Roberto Bruni, Ugo Montanari

Models of Computation

– Monograph –

May 18, 2016

Springer

Mathematical reasoning may be regarded rather schematically as the exercise of a combination of two facilities, which we may call intuition and ingenuity.

Alan Turing[1]

[1] The purpose of ordinal logics (from Systems of Logic Based on Ordinals), Proceedings of the London Mathematical Society, series 2, vol. 45, 1939.

Preface

The origins of this book lie their roots on more than 15 years of teaching a course on formal semantics to graduate Computer Science to students in Pisa, originally called *Fondamenti dell'Informatica: Semantica* (*Foundations of Computer Science: Semantics*) and covering models for imperative, functional and concurrent programming. It later evolved to *Tecniche di Specifica e Dimostrazione* (*Techniques for Specifications and Proofs*) and finally to the currently running *Models of Computation*, where additional material on probabilistic models is included.

The objective of this book, as well as of the above courses, is to present different *models of computation* and their basic *programming paradigms*, together with their mathematical descriptions, both *concrete* and *abstract*. Each model is accompanied by some relevant formal techniques for reasoning on it and for proving some properties.

To this aim, we follow a rigorous approach to the definition of the *syntax*, the *typing* discipline and the *semantics* of the paradigms we present, i.e., the way in which well-formed programs are written, ill-typed programs are discarded and the way in which the meaning of well-typed programs is unambiguously defined, respectively. In doing so, we focus on basic proof techniques and do not address more advanced topics in detail, for which classical references to the literature are given instead.

After the introductory material (Part I), where we fix some notation and present some basic concepts such as term signatures, proof systems with axioms and inference rules, Horn clauses, unification and goal-driven derivations, the book is divided in four main parts (Parts II-V), according to the different styles of the models we consider:

IMP: imperative models, where we apply various incarnations of well-founded induction and introduce λ-notation and concepts like structural recursion, program equivalence, compositionality, completeness and correctness, and also complete partial orders, continuous functions, fixpoint theory;

HOFL: higher-order functional models, where we study the role of type systems, the main concepts from domain theory and the distinction between lazy and eager evaluation;

CCS, π: concurrent, non-deterministic and interactive models, where, starting from operational semantics based on labelled transition systems, we introduce the notions of bisimulation equivalences and observational congruences, and overview some approaches to name mobility, and temporal and modal logics system specifications;

PEPA: probabilistic/stochastic models, where we exploit the theory of Markov chains and of probabilistic reactive and generative systems to address quantitative analysis of, possibly concurrent, systems.

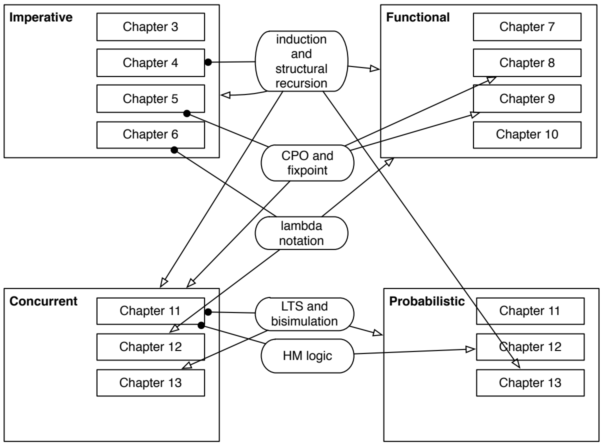
Each of the above models can be studied in separation from the others, but previous parts introduce a body of notions and techniques that are also applied and extended in later parts.

Parts I and II cover the essential, classic topics of a course on formal semantics.

Part III introduces some basic material on process algebraic models and temporal and modal logic for the specification and verification of concurrent and mobile systems. CCS is presented in good detail, while the theory of temporal and modal logic, as well as π-calculus, are just overviewed. The material in Part III can be used in conjunction with other textbooks, e.g., on model checking or π-calculus, in the context of a more advanced course on the formal modelling of distributed systems.

Part IV outlines the modelling of probabilistic and stochastic systems and their quantitative analysis with tools like PEPA. It poses the basis for a more advanced course on quantitative analysis of sequential and interleaving systems.

The diagram that highlights the main dependencies is represented below:



The diagram contains a squared box for each chapter / part and a rounded-corner box for each subject: a line with a filled-circle end joins a subject to the chapter where it is introduced, while a line with an arrow end links a subject to a chapter or part where it is used. In short:

Induction and recursion: various principles of induction and the concept of structural recursion are introduced in Chapter 4 and used extensively in all subsequent chapters.

CPO and fixpoint: the notion of complete partial order and fixpoint computation are first presented in Chapter 5. They provide the basis for defining the denotational semantics of IMP and HOFL. In the case of HOFL, a general theory of product and functional domains is also introduced (Chapter 8). The notion of fixpoint is also used to define a particular form of equivalence for concurrent and probabilistic systems, called bisimilarity, and to define the semantics of modal logic formulas.

Lambda-notation: λ-notation is a useful syntax for managing anonymous functions. It is introduced in Chapter 6 and used extensively in Part III.

LTS and bisimulation: Labelled transition systems are introduced in Chapter 11 to define the operational semantics of CCS in terms of the interactions performed. They are then extended to deal with name mobility in Chapter 13 and with probabilities in Part V. A bisimulation is a relation over the states of an LTS that is closed under the execution of transitions. The before mentioned bisimilarity is the coarsest bisimulation relation. Various forms of bisimulation are studied in Part IV and V.

HM-logic: Hennessy-Milner logic is the logic counterpart of bisimilarity: two state are bisimilar if and only if they satisfy the same set of HM-logic formulas. In the context of probabilistic system, the approach is extended to Larsen-Skou logic in Chapter 15.

Each chapter of the book is concluded by a list of exercises that span over the main techniques introduced in that chapter. Solutions to selected exercises are collected at the end of the book.

Pisa,
February 2016

Roberto Bruni
Ugo Montanari

Acknowledgements

We want to thank our friend and colleague Pierpaolo Degano for encouraging us to prepare this book and submit it to the EATCS monograph series. We thank Ronan Nugent and all the people at Springer for their editorial work. We acknowledge all the students of the course on *Models of Computation (MOD)* in Pisa for helping us to refine the presentation of the material in the book and to eliminate many typos and shortcomings from preliminary versions of this text. Last but not least, we thank Lorenzo Galeotti, Andrea Cimino, Lorenzo Muti, Gianmarco Saba, Marco Stronati, former students of the course on *Models of Computation*, who helped us with the LaTeX preparation of preliminary versions of this book, in the form of lecture notes.

Contents

Part I Preliminaries

Part II IMP: a simple imperative language

Part III HOFL: a higher-order functional language

Part IV Concurrent Systems

Part V Probabilistic Systems

Acronyms

$\sim$	operational equivalence in IMP (see Definition 3.3)
$\equiv_{den}$	denotational equivalence in HOFL (see Definition 10.4)
$\equiv_{op}$	operational equivalence in HOFL (see Definition 10.3)
$\simeq$	CCS strong bisimilarity (see Definition 11.5)
$\approx$	CCS weak bisimilarity (see Definition 11.16)
$\approxeq$	CCS weak observational congruence (see Section 11.8.2)
$\cong$	CCS dynamic bisimilarity (see Definition 11.18)
$\sim_E$	π-calculus strong early bisimilarity (see Definition 13.3)
$\sim_L$	π-calculus strong late bisimilarity (see Definition 13.4)
$\simeq_E$	π-calculus strong early full bisimilarity (see Section 13.5.3)
$\simeq_L$	π-calculus strong late full bisimilarity (see Section 13.5.3)
$\approx_E$	π-calculus weak early bisimilarity (see Section 13.5.4)
$\approx_L$	π-calculus weak late bisimilarity (see Section 13.5.4)
$\mathscr{A}$	interpretation function for the denotational semantics of IMP arithmetic expressions (see Section 6.2.1)
ack	Ackermann function (see Example 4.18)
Aexp	set of IMP arithmetic expressions (see Chapter 3)
$\mathscr{B}$	interpretation function for the denotational semantics of IMP boolean expressions (see Section 6.2.2)
Bexp	set of IMP boolean expressions (see Chapter 3)
$\mathbb{B}$	set of booleans
$\mathscr{C}$	interpretation function for the denotational semantics of IMP commands (see Section 6.2.3)
CCS	Calculus of Communicating Systems (see Chapter 11)
Com	set of IMP commands (see Chapter 3)
CPO	Complete Partial Order (see Definition 5.11)
$\mathrm{CPO}_\perp$	Complete Partial Order with bottom (see Definition 5.12)
CSP	Communicating Sequential Processes (see Section 16.2)
CTL	Computation Tree Logic (see Section 12.2.2)
CTMC	Continuous Time Markov Chain (see Definition 14.15)
DTMC	Discrete Time Markov Chain (see Definition 14.14)

Env	set of HOFL environments (see Chapter 9)
fix	(least) fixpoint (see Definition 5.2.2)
FIX	(greatest) fixpoint
gcd	greatest common divisor
HML	Hennessy-Milner modal Logic (see Section 11.6)
HM-Logic	Hennessy-Milner modal Logic (see Section 11.6)
HOFL	A Higher-Order Functional Language (see Chapter 7)
IMP	A simple IMPerative language (see Chapter 3)
int	integer type in HOFL (see Definition 7.2)
Loc	set of locations (see Chapter 3)
LTL	Linear Temporal Logic (see Section 12.2.1)
LTS	Labelled Transition System (see Definition 11.2)
lub	least upper bound (see Definition 5.7)
$\mathbb{N}$	set of natural numbers
$\mathscr{P}$	set of closed CCS processes (see Definition 11.1)
PEPA	Performance Evaluation Process Algebra (see Chapter 16)
Pf	set of partial functions on natural numbers (see Example 5.13)
PI	set of partial injective functions on natural numbers (see Problem 5.12)
PO	Partial Order (see Definition 5.1)
PTS	Probabilistic Transition System (see Section 14.4.2)
$\mathbb{R}$	set of real numbers
$\mathscr{T}$	set of HOFL types (see Definition 7.2)
Tf	set of total functions from $\mathbb{N}$ to $\mathbb{N}_\bot$ (see Example 5.14)
Var	set of HOFL variables (see Chapter 7)
$\mathbb{Z}$	set of integers

Part V
Probabilistic Systems

This part focuses on models and logics for probabilistic and stochastic systems. Chapter 14 presents the theory of random processes and Markov chains. Chapter 15 studies (reactive and generative) probabilistic models of computation with observable actions and sources of non-determinism together with a specification logic. Chapter 16 defines the syntax, operational and abstract semantics of PEPA, a well-known high-level language for the specification and analysis of stochastic, interactive systems.

Chapter 14
Measure Theory and Markov Chains

The future is independent of the past, given the present. (Markov property as folklore)

Abstract Future is largely unpredictable. Non-determinism accounts for modelling some phenomena arising in reactive systems, but it does not allow a quantitative estimation of how likely is one event w.r.t. another. We use the term *random* or *probability* to denote systems where the quantitative estimation is possible. In this chapter we present well-studied models of probabilistic systems, called *random processes* and *Markov chains* in particular. The second come in two flavours, depending on the underlying model of time (discrete or continuous). Their key feature is called *Markov property* and it allows to develop an elegant theoretical setting, where it can be conveniently estimated, e.g., how long a system will sojourn in a given state, or the probability of finding the system in a given state at a given time or in the long run. We conclude the chapter by discussing how bisimilarity equivalences can be extended to Markov chains.

14.1 Probabilistic and Stochastic Systems

In previous chapters we have exploited non-determinism to represent choices and parallelism. Probability can be viewed as a refinement of non-determinism, where it can be expressed that some choices are more likely or more frequent than others. We distinguish two main cases: *probabilistic* and *stochastic* models.

Probabilistic models associate a probability to each operation. If many operations are enabled at the same time, then the system uses the probability measure to choose the action that will be executed next. As we will see in Chapter 15, models with many different combinations of probability, non-determinism and observable actions have been studied.

In *stochastic* models each event has a duration. The model binds a random variable to each operation. This variable represents the time necessary to execute the operation. The models we will study use exponentially distributed variables, associating a rate to each event. Often in stochastic systems there is no explicit non-deterministic choice: when a race between events is enabled, the fastest operation is actually chosen.

We start this chapter by introducing some basic concepts of measure theory on which we will rely in order to construct probabilistic and stochastic models. Then we will present one of the most used stochastic models, called *Markov chains*. A Markov chain, named after the Russian mathematician Andrey Markov (1856–1922), is characterised by the the fact that the probability to evolve from one state to another depends only on the current state and not on the sequence of events that preceded it (e.g., it does not depend on the states traversed before reaching the current one). This feature, called the *Markov property*, essentially states that the system is memoryless, or rather that the relevant information about the past is entirely contained in the present state. A Markov chain allows to predict important statistical properties about the future behaviour of a system. We will discuss both the discrete time and the continuous time variants of Markov chains and we will examine some interesting properties which can be studied relying on probability theory.

14.2 Probability Space

A probability space accounts for modelling experiments with some degree of randomness. It comprises a set Ω of all possible outcomes (called *elementary events*) and a set $\mathscr{A}$ of *events* that we are interested in. An event is just a set of outcomes, i.e., $\mathscr{A} \subseteq \wp(\Omega)$, but in general we are not interested in the whole powerset $\wp(\Omega)$, especially because when Ω is infinite, then we would not be able to assign reasonable probabilities to all events in $\wp(\Omega)$. However, the set $\mathscr{A}$ should include at least the impossible event $\varnothing$ and the certain event Ω. Moreover, since events are sets, it is convenient to require that $\mathscr{A}$ is closed under the usual set operations. Thus if A and B are events, then also their intersection $A \cap B$, their union $A \cup B$ and complement $\overline{A}$ should be event, so that we can express, e.g., probabilities about the fact that two events will happen together, or about the fact that some event is not going to happen. If this is the case, then $\mathscr{A}$ is called a *field*. We call it a σ-field if it is also closed under countable union of events. A σ-field is indeed the starting point to define measurable spaces and hence probability spaces.

Definition 14.1 (σ-field). Let Ω be a set of elementary events and $\mathscr{A} \subseteq \wp(\Omega)$ be a family of subsets of Ω, then $\mathscr{A}$ is a σ-field if all of the following hold:

1. $\varnothing \in \mathscr{A}$ (the impossible event is in $\mathscr{A}$);
2. $\forall A \in \mathscr{A} \Rightarrow (\Omega \setminus A) \in \mathscr{A}$ ($\mathscr{A}$ is closed under complement);
3. $\forall \{A_n\}_{n\in\mathbb{N}} \subseteq \mathscr{A}.\ \bigcup_{i\in\mathbb{N}} A_i \in \mathscr{A}$ ($\mathscr{A}$ is closed under countable union).

The elements of $\mathscr{A}$ are called *events*.

Remark 14.1. It is immediate to see that $\mathscr{A}$ must include the certain event (i.e., $\Omega \in \mathscr{A}$, by 1 and 2) and that also the intersection of a countable sequence of elements of $\mathscr{A}$ is in $\mathscr{A}$, i.e., $\bigcap_{i\in\mathbb{N}} A_i = \Omega \setminus (\bigcup_{i\in\mathbb{N}} (\Omega \setminus A_i))$ (it follows by 2, 3 and the De Morgan property).

Let us illustrate the notion of σ-field by showing a simple example over a finite set of events.

Example 14.1. Let $\Omega = \{a, b, c, d\}$, we define a σ-field on Ω by setting $\mathscr{A} \subseteq \wp(\Omega)$:

$$\mathscr{A} = \{\varnothing, \{a, b\}, \{c, d\}, \{a, b, c, d\}\}$$

The smallest σ-field associated with a set Ω is $\{\varnothing, \Omega\}$ and the smallest σ-field that includes an event A is $\{\varnothing, A, \Omega \setminus A, \Omega\}$. More generally, given any subset $\mathscr{B} \subseteq \wp(\Omega)$ there is a least σ-field that contains $\mathscr{B}$.

σ-fields fix the domain on which we define a particular class of functions called *measures*, which assign a real number to each measurable set of the space. Roughly a measure can be seen as a notion of size that we wish to attach to sets.

Definition 14.2 (Measure). Let $(\Omega, \mathscr{A})$ be a σ-field. A function $\mu : \mathscr{A} \to [0, +\infty]$ is a *measure* on $(\Omega, \mathscr{A})$ if all of the following hold:

1. $\mu(\varnothing) = 0$;
2. for any countable collection $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathscr{A}$ of pairwise disjoint sets we have $\mu(\bigcup_{i\in\mathbb{N}} A_i) = \sum_{i\in\mathbb{N}} \mu(A_i)$.

A set contained in $\mathscr{A}$ is then called a *measurable set*, and the pair $(\Omega, \mathscr{A})$ is called *measurable space*. We are interested to a particular class of measures called *probabilities*. A probability is a essentially a “normalised” measure.

Definition 14.3 (Probability). A measure P on $(\Omega, \mathscr{A})$ is a *probability* if $P(\Omega) = 1$.

It is immediate from the definition of probability that the codomain of P cannot be the whole set $\mathbb{R}$ of real numbers but it is just the interval of reals $[0, 1]$.

Definition 14.4 (Probability space). Let $(\Omega, \mathscr{A})$ be a measurable space and P be a probability on $(\Omega, \mathscr{A})$, then $(\Omega, \mathscr{A}, P)$ is called a *probability space*.

14.2.1 Constructing a σ-field

Obviously one can think that in order to construct a σ-field that contains some sets equipped with a probability it is enough to construct the closure of these sets (together with top and bottom elements) under complement and countable union. But it comes out from set theory that not all sets are measurable. More precisely, it has been shown that it is not possible to define (in ZFC set theory) a probability for all the subsets of Ω when its cardinality is[1] $2^{\aleph_0}$ (i.e., there is no function $P : \wp(\mathbb{R}) \to [0, 1]$ that satisfies Definition 14.4). So we have to be careful in defining σ-field on a set Ω of elementary events that is uncountable.

The next example shows how this problem can be solved in a special case.

[1] The symbol $\aleph_0$, called *aleph zero*, is the smallest infinite cardinal, i.e., it denotes the cardinality of $\mathbb{N}$. Thus $2^{\aleph_0}$ is the cardinality of the powerset $\wp(\mathbb{N})$ as well as of the continuum $\mathbb{R}$.

Example 14.2 (Coin tosses). Let us consider the classic coin toss experiment. We have a fair coin and we want to model sequences of coin tosses. We would like to define Ω as the set of infinite sequences of head (H) and tail (T):

$$\Omega = \{H, T\}^{\infty}.$$

Unfortunately this set has cardinality $2^{\aleph_0}$. As we have just said a measure on uncountable sets does not exist. So we can restrict our attention to a countable set: the set $\mathscr{C}$ of finite sequences of coin tosses. In order to define a σ-field which can account for almost all the events that we could express in words, we define the following set for each $\alpha \in \mathscr{C}$ called the *shadow* of α:

$$[\alpha] = \{\ \omega \in \Omega \mid \exists \omega' \in \Omega.\ \alpha\omega' = \omega\ \}$$

The shadow of α is the set of infinite sequences of which α is a prefix. The right hand side of Figure 14.1 shows graphically the set $[\alpha]$ of infinite paths corresponding to the finite sequence α.

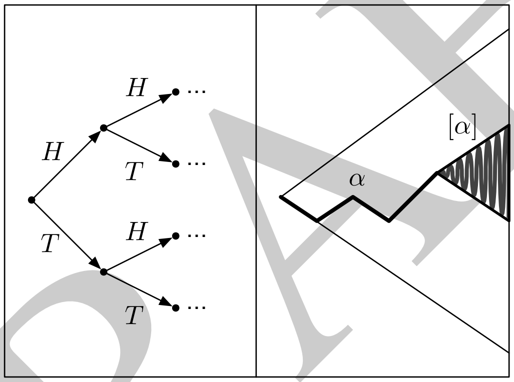


Fig. 14.1: The shadow of α

Now the σ-field which we were looking for is the one generated by the shadows of the sequences in $\mathscr{C}$. In this way we can start by defining a probability measure P on the σ-field generated by the shadows of $\mathscr{C}$, then we can assign a non-zero probability to (all finite sequences and) some infinite sequences of coin tosses by setting:

$$p(\omega) = \begin{cases} P(\,[\omega]\,) & \text{if } \omega \text{ is finite} \\ P\left(\bigcap\limits_{\alpha \in \mathscr{C},\ \omega \in [\alpha]} [\alpha] \right) & \text{if } \omega \text{ is infinite} \end{cases}$$

For the second case, remind that the definition of σ-field ensures that countable intersection of measurable sets is measurable. Measure theory results show that this measure exists and is unique.

Very often we have structures that are associated with a topology (e.g. there exists a standard topology, called Scott topology, associated to each CPO) so it is useful to define a standard method to obtain a σ-field from a topology.

Definition 14.5 (Topology). Let T be a set and $\mathscr{T} \subseteq \wp(T)$ be a family of subsets of T. Then $\mathscr{T}$ is said to be a *topology* on T if:

- $T, \varnothing \in \mathscr{T}$;
- $A, B \in \mathscr{T} \Rightarrow A \cap B \in \mathscr{T}$, i.e., the topology is closed under finite intersection;
- let $\{A_i\}_{i \in I}$ be any family of sets in $\mathscr{T}$ then $\bigcup_{i \in I} A_i \in \mathscr{T}$, i.e., the topology is closed under finite and infinite union.

The pair $(X, \mathscr{T})$ is said to be a *topological space*.

We call A an *open* set if it is in $\mathscr{T}$ and it is a *closed* set if $T \setminus A$ is open.

Remark 14.2. Note that in general a set can be open, closed, both or neither. For example, T and $\varnothing$ are open and also closed sets. Open sets should not be confused with measurable sets, because measurable sets are closed under complement and countable intersection. This difference makes the notion of measurable function very different from that of continuous function.

Definition 14.6 (Borel σ-field). Let $\mathscr{T}$ be a topology, we call the *Borel* σ-field of $\mathscr{T}$ the smallest σ-field that contains $\mathscr{T}$.

It turns out that the σ-field generated by the shadows which we have seen in the previous example is the Borel σ-field generated by the topology associated with the CPO of sets of infinite paths ordered by inclusion.

Example 14.3 (Euclidean topology). The *euclidean topology* is a topology on real numbers whose open sets are open intervals of real numbers:

$$]a,b[\quad = \quad \{x \in \mathbb{R} \mid a < x < b\}$$

We can extend the topology to the correspondent Borel σ-field, then associating to each open interval its length we obtain the usual *Lebesgue* measure.

It is often convenient to work with a generating collection, because Borel σ-fields are difficult to describe directly.

14.3 Continuous Random Variables

Stochastic processes associate a(n exponentially distributed) *random variable* to each event in order to represent its timing. So the concept of random variable and distribution will be central to the development in this chapter.

Suppose that an experiment has been performed and its outcome $\omega \in \Omega$ is known. A (continuous) random variable associates a real number to ω, e.g., by observing some of its features. For example, if ω is a finite sequence of coin tosses, a random variable X can count how many heads appear in ω. Then we can try to associate a probability measure on the possible values of X. However, it turns out that in general we cannot define a function $f : \mathbb{R} \to [0,1]$ such that $f(x)$ is the probability that X is x, because the set $\{\omega \mid X(\omega) = x\}$ is not necessarily an element of a measurable space. We consider instead (measurable) sets of the form $\{\omega \mid X(\omega) \leq x\}$.

Definition 14.7 (Random variable). Let $(\Omega, \mathscr{A}, P)$ be a probability space, a function $X : \Omega \to \mathbb{R}$ is said to be a *random variable* if

$$\forall x \in \mathbb{R}.\ \{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathscr{A}.$$

The condition expresses the fact that for each real number x, we can assign a probability to the set $\{\omega \in \Omega \mid X(\omega) \leq x\}$, because it is included in a measurable space. Notice that if we take as $(\Omega, \mathscr{A})$ the measurable space of the real numbers with the Lebesgue measure, the identity $id : \mathbb{R} \to \mathbb{R}$ satisfies the above condition. As another example, we can take sequences of coin tosses, assign the digit 0 to head and 1 to tail and see the sequences as binary representations of decimals in $[0,1)$.

Random variables can be classified by considering the set of their values. We call *discrete* a random variable that has a numerable or finite set of possible values. We say that a random variable is *continuous* if the set of its values is continuous. In the remainder of this section we will consider mainly continuous variables.

A random variable is completely characterised by its *probability law* which describes the probability that the variable will be found in a value less than or equal to the parameter.

Definition 14.8 (Cumulative distribution function). Let $S = (\Omega, \mathscr{A}, P)$ be a probability space, $X : \Omega \to \mathbb{R}$ be a continuous random variable over S. We call *cumulative distribution function* (also *probability law*) of X the image of P through X and denote it by $F_X : \mathbb{R} \to [0,1]$, i.e.:

$$F_X(x) \stackrel{\text{def}}{=} P(\{\omega \in \Omega \mid X(\omega) \leq x\}).$$

Note that the definition of random variable guarantees that, for any $x \in \mathbb{R}$, the set $\{\omega \in \Omega \mid X(\omega) \leq x\}$ is assigned a probability. Moreover, if $x < y$ then $F_X(x) \leq F_X(y)$.

As a matter of notation, we write $P(X \leq a)$ to mean $F_X(a)$, from which we derive:

$$P(X > a) \stackrel{\text{def}}{=} P(\{\omega \in \Omega \mid X(\omega) > a\}) = 1 - F_X(a)$$
$$P(a < X \leq b) \stackrel{\text{def}}{=} P(\{\omega \in \Omega \mid a < X(\omega) \leq b\}) = F_X(b) - F_X(a).$$

The other important function which describes the relative probability of a continuous random variable to take a specified value is the *probability density*.

Definition 14.9 (Probability density). Let $X : \Omega \to \mathbb{R}$ be a continuous random variable on the probability space $(\Omega, \mathscr{A}, P)$. We call the integrable function $f_X : \mathbb{R} \to [0, \infty)$ the *probability density* of X if:

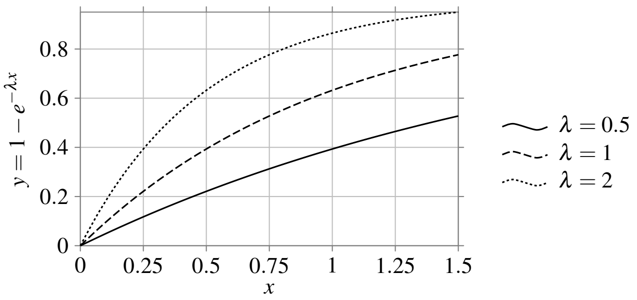


Fig. 14.2: Exponential probability laws with different rates λ

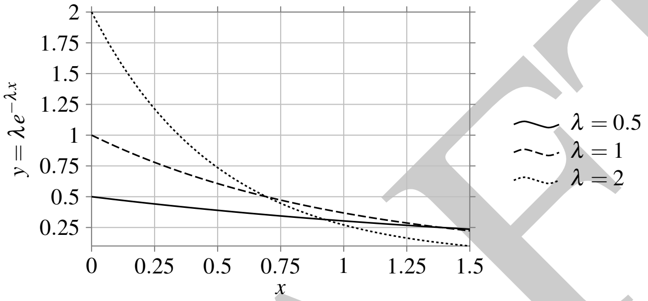


Fig. 14.3: Exponential density distributions with different rates λ

$$\forall a, b \in \mathbb{R}.\ P(a < X \leq b) = \int_a^b f_X(x)dx$$

So we can define the probability law F_X of a variable X with density f_X as follows:

$$F_X(a) = \int_{-\infty}^a f_X(x)dx$$

Note that $P(X = a) \overset{\text{def}}{=} P(\{\omega \mid X(\omega) = a\})$ is usually 0 when continuous random variables are considered. In case X is a discrete random variable, then its distribution function has jump discontinuities and the function $f_X : \mathbb{R} \to [0,1]$ given by $f_X(x) \overset{\text{def}}{=} P(X = x)$ is called *probability mass function*.

We are particularly interested in exponentially distributed random variables.

Definition 14.10 (Exponential distribution). A continuous random variable X is said to be *exponentially distributed* with parameter λ if its probability law and density function are defined as follows:

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x \geqslant 0 \\ 0 & x < 0 \end{cases} \qquad f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geqslant 0 \\ 0 & x < 0 \end{cases}$$

The parameter λ is called the *rate* of X and it characterises the expected value (mean) of X, which is $1/\lambda$, and the variance of X, which is $1/\lambda^2$. Some plottings of the functions F_X and f_X associated with exponential distributions with different rates are illustrated in Figure 14.2 and 14.3.

One of the most important features of exponentially distributed random variables is that they are memoryless, meaning that the current value of the random variable does not depend on the previous values.

Example 14.4 (Radioactive Atom). Let us consider a radioactive atom, which due to its instability can easily loose energy. It turns out that the probability that an atom will decay is constant over the time. So this system can be modelled by using an exponentially distributed, continuous random variable whose rate is the decay rate of the atom. Since the random variable is memoryless we have that the probability that the atom will decay at time $t_0 + t$ knowing that it is not decaying yet at time t_0 is the same for any choice of t_0, as it depends just on t.

In the following we denote by $P(A \mid B)$ the conditional probability of the event A given the event B, with

$$P(A \mid B) \overset{\text{def}}{=} \frac{P(A \cap B)}{P(B)}.$$

Theorem 14.1 (Memoryless). *Let X be an exponentially distributed (continuous) random variable with rate λ. Then:*

$$P(X \leq t_0 + t \mid X > t_0) = P(X \leq t).$$

Proof. Since X is exponentially distributed, its probability law is:

$$F_X(t) = \int_0^t \lambda e^{-\lambda x} dx$$

so we need to prove:

$$\frac{P(t_0 < X \leq t_0 + t)}{P(X > t_0)} = \frac{\int_{t_0}^{t_0+t} \lambda e^{-\lambda x} dx}{\int_{t_0}^{\infty} \lambda e^{-\lambda x} dx} \overset{?}{=} \int_0^t \lambda e^{-\lambda x} dx = P(X \leq t)$$

Since $\int_a^b \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_a^b = \left[e^{-\lambda x}\right]_b^a$ it follows that:

$$\frac{\int_{t_0}^{t_0+t} \lambda e^{-\lambda x} dx}{\int_{t_0}^{\infty} \lambda e^{-\lambda x} dx} = \frac{\left[e^{-\lambda x}\right]_{t_0+t}^{t_0}}{\left[e^{-\lambda x}\right]_{\infty}^{t_0}} = \frac{e^{-\lambda t_0} - e^{-\lambda t} \cdot e^{-\lambda t_0}}{e^{-\lambda t_0}} = \frac{\cancel{e^{-\lambda t_0}}(1 - e^{-\lambda t})}{\cancel{e^{-\lambda t_0}}} = 1 - e^{-\lambda t}$$

We conclude by:

$$\int_0^t \lambda e^{-\lambda x} dx = \left[e^{-\lambda x}\right]_t^0 = 1 - e^{-\lambda t}.$$

□

Another interesting feature of exponentially distributed random variables is the easy way in which we can compose information in order to find the probability of more complex events. For example if we have two random variables X_1 and X_2 which represent the delay of two events e_1 and e_2, we can try to calculate the probability that either of the two events will be executed before a specified time t. As we will see it happens that we can define an exponentially distributed random variable whose cumulative probability is the probability that either e_1 or e_2 executes before a specified time t.

Theorem 14.2. *Let X_1 and X_2 be two exponentially distributed continuous random variables with rate respectively λ_1 and λ_2 then:*

$$P(\min\{X_1, X_2\} \le t) = 1 - e^{-(\lambda_1+\lambda_2)t}$$

Proof. We recall that for any two events (not necessarily disjoint) we have:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

and that for two independent events we have

$$P(A \cap B) = P(A) \times P(B).$$

Then:

$$\begin{aligned}
P(\min\{X_1, X_2\} \le t) &= P(X_1 \le t \vee X_2 \le t) \\
&= P(X_1 \le t) + P(X_2 \le t) - P(X_1 \le t \wedge X_2 \le t) \\
&= P(X_1 \le t) + P(X_2 \le t) - P(X_1 \le t) \times P(X_2 \le t) \\
&= (1 - e^{-\lambda_1 t}) + (1 - e^{-\lambda_2 t}) - (1 - e^{-\lambda_1 t})(1 - e^{-\lambda_2 t}) \\
&= 1 - e^{-\lambda_1 t} e^{-\lambda_2 t} \\
&= 1 - e^{-(\lambda_1+\lambda_2)t}
\end{aligned}$$

□

Thus $X = \min\{X_1, X_2\}$ is also an exponentially distributed random variable, whose rate is $\lambda_1 + \lambda_2$. We will exploit this property to define, e.g., the sojourn time in continuous time Markov chains (see Section 14.4.4).

A second important value that we can calculate is the probability that an event will be executed before another. This corresponds in our view to calculate the probability that X_1 will take a value smaller than the one taken by X_2, namely that the action associated with X_1 is chosen instead of the one associated with X_2.

Theorem 14.3. *Let X_1 and X_2 be two exponentially distributed, continuous random variables with rate respectively λ_1 and λ_2 then:*

$$P(X_1 < X_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

Proof. Imagine you are at some time t and neither of the two variables has fired. The probability that X_1 fires in the infinitesimal interval dt while X_2 fires in any successive instant is

$$\lambda_1 e^{-\lambda_1 t}\left(\int_t^\infty \lambda_2 e^{-\lambda_2 t_2} dt_2\right) dt$$

From which we derive:

$$\begin{aligned}
P(X_1 < X_2) &= \int_0^\infty \lambda_1 e^{-\lambda_1 t_1}\left(\int_{t_1}^\infty \lambda_2 e^{-\lambda_2 t_2} dt_2\right) dt_1 \\
&= \int_0^\infty \lambda_1 e^{-\lambda_1 t_1}\left[e^{-\lambda_2 t_2}\right]_\infty^{t_1} dt_1 \\
&= \int_0^\infty \lambda_1 e^{-\lambda_1 t_1} \cdot e^{-\lambda_2 t_1} dt_1 \\
&= \int_0^\infty \lambda_1 e^{-(\lambda_1+\lambda_2)t_1} dt_1 \\
&= \left[\frac{\lambda_1}{\lambda_1+\lambda_2} e^{-(\lambda_1+\lambda_2)t}\right]_\infty^0 \\
&= \frac{\lambda_1}{\lambda_1+\lambda_2}.
\end{aligned}$$

□

We will exploit this property when presenting the process algebra PEPA, in Chapter 16.

As a special case, when the rates of the two variables are equal, i.e., $\lambda_1 = \lambda_2$, then $P(X_1 < X_2) = 1/2$.

14.3.1 Stochastic Processes

Stochastic processes are a very powerful mathematical tool that allows us to describe and analyse a wide variety of systems.

Definition 14.11 (Stochastic process). Let $(\Omega, \mathscr{A}, P)$ be a probability space and T be a set, then a family $\{X_t\}_{t\in T}$ of random variables over Ω is said to be a *stochastic process*.

A stochastic process can be identified with a function $X : \Omega \times T \to \mathbb{R}$ such that:

$$\forall t \in T.\, X(\cdot,t) : \Omega \to \mathbb{R} \text{ is a random variable.}$$

Usually the values in $\mathbb{R}$ that each random variable can take are called *states* and the element of T are interpreted as times.

Obviously the set T strongly characterises the process. A process in which T is $\mathbb{N}$ or a subset of $\mathbb{N}$ is said to be a *discrete time* process; on the other hand if $T = \mathbb{R}$

(or $T = [0,\infty)$) then the process is a *continuous time* process. The same distinction is usually done on the value that each random variable can assume: if this set has a countable or finite cardinality then the process is *discrete*; otherwise it is *continuous*. We will focus only on discrete processes with both discrete and continuous time. When the set $S = \{x \mid \exists \omega \in \Omega, t \in T . X(\omega,t) = x\}$ of states is finite, with cardinality N, without loss of generality, we can assume that $S = \{1,2,...,N\}$ is just the set of the first N positive natural numbers and we read $X_t = i$ as "the stochastic process X is in the ith state at time t".

14.4 Markov Chains

Stochastic processes studied by classical probability theory often involve only independent variables, namely the outcomes of the process are totally independent from the past. *Markov chains* extend the classic theory by dealing with processes where each variable is influenced by the previous one. This means that in Markov processes the next outcome of the system is influenced only by the previous state. One could think to extend this theory in order to allow general dependencies between variables, but it turns out that it is very difficult to prove general results on processes with dependent variables. We are interested in Markov chains since they provide an expressive mathematical framework to represent and analyse important interleaving and sequential systems.

Definition 14.12 (Markov chain). Let $(\Omega, \mathscr{A}, P)$ be a probability space, T be a totally ordered set and $\{X_t\}_{t\in T}$ be a stochastic process. Then, $\{X_t\}_{t\in T}$ is said to be a *Markov chain* if for each sequence $t_0 < ... < t_n < t_{n+1}$ of times in T and for all states $x, x_0, x_1, ..., x_n \in \mathbb{R}$:

$$P(X_{t_{n+1}} = x \mid X_{t_n} = x_n, \ldots, X_{t_0} = x_0) = P(X_{t_{n+1}} = x \mid X_{t_n} = x_n).$$

The previous proposition is usually referred to as *Markov property*.

An important characteristic of a Markov chain is the way in which it is influenced by the time. We have two types of Markov chains, *inhomogeneous* and *homogeneous*. In the first case the state of the system depends on the time, namely the probability distribution changes over time. In homogeneous chains on the other hand the time does not influence the distribution, i.e., the transition probability does not change during the time. We will consider only the simpler case of homogeneous Markov chains, gaining the possibility to shift the time axis back and forward.

Definition 14.13 (Homogeneous Markov chain). Let $\{X_t\}_{t\in T}$ be a Markov chain; it is *homogeneous* if for all states $x, x' \in \mathbb{R}$ and for all times $t, t' \in T$ with $t < t'$ we have:

$$P(X_{t'} = x' | X_t = x) = P(X_{t'-t} = x' | X_0 = x).$$

In what follows we use the term "Markov chain" as a synonym for "homogeneous Markov chain".

14.4.1 Discrete and Continuous Time Markov Chain

As we said, one of the most important things about stochastic processes in general, and about Markov chains in particular, is the choice of the set of times. In this section we will introduce two kinds of Markov chains, those in which $T = \mathbb{N}$, called *discrete time Markov chain* (DTMC), and those in which $T = \mathbb{R}$, referred to as *continuous time Markov chain*.

Definition 14.14 (Discrete time Markov Chain (DTMC)). Let $\{X_t\}_{t\in\mathbb{N}}$ be a stochastic process; then, it is a *discrete time Markov chain* (DTMC) if for all $n \in \mathbb{N}$ and for all states $x, x_0, x_1, ..., x_n \in \mathbb{R}$:

$$P(X_{n+1} = x \mid X_n = x_n, \ldots, X_0 = x_0) = P(X_{n+1} = \mid X_n = x_n).$$

Since we are restricting our attention to homogeneous chains then we can reformulate the Markov property as follows:

$$P(X_{n+1} = x \mid X_n = x_n, \ldots, X_0 = x_0) = P(X_1 = x \mid X_0 = x_n)$$

Assuming the possible states are $1, ..., N$, the DTMC is entirely determined by the transition probabilities $a_{i,j} = P(X_1 = j \mid X_0 = i)$ for $i, j \in \{1, ..., N\}$.

Definition 14.15 (Continuous time Markov Chain (CTMC)). Let $\{X_t\}_{t\in\mathbb{R}}$ be a stochastic process; then, it is a *continuous time Markov chain* (CTMC) if for all states $x, x_0, ..., x_n$, for any $\Delta_t \in [0, \infty)$ and any sequence of times $t_0 < ... < t_n$ we have:

$$P(X_{t_n+\Delta_t} = x \mid X_{t_n} = x_n, \ldots, X_{t_0} = x_0) = P(X_{t_n+\Delta_t} = x \mid X_{t_n} = x_n).$$

As for the discrete case, the homogeneity allows to reformulate the Markov property as follows:

$$P(X_{t_n+\Delta_t} = x \mid X_{t_n} = x_n, \ldots, X_{t_0} = x_0) = P(X_{\Delta_t} = x \mid X_0 = x_n).$$

Assuming the possible states are $1, ..., N$, the CTMC is entirely determined by the rates $\lambda_{i,j}$ that govern the probability $P(X_1 = j \mid X_0 = i) = 1 - e^{-\lambda_{i,j} t}$.

We remark that the exponential random variable is the only continuous random variable with the memoryless property, i.e., CTMC are necessarily exponentially distributed.

14.4.2 DTMC as LTS

A DTMC can be viewed as a particular LTS whose labels are probabilities. Usually such LTS are called *probabilistic transition systems* (PTS).

A difference between LTS and PTS is that in LTS we can have structures like the one shown in Figure 14.4(a), with two transitions that are co-initial and co-final and carry different labels. In PTS we cannot have this kind of situation since two different transitions between the same pair of states have the same meaning of a single transition labeled with the sum of the probabilities, as shown in Figure 14.4(b).

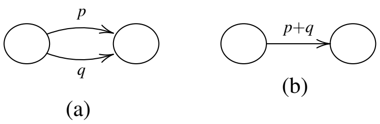


Fig. 14.4: Two equivalent DTMCs

The PTS (S,α) associated with a DTMC has a set of states S and a transition function $\alpha : S \rightarrow (D(S)+1)$ where $D(S)$ denotes the set of discrete probability distributions over S and $1 = \{*\}$ is a singleton used to represent the deadlock states. We recall that a discrete probability distribution over a set S is a function $D : S \rightarrow [0,1]$ such that $\sum_{s\in S} D(S) = 1$.

Definition 14.16 (PTS of a DTMC). Let $\{X_t\}_{t\in\mathbb{N}}$ be a DTMC whose set of states is S. Its corresponding PTS has set of states S and transition function $\alpha : S \rightarrow (D(S)+1)$ defined as follows:

$$\alpha(s) = \begin{cases} \lambda s'.\ P(X_1 = s' \mid X_0 = s) & \text{if } s \text{ is not a deadlock state} \\ * & \text{otherwise.} \end{cases}$$

Note that for each non-deadlock state s it holds:

$$\sum_{s'\in S} \alpha(s)(s') = 1.$$

Usually the transition function is represented through a matrix P whose indices i,j represent states s_i, s_j and each element $a_{i,j}$ is the probability that knowing that the system is in the state i it would be in the state j in the next time instant, namely $\forall i,j \leq |S|.\ a_{i,j} = \alpha(s_i)(s_j)$, note that in this case each row of P must sum to one. This representation allows us to study the system by relaying on linear algebra. In fact we can represent the present state of the system by using a row vector $\pi^{(t)} = [\pi_i^{(t)}]_{i\in S}$ where $\pi_i^{(t)}$ represents the probability that the system is in the state s_i at the time t. If we want to calculate how the system will evolve (i.e., the next state distribution) starting from this state we can simply multiply the vector with the matrix which represents the transition function, as the following example of a three state system shows:

$$\pi^{(t+1)} = \pi^{(t)}P = \begin{vmatrix} \pi_1^{(t)} & \pi_2^{(t)} & \pi_3^{(t)} \end{vmatrix} \begin{vmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{vmatrix} = \begin{vmatrix} a_{1,1}\pi_1^{(t)} + a_{2,1}\pi_2^{(t)} + a_{3,1}\pi_3^{(t)} \\ a_{1,2}\pi_1^{(t)} + a_{2,2}\pi_2^{(t)} + a_{3,2}\pi_3^{(t)} \\ a_{1,3}\pi_1^{(t)} + a_{2,3}\pi_2^{(t)} + a_{3,3}\pi_3^{(t)} \end{vmatrix}^{\mathsf{T}}$$

where the resulting row vector is transposed for space matter.

For some special class of DTMCs we can prove the existence of a limit vector for $t \to \infty$, that is to say the probability that the system is found in a particular state is stationary in the long run (see Section 14.4.3).

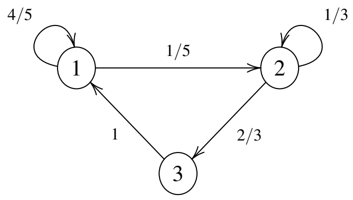


Fig. 14.5: A DTMC

Example 14.5 (DTMC). Let us consider the DTMC in Figure 14.5. We represent the chain algebraically by using the following matrix:

$$P = \begin{vmatrix} 4/5 & 1/5 & 0 \\ 0 & 1/3 & 2/3 \\ 1 & 0 & 0 \end{vmatrix}$$

Now suppose that we do not know the state of the system at time t, thus we assume the system has equal probability $\frac{1}{3}$ of being in any of the three states. We represent this situation with the following vector:

$$\pi^{(t)} = \begin{vmatrix} 1/3 & 1/3 & 1/3 \end{vmatrix}$$

Now we can calculate the state distribution at time $t+1$ as follows:

$$\pi^{(t+1)} = \begin{vmatrix} 1/3 & 1/3 & 1/3 \end{vmatrix} \begin{vmatrix} 4/5 & 1/5 & 0 \\ 0 & 1/3 & 2/3 \\ 1 & 0 & 0 \end{vmatrix} = \begin{vmatrix} 3/5 & 8/45 & 2/9 \end{vmatrix}$$

Notice that the sum of probabilities in the result $3/5 + 8/45 + 2/9$ is again 1. Obviously we can iterate this process in order to simulate the evolution of the system.

Since we have represented a Markov chain by using a transition system it is quite natural to ask for the probability of a finite path.

Definition 14.17 (Finite path probability). Let $\{X_t\}_{t\in\mathbb{N}}$ be a DTMC and $s_1 \cdots s_n$ a finite path of its PTS (i.e., $\forall i.\ 1 \le i < n \Rightarrow \alpha(s_i)(s_{i+1}) > 0$) we define the probability $P(s_1 \cdots s_n)$ of the path $s_1 \cdots s_n$ as follows:

$$P(s_1 \cdots s_n) = \prod_{i=1}^{n-1} \alpha(s_i)(s_{i+1}) = \prod_{i=1}^{n-1} a_{i,i+1}.$$

Example 14.6 (Finite paths). Let us consider the DTMC of Example 14.5 and take the path 1 2 3 1. We have:

$$P(1\ 2\ 3\ 1) = a_{1,2} \times a_{2,3} \times a_{3,1} = \frac{1}{5} \times \frac{2}{3} \times 1 = \frac{2}{15}$$

Note that if we consider the sequence of states 1 1 3 1:

$$P(1\ 1\ 3\ 1) = a_{1,1} \times a_{1,3} \times a_{3,1} = \frac{4}{5} \times 0 \times 1 = 0$$

In fact there is no transition allowed from state 1 to 3.

Note that it would make no sense to define the probability of infinite paths as the product of the probabilities of all choices, because any infinite sequence would have a null probability. We can overcome this problem by using the Borel σ-field generated by the shadows, as seen in Example 14.2.

14.4.3 DTMC Steady State Distribution

In this section we will present a special class of DTMCs which guarantees that the probability that the system is found in a state can be estimated on the long term. This means that the probability distribution of each state of the DTMC (i.e., the corresponding value in the vector $\pi^{(t)}$) reaches a *steady state distribution* which does not change in the future, namely if π_i is the steady state distribution for the state i, if $\pi_i^{(0)} = \pi_i$ then $\pi_i^{(t)} = \pi_i$ for each $t > 0$.

Definition 14.18 (Steady state distribution). We define the *steady state distribution* (or *stationary distribution*) $\pi = \left|\pi_1 \dots \pi_n\right|$ of a DTMC as the limit distribution:

$$\forall i \in [1,n].\ \pi_i = \lim_{t\to\infty} \pi_i^{(t)}$$

when such limit exists.

In order to guarantee that the limit exists we will restrict our attention to a subclass of Markov chains.

Definition 14.19 (Ergodic Markov chain). Let $\{X_t\}_{t\in T}$ be a Markov chain then it is said to be *ergodic* if it is both:

irreducible: each state is reachable from each other; and
aperiodic the greatest common divisor (gcd) of the lengths of all paths from any state to itself must be 1.

Theorem 14.4. *Let $\{X_t\}_{t\in T}$ be an ergodic (homogeneous) Markov chain. Then the steady state probability π always exists and it is independent from the initial state probability distribution.*

The steady state probability distribution π can be computed by solving the system of linear equations:

$$\pi = \pi\, P$$

where P is the matrix associated to the chain, under the additional constraint that the sum of all probabilities is 1.

Example 14.7 (Steady state distribution). Let us consider the DTMC of Example 14.5. It is immediate to check that it is ergodic. To find the steady state distribution we need to solve the following linear system:

$$\begin{vmatrix} \pi_1 & \pi_2 & \pi_3 \end{vmatrix} \begin{vmatrix} 4/5 & 1/5 & 0 \\ 0 & 1/3 & 2/3 \\ 1 & 0 & 0 \end{vmatrix} = \begin{vmatrix} \pi_1 & \pi_2 & \pi_3 \end{vmatrix}$$

The corresponding system of linear equations is

$$\begin{cases} \frac{4}{5}\pi_1 + \pi_3 = \pi_1 \\ \frac{1}{5}\pi_1 + \frac{1}{3}\pi_2 = \pi_2 \\ \frac{2}{3}\pi_2 = \pi_3 \end{cases}$$

Note that the equations express the fact that the probability to be in the state i is given by the sum of the probabilities to be in any other state j weighted by the probability to move from j to i. By solving the system of linear equations we obtain the solution:

$$\begin{vmatrix} 10\pi_2/3 & \pi_2 & 2\pi_2/3 \end{vmatrix}$$

i.e., $\pi_1 = \frac{10}{3}\pi_2$ and $\pi_3 = \frac{2}{3}\pi_2$.

Now by imposing $\pi_1 + \pi_2 + \pi_3 = 1$ we have $\pi_2 = 1/5$ thus:

$$\pi = \begin{vmatrix} 2/3 & 1/5 & 2/15 \end{vmatrix}$$

So, independently from the initial state, in the long run it is more likely to find the system in the state 1 than in states 2 or 3, because the steady state probability of being in state is much larger than the other two probabilities.

14.4.4 CTMC as LTS

Also continuous time Markov chains can be represented as LTSs, but in this case the labels are rates and not probabilities. We have two equivalent definitions for the transition function:

$$\alpha : S \to S \to \mathbb{R} \qquad \text{or} \qquad \alpha : (S \times S) \to \mathbb{R}$$

where S is the set of states of the chain and any real value $\lambda = \alpha(s)(s')$ (or $\lambda = \alpha(s_1, s_2)$) represents the rate which labels the transition $s \xrightarrow{\lambda} s'$. Also in this case, likewise DTMC, we have that two different transitions between the same two states are merged in a single transition whose label is the sum of the rates. We write $\lambda_{i,j}$ for the rate $\alpha(s_i, s_j)$ associated with the transition from state s_i to state s_j. A difference here is that the self loops can be ignored: this is due to the fact that in continuous time we allow the system to *sojourn* in a state for a period and staying in a state is indistinguishable from moving to the same state via a loop.

The probability that some transition happens from state s_i in some time t can be computed by taking the minimum of the continuous random variables associated with the possible transitions: by Theorem 14.2 we know that such probability is also exponentially distributed and has a rate that is given by the sum of rates of all the transitions outgoing from s_i.

Definition 14.20 (Sojourn time). Let $\{X_t\}$ a CTMC. The probability that no transition happens from a state s_i in some (sojourn) time t is 1 minus the probability that some transition happens:

$$\forall t \in (0, \infty).\ P(X_t = s_i \mid X_0 = s_i) = e^{-\lambda t} \text{ with } \lambda = \sum_{j \neq i} \lambda_{i,j}.$$

As for DTMCs we can represent a CTMC by using linear algebra. In this case the matrix Q which represents the system is defined by setting $q_{i,j} = \alpha(s_i, s_j) = \lambda_{i,j}$ when $i \neq j$ and $q_{i,i} = -\sum_{j \neq i} q_{i,j}$. This matrix is usually called *infinitesimal generator*. This definition is convenient for steady state analysis, as explained by the end of the next section.

14.4.5 Embedded DTMC of a CTMC

Often the study of a CTMC results very hard particularly in term of computational complexity. So it is useful to have a standard way to discretise the CTMC by synthesising a DTMC, called *embedded DTMC*, in order to simplify the analysis.

Definition 14.21 (Embedded DTMC). Let α_C be the transition function of a CTMC. Its *embedded DTMC* has the same set of states S and transition function α_D defined by taking:

$$\alpha_D(s_i)(s_j) = \begin{cases} \frac{\alpha_c(s_i, s_j)}{\sum_{s \neq s_i} \alpha_c(s_i, s)} & \text{if } s_i \neq s_j \\ 0 & \text{otherwise.} \end{cases}$$

As we can see, the previous definition simply normalises to 1 the rates in order to calculate a probability.

While the embedded DTMC completely determines the probabilistic behaviour of the system, it does not fully capture the behaviour of the continuous time process because it does not specify the rates at which transitions occur.

Regarding the steady state analysis, since in the infinitesimal generator matrix Q describing the CTMC we have $q_{i,i} = -\sum_{j \neq i} q_{i,j}$ for any state index i, the steady state distribution can equivalently be computed by solving the system of (homogeneous, normalised) linear equations $\pi\, Q = 0$ (see Problem 14.11).

14.4.6 CTMC Bisimilarity

Obviously, since Markov chains can be seen as a particular type of LTS, one could think to modify the notion of bisimilarity in order to study the equivalence between stochastic systems.

Let us start by revisiting the notion of LTS bisimilarity in a slightly different way from that seen in Chapter 11.

Definition 14.22 (Reachability predicate). Given and LTS $(S, L, \rightarrow)$, we define a function $\gamma : S \times L \times \wp(S) \rightarrow \{true, false\}$ which takes a state p, an action ℓ and a set of states I and returns *true* if there exists a state $q \in I$ reachable from p with a transition labelled by ℓ, and *false* otherwise. Formally, given an equivalence class of states I we define:

$$\gamma(p, \ell, I) \stackrel{\text{def}}{=} \exists q \in I.\ p \xrightarrow{\ell} q.$$

Suppose we are given a (strong) bisimulation relation R. We know that its induced equivalence relation $\equiv_R$ is also a bisimulation. Let I be an equivalence classes induced by R. By definition of bisimulation we have that taken any two states $s_1, s_2 \in I$ if $s_1 \xrightarrow{\ell} s_1'$ for some ℓ and s_1' then it must be the case that there exists s_2' such that $s_2 \xrightarrow{\ell} s_2'$ and s_2' is in the same equivalence class I' as s_1' (and vice versa).

Now consider the function $\Phi : \wp(S) \rightarrow \wp(S)$ defined by letting:

$$p\ \Phi(R)\ q \stackrel{\text{def}}{=} (\ \forall \ell \in L, I \in R.\ \gamma(p, \mu, I) \Leftrightarrow \gamma(q, \mu, I)\ )$$

where I ranges over the equivalence classes induced by the relation R. Then, by the argument above, a (strong) bisimulation is just a relation such that $R \subseteq \Phi(R)$ and the largest bisimulation is the bisimilarity relation

$$\simeq \stackrel{\text{def}}{=} \bigcup_{R \subseteq \Phi(R)} R.$$

The construction Φ can be extended to the case of CTMCs. The idea is that equivalent states will fall into the same equivalence class and that if a state has multiple transitions with rates $\lambda_1, \ldots, \lambda_n$ to different states $s_1, \ldots, s_n$ that are in the same equivalence class, then we can represent all such transitions by a single transitions that carries the rate $\sum_{i=1}^{n} \lambda_i$. To this aim, we define a function $\gamma_C : S \times \wp(S) \rightarrow \mathbb{R}$ simply by extending the transition function to sets of states as follows:

$$\gamma_C(s, I) = \sum_{s' \in I} \alpha(s, s')$$

As we have done above for LTSs, we define the function $\Phi : \wp(S) \to \wp(S)$ by:

$$s_1 \; \Phi(R) \; s_2 \overset{\text{def}}{=} \forall I \in S_{/\equiv_R}. \; \gamma_C(s_1, I) = \gamma_C(s_2, I)$$

meaning that the total rate of reaching any equivalence class of R from s_1 is the same as that of s_2.

Definition 14.23 (CTMC bisimilarity). A *CTMC bisimulation* is a relation R such that $R \subseteq \Phi(R)$ and the *CTMC bisimilarity* is the relation

$$\simeq \overset{\text{def}}{=} \bigcup_{R \subseteq \Phi(R)} R.$$

Let us show how this construction works with an example. Abusing the notation, in the following we write $\alpha(s, I)$ instead of $\gamma_C(s, I)$.

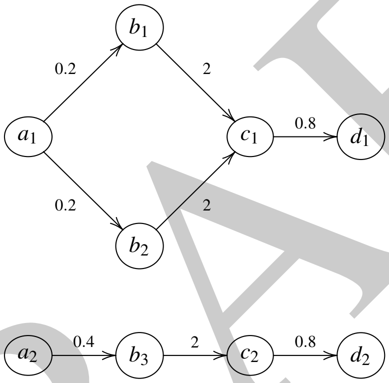


Fig. 14.6: CTMC bisimilarity

Example 14.8. Let us consider the two CTMCs in Figure 14.6. We argue that the following equivalence relation R identifies bisimilar states:

$$R = \{ \, \{a_1, a_2\}, \{b_1, b_2, b_3\}, \{c_1, c_2\}, \{d_1, d_2\} \, \}.$$

Let us show that R is a CTMC bisimulation: whenever two states are related, we must check that the sum of the rates from them to the states on any equivalence class coincide. For a_1 and a_2, we have

$$\begin{aligned}
\alpha(a_1, \{a_1, a_2\}) &= \alpha(a_2, \{a_1, a_2\}) &&= 0 \\
\alpha(a_1, \{b_1, b_2, b_3\}) &= \alpha(a_2, \{b_1, b_2, b_3\}) &&= 0.4 \\
\alpha(a_1, \{c_1, c_2\}) &= \alpha(a_2, \{c_1, c_2\}) &&= 0 \\
\alpha(a_1, \{d_1, d_2\}) &= \alpha(a_2, \{d_1, d_2\}) &&= 0.
\end{aligned}$$

For b_1, b_2, b_3 we have

$$\alpha(b_1,\{c_1,c_2\}) = \alpha(b_2,\{c_1,c_2\}) = \alpha(b_3,\{c_1,c_2\}) = 2.$$

Note that we no longer mention all remaining trivial cases concerned with the other equivalence classes, where α returns 0, because there are no transitions to consider.

Finally, we have one last non trivial case to check:

$$\alpha(c_1,\{d_1,d_2\}) = \alpha(c_2,\{d_1,d_2\}) = 0.8.$$

14.4.7 DTMC Bisimilarity

One could think that the same argument about bisimilarity that we have exploited for CTMCs can be also extended to DTMCs. It is easy to show that if a DTMC has no deadlock states, in particular if it is ergodic, then bisimilarity becomes trivial (see Problem 14.1). This does not mean that the concept of bisimulation on ergodic DTMCs is useless, in fact these relations (finer than bisimilarity) can be used to factorise the chain (lumping) in order to study particular properties.

If we consider DTMCs with some deadlock states, then bisimilarity can be non trivial. Let us define the function $\gamma_D : S \to \wp(S) \to (\mathbb{R}+1)$ as follows:

$$\gamma_M(s)(I) = \begin{cases} * & \text{if } \alpha(s) = * \\ \sum_{s' \in I} \alpha(s)(s') & \text{otherwise} \end{cases}$$

Correspondingly, we set $\Phi : \wp(S) \to \wp(S)$ to be defined as:

$$s_1\ \Phi(R)\ s_2 \overset{\text{def}}{=} \forall I \in S_{/\equiv_R}.\ \gamma_D(s_1)(I) = \gamma_D(s_2)(I).$$

In this case any two deadlock states s_1, s_2 are bisimilar, because

$$\forall I.\ \gamma_D(s_1)(I) = \gamma_D(s_2)(I) = *$$

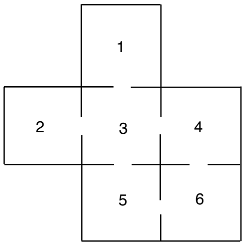
and they are separated from any non deadlock state s, as

$$\forall I.\ \gamma_D(s_1)(I) = * \neq \gamma_D(s)(I) \in \mathbb{R}.$$

Problems

14.1. Prove that the bisimilarity relation in a DTMC $\alpha : S \to (D(S)+1)$ without deadlock states (and in particular, when it is ergodic) is always the universal relation $S \times S$.

14.2. A mouse runs through the maze shown below.



At each step it stays in the room or it leaves the room by choosing at random one of the doors (all choices have equal probability).

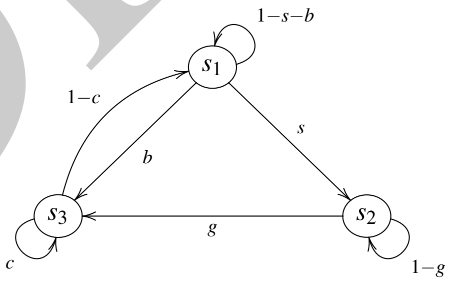
1. Draw the transition graph and give the matrix P for this DTMC.
2. Show that it is ergodic and compute the steady state distribution.
3. Assuming the mouse is initially in room 1, what is the probability that it is in room 6 after three steps?

14.3. Show that the DTMC described by the matrix

$$\begin{vmatrix} \frac{1}{4} & 0 & \frac{3}{4} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}$$

has more than one stationary distribution. Explain why it is so.

14.4. With the Markov chain below we intend to represent the scenario where Mario, a taxi driver, is looking for costumers. In state s_1, Mario is in the parking place waiting for costumers, which arrive with probability b. Then Mario moves to the busy state s_3, with probabilities c of staying there and $1-c$ of moving back to s_1. Alternatively, Mario may decide, with probability s, of moving around (state s_2), driving in the busiest streets of town looking for clients, which may show up with probability g.



1. Check that the Markov chain above is ergodic.
2. Compute the steady state probabilities π_1, π_2 and π_3 for the three states s_1, s_2 and s_3 as functions of the parameters b, c, g and s.

3. Evaluate the probabilities for suitable values of the parameters, e.g.,

$$b = 0.5, \quad c = 0.5, \quad g = 0.8, \quad s = 0.3$$

4. Prove that, when it is very likely to find costumers on the streets (i.e., when $g = 1$), in order to maximise π_3, Mario must always move around (i.e., he must choose $s = 1 - b$).

14.5. A state s_i of a Markov chain is called *absorbing* if $\alpha(s_i)(s_i) = 1$, and a Markov chain is *absorbing* if it has at least one absorbing state. Can an absorbing Markov chain be ergodic? Explain.

14.6. A machine can be described as being in three different states: (R) under repair, (W) waiting for a new job, (O) operating.

- While the machine is operating the probability to break down is $\frac{1}{20} = 0.05$ and the probability to finish the task (and go to waiting) is $\frac{1}{10} = 0.1$.
- If the machine is under repair there is a $\frac{1}{10} = 0.1$ probability to get repaired, and then the machine will become waiting.
- A broken machine is never brought directly (in one step) to operation.
- If the machine is waiting, there is a $\frac{9}{10} = 0.9$ probability to get into operation.
- A waiting machine does not break.

1. Describe the system as a DTMC, draw the corresponding transition system and define the transition probability matrix. Is it ergodic?
2. Assume that the machine is waiting at time t. What is the probability to be operating at time $t + 1$? Explain.
3. What is the probability that the machine is operating after a long time? Explain.

14.7. A certain calculating machine uses only the digits 0 and 1. It is supposed to transmit one of these digits through several stages. However, at every stage, there is a probability p that the digit that enters this stage will be changed when it leaves and a probability $q = 1 - p$ that it won't.

1. Form a Markov chain to represent the process of transmission. What are the states? What is the matrix of transition probabilities?
2. Assume that the digit 0 enters the machine: what is the probability that the machine, after two stages, produces the digit 0? For which value of p is this probability minimal?

14.8. Consider a CTMC with state space $S = \{0, 1\}$. The only possible transitions are described by the rates $q_{0,1} = \lambda$ and $q_{1,0} = \mu$. Compute the following:

1. the embedded DTMC;
2. the state probabilities $\pi^{(t)}$ in terms of the initial distribution $\pi^{(0)}$;
3. the steady state probability distribution.

14.9. Consider a CTMC with $N+1$ states representing the number of possible active instances of a service, from 0 to a maximum N. Let i denote the number of currently active instances. A new instance can be spawn with rate

$$\lambda_i \overset{\text{def}}{=} (N-i) \times \lambda$$

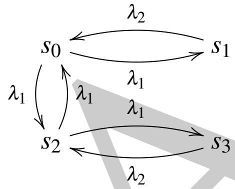
for some fixed λ, i.e., the rate decreases as there are more instances already running,[2] while an instance is terminated with rate

$$\mu_i \overset{\text{def}}{=} i \times \mu$$

for some fixed μ, i.e., the rate increases as there are more active instances to be terminated.

1. Model the system as a CTMC;
2. Compute the infinitesimal generator matrix;
3. Find the steady state probability distribution.

14.10. Let us consider the CTMC



1. What is the probability to sojourn in s_0 for some time t?
2. Assume $\lambda_2 > 2\lambda_1$: are there some bisimilar states?

14.11. Prove that computing the steady state distribution of a CTMC by solving the system of (homogeneous, normalised) linear equations $\pi\, Q = 0$ gives the same result as computing the steady state distribution of the embedded DTMC.

[2] Imagine the number of client is fixed, when i instances of the service are already active to serve i clients, then the number of clients that can require a new instance of the service is decreased by i.