Lecture 4a: Continuous-Time Markov Chain Models

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 Poisson process.

1 Preliminaries

Before discussing Poisson and death and birth processes, we review and take a look at theories and mathematical foundations needed to better understand this topic.

Please note that throughout this lecture, we use the notation $\text{Prob}\{\cdot\}$ in place of $P\{\cdot\}$ to emphasize that a probability is being computed, for example

$$P\{X \leq x\} = \text{Prob}\{X \leq x\} = F(x)\,, \qquad \text{or} \qquad P\{X = x\} = \text{Prob}\{X = x\} = f(x)\,.$$

1.1 The Exponential Distribution

In stochastic modeling, it is often assumed that certain random variables are exponentially distributed. The reason for this is that the exponential distribution is both relatively easy to work with and is often a good approximation to the actual distribution. One of the properties of the exponential distribution is that it does not deteriorate with time. It means that if the lifetime of an item is exponentially distributed, then an item that has been in use for ten (or any number of) hours is as good as a new item in regards to the amount of time remaining until the item fails [7].

Let us consider a continuous random variable X. The random variable is said to have an exponential distribution with parameter λ with $\lambda > 0$, if its probability density function (p.d.f.) is given by

$$f(x) = \begin{cases} \lambda\, e^{-\lambda x} & x \geq 0 \\ 0 & x < 0\,, \end{cases}$$

or, equivalently, if its cumulative distribution function (c.d.f.) is given by

$$F(x) = \int_{-\infty}^{x} f(y)\, dy = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0\,. \end{cases}$$

The expectation or mean of the exponential distribution, $E[X]$ is given by

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x\, f(x)\, dx \\ &= \int_{0}^{\infty} \lambda\, x\, e^{-\lambda x}\, dx\,. \end{aligned}$$

By letting $u = x$, $dv = \lambda\, e^{-\lambda x}$, and using these for integration by parts, we get

$$E[X] = -x\, e^{-\lambda x}\big|_0^\infty + \int_0^\infty e^{-\lambda x}\, dx = \frac{1}{\lambda}\,.$$

If X is exponentially distributed with mean $1/\lambda$, then it can be shown (see its proof in the next subsection) that

$$\text{Prob}\{X > x\} = e^{-\lambda x}\,.$$

The moment generating function (m.g.f.) of the exponential distribution is given by

$$\begin{aligned} \phi(t) &= E[e^{tX}] \\ &= \int_0^\infty e^{tx}\, \lambda\, e^{-\lambda x}\, dx \\ &= \frac{\lambda}{\lambda - t}\,, \qquad \text{for } t < \lambda\,. \end{aligned} \tag{1}$$

All the moments of X can be obtained by differentiating Eq. (1),

$$\begin{aligned} E[X^2] &= \frac{d^2}{dt^2}\phi(t)\bigg|_{t=0} \\ &= \frac{2\lambda}{(\lambda - t)^3}\bigg|_{t=0} \\ &= \frac{2}{\lambda^2}\,, \end{aligned} \tag{2}$$

from which we obtain the variance,

$$\begin{aligned} \mathrm{Var}(X) &= E[X^2] - (E[X])^2 \\ &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} \\ &= \frac{1}{\lambda^2}\,. \end{aligned} \tag{3}$$

1.2 Properties of the Exponential Distribution

1.2.1 Memoryless

The usefulness of exponential random variable derives from the fact that they possess the memoryless property [8], where a random variable X is said to be memoryless, or without memory, if [7]:

$$\mathrm{Prob}\{X > s + t \,|\, X > t\} = \mathrm{Prob}\{X > s\} \qquad \forall\, s, t \geq 0\,. \tag{4}$$

If we think of X as being the lifetime of some instrument, then Eq. (4) describes the probability that the instrument lives for at least $s + t$ hours given that it has survived t hours is the same as the initial probability that it lives for at least s hours [7].

By definition of the conditional probability on the left hand side of Eq. (4), where

$$\mathrm{Prob}\{X > s + t \,|\, X > t\} = \frac{\mathrm{Prob}\{X > s + t\,,\; X > t\}}{\mathrm{Prob}\{X > t\}}\,, \tag{5}$$

and since $\mathrm{Prob}\{X > s+t\}$ is included in $\mathrm{Prob}\{X > t\}$, the condition in Eq. (4) is equivalent to

$$\mathrm{Prob}\{X > s + t \,|\, X > t\} = \frac{\mathrm{Prob}\{X > s + t\}}{\mathrm{Prob}\{X > t\}} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s} = \mathrm{Prob}\{X > s\}\,,$$

or

$$\mathrm{Prob}\{X > s + t\} = \mathrm{Prob}\{X > s\} \cdot \mathrm{Prob}\{X > t\}\,. \tag{6}$$

Since Eq. (6) is satisfied when X is exponentially distributed, such as $e^{-\lambda(s+t)} = e^{-\lambda s} e^{-\lambda t}$, it follows that exponentially distributed random variables are memoryless [7].

1.2.2 Exponential Distribution

Aside from being memoryless, the exponential distribution is also the unique distribution possessing the memoryless property. Let us assume that X is memoryless and let $\bar{F}(x) = \text{Prob}\{X > x\}$. Then by Eq. (6), it follows that [7]

$$\bar{F}(s+t) = \bar{F}(s)\,\bar{F}(t)\,.$$

That is, $\bar{F}(x)$ satisfies the functional equation

$$g(s+t) = g(s)\,g(t)\,.$$

However, the only solutions of the above equation that satisfy any sort of reasonable condition (such as monotonicity, right or left continuity, or even measurability), are of the form [8]

$$g(x) = e^{-\lambda x}\,,$$

which is the only right continuous solution, for some suitable value of λ. For proof of this solution, see [7, 8]. Since a distribution function is always right continuous, we have

$$\bar{F}(x) = e^{-\lambda x}\,,$$

or

$$F(x) = \text{Prob}\{X \leq x\} = 1 - e^{-\lambda x}\,,$$

which shows that the random variable X is exponentially distributed.

Alternatively,

$$\text{Prob}\{X > x\} = 1 - F(x) = e^{-\lambda x}\,. \tag{7}$$

1.2.3 Gamma Distribution

Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed exponential random variables having mean $1/\lambda$. Let us assume that $X_1 + X_2 + \cdots + X_n$ has density given by [7]

$$\begin{aligned} f_{X_1+X_2+\cdots+X_n}(t) &= \int_0^\infty f_{X_n}(t-s)\,f_{X_1+X_2+\cdots+X_{n-1}}(s)\,ds \\ &= \int_0^t \lambda\,e^{-\lambda(t-s)}\,\lambda\,e^{-\lambda s}\,\frac{(\lambda s)^{n-2}}{(n-2)!}\,ds \\ &= \lambda\,e^{-\lambda t}\,\frac{(\lambda t)^{n-1}}{(n-1)!}\,, \end{aligned}$$

for $t \geq 0$, which shows a gamma distribution with parameters n and λ where $\lambda > 0$. Hence the density for $X_1, X_2, \cdots, X_{n-1}$ is given by [7]

$$f_{X_1+X_2+\cdots+X_{n-1}}(t) = \lambda\,e^{-\lambda t}\,\frac{(\lambda t)^{n-2}}{(n-2)!}\,.$$

Another useful calculation is to determine the probability that one exponential random variable is smaller than another. Suppose that X_1 and X_2 are independent exponential random variables with respective means $1/\lambda_1$ and $1/\lambda_2$. Then [7]

$$\begin{aligned}
\text{Prob}\{X_1 < X_2\} &= \int_0^\infty \text{Prob}\{X_1 < X_2 \,|\, X_1 = x\}\, \lambda_1 \, e^{-\lambda_1 x}\, dx \\
&= \int_0^\infty \text{Prob}\{x < X_2\}\, \lambda_1 \, e^{-\lambda_1 x}\, dx \\
&= \int_0^\infty e^{-\lambda_2 x}\, \lambda_1 \, e^{-\lambda_1 x}\, dx \\
&= \int_0^\infty \lambda_1 \, e^{-(\lambda_1+\lambda_2) x}\, dx \\
&= \frac{\lambda_1}{\lambda_1+\lambda_2}\,. \qquad (8)
\end{aligned}$$

Likewise,

$$\text{Prob}\{X_1 > X_2\} = \frac{\lambda_2}{\lambda_1+\lambda_2}\,.$$

1.2.4 Hyperexponential Distribution

Suppose $X_1, X_2, \cdots, X_n$ are independent exponential random variables. If they are identically distributed where the common mean is $1/\lambda$, then the sum $X_1 + X_2 + \cdots + X_n$ follows the gamma distribution with shape parameter n and rate parameter λ.

If the exponential random variables $X_1, X_2, \cdots, X_n$ are not identically distributed, where the respective rates of the random variables are $\lambda_1, \lambda_2, \cdots, \lambda_n$, or the rate parameter of X_i is λ_i, such that $\lambda_i \neq \lambda_j$ for $i \neq j$, then the sum $X_1 + X_2 + \cdots + X_n$ is said to follow a *hyperexponential* distribution.

Let T be independent of these random variables and suppose that

$$\sum_{j=1}^{n} P_j = 1\,, \qquad \text{where} \;\; P_j = \text{Prob}\{T = j\}\,.$$

The random variable X_T is said to be a *hyperexponential* random variable. The hyperexponential distribution is the mixture of a set of independent exponential distributions. To see how such a random variable might originate, imagine that a bin contains n different types of batteries, with a type j battery lasting for an exponential distributed time with rate λ_j, $j = 1, 2, \cdots, n$. Suppose further that P_j is the proportion of batteries in the bin that are type j for each $j = 1, 2, \cdots, n$. If a battery is randomly chosen, in the sense that it is equally likely to be any of the batteries in the bin, then the lifetime of the battery selected will have the hyperexponential distribution [7].

The cumulative distribution function F of $X = X_T$ is obtained by condition on T, following Eq. (7),

$$\begin{aligned} 1 - F(t) &= \text{Prob}\{X > t\} \\ &= \sum_{i=1}^{n} \text{Prob}\{X > t \,|\, T = i\} \cdot \text{Prob}\{T = i\} \\ &= \sum_{i=1}^{n} P_i \, e^{-\lambda_i t} . \end{aligned}$$

Differentiation of the preceding equation yields the probability density function f of X,

$$\begin{aligned} \frac{d}{dt}(1 - F(t)) &= \frac{d}{dt} \sum_{i=1}^{n} P_i \, e^{-\lambda_i t} \\ -f(t) &= \sum_{i=1}^{n} -\lambda_i \, P_i \, e^{-\lambda_i t} \\ f(t) &= \sum_{i=1}^{n} \lambda_i \, P_i \, e^{-\lambda_i t} . \end{aligned} \tag{9}$$

1.2.5 Convolutions of Exponential Random Variables

Let X_i, $i = 1, 2, \cdots, n$ be independent exponentially distributed random variables with respective rates λ_i, $i = 1, 2, \cdots, n$. Suppose that for the rates, $\lambda_i \neq \lambda_j$ for $i \neq j$. The random variable X_i is said to be a *hypoexponential* random variable [7]. To compute its probability density function, let us start with the case $n = 2$,

$$\begin{aligned} f_{X_1+X_2}(t) &= \int_0^t f_{X_1}(s) f_{X_2}(t-s) \, ds \\ &= \int_0^t \lambda_1 e^{-\lambda_1 s} \lambda_2 e^{-\lambda_2 (t-s)} \, ds \\ &= \lambda_1 \lambda_2 e^{-\lambda_2 t} \int_0^t e^{-(\lambda_1 - \lambda_2) s} \, ds \\ &= \frac{\lambda_1}{\lambda_1 - \lambda_2} \lambda_2 e^{-\lambda_2 t} \left(1 - e^{-(\lambda_1 - \lambda_2) t}\right) \\ &= \frac{\lambda_1}{\lambda_1 - \lambda_2} \lambda_2 e^{-\lambda_2 t} + \frac{\lambda_2}{\lambda_2 - \lambda_1} \lambda_1 e^{-\lambda_1 t} . \end{aligned}$$

For the case $n = 3$,

$$\begin{aligned}
f_{X_1+X_2+X_3}(t) &= \int_0^t f_{X_3}(t-s) f_{X_1+X_2}(s)\, ds \\
&= \int_0^t \lambda_3 e^{-\lambda_3(t-s)} \frac{\lambda_1}{\lambda_1-\lambda_2} \lambda_2 e^{-\lambda_2 s} \left(1 - e^{-(\lambda_1-\lambda_2)s}\right) ds \\
&= \frac{\lambda_1\lambda_2\lambda_3}{\lambda_1-\lambda_2} e^{-\lambda_3 t} \int_0^t e^{(\lambda_3-\lambda_2)s}\, ds + \frac{\lambda_1\lambda_2\lambda_3}{\lambda_2-\lambda_1} e^{-\lambda_3 t} \int_0^t e^{(\lambda_3-\lambda_1)s}\, ds \\
&= \frac{\lambda_1\lambda_2\lambda_3}{(\lambda_1-\lambda_2)(\lambda_3-\lambda_2)} e^{-\lambda_2 t} - \frac{\lambda_1\lambda_2\lambda_3}{(\lambda_1-\lambda_2)(\lambda_3-\lambda_2)} e^{-\lambda_3 t} + \\
&\quad \frac{\lambda_1\lambda_2\lambda_3}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)} e^{-\lambda_1 t} - \frac{\lambda_1\lambda_2\lambda_3}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)} e^{-\lambda_3 t} \\
&= \left(\frac{\lambda_2}{\lambda_2-\lambda_1}\right)\left(\frac{\lambda_3}{\lambda_3-\lambda_1}\right) \lambda_1 e^{-\lambda_1 t} + \left(\frac{\lambda_1}{\lambda_1-\lambda_2}\right)\left(\frac{\lambda_3}{\lambda_3-\lambda_2}\right) \lambda_2 e^{-\lambda_2 t} \\
&\quad + \left(\frac{\lambda_1}{\lambda_1-\lambda_3}\right)\left(\frac{\lambda_2}{\lambda_2-\lambda_3}\right) \lambda_3 e^{-\lambda_3 t} \\
&= \sum_{i=1}^{3} \lambda_i e^{-\lambda_i t} \left(\prod_{j\neq i} \frac{\lambda_j}{\lambda_j-\lambda_i}\right).
\end{aligned}$$

In general,

$$f_{X_1+\cdots+X_n}(t) = \sum_{i=1}^{n} \lambda_i e^{-\lambda_i t} \left(\prod_{j\neq i} \frac{\lambda_j}{\lambda_j-\lambda_i}\right).$$

If $T = \sum_{i=1}^n X_i$, then

$$f_T(t) = \sum_{i=1}^{n} C_{i,n}\, \lambda_i\, e^{-\lambda_i t}, \tag{10}$$

where

$$C_{i,n} = \prod_{j\neq i} \frac{\lambda_j}{\lambda_j-\lambda_i}.$$

Integrating both sides of the expression for f_T from t to ∞, see Eq. (9), yields the tail distribution function of T, that is

$$\begin{aligned}
\int_t^\infty f_T(s)ds &= \int_t^\infty \sum_{i=1}^n C_{i,n}\,\lambda_i\,e^{-\lambda_i s}ds \\
1 - F(t) &= \sum_{i=1}^n C_{i,n}e^{-\lambda_i t} \\
\text{Prob}\{T > t\} &= \sum_{i=1}^n C_{i,n}\,e^{-\lambda_i t}\,. \qquad (11)
\end{aligned}$$

1.3 The Law of Total Probability

In probability theory, the law (or formula) of total probability is a fundamental rule relating marginal probabilities to conditional probabilities [11]. The law states that:

> *Given n mutually exclusive events $B_1, B_2, \cdots, B_n$ whose probabilities sum to one, then for any event A we have*
>
> $$\text{Prob}(A) = \text{Prob}(A|B_1)\,\text{Prob}(B_1) + \cdots + \text{Prob}(A|B_n)\,\text{Prob}(B_n)$$
>
> *where Prob$(A|B_i)$ is the conditional probability.*

Example 4a.1

Let X and Y be independent random variables having Poisson distributions with parameters μ and ν, respectively. Calculate the distribution of the sum $X + Y$.

Solution:

Recall that the Poisson distribution with parameter $\lambda > 0$ is given by

$$P_k = \text{Prob}\{X = k\} = \frac{e^{-\lambda}\lambda^k}{k!}\,, \qquad \text{for } \; k = 0, 1, 2, \cdots \qquad (12)$$

The sum $X + Y$ has a Poisson distribution with parameter $\mu + \nu$. By the law of total probability,

$$\begin{aligned}
\text{Prob}\{X + Y = n\} &= \sum_{k=0}^n \text{Prob}\{X = k,\; Y = n - k\} \\
&= \sum_{k=0}^n \text{Prob}\{X = k\}\,\text{Prob}\{Y = n - k\}\,,
\end{aligned}$$

provided that X and Y are independent. Then

$$\begin{aligned}\text{Prob}\{X+Y=n\} &= \sum_{k=0}^{n}\left(\frac{\mu^k e^{-\mu}}{k!}\right)\left(\frac{\nu^{n-k}e^{-\nu}}{(n-k)!}\right) \\ &= \frac{e^{\mu+\nu}}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\mu^k\nu^{n-k}\,. \end{aligned} \tag{13}$$

The binomial expansion of $(\mu+\nu)^n$ is

$$(\mu+\nu)^n = \sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\mu^k\nu^{n-k}\,,$$

so Eq. (13) simplifies to

$$\text{Prob}\{X+Y=n\} = \frac{e^{-(\mu+\nu)}(\mu+\nu)^n}{n!}\,, \qquad n=0,1,2,\cdots$$

the desired Poisson distribution. ■

1.4 Negative Binomial Distribution

The negative binomial distribution is a discrete probability distribution. Assume Bernoulli trials, where:

(1) there are only two possible outcomes,

(2) each trial is independent,

(3) the number of trials is not fixed,

(4) the probability of success p for each trial is constant.

Let X denote the number of trials until the rth success occurs. Then, the probability mass function of X is given by

$$f(x) = \text{Prob}\{X=x\} = \binom{x-1}{r-1}(1-p)^{x-r}p^r\,, \tag{14}$$

for $x = r,\ r+1,\ r+2,\ \cdots$. Hence we say that X follows a negative binomial distribution.

The negative binomial distribution is sometimes defined in terms of failure, *e.g.*, the random variable Y = number of failures before rth success. This formulation is statistically equivalent to the $X =$ number of trials until the rth success occurs, by $Y = X - r$. Hence, the negative binomial can be expressed in its alternative form

$$\text{Prob}\{Y = y\} = \binom{r+y-1}{y}(1-p)^y p^r\,, \qquad y = 0, 1, 2, \cdots \tag{15}$$

The moment generating function of a negative binomial random variable X is

$$M(t) = E[e^{tX}] = \frac{(pe^t)^r}{(1-(1-p)e^t)^r}\,,$$

for $(1-p)e^t < 1$.

The mean of a negative binomial random variable X is

$$\mu = E[X] = \frac{r}{p}\,.$$

The variance of a negative binomial random variable X is

$$\sigma^2 = \text{Var}(x) = \frac{r(1-p)}{p^2}\,.$$

1.5 Properties of the Continuous-Time Markov Chains

Let $\{X(t) : 0 \leq t < \infty\}$ be a collection of discrete random variables, where the possible values of $X(t)$ are the nonnegative integers in a finite, $\{0, 1, 2, \cdots, N\}$, or infinite set, $\{0, 1, 2, \cdots\}$. The index set is continuous, $0 \leq t < \infty$.

Definition 1. *The stochastic process $\{X(t) : 0 \leq t < \infty\}$ is called a continuous-time Markov chain if it satisfies the following condition:*

For any sequence of real numbers satisfying $0 \leq t_0 < t_1 < \cdots < t_n < t_{n+1}$,

$$Prob\{X(t_{n+1}) = i_{n+1} \,|\, X(t_0) = i_0\,, X(t_1) = i_1, \cdots, X(t_n) = i_n\} = Prob\{X(t_{n+1}) = i_{n+1} \,|\, X(t_n) = i_n\}\,. \tag{16}$$

This condition is known as the Markov property. The transition to the future state i_{n+1} at time t_{n+1} depends only on the value of the current state i_n at the current time t_n, does not depend on the past states. Each random variable $X(t)$ has an associated probability distribution $\{P_i(t)\}_{i=0}^{\infty}$, where [1]

$$P_i(t) = \text{Prob}\{X(t) = i\}\,. \tag{17}$$

We shall restrict our attention to the case where $\{X(t)\}$ is a Markov process with *stationary transition probabilities*, which is defined by

$$P_{ij}(t) = \text{Prob}\{X(t+s) = j \,|\, X(s) = i\}\,, \tag{18}$$

for $i, j = 0, 1, 2, \cdots$. The transition probability function in Eq. (47) states that a process presently in state i will be in state j at time $t - s$ later. These transition probabilities are called stationary or homogeneous transition probabilities if they depends only on the length of the time interval $t - s$, not explicitly on t or s. Otherwise, they are referred to as non-stationary or non-homogeneous. The transition probabilities are stationary if

$$\begin{aligned} P_{ij}(t-s) &= \text{Prob}\{X(t) = j \,|\, X(s) = i\} \\ &= \text{Prob}\{X(t-s) = j \,|\, X(0) = i\}\,, \end{aligned}$$

for $t > s$.

Example 4a.2

Taken from [7]:

Suppose that a continuous-time Markov chain enters state i at some time, say, time 0, and suppose that the process does not leave state i (or, a transition does not occur) during the next ten minutes. What is the probability that the process will not leave state i during the following five minutes?

Solution:

Now since the process is in state i at time 10, it follows, by the Markov property, that the probability that it remains in that state during the interval $[10, 15]$ is just the (unconditional) probability that it stays in state i for at least five minutes. That is, if we let T_i denote the amount of time that the process stays in state i before making a transition into a different state, then

$$\text{Prob}\{T_i > 15 \,|\, T_i > 10\} = \text{Prob}\{T_i > 5\}\,,$$

or, recall Eq. (4), in general,

$$\text{Prob}\{T_i > s + t \,|\, T_i > s\} = \text{Prob}\{T_i > t\}\,, \qquad \forall\, s, t \geq 0\,. \qquad \blacksquare$$

The Markov property asserts that $P_{ij}(t)$ satisfies [3, 4, 9]

(a) The probabilities are positive:

$$P_{ij}(t) \geq 0\,.$$

(b) The sum of probabilities that there is a transition from state i to some other state (j) in time $[0, t]$ equals one:

$$\sum_{j=0}^{N} P_{ij} = 1\,, \qquad i, j = 0, 1, 2, \cdots .$$

(c) The transition probabilities $P_{ij}(t)$ are solutions of the Chapman-Kolmogorov equations:

$$P_{ij}(t+s) = \sum_{k=0}^{\infty} P_{ik}(t) P_{kj}(s)\,, \tag{19}$$

which states that in order to move from state i to state j in time $t+s$, $X(t)$ moves to some state k in time t and then from state k to state j in the remaining time s.

(d) In addition, we postulate that

$$\lim_{t \to 0^+} P_{ij}(t) = \begin{cases} 1, & i = j \\ 0, & i \neq j\,. \end{cases}$$

If $\mathbf{P}(t)$ denotes the matrix of transition probabilities or the transition matrix, defined as

$$\mathbf{P}(t) = \left(P_{ij}(t)\right)\,, \tag{20}$$

then property (c) or Eq. (19) can be written compactly in matrix notation as

$$\mathbf{P}(t+s) = \mathbf{P}(t)\mathbf{P}(s)\,, \tag{21}$$

for all $t \geq 0$, $s \geq 0$.

These definitions are the continuous analogues of the definitions given for discrete-time Markov chains [1].

1.6 Limiting Probabilities

Here we present a theorem that enables us to calculate the limit $\lim_{n\to\infty} P_{ij}^{(n)}$ by solving a system of linear equations. The theorem in question is valid if the chain is irreducible and ergodic, as defined below [6].

Consider state i of a Markov chain process that has period d, where $P_{ii}^{n} = 0$ whenever n is not divisible by d, and d is the largest integer. For example, starting in i, it may be possible for the process to enter state i only at the times $2, 4, 6, 8, \cdots$. In this case, state i has period 2. A state with period 1 is said to be *aperiodic*. State i is recurrent, if, starting in i the expected time until the process returns to state i is finite. Then state i is also said to be *positive recurrent*. Positive recurrent, aperiodic states are called *ergodic* [7].

Theorem 1. *The basic limit theorem of Markov chains*

(a) Consider a recurrent, irreducible, aperiodic Markov chain. Let $P_{ii}^{(n)}$ be the probability of entering state i at the nth transition, $n = 0, 1, 2, \cdots$, given that $X_0 = i$ (the initial state is i), thus $P_{ii}^{(0)} = 1$. Let $f_{ii}^{(n)}$ be the probability of first returning to state i at the nth transition, $n = 0, 1, 2, \cdots$, where $f_{ii}^{(0)} = 0$. Then,

$$\lim_{n\to\infty} P_{ii}^{(n)} = \frac{1}{\sum_{n=0}^{\infty} n f_{ii}^{(n)}} = \frac{1}{m_i} \,. \tag{22}$$

(b) Under the same conditions as in (a),

$$\lim_{n\to\infty} P_{ji}^{(n)} = \lim_{n\to\infty} P_{ii}^{(n)} \qquad \textit{for all states } j \,.$$

Theorem 2. *For an irreducible, positive recurrent, and aperiodic (ergodic) Markov chain, $\lim_{n\to\infty} P_{ij}^{(n)}$ exists and is independent of i. Furthermore, letting*

$$\pi_j = \lim_{n\to\infty} P_{ij}^{(n)} \,, \qquad j \geq 0 \,,$$

then π_j is the unique nonnegative solution of

$$\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij} \,, \qquad \textit{for} \quad j \geq 0 \quad \textit{and} \quad \sum_{j=0}^{\infty} \pi_j = 1 \,, \tag{23}$$

where $\pi = (\pi_0, \pi_1, \pi_2, \cdots)$ is a vector.

Example 4a.3

Taken from Example on pp. 173–174 in [10]:

Let the transition probability matrix of a Markov chain be

$$\mathbf{P} = \begin{bmatrix} 0.4 & 0.6 \\ 0.2 & 0.8 \end{bmatrix} .$$

Computing successive powers of $\mathbf{P}$ gives

$$\begin{aligned}
\mathbf{P}^{(2)} &= \begin{bmatrix} 0.28 & 0.72 \\ 0.24 & 0.76 \end{bmatrix} \\
\mathbf{P}^{(4)} &= \begin{bmatrix} 0.2512 & 0.7488 \\ 0.2496 & 0.7504 \end{bmatrix} \\
\mathbf{P}^{(8)} &= \begin{bmatrix} 0.25 & 0.75 \\ 0.25 & 0.75 \end{bmatrix}
\end{aligned}$$

Note that the matrix $\mathbf{P}^{(8)}$ is almost identical to the matrix $\mathbf{P}^{(4)}$, and secondly, that each of the rows of $\mathbf{P}^{(8)}$ has almost identical entries. In terms of matrix elements, it would seem that $P_{ij}^{(n)}$ is converging to some value as $n \to \infty$, that is

$$\lim_{n\to\infty} P_{ij}^{(n)} = \pi_j \,,$$

which is the same for all i. Here, there seems to exist a limiting probability that the process will be in state j after a large number of transitions, and this value is independent of the initial state.

Equivalently,

$$\lim_{n\to\infty} \mathbf{P}^{(n)} = \widehat{\mathbf{P}} \,.$$

Note that in general each row of $\widehat{\mathbf{P}}$ has the same elements,

$$\widehat{\mathbf{P}} = \begin{bmatrix} \pi_0 & \pi_1 & \pi_2 & \cdots \\ \pi_0 & \pi_1 & \pi_2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} ,$$

and that $\widehat{\mathbf{P}}$ is a matrix whose rows are identical and equal to a vector

$$\pi = (\pi_0,\ \pi_1,\ \pi_2,\ \cdots) \,. \qquad \blacksquare$$

The vector π is called a stationary probability distribution.

For general cases, let us first define the following [9].

Definition 2. *Suppose that a transition probability matrix* $\mathbf{P}$ *on a finite number of states labeled* $0, 1, 2, \cdots, N$ *has the property that when raised to some power* n*, the matrix* $\mathbf{P}^n$ *has all of its elements strictly positive. Such a transition probability matrix, or the corresponding Markov chain, is called regular.*

Based on the preceding definition, we proceed to the following theorem [10].

Theorem 3. *Let* $X = \{X_0, X_1, X_2, \cdots\}$ *be a regular Markov chain with a finite number* N *of states and transition probabilitty matrix* $\mathbf{P}$*. Then*

(i) Regardless of the value of $i = 0, 1, 2, \cdots, N$*,*

$$\lim_{n\to\infty} P_{ij}^{(n)} = \pi_j > 0 \,, \qquad j = 0, 1, 2, \cdots, N \,,$$

(ii) *or equivalently, in terms of the Markov chain* $\{X_n\}$,

$$\lim_{n\to\infty} Prob\{X_n = j \,|\, X_0 = i\} = \pi_j > 0\,, \qquad j = 0, 1, 2, \cdots, N\,,$$

(iii) *or equivalently, in matrix notation,*

$$\lim_{n\to\infty} \mathbf{P}^n = \widehat{\mathbf{P}}$$

where $\widehat{\mathbf{P}}$ *is a matrix whose rows are identical and equal to the stationary probability distribution vector*

$$\pi = (\pi_0, \pi_1, \pi_2, \cdots, \pi_N)\ .$$

(iv) *The convergence means that, no matter in which state the chain begins at time* 0 *(or, regardless the initial probability distribution), the probability distribution approaches* π *as* $n \to \infty$.

(v) *The stationary probability distribution* π *is the unique solution of*

$$\mathbf{P}\pi = \pi$$

satisfying $\pi > 0$ *and* $\sum_k \pi_k = 1$.

The limiting distribution is a stationary probability distribution if the continuous-time Markov chain is:

- positive recurrent,
- aperiodic, and
- irreducible.

Theorem 4. *In a positive recurrent aperiodic class with states* $j = 0, 1, 2, \cdots$,

$$\lim_{n\to\infty} P_{ij}^{(n)} = \pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}\,, \qquad \sum_{i=0}^{\infty} \pi_i = 1\,,$$

and the π*'s are uniquely determined by the set of equations*

$$\pi_i \geq 0\,, \quad \sum_{i=0}^{\infty} \pi_i = 1\,, \quad \text{and} \quad \pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij} \quad \text{for } j = 0, 1, 2, \cdots \tag{24}$$

Any set $\{\pi_i\}_{i=0}^{\infty}$ satisfying Eq. (24) is called a *stationary probability distribution* of the Markov chain. A Markov chain that starts according to a stationary probability distribution will follow this distribution at all points of time. Formally, if $\text{Prob}\{X_0 = i\} = \pi$, then $\text{Prob}\{X_n = i\} = \pi_i$ for all $n = 1, 2, \cdots$ [9]. For example, for the case $n = 1$,

$$\begin{aligned}\text{Prob}\{X_1 = i\} &= \sum_{k=0}^{\infty} \text{Prob}\{X_0 = k\}\,\text{Prob}\{X_1 = i \,|\, X_0 = k\} \\ &= \sum_{k=0}^{\infty} \pi_k P_{ki} \\ &= \pi_i\,.\end{aligned}$$

Example 4a.4

Taken from Example on pp. 176–177 in [10]:

Genes can be changed by certain stimuli, such as radiation. Sometimes in the 'natural' course of events, chemical accidents may occur which change one allele to another. Such alteration of genetic material is called mutation. Suppose that each A_1 allele mutates to A_2 with probability α_1 per generation and that each A_2 allele mutates to A_1 with probability α_2 per generation. By considering the ways in which a gene may be of type A_1 we see that

$$\begin{aligned}\text{Prob}\{\text{a gene is } A_1 \text{ after mutation}\} &= \text{Prob}\{\text{gene is } A_1 \text{ before mutation}\} \\ &\quad \times \text{Prob}\{\text{gene does not mutate from } A_1 \text{ to } A_2\} \\ &\quad + \text{Prob}\{\text{gene is } A_2 \text{ before mutation}\} \\ &\quad \times \text{Prob}\{\text{gene mutates from } A_2 \text{ to } A_1\}\,.\end{aligned}$$

Thus, if there were j genes of type A_1 in the parent population before mutation, the probability p_j of choosing an A_1 gene when we form the next generation is

$$p_j = \frac{j}{2N}(1 - \alpha_1) + \left(1 - \frac{j}{2N}\right)\alpha_2\,. \tag{25}$$

Let X_n be the number of A_1 alleles at generation n. Suppose that there were j genes of type A_1 in generation n. We choose $2N$ genes to form generation $n + 1$ with probability p_j of an A_1 and probability $1 - p_j$ of an A_2 at each trial. Then $\{X_n,\ n = 0, 1, 2, \cdots\}$ is a temporarily homogeneous Markov chain with one-step transition probabilities

$$\begin{aligned}P_{jk} &= \text{Prob}\{X_{n+1} = k \,|\, X_n = j\} \\ &= \binom{2N}{k} p_j^k (1 - p_j)^{2N-k}\,, \qquad j, k = 0, 1, 2, \cdots, 2N,\end{aligned} \tag{26}$$

with p give by Eq. (25). Now if we let $N = 1$, $\alpha_1 = \alpha_2 = \frac{1}{4}$, then substituting in Eq. (25),

$$p_j = \frac{j}{4} + \frac{1}{4}\,.$$

This gives $p_0 = \frac{1}{4}$, $p_1 = \frac{1}{2}$, $p_2 = \frac{3}{4}$. The elements of $\mathbf{P}$ are, for $k = 0, 1, 2$,

$$\begin{aligned} p_{0k} &= \binom{2}{k}\left(\frac{1}{4}\right)^k\left(\frac{3}{4}\right)^{2-k} \\ p_{1k} &= \binom{2}{k}\left(\frac{1}{2}\right)^2 \\ p_{2k} &= \binom{2}{k}\left(\frac{3}{4}\right)^k\left(\frac{1}{4}\right)^{2-k}\,, \end{aligned}$$

obtained from Eq. (26). Evaluating these we get

$$\mathbf{P} = \frac{1}{16}\begin{bmatrix} 9 & 6 & 1 \\ 4 & 8 & 4 \\ 1 & 6 & 9 \end{bmatrix}.$$

All states of the Markov chain communicate with each other, thus there is one class and the chain is irreducible. The chain has no absorbing states as there are no ones on the principal diagonal. All elements of $\mathbf{P}$ are non-zero, so at any time step a transition is possible from any state to another.

Computing successive powers of $\mathbf{P}$ results in

$$\lim_{n\to\infty} \mathbf{P}^n = \begin{bmatrix} 0.285 & 0.428 & 0.285 \\ 0.285 & 0.428 & 0.285 \\ 0.285 & 0.428 & 0.285 \end{bmatrix},$$

where the rows are identical and equal to

$$\pi = (0.285 \;\; 0.428 \;\; 0.285)\,,$$

which is the stationary probability distribution.

We can also find π by solving $\mathbf{P}\pi = \pi$, or

$$\begin{bmatrix}\pi_0 \; \pi_1 \; \pi_2\end{bmatrix}\begin{bmatrix} 9 & 6 & 1 \\ 4 & 8 & 4 \\ 1 & 6 & 9 \end{bmatrix} = 16\begin{bmatrix}\pi_0 \; \pi_1 \; \pi_2\end{bmatrix}$$

gives three equations, where the first two equations are

$$\begin{aligned} 9\pi_0 + 4\pi_1 + \pi_2 &= 16\pi_0 \\ 6\pi_0 + 8\pi_1 + 6\pi_2 &= 16\pi_1 \end{aligned}$$

or

$$\begin{aligned} 7\pi_0 - 4\pi_1 - \pi_2 &= 0 \\ -6\pi_0 + 8\pi_1 - 6\pi_2 &= 0\,. \end{aligned}$$

Since of the components of π is arbitrary, setting $\pi_2 = 1$ yields $\pi_0 = 1$ and $\pi_1 = \frac{3}{2}$. The sum of the components of π is $\frac{7}{2}$, so dividing each element of π by $\frac{7}{2}$ gives us the required stationary probability vector

$$\pi = (0.285 \;\; 0.428 \;\; 0.285)\,. \qquad \blacksquare$$

Example 4a.5

Taken from Example on pp. 177–178 in [10], which shows that a stationary distribution may exist but it does not imply that a steady-state is approached as $n \to \infty$.

Consider a Markov chain with two states 0 and 1, that is

$$\mathbf{P} = \begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

so that transitions are only possible from one state to the other. Solving

$$\mathbf{P}\pi = \pi$$

or,

$$\begin{bmatrix} \pi_0 \; \pi_1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \pi_0 \; \pi_1 \end{bmatrix}$$

gives $\pi_0 = \pi_1$. If we set $\pi_0 = 1$, then $\pi_1 = 1$. Since the sum of the components $(\pi_0 + \pi_1)$ is 2, so dividing each element of π by 2 we obtain the stationary probability distribution vector

$$\pi = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

However, as $n \to \infty$, $\mathbf{P}^n$ does not approach a constant matrix, because

$$\mathbf{P}^n = \begin{cases} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, & n = 1, 3, 5, \cdots \\[2ex] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, & n = 2, 4, 6, \cdots \end{cases}$$

It can be seen that state 1 can only be entered from state 0 ($P_{01} \neq 0$) and vice versa ($P_{10} \neq 0$) on time steps $1, 3, 5, \cdots$. This Markov chain is not regular and periodic with period 2. Since state 0 and state 1 communicate with each other or $0 \leftrightarrow 1$, then there is only one class $\{0, 1\}$ and the chain is irreducible. From the recurrence theory in Lecture 3, we see that states 0 and 1 are recurrent, because

$$\sum_{n=1}^{\infty} f_{00}^{(n)} = P_{00} + P_{01} = 0 + 1 = 1\,,$$

and

$$\sum_{n=1}^{\infty} f_{11}^{(n)} = P_{11} + P_{10} = 0 + 1 = 1\,. \qquad \blacksquare$$

If a Markov chain is reducible and/or periodic, then there can be many stationary distributions, as in the preceding example.

In addition to being the limiting probabilities, the π_j's also represent the proportion of time that the process spends in state j, over a long period of time. That is, if the chain is positive recurrent but periodic, then this is the only interpretation of the π_j's [6]. On average, the process spends one time unit in state j for μ_j time units, from which

$$\pi_j = \frac{1}{\mu_j}\,,$$

where

$$\mu_j = \sum_{n=1}^{\infty} n\, f_{jj}^{(n)}\,,$$

which is the mean (average number) of transitions needed by the process, starting from state j, to return to j for the first time [6], or also known as *first return probabilities* (see Lecture 3).

1.7 Counting Process

Let $\{X(t) : t \geq 0\}$ be a stochastic process. If $X(t)$ represents the total number of "events" that occur by time t, the stochastic process is said to be a *counting process*. Some examples of counting processes are as follows [7]:

(1) In a store, an event occurs whenever a customer enters the store. If we let $X(t)$ equal the total number of customers who enter the store at or prior to time t, then $\{X(t), t \geq 0\}$ is a counting process. However, if we let $X(t)$ equal the number of customers who are already in the store at time t, then $\{X(t), t \geq 0\}$ is not a counting process.

(2) If we say that an event occurs whenever a child is born, then $\{X(t), t \geq 0\}$ is a counting process when $X(t)$ equals the total number of babies who were born by time t.

(3) In a soccer/football game, an event occurs whenever a given soccer player scores a goal. If $X(t)$ equals the number of goals that the soccer player scores by time t, then $\{X(t), t \geq 0\}$ is a counting process.

From its definition, a counting process $X(t)$ must satisfy the following:

(i) $X(t) \geq 0$.

(ii) $X(t)$ is integer valued.

(iii) If $s < t$, then $X(s) \leq X(t)$.

(iv) For $s < t$, $X(t) - X(s)$ equals the number of events that occur in the interval $(s, t]$.

A counting process is said to possess:

- *independent increments* if the numbers of events that occur in disjoint time intervals are independent. For example, in the customer arrival counting process of example (1), the number of events that occur by time $t = 10$, that is, $X(10)$, must be independent of the number of events that occur between times $t = 10$ and $t = 15$, that is, $X(15) - X(10)$. Also, the number of customers arriving during one time interval, *e.g.*, $X(25) - X(20)$ does not affect the number arriving during a different time interval, *e.g.*, $X(15) - X(10)$.

  Definition 3. *Let $\{X(t),\ t \in [0, \infty)\}$ be a continuous-time random process. We say that $X(t)$ has* independent increments *if, for all $0 \leq t_1 < t_2 < \cdots < t_n$, the random variables*

  $$X(t_2) - X(t_1)\,,\ X(t_3) - X(t_2)\,,\ \cdots, X(t_n) - X(t_{n-1})$$

  are independent [13].

- *stationary increments* if the distribution of the number of events that occur in any time interval depends only on the length of the time interval. In other words, a counting process is said to have stationary increments if, for all $t_2 > t_1 \geq 0$,

$$X(t_2) - X(t_1) \quad \text{has the same distribution as} \quad X(t_2 - t_1)\,.$$

2 The Poisson Process

One of the most important counting processes is the Poisson process.

Definition 4. *The counting process $\{X(t), t \geq 0\}$ is said to be a Poisson process having rate λ, where $\lambda > 0$, if [7],*

(i) For $t = 0$, $X(0) = 0$.

(ii) The process has independent increments.

(iii) The number of events in any interval of length t is Poisson distributed with mean λt. That is, for all s, $t \geq 0$:

$$Prob\{X(t+s) - X(s) = n\} = e^{-\lambda t}\frac{(\lambda t)^n}{n!}\,, \qquad n = 0, 1, 2, \cdots \tag{27}$$

Related to the customer arrival counting process of example (1) above, we assume the following about the rate at which customers arrive [5]:

(1) The number of customers arriving during one time interval does not affect the number of customers arriving during a different time interval $\rightarrow$ independent increments

(2) The "average" rate at which customers arrive remains constant.

(3) Customers arrive one at a time.

The first assumption implies that for $0 \leq t_1 < t_2 < t_3 < \cdots < t_n$, the random variables $X(t_2) - X(t_1), \cdots, X(t_n) - X(t_{n-1})$ are independent.

By the second assumption, with λ is the rate at which customers arrive, then the process has stationary increments. In other words, on average we expect λt customers in time t, hence

$$E[X(t)] = \lambda\, t\,,$$

which explains why λ is called the rate of the process.

The rate λ in a Poisson process $X(t)$ is the proportionality constant in the probability of an event occurring during an arbitrarily small interval [9].

In a small time interval $[t, t + \Delta t]$, we expect that customers arrive one at a time, and a new customer arrives with probability about $\lambda \Delta t$. Mathematically, this can be precisely explained from Eq. (27), with $n = 1$ and using the Maclaurin's (or power) series for the expansion of $e^{-\lambda \Delta t}$,

$$\begin{aligned} \text{Prob}\{X(t+\Delta t) - X(t) = 1\} &= e^{-\lambda \Delta t} \frac{(\lambda \Delta t)}{1!} \\ &= (\lambda \Delta t) \left(1 - \lambda \Delta t + \frac{1}{2} \lambda^2 \Delta t^2 - \cdots \right) \\ &= \lambda \Delta t - \lambda^2 \Delta t^2 + \frac{1}{2} \lambda^3 \Delta t^3 - \cdots \\ &= \lambda \Delta t + o(\Delta t) \,, \end{aligned}$$

where the terms with Δt^k, $k \geq 2$ are very small and all very small terms are group together in the notation $o(\Delta t)$. Note that the formula of Maclaurin's series for the expansion of $f(\Delta t) = e^{-\lambda \Delta t}$ around $\Delta t = 0$ is given by

$$f(0) = f(0) + \Delta t f'(0) + \frac{\Delta t^2}{2!} f''(0) + \frac{\Delta t^3}{3!} f'''(0) + \cdots$$

with $f(0) = e^0 = 1$, $f'(0) = -\lambda$, $f''(0) = \lambda^2$, and so on.

If within the small time interval there is no new customers arrive, $n = 0$, then the probability is mathematically explained as,

$$\begin{aligned} \text{Prob}\{X(t+\Delta t) - X(t) = 0\} &= e^{-\lambda \Delta t} \frac{(\lambda \Delta t)^0}{0!} \\ &= e^{-\lambda \Delta t} \\ &= 1 - \lambda \Delta t + \frac{1}{2} \lambda^2 \Delta t^2 - \cdots \\ &= 1 - \lambda \Delta t + o(\Delta t) \,. \end{aligned}$$

The third assumption states that the probability that more than one customer arrives during a small time interval is significantly small. Mathematically, using Eq. (27) with $n = 2$, it also results in terms with Δt^2 and even smaller, hence the transition probability only equals $o(\Delta t)$,

$$\begin{aligned} \text{Prob}\{X(t+\Delta t) - X(t) = 2\} &= e^{-\lambda \Delta t} \frac{(\lambda \Delta t)^2}{2!} \\ &= \frac{1}{2!} (\lambda \Delta t)^2 \left(1 - \lambda \Delta t + \frac{1}{2} \lambda^2 \Delta t^2 - \cdots \right) \\ &= o(\Delta t) \,. \end{aligned}$$

Obviously, equations with $n = 3, 4, \cdots$ also result in only very small terms, hence it is appropriate to write the transition probabilities as

$$\text{Prob}\{X(t + \Delta t) - X(t) \geq 2\} = o(\Delta t)\,.$$

All these lead to an alternate definition of a Poisson process for phenomena with infinitesimal probabilities.

Definition 5. *The counting process $\{X(t), t \geq 0\}$ is said to be a Poisson process having rate λ where $\lambda > 0$, if [7]*

(i) For $t = 0$, $X(0) = 0$.

(ii) The process has stationary and independent increments.

(iii) For Δt sufficiently small, the transition probabilities are [7]

$$\begin{aligned}
P_{ii}(\Delta t) &= Prob\{X(t + \Delta t) - X(t) = 0\} &&= 1 - \lambda\Delta t + o(\Delta t)\,, \\
P_{i,i+1}(\Delta t) &= Prob\{X(t + \Delta t) - X(t) = 1\} &&= \lambda\Delta t + o(\Delta t)\,, \\
P_{ij}(\Delta t) &= Prob\{X(t + \Delta t) - X(t) \geq 2\} &&= o(\Delta t)\,,
\end{aligned} \tag{28}$$

where the notation $o(\Delta t)$ represents some function that is much smaller than Δt for small Δt.

Another way of writing these transition probabilities is:

$$\begin{aligned}
P_{ii}(\Delta t) &= \text{Prob}\{X(t + \Delta t) = i \,|\, X(t) = i\} &&= 1 - \lambda\Delta t + o(\Delta t)\,, \\
P_{i,i+1}(\Delta t) &= \text{Prob}\{X(t + \Delta t) = i + 1 \,|\, X(t) = i\} &&= \lambda\Delta t + o(\Delta t)\,, \\
P_{ij}(\Delta t) &= \text{Prob}\{X(t + \Delta t) = j \,|\, X(t) = i\} &&= o(\Delta t)\,, \quad \text{if } j \geq i + 2 \\
P_{ij}(\Delta t) &= 0\,, \quad \text{if } j < i
\end{aligned} \tag{29}$$

The transition probabilities in Eq. (29) are obtained after using the conditional probabilities and making use of independent increments. For example,

$$\begin{aligned}
\text{Prob}\{X(t + \Delta t) = i \,|\, X(t) = i\} &= \frac{\text{Prob}\{X(t + \Delta t) = i\,,\ X(t) = i\}}{\text{Prob}\{X(t) = i\}} \\
&= \text{Prob}\{X(t + \Delta t) = i\} \\
&= \text{Prob}\{X(t + \Delta t) = X(t)\}\,, \quad \text{by the condition } \ i = X(t) \\
&= \text{Prob}\{X(t + \Delta t) - X(t) = 0\}\,.
\end{aligned}$$

Likewise, for the other two equations.

Definition 6. *The function $f(\Delta t)$ is said to be $o(\Delta t)$, or $f(\Delta t) = o(\Delta t)$, as $\Delta t \to 0$ if*

$$\lim_{\Delta t \to 0} \frac{f(\Delta t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0\,.$$

The following are examples and properties of $o(\Delta t)$ [7]:

(i) The function $f(x) = x^2$ is $o(\Delta t)$ since

$$\lim_{\Delta t \to 0} \frac{f(\Delta t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{(\Delta t)^2}{\Delta t} = \lim_{\Delta t \to 0} \Delta t = 0$$

(ii) The function $f(x) = x$ is not $o(\Delta t)$ since

$$\lim_{\Delta t \to 0} \frac{f(\Delta t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{\Delta t}{\Delta t} = \lim_{\Delta t \to 0} 1 = 1 \neq 0$$

(iii) If $f(\cdot)$ is $o(\Delta t)$ and $g(\cdot)$ is $o(\Delta t)$, then so is $f(\cdot) + g(\cdot)$, since

$$\lim_{\Delta t \to 0} \frac{f(\Delta t) + g(\Delta t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{f(\Delta t)}{\Delta t} + \lim_{\Delta t \to 0} \frac{g(\Delta t)}{\Delta t} = 0 + 0 = 0$$

(iv) If $f(\cdot)$ is $o(\Delta t)$, then so is $g(\cdot) = c\,f(\cdot)$, since

$$\lim_{\Delta t \to 0} \frac{c\,f(\Delta t)}{\Delta t} = c \lim_{\Delta t \to 0} \frac{f(\Delta t)}{\Delta t} = c \cdot 0 = 0$$

(v) From (iii) and (iv), it follows that any finite linear combination of functions, each of which is $o(\Delta t)$, is $o(\Delta t)$.

The transition probabilities in Eq. (28) are $o(\Delta t)$ as $\Delta t \to 0$, in particular

$$\begin{aligned}
\lim_{\Delta t \to 0} \frac{P_{i,i+1}(\Delta t) - \lambda \Delta t}{\Delta t} &= \lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0 \\
\lim_{\Delta t \to 0} \frac{P_{ii}(\Delta t) - 1 + \lambda \Delta t}{\Delta t} &= \lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0 \\
\lim_{\Delta t \to 0} \frac{P_{ij}(\Delta t)}{\Delta t} &= \lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0\,, \quad j \geq i + 2\,.
\end{aligned}$$

In order for the function $f(\cdot)$ to be $o(\Delta t)$, it is necessary that $\frac{f(\Delta t)}{\Delta t}$ go to zero as Δt goes to zero. But if Δt goes to zero, the only way for $\frac{f(\Delta t)}{\Delta t}$ to go to zero is for $f(\Delta t)$ to go to zero faster than Δt does. That is, for Δt small, $f(\Delta t)$ must be small compared with Δt [7].

In a small time interval Δt, the Poisson process can either stay in the same state or move to the next larger state, $i \to i+1$; it cannot move to a smaller state. Note that the transition probabilities $P_{ij}(\Delta t)$ are independent of i and j and only depend on the length of the interval Δt, see Eq. (28). For the Poisson process to have independent and stationary increments, the intervals $[s, s+\Delta t]$ and $[t, t+\Delta t]$ must be non-overlapping, that is $s + \Delta t \leq t$. Then these random variables from the Poisson process, $X(t+\Delta t) - X(t)$ and $X(s+\Delta t) - X(s)$ are independent and have the same probability distributions (*i.e.*, the Poisson process has stationary and independent increments) [1].

Now we derive a system of differential equations for $P_i(t)$ for $i = 0, 1, 2, \cdots$. Note that in the Poisson process $X(0) = 0$:

- For $\mathbf{i} = \mathbf{0}$:
  From the definition of a probability distribution in Eq. (17), that is

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

then

$$P_0(t+\Delta t) = \text{Prob}\{X(t+\Delta t) = 0\}$$

and by

$$X(t+\Delta t) = X(t) - X(0) + X(t+\Delta t) - X(t)$$

where $X(t) - X(0) = X(t)$ and $X(t+\Delta t) - X(t)$ are independent, and if we let $P_{00}(t+\Delta t) = P_0(t+\Delta t)$, by the law of total probability

$$\begin{aligned} P_0(t+\Delta t) &= \text{Prob}\{X(t+\Delta t) = 0\} \\ &= \text{Prob}\{X(t) = 0, X(t+\Delta t) - X(t) = 0\} \\ &= \text{Prob}\{X(t) = 0\} \cdot \text{Prob}\{X(t+\Delta t) - X(t) = 0\} \\ &= \text{Prob}\{X(t) = 0\} \cdot \text{Prob}\{X(\Delta t) = 0\}\,. \end{aligned} \tag{30}$$

Due to $P_i(t) = \text{Prob}\{X(t) = i\}$, then Eq. (30) becomes

$$P_0(t+\Delta t) = P_0(t) \cdot P_0(\Delta t)\,. \tag{31}$$

Since

$$\begin{aligned} P_0(\Delta t) &= \text{Prob}\{X(\Delta t) = 0\} \\ &= \text{Prob}\{X(t+\Delta t) - X(t) = 0\} \\ &= \text{Prob}\{X(t+\Delta t) = X(t)\} \\ &= 1 - \lambda\Delta t + o(\Delta t)\,, \end{aligned}$$

which, after being substituted into Eq. (31)

$$P_0(t + \Delta t) = P_0(t)\big[1 - \lambda \Delta t + o(\Delta t)\big] \, .$$

Subtracting $P_0(t)$ from both sides of the equation and dividing by Δt yields

$$\frac{P_0(t + \Delta t) - P_0(t)}{\Delta t} = -\lambda \, P_0(t) + P_0(t) \, \frac{o(\Delta t)}{\Delta t} \, .$$

Taking the limit as $\Delta t \to 0$,

$$\lim_{\Delta t \to 0} \frac{P_0(t + \Delta t) - P_0(t)}{\Delta t} = - \lim_{\Delta t \to 0} \lambda \, P_0(t) + \underbrace{\lim_{\Delta t \to 0} P_0(t) \, \frac{o(\Delta t)}{\Delta t}}_{=0} \, ,$$

gives us the following first-order linear differential equation for $i = 0$:

$$\frac{dP_0(t)}{dt} = -\lambda \, P_0(t) \, . \tag{32}$$

The general solution of the differential equation is

$$P_0(t) = c_0 \, e^{-\lambda t} \, ,$$

where c_0 is a constant determined from the equation's initial condition.

Given the initial condition $P_0(0) = \text{Prob}\{X(0) = 0\} = 1$, the particular solution of the differential equation is

$$P_0(t) = e^{-\lambda t} \, .$$

- For $\mathbf{i \geq 1}$:
  Similarly, by the law of total probability,

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

However, by the last definition of Eq. (28), the last term on the right hand side is $o(\Delta t)$, since

$$\begin{aligned}
\text{Prob}\{X(t + \Delta t) = i, X(t + \Delta t) - X(t) \geq 2\} &= \text{Prob}\{X(t + \Delta t) = i\} \cdot \\
&\quad\; \text{Prob}\{X(t + \Delta t) - X(t) \geq 2\} \\
&= \text{Prob}\{X(t + \Delta t) = i\} \cdot o(\Delta t) \\
&= P_i(t + \Delta t) \cdot o(\Delta t) \\
&= o(\Delta t)
\end{aligned}$$

Applying the definition of the transition probabilities of Eq. (28) and the independence of the increments,

$$\begin{aligned} P_i(t+\Delta t) &= \text{Prob}\{X(t)=i\}\cdot\text{Prob}\{X(t+\Delta t)-X(t)=0\}+ \\ &\quad \text{Prob}\{X(t)=i-1\}\cdot\text{Prob}\{X(t+\Delta t)-X(t)=1\}+o(\Delta t) \\ &= P_i(t)\cdot P_0(\Delta t)+P_{i-1}(t)\cdot P_{i,i+1}(\Delta t)+o(\Delta t) \\ &= P_i(t)\big[1-\lambda\Delta t+o(\Delta t)\big]+P_{i-1}(t)\big[\lambda\Delta t+o(\Delta t)\big]+o(\Delta t) \end{aligned}$$

Rearranging

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

where all terms with a factor of $o(\Delta t)$ are put in the $o(\Delta t)$ expression, and dividing both sides by Δt leads to

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

Taking the limit as $\Delta t\to 0$ gives us the following system of difference-differential equations for $i\geq 1$:

$$\frac{dP_i(t)}{dt} = -\lambda P_i(t)+\lambda P_{i-1}(t)\,. \tag{33}$$

where, difference equations in the variable i and differential equations in the variable t.

If we multiply both sides of Eq. (33) by $e^{\lambda t}$ (known as an *integrating factor*) and equate $P_i(t)$ terms to the left, we get

$$e^{\lambda t}\left(\frac{dP_i(t)}{dt}+\lambda P_i(t)\right) = \lambda\, e^{\lambda t}\, P_{i-1}(t)\,,$$

which is equivalent to

$$\frac{d}{dt}\left(e^{\lambda t}\, P_i(t)\right) = \lambda\, e^{\lambda t}\, P_{i-1}(t)\,.$$

Now if $i=1$, using the solution $P_0(t)=e^{-\lambda t}$ for $i=0$ above, we get

$$\frac{d}{dt}\left(e^{\lambda t}\, P_1(t)\right) = \lambda\,.$$

Integrating both sides from 0 to t, yields the general solution

$$P_1(t) = e^{-\lambda t}(\lambda t+c)\,.$$

Applying the initial conditions $P_i(0) = 0$ for $i \geq 1$ gives $c = 0$, and thus the particular solution is

$$P_1(t) = \lambda t e^{-\lambda t}.$$

Next for $i = 2$, the differential equation is

$$\frac{dP_2(t)}{dt} = -\lambda P_2(t) + \lambda P_1(t) = -\lambda P_2(t) + \lambda^2 t e^{-\lambda t}$$

with the initial condition $P_2(0) = 0$. Applying the same technique to $P_2(t)$ as for $P_1(t)$ yields the solution

$$P_2(t) = (\lambda t)^2 \frac{e^{-\lambda t}}{2!}.$$

By induction, it can be shown that

$$P_i(t) = e^{-\lambda t} \frac{(\lambda t)^i}{i!}, \qquad i = 0, 1, 2, \cdots \tag{34}$$

Thus, Definition 5 implies Definition 4.

The probabilities $P_i(t)$ plotted against successive states $i = 0, 1, 2, \cdots, 50$ with real-valued times $t = 1.1$ (blue), $t = 2.5$ (magenta), $t = 3.8$ (dark green), and $t = 4.6$ (yellow) are shown in Fig. 1, where $\lambda = 1.0$ (left) and $\lambda = 3.5$ (right). Observe that the probabilities decrease as time progresses, also as λ is increased. The same results of probabilities $P_i(t)$ plotted against i show the same pattern as in Fig. 2.

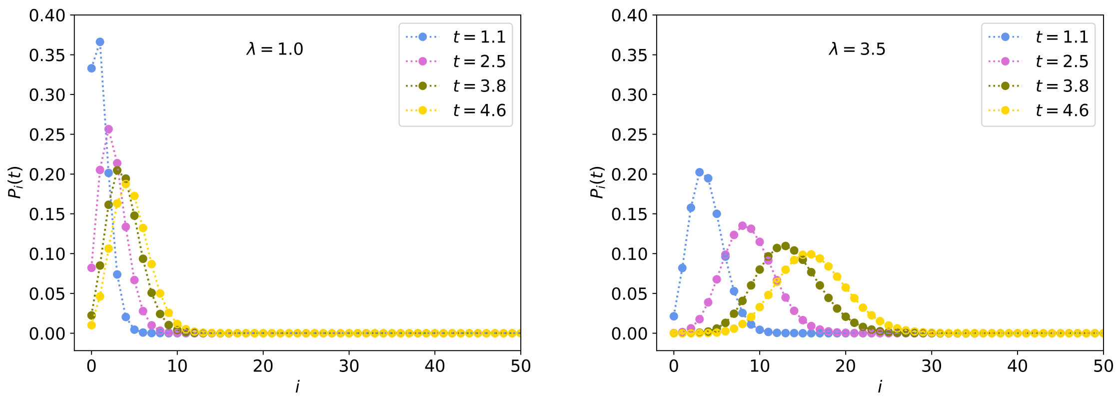

Figure 1: The probability distributions of $X(t)$ for the Poisson process from Eq. (34) with $t = 1.1$ (blue), $t = 2.5$ (magenta), $t = 3.8$ (dark green), and $t = 4.6$ (yellow) at low $\lambda = 1.0$ (left) and higher $\lambda = 3.5$ (right).

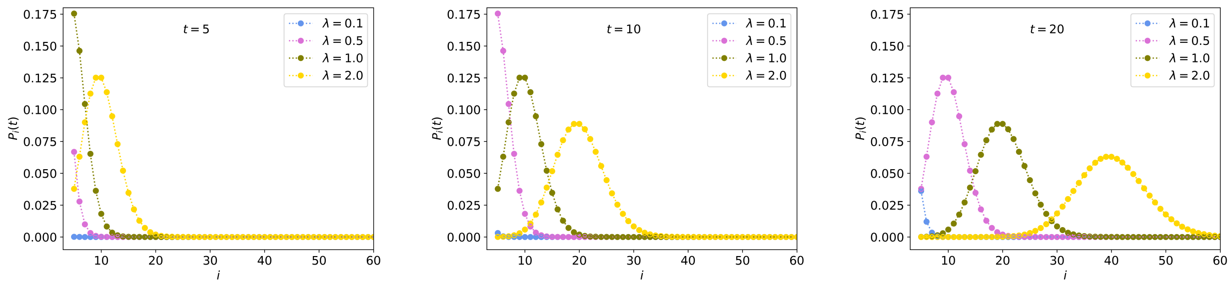

Figure 2: The probability distributions of $X(t)$ for the Poisson process from Eq. (34) with $\lambda = 0.1$ (blue), $\lambda = 0.5$ (magenta), $\lambda = 1.0$ (dark green), and $\lambda = 2.0$ (yellow) at $t = 5$ (left), $t = 10$ (middle), and $t = 20$ (right).

2.1 The Poisson Point Process

The Poisson point process is the simplest and most ubiquitous example of a point process. This process often arises in a form where the time parameter is replaced by a suitable spatial parameter. The Poisson point process is also said to be a spatial generalization of the Poisson process. A Poisson (counting) process on the line can be characterized by two properties: the number of points (or events) in disjoint intervals are independent and it has a Poisson distribution [12].

Consider events occurring along the positive axis $[0, \infty)$, as shown in Fig. 3.

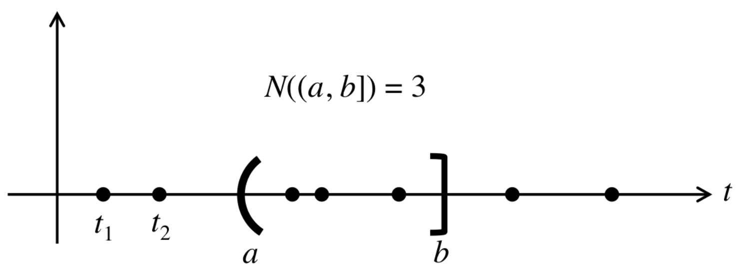

Figure 3: A Poisson point process.

Let $N((a, b])$ denote the number of events that occur during the interval $(a, b]$. That is, if $t_1 < t_2 < t_3 < \cdots$ denote the times (or locations, etc.) of successive events, then $N((a, b])$ is the number of values t_i for which $a < t_i \leq b$. In other words, the $N((a, b])$ is a *Poisson point process* that counts the number of events occurring in an interval $(a, b]$. A Poisson process $X(t)$ counts the number of events occurring up to time t, or formally $X(t) = N((0, t])$.

Definition 7. *Let $N((s, t])$ be a random variable counting the number of events occurring in an interval $(s, t]$. Then $N((s, t])$ is a Poisson point process of intensity $\lambda > 0$ if* *(i) for every $m = 2, 3, \cdots$ and distinct time points $t_0 = 0 < t_1 < t_2 < \cdots < t_m$, the random variables*

$$N((t_0, t_1]),\ N((t_1, t_2]),\ \cdots\ N((t_{m-1}, t_m])$$

are independent,

(ii) for any times $s < t$, the random variable $N((s, t])$ has the Poisson distribution

$$\text{Prob}\{N((s,t]) = k\} = \frac{\left(\lambda(t-s)\right)^k\, e^{-\lambda(t-s)}}{k!}\,, \qquad k = 0, 1, 2, \cdots$$

2.2 Interevent and Waiting Time Distributions

Another way to view the Poisson process is to consider the waiting times between customers. Let T_i denote the time between the arrivals of the $(i-1)$st and the ith customers, or the elapsed time between the $(i-1)$st and the ith events. The sequence $\{T_i,\ i = 1, 2, \cdots\}$ is called the sequence of the *interevent times* or *sojourn times* or *holding times*. For example, if $T_1 = 5$ and $T_2 = 10$, then the first event of the Poisson process would have occurred at time 5 and the second event at time 15 [7].

The fundamental difference between a discrete-time Markov chain and a continuous-time Markov chain is that in discrete-time Markov chains, there is a "jump" to a new state at discrete times $1, 2, \cdots$, but in continuous-time Markov chains, the "jump" to a new state may occur at any real-valued time $t \geq 0$ [1]. In a continuous-time chain, beginning at state $X(0)$, the process stays in state $X(0)$ for a random amount of time W_1 until it jumps to a new state, $X(W_1)$. Then it stays in state $X(W_1)$ for a random amount of time until it jumps to a new state at time W_2, $X(W_2)$, and so on. In general, W_i is the random variable for the time of the ith jump [1].

Fig. 4 shows a typical sample path of a Poisson process, where the random variable W_i is referred to as the time of occurrence of the ith event, or the so-called *jump time* or *waiting time* of the process. It is often convenient to set $W_0 = 0$. The differences $T_i = W_{i+1} - W_i$ are the ones that are referred to as the interevent/holding/sojourn times. The random variable T_i measures the duration that the Poisson process sojourns in state i.

The random variables T_i should be independent and identically distributed. The T_i should also satisfy the loss of memory property, that is, if we have waited s time units for a customer and no one has arrived, the chance that a customer will come in the next t time units is exactly the same as if there had been some customers before [5]. Mathematically, the property is written

$$\text{Prob}\{T_i \geq s + t \,|\, T_i \geq s\} = \text{Prob}\{T_i \geq t\}\,.$$

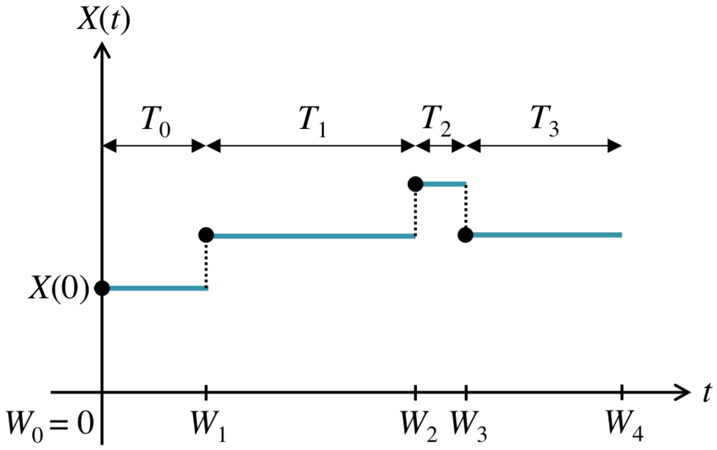

Figure 4: Sample path of a continuous-time Markov chain, illustrating waiting times (W_1, W_2, W_3, and W_4) and interevent/sojourn times (T_0, T_1, T_2, and T_3).

It is due to this memoryless property that Markov processes have an interevent/sojourn time T_i that is exponential.

To determine the distribution of the T_i, we note that the event $\{T_1 > t\}$ takes place if and only if no events of the Poisson process occur in the interval $[0, t]$, where T_1 denotes the random variable for the time until the process reaches state 1, or the time of the first event, which is the holding time until the first jump. Thus [1],

$$\text{Prob}\{T_1 > t\} = \text{Prob}\{X(t) = 0\} = e^{-\lambda t}\,.$$

Alternatively,

$$F(t) = \text{Prob}\{T_1 \le t\} = 1 - e^{-\lambda t}\,,$$

which is the cumulative distribution function for an exponential random variable with parameter λ. Hence, T_1 is an exponential random variable with parameter λ and having an exponential distribution with mean $1/\lambda$, with:

- ▷ Cumulative distribution function $F(t) = 1 - e^{-\lambda t}$, and
- ▷ Probability distribution function $f(t) = \dfrac{dF(t)}{dt} = \lambda e^{-\lambda t}$.

Now, for $i = 2$, to obtain the distribution of T_2 condition on T_1,

$$\begin{aligned}\text{Prob}\{T_2 > t \,|\, T_1 = s\} &= \text{Prob}\{0 \text{ events in } (s, s+t] \,|\, T_1 = s\} \\ &= \text{Prob}\{0 \text{ events in } (s, s+t]\} \qquad \text{(by independent increments)} \\ &= e^{-\lambda t} \qquad \text{(by stationary increments)}\end{aligned} \tag{35}$$

where the last two equations follow from independent and stationary increments. Thus, T_2 is also an exponential random variable with parameter λ, and T_2 is independent of T_1. In general, $T_i,\ i = 1, 2, \cdots$ are independent and identically distributed exponential random variables with parameter λ and mean $1/\lambda$ [7, 8]. In other words, it takes an exponential amount of time to move from state i to state $i+1$, or, the random variable for the time between jumps i and $i+1$, $W_{i+1} - W_i$, has an exponential distribution with parameter λ. In fact, sometimes in the definition of the Poisson process it is stated that the sojourn times, $T_i = W_{i+1} - W_i$, are independent exponential random variables with parameter λ [1].

In general, we have the following theorem [9]

Theorem 5. *The sojourn/interevent/holding times $T_0, T_1, \cdots, T_{n-1}$ are independent random variables, each having the exponential probability density function*

$$f_{T_k}(t) = \lambda e^{-\lambda t}\,, \qquad t \geq 0\,.$$

Proof. We need to show that the joint probability density function of $T_0, T_1, \cdots, T_{n-1}$ is the product of the exponential densities given by

$$f_{T_0,T_1,\cdots,T_{n-1}}(t_0, t_1, \cdots, t_{n-1}) = (\lambda e^{-\lambda t_0})(\lambda e^{-\lambda t_1}) \cdots (\lambda e^{-\lambda t_{n-1}})\,. \tag{36}$$

Because the general case is entirely similar, here we only give the proof for the case $n = 2$. Referring to Fig. 5, the joint occurrence of

$$t_1 < T_1 < t_1 + \Delta t_1 \qquad \text{and} \qquad t_2 < T_2 < t_2 + \Delta t_2$$

corresponds to no events in the intervals $(0, t_1]$ and $(t_1+\Delta t_1,\ t_1+\Delta t_1+t_2]$, and exactly one event occurs in each of the intervals $(t_1,\ t_1+\Delta t_1]$ and $(t_1+\Delta t_1+t_2,\ t_1+\Delta t_1+t_2+\Delta t_2]$. From its definition, the Poisson process $X(t)$ counts the number of events occurring up to time t, formally $X(t) = N((0, t])$. Thus

$$\begin{aligned} f_{T_1,T_2}(t_1, t_2)\Delta t_1 \Delta t_2 &= \text{Prob}\{t_1 < T_1 < t_1 + \Delta t_1,\ t_2 < T_2 < t_2 + \Delta t_2\} + o(\Delta t_1 \Delta t_2) \\ &= \text{Prob}\{X(t_1) = 0\} \times \text{Prob}\{X(t_2) = 0\} \times \text{Prob}\{X(\Delta t_1) = 1\} \\ &\quad \times \text{Prob}\{X(\Delta t_2) = 1\} + o(\Delta t_1 \Delta t_2) \\ &= e^{-\lambda t_1} e^{-\lambda t_2} e^{-\lambda \Delta t_1} e^{-\lambda \Delta t_2} \lambda(\Delta t_1)\lambda(\Delta t_2) + o(\Delta t_1 \Delta t_2) \\ &= (\lambda e^{-\lambda t_1})(\lambda e^{-\lambda t_2})(\Delta t_1)(\Delta t_2) + o(\Delta t_1 \Delta t_2) \end{aligned}$$

where

$$\begin{aligned} X(t_1) &= N((0, t_1])\,, \\ X(t_2) &= N((t_1 + \Delta t_1,\ t_1 + \Delta t_1 + t_2])\,, \\ X(\Delta t_1) &= N((t_1,\ t_1 + \Delta t_1])\,, \\ X(\Delta t_2) &= N((t_1 + \Delta t_1 + t_2,\ t_1 + \Delta t_1 + t_2 + \Delta t_2])\,. \end{aligned}$$

Dividing both sides by $\Delta t_1 \Delta t_2$ and taking the limits as $\Delta t_1 \to 0$ and $\Delta t_2 \to 0$, we obtain

$$f_{T_1,T_2}(t_1, t_2) = (\lambda e^{-\lambda t_1})(\lambda e^{-\lambda t_2})$$

which is Eq. (36) in the case $n = 2$. □

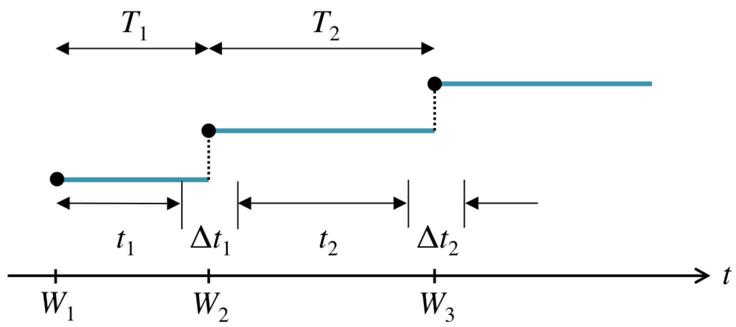

Figure 5:

Let W_i, $i = 1, 2, \cdots$ be the arrival time of the ith event, also referred to as the waiting time until the ith event. We can see that

$$W_i = \sum_{k=1}^{i} T_k \,, \qquad i \geq 1\,.$$

Theorem 6. *The waiting time W_i has the gamma distribution whose probability density function is*

$$f_{W_i}(t) = \frac{\lambda^i t^{i-1}}{(i-1)!} e^{-\lambda t}\,, \qquad i = 1, 2, \cdots, t \geq 0\,. \tag{37}$$

In particular, W_1, the time to the first event, is exponentially distributed,

$$f_{W_1}(t) = \lambda e^{-\lambda t}\,, \qquad t \geq 0\,. \tag{38}$$

Proof. The event $W_i \leq t$ occurs if and only if there are at least i events in the interval $(0, t]$. Since the number of events in $(0, t]$ has a Poisson distribution with mean λt, we obtain the cumulative distribution function (c.d.f.) of W_n by

$$\begin{aligned} F_{W_i}(t) &= \text{Prob}\{W_i \leq t\} = \text{Prob}\{X(t) \geq i\} \\ &= \sum_{k=i}^{\infty} \frac{(\lambda t)^k e^{-\lambda t}}{k!} \\ &= 1 - \sum_{k=0}^{i-1} \frac{(\lambda t)^k e^{-\lambda t}}{k!}\,, \qquad i = 1, 2, \cdots, \quad t \geq 0\,. \end{aligned}$$

The probability distribution function (p.d.f.) is obtained by differentiating the c.d.f. $F_{W_i}(t)$, that is

$$\begin{aligned}
f_{W_i}(t) &= \frac{d}{dt} F_{W_i}(t) \\
&= \frac{d}{dt}\left(1 - e^{-\lambda t}\left(1 - \frac{\lambda t}{1!} + \frac{(\lambda t)^2}{2!} - \cdots + \frac{(\lambda t)^{i-1}}{(i-1)!}\right)\right) \\
&= -e^{-\lambda t}\left(\lambda + \lambda\frac{\lambda t}{1!} + \lambda\frac{(\lambda t)^2}{2!} + \cdots + \lambda\frac{(\lambda t)^{i-2}}{(i-2)!}\right) \\
&\quad + \lambda e^{-\lambda t}\left(1 + \frac{\lambda t}{1!} + \frac{(\lambda t)^2}{2!} + \cdots + \frac{(\lambda t)^{i-1}}{(i-1)!}\right) \\
&= \frac{\lambda^i t^{i-1}}{(i-1)!} e^{-\lambda t}, \qquad i = 1, 2, \cdots, t \geq 0\,.
\end{aligned}$$

□

Next, it is shown that the random variable T_i can be expressed in terms of the cumulative distribution function (c.d.f.) $F_i(t)$ and a uniform random variable U. This relationship is very useful for computational purposes [1].

Theorem 7. *Let U be a uniform random variable defined on $[0, 1]$ and T be a continuous random variable defined on $[0, \infty)$. Then $T = F^{-1}(U)$, where F is the cumulative distribution of the random variable T.*

Proof. Since $\text{Prob}\{T \leq t\} = F(t)$, we want to show that $\text{Prob}\{F^{-1}(U) \leq t\} = F(t)$. First note that $F : [0, \infty) \to [0, 1)$ is strictly increasing, so that F^{-1} exists. In addition, for $t \in [0, \infty)$,

$$\begin{aligned}
\text{Prob}\{F^{-1}(U) \leq t\} &= \text{Prob}\{F(F^{-1}(U)) \leq F(t)\} \\
&= \text{Prob}\{U \leq F(t)\}\,.
\end{aligned}$$

Because U is a uniform random variable, then $\text{Prob}\{U \leq y\} = y$ for $y \in [0, 1]$. Thus, $\text{Prob}\{U \leq F(t)\} = F(t)$. □

In the Poisson process, the only change in state is a birth that occurs with probability $\lambda\Delta t + o(\Delta t)$ in a small interval of time Δt. The cumulative distribution function for the sojourn/interevent time is

$$F_i(t) = \text{Prob}\{T_i \leq t\} = 1 - e^{-\lambda t}\,.$$

But because λ is independent of the state of the process, the sojourn/interevent time is the same for every jump i, thus $T_i \equiv T$. The sojourn/interevent time T, expressed in terms of the uniform random variable U, is $T = F^{-1}(U)$. The function $F^{-1}(U)$ is found by solving

$$F(T) = 1 - e^{-\lambda T} = U$$

to get

$$T = -\frac{\ln(1-U)}{\lambda} = F^{-1}(U)\,.$$

However, because U is a uniform random variable on $[0,1]$, so is $1-U$. It follows that the sojourn/interevent time can be expressed in terms of a uniform random variable U as follows

$$T = -\frac{\ln(U)}{\lambda}\,. \tag{39}$$

where U is a uniform random variable on $[0,1]$. An example of Poisson process plot using Python is shown in Fig. 6. The Python code for this figure is provided in the Python Implementation section.

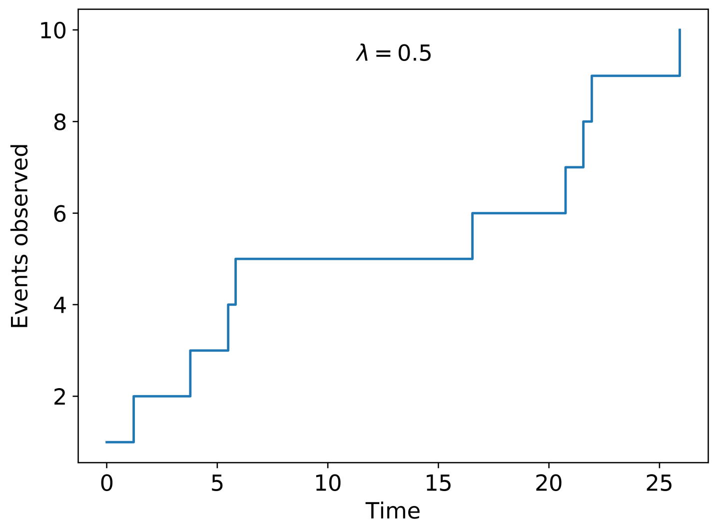

Figure 6: A sample path of a Poisson process with $\lambda = 0.5$ and 10 events.

For more general processes, the formula given in Eq. (39) for the sojourn/interevent time depends on the state of the process.

The formula given in Eq. (39) for the sojourn/interevent time is applied to three simple birth and death processes, known as simple birth process (or Yule process), simple death process (or linear death process), and simple birth and death process. In each of these processes, probabilities of births and deaths are linear functions of the population size. In all cases, $X(t)$ is the random variable for the total population size at time t.

2.3 Transition Rates and Infinitesimal Generator Matrix

In continuous-time Markov chains, the corresponding key quantities are the transition rates, denoted by q_{ij}. The transition probabilities P_{ij} are used to derive transition rates q_{ij}. The transition rates are used to define the forward and backward Kolmogorov differential equations, which we discuss in the next section on birth and death processes.

Recall the four properties of the transition probabilities $P_{ij}(t)$ given in subsection 1.5 on properties of the continuous-time Markov chains, rewrite here:

(a) $P_{ij}(t) \geq 0\,,$

(b) $\sum_{j=0}^{N} P_{ij} = 1\,, \qquad i, j = 0, 1, 2, \cdots\,,$

(c) The transition probabilities $P_{ij}(t)$ are solutions of the Chapman-Kolmogorov equations:

$$P_{ij}(t+s) = \sum_{k=0}^{\infty} P_{ik}(t) P_{kj}(s)\,,$$

or in the matrix form

$$\mathbf{P}(t+s) = \mathbf{P}(t)\mathbf{P}(s)\,, \qquad t, s \geq 0\,. \tag{40}$$

(d) In addition, we postulate that

$$\lim_{t \to 0^+} P_{ij}(t) = \begin{cases} 1, & i = j \\ 0, & i \neq j\,. \end{cases}$$

Property (d) asserts that $P_{ij}(t)$ are continuous and differentiable for $t \geq 0$, and at $t = 0$:

$$\begin{aligned} P_{ij}(0) &= \text{Prob}\{X(s) = j \,|\, X(s) = i\} \\ &= \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \end{aligned}$$

or

- $P_{ij}(0) = 0$, for $i \neq j$.
- $P_{ii}(0) = 1$.

In the matrix form, it leads to $\mathbf{P}(0) = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. Hence, $\mathbf{P}(t)$ is continuous at $t = 0$. It follows from Eq. (40) that $\mathbf{P}(t)$ is continuous for all $t > 0$:

♠ If $s = h > 0$, then due to the property (d), we have [3, 4, 9]

$$\lim_{h\to 0^+} \mathbf{P}(t+h) = \mathbf{P}(t) \lim_{h\to 0^+} \mathbf{P}(h) = \mathbf{P}(t)\,\mathbf{I} = \mathbf{P}(t)\,. \tag{41}$$

♠ Whereas when $t > 0$ and $0 < h < t$, then

$$\mathbf{P}(t) = \mathbf{P}(t-h)\,\mathbf{P}(h)\,. \tag{42}$$

♠ But when $s = \Delta t$ and Δt is sufficiently small, $\mathbf{P}(\Delta t)$ is near the identity and so matrix $\mathbf{P}(\Delta t)$ is invertible, or $(\mathbf{P}(\Delta t))^{-1}$ exists, and also approaches the identity $\mathbf{I}$. Therefore [3, 4, 9]

$$\mathbf{P}(t) = \mathbf{P}(t) \lim_{\Delta t\to 0^+} (\mathbf{P}(\Delta t))^{-1} = \lim_{\Delta t\to 0^+} \mathbf{P}(t-\Delta t)\,. \tag{43}$$

Now we look at the transition rates. For any pair of states i and j, the rate of the process, when in state i, makes a transition into state j, denoted by q_{ij}, is defined by [7]

$$q_{ij} = v_i P_{ij}\,,$$

where v_i is the rate at which the process makes a transition when in state i and P_{ij} is the probability that this transition is into state j. The quantities q_{ij} are called the *instantaneous transition rates*, and $q_{ij} \geq 0$.

We can obtain v_i from

$$v_i = v_i \sum_j P_{ij} = \sum_j v_i P_{ij} = \sum_j q_{ij}\,.$$

Thus, the transition probabilities can be written as

$$P_{ij} = \frac{q_{ij}}{v_i} = \frac{q_{ij}}{\sum_j q_{ij}}\,.$$

Hence, specifying the instantaneous transition rates q_{ij} determines the parameters of the continuous-time Markov chain.

The rates q_{ij} and v_i furnish an infinitesimal description of the process with [9]

$$\begin{aligned} P_{ii}(\Delta t) &= \text{Prob}\{X(t+\Delta t) = i \,|\, X(t) = i\} &&= 1 - v_i\Delta t + o(\Delta t)\,, \\ P_{ij}(\Delta t) &= \text{Prob}\{X(t+\Delta t) = j \,|\, X(t) = i\} &&= q_{ij}\Delta t + o(\Delta t)\,, \quad \text{for } \; i \neq j \end{aligned} \tag{44}$$

The limit relations in Eqs. (41) and (43) together show that $P_{ij}(t)$ are continuous. Theorems 1.1 and 1.2 in [4] (pp. 139–142) show that $P_{ij}(t)$ are differentiable with right-hand derivatives. We use it to define the following:

(1) The probability that a process in state i at time 0 will not be in state i at time Δt, which is $1 - P_{ii}(\Delta t)$, equals the probability that a transition occurs within time Δt plus something small compared to Δt [7]. Hence,

$$1 - P_{ii}(\Delta t) = v_i \Delta t + o(\Delta t)\,, \tag{45}$$

and dividing by Δt and taking the limit as $\Delta t \to 0^+$ give us

$$\lim_{\Delta t \to 0} \frac{1 - P_{ii}(\Delta t)}{\Delta t} = v_i\,, \qquad \forall\, i\,. \tag{46}$$

(2) The probability that the process goes from state i to state j in a time Δt, which is $P_{ij}(\Delta t)$, equals the probability that a transition occurs in this time multiplied by the probability that the transition is into state j, plus something small compared to Δt [7]. That is,

$$P_{ij}(\Delta t) = v_i P_{ij} \Delta t + o(\Delta t) = q_{ij} \Delta t + o(\Delta t) \tag{47}$$

thus, the limit,

$$\lim_{\Delta t \to 0} \frac{P_{ij}(\Delta t)}{\Delta t} = q_{ij}\,, \qquad \forall\, i, j;\; i \neq j\,. \tag{48}$$

Note that since the amount of time until a transition occurs is exponentially distributed it follows that the probability of two or more transitions in a time Δt is $o(\Delta)$.

From Eqs. (45) and (47), it follows that

$$P_{ij}(\Delta t) = \delta_{ij} + q_{ij} \Delta t + o(\Delta t)\,, \tag{49}$$

where

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

and

$$\begin{aligned} q_{ii} &= \lim_{\Delta t \to 0^+} \frac{P_{ii}(\Delta t) - P_{ii}(0)}{\Delta t} = \lim_{\Delta t \to 0} \frac{P_{ii}(\Delta t) - 1}{\Delta t} && (50) \\ &= \lim_{\Delta t \to 0^+} \frac{-\sum_{j=0, j\neq i}^{\infty} (q_{ij}\Delta t + o(\Delta t))}{\Delta t} \\ &= -\sum_{j=0, j\neq i}^{\infty} q_{ij} \\ &= -v_i\,. && (51) \end{aligned}$$

The transition rates q_{ij} and q_{ii} form a matrix known as the infinitesimal generator matrix $\mathbf{Q}$. Matrix $\mathbf{Q}$ defines a relationship between the rates of change of the transition probabilities [1].

Individual limits in Eqs. (46) and (48) can be expressed more as the entries in a matrix limit. Recall the matrix of transition probabilities defined in Eq. (20), and now we define

$$\mathbf{P}(\Delta t) = (P_{ij}(\Delta t)) \tag{52}$$

as the *infinitesimal transition matrix*. Let $\mathbf{I}$ be the identity matrix having the same dimension as $\mathbf{P}(\Delta t)$. Then matrix $\mathbf{Q}$ is defined as the one-sided derivative [1, 9]

$$\mathbf{Q} = \lim_{\Delta t \to 0^+} \frac{\mathbf{P}(\Delta t) - \mathbf{I}}{\Delta t}, \tag{53}$$

which shows that $\mathbf{Q}$ is the matrix derivative of $\mathbf{P}(t)$ at $t = 0$. Formally, $\mathbf{Q} = \mathbf{P}'(0)$.

The matrix $\mathbf{Q} = (q_{ij})$ is called the *matrix of transition rates* and whose elements are

$$\mathbf{Q} = \begin{pmatrix} q_{00} & q_{01} & q_{02} & \cdots \\ q_{10} & q_{11} & q_{12} & \cdots \\ q_{30} & q_{31} & q_{32} & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix} = \begin{pmatrix} -\sum\limits_{i=1}^{\infty} q_{0i} & q_{01} & q_{02} & \cdots \\ q_{10} & -\sum\limits_{i=0, i\neq 1}^{\infty} q_{1i} & q_{12} & \cdots \\ q_{20} & q_{21} & -\sum\limits_{i=0, i\neq 2}^{\infty} q_{2i} & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix}.$$

Sometimes, the transition rate matrix $\mathbf{Q}$ is also referred to as the *infinitesimal generator matrix* or simply the *generator matrix*.

Matrix $\mathbf{Q}$ has the following properties:

(1) the diagonal entries are nonpositive,

(2) the nondiagonal entries of $\mathbf{Q}$ are nonnegative,

(3) the sum of each row equals zero.

By Eq. (53) and referring to Eq. (40), we have

$$\begin{aligned}\frac{\mathbf{P}(t+\Delta t)-\mathbf{P}(t)}{\Delta t} &= \frac{\mathbf{P}(t)\mathbf{P}(\Delta t)-\mathbf{P}(t)}{\Delta t} \\ &= \frac{\mathbf{P}(t)\big[\mathbf{P}(\Delta t)-\mathbf{I}\big]}{\Delta t} \\ &= \mathbf{P}(t)\,\frac{\mathbf{P}(\Delta t)-\mathbf{I}}{\Delta t} \quad \text{or} \quad \frac{\mathbf{P}(\Delta t)-\mathbf{I}}{\Delta t}\,\mathbf{P}(t)\,. \qquad (54)\end{aligned}$$

If the limit on the right exists, it leads to the matrix of differential equation,

$$\frac{d}{dt}\mathbf{P}(t) = \lim_{\Delta t\to 0^+} \frac{\mathbf{P}(t+\Delta t)-\mathbf{P}(t)}{\Delta t}\,, \qquad (55)$$

where the derivative is taken entry-wise. As we have seen in Eq. (54), we have two choices factoring $\mathbf{P}(t)$ out, on the right or on the left. Therefore, the system of differential equations for the chain can then be written in two ways,

$$\frac{d}{dt}\mathbf{P}(t) = \mathbf{Q}\,\mathbf{P}(t)\,, \qquad (56)$$

which is known as the *backward Kolmogorov equation*, and

$$\frac{d}{dt}\mathbf{P}(t) = \mathbf{P}(t)\,\mathbf{Q}\,, \qquad (57)$$

which is known as the *forward Kolmogorov equation*.

If $\mathbf{P}(0) = \mathbf{I}$, then the solution of the matrix differential equations (56) and (57) is given by

$$\mathbf{P}(t) = e^{\mathbf{Q}t} = \mathbf{I} + \sum_{n=1}^{\infty} \frac{\mathbf{Q}^n t^n}{n!}\,. \qquad (58)$$

Example 4a.6

Taken from Example on pp. 396–397 in [9]:

Consider a chain with two states $\{0, 1\}$. Assume $q_{01} = \alpha$ and $q_{10} = \beta$. Then the infinitesimal generator matrix is

$$\mathbf{Q} = \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix}.$$

The process alternates between states 0 and 1. The waiting times in state 0 are independent and exponentially distributed with parameter α. Those in state 1 are independent and exponentially distributed with parameter β. The matrix multiplication,

$$\begin{aligned}
\begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix} \times \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix} &= \begin{bmatrix} \alpha^2 + \alpha\beta & -\alpha^2 - \alpha\beta \\ -\beta^2 - \alpha\beta & \beta^2 + \alpha\beta \end{bmatrix} \\
&= -(\alpha+\beta) \begin{bmatrix} -\alpha & \alpha \\ \beta & -\beta \end{bmatrix},
\end{aligned}$$

from which we obtain that

$$\mathbf{Q}^2 = -(\alpha+\beta)\mathbf{Q}\,.$$

Repeated multiplication by $\mathbf{Q}$ yields

$$\mathbf{Q}^n = \big[-(\alpha+\beta)\big]^{n-1}\mathbf{Q}\,,$$

which when inserted into Eq. (58) simplifies the sum according to

$$\begin{aligned}
\mathbf{P}(t) &= \mathbf{I} - \frac{1}{\alpha+\beta}\sum_{n=1}^{\infty}\frac{[-(\alpha+\beta)]^n}{n!}\,\mathbf{Q} \\
&= \mathbf{I} - \frac{1}{\alpha+\beta}\left[e^{-(\alpha+\beta)t} - 1\right]\mathbf{Q} \\
&= \mathbf{I} + \frac{1}{\alpha+\beta}\,\mathbf{Q} - \frac{1}{\alpha+\beta}\,\mathbf{Q}\,e^{-(\alpha+\beta)t}\,. \qquad \blacksquare \qquad (59)
\end{aligned}$$

Example 4a.7

Taken from Example 1 pp. 70–71 in [5]:

Consider a chain with two states $\{0,1\}$. Assume $q_{01} = 1$ and $q_{10} = 2$. Then the infinitesimal generator matrix is

$$\mathbf{Q} = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}.$$

Using Eq. (59), we obtain

$$\begin{aligned}
\mathbf{P}(t) &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{3}\begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix} - \frac{1}{3}e^{-3t}\begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix} \\
&= \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix} + e^{-3t}\begin{bmatrix} \frac{1}{3} & -\frac{1}{3} \\ -\frac{2}{3} & \frac{2}{3} \end{bmatrix}. \qquad \blacksquare
\end{aligned}$$

Example 4a.8

Taken from Example 5.1 pp. 206–207 in [1]:

Show that the system of differential-difference equations for the Poisson process in Eq. (33) can be expressed in terms of the generator matrix $\mathbf{Q}$.

Solution:

The infinitesimal generator matrix for the Poisson process from Eqs. (32) and (33) is obtained by

$$\begin{aligned}
i = 0: &\quad \frac{dP_0(t)}{dt} = -\lambda P_0(t) \\
i = 1: &\quad \frac{dP_1(t)}{dt} = -\lambda P_1(t) + \lambda P_0(t) \\
i = 2: &\quad \frac{dP_2(t)}{dt} = -\lambda P_2(t) + \lambda P_1(t) \\
\cdots &\quad \cdots
\end{aligned}$$

and arranging into a matrix form,

$$\begin{bmatrix} dP_0(t)/dt \\ dP_1(t)/dt \\ dP_2(t)/dt \\ \vdots \end{bmatrix} = \begin{bmatrix} P_0(t) & P_1(t) & P_2(t) & \cdots \end{bmatrix} \begin{bmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ 0 & -\lambda & \lambda & 0 & \cdots \\ 0 & 0 & -\lambda & \lambda & \cdots \\ \vdots & \vdots & \vdots & & \ddots \end{bmatrix}.$$

The system of differential-difference equations for the Poisson process, Eqs. (32) and (33), can be expressed similarly as the forward Kolmogorov equation $d\mathbf{P}(t)/dt = \mathbf{P}(t)\mathbf{Q}$, where the infinitesimal generator matrix is

$$\mathbf{Q} = \begin{bmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ 0 & -\lambda & \lambda & 0 & \cdots \\ 0 & 0 & -\lambda & \lambda & \cdots \\ \vdots & \vdots & \vdots & & \ddots \end{bmatrix}. \quad \blacksquare \tag{60}$$

2.4 Stationary Probability Distribution

A stationary probability distribution of a Markov chain is a probability distribution that remains unchanged in the Markov chain as time progresses. Typically, it is represented by a row vector, usually denoted by π, whose entries are probabilities summing to one. We have discussed the basics of stationary probability distribution in a great detail in Section 1.6.

A stationary probability distribution can be defined in terms of the generator matrix $\mathbf{Q}$ and the transition probability matrix $\mathbf{P}(t)$:

(1) Let $\{X(t) : \; t \in [0, \infty)\}$ be a continuous-time Markov chain with generator matrix $\mathbf{Q}$. Suppose $\pi = (\pi_0, \pi_1, \cdots)^T$ is nonnegative, *i.e.*, $\pi_i \geq 0$ for $i = 0, 1, 2, \cdots$,

$$\mathbf{Q}\pi = 0\,, \qquad \text{and} \qquad \sum_{i=0}^{\infty} \pi_i = 1\,.$$

Then π is called a *stationary probability distribution* of the continuous-time Markov chain.

(2) A constant solution π is called a stationary probability distribution if

$$\mathbf{P}(t)\pi = \pi\,, \quad \text{for} \;\; t \geq 0\,, \quad \sum_{i=0}^{\infty} \pi_i = 1\,, \quad \text{and} \;\; \pi_i \geq 0\,.$$

Note that the equation $\mathbf{P}(t)\pi = \pi$ looks similar to the equation $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ for eigenvalues and eigenvectors with $\lambda = 1$, where the product of a square matrix $\mathbf{A}$ and a column vector $\mathbf{v}$ equals to the scaling of the column vector by a factor λ.

Using the Kolmogorov differential equations, we can prove that the two definitions of the stationary probability distribution involving $\mathbf{Q}$ and $\mathbf{P}(t)$ above are equivalent if [1]:

(1) the transition matrix $\mathbf{P}(t)$ is a solution of the forward Kolmogorov differential equation. That is, if we multiply $d\mathbf{P}(t)/dt = \mathbf{P}(t)\mathbf{Q}$ by π and using the first definition where $\mathbf{Q}\pi = 0$, then

$$\frac{d\mathbf{P}(t)}{dt}\pi = \frac{d}{dt}\mathbf{P}(t)\pi = \mathbf{P}(t)\mathbf{Q}\pi = 0$$

implies that $\mathbf{P}(t)\pi =$ constant for all t.

(2) the transition matrix $\mathbf{P}(t)$ is a solution of the backward Kolmogorov differential equation. That is, if we multiply $d\mathbf{P}(t)/dt = \mathbf{Q}\mathbf{P}(t)$ by π and using the second definition where $\mathbf{P}(t)\pi = \pi$, then

$$\frac{d\mathbf{P}(t)}{dt}\pi = \mathbf{Q}\mathbf{P}(t)\pi = \mathbf{Q}\pi = 0\,.$$

The following theorem adds that *nonexplosive* is another property for the limiting distribution to be considered a stationary probability distribution [1], hence the properties now:

- nonexplosive,
- positive recurrent,
- aperiodic, and
- irreducible.

Theorem 8. *Let $\{X(t) : t \in [0, \infty)\}$ be a nonexplosive, positive recurrent, and irreducible continuous-time Markov chain with transition matrix $\mathbf{P}(t) = (P_{ij}(t))$ and generator matrix $\mathbf{Q} = (q_{ij})$. Then there exists a unique positive stationary probability distribution π, $\mathbf{Q}\pi = 0$, such that*

$$\lim_{n\to\infty} P_{ij}(t) = \pi_i\,, \qquad i, j = 1, 2, \cdots$$

It follows that the mean recurrence time is given by

$$\pi_i = -\frac{1}{q_{ii}\mu_{ii}} > 0\,. \tag{61}$$

3 Python Implementation

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

Following Eq. (39), to generate a uniformly distributed random number from the uniform distribution $U(0, 1)$ in Python, we use

```
import numpy as np

U = np.random.rand()
```

from which we obtain T:

```
T = -np.log(U)/lam
```

where `lam` is the parameter λ.

The whole code to generate Fig. 6 is as follows:

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

np.random.seed(100)

fig, ax = plt.subplots()

b = 0.5 # birth rate
N = 10 # maximal population size
N0 = 1 # initial population size

samplepaths = 1

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
        s[j,i+1] = s[j,i] + h
        X[j,i+1] = X[j,i] + 1



## Sets axis ranges for plotting
xaxis = max([max(s[k,:]) for k in range(samplepaths)])
ymin = min([min(X[k,:]) for k in range(samplepaths)])
ymax = max([max(X[k,:]) for k in range(samplepaths)])

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

#plt.axis([-0.5, xaxis+(.01*xaxis), ymin-1, ymax+(0.01*ymax)])
ax.set_xlabel('Time', fontsize=14)
ax.set_ylabel('Events observed', fontsize=14)
plt.xticks(fontsize=14)
plt.yticks(fontsize=14)
plt.tight_layout()
plt.legend(loc=2)
plt.show()
```

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