Analysis of a large number of Markov chains competing for transitions

Emmanuelle Anceaume, François Castella, Bruno Sericola

Analysis of a large number of Markov chains competing for transitions

Emmanuelle Anceaume*, IRISA/CNRS

François Castella*, Université de Rennes 1

Bruno Sericola*, INRIA Rennes - Bretagne Atlantique

January 21, 2011

Abstract We consider the behavior of a stochastic system composed of several identically distributed, but non independent, discrete-time absorbing Markov chains competing at each instant for a transition. The competition consists in determining at each instant, using a given probability distribution, the only Markov chain allowed to make a transition. We analyze the first time at which one of the Markov chains reaches its absorbing state. When the number of Markov chains goes to infinity, we analyze the asymptotic behavior of the system for an arbitrary probability mass function governing the competition. We give conditions for the existence of the asymptotic distribution and we show how these results apply to cluster-based distributed systems when the competition between the Markov chains is handled by using a geometric distribution.

Keywords. Asymptotic Analysis, Competing Markov Chains, Cluster-Based Distributed Systems, Markov Chains, Geometric Distribution.

*Campus de Beaulieu, 35042 Rennes Cedex, France

1 Introduction

Competing Markov chains generally compete over a set of resources, see for instance [2] and the references therein. The resulting process is then a multidimensional Markov chain based on the Cartesian product of the states spaces and on competition rules over resources. In this paper, the Markov chains do not compete for resources but for transitions. More precisely, we consider a stochastic system composed of n identically distributed, but non independent, discrete-time absorbing Markov chains competing at each instant for a transition. The competition consists in determining at each instant, using a given probability mass function of dimension n, the only Markov chain allowed to make a transition.

For this system, we analyze the first time Θ_n^γ at which one of the n Markov chains reaches its absorbing state, when the probability mass function is $\gamma(n)$. The distribution of this random variable has already been studied in [1] in particular when the probability mass function $\gamma(n)$ handling the competition is uniform. In that case, we exhibited the asymptotic behavior of the system when the number n of Markov chains goes to infinity and we applied these results to the analysis of large-scale distributed systems.

We propose here the study of the asymptotic behavior of the system when the number n of Markov chains goes to infinity, for an arbitrary probability mass function $\gamma(n)$ governing the competition. More precisely, we give conditions on probability mass function $\gamma(n)$ governing the competition for the existence of the limiting distribution of Θ_n^γ. We apply these results to the case where the competition is governed by a geometric distribution and we study the effects of this distribution on a model of a cluster-based systems distributed, when the number of clusters increases.

The remainder part of the paper is organized as follows. In the next section, we describe the model, the notation and we give the transient state distribution of the global Markov chain composed of the n joined identically distributed local Markov chains. We also extend a result obtained in [1]. This result is a recurrence relation that allows us not only to compute the distributions of Θ_n^γ but also to compute the limiting distribution, when it exists, of Θ_n^γ. In Section 3, we study the asymptotic behavior of the system when n goes to infinity and we give conditions on the probability mass function γ_n governing the competition for the existence of the limiting distribution of Θ_n^γ. We also show how to compute this limiting distribution. We apply these results in Section 4 to case where the probability mass function γ_n governing the competition is a geometric distribution. Section 5 is devoted to an application of these results to a model of a cluster-based distributed system.

2 Transient Analysis

We consider a homogeneous discrete-time Markov chain $X = \{X_k, k \geq 0\}$ with finite state space S composed of a set of transient states denoted by B and an absorbing state denoted by a. The transition probability matrix P of can thus be decomposed as

$$P = \begin{pmatrix} Q & v \\ 0 & 1 \end{pmatrix},$$

where Q is the submatrix of dimension $|B| \times |B|$ containing the transitions between states of B. In the same way, v is the column vector with $|B|$ entries representing the transitions from the transient states to the absorbing state. We suppose that the initial state is in B, i.e. $\mathbb{P}\{X_0 \in B\} = 1$, and we denote by α the row vector of dimension $|B|$ containing the initial probability distribution, i.e. for every $i \in B$, $\alpha_i = \mathbb{P}\{X_0 = i\}$. We denote by Θ_1 the total time spent in B before reaching the absorbing state or equivalently the first instant at which the absorbing state a is reached. We have

$$\Theta_1 = \inf\{k \geq 0 \mid X_k = a\}.$$

The complementary cumulative distribution function of Θ_1 is easily derived as, see for instance [4] or [3],

$$\mathbb{P}\{\Theta_1 > k\} = \mathbb{P}\{X_k \in B\} = \alpha Q^k \mathbb{1}, \tag{1}$$

where $\mathbb{1}$ is the column vector of dimension $|B|$ with all entries equal to 1 and I is the identity matrix of the right dimension. Since all the states of B are transient the matrix $I - Q$ is invertible and the expectation of Θ_1 is given by

$$E(\Theta_1) = \alpha(I - Q)^{-1}\mathbb{1}. \tag{2}$$

Let us now consider, for $n \geq 1$, n Markov chains denoted by $X^{(1)}, \ldots, X^{(n)}$ identical to X, i.e. with the same state space S, the same transition probability matrix P and the same initial probability distribution α. These n Markov chains are in competition at each instant to make a transition using the probability mass function $\gamma(n) = (p_{1,n}, \ldots, p_{n,n})$.

From these n Markov chains, we construct a new Markov chain denoted by $Y = \{Y_k, k \geq 0\}$ as follows. The state space of Y is equal to S^n and $Y_k = (X_k^{(1)}, \ldots, X_k^{(n)})$. A transition in the Markov chain Y corresponds to a transition in only one of the Markov chains $X^{(1)}, \ldots, X^{(n)}$, all the others staying in the same state. The Markov chain that makes the transition is chosen with the probability mass function $\gamma(n)$, which means that Markov chain $X^{(\ell)}$ makes the transition with probability $p_{\ell,n}$. We suppose without any loss of generality that, for every $\ell = 1, \ldots, n$, we have $0 < p_{\ell,n} < 1$.

The transition probability matrix of Y is detailed in [1] where we give the proof of the following theorem giving the transient distribution of the Markov chain Y. For every $k \geq 0$ and $\ell \geq 1$, we introduce the set $S_{k,\ell}$ defined by

$$S_{k,\ell} = \{\underline{k} = (k_1, \ldots, k_\ell) \in \mathbb{N}^\ell \mid k_1 + \cdots + k_\ell = k\}.$$

Theorem 1 *For every $k \geq 0$, $n \geq 1$ and $(j_1, \ldots, j_n) \in S^n$, we have*

$$\mathbb{P}\{Y_k = (j_1, \ldots, j_n)\} = \sum_{\underline{k} \in S_{k,n}} \frac{k!}{k_1! \cdots k_n!} \prod_{r=1}^{n} (p_{r,n})^{k_r} \mathbb{P}\{X_{k_r} = j_r\}. \tag{3}$$

The following corollary, which is proved in [1], provides the distribution of the first instant Θ_n^γ at which one of the n Markov chains $X^{(1)}, \ldots, X^{(n)}$ gets absorbed when the probability mass function is $\gamma(n)$. More formally, this instant denoted by Θ_n^γ is defined as

$$\Theta_n^\gamma = \inf\{k \geq 0 \mid \exists r \text{ such that } X_k^{(r)} = a\}.$$

When $n = 1$ we have $\gamma(1) = 1$ and, thus $\Theta_1^\gamma = \Theta_1$.

Corollary 2 *For every $k \geq 0$ and $n \geq 1$, we have*

$$\mathbb{P}\{\Theta_n^\gamma > k\} = \sum_{\underline{k} \in S_{k,n}} \frac{k!}{k_1! \cdots k_n!} \prod_{r=1}^{n} (p_{r,n})^{k_r} \alpha Q^{k_r} \mathbb{1}. \tag{4}$$

Clearly the complexity for the computation of $\mathbb{P}\{\Theta_n^\gamma > k\}$ using relation (4) is exponential. A solution to this problem is given by the following theorem which generalizes a previous result obtained in [1].

Theorem 3 *For every $k \geq 0$, $n \geq 2$ and $h = 1, \ldots, n$, we have*

$$\mathbb{P}\{\Theta_n^\gamma > k\} = \sum_{\ell=0}^{k} \binom{k}{\ell} (p_{h,n})^\ell (1 - p_{h,n})^{k-\ell} \alpha Q^\ell \mathbb{1} \mathbb{P}\{\Theta_{n-1}^{\gamma'} > k - \ell\}, \tag{5}$$

where the probability mass distribution $\gamma'(n-1) = (p'_{1,n-1}, \ldots, p'_{n-1,n-1})$ associated with $\Theta_{n-1}^{\gamma'}$ is defined, by

$$p'_{r,n-1} = \frac{p_{r,n}}{1 - p_{h,n}} \textit{ for } r = 1, \ldots, h-1 \textit{ and } p'_{r,n-1} = \frac{p_{r+1,n}}{1 - p_{h,n}} \textit{ for } r = h, \ldots, n-1.$$

Proof. For every $k \geq 0$ and $n \geq 2$, we fix a value of h with $1 \leq h \leq n$. In Relation 4, we extract in the multiple sum indexed by $\underline{k} \in S_{k,n}$ the index k_h, we rename it ℓ and next, if $h < n$, we perform the variable changes $k_{h+1} := k_h, \ldots, k_n := k_{n-1}$. We thus obtain

$$\mathbb{P}\{\Theta_n^\gamma > k\} = \sum_{\ell=0}^{k} \frac{(p_{h,n})^\ell \alpha Q^\ell \mathbb{1}}{\ell!} \sum_{\underline{k} \in S_{k-\ell,n-1}} \frac{k!}{k_1! \cdots k_{n-1}!} \prod_{r=1}^{h-1} (p_{r,n})^{k_r} \alpha Q^{k_r} \mathbb{1} \prod_{r=h}^{n-1} (p_{r+1,n})^{k_r} \alpha Q^{k_r} \mathbb{1}.$$

Multiplying and dividing respectively by $(k-\ell)!$ and $(1 - p_{h,n})^{k-\ell}$, we get

$$\begin{aligned} \mathbb{P}\{\Theta_n^\gamma > k\} &= \sum_{\ell=0}^{k} \binom{k}{\ell} (p_{h,n})^\ell (1 - p_{h,n})^{k-\ell} \alpha Q^\ell \mathbb{1} \\ &\quad \times \sum_{\underline{k} \in S_{k-\ell,n-1}} \frac{(k-\ell)!}{k_1! \cdots k_{n-1}!} \prod_{r=1}^{h-1} \left(\frac{p_{r,n}}{1 - p_{h,n}}\right)^{k_r} \alpha Q^{k_r} \mathbb{1} \prod_{r=h}^{n-1} \left(\frac{p_{r+1,n}}{1 - p_{h,n}}\right)^{k_r} \alpha Q^{k_r} \mathbb{1}. \end{aligned}$$

If the probability mass function $\gamma'(n-1) = (p'_{1,n-1}, \ldots, p'_{n-1,n-1})$ associated with $\Theta_{n-1}^{\gamma'}$ is defined by

$$p'_{r,n-1} = \frac{p_{r,n}}{1 - p_{h,n}} \text{ for } r = 1, \ldots, h-1 \text{ and } p'_{r,n-1} = \frac{p_{r+1,n}}{1 - p_{h,n}} \text{ for } r = h, \ldots, n-1,$$

we obtain, from Relation (4),

$$\begin{aligned}
\mathbb{P}\{\Theta_n^\gamma > k\} &= \sum_{\ell=0}^{k} \binom{k}{\ell} (p_{h,n})^\ell (1-p_{h,n})^{k-\ell} \alpha Q^\ell \mathbb{1} \sum_{\underline{k}\in S_{k-\ell,n-1}} \frac{(k-\ell)!}{k_1!\cdots k_{n-1}!} \prod_{r=1}^{n-1} (p'_{r,n-1})^{k_r} \alpha Q^{k_r} \mathbb{1} \\
&= \sum_{\ell=0}^{k} \binom{k}{\ell} (p_{h,n})^\ell (1-p_{h,n})^{k-\ell} \alpha Q^\ell \mathbb{1} \mathbb{P}\{\Theta_{n-1}^{\gamma'} > k-l\},
\end{aligned}$$

which completes the proof. ■

This result shows that the computation of $\mathbb{P}\{\Theta_n^\gamma > k\}$ can be done using a simple recurrence with a polynomial complexity. The expectation of Θ_n^γ is then obtained by

$$E(\Theta_n^\gamma) = \sum_{n=0}^{\infty} \mathbb{P}\{\Theta_n^\gamma > k\}. \tag{6}$$

3 Asymptotic Analysis

This section is devoted to the analysis of the distribution on Θ_n^γ when n is large. This is generally the case in practice for large-scale distributed systems which are studied in the last section. We consider the following transform. For every $n \geq 1$ and $x \in \mathbb{R}$, we introduce the function $F_n(x)$ defined by

$$F_n(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} \mathbb{P}\{\Theta_n^\gamma > k\}.$$

The function F_n is defined for every $x \in \mathbb{R}$ and an explicit expression is given in the following theorem, which is proved in [1].

Theorem 4 *For every $n \geq 1$ and $x \in \mathbb{R}$, we have*

$$F_n(x) = \prod_{\ell=1}^{n} \alpha e^{Qxp_{\ell,n}} \mathbb{1}, \tag{7}$$

and, for every $k \in \mathbb{N}$,

$$\mathbb{P}\{\Theta_n^\gamma > k\} = F_n^{(k)}(0), \tag{8}$$

where $F_n^{(k)}$ is the k-th derivative of function F_n with respect to x.

This result not only shows that $\mathbb{P}\{\Theta_n^\gamma > 0\} = 1$ as expected, but also that, for every $n \geq 1$, we have

$$\mathbb{P}\{\Theta_n^\gamma > 1\} = \alpha Q \mathbb{1} = F_n'(0).$$

It also gives access to an expression of $\mathbb{P}\{\Theta_n^\gamma > k\}$ for any k. Adopting this point of view, our strategy in order to compute $\lim_{n\to\infty} \mathbb{P}\{\Theta_n^\gamma > k\}$ is to compute $F(x) = \lim_{n\to\infty} F_n(x)$, an analytic function of x, so as to deduce the value $\lim_{n\to\infty} \mathbb{P}\{\Theta_n^\gamma > k\} = F^{(k)}(0)$.

In order to pass to the limit in a clean fashion, we need the following

Hypothesis (H) – Limiting value of the powers sums of the $p_{\ell,n}$'s.

For any $k \geq 1$, the following limit exists :

$$V_k := \lim_{n\to\infty} \sum_{\ell=1}^{n} p_{\ell,n}^k.$$

Important remark. *The above assumption is harmless. Indeed, introducing the quantities $V_{n,k} = \sum_{\ell=1}^{n} (p_{\ell,n})^k$, it is clear that $0 \leq V_{n,k} \leq 1$ for any value of $n \geq 1$ and $k \geq 1$. Therefore, there exists a subsequence in n, say n_j with $n_j \to \infty$ as $j \to \infty$, such that $V_{n_j,k}$ has a limit as $j \to \infty$ for any $k \geq 1$. We are here merely assuming that the limit V_k is well defined without refering to taking a subsequence in the original $V_{n,k}$'s. To give but an example, one may imagine for instance that the $p_{\ell,2n}$'s are uniformly distributed, i.e. $p_{\ell,2n} = 1/(2n)$, in which case $V_{2n,k} = 1/(2n)^{k-1} \to 0$ whenever $k \geq 2$, and $V_{2n,1} = 1$, while the $p_{\ell,2n+1}$'s are geometrically distributed with parameter b and truncation at step $2n+1$, i.e. $p_{\ell,2n+1} = (1-b)^{\ell-1}b$, for $\ell \leq 2n$ and $p_{2n+1,2n+1} = (1-b)^{2n}$, in which case $V_{2n+1,k} \to 1/(2^k - 1)$. In that case it clearly does not make sense to study the whole sequence $V_{n,k}$ itself, and we need to separate the case when n is odd and the case when n is even.*

We assert here that this situation is generic, and that, up to extracting a subsequence, one may always assume that the original sequence $V_{n,k}$ itself possesses a limit V_k for any k.

With this assumption in mind, the following theorem gives the limit of the transform $F_n(x)$ when n goes to infinity.

Theorem 5 *Under hypothesis (H), the limit $F(x) = \lim_{n\longrightarrow\infty} F_n(x)$ exists, whenever $|x| < \ln 2$, and the limit is uniform on compact subsets of $\{x \mid |x| < \ln 2\}$. Besides, we have the explicit value*

$$F(x) = \exp\left(\sum_{m\geq 1}\sum_{k_1\geq 1}\cdots\sum_{k_m\geq 1}\frac{(-1)^{m+1}}{m}\frac{\alpha Q^{k_1}\mathbb{1}\ldots\alpha Q^{k_m}\mathbb{1}}{k_1!\ldots k_m!}x^{k_1+\cdots+k_m}V_{k_1+\cdots+k_m}\right).$$

Proof. Starting from Relation (7), we recover, expanding into power series in x, the value

$$\ln\left(F_n(x)\right) = \sum_{\ell=1}^{n}\ln\left(\alpha\, e^{Qxp_{\ell,n}}\mathbb{1}\right) = \sum_{\ell=1}^{n}\ln\left(1+\sum_{k\geq 1}\frac{\alpha Q^k\mathbb{1}}{k!}x^k(p_{\ell,n})^k\right)$$

Hence, using the fact that $0\leq \alpha Q^k\mathbb{1}\leq 1$ whenever $k\geq 0$, and deducing the bound

$$\left|\sum_{k\geq 1}\frac{\alpha Q^k\mathbb{1}}{k!}x^k(p_{\ell,n})^k\right| \leq \sum_{k\geq 1}\frac{|x|^k}{k!} = e^{|x|}-1<1,$$

whenever $|x| < \ln 2$, we may expand further and obtain

$$\begin{aligned}
&\ln\left(F_n(x)\right)\\
&= \sum_{\ell=1}^{n}\sum_{m\geq 1}\frac{(-1)^{m+1}}{m}\left(\sum_{k\geq 1}\frac{\alpha Q^k\mathbb{1}}{k!}x^k(p_{\ell,n})^k\right)^m\\
&= \sum_{\ell=1}^{n}\sum_{m\geq 1}\frac{(-1)^{m+1}}{m}\sum_{k_1\geq 1}\cdots\sum_{k_m\geq 1}\frac{\alpha Q^{k_1}\mathbb{1}\ldots\alpha Q^{k_m}\mathbb{1}}{k_1!\ldots k_m!}x^{k_1+\cdots+k_m}(p_{\ell,n})^{k_1+\cdots+k_m}\\
&= \sum_{m\geq 1}\sum_{k_1\geq 1}\cdots\sum_{k_m\geq 1}\frac{(-1)^{m+1}}{m}\frac{\alpha Q^{k_1}\mathbb{1}\ldots\alpha Q^{k_m}\mathbb{1}}{k_1!\ldots k_m!}x^{k_1+\cdots+k_m}\left(\sum_{\ell=1}^{n}(p_{\ell,n})^{k_1+\cdots+k_m}\right).
\end{aligned}$$

The above expansions clearly converge in any desirable sense whenever $|x| < \ln 2$ (say, for instance, uniformly in x on compact subsets of $\{x \mid |x| < \ln 2\}$). The existence of the limiting values V_k, together with the pointwise bound

$$\left|\frac{(-1)^{m+1}}{m}\frac{\alpha Q^{k_1}\mathbb{1}\ldots\alpha Q^{k_p}\mathbb{1}}{k_1!\ldots k_m!}x^{k_1+\cdots+k_m}\left(\sum_{\ell=1}^{n}(p_{\ell,n})^{k_1+\cdots+k_m}\right)\right| \leq \frac{|x|^{k_1+\cdots+k_m}}{k_1!\ldots k_m!},$$

a converging series whenever $|x| < \ln 2$, therefore provides the limit

$$\begin{aligned}
&\lim_{n\to\infty}\ln\left(F_n(x)\right) =\\
&\sum_{m\geq 1}\sum_{k_1\geq 1}\cdots\sum_{k_m\geq 1}\frac{(-1)^{m+1}}{m}\frac{\alpha Q^{k_1}\mathbb{1}\ldots\alpha Q^{k_p}\mathbb{1}}{k_1!\ldots k_m!}x^{k_1+\cdots+k_m}V_{k_1+\cdots+k_m},
\end{aligned}\tag{9}$$

and the above convergence is uniform on compact subsets of $\{x \mid |x| < \ln 2\}$, which completes the proof. $\blacksquare$

Armed with the above theorem, we are able to deduce the limiting behavior of $\mathbb{P}\{\Theta_n^\gamma > k\}$ in the following theorem.

Theorem 6 *Under hypothesis (H), for every $k \geq 0$, we have*

$$\lim_{n\longrightarrow\infty} \mathbb{P}\{\Theta_n^\gamma > k\} = F^{(k)}(0),$$

and

$$\lim_{n\longrightarrow\infty} E(\Theta_n^\gamma) = \sum_{m\geq 1}\sum_{k_1\geq 1}\cdots\sum_{k_m\geq 1} \frac{(-1)^{p+1}}{p}\frac{k!}{k_1!\dots k_p!}\left(\alpha Q^{k_1}\mathbb{1}\right)\dots\left(\alpha Q^{k_p}\mathbb{1}\right)V_{k_1+\cdots+k_p}.$$

Proof. The argument is standard. The function $F_n(x)$ being clearly analytic on the disk $\{z \in \mathbb{C} \mid |z| < \ln 2\}$, we may write, for any $0 < r < \ln 2$ and $k \geq 0$, the relation

$$\mathbb{P}\{\Theta_n > k\} = F_n^{(k)}(0) = \frac{k!}{2i\pi}\int_{|z|=r}\frac{F_n(z)}{z^{k+1}}dz.$$

Hence, using the above-mentioned uniform convergence of F_n towards F, we recover

$$\lim_{n\to\infty}\mathbb{P}\{\Theta_n > k\} = \lim_{n\to\infty}\frac{k!}{2i\pi}\int_{|z|=r}\frac{F_n(z)}{z^{k+1}}dz = \frac{k!}{2i\pi}\int_{|z|=r}\frac{F(z)}{z^{k+1}}dz.$$

On the other hand, since the function F itself clearly is analytic on the disk $\{z \in \mathbb{C} \mid |z| < \ln 2\}$ as well, we may write similarly

$$\frac{k!}{2i\pi}\int_{|z|=r}\frac{F(z)}{z^{k+1}}\,dz = F^{(k)}(0),$$

which completes the proof. $\blacksquare$

We denote by Θ^γ the random variable having as distribution the limiting distribution of Θ_n^γ. We then have, for every $k \geq 0$,

$$\lim_{n\longrightarrow\infty}\mathbb{P}\{\Theta_n^\gamma > k\} = \mathbb{P}\{\Theta^\gamma > k\}.$$

The following corollary shows how to compute recursively the limiting distribution Θ_n^γ.

Corollary 7 *If, for a fixed $h \geq 1$, we have $\lim_{n \longrightarrow \infty} p_{h,n} = b > 0$, then, we have $\mathbb{P}\{\Theta^\gamma > 0\} = 1$ and, for every $k \geq 1$,*

$$\mathbb{P}\{\Theta^\gamma > k\} = \frac{1}{1-(1-b)^k} \sum_{\ell=0}^{k-1} \binom{k}{\ell} (1-b)^\ell b^{k-\ell} \mathbb{P}\{\Theta^\gamma > \ell\} \alpha Q^{k-\ell} \mathbb{1}. \tag{10}$$

Proof. Since $\mathbb{P}\{\Theta_n^\gamma > k\} = 1$ for every $n \geq 1$, we have $\mathbb{P}\{\Theta^\gamma > 0\} = 1$. If, for a fixed $h \geq 1$, we have $\lim_{n \longrightarrow \infty} p_{h,n} = b > 0$, then, using Theorem 3 and taking the limit in Relation (5), we get

$$\mathbb{P}\{\Theta^\gamma > k\} = \sum_{\ell=0}^{k} \binom{k}{\ell} b^\ell (1-b)^{k-\ell} \alpha Q^\ell \mathbb{1} \mathbb{P}\{\Theta^\gamma > k-\ell\}. \tag{11}$$

Extracting the term containing $\mathbb{P}\{\Theta^\gamma > k\}$, which corresponds to index $\ell = 0$, from the right hand side, we get the desired relation. ■

Without any loss of generality, by renumbering the Markov chains, we take in the rest of the paper $h = 1$. This means, from Theorem 3, that the probability mass function $\gamma(n-1) = (p_{1,n-1}, \ldots, p_{n-1,n-1})$ associated with Θ_{n-1}^γ is given, for $r = 1, \ldots, n-1$, by

$$p_{r,n-1} = \frac{p_{r+1,n}}{1 - p_{1,n}}. \tag{12}$$

For a fixed value of $n \geq 2$, the computation of the distribution of $Theta_n^\gamma$ with a given probability mass distribution $\gamma(n)$ necessites the computation of the distribution of Θ_n^γ with the probability mass distribution $\gamma(n-1)$ given by Relation (12). Let ε be a predetermined error tolerance. If we want to compute $\mathbb{P}\{\Theta_n^\gamma > k\}$ for every k such that $\mathbb{P}\{\Theta_n^\gamma > k\} > \varepsilon$ we need to determine an integer K such that, for every $i = 1, \ldots, n$, $\mathbb{P}\{\Theta_i^\gamma > K\} \leq \varepsilon$ and then compute, for $i = 1, \ldots, n$, the values of $\mathbb{P}\{\Theta_i^\gamma > k\}$ for $k = 0, \ldots, K-1$. The following lemma will be used in the next theorem where we propose a value of K. An inequality between vectors is meant entrywise.

Lemma 8 *For every $k \geq 1$, the vector function $f(x)$ defined, for $x \in [0,1]$, by*

$$f(x) = (xQ + (1-x)I)^k \mathbb{1}$$

is decreasing.

Proof. The function f is differentiable on the interval $(0,1)$ and its derivative f' is given by

$$f'(x) = k(xQ + (1-x)I)^{k-1}(Q\mathbb{1} - \mathbb{1}).$$

The matrix Q being substochastic, we have $Q\mathbb{1} - \mathbb{1} \le 0$ with strict inequality for at least one entry. We thus have $f'(x) \le 0$ which means that fiunction f is decreasing on interval $[0,1]$. ■

For every $n \ge 1$, we introduce the numbers m_n defined by

$$m_n = \min_{i=1,\ldots,n} p_{1,i}.$$

Theorem 9 *For every $n \ge 1$, for every $\varepsilon \in (0,1)$, we have*

$$\max_{i=1,\ldots,n} \mathbb{P}\{\Theta_i^\gamma > k\} \le \varepsilon \text{ for every } k \ge K,$$

where

$$K = \inf\left\{k \ge 0 \;\middle|\; \alpha\left(m_n Q + (1-m_n)I\right)^k \mathbb{1} \le \varepsilon\right\}.$$

Proof. For every $i = 1,\ldots,n$, we have

$$\begin{aligned}
\mathbb{P}\{\Theta_i^\gamma > k\} &= \sum_{\ell=0}^{k} \binom{k}{\ell} p_{1,i}^\ell (1-p_{1,i})^{k-\ell} \alpha Q^\ell \mathbb{1} \mathbb{P}\{\Theta_{i-1}^\gamma > k-\ell\} \\
&\le \sum_{\ell=0}^{k} \binom{k}{\ell} p_{1,i}^\ell (1-p_{1,i})^{k-\ell} \alpha Q^\ell \mathbb{1} \\
&= \alpha\left(p_{1,i} Q + (1-p_{1,i})I\right)^k \mathbb{1} \\
&\le \alpha\left(m_n Q + (1-m_n)I\right)^k \mathbb{1} \qquad \text{(from Lemma 8)} \\
&= \sum_{\ell=0}^{k} \binom{k}{\ell} m_n^\ell (1-m_n)^{k-\ell} \alpha Q^\ell \mathbb{1}.
\end{aligned}$$

Note that matrix $m_n Q + (1-m_n)I$ is substochastic, i. e. $(m_n Q + (1-m_n)I)\mathbb{1} \le \mathbb{1}$ with the strict inequality for at least one entry. This means in particular that $\alpha\left(m_n Q + (1-m_n)I\right)^k \mathbb{1}$ is decreasing with k and

$$\lim_{k \longrightarrow \infty} \alpha\left(m_n Q + (1-m_n)I\right)^k \mathbb{1} = 0,$$

So, for a fixed $\varepsilon \in (0,1)$ and by definition of integer K we have that for every $i = 1,\ldots,n$,

$$\mathbb{P}\{\Theta_i^\gamma > k\} \le \varepsilon, \text{ for every } k \ge K,$$

which completes the proof. ■

In the same way, we obtain a similar result for the computation of the expected values $E(\Theta_i^\gamma)$, for $i = 1, \ldots, n$, for which the truncation of the series (6) is needed.

Theorem 10 *For every $n \geq 1$, for every $\varepsilon \in (0,1)$,*

$$0 \leq \max_{i=1,\ldots,n} \left(E(\Theta_i^\gamma) - \sum_{k=0}^{L-1} \mathbb{P}\{\Theta_i^\gamma > k\} \right) \leq \varepsilon,$$

where

$$L = \inf \left\{ k \geq 0 \;\middle|\; \frac{1}{m_n} \alpha (I-Q)^{-1} \left(m_n Q + (1-m_n) I\right)^k \mathbb{1} \leq \varepsilon \right\}.$$

Proof. We introduce the notation

$$r_i = E(\Theta_i^\gamma) - \sum_{k=0}^{L-1} \mathbb{P}\{\Theta_i^\gamma > k\}.$$

We then have, for every $i = 1, \ldots, n$,

$$\begin{aligned}
r_i &= \sum_{k=L}^{\infty} \mathbb{P}\{\Theta_i^\gamma > k\} \\
&= \sum_{k=L}^{\infty} \sum_{\ell=0}^{k} \binom{k}{\ell} p_{1,i}^\ell (1-p_{1,i})^{k-\ell} \alpha Q^\ell \mathbb{1} \mathbb{P}\{\Theta_{i-1}^\gamma > k-\ell\} \\
&\leq \sum_{k=L}^{\infty} \sum_{\ell=0}^{k} \binom{k}{\ell} p_{1,i}^\ell (1-p_{1,i})^{k-\ell} \alpha Q^\ell \mathbb{1} \\
&= \sum_{k=L}^{\infty} \alpha \left(p_{1,i} Q + (1-p_{1,i}) I\right)^k \mathbb{1} \\
&\leq \sum_{k=L}^{\infty} \alpha \left(m_n Q + (1-m_n) I\right)^k \mathbb{1} \qquad \text{(from Lemma 8)} \\
&= \alpha \left(I - \left(m_n Q + (1-m_n) I\right)\right)^{-1} \left(m_n Q + (1-m_n) I\right)^L \mathbb{1} \\
&= \frac{1}{m_n} \alpha (I-Q)^{-1} \left(m_n Q + (1-m_n) I\right)^L \mathbb{1} \\
&\leq \varepsilon \text{ by definition of integer } L_2.
\end{aligned}$$

which means that $\max_{i=1,\ldots,n} r_i \leq \varepsilon$. ■

It is easily checked, from Relation (11), that the same result holds for the limiting expected value $E(\Theta^\gamma)$. More precisely, if $\lim_{n\longrightarrow\infty} p_{1,n} = b > 0$, then, for every $\varepsilon \in (0,1)$, we have

$$0 \leq \left(E(\Theta^\gamma) - \sum_{k=0}^{H-1} \mathbb{P}\{\Theta^\gamma > k\} \right) \leq \varepsilon,$$

where

$$H = \inf \left\{ k \geq 0 \;\middle|\; \frac{1}{b}\alpha(I-Q)^{-1}\left(bQ + (1-b)\,I\right)^k \mathbb{1} \leq \varepsilon \right\}.$$

4 Geometric distribution

We suppose in the section that the probability mass distribution $\gamma(n)$ is the geometric distribution with parameter b, with $0 < b < 1$, truncated at step n, i.e. given, for $n \geq 2$ and $r = 1, \ldots, n-1$ by

$$p_{r,n} = (1-b)^{r-1} b \text{ and } p_{n,n} = (1-b)^{n-1}.$$

From Relation (12), we have $p_{1,i} = b$ for every $i \geq 2$ and thus $\lim_{n\longrightarrow\infty} p_{1,n} = b$. We then have from Theorem 3, for every $n \geq 2$,

$$\mathbb{P}\{\Theta_n^\gamma > k\} = \sum_{\ell=0}^{k} \binom{k}{\ell} b^\ell (1-b)^{k-\ell} \alpha Q^\ell \mathbb{1} \mathbb{P}\{\Theta_{n-1}^\gamma > k - \ell\},$$

IIIIIIIIIIIIIIIIIIIIICCCCCCCCCCCCCCCCCCCIIIIIIIIIIIIIIII

5 Application to cluster-based distributed storage

A cluster-based distributed storage peer-to-peer system guarantees durable access to large scale applications such as file sharing, streaming, or video-on-demand. It is achievable by harnessing the very large storage space globally provided by the many unused or idle nodes connected to the network. A common approach to handle these nodes is by having nodes that are close to each other according to a given proximity metric to self-organize into clusters. Specifically, each object (*e.g.* data stream, file) is divided into k equal size fragments, and recoded into a potentially unlimited number of independent check blocks through a rateless-erasure coding (also called Fountain) schema (*e.g.* [5]). Fundamental property of erasure coding is that one may recover an initial object by collecting k' distinct check blocks generated by different sources, with k' slightly greater than k. During the coding phase, each check block c_i is generated by *(i)* choosing a *degree* d_i from a particular degree distribution, *(ii)* randomly choosing d_i distinct input symbols (called neighbors of c_i) among the k input symbols, and *(iii)* combining the d_i neighbors into a check block c_i by performing a bitwise XOR operation. The key idea of the decoding process is to build the Tanner graph based on the set of received check blocks. Upon receipt of check blocks, the decoder *(i)* finds any check block c_i with degree equal to one *(ii)* removes the edge between c_i and k_i in the Tanner graph, and *(iii)* executes a bitwise XOR operation between k_i and any remaining check block c_r that has k_i as neighbor, and remove the edge between c_r and k_i. These steps are repeated until all k input symbols are successfully recovered. To guarantee the success of the decoding, the degree distribution is designed so that as few as possible check blocks are needed to ensure minimum redundancy among them, and the average degree is as low as possible to reduce the average number of symbol operations to recover the original data. This amounts to generating check blocks so that in average no more than 1/4 of them are degree one to start the decoding and to prevent a too high amount of redundancy among these check blocks, 1/2 of them are of degree 2 so that combined with degree 1 check blocks they allow to cover a large proportion of input blocks, and 1/8 of them are of degree 3 so that the decoding process is unlikely to be get stuck. The repartition of the other check blocks classically shows a steep decline, *i.e.* $1/2^i$ of them are of degree i. These check blocks are disseminated to the nodes of the system so that all the nodes that receive degree 1 check blocks self-organize in a cluster, those that receive degree two check blocks self organize in another cluster, and so on and so forth. Nodes can freely join and leave a cluster. For scalability and reliability reasons the number of nodes in a cluster is constrained. When the cluster size undershoots m nodes, then new check blocks are generated so that new nodes will join the cluster. Similarly when it exceeds M then generation of check blocks is suspended.

We model the effect of join and leave events using a homogeneous discrete-time Markov chain denoted by $X = \{X_n, n \geq 0\}$. Markov chain X represents the evolution of the number of nodes in the system. The Markov chain X modeling the behavior of one cluster is depicted in Figure **??** in which $q = 1 - p$ and $p \in (0, 1)$.

gam.eepicMarkov chain model of one cluster.fig:gambler

The transition probability p means that a new peer joins the cluster while the transition probability q means that a peer leaves the cluster. The transition from state $m+1$ to the absorbing state expresses that the cluster has reached its minimal size m and that the coding process has to be activated. In the same way the transition from state $M-1$ to the absorbing state means that the cluster has reached its maximal size M and that the coding process has to be suspended. The initial distributions α that we consider are the unit row vectors e_j for $j = m+1, \ldots, M-1$. So, the initial distribution $\alpha = e_j$ means that $X_0 = j$ with probability 1.

The matrix Q which gives the transitions between the transient states of X is thus a tri-diagonal matrix where non-zero entries are $Q_{i,i+1} = p$ and $Q_{i,i-1} = q = 1 - p$. The probability mass function $\pi(n)$ is such that $p_{i,n} = 1/n$, for every $i = 1, \ldots, n$. With these values, the limiting behavior of respectively the distribution and the expectation of Θ_n are given from Theorem 6 for every $k \geq 0$, by

$$\lim_{n \longrightarrow \infty} \mathbb{P}\{\Theta_n > k\} = \begin{cases} p^k & \text{if} \quad X_0 = S_{\min} + 1 \\ 1 & \text{if} \quad X_0 = j, \ \text{for } S_{\min} + 2 \leq j \leq S_{\max} - 2 \\ (1-p)^k & \text{if} \quad X_0 = S_{\max} - 1 \end{cases}$$

and

$$\lim_{n \longrightarrow \infty} E(\Theta_n) = \begin{cases} \dfrac{1}{1-p} & \text{if} \quad X_0 = S_{\min} + 1 \\ \infty & \text{if} \quad X_0 = j, \ \text{for } S_{\min} + 2 \leq j \leq S_{\max} - 2 \\ \dfrac{1}{p} & \text{if} \quad X_0 = S_{\max} - 1. \end{cases}$$

For the numerical evaluations, we have chosen $p = 1/2$. With this value, we easily get, when $\alpha = e_j$,

$$E(\Theta_1) = (j - S_{\min})(S_{\max} - j).$$

We have also chosen $S_{\min} = 4$ and $S_{\max} = 16$ which implies that the number of transient states is equal to 11.

$$\alpha(I - Q)^{-1} = (1, 2, 3, 4, 5, 6, 5, 4, 3, 2, 1)$$

Figure 1: From bottom to the top: $\mathbb{P}\{\Theta_1^\gamma > k\}$, $\mathbb{P}\{\Theta_2^\gamma > k\}$, $\mathbb{P}\{\Theta_3^\gamma > k\}$, $\mathbb{P}\{\Theta_4^\gamma > k\}$, $\mathbb{P}\{\Theta_5^\gamma > k\}$, $\mathbb{P}\{\Theta^\gamma > k\}$ when $X_0 = 6$ ou 10, as functions of k.

$\varepsilon = 10^{-4}$.

$E(\Theta_1^\gamma)$	$E(\Theta_2^\gamma)$	$E(\Theta_3^\gamma)$	$E(\Theta_4^\gamma)$	$E(\Theta_5^\gamma)$	$E(\Theta_6^\gamma)$	$E(\Theta_7^\gamma)$	$E(\Theta_8^\gamma)$	$E(\Theta^\gamma)$
36	42.2046	47.4027	50.6644	52.0177	52.3487	52.3928	52.3960	52.3961

Fig. 1: Values of k^* for different values of ε and n.

References

[1] Anceaume, E., Castella, F., Ludinard, R. and Sericola, B.(2010). Markov Chains Competing for Transitions: Application to Large-Scale Distributed Systems. Research Report n^o 1953, http://hal.inria.fr/inria-00485667/PDF/RR-1953.pdf

[2] Fourneau, J.-M. (2008). Discrete time Markov chains competing over resources: product form steady-state distribution. In *Proceedings of the 5th International Conference on the Quantitative Evaluation of Systems (QEST'08)*, Saint-Malo, France, 147-156.

[3] Karlin, S. and Taylor, H. M. (1981). A Second Course in Stochastic Processes. Academic Press.

[4] Neuts, M. F. (1981). Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach. Johns Hopkins University Press.

[5] M. Luby. LT codes. In *Proceedings of the IEEE International Symposium on Foundations of Computer Science (SFCS)*, 2002.

[6] P. Maymounkov. Online codes. *Research Report TR2002-833, New York University*, 2002.

[7] A. Shokrollahi. Raptor codes. *IEEE/ACM Transactions on Networking*, pages 2551–2567, 2006.

[8] H.R. Ribeiro and E. Anceaume. A Comparative Study of Rateless Codes for P2P Distributed Storage. *Proceedings of the 12th International Symposium on Stabilization, Safety, and Security of Distributed Systems (SSS 2010)*, 2010.