Cambridge University Press & Assessment
978-1-107-00757-4 — Markov Chains and Dependability Theory
Gerardo Rubino, Bruno Sericola
Excerpt
More Information

1 Introduction

1.1 Preliminary words

From the theoretical point of view, Markov chains are a fundamental class of stochastic processes. They are the most widely used tools for solving problems in a large number of domains. They allow the modeling of all kinds of systems and their analysis allows many aspects of those systems to be quantified. We find them in many subareas of operations research, engineering, computer science, networking, physics, chemistry, biology, economics, finance, and social sciences. The success of Markov chains is essentially due to the simplicity of their use, to the large set of theoretical associated results available, that is, the high degree of understanding of the dynamics of these stochastic processes, and to the power of the available algorithms for the numerical evaluation of a large number of associated metrics.

In simple terms, the Markov property means that given the present state of the process, its past and future are independent. In other words, knowing the present state of the stochastic process, no information about the past can be used to predict the future. This means that the number of parameters that must be taken into account to represent the evolution of a system modeled by such a process can be reduced considerably. Actually, many random systems can be represented by a Markov chain, and certainly most of the ones used in practice. The price to pay for imposing the Markov property on a random system consists of cleverly defining the present of the system or equivalently its state space. This can be done by adding a sufficient amount of information about the past of the system into the definition of the states. The theory of Markov models is extremely rich, and it is completed by a large set of numerical procedures that allow the analysis in practice of all sorts of associated problems.

Markov chains are at the heart of the tools used to analyze many types of systems from the point of view of their dependability, that is, of their ability to behave as specified when they were built, when faced with the failure of their components. The reason why a system will not behave as specified can be, for instance, some fault in its design, or the failure of some of its components when faced with unpredicted changes in the system's environment [3]. The area where this type of phenomenon is analyzed is globally called *dependability*. The two main associated keywords are *failures* and *repairs*. Failure is the transition from a state where the system behaves as specified to a state where this is not true anymore. Repair is the name of the opposite transition. Markov chains play a central role in the quantitative analysis of the behavior of a system that faces failure occurrences and possibly the repair of failed components, or at least of part of them. This book develops a selected set of topics where different aspects of these mathematical objects are analyzed, having in mind mainly applications in the dependability analysis of multicomponent systems.

In this chapter, we first introduce some important dependability metrics, which also allow us to illustrate in simple terms some of the concepts that are used later. At the same time, small examples serve not only to present basic dependability concepts but also some of the Markovian topics that we consider in this book. Then, we highlight the central pattern that can be traced throughout the book, the fact that in almost all chapters, some aspect of the behavior of the chains in subsets of their state spaces is considered, from many different viewpoints. We finish this Introduction with a description of the different chapters that compose the book, while commenting on their relationships.

1.2 Dependability and performability models

In this section we introduce the main dependability metrics and their extensions to the concept of performability. At the same time, we use small Markov models that allow us to illustrate the type of problems this book is concerned with. This section also serves as an elementary refresher or training in Markov analysis techniques.

1.2.1 Basic dependability metrics

Let us start with a single-component system, that is, a system for which the analyst has no structural data, and let us assume that the system can not be repaired. At time 0, the system works, and at some random time T, the system's *lifetime*, a failure occurs and the system becomes forever failed. We obviously assume that T is finite and that it has a finite mean. The two most basic metrics defined in this context are the Mean Time To Failure, MTTF, which is the expectation of T, MTTF = $\mathbb{E}\{T\}$, and the *reliability at time t*, $R(t)$, defined by

$$R(t) = \mathbb{P}\{T > t\},$$

that is, the tail of the distribution of the random variable, T. Observe that we have

$$\mathbb{E}\{T\} = \text{MTTF} = \int_0^\infty R(t)\,dt.$$

The simplest case from our Markovian point of view is when T is an exponentially distributed random variable with rate λ. We then have MTTF = $1/\lambda$ and $R(t) = e^{-\lambda t}$. Defining a stochastic process $X = \{X_t,\ t \in \mathbb{R}^+\}$ on the state space $S = \{1, 0\}$ as $X_t = 1$ when the system is working at time t, 0 otherwise, X is a continuous-time Markov chain whose dynamics is represented in Figure 1.1.

Let us assume now that the system (always seen as made of a single component) can be repaired. After a repair, it becomes operational again as it was at time 0. This behavior then cycles forever, alternating periods where the system works (called *operational* or

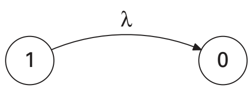

Figure 1.1 A single component with failure rate λ and $X_0 = 1$

up periods) and those where it is being repaired and thus does not provide any useful work (called *nonoperational* or *down* periods). Thus, after a first failure at some time F_1, the system becomes nonoperational until it is repaired at some time $R_1 \geq F_1$, then it works until the occurrence of a second failure at some time $F_2 \geq R_1$, etc. Let us call $U_1 = F_1$ the *length of the first up period*, $D_1 = R_1 - F_1$ the *length of the first down period*, $U_2 = F_2 - R_1$ *the length of the second up period*, etc. Let us consider now the main case for this framework, which occurs when the two sequences $(U_i)_{i\geq 1}$ and $(D_j)_{j\geq 1}$ are both i.i.d. and independent of each other (this is called an *alternating renewal process* in some contexts).

In this model, there is an infinite number of failures and repairs. By definition, the MTTF is the mean time until the first system's failure:

$$\text{MTTF} = \mathbb{E}\{U_1\},$$

and

$$R(t) = \mathbb{P}\{U_1 > t\}.$$

We may now consider other relevant metrics. First, the Mean Time To Repair, MTTR, is given by

$$\text{MTTR} = \mathbb{E}\{D_1\},$$

and the Mean Time Between Failures, MTBF, is given by MTBF = MTTF + MTTR. The reliability at time t measures the *continuity* of the service associated with the system, but one may also need to know if the system will be operational *at time t*. We define the *point availability at time t*, $PAV(t)$, as the probability that the system will be working at t.

Assume now that the U_i are exponentially distributed with rate λ and that the D_j are also exponentially distributed with rate μ. We then have MTTF = $1/\lambda$, MTTR = $1/\mu$, and $R(t) = e^{-\lambda t}$. If we define a stochastic process $X = \{X_t,\ t \in \mathbb{R}^+\}$ such that $X_t = 1$ if the system works at time t, and $X_t = 0$ otherwise, X is the continuous-time Markov chain whose associated graph is depicted in Figure 1.2. Let us denote $p_i(t) = \mathbb{P}\{X_t = i\}$, $i = 1, 0$. In other words, $(p_1(t), p_0(t))$ is the distribution of the random variable X_t, seen as a row vector (a convention that is followed throughout the book). Solving the Chapman–Kolmogorov differential equations in the $p_i(t)$ and adding the initial condition $X_0 = 1$, we get

$$PAV(t) = \mathbb{P}\{X_t = 1\} = p_1(t) = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu}e^{-(\lambda+\mu)t}.$$

This example allows us to introduce the widely used *asymptotic availability* of the system, which we denote here by $PAV(\infty)$, defined as $PAV(\infty) = \lim_{t\to\infty} PAV(t)$. Taking

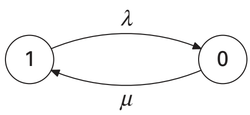

Figure 1.2 A single component with failure rate λ and repair rate μ; $X_0 = 1$

the limit in $p_1(t)$, we get $PAV(\infty) = \mu/(\lambda+\mu)$. Of course, if $\pi = (\pi_1, \pi_0)$ is the stationary distribution of X, we have $PAV(\infty) = \pi_1$. The stationary distribution, π, can be computed by solving the linear system of equilibrium equations of the chain: $\pi_1\lambda = \pi_0\mu$, $\pi_1 + \pi_0 = 1$.

1.2.2 More complex metrics

Let us now illustrate the fact that things can become more complex when dealing with more sophisticated metrics. Suppose we are interested in the behavior of the system in the interval $[0,t]$, and that we want to focus on how much time the system works in that interval. This is captured by the *interval availability on the interval* $[0,t]$, $IA(t)$, defined by the fraction of that interval during which the system works. Formally,

$$IA(t) = \frac{1}{t}\int_0^t 1_{\{X_s=1\}} ds.$$

Observe that $IA(t)$ is itself a random variable. We can be interested just in its mean, the *expected interval availability on* $[0,t]$. In the case of the previous two-state example, it is given by

$$\mathbb{E}\{IA(t)\} = \frac{1}{t}\int_0^t PAV(s)\,ds = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{(\lambda+\mu)^2 t}\left(1 - e^{-(\lambda+\mu)t}\right).$$

If, at the other extreme, we want to evaluate the distribution of this random variable, things become more complex. First, see that $\mathbb{P}\{IA(t) = 1\} = e^{-\lambda t}$, that is, there is a *mass* at $t = 1$. Then, for instance in [5], building upon previous work by Takàks, it is proved that if $x < 1$,

$$\mathbb{P}\{IA(t) \le x\} = 1 - e^{-\lambda xt}\left(1 + \sqrt{\lambda\mu xt}\int_0^{(1-x)t} \frac{e^{-\mu y}}{\sqrt{y}} I_1(2\sqrt{\lambda\mu xty})\,dy\right), \tag{1.1}$$

where I_1 is the modified Bessel function of the first kind defined, for $z \ge 0$, by

$$I_1(z) = \sum_{j\ge 0}\left(\frac{z}{2}\right)^{2j+1}\frac{1}{j!(1+j)!}.$$

In the well-known book by Gnedenko *et al.* [39], the following expression is proposed:

$$\mathbb{P}\{IA(t) \le x\} = \sum_{n\ge 0} e^{-\mu(1-x)t}\frac{(\mu(1-x)t)^n}{n!}\sum_{k=n+1}^{\infty} e^{-\lambda xt}\frac{(\lambda xt)^k}{k!}. \tag{1.2}$$

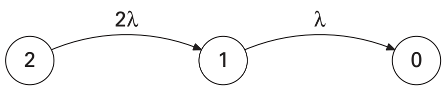

Figure 1.3 Two identical components in parallel with failure rate λ and no repair; $X_0 = 2$

Actually, there is an error in [39] on the starting index value of the embedded sum. The expression given here is the correct one.

In the book [85] by S. Ross, using the uniformization method (see Chapter 3 of this book if you are not familiar with this technique), the following expression is derived:

$$\mathbb{P}\{IA(t) \le x\} = \sum_{n \ge 1} e^{-\nu t} \frac{(\nu t)^n}{n!} \sum_{k=1}^{n} \binom{n}{k-1} p^{n-k+1} q^{k-1} \sum_{i=k}^{n} \binom{n}{i} x^i (1-x)^{n-i}, \qquad (1.3)$$

where $p = \lambda/(\lambda+\mu) = 1-q$ and $\nu = \lambda + \mu$. In Chapter 6 this approach is followed for the analysis of the interval availability metric in the general case. This discussion illustrates that even for elementary stochastic models (here, a simple two-state Markov chain), the evaluation of a dependability metric can involve some effort.

The previous model is irreducible. Let us look at simple systems modeled by absorbing chains. Consider a computer system composed of two identical processors working in parallel. Assume that the behavior of the processors, with respect to failures, is independent of each other, and that the lifetime of each processor is exponentially distributed, with rate λ. When one of the processors fails, the system continues to work with only one unit. When this unit fails, the system is dead, that is, failed forever. If X_t is the number of processors working at time t, then $X = \{X_t, t \in \mathbb{R}^+\}$ is a continuous-time Markov chain on the state space $S = \{2, 1, 0\}$, with the dynamics shown in Figure 1.3. The system is considered operational at time t if $X_t \ge 1$.

There is no repair here. The MTTF of the system, the mean time to go from the initial state 2 to state 0, is the sum of the mean time spent in state 2 plus the mean time spent in state 1, that is,

$$\text{MTTF} = \frac{1}{2\lambda} + \frac{1}{\lambda} = \frac{3}{2\lambda}.$$

To evaluate the reliability at time t, which is given by

$$R(t) = \mathbb{P}\{U_1 > t\} = \mathbb{P}\{X_t \ge 1\},$$

we need the transient distribution of the model, the distribution $p(t)$ of the random variable X_t, that is, the row vector

$$p(t) = (p_2(t),\ p_1(t),\ p_0(t)).$$

After solving the Chapman–Kolmogorov differential equations satisfied by vector $p(t)$, we have

$$p_2(t) = e^{-2\lambda t}, \quad p_1(t) = 2e^{-\lambda t}(1 - e^{-\lambda t}), \quad p_0(t) = 1 - 2e^{-\lambda t} + e^{-2\lambda t}.$$

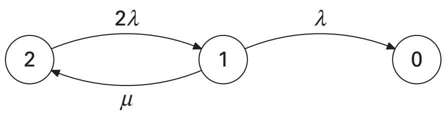

Figure 1.4 Two identical components in parallel, each with failure rate λ; a working component can repair a failed one, with repair rate μ; if both components are failed, the system is dead; $X_0 = 2$

We obtain

$$R(t) = p_2(t) + p_1(t) = 1 - p_0(t) = 2e^{-\lambda t} - e^{-2\lambda t}.$$

Observe that in this particular case, there is another elementary method to obtain the reliability function. Since the lifetime T of the system is the sum of two independent and exponentially distributed random variables (the sojourn times of X in states 2 and 1), the convolution of the two density functions of these sojourn times gives the density function of T. Integrating the latter, we obtain the cumulative distribution function of T:

$$\mathbb{P}\{T \le t\} = 1 - R(t) = \int_0^t \int_0^s 2\lambda e^{-2\lambda x} \lambda e^{-\lambda(s-x)} dx\, ds.$$

Now, suppose that when a processor fails, the remaining operational one can repair the failed unit, while doing its normal work at the same time. The repair takes a random amount of time, exponentially distributed with rate μ, and it is independent of the processors' lifetimes. If, while repairing the failed unit, the working one fails, then the system is dead since there is no operational unit able to perform a repair. With the same definition of X_t, we obtain the continuous-time Markov chain depicted in Figure 1.4.

The best way to evaluate the MTTF = $\mathbb{E}\{T\}$ is to define the conditional expectations, $x_i = \mathbb{E}\{T \mid X_0 = i\}$ for $i = 2, 1$, and write the equations

$$x_2 = \frac{1}{2\lambda} + x_1, \qquad x_1 = \frac{1}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} x_2,$$

leading to

$$\text{MTTF} = x_2 = \frac{3\lambda + \mu}{2\lambda^2}.$$

For the reliability at time t, we must again solve for the distribution $p(t)$ of X_t. We obtain

$$p_2(t) = \frac{(\lambda - \mu + G)e^{-a_1 t} - (\lambda - \mu - G)e^{-a_2 t}}{2G},$$

$$p_1(t) = 2\lambda \frac{e^{-a_2 t} - e^{-a_1 t}}{G},$$

where

$$G = \sqrt{\lambda^2 + 6\lambda\mu + \mu^2},$$

and

$$a_1 = \frac{3\lambda + \mu + G}{2} > a_2 = \frac{3\lambda + \mu - G}{2} > 0.$$

This leads to

$$R(t) = \mathbb{P}\{U_1 > t\} = \mathbb{P}\{X_t \geq 1\} = p_2(t) + p_1(t) = \frac{a_1 e^{-a_2 t} - a_2 e^{-a_1 t}}{G}.$$

Let us include here a brief introduction to the concept of *quasi-stationary distribution*, used in Chapter 4. When the model is absorbing, as in the last example, the limiting distribution is useless: at ∞, the process will be in its absorbing state (with probability 1), meaning that $p(t) \to (0,0,1)$ as $t \to \infty$. But we can wonder if the conditional distribution of X_t knowing that the process is not absorbed at time t has a limit. When it exists, this limit is called the quasi-stationary distribution of process X. In the example, we have

$$\lim_{t\to\infty} \mathbb{P}\{X_t = 2 \,|\, X_t \neq 0\} = \lim_{t\to\infty} \frac{p_2(t)}{p_2(t) + p_1(t)} = \frac{G - \lambda + \mu}{3\lambda + \mu + G} = \frac{G - \lambda - \mu}{2\lambda},$$

and

$$\lim_{t\to\infty} \mathbb{P}\{X_t = 1 \,|\, X_t \neq 0\} = \lim_{t\to\infty} \frac{p_1(t)}{p_2(t) + p_1(t)} = \frac{4\lambda}{3\lambda + \mu + G} = \frac{3\lambda + \mu - G}{2\lambda}.$$

When the system is not repairable, the point availability and the reliability functions coincide ($PAV(t) = R(t)$ for all t). At the beginning of this chapter we used the elementary model given in Figure 1.2 where the system could be repaired, and thus $PAV(t) \neq R(t)$.

To conclude this section, let us consider the example given in Figure 1.5, and described in the figure's caption. Observe that the topology of the model (its Markovian graph) is the same as in the model of Figure 1.4 but the transition rates and the interpretation are different.

First of all, we have MTTF = $1/\lambda$. The mean time until the system is dead, the mean absorption time of the Markov chain, can be computed as follows. If W is the absorption time, and if we denote $w_i = \mathbb{E}\{W \mid W_0 = i\}$, $i = 1, 0$, we have

$$w_1 = \frac{1}{\lambda} + w_0, \qquad w_0 = \frac{1}{\mu} + c w_1,$$

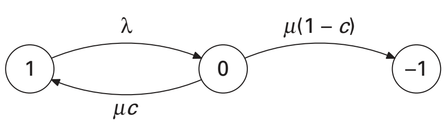

Figure 1.5 A single component with failure rate λ; there is a repair facility with repair rate μ; when being repaired, the system does not work; the repair can fail, and this happens with probability $1 - c$; the repair is successful with probability c, called the *coverage factor*, usually close to 1; if the repair fails, the system is dead; if the repair is successful, the system restarts as new

leading to

$$\mathbb{E}\{W\} = w_1 = \frac{1}{1-c}\left(\frac{1}{\lambda} + \frac{1}{\mu}\right).$$

The system can be repaired. This leads to $R(t) = e^{-\lambda t}$ and $PAV(t) = p_1(t)$. Solving the differential equations $p_1'(t) = -\lambda p_1(t) + \mu c p_0(t)$, $p_0'(t) = -\mu p_0(t) + \lambda p_1(t)$ with $p_1(0) = 1$, $p_0(0) = 0$, we obtain

$$PAV(t) = \frac{\mu - a_1}{H} e^{-a_1 t} + \frac{a_2 - \mu}{H} e^{-a_2 t},$$

where $H = \sqrt{(\lambda - \mu)^2 + 4(1-c)\lambda\mu}$, $a_1 = (\lambda + \mu - H)/2$ and $a_2 = (\lambda + \mu + H)/2$.

We are also interested in the total time, TO, during which the system is operational. Let us first compute its mean. Observing that the number of visits that the chain makes to state 1 (that is, the number of operational periods) is geometric, we have

$$\mathbb{E}\{TO\} = \sum_{n \geq 1} \frac{n}{\lambda}(1-c)c^{n-1} = \frac{1}{(1-c)\lambda}.$$

The distribution of TO is easy to derive using Laplace transforms. If $\widetilde{TO}$ denotes the Laplace transform of TO, we have

$$\widetilde{TO}(s) = \sum_{n \geq 1} \left(\frac{\lambda}{\lambda + s}\right)^n (1-c)c^{n-1} = \frac{(1-c)\lambda}{(1-c)\lambda + s},$$

that is, TO has the exponential distribution with rate $(1-c)\lambda$.

1.2.3 Performability

Consider the model of Figure 1.4 and assume that when the system works with only one processor, it generates r \$ per unit of time, while when there are two processors operational, the reward per unit of time is equal to αr, with $1 < \alpha < 2$. The reward is not equal to $2r$ because there is some capacity cost (some *overhead*) in being able to work with two parallel units at the same time. We can now look at the amount of money produced by the system until the end, T, of its lifetime. Let us call it R, and name it the *accumulated reward until absorption*. Looking for the expectation of R is easy. As for the evaluation of the MTTF, we use the conditional expectations, $y_i = \mathbb{E}\{R \mid X_0 = i\}$, $i = 2, 1$, which must satisfy

$$y_2 = \frac{\alpha r}{2\lambda} + y_1, \qquad y_1 = \frac{r}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} y_2.$$

This leads to

$$\mathbb{E}\{R\} = y_2 = r\frac{2\lambda + \alpha(\lambda + \mu)}{2\lambda^2}.$$

This is an example of a performability metric: instead of looking at the system as either working or not, we now distinguish between two different levels when it works, since it does not produce the same reward when it works with two units or with only one.

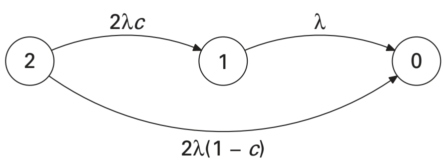

Figure 1.6 Two identical components in parallel; the failure rate is λ, the failure coverage factor is c; $X_0 = 2$

We can of course look for much more detailed information. To illustrate this and, at the same time, to provide more examples of Markovian models for the reader, we will use a variation of a previous example. We always have two identical processors working in parallel. When one of the two units fails, the system launches a procedure to try to continue working using the remaining one. This procedure works in general, but not always, and it takes very little time. We neglect this delay in the model, and we take into account the fact that the recovering procedure is not always successful by stating that when one of the units fails, the system continues to work with the other one with some fixed probability, c, and ends its life with probability $1 - c$. In the former case, when a new failure occurs, the remaining processor stops working and the whole system is dead. The parameter c is sometimes called the *coverage factor* in the dependability area, and its value is usually close to 1. Figure 1.6 shows the resulting absorbing model (to the previous assumptions we add the usual independence conditions on the events controlling the system's dynamics).

Let us first look at the previously considered metrics. If we just want to evaluate the MTTF of the system, we simply have

$$\text{MTTF} = \frac{1}{2\lambda} + c\frac{1}{\lambda} = \frac{1+2c}{2\lambda}.$$

The reader can check that this system behaves better than a single component from the point of view of the MTTF only if $c > 1/2$. The evaluation of the reliability at time t (or the point availability at time t: here, both metrics are identical) needs the transient distribution of the system. Solving the Chapman–Kolmogorov differential equations, we obtain

$$p_2(t) = e^{-2\lambda t}, \quad p_1(t) = 2ce^{-\lambda t}(1 - e^{-\lambda t}), \quad p_0(t) = 1 - 2ce^{-\lambda t} - (1-2c)e^{-2\lambda t}.$$

This gives

$$R(t) = PAV(t) = p_2(t) + p_1(t) = 2ce^{-\lambda t} + (1-2c)e^{-2\lambda t}.$$

Now, assume again that a reward of r \$ per unit of time is earned when the system works with one processor, and αr \$ per unit of time when it works with two processors, where $1 < \alpha < 2$. Denote the accumulated reward until absorption by R. The mean accumulated reward until absorption, $\mathbb{E}\{R\}$, is, using the same procedure as before,

$$\mathbb{E}\{R\} = \frac{\alpha r}{2\lambda} + c\frac{r}{\lambda} = r\frac{\alpha+2c}{2\lambda}.$$

But what about the distribution of the random variable R? In Chapter 7, we show why, for any $x \geq 0$ and for $r = 1$,

$$\mathbb{P}\{R > x\} = \frac{2c}{2-\alpha}e^{-\lambda x} + \left(1 - \frac{2c}{2-\alpha}\right)e^{-2\lambda x/\alpha}.$$

Obtaining this distribution is much more involved. For instance, observe that this expression does not hold if $\alpha = 2$. In Chapter 7, the complete analysis of this metric is developed, for this small example, and of course in the general case.

1.2.4 Some general definitions in dependability

As we have seen in the previous small examples, we consider systems modeled by Markov chains in continuous time, where we can distinguish two main cases: either the model is irreducible (as in Figure 1.2), or absorbing (as in Figures 1.1, 1.3, 1.4, and 1.6). In all cases, we have a partition $\{U, D\}$ of the state space S; U is the set of *up* states (also called *operational* states), where the system works, and D is the set of *down* states (also called *nonoperational* states), where the system is failed and does not provide the service it was built for. For instance, in Figure 1.1, $U = \{1\}$ and $D = \{0\}$; in Figure 1.6, $U = \{2, 1\}$ and $D = \{0\}$.

We always have $X_0 \in U$ and, if a is an absorbing state, $a \in D$. Otherwise, the possible interest of the model in dependability is marginal. As we have already stated, transitions from U to D are called failures, and transitions from D to U are called repairs. Observe that this refers to the global system. For instance, if we look at Figure 1.4, the transition from state 1 to state 2 corresponds to the repair of a *component*, not of the whole system. We always have at least one failure in the model; we may have models without repairs (as in Figure 1.1 or in Figure 1.3).

With the previous assumptions, we always have at least one first sojourn of X on the set of up states, U. As when we described the example of Figure 1.2, the lengths in time of the successive sojourns of X in U are denoted by U_1, U_2, etc.; these sojourns are also called operational (or up) periods. The corresponding unoperational (or down, or nonoperational) periods (if any), have lengths denoted by D_1, D_2, etc. A first remark here is that these sequences of random variables need not be independent and identically distributed anymore, as they were in the model of Figure 1.2, neither do they need to be independent of each other. The analysis of these types of variables is the object of the whole of Chapter 5 in this book. To see an example where these sequences are not independent and identically distributed, just consider the one in Figure 1.4 but assume now that we add another repair facility that is activated when both processors are failed, that is, in state 0. This means that we add a transition from 0 to 1 with some rate, η. The new model is given in Figure 1.7.

We can observe that the first sojourn in $U = \{2, 1\}$ starts in state 2, while the remaining sojourns in U start in state 1. It is easy to check that, in distribution, we have $U_1 \neq U_2 = U_3 = \cdots$ (see next section where some details are given). Here, just observe that the mean sojourn times have already been computed on page 6, when the model in