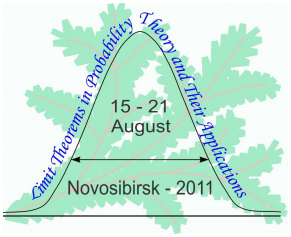

Asymptotic expansion for the distributions of canonical U-statistics

N.V. Volodko

Sobolev Institute of Mathematics of the Russian Academy of Sciences
630090 Novosibirsk, RUSSIA. E-mail: nvolodko@gmail.com

The second term of the asymptotic expansion is obtained for the distribution function of a canonical U-statistic of third order. The result is mainly based on the two papers by I. S. Borisov and E. A.'Solov'ev (1993) and by I. S. Borisov and N. V. Volodko (2008).

Let $\{X_i\}$ be a sequence of i.i.d. random variables in a measurable space $\mathfrak{X}$. Denote

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

where $\mathbf{E}f^2(X_1,...,X_d) < \infty$ and $\mathbf{E}f(t_1,\ldots,t_{k-1},X_k,t_{k+1},\ldots,t_d) = 0$ for every $k \le d$ and all $t_j \in \mathfrak{X}$; in this case, the kernel $f(t_1,...,t_d)$ and U-statistic $U_n(f)$ are called *canonical*. Let $\{e_i(t);\, i \ge 0\}$ be an orthonormal basis in the separable Hilbert space $L_2(\mathfrak{X})$ such that $e_0(t) \equiv 1$. Then every canonical kernel from $L_2(\mathfrak{X}^d, P^d)$ admits the representation

$$f(t_1,...,t_d) = \sum_{i_1,\ldots,i_d=1}^{\infty} f_{i_1\ldots i_d} e_{i_1}(t_1)\ldots e_{i_d}(t_d),$$

where the multiple series $L_2(\mathfrak{X}^d, P^d)$-converges.

Let $x \in \mathbb{R}^\infty$. Introduce the notation (well defined under the conditions below)

$$\|x\| = \big(\sum_{i,j,k} |f_{ijk}|(|x_i|^3 + |x_j|^3 + |x_k|^3)\big)^{1/3}, \quad \xi_i = \{e_k(X_i)\}_k, \quad \sigma^2 := \mathbb{E}\|\xi_1\|^2.$$

Theorem. *Let the following conditions be fulfilled:*
(I) $\sum_{i,j,k=1}^{\infty} |f_{ijk}| < \infty$;
(II) $\sup_i \mathbb{E}e_i^6(X_1) < \infty$,
(III) $\forall i$ *the infinite-dimensional vector* $\{f_{iik}\}_k$ *can not be represented as a linear combination of vectors* $\{f_{ljk}\}_k$, *where* $\min(l,j) < i$.

Then

$$\mathbb{P}\big(U_n(f) \le z\big) = \mathbb{P}\big(F(\tau) \le z\big) + \frac{1}{6\sigma\sqrt{n}} \mathbb{E} g_z^{(3)}(0)[\xi_1^3] + R_n(z),$$

where $\tau = \{\tau_i\}$ *is a sequence of independent random variables with the standard normal distribution,*

$$F(\tau) = \sum_{i\ne j\ne k} f_{ijk}\tau_i\tau_j\tau_k + \sum_{i\ne j} f_{iij}(\tau_i^2 - 1)\tau_j + \sum_i f_{iii}(\tau_i^3 - 3\tau_i),$$

and the above multiple series converge almost surely, $g_z^{(3)}(x)[\cdot]$ *is the third Frechet derivative of* $g_z(x) := \mathbb{P}(F(\tau + x) < z)$ *in the Banach space* $l_f^3 := \{x \in \mathbb{R}^\infty : \|x\| < \infty\}$ *(under the conditions of Theorem this derivative exists), and, for any* $\nu > 0$,

$$\sup_{z\in\mathbb{R}} |R_n(z)| \le C(\nu, f) n^{-1+\nu}.$$