Chapter 4
Renewal and Regenerative Processes

4.1 Basic Theory of Renewal Processes

Let $\{N(t),\ t \geq 0\}$ be a nonnegative-integer-valued stochastic process that counts the occurrences of a given event. That is, $N(t)$ is the number of events in the time interval $[0, t)$. For example, $N(t)$ can be the number of bulb replacements in a lamp that is continuously on, and the dead bulbs are immediately replaced (Fig. 4.1).

Let $0 \leq t_1 \leq t_2 \leq \ldots$ be the times of the occurrences of consecutive events and $t_0 = 0$ and $T_i = t_i - t_{i-1}$, $i = 1, 2, 3, \ldots$ be the time intervals between consecutive events.

Definition 4.1. $t_1 \leq t_2 \leq \ldots$ is a **renewal process** if the time intervals between consecutive events $T_i = t_i - t_{i-1}$, $i = 2, 3, \ldots$, are independent and identically distributed (i.i.d.) random variables with CDF

$$F(x) = \mathbf{P}(T_k \leq x), \quad k = 1, 2, \ldots .$$

The nth event time, t_n, $n = 1, 2, \ldots$, is referred to as the n**th renewal point** or renewal time. According to the definition, the first time interval might have a different distribution.

We assume that $F(0) = 0$ and $F(+0) = \mathbf{P}(T_k = 0) < 1$. In this case

$$t_0 = 0, \quad t_n = T_1 + \ldots + T_n, \quad n = 1, 2, \ldots,$$

$$N(0) = 0, \quad N(t) = \sup\{n : t_n \leq t,\ n \geq 0\} = \sum_{i=1}^{\infty} \mathcal{I}_{\{t_i \leq t\}}, \quad t > 0.$$

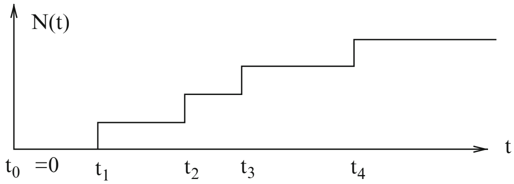
Remark 4.2. $\{N(t),\ t \geq 0\}$ and $\{t_n,\ n \geq 1\}$ mutually and univocally determine each other because for arbitrary $t \geq 0$ and $k \geq 1$ we have

$$N(t) \geq k \quad \Leftrightarrow \quad t_k \leq t.$$

L. Lakatos et al., *Introduction to Queueing Systems with Telecommunication Applications*,

DOI 10.1007/978-1-4614-5317-8_4,

Fig. 4.1 Renewal process



Definition 4.3. When $\mathbf{P}(T_k \le x) = F(x), \quad k = 2, 3, \ldots,$ but $F_1(x) = \mathbf{P}(T_1 \le x) \not\equiv F(x)$, the process is referred to as a **delayed renewal process**.

Remark 4.4. $T_1, T_2, \ldots$ are i.i.d. random variables and from $t_n = T_1 + \ldots + T_n$, and we can compute the distribution of the time of the nth event $F^{(n)}(x) = \mathbf{P}(t_n \le x)$ using the convolution formula

$$F^{(n)}(x) = \int_0^\infty F^{(n-1)}(x-y)\mathrm{d}F(y) = \int_0^x F^{(n-1)}(x-y)\mathrm{d}F(y), \quad n \ge 2, \quad x \ge 0,$$

$$F^{(n)}(x) \equiv 0, \quad \text{if } x \le 0 \text{ and } n \ge 1.$$

Starting from $F^{(1)}(x) = F_1(x)$ the same formula applies in the delayed case.

Definition 4.5. The function $H(t) = \mathbf{E}(N(t)), \quad t \ge 0$, is referred to as a **renewal function**.

One of the main goals of renewal theory is the analysis of the renewal function $H(t)$ and the description of its asymptotic behavior. Below we discuss the related results for regular renewal processes. The properties of delayed renewal processes are similar, and we do not provide details on them here. We will show that the law of large numbers and the central limit theorem hold for the renewal process (see also Ch. 5. in [48]).

Theorem 4.6. *If* $\{T_n, \quad n = 1, 2, \ldots\}$ *is a series of nonnegative i.i.d. random variables and* $\mathbf{P}(T_1 = 0) < 1$, *then there exists* $\rho_0 > 0$ *such that for all* $0 < \rho < \rho_0$ *and* $t \ge 0$

$$\mathbf{E}\left(\mathrm{e}^{\rho N(t)}\right) < \infty$$

holds.

Proof (Proof 1 of Theorem 4.6). From the Markov inequality (Theorem 1.35) we have

$$\begin{aligned}\mathbf{E}\left(\mathrm{e}^{\rho N(t)}\right) &= \sum_{k=0}^\infty \mathrm{e}^{\rho k}\mathbf{P}(N(t) = k) \le \sum_{k=0}^\infty \mathrm{e}^{\rho k}\mathbf{P}(N(t) \ge k) \\ &= \sum_{k=0}^\infty \mathrm{e}^{\rho k}\mathbf{P}(t_k \le t) \le \sum_{k=0}^\infty \mathrm{e}^{\rho k}\mathrm{e}^{t-k\kappa} = \mathrm{e}^t(1-\mathrm{e}^{-(\kappa-\rho)})^{-1},\end{aligned}$$

where $\rho < \rho_0 = \kappa$, $\kappa = \log \frac{1}{h}$, and $h = \mathbf{E}\left(\mathrm{e}^{-T_1}\right)$. Additionally, $h < 1$ because $F(0) = 0$ and $\mathbf{P}(T_1 = 0) < 1$. □

Proof (Proof 2 of Theorem 4.6). According to the condition of the theorem, there exist ϵ and δ positive numbers such that $\mathbf{P}(T_k \geq \delta) > \epsilon$. Introducing $\{T_k' = \delta \mathcal{I}_{\{T_k \geq \delta\}},\ k = 1, 2, \ldots\}$ (where T_n', $n = 1, 2, \ldots$, is a series of i.i.d. random variables) and the related $\{N'(t),\ t \geq 0\}$ renewal process we have that $\mathbf{P}\left(T_k' \leq T_k\right) = 1$, $k \geq 1$, and consequently $\mathbf{P}(N'(t) \geq N(t)) = 1$, $t \geq 0$. The distribution of $N'(t)$ is negative binomial with the parameter $p = \mathbf{P}\left(T_k' \geq \delta\right)$ and order $r = \lfloor t/\delta \rfloor$,

$$\mathbf{P}\left(N'(t) = k + r\right) = \binom{k + r - 1}{r - 1} p^r (1 - p)^k, \quad k = 0, 1, 2, \ldots,$$

from which the statement of the theorem follows. □

Corollary 4.7. *All moments of $N(t)$ $(t \geq 0)$ are finite, and the renewal function $H(t)$ is also finite for all $t \geq 0$.*

Proof. The corollary comes from Theorem 4.6 and the inequality $x^n \leq n!\mathrm{e}^x$, $n \geq 1,\ x \geq 0$. □

Before conducting an analysis of the renewal function we recall some properties of convolution.

Let $A(t)$ and $B(t)$ be monotonically nondecreasing right-continuous functions such that $A(0) = B(0) = 0$.

Definition 4.8. The **convolution** of $A(t)$ and $B(t)$ [denoted by $A * B(t)$] is

$$A * B(t) = \int_0^t B(t - y)\mathrm{d}A(y), \quad t \geq 0.$$

Lemma 4.9. $A * B(t) = B * A(t)$.

Proof. From $B(0) = 0$ we have $B(t - y) = \int_0^{t-y} \mathrm{d}B(s)$, and consequently

$$\begin{aligned} A * B(t) &= \int_0^t \left\{ \int_0^{t-y} \mathrm{d}B(s) \right\} \mathrm{d}A(y) = \int_0^t \int_0^t \mathcal{I}_{\{s<t-y\}} \mathrm{d}A(y)\mathrm{d}B(s) \\ &= \int_0^t \int_0^t \mathcal{I}_{\{y<t-s\}} \mathrm{d}A(y)\mathrm{d}B(s) = \int_0^t \left\{ \int_0^{t-s} \mathrm{d}A(y) \right\} \mathrm{d}B(s) \\ &= B * A(t). \end{aligned}$$

□

Remark 4.10. The definition of the renewal function $H(t)$

$$H(t) = \mathbf{E}(N(t)) = \mathbf{E}\left(\sum_{i=1}^{\infty} \mathcal{I}_{\{t_i \le t\}}\right) = \sum_{i=1}^{\infty} \mathbf{P}(T_1 + \ldots + T_i \le t)$$

immediately determines the relation between the renewal function and the order k of the convolutions of the event time distribution

$$H(t) = \sum_{k=1}^{\infty} F^{(k)}(t).$$

Theorem 4.11. *If $\{T_n,\ n = 1, 2, \ldots\}$ is a series of i.i.d. random variables and $\mathbf{P}(T_1 < 0) = 0,\ \ \mathbf{P}(T_1 = 0) < 1$, then $H(t)$ satisfies the* ***renewal equation***

$$H(t) = F(t) + \int_0^t H(t-y)\mathrm{d}F(y), \quad t \ge 0.$$

Proof. According to Remarks 4.4 and 4.10, the renewal function can be written as

$$\begin{aligned} H(t) &= F^{(1)}(t) + \sum_{k=1}^{\infty} \int_0^t F^{(k)}(t-y)\mathrm{d}F(y) \\ &= F(t) + \int_0^t \left(\sum_{k=1}^{\infty} F^{(k)}(t-y)\right) \mathrm{d}F(y) \\ &= F(t) + \int_0^t H(t-y)\mathrm{d}F(y)\,, \end{aligned}$$

where the order of the summation and the integration are interchanged based on Corollary 4.7. □

In the case of a delayed renewal process, the renewal function is denoted by $H_1(t)$, and the same composition holds as for the regular renewal process (Remark 4.10)

$$H_1(t) = \sum_{k=1}^{\infty} F^{(k)}(t),\ t \ge 0, \qquad (F^{(k)}(t) = \mathbf{P}(t_k \le t)),$$

but in this case $F_1 \not\equiv F$.

Theorem 4.12. *The renewal function can be written in the following forms:*

$$H_1(t) = F_1(t) + H_1 * F(t) = F_1(t) + F * H_1(t),$$
$$H_1(t) = F_1(t) + H * F_1(t) = F_1(t) + F_1 * H(t),$$
$$H(t) = F(t) + H * F(t) = F(t) + F * H(t).$$

Renewal Equations

Definition 4.13. An integral equation of the type

$$A(t) = a(t) + \int_0^t A(t-x)\mathrm{d}F(x),\ t \geq 0,$$

where $a(t)$ and $F(t)$ are known functions and $A(t)$ is unknown, is referred to as a **renewal equation** (see also Theorem 4.1 of Ch. 5. in [48]).

Theorem 4.14. *If $a(t),\ t \geq 0$, is a bounded real function that is Riemann–Stieltjes integrable according to $H(t)$ over any finite interval, then there uniquely exists the function $A(t),\ t \geq 0$, which is finite over any finite interval and satisfies the renewal equation*

$$(i) \qquad A(t) = a(t) + \int_0^t A(t-x)\mathrm{d}F(x), \quad t \geq 0,$$

and furthermore it satisfies

$$(ii) \qquad A(t) = a(t) + \int_0^t a(t-x)\mathrm{d}H(x), \quad t \geq 0,$$

where $H(t) = \sum_{k=1}^{\infty} F^{(k)}(t),\ t \geq 0$, is the renewal function.

Proof. First we show that the function $A(t)$, $t \geq 0$, defined by equation (ii), is (a) bounded on the $[0, T]$ interval for all $T > 0$ and (b) satisfies (i). Next we prove that (c) all solutions of (i) that are bounded on $[0, T]$ can be given in form (ii), i.e., the solution is unique.

(a) Since $a(t)$ is bounded and $H(t)$ is monotonically nondecreasing, we have

$$\sup_{0\le t\le T}|A(t)| \le \sup_{0\le t\le T}|a(t)| + \int_0^T [\sup_{0\le y\le T}|a(y)|\mathrm{d}H(x)$$
$$\le \sup_{0\le t\le T}|a(t)|(1+H(T)) < \infty.$$

(b) Furthermore, we have

$$\begin{aligned} A(t) = a(t) + H * a(t) &= a(t) + \left(\sum_{k=1}^{\infty} F^{(k)}\right) * a(t) \\ &= a(t) + F * a(t) + \left(\sum_{k=2}^{\infty} F^{(k)}\right) * a(t) \\ &= a(t) + F * [a(t) + \left(\sum_{k=1}^{\infty} F^{(k)}\right) * a(t)] \\ &= a(t) + F * A(t). \end{aligned}$$

(c) We prove this by successive approximation. According to equation (i), $A = a + F * A$. Substituting this into (i) we have

$$\begin{aligned} A = a(t) + F * (a + F * A) &= a + F * a + F * (F * A) \\ &= a + F * a + F^{(2)} * A. \end{aligned}$$

Continuously substituting equation (i) we obtain for $n \ge 1$ that

$$A = a + F * a + F^{(2)} * (a + F * A) = \ldots = a + \sum_{k=1}^{n-1}(F^{(k)} * a) + F^{(n)} * A.$$

Since $A(t)$ is bounded on every finite interval according to (a), $F^{(n)}(0-) = 0$, $F^{(n)}(y)$ is monotonically nondecreasing, and $F^{(n)}(t) \to 0,\ n \to \infty$, for all fixed t, we have that for a fixed t

$$|F^{(n)} * A(t)| = \left|\int_0^t A(t-y)\mathrm{d}F^{(n)}(y)\right| \le \sup_{0\le y\le t}|A(t-y)|F^{(n)}(t) \to 0,\ n \to \infty.$$

From the fact that $a(t)$ is bounded it follows that

$$\lim_{n\to\infty}\left(\sum_{k=1}^{n-1} F^{(k)}\right) * a(t) = \left(\sum_{k=1}^{\infty} F^{(k)}\right) * a(t) = H * a(t),$$

and consequently

$$A(t) = a(t) + \lim_{n\to\infty}\left[\left(\sum_{k=1}^{n-1} F^{(k)}\right) * a(t) + F^{(n)}A(t)\right] = a(t) + H * a(t).$$

This means that if A is a bounded solution of (i), then it is identical with (ii).

□

Analysis of the Renewal Function

One of the main goals of the renewal theorem is the analysis of the renewal function. According to Theorem 4.12, in the case of delayed renewal processes the renewal function $H_1(t)$ can be obtained from $F_1(t)$ and $H(t)$. In the rest of this section we focus on the analysis of the renewal function of an ordinary renewal process, $H(t)$, that is, $F_k = F,\ k \geq 1$. During the subsequent analysis we assume that $F(t)$ is such that $F(0-) = 0$ and $F(0+) < 1$.

Theorem 4.15 (*Elementary renewal theorem*). *There exists the limit*

$$\lim_{t\to\infty} \frac{H(t)}{t} = \frac{1}{\mathbf{E}(T_1)},$$

and it is 0 *if* $\mathbf{E}(T_1) = \infty$.

Definition 4.16. The random variable X has a lattice distribution if there exists $d > 0$ and $r \in R$ such that the random variable $\frac{1}{d}(X - r)$ is distributed on the integer numbers, that is, $\mathbf{P}\left(\frac{1}{d}(X - r) \in \mathbf{Z}\right) = 1$. The largest d with that property is referred to as the step size of the distribution.

Remark 4.17. If X has a lattice distribution with step size d, then

$$d = \min\{s : |\psi(2\pi/s)| = 1\},$$

where $\psi(u) = \mathbf{E}\left(e^{iuX}\right),\ u \in R$, denotes the characteristic function of X. In this case, $|\psi(u)| < 1$ if $0 < |u| < 2\pi/d$. If the distribution of X is not lattice, then $|\psi(u)| < 1$ if $u \neq 0$.

Theorem 4.18 (*Blackwell's theorem*). *If $F(t)$ is a lattice distribution with step size d, then*

$$\lim_{n\to\infty} q_n = \frac{d}{\mathbf{E}(T_1)},$$

where $q_n = H(nd) - H((n-1)d)$*. If* $F(t)$ *is not a lattice distribution, then for all* $h > 0$

$$\lim_{t\to\infty} (H(t+h) - H(t)) = \frac{h}{\mathbf{E}(T_1)}$$

holds.

The following theorems require the introduction of *direct Riemann integrability*, which is more strict than Riemann integrability.

Let g be a nonnegative function on the interval $[0, \infty)$ and

$$s(\delta) = \delta \sum_{n=1}^{\infty} \inf\{g(x) :\ (n-1)\delta \le x \le n\delta\},$$

$$S(\delta) = \delta \sum_{n=1}^{\infty} \sup\{g(x) :\ (n-1)\delta \le x \le n\delta\}.$$

Definition 4.19. The function g is *directly Riemann integrable* if $s(\delta)$ and $S(\delta)$ are finite for all $\delta > 0$ and

$$\lim_{\delta\to 0} [S(\delta) - s(\delta)] = 0.$$

Remark 4.20. If the function g is directly Riemann integrable, then g is bounded, and the limit of $s(\delta)$ and $S(\delta)$ at $\delta \to 0$ is equal to the infinite Riemann integral, that is,

$$\lim_{\delta\to 0} s(\delta) = \lim_{\delta\to 0} S(\delta) = \int_0^{\infty} g(x)\mathrm{d}x = \lim_{y\to\infty} \int_0^{y} g(x)\mathrm{d}x.$$

Sufficient and necessary conditions for direct Riemann integrability:

(a) There exists $\delta > 0$ such that $S(\delta) < \infty$.
(b) g is almost everywhere continuous along the real axes according to the Lebesgue measure (that is, equivalent to Riemann integrability on every finite interval).

Sufficient conditions for direct Riemann integrability:

g is bounded and has a countable number of discontinuities, and at least either condition (a) or (b) holds:

(a) g equals 0 apart from a finite interval.
(b) g is monotonically decreasing and $\int_0^{\infty} g(x)\mathrm{d}x < \infty$.

Theorem 4.21 (*Smith's renewal theorem*). *If* $g(x) \ge 0$*,* $x \ge 0$*, is a nonincreasing directly Riemann integrable function on the interval* $[0, \infty)$*, then for* $t \to \infty$ *one of the following identities holds:*

(a) If F is a nonlattice distribution, then

$$\lim_{t\to\infty} H * g(t) = \lim_{t\to\infty} \int_0^t g(t-u)\mathrm{d}H(u) = \frac{1}{\mathbf{E}(T_1)} \int_0^\infty g(u)\mathrm{d}u.$$

(b) If F is a lattice distribution with step size d, then

$$\lim_{n\to\infty} H*g(x+nd) = \lim_{n\to\infty} \int_0^{x+nd} g(x+nd-u)\mathrm{d}H(u) = \frac{d}{\mathbf{E}(T_1)} \sum_{k=0}^\infty g(x+kd).$$

Remark 4.22. Blackwell's theorem (Theorem 4.18) follows from Smith's renewal theorem (Theorem 4.21) assuming that $g(u) = \mathcal{I}_{\{0<u\le h\}}$. The reverse direction is an implicit consequence of the proof of Blackwell's theorem provided by Feller in [31].

Before proving Theorem 4.21 we collect some simple properties of the renewal function $H(t)$.

Lemma 4.23. *H is monotonically nondecreasing and continuous from the right.*

Proof. $F^{(k)}(t)$ is monotonically nondecreasing and continuous from the right for all $k \ge 1$, and the series $\sum_{k=1}^\infty F^{(k)}(t)$ is uniformly convergent on every finite interval, from which the lemma follows. □

Lemma 4.24. *The function H is subadditive, that is,*

$$H(t+h) \le H(t) + H(h) \tag{4.1}$$

for $t, h \ge 0$.

Proof. Since $H(0) = 0$, it is enough to consider the case where $t, h > 0$. Let $n(t) = \inf\{n : t_n \ge t,\ n \ge 0\}$. If $t_n \le t$ for all $n \ge 0$, then let $n(t) = \infty$. This case can occur only on a set with measure 0.

Due to the fact that $\mathbf{P}(T_1 = 0)$ might be positive, the relation of $n(t)$ and $N(t)$ is not deterministic. It holds that $n(t) \ge N(t) + 1$ and the right continuity of $N(t)$ implies $N(t_{n(t)}) = N(t),\ t \ge 0$. Using that we have

$$N(t+h) - N(t) = N(t+h) - N(t_{n(t)}) \le N(t_{n(t)} + h) - N(t_{n(t)}),$$

and using the total probability theorem, we obtain

$$\begin{aligned}
\mathbf{E}(N(t+h) - N(t)) &\le \mathbf{E}\big(N(t_{n(t)} + h) - N(t_{n(t)})\big) \\
&= \sum_{k=1}^\infty \mathbf{E}\big(N(t_{n(t)} + h) - N(t_{n(t)}) | n(t) = k\big)\mathbf{P}(n(t) = k) \\
&= \sum_{k=1}^\infty \mathbf{E}(N(t_k + h) - N(t_k) | n(t) = k)\mathbf{P}(n(t) = k).
\end{aligned}$$

Since t_k is a renewal point, the conditional expected value in the last summation does not depend on the condition

$$\mathbf{E}\left(N(t_k+h)-N(t_k)|n(t)=k\right)=\mathbf{E}\left(N(h)-N(0)\right)=\mathbf{E}\left(N(h)\right),$$

and in this way we have

$$\begin{aligned}\mathbf{E}\left(N(t+h)-N(t)\right) &\leq \sum_{k=1}^{\infty}\mathbf{E}\left(N(h)\right)\mathbf{P}\left(n(t)=k\right)\\ &= \mathbf{E}\left(N(h)\right)\sum_{k=1}^{\infty}\mathbf{P}\left(n(t)=k\right)=\mathbf{E}\left(N(h)\right),\end{aligned}$$

from which the lemma follows. □

Lemma 4.25. *For the renewal function H the following inequality holds:*

$$H(t)\leq H(1)(1+t),\quad t\geq 0. \tag{4.2}$$

Proof. From the previous statement and the monotonicity of H

$$\begin{aligned}H(t)&\leq H(\lfloor t\rfloor+1)\leq H(1)+H(\lfloor t\rfloor)\leq H(1)+(H(1)+H(\lfloor t\rfloor-1))\leq\\ &\leq\ldots\leq H(1)+\lfloor t\rfloor H(1)\leq H(1)+tH(1)=H(1)(1+t).\end{aligned}$$

□

Remark 4.26. The nonnegative subadditive functions can be estimated from the preceding expression by a linear function.

Lemma 4.27. *For arbitrary $\lambda>0$ the Laplace–Stieltjes transform $H^{\sim}(\lambda)=\int_0^{\infty}\mathrm{e}^{-\lambda t}\mathrm{d}H(t)$, $\lambda\geq 0$, of the function H can be represented in the Laplace–Stieltjes transform as*

$$H^{\sim}(\lambda)=(1-\varphi^{\sim}(\lambda))^{-1},$$

where $\varphi^{\sim}(\lambda)=\mathbf{E}\left(\mathrm{e}^{-\lambda T_1}\right)$ is the Laplace–Stieltjes transform of the distribution function F.

Proof. For $\lambda>0$ there obviously exists $H^{\sim}(\lambda)$ since, according to Eqs. (1.3) and (4.2),

$$H^{\sim}(\lambda)=\lambda\int_0^{\infty}\mathrm{e}^{-\lambda t}H(t)\mathrm{d}t\leq\lambda H(1)\int_0^{\infty}\mathrm{e}^{-\lambda t}(1+t)\mathrm{d}t<\infty.$$

It is clear that

$$\int_0^{\infty}\mathrm{e}^{-\lambda t}\mathrm{d}N(t)=\sum_{k=0}^{\infty}\mathrm{e}^{-\lambda t_k}=1+\sum_{k=1}^{\infty}\prod_{i=1}^{k}\mathrm{e}^{-\lambda T_i}.$$

Using this equality we obtain

$$\begin{aligned}\mathbf{E}\left(\int_0^\infty \mathrm{e}^{-\lambda t}\mathrm{d}N(t)\right) &= \mathbf{E}\left(\lambda\int_0^\infty N(t)\mathrm{e}^{-\lambda t}\mathrm{d}t\right)\\ &= \lambda\int_0^\infty H(t)\mathrm{e}^{-\lambda t}\mathrm{d}t = \int_0^\infty \mathrm{e}^{-\lambda t}\mathrm{d}H(t) = (h(\lambda) =)\\ &= \mathbf{E}\left(1+\sum_{k=1}^\infty\prod_{i=1}^k \mathrm{e}^{-\lambda T_i}\right) = 1+\sum_{k=1}^\infty(\varphi(\lambda))^k = \frac{1}{1-\varphi(\lambda)},\end{aligned}$$

where $0 < \varphi(\lambda) < 1$ if $\lambda > 0$. □

Proof of Elementary Renewal Theorem. First we prove that the limit exists. If $t \geq 1$, then we have that $0 \leq \frac{H(t)}{t} \leq \frac{1+t}{t}H(1) \leq 2H(1)$ is bounded. Let $c = \inf\limits_{t\geq 1}\frac{H(t)}{t}$. Then for arbitrary $\epsilon > 0$ there exists a number $t_0 > 0$ such that

$$\frac{H(t_0)}{t_0} < c + \epsilon.$$

Moreover, for all integers $k \geq 1$ and $\tau \geq 0$

$$\frac{H(kt_0+\tau)}{kt_0+\tau} \leq \frac{kH(t_0)+H(\tau)}{kt_0} \leq c+\epsilon+\frac{H(\tau)}{kt_0},$$

and consequently

$$\limsup_{t\to\infty}\frac{H(t)}{t} \leq c+\epsilon,$$

and

$$c \leq \liminf_{t\to\infty}\frac{H(t)}{t} \leq \limsup_{t\to\infty}\frac{H(t)}{t} \leq c$$

follows. We have proved the existence of the limit.

Using the preceding expression for the Laplace–Stieltjes transform $h(\lambda)$,

$$\int_0^\infty \mathrm{e}^{-\lambda t}H(t)\mathrm{d}t = \frac{1}{\lambda}\int_0^\infty \mathrm{e}^{-\lambda t}\mathrm{d}H(t) = \frac{1}{\lambda}h(\lambda) = \frac{1}{\lambda}\frac{1}{1-\varphi(\lambda)},$$

and we obtain

$$\frac{\lambda}{1-\varphi(\lambda)} = \lambda^2\int_0^\infty \mathrm{e}^{-\lambda t}H(t)\mathrm{d}t = \int_0^\infty \mathrm{e}^{-t}\lambda H\left(\frac{t}{\lambda}\right)\mathrm{d}t. \tag{4.3}$$

By means of the relation for the derivative of the Laplace–Stieltjes transform

$$\lim_{\lambda\to+0}\frac{\lambda}{1-\varphi(\lambda)} = \lim_{\lambda\to+0}\left(\mathbf{E}\left(\frac{1-\mathrm{e}^{-\lambda T_1}}{\lambda}\right)\right)^{-1} = \begin{cases} 0, & if\ \ \mathbf{E}\,(T_1)=\infty,\\ \frac{1}{\mathbf{E}(T_1)}, & if\ \ \mathbf{E}\,(T_1)<\infty. \end{cases}$$

On the other hand, in the case $0<\lambda\le 1$, we can give a uniform upper estimation for the integrand in Eq. (4.3):

$$\mathrm{e}^{-t}\lambda H\left(\frac{t}{\lambda}\right) \le \mathrm{e}^{-t}\lambda\left(1+\frac{t}{\lambda}\right)H(1) \le \mathrm{e}^{-t}(1+t)H(1);$$

furthermore,

$$\lim_{\lambda\to+0}\lambda H\left(\frac{t}{\lambda}\right) = t\lim_{\lambda\to+0}\frac{H\left(\frac{t}{\lambda}\right)}{\frac{t}{\lambda}} = tc,$$

so from the Lebesgue majorated convergence theorem

$$\lim_{\lambda\to+0}\int_0^\infty \mathrm{e}^{-t}\lambda H\left(\frac{t}{\lambda}\right)\mathrm{d}t = \int_0^\infty \mathrm{e}^{-t}ct\,\mathrm{d}t = c.$$

Summing up the previous results we obtain

$$c = \lim_{\lambda\to+0}\frac{\lambda}{1-\varphi(\lambda)} = \begin{cases} 0 & if\ \ \mathbf{E}\,(T_1)=\infty,\\ \frac{1}{\mathbf{E}(T_1)} & if\ \ \mathbf{E}\,(T_1)<\infty. \end{cases}$$

□

4.1.1 Limit Theorems for Renewal Processes

Theorem 4.28. *Let* $0 < \mathbf{E}\,(T_1) = \mu < \infty$*; then the following stochastic convergence holds:*

$$\frac{N(t)}{t}\xrightarrow{P}\frac{1}{\mu}, \qquad t\to\infty.$$

Proof. The proof of Theorem 4.28 is based on the relation

$$\{N(t)>k\} = \{t_k\le t\}$$

from Comment 4.2. Let us estimate the probability $\mathbf{P}\,(|N(t)/t-1/\mu|>\epsilon)$ for arbitrary $\epsilon>0$. Let $n=n(t)=\lfloor t/\mu+\epsilon t\rfloor$; then

$$\begin{aligned}
\mathbf{P}\left(\frac{N(t)}{t}-\frac{1}{\mu}>\epsilon\right) &= \mathbf{P}\left(N(t)>\frac{t}{\mu}+\epsilon t\right) \leq \mathbf{P}\left(N(t)>n\right) \\
&= \mathbf{P}\left(t_n \leq t\right) = \mathbf{P}\left(\frac{t_n}{n} \leq \frac{t}{\lfloor t/\mu+\epsilon t \rfloor}\right) \\
&\leq \mathbf{P}\left(\frac{t_n}{n} \leq \frac{t}{t/\mu+\epsilon t-1}\right) \\
&= \mathbf{P}\left(\frac{t_n}{n} \leq \frac{1}{1/\mu+\epsilon-1/t}\right) \\
&\leq \mathbf{P}\left(\frac{t_n}{n} \leq \frac{\mu}{1+\mu\epsilon/2}\right) \quad \text{if } t \geq 2/\epsilon,
\end{aligned}$$

which by Bernoulli's law of large numbers tends to 0 for the sequence t_n, $n = 1, 2, \ldots$, as $t \to \infty$. The probability $\mathbf{P}\left(N(t)/t - 1/\mu < -\epsilon\right)$ is estimated in a similar way. □

Remark 4.29. By the strong law of large numbers, $\frac{t_k}{k} \to \mu$, $k \to \infty$, with probability 1. Using this fact one can prove that with probability 1

$$\frac{N(t)}{t} \to \frac{1}{\mu}, \quad t \to \infty.$$

The convergence with probability 1 remains valid for delayed renewal processes if the first time interval is finite with probability 1.

Theorem 4.30. *If* $\mathbf{E}\left(T_1\right) = \mu > 0,\ \mathbf{D}^2\left(T_1\right) = \sigma^2 < \infty$, *then as* $t \to \infty$

$$\lim_{t\to\infty} \mathbf{P}\left(\frac{N(t)-t/\mu}{\sqrt{t\sigma^2/\mu^3}} \leq x\right) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-u^2/2} \mathrm{d}u.$$

Proof. Let x be a real number and denote

$$r(t) = \lfloor t/\mu + x\sqrt{t\sigma^2/\mu^3} \rfloor.$$

Note that $r(t) \geq 1$ if $\sqrt{t} + x\sigma/\sqrt{\mu} - \mu/\sqrt{t} \geq 0$. Since $r(t) \to \infty$ as $t \to \infty$, then from the central limit theorem it follows that for all $x \in \mathbb{R}$

$$\mathbf{P}\left(\frac{t_{r(t)} - \mu r(t)}{\sigma\sqrt{r(t)}} \leq x\right) \to \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-u^2/2} du, \quad t \to \infty. \tag{4.4}$$

Using the relation $\{N(t) \leq r(t)\} = \{t_{r(t)} > t\}$ we have

$$\begin{aligned}
\mathbf{P}\left(\frac{N(t)-t/\mu}{\sqrt{t\sigma^2/\mu^3}} \leq x\right) &= \mathbf{P}\left(N(t) \leq t/\mu + x\sqrt{t\sigma^2/\mu^3}\right) \\
&= \mathbf{P}\left(N(t) \leq r(t)\right) = \mathbf{P}\left(t_{r(t)} > t\right) \\
&= \mathbf{P}\left(\frac{t_{r(t)} - \mu r(t)}{\sigma\sqrt{r(t)}} > \frac{t-\mu r(t)}{\sigma\sqrt{r(t)}}\right) \\
&= 1 - \mathbf{P}\left(\frac{t_{r(t)} - \mu r(t)}{\sigma\sqrt{r(t)}} \leq \frac{t-\mu r(t)}{\sigma\sqrt{r(t)}}\right).
\end{aligned}$$

It can be easily checked that

$$\frac{t-\mu r(t)}{\sigma\sqrt{r(t)}} \to -x, \quad t \to \infty,$$

and the continuity of the standard normal distribution function implies the convergence

$$\mathbf{P}\left(\frac{N(t)-t/\mu}{\sqrt{t\sigma^2/\mu^3}} \leq x\right) \to 1 - \Phi(-x) = \Phi(x), \quad t \to \infty.$$

The equation $1-\Phi(-x) = \Phi(x)$ follows from the symmetry of the standard normal distribution. □

The following results (without proof) concerning the mean value and variance of the renewal process $N(t)$ are a generalization of previous results and are valid for the renewal processes with delay, too.

Theorem 4.31. *If $\mu_2 = \mathbf{E}\left(T_1^2\right) < \infty$ and T_1 has a nonlattice distribution, then as $t \to \infty$ [31, XIII-12§]*

$$\mathbf{E}\left(N(t)\right) - \frac{t}{\mu} = H(t) - \frac{t}{\mu} \to \frac{\mu_2}{2\mu^2} - 1,$$

$$\mathbf{D}^2\left(N(t)\right) = \frac{\mu_2 - \mu^2}{\mu^3} t + o(t).$$

If, additionally, $\mu_3 = \mathbf{E}\left(T_1^3\right) < \infty$, then [31]

$$\mathbf{D}^2\left(N(t)\right) = \frac{\mu_2 - \mu^2}{\mu^3} t + \left(\frac{5\mu_2^2}{4\mu^4} - \frac{2\mu_3}{3\mu^3} - \frac{\mu_2}{2\mu^2}\right) + o(1).$$

4.2 Regenerative Processes

Many queueing systems can be described by means of regenerative processes. This property makes it possible to prove the limit and stability theorems in order to use the method of simulation.

Definition 4.32. Let T be a nonnegative random variable and $Z(t),\ t \in [0, T)$ be a stochastic process. The pair $(T, Z(t))$, taking on values in the measurable space $(\mathcal{Z}, \mathcal{B})$, is called a cycle of length T.

Definition 4.33. The stochastic process $Z(t),\ t \geq 0$, taking on values in the measurable space $(\mathcal{Z}, \mathcal{B})$, is called a **regenerative process** with moments of regeneration $t_0 = 0 < t_1 < t_2 < \ldots$ if there exists a sequence of independent cycles $(T_k, Z_k(t)),\ \ k \geq 1$, such that

(1) $T_k = t_k - t_{k-1},\ \ k \geq 1$;
(2) $\mathbf{P}(T_k > 0) = 1,\ \ \mathbf{P}(T_k < \infty) = 1$;
(3) All cycles are stochastically equivalent.
(4) $Z(t) = Z_k(t - t_{k-1})$ if $t \in [t_{k-1}, t_k),\ k \geq 1$.

Definition 4.34. If property (3) is fulfilled only starting with the second cycle (analogously to the renewal processes), then we have a **delayed regenerative process**.

Remark 4.35. $t_k,\ k \geq 1$, is a renewal process.

In the case of regenerative processes, an important task is to find conditions assuring the existence and possibility of determining the limit

$$\lim_{t\to\infty} \mathbf{P}(Z(t) \in B), \quad B \in \mathcal{B}.$$

It is also important to estimate the rate of convergence (especially upon examination of the stability problems of queueing systems and simulation procedures).

Let $\{Z(t),\ t \geq 0\}$ be a regenerative process taking on values in the measurable space $(\mathcal{Z}, \mathcal{B})$ with regeneration points $t_0 = 0 < t_1 < t_2 < \ldots$, $T_n = t_n - t_{n-1},\ \ n = 1, 2, \ldots$. Assume that $Z(t)$ is right continuous and there exists a limit from the left. Then the cycles $\{T_n,\ \{Z(t_{n-1} + u) :\ 0 \leq u < T_n\}\},\ n = 1, 2, \ldots$, are independent and stochastically equivalent; $\{t_n,\ n \geq 1\}$; and the corresponding counting process $\{N(t),\ t \geq 0\}$ is a renewal process. Let F denote the common distribution of random variables $\{T_n,\ n \geq 1\}$.

The most important application of Smith's theorem is the determination of limit values $\lim_{t\to\infty} \mathbf{E}(W(t))$ for the renewal and regenerative processes, where $W(t) = \Psi(t, N, Z)$ is the function of t, the renewal process N, and the regenerative process Z. The determination of the limit value is based on a more general theorem.

Theorem 4.36. *Let $\{V(t),\ t \geq 0\}$ be a real-valued stochastic process on the same probability space as the process $\{N(t),\ t \geq 0\}$, and for which the mean value $f(t) = \mathbf{E}\,(V(t))$ is bounded on each finite interval. Let*

$$g(t) = \mathbf{E}\left(V(t)\mathcal{I}_{\{T_1 > t\}}\right) + \int_0^t \left[\mathbf{E}\,(V(t)|T_1 = s) - \mathbf{E}\,(V(t-s))\right] \mathrm{d}F(s),\quad t \geq 0.$$

Assume that the positive and negative parts of g are directly Riemann integrable. If F is a nonlattice distribution, then

$$\lim_{t\to\infty} f(t) = \lim_{t\to\infty} \mathbf{E}\,(V(t)) = \frac{1}{\mu}\int_0^\infty g(x)\mathrm{d}x.$$

A similar result is valid if F is a lattice distribution.

Remark 4.37. In the theorem, the property of direct Riemann integrability was required separately for the positive and negative parts of the function g. The reason is that the property is defined only for nonnegative functions.

Proof. It is clear that

$$\begin{aligned} f(t) &= \mathbf{E}\left(V(t)\mathcal{I}_{\{T_1 > t\}}\right) + \mathbf{E}\left(V(t)\mathcal{I}_{\{T_1 \leq t\}}\right) \\ &= \mathbf{E}\left(V(t)\mathcal{I}_{\{T_1 > t\}}\right) + \int_0^t \mathbf{E}\,(V(t)|T_1 = s)\mathrm{d}F(s). \end{aligned}$$

Let us add and subtract $F * f(t)$; then we get the renewal equation

$$f = g + F * f.$$

The solution of the equation is $f(t) = g + H * g(t)$, which because of the convergence $g(t) \to 0,\ t \to \infty$, and the elementary renewal theorem as a simple consequence of direct Riemann integrability tends to $\frac{1}{\mu}\int_0^\infty g(x)\mathrm{d}x$ as $t \to \infty$. □

Remark 4.38. From the proof it is clear that under the condition of Theorem 4.36 for an arbitrary process $V(t)$ there exists the representation $\mathbf{E}\,(V(t)) = H * g(t)$ and for the existence of the limit the direct Riemann integrability is required. This representation is interesting if $V(t)$ depends on $Z(t)$.

Special Case Let $h : \mathcal{Z} \to R$ be a measurable function for which, for all t, $\mathbf{E}\,(|h(Z(t))|) < \infty$. $Z(t)$ is a regenerative process, and the part starting with the second cycle is independent of the first cycle of length T_1, so for arbitrary $0 < s < t$

$$\mathbf{E}\,((h(Z(t))|T_1 = s)) = \mathbf{E}\,(h(Z(t-s))).$$

Using the previous notation

$$g(t) = \mathbf{E}\left(h(Z(t))\mathcal{I}_{\{T_1>t\}}\right).$$

Theorem 4.39. *If g_+ and g_- are directly Riemann integrable, then*

$$\begin{aligned}\lim_{t\to\infty} \mathbf{E}\left(h(Z(t))\right) &= \mu^{-1}\int_0^\infty g(s)\,\mathrm{d}s \\ &= \mu^{-1}\int_0^\infty \mathbf{E}\left(h(Z(s)\mathcal{I}_{\{T_1>s\}}\right)\mathrm{d}s \\ &= \mu^{-1}\mathbf{E}\left(\int_0^{T_1} h(Z(s))\right)\mathrm{d}s.\end{aligned}$$

For arbitrary $A \in \mathcal{B}$ the following equality holds:

$$\begin{aligned}\lim_{t\to\infty} \mathbf{P}\left(Z(t) \in A\right) &= \mu^{-1}\int_0^\infty \mathbf{P}\left(Z(s) \in A,\ T_1 > s\right)\mathrm{d}s \\ &= \mu^{-1}\mathbf{E}\left(\int_0^{T_1} \mathcal{I}_{\{Z(s)\in A\}}\,\mathrm{d}s\right).\end{aligned}$$

Proof. The first relation follows from the previous theorem, and for the second one it is necessary to mention that, since the trajectories of Z are right continuous and have left limits, the (integrable, bounded) function $\mathbf{P}\left(Z(s) \in A,\ T_1 > s\right)$ has a countable number of discontinuities and, consequently, is directly Riemann integrable. □

We give one more limit theorem (without proof) that is often useful in practice.

Theorem 4.40. *Let F be a nonlattice distribution, and let at least one of the following conditions be fulfilled:*

(a) $\mathbf{P}\left(Z(t) \in A\right)$ *is Riemann integrable on an arbitrary finite interval, and* $\mu = \int_0^\infty x\,\mathrm{d}F(x) < \infty$ *holds.*
(b) Starting with a certain integer $n \geq 1$ *the distribution functions defined by* $F^{(1)} = F$, $F^{(n+1)} = F^{(n)} * F$, *are absolute continuous and* $\mu = \int_0^\infty x\,\mathrm{d}F(x) < \infty$.

Then the following relation holds:

$$\lim_{t\to\infty} \mathbf{P}(Z(t) \in A) = \mu^{-1} \int_0^\infty \mathbf{P}(Z(s) \in A, T_1 > s)\,\mathrm{d}s$$
$$= \mu^{-1}\mathbf{E}\left(\int_0^{T_1} \mathcal{I}_{\{Z(s)\in A\}}\,\mathrm{d}s\right).$$

Example 4.41. Let us consider the renewal process $\{N(t),\ t \ge 0\}$; the renewal moments are

$$t_0 = 0,\ \ t_n = T_1 + T_2 + \ldots + T_n,\ \ n \ge 1,$$

and, furthermore, $\mathbf{P}(T_k \le x) = F(x),\ k \ge 1$, $\mu = \int_0^\infty x\,\mathrm{d}F(x)$. For arbitrary $t > 0$ we define

$$\begin{array}{ll} \delta(t) = t - t_{N(t)}, & \text{the age,} \\ \gamma(t) = t_{N(t)+1} - t, & \text{the residual lifetime,} \\ \beta(t) = \gamma(t) - \delta(t) = t_{N(t)+1} - t_{N(t)}, & \text{the total lifetime.} \end{array}$$

(For example, at instant t, $\delta(t)$ indicates how much time passed without a car arriving at the station, and $\gamma(t)$ indicates how long it was necessary to wait till the arrival of the next car, on the condition that the interarrival times are i.i.d. random variables with the common distribution function F.)

Theorem 4.42. $\{\delta(t),\ t \ge 0\}$ *and* $\{\gamma(t),\ t \ge 0\}$ *are regenerative processes, and in the case of the nonlattice distribution* F*,*

$$\lim_{t\to\infty} \mathbf{P}(\delta(t) \le x) = \lim_{t\to\infty} \mathbf{P}(\gamma(t) \le x) = \frac{1}{\mu}\int_0^x (1 - F(u))\,\mathrm{d}u,$$

$$\lim_{t\to\infty} \mathbf{P}(\beta(t) \le x) = \frac{1}{\mu}\int_0^x s\,\mathrm{d}F(s).$$

Proof. Both processes are obviously regenerative with common regeneration points $t_n,\ n \ge 1$. By our previous theorem,

$$\lim_{t\to\infty} \mathbf{P}(\delta(t) \le x) = \frac{1}{\mu}\int_0^\infty \mathbf{P}(\delta(s) \le x, T_1 > s)\,\mathrm{d}s;$$

furthermore,

$$\mathbf{P}(\delta(s) \le x, T_1 > s) = \mathbf{P}(s \le x, T_1 > s) = \begin{cases} 1 - F(s), & \text{if } s < x, \\ 0, & \text{if } s \ge x, \end{cases}$$

so

$$\lim_{t\to\infty} \mathbf{P}(\delta(t) \le x) = \frac{1}{\mu}\int_0^x (1-F(s))\mathrm{d}s = \frac{1}{\mu}\int_0^x (1-F(s))\mathrm{d}s$$

using the identity $\mu = \int_0^\infty (1-F(s))\mathrm{d}s$ (Exercise 1.5). Similarly, for the process $\{\gamma(t),\ t \ge 0\}$ we obtain

$$\begin{aligned}
&\lim_{t\to\infty} \mathbf{P}(\gamma(t) \le x) \\
&= \frac{1}{\mu}\int_0^\infty \mathbf{P}(\gamma(s) \le x, T_1 > s)\mathrm{d}s = \frac{1}{\mu}\int_0^\infty \mathbf{P}(T_1 - s \le x, T_1 > s)\mathrm{d}s \\
&= \frac{1}{\mu}\int_0^\infty \mathbf{P}(s \le T_1 < s + x)\mathrm{d}s = \frac{1}{\mu}\int_0^\infty (F(s+x) - F(s))\mathrm{d}s \\
&= -\frac{1}{\mu}\left(\int_x^\infty (1-F(s))\mathrm{d}s - \int_0^\infty (1-F(s))\mathrm{d}s\right) = \frac{1}{\mu}\int_0^x (1-F(s))\mathrm{d}s.
\end{aligned}$$

The statement for $\{\gamma(t),\ t \ge 0\}$ can be obtained analogously. □

Similarly to the renewal processes, the law of large numbers and the central limit theorem can be proved for the regenerative processes, too. Here we will not deal with these questions.

4.2.1 Estimation of Convergence Rate for Regenerative Processes

For a wide class of regenerative processes (e.g., stochastic processes describing queueing systems) one can estimate the rate of convergence of distributions of certain parameters to a stationary distribution by means of the so-called coupling method [65].

Lemma 4.43 (*Coupling lemma*). *For the arbitrary random variables X and Y and an arbitrary Borel set A of the real line the following statements hold:*

(i) $|\mathbf{P}(X \in A) - \mathbf{P}(Y \in A)| \le \mathbf{P}(X \ne Y)$.
(ii) *If* $X = X_1 + \ldots + X_n$ *and* $Y = Y_1 + \ldots + Y_n$, *then* $|\mathbf{P}(X \in A) - \mathbf{P}(Y \in A)| \le \sum_{k=1}^{n} \mathbf{P}(X_k \ne Y_k)$.

Proof. If $\mathbf{P}(X \in A) = \mathbf{P}(Y \in A)$, then (i) is obviously true.

Suppose that $\mathbf{P}(X \in A) > \mathbf{P}(Y \in A)$ (if one changes the notation, then this can always be done if the two probabilities differ). Then

$$\begin{aligned} |\mathbf{P}(X \in A) - \mathbf{P}(Y \in A)| &= \mathbf{P}(X \in A) - \mathbf{P}(Y \in A) \\ &\leq \mathbf{P}(X \in A) - \mathbf{P}(Y \in A, X \in A) \\ &= \mathbf{P}(X \in A, Y \in A^c) \leq \mathbf{P}(X \neq Y). \end{aligned}$$

Proof of relation (ii). Since $\{X \neq Y\} \subset \bigcup\limits_{k=1}^{n} \{X_k \neq Y_k\}$, we have

$$\mathbf{P}(X \neq Y) \leq \mathbf{P}\left(\bigcup_{k=1}^{n} \{X_k \neq Y_k\}\right) \leq \sum_{k=1}^{n} \mathbf{P}(X_k \neq Y_k).$$

□

Application of Coupling Lemma Let $Z = \{Z(j),\ j \geq 1\}$ be the discrete-time, real-valued regenerative process under consideration. Assume that there exists the weak stationary limit of the process $\tilde{Z} = \{Z(j+n),\ j \geq 1\}$ as $n \to \infty$ (its finite-dimensional distributions weakly converge to the finite-dimensional distributions of a stationary process), which is also regenerative, and let $Y = \{Y(j),\ j \geq 1\}$ be its realization, not necessarily different from Z on the same probability space. Let τ denote the first instant when the processes Z and Y are regenerated at the same time (in many concrete cases the distribution of τ can be easily estimated). Then the convergence rate of the distribution of $Z(j)$ can be estimated by means of the distribution of τ as follows: if after the regeneration point τ the process Z is replaced by the next part of process Y following the common regeneration point τ, then the finite-dimensional distributions of process Z do not change. It is clear that $\{\tau < j\} \subseteq \{Z(j) = Y(j)\}$, i.e., $\{Z(j) \neq Y(j)\} \subseteq \{\tau \geq j\}$, from which, using the coupling lemma for the arbitrary Borel set A of the real line, the estimation

$$|\mathbf{P}(Z(j) \in A) - \mathbf{P}(Y(j) \in A)| \leq \mathbf{P}(Z(j) \neq Y(j)) \leq \mathbf{P}(\tau \geq j)$$

holds.

4.3 Analysis Methods Based on Markov Property

Definition 4.44. A discrete-state, continuous-time stochastic process, $X(t)$, possesses the **Markov propety** at time t_n if for all $n, m \geq 1$, $0 \leq t_0 < t_1 < \ldots < t_n < t_{n+1} < \ldots < t_{n+m}$, and $x_0, x_1, \ldots, x_n, x_{n+1}, \ldots, x_{n+m} \in S$ we have

$$\begin{aligned} &\mathbf{P}(X(t_{n+m}) = x_{n+m}, \ldots, X(t_{n+1}) = x_{n+1} | X(t_n) = x_n, \ldots, X(t_0) = x_0) \\ &\quad = \mathbf{P}(X(t_{n+m}) = x_{n+m}, \ldots, X(t_{n+1}) = x_{n+1} | X(t_n) = x_n). \end{aligned} \tag{4.5}$$

In this case t_n is referred to as a regenerative point.

A commonly applied interpretation of the Markov property is as follows. Assuming that the current time is t_n (present), which is a regenerative point, and we know the current state of the process $X(t_n)$, then the future of the stochastic process $X(t)$ for $t_n \leq t$ is independent of the past history of the process $X(t)$ for $0 \leq t < t_n$, and it only depends on the current state of the process $X(t_n)$. That is, if one knows the present state, the future is independent of the past.

In the case of discrete-time processes, it is enough to check if the one-step state transitions are independent of the past, i.e., it is enough to check the condition for $m = 1$.

Usually, we restrict our attention to stochastic processes with nonnegative parameters (positive half of the time axes), and in these cases we assume that $t = 0$ is a regenerative point.

4.3.1 Time-Homogeneous Behavior

Definition 4.45. The stochastic process $X(t)$ is *time homogeneous* if the stochastic behavior of $X(t)$ is invariant for time shifting, that is, the stochastic behavior of $X(t)$ and $X'(t) = X(t+s)$ are identical in distribution $X(t) \stackrel{d}{=} X'(t)$.

Corollary 4.46. *If the time-homogeneous stochastic process $X(t)$ possesses the Markov property at time T and $X(T) = i$, then $X(t) \stackrel{d}{=} X(t-T)$ if $X(0) = i$.*

The corollary states that starting from two different Markov points with the same state results in stochastically identical processes.

4.4 Analysis of Continuous-Time Markov Chains

Definition 4.47. The discrete-state, continuous-time stochastic process $X(t)$ is a *continuous-time Markov chain* (CTMC) if it possesses the Markov property for all $t \geq 0$.

Based on this definition and assuming time-homogeneous behavior we obtain the following properties.

Corollary 4.48. *An arbitrary finite-dimensional joint distribution of a CTMC is composed of the product of transition probabilities multiplied by an initial probability.*

Corollary 4.49. *For the time points $t < u < v$ the following Chapman–Kolmogorov equation holds:*

$$\hat{p}_{ij}(t,v) = \sum_{l \in S} \hat{p}_{il}(t,u)\hat{p}_{lj}(u,v); \quad \hat{\Pi}(t,v) = \hat{\Pi}(t,u)\hat{\Pi}(u,v), \tag{4.6}$$

where $\hat{p}_{ij}(t,u) = \mathbf{P}(X(u) = j \mid X(t) = i)$ *for all* $i,\ j \in S, 0 \le t \le u$, $\hat{\Pi}(t,u) = \left[\hat{p}_{ij}(t,u)\right]$. *In the case of time-homogeneous processes the time shifts* $u - t = \tau_1$ *and* $v - u = \tau_2$ *play a role:*

$$p_{ij}(\tau_1 + \tau_2) = \sum_{l \in S} p_{il}(\tau_1) p_{lj}(\tau_2); \quad \Pi(\tau_1 + \tau_2) = \Pi(\tau_1)\Pi(\tau_2), \tag{4.7}$$

where $\Pi(\tau) = \left[p_{ij}(\tau)\right]$, $p_{ij}(\tau) = \mathbf{P}(X(\tau) = j \mid X(0) = i)$, *for all* $i,\ j \in S, 0 \le \tau$.

Definition 4.50. The stochastic evolution of a CTMC is commonly characterized by an *infinitesimal generator* matrix (commonly denoted by Q) that can be obtained from the derivative of the state-transition probabilities as follows:

$$\frac{\mathrm{d}}{\mathrm{d}t}\Pi(t) = \lim_{\delta \to 0} \frac{\Pi(t+\delta) - \Pi(t)}{\delta} = \Pi(t) \underbrace{\lim_{\delta \to 0} \frac{\Pi(\delta) - I}{\delta}}_{Q} = \Pi(t)Q. \tag{4.8}$$

Corollary 4.51. *The sojourn time of a CTMC in a given state* i *is exponentially distributed with the parameter* $q_i = -q_{ii}$. *The probability that after state* i *the next visited state will be state* j *is* q_{ij}/q_i, *and it is independent of the sojourn time in state* i.

Remark 4.52. Based on Corollary 4.51 and the properties of the exponential distribution, the state transitions of a CTMC can also be interpreted in the following way. When the CTMC moves to state i, several exponentially distributed activities start, exactly one for each nonzero transition rate. The time of the activity associated with the state transition from state i to state j is exponentially distributed with the parameter q_{ij}. The CTMC leaves state i and moves to the next state when the first one of these activities completes. The next visited state is the state whose associated activity finishes first.

Corollary 4.53 (*Short-term behavior of CTMCs*). *During a short time period* Δ, *the behavior of a CTMC is characterized by the following transition probabilities:*

(a) $\mathbf{P}(X(t+\Delta) = i | X(t) = i) = 1 - q_i\Delta + o(\Delta)$;
(b) $\mathbf{P}(X(t+\Delta) = j | X(t) = i) = q_{ij}\Delta + o(\Delta)$ *for* $i \ne j$;
(c) $\mathbf{P}(X(t+\Delta) = j, X(u) = k | X(t) = i) = o(\Delta)$ *for* $i \ne k,\ j \ne k$, *and* $t < u < t + \Delta$,

where $o(x)$ *denotes the set of functions with the property* $\lim_{x \to 0} o(x)/x = 0$.

According to the corollary, two main events can happen with significant probability during a short time period:

- The CTMC stays in the initial state during the whole period [(a)].
- It moves from state i to j [(b)].

The event that more than one state transition happens during a short time period [(c)] has a negligible probability as $\Delta \to 0$.

Corollaries 4.51 and 4.53 allow different analytical approaches for the description of the transient behavior of CTMCs.

4.4.1 Analysis Based on Short-Term Behavior

Let $X(t)$ be a CTMC with state space S, and let us consider the change in state probability $P_i(t+\Delta) = \mathbf{P}(X(t+\Delta) = i)$ $(i \in S)$ considering the possible events during the interval $(t, t+\Delta)$. The following cases must be considered:

- There is no state transition during the interval $(t, t+\Delta)$. In this case $P_i(t+\Delta) = P_i(t)$, and the probability of this event is $1 - q_i\Delta + o(\Delta)$.
- There is one state transition during the $(t, t+\Delta)$ interval from state k to state i. In this case $P_i(t+\Delta) = P_k(t)$, and the probability of this event is $q_{ki}\Delta + o(\Delta)$.
- The process stays in state i at time $t+\Delta$ such that there is more than one state transition during the interval $(t, t+\Delta)$. The probability of this event is $o(\Delta)$.

Considering these cases we can compute $P_i(t+\Delta)$ from $P_k(t)$, $k \in S$, as follows:

$$\begin{aligned} P_i(t+\Delta) &= (1 - q_i\Delta + o(\Delta))P_i(t) + \sum_{k \in S, k \neq i} (q_{ki}\Delta + o(\Delta))P_k(t) + o(\Delta) \\ &= (1 - q_i\Delta)P_i(t) + \sum_{k \in S, k \neq i} (q_{ki}\Delta)P_k(t) + o(\Delta), \end{aligned}$$

from which

$$\frac{P_i(t+\Delta) - P_i(t)}{\Delta} = -q_i P_i(t) + \sum_{k \in S, k \neq i} q_{ki} P_k(t) + \frac{o(\Delta)}{\Delta} = \sum_{k \in S} q_{ki} P_k(t) + \frac{o(\Delta)}{\Delta}.$$

Finally, setting the limit $\Delta \to 0$ we obtain that

$$\frac{\mathrm{d}P_i(t)}{\mathrm{d}t} = \sum_{k \in S} q_{ki} P_k(t).$$

Introducing the row vector of state probabilities $P(t) = \{P_i(t)\}, i \in S$, we obtain the vector-matrix form of the previous equation:

$$\frac{\mathrm{d}}{\mathrm{d}t} P(t) = P(t)\mathbf{Q}. \tag{4.9}$$

A differential equation describes the evolution of a transient state probability vector. To define the state probabilities, we additionally need to have an initial condition. In practical applications, the initial condition is most often the state probability distribution at time 0, i.e., $P(0)$. The solution of Eq. (4.9) with initial condition $P(0)$ is [55]

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

Transform Domain Description The Laplace transform of the two sides of Eq. (4.9) gives

$$s\,P^*(s) - P(0) = P^*(s)\mathbf{Q},$$

from which we can express $P^*(s)$ in the following form:

$$P^*(s) = P(0)[s\mathbf{I} - \mathbf{Q}]^{-1}.$$

Comparing the time and transform domain expressions we have that $\mathrm{e}^{\mathbf{Q}t}$ and $[s\mathbf{I} - \mathbf{Q}]^{-1}$ are Laplace transform pairs of each other.

Stationary Behavior If $\lim_{t\to\infty} P_i(t)$ exists, then we say that $\lim_{t\to\infty} P_i(t) = P_i$ is the stationary probability of state i. In this case, $\lim_{t\to\infty} \mathrm{d}P_i(t)/\mathrm{d}t = 0$, and the stationary probability satisfies the system of linear equations $\sum_{k\in S} q_{ki}\, P_k(t) = 0$ for all $k \in S$.

4.4.2 *Analysis Based on First State Transition*

Let $X(t)$ be a CTMC with state space S, and let $T_1, T_2, T_3, \ldots$ denote the time of the first, second, etc. state transitions of the CTMC. We assume that $T_0 = 0$, and $\tau_1, \tau_2, \tau_3, \ldots$ are the sojourn times spent in the consecutively visited states ($\tau_i = T_i - T_{i-1}$). We compute the state-transition probability $\pi_{ij}(t) = \mathbf{P}\,(X(t) = j \mid X(0) = i)$ assuming that $T_1 = h$, i.e., we are interested in

$$\pi_{ij}(t|T_1 = h) = \mathbf{P}\,(X(t) = j \mid X(0) = i, T_1 = h).$$

We have

$$\pi_{ij}(t|T_1 = h) = \begin{cases} \delta_{ij}, & h \ge t, \\ \displaystyle\sum_{k\in S, k\neq i} \frac{q_{ik}}{-q_{ii}}\pi_{kj}(t-h), & h < t, \end{cases} \tag{4.10}$$

where δ_{ij} is the Kronecker delta ($\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$), and $\frac{q_{ik}}{-q_{ii}}$ is the probability that after visiting state i the Markov chain moves to state k. In the case of general stochastic processes, this probability might depend on the sojourn time in state i, but in the case of CTMCs, it is independent.

Equation (4.10) has two cases:

- If the time point of interest, t, is before the first state transition of the CTMC, $h \geq t$, then the conditional state-transition probability is either 1 (if the initial and final states are identical $i = j$) or 0 (if $i \neq j$).
- If the time point of interest, t, is after the first state transition of the CTMC, $T_1 < t$, then we can analyze the evolution of the process from T_1 to t using the fact that the process possesses the Markov property at time T_1. In this case we need to consider all possible states that might be visited at time T_1, $k \in S, k \neq i$, with the associated probability $\frac{q_{ik}}{-q_{ii}}$. The state-transition probabilities from T_1 to t are identical with the state-transition probabilities of the original process from 0 to $T_1 - t$, assuming that the original process starts from state k.

The distribution of T_1 is known. It is exponentially distributed with the parameter $-q_{ii}$. Its cumulated and probability density functions are $F_{T_1}(x) = 1 - \mathrm{e}^{q_{ii}x}$ and $f_{T_1}(x) = -q_{ii}\mathrm{e}^{q_{ii}x}$, respectively. With that we can apply the total probability theorem to compute the (unconditional) state-transition probability $\pi_{ij}(t)$:

$$
\begin{aligned}
\pi_{ij}(t) &= \int_{h=0}^{\infty} \pi_{ij}(t|T_1 = h)\; f_{T_1}(h)\,\mathrm{d}h \\
&= \int_{h=t}^{\infty} \delta_{ij}\; f_{T_1}(h)\,\mathrm{d}h + \int_{h=0}^{t} \sum_{k\in S,k\neq i} \frac{q_{ik}}{-q_{ii}}\,\pi_{kj}(t-h)\; f_{T_1}(h)\,\mathrm{d}h \\
&= \delta_{ij}\;(1 - F_{T_1}(t)) + \int_{h=0}^{t} \sum_{k\in S,k\neq i} \frac{q_{ik}}{-q_{ii}}\,\pi_{kj}(t-h)\; f_{T_1}(h)\,\mathrm{d}h \\
&= \delta_{ij}\;\mathrm{e}^{q_{ii}t} + \sum_{k\in S,k\neq i} q_{ik} \int_{h=0}^{t} \pi_{kj}(t-h)\,\mathrm{e}^{q_{ii}h}\,\mathrm{d}h.
\end{aligned}
\tag{4.11}
$$

The obtained integral equation is commonly referred to as a Volterra integral equation. Its only unknown is the state-transition probability function $\pi_{ij}(t)$. The numerical methods developed for the numerical analysis of Volterra integral equations can be used to compute the state-transition probabilities of a CTMC.

Relation of Analysis Methods We can rewrite Eq. (4.11) in the following form:

$$
\begin{aligned}
\pi_{ij}(t) &= \delta_{ij}\;\mathrm{e}^{q_{ii}t} + \sum_{k\in S,k\neq i} q_{ik} \int_{h=0}^{t} \pi_{kj}(t-h)\,\mathrm{e}^{q_{ii}h}\,\mathrm{d}h \\
&= \delta_{ij}\;\mathrm{e}^{q_{ii}t} + \sum_{k\in S,k\neq i} q_{ik} \int_{h=0}^{t} \pi_{kj}(h)\,\mathrm{e}^{q_{ii}(t-h)}\,\mathrm{d}h \\
&= \delta_{ij}\;\mathrm{e}^{q_{ii}t} + \sum_{k\in S,k\neq i} q_{ik}\,\mathrm{e}^{q_{ii}t} \int_{h=0}^{t} \pi_{kj}(h)\,\mathrm{e}^{-q_{ii}h}\,\mathrm{d}h.
\end{aligned}
\tag{4.12}
$$

The derivation of the two sides of Eq. (4.12) according to t is as follows:

$$\begin{aligned}
\pi'_{ij}(t) &= \delta_{ij}\; q_{ii}\; \mathrm{e}^{q_{ii}t} + \sum_{k\in S, k\neq i} q_{ik} \left(q_{ii}\; \mathrm{e}^{q_{ii}t} \int_{h=0}^{t} \pi_{kj}(h)\; \mathrm{e}^{-q_{ii}h}\; \mathrm{d}h + \mathrm{e}^{q_{ii}t}\; \pi_{kj}(t)\; \mathrm{e}^{-q_{ii}t} \right) \\
&= \sum_{k\in S, k\neq i} q_{ik}\; \pi_{kj}(t) + q_{ii} \underbrace{\left(\delta_{ij}\; \mathrm{e}^{q_{ii}t} + \sum_{k\in S, k\neq i} q_{ik}\; \mathrm{e}^{q_{ii}t} \int_{h=0}^{t} \pi_{kj}(h)\; \mathrm{e}^{-q_{ii}h}\; \mathrm{d}h \right)}_{\pi_{ij}(t)} \\
&= \sum_{k\in S} q_{ik} \pi_{kj}(t),
\end{aligned}$$

where we used Eq. (4.11) for the substitution of the integral expression. The obtained differential equation is similar to that provided by the analysis of the short-term behavior.

Transform Domain Description To relate the two transient descriptions of the CTMC, one with a differential equation and one with an integral equation, we transform these descriptions into a Laplace transform domain. It is easy to take the Laplace transform from the last line of Eq. (4.11) because the second term of the right-hand side is a convolution integral. That is,

$$\pi^*_{ij}(s) = \delta_{ij}\; \frac{1}{s - q_{ii}} + \sum_{k\in S, k\neq i} q_{ik}\; \pi^*_{kj}(s)\; \frac{1}{s - q_{ii}}.$$

Multiplying by the denominator and using that $-q_{ii} = \sum_{k\in S, k\neq i} q_{ik}$ we obtain

$$s\; \pi^*_{ij}(s) = \delta_{ij} + \sum_{k\in S} q_{ik}\; \pi^*_{kj}(s),$$

which can be written in the matrix form

$$s\; \Pi^*(s) = \mathbf{I} + \mathbf{Q}\Pi^*(s).$$

Finally, we have

$$\Pi^*(s) = [s\mathbf{I} - \mathbf{Q}]^{-1},$$

which is identical to the Laplace transform expression obtained from the differential equation.

Embedded Markov Chain at State Transitions Let $X_i \in S, i = 0, 1, \ldots,$ denote the ith visited state of the Markov chain $X(t)$, which is the state of the Markov chain in the interval (T_i, T_{i+1}) (Fig. 4.2). The $X_0, X_1, \ldots$ series of random variables is a discrete-time Markov chain (DTMC) due to the Markov property of $X(t)$. This DTMC is commonly referred to as a Markov chain embedded at the state transitions or simply an *embedded Markov chain* (EMC). The state-transition probability matrix of the EMC is

$$\Pi_{ij} = \begin{cases} \dfrac{q_{ij}}{-q_{ii}}, & i \neq j, \\ 0, & i = j. \end{cases}$$

Stationary Analysis Based on the EMC The stationary distribution of the EMC $\hat{P}$ (which is the solution of $\hat{P} = \hat{P}\Pi, \sum_i \hat{P}_i = 1$) defines the relative frequency of the visits to the state of the Markov chain. The higher the stationary probability is, the more frequently the state is visited. The stationary behavior of the CTMC $X(t)$ is characterized by two main factors: how often the state is visited (represented by $\hat{P}_i$) and how long a visit lasts. If state i is visited twice as frequently as state j but the mean time of a visit to state i is half the mean time of a visit to j, then the stationary probabilities of states i and j are identical. This intuitive behavior is summarized in the following general rule of renewal theory [58]:

$$P_i = \frac{\hat{P}_i \hat{\tau}_i}{\sum_j \hat{P}_j \hat{\tau}_j},$$

where $\hat{\tau}_j$ is the mean time spent in state j, which is known from the diagonal element of the infinitesimal generator, $\hat{\tau}_j = -1/q_{jj}$.

Discrete-Event Simulation of CTMCs There are at least two possible approaches.

- When the CTMC is in state i, first draw an exponentially distributed random sample with parameter $-q_{ii}$ for the sojourn time in state i, then draw a discrete random sample for deciding the next visited state with distribution Π_{ij}, $j \in S$.
- When the CTMC is in state i, draw an exponentially distributed random sample with parameter q_{ij}, say τ_{ij}, for all positive transition rates of row i of the infinitesimal generator matrix. Find the minimum of these samples, $\min_j \tau_{ij}$. The sojourn time in state i is this minimum, and the next state is the one whose associated random sample is minimal.

4.5 Semi-Markov Process

Definition 4.54. The discrete-state, continuous-time random process $X(t)$ is a semi-Markov process if it is time homogeneous and it possesses the Markov property at the state-transition instances (Fig. 4.2).

The name semi-Markov process comes from the fact that such processes do not always possess the Markov property (during its sojourn in a state), but there are particular instances (state-transition instances) when they do.

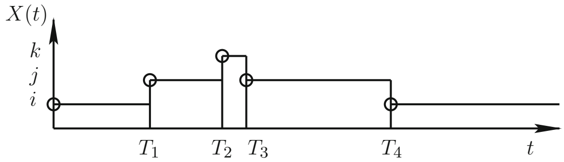
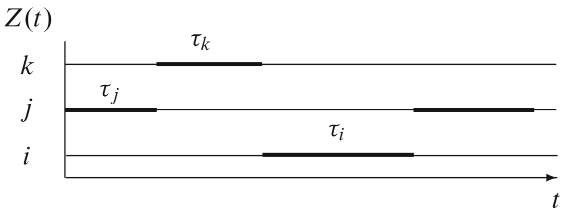


Fig. 4.2 Semi-Markov process that possesses the Markov property at the indicated time points

Corollary 4.55. *The sojourn time in state i can be any general real-valued positive random variable. During a sojourn in state i, both the remaining time in both that state and the next visited state depend on the elapsed time since the process entered state i.*



Example 4.56. A two-state (up/down) system fails at a rate λ (the up time of the system is exponentially distributed with parameter λ) and gets repaired at a rate μ. To avoid long down periods, the repair process is stopped and a replacement process is initialized after a deterministic time limit d. The time of the replacement is a random variable with a distribution $G(t)$. Define a system model and check if it is a semi-Markov process.

Because a CTMC always possesses the Markov property, it follows that the sojourn time in a state is exponentially distributed and that the distribution of the next state is independent of the sojourn time. For example, considering the first state transition and the sojourn time in the first state we have

$$\mathbf{P}(X_1 = j, T_1 = c | X_0 = i) = \mathbf{P}(X_1 = j | X_0 = i)\mathbf{P}(T_1 = c | X_0 = i).$$

This property does not hold for semi-Markov processes in general. The most important consequences of the definition of semi-Markov processes are the following ones. The sojourn time in a state can have any positive distribution, and the distribution of the next state and the time spent in a state are not independent in general. Consequently, to define a semi-Markov process, this joint distribution must be given. This is usually done by defining the kernel matrix of a process whose i, j element is

$$Q_{ij}(t) = \mathbf{P}(X(T_{i+1}) = j, \tau_{i+1} \le t \mid X(T_i) = i).$$

Utilizing the time homogeneity of the process we further have for T_i that

$$Q_{ij}(t) = \mathbf{P}(X(T_{i+1}){=}j, \tau_{i+1} \le t \mid X(T_i){=}i){=}\mathbf{P}(X(T_1){=}j, T_1 \le t \mid X(0) = i).$$

The analysis of semi-Markov processes is based on the results of renewal theory and the analysis of an EMC (of state-transition instances). The definition of a semi-Markov process requires knowledge of the kernel matrix $Q(t) = \{Q_{ij}(t)\}$ (for $t \geq 0$) and an initial distribution. It is commonly assumed that $X(t)$ possesses the Markov property at time $t = 0$.

4.5.1 Analysis Based on State Transitions

Let $X(t) \in S$ be a continuous-time semi-Markov process, $T_1, T_2, T_3, \dots$ the state-transition instances, and $\tau_1, \tau_2, \tau_3, \dots$ the consecutive sojourn times ($\tau_i = T_i - T_{i-1}$). We assume $T_0 = 0$. We intend to compute the state-transition probability $\pi_{ij}(t) = \mathbf{P}(X(t) = j \mid X(0) = i)$ assuming that the sojourn in the first state finishes at time h ($T_1 = h$), that is,

$$\pi_{ij}(t|T_1 = h) = \mathbf{P}(X(t) = j \mid X(0) = i, T_1 = h).$$

In this case

$$\pi_{ij}(t|T_1 = h) = \begin{cases} \delta_{ij}, & h \geq t, \\ \sum\limits_{k \in S} \mathbf{P}(X(T_1) = k \mid X(0) = i, T_1 = h)\, \pi_{kj}(t - h), & h < t, \end{cases} \tag{4.13}$$

where $\mathbf{P}(X(T_1) = j \mid X(0) = i, T_1 = h)$ is the probability that the process will start from state i at time 0 and is in state j right after the state transition at time T_1 assuming $T_1 = h$. In contrast with CTMCs, this probability depends on the sojourn time in state i:

$$\begin{aligned} \mathbf{P}(X(T_1) &= j \mid X(0) = i, T_1 = h) \\ &= \lim_{\Delta \to 0} \frac{\mathbf{P}(X(T_1) = j, h < T_1 \leq h + \Delta \mid X(0) = i)}{\mathbf{P}(h < T_1 \leq h + \Delta \mid X(0) = i)} \\ &= \lim_{\Delta \to 0} \frac{Q_{ij}(h + \Delta) - Q_{ij}(h)}{Q_i(h + \Delta) - Q_i(h)} = \frac{dQ_{ij}(h)}{dQ_i(h)}\ , \end{aligned} \tag{4.14}$$

where $Q_i(h)$ denotes the distribution of time spent in state i,

$$Q_i(t) = \mathbf{P}(T_1 \leq t \mid Z(0)=i) = \sum_j \mathbf{P}(Z(T_1)=j, T_1 \leq t \mid Z(0)=i) = \sum_j Q_{ij}(t).$$

It is commonly assumed that state transitions are real, which means that after staying in state i a state transition moves the process to a different state. This means that $Q_{ii}(t) = 0,\ \forall i \in S$. It is also possible to consider virtual state transitions from state i to state i, but this does not expand the set of semi-Markov processes and we do not consider it here. Note that the meaning of a diagonal element of a semi-Markov kernel matrix is completely different from that of a diagonal element of an infinitesimal generator of a CTMC. One of the technical consequences of this difference is the fact that we do not need to exclude the diagonal element from the summations over the set of states.

Two cases are considered in Eq. (4.13):

- If the time point of interest, t, is before the first state transition of the process ($h \geq t$), then the conditional state-transition probability is either 0 or 1 depending on the initial and final states. If the initial state i is identical with the final state j, then the transition probability is 1 because there is no state transition up to time t, otherwise it is 0.
- If the time point of interest, t, is after the first state transition of the process ($h < t$), then we need to evaluate the distribution of the next state k, assuming that the state transition occurs at time h, and after that the state-transition probability from the new state k to the final state j during time $t-h$, using the Markov property of the process at time h. The probability that the process moves to state k assuming it occurs at time h is $\frac{dQ_{ij}(h)}{dQ_i(h)}$, and the probability of its moving from state k to state j during an interval of length $t-h$ is $\pi_{ij}(t-h)$.

The distribution of the condition of Eq. (4.13) is known. The distribution of the sojourn time in state i is $Q_i(h)$. Using the law of total probability we obtain

$$\begin{aligned}
\pi_{ij}(t) &= \int_{h=0}^{\infty} \pi_{ij}(t|T_1 = h)\, \mathrm{d}F_{T_1}(h) \\
&= \int_{h=t}^{\infty} \delta_{ij}\, \mathrm{d}Q_i(t) + \int_{h=0}^{t} \sum_{k \in S} \frac{dQ_{ik}(h)}{dQ_i(h)}\, \pi_{kj}(t-h)\, \mathrm{d}Q_i(h) \\
&= \delta_{ij}\, (1 - Q_i(t)) + \int_{h=0}^{t} \sum_{k \in S} \pi_{kj}(t-h)\, \mathrm{d}Q_{ik}(h). \qquad (4.15)
\end{aligned}$$

Similar to the case of CTMCs, analysis based on the first state transition resulted in a Volterra integral equation also in the case of semi-Markov processes. The transient behavior of semi-Markov processes can be computed using the same numerical procedures.

Transform Domain Description We take the Laplace transform of both sides of the Volterra integral Eq. (4.15). The only nontrivial term is a convolution integral on the right-hand side:

$$\pi_{ij}^*(s) = \delta_{ij}\, (1 - Q_i^*(s)) + \sum_{k \in S} q_{ik}^*(s)\, \pi_{kj}^*(s),$$

where $q_{ik}(t) = \mathrm{d}Q_{ik}(t)/\mathrm{d}t$ and the transform domain functions are defined as $f^*(s) = \int_0^{\infty} f(t)\mathrm{e}^{-st}\mathrm{d}t$.

Introducing the diagonal matrix $D^*(s)$ composed of the elements $1 - Q_i^*(s)$, that is , $D^*(s) = \text{diag}\langle 1 - Q_i^*(s)\rangle$, the Laplace transforms of the state transition probabilities are obtained in matrix form,

$$\Pi^*(s) = D^*(s) + q^*(s)\Pi^*(s),$$

from which

$$\Pi^*(s) = [\mathbf{I} - q^*(s)]^{-1} D^*(s).$$

Stationary Behavior The stationary analysis of a semi-Markov process is very similar to the stationary analysis of a CTMC based on an EMC. Let the transition probability matrix of the EMC be Π. It is obtained from the kernel matrix through the following relation:

$$\Pi_{ij} = \mathbf{P}(Z(T_1) = j \mid Z(0) = i) = \lim_{t\to\infty} \mathbf{P}(Z(T_1) = j, T_1 \le t \mid Z(0) = i) = \lim_{t\to\infty} Q_{ij}(t).$$

The stationary distribution of the EMC $\hat{P}$ is the solution of the linear system $\hat{P} = \hat{P}\Pi, \sum_i \hat{P}_i = 1$. The stationary distribution of the semi-Markov process is

$$P_i = \frac{\hat{P}_i \hat{\tau}_i}{\sum\limits_j \hat{P}_j \hat{\tau}_j}, \tag{4.16}$$

where $\hat{\tau}_i$ is the mean time spent in state i. It can be computed from a kernel matrix using $\hat{\tau}_i = \int_0^\infty (1 - Q_i(t))\mathrm{d}t$.

Discrete-Event Simulation of Semi-Markov Processes The initial distribution and the $Q_{ij}(t)$ kernel completely define the stochastic behavior of a semi-Markov process. As a consequence, it is possible to simulate the process behavior based on them.

The key step of the simulation is to draw dependent samples for the sojourn time and the next visited state. This can be done based on the marginal distribution of one of the two random variables and a conditional distribution of the other one. Depending on which random variable is sampled first, there are two ways to simulate a semi-Markov process:

- When the process is in state i, first draw a $Q_i(t)$ distributed sample for the sojourn time, denoted by τ, then draw a sample for the next state assuming that the sojourn is τ based on the discrete probability distribution $\mathbf{P}(X(T_1) = j \mid X(0) = i, T_1 = \tau)$ $(\forall j \in S)$ given in Eq. (4.14).
- When the process is in state i, first draw a sample for the next visited state based on the discrete probability distribution $\Pi_{ij} = \mathbf{P}(X(T_1) = j \mid X(0) = i)$ $(\forall j \in S)$, then draw a sample for the sojourn time given in the next state with a distribution
$$\mathbf{P}(T_1 \le t \mid Z(0) = i, Z(T_1) = j) = \frac{Q_{ij}(t)}{\Pi_{ij}}. \tag{4.17}$$

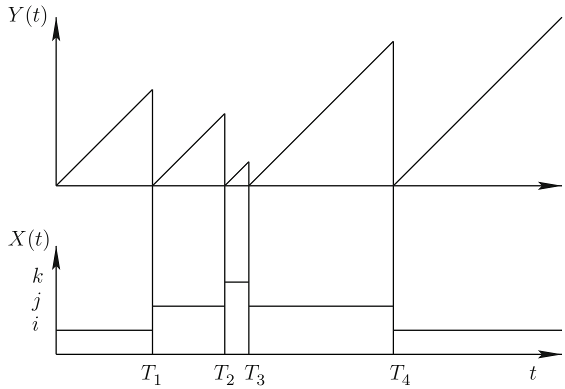


Fig. 4.3 Analysis of semi-Markov process with supplementary variable

4.5.2 Transient Analysis Using the Method of Supplementary Variables

A semi-Markov process does not possess the Markov property during its sojourn in a state. For example, the distribution of the time till the next state transition may depend on the amount of time that has passed since the last state transition. It is possible to extend the analysis of semi-Markov processes so that all information that makes the future evolution of the process conditionally independent of its past history is involved in the process description for $\forall t \geq 0$. It is indeed the Markov property for $\forall t \geq 0$. In the case of semi-Markov processes, this means that the discrete state of the process $X(t)$ and the time passed since the last state transition $Y(t) = t - \max(T_i \leq t)$ need to be considered together because the vector-valued stochastic process $\{X(t), Y(t)\}$ is already such that the future behavior of this vector process is conditionally independent of its past given the current value of the vector. That is, the $\{X(t), Y(t)\}$ process possesses the Markov property for $\forall t \geq 0$. The behavior of the $\{X(t), Y(t)\}$ process is depicted in Fig. 4.3.

This extension of a random process with an additional variable such that the obtained vector-valued process possesses the Markov property is referred to as the method of supplementary variables [24].

With $X(t)$ and $Y(t)$ and the kernel matrix of the process we can compute the distribution of time till the next state transition at any time instant; this is commonly referred to as the remaining sojourn time in the given state. If at time t the process stays in state i for a period of τ $[X(t) = i, Y(t) = \tau]$ and the distribution of the total sojourn time in state i is $Q_i(t)$, then the distribution of the remaining sojourn time in state i, denoted by γ, is

$$\mathbf{P}(\gamma \leq t) = \mathbf{P}(\gamma_t \leq t + \tau \mid \gamma_t > \tau) = \frac{Q_i(t+\tau) - Q_i(\tau)}{1 - Q_i(\tau)},$$

where γ_t denotes the total time spent in state i during this visit in state i.

To analyze the $\{X(t), Y(t)\}$ process, we need to characterize the joint distribution of the following two quantities:

$$h_i(t,x) = \frac{\mathbf{P}\left(X(t) = i, x \le Y(t) < x + \Delta\right)}{\Delta}.$$

It is possible to obtain $h_i(t,x)$ based on the analysis of the short-term behavior of CTMCs:

$$\begin{aligned}
&h_i(t+\Delta, x) \\
&\quad = \mathbf{P}[\text{there is no state transition in the interval } (t, t+\Delta)] \\
&\qquad \times h_i(t+\Delta, x \mid \text{there is no state transition}) \\
&\qquad + \mathbf{P}[\text{there is one state transition in the interval } (t, t+\Delta)] \\
&\qquad \times h_i(t+\Delta, x \mid \text{there is one state transition}) + o(\Delta),
\end{aligned}$$

where $h_i(t+\Delta, x \mid \text{condition})$ denotes $\dfrac{\mathbf{P}\left(X(t) = i, x \le Y(t) < x+\Delta \mid \text{condition}\right)}{\Delta}$. The probability of the state transition can be computed based on the distribution of the remaining sojourn time:

$$\begin{aligned}
&\mathbf{P}[\text{there is one state transition in the interval } (t, t+\Delta)] \\
&\quad = \mathbf{P}\left(\text{remaining sojourn time} \le \Delta\right) = \frac{Q_i(x+\Delta) - Q_i(x)}{1 - Q_i(x)},
\end{aligned}$$

from which

$$\mathbf{P}[\text{there is no state transition in the interval } (t, t+\Delta)] = \frac{1 - Q_i(x+\Delta)}{1 - Q_i(x)}.$$

Immediately following a state transition $Y(t)$ is reset to zero. Consequently, the probability that $Y(t+\Delta) = x$ for a fixed $x > 0$ is zero when Δ is sufficiently small. That is,

$$h_i[t+\Delta, x \mid \text{there is one state transition in the interval } (t, t+\Delta)] = 0 \ \text{ if } \ x > 0.$$

It follows that

$$\begin{aligned}
&h_i(t+\Delta, x) \\
&\quad = \mathbf{P}[\text{there is no state transition in the interval } (t, t+\Delta)] \\
&\qquad \times h_i[t+\Delta, x \mid \text{there is no state transition in the interval } (t, t+\Delta)] \\
&\quad = \frac{1 - Q_i(x+\Delta)}{1 - Q_i(x)} \cdot h_i(t, x-\Delta).
\end{aligned}$$

Analysis of the process $\{X(t), Y(t)\}$ is made much simpler by the use of the transition rate of α_i instead of its distribution $Q_i(t)$. The transition rate is defined by

$$\lambda_i(t) = \lim_{\Delta\to 0} \frac{\mathbf{P}(\alpha_i \le t + \Delta \mid \alpha_i > t)}{\Delta} = \lim_{\Delta\to 0} \frac{Q_i(t+\Delta) - Q_i(t)}{\Delta\,(1 - Q_i(t))} = \frac{Q_i'(t)}{1 - Q_i(t)} .$$

It is also referred to as the hazard rate in probability theory. The probability of a state transition can be written in the following form:

$$\begin{aligned} &\mathbf{P}[\text{there is one state transition in the interval } (t, t+\Delta)] \\ &= \frac{Q_i(x+\Delta) - Q_i(x)}{1 - Q_i(x)} = \lambda_i(x)\Delta + o(\Delta) , \end{aligned}$$

from which

$$\mathbf{P}[\text{there is no state transition in the interval } (t, t+\Delta)] = 1 - \lambda_i(x)\Delta + o(\Delta) .$$

Based on all of these expressions, $h_i(t, x)$ satisfies

$$h_i(t+\Delta, x) = \Big(1 - \lambda_i(x)\Delta + o(\Delta)\Big)\, h_i(t, x-\Delta).$$

From this difference equation we can go through the usual steps to obtain the partial differential equation for $h_i(t, x)$. First we move $h_i(t, x-\Delta)$ to the other side,

$$h_i(t+\Delta, x) - h_i(t, x-\Delta) = \Big(-\lambda_i(x)\Delta + o(\Delta)\Big)\, h_i(t, x-\Delta),$$

then we add and subtract $h_i(t, x)$,

$$h_i(t+\Delta, x) - h_i(t, x) + h_i(t, x) - h_i(t, x-\Delta) = \Big(-\lambda_i(x)\Delta + o(\Delta)\Big)\, h_i(t, x-\Delta),$$

and reorder the terms,

$$\frac{h_i(t+\Delta, x) - h_i(t, x)}{\Delta} + \frac{h_i(t, x) - h_i(t, x-\Delta)}{\Delta} = \left(-\lambda_i(x) + \frac{o(\Delta)}{\Delta}\right) h_i(t, x-\Delta).$$

Finally, the $\Delta \to 0$ transition results in

$$\frac{\partial h_i(t, x)}{\partial t} + \frac{\partial h_i(t, x)}{\partial x} = -\lambda_i(x)\, h_i(t, x). \tag{4.18}$$

This partial differential equation describes $h_i(t, x)$ for $x > 0$. The case of $x = 0$ requires a different treatment:

$$\begin{aligned}&\mathbf{P}\left(X(t+\Delta)=i, Y(t)\le \Delta\right)\\&=\sum_{k\in S,k\neq i}\int_{x=0}^{\infty}\mathbf{P}\left(X(t)=k, Y(t)=x, \text{one transition to state } i \text{ in } (t,t+\Delta)\right)\mathrm{d}x.\end{aligned}$$

The probability that in the interval $(t, t+\Delta)$ the process moves from state k to state i is

$$\begin{aligned}&\mathbf{P}[\text{there is one state transition in the interval } (t,t+\Delta) \text{ from } k \text{ to } i]\\&\quad=\mathbf{P}[\text{one state transition in the interval } (t,t+\Delta)]\\&\qquad\times\mathbf{P}[\text{state transition from } k \text{ to } i \mid \text{one state transition in the interval } (t,t+\Delta)]\\&\quad=\frac{Q_k(x+\Delta)-Q_k(x)}{1-Q_k(x)}\cdot\frac{Q_{ki}(x+\Delta)-Q_{ki}(x)}{Q_k(x+\Delta)-Q_k(x)},\end{aligned}$$

where the second term is already known from Eq. (4.14). We can also introduce the intensity of transition from k to i:

$$\begin{aligned}\lambda_{ki}(x)&=\lim_{\Delta\to 0}\frac{\mathbf{P}[\text{there is a transition in the interval } (t,t+\Delta) \text{ from } k \text{ to } i]}{\Delta}\\&=\lim_{\Delta\to 0}\frac{Q_{ki}(x+\Delta)-Q_{ki}(x)}{\Delta(1-Q_k(x))}=\frac{Q'_{ki}(x)}{1-Q_k(x)}.\end{aligned}$$

The transition probability can be written in the form

$$\mathbf{P}[\text{there is a transition in the interval } (t,t+\Delta) \text{ from } k \text{ to } i]=\lambda_{ki}(x)\Delta+o(\Delta).$$

Using this we can write

$$\begin{aligned}\mathbf{P}(X(t+\Delta)=i, Y(t)\le\Delta)&=h_i(t+\Delta,0)\Delta\\&=\sum_{k\in S,k\neq i}\int_{x=0}^{\infty}\left(\lambda_{ki}(x)\Delta+o(\Delta)\right)h_k(t,x)\,\mathrm{d}x,\end{aligned}$$

from which a multiplication with Δ and the $\Delta\to 0$ transition result in

$$h_i(t,0)=\sum_{k\in S,k\neq i}\int_{x=0}^{\infty}\lambda_{ki}(x)\,h_k(t,x)\,\mathrm{d}x. \tag{4.19}$$

In summary, the method of supplementary variable allows for the analysis of the process $\{X(t), Y(t)\}$ through the function $h_i(t,x)$, which is given by a partial differential equation (4.18) for $x>0$ and a boundary equation (4.19) for $x=0$. Based on these equations and the initial distributions of $h_i(0,x)$ for $\forall i\in S$ numerical partial differential solutions methods can be applied to compute the transient behavior of a semi-Markov process.

Stationary Behavior If the limit $\lim_{t\to\infty} h_i(t,x) = h_i(x)$ exists for all states $i \in S$, then we can evaluate the limit $t \to \infty$ of Eqs. (4.18) and (4.19)

$$\frac{\mathrm{d}h_i(x)}{\mathrm{d}x} = -\lambda_i(x)\, h_i(x), \tag{4.20}$$

$$h_i(0) = \sum_{k\in S, k\neq i} \int_{x=0}^{\infty} \lambda_{ki}(x)\, h_k(x)\, \mathrm{d}x. \tag{4.21}$$

The solution of ordinary differential Eq. (4.20) is

$$h_i(x) = h_i(0)\mathrm{e}^{\int_{u=0}^{x} -\lambda_i(u)\, \mathrm{d}u},$$

where the unknown quantity is $h_i(0)$. It can be obtained from Eq. (4.21) as follows:

$$h_i(0) = \sum_{k\in S, k\neq i} \int_{x=0}^{\infty} \lambda_{ki}(x)\, h_k(0)\mathrm{e}^{\int_{u=0}^{x} -\lambda_k(u)\, \mathrm{d}u}\, \mathrm{d}x$$

$$= \sum_{k\in S, k\neq i} h_k(0) \int_{x=0}^{\infty} \lambda_{ki}(x)\mathrm{e}^{\int_{u=0}^{x} -\lambda_k(u)\, \mathrm{d}u}\, \mathrm{d}x,$$

where

$$\int_{x=0}^{\infty} \lambda_{ki}(x)\mathrm{e}^{\int_{u=0}^{x} -\lambda_k(u)\, \mathrm{d}u}\, \mathrm{d}x = \mathbf{P}\,(\text{after state } k \text{ the process moves to state } i) = \Pi_{ki}.$$

That is, we are looking for the solution of the linear system

$$h_i(0) = \sum_{k\in S, k\neq i} h_k(0)\, \Pi_{ki} \quad \forall i \in S$$

with the normalizing condition

$$\sum_{i\in S} \int_{x=0}^{\infty} h_i(x)\, \mathrm{d}x = 1,$$

where the normalizing condition is the sum of the stationary-state probabilities. From

$$\sum_{i\in S} \int_{x=0}^{\infty} h_i(x)\, \mathrm{d}x = \sum_{i\in S} h_i(0) \int_{x=0}^{\infty} \mathrm{e}^{\int_{u=0}^{x} -\lambda_i(u)\, \mathrm{d}u}\, \mathrm{d}x = \sum_{i\in S} h_i(0)\, \hat{\tau}_i = 1$$

and Eq. (4.16) we have that the required solution is

$$h_i(0) = \frac{\hat{P}_i}{\sum_j \hat{P}_j \hat{\tau}_j}.$$

4.6 Markov Regenerative Process

Definition 4.57. The $X(t)$ discrete-state, continuous-time, time-homogeneous stochastic process is a **Markov regenerative process** if there exists a random time series $T_0, T_1, T_2, \ldots$ ($T_0 = 0$) such that the $X(t)$ process possesses the Markov property at time $T_0, T_1, T_2, \ldots$ [23, 58] (Fig. 4.4).

Compared to the properties of semi-Markov processes, where the process possesses the Markov property at all state-transition points, the definition of Markov regenerative processes is less restrictive. It allows that at some state-transition point the process does not possess the Markov property, but the analysis of Markov regenerative processes is still based on the occurrence of time points where the process possesses the Markov property.

Since Definition 4.57 does not address the behavior of the process between the consecutive time points $T_0, T_1, T_2, \ldots$, Markov regenerative processes can be fairly general stochastic processes. In practice, the use of a renewal theorem for the analysis of these processes is meaningful only when the stochastic behavior between the consecutive time points $T_0, T_1, T_2, \ldots$ is easy to analyze.

A common method for analyzing Markov regenerative processes is based on the next time point with the Markov property (T_1).

Definition 4.58. The series of random variables $\{Y_n, T_n; n \geq 0\}$ is a time-homogeneous **Markov renewal series** if

$$\mathbf{P}(Y_{n+1} = y, T_{n+1} - T_n \leq t \mid Y_0, \ldots, Y_n, T_0, \ldots, T_n)$$

$$= \mathbf{P}(Y_{n+1} = y, T_{n+1} - T_n \leq t \mid Y_n)$$

$$= \mathbf{P}(Y_1 = y, T_1 - T_0 \leq t \mid y_0)$$

for all $n \geq 0$, $y \in S$, and $t \geq 0$.

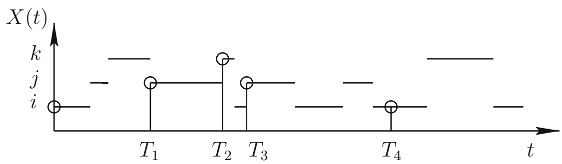


Fig. 4.4 Markov regenerative process; *circles* denote points with Markov property

It can be seen from the definition of Markov renewal series that the series $Y_0, Y_1, \ldots$ is a DTMC. According to Definition 4.57, the sequence of states $X(T_i)$ of a Markov regenerative process at the time sequence T_i instants with the Markov property and the time sequence T_i instants with the Markov property form a Markov renewal sequence $\{X(T_i), T_i\}$ $(i = 0, 1, \ldots)$.

Analysis of Markov regenerative processes is based on this *embedded* Markov renewal series. To this end the joint distribution of the next time point and the state in that time point must be known. In contrast with the similar kernel of semi-Markov processes, in the case of Markov regenerative processes, the kernel is denoted by

$$K_{ij}(t) = \mathbf{P}\left(X_1 = j, T_1 - T_0 \le t \mid X_0 = i\right), \quad i, j \in S,$$

and the matrix $K(t) = \{K_{ij}(t)\}$ is referred to as the global kernel of a Markov regenerative process. The global kernel of a Markov regenerative process completely defines the stochastic properties of the Markov regenerative process at time points with the Markov property. The description of the process between those time points is complex, but for a transient analysis of the process (more precisely for computing transient-state probabilities) it is enough to know the transient-state probabilities between consecutive time points with the Markov property. This is given by the local kernel matrix of the Markov regenerative process $E(t) = \{E_{ij}(t)\}$ whose elements are

$$E_{ij}(t) = \mathbf{P}\left(X(t) = j,\ T_1 > t,\ \mid Z(0) = i\right),$$

where $E_{ij}(t)$ is the probability that the process will start in state i, the first point with the Markov property will be later than t, and the process will stay in state j at time t.

4.6.1 *Transient Analysis Based on Embedded Markov Renewal Series*

Let the transient-state transition probability matrix be $\Pi(t)$ whose elements are

$$\Pi_{ij}(t) = \mathbf{P}\left(X(t) = j \mid X(0) = i\right).$$

Assuming that $T_1 = h$, we can compute the conditional state-transition probability as follows:

$$\Pi_{ij}(t \mid T_1 = h) = \begin{cases} \mathbf{P}\left(X(t) = j \mid T_1 = h,\ X(0) = i\right), & h > t, \\ \displaystyle\sum_{k \in S} \mathbf{P}\left(X(T_1) = k \mid X(0) = i, T_1 = h\right) \cdot \Pi_{kj}(t - h), & h \le t. \end{cases} \tag{4.22}$$

Similar to the transient analysis of semi-Markov processes, Eq. (4.22) describes two exclusive cases: $h \le t$ and $h > t$. In the case of semi-Markov processes, the $h > t$ case results in 0 or 1; in the case of a Markov regenerative process, the conditional probability for $h > t$ can be different from 0 or 1 because the process can have state transitions also before T_1.

Using the distribution of T_1 and the formula of total probability we obtain

$$\begin{aligned}\Pi_{ij}(t) = &\int_{h=t}^{\infty} \mathbf{P}\left(X(t) = j \mid T_1 = h,\ X(0) = i\right) \mathrm{d}K_i(h) \\ &+ \int_{h=0}^{t} \sum_{k \in S} \frac{dK_{ik}(t)}{dK_i(t)}\ \Pi_{kj}(t-h)\, \mathrm{d}K_i(h)\,. \end{aligned} \tag{4.23}$$

Let us consider the first term on the right-hand side:

$$\begin{aligned}&\int_{h=t}^{\infty} \mathbf{P}\left(X(t) = j \mid T_1 = h,\ X(0) = i\right) \mathrm{d}K_i(h) \\ &= \int_{h=t}^{\infty} \lim_{\Delta \to 0} \mathbf{P}\left(X(t) = j \mid h \le T_1 < h + \Delta,\ X(0) = i\right)\ \mathrm{d}K_i(h) \\ &= \int_{h=t}^{\infty} \lim_{\Delta \to 0} \frac{\mathbf{P}\left(X(t) = j,\ h \le T_1 < h + \Delta \mid X(0) = i\right)}{\mathbf{P}\left(h \le T_1 < h + \Delta,\ \mid X(0) = i\right)}\ \mathrm{d}K_i(h) \\ &= \int_{h=t}^{\infty} \frac{d_h\, \mathbf{P}\left(X(t) = j,\ T_1 < h \mid X(0) = i\right)}{dK_i(h)}\ \mathrm{d}K_i(h) \\ &= \mathbf{P}\left(X(t) = j,\ t < T_1 \mid X(0) = i\right),\end{aligned}$$

from which

$$\Pi_{ij}(t) \ = \ E_{ij}(t) \ + \sum_{k \in S} \int_{h=0}^{t} \Pi_{kj}(t-h)\, \mathrm{d}K_{ik}(h)\,. \tag{4.24}$$

Assuming that $K(t)$ is derivable and $dK(t)/dt = k(t)$ we have

$$\Pi_{ij}(t) \ = \ E_{ij}(t) \ + \sum_{k \in S} \int_{h=0}^{t} \Pi_{kj}(t-h)\, k_{ik}(h)\, \mathrm{d}h. \tag{4.25}$$

Similar to the transient analysis of CTMCs and semi-Markov processes we obtain a Volterra equation for the transient analysis of Markov regenerative processes.

Transform Domain Description The Laplace transform of Eq. (4.25) is

$$\Pi^*_{ij}(s) \ = \ E^*_{ij}(s) + \sum_{k \in \Omega} k^*_{ik}(s)\, \Pi^*_{kj}(s), \tag{4.26}$$

which can be written in matrix form:

$$\Pi^*(s) \;=\; E^*(s) \;+\; k^*(s)\,\Pi^*(s). \tag{4.27}$$

The solution of Eq. (4.27) is

$$\Pi^*(s) \;=\; [\mathbf{I} - k^*(s)]^{-1}\, E^*(s). \tag{4.28}$$

Based on Eq. (4.28), numerical inverse Laplace methods can also be used for the transient analysis of Markov regenerative processes.

Stationary Behavior Despite the differences between semi-Markov and Markov regenerative processes, their stationary analysis follows the same steps. The state-transition probability of the DTMC embedded in time points with the Markov property is

$$\Pi_{ij} = \mathbf{P}\left(Z(T_1) = j \mid Z(0) = i\right) = \lim_{t\to\infty} \mathbf{P}\left(Z(T_1) = j, T_1 \le t \mid Z(0) = i\right) = \lim_{t\to\infty} K_{ij}(t).$$

The stationary distribution of the EMC is the solution of $\hat{P} = \hat{P}\Pi, \sum_i \hat{P}_i = 1$. Now we need to compute the mean time spent in the different states during the interval (T_0, T_1). Fortunately, the local kernel carries the necessary information. Let τ_{ij} be the mean time the process spends in state j during the interval (T_0, T_1) assuming that it starts from state i $(X(T_0) = i)$. Then

$$\begin{aligned}
\tau_{ij} &= \mathbf{E}\left(\int_{t=0}^{\infty} \mathcal{I}_{\{X(t)=j, T_1>t \;\mid\; X(0)=i\}} \mathrm{d}t\right) \\
&= \int_{t=0}^{\infty} \mathbf{P}\left(X(t) = j, T_1 > t \mid X(0) = i\right)\mathrm{d}t \\
&= \int_{t=0}^{\infty} E_{ij}(t)\mathrm{d}t,
\end{aligned}$$

where $\mathcal{I}_{\{\bullet\}}$ is the indicator of event $\bullet$. The mean length of the interval (T_0, T_1) is

$$\tau_i = \sum_{j\in S} \tau_{ij}\,.$$

Finally, the stationary distribution of the process can be computed as

$$P_i = \frac{\sum_{j\in S} \hat{P}_j\ \tau_{ji}}{\sum_{j\in S} \hat{P}_j\ \tau_j}.$$

4.7 Exercises

Exercise 4.1. Applying Theorem 4.42, find the limit (stationary) distributions of age, residual lifetime, and total lifetime $[\delta(t) = t - t_{N(t)},\ \gamma(t) = t_{N(t)+1} - t,\ \beta(t) = t_{N(t)+1} - t_{N(t)}]$ if the interarrival times are independent random variables having a joint exponential distribution with the parameter λ. Show the expected values for the limit distributions.

Exercise 4.2 (*Ergodic property of semi-Markov processes).* Consider a system with the finite state space $\mathcal{X} = \{1, \ldots, N\}$. The system begins to work at the moment $T_0 = 0$ in a state $X_0 \in \mathcal{X}$ and changes states at the random moments $0 < T_1 < T_2 < \ldots$. Denote by $X_1, X_2, \ldots$ the sequence of consecutive states of the system, and suppose that it constitutes a homogeneous, irreducible, and aperiodic Markov chain with initial distribution $(p_i = \mathbf{P}(X_0 = i),\ 1 \le i \le N)$ and probability transition matrix $\Pi = \left(p_{ij}\right)_{i,j=1}^{n}$. Define the process $X(t) = X_{n-1},\ T_{n-1} \le t < T_n,\ n = 1, 2, \ldots$, assume that the sequence of holding times $Y_k = T_k - T_{k-1},\ k = 1, 2, \ldots$, depends only conditionally on the states $X_{k-1} = i$ and $X_k = j$, and denote $F_{ij}(x) = \mathbf{P}(Y_k \le x \mid X_{k-1} = i,\ X_k = j)$ if $p_{ij} > 0$, where $\nu_{ij} = \int_0^\infty x \mathrm{d}F_{ij}(x) < \infty$.

Find the limits for

(a) The average number of transitions/time;
(b) The relative frequencies of states i in the sequence $X_0, X_1, \ldots$;
(c) The limit distribution $\mathbf{P}(X_t = i),\ i \in X$;
(d) The average time spent in a state $i \in X$.