J. Eur. Math. Soc. 8, 49–75

Giacomo Aletti · Ely Merzbach

Stopping Markov processes and first path on graphs

Received November 21, 2004 and in revised form June 3, 2005

Abstract. Given a strongly stationary Markov chain (discrete or continuous) and a finite set of stopping rules, we show a noncombinatorial method to compute the law of stopping. Several examples are presented. The problem of embedding a graph into a larger but minimal graph under some constraints is studied. Given a connected graph, we show a noncombinatorial manner to compute the law of a first given path among a set of stopping paths. We prove the existence of a minimal Markov chain without oversized information.

Keywords. Markov chains, stopping rules, directed graph

1. Introduction

Let X_n be a stationary Markov chain (process) on a finite set E with transition matrix P (intensity Q). The Markov process stops when one of the given stopping rules applies. We assume the rules take a very general form that will be described in the following. The problem of finding the stopping law may be solved by embedding the Markov chain into another Markov chain on a larger state set (the tree made by both the states and the stopping rules). The desired law is then obtained from the transition matrix of the new Markov chain.

Unfortunately, this new Markov chain may be so big that numerical computations can be not practicable. The problem here is to find a way of compressing the oversized information. In this paper, we present a new method permitting to obtain a projection of the Markov chain into a "minimal" Markov chain which preserves probabilities.

The problem of finding general closed forms for different kinds of waiting problems is widely studied. As an example, Ebneshahrashoob and Sobel [6] derived distributional results for the random variables in the case of Bernoulli trials. Several extensions have

Work done while the second author stayed at the University of Milano. He wishes to thank G. Aletti for the invitation and warm hospitality.

G. Aletti: Dipartimento di Matematica, Università di Milano, 20133 Milan, Italy;
e-mail: giacomo.aletti@unimi.it

E. Merzbach: Department of Mathematics, Bar-Ilan University, Ramat-Gan, Israel;
e-mail: merzbach@macs.biu.ac.il

Mathematics Subject Classification (2000): Primary 60J10; Secondary 60J05, 05C20

appeared recently to Markov-dependent trials via combinatorial or Markov chain embedding (see, e.g. Aki and Hirano [2]; Antzoulakos and Philippou [3]; Koutras and Alexandrou [7]) and in general closed forms by Stefanov [10].

Stefanov and Pakes [11] explicitly derive the joint distributions of various quantities associated with the time of reaching an arbitrary pattern of zeros and ones in a sequence of Bernoulli (or dependent) trials. The methodology is based on first embedding the problem into a more general framework for an appropriate finite-state Markov chain with one absorbing state and then treating that chain using tools of exponential families.

The motivation of our work comes from many situations.

(i) In finance the filter rule for trading is a special case of the Markov chain stopping rule suggested by the authors (see [8]).

(ii) "When enough is enough"! For example, an insured has an accident only occasionally. How many accidents in a specified number of years should be used as a stopping time for the insured (in other words, when should the insurance contract be discontinued)?

(iii) *State dependent Markov chains.* Namely, the transition probabilities are given in terms of the history. For simplicity consider the decision to stop if we get two identical throws $(11, 22, 33, \dots, nn)$ (for example, when $n = 2$, an insured has two kinds of accidents in a row, one each year, and his contract is discontinued, or an insured has no accidents two years in a row and therefore he is "promoted" to a better class of insured). If the probability of a switch from hm to mk is denoted by $p_{hm,mk}$ then the Markov transition matrix has the form:

$$
\begin{array}{c|ccccccccccccc}
 & 11 & 12 & \dots & 1n & 21 & 22 & \dots & 2n & \dots & n1 & n2 & \dots & nn \\
\hline
11 & p_{11,11} & p_{11,12} & \cdots & p_{11,1n} & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 & \\
12 & 0 & 0 & \dots & 0 & p_{12,21} & p_{12,22} & \cdots & p_{12,2n} & \cdots & 0 & 0 & \dots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1n & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & p_{1n,n1} & p_{1n,n2} & \cdots & p_{1n,nn} \\
21 & p_{21,11} & p_{21,12} & \cdots & p_{21,1n} & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\
22 & 0 & 0 & \dots & 0 & p_{22,21} & p_{22,22} & \cdots & p_{22,2n} & \cdots & 0 & 0 & \dots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
2n & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & p_{2n,n1} & p_{2n,n2} & \cdots & p_{2n,nn} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
n1 & p_{n1,11} & p_{n1,12} & \cdots & p_{n1,1n} & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\
n2 & 0 & 0 & \dots & 0 & p_{n2,21} & p_{n2,21} & \cdots & p_{n2,2n} & \cdots & 0 & 0 & \dots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
nn & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \cdots & p_{nn,n1} & p_{nn,n2} & \cdots & p_{nn,nn}
\end{array}
\tag{1}
$$

which can be analyzed for the stopping time by the usual methods. Obviously, in many situations (e.g., if $p_{hm,mk} = p_{m,k}\ \forall h \neq m$), this matrix has a special structure and can be reduced. This is the applied part of the paper: When a big matrix can be shrinked then the paper provides a mechanism for handling the stopping time issue.

(iv) Medical sciences: given that the length of a menstrual cycle has a known distribution, what is the probability that the length of a woman's menstrual cycle is the same three consecutive times?

(v) *Small-world networks*. Given one of the networks as in Figure 1, is it possible to reduce it and to preserve the law of reaching a given absorbing state?

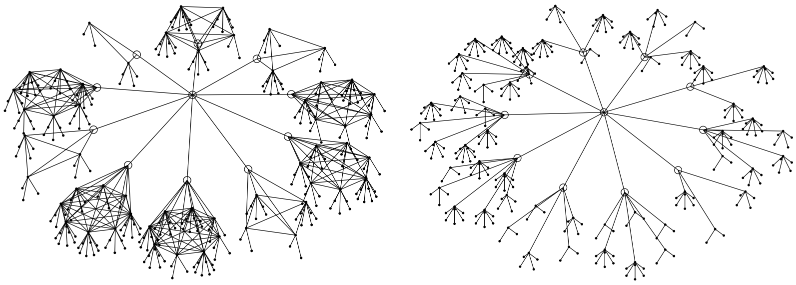


Fig. 1. Networks that may be shrinked.

There are of course many other such examples (e.g., records: Arnold, Balakrishnan, and Nagaraja [4] and optimization: Cairoli and Dalang [5]).

In the next section, we present some examples in which our method is used, giving elegant solutions to some cumbersome combinatorial problems. It happens that this framework is very well adapted to the language of graphs.

In Section 3, using some tools from Pattern-Matching Algorithms, we discuss the problem of embedding a graph into a larger but minimal graph when some constraints are imposed.

Section 4 is devoted to finding a necessary and sufficient condition for a projection to be compatible. Moreover, we can prove the existence of a minimal Markov chain without oversized information. Finally, in the last section, we show that our results can be translated very easily into the language of Category Theory, leading to a neat and concise formalism.

2. A combinatorial problem

Let $X = \{X_n, n \in \mathbb{N}\}$ be a Markov chain on a finite state space $E = \{e_1, \dots, e_n\}$:

$$\begin{array}{c|ccc} & e_1 & \dots & e_n \\ \hline e_1 & p_{1,1} & \dots & p_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ e_n & p_{n,1} & \dots & p_{n,n} \end{array} \left. \vphantom{\begin{array}{c} e_1 \\ \vdots \\ e_n \end{array}} \right\} =: P$$

The process is stopped when it reaches one of some given states $F := (e_{n_i})_{i=1}^{k} \subseteq E$. We can permute the order of the states so that $F = \{e_1, \dots, e_k\}$ and $E \setminus F = \{e_{k+1}, \dots, e_n\}$. To compute the law of stopping, we may consider a new Markov chain $X' = \{X'_n, n \in \mathbb{N}\}$ on $F \cup (E \setminus F)$:

$$
\left.\begin{array}{cc|cccccc}
 & & F & \dots & F & E\setminus F & \dots & E\setminus F \\
 & & e_1 & \dots & e_k & e_{k+1} & \dots & e_n \\
\hline
F & e_1 & \mathbf{1} & \dots & \mathbf{0} & 0 & \dots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
F & e_k & \mathbf{0} & \dots & \mathbf{1} & 0 & \dots & 0 \\
E\setminus F & e_{k+1} & p_{k+1,1} & \cdots & p_{k+1,k} & p_{k+1,k+1} & \cdots & p_{k+1,n} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
E\setminus F & e_n & p_{n,1} & \dots & p_{n,k} & p_{n,k+1} & \cdots & p_{n,n}
\end{array}\right\} =: P'
$$

Thus, the probability of reaching F by time n is reduced to the computation of the n-th power of P':

$$
\mathcal{P}\Big(\bigcup_{i=1}^{n}\{X_i \in F\}\Big) = \mathcal{P}(\{X'_n \in F\}) = \overbrace{[p_1^0, \dots, p_0^n]}^{\mathbf{p}_0}(P')^n \begin{bmatrix} 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}1\\ \vdots\\ 1\end{matrix}}\right\} k \text{ terms} \\ \left.\vphantom{\begin{matrix}0\\ \vdots\\ 0\end{matrix}}\right\} n-k \text{ terms} \end{matrix} . \tag{2}
$$

Obviously, there may be some oversized information in (2). In fact, there exists a trivial reduction (see Corollary 28) which preserves the above calculation for any $n \in \mathbb{N}$ and initial distribution $\mathbf{p}_0$:

$$
\mathcal{P}\Big(\bigcup_{i=1}^{n}\{X_i \in F\}\Big) =
$$

$$
\Big[\sum_{i=1}^{k} p_i^0, p_i^{k+1}, \dots, p_0^n\Big] \begin{pmatrix} \mathbf{1} & 0 & \dots & 0 \\ \sum_{i=1}^{k} p_{k+1,i} & p_{k+1,k+1} & \cdots & p_{k+1,n} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{k} p_{n,i} & p_{n,k+1} & \dots & p_{n,n} \end{pmatrix}^n \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} .
$$

Example 1. The process X' for the problem of state dependent Markov chains given in (1) is accordingly defined by the transition matrix

$$
\begin{array}{c|cccccccccccc}
 & 11 & 12 & \dots & 1n & 21 & 22 & \dots & 2n & \dots & n1 & n2 & \dots & nn \\
\hline
11 & \mathbf{1} & \mathbf{0} & \dots & \mathbf{0} & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 & \\
12 & 0 & 0 & \dots & 0 & p_{12,21} & p_{12,22} & \cdots & p_{12,2n} & \cdots & 0 & 0 & \dots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
1n & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \cdots & p_{1n,n1} & p_{1n,n2} & \cdots & p_{1n,nn} \\
21 & p_{21,11} & p_{21,12} & \cdots & p_{21,1n} & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\
22 & 0 & 0 & \dots & 0 & \mathbf{0} & \mathbf{1} & \dots & \mathbf{0} & \dots & 0 & 0 & \dots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
2n & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \cdots & p_{2n,n1} & p_{2n,n2} & \cdots & p_{2n,nn} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
n1 & p_{n1,11} & p_{n1,12} & \cdots & p_{n1,1n} & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\
n2 & 0 & 0 & \dots & 0 & p_{n2,21} & p_{n2,21} & \cdots & p_{n2,2n} & \cdots & 0 & 0 & \dots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
nn & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & \mathbf{0} & \mathbf{0} & \dots & \mathbf{1}
\end{array} \tag{3}
$$

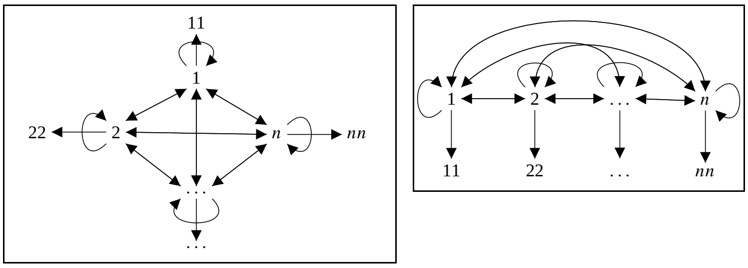


Fig. 2. Two representations of the Markov network given in Example 1. The final states ii are generated by the stopping rule $S = \inf\{i : X_{i-1} = X_i\}$.

When $p_{hm,mk} = p_{m,k}$ $\forall h \neq m$, X is a Markov chain. Moreover, the solution is the cumulative distribution of the stopping time

$$S = \inf\{i \in \mathbb{N} \colon X_{i-1} = X_i\}$$

and therefore the previous matrix may be reduced to

	11	22	...	nn	1	2	...	n
11	**1**	**0**	...	**0**	0	0	...	0
22	**0**	**1**	...	**0**	0	0	...	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
nn	**0**	**0**	...	**1**	0	0	...	0
1	$p_{1,11}$	0	...	0	0	$p_{1,2}$	...	$p_{1,n}$
2	0	$p_{2,22}$	...	0	$p_{2,1}$	0	...	$p_{2,n}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	0	0	...	$p_{n,nn}$	$p_{n,1}$	$p_{1,2}$	...	0

	T	1	2	...	n
T	**1**	0	0	...	0
1	$p_{1,11}$	0	$p_{1,2}$	...	$p_{1,n}$
2	$p_{2,22}$	$p_{2,1}$	0	...	$p_{2,n}$
⋮	⋮	⋮	⋮	⋮	⋮
n	$p_{n,nn}$	$p_{n,1}$	$p_{1,2}$	...	0

The process X' given by (3) is therefore the extension of the process X on the graph of Figure 2. The couple (X, S) "builds" a new process X' on a tree (see Theorem 17). The process X' is "tree-adapted" (specified in Definition 16).

The Markov chain X' may have a lot of oversized information, as the following example shows.

Example 2. What is the probability that the length of a woman's menstrual cycle is the same three consecutive times? If the length of a menstrual cycle is uniformly distributed between 25 and 35 days (and the lengths of menstrual cycles are independent), then the process may be seen as a Markov chain on $E = \{25, \dots, 35\}$, where

$$P = \begin{pmatrix} 1/10 & \dots & 1/10 \\ \vdots & \ddots & \vdots \\ 1/10 & \dots & 1/10 \end{pmatrix}$$

and the problem is related to the stopping time defined by

$$S = \inf\{i \in \mathbb{N} \colon X_{i-2} = X_{i-1} = X_i\}.$$

The process X' has 21 states (see Corollary 28) and its transition matrix is defined in (4).

	T	25.25	26.26	...	35.35	25	26	...	35
T	**1**	0	0	...	0	0	0	...	0
25.25	1/10	0	0	...	0	0	1/10	...	1/10
26.26	1/10	0	0	...	0	1/10	0	...	1/10
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
35.35	1/10	0	0	...	0	1/10	1/10	...	0
25	0	1/10	0	...	0	0	1/10	...	1/10
26	0	0	1/10	...	0	1/10	0	...	1/10
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
35	0	0	0	...	1/10	1/10	1/10	...	0

(4)

The previous example can be simplified by considering the process Z_n with the following three states:

$$\widehat{E} = \begin{cases} 1 & \text{if } X_{n-1} \neq X_n \text{ and } \nexists N \leq n \colon X_{N-2} = X_{N-1} = X_N, \\ 2 & \text{if } X_{n-2} \neq X_{n-1} = X_n \text{ and } \nexists N \leq n \colon X_{N-2} = X_{N-1} = X_N, \\ 3 = T & \text{if } \exists N \leq n \colon X_{N-2} = X_{N-1} = X_N, \end{cases}$$

with initial distribution $\widehat{p}_1 = [1, 0, 0]$ and matrix

	1	2	$3 = T$
1	p	$1 - p$	0
2	p	0	$1 - p$
$3 = T$	0	0	1

$=: \widehat{P}$

where $p = 9/10$. In general, for solving this problem it is sufficient to note that

$$\widehat{P} = ADA^{-1},$$

where

$$A = \begin{pmatrix} 1 & \frac{p+\sqrt{-3p^2+4p}}{2p} & \frac{p-\sqrt{-3p^2+4p}}{2p} \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix},$$

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{p+\sqrt{-3p^2+4p}}{2} & 0 \\ 0 & 0 & \frac{p-\sqrt{-3p^2+4p}}{2} \end{pmatrix},$$

$$A^{-1} = \begin{pmatrix} 0 & 0 & 1 \\ p & -\frac{p-\sqrt{-3p^2+4p}}{2} & -\frac{p+\sqrt{-3p^2+4p}}{2} \\ -p & \frac{p+\sqrt{-3p^2+4p}}{2} & \frac{p-\sqrt{-3p^2+4p}}{2} \end{pmatrix} \frac{1}{\sqrt{-3p^2+4p}}.$$

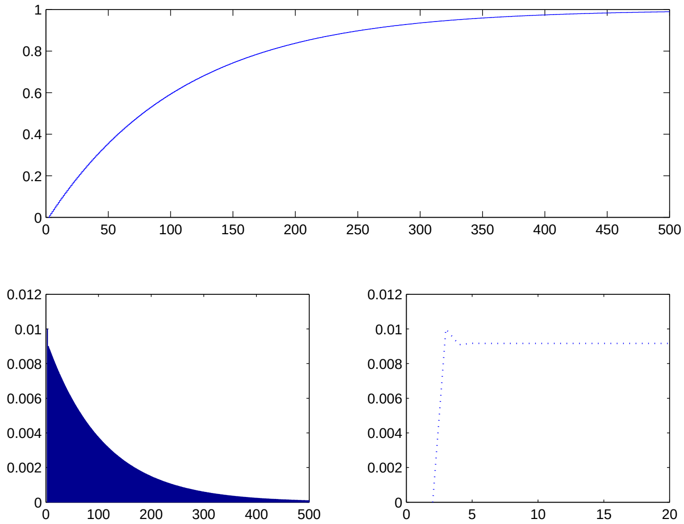


Fig. 3. Cumulative probability, probability density and hazard function for Example 2.

Hence

$$F_S(n) = 1 - \frac{\left(\frac{p+\sqrt{-3p^2+4p}}{2}\right)^{n+1}}{p\sqrt{-3p^2+4p}} + \frac{\left(\frac{p-\sqrt{-3p^2+4p}}{2}\right)^{n+1}}{p\sqrt{-3p^2+4p}} \tag{5}$$

and the corresponding hazard rate is

$$\begin{aligned} H_S(n) = \mathcal{P}(S = n \mid S \geq n) &= \frac{F_S(n) - F_S(n-1)}{1 - F_S(n-1)} \\ &= 1 - \frac{\left(\frac{p+\sqrt{-3p^2+4p}}{2}\right)^{n+1} - \left(\frac{p-\sqrt{-3p^2+4p}}{2}\right)^{n+1}}{\left(\frac{p+\sqrt{-3p^2+4p}}{2}\right)^{n} - \left(\frac{p-\sqrt{-3p^2+4p}}{2}\right)^{n}}. \end{aligned} \tag{6}$$

Equations (5) and (6) may be applied with $p = 9/10$, leading to (see Figure 3)

$$F_S(n) = 1 - \frac{100}{9\sqrt{117}}\left(\left(\frac{9+\sqrt{117}}{20}\right)^{n+1} - \left(\frac{9-\sqrt{117}}{20}\right)^{n+1}\right)$$

and

$$H_S(n) = 1 - \frac{1}{20}\frac{(9+\sqrt{117})^{n+1} - (9-\sqrt{117})^{n+1}}{(9+\sqrt{117})^{n} - (9-\sqrt{117})^{n}}.$$

What are we doing? In fact, we made a projection from a big Markov chain on $E' := \{25, 26, \ldots, 35, 25.25, 26.26, \ldots, 35.35, T\}$ (with the matrix given by (4)) to a smaller Markov chain on $\widehat{E} = \{1, 2, 3\}$ which preserves probability, as stated in Theorem 34. This projection is minimal (see Remark 32) and is unique, as stated in Theorem 39.

This result ensures that there exists a minimum Markov chain to which a large class of stopping problems may be reduced. Numerical efficient algorithms for such reduction and for ε-approximations of the problem in suitable spaces would be a real "chaos reduction" algorithm.

3. A graph generalization

Let X be a Markov chain on a set E. We denote by $\mathcal{E}$ the free semigroup with generators the elements of E, i.e.

$$\mathcal{E} := \{\mathbf{e} = (e_1, \ldots, e_n) \colon e_l \in E,\ n \in \mathbb{N}\}.$$

We denote by ϵ the unity (empty element) of $\mathcal{E}$. A "word" $\mathbf{e}$ will also be briefly denoted by $\mathbf{e} = (e_1, \ldots, e_n) = \prod_{i=1}^n e_i$.

In another context (Pattern-Matching Algorithms, here PMA, see [1]), the set E is called the *alphabet*, ϵ the *empty string*, and if we denote by L the language given by the strings in E then $\mathcal{E}$ is the *Kleene closure* of L: $\mathcal{E} = \{E\}^*$. Hence $\mathcal{E}$ is a regular language.

The *stopping rule* we consider here is a finite subset $\mathcal{A}$ of $\mathcal{E}$: the Markov process ends when an element of $\mathcal{A}$ occurs for the first time (in PMA, the problem is to find the first time we obtain $\{E\}^*\mathcal{A}$). Note that a stopping rule cannot be a substring of another stopping rule (otherwise, the latter can never occur: the process is stopped when its substring occurs). Thus, we will require that the elements in $\mathcal{A}$ are not comparable with respect to the relation $\lhd$ of "being a substring of".

The problem is solved by embedding E in a directed tree with root ϵ, first generation $\{E\}$ and tree leaves $\mathcal{A}$. We denote this set by $\overrightarrow{\mathcal{A} \cup E}$.

In this context, a tree is a particular subset of $\mathcal{E}$. Its nodes are identified by the fact that if a string belongs to the tree nodes, any of its substrings will also belong to it. The relation "being a prefix of" will be denoted by $\sqsubset$. The nodes in a tree are hence partially ordered by $\sqsubset$. The edges of the tree will be defined by means of the partial order $\sqsubset$ (successors, see (7)).

The first problem we study here is: given $\mathcal{A}$, what is the size (number of nodes) of $\overrightarrow{\mathcal{A} \cup E}$? Equivalently, we are asking how many states the resolving Markov problem has.

3.1. *Size of a tree*

Two partial order relations are naturally defined on $\mathcal{E}$: for all $\mathbf{e}_1, \mathbf{e}_2 \in \mathcal{E}$,

$$\begin{aligned} \mathbf{e}_1 \sqsubset \mathbf{e}_2 \quad &\text{if} \quad \exists \mathbf{e}_3 \in \mathcal{E} \colon \mathbf{e}_1\mathbf{e}_3 = \mathbf{e}_2, \\ \mathbf{e}_1 \lhd \mathbf{e}_2 \quad &\text{if} \quad \exists \mathbf{e}_3 \in \mathcal{E} \colon \mathbf{e}_3\mathbf{e}_1 \sqsubset \mathbf{e}_2. \end{aligned}$$

Lemma 3. *The following conditions are equivalent:*

- $\mathbf{e}_1$ *and* $\mathbf{e}_2$ *are comparable* ($\mathbf{e}_1 \sqsubset \mathbf{e}_2$ *or* $\mathbf{e}_2 \sqsubset \mathbf{e}_1$)*;*
- $\exists \mathbf{e}_3 \in \mathcal{E} \colon \mathbf{e}_1 \sqsubset \mathbf{e}_3$ *and* $\mathbf{e}_2 \sqsubset \mathbf{e}_3$.

Proof. $\Rightarrow$: Suppose $\mathbf{e}_1 \sqsubset \mathbf{e}_2$. Then $\mathbf{e}_1 \sqsubset \mathbf{e}_2$ and $\mathbf{e}_2 \sqsubset \mathbf{e}_2$ (take $\mathbf{e}_3 = \mathbf{e}_2$).

$\Leftarrow$: Let $\mathbf{e}_3 = (e_{31}, \ldots, e_{3n})$. Since $\mathbf{e}_i \sqsubset \mathbf{e}_3$ $(i = 1, 2)$, we have $\mathbf{e}_i = (e_{31}, \ldots, e_{3n_i})$ with $n_i \leq n$, $i = 1, 2$. If $n_1 \leq n_2$, then $\mathbf{e}_1 \sqsubset \mathbf{e}_2$. Otherwise, $\mathbf{e}_2 \sqsubset \mathbf{e}_1$. □

In PMA, $\mathbf{e}_1 \sqsubset \mathbf{e}_2$ iff $\mathbf{e}_1$ is a prefix of $\mathbf{e}_2$, while $\mathbf{e}_1 \lhd \mathbf{e}_2$ iff $\mathbf{e}_1$ is a substring of $\mathbf{e}_2$. These relations extend to subsets of $\mathcal{E}$. For any $\mathcal{A}_1, \mathcal{A}_2 \subseteq \mathcal{E}$, we say

$$\begin{aligned} \mathcal{A}_1 \sqsubset \mathcal{A}_2 &\quad \text{if} \quad \forall \mathbf{a}_1 \in \mathcal{A}_1, \exists \mathbf{a}_2 \in \mathcal{A}_2 \colon \mathbf{a}_1 \sqsubset \mathbf{a}_2, \\ \mathcal{A}_1 \lhd \mathcal{A}_2 &\quad \text{if} \quad \forall \mathbf{a}_1 \in \mathcal{A}_1, \exists \mathbf{a}_2 \in \mathcal{A}_2 \colon \mathbf{a}_1 \lhd \mathbf{a}_2. \end{aligned}$$

Clearly, $\subseteq$ implies $\sqsubset$ which itself implies $\lhd$. Given $\mathbf{e} = (e_1, \ldots, e_n) \in \mathcal{E}$, we define $|\mathbf{e}| = n$. Given $\mathcal{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$, we define $\mu(\mathcal{A}) = n$ (counting measure), $|\mathcal{A}| = \sum_{i=1}^{n} |\mathbf{a}_i|$ and $\mathbb{A} = \{\mathcal{A} \subseteq \mathcal{E} \colon |\mathcal{A}| < \infty\}$. Then $\mathbb{A}$ is a ring with respect to the usual binary operations $\cup$ and $\triangle$ (symmetric difference). Moreover, μ is an additive measure on the ring $\mathbb{A}$.

Remark 4. $\sqsubset$ is a partial order on $\mathcal{E}$ but not on $\mathbb{A}$.

Definition 5. *A subset $\mathcal{A} \in \mathbb{A}$ is called* admissible with respect to $\sqsubset$ (resp. $\lhd$) *if for any $\mathbf{a}_1, \mathbf{a}_2 \in \mathcal{A}$, $\mathbf{a}_1 \sqsubset \mathbf{a}_2$ (resp. $\mathbf{a}_1 \lhd \mathbf{a}_2$) implies $\{\mathbf{a}_1, \mathbf{a}_2\} \not\subseteq \mathcal{A}$ (i.e. the elements of an admissible set $\mathcal{A}$ are not comparable). We denote by $\mathbb{A}_{\sqsubset}$ (resp. $\mathbb{A}_{\lhd}$) the collection of admissible sets with respect to $\sqsubset$ (resp. $\lhd$).*

The operator $|\cdot| : \mathbb{A} \to \mathbb{N}$ is clearly monotone with respect to $\subseteq$. When its domain is restricted to $\mathbb{A}_{\sqsubset}$, $|\cdot| : \mathbb{A}_{\sqsubset} \to \mathbb{N}$ is also monotone with respect to $\sqsubset$, as the following lemma shows.

Lemma 6. *Let $\mathcal{A}, \mathcal{A}' \in \mathbb{A}$ with $\mathcal{A} \sqsubset \mathcal{A}'$ and $\mathcal{A} \in \mathbb{A}_{\sqsubset}$. Then $|\mathcal{A}| \leq |\mathcal{A}'|$.*

Proof. Since $\mathcal{A} \sqsubset \mathcal{A}'$, there is $\phi : \mathcal{A} \to \mathcal{A}'$ such that $\mathbf{a} \sqsubset \phi(\mathbf{a})$ for all $\mathbf{a} \in \mathcal{A}$. The assertion will follow from the fact that ϕ is injective. In order to prove this fact, let $\mathbf{a}_1, \mathbf{a}_2 \in \mathcal{A}$ be such that $\phi(\mathbf{a}_1) = \phi(\mathbf{a}_2) =: \mathbf{a}'$. By Lemma 3, $\mathbf{a}_1 \sqsubset \mathbf{a}'$ and $\mathbf{a}_2 \sqsubset \mathbf{a}'$ imply that $\mathbf{a}_1$ and $\mathbf{a}_2$ are comparable. Since $\mathcal{A} \in \mathbb{A}_{\sqsubset}$, it follows that $\mathbf{a}_1 = \mathbf{a}_2$. □

Proposition 7. *For any $\mathcal{A} \in \mathbb{A}$, there exists a unique set $\mathcal{A}_{\sqsubset} \in \mathbb{A}_{\sqsubset}$ such that $\mathcal{A} \sqsubset \mathcal{A}_{\sqsubset}$ and $\mathcal{A}_{\sqsubset} \subseteq \mathcal{A}$ (in particular, $|\mathcal{A}_{\sqsubset}| \leq |\mathcal{A}|$). Moreover, if $\mathcal{A} \sqsubset \mathcal{A}' \in \mathbb{A}$, then $\mathcal{A}_{\sqsubset} \sqsubset \mathcal{A}' \sqsubset \mathcal{A}'_{\sqsubset}$ (in particular, $|\mathcal{A}_{\sqsubset}| \leq |\mathcal{A}'_{\sqsubset}| \leq |\mathcal{A}'|$).*

Proof. Define

$$\mathcal{A}_{\sqsubset} := \{\mathbf{a} \in \mathcal{A} \colon \mathbf{a} \not\sqsubset \mathbf{a}',\ \forall \mathbf{a}' \in \mathcal{A}, \mathbf{a}' \neq \mathbf{a}\}.$$

By definition, $\mathcal{A}_{\sqsubset} \in \mathbb{A}_{\sqsubset}$ and $\mathcal{A}_{\sqsubset} \subseteq \mathcal{A}$. Since $\mathcal{A} \in \mathbb{A}$, it follows that $|\mathcal{A}| < \infty$ and hence $\mathcal{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$. We will show that $\mathbf{a}_1 \sqsubset \mathcal{A}_{\sqsubset}$. We have two possibilities:

Case A: $\mathbf{a}_1 \not\sqsubset \mathbf{a}_i, \forall i = 2, \ldots, n$. In this case $\mathbf{a}_1 \in \mathcal{A}_{\sqsubset}$, and hence $\mathbf{a}_1 \sqsubset \mathcal{A}_{\sqsubset}$.

Case B: $\exists m > 1 \colon \mathbf{a}_1 \sqsubset \mathbf{a}_m$. In this case $\mathbf{a}_1 \sqsubset \mathcal{A}^1$, where

$$\mathcal{A}^1 = \{\mathbf{a}_m, \mathbf{a}_2, \ldots, \mathbf{a}_{m-1}, \mathbf{a}_{m+1}, \ldots, \mathbf{a}_n\}.$$

Note that $\mathcal{A}^1_{\sqsubset} = \mathcal{A}_{\sqsubset}$ and hence we may act again in the same way with $\mathcal{A}^1$ instead of $\mathcal{A}$ (this process ends in at most n steps, since $\mu(\mathcal{A}^1) = n - 1$ and $\{\mathbf{e}\}_{\sqsubset} = \{\mathbf{e}\}, \forall \mathbf{e} \in \mathcal{E}$). □

Lemma 8. *Let $\mathcal{A} \neq \{\epsilon\}$. $\mathcal{A} \in \mathbb{A}_{\sqsubset}$ iff* (a) $\{\epsilon\} \notin \mathcal{A}$ *and* (b) $\mathcal{A} \sqsubset \mathcal{A}'$ *implies* $|\mathcal{A}| \leq |\mathcal{A}'|$.

Proof. $\Leftarrow$: Suppose $\{\mathbf{a}_1, \mathbf{a}_2\} \subseteq \mathcal{A}$ and $\mathbf{a}_1 \sqsubset \mathbf{a}_2$. Take $\mathcal{A}' = \mathcal{A} \setminus \{\mathbf{a}_1\}$. Then $\mathcal{A} \sqsubset \mathcal{A}'$ and $|\mathcal{A}| = |\mathcal{A}'| + |\mathbf{a}_1| > |\mathcal{A}'|$, since $\mathbf{a}_1 \neq \epsilon$.

$\Rightarrow$: (a) Since $\forall \mathbf{e} \in \mathcal{E}$, $\epsilon \sqsubset \mathbf{e}$, it follows that $\mathcal{A} \in \mathbb{A}_{\sqsubset}$, $\mathcal{A} \neq \{\epsilon\}$ implies $\epsilon \notin \mathcal{A}$.

(b) Take $\mathcal{A} \sqsubset \mathcal{A}'$ and let $\mathcal{A}'_{\sqsubset}$ be as in Proposition 7. By the same proposition, $|\mathcal{A}'_{\sqsubset}| \leq |\mathcal{A}'|$. By Lemma 6, $|\mathcal{A}| \leq |\mathcal{A}'_{\sqsubset}|$. ◻

Now, we define a rooted tree as the collection of words contained in its tree leaves. For example, if $E = \{1, 2\}$, then $\{\epsilon, (1), (2), (1, 1), (1, 1, 2)\}$ is a tree with tree leaves $(1, 1, 2)$ and (2), while $\{(2), (1, 1), (1, 1, 2)\}$ is not a tree, since (1) and ϵ are not contained in it. Hence, for any $\mathcal{A} \in \mathbb{A}$, let

$$\overrightarrow{\mathcal{A}} := \{\mathbf{e} \in \mathcal{E} \colon \{\mathbf{e}\} \sqsubset \mathcal{A}\}. \tag{7}$$

Definition 9. *$\mathcal{A} \in \mathbb{A}$ such that $\mathcal{A} = \overrightarrow{\mathcal{A}}$ is called a* rooted tree with root ϵ. *The set of all trees will be denoted by $\overrightarrow{\mathbb{A}}$, i.e.: $\overrightarrow{\mathbb{A}} := \{\mathcal{A} \in \mathbb{A} \colon \mathcal{A} = \overrightarrow{\mathcal{A}}\}$.*

Proposition 10. (i) *The set $\overrightarrow{\mathbb{A}}$ is closed under intersections and finite unions:*

$$\overrightarrow{\mathcal{A}_1} \cup \overrightarrow{\mathcal{A}_2} = \overrightarrow{\mathcal{A}_1 \cup \mathcal{A}_2} \quad \text{and} \quad \overrightarrow{\mathcal{A}_1} \cap \overrightarrow{\mathcal{A}_2} = \overrightarrow{\mathcal{A}_1 \cap \mathcal{A}_2}.$$

(ii) *Any tree $\overrightarrow{\mathcal{A}}$ is identified by the extreme values $\mathcal{A}_{\sqsubset}$ (tree leaves) and vice versa: for any $\mathcal{A} \in \mathbb{A}$,*

$$\overrightarrow{(\overrightarrow{\mathcal{A}})_{\sqsubset}} = \overrightarrow{\mathcal{A}} \quad \text{and} \quad (\overrightarrow{\mathcal{A}_{\sqsubset}})_{\sqsubset} = \mathcal{A}_{\sqsubset},$$

i.e., there exists a natural bijection between $\overrightarrow{\mathbb{A}}$ and $\mathbb{A}_{\sqsubset}$.

(iii) *μ is the unique function on $\overrightarrow{\mathbb{A}}$ identified by:*

- $\mu(\overrightarrow{\{\mathbf{a}\}}) = |\mathbf{a}| + 1$ *(i.e. the number of nodes of a tree with only one tree leaf is the length of its tree leaf plus* 1*, the root);*
- *for any $\mathcal{A}_1, \mathcal{A}_2 \in \mathcal{A}$,*

$$\mu(\overrightarrow{\mathcal{A}_1 \cup \mathcal{A}_2}) = \mu(\overrightarrow{\mathcal{A}_1}) + \mu(\overrightarrow{\mathcal{A}_2}) - \mu(\overrightarrow{\mathcal{A}_1 \cap \mathcal{A}_2}).$$

Proof. The proof of this proposition is trivial, since an element of $\mathbb{A}$ (and hence of $\overrightarrow{\mathbb{A}}$) is a finite collection of finite-length words. ◻

Corollary 11. *For any $\mathcal{A} \in \mathbb{A}$, $\mu(\overrightarrow{\mathcal{A}}) \leq |\mathcal{A}_{\sqsubset}| + 1$.*

Now, we want to find a formula for the number of nodes of $\overrightarrow{\mathcal{A}}$, i.e., $\mu(\overrightarrow{\mathcal{A}})$. This formula is quite simple, if we refer to a suitable function $M_{\mathcal{A}}$ (see Lemma 15).

3.2. Interior of a tree

The first step is to find the "interior" $\mathcal{A}^{\circ}$ of the set $\mathcal{A}$, which is the set of vertices in $\overrightarrow{\mathcal{A}}$ which have at least two children (see Figure 4).

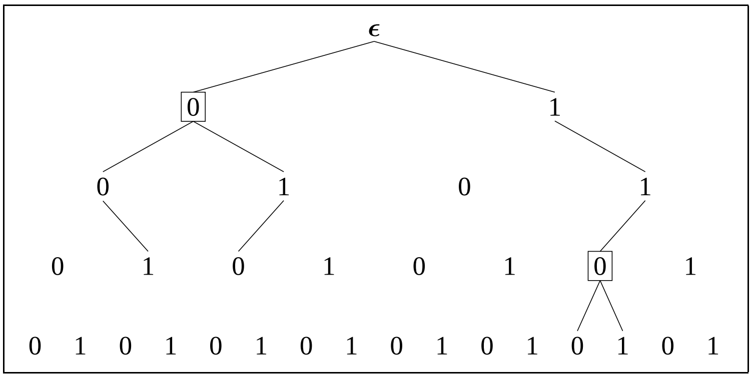


Fig. 4. On $E = \{0, 1\}$, the binary tree formed by the set $\mathcal{A} = \{(1, 1, 0, 1), (1, 1, 0, 0), (0, 0, 1), (0, 1, 0)\}$ and its interior $\mathcal{A}^\circ = \{(0), (1, 1, 0)\}$.

For this purpose, we define the function $\Phi_{\mathcal{A}} : \mathcal{E} \to \mathbb{N} \cup \{0\}$ by

$$\Phi_{\mathcal{A}}(\mathbf{e}) := \#\{\mathbf{a} \in \mathcal{A}_{\sqsubset} : \mathbf{e} \sqsubset \mathbf{a}\}.$$

Then $\Phi_{\mathcal{A}}$ maps each word $\mathbf{e}$ of $\mathcal{E}$ to the number of different tree leaves of $\overrightarrow{\mathcal{A}}$ that descend from $\mathbf{e}$. It is clear that a vertex $\mathbf{e} \in \mathcal{A}^\circ$ will be characterized by $2 \le \Phi_{\mathcal{A}}(\mathbf{e}) > \Phi_{\mathcal{A}}(\mathbf{e}')$, $\forall \mathbf{e} \sqsubset \mathbf{e}'$, $\mathbf{e} \neq \mathbf{e}'$. Moreover, we have the following

Lemma 12. *For any* $\mathbf{e} \in \mathcal{E} \setminus \mathcal{A}_{\sqsubset}$, $\Phi_{\mathcal{A}}(\mathbf{e}) = \sum_{e \in E} \Phi_{\mathcal{A}}(\mathbf{e}e)$.

Proof. It is sufficient to note that $\mathbf{e}e \sqsubset \mathbf{a}$ implies $\mathbf{e}e' \not\sqsubset \mathbf{a}$, for any distinct $e, e' \in E$. □

Define the function of the level sets of $\Phi_{\mathcal{A}}$,

$$\mathrm{Lev}_{\mathcal{A}} : \mathbb{N} \cup \{0\} \to \mathcal{E},$$

as the counter image of $\Phi_{\mathcal{A}}$ at value n:

$$\mathrm{Lev}_{\mathcal{A}}(n) := \{\mathbf{e} \in \mathcal{E} \colon \Phi_{\mathcal{A}}(\mathbf{e}) \ge n\} = \Phi_{\mathcal{A}}^{-1}([n, \infty)).$$

The set $\mathcal{A}^\circ$ will be the set of extremal values of the n^{th}-level sets ($n \ge 2$), i.e.

$$\mathcal{A}^{(n)} := \{\mathrm{Lev}_{\mathcal{A}}(n)\}_{\sqsubset}, \quad n > 0, \quad \mathcal{A}^\circ := \bigcup_{n=2}^{\infty} \mathcal{A}^{(n)}.$$

The following lemma shows that $\mathcal{A}^{(n)}$, $n > 0$, is well defined whenever $\mathcal{A} \in \mathbb{A}$.

Lemma 13. *Let* $\mathcal{A} \in \mathbb{A}$. *Then*

- (i) $\bigcup_{n=0}^{\infty} \mathrm{Lev}_{\mathcal{A}}(n) = \mathcal{E}$;
- (ii) $\forall n \ge m$, $\mathrm{Lev}_{\mathcal{A}}(n) \subseteq \mathrm{Lev}_{\mathcal{A}}(m)$;
- (iii) $\mathrm{Lev}_{\mathcal{A}}(n) \in \mathbb{A}$ *iff* $n > 0$;
- (iv) $\forall n \ge m > 1$, $\mathrm{Lev}_{\mathcal{A}}(n) \sqsubset \mathrm{Lev}_{\mathcal{A}}(m)$;
- (v) $\exists n_0 = n_0(\mathcal{A}) \colon \forall n > n_0$, $\mathrm{Lev}_{\mathcal{A}}(n) = \emptyset$.

Proof. (i)–(ii): Obvious. (iii): Let $\mathcal{A}^* = \{\mathbf{e} \in \mathcal{E} : |\mathbf{e}| > |\mathcal{A}|\}$. Then $\mathcal{A}^* \subseteq \mathrm{Lev}_{\mathcal{A}}(0) \setminus \mathrm{Lev}_{\mathcal{A}}(1)$, and hence $\infty = |\mathcal{A}^*| \leq |\mathrm{Lev}_{\mathcal{A}}(0)|$. By (i)–(ii), we have $\mathrm{Lev}_{\mathcal{A}}(n) \subseteq \mathcal{E} \setminus \mathcal{A}^*$, which implies $|\mathrm{Lev}_{\mathcal{A}}(n)| \leq |\mathcal{E} \setminus \mathcal{A}^*| < \infty$. (iv): This is a consequence of (ii)–(iii). (v): Let $n_0 = \#\{\mathcal{A}\}$. Then $n > n_0$ implies $\Phi_{\mathcal{A}}^{-1}(n) = \emptyset$. □

Now, given a set $\mathcal{A} \in \mathbb{A}$ and $n > 0$, we define, for any $\mathbf{e} \in \mathcal{E}$,

$$M_{\mathcal{A}}(\mathbf{e}) := \#\{e \in E : \mathbf{e}e \sqsubset \mathcal{A}\}$$

as the number of children in $\overrightarrow{\mathcal{A}}$ of the point $\mathbf{e}$. In fact, it is obvious that

$$M_{\mathcal{A}}(\mathbf{e}) = \#\{e \in E : \mathbf{e}e \sqsubset \overrightarrow{\mathcal{A}}\} = \#\{e \in E : \mathbf{e}e \sqsubset \mathcal{A}_{\sqsubset}\}$$

since $M_{\cdot}(\mathbf{e})$ is nondecreasing with respect to $\sqsubset$ and

$$\overrightarrow{\mathcal{A}} \sqsubset \mathcal{A}_{\sqsubset} \sqsubset \mathcal{A} \sqsubset \overrightarrow{\mathcal{A}}.$$

The next lemma shows that the "interior" $\mathcal{A}^\circ$ of the set $\mathcal{A}$ is the set of those vertices of $\overrightarrow{\mathcal{A}}$ which have at least two children.

Lemma 14. *Let $\mathcal{A} \in \mathbb{A}$. Then $M_{\mathcal{A}}(\mathbf{e}) > 0 \Leftrightarrow \mathbf{e} \in \overrightarrow{\mathcal{A}} \setminus \mathcal{A}_{\sqsubset}$. Moreover, $\mathbf{a} \in \mathcal{A}^\circ \Leftrightarrow M_{\mathcal{A}}(\mathbf{a}) \geq 2$.*

Proof. Let $\mathbf{e} \in \mathcal{E} \setminus (\overrightarrow{\mathcal{A}} \setminus \mathcal{A}_{\sqsubset}) = \mathcal{A}_{\sqsubset} \cup (\mathcal{E} \setminus \overrightarrow{\mathcal{A}})$. If $\mathbf{e} \in \mathcal{A}_{\sqsubset}$, then $\mathbf{e}e \not\sqsubset \mathcal{A}_{\sqsubset}$, $\forall e \in E$, and hence $M_{\mathcal{A}}(\mathbf{e}) = 0$. If $\mathbf{e} \notin \overrightarrow{\mathcal{A}}$, then $\mathbf{e}e \notin \overrightarrow{\mathcal{A}}$, and hence $M_{\mathcal{A}}(\mathbf{e}) = 0$. Thus, $M_{\mathcal{A}}(\mathbf{e}) > 0 \Rightarrow \mathbf{e} \in \overrightarrow{\mathcal{A}} \setminus \mathcal{A}_{\sqsubset}$. Conversely, suppose that $\mathbf{e} \in \overrightarrow{\mathcal{A}} \setminus \mathcal{A}_{\sqsubset}$. Then $\exists \mathbf{a} \in \mathcal{A}_{\sqsubset} : \mathbf{a} \neq \mathbf{e}, \mathbf{e} \sqsubset \mathbf{a}$, and hence $\exists e \in E : \mathbf{e}e \sqsubset \mathbf{a}$. Thus $M_{\mathcal{A}}(\mathbf{e}) > 0$.

Let $\mathbf{a} \in \mathcal{A}^\circ$. Then $\exists n_0 \geq 2 : \mathbf{a} \in \{\mathrm{Lev}_{\mathcal{A}}(n_0)\}_{\sqsubset}$. Then $\Phi_{\mathcal{A}}(\mathbf{a}) \geq n_0$, while $\Phi_{\mathcal{A}}(\mathbf{a}e) < n_0$, $\forall e \in E$ (otherwise

$$\mathrm{Lev}_{\mathcal{A}}(n_0) \ni \mathbf{a}e \sqsupset \mathbf{a} \in \{\mathrm{Lev}_{\mathcal{A}}(n_0)\}_{\sqsubset},$$

which is a contradiction). Hence $M_{\mathcal{A}}(\mathbf{a}) \geq 2$. Conversely, $M_{\mathcal{A}}(\mathbf{a}) \geq 2$ implies that $\exists e_1, e_2 \in E : e_1 \neq e_2, \mathbf{a}e_1 \in \overrightarrow{\mathcal{A}}, \mathbf{a}e_2 \in \overrightarrow{\mathcal{A}}$. Since $\mathbf{a}e_i \in \overrightarrow{\mathcal{A}}$ $(i = 1, 2)$, we have $\Phi_{\mathcal{A}}(\mathbf{a}e_i) > 0$. By Lemma 12, $\Phi_{\mathcal{A}}(\mathbf{a}) \geq \Phi_{\mathcal{A}}(\mathbf{a}e_1) + \Phi_{\mathcal{A}}(\mathbf{a}e_2) > \Phi_{\mathcal{A}}(\mathbf{a}e)$, $\forall e \in E$, and hence $\mathbf{a} \in \{\mathrm{Lev}_{\mathcal{A}}(\Phi_{\mathcal{A}}(\mathbf{a}))\}_{\sqsubset} \subseteq \mathcal{A}^\circ$. □

Proposition 15. *For any $\mathcal{A} \in \mathbb{A}$, we have*

$$\mu(\overrightarrow{\mathcal{A}}) = 1 + \sum_{\mathbf{a} \in \overrightarrow{\mathcal{A}}} (1 - M_{\mathcal{A}}(\mathbf{a}))|\mathbf{a}|. \tag{8}$$

Proof. Let $m = \#\{\mathcal{A}_{\sqsubset}\}$. We prove (8) by induction on m.

For $m = 1$, see Proposition 10(iii).

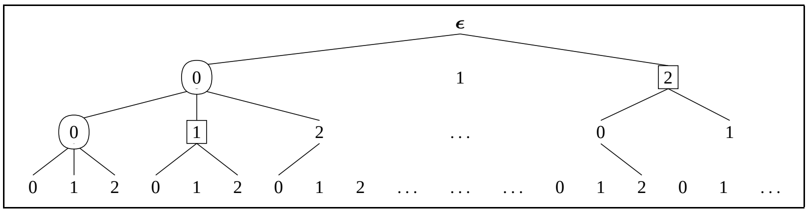


Fig. 5. Nodes with different numbers of children in a tree. Boxed: $\{\mathbf{a}\colon M_{\mathcal{A}}(\mathbf{a}) = 2\}$, circled: $\{\mathbf{a}\colon M_{\mathcal{A}}(\mathbf{a}) = 3\}$

For the induction step, let $\mathcal{A}_{\sqsubset} = \{\mathbf{a}_1, \ldots, \mathbf{a}_{m+1}\}$ and define $\mathcal{A}_1 = \{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$, $\mathcal{A}_2 = \{\mathbf{a}_{m+1}\}$. If $\mathbf{a}_{m+1} = \prod_{i=1}^{k} e_i$, we note that

$$\overrightarrow{\mathcal{A}_2} = \Big\{\epsilon, e_1, e_1 e_2, \ldots, \prod_{i=1}^{k} e_i = \mathbf{a}_{m+1}\Big\}.$$

Hence there exists $0 \leq l < k$ such that

$$\overrightarrow{\mathcal{A}_1 \cap \mathcal{A}_2} = \Big\{\epsilon, e_1, e_1 e_2, \ldots, \prod_{i=1}^{l} e_i =: \mathbf{a}\Big\}$$

(otherwise $\mathbf{a}_{m+1} \sqsubset \mathcal{A}_1$). Note that $M_{\mathcal{A}}(\mathbf{a}) = M_{\mathcal{A}_1}(\mathbf{a}) + 1$, while $M_{\mathcal{A}}(\mathbf{e}) = M_{\mathcal{A}_1}(\mathbf{e})$ for any $\mathbf{e} \in \overrightarrow{\mathcal{A}_1 \cap \mathcal{A}_2}$, $\mathbf{e} \neq \mathbf{a}$. Moreover, for any $l < p < k$, $M_{\mathcal{A}}(\prod_{i=l+1}^{p} e_i) = 1$ and $M_{\mathcal{A}}(\prod_{i=l+1}^{k} e_i) = 0$. By Proposition 10(iii),

$$\begin{aligned}
\mu(\overrightarrow{\mathcal{A}_1 \cup \mathcal{A}_2}) &= \mu(\overrightarrow{\mathcal{A}_1}) + \mu(\overrightarrow{\mathcal{A}_2}) - \mu(\overrightarrow{\mathcal{A}_1 \cap \mathcal{A}_2}) \\
&= \mu(\overrightarrow{\mathcal{A}_1}) + (k+1) - (l+1) \\
&= \Big(1 + \sum_{\mathbf{a} \in \overrightarrow{\mathcal{A}_1}} (1 - M_{\mathcal{A}_1}(\mathbf{a}))|\mathbf{a}|\Big) + k - l \qquad \text{(by induction hypothesis)} \\
&= 1 + \sum_{\mathbf{a} \in \overrightarrow{\mathcal{A}}} (1 - M_{\mathcal{A}}(\mathbf{a}))|\mathbf{a}|.
\end{aligned}$$ □

Definition 16. *Let $X = \{X_n, n \in \mathbb{N}\}$ be a process on a finite set of states $E = \{e_1, \ldots, e_m\}$ and adapted to the filtration $\{\mathcal{F}_n = \sigma(X_i, i \leq n), n \in \mathbb{N}\}$. Let $\mathcal{A} \in \mathbb{A}_{\lhd}$ be a set of admissible stopping rules with respect to $\lhd$. Then X is said to be an* $\mathcal{A}$-dependent Markov chain on *E if*

$$\forall e_j \in E, \qquad \mathcal{P}(X_{n+1} = e_j \mid \mathcal{F}_n) = \mathcal{P}(X_{n+1} = e_j \mid \mathcal{G}_n),$$

where $\mathcal{G}_n = \sigma(\{(X_{n-m}, X_{n-m+1}, \ldots, X_n) \in \overrightarrow{\mathcal{A} \cup E}\}, m \leq n)$.

An $\mathcal{A}$-dependent Markov chain is therefore a process such that the transition probability may depend only on the longest last path on the tree $\overrightarrow{\mathcal{A} \cup E}$. Obviously, *a stationary Markov chain is an $\mathcal{A}$-dependent Markov chain, for any $\mathcal{A}$.*

Theorem 17. *Let E be a finite set of states, $\mathcal{A} \in \mathbb{A}_{\lhd}$ be a set of admissible stopping rules with respect to $\lhd$ and X be an $\mathcal{A}$-dependent Markov chain on E. Then X can be embedded in a Markov chain X' on E', where $E \hookrightarrow E'$ and the cardinality of E' is $\mu(\overrightarrow{\mathcal{A} \cup E}) - 1 = \sum_{\mathbf{a} \in \overrightarrow{\mathcal{A} \cup E}} (1 - M_{\mathcal{A}}(\mathbf{a}))|\mathbf{a}|$ (see Corollary 28 for a trivial reduction of this number).*

Proof. Let $\mathcal{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_n\} \in \mathbb{A}$ be a set of stopping rules. If $\mathbf{a}_1, \mathbf{a}_2 \in \mathcal{A}$ and $\mathbf{a}_1 \lhd \mathbf{a}_2$, we can delete $\mathbf{a}_2$ as stated at the beginning of this section, and hence $\mathcal{A}$ must be a set of admissible stopping rules with respect to $\lhd$.

Now, let $E' := \overrightarrow{\mathcal{A} \cup E}$. Since X is an $\mathcal{A}$-dependent Markov chain, it follows that X can be embedded in a Markov chain X' on E'. Since $\mathcal{A} \in \mathbb{A}_{\lhd}$ implies $\mathcal{A} \in \mathbb{A}$, the number of states in E' arises from Proposition 15, as ϵ is not a possible state of the process.[1]

More precisely, for any state $\mathbf{e}_i \in E'$, we will find the states of E' reachable from $\mathbf{e}_i$ and their transition probability. We distinguish the following cases:

Case A: $\mathbf{e}_i \in \mathcal{A}$. In this case we have reached the stopping state defined by the corresponding rule. Then we will remain at $\mathbf{e}_i$: $q_{\mathbf{e}_i,\mathbf{e}_j} = \delta_{\mathbf{e}_i}(\mathbf{e}_j)$.

Case B: $\mathbf{e}_i \notin \mathcal{A}$. Let $\mathbf{e}_i = (e_{i_1}, \ldots, e_{i_m})$. The Markov process X has reached the state e_{i_m} (it will be consistent with what follows). Now, suppose that X jumps from the state e_{i_m} to the state e_j. We must find the corresponding state $\mathbf{e} \in \overrightarrow{\mathcal{A} \cup E}$ where X' jumps to. The idea is that X' will jump to $(\mathbf{e}_i, e_j)$ if $(\mathbf{e}_i, e_j)$ belongs to $\overrightarrow{\mathcal{A} \cup E}$. Otherwise, it will jump to the "maximum" available site. For example, if $E = \{1, 2, 3\}$, $\mathbf{e}_i = (1, 2, 1)$ and $e_j = 3$ the process will jump to $(1, 2, 1, 3)$ if $(1, 2, 1, 3) \in \overrightarrow{\mathcal{A} \cup E}$, otherwise it will try to jump to $(2, 1, 3)$. If $(2, 1, 3) \notin \overrightarrow{\mathcal{A} \cup E}$, then the process will reach the site $(1, 3)$—if $(1, 3) \in \overrightarrow{\mathcal{A} \cup E}$—or at least (3).

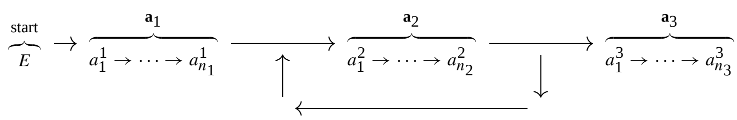
Thus, for $l = 1, \ldots, m + 1$, let

- $e_{i_{m+1}} = e_j$;
- $\mathbf{f}_j^l := (e_{i_l}, \ldots, e_{i_{m+1}})$;
- $r = \min\{l : \mathbf{f}_j^l \in \overrightarrow{\mathcal{A} \cup E}\}$;

and let $\mathbf{e} := \mathbf{f}_j^r$ be the state reachable from $\mathbf{e}_i$ "via e_j". We set $q_{\mathbf{e}_i,\mathbf{e}} = p_{\mathbf{e}_i,e_j}$.

We have $\sum_{\mathbf{e}} q_{\mathbf{e}_i,\mathbf{e}} = 1$ for any $\mathbf{e}_i \in \overrightarrow{\mathcal{A} \cup E}$, since $\sum_j p_{\mathbf{e}_i,e_j} = 1$ for any i. Moreover, $q_{\mathbf{e}_i,\mathbf{e}} \geq 0$ and hence the proof is complete once we have proved that the new process X' is a Markov chain on E'. This is a trivial consequence of the fact that X is an $\mathcal{A}$-dependent Markov chain on E (and so the process we have defined is a Markov process on E'). □

Remark 18. Note that the PMA framework suggests a trivial extension to some "possibly infinite" rules. The rule $\mathbf{a}_1\{\mathbf{a}_2\}^*\mathbf{a}_3$ may be modelled as follows:



[1] Note that $\mathcal{A}$ can be considered as a unique target absorbing state T (see Corollary 28).

Even if the stopping rule is not finite (it contains the stopping rules of the form

$$\{\mathbf{a}_1, \underbrace{\mathbf{a}_2, \ldots, \mathbf{a}_2}_{n \text{ times}}, \mathbf{a}_3\},$$

for all n), it defines an embedding in $E' = E \cup \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$.

4. A reduced graph Markov problem

Now, let X be a stationary Markov chain on an at most countable set E and let $T \subseteq E$ be the "target" absorbing states. We wish to compute the probability of reaching T by time n, given the initial distribution μ on E. If $(\Omega, \mathcal{F}, \mathcal{P})$ is the underlying probability space, we are accordingly interested in

$$\mathcal{P}^X_{E,\mu,T}(n) := \mathcal{P}\Big(\bigcup_{i=0}^{n}\{\omega \in \Omega \colon X_i(\omega) \in T\}\Big),$$

under the assumption that $\mathcal{P}(\{X^0 = i\}) = \mu(i)$.

The problem is the following: is there a "minimum" set F such that the problem may be projected to a problem on a Markov chain on F, for any initial distribution μ on E?

Note that, by the Total Probability Theorem, the previous problem is equivalent to the following:

Problem 19 *Let X be a stationary Markov chain on an at most countable set E and let $T \subseteq E$ be the "target" absorbing states.*

- *Is there a surjective function $\pi : E \to F$ such that $\pi(X)$ is a Markov chain on F and*

$$\mathcal{P}^X_{E,\delta_e,T}(n) = \mathcal{P}^{\pi(X)}_{\pi(E),\delta_{\pi(e)},\pi(T)}(n), \qquad \forall e \in E,\ \forall n \in \mathbb{N} \cup \{0\}\ ? \tag{9}$$

- *Is there a minimum set F satisfying this relation?*

Remark 20. In this problem, we are interested in the time of first entering the target set T. Thus, without loss of generality, we may (and will—see Remark 23) assume that T is an absorbing set:

$$\mathcal{P}(\{X_{n+1} \in T\} \mid \{X_n \in T\}) = 1.$$

First, let us see the equivalent problem in the network framework.

Definition 21. *A* complete directed graph *is a pair $(E, E \times E)$, where E is a nonempty at most countable set. A* complete network *is a triple $(E, E \times E, P)$ where $(E, E \times E)$ is a complete directed graph and $P : E \times E \to \mathbb{R}_+ \cup \{0\}$. A* Markov network *is a complete network $(E, E \times E, P)$ such that $\sum_{e_i \in E} P(e, e_i) = 1$ for any $e \in E$.*

Obviously, to any stationary Markov chain X on a space set E may be associated a Markov network on E since the $p_{i,j}$-matrix associated with the process is linked to the nonnegative function P by the relation $P(e_i, e_j) = p_{i,j}$, and vice versa. Note that no probability space is required for Markov networks: the only important tool is the transition matrix P.

First, we extend P to $\mathbb{P}$ in the classical way:

$$\mathbb{P} : E \times \mathfrak{P}(E) \to \mathbb{R}_+ \cup \{0\}$$

where $\mathfrak{P}(E)$ is the set of subsets of E and $\mathbb{P}(e, A) = \sum_{e_i \in A} P(e, e_i)$. Obviously, for each $e \in E$, $\mathbb{P}(e, \cdot)$ is a probability on $(E, \mathfrak{P}(E))$ that gives the conditional probability of reaching $\cdot$ given that we are in the state e: $\mathbb{P}(e, A) = \mathcal{P}(\{X_{n+1} \in A\} \,|\, \{X_n = e\})$.

Definition 22. *Let $(E, E \times E, P)$ be a Markov network. A* target set *T is a subset of E such that $\mathbb{P}(t, T) = 1$ for all $t \in T$.*

Remark 23. Any subset of E may be chosen as a target set T by changing the Markov network. In fact, if

$$\widetilde{P}(e_1, e_2) = \begin{cases} P(e_1, e_2) & \text{if } e_1 \notin T, \\ \delta_{e_1}(e_2) & \text{if } e_1 \in T, \end{cases}$$

then $(E, E \times E, \widetilde{P})$ is a Markov network. Moreover, (9) trivially holds (see Remark 20). In the framework of stochastic processes, $\widetilde{P}$ is the conditional probability of the stopped chain, where the stopping time is

$$\tau_T = \begin{cases} \inf\{n \in \mathbb{N} \colon X_n \in T\} & \text{if } \inf\{n \in \mathbb{N} \colon X_n \in T\} \neq \emptyset; \\ +\infty & \text{otherwise.} \end{cases}$$

The choice of T to be a target set will simplify the notation (see the definition of U_n in (10) compared with Lemma 25).

Remark 24. Any target set T of E identifies two extended sequences $(U_n, T_n)_{n \in \mathbb{N} \cup \{0,\infty\}}$ of subsets of E:

$$U_0 = T_0 := T,$$

$$U_n := \operatorname{supp}(\mathbb{P}(\cdot, U_{n-1})) = \{e \in E \colon \exists e^* \in U_{n-1} \colon P(e, e^*) > 0\}, \tag{10}$$

$$T_n := U_n \setminus U_{n-1}, \tag{11}$$

$$T_\infty := E \setminus \bigcup_{n=0}^{\infty} U_n = E \setminus \bigcup_{n=0}^{\infty} T_n. \tag{12}$$

The last equality may not be obvious and it will be proven below (see Lemma 25), as a consequence of monotonicity of U_n's. Note that U_n is the set of states that can reach T with positive probability in fewer than $n + 1$ jumps (there exists a path of positive probability of length at most n). Moreover, T_n is the set of states that can reach T with positive probability in n jumps and not less (the shortest path of positive probability has length n). Thus, a state $t \in T_n$ may jump only to T_{n-1} or to $E \setminus U_{n-1}$. Finally, T_∞ is the set of states that cannot reach T; it is an absorbing set. The following Lemma 25 summarizes these statements. To compare with [12], T_n may be seen as the "parking places at distance n from our destinations" (in [12], the parking places at distance n are the sites that can reach the destination T in n jumps and not less). This concept arises in a very natural way.

Lemma 25. *Let $(E, E \times E, P)$ be a Markov network and T be a target set. For any $n \in \mathbb{N}$, $U_n \subseteq U_{n+1}$. Moreover, we have*

$$\mathbb{P}(t, T_n) = 0 \quad \Leftarrow \quad t \in \bigcup_{i \geq 2} T_{n+i} \cup T_\infty, \tag{13}$$

$$\mathbb{P}(t, T_n) > 0 \quad \Leftarrow \quad t \in T_{n+1}, \tag{14}$$

$$\mathbb{P}(t, T_\infty) = 1 \quad \Leftrightarrow \quad t \in T_\infty. \tag{15}$$

Proof. We prove the assertion by induction on n. For $n=0$, since $\mathbb{P}(t, U_0) = \mathbb{P}(t, T_0) = 1$ for any $t \in U_0 = T_0$, we have $U_0 \subseteq \operatorname{supp}(\mathbb{P}(\cdot, U_0)) = U_1$. For the induction step, since $U_{n-1} \subseteq U_n$ (and hence $\mathbb{P}(\cdot, U_{n-1}) \leq \mathbb{P}(\cdot, U_n)$), we have $\operatorname{supp}(\mathbb{P}(\cdot, U_{n-1})) \subseteq \operatorname{supp}(\mathbb{P}(\cdot, U_n))$.

(13): $T_n \subseteq U_n$ implies $\mathbb{P}(t, T_n) = 0$ if $t \in E \setminus U_{n+1}$, by definition of U_{n+1}. The statement is a consequence of the previous result: $T_{n+1+k} = U_{n+1+k} \setminus U_{n+k} \subseteq E \setminus U_{n+1}$, $\forall k \geq 1$.

(14): By (13), $\mathbb{P}(t, U_{n-1}) = 0$. Moreover, $T_{n+1} \subseteq U_{n+1}$ implies $t \in U_{n+1}$. By definition of U_{n+1}, $\mathbb{P}(t, U_n) > 0$. Then $\mathbb{P}(t, T_n) = \mathbb{P}(t, U_n \setminus U_{n-1}) > 0$.

(15) $\Rightarrow$: By contradiction, suppose that $\exists t \in U_n \colon \mathbb{P}((t, T_\infty)) = 1$. Since $U_{n-1} \subseteq E \setminus T_\infty$, it follows that $\mathbb{P}(t, U_{n-1}) = 0$, which is a contradiction.

$\Leftarrow$: By contradiction, assume that $\exists t \in T_\infty, t^* \in E \setminus T_\infty \colon P(t, t^*) > 0$. But $E \setminus T_\infty = \bigcup_{n \geq 0} U_n$, so $\exists n \geq 0 \colon t^* \in U_n$. By definition, $t \in U_{n+1}$, which is a contradiction. □

Remark 26. If $T_n = \emptyset$, then $U_n = U_{n-1}$ and hence $U_{n+i} = U_n$ for any $i \geq 0$. Thus, since $E \neq \emptyset$, we have $T \cup T_\infty \neq \emptyset$. In fact, if $T = \emptyset$, then $U_n = \emptyset, \forall n \geq 0$, and hence $T_\infty = E$. Moreover, $T \cup T_\infty$ can always be mapped to at most two states, as the following lemma shows.

Lemma 27. *Let $(E, E \times E, P)$ be a Markov network, T be a target set, T_∞ as in* (12), *and t and t_∞ be extra points. Define*

$$F := \begin{cases} \{t_\infty\} & \text{if } T = \emptyset, \\ (E \setminus T) \cup \{t\} & \text{if } T \neq \emptyset \text{ and } T_\infty = \emptyset, \\ (E \setminus (T \cup T_\infty)) \cup \{t, t_\infty\} & \text{if } T \neq \emptyset \text{ and } T_\infty \neq \emptyset. \end{cases}$$

Then there exists a function π such that (9) *holds.*

Proof. We will give the proof for the case $T \neq \emptyset$ and $T_\infty \neq \emptyset$, since the first case is trivial ($\mathcal{P}^X_{E,\delta_e,\emptyset}(n) = 0$ for any $n \in \mathbb{N}$ and $e \in E$) and the second case is a special case of the third one.

Let $\pi : E \to F$ be the “trivial” projection:

$$\pi(e) = \begin{cases} e & \text{if } e \in E \setminus (T \cup T_\infty), \\ t_\infty & \text{if } e \in T_\infty, \\ t & \text{if } e \in T. \end{cases} \tag{16}$$

Note that, by Remark 26, π is well defined also for the first two cases.

On $F \times F$, we consider the function $P_\pi : F \times F \to \mathbb{R}_+ \cup \{0\}$,

$$P_\pi(f_1, f_2) := \begin{cases} \mathbb{P}(f_1, \pi^{-1}(f_2)) & \text{if } f_1 \notin \{t, t_\infty\}, \\ \delta_{f_1}(f_2) & \text{if } f_1 \in \{t, t_\infty\}. \end{cases} \tag{17}$$

Note that P_π is well defined, since $\pi : E \setminus (T \cup T_\infty) \to E \setminus (T \cup T_\infty)$ is the identity function. Thus, $(F, F \times F, P_\pi)$ is a Markov network and we need to prove that $\pi(X)$ has transition matrix P_π to prove the assertion. We must therefore check that, for any $n \in \mathbb{N}$ and any $(f_0, \ldots, f_n) \in F^{n+1}$,

$$\begin{aligned} &\mathcal{P}(\{\pi(X_0) = f_0, \pi(X_1) = f_1, \ldots, \pi(X_{n-1}) = f_{n-1}, \pi(X_n) = f_n\}) \\ &\qquad = \mathcal{P}(\{\pi(X_0) = f_0\}) \prod_{i=1}^{n} P_\pi(f_{i-1}, f_i). \end{aligned} \tag{18}$$

We prove it by induction on n. We will use (without citing them) the obvious facts that $a \in T \Leftrightarrow \pi(a) = t$ and $a \in T_\infty \Leftrightarrow \pi(a) = t_\infty$ (consequences of the fact that $T = \pi^{-1}(\pi(T))$).

For $n = 0$, the statement is obvious. For the induction step, the case of $\mathcal{P}(\{\pi(X_0) = f_0, \pi(X_1) = f_1, \ldots, \pi(X_{n-1}) = f_{n-1}\}) = 0$ is trivial, hence we deal with $\mathcal{P}(\{\pi(X_0) = f_0, \pi(X_1) = f_1, \ldots, \pi(X_{n-1}) = f_{n-1}\}) > 0$. We may have three cases.

Case A: $f_{n-1} \notin \{t, t_\infty\}$. Since $\pi(e) = e$ when $e \in E \setminus (T \cup T_\infty)$, it follows that $\pi(X_{n-1}) = f_{n-1} \Leftrightarrow X_{n-1} = f_{n-1}$, and hence

$$\begin{aligned} &\mathcal{P}(\{\pi(X_0) = f_0, \pi(X_1) = f_1, \ldots, \pi(X_{n-1}) = f_{n-1}, \pi(X_n) = f_n\}) \\ &\qquad = \mathcal{P}(\{\pi(X_0) = f_0, \ldots, X_{n-1} = f_{n-1}, \pi(X_n) = f_n\}) \\ &\qquad = \underbrace{\mathcal{P}(\{\pi(X_n) = f_n\} \mid \{X_0 \in \pi^{-1}(f_0), \ldots, X_{n-1} = f_{n-1}\})}_{a} \\ &\qquad\quad \cdot \underbrace{\mathcal{P}(\{X_0 \in \pi^{-1}(f_0), \ldots, X_{n-1} = f_{n-1}\})}_{b}. \end{aligned}$$

Since X is Markov, we have

$$\begin{aligned} a &= \mathcal{P}(\{\pi(X_n) = f_n\} \mid \{X_0 \in \pi^{-1}(f_0), \ldots, X_{n-1} = f_{n-1}\}) \\ &= \mathcal{P}(\{X_n \in \pi^{-1}(f_n)\} \mid \{X_{n-1} = f_{n-1}\}) = \mathbb{P}(f_{n-1}, \pi^{-1}(f_n)). \end{aligned}$$

Moreover, by the induction hypothesis,

$$\begin{aligned} b &= \mathcal{P}(\{\pi(X_0) = f_0, \pi(X_1) = f_1, \dots, X_{n-1} = f_{n-1}\}) \\ &= \mathcal{P}(\{\pi(X_0) = f_0\}) \prod_{i=1}^{n-1} P_\pi(f_{i-1}, f_i). \end{aligned}$$

Hence

$$\begin{aligned} &\mathcal{P}(\{\pi(X_0) = f_0, \pi(X_1) = f_1, \dots, \pi(X_{n-1}) = f_{n-1}, \pi(X_n) = f_n\}) \\ &\qquad = \mathbb{P}(f_{n-1}, \pi^{-1}(f_n))\mathcal{P}(\pi(X_0) = f_0) \prod_{i=1}^{n-1} P_\pi(f_{i-1}, f_i) \\ &\qquad = \mathcal{P}(\{\pi(X_0) = f_0\}) \prod_{i=1}^{n} P_\pi(f_{i-1}, f_i). \end{aligned}$$

Case B: $f_{n-1} = t$. We have $X_{n-1} \in T$. By Definition 22, $\{X_{n-1} \in T\} \subseteq \{X_n \in T\}$ a.s.

Now assume that $f_n = t$. Since $\{X_{n-1} \in T\} \subseteq \{X_n \in T\}$,

$$\begin{aligned} &\mathcal{P}(\{\pi(X_0) = f_0, \pi(X_1) = f_1, \dots, \pi(X_{n-1}) = f_{n-1}, \pi(X_n) = f_n\}) \\ &\qquad = \mathcal{P}(\{\pi(X_0) = f_0, \pi(X_1) = f_1, \dots, X_{n-1} \in T, X_n \in T\}) \\ &\qquad = \mathcal{P}(\{\pi(X_0) = f_0, \pi(X_1) = f_1, \dots, X_{n-1} \in T\}) \\ &\qquad = \mathcal{P}(\{\pi(X_0) = f_0\}) \prod_{i=1}^{n-1} P_\pi(f_{i-1}, f_i), \end{aligned}$$

where the last equality is a consequence of the induction hypothesis. By (17), $P_\pi(t, t) = 1$, and hence (18) holds.

Next, assume that $f_n \neq t$. Since $\{X_{n-1} \in T\} \subseteq \{X_n \in T\}$,

$$\mathcal{P}(\{\pi(X_0) = f_0, \pi(X_1) = f_1, \dots, \pi(X_{n-1}) = f_{n-1}, \pi(X_n) = f_n\}) = 0.$$

By (17), $P_\pi(t, f) = 0, \forall f \neq t$, and hence (18) trivially holds.

Case C: $f_{n-1} = t_\infty$. By (15), $\{X_{n-1} \in T_\infty\} \subseteq \{X_n \in T_\infty\}$ a.s. The proof is the same as above, after replacing t and T with t_∞ and T_∞, respectively.

Thus $(\pi(E), \pi(E) \times \pi(E), P_\pi)$ is a Markov network associated to the Markov chain $\pi(X)$ on $\pi(E)$. The equality (9) is now obvious, since

$$\bigcup_{i=0}^{n} \{\omega \in \Omega \colon X_i(\omega) \in T\} = \bigcup_{i=0}^{n} \{\omega \in \Omega \colon \pi(X_i)(\omega) \in \pi(T)\}$$

and hence

$$\mathcal{P}^X_{E,\delta_e,T}(n) = \left\{ \begin{array}{ll} 1 & \text{if } e \in T \\ 0 & \text{if } e \in T_\infty \\ \mathcal{P}^X_{E,\pi(\delta_e),T}(n) & \text{if } e \notin T \cup T_\infty \end{array} \right\} = \mathcal{P}^{\pi(X)}_{E,\pi(\delta_e),\pi(T)}(n). \qquad \square$$

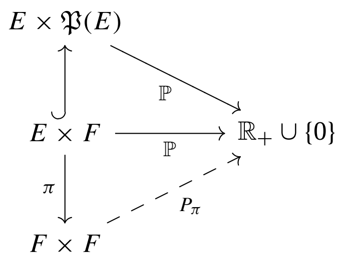
Corollary 28. *With the same hypotheses as in Theorem* 17*, if we are only interested in the probability of reaching the target set $\mathcal{A}$, then the cardinality of E' may be reduced at least to $\mu(\overrightarrow{\mathcal{A} \cup E}) - \mu(\mathcal{A})$.*

The proof of Lemma 27 suggests the following definition, which will be a characterization of π as given in (9) for the network framework.

Definition 29. *Let $(E, E \times E, P)$ be a Markov network, T be a target set and F be a nonempty set. A function $\pi = \pi^T : E \to F$ is called a* projection *if the following two properties hold:*

- *π is surjective: $F = \pi(E)$;*
- *$T = \pi^{-1}(\pi(T))$.*

A projection $\pi : E \to F$ is said to be compatible with respect to P *if there exists $P_\pi : F \times F \to \mathbb{R}_+ \cup \{0\}$ such that the following diagram commutes:*



(19)

Remarks 30. • A projection π^T divides the target set T from the rest of the states. This is the second characteristic in the projection's definition, and will ensure (9).

- Note that F may be embedded in $\mathfrak{P}(E)$ via π^{-1}. For simplicity, we have denoted this embedding by $F \hookrightarrow \mathfrak{P}(E)$ and we have considered $\mathbb{P}$ to be defined also on $E \times F$: $\mathbb{P}(e, f) := \mathbb{P}(e, \pi^{-1}(f))$.
- A compatible projection π is a projection such that $\mathbb{P}$ is well defined on the quotient set F: for every $f \in F$, if $\pi(e_1) = \pi(e_2)$, then $\mathbb{P}(e_1, \pi^{-1}(f)) = \mathbb{P}(e_2, \pi^{-1}(f))$ (this will ensure that $\pi(X)$ is a Markov chain if X is a Markov chain on E with transition matrix P).
- The function $\pi : E \to \pi(E)$ defined in (16) is a compatible projection (also called the *trivial projection*), where P_π that makes the diagram commute is just that of the proof of Lemma 27.
- A compatible projection divides the sets T_n, which pass to the quotient, as the following proposition states.

Proposition 31. *Let $(E, E \times E, P)$ be a Markov network, T a target set and $\pi^T : E \to F$ a compatible projection. Then, for any $n \in \mathbb{N} \cup \{0\}$, $U_n = \pi^{-1}(\pi(U_n))$, and hence $T_n = \pi^{-1}(\pi(T_n))$. Moreover, $\pi(U_n) = (\pi(U_0))_n$, and hence $\pi(T_n) = (\pi(T))_n$.*

Proof. We prove the first part by induction on n. For $n = 0$, by definition of projection, $U_0 = T = \pi^{-1}(\pi(T)) = \pi^{-1}(\pi(U_0))$. For the induction step, by contradiction, suppose that there exists $e \in E \setminus U_{n+1}$ such that $\pi(e) \in \pi(U_{n+1})$. Hence, there exists $e^* \in U_{n+1}$ such that $\pi(e) = \pi(e^*) =: f$. But then we have

$$\begin{aligned} 0 &= \mathbb{P}(e, U_n) = \mathbb{P}(e, \pi^{-1}(\pi(U_n))) = \sum_{g \in \pi(U_n)} \mathbb{P}(e, \pi^{-1}(g)) \\ &= \sum_{g \in \pi(U_n)} P_\pi(\pi(e), g) = \sum_{g \in \pi(U_n)} P_\pi(f, g) = \sum_{g \in \pi(U_n)} P_\pi(\pi(e^*), g) \\ &= \sum_{g \in \pi(U_n)} \mathbb{P}(e^*, \pi^{-1}(g)) = \mathbb{P}(e^*, \pi^{-1}(\pi(U_n))) = \mathbb{P}(e^*, U_n) > 0. \end{aligned}$$

The fact that $T_n = \pi^{-1}(\pi(T_n))$ is now trivial. We prove the second part also by induction on n. For $n = 0$, by definition of $U_0 := T$, $(\pi(U_0))_0 = \pi(U_0)$. For the induction step, by (10),

$$\begin{aligned} (\pi(U_0))_{n+1} &= \{f \in F : \exists f^* \in (\pi(U_0))_n : P_\pi(f, f^*) > 0\} \\ &= \{f \in F : \exists f^* \in \pi(U_n) : P_\pi(f, f^*) > 0\} \\ &= \{f \in F : \exists e \in \pi^{-1}(f), \exists e^* \in \pi^{-1}(\pi(U_n)) : P(e, e^*) > 0\} \\ &= \pi(\{e \in E : \exists e^* \in U_n : P(e, e^*) > 0\}) = \pi(U_{n+1}). \end{aligned}$$

□

Remark 32 (Back to the examples)**.** By Proposition 31, Example 2 in Section 2 cannot be projected on Markov processes with fewer states, since each state corresponds to a different T_n.

The following theorem relates the projections on Markov chains and networks.

Theorem 33. *Let X be a Markov chain on E and let $\pi : E \to F$ be a surjective function. Then $\pi(X)$ is a Markov chain on F iff* (19) *holds.*

Proof. $\Rightarrow$: Let $Y = \pi(X)$. Since Y is a Markov process on F, there exists a "usual" transition matrix Q. By contradiction, suppose there exist $e_1, e_2 \in E$ and $f \in F$ such that $\pi(e_1) = \pi(e_2) = f^*$ and $\mathbb{P}(e_1, \pi^{-1}(f)) \neq \mathbb{P}(e_2, \pi^{-1}(f))$.

This contradicts the fact that Y is a Markov process: take the two initial distributions δ_{e_i}, $i = 1, 2$ (i.e. $X_0 = e_i$ a.s., $i = 1, 2$). For both starting points, we have $Y_0 = f^*$ a.s. Since Y is a Markov chain, $\mathcal{P}(\{Y_0 = f^*, Y_1 = f\} \mid \{Y_0 = f^*\}) = Q_{f^*,f}$, which does not depend on $i = 1, 2$.

$\Leftarrow$: First, we have to prove that $P_\pi(f_1, f_2) \geq 0$, $\forall f_1, f_2 \in F$. This is a consequence of the fact that $P_\pi(f_1, f_2) = \mathbb{P}(e_1, \pi^{-1}(f_2))$, where $e_1 \in \pi^{-1}(f_1)$ (the existence of such an e_1 is a consequence of the fact that π is surjective). Second, we prove that $\sum_{f \in F} P_\pi(f_1, f) = 1$, $\forall f_1 \in F$: since $\pi^{-1}(f_i) \cap \pi^{-1}(f_j) = \emptyset$, $\forall f_i \neq f_j$ and $E = \bigcup_{f \in F} \pi^{-1}(f)$, we have

$$\sum_{f \in F} P_\pi(f_1, f) \overset{e_1 \in \pi^{-1}(f_1)}{=} \sum_{f \in F} \mathbb{P}(e_1, \pi^{-1}(f)) = \mathbb{P}(e_1, T) = 1.$$

We must check that, for any $n \in \mathbb{N}$ and any $(f_0, \dots, f_n) \in F^{n+1}$,

$$\mathcal{P}(\{\pi(X_0) = f_0, \dots, \pi(X_n) = f_n\}) = \mathcal{P}(\{\pi(X_0) = f_0\}) \prod_{i=1}^{n} P_\pi(f_{i-1}, f_i).$$

We prove this by induction on n. For $n = 0$, it is obvious. For the induction step,

$$\begin{aligned}
\mathcal{P}(\{\pi(X_0) &= f_0, \dots, \pi(X_n) = f_n\}) \\
&= \mathcal{P}(\{X_0 \in \pi^{-1}(f_0), \dots, X_n \in \pi^{-1}(f_n)\}) \\
&= \mathcal{P}\Big(\bigcup_{\substack{(x_0,\dots,x_n)\in \\ \pi^{-1}(f_0)\times\cdots\times\pi^{-1}(f_n)}} \{X_0 = x_0, \dots, X_n = x_n\} \Big) \\
&= \sum_{\substack{(x_0,\dots,x_n)\in \\ \pi^{-1}(f_0)\times\cdots\times\pi^{-1}(f_n)}} \mathcal{P}(\{X_0 = x_0, \dots, X_n = x_n\}) \\
&= \sum_{\substack{(x_0,\dots,x_{n-1})\in \\ \pi^{-1}(f_0)\times\cdots\times\pi^{-1}(f_{n-1})}} \sum_{x_n \in \pi^{-1}(f_n)} \mathcal{P}(\{X_0 = x_0, \dots, X_n = x_n\}). \qquad (20)
\end{aligned}$$

Since X is Markov, it follows that

$$\begin{aligned}
\sum_{x_n \in \pi^{-1}(f_n)} \mathcal{P}(\{X_0 = x_0, \dots, X_n = x_n\}) &= \sum_{x_n \in \pi^{-1}(f_n)} \mathcal{P}(\{X_0 = x_0\}) \prod_{i=1}^{n} P(x_{i-1}, x_i) \\
&= \mathcal{P}(\{X_0 = x_0\}) \prod_{i=1}^{n-1} P(x_{i-1}, x_i) \cdot \sum_{x_n \in \pi^{-1}(f_n)} P(x_{n-1}, x_n) \\
&= \mathcal{P}(\{X_0 = x_0\}) \prod_{i=1}^{n-1} P(x_{i-1}, x_i) \cdot \mathbb{P}(x_{n-1}, \pi^{-1}(f_n)) \\
&= \mathcal{P}(\{X_0 = x_0\}) \prod_{i=1}^{n-1} P(x_{i-1}, x_i) \cdot P_\pi(f_{n-1}, f_n), \qquad (21)
\end{aligned}$$

the last equation being true by (19) since $x_{n-1} \in \pi^{-1}(f_{n-1})$. Now, if we substitute (21) in (20), we obtain the assertion by the induction hypothesis:

$$\begin{aligned}
\mathcal{P}(\{\pi(X_0) &= f_0, \dots, \pi(X_n) = f_n\}) \\
&= P_\pi(f_{n-1}, f_n) \sum_{\substack{(x_0,\dots,x_{n-1})\in \\ \pi^{-1}(f_0)\times\cdots\times\pi^{-1}(f_{n-1})}} \mathcal{P}(\{X_0 = x_0\}) \prod_{i=1}^{n-1} P(x_{i-1}, x_i) \\
&= P_\pi(f_{n-1}, f_n) \cdot \mathcal{P}(\{\pi(X_0) = f_0, \dots, \pi(X_{n-1}) = f_{n-1}\}) \\
&= \mathcal{P}(\{X_0 = x_0\}) \prod_{i=1}^{n} P_\pi(f_{i-1}, f_i). \qquad \square
\end{aligned}$$

Theorem 34. *Let $(E, E \times E, P)$ be a Markov network and T be a target set. Then $\pi = \pi^T : E \to F$ is a compatible projection iff (9) holds.*

Proof. We will make use of the following trivial fact:

$$(a \in T \Leftrightarrow \pi(a) \in \pi(T)) \ \Leftrightarrow \ (T = \pi^{-1}(\pi(T))). \tag{22}$$

$\Leftarrow$: We must prove that $T = \pi^{-1}(\pi(T))$. Since

$$\delta_T(e) = \mathcal{P}^X_{E,\delta_e,T}(0) = 1 \quad \Leftrightarrow \quad e \in T,$$
$$\delta_{\pi(T)}(\pi(e)) = \mathcal{P}^Y_{F,\delta_{\pi(e)},\pi(T)}(0) = 1 \quad \Leftrightarrow \quad \pi(e) \in \pi(T),$$

it follows that (9) implies that $e \in T \Leftrightarrow \pi(e) \in \pi(T)$. By (22), we have the assertion.

$\Rightarrow$: We prove (9) by induction on n. For $n = 0$, by (22),

$$\mathcal{P}^X_{E,\delta_e,T}(0) = \delta_T(e) \overset{(22)}{=} \delta_{\pi(T)}(\pi(e)) = \mathcal{P}^Y_{F,\delta_{\pi(e)},\pi(T)}(0).$$

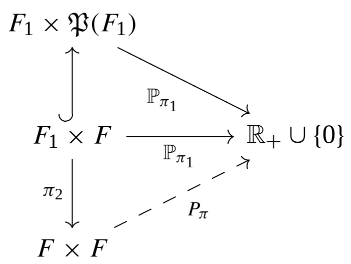
For the induction step, by Theorem 33, X and $Y = \pi(X)$ are both Markov chains, and hence

$$\begin{aligned}
\mathcal{P}^X_{E,\delta_e,T}(n+1) &= \sum_{e^* \in E} P(e, e^*)\, \mathcal{P}^X_{E,\delta_{e^*},T}(n) = \sum_{f \in F} \sum_{e_f \in \pi^{-1}(f)} P(e, e_f)\, \mathcal{P}^X_{E,\delta_{e^*},T}(n) \\
&= \sum_{f \in F} \mathcal{P}^Y_{F,f,\pi(T)}(n) \sum_{e_f \in \pi^{-1}(f)} P(e, e_f) \qquad \text{(by induction)} \\
&= \sum_{f \in F} \mathcal{P}^Y_{F,f,\pi(T)}(n)\, \mathbb{P}(e, \pi^{-1}(f)) \\
&= \sum_{f \in F} P_\pi(\pi(e), f)\, \mathcal{P}^Y_{F,f,\pi(T)}(n) \qquad \text{(by (16))} \\
&= \mathcal{P}^Y_{F,\pi(\delta_e),\pi(T)}(n+1).
\end{aligned}$$

□

Lemma 35. *Let $\pi_1 = \pi_1^T : E \to F_1$ be a compatible projection with respect to P and $\pi_2 : F_1 \to F$ be a function. Then $\pi^T := \pi_1 \circ \pi_2 : E \to F$ is a compatible projection with respect to P iff $\pi_2 = \pi_2^{\pi_1^T(T)}$ is a compatible projection with respect to $P_{\pi_1^T}$.*

Proof. $\Rightarrow$: Since π is surjective, π_2 is trivially surjective. Moreover, $T = \pi^{-1}(\pi(T))$ means $T = \pi_1^{-1}(\pi_2^{-1}(\pi_2(\pi_1(T))))$, which implies $\pi_1(T) = \pi_2^{-1}(\pi_2(\pi_1(T)))$. Note that the following diagram commutes:



$\Leftarrow$: This is just a consequence of the compatible projection's definition. □

Remark 36. We may say that two compatible projections $\pi_1 = \pi_1^T : E \to F_1$ and $\pi_2 : E \to F_2$ are equivalent if there exists a compatible bijection $\pi : F_1 \to F_2$. Then there exists a natural partial order $\preccurlyeq$ on the set of compatible projections: we say that $\pi : E \to F \preccurlyeq \pi' : E \to F'$ iff there exists a compatible projection $\pi^* : F' \to F$ such that $\pi = \pi^* \circ \pi'$.

Note that a minimal set F for Problem 19 must be minimum in the set of compatible projections. The following part will ensure that for any Markov network on E and any target set T, there exists "the" minimal set F on which the Markov network can be projected.

We denote by $\widetilde{E}$ the set of all equivalence relations on E. We introduce a partial order $\vDash$ on $\widetilde{E}$. Let $R, S \in \widetilde{E}$. We say that $R \vDash S$ if $e_1 \, R \, e_2$ implies $e_1 \, S \, e_2$ (if you think E as the set of all men and R is "belonging to the same state" while S is "belonging to the same continent", then $R \vDash S$). The relation $\vDash$ is just set-theoretic inclusion between equivalence relations, since any relation is a subset of $E \times E$.

Lemma 37. *Let $K \subseteq \widetilde{E}$ be fixed. Then there exists $S_K \in \widetilde{E}$ such that $R \vDash S_K$ for all $R \in K$ and if $S' \in \widetilde{E}$ is such that $R \vDash S'$ for all $R \in K$, then $S_K \vDash S'$ (S_K is called* the least majorant *of K).*

Proof. First, note that the trivial relation $T = E \times E$ (defined by $e_1 \, T \, e_2$ for all $(e_1, e_2) \in E \times E$) is a majorant of K. Set

$$S_K = \bigcap \{S \in \widetilde{E} : S \supseteq R \text{ for all } R \in K\}.$$

Then S_K is trivially an equivalence such that $R \vDash S_K, \forall R \in K$, and if $S' \in \widetilde{E}$ is such that $R \vDash S', \forall R \in K$, then $S_K \vDash S'$. □

For any set $E' \subseteq E$, let $R_{E'}$ be the equivalence relation in E corresponding to the partition of E into E' and $E \setminus E'$.

Lemma 38. *Let $(E, E \times E, P)$ be a Markov network, T be a target set, and let $\mathcal{K} = \{\pi_\alpha = \pi_\alpha^T : E \to F_\alpha, \alpha \in A\}$ be a family of compatible projections with respect to the same P. Then there exists a family $J_{\mathcal{K}} = \{\pi'_\alpha = {\pi'_\alpha}^{\pi_\alpha(T)} : F_\alpha \to F, \alpha \in A\}$ of compatible projections with respect to the induced P_{π_α} such that $\pi = \pi'_\alpha \circ \pi_\alpha$ does not depend on α.*

Proof. First, note that any projection π induces trivially an equivalence relation R_π on $E \times E$ given by $e_1 \, R_\pi \, e_2 \Leftrightarrow \pi(e_1) = \pi(e_2)$. We set $K = \{R_{\pi_\alpha}, \alpha \in A\}$ and denote by S_K the least majorant of K defined in Lemma 37. Let $F := E/S_K$ and π be the canonical projection of E onto E/S_K. We want to prove that π is a compatible projection.

First, let us prove that $T = \pi^{-1}(\pi(T))$. Since $T = \pi_\alpha^{-1}(\pi_\alpha(T)), \forall \alpha$, it is sufficient to note that $R_{\pi_\alpha} \vDash R_T, \forall \alpha$ (R_T is the partition of E into T and $E \setminus T$), and hence $S_K \vDash R_T$.

Second, (19) also holds. Let $e \in E$ be fixed. We want to characterize the set $\pi^{-1}(\pi(e))$. Let $\boldsymbol{\alpha} \in A^n$ be a vector of $A^n = \underbrace{A \times \cdots \times A}_{n \text{ times}}$. We define:

$$E^e_{\boldsymbol{\alpha},n} = \{f \in E : \exists (e = e_0, e_1, \ldots, e_n = f) \in E^n : e_{i-1} \, R_{\pi_{\alpha_i}} \, e_i, i = 1, \ldots, n\},$$

$$E_n^e = \bigcup_{\boldsymbol{\alpha} \in A^n} E_{\boldsymbol{\alpha},n}^e, \qquad E^e = \bigcup_{n \in \mathbb{N}} E_n^e.$$

Now, we are going to prove that $S_K \vDash R_{E^e}$ and $E^e = \pi^{-1}(\pi(e))$.

For $S_K \vDash R_{E^e}$, we will prove that, for any α, $R_{\pi_\alpha} \vDash R_{E^e}$. By contradiction, suppose there exist $e_1 \in E^e$, $e_2 \in E \setminus E^2$ and $\alpha \in A$ such that $e_1 \, R_{\pi_\alpha} \, e_2$. Since $e_1 \in E^e$ we have $e_1 \in E_n^e$ for some n, say n_0, and therefore there is $\boldsymbol{\alpha} \in A^{n_0}$ such that $e_1 \in E_{\boldsymbol{\alpha},n_0}^e$. Take $\boldsymbol{\alpha}' = (\boldsymbol{\alpha}, \alpha) \in A^{n_0+1}$. Then $e_2 \in E_{\boldsymbol{\alpha}',n_0+1}^e$, which is a contradiction.

For $E^e = \pi^{-1}(\pi(e))$, since $R_{\pi_\alpha} \vDash S_K$, we have $E^e \subseteq \pi^{-1}(\pi(e))$ (S_K is transitive), and $E^e \supseteq \pi^{-1}(\pi(e))$ is just a consequence of $S_K \vDash R_{E^e}$.

For any $e \in E$ and $f \in F$, let $p_{e,f} = \mathbb{P}(e, \pi^{-1}(f))$. We have to prove that $p_{e,f} = p_{e',f}$ if $\pi(e) = \pi(e')$ or, equivalently, if $e \, S_K \, e'$. Under this hypothesis, since $E^e = \pi^{-1}(\pi(e))$, there exists $\boldsymbol{\alpha} \in A^n$ such that $e' \in E_{\boldsymbol{\alpha},n}^e$, i.e.

$$\exists (e = e_0, e_1, \ldots, e_n = e') \in E^n : e_{i-1} \, R_{\pi_{\boldsymbol{\alpha}_i}} \, e_i, \; i = 1, \ldots n.$$

Note that $\pi^{-1}(f) \subseteq \pi_\alpha^{-1}(\mathfrak{P}(F_\alpha))$ for all α, since $R_{\pi_\alpha} \vDash S_K$ (and hence we have $\pi^{-1}(f) = \pi_\alpha^{-1}(\pi_\alpha(\pi^{-1}(f)))$, $\forall \alpha$). Therefore, for any $i = 1, \ldots, n$,

$$p_{e_{i-1},f} = \mathbb{P}_{\pi_{\boldsymbol{\alpha}_i}}(\pi_{\boldsymbol{\alpha}_i}(e_{i-1}), \pi_{\boldsymbol{\alpha}_i}(\pi^{-1}(f))) = \mathbb{P}_{\pi_{\boldsymbol{\alpha}_i}}(\pi_{\boldsymbol{\alpha}_i}(e_i), \pi_{\boldsymbol{\alpha}_i}(\pi^{-1}(f))) = p_{e_i,f},$$

which leads to the statement of the lemma: the existence of $J_{\mathcal{K}} = \{\pi'_\alpha, \alpha \in A\}$ is ensured by Lemma 35. □

Theorem 39. *Let $(E, E \times E, P)$ be a Markov network and T be a target set. There exists a minimal set F as discussed in Problem* 19*. If F' is another minimal set, there exists a bijection $\Phi : F \to F'$ which is a compatible projection.*

Proof. Let $K = \{\pi_\alpha = \pi_\alpha^T : E \to F_\alpha, \alpha \in A\}$ be the set of all compatible projections with respect to P. Then there exists a compatible projection $\pi = \pi^T : E \to F$ such that $\pi \preccurlyeq \pi_\alpha, \forall \alpha$, by Lemma 38. Hence F is a minimal set.

Now, suppose that F' is another minimal set (there exists $\pi' : E \to F'$ which satisfies the hypothesis in Problem 19). Note that $\pi' \in K$ and hence $\pi \preccurlyeq \pi'$. Conversely, since π' is minimal, $\pi' \preccurlyeq \pi$ by Lemma 35, and the result follows by Remark 36. □

5. Translation into category theory

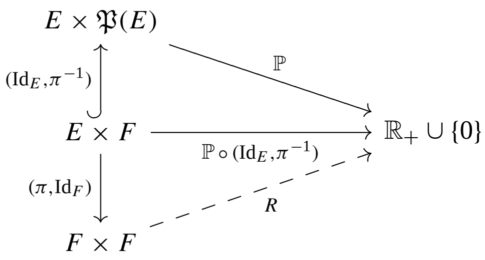
What we have obtained in the previous section may be reread in terms of categories (see [9]).

Let $\mathcal{O}$ be the objects given by the triples (E, T, P), where:

- E is an at most countable set;
- T is a subset of E;
- $P : E \times E \to \mathbb{R}_+$ with the following properties:
  - $\forall e \in E, \sum P(e, \cdot) = 1$;
  - $P^{-1}(0, \infty) \cap (T \times E) \subseteq T \times T$.

Let $\mathcal{A}$ be the morphisms given by the maps $(E, T, P) \xrightarrow{\pi} (F, S, R)$, where:

- π is a surjective set function from E to F;
- $S = \pi(T)$, $T = \pi^{-1}(S)$;
- the following diagram commutes:



where $\mathrm{Id}_E : E \to E$ is the identity map on E, $\mathfrak{P}(E)$ is the power set of E and $\mathbb{P}(e, A) = \sum_{e_i \in A} P((e, e_i))$.

Obviously we have the two functions

$$\mathcal{A} \underset{\mathrm{cod}}{\overset{\mathrm{dom}}{\rightrightarrows}} \mathcal{O}$$

where, if $f := (E, T, P) \xrightarrow{\pi} (F, S, R) \in \mathcal{A}$, then $\mathrm{dom}(f) = (E, T, P)$ and $\mathrm{cod}(f) = (F, S, R)$. The couple $(\mathcal{O}, \mathcal{A})$ is hence a graph. Let $\mathcal{A} \times_{\mathcal{O}} \mathcal{A}$ be the set of composable morphisms:

$$\mathcal{A} \times_{\mathcal{O}} \mathcal{A} = \{(g, f) \colon g, f \in \mathcal{A} \text{ and } \mathrm{dom}(g) = \mathrm{cod}(f)\}.$$

Note that the identity set function Id_E defines a morphism

$$\mathrm{id}_{(E,T,P)} := (E, T, P) \xrightarrow{\mathrm{Id}_E} (E, T, P)$$

which is trivially in $\mathcal{A}$, and hence we may define the function $\mathcal{O} \xrightarrow{\mathrm{id}} \mathcal{A}$ such that

$$\mathrm{dom}(\mathrm{id}_{(E,T,P)}) = (E, T, P) = \mathrm{cod}(\mathrm{id}_{(E,T,P)}).$$

Moreover, Lemma 35 states that there exists a function $\mathcal{A} \times_{\mathcal{O}} \mathcal{A} \xrightarrow{\circ} \mathcal{A}$ where, if $f := (E, T, P) \xrightarrow{\pi_1} (F_1, S_1, R_1) \in \mathcal{A}$ and $g := (F_1, S_1, R_1) \xrightarrow{\pi_2} (F, S, R) \in \mathcal{A}$, we have

$$f \circ g := (E, T, P) \xrightarrow{\pi_2 \circ \pi_1} (F, S, R).$$

Since $\pi_3 \circ (\pi_2 \circ \pi_1) = (\pi_3 \circ \pi_2) \circ \pi_1$, it is straightforward to prove that $\mathcal{A} \times_{\mathcal{O}} \mathcal{A} \xrightarrow{\circ} \mathcal{A}$ is associative. For any $f, g \in \mathcal{A}$ with $\mathrm{dom}(g) = \mathrm{cod}(f) = (E, T, P)$ we have

$$\mathrm{id}_{(E,T,P)} \circ f = f, \qquad g \circ \mathrm{id}_{(E,T,P)} = g,$$

and hence $(\mathcal{O}, \mathcal{A})$ is a category. We call it BIGNET.

Now, let PRENET $=$ BIGNET$/R$, where $R = R_{a,b}$ is the relation between the morphisms on BIGNET given by

$$(a \xrightarrow{f} b)\; R_{a,b}\; (a \xrightarrow{f'} b) \;\Leftrightarrow\; \exists (b \xrightarrow{g} b) \colon f = f' \circ g.$$

As a consequence of Lemma 38, it is straightforward to see that PRENET is a preorder.

Remark 36 states that PRENET may be reduced to the ordered universal category NET, by identifying the objects a, b where $\exists (a \xrightarrow{f} b), (b \xrightarrow{f'} a)$. The order $\preccurlyeq$ is given by $b \preccurlyeq a \Leftrightarrow \exists (a \xrightarrow{f} b)$. Theorem 39 states that NET has the least majorant property.

Acknowledgements. We wish to thank the referee for suggesting interesting examples. The first author also thanks Jerzy Trzeciak for his precious comments and remarks and for his kind assistance.

References

[1] Aho, A. V., Hopcroft, J. E., Ullman, J. D.: The Design and Analysis of Computer Algorithms. Addison-Wesley, Reading, MA (1975) Zbl 0326.68005 MR 0413592

[2] Aki, S., Hirano, K.: Sooner and later waiting time problems for runs in Markov dependent bivariate trials. Ann. Inst. Statist. Math. **51**, 17–29 (1999) Zbl 0952.60071 MR 1704644

[3] Antzoulakos, D. L., Philippou, A. N.: Probability distribution functions of succession quotas in the case of Markov dependent trials. Ann. Inst. Statist. Math. **49**, 531–539 (1997) Zbl 0947.60503 MR 1482371

[4] Arnold, B. C., Balakrishnan, N., Nagaraja, H. N.: Records. Wiley, New York (1998) Zbl 0914.60007 MR 1628157

[5] Cairoli, R., Dalang, R. C.: Sequential Stochastic Optimization. Wiley, New York (1996) Zbl 0856.62070 MR 1369770

[6] Ebneshahrashoob, M., Sobel, M.: Sooner and later waiting time problems for Bernoulli trials: frequency and run quotas. Statist. Probab. Lett. **9**, 5–11 (1990) Zbl 0695.60016 MR 1035984

[7] Koutras, M. V., Alexandrou, V. A.: Sooner waiting time problems in a sequence of trinary trials. J. Appl. Probab. **34**, 593–609 (1997) Zbl 0891.60020 MR 1464596

[8] Lam, K., Yam, H. C.: Cusum techniques for technical trading in financial markets. Financial Engineering & Japanese Markets **4**, 257–274 (1997)

[9] Mac Lane, S.: Categories for the Working Mathematician, Grad. Texts Math. 5, Springer, New York (1998) Zbl 0906.18001 MR 1712872

[10] Stefanov, V. T.: On some waiting time problems. J. Appl. Probab. **37**, 756–764 (2000) Zbl 0969.60021 MR 1782451

[11] Stefanov, V. T., Pakes, A. G.: Explicit distributional results in pattern formation. Ann. Appl. Probab. **7**, 666–678 (1997) Zbl 0893.60005 MR 1459265

[12] Yang, J.-R., Hsiau, S.-R.: On Markov-dependent parking problems. J. Appl. Probab. **41**, 579–586 (2004) Zbl pre02103388 MR 2052594