ESAIM: PS
April 2005, Vol. 9, p. 98–115
DOI: 10.1051/ps:2005004

ESAIM: Probability and Statistics

LIMIT THEOREMS FOR U-STATISTICS INDEXED BY A ONE DIMENSIONAL RANDOM WALK

Nadine Guillotin-Plantard[1] and Véronique Ladret[1]

Abstract. Let $(S_n)_{n\geq 0}$ be a $\mathbb{Z}$-random walk and $(\xi_x)_{x\in\mathbb{Z}}$ be a sequence of independent and identically distributed $\mathbb{R}$-valued random variables, independent of the random walk. Let h be a measurable, symmetric function defined on $\mathbb{R}^2$ with values in $\mathbb{R}$. We study the weak convergence of the sequence $\mathcal{U}_n, n \in \mathbb{N}$, with values in $D[0,1]$ the set of right continuous real-valued functions with left limits, defined by

$$\sum_{i,j=0}^{[nt]} h(\xi_{S_i}, \xi_{S_j}), t \in [0,1].$$

Statistical applications are presented, in particular we prove a strong law of large numbers for U-statistics indexed by a one-dimensional random walk using a result of [1].

Mathematics Subject Classification. 60F05, 60J15.

Received August 15, 2004.

1. Introduction

In this paper, we focus on "U-statistics indexed by a one dimensional random walk". Our problem concerns the asymptotic behavior, as $n \to +\infty$, of partial sums of the form:

$$U_n = \sum_{i,j=0}^{n} h(\xi_{S_i}, \xi_{S_j}), n \in \mathbb{N}.$$

Here $h : \mathbb{R}^2 \to \mathbb{R}$ is a measurable, symmetric function, $(X_i)_{i\geq 1}$ is a sequence of centered, i.i.d. $\mathbb{Z}$-valued random variables and for $n \geq 1$, $S_n = X_1 + \ldots + X_n$ is the associated random walk starting at $S_0 = 0$ and $(\xi_x)_{x\in\mathbb{Z}}$ is a sequence of i.i.d. real-valued random variables with probability measure μ, independent of the random walk $(S_n)_{n\geq 0}$. Let Z_α be a one dimensional α-stable random variable with index $1 < \alpha \leq 2$. We assume that the X's belong to the domain of attraction of Z_α, namely $\left(\frac{1}{n^{1/\alpha}} S_n\right)_{n\geq 1}$ converges in distribution to Z_α as $n \to +\infty$. Let $(D[0,1], \mathcal{D})$ denote the set of right continuous real-valued functions with left limits endowed with the Skorohod J_1-topology $\mathcal{D}$. In the case where S_n is a $\mathbb{Z}^m$-valued random walk with $m \geq 2$, Cabus and

Keywords and phrases. Random walk, random scenery, U-statistics, functional limit theorem.

[1] Université Claude Bernard, Lyon 1, 50 av. Tony-Garnier, 69366 Lyon Cedex 07, France; nadine.guillotin@univ-lyon1.fr; veronique.ladret@univ-lyon1.fr

© EDP Sciences, SMAI 2005

Article published by EDP Sciences and available at http://www.edpsciences.org/ps or http://dx.doi.org/10.1051/ps:2005004

Guillotin [6] studied the weak convergence in $(D[0,1],\mathcal{D})$ of the sequence $U_{[nt]}, t\in[0,1]$, as $n\to+\infty$. In this paper we solve the case $m=1$.

"U-statistics indexed by a random walk" appear as a natural extension of widely studied random walks in random scenery, namely, partial sums of the form

$$Z_n=\sum_{k=0}^{n}\xi_{S_k},\ n\geq 0.$$

For $m=1$, Kesten and Spitzer [12] proved that when X and ξ belong to the domains of attraction of different stable laws of indices $1<\alpha\leq 2$ and $0<\beta\leq 2$, respectively, then there exists $\delta>\frac{1}{2}$ such that $\left(n^{-\delta}Z_{[nt]}\right)$ converges weakly as $n\to\infty$ to a self-similar process with stationary increments, δ being related to α and β by $\delta=1-\alpha^{-1}+(\alpha\beta)^{-1}$. The case $0<\alpha<1$ and β arbitrary is easier; they showed then that $\left(n^{-\frac{1}{\beta}}Z_{[nt]}\right)$ converges weakly, as $n\to\infty$, to a stable process with index β. Bolthausen [3] gave a method to solve the more difficult case $\alpha=1$ and $\beta=2$ and especially, he proved that when $(S_n)_{n\in\mathbb{N}}$ is a recurrent $\mathbb{Z}^2$-random walk, $(n\log n)^{-\frac{1}{2}}Z_{[nt]}$ satisfies a functional central limit theorem. For an arbitrary transient random walk, $n^{-\frac{1}{2}}Z_n$ is asymptotically normal (see [19], p. 53). Maejima [14] generalized the result of Kesten and Spitzer [12] in the case where $(\xi_x)_{x\in\mathbb{Z}}$ are i.i.d. $\mathbb{R}^d$-valued random variables which belong to the domain of attraction of an operator stable random vector with exponent B. If we denote by D the linear operator on $\mathbb{R}^d$ defined by $D=(1-\frac{1}{\alpha})I+\frac{1}{\alpha}B$, he proved that $\left(n^{-D}Z_{[nt]}\right)$ converges weakly to an operator self similar with exponent D and having stationary increments.

In general, in the theory of U-statistics (of order 2), the classical object for study is the sequence

$$\frac{1}{n(n-1)}\sum_{1\leq i\neq j\leq n}h(\xi_i,\xi_j)$$

where $(\xi_k)_{k\geq 1}$ is a sequence of independent and identically distributed random variables and we usually assume that

$$E(h(\xi_1,\xi_2))=0\quad\text{ and }\quad E(h^2(\xi_1,\xi_2))<\infty.$$

When $E(h(\xi_1,\xi_2)|\xi_1)=0$, the U-statistic is called *degenerate*. U-statistics were introduced by Hoeffding (1948) who obtained many properties of U-statistics and their generalizations, in particular their asymptotic normality. Definitions, results and applications of U-statistics can be found in Serfling's book [18] and also in the more recent book of Lee [13]. The study of U-statistics permits us to construct statistical tests and to estimate integral from samples $\xi_i, i\geq 1$. Therefore the asymptotic study of the random variable U_n is quite natural and should give us information about the statistic of the sample $(\xi_{S_k})_{0\leq k\leq n}$, *i.e.* observations of a random landscape by a random walker. A way of estimating functionals from a large sample of this sequence of random variables with control of the error will be presented in the last section.

Moreover, let us also mention that the model we consider here should be of some relevance in the study of a polymer represented by the random walk $(S_k)_{k\geq 0}$ evolving in a disordered medium which creates random electrical charges on the polymer. These random electrical charges correspond to the random variables $\xi_i, i\in\mathbb{Z}^m$. We are interested in the total electrical energy $\sum_{i,j}h(\xi_{S_i},\xi_{S_j})$ describing the interactions of the polymer with itself. This physical model is quite realistic as the proteins considered are very long charged molecules and their stereochemical shape is the one minimizing the total electrical energy. The results obtained in this paper constitutes a first step for the study of this problem (see [5, 15]).

In this paper the kernel h is L^4-integrable with respect to the product measure $\mu\otimes\mu$:

$$h\in L^4(\mu\otimes\mu).$$

The first result concerns the case where h is degenerate, *i.e.*

$$E(h(x,\xi_0))=0,\forall x\in\mathbb{R}.$$

Article published by EDP Sciences and available at http://www.edpsciences.org/ps or http://dx.doi.org/10.1051/ps:2005004

Since $h \in L^2(\mu \otimes \mu)$, it induces a Hilbert-Schmidt operator T_h on $L^2(\mu)$:

$$\begin{aligned} T_h : L^2(\mu) &\to L^2(\mu) \\ f &\to T_h f(x) = E(h(\xi_0, x) f(\xi_0)). \end{aligned}$$

From the theory of compact hermitian operators, there exists an orthonormal sequence of eigenfunctions $(\phi_\nu)_{\nu \geq 1}$ and eigenvalues $(\lambda_\nu)_{\nu \geq 1}$ such that

$$h(x,y) = \sum_{\nu=1}^{+\infty} \lambda_\nu \phi_\nu(x) \phi_\nu(y) \qquad \text{in } L^2(\mu \otimes \mu). \tag{1}$$

We notice that since h is degenerate, and $(\phi_\nu)_{\nu \geq 1}$ is an orthonormal sequence of $L^2(\mu)$, we get that, for all $\nu, \sigma \geq 1$,

$$E(\phi_\nu(\xi_0)) = 0, \quad E(\phi_\nu(\xi_0)\phi_\sigma(\xi_0)) = \delta_{\nu,\sigma} \quad \text{where } \delta_{\nu,\sigma} = \begin{cases} 0 \text{ if } \nu \neq \sigma \\ 1 \text{ if } \nu = \sigma. \end{cases} \tag{2}$$

Since $h \in L^4(\mu \otimes \mu)$, it implies, by Cauchy-Schwarz inequality, that for all $\nu \geq 1$,

$$E((\phi_\nu(\xi_0))^4) < \infty. \tag{3}$$

Moreover, let $\tilde{h} : \mathbb{R} \to \mathbb{R}$ be the function defined by $\tilde{h}(x) = h(x,x)$. We shall impose the following conditions on $\tilde{h}$

$$\tilde{h} \in L^4(\mu) \quad \text{and} \quad \tilde{h}(x) = \sum_{\nu=1}^{+\infty} \lambda_\nu {\phi_\nu}^2(x) \quad \text{in } L^1(\mu). \tag{4}$$

Let $\delta = 1 - (2\alpha)^{-1}$ $(\delta > 1/2)$ and let $(Y(t))_{t \in [0,1]}$ denote an α-stable Levy process with right continuous sample paths, such that $Y(1)$ has the same distribution as Z_α. Then there exists a version of the *local time* $L_t(x)$ of Y which is continuous in (t,x) [4, 10]. Let $(B_+^{(i)}(x))_{x \geq 0}$ and $(B_-^{(i)}(x))_{x \geq 0}$, $i \geq 1$, be a pair of sequences of independent Brownian motions, independent of each other and of $(Y_t)_{t \in [0,1]}$. Then $B_\pm^{(i)}(x), i \geq 1$, is also independent of $L_t(x)$. Now, since $B_\pm^{(i)}(x)$ is a semimartingale the following stochastic integrals can be defined as in Kesten and Spitzer [12]

$$\Delta_t^{(i)} = \int_0^\infty L_t(x) \, \mathrm{d}B_+^{(i)}(x) + \int_0^\infty L_t(-x) \, \mathrm{d}B_-^{(i)}(x). \tag{5}$$

In order to simplify the notations, throughout the paper, these processes will be respectively denoted by $(B(x))_{x \in \mathbb{R}}$ and $\Delta_t = \int_{\mathbb{R}} L_t(x) \, \mathrm{d}B(x)$.

For all $i \geq 1$, the process $(\Delta_t^{(i)})_{t \in [0,1]}$ has the following two properties:

(i) $\Delta_t^{(i)}$ has stationary increments.
(ii) $\Delta_t^{(i)}$ is self-similar with index δ.

Moreover, the characteristic function of the process $(\Delta_t^{(i)})_{t \in [0,1]}$ is given by

$$E\left(\mathrm{e}^{i\theta \Delta_t^{(i)}}\right) = E\left(\mathrm{e}^{-\frac{\theta^2}{2} \int_{\mathbb{R}} L_t^{(i)}(x)^2 \, \mathrm{d}x}\right), \ \theta \in \mathbb{R}$$

(see Lem. 5 in [12]).

Article published by EDP Sciences and available at http://www.edpsciences.org/ps or http://dx.doi.org/10.1051/ps:2005004

We are interested in the behavior of $\big(U_{[nt]}\big)_{t\in[0,1]}$ as $n\to+\infty$. The first result is as follows.

Theorem 1.1. *The sequence* $\big(\mathcal{U}_n(t)\big)_{t\in[0,1]} = \big(n^{-2\delta}U_{[nt]}\big)_{t\in[0,1]}$ *converges weakly as* $n\to+\infty$ *in* $\big(D[0,1],\mathcal{D}\big)$ *to the process* $(W_t)_{t\in[0,1]}$ *defined by*

$$W_t = \sum_{i=1}^{\infty} \lambda_i \big(\Delta_t^{(i)}\big)^2.$$

Remarks.
1. We shall see from the proof of the theorem, that the convergence of the finite dimensional distributions still holds if we remove the L^4-integrability condition concerning both h and $\tilde{h}$. It is only used for showing the tightness of the sequence $\mathcal{U}_n$.
2. Let V_n denote the number of self-intersections of the random walk $(S_k)_{k\geq 0}$ up to time n,

$$V_n = \sum_{i,j=0}^{n} 1_{\{S_i=S_j\}}. \tag{6}$$

In the case where $(S_n)_{n\geq 0}$ is a recurrent $\mathbb{Z}^2$-random walk, Cabus and Guillotin [6] showed that if the covariance matrix Q of X_1 is nonsingular, then there exists a sequence $\big((B_t^{(i)})_{t\in[0,1]}\big)_{i\geq 1}$ of independent Brownian motions such that

$$\left(\frac{\pi(\det\, Q)^{1/2}}{n\log n}U_{[nt]}\right)_{t\in[0,1]} \text{ converges weakly in } (D[0,1],\mathcal{D}) \text{ to } \left(\sum_{i=1}^{+\infty}\lambda_i (B_t^{(i)})^2\right)_{t\in[0,1]} \text{ as } n\to+\infty.$$

The proof mainly uses the fact that $V_n/n\log n$ converges almost surely to some constant (see [6]). When S_n is $\mathbb{Z}$-valued, there is no analogous result. That is why the proof of Cabus and Guillotin [6] does not apply in the one-dimensional case.

The following result concerns the case where h is non degenerate. Here we only assume $h\in L^4(\mu\otimes\mu)$ and $\tilde{h}\in L^4(\mu)$, without assuming condition (4) concerning the decomposition of $\tilde{h}$ in $L^1(\mu)$.

Theorem 1.2. *Let* $m = E\big(h(\xi_0,\xi_1)\big)$ *and let* $\sigma^2 = E\big(h(\xi_0,\xi_1)h(\xi_0,\xi_{-1})\big) - m^2$. *If* h *is non degenerate, then*

$$\left(\frac{U_{[nt]} - m[nt]^2}{2t\sigma n^{2-1/2\alpha}}\right)_{t\in[0,1]}$$

converges weakly, as $n\to+\infty$, *in* $\big(D[0,1],\mathcal{D}\big)$ *to* $\big(\Delta_t\big)_{t\in[0,1]}$.

The paper is organized as follows: we first establish some results concerning the asymptotic behavior of the occupation times of a recurrent one dimensional random walk. Then, we prove Theorems 1.1 and 1.2. Finally, in the last section a method to estimate functionals of observations given by a random walker is described. In particular, we give sufficient conditions for which a strong law of large numbers holds for U-statistics indexed by a one-dimensional random walk. We also establish a link between the ergodic properties of the sequence $(\xi_{S_n})_n$ (for instance, Bernoullicity, Weak Bernoullicity, ...) and the existence of a strong law of large numbers for U_n.

2. Properties of occupation times

In order to prove Theorem 1.1, we need some preliminary results concerning the asymptotic behavior of occupation times $N_n(x)$ and self-intersections times V_n defined in (6), as $n\to+\infty$. Here $N_n(x)$ is the number

Article published by EDP Sciences and available at http://www.edpsciences.org/ps or http://dx.doi.org/10.1051/ps:2005004

of visits of the random walk to the point x up to time n:

$$N_n(x) = \sum_{i=0}^{n} 1_{\{S_i = x\}}.$$

Then,

$$V_n = \sum_{x\in\mathbb{Z}} N_n(x)^2.$$

Throughout the paper,

$$Q_n^{(p)} = \sum_{x\in\mathbb{Z}} N_n(x)^p.$$

Kesten and Spitzer [12] proved that, for all $x \in \mathbb{Z}$,

$$E\big(N_n(x)^p\big) = \mathcal{O}\left(n^{p(1-\frac{1}{\alpha})}\right) \tag{7}$$

and that for some $C > 0$,

$$E\big(V_n\big) \sim Cn^{2-\frac{1}{\alpha}}, \quad n \to +\infty.$$

Lemma 2.1. *For all $p \geq 1$ and all $k \geq 1$,*

$$E\big((Q_n^{(p)})^k\big) = \mathcal{O}\left(n^{kp(1-\frac{1}{\alpha})+\frac{k}{\alpha}}\right).$$

In particular ($p = 2$), for all $k \geq 1$,

$$E\big({V_n}^k\big) = \mathcal{O}\big(n^{2k\delta}\big).$$

Proof. Let (Ω, Σ, P) be the probability space on which all the random variables are defined. We denote by $||.||_k$ the L^k-norm on the space $L^k(P)$, *i.e.* $||X||_k = E\big(|X|^k\big)^{\frac{1}{k}}$ for all X in $L^k(P)$.

$$\begin{aligned} Q_n^{(p)} &= \sum_{0\leq i_1,i_2,\ldots,i_p\leq n} 1_{\{S_{i_1}=S_{i_2}=\ldots=S_{i_p}\}} \\ &\leq p! \sum_{0\leq i_1\leq i_2\leq\cdots\leq i_p\leq n} 1_{\{S_{i_1}=S_{i_2}=\ldots=S_{i_p}\}}. \end{aligned}$$

Thus,

$$||Q_n^{(p)}||_k \leq p! \sum_{0\leq i_1\leq n} \left\| \sum_{i_1\leq i_2\leq\cdots\leq i_p\leq n} 1_{\{S_{i_1}=S_{i_2}=\ldots=S_{i_p}\}} \right\|_k. \tag{8}$$

Now, i_1 being fixed,

$$E\left(\left(\sum_{i_1\leq i_2\leq\ldots i_p\leq n} 1_{\{S_{i_1}=S_{i_2}=\ldots=S_{i_p}\}}\right)^k\right) \leq E\left(\sum_{i_1\leq i_2,\ldots,i_{k(p-1)}\leq n} 1_{\{S_{i_1}=S_{i_2}=\ldots=S_{i_{k(p-1)+1}}\}}\right).$$

Here the Markov property applies and it leads to

$$E\left(\left(\sum_{i_1\leq i_2,\ldots,\leq i_p\leq n} 1_{\{S_{i_1}=S_{i_2}=\ldots=S_{i_p}\}}\right)^k\right) \leq E\Big(N_n(0)^{k(p-1)}\Big).$$

Article published by EDP Sciences and available at http://www.edpsciences.org/ps or http://dx.doi.org/10.1051/ps:2005004

It follows from this inequality and (7) that

$$\left\| \sum_{i_1 \le i_2 \le \dots \le i_p \le n} 1_{\{S_{i_1}=S_{i_2}=\ldots=S_{i_p}\}} \right\|_k = \mathcal{O}\Big(n^{(p-1)(1-\frac{1}{\alpha})}\Big). \tag{9}$$

Finally we conclude the proof by combining (8) and (9). □

3. The degenerate case

We first show the convergence of the finite dimensional distributions.

Proposition 3.1. *Any finite dimensional distribution of* $\left(n^{-2\delta}U_{[nt]}\right)_{t\in[0,1]}$ *converges to the finite dimensional distribution of* $\left(W_t\right)_{t\in[0,1]}$.

Before we begin the proof of Proposition 3.1 we need to give details of one of the results of Maejima [14]. Let $\left(\xi_x\right)_{x\in\mathbb{Z}}$ be a collection of i.i.d. $\mathbb{R}^d$-valued random vectors, independent of the random walk $(S_n)_{n\ge0}$. We suppose that the $\left(\xi_x\right)_{x\in\mathbb{Z}}$ belong to the domain of attraction of a Gaussian vector $Z_{1/2}$. Then, Maejima [14] proved that for the linear operator $D = (1-\frac{1}{2\alpha})I$,

$$n^{-D}\sum_{k=0}^{[nt]}\xi_{S_k} \text{ converges in } (D[0,1],\mathcal{D}) \text{ to the limit process } W_t = \int_{-\infty}^{\infty} L_t(x)\ \mathrm{d}B(x)$$

where $L_t(x)$ is the local time at the point $x\in\mathbb{R}$ of an α-stable Levy process $\left(Y(t)\right)_{t\in[0,1]}$ with right continuous sample paths, such that $Y(1)$ has the same distribution as Z_α, and $\left(B(x)\right)_{x\in\mathbb{R}}$ is a $\mathbb{R}^d$-valued Brownian motion independent of $Y(.)$ such that the distribution of $B(1)$ is the same as the one of $Z_{1/2}$. Here

$$\int_{-\infty}^{\infty} L_t(x)\ \mathrm{d}B(x) = \left(\int_{-\infty}^{\infty} L_t(x)\ \mathrm{d}B^{(1)}(x),\ldots,\int_{-\infty}^{\infty} L_t(x)\ \mathrm{d}B^{(d)}(x)\right)$$

where, for each $i=1,\ldots,d$, the real valued process $B^{(i)}(x)$ denotes the i-th component of $B(x)$. Each stochastic integral can be rigorously defined as (5).

Now we are ready for the proof of Proposition 3.1.

Proof. Here and later in the proof we fix $\theta=(\theta_1,\ldots,\theta_m)\in\mathbb{R}^m$ and $0\le t_1<\ldots<t_m\le1$. First we choose some arbitrary integer $K\ge1$ and we define the kernel $h_K(x,y)$ by

$$h_K(x,y) = \sum_{\nu=1}^{K}\lambda_\nu\phi_\nu(x)\phi_\nu(y). \tag{10}$$

We denote by $U_{n,K}$ the corresponding U-statistic indexed by $\left(S_n\right)_{n\ge0}$, *i.e.*

$$U_{n,K} = \sum_{i,j=0}^{n} h_K(\xi_{S_i},\xi_{S_j}).$$

Then we set

$$\mathcal{U}_{n,K}(t) = U_{[nt],K}.$$

Article published by EDP Sciences and available at http://www.edpsciences.org/ps or http://dx.doi.org/10.1051/ps:2005004

Step 1. From the definition of h_K (10), we derive the following expression

$$\mathcal{U}_{n,K}(t) = \sum_{\nu=1}^{K} \lambda_\nu \left(\frac{1}{n^\delta} \sum_{i=1}^{[nt]} \phi_\nu(\xi_{S_i}) \right)^2. \tag{11}$$

For all $x \in \mathbb{Z}$, we define the random vector $\widetilde{\xi}_x^{(K)}$ by

$$\widetilde{\xi}_x^{(K)} = \big(\phi_1(\xi_x), \ldots, \phi_K(\xi_x)\big) \in \mathbb{R}^K.$$

For each $\nu \in \{1, \ldots, K\}$, the sequence $(\phi_\nu(\xi_x))_{x\in\mathbb{Z}}$ is a sequence of i.i.d. random variables. It follows from (2) that, for each $x \in \mathbb{Z}$, the components of $\widetilde{\xi}_x^{(K)}$ are uncorrelated random variables with expectation 0 and variance 1. Hence the central limit theorem applies, *i.e.*

$$\frac{1}{\sqrt{n}} \sum_{x=1}^{n} \widetilde{\xi}_x^{(K)} \xrightarrow{\mathcal{L}} Z_{1/2},$$

where $Z_{1/2}$ is a K-dimensional standard Gaussian vector.

For all $n \geq 1$, let

$$\mathcal{Z}_{n,K}(t) = \sum_{i=0}^{[nt]} \widetilde{\xi}_{S_i}^{(K)}.$$

Here, we apply the result of Maejima [14], *i.e.*

$$\big(\mathcal{Z}_{n,K}(t)\big)_{t\in[0,1]} \text{ converges weakly in } D[0,1], \text{ as } n \to +\infty, \text{ to } \big(\Delta_t^{(1)}, \ldots, \Delta_t^{(K)}\big)_{t\in[0,1]}, \tag{12}$$

with

$$\Delta_t^{(i)} = \int_{-\infty}^{\infty} L_t(x)\, \mathrm{d}B^{(i)}(x), \quad i = 1, \ldots, K,$$

where $L_t(x)$ is the local time at $x \in \mathbb{R}$ of the process $Y(.)$ defined above and $B(x) = (B^{(1)}(x), \ldots, B^{(K)}(x))$, $x \in \mathbb{R}$ is a K-dimensional Brownian motion, independent of $Y(.)$, with independent components.

Now, let

$$L : \mathbb{R}^K \to \mathbb{R}$$
$$(x_1, \ldots, x_K) \mapsto \sum_{\nu=1}^{K} \lambda_\nu x_\nu^2.$$

Then, by (11),

$$\mathcal{U}_{n,K} = L(\mathcal{Z}_{n,K}). \tag{13}$$

Since $\mathcal{Z}_{n,K}(.)$ converges weakly to $\big(\Delta_t^{(1)}, \ldots, \Delta_t^{(K)}\big)_{t\in[0,1]}$ as n goes to infinity and L is continuous, it follows that, for all integer $K \geq 1$,

$$\big(\mathcal{U}_{n,K}(t)\big)_{t\in[0,1]} \text{ converges in } (D[0,1], \mathcal{D}) \text{ to } \left(W_K(t) := \sum_{\nu=1}^{K} \lambda_\nu \big(\Delta_t^{(\nu)}\big)^2 \right)_{t\in[0,1]}, \text{ as } n \to +\infty.$$

Hence, the finite dimensional laws of $\mathcal{U}_{n,K}(.)$ converge to the finite dimensional laws of $W_K(.)$. If we denote by ϕ_K and $\phi_{n,K}$ the characteristic functions of the random vectors $(W_K(t_1), \ldots, W_K(t_m))$ and $(\mathcal{U}_{n,K}(t_1), \ldots, \mathcal{U}_{n,K}(t_m))$

Article published by EDP Sciences and available at http://www.edpsciences.org/ps or http://dx.doi.org/10.1051/ps:2005004

respectively, then for every $\theta \in \mathbb{R}^m$,

$$|\phi_{n,K}(\theta) - \phi_K(\theta)| \to 0, \; n \to +\infty. \tag{14}$$

Step 2. Next we turn to the analysis of $R_{n,K}$ defined by

$$\begin{aligned} R_{n,K}(t) &= U_{[nt]} - U_{[nt],K} \\ &= \sum_{x,y\in\mathbb{Z}} N_{[nt]}(x)N_{[nt]}(y)\big(h - h_K\big)(\xi_x, \xi_y) \\ &= R^{(1)}_{n,K}(t) + R^{(2)}_{n,K}(t) \end{aligned}$$

with

$$R^{(1)}_{n,K}(t) = \sum_{x\neq y} N_{[nt]}(x)N_{[nt]}(y)\big(h - h_K\big)(\xi_x, \xi_y)$$

and

$$R^{(2)}_{n,K}(t) = \sum_{x\in\mathbb{Z}} N^2_{[nt]}(x)\big(h - h_K\big)(\xi_x, \xi_x).$$

Let $\tilde{h}_K(x) = h_K(x, x)$, then

$$R^{(2)}_{n,K}(t) = \sum_{x\in\mathbb{Z}} N^2_{[nt]}(x)\big(\tilde{h} - \tilde{h}_K\big)(\xi_x).$$

First we focus on $E\Big(|R^{(1)}_{n,K}(t)|^2\Big)$. From the independence of $\big(S_n\big)_{n\geq 0}$ and $\big(\xi_x\big)_{x\in\mathbb{Z}}$ and the fact that h is degenerate, we get that

$$\begin{aligned} E\Big(|R^{(1)}_{n,K}(t)|^2\Big) &= \sum_{x\neq y} E\Big(N^2_{[nt]}(x)N^2_{[nt]}(y)\Big)||h - h_K||^2_{L^2(\mu\otimes\mu)} \\ &\leq E\big(V^2_{[nt]}\big)||h - h_K||^2_{L^2(\mu\otimes\mu)}. \end{aligned} \tag{15}$$

Then it follows from Lemma 2.1 that there exists a constant $C_1 > 0$ such that for all $n \geq 1$ and $t \in [0, 1]$,

$$E\left(\left|\frac{R^{(1)}_{n,K}(t)}{n^{2\delta}}\right|^2\right) \leq C_1||h - h_K||^2_{L^2(\mu\otimes\mu)}. \tag{16}$$

Now, we focus on $R^{(2)}_{n,K}(t)$.

$$E\Big(\big|R^{(2)}_{n,K}(t)\big|\Big) \leq E\left(\sum_x N^2_{[nt]}(x)\big|\big(\tilde{h} - \tilde{h}_K\big)(\xi_x)\big|\right).$$

The independence of $(S_n)_{n\geq 0}$ and $(\xi_x)_{x\in\mathbb{Z}}$ implies that

$$E\big(|R^{(2)}_{n,K}(t)|\big) \leq E\big(V_{[nt]}\big)||\tilde{h} - \tilde{h}_K||_{L^1(\mu)}.$$

Thus, from Lemma 2.1 it follows that there exists a constant $C_2 > 0$ such that for all $n \geq 1$ and $t \in [0, 1]$,

$$E\left(\left|\frac{R^{(2)}_{n,K}(t)}{n^{2\delta}}\right|\right) \leq C_2||\tilde{h} - \tilde{h}_K||_{L^1(\mu)}. \tag{17}$$

Article published by EDP Sciences and available at http://www.edpsciences.org/ps or http://dx.doi.org/10.1051/ps:2005004

Put $C = \max(\sqrt{C_1}, C_2)$ and set

$$\mathcal{R}_{n,K}(t) = R_{n,K}(t).$$

By virtue of (16) and (17), we have that for all $n \geq 1$ and for all $t \in [0,1]$,

$$E\Big(|\mathcal{R}_{n,K}(t)|\Big) \leq C\Big(||h - h_K||_{L^2(\mu\otimes\mu)} + ||\tilde{h} - \tilde{h}_K||_{L^1(\mu)}\Big). \tag{18}$$

Now set

$$\Theta_n = \sum_{j=1}^m \theta_j \mathcal{U}_n(t_j), \quad \Theta_{n,K} = \sum_{j=1}^m \theta_j \mathcal{U}_{n,K}(t_j).$$

Let ϕ_n be the characteristic function of the random vector $(\mathcal{U}_n(t_1), \ldots, \mathcal{U}_n(t_m))$.

Then, for every $\theta \in \mathbb{R}^m$,

$$\begin{aligned} |\phi_n(\theta) - \phi_{n,K}(\theta)| &= |E(\mathrm{e}^{i\Theta_n} - \mathrm{e}^{i\Theta_{n,K}})| \\ &\leq E|\mathrm{e}^{i(\Theta_n - \Theta_{n,K})} - 1| \\ &\leq E|\Theta_n - \Theta_{n,K}| \\ &= E|\sum_{j=1}^m \theta_j \mathcal{R}_{n,K}(t_j)|. \end{aligned}$$

By (18) we have that for all integer $n \geq 1$,

$$|\phi_n(\theta) - \phi_{n,K}(\theta)| \leq C \sum_{j=1}^m |\theta_j| \left(||h - h_K||_{L^2(\mu\otimes\mu)} + ||\tilde{h} - \tilde{h}_K||_{L^1(\mu)}\right).$$

Hence, for all $\varepsilon > 0$ there exists a $K_0 > 0$ such that for any $K > K_0$,

$$\sup_{n\geq 1} |\phi_n(\theta) - \phi_{n,K}(\theta)| < \varepsilon. \tag{19}$$

Step 3. Let

$$W_t = \sum_{\nu=1}^{+\infty} \lambda_\nu \big(\Delta_t^{(\nu)}\big)^2.$$

We denote by ϕ the characteristic function of the random vector $(W_{t_1}, \ldots, W_{t_m})$.

Let us first prove that $E(\Delta_t^4) < \infty$. By Burkholder-Davis-Gundy's inequality (see [17]),

$$\begin{aligned} E\left(\left(\int_0^{+\infty} L_1(x)\ \mathrm{d}B_+(x)\right)^4\right) &\leq C\ E\left(\left\langle \int_0^{+\infty} L_1(x)\ \mathrm{d}B_+(x), \int_0^{+\infty} L_1(x)\ \mathrm{d}B_+(x)\right\rangle^2\right) \\ &\leq C\ E\left(\left(\int_0^{\infty} L_1^2(x)\ \mathrm{d}x\right)^2\right) \\ &\leq C\ E\left(\left(\int_{-\infty}^{\infty} L_1^2(x)\ \mathrm{d}x\right)^2\right), \end{aligned}$$

for some constant $C > 0$. The same inequality holds for $E\Big(\big(\int_0^{+\infty} L_1(-x)\ \mathrm{d}B_-(x)\big)^4\Big)$. From Kesten and Spitzer [12], we know that the sequence $\frac{V_n}{n^{2\delta}}$ weakly converges to $\int_{\mathbb{R}} L_1(x)^2\ \mathrm{d}x$, hence, $\left(\frac{V_n}{n^{2\delta}}\right)^2$ weakly converges

Article published by EDP Sciences and available at http://www.edpsciences.org/ps or http://dx.doi.org/10.1051/ps:2005004

to $\left(\int_{\mathbb{R}} L_1(x)^2 \, \mathrm{d}x\right)^2$,

$$\left(\frac{V_n}{n^{2\delta}}\right)^2 \xrightarrow{\mathcal{L}} \left(\int_{\mathbb{R}} L_1^2(x) \, \mathrm{d}x\right)^2. \tag{20}$$

Moreover, by Lemma 2.1,

$$\sup_{n\geq 1} E\left(\left(\frac{V_n}{n^{2\delta}}\right)^4\right) \leq C,$$

where $C > 0$ denotes a constant.

So, the sequence $\left(V_n^2/n^{4\delta}\right)_{n\geq 1}$ is uniformly integrable and Theorem 25.12 of Billingsley [2] applies:

$$E\left(\left(\int_{\mathbb{R}} L_1^2(x)\mathrm{d}x\right)^2\right) \leq C.$$

The self-similarity of Δ ($\Delta_t \stackrel{d}{=} t^\delta \Delta_1$) implies that for all $t \in [0,1]$,

$$E(\Delta_t^4) = t^{4\delta} E(\Delta_1^4) \leq C < \infty.$$

Now, we come back to the convergence of the finite dimensional distributions of $W_{K,t}$ to the finite dimensional distributions of W_t.

We compute the following L^2-norm:

$$\Big|\Big|\sum_{j=1}^m \theta_j (W_{t_j} - W_{K,t_j})\Big|\Big|_2 \leq \sum_{j=1}^m |\theta_j|.\Big|\Big|W_{t_j} - W_{K,t_j}\Big|\Big|_2. \tag{21}$$

From the Cauchy-Schwarz inequality, it follows that

$$\begin{aligned}\Big|\Big|W_{t_j} - W_{K,t_j}\Big|\Big|_2^2 &= \sum_{\nu=K+1}^{\infty} \lambda_\nu^2 E\big((\Delta_{t_j}^{(\nu)})^4\big) + \sum_{\nu\neq\mu=K+1}^{\infty} \lambda_\nu\lambda_\mu E\big((\Delta_{t_j}^{(\nu)})^2(\Delta_{t_j}^{(\mu)})^2\big) \\ &\leq {t_j}^{4\delta} E\big((\Delta_1)^4\big)\left(\sum_{\nu=K+1}^{\infty} \lambda_\nu^2 + \left(\sum_{\nu=K+1}^{\infty}\lambda_\nu\right)^2\right) \\ &\leq C\left(\sum_{\nu=K+1}^{\infty} \lambda_\nu^2 + \left(\sum_{\nu=K+1}^{\infty}\lambda_\nu\right)^2\right).\end{aligned}$$

From the theory of compact operators, we know that, since $h \in L^2(\mu\otimes\mu)$,

$$\sum_{\nu=1}^{\infty} \lambda_\nu^2 = E(h(\xi_0,\xi_1)).$$

Moreover, the condition (4) implies that

$$\left|\sum_{\nu=1}^{\infty}\lambda_\nu\right| < +\infty, \quad \text{with} \quad \sum_{\nu=1}^{\infty}\lambda_\nu = E(\tilde{h}(\xi_0)).$$

Article published by EDP Sciences and available at http://www.edpsciences.org/ps or http://dx.doi.org/10.1051/ps:2005004

Hence, the convergence of the series $\sum_{\nu=1}^{\infty} \lambda_\nu^2$ and $\sum_{\nu=1}^{\infty} \lambda_\nu$ implies the convergence in L^2 of the sequence $\sum_{j=1}^{m} \theta_j W_{K,t_j}$ to $\sum_{j=1}^{m} \theta_j W_{t_j}$, as K goes to infinity. Thus,

$$|\phi(\theta) - \phi_K(\theta)| \to 0, \qquad K \to +\infty. \tag{22}$$

Finally, by (19) and (22), given $\varepsilon > 0$ we can choose K_1 large enough so that

$$\sup_{n\geq 1} |\phi_n(\theta) - \phi_{n,K_1}(\theta)| < \varepsilon/3 \quad \text{and} \quad |\phi(\theta) - \phi_{K_1}(\theta)| < \varepsilon/3.$$

By (14), it follows that we can find N such that for all $n > N$

$$|\phi_{K_1}(\theta) - \phi_{n,K_1}(\theta)| < \varepsilon/3.$$

Then, for all $n > N$,

$$|\phi(\theta) - \phi_n(\theta)| \leq |\phi(\theta) - \phi_{K_1}(\theta)| + |\phi_{K_1}(\theta) - \phi_{n,K_1}(\theta)| + |\phi_n(\theta) - \phi_{n,K_1}(\theta)| < \varepsilon$$

proving the convergence of the finite dimensional laws. □

We still have to show the tightness of the family $\big(\mathcal{U}_n(t)\big)_{t\in[0,1]}$. By Theorem 15.6 of [2] it is enough to prove the following

Proposition 3.2. *There exists a strictly positive constant C and a constant $\gamma = 4 - 2/\alpha > 1$ such that for every $0 \leq t_1 < t < t_2 \leq 1$,*

$$E\Big(\big(\mathcal{U}_{[nt_2]} - \mathcal{U}_{[nt]}\big)^2 \big(\mathcal{U}_{[nt]} - \mathcal{U}_{[nt_1]}\big)^2\Big) \leq C\big(t_2 - t_1\big)^\gamma.$$

We first fix some notations. In what follows, for all s, t in $[0,1]$ and x, y in $\mathbb{Z}$,

$$D_{s,t}(x,y) = N_{[nt]}(x)N_{[nt]}(y) - N_{[ns]}(x)N_{[ns]}(y) \tag{23}$$

and

$$A_{s,t}(x) = N_{[nt]}(x) - N_{[ns]}(x), \quad A_t(x) = A_{0,t}(x) = N_{[nt]}(x). \tag{24}$$

Proof of Proposition 3.2. The sequences $(S_n)_{n\geq 0}$ and $(\xi_x)_{x\in\mathbb{Z}}$ are independent, thus

$$\begin{aligned} E\big((\mathcal{U}_{[nt_2]} - \mathcal{U}_{[nt]})^2(\mathcal{U}_{[nt]} - \mathcal{U}_{[nt_1]})^2\big) &= \frac{1}{n^{8\delta}} \sum_{(x_i,y_i)_{1\leq i\leq 4}} E\left(\prod_{i=1}^{2} D_{t,t_2}(x_i,y_i) \prod_{i=3}^{4} D_{t_1,t}(x_i,y_i)\right) \\ &\quad \times E\left(\prod_{i=1}^{4} h(\xi_{x_i}, \xi_{y_i})\right) \\ &\leq \frac{1}{n^{8\delta}} \sum_{(x_i,y_i)_{1\leq i\leq 4}} E\left(\prod_{i=1}^{4} D_{t_1,t_2}(x_i,y_i)\right) E\left(\prod_{i=1}^{4} h(\xi_{x_i}, \xi_{y_i})\right). \end{aligned} \tag{25}$$

Put

$$\alpha(x_1, y_1, \ldots, x_4, y_4) = E\left(\prod_{i=1}^{4} h(\xi_{x_i}, \xi_{y_i})\right)$$

Article published by EDP Sciences and available at http://www.edpsciences.org/ps or http://dx.doi.org/10.1051/ps:2005004

and denote by $\mathcal{S}$ the support of α,

$$\mathcal{S} = \big\{(x_1,\dots,y_4) \mid \alpha(x_1,\dots,y_4) = 0\big\}. \qquad \square$$

Notations. In order to simplify the notation, we shall denote by both $(x_i,y_i)_{1\le i\le 4}$ and $(x_1,y_1)(x_2,y_2)(x_3,y_3)(x_4,y_4)$ the 8-tuple $\big(x_1,y_1,\dots,x_4,y_4\big)$. For all $k\in\{2,3,4\}$, $(x,y)^k$ will stand for $\underbrace{(x,y)\cdots(x,y)}_{k\text{ times}}$.

The fact that h is degenerate implies that α vanishes on the 8-tuples $(x_1,\dots,y_4)$ in which one coordinate appears only once. In consequence, all the elements of $\mathcal{S}$ contain each coordinate at least twice. Now we focus on the following subsets of $\mathcal{S}$:
for all $i\in\{3,\dots,8\}$,

$$\mathcal{S}_i = \big\{(x_j,y_j)_{1\le j\le 4}\in\mathcal{S} \mid \text{ one of the coordinates appears exactly } i \text{ times}\big\},$$

and

$$\mathcal{S}_2 = \big\{(x_i,y_i)_{1\le i\le 4}\in\mathcal{S} \mid \text{ all the coordinates appear exactly twice}\big\}.$$

We notice that $\mathcal{S} = \cup_{i=2}^{8}\mathcal{S}_i$. Here, as above, the argument of the degeneracy of h applies. And, even if it means reordering the coordinates, it leads to

$$\mathcal{S}_8 = \big\{(x_1,x_1)^4\big\},$$

$$\mathcal{S}_2 = \big\{(x_1,x_1)(x_2,x_2)(x_3,x_3)(x_4,x_4)\big\}\cup\big\{(x_1,x_1)(x_2,x_2)(x_3,y_3)^2\big\}\cup\big\{(x_1,y_1)^2(x_2,y_2)^2\big\}, \tag{26}$$

$$\mathcal{S}_3 = \mathcal{S}_5 = \big\{(x_1,y_1)^3(x_4,x_4) \text{ with } x_1\notin\{x_4,y_1\}\big\}, \tag{27}$$

$$\begin{aligned}\mathcal{S}_4 = \big\{&(x_1,x_1)^2(x_2,x_2)(x_3,x_3),\ (x_1,x_1)^2(x_2,x_3)^2,\ (x_1,x_1)(x_1,x_2)^2(x_3,x_3),\ (x_1,x_2)^2(x_1,x_3)^2,\\ &\text{with } x_1\notin\{x_2,x_3\}\big\},\end{aligned} \tag{28}$$

$$\mathcal{S}_6 = \big\{(x_1,x_1)^3(x_2,x_2) \text{ with } x_1\ne x_2\big\}\cup\big\{(x_1,x_1)^2(x_1,x_2)^2 \text{ with } x_1\ne x_2\big\}. \tag{29}$$

We put

$$C = \max_{\{(x_i,y_i)_{1\,\le\, i\,\le\, 4}\in\mathcal{S}\}} E\left(\prod_{i=1}^{4} h(\xi_{x_i},\xi_{y_i})\right).$$

Since $h\in L^4(\mu\otimes\mu)$ and $\tilde h\in L^4(\mu)$, C is finite.

Then, from (25), it follows that

$$E\Big(\big(\mathcal{U}_{[nt_2]}-\mathcal{U}_{[nt]}\big)^2\big(\mathcal{U}_{[nt]}-\mathcal{U}_{[nt_1]}\big)^2\Big) \le \frac{C}{n^{8\delta}}\sum_{(x_i,y_i)_{i\le 4}\in\mathcal{S}} E\left(\prod_{i=1}^{4} D_{t_1,t_2}(x_i,y_i)\right). \tag{30}$$

Now we still have to compute the terms

$$\sum_{(x_i,y_i)_{i\le 4}\in\mathcal{S}_k} E\left(\prod_{i=1}^{4} D_{t_1,t_2}(x_i,y_i)\right),\ k=2,3,\dots,8.$$

As the calculations of these quantities rely on the same basic arguments, we shall concentrate on the case where $k=3$.

Article published by EDP Sciences and available at http://www.edpsciences.org/ps or http://dx.doi.org/10.1051/ps:2005004

By the Cauchy-Schwarz inequality, by (27) and the fact that D_{t_1,t_2} is non-negative, we get that

$$\sum_{(x_i,y_i)_{i\le 4}\in\mathcal{S}_3} E\left(\prod_{i=1}^{4} D_{t_1,t_2}(x_i,y_i)\right) \le E\left(\sum_{x_1,y_1} D^3_{t_1,t_2}(x_1,y_1)\sum_{x_4,y_4} D_{t_1,t_2}(x_4,y_4)\right)$$

$$\le E\left(\left(\sum_{x_1,y_1} D^3_{t_1,t_2}(x_1,y_1)\right)^2\right)^{1/2} E\left(\left(\sum_{x_4,y_4} D_{t_1,t_2}(x_4,y_4)\right)^2\right)^{1/2}. \quad (31)$$

First we focus on the second term equal to $E\Big(\Big(\sum_{x_4,y_4} D_{t_1,t_2}(x_4,y_4)\Big)^2\Big)^{1/2}$.

$$\sum_{x_4,y_4} D_{t_1,t_2}(x_4,y_4) = \left(\sum_{x_4} N_{[nt_2]}(x_4)\right)^2 - \left(\sum_{x_4} N_{[nt_1]}(x_4)\right)^2 = [nt_2]^2 - [nt_1]^2. \quad (32)$$

Since $0 \le t_1 < t_2 \le 1$,

$$[nt_2]^2 - [nt_1]^2 \le 2n([nt_2] - [nt_1]).$$

If $1 \le n(t_2-t_1)$, then $[nt_2]-[nt_1] \le n(t_2-t_1)+1 \le 2n(t_2-t_1)$. Otherwise $[nt_2]-[nt_1] = 0$. Therefore,

$$[nt_2]^2 - [nt_1]^2 \le 4n^2(t_2-t_1). \quad (33)$$

From (33), it follows that

$$E\left(\left(\sum_{x_4,y_4} D_{t_1,t_2}(x_4,y_4)\right)^2\right)^{1/2} \le 4n^2(t_2-t_1). \quad (34)$$

Now we focus on the integral

$$E\left(\left(\sum_{x_1,y_1} D^3_{t_1,t_2}(x_1,y_1)\right)^2\right)^{1/2}. \quad (35)$$

We recall from the definitions (23) and (24) of D and A that

$$D_{t_1,t_2}(x,y) = A_{t_1,t_2}(x)A_{t_1,t_2}(y) + A_{t_1}(x)A_{t_1,t_2}(y) + A_{t_1}(y)A_{t_1,t_2}(x).$$

In order to estimate (35) we shall focus on the L^2-norm (denoted by $||.||_2$) of the following terms

$$\alpha_1 = \sum_{x,y} A^3_{t_1,t_2}(x)A^3_{t_1,t_2}(y),\ \alpha_2 = \sum_{x,y} A^3_{t_1}(x)A^3_{t_1,t_2}(y),\ \alpha_3 = \sum_{x,y} A_{t_1}(x)A^2_{t_1,t_2}(x)A^3_{t_1,t_2}(y),$$

$$\alpha_4 = \sum_{x,y} A^2_{t_1}(x)A_{t_1,t_2}(x)A^3_{t_1,t_2}(y),\ \alpha_5 = \sum_{x,y} A^2_{t_1}(x)A_{t_1,t_2}(x)A_{t_1}(y)A^2_{t_1,t_2}(y).$$

We first evaluate $||\alpha_1||_2$:

$$||\alpha_1||_2 = \Big|\Big|\Big(\sum_x (N_{[nt_2]}(x) - N_{[nt_1]}(x))^3\Big)^2\Big|\Big|_2.$$

It follows from the Markov property that

$$\sum_x (N_{[nt_2]}(x) - N_{[nt_1]}(x))^3$$

Article published by EDP Sciences and available at http://www.edpsciences.org/ps or http://dx.doi.org/10.1051/ps:2005004

has the same distribution as $Q^{(3)}_{[nt_2]-[nt_1]}$. Thus, from Lemma 2.1,

$$||\alpha_1||_2 = \left|\left|(Q^{(3)}_{[nt_2]-[nt_1]})^2\right|\right|_2 = \mathcal{O}\Big((n(t_2-t_1))^{6-4/\alpha}\Big). \tag{36}$$

Now we focus on $||\alpha_2||_2$:

$$||\alpha_2||_2 = \left(E\left(\sum_x N^3_{[nt_1]}(x)\sum_y A^3_{t_1,t_2}(y)\right)^2\right)^{1/2}.$$

Here the Markov property applies again: $\sum_x N^3_{[nt_1]}(x)$ and $\sum_y A^3_{t_1,t_2}(y)$ are independent and distributed as $Q^{(3)}_{[nt_1]}$ and $Q^{(3)}_{[nt_2]-[nt_1]}$ respectively. Thus,

$$\begin{aligned} ||\alpha_2||_2 &= ||Q^{(3)}_{[nt_1]}||_2 \; ||Q^{(3)}_{[nt_2]-[nt_1]}||_2 \\ &= \mathcal{O}\Big(n^{6-4/\alpha}(t_2-t_1)^{3-2/\alpha}\Big). \end{aligned} \tag{37}$$

In the calculation of the square of $||\alpha_3||_2$, the Markov property leads to

$$\begin{aligned} ||\alpha_3||_2^2 &= E\left(\sum_{x,y} A_{t_1}(x)A^2_{t_1,t_2}(x)A^3_{t_1,t_2}(y)\right)^2 \\ &= \sum_{x_1,y_1,x_2,y_2} E\Big(A_{t_1}(x_1)A_{t_2}(x_2)\Big)E\Big(A^2_{t_1,t_2}(x_1)A^2_{t_1,t_2}(x_2)A^3_{t_1,t_2}(y_1)A^3_{t_1,t_2}(y_2)\Big). \end{aligned} \tag{38}$$

By the Cauchy-Schwarz inequality and (7) it follows that for all x_1 and x_2 in $\mathbb{Z}$,

$$E\Big(A_{t_1}(x_1)A_{t_1}(x_2)\Big) = \mathcal{O}\big(n^{2-2/\alpha}\big). \tag{39}$$

Then, the Cauchy-Schwarz inequality still applies in the estimate of the left term

$$\begin{aligned} E\left(\left(\sum_{x_1,y_1} A^2_{t_1,t_2}(x_1)A^3_{t_1,t_2}(y_1)\right)^2\right) &= E\left(\left(\sum_{x_1} A^2_{t_1,t_2}(x_1)\right)^2\left(\sum_{y_1} A^3_{t_1,t_2}(y_1)\right)^2\right) \\ &\le ||Q^{(2)}_{t_1,t_2}||_4^2 \; ||Q^{(3)}_{t_1,t_2}||_4^2. \end{aligned} \tag{40}$$

Now, from (38), (39), (40) and Lemma 2.1, it follows that

$$||\alpha_3||_2 = \mathcal{O}\Big(n^{6-4/\alpha}\big(t_2-t_1\big)^{5-3/\alpha}\Big). \tag{41}$$

The same basic ideas apply in the computation of the two left terms $||\alpha_4||_2$ and $||\alpha_5||_2$. Thus it is easy to see that

$$||\alpha_4||_2 = \mathcal{O}\left(n^{6-4/\alpha}(t_2-t_1)^{4-2/\alpha}\right), \quad ||\alpha_5||_2 = \mathcal{O}\left(n^{6-4/\alpha}(t_2-t_1)^{3-1/\alpha}\right). \tag{42}$$

Thus we derive, from (36), (37), (41), (42) that

$$\left(E\left(\sum_{x,y} D^3_{t_1,t_2}(x,y)\right)^2\right)^{1/2} = \mathcal{O}\Big(n^{6-4/\alpha}\big(t_2-t_1\big)^{3-2/\alpha}\Big). \tag{43}$$

Article published by EDP Sciences and available at http://www.edpsciences.org/ps or http://dx.doi.org/10.1051/ps:2005004

Finally, using (43) and (34), (31) reads

$$\sum_{(x_i,y_i)_{i\le 4}\in\mathcal{S}_3} E\left(\prod_{i=1}^{4} D_{t_1,t_2}(x_i,y_i)\right) = \mathcal{O}\Big(n^{8\delta}(t_2-t_1)^{4-2/\alpha}\Big).$$

With the same techniques it is possible to compute the terms in the case where S_3 is replaced by S_i with $i = 2, 4, \ldots, 8$ and it can be checked that

$$\sum_{(x_i,y_i)_{i\le 4}\in\mathcal{S}_k} E\left(\prod_{i=1}^{4} D_{t_1,t_2}(x_i,y_i)\right) = \mathcal{O}\Big(n^{8\delta}\big(t_2-t_1\big)^{4-2/\alpha}\Big),\quad k=2,\ldots,6 \tag{44}$$

and

$$\sum_{(x_i,y_i)_{i\le 4}\in\mathcal{S}_8} E\left(\prod_{i=1}^{4} D_{t_1,t_2}(x_i,y_i)\right) = \mathcal{O}\Big(n^{8\delta-2/\alpha}(t_2-t_1)^{4-2/\alpha}\Big). \tag{45}$$

Thus, if we put together (30), (44) and (45), it follows that there exist a positive constant C and a $\gamma = 4-2/\alpha$, $\gamma > 1$, such that, for every $0 \le t_1 \le t_2 \le 1$

$$E\Big(\big(\mathcal{U}_{[nt_2]} - \mathcal{U}_{[nt]}\big)^2\big(\mathcal{U}_{[nt]} - \mathcal{U}_{[nt_1]}\big)^2\Big) \le C\big(t_2-t_1\big)^{\gamma}.$$

4. The non-degenerate case

From Hoeffding's decomposition, we have, for $x \neq y$,

$$h(\xi_x,\xi_y) = m + (g(\xi_x)-m) + (g(\xi_y)-m) + \phi(\xi_x,\xi_y)$$

where $g(\xi_x) = E(h(\xi_x,\xi_y)|\xi_x)$ and ϕ is a degenerate kernel.

Using this decomposition, we get:

$$\begin{aligned} U_{[nt]} &= m([nt]+1)^2 - mV_{[nt]} + \sum_{x\in\mathbb{Z}} N_{[nt]}(x)^2(h(\xi_x,\xi_x) - 2(g(\xi_x)-m)) \\ &\quad +2([nt]+1)\sum_{x\in\mathbb{Z}} N_{[nt]}(x)(g(\xi_x)-m) + \sum_{x\neq y} N_{[nt]}(x)N_{[nt]}(y)\phi(\xi_x,\xi_y) \\ &= m([nt]+1)^2 - mV_{[nt]} + \sum_{x\in\mathbb{Z}} N_{[nt]}(x)^2(h(\xi_x,\xi_x) - 2(g(\xi_x)-m)) \\ &\quad +2([nt]+1)\sum_{k=0}^{[nt]}(g(\xi_{S_k})-m) + \sum_{x\neq y} N_{[nt]}(x)N_{[nt]}(y)\phi(\xi_x,\xi_y). \end{aligned}$$

Now, from Lemma 2.1,

$$E(V_n) = \mathcal{O}(n^{2-\frac{1}{\alpha}})$$

so $V_n/n^{2-1/2\alpha}$ converges in probability to 0 as $n\to+\infty$:

$$\frac{V_n}{n^{2-1/2\alpha}} \xrightarrow{P} 0.$$

Article published by EDP Sciences and available at http://www.edpsciences.org/ps or http://dx.doi.org/10.1051/ps:2005004

Using the results of the previous section, it is easy to prove that

$$\frac{1}{n^{2-\frac{1}{2\alpha}}}\left[\sum_{x\neq y} N_{[nt]}(x)N_{[nt]}(y)\phi(\xi_x,\xi_y)+\sum_{x\in\mathbb{Z}} N_{[nt]}(x)^2(h(\xi_x,\xi_x)-2(g(\xi_x)-m))\right]$$

converges in probability to 0 as $n\to\infty$.

Finally, from Kesten and Spitzer's Theorem (see [12]),

$$\left(\frac{1}{\sigma n^{1-\frac{1}{2\alpha}}}\sum_{k=0}^{[nt]}(g(\xi_{S_k})-m)\right)_{t\in[0,1]} \xrightarrow{\mathcal{D}} (\Delta_t)_{t\in[0,1]},$$

and the theorem follows.

5. Statistical estimation of m

In this section we give some techniques to estimate functionals of the random scenery viewed by a random walker. Consider the recurrent one-dimensional random walk $(S_n)_{n\geq 0}$ and the random scenery $(\xi_x)_{x\in\mathbb{Z}}$ as defined in the introduction. The random walk is also assumed to be strongly aperiodic (see the definition in [19]). Our problem is to estimate $m = E(h(\xi_0,\xi_1))$ from observations $(\xi_{S_k})_{0\leq k\leq n}$ up to time n given by a random walker. The distribution μ of the ξ's is unknown as well as the paths of the random walker which are assumed to belong to a domain of attraction of a stable law with index α. The hypotheses on the function h are the ones given in the introduction. When h is non-degenerate, thanks to Theorem 1.2, the sequence of random variables

$$\frac{1}{n^2}\sum_{1\leq i,j\leq n} h(\xi_{S_i},\xi_{S_j})$$

converges in probability to m as $n\to+\infty$. Moreover, Theorem 1.2 allows us to give a confidence interval of our estimation since

$$\frac{U_n - mn^2}{2\sigma n^{2-1/2\alpha}}\to\Delta_1$$

where $\sigma^2 = E\big(h(\xi_0,\xi_1)h(\xi_0,\xi_{-1})\big)-m^2$. The expectation $E\big(h(\xi_0,\xi_1)h(\xi_0,\xi_{-1})\big)$ can be estimated by a realization for large n of the sequence of random variables

$$\frac{1}{n^3}\sum_{i,j,k=1}^{n} h(\xi_{S_i},\xi_{S_j})h(\xi_{S_j},\xi_{S_k}).$$

The distribution of the random variable $\Delta_1 = \int_{\mathbb{R}} L_1(x)\,\mathrm{d}B(x)$ can be simulated, for instance by using Kesten and Spitzer's theorem.

The next natural question is to know if the parameter m can be almost surely approximated by the U-statistic sampled by the one-dimensional random walk. Due to the strong correlations of the sequence of observations, the strong law of large numbers for U-statistics of i.i.d. observations due to Hoeffding [11] is clearly not applicable. Aaronson *et al.* [1] proved a strong law of large numbers for U-statistics for ergodic stationary processes (see Theorem U in [1]). Let us extend our random walk in the following natural way: let $(X_i)_{i\in\mathbb{Z}}$ be a sequence of centered, i.i.d. $\mathbb{Z}$-valued random variables with distribution ρ. The associated $\mathbb{Z}$-random walk $(S_n)_{n\in\mathbb{Z}}$ is now defined as

$$S_0 = 0,\ S_n = \sum_{i=1}^{n} X_i \text{ for } n\geq 1$$

Article published by EDP Sciences and available at http://www.edpsciences.org/ps or http://dx.doi.org/10.1051/ps:2005004

and

$$S_n = -\sum_{i=n+1}^{0} X_i \text{ for } n \leq -1.$$

The sequence $(S_n)_{n\in\mathbb{Z}}$ is assumed strongly aperiodic. Let us define the transformation

$$\begin{aligned} U : \Omega \times \Omega' &\to \Omega \times \Omega' \\ (\omega, \omega') &\to (\sigma\omega, T^{\omega_1}\omega') \end{aligned}$$

where σ is the shift operator on the space $\Omega = \mathbb{Z}^{\mathbb{Z}}$ (the random walk) with the product measure ρ^∞ and T is the shift operator on the space $\Omega' = \mathbb{R}^{\mathbb{Z}}$ (the random scenery) with the product measure μ^∞. The transformation U preserves the measure $\rho \otimes \mu$ and is a K-automorphism (see Meilijson [16]). For every $n \in \mathbb{Z}$, each random variable ξ_{S_n} can be written as $f \circ U^n$ where f is the composition of two (measurable) projections, so $(\xi_{S_n})_{n\in\mathbb{Z}}$ is a stationary and ergodic sequence. We can now apply Aaronson's Theorem U as follows: Assume that the function $h : \mathbb{R}^2 \to \mathbb{R}$ is measurable, symmetric and bounded by an μ-integrable product, *i.e.* $|h(x_1, x_2)| \leq f_1(x_1) \times f_2(x_2)$ for every $(x_1, x_2) \in \mathbb{R}^2$ for some $f_1, f_2 \in L^1(\mu)$.

If any of the following three conditions hold:

- μ is a discrete law;
- The function h is continuous at $\mu \otimes \mu$-almost every point;
- $(\xi_{S_n})_n$ is weakly Bernoulli;

then

$$\lim_{n\to+\infty} \frac{1}{n(n-1)} \sum_{1\leq i\neq j\leq n} h(\xi_{S_i}, \xi_{S_j}) = m.$$

The weak Bernoullicity of the process $(X_n, \xi_{S_n}), n \in \mathbb{Z}$ has been studied in great details in a collection of nice papers [7–9] when the ξ's take their values in a finite set. In particular, in [8], a necessary and sufficient condition is given in order to get the weak Bernoullicity of this process. This condition is strongly related to the intersection properties of the random walk. As it is mentioned in Section 8 of [9], the weak (or *not*) Bernoullicity of the second coordinate of $(X_n, \xi_{S_n}), n \in \mathbb{Z}$ can not be deduced from the fact that $(X_n, \xi_{S_n}), n \in \mathbb{Z}$ is (or *is not*) weak Bernoulli and to our knowledge, it is still an open problem. One way of proving it would be to find a function h satisfying all the conditions of Theorem U from [1] but for which the strong law of large numbers for the U-statistics indexed by the one dimensional random walk does not hold. Necessarily, this would imply that the sequence $\xi_{S_n}, n \in \mathbb{Z}$ is not weakly Bernoulli.

6. Conclusion

We have proved a functional limit theorem for U-statistics indexed by a one-dimensional $\mathbb{Z}$-random walk and solved the conjecture stated in [6]. We have also introduced statistics of a sequence of random variables indexed by a one-dimensional random walk. More results on limit theorems for U-statistics indexed by a one-dimensional random walk are expected as well as statistical applications. The case of a kernel h of order k with $k \geq 3$ is presently under investigation.

References

[1] J. Aaronson, R. Burton, H. Dehling, D. Gilat, T. Hill and B. Weiss, Strong laws for L- and U-statistics. *Trans. Amer. Math. Soc.* **348** (1996) 2845–2866.

[2] P. Billingsley, *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition. A Wiley-Interscience Publication (1999).

[3] E. Bolthausen, A central limit theorem for two-dimensional random walks in random sceneries. *Ann. Probab.* **17** (1989) 108–115.

[4] E. Boylan, Local times for a class of Markoff processes. *Illinois J. Math.* **8** (1964) 19–39.

[5] E. Buffet and J.V. Pulé, A model of continuous polymers with random charges. *J. Math. Phys.* **38** (1997) 5143–5152.

Article published by EDP Sciences and available at http://www.edpsciences.org/ps or http://dx.doi.org/10.1051/ps:2005004

[6] P. Cabus and N. Guillotin-Plantard, Functional limit theorems for U-statistics indexed by a random walk. *Stochastic Process. Appl.* **101** (2002) 143–160.

[7] F. den Hollander, Mixing properties for random walk in random scenery. *Ann. Probab.* **16** (1988) 1788–1802.

[8] F. den Hollander, M.S. Keane, J. Serafin and J.E. Steif, Weak bernoullicity of random walk in random scenery. *Japan. J. Math. (N.S.)* **29** (2003) 389–406.

[9] F. den Hollander and J.E. Steif, Mixing properties of the generalized T, T^{-1}-process. *J. Anal. Math.* **72** (1997) 165–202.

[10] R.K. Getoor and H. Kesten, Continuity of local times for Markov processes. *Comp. Math.* **24** (1972) 277–303.

[11] W. Hoeffding, The strong law of large numbers for U-statistics. *Univ. N. Carolina, Institue of Stat. Mimeo series* **302** (1961).

[12] H. Kesten and F. Spitzer, A limit theorem related to a new class of self-similar processes. *Z. Wahrsch. Verw. Gebiete* **50** (1979) 5–25.

[13] A.J. Lee, *U-statistics. Theory and practice.* Marcel Dekker, Inc., New York (1990).

[14] M. Maejima, Limit theorems related to a class of operator-self-similar processes. *Nagoya Math. J.* **142** (1996) 161–181.

[15] S. Martínez and D. Petritis, Thermodynamics of a Brownian bridge polymer model in a random environment. *J. Phys. A* **29** (1996) 1267–1279.

[16] I. Meilijson, Mixing properties of a class of skew-products. *Israel J. Math.* **19** (1974) 266–270.

[17] D. Revuz and M. Yor, *Continuous martingales and Brownian motion*. Springer-Verlag, Berlin. Fundamental Principles of Mathematical Sciences **293** (1999).

[18] R.J. Serfling, *Approximation theorems of mathematical statistics*. John Wiley & Sons Inc., New York. Wiley Series in Probability and Mathematical Statistics (1980).

[19] F. Spitzer, *Principles of random walks*. Springer-Verlag, New York, second edition. Graduate Texts in Mathematics **34** (1976).

Article published by EDP Sciences and available at http://www.edpsciences.org/ps or http://dx.doi.org/10.1051/ps:2005004