U-statistics on Associated Random Variables

Tasos C. Christofides
Department of Mathematics and Statistics
University of Cyprus
P.O. Box 20537
CY 1678 Nicosia
CYPRUS
tasos@ucy.ac.cy

1. Introduction

Assume that $X_1, \ldots, X_n$ is a collection of random variables and for $m \leq n$ let $h : \Re^m \to \Re$ be a real valued function which is symmetric in its arguments i.e., invariant under permutations of arguments. We can define the U-statistic U_n as

$$U_n = \binom{n}{m}^{-1} \sum_c h(X_{i_1}, \ldots, X_{i_m})$$

where $\sum_c$ denotes summation over the $\binom{n}{m}$ combinations of m distinct elements $\{i_1, \ldots i_m\}$ from $\{1, \ldots, n\}$. In the case where the observations $X_1, \ldots, X_n$ are independent and identically distributed, we have the class of U-statistics introduced by Hoeffding (1948). For the theory and applications of U-statistics see for example Serfling (1980) or Lee (1990).

For application purposes it is important to depart from the assumption that the observations are independent. Various results can be found in the literature on U-statistics based on observations which are dependent. See for example Sen (1963) and Becker and Utev (2001). In this paper we consider U-statistics which are based on a collection of associated random variables and we consider their strong convergence.

2. A strong law of large numbers

The concept of associated random variables was first introduced by Esary, Proschan and Walkup (1967) and have found many applications especially in reliability. Many authors have studied this concept providing interesting results and applications.

Definition. A finite collection of random variables $X_1, \ldots, X_k$ is said to be associated if

$$Cov\{f(X_1, \ldots, X_k)g(X_1, \ldots, X_k)\} \geq 0$$

for any two coordinatewise nondecreasing functions f, g on $\Re^k$ such that the covariance is defined. An infinite collection is associated if every finite subcollection is associated.

Theorem. Let U_n be a U-statistic based on associated random variables and on the kernel h. Assume that h is a function of bounded variation in its m arguments and $E(h) = 0$. If for some $\nu > 1$

$$\sum_{k=1}^{\infty} [1 - (1 - m/(k+1))^{\nu}]\, E|U_k|^{\nu} < \infty$$

then

$$U_n \stackrel{a.s.}{\to} 0 \qquad\qquad n \to \infty.$$

The previous result requires that h is a function of bounded variation. This assumption is not very restrictive since the class of functions of bounded variation is quite large. Trivially all nondecreasing functions are included, but the class also includes other functions which are useful in the theory of U-statistics. For example the kernel $h(x, y) = |x - y|$ is a function of bounded variation.

REFERENCES

Becker, N.G. and Utev, S. (2001). Threshold results for U-statistics of dependent binary variables. *J. Theor. Prob.*, **14**, 97-114.

Esary, J., Proschan, F., Walkup, D. (1967). Association of random variables with applications. *Ann. Math. Statist.*, **38**, 1466-1474.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.*, **19**, 293-325.

Lee, A.J. (1990). *U-statistics, Theory and Methods. Marcel Dekker.*

Sen, P.K. (1963). On the properties of U-statistics when the observations are not independent. *Cal. Statist. Assoc. Bull.*, **12**, 69-92.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics. Wiley.*

RESUME

On considère les U-statistiques basées sur des variables aléatoires associées. Cette classe inclut les U-statistiques basées sur des variables indépendantes et identiquement distribuées. Cependant, les résultats connus pour les U-statistiques ne sont pas appliqués à cette classe. Dans cet article on présente la loi forte des grands nombres.