Orthogonal series and asymptotic analysis of canonical U-statistics based on dependent observations

I. S. Borisov and N. V. Volodko
Sobolev Institute of Mathematics, Novosibirsk, 630090 Russia

Let $\{X_i;\ i \in \mathbb{Z}\}$ be a stationary sequence of r.v.'s. in a separable metric space $\mathfrak{X}$, and let P be the Borel distribution of X_1. We study normalized m-variate U-statistics

$$U_n(f) := n^{-m/2} \sum_{1 \le i_1 \ne \ldots \ne i_m \le n} f(X_{i_1}, ..., X_{i_m}), \quad m \ge 2,$$

where $f \in L_2(\mathfrak{X}^m, P^m)$ is *canonical*, i.e., $\mathbf{E} f(t_1, \ldots, t_{k-1}, X_k, t_{k+1}, \ldots, t_m) = 0$ for every $k \le m$ and all $t_j \in \mathfrak{X}$.

Let $\{e_i(t);\ i \ge 0\}$ be an orthonormal basis of the separable Hilbert space $L_2(\mathfrak{X})$ such that $e_0(t) \equiv 1$. Every canonical kernel from $L_2(\mathfrak{X}^m, P^m)$ admits the representation

$$f(t_1, ..., t_m) = \sum_{i_1, ..., i_m = 1}^{\infty} f_{i_1 \ldots i_m} e_{i_1}(t_1) \ldots e_{i_m}(t_m), \tag{1}$$

where the multiple series $L_2(\mathfrak{X}^m, P^m)$-converges. Notice that the multiple sum in (1) does not contain the element $e_0(t)$. We also introduce the following restriction on all m-dimensional distributions of $\{X_i\}$:
(AC) *For all natural $j_1 < \ldots < j_m$, the distribution of the vector $(X_{j_1}, ..., X_{j_m})$ is absolutely continuous relative to the distribution P^m.*

Theorem. *Let $\{X_i\}$ be a φ-mixing stationary sequence with $\sum_k \varphi^{1/2}(k) < \infty$. Let a canonical kernel $f(t_1, ..., t_m)$ satisfy the conditions: $\sum_{i_1, ..., i_m = 1}^{\infty} |f_{i_1 \ldots i_m}| < \infty$ and $\sup_i \mathbb{E}|e_i(X_1)|^m < \infty$. If, in addition, (**AC**) is fulfilled then*

$$U_n(f) \xrightarrow{d} \sum_{i_1, ..., i_m = 1}^{\infty} f_{i_1 \ldots i_m} \prod_{j=1}^{\infty} H_{\nu_j(i_1, ..., i_m)}(\tau_j), \tag{2}$$

where the multiple series in (2) converges almost surely, $\{\tau_i\}$ is a Gaussian sequence of centered random variables with covariance matrix

$$\mathbb{E}\tau_k \tau_l = \mathbb{E} e_k(X_1) e_l(X_1) + \sum_{j=1}^{\infty} \big[\mathbb{E} e_k(X_1) e_l(X_{j+1}) + \mathbb{E} e_l(X_1) e_k(X_{j+1})\big],$$

$\nu_j(i_1, ..., i_m) := \sum_{k=1}^{m} \delta_{i_k, j}$ (here $\delta_{i,j}$ is the Kronecker symbol), and $H_k(x) := (-1)^k e^{x^2/2} \frac{d^k}{dx^k} e^{-x^2/2}$ are the Hermite polynomials.

In the case of i.i.d. $\{X_i\}$, an analog of this theorem was proved by H. Rubin and R. A. Vitale (*Ann. Statist.*, 1980, **8**(1), 165-170). Notice that, in general, the restriction of the Theorem on the coefficients $\{f_{i_1 \ldots i_m}\}$ as well as condition (**AC**) cannot be omitted.

We also discuss the Hoeffding-type exponential inequalities for the distribution tails of the canonical U-statistics under consideration.