Stable Limit Theorem for U-Statistics Processes Indexed by a Random Walk

Martin Wendler
joint work with Brice Franke and Françoise Pène

RUHR-UNIVERSITÄT BOCHUM

ÖMG Congress and Annual DMV Meeting
Innsbruck, September 23 - 27, 2013

Outline

Introduction
- Random Walks in Random Scenery
- *U*-Statistics

Main Results
- Limit Theorems
- Sketch of Proofs

Partial Sum Processes

for $(\xi_n)_{n\in\mathbb{N}}$ iid with $E\xi_i = 0$ and $E\xi_i^2 < \infty$

$$\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\xi_i\right)_t \Rightarrow B$$

for a Brownian motion B (self-similar process with index $1/2$)

different scaling and different self-similarity of the limes process for

- heavy tailed random variables
- long range dependent random variables

Definition

random walk: $(X_n)_{n\in\mathbb{N}}$ iid $\mathbb{Z}$-valued random variables in the normal domain of attraction of an α-stable law F_α with $0 < \alpha \leq 2$

$$S_n := \sum_{m=1}^{n} X_m$$

scenery: $(\xi(i))_{i\in\mathbb{Z}}$ iid $\mathbb{R}$-valued random variables in the normal domain of attraction of an β-stable law F_β with $0 < \beta \leq 2$

Definition

random walk: $(X_n)_{n\in\mathbb{N}}$ iid $\mathbb{Z}$-valued random variables in the normal domain of attraction of an α-stable law F_α with $0 < \alpha \leq 2$

$$S_n := \sum_{m=1}^{n} X_m$$

scenery: $(\xi(i))_{i\in\mathbb{Z}}$ iid $\mathbb{R}$-valued random variables in the normal domain of attraction of an β-stable law F_β with $0 < \beta \leq 2$
random walk in random scenery: partial sum process

$$W_n(t) = \sum_{i=1}^{\lfloor nt \rfloor} \xi(S_i)$$

(or linear interpolated version)

Limit Theorem:

Theorem (Kesten, Spitzer, 1979)

For $1 < \alpha \leq 2$ *in* $C[0,1]$

$$n^{-1+\frac{1}{\alpha}-\frac{1}{\alpha\beta}} W_n(t) \Rightarrow \Delta_t$$

Limit Theorem:

Theorem (Kesten, Spitzer, 1979)

For $1 < \alpha \leq 2$ *in* $C[0,1]$

$$n^{-1+\frac{1}{\alpha}-\frac{1}{\alpha\beta}} W_n(t) \Rightarrow \Delta_t$$

properties of Δ

- self-similar with index $\delta = 1 - \frac{1}{\alpha} + \frac{1}{\alpha\beta}$
- non Gaussian even if $\beta = 2$ (scenery has Gaussian limit)
- continuous even if $\beta < 2$ (scenery has limit with jumps)

Definition of Limit Process:

Rewrite $W_n = W_n(1)$ as

$$W_n = \sum_{i=1}^{n} \xi(S_i) = \sum_{x \in \mathbb{Z}} N_n(x)\xi(x)$$

with occupation times

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

Definition of Limit Process:

Rewrite $W_n = W_n(1)$ as

$$W_n = \sum_{i=1}^{n} \xi(S_i) = \sum_{x \in \mathbb{Z}} N_n(x)\xi(x)$$

with occupation times

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

$$\Delta_t = \int_{-\infty}^{\infty} L_t(x) dZ(x)$$

with two-sided β-stable process Z, local time L of an α-stable process $S^\star$, i.e.

$$\int_0^t \mathbb{1}_{[a,b)}(S_s^\star) ds = \int_a^b L_t(x) dx$$

Transient Random Walks

Theorem (Castell, Guillotin-Plantard, Pène, 2013)

If $\alpha = 1$, $\beta \neq 1$, *then in* $D[0,1]$ *(equipped with* M_1 *topology)*

$$\left(n^{-\frac{1}{\beta}} (\log n)^{\frac{1-\beta}{\beta}} W_n(t) \right)_t \Rightarrow C_1 Z$$

- β-stable process Z (discontinuous)
- $S_i = S_j$ less often, so less dependence

Transient Random Walks

Theorem (Castell, Guillotin-Plantard, Pène, 2013)

If $\alpha = 1$, $\beta \neq 1$, *then in* $D[0,1]$ *(equipped with* M_1 *topology)*

$$\left(n^{-\frac{1}{\beta}}(\log n)^{\frac{1-\beta}{\beta}}W_n(t)\right)_t \Rightarrow C_1 Z$$

- β-stable process Z (discontinuous)
- $S_i = S_j$ less often, so less dependence

Theorem (Castell, Guillotin-Plantard, Pène, 2013)

If $\alpha < 1$, *then*

$$\left(n^{-\frac{1}{\beta}}W_n(t)\right)_t \Rightarrow C_2 Z$$

Definition

- $(X_n)_{n\in\mathbb{N}}$ stationary sequence of random variables
- $h : \mathbb{R}^2 \to \mathbb{R}$ measurable and symmetric

Definition

The (bivariate) *U-statistic* $U_n(h)$ with kernel h is defined as

$$U_n(h) := \sum_{1\leq i<j\leq n} h(X_i, X_j).$$

Examples

1. $h(x, y) = |x - y|$

$$\frac{2}{n(n-1)} U_n(h) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} |X_i - X_j|,$$

   Gini's mean difference

2. $h(x, y) = \frac{1}{2}(x - y)^2$

$$\frac{2}{n(n-1)} U_n(h) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

   sample variance

Hoeffding-Decomposition

$U_n(h)$ can be decomposed into a *linear part* and a *degenerate part*

$$U_n(h) = \frac{n(n-1)}{2}\theta + (n-1)\sum_{i=1}^{n} h_1(X_i) + U_n(h_2)$$

with

$$\begin{aligned} \theta &:= Eh(X_1, X_2) \\ h_1(x) &:= Eh(x, X_2) - \theta \\ h_2(x, y) &:= h(x, y) - h_1(x) - h_1(y) - \theta. \end{aligned}$$

Note that h_2 is degenerate, i.e. for $i < j$ and X_i, X_j independent

$$E[h_2(X_i, X_j)|X_i] = 0.$$

Hoeffding-Decomposition II

Hence for independent data and $i < j < k$

$$E[h_2(X_i, X_j)h_2(X_i, X_k)] = E[E[h_2(X_i, X_j)h_2(X_i, X_k)|X_i, X_j]]$$
$$= E[h_2(X_i, X_j)E[h_2(X_i, X_k)|X_i, X_j]] = 0 = E[h_2(X_i, X_j)] \cdot E[h_2(X_i, X_k)]$$

So the summands of $U_n(h_2)$ are uncorrelated and we have

$$\operatorname{Var} U_n(h_2) = O(n^2)$$

Hoeffding-Decomposition II

Hence for independent data and $i < j < k$

$$E[h_2(X_i, X_j)h_2(X_i, X_k)] = E[E[h_2(X_i, X_j)h_2(X_i, X_k)|X_i, X_j]]$$
$$= E[h_2(X_i, X_j)E[h_2(X_i, X_k)|X_i, X_j]] = 0 = E[h_2(X_i, X_j)] \cdot E[h_2(X_i, X_k)]$$

So the summands of $U_n(h_2)$ are uncorrelated and we have

$$\operatorname{Var} U_n(h_2) = O(n^2)$$

Theorem (Hoeffding, 1948)

If $(X_n)_{n\in\mathbb{N}}$ *is a sequence of iid random variables and* $\operatorname{Var} h_1(X_1) < \infty$, *then*

$$n^{-\frac{3}{2}}\left(U_n(h) - \theta\right) \xrightarrow{\mathcal{D}} N\left(0, \operatorname{Var} h_1(X_1)\right).$$

Existing Results

under second moments (or higher) and dependence:

- Sen (1963): *m*-dependence
- Yoshihara (1976): absolute regularity
- Dehling, Taqqu (1989): long-range dependent Gaussian processes
- Guillotin-Plantard, Ladret (2005): random walk in random scenery
- Dehling, Wendler (2010): strong mixing

without second moments, under independence:

- Heinrich, Wolf (1993)
- Dabrowski, Dehling, Mikosch, Sharipov (2002)

U-Statistic indexed by a random walk

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

Assumption: $E|h(\xi(1), \xi(2))| < \infty$, $E|h(\xi(1), \xi(1))| < \infty$
without loss of generality $Eh(\xi(1), \xi(2)) = 0$

U-Statistic indexed by a random walk

$$U_n = \sum_{1 \le i < j \le n} h\left(\xi(S_i), \xi(S_j)\right)$$

Assumption: $E|h(\xi(1), \xi(2))| < \infty$, $E|h(\xi(1), \xi(1))| < \infty$
without loss of generality $Eh(\xi(1), \xi(2)) = 0$

Hoeffding decomposition

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

$$\text{with} \quad h_1(x) := E(h(x, \xi(1)))$$
$$h_2(x, y) := h(x, y) - h_1(x) - h_1(y).$$

Weak Convergence

- $h_1(\xi(1))$ in normal domain of attraction of β-stable law, $1 < \beta \le 2$
- $E|h(\xi(1), \xi(2))|^\eta < \infty$ with $\eta = \frac{2\beta'}{1+\beta'}$ for a $\beta' > \beta$
- X_1 in normal domain of attraction of an α-stable law F_α

Theorem (Franke, Pène, W.)

If $1 < \alpha \le 2$*, then in* $C[0,1]$

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

$U_n(t) = U_k$ if $nt = k$, linear interpolated in between

Transient Random Walks

Theorem (Franke, Pène, W.)

If $\alpha = 1$, *then in* $D[0,1]$ *(equipped with* M_1 *topology)*

$$\left(n^{1-\frac{1}{\beta}}(\log n)^{\frac{1-\beta}{\beta}} U_{\lfloor nt \rfloor}\right)_{t\in[0,1]} \Rightarrow (C_3 Z_t)_{t\in[0,1]},$$

Z β-stable process which is the limit of $n^{-1/\beta}\sum_{i=1}^{\lfloor nt \rfloor} h_1(\xi_i)$

Transient Random Walks

Theorem (Franke, Pène, W.)

If $\alpha = 1$*, then in* $D[0,1]$ *(equipped with* M_1 *topology)*

$$\left(n^{1-\frac{1}{\beta}}(\log n)^{\frac{1-\beta}{\beta}}U_{\lfloor nt\rfloor}\right)_{t\in[0,1]} \Rightarrow (C_3 Z_t)_{t\in[0,1]},$$

Z β-stable process which is the limit of $n^{-1/\beta}\sum_{i=1}^{\lfloor nt\rfloor} h_1(\xi_i)$

Theorem (Franke, Pène, W.)

If $0 < \alpha < 1$*, then in* $D[0,1]$

$$\left(n^{-1-\frac{1}{\beta}}U_{\lfloor nt\rfloor}\right)_{t\in[0,1]} \Rightarrow (C_4 Z_t)_{t\in[0,1]},$$

Law of the iterated logarithm

- $E|h_1(\xi(i))|^p < \infty$
- $E|X_i|^p < \infty$ for some $p > 2$

Theorem (Franke, Pène, W.)

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

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

almost surely

Lemma for Degenerate Part

Hoeffding decomposition

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

Lemma

for $0 < \alpha \le 2$

$$\max_{k\le n} \sum_{1\le i<j\le k} h_2(\xi(S_i),\xi(S_j)) = o(n^{2-\frac{1}{\alpha'}+\frac{1}{\alpha'\beta}})$$

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

Truncation of Kernel

$$h_{0,n}(x,y) := \begin{cases} h(x,y) & \text{if } |h(x,y)| \le a_n = n^{\frac{1+\beta'}{\alpha'\beta'}} \\ 0 & \text{else} \end{cases}.$$

Hoeffding decomposition of truncated kernel

$$h_{1,n}(x) := E(h_{0,n}(x,\xi(1)))$$
$$h_{2,n}(x,y) := h_{0,n}(x,y) - h_{1,n}(x) - h_{1,n}(y).$$

$$\sum_{1\le i<j\le n} h_{0,n}(\xi(S_i),\xi(S_j)) = (n-1)\sum_{i=1}^{n} h_{1,n}(\xi(S_i)) + \sum_{1\le i<j\le n} h_{2,n}(\xi(S_i),\xi(S_j))$$

Hoeffding Decomposition Read Backwards

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

$$= \Big(\sum_{1\leq i<j\leq k} h(\xi(S_i),\xi(S_j)) - \sum_{1\leq i<j\leq k} h_{0,n}(\xi(S_i),\xi(S_j))\Big)$$

$$- (n-1)\sum_{i=1}^{n} h_1(\xi(S_i)) + \sum_{1\leq i<j\leq k} h_{0,n}(\xi(S_i),\xi(S_j))$$

$$= \sum_{1\leq i<j\leq k} \big(h(\xi(S_i),\xi(S_j)) - h_{0,n}(\xi(S_i),\xi(S_j))\big)$$

$$- (n-1)\sum_{i=1}^{n} \big(h_1(\xi(S_i)) - h_{1,n}(\xi(S_i))\big) + \sum_{1\leq i<j\leq k} h_{2,n}(\xi(S_i),\xi(S_j))$$

Truncation of Linear Part

$$E|h_1(\xi(1)) - h_{1,n}(\xi(1))| \leq \frac{1}{a_n^{\eta-1}} E|h(\xi(1), \xi(2))|^{\eta}.$$

with triangular inequality

$$E\left|(n-1)\sum_{i=1}^{n}\left(h_1(\xi(i)) - h_{1,n}(\xi(i))\right)\right| \leq Cn^2 \frac{1}{a_n^{\eta-1}} = o(n^{2-\frac{1}{\alpha'}+\frac{1}{\alpha'\beta}})$$

with Markov inequality

$$(n-1)\sum_{i=1}^{n}\left(h_1(\xi(i)) - h_{1,n}(\xi(i))\right) = o_p(n^{2-\frac{1}{\alpha'}+\frac{1}{\alpha'\beta}})$$

similar: truncation of U_n

Degenerate Part

$$E\left(h_{2,n}(\xi(1),\xi(2))\right)^2 \le C a_n^{2-\eta} E|h_2(\xi(1),\xi(2))|^\eta = C n^{\frac{2}{\alpha'\beta'}}$$

as summands are uncorrelated

$$E\Big(\sum_{\substack{1\le i<j\le k\\ S(i)\neq S(j)}} h_{2,n}(\xi(S_i),\xi(S_j))\Big)^2 \le C n^{\frac{2}{\alpha'\beta'}} E\Big(\sum_{x\in\mathbb{Z},y\in\mathbb{Z}} N_n^2(x)N_n^2(y)\Big)$$

$$= o(n^{4-\frac{2}{\alpha'}+\frac{2}{\alpha'\beta}})$$

because for occupation times $N_n(x)$

$$E\left(\sum_{x\in\mathbb{Z}} N_n^2(x)\right)^2 = O(n^{4-\frac{2}{\alpha}}\log^2 n)$$

Linear Part

$$(n-1)\sum_{i=1}^{n} h_1(\xi(S_i))$$

weak convergence

- $1 < \alpha \leq 2$: Kesten, Spitzer (1979)
- $0 < \alpha \leq 1$: Castell, Guillotin-Plantard, Pène (2013)

law of the iterated logarithm

- $\alpha = \beta = 2$: Zhang (2001)

Linear Part

$$(n-1)\sum_{i=1}^{n} h_1(\xi(S_i))$$

weak convergence

- $1 < \alpha \leq 2$: Kesten, Spitzer (1979)
- $0 < \alpha \leq 1$: Castell, Guillotin-Plantard, Pène (2013)

law of the iterated logarithm

- $\alpha = \beta = 2$: Zhang (2001)

Open Questions:

- limit for $\beta \leq 1$ (Hoeffding decomposition not possible)
- limit if $h_1 = 0$ (degenerate U-statistic)

Acknowledgement

Thank you for listening!

RUHR-UNIVERSITÄT BOCHUM

References

- F. Castell, N. Guillotin-Plantard, F. Pène, Limit theorems for one and two-dimensional random walks in random scenery, *Ann. Inst. H. Poincaré* **49**, (2013), 506-528.
- B. Franke, F. Pène, M. Wendler, Stable Limit Theorem for U-Statistic Processes Indexed by a Random Walk, *preprint* (2013) arXiv:1212.2133.
- N. Guillotin-Plantard, V. Ladret, Limit theorems for U-statistics indexed by a one dimensional random walk, *ESAIM* **9** (2005) 95-115.
- L. Heinrich, W. Wolf, On the convergence of U-statistics with stable limit distribution, *Journal of Multivariate Analysis* **44** (1993) 266-278.
- H. Kesten, F. Spitzer, A limit theorem related to an new class of self similar processes, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **50** (1979) 5-25.
- L. Zhang, Strong approximation for the general Kesten-Spitzer random walk in independent random scenery, *Science in China, series A* **44** (2001) 619-630.