# 1

# Introduction

*This introductory chapter has a twofold purpose: to answer the basic question "Why do we think that the topic presented in the book is interesting and worth studying?" and to give an overall presentation of the main features discussed in the book.*

## 1.1 Object of the Study

First, we want to look at the main reasons which motivated us to study the topic of this book. Basically, here we answer two questions: "Why semi-Markov?" and "Why work in discrete time?"

### 1.1.1 The Underlying Idea in Semi-Markov Models

Much work has been carried out in the field of Markov processes, and a huge amount of Markov process applications can be found in the literature of the last 50 years. One of the reasons for applying Markov process theory in various fields is that the Markovian hypothesis is very intuitive and convenient when dealing with applications and the underlying computations are quite simple. One can formulate this hypothesis as follows: if the past and the present of a system are known, then the future evolution of the system is determined only by its present state, or equivalently, the past and the future are conditionally independent given the present (state). Thus the past history of a system plays no role in its future evolution, which is usually known as the "memoryless property of a Markov process." But the Markovian hypothesis imposes restrictions on the distribution of the sojourn time in a state, which should be exponentially distributed (continuous case) or geometrically distributed (discrete case). This is the main drawback when applying Markov processes in real applications.



V.S. Barbu, N. Limnios, *Semi-Markov Chains and Hidden Semi-Markov Models toward Applications*, DOI: 10.1007/978-0-387-73173-5_1,

What came naturally was to relax the underlying Markov assumption in order to:

- Allow arbitrarily distributed sojourn times in any state;
- Still have the Markovian hypothesis, but in a more flexible manner.

A process that has these two properties will be called a *semi-Markov process*.

To be more specific, let us consider a random system with finite state space $E = \{1, \ldots, s\}$, whose evolution in time is governed by a stochastic process $Z = (Z_k)_{k\in\mathbb{N}}$. We note here that all stochastic processes taken into account throughout this book are considered to evolve in discrete time. We will use the term *chain* for a discrete-time stochastic process.

Let us also denote by $S = (S_n)_{n\in\mathbb{N}}$ the successive time points when state changes in $(Z_n)_{n\in\mathbb{N}}$ occur and by $J = (J_n)_{n\in\mathbb{N}}$ the chain which records the visited states at these time points. Let $X = (X_n)_{n\in\mathbb{N}}$ be the successive sojourn times in the visited states. Thus, $X_n = S_n - S_{n-1}$, $n \in \mathbb{N}^*$, and, by convention, we set $X_0 = S_0 = 0$.

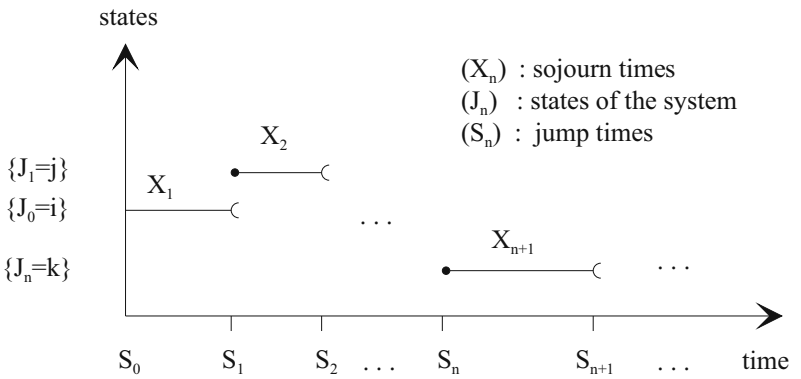


**Fig. 1.1.** Sample path of a semi-Markov chain

The relation between process $Z$ and process $J$ of the successively visited states is given by

$$Z_k = J_{N(k)}, \text{ or, equivalently, } J_n = Z_{S_n}, n, k \in \mathbb{N},$$

where $N(k) := \max\{n \in \mathbb{N} \mid S_n \leq k\}$ is the discrete-time counting process of the number of jumps in $[1, k] \subset \mathbb{N}$. Thus $Z_k$ gives the system's state at time $k$.

Suppose the following conditional independence relation holds true almost surely:

$$\begin{aligned}&\mathbb{P}(J_{n+1}=j, S_{n+1}-S_n=k \mid J_0,\ldots,J_n; S_0,\ldots,S_n)\\ &=\mathbb{P}(J_{n+1}=j, S_{n+1}-S_n=k \mid J_n). \qquad (1.1)\end{aligned}$$

This equation means: if we know the past visited states and jump times of the system, as well as its present state, the future visited state and the sojourn time in the present state depend only on the present state. In other words, we basically have a Markovian-type hypothesis, with the difference that the memoryless property does not act on the calendar time $(0,1,\ldots,k,k+1,\ldots)$ but on a time governed by the jump time process $J$, $(J_0,J_1,\ldots,J_n,J_{n+1},\ldots)$. This is what we called before a more flexible Markovian hypothesis.

If Equation (1.1) holds true, then $Z=(Z_n)_{n\in\mathbb{N}}$ is called a *semi-Markov chain* (SMC) and the couple $(J,S)=(J_n,S_n)_{n\in\mathbb{N}}$ is called a *Markov renewal chain* (MRC). Moreover, if the right-hand-side term of Relation (1.1) is independent of $n$, then $Z$ and $(J,S)$ are said to be (time) homogeneous and we define the *discrete-time semi-Markov kernel* $\mathbf{q}=(q_{ij}(k);\ i,j\in E,k\in\mathbb{N})$ by

$$q_{ij}(k):=\mathbb{P}(J_{n+1}=j, X_{n+1}=k \mid J_n=i).$$

The semi-Markov kernel $\mathbf{q}$ is the essential quantity which defines a semi-Markov chain (together with an initial distribution $\boldsymbol{\alpha}$, $\alpha_i:=\mathbb{P}(J_0=i), i\in E$). All the work in this book will be carried out for homogeneous Markov renewal/semi-Markov chains.

A few remarks are in order at this point.

First, Equation (1.1), together with the hypothesis of time homogeneity, shows that the visited-state chain $(J_n)_{n\in\mathbb{N}}$ is a homogeneous Markov chain (MC), called the embedded Markov chain (EMC). We denote by $\mathbf{p}=(p_{ij})_{i,j\in E}$ its transition matrix,

$$p_{ij}=\mathbb{P}(J_{n+1}=j \mid J_n=i),\ i,j\in E,\ n\in\mathbb{N}.$$

We do not allow transitions to the same state, i.e., we set $p_{ii}=0$ for any $i\in E$.

Second, let us set $\mathbf{f}=(f_{ij}(k);\ i,j\in E,k\in\mathbb{N})$ for the sojourn time distributions, conditioned by the next state to be visited, $f_{ij}(k)=\mathbb{P}(X_{n+1}=k \mid J_n=i,J_{n+1}=j)$. Obviously, for any states $i,j\in E$ and nonnegative integer $k$ we have

$$q_{ij}(k)=p_{ij}f_{ij}(k).$$

We want the chain to spend at least one time unit in a state, that is, $f_{ij}(0)=q_{ij}(0)=0$, for any states $i,j$.

Consequently, the evolution of a sample path of a semi-Markov chain can be described as follows: the first state $i_0$ is chosen according to the initial distribution $\boldsymbol{\alpha}$; then, the next visited state $i_1$ is determined according to the transition matrix $\mathbf{p}$ and the chain stays in state $i_0$ for a time $k$ determined by the sojourn time distribution in state $i_0$ before going to state $i_1$,

$(f_{i_0 i_1}(k);\ k \in \mathbb{N})$. Note that the sojourn time distributions $(f_{ij}(k);\ k \in \mathbb{N})$, $i, j \in E$ can be any discrete distribution (or continuous distribution in the continuous case), as opposed to the sojourn time constraints in the Markov case. This is why the semi-Markov processes are much more general and better adapted to applications than the Markov ones.

We note here that a Markov chain is a particular case of a semi-Markov chain (Example 3.2).

### 1.1.2 Discrete Time

As mentioned above, all the work presented in this book is carried out in discrete time. Here we simply want to explain why we think that discrete time is important.

Since the introduction of semi-Markov processes in the mid-1950s, their probabilistic and statistical properties have been widely studied in the continuous-time case. By contrast, discrete-time semi-Markov processes are almost absent from the literature (see Chapter 3 for references). In the authors' opinion, there are at least two reasons why discrete-time semi-Markov processes are interesting and worth studying.

The first reason for our interest in discrete time comes from specific semi-Markov applications where the time scale is intrinsically discrete. For instance, in reliability theory, one could be interested in the number of cycles done by a system or in the number of times (hours, days, etc.) that a specific event occurs. Note also that in DNA analysis, any natural (hidden) semi-Markov approach is based on discrete time because we are dealing with discrete sequences of, say, the four bases A, C, G, T.

The second reason relies on the simplicity of modeling and calculus in discrete time. A discrete-time semi-Markov process makes only a bounded number of jumps in a finite time interval (a fortiori, it does not explode). For this reason, any quantity of interest in a discrete-time semi-Markov model can be expressed as a finite series of semi-Markov kernel convolution products instead of an infinite series as in the continuous case.

Let us briefly explain this phenomenon. As will be seen in Chapter 3, the functionals of the semi-Markov kernel that we are interested in can be expressed as finite combinations of $\sum_{n=0}^{\infty} \mathbf{q}^{(n)}(k)$, $k \in \mathbb{N}$, with $\mathbf{q}^{(n)}$ the $n$-fold convolution of the semi-Markov kernel. In Proposition 3.1 we will see that for any states $i, j \in E$, the $(i, j)$ element of the $n$th kernel convolution power has the probabilistic expression

$$q_{ij}^{(n)}(k) = \mathbb{P}(J_n = j, S_n = k \mid J_0 = i),$$

which means: $q_{ij}^{(n)}(k)$ is the probability that, starting from state $i$ at initial time, the semi-Markov chain will do the $n$th jump at time $k$ to state $j$. But the

SMC stays at least one unit of time in a state, so we obviously have $\mathbf{q}^{(n)}(k) = 0$ for any $n \geq k+1$ (Lemma 3.2). Consequently, the infinite series $\sum_{n=0}^{\infty} \mathbf{q}^{(n)}(k)$ becomes the finite series $\sum_{n=0}^{k} \mathbf{q}^{(n)}(k)$, and, as previously mentioned, all the quantities of interest can be written as finite series.

For these reasons, we think that it is interesting and legitimate to have probabilistic and statistical tools for (hidden) semi-Markov models (SMMs), adapted to the discrete case. This is also the case in Markov process studies, where we have a discrete-time theory and a continuous-time theory.

## 1.2 Discrete-Time Semi-Markov Framework

Our main objective concerning the semi-Markov models is to provide estimators for the main characteristics and to investigate their asymptotic properties. In order to achieve this, we first need to present briefly the classical theory of discrete-time renewal processes and then give the basic definitions and properties of a semi-Markov chain. We also need some basic definitions and results on Markov chains, which can be found in Appendix D.

### 1.2.1 Discrete-Time Renewal Processes

Roughly speaking, a *renewal process* (RP) $(S_n)_{n\in\mathbb{N}}$ represents the successive instants when a specific (fixed but random) event occurs. The term renewal comes from the assumption that when this event occurs, the process starts anew (this is a regeneration point of time). For this reason, that specific event will be called a *renewal* and $(S_n)_{n\in\mathbb{N}}$ will be called a *renewal process*. Since we will be concerned only with discrete-time renewal processes, we will generally use the term *renewal chain* (RC).

For a renewal process we have

$$S_n := X_0 + X_1 + \ldots + X_n,$$

where $X_n,\ n \in \mathbb{N}$, called *waiting times*, represent the times between two successive renewals. Note that the fundamental fact that the chain starts anew each time a renewal occurs means that $(X_n)_{n\in\mathbb{N}^*}$ is a sequence of i.i.d. random variables (in the simplest case, we suppose $X_0 = S_0 = 0$).

In Figure 1.2 we present the *counting process of the number of renewals* in the time interval $[1, n]$, denoted by $N(n)$.

The need for studying renewal chains before investigating the probabilistic behavior of semi-Markov chains comes from the fact that, as will be seen in Chapters 3 and 4, a semi-Markov chain can be analyzed through the so-called embedded renewal chains. That means that by taking into account only some particular aspects of the evolution of a SMC (successive visits of a specific

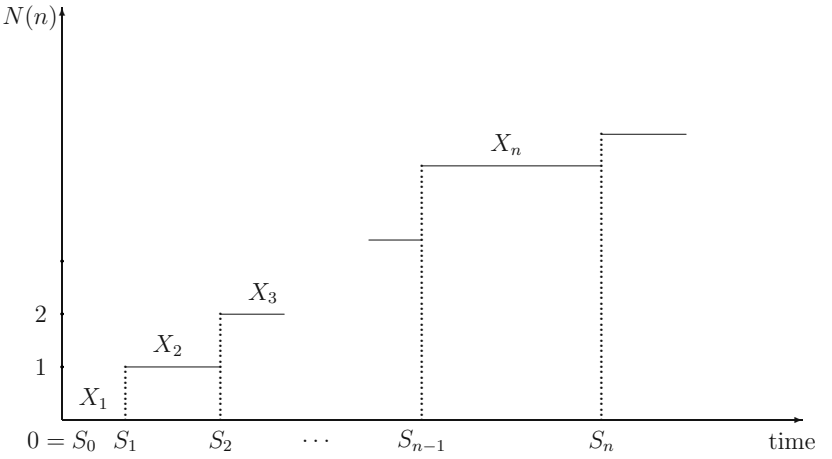


**Fig. 1.2.** Sample path of a renewal chain

state, for instance), we obtain a renewal chain. Due to this property, results on RCs will be of great help when investigating the behavior of SMCs.

For these reasons, Chapter 2 is devoted to renewal chains. First, we give some definitions and notation on RCs and introduce the basic notion of *renewal equation in discrete time*, given by

$$g_n = b_n + \sum_{k=0}^{n} f_k g_{n-k}, \quad n \in \mathbb{N}, \tag{1.2}$$

where $b = (b_n)_{n\in\mathbb{N}}$ is a known sequence, $g = (g_n)_{n\in\mathbb{N}}$ is an unknown sequence, and $f_k := \mathbb{P}(X_1 = k)$ is the *waiting time distribution of the renewal chain* (i.e., the distribution of the time elapsed between two successive renewals). We also suppose that the chain stays at least one time unit in a state, i.e., $f_0 = 0$.

Under some conditions, the solution of the renewal equation is given by (Theorem 2.2)

$$g_n = (u * b)_n, \quad n \geq 1, \tag{1.3}$$

where $u_n$ represents the probability that a renewal will occur at time $n$ and $*$ is the usual sequence convolution product (Definition 2.2).

Second, we study the asymptotic properties of quantities related to a renewal chain. Two cases – one periodic and the other aperiodic – are considered (Definitions 2.4 and 2.5). In this introductory part we consider only the aperiodic case. The main results can be summarized as follows.

- We give the strong law of large numbers (SLLN) and the central limit theorem (CLT) for the counting chain $(N(n))_{n\in\mathbb{N}^*}$ of the number of renewals in the time interval $[1, n]$.
- Another result consists in the asymptotic behavior of the expected number of renewals up to time $n$, denoted by $\Psi(n)$ and called *renewal function*. This is the so-called *elementary renewal theorem*, which states that

$$\lim_{n\to\infty} \frac{\Psi(n)}{n} = \frac{1}{\mu},$$

  where $\mu := \mathbb{E}(X_1)$ is the expected waiting time between two successive renewals.
- Similarly, the limit probability that a renewal will ever occur is provided by the *renewal theorem*:

$$\lim_{n\to\infty} u_n = \frac{1}{\mu}.$$

- A crucial result is the *key renewal theorem*, which gives the limit of the solution of renewal equation (1.2):

$$\lim_{n\to\infty} \sum_{k=0}^{n} b_k u_{n-k} = \frac{1}{\mu} \sum_{n=0}^{\infty} b_n.$$

  The importance of this theorem stems from the fact that quantities related to a RC verify associated renewal equations. After solving these equations (via Theorem 2.2), the key renewal theorem immediately provides the limit of those quantities.

In the same chapter, we also introduce delayed renewal chains, stationary renewal chains, and alternating renewal chains and we give some useful related results.

### 1.2.2 Semi-Markov Chains

In order to undertake the estimation problems on semi-Markov chains, it is important to investigate them first from a probabilistic point of view.

Generally speaking, the probabilistic study of semi-Markov chains follows approximatively the same steps as for renewal chains, but with a different degree of complexity. Obviously, this complexity comes from the structural difference between the renewal framework and the semi-Markov one. We pass from an i.i.d. context to a semi-Markov one, introducing the dependence on a randomly chosen state.

As was done for renewal chains, the first part of this study developed in Chapter 3 is centered around a type of equation similar to the renewal equation. For a discrete-time semi-Markov framework, the variables in such an equation are not real sequences, like for renewal chains, but matrix-valued

functions defined on $\mathbb{N}$ and denoted by $\mathbf{A} = (\mathbf{A}(k);\ k \in \mathbb{N})$, where $\mathbf{A}(k) = (A_{ij}(k);\ i,j \in E),\ k \in \mathbb{N}$, are matrices on $E \times E$. Thus we define the *discrete-time Markov renewal equation* (DTMRE or MRE) by

$$\mathbf{L}(k) = \mathbf{G}(k) + \mathbf{q} * \mathbf{L}(k),\ k \in \mathbb{N}, \tag{1.4}$$

where $\mathbf{q}$ is the semi-Markov kernel, $\mathbf{L} = (L_{ij}(k);\ i,j \in E, k \in \mathbb{N})$ is an unknown matrix-valued function, $\mathbf{G} = (G_{ij}(k);\ i,j \in E, k \in \mathbb{N})$ is a known matrix-valued function, and $*$ denotes the discrete-time matrix convolution product (Definition 3.5).

First, the solution of such an equation is given in Theorem 3.1 by

$$\mathbf{L}(k) = \left( \sum_{n=0}^{k} \mathbf{q}^{(n)}(k) \right) * \mathbf{G}(k), \tag{1.5}$$

where $\mathbf{q}^{(n)}$ represents the $n$-fold convolution of the semi-Markov kernel $\mathbf{q}$. We see here what we already stated above: considering a quantity of interest of a discrete-time semi-Markov process, after solving the corresponding Markov renewal equation, we are able to express it as a series involving only a finite number of terms.

Second, let us denote by $\boldsymbol{\psi}$ this finite series of matrix convolution powers, $\boldsymbol{\psi}(k) := \sum_{n=0}^{k} \mathbf{q}^{(n)}(k)$. As will be seen in Chapter 3, the $(i,j)$ element of $\mathbf{q}^{(n)}(k)$ has a simple probabilistic meaning: $q_{ij}^{(n)}(k)$ is the probability that, starting from state $i$ at time zero, the semi-Markov chain will do the $n$th jump at time $k$ to state $j$. Thus $\psi_{ij}(k)$ represents the probability that, starting from state $i$ at time zero, the semi-Markov chain will do a jump at time $k$ to state $j$. Again, recalling that in a renewal context we denoted by $u_n$ the probability that a renewal will occur at time $n$, we see that we have a perfect correspondence between the solution of a renewal equation ($g_n = (u * b)_n$)) and the solution of a Markov renewal equation.

As an application of the Markov renewal theory, we obtain the explicit form of an important quantity of a semi-Markov chain, the semi-Markov transition function (matrix) $\mathbf{P} = (P_{ij}(k);\ i,j \in E, k \in \mathbb{N})$, defined by

$$P_{ij}(k) := \mathbb{P}(Z_k = j \mid Z_0 = i),\ i,j \in E,\ k \in \mathbb{N}.$$

First, we obtain its associated Markov renewal equation (Proposition 3.2)

$$\mathbf{P} = \mathbf{I} - \mathbf{H} + \mathbf{q} * \mathbf{P},$$

where:

$\mathbf{I} := (\mathbf{I}(k);\ k \in \mathbb{N})$, with $\mathbf{I}(k)$ the identity matrix for any $k \in \mathbb{N}$;
$\mathbf{H} := (\mathbf{H}(k);\ k \in \mathbb{N})$, $\mathbf{H}(k) := diag(H_i(k);\ i \in E)$, with $H_i(k)$ the cumulative sojourn time distribution in state $i$ of the SMC (Definition 3.4).

Second, solving this Markov renewal equation we obtain the explicit expression of the semi-Markov transition function

$$\mathbf{P}(k) = \boldsymbol{\psi} * (\mathbf{I} - \mathbf{H})(k),\ k \in \mathbb{N}. \tag{1.6}$$

The second part of the probabilistic study of semi-Markov chains presented in Chapter 3 consists in limit results.

- We first have the analogy of the renewal theorem, called *Markov renewal theorem* (Theorem 3.2), which states that, under some conditions, for any states $i$ and $j$ we have
$$\lim_{k\to\infty} \psi_{ij}(k) = \frac{1}{\mu_{jj}},$$
where $\mu_{jj}$ is the mean recurrence time of state $j$ for the semi-Markov chain.
- Similarly, we have the *key Markov renewal theorem* (Theorem 3.3). Let us associate to each state $j \in E$ a real-valued function $v_j(n)$, defined on $\mathbb{N}$, with $\sum_{n\geq 0} \mid v_j(n) \mid < \infty$. Then, for any states $i$ and $j$ we have
$$\lim_{k\to\infty} \psi_{ij} * v_j(k) = \frac{1}{\mu_{jj}} \sum_{n\geq 0} v_j(n).$$
- As an application of the key Markov renewal theorem, we obtain the limit distribution of a semi-Markov chain (Proposition 3.9). Under some mild conditions, for any states $i$ and $j$ we have
$$\lim_{k\to\infty} P_{ij}(k) = \frac{\nu(j)m_j}{\sum_{l\in E} \nu(l)m_l},$$
where $(\nu(1), \ldots, \nu(s))$ is the stationary distribution of the EMC and $m_j$ is the mean sojourn time of the SMC in state $j$.
- For each $M \in \mathbb{N}$, we consider the following functional of the semi-Markov chain
$$W_f(M) := \sum_{n=1}^{N(M)} f(J_{n-1}, J_n, X_n),$$
where $f$ is a real measurable function defined on $E \times E \times \mathbb{N}^*$. For the functional $W_f(M)$ we have the SLLN and the CLT (Theorems 3.4 and 3.5, respectively). This CLT will be extensively used in Chapters 4 and 5 for proving the asymptotic normality of different estimators.

In Section 3.5 we give a Monte Carlo algorithm for obtaining a trajectory of a given SMC in the time interval $[0, M]$. Another Monte Carlo algorithm is proposed in Exercise 3.6.

### 1.2.3 Semi-Markov Chain Estimation

Let us consider a sample path of a semi-Markov chain censored at fixed arbitrary time $M \in \mathbb{N}^*$, that is, a sequence of successively visited states and sojourn times

$$\mathcal{H}(M) := (J_0, X_1, \ldots, J_{N(M)-1}, X_{N(M)}, J_{N(M)}, u_M), \tag{1.7}$$

where $u_M := M - S_{N(M)}$ represents the censored sojourn time in the last visited state $J_{N(M)}$ (recall that $N(M)$ is the discrete-time counting process of the number of jumps in $[1, M]$).

Our objective was, first, to obtain estimators for any quantity of a semi-Markov model and, second, to investigate the asymptotic properties of these estimators.

The likelihood function corresponding to the sample path $\mathcal{H}(M)$ is

$$L(M) = \alpha_{J_0} \prod_{k=1}^{N(M)} p_{J_{k-1}J_k} f_{J_{k-1}J_k}(X_k) \overline{H}_{J_{N(M)}}(u_M),$$

where $\overline{H_i}(\cdot) := \mathbb{P}(X_1 > \cdot \mid J_0 = i)$ is the survival function in state $i$ and $\alpha_i$ is the initial distribution of state $i$.

By Lemma 4.1 we will show that $u_M/M \xrightarrow[M\to\infty]{a.s.} 0$. Consequently, the term $\overline{H}_{J_{N(M)}}(u_M)$ corresponding to $u_M$ has no contribution to the likelihood when $M$ tends to infinity, and for this reason it can be neglected. On the other side, the information on $\alpha_{J_0}$ contained in the sample path $\mathcal{H}(M)$ does not increase with $M$, because $\mathcal{H}(M)$ contains only one observation of the initial distribution $\boldsymbol{\alpha}$ of $(J_n)_{n\in\mathbb{N}}$. As we are interested in large-sample theory of semi-Markov chains, the term $\alpha_{J_0}$ will be also neglected in the expression of the likelihood function. For this reasons, instead of maximizing $L(M)$ we will maximize the *approached likelihood function* defined by

$$L_1(M) = \prod_{k=1}^{N(M)} p_{J_{k-1}J_k} f_{J_{k-1}J_k}(X_k)$$

and we will call the obtained estimators "approached maximum-likelihood estimators."

Starting from a semi-Markov sample path $\mathcal{H}(M)$, we define for any states $i, j$, and positive integer $k$, $1 \le k \le M$:

- $N_i(M)$: the number of visits to state $i$ of the EMC, up to time $M$;
- $N_{ij}(M)$: the number of transitions of the EMC from $i$ to $j$, up to time $M$;
- $N_{ij}(k, M)$: the number of transitions of the EMC from $i$ to $j$, up to time $M$, with sojourn time in state $i$ equal to $k$.

For any states $i$ and $j$, the estimators of the $(i, j)$ element of the transition matrix $\mathbf{p}$, of the conditional sojourn time distributions $\mathbf{f}$, and of the discrete-time semi-Markov kernel $\mathbf{q}$ are given by (Proposition 4.1 and Relations (4.1)–(4.3))

$$\begin{aligned}\widehat{p}_{ij}(M) &= N_{ij}(M)/N_i(M),\\ \widehat{f}_{ij}(k, M) &= N_{ij}(k, M)/N_{ij}(M),\\ \widehat{q}_{ij}(k, M) &= N_{ij}(k, M)/N_i(M).\end{aligned}$$

Once the estimator of the kernel is obtained, any quantity of the semi-Markov chain can be estimated, after being expressed as a function of the semi-Markov kernel.

The second step in the semi-Markov estimation is to derive the asymptotic properties of the proposed estimators.

First, we need asymptotic results on $N_i(M), N_{ij}(M)$ and $N(M)$, as $M$ tends to infinity. Using the fact that $(J_n)_{n\in\mathbb{N}}$ is a Markov chain, some results are directly obtained from Markov chain theory (Convergences (3.34) and (3.35)), whereas others are specific to the semi-Markov context (Convergences (3.36)–(3.38)).

Second, we look at the strong convergence of the estimators $\widehat{p}_{ij}(M)$, $\widehat{f}_{ij}(k, M)$, $\widehat{q}_{ij}(k, M)$, $\widehat{P}_{ij}(k, M)$, etc. Results of this type are given in Corollary 4.1, Proposition 4.2, and Theorems 4.1, 4.3, 4.4, and 4.6.

Third, we are interested in the asymptotic normality of the estimators. Two different proofs of this kind of result can be given. The first one is based on the CLT for Markov renewal chains (Theorem 3.5) and on Lemmas A.1–A.3. The second uses the Lindeberg–Lévy CLT for martingales (Theorem E.4). When proving the asymptotic normality of the kernel estimator (Theorem 4.2), we give complete proofs based on both methods. In the other cases (Theorems 4.5 and 4.7), we only present the main steps of the proof based on the martingale approach. Note equally that the asymptotic normality allows us also to construct asymptotic confidence intervals for the estimated quantities.

## 1.3 Reliability Theory of Discrete-Time Semi-Markov Systems

Much work has been conducted in recent decades on probabilistic and statistical methods in reliability. Most of the existing mathematical models for system reliability suppose that a system's evolution is in continuous time. As mentioned before, there are particular applications where it is natural to consider that the time is discrete. For example, we think of systems working on demand, like electric gadgets, in which case we are interested in the number of times the system functioned up to failure. Other examples are those

systems whose lifetimes are expressed as the number of cycles/days/months up to failure. For these kinds of problems we think that it is important to have discrete-time mathematical models of reliability. But even when the initial problem is in continuous time, we can discretize it and handle it in discrete time. Since in computer implementation we always need to discretize, discrete-time semi-Markov systems can be useful as schemes of discretization of continuous-time systems.

Among examples of existing works on discrete-time reliability, we cite Nakagawa and Osaki (1975), Roy and Gupta (1992), Xie et al. (2002), Bracquemond and Gaudoin (2003) (in a general i.i.d. context), Balakrishnan et al. (2001), Platis et al. (1998) (reliability modeling via homogeneous and nonhomogeneous Markov chains), Sadek and Limnios (2002) (reliability metric estimation of discrete-time Markov systems), Csenki (2002), Barbu et al. (2004) (reliability modeling via semi-Markov chains), and Barbu and Limnios (2006a,b) (reliability metric estimation of discrete-time semi-Markov systems).

In the work presented in Chapter 5 we consider a system $\mathcal{S}$ whose possible states during its evolution in time are $E = \{1, \ldots, s\}$. We denote by $U = \{1, \ldots, s_1\}$ the subset of working states of the system (the up states) and by $D = \{s_1 + 1, \ldots, s\}$ the subset of failure states (the down states), with $0 < s_1 < s$. We suppose that $E = U \cup D$ and $U \cap D = \emptyset$, $U \neq \emptyset$, $D \neq \emptyset$.

The first part of the chapter is concerned with the modeling of reliability for discrete-time semi-Markov systems. We obtain explicit forms for reliability indicators: reliability, availability, maintainability, failure rates, and mean hitting times (mean time to repair, mean time to failure, mean up time, mean down time, mean time between failures) (Propositions 5.1–5.6).

NB: Generally, there are two different ways of obtaining these explicit forms. The first one is the straight Markov renewal theoretic way: first we find the Markov renewal equation associated to the respective quantity; then we solve this equation (via Theorem 3.1) and get the desired result. The second way is a case-adapted one, constructed for each reliability indicator. As an example, see the two proofs of Proposition 5.3, where we use both methods in order to find the explicit form of reliability.

From a theoretical point of view, the method based on Markov renewal theory is more attractive due to its generality. But the direct way can be more intuitive and can tell us information that is kept "hidden" by the general method. An example justifying this assertion can be seen in the two different ways of obtaining the reliability (Proposition 5.3).

The second part of the chapter looks at reliability indicator estimation. Just like for the estimation of the SM model, we start with a semi-Markov path censored at fixed arbitrary time $M \in \mathbb{N}^*$, as defined in (1.7). As the reliability indicators have explicit formulas in terms of the basic quantities of

the semi-Markov chain, the estimators of these quantities obtained in Chapter 4 allow us to immediately derive plug-in estimators of reliability, availability, failure rates, and mean times. For instance, the reliability of a discrete-time semi-Markov system at time $k \in \mathbb{N}$ is given by

$$R(k) = \boldsymbol{\alpha}_1 \, \mathbf{P}_{11}(k) \, \mathbf{1}_{s_1},$$

where $\boldsymbol{\alpha}_1$ and $\mathbf{P}_{11}$ are (roughly speaking) partitions of the initial distribution vector $\boldsymbol{\alpha}$ and of the semi-Markov transition matrix $\mathbf{P}(k)$ according to the state partition $E = U \cup D$, and $\mathbf{1}_{s_1}$ denotes the $s_1$-column vector whose elements are all 1. Using this expression of reliability, we immediately have its estimator

$$\widehat{R}(k, M) = \boldsymbol{\alpha}_1 \cdot \widehat{\mathbf{P}}_{11}(k, M) \cdot \mathbf{1}_{s_1}.$$

We also investigate the consistency and the asymptotic normality of these estimators as the sample size $M$ tends to infinity (Theorems 5.1–5.4 and Corollary 5.1). The techniques are the same as those previously used for the asymptotic properties of the estimators of quantities associated to semi-Markov chains. The only difference is the greater complexity of computations. Anyway, due to the particularities of discrete-time semi-Markov systems (finite-series expressions), the numerical computations of complicated quantities are fast.

## 1.4 Hidden Semi-Markov Models

The basic idea of a hidden model is the following: we observe the evolution in time of a certain phenomenon (observed process), but we are interested in the evolution of another phenomenon, which we are not able to observe (hidden process). The two processes are related in the sense that the state occupied by the observed process depends on the state that the hidden process is in.

To get one of the most intuitive and general examples of a hidden model, one can think the observed process as a received signal and the hidden process as the emitted signal.

If we add to this elementary hidden model the assumption that the hidden process (denoted by $Z$) is a Markov process of a certain order, we obtain a hidden Markov model (HMM). The observed process (denoted by $Y$) is either a conditional Markov chain of order $k$ or a sequence of conditionally independent random variables. A schematic representation of this kind of model is given in Figure 1.3. The arrows in $Y$ or $Z$ lines denote the dependence on the past ($k$ arrows if Markov of order $k$).

Since being introduced by Baum and Petrie (1966), hidden Markov models have become very popular, being applied in a wide range of areas. The reason is that many types of concrete applications can be investigated using this kind of model. Some examples: $Y$ is the GPS position of a car, while $Z$ represents the real (and unknown) position of the car; in reliability and maintenance, $Y$

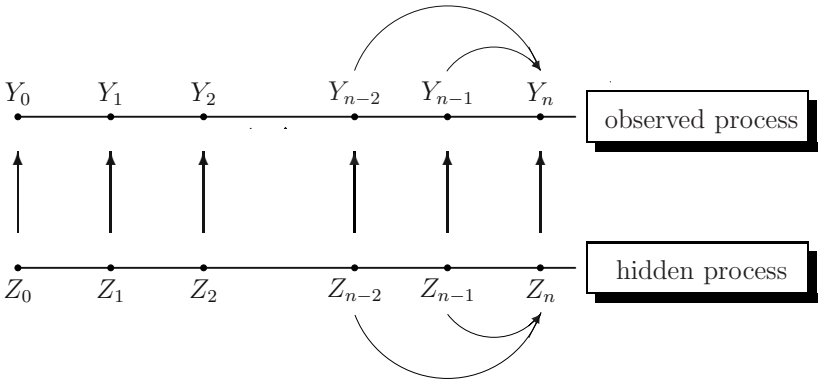


**Fig. 1.3.** A hidden (semi-)Markov model

can be any indicator of the state of an engine (temperature, pressure, noise), while $Z$ is the real state of the engine; $Y$ is an observed DNA sequence and $Z$ is an unknown chain related to a DNA coding mechanism (for example, $Z$ can be the sequence of indicators for CpG islands, as in Example 6.1).

The main drawback of hidden Markov models comes from the Markov property, which requires that the sojourn time in a state be geometrically distributed. This makes the hidden Markov models too restrictive from a practical point of view. In order to solve this kind of problem in the field of speech recognition, Ferguson (1980) proposed a model that allows arbitrary sojourn time distributions for the hidden process. This is called a hidden semi-Markov model, represented by a coupled process $(Z, Y) = (Z_n, Y_n)_{n\in\mathbb{N}}$, where $Z$ is a semi-Markov process and $Y$ is the observed process, depending on $Z$. Thus we have a model that combines the flexibility of semi-Markov processes with the modeling capacity of hidden-Markov models.

An important example of practical interest of hidden semi-Markov models (also known as explicit/variable state duration hidden Markov models), is GENSCAN, a program for gene identification (see Burge and Karlin, 1997) developed by Chris Burge at Stanford University. The underlying mathematical model is based on hidden semi-Markov chains.

Chapter 6 focuses on hidden semi-Markov model estimation. In the first part, we introduce two types of hidden semi-Markov models. First, we consider that the observed process $Y$, with finite state space $A := \{1, \ldots, d\}$, is a sequence of conditionally independent random variables, given a sample path of the hidden semi-Markov process $Z$, i.e., for all $a \in A, j \in E, n \in \mathbb{N}^*$, we have

$$\begin{aligned}&\mathbb{P}(Y_n = a \mid Y_{n-1} = \cdot, \ldots, Y_0 = \cdot, Z_n = i, Z_{n-1} = \cdot, \ldots, Z_0 = \cdot)\\ &= \mathbb{P}(Y_n = a \mid Z_n = i) =: R_{i;a}. \end{aligned} \tag{1.8}$$

The process $(Z, Y) = (Z_n, Y_n)_{n\in\mathbb{N}}$ is called a *hidden semi-Markov chain of type SM-M0*, where the index 0 stands for the order of $Y$ regarded as a conditional Markov chain.

Second, we suppose that $Y$ is a homogeneous Markov chain of order $k, k \geq 1$, conditioned on $Z$, i.e., Relation (6.3) holds. In this case, the process $(Z, Y)$ is called a *hidden semi-Markov chain of type SM-Mk*, where the index $k$ stands for the order of the conditional Markov chain $Y$.

The second part of the chapter is devoted to nonparametric maximum-likelihood estimation for a hidden semi-Markov chain $(Z, Y)$ of type SM-M0. We present the strong consistency and the asymptotic normality of the MLEs.

The basic idea is to associate a hidden Markov model to the initial model in order to be able to apply the classical estimation results of hidden Markov models. To this purpose, we consider $U = (U_n)_{n\in\mathbb{N}}$ the backward-recurrence times of the semi-Markov chain $(Z_n)_{n\in\mathbb{N}}$ defined by $U_n := n - S_{N(n)}$.

It is known that the couple $(Z, U)$ is a Markov chain (Proposition 6.1), so we have a hidden Markov chain $((Z, U), Y)$ associated to the initial hidden semi-Markov chain. The results are obtained under the assumption that all the conditional sojourn time distributions $f_{ij}(\cdot), i, j \in E$, have bounded support. From a practical point of view, this assumption does not represent an important drawback because in applications we always take into account only a finite support.

Then, applying the classical results of asymptotic properties of the MLEs for a hidden Markov model (Baum and Petrie, 1966; Bickel et al., 1998), we obtain the strong consistency of the model's estimators (Theorems 6.1–6.3) and the asymptotic normality (Theorems 6.4–6.6).

When performing maximum-likelihood estimation on a hidden (semi-) Markov model, we cannot obtain explicit expressions for the MLEs of the model parameter, $\boldsymbol{\theta} = ((q_{ij}(k))_{i,j;k}, (R_{i;a})_{i;a})$, since we do not observe the hidden process. For this reason, in the last part of Chapter 6 we propose an EM algorithm for obtaining the nonparametric MLEs of a hidden semi-Markov chain $(Z, Y)$ of type SM-M0 or SM-M1. We present here the main features of the SM-M0 case.

Our interest in developing an EM algorithm comes, first, from the fact that the existing literature on the EM algorithms for hidden semi-Markov models (Ferguson, 1980; Levinson, 1986; Guédon and Cocozza-Thivent, 1990; Sansom and Thomson, 2001; Guédon, 2003) considers, generally, parametric families for sojourn time distributions. Second, the semi-Markov chain we consider is general (in the sense that the kernel has the form $q_{ij}(k) = p_{ij} f_{ij}(k)$), while most of the hidden semi-Markov models in the literature assume particular

cases of semi-Markov processes (see Remark 3.3 for more details).

Let us consider a sample path of the hidden semi-Markov chain, censored at fixed arbitrary time $M \in \mathbb{N}^*$, $(Z_0^M, Y_0^M) := (Z_0, \ldots, Z_M, Y_0, \ldots, Y_M)$. Set also $\mathbf{y}_0^M := \{Y_0^M = y_0^M\}$. The likelihood function of the complete data for a hidden system SM-M0 is given by

$$f_M(Y_0^M, Z_0^M \mid \boldsymbol{\theta}) = \mu(J_0, Y_0) \prod_{k=1}^{N(M)} \left[ p_{J_{k-1}J_k} f_{J_{k-1}J_k}(X_k) \prod_{l=S_{k-1}}^{S_k - 1} R_{J_{k-1}, Y_l} \right]$$

$$\times \overline{H}_{J_{N(M)}}(U_M) \prod_{l=S_{N(M)}}^{M} R_{J_{N(M)}, Y_l},$$

where $\mu(i,a)$ is the initial distribution of the the state $(i,a) \in E \times A$. Let us also denote by $g_M(Y_0^M \mid \boldsymbol{\theta})$ the likelihood function of the incomplete data (the known data) $Y_0^M$,

$$g_M(Y_0^M \mid \boldsymbol{\theta}) = \sum_{(Z_0, \ldots, Z_M) \in E^{M+1}} f_M(Y_0^M, Z_0^M \mid \boldsymbol{\theta}).$$

The basic idea of the EM algorithm (cf. Dempster et al., 1977; Baum et al., 1970) is to start with a given value $\boldsymbol{\theta}^{(m)}$ of the parameter and to maximize with respect to $\boldsymbol{\theta}$ the expectation of the log-likelihood function of the complete data, conditioned by $\mathbf{y}_0^M$,

$$\mathbb{E}_{\boldsymbol{\theta}^{(m)}} \left[ \log(f_M(Y_0^M, Z_0^M \mid \boldsymbol{\theta})) \mid \mathbf{y}_0^M \right],$$

instead of maximizing the likelihood function of observed data $g_M(Y_0^M \mid \boldsymbol{\theta})$. The maximum, denoted by $\boldsymbol{\theta}^{(m+1)}$, is obtained by an iterative scheme, alternating two steps, called *expectation* and *maximization*.

So, we start with an initial value of the parameter, denoted by $\boldsymbol{\theta}^{(0)}$. After the first iteration of the algorithm we obtain an update $\boldsymbol{\theta}^{(1)}$ of the parameter and we continue recurrently up to the moment when a particular stopping condition is fulfilled. The sequence $\left(\boldsymbol{\theta}^{(m)}\right)_{m \in \mathbb{N}}$ of successively obtained parameters satisfies

$$g_M(Y_0^M \mid \boldsymbol{\theta}^{(m+1)}) \geq g_M(Y_0^M \mid \boldsymbol{\theta}^{(m)}),\ m \in \mathbb{N},$$

so the EM algorithm provides updates of the parameter that increase the likelihood function of the incomplete data.

Other algorithmic approaches, not discussed in this book, could be developed using Markov Chain Monte Carlo (MCMC) methods (see, e.g., Muri-Majoube, 1997, for the use of MCMC methods in a hidden Markov framework).