A simple Markovian individual-based model as a means of understanding forest dynamics

Khader Khadraoui

Laval University, Department of Mathematics and Statistics, Quebec city, G1V 0A6, Canada.

Abstract

The forests are ecological systems of great complexity which present interaction phenomena associated with the competition between individuals of the same and of different species. This competition is about access to resources (light, water, nutrients, ...). The scales of forest ecosystems are very long. For this reason we use in this framework Markovian modeling for space and time evolution of population distribution thanks to an individual-based model in the form of a stochastic branching process. Understanding the behaviour of individual-based models of forest dynamics becomes difficult as their complexity increases. A useful strategy consists in simplifying parts of the original model in order to simulate a simplified version of forest dynamics. This strategy is adopted to understand the spatial pattern structured population in a simple individual-based model. The model is made of two components: birth or recruitment and death (natural mortality and mortality due to competition). The interplay between the spatial pattern of trees and the competition/dispersion level is thus understood and the excessive impact of the low dispersion that favours the establishment of clusters is diagnosed.

Keywords: forests dynamics, spatial pattern, interaction, competition, branching stochastic process, individual-based model (IBM), Monte Carlo.

1. Introduction

Our aim is to present a Markovian individual-based model for forest dynamics. This framework does not pretend to offer a faithful model of forest ecological systems, but to propose a simplified modeling taking into account two key features. Firstly, birth and dispersal: each individual is capable of give birth to a new tree located in the vicinity of the parent tree (Dispersal). Secondly, death or mortality: the death is due to natural causes or to competition for access to natural resources. The growth will not be discussed in this framework.

Developments in forest dynamics modelling have led to complex individual-based models, Deutschman [8, 1997]. This tendency is not limited to forestry but more generally occurs in ecology (12, 13...).
and [23]). In this kind of dynamics modelling, the level of description is the tree or, more generally, the individual and the environment is spatially explicit so that a local neighbourhood is defined around each tree. A tree then interacts with its neighbours. However, analysing and understanding such complex models is difficult [23]. As mathematical analysis often is intractable, the study of the model relies on tedious simulation runs. Moreover, we can lose sight of the important phenomena. For example, it becomes harder to understand the influence of spatial interactions on the formation of patterns, Parrott [21, 2005]. To better understand the behaviour of complex models, model simplification has arisen as a tool, together with sensitivity analysis, [18], Garcia [12, 2003]. Model simplification consists in replacing the model with a simplified version. It thus permits the identification of processes, variables or parameters that play a significant role in the model. Moreover the simplified model may sometimes be amenable to mathematical analysis.

In this study, model simplification will be treated in the context of a Markovian model of particles interacting in continuous time and space. It means that branching processes (birth/death) are particularly adapted to this situation. We are particularly interested to a model developed by Fournier and Méleard [11]. This model had originally been proposed by Bolker and Pacala [3]. The treatments in these two items are different. Bolker and Pacala [3] described the Bolker equation, called the Kolmogorov equation, which describes the distribution law of the ecosystem process. This equation is complex and very useful in many contexts. Then, the authors propose a method to approximate the moments which allows to find the first two moments of the exact solution. Fournier and Méleard [11] used this model and provides a rigorous mathematical formulation. They also produce a numerical algorithm to simulate the underlying Markov process by using a Monte Carlo scheme.

The goal of this study is to present a family of individuals whose state includes only their position. Each individual is subject to random point processes (reproduction and mortality) [5], [6], [7]. The individuals interact through the competition mechanism [2], [19]. In Section 2 we describe the model in detail. In Section 3 we study the Markovian representation of the process. We describe the representation of the model in the simple form in Section 4 and we study some numerical Monte Carlo simulation tests in Section 5. Finally, we propose to estimate the model parameters thanks to a stochastic algorithm and we prove the almost sure convergence of the estimator. The numerical simulations are made in MatLab.
2. The model

Let us first describe in detail the model, originally proposed by Bolker and Pacala [3] then developed by Fournier and Mélaërd [11]. The plants are supposed to be a family of individuals whose state includes only their spatial location. We assume that the spatial domain is

\[ \mathcal{X} = \mathcal{S} \times \mathbb{R}^d, \]

where \( \mathcal{S} \) is a measurable connected subset of \( \mathbb{R}^n \) for some \( n, d \geq 1 \). The state \( x \in \mathcal{X} \) of a plant represents its spatial location in the space \( \mathcal{S} \). We note here that Campillo and Joannides [4] considered a more general state \( x = (p, r) \in \mathcal{X} \) where \( p \) represents the position of individual and \( r \) an associated characteristic that could represent its size or its maturation age. In the following, we represent an individual at state \( x \in \mathcal{X} \) as the Dirac measure \( \delta_x \). Hence the population at time \( t \) will be:

\[ \nu_t(dx) = \sum_{i=1}^{N_t} \delta_{x^i}(dx), \]

where \( N_t \in \mathbb{N} \) stands for the number of plants alive at time \( t \) and \( x_1^1, \ldots, x_{N_t}^t \) describe their locations in \( \mathcal{S} \). We denote by \( \mathcal{M}_F(\mathcal{X}) \) the set of finite nonnegative measures on \( \mathcal{X} \) and \( \mathcal{M} \subset \mathcal{M}_F(\mathcal{X}) \) that consists of all finite point measures on \( \mathcal{X} \):

\[ \mathcal{M} = \left\{ \sum_{i=1}^{N} \delta_{x^i}, \ N \geq 0, \ x^i \in \mathcal{X} \right\}. \]

From now on, we use the angle brackets to denote integration, thus for any \( m = \sum_{i=1}^{N} \delta_{x^i} \in \mathcal{M} \) and any measurable function \( f \) on \( \mathcal{X} \),

\[ \langle m, f \rangle = \int_{\mathcal{X}} f dm = \sum_{i=1}^{N} f(x^i). \]

Hence, \( (\nu_t)_{t \geq 0} \) is a stochastic process, taking its values in \( \mathcal{M} \) and describing the distribution of plants at time \( t \). According to [2], we define the size of the population by

\[ N_t := \langle \nu_t, 1 \rangle. \]

To make the following description of the model clear, some notations that we will use in the sequel are as follow: considering the state \( \nu = \sum_{i=1}^{N} \delta_{x^i} \) of the population at a given time, each plant located at some \( x \in \mathcal{X} \) has three independent exponential clocks. A seed production clock with parameter \( \lambda^b > 0 \), a natural death clock with parameter \( \lambda^d > 0 \) and a competition mortality clock
with parameter $\lambda^c(x, \nu) \in [0, \infty)$. Let $D(x, dx')$ be the dispersion law of the seeds of plants located at $x$. It is assumed to satisfy, for each $x \in \mathcal{X}$,

$$
\int_{x' \in \mathcal{S}, x + x' \in \mathcal{X}} D(x, dx') = 1, \quad \int_{x' \in \mathcal{S}, x + x' \notin \mathcal{X}} D(x, dx') = 0.
$$

For $x, y \in \mathcal{X}$, $u(x, y) = u(y, x) \in [0, \infty)$ is the competition kernel which describes the strength of competition between plants located at $x$ and $y$. Obviously, an individual will be subject to 3 types of punctual events occurring at specific clocks:

- **Birth and dispersal**: The individual in state $x$ gives birth to a new individual at a rate $\lambda^b$ which does not depend on its state. This seed immediately becomes a mature plant. Its location is given by $x + x'$ where $x'$ is randomly chosen according to the dispersion kernel $D(x, dx')$.

- **Natural death**: The individual in state $x$ disappears at a rate $\lambda^d$ which does not depend on its state. This death is called natural or intrinsic as it does not depend on the state of all the population.

- **Competition death**: The individual in state $x$ disappears at a rate $\lambda^c(x; \nu)$ which may depend on its state and on the state of the population $(x, \nu)$. In this paper, we consider the competition rate:

$$
\lambda^c(x, \nu) = \sum_{y \in \nu} u(x, y) = \int_{\mathcal{X}} u(x, y)\nu(dy),
$$

where $u(x, y)$ is the contribution of an individual in state $y$ to the competition affecting an individual in state $x$.

A natural hypothesis will be considered in the sequel which consists on supposing that these mechanisms of birth with dispersion, natural or intrinsic death and competition are mutually independent. We also suppose that the dispersion kernel induces a density with respect to the Lebesgue measure on $\mathcal{S}$. This density is given by

$$
D(x, dx') = D(x, x')dx'.
$$

Hence, the model considered in this section allows to describe a spatial ecological system that consists of motionless individuals (such as plants). Individuals are characterized by their location. We assume that each plant produces seeds at a given rate. When a seed is born, it immediately disperses from its mother and becomes a mature plant (we do not consider growth in this model). For birth with dispersion, natural death and death due to competition, all these events occur randomly in continuous time. In the following Section 3, we wish to give a rigorous mathematical analysis for the Markovian representation of the process $(\nu_t)_{t \geq 0}$ by studying its behavior and some properties.
3. Markovian representation of the process

Let \((\Omega, \mathcal{A}, \mathbb{P})\) be a probability space. On this space, we consider that the parameters \(\lambda^c(x, \nu)\) and \(D(x, x')\) of independent random punctual events were assumed to be space-dependent. We have left open so far the precise distribution of the clocks, other than to specify that it be independent. As the exponential distribution is memoryless, note that assuming that all the clocks of events are exponentially distributed allows us to reset all the clocks to zero each time one clock rings. For reasons of simplicity, we assume that the local clock of death due to competition is uniformly bounded (unless otherwise stated) as the hypothesis \((H)\) is:

\[ H: \]

(i) We assume that there exist some positive real \(C_{\text{max}}\) such that:

\[
 u(x, y) \leq C_{\text{max}}. \tag{5}
\]

Hence:

\[
 \lambda^c(x, \nu) \leq C_{\text{max}} \langle \nu, 1 \rangle.
\]

(ii) We also assume that there exist a constant \(\alpha > 0\) and a probability density \(\tilde{D}\) on \(S\) such that, for all \(x \in \mathcal{X}\),

\[
 D(x, x') \leq \alpha \tilde{D}(x').
\]

The assumption \((H)(i)\) is convenient since it avoids explosion phenomena (due to the accumulation of many events at a given time). The assumption \((H)(ii)\) allows to control the dispersion in the vicinity of the plant mother.

We aim to study the stochastic process \((\nu_t)_{t \geq 0}\), taking its values in \(\mathcal{M}\) and describing the distribution of plants at time \(t\). For this reason, we give a pathwise representation of the model in terms of Poisson point measures.

**Definition 3.1.** We set \(\lambda^G = \lambda^b + \lambda^d + C_{\text{max}}\langle \nu, 1 \rangle\). Let:

(i) \(N^b\) be a Poisson random measure on \([0, \infty) \times \mathbb{N}^* \times S \times [0, 1]\) of intensity measure:

\[
 n^b(d\tau, di, dx', d\theta) := \lambda^G \alpha \tilde{D}(x') d\tau dx' d\theta.
\]

where \(d\tau, dx'\) and \(d\theta\) are the Lebesgues measures on \([0, \infty), S\) and \([0, 1]\) and \(di\) is the counting measure on \(\mathbb{N}^*\).

(ii) \(N^d\) be a Poisson random measure on \([0, \infty) \times \mathbb{N}^* \times [0, 1]\) of intensity measure:

\[
 n^d(d\tau, di, d\theta) := \lambda^G d\tau di d\theta.
\]
(iii) \( N^c \) be a Poisson random measure on \([0, \infty) \times \mathbb{N}^* \times \mathbb{N}^* \times [0, 1]\) of intensity measure:

\[
n^c(d\tau, di, dj, d\theta) := C \max d\tau didjd\theta.
\]

We consider the canonical filtration \((F_t)_{t \geq 0}\) generated by stochastic processes \(N^b, N^d\) and \(N^c\). We aim to give a pathwise description in terms of Poisson point measures.

3.1. The infinitesimal generator of the process

We wish to describe the system by the evolution in time of the empirical measure \(\nu_t\) with infinitesimal generator \(L\), defined for a large class of functions \(\psi\) from \(\mathcal{M}\) into \(\mathbb{R}\), for all \(\nu \in \mathcal{M}\). To give an explicit expression for the infinitesimal generator \(L\) of the process \((\nu_t)_{t \geq 0}\) in the following Proposition \ref{prop0}, we consider the path space \(T \subset \mathbb{D}([0, \infty), \mathcal{M}_F(X))\) defined by

\[
T := \left\{(\nu_t)_{t \geq 0} \mid \forall t \geq 0, \nu_t \in \mathcal{M}, \text{ and } \exists 0 = t_0 < t_1 < t_2 < \cdots, \lim_{n \to \infty} t_n = \infty \text{ and } \nu_t = \nu_{t_i} \forall t \in [t_i, t_{i+1}) \right\};
\]

and for \((\nu_t)_{t \geq 0} \in T\) and \(t \geq 0\), we can define \(\nu_{t-}\) by

\[
\nu_{t-} := \begin{cases} 
\nu_t & \text{if } t \notin \cup_i \{t_i\}, \\
\nu_{t_{i-1}} & \text{while if } t = t_i \text{ for some } i \geq 1.
\end{cases}
\]

Before presenting the following statement, we consider the notation \(N_{t-} = \langle \nu_{t-}, 1 \rangle\).

**Proposition 3.2.** Admit the hypothesis \(H\). A \((F_t)_{t \geq 0}\)-adapted stochastic process \((\nu_t)_{t \geq 0}\) that belongs a.s. to \(T\). The process \((\nu_t)_{t \geq 0}\) is a Markov process and its infinitesimal generator \(L\) is defined for the following set \(S\) of test functions \(\psi : \mathcal{M}_F(X) \mapsto \mathbb{R}\) of the form \(\psi(\nu) = F(\langle \nu, f \rangle)\) for any function \(f : X \mapsto \mathbb{R}\) and \(F : \mathbb{R} \mapsto \mathbb{R}\) twice continuously differentiable, bounded with bounded derivatives. The infinitesimal generator \(L\) is defined by:

\[
L \psi(\nu) = L^b + L^d + L^c
\]

where

\[
L^b := \lambda_b \int_X \left\{ \int_S \left\{ \psi(\nu + \delta_x + x') - \psi(\nu) \right\} D(x, x') dx' \right\} \nu(dx);
\]

\[
L^d := \lambda_d \int_X \left\{ \psi(\nu - \delta_x) - \psi(\nu) \right\} \nu(dx);
\]

\[
L^c := \int_X \left\{ \psi(\nu - \delta_x) - \psi(\nu) \right\} \left\{ \int_X u(x, y) \nu(dy) \right\} \nu(dx).
\]
Proof. The different random events (birth, intrinsic death, death due to competition) described in Section 2 implies that, for all $t \geq 0$, the process $\nu_t$ is given a.s. by

$$\nu_t = \nu_0 + \int_0^t \int_{N^*} \int_S \int_0^1 1_{\{i \leq N_{t-}\}} 1_{\{\theta \leq \langle \lambda^B D(x^1_{t-}, x^*) \rangle / (\lambda^G a D(x^*) \rangle \}} \times \delta_{(x^1_{t-} + x^*')}\mathcal{N}_d(d\tau, di, dx', d\theta)$$

$$- \int_0^t \int_{N^*} \int_S \int_0^1 1_{\{i \leq N_{t-}\}} 1_{\{\theta \leq \langle \lambda^G D(x^1_{t-}, x^*) \rangle \}} \delta_{(x^1_{t-})}\mathcal{N}_d(d\tau, di, d\theta)$$

$$- \int_0^t \int_{N^*} \int_S \int_0^1 1_{\{i \leq N_{t-}\}} 1_{\{i \leq N_{t}\}} 1_{\{j \leq N_{t-}\}} 1_{\{\theta \leq \langle u(x^1_{t-}, x^*_{t-}) \rangle / (C_{max}) \}} \delta_{(x^1_{t-})}\mathcal{N}_c(d\tau, di, dj, d\theta),$$

where $1_A$ denotes the indicator function of the set $A$ and the three terms of integrals are associated with the three basic independent mechanisms (respectively birth, intrinsic death and death due to competition). In interpreting the expression of $\nu_t$, it is very simple to see that the indicator functions that involve $\theta$ are related to the individual chosen at random ($i \in \{1, \ldots, \langle \nu_t \rangle, 1\}$) and the type of the punctual mechanism to be considered is determined thanks to an acceptance/rejection sampling technique. The proof for the fact that $(\nu_t)_{t \geq 0}$ is a Markov process is classical and for this reason it will be omitted to save place. For any function $\psi = (F, f) \in \mathcal{S}$ and using the fact that a.s. $\psi(\nu_t) = \psi(\nu_0) + \sum_{t \leq \tau} [\psi(\nu_{\tau-} + \{\nu_{\tau} - \nu_{\tau-}\}) - \psi(\nu_{\tau-})]$ where $\nu_0 = \sum_{i=1}^{N_0} \delta_{x_0^i}$, we can write:

$$\psi(\nu_t) = \psi(\nu_0) + \int_0^t \int_{N^*} \int_S \int_0^1 1_{\{i \leq N_{t-}\}} 1_{\{\theta \leq \langle \lambda^B D(x^1_{t-}, x^*) \rangle / (\lambda^G a D(x^*) \rangle \}} \times [\psi(\nu_{\tau-} + \{\delta_{(x^1_{\tau-} + x^*')}\}) - \psi(\nu_{\tau-})]\mathcal{N}_d(d\tau, di, dx', d\theta)$$

$$+ \int_0^t \int_{N^*} \int_S \int_0^1 1_{\{i \leq N_{t-}\}} 1_{\{\theta \leq \langle \lambda^G D(x^1_{t-}, x^*) \rangle \}} \times [\psi(\nu_{\tau-} - \{\delta_{(x^1_{\tau-})}\}) - \psi(\nu_{\tau-})]\mathcal{N}_d(d\tau, di, d\theta)$$

$$+ \int_0^t \int_{N^*} \int_S \int_0^1 1_{\{i \leq N_{t-}\}} 1_{\{j \leq N_{t-}\}} 1_{\{\theta \leq \langle u(x^1_{t-}, x^*_{t-}) \rangle / (C_{max}) \}} \times [\psi(\nu_{\tau-} - \{\delta_{(x^1_{\tau-})}\}) - \psi(\nu_{\tau-})]\mathcal{N}_c(d\tau, di, dj, d\theta).$$
Let us now consider the Definition 3.1 and taking expectations, we shows that

\[ \mathbb{E}[\psi(\nu_t)] = \mathbb{E}[\psi(\nu_0)] + \int_0^t \mathbb{E} \left[ \lambda^G \alpha \dot{D}(x') \sum_{i=1}^{N_r} \frac{\lambda^b}{\lambda^G} \beta \delta \{x' \} \right] d\tau \]

\[ \times \int_S \left\{ \psi(\nu_{r-} + \{\delta(x_i', x')\}) - \psi(\nu_{r-}) \right\} D(x_i', x') dx' d\tau \]

\[ + \int_0^t \mathbb{E} \left[ \lambda^G \sum_{i=1}^{N_r} \lambda^b \frac{\lambda^G}{\lambda^b} \right] \left\{ \psi(\nu_{r-} - \{\delta(x_i', x')\}) - \psi(\nu_{r-}) \right\} d\tau \]

\[ + \int_0^t \mathbb{E} \left[ C_{\max} \sum_{i=1}^{N_r} \lambda^b \frac{\lambda^G}{\lambda^b} \sum_{j=1}^{N_r} \frac{u(x_i', x_j')}{C_{\max}} \left\{ \psi(\nu_{r-} - \{\delta(x_i', x')\}) - \psi(\nu_{r-}) \right\} \right] d\tau \]

\[ = \mathbb{E}[\psi(\nu_0)] + \lambda^b \int_0^t \mathbb{E} \left[ \int_X \nu_r(dx) \right] \]

\[ \times \int_S \left\{ \psi(\nu_r + \{\delta(x_i', x')\}) - \psi(\nu_r) \right\} D(x, x') dx' d\tau \]

\[ + \int_0^t \mathbb{E} \left[ \int_X \left\{ \psi(\nu_r - \delta(x)) - \psi(\nu_r) \right\} \lambda^d \right] d\tau. \]  

(10)

We recall that \( \mathcal{L}\psi(\nu_0) = \partial_t \mathbb{E}[\psi(\nu_t)] \). Now differentiating the expression (10) at \( t = 0 \) leads to

\[ \mathcal{L}\psi(\nu) = \lambda^b \int_X \nu(dx) \int_S \left\{ \psi(\nu + \delta(x')) - \psi(\nu) \right\} D(x, x') dx' \]

\[ + \int_X \left\{ \psi(\nu - \delta(x)) - \psi(\nu) \right\} \lambda^d \]

\[ \times \int_X u(x, y) \nu(dy) \]  

This completes the proof.

3.2. Some properties of the process

Studying the properties of the process \((\nu_t)_{t \geq 0}\) is of considerable interest. Obviously, we focus on the size \( N_t \) of the population, the extinction and survival properties of the process \((\nu_t)_{t \geq 0}\). We precise that these properties has been studied by many mathematicians: for the contact process we cite [17] and for the extinction/persistence results of the Bolker and Pacala model we cite [10].

Before giving our results about extinction/survival of the process \((\nu_t)_{t \geq 0}\), let us point out the following obvious statement shows that the population will grow with bound.

**Theorem 3.3.** Under the hypothesis \( \mathbb{H} \) and assuming that \( \mathbb{E}[N_0^p] < \infty \) for some \( p \geq 1 \), then for any \( \tau > 0 \) we have

\[ \mathbb{E} \left[ \sup_{t \in [0, \tau]} N_t^p \right] < \infty. \]

(11)

8
Proof. In the sequel $C_p$ denotes some positive constant which may change from one line to another line. We also assume without loss of generality that $\psi(\nu) = (\nu, 1)^p$ (by setting $f = 1$ and $F(a) = a^p$). To make the proof clear, it is easy to remark from (9) that the natural death term

$$\int_0^t \int_{\mathbb{N}^p} \int_{\mathbb{R}^p} 1_{\{t \leq N_{r-}\}} 1_{\theta \leq (\lambda^b D(x_{\nu-}^\ell))} \times [(N_{r-} - 1)^p - N_{r-}^p] N^d (d\tau, di, d\theta) < 0$$

and the death due to competition term

$$\int_0^t \int_{\mathbb{N}^p} \int_{\mathbb{R}^p} 1_{\{t \leq N_{r-}\}} 1_{\{\nu(x_{\nu-}^\ell)\}} \times [(N_{r-} - 1)^p - N_{r-}^p] N^b (d\tau, di, dj, d\theta) < 0,$$

which immediately yields

$$N_t^p \leq N_0^p + \int_0^t \int_{\mathbb{N}^p} \int_{\mathbb{R}^p} 1_{\{t \leq N_{r-}\}} 1_{\theta \leq (\lambda^b D(x_{\nu-}^\ell))} \times [(N_{r-} - 1)^p - N_{r-}^p] N^b (d\tau, di, dx', d\theta)$$

We consider, for each $n \geq 1$, the stopping time $\tau_n = \inf\{t \geq 0, N_t \geq n\}$. Then,

$$\sup_{t \leq [0, \tau \land \tau_n]} N_t^p \leq N_0^p + \int_0^{\tau \land \tau_n} \int_{\mathbb{N}^p} \int_{\mathbb{R}^p} 1_{\{t \leq N_{r-}\}} 1_{\theta \leq (\lambda^b D(x_{\nu-}^\ell))} \times [(N_{r-} - 1)^p - N_{r-}^p] N^b (d\tau, di, dx', d\theta)$$

$$\leq N_0^p + C_p \int_0^{\tau \land \tau_n} \int_{\mathbb{N}^p} \int_{\mathbb{R}^p} 1_{\{t \leq N_{r-}\}} 1_{\theta \leq (\lambda^b D(x_{\nu-}^\ell))} \times [N_{r-}^p - 1] N^b (d\tau, di, dx', d\theta),$$

because $(\zeta + 1)^p - \zeta^p \leq C_p (\zeta^{p-1} + 1)$ for some constant $C_p$. Let us now take expectation,

$$\mathbb{E}\left[ \sup_{t \in [0, \tau \land \tau_n]} N_t^p \right] \leq \mathbb{E}[N_0^p] + C_p \mathbb{E}\left[ \int_0^{\tau \land \tau_n} \int_{\mathbb{N}^p} 1_{\{t \leq N_{r-}\}} [N_{r-}^p - 1] d\tau d\theta \right]$$

$$\leq \mathbb{E}[N_0^p] + C_p \int_0^\tau \mathbb{E}\left[ 1_{\{\tau \leq \tau \land \tau_n\}} N_{r-} (N_{r-}^p - 1) \right] d\tau$$

$$\leq \mathbb{E}[N_0^p] + C_p \int_0^\tau \mathbb{E}\left[ 1_{\{\tau \leq \tau \land \tau_n\}} (N_{r-}^p + 1) \right] d\tau$$

$$\leq \mathbb{E}[N_0^p] + 2C_p \int_0^\tau \mathbb{E}\left[ N_{r-}^p \right] d\tau.$$ 

Now, thanks to the Gronwall’s lemma, we deduce that for all $n$ and for any $\tau < \infty$, there exists a constant $C_{p,\tau}$ such that

$$\mathbb{E}\left[ \sup_{t \in [0, \tau \land \tau_n]} N_t^p \right] \leq C_{p,\tau}. \quad (12)$$
Finally, for $\tau_n$ tends to infinity and thanks to the Fatou lemma, we prove (11). There may be some doubt as to why $\tau_n$ tends to infinity. If $\tau_n$ does not tend to infinity, we can find $\tau_0 < \infty$ such that $P(\sup_n \tau_n \leq \tau_0) = \epsilon_0 > 0$, then $E\left[\sup_{t \in [0, \tau_0 \wedge \tau_n]} \nu^\prime \right] \geq n^p \epsilon_0$, which contradicts (12). Hence, $\tau_n$ tends a.s. to infinity which completes the proof.

For the model introduced recently by Bolker and Pacala, Etheridge [10] has shown in his paper that if individuals in the population disperse sufficiently quickly, then the population has positive chance of survival, whereas, if they do not, then the population will die out. The techniques used in [10] are specific to continuous processes and cannot be adapted to discontinuous process. We are not able to handle a proof shown that the process $(\nu_t)_{t \geq 0}$ survives with positive probability because the problem seems more difficult than the problem of survival for the contact process. In the sequel, we prove that the process $(\nu_t)_{t \geq 0}$ does a.s. not survive (that is, $P(\exists \tau, \nu_\tau = 0) = 1$) where the state space $X$ is compact.

**Lemma 3.4.** If $X$ is compact in $\mathbb{R}^d$ and if there exist $\epsilon > 0$ and $\delta > 0$ such that $u(x, y) \geq \epsilon 1_{|x-y| \leq \delta}$, then, there exists a nondecreasing function $\phi : \mathbb{R}_+ \to \mathbb{R}_+$, satisfying $\phi(0) = 0$, such that $\lim_{x \to \infty} \phi(x) = \infty$, such that the map $x\phi(x)$ is convex on $[0, \infty)$ and for all $\nu \in \mathcal{M}$

$$\langle \nu \otimes \nu, u \rangle \geq \langle \nu, 1 \rangle \phi(\langle \nu, 1 \rangle).$$

We do not claim originality of the Lemma 3.4 and, in fact, similar result have been proved in [11]. Since we have found this particular result in the literature ([11], Remark 6.2 p.1907), we not provide a detailed proof.

Using the Lemma 3.4, we check now a result shown essentially that the population does almost surely not survive.

**Theorem 3.5.** Under the hypothesis $\mathbb{H}$ and assuming that $E[N_0] < \infty$ and $u(x, x)$ is bounded below, then

$$P(\exists \tau \geq 0, \nu_\tau = 0) = 1.$$
Proof. We assume without loss of generality that \( \psi(\nu) = \langle \nu, 1 \rangle \) (by setting \( f = 1 \) and \( F(a) = a \)). We recall that

\[
E[\langle \nu_t, 1 \rangle] = E[\langle \nu_0, 1 \rangle] + \int_0^t E\left[ \int_X \lambda^b \nu_\tau(dx) \right] d\tau = E[\langle \nu_t, 1 \rangle] + \int_0^t E\left[ \int_X \lambda^b \nu_\tau(dx) \right] d\tau
\]

As \( E[\langle \nu_t, 1 \rangle] \) is differentiable, we set \( |\Delta| = |\lambda^b - \lambda^d| \) and by differentiating \( E[\langle \nu_t, 1 \rangle] \) we obtain

\[
\partial_t E[\langle \nu_t, 1 \rangle] \leq |\Delta| E[\langle \nu_t, 1 \rangle] - E[\langle \nu_t \otimes \nu_t \rangle, u]
\]

where inequality \((14)\) is obtained from Lemma \(3.4\) and the Jensen inequality. We recall that \( \phi(x) \) is nondecreasing and tends to infinity with \( x \) and \( \phi(0) = 0 \). We consider \( x_0 \) the greatest solution of \( |\Delta|x_0 = x_0 \phi(x_0) \) and using \((14)\) we obtain for any \( t \geq 0 \), \( E[\langle \nu_t, 1 \rangle] \leq E[\langle \nu_0, 1 \rangle] \vee x_0 \). Consequently

\[
\sup_{t \geq 0} E[\langle \nu_t, 1 \rangle] < \infty.
\]

Let us deal about the \( \mathbb{N} \)-valued process \( N_t = \langle \nu_t, 1 \rangle \). For this purpose, we have for any \( n \in \mathbb{N}^* \),

\[
P \left( \lim \inf_{t \to \infty} N_t = n \right) = 0.
\]

We precise that is not hard to check \((16)\). In the case of \( \lim \inf_{t \to \infty} N_t = n \), the process \( N_t \) visits infinitely often the state \( n \) and visits a finite number of times the state \( n - 1 \). Or this is almost surely impossible because each time \( N_t \) visits the state \( n \), the probability that its next state is \( n - 1 \) is bounded from below by \( (n \gamma_0/(n \lambda^b + n \lambda^d + n^2 C_{\max})) > 0 \), where \( \gamma_0 = \lambda^d + \inf_{x \in X}(u(x, x)) > 0 \). Thus, we have almost surely

\[
\lim \inf_{t \to \infty} N_t \in \{0, \infty\}.
\]
As $N_t$ is $\mathbb{N}$-valued process and 0 is an absorbing state, we remark from (17) that a.s. $\lim_{t \to \infty} N_t$ exists and $\lim_{t \to \infty} N_t \in \{0, \infty\}$. Finally, thanks to the Fatou’s lemma and (15) we deduce
\[
\mathbb{E}\left[ \lim_{t \to \infty} N_t \right] = \mathbb{E}\left[ \liminf_{t \to \infty} N_t \right] \leq \liminf_{t \to \infty} \mathbb{E}[N_t] \leq \sup_{t \geq 0} \mathbb{E}[N_t] < \infty.
\]
Then, we conclude from (17) and (18) that a.s. $\lim_{t \to \infty} N_t < \infty$ and that a.s. $\lim_{t \to \infty} N_t = 0$ which completes the proof.

In the sequel, we take the model described in the Section 2 and we propose how to use it to simulate efficiently the forest dynamics and to present numerical tests.

4. Modeling forest dynamics

We now describe the dynamic of the process starting from an initial population state $\nu$. We consider a family of individuals that live in a set of the form:
\[
\mathcal{X} = [x_{\min}^1, x_{\max}^1] \times [x_{\min}^2, x_{\max}^2] \subset \mathbb{R}^2.
\]
For simplicity we use one hectare of the forest (bounded squared parcel). To easily manage the boundary conditions we assume that $\mathcal{X}$ is a torus.

4.1. Simplified version of forest dynamics

We aim to propose a simple model for forest dynamics that consist of a single species and that only consider the space position of trees (individuals). The forest will be subject to random point processes and an individual with state $x$ in the forest $\nu$ has three independent exponential clocks: natural mortality clock with constant rate $\lambda^d$ and mortality due to competition clock with rate $\lambda^c(x, \nu)$ and reproduction clock with constant fertility rate $\lambda^b$.

(i) The individual in state $x$ disappears (natural death) with rate $\lambda^d$ which not depends on its position. This intrinsic death does not depend on the state of forest $\nu$ and then:
\[
\nu \mapsto \nu - \delta_x.
\]
The forest loses the individual with state $x$ after each natural death.

(ii) The individual with state $x$ gives birth to a new individual with rate $\lambda^b$ which not depends on its position and does not depend on the state of forest $\nu$. The position $x'$ of the new individual will be determined by a given dispersal kernel $D(x, dx')$ and then:
\[
\nu \mapsto \nu + \delta_{x+x'}.
\]
The forest earns new individual with state $x+x'$ after each birth event.
(iii) The individual with state \( x \) disappears (competition death) with rate \( \lambda^c(x, \nu) \) which depends on its state \( x \) and the state of the population \( \nu \). This death is due to competition and then:

\[ \nu \mapsto \nu - \delta_x. \]

After this event the forest loses the individual with state \( x \).

We assume that these basic mechanisms are independent. We can consider more complex models by adding spatial dependence (for example the birth and intrinsic death rates may depend on the space position) or we can also consider a growth part in the model (in addition to his position, each individual is also characterized by its size). We prefer the study of the simple model described above in this framework.

4.2. Dispersal kernel and competition model

In our model, birth occurs at a constant fertility rate \( \lambda^b \). The birth model can be understood as all individual alive in the forest \( \nu \) can give birth. According to this principle, an individual in state \( x \) will give birth to a new individual in state \( z = x + x' \) with \( x' \) will be determined by a given dispersal kernel. In this paper we suppose that the dispersal kernel is homogeneous for all the individuals: \( D(x, x') = D(x') \) and we consider for \( D(x') \) a Gaussian law such as \( D(x, x') = D(x') = \mathcal{N}(0, \sigma^2 I) \) where \( \sigma^2 < \infty \).

For competition model we assume that \( u(x, y) \) is the strength of competition exercised on the individual in the state \( x \) from an individual in the state \( y \) in a parcel \( \mathcal{X} \). For the competition rate which depends on the position we consider:

\[ \lambda^c(x, \nu) = \sum_{y \in \nu} u(x, y), \]

where the local interaction kernel \( u(x, y) \) is of the form:

\[ u(x, y) = \begin{cases} 
C_{\text{max}} \left( 1 - \frac{1}{r_{\text{max}}} \| x - y \| \right)^+ & \text{if } x \neq y, \\
0 & \text{else},
\end{cases} \tag{19} \]

where \( (\cdot)^+ = \max(\cdot, 0) \) is the positive part, \( \| \cdot \| \) is the Euclidean norm, \( C_{\text{max}} \) is the maximal competition introduced in \( [5] \) and \( r_{\text{max}} \) is the radius of the zone of influence. This zone of influence symbolizes the portion of the ground that the individual needs to ensure its influence on their neighbors. Hence, the interaction kernel \( [19] \) means that: the more the individual in state \( x \) is surrounded by neighbors (such as \( \| x - y \| < r_{\text{max}} \)), the higher the strength of competition is, and the greater the associated death rate is and the more it is subject of disappears (see Fig. [1]).
The radius $r_{\text{max}}$ of the zone of influence associated with an individual in state $x$ in $\nu$ represents the amount of resources (e.g. nutrients, sunlight, water,...) needed for the survival of the tree and $\lambda^c(x, \nu)$ is the computation of the strength of competition experienced by the tree in the state $x$ from all alive trees in the population $\nu$.

5. Monte Carlo simulation

We describe in this section the forest dynamics through Monte Carlo simulation. Individuals in the population $\nu$ are independent of each other. Each individual in the state $x$ has three independent exponential clocks that control the occurrence of the events. A more efficient Monte Carlo procedure consist on the existence of a "global clock" that models all punctual mechanisms (birth, competition death and natural death). The global clock can be computed easily thanks to the properties of the exponential distribution. Then, given an individual chosen randomly in the population, the event to be considered is determined by sampling technique, more precisely it mean that the chosen punctual phenomenon is actually applied or not thanks to an acceptance/rejection sampling technique. Let the time of the last event $T_{k-1}$ and the corresponding forest $\nu_{T_{k-1}}$, we simulate $t_k$ and $\nu_{t_k}$ as follows: we calculate the "global clock" events given by

$$m_k = m_k^b + m_k^d + m_k^c$$
with
\[
    m_k^b = \lambda_b \langle \nu_{T_k-1}, 1 \rangle, \quad m_k^d = \lambda_d \langle \nu_{T_k-1}, 1 \rangle, \quad m_k^c = C_{\text{max}} \langle \nu_{T_k-1}, 1 \rangle^2.
\]

We set \( T_k = T_{k-1} + S_k \) with \( S_k \sim \text{Exp}(m_k) \), and \( \nu_t = \nu_{T_k-1} \) for \( t \in [T_{k-1}, T_k] \).

We calculate the probabilities:
\[
    \alpha_k^b = m_k^b / m_k, \quad \alpha_k^d = m_k^d / m_k, \quad \alpha_k^c = m_k^c / m_k.
\]

We draw the event as follows:

- with probability \( \alpha_k^b \): a birth event. We draw the individual \( x^i_{T_k-1} \) in \( \{1, \ldots, \langle \nu_{T_k-1}, 1 \rangle \} \). We draw \( x' \in \mathbb{R}^2 \) with the dispersal kernel \( D(x') = \mathcal{N}(0, \sigma^2 I) \). We adds a new individual to the position \( z = x^i_{T_k-1} + x' \) and we set:
  \[
  \nu_{T_k} = \nu_{T_k-1} + \delta_z.
  \]

- with probability \( \alpha_k^d \): a natural death event. We draw the individual \( x^i_{T_k-1} \) in \( \{1, \ldots, \langle \nu_{T_k-1}, 1 \rangle \} \) and we set:
  \[
  \nu_{T_k} = \nu_{T_k-1} - \delta_{\{x^i_{T_k-1}\}}.
  \]

- with probability \( \alpha_k^c \): a death due to competition event. We draw two individuals \( x^i_{T_k-1} \) and \( x^j_{T_k-1} \) in \( \{1, \ldots, \langle \nu_{T_k-1}, 1 \rangle \} \). We rejects the event of death due to competition with probability
  \[
  1 - \left( u(x^i_{T_k-1}, x^j_{T_k-1}) / C_{\text{max}} \right),
  \]
  otherwise the individual \( i \) disappears and we set:
  \[
  \nu_{T_k} = \nu_{T_k-1} - \delta_{\{x^i_{T_k-1}\}}.
  \]

The simulation procedure is detailed in Algorithm 1. In the sequel, we present numerical tests using the Monte Carlo simulation described below.

### 5.1. Numerical tests

Our aim is to understand the behaviour of the individual-based model for forest dynamics through four numerical tests using the simulation procedure proposed in Algorithm 1. Especially we are interested in understanding the influence of spatial interactions (competition, dispersion, ...) on the formation of patterns and other aspects of the forest dynamics. To better understand the behaviour of the complex process, model simplification was considered as a tool in this framework. This strategy permits the identification of variables and parameters that play a significant role in the dynamics of the process. Moreover the simplified model has led to mathematical analysis, together with sensitivity analysis. According to the model proposed, we can identifies the following
parameters $\lambda^b, \lambda^d, C_{\text{max}}, r_{\text{max}}$ and $\sigma$ that play a significant role in the numerical simulation. For example, according to the values taken by the two parameters $C_{\text{max}}$ (competition parameter) and $\sigma^2$ (variance of the dispersal kernel) we can test the effect of the competition and dispersal level on the dynamics. The forest of high dispersion (Fig. 2), of low dispersion (Fig. 3), of high competition (Fig. 4) and of low competition (Fig. 5) have been simulated using numerical values presented in the following table:

<table>
<thead>
<tr>
<th>Parameters:</th>
<th>$\lambda^b$</th>
<th>$\lambda^d$</th>
<th>$C_{\text{max}}$</th>
<th>$r_{\text{max}}$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>High dispersion:</td>
<td>2</td>
<td>1</td>
<td>$2.5 \times 10^{-3}$</td>
<td>100</td>
<td>$\sqrt{5}$</td>
</tr>
<tr>
<td>Low dispersion:</td>
<td>2</td>
<td>1</td>
<td>$2.5 \times 10^{-3}$</td>
<td>100</td>
<td>$\sqrt{0.01}$</td>
</tr>
<tr>
<td>High competition:</td>
<td>2</td>
<td>1</td>
<td>$3 \times 10^{-3}$</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>Low competition:</td>
<td>2</td>
<td>1</td>
<td>$3 \times 10^{-4}$</td>
<td>100</td>
<td>1</td>
</tr>
</tbody>
</table>

The high dispersion test and the low dispersion test showed that the dynamics of forest simulated by Algorithm 1 is dependent on the dispersion of new trees. The analysis of figure 2(b) shows that with high variance ($\sigma^2 = 5$) new trees tend to occupy all the space and their dispersion is uniformly in the vicinity of trees alive in the population $\nu$. We can conclude that after their births, the greater the standard deviation of Gaussian kernel is, the greater the dispersion of new trees is.

The analysis of figure 3(b) shows that with low variance ($\sigma^2 = 0.01$) new trees tend to occupy only the space located in the vicinity of parent trees. The impact of the low dispersion level is strongest. Two approximations may explain the behaviour of the forest in the figure 3(b). The
first approximation relates to the location of newly recruited trees. In our model, places with a low local stocking of trees have a smaller ability to give birth which draws the spatial pattern towards randomness. This is no longer true with mechanism of birth, as new births trees are located at random: highly stocked places thus recruit as much as places with low basal areas. This favours the formation of clusters. Second, the number of trees in the population at the stationary state (a burn-in period of 10 time steps was thrown away) has significantly decreased when we reduce the dispersion. The empirical size of trees number at stationary state can be computed as \( \bar{\nu} = \frac{1}{N_{iter} - \tau_b + 1} \sum_{t=\tau_b}^{N_{iter}} \nu_t, 1 \), where \( \tau_b \) is the time of burn-in period and \( N_{iter} \) is the total number of iterations. We note that no other analytical method is available to compute \( \bar{\nu} \). In case the variance equal to 5, \( \bar{\nu} = 498.56 \) and when the variance equal to 0.01, \( \bar{\nu} = 208.41 \). Indeed, the size of population was computed at each time for the count of trees. In fact, as the spatial pattern is clustered (in case \( \sigma^2 = 0.01 \)), the count of trees is over-dispersed with a greater occurrence of competition death events expected in simulations of events. This mortality due to competition explains the small number of trees. So when \( \sigma \) is small, the forest dynamics evolves through a process of clustering.

The analysis of figure 4(b) shows that with high competition (\( C_{max} = 0.003 \)) trees tend to occupy uniformly all the forest space. We note that the distance between trees is large which is due to the effect exerted by the most dominant trees in their area of influence. We also note that the number of trees is low (\( \bar{\nu} = 113.36 \)) reflecting an intense processes of death mostly due to competition. So, if competition is strong, we get a phenomenon of desertification of the forest, which prevents the growth of the trees density on \( \mathcal{X} \). The figure 5(b) shows that with low competition (\( C_{max} = 0.0003 \)) trees tend to invade the most forest area. The distance between trees is low and the number of trees is very high (\( \bar{\nu} = 5469.98 \)) because the death process due to competition is negligible. So, if competition is low, while the forest invaded the area of \( \mathcal{X} \) which leads to strong growth of trees density.

5.2. Parameters estimation using stochastic algorithm

In this study, model simplification was be treated in the context of understanding the complexity of spatial pattern for forest dynamics with the most simple model. As we have seen in the numerical tests, although the strategy of simplification, several parameters in the proposed model have a significant effect on the dynamic simulated. Thus, the model proposed in this framework contains parameters \( \theta = (\lambda^b, \lambda^d, C_{max}, r_{max}, \sigma^2)' \) to estimate. To do this, we set a target characteristic such as population size \( N^* \). The problem is whether it is possible to estimate the parameters \( \theta \) given the target characteristic \( N^* \). The aim is obviously to do that automatically. Formally, we aim to
draw $\nu_0$ and set $T_0 = 0$

for $k = 1, 2, \ldots, N_{\text{iter}}$ do

$N \leftarrow (\nu_{T_{k-1}}, 1)$

% setting clocks

$m^b \leftarrow \lambda^b N$

$m^d \leftarrow \lambda^d N$

$m^c \leftarrow C_{\text{max}} N^2$

$m \leftarrow m^b + m^d + m^c$

% new instant event

$S \sim \text{Exp}(m)$

$T_k \leftarrow T_{k-1} + S$

$u \sim U[0, 1]$

$i \sim U\{1, \ldots, N\}$

if $u \leq m^b/m$ then

$x' \sim N(0, \sigma^2 I)$

$\nu_{T_k} \leftarrow \nu_{T_{k-1}} + \delta_{\{x_{T_{k-1}} + x'\}}$ \% birth

else if $u \leq (m^b + m^d)/m$ then

$\nu_{T_k} \leftarrow \nu_{T_{k-1}} - \delta_{\{x_{T_{k-1}}\}}$ \% natural death

else

$j \sim U\{1, \ldots, N\}$

$\beta \sim U[0, 1]$

if $\beta \leq u(x_{T_{k-1}}^i, x_{T_{k-1}}^j)/C_{\text{max}}$ then

$\nu_{T_k} \leftarrow \nu_{T_{k-1}} - \delta_{\{x_{T_{k-1}}\}}$ \% death by competition

end if

end if

end if

end for

Alg. 1: Monte Carlo simulation algorithm of the process $\nu_t$. 
Fig. 3: Low dispersion test: dynamics of individuals number in function of time and spatial pattern of forest.

Fig. 4: High competition test: dynamics of individuals number in function of time and spatial pattern of forest.
Fig. 5: Low competition test: dynamics of individuals number in function of time and spatial pattern of forest.

minimize, for $k$ large, the criterion:

$$J(\theta) = \mathbb{E}_\theta((\nu_{T_k}, 1) - N^*)^2,$$

where $\mathbb{E}_\theta$ means that the process is generated by the value of $\theta$. To be more rigorous, it must be understood in a ergodic sense that, for $k$ large, the rigorous criterion is $\lim_{T \to \infty} \frac{1}{T} \int_0^T ((\nu_{T_k}, 1) - N^*)^2 dT_k$. Hence, we are looking for the minimum $\theta^* \in \mathbb{R}^5_+ \setminus \{\infty\}$ such that

$$J(\theta^*) = \inf_{\theta \in \mathbb{R}^5_+ \setminus \{\infty\}} J(\theta), \quad \theta \in \mathbb{R}^5_+ \setminus \{\infty\}. \quad (21)$$

We precise that no explicit solution is available for the problem (21) and we propose to solve this by using a stochastic algorithm. For an account on stochastic algorithm, we refer the reader to [16], [9] and [1]. As the gradient of the function $J$ is not explicitly computable, we use a Kifer-Wolfowitz approach by approximating the gradient recursively. We construct a stochastic algorithm to estimate $\theta^*$. Under a moment condition (order $> 2$) we prove the almost sure convergence of the estimator.

5.2.1. Almost sure convergence of the estimator

Let us first prove that the criterion (20) is differentiable and convex. The fact that the function $J$ is convex is crucial to ensure the convergence of the stochastic algorithm (as Kiefer-Wolfowitz or Robbins-Monro). We now introduce some notation that is useful, let $\nabla_\theta$ be the vector of first derivatives with respect to $\theta$.

**Proposition 5.1.** Under the hypothesis $\mathbb{H}$ and assuming that $\mathbb{E}[N_0] < \infty$ and $u(x, x)$ is bounded
below, then the criterion (20) is convex on the convex set \( \mathcal{D}_0 = \{(\lambda^b, \lambda^d, C_{\text{max}}, r_{\text{max}}, \sigma^2) \in \mathbb{R}_+^5 \setminus \{\infty\}\} \).

**Proof.** In order to prove the convexity of the criterion (20), it is sufficient to prove that

\[
(\theta - \theta')^t \nabla_{\theta} J(\theta') \leq J(\theta) - J(\theta').
\]

(22)

We have:

\[
(\theta - \theta')^t \nabla_{\theta} J(\theta') = \sum_{\kappa=1}^{5} (\theta_{\kappa} - \theta'_{\kappa}) \nabla_{\theta_{\kappa}} \mathbb{E}_{\theta_{\kappa}} \left[ (\nu_t, 1)^2 - 2N^* \langle \nu_t, 1 \rangle \right].
\]

Or, by assuming without loss of generality that \( \psi(\nu) = \langle \nu, 1 \rangle^2 \) (by setting \( f = 1 \) and \( F(a) = a^2 \)), we have

\[
\mathbb{E}[\langle \nu_t, 1 \rangle^2] - \mathbb{E}[\langle \nu_0, 1 \rangle^2] + \int_0^t \mathbb{E} \left[ \int_X \lambda^b \nu_\tau(dx) \times \int_S \left\{ \psi(\nu_\tau + \{\delta_{(x+x')}\}) - \psi(\nu_\tau) \right\} D(x, x') d\tau \right.

\[
+ \left. \int_0^t \mathbb{E} \left[ \int_X \left\{ \psi(\nu_\tau) - \delta(x) \right\} \cdot \psi(\nu_\tau) \right] \right) \nu_\tau(dx) d\tau \right)

\[
= \mathbb{E}[\langle \nu_0, 1 \rangle^2] + \int_0^t \mathbb{E} \left[ \int_X \lambda^b \nu_\tau(dx) \{2\langle \nu_\tau, 1 \rangle + 1\} \right] d\tau

\[
+ \int_0^t \mathbb{E} \left[ \int_X \left\{ -2\langle \nu_\tau, 1 \rangle + 1 \right\} \{\lambda^d + \int_X u(x, y) \nu_\tau(dy) \} \nu_\tau(dx) \right] d\tau

\[
= \mathbb{E}[\langle \nu_0, 1 \rangle^2] + \int_0^t \mathbb{E} \left[ 2(\lambda^b - \lambda^d) \langle \nu_\tau, \langle \nu_\tau, 1 \rangle \rangle + (\lambda^b + \lambda^d) \langle \nu_\tau, 1 \rangle \right]

\[
+ (1 - 2\langle \nu_\tau, 1 \rangle) \langle \nu_\tau \otimes \nu_\tau, u \rangle \right] d\tau.
\]
Then, by assuming without loss of generality that $\lambda^b = \lambda^d = \lambda$ and $\langle \nu_t \otimes \nu_t, u \rangle = \langle \nu_t, 1 \rangle \phi(\langle \nu_t, 1 \rangle)$ from Lemma 3.4, we have

$$(\theta - \theta')^t \nabla_{\theta'} J(\theta') = \sum_{\kappa=1}^{5} (\theta_{\kappa} - \theta'_{\kappa}) \nabla_{\theta'_{\kappa}} \left( \int_{0}^{t} E_{\theta_{\kappa}} \left[ 2\lambda \langle \nu_{\kappa}', 1 \rangle + (1 - 2\langle \nu_{\kappa}', 1 \rangle) \langle \nu_{\kappa}', 1 \rangle \phi(\langle \nu_{\kappa}', 1 \rangle) \right] d\tau \right)$$

$$+ 2N^* \int_{0}^{t} E_{\theta_{\kappa}} \left[ \langle \nu_{\kappa}', 1 \rangle \phi(\langle \nu_{\kappa}', 1 \rangle) \right] d\tau$$

$$= \sum_{\kappa=1}^{5} (\theta_{\kappa} - \theta'_{\kappa}) \nabla_{\theta'_{\kappa}} \left( \int_{0}^{t} E_{\theta_{\kappa}} \left[ 2\lambda \langle \nu_{\kappa}', 1 \rangle + (1 - 2\langle \nu_{\kappa}', 1 \rangle + 2N^*) \langle \nu_{\kappa}', 1 \rangle \phi(\langle \nu_{\kappa}', 1 \rangle) \right] d\tau \right)$$

$$\leq \int_{0}^{t} \left( 2\lambda E_{\theta'} \langle \nu_{\tau}, 1 \rangle + (1 - 2E_{\theta'} \langle \nu_{\tau}, 1 \rangle + 2N^*) E_{\theta'} \langle \nu_{\tau}, 1 \rangle \phi(\langle E_{\theta'} \langle \nu_{\tau}, 1 \rangle) \right) d\tau$$

$$- 2\lambda E_{\theta'} \langle \nu_{\tau}', 1 \rangle - (1 - 2E_{\theta'} \langle \nu_{\tau}', 1 \rangle + 2N^*) E_{\theta'} \langle \nu_{\tau}', 1 \rangle \phi(\langle E_{\theta'} \langle \nu_{\tau}', 1 \rangle) \right) d\tau$$

$$\leq J(\theta) - J(\theta')$$

where (23) is obtained by using the differentiation under the integral sign method and inequality (24) is obtained from the Jensen inequality (as $\phi$ is convex) and the convexity of the function $h$ (that satisfies the gradient inequality) given by

$$h(\theta') = \int_{0}^{t} \left( 2\lambda E_{\theta'} \langle \nu_{\tau}', 1 \rangle + (1 - 2E_{\theta'} \langle \nu_{\tau}', 1 \rangle + 2N^*) E_{\theta'} \langle \nu_{\tau}', 1 \rangle \phi(\langle E_{\theta'} \langle \nu_{\tau}', 1 \rangle) \right) d\tau,$$

where $E_{\theta'} \langle \nu_{\tau}', 1 \rangle \leq E_{\theta'} \langle \nu_{\tau}', 1 \rangle \forall x_0$ and $x_0$ is the greatest solution of $|\Delta x_0 = x_0 \phi(x_0)$. Thus, we conclude that the gradient inequality (22) is satisfied which concludes the proof. □

We now study the convergence of our algorithm which is a way to solve $\nabla J(\theta) = 0$. We shall prove its convergence under general assumptions that are satisfied in particular for our specific criterion $J$. We consider that $\mathbb{R}^5$ is endowed with the $L^1$ norm $\| \cdot \|_{L^1}$ and the dual space $(\mathbb{R}^5)^*$ is endowed with the dual norm $\| \cdot \|_{*} = \sup_{v \in S} \| v \|_{*}$ and $\| v \|_{*} = \sup_{v \in S \subseteq [0,1]} | v |_{*}$. Let $S$ be a compact convex subset of $\mathbb{R}^5$. Recall that a convex function $\omega : S \mapsto \mathbb{R}$ is strongly convex with parameter $\alpha > 0$ with respect to the norm $\| \cdot \|_{L^1}$ if for any $\mu \in [0,1]$ and any $v, v'$ belonging to $S$,

$$\omega(\mu v + (1 - \mu)v') \leq \mu \omega(v) + (1 - \mu)\omega(v') - \frac{\alpha}{2} \mu(1 - \mu)\| v - v' \|^2_{L^1}.$$

The stochastic algorithm uses an $\alpha$-convex function $\omega$ differentiable on a point $v_0 \in S$ to define the auxiliary $\alpha$-convex function $\omega$:

$$\omega(v) = \omega(v) - \omega(v_0) - (\nabla \omega(v_0), v - v_0).$$

Setting a positive $\beta$, $V_\beta$ denote the Legendre-Fenchel transform of $\beta \omega$ with

$$V_\beta(v) = \sup_{v' \in S} \left( \langle v, v' \rangle - \beta \omega(v') \right).$$
We know that the function \( V_\beta \) is continuously differentiable and that its gradient takes its values in \( S \) thanks to the properties of Legendre-Fenchel transform of \( \alpha \)-convex functions. We precise also that the gradient of \( V_\beta \) has a Lipschitz constant equal to \((\alpha \beta)^{-1}\). We consider a function \( g \) which is an approximation of the gradient of \( J \) with the following decomposition \( g = \nabla J + X + w \), where \( X \) is a martingale difference and \( w \) is negligible with respect to the gradient of \( J \) and \( X \). The stochastic algorithm uses the two positive sequences \( a = (a_k)_{k=1}^K \) and \( \beta = (\beta_k)_{k=1}^K \) and a sequence of iid random variable time \( T \).

Let us denote a realization of \( T \) by \( t \) and in the following we consider \( K \) independent copies \( t_1, \ldots, t_K \) of the random variable time \( T \). We first notice that \( J(\lambda^b, \lambda^d, C_{\max}, r_{\max}, \sigma^2) \) is equal to \( \mathbb{E}[J(\lambda^b, \lambda^d, C_{\max}, r_{\max}, \sigma^2,t)] \) where

\[
J(\lambda^b, \lambda^d, C_{\max}, r_{\max}, \sigma^2,t) = \mathbb{E}_\theta(\nu_{t}N) - N^*.
\]

For reasons of simplicity, we introduce the following notations:

\[
J^m_{\pm}(c_k, T) = J(\theta_{1,k-1}, \ldots, \theta_{m-1,k-1}, \theta_{m,k-1} \pm c_k, \theta_{m+1,k-1}, \ldots, \theta_{5,k-1}, T)
\]

\[
J^m_{\pm}(c_k) = J(\theta_{1,k-1}, \ldots, \theta_{m-1,k-1}, \theta_{m,k-1} \pm c_k, \theta_{m+1,k-1}, \ldots, \theta_{5,k-1}),
\]

where \( \theta = \nabla V_\beta(\theta) \). We consider \( d_{c_k}J \) (resp. \( d_{c_k}J \)) the random vector whose \( m \)-th coordinate \( d_{c_k}J(\lambda^b, \lambda^d, C_{\max}, r_{\max}, \sigma^2, T) \) (resp. \( d_{c_k}J(\lambda^b, \lambda^d, C_{\max}, r_{\max}, \sigma^2) \)) is defined by

\[
\frac{J^m_{+}(c_k, T) - J^m_{-}(c_k, T)}{2c_k} \quad \text{(resp.} \quad \frac{J^m_{+}(c_k) - J^m_{-}(c_k)}{2c_k} \text{).}
\]

Then, we put \( g_{c_k}(\theta_{k-1}, t_k) = d_{c_k}J(\theta_{k-1}, t_k) \) and we define

\[
X_{c_k}(\theta_{k-1}, t_k) := d_{c_k}J(\theta_{k-1}, t_k) - d_{c_k}J(\theta_{k-1}),
\]

and

\[
w_{c_k}(\theta_{k-1}) := d_{c_k}J(\theta_{k-1}) - \nabla J(\theta_{k-1}).
\]

Thus, it enhances the decomposition

\[
g_{c_k}(\theta_{k-1}, t_k) = \nabla J(\theta_{k-1}) + X_{c_k}(\theta_{k-1}, t_k) + w_{c_k}(\theta_{k-1}). \tag{25}
\]

The idea of stochastic algorithm consists in a rapid movement on the surface of the function \( J \) to minimize and avoid the intense attraction of local minima. Given a parameters \( a_k \) and \( c_k \) and \( \beta_k \), the stochastic algorithm modifies these parameters at each iteration. Then, it is of the form

Now, under assumptions on the function \( g \), we shall prove that \( \hat{\theta} \) (see output of Alg. 2) goes almost surely to the unique minimum \( \theta^* \) of the function \( J \).
choose $\theta_0 = 0 \in (\mathbb{R}^5)^*$ and $\theta_0 \in S$

for $k = 1, 2, \ldots, K$ do

$\theta_k \leftarrow \theta_{k-1} - a_k g_{c_k} (\theta_{k-1}, t_k)$

$\theta_k \leftarrow \nabla V_{\beta_k} (\theta_k)$

end for

% output:

$\hat{\theta} \leftarrow \sum_{k=1}^{K} a_k \theta_{k-1} \sum_{k=1}^{K} a_k$

 Alg. 2: Kiefer-Wolfowitz version of the stochastic algorithm.

Let us consider $S$ is compact convex subset of $\mathbb{R}^5$ and we assume that $J$ is a convex function as the hypothesis ($\mathbb{H}^{(1)}$) is:

$\mathbb{H}^{(1)}$: (i) $J : \mathbb{R}^5 \mapsto \mathbb{R}$ is a convex function,
(ii) The function $J$ is of $C^2$ class,
(iii) The function $J$ admits a unique minimum $\theta^*$ on $S$.

Moreover, we have for any $\zeta \in \mathbb{R}^5$ and any $c > 0$:

$$E[X_c(\zeta, T)] = 0 \quad \text{and} \quad w_c(\zeta) \xrightarrow{c \to 0} 0.$$ 

In the algorithm 2, $\theta_0$ being any random vector taking values in $S$, and $\mathcal{F}_k^{(2)}$ being the $\sigma$-field generated by $\theta_0, t_1, \ldots, t_k$. It is clear that $\theta_k$ is $\mathcal{F}_k^{(2)}$ measurable and $t_k$ is independent of $\mathcal{F}_{k-1}^{(2)}$. We have therefore

$$E[X_{c_k}(\theta_{k-1}, t_k)|\mathcal{F}_{k-1}^{(2)}] = 0.$$ 

In order to prove the $L^1$ convergence, we shall need an additional technical assumptions as the hypothesis ($\mathbb{H}^{(2)}$) is:

$\mathbb{H}^{(2)}$: (i) There exists some non negative real number $\rho$ such that for all vector $\zeta \in \mathbb{R}^5$, the random variable $J(\zeta, T)$ satisfies the moment condition:

$$E[|J(\zeta_1, \ldots, \zeta_5, t_k)|^{2|\mathcal{F}_{k-1}^{(2)}]} \leq \rho^2, \quad \text{for all } k \geq 1.$$ 

(ii) Let $a_k, c_k, \beta_k$, for $k \geq 0$, be sequences in $(\mathbb{R}^+)^*$. The sequence $(\beta_k)_{k \geq 0}$ is non decreasing and
we have

\[ a) \quad \frac{\beta_k K}{\sum_{k=1}^{K} a_k} \xrightarrow{K \to +\infty} 0, \]
\[ b) \quad \sum_{k=1}^{K} a_k c_k \xrightarrow{K \to +\infty} 0, \]
\[ c) \quad \frac{\sum_{k=1}^{K} a_k^2}{\sum_{k=1}^{K} a_k} \xrightarrow{K \to +\infty} 0, \]
\[ d) \quad \sum_{k=1}^{+\infty} \frac{a_k^2}{c_k} < \infty. \]

The assumption \((\mathbb{H}(2))(i)\) allows us to overcome the following bound problem: in fact, we have

\[
\mathbb{E} \left[ \|X_{\alpha k} (\theta_{k-1}, t_k)\|^2 | \mathcal{F}_{k-1}^{(2)} \right] = \sup_{1 \leq m \leq 5} \left( \frac{\mathbb{V}[J^m (c_k, t_k)] | \mathcal{F}_{k-1}^{(2)} | + \mathbb{V}[J^m (c_k, t_k)] | \mathcal{F}_{k-1}^{(2)} |}{4c_k^2} \right)
- \frac{1}{2c_k} \mathbb{E} \left[ (J^m (c_k, t_k) - J^m (c_k)) (J^m (c_k, t_k) - J^m (c_k)) \right],
\]

and consequently thanks to the assumption \((\mathbb{H}(2))(i)\) and using the Cauchy-Schwartz inequality we obtain the following bound

\[
\mathbb{E} \left[ \|X_{\alpha k} (\theta_{k-1}, t_k)\|^2 | \mathcal{F}_{k-1}^{(2)} \right] \leq \frac{\alpha^2}{c_k^2} \quad \text{for any } k = 1, \ldots, K. \tag{26}
\]

This condition (26) is sufficient to prove the \(L^1\)-convergence. In the sequel, to make the proof of the convergence clearer, we introduce an estimation of \(\epsilon = J(\tilde{\theta}) - J(\theta^*)\) in a similar manner as [22].

**Lemma 5.2.** Assume that the sequence \((\beta_k)_{k \geq 0}\) is non decreasing. For a purely notational reason, we put \(A_k = \langle X_{\alpha k} (\theta_{k-1}, t_k), \theta_{k-1} - \theta^* \rangle\) and \(B_k = \langle w_{\alpha k} (\theta_{k-1}), \theta_{k-1} - \theta^* \rangle\) and \(C_k = \|g_{\alpha k} (\theta_{k-1}, t_k)\|^2\).

Then, we have: (i) \(\epsilon \sum_{k=1}^{K} a_k \geq 0\) and

\[
(ii) \quad \epsilon \sum_{k=1}^{K} a_k \leq \beta_k \omega(\theta^*) - \sum_{k=1}^{K} a_k A_k - \sum_{k=1}^{K} a_k B_k + \sum_{k=1}^{K} \frac{a_k^2}{2\alpha \beta_k} C_k.
\]

**Proof.** To prove (i) its clear that, for any \(k = 1, \ldots, K\), \(J(\theta_{k-1}) - J(\theta^*) \geq 0\) because \(J(\theta^*) = \inf_{\theta \in S} J(\theta)\) and consequently \(\epsilon \geq 0\). For the proof of (ii) we use the convexity and the differentiability of the function \(J\). We recall that the the function \(V_{\beta_{k-1}}\) is a Legendre transform of a strictly convex function and for this reason its convex and differentiable. As the sequence \((\beta_k)_{k \geq 0}\) is non decreasing, we have

\[
V_{\beta_k} (\theta_k) - V_{\beta_{k-1}} (\theta_{k-1}) \leq V_{\beta_{k-1}} (\theta_k) - V_{\beta_{k-1}} (\theta_{k-1})
\]
\[
\leq \langle \nabla V_{\beta_{k-1}} (\theta_{k-1}), \theta_k - \theta_{k-1} \rangle + \frac{1}{2\alpha \beta_{k-1}} \|\theta_{k-1} - \theta_k\|^2 \quad \text{(27)}
\]
\[
= -a_k \langle \theta_{k-1}, g_{\alpha k} (\theta_{k-1}, t_k) \rangle + \frac{a_k^2 \|g_{\alpha k} (\theta_{k-1}, t_k)\|^2}{2\alpha \beta_{k-1}}, \quad \text{(28)}
\]

25
where (27) is obtained using the convexity and the differentiability of \( V_{\beta_{k-1}} \) and using the item 2 of Proposition A.3 in [22, p.113] (we recall that the gradient of \( V_\beta \) has a Lipschitz constant equal to \( 1/\alpha \)). Then, by taking the sum, we obtain

\[
V_{\beta_K}(\theta_K) - V_{\beta_0}(\theta_0) \leq -\sum_{k=1}^{K} a_k \langle \theta_{k-1}, g_{c_k}(\theta_{k-1}, t_k) \rangle + \sum_{k=1}^{K} \frac{a_k^2 \| g_{c_k}(\theta_{k-1}, t_k) \|^2}{2\alpha \beta_{k-1}}.
\]

(29)

As \( V_{\beta_0}(\theta_0) = 0 \) and \( \langle \theta^*, \theta_K \rangle - V_{\beta_K}(\theta_K) \leq \beta_K \omega(\theta^*) \), we have from (29):

\[
\sum_{k=1}^{K} a_k \langle \theta_{k-1}, g_{c_k}(\theta_{k-1}, t_k) \rangle \leq -V_{\beta_K}(\theta_K) + \sum_{k=1}^{K} \frac{a_k^2 \| g_{c_k}(\theta_{k-1}, t_k) \|^2}{2\alpha \beta_{k-1}} \\
\leq \beta_K \omega(\theta^*) - \langle \theta^*, \theta_K \rangle + \sum_{k=1}^{K} \frac{a_k^2 \| g_{c_k}(\theta_{k-1}, t_k) \|^2}{2\alpha \beta_{k-1}}.
\]

(30)

Now, as \( J \) is convex and differentiable (Proposition 5.1) and using decomposition (25), we have:

\[
e \sum_{k=1}^{K} a_k \leq \sum_{k=1}^{K} a_k \left( J(\theta_{k-1}) - J(\theta^*) \right) \leq \sum_{k=1}^{K} a_k \langle \nabla J(\theta_{k-1}), \theta_{k-1} - \theta^* \rangle \\
\leq \sum_{k=1}^{K} a_k \langle g_{c_k}(\theta_{k-1}, t_k), \theta_{k-1} - \theta^* \rangle - \sum_{k=1}^{K} a_k \langle X_{c_k}(\theta_{k-1}, t_k), \theta_{k-1} - \theta^* \rangle \\
- \sum_{k=1}^{K} a_k \langle w_{c_k}(\theta_{k-1}, t_k), \theta_{k-1} - \theta^* \rangle \\
= \langle \theta_K, \theta^* \rangle + \sum_{k=1}^{K} a_k \langle g_{c_k}(\theta_{k-1}, t_k), \theta_{k-1} \rangle - \sum_{k=1}^{K} a_k \langle X_{c_k}(\theta_{k-1}, t_k), \theta_{k-1} - \theta^* \rangle \\
- \sum_{k=1}^{K} a_k \langle w_{c_k}(\theta_{k-1}, t_k), \theta_{k-1} - \theta^* \rangle.
\]

(31)

Finally, using the inequality (30) in (31), we obtain

\[
e \sum_{k=1}^{K} a_k \leq \beta_K \omega(\theta^*) - \sum_{k=1}^{K} a_k \langle X_{c_k}(\theta_{k-1}, t_k), \theta_{k-1} - \theta^* \rangle \\
- \sum_{k=1}^{K} a_k \langle w_{c_k}(\theta_{k-1}, t_k), \theta_{k-1} - \theta^* \rangle + \sum_{k=1}^{K} \frac{a_k^2 \| g_{c_k}(\theta_{k-1}, t_k) \|^2}{2\alpha \beta_{k-1}},
\]

(32)

which concludes the proof of the Lemma 5.2.

Thanks to our assumptions, we deduce that \( E[\langle X_{c_k}(\theta_{k-1}, t_k), \theta_{k-1} - \theta^* \rangle] = 0 \) and consequently \( \langle X_{c_k}(\theta_{k-1}, t_k), \theta_{k-1} - \theta^* \rangle \) is a martingale difference. In the following, we give a Lemma that may be deduced from Taylor inequality.

**Lemma 5.3.** For \( \theta_k \in S \), for all \( k \), there exists a constant \( \xi > 0 \) such that, for any \( k \)

\[
\| w_{c_k}(\theta_{k-1}) \|_s \leq \xi c_k.
\]

(33)
Proof. The proof follows from the fact that the Hessian matrix of $J$ is bounded on $S$. \qed

Theorem 5.4. Under the assumptions $H^{(1)}$ and $H^{(2)} (i)$ and $H^{(2)} (ii)$ a and $H^{(2)} (ii)$ b and $H^{(2)} (ii)$ c we have the $L^1$-convergence of the estimator $\hat{\theta}$ defined in Alg. 2

\( \hat{\theta} - \theta^* \overset{L^1}{\longrightarrow} 0. \)

Proof. Using the inequality (32) and (33) proved in Lemma 5.2 and Lemma 5.3 and after taking expectation we obtain

\[
0 \leq \mathbb{E}[\epsilon] \leq \beta K \omega(\theta^*) + \frac{2 \xi d \sum_{k=1}^{K} a_k c_k}{\sum_{k=1}^{K} a_k} + \frac{\sum_{k=1}^{K} b a_k^2 c_k}{\sum_{k=1}^{K} a_k},
\]

where $d$ is the diameter of $S$ and $b/c_k^2$ denotes a bound of $\mathbb{E}[\|g_{c_k}(\theta_k-1, t_k)\|_2^2]$ and we recall that $\mathbb{E}[\|X_{c_k}(\theta_{k-1}, t_k, \theta_k-1 - \theta^*)\|_2] = 0$. There may be some doubt as to why $\mathbb{E}[\|g_{c_k}\|_2^2]$ is bounded. From the decomposition (25), it would appear that $\|g_{c_k}\|_2^2 \leq \|\nabla J\|_*^2 + \|X_{c_k}\|_*^2 + \|w_{c_k}\|_*^2 + 2\left(\|\nabla J\|_* + \|w_{c_k}\|_*\right)\|X_{c_k}\|_* + 2\|\nabla J\|_* \|w_{c_k}\|_*,$

where $\|\nabla J\|_*$ is bounded as $S$ is compact and $\nabla J$ is continuous and $\|w_{c_k}\|_* \leq \xi c_k$ from Lemma 5.3 and $\mathbb{E}[\|X_{c_k}\|_*^2] \leq \frac{\beta^2}{c_k^2}$ (with $c_k \to 0$ as $k \to \infty$) from (26). Thus, thanks to the inequality (34), we conclude that using the sufficient conditions $H''(ii)$ a and $H''(ii)$ b and $H''(ii)$ c we have the $L^1$-convergence of the estimator $\hat{\theta}$ to the unique minimum $\theta^*$ of the function $J$. \qed

To make proofs for the almost surely convergence clearer, some definitions that we will use in the sequel are as follows: we use $X_K$ to denote a square integrable martingale written as

\[ X_K := \sum_{k=1}^{K} x_k, \]

where $x$ is a martingale difference. For $\mathcal{F}_k^{(3)}$ being the $\sigma$-field generated by $x_1, \ldots, x_k$, we defines

\[ (X_K) := \sum_{k=1}^{K} \mathbb{E}\left[ x_k^2 | \mathcal{F}_k^{(3)} \right]. \]

To prove the almost surely convergence, we use the Law of large numbers for martingales and the Chow Lemma (see [20] and [9, p.22] respectively for instance). To make proofs clearer, we recall the following Theorem 5.5 and Theorem 5.6 that we will use in the sequel.
Theorem 5.5 (Law of large numbers for martingales). Let $X_K$ denotes a square integrable martingale if
\[
\langle X \rangle_\infty = \lim_{K \to \infty} \langle X_K \rangle < \infty \quad \text{a.s.,}
\]
then $(X_K)_{K \in \mathbb{N}}$ converges a.s. to $X_\infty$ square integrable.

Theorem 5.6 (Chow Lemma). Suppose $(e_K)_{K \in \mathbb{N}}$ is a bounded sequence of positive numbers, suppose that $1 < p \leq 2$. For $K \in \mathbb{N}$, let $E_K = 1 + \sum_{k=0}^{K} e_k$ and $E_\infty = \lim_{K \to \infty} E_K$. Suppose that $(Y_K)_{K \in \mathbb{N}}$ is a positive sequence of random variables adapted to $(\mathcal{F}_K^{(3)})_{K \in \mathbb{N}}$ and $C$ is a constant such that
\[
\mathbb{E}[Y_{K+1}|\mathcal{F}_K^{(3)}] \leq C \quad \text{and} \quad \sup_K \mathbb{E}[Y_{K+1}^p|\mathcal{F}_K^{(3)}] < \infty,
\]
then we have the following properties almost surely:
\[
\sum_{k=1}^{\infty} e_k Y_{k+1} \text{ converges on } \{E_\infty < \infty\}
\]
\[
E_K^{−1} \sum_{k=1}^{\infty} e_k Y_{k+1} \leq C \quad \text{on } \{E_\infty = \infty\}.
\]

In order to give our main result, we shall need a higher order moment assumption as the hypothesis $(\mathbb{H}^{(3)})$ is:

\(\mathbb{H}^{(3)}\): We assume that there exists $p > 2$ such that almost surely:
\[
\sup_{k > 0} \mathbb{E} \left[\left|\mathcal{J}(\zeta_1, \ldots, \zeta_5, t_k)\right|^p |\mathcal{F}_k^{(1)}\right] < \infty.
\]

We precise that assumption $\mathbb{H}^{(3)}$ yields the existence of a real $p > 2$ such that
\[
\sup_{k > 0} \mathbb{E} \left[\left\|c_k X_{c_k}(\theta_{k-1}, t_k)\right\|^p |\mathcal{F}_{k-1}^{(1)}\right] < \infty, \quad \text{a.s.}
\]

Theorem 5.7. Under assumption $\mathbb{H}^{(1)}$ and $\mathbb{H}^{(2)}$ and $\mathbb{H}^{(3)}$, we have almost surely the convergence of the estimator $\hat{\theta}$
\[
(\hat{\theta} - \theta^*) \xrightarrow{a.s.} 0.
\]
Proof. To prove (35), we put the martingale $X_K$ equal to

$$X_K = \sum_{k=1}^{K} a_k (X_{c_k} (\theta_{k-1}, t_k), \theta_{k-1} - \theta^*),$$

and we apply the law of large numbers for martingales. Under the hypothesis $\mathbb{H}^{(2)} (ii)$, $\langle X_K \rangle$ is bounded and converges as it is increasing. Consequently $X_K$ a.s. converges. Thanks to (ii) in Lemma 5.2, we can write

$$0 \leq \epsilon \leq \frac{\beta_K \omega(\theta^*)}{\sum_{k=1}^{K} a_k} + \frac{2 \xi d \sum_{k=1}^{K} a_k c_k}{\sum_{k=1}^{K} a_k} + \frac{\sum_{k=1}^{K} \frac{a_k^2}{2a_k \beta_{k-1}} \|g_{c_k}(\theta_{k-1}, t_k)\|^2}{\sum_{k=1}^{K} a_k}$$

(36)

where in the inequality (36) it is easy to see that the three terms goes a.s. to zero as $\sum_{k=1}^{K} a_k \rightarrow +\infty$ (by using the convergence of $X_K$ and under assumption $\mathbb{H}^{(2)} (i)$ and $\mathbb{H}^{(2)} (ii)$). Furthermore, by using the decomposition (25) and by applying the Chow Lemma twice to $Y_K = c_K^2 \|X_{c_k}\|^2$ and $e_K = a_k^2 / (\beta_{K-1} c^2_K)$ (resp. $Y_K = c_K \|X_{c_k}\|$ and $e_K = a_k^2 / (\beta_{K-1} c_K)$) it is straightforward to prove that the term in (37) converges to zero.

Thus, we have proved that $(\hat{\theta} - \theta^*) \xrightarrow{a.s.} 0$. It is worth noting that one gets $\hat{\theta} \xrightarrow{a.s.} \theta^*$ as $\hat{\theta}$ belongs to the compact set $S$. If $\theta^*$ is an accumulation point then as $(\hat{\theta} - \theta^*) \rightarrow 0$, $\theta^*$ is also a minimum of the function $J$. So that, we have $\theta^* = \theta^*$ thanks to the uniqueness of the minimum which concludes the proof of Theorem 5.7.

Clearly, if some higher order moment assumption holds we could establish an estimation for the rate of the convergence thanks to Markov inequality and the Chow Lemma.

5.2.2. Computation by use of stochastic algorithm

We use now the stochastic gradient algorithm, given by Alg 2, in Kiefer-Wolfowitz version to estimate the numerical solution of problem (21). In order to run our algorithm, we propose an auxiliary function $\omega^c$, whose Legendre-Fenchel transform is computable. In our computation, we used the following computable entropy function

$$\omega_q^c (\theta) = \sum_{l=1}^{5} \frac{\theta_l}{q} \ln \left( \frac{\theta_l}{q} \right),$$

which is a natural strongly convex function (it is $\frac{1}{q}$-convex and admits a unique minimum $\theta_0 = (\frac{q}{5}, \ldots, \frac{q}{5})$). From this choice of $\omega_q^c (\theta)$, we used the function

$$\omega^c (\theta) = \omega_q^c (\theta) - \omega_q^c (\theta_0) - \langle \nabla \omega_q^c (\theta_0), \theta - \theta_0 \rangle$$

$$= \ln(5) + \sum_{l=1}^{5} \frac{\theta_l}{q} \ln \left( \frac{\theta_l}{q} \right),$$

29
which is also \( \frac{1}{q} \)-convex and its Legendre-Fenchel transform is

\[
V_{\beta}^c(\theta) = \beta \ln \left( \frac{1}{5} \sum_{i=1}^{5} \exp \{ \theta_{i} \} \right).
\]

We note that the choice of the above computable functions is inspired from [22]. Also note that the Legendre-Fenchel transform \( V_{\beta}^c \) involves an exponential part which could involve numerical difficulties when \( q \) takes large values. Numerical problems can be avoided by using the following choice \( q = \beta \) in computations. It is straightforward that

\[
\nabla V_{\beta}^c(\theta) = \beta \exp \{ \theta \} \left( \sum_{l=1}^{5} \exp \{ \theta_l \} \right)^{-1}.
\]

As we consider the classical Kiefer-Wolfowitz version of the stochastic algorithm, the assumptions on the gains are particular cases of our assumption with a constant sequence \( \beta_k = 1 \) and

\[
\sum_k a_k = \infty, \quad \text{and} \quad \sum_k a_k c_k < \infty \quad \text{and} \quad \sum_k a_k^2 / c_k^2 < \infty.
\]

Below, the algorithm has been performed with the following sequences \( a_k = a \times (k + 1)^{-(\frac{3}{4} + \frac{1}{10})} \) and \( c_k = c \times (k + 1)^{-1/4} \) with \( a \) and \( c \) are two positive constants [14].

The Kiefer-Wolfowitz algorithm considered is performed as follows:

choose \( \theta^0 = (\lambda^0, \lambda_d^0, C_{0\max}^0, r_{0\max}^0, \sigma_0^2)' \) and \( \theta^0 \)

for \( k = 1, 2, \ldots \) do

% simulation of forests (see. Alg. 1):

For \( l = 1, \ldots, 5 \) and \( \ell = 1, \ldots, N_{\text{iter}} \), we simulate the forests \( \nu_{\ell}^{\theta_k^{-1} \pm c_k} \)

% empirical population sizes:

\( \hat{N}^{\pm} = \frac{1}{2} \sum_{l=N_{\text{iter}} - \ell}^{N_{\text{iter}}} (\nu_{\ell}^{\theta_k^{-1} \pm c_k}, 1) \)

% Kiefer-Wolfowitz iteration:

\( \theta_k^{\pm} - \theta_k^{\pm-1} = a_k \frac{ (\hat{N}^{+} - N^*)^2 - (\hat{N}^{-} - N^*)^2}{4 \sigma_k} \)

\( \theta_k^{\pm} \leftarrow \nabla V_{\beta_k}^c(\theta_k^{\pm}) \)

end for

output: \( \hat{\theta}_l \leftarrow \frac{\sum_{k=1}^{K} a_k \theta_k^{\pm-1}}{\sum_{k=1}^{K} a_k} \)

---

**Alg. 3:** Kiefer-Wolfowitz Algorithm for parameters estimation.

30
We run the algorithm for $K$ iterations. The suites $a_k$ and $c_k$ depend on the number of algorithm iteration $k$. We have chosen to perform the simulations with $a = c = 1$. Below we have plotted, for the birth, death, competition, radius of the influence zone and the standard deviation (of the dispersion) parameters, the trajectories $(\theta^k_l)_{l=1}^5$ where $k = 1, \ldots, 10000$ and using a target $N^* = 300$.

Fig. 6: Figures shown the convergence diagnostics, the whole trajectory of $(\theta^k_l)_{l=1}^5$ simulated at each iteration $k$, for the estimation of the minimum using the Kiefer-Wolfowitz algorithm and a target $N^* = 300$.

6. Discussion

In this framework, we describe and simulate an individual-based model for understanding forest dynamics. The approach considered is simple where each tree (individual) is explicitly represented by their position in the population. The basic mechanisms (birth, dispersal, natural death, death by competition) are explicitly described. In fact, the level of competition is explained as a death rate $\lambda^c(x, \nu)$ which depends on the position and a function $u(x, y)$ which represents the strength of competition exercised on the individual in the state $x$ from an individual in the state $y$ in the population $\nu$. Although very simple, the resulting model presents very interesting features. It helps to account for different behaviours depending on: low or high dispersion and low or high
competition.

The previous Monte Carlo numerical simulations suggest that despite of its simplicity, this model can account for a large variety of aggregations. The relationship between the spatial pattern of trees and the competition/ dispersion level is thus understood: the excessive impact of the low dispersion that favours the establishment of clusters is diagnosed and the impact of the high competition that favours the desertification of the forest is observed.

Finally, to estimate the model parameters, we are interested in a stochastic gradient algorithm in the classical Kiefer-Wolfowitz version which allows to solve a minimization problem under several parameters simultaneously and given a target feature. In conclusion, we can say that, on one hand, convergence of the process described in this paper as time tends to infinity and conditioned on nonextinction should be more investigated to study the random measure of this situation. On the other hand, there is clearly a deterministic approximation of the process that could be interesting to study.

Acknowledgments. This article was mainly prepared while the author was working at UMR Mistea of SupAgro Montpellier. I would like to thank an Associate Editor and two anonymous referees for their insightful comments on a first draft of this work.

References


