Functional limit theorem for canonical U-processes of dependent observations

Igor S. Borisov, Vasiliy Zhechev
Sobolev Institute of Mathematics, Novosibirsk

April 11, 2011

Outline

1. Introduction
2. History
3. Results

$X_1, X_2, ...$ is a stationary sequence of random variables defined on $(\Omega, F, \mathbb{P})$.

$X_1, X_2, ...$ is a stationary sequence of random variables defined on $(\Omega, F, \mathbb{P})$.
Denote by $\{X_i^*\}$ a sequence of independent copies of random variables X_1.

$X_1, X_2, ...$ is a stationary sequence of random variables defined on $(\Omega, F, \mathbb{P})$.
Denote by $\{X_i^*\}$ a sequence of independent copies of random variables X_1.
Denote by F the distribution of X_1.

Kernel f

Let $f \in L_2(\mathbb{R}^m, F^m)$,

Kernel f

Let $f \in L_2(\mathbb{R}^m, F^m)$,
then we have

$$f(t_1, ..., t_m) = \sum_{k_1=0}^{\infty} ... \sum_{k_m=0}^{\infty} f_{k_1...k_m} e_{k_1}(t_1)...e_{k_m}(t_m),$$

where $\{e_{k_j}\}$ is an orthonormal basis in $L_2(\mathbb{R}, F)$.

Kernel f

Let $f \in L_2(\mathbb{R}^m, F^m)$,
then we have

$$f(t_1, ..., t_m) = \sum_{k_1=0}^{\infty} ... \sum_{k_m=0}^{\infty} f_{k_1...k_m} e_{k_1}(t_1)...e_{k_m}(t_m),$$

where $\{e_{k_j}\}$ is an orthonormal basis in $L_2(\mathbb{R}, F)$.

$$\sum_{k_1=1}^{\infty} ... \sum_{k_m=1}^{\infty} |f_{k_1...k_m}| < \infty$$

Kernel f

Let $f \in L_2(\mathbb{R}^m, F^m)$,
then we have

$$f(t_1, ..., t_m) = \sum_{k_1=0}^{\infty} ... \sum_{k_m=0}^{\infty} f_{k_1...k_m} e_{k_1}(t_1)...e_{k_m}(t_m),$$

where $\{e_{k_j}\}$ is an orthonormal basis in $L_2(\mathbb{R}, F)$.

$$\sum_{k_1=1}^{\infty} ... \sum_{k_m=1}^{\infty} |f_{k_1...k_m}| < \infty$$

$f(t_1, ..., t_m) \in L_2(\mathbb{R}^m, F^m)$ is *canonical*, i.e.,

$$\mathbb{E} f(y_1, ..., y_{i-1}, X_1, y_{i+1}, ..., y_m) = 0$$

for all $y_j \in \mathbb{R}$ and $i \in \{1...m\}$.

U-process

We consider the sequence of U-statistics

$$U_n(t) := n^{-m/2} \sum_{1 \le i_1 \ne} \dots \sum_{\ne i_m \le [nt]} f(X_{i_1}, \dots, X_{i_m}), \quad t \in [0, 1],$$

as a random process in $D[0, 1]$.

U-process

We consider the sequence of U-statistics

$$U_n(t) := n^{-m/2} \sum_{1\le i_1 \ne} \dots \sum_{\ne i_m \le [nt]} f(X_{i_1}, \dots, X_{i_m}), \quad t \in [0,1],$$

as a random process in $D[0,1]$.
For i.i.d. observations,

$$U_n \xrightarrow{d} \sum_{k_1=0}^{\infty} \dots \sum_{k_m=0}^{\infty} f_{k_1 \dots k_m} \Pi_{j=1}^{\infty} H_{v_j(i_1,\dots,i_m)}(\tau_j),$$

U-process

We consider the sequence of U-statistics

$$U_n(t) := n^{-m/2} \sum_{1 \le i_1 \ne} \dots \sum_{\ne i_m \le [nt]} f(X_{i_1}, \dots, X_{i_m}), \quad t \in [0, 1],$$

as a random process in $D[0, 1]$.
For i.i.d. observations,

$$U_n \xrightarrow{d} \sum_{k_1=0}^{\infty} \dots \sum_{k_m=0}^{\infty} f_{k_1 \dots k_m} \Pi_{j=1}^{\infty} H_{\nu_j(i_1, \dots, i_m)}(\tau_j),$$

where $\{\tau_j\}$ is a sequence of independent variables with the standard normal distribution,

U-process

We consider the sequence of U-statistics

$$U_n(t) := n^{-m/2} \sum_{1 \le i_1 \ne} \dots \sum_{\ne i_m \le [nt]} f(X_{i_1}, \dots, X_{i_m}), \quad t \in [0,1],$$

as a random process in $D[0,1]$.
For i.i.d. observations,

$$U_n \xrightarrow{d} \sum_{k_1=0}^{\infty} \dots \sum_{k_m=0}^{\infty} f_{k_1 \dots k_m} \Pi_{j=1}^{\infty} H_{\nu_j(i_1,\dots,i_m)}(\tau_j),$$

where $\{\tau_j\}$ is a sequence of independent variables with the standard normal distribution,
$\nu_j(i_1, \dots, i_m)$ is the number of the subscripts $i_1, \dots, i_m$ equal to j,

U-process

We consider the sequence of U-statistics

$$U_n(t) := n^{-m/2} \sum_{1\le i_1 \ne} \dots \sum_{\ne i_m \le [nt]} f(X_{i_1}, ..., X_{i_m}), \quad t \in [0,1],$$

as a random process in $D[0,1]$.
For i.i.d. observations,

$$U_n \xrightarrow{d} \sum_{k_1=0}^{\infty} \dots \sum_{k_m=0}^{\infty} f_{k_1 \dots k_m} \Pi_{j=1}^{\infty} H_{\nu_j(i_1,\dots,i_m)}(\tau_j),$$

where $\{\tau_j\}$ is a sequence of independent variables with the standard normal distribution,
$\nu_j(i_1, ..., i_m)$ is the number of the subscripts $i_1, ..., i_m$ equal to j,
and $H_k(x)$ are the Hermite polynomials defined by the formula

$$H_k(x) = (-1)^k \exp(x^2/2) \frac{d^k}{dx^k} \exp(-x^2/2), \quad k \ge 0.$$

φ-mixing

$X_1, X_2, \ldots$ satisfy φ-mixing, i.e.,

φ-mixing

$X_1, X_2, \ldots$ satisfy φ-mixing, i.e.,

$$\varphi(i) := \sup_{k\geq 1} \sup_{A\in\mathfrak{M}_1^k, B\in\mathfrak{M}_{k+i}^\infty, \mathbb{P}(A)>0} \frac{|\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)|}{\mathbb{P}(A)} \to 0 \ \text{ for } \ i \to \infty.$$

φ-mixing

$X_1, X_2, ...$ satisfy φ-mixing, i.e.,

$$\varphi(i) := \sup_{k\geq 1} \sup_{A\in\mathfrak{M}_1^k, B\in\mathfrak{M}_{k+i}^\infty, \mathbb{P}(A)>0} \frac{|\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)|}{\mathbb{P}(A)} \to 0 \text{ for } i \to \infty.$$

$\mathfrak{M}_j^k$ is the σ-algebra of events generated by $X_j, ..., X_k$.

φ-mixing

$X_1, X_2, \ldots$ satisfy φ-mixing, i.e.,

$$\varphi(i) := \sup_{k \geq 1} \sup_{A \in \mathfrak{M}_1^k, B \in \mathfrak{M}_{k+i}^\infty, \mathbb{P}(A) > 0} \frac{|\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)|}{\mathbb{P}(A)} \to 0 \text{ for } i \to \infty.$$

$\mathfrak{M}_j^k$ is the σ-algebra of events generated by $X_j, \ldots, X_k$.
We assume that

$$\sum_{k=1}^\infty \varphi(k)^{1/2} < \infty.$$

This condition provides the corresponding central limit theorem.

Restrictions on the joint distributions of the sample elements.

(AC) For any set of pairwise different subscripts $(j_1, ..., j_m)$, the distribution of the vector $(X_{j_1}, ..., X_{j_m})$ is absolutely continuous with respect to the distribution of the vector $(X_1^*, ..., X_m^*)$.

Restrictions on the joint distributions of the sample elements.

(AC) For any set of pairwise different subscripts $(j_1, ..., j_m)$, the distribution of the vector $(X_{j_1}, ..., X_{j_m})$ is absolutely continuous with respect to the distribution of the vector $(X_1^*, ..., X_m^*)$.
Then U-statistics can be represented as the multiple series that converges with probability 1:

Restrictions on the joint distributions of the sample elements.

(AC) For any set of pairwise different subscripts $(j_1, ..., j_m)$, the distribution of the vector $(X_{j_1}, ..., X_{j_m})$ is absolutely continuous with respect to the distribution of the vector $(X_1^*, ..., X_m^*)$.
Then *U*-statistics can be represented as the multiple series that converges with probability 1:

$$U_n(t) = n^{-m/2} \sum_{k_1=0}^{\infty} \dots \sum_{k_m=0}^{\infty} f_{k_1 \dots k_m} \sum_{1 \le i_1 \ne} \dots \sum_{\ne i_m \le [nt]} e_{k_1}(X_{i_1}) \dots e_{k_m}(X_{i_m}).$$

Assume that the basis $\{e_j(t)\}$ satisfies the restriction:

$$\sup_i \mathbb{E}|e_i(X_1)|^m < \infty.$$

Assume that the basis $\{e_j(t)\}$ satisfies the restriction:

$$\sup_i \mathbb{E}|e_i(X_1)|^m < \infty.$$

Introduce a sequence of dependent Wiener processes $\{w_i(t)\}$ with covariances

Assume that the basis $\{e_j(t)\}$ satisfies the restriction:

$$\sup_i \mathbb{E}|e_i(X_1)|^m < \infty.$$

Introduce a sequence of dependent Wiener processes $\{w_i(t)\}$ with covariances

$$\mathbb{E}w_k(t_1)w_k(t_2) = min(t_1, t_2)\Big(1 + 2\sum_{j=1}^{\infty} \mathbb{E}e_k(X_1)e_k(X_{j+1})\Big);$$

Assume that the basis $\{e_j(t)\}$ satisfies the restriction:

$$\sup_i \mathbb{E}|e_i(X_1)|^m < \infty.$$

Introduce a sequence of dependent Wiener processes $\{w_i(t)\}$ with covariances

$$\mathbb{E}w_k(t_1)w_k(t_2) = min(t_1, t_2)\Big(1 + 2\sum_{j=1}^{\infty}\mathbb{E}e_k(X_1)e_k(X_{j+1})\Big);$$

$$\mathbb{E}w_k(t_1)w_l(t_2) = min(t_1, t_2)\Big(\sum_{j=1}^{\infty}\mathbb{E}e_k(X_1)e_l(X_{j+1}) + \sum_{j=1}^{\infty}\mathbb{E}e_l(X_1)e_k(X_{j+1})\Big)$$

Introduce the process

$$U(t) := \sum_{k_1=1}^{\infty} \dots \sum_{k_m=1}^{\infty} f_{k_1 \dots k_m} t^{m/2} \Pi_{j=1}^{\infty} H_{v_j(i_1,\dots,i_m)}(t^{-1/2} w_j(t))$$

results on the limiting behavior of U-statistics with canonical kernels: history

(I)i.i.d. observations:
A. F. Ronzhin, 1986 (polynomial form)
Rubin H., Vitale R., 1980 (integral form)
Denker M., Grillenberger C., Kassel and Keller G., Heidelberg., 1985 (integral form)
Dehling H., Denker M., Philipp W., 1987 (integral form)

results on the limiting behavior of U-statistics with canonical kernels: history

(I)i.i.d. observations:
A. F. Ronzhin, 1986 (polynomial form)
Rubin H., Vitale R., 1980 (integral form)
Denker M., Grillenberger C., Kassel and Keller G., Heidelberg., 1985 (integral form)
Dehling H., Denker M., Philipp W., 1987 (integral form)
(II) Stationary connected observations:
I.S. Borisov, N.V. Volodko (limit behavior, polynomial form)

the introduced conditions

The stationary sequence X_i satisfies φ-mixing

the introduced conditions

The stationary sequence X_i satisfies φ-mixing

$$\sup_i \mathbb{E}|e_i(X_1)|^m < \infty.$$

the introduced conditions

The stationary sequence X_i satisfies φ-mixing

$$\sup_i \mathbb{E}|e_i(X_1)|^m < \infty.$$

$$\sum_{k=1}^{\infty} \varphi(k)^{1/2} < \infty$$

the introduced conditions

The stationary sequence X_i satisfies φ-mixing

$$\sup_i \mathbb{E}|e_i(X_1)|^m < \infty.$$

$$\sum_{k=1}^{\infty} \varphi(k)^{1/2} < \infty$$

$$f \in L_2(\mathbb{R}^m, F^m) \text{ and } \sum_{k_1=1}^{\infty} \dots \sum_{k_m=1}^{\infty} |f_{k_1 \dots k_m}| < \infty$$

the introduced conditions

The stationary sequence X_i satisfies φ-mixing

$$\sup_i \mathbb{E}|e_i(X_1)|^m < \infty.$$

$$\sum_{k=1}^{\infty} \varphi(k)^{1/2} < \infty$$

$f \in L_2(\mathbb{R}^m, F^m)$ and $\sum\limits_{k_1=1}^{\infty} \dots \sum\limits_{k_m=1}^{\infty} |f_{k_1 \dots k_m}| < \infty$

(AC) The distribution of $(X_{j_1}, \dots, X_{j_m})$ is absolutely continuous with respect to the distribution of $(X_1^*, \dots, X_m^*)$

Theorem (Functional limit theorem for U-processes)

If the above conditions are met, then for every measurable functional $g(\cdot)$ in $D[0,1]$, continuous at points of $C[0,1]$ in the uniform topology, the sequence $g(U_n)$ converges in distribution to the random variable $g(U)$, where the random process $U(t)$ defined above and the corresponding multiple series converges almost surely for each $t \in [0,1]$ and is a.s. continuous in t.

Thank you for your attention!