



**HAL**  
open science

# A Poisson limit theorem for reliability models based on Markov chains

James Ledoux

► **To cite this version:**

James Ledoux. A Poisson limit theorem for reliability models based on Markov chains. *Communications in Statistics - Theory and Methods*, 2006, 35 (1), pp.173-196. 10.1080/03610920500439331 . hal-00452069

**HAL Id: hal-00452069**

**<https://hal.science/hal-00452069>**

Submitted on 26 Aug 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A Poisson limit theorem for reliability models based on Markov chains

James Ledoux\*

November 24, 2005

## Abstract

In this article, we deal with a class of discrete-time reliability models. The failures are assumed to be generated by an underlying time inhomogeneous Markov chain. The multivariate point process of failures is proved to converge to a Poisson-type process when the failures are rare. As a result, we obtain a Compound Poisson approximation of the cumulative number of failures. A rate of convergence is provided.

**Key-Words** Discrete-time multivariate point process, time inhomogeneous Markov chain, Compound Poisson approximation, filtering

(AMS 2000) Primary : 60F99, 60G55, 60J10  
Secondary : 60F17, 60G42, 60K15, 93E11

## 1 Introduction

A basic issue in reliability modeling is what happens when the system becomes reliable. In other words, what becomes the model when the failure parameters tend to be smaller and smaller? We discuss such an issue for a class of reliability models which are based on discrete-time Markov chains. Specifically, we consider a failure process which is generated by the dynamics of a non-homogeneous Markov chain. Clumping of failures is also considered. Such a failure process may be thought of as a discrete-time multivariate point process. Then, we investigate its asymptotic distribution when the failure parameters converge to 0. As it can be expected, the asymptotic distribution is of Poisson-type. Under a condition on the rate of the convergence of the underlying non-homogeneous Markov chain to stationarity, we provide a rate of convergence for the distance in variation. As a result, we obtain a Compound Poisson approximation of the cumulative number of failures. The ergodic and irreducible cases for the limit Markov chain are discussed. The derivation

---

\*James Ledoux, INSA 20 avenue des Buttes de Coësmes, 35043 Rennes Cedex, FRANCE; email: james.ledoux@insa-rennes.fr

of the main results is based on elementary discrete-time stochastic calculus and on basic convergence results for non-homogeneous Markov chains reported in [HIV76, BDI77]. It is not intended to deal with the mildest assumptions for deriving Poisson limit theorems. The assumptions here, are expected to be easily checked and general enough to be of value in the reliability framework.

The problem of the Poisson approximation of a flow of rare events is a very old topic. Our purpose is not to review the huge literature on this topic. We limit ourselves to mentioning standard methods for deriving such Poisson approximations and only discuss the closest works to ours. The counting process associated with a discrete-time point process is nothing else but a sum of dependent random variables. There is a vast literature on limit theorems for such processes, especially on the central limit theorem and functional theorems. In a much more general setting, a complete account for such an “limit theorems approach” is in [JS89]. For a martingale approach of the Poisson approximation of point processes, some specific references are [Bro82, Bro83, KLS83] (see also the references therein). The basic tool is the compensator associated with a point process. An instance of Poisson approximation of a Poisson process driven by a Markov process is provided by Kabanov, Liptser and Shiryaev [KLS83] (see also [DK93]). Basically, the idea is to combine a bound on the distance in variation between the respective probability distributions of the point process and of an appropriate Poisson process, with estimates on a filtered Markov process. The bound is in terms of compensators. This “KLS approach” was used in [GL04] for deriving Poisson approximation of a complex continuous-time univariate point process for which the underlying Markov model was time-homogeneous. Here, we also use this approach in a discrete-time context. To the best of our knowledge, the present article and [GL04] are the only papers in which the “KLS approach” is made hard. The objective of this article is not only to derive Poisson limit theorems, but also to obtain a convergence rate under conditions that are expected to be easily checked. Moreover, we try to give as elementary as possible discrete-time oriented proofs, so that the statements only concern the Poisson approximation of random variables. However, we mention that our results are easily extended to a functional framework as reported in Remark 8.

At the present time, we can not discuss Poisson approximation without paying attention to the well-known Stein-Chen’s method for estimating the closeness of a probability distribution to the Poisson distribution (e.g. see [BHJ92] for an account on this approach). Basically, this method consists in using a bound on the deviation between the two probability distributions with respect to an appropriate distance. In general, the distance in variation is chosen. Barbour, Brown, Xia among other, have developed the Stein-Chen method to discuss Poisson process approximation of the distribution of point processes. In particular, they have combined the martingale method with the Stein-Chen’s method. We refer to [Xia00] and references therein for an account in this direction. For an effective computation of the bounds provided by Stein-Chen’s method, we have to know a form of local dependence or to produce couplings which is known to be a difficult task when dealing with complex point processes (e.g. see the recent paper [CX04] and references therein). Here, the computation of the bound involves the solution of an elementary fil-

tering problem. Thus, the “KLS approach” used here, is different from Stein-Chen’s one, and is quite effective for our class of sums of dependent random variables. However, note that our bounds depend on the length of the time interval on which the approximation is performed (see Remark 7).

The closest results to ours are those by Anisimov [Ani83]. Specifically, he obtained functional theorems for sums of conditionally independent random variables on a Markov chain that satisfied an uniform strong mixing condition. But, the method used in the present article is different from that used in [Ani83]. We also mention that Markov approximation for related processes have been obtained by Anisimov (e.g. see [Ani02] and references therein). In particular, limit theorems for a class of continuous-time bivariate Markov chains that can be represented as switching processes are derived from general limit theorems on switching processes. The switching processes have been studied by Anisimov and Korolyuk among others (see [Ani95] and references therein). This kind of results is relevant to our context, because we deal with a discrete-time counterpart of such a bivariate Markov process. Relevant references with Poisson approximation in reliability theory are also reported in [Kov94].

The article is organized as follows. The specific failure model is introduced in Section 2. In Section 3, we state the Poisson-type approximation of the corresponding multivariate failure point process. The cumulative number of failures is shown to have asymptotically a Compound Poisson distribution. Section 4 is devoted to the proof of our main results. We report some basic facts on the convergence of non-homogeneous Markov chains in Appendix A.

## 2 Basic facts on the model

### Main notation

- By convention, vectors are row vectors. Column vectors are denoted by means of the transpose operator  $(\cdot)^\top$ .
- $\mathbf{1}$  (resp.  $\mathbf{0}$ ) is a vector with each entry equals to one (resp. 0). Its dimension is defined by the context.
- For any  $x, y \in \mathbb{R}^n$ ,  $\langle x, y \rangle := x y^\top$ .
- For any vector  $x \in \mathbb{R}^n$ ,  $\|x\|_1$  is its  $l_1$ -norm, i.e.  $\|x\|_1 := \sum_{i=1}^n |x(i)|$ . The  $l_1$ -norm of any  $n \times n$ -matrix  $M$  is defined by  $\|M\|_1 := \max_i (\sum_{j=1}^n |M(i, j)|)$ . Note that  $\|M\|_1 = 1$  for a stochastic matrix  $M$ .  
Let us recall that  $\|M_1 M_2\|_1 \leq \|M_1\|_1 \|M_2\|_1$  and  $|\langle x, y \rangle| \leq \|x\|_1 \|y\|_1$  for any vectors  $x, y$  and matrices  $M_1, M_2$ .

## 2.1 The model

The model of the system is assumed to be a discrete time non-homogeneous Markov chain  $X := (X_t)_{t \in \mathbb{N}}$  on the finite state space  $\mathcal{U} = \{e_i, i = 1, \dots, N\}$ . The different states can represent the load on the system as it is usually done in performance evaluation, or some measure of performance level, or the active module of a software system, etc. This Markov chain is specified by its transition probability matrices  $P(t) := (P(t; i, j))_{i, j=1, \dots, N}$ ,  $t \geq 0$  and the probability distribution of  $X_0$ .  $P(t; i, j)$  is the probability that  $X$  jumps to state  $e_j$  at time  $t + 1$  given  $X$  is in state  $e_i$  at time  $t$ .

Assume now that the system is subject to failures. Let us describe the failure process. We distinguish two types of failure: the first one is associated with the visits to states, the second one with the transitions between states. When the model is in state  $e_i$ , a failure occurs with probability  $p_i$ , thus depending on the identity of the state. For the sake of simplicity, the delay to recover a safe state is neglected. Then, state  $e_j$  is entered with constant probability  $\alpha(i, j)$  (with  $\sum_{j=1}^N \alpha(i, j) = 1$ ). In some applications, it can be useful to associate failures directly with transitions. Thus, suppose that a failure does not occur during a visit to state  $e_i$  at time  $t$  (this event has probability  $1 - p_i$ ). If the next state to be visited is state  $e_j$  (that happens with probability  $P(t; i, j)$ ), a *transfer failure* may happen with probability  $\lambda_{i, j}$ , thus depending on both the original and the next states. Then state  $e_l$  is entered with constant probability  $\alpha^{i, j}(i, l)$  (with  $\sum_{l=1}^N \alpha^{i, j}(i, l) = 1$ ). The parameters  $\lambda_{i, j}, p_i$  are assumed to be strictly less than 1 for every  $i, j = 1, \dots, N$ .

A basic way for modeling failures which appear in clusters is to suppose that several failures may appear at each failure event. It is assumed that at most  $K < +\infty$  failures may be observed at a specific failure epoch. For the failures associated with the visits to state  $e_i$ , the numbers of simultaneous failures are i.i.d.  $\{1, \dots, K\}$ -valued random variables with probability distribution  $(b_i^{(k)}, k = 1, \dots, K)$ . For the transfer failures corresponding to transitions from state  $e_i$  to state  $e_j$ , the associated numbers of simultaneous failures are i.i.d.  $\{1, \dots, K\}$ -valued random variables with probability distribution  $(b_{i, j}^{(k)}, k = 1, \dots, K)$ . All these random variables are also assumed to be independent of the process of failures events. In other words, we have an independent marking of the failure events.

Let us define now the process  $X^* := (X_t^*)_{t \in \mathbb{N}}$  where  $X_t^*$  is the occupied state at the discrete epoch  $t$ . The random variable  $X_t^*$  is  $\mathcal{U}$ -valued. Let  $C_t$  be the cumulative number of failures observed up to time  $t$ . It follows from the independence assumptions on failure processes that the process  $(C, X^*) := (C_t, X_t^*)_{t \in \mathbb{N}}$  is a non-homogeneous Markov chain with state space  $\mathbb{N} \times \mathcal{U}$  and initial distribution: for  $k \geq 1$  and  $i = 1, \dots, N$

$$\mathbb{P}(C_0 = 0, X_0^* = e_i) = p_0(i), \quad \mathbb{P}(C_0 = k, X_0^* = e_i) = 0$$

where  $p_0$  is the probability distribution of  $X_0$ . Moreover, the transition probabilities of  $(C, X^*)$  satisfy: for all  $t \geq 0$ ,  $m \geq 0$ ,  $i, j = 1, \dots, N$  and  $k = 0, \dots, K$

$$\begin{aligned} & \mathbb{P}\left( (C_{t+1}, X_{t+1}^*) = (m + k, e_j) \mid (C_t, X_t^*) = (m, e_i) \right) \\ & = \mathbb{P}\left( (C_{t+1} - C_t, X_{t+1}^*) = (k, e_j) \mid X_t^* = e_i \right) = D_k(t; i, j) \end{aligned} \tag{1}$$

where the matrices  $D_k(t)$  are defined by: for  $k = 1, \dots, K$  and  $i, j = 1, \dots, N$

$$\begin{aligned} D_0(t; i, j) &:= (1 - p_i)P(t; i, j)(1 - \lambda_{i,j}) \\ D_k(t; i, j) &:= p_i b_i^{(k)} \alpha(i, j) + (1 - p_i) \left[ \sum_{l=1}^N P(t; i, l) \lambda_{i,l} b_{i,l}^{(k)} \alpha^{i,l}(i, j) \right]. \end{aligned} \quad (2)$$

The other transition probabilities for  $(C, X^*)$  are zero. Then, it follows from (1) that the process  $X^*$  is also a non-homogeneous Markov chain with transition probability matrices

$$P^*(t) = \sum_{k=0}^K D_k(t). \quad (3)$$

The nonnegative number  $D_0(t; i, j)$  represents the probability that  $X^*$  jumps from state  $e_i$  to  $e_j$  at time  $t + 1$  with no failure event. If  $k \geq 1$ , then the entry  $D_k(t; i, j)$  is the probability that  $X^*$  jumps from state  $e_i$  to state  $e_j$  at time  $t + 1$  with the occurrence of  $k$  failures.

**Example 1** *A basic model is when there is no feedback of the failure process on the execution process. Set  $\alpha(i, j) := P(t; i, j)$  and  $\alpha^{j,l}(i, j) := \delta_{jl}$  in the formula (2) for  $l, i, j = 1, \dots, N$ :*

$$\begin{aligned} D_0(t; i, j) &= (1 - p_i)P(t; i, j)(1 - \lambda_{i,j}) \\ D_k(t; i, j) &= [ p_i b_i^{(k)} + (1 - p_i) \lambda_{i,j} b_{i,j}^{(k)} ] P(t; i, j) \quad k = 1, \dots, K. \end{aligned}$$

*Note that, for every  $t \geq 0$ ,  $P^*(t) = P(t)$  as it can be expected. This model is a discrete-time counterpart of the Littlewood's software reliability model [Lit75]*

*A special instance of the previous model is obtained by setting  $\lambda_{i,j} = 0, i, j = 1, \dots, N$  in the above definition of matrices  $D_k(t), k = 0, \dots, K$ . In such a case, we have*

$$D_0(t) = (I - \text{diag}(p_i))P(t) \quad D_k(t) = \text{diag}(p_i b_i^{(k)}) P(t).$$

*where  $\text{diag}(c_i)$  is a diagonal matrix with the scalar  $c_i$  as  $(i, i)$ -entry and  $I$  is the  $N \times N$  identity matrix. When considering no multiple failures, the model may be interpreted as an elaboration of Cheung's model [Che80]. We emphasize that Cheung's model is integrated in the Cleanroom Reliability Manager [PMM93] whose aim is planning and certification of component-based software system reliability.*

**Remark 2** Let us consider a time-independent matrix  $P$  in the previous models. These models are instances of the discrete time counterpart of the Batch Markovian Arrival Process as defined by Neuts. Indeed, our matrices  $D_k$  ( $k = 0, \dots, K$ ) agree with those of Lucantoni's formalism for defining the finite BMAP [LR99]. The transition probability matrix of the underlying Markov chain is  $P^* = \sum_{k=0}^K D_k$ . More generally, if we consider the bivariate Markov chain  $(X_t^*, C_t)_{t \in \mathbb{N}}$  with transition probabilities satisfying (1), then such a process is a discrete-time instance of a Markov process homogeneous in the second component as defined in [ES69] (or a Markov additive process as studied by Çinlar [Çin72], Grigelionis [Gri79],...). Note that these processes are also special instances of switching processes as studied by Anisimov (see [Ani02] and the references therein).

## 2.2 The multivariate point process

A basic way to consider the failure process is as follows. Let  $N(k)_t$  be the number of  $k$ -failures events up to time  $t$  ( $k = 1, \dots, K$ ). Then, the failure process is specified by the  $K$ -variate process  $\mathbf{N} = (\mathbf{N}_t)_t$  where

$$\mathbf{N}_t := (N(1)_t, \dots, N(K)_t).$$

Note that the  $N(k)$ 's do not have common jumps. Let us introduce the  $K$ -variate process  $\Delta\mathbf{N}$  defined by

$$\Delta\mathbf{N}_0 := \mathbf{0}, \quad \Delta\mathbf{N}_t := \mathbf{N}_t - \mathbf{N}_{t-1}, \quad t \geq 1.$$

The random variable  $\Delta\mathbf{N}_t$  is  $\{\mathbf{0}, f_k, k = 1, \dots, K\}$ -valued, where  $f_k$  is  $k$ th vector of the canonical basis of  $\mathbb{R}^K$ . We encode  $\mathbf{0}$  as  $f_0$ , so that  $f_k$  will stand for a  $k$ -failures event ( $k = 0, \dots, K$ ). From now on, we suppose without loss of generality, that the state space of  $X^*$  is  $\mathcal{U} = \{e_i, i = 1, \dots, N\}$ , where  $e_i$  is the  $i$ th vector of the canonical basis of  $\mathbb{R}^N$ . Note that with these conventions,

$$1_{\{X_t^* = e_i\}} = \langle X_t^*, e_i \rangle, \quad 1_{\{\Delta\mathbf{N}_t = f_k\}} = \langle \Delta\mathbf{N}_t, f_k \rangle, \quad \text{and } X_t^* \mathbf{1}^\top = 1.$$

We suppose that all processes are defined on the same probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . We denote by  $\mathbb{F}^{X^*} = (\mathbb{F}_t^{X^*})_{t \in \mathbb{N}}$ ,  $\mathbb{F}^{\mathbf{N}} = (\mathbb{F}_t^{\mathbf{N}})_{t \in \mathbb{N}}$ ,  $\mathbb{F}^{\mathbf{N}, X^*} = (\mathbb{F}_t^{\mathbf{N}, X^*})_{t \in \mathbb{N}}$  the internal histories of processes  $X^*$ ,  $\mathbf{N}$  and  $(\mathbf{N}, X^*)$ , respectively. That is,  $\mathbb{F}_t^{X^*} := \sigma(X_s^*, 0 \leq s \leq t)$ ,  $\mathbb{F}_t^{\mathbf{N}} := \sigma(\mathbf{N}_s, 0 \leq s \leq t)$ ,  $\mathbb{F}_t^{\mathbf{N}, X^*} := \sigma((\mathbf{N}_s, X_s^*), 0 \leq s \leq t)$ . These histories are assumed to be complete, that is any  $\sigma$ -algebra contains all the sets of  $\mathcal{F}$  of  $\mathbb{P}$ -probability zero.

It follows from the independence assumptions on the failure and marking processes that: for  $k = 0, \dots, K$  and  $i, j = 1, \dots, N$

$$\mathbb{P}((\Delta\mathbf{N}_{t+1}, X_{t+1}^*) = (f_k, e_j) \mid \mathbb{F}_t^{\mathbf{N}, X^*}) = \mathbb{P}((\Delta\mathbf{N}_{t+1}, X_{t+1}^*) = (f_k, e_j) \mid X_t^*) \quad (4a)$$

and

$$\mathbb{P}((\Delta\mathbf{N}_{t+1}, X_{t+1}^*) = (f_k, e_j) \mid X_t^* = e_i) = D_k(t; i, j). \quad (4b)$$

Recall that  $X^*$  is a non-homogeneous Markov chain with transition probability matrices  $(P^*(t))_{t \in \mathbb{N}}$  given in (3).

The  $\mathbb{F}^{\mathbf{N}}$ -intensity of the univariate counting process  $N(k)$  is defined by: for  $t \geq 1$

$$\widehat{\lambda}(k)_t := \mathbb{P}(N(k)_t - N(k)_{t-1} = 1 \mid \mathbb{F}_{t-1}^{\mathbf{N}}) = \mathbb{P}(\Delta\mathbf{N}_t = f_k \mid \mathbb{F}_{t-1}^{\mathbf{N}}).$$

Note that  $\widehat{\lambda}(k)_t$  is  $\mathbb{F}_{t-1}^{\mathbf{N}}$ -measurable for each  $t \geq 1$ , or the process  $(\widehat{\lambda}(k))_{t \geq 1}$  is  $\mathbb{F}^{\mathbf{N}}$ -predictable. In our framework, the  $\mathbb{F}^{\mathbf{N}}$ -intensity of  $N(k)$  has the form: for  $t \geq 1$

$$\begin{aligned} \widehat{\lambda}(k)_t &= \mathbb{E}[\mathbb{P}(\Delta\mathbf{N}_t = f_k \mid \mathbb{F}_{t-1}^{\mathbf{N}, X^*}) \mid \mathbb{F}_{t-1}^{\mathbf{N}}] \\ &= \mathbb{E}[X_{t-1}^* D_k(t-1) \mathbf{1}^\top \mid \mathbb{F}_{t-1}^{\mathbf{N}}] \quad \text{from (4a, 4b)} \\ &= \widehat{X}_{t-1}^* D_k(t-1) \mathbf{1}^\top \end{aligned} \quad (5)$$

where  $\widehat{X}_t^* := \mathbb{E}[X_t^* | \mathbb{F}_t^{\mathbf{N}}]$ . The  $\mathbb{F}^{\mathbf{N}}$ -intensity  $\widehat{\boldsymbol{\lambda}}$  of the multivariate counting process  $\mathbf{N}$  is the  $\mathbb{F}^{\mathbf{N}}$ -predictable stochastic process defined by: for  $t \geq 1$ ,

$$\widehat{\boldsymbol{\lambda}}_t := (\widehat{\lambda}(1)_t, \dots, \widehat{\lambda}(K)_t),$$

so that  $\Delta \mathbf{N} - \widehat{\boldsymbol{\lambda}}$  is a  $\mathbb{F}^{\mathbf{N}}$ -martingale difference.

### 3 Poisson-type limit theorems

In this section, we claim that the  $K$ -variate counting variable  $\mathbf{N}_t$  converges in total variation to a vector of  $K$  independent Poisson distributed random variables when the failure parameters tend to zero. This convergence takes place at a specific time scale. Roughly speaking, we introduce a small parameter  $\varepsilon$  in the failure parameters and the convergence takes place at time scale  $t/\varepsilon$ . A rate of convergence is provided. It is based on an estimation of the total variation distance between the respective distributions of a  $K$ -variate counting variable and a vector of  $K$  independent distributed Poisson random variables reported in [Bro83, KLS83]. As a result, we obtain a Compound Poisson approximation of the cumulative number of failures.

#### 3.1 The model of perturbation

We multiply each failure parameter by  $\varepsilon$ :  $i, j = 1, \dots, N$

$$\varepsilon p_i, \quad \varepsilon \lambda_{i,j}.$$

Then, we consider the  $K$ -variate counting variable

$$\mathbf{N}_0^{(\varepsilon)} = \mathbf{0}, \quad \mathbf{N}_t^{(\varepsilon)} = (N^{(\varepsilon)}(1)_t, \dots, N^{(\varepsilon)}(K)_t) \quad t \geq 1, \quad (6)$$

associated with perturbed failure parameters (see (2)):  $k = 1, \dots, K$  and  $i, j = 1, \dots, N$

$$\begin{aligned} D_0^{(\varepsilon)}(t; i, j) &= (1 - p_i \varepsilon) P(t; i, j) (1 - \lambda_{i,j} \varepsilon) \\ D_k^{(\varepsilon)}(t; i, j) &= \varepsilon [p_i b_i^{(k)} \alpha(i, j) + (1 - p_i \varepsilon) \sum_l P(t; i, l) \lambda_{i,l} b_{i,l}^{(k)} \alpha^{i,l}(i, j)]. \end{aligned} \quad (7)$$

The transition probability matrices of the Markov process  $X^{*,\varepsilon}$  have the form for  $t \geq 0$

$$P^{*,\varepsilon}(t) = P(t) + \varepsilon Q(t) + \varepsilon^2 R(t) \quad (8a)$$

where

$$\begin{aligned} Q(t; i, j) &= -P(t; i, j)(p_i + \lambda_{i,j}) + p_i \alpha(i, j) + \sum_l P(t; i, l) \lambda_{i,l} \alpha^{i,j}(i, l), \\ R(t; i, j) &= p_i P(t; i, j) \lambda_{i,j} - p_i \left[ \sum_l P(t; i, l) \lambda_{i,l} \alpha^{i,j}(i, l) \right] \quad i, j = 1, \dots, N. \quad (8b) \\ Q(t) \mathbf{1}^\top &= R(t) \mathbf{1}^\top = \mathbf{0}. \end{aligned}$$

As it can be expected, the convergence of  $\mathbf{N}_t^{(\varepsilon)}$  to a non trivial limit fails at the current time scale. It is easily seen that the correct time scale is  $t/\varepsilon$ . Therefore, we investigate the asymptotic distribution of

$$\mathbf{N}_{t/\varepsilon}^{(\varepsilon)}$$

as  $\varepsilon$  tends to 0. For notational simplicity, we write  $t/\varepsilon$  for  $\lfloor t/\varepsilon \rfloor$ .

Note that  $D^{(\varepsilon)}(t)$ 's have the following form from (7): for  $k = 1, \dots, K$ ,  $i, j = 1, \dots, N$

$$D_k^{(\varepsilon)}(t) = \varepsilon B_k(t) + \varepsilon^2 L_k(t)$$

$$\text{with } \begin{cases} B_k(t; i, j) &= p_i b_i^{(k)} \alpha(i, j) + \sum_l P(t; i, l) b_{i,l}^{(k)} \lambda_{i,l} \alpha^{i,l}(i, j) \\ L_k(t; i, j) &= -p_i [\sum_l P(t; i, l) \lambda_{i,l} b_{i,l}^{(k)} \alpha^{i,l}(i, j)]. \end{cases} \quad (9)$$

## 3.2 The results

Let us recall that the total variation distance between two probability distributions  $\mathbb{P}_1$  and  $\mathbb{P}_2$  on  $\mathbb{N}^K$  is

$$d_{\text{TV}}(\mathbb{P}_1, \mathbb{P}_2) := \frac{1}{2} \sum_{\mathbf{n} \in \mathbb{N}^K} |\mathbb{P}_1(\mathbf{n}) - \mathbb{P}_2(\mathbf{n})|. \quad (10)$$

Let  $T$  be a positive integer. A vector of  $K$  independent Poisson distributed random variables with respective parameter  $\mu(k)_T$  ( $k = 1, \dots, K$ ) is denoted by  $\Pi(\boldsymbol{\mu}_T)$ , where  $\boldsymbol{\mu}_T$  stands for the vector  $(\mu(1)_T, \dots, \mu(K)_T)$ . Let us consider the vector  $\boldsymbol{\mu}_T$  defined by

$$\mu(k)_T := \pi B_k \mathbf{1}^\top T, \quad (11)$$

for  $k = 1, \dots, K$ , where  $B_k$ 's are defined by (14). For the random variables  $\mathbf{N}_{T/\varepsilon}^{(\varepsilon)}$  and  $\Pi(\boldsymbol{\mu}_T)$ ,  $d_{\text{TV}}(\mathbf{N}_{T/\varepsilon}^{(\varepsilon)}, \Pi(\boldsymbol{\mu}_T))$  will denote the total variation distance between their respective distributions. We have the following estimate from [Bro83, Th 1] or [KLS83, Th 3.1]

$$d_{\text{TV}}(\mathbf{N}_{T/\varepsilon}^{(\varepsilon)}, \Pi(\boldsymbol{\mu}_T)) \leq \sum_{k=1}^K \mathbb{E} \left| \sum_{t=1}^{T/\varepsilon} \widehat{\lambda}^{(\varepsilon)}(k)_t - \mu(k)_T \right| + \sum_{k=1}^K \mathbb{E} \sum_{t=1}^{T/\varepsilon} \widehat{\lambda}^{(\varepsilon)}(k)_t^2 \quad (12)$$

with (see (5) and (9))

$$\widehat{\lambda}^{(\varepsilon)}(k)_t = \widehat{X}^{*,\varepsilon}_{t-1} D_k^{(\varepsilon)}(t-1) \mathbf{1}^\top \quad \text{and} \quad \widehat{X}^{*,\varepsilon}_t := \mathbb{E}[X_t^{*,\varepsilon} \mid \mathbb{F}_t^{\mathbf{N}^{(\varepsilon)}}]. \quad (13)$$

### 3.2.1 The ergodic case

The main assumption is on the rate of convergence of  $P(t)$  in (8a) to some ergodic matrix  $P$ . We recall that  $P$  is said to be an (strongly) ergodic stochastic matrix if there exists an unique stochastic vector  $\pi$  such that, for any stochastic vector  $p_0$ ,

$$\lim_{t \rightarrow +\infty} \|p_0 P^t - \pi\|_1 = 0.$$

Note that some components of vector  $\pi$  may be zero, so that  $P$  has not to be irreducible and aperiodic.

(ASe) There exist an ergodic stochastic matrix  $P$  with invariant distribution  $\pi$ , and a real  $\alpha > 1$  such that

$$\lim_{t \rightarrow +\infty} t^\alpha \|P(t) - P\|_1 = 0.$$

Under (ASe), we deduce from (9) that there exist matrices  $(B_k)_{k=1}^K$  such that

$$\lim_{t \rightarrow +\infty} t^\alpha \|B_k(t) - B_k\|_1 = 0. \quad (14)$$

Moreover, matrix  $P^{*,\varepsilon}(t)$  in (8a) converges as  $t$  tends to  $\infty$ .

**Theorem 3** *Assume (ASe). Let  $T$  be a positive integer. Consider the multivariate counting variable  $\mathbf{N}_T^{(\varepsilon)}$  defined by (6). Let  $\Pi(\boldsymbol{\mu}_T)$  be a vector of  $K$  independent distributed Poisson random variables with vector of parameters  $\boldsymbol{\mu}_T = (\mu(1)_T, \dots, \mu(K)_T)$  defined in (11). Then, there is a positive constant  $\kappa_T$  such that*

$$d_{\text{TV}}(\mathbf{N}_{T/\varepsilon}^{(\varepsilon)}, \Pi(\boldsymbol{\mu}_T)) \leq \kappa_T \varepsilon. \quad (15)$$

The above theorem states that  $\mathbf{N}_{T/\varepsilon}^{(\varepsilon)}$  converges in variation to the random vector  $\Pi(\boldsymbol{\mu}_T)$  and this convergence takes place at the rate  $O(\varepsilon)$ . We recall that for discrete-valued random variables, the convergence in total variation is equivalent to the convergence in distribution [Bil68]. We mention that the convergence to a multivariate Poisson distribution and a convergence rate may be obtained with  $\alpha \in ]0, 1]$  in (ASe) (see Remark 17).

In the univariate case, we have the following Poisson approximation of the counting variable of failures.

**Corollary 4** *Assume (ASe) and  $K = 1$ .  $(N_t^{(\varepsilon)})_{t \in \mathbb{N}}$  is the counting process defined in (6). Let  $\Pi(\mu_T)$  a distributed Poisson random variable with parameter  $\mu_T = \pi B_1 \mathbf{1}^\top T$ , where matrix  $B_1$  is defined by (14). Then, there is a positive constant  $\kappa_T$  such that*

$$d_{\text{TV}}(N_{T/\varepsilon}^{(\varepsilon)}, \Pi(\mu_T)) \leq \kappa_T \varepsilon.$$

**Remark 5** The basic inequality (12) on  $d_{\text{TV}}(N_{T/\varepsilon}^{(\varepsilon)}, \Pi(\mu_T))$  may be replaced by another one provided by Serfling in [Ser75]. Although it has been shown that Serfling's bound cannot be compared to that of Brown, here, under (ASe) with  $\alpha > 1$ , Serfling's inequality also leads to the upper bound reported in Corollary 4. However, we do not know if the bound that we obtain for  $\alpha \in ]0, 1]$  (see Remark 17), can be derived from Serfling's result. This last comment holds for the irreducible case (ASi) below as well.

**Remark 6** Let  $P$  be an ergodic matrix and  $0 < r < 1$  denotes the geometric rate of the convergence of  $\|P^t - \mathbf{1}^\top \pi\|_1$  to 0. Assume that  $\lambda \in ]1, 1/\sqrt{r}[$ . Condition (ASe) may be weakened in

$$\lim_{t \rightarrow +\infty} g(2t) \|P(t) - P\|_1 = 0$$

where  $g(2t)$  is any strictly monotone function on  $\mathbb{N}$  such that  $g(2t) \leq \lambda^t$  for  $t$  large enough and  $\sum_t 1/g(2t) < +\infty$  (see Remark 23).

**Remark 7** The order of the bound in (15) can not be improved in general. Indeed, suppose that  $K = 1$ . We have from (10)

$$d_{\text{TV}}(N_{T/\varepsilon}^{(\varepsilon)}, \Pi(\mu_T)) \geq \frac{1}{2} |\mathbb{P}(N_{T/\varepsilon}^{(\varepsilon)} = 0) - \mathbb{P}(\Pi(\mu_T) = 0)|.$$

Let us consider the following simple time-homogeneous model

$$P := \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, D_0^{(\varepsilon)} := \begin{pmatrix} \frac{1}{2} - \frac{\varepsilon}{2} & \frac{1}{2} \\ \frac{1}{2} - \frac{\varepsilon}{2} & \frac{1}{2} \end{pmatrix}, D_1^{(\varepsilon)} := \begin{pmatrix} \frac{\varepsilon}{2} & 0 \\ \frac{\varepsilon}{2} & 0 \end{pmatrix}$$

with  $(1/2, 1/2)$  as the probability distribution of  $X_0$ .  $\mathbb{P}(N_{T/\varepsilon}^{(\varepsilon)} = 0)$  is

$$\mathbb{P}(N_{T/\varepsilon}^{(\varepsilon)} = 0) = (1/2 \ 1/2)(D_0^{(\varepsilon)})^{T/\varepsilon} \mathbf{1}^\top = \left(1 - \frac{1}{2} \varepsilon\right)^{T/\varepsilon}.$$

It follows from the last equality that

$$\begin{aligned} \mathbb{P}(\Pi(\mu_T) = 0) - \mathbb{P}(N_{T/\varepsilon}^{(\varepsilon)} = 0) &= \exp(-T/2) - (1 - \varepsilon/2)^{T/\varepsilon} \\ &= c_T \varepsilon + o(\varepsilon) \quad \text{with } c_T > 0. \end{aligned}$$

Then  $d_{\text{TV}}(N_{T/\varepsilon}^{(\varepsilon)}, \Pi(\mu_T)) \geq \kappa_T \varepsilon + o(\varepsilon)$  for  $\varepsilon$  small enough and some constant  $\kappa_T$ .

It is known that the estimates of the discrepancy between probability distributions provided by a martingale approach produce bounds which are  $T$ -dependent. Here,  $\kappa_T$  is quadratic in  $T$  in general. The dependence is linear when we have  $P^{*,\varepsilon}(t) = P(t), t \geq 0$ . This happens for the special models in Example 1.

**Remark 8** Let us consider the continuous-time multivariate counting process  $(\mathbf{N}_{\lfloor t/\varepsilon \rfloor}^{(\varepsilon)})_{t \in \mathbb{R}_+}$ . The compensator at time  $T$  of this counting process with respect to its internal filtration is given by  $\sum_{t=1}^{\lfloor T/\varepsilon \rfloor} \widehat{\boldsymbol{\lambda}}^{(\varepsilon)}_t$ , where  $\widehat{\boldsymbol{\lambda}}^{(\varepsilon)}_t := (\widehat{\lambda}_t^{(\varepsilon)}(k))_{k=1, \dots, K}$  is defined from (13).  $\boldsymbol{\mu}_T$  defined from (11) is the compensator at time  $T$  of a multivariate Poisson process with parameter vector  $(\pi B_k \mathbf{1}^\top)_{k=1, \dots, K}$ . In such a case, the results above are statements on the one-dimensional distribution of this continuous-time process. Let  $0 = t_0 < t_1 < \dots < t_k = T$ . Then, using [Bro83, Cor 1] or [KLS83, Th 3.2] we also have that the distance in variation between the finite dimensional distributions  $(\mathbf{N}_{\lfloor t_1/\varepsilon \rfloor}^{(\varepsilon)}, \dots, \mathbf{N}_{\lfloor T/\varepsilon \rfloor}^{(\varepsilon)})$  and  $(\Pi(\boldsymbol{\mu}_{t_1}), \dots, \Pi(\boldsymbol{\mu}_T))$  are bounded from above by the right hand side member of (12). Hence, the results above also hold when considering the problem of finite-dimensional convergence in variation of the continuous-time process  $(\mathbf{N}_{\lfloor t/\varepsilon \rfloor}^{(\varepsilon)})_{t \in \mathbb{R}_+}$ . In fact, the convergence also takes place at rate  $\varepsilon$  for the Skorokhod topology on the space of the càdlàg functions [Bro83, Cor 1], [KLS83, Th 4.1].

Before proving Theorem 3, we discuss a Compound approximation of the cumulative number of failures. This random variable is the following function of the vector  $\mathbf{N}_T^{(\varepsilon)}$

$$f(\mathbf{N}_T^{(\varepsilon)}) := \sum_{k=1}^K k N^{(\varepsilon)}(k)_T.$$

Consider  $f(\Pi(\boldsymbol{\mu}_T)) := \sum_{k=1}^K k \Pi(\mu(k)_T)$  with  $\mu(k)_T$ 's defined in (11). Then  $f(\Pi(\boldsymbol{\mu}_T))$  has a Compound Poisson distribution  $\text{CP}(\boldsymbol{\mu}_T \mathbf{1}^\top, \mathbf{p})$ , where the vector  $\mathbf{p}$  is

$$\mathbf{p} := \left( \frac{\mu(k)_T}{\sum_{k=1}^K \mu(k)_T} \right)_{k=1, \dots, K} = \left( \frac{\pi B_k \mathbf{1}^\top}{\sum_{k=1}^K \pi B_k \mathbf{1}^\top} \right)_{k=1, \dots, K}. \quad (16)$$

That is,

$$f(\Pi(\boldsymbol{\mu}_T)) \stackrel{d}{=} \sum_{j=1}^{P_T} \xi_j$$

where the random variable  $P_T := \Pi(\boldsymbol{\mu}_T) \mathbf{1}^\top$  has a Poisson distribution with parameter  $\boldsymbol{\mu}_T \mathbf{1}^\top$  and is independent of the  $\{1, \dots, K\}$ -valued random variables  $\xi_1, \xi_2, \dots$ . The  $\xi_j$ 's are i.i.d. with common distribution specified by vector  $\mathbf{p}$ . Since we have

$$d_{\text{TV}}(f(\mathbf{N}_{T/\varepsilon}^{(\varepsilon)}), f(\Pi(\boldsymbol{\mu}_T))) \leq d_{\text{TV}}(\mathbf{N}_{T/\varepsilon}^{(\varepsilon)}, \Pi(\boldsymbol{\mu}_T))$$

the following Compound Poisson approximation of  $f(\mathbf{N}_{T/\varepsilon}^{(\varepsilon)})$  results from Theorem 3.

**Corollary 9** *Assume (ASe). Let  $\text{CP}(\boldsymbol{\mu}_T \mathbf{1}^\top, \mathbf{p})$  be a Compound Poisson distributed random variable with  $\boldsymbol{\mu}_T, \mathbf{p}$  defined by (11) and (16) respectively.  $\mathbf{N}^{(\varepsilon)}$  is the  $K$ -variate counting process defined in (6). Let  $f(\mathbf{N}_T^{(\varepsilon)})$  be the cumulative counting process  $\sum_{k=1}^K k N^{(\varepsilon)}(k)$ . Then, there is a positive constant  $\kappa_T$  such that*

$$d_{\text{TV}}(f(\mathbf{N}_{T/\varepsilon}^{(\varepsilon)}), \text{CP}(\mathbf{1}^\top \boldsymbol{\mu}_T, \mathbf{p})) \leq \kappa_T \varepsilon.$$

**Remark 10** For the generic model considered in Section 2, the parameters specifying the Compound distribution in Corollary 9 have the following form from (9), (11) and (16),

$$\begin{aligned} \mu(k)_T &= \sum_i \pi(i) (p_i b_i^{(k)} + \sum_j P(i, j) \lambda_{i,j} b_{i,j}^{(k)}), \quad k = 1, \dots, K \\ \boldsymbol{\mu}_T \mathbf{1}^\top &= \sum_i \pi(i) (p_i + \sum_j P(i, j) \lambda_{i,j}) \\ \text{and } \mathbf{p} &= \left( \frac{\sum_i \pi(i) (p_i b_i^{(k)} + \sum_j P(i, j) \lambda_{i,j} b_{i,j}^{(k)})}{\sum_i \pi(i) (p_i + \sum_j P(i, j) \lambda_{i,j})} \right)_{k=1, \dots, K}. \end{aligned}$$

Note that  $\alpha(i, j), \alpha^{i,l}(i, j)$ 's representing the feedback of failures on the dynamics of  $X$  vanish in the definition of parameters of the asymptotic Compound distribution. This can be interpreted as follows. The failures tend to be rare and then, the effects of feedback on the dynamics of  $X$  may be neglected. Hence, from the asymptotic point of view, the models with or without feedback are equivalent.

### 3.2.2 The irreducible case

In this part, the case of an irreducible limit matrix  $P$  is discussed. Let us introduce our assumption on the convergence rate of the sequence  $(P(t))_{t \in \mathbb{N}}$  to  $P$ .

**(ASi)** There exist an irreducible stochastic matrix  $P$  with invariant distribution  $\pi$ , and a real  $\alpha > 1$  such that

$$\lim_{t \rightarrow +\infty} t^\alpha \|P(t) - P\|_1 = 0.$$

Then, our main result is as follows.

**Theorem 11** *Assume (ASi). The notation and definition of Theorem 3 are used. Then, there is, for any  $\delta \in ]0, 1[$ , a positive constant  $\kappa_{T,\delta}$  such that*

$$d_{\text{TV}}(\mathbf{N}_{T/\varepsilon}^{(\varepsilon)}, \Pi(\boldsymbol{\mu}_T)) \leq \kappa_{T,\delta} \varepsilon^\delta. \quad (17)$$

In particular, the result states that the convergence in distribution of the random variable  $\mathbf{N}_{T/\varepsilon}^{(\varepsilon)}$  to  $\Pi(\boldsymbol{\mu}_T)$ . Analogue of Corollaries 4,9 and Remarks 8,10 may be stated. We omit the details.

## 4 Proof of the Poisson approximation

### 4.1 Ergodic case

For any sequence of matrices  $(M(t))_{t \in \mathbb{N}}$  and  $t \geq n \geq 0$ , the forward-product matrix  $M^{(n,t)}$  is defined by

$$M^{(n,t)} := \begin{cases} M(n) \times \cdots \times M(t-1) & t > n \geq 0 \\ I & t = n. \end{cases} \quad (18)$$

The following lemma is essentially based on an estimate of the distance between the forward-product matrices generated by the sequences of perturbed and non-perturbed stochastic matrices  $(P^{*,\varepsilon}(t))_{t \in \mathbb{N}}$  and  $(P(t))_{t \in \mathbb{N}}$ . This lemma is related to estimates reported in [YZB03, Lem 2.1, Rem 2.2], even if the perturbation  $P^{*,\varepsilon}(t)$  of  $P^*(t)$  is not of the form assumed in this recent article. A related reference is also [Ani88]. A direct proof is reported below for completeness.

**Lemma 12** *Let  $(P^{(\varepsilon)}(t))_{t \in \mathbb{N}}$  be the sequence of stochastic matrices defined in (8a-8b).  $P^{\varepsilon,(n,t)}$  (resp.  $P^{(n,t)}$ ) is the forward-product matrix defined by (18) from  $(P^{(\varepsilon)}(t))_{t \in \mathbb{N}}$  (resp.  $(P(t))_{t \in \mathbb{N}}$ ). Let  $\pi$  be any stochastic vector.*

*If  $0 < \varepsilon \leq 1$  then there exists a positive constant  $\kappa$  such that: for  $t \geq n \geq 0$ ,*

$$\|P^{\varepsilon,(n,t)} - \mathbf{1}^\top \pi\|_1 \leq \|P^{(n,t)} - \mathbf{1}^\top \pi\|_1 + \kappa \varepsilon \sum_{k=n+1}^t \|P^{(k,t)} - \mathbf{1}^\top \pi\|_1 \quad (19)$$

*where  $\kappa$  does not depend on  $\varepsilon$  and  $t, n$ . We adopt the convention that  $\sum_{k=t+1}^t$  is null.*

**Proof .** Let us show that inequality (19) is valid. Set, for any  $t \geq n \geq 0$ ,

$$K(n, t) := P^{\varepsilon, (n, t)} - P^{(n, t)}. \quad (20)$$

Note that  $K(n, n) = \mathbf{0}$  with our conventions. We have for  $t > n \geq 0$

$$\begin{aligned} K(n, t) &= (P^{\varepsilon, (n, t-1)} - P^{(n, t-1)})P(t-1) + P^{\varepsilon, (n, t-1)}(P^{*, \varepsilon}(t-1) - P(t-1)) \\ &= K(n, t-1)P(t-1) + P^{\varepsilon, (n, t-1)}(P^{*, \varepsilon}(t-1) - P(t-1)). \end{aligned}$$

Then, solving this difference equation, using the special form of  $(P^{*, \varepsilon}(t))_{t \in \mathbb{N}}$  and  $(Q(l) + \varepsilon R(l))\mathbf{1}^\top = \mathbf{0}$ , we can write

$$\begin{cases} K(n, t) &= \varepsilon \sum_{l=n+1}^t P^{\varepsilon, (n, l-1)}(Q(l-1) + \varepsilon R(l-1))(P^{(l, t)} - \mathbf{1}^\top \pi) & t > n \geq 0 \\ K(n, n) &= \mathbf{0}. \end{cases} \quad (21)$$

Since  $\|P^{*, \varepsilon}(t)\|_1 = 1$  and  $\|Q(l) + \varepsilon R(l)\|_1$  is uniformly bounded for  $\varepsilon \leq 1$ , we find that, for  $t > n$ ,

$$\begin{aligned} \|K(n, t)\|_1 &\leq \varepsilon \sum_{l=n+1}^t \|Q(l-1) + \varepsilon R(l-1)\|_1 \|P^{(l, t)} - \mathbf{1}^\top \pi\|_1 \\ &\leq \varepsilon \kappa \sum_{l=n+1}^t \|P^{(l, t)} - \mathbf{1}^\top \pi\|_1. \end{aligned} \quad (22)$$

The final step consist in using the triangle inequality

$$\|P^{\varepsilon, (n, t)} - \mathbf{1}^\top \pi\|_1 \leq \|P^{\varepsilon, (n, t)} - P^{(n, t)}\|_1 + \|P^{(n, t)} - \mathbf{1}^\top \pi\|_1.$$

Then, inequality (19) follows from inequality (22).  $\square$

For any vector  $\mathbf{a}$  and scalar  $b \in \mathbb{R}$ , the vector  $(\mathbf{a}/b)^+$  is defined by  $\mathbf{0}$  if  $b = 0$  and  $\mathbf{a}/b$  otherwise. The process  $(\widehat{X}^{*, \varepsilon}_t := \mathbb{E}[X_t^{*, \varepsilon} \mid \mathbb{F}_t^{\mathbf{N}^{(\varepsilon)}}])_{t \in \mathbb{N}}$  may be shown to satisfy the following recursive formula from [Led05, Cor 3.1, Rem 3.2]:

$$\widehat{X}^{*, \varepsilon}_0 = p_0 \quad \text{and} \quad \widehat{X}^{*, \varepsilon}_t = \widehat{X}^{*, \varepsilon}_{t-1} P^{*, \varepsilon}(t) + (\Delta \mathbf{N}_t^{(\varepsilon)} - \widehat{\boldsymbol{\lambda}}^{(\varepsilon)}_t) K_{t-1} \quad t \geq 1, \quad (23)$$

where

$$\begin{aligned} \Delta \mathbf{N}_0^{(\varepsilon)} &:= \mathbf{0}, \quad \Delta \mathbf{N}_t^{(\varepsilon)} := \mathbf{N}_t^{(\varepsilon)} - \mathbf{N}_{t-1}^{(\varepsilon)}, \quad \widehat{\boldsymbol{\lambda}}^{(\varepsilon)}_t := (\widehat{\lambda}^{(\varepsilon)}(1)_t, \dots, \widehat{\lambda}^{(\varepsilon)}(K)_t) \\ K_{t-1} &:= \begin{pmatrix} \left( \frac{\widehat{X}^{*, \varepsilon}_{t-1} D_1^{(\varepsilon)}(t-1)}{\widehat{\lambda}^{(\varepsilon)}(1)_t} \right)^+ - \frac{\widehat{X}^{*, \varepsilon}_{t-1} D_0^{(\varepsilon)}(t-1)}{1 - \sum_{k=1}^K \widehat{\lambda}^{(\varepsilon)}(k)_t} \\ \vdots \\ \left( \frac{\widehat{X}^{*, \varepsilon}_{t-1} D_K^{(\varepsilon)}(t-1)}{\widehat{\lambda}^{(\varepsilon)}(K)_t} \right)^+ - \frac{\widehat{X}^{*, \varepsilon}_{t-1} D_0^{(\varepsilon)}(t-1)}{1 - \sum_{k=1}^K \widehat{\lambda}^{(\varepsilon)}(k)_t} \end{pmatrix} \end{aligned}$$

and the stochastic vector  $p_0$  corresponds to the probability distribution of  $X_0^*$ .

**Remark 13** Since  $\mathbb{E}[X_{t+1}^{*,\varepsilon} | X_t^{*,\varepsilon}] = X_t^{*,\varepsilon} P^{*,\varepsilon}(t)$ , the Markov chain  $X^{*,\varepsilon}$  has the following representation: for  $t \geq 0$

$$X_{t+1}^{*,\varepsilon} = X_t^{*,\varepsilon} P^{*,\varepsilon}(t) + V_{t+1} \quad (24)$$

where  $(V_t)_{t \geq 1}$  is a  $\mathbb{F}^{X^{*,\varepsilon}}$ -martingale difference. In fact, Formula (23) is the Doob-Meyer decomposition of the process  $\widehat{X}^*$  and can be thought of as the “projection of the decomposition (24) for  $X^{*,\varepsilon}$  onto the internal filtration  $\mathbb{F}^{\mathbf{N}^{(\varepsilon)}}$ ”. A related reference is [Seg76].

The next lemma gives an estimate of the deviation of the filtered state  $\widehat{X}^{*,\varepsilon}_t$  from any stochastic vector  $\pi$ .

**Lemma 14** *Let  $\pi$  be any stochastic vector and  $0 < \varepsilon \leq 1$ . Then, for  $t \geq 0$ ,*

$$\mathbb{E} \|\widehat{X}^{*,\varepsilon}_t - \pi\|_1 \leq \|P^{(0,t)} - \mathbf{1}^\top \pi\|_1 + \kappa \varepsilon \sum_{l=1}^t \|P^{(l,t)} - \mathbf{1}^\top \pi\|_1 + \kappa \varepsilon^2 \sum_{l=1}^t \sum_{k=l+1}^t \|P^{(k,t)} - \mathbf{1}^\top \pi\|_1 \quad (25)$$

for some constant  $\kappa$  that does not depend on  $\varepsilon, t$ . We adopt the convention that  $\sum_{k=t+1}^t$  is null.

**Proof .** When  $t = 0$ , inequality (25) follows from (19) because  $\|\widehat{X}^{*,\varepsilon}_0 - \pi\|_1 = \|\widehat{X}^{*,\varepsilon}_0(P^{(0,0)} - \mathbf{1}^\top \pi)\|_1 \leq \|\widehat{X}^{*,\varepsilon}_0\|_1 \|P^{(0,0)} - \mathbf{1}^\top \pi\|_1 = \|P^{(0,0)} - \mathbf{1}^\top \pi\|_1$  since  $\|\widehat{X}^{*,\varepsilon}_0\|_1 = 1$ .

Relation (23) may be rewritten as follows

$$\widehat{X}^{*,\varepsilon}_t = \widehat{X}^{*,\varepsilon}_{t-1} P^{*,\varepsilon}(t) + \sum_{k=1}^K \langle \Delta \mathbf{N}_t^{(\varepsilon)} - \widehat{\boldsymbol{\lambda}}^{(\varepsilon)}_t, f_k \rangle H(k)_{t-1}, \quad t \geq 1,$$

with

$$H(k)_{t-1} := \widehat{X}^{*,\varepsilon}_{t-1} \left( \left( \frac{D_k^{(\varepsilon)}(t-1)}{\widehat{\lambda}^{(\varepsilon)}(k)_t} \right)^+ - \frac{D_0^{(\varepsilon)}(t-1)}{1 - \sum_k \widehat{\lambda}^{(\varepsilon)}(k)_t} \right).$$

For any  $k = 1, \dots, K$ ,  $H(k)_{t-1}$  is  $\mathbb{F}^{\mathbf{N}^{(\varepsilon)}}_{t-1}$ -measurable for all  $t \geq 1$ , i.e. the matrix-valued random process  $(H(k)_t)_{t \in \mathbb{N}}$  is  $\mathbb{F}^{\mathbf{N}^{(\varepsilon)}}$ -predictable. Iterating the previous relation, we obtain

$$\widehat{X}^{*,\varepsilon}_t = \widehat{X}^{*,\varepsilon}_0 P^{\varepsilon,(0,t)} + \sum_{k=1}^K \sum_{l=1}^t \langle \Delta \mathbf{N}_l^{(\varepsilon)} - \widehat{\boldsymbol{\lambda}}^{(\varepsilon)}_l, f_k \rangle H_{l-1}(k) P^{\varepsilon,(l,t)} \quad t \geq 1$$

where  $P^{\varepsilon,(n,t)}$  is the forward-product matrix associated with the sequence  $(P^{*,\varepsilon}(t))_{t \in \mathbb{N}}$ . Because  $H_{t-1}(k) \mathbf{1}^\top = \mathbf{0}$  for  $t \geq 1$  and  $k = 1, \dots, K$ , and  $\widehat{X}^{*,\varepsilon}_0 \mathbf{1}^\top = 1$ , we find for  $t \geq 1$

$$\widehat{X}^{*,\varepsilon}_t - \pi = \widehat{X}^{*,\varepsilon}_0 (P^{\varepsilon,(0,t)} - \mathbf{1}^\top \pi) + \sum_{k=1}^K \sum_{l=1}^t \langle \Delta \mathbf{N}_l^{(\varepsilon)} - \widehat{\boldsymbol{\lambda}}^{(\varepsilon)}_l, f_k \rangle H_{l-1}(k) (P^{\varepsilon,(l,t)} - \mathbf{1}^\top \pi) \quad (26)$$

and

$$\|\widehat{X}^{*,\varepsilon}_t - \pi\|_1 \leq \|P^{\varepsilon,(0,t)} - \mathbf{1}^\top \pi\|_1 + \sum_{k=1}^K \sum_{l=1}^t \langle \Delta \mathbf{N}_l^{(\varepsilon)} + \widehat{\boldsymbol{\lambda}}^{(\varepsilon)}_l, f_k \rangle \|H_{l-1}(k) (P^{\varepsilon,(l,t)} - \mathbf{1}^\top \pi)\|_1.$$

using basic facts for  $\|\cdot\|_1$  and  $\|\widehat{X}^{*,\varepsilon_0}\|_1 = 1$ .

Since  $(H_t(k))$  is a  $\mathbb{F}^{\mathcal{N}^{(\varepsilon)}}$ -predictable process and  $\widehat{\lambda}^{(\varepsilon)}(k) = \langle \widehat{\lambda}^{(\varepsilon)}, f_k \rangle$  is the  $\mathbb{F}^{\mathcal{N}^{(\varepsilon)}}$ -intensity of the counting process  $N^{(\varepsilon)}(k)$ , it follows taking the expectation on both side of the last inequality that, for  $t \geq 1$ ,

$$\begin{aligned} \mathbb{E}\|\widehat{X}^{*,\varepsilon}_t - \pi\|_1 &\leq \|P^{\varepsilon,(0,t)} - \mathbf{1}^\top \pi\|_1 + 2 \sum_{k=1}^K \sum_{l=1}^t \|H_{l-1}(P^{\varepsilon,(l,t)} - \mathbf{1}^\top \pi)\|_1 \widehat{\lambda}^{(\varepsilon)}(k)_l \\ &\leq \|P^{\varepsilon,(0,t)} - \mathbf{1}^\top \pi\|_1 + 4 \sum_{k=1}^K \sum_{l=1}^t \|P^{\varepsilon,(l,t)} - \mathbf{1}^\top \pi\|_1 \widehat{\lambda}^{(\varepsilon)}(k)_l. \end{aligned}$$

The last inequality holds since  $\|H_{l-1}(k)\|_1 \leq 2$  for every  $t \geq 1$  as the  $l_1$ -norm of the difference of two stochastic vectors.

Hereafter,  $\kappa$  stands for a generic positive constant. As it can be seen from the proof of Theorem 3,  $\widehat{\lambda}^{(\varepsilon)}(k)_t \leq \kappa \varepsilon$  for every  $t \geq 1$  (with  $\varepsilon \leq 1$ ). Then, we obtain for  $t \geq 1$

$$\mathbb{E}\|\widehat{X}^{*,\varepsilon}_t - \pi\|_1 \leq \|P^{\varepsilon,(0,t)} - \mathbf{1}^\top \pi\|_1 + \kappa \varepsilon \sum_{l=1}^t \|P^{\varepsilon,(l,t)} - \mathbf{1}^\top \pi\|_1.$$

Finally, we find from Lemma 12-(19) that, for  $t \geq 1$

$$\begin{aligned} \mathbb{E}\|\widehat{X}^{*,\varepsilon}_t - \pi\|_1 &\leq \|P^{(0,t)} - \mathbf{1}^\top \pi\|_1 + \kappa \varepsilon \sum_{l=1}^t \|P^{(l,t)} - \mathbf{1}^\top \pi\|_1 \\ &\quad + \kappa \varepsilon \sum_{l=1}^t \left( \|P^{(l,t)} - \mathbf{1}^\top \pi\|_1 + \kappa \varepsilon \sum_{k=l+1}^t \|P^{(k,t)} - \mathbf{1}^\top \pi\|_1 \right). \end{aligned}$$

for some constant  $\kappa$  which does not depend on  $t, \varepsilon$ . Inequality (25) follows easily.  $\square$

**Proof of Theorem 3.** First, we consider the second term in the right hand side member of (12). We obtain from (5) that

$$\begin{aligned} \mathbb{E} \sum_{t=1}^{T/\varepsilon} (\widehat{\lambda}^{(\varepsilon)}(k)_t)^2 &= \mathbb{E} \sum_{t=1}^{T/\varepsilon} \left| \langle \widehat{X}^{*,\varepsilon}_{t-1}, D_k^{(\varepsilon)}(t-1) \mathbf{1}^\top \rangle \right|^2 \\ &\leq \mathbb{E} \sum_{t=1}^{T/\varepsilon} \|\widehat{X}^{*,\varepsilon}_{t-1}\|_1^2 \|D_k^{(\varepsilon)}(t-1) \mathbf{1}^\top\|_1^2 = \sum_{t=1}^{T/\varepsilon} \|D_k^{(\varepsilon)}(t-1) \mathbf{1}^\top\|_1^2 \end{aligned}$$

since  $\|\widehat{X}^{*,\varepsilon}_{t-1}\|_1 = 1$ . Now, we know that  $\|D_k^{(\varepsilon)}(t-1) \mathbf{1}^\top\|_1^2 = c\varepsilon^2 + o(\varepsilon^2)$ . Then, with  $\varepsilon \leq 1$ ,

$$\mathbb{E} \sum_{t=1}^{T/\varepsilon} (\widehat{\lambda}^{(\varepsilon)}(k)_t)^2 \leq \kappa_{T,1} \varepsilon. \quad (27)$$

Second, consider the first term in the right hand side member of (12). We can write from (5),(7),(11)

$$\begin{aligned}
\sum_{t=1}^{T/\varepsilon} \widehat{\lambda}^{(\varepsilon)}(k)_t - \mu(k)_T &= \sum_{t=1}^{T/\varepsilon} \widehat{X}^{*,\varepsilon}_{t-1} D_k^{(\varepsilon)}(t-1) \mathbf{1}^\top - T\pi B_k \mathbf{1}^\top \\
&= \sum_{t=1}^{T/\varepsilon} \varepsilon \widehat{X}^{*,\varepsilon}_{t-1} B_k(t-1) \mathbf{1}^\top - T\pi B_k \mathbf{1}^\top + \sum_{t=1}^{T/\varepsilon} \varepsilon^2 \widehat{X}^{*,\varepsilon}_{t-1} L_k(t-1) \mathbf{1}^\top \\
&= \varepsilon \sum_{t=1}^{T/\varepsilon} (\widehat{X}^{*,\varepsilon}_{t-1} - \pi) B_k \mathbf{1}^\top + \varepsilon \sum_{t=1}^{T/\varepsilon} \widehat{X}^{*,\varepsilon}_{t-1} (B_k(t-1) - B_k) \mathbf{1}^\top \\
&\quad + \varepsilon^2 \sum_{t=1}^{T/\varepsilon} \widehat{X}^{*,\varepsilon}_{t-1} L_k(t-1) \mathbf{1}^\top.
\end{aligned}$$

We clearly have from the last term in the equality above,

$$\varepsilon^2 \left| \sum_{t=1}^{T/\varepsilon} \widehat{X}^{*,\varepsilon}_{t-1} L_k(t-1) \mathbf{1}^\top \right| \leq \varepsilon^2 \sum_{t=1}^{T/\varepsilon} |\langle \widehat{X}^{*,\varepsilon}_{t-1}, L_k(t-1) \mathbf{1}^\top \rangle| \leq \varepsilon^2 \sum_{t=1}^{T/\varepsilon} \|L_k(t-1) \mathbf{1}^\top\|_1.$$

Since  $\|L_k(t-1) \mathbf{1}^\top\|_1$  is uniformly bounded in  $t$  (see (9)), it follows that

$$\left| \sum_{t=1}^{T/\varepsilon} \widehat{\lambda}^{(\varepsilon)}(k)_t - \mu_T(k) \right| \leq \varepsilon \left| \sum_{t=1}^{T/\varepsilon} (\widehat{X}^{*,\varepsilon}_{t-1} - \pi) B_k \mathbf{1}^\top \right| + \varepsilon \left| \sum_{t=1}^{T/\varepsilon} \widehat{X}^{*,\varepsilon}_{t-1} (B_k(t-1) - B_k) \mathbf{1}^\top \right| + \kappa_{T,2} \varepsilon.$$

Next, in a similar way than for the previous estimate, we get

$$\begin{aligned}
\varepsilon \left| \sum_{t=1}^{T/\varepsilon} \widehat{X}^{*,\varepsilon}_{t-1} (B_k(t-1) - B_k) \mathbf{1}^\top \right| &\leq \varepsilon \sum_{t=1}^{T/\varepsilon} \|(B_k(t-1) - B_k) \mathbf{1}^\top\|_1 \\
&\leq \varepsilon N \sum_{t=1}^{+\infty} \|B_k(t-1) - B_k\|_1.
\end{aligned}$$

Under (ASe), this last term is bounded from above by  $\kappa_3 \varepsilon$  for some  $\kappa_3$ . Therefore, we find that

$$\mathbb{E} \left| \sum_{t=1}^{T/\varepsilon} \widehat{\lambda}^{(\varepsilon)}(k)_t - \mu_T(k) \right| \leq \varepsilon \|B_k \mathbf{1}^\top\|_1 \sum_{t=1}^{T/\varepsilon} \mathbb{E} \|\widehat{X}^{*,\varepsilon}_{t-1} - \pi\|_1 + \kappa_3 \varepsilon + \kappa_{T,2} \varepsilon.$$

The proof will be complete from (27) and the above inequality, if we show that

$$\sum_{t=1}^{T/\varepsilon} \mathbb{E} \|\widehat{X}^{*,\varepsilon}_{t-1} - \pi\|_1 \leq \kappa_{T,4} \tag{28}$$

where  $\kappa_{T,4}$  does not depend upon  $\varepsilon$ . Using Lemma 14, we have from (25)

$$\begin{aligned} \sum_{t=1}^{T/\varepsilon} \mathbb{E} \|\widehat{X}^{*,\varepsilon}_{t-1} - \pi\|_1 &= \sum_{t=0}^{T/\varepsilon-1} \mathbb{E} \|\widehat{X}^{*,\varepsilon}_t - \pi\|_1 \\ &\leq \sum_{t=0}^{T/\varepsilon-1} \|P^{(0,t)} - \mathbf{1}^\top \pi\|_1 + \kappa \varepsilon \sum_{t=0}^{T/\varepsilon-1} \sum_{l=1}^t \|P^{(l,t)} - \mathbf{1}^\top \pi\|_1 \\ &\quad + \kappa \varepsilon^2 \sum_{t=0}^{T/\varepsilon-1} \sum_{l=1}^t \sum_{k=l+1}^t \|P^{(k,t)} - \mathbf{1}^\top \pi\|_1. \end{aligned}$$

Under (ASe), we have that  $P(t)$  converges to the ergodic matrix  $P$  with invariant distribution  $\pi$ . The following inequalities are shown to be valid in Appendix A (see (A.3-A.5))

$$\begin{aligned} \sum_{t=0}^{T/\varepsilon-1} \|P^{(0,t)} - \mathbf{1}^\top \pi\|_1 &\leq C & \sum_{t=0}^{T/\varepsilon-1} \sum_{l=1}^t \|P^{(l,t)} - \mathbf{1}^\top \pi\|_1 &\leq \frac{T}{\varepsilon} C \\ \sum_{t=0}^{T/\varepsilon-1} \sum_{l=1}^t \sum_{k=l+1}^t \|P^{(k,t)} - \mathbf{1}^\top \pi\|_1 &\leq \left(\frac{T}{\varepsilon}\right)^2 C. \end{aligned}$$

The proof is easily completed.  $\square$

## 4.2 The irreducible case

The structure of the proof of Theorem 11 is similar to that of Theorem 3. Only the relevant steps are given below.

**Lemma 15** *Suppose that the conditions of Lemma 12 are satisfied. Then,*

$$\left\| \sum_{k=n}^t (P^{\varepsilon,(n,k)} - \mathbf{1}^\top \pi) \right\|_1 \leq \left\| \sum_{k=n}^t (P^{(n,k)} - \mathbf{1}^\top \pi) \right\|_1 + \kappa \varepsilon \sum_{l=n+1}^t \left\| \sum_{k=l}^t (P^{(l,k)} - \mathbf{1}^\top \pi) \right\|_1$$

where  $\kappa$  does not depend on  $\varepsilon$  and  $t, n$ . The convention for  $\sum_{k=t+1}^t$  is as in Lemma 12.

**Proof .** Set, for any  $t \geq n \geq 0$ ,

$$C(n, t) := \sum_{k=n}^t (P^{\varepsilon,(n,k)} - P^{(n,k)}) = \sum_{k=n}^t K(k, n)$$

where  $K(k, n)$  is defined by (20). It follows from (21) that, for  $t \geq n$

$$\begin{aligned} C(n, t) &= \varepsilon \sum_{k=n}^t \sum_{l=n+1}^k P^{\varepsilon,(n,l-1)} (Q(l-1) + \varepsilon R(l-1)) (P^{(l,k)} - \mathbf{1}^\top \pi) \\ &= \varepsilon \sum_{l=n+1}^t P^{\varepsilon,(n,l-1)} (Q(l-1) + \varepsilon R(l-1)) \sum_{k=l}^t (P^{(l,k)} - \mathbf{1}^\top \pi) \end{aligned}$$

Then, we obtain as for (22)

$$\|C(n, t)\|_1 \leq \varepsilon \kappa \sum_{l=n+1}^t \left\| \sum_{k=l}^t (P^{(l,k)} - \mathbf{1}^\top \pi) \right\|_1. \quad (29)$$

The final step consist in using the triangle inequality as in Lemma 12.  $\square$

The recursive equation (23) for the filter is valid under no special assumption on the underlying Markov chain [Led05]. The analogue of Lemma 14 is given below.

**Lemma 16** *Suppose that the conditions of Lemma 14 are satisfied. Then, for  $t \geq 0$*

$$\begin{aligned} \mathbb{E} \left\| \sum_{k=0}^t (\widehat{X}^{*,\varepsilon}_k - \pi) \right\|_1 &\leq \left\| \sum_{k=0}^t (P^{(0,k)} - \mathbf{1}^\top \pi) \right\|_1 + \kappa \varepsilon \sum_{l=1}^t \left\| \sum_{k=l}^t (P^{(l,k)} - \mathbf{1}^\top \pi) \right\|_1 \\ &\quad + \kappa \varepsilon^2 \sum_{l=1}^t \sum_{m=l+1}^t \left\| \sum_{n=m}^t (P^{(m,n)} - \mathbf{1}^\top \pi) \right\|_1 \end{aligned} \quad (30)$$

for some constant  $\kappa$  which does not depend on  $\varepsilon, t$  (the convention are as in Lemma 14).

**Proof .** When  $t = 0$  inequality (30) is nothing else than (25) for  $t = 0$ .

We know from (26) that, for  $t \geq 1$

$$\begin{aligned} \sum_{m=0}^t (\widehat{X}^{*,\varepsilon}_m - \pi) &= \sum_{m=0}^t \widehat{X}^{*,\varepsilon}_0 (P^{\varepsilon,(0,m)} - \mathbf{1}^\top \pi) \\ &\quad + \sum_{k=1}^K \sum_{m=0}^t \sum_{l=1}^m \langle \Delta \mathbf{N}_l^{(\varepsilon)} - \widehat{\lambda}^{(\varepsilon)}_{l, f_k} \rangle H_{l-1}(k) (P^{\varepsilon,(l,m)} - \mathbf{1}^\top \pi) \\ &= \widehat{X}^{*,\varepsilon}_0 \sum_{k=0}^t (P^{\varepsilon,(0,k)} - \mathbf{1}^\top \pi) \\ &\quad + \sum_{k=1}^K \sum_{l=1}^t \langle \Delta \mathbf{N}_l^{(\varepsilon)} - \widehat{\lambda}^{(\varepsilon)}_{l, f_k} \rangle H_{l-1}(k) \sum_{m=l}^t (P^{\varepsilon,(l,m)} - \mathbf{1}^\top \pi) \end{aligned}$$

Then, we obtain using basic facts for  $\|\cdot\|_1$  and  $\|\widehat{X}^{*,\varepsilon}_0\|_1 = 1$

$$\begin{aligned} \left\| \sum_{m=0}^t (\widehat{X}^{*,\varepsilon}_m - \pi) \right\|_1 &\leq \left\| \sum_{k=0}^t (P^{\varepsilon,(0,k)} - \mathbf{1}^\top \pi) \right\|_1 \\ &\quad + \sum_{k=1}^K \sum_{l=1}^t \langle \Delta \mathbf{N}_l^{(\varepsilon)} + \widehat{\lambda}^{(\varepsilon)}_{l, f_k} \rangle \|H_{l-1}(k)\|_1 \left\| \sum_{m=l}^t (P^{\varepsilon,(l,m)} - \mathbf{1}^\top \pi) \right\|_1. \end{aligned}$$

The proof is easily completed along with the line of the proof of Lemma 14 ( using Lemma 15 in place of of Lemma 12).  $\square$

**Proof of Theorem 11.** A close inspection of the proof of Theorem 3, shows that, under (ASi), the only changes concern the term

$$\begin{aligned} \varepsilon \mathbb{E} \left| \sum_{t=1}^{T/\varepsilon} (\widehat{X}^{*,\varepsilon}_{t-1} - \pi) B_k \mathbf{1}^\top \right| &= \varepsilon \mathbb{E} \left| \left\langle \sum_{t=0}^{T/\varepsilon-1} (\widehat{X}^{*,\varepsilon}_t - \pi), B_k \mathbf{1}^\top \right\rangle \right| \\ &\leq \varepsilon \|B_k \mathbf{1}^\top\|_1 \mathbb{E} \left\| \sum_{t=0}^{T/\varepsilon-1} (\widehat{X}^{*,\varepsilon}_t - \pi) \right\|_1 \end{aligned}$$

It follows from Lemma 16 that

$$\begin{aligned} \mathbb{E} \left\| \sum_{t=0}^{T/\varepsilon-1} (\widehat{X}^{*,\varepsilon}_t - \pi) \right\|_1 &\leq \left\| \sum_{t=0}^{T/\varepsilon-1} (P^{(0,t)} - \mathbf{1}^\top \pi) \right\|_1 + \kappa \varepsilon \sum_{l=1}^{T/\varepsilon-1} \left\| \sum_{k=l}^{T/\varepsilon-1} (P^{(l,k)} - \mathbf{1}^\top \pi) \right\|_1 \\ &\quad + \kappa \varepsilon^2 \sum_{l=1}^{T/\varepsilon-1} \sum_{m=l+1}^{T/\varepsilon-1} \left\| \sum_{n=m}^{T/\varepsilon-1} (P^{(m,n)} - \mathbf{1}^\top \pi) \right\|_1 \end{aligned}$$

for some constant  $\kappa$  which does not depend on  $\varepsilon, t$ . We easily deduce from (A.7)–(A.9) that there exists, for any  $\delta \in ]0, 1[$ , a  $\kappa_{T,\delta}$  such that

$$\mathbb{E} \left\| \sum_{t=0}^{T/\varepsilon-1} (\widehat{X}^{*,\varepsilon}_t - \pi) \right\|_1 \leq \kappa_{T,\delta} \varepsilon^{-\delta}. \quad (31)$$

Multiplying the inequality above by  $\varepsilon$  gives the final form of (17).  $\square$

**Remark 17** We emphasize that Lemmas 15,16 are valid under no specific assumption on  $P$ . Then, it is not difficult to see that the convergence in Theorem 3, with  $\alpha \in ]0, 1[$  in (ASe), may be derived from a similar proof than that of Theorem 11. We also obtain a convergence rate. For  $\alpha \in ]0, 1[$ , the order is in  $\varepsilon^{\alpha-\delta}$  for any  $\delta$  such that  $0 < \delta < \alpha$  and  $\varepsilon$  small enough. The details are omitted.

**Remark 18** It follows from the proofs of Theorems 3,11 (see (28) and (31)) that the following functional of the filtered state converges in  $L_1$  norm to  $\pi$

$$\frac{1}{T/\varepsilon} \sum_{t=1}^{T/\varepsilon} \widehat{X}^{*,\varepsilon}_{t-1} \xrightarrow{\varepsilon \rightarrow 0} \pi$$

with a convergence rate of order  $\varepsilon$  under (ASe), of  $\varepsilon^\delta$  for any  $\delta \in ]0, 1[$  under (ASi). Under (ASe) with  $\alpha \in ]0, 1[$ , a similar statement also holds from Remark 17.

## 5 Conclusion

In this article, we deal with a reliability model in which the failures are generated by an underlying non-homogeneous Markov chain. The main contribution of this article is to

show that a Poisson-type approximation exists for the corresponding multivariate failure point process, when the failures tend to be rare. As a corollary, we obtain a Compound Poisson approximation for the cumulative number of failures. In this article, we only use simple stochastic calculus in discrete-time and the assumptions used here, are expected to be easily checked. It is intended that this approach can lead to similar results for much more general marking of failure events than the independence case considered here.

## A Rate of convergence of non-homogeneous Markov chains

For a sequence of stochastic matrices  $(P(t))_{t \in \mathbb{N}}$  and  $t \geq n \geq 0$ , the forward-product matrix  $P^{(n,t)}$  is

$$P^{(n,t)} := \begin{cases} P(n) \times \cdots \times P(t-1) & t > n \geq 0 \\ I & t = n. \end{cases} \quad (\text{A.1})$$

A non-homogeneous Markov chain with transition matrices  $(P(t))_{t \in \mathbb{N}}$  and initial distribution  $p_0$ , is said to be *strongly ergodic* if there exists a stochastic vector  $\pi$  such that

$$\forall n \geq 0, \quad \lim_{t \rightarrow +\infty} \sup_{p_0} \|p_0 P^{(n,n+t)} - \pi\|_1 = 0.$$

We know from [IM76, Th V.4.1] that  $(P(t))_{t \in \mathbb{N}}$  is strongly ergodic iff there exists a stochastic vector  $\pi$  such that

$$\forall n \geq 0, \quad \lim_{t \rightarrow +\infty} \|P^{(n,n+t)} - \mathbf{1}^\top \pi\|_1 = 0.$$

If the previous limit is uniform in  $n$ , then  $(P(t))_{t \in \mathbb{N}}$  is said to be *uniformly strongly ergodic*. We have the following basic result [HIV76].

**Lemma 19** *Let  $P$  an ergodic matrix with invariant distribution  $\pi$ . If the sequence  $(P(t))_{t \in \mathbb{N}}$  converges to  $P$  in norm  $l_1$  then  $(P(t))_{t \in \mathbb{N}}$  is uniformly strongly ergodic*

$$\lim_{t \rightarrow +\infty} \sup_{n \geq 0} \|P^{(n,n+t)} - \mathbf{1}^\top \pi\|_1 = 0$$

**Remark 20** In our finite-dimensional context, note that the  $l_1$  convergence of matrices is equivalent to the component-wise convergence.

**Remark 21** We mention that there are standard conditions for the (uniform) weak or strong ergodicity are in terms of spectral properties of each matrix  $P(t)$  (e.g. [IM76, Ios80]). This is the point of view used in [YZB03]. The convergent case is favored here, from the reliability context.

We are concerned here with rate of convergence. For a strongly ergodic matrix  $P$  according to the definition above, it is well known that the convergence of the forward-product  $P^{(n,n+t)} = P^t$  is geometric: there is a constant  $\kappa$  such that for all  $t \geq 0$ ,

$$\|P^t - \mathbf{1}^\top \pi\|_1 \leq \kappa r^t.$$

In the non-stationary case, we will use the following form of the main result of [HIV76].

**Lemma 22** *Let  $P$  be an ergodic matrix with invariant distribution  $\pi$ . If the sequence of stochastic matrices  $(P(t))_{t \in \mathbb{N}}$  converges in norm  $l_1$  to  $P$  with*

$$\lim_{t \rightarrow +\infty} (2t)^\alpha \|P(t) - \mathbf{1}^\top \pi\|_1 = 0$$

for some  $\alpha > 0$ , then

$$\lim_{t \rightarrow +\infty} t^\alpha \sup_{n \geq 0} \|P^{(n, n+t)} - \mathbf{1}^\top \pi\|_1 = 0$$

If  $\alpha$  is greater than 1 in the previous lemma, then it is clear that the following series converge:

$$\sup_{n \geq 0} \sum_{t \geq 0} \|P^{(n, n+t)} - \mathbf{1}^\top \pi\|_1 \leq \sum_{t \geq 0} \sup_{n \geq 0} \|P^{(n, n+t)} - \mathbf{1}^\top \pi\|_1 \leq C < \infty \quad (\text{A.2})$$

This gives our basic assumption (ASe) on the convergence of  $P(t)$  (see page 9). The following properties are deduced from the previous result and are used in the proof of Theorem 3

$$\sum_{t \geq 0} \|P^{(0, t)} - \mathbf{1}^\top \pi\|_1 \leq C \quad (\text{A.3})$$

$$\sum_{t=0}^{T/\varepsilon-1} \sum_{n=1}^t \|P^{(n, t)} - \mathbf{1}^\top \pi\|_1 \leq \frac{T}{\varepsilon} C \quad (\text{A.4})$$

$$\sum_{t=0}^{T/\varepsilon-1} \sum_{l=1}^t \sum_{k=l+1}^t \|P^{(k, t)} - \mathbf{1}^\top \pi\|_1 \leq \left(\frac{T}{\varepsilon}\right)^2 C \quad (\text{A.5})$$

where  $C$  does not depend on  $\varepsilon, T$ . In the inequalities above, we adopt the convention that  $\sum_{k=t+1}^t$  for  $t \geq 0$  is null.

The first inequality is immediate from (A.2). We find the second inequality from

$$\begin{aligned} \sum_{t=0}^{T/\varepsilon-1} \sum_{n=1}^t \|P^{(n, t)} - \mathbf{1}^\top \pi\|_1 &= \sum_{n=1}^{T/\varepsilon-1} \sum_{t=n}^{T/\varepsilon} \|P^{(n, t)} - \mathbf{1}^\top \pi\|_1 \\ &\leq \sum_{n=1}^{T/\varepsilon-1} \sum_{t=n}^{+\infty} \|P^{(n, t)} - \mathbf{1}^\top \pi\|_1 \\ &\leq \frac{T}{\varepsilon} C \quad \text{from (A.2)} \end{aligned}$$

Finally, inequality (A.5) is proved as follows. First, we have for  $t \geq 2$ :

$$\begin{aligned} \sum_{l=1}^t \sum_{k=l+1}^t \|P^{(k, t)} - \mathbf{1}^\top \pi\|_1 &= \sum_{l=1}^{t-1} \sum_{k=l+1}^t \|P^{(k, t)} - \mathbf{1}^\top \pi\|_1 \\ &= \sum_{k=2}^t (k-1) \|P^{(k, t)} - \mathbf{1}^\top \pi\|_1. \end{aligned}$$

Second, summing over  $t$  we obtain

$$\begin{aligned}
& \sum_{t=0}^{T/\varepsilon-1} \sum_{l=1}^t \sum_{k=l+1}^t \|P^{(k,t)} - \mathbf{1}^\top \pi\|_1 = \sum_{t=2}^{T/\varepsilon-1} \sum_{l=1}^t \sum_{k=l+1}^t \|P^{(k,t)} - \mathbf{1}^\top \pi\|_1 \\
&= \sum_{t=2}^{T/\varepsilon-1} \sum_{k=2}^t (k-1) \|P^{(k,t)} - \mathbf{1}^\top \pi\|_1 = \sum_{k=2}^{T/\varepsilon-1} \sum_{t=k}^{T/\varepsilon-1} (k-1) \|P^{(k,t)} - \mathbf{1}^\top \pi\|_1 \\
&\leq \sum_{k=2}^{T/\varepsilon-1} (k-1) \sum_{t=k}^{+\infty} \|P^{(k,t)} - \mathbf{1}^\top \pi\|_1 \\
&\leq C \sum_{k=2}^{T/\varepsilon-1} (k-1) \quad \text{from (A.2)}.
\end{aligned}$$

**Remark 23** Let  $0 < r < 1$  be the geometric rate of convergence of  $\|P^t - \mathbf{1}^\top \pi\|_1$  to 0. In Lemma 22, let us replace the function  $(2t)^\alpha$  by any strictly increasing function  $g(2t)$ . Then, the conclusion of the lemma is now [HIV76]

$$\lim_{t \rightarrow +\infty} \min(g(t), \lambda^t) \sup_{n \geq 0} \|P^{(n, n+t)} - \mathbf{1}^\top \pi\|_1 = 0$$

for any  $\lambda \in ]1, 1/\sqrt{r}[$ . In such a case, if function  $g(2t)$  satisfies  $g(2t) \leq \lambda^t$  for  $t$  large enough and  $\sum_t 1/g(2t) < +\infty$ , then Property (A.2) is still valid and (A.3-A.5) as well. It easily follows that Theorem 3 holds under the condition above.

Next, we consider the case where the limit matrix  $P$  is only assumed to be irreducible. Let  $\pi$  be the invariant distribution of  $P$ . In that case, we consider the Cesaro sums of the forward product (A.1), that is

$$C(n, t) := \frac{1}{t} \sum_{k=1}^t P^{(n, n+k)}. \tag{A.6}$$

If the sequence of non-stationary stochastic matrices  $(P(t))_{t \in \mathbb{N}}$  converges to an irreducible matrix  $P$ , it has been shown in [BDI77] that the Cesaro sums (A.6) converge to  $\mathbf{1}^\top \pi$  uniformly in  $n$ , with an upper bound on the convergence rate which is reported in the next result.

**Lemma 24** *Let  $P$  be an irreducible matrix with invariant distribution  $\pi$ . If the sequence of stochastic matrices  $(P(t))_{t \in \mathbb{N}}$  converges in norm  $l_1$  to  $P$  with*

$$\lim_{t \rightarrow +\infty} t^\alpha \|P(t) - \mathbf{1}^\top \pi\|_1 = 0$$

for some  $\alpha > 1$ , then there is, for any  $\delta > 0$ , a  $D(\delta)$  such that

$$\sup_{n \geq 0} \|C(n, t) - \mathbf{1}^\top \pi\|_1 \leq \frac{D(\delta)}{t^{1-\delta}} \quad t \geq 1.$$

This leads to our basic assumption (ASi) on the convergence of  $P(t)$  (see page 12). In fact, the previous estimate has the following form: for any  $t \geq 1$

$$\sup_{n \geq 0} \left\| \sum_{k=1}^t (P^{(n, n+k)} - \mathbf{1}^\top \pi) \right\|_1 \leq D(\delta) t^\delta.$$

In a similar way than for (A.3-A.5), the following properties may be deduced from the estimate above

$$\left\| \sum_{k=0}^{T/\varepsilon-1} (P^{(0, k)} - \mathbf{1}^\top \pi) \right\|_1 \leq D(\delta) \left( \frac{T}{\varepsilon} \right)^\delta \quad (\text{A.7})$$

$$\sum_{l=1}^{T/\varepsilon-1} \left\| \sum_{k=l}^{T/\varepsilon-1} (P^{(l, k)} - \mathbf{1}^\top \pi) \right\|_1 \leq D(\delta) \left( \frac{T}{\varepsilon} \right)^{1+\delta} \quad (\text{A.8})$$

$$\sum_{l=1}^{T/\varepsilon-1} \sum_{m=l+1}^{T/\varepsilon-1} \left\| \sum_{n=m}^{T/\varepsilon-1} (P^{(m, n)} - \mathbf{1}^\top \pi) \right\|_1 \leq D(\delta) \left( \frac{T}{\varepsilon} \right)^{2+\delta} \quad (\text{A.9})$$

where  $D(\delta)$  does not depend on  $\varepsilon, T$ . In the inequalities above, we adopt the convention that  $\sum_{k=t+1}^t$  for  $t \geq 0$  is null.

## Acknowledgment

The author thanks the referees for their detailed comments which led to a substantial improvement in the presentation of the article.

## References

- [Ani83] Anisimov, V.V. Limits theorems for nonhomogeneous weakly dependent summations. *Theory Probab. Math. Statist.*, 27:9–21, 1983.
- [Ani88] Anisimov, V.V. Estimates for deviations of transient characteristics of nonhomogeneous Markov processes. *Ukrainian Math. J.*, 40:588–592, 1988.
- [Ani95] Anisimov, V.V. Switching processes: Averaging principle, diffusion approximation and applications. *Acta Appl. Math.*, 40:95–141, 1995.
- [Ani02] Anisimov, V.V. Averaging in models with fast Markov switches and applications to queueing models. *Ann. Oper. Res.*, 112(1):63–82, 2002.
- [BDI77] Bowerman, B., David, H.T., and Isaacson, D. The convergence of Cesaro averages for certain nonstationary Markov chains. *Stochastic Process. Appl.*, 5:221–230, 1977.

- [BHJ92] Barbour, A.D., Holst, L., and Janson, S. *Poisson Approximation*. Oxford Studies in Probability –2. Oxford Science Publications, 1992.
- [Bil68] Billingsley, P. *Convergence of probability measures*. John Wiley & Sons, 1968.
- [Bro82] Brown, T.C. Poisson approximations and exchangeable random variables. In Koch, G and Spizzichino, F., editors, *Poisson approximations and exchangeable random variables*, pages 177–183. North-Holland, 1982.
- [Bro83] Brown, T.C. Some Poisson approximations using compensators. *Ann. Probab.*, 11:726–744, 1983.
- [Che80] Cheung, R.C. A user-oriented software reliability model. *IEEE Trans. Software Eng.*, 6:118–125, 1980.
- [CX04] Chen, L.H.Y. and Xia, A. Stein’s method, Palm theory and Poisson process approximation. *Ann. Probab.*, 32:2545–2569, 2004.
- [Çin72] E. Çinlar. Markov additive processes Part I. *Z. Wahrsch. Verw. Gebiete*, 24:85–93, 1972.
- [DK93] Di Masi, G.B. and Kabanov, Y.M. The strong convergence of two-scale stochastic systems and singular perturbations of filtering equations. *J. Math. Systems, Estimation, and Control*, 3:207–224, 1993.
- [ES69] Ezhov, I.I. and Skorokhod, A.V. Markov processes with homogeneous second component II. *Theory Probab. Appl.*, 14:652–667, 1969.
- [GL04] Gravereaux, J.-B. and Ledoux, J. Poisson approximation for some point processes in reliability. *Adv. in Appl. Probab.*, 36:455–470, 2004.
- [Gri79] B. Grigelionis. Additive Markov processes. *Lithuanian Math. J.*, 18:340–342, 1979.
- [HIV76] Huang, C.-C., Isaacson, D., and Vinograd, B. The rate of convergence of certain nonhomogeneous Markov chains. *Z. Wahrsch. Verw. Gebiete*, 35:141–146, 1976.
- [IM76] Isaacson, D.L. and Madsen, R.W. *Markov chains theory and applications*. Robert E. Krieger Publishing Company, 1976.
- [Ios80] Iosifescu, M. *Finite Markov Processes and Their Applications*. John Wiley & Sons, 1980.
- [JS89] Jacod, J. and Shiryaev, A.N. *Limit Theorems for Stochastic Processes*. Springer, 1989.
- [KLS83] Kabanov, Y.M., Liptser, R.S., and Shiryaev, A.N. Weak and strong convergence of the distributions of counting processes. *Theory Probab. Appl.*, 28:303–336, 1983.

- [Kov94] Kovalenko, I.N. Rare events in queueing systems—a survey. *Queueing Syst. Theory Appl.*, 16:1–49, 1994.
- [Led05] Ledoux, J. Recursive filters for partially observable finite Markov chains. *J. Appl. Probab.*, 42: 684–697, 2005.
- [Lit75] Littlewood, B. A reliability model for systems with Markov structure. *Appl. Statist.*, 24:172–177, 1975.
- [LR99] Latouche, G. and Ramaswami, V. *Introduction to Matrix Analytic Methods in Stochastic Modeling*. series on Statistics and Applied Probability. SIAM, 1999.
- [PMM93] Poore, J. H., Mills, H. D., and Mutchler, D. Planning and certifying software system reliability. *IEEE Software*, 10:89–99, 1993.
- [Seg76] Segall, A. Recursive estimation from discrete-time point process. *IEEE Trans. Inf. Theory*, 22:422–431, 1976.
- [Ser75] Serfling, R.J. A general Poisson approximation theorem. *Ann. Probab.*, 3:726–731, 1975.
- [Xia00] Xia, A. Poisson approximation, compensators and coupling. *Stochastic Anal. Appl.*, 18:159–177, 2000.
- [YZB03] Yin, G., Zhang, Q., and Badowski, G. Discrete-time singularly perturbed Markov chains: aggregation, occupation measures, and switching diffusion limit. *Adv. in Appl. Probab.*, 35:449–476, 2003.