Computing Absorbing Times via Fluid Approximations

Nicolas Gast, Bruno Gaujal

Applied Probability Trust (June 12, 2017)

COMPUTING ABSORBING TIMES VIA FLUID APPROXIMATIONS

NICOLAS GAST,* *Inria*

BRUNO GAUJAL,** *Inria*

Abstract

In this paper, we compute the absorbing time T_n of a n-dimensional discrete time Markov chain made of n components, each with an absorbing state and evolving in mutual exclusion. We show that the random absorbing time T_n is well approximated by a deterministic time t_n that is the first time when a fluid approximation of the chain approaches the absorbing state at a distance $1/n$. We provide an asymptotic expansion of t_n that uses the spectral decomposition of the kernel of the chain as well as the asymptotic distribution of T_n, relying on extreme values theory. We show the applicability of this approach with three different problems: the coupon collector, the erasure channel lifetime and the coupling times of random walks in high dimensional spaces.

Keywords: Markov chain, Fluid approximation, Extreme values, Coupon collector, Coupling time

2010 Mathematics Subject Classification: Primary 05D40
Secondary 60J10;37A25

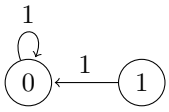
1. Introduction.

Analyzing stochastic discrete event dynamic systems by constructing deterministic fluid approximations (sometimes called hydrodynamic limits) has been very popular in recent years. This technique is very powerful to get fast and accurate performance evaluation of computer based systems [2, 6, 13]. It can be applied quite generally to study the transient behavior or the convergence to stationary or quasi-stationary distributions [4, 14, 7].

In this paper, we rather consider the case where we want to approximate a finite (but large) stochastic system over a *stopping time* of the process. We consider a Markov chain $\mathbf{X} := (X_1 \dots X_n)$, where X_i is called the ith component. The components are all Markov chains that have an absorbing state (say state 0) and respective transition matrices $P_1, \dots, P_n$. The evolution of $\mathbf{X}$ is as follows: At step t, a single component is chosen according to a fixed probability vector $(p_1, \dots, p_n)$ and the chosen component (say i) makes one transition, according to its kernel P_i. Our goal is to compute the first instant T_n when all components have reached their absorbing state.

The simplest example of such a system is the coupon collector. Each coupon is seen as a Markov chain with two states. The absorbing state (0) corresponds to the presence of the coupon in the current collection and the other state (1) to its absence:

* Postal address: University of Grenoble Alpes and Inria, Grenoble, France
** Postal address: University of Grenoble Alpes and Inria, Grenoble, France



In the original coupon collect, all the chains $X_1 \dots X_n$ have this transition matrix and the choice probability vector is uniform: $p_i = 1/n$ for $i \in \{1 \dots n\}$.

The coupon collector, or one of its generalizations, appears naturally in the analysis of distributed algorithms, for example in [11, 12]. Given its importance, several generalizations of the coupon collector have been studied in the literature. Many of them fall in our general model. In [20], the author studies the asymptotic distribution of the tail-distribution of the time to collect a collection under a non-uniform *i.i.d.* distribution of coupons. In [9], the authors study a non-uniform coupon (coupons are *i.i.d.* with non-uniform probabilities) and obtain the asymptotic behavior of the variance of the time to collect n coupons. In [1] the authors study the time to collect $c < n$ different coupons. They show that the uniform coupon is often a best-case scenario for the hitting time in many cases. Our model of heterogeneous coupon is close to the one of [8, 9] that studies asymptotic properties of the problem of collecting n different coupons, where coupon i has probability $p_i = a_i / \sum_{j=1}^{n} a_j$. They fix a sequence a_i and study asymptotic properties. Our problem can also be seen as a special case of the coupon collector with random quotas studied in [19, 23] where the authors study the time to collect m_i coupons of type i, where m_i is a random variable. Our paper differs on two points: (i) from a modeling point of view, we focus on the case where m_i is the hitting time of a Markov chain, and (ii) the authors use generating functions while we use a fluid approximation. With our model, we are able to use extreme value theory to obtain close-form expression for asymptotic distribution and moments.

Contributions In this paper, we compute an asymptotic equivalent t_n of the absorbing time T_n under weak assumptions on p_n (see Section 3.2 for a precise statement). The approximation t_n is obtained by defining a fluid approximation of $\mathbf{X} := (X_1 \dots X_n)$ and by computing the time when this fluid approximation is at a distance $1/n$ from the absorbing state. Our main result is to show that T_n is asymptotic equal (in distribution) to t_n plus a Gumbel distribution with mode 0 and scale $1/\nu \min_i p_i$, plus a negligible sub-linear term $o(n)$. The exponent ν depends on the eigenvalues of the transition matrices. We also provide closed forms expressions for the asymptotic expectation of T_n and its moments.

The deterministic time t_n is easier to compute that its stochastic counterpart T_n. When all chains have the same transition matrix and $p_i = 1/n$, we show that $t_n = (1/\nu) n \log n + ((d-1)/\nu) n \log\log n + O(n)$, where $-\nu$ is the eigenvalue of Q with the largest real part (which is a real negative number) and d depends on its multiplicity. We apply this to the coupon collector in which case $-\nu = 1$ (see Section 4) as well as to random walks in finite grids (see Section 5 where we give an equivalent to the coupling time in grids with or without drifts). We also provide bounds on t_n based on the absorbing time of a single component in Section 4.4. In the heterogeneous and/or non-uniform cases, t_n can often only be computed numerically (see Section 4.3 for a numerical application to the coupon collector with rare coupons).

Our proofs rely on two ingredients. First, we use a classical "Poissonization tech-

nique" to transform the system of n coupled Markov chains into a continuous time Markov chain made of n independent components. Second, we use ideas from extreme value theory to relate T_n and the time when the fluid approximation approaches $1/n$.

Road-map The rest of the paper is organized as follows. We introduce the model and the definition of t_n in Section 2. We develop the equivalence theorems between t_n and T_n is Section 3. We present the applications to the coupon collector in Section 4, where we also show how to compute a bound (Theorem 4.1). Next, we show how this can be applied to compute the coupling time of random walks in Section 5. Finally, we conclude in Section 6.

2. Problem statement.

We consider a discrete-time finite Markov chain $\mathbf{X}$ made of n chains: $X_1, \ldots, X_n$. The chain X_i called component i in the following. All components have an absorbing state 0 and respective transition matrices $P_1, \ldots, P_n$. The component X_i lives in a finite state space $\{0 \ldots k\}$ and 0 is an absorbing state: $P_i(0,0) = 1$. We assume that the state 0 is reachable from any initial state, which implies that with probability one, each component hits 0 in finite time. For reasons that will become clear in the next section, the transition matrix P_i will be decomposed as

$$P_i = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{q}_i & I + Q_i \end{bmatrix}, \tag{1}$$

where Q_i is a $k \times k$ rate matrix with the following properties: Q_i is non-singular, for all a, $\sum_b Q_i(a,b) \leq 0$ and there exists at least one state a such that $\sum_b Q_i(a,b) < 0$.

Let $p = (p_1 \geq \cdots \geq p_n)$ a probability distribution on $\{1 \ldots n\}$. The discrete-time Markov chain $\mathbf{X} = (X_1 \ldots X_n)$ is defined by the following dynamics. At time 0, each $X_i(0)$ is picked independently according to an initial distribution α_i ($(\alpha_i)_0 = 0$, because component i is not in its absorbing state initially, so that we denote by α_i the k-dimensional row vector $((\alpha_i)_1 \ldots (\alpha_i)_k)$. Then at each time step t, one coordinate i is picked with probability p_i, independently of the past. Then, the ith coordinate, X_i, makes one transition according to the matrix P_i.

In what follows, we characterize T_n, the hitting time of $(0, 0, \ldots, 0)$:

$$T_n := \inf\{t \text{ such that } \forall i \in \{1 \ldots n\} : X_i(t) = 0\}$$

We distinguish several cases in the analysis of this problem. We say that the chain is *homogeneous* if all components $X_1 \ldots X_n$ have the same transition matrix P, and the same initial distribution α. From this point on, we assume homogeneity, unless it is specified otherwise. The non-homogeneous case is treated in Section 3.3. We say that $\mathbf{X}$ is *uniform* if the choice probabilities among the components are all equal: $(p_1, \ldots, p_n) = (1/n, \ldots, 1/n)$. In the following, we will show that when n is large, T_n is well approximated by a deterministic time t_n that we define below.

2.1. Definition of t_n in the uniform and homogeneous case.

In the uniform and homogeneous case, we define t_n to be the unique solution of the following equation, using $Q = Q_i$ for all i, as defined in (1), α, the initial distribution

vector of each component and $\mathbf{1}$, the k-dimensional column vector whose elements are all equal to 1.

$$\frac{1}{n} = \alpha \exp(Qt_n/n)\mathbf{1}, \tag{2}$$

where $\mathbf{1}$ denotes the k-dimensional column vector whose elements are all equal to 1.

This definition is justified by using a dynamical system that is a mean field limit of the Markov chain. We define the empirical measure, $(M_0, M_1, \ldots, M_k)$ of the chain:

$$M_a(t) := \frac{1}{n}\sum_{i=1}^{n} \delta_{\{X_i(t)=a\}}.$$

By the definition of Q given in (1), for all $a \neq 0$, we have

$$\mathbb{E}\left[M_a(t+1) - M_a(t) | \mathbf{M}(t)\right] = \frac{1}{n}\sum_b M_b(t)Q(b,a).$$

When the number of components becomes large, the empirical measure $(M_0, M_1, \ldots, M_k)$ converges to a deterministic population $(m_0, m_1, \ldots, m_k)$ whose dynamics is given by an ordinary differential equation (ODE) (see [16] for example). In matrix form, and focusing on the k-dimensional vector $\mathbf{m} := (m_1, \ldots, m_k)$ (removing the coordinate m_0), this equation can be written

$$\dot{\mathbf{m}}(t) = \frac{1}{n}\mathbf{m}(t)Q. \tag{3}$$

This ODE is linear and its solution can be computed explicitly. The sum $m(t) := \mathbf{m}\cdot\mathbf{1}$ is equal to $m(t) = \alpha\exp(Qt/n)\mathbf{1}$.

When n is large, the proportion of components not yet in state 0 can be approximated by $m(t)$.

Notice that, unlike in the finite case, $m(t)$ never reaches 0. However, since in the system with n components $M_1(t) + M_2(t) + \cdots + M_k(t)$ reaches 0 as soon as it becomes smaller than $1/n$, a good fluid approximation of the stopping time T_n should be obtained by setting

$$t_n := \inf\{t > 0 \text{ such that } m(t) = \frac{1}{n}\}. \tag{4}$$

This definition implies that t_n is the solution of Equation (2).

2.2. Definition of t_n in the non-uniform or non-homogeneous case.

Let us now consider the non-homogeneous, non-uniform case, with a general choice vector $(p_1, \ldots, p_n)$. We always assume with no loss of generality that $p_1 \geq p_2 \geq \cdots \geq p_n$. We do not construct explicitly the population dynamics but jump directly to the analog of Equation (2).

If one considers the evolution of the ith chain, independently of the rest, it is a Markov chain with transition matrix

$$\begin{bmatrix} 1 & \mathbf{0} \\ p_i\mathbf{q}_i & I + p_iQ_i \end{bmatrix}.$$

In particular, the probability for X_i to be equal to 0 at time t is: $\mathbb{P}(X_i(t) = 0) = \alpha_i(I + p_iQ_i)^t\mathbf{1}$, which is close to $\alpha_i \exp(p_iQ_it)\mathbf{1}$ as p_i is small and t is large.

We define $m(t) = n^{-1}\sum_{i=1}^{n} \alpha_i \exp(p_iQ_it)\mathbf{1}$ the *fluid approximation* of the process $\mathbf{X}$ and define as t_n the time for the fluid approximation to reach $1/n$, as in the uniform case:

$$t_n := \inf\{t > 0 \text{ such that } m(t) = \frac{1}{n}\}. \tag{5}$$

In the non-uniform case, the behavior of T_n depends on how the choice probabilities $(p_1, \ldots, p_n)$ evolve when n goes to infinity. In Section 3.2, we provide general conditions on $(p_1, \ldots, p_n)$ under which t_n is a good approximation of T_n.

3. Convergence results.

This section contains the main results of our paper, namely Theorems 3.1, 3.2 and 3.3. We will first remove the difficulty in dealing with the dependencies between the components by using a classical Poissonization trick. Once this is done, we focus on the homogeneous case in Section 3.2. The heterogeneous case follows from the homogeneous case by using a continuity arguments (Theorem 3.3).

3.1. Removing the dependence ("Poissonization" trick)

The first difficulty in dealing with the discrete-time Markov chain $\mathbf{X}$ is that the components $X_1, \ldots, X_n$ are not independent: If component X_i takes a step, then all the others remain still, by the mutual exclusion principle. We remove this difficulty by using the classical trick that consists in constructing a continuous time Markov chain that approximates the discrete time evolution of $\mathbf{X}$ and makes all components independent, in continuous time. We detail this construction below.

Let $(Z_1, Z_2, \ldots)$ be a real Poisson process with rate 1. Let us define a continuous time Markov chain $\mathbf{Y}$, associated with $\mathbf{X}$, whose jumps occur at times $Z_1, Z_2, \ldots$. More precisely, $\mathbf{Y}$ is a piecewise constant cadlag function such that for all n, $\mathbf{Y}(Z_n) = \mathbf{X}(n)$. Let $\widetilde{T}_n$ be the time when $\mathbf{Y}$ gets absorbed. By definition, $\widetilde{T}_n = Z_{T_n}$.

By construction, $\widetilde{T}_n$ can be viewed as the maximum of n independent random variable, which makes its analysis simpler compared to the one of T_n. Indeed, $\mathbf{Y}$ has the same law as a process composed of n independent continuous time Markov chains $(Y_1, \cdots, Y_n)$, where Y_i is a continuous time Markov chain of kernel p_iQ_i. In particular, $\mathbb{P}(\widetilde{T}_n \leq t) = \prod_{i=1}^{n} \mathbb{P}(Y_i(t) = 0) = \prod_{i=1}^{n}(1 - \alpha_i \exp(p_iQ_it)\mathbf{1})$. In the proofs of Theorem 3.1 and 3.2, we use the fact that the variables T_n and $\widetilde{T}_n$ are close to compute the behavior of T_n.

3.2. Convergence in the homogeneous case.

To prove the concentration of T_n around t_n, we make the following scaling assumptions on the way the probability vector evolves as n grows. To emphasize the dependence on n, let $(p_1^{(n)} \ldots p_n^{(n)})$ denotes the probability vector for the chain with n components. We assume that:

$$\text{For each } n\text{: } p_i^{(n)} \geq p_{i+1}^{(n)} \text{ and } p_{i-1}^{(n)}/p_i^{(n)} \text{ decreases with i} \tag{6}$$

$$\text{For any } \ell\text{: } \lim_{n\to\infty} \frac{p_n^{(n)}}{p_{n-\ell}^{(n)}} = 1. \tag{7}$$

These two assumptions are satisfied by any distribution that is close enough to the uniform distribution ($p_i^{(n)} = 1/n$), or the almost uniform distribution with a popular element used for example in [1] ($p_1^{(n)} = c > 0$ and $p_i^{(n)} = (1-c)/(n-1)$ for $i > 1$). They are also satisfied by distributions of the form $p_i^{(n)} = a_i / \sum_{j=1} a_j$, where $a_i > 0$ is a positive sequence that decreases slower than any exponential, for example with $p_i^{(n)} := i^r / \sum_{j=1}^n j^r$ or $p_i^{(n)} := (\log i)^r / \sum_{j=1}^n (\log j)^r$ for $r \leq 0$. Note that Assumptions (6)-(7) generalize the conditions (2.23) of [8].

Theorem 3.1. *Assume that the probability distribution $(p_1^{(n)} \dots p_n^{(n)})$ satisfies (6) and (7). Then, for any x, the hitting time T_n for the stochastic system composed of the n chains satisfies:*

$$\lim_{n\to\infty} \mathbb{P}\left(T_n \leq t_n + \frac{x}{p_n^{(n)}} \right) = \exp(-e^{-\nu x}). \tag{8}$$

This theorem essentially shows that the asymptotic distribution of $p_n^{(n)}(T_n - t_n)$ converges to a Gumbel distribution with mode 0 and scale $1/\nu$. An interesting property of this result is that the shape of the distribution only depends on the smallest value of the distribution, $p_n^{(n)}$ (even when this distribution is far from uniform, as in the case $p_i^{(n)} = 1/\sum_{j=1}^n (i/j)$). The proof is postponed to Appendix A.

Next, we show in Theorem 3.2 that the convergence also holds for the moments.

Theorem 3.2. *Assume that the sequence of distributions $(p_1^{(n)} \dots p_n^{(n)})$ satisfies Equations (6) and (7). Let M_m be the mth moment of the standard Gumbel distribution. Then, the hitting time T_n for the stochastic system composed of the n chains satisfies:*

$$\lim_{n\to\infty} \mathbb{E}\left[(p_n^{(n)}(T_n - t_n))^m \right] = \nu^{-m} M_m. \tag{9}$$

In particular, this theorem implies that:

$$\lim_{n\to\infty} p_n^{(n)}(\mathbb{E}[T_n] - t_n) = \frac{\gamma}{\nu},$$
$$\lim_{n\to\infty} \operatorname{var}\left[p_n^{(n)} T_n \right] = \frac{\pi^2}{6\nu^2},$$

where γ is the Euler–Mascheroni constant $\gamma = \lim_{n\to\infty} (\sum_{k=1}^n (1/k) - \log(n))$.

This theorem provides asymptotic closed form values for the absorbing time of the chain, not only bounds, or $O(.)$ limits, as it is sometimes done. Again, as for the distribution, the moments only depend on the smallest probability $p_n^{(n)}$. The proof is postponed to Appendix B.

3.3. Heterogeneous case.

To deal with different kernels, we simplify the assumptions (6) and (7) on the choice vector. They are replaced by Assumption (10): We assume that there exists a constant $c > 0$ such that for any component i,

$$p_i^{(n)} \geq c/n. \tag{10}$$

We now consider a collection of matrix $(Q^{(\theta)})$, indexed by a parameter $\theta \in \Theta$, where $\Theta \subset \mathbb{R}^d$ is a compact set. We assume that $Q_{ij}^{(\theta)}$ is a continuous function of θ and that for all $\theta \in \Theta$, Q^θ is non-singular and such that for all $a \neq b$ $Q^\theta(a,b) \geq 0$, for all a: $\sum_b Q^\theta(a,b) \leq 0$ and for at least one a: $\sum_b Q^\theta(a,b) < 0$. The evolution of the component i depends on the matrix $Q^{(\theta_i)}$, for a given θ_i. If the component i starts with a distribution α_i, the probability that this component i has reached its absorbing state at time-slot nt is therefore $\alpha_i(1+Q^{(\theta_i)}/n)^{nt}\mathbf{1}$. We define the quantity $m_{\theta_1\dots\theta_n}(t)$ as

$$m_{\theta_1\dots\theta_n}(t) := \frac{1}{n}\sum_{i=1}^{n} \alpha_i \exp(Q^{(\theta_i)}t)\mathbf{1}.$$

The hitting time $t_{\theta_1\dots\theta_n}$ is defined similarly to t_n:

$$t_{\theta_1\dots\theta_n} = \inf\left\{t : m_{\theta_1\dots\theta_n}(t) \leq \frac{1}{n}\right\}. \tag{11}$$

For each $\theta \in \Theta$, let $-\nu_\theta$ be the eigenvalue of Q^θ with the greatest real part. ν_θ is a continuous function of θ. Let $\nu_{\max} = \max_{\theta\in\Theta} \nu_\theta$ and $\nu_{\min} = \min_\theta \nu_\theta$.

Theorem 3.3. *For all x, there exists $\nu_{\theta_1\dots\theta_n}(x) \in [\nu_{\min}, \nu_{\max}]$ such that the hitting time $T_{\theta_1\dots\theta_n}$ satisfies*

$$\lim_{n\to\infty} \left|\mathbb{P}\left(T_{\theta_1\dots\theta_n} \geq n(t_{\theta_1\dots\theta_n} + x)\right) - \exp(-e^{-\nu_{\theta_1\dots\theta_n}(x)x})\right| = 0$$

Moreover, there exists $\nu_{\theta_1\dots\theta_n} \in [\nu_{\min}, \nu_{\max}]$ such that

$$\lim_{n\to\infty} \left|\frac{\mathbb{E}[T_{\theta_1\dots\theta_n}] - \left(t_{\theta_1\dots\theta_n} + \frac{\gamma}{\nu_{\theta_1\dots\theta_n}}\right)}{n}\right| = 0$$

The proof is postponed to Appendix C

4. Why approximating T_n by t_n helps ?

We claim that t_n is a simpler quantity to work with than T_n. In this section, we illustrate this claim in four different contexts.

1. First we show that t_n admits an asymptotic expansion when n goes to infinity that can be expressed using the spectral decomposition of Q (Subsection 4.1).

2. We also show in Subsection 4.2 that sometimes t_n admits an asymptotic closed form. This is the case of the coupon collector problem, for which we retrieve the classical formula due to Erdos [10], rather effortlessly.

3. In Subsection 4.3, we show that t_n can be computed numerically with a high precision, beating the computing effort needed to sample T_n by simulation.

4. Finally, Subsection 4.4 shows that efficient bounds on t_n can be found by using simple linear algebra properties.

4.1. Asymptotic expansion of t_n.

Let us recall that in the homogeneous and uniform case, t_n is the solution of

$$\frac{1}{n} = \alpha \exp(Qt_n/n)\mathbf{1}.$$

Without loss of generality, assume that any state can be reached from the support of the initial measure α (if some states are not reachable, they can be suppressed). All the lines in Q sum to non-positive numbers (with at least one line summing to a negative number), hence the eigenvalue of Q with the largest real part is a real negative number, that we denote $-\nu$. According to Theorem 2.7.2 of [17, Chapter 2], there exists a constant $a > 0$ and an integer $d \in \{1 \dots k\}$ such that:

$$\alpha \exp(Qt)\mathbf{1} = at^{d-1}e^{-\nu t}(1 + O(t^{-1})). \tag{12}$$

Together, (2) and (12) yield almost directly

$$t_n = \frac{n}{\nu}\log(n) + \frac{d-1}{\nu} n \log\log(n) + \frac{n(d-1)}{\nu}\log(\nu) - \frac{n}{\nu}\log a + o(n). \tag{13}$$

It should be noted that, as long as all states are reachable from the initial measure α, the only quantity in (12) that depends on the initial measure is a. In the asymptotic expansion (13), the initial measure α appears only in the $O(n)$-term but does not affect the first two terms in $n \log n$ and $n \log\log n$.

We call d the *degree* of ν. In linear algebraic terms, d corresponds to the dimension of the largest Jordan block in Q for eigenvalue $-\nu$, that can be reached under α. It is smaller or equal to the multiplicity of the largest eigenvalue of Q. The degree d can also be computed in the following way. Using an adequate permutation of the states, Q can be recomposed into a block upper triangular matrix.

$$Q = \begin{bmatrix} Q_{1,1} & Q_{1,2} & \cdots & Q_{1,\ell} \\ 0 & Q_{2,2} & \cdots & Q_{2,\ell} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Q_{\ell,\ell} \end{bmatrix} \tag{14}$$

where each $Q_{j,j}$ is an irreducible square matrix. We denote by $-\nu_j$ the maximal eigenvalue of the jth block $Q_{j,j}$. Since $Q_{j,j}$ is irreducible, $-\nu_j$ is real, negative, and simple by the Perron-Frobenius Theorem.

We say that block $Q_{j,j}$ is *communicating* with block $Q_{v,v}$, $v > j$, if the block $Q_{j,v}$ is not null. By definition, the communication graph of Q is acyclic. Let us *tag* the components whose maximal eigenvalues are maximal among all $-\nu_j$'s. The degree d of Q is defined as the maximal number of tagged components over any path in the communication graph.

To illustrate this definition, let us consider the transition graph of Figure 1 and assume that $-\nu_2 = -\nu_3 = -\nu_4 = -\nu$ are maximal. The degree is $d = 3$ because the path $Q_{1,1} \to Q_{2,2} \to Q_{3,3} \to Q_{4,4}$ contains 3 tagged components.

4.2. Classical coupon collector and double-dixie problems.

As mentioned in the introduction, the coupon collector can be seen as the simplest instance of our problem. Here, we use it as an illustration of how to use Theorems 3.1

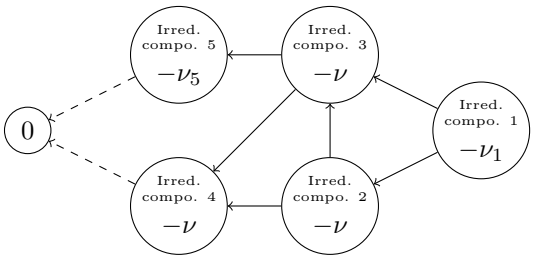
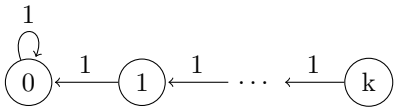


FIGURE 1: Example of a communication graph over the irreducible components of a matrix Q. Each component j has a maximal eigenvalue $-\nu_j$ that is simple. The dashed lines to 0 are just a reminder that the sums over some lines in Q are negative, corresponding to a transition to the absorbing state (0) in matrix P.

and 3.2. It shows the usefulness of our approach: the well-known formulas for the classical coupon collector are obtained directly.

We consider the classical coupon collector problem: there are n different types of coupon. At each time step, a coupon of type i is picked at random, uniformly among all coupon types, until k coupons of each kind are collected. It is proved in [21] that the expectation of T_n (average time to collect k coupon of each type is bounded by $n(\log n+(k-1)\log\log n+O(1))$. In this section, we show that our approach allows one to retrieve this result directly and even get the more precise statement given in [10]: We also provide the asymptotic distribution of T_n and the asymptotics of its moments.

At time 0, all chains start in state k: $\alpha = (0,\dots,0,1)$. The transition graph of each coupon type is



This is certainly the simplest chain with an absorbing state and, in isolation, the absorbing time for each component is trivial and is equal to k. This case is homogeneous (all coupon types have the same transition matrix) and uniform (uniform choice among all coupon types).

The matrix Q (defined in (1)) corresponding to this Markov chain is a $k\times k$ matrix that has -1 on its diagonal and 1 on its sub-diagonal. It has one eigenvalue $-\nu=-1$ that has a degree k. In the homogeneous and uniform case, the ODE (3) becomes

$$\begin{cases} \dfrac{d}{dt}M_k(t) &= -\dfrac{1}{n}M_k(t) \\ \dfrac{d}{dt}M_i(t) &= -\dfrac{1}{n}M_i(t)+\dfrac{1}{n}M_{i+1}(t) \qquad \text{for } 0<i<k \end{cases}$$

with $M_k(0)=1$ and $M_i(t)=0$ for $i\in\{0\dots k-1\}$.

$M_0(t)$ is the cumulative distribution function of an Erlang variable of parameter $(k,1)$ (*i.e.* the sum of k *i.i.d.* exponential variables of parameter 1) which can be written:

$$M_0(t)=1-\sum_{i=0}^{i-1}\frac{(t/n)^i e^{-t/n}}{i!}=1-\frac{(t/n)^{k-1}e^{-t/n}}{(k-1)!}+O((t/n)^{k-2}e^{-t/n}).$$

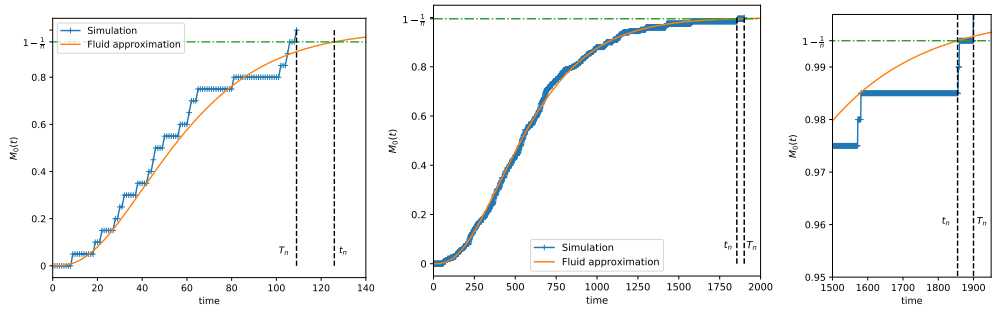
The solution in t of $\frac{(t/n)^{k-1}e^{-t/n}}{(k-1)!} = 1/n$ is $-n(k-1)W_{-1}\left(\frac{-1}{k-1}\left(\frac{(k-1)!}{n}\right)^{1/n}\right)$, where W_{-1} is the second branch of the Lambert W function ($W_{-1}(z)$ is the unique solution of $W_{-1}(ze^z) = z$ for $z \leq -1$). Since $W_{-1}(x) = \log(-x) - \log(-\log(-x)) + o(1)$ when $x < 0$ approaches 0, one gets $t_n = n(\log n + (k-1)\log\log n - \log((k-1)!)) + o(n)$. Using Theorems 3.1, and 3.2, this shows that the time T_n to collect k coupons of each type satisfies:

$$\lim_{n\to\infty} \mathbb{P}\left(T_n \leq n[\log n + (k-1)\log\log n + x]\right) = e^{-\frac{e^{-x}}{(k-1)!}}$$
$$\mathbb{E}\left[T_n\right] = n\log n + (k-1)n\log\log n + n(\gamma - \log((k-1)!) + o(n)$$
$$\text{var}\left[T_n\right] = \frac{\pi^2 n^2}{6} + o(n^2).$$

The first two results were first established in [10]. For the special case $k = 1$, a slightly stronger form of the third formula has been proven is [5], where the $o(n^2)$ is shown to be $O(n^2 \log\log\log n/\log\log n)$. In the case of general k, the third formula had been conjectured in [8].

It is interesting to notice that the variance of T_n does not depend on k when n is large, and more generally that the asymptotic distribution of T_n is Gumbel and its shape does not depend on k.

To illustrate this result and give a grasp on the respective behaviors of T_n and t_n with n, we ran several simulations of the coupon collector (with $k = 3$) and display the corresponding values of T_n and t_n. Figure 2 shows the empirical proportion of coupons collected up to time t, $M_0(t)$ and its fluid approximation $1 - m(t)$ (using the notations of Section 2.1). Although these plots only show a single trajectory and its fluid limit, they are rather typical. They illustrate the fact that T_n and t_n are very close to each other.

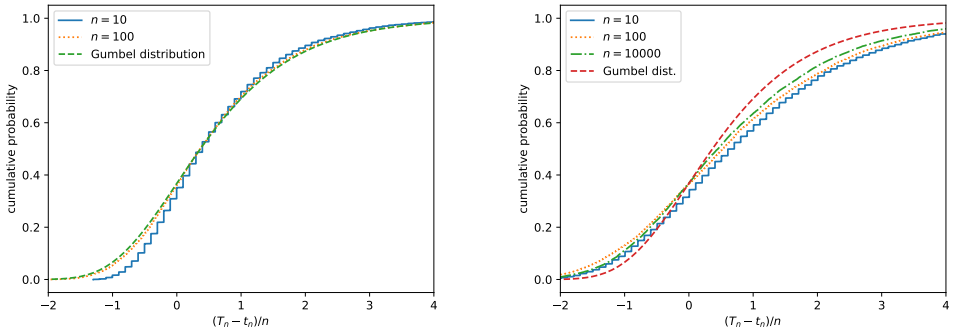


(a) $n = 20$ (b) $n = 200$ (c) $n = 200$ (zoom)

FIGURE 2: Illustration of the fraction of the n coupons that have been collected $k = 3$ times as a function of time. We also compare the stochastic hitting time T_n with the fluid approximation t_n for $n = 20$ and $n = 200$.

We have also computed the empirical distribution of T_n and tested it against its limit Gumbel distribution. Figure 3 shows the cumulative empirical distribution of T_n as well as the cumulative distribution of a Gumbel distribution with different values

of n. In the case $k = 1$, the convergence appears to be quick and the cumulative distribution of $(T_n - t_n)/n$ seems to coincide with the one of the Gumbel distribution for $n \geq 100$. For $k = 5$, even if one can convince oneself that there is convergence to the limit distribution, this convergence appears to be slow.



$k = 1$ $k = 5$

FIGURE 3: Empirical cumulative distribution of $(T_n - t_n)/n$ and its limit Gumbel distribution.

4.3. Numerical methods for t_n: an example with non-uniform probabilities.

The case when probabilities to get the different coupons are not uniform is also well studied (see for example [3] and references therein). Most papers provide bounds on the collection time. When the probability vector $(p_1, \ldots, p_n)$ satisfy Assumptions (6) and (7)), Theorem 3.2 shows that $\mathbb{E}[T_n] = t_n + n\gamma/\nu + o(n)$, where t_n satisfies

$$\sum_{i=1}^{n} \alpha \exp(p_i Q t_n)\mathbf{1} = 1. \tag{15}$$

There is in general no closed-form formula for t_n. Yet, Equation (15) is easy to solve numerically (using a classical Newton's method for example) with a very good precision. This can be done by using a numerical computation of the exponential of the matrix Q. Similarly, the value of the largest eigenvalue of Q can be computed numerically. This numerical computation provides an alternative method to estimate T_n, much faster than the repetition of simulations.

We illustrate this method and show its accuracy by computing the time to collect k coupons that have non-uniform probabilities, under the two following probability distributions, that we call $p^{\log}$ and p^{square}. These distributions $p^{\log}$ and p^{square} select the chain i with probabilities respectively proportional to $(\log(1+i))^{-1}$ and i^{-2}. For a system with n chains, these probabilities are:

$$p_i^{\log} = \frac{(\log(1+i))^{-1}}{\sum_{j=1}^{n}(\log(1+j))^{-1}}$$

$$p_i^{\text{square}} = \frac{i^{-2}}{\sum_{j=1}^{n} j^{-2}}$$

We implemented a simulator of the non-uniform coupon collector and a numerical algorithm that computes the quantity $t_n + \gamma/p_n$. We compare the value $\mathbb{E}[T_n]$ (com-

puted by averaging 1000 runs of the simulator) with the quantity $t_n + \gamma/p_n$. We report the results in Table 1 for three different probability distribution: uniform ($p_i = 1/n$), log ($p_i \propto 1/\log(1+i)$) and square ($p_i \propto 1/i^2$). We observe that in all cases, there is a good match between the approximation and the simulation. Also, as expected, the average hitting time $\mathbb{E}[T_n]$ is the smallest for the uniform distribution and is the largest for the square distribution.

The time taken by our simulator to compute 1000 simulations for $n = 50$ and $k = 5$ is about ten minutes. In contrast, it takes less than a second to compute the value t_n with a non-optimize implementation of the numerical algorithm that uses a dichotomic search and the standard matrix exponential function of `python numpy` to compute t_n. The most costly operation of our implementation is the evaluation of the left hand side of Equation (15) that requires the computation of n exponentials of $k \times k$ matrices. Moreover, by Theorem 3.2, the variance of the simulation for the square distribution grows as n^2, which leads to confidence intervals that grow with n. This makes the numerical computation of t_n much faster than the simulation for large values of n.

	$n = 10$		$n = 20$		$n = 50$	
	Simulation	$t_n + \frac{\gamma}{p_n}$	Simulation	$t_n + \frac{\gamma}{p_n}$	Simulation	$t_n + \frac{\gamma}{p_n}$
Unif ($k = 5$)	29.4 ± 0.2	28.8	71.8 ± 0.5	71.5	224.2 ± 1.2	224.5
log ($k = 1$)	35.8 ± 1.0	35.9	88.3 ± 2.0	90.2	275.5 ± 5.0	277.6
square ($k = 1$)	247.6 ± 9.3	271.5	1255 ± 39	1346	10590 ± 262	10950
Unif ($k = 5$)	89.4 ± 0.4	85.7	199.7 ± 0.7	194.6	569.8 ± 1.8	557.9
log ($k = 5$)	116.3 ± 1.8	109.9	257.3 ± 3.2	251.1	720.9 ± 7.8	704.0
square ($k = 5$)	922 ± 18	834.5	4209 ± 77	3987	31277 ± 455	30214

TABLE 1: Coupon collector and non-uniform probabilities: Comparison between the approximation $t_n + \gamma/p_n$ and the empirical average of $\mathbb{E}[T_n]$ computed over 1000 runs of the simulator. The value $\pm x$ indicate the 95% confidence intervals on the mean.

4.4. Bounding T_n by using T_1: Example of erasure channels.

The general formula for t_n involves the largest eigenvalue of Q. For the classical coupon collector, this eigenvalue is easy to compute in closed form. This is not true in general. As for non-uniform probabilities, a numerical algorithm can always be used to compute t_n. In this section, we provide an alternative by showing that an upper bound on t_n can be computed by using T_1, the hitting time of 0 for a single component. This quantity is often much easier to compute.

We consider the system made of a single component. The quantity $\mathbb{E}[T_1 \mid X_1(0) = x]$ is the expected hitting time of 0 starting from a state x. We define the quantity $\mathbf{T}_1^{\max}$ to be the largest expected hitting time, over all possible initial states x:

$$\mathbf{T}_1^{\max} = \max_{x \in \{1 \dots k\}} \mathbb{E}\left[T_1 \mid X_1(0) = x\right]. \tag{16}$$

Our next result uses the fact that $\mathbf{T}_1^{\max}$ is an upper bound on $-1/\nu$, with possible equality only if the degree is one. It implies that in particular that the expectation of the hitting time T_n satisfies

$$\mathbb{E}\left[T_n\right] \leq \mathbf{T}_1^{\max}(n \log n + n\gamma) + o(n). \tag{17}$$

This inequality is most of the time strict. For example, for the coupon collector with k coupons, $\mathbf{T}_1^{\max} = k$ and $t_n = n\log n + (k-1)n\log\log n + O(n) \ll kn\log n + O(n)$. Also, the inequality is in general strict, even when the matrix Q is irreducible. An example is the case of the erasure channel, shown in this section: The upper bound reported in Figure 4 is strict while the original matrix Q is irreducible.

Theorem 4.1. *Let $\mathbf{T}_1^{\max}$ be defined as in Equation (16). Then, in the case of uniform probabilities $p_i = 1/n$:*

$$t_n \leq \mathbf{T}_1^{\max} n\log n + o(n).$$

Proof. Theorem 3.2 combined with Equation (13) shows that $\mathbb{E}[T_n] = (1/\nu)n\log n + ((d-1)/\nu)n\log\log n - n\log(a\nu^{1-d})/\nu + n\gamma/\nu + o(n)$, where $-\nu$ is the largest eigenvalue of Q and d is the degree of the eigenvalue $-\nu$. We now show that $1/\nu \leq \mathbf{T}_1^{\max}$ with possible equality only in a special case, treated later.

As Q is non-singular, we have:

$$\begin{aligned}\mathbb{E}\left[T_1 \mid X_1(0) = x\right] &= \sum_{t=0}^{\infty} \mathbb{P}\left(T_1 \geq t \mid X_1(0) = x\right) \\ &= \sum_{t=0}^{\infty}\sum_{j=1}^{k}(I+Q)^t_{xj} = -\sum_{j=1}^{k}(Q^{-1})_{xj}.\end{aligned}$$

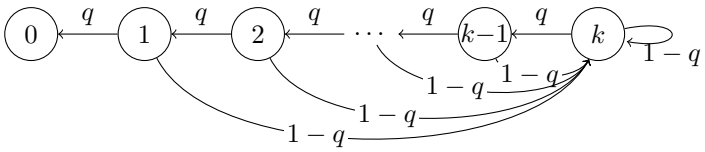
The largest eigenvalue of Q^{-1} is $-1/\nu$. Moreover, the term Q^{-1}_{xj} is a sum of non-negative terms and is therefore non-negative. By Perron-Frobenius theorem, this implies that $-1/\nu \leq \max_x \sum_j -Q^{-1}_{xj} = \mathbf{T}_1^{\max}$.

If the inequality is strict, we get the desired result, namely $t_n \leq \mathbf{T}_1^{\max} n\log n + o(n)$.

Now, the inequality becomes an equality only when the eigenvalue $-1/\nu$ of Q^{-1} admits $\mathbf{1}$ as an eigenvector. This implies that $\exp(Qt)$ also admits $\mathbf{1}$ as an eigenvector for eigenvalue $\exp(-\nu t)$. In this case, the quantity $\alpha\exp(Qt)\mathbf{1}$ used in the definition of t_n can be computed exactly and is equal to $\alpha\exp(-\nu t)\mathbf{1} = \exp(-\nu t)$, because $\alpha\mathbf{1} = 1$. Therefore t_n admits an exact form, $t_n = \frac{1}{\nu}n\log n = \mathbf{T}_1^{\max} n\log n$.

□

Let us illustrate this theorem with the following chain for each component:



This is a natural model of the lifetime of an *erasure channel* system. An erasure channel transmits bits and each of them gets erased with probability p. The state of the channel is the current number of consecutive bits that have been erased. When this number reaches some value k (that depends on the forward error correction that is used), the message transmitted on the channel cannot be corrected anymore and the channel is declared faulty. Here we consider a communication system made of n parallel independent erasure channels and we look at its lifetime: When all channels become faulty, the communication stops.

For this example, the quantity $\mathbf{T}_1^{\max}$ can be easily computed in closed form. Indeed, let $h_j := \mathbb{E}[T_1 \mid X_1(0) = j]$ be the expected hitting time of 0 starting from j. We have $h_0 = 0$ and for $j \in \{1, \dots, k\}$:

$$h_j = 1 + qh_{j-1} + (1-q)h_k.$$

One can verify that the vector h such that $h_j = \sum_{i=0}^{j-1} q^{i-k}$ is the unique solution of the above system of equations. This implies that $\mathbf{T}_1^{\max} = \max_{j\in\{0,\dots,d\}} h_j = \sum_{i=0}^{k-1} q^{i-k} = (q^{-k}-1)/(1-q)$. Hence, Theorem 4.1 implies that for this example:

$$\mathbb{E}\left[T_n\right] \leq \frac{q^{-k}-1}{(1-q)}(n\log n + \gamma) + o(n). \tag{18}$$

We implemented a simulator of the erasure model as well as a numerical algorithm to compute the asymptotically exact approximation $t_n + n\gamma/\nu$. We restrict our comparison to the case of uniform probabilities. To compute numerically t_n, we use a dichotomic algorithm similar to the one of Section 4.3 and we use functions from `numpy` standard library to compute the eigenvalue $-\nu$. In Figure 4, we compare numerically the bound provided by Equation (18), with the asymptotically exact bound of Theorem 3.2. We report the results for $n = 10$. We also computed the results for larger values of n. They are similar and not reported in the paper.

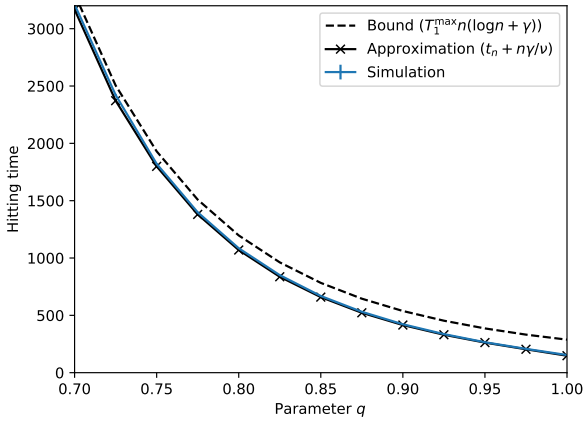


FIGURE 4: Erasure channel: Comparison of an estimation of $\mathbb{E}[T_n]$ computed by simulation, the asymptotic result of Theorem 3.2 and the upper bound of Theorem 4.1 as a function of the success probability p. The simulation results are averages over 10^4 values. The confidence intervals on the mean are plotted on the figure but they are too small to be visible.

We observe that, for this model, the approximation $t_n + n\gamma/\nu$ is very close to the value obtained by simulation, even when $n = 10$. Second, as expected, $T_1^{\max}(n\log n + \gamma)$ is an upper bound on this value. For large values of q, this upper bound is loose (roughly twice as large as $t_n + n\gamma/\nu$). This bound becomes tighter as q decreases (about 3% off the correct value when $q = 0.7$).

5. Coupling time of random walks.

In general, computing coupling times of Markov chains is useful because they provide bounds on their mixing times, [18]. They also correspond to the average time complexity of perfect simulation algorithms [22]. As for random walks, they are often used to model discrete event systems, such as queuing networks. Here we consider a random walk on a finite grid $\{1,\dots,k\}^n$. We first analyze a non-lazy random walk. It is usually more difficult to deal with coupling times in non-lazy random walks, while it makes no difference with our approach.

The random walk is defined as follows. From a position $(x_1,\dots,x_i,\dots,x_n)$ the walker moves to a neighboring position $(x_1,\dots,x_i+1,\dots,x_n)$ with probability q_i and to $(x_1,\dots,x_i-1,\dots,x_n)$ with probability r_i. On the boundaries, with probability q_i (resp. r_i) the walker remains at the same position if $x_i = k$ (resp. $x_i = 1$). In the following we will consider several possible values of (q_i) and (r_i) satisfying assumptions (6) and (7). The random walk in dimension two is displayed in Figure 5.

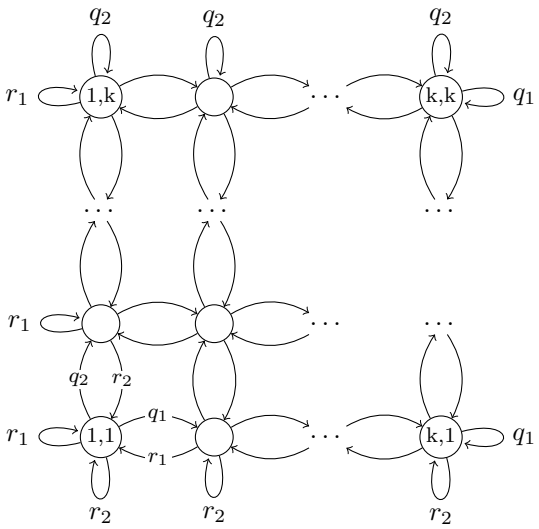


FIGURE 5: Random walk in two dimensions with borders.

We consider a coupling of k^n versions $(\mathbf{X}_x)_{x\in\{1\dots k\}^n}$, of the chain $\mathbf{X}$, one per possible initial state, under which all chains take the same transitions. We are interested in computing the first time T_n such that the state of the system does not depend on the initial state, *i.e.*, such that for any initial state x and y: $\mathbf{X}_x(T_n) = \mathbf{X}_y(T_n)$ (note that by assumption on the coupling, this implies that $\mathbf{X}_x(t) = \mathbf{X}_y(t)$ for $t \geq T_n$).

By monotonocity of the random walk, the coupling time T_n is equal to the coupling time of the two extreme chains $\mathbf{L}$ and $\mathbf{U}$ starting respectively in the extreme points, namely $\mathbf{L}(0) = (1,\dots,1)$ and $\mathbf{U}(0) = (k,\dots,k)$. Another property of this coupling is that whenever coalescence occurs in one coordinate, say i, $(L_i(t_0) = U_i(t_0))$ then this partial coalescence remains: $\forall t \geq t_0$, $L_i(t) = U_i(t)$.

We consider the Markov chain $\mathbf{X} := (\mathbf{L},\mathbf{U})$. Each component, corresponding to one coordinate, $X_i := (L_i, U_i)$, is selected with probability $p_i := q_i + r_i$. Once a coordinate

has been chosen, the two walks U_i and L_i take the same step: They both increase (resp. decrease) their i-th coordinate with probability q_i/p_i (resp. r_i/p_i), if possible. Each component has two absorbing states corresponding to the coalescence of L_i and U_i, namely $(1,1)$ and (k,k). Note that it makes no difference with the case with a single absorbing state because one can always merge all absorbing states into one.

All this implies that this chain fits in our general framework by using the transition matrix for each component depicted in Figure 6. At time 0, the difference is maximum: $(L_i(0), U_i(0)) = (1,k)$. If the chain is a state (a,b) (with $a < b$), it jumps to the state $(a+1, \min(b+1,k))$ with probability q_i/p_i and to the state $(\max(a-1,0), b-1)$ with probability $1 - q_i/p_i$.

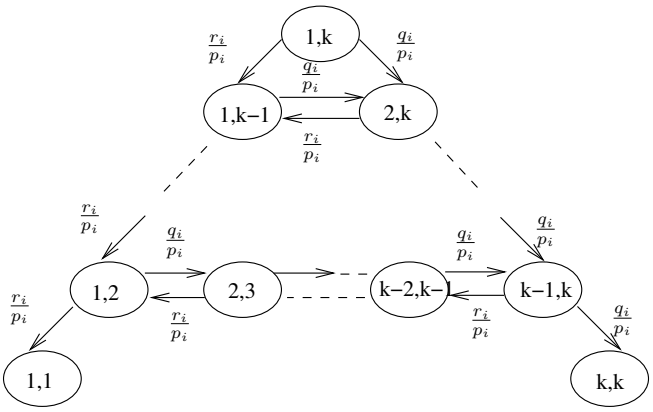


FIGURE 6: transition matrix of the chain $X_i = (L_i, U_i)$

5.1. Homogeneous and uniform case.

Let us first consider the homogeneous and uniform case. In our context, this means that q_i and r_i do not depend on i. Of course this implies that $p_i = q_i + r_i = 1/n$. However homogeneity and uniformity do not imply that $q_i = r_i$. Therefore, the walk can still have a drift. In the rest of this subsection, the index i is useless and will be dropped.

The asymptotics of T_n depend on the largest eigenvalue of Q. As seen in Figure 6, this matrix is block lower triangular because the difference $U - L$ cannot increase. Therefore, the largest eigenvalue of Q is the largest eigenvalue of one of its block diagonal matrices. It should be clear that the block with the largest eigenvalue is the largest block, corresponding to the level $U_i - L_i = 1$ in Figure 6. The sub-matrix of Q corresponding to this $(k-1) \times (k-1)$ block is

$$A = \begin{bmatrix} -(c+d) & c & & & & \\ d & -(c+d) & c & & & \\ & d & \ddots & \ddots & & \\ & & \ddots & \ddots & c & \\ & & & d & -(c+d) \end{bmatrix}, \tag{19}$$

with $c = nq$ and $d = nr$.

It is shown in [15] that the eigenvalues of this matrix are $-(d+c)+2\sqrt{cd}\cos(j\pi/k) = -1+2\sqrt{n^2qr}\cos(j\pi/k), \quad j \in \{1,\dots,k\}$. Hence, the eigenvalue with the largest real part is $-1+2n\sqrt{qr}\cos(\pi/k)$ and its multiplicity is 1.

- When $p \neq q$, Theorem 3.2, together with $\cos(\pi/k) \leq 1$, imply that the coupling time T_n satisfies

$$\begin{aligned}\mathbb{E}[T_n] &= \frac{1}{1-2n\sqrt{qr}\cos(\pi/k)} n\log n + O(n)\\ &\lesssim \frac{1}{1-2\sqrt{cd}} n\log n + O(n),\end{aligned}$$

  where $c := nq$ and $d := nr$ are the normalized probabilities that do not depend on n. Note that the approximation $\cos(\pi/k) \approx 1$ becomes more accurate as k grows.

- The case $q = r = \frac{1}{2n}$ corresponds to the random walk with no drift. In this case one can use the second order Taylor expansion $1-\cos(\pi/k) \approx \pi^2/(2k^2)$. This implies

$$\begin{aligned}\mathbb{E}[T_n] &= \frac{1}{1-\cos\left(\frac{\pi}{k}\right)} n\log n + O(n)\\ &\approx \frac{2k^2}{\pi^2} n\log n + O(n).\end{aligned}$$

At this point, one can notice that this approach provides an equivalent to the coupling time rather than an upper bound, as usually done in the literature [18]. Also note that the asymptotic behavior of the random walk with a drift is very different from the case with no drift: in the drift case, the asymptotic coupling time does not depend on k, the size of the grid, while in the no-drift case, the coupling time grows as k^2.

5.2. Heterogeneous cases.

We first consider the heterogeneous case with uniform choices, *i.e.*, $q_i + r_i = 1/n$ so that the coordinate i is chosen with probability $1/n$. To comply with conditions (6) and (7), and with the assumptions of Theorem 3.3, we assume that q_i is of the form c_i/n for all i (where c_i is a constant). Uniformity implies that $r_i = (1-c_i)/n$.

By applying the results of Theorem 3.3 and the general form of the eigenvalues of the matrix of Equation (19), there exists $\bar{c} \in [\min c_i, \max c_i]$ such that the coupling time is:

$$\mathbb{E}[T_n] = \frac{1}{1-2\sqrt{\bar{c}(1-\bar{c})}\cos\left(\frac{\pi}{k}\right)} n\log n + O(n).$$

Let us now consider a random walk with heterogeneous as well as non-uniform transitions. This is the case when, for example, for all i, $p_i = c_i/n$ and $q_i = d_i/n$ but $c_i + d_i$ is not necessarily equal to 1. In that case, Theorem 3.3 says that there exist $\bar{c} \in [\min c_i, \max c_i]$ and $\bar{d} \in [\min d_i, \max d_i]$ such that the coupling time is:

$$\mathbb{E}[T_n] = \frac{1}{\bar{d}+\bar{c}-2\sqrt{\bar{c}\bar{d}}\cos\left(\frac{\pi}{k}\right)} n\log n + O(n).$$

This result calls for several comments.

- First, the behavior of the heterogeneous and/or non-uniform case is very similar to the homogeneous and uniform random walk with a drift: when n is large: $\mathbb{E}[T_n] \approx Cn\log n$ where C is a constant that does not depend on k, the size of the state space.
- Up to our knowledge, this is the first time an asymptotic of the coupling of a random walk with general probabilities (not dependent on the position) is computed in closed form.

5.3. Lazy random walk on the torus.

Other random walks can be analyzed with this approach. In this section, we compute the coupling time of the uniform lazy chain on the torus $\{0\dots k-1\}^n \mod (k,\dots,k)$, analyzed in Chapter 5 of [18]. At each step, the lazy random walk remains at its current position with probability $1/2$. With probability $1/2$, one coordinate i is picked and $X_i(t+1) = X_i(t)+1$, or $X_i(t)-1$, with probability $1/2$. We show that the coupling time of the lazy chain is similar to the previous one.

We consider the following coupling of two lazy random walks $\mathbf{X}$ and $\mathbf{Y}$ that start at x and y respectively. At each step, we first pick one of the n coordinates at random. If the positions of the two walks agree in the chosen coordinate, we move both of the walks by $+1, -1$ or 0 in that coordinate, with respective probabilities $1/4, 1/4$ and $1/2$. If the positions of the two walks differ in the chosen coordinate, we randomly choose one of the chains to move, leaving the other fixed. We then move the selected walk by $+1$ or -1 in the chosen coordinate, with probability $1/2$.

The coupling time of the chain is the first time when $\mathbf{X}$ and $\mathbf{Y}$ agree in all coordinates. For each coordinate, the clockwise difference $\min\{(X_i - Y_i) \mod k, (Y_i - X_i) \mod k\}$ has the transition matrix displayed in Figure 7.

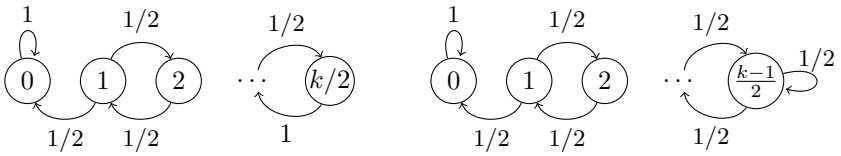


FIGURE 7: transition matrix of the chain $\min\{(X_i - Y_i) \mod k, (Y_i - X_i) \mod k\}$. The cases when k is even (left) and k is odd (right) are different. However, they both correspond to the transition matrix A folded in two.

This transition matrix is the same as A, the largest diagonal block of the previous chain, given in (19), but with size k instead of $k-1$, and folded in two by the middle. Therefore, the coupling time of the lazy walk on the torus is asymptotically the same:

$$\mathbb{E}[T_n] = \frac{1}{1-\cos\left(\frac{\pi}{k+1}\right)} n\log n + O(n) \approx \frac{2(k+1)^2}{\pi^2} n\log n + O(n).$$

This is a more precise statement than the results given in [18] that only give bounds.

6. Conclusions and Future work.

In this paper, we have shown how fluid approximations can be used to get precise and efficient estimations of the absorbing time of discrete time Markov chains made of several components. Our approach allows us to obtain closed form results or numerical methods that have shown to be accurate as the number of chains n goes to infinity. We applied this approach to get closed form formulas for the coupon collector, the lifetime of erasure channels and coupling times in finite random walks.

One of the main features of our model is the fact that the components are independent of each other (up to the mutual exclusion). One could think of extending our approach to interacting components, with transition matrices that depend on the state of other components. We believe that the absorbing time of all the components can still be computed using a similar approach. One natural application of this would be the computation of the time of extinction in several population dynamics problems.

Acknowledgements

We would like to thank Agnès Coquio who suggested the use of Poissonization and carefully read a preliminary version of this paper.

This project was partially supported by the EU project QUANTICOL 600708.

Appendix A. Proof of Theorem 3.1

We use the notations introduced in the paper before Theorem 3.1, without recalling them here. Note that to ease the notations, we avoid to write explicitly the dependence of p on n and write $(p_1 \dots p_n)$ instead of $(p_1^{(n)} \dots p_n^{(n)})$.

For a fixed n, the distribution p_i satisfies $p_1 \geq p_2 \cdots \geq p_n$. Let $f(t) = \alpha \exp(Qt)\mathbf{1}$. The function f is non-increasing and positive. Hence, for all ℓ, by definition of t_n in Equation (5), we have:

$$1 = \sum_{i=1}^{n} f(p_i t_n) \geq \sum_{i=n-\ell+1}^{n} f(p_i t_n) \geq \ell f(p_{n-\ell} t_n).$$

This shows that for all ℓ, $f(p_{n-\ell}t_n) \leq 1/\ell$. By Assumption (7), the quantity $p_{n-\ell}/p_n$ converges to 1 as n goes to infinity. This implies that $f(p_n t_n)$ converges to 0 and therefore that $p_n t_n$ converges to infinity as n grows. Moreover, by Equation (12), there exists a constant a and an integer $d \geq 1$ such that

$$f(t) = a t^{d-1} e^{-\nu t}(1 + O(t^{-1})). \tag{20}$$

Let $\tilde{g}_n(x) = \mathbb{P}(\widetilde{T}_n \leq t_n + x/p_n)$. $\tilde{T}_n$ is the maximum of independent variables. Hence

$$\log(\tilde{g}_n(x)) = \log \prod_{i=1}^{n} (1 - f(p_i(t_n + x/p_n))) = \sum_{i=1}^{n} \log(1 - f(p_i(t_n + x/p_n))).$$

By using $-\log(1-s) = s + o_{s\to 0}(s)$ and Equation (20), we have:

$$\begin{aligned}
-\log \tilde{g}_n(x) &= \sum_{i=1}^{n} f(p_i(t_n + x/p_n)) + o(1) \\
&= a\sum_{i=1}^{n} (p_i(t_n + \frac{x}{p_n}))^{d-1} e^{-\nu p_i(t_n+\frac{x}{p_n})} + o(1) \\
&= a\sum_{i=1}^{n} (p_i t_n)^{d-1} e^{-\nu p_i(t_n+\frac{x}{p_n})}(1 + \frac{x}{t_n p_n})^{d-1} + o(1) \\
&= a\sum_{i=1}^{n} (p_i t_n)^{d-1} e^{-\nu p_i t_n} e^{-\nu x\frac{p_i}{p_n}} + o(1). \qquad (21)
\end{aligned}$$

Assume that $x \geq 0$. The proof for $x \leq 0$ is similar, by inverting most of the signs "$\geq$" and "$\leq$". As $p_i \geq p_n$, we have $e^{-\nu x p_i/p_n} \leq e^{-\nu x}$.

By definition of t_n and Equation (20), we have $a\sum_{i=1}^{n}(p_i t_n)^{d-1}e^{-\nu p_i t_n} = 1 + o(1)$. This implies

$$-\log \tilde{g}_n(x) \leq e^{-\nu x} + o(1).$$

Let $\varepsilon > 0$. Let ℓ_n be such that $p_i/p_n \geq (1+\varepsilon)$ if and only if $i \leq n - \ell_n$ (which implies that $e^{-\nu x p_i/p_n} \geq e^{-\nu x(1+\varepsilon)}$ for $i \geq \ell_n$). The sum in Equation (21) is greater than the sum of the last ℓ_n terms. Hence:

$$\begin{aligned}
-\log \tilde{g}_n(x) &\geq a \sum_{i=n-\ell_n+1}^{n} (p_i t_n)^{d-1} e^{-\nu p_i(t_n+\frac{x}{p_n})} + o(1) \\
&= a \sum_{i=n-\ell_n+1}^{n} (p_i t_n)^{d-1} e^{-\nu p_i t_n} e^{-\nu x\frac{p_i}{p_n}} + o(1) \\
&\geq a \sum_{i=n-\ell_n+1}^{n} (p_i t_n)^{d-1} e^{-\nu p_i t_n} e^{-\nu x(1+\varepsilon)} + o(1) \\
&= e^{-\nu x(1+\varepsilon)} - e^{-\nu x(1+\varepsilon)} a \sum_{i=1}^{n-\ell_n} (p_i t_n)^{d-1} e^{-\nu p_i t_n} + o(1) \qquad (22)
\end{aligned}$$

where the last equality comes from the definition of t_n.

Therefore, to show the result, it suffices to show that the second term in Equation (22) converges to 0 as n grows. By Assumption (6) and the definition of ℓ_n, we have $p_{i-\ell_n}/p_i \geq p_{n-\ell_n}/p_n \geq 1+\varepsilon$. Moreover, as $p_i t_n$ converges to infinity, we have $(\frac{p_{i-\ell_n}}{p_i})^{d-1} e^{-\nu t_n(p_{i-\ell_n}-p_i)} \leq (1+\varepsilon)^{d-1} e^{-\nu t_n p_n \varepsilon}$. Hence, the last factor of the second

term of Equation (22) is:

$$\begin{aligned} a\sum_{i=1}^{n-\ell_n}(p_i t_n)^{d-1}e^{-\nu p_i t_n} &= a\sum_{i=\ell_n}^{n}(p_{i-\ell_n} t_n)^{d-1}e^{-\nu p_{i-\ell_n} t_n} \\ &\leq a\sum_{i=\ell_n}^{n}(p_i t_n)^{d-1}e^{-\nu p_i t_n}(1+\varepsilon)e^{-\nu t_n p_n \varepsilon} \\ &\leq a\sum_{i=1}^{n}(p_i t_n)^{d-1}e^{-\nu p_i t_n}(1+\varepsilon)e^{-\nu t_n p_n \varepsilon} \\ &= (1+\varepsilon)e^{-\nu t_n p_n \varepsilon} + o(1). \end{aligned}$$

that converges to 0 as n goes to infinity.

The proof for $x \leq 0$ is similar, by inverting most of the signs "$\geq$" and "$\leq$".

We conclude by showing that T_n and $\tilde{T}_n$ are close (in probability). Indeed, as mentioned in Section 3.1, $\tilde{T}_n = Z_{T_n}$, where $Z_k = \sum_{i=1}^{k} E_i$ and E_i are *i.i.d* exponential random variable of mean 1. Applying Chebyshev's inequality shows that for all $c > 0$, $\mathbb{P}(|Z_k - k| \geq c) \leq \mathrm{var}[Z_k]/c^2 = k/c^2$. By conditioning on T_n, and using $\mathbb{E}[T_n] = \mathbb{E}[\tilde{T}_n]$, this implies that $\mathbb{P}(|\tilde{T}_n - T_n| \geq c) \leq \mathbb{E}[T_n]/c^2 = \mathbb{E}[\tilde{T}_n]/c^2$.

Let $\varepsilon > 0$. Applying the previous bound with $c = \varepsilon/p_n$ shows that:

$$\mathbb{P}\left(\tilde{T}_n \leq t_n + \frac{x-\varepsilon}{p_n}\right) - \frac{p_n^2\mathbb{E}[\tilde{T}_n]}{\varepsilon^2} \leq \mathbb{P}\left(T_n \leq t_n + \frac{x}{p_n}\right) \leq \mathbb{P}\left(\tilde{T}_n \leq t_n + \frac{x+\varepsilon}{p_n}\right) + \frac{p_n^2\mathbb{E}[\tilde{T}_n]}{\varepsilon^2}$$

In Step 2 of the proof of Theorem 3.2, we show that $\mathbb{E}[\tilde{T}_n] = t_n + O(1)$. This means that it suffices to shows that $\lim_{n\to\infty}(p_n)^2 t_n = 0$ to show the result.

To show this, we first remark that for any $x \geq 0$ and $\alpha \geq 1$, we have $e^{-\alpha x} \leq \alpha e^{-x}$. Hence, using Equation (20) and the fact that $p_n t_n$ converges to infinity, we have $f(t_n p_i) \leq (p_i/p_n) f(t_n p_n)$ for n large enough. This implies that

$$1 = \sum_{i=1}^{n} f(t_n p_i) \leq \sum_{i=1}^{n} \frac{p_i}{p_n} f(t_n p_n) = \frac{1}{p_n} f(t_n p_n).$$

Similarly to Equation (13), Equation (20) implies that $t_n \leq \frac{1}{\nu p_n}(\frac{1}{p_n} + (d-1)\log\log\frac{1}{p_n}) + O(1/p_n)$. This in particular implies that $\lim_{n\to\infty} t_n (p_n)^2 = 0$.

Appendix B. Proof of Theorem 3.2

The proof consists of three steps. We first obtain uniform upper and lower bounds on $g_n(x) := \mathbb{P}(\tilde{T}_n \leq t_n + x/p_n)$ on the intervals $[-p_n t_n/2; 0]$ and $[0; +\infty)$. We then use these bounds to show that for all $m \in \{1, 2 \dots\}$, the sequence $[p_n(\tilde{T}_n - t_n)]^m$ is uniformly integrable, which shows that the moments of $p_n(\tilde{T}_n - t_n)$ converge to the one of a Gumbel distribution. Finally, we show that the moments of $p_n(T_n - t_n)$ converge to the ones of $p_n(\tilde{T}_n - t_n)$.

Note that Step 1 and Step 2 of the proof of Theorem 3.2 concern the convergence of $\tilde{T}_n$ and do not use that T_n is close to $\tilde{T}_n$.

Step 1. Recall that $f(t) = at^{d-1}e^{-\nu t}(1+\epsilon(t))$ with $\lim_{t\to\infty}\epsilon(t) = 0$. By definition of t_n, this implies $1 = \sum_{i=1}^n f(p_i t_n) = \sum_{i=1}^n a(p_i t_n)^{d-1}e^{-\nu p_i t_n}(1+\varepsilon(p_i t_n))$. As $p_i \geq p_n$ and $\lim_{n\to\infty} p_n t_n = \infty$, this implies that

$$\sum_{i=1}^n a(p_i t_n)^{d-1}e^{-\nu p_i t_n} = 1+\varepsilon_n,$$

with $\lim_{n\to\infty}\varepsilon_n = 0$.

Moreover, for $x \in [0; p_n t_n/2]$, we have:

$$\begin{aligned} f(p_i(t_n - x/p_n)) &= a(p_i(t_n - x/p_n)^{d-1})e^{-\nu p_i(t_n - x/p_n)}(1+\epsilon(p_i(t_n - x/p_n))) \\ &\geq a(p_i t_n)^{d-1}e^{-\nu p_i t_n}e^{\nu x}2^{1-d}(1+\epsilon(p_i(t_n - x/p_n))), \end{aligned}$$

which implies that $\sum_{i=1}^n f(p_i(t_n - x/p_n)) \geq e^{\nu x}2^{1-d}(1+\tilde{\varepsilon}_n(x))$, where the functions $\tilde{\varepsilon}_n(x)$ satisfy $\lim_{n\to\infty}\sup_{x\in[0;p_n t_n/2]}\tilde{\varepsilon}_n(x) = 0$.

By using $\log(1-x) \leq -x$, this implies that $\tilde{g}_n(x) = \mathbb{P}(\widetilde{T}_n \leq t_n + x/p_n)$ satisfies

$$\begin{aligned} \tilde{g}_n(-x) = \prod_{i=1}^n (1 - f(p_i(t_n - x/p_n))) &\leq \exp(-\sum_{i=1}^n f(p_i(t_n - x/p_n))) \\ &\leq \exp(-e^{\nu x}2^{1-d}[1+\tilde{\varepsilon}_n(x)]) \end{aligned} \tag{23}$$

Similarly, by using the union bound, for $x > 0$, we have:

$$\begin{aligned} \tilde{g}_n(x) = \prod_{i=1}^n (1 - f(p_i(t_n + x/p_n))) &\geq 1 - \sum_{i=1}^n f(p_i(t_n + x/p_n)) \\ &\geq 1 - e^{-\nu x}(1+\hat{\varepsilon}_n(x)), \end{aligned} \tag{24}$$

with $\lim_{n\to\infty}\sup_{x>0}|\hat{\varepsilon}_n(x)| = 0$.

Let $e_n = \max(\varepsilon_n, \sup_{x>0}|\hat{\varepsilon}_n(x)|, \sup_{x\in[0,t_n p_n/2]}|\tilde{\varepsilon}_n(x)|)$, we have $\lim_{n\to\infty} e_n = 0$.

Step 2. Let $A > 0$ and $m \in \{1, 2\dots\}$. By Equation (24), we have

$$\begin{aligned} \mathbb{E}\left[\left(p_n\left|\tilde{T}_n - t_n\right|\right)^m \mathbf{1}_{p_n(\tilde{T}_n - t_n)\geq A}\right] &= \int_A^\infty \frac{x^{m-1}}{m}(1 - g_n(x))dx \\ &\leq \int_A^\infty \frac{x^{m-1}}{m}e^{-\nu x}(1+\hat{\varepsilon}_n(x))dx \\ &\leq (1+e_n)\int_A^\infty \frac{x^{m-1}}{m}e^{-\nu x}dx. \end{aligned} \tag{25}$$

Similarly, by Equation (23), we have

$$\begin{aligned} \mathbb{E}\left[\left(p_n\left|\tilde{T}_n - t_n\right|\right)^m \mathbf{1}_{\{p_n(\tilde{T}_n - t_n)\leq -A\}}\right] &= \int_A^{p_n t_n} \frac{x^{m-1}}{m}g_n(-x)dx \\ &= \int_A^{p_n t_n/2} \frac{x^{m-1}}{m}g_n(-x)dx + \int_{p_n t_n/2}^{p_n t_n} \frac{x^{m-1}}{m}g_n(-x)dx \\ &\leq \int_A^{p_n t_n/2} \frac{x^{m-1}}{m}\exp(-2^{1-d}e^{\nu x}(1-e_n))dx \qquad (26) \\ &\quad + \frac{(p_n t_n)^{m-1}}{m}g_n(-p_n t_n/2)\left(p_n t_n - \frac{p_n t_n}{2}\right) \qquad (27) \end{aligned}$$

It should be clear that Equation (25) and (26) are uniformly bounded in n by a bound that converges to 0 as A goes to infinity. Moreover (27) is equal to

$$\frac{(p_n t_n)^m}{2m} g_n\Big(-\frac{p_n t_n}{2}\Big) \leq a \frac{(p_n t_n)^m}{2m} \left(\frac{p_n t_n}{2}\right)^{d-1} e^{-\nu p_n t_n/2}(1+e_n). \tag{28}$$

When n goes to infinity, $p_n t_n$ goes to infinity, which implies that Equation (28) converges to 0 as n goes to infinity.

Combining Equations (25,26,28) shows that for all $m \geq 1$, the quantity $[p_n(\tilde{T}_n - t_n)]^m$ is uniformly integrable. By Theorem 3.1 this implies that all moments of $p_n(\tilde{T}_n - t_n)$ converges to the moment the Gumbel distribution of parameter ν.

Step 3. We conclude by showing that the moments of $p_n(\tilde{T}_n - t_n)$ and the ones of $p_n(T_n - t_n)$ are asymptotically equal.

Recall that $\tilde{T}_n$ is an Erlang variable of parameters $(T_n, 1)$. The mth moment of an Erlang distribution of parameters $(k,1)$ is $k(k+1)\dots(k+m-1) = k^m + O(k^{m-1})$.

$$\begin{aligned}
\mathbb{E}\left[(\tilde{T}_n)^m\right] &= \mathbb{E}\left[\mathbb{E}\left[(\tilde{T}_n)^m \mid T_n\right]\right] \\
&= \mathbb{E}\left[T_n(T_n+1)\dots(T_n+m-1)\right] &(29)\\
&= \mathbb{E}\left[(T_n)^m\right] + O(\mathbb{E}\left[(T_n)^{m-1}\right]) &(30)
\end{aligned}$$

Equation (29) implies that for $m \geq 1$:

$$\begin{aligned}
\mathbb{E}\left[[p_n(\tilde{T}_n - t_n)]^m\right] &= (p_n)^m \sum_{\ell=0}^{m} \binom{m}{\ell} \mathbb{E}\left[(\tilde{T}_n)^\ell\right] (-t_n)^{m-\ell} \\
&= (p_n)^m \sum_{\ell=0}^{m} \binom{m}{\ell} \mathbb{E}\left[T_n(T_n+1)\dots(T_n+\ell-1)\right] (-t_n)^{m-\ell} &(31)
\end{aligned}$$

By Step 2 of this proof, $\lim_{n\to\infty} \mathbb{E}[(\tilde{T}_n)^{m-1}]/\mathbb{E}[(\tilde{T}_n)^m] = 0$. Hence, using Equation (30), a direct recurrence on m shows that $\lim_{n\to\infty} \mathbb{E}[(T_n)^{m-1}]/\mathbb{E}[(T_n)^m] = 0$ which implies that $\lim_{n\to\infty} \mathbb{E}[(T_n)^m]/\mathbb{E}[(\tilde{T}_n)^m] = 1$. Combining this with Equation (31) shows that $\mathbb{E}[[p_n(\tilde{T}_n - t_n)]^m]$ and $\mathbb{E}[[p_n(T_n - t_n)]^m]$ are asymptotically equal.

Appendix C. Proof of Theorem 3.3

The proof is similar to the proof of Theorem 3.1. We only provide a sketch of proof. Up to replacing Q_i by $p_i Q_i$, by assumption (10), one can assume that the probabilities are uniform: $p_i = 1/n$.

We define $\tilde{g}$ similarly to the $\tilde{g}$ of the proof of Theorem 3.1:

$$\tilde{g}_{\theta_1\dots\theta_n}(x) = \mathbb{P}\left(T_{\theta_1\dots\theta_n} \geq t_{\theta_1\dots\theta_n} + x\right).$$

As in Equation (21), we have:

$$\log \tilde{g}(x) = -\sum_{i=1}^{n} (t_n/n)^{d-1} e^{-\nu_i t_n/n} e^{-\nu_i x} + o(1).$$

By the intermediate value theorem and the definition of t_n, this implies that there exists $\nu(\theta_1\dots\theta_n) \in [\nu_{\min}, \nu_{\max}]$ such that $\log \tilde{g}(x) = -e^{-\nu_{\theta_1\dots\theta_n} x} + o(1)$.

The proof is similar for the moments and makes again use of the intermediate value theorem.

References

[1] Anceaume, E., Busnel, Y. and Sericola, B. (2015). New results on a generalized coupon collector problem using markov chains. *J. Appl. Probab.* **52,** 405–418.

[2] Benaïm, M. and Boudec, J.-Y. L. (2008). A class of mean field interaction models for computer and communication systems. *Performance Evaluation* **65,** 823–838.

[3] Berenbrink, P. and Sauerwald, T. (2009). The weighted coupon collector's problem and applications. In *Proceedings of the 15th Annual International Conference on Computing and Combinatorics*. COCOON '09. Springer-Verlag. pp. 449–458.

[4] Boudec, J.-Y. L. (2013). The stationary behaviour of fluid limits of reversible processes is concentrated on stationary points. *Networks and Heterogeneous Media* **8,** 529–540.

[5] Brayton, R. K. (1963). On the asymptotic behavior of the number of trials necessary to complete a set with random selection. *Journal of Mathematical Analysis and Applications* **7,** 31 – 61.

[6] Chaintreau, A., Le Boudec, J.-Y. and Ristanovic, N. (2009). The age of gossip: Spatial mean field regime. *SIGMETRICS Perform. Eval. Rev.* **37,** 109–120.

[7] Chazottes, J.-R., Collet, P. and Méléard, S. (2016). Sharp asymptotics for the quasi-stationary distribution of birth-and-death processes. *Probability Theory and Related Fields* **164,** 285–332.

[8] Doumas, A. V. and Papanicolaou, V. G. (2014). The coupon collector's problem revisited: Generalizing the double dixie cup problem of Newman and Shepp. *arXiv preprint arXiv:1412.3626*.

[9] Doumas, A. V., Papanicolaou, V. G. et al. (2012). The coupon collector's problem revisited: asymptotics of the variance. *Advances in Applied Probability* **44,** 166–195.

[10] Erdös, P. and Rényi, A. (1961). On a classical problem of probability theory. *Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei* **6,** 215–220.

[11] Flajolet, P., Gardy, D. and Thimonier, L. (1992). Birthday paradox, coupon collectors, caching algorithms and self-organizing search. *Discrete Applied Mathematics* **39,** 207–229.

[12] Ganesh, A., Lilienthal, S., Manjunath, D., Proutiere, A. and Simatos, F. (2010). Load balancing via random local search in closed and open systems. *SIGMETRICS Perform. Eval. Rev.* **38,** 287–298.

[13] Gast, N. and Bruno, G. (2010). A mean field model of work stealing in large-scale systems. *SIGMETRICS Perform. Eval. Rev.* **38,** 13–24.

[14] Kang, W. and Ramanan, K. (2012). Asymptotic approximations for the stationary distributions of many-server queues. *Ann. Appl. Probab.* **22,** 477–521.

[15] Kulkarni, D., Schmidt, D. and Tsui, S.-K. (1999). Eigenvalues of tridiagonal pseudo-toeplitz matrices. *Linear Algebra and its Applications* **297,** 63–80.

[16] Kurtz, T. G. (1978). Strong approximation theorems for density dependent markov chains. *Stochastic Processes and their Applications* **6,** 223 – 240.

[17] Latouche, G. and Ramaswami, V. (1999). *Introduction to Matrix Analytic Methods in Stochastic Modeling*. Society for Industrial and Applied Mathematics.

[18] Levin, D. A., Peres, Y. and Wilmer, E. L. (2009). *Markov Chains and Mixing Times*. AMS.

[19] May, R. (2008). Coupon collecting with quotas. *the electronic journal of combinatorics* **15,** N31.

[20] Neal, P. (2008). The generalised coupon collector problem. *Journal of Applied Probability* 621–629.

[21] Newman, D. and Shepp, L. (1960). The double dixie cup problem. *American Mathematical Monthly* **67,** 58–61.

[22] Propp, J. G. and Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Rand. Struct. Alg.* **9,** 223–252.

[23] Shank, N. B. and Yang, H. (2013). Coupon collector problem for non-uniform coupons and random quotas. *the electronic journal of combinatorics* **20,** P33.