Int. J. Anal. Appl. (2025), 23:15

International Journal of Analysis and Applications

Central Limit Theorem for Markov Chains with Variable Memory via the Chen-Stein Method

Victorien Konané[1,2,*], **Claude Yaméogo**[2], **Wahabo Baguian**[2]

[1]*Mathematics Department, Université Joseph KI-ZERBO, 03 B.P. 7021, Ouagadougou 03, Burkina Faso*
[2]*Laboratoire de Mathématique et Informatique, Burkina Faso*

[*]*Corresponding author:* kfourtoua@gmail.com

Abstract. In this paper, we studied Markov chains of variable length and the convergence of persistent walk.We, also, looked at the rate of convergence of such process. We also provide the use of variable-memory stochastic chains in risk models.

1. Introduction

In modeling stochastic phenomena, several theories have been developed over time. Among others, phenomena having long memory in the past, like Markov chains with variable memory saw the light of day with [9]. This scientific approach was originally developed to overcome the limitations of modeling using, up to now, Markov's chain of fixed-order. Indeed, using Markov's chains of fixed-order to fit complex data would require a very high order. As a result, the variable-memory chains permit to avoid the need to estimate an exponentially increasing number of parameters required to describe a Markov's chain of unknown order.

Markov's variable-memory chains are very useful in several fields. Among others, in biology, more specifically in genetics, in the reproduction process of plants or in molecular biology in converting genes into proteins. For example, if we look at the reproductive process of a plant in which each seedling is crossed with a hybrid plant Gg, we can define a Markov chain whose

Received: Nov. 24, 2024.

2020 *Mathematics Subject Classification.* 60J10, 60J20, 60G50, 91B05,

Key words and phrases. Markov chains of variable length; persistent random walk; Chen-Stein method, counting process, risk model.

https://doi.org/10.28924/2291-8639-23-2025-15
ISSN: 2291-8639

transition matrix is:

$$\Sigma = \begin{pmatrix} & \mathbf{GG} & \mathbf{Gg} & \mathbf{gg} \\ \mathbf{GG} & 1/2 & 1/2 & 0 \\ \mathbf{Gg} & 1/4 & 1/2 & 1/4 \\ \mathbf{gg} & 0 & 1/2 & 1/2 \end{pmatrix}$$

On the other hand, if reproduction takes place with a dominant GG plant, then the transition matrix is:

$$\Delta = \begin{pmatrix} & \mathbf{GG} & \mathbf{Gg} & \mathbf{gg} \\ \mathbf{GG} & 1 & 0 & 0 \\ \mathbf{Gg} & 1/2 & 1/2 & 0 \\ \mathbf{gg} & 0 & 1 & 0 \end{pmatrix}$$

Furthermore, Markov chains are used in actuarial science. For example, the contribution model for a policyholder might look like this: the contribution paid for the $(n+1)$-th year depends on the policyholder's contribution for the n-th year and the number of accidents he or she has had in that same n-th year.

On the other hand, from a statistical point of view, the partial sum of these chains with a short memory defines a stochastic process known as a persistent random walk. These type of processes were first introduced by Kac and are still referred to as Kac walk or correlated random procedure. More precisely, let $(\xi_i)_{i\in\mathbb{N}}$ be a sequence of random variables with value in $\{0;1\}$ such that

$$\begin{aligned} \xi_0 &= \cdots \hbar_{-2}\hbar_{-1}\hbar_0 \\ \xi_1 &= \cdots \hbar_{-2}\hbar_{-1}\hbar_0\hbar_1 \\ &\cdots \\ &\cdots \\ &\cdots \\ \xi_n &= \cdots \hbar_{-2}\hbar_{-1}\hbar_0\hbar_1 \times \cdots \times \hbar_n, \end{aligned}$$

where the sequence $(\hbar_i)_{i\in\mathbb{Z}}$ represents the increments of (ξ_k). By introducing a memory random variable $(\mathcal{M}_n)$ on $\hbar_i$ such that

$$\mathcal{M}_n = 1 + \sup\{0 \leq i \leq n, \hbar_{n-j} = \hbar_n, \forall j \in \{0, \cdots, i\}\} = \inf\{0 \leq i \leq n, \hbar_{n-1} \neq \hbar_n\}, \tag{1.1}$$

the sequence $(\hbar_i)$ is a Markov chain with transition probability

$$P(\hbar_{i+1} = \chi_{i+1} | \hbar_i = \chi_i) \tag{1.2}$$

with transition matrix

$$\nabla = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}, \ \ 0 \leq \alpha \leq 1, \ \ 0 \leq \beta \leq 1. \tag{1.3}$$

Set

$$S_n(\xi) = \sum_{i=0}^{n} \hbar_i \text{ with } S_0(\xi) = 1 \text{ or } S_0(\xi) = -1. \tag{1.4}$$

The partial sum $S_n(\xi)$ is then called a persistent random walk. If $\beta = 1 - \alpha$, it is called a classical random walk. It is called a Kac walk when $\alpha = \beta$.

A vast literature on this particular type of stochastic process has appeared since the 20th century. Under the assumptions $\alpha + \beta = 1$ and $(\hbar_i)$ independent, we obtain a Bernoulli random walk. Furthermore, assumptions $\alpha = \beta$ are studied in [3]. The limiting process of $(\hbar_i)$ is the Integrated Telegrated Noise (ITN) (see [4]). By asymptotic analysis in [5] and by Fourier transform in [6], we show that this limit process allows us to represent the solution of the telegraph equation:

$$\begin{cases} \frac{\partial^2 u}{\partial t^2}(x,t) = c^2 \frac{\partial^2 u}{\partial x^2}(x,t) - 2\lambda \frac{\partial u}{\partial t}(x,t) \quad \text{for all} \quad x \in \mathbb{R},\ t \geq 0 \\ \\ u(x,0) = f(x),\ \dfrac{\partial u}{\partial t}(x,0) = 0. \end{cases} \tag{1.5}$$

where $u(x,t)$ represents the voltage in a cable at point x at time t, f is a function of class C^2 in $\mathbb{R}$, with bounded first and second derivatives. λ is the parameter of the process $(\hbar_i)$.

The calculation of the probability distribution of the process $S_n(\xi)$, and the study of the influence of the persistence phenomenon on the first and second-order moments of the process $(\hbar_i)$ when it has a value in $\{0;1\}$ are studied in [1] and [2] respectively. As for the characteristic function of the $S_n(\xi)$ process, it is determined in [19] by Proposition 6.4 in Section 6. Moreover, several research questions on this stochastic process remain deeply unexplored, such as the acceleration of its convergence or the existence of its infinite-length chains. Recently, S. Herrmann and P. Vallois studied the limit law of the $S_n(\xi)$ random process by constructing a chain of large memory for the random sequence $(\hbar_i)$.

In this work, we extend the work in [] by constructing a Markov process of infinite length. We justify the existence of such a stochastic process. We evaluate the rate of convergence of the process $S_n(\xi)$ in the continuous case by mean of the Chen-Stein method. Thus, the paper is organized as follows: in section 2 we make preliminaries. In these preliminaries, we present notions and properties on Markov chains of variable lengths and on the Kac walk. We also present the Chen-Stein method for evaluating the rate of convergence of stochastic processes. In section 3, we construct and justify the existence of a Markov chain of infinite length for a persistent random walk. Section 4 is devoted to the study of the law-like convergence of the persistent random walk via the Chen-Stein method. The rate of this convergence is evaluated. In section 5, we will discuss the use of variable-memory stochastic chains in the study of counting processes for risk models.

2. Preliminaries

2.1. **Markov chain with variable memory.** This stochastic process was first introduced in literature, in 1983, by Rissanen [9]. Rissanen's ingenious idea was the basis for large-scale data compression in information theory. Today, his theory is very useful for modeling in many fields. We gave a few examples in our introduction.

Definition 2.1 (Markov chain with variable memory)**.** *A Markov Chain with Variable Memory (VLMC) is a markov chain for which the dependence on the past is unbounded.*
We call context the part of the past of the process (ξ_i) needed to predict the next symbol. In the following constructions, for any $x \in \{-1; 1\}$, we denote the context of process (ξ_i) by $\overleftarrow{pref}(x_j)_{j\in -\mathbb{N}}$.

Remark 2.1. *In this definition, the context length is not limited. However, it is a deterministic function of the string of passed symbols. The main mixing class for VLMC is ψ − mixing. In particular, non-uniform polynomial blending when the deterministic function of the past symbol string is the natural logarithm. Mixing is uniform exponential when the function is exponential.*

2.2. **Persistent random walk.** We consider equation 1.4 in its entirety.

Proposition 2.1. .
The trajectory of the Kac walk is a straight line and the jump times are random variables with geometric distribution.

Proof. For the proof, see []. □

The time spent in a state before leaving it leads us to define the stopping time of the process (τ_n) by

$$\tau_n = \inf\{i \geq \tau_{n-1}, \xi_i \neq \xi_{\tau_{n-1}}\}. \tag{2.1}$$

We introduce the counting process $(\mathcal{W}_t,\ t \in \mathbb{R}_+)$ with its stopping times 2.1 by :

$$\mathcal{W}_t = \sup\{n \geq 1, \tau_n \leq t\}. \tag{2.2}$$

The process $(\mathcal{W}_t)$ is a counting process and follows a Poisson distribution of intensity $\lambda > 0$ (see [7]). An illustration of its trajectory is given in Figure 1 in [].

Proposition 2.2. *Let's assume $(\hbar_0, \mathcal{M}_0) = (-1; 1)$. For all $n \geq 1$ we have*

$$\mathcal{M}_t = 1 + \sup\{n \geq 0, \mathcal{W}_{t-n} = \mathcal{W}_t,\ \forall t \in \mathbb{N}\} \tag{2.3}$$

and

$$S_n(\xi) = 1 + 2\sum_{i=1}^{n} \mathbb{1}_{\{\hbar_i = 1\}} - n. \tag{2.4}$$

Proof. It can be seen that

$$\{\mathcal{M}_p = 1 : p \in \mathbb{N}^*\} = \{\tau_n; n \geq 0\}. \tag{2.5}$$

By combining 1.1, 2.2 and 2.5, we obtain 2.3.

$$\begin{aligned} S_n(\xi) =& 1+\sum_{i=1}^{n}\mathbb{1}_{\{\hbar_i=1\}}-\sum_{i=1}^{n}\mathbb{1}_{\{\hbar_i=-1\}} \\ =& 1+\sum_{i=1}^{n}\mathbb{1}_{\{\hbar_i=1\}}-\left(n-\sum_{i=1}^{n}\mathbb{1}_{\{\hbar_i=1\}}\right) \\ =& 1+2\sum_{i=1}^{n}\mathbb{1}_{\{\hbar_i=1\}}-n. \end{aligned}$$

□

2.3. **Chen-Stein method.** The Central Limit Theorem states that if the process $(\xi_i)_{i\geq 1}$ is independent and identically distributed $(i.i.d)$, admitting a moment of order 2, with $\mathbb{E}(\xi_1)=\mu$ and $\mathbb{V}(\xi_1)=\sigma^2>0$ then:

$$\frac{(\xi_1+\cdots+\xi_n)-n\mu}{\sigma\sqrt{n}}\xrightarrow{\mathcal{L}}\mathcal{N}(0;1). \tag{2.6}$$

For a long time, the approach used to prove this theorem was to demonstrate convergence in distribution. This approach relied heavily on Fourier methods and the study of characteristic functions. Later, with Andrew Campbell Berry and Carl-Gustav Esseen [10] a more explicit approximation to the normal distribution was obtained, using so-called Berry-Esseen inequalities. However, since random variables being not always independent, new methods were developed in 1972 by Charles Stein and in 1975 by Chen for such cases.

2.3.1. *Description of Stein's method.* The starting point for Stein's method is a characterization of the normal distribution using expectation and an absolutely continuous function. The determination of a so-called Stein operator linked to a differential equation and its resolution is then established. Finally, we look for an increase in the norm of the solution obtained, as well as that of its successive derivatives. The method makes it possible to obtain bounds between the distribution of a sum of dependent random variables and the distribution of a random variable that follows a normal distribution for the Kolmogorov metric:

$$\sup_{h\in\mathcal{H}}\left|\int hdP-\int hdQ\right|=\sup_{h\in\mathcal{H}}\left|\mathbb{E}[h(X)]-\mathbb{E}[h(Y)]\right| \tag{2.7}$$

where P and Q are probability measures on the same measurable space. X and Y are random variables with laws P and Q respectively. $\mathcal{H}$ is a set of functions defined on the measurable set and having values in $\mathbb{R}$.

For a more detailed account of Stein's method, we refer readers to the article [8].

2.3.2. *Description of Chen's method.* Stein's method is similar to Chen's method. Both methods generalize the rare event theorem to the case where events are non-independent and the dependencies between them are small. However, Chen's method relies on Stein's to obtain results on the approximation of Poisson's law in terms of expectation. An operator and then a Chen equation

are first considered. The search for a limit in terms of the total variation distance between two distributions is then undertaken.
For any function h defined on $\{-1;1\}$, we define the total variation distance as:

$$||h|| = \sup_k |h(k)| \tag{2.8}$$

where $||.||$ denotes the Euclidean norm: $\forall x \in \{0,1\},\ ||x|| = \sqrt{x.x}$.

3. Construction of a Markov chain with variable memory

Knowing that with Gibbs measures with continuous Holdierian interactions, we can obtain stochastic chains of infinite order (see [12]), we are now witnessing new approaches to the Markov process. Since 2004, new constructions of Markov chains have appeared. For more details on these constructions, we refer you dear readers to the following articles: [13], [11] and [17]. In this part of our work, starting from a stochastic chain of infinite order, we prove the existence of a Markov chain of variable length in $\{0;1\}^{-\mathbb{N}}$ with a stationary distribution. Our results pursue the following objective: Extend Hermann's work by redefining the persistent random walk with infinite memory length for an irreducible aperiodic chain.
In the remainder of this paper, we denote by $\eth$ the set of strings in the process $(\hbar_i)$. The following construction is inspired by [17] (see definitions 5.2 and 5.3 in Section 5).
For all $x \in \{-1;1\}$, let

$$(P\Big(x|\overleftarrow{pref}(x_j)_{j\in-\mathbb{N}}\Big)\Big)_{(x,\overleftarrow{pref}(x_j)\in\{-1;1\}^2} \tag{3.1}$$

a family of positive real numbers satisfying

$$\sum_x P\Big(x|\overleftarrow{pref}(x_j)_{j\in-\mathbb{N}}\Big) = 1. \tag{3.2}$$

For all $\aleph \in \eth$, let us note

$$P\Big(x|\overleftarrow{pref}(x_j)_{j\in-\mathbb{N}}\Big) = P(x|\aleph). \tag{3.3}$$

We redefine 1.2 by positing $\chi_{i+1} = x$ and $\chi_i = \overleftarrow{pref}(x_i)$ then consider 1.4 and the stochastic process $(\xi_i)_{i\in\mathbb{N}}$ defined in the introduction. The stochastic process $(\hbar_i)$ is a stochastic chain of infinite order with variable-length memory.
We make the following assumptions:

$$(1-\beta)(1-\alpha) - \alpha^2 \neq 0; \tag{3.4}$$

$$\exists n \in \mathbb{N} \text{ such that } \nabla^n \text{ is strictly positive.} \tag{3.5}$$

These assumptions imply that the transition matrix of the chain $(\hbar_i)$ is invertible and, consequently, that the chain is irreducible.
For probability family 3.1, , we have the following result

Theorem 3.1.
Based on the above assumptions, the family $\left(P\left(x|\overleftarrow{pref}(x_j)_{j\in -\mathbb{N}}\right)\right)_{(x,\overleftarrow{pref}(x_j)\in\{-1;1\}^2}$ *admit a unique stationary law on* $(\{-1;1\},\{0;1\}^{-\mathbb{N}},\mathcal{F},P)$.

Proof. Let's consider $P_\triangledown(\lambda)$ as the characteristic polynomial of $\triangledown$.
We have

$$\begin{aligned} P_\triangledown(\lambda) &= det(\triangledown - \lambda I_2) \\ &= \begin{vmatrix} 1-\alpha-\lambda & \alpha \\ \beta & 1-\beta-\lambda \end{vmatrix} \\ &= 1-\beta-\lambda-\alpha+\alpha\lambda-\lambda+\lambda\beta+\lambda^2 \end{aligned}$$

We note that 1 is an eigenvalue of $\triangledown$. We can also check that 1 is an eigenvalue of ${}^t\triangledown$(${}^t\triangledown$ denotes the transpose of $\triangledown$.) Now we just need to show that there is a single probability vector such that the product of it and the transition matrix $\triangledown$ is equal to itself.
Let $\vec{\vartheta}\begin{pmatrix}\vartheta_1\\ \vartheta_2\end{pmatrix}$ be the strictly positive eigenvector associated with the eigenvalue $\lambda = 1$.(Such a vector exists. Simply take the absolute value of the coordinates of the eigenvector associated with eigenvalue $\lambda = 1$ for matrix ${}^t\triangledown$).
We have

$$\vartheta_1 P\left(x|\overleftarrow{pref}(x_j)_{j=\overline{-\infty,0}}\right) + \vartheta_2 P\left(x|\overleftarrow{pref}(x_j)_{j\in-\mathbb{N}}\right) - (\vartheta_1 - \vartheta_2) = 0. \tag{3.6}$$

As a result

$$\vec{\vartheta}\triangledown = \vec{\vartheta}. \tag{3.7}$$

□

Remark 3.1. *The unique stationary law* $\nu\begin{pmatrix}\nu_1\\ \nu_2\end{pmatrix}$ *is therefore:*

$$\nu = \lim_{n\to+\infty} \frac{1}{n+1}\sum_{k=1}^{n} \triangledown^k \tag{3.8}$$

and it verifies

$$\begin{cases} (1-\alpha)\nu_1 + \beta\nu_2 = \nu_1 \\ \alpha\nu_1 + (1-\beta)\nu_2 = \nu_2 \\ \nu_1 + \nu_2 = 1. \end{cases} \tag{3.9}$$

We obtain the following corollary from the previous theorem:

Corollary 3.1. *There exists a unique stationary measure* ν *for the process* (ξ_i).

Justification of the Markov process: By associating the process (ξ_i) with the transition probability family of Theorem 3.1, equality 3.3 allows us to see that the process $(\xi_i)_{i\in\mathbb{N}}$ eis markovian, constructed from a stochastic chain with variable memory.

4. Evaluation of the rate of convergence by the Chen-Stein method for a Markov chain of infinite length.

The need to evaluate and/or accelerate the convergence of sequences has long been fundamental to research in applied mathematics. Today, it remains central to research in probability theory. Indeed, in many fields we are sometimes called upon to use sequences while they show slow convergence. Pioneers in this field of research for numerical sequences, Richardson and Aitken followed in their footsteps for stochastic sequences. Since then, a vast literature has appeared in this new field of research, led by Berry [14]. For a more detailed account of the historical evolution, see [15]. With regard to linear and nonlinear stochastic processes in separable Hilbert spaces, notable results were obtained with Jirak [16]. We have also made our contribution in this direction in ours previous articles [20] and [23].
In this part of the paper, we focus on convergence. Thus, under the functional assumptions of the stochastic chain $(\hbar_i)$, we study the convergence in law of the persistent random walk 1.4. By a conditioning argument similar to the mind of P. Cénac in [] and contrary to their approach, we apply the Chen-Stein method to extend the previous results by evaluating the rate of convergence. As we have already pointed out in the introduction, we consider the variable memory context of the process $(\hbar_i)$. Therefore, we first control the process (ξ_i) through its stopping times.

4.1. **Control over the Markov process.** Let $P^i_{a,b}$ be the probability of going from state a to state b if we make a decision $d \in \{0;1\}^{-\mathbb{N}}$. For all $i \in \mathbb{N}$, we denote control by the random variable O_i. Considering a given strategy S, we define a probability on the trajectory of the controlled chain. Let $P_{O_i}(d)$ denote the probability law on the random variable O_i. Let $\mathcal{R}_i(a)$, denote the set of feedback strategies and $\mathcal{U}_i(a)$ the set of deterministic Markov strategies. Optimal control therefore consists in optimizing the trajectory of the process (ξ_i) through the following problem:

$$\min_{s\in\mathbb{S}(.)} E\left[\sum_{i=0}^{n-1} c^{O_i}(\xi_i)|\xi_0 = a\right] \tag{4.1}$$

where $\mathbb{S}(.)$ is the set of strategies and c is the instantaneous cost function. By virtue of Markov strategies, let us set

$$\begin{cases} c^W(a) = \displaystyle\sum_{d\in\{0;1\}^{-\mathbb{N}}} c^d(a) P_{O_i(a)}(d) \quad \text{si} \quad W \in \mathcal{R}_i(a) \\ \\ c^W(a) = c^{W(a)}(a) \quad \text{si} \quad W \in \mathcal{U}_i(a). \end{cases} \tag{4.2}$$

The following result controls the stopping times of the process ξ_i.

Proposition 4.1. *Let $\mathcal{J}_i(x)$ be a function on $\{-1;1\}$; $0 \le i \le n$. Let's note*

$$\mathcal{R}_i(x) = \sup_{\mathcal{D},\mathcal{F}_i, i\le\mathcal{D}\le n} \mathbb{E}\left[\sum_{k=1}^{\mathcal{D}-1} c^{(k)}(\xi_k) + \mathcal{J}_{\mathcal{D}_i}(\xi_{\mathcal{D}_i})|\xi_i = a\right] \tag{4.3}$$

and let's rewrite the stop delay 2.1 by

$$\mathcal{D}_i = \inf\{n \geq k \geq i, \mathcal{R}_k(\xi_k) = \mathcal{J}_k(\xi_k)\}. \tag{4.4}$$

$\mathcal{D}_i$ is an optimal stopping time for the process $(\xi_i)_{i\geq 0}$. Furthermore

$$\mathcal{R}_i(\xi_i) = \mathbb{E}\Big[\sum_{k=i}^{\mathcal{D}_i-1} c^{(k)}(\xi_k) + \mathcal{J}_{\mathcal{D}_i}(\xi_{\mathcal{D}_i})|\xi_i\Big] \tag{4.5}$$

Proof. Let's consider the process $(Z_\ell)_{i\leq\ell\leq n}$ defined by:

$$Z_\ell = \mathcal{R}_l(\xi_l)\mathbb{1}_{\{\ell<\mathcal{D}_i\}} + \mathcal{J}_{\mathcal{D}_i}(\xi_{\mathcal{J}_i})\mathbb{1}_{\{\mathcal{D}_i\leq\ell\}} + \sum_{j=i}^{\ell-1} c^{(j)}(\xi_j)\mathbb{1}_{\{j<\mathcal{D}_i\}} \tag{4.6}$$

Since $\mathcal{D}_i \leq n$, we have

$$\begin{aligned} Z_i =& \mathcal{J}_i(\xi_{\mathcal{D}_i}) + \sum_{j=1}^{n-1} c^{(j)}(\xi_j)\mathbb{1}_{\{j<\mathcal{D}_i\}} \\ =& \mathcal{J}_{\mathcal{D}_i}(\xi_{\mathcal{D}_i}) + \sum_{j=i}^{\mathcal{D}_i-1} c^{(j)}(\xi_j). \end{aligned}$$

Also, we have

$$Z_i = \mathcal{J}_i(\xi_i)\mathbb{1}_{\{i=\mathcal{D}_i\}} + \mathcal{J}_{\mathcal{D}_i}(\xi_{\mathcal{D}_i})\mathbb{1}_{\{\mathcal{D}_i\leq n\}} \tag{4.7}$$

as $i \leq \mathcal{D}_i$ and $\mathcal{J}_{\mathcal{D}_i}(\xi_i)\mathbb{1}_{\{\mathcal{D}_i=i\}} = \mathcal{R}_i(\xi_i)\mathbb{1}_{\{\mathcal{D}_i=i\}}$ then $Z_i = \mathcal{R}_i(\xi_i)$.
Therefore,

$$\begin{aligned} Z_\ell =& \Big(\triangledown\,\mathcal{R}_{\ell+1}(\xi_\ell) + c^{(\ell)}(\xi_\ell)\Big)\mathbb{1}_{\{\ell<\mathcal{D}_i\}} + \sum_{j=i}^{l-1} c^{(j)}(\xi_j)\mathbb{1}_{\{j<\mathcal{D}_i\}} + \mathcal{J}(\xi_{\mathcal{D}_i})\mathbb{1}_{\{\mathcal{D}_i\leq\ell\}} \\ =& E\Big[\mathcal{R}_{\ell+1}(\xi_{\ell+1})|\mathcal{F}_\ell\Big]\mathbb{1}_{\{\ell<\mathcal{D}_i\}} + \sum_{j=i}^{\ell} c(j)(\xi_j)\mathbb{1}_{\{j<\mathcal{D}_i\}} + \mathcal{J}_{\mathcal{D}_i}(\xi_{\mathcal{D}_i}\mathbb{1}_{\{\mathcal{D}_i\leq\ell\}} \\ =& E\Big[\mathcal{R}_{\ell+1}(\xi_{\ell+1})|\mathcal{F}_\ell\Big] + E\Big[\mathcal{J}_{\mathcal{D}_i}(\xi_{\mathcal{D}_i})\mathbb{1}_{\{\mathcal{D}_i\leq\ell\}} + \sum_{j=i}^{\ell} c^j(\xi_j)\mathbb{1}_{\{j<\mathcal{D}_i\}}|\mathcal{F}_l\Big] \\ =& E\Big[\mathcal{J}_{\ell+1}\mathbb{1}_{\{\ell+1<\mathcal{D}_i\}} + \mathcal{J}_{\mathcal{D}_i}(\xi_{\mathcal{D}_i})\mathbb{1}_{\{\mathcal{D}_i<\ell+1\}} + \sum_{j=i}^{\ell} c^{(j)}(\xi_j)\mathbb{1}_{\{j<\mathcal{D}_i\}}|\mathcal{F}_\ell\Big] \\ =& E[Z_{\ell+1}|\mathcal{F}_\ell] \end{aligned}$$

The process Z_ℓ is therefore a discrete martingale. Hence $Z_i = E[Z_n|\mathcal{F}_i]$, for all $n \geq i$.
We conclude that

$$\mathcal{R}_i(\xi_i) = E\Big[\mathcal{J}_{\mathcal{D}_i}(\xi_{\mathcal{D}_i}) + \sum_{j=i}^{\mathcal{D}_i-1} c^{(j)}(\xi_j)|\mathcal{F}_i\Big] \tag{4.8}$$

□

4.2. **Rate of convergence using the Chen-Stein method.** .

Notable results establishing convergence or evaluating the rate of convergence for Markov chains are given by [] and []. In [], the result is that of the convergence of the persistent random walk, where it is shown under the memory process condition ($\mathcal{M}_0 = 1$) that $S_n(\xi)$ in the continuous case converges in law to the sum of a certain independent random variable. As for the result in [], the chain is assumed to be strongly ergodic, then by a Nagaev method coupled with a perturbation theorem of Keller and Liverani, the rate of convergence in $n^{-1/2}$ is obtained.

For our part, we evaluate the rate of convergence of the persistent random walk by applying the Chen-Stein method and the mind of Jirak's approach [16] to boundary setting. Recall always that we consider 1.4 in its entire construction as well as the counting process $\mathcal{W}_t$. For all $k \in \mathbb{N}$,

$$P(\mathcal{W}_{\lambda t} = k) = e^{-\lambda \times t} \frac{(\lambda t)^k}{k!} \tag{4.9}$$

Before reporting our convergence result, we make the following assumptions:

($\mathbf{H_1}$) The stopping time process (τ_n) is an independent random variable.
($\mathbf{H_2}$) The probability measure P is ergodic.
($\mathbf{H_3}$) There exists a function h that is ν-integrable from $\{-1;1\}$ into $\mathbb{R}$ such that $\nu(h) = 0$.
($\mathbf{H_4}$) $\liminf_{n \to +\infty} E[|S_n h(\xi)|^2] > 0$.

Remark 4.1. *Assumption* ($\mathbf{H_3}$) *allows us to hold 1.4 in the continuous case. Thus, we display* $S_n h(\xi) = \sum_{i=1}^{n} h(\hbar_i)$*. This hypothesis* ($\mathbf{H_3}$) *is crucial for showing convergence of the partial sum* $S_n h(\xi)$ *with Chen and Stein's method. The assumption* ($\mathbf{H_4}$) *is a non-degeneracy condition.* ($\mathbf{H_1}$) *plays the role of determining the limit process of* $S_n h(\xi)$*. It allows us to approximate* $S_n h(\xi)$ *to a Poison distribution.*
Since the process $(\hbar_i)$ *is a stochastic chain of variable length, it is necessary to guarantee the computation of its moments. Assumption* ($\mathbf{H_2}$) *assures us of the existence of the moments of the stochastic process* $S_n h(\xi)$*.*

Now, we state the theorem that gives the main result of our paper and also the second objective of our work.

Theorem 4.1. *Under assumptions* ($\mathbf{H_1}$) − ($\mathbf{H_4}$)*, the persistent random walk* $S_n h(\xi)$ *converges to the limit process* $\mathcal{Z} = \int_0^t \Big(\sum_{k=0}^{+\infty} (-1)^k e^{-\lambda t} \frac{(\lambda t)^k}{k!} \Big) dt$ *with a convergence rate in the order of* $n^{-1/2}$*. That is,* $\sup_{a \in \mathbb{R}} \Big| P\Big(S_n h(\xi)\Big) \leq a - \mathcal{Z} \Big| = o(n^{-1/2})$.

Proof. For the proof, we proceed in two steps. We begin by adopting the following notations: First, we denote parameter-dependent constants by $c(:)$.

For all $n \in \mathbb{N}^*$, let $I_1, \cdots, I_n$ be independent events and $\Theta_1, \cdots, \Theta_n$ the number of its events that occur.

Then, denote by $\nu_{S_n h(\xi)}$ the law of $S_n h(\xi)$ and by $\nu_{\mathcal{P}(\lambda)}$ the law of $\sum_{k=0}^{+\infty} (-1)^k e^{-\lambda t} \frac{(\lambda t)^k}{k!}$.

We have

$$\nu_{S_n h(\xi)} = P\Big(S_n h(\xi) \leq a\Big) \text{ and } \nu_{\mathcal{P}(\lambda)} = \mathcal{Z}. \tag{4.10}$$

Step 1: We prove the convergence in distribution of $S_n h(\xi)$ to $\mathcal{Z}$ by determining the upper bound of the distance in total variation between $\nu_{S_n h(\xi)}$ and $\nu_{\mathcal{P}(\lambda)}$. To do that, we first consider equation

$$\lambda\varphi(n+1) - k\varphi(n) = \Theta_I - \sum_{k=1}^{+\infty} \Theta_I e^{-\lambda}\frac{\lambda^k}{k!},\ \forall I \subset \mathbb{N}. \tag{4.11}$$

and we denote its solution by φ_I. It is easy to see that

$$\varphi_I(n) = -\frac{(n-1)!}{\lambda^n}\sum_{k=n}^{+\infty}\frac{\lambda^k}{k!}\Big(\Theta_I - \sum_{k=1}^{+\infty}\Theta_I e^{-\lambda}\frac{\lambda^k}{k!}\Big). \tag{4.12}$$

Then, we have

$$\begin{aligned}
\nu_{S_n h(\xi)}(I) - \sum_{k=1}^{+\infty}\Theta_I e^{-\lambda}\frac{\lambda^k}{k!} &= \int_{\mathbb{N}}\Big(\Theta_I - \sum_{k=1}^{+\infty}\Theta_I e^{-\lambda}\frac{\lambda^k}{k!}\Big)d\nu_{S_n h(\xi)} \\
&= E\Big[\lambda\varphi_I\Big(\sum_{i=1}^{n} h(\hbar_i)+1\Big) - \sum_{i=1}^{n} h(\hbar_i)\varphi_I\Big(\sum_{i\in\Theta} h(\hbar_i)\Big)\Big] \\
&= \sum_{i=1}^{n} E\Big[\frac{\lambda}{n}\varphi_I\Big(\sum_{j\neq i} h(\hbar_j)+\hbar_i+1\Big) - \hbar_i\varphi_I\Big(\sum_{j\neq i} h(\hbar_j)+1\Big)\Big].
\end{aligned}$$

Since the stochastic process $(\hbar_i)$ is non-independent, we consider both strong and weak dependencies within it. Let (o) be the set of indices of variables $\hbar_j$ strongly dependent on $\hbar_i$ and (e) the set of indices of variables $\hbar_j$ weakly dependent on $\hbar_i$ with

$$S_n^{(o)}(\xi) = \sum_{i\in(o)} h(\hbar_i) \text{ and } S_n^{(e)}(\xi) = \sum_{i\in(e)} h(\hbar_i) \tag{4.13}$$

Then, we have the following decomposition

$$S_n h(\xi) = S_n^{(o)}(\xi) + S_n^{(e)}(\xi) \tag{4.14}$$

$$\frac{\lambda}{n}\varphi_I\Big(\sum_{j\neq i} h(\hbar_j)+\hbar_i+1\Big) - \hbar_i\varphi_I\Big(\sum_{j\neq i} h(\hbar_j)+1\Big) = A(S_n^{(e)}(\xi)) + B(S_n^{(e)}(\xi)) + C(S_n^{(e)}(\xi)) \tag{4.15}$$

with:

$$A(S_n^{(e)}(\xi)) = \frac{\lambda}{n}\varphi_I\Big(\sum_{j\neq i} h(\hbar_j)+\hbar_i+1\Big) - \frac{\lambda}{n}\varphi_I\Big(S_n^{(e)}(\xi)+1\Big)$$

$$B(S_n^{(e)}(\xi)) = \frac{\lambda}{n}\varphi_I\Big(S_n^{(e)}(\xi)+1\Big) - \hbar_i\varphi_I\Big(S_n^{(e)}(\xi)+1\Big)$$

$$C(S_n^{(e)}(\xi)) = \hbar_i\varphi_I\Big(S_n^{(e)}(\xi)+1\Big) - \hbar_i\varphi_I\Big(\sum_{j\neq i} h(\hbar_j)+\hbar_i+1\Big)$$

Based on expression (4.15), we have

$$E\left[\frac{\lambda}{n}\varphi_I\Big(\sum_{j\neq i} h(\hbar_j)+\hbar_i+1\Big)-\hbar_i\varphi_I\Big(\sum_{j\neq i} h(\hbar_j)+1\Big)\right]$$
$$\leq \frac{\lambda}{n}E\left[\left|\varphi_I\Big(\sum_{j\neq i} h(\hbar_j)+\hbar_i+1\Big)-\varphi_I(S_n^{(e)}(\xi)+1)\right|\right]$$
$$+\left|E\left[\Big(\frac{\lambda}{n}-\hbar_i\Big)\varphi_I\Big((S_n^{(e)}(\xi)+1\Big)\right]\right|$$
$$+E\left[\hbar_i\left|\varphi_I\Big((S_n^{(e)}(\xi)+1\Big)-\varphi_I\Big(\sum_{j\neq i} h(\hbar_j)+1\Big)\right|\right]$$

For the rest of the proof, we will give the bounds of $A(S_n^{(e)}(\xi))$, $B(S_n^{(e)}(\xi))$ and $C(S_n^{(e)}(\xi))$.

Bound for $A(S_n^{(e)}(\xi))$:
We have

$$\left|\varphi_I\Big(\sum_{j\neq i} h(\hbar_j)+\hbar_i+1\Big)-\varphi_I\Big(S_n^{(e)}(\xi)+1\Big)\right| \leq \|\varphi_I(n+1)-\varphi_I(n)\|(\hbar_i-S_n^{(o)}(\xi)) \tag{4.16}$$

Therefore,

$$A(S_n^{(e)}(\xi) \leq \frac{\lambda}{n}E\left[\|\varphi_I(n+1)-\varphi_I(n)\|(\hbar_i-S_n^{(o)}(\xi))\right]$$
$$\leq \frac{1-e^{-\lambda}}{\lambda}E[(\hbar_i-S_n^{(o)}(\xi))]$$

because, according to [22], $\|\varphi_I(n+1)-\varphi_I(n)\| \leq \frac{1-e^{-\lambda}}{\lambda}$.

Bound for $B(S(S_n^{(e)}(\xi))$:
Increasing $B(S_n^{(e)}(\xi))$ is based on Jensen's inequality:

$$\left\|\frac{\lambda}{n}-E[\hbar_i|S_n^{(e)}(\xi)]\right\| \leq \left\|\frac{\lambda}{n}-E[\hbar_i|\sigma(X_j)_{j\in(e)}]\right\|. \tag{4.17}$$

where $\sigma(X_j)$ denotes the tribe on (X_j). Thus, according to 4.17, we have

$$\left|E\left[\Big(\frac{\lambda}{n}-\hbar_i\Big)\varphi_I\Big(S_n^{(e)}(\xi)+1\Big)\right]\right| = \left|E\left[\Big(\frac{\lambda}{n}-E[\hbar_i|S_n^{(e)}(\xi)]\Big)\varphi_I\Big(S_n^{(e)}(\xi)+1\Big)\right|$$
$$\leq \|\varphi_I\|\times E\left[\left|\frac{\lambda}{n}-E[\hbar_i|S_n^{(e)}(\xi)]\right|\right]$$
$$\leq \min\left(1,\sqrt{\frac{2}{e\lambda}}\right)\times E\left[\left|\frac{\lambda}{n}-E[\hbar_i|S_n^{(e)}(\xi)]\right|\right].$$

Bound for $C(S(S_n^{(e)}(\xi))$:

Similar to the case $A(S(S_n^{(e)}(\xi))$ we have

$$\left|\varphi_I\Big(S(S_n^{(e)}(\xi)+1\Big)-\varphi_I\Big(\sum_{j\neq i}h(\hbar_j)+\hbar_i+1\Big)\right|\leq\frac{1-e^{-\lambda}}{\lambda}S_n^{(o)}(\xi). \tag{4.18}$$

This gives

$$|\nu_{S_nh(\xi)}-\nu_{\mathcal{P}(\lambda)}|\leq c(\lambda)\bigg(2\frac{\lambda}{n}+\sum_{i\in\ominus\ ;\ j\in(o)\setminus\{i\}}E(\hbar_i\hbar_j+\sum_{i\in\ominus}\Big\|\frac{\lambda}{n}-E(\hbar_i|\sigma(\hbar_j)_{j\in(e)})\Big\|\bigg) \tag{4.19}$$

Step 2: Regarding the proof of the convergence rate evaluation in $n^{-1/2}$, according to hypothesis **H3**, the process $(\hbar_i)$ admits a covariance operator. Therefore, we can adapt theorem 3.2 in [16] and write:

$$|\nu_{S_nh(\xi)}-\nu_{\mathcal{P}(\lambda)}|\leq c(\lambda)mn^{-1/2},\ \forall m\geq E(\|\hbar_i\|^r,\ 2<r\leq 3 \tag{4.20}$$

□

5. Counting processes and the infinite-time risk model

Nowadays, with the multiplicity of claims, it is more necessary than ever for the actuary to focus more on the vulnerability of policyholders. Despite Lundberg's early work in 1903, the theory of ruin was born in probability theory. Numerous models have also emerged from this new theory. These include the Cramer-Lundberg and Sparre-Anderson models. These models are mainly characterized by two components: the number of claims (frequency) and the cost of each event (severity).

In this part of our paperwork, we discuss the (ξ_i) and $(\mathcal{W}_t)$ processes in the following ruin model:

$$\begin{cases} R(t)=u+ct-S_t\\ S_t=\sum_{i=1}^{N_t}X_i. \end{cases} \tag{5.1}$$

where:

-$R(t)$ represents the amount of an insurance company's reserves at time t;

- N_t the number of claims at time t;
- u is the initial capital;
- c is the rate at which premiums are received per unit of time;
- X_i is the cost of the ith claim;
- S_t is the insurance company's aggregate losses.

The main questions we ask ourselves in this model are as follows: Can the processes (ξ_i) and $(\mathcal{W}_t)$ respectively model the cost of claims and the number of claims in model (5.1) ? What are their limitations? The following proposition on our first objective will lead us to a discussion for the second question.

Proposition 5.1. *Given 2.2 and the process (ξ_i), model (5.1) over an infinite horizon can be written:*

$$R(t) = u + ct - \sum_{i=1}^{\mathcal{W}_t} \xi_i. \tag{5.2}$$

Proof. The proof of this proposition derives directly from the nature and law of the process $(\mathcal{W}_t)$ The proof of this proposition derives directly from the nature and law of the process (ξ_i). □

5.1. **Discussion.** While the result on the ruin model obtained through the previous proposal is attractive at first sight, it also highlights a problem that we present and discuss.
The process $(\mathcal{W}_t)$ is a counting process. It is a Poisson process of intensity $\lambda > 0$ (see [7]).

$$P(\mathcal{W}_{dt} = k) = \begin{cases} 1 - \lambda dt + o(dt) \quad \text{si} \quad k = 0 \\ \lambda dt + o(dt) \quad \text{si} \quad k = 1 \\ o(dt) \quad \text{si} \quad k \geq 2 \\ \lim_{t \to 0} o(dt)/t = 0. \end{cases} \tag{5.3}$$

Our model is supposed to represent the ruin of an insurance company in an infinite horizon. However, we note that :
(i) The process (ξ_i) in the model is Markovian, built from a stochastic chain of variable length.
(ii) The counting process $(\mathcal{W}_t)$ can be a renewal process which we denote by $(\mathcal{B}_k)$ (i.e. a point process: representing the instants of successive occurrences of a loss such that the inter-successive occurrence durations are independent real random variables with the same law) verifying relation

$$\forall k \geq 0, (\mathcal{W}_t \geq k) = (\mathcal{B}_k \leq t). \tag{5.4}$$

Let's develop case (i)): markovian processes are used in modeling in the field of actuarial science (e.g. model of the contribution to be paid by an insured). However, it should be noted that the class of stochastic processes best developed in risk modeling are independent and identically distributed random variables (Cramer-Lundberg model, Sparre-Andersen model). Would the use of a Markov process in ruin models therefore require homogeneous Markov chains?
In case (ii), we study the distribution of the process $(\mathcal{B}_k)$. The probability distribution of $(\mathcal{B}_k)$ is also the probability distribution of $(\mathcal{W}_t)$. Indeed, if Ψ_k is the distribution function of $(\mathcal{B}_k)$, we can show by recurrence on k that:

$$\Psi_k(x) = \int_0^x \Psi_{k-1}(x-y) d\Psi(y) \tag{5.5}$$

where F is the distribution function of the random variable representing successive inter-occurrence times.
According to (5.4),

$$\forall k \geq 0, P(\mathcal{W}_t \geq k) = P(\mathcal{B}_k \leq t). \tag{5.6}$$

That is to say

$$P(\mathcal{W}_t \geq k) = \Psi_k(t). \tag{5.7}$$

The above considerations indicate that model (5.1) is not at all a classical model in general ruin theory. The following question arises: what type of dependence in the chain of the Markov process (ξ_i) would make model (5.1) be of Cramer-Lundberg type, and what about the Gerber-Shiude function?

6. Conclusion

In this work, we have studied stochastic chains with variable memory. We first constructed a Markov chain from a variable-memory random chain. Then, we defined a persistent random walk in the continuous case and showed that the limit process of this persistent walk is a Poisson process. For this purpose, the Chen-Stein method was used. We, likewise, established a rate of convergence in the order of $\sqrt{n}$ $\forall n \in \mathbb{N}$ for the persistent random walk.
The issue of variable-memory Markov chains in risk models was also discussed in this paper. Indeed, we discussed the use of variable-memory Markov chains in a Crammer-Lundberg type ruin model and on the Gerber-Shiude expected penalty function.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] G.H. Weiss, Aspects and Applications of the Random Walk, North-Holland, Amsterdam, 1994. https://cir.nii.ac.jp/crid/1130000796986494208
[2] P. Vallois, C.S. Tapiero, Memory-Based Persistence in a Counting Random Walk Process, Physica A: Stat. Mech. Appl. 386 (2007), 303–317. https://doi.org/10.1016/j.physa.2007.08.027.
[3] G.I. Taylor, Diffusion by Continuous Movements, Proc. Lond. Math. Soc. 20 (1921), 196–212.
[4] S. Herrmann, P. Vallois, From Persistent Random Walk to the Telegraph Noise, Stoch. Dyn. 10 (2010), 161–196. https://doi.org/10.1142/S0219493710002905.
[5] S. Goldstein, On Diffusion by Discontinuous Movements, and on the Telegraph Equation, Q. J. Mech. Appl. Math. 4 (1951), 129–156. https://doi.org/10.1093/qjmam/4.2.129.
[6] G.H. Weiss, Some Applications of Persistent Random Walks and the Telegrapher's Equation, Physica A: Stat. Mech. Appl. 311 (2002), 381–410. https://doi.org/10.1016/S0378-4371(02)00805-1.
[7] M. Kac, A Stochastic Model Related to the Telegrapher's Equation, Rocky Mt. J. Math. 4 (1974), 497-509. https://www.jstor.org/stable/44236399.
[8] L.H.Y. Chen, L. Goldstein, Q.M. Shao, Normal Approximation by Stein's Method, Springer, Berlin Heidelberg, 2011.
[9] J. Rissanen, A Universal Data Compression System, IEEE Trans. Inf. Theory 29 (1983), 656–664. https://doi.org/10.1109/TIT.1983.1056741.
[10] A.C. Berry, The Accuracy of the Gaussian Approximation to the Sum of Independent Variates, Trans. Amer. Math. Soc. 49 (1941), 122–136. https://doi.org/10.1090/S0002-9947-1941-0003498-3.
[11] S. Gallo, N.L. Garcia, Perfect Simulation for Stochastic Chains of Infinite Memory: Relaxing the Continuity Assumption, arXiv:1005.5459 [math.PR] (2010). https://doi.org/10.48550/ARXIV.1005.5459.
[12] R. Fernandez, G. Maillard, Construction of a Specification from Its Singleton Part, arXiv:math/0409539 [math.PR] (2004). https://doi.org/10.48550/ARXIV.MATH/0409539.
[13] T.E. Harris, On Chains of Infinite Order, Pac. J. Math. 5 (1955), 707-724.

[14] A.C. Berry, The Accuracy of the Gaussian Approximation to the Sum of Independent Variates, Trans. Amer. Math. Soc. 49 (1941), 122–136. https://doi.org/10.1090/S0002-9947-1941-0003498-3.

[15] B.A. Zalesski, V.V. Sazonov, V.V. Ulyanov, A Sharp Estimate for the Accuracy of the Normal Approximation in a Hilbert Space, Theory Probab. Appl. 33 (1988), 700-701.

[16] M. Jirak, Rate of Convergence for Hilbert Space Valued Processes, Bernoulli 24 (2018), 202-230. https://doi.org/10.3150/16-BEJ870.

[17] P. Collet, A. Galves, Chains of Infinite Order, Chains with Memory of Variable Length, and Maps of the Interval, J. Stat. Phys. 149 (2012), 73–85. https://doi.org/10.1007/s10955-012-0579-6.

[18] P. Cénac, B. Chauvin, S. Herrmann, P. Vallois, Persistent Random Walks, Variable Length Markov Chains and Piecewise Deterministic Markov Processes, arXiv:1208.3358 [math.PR] (2012). https://doi.org/10.48550/ARXIV.1208.3358.

[19] S. Herrmann, P. Vallois, From Persistent Random Walk to the Telegraph Noise, Stoch. Dyn. 10 (2010), 161–196. https://doi.org/10.1142/S0219493710002905.

[20] V.F. Konane, C. Yaméogo, W. Baguian, Asymptotic Probability Expansions for Random Elements in a Hilbert Space, Contemp. Math. 4 (2023), 1048–1061. https://doi.org/10.37256/cm.4420232651.

[21] L. Hervé, Vitesse de Convergence Dans Le Théorème Limite Central Pour Des Chaînes de Markov Fortement Ergodiques, Ann. Inst. Henri Poincaré, Probab. Stat. 44 (2008), 280-292. https://doi.org/10.1214/07-AIHP101.

[22] A.D. Barbour, L.H.Y. Chen, An Introduction to Stein's Method, World Scientific, 2005. https://doi.org/10.1142/5792.

[23] C. Yameogo, V.F. Konane, W. Baguian, On Hilbertian Stochastic Processes: Convergence in the Central Limit Theorem, Far East J. Theor. Stat. 68 (2024), 305–333. https://doi.org/10.17654/0972086324018.