

## Stable limit theorem for $U$ -statistic processes indexed by a random walk

Brice Franke\*      Françoise Pène†      Martin Wendler‡

### Abstract

Let  $(S_n)_{n \in \mathbb{N}}$  be a  $\mathbb{Z}$ -valued random walk with increments from the domain of attraction of some  $\alpha$ -stable law and let  $(\xi(i))_{i \in \mathbb{Z}}$  be a sequence of iid random variables. We want to investigate  $U$ -statistics indexed by the random walk  $S_n$ , that is  $U_n := \sum_{1 \leq i < j \leq n} h(\xi(S_i), \xi(S_j))$  for some symmetric bivariate function  $h$ . We will prove the weak convergence without assumption of finite variance. Additionally, under the assumption of finite moments of order greater than two, we will establish a law of the iterated logarithm for the  $U$ -statistic  $U_n$ .

**Keywords:** random walk; random scenery;  $U$ -statistics; stable limits; law of the iterated logarithm.

**AMS MSC 2010:** 60F05; 60F17; 60G50; 60K37.

Submitted to ECP on March 9, 2015, final version accepted on March 16, 2016.

## 1 Introduction

Random walks in random scenery were introduced by Kesten and Spitzer [14]. They studied the partial sum process  $\sum_{k=1}^n \xi(S_k)$ , where  $S_k := \sum_{m=1}^k X_m$  is a random walk based on some sequence of  $\mathbb{Z}$ -valued i.i.d. random variables  $(X_m)_{m \in \mathbb{N}}$  and  $(\xi(i))_{i \in \mathbb{Z}}$  is a sequence of real valued i.i.d. random variables which are supposed to be independent of the random walk  $(S_k)_{k \in \mathbb{N}}$ . The law of the random variable  $X_1$  is supposed to belong to the normal domain of attraction of an  $\alpha$ -stable law  $F_\alpha$  with  $0 < \alpha \leq 2$ , i.e.: one has

$$P\left(n^{-\frac{1}{\alpha}} S_n \leq x\right) \rightarrow F_\alpha(x).$$

It is then well known that the sequence of stochastic processes

$$S_t^{(n)} := n^{-\frac{1}{\alpha}} S_{[nt]}; \quad t \geq 0, n \in \mathbb{N}$$

converges in distribution towards an  $\alpha$ -stable Lévy process  $S_t^*$  (see Skorokhod [20], Theorem 2.7). It is further assumed that the random walk  $(S_n)_{n \in \mathbb{N}}$  is irreducible and strongly aperiodic. For the case  $\alpha > 1$ , Kesten and Spitzer [14] showed that if the scenery variable  $\xi(1)$  is in the normal domain of attraction of some  $\beta$ -stable law  $F_\beta$  with  $1 < \beta \leq 2$ , then the partial sum process converges after some suitable renormalization toward some

---

\*Université de Brest, LMBA, UMR CNRS 6205, 29238 Brest cedex, France. E-mail: brice.franke@univ-brest.fr

†Université de Brest and IUF, LMBA, UMR CNRS 6205, 29238 Brest cedex, France. E-mail: francoise.pene@univ-brest.fr

‡Ernst-Moritz-Arndt-Universität, 17487 Greifswald, Germany. E-mail: martin.wendler@uni-greifswald.de

specific continuous self-similar process  $(\Delta_t)_{t \geq 0}$  with stationary increments. Since the random walk  $(S_k)_{k \in \mathbb{N}}$  can visit the same location several times, the sequence of random variables  $(\xi(S_k))_{k \in \mathbb{N}}$  shows some long range dependence. For the limit process  $(\Delta_t)_{t \geq 0}$  this imposes some non-classical scaling index and also some non-trivial dependence of the stationary increments. The construction of the process  $(\Delta_t)_{t \geq 0}$  is given after Theorem 2.1. The case  $\alpha \leq 1$  was studied extensively in [2, 9, 4].

A natural and widely applicable generalization of partial sums are *U*-statistics. In this paper we want to investigate the asymptotic behavior of *U*-statistics indexed by some random walk  $(S_k)_{k \in \mathbb{N}}$ , which are defined as follows:

$$U_n := \sum_{1 \leq i < j \leq n} h(\xi(S_i), \xi(S_j)).$$

In what follows we will assume that  $h$  is a bivariate, measurable and symmetric function such that

$$E[|h(\xi(1), \xi(2))|] < \infty \text{ and } E[|h(\xi(1), \xi(1))|] < \infty.$$

We further assume that  $Eh(\xi(1), \xi(2)) = 0$ . The random walk  $(S_k)_{k \in \mathbb{N}}$  is supposed to fulfill the same assumptions as in the paper of Kesten and Spitzer [14] described above.

A classic approach for *U*-statistics in the case of finite second moments is the Hoeffding decomposition [13]. We can write

$$U_n = (n - 1) \sum_{i=1}^n h_1(\xi(S_i)) + \sum_{1 \leq i < j \leq n} h_2(\xi(S_i), \xi(S_j))$$

with

$$h_1(x) := E(h(x, \xi(1)))$$

$$h_2(x, y) := h(x, y) - h_1(x) - h_1(y).$$

We call  $L_n := \sum_{i=1}^n h_1(\xi(S_i))$  the linear part of the *U*-statistic  $U_n$  and  $R_n$  the remainder term:

$$R_n = \sum_{1 \leq i < j \leq n} h_2(\xi(S_i), \xi(S_j)).$$

The idea of using the Hoeffding decomposition in the present context is based on the following crucial argument. Under second moment assumptions, the family of random variables  $(h_2(\xi_x, \xi_y))_{x \neq y}$  is pairwise uncorrelated. This leads to the following expression of the conditional variance of the remainder term, when the random walk is given

$$\text{Var} [R_n | (S_k)_{k \in \mathbb{N}}] = \sum_{x, y \in \mathbb{Z}} \text{Var}[h_2(\xi_x, \xi_y)] (\#\{(i, j) | i, j \leq n, i \neq j, S_i = x, S_j = y\})^2.$$

By using moment bounds for the occupation times of the random walk  $(S_n)_n$ , that is the number of indices  $i \leq n$  such that  $S_i = x$  (for  $x \in \mathbb{Z}$ ), the asymptotic domination of the linear part  $(n - 1)L_n$  will follow. In the particular case when  $S_k = k$ , our process  $U_n$  corresponds to the classical *U*-statistic with independent entries  $\xi_i$ . There under second moment assumptions on  $\xi$  one has  $\text{Var}[R_n] = O(n^2)$  while  $\text{Var}[(n - 1)L_n]$  is of order  $n^3$ . It turns out that for the random walk in random scenery, considered here, the Hoeffding decomposition is helpful in spite of the dependence.

The *U*-statistic indexed by a random walk was examined by Cabus and Guillotin-Plantard [3] and Guillotin-Plantard and Ladret [11], but only in the case of finite fourth moments. We will extend their results to the case when  $h_1(\xi_1)$  is in the normal domain of attraction of a stable distribution. However, we then have to overcome another obstacle

which results from the lack of second moments, which was crucial in the above argument. To deal with this problem, we will adapt a truncation method Heinrich and Wolf [12] introduced to study  $U$ -statistics with stable limits under independent random variables.

Note that for other models of long range dependence (e.g. Gaussian sequences with slowly decaying covariances), both the linear part and the remainder term might contribute to the limit distribution, see Beutner and Zähle [1]. Because of this, other methods, like representing the  $U$ -statistics as a functional of the empirical distribution function, are appropriate, see Dehling and Taqqu [8].

Another approach to study  $U$ -statistics with heavy tails was presented by Dabrowski, Dehling, Mikosch, Sharipov [7]. They use a method based on functional convergence of suitably defined point processes to handle the existence of second moments. This alternative approach was used in Franke, Pène and Wendler [10] to prove distributional convergence of  $U$ -statistics indexed by a random walk. There it is proved that under some different assumptions on the scenery variables and also under some different scaling the  $U$ -statistics process converges toward some stochastic integral with respect to some Lévy sheet.

We complete our study with a law of the iterated logarithm for the  $U$ -statistic process indexed by  $S_n$ , extending results from Lewis [17] and Zhang [21] for the partial sum indexed by a random walk.

## 2 Main results

Our first theorem will establish the weak convergence of the  $U$ -statistic process without assuming that the summands of the linear part have second (or even higher) moments. More precisely, we will assume that the law  $\mathcal{L}(h_1(\xi(1)))$  is in the normal domain of attraction of a  $\beta$ -stable law  $F_\beta$  with  $1 < \beta \leq 2$

For  $1 < \alpha \leq 2$  or  $\beta = 2$ , the random walk in random scenery converges to a continuous limit (see [14, 2, 9]), even if the scenery contains jumps, so we define the continuous version of the  $U$ -statistics process

$$U_n(t) = U_k \quad \text{if } nt = k \in \mathbb{N}$$

and linear interpolated in between. We will prove weak convergence in the space of continuous functions  $C[0, 1]$  equipped with the supremum norm. For  $\alpha \leq 1$  and  $\beta < 2$ , the limit process has a discontinuous limit (see Castell, Guillotin-Plantard, Pène [4]), so we will consider the space of càdlàg-functions  $D[0, 1]$  endowed with the Skorohod  $M_1$ -topology (see Skorokhod [19]).

**Theorem 2.1.** *Let  $(S_n)_{n \in \mathbb{Z}}$  be a two-sided random walk such that  $S_0 = 0$  and with identically distributed increments  $X_n = S_n - S_{n-1}$ . Assume that*

$$E[|h(\xi(1), \xi(2))|] < \infty \quad \text{and} \quad E[|h(\xi(1), \xi(1))|] < \infty,$$

*and that the law  $\mathcal{L}(h_1(\xi(1)))$  is in the normal domain of attraction of a  $\beta$ -stable law  $F_\beta$  with  $1 < \beta \leq 2$ . Assume furthermore that there exists  $\beta' > \beta$  such that  $E|h(\xi(1), \xi(2))|^\eta < \infty$  with  $\eta = \frac{2\beta'}{1+\beta'}$ . Assume that the law  $\mathcal{L}(X_1)$  is in the normal domain of attraction of an  $\alpha$ -stable law  $F_\alpha$ .*

- *If  $1 < \alpha \leq 2$ , then we have the weak convergence in  $C[0, 1]$*

$$\left( n^{-2+\frac{1}{\alpha}-\frac{1}{\alpha\beta}} U_n(t) \right)_{t \in [0,1]} \Rightarrow (\Delta_t)_{t \in [0,1]},$$

*with  $\Delta_t$  as defined in Kesten and Spitzer [14].*

- If  $\alpha = 1$  and if  $\frac{1}{n} \sum_{i=1}^n X_i$  converges in distribution to  $aZ$  for a Cauchy-distributed random variable  $Z$ , then we have the following weak convergence in the sense of the finite dimensional distributions

$$\left( n^{-1-\frac{1}{\beta}} (\log n)^{\frac{1-\beta}{\beta}} U_{[nt]} \right)_{t \in [0,1]} \Rightarrow \left( \frac{(\Gamma(\beta + 1))^{\frac{1}{\beta}}}{(a\pi)^{\frac{\beta-1}{\beta}}} Y_t \right)_{t \in [0,1]},$$

where  $Y$  is the  $\beta$ -stable Lévy-Process which is the limit of  $n^{-1/\beta} \sum_{i=1}^{[nt]} \xi_i$ . This weak convergence also holds in  $(D[0, 1], M_1)$  if  $1 < \beta < 2$  and in  $(D[0, 1], J_1)$  if  $\beta = 2$ .

- If  $0 < \alpha < 1$ , then we have the weak convergence in the sense of the finite dimensional distributions

$$\left( n^{-1-\frac{1}{\beta}} U_{[nt]} \right)_{t \in [0,1]} \Rightarrow (bY_t)_{t \in [0,1]},$$

where  $Y$  is the  $\beta$ -stable Lévy-Process which is the limit of  $n^{-1/\beta} \sum_{i=1}^{[nt]} \xi_i$  and with  $b = \left( E \left| \sum_{i \in \mathbb{Z}} \mathbf{1}_{\{S_i=0\}} \right|^{\beta-1} \right)^{1/\beta}$ . This weak convergence also holds in  $(D[0, 1], M_1)$  if  $1 < \beta < 2$  and in  $(D[0, 1], J_1)$  if  $\beta = 2$ .

Observe that we can choose  $\beta'$  such that  $\eta < \beta$ , that means that the summands  $h$  of the  $U$ -statistic might have less moments than  $h_1$ , so without loss of generality, we can assume that  $E|h_1(\xi(1))|^\eta < \infty$ .

To give the definition of the process  $(\Delta_t)_{t \in [0,1]}$ , we have to introduce some notation. Let  $(T_t(x))_{t \geq 0}$  be the local time of the limit process  $(S_t^*)_{t \geq 0}$  of the rescaled partial sum  $(n^{-\frac{1}{\alpha}} \sum_{i=1}^{[nt]} X_i)_{t \geq 0}$ , that means

$$\int_0^t \mathbf{1}_{[a,b)}(S_s^*) ds = \int_a^b T_t(x) dx$$

almost surely. Let  $(Z_+(t))_{t \geq 0}$  and  $(Z_-(t))_{t \geq 0}$  be two independent copies of the limit process of the rescaled partial sum process  $(n^{-\frac{1}{\beta}} \sum_{i=1}^{[nt]} h_1(\xi(i)))_{t \geq 0}$ . Then the limit process of the random walk in random scenery is defined as

$$\Delta_t = \int_0^\infty T_t(x) dZ_+(x) + \int_0^\infty T_t(-x) dZ_-(x).$$

For random walks in random scenery, Lewis [17] and Khoshnevisan and Lewis [15] proved the law of the iterated logarithm. This was improved by Csáki, König, Shi [6] and Zhang [21] using strong approximation methods. In our second theorem, we will extend these results to  $U$ -statistics:

**Theorem 2.2.** *Let the assumption of Theorem 1 hold with  $\alpha = \beta = 2$  and additional  $E|h_1(\xi(i))|^p < \infty$  and  $E|X_i|^p < \infty$  for some  $p > 2$ . Then*

$$\limsup_{n \rightarrow \infty} \frac{U_n}{n^{\frac{7}{4}} (\log \log n)^{\frac{3}{4}}} = \frac{2^{\frac{1}{4}} \text{Var}(h_1(\xi(1)))^{\frac{1}{2}}}{3 \text{Var}(X)^{\frac{1}{4}}}$$

$$\liminf_{n \rightarrow \infty} \frac{U_n}{n^{\frac{7}{4}} (\log \log n)^{\frac{3}{4}}} = -\frac{2^{\frac{1}{4}} \text{Var}(h_1(\xi(1)))^{\frac{1}{2}}}{3 \text{Var}(X)^{\frac{1}{4}}}$$

almost surely.

### 3 Auxiliary results

We define the occupation times  $N_n(x) := \sum_{i=1}^n \mathbb{1}_{\{S_i=x\}}$ . Let us write  $V_n := \sum_{x \in \mathbb{Z}} N_n^2(x)$ . Observe that

$$V_n = \sum_{x \in \mathbb{Z}} \left( \sum_{k=1}^n \mathbb{1}_{\{S_k=x\}} \right)^2 = \sum_{x \in \mathbb{Z}} \sum_{k,l=1}^n \mathbb{1}_{\{S_k = S_l = x\}} = \sum_{k,l=1}^n \mathbb{1}_{\{S_k=S_l\}}$$

equals  $n$  plus the number of self intersections of the random walk.

**Proposition 3.1.** *If  $0 < \alpha \leq 2$ , then*

$$\begin{aligned} E[V_n] &= O(n^{2-\frac{1}{\alpha'}} \log n), \\ E[V_n^2] &= O(n^{4-\frac{2}{\alpha'}} \log^2 n), \end{aligned}$$

with  $\alpha' := \max(1, \alpha)$ .

For the case  $1 < \alpha \leq 2$ , this follows from Lemma 2.1 of Guillotin-Plantard, Ladrat [11]. For the case  $\alpha = 1$  and more precise results, we refer to Deligiannidis, Utev [9]. The bound for the case  $0 < \alpha < 1$  comes for example from the proof of Lemma 19 in Castell, Guillotin-Plantard, Pène, Schapira [5] in which it is proven that there exists  $\gamma > 0$  such that  $\sup_n \mathbb{E}[\exp(\gamma V_n/n)] < \infty$ .

**Proposition 3.2.** *Under the conditions of Theorem 1 for  $0 < \alpha \leq 2$ , we have that*

$$\max_{k \leq n} R_k = o(n^{2-\frac{1}{\alpha'} + \frac{1}{\alpha'\beta}})$$

almost surely with  $\alpha' := \max(1, \alpha)$ .

*Proof.* We define  $a_l = 2^{l \frac{1+\beta'}{\alpha'\beta'}}$  and the truncated kernel

$$h_{0,l}(x, y) := \begin{cases} h(x, y) & \text{if } |h(x, y)| \leq a_l \\ 0 & \text{else} \end{cases}.$$

We also need the Hoeffding decomposition for the truncated kernel:

$$\begin{aligned} h_{1,l}(x) &:= E(h_{0,l}(x, \xi(1))) \\ h_{2,l}(x, y) &:= h_{0,l}(x, y) - h_{1,l}(x) - h_{1,l}(y). \end{aligned}$$

We introduce the following notation:

$$\begin{aligned} \tilde{L}_{l,n} &:= \sum_{i=1}^n h_{1,l}(\xi(S_i)) \\ \tilde{U}_{l,n} &:= \sum_{1 \leq i < j \leq n} h_{0,l}(\xi(S_i), \xi(S_j)) \\ \tilde{R}_{l,n} &:= \sum_{1 \leq i < j \leq n} h_{2,l}(\xi(S_i), \xi(S_j)). \end{aligned}$$

Recall the Hoeffding decomposition

$$U_n = (n-1)L_n + R_n.$$

Similar, we have that

$$\tilde{U}_{l,n} = (n-1)\tilde{L}_{l,n} + \tilde{R}_{l,n}.$$

We now obtain the following representation for the remainder term:

$$\begin{aligned} R_n &= U_n - (n-1)L_n = (U_n - \tilde{U}_{l,n}) - (n-1)L_n + \tilde{U}_{l,n} \\ &= (U_n - \tilde{U}_{l,n}) - (n-1)L_n + (n-1)\tilde{L}_{l,n} + \tilde{R}_{l,n} = (U_n - \tilde{U}_{l,n}) - (n-1)(L_n - \tilde{L}_{l,n}) + \tilde{R}_{l,n} \end{aligned}$$

We will treat the three summands separately. To do this we have to show

- (1)  $\max_{n \leq 2^l} |U_n - \tilde{U}_{l,n}| = o(2^{l(2 - \frac{1}{\alpha'} + \frac{1}{\alpha'\beta})})$  almost surely,
- (2)  $\max_{n \leq 2^l} |L_n - \tilde{L}_{l,n}| = o(2^{l(2 - \frac{1}{\alpha'} + \frac{1}{\alpha'\beta})})$  almost surely,
- (3)  $\max_{n \leq 2^l} |\tilde{R}_{l,n}| = o(2^{l(2 - \frac{1}{\alpha'} + \frac{1}{\alpha'\beta})})$  almost surely.

In the proof of (1), we have to deal with the problem that we might have  $S_i = S_j$  for  $i \neq j$ . For this reason we will treat separately the cases  $S_i = S_j$  and  $S_i \neq S_j$ :

$$|U_n - \tilde{U}_{l,n}| \leq \left| \sum_{\substack{1 \leq i < j \leq n \\ S_i \neq S_j}} (h(\xi(S_i), \xi(S_j)) - h_{0,l}(\xi(S_i), \xi(S_j))) \right| + \left| \sum_{\substack{1 \leq i < j \leq n \\ S_i = S_j}} (h(\xi(S_i), \xi(S_j)) - h_{0,l}(\xi(S_i), \xi(S_j))) \right| = |A_{l,n}| + |B_{l,n}|.$$

In order to establish bounds for the maximum, we have to control the increments of  $A_{l,n}$ . Let  $n_1, n_2 \in \mathbb{N}$  with  $n_1 \leq n_2 \leq 2^l$ , then

$$A_{l,n_2} - A_{l,n_1} = \sum_{\substack{1 \leq i < j \leq n \\ n_1 < j \leq n_2 \\ S_i \neq S_j}} (h(\xi(S_i), \xi(S_j)) - h_{0,l}(\xi(S_i), \xi(S_j))),$$

so we have at most  $2^l(n_2 - n_1)$  summands of  $A_{l,n_2} - A_{l,n_1}$  and for every summand

$$\begin{aligned} E|h(\xi(S_i), \xi(S_j)) - h_{0,l}(\xi(S_i), \xi(S_j))| &= E|h(\xi(1), \xi(2))\mathbb{1}_{\{|h(\xi(1), \xi(2))| > a_l\}}| \\ &\leq a_l^{1-\eta} E|h(\xi(1), \xi(2))\mathbb{1}_{\{|h(\xi(1), \xi(2))| > a_l\}}|^\eta \\ &\leq a_l^{1-\eta} E|h(\xi(1), \xi(2))|^\eta. \end{aligned}$$

Consequently, we have by the triangular inequality that

$$\begin{aligned} E|A_{l,n_2} - A_{l,n_1}| &\leq 2^l(n_2 - n_1) E|h(\xi(1), \xi(2))\mathbb{1}_{\{|h(\xi(1), \xi(2))| > a_l\}}| \\ &\leq 2^l(n_2 - n_1) a_l^{1-\eta} E|h(\xi(1), \xi(2))|^\eta \end{aligned}$$

Fix some  $\theta$  such that  $1 < \theta < 1 + \frac{1}{\alpha'\beta} - \frac{1}{\alpha'\beta'}$  (such a  $\theta$  exists since  $\beta' > \beta$ ). Observe that

$$E|A_{l,n_2} - A_{l,n_1}| \leq C_0 2^l(n_2 - n_1)^\theta a_l^{1-\eta}$$

with the constant  $C_0 = E|h(\xi(1), \xi(2))|^\eta$ . We can write  $A_{l,n} = \sum_{i=1}^n (A_{l,i} - A_{l,i-1})$  (with  $A_{l,0} := 0$ ) and in the same way  $A_{l,n_2} - A_{l,n_1} = \sum_{i=n_1+1}^{n_2} (A_{l,i} - A_{l,i-1})$ , so due to the maximal inequality given in Theorem 1 of Móricz [18] (applied with  $\gamma = 1$ ,  $\theta$  instead of  $\alpha$  and  $(g(F_{b,n}))^\theta = C_0 2^l n^\theta a_l^{1-\eta}$ ), we obtain

$$E \left| \max_{n \leq 2^l} A_{l,n} \right| \leq C_0 2^{l(1+\theta)} a_l^{1-\eta}.$$

Recall that  $a_l = 2^{\frac{l(1+\beta')}{\alpha'\beta'}}$  and  $\eta = \frac{2\beta'}{1+\beta'}$ , so  $1 - \eta = \frac{1-\beta'}{1+\beta'}$ . It follows from the Markov inequality that

$$\begin{aligned} \sum_{l=1}^{\infty} P \left( \frac{1}{2^{l(2 - \frac{1}{\alpha'} + \frac{1}{\alpha'\beta})}} \max_{n \leq 2^l} A_{l,n} \geq \epsilon \right) &\leq \frac{C_0}{\epsilon} \sum_{l=1}^{\infty} \frac{2^{l(1+\theta)} a_l^{1-\eta}}{2^{l(2 - \frac{1}{\alpha'} + \frac{1}{\alpha'\beta})}} \\ &= \frac{C_0}{\epsilon} \sum_{l=1}^{\infty} \frac{2^{l(\frac{1+\beta'}{\alpha'\beta'} - \frac{1-\beta'}{1+\beta'})}}{2^{l(1-\theta - \frac{1}{\alpha'} + \frac{1}{\alpha'\beta})}} = \frac{C_0}{\epsilon} \sum_{l=1}^{\infty} 2^{l(\frac{1}{\alpha'\beta'} - \frac{1}{\alpha'\beta} + \theta - 1)} < \infty, \end{aligned}$$

as  $\frac{1}{\alpha'\beta'} - \frac{1}{\alpha'\beta} + \theta - 1 < 0$ . With the Borel-Cantelli lemma, we can now conclude that

$$P\left(\frac{1}{2^{l(2-\frac{1}{\alpha'}+\frac{1}{\alpha'\beta})}} \max_{n \leq 2^l} A_{l,n} \geq \epsilon \text{ infinitely often}\right) = 0$$

and thus  $\max_{n \leq 2^l} A_{l,n} = o(2^{l(2-\frac{1}{\alpha'}+\frac{1}{\alpha'\beta})})$  almost surely. For  $B_{l,n}$ , we use the fact that the sequences  $(S_n)_{n \in \mathbb{N}}$  and  $(\xi(n))_{n \in \mathbb{N}}$  are independent and observe that

$$\begin{aligned} E|\max_{n \leq 2^l} B_{l,n}| &\leq E \max_{n \leq 2^l} \sum_{\substack{1 \leq i < j \leq n \\ S_i = S_j}} |(h(\xi(S_i), \xi(S_j)) - h_{0,l}(\xi(S_i), \xi(S_j)))| \\ &\leq E \sum_{\substack{1 \leq i < j \leq 2^l \\ S_i = S_j}} |(h(\xi(S_i), \xi(S_j)) - h_{0,l}(\xi(S_i), \xi(S_j)))| \\ &\leq E \#\{(i, j) : 1 \leq i < j \leq 2^l | S_i = S_j\} E|h(\xi(1), \xi(1))\mathbb{1}_{\{|h(\xi(1), \xi(1))| > a_l\}}| \\ &\leq E \left| \sum_{x \in \mathbb{Z}} N_{2^l}^2(x) \right| E|h(\xi(1), \xi(1))| \leq C 2^{l(2-\frac{1}{\alpha'})} \end{aligned}$$

for some constant  $C$ , where we used Proposition 3.1 for the occupation times  $N_n(x) := \sum_{i=1}^n \mathbb{1}_{\{S_i=x\}}$ . Again using the Markov inequality, we arrive at

$$\sum_{l=1}^{\infty} P\left(\frac{1}{2^{l(2-\frac{1}{\alpha'}+\frac{1}{\alpha'\beta})}} \max_{n \leq 2^l} B_{l,n} \geq \epsilon\right) \leq \frac{C}{\epsilon} \sum_{l=1}^{\infty} \frac{2^{l(2-\frac{1}{\alpha'})} l}{2^{l(2-\frac{1}{\alpha'}+\frac{1}{\alpha'\beta})}} \leq \frac{C}{\epsilon} \sum_{l=1}^{\infty} 2^{-\frac{l}{\alpha'\beta}} l < \infty$$

and, as above, the Borel-Cantelli lemma leads to  $\max_{n \leq 2^l} B_{l,n} = o(2^{l(2-\frac{1}{\alpha'}+\frac{1}{\alpha'\beta})})$  almost surely, which completes the proof of (1). To prove (2), note that

$$\begin{aligned} E|h_1(\xi(1)) - h_{1,l}(\xi(1))| &= E|E[h(\xi(1), \xi(2))|\xi(1)] - E[h(\xi(1), \xi(2))\mathbb{1}_{\{|h(\xi(1), \xi(2))| \leq a_l\}}|\xi(1)]| \\ &= E|E[h(\xi(1), \xi(2))\mathbb{1}_{\{|h(\xi(1), \xi(2))| > a_l\}}|\xi(1)]| \\ &\leq E[E[|h(\xi(1), \xi(2))\mathbb{1}_{\{|h(\xi(1), \xi(2))| > a_l\}}| |\xi(1)]] \\ &= E|h(\xi(1), \xi(2))\mathbb{1}_{\{|h(\xi(1), \xi(2))| > a_l\}}| \leq \frac{1}{a_l^{\eta-1}} E|h(\xi(1), \xi(2))|^\eta. \end{aligned}$$

With the triangular inequality and the assumption that  $E|h(\xi(1), \xi(2))|^\eta < \infty$ , it follows that for some constant  $C$  and any  $n_1, n_2 \in \mathbb{N}$  with  $n_1 \leq n_2$

$$E\left|\sum_{i=n_1+1}^{n_2} (h_1(\xi(S_i)) - h_{1,l}(\xi(S_i)))\right| \leq C(n_2 - n_1)a_l^{1-\eta}.$$

Again, we apply the maximal inequality in Theorem 1 of Móricz [18] and obtain

$$E \max_{n \leq 2^l} 2^l |L_n - \tilde{L}_{l,n}| \leq C 2^{l(1+\theta)} a_l^{1-\eta}$$

for every  $\theta > 1$  and we can proceed in the same way as we proved almost sure convergence for  $A_{l,n}$ . So it remains to show the last part (3). We will prove that

$$\begin{aligned} \max_{n \leq 2^l} |E\tilde{R}_{l,n}| &= o(2^{l(2-\frac{1}{\alpha'}+\frac{1}{\alpha'\beta})}) \\ \max_{n \leq 2^l} |\tilde{R}_{l,n} - E\tilde{R}_{l,n}| &= o(2^{l(2-\frac{1}{\alpha'}+\frac{1}{\alpha'\beta})}) \text{ almost surely.} \end{aligned}$$

We obtain with a short calculation that

$$\tilde{R}_{l,n} = \tilde{U}_{l,n} - (n-1)\tilde{L}_{l,n} = (\tilde{U}_{l,n} - U_n) + (n-1)(L_n - \tilde{L}_{l,n}) + R_n$$

and consequently

$$\begin{aligned} \max_{n \leq 2^l} |E\tilde{R}_{l,n}| &= \max_{n \leq 2^l} \left| E(\tilde{U}_{l,n} - U_n) + (n-1)E(L_n - \tilde{L}_{l,n}) + ER_n \right| \\ &\leq E \max_{n \leq 2^l} |U_n - \tilde{U}_{l,n}| + (n-1)E \max_{n \leq 2^l} |L_n - \tilde{L}_{l,n}| + \max_{n \leq 2^l} |ER_n|. \end{aligned}$$

We have already shown in (1) and (2) that the first two summands are of order  $o(2^{l(2-\frac{1}{\alpha'}+\frac{1}{\alpha'\beta})})$ . For the last summand, we use the fact that

$$Eh_2(\xi(1), \xi(2)) = Eh(\xi(1), \xi(2)) - Eh_1(\xi(1)) - Eh_1(\xi(2)) = 0$$

to see that only the indices with  $S_i = S_j$  contribute to the expectation

$$ER_n = E \left[ \sum_{\substack{1 \leq i < j \leq n \\ S_i = S_j}} h_2(\xi(S_i), \xi(S_j)) \right]$$

and due to Proposition 3.1 we have

$$\begin{aligned} \max_{n \leq 2^l} |ER_n| &\leq \max_{n \leq 2^l} E\#\{(i, j) : 1 \leq i < j \leq n | S_i = S_j\} E|h_2(\xi(1), \xi(1))| \\ &\leq E\#\{(i, j) : 1 \leq i < j \leq 2^l | S_i = S_j\} E|h_2(\xi(1), \xi(1))| \\ &\leq C2^{l(2-\frac{1}{\alpha'})}l = o(2^{l(2-\frac{1}{\alpha'}+\frac{1}{\alpha'\beta})}). \end{aligned}$$

To show the convergence of the remaining part, we first decompose it as

$$\begin{aligned} \tilde{R}_{l,n} - ER_{l,n} &= \sum_{\substack{1 \leq i < j \leq n \\ S_i \neq S_j}} (h_{2,l}(\xi(S_i), \xi(S_j)) - Eh_{2,l}(\xi(1), \xi(2))) \\ &\quad + \sum_{\substack{1 \leq i < j \leq n \\ S_i = S_j}} (h_{2,l}(\xi(S_i), \xi(S_j)) - Eh_{2,l}(\xi(1), \xi(1))) =: C_{l,n} + D_{l,n}. \end{aligned}$$

For  $D_{l,n}$ , we have by the independence of  $(S_n)_{n \in \mathbb{N}}$  and  $(\xi(n))_{n \in \mathbb{N}}$  and the fact

$$\begin{aligned} E|h_{2,l}(\xi(1), \xi(1))| &\leq E|h_{0,l}(\xi(1), \xi(1))| + 2E|h_{1,l}(\xi(1))| \\ &\leq E|h(\xi(1), \xi(1))| + 2E|h(\xi(1), \xi(2))| < \infty, \end{aligned}$$

that

$$E \max_{n \leq 2^l} |D_{l,n}| \leq E\#\{(i, j) : 1 \leq i < j \leq 2^l | S_i = S_j\} 2E|h_{2,l}(\xi(1), \xi(1))| \leq C2^{l(2-\frac{1}{\alpha'})}l.$$

In the same way as for  $B_{l,n}$  we now can conclude that  $\max_{n \leq 2^l} D_{l,n} = o(2^{l(2-\frac{1}{\alpha'}+\frac{1}{\alpha'\beta})})$  almost surely. Finally, we will deal with  $C_{l,n}$ . Recall that  $h_{0,l}$  is bounded by  $a_l$ , so  $h_{2,l}$  is bounded by  $3a_l$ . By the triangular inequality for the  $L_\eta$ -norm, we have that

$$\begin{aligned} E|h_{2,l}(\xi(1), \xi(2))|^\eta &\leq \left( \|h_{0,l}(\xi(1), \xi(2))\|_\eta + 2\|h_{1,l}(\xi(1))\|_\eta \right)^\eta \\ &\leq \left( 3\|h(\xi(1), \xi(2))\|_\eta \right)^\eta, \end{aligned}$$

and as a consequence for some constant  $C > 0$

$$E(h_{2,l}(\xi(1), \xi(2)))^2 \leq (3a_l)^{2-\eta} E|h_{2,l}(\xi(1), \xi(2))|^\eta \leq C2^{l\frac{1+\beta'}{\alpha'\beta'}(2-\frac{2\beta'}{1+\beta'})} = C2^{l\frac{2}{\alpha'\beta'}}.$$

Furthermore we have the property of the Hoeffding decomposition that the random variables  $h_{2,l}(\xi(1), \xi(2))$  and  $h_{2,l}(\xi(1), \xi(3))$  are uncorrelated, see Lee [16], page 30. So we can find bounds for the conditional variance of the increments of  $C_{l,n}$ . To simplify the notation, we write

$$Y(i, j) := h_{2,l}(\xi(i), \xi(j))\mathbf{1}_{\{i \neq j\}} - (Eh_{2,l}(\xi(i), \xi(j))) \mathbf{1}_{\{i \neq j\}}$$

and obtain for  $n_1 \leq n_2 \leq 2^l$

$$\begin{aligned} E \left[ (C_{l,n_2} - C_{l,n_1})^2 \mid (X_k)_{k \in \mathbb{N}} \right] &= E \left[ \left( \sum_{\substack{1 \leq i < j \leq n_2 \\ n_1 < j \leq n_2}} Y(S_i, S_j) \right)^2 \mid (X_k)_{k \in \mathbb{N}} \right] \\ &= \sum_{\substack{x, y \in \mathbb{Z} \\ x < y}} \left( \#\{(i, j) : 1 \leq i < j \leq n_2 \mid n_1 < j \leq n_2, S_i = x, S_j = y\} \right. \\ &\quad \left. + \#\{(i, j) : 1 \leq i < j \leq n_2 \mid n_1 < j \leq n_2, S_i = y, S_j = x\} \right)^2 E(Y(x, y))^2 \\ &\leq C 2^{l \frac{2}{\alpha^l \beta^l}} 2 \sum_{x, y \in \mathbb{Z}} (N_{n_2}(x)(N_{n_2}(y) - N_{n_1}(y)))^2. \end{aligned}$$

Due to Theorem 3 of Móricz [18] applied with  $\gamma = 2$  and the (random) superadditive function

$$g(F_{b,n}) = C 2^{\ell \frac{2}{\alpha^l \beta^l} + 1} \sum_{x, y \in \mathbb{Z}} N_{2^\ell}(x)^2 (N_{b+n}(y) - N_b(y))^2.$$

It follows that

$$E \left[ \max_{n \leq 2^l} \left( \sum_{1 \leq i < j \leq n} Y(S_i, S_j) \right)^2 \mid (X_k)_{k \in \mathbb{N}} \right] \leq C 2^{l \frac{2}{\alpha^l \beta^l}} \sum_{x \in \mathbb{Z}} N_{2^l}^2(x) \sum_{y \in \mathbb{Z}} N_{2^l}^2(y) l^2.$$

Taking the expectation with respect to  $(X_k)_{k \in \mathbb{N}}$ , we get the following bound using Proposition 3.1 at

$$\begin{aligned} E \left( \max_{n \leq 2^l} C_{l,n} \right)^2 &= E \left[ E \left[ \max_{n \leq 2^l} \left( \sum_{1 \leq i < j \leq n} Y(S_i, S_j) \right)^2 \mid (X_k)_{k \in \mathbb{N}} \right] \right] \\ &\leq C 2^{l \frac{2}{\alpha^l \beta^l}} l^2 E \left( \sum_{x \in \mathbb{Z}} N_{2^l}^2(x) \right)^2 \\ &\leq C 2^{l \frac{2}{\alpha^l \beta^l}} l^4 2^{l(4 - \frac{2}{\alpha^l})} \\ &= C 2^{l(2 - \frac{1}{\alpha^l} + \frac{1}{\alpha^l \beta^l})} l^4. \end{aligned}$$

We can now use the Chebyshev inequality and arrive at

$$\begin{aligned} \sum_{l=1}^{\infty} P \left( \frac{1}{2^{l(2 - \frac{1}{\alpha^l} + \frac{1}{\alpha^l \beta^l})}} \max_{n \leq 2^l} C_{l,n} \geq \epsilon \right) &\leq \frac{C}{\epsilon^2} \sum_{l=1}^{\infty} \frac{2^{l(2 - \frac{1}{\alpha^l} + \frac{1}{\alpha^l \beta^l})} l^4}{2^{l(2 - \frac{1}{\alpha^l} + \frac{1}{\alpha^l \beta^l})}} \\ &\leq \frac{C}{\epsilon^2} \sum_{l=1}^{\infty} 2^{l(2(\frac{1}{\alpha^l \beta^l} - \frac{1}{\alpha^l \beta})} l^4 < \infty \end{aligned}$$

and the Borel-Cantelli lemma completes the proof. □

## 4 Proofs of main results

*Proof of Theorem 2.1.* Recall the Hoeffding decomposition

$$n^{-2 + \frac{1}{\alpha^l} - \frac{1}{\alpha^l \beta}} U_{[nt]} = \frac{n-1}{n} n^{-1 + \frac{1}{\alpha^l} - \frac{1}{\alpha^l \beta}} L_{[nt]} + n^{-2 + \frac{1}{\alpha^l} - \frac{1}{\alpha^l \beta}} R_{[nt]}.$$

For the linear part in the case  $1 < \alpha \leq 2$ , we apply Theorem 1.1 of Kesten and Spitzer [14] to the random variables  $h_1(\xi(i))$  and conclude that  $\left(n^{-1+\frac{1}{\alpha}-\frac{1}{\alpha\beta}} L_{[nt]}\right)_{t \in [0,1]}$  converges weakly to  $(\Delta_t)_{t \in [0,1]}$ .

In the case  $\alpha = 1$  and  $\beta = 2$ , due to Theorem 2.3 in [2, 9],  $(L_{[nt]}/\sqrt{n \log n})_{t \in [0,1]}$  converges in distribution to  $\left(\frac{\sqrt{2}}{\sqrt{a\pi}} Y_t\right)_{t \in [0,1]}$  as  $n$  goes to infinity.

In the case  $\alpha = 1$  and  $1 < \beta < 2$ , due to Theorem 1 of [4],  $\left(L_{[nt]}/(n^{\frac{1}{\beta}}(\log n)^{\frac{\beta-1}{\beta}})\right)_{t \in [0,1]}$  converges in distribution (with respect to the  $M_1$ -metric) to  $\left(\frac{(\Gamma(\beta+1))^{\frac{1}{\beta}}}{(a\pi)^{\frac{\beta-1}{\beta}}} Y_t\right)_{t \in [0,1]}$  as  $n$  goes to infinity.

In the case  $\alpha < 1$  and  $1 < \beta \leq 2$ , the convergence of  $(L_{[nt]}/n^{\frac{1}{\beta}})_{t \in [0,1]}$  is proved namely in Remark 2 of [4].

For the tightness in  $(D([0,1], J_1))$  when  $\alpha < 1$  and  $\beta = 2$ , we follow the proof of tightness Bolthausen in [2]. For completeness, we explain the adaptations to make. Using the fact that  $L_n(0) \equiv 0$  and that  $(L_i)_{i \geq 0}$  is a sequence of partial sums of a stationary sequence, it is enough to prove that, for every  $\varepsilon > 0$ , there exists  $\lambda > 0$  such that

$$\exists n_0 \geq 1, \quad \forall n \geq n_0, \quad \lambda^2 \mathbb{P}\left(\max_{j \leq n} |L_j| > \lambda \sqrt{n}\right) \leq \varepsilon. \tag{4.1}$$

Recall that  $(V_n/n)_n$  converges almost surely to  $b^2$  (see for example [14, p. 10]) and write

$$\mathbb{P}\left(\max_{j \leq n} |L_j| > \lambda \sqrt{n}\right) \leq \mathbb{P}\left(\max_{j \leq n} |L_j| > \lambda \sqrt{V_n}/(2b)\right) + \mathbb{P}(V_n > 4b^2 n).$$

Let  $\rho > \sqrt{2}$ . Following [2], we obtain that

$$\mathbb{P}\left(\max_{j \leq n} |L_j| \geq \sigma \rho \sqrt{V_n}\right) \leq 2\mathbb{P}\left(|L_n| > (\rho - \sqrt{2})\sigma \sqrt{V_n}\right),$$

with  $\sigma := \sqrt{\text{Var}(h_1(\xi_1))}$ . But we know that  $(L_n/\sqrt{V_n})_n$  converges in distribution to  $Y$ , so

$$\limsup_{n \rightarrow +\infty} \mathbb{P}\left(\max_{j \leq n} |L_j| \geq \rho \sigma \sqrt{V_n}\right) \leq 2\mathbb{P}\left(|Y| > (\rho - \sqrt{2})\sigma\right).$$

Choose  $\rho > \sqrt{2}$  such that  $2(2b\rho\sigma)^2 \mathbb{P}\left(|Y| > (\rho - \sqrt{2})\sigma\right) < \varepsilon$  (this is possible since  $Y$  is gaussian). The tightness criteria (4.1) is satisfied with  $\lambda = 2b\rho\sigma$ .

For the remainder term, we have proved in Proposition 3.2 that

$$\sup_{t \in [0,1]} \left| n^{-2+\frac{1}{\alpha'}-\frac{1}{\alpha'\beta}} R_{[nt]} \right| \longrightarrow 0$$

in probability. The statement of the theorem follows by Slutsky's theorem.

In the cases  $\alpha \leq 1$ , note that the uniform convergence of the remainder  $R_{[nt]}$  implies the convergence with respect to the  $M_1$ -topology. □

*Proof of Theorem 2.2.* We use again the Hoeffding decomposition

$$\frac{U_n}{n^{\frac{7}{4}}(\log \log n)^{\frac{3}{4}}} = \frac{n-1}{n} \frac{L_n}{(n \log \log n)^{\frac{3}{4}}} + \frac{R_n}{n^{\frac{7}{4}}(\log \log n)^{\frac{3}{4}}}.$$

For the remainder term we use Proposition 3.2 with  $\alpha = \beta = 2$ :

$$R_n = o(n^{2-\frac{1}{2}+\frac{1}{4}}) = o(n^{\frac{7}{4}}(\log \log n)^{\frac{3}{4}}) \quad \text{almost surely.}$$

As  $L_n = \sum_{i=1}^n h_1(\xi(S_i))$ , we can apply Corollary 1 of Zhang [21] and we obtain

$$\limsup_{n \rightarrow \infty} \pm \frac{L_n}{(n \log \log n)^{\frac{3}{4}}} = \frac{2^{\frac{1}{4}} \text{Var}(\xi)^{\frac{1}{2}}}{3 \text{Var}(X)^{\frac{1}{4}}}$$

almost surely, which leads to the statement of the theorem.  $\square$

## References

- [1] Beutner, E. and Zähle, H.: Continuous mapping approach to the asymptotics of U- and V-statistics. *Bernoulli* **20**, (2014), no. 2, 846–877. MR-3178520
- [2] Bolthausen, E.: A central limit theorem for two-dimensional random walks in random sceneries. *Ann. Probab.* **17**, (1989), no. 1, 108–115. MR-0972774
- [3] Cabus, P. and Guillin-Plantard, N.: Functional limit theorems for U-statistics indexed by a random walk, *Stochastic Process. Appl.* **101**, (2002), no. 1, 143–160. MR-1921446
- [4] Castell, F and Guillin-Plantard, N. and Pène, F.: Limit theorems for one and two-dimensional random walks in random scenery, *Ann. Inst. Henri Poincaré Probab. Stat.* **49**, (2013), no. 2, 506–528. MR-3088379
- [5] Castell, F. and Guillin-Plantard, N. and Pène, F. and Schapira, Br.: A local limit theorem for random walks in random scenery and on randomly oriented lattices, *Ann. Probab.*, **39**, (2011), no. 6, 2079–2118. MR-2932665
- [6] Csáki, E. and König, W. and Shi, Z.: An embedding for the Kesten-Spitzer random walk in random scenery, *Stochastic Processes and their Application* **82**, (1999), no. 2, 283–292. MR-1700010
- [7] Dabrowski, A. and Dehling, H. and Mikosch, T. and Sharipov, O.: Poisson limits for U-statistics, *Stochastic Process. Appl.* **99**, (2002), no. 1, 137–157. MR-1894255
- [8] Dehling, H. and Taqqu, M.S.: The empirical process of some long-range dependent sequences with an application to U-statistics, *Ann. Statist.* **17**, (1989), no. 4, 1767–1783. MR-1026312
- [9] Deligiannidis, G. and Utev, S.: Computation of the asymptotics of the variance of the number of self-intersections of stable random walks using the Wiener-Darboux theory. (Russian) *Sibirsk. Mat. Zh.* **52**, (2011), no. 4, 809–822; translation in *Sib. Math. J.* **52**, (2011), no. 4, 639–650. MR-2883216
- [10] Franke, B. and Pène, F. and Wendler, M.: Convergence of U-statistics indexed by a random walk to stochastic integrals of a Levy sheet, *Bernoulli*, **23**, (2017), 329–378. MR-3556775
- [11] Guillin-Plantard, N. and Ladret, V.: Limit theorems for U-statistics indexed by a one dimensional random walk, *ESAIM Probab. Stat.* **9**, (2005), 98–115. MR-2148962
- [12] Heinrich, L. and Wolf, W.: On the convergence of U-statistics with stable limit distribution, *J. Multivariate Anal.* **44**, (1993), no. 2, 266–278. MR-1219207
- [13] Hoeffding, W.: A class of statistics with asymptotically normal distribution, *Ann. Math. Stat.* **19**, (1948), 293–325. MR-0026294
- [14] Kesten, H. and Spitzer, F.: A limit theorem related to an new class of self-similar processes, *Z. Wahrsch. Verw. Gebiete* **50**, (1979), no. 1, 5–25. MR-0550121
- [15] Khoshnevisan, D. and Lewis, T.M.: A law of the iterated logarithm for stable processes in random scenery, *Stochastic Process. Appl.* **74**, (1998), no. 1, 89–121. MR-1624017
- [16] Lee, A.J.: U-Statistics: Theory and Practice, *Statistics: Textbooks and Monographs*, **110**, Marcel Dekker, New York (1990). MR-1075417
- [17] Lewis, T.M.: A law of the iterated logarithm for random walk in random scenery with deterministic normalizer, *J. Theoret. Probab.* **6**, (1993), no. 2, 209–230. MR-1215656
- [18] Móricz, F.: Moment inequalities and the strong law of large numbers, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **35**, (1976), no. 4, 299–314. MR-0407950
- [19] Skorohod, A.V.: Limit theorems for stochastic processes (Russian), *Teor. Veroyatnost. i Primenen* **1**, (1956), 289–319. MR-0084897
- [20] Skorohod, A.V.: Limit theorems for stochastic processes with independent increments (Russian), *Teor. Veroyatnost. i Primenen* **2**, (1957), 145–177. MR-0094842

- [21] Zhang, L.: Strong approximation for the general Kesten-Spitzer random walk in independent random scenery, *Sci. in China Ser. A* **44**, (2001), no. 5, 619–630. MR-1843757

**Acknowledgments.** The research was supported by the DFG Sonderforschungsbereich 823 (Collaborative Research Center) *Statistik nichtlinearer dynamischer Prozesse*.