BULL. AUSTRAL. MATH. SOC. 60F05
VOL. 21 (1980), 1-5. (60F99)

RATES OF CONVERGENCE FOR U-STATISTICS WITH VARYING KERNELS

N.C. Weber

Let U_n be a U-statistic whose kernel depends on the size n of the sample under consideration. It is shown that when U_n is suitably normalised its distribution function differs in L_p norm from the distribution function of a standard normal variable by a term of $O\left(n^{-\frac{1}{2}}\right)$.

1. Introduction and notation

U-statistics with kernels which depend on the sample size as well as the sample values have attracted some attention recently, particularly in connection with the analysis of spatial data. In this paper an L_p rate of convergence for the central limit theorem is given for these statistics.

Let $\{X_n\}$ be a sequence of independent and identically distributed random variables and let $\{h_n\}$ be a sequence of real valued functions, $h_n : R^m \to R$, which are symmetric in their arguments. Define

$$U_n = \binom{n}{m}^{-1} \Sigma_n' h_n\left(X_{i_1}, X_{i_2}, \ldots, X_{i_m}\right) ,$$

where Σ_n' denotes summation over $1 \le i_1 < i_2 < \ldots < i_m \le n$. Then $\{U_n\}$ are the U-statistics of order m with kernel sequence $\{h_n\}$ based on

Received 9 May 1979.

$\{X_n\}$.

As an example of such statistics consider $X_1, X_2, \ldots$ independent random variables uniformly distributed over the unit square in R^2 . Let d denote the Euclidean metric in R^2 and let

$$h_n(X_i, X_j) = I\left(d(X_i, X_j) \leq \beta n^{-\alpha}\right) = 1 \text{ if } d(X_i, X_j) \leq \beta n^{-\alpha}$$
$$= 0 \text{ otherwise,}$$

where $\alpha > 0$ and $\beta > 0$. Given a set of n observations on the unit square, U_n is the proportion of pairs of points which are within distance $\beta n^{-\alpha}$ of each other. This statistic has been suggested as a basis for testing the hypothesis that the points are placed on the unit square independently against an alternative which favours clustering or repulsion. The asymptotic behaviour of this statistic for various values of α has been considered by Kester [4] and Silverman and Brown [5].

Without loss of generality suppose that $Eh_n(X_1, X_2, \ldots, X_m) = 0$ and set $\rho_n = E[h_n(X_1, X_2, \ldots, X_m)h_n(X_1, X_{m+1}, \ldots, X_{2m-1})]$. Let

$$g_n(X_i) = E[h_n(X_i, X_{j_1}, X_{j_2}, \ldots, X_{j_{m-1}})|X_i] , \quad i \neq j_1, j_2, \ldots, j_{m-1}$$

and let

$$Y_{i_1,i_2,\ldots,i_m} = h_n(X_{i_1}, \ldots, X_{i_m}) - g_n(X_{i_1}) - g_n(X_{i_2}) \ldots - g_n(X_{i_m}) .$$

Notice that for each n , $\{g_n(X_i)\}$ is a sequence of independent and identically distributed random variables with $Eg_n(X_1) = 0$ and $E(g_n(X_1))^2 = \rho_n$. Let

$$\hat{U}_n = n^{-1} \sum_{j=1}^{n} g_n(X_j)$$

and

https://doi.org/10.1017/S0004972700011230 Published online by Cambridge University Press

$$\Delta_n = \left(n/m^2\rho_n\right)^{\frac{1}{2}}\left(U_n-\hat{U}_n\right)$$
$$= \left(n/m^2\rho_n\right)^{\frac{1}{2}}\binom{n}{m}^{-1}\Sigma_n' Y_{i_1,i_2,\ldots,i_m} .$$

Let $\Phi(x)$ be the distribution function of a standard normal variable.

2. Results

The first bound that we shall consider is the L_∞ bound or Berry-Esseen bound for U-statistics with varying kernels.

THEOREM 1. *Given* $\{X_i\}$ *and a sequence of kernels* $\{h_n\}$ *with* $\rho_n > 0$ *then*

$$\sup_x \left|P\left(\sqrt{n}U_n \le m\rho_n^{\frac{1}{2}}x\right)-\Phi(x)\right| \le C_m n^{-\frac{1}{2}}\rho_n^{-3/2}E\left|h_n\left(X_1,\ X_2,\ \ldots,\ X_m\right)\right|^3 ,$$

where C_m *is a function of* m *and does not depend on* n .

Proof. This result follows by arguing as in Callaert and Janssen [*1*] since for fixed n , $\{g_n(X_i)\}$ are independent and identically distributed random variables and so the only changes required in Callaert and Janssen's proof is that the moments in the various inequalities now depend on n .

Using Theorem 1 we now prove the general L_p result.

THEOREM 2. *Given* $\{X_i\}$ *and a sequence of kernels* $\{h_n\}$ *with* $\rho_n > 0$ *then for* $p \ge 1$,

$$\left(\int_{-\infty}^{\infty} \left|P\left(\sqrt{n}U_n \le m\rho_n^{\frac{1}{2}}x\right)-\Phi(x)\right|^p dx\right)^{1/p} \le C_m n^{-\frac{1}{2}}\rho_n^{-3/2}E\left|h_n\left(X_1,\ \ldots,\ X_m\right)\right|^3 ,$$

where C_m *is a function of* m *and does not depend on* n .

Proof. For an L_p-integrable function f ,

$$\left(\int |f(x)|^p dx\right)^{1/p} \le \left(\sup_x |f(x)|\right)^{(p-1)/p}\cdot\left(\int |f(x)|dx\right)^{1/p}$$

and so in view of Theorem 1, the result will follow once we establish the result for $p = 1$.

From Lemma 4.1 (a) of Erickson and Koul [3] we have that

$$(1)\quad \int_{-\infty}^{\infty}\left|P\left(\sqrt{n}U_n \le m\rho_n^{\frac{1}{2}}x\right)-\Phi(x)\right|dx \le \int_{-\infty}^{\infty}\left|P\left(\sqrt{n}\hat{U}_n \le m\rho_n^{\frac{1}{2}}x\right)-\Phi(x)\right|dx + E^{\frac{1}{2}}\left(\Delta_n^2\right)$$

and from Theorem 1 of Erickson [2],

$$(2)\quad \int_{-\infty}^{\infty}\left|P\left(\sqrt{n}\hat{U}_n \le m\rho_n^{\frac{1}{2}}x\right)-\Phi(x)\right|dx \le Cm^{-3}\rho_n^{-3/2}n^{-\frac{1}{2}}E|g_n(X_1)|^3$$

$$\le Cm^{-3}\rho_n^{-3/2}n^{-\frac{1}{2}}E|h_n(X_1, X_2, \ldots, X_m)|^3$$

for some constant $C < 72$. Also,

$$(3)\quad E\Delta_n^2 \le C_m'n^{-1}\rho_n^{-1}E(Y_{1,2,\ldots,m})^2$$

$$\le C_m'n^{-1}\rho_n^{-3/2}E|h_n(X_1, \ldots, X_m)|^3 ,$$

where C_m' is a function of m but does not depend on n . The result follows by substituting (3) and (2) into (1) and noticing that

$$\rho_n^{-3/2}E|h_n(X_1, X_2, \ldots, X_m)|^3 \ge 1 .$$

References

[1] Herman Callaert and Paul Janssen, "The Berry-Esseen theorem for U-statistics", *Ann. Statist.* 6 (1978), 417-421.

[2] R.V. Erickson, "On an L_p version of the Berry-Esseen theorem for independent and m-dependent variables", *Ann. Prob.* 1 (1973), 497-503.

[3] R.V. Erickson and H.L. Koul, "L_1 rates of convergence for linear rank statistics", *Ann. Statist.* 4 (1976), 771-774.

[4] Adri Kester, "Asymptotic normality of the number of small distances between random points in a cube", *Stochastic Process. Appl.* 3 (1975), 45-54.

https://doi.org/10.1017/S0004972700011230 Published online by Cambridge University Press

[5] Bernard Silverman, Tim Brown, "Short distances, flat triangles and Poisson limits", *J. Appl. Probab.* **15** (1978), 815-825.

Department of Mathematical Statistics,
University of Sydney,
Sydney,
New South Wales 2006,
Australia.