\infty} \pi_i &= 1 \\ \pi_0 \left(1 + \sum_{i=1}^{\infty} \frac{\pi_i}{\pi_0} \right) &= 1, \end{aligned}$$

from which we solve for π_0 by invoking Eq. (10),

$$\pi_0 = \frac{1}{1 + \sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i}}. \quad (12)$$

If the state space is finite, $\{0, 1, 2, \dots, N\}$, then a unique positive stationary probability distribution π exists if and only if

$$\mu_i > 0 \quad \text{and} \quad \lambda_{i-1} > 0 \quad \text{for } i = 1, 2, \dots, N.$$

The stationary probability distribution is given by Eqs. (10) and (12), where the index and the summation extend from $i = 1, 2, \dots, N$.

2.3 Pure Birth

In a pure birth process, the chance of an event occurring at a given instant of time depends upon the number of events that have already occurred. An example of this phenomenon is the reproduction of living organisms, hence the name of the process. This process considers only the birth or arrival of new members, permitted with certain conditions, such as, sufficient food, no mortality, and no migration out of the system [9].

Let $X(t)$ represent the number of births in the time interval $(0, t]$. We define a pure birth process as a Markov process satisfying the following:

$$\begin{aligned} P_{i,i+j}(\Delta t) &= \text{Prob}\{X(t + \Delta t) - X(t) = j | X(t) = i\} \\ &= \begin{cases} \lambda_i \Delta t + o(\Delta t), & j = 1 \\ 1 - \lambda_i \Delta t + o(\Delta t), & j = 0 \\ 0, & j < 0. \end{cases} \end{aligned} \tag{13}$$

By analyzing the possibilities at time t just prior to time $t + \Delta t$, for Δt is sufficiently small, we derive a system of Kolmogorov's forward differential equations for a pure birth process. The probabilities $P_n(t) = \text{Prob}\{X(t) = n\}$ are solutions of the differential equations

$$\begin{aligned} \frac{dP_0(t)}{dt} &= -\lambda_0 P_0(t), \\ \frac{dP_n(t)}{dt} &= -\lambda_n P_n(t) + \lambda_{n-1} P_{n-1}(t), \end{aligned} \tag{14}$$

with initial conditions

$$P_0(0) = 1, \quad P_n(0) = 0, \quad n > 0.$$

The first equation of Eq. (14) can be solved immediately. With $P_0(0) = 1$, then

$$P_0(t) = e^{-\lambda_0 t}, \quad \text{for } t > 0.$$

For the second equation of Eq. (14), there are several ways to compute the solutions:

(1) To solve the differential equations recursively, introduce

$$R_n(t) = e^{\lambda_n t} P_n(t), \quad \text{for } n = 0, 1, 2, \dots$$

and rearrange the second equation of Eq. (14) to get

$$\lambda_{n-1} P_{n-1}(t) = \frac{dP_n(t)}{dt} + \lambda_n P_n(t),$$

which is used for the following. Differentiating R_n with respect to t yields

$$\begin{aligned} \frac{dR_n(t)}{dt} &= \lambda_n e^{\lambda_n t} P_n(t) + e^{\lambda_n t} \frac{dP_n(t)}{dt} \\ &= e^{\lambda_n t} \left[\lambda_n P_n(t) + \frac{P_n(t)}{dt} \right] \\ &= e^{\lambda_n t} \lambda_{n-1} P_{n-1}(t). \end{aligned}$$

Integrating both sides of these equations and using the initial condition $R_n(0) = 0$ gives

$$R_n(t) = \int_0^t e^{\lambda_n x} \lambda_{n-1} P_{n-1}(x) dx, \quad \text{for } n \geq 1.$$

Substituting $R_n(t)$ on the left hand side,

$$e^{\lambda_n t} P_n(t) = \int_0^t e^{\lambda_n x} \lambda_{n-1} P_{n-1}(x) dx,$$

or

$$\begin{aligned} P_n(t) &= e^{-\lambda_n t} \int_0^t e^{\lambda_n x} \lambda_{n-1} P_{n-1}(x) dx \\ &= \lambda_{n-1} e^{-\lambda_n t} \int_0^t e^{\lambda_n x} P_{n-1}(x) dx, \quad n = 1, 2, \dots \end{aligned} \tag{15}$$

It can be seen that all $P_k(t) \geq 0$, but there is still a possibility that [9]

$$\sum_{n=0}^{\infty} P_n(t) < 1.$$

To ensure that $\sum_{n=0}^{\infty} P_n(t) = 1$ for all t , we must restrict the λ_k according to the following [9]

$$\sum_{n=0}^{\infty} P_n(t) = 1 \quad \text{if and only if} \quad \sum_{n=0}^{\infty} \frac{1}{\lambda_n} = \infty.$$

The integral equation (15) is solved as follows

$$\begin{aligned}
 n = 1 : \quad P_1(t) &= \lambda_0 e^{-\lambda_1 t} \int_0^t e^{\lambda_1 x} P_0(x) dx \\
 &= \lambda_0 \left(\frac{e^{-\lambda_0 t}}{\lambda_1 - \lambda_0} + \frac{e^{-\lambda_1 t}}{\lambda_0 - \lambda_1} \right), \\
 n = 2 : \quad P_2(t) &= \lambda_1 e^{-\lambda_2 t} \int_0^t e^{\lambda_2 x} P_1(x) dx \\
 &= \lambda_0 \lambda_1 \left(\frac{e^{-\lambda_0 t}}{(\lambda_1 - \lambda_0)(\lambda_2 - \lambda_0)} + \frac{e^{-\lambda_1 t}}{(\lambda_0 - \lambda_1)(\lambda_2 - \lambda_1)} + \frac{e^{-\lambda_2 t}}{(\lambda_0 - \lambda_2)(\lambda_1 - \lambda_2)} \right).
 \end{aligned}$$

Continuing the computations, we obtain that for $n > 1$ and initial condition

$$P_n(0) = 0,$$

$$\begin{aligned}
 P_n(t) &= \text{Prob}\{X(t) = n \mid X(0) = 0\} \\
 &= \lambda_0 \cdots \lambda_{n-1} (B_{0,n} e^{-\lambda_0 t} + \cdots + B_{n,n} e^{-\lambda_n t}),
 \end{aligned} \tag{16}$$

where

$$\begin{aligned}
 B_{0,n} &= \frac{1}{(\lambda_1 - \lambda_0)(\lambda_2 - \lambda_0) \cdots (\lambda_n - \lambda_0)} \\
 B_{k,n} &= \frac{1}{(\lambda_0 - \lambda_k) \cdots (\lambda_{k-1} - \lambda_k)(\lambda_{k+1} - \lambda_k) \cdots (\lambda_n - \lambda_k)} \\
 B_{n,n} &= \frac{1}{(\lambda_0 - \lambda_n)(\lambda_1 - \lambda_n) \cdots (\lambda_{n-1} - \lambda_n)}.
 \end{aligned} \tag{17}$$

- (2) Alternatively, we can also explicitly determine $P_{ij}(t)$ in the case of a pure birth process, by letting X_k denote the time the process spends in state k before making a transition into state $k+1$ for $k \geq 1$. Suppose that the process is presently in state i , and let state $j > i$. Then X_i is the time it spends in state i before moving to state $i+1$, and X_{i+1} is the time it then spends in state $i+1$ before moving to state $i+2$, and so on. It follows that $\sum_{k=i}^{j-1} X_k$ is the time it takes until the process enters state j . If the process has not yet entered state j by time t , then its state at time t is smaller than j , and vice versa [7], that is

$$X(t) < j \quad \text{if and only if } X_i + \cdots + X_{j-1} > t.$$

Therefore, for $i < j$, we have for a pure birth process

$$P_{ij}(t) = \text{Prob}\{X(t) < j \mid X(0) = i\} = \text{Prob}\left\{\sum_{k=i}^{j-1} X_k > t\right\}.$$

Since the random variables are hyperexponentially distributed (see hyperexponential distribution and convolutions of distribution in Lecture 4a on properties of exponential distribution), we obtain the tail distribution function of $\sum_{k=i}^{j-1} X_k$, that [7]

$$\text{Prob}\{X(t) < j | X(0) = i\} = \sum_{k=i}^{j-1} e^{-\lambda_k t} \prod_{r \neq k, r=i}^{j-1} \frac{\lambda_r}{\lambda_r - \lambda_k}.$$

Replacing j by $j + 1$ in the preceding gives

$$\text{Prob}\{X(t) < j + 1 | X(0) = i\} = \sum_{k=i}^j e^{-\lambda_k t} \prod_{r \neq k, r=i}^j \frac{\lambda_r}{\lambda_r - \lambda_k}.$$

Since

$$\text{Prob}\{X(t) = j | X(0) = i\} = \text{Prob}\{X(t) < j + 1 | X(0) = i\} - \text{Prob}\{X(t) < j | X(0) = i\},$$

and since $P_{ii}(t) = \text{Prob}\{X_i > t\} = e^{-\lambda_i t}$, we have for a pure birth process having $\lambda_i \neq \lambda_j$ when $i \neq j$ [7]

$$\begin{aligned} P_{ij}(t) &= \sum_{k=i}^j e^{-\lambda_k t} \prod_{r \neq k, r=i}^j \frac{\lambda_r}{\lambda_r - \lambda_k} - \sum_{k=i}^{j-1} e^{-\lambda_k t} \prod_{r \neq k, r=i}^{j-1} \frac{\lambda_r}{\lambda_r - \lambda_k}, & i < j \\ P_{ii}(t) &= e^{-\lambda_i t}. \end{aligned} \tag{18}$$

2.4 The Yule Process

The *Yule process* is an example of a pure birth process that commonly arises in physics and biology. Let $X(t)$ represent the population size at time t and the initial population is $X(0) = N$, so that $P_i(0) = \delta_{iN}$. The population size can only increase in size because there are only births. The transition probabilities for sufficiently small Δt are

$$\begin{aligned} P_{i,i+j}(\Delta t) &= \text{Prob}\{X(t + \Delta t) - X(t) = j | X(t) = i\} \\ &= \begin{cases} 1 - i\lambda\Delta t + o(\Delta t) & j = 0 \\ i\lambda\Delta t + o(\Delta t) & j = 1 \\ o(\Delta t) & j \geq 2 \\ 0 & j < 0. \end{cases} \end{aligned} \tag{19}$$

Converting these transition probabilities into differential equations for $P_i(t) = \text{Prob}\{X(t) = i\}$ gives

$$\begin{aligned}
 P_i(t + \Delta t) &= \text{Prob}\{X(t + \Delta t) = i\} \\
 &= \text{Prob}\{X(t) = i, X(t + \Delta t) - X(t) = 0\} \\
 &\quad + \text{Prob}\{X(t) = i - 1, X(t + \Delta t) - X(t) = 1\} \\
 &\quad + \text{Prob}\{X(t) = i - 2, X(t + \Delta t) - X(t) \geq 2\} \\
 &= \text{Prob}\{X(t) = i\} \cdot \text{Prob}\{X(t + \Delta t) - X(t) = 0\} \\
 &\quad + \text{Prob}\{X(t) = i - 1\} \cdot \text{Prob}\{X(t + \Delta t) - X(t) = 1\} \\
 &\quad + \text{Prob}\{X(t) = i - 2\} \cdot \text{Prob}\{X(t + \Delta t) - X(t) \geq 2\} \\
 &= P_i(t) \cdot [1 - i\lambda\Delta t + o(\Delta t)] + P_{i-1}(t) \cdot [(i-1)\lambda\Delta t + o(\Delta t)] \\
 &\quad + P_{i-2}(t) \cdot o(\Delta t).
 \end{aligned}$$

Rearranging and equating terms with $o(\Delta t)$ yield

$$P_i(t + \Delta t) - P_i(t) = (i-1)\lambda\Delta t P_{i-1}(t) - i\lambda\Delta t P_i(t) + o(\Delta t).$$

Dividing both sides by Δt and taking the limit as $\Delta t \rightarrow 0$ lead to the system of forward Kolmogorov equations

$$\begin{aligned}
 \frac{dP_i(t)}{dt} &= (i-1)\lambda P_{i-1}(t) - i\lambda P_i(t), \quad i = N, N+1, N+2, \dots \\
 \frac{dP_i(t)}{dt} &= 0, \quad i = 0, 1, 2, \dots, N-1
 \end{aligned} \tag{20}$$

Its solution is given by

$$P_i(t) = e^{-\lambda t} (1 - e^{-\lambda t})^{i-1}, \quad i \geq 1. \tag{21}$$

Several ways of solving the differential-difference equation of the Yule process Eq. (20) are as follows:

(1) Using the solution of the recursive method of Eq. (16):

The general solution is analogous to Eq. (16), but for pure birth processes starting from $X(0) = 1$, [9]

$$P_i(t) = \lambda_1 \cdots \lambda_{i-1} (B_{1,i} e^{-\lambda_1 t} + \cdots + B_{i,i} e^{-\lambda_i t}), \quad i > 1. \tag{22}$$

When $\lambda_i = i\lambda$,

$$\begin{aligned}
 B_{1,i} &= \frac{1}{(\lambda_2 - \lambda_1)(\lambda_3 - \lambda_1)(\lambda_4 - \lambda_1) \cdots (\lambda_i - \lambda_1)} \\
 &= \frac{1}{(2\lambda - \lambda)(3\lambda - \lambda)(4\lambda - \lambda) \cdots (i\lambda - \lambda)} \\
 &= \frac{1}{\lambda(1)\lambda(2)\lambda(3) \cdots \lambda(i-1)} \\
 &= \frac{1}{\lambda^{i-1}(1)(2)(3) \cdots (i-1)} \\
 &= \frac{1}{\lambda^{i-1}(i-1)!},
 \end{aligned}$$

and

$$\begin{aligned}
 B_{2,i} &= \frac{1}{(\lambda_1 - \lambda_2)(\lambda_3 - \lambda_2) \cdots (\lambda_i - \lambda_2)} \\
 &= \frac{1}{\lambda^{i-1}(-1)(1)(2) \cdots (i-2)} \\
 &= \frac{-1}{\lambda^{i-1}(i-2)!},
 \end{aligned}$$

so, in general,

$$\begin{aligned}
 B_{k,i} &= \frac{1}{(\lambda_1 - \lambda_k) \cdots (\lambda_{k-1} - \lambda_k)(\lambda_{k+1} - \lambda_k) \cdots (\lambda_i - \lambda_k)} \\
 &= \frac{(-1)^{k-1}}{\lambda^{i-1}(k-1)!(i-k)!}.
 \end{aligned}$$

Thus the solution is

$$\begin{aligned}
 P_i(t) &= \lambda^{i-1}(i-1)! \left(B_{1,i} e^{-\lambda t} + B_{2,i} e^{-2\lambda t} + \cdots + B_{i,i} e^{-i\lambda t} \right) \\
 &= \sum_{k=1}^i \frac{(i-1)!}{(k-1)!(i-k)!} (-1)^{k-1} e^{-k\lambda t} \\
 &= e^{-\lambda t} \sum_{j=0}^{n-1} \frac{(i-1)!}{j!(i-1-j)!} (e^{-\lambda t})^j \\
 &= e^{-\lambda t} \left(1 - e^{-\lambda t} \right)^{i-1}, \quad i \geq 1. \tag{23}
 \end{aligned}$$

(2) Alternatively, using Eq. (18), for the case $i = 1$:

$$\begin{aligned}
P_{1j}(t) &= \text{Prob}\{X(t) \leq j | X(0) = 1\} \\
&= \sum_{k=1}^j e^{-k\lambda t} \prod_{r \neq k, r=1}^j \frac{r}{r-k} - \sum_{k=1}^{j-1} e^{-k\lambda t} \prod_{r \neq k, r=1}^{j-1} \frac{r}{r-k} \\
&= e^{-j\lambda t} \prod_{r=1}^{j-1} \frac{r}{r-j} + \sum_{k=1}^{j-1} e^{-k\lambda t} \left(\prod_{r \neq k, r=1}^j \frac{r}{r-k} - \prod_{r \neq k, r=1}^{j-1} \frac{r}{r-k} \right) \\
&= e^{-j\lambda t} (-1)^{j-1} + \sum_{k=1}^{j-1} e^{-k\lambda t} \binom{j}{j-k-1} \prod_{r \neq k, r=1}^{j-1} \frac{r}{r-k}.
\end{aligned}$$

Now,

$$\begin{aligned}
\frac{k}{j-k} \prod_{r \neq k, r=1}^{j-1} \frac{r}{r-k} &= \frac{(j-1)!}{(1-k)(2-k)\cdots(k-1-k)(j-k)!} \\
&= (-1)^{k-1} \binom{j-1}{k-1}.
\end{aligned}$$

Hence,

$$\begin{aligned}
P_{1j}(t) &= \sum_{k=1}^j \binom{j-1}{k-1} e^{-k\lambda t} (-1)^{k-1} \\
&= e^{-\lambda t} \sum_{i=0}^{j-1} \binom{j-1}{i} e^{-i\lambda t} (-1)^i \\
&= e^{-\lambda t} (1 - e^{-\lambda t})^{j-1}, \quad j \geq 1,
\end{aligned} \tag{24}$$

which is exactly the same as the solution in Eq. (23).

The solution in Eq. (24) states that starting with one individual, the population size at time t has a geometric distribution with mean $e^{-\lambda t}$. If the population starts with i individuals, then we can regard each of these individuals as starting their own independent Yule process, and so the population at time t will be the sum of i independent and identically distributed geometric random variables with parameter $e^{-\lambda t}$. But this means that the conditional distribution of $X(t)$, given that $X(0) = i$, is the same as the distribution of the number of times that a coin that lands heads on each flip with probability $e^{-\lambda t}$ must be flipped to amass a total of i heads. Hence, the population size at time t has a negative binomial distribution with parameters i and $e^{-\lambda t}$, thus

$$P_{ij}(t) = \binom{j-1}{i-1} e^{-i\lambda t} (1 - e^{-\lambda t})^{j-i}, \quad j \geq i \geq 1. \tag{25}$$

(3) Using the probability generating function (p.g.f.):

Observe that the state space for this process is $\{1, 2, 3, \dots\}$. All of the states are transient and the transition probabilities $P_n(t)$ all depend on time t or non-stationary probability distribution.

Recall that the p.g.f., denoted $\mathbb{P}(t)$, is given by

$$\mathbb{P}(z) = \sum_{i=0}^{\infty} P_i z^i.$$

For the sake of generality, here we alter the initial conditions where now $P_N(0) = 1$ and $P_i(0) = 0$. We derive the partial differential equation for the p.g.f. by multiplying the differential equations by z^i and sum over i ,

$$\begin{aligned} \frac{\partial \mathbb{P}(z, t)}{\partial t} &= \sum_{i=N+1}^{\infty} (N+1-i)\lambda P_{N+1-i} z^{N+1} - \sum_{i=N}^{\infty} N\lambda P_N z^N \\ &= \sum_{i=N+1}^{\infty} N\lambda P_N z^{N-1} z^2 - \sum_{i=N}^{\infty} N\lambda P_N z^{N-1} z \\ &= \lambda z^2 \sum_{i=N}^{\infty} i P_i z^{i-1} - \lambda z \sum_{i=N}^{\infty} i P_i z^{i-1} \\ &= \lambda z(z-1) \sum_{i=N}^{\infty} i P_i z^{i-1}. \end{aligned}$$

Recall that the first derivative of \mathbb{P} with respect to z is

$$\frac{\partial \mathbb{P}}{\partial z} = \sum_{i=1}^{\infty} i P_i z^{i-1},$$

then

$$\frac{\partial \mathbb{P}(z, t)}{\partial t} = \lambda z(z-1) \frac{\partial \mathbb{P}}{\partial z}.$$

The initial condition is $\mathbb{P}(z, 0) = z^N$.

The partial differential equation for the moment generating function (m.g.f.) can be derived by a change of variable. Let $z = e^\theta$, where $d\theta/dz = 1/e^\theta = 1/z$. There exists a close relationship between the probability generating function $\mathbb{P}_X(\cdot)$ and the moment generating function $\mathbb{M}_X(\cdot)$, that is

$$\mathbb{M}(\theta) = E[e^{\theta X}] = \mathbb{P}(e^\theta).$$

Then $\mathbb{P}(e^\theta, t) = \mathbb{M}(\theta, t)$. Following the chain rule,

$$\frac{\partial \mathbb{P}}{\partial z} = \frac{\partial \mathbb{M}}{\partial \theta} \frac{d\theta}{dz} = \frac{1}{z} \frac{\partial \mathbb{M}}{\partial \theta}.$$

The m.g.f. is a solution of the following partial differential equation

$$\frac{\partial \mathbb{M}}{\partial t} = \lambda(e^\theta - 1) \frac{\partial \mathbb{M}}{\partial \theta},$$

with initial condition $\mathbb{M}(\theta, 0) = e^{N\theta}$.

The method of characteristics is applied to find the solution of $\mathbb{M}(\theta, t)$. Rewrite the differential equation for $\mathbb{M}(\theta, t)$

$$\frac{\partial \mathbb{M}}{\partial t} + \lambda(1 - e^\theta) \frac{\partial \mathbb{M}}{\partial \theta} = 0.$$

The solution is found by introducing the characteristic equations

$$\frac{d\theta}{d\tau} = \lambda(1 - e^\theta), \quad \frac{dt}{d\tau} = 1, \quad \text{and} \quad \frac{dv}{d\tau} = 0,$$

with initial conditions

$$\theta(s, 0) = s, \quad t(s, 0) = 0, \quad \text{and} \quad v(s, 0) = e^{Ns}.$$

For the first characteristic equation, separation of variables and simplification lead to

$$\frac{d\theta}{1 - e^\theta} = \lambda d\tau \quad \text{and} \quad \frac{e^{-\theta} d\theta}{e^{-\theta} - 1} = \lambda d\tau.$$

By using a substitution $u = e^{-\theta} - 1$, from which $du = -e^{-\theta} d\theta$, we integrate both sides of

$$-\frac{du}{u} = \lambda d\tau \quad \text{or} \quad \frac{du}{u} = -\lambda d\tau$$

which results in a general solution

$$\ln(u) = -\lambda\tau + c_1 \quad \text{or} \quad \ln(e^{-\theta} - 1) = -\lambda\tau + c_1.$$

The general solution can also be written

$$e^{-\theta} - 1 = e^{-\lambda\tau+c_1} = C_0 e^{-\lambda\tau}.$$

Applying the initial condition $\theta(s, 0) = s$ solves for the constant $C_0 = e^{-s} - 1$. Thus the particular solution of the first characteristic equation is

$$e^{-\theta} - 1 = (e^{-s} - 1)e^{-\lambda\tau}.$$

The second characteristic equation can be immediately integrated

$$t = \tau + c_2,$$

and using the initial condition $t(s, 0) = 0$, the solution is simply

$$t = \tau.$$

And the solution of the third characteristic equation

$$v(s, \tau) = e^{Ns}.$$

Letting $v(s, \tau) = \mathbb{M}(s, \tau)$, we can write for the m.g.f.

$$\mathbb{M}(s, \tau) = e^{Ns}.$$

The solution \mathbb{M} must be expressed in terms of the original variables θ and t . We do this by expressing e^{-s} in terms of θ and t , that is

$$e^{-s} = 1 - e^{\lambda\tau}(1 - e^{-\theta}).$$

Since $e^{Ns} = (e^{-s})^{-N}$, the m.g.f. for the Yule process is

$$\mathbb{M}(\theta, t) = \left(1 - e^{\lambda\tau}(1 - e^{-\theta})\right)^{-N}.$$

The p.g.f. can be found directly from the m.g.f. by replacing $e^{-\theta} = z^{-1}$ since $e^\theta = z$,

$$\begin{aligned} \mathbb{P}(z, t) &= \left(1 - e^{\lambda\tau}(1 - z^{-1})\right)^{-N} \\ &= z^N e^{-\lambda N t} \left(z e^{-\lambda t} - (z - 1)\right)^{-N} \\ &= \frac{z^N e^{-\lambda N t}}{(1 - z(1 - e^{-\lambda t}))^N}. \end{aligned}$$

Letting $p = e^{-\lambda t}$ and $q = 1 - e^{-\lambda t}$, the p.g.f. is

$$\mathbb{P}(z, t) = \frac{(pz)^N}{(1 - qz)^N}. \quad (26)$$

Taking $N = 1$, Eq. (26) can be solved trivially, that is, with

$$\mathbb{P}(z, t) = \frac{pz}{1 - qz},$$

then

$$\begin{aligned}
\frac{\partial \mathbb{P}(z)}{\partial z} &= \left. \frac{p}{(1 - qz)^2} \right|_{z=0} = p \\
\frac{\partial^2 \mathbb{P}(z)}{\partial z^2} &= \left. \frac{2pq}{(1 - qz)^3} \right|_{z=0} = 2pq \\
\frac{\partial^3 \mathbb{P}(z)}{\partial z^3} &= \left. \frac{6pq^2}{(1 - qz)^4} \right|_{z=0} = 6pq^2 \\
\frac{\partial^4 \mathbb{P}(z)}{\partial z^4} &= \left. \frac{24pq^3}{(1 - qz)^5} \right|_{z=0} = 24pq^3 \\
&\vdots
\end{aligned}$$

and from Eq. (2),

$$P_i(t) = \frac{1}{1!}p + \frac{1}{2!}2!pq^{2-1} + \frac{1}{3!}3!pq^{3-1} + \frac{1}{4!}4!pq^{4-1},$$

or

$$P_i(t) = \binom{i-1}{i-N} p^N q^{i-N}, \quad \text{for } N = 1 \text{ and } i = 1, 2, 3, \dots$$

In general, for $N \geq 1$, the p.g.f. and m.g.f. for the Yule process correspond to a negative binomial distribution. From the binomial coefficient, i must be greater than or equal to N , or $i \geq N$. Then taking $i = i + N$ and replacing p by $e^{-\lambda t}$ and q by $1 - e^{-\lambda t}$, the probabilities $P_i(t)$ can also be written as,

$$P_{i+N}(t) = \binom{N+i-1}{i} e^{-\lambda N t} (1 - e^{-\lambda t})^i, \quad i = 0, 1, 2, \dots \quad (27)$$

By letting $k = i + N$, another way to write the probabilities is

$$P_k(t) = \binom{k-1}{k-N} e^{-\lambda N t} (1 - e^{-\lambda t})^{k-N}, \quad k = N, N+1, \dots$$

When $N = 1$ or assuming that the population starts with one individual, then $k = i + 1$ and Eq. (27) becomes

$$P_{i+1}(t) = e^{-\lambda t} (1 - e^{-\lambda t})^i \quad \text{or} \quad P_i(t) = e^{-\lambda t} (1 - e^{-\lambda t})^{i-1}. \quad (28)$$

The mean and variance for the Yule process are

$$m(t) = \frac{N}{p} = N e^{\lambda t} \quad \text{and} \quad \sigma^2(t) = \frac{Nq}{p^2} = N e^{2\lambda t} (1 - e^{-\lambda t}).$$

The probabilities with four different values of t using Eq. (27) and three sample paths of the Yule process are shown in Fig. 1. Python code for these figures are provided in the Python Implementation section 3.1.

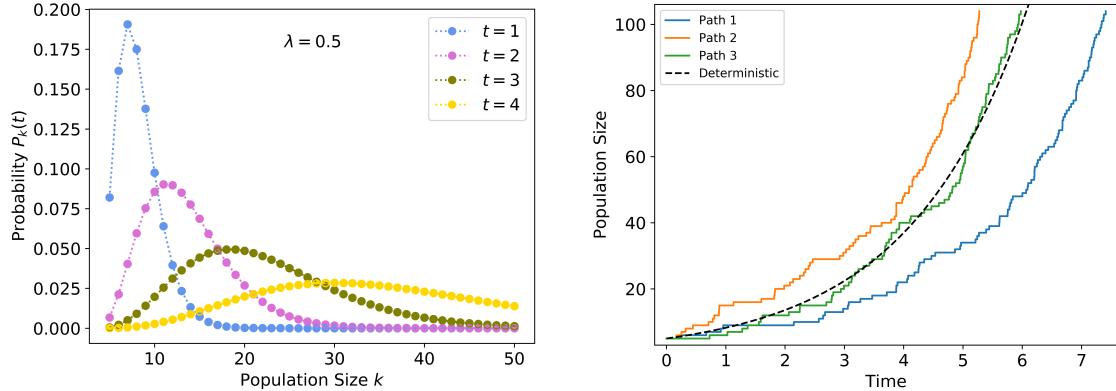


Figure 1: Plots of probability distributions $P_{i+N}(t)$ for the Yule process using Eq. (27) with $N = 5$ (left) and population size over time with three sample paths (right), both when $\lambda = 0.5$ and $X(0) = 5$. The mean and variance are $m(t) = 5e^{0.5t}$ and $\sigma^2(t) = 5e^t(1 - e^{-0.5t})$.

2.5 Pure Death

Complementing the increasing pure birth process is the pure death process that is decreasing. Since the only event in the pure death process is a death, the process decreases successively through states $N, N - 1, \dots, 2, 1$ and ultimately is absorbed in state 0 (extinction). The pure death process is specified by the death parameters $\mu_k > 0$ for $k = 1, 2, \dots, N$, where the sojourn/interevent time in state k is exponentially distributed with parameter μ_k .

The infinitesimal transition probabilities of the pure death process whose state space is $0, 1, 2, \dots, N$ are

$$\begin{aligned} P_{i,i+j}(\Delta t) &= \text{Prob}\{X(t + \Delta t) - X(t) = j \mid X(t) = i\} \\ &= \begin{cases} \mu_i \Delta t + o(\Delta t), & j = -1 \\ 1 - \mu_i \Delta t + o(\Delta t), & j = 0 \\ 0, & j > 0 \end{cases} \end{aligned} \quad (29)$$

for $i = 1, 2, \dots, N$ for the first and second equations and $i = 0, 1, \dots, N$ for the third equation. As we have previously discussed in the general birth and death process, it is common to assign $\mu_0 = 0$.

If the death parameters $\mu_1, \mu_2, \dots, \mu_N$ are distinct, that is $\mu_i \neq \mu_j$ if $i \neq j$, taking from Eq. (4), the forward Kolmogorov equation for the pure death process at $i = N$ is given by

$$\frac{dP_N(t)}{dt} = -\mu_N P_N(t).$$

If the initial condition is $X(0) = N$, or $P_N(0) = \text{Prob}\{X(0) = N\} = 1$, thus the solution, or the explicit transition probabilities are

$$P_N(t) = e^{-\mu_N t}.$$

For $i < N$,

$$\frac{dP_i(t)}{dt} = \mu_{i+1}P_{i+1}(t) - \mu_iP_i(t),$$

and the solutions are given by

$$\begin{aligned} P_i(t) &= \text{Prob}\{X(t) = i | X(0) = N\} \\ &= \mu_{i+1}\mu_{i+2}\cdots\mu_N (A_{i,i}e^{-\mu_it} + \cdots + A_{N,i}e^{-\mu_N t}) \end{aligned} \tag{30}$$

where

$$A_{k,i} = \frac{1}{(\mu_N - \mu_k)\cdots(\mu_{k+1} - \mu_k)(\mu_{k-1} - \mu_k)\cdots(\mu_i - \mu_k)}.$$

2.6 Linear Death

As the pure death process complements the pure birth process, the linear death process complements the Yule process. In the linear death process, or, also called the simple death process, the death rates are proportional to population size, $\mu_i = i\mu$. The infinitesimal transition probabilities of the linear death process are

$$\begin{aligned} P_{i,i+j}(\Delta t) &= \text{Prob}\{X(t + \Delta t) - X(t) = j | X(t) = i\} \\ &= \begin{cases} i\mu\Delta t + o(\Delta t), & j = -1 \\ 1 - i\mu\Delta t + o(\Delta t), & j = 0 \\ o(\Delta t), & j \leq -2 \\ 0, & j > 0 \end{cases} \end{aligned} \tag{31}$$

Since the process begins in state N , the state space is $\{0, 1, 2, \dots, N\}$.

Converting these transition probabilities into differential equations

for $P_i(t) = \text{Prob}\{X(t) = i\}$ gives

$$\begin{aligned}
 P_i(t + \Delta t) &= \text{Prob}\{X(t + \Delta t) = i\} \\
 &= \text{Prob}\{X(t) = i, X(t + \Delta t) - X(t) = 0\} \\
 &\quad + \text{Prob}\{X(t) = i + 1, X(t + \Delta t) - X(t) = -1\} \\
 &\quad + \text{Prob}\{X(t) = i + 2, X(t + \Delta t) - X(t) \leq -2\} \\
 &= \text{Prob}\{X(t) = i\} \cdot \text{Prob}\{X(t + \Delta t) - X(t) = 0\} \\
 &\quad + \text{Prob}\{X(t) = i + 1\} \cdot \text{Prob}\{X(t + \Delta t) - X(t) = -1\} \\
 &\quad + \text{Prob}\{X(t) = i + 2\} \cdot \text{Prob}\{X(t + \Delta t) - X(t) \leq -2\} \\
 &= P_i(t) \cdot [1 - i\mu\Delta t + o(\Delta t)] + P_{i+1}(t) \cdot [(i+1)\mu\Delta t + o(\Delta t)] \\
 &\quad + P_{i+2}(t) \cdot o(\Delta t).
 \end{aligned}$$

Rearranging and equating terms with $o(\Delta t)$ yield

$$P_i(t + \Delta t) - P_i(t) = (i + 1)\mu\Delta t P_{i+1}(t) - i\mu\Delta t P_i(t) + o(\Delta t).$$

Dividing both sides by Δt and taking the limit as $\Delta \rightarrow 0$ lead to the system of forward Kolmogorov equations

$$\begin{aligned}
 \frac{dP_i(t)}{dt} &= (i + 1)\mu P_{i+1}(t) - i\mu P_i(t) \\
 \frac{dP_N(t)}{dt} &= -N\mu P_N(t),
 \end{aligned}$$

for $i = 0, 1, 2, \dots, N - 1$ and with initial conditions $P_N(0) = 1$ and $P_i(0) = 0$ for $i \neq N$.

As in the Yule process, the probability distribution may be found by using the probability generating function and method of characteristics. First, we multiply the forward Kolmogorov equations

by z^i and sum over i ,

$$\begin{aligned}
 \frac{\partial \mathbb{P}(z, t)}{\partial t} &= \sum_{i=N-1}^{\infty} (i+1)\mu P_{i+1}(t)z^i - \sum_{i=N}^{\infty} i\mu P_i(t)z^i \\
 &= \sum_{i=N-1}^{\infty} (N-1+1)\mu P_{N-1+i}(t)z^{N-1} - \sum_{i=N}^{\infty} N\mu P_N(t)z^N \\
 &= \sum_{i=N}^{\infty} N\mu P_N(t)z^{N-1} - \sum_{i=N}^{\infty} N\mu P_N(t)z^{N-1}z \\
 &= \mu \sum_{i=N}^{\infty} iP_i(t)z^{i-1} - \mu z \sum_{i=N}^{\infty} iP_i(t)z^{i-1} \\
 &= \mu(1-z) \sum_{i=N}^{\infty} iP_i(t)z^{i-1}
 \end{aligned}$$

to get the partial differential equation for the p.g.f.,

$$\frac{\partial \mathbb{P}(z, t)}{\partial t} = \mu(1-z) \frac{\partial \mathbb{P}}{\partial z}, \quad \mathbb{P}(z, 0) = z^N.$$

Substituting $z = e^\theta$, the m.g.f. is

$$\frac{\partial \mathbb{M}}{\partial t} = \mu(e^{-\theta} - 1) \frac{\partial \mathbb{M}}{\partial \theta}, \quad \mathbb{M}(\theta, 0) = e^{N\theta}.$$

After applying the method of characteristics, we obtain the solutions

$$\mathbb{P}(z, t) = (1 - e^{-\mu t}(1 - z))^N = (1 - e^{-\mu t} + e^{-\mu t}z)^N$$

and

$$\mathbb{M}(\theta, t) = (1 - e^{-\mu t}(1 - e^\theta))^N.$$

By letting $p = e^{-\mu t}$ and $q = 1 - p = 1 - e^{-\mu t}$, the p.g.f. is

$$\mathbb{P}(z, t) = (q + pz)^N,$$

which corresponds to a binomial distribution $b(N, p)$. Since

$$P_i(t) = \frac{1}{i!} \left. \frac{\partial^i \mathbb{P}}{\partial z^i} \right|_{z=0},$$

from which we show that,

$$\begin{aligned}
 k = 0 &\rightarrow P_0(t) = q^N \\
 k = 1 &\rightarrow P_1(t) = \frac{1}{1!} \left. \frac{\partial \mathbb{P}}{\partial z} \right|_{z=0} = \frac{1}{1!} N q^{N-1} p \\
 k = 2 &\rightarrow P_2(t) = \frac{1}{2!} \left. \frac{\partial^2 \mathbb{P}}{\partial z^2} \right|_{z=0} = \frac{1}{2!} N(N-1) q^{N-2} p^2 \\
 k = 3 &\rightarrow P_3(t) = \frac{1}{3!} \left. \frac{\partial^3 \mathbb{P}}{\partial z^3} \right|_{z=0} = \frac{1}{3!} N(N-1)(N-2) q^{N-3} p^3 \\
 &\vdots
 \end{aligned}$$

thus, the probabilities are

$$P_i(t) = \frac{N!}{i!(N-i)!} p^i q^{N-i} = \binom{N}{i} e^{-\mu t} (1 - e^{-\mu t})^{N-i}, \quad (32)$$

for $i = 0, 1, 2, \dots, N$. The mean is obtained by first differentiating the p.g.f. with respect to z , that is

$$\frac{d\mathbb{P}}{dz} = N(q + pz)^{N-1} p,$$

and the mean, now denoted by $m(t)$, is

$$m(t) = \mathbb{P}'(1) = Np,$$

or, by $p = e^{-\mu t}$

$$m(t) = Ne^{-\mu t},$$

which corresponds to exponential decay. Whereas, the variance is

$$\sigma^2(t) = \mathbb{P}''(1) + \mathbb{P}'(1) - (\mathbb{P}'(1))^2 = Ne^{-\mu t} (1 - e^{-\mu t}).$$

Example 4b.1: Linear Death Process

Taken from Example 5.13 in [1]:

In a linear death process, the only event is a death, that is, state $i \rightarrow i - 1$. For Δt sufficiently

small, the transition probabilities are

$$\begin{aligned}
P_{i,i+j}(\Delta t) &= \text{Prob}\{X(t + \Delta t) - X(t) = j | X(t) = i\} \\
&= \begin{cases} di\Delta t + o(\Delta t) & j = -1 \\ 1 - di\Delta t + o(\Delta t) & j = 0 \\ o(\Delta t) & j \leq -2 \\ 0 & j > 0. \end{cases}
\end{aligned}$$

Given $X(W_i) = n$, then $\alpha(n) = dn$. Therefore, the interevent time T_i is

$$T_i = -\frac{\ln(U)}{dn}.$$

The next event is a death $n \rightarrow n - 1$, $X(W_{i+1}) = n - 1$. The deterministic analogue of this simple death process is the differential equation $dn/dt = -dn$, with initial condition $n(0) = N$, and the solution

$$n(t) = N e^{-dt}. \quad \blacksquare$$

The probability distributions and population size over time with three sample paths of the linear death process are shown in Fig. 2. The Python code to produce these results are given in the Python Implementation section 3.2. The code are also available on the course website.

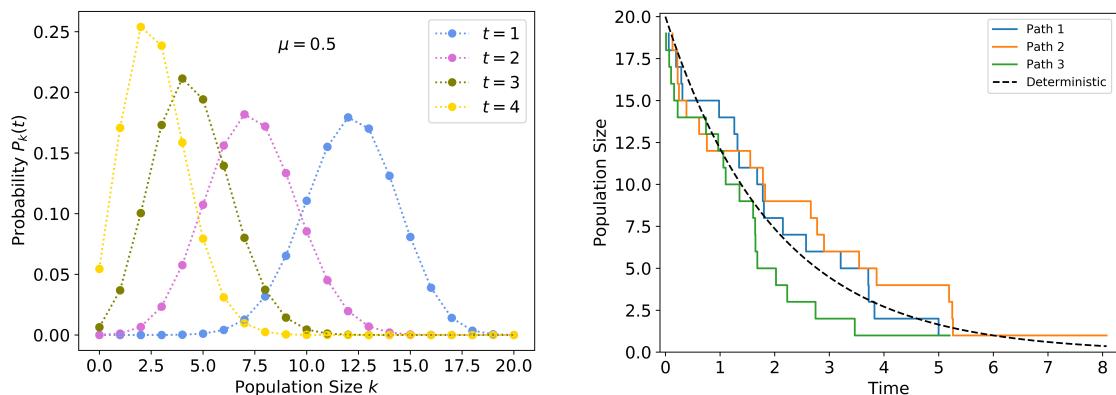


Figure 2: Plots of probability distributions $P_k(t)$ for the linear death process using Eq. (32) (left) and population size over time with three sample paths (right), both when $\mu = 0.5$ and $X(0) = 20$. The mean and variance are $m(t) = 20e^{-0.5t}$ and $\sigma^2(t) = 20e^{-0.5t}(1 - e^{-0.5t})$.

2.7 Simple Birth and Death

In a simple birth and death process, $X(t)$ is allowed to increase and decrease. If at time t the process is in state n , after a random sojourn time, it may move to either of the neighboring states

$n + 1$ or $n - 1$. The birth and death process is irreducible if all the states communicate, and this happens if and only if all the λ_n ($n \geq 0$) and all the μ_n ($n \geq 1$) are positive.

If we let $X(t)$ denote the state of the chain at time t , then

$$\begin{aligned} P_{i,i+j}(\Delta t) &= \text{Prob}\{X(t + \Delta t) - X(t) = j | X(t) = i\} \\ &= \begin{cases} 1 - i(\mu + \lambda)\Delta t + o(\Delta t) & j = 0 \\ i\lambda\Delta t + o(\Delta t) & j = 1 \\ i\mu\Delta t + o(\Delta t) & j = -1 \\ o(\Delta t) & j \neq -1, 0, 1. \end{cases} \end{aligned} \quad (33)$$

Simple birth and death processes are also known as birth and death processes with absorbing states. For these processes, the zero state is an absorbing state, where when the population size becomes zero, it remains zero thereafter.

As for the formulations in the Poisson process, we can convert the transition probabilities in Eq. (35) into differential equations for $P_i(t) = \text{Prob}\{X(t) = i\}$, hence

$$\begin{aligned} P_i(t + \Delta t) &= \text{Prob}\{X(t + \Delta t) = i\} \\ &= \text{Prob}\{X(t) = i, X(t + \Delta t) - X(t) = 0\} \\ &\quad + \text{Prob}\{X(t) = i - 1, X(t + \Delta t) - X(t) = 1\} \\ &\quad + \text{Prob}\{X(t) = i + 1, X(t + \Delta t) - X(t) = -1\} \\ &= P_i(t) \cdot \text{Prob}\{X(t + \Delta t) = X(t)\} \\ &\quad + \text{Prob}\{X(t) = i - 1\} \cdot \text{Prob}\{X(t + \Delta t) - X(t) = 1\} \\ &\quad + \text{Prob}\{X(t) = i + 1\} \cdot \text{Prob}\{X(t + \Delta t) - X(t) = -1\} \\ &= P_i(t) \cdot [1 - i(\mu + \lambda)\Delta t + o(\Delta t)] \\ &\quad + P_{i-1}(t) \cdot [(i - 1)\lambda\Delta t + o(\Delta t)] + P_{i+1}(t) \cdot [(i + 1)\mu\Delta t + o(\Delta t)]. \end{aligned}$$

Rearranging and equating terms with $o(\Delta t)$ give us

$$P_i(t + \Delta t) - P_i(t) = (i + 1)\mu\Delta t P_{i+1}(t) + (i - 1)\lambda\Delta t P_{i-1}(t) - i(\mu + \lambda)\Delta t P_i(t) + o(\Delta t).$$

Dividing both sides by Δt and taking the limit as $\Delta t \rightarrow 0$ lead to the following system of

differential-difference equations

$$\begin{aligned}\frac{dP_i(t)}{dt} &= (i-1)\lambda P_{i-1}(t) + (i+1)\mu P_{i+1}(t) - i(\mu + \lambda)P_i(t), \\ \frac{dP_0(t)}{dt} &= \mu P_1(t)\end{aligned}\tag{34}$$

for $i = 1, 2, \dots$ with initial conditions $P_i(0) = \delta_{iN}$. Multiplying the first equation by z^i and summing over i we obtain

$$\begin{aligned}\frac{\partial \mathbb{P}(z, t)}{\partial t} &= \sum_{i=N+1}^{\infty} (N+1-i)\lambda P_{N+1-i}(t)z^{N+1} + \sum_{i=N-1}^{\infty} (N-1+i)\mu P_{N-1+i}(t)z^{N-1} \\ &\quad - \sum_{i=N}^{\infty} N(\mu + \lambda)P_N(t)z^N \\ &= \sum_{i=N+1}^{\infty} N\lambda P_N(t)z^{N-1}z^2 + \sum_{i=N-1}^{\infty} N\mu P_N(t)z^{N-1} - \sum_{i=N}^{\infty} N(\mu + \lambda)P_N(t)z^{N-1}z \\ &= \lambda z^2 \sum_{i=N}^{\infty} iP_i(t)z^{i-1} + \mu \sum_{i=N}^{\infty} iP_i(t)z^{i-1} - z(\mu + \lambda) \sum_{i=N}^{\infty} iP_i(t)z^{i-1} \\ &= [\lambda z^2 - (\mu + \lambda)z + \mu] \sum_{i=N}^{\infty} iP_i(t)z^{i-1}.\end{aligned}$$

The p.g.f. is a solution of

$$\frac{\partial \mathbb{P}(z, t)}{\partial t} = [\mu(1-z) + \lambda z(z-1)] \frac{\partial \mathbb{P}}{\partial z}, \quad \mathbb{P}(z, 0) = z^N$$

and the m.g.f. is a solution is

$$\frac{\partial \mathbb{M}}{\partial t} = [\mu(e^{-\theta} - 1) + \lambda(e^\theta - 1)] \frac{\partial \mathbb{M}}{\partial \theta}, \quad \mathbb{M}(\theta, 0) = e^{\theta N}.$$

The characteristic equations are

$$\begin{aligned}\frac{dt}{d\tau} &= 1, \\ \frac{d\theta}{d\tau} &= -[\mu(e^{-\theta} - 1) + \lambda(e^\theta - 1)], \\ \frac{d\mathbb{M}}{d\tau} &= 0,\end{aligned}$$

after immediately letting $v(s, \tau) = \mathbb{M}(s, \tau)$. The initial conditions are

$$t(s, 0) = 0, \quad \theta(s, 0) = s, \quad \text{and} \quad \mathbb{M}(s, 0) = e^{Ns}.$$

The solutions satisfy

$$t = \tau \quad \text{and} \quad \mathbb{M}(s, \tau) = e^{Ns}.$$

We solve for τ in the following ways

- If $\mu = \lambda$:

By a change of variable $x = e^\theta$, the second characteristic equation becomes

$$\begin{aligned} \frac{d\theta}{\lambda(e^{-\theta} - 1) + \lambda(e^\theta - 1)} &= -d\tau \\ \frac{d\theta}{\lambda(e^\theta + e^{-\theta}) - 2\lambda} &= -d\tau \\ \frac{dx}{x[\lambda(x + x^{-1}) - 2\lambda]} &= -d\tau \\ \frac{dx}{\lambda(x^2 - 2x + 1)} &= -d\tau \\ \frac{dx}{\lambda(x - 1)^2} &= -d\tau \end{aligned}$$

Using a substitution $u = x - 1$, we integrate both sides

$$\begin{aligned} \frac{1}{\lambda} \int \frac{du}{u^2} &= - \int d\tau \\ -\frac{1}{\lambda u} &= -\tau + c_1, \end{aligned}$$

or

$$\tau = \frac{1}{\lambda u} + c_1 = \frac{1}{\lambda(e^\theta - 1)} + c_1,$$

where c_1 is the arbitrary constant of integration. Applying the initial condition $\theta(s, 0) = s$, we obtain

$$0 = \frac{1}{\lambda(e^s - 1)} + c_1,$$

or

$$c_1 = -\frac{1}{\lambda(e^s - 1)}.$$

Hence,

$$\tau = \frac{1}{\lambda(e^\theta - 1)} - \frac{1}{\lambda(e^s - 1)}.$$

- If $\mu \neq \lambda$:

Similar way as the above,

$$\begin{aligned} \frac{d\theta}{\mu(e^{-\theta}-1) + \lambda(e^\theta-1)} &= -d\tau \\ \frac{dx}{x[\mu(x^{-1}-1) + \lambda(x-1)]} &= -d\tau \\ \frac{dx}{\mu(1-x) + \lambda(x^2-x)} &= -d\tau \\ \frac{dx}{\lambda x^2 - (\lambda+\mu)x + \mu} &= -d\tau. \end{aligned}$$

The left hand side equation can be resolved into partial fractions as follows

$$\begin{aligned} \frac{1}{\lambda x^2 - (\lambda+\mu)x + \mu} &\equiv \frac{1}{(\lambda x - \mu)(x - 1)} \\ &\equiv \frac{A}{\lambda x - \mu} + \frac{B}{x - 1}, \end{aligned}$$

where A and B are constants to be determined by algebraic addition, *i.e.*,

$$\frac{1}{(\lambda x - \mu)(x - 1)} \equiv \frac{A(x - 1) + B(\lambda x - \mu)}{(\lambda x - \mu)(x - 1)}.$$

Since the denominators are the same on each side of the identity, then the numerators are equal to each other. Thus,

$$1 \equiv A(x - 1) + B(\lambda x - \mu).$$

The constants A and B are determined by choosing the values of x to make the term in A or B equal to zero. That is, when $x = 1$, then

$$\begin{aligned} 1 &\equiv A(0) + B(\lambda - \mu) \\ B &= \frac{1}{\lambda - \mu}, \end{aligned}$$

and when $x = \mu/\lambda$, then

$$\begin{aligned} 1 &\equiv A\left(\frac{\mu}{\lambda} - 1\right) + B(0) \\ A &= \frac{\lambda}{\mu - \lambda}. \end{aligned}$$

The characteristic equation becomes

$$\left(\frac{\lambda}{\mu - \lambda} \cdot \frac{1}{\lambda x - \mu} + \frac{1}{\lambda - \mu} \cdot \frac{1}{x - 1} \right) dx = -d\tau.$$

Integrating both sides and using substitutions

$$\begin{aligned}\frac{\lambda}{\mu - \lambda} \int \frac{dx}{\lambda x - \mu} + \frac{1}{\lambda - \mu} \int \frac{dx}{x - 1} &= - \int d\tau \\ \frac{1}{\mu - \lambda} \ln u + \frac{1}{\lambda - \mu} \ln v &= -\tau + c_2 \\ -\frac{1}{\mu - \lambda} \ln \left(\frac{v}{u} \right) &= -\tau + c_2\end{aligned}$$

where $u = \lambda x - \mu$ and $v = x - 1$. Thus,

$$\tau = \frac{1}{\mu - \lambda} \ln \left(\frac{e^\theta - 1}{\lambda e^\theta - \mu} \right) + c_2.$$

Applying the initial condition $\theta(s, 0) = s$ gives

$$c_2 = -\frac{1}{\mu - \lambda} \ln \left(\frac{e^s - 1}{\lambda e^s - \mu} \right),$$

and the particular solution of the characteristic equation is

$$\begin{aligned}\tau &= \frac{1}{\mu - \lambda} \ln \left(\frac{e^\theta - 1}{\lambda e^\theta - \mu} \right) - \frac{1}{\mu - \lambda} \ln \left(\frac{e^s - 1}{\lambda e^s - \mu} \right) \\ &= \frac{1}{\mu - \lambda} \left[\ln \left(\frac{(e^\theta - 1)(\lambda e^s - \mu)}{(\lambda e^\theta - \mu)(e^s - 1)} \right) \right].\end{aligned}$$

Because $\mathbb{M}(s, \tau) = (e^s)^N$, these relations are solved for e^s ,

$$e^s = \begin{cases} \frac{e^{\tau(\mu-\lambda)}(\lambda e^\theta - \mu) - \mu(e^\theta - 1)}{e^{\tau(\mu-\lambda)}(\lambda e^\theta - \mu) - \lambda(e^\theta - 1)}, & \text{if } \mu \neq \lambda \\ \frac{1 - (\lambda\tau - 1)(e^\theta - 1)}{1 - \lambda\tau(e^\theta - 1)}, & \text{if } \mu = \lambda. \end{cases}$$

The m.g.f. now can be expressed in terms of θ and t , that is

$$\mathbb{M}(\theta, t) = \begin{cases} \left(\frac{e^{t(\mu-\lambda)}(\lambda e^\theta - \mu) - \mu(e^\theta - 1)}{e^{t(\mu-\lambda)}(\lambda e^\theta - \mu) - \lambda(e^\theta - 1)} \right)^N, & \text{if } \mu \neq \lambda \\ \left(\frac{1 - (\lambda t - 1)(e^\theta - 1)}{1 - \lambda t(e^\theta - 1)} \right)^N, & \text{if } \mu = \lambda. \end{cases}$$

Making the change of variable $\theta = \ln(z)$, we obtain the p.g.f.,

$$\mathbb{P}(z, t) = \begin{cases} \left(\frac{e^{t(\mu-\lambda)}(\lambda z - \mu) - \mu(z - 1)}{e^{t(\mu-\lambda)}(\lambda z - \mu) - \lambda(z - 1)} \right)^N, & \text{if } \mu \neq \lambda \\ \left(\frac{1 - (\lambda t - 1)(z - 1)}{1 - \lambda t(z - 1)} \right)^N, & \text{if } \mu = \lambda. \end{cases}$$

Obtaining the probabilities $P_i(t)$ is not as straightforward as it was for the Yule and linear death processes. We follow calculation results in [1]. The first term in the expansion of $\mathbb{P}(z, t)$ is $P_0(t) = \mathbb{P}(0, t)$:

$$P_0(t) = \begin{cases} \left(\frac{\mu - \mu e^{(\mu-\lambda)t}}{\lambda - \mu e^{(\mu-\lambda)t}} \right)^N, & \text{if } \mu \neq \lambda \\ \left(\frac{\lambda t}{1 + \lambda t} \right)^N, & \text{if } \mu = \lambda. \end{cases}$$

The probability of extinction, $P_0(t)$, has a simple expression when $t \rightarrow \infty$,

$$P_0(\infty) = \lim_{t \rightarrow \infty} P_0(t) = \begin{cases} 1, & \text{if } \lambda \leq \mu \\ \left(\frac{\mu}{\lambda} \right)^N & \text{if } \lambda > \mu. \end{cases}$$

If the probability of death is greater than or equal to the probability of birth, or $\lambda \leq \mu$, then in the long run as $t \rightarrow \infty$ the probability of losing all of the initial population N approaches one. But if the probability of birth is greater than the probability of death, or $\lambda > \mu$, then as $t \rightarrow \infty$ the probability of losing all of the initial population is $(\mu/\lambda)^N$. The result when $\lambda > \mu$ as $t \rightarrow \infty$ is reminiscent of a semi-infinite random walk with an absorbing barrier at $x = 0$, or the gambler's ruin problem, where the probability of losing a game is μ and the probability of winning a game is λ [1].

The mean $m(t)$ and variance $\sigma^2(t)$ of the simple birth and death process are:

- For $\mu \neq \lambda$

$$m(t) = N e^{(\lambda-\mu)t} \quad \text{and} \quad \sigma^2(t) = N \frac{(\lambda + \mu)}{(\lambda - \mu)} e^{(\lambda-\mu)t} (e^{(\lambda-\mu)t} - 1),$$

where the mean corresponds to exponential growth when $\lambda > \mu$ and exponential decay when $\lambda < \mu$.

- For $\mu = \lambda$

$$m(t) = N \quad \text{and} \quad \sigma^2(t) = 2N\lambda t.$$

The population over time plots of the simple birth and death process are shown in Fig. 3. The Python code to produce these results are given in the Python Implementation section 3.3. The code are also available on the course website.

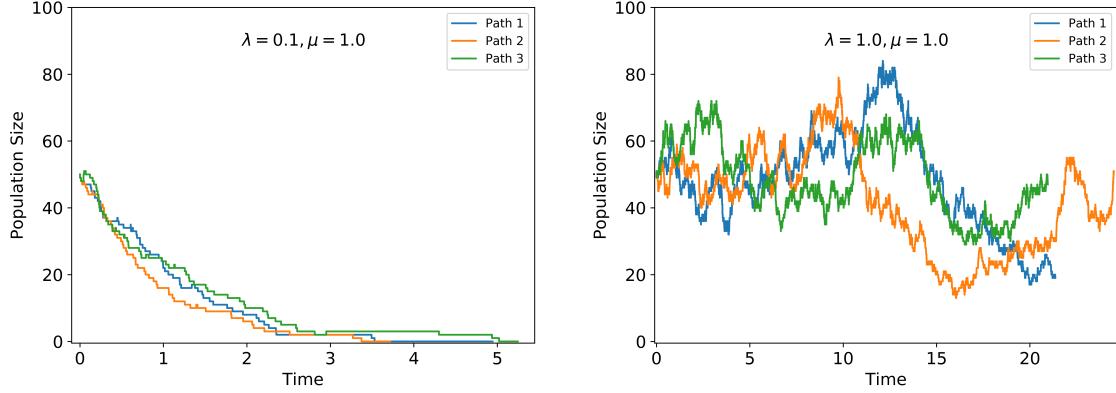


Figure 3: Plots of three sample paths for the simple birth and death process when $\mu = 1.0$, $X(0) = 50$, and $\lambda = 0.1$ (left) and $\lambda = 1.0$ (right). In the case $\lambda \ll \mu$ (left figure), the population size becomes zero and it remains in zero forever.

2.8 Simple Birth and Death Process with Immigration

A birth and death process is called a linear growth process if $\lambda_i = i\lambda + \beta$ and $\mu_i = i\mu$ with $\lambda > 0$, $\mu > 0$, and $\beta > 0$. If the state i describes the current population size, then the average instantaneous rate of growth is $i\lambda + \beta$, where the factor $i\lambda$ represents the natural growth of the population owing to its current size and the factor β may be interpreted as the infinitesimal rate of increase of the population due to an external source, such as immigration. Note that the immigration rate does not depend linearly on i , which looks like the Poisson process. Similarly, the probability of the state of the process decreasing by one after the elapse of a small duration of time Δt is $i\mu\Delta t + o(\Delta t)$, where the factor $i\mu$ gives the mean infinitesimal death rate of the present population.

For the general case of a simple birth and death process with immigration and emigration where the rates are $\lambda_i = i\lambda + \beta$ and $\mu_i = i\mu + \rho$, state transitions are a combination of:

- birth: $i \rightarrow i + 1$ at rate $i\lambda$
- death: $i \rightarrow i - 1$ at rate $i\mu$
- immigration: $i \rightarrow i + 1$ at rate β
- emigration: $i \rightarrow i - 1$ at rate ρ

The infinitesimal transition probabilities for the simple birth and death process with immigrations are

$$\begin{aligned}
 P_{i,i+j}(\Delta t) &= \text{Prob}\{X(t + \Delta t) - X(t) = j \mid X(t) = i\} \\
 &= \begin{cases} 1 - [\beta + i(\mu + \lambda)]\Delta t + o(\Delta t) & j = 0 \\ (\beta + i\lambda)\Delta t + o(\Delta t) & j = 1 \\ i\mu\Delta t + o(\Delta t) & j = -1 \\ o(\Delta t) & j \neq -1, 0, 1. \end{cases} \tag{35}
 \end{aligned}$$

Because of the immigration term, the value of $\lambda_0 = \beta > 0$.

The forward Kolmogorov differential equations for the process are

- For $i = 0$, where $\lambda_0 = \beta$ and $\mu_1 = \mu$,

$$\frac{dP_0(t)}{dt} = -\beta P_0(t) + \mu P_1(t)$$

- For $i \geq 1$, where $\lambda_{i-1} = (i-1)\lambda + \beta$, $\lambda_i = i\lambda + \beta$, $\mu_{i+1} = (i+1)\mu$, and $\mu_i = i\mu$,

$$\frac{dP_i(t)}{dt} = [\lambda(i-1) + \beta]P_{i-1}(t) - [(\lambda + \mu)i + \beta]P_i(t) + \mu(i+1)P_{i+1}(t).$$

Letting $X(0) = N$, for the second equation with $i \neq 1$, we apply the generating function technique by multiplying the equation by z^i and summing over i ,

$$\begin{aligned}
 \frac{\partial \mathbb{P}}{\partial t} &= \sum_{i=N+1}^{\infty} [\lambda(i-1) + \beta]P_{i-1}(t)z^i - \sum_{i=N}^{\infty} [i\lambda + i\mu + \beta]P_i(t)z^i + \sum_{i=N-1}^{\infty} \mu(i+1)P_{i+1}(t)z^i \\
 &= \lambda z^2 \sum_{i=N}^{\infty} iP_i(t)z^{i-1} + \beta z \sum_{i=N}^{\infty} P_i(t)z^i - (\lambda + \mu)z \sum_{i=N}^{\infty} iP_i(t)z^{i-1} - \beta \sum_{i=N}^{\infty} P_i(t)z^i \\
 &\quad + \mu \sum_{i=N}^{\infty} iP_i(t)z^{i-1} \\
 &= [\lambda z(z-1) + \mu(1-z)] \sum_{i=N}^{\infty} iP_i(t)z^{i-1} + \beta(z-1) \sum_{i=N}^{\infty} P_i(t)z^i,
 \end{aligned}$$

we obtain the p.g.f.,

$$\frac{\partial \mathbb{P}}{\partial t} = [\lambda z(z-1) + \mu(1-z)] \frac{\partial \mathbb{P}}{\partial z} + \beta(z-1)\mathbb{P},$$

and subsequently, the m.g.f. [1],

$$\frac{\partial \mathbb{M}}{\partial t} = [\lambda(e^\theta - 1) + \mu(e^{-\theta} - 1)] \frac{\partial \mathbb{M}}{\partial \theta} + \beta(e^\theta - 1)\mathbb{M},$$

with initial condition $\mathbb{M}(\theta, 0) = e^{N\theta}$. The solution of the m.g.f [1] is

$$\mathbb{M}(\theta, t) = \frac{(\lambda - \mu)^{\beta/\lambda} [\mu(\rho - 1) - e^\theta(\mu\rho - \lambda)]^N}{[(\lambda\rho - \mu) - \lambda e^\theta(\rho - 1)]^{N+\beta/\lambda}},$$

and the p.g.f.,

$$\mathbb{P}(z, t) = \frac{(\lambda - \mu)^{\beta/\lambda} [\mu(\rho - 1) - z(\mu\rho - \lambda)]^N}{[\lambda\rho - \mu - \lambda z(\rho - 1)]^{N+\beta/\lambda}},$$

where $\rho = e^{(\lambda-\mu)t}$, $X(0) = N$, and $\lambda \neq \mu$.

The mean can be calculated by differentiating $\mathbb{M}(\theta, t)$ with respect to θ and evaluating at $\theta = 0$,

$$m(t) = \begin{cases} \frac{\rho(N\mu - N\lambda - \beta) + \beta}{\mu - \lambda}, & \text{if } \lambda \neq \mu \\ \beta t + N, & \text{if } \lambda = \mu. \end{cases} \quad (36)$$

For the cases $\lambda > \mu$ and $\lambda = \mu$, the process is explosive, where when $\lambda > \mu$, the mean increases exponentially in time and when $\lambda = \mu$, the mean increases linearly in time. However, the process is nonexplosive in the case $\lambda < \mu$, where the mean approaches a constant,

$$m(\infty) = \frac{\beta}{\mu - \lambda}, \quad \lambda < \mu.$$

Also, for the case $\lambda < \mu$, the process is irreducible and has a unique positive stationary distribution. These conditions of being nonexplosive, irreducible, and having a positive stationary distribution imply that the process is positive recurrent [1]. Then, Theorem 8 in Lecture 4a implies the limiting distribution exists and equals the (unique positive) stationary distribution.

Example 4b.1

Taken from Example 6.1 on pp. 246–247 in [1]:

A continuous-time birth and death Markov chain has a birth rate $\lambda_i = b$ and a death rate $\mu_i = id$ for $i = 0, 1, 2, \dots$. Find the stationary probability distribution.

Solution:

To find the stationary probability distribution π , we use formulations derived in subsection 2.2.

For:

$$\lambda_i = b, \rightarrow \text{then } \lambda_0 \lambda_1 \cdots \lambda_{i-1} = b \cdot b \cdots b = b^i,$$

$$\mu_i = id, \rightarrow \text{then } \mu_1 \mu_2 \mu_3 \cdots \mu_i = 1d \cdot 2d \cdot 3d \cdots id = d^i i!,$$

from which,

$$\frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \mu_3 \cdots \mu_i} = \frac{b^i}{d^i i!} = \frac{(b/d)^i}{i!},$$

then, the general equation of the stationary probability distribution is given by

$$\pi_i = \frac{(b/d)^i}{i!} \pi_0. \quad (37)$$

Since, according to the power series of an exponential function,

$$\begin{aligned} 1 + \sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \mu_3 \cdots \mu_i} &= 1 + \sum_{i=1}^{\infty} \frac{(b/d)^i}{i!} \\ &= 1 + \frac{(b/d)}{1!} + \frac{(b/d)^2}{2!} + \frac{(b/d)^3}{3!} + \cdots \\ &= e^{b/d}, \end{aligned}$$

from Eq. (12),

$$\begin{aligned} \pi_0 &= \frac{1}{1 + \sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \mu_3 \cdots \mu_i}} \\ &= \frac{1}{e^{b/d}} \\ &= e^{-b/d}, \end{aligned}$$

and from Eq. (10), the unique stationary probability distribution is

$$\begin{aligned} \pi_i &= \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \mu_3 \cdots \mu_i} \pi_0 \\ &= \frac{(b/d)^i}{i!} e^{-b/d}, \end{aligned}$$

which is a Poisson distribution with parameter b/d , for $i = 0, 1, 2, 3, \dots$. The Markov chain is positive recurrent for all λ and μ .

Example 4b.2**Taken from Example 6.3 on pp. 256 in [1]:**

Consider a simple birth and death process with immigration. Let $\lambda = 0.5$ and $\mu = 1$. Find the stationary probability distribution when $\beta = 0.5, \beta = 1.0, \beta = 1.5$, and $\beta = 5.0$.

Solution:

The death rate is $\mu_i = i\mu = i$, and
the average instantaneous rate of growth is $\lambda_i = i\lambda + \beta = 0.5i + \beta$,
with various values of β as above.

Using the same steps as in Example 4b.1, that is

(1) When $\beta = 0.5$ * $\lambda_i = 0.5(i + 1)$, then

$$\begin{aligned}\lambda_0\lambda_1 \cdots \lambda_{i-1} &= [0.5(0+1)] \cdot [0.5(1+1)] \cdot [0.5(2+1)] \cdots [0.5(i-1+1)] \\ &= [0.5(1)] \cdot [0.5(2)] \cdot [0.5(3)] \cdots [0.5(i)] \\ &= 0.5^i i!\end{aligned}$$

* $\mu_i = i\mu$, then

$$\mu_1\mu_2\mu_3 \cdots \mu_i = 1 \cdot 2 \cdot 3 \cdots i = i!.$$

Combining the above,

$$\begin{aligned}\pi_i &= \frac{\lambda_0\lambda_1 \cdots \lambda_{i-1}}{\mu_1\mu_2 \cdots \mu_i} \pi_0 \\ &= \left(\frac{0.5(1)}{1}\right)\left(\frac{0.5(2)}{2}\right)\left(\frac{0.5(3)}{3}\right) \cdots \left(\frac{0.5(i)}{i}\right) \pi_0 \\ &= \frac{0.5^i (i+0)!}{0! i!} \pi_0 \\ &= \frac{0.5^i i!}{i!} \pi_0 \\ &= (0.5)^i \pi_0.\end{aligned}$$

(2) When $\beta = 1$ * $\lambda_i = 0.5i + 1$,

$$\begin{aligned}\lambda_0\lambda_1 \cdots \lambda_{i-1} &= [0.5(0)+1] \cdot [0.5(1)+1] \cdot [0.5(2)+1] \cdots [0.5(i-1)+1] \\ &= [0.5(2)] \cdot [0.5(3)] \cdot [0.5(4)] \cdots [0.5(i+1)]\end{aligned}$$

$$* \mu_i = i\mu, \\ \mu_1\mu_2\mu_3 \cdots \mu_i = 1 \cdot 2 \cdot 3 \cdots i = i!.$$

Then

$$\begin{aligned} \pi_i &= \frac{\lambda_0\lambda_1 \cdots \lambda_{i-1}}{\mu_1\mu_2 \cdots \mu_i} \pi_0 \\ &= \left(\frac{0.5(2)}{1}\right)\left(\frac{0.5(3)}{2}\right)\left(\frac{0.5(4)}{3}\right) \cdots \left(\frac{0.5(i+1)}{i}\right) \pi_0 \\ &= \frac{0.5^i (i+1)!}{1! i!} \pi_0 \\ &= \frac{0.5^i (i+1)!}{i!} \pi_0. \end{aligned}$$

(3) When $\beta = 1.5$

$$\begin{aligned} * \lambda_i &= 0.5i + 1.5, \\ \lambda_0\lambda_1 \cdots \lambda_{i-1} &= [0.5(0) + 1.5] \cdot [0.5(1) + 1.5] \cdot [0.5(2) + 1.5] \cdots [0.5(i-1) + 1.5] \\ &= [0.5(3)] \cdot [0.5(4)] \cdot [0.5(5)] \cdots [0.5(i+2)] \\ * \mu_i &= i\mu, \\ \mu_1\mu_2\mu_3 \cdots \mu_i &= 1 \cdot 2 \cdot 3 \cdots i = i!. \end{aligned}$$

Then

$$\begin{aligned} \pi_i &= \frac{\lambda_0\lambda_1 \cdots \lambda_{i-1}}{\mu_1\mu_2 \cdots \mu_i} \pi_0 \\ &= \left(\frac{0.5(3)}{1}\right)\left(\frac{0.5(4)}{2}\right)\left(\frac{0.5(5)}{3}\right) \cdots \left(\frac{0.5(i+2)}{i}\right) \pi_0 \\ &= \frac{0.5^i (i+2)!}{2! i!} \pi_0. \end{aligned}$$

(4) When $\beta = 5$

$$\begin{aligned} * \lambda_i &= 0.5i + 5, \\ \lambda_0\lambda_1 \cdots \lambda_{i-1} &= [0.5(0) + 5] \cdot [0.5(1) + 5] \cdot [0.5(2) + 5] \cdots [0.5(i-1) + 5] \\ &= [0.5(10)] \cdot [0.5(11)] \cdot [0.5(12)] \cdots [0.5(i+9)] \\ * \mu_i &= i\mu, \\ \mu_1\mu_2\mu_3 \cdots \mu_i &= 1 \cdot 2 \cdot 3 \cdots i = i!. \end{aligned}$$

Then

$$\begin{aligned}\pi_i &= \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i} \pi_0 \\ &= \left(\frac{0.5(10)}{1} \right) \left(\frac{0.5(11)}{2} \right) \left(\frac{0.5(12)}{3} \right) \cdots \left(\frac{0.5(i+9)}{i} \right) \pi_0 \\ &= \frac{0.5^i (i+9)!}{9! i!} \pi_0.\end{aligned}$$

We can deduce that, the general equation of the stationary probability distribution π_i , with birth rate λ , migration rate β , and death rate μ , is given by

$$\pi_i = \frac{\lambda^i (i + \beta/\lambda - 1)!}{(\beta/\lambda - 1)! \mu^i i!} \pi_0,$$

or

$$\pi_i = (\lambda/\mu)^i \cdot \frac{(i + \beta/\lambda - 1)!}{i! (\beta/\lambda - 1)!} \pi_0. \quad (38)$$

Taking $\beta/\lambda = n$, the term

$$\frac{(i+n-1)!}{i! (n-1)!} = \binom{i+n-1}{i},$$

is a binomial coefficient. Hence, we can write (38) as

$$\pi_i = (\lambda/\mu)^i \binom{i+n-1}{i} \pi_0. \quad (39)$$

To calculate π_0 , we may use the following

$$\pi_0 = \frac{1}{1 + \sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i}},$$

(1) when $\beta = 0.5$, and $n = 0.5/0.5 = 1$,

$$\begin{aligned}1 + \sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i} &= 1 + \sum_{i=1}^{\infty} (0.5)^i \\ &= 1 + 0.5 + (0.5)^2 + (0.5)^3 + \cdots \quad (\text{geometric series}) \\ &= \frac{1}{1 - 0.5} \\ &= 2,\end{aligned}$$

from which,

$$\pi_0 = \frac{1}{2}.$$

The stationary distribution is a geometric distribution,

$$\pi_i = (0.5)^{i+1}, \quad i = 0, 1, 2, \dots$$

(2) when $\beta = 1$, and $n = 1/0.5 = 2$,

$$\begin{aligned} 1 + \sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i} &= 1 + \sum_{i=1}^{\infty} \frac{(0.5)^i (i+1)!}{i!} \\ &= 1 + \frac{(0.5) 2!}{1!} + \frac{(0.5)^2 3!}{2!} + \frac{(0.5)^3 4!}{3!} + \dots \quad (\text{binomial series}) \\ &= \sum_{i=0}^{\infty} (0.5)^i \binom{2+i-1}{i} \\ &= \frac{1}{(1-0.5)^2} \\ &= 4, \end{aligned}$$

from which,

$$\pi_0 = \frac{1}{4}.$$

The unique stationary distribution is

$$\pi_i = \frac{1}{4} \cdot \frac{(0.5)^i (i+1)!}{i!} = \frac{(0.5)^{i+2} (i+1)!}{i!}, \quad i = 0, 1, 2, \dots$$

(3) when $\beta = 1.5$, and $n = 1.5/0.5 = 3$,

$$\begin{aligned} 1 + \sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i} &= 1 + \sum_{i=1}^{\infty} \frac{(0.5)^i (i+2)!}{2! i!} \\ &= 1 + \frac{(0.5) 3!}{2! 1!} + \frac{(0.5)^2 4!}{2! 2!} + \frac{(0.5)^3 5!}{2! 3!} + \dots \quad (\text{binomial series}) \\ &= \sum_{i=0}^{\infty} (0.5)^i \binom{3+i-1}{i} \\ &= \frac{1}{(1-0.5)^3} \\ &= 8, \end{aligned}$$

from which,

$$\pi_0 = \frac{1}{8}.$$

The unique stationary distribution is

$$\pi_i = \frac{1}{8} \cdot \frac{(0.5)^i (i+2)!}{2! i!} = \frac{(0.5)^{i+3} (i+2)!}{2! i!}, \quad i = 0, 1, 2, \dots$$

(4) when $\beta = 5$, and $n = 5/0.5 = 10$,

$$\begin{aligned} 1 + \sum_{i=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i} &= 1 + \sum_{i=1}^{\infty} \frac{(0.5)^i (i+9)!}{9! i!} \\ &= 1 + \frac{(0.5) 10!}{9! 1!} + \frac{(0.5)^2 11!}{9! 2!} + \frac{(0.5)^3 12!}{9! 3!} + \dots \quad (\text{binomial series}) \\ &= \sum_{i=0}^{\infty} (0.5)^i \binom{10+i-1}{i} \\ &= \frac{1}{(1-0.5)^{10}} \\ &= 2^{10}, \end{aligned}$$

from which,

$$\pi_0 = \left(\frac{1}{2}\right)^{10}.$$

The unique stationary distribution is

$$\pi_i = \left(\frac{1}{2}\right)^{10} \cdot \frac{(0.5)^i (i+9)!}{9! i!} = \frac{(0.5)^{i+10} (i+9)!}{9! i!}, \quad i = 0, 1, 2, \dots$$

Hence, we can deduce that

$$\pi_0 = (0.5)^n.$$

However, for general parameter values, π_0 is actually given by

$$\pi_0 = \left(1 - \frac{\lambda}{\mu}\right)^n.$$

Since, in this problem, $\beta/\lambda = n$ is an integer, then the m.g.f. corresponds to a negative binomial distribution with parameter $p = 1 - \lambda/\mu$, if and only if $\lambda < \mu$ and $\beta > 0$. The general formula for the stationary probability distribution is then given by

$$\pi_i = \binom{i+n-1}{i} p^n (1-p)^i, \quad i = 0, 1, 2, \dots \quad (40)$$

In the case $n = 1$, the negative binomial distribution is the same as the geometric distribution, as we have seen with $\beta = 0.5$ and $\lambda = 0.5$. The graphs of the stationary probability distributions for four different sets of the immigration rate β is shown in Fig. 4.

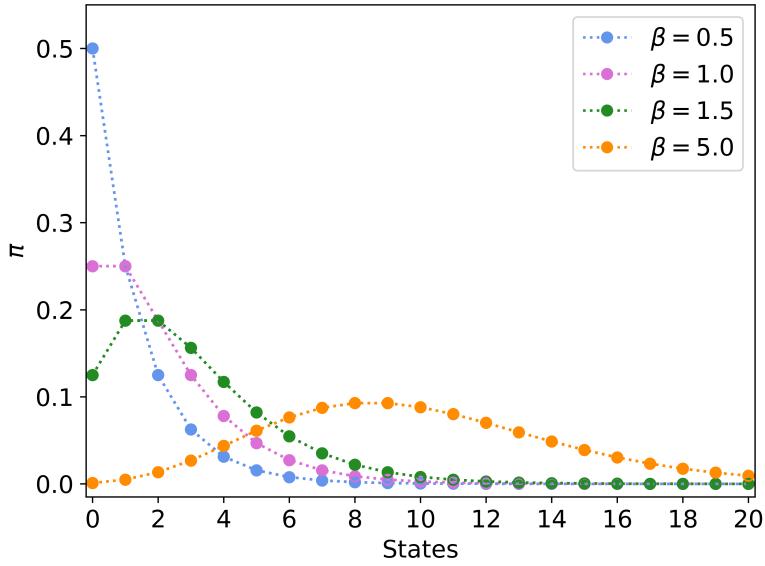


Figure 4: Stationary probability distributions for the simple birth, death, and immigration process with $\lambda = 0.5$, $\mu = 1$, and $\beta = 0.5$ (blue), 1.0 (magenta), 1.5 (green), and 5 (orange).

Using

$$m = \frac{\beta}{\mu - \lambda} \quad \text{and} \quad \sigma^2 = \frac{\beta\mu}{(\mu - \lambda)^2}$$

for the mean and variance, respectively, then the mean values for the stationary probability distributions corresponding to each parameter set are $m = 1, 2, 3$, and 10 , and the variance $\sigma^2 = 2, 4, 6$, and 20 .

The plots of population size over time with four sample paths of the birth and death process with migration are shown in Fig. 5. The Python code to produce these results are given in the Python Implementation section 3.4. The code are also available on the course website.

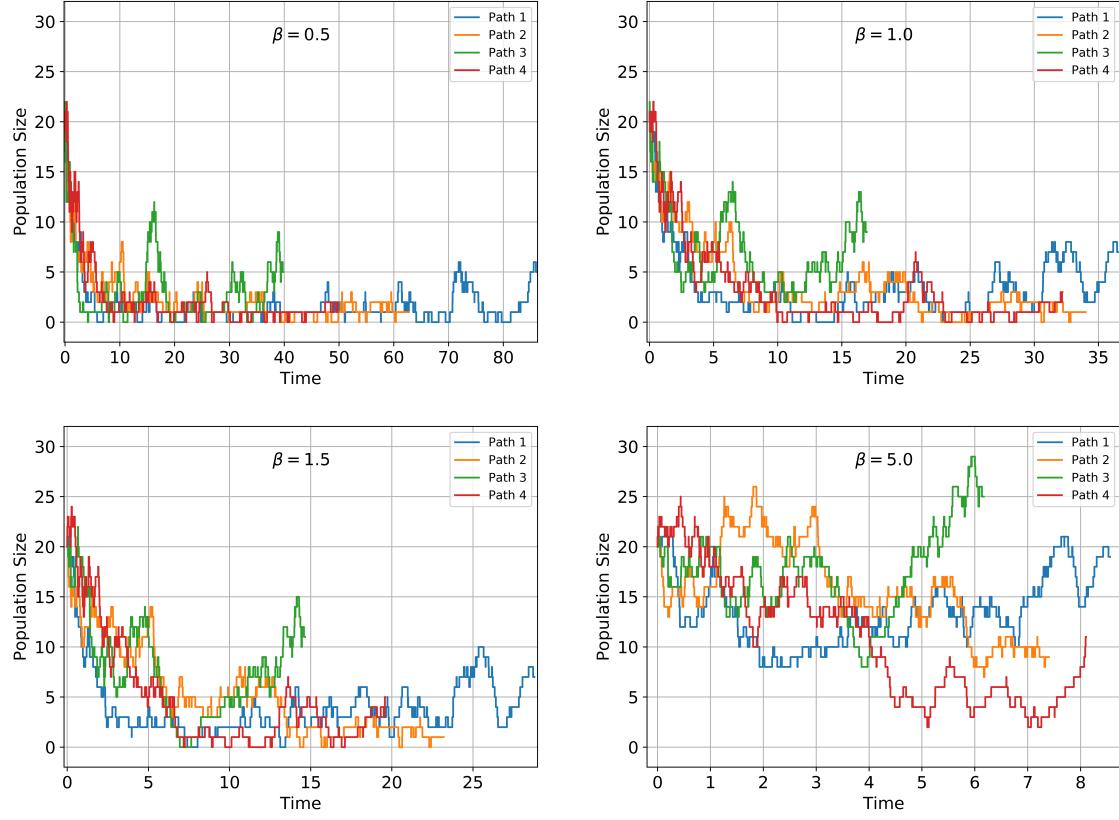


Figure 5: Four sample paths corresponding to the birth, death, and immigration process when $X(0) = 20$, $\lambda = 0.5$, $\mu = 1$, and $\beta = 0.5, 1.0, 1.5, 5$.

3 Python Implementation

All simulations were performed using Python v 3.6. All code presented here are available on the course website.

3.1 Yule Process Simulation

To simulate the Yule process, it is necessary to know the random variable for the time between births or interevent/sojourn time. It is shown that the random variable for the interevent/sojourn time is exponentially distributed; if the population is of size N , then the time T to the next event (or interevent/sojourn time) has a distribution equal to [1]

$$\text{Prob}\{T \geq h\} = \exp(-bNs).$$

To simulate a value $h \in T$, a uniformly distributed random number Y from the uniform distribu-

tion $U(0, 1)$ is selected in the range $0 \leq Y \leq 1$ [1]. Then

$$Y = \exp(-bNh),$$

and solving for h yields

$$h = -\frac{\ln(Y)}{bN}.$$

Note that in the simple birth process, as N increases, $h = -\ln(Y)/bN$ decreases, or, the interevent/sojourn time decreases as the population size increases [1].

In Python, to generate a random sample Y from the uniform distribution, we use:

```
import numpy as np
Y = np.random.rand()
```

which is used to get h :

```
h = -np.log(Y)/(b*N)
```

or:

```
h = -np.log(np.random.rand())/(b*N)
```

What we are calculating here is the interevent/sojourn times as the population size increases one by one, by using:

```
for j in range(numberofpaths):
    for i in range(N-1):
        h = - np.log(np.random.rand())/(b*X[i])
        s[j,i+1] = s[j,i] + h
```

and after which, the population is increased:

```
X[j,i+1] = X[j,i] + 1
```

The whole Python code for the Yule process can be written as follows:

```
import numpy as np
import matplotlib.pyplot as plt

np.random.seed(100)

def simplebirth(N, samplepaths, b, N0):
    s = np.zeros((samplepaths, N))
    X = np.zeros((samplepaths, N))
    X[:,0] = N0
```

```

for j in range(samplepaths):
    for i in range(N-1):
        h = - np.log(np.random.rand())/(b*X[j,i])
        s[j,i+1] = s[j,i] + h
        X[j,i+1] = X[j,i] + 1

return [s, X]

def deterministic(s, N, b, N0):
    tmax = max(s[:, -1])
    t = np.linspace(0, tmax, 10*N)
    yd = N0*np.exp(b*t)

    return[t, yd]

def solve_and_plot(N, paths, b, X0):
    fig, ax = plt.subplots()

    [sojourn, population] = simplebirth(N, paths, b, X0)
    [time, detsol] = deterministic(sojourn, N, b, X0)

    ## Sets axis ranges for plotting
    xmax = max([max(sojourn[k,:]) for k in range(paths)])
    ymin = min([min(population[k,:]) for k in range(paths)])
    ymax = max([max(population[k,:]) for k in range(paths)])

    ## Generates plots
    for r in range(walkers):
        plt.step(sojourn[r,:], population[r,:], where='post', label="Path %s"
                 " % str(r+1))

    plt.plot(time, detsol, 'k--', label="Deterministic")
    plt.axis([-0.2, xmax+0.2, ymin-2, ymax+2])
    ax.set_xlabel('Time', fontsize=14)
    ax.set_ylabel('Population Size', fontsize=14)
    plt.xticks(fontsize=14)
    plt.yticks(fontsize=14)
    plt.tight_layout()
    plt.legend(loc=2)
    plt.show()

maxPop = 100
walkers = 3
birthrate = 0.5
initPop = 5

solve_and_plot(maxPop, walkers, birthrate, initPop)

```

The code for calculating probability distributions of the Yule process given in Eq. (27) is

```

import numpy as np
import matplotlib.pyplot as plt
import scipy.misc
import scipy.special
from matplotlib.lines import Line2D
from cycler import cycler

def probdist(N, b):
    t = np.arange(1, N)
    k = np.arange(0, 51) # population vector
    p = np.zeros((len(t), len(k)))

    for i in range(len(t)):
        for j in range(len(k)):
            p[i][j] = scipy.special.binom(N+k[j]-1, k[j])\
                *np.exp(-b*N*t[i])\
                *(1-np.exp(-b*t[i]))**k[j]

    return [p, t, k]

def solve_and_plot(N, b):
    fig, ax = plt.subplots()
    mycolors = ['cornflowerblue', 'orchid', 'olive', 'gold']
    colors = ax.set_prop_cycle(cycler('color', mycolors))
    marker_style = dict(color=colors, linestyle=':',
                         marker='o', markersize=6)

    [p, t, k] = probdist(N, b)

    for y, fill_style in enumerate(Line2D.fillStyles):
        for r in range(len(t)):
            ax.plot(k, p[r][:], fillstyle=fill_style, **marker_style)

    ax.set_xlabel('Population Size $k$', fontsize=14)
    ax.set_ylabel('Probability $P_k(t)$', fontsize=14)

    plt.legend(['$t=%s$' % str(i+1) for i in range(len(t))], fontsize=14,
               loc=1)
    plt.xticks(fontsize=14)
    plt.yticks(fontsize=14)
    plt.text(0.5, 0.9, '$\lambda = %s$' % b, fontsize=14,
             horizontalalignment='center',
             verticalalignment='center',
             transform=ax.transAxes)
    plt.tight_layout()
    plt.show()

N = 5

```

```
birth = 0.5

solve_and_plot(N, birth)
```

3.2 Linear Death Process Simulation

```
import numpy as np
import matplotlib.pyplot as plt

np.random.seed(100)

def simpledeath(N, samplepaths, d, N0):
    s = np.zeros((samplepaths, N))
    X = np.zeros((samplepaths, N))
    X[:,0] = N0

    for j in range(samplepaths):
        for i in range(N-1):
            h = - np.log(np.random.rand())/(d*X[j,i])
            s[j,i+1] = s[j,i] + h
            X[j,i+1] = X[j,i] - 1

    return [s, X]

def deterministic(s, N, d, N0):
    tmax = max(s[:, -1])
    t = np.linspace(0, tmax, 10*N)
    yd = N0*np.exp(-d*t)

    return [t, yd]

def solve_and_plot(N, paths, d, X0):
    fig, ax = plt.subplots()

    [sojourn, population] = simpledeath(N, paths, d, X0)
    [time, detsol] = deterministic(sojourn, N, d, X0)

    ## Sets axis ranges for plotting
    xmax = max([max(sojourn[k,:]) for k in range(paths)])
    ymax = max([max(population[k,:]) for k in range(paths)])

    ## Generates plots
    for r in range(walkers):
        plt.step(sojourn[r,:], population[r,:], where='pre',
                 label="Path %s" % str(r+1))

    plt.plot(time, detsol, 'k--', label="Deterministic")
```

```

plt.axis([-0.1, xmax+0.2, 0, ymax+(ymax*0.01)])
ax.set_xlabel('Time', fontsize=14)
ax.set_ylabel('Population Size', fontsize=14)
plt.xticks(fontsize=14)
plt.yticks(fontsize=14)
plt.tight_layout()
plt.legend(loc=1)
plt.show()

maxPop = 20
walkers = 3
deathrate = 0.5
initPop = maxPop

solve_and_plot(maxPop, walkers, deathrate, initPop)

```

Note that the code for the Yule and linear death processes are almost identical, except for the following:

- Initial population in the linear death process must be equal to its maximal population size, whereas in the Yule process the initial population must be smaller than the maximal population size,
- In the Yule process the population increases, $X[j, i+1]=X[j, i]+1$, and in the linear death process the population decreases, $X[j, i+1]=X[j, i]-1$.

The code for probability distributions of linear death process using Eq. (32) is as follows:

```

import numpy as np
import matplotlib.pyplot as plt
import scipy.misc
import scipy.special
from matplotlib.lines import Line2D
from cycler import cycler

def probdist(N, mu):
    t = np.arange(1, 5)
    k = np.arange(0, N+1)
    p = np.zeros((len(t), len(k)))

    for i in range(len(t)):
        for j in range(len(k)):
            p[i][j] = scipy.special.binom(N, j) \
                *np.exp(-j*mu*t[i]) \
                *(1-np.exp(-mu*t[i]))**(N-j)

    return [p, t, k]

```

```

def solve_and_plot(N, mu):
    fig, ax = plt.subplots()
    mycolors = ['cornflowerblue', 'orchid', 'olive', 'gold']
    colors = ax.set_prop_cycle(cycler('color', mycolors))
    marker_style = dict(color=colors, linestyle=':',
                         marker='o', markersize=6)

    [p, t, k] = probdist(N, mu)

    for y, fill_style in enumerate(Line2D.fillStyles):
        for r in range(len(t)):
            ax.plot(k, p[r][:], fillstyle=fill_style, **marker_style)

    ax.set_xlabel('Population Size $k$', fontsize=14)
    ax.set_ylabel('Probability $P_k(t)$', fontsize=14)

    plt.legend(['$t=%s$' % str(i+1) for i in range(len(t))],
               fontsize=14, loc=1)
    plt.xticks(fontsize=14)
    plt.yticks(fontsize=14)
    plt.text(0.5, 0.9, '$\mu = %s$' % mu, fontsize=14,
             horizontalalignment='center',
             verticalalignment='center',
             transform=ax.transAxes)
    plt.tight_layout()
    plt.show()

N = 20
death = 0.5

solve_and_plot(N, death)

```

3.3 Birth and Death Process Simulation

For birth and death processes, we first look at the transition probabilities

$$\begin{aligned}
 P_{i,i+j}(\Delta t) &= \text{Prob}\{X(t + \Delta t) - X(t) = j | X(t) = i\} \\
 &= \begin{cases} 1 - (\mu_i + \lambda_i)\Delta t + o(\Delta t) & j = 0 \\ \lambda_i \Delta t + o(\Delta t) & j = 1 \\ \mu_i \Delta t + o(\Delta t) & j = -1 \\ o(\Delta t) & j \neq -1, 0, 1. \end{cases} \tag{41}
 \end{aligned}$$

According to Eq. (41), during a small time duration of length Δt , a transition occurs from state i to state $i+1$ with probability $\lambda_i \Delta t + o(\Delta t)$ and from state i to state $i-1$ with probability $\mu_i \Delta t + o(\Delta t)$. Given that a transition occurs at time t , the probability of:

- moving to state $i+1$ is

$$\frac{\lambda_i}{\lambda_i + \mu_i}$$

- moving to state $i-1$ is

$$\frac{\mu_i}{\lambda_i + \mu_i}$$

Recall that the random variable for the interevent/sojourn time is continuous and nonnegative, that is $T_i = W_{i+1} - W_i \geq 0$, where W_i is the time of the i th jump. In the linear birth and death process, if the events consist of births with rate $\lambda_i = i\lambda$ and deaths with rate $\mu_i = i\mu$, the interevent time T_i is given by

$$T_i = -\frac{\ln(U_1)}{(\lambda + \mu)i},$$

where U_1 is a generated random number.

Since there are two events, to determine whether there is a birth or a death, the unit interval $[0, 1]$ is divided into two subintervals:

- one subinterval has probability

$$\frac{\lambda_i}{\lambda_i + \mu_i} = \frac{i\lambda}{i\lambda + i\mu} = \frac{\lambda}{\lambda + \mu},$$

in which a birth occurs, and

- the other has probability

$$\frac{\mu_i}{\lambda_i + \mu_i} = \frac{i\mu}{i\lambda + i\mu} = \frac{\mu}{\lambda + \mu}$$

in which a death occurs.

We generate another random number, say U_2 , and if $U_2 < \lambda/(\lambda + \mu)$, then this random number lies in the first subinterval and a birth occurs. Otherwise, if $U_2 \geq \lambda/(\lambda + \mu)$, the random number lies in the second subinterval and a death occurs.

The deterministic analogue of this simple birth and death process is the differential equation $dn/dt = (\lambda - \mu)n$, with initial condition $n(0) = N$, and the solution is given by

$$n(t) = N e^{(\lambda-\mu)t}.$$

```

import numpy as np
import matplotlib.pyplot as plt

np.random.seed(100)

def simplebirthdeath(N, samplepaths, b, d):
    X0 = 50 # initial population size

    s = np.zeros((samplepaths, N))
    X = np.zeros((samplepaths, N))
    X[:,0] = X0
    s[:,0] = 0.001

    for j in range(samplepaths):
        i = 0
        while X[j,i] > 0 and i < (N-1):
            U1 = np.random.rand()
            U2 = np.random.rand()
            h = - np.log(U1)/((b+d)*X[j,i])
            s[j,i+1] = s[j,i] + h

            if U2 < b/(b+d):
                X[j,i+1] = X[j,i] + 1 # birth occurs
            else:
                X[j,i+1] = X[j,i] - 1 # death occurs

            i += 1

    return [s, X]

def solve_and_plot(N, paths, b, d):
    fig, ax = plt.subplots()

    [sojourn, population] = simplebirthdeath(N, paths, b, d)

    ## Sets axis ranges for plotting
    xmax = max([max(sojourn[k,:]) for k in range(paths)])
    ymax = max([max(population[k,:]) for k in range(paths)])

    sojourn[sojourn==0] = np.nan

    ## Generates plots
    for r in range(paths):
        plt.step(sojourn[r,:], population[r,:], where='pre', label="Path %s" % str(r+1))

    plt.axis([-0.1, xmax+0.2, 0, ymax+2])
    ax.set_xlabel('Time', fontsize=14)
    ax.set_ylabel('Population Size', fontsize=14)
    plt.xticks(fontsize=14)

```

```

plt.yticks(fontsize=14)
plt.tight_layout()
plt.legend(loc=1)
plt.show()

maxPop = 2000
walkers = 3
birth = 1.0
death = 1.0

solve_and_plot(maxPop, walkers, birth, death)

```

3.4 Birth and Death Process with Migration Simulation

A birth and death process with (im)migration is called a linear growth process if the increase of a population is determined by $\lambda_i = i\lambda + \beta$ and the decrease of population by $\mu_i = i\mu$, with $\lambda > 0$, $\beta > 0$, and $\mu > 0$. Here β represents the rate of immigration into the population and λ represents the individual birth rate.

For this process, there are three events and to determine whether population increases or decreases, the unit interval $[0, 1]$ is divided into three subintervals:

- the first subinterval has probability

$$\frac{i\lambda}{i\lambda + \beta + i\mu}$$

in which a birth occurs,

- the first subinterval has probability probability

$$\frac{i\mu}{i\lambda + \beta + i\mu}$$

in which a death occurs, and

- the third subinterval has probability

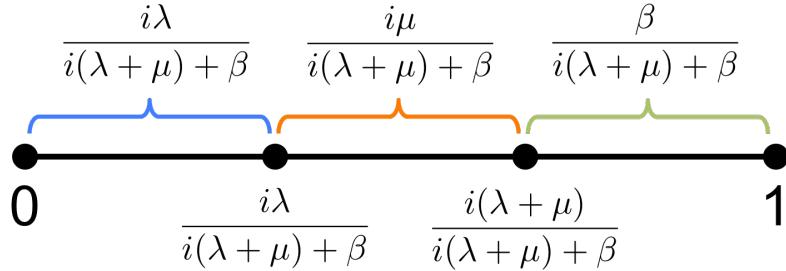
$$\frac{\beta}{i\lambda + \beta + i\mu}$$

in which an immigration occurs.

The endpoints of the unit interval are:

$$0, \quad \frac{i\lambda}{i(\lambda + \mu) + \beta}, \quad \frac{i(\lambda + \mu)}{i(\lambda + \mu) + \beta}, \quad 1.$$

The unit interval $[0, 1]$ having three subintervals with each probability and four end points is depicted in the following figure:



If a generated random number, say U_2 , is less than $i\lambda/(i(\lambda + \mu) + \beta)$, then the random number lies in the first subinterval and a birth occurs. If $U_2 \geq i\lambda/(i(\lambda + \mu) + \beta)$ and $U_2 < i(\lambda + \mu)/(i(\lambda + \mu) + \beta)$, the random number lies in the second subinterval and a death occurs. Otherwise, if $U_2 \geq i(\lambda + \mu)/(i(\lambda + \mu) + \beta)$, the random number lies in the third subinterval and an immigration occurs.

In general, if there are k events with rates $\alpha_k(n)$, $k = 1, 2, 3, \dots$, where n is the current state, the unit interval must be divided into k subintervals (and $k + 1$ endpoints). The probabilities of these events are

$$\frac{\alpha_k(n)}{\alpha(n)}, \quad \alpha(n) = \sum_k \alpha_k(n), \quad k = 1, 2, 3, \dots$$

The python code for the simple birth and death process with immigration is as follows

```
import numpy as np
import matplotlib.pyplot as plt

np.random.seed(10)

def birthdeathmigration(b, d, c, paths):
    X0 = 20 # initial population size
    N = 200 # maximal population size

    # Pre-allocation and initialization
    s = np.zeros((paths, N))
    X = np.zeros((paths, N))
    X[:, 0] = X0
    s[:, 0] = 0.00

    for j in range(paths):
        i = 0
        while X[j, i] >= 0 and i < (N-1):
            U1 = np.random.rand()
            U2 = np.random.rand()
            h = - np.log(U1)/((b+d)*X[j, i]+c)
            s[j, i+1] = s[j, i] + h
            i += 1
        X[j, i] = 0
```

```

    if U2 < b*X[j,i]/((b+d)*X[j,i]+c):
        X[j,i+1] = X[j,i] + 1 # a birth occurs
    elif U2 >= b*X[j,i]/((b+d)*X[j,i]+c) \
        and U2 < (b+d)*X[j,i]/((b+d)*X[j,i]+c):
        X[j,i+1] = X[j,i] - 1 # a death occurs
    elif U2 >= (b+d)*X[j,i]/((b+d)*X[j,i]+c):
        X[j,i+1] = X[j,i] + 1 # immigration occurs

    i += 1

    return [X, s]

def solve_and_plot(b, d, c, paths):
    fig, ax = plt.subplots()

    [population, sojourn] = birthdeathmigration(b, d, c, paths)

    ## Sets axis ranges for plotting
    xmax = max([max(sojourn[k,:]) for k in range(paths)])

    ## Generates plots
    for r in range(paths):
        plt.step(sojourn[r,:], population[r,:],
                 where='pre', label="Path %s" % str(r+1))

    plt.axis([-0.2, xmax+0.2, -2, 32]) #ymax+2])
    ax.set_xlabel('Time', fontsize=14)
    ax.set_ylabel('Population Size', fontsize=14)
    plt.text(0.5, 0.9, '$\beta = %s$' % c, fontsize=14,
             horizontalalignment='center',
             verticalalignment='center',
             transform=ax.transAxes)
    plt.xticks(fontsize=14)
    plt.yticks(fontsize=14)
    plt.tight_layout()
    plt.legend(loc=1)
    plt.grid(True)
    plt.show()

    ## Uncomment to save the figure
    #fig.savefig('Figures/bd_mig_1p0.png', dpi=600)

b = 0.5 # birth rate
d = 1.0 # death rate
c = 1.0 # immigration rate
walkers = 4

solve_and_plot(b, d, c, walkers)

```

References

- [1] Linda J. S. Allen. *An Introduction to Stochastic Processes with Applications to Biology 2nd Edition*. CRC Press, 2010.
- [2] Paul G. Hoel, Sidney C. Port, Charles J. Stone. *Introduction to Stochastic Processes*. Houghton Mifflin Company, 1972.
- [3] Samuel Karlin, Howard M. Taylor. *A First Course in Stochastic Processes*. Academic Press, 1975.
- [4] Samuel Karlin, Howard M. Taylor. *A Second Course in Stochastic Processes*. Academic Press, 1981.
- [5] Gregory F. Lawler. *Introduction to Stochastic Processes 2nd Edition*. Chapman & Hall, 2006.
- [6] Mario Lefebvre. *Applied Stochastic Processes*. Springer, 2007.
- [7] Sheldon M. Ross. *Introduction to Probability Models 9th Edition*. Elsevier, 2007.
- [8] Sheldon M. Ross. *Stochastic Processes 2nd Edition*. John Wiley & Sons Inc., 1996.
- [9] Howard M. Taylor, Samuel Karlin. *An Introduction to Stochastic Modeling 3rd Edition*. Academic Press, 1998.
- [10] Henry C. Tuckwell. *Elementary Applications of Probability Theory 2nd Edition*. Springer-Science+Business Media, BV, 1995.
- [11] Law of Total Probability,
https://en.wikipedia.org/wiki/Law_of_total_probability
- [12] Point Process,
https://en.wikipedia.org/wiki/Point_process
- [13] Counting Processes
https://www.probabilitycourse.com/chapter11/11_1_1_counting_processes.php