Richard J. Boucherie

Markovian queueing networks

September 11, 2022

Springer

Contents

Part I General solution concepts

Part I
General solution concepts

Chapter 1
Preliminaries

This chapter reviews and discusses the basic assumptions and techniques that will be used in this monograph. Proofs of results given in this chapter are omitted, but can be found in standard textbooks on Markov chains and queueing theory, e.g., [1, 2, 6, 8, 12, 13, 14, 15]. Results from these references are used in this chapter without reference except for cases where a specific result (e.g. theorem) is inserted into the text.

1.1 Basic results for Markov chains

Consider a stochastic process $\{N(t),\ t \in T\}$ taking values in a countable state space S. Applications will usually assume that $S \subseteq \mathbb{N}_0^J$ and that t represents time.[1] Then, a state $\mathbf{s} = (n_1, \ldots, n_J) \in S$ is a vector with components $n_i \in \mathbb{N}_0$, $i = 1, \ldots, J$. For a *discrete-time* stochastic process T is the set of integers: $T = \mathbb{N}_0$, or $T = \mathbb{Z}$, whereas for a *continuous-time* stochastic process T is the positive real line: $T = \mathbb{R}_0^+$ or the real line $T = \mathbb{R}$. A vector $\mathbf{s} \in \mathbb{R}^J$ is called non-negative if $n_i \geq 0$, $i = 1, \ldots, J$, and positive if it is non-negative and non-null. In this monograph emphasis will be on continuous-time stochastic processes. Therefore, in the sequel all results are given for continuous-time stochastic processes only. The exposition in this section focusses on continuous-time Markov chains with countable state space S. In this section, we will not impose further structure on the states $\mathbf{s} \in S$.

A stochastic process is a *stationary process* if $(N(t_1), N(t_2), \ldots, N(t_k))$ has the same distribution as $(N(t_1+\tau), N(t_2+\tau), \ldots, N(t_k+\tau))$ for all $k \in \mathbb{N}$, $t_1, t_2, \ldots, t_k \in T$, $\tau \in T$. The stochastic process $\{N(t),\ t \in T\}$ is a *Markov process* if for every $k \geq 1$, $t_1 < \cdots < t_k < t_{k+1}$, and any $\mathbf{s}_1, \ldots, \mathbf{s}_{k+1}$ in S, the joint distribution of $(N(t_1), \ldots, N(t_{k+1}))$ is such that

[1] $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$, $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$, $\mathbb{R} = (-\infty, \infty)$, $\mathbb{R}_0^+ = [0, \infty)$.

$$\mathbb{P}\{N(t_{k+1}) = \mathbf{s}_{k+1} | N(t_1) = \mathbf{s}_1, \ldots, N(t_k) = \mathbf{s}_k\}$$
$$= \mathbb{P}\{N(t_{k+1}) = \mathbf{s}_{k+1} | N(t_k) = \mathbf{s}_k\}, \qquad (1.1)$$

whenever the conditioning event $(N(t_1) = \mathbf{s}_1, \ldots, N(t_k) = \mathbf{s}_k)$ has positive probability. In words, for a Markov process the state at a given time contains all information about the past evolution necessary to probabilistically predict the future evolution of the process.

A Markov process is *time-homogeneous* if the conditional probability $\mathbb{P}\{N(s+t) = \mathbf{s}' | N(s) = \mathbf{s}\}$ is independent of s for all $s,t > 0$, $\mathbf{s},\mathbf{s}' \in S$. For a time-homogeneous Markov process the *transition probability* from state $\mathbf{s}$ to state $\mathbf{s}'$ in time t is defined as

$$P(\mathbf{s},\mathbf{s}';t) = \mathbb{P}\left\{N(s+t) = \mathbf{s}' | N(s) = \mathbf{s}\right\}, \quad s,t > 0.$$

The *transition matrix* $P(t) = (P(\mathbf{s},\mathbf{s}';t),\ \mathbf{s},\mathbf{s}' \in S)$ has non-negative entries (1.2) and row sums equal to one (1.3). The *Markov property* (1.1) implies that the transition probabilities satisfy the *Chapman-Kolmogorov equations* (1.4). In addition, we assume that the transition matrix is *standard* (1.5). For all $\mathbf{s},\mathbf{s}' \in S$, $s,t > 0$, a *standard transition matrix* satisfies:

$$P(\mathbf{s},\mathbf{s}';t) \geq 0; \qquad (1.2)$$

$$\sum_{\mathbf{s}' \in S} P(\mathbf{s},\mathbf{s}';t) = 1; \qquad (1.3)$$

$$P(\mathbf{s},\mathbf{s}';s+t) = \sum_{\mathbf{s}'' \in S} P(\mathbf{s},\mathbf{s}'';s)P(\mathbf{s}'',\mathbf{s}';t); \qquad (1.4)$$

$$\lim_{t \downarrow 0} P(\mathbf{s},\mathbf{s}';t) = \delta_{\mathbf{s},\mathbf{s}'}, \qquad (1.5)$$

where $\delta_{\mathbf{s},\mathbf{s}'}$ is the *Kronecker-delta*, $\delta_{\mathbf{s},\mathbf{s}'} = 1$ if $\mathbf{s} = \mathbf{s}'$ and $\delta_{\mathbf{s},\mathbf{s}'} = 0$ if $\mathbf{s} \neq \mathbf{s}'$. For a standard transition matrix it is natural to extend the definition of $P(\mathbf{s},\mathbf{s}';\cdot)$ to $[0,\infty)$ by setting $P(\mathbf{s},\mathbf{s}';0) = \delta_{\mathbf{s},\mathbf{s}'}$. Then for all $\mathbf{s},\mathbf{s}'$ the transition probabilities are *uniformly continuous* on $[0,\infty)$. Furthermore, each $P(\mathbf{s},\mathbf{s}';t)$ is either identically zero for all $t > 0$ or never zero for $t > 0$ (Lévy's dichotomy [2, Theorem II.5.2]).

For a standard transition matrix the *transition rate* from state $\mathbf{s}$ to state $\mathbf{s}'$ is defined as

$$q(\mathbf{s},\mathbf{s}') = \lim_{h \downarrow 0} \frac{P(\mathbf{s},\mathbf{s}';h) - \delta_{\mathbf{s},\mathbf{s}'}}{h}.$$

For all $\mathbf{s},\mathbf{s}' \in S$ this limit exists. For $\mathbf{s} \neq \mathbf{s}'$ this limit is finite (1.6), whereas for $\mathbf{s} = \mathbf{s}'$ the limit may be infinite. For practical systems the limit for $\mathbf{s} = \mathbf{s}'$ is finite too. In the sequel, we assume that the limit exists for $\mathbf{s} = \mathbf{s}'$: (1.7). A Markov process is called a continuous-time *Markov chain* if for all $\mathbf{s},\mathbf{s}' \in S$ the limit exists and is finite (1.6), (1.7). In addition, we assume that the rate matrix is *conservative* (1.8). Then for all $\mathbf{s},\mathbf{s}'$ the rate matrix satisfies

$$0 \le q(\mathbf{s},\mathbf{s}') < \infty, \quad \mathbf{s}' \neq \mathbf{s}; \tag{1.6}$$

$$0 \le q(\mathbf{s}) := -q(\mathbf{s},\mathbf{s}) < \infty; \tag{1.7}$$

$$\sum_{\mathbf{s}' \in S} q(\mathbf{s},\mathbf{s}') = 0. \tag{1.8}$$

For a *rate matrix* $Q = (q(\mathbf{s},\mathbf{s}'),\ \mathbf{s},\mathbf{s}' \in S)$ that satisfies (1.6), (1.7), the definition of the transition rates implies that the transition probabilities can be expressed in the transition rates. This gives, for $\mathbf{s},\mathbf{s}' \in\ S$,

$$P(\mathbf{s},\mathbf{s}';h) = \delta_{\mathbf{s},\mathbf{s}'} + q(\mathbf{s},\mathbf{s}')h + \mathrm{o}(h) \quad \text{for } h \downarrow 0, \tag{1.9}$$

where $\mathrm{o}(h)$ denotes a function $g(h)$ with the property that $g(h)/h \to 0$ as $h \downarrow 0$. For small positive values of h, for $\mathbf{s}' \neq \mathbf{s}$, the term $q(\mathbf{s},\mathbf{s}')h$ may be interpreted as the conditional probability, up to order $\mathrm{o}(h)$, that the Markov chain $\{N(t)\}$ makes a transition to state $\mathbf{s}'$ during $(t,t+h)$ given that the process is in state $\mathbf{s}$ at time t. From (1.7), (1.8), note that $q(\mathbf{s}) = \sum_{\mathbf{s}' \neq \mathbf{s}} q(\mathbf{s},\mathbf{s}')$. If $q(\mathbf{s})$ is finite, $q(\mathbf{s})h$ is the conditional probability that $\{N(t)\}$ leaves this state during $(t,t+h)$ given that $\{N(t)\}$ is in state $\mathbf{s}$ at time t. As a consequence, $q(\mathbf{s},\mathbf{s}')$ can be interpreted as the rate at which transitions occur, i.e., as transition rates. To elaborate on the transition rates and on the role of stability, consider the conditional probability that the process remains in $\mathbf{s}$ during $(t,t+h)$ if the process is in $\mathbf{s}$ at time t. This conditional probability is

$$\mathbb{P}\{N(\tau) = \mathbf{s},\ t < \tau < t+h | N(t) = \mathbf{s}\} = \mathrm{e}^{-q(\mathbf{s})h}, \quad h > 0.$$

The *exit-time* from state $\mathbf{s}$, $\xi(\mathbf{s})$, defined as

$$\xi(\mathbf{s}) = \inf\{s :\ s > 0,\ N(t+s) \neq \mathbf{s}\}$$

given that the process is in state $\mathbf{s}$ at time s, has an exponential distribution with mean $q(\mathbf{s})^{-1}$.

For every initial state $N(0) = \mathbf{s}$, $\{N(t),\ t \in T\}$ is a *pure-jump process*, which means that the process jumps from state to state and remains in each state a *strictly positive* sojourn-time with probability 1. For the Markovian case, the process remains in state $\mathbf{s}$ for an exponentially distributed sojourn-time with mean $q(\mathbf{s})^{-1}$. In addition, conditional on the process departing from state $\mathbf{s}$ it jumps to state $\mathbf{s}'$ with probability $p(\mathbf{s},\mathbf{s}') = q(\mathbf{s},\mathbf{s}')/q(\mathbf{s})$. This second interpretation is sometimes used as a definition of a continuous-time Markov chain and is used to construct such processes. The Markov chain represented via the exponentially distributed holding times with mean $q(\mathbf{s})^{-1}$ and transition probabilities $p(\mathbf{s},\mathbf{s}')$, $\mathbf{s},\mathbf{s}' \in S$, is referred to as the *Markov jump chain* of the Markov chain $\{N(t)\}$. Note that we obtain the Markov chain with transition rates $q(\mathbf{s},\mathbf{s}')$ from the Markov jump chain with holding times with mean $q(\mathbf{s})^{-1}$ and transition probabilities $p(\mathbf{s},\mathbf{s}')$ as $q(\mathbf{s},\mathbf{s}') = q(\mathbf{s})p(\mathbf{s},\mathbf{s}')$, $\mathbf{s},\mathbf{s}' \in S$.

From the Chapman-Kolmogorov equations (1.4) two systems of differential equations for the transition probabilities can be obtained. To this end, observe that for a standard transition matrix every element $P(\mathbf{s},\mathbf{s}';\cdot)$ has a continuous derivative in $(0,\infty)$, which is continuous at zero if the rate matrix satisfies (1.6), (1.7) [2, Theorem II.12.8]. Conditioning on the first jump of the Markov chain in $(0,t]$ yields the *Kolmogorov backward equations* (1.10), whereas conditioning on the last jump in $(0,t]$ gives the *Kolmogorov forward equations* (1.11). The validity of this method is discussed below. These equations read for $\mathbf{s},\mathbf{s}' \in S$, $t \geq 0$,

$$\frac{dP(\mathbf{s},\mathbf{s}';t)}{dt} = \sum_{\mathbf{s}''\in S} q(\mathbf{s},\mathbf{s}'')P(\mathbf{s}'',\mathbf{s}';t), \tag{1.10}$$

$$\frac{dP(\mathbf{s},\mathbf{s}';t)}{dt} = \sum_{\mathbf{s}''\in S} P(\mathbf{s},\mathbf{s}'';t)q(\mathbf{s}'',\mathbf{s}'). \tag{1.11}$$

If the rate matrix satisfies (1.6), (1.7), then starting from the initial state $N(0)=\mathbf{s}$, a first jump of the Markov chain exists for $t>0$. As a consequence conditioning on this first jump is allowed. In contrast, the last jump of the Markov chain in $(0,t]$ is not properly defined. It may be that also for a rate matrix that satisfies (1.6), (1.7) jumps will accumulate in such a way that $\{N(t)\}$ will make infinitely many jumps in finite time. In this case $\{N(t)\}$ is not properly defined for all $t>0$ from the rate matrix.

Example 1.1.1 (Explosion in a pure birth process) Consider the Markov chain $\{N(t),\ t\in[0,\infty)\}$, at state space $S=\mathbb{N}_0$ with transition rates

$$q(\mathbf{s},\mathbf{s}') = \begin{cases} q(\mathbf{s}), & \text{if } \mathbf{s}'=\mathbf{s}+1,\\ -q(\mathbf{s}), & \text{if } \mathbf{s}'=\mathbf{s},\\ 0, & \text{otherwise,}\end{cases}$$

with initial distribution $\mathbb{P}(N(0)=\mathbf{s})=\delta(\mathbf{s},0)$. Then $\{N(t)\}$ is a pure birth process that spends an exponentially distributed time with rate $q(\mathbf{s})$ in state $\mathbf{s}$ and then jumps to state $\mathbf{s}+1$ with probability 1, $\mathbf{s}\in S$. Let $\xi(\mathbf{s})$ denote the time spent in state $\mathbf{s}$, and $\xi=\sum_{\mathbf{s}=0}^{\infty}\xi(\mathbf{s})$ the time spent in the states $0,1,2,\ldots$. Let $q(\mathbf{s})=2^{\mathbf{s}}$, then

$$\mathbb{E}\{\xi\} = \sum_{\mathbf{s}=0}^{\infty}\mathbb{E}\{\xi(\mathbf{s})\} = \sum_{\mathbf{s}=0}^{\infty}2^{-\mathbf{s}} = 2$$

by monotone convergence. As $\mathbb{E}\{\xi\}<\infty$ it must be that $\mathbb{P}(\xi<\infty)=1$ and therefore $\{N(t)\}$ is explosive (diverges to infinity in finite time).[2] □

An additional assumption on the rate matrix guaranteeing the existence of a last jump in $(0,t]$ is regularity. A pure-jump Markov chain is *regular* if for every initial state $N(0)=\mathbf{s}$ the number of transitions in finite time is finite with probability 1. For

[2] We may actually show the following stronger result: *A pure birth process is explosive if and only if* $\sum_{\mathbf{s}=0}^{\infty}q(\mathbf{s})^{-1}<\infty$.

a regular Markov chain the last jump before t is well-defined and conditioning on the last jump before t is allowed. Thus if a pure-jump Markov chain satisfies (1.6), (1.7) and is regular, then for all $t > 0$ the evolution of the process is uniquely determined by the transition rates, that is specification of the transition rates is sufficient to completely characterize the process.

Regularity is a property of the rate matrix. It can be shown [11] that the rate matrix is regular if and only for some $\nu > 0$ the system of equations

$$\sum_{\mathbf{s}' \in S} q(\mathbf{s},\mathbf{s}')x(\mathbf{s}') = \nu x(\mathbf{s}), \quad \mathbf{s} \in S,$$

has no bounded solution other than $\{x(\mathbf{s}) = 0,\ \mathbf{s} \in S\}$. This characterization of regularity may be difficult to apply in practical situations. A simple sufficient condition ensuring regularity of a Markov chain is the existence of a uniform finite upper bound on $q(\mathbf{s})$. If such a bound exists, i.e., if a constant C exists such that for all $\mathbf{s} \in S$

$$q(\mathbf{s}) \le C < \infty,$$

then the Markov chain is said to be *uniformizable* and the forward and backward equations have the same solution. Uniformizability can be too strong for practical applications as it excludes, for example, the infinite-server queue (see Example **??**). More general sufficient conditions can be found in, e.g., [15, Section 4-3]. A detailed discussion of regularity is beyond the scope of this monograph. The behaviour of irregular Markov chains is, for example, discussed in [9, 10].

The following theorem summarizes the results on regularity and the forward and backward equations stated above.

Theorem 1.1.2 ([2, Theorem II.18.3]) *For a conservative, regular, continuous-time Markov chain the forward equations (1.11) and the backward equations (1.10) have the same unique solution* $\{P(\mathbf{s},\mathbf{s}';t),\ \mathbf{s},\mathbf{s}' \in S,\ t \ge 0\}$. *Moreover, this unique solution is the transition matrix of the Markov chain.*

In particular, Theorem 1.1.2 states that either the forward or the backward equations can be solved to find the transition matrix

$$P(t) = \mathrm{e}^{Qt} = \sum_{n=0}^{\infty} \frac{(Qt)^n}{n!}, \quad t \ge 0.$$

Usually the forward equations are easier to use in practical cases as they allow for an interpretation using probability fluxes (see below).

For any *initial distribution* $\{p_{(0)}(\mathbf{s}),\ \mathbf{s} \in S\}$ defined as

$$p_{(0)}(\mathbf{s}) = \mathbb{P}\{N(0) = \mathbf{s}\}, \quad \sum_{\mathbf{s} \in S} p_{(0)}(\mathbf{s}) = 1,$$

the *time-dependent distribution* $\{p(\mathbf{s},t),\ \mathbf{s} \in S\}$ defined as

$$p(\mathbf{s},t) = \mathbb{P}\{N(t) = \mathbf{s}\}, \quad \sum_{\mathbf{s} \in S} p(\mathbf{s},t) = 1,$$

can be obtained from the forward equations (1.11). Pre-multiplication of the forward equations (1.11) with the initial distribution $\{p_{(0)}(\mathbf{s}),\ \mathbf{s} \in S\}$ gives the following version of the *Kolmogorov forward equations* for the time-dependent distribution, for $\mathbf{s}' \in S$, $t \geq 0$,

$$\begin{cases} \dfrac{dp(\mathbf{s}',t)}{dt} = \displaystyle\sum_{\mathbf{s}\neq\mathbf{s}'} \big\{p(\mathbf{s},t)q(\mathbf{s},\mathbf{s}') - p(\mathbf{s}',t)q(\mathbf{s}',\mathbf{s})\big\}, \\ \\ p(\mathbf{s}',0) = p_{(0)}(\mathbf{s}'). \end{cases} \tag{1.12}$$

From the interpretation of the transition rates obtained from (1.9), for $\mathbf{s} \neq \mathbf{s}'$, the probability that the process jumps from $\mathbf{s}$ to $\mathbf{s}'$ in the interval $(t, t+h)$ is $p(\mathbf{s},t)q(\mathbf{s},\mathbf{s}')h + \mathrm{o}(h)$. Therefore, $p(\mathbf{s},t)q(\mathbf{s},\mathbf{s}')$ may be called the *probability flux* or *probability flow* from state $\mathbf{s}$ to state $\mathbf{s}'$. The forward equations now express that the rate of change of the *probability mass* of state $\mathbf{s}'$, $\frac{dp(\mathbf{s}',t)}{dt}$, equals the net probability flux from $S \setminus \{\mathbf{s}'\}$ to $\mathbf{s}'$. Thus the Kolmogorov forward equations express an intuitively obvious relation for the time-dependent distribution. A similar straightforward interpretation of the backward equations is not available.

Remark 1.1.3 (Uniformization) The *embedded Markov chain* of $\{N(t),\ t \in \mathbb{R}_0^+\}$ is the discrete-time Markov chain $\{Y(t),\ t \in \mathbb{N}_0\}$ at state space S with transition probabilities $p(\mathbf{s},\mathbf{s}') = q(\mathbf{s},\mathbf{s}')/q(\mathbf{s})$, $\mathbf{s},\mathbf{s}' \in S$, that follows the transitions of $\{N(t)\}$. If $q(\mathbf{s}) = q$ for all $\mathbf{s} \in S$ then $\{N(t)\}$ makes transitions at constant rate q and the state after k transitions is determined by the k-step transition probabilities of $\{Y(t)\}$.

If $\{N(t)\}$ is uniformizable with $\sup_{\mathbf{s}\in S} q(\mathbf{s}) \leq C < \infty$ we may define the discrete-time Markov chain $\{X(t),\ t \in \mathbb{N}_0\}$ at state space S with transition probabilities, for $\mathbf{s},\mathbf{s}' \in S$,

$$p_u(\mathbf{s},\mathbf{s}') = \begin{cases} q(\mathbf{s},\mathbf{s}')/C, & \text{if } \mathbf{s}' \neq \mathbf{s}, \\ 1 - q(\mathbf{s})/C, & \text{if } \mathbf{s}' = \mathbf{s}. \end{cases}$$

Note that $p_u(\mathbf{s},\mathbf{s}') = p(\mathbf{s},\mathbf{s}')q(\mathbf{s})/C$ for $\mathbf{s}' \neq \mathbf{s}$. Thus, $\{X(t)\}$ is an embedded Markov chain with transitions occurring at the event times of a Poisson process with rate C. In state $\mathbf{s} \in S$ with probability $1 - q(\mathbf{s})/C$ the Markov chain makes a self-transition, and with probability $q(\mathbf{s})/C$ the Markov chain makes a transition to another state, and this state is $\mathbf{s}'$ with probability $p(\mathbf{s},\mathbf{s}')$.[3] Let $P_u = (p_u(\mathbf{s},\mathbf{s}'),\ \mathbf{s},\mathbf{s}' \in S)$. Then for all $\mathbf{s},\mathbf{s}' \in S$ and $t > 0$

$$P(t) = \sum_{k=0}^{\infty} \frac{(Ct)^k}{k!} \mathrm{e}^{-Ct} \left(P_u\right)^k. \tag{1.13}$$

Uniformization transforms the continuous-time Markov chain $\{N(t)\}$ into the discrete-time Markov chain $\{X(t)\}$. Evaluation of $P(t)$ for fixed t via (1.13) is efficient as $(P_u)^k$ can be computed efficiently. Observe, however, that the sum must be evaluated

[3] Observe that for $\{X(t)\}$ the exit-time from state $\mathbf{s}$ is $\xi(\mathbf{s}) = \sum_{k=1}^{K(\mathbf{s})} X_k$, where $K(\mathbf{s})$ has a geometric distribution with succes probability $q(\mathbf{s})/C$, and the X_k, $k = 1, 2, \ldots$, are i.i.d. exponentially distributed with rate C. Hence $\xi(\mathbf{s})$ has an exponential distribution with rate $q(\mathbf{s})$.

for each t separately, so that uniformization does not provide an elegant construction for $P(t)$ for all t. See [4] for a survey on uniformization. □

The remaining part of this section considers the stationary or equilibrium behaviour of Markov chains. Throughout it will be assumed that the rate matrix satisfies (1.6), (1.7), is conservative and regular. Although these assumptions are not necessary for a large part of the discussion below, the discussion particularizes to conservative, regular Markov chains when the stationary distribution is related to the invariant distribution (the equilibrium solution of the Kolmogorov forward equations). When the assumptions are crucial to the theory they will be explicitly repeated.

If $P(t) = (p(\mathbf{s},\mathbf{s}';t),\ \mathbf{s},\mathbf{s}' \in S)$ is a transition matrix then the following limit exists for all $\mathbf{s},\mathbf{s}' \in S$

$$\lim_{t\to\infty} p(\mathbf{s},\mathbf{s}';t) = \upsilon(\mathbf{s},\mathbf{s}').$$

The matrix $\Upsilon = (\upsilon(\mathbf{s},\mathbf{s}'),\ \mathbf{s},\mathbf{s}' \in S)$ satisfies for all $\mathbf{s},\mathbf{s}' \in S$, $s > 0$,

$$\begin{aligned}\upsilon(\mathbf{s},\mathbf{s}') &= \sum_{\mathbf{s}''\in S} \upsilon(\mathbf{s},\mathbf{s}'')p(\mathbf{s}'',\mathbf{s}';s)\\ &= \sum_{\mathbf{s}''\in S} p(\mathbf{s},\mathbf{s}'';s)\upsilon(\mathbf{s}'',\mathbf{s}') = \sum_{\mathbf{s}''\in S} \upsilon(\mathbf{s},\mathbf{s}'')\upsilon(\mathbf{s}'',\mathbf{s}').\end{aligned}$$

Furthermore, $\upsilon(\mathbf{s},\mathbf{s}') \geq 0$ for all $\mathbf{s},\mathbf{s}' \in S$, and if $\upsilon(\mathbf{s},\mathbf{s}) \neq 0$ then $\sum_{\mathbf{s}'\in S} \upsilon(\mathbf{s},\mathbf{s}') = 1$. Therefore, Υ characterizes the stationary behaviour, but cannot be immediately associated with the stationary distribution. For Υ to be the stationary distribution additional assumptions guaranteeing that $\upsilon(\mathbf{s},\mathbf{s}) \neq 0$ must be made.

A state $\mathbf{s}$ is *absorbing* if the process cannot leave state $\mathbf{s}$, that is $p(\mathbf{s},\mathbf{s};t) = 1$ for all $t \geq 0$. For a non-absorbing state $\mathbf{s}$ the *recurrence-time* $\varepsilon(\mathbf{s})$ is defined as

$$\varepsilon(\mathbf{s}) = \inf\{t :\ t > \xi(\mathbf{s}),\ N(t) = \mathbf{s} \text{ if } N(0) = \mathbf{s}\},$$

where $\xi(\mathbf{s})$ is the exit-time from state $\mathbf{s}$. $\varepsilon(\mathbf{s})$ is the time it takes the process to return to state $\mathbf{s}$ if it starts at $\mathbf{s}$. A state $\mathbf{s}$ is called *recurrent* if recurrence to $\mathbf{s}$ is certain, i.e., if $\mathbb{P}\{\varepsilon(\mathbf{s}) < \infty\} = 1$. Otherwise it is *transient*. A recurrent state is *positive-recurrent* if $\mathbb{E}\{\varepsilon(\mathbf{s})\} < \infty$, that is if the expected return-time to state $\mathbf{s}$ is finite. Otherwise it is *null-recurrent*.

State $\mathbf{s}$ is *reachable* from state $\mathbf{s}'$ if passage from $\mathbf{s}$ to $\mathbf{s}'$ is possible, that is if $P(\mathbf{s},\mathbf{s}';t) > 0$ for some positive t. Two states *communicate* if each one is reachable from the other. A set $V \subset S$ is *closed* if the process cannot leave V, so that $q(\mathbf{s},\mathbf{s}') = 0$ for $\mathbf{s} \in V$, $\mathbf{s}' \in S\setminus V$. A set $V \subset S$ is *irreducible* if it is closed and all its states communicate. Two irreducible sets are disjoint, so the state space S can be decomposed into disjoint irreducible sets $V_1, V_2, \ldots$, and a non-irreducible set W. For the equilibrium behaviour of $\{N(t)\}$ the process may be analysed at each irreducible set separately. Therefore, without loss of generality, for equilibrium analysis the Markov chain may be assumed irreducible at S, that is S is an irreducible set. In this case all states $\mathbf{s} \in S$ are of the same type (transient, null-recurrent, positive-recurrent).

A measure $m = (m(\mathbf{s}),\ \mathbf{s} \in S)$ such that $0 \le m(\mathbf{s}) < \infty$ for all $\mathbf{s} \in S$ and $m(\mathbf{s}) > 0$ for some $\mathbf{s} \in S$ is called a *stationary measure* if for all $\mathbf{s}' \in S$, $t \ge 0$,

$$m(\mathbf{s}') = \sum_{\mathbf{s} \in S} m(\mathbf{s}) P(\mathbf{s}, \mathbf{s}'; t),$$

and is called an *invariant measure* if for all $\mathbf{s} \in S$,

$$\sum_{\mathbf{s}' \neq \mathbf{s}} \left\{ m(\mathbf{s}) q(\mathbf{s}, \mathbf{s}') - m(\mathbf{s}') q(\mathbf{s}', \mathbf{s}) \right\} = 0. \tag{1.14}$$

The relation between stationary and invariant measures is rather involved [10]. Based on regularity of the rate matrix a simple relation between these measures can be obtained. If the Markov chain is irreducible and positive-recurrent at S then there exists a unique (up to a multiplicative factor) stationary measure m which is positive ($m(\mathbf{s}) > 0$ for all $\mathbf{s} \in S$). From this result, for a regular and irreducible pure-jump process, if a finite mass ($\sum_{\mathbf{s} \in S} m(\mathbf{s}) < \infty$) invariant measure m exists then the process is positive-recurrent and m is the unique stationary measure. In the literature, an irreducible positive-recurrent process with invariant measure having finite mass is called *ergodic*.

Ergodicity is an important property of a process as it guarantees the existence of a unique *stationary distribution* π, that is a stationary measure summing to unity. Furthermore, if $\{N(t)\}$ is ergodic and π is the stationary distribution then $P(\mathbf{s}, \mathbf{s}'; t) \to \pi(\mathbf{s}')$ $(t \to \infty)$ for all $\mathbf{s}, \mathbf{s}' \in S$, or equivalently, $P(\mathbf{s}', t) \to \pi(\mathbf{s}')$ $(t \to \infty)$ for all $\mathbf{s}' \in S$ for any initial distribution $P_{(0)}$. As a consequence π may be called *equilibrium distribution*. Moreover, if $\{N(t)\}$ is ergodic then for any $f : S \to [0, \infty)$ such that $\sum_{\mathbf{s} \in S} f(\mathbf{s}) \pi(\mathbf{s}) < \infty$, with probability 1

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T f(N(t)) dt = \mathbb{E}_\pi \{ f(N) \} \equiv \sum_{\mathbf{s} \in S} \pi(\mathbf{s}) f(\mathbf{s}).$$

In particular, for $f(N(t)) = \mathbb{1}\{N(t) = \mathbf{s}\}$, the *indicator* of the event $\{N(t) = \mathbf{s}\}$, i.e., $\mathbb{1}\{A\} = 1$ if A occurs and 0 otherwise,

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T \mathbb{1}\{N(t) = \mathbf{s}\} dt = \pi(\mathbf{s}).$$

Thus $\pi(\mathbf{s})$ is the *long-run fraction of time* the process spends in state $\mathbf{s}$. The result may be extended to a function $h : S \times S \to [0, \infty)$ on the transitions of $\{N(t)\}$. If $\sum_{\mathbf{s}, \mathbf{s}' \in S} \pi(\mathbf{s}) q(\mathbf{s}, \mathbf{s}') h(\mathbf{s}, \mathbf{s}') < \infty$, then with probability 1

$$\lim_{T \to \infty} \frac{1}{T} \sum_{k=1}^{\infty} h(N(\tau_{k-1}), N(\tau_k)) \mathbb{1}(\tau_k \in (0, T]) = \sum_{\mathbf{s}, \mathbf{s}' \in S} \pi(\mathbf{s}) q(\mathbf{s}, \mathbf{s}') h(\mathbf{s}, \mathbf{s}'), \tag{1.15}$$

where $0 = \tau_0 < \tau_1 < \tau_2 < \cdots$ are the transition epochs of $\{N(t)\}$. Conditions for the process to be ergodic can be found, for example, in [3, 5].

The following theorem summarizes the relation between stationary, invariant and equilibrium distributions, and is the basis for determining the stationary or equilibrium distribution.

Theorem 1.1.4 (Equilibrium distribution) *Let* $\{N(t),\ t \geq 0\}$ *be a conservative, regular, irreducible continuous-time Markov chain.*

(i) If a positive finite mass invariant measure m exists then the Markov chain is positive-recurrent (ergodic). In this case $\pi = (\pi(\mathbf{s}),\ \mathbf{s} \in S)$ *defined as* $\pi(\mathbf{s}) = m(\mathbf{s})\left[\sum_{\mathbf{s}\in S} m(\mathbf{s})\right]^{-1}$, $\mathbf{s} \in S$, *is the unique stationary distribution and* π *is the equilibrium distribution, i.e., for all* $\mathbf{s}, \mathbf{s}' \in S$,

$$\lim_{t\to\infty} P(\mathbf{s}, \mathbf{s}'; t) = \pi(\mathbf{s}'),$$

independent of the initial distribution.

(ii) If a positive finite mass invariant measure does not exist then for all $\mathbf{s}, \mathbf{s}' \in S$,

$$\lim_{t\to\infty} P(\mathbf{s}, \mathbf{s}'; t) = 0.$$

The main result of Theorem 1.1.4 is that the stationary or equilibrium distribution can be obtained as the unique probability solution to (1.14). The equations (1.14) for $m = \pi$, the invariant distribution, can be obtained from the Kolmogorov forward equations. To this end note that the transition matrix $P(t)$ is the unique solution to (1.11). Furthermore, for a standard transition matrix $\frac{dP(\mathbf{s},\mathbf{s}';t)}{dt} \to 0$ $(t \to \infty)$ for all $\mathbf{s}, \mathbf{s}' \in S$. Thus for $t \to \infty$ (1.11) reduces to (1.14). Similar to the interpretation of (1.12), the equations (1.14) for $m = \pi$ can be interpreted as balancing the flow of probability mass on S. To this end $\pi(\mathbf{s})$ is interpreted as the probability mass at state $\mathbf{s}$ and $q(\mathbf{s}, \mathbf{s}')$ as the conductance of the direct path from $\mathbf{s}$ to $\mathbf{s}'$. Then $\pi(\mathbf{s})q(\mathbf{s}, \mathbf{s}')$ is the flow of probability mass from $\mathbf{s}$ to $\mathbf{s}'$ and (1.14) states that the flow of probability mass leaving $\mathbf{s}$ is balanced by the flow of probability mass entering $\mathbf{s}$. Therefore, (1.14) is usually referred to as *global balance equations*.

1.2 Three general solution concepts

This section introduces three approaches to obtain the stationary or equilibrium distribution that will form the basis for the analysis in Chapters **??**, **??**, and **??**, respectively: reversibility, partial balance, and Kelly's lemma.

Assumption 1.2.1 *Throughout this monograph, let* $\{N(t),\ t \geq 0\}$ *be a conservative, ergodic, continuous-time Markov chain with initial distribution* $\mathbb{P}(\mathbf{s}, 0) = \pi(\mathbf{s})$, $\mathbf{s} \in S$. *Let N be the random variable recording the state of* $\{N(t),\ t \geq 0\}$ *in equilibrium with distribution* π.

As is discussed in Section 1.1, under Assumption 1.2.1 the equilibrium distribution or stationary distribution, $\pi = (\pi(\mathbf{s}),\ \mathbf{s} \in S)$, can be obtained as the unique solution to the *global balance equations*

$$\sum_{\mathbf{s}'\neq\mathbf{s}} \left\{\pi(\mathbf{s})q(\mathbf{s},\mathbf{s}') - \pi(\mathbf{s}')q(\mathbf{s}',\mathbf{s})\right\} = 0, \quad \mathbf{s}\in S, \tag{1.16}$$

also called full balance equations or total balance equations as these equations express balance of the total probability flow in and out of each state $\mathbf{s}$. Solving the global balance equations is often very hard. Almost all solutions available in literature satisfy more stringent balance relations.

Note that under Assumption 1.2.1 the Markov chain $\{N(t),\ t\geq 0\}$ is stationary:

Theorem 1.2.2 *If Markov chain* $\{N(t),\ t\geq 0\}$ *has initial distribution* $\mathbb{P}(\mathbf{s},0) = \pi(\mathbf{s})$, $\mathbf{s}\in S$, *then* $\{N(t),\ t\geq 0\}$ *is stationary and* $\mathbb{P}(\mathbf{s},t) = \pi(\mathbf{s})$, $\mathbf{s}\in S$, *for all* $t\geq 0$.

1.2.1 Reversibility

The most stringent balance relation is transition balance. A Markov chain satisfies *transition balance* if for all $\mathbf{s},\mathbf{s}'\in S$ the transition rate from $\mathbf{s}$ to $\mathbf{s}'$ equals the transition rate from $\mathbf{s}'$ to $\mathbf{s}$, that is for all $\mathbf{s},\mathbf{s}'\in S$

$$q(\mathbf{s},\mathbf{s}') = q(\mathbf{s}',\mathbf{s}).$$

If a Markov chain satisfies transition balance then $m(\mathbf{s}) = 1$ for all $\mathbf{s}\in S$ satisfies the global balance equations (1.16). The equilibrium distribution π exists only if S is finite, in which case $\pi(\mathbf{s}) = |S|^{-1}$, $\mathbf{s}\in S$, with $|S|$ the *cardinality* of S.

A less restrictive form of balance often encountered in physical systems is detailed balance [7, 8, 14]. A Markov chain satisfies *detailed balance* if a distribution $\pi = (\pi(\mathbf{s}),\ \mathbf{s}\in S)$ exists that satisfies the *detailed balance equations* (1.17), for all $\mathbf{s},\mathbf{s}'\in S$,

$$\pi(\mathbf{s})q(\mathbf{s},\mathbf{s}') - \pi(\mathbf{s}')q(\mathbf{s}',\mathbf{s}) = 0. \tag{1.17}$$

Detailed balance is an important equilibrium concept. Summing (1.17) over all $\mathbf{s}'\in S$ yields that a distribution π that satisfies the detailed balance equations is the stationary distribution. The detailed balance equations state that the probability flow between each pair of states is balanced.

Detailed balance is related to reversibility. A stochastic process $\{N(t),\ -\infty < t < \infty\}$ is *reversible* if $(N(t_1), N(t_2), \ldots, N(t_n))$ has the same distribution as $(N(\tau - t_1), N(\tau - t_2), \ldots, N(\tau - t_n))$ for all $n\in\mathbb{N}$, $t_1,t_2,\ldots,t_n\in\mathbb{R}$, $\tau\in\mathbb{R}$. If a stochastic process is reversible and the direction of time is reversed, then the probabilistic behaviour of the process remains the same. The algebraic detailed balance property and the probabilistic reversibility property are the basis for the analysis in Chapter **??**.

Theorem 1.2.3 (Reversibility and detailed balance) *Let* $\{N(t),\ t\in T\}$, $T=\mathbb{R}$, *be a stationary Markov chain with transition rates* $q(\mathbf{s},\mathbf{s}')$, $\mathbf{s},\mathbf{s}'\in S$. $\{N(t)\}$ *is reversible* *if and only if there exists a distribution* $\pi = (\pi(\mathbf{s}),\ \mathbf{s} \in S)$ *that satisfies the detailed balance equations. When there exists such a distribution* π*, then* π *is the equilibrium distribution of* $\{N(t)\}$.

Proof. See Chapter **??**. □

1.2.2 Partial balance

Partial balance is less restrictive than detailed balance. Define for $\mathbf{s} \in S$ a collection of mutually exclusive sets $\{A_k(\mathbf{s}),\ k \in I(\mathbf{s})\}$, $I(\mathbf{s}) \subseteq \mathbb{N}$, such that $\bigcup_{k \in I(\mathbf{s})} A_k(\mathbf{s}) = S$. A Markov chain is *partially balanced over* $\{A_k(\mathbf{s}),\ k \in I(\mathbf{s})\}$ if a distribution $\pi = (\pi(\mathbf{s}),\ \mathbf{s} \in S)$ exists such that for all $\mathbf{s} \in S$, $k \in I(\mathbf{s})$,

$$\sum_{\mathbf{s}' \in A_k(\mathbf{s})} \left\{ \pi(\mathbf{s}) q(\mathbf{s}, \mathbf{s}') - \pi(\mathbf{s}') q(\mathbf{s}', \mathbf{s}) \right\} = 0. \tag{1.18}$$

The following result follows by summation of (1.18) over $k \in I(\mathbf{s})$.

Theorem 1.2.4 (Partial balance) *A distribution* $\pi = (\pi(\mathbf{s}),\ \mathbf{s} \in S)$ *satisfying the partial balance equations (1.18) is a stationary distribution.*

Chapter **??** explores partial balance as a means to obtain the equilibrium distribution of Markov chains.

1.2.3 Kelly's lemma

The transition rates of the time-reversed Markov chain are given in the following theorem.

Theorem 1.2.5 *Let* $\{N(t),\ t \in T\}$, $T = \mathbb{R}$, *be a stationary Markov chain with transition rates* $q(\mathbf{s}, \mathbf{s}')$, $\mathbf{s}, \mathbf{s}' \in S$, *and equilibrium distribution* $\pi = (\pi(\mathbf{s}),\ \mathbf{s} \in S)$. *The time-reversed process* $\{N(\tau - t),\ t \in T\}$ *is a conservative, regular, irreducible continuous-time stationary Markov chain with transition rates* $q^r(\mathbf{s}, \mathbf{s}')$, $\mathbf{s}, \mathbf{s}' \in S$, *given by*

$$q^r(\mathbf{s}, \mathbf{s}') = \frac{\pi(\mathbf{s}')}{\pi(\mathbf{s})} q(\mathbf{s}', \mathbf{s})$$

and the same equilibrium distribution $\pi = (\pi(\mathbf{s}),\ \mathbf{s} \in S)$.

Proof. See Chapter **??**. □

An important consequence of Theorem 1.2.5 is Kelly's lemma that will be the basis for the analysis in Chapter **??**.

Theorem 1.2.6 (Kelly's lemma) *Let* $\{N(t),\ t \in T\}$, $T = \mathbb{R}$, *be a stationary Markov chain with transition rates* $q(\mathbf{s},\mathbf{s}')$, $\mathbf{s},\mathbf{s}' \in S$. *If we can find a collection of numbers* $q^r(\mathbf{s},\mathbf{s}')$, $\mathbf{s},\mathbf{s}' \in S$, *such that*

$$\sum_{\mathbf{s}' \neq \mathbf{s}} q(\mathbf{s},\mathbf{s}') = \sum_{\mathbf{s}' \neq \mathbf{s}} q^r(\mathbf{s},\mathbf{s}'), \quad \mathbf{s} \in S,$$

and a distribution $\pi = (\pi(\mathbf{s}),\ \mathbf{s} \in S)$ *such that*

$$\pi(\mathbf{s}) q^r(\mathbf{s},\mathbf{s}') = \pi(\mathbf{s}') q(\mathbf{s}',\mathbf{s}), \quad \mathbf{s},\mathbf{s}' \in S,$$

then $q^r(\mathbf{s},\mathbf{s}')$, $\mathbf{s},\mathbf{s}' \in S$, *are the transition rates of the time-reversed Markov chain* $\{N(\tau - t),\ t \in T\}$ *and* $\pi = (\pi(\mathbf{s}),\ \mathbf{s} \in S)$, *is the equilibrium distribution of both Markov chains.*

Proof. See Chapter **??** □

References

1. Asmussen, S.: Applied probability and queues. Springer, New York (2010).
2. Chung, K..L.: Markov chains with stationary transition probabilities. Springer Berlin, Heidelberg (1967).
3. Cohen, J.W.: The single server queue. North-Holland, Amsterdam (1982).
4. Dijk, N.M. van, Brummelen, S.P.J. van, Boucherie, R.J.: Uniformization: basics, extensions and applications. Perf. Eval. **118**, 8–32 (2018).
5. Foster, F.G. On the stochastic matrices associated with certain queueing processes. Ann. Math. Statist. **24**, 355–360 (1953).
6. Gross, D., Shortle, J.F., Thompson, J.M., Harris, C.M.: Fundamentals of Queueing Theory. Wiley (2011).
7. Kampen, N.G. van: Stochastic processes in physics and chemistry. North-Holland, Amsterdam (1981)
8. Kelly, F.P.: Reversibility and stochastic networks. Wiley, New York (1979).
9. Pollett, P.K.: A note on the classification of Q-processes when Q is not regular. J. Appl. Prob. **27**, 278–290 (1990)
10. Pollett, P.K.: Invariant measures for Q-processes when Q is not regular. Adv. Appl. Prob. **23**, 277–292 (1991)
11. Reuter, G.E.H.: Denumerable Markov processes and the associated contraction semigroups on l. Acta Math. **97**, 1–46 (1957)
12. Serfozo, R.F.: Introduction to stochastic networks. Springer-Verlag, New York (1999).
13. Walrand, J.: An introduction to queueing networks. Prentice-Hall, New Jersey. (1988)
14. Whittle, P.: Systems in stochastic equilibrium. Wiley, New York (1986) .
15. Wolff, R.W.: Stochastic modeling and the theory of queues. Prentice Hall, Englewood Cliffs (1989).