EM509: Individual Project
Birth-death Process

Y.D. Sumanasekera (E/14/337)

Introduction

Let $\{X(t)\ ;\ t \geq 0\}$ be a continuous-time Markov chain with state space $\{0, 1, 2, \ldots\}$. If one interprets $X(t)$ as the size of a randomly varying population, and the only possible transitions within an infinitesimal interval are limited to either a 'birth' (which increases the population size by 1) or a 'death' (which decreases the population size by 1), then the process $\{X(t)\}$ is called a (finite) birth-death (BD) process.

$*$ Note that, a continuous-time BD Markov chain $X(t)$ can have either a finite state space $\{0, 1, 2, \ldots, N\}$ or an infinite state space $\{0, 1, 2, \ldots\}$. In this text, I have considered the case where the state space is infinite.

The transition rate λ_i is viewed as the birth rate while the transition rate μ_i is viewed as the death rate when the process is in state i. When a birth occurs, the process goes from state i to state $i + 1$. Similarly, when a death occurs, the process goes from state i to state $i - 1$. It is assumed that all births and deaths occur independently.

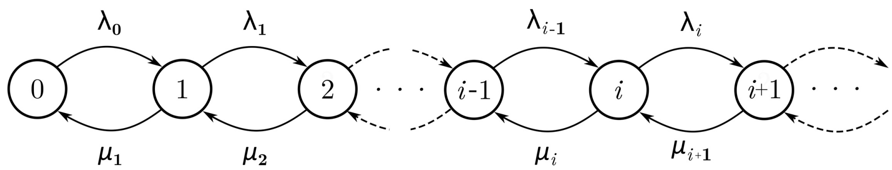

Figure 1: State diagram for a BD process. Note that in one time step, the chain may move at most one unit to the left or to the right (i.e. $\{X(t)\}$ can only jump to adjacent states).

Markov Property. As mentioned, BD processes are characterized by the property that whenever a transition occurs from one state to another, then this transition (which represents a birth or death) can only be to a neighbouring state. Further, it is assumed that all births and deaths occur independently. Suppose that the state space $\mathcal{S} = \{0, 1, 2, \ldots, i, \ldots\}$, if the current state (at time instance t) is $X(t) = i$, then the state at the next instance can only be $i+1$ or $i-1$. This means that the current value $X(t)$ suffices to determine the distribution of the future value $X(t + s)$ for all $s > 0$. Therefore, the BD process $\{X(t)\}$ is Markovian in continuous time.

A BD process is characterized by the birth rate $\{\lambda_n\}_{n=0}^{\infty}$ and death rate $\{\mu_n\}_{n=0}^{\infty}$ which vary according to state i of the system. If λ_i and μ_i are independent of i, it is known as a homogeneous BD process. If the state space has a minimum element i, then we must have $\mu_i = 0$. In addition, if $\lambda_i = 0$ as well, then the boundary point i is absorbing. Similarly, if the state space has a maximum element i, then we must have $\lambda_i = 0$. In addition, if $\mu_i = 0$ as well, then the boundary point i is absorbing. If some state i is not a boundary point, then typically we have $\lambda_i + \mu_i > 0$, so that i is stable.

Further, we can define a *pure birth process* as a BD process with $\mu_i = 0$ for all i. Similarly, a *pure death process* corresponds to a BD process with $\lambda_i = 0$ for all i. Thus, it can be observed that the Poisson process is a special case of the BD process, as it is a pure birth process, where the birth rate is homogeneous (i.e. $\lambda_i = \lambda$ for all $i \geq 0$), and no deaths occur (i.e. $\mu_i = 0$ for all $i \geq 0$).

1. Construction of a General BD Process

Assume $X(t) = i \geq 1$ at time $t \geq 0$. Let $B(i)$ and $D(i)$ denote two exponentially distributed random variables with parameters λ_i and μ_i respectively. These random variables describe the holding time in the state i. Intuitively, we can think of $B(i)$ as the time until a birth and $D(i)$ as the time until a death (when a population size is i). The population increases by one if the birth occurs prior to death and decreases by one otherwise. If $B(i)$ and $D(i)$ are independent, exponentially distributed random variables, then their minimum is exponentially distributed with parameter $(\lambda_i + \mu_i)$[1]. Once the process enters state i, it holds (sojourns) in the given state for

[1] Check Appendix: Section A for the corresponding proof.

some random length of time, exponentially distributed with parameter $(\lambda_i + \mu_i)$. A transition from i to $i+1$ is made if $B(i) < D(i)$, which occurs with probability,

$$P[B(i) < D(i)] = \frac{\lambda_i}{\lambda_i + \mu_i}$$

Similarly, a transition from i to $i-1$ is made if $B(i) > D(i)$, which occurs with probability,

$$P[B(i) > D(i)] = \frac{\mu_i}{\lambda_i + \mu_i}$$

where $\lambda_i, \mu_{i+1} > 0$ for all i. Note that we set $\mu_0 = 0$ so that there is a reflecting boundary at the origin and the chain remains on the positive integers at all times.

Thus, we look at the two exponential random variables and decide for a jump to the left or to the right depending on which random variable occurs first. If the next state is chosen to be $i+1$, then the process sojourns in this state according to the exponential distribution with parameter $(\lambda_{i+1} + \mu_{i+1})$. Iterating this procedure gives a construction of a continuous-time BD process. The motion is analogous to that of a random walk except that transitions occur at random times rather than at fixed time periods. Note that, the number of visits back to the same state is ignored since in a continuous-time process, transitions from state i back to i would not be identifiable.

Definition. The infinitesimal transition probabilities of a BD process with birth and death rates λ_i and μ_i, respectively, have the following properties,

$$p_{i,j}(\Delta t) = \begin{cases} \lambda_i \Delta t + o(\Delta t), & j = i+1 \\ \mu_i \Delta t + o(\Delta t), & j = i-1 \\ 1 - (\lambda_i + \mu_i)\,\Delta t + o(\Delta t), & j = i \\ o(\Delta t), & |j-i| \geq 2 \end{cases}$$

for a sufficiently small Δt, $\mu_0 = 0$, and $\lambda_i, \mu_{i+1} > 0$ for $i = 0, 1, 2, \ldots$ Here, $o(\Delta t)$ designate higher-order terms in Δt such that $\lim_{\Delta t \to \infty} \frac{o(\Delta t)}{\Delta t} = 0$. This indicates that the probability of two or more events occurring within the same Δt interval is negligible. That is, in a small interval of time Δt, at most only one change in a state can occur; either a birth or a death.

The forward Kolmogorov differential equations for $p_{i,j}(t)$ can be derived directly from the above properties[2] and can be expressed as,

$$\begin{aligned} \frac{dp_{i,0}(t)}{dt} &= -\lambda_0 p_{i,0}(t) + \mu_1 p_{i,1}(t) \\ \frac{dp_{i,j}(t)}{dt} &= \lambda_{j-1} p_{i,j-1}(t) - (\lambda_j + \mu_j)\, p_{i,j}(t) + \mu_{j+1} p_{i,j+1}(t), \quad \text{for } j \geq 1 \end{aligned} \tag{1}$$

with initial conditions,

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

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

$$\frac{d}{dt}(\boldsymbol{P}(t)) = \boldsymbol{P}(t)\boldsymbol{Q}$$

where $\boldsymbol{Q}$ is the transition rate matrix or infinitesimal generator matrix (for an infinite state space) given by,

$$\boldsymbol{Q} = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdots \\ \mu_1 & -\lambda_1 - \mu_1 & \lambda_1 & 0 & \cdots \\ 0 & \mu_2 & -\lambda_2 - \mu_2 & \lambda_2 & \cdots \\ 0 & 0 & \mu_3 & -\lambda_3 - \mu_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

The transition matrix $\boldsymbol{P}$ for the associated embedded Markov chain $\{Y(n)\}$ is easily defined from $\boldsymbol{Q}$.

$$\boldsymbol{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots \\ \frac{\mu_1}{\lambda_1+\mu_1} & 0 & \frac{\lambda_1}{\lambda_1+\mu_1} & 0 & \cdots \\ 0 & \frac{\mu_2}{\lambda_2+\mu_2} & 0 & \frac{\lambda_2}{\lambda_2+\mu_2} & \cdots \\ 0 & 0 & \frac{\mu_3}{\lambda_3+\mu_3} & 0 & \ddots \\ \vdots & \vdots & & \ddots & \ddots \end{bmatrix}$$

[2]Check Appendix: Section B for the complete derivation

It is easy to see that $\{Y(n)\}$ is a discrete-time BD chain. Note that $\{X(t)\}$ and $\{Y(n)\}$ start in the same state and visit the same states in the same order. The difference between the two processes is that the time between two jumps is random for $\{X(t)\}$ while it is always one for $\{Y(n)\}$. In both cases, it is assumed that $\lambda_i + \mu_i > 0$ for $i = 0, 1, 2, \dots$ If, for any $i, \lambda_i + \mu_i = 0$, then state i is absorbing.

2. Irreducibility

A continuous-time Markov chain is irreducible if and only if its embedded discrete-time Markov chain is irreducible. Thus, it can be verified easily from the transition matrix $\boldsymbol{P}$ that the chain is irreducible if and only if $\lambda_i > 0$ and $\mu_{i+1} > 0$ for $i = 0, 1, 2, \dots$. If any $\lambda_i = 0$, then $p^{(n)}_{i,i+1} = 0$ for all n, and if any $\mu_i = 0$, then $p^{(n)}_{i,i-1} = 0$ for all n. Alternatively, we can use the following definition to determine the irreducibility of a given BD process.

Definition. A chain is said to be irreducible if its transition probabilities $p_{i,j}(t)$ are strictly positive for all $t > 0$ and all states i and j.

That is, an irreducible chain can go from any given state to another state in some fixed time. Note that, for a BD process with $\mu_0 = 0$ and $\lambda_i, \mu_{i+1} > 0$ for $i = 0, 1, 2, \dots$, this condition holds. That is, for any two states i and j, exactly one of the following two conditions holds,

1. $p_{i,j}(t) = 0$ for all $t > 0$, or
2. $p_{i,j}(t) > 0$ for all $t > 0$

3. Limiting Behaviour and Stationary Probability Distribution

For a general BD process with an infinite state space, a unique positive stationary probability distribution $\pi = (\pi_0, \pi_1, \pi_2, \dots)^\top$ exists if,

$$\mu_0 = 0 \quad \text{and} \quad \mu_{i+1}, \lambda_i > 0 \quad \text{for } i = 0, 1, 2, \dots$$

and

$$\sum_{i=1}^{\infty} \prod_{j=0}^{i-1} \frac{\lambda_j}{\mu_{j+1}} < \infty$$

The stationary probability distribution satisfies[3],

$$\pi_i = \pi_0 \prod_{j=0}^{i-1} \frac{\lambda_j}{\mu_{j+1}} \qquad \text{for } i = 1, 2, \dots$$

and

$$\pi_0 = \frac{1}{1 + \sum_{i=1}^{\infty} \prod_{j=0}^{i-1} \frac{\lambda_j}{\mu_{j+1}}}$$

Further, from the above equations we can deduce,

$$\mu_i \pi_i = \lambda_{i-1} \pi_{i-1}, \qquad \text{for all } i = 1, 2, 3, \dots \tag{2}$$

The left-hand side of equation (2) represents the rate of transition from state i to state $i-1$ and the right-hand side is the transition rate from state $i-1$ to state i, and these two quantities should balance out each other in the steady state. The equations (2) are called *detailed balance equations*, which hold between every pair of adjacent states; $i-1$ and i of the BD process for all $i = 1, 2, 3 \dots$

4. Classification of States

Definition. An irreducible continuous-time Markov chain is recurrent (or transient, respectively) if and only if its embedded discrete-time Markov chain is recurrent (or transient). Note that an absorbing state will always be recurrent.

We introduce the following notation,

$$A = \sum_{i=1}^{\infty} \prod_{j=0}^{i-1} \frac{\mu_{j+1}}{\lambda_{j+1}} \qquad\qquad B = \sum_{i=1}^{\infty} \prod_{j=0}^{i-1} \frac{\lambda_j}{\mu_{j+1}}$$

A BD process $\{X(t)\}$ is recurrent if and only if $A = \infty$. Equivalently, $\{X(t)\}$ is transient if and only if $A < \infty$. Further, when $\{X(t)\}$ is recurrent, it is positive recurrent if and only if $B < \infty$ and null recurrent if and only if $B = 0$.

[3] Check Appendix: Section C for the corresponding proof

5. Application of BD processes in Queueing Theory (M/M/1 Queue)

A BD process can be viewed as a generalization of the M/M/1 queuing system, which is one of the most basic queueing models in queueing theory. A canonical example is a single processor system where the jobs are processed in order of arrival (i.e. on a first-come-first-served basis). Jobs that arrive to a non-empty system will line up in a queue (waiting buffer) for service. After a job is processed (i.e. the job completes its service) it leaves the system. The jobs are assumed to arrive at the processor according to a Poisson process with some rate $\lambda > 0$. Equivalently, the inter-arrival times of succeeding jobs follow an i.i.d. exponential distribution with parameter λ. The service times of jobs are i.i.d exponential random variables with some rate $\mu > 0$. It is also implicitly assumed that the service times are independent of the arrival process. Further, the capacity of the queue is considered to be infinite. With these assumptions we have,

$$\mathbb{E}[\text{inter-arrival time}] = \frac{1}{\text{arrival rate}} = \frac{1}{\lambda} \qquad \mathbb{E}[\text{service time}] = \frac{1}{\text{service rate}} = \frac{1}{\mu}$$

In queueing theory, the ratio of arrival rate to service rate plays a significant role in measuring the performance of queueing systems. Therefore, we define,

$$\text{Average utilization of the system } = \rho = \frac{\lambda}{\mu}$$

We are interested in the behavior of $X(t)$, the number of jobs in the system at time $t, t \geq 0$. As the inter-arrival times and the service times are all independent, exponentially distributed random variables, the process $\{X(t) : t \geq 0\}$ is a continuous-time Markov chain with an infinite state space $\{0, 1, 2, \ldots\}$. The notation "M/M/1" is the standardized Kendall's notation used to describe the underlying queueing model as a Markov input and Markov output with a single server (processor).

Clearly M/M/1 queue is equivalent to a homogeneous BD process with birth rates $\lambda_i = \lambda$ for $i \geq 0$, and death rates $\mu_i = \mu$ for $i \geq 1$ and $\mu_0 = 0$. A birth or a death may be considered as an arrival or departure respectively, in a queueing system. However, even for this relatively simple queueing system, the distribution of the time-dependent Markov process is quite complicated. Therefore, we restrict ourselves to the stationary (limiting) distribution. It is easily seen that the Markov process is irreducible as all states can be reached from each other. Following the discussion in Section 3, the condition for the stationary distribution to exist becomes,

$$\sum_{i=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^i < \infty$$

The sum on the left-hand side is simply a geometric series, and thus, converges if and only if $\rho = \frac{\lambda}{\mu} < 1$. This condition for the stationary distribution to exist is equivalent to $\lambda < \mu$, and has the intuitive interpretation that the arrival rate to the system (λ) must be less than the service rate of the processor (μ). If $\lambda > \mu$, then jobs are arriving to the system at a faster rate than the processor can serve them, and thus, the length of the queue grows infinitely. Such a system is said to be unstable. Thus, $\rho < 1$ is necessary and sufficient for the steady-state solution. Alternatively, it can be said that the Markov process is positive recurrent if and only if $\rho < 1$. This is intuitively obvious as every finite state i (i jobs) will be visited infinitely often, with finite mean inter-visit times, as long as the processor is offered less jobs per time unit than it can handle. Thus, given $\rho < 1$, the stationary probability distributions can be expressed as follows,

$$\begin{aligned}
\pi_0 &= \frac{1}{1 + \sum_{i=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^i} = \frac{1}{1 + \left(\frac{\lambda}{\mu}\right) + \left(\frac{\lambda}{\mu}\right)^2 + \ldots} \\
&= 1 - \frac{\lambda}{\mu} = 1 - \rho \\
\pi_0 &= 1 - \rho
\end{aligned}$$

and

$$\pi_i = \left(\frac{\lambda}{\mu}\right)^i \pi_0 = \rho^i (1 - \rho) \quad \text{for } i = 1, 2, 3, \ldots$$

which is the probability function of a geometric distribution. Thus, $\pi = (\pi_0, \pi_1, \ldots \pi_i, \ldots)^\top$ forms the stationary distribution, where π_i represents the probability that there are i jobs in the system when the system is in operation for a long period of time and the transient behaviour subsides and the conditions for stationarity are satisfied. In other words, as $t \to \infty$, the probability that the system is in a particular state, does not depend on t nor on the initial state.

Appendix

• Section A

Proposition: Let $T_1, T_2, \dots, T_n$ be random variables having independent exponential distributions with rates $\alpha_1, \alpha_2, \dots, \alpha_n$. The random variable: $\min(T_1, T_2, \dots, T_n)$ is also exponentially distributed with a rate $(\alpha_1 + \alpha_2 + \cdots + \alpha_n)$. Moreover, the probability,

$$P(\min(T_1, T_2, \dots, T_n) = T_k) = \frac{\alpha_k}{\alpha_1 + \alpha_2 + \cdots + \alpha_n}$$

Proof: Note that,

$$P(\min(T_1, T_2, \dots, T_n) > t) = P(T_1 > t; T_2 > t; \dots; T_n > t)$$

and by independence we get,

$$\begin{aligned} P(\min(T_1, T_2, \dots, T_n) > t) &= P(T_1 > t) P(T_2 > t) \dots P(T_n > t) \\ &= e^{-\alpha_1 t} e^{-\alpha_2 t} \dots e^{-\alpha_n t} \\ &= e^{-(\alpha_1 + \alpha_2 + \dots \alpha_n) t} \end{aligned}$$

This is sufficient to prove that $\min(T_1, T_2, \dots, T_n)$ is exponentially distributed with a rate $(\alpha_1 + \alpha_2 + \cdots + \alpha_n)$.

Let $S_k = \min_{j \neq k} T_j)$. Recall that the probability that two continuous and independent random variables being equal, is zero. Thus,

$$P(\min(T_1, T_2, \dots, T_n) = T_k) = P(T_k < S_k)$$

However, according to the computation above, S_k is exponentially distributed with rate $\beta_k = \sum_{j \neq k} \alpha_j$ and S_k and T_k are independent. Thus,

$$\begin{aligned} P(T_k < S_k) &= \iint_{0<t<s} \alpha_k e^{-\alpha_k t} \beta_k e^{-\beta_k s} dt ds \\ &= \frac{\alpha_k}{\alpha_k + \beta_k} \\ &= \frac{\alpha_k}{\alpha_1 + \alpha_2 + \cdots + \alpha_n} \end{aligned}$$

• Section B

Proposition: The transition probabilities of a BD process satisfy a system of differential equations known as the backward Kolmogorov differential equations given by,

$$\frac{dp_{i,j}(t)}{dt} = \mu_i p_{i-1,j}(t) - (\lambda_i + \mu_i) p_{i,j}(t) + \lambda_i p_{i+1,j}(t) \quad \text{for } i \geq 1 \text{ and } j \geq 0$$

and

$$\frac{dp_{0,j}(t)}{dt} = -\lambda_0 p_{0,j}(t) + \lambda_0 p_{1,j}(t) \quad \text{for } j \geq 0$$

together with the initial conditions,

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

Proof: Assume $\Delta t > 0$ is an infinitesimal time interval and consider the transition probability $p_{i,j}(t + \Delta t)$. Applying the Chapman-Kolmogorov equation we have,

$$\begin{aligned} p_{i,j}(t + \Delta t) &= \sum_{k=0}^{\infty} p_{i,k}(\Delta t) p_{k,j}(t) \\ &= p_{i,i-1}(\Delta t) p_{i-1,j}(t) + p_{i,i+1}(\Delta t) p_{i+1,j}(t) + p_{i,i}(\Delta t) p_{i,j}(t) + \sum_{k \notin I} p_{i,k}(\Delta t) p_{k,j}(t) \end{aligned}$$

However, the transition probability $p_{i,j}(t + \Delta t)$ can be expressed in terms of the transition probabilities at time t as follows,

$$\begin{aligned} p_{i,j}(t + \Delta t) = [\mu_i \Delta t + o(\Delta t)] p_{i-1,j}(t) + [\lambda_i \Delta t + o(\Delta t)] p_{i+1,j}(t) + [1 - (\lambda_i + \mu_i) \Delta t + o(\Delta t)] p_{i,j}(t) \\ + \sum_{k \notin I} p_{i,k}(\Delta t) p_{k,j}(t) \end{aligned} \tag{3}$$

where $I = \{i-1, i, i+1\}$.

Consider the final term in equation (3),

$$\sum_{k \notin I} p_{i,k}(\Delta t) p_{k,j}(t) \leq \sum_{k \notin I} p_{i,k}(\Delta t) = 1 - \sum_{k \in I} p_{i,k}(\Delta t)$$

$$\begin{aligned} 1 - \sum_{k \in I} p_{i,k}(\Delta t) &= 1 - \left[1 - (\lambda_i + \mu_i)\,\Delta t + o(\Delta t) + \lambda_i \Delta t + o(\Delta t) + \mu_i \Delta t + o(\Delta t)\right] \\ &= o(\Delta t) \end{aligned}$$

Therefore, equation (3) reduces to,

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

By re-arranging the terms, we obtain,

$$\begin{aligned} \frac{p_{i,j}(t+\Delta t) - p_{i,j}(t)}{\Delta t} &= \frac{\mu_i p_{i-1,j}(t)\Delta t + \lambda_i p_{i+1,j}(t)\Delta t - p_{i,j}(t)\Delta t\,(\lambda_i + \mu_i) + o(\Delta t)}{\Delta t} \\ &= \mu_i p_{i-1,j}(t) + \lambda_i p_{i+1,j}(t) - p_{i,j}(t)\,(\lambda_i + \mu_i) + \frac{o(\Delta t)}{\Delta t} \end{aligned}$$

Taking the limit on both sides as $\Delta t \to 0$, the above equation reduces to,

$$\frac{dp_{i,j}(t)}{dt} = \mu_i p_{i-1,j}(t) + \lambda_i p_{i+1,j}(t) - p_{i,j}(t)\,(\lambda_i + \mu_i)$$

The equation corresponding to $i = 0$ is a special case of this derivation with $\mu_0 = 0$,

$$\begin{aligned} \frac{dp_{0,j}(t)}{dt} &= \mu_0 p_{i-1,j}(t) + \lambda_0 p_{0+1,j}(t) - p_{0,j}(t)\,(\lambda_0 + \mu_0) \\ &= \lambda_0 p_{1,j}(t) - \lambda_0 p_{0,j}(t) \end{aligned}$$

Thus, the backward Kolmogorov differential equations for a general BD process can be expressed as,

$$\frac{dp_{i,j}(t)}{dt} = \mu_i p_{i-1,j}(t) - (\lambda_i + \mu_i)\, p_{i,j}(t) + \lambda_i p_{i+1,j}(t) \qquad \text{for } i \geq 1 \text{ and } j \geq 0$$

and

$$\frac{dp_{0,j}(t)}{dt} = -\lambda_0 p_{0,j}(t) + \lambda_0 p_{1,j}(t) \qquad \text{for } j \geq 0$$

The backward Kolmogorov differential equations are deduced by decomposing the time interval $(0, t+\Delta t)$, where $\Delta t > 0$ is sufficiently small, into two sub-intervals $(0, \Delta t)$ and $(\Delta t, t+\Delta t)$ and examining the transition in each period separately. In this sense, the backward equations result from a 'first-step analysis', the first step being over the short time interval of duration Δt. Another set of equations arises from a 'last-step analysis' which proceeds by splitting the time interval $(0, t+\Delta t)$ into two sub-intervals $(0, t)$ and $(t, t+\Delta t)$ and adapting the preceding reasoning. From this viewpoint, under more stringent conditions, we can derive a further system of differential equations known as the forward Kolmogorov differential equations given by,

$$\frac{dp_{i,j}(t)}{dt} = \lambda_{j-1} p_{i,j-1}(t) - (\lambda_j + \mu_j)\, p_{i,j}(t) + \mu_{j+1} p_{i,j+1}(t) \qquad \text{for } i \geq 0 \text{ and } j \geq 1$$

and

$$\frac{dp_{i,0}(t)}{dt} = -\lambda_0 p_{i,0}(t) + \mu_1 p_{i,1}(t) \qquad \text{for } i \geq 0$$

together with the initial conditions,

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

- **Section C**

If for a general BD process, a stationary probability distribution $\pi = (\pi_0, \pi_1, \pi_2, \ldots)^\top$ exists, it is unique, and for each state j,

$$\lim_{t\to\infty} p_{i,j}(t) = \pi_j \geq 0 \tag{4}$$

exist and are independent of the initial state i.

To determine if a limiting distribution exists and to determine its corresponding values, we consider the forward Kolmogorov differential equations,

$$\frac{dp_{i,0}(t)}{dt} = -\lambda_0 p_{i,0}(t) + \mu_1 p_{i,1}(t) \tag{5}$$

$$\frac{dp_{i,j}(t)}{dt} = \lambda_{j-1} p_{i,j-1}(t) - (\lambda_j + \mu_j)\, p_{i,j}(t) + \mu_{j+1} p_{i,j+1}(t), \quad \text{for } j \geq 1 \tag{6}$$

with initial conditions,

$$p_{i,j}(0) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

Since we are interested only in the statistical equilibrium properties of the system, we take the limits as $t \to \infty$ in both equations (5) and (6). The limits on the right-hand side exists according to equation (4). The limits on the left-hand side; the derivatives $\frac{dp_{i,j}(t)}{dt}$, also exist as the probabilities converge to a constant. Thus,

$$\lim_{t\to\infty} \frac{dp_{i,j}(t)}{dt} = 0$$

Taking the corresponding limits of equation (5),

$$\begin{aligned} \lim_{t\to\infty} \frac{dp_{i,0}(t)}{dt} &= -\lambda_0 \lim_{t\to\infty} p_{i,0}(t) + \mu_1 \lim_{t\to\infty} p_{i,1}(t) \\ 0 &= -\lambda_0 \pi_0 + \mu_1 \pi_1 \\ \pi_1 &= \frac{\lambda_0}{\mu_1} \pi_0 \end{aligned} \tag{7}$$

Subsequently, taking the limits of equation (6),

$$\begin{aligned} \lim_{t\to\infty} \frac{dp_{i,j}(t)}{dt} &= \lambda_{j-1} \lim_{t\to\infty} p_{i,j-1}(t) - (\lambda_j + \mu_j) \lim_{t\to\infty} p_{i,j}(t) + \mu_{j+1} \lim_{t\to\infty} p_{i,j+1}(t) \\ 0 &= \lambda_{j-1}\pi_{j-1} - (\lambda_j + \mu_j)\,\pi_j + \mu_{j+1}\pi_{j+1} \end{aligned} \tag{8}$$

Equations (7) and (8) can be solved recursively.

$$\begin{aligned} \pi_2 &= \frac{\lambda_1}{\mu_2}\pi_1 = \frac{\lambda_1\lambda_0}{\mu_2\mu_1}\pi_0 \\ \pi_3 &= \frac{\lambda_2}{\mu_3}\pi_2 = \frac{\lambda_2\lambda_1\lambda_0}{\mu_3\mu_2\mu_1}\pi_0 \end{aligned}$$

Thus, the general formula for π_j can be proved by induction. Assume π_j has been defined for $j = 1, 2, \ldots$.

$$\pi_j = \frac{\lambda_{j-1}}{\mu_j}\pi_{j-1} = \frac{\lambda_{j-1}\ldots\lambda_1\lambda_0}{\mu_j\ldots\mu_2\mu_1}\pi_0$$

By re-arranging the terms in the above equation, we obtain,

$$\pi_j = \frac{\lambda_{j-1}}{\mu_j}\pi_{j-1} = \pi_0 \prod_{i=0}^{j-1} \frac{\lambda_i}{\mu_{i+1}}$$

Thus, if π_0 is known, the remaining values π_j for $j = 1, 2, \ldots$ may be determined recursively. π_0 can be determined using the law of total probability,

$$\sum_{j=0}^{\infty} \pi_j = 1$$

$$\pi_0 \left(1 + \sum_{j=1}^{\infty} \frac{\pi_j}{\pi_0}\right) = 1$$

It follows that,

$$\pi_0 = \frac{1}{1 + \sum_{j=1}^{\infty} \prod_{i=0}^{j-1} \frac{\lambda_i}{\mu_{i+1}}}$$

A unique positive stationary distribution exists if and only if the summation is positive and finite,

$$0 < \sum_{j=1}^{\infty} \prod_{i=0}^{j-1} \frac{\lambda_i}{\mu_{i+1}} < \infty$$

Note that if $\lambda_i = 0$ for some i and $\mu_i > 0$ for $i \geq 1$, a stationary distribution still exists, but it is not positive. If $\lambda_0 = 0$ and $\mu_i > 0$ for $i \geq 1$, then $\pi_0 = 1$ and $\pi_i = 0$ for $i \geq 1$

References

1. Kijima, M., 1997. Markov processes for stochastic modeling (Vol. 6). CRC Press.
2. Schinazi, R.B., 2014. Classical and spatial stochastic processes: with applications to biology. Springer.
3. Kobayashi, H., Mark, B.L. and Turin, W., 2011. Probability, random processes, and statistical analysis: applications to communications, signal processing, queueing theory and mathematical finance. Cambridge University Press.
4. Osaki, S., 2012. Applied stochastic system modeling. Springer Science & Business Media.
5. Continuous-Time Markov Chains - Introduction, viewed 15 June 2020, <https://mast.queensu.ca/~stat455/lecturenotes/set5.pdf>.