DEPARTMENT OF MATHEMATICS AND STATISTICS

Probability Inequalities and Related Asymptotic Results for U-statistics Based on Associated and Negatively Associated Random Variables

DOCTOR OF PHILOSOPHY DISSERTATION

Charalambos Charalambous

NICOSIA, CYPRUS

December 2021

DEPARTMENT OF MATHEMATICS AND STATISTICS

Probability Inequalities and Related Asymptotic Results for U-statistics Based on Associated and Negatively Associated Random Variables

Charalambos Charalambous

A Dissertation Submitted to the University of Cyprus in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

NICOSIA, CYPRUS

December 2021

VALIDATION PAGE

Doctoral Candidate: Charalambos Charalambous

Doctoral Thesis Title: Probability Inequalities and Related Asymptotic Results for U-statistics Based on Associated and Negatively Associated Random Variables

The present Doctoral Dissertation was submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in the ***Department of Mathematics and Statistics*** *and was approved on the 6th December, 2021 by the members of the* ***Examination Committee.***

Examination Committee:

Committee Chair: Anastassia Baxevani, Associate Professor,
Department of Mathematics and Statistics,
University of Cyprus

Committee Member: Fazekas Istvan, Professor,
Faculty of Informatics,
Department of Applied Mathematics and Probability Theory,
University of Debrecen

Committee Member: Rita Giuliano, Professor,
Department of Mathematics,
University of Pisa

Committee Member: Sergios Agapiou, Assistant Professor,
Department of Mathematics and Statistics,
University of Cyprus

Research Supervisor: Tasos Christofides, Professor,
Department of Mathematics and Statistics,
University of Cyprus

DECLARATION OF DOCTORAL CANDIDATE

The present doctoral dissertation was submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy of the University of Cyprus. It is a product of original work of my own, unless otherwise mentioned through references, notes, or any other statements.

Charalambos Charalambous

Περίληψη

Η παρούσα διπλωματική εργασία εκπονήθηκε στα πλαίσια του διδακτορικού προγράμματος στατιστικής του Τμήματος Μαθηματικών και Στατιστικής του Πανεπιστημίου Κύπρου.

Η θεωρία των στατιστικών συναρτήσεων U έχει εισαχθεί από τον *W. Hoeffding* (1948). Πολλοί ερευνητές μελέτησαν και συνεχίζουν να μελετούν την ασυμπτωτική συμπεριφορά τους όταν αυτές κατασκευάζονται με ανεξάρτητες και ισόνομες παρατηρήσεις. Στην πιο πρόσφατη βιβλιογραφία όμως παρουσιάζει μεγάλο ενδιαφέρον η μελέτη αυτών των στατιστικών συναρτήσεων όταν αυτές κατασκευάζονται με εξαρτημένες παρατηρήσεις. Ένα είδος εξάρτησης είναι η θετική και αρνητική σύνδεση (*association*), οι οποίες έχουν εισαχθεί από τους *Esary et al.* (1967) και *Joag-Dev and Proschan* (1983) αντίστοιχα.

Ο στόχος της παρούσας διατριβής επικεντρώνεται κυρίως στην ασυμπτωτική συμπεριφορά των στατιστικών συναρτήσεων U που κατασκευάζονται με θετικά ή αρνητικά συνδεδεμένες τυχαίες μεταβλητές. Για τον σκοπό αυτό αποδεικνύονται άνω φράγματα ροπών και χρήσιμες εκθετικές, μεγιστικές ανισότητες που αποτελούν βασικά εργαλεία για την απόδειξη ασυμπτωτικών αποτελεσμάτων. Εισάγουμε την έννοια των *reverse demimartingales* και *reverse demisubmartingales* που είναι γενίκευση των *reverse martingales* και *reverse submartingales* αντίστοιχα, και αποδεικνύουμε μεγιστικές ανισότητες τύπου *Chow* και *Doob.*

Η στενή σύνδεση μεταξύ *martingales*, *demimartingales*, *N-demimartingales* και των στατιστικών συναρτήσεων U αποτελούν επίσης μέρος της μελέτης μας. Τα ασυμπτωτικά αποτελέσματα που ισχύουν για *demimartingales* και *N-demimartingales* μπορούν να χρησιμοποιηθούν για την απόδειξη νόμων των μεγάλων αριθμών για στατιστικές συναρτήσεις U που βασίζονται σε εξαρτημένες τυχαίες μεταβλητές με πολυδιάστατους δείκτες καθώς και για στατιστικές συναρτήσεις U που βασίζονται σε πολλαπλά δείγματα.

Χρησιμοποιούμε επίσης ανισότητες και άλλα αποτελέσματα που ισχύουν για ανεξάρτητες και ισόνομες παρατηρήσεις για να βρούμε την απόσταση μεταξύ των στατιστικών

συναρτήσεων U που βασίζονται σε ανεξάρτητες και ισόνομες παρατηρήσεις και μιας τυχαίας μεταβλητής που ακολουθεί την τυποποιημένη κανονική κατανομή. Προφανώς τα αποτελέσματα που ισχύουν για ανεξάρτητες και ισόνομες παρατηρήσεις δεν ισχύουν και για την περίπτωση των εξαρτημένων παρατηρήσεων. Τα αποτελέσματα αυτά τροποποιούνται ή αντικαθίστανται εντελώς για να μελετηθεί η απόσταση μεταξύ των στατιστικών συναρτήσεων U με ανεξάρτητες και ισόνομες παρατηρήσεις και των στατιστικών συναρτήσεων U με εξαρτημένες τυχαίες μεταβλητές που έχουν την ίδια κατανομή. Παρουσιάζεται επίσης η ασυμπτωτική κανονικότητα των στατιστικών συναρτήσεων U που κατασκευάζονται από θετικά συνδεδεμένες τυχαίες μεταβλητές, όπως και η απόσταση μεταξύ των στατιστικών συναρτήσεων U που κατασκευάζονται από μια συλλογή ισόνομων και αρνητικά συνδεδεμένων τυχαίων μεταβλητών από μια τυχαία μεταβλητή που ακολουθεί την τυποποιημένη κανονική κατανομή χρησιμοποιώντας τη μετρική *Zolotarev*. Τα αποτελέσματα αυτά αποτελούν κεντρικά οριακά θεωρήματα τα οποία μελετώνται με μια εναλλακτική τεχνική και συγκεκριμένα με τη βοήθεια μετρικών αποστάσεων. Παράλληλα παρουσιάζονται αντίστοιχα αποτελέσματα για μια άλλη συγγενική κατηγορία στατιστικών συναρτήσεων, τις στατιστικές συναρτήσεις V, (*V-statistics*).

Τέλος εφαρμόζεται η τεχνική επαναδειγματοληψίας *Jackknife* σε στατιστικές συναρτήσεις U που βασίζονται σε θετικά ή αρνητικά συνδεδεμένες τυχαίες μεταβλητές με σκοπό την εκτίμηση του τυπικού σφάλματος και της μεροληψίας της εκτιμήτριας της διασποράς των συναρτήσεων αυτών.

Abstract

The basic theory of U-statistics was developed by W. Hoeffding (1948). U-statistics are generalized averages and include among others the sample mean and the unbiased sample variance as special cases. Detailed expositions of the general topic may be found in Denker (1985), Lee (1990). See also Fraser (1957) Chapter 6, Serfling (1980) Chapter 5, and Lehmann (1999) Chapter 6. The closely related class of V-statistics has been introduced by von Mises (1947).

U-statistics were originally defined on i.i.d. observations and many authors study their asymptotic behavior. However, many authors have also studied U-statistics based on dependent observations since the theoretical results which are valid for U-statistics based on i.i.d. random variables cannot automatically be applied to the case of U-statistics based on dependent random variables. One type of dependence is association (negative or positive). Positively associated, or simply associated random variables were introduced by Esary et al. (1967) and negative association was introduced by Joag-Dev and Proschan (1983).

Our study is mainly focused on the asymptotic behavior of U-statistics based on associated and negatively associated random variables. Although some results have been established, the conditions imposed are restrictive and in some cases unrealistic. Our aim is to study the asymptotic behavior under conditions which are applicable and verifiable. Among our objectives in this thesis, is to prove moment and exponential inequalities for this type of U-statistics. We introduce the concept of a reverse demimartingale and a reverse demisubmartingale as a generalization of the notion of reverse (backward) martingales and reverse submartingales, and we establish Chow and Doob type maximal inequalities.

The close connection between martingales, demimartingales, N-demimartingales and U-statistics is fully exploited. The asymptotic results derived from demimartingales and N-demimartingales can be applied to U-statistics, to obtain strong laws for

U-statistics based on multidimensionally indexed associated random variables and multisample U-statistics on collections of associated random variables that are introduced for the first time, as a natural generalization of one sample U-statistics.

We also use tools such as inequalities and results valid for U-statistics on i.i.d. observations, to find the distance between U-statistics on i.i.d. observations and a normal random variable. It is obvious that results proved for the classical setup (i.e. for i.i.d. observations) which are not applicable for the case of associated observations are modified or replaced altogether by results on associated random variables. The distance between U-statistics on i.i.d. observations and U-statistics on identically distributed associated random variables having the same distribution is also investigated and exploited. Asymptotic normality for U-statistics based on associated random variables is also presented. We also investigate the distance between a U-statistic based on a collection of identically distributed negatively associated random variables and a normal random variable using the Zolotarev's ideal metric. Those results also provide a central limit theorem for U-statistics with an alternative technique using probability metrics. Finally, it is natural that we also investigate another related class of statistics, the von Mises statistics or V-statistics. Corresponding results are also proved for this type of statistics.

Finally, jackknifing U-statistics based on associated and negatively associated random variables is also part of our study.

Acknowledgments

First and foremost, I should like to express my heartfelt thanks to Prof. Tasos Christofides, for his guidance, encouragement and support during the running of my PhD study. The meetings, conversations and his feedback were vital in inspiring me to complete this thesis.

I would also like to thank the members of my committee: Prof. Fazekas Istvan, Prof. Rita Giuliano, Associate Prof. Anastassia Baxevani and Assistant Prof. Sergios Agapiou for their careful reading of this thesis and their valuable comments.

Moreover, I would like to express my gratitude to my family and friends whose support has been an important contribution in finishing this work.

Contents

Chapter 1

Introduction

1.1 Literature review and outline of the thesis

According to Lee (1990) the class of U-statistics is important for at least three reasons. First, a great many statistics in common use are in fact members of this class, so that the theory provides a unified paradigm for the study of the distributional properties of many well-known test statistics and estimators, particularly in the field of non-parametric statistics. Second, the simple structure of U-statistics makes them ideal for studying general estimation processes such as bootstrapping and jackknifing, and for generalizing those parts of asymptotic theory concerned with the behavior of the sequence of sample means. Third, application of the theory often generates new statistics useful in particular estimation problems.

U-statistics were originally defined on i.i.d. observations. However, many authors have studied U-statistics based on dependent observations. Among them Sen (1963), Nandi and Sen (1963), Serfling (1968), Denker and Keller (1983) and Becker and Utev (2001). One notion of dependence is the concept of association. Since independent random variables are associated, the class of U-statistics introduced by Hoeffding (1948) is included as a special case in the class of U-statistics constructed on associated random variables. Since many of the theoretical results which are valid for Hoeffding's (1948) U-statistics depend on the assumption of independence, the theory known cannot automatically be applied to the case of U-statistics based on associated random variables.

A few authors have studied the limiting behavior of U-statistics based on associated observations. Dewan and Prakasa Rao (2001) established a central limit theorem for U-statistics based on stationary associated random variables using an orthogonal expansion for the underlying kernel. Dewan and Prakasa Rao (2002) and its corrigendum Dewan and Prakasa Rao (2015) give a central limit theorem for U-statistics using Hoeffding's decomposition with continuous component-wise monotonic kernels of degree two. Furthermore, Christofides (2004) studied this class of U-statistics in a different way, using the concept of demimartingales. He has shown the connection between one sample U-statistics and demimartingales and also has proved a strong law of large numbers for U-statistics based on associated random variables in the case where the kernel of the statistic belongs to a large family of functions called *kernels of bounded variation*. Garg and Dewan (2015) obtained the limiting distribution of U-statistics based on kernels of bounded Hardy-Krause variation when the underlying sample consists of stationary associated observations. Continuing with the investigations, Garg and Dewan (2018) discussed a central limit theorem for U-statistics based on associated random variables on differentiable kernels of degree two or higher.

Some other authors have studied the limiting behavior of U-statistics based on negatively associated random variables. Huang and Zhang (2006) studied the asymptotic normality of those U-statistics, when the U-statistic is degenerate or non-degenerate. Budsaba et al. (2009) established the Marcinkiewicz-Zygmund type strong laws of large numbers for certain class of multilinear U-statistics based also on negatively associated random variables.

Our study mainly focuses on the asymptotic behavior of U-statistics based on observations which are dependent and specifically on observations which are associated or negatively associated. Also we investigate U-statistics based on multidimensionally indexed associated random variables and multisample U-statistics on collections of associated random variables that are introduced for the first time, as a natural generalization of one sample U-statistics. Jackknifing U-statistics based on associated and negatively associated random variables is also part of our study. Finally, it is natural that we also investigate another related class of statistics, the von Mises statistics or V-statistics.

This thesis is organized as follows. In Chapter 1 we present the necessary literature review, the definitions and some auxiliary results of associated and negatively associated random variables with some properties included. The concept of demimartingales, N-demimartingales are also presented and we define one sample U-statistics with associated and negatively associated random variables.

In Chapter 2, we discuss some inequalities for U-statistics based on associated and negatively associated random variables. Generally, in probability theory, moment and exponential inequalities play an important role in various proofs of limit theorems. In particular they provide a measure of convergence rate for the strong law of large numbers. Moment inequalities for sums of associated random variables were studied by Birkel (1988) and Oliveira (2012). Exponential inequalities for negatively associated random variables were obtained by Kim and Kim (2007), Nooghabi and Azarnoosh (2009), Xing et al. (2009), Sung (2009), Xing and Shanchao (2010). The chapter is organized as follows. In Section 2.1 we establish moment inequalities for U-statistics and V-statistics based on negatively associated random variables. In Section 2.2 we present an exponential bound for U-statistics based on the same type of random variables. Wang and Hu (2009) generalized the results of Christofides (2000) for demimartingales and demisubmartingales. In Section 2.3 we introduce the concept of a reverse demimartingale and a reverse demisubmartingale as a generalization of the notion of reverse (backward) martingales and reverse submartingales. Chow (1960) proved a maximal inequality for submartingales. Christofides (2000) showed that Chow's inequality is valid for the more general glass of demisubmartingales. In this chapter we give a Chow type maximal inequality for reverse demisubmartingales and we establish a Doob's maximal inequality for reverse demisubmartingales. Finally, we show the connection between U-statistics based on associated random variables and reverse demimartingales and we give some examples.

Christofides (2004) established a strong law of large numbers for U-statistics based on associated random variables. In Chapter 3 we introduced the definition of U-statistics on associated multidimensionally indexed random variables and multisample U-statistics on collections of associated random variables. We focus on their connection with multidimensionally indexed demimartingales, and we establish strong laws for this type of U-statistics.

Probability metrics play an important role in asymptotic statistics. Generally speaking, a probability metric is a functional that measures the distance between two random quantities and are very useful in investigating the asymptotic behavior of a statistical function or estimator. The metric approach to problems on the accuracy of approximations of distributions appeared in the theory of probability in the mid 1930s. In Chapter 4, we use some metrics that are commonly found in probability and statistics. One useful metric utilized in this paper is the so called Zolotarev's ideal metric (Zolotarev (1983)). In Section 4.1, we give the distance between a U-statistic U_n based on associated random variables and a U-statistic U_n^* based on i.i.d. random variables under Zolotarev's ideal metric. Asymptotic normality for U-statistics based on associated random variables is also presented in Section 4.2, with an alternative way to prove asymptotic normality for this type of U-statistics to the approach of Garg and Dewan (2015). We also investigate the distance between a U-statistic based on a collection of independent identically distributed random variables with a distribution function F, and a normal random variable using the Zolotarev's ideal metric. Sharakhmetov (2004) proved limit theorems for U-statistics using the mean metric κ_1. In Section 4.3, based on Sharakhmetov (2004) we improve his results using a higher order metric, the Zolotarev's ideal metric. It is worth mentioning that limit theorems for U-statistics are usually considered for the uniform (Kolmogorov) metric (see Serfling (1980) or Korolyuk and Borovskikh (1989)). The rate of convergence in the central limit theorem in the form of a uniform Berry-Esseen bound for U-statistics has been investigated among others by Filippova (1962), Grams and Serfling (1973), Bickel (1974), Chan and Wierman (1977), Callaert and Janssen (1978), van Zwet (1984) and Friedrich (1989). Our goal is to depart from the approach of a uniform Berry-Esseen bound for U-statistics and provide an alternative approach for the distance between a U-statistic and a normal random variable. This result is used to establish a "central limit theorem" for U-statistics. Corresponding results are also investigated for von Mises statistics. Finally in Section 4.3 we discuss an alternative way to prove central limit theorems for U-statistics based on dependent random variables to the approach of Garg and Dewan (2015, 2018) and Huang and Zhang (2006), using Zolotarev's ideal metric.

In Chapter 5 we apply the jackknife technique on U-statistics based on associated and negatively associated random variables. The jackknife is a technique used to estimate the variance and bias of a large population. This resampling method was originally proposed by Quenouille (1949) as a method of reducing the bias of an estimator of a serial correlation coefficient. The same author expanded the technique in Quenouille (1956) and explored its general bias reduction properties in an infinite-population context. Later, the technique was refined and given its current name by Tukey (1958). Tukey (1958) described its use in constructing confidence limits for a large class of estimators. In the case of U-statistics, this concept has been studied by few authors in the past. In particular, the problem of estimating the standard error of U-statistics was first considered by Arvesen (1969) although an equivalent formulation appears in Sen (1960). Majumdar and Sen (1978) studied the invariance principles for jackknifing U-statistics for finite population sampling. Krewski (1978) applied the jackknifing technique on U-statistics in finite populations. Yamato, Toda and Nomachi (2007) investigated the jackknifing method on a convex combination of one-sample U-statistics. Chapter 5 is organized in two sections. In Section 5.1 we jackknife U-statistics based on associated random variables and in Section 5.2 we jackknife U-statistics based on negatively associated random variables.

Finally, in Chapter 6 we discuss our future research plan which can be initiated based on the results presented in this thesis.

1.2 Associated and negatively associated random variables

The basic concepts of association, U-statistics and demimartingales are crucial for our investigation. We briefly introduce each one and offer some elementary examples.

Positively associated, or simply associated random variables are of considerable interest in reliability theory, percolation theory and statistical mechanics. For a review of several probabilistic and statistical inferential results for associated sequences, see for example Newman (1984), Cox and Grimmett (1984), Birkel (1988), Birkel (1989), Roussas (1993), Matula (1998), Roussas (1999) and Dewan and Prakasa Rao (2001).

All random variables appearing in this thesis are defined on a probability space $(\Omega, \mathcal{A}, \mathcal{P})$. The definition of associated random variables is given below.

Definition 1.2.1. *Let $\{X_i, i \geq 1\}$ be a sequence of random variables. Every finite collection $\{X_1, X_2, ..., X_n\}$ is said to be associated if for any real valued, coordinatewise nondecreasing functions $f, g : \mathbb{R}^n \to \mathbb{R}$, we have*

$$cov\left[f\left(X_1, X_2, ..., X_n\right), g\left(X_1, X_2, ..., X_n\right)\right] \geq 0$$

whenever the covariance is defined. An infinite collection is associated if every finite subcollection is associated.

A weaker concept of association is presented in the following definition.

Definition 1.2.2. *Let $\{X_i, i \geq 1\}$ be a sequence of random variables. A finite collection $\{X_1, X_2, ..., X_n\}$ is said to be weakly associated if for any nonempty disjoint subsets A and B of $\{1, ..., n\}$ and for any real valued, coordinatewise nondecreasing functions $f : \mathbb{R}^{|A|} \to \mathbb{R}$ and $g : \mathbb{R}^{|B|} \to \mathbb{R}$, where $|A|$ is the cardinality of A, we have*

$$cov\left[f\left(X_i, i \epsilon A\right), g\left(X_j, j \epsilon B\right)\right] \geq 0.$$

An infinite collection is weakly associated if every finite subcollection is weakly associated.

Associated random variables were introduced by Esary et al. (1967). Some properties of association are the following:

- Any subset of associated random variables is a set of associated random variables.
- If two sets of associated random variables are independent of one another, then their union is a set of associated random variables.
- The set consisting of a single random variable is associated.
- Nondecreasing functions of associated random variables are associated.
- Independent random variables are associated.

The covariance structure of an associated sequence $\{X_i, i \geq 1\}$ presented below, plays a significant role in studying the probabilistic properties of the associated sequence.

Notation 1.2.3. *(Oliveira (2012), p. 41). Let $\{X_i, i \in \mathbb{N}\}$, be a sequence of random variables. Denote*

$$u(n) = \sup_{k \in \mathbb{N}} \sum_{j:|j-k| \geq n} \operatorname{cov}\left(X_j, X_k\right), \quad n \geq 0.$$

Remark 1.2.4. *Notice that if we assume the random variables to be stationary, then*

$$u(n) = 2 \sum_{j=n+1}^{\infty} \operatorname{cov}\left(X_1, X_j\right), \quad n \geq 0.$$

One can recognize this expression as the asymptotic variance in central limit theorems for dependent variables if we choose $n = 0$.

Birkel (1988) observed that moment bounds for partial sums of associated sequences also depend on the rate of decrease of $u(n)$.

Proposition 1.2.5. *(Birkel (1988)). Let $\{X_j, j \in \mathbb{N}\}$, be a sequence of associated random variables satisfying $E(X_j) = 0$ for every j and*

$$\sup_{j\in\mathbb{N}} E\left|X_j\right|^{r+\delta} < \infty \quad for\ some\ r > 2,\ \delta > 0.$$

Assume

$$u(n) = O\left(n^{-\frac{(r-2)(r+\delta))}{2\delta}}\right).$$

Then there is a constant B not depending on n such that for all $n \in \mathbb{N}$

$$\sup_{m\geq 0} E\left|\sum_{j=m+1}^{n+m} X_j\right|^r \leq Bn^{\frac{r}{2}}.$$

The definition of negatively associated random variables is given below.

Definition 1.2.6. *Let $\{X_i, i \geq 1\}$ be a sequence of random variables. A finite collection $\{X_1, X_2, ..., X_n\}$ is said to be negatively associated (NA) if for any nonempty disjoint subsets A and B of $\{1, ..., n\}$ and for any real valued, coordinatewise nondecreasing bounded functions $f : \mathbb{R}^{|A|} \to \mathbb{R}$ and $g : \mathbb{R}^{|B|} \to \mathbb{R}$, where $|A|$ is the cardinality of A, we have*

$$cov\left[f\left(X_i, i \epsilon A\right), g\left(X_j, j \epsilon B\right)\right] \leq 0.$$

An infinite collection is negatively associated if every finite subcollection is negatively associated.

Negative association was introduced by Joag-Dev and Proschan (1983). Some properties of negatively associated random variables are the following.

- A subset of two or more NA random variables is a subset of NA random variables.
- A set of independent random variables is a set of NA random variables.
- Increasing functions defined on disjoint subsets of a set of NA random variables are NA.
- The union of independent sets of NA random variables is a set of NA random variables.

We present now some important moment bounds for sums of negatively associated random variables.

Proposition 1.2.7. *(Shao (2000)). Let* $p \geq 1$, $\{X_i,\ 1 \leq i \leq n\}$ *be a collection of negatively associated mean zero random variables with* $E|X_i|^p < \infty$ *for every* $1 \leq i \leq n$, *and let* $\{X_i^*,\ 1 \leq i \leq n\}$ *be a collection of independent random variables such that* X_i *and* X_i^* *have the same distribution for each* $1 \leq i \leq n$. *Then*

$$E\left|\sum_{i=1}^{n} X_i\right|^p \leqslant E\left|\sum_{i=1}^{n} X_i^*\right|^p.$$

Proposition 1.2.8. *(Shao (2000)). Let* $\{X_i,\ 1 \leq i \leq n\}$ *be a collection of negatively associated mean zero random variables with* $E|X_i|^p < \infty$ *for every* $1 \leq i \leq n$ *and* $1 < p \leq 2$. *Then*

$$E\left|\sum_{i=1}^{n} X_i\right|^p \leqslant 2^{3-p}\sum_{i=1}^{n} E\left|X_i\right|^p.$$

Proposition 1.2.9. *(Su et al. (1997)). Let* $\{X_i,\ 1 \leq i \leq n\}$ *be a collection of negatively associated mean zero random variables and* $E|X_i|^p < \infty$, *for* $i = 1, ..., n$ *and for* $p \geq 2$. *Then there exists a positive constant* C_p *which only depends on* p *such that*

$$E\left|\sum_{i=1}^{n} X_i\right|^p \leqslant C_p n^{p/2-1}\sum_{i=1}^{n} E\left|X_i\right|^p,$$

where $C_p = \max\left\{p^p, p^{1+\frac{p}{2}} e^p B(\frac{p}{2}, \frac{p}{2})\right\}$ *with* $B(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx$.

1.3 Demimartingales and N-demimartingales

Relevant to the notion of positively associated random variables is the notion of demimartingales. Below we give the definition.

Definition 1.3.1. *Let* $S_1, S_2, ...$ *be an* L^1 *sequence of random variables. Assume that for* $j = 1, 2, ...$

$$E\left\{(S_{j+1} - S_j)\, f\left(S_1, S_2, ..., S_j\right)\right\} \geq 0$$

for all coordinatewise nondecreasing functions f *such that the expectation is defined. Then* $\{S_j\}_{j\geq 1}$ *is called a demimartingale. If in addition the function* f *is assumed to be nonnegative, the sequence* $\{S_j\}_{j\geq 1}$ *is called a demisubmartingale.*

One can easily verify that the partial sum of mean zero positively associated random variables is a demimartingale. It is worth mentioning that a martingale with the natural choice of σ-algebras is a demimartingale. Furthermore, it can also be verified that a submartingale (with the natural choice of σ-algebras) is a demisubmartingale.

Motivated by the definition of a demimartingale, the idea of a similar generalization for negatively associated random variables leads to the concept of the so-called N-demimartingales and N-demisupermartingales. The definition is as follows.

Definition 1.3.2. *Let $S_1, S_2, ...$ be an L^1 sequence of random variables. Assume that for $j = 1, 2, ...$*

$$E\left\{(S_{j+1} - S_j)\, f\left(S_1, S_2, ..., S_j\right)\right\} \leq 0$$

for all coordinatewise nondecreasing functions f whenever the expectation is defined. Then the sequence $\left\{S_j\right\}_{j\geq 1}$ is called an N-demimartingale. If the inequality holds for nonnegative coordinatewise nondecreasing functions f, then the sequence $\left\{S_j\right\}_{j\geq 1}$ is called an N-demisupermartingale.

It is trivial to verify that the partial sum of mean zero negatively associated random variables is an N-demimartingale. It is also worth mentioning that a martingale with the natural choice of σ-algebras is an N-demimartingale. Furthermore, it can be verified that a supermartingale (with the natural choice of σ-algebras) is an N-demisupermartingale.

Newman and Wright (1982) introduced the concept of a demimartingale and a demisubmartingale as a generalization of martingales and submartingales respectively. For a review of some probabilistic results see Christofides (2000), Wang (2004), Wang et al. (2009, 2010) and Prakasa Rao (2012). The notion of N-demimartingales was introduced later by Christofides (2003). Various results and examples of N-demimartingales and N-demisupermartingales can be found in Christofides (2003), Prakasa Rao (2004, 2007), Hadjikyriakou (2010) and Wang et al. (2011).

1.4 U-statistics based on associated and negatively associated random variables

U-statistics were introduced by Hoeffding (1948) following an idea of Halmos (1946). They are generalized averages containing some classical statistics as special cases such as the sample mean and the sample variance. In what follows we give the definition of U-statistics defined not on i.i.d. random variables as in the original construction of Hoeffding but on associated random variables.

Definition 1.4.1. *Let $\{X_1, X_2, ..., X_n\}$ be a collection of identically distributed associated (or negatively associated) random variables. In a large class of problems, parameters to be estimated are of the form $\vartheta = E\left[h\left(X_1, X_2, ..., X_m\right)\right]$ where m is a positive integer $m \leq n$ and h is a symmetric mapping from $\mathbb{R}^m$ to $\mathbb{R}$ called a "kernel". An unbiased estimator of ϑ is*

$$U_n = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_m \leq n} h\left(X_{i_1}, X_{i_2}, ..., X_{i_m}\right)$$

where $\sum_{1 \leq i_1 < i_2 < \cdots < i_m \leq n}$ denotes summation over all $\binom{n}{m}$ combinations of the m distinct elements $\{i_1, i_2, ..., i_m\}$ from $\{1, ..., n\}$. The estimator U_n is called a U-statistic based on the kernel h and the given observations.

Some closely related statistics are the V-statistics. A V-statistic (von Mises (1947)) based on the symmetric kernel h of degree m is defined by

$$V_n = n^{-m} \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} h\left(X_{i_1}, X_{i_2}, ..., X_{i_m}\right).$$

The next proposition shows the asymptotic connection beetween U and V-statistics.

Proposition 1.4.2. *(Prakasa Rao (2012), p. 180). Let $\{X_n, n \geq 1\}$ be a stationary associated sequence. Let U_n and V_n be the U-statistic and the V-statistic respectively based on these observations and on a symmetric kernel h of degree 2. Assume that $h\left(X_1, X_2\right)$ is monotonic in X_1. Furthermore, suppose that*

$$E\left[\left|h\left(X_1, X_2\right)\right|^{r+\delta}\right] < \infty \quad for \quad r > 2,\ \delta > 0,$$

and

$$2 \sum_{j=n+1}^{\infty} \operatorname{cov}\left(h\left(X_1, X_2\right), h\left(X_{2j-1}, X_{2j}\right)\right) = O\left(n^{-\frac{(r-2)(r+\delta))}{2\delta}}\right).$$

Then

$$E\left|U_n\right|^r = O\left(n^{-\frac{r}{2}}\right), \qquad n \to \infty,$$

and

$$E\left|U_n - V_n\right|^r = O\left(n^{-\frac{r}{2}}\right), \qquad n \to \infty.$$

Suppose that $\{X_1, X_2, ..., X_n\}$ is a collection of associated (or negatively associated) random variables identically distributed with distribution function F. Below we present some examples of U-statistics.

Example 1.4.3. *If $m = 1$, U_n is simply the sample mean. Consider the estimation of $\vartheta = \mu^m$, where $\mu = E(X_1)$ and m is a positive integer. Using $h(x_1, x_2, ..., x_m) = x_1 x_2 \cdots x_m$, we obtain the following U-statistic as an unbiased estimator of $\vartheta = \mu^m$:*

$$U_n = \binom{n}{m}^{-1} \sum_{1 \le i_1 < i_2 < \cdots < i_m \le n} X_{i_1} X_{i_2} \cdots X_{i_m}.$$

Example 1.4.4. *Consider the estimation of $\vartheta = \sigma^2 = Var(X_1)$. Since*

$$\sigma^2 = [Var(X_1) + Var(X_2)]/2 = E[(X_1 - X_2)^2/2],$$

we obtain the following U-statistic with kernel $h(x_1, x_2) = (x_1 - x_2)^2/2$:

$$U_n = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \frac{(X_i - X_j)^2}{2} = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) = S^2$$

which is the sample variance.

Example 1.4.5. *We obtain the following U-statistic with kernel $h(x_1) = \mathbb{I}_{\{x_1 \le x\}}$ where $\mathbb{I}_{\{.\}}$ is the indicator function:*

$$U_n = \binom{n}{1}^{-1} \sum_{i=1}^{n} h(X_i) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{\{X_i \le x\}} = \hat{F}(x)$$

which is the empirical distribution function.

Example 1.4.6. *In some cases, we would like to estimate $\vartheta = E|X_1 - X_2|$, a measure of concentration. Using the kernel $h(x_1, x_2) = |x_1 - x_2|$, we obtain the following U-statistic as an unbiased estimator of $\vartheta = E|X_1 - X_2|$:*

$$U_n = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} |X_i - X_j|$$

which is known as Gini's mean difference. This statistic has a number of applications in studying the income of human populations.

Example 1.4.7. *Let* $\vartheta = P(X_1+X_2 \leq 0)$*. Using the kernel* $h(x_1, x_2) = \mathbb{I}_{(-\infty,0]}(x_1+x_2)$*, we obtain the following U-statistic:*

$$U_n = \frac{2}{n(n-1)} \sum_{1\leq i<j\leq n} \mathbb{I}_{(-\infty,0]}(X_i + X_j)$$

which is known as the one-sample Wilcoxon statistic.

1.5 Hoeffding decomposition of a U-statistic based on a kernel of degree two

The Hoeffding decomposition was introduced by Hoeffding (1961). This result is very useful in providing asymptotic results for U-statistics. Next we present the Hoeffding decomposition of a U-statistic based on a kernel of degree two.

Let $\{X_n,\ n \geq 1\}$ be a sequence of identically distributed associated or negatively associated random variables. Define the U-statistic of dimension two by

$$U_n = \binom{n}{2}^{-1} \sum_{1\leq i_1<i_2\leq n} h\left(X_{i_1}, X_{i_2}\right)$$

where h is a real-valued function symmetric in its arguments. Furthermore, the von Mises statistic V_n of dimension two is defined by

$$V_n = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} h\left(X_i, X_j\right).$$

One can express V_n in terms of U_n in the form

$$V_n = \frac{n-1}{n} U_n + \frac{1}{n^2} \sum_{i=1}^{n} h\left(X_i, X_i\right).$$

Let

$$\theta = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x,y) dF(x) dF(y),$$

$$h_1(x_1) = E[h(x_1, X_2)] = \int_{-\infty}^{+\infty} h(x_1, x_2)\, dF(x_2),$$

$$h^{(1)}(x_1) = h_1(x_1) - \theta$$

and

$$h^{(2)}(x_1, x_2) = h(x_1, x_2) - h_1(x_1) - h_1(x_2) + \theta.$$

Then, the Hoeffding decomposition (H-decomposition) for U_n is given by (see Lee, 1990)

$$U_n = \theta + 2H_n^{(1)} + H_n^{(2)},$$

where $H_n^{(j)}$ is the U-statistic of degree j based on the kernel $h^{(j)}$, $j = 1, 2$, that is,

$$H_n^{(1)} = \frac{1}{n}\sum_{i=1}^{n} h^{(1)}(X_i) \quad and \quad H_n^{(2)} = \frac{1}{\binom{n}{2}} \sum_{1 \le i_1 < i_2 \le n} h^{(2)}(X_{i_1}, X_{i_2}).$$

The well-known H-decomposition is described in Hoeffding (1961). The importance of this decomposition is that a U-statistic can be expressed as a sum of two new uncorrelated U-statistics. To simplify our calculation let also $E(h(X_1, X_2)) = 0$, in short, $E(h) = 0$. Using the Hoeffding decomposition we can write

$$\frac{n^{\frac{1}{2}} U_n}{2\sigma_1} = \frac{1}{n^{\frac{1}{2}}\sigma_1}\sum_{i=1}^{n} h^{(1)}(X_i) + \frac{1}{n^{\frac{1}{2}}(n-1)\sigma_1} \sum_{1 \le i < j \le n} h^{(2)}(X_i, X_j), \tag{1.5.1}$$

where $\sigma_1^2 = \text{Var}[h_1(X_1)] < \infty$. A similar Hoeffding decomposition for a von Mises statistic of dimension two is given by

$$V_n = \frac{1}{n^2}\sum_{i=1}^{n} h(X_i, X_i) + \frac{2(n-1)}{n^2}\sum_{i=1}^{n} h^{(1)}(X_i) + \frac{2}{n^2}\sum_{1 \le i < j \le n} h^{(2)}(X_i, X_j),$$

and thus

$$\frac{n^{\frac{1}{2}} V_n}{2\sigma_1} = \frac{1}{2n^{\frac{3}{2}}\sigma_1}\sum_{i=1}^{n} h(X_i, X_i) + \frac{n-1}{n^{\frac{3}{2}}\sigma_1}\sum_{i=1}^{n} h^{(1)}(X_i) + \frac{1}{n^{\frac{3}{2}}\sigma_1}\sum_{1 \le i < j \le n} h^{(2)}(X_i, X_j).$$

We now present the Hoeffding-decomposition for some U-statistics. Let $\{X_n,\ n \ge 1\}$ be a sequence of identically distributed associated or negatively associated random variables with $E(X_1) = \mu_1$, $E(X_1^2) = \mu_2$ and $\text{Var}(X_1) = \sigma^2$.

Example 1.5.1. *(Estimator of second moment). Consider the estimation of* $\theta = E(X_1^2) = \sigma^2 + \mu^2$. U_n *is based on the kernels*

$$\begin{aligned} h(x_1, x_2) &= x_1 x_2, \\ h^{(1)}(x_1) &= \mu x_1 - (\sigma^2 + \mu^2), \\ h^{(2)}(x_1, x_2) &= x_1 x_2 - \mu(x_1 + x_2) + (\sigma^2 + \mu^2). \end{aligned}$$

Example 1.5.2. *(Estimator of variance). Consider the estimation of* $\theta = Var(X_1) = \sigma^2$. U_n *is based on the kernels*

$$\begin{aligned} h(x_1, x_2) &= \frac{1}{2}(x_1 - x_2)^2, \\ h^{(1)}(x_1) &= \frac{1}{2}x_1^2 - \mu x_1 + \frac{1}{2}(\mu^2 - \sigma^2), \\ h^{(2)}(x_1, x_2) &= \mu(x_1 + x_2) - x_1 x_2 - \mu^2. \end{aligned}$$

Example 1.5.3. *(Estimator of third central moment). Consider the estimation of the third central moment* $\theta = E(X_1 - \mu_1)^3$. *Then*

$$U_n = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} (X_i - \bar{X}_n)^3 = \binom{n}{3}^{-1} \sum_{1 \le i_1 < i_2 < i_3 \le n} h\left(X_{i_1}, X_{i_2}, X_{i_3}\right)$$

is based on the kernels

$$\begin{aligned} h(x_1, x_2, x_3) &= \frac{x_1^3 + x_2^3 + x_3^3}{3} - \frac{x_1^2(x_2 + x_3) + x_2^2(x_1 + x_3) + x_3^2(x_1 + x_2)}{2} + 2x_1 x_2 x_3, \\ h^{(1)}(x_1) &= \frac{2\mu_3 + x_1^3}{3} - \mu_2 x_1, \\ h^{(2)}(x_1, x_2) &= \frac{\mu_3 + x_1^3 + x_2^3}{3} - \frac{x_1^2 x_2 + x_2^2 x_1 + \mu_2(x_1 + x_2)}{2}. \end{aligned}$$

where $E(X_1^2) = \mu_2$ *and* $E(X_1^3) = \mu_3$, *if we assume that* $E(X_1) = 0$.

Example 1.5.4. *(Wilcoxon's one sample rank statistic). Recall Wilcoxon's one sample test, which is used to test if a distribution* F *is symmetric about zero. Let* $\vartheta = P(X_1 + X_2 > 0)$. *Using the kernel* $h(x_1, x_2) = \mathbb{I}_{\{x_1 + x_2 > 0\}}$, *we obtain the following U-statistic:*

$$\hat{\vartheta} = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \mathbb{I}_{\{X_i + X_j > 0\}}$$

which is known as the one-sample Wilcoxon statistic. The Hoeffding-decomposition for

$\hat{\vartheta}$ is based on kernels

$$
\begin{aligned}
&h(x_1, x_2) = \mathbb{I}_{\{x_1 + x_2 > 0\}}, \\
&h_1(x_1) = E[\mathbb{I}_{\{x_1 + X_2 > 0\}}] = 1 - P(X_1 \leq -x_1), \\
&h^{(1)}(x_1) = 1 - P(X_1 \leq -x_1) - \vartheta, \\
&h^{(2)}\left(x_1, x_2\right) = h\left(x_1, x_2\right) - h_1\left(x_1\right) - h_1\left(x_2\right) + \vartheta.
\end{aligned}
$$

Example 1.5.5. *(Estimator Gini's mean difference). Gini's mean difference is an index that measures the variability for observations from a distribution F. Assume that we have a finite population with N elements and a sample with $n < N$ observations is drawn without replacement. An unbiased estimator of*

$$
\theta = \iint |x - y| dF(x) dF(y),
$$

is the U-statistic

$$
U_n = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} |X_i - X_j|.
$$

We rewrite U_n in the form

$$
U_n = \binom{n}{2}^{-1} \sum_{j=1}^{n} (2j - n - 1) X_{j:n},
$$

where $X_{1:n} \leqslant \ldots \leqslant X_{n:n}$ are the order statistics of the observations of the sample. Denote $a_i = \frac{2i-N}{N}$ and $\Delta_i = X_{i+1} - X_i$ with $1 \leq i \leq N-1$. The Hoeffding decomposition of U_n is

$$
U_n = \theta + \sum_{k=1}^{n} h_1\left(X_k\right) + \sum_{1 \leq k < l \leq n} h_2\left(X_k, X_l\right)
$$

with

$$
h_1\left(X_k\right) = -\frac{2}{n} \frac{N}{N-2} \sum_{i=1}^{N-1} \left(I_{\{i \geq k\}} - \frac{i}{N} \right) a_i \Delta_i, \quad \textit{for } 1 \leq k \leq N
$$

$$
h_2\left(X_k, X_l\right) = -\frac{4}{n(n-1)} \sum_{i=1}^{N-1} \Phi_{k,l}\left(i\right) \Delta_i, \quad \textit{for } 1 \leq k < l \leq N
$$

where

$$\Phi_{k,l}(i) = \begin{cases} i(i-1)/A, & \text{if } 1 \le i \le k \\ -(i-1)(N-i-1)/A, & \text{if } k \le i \le l \\ (N-i-1)(N-i)/A, & \text{if } l \le i \le k \end{cases}$$

with $A = (N-1)(N-2)$.

1.6 Demimartingale and N-demimartingale approach

Classical U-statistics based on independent random variables can be expressed in terms of martingales. This follows from the H-decomposition that we described in Section 1.5. In the case of U-statistics which are constructed using a collection of associated random variables this result is not true. However, for a special class of kernels, a U-statistic can be expressed in terms of a demimartingale as the following result shows.

Proposition 1.6.1. *(Christofides (2004)). Let* U_n *be a U-statistic based on associated random variables and on the kernel* h*. Assume that* h *is componentwise nondecreasing and* $E(h) = 0$*. Then* $\left\{S_n = \binom{n}{m} U_n,\ n \ge m\right\}$ *is a demimartingale.*

Proof. We can write

$$\begin{aligned} S_{n+1} - S_n &= \sum_{1 \le i_1 < \cdots < i_m \le n+1} h\left(X_{i_1}, ..., X_{i_m}\right) - \sum_{1 \le i_1 < \cdots < i_m \le n} h\left(X_{i_1}, ..., X_{i_m}\right) \\ &= \sum_{1 \le i_1 < \cdots < i_{m-1} \le n} h\left(X_{i_1}, ..., X_{i_{m-1}}, X_{n+1}\right). \end{aligned}$$

Then for f componentwise nondecreasing function

$$\begin{aligned} &E\{(S_{n+1} - S_n) f(S_m, ..., S_n)\} \\ &= E\left\{ \sum_{1 \le i_1 < \cdots < i_{m-1} \le n} h\left(X_{i_1}, ..., X_{i_{m-1}}, X_{n+1}\right) f(S_m, ..., S_n) \right\} \\ &= \sum_{1 \le i_1 < \cdots < i_{m-1} \le n} E\{h\left(X_{i_1}, ..., X_{i_{m-1}}, X_{n+1}\right) f(S_m, ..., S_n)\} \\ &= \sum_{1 \le i_1 < \cdots < i_{m-1} \le n} E\{h\left(X_{i_1}, ..., X_{i_{m-1}}, X_{n+1}\right) g(X_1, ..., X_n)\} \\ &\ge 0, \end{aligned}$$

where the function g defined as

$$g(x_1,...,x_n) = f(h(x_1,...,x_m), \sum_{1\le i_1<\cdots<i_m\le m+1} h\left(x_{i_1},...,x_{i_m},\right),..., \sum_{1\le i_1<\cdots<i_m\le n} h\left(x_{i_1},...,x_{i_m}\right))$$

is componentwise nondecreasing since h, f are componentwise nondecreasing and the last inequality follows from the nondecreasingness of g and the fact that $\{X_i\}_{i\geq 1}$ is a sequence of associated random variables. ■

The following proposition from Prakasa Rao (2012) shows that a U-statistic based on negatively associated random variables and having a specific structure can be expressed in terms of an N-demimartingale for a particular class of kernels.

Proposition 1.6.2. *Suppose that $\{X_i, i \geq 1\}$ is a sequence of negatively associated random variables. For any fixed integer $m \leq n$, let $h\left(x_1, x_2, ..., x_m\right) = \widetilde{h}\left(x_1\right)\widetilde{h}\left(x_2\right)\cdots\widetilde{h}\left(x_m\right)$ be a kernel mapping $\mathbb{R}^m$ to $\mathbb{R}$ for some nondecreasing function $\widetilde{h}\left(.\right)$ with $E\left[\widetilde{h}\left(X_1\right)\right] = 0$. Then the sequence $\left\{S_n = \binom{n}{m}U_n, n \geq m\right\}$ is an N-demimartingale.*

More results and examples for N-demimartingales can be found in Prakasa Rao (2012).

1.7 Probability metrics and distances

Probability metrics play an important role in asymptotic statistics. Generally speaking, a probability metric is a functional that measures the distance between two random quantities and is very useful for investigating the asymptotic behavior of a statistical function or estimator. The definitions of probability and ideal probability metrics are given below.

Definition 1.7.1. *A probability metric $\mu(X,Y)$ is a functional which measures the closeness between the random variables X and Y, and satisfies the following three properties:*

Property 1. $\mu(X,Y) \geq 0$ *for any* $X,\ Y$ *and* $\mu(X,X) = 0$.

Property 2. $\mu(X,Y) = \mu(Y,X)$ *for any* $X,\ Y$.

Property 3. $\mu(X,Y) \leq \mu(X,Z) + \mu(Z,Y)$ *for any* $X,\ Y,\ Z$.

The three properties are called the identity axiom, the symmetry axiom, and the triangle inequality, respectively.

The ideal probability metrics are probability metrics which satisfy two additional properties which make them uniquely positioned to study problems related to the generalized central limit theorem (CLT). The two additional properties are the homogeneity of order r property and the regularity property.

The homogeneity property is

Property 4. $\mu(cX,cY) = |c|^r \mu(X,Y)$ *for any* $X,\ Y$ *and constants* $c \in \Re$ *and* $r \in \Re$.

The regularity property is

Property 5. $\mu(X+Z,Y+Z) \leq \mu(Y,X)$ *for any* $X,\ Y$ *and* Z *independent of* X *and* Y.

Next we give some various metrics that are needed for our study.

Definition 1.7.2. *Let* X,Y *random variables. The uniform or Kolmogorov distance is defined as*

$$\rho(X,Y) = \sup_{x \in \Re} \left| F_X(x) - F_Y(x) \right|.$$

Definition 1.7.3. *Let* $s \in \mathbb{N}$. *For two random variables* X *and* Y *denote by* κ_s *the mean metric, that is,*

$$\kappa_s(X,Y) = s \int |t|^{s-1} \left| F_X(t) - F_Y(t) \right| dt.$$

Definition 1.7.4. *Let* X,Y *random variables. The Levy metric is defined as*

$$L(X,Y) = \inf \left\{ \varepsilon > 0 : F_X(x-\varepsilon) - \varepsilon \leqslant F_Y(x) \leqslant F_X(x+\varepsilon) + \varepsilon, \quad \forall\ x \in \Re \right\}.$$

An ideal probability metric is the *Zolotarev's ideal metric* introduced by Zolotarev (1983).

Definition 1.7.5. *The Zolotarev's ideal metric is defined as*

$$\zeta_s(X,Y) = \frac{1}{(s-1)!} \int_{-\infty}^{+\infty} \left| E(X-t)_+^{s-1} - E(Y-t)_+^{s-1} \right| dt, \ \ s \in \mathbb{N}$$

where $E|X|^{s-1} < \infty$, $E|Y|^{s-1} < \infty$ *and* $X_+ = \max\{0, X\}$.

1.8 Functions of bounded variation

The concept of functions of bounded variation is presented in the following definitions and propositions.

Definition 1.8.1. *A partition of an interval* $[a,b]$ *is a set of points* $\{x_0, x_1, ..., x_n\}$ *such that* $a = x_0 < x_1 < \cdots < x_n = b$.

Definition 1.8.2. *Let* $f : [a,b] \to \mathbb{R}$ *be a function and* $\Pi = \{x_0, x_1, ..., x_n\}$ *a partition of the interval* $[a,b]$. *We denote*

$$C_{\Pi}(f) = \sum_{k=1}^{n} |f(x_k) - f(x_{k-1})|$$

and set

$$C^f_{[a,b]} = \sup_{\Pi} C_{\Pi}(f),$$

where the supremum is taken over all partitions Π *of the interval* $[a,b]$.

Definition 1.8.3. *A function* $f : [a,b] \to \mathbb{R}$ *is said to be of bounded variation on* $[a,b]$ *if* $C^f_{[a,b]}$ *is finite.*

Similarly, we present the concept of functions of bounded variation on a rectangle $[a,b] \times [c,d]$.

Definition 1.8.4. *Let* $f : [a,b] \times [c,d] \to \mathbb{R}$ *be a function defined on the rectangle* $[a,b] \times [c,d]$, $\Pi_1 = \{x_0, x_1, ..., x_n\}$ *a partition of the interval* $[a,b]$ *and* $\Pi_2 = \{y_0, y_1, ..., y_m\}$ *a partition of the interval* $[c,d]$. *We denote*

$$C_{\Pi_1 \times \Pi_2}(f) = \sum_{i=1}^{n} \sum_{j=1}^{m} |f(x_i, y_j) - f(x_{i-1}, y_j) - f(x_i, y_{j-1}) + f(x_{i-1}, y_{j-1})|$$

and set

$$C^f_{[a,b] \times [c,d]} = \sup_{\Pi_1 \times \Pi_2} C_{\Pi_1 \times \Pi_2}(f),$$

where the supremum is taken over all possible subdivisions of the rectangle $[a,b] \times [c,d]$.

Definition 1.8.5. *A function* $f : [a,b] \times [c,d] \to \mathbb{R}$ *is said to be of bounded variation on the rectangle* $[a,b] \times [c,d]$ *if* $C^f_{[a,b] \times [c,d]}$ *is finite.*

We now consider some of the properties of functions of bounded variation on $[a, b]$ that are still valid on the rectangle $[a, b] \times [c, d]$.

Proposition 1.8.6. *Let f and g be functions of bounded variation on $[a, b]$ and let k be a constant. Then*
(1) f is bounded on $[a, b]$;
(2) f is of bounded variation on every closed subinterval of $[a, b]$;
(3) kf is of bounded variation on $[a, b]$;
(4) $f + g$ and $f - g$ are of bounded variation on $[a, b]$;
(5) fg is of bounded variation on $[a, b]$;
(6) if $1/g$ is bounded on $[a, b]$, then f/g is of bounded variation on $[a, b]$;
(7) if f is constant on $[a, b]$, then f is of bounded variation on $[a, b]$;
(8) if f is monotone on $[a, b]$, then f is of bounded variation on $[a, b]$.

Remark 1.8.7. *The results in Proposition 1.8.6 are still valid for functions f and g of bounded variation on the rectangle $[a, b] \times [c, d]$.*

Another very useful property is the fact that a function of bounded variation can be written as the difference of two increasing functions.

Proposition 1.8.8. *If $f : [a, b] \to \mathbb{R}$ is a function of bounded variation on $[a, b]$ then there exist two nondecreasing functions on $[a, b]$, f_1 and f_2, such that $f = f_1 - f_2$.*

Proposition 1.8.9. *If $f : [a, b] \times [c, d] \to \mathbb{R}$ is a function of bounded variation on $[a, b] \times [c, d]$ then there exist two nondecreasing functions on $[a, b] \times [c, d]$, f_1 and f_2, such that $f = f_1 - f_2$.*

Here are some examples of functions of bounded variation:
(1) $f(x) = c$, where c is a constant ;
(2) $f(x) = x$;
(3) $f(x, y) = xy$;
(4) $f(x, y) = x \pm y$;
(5) $f(x, y) = |x - y|$;
(6) $f(x, y) = \mathbb{I}_{\{x \geq y\}}$;
where $\mathbb{I}$ is the indicator function.

It is worth noticing that for some U-statistics, the functions h, $h^{(1)}$ and $h^{(2)}$ involved in the Hoeffding decomposition are functions of bounded variation. We present here a few examples. Let $\{X_1, X_2, ..., X_n\}$ be a collection of identically distributed negatively associated random variables with $E(X_1) = \mu$ and $Var(X_1) = \sigma^2$.

Example 1.8.10. *Consider the estimation of* $\theta = Var(X_1) = \sigma^2$, U_n *is based on the kernels*

$$h(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2,$$
$$h^{(1)}(x_1) = \frac{1}{2}x_1^2 - \mu x_1 + \frac{1}{2}(\mu^2 - \sigma^2),$$
$$h^{(2)}(x_1, x_2) = \mu(x_1 + x_2) - x_1x_2 - \mu^2.$$

One can verify that h, $h^{(1)}$ *and* $h^{(2)}$ *are functions of bounded variation.*

Example 1.8.11. *Consider the estimation of* $\theta = E(X_1^2) = \sigma^2 + \mu^2$. U_n *is based on the kernels*

$$h(x_1, x_2) = x_1x_2,$$
$$h^{(1)}(x_1) = \mu x_1 - (\sigma^2 + \mu^2),$$
$$h^{(2)}(x_1, x_2) = x_1x_2 - \mu(x_1 + x_2) + (\sigma^2 + \mu^2).$$

One can verify that h, $h^{(1)}$ *and* $h^{(2)}$ *are functions of bounded variation.*

Chapter 2

Inequalities for U-statistics based on associated and negatively associated random variables

In Chapter 2, we discuss some inequalities for U-statistics based on associated and negatively associated random variables. The chapter is organized as follows. In Section 2.1 we obtain moment inequalities for U-statistics based on negatively associated random variables and in Section 2.2 we present an exponential bound for U-statistics based on the same class of random variables. In Section 2.3 we introduce the concept of a reverse demimartingale and a reverse demisubmartingale as a generalization of the notion of reverse (backward) martingales and reverse submartingales. Furthermore, we give a Chow type maximal inequality for reverse demisubmartingales and we establish a Doob's maximal inequality for reverse demisubmartingales. Finally, we show the connection between U-statistics based on associated random variables and reverse demimartingales and we give some examples.

2.1 Moment inequalities

Moment inequalities are useful tools for studying asymptotic results in statistics. Below we give some inequalities for the partial sum of random variables. The following inequality is a crude one, valid for any random variables but can be used in cases where no other inequality is available.

Proposition 2.1.1. *Let* $\{X_i,\ 1 \leq i \leq n\}$ *be any random variables. Then*

$$E\left|X_1 + X_2 + \cdots + X_n\right|^l \leq n^{l-1} \sum_{i=1}^{n} E\left|X_i\right|^l, \quad l \geq 1.$$

The following inequality can be applied to the case of independent random variables.

Proposition 2.1.2. *(Petrov (1995), p. 62). Let* $X_1, X_2, ..., X_n$ *be independent random variables with zero means, and let* $p \geq 2$. *Then*

$$E\left|\sum_{i=1}^{n} X_i\right|^p \leq C(p)\, n^{\frac{p}{2}-1} \sum_{i=1}^{n} E\left|X_i\right|^p,$$

where $C(p)$ *is a positive constant depending only on* p.

2.1.1 Moment inequalities for U-statistics based on negatively associated random variables

A number of authors have studied moment inequalities for negatively associated random variables. The most interesting results are obtained in Shao (2000) and Su et al. (1997). However, moment inequalities for U-statistics based on negatively associated random variables are rarely discussed in the literature. In this section we give an upper bound for $E\left|U_n\right|^p$ when p is a real number $1 < p \leq 2$ or $p \geq 2$, when U_n is based on negatively associated random variables.

Lemma 2.1.3 that follows, provides a moment bound of a U-statistic based on negatively associated random variables.

Lemma 2.1.3. *Let* U_n *be a U-statistic based on negatively associated random variables and on the kernel* h *of dimension two. Assume that* h *is componentwise nondecreasing and* p *is a real number,* $p \geq 2$. *Further assume that* $E\left|h\left(X_i, X_j\right)\right|^p \leq c < \infty$ *for all* $1 \leq i < j \leq n$. *Then*

$$E\left|U_n\right|^p < 2C_p 3^{p-1} n^{-p}(n-1)^{p/2-2} \sum_{1 \leq i < j \leq n} E\left|h\left(X_i, X_j\right)\right|^p,$$

where $C_p = \max\left\{p^p, p^{1+\frac{p}{2}} e^p B(\frac{p}{2}, \frac{p}{2})\right\}$ *with* $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx$. *Moreover*

$$E\left|U_n\right|^p = O\left(n^{-\frac{p}{2}}\right), \qquad n \to \infty.$$

Proof. It can be easily verified that the sum $\sum_{1\leq i<j\leq n} h\left(X_i, X_j\right)$ can be written as a sum of $n-1$ blocks

$$\sum_{1\leq i<j\leq n} h\left(X_i, X_j\right) = \sum_{k=1}^{n-1} A_k, \tag{2.1.1}$$

where $A_k = \sum_{i=1}^{n-k} h(X_i, X_{i+k})$ for $k = 1, 2, ..., n-1$.

It is obvious that each block A_k when $k \geq \left\lfloor \frac{n+1}{2} \right\rfloor$, where $\lfloor x \rfloor$ is the floor function, is a sum of negatively associated random variables since increasing functions defined on disjoint subsets of a set of negatively associated random variables are negatively associated random variables (see Property 5 of Joag-Dev and Proschan (1983)). On the other hand, when $k < \left\lfloor \frac{n+1}{2} \right\rfloor$, this statement is not true. But we observe that each block A_k, when $k < \left\lfloor \frac{n+1}{2} \right\rfloor$, can be divided into two new blocks with each one now being a sum of negatively associated random variables. The result of this remark is that the sum in the right-hand side of (2.1.1) is a sum of blocks where each block $B_s,\ s = 1, 2, ..., m$ is a sum of negatively associated random variables and $m = \left\lfloor \frac{3(n-1)}{2} \right\rfloor$ denotes the number of the blocks. Thus,

$$\sum_{1\leq i<j\leq n} h\left(X_i, X_j\right) = \sum_{k=1}^{n-1} A_k = \sum_{s=1}^{m} B_s.$$

The p-th moment of the U-statistic can be written now as

$$E\left|U_n\right|^p = \binom{n}{2}^{-p} E\left|\sum_{s=1}^{m} B_s\right|^p \tag{2.1.2}$$

Applying now Proposition 2.1.1 we have that

$$E\left|U_n\right|^p \leq \binom{n}{2}^{-p} m^{p-1} \sum_{s=1}^{m} E\left|B_s\right|^p.$$

Since each block B_s $(s = 1, 2, ..., m)$ is a sum of negatively associated random variables from Proposition 1.2.9 we have that

$$E\left|U_n\right|^p \leq \binom{n}{2}^{-p} m^{p-1} \sum_{s=1}^{m} \left[C_p n_s^{p/2-1} \sum_{\substack{1\leq i<j\leq n\ : \\ h(X_i,X_j)\in B_s}} E\left|h\left(X_i, X_j\right)\right|^p \right],$$

where n_s is the number of elements of the s^{th} block. Since $n_s \leq n-1$ for all $n \geq 2$, we get that

$$E|U_n|^p \leq \binom{n}{2}^{-p} m^{p-1} C_p (n-1)^{p/2-1} \sum_{s=1}^{m} \sum_{\substack{1\leq i<j\leq n\ : \\ h(X_i,X_j)\in B_s}} E|h(X_i,X_j)|^p$$

$$< 2C_p 3^{p-1} n^{-p} (n-1)^{p/2-2} \sum_{1\leq i<j\leq n} E|h(X_i,X_j)|^p ,$$

where the positive constant C_p depends only on p.

Under the assumption of $E|h(X_i,X_j)|^p \leq c < \infty$ for all $1 \leq i < j \leq n$ it follows that

$$E|U_n|^p = O\left(n^{-\frac{p}{2}}\right), \quad n \to \infty. \ \blacksquare$$

Note that the previous result, Lemma 2.1.3, is still valid for U-statistics based on a kernel h of bounded variation as the next result shows because of the fact that a function of bounded variation can be written as the difference of two nondecreasing functions.

Corollary 2.1.4. *Let U_n be a U-statistic based on negatively associated random variables and on the kernel h of dimension two. Assume that h is a function of bounded variation and p is a real number, $p \geq 2$. Furthermore, assume that $E|h(X_i,X_j)|^p \leq c < \infty$ for all $1 \leq i < j \leq n$. Then*

$$E|U_n|^p = O\left(n^{-\frac{p}{2}}\right), \quad n \to \infty.$$

Proof.

$$E|U_n|^p = \binom{n}{2}^{-p} E\left|\sum_{1\leq i<j\leq n} h(X_i,X_j)\right|^p .$$

Since the kernel h is a function of bounded variation there are two nondecreasing functions h_1 and h_2 such that $h = h_1 - h_2$. Therefore U_n can be expressed as

$$U_n = U_n^{(1)} - U_n^{(2)},$$

where $U_n^{(1)}$ and $U_n^{(2)}$ are U-statistics based on the componentwise nondecreasing kernels

h_1 and h_2 respectively. Then from Proposition 2.1.1 and Lemma 2.1.3 it follows that

$$E\left|U_n\right|^p \leq 2^{p-1}E\left|U_n^{(1)}\right|^p + 2^{p-1}E\left|U_n^{(2)}\right|^p = O\left(n^{-\frac{p}{2}}\right), \qquad n\to\infty. \blacksquare$$

Lemma 2.1.5 that follows, provides a moment bound for a U-statistic based on negatively associated random variables when p is a real number $1 < p \leq 2$.

Lemma 2.1.5. *Let U_n be a U-statistic based on negatively associated random variables and on the kernel h of dimension two. Assume that h is componentwise nondecreasing and p is a real number $1 < p \leq 2$. Furthermore, assume that $E\left|h\left(X_i, X_j\right)\right|^p \leq c < \infty$ for all $1 \leq i < j \leq n$. Then*

$$E\left|U_n\right|^p < 3^{p-1}2^{4-p}n^{-p}(n-1)^{-1}\sum_{1\leq i<j\leq n} E\left|h\left(X_i, X_j\right)\right|^p,$$

and thus,

$$E\left|U_n\right|^p = O\left(n^{1-p}\right), \qquad n\to\infty.$$

Proof. Using the same steps as in the proof of Lemma 2.1.3 we have that

$$E\left|U_n\right|^p = \binom{n}{2}^{-p} E\left|\sum_{1\leq i<j\leq n} h\left(X_i, X_j\right)\right|^p = \binom{n}{2}^{-p} E\left|\sum_{s=1}^{m} B_s\right|^p,$$

where $m = \left\lfloor\frac{3(n-1)}{2}\right\rfloor$ denotes the number of the blocks.

Applying now Proposition 2.1.1 we have that

$$E\left|U_n\right|^p \leq \binom{n}{2}^{-p} m^{p-1}\sum_{s=1}^{m} E\left|B_s\right|^p.$$

Since every block B_s $(s = 1, 2, ..., m)$ is a sum of negatively associated random variables from Proposition 1.2.8 we have that

$$E\left|U_n\right|^p < \binom{n}{2}^{-p}\left[\frac{3(n-1)}{2}\right]^{p-1}\sum_{s=1}^{m}\left[2^{3-p}\sum_{\substack{1\leq i<j\leq n\,:\\ h(X_i,X_j)\in B_s}} E\left|h\left(X_i, X_j\right)\right|^p\right]$$

$$= 3^{p-1}2^{4-p}n^{-p}(n-1)^{-1}\sum_{1\leq i<j\leq n} E\left|h\left(X_i, X_j\right)\right|^p.$$

Under the assumption $E\left|h\left(X_i,X_j\right)\right|^p \leq c < \infty$ for all $1 \leq i < j \leq n$ it follows that

$$E\left|U_n\right|^p = O\left(n^{1-p}\right), \qquad n \to \infty. \blacksquare$$

Corollary 2.1.6. *Let U_n be a U-statistic based on negatively associated random variables and on the kernel h of dimension two. Assume that h is a function of bounded variation and p is a real number $1 < p \leq 2$. Furthermore, assume that $E\left|h\left(X_i,X_j\right)\right|^p \leq c < \infty$ for all $1 \leq i < j \leq n$. Then*

$$E\left|U_n\right|^p = O\left(n^{1-p}\right), \qquad n \to \infty.$$

Proof.

$$E\left|U_n\right|^p = \binom{n}{2}^{-p} E\left|\sum_{1\leq i<j\leq n} h\left(X_i,X_j\right)\right|^p.$$

Since the kernel h is a function of bounded variation there are two nondecreasing functions h_1 and h_2 such that $h = h_1 - h_2$. Therefore U_n can be expressed as

$$U_n = U_n^{(1)} - U_n^{(2)},$$

where $U_n^{(1)}$ and $U_n^{(2)}$ are U-statistics based on the componentwise nondecreasing kernels h_1 and h_2 respectively. Then from Proposition 2.1.1 and Lemma 2.1.5 it follows that

$$E\left|U_n\right|^p \leq 2^{p-1}E\left|U_n^{(1)}\right|^p + 2^{p-1}E\left|U_n^{(2)}\right|^p = O\left(n^{1-p}\right), \qquad n \to \infty. \blacksquare$$

2.1.2 Moment inequalities for V-statistics based on negatively associated random variables

The corresponding theorem for V-statistics can be proved similarly. Lemma 2.1.7 that follows, provides a bound for the p-th absolute moment of a V-statistic based on negatively associated random variables when p is a real number, $p \geq 2$.

Lemma 2.1.7. *Let V_n be a V-statistic based on negatively associated random variables and on the kernel h of dimension two. Assume that h is componentwise nondecreasing and p is a real number, $p \geq 2$. Furthermore, assume that $E\left|h\left(X_i,X_j\right)\right|^p \leq c < \infty$ for* *all* $1 \leq i \leq n$, $1 \leq j \leq n$. *Then*

$$E|V_n|^p < 2^p 3^{p-1} C_p n^{-2p}(n-1)^{3p/2-2} \sum_{1\leq i<j\leq n} E|h(X_i,X_j)|^p$$
$$+ 2^{p-1} C_p n^{-3p/2-1} \sum_{i=1}^{n} E|h(X_i,X_i)|^p,$$

where $C_p = \max\left\{p^p, p^{1+\frac{p}{2}} e^p B(\frac{p}{2},\frac{p}{2})\right\}$ *with* $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx$. *Moreover*

$$E|V_n|^p = O\left(n^{-\frac{p}{2}}\right), \quad n \to \infty.$$

Proof.

$$\begin{aligned} E|V_n|^p &= E\left|\frac{n-1}{n}U_n + \frac{1}{n^2}\sum_{i=1}^{n} h(X_i,X_i)\right|^p \\ &\leq 2^{p-1}E\left|\frac{n-1}{n}U_n\right|^p + 2^{p-1}E\left|\frac{1}{n^2}\sum_{i=1}^{n} h(X_i,X_i)\right|^p. \end{aligned} \tag{2.1.3}$$

Applying Proposition 1.2.9 and Lemma 2.1.3, since increasing functions defined on disjoint subsets of a set of negatively associated random variables are negatively associated random variables (see Property 5 of Joag-Dev and Proschan (1983)), we have that the first term of the RHS of (2.1.3) by Lemma 2.1.3 is $O\left(n^{-\frac{p}{2}}\right)$. Now consider the second term. Using Proposition 1.2.9

$$\begin{aligned} E\left|\frac{1}{n^2}\sum_{i=1}^{n} h(X_i,X_i)\right|^p &\leq C_p n^{-\frac{3p}{2}-1}\sum_{i=1}^{n} E|h(X_i,X_i)|^p \\ &= O\left(n^{-\frac{3p}{2}}\right). \end{aligned}$$

Combining the two, the result follows. ■

Remark 2.1.8. *The previous result, Lemma 2.1.7, is still valid for V-statistics based on a kernel h of bounded variation because a function of bounded variation can be written as the difference of two nondecreasing functions.*

Corollary 2.1.9. *Let* V_n *be a V-statistic based on negatively associated random variables and on the kernel h of dimension two. Assume that h is a function of bounded variation and p is a real number,* $p \geq 2$. *Furthermore, assume that*
$E|h(X_i,X_j)|^p \leq c < \infty$ *for all for all* $1 \leq i \leq n$, $1 \leq j \leq n$. *Then*

$$E|V_n|^p = O\left(n^{-\frac{p}{2}}\right), \quad n \to \infty.$$

Proof. Since h can be expressed as the difference of two nondecreasing functions h_1 and h_2, then V_n can be expressed as the difference of two V-statistics,

$$V_n = V_n^{(1)} - V_n^{(2)},$$

where $V_n^{(1)}$ and $V_n^{(2)}$ are the V-statistics based on the nondecreasing kernels h_1 and h_2 respectively. The result now follows using Lemma 2.1.7. ■

Lemma 2.1.10 that follows, provides a moment bound for a V-statistic based on negatively associated random variables for a real number $1 < p \leq 2$.

Lemma 2.1.10. *Let V_n be a V-statistic based on negatively associated random variables and on the kernel h of dimension two. Assume that h is componentwise nondecreasing and p is a real number, $1 < p \leq 2$. Furthermore, assume that $E\left|h\left(X_i, X_j\right)\right|^p \leq c < \infty$ for all $1 \leq i \leq n$, $1 \leq j \leq n$. Then*

$$\begin{aligned} E\left|V_n\right|^p < & \, 2^3 3^{p-1}(n-1)^{p-1} n^{-2p} \sum_{1 \leq i < j \leq n} E\left|h\left(X_i, X_j\right)\right|^p \\ & + 4n^{-2p} \sum_{i=1}^{n} E\left|h\left(X_i, X_i\right)\right|^p . \end{aligned}$$

and

$$E\left|V_n\right|^p = O\left(n^{1-p}\right), \qquad n \to \infty.$$

Proof.

$$\begin{aligned} E\left|V_n\right|^p &= E\left|\frac{n-1}{n} U_n + \frac{1}{n^2} \sum_{i=1}^{n} h\left(X_i, X_i\right)\right|^p \\ &\leq 2^{p-1} E\left|\frac{n-1}{n} U_n\right|^p + 2^{p-1} E\left|\frac{1}{n^2} \sum_{i=1}^{n} h\left(X_i, X_i\right)\right|^p . \end{aligned} \tag{2.1.4}$$

Applying Proposition 1.2.8 and Lemma 2.1.5, since increasing functions defined on disjoint subsets of a set of negatively associated random variables are negatively associated random variables (see Property 5 of Joag-Dev and Proschan (1983)), we have that the first term of the RHS of (2.1.4) by Lemma 2.1.5 is $O\left(n^{1-p}\right)$. Now consider the second term. Using Proposition 1.2.8

$$E\left|\frac{1}{n^2}\sum_{i=1}^{n}h\left(X_i,X_i\right)\right|^p \leq n^{-2p}2^{3-p}\sum_{i=1}^{n}E\left|h\left(X_i,X_i\right)\right|^p$$
$$= O\left(n^{1-2p}\right).$$

Combining the two, the result follows. ■

Remark 2.1.11. *The previous result, Lemma 2.1.10, is still valid for V-statistics based on a kernel h of bounded variation because a function of bounded variation can be written as the difference of two nondecreasing functions.*

Corollary 2.1.12. *Let V_n be a V-statistic based on negatively associated random variables and on the kernel h of dimension two. Assume that h is a function of bounded variation and p is a real number, $1 < p \leq 2$. Furthermore, for all $1 \leq i \leq n$ and $1 \leq j \leq n$ assume that $E\left|h\left(X_i,X_j\right)\right|^p \leq c < \infty$. Then*

$$E\left|V_n\right|^p = O\left(n^{1-p}\right), \qquad n \to \infty.$$

Proof. Since h can be expressed as the difference of two nondecreasing functions h_1 and h_2, then V_n can be expressed as the difference of two V-statistics,

$$V_n = V_n^{(1)} - V_n^{(2)},$$

where $V_n^{(1)}$ and $V_n^{(2)}$ are the V-statistics based on the nondecreasing kernels h_1 and h_2 respectively. The result now follows using Lemma 2.1.4. ■

2.2 An exponential inequality for U-Statistics based on negatively associated random variables

Generally, in probability theory, exponential inequalities play an important role in various proofs of limit theorems. In particular they provide a measure of convergence rate for the strong law of large numbers. Exponential inequalities for negatively associated random variables were obtained by Kim and Kim (2007), Nooghabi and Azarnoosh (2009), Xing et al. (2009), Sung (2009), Xing and Shanchao (2010).

Consider now the class of U-statistics which are based on a collection of negatively associated random variables. In this section, we establish an exponential inequality for identically distributed negatively associated random variables. First, we state some propositions required to prove the main result given in Theorem 2.2.4.

Proposition 2.2.1. *Let $x \in \mathbb{R}$. Then*

$$e^x \leq 1 + x + \frac{x^2}{2} e^{|x|}.$$

Proposition 2.2.2. *Let $a > 0,\ b > 0,\ r > 1$. Then*

$$ab \leq \frac{a^r}{r} + \frac{b^{r'}}{r'},$$

where $\frac{1}{r} + \frac{1}{r'} = 1$.

Lemma 2.2.3. *Let U_n be a U-statistic based on negatively associated random variables and on a kernel h of dimension two with $E(h) = 0$. Then for $r > 1$ and $t > 0$,*

$$E(e^{tU_n}) < exp\left\{ \frac{t^{2r}}{n^r} \frac{3^{2r-1}}{r2^{r-1}} \frac{C_{2r}}{n(n-1)} \sum_{1 \leq i < j \leq n} E\left|h\left(X_i, X_j\right)\right|^{2r} + \frac{(r-1)}{r} E\left(e^{\frac{tr}{r-1}|U_n|}\right) \right\}$$

where $C_p = \max\left\{p^p, p^{1+\frac{p}{2}} e^p B(\frac{p}{2}, \frac{p}{2})\right\}$ with $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx$.

Proof. Using Propositions 2.2.1 and 2.2.2 and the representation of U_n as a sum of blocks, which every block consists of negatively associated random variables we may write

$$E(e^{tU_n}) \leq 1 + E(tU_n) + E\left(\frac{t^2}{2} U_n^2 e^{t|U_n|}\right)$$

$$\leq 1 + \frac{t^{2r}}{r2^r} E(U_n)^{2r} + \frac{1}{r'} E\left(e^{tr'|U_n|}\right)$$

$$= 1 + \frac{t^{2r}}{r2^r} \binom{n}{2}^{-2r} E\left(\sum_{s=1}^{m} B_s\right)^{2r} + \frac{1}{r'} E\left(e^{tr'|U_n|}\right)$$

$$\leq 1 + \frac{t^{2r}}{r2^r} \binom{n}{2}^{-2r} m^{2r-1} \sum_{s=1}^{m} E\left|B_s\right|^{2r} + \frac{1}{r'} E\left(e^{tr'|U_n|}\right)$$

$$< 1 + \frac{t^{2r}}{r2^r} \binom{n}{2}^{-2r} \left[\frac{3(n-1)}{2} \right]^{2r-1} C_{2r} n^{r-1} \sum_{1 \le i < j \le n} E \left| h \left(X_i, X_j \right) \right|^{2r} + \frac{1}{r'} E \left(e^{tr'|U_n|} \right),$$

where the two inequalities follow form Propositions 2.1.1 and 1.2.9 respectively.

$$E(e^{tU_n}) < 1 + \frac{t^{2r}}{n^r} \frac{3^{2r-1}}{r2^{r-1}} \frac{C_{2r}}{n(n-1)} \sum_{1 \le i < j \le n} E \left| h \left(X_i, X_j \right) \right|^{2r} + \frac{1}{r'} E \left(e^{tr'|U_n|} \right)$$

$$\le exp \left\{ \frac{t^{2r}}{n^r} \frac{3^{2r-1}}{r2^{r-1}} \frac{C_{2r}}{n(n-1)} \sum_{1 \le i < j \le n} E \left| h \left(X_i, X_j \right) \right|^{2r} + \frac{(r-1)}{r} E \left(e^{\frac{tr}{r-1}|U_n|} \right) \right\},$$

where the last inequality follows from $1 + x \le e^x, \ \forall \ x \in \Re$. ■

The following theorem gives the exponential bound for U-Statistics based on negatively associated random variables.

Theorem 2.2.4. *Let U_n be a U-statistic based on negatively associated random variables and on the kernel h with $E(h) = 0$. Assume that $E \left(e^{\frac{tr}{r-1}|U_n|} \right) \le d$ for $r > 1$ and $t > 0$, where d is a positive constant. Then for $\varepsilon > 0$,*

$$P \left(U_n > \varepsilon \right) < exp \left\{ - \frac{\varepsilon^{\frac{2r}{2r-1}} n^{\frac{r}{2r-1}}}{\lambda^{\frac{1}{2r-1}}} \left(\frac{1}{(2r)^{\frac{1}{2r-1}}} - \frac{1}{(2r)^{\frac{2r}{2r-1}}} \right) + \frac{d(r-1)}{r} \right\}.$$

where

$$\lambda = \frac{3^{2r-1}}{r2^{r-1}} \frac{C_{2r}}{n(n-1)} \sum_{1 \le i < j \le n} E \left| h \left(X_i, X_j \right) \right|^{2r}.$$

Proof. Assume that $r > 1$ and $t > 0$. Applying Markov's inequality and Lemma 2.2.3 we have that

$$\begin{aligned} P \left(U_n > \varepsilon \right) &= P \left(e^{tU_n} > e^{t\varepsilon} \right) \\ &\le e^{-t\varepsilon} E \left(e^{tU_n} \right) \\ &= exp \left\{ -t\varepsilon + \frac{t^{2r}}{n^r} \frac{3^{2r-1}}{r2^{r-1}} \frac{C_{2r}}{n(n-1)} \sum_{1 \le i < j \le n} E \left| h \left(X_i, X_j \right) \right|^{2r} + \frac{1}{r'} d \right\} \\ &= exp \left\{ g(t) \right\}, \end{aligned}$$

where $g(t) = -t\varepsilon + \frac{t^{2r}}{n^r}\lambda + \frac{d}{r'}$. The function $g(t)$ is minimized at

$$t^* = \left(\frac{\varepsilon n^r}{2r\lambda} \right)^{\frac{1}{2r-1}}.$$

Then

$$g(t^*) = -\frac{\varepsilon^{\frac{2r}{2r-1}} n^{\frac{r}{2r-1}}}{\lambda^{\frac{1}{2r-1}}} \left(\frac{1}{(2r)^{\frac{1}{2r-1}}} - \frac{1}{(2r)^{\frac{2r}{2r-1}}} \right) + \frac{d(r-1)}{r}$$

$$= -n^{\frac{r}{2r-1}} \left(\frac{r 2^{r-1} \varepsilon^{2r}}{3^{2r-1} C_{2r} \frac{1}{n(n-1)} \sum_{1 \le i < j \le n} E \left| h\left(X_i, X_j\right) \right|^{2r}} \right)^{\frac{1}{2r-1}} \left(\frac{1}{(2r)^{\frac{1}{2r-1}}} - \frac{1}{(2r)^{\frac{2r}{2r-1}}} \right)$$

$$+ \frac{d(r-1)}{r}.$$

Hence we have that

$$P\left(U_n > \varepsilon\right) < exp \left\{ -\frac{\varepsilon^{\frac{2r}{2r-1}} n^{\frac{r}{2r-1}}}{\lambda^{\frac{1}{2r-1}}} \left(\frac{1}{(2r)^{\frac{1}{2r-1}}} - \frac{1}{(2r)^{\frac{2r}{2r-1}}} \right) + \frac{d(r-1)}{r} \right\}. \quad \blacksquare$$

Example 2.2.5. *Let $\{X_n, n \ge 1\}$ be a sequence of i.i.d. standard normal random variables. Then $\{X_n, n \ge 1\}$ are negatively associated random variables with $E\left(e^{\delta|X_1|}\right) < \infty$ for and real number δ. Let U_n be a U-statistic based on the previous random variables with $E(h) = 0$. Then for $\delta = \frac{tr}{r-1} > 0$, U_n satisfies the conditions of Theorem 2.2.4.*

2.3 Reverse Demimartingales and N-demimartingales

In this section we introduce the concept of reverse demimartingales and reverse demisubmartingales as a generalization of the notion of reverse (backward) martingales and reverse submartingales.

Definition 2.3.1. *Let $S_1, S_2, ...$ be an L^1 sequence of random variables with $S_0 \equiv 0$. Assume that for all $j = 1, 2, ...$ and all $k \ge j+1$*

$$E\left\{\left(S_j - S_{j+1}\right) f\left(S_{j+1}, S_{j+2}, ..., S_k\right)\right\} \ge 0$$

for all coordinatewise nondecreasing functions f such that the expectation is defined. Then $\{S_j\}_{j \ge 1}$ is called a reverse demimartingale. If in addition the function f is assumed to be nonnegative, the sequence $\{S_j\}_{j \ge 1}$ is called a reverse demisubmartingale.

Definition 2.3.2. *Let $S_1, S_2, ...$ be an L^1 sequence of random variables with $S_0 \equiv 0$. Assume that for all $j = 1, 2, ...$ and all $k \ge j+1$*

$$E\left\{\left(S_j - S_{j+1}\right) f\left(S_{j+1}, S_{j+2}, ..., S_k\right)\right\} \le 0$$

for all coordinatewise nondecreasing functions f such that the expectation is defined. Then $\{S_j\}_{j\geq 1}$ *is called a reverse N-demimartingale. If in addition the function f is assumed to be nonnegative, the sequence* $\{S_j\}_{j\geq 1}$ *is called a reverse N-demisupermartingale.*

Remark 2.3.3. *Let* $S_1, S_2, ...$ *be an* L^1 *sequence of random variables with* $S_0 \equiv 0$ *and let* $\mathcal{A} = \sigma(S_{j+1}, S_{j+2}, ...)$*. A reverse martingale under the* σ*-algebra* $\mathcal{A}$ *is a reverse demimartingale. Furthermore, it can also be verified that a reverse submartingale (with the same choice of* σ*-algebra* $\mathcal{A}$*) is a reverse demisubmartingale.*

We can verify that the sample mean of associated random variables (with zero mean) is a reverse demimartingale under an appropriate assumption.

Proposition 2.3.4. *Let* $\{X_1, X_2, ..., X_n\}$ *be a sequence of random variables with* $E(X_i) = 0$ *for* $i \geq 1$*. Then the sample mean* $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ *is a reverse demimartingale if*

$$\operatorname{Cov}\left[\bar{X}_n, f\left(\bar{X}_{n+1}, ..., \bar{X}_k\right)\right] \geq \operatorname{Cov}\left[X_{n+1}, f\left(\bar{X}_{n+1}, ..., \bar{X}_k\right)\right] \tag{2.3.1}$$

for all coordinatewise nondecreasing functions f such that the expectation is defined.

Proof. We need to show that

$$E\left\{\left(\bar{X}_n - \bar{X}_{n+1}\right) f\left(\bar{X}_{n+1}, \bar{X}_{n+2}, ..., \bar{X}_k\right)\right\} \geq 0, \quad k \geq n+1$$

for all coordinatewise nondecreasing functions *f*.

Note that

$$\begin{aligned} \bar{X}_n - \bar{X}_{n+1} &= \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n+1}\sum_{i=1}^{n+1} X_i \\ &= \frac{(n+1)\sum_{i=1}^{n} X_i - n\sum_{i=1}^{n+1} X_i}{n(n+1)} \\ &= \frac{\sum_{i=1}^{n} X_i - nX_{n+1}}{n(n+1)} = \frac{\bar{X}_n - X_{n+1}}{n+1}. \end{aligned}$$

Under assumption (2.3.1) we can easily have that

$$\begin{aligned} &E\left[\left(\bar{X}_n - \bar{X}_{n+1}\right) f\left(\bar{X}_{n+1}, \bar{X}_{n+2}, ..., \bar{X}_k\right)\right] \\ &= \frac{1}{n+1} E\left[\left(\bar{X}_n - X_{n+1}\right) f\left(\bar{X}_{n+1}, \bar{X}_{n+2}, ..., \bar{X}_k\right)\right] \\ &\geq 0. \quad \blacksquare \end{aligned}$$

Example 2.3.5. *Let X be a random variable with $E|X| < \infty$ and let $X_i = i^{-1}X,\ i \geq 1$. Then $\{X_i,\ i \geq 1\}$ are associated random variables by properties (P3) and (P4) of Esary et al. (1967). We can easily prove that assumption (2.3.1) is satisfied:*

$$\begin{aligned}
\operatorname{Cov}\left[\bar{X}_n, f\left(\bar{X}_{n+1}, ..., \bar{X}_k\right)\right] =& \operatorname{Cov}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{i}X, f\left(\bar{X}_{n+1}, ..., \bar{X}_k\right)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{i}\operatorname{Cov}\left[X, f\left(\bar{X}_{n+1}, ..., \bar{X}_k\right)\right] \\
&= \frac{1}{n}\left(1+\frac{1}{2}+\cdots+\frac{1}{n}\right)\operatorname{Cov}\left[X, f\left(\bar{X}_{n+1}, ..., \bar{X}_k\right)\right] \\
&\geq \frac{1}{n+1}\operatorname{Cov}\left[X, f\left(\bar{X}_{n+1}, ..., \bar{X}_k\right)\right] \\
&= \operatorname{Cov}\left[X_{n+1}, f\left(\bar{X}_{n+1}, ..., \bar{X}_k\right)\right].
\end{aligned}$$

Thus the sample mean of these random variables is a reverse demimartingale.

Remark 2.3.6. *The previous example is a special case of a collection of random variables satisfying assumption (2.3.1). Let X be a random variable with $E(X) = 0$ and let $\{X_n = c_nX,\ n \geq 1\}$, where $\{c_n, n \geq 1\}$ is a sequence of positive numbers satisfying $\frac{1}{n}\sum_{i=1}^{n} c_i \geq c_{n+1}$. Then the sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ is a reverse demimartingale.*

Remark 2.3.7. *A nonincreasing sequence $\{c_n, n \geq 1\}$ clearly satisfies the assumption*

$$\frac{1}{n}\sum_{i=1}^{n} c_i \geq c_{n+1}.$$

2.3.1 Chow and Doob type maximal inequality for reverse demisubmartingales

Chow (1960) proved a maximal inequality for submartingales. Christofides (2000) showed that Chow's inequality is valid for the more general class of demisubmartingales. Theorem 2.3.8 presents a Chow type maximal inequality for reverse demisubmartingales, and will be used later in this chapter to establish a Doob's maximal inequality for reverse demisubmartingales.

Theorem 2.3.8. *Let $S_1, S_2, ...$ be a reverse demisubmartingale and $\{c_k,\ k \geq 1\}$ a nondecreasing sequence of positive numbers. Then for $\varepsilon > 0$,*

$$\varepsilon P\left\{\max_{n\leq k\leq N} c_kS_k \geqslant \varepsilon\right\} \leqslant c_NE(S_N^+) + \sum_{j=n+1}^{N} c_jE(S_{j-1}^+ - S_j^+)$$

where $X^+ = \max\{0, X\}$.

Proof. Let, $A = \{\max_{n \le k \le N} c_k S_k \geqslant \varepsilon\}$. Then A can be written as $A = \bigcup_{j=n}^{N} A_j$, where $A_j = \{c_i S_i < \varepsilon,\ n \le j < i \le N,\ c_j S_j \ge \varepsilon\}$ and $A_i \bigcap A_j = \emptyset$ when $i \neq j$. Then for $\varepsilon > 0$,

$$\begin{aligned}
\varepsilon P(A) = \varepsilon \sum_{j=n}^{N} P(A_j) &= \sum_{j=n}^{N} E(\varepsilon I_{A_j}) \le \sum_{j=n}^{N} E(c_j S_j I_{A_j}) = \sum_{j=n}^{N} E(c_j S_j^+ I_{A_j}) \\
&= c_N E(S_N^+) - c_N E(S_N^+ I_{A_N^c}) + c_{N-1} E(S_{N-1}^+ I_{A_{N-1}}) + \sum_{j=n}^{N-2} c_j E(S_j^+ I_{A_j}) \\
&= c_N E(S_N^+) - c_N E(S_N^+ I_{A_N^c}) + c_{N-1} E(S_{N-1}^+ I_{A_N^c}) \\
&\quad - c_{N-1} E(S_{N-1}^+ I_{A_N^c \cap A_{N-1}^c}) + \sum_{j=n}^{N-2} c_j E(S_j^+ I_{A_j}) \qquad (2.3.2)
\end{aligned}$$

where (2.3.2) follows from the fact that $I_{A_{N-1}} = I_{A_N^c} - I_{A_N^c \cap A_{N-1}^c}$ which in turn holds since $A_{N-1} \subseteq A_N^c$. By the nondecreasingness of the sequence $\{c_k,\ k \ge 1\}$ the quantity on the right-hand side of (2.3.2) is less than or equal to

$$\begin{aligned}
&c_N E(S_N^+) + c_N E\{(S_{N-1}^+ - S_N^+) I_{A_N^c}\} - c_{N-1} E(S_{N-1}^+ I_{A_N^c \cap A_{N-1}^c}) \\
&+ \sum_{j=n}^{N-2} c_j E(S_j^+ I_{A_j}). \qquad (2.3.3)
\end{aligned}$$

Let $h(y) = \lim_{x \to y^-} (x^+ - y^+) / (x - y)$. Then h is a nondecreasing function. By the convexity of the function $x^+ = \max\{0, x\}$ we have

$$S_{N-1}^+ - S_N^+ \geqslant (S_{N-1} - S_N) h(S_N)$$

and we get that

$$E\{(S_{N-1}^+ - S_N^+) I_{A_N}\} \geqslant E\{(S_{N-1} - S_N) h(S_N) I_{A_N}\}. \qquad (2.3.4)$$

Since $h(S_N) I_{A_N}$ is a nonnegative and componentwise nondecreasing function of S_N by the reverse demisubmartingale property the expression on the right-hand side of (2.3.4) is nonnegative. Thus,

$$E\{(S_{N-1}^+ - S_N^+) I_{A_N}\} \geqslant 0$$

and the right-hand side of (2.3.3) is bounded by

$$B = c_N E(S_N^+) + c_N E(S_{N-1}^+ - S_N^+) - c_{N-1} E(S_{N-1}^+ I_{A_N^c \cap A_{N-1}^c}) + \sum_{j=n}^{N-2} c_j E(S_j^+ I_{A_j}).$$

Furthermore,

$$\begin{aligned}
B &= c_N E(S_N^+) + c_N E(S_{N-1}^+ - S_N^+) - c_{N-1} E(S_{N-1}^+ I_{A_N^c \cap A_{N-1}^c}) \\
&\quad + c_{N-2} E(S_{N-2}^+ I_{A_{N-2}}) + \sum_{j=n}^{N-3} c_j E(S_j^+ I_{A_j}) \\
&= c_N E(S_N^+) + c_N E(S_{N-1}^+ - S_N^+) - c_{N-1} E(S_{N-1}^+ I_{A_N^c \cap A_{N-1}^c}) + c_{N-2} E(S_{N-2}^+ I_{A_N^c \cap A_{N-1}^c}) \\
&\quad - c_{N-2} E(S_{N-2}^+ I_{A_N^c \cap A_{N-1}^c \cap A_{N-2}^c}) + \sum_{j=n}^{N-3} c_j E(S_j^+ I_{A_j}) \qquad (2.3.5)
\end{aligned}$$

where (2.3.5) follows from the fact that $I_{A_{N-2}} = I_{A_N^c \cap A_{N-1}^c} - I_{A_N^c \cap A_{N-1}^c \cap A_{N-2}^c}$ since $A_{N-2} \subseteq A_N^c \cap A_{N-1}^c$. Again by the nondecreasingness of the sequence $\{c_k,\ k \geq 1\}$, now the right-hand side of (2.3.5) is less than or equal to

$$\begin{aligned}
&c_N E(S_N^+) + c_N E(S_{N-1}^+ - S_N^+) + c_{N-1} E\{(S_{N-2}^+ - S_{N-1}^+) I_{A_N^c \cap A_{N-1}^c}\} \\
&\quad - c_{N-2} E(S_{N-2}^+ I_{A_N^c \cap A_{N-1}^c \cap A_{N-2}^c}) + \sum_{j=n}^{N-3} c_j E(S_j^+ I_{A_j}) \\
&= c_N E(S_N^+) + \sum_{j=N-1}^{N} c_j E(S_{j-1}^+ - S_j^+) - c_{N-1} E\{(S_{N-2}^+ - S_{N-1}^+) I_{A_N \cup A_{N-1}}\} \\
&\quad - c_{N-2} E(S_{N-2}^+ I_{A_N^c \cap A_{N-1}^c \cap A_{N-2}^c}) + \sum_{j=n}^{N-3} c_j E(S_j^+ I_{A_j}). \qquad (2.3.6)
\end{aligned}$$

Again by the convexity of the function x^+, it follows that

$$E\{(S_{N-2}^+ - S_{N-1}^+) I_{A_N \cup A_{N-1}}\} \geqslant E\{(S_{N-2} - S_{N-1}) h(S_{N-1}) I_{A_N \cup A_{N-1}}\}. \qquad (2.3.7)$$

Since $h(S_{N-1}) I_{A_N \cup A_{N-1}}$ is a nonnegative and componentwise nondecreasing function of $\{S_{N-1}, S_N\}$ by the reverse demisubmartingale property the right-hand side of (2.3.7) is nonnegative and thus the quantity in (2.3.6) is bounded by

$$c_N E(S_N^+) + \sum_{j=N-1}^{N} c_j E(S_{j-1}^+ - S_j^+) - c_{N-2} E(S_{N-2}^+ I_{A_N^c \cap A_{N-1}^c \cap A_{N-2}^c}) + \sum_{j=n}^{N-3} c_j E(S_j^+ I_{A_j}).$$

Working in this way we prove that

$$\begin{aligned}\varepsilon P(A) &\leq c_N E(S_N^+) + \sum_{j=n+1}^{N} c_j E(S_{j-1}^+ - S_j^+) - c_n E(S_n^+ I_{A^c}) \\ &\leq c_N E(S_N^+) + \sum_{j=n+1}^{N} c_j E(S_{j-1}^+ - S_j^+). \quad \blacksquare\end{aligned}$$

The following result shows that the function of a reverse demisubmartingale (or a reverse demimartingale) is a reverse demisubmartingale under a monotonicity assumption.

Lemma 2.3.9. *Let $S_1, S_2, ...$ be a reverse demisubmartingale (or a reverse demimartingale) and g be a nondecreasing convex function. Then $g(S_1), g(S_2), ...$ is a reverse demisubmartingale.*

Proof. We need to show that

$$E\left\{\left(g(S_j) - g(S_{j+1})\right) f\left(g(S_{j+1}), g(S_{j+2}), ..., g(S_k)\right)\right\} \geq 0,$$

for every f nonnegative and coordinatewise nondecreasing function. Since g is a nondecreasing convex function

$$g(S_j) \geq g(S_{j+1}) + (S_j - S_{j+1})\, h(S_{j+1}),$$

where h is the left derivative of g. Obviously h is a nonnegative nondecreasing function. Then for every f nonnegative and coordinatewise nondecreasing function we have that

$$\begin{aligned}&E\left\{\left(g(S_j) - g(S_{j+1})\right) f\left(g(S_{j+1}), g(S_{j+2}), ..., g(S_k)\right)\right\} \\ &\geq E\left\{(S_j - S_{j+1})\, h(S_{j+1}) f\left(g(S_{j+1}), g(S_{j+2}), ..., g(S_k)\right)\right\} \\ &= E\left\{(S_j - S_{j+1})\, f^*\left(S_{j+1}, S_{j+2}, ..., S_k\right)\right\} \\ &\geq 0,\end{aligned}$$

since $\{S_j,\ j \geq 1\}$ is a reverse demimartingale and

$$f^*\left(S_{j+1}, S_{j+2}, ..., S_k\right) = h(S_{j+1}) f\left(g(S_{j+1}), g(S_{j+2}), ..., g(S_k)\right)$$

is a nonnegative componentwise nondecreasing function $f^* : \Re^{k-j} \to \Re$. $\blacksquare$

Lemma 2.3.10. *If $S_1, S_2, ...$ is a reverse demimartingale then $S_1^+, S_2^+, ...$ is a reverse demisubmartingale and $S_1^-, S_2^-, ...$ is a reverse demisubmartingale.*

Proof. Obviously the function $g(x) = \max\{0, x\}$ is nondecreasing and convex. Applying Lemma 2.3.9, we have that $S_1^+, S_2^+, ...$ is a reverse demisubmartingale. Now let $Y_i = -S_i$, $i = 1, 2, ...$. Then by the reverse demimartingale property we note that $Y_1, Y_2, ...$ is also a reverse demimartingale. Furthermore, by Lemma 2.3.9 $Y_1^+, Y_2^+, ...$ is a reverse demisubmartingale. Clearly $Y_i^+ = S_i^-$. Therefore $S_1^-, S_2^-, ...$ is a reverse demisubmartingale. ■

A corollary to the Chow type maximal inequality is the following Doob type inequality.

Corollary 2.3.11. *(Doob's inequality). Let $S_1, S_2, ...$ be a reverse demisubmartingale. Then, for any $\varepsilon > 0$,*

$$P\left\{\max_{n\le k\le N} S_k \geqslant \varepsilon\right\} \leqslant \frac{1}{\varepsilon}\int_{\{\max_{n\le k\le N} S_k \geqslant \varepsilon\}} S_n dP.$$

Proof. In the proof of Theorem 2.3.8, in (2.3.2) we departed from the reverse demisubmartingale $S_1, S_2, ...$ to the reverse demisubmartingale $S_1^+, S_2^+, ...$ because in (2.3.3) we needed to bound $c_{N-1}E(S_{N-1}^+ I_{A_{N-1}^c})$ by $c_N E(S_{N-1}^+ I_{A_{N-1}^c})$. This is correct since $E(S_{N-1}^+ I_{A_{N-1}^c})$ is nonegative and $c_N \geq c_{N-1}$. However, if all the c_i's are equal, such a need does not arise. Therefore, we can stay with the original reverse demisubmartingale. Then in such a case we have

$$\begin{aligned}
\varepsilon P\left\{\max_{n\le k\le N} S_k \geqslant \varepsilon\right\} &\leqslant E(S_N) + \sum_{j=n+1}^{N} E(S_{j-1} - S_j) - E(S_n I_{A^c}) \\
&= E(S_n) - E(S_n I_{A^c}) \\
&= E(S_n I_A) \\
&= \int_{\{\max_{n\le k\le N} S_k \geqslant \varepsilon\}} S_n dP. \quad ■
\end{aligned}$$

Corollary 2.3.12 that follows generalizes the result in Corollary 2.3.11.

Corollary 2.3.12. *(Doob's inequality). Let $S_1, S_2, ...$ be a reverse demisubmartingale and g be a nondecreasing convex function on $\Re$. Then, for any $\varepsilon > 0$,*

$$P\left\{\max_{n\le k\le N} g(S_k) \geqslant \varepsilon\right\} \leqslant \frac{1}{\varepsilon}\int_{\{\max_{n\le k\le N} g(S_k)\geqslant\varepsilon\}} g(S_n)dP.$$

Proof. This result follows from the fact that the sequence $g(S_1), g(S_2), ...$ is a reverse demisubmartingale (Lemma 2.3.9) and by applying Theorem 2.3.8 on the Doob type maximal inequality for reverse demisubmartingales. ■

Remark 2.3.13. *Observe that* $S_i^2 = \left(S_i^+\right)^2 + \left(S_i^-\right)^2$. *In general*

$$|S_i|^\nu = \left(S_i^+\right)^\nu + \left(S_i^-\right)^\nu,\ \nu \ge 1.$$

This observation will be useful to prove Theorem 2.3.14.

Now if we take $g(x) = |x|^\nu,\ \nu \ge 1$ in Corollary 2.3.12 we can obtain the following result.

Theorem 2.3.14. *Let* $S_1, S_2, ...$ *be a reverse demisubmartingale and* $E|S_n|^\nu < \infty$, $n \ge 1$ *for some* $\nu \ge 1$. *Then, for any* $\varepsilon > 0$,

$$P\left\{\max_{n\le k\le N} |S_k| \geqslant \varepsilon\right\} \leqslant \frac{1}{\varepsilon^\nu} E|S_n|^\nu.$$

Proof. Let $\varepsilon > 0$. Using the observation in Remark 2.3.13

$$P\left\{\max_{n\le k\le N} |S_k| \geqslant \varepsilon\right\} \leqslant P\left\{\max_{n\le k\le N} \left(S_k^+\right)^\nu \geqslant \frac{\varepsilon^\nu}{2}\right\} + P\left\{\max_{n\le k\le N} \left(S_k^-\right)^\nu \geqslant \frac{\varepsilon^\nu}{2}\right\}. \quad (2.3.8)$$

Since $\left(S_1^+\right)^\nu, \left(S_2^+\right)^\nu, ...$ and $\left(S_1^-\right)^\nu, \left(S_2^-\right)^\nu, ...$ are reverse demisubmartingales, then from the Doob type maximal inequality for reverse demisubmartingales (Corollary 2.3.11), we have that the right-hand side of (2.3.8) is bounded by

$$\frac{2}{\varepsilon^\nu}\int_{\{\max_{n\le k\le N}(S_k^+)^\nu\geqslant\varepsilon^\nu\}} (S_n^+)^\nu dP + \frac{2}{\varepsilon^\nu}\int_{\{\max_{n\le k\le N}(S_k^-)^\nu\geqslant\varepsilon^\nu\}} (S_n^-)^\nu dP$$

$$\leqslant \frac{1}{\varepsilon^\nu}\int_{\{\max_{n\le k\le N}|S_k|^\nu\geqslant\varepsilon^\nu\}} |S_n|^\nu dP$$

$$\leqslant \frac{1}{\varepsilon^\nu} E|S_n|^\nu.$$ ■

2.3.2 Maximal inequalities for functions of Reverse Demisubmartingales

Wang and Hu (2009) generalized the results of Christofides (2000) for demimartingales and demisubmartingales. Here, we present a similar theorem which generalizes Theorem 2.3.8 for reverse demisubmartingales.

Theorem 2.3.15. *Let $S_1, S_2, ...$ be a reverse demimartingale, g be a nonnegative convex function on $\Re$ with $g(0) = 0$ and $g(S_i) \in L^1,\ i \geq 1$. Let $\{c_k,\ k \geq 1\}$ a nondecreasing sequence of positive numbers. Then for $\varepsilon > 0$,*

$$\varepsilon P\left\{\max_{n\leq k\leq N} c_k g(S_k) \geqslant \varepsilon\right\} \leqslant c_N E(g(S_N)) + \sum_{j=n+1}^{N} c_j E(g(S_{j-1}) - g(S_j)).$$

Proof. Define the functions

$$u(x) = g(x)I_{[x\geq 0]} \ \ \textit{and} \ \ v(x) = g(x)I_{[x<0]}.$$

Note that

$$g(x) = u(x) + v(x) = \max\{u(x), v(x)\}, \quad x \in \Re, \tag{2.3.9}$$

where $u(x)$ is a nonnegative nondecreasing convex function and $v(x)$ is a nonnegative nonincreasing convex function.

Then we have that

$$\begin{aligned}
&\varepsilon P\left\{\max_{n\leq k\leq N} c_k g(S_k) \geqslant \varepsilon\right\} \\
&= \varepsilon P\left\{\max_{n\leq k\leq N} c_k \max(u(S_k), v(S_k)) \geqslant \varepsilon\right\} \\
&= \varepsilon P\left\{\max(c_n \max(u(S_n), v(S_n)), ..., c_N \max(u(S_N), v(S_N))) \geqslant \varepsilon\right\} \\
&\leqslant \varepsilon P\left\{\max(c_n u(S_n), ..., c_N u(S_N)) \geqslant \varepsilon\right\} + \varepsilon P\left\{\max(c_n v(S_n), ..., c_N v(S_N)) \geqslant \varepsilon\right\} \\
&= \varepsilon P\left\{\max_{n\leq k\leq N} c_k u(S_k) \geqslant \varepsilon\right\} + \varepsilon P\left\{\max_{n\leq k\leq N} c_k v(S_k) \geqslant \varepsilon\right\}. \qquad (2.3.10)
\end{aligned}$$

Furthermore, combining Theorem 2.3.8 and Lemma 2.3.9 we have that

$$\varepsilon P\left\{\max_{n\leq k\leq N} c_k u(S_k) \geqslant \varepsilon\right\} \leqslant c_N E(u(S_N)) + \sum_{j=n+1}^{N} c_j E(u(S_{j-1}) - u(S_j)), \tag{2.3.11}$$

since $u(x)$ is a nondecreasing convex function.

Let $A = \{\max_{n\leq k\leq N} c_k v(S_k) \geqslant \varepsilon\}$. Then A can be written as $A = \bigcup_{j=n}^{N} A_j$, where $A_j = \{c_i v(S_i) < \varepsilon,\ j < i \leq N,\ c_j v(S_j) \geq \varepsilon\}$, $n \leq j \leq N$, and the A_j's are disjoint. Therefore,

$$\begin{aligned}
\varepsilon P(A) &= \varepsilon \sum_{j=n}^{N} P(A_j) = \sum_{j=n}^{N} E(\varepsilon I_{A_j}) \leq \sum_{j=n}^{N} E(c_j v(S_j) I_{A_j}) = \sum_{j=n}^{N} E(c_j v(S_j) I_{A_j}) \\
&= c_N E(v(S_N)) - c_N E(v(S_N) I_{A_N^c}) + c_{N-1} E(v(S_{N-1}) I_{A_{N-1}}) + \sum_{j=n}^{N-2} c_j E(v(S_j) I_{A_j}) \\
&= c_N E(v(S_N)) - c_N E(v(S_N) I_{A_N^c}) + c_{N-1} E(v(S_{N-1}) I_{A_N^c}) \\
&\quad - c_{N-1} E(v(S_{N-1}) I_{A_N^c \cap A_{N-1}^c}) + \sum_{j=n}^{N-2} c_j E(v(S_j) I_{A_j})
\end{aligned} \tag{2.3.12}$$

where (2.3.12) follows from the fact that $I_{A_{N-1}} = I_{A_N^c} - I_{A_N^c \cap A_{N-1}^c}$ which in turn holds since $A_{N-1} \subseteq A_N^c$. By the nondecreasingness of the sequence $\{c_k,\ k \geq 1\}$ the right-hand side of (2.3.12) is less than or equal to

$$\begin{aligned}
&c_N E(v(S_N)) + c_N E\{(v(S_{N-1}) - v(S_N)) I_{A_N^c}\} - c_{N-1} E(v(S_{N-1}) I_{A_N^c \cap A_{N-1}^c}) \\
&+ \sum_{j=n}^{N-2} c_j E(v(S_j) I_{A_j}).
\end{aligned} \tag{2.3.13}$$

Let $h(y) = \lim_{x\to y^-} (v(x) - v(y)) / (x - y)$. Then h is a nondecreasing non-positive function. By the convexity of $v(x)$, we have

$$v(S_{N-1}) - v(S_N) \geqslant (S_{N-1} - S_N) h(S_N)$$

and, therefore,

$$E\{(v(S_{N-1}) - v(S_N)) I_{A_N}\} \geqslant E\{(S_{N-1} - S_N) h(S_N) I_{A_N}\}. \tag{2.3.14}$$

Since I_{A_N} is a nonincreasing function of S_N and $h(.)$ is a non-positive nondecreasing function, it follows that $h(S_N)I_{A_N}$ is a nondecreasing function of S_N. By the reverse demimartingale property the right-hand side of (2.3.14) is nonnegative. Thus,

$$E\{(v(S_{N-1}) - v(S_N))I_{A_N}\} \geqslant 0$$

and the right-hand side of (2.3.13) is bounded by

$$\begin{aligned}B =& c_N E(v(S_N)) + c_N E(v(S_{N-1}) - v(S_N)) - c_{N-1} E(v(S_{N-1}) I_{A_N^c \cap A_{N-1}^c}) \\ &+ \sum_{j=n}^{N-2} c_j E(v(S_j) I_{A_j}).\end{aligned}$$

Furthermore,

$$\begin{aligned}B &= c_N E(v(S_N)) + c_N E(v(S_{N-1}) - v(S_N)) - c_{N-1} E(v(S_{N-1}) I_{A_N^c \cap A_{N-1}^c}) \\ &\quad + c_{N-2} E(v(S_{N-2}) I_{A_{N-2}}) + \sum_{j=n}^{N-3} c_j E(v(S_j) I_{A_j}) \\ &= c_N E(v(S_N)) + c_N E(v(S_{N-1}) - v(S_N)) - c_{N-1} E(v(S_{N-1}) I_{A_N^c \cap A_{N-1}^c}) \\ &\quad + c_{N-2} E(v(S_{N-2}) I_{A_N^c \cap A_{N-1}^c}) - c_{N-2} E(v(S_{N-2}) I_{A_N^c \cap A_{N-1}^c \cap A_{N-2}^c}) \\ &\quad + \sum_{j=n}^{N-3} c_j E(v(S_j) I_{A_j}), \qquad (2.3.15)\end{aligned}$$

where (2.3.15) follows from the fact that $I_{A_{N-2}} = I_{A_N^c \cap A_{N-1}^c} - I_{A_N^c \cap A_{N-1}^c \cap A_{N-2}^c}$ since $A_{N-2} \subseteq A_N^c \cap A_{N-1}^c$. Again by the nondecreasingness of the sequence $\{c_k,\ k \geq 1\}$, the right-hand side of (2.3.15) is less than or equal to

$$\begin{aligned}& c_N E(v(S_N)) + c_N E(v(S_{N-1}) - v(S_N)) + c_{N-1} E\{(v(S_{N-2}) - v(S_{N-1})) I_{A_N^c \cap A_{N-1}^c}\} \\ &\quad - c_{N-2} E(v(S_{N-2}) I_{A_N^c \cap A_{N-1}^c \cap A_{N-2}^c}) + \sum_{j=n}^{N-3} c_j E(v(S_j) I_{A_j}) \\ &= c_N E(v(S_N)) + \sum_{j=N-1}^{N} c_j E(v(S_{j-1}) - v(S_j)) \\ &\quad - c_{N-1} E\{(v(S_{N-2}) - v(S_{N-1})) I_{A_N \cup A_{N-1}}\} - c_{N-2} E(v(S_{N-2}) I_{A_N^c \cap A_{N-1}^c \cap A_{N-2}^c}) \\ &\quad + \sum_{j=n}^{N-3} c_j E(v(S_j) I_{A_j}). \qquad (2.3.16)\end{aligned}$$

Again by the convexity of $v(x)$ we have that

$$E\{(v(S_{N-2}) - v(S_{N-1}))I_{A_N \cup A_{N-1}}\} \geqslant E\{(S_{N-2} - S_{N-1})h(S_{N-1})I_{A_N \cup A_{N-1}}\}. \quad (2.3.17)$$

Since $I_{A_N \cup A_{N-1}}$ is a nonnegative and componentwise nonincreasing function of $\{S_{N-1}, S_N\}$ and $h(.)$ is a non-positive nondecreasing function, it follows that $h(S_{N-1})I_{A_N \cup A_{N-1}}$ is a nondecreasing function of $\{S_{N-1}, S_N\}$. By the reverse demimartingale property the right-hand side of (2.3.17) is nonnegative and thus the quantity in (2.3.16) is bounded by

$$c_N E(v(S_N)) + \sum_{j=N-1}^{N} c_j E(v(S_{j-1}) - v(S_j)) - c_{N-2} E(v(S_{N-2}) I_{A_N^c \cap A_{N-1}^c \cap A_{N-2}^c})$$
$$+ \sum_{j=n}^{N-3} c_j E(v(S_j) I_{A_j}).$$

Working in this manner we prove that

$$\begin{aligned} \varepsilon P(A) &\leq c_N E(v(S_N)) + \sum_{j=n+1}^{N} c_j E(v(S_{j-1}) - v(S_j)) - c_n E(v(S_n) I_{A^c}) \\ &\leq c_N E(v(S_N)) + \sum_{j=n+1}^{N} c_j E(v(S_{j-1}) - v(S_j)). \end{aligned} \quad (2.3.18)$$

Finally, by (2.3.10), (2.3.11) and (2.3.18), we have that

$$\begin{aligned} &\varepsilon P\left\{\max_{n \leq k \leq N} c_k g(S_k) \geqslant \varepsilon\right\} \\ &\leqslant \varepsilon P\left\{\max_{n \leq k \leq N} c_k u(S_k) \geqslant \varepsilon\right\} + \varepsilon P\left\{\max_{n \leq k \leq N} c_k v(S_k) \geqslant \varepsilon\right\} \\ &\leqslant c_N E(u(S_N)) + \sum_{j=n+1}^{N} c_j E(u(S_{j-1}) - u(S_j)) \\ &\quad + c_N E(v(S_N)) + \sum_{j=n+1}^{N} c_j E(v(S_{j-1}) - v(S_j)) \\ &= c_N E(g(S_N)) + \sum_{j=n+1}^{N} c_j E(g(S_{j-1}) - g(S_j)). \quad \blacksquare \end{aligned} \quad (2.3.19)$$

Observe that for $g(x) = |x|$ in Theorem 2.3.15 we obtain the following result.

Remark 2.3.16. *Since $g(x) = |x|$ is a nonnegative convex function with $g(0) = 0$, applying Theorem 2.3.15 for every $\varepsilon > 0$, we have that*

$$\varepsilon P\left\{\max_{n\le k\le N} c_k|S_k| \geqslant \varepsilon\right\} \leqslant c_N E|S_N| + \sum_{j=n+1}^{N} c_j E(|S_{j-1}| - |S_j|).$$

A corollary to the Chow type maximal inequality is the following Doob type inequality.

Corollary 2.3.17. *(Doob's inequality). Let $S_1, S_2, \ldots$ be a reverse demimartingale, g be a nonnegative convex function on $\Re$ with $g(0) = 0$ and $g(S_i) \in L^1,\ i \ge 1$. Then for $\varepsilon > 0$,*

$$P\left\{\max_{n\le k\le N} g(S_k) \geqslant \varepsilon\right\} \leqslant \frac{1}{\varepsilon}\int_{\{\max_{n\le k\le N} g(S_k)\geqslant\varepsilon\}} g(S_n)dP.$$

Proof. Let $u(x)$ and $v(x)$ be as defined in Theorem 2.3.15. Since $u(x)$ is a nondecreasing convex function, by Corollary 2.3.12 we have

$$\begin{aligned}\varepsilon P\left\{\max_{n\le k\le N} u(S_k) \geqslant \varepsilon\right\} &\leqslant \int_{\{\max_{n\le k\le N} u(S_k)\geqslant\varepsilon\}} u(S_n)dP \\ &\leqslant \int_{\{\max_{n\le k\le N} g(S_k)\geqslant\varepsilon\}} u(S_n)dP. \qquad (2.3.20)\end{aligned}$$

Furthermore, from Theorem 2.3.15 with $c_k \equiv 1,\ k \ge 1$

$$\begin{aligned}\varepsilon P\left\{\max_{n\le k\le N} v(S_k) \geqslant \varepsilon\right\} &\leqslant E(v(S_N)) + \sum_{j=n+1}^{N} E(v(S_{j-1}) - v(S_j)) - E(v(S_n)I_{A^c}) \\ &= E(v(S_n)) - E(v(S_n)I_{A^c}) \\ &= E(v(S_n)I_A) \\ &= \int_{\{\max_{n\le k\le N} v(S_k)\geqslant\varepsilon\}} v(S_n)dP \\ &\leqslant \int_{\{\max_{n\le k\le N} g(S_k)\geqslant\varepsilon\}} v(S_n)dP. \qquad (2.3.21)\end{aligned}$$

Finally, by (2.3.10), (2.3.20) and (2.3.21), we have that

$$\begin{aligned}
&\varepsilon P\left\{\max_{n\leq k\leq N} g(S_k)\geqslant\varepsilon\right\}\\
&\leqslant \varepsilon P\left\{\max_{n\leq k\leq N} u(S_k)\geqslant\varepsilon\right\}+\varepsilon P\left\{\max_{n\leq k\leq N} v(S_k)\geqslant\varepsilon\right\}\\
&\leqslant \int_{\{\max_{n\leq k\leq N} g(S_k)\geqslant\varepsilon\}} u(S_n)dP+\int_{\{\max_{n\leq k\leq N} g(S_k)\geqslant\varepsilon\}} v(S_n)dP\\
&\leqslant \int_{\{\max_{n\leq k\leq N} g(S_k)\geqslant\varepsilon\}} g(S_n)dP. \ \blacksquare \qquad (2.3.22)
\end{aligned}$$

Again, taking $g(x)=|x|^\nu,\ \nu\geq 1$ in Corollary 2.3.17 we get the following result.

Corollary 2.3.18. *Let $S_1, S_2, ...$ be a reverse demimartingale, $\nu\geq 1$. Then for any $\varepsilon>0$,*

$$P\left\{\max_{n\leq k\leq N}|S_k|\geqslant\varepsilon\right\}\leqslant\frac{1}{\varepsilon^\nu}\int_{\{\max_{n\leq k\leq N}|S_k|^\nu\geqslant\varepsilon^\nu\}}|S_n|^\nu dP\leqslant\frac{1}{\varepsilon^\nu}E|S_n|^\nu.$$

Proof. By taking $g(x)=|x|^\nu$, $\nu\geq 1$ in Corollary 2.3.17, then for any $\varepsilon>0$,

$$\begin{aligned}
P\left\{\max_{n\leq k\leq N}|S_k|\geqslant\varepsilon\right\} &= P\left\{\max_{n\leq k\leq N}|S_k|^\nu\geqslant\varepsilon^\nu\right\}\\
&\leqslant\frac{1}{\varepsilon^\nu}\int_{\{\max_{n\leq k\leq N}|S_k|^\nu\geqslant\varepsilon^\nu\}}|S_n|^\nu dP\\
&\leqslant\frac{1}{\varepsilon^\nu}E|S_n|^\nu. \ \blacksquare
\end{aligned}$$

2.3.3 Doob type maximal inequality for reverse Demimartingales

Some results that Wang and Hu (2009) established for demimartingales (see their Theorem 3.1 and Corollary 3.1), with appropriate modification are still valid for reverse demimartingales.

Lemma 2.3.19. *Let $S_1, S_2, ...$ be a reverse demimartingale and g be a nonnegative function with $g(0)=0$. Assume that $E(g(S_k))^p<\infty$ for $p>1$. Then*

$$E\left(\max_{n\le k\le N} g(S_k)\right)^p \leqslant \left(\frac{p}{1-p}\right)^p E(g(S_n))^p.$$

Proof. Applying Holder's inequality and Corollary 2.3.18, we can obtain that

$$E\left(\max_{n\le k\le N} g(S_k)\right)^p = p\int_0^\infty x^{p-1} P\left\{\max_{n\le k\le N} g(S_k) \ge x\right\} dx$$

$$\leqslant p\int_0^\infty x^{p-2} E\left(g(S_n) I\left\{\max_{n\le k\le N} g(S_k) \ge x\right\}\right) dx$$

$$= pE\left(g(S_n)\int_0^{\max_{n\le k\le N} g(S_k)} x^{p-2} dx\right)$$

$$= \frac{p}{p-1} E\left(g(S_n)\left(\max_{n\le k\le N} g(S_k)\right)^{p-1}\right)$$

$$\leqslant \frac{p}{p-1}\left[E\left(g(S_n)\right)^p\right]^{1/p}\left[E\left(\max_{n\le k\le N} g(S_k)\right)^p\right]^{1/q},$$

where $1/p + 1/q = 1$. Since $E(g(S_k))^p < \infty$ for $p > 1$ we have that

$$E\left[\left(\max_{n\le k\le N} g(S_k)\right)^p\right]^{1/p} \le \frac{p}{p-1}\left[E\left(g(S_n)\right)^p\right]^{1/p},$$

and

$$E\left(\max_{n\le k\le N} g(S_k)\right)^p \leqslant \left(\frac{p}{1-p}\right)^p E(g(S_n))^p. \quad ■$$

Applying now $g(x) = |x|$ in Lemma 2.3.19 we get the following corollary.

Corollary 2.3.20. *Let $S_1, S_2, \ldots$ be a reverse demimartingale. Assume that $E|S_k|^p < \infty$, $k \ge 1$, for $p > 1$. Then*

$$E\left(\max_{n\le k\le N} |S_k|\right)^p \leqslant \left(\frac{p}{1-p}\right)^p E(|S_n|)^p.$$

Proof. If we take $g(x) = |x|$ in Lemma 2.3.19 we immediately have the result. ■

2.3.4 U-statistics and reverse Demimartingales

Proposition 2.3.21. *Let U_n be a U-statistic based on a collection of random variables and on the kernel h with $E(h) = 0$. Assume that*

$$m \sum_{1\leq i_1<\cdots<i_m\leq n+1} \mathrm{Cov}[h(X_{i_1},...,X_{i_m}), f(U_{n+1},...,U_k)]$$
$$\geqslant (n+1) \sum_{1\leq i_1<\cdots<i_{m-1}\leq n} \mathrm{Cov}[h(X_{i_1},...,X_{i_{m-1}}, X_{n+1}), f(U_{n+1},...,U_k)], \tag{2.3.23}$$

for all coordinatewise nondecreasing functions f such that the expectation is defined. Then $\{U_n, n \geq m\}$ is a reverse demimartingale.

Proof. We need to show that

$$E\left\{(U_n - U_{n+1}) f\left(U_{n+1}, U_{n+2}, ..., U_k\right)\right\} \geq 0, \quad k \geq n+1$$

for all coordinatewise nondecreasing functions *f*.

Let

$$S_n = \sum_{1\leq i_1<i_2<\cdots<i_m\leq n} h\left(X_{i_1}, X_{i_2}, ..., X_{i_m}\right).$$

Then note that

$$U_n - U_{n+1}$$
$$= \frac{S_n}{\binom{n}{m}} - \frac{S_{n+1}}{\binom{n+1}{m}} = \frac{\binom{n+1}{m}S_n - \binom{n}{m}S_{n+1}}{\binom{n+1}{m}\binom{n}{m}} = \frac{\frac{n+1}{n+1-m}\binom{n}{m}S_n - \binom{n}{m}S_{n+1}}{\binom{n+1}{m}\binom{n}{m}}$$

$$= \frac{(n+1)S_n - (n+1-m)S_{n+1}}{\binom{n+1}{m}(n+1-m)} = \frac{(n+1)[S_n - S_{n+1}] + mS_{n+1}}{\binom{n+1}{m}(n+1-m)}$$

$$= \frac{mS_{n+1} - (n+1)\sum_{1\leq i_1<\cdots<i_{m-1}\leq n} h(X_{i_1},...,X_{i_{m-1}}, X_{n+1})}{\binom{n+1}{m}(n+1-m)}.$$

Under (2.3.23) we can easily have that

$$E\left\{(U_n - U_{n+1}) f\left(U_{n+1}, U_{n+2}, ..., U_k\right)\right\} \tag{2.3.24}$$

$$= E\left\{\left(\frac{mS_{n+1}-(n+1)\sum_{1\le i_1<\cdots<i_{m-1}\le n} h(X_{i_1},...,X_{i_{m-1}},X_{n+1})}{\binom{n+1}{m}(n+1-m)}\right) f\left(U_{n+1},U_{n+2},...,U_k\right)\right\}$$

$\geq 0.$ ■

Finally, we give an example of a reverse demimartingale.

Example 2.3.22. *Let X be a random variable with $E|X| < \infty$ and assume that $\{X_i = i^{-1}X,\ i \geq 1\}$. Let U_n be the U-statistic based on $\{X_n\}_{n\geq 1}$ with kernel $h(x,y) = xy$. Then the previous assumption is true, since,*

$$m \sum_{1\le i_1<\cdots<i_m\le n+1} \mathrm{Cov}[h(X_{i_1},...,X_{i_m}), f(U_{n+1},...,U_k)]$$

$$= 2 \sum_{1\le i<j\le n+1} \mathrm{Cov}[\frac{1}{ij}X^2, f(U_{n+1},...,U_k)]$$

$$= 2\mathrm{Cov}[X^2, f(U_{n+1},...,U_k)] \sum_{1\le i<j\le n+1} \frac{1}{ij}$$

$$\geqslant \mathrm{Cov}[X^2, f(U_{n+1},...,U_k)] \sum_{1\le i\le n+1} \frac{1}{i}$$

$$= (n+1) \sum_{1\le i\le n+1} \mathrm{Cov}[\frac{1}{i(n+1)}X^2, f(U_{n+1},...,U_k)]$$

$$= (n+1) \sum_{1\le i_1<i_2\le n} \mathrm{Cov}[h(X_{i_1},X_{i_2}), f(U_{n+1},...,U_k)].$$

Chapter 3

Strong convergence for U-statistics based on associated random variables

In Chapter 3 we study strong convergence for U-statistics based on associated random variables. The chapter is organized as follows. In Section 3.1 we present a strong law for one sample U-statistics based on associated random variables which can be found in Christofides (2000). In Section 3.2 we obtain a strong law for U-statistics based on associated multidimensionally indexed random variables. Furthermore in Section 3.3, we prove a strong law for multi-sample U-statistics based on collections of associated random variables.

3.1 Strong law for one sample U-statistics based on associated random variables

The following result gives a strong law of large numbers for demimartingales and can be found in Christofides (2000). Christofides (2000) gives a strong law of large numbers for U-Statistics based on associated random variables in the case where the kernel of the U-statistic belongs to a large family of functions, called kernels of bounded variation.

Lemma 3.1.1. *Let $S_0, S_1, S_2, ...$ be a demimartingale, with $S_0 = 0$. Let $\{c_k, k \geq 1\}$ be a nonincreasing sequence of positive numbers with $\lim_{k\to\infty} c_k = 0$. Assume that*

$E|S_k|^\nu <$ *for* $\nu \geq 1$, *for all* k. *If*

$$\sum_{k=1}^{\infty} c_k^{\nu} E\left(|S_k|^{\nu} - |S_{k-1}|^{\nu}\right) < \infty$$

then

$$c_n S_n \xrightarrow{a.s.} 0, \quad n \to \infty.$$

Combining Lemma 3.1.1 and Proposition 1.6.1 we get the following strong law of large numbers.

Theorem 3.1.2. *(Christofides (2002)). Let* U_n *be a U-statistic based on a collection of associated random variables and on the kernel* h. *Assume that* $E|U_k|^{\nu} < \infty$ *for* $\nu \geq 1$ *and all* $k \geq m$. *Furthermore assume that* h *is componentwise nondecreasing. If*

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

then

$$U_n - E(U_n) \xrightarrow{a.s.} 0, \quad n \to \infty.$$

A U-statistic based on associated random variables is a demimartingale if the kernel h is componentwise nondecreasing. But not all kernels of interest fullfill this requirement. For example the kernel $h(x;y) =| x-y |$ or $h(x;y) = 1_{\{x>y\}}$. In that case, the U-statistic defined is not necessarily a demimartingale. However, if the kernel h belongs to a large family of functions which includes all nondecreasing functions then the U-statistic can be expressed as the difference of two U-statistics, each of which has a componentwise nondecreasing kernel. Christofides (2004) introduced this type of kernels termed as kernels of bounded variation and also proves a strong law of large numbers for this family of kernels.

Definition 3.1.3. *(Christofides (2004)). Assume that* $[a,b],[c,d] \subseteq \Re$ *and* f *is a real-valued function defined on the rectangle* $[a,b]\times[c,d]$. *For simplicity by* $\Delta f((q,r),(s,t))$ *we denote the quantity* $f(r,t) - f(q,t) - f(r,s) + f(q,s)$ *with* $a \leq q < r \leq b$ *and* $c \leq s < t \leq d$. *Let*

$$a < x_0 < x_1 < \cdots < x_k = b$$

and

$$c < y_0 < y_1 < \cdots < y_l = d$$

be any subdivisions of the intervals $[a,b]$ *and* $[c,d]$*, respectively. Let*

$$C \equiv \sum_{i=1}^{k} \sum_{j=1}^{l} \left| \Delta f((x_{i-1}, x_i), (y_{j-1}, y_j)) \right|.$$

The function $f : [a,b] \times [c,d] \to \Re$ *is called a function of bounded variation on* $[a,b] \times [c,d]$ *if*

$$C^f_{[a,b]\times[c,d]} = \sup C < \infty.$$

Theorem 3.1.2 is extended to the case of kernels of bounded variation.

Theorem 3.1.4. *(Christofides (2004)). Let* U_n *be a U-statistic based on a collection of associated random variables and on the kernel h. Assume that* $E|U_k|^\nu < \infty$ *for* $\nu \geq 1$ *and all* $k \geq m$*. Furthermore assume that h is a function of bounded variation. If*

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

then

$$U_n - E(U_n) \xrightarrow{a.s.} 0, \quad n \to \infty.$$

3.2 U-statistics based on associated multidimensionally indexed random variables

For a positive integer d let N^d denote the d-dimensional positive integer lattice. Furthermore, for $\mathbf{n} = (n_1, ..., n_d)$, we put $|\mathbf{n}| = \prod_{i=1}^{d} n_i$, and by $\mathbf{n} \to \infty$ we mean that $|\mathbf{n}| \to \infty$ (equivalently, $\max\{n_1, ..., n_d\} \to \infty$). For $\mathbf{n},\ \mathbf{m} \in N^d$ with $\mathbf{n} = (n_1, ..., n_d)$ and $\mathbf{m} = (m_1, ..., m_d)$ the notation $\mathbf{n} \leq \mathbf{m}$ means that $n_i \leq m_i\ \forall\ i = 1, ..., d$ while the notation $\mathbf{n} < \mathbf{m}$ means that $n_i \leq m_i\ \forall\ i = 1, ..., d$ with at least one inequality strict.

Definition 3.2.1. *A collection of multidimensionally indexed random variables* $\{X_{\boldsymbol{k}},\ \boldsymbol{k} \leq \boldsymbol{n},\ \boldsymbol{n} \in \mathbb{N}^d\}$ *is said to be associated if for any two coordinatewise nondecreasing functions f and g*

$$cov(f(X_{\boldsymbol{k}}, \boldsymbol{k} \leq \boldsymbol{n}), g(X_{\boldsymbol{k}}, \boldsymbol{k} \leq \boldsymbol{n})) \geq 0,$$

provided that the covariance is defined. An infinite collection is associated if every finite subcollection is associated.

The above definition is just the classical definition of association stated for the case of multidimensionally indexed random variables. The index of the variables in no way affects the qualitative property of association, i.e., that nondecreasing functions of all (or some) of the variables are nonnegatively correlated.

Definition 3.2.2. *Let* $\left\{S_{\boldsymbol{n}}, \boldsymbol{n} \in \mathbb{N}^d\right\}$ *be a collection of multidimensionally indexed random variables. The collection is called a multidimensionally indexed demimartingale if*

$$E\left[\left(S_{\boldsymbol{t}}-S_{\boldsymbol{r}}\right) f\left(S_{\boldsymbol{k}}, \boldsymbol{k} \leq \boldsymbol{r}\right)\right] \geq 0$$

for all $\boldsymbol{t}, \boldsymbol{r} \in \mathbb{N}^d$ *with* $\boldsymbol{r} \leq \boldsymbol{t}$ *and for all componentwise nondecreasing functions f.*

It is easy to verify that the partial sum of mean zero associated multidimensionally indexed random variables is a multidimensionally indexed demimartingale.

A U-statistic on multidimensionally indexed random variables can be defined as follows:

Definition 3.2.3. *Let* $\{X_{\boldsymbol{k}},\ \boldsymbol{k} \leq \boldsymbol{n},\ \boldsymbol{n} \in N^d\}$ *be a collection of identically distributed associated multidimensionally indexed random variables. Let h be a symmetric mapping from* $\mathbb{R}^m$ *to* $\mathbb{R}$ *with* $m \leq |\boldsymbol{n}|$. *We define the corresponding U-statistic*

$$U_{\boldsymbol{n}} = \binom{|\ \boldsymbol{n}\ |}{m}^{-1} \sum_{c} h\left(X_{\boldsymbol{i}_1}, X_{\boldsymbol{i}_2}, ..., X_{\boldsymbol{i}_m}\right)$$

where $\sum_c$ *denotes the summation over the* $\binom{|\boldsymbol{n}|}{m}$ *combinations of the m distinct elements* $\{\boldsymbol{i}_1, \boldsymbol{i}_2, ..., \boldsymbol{i}_m\}$ *from* $\{(1, ..., 1), ..., (n_1, ..., n_d)\}$.

A U-statistic as given by Definition 3.2.3 can be shown to be expressed in terms of a multidimensionally indexed demimartingale under the usual assumptions on the kernel h.

Proposition 3.2.4. *Assume for simplicity that d=2. Let* $U_{\boldsymbol{n}}$ *be a U-statistic with* $m = 2$ *based on a sample of associated random variables and on the kernel h. Let* $S_{\boldsymbol{n}} = \binom{|\boldsymbol{n}|}{2} U_{\boldsymbol{n}}$ *where* $\boldsymbol{n} = (n_1, n_2),\ n_1 \geq 1,\ n_2 \geq 1$ *and* $|\boldsymbol{n}| \geq 2$. *Assume that h is componentwise nondecreasing and that* $E(h) = 0$. *Then the collection* $\{S_{\boldsymbol{n}},\ \boldsymbol{n} > \boldsymbol{1}\}$ *is a multidimensionally indexed demimartingale.*

Proof. Let

$$S_{\mathbf{t}} = \sum_{c_{\mathbf{t}}} h\left(X_{\mathbf{i}}, X_{\mathbf{j}}\right),\ \ S_{\mathbf{r}} = \sum_{c_{\mathbf{r}}} h\left(X_{\mathbf{i}}, X_{\mathbf{j}}\right)$$

where $\mathbf{t} = (t_1, t_2)$ and $\mathbf{r} = (r_1, r_2)$ with $\mathbf{r} \leq \mathbf{t}$. Also $c_{\mathbf{r}}$ and $c_{\mathbf{t}}$ denote the summation over the $\binom{|\mathbf{n}|}{2}$ combinations of the m=*2* distinct elements $\{\mathbf{i}, \mathbf{j}\}$ from $\{(1,1), ..., (r_1, r_2)\}$ and $\{(1,1), ..., (t_1, t_2)\}$ respectively.

We can write

$$S_{\mathbf{t}} - S_{\mathbf{r}} = \sum_{c_{\mathbf{t}\text{-}\mathbf{r}}} h\left(X_{\mathbf{i}}, X_{\mathbf{j}}\right).$$

Then for any componentwise nondecreasing function f we have

$$\begin{aligned} E\left[\left(S_{\mathbf{t}} - S_{\mathbf{r}}\right) f\left(S_{\mathbf{k}}, \mathbf{k} \leq \mathbf{r}\right)\right] &= \sum_{c_{\mathbf{t}\text{-}\mathbf{r}}} E[h\left(X_{\mathbf{i}}, X_{\mathbf{j}}\right) f\left(S_{\mathbf{k}}, \mathbf{k} \leq \mathbf{r}\right)] \\ &= \sum_{c_{\mathbf{t}\text{-}\mathbf{r}}} E[h\left(X_{\mathbf{i}}, X_{\mathbf{j}}\right) g\left(X_{\mathbf{k}}, \mathbf{k} \leq \mathbf{r}\right)] \\ &\geq 0 \end{aligned}$$

where the function g defined as:

$$\begin{aligned} g\left(X_{\mathbf{k}}, \mathbf{k} \leq \mathbf{r}\right) = f(&\sum_{(1,1)<\mathbf{j}\leq\mathbf{r}} h(x_{1,1}, x_{\mathbf{j}}), \sum_{(1,2)<\mathbf{j}\leq\mathbf{r}} h(x_{1,2}, x_{\mathbf{j}}), \sum_{(2,1)<\mathbf{j}\leq\mathbf{r}} h(x_{2,1}, x_{\mathbf{j}}), ..., \\ &\sum_{r_1-1,r_2<\mathbf{j}\leq\mathbf{r}} h(x_{r_1-1,r_2}, x_{\mathbf{j}}), \sum_{r_1,r_2-1<\mathbf{j}\leq\mathbf{r}} h(x_{r_1,r_2-1}, x_{\mathbf{j}})), \end{aligned}$$

is componentwise nondecreacing since h, f are componentwise nondecreacing, (see property 4 in Esary et al. (1967)). ■

3.2.1 A strong law in the case of nondecreasing kernels

The following result gives a strong law of large numbers for multidimensionally indexed demimartingales and can be found in Christofides and Hadjikyriakou (2011).

Lemma 3.2.5. *Assume that* $\{Y_{\boldsymbol{k}},\ \boldsymbol{k} \in \mathbb{N}^d\}$, $\{c_{\boldsymbol{k}},\ \boldsymbol{k} \in \mathbb{N}^d\}$ *and the function* g *is a nonnegative convex function on* $\Re$ *with* $g(0) = 0$. *Further assume that there exists a number* $p \geq 1$ *such that* $E[g(Y_{\boldsymbol{k}})]^p < \infty$ *for all* $\boldsymbol{k}$ *and for some* $1 \leq s \leq d$, $\sum_{\boldsymbol{k}} c_{\boldsymbol{k}}^p E([g(Y_{\boldsymbol{k}})]^p - [g(Y_{\boldsymbol{k};s;k_s-1})]^p) < \infty$ *and* $\sum_{k_i, i\neq s} c_{\boldsymbol{k};s;N}^p E([g(Y_{\boldsymbol{k};s;N})]^p) < \infty$ *for each* $N \in \mathbb{N}$. *Then*

$$c_{\boldsymbol{k}} g(Y_{\boldsymbol{k}}) \xrightarrow{a.s.} 0, \quad as\ \boldsymbol{k}\to\infty,$$

where $Y_{\boldsymbol{k};s;i} = Y_{k_1\dots k_{s-1}ik_{s+1}\dots k_d}$, *i.e., at the* s^{th} *position of the index* $\boldsymbol{k}$ *the component* k_s *is equal to* i, *and where* $Y_{\boldsymbol{k}}$ *should be taken to be zero if at least one of* $k_1, \dots, k_d$ *is zero.*

In the simple case where $g(x) = x$ and $d = 2$ we have the following result.

Lemma 3.2.6. *Let* $\{S_{\boldsymbol{n}},\ \boldsymbol{n} \in \mathbb{N}^2\}$ *be a multidimensionally indexed demimartingale and* $\{c_{\boldsymbol{n}},\ \boldsymbol{n} \in \mathbb{N}^2\}$ *a nonincreasing array of positive numbers. Assume that there exists a number* $p \geq 1$ *such that* $E\,|S_{ij}|^p < \infty$, *for all* $i \geq 1,\ j \geq 1$ *and*

$$\sum_{i=1}^{\infty} c_{iN}^p E\,|S_{iN}|^p < \infty \ \ \forall \ \ N \in \mathbb{N} \qquad \text{and} \qquad \sum_{i=2}^{\infty}\sum_{j=1}^{\infty} c_{ij}^p E(|S_{ij}|^p - |S_{i-1j}|^p) < \infty$$

or

$$\sum_{j=1}^{\infty} c_{Nj}^p E\,|S_{Nj}|^p < \infty \ \ \forall \ \ N \in \mathbb{N} \qquad \text{and} \qquad \sum_{i=1}^{\infty}\sum_{j=2}^{\infty} c_{ij}^p E(|S_{ij}|^p - |S_{ij-1}|^p) < \infty.$$

Then

$$c_{n_1n_2} S_{n_1n_2} \xrightarrow{a.s.} 0, \ \ as \ (n_1, n_2) \to \infty.$$

Combining Lemma 3.2.6 and Proposition 3.2.4 we get the following strong law of large numbers for a U-statistics based on multidimensionally indexed associated random variables. First a simple auxiliary result is needed.

Lemma 3.2.7.

$$\left[\binom{ij}{m}^{-p} - \binom{i(j+1)}{m}^{-p}\right]\binom{ij}{m}^{p} = O(j^{-1})$$

for $i \geq 1, j \geq 1, m \geq 1$ *and* $p \geq 1$.

Proof.

$$\begin{aligned}
&\left[\binom{ij}{m}^{-p} - \binom{i(j+1)}{m}^{-p}\right]\binom{ij}{m}^{p} \\
\leq\ & p\binom{ij}{m}^{-p+1}\left[\binom{ij}{m}^{-1} - \binom{i(j+1)}{m}^{-1}\right]\binom{ij}{m}^{p} \qquad (3.2.1) \\
=\ & p\binom{ij}{m}\left[\binom{ij}{m}^{-1} - \binom{i(j+1)}{m}^{-1}\right]
\end{aligned}$$

$$\begin{aligned}
&= \left[\binom{i(j+1)}{m} - \binom{ij}{m}\right]\binom{i(j+1)}{m}^{-1} \\
&= \frac{i(j+1)[i(j+1)-1]\cdots[i(j+1)-m+1] - ij(ij-1)\cdots(ij-m+1)}{i(j+1)[i(j+1)-1]\cdots[i(j+1)-m+1]} \\
&< \frac{[i(j+1)]^m - (ij-m+1)^m}{[i(j+1)-m+1]^m} \\
&< \frac{m[i(j+1)]^{m-1}(i+m-1)}{[i(j+1)-m+1]^m} \qquad (3.2.2) \\
&= O(j^{-1})
\end{aligned}$$

where (3.2.1) and (3.2.2) follow from the elementary inequality $x^r - y^r \leq rx^{r-1}(x-y)$ which is valid for $x, y > 0, r \geq 1$. ■

Theorem 3.2.8. *Let $U_{\mathbf{n}}$ be a U-statistic based on a collection of multidimensionally indexed associated random variables ($d = 2$) and on the kernel h. Assume that h is componentiwise nondecreasing and $E|U_{ij}|^p < \infty$ for $p \geq 1$ and for all $i \geq m_1,\ j \geq m_2$. Furthermore assume that*

$$\sum_{i=m_1}^{\infty} E|U_{iN}|^p < \infty \quad \forall \quad N \in \mathbb{N} \qquad \textit{and} \qquad \sum_{i=m_1}^{\infty}\sum_{j=m_2+1}^{\infty} j^{-1}E(|U_{ij}|^p) < \infty$$

or

$$\sum_{j=m_2}^{\infty} E|U_{Nj}|^p < \infty \quad \forall \quad N \in \mathbb{N} \qquad \textit{and} \qquad \sum_{i=m_1+1}^{\infty}\sum_{j=m_2}^{\infty} i^{-1}E(|U_{ij}|^p) < \infty$$

then

$$U_{\mathbf{n}} - E(U_{\mathbf{n}}) \xrightarrow{a.s.} 0, \qquad \mathbf{n} \to \infty.$$

Proof. Let $S_{\mathbf{n}} = \binom{|\mathbf{n}|}{m}U_{\mathbf{n}}$ for $m \leq |\mathbf{n}|$ and $S_{\mathbf{n}} = 0$ for $m > |\mathbf{n}|$. Since h is nondecreasing, $S_{\mathbf{n}}$ is a multidimensionally indexed demimartingale. Clearly $c_{\mathbf{n}} = \binom{|\mathbf{n}|}{m}^{-1}$ is a decreasing sequence of positive numbers.

By Lemma 2.1 of Christofides and Hadjikyriakou (2011) which obtain a strong law of large numbers for multidimensionally indexed demimartingales, where $g(x) = x$ and $d = 2$, we observe that

$$\binom{|\mathbf{n}|}{m}^{-1}(S_{\mathbf{n}}-E(S_{\mathbf{n}}))\xrightarrow{a.s.}0, \qquad \mathbf{n}\to\infty$$

provided that

$$\sum_{i=m_1}^{\infty} c_{iN}^p E|S_{iN}|^p<\infty$$

for all $N\in\mathbb{N}$, and

$$\sum_{i=m_1}^{\infty}\sum_{j=m_2+1}^{\infty} c_{ij}^p E(|S_{ij}|^p-|S_{ij-1}|^p)<\infty.$$

Observe that

$$\sum_{i=m_1}^{\infty} c_{iN}^p E|S_{iN}|^p=\sum_{i=m_1}^{\infty}\binom{iN}{m}^{-p}E\left|\binom{iN}{m}U_{iN}\right|^p=\sum_{i=m_1}^{\infty}E|U_{iN}|^p<\infty$$

for all $N\in\mathbb{N}$, and

$$\sum_{i=m_1}^{\infty}\sum_{j=m_2+1}^{\infty} c_{ij}^p E(|S_{ij}|^p-|S_{ij-1}|^p)$$

$$=\sum_{i=m_1}^{\infty}\sum_{j=m_2+1}^{\infty}\binom{ij}{m}^{-p}E(|S_{ij}|^p-|S_{ij-1}|^p)$$

$$=\sum_{i=m_1}^{\infty}\sum_{j=m_2+1}^{\infty}\left[\binom{ij}{m}^{-p}-\binom{i(j+1)}{m}^{-p}\right]E(|S_{ij}|^p)$$

$$=\sum_{i=m_1}^{\infty}\sum_{j=m_2+1}^{\infty}\left[\binom{ij}{m}^{-p}-\binom{i(j+1)}{m}^{-p}\right]\binom{ij}{m}^{p}E(|U_{ij}|^p)$$

$$<\sum_{i=m_1}^{\infty}\sum_{j=m_2+1}^{\infty} j^{-1}E|U_{ij}|^p<\infty$$

where the last inequality follows from Lemma 3.2.7. ■

Usually strong laws for U-statistics are stated with conditions on the moments of kernel h. Lemma 3.2.10 gives conditions on h under which the moment condition of Theorem 3.2.8 holds true. Before we obtain Lemma 3.2.10 we present a moment bound for associated random fields by Bulinski (1994).

Proposition 3.2.9. *(Bulinski (1994)). Let Ψ be the class of blocks in $\mathbb{Z}^2$, that is, of sets $F = ((a_1, b_1] \times (a_2, b_2]) \cap \mathbb{Z}^2$, $a_i < b_i$, a_i, $b_i \in \Re$, $i = 1, 2$. Let $F = \{(1-\varepsilon, n_1] \times (1-\varepsilon, n_2]\} \cap \mathbb{Z}^2$ where $0 < \varepsilon < 1$ and $|F|$ is the cardinality of F. Let $\{X_{\mathbf{j}},\ \mathbf{j} \in \mathbb{N}^2\}$ be a collection of identically distributed associated multidimensionally indexed random variables. Assume that $E(X_{\mathbf{j}}) = 0$ and for some $p > 2$, $\delta > 0$, $0 \leq \mu < \frac{1+\delta/\kappa}{2}$ where $\kappa = \delta + (p+\delta)(p-2)$ the collection satisfies*

$$\sup_{\mathbf{j} \in \mathbb{Z}^2} E\left|X_{\mathbf{j}}\right|^{p+\delta} < \infty,$$

$$\sup_{\substack{F \in \Psi \\ |F| = |\mathbf{n}|}} \left\{ \sum_{\mathbf{j} \in \mathbf{F}} \left[\sum_{\mathbf{r} \notin \mathbf{F}} cov\left(X_{\mathbf{i}}, X_{\mathbf{j}}\right) \right]^{\delta/\kappa} \right\} = O\left(|\mathbf{n}|\right)^{\mu}.$$

Then

$$\sup_{\substack{F \in \Psi \\ |F| = |\mathbf{n}|}} \left\{ E\left| \sum_{\mathbf{j} \in F} X_{\mathbf{j}} \right|^p \right\} = O\left(|\mathbf{n}|^{p/2}\right).$$

Lemma 3.2.10. *Let Ψ be the class of blocks in $\mathbb{Z}^2$, that is, of sets $F = ((a_1, b_1] \times (a_2, b_2]) \cap \mathbb{Z}^2$, $a_i < b_i$, a_i, $b_i \in \Re$, $i = 1, 2$. Let $F = \{(1-\varepsilon, n_1] \times (1-\varepsilon, n_2]\} \cap \mathbb{Z}^2$ where $0 < \varepsilon < 1$ and $|F|$ is the cardinality of F. Let $\{X_{\mathbf{j}},\ \mathbf{j} \in \mathbb{N}^2\}$ be a collection of identically distributed associated multidimensionally indexed random variables. Let $U_{\boldsymbol{n}}$ be a U-statistic based on the collection of multidimensionally indexed associated random variables and on the kernel h. Assume that h is componentwise nondecreasing, $E(h) = 0$ and for some $p > 2$, $\delta > 0$, $0 \leq \mu < \frac{1+\delta/\kappa}{2}$ where $\kappa = \delta + (p+\delta)(p-2)$ the collection satisfies*

$$\sup_{\mathbf{i}, \mathbf{j} \in \mathbb{Z}^2} E\left|h(X_{\mathbf{i}}, X_{\mathbf{j}})\right|^{p+\delta} < \infty,$$

and

$$\sup_{\substack{F \in \Psi \\ |F| = |\mathbf{n}|}} \left\{ \sum_{\mathbf{i}, \mathbf{j} \in \mathbf{F}} \left[\sum_{\mathbf{t}, \mathbf{r} \notin \mathbf{F}} cov\left[h(X_{\mathbf{i}}, X_{\mathbf{j}}), h(X_{\mathbf{t}}, X_{\mathbf{r}})\right] \right]^{\delta/\kappa} \right\} = O\left(|\mathbf{n}|\right)^{\mu}.$$

Then

$$E\left|U_{\mathbf{n}}\right|^p = O\left(|\mathbf{n}|^{-p/2}\right).$$

Proof.

$$E\left|S_{\mathbf{n}}\right|^p = E\left| \sum_{c} h(X_{\mathbf{i}}, X_{\mathbf{j}}) \right|^p = E\left| \frac{1}{2} \sum_{\mathbf{i} \leq \mathbf{n}} \sum_{\substack{\mathbf{j} \leq \mathbf{n} \\ \mathbf{i} \neq \mathbf{j}}} h(X_{\mathbf{i}}, X_{\mathbf{j}}) \right|^p, \tag{3.2.3}$$

where $\sum_c$ denotes the summation over the $\binom{|\mathbf{n}|}{2}$ combinations of the two distinct elements $\{\mathbf{i}, \mathbf{j}\}$ from $\{(1,1), ..., (n_1, n_2)\}$.

$$E\left|S_{\mathbf{n}}\right|^p \leq \frac{1}{2^p}\left|\mathbf{n}\right|^{p-1}\sum_{\mathbf{i}\leq\mathbf{n}} E\left|\sum_{\mathbf{j}\leq\mathbf{n}} h(X_{\mathbf{i}}, X_{\mathbf{j}})\right|^p, \tag{3.2.4}$$

where (3.2.4) follows from (3.2.3) and Proposition 2.1.1. Since h is a nondecreasing function, the collection of random variables $\{h(X_{\mathbf{i}}, X_{\mathbf{k}}),\ \mathbf{k} \leq \mathbf{n},\ \mathbf{n} \in \mathbb{N}^d\}$ is associated (see property 4 in Esary et al. (1967)). By Proposition 3.2.9 we have that

$$E\left|S_{\mathbf{n}}\right|^p \leq \frac{1}{2^p}\left|\mathbf{n}\right|^{p-1}\left|\mathbf{n}\right| A \left|\mathbf{n}\right|^{p/2}$$

where A is a constant which does not depend on $|\mathbf{n}|$. Finally

$$E\left|U_{\mathbf{n}}\right|^p \leq \left|\mathbf{n}\right|^{-2p} A \left|\mathbf{n}\right|^{3p/2} \leq A\left|\mathbf{n}\right|^{-p/2}$$

$$\Rightarrow E\left|U_{\mathbf{n}}\right|^p = O\left(\left|\mathbf{n}\right|^{-p/2}\right). \quad \blacksquare$$

3.2.2 A strong law in the case of kernels of bounded variation

Recall the definition of kernels of bounded variation of dimension two introduced in Section 3.1.

Theorem 3.2.11. *(Christofides (2004)). A function $f : [a,b] \times [c,d] \to \Re$ is of bounded variation on $[a,b]\times[c,d]$ if and only if it can be written as the difference of two componentwise nondecreasing functions $G,\ H$ on $[a,b]\times[c,d]$ with $\Delta G((q,r),(s,t)) \geq 0$ and $\Delta H((q,r),(s,t)) \geq 0$ for $a \leq q < r \leq b,\ c \leq s < t \leq d$.*

Remark 3.2.12. *The results for functions of bounded variation are presented in the case of functions defined on $\Re^2$. The extension to higher dimensions is straightforward by using induction. However, the notation becomes cumbersome. For example, in the case of a real function of bounded variation defined on the parallelepiped $[a_1,b_1]\times[a_2,b_2]\times[a_3,b_3]$ the statement of Theorem 3.2.11 involves the quantity $\Delta G((q,r),(s,t),(u,v))$ defined as*

$$G(r,t,v)-G(q,t,v)-G(r,s,v)+G(q,s,v)-G(r,t,u)+G(q,t,u)+G(r,s,u)-G(q,s,u)$$

for $a_1 \leq q < r \leq b_1,\ a_2 \leq s < t \leq b_2,\ a_3 \leq u < v \leq b_3$.

Theorem 3.2.13. *Let $U_{\mathbf{n}}$ be a U-statistic based on a collection of multidimensionally indexed associated random variables ($d = 2$) and on the kernel h. Assume that h is a function of bounded variation and $E\,|U_{ij}|^p < \infty$ for $p \geq 1$ and for all $i \geq m_1,\ j \geq m_2$. Furthermore assume that*

$$\sum_{i=m_1}^{\infty} E\,|U_{iN}|^p < \infty \quad \forall \quad N \in \mathbb{N} \qquad \text{and} \qquad \sum_{i=m_1}^{\infty} \sum_{j=m_2+1}^{\infty} j^{-1} E(|U_{ij}|^p) < \infty \tag{3.2.5}$$

or

$$\sum_{j=m_2}^{\infty} E\,|U_{Nj}|^p < \infty \quad \forall \quad N \in \mathbb{N} \qquad \text{and} \qquad \sum_{i=m_1+1}^{\infty} \sum_{j=m_2}^{\infty} i^{-1} E(|U_{ij}|^p) < \infty.$$

Then

$$U_{\mathbf{n}} - E\,(U_{\mathbf{n}}) \xrightarrow{a.s.} 0, \qquad \mathbf{n} \to \infty.$$

Proof. By Theorem 3.1 of Christofides (2004) the kernel h can be written as the difference of two componentwise nondecreasing functions, say G and H whose explicit expressions can be found in the proof of the theorem. Therefore $U_{\mathbf{n}}$ can be expressed as

$$U_{\mathbf{n}} = U_{\mathbf{n}}^{(1)} - U_{\mathbf{n}}^{(2)}, \tag{3.2.6}$$

where $U_{\mathbf{n}}^{(1)}$ and $U_{\mathbf{n}}^{(2)}$ are U-statistics based on the componentwise nondecreasing kernels G and H, respectively. From Theorem 3.2.8 it follows that

$$U_{\mathbf{n}}^{(1)} - E(U_{\mathbf{n}}^{(1)}) \xrightarrow{a.s.} 0, \quad \boldsymbol{n} \to \infty \tag{3.2.7}$$

and

$$U_{\mathbf{n}}^{(2)} - E(U_{\mathbf{n}}^{(2)}) \xrightarrow{a.s.} 0, \quad \boldsymbol{n} \to \infty \tag{3.2.8}$$

provided that

$$\sum_{i=m_1}^{\infty} \sum_{j=m_2+1}^{\infty} j^{-1} E|U_{ij}^{(1)}|^p < \infty \tag{3.2.9}$$

and

$$\sum_{i=m_1}^{\infty} \sum_{j=m_2+1}^{\infty} j^{-1} E|U_{ij}^{(2)}|^p < \infty. \tag{3.2.10}$$

By the construction of the functions G and F both (3.2.9) and (3.2.10) are implied by (3.2.5). The result now follows from (3.2.7) and (3.2.8). ■

Remark 3.2.14. *Condition (3.2.5) is stated in terms of the moments of the U-statistic. This in a way says that convergence depends not only on the kernel h but also on the* *nature of the observations. Let us consider the following example.*

Example 3.2.15. *Let X be a random variable with $E|X| < \infty$ and let*

$$X_{ij} = \frac{\log(ij+1)}{ij} X, \quad i \geq 1, \; j \geq 1.$$

Then $\{X_{ij}, i \geq 1, j \geq 1\}$ are associated by properties (P3) and (P4) of Esary et al. (1967). Let $h(x;y) = |x-y|$ and consider the U-statistic based on $\{X_{ij}, i \geq 1, j \geq 1\}$ and on the kernel h, which is a function of bounded variation. Then

$$U_{\boldsymbol{n}} = |X| \binom{n_1 n_2}{2}^{-1} \sum_{i=1}^{n_1-1} \sum_{j=1}^{n_2-1} \sum_{k=i+1}^{n_1} \sum_{l=j+1}^{n_2} \left| \frac{\log(ij+1)}{ij} - \frac{\log(kl+1)}{kl} \right|.$$

Direct computation shows that condition (3.2.5) for $p=1$ is satisfied and by Theorem 3.2.13 the strong law of large numbers holds. Consider now the U-statistic based on the same kernel, but on the observations $\{Y_{ij}, i \geq 1, j \geq 1\}$ where

$$Y_{ij} = ij \log(ij) X, \quad i \geq 1, \; j \geq 1.$$

Then

$$U_{\boldsymbol{n}} = |X| \binom{n_1 n_2}{2}^{-1} \sum_{i=1}^{n_1-1} \sum_{j=1}^{n_2-1} \sum_{k=i+1}^{n_1} \sum_{l=j+1}^{n_2} |ij \log(ij) - kl \log(kl)|$$

It can easily be shown by direct computation that U_n does not converge. In addition, condition (3.2.5) for $p=1$ is violated.

3.3 Multi-sample U-statistics on collections of associated random variables

U-statistics can also be extended to multisample setups. For example, in a two-sample model, let $\{X_1, X_2, ..., X_{n_1}\}$ be a finite collection of identically distributed associated random variables with distribution $\mathbb{F}$ and $\{Y_1, Y_2, ..., Y_{n_2}\}$ be another finite collection of identically distributed associated random variables with distribution function $\mathbb{G}$. Assume that the two samples are independent. We write $\vartheta = \vartheta(\mathbb{F}, \mathbb{G})$ as

$$\vartheta = E_{\mathbb{F},\mathbb{G}} \left[h\left(X_1, X_2, ..., X_{m_1}; Y_1, Y_2, ..., Y_{m_2}\right) \right]$$

where m_1, m_2 are positive integers $m_1 \leq n_1$, $m_2 \leq n_2$ and the kernel h is a symmetric, in each set of $X_1, X_2, ..., X_{m_1}$ and $Y_1, Y_2, ..., Y_{m_2}$ mapping $\mathbb{R}^{m_1+m_2}$ to $\mathbb{R}$.

We define the generalized or two-sample U-statistic as

$$U_{n_1,n_2} = \binom{n_1}{m_1}^{-1} \binom{n_2}{m_2}^{-1} \sum_{\substack{1 \leq i_1 < i_2 < \cdots < i_{m_1} \leq n_1 \\ 1 \leq j_1 < j_2 < \cdots < j_{m_2} \leq n_2}} h\left(X_{i_1}, X_{i_2}, ..., X_{i_{m_1}}; Y_{j_1}, Y_{j_2}, ..., Y_{j_{m_2}}\right)$$

where $\sum_{\substack{1 \leq i_1 < i_2 < \cdots < i_{m_1} \leq n_1 \\ 1 \leq j_1 < j_2 < \cdots < j_{m_2} \leq n_2}}$ denotes the summation over the $\binom{n_1}{m_1}$ combinations of the m_1 distinct elements $\{i_1, i_2, \ldots, i_{m_1}\}$ from $\{1, ..., n_1\}$ and the $\binom{n_2}{m_2}$ combinations of the m_2 distinct elements $\{j_1, j_2, \ldots, j_{m_2}\}$ from $\{1, ..., n_2\}$ respectively.

U-statistics on more than two independent samples of associated random variables can be defined in a similar way. In the simplest case where $m_1 = m_2 = 1$ we have

$$U_{n_1,n_2} = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} h\left(X_i; Y_j\right).$$

Example 3.3.1. *(Lee (1990)). Let $\{X_1, X_2, ..., X_{n_1}\}$ be a finite collection of associated identically distributed random variables with distribution function $\mathbb{F}$ and $\{Y_1, Y_2, ..., Y_{n_2}\}$ be another finite collection of associated identically distributed random variables with distribution function $\mathbb{G}$. The two samples are independent. Let R_j denote the rank of Y_j in the combined sample. Then the Wilcoxon rank sum statistic is*

$$W = \sum_{j=1}^{n_2} R_j.$$

If we define

$$h(x; y) = \begin{cases} 1 & if\ x < y \\ 0 & otherwise; \end{cases}$$

and

$$U_{n_1,n_2} = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} h\left(X_i; Y_j\right)$$

then in the absence of ties it can be shown that

$$W = n_1 n_2 U_{n_1,n_2} + n_2(n_2 + 1)/2.$$

Here the generalized U-statistic which is the two sample Wilcoxon (Mann-Whitney) statistic is the estimator of $\vartheta = P(X_1 < Y_1)$.

Example 3.3.2. *(Kowalski (2008)). Let* $\{X_1, X_2, ..., X_{n_1}\}$ *be a finite collection of identically distributed associated random variables with a continuous distribution with variance* σ_1^2 *and* $\{Y_1, Y_2, ..., Y_{n_2}\}$ *be another finite collection of identically distributed associated random variables with a continuous distribution function with variance* σ_2^2. *The two samples are independent.*

For the two sample U-statistic, in the case of $m_1 = m_2 = 2$, *let us define a symmetric kernel as follows:*

$$h(x_1, x_2; y_1, y_2) = \frac{1}{2}(x_1 - x_2)^2 - \frac{1}{2}(y_1 - y_2)^2.$$

Then, it follows that

$$\vartheta = E\left[h(X_1, X_2; Y_1, Y_2)\right] = \sigma_1^2 - \sigma_2^2.$$

If $\vartheta = 0$, *then the two populations have the same variance. Thus the two sample U-statistic given by*

$$U_{n_1,n_2} = \binom{n_1}{2}^{-1}\binom{n_2}{2}^{-1} \sum_{1 \le i_1 < i_2 \le n_1} \sum_{1 \le j_1 < j_2 \le n_2} h(X_{i_1}, X_{i_2}; Y_{j_1}, Y_{j_2})$$

can be used to test the null hypothesis of equal variances between the two populations.

The following result shows that a multisample U-statistic on collections of associated random variables, under certain conditions on the kernel, is a multidimensionally indexed demimartingale.

Proposition 3.3.3. *Assume for simplicity that d=2. Let* U_{n_1,n_2} *be a U-statistic with* $m_1 = m_2 = 1$ *based on two independent samples with associated random variables and on the kernel h. Let* $S_{n_1,n_2} = n_1 n_2 U_{n_1,n_2}$ *where* $n_1 \geq m_1, n_2 \geq m_2$. *Assume that h is componentwise nondecreasing and that* $E(h) = 0$. *Then the sequence*

$$\{S_{n_1,n_2},\ n_1 \geq m_1,\ n_2 \geq m_2\}$$

is a multidimensionally indexed demimartingale.

Proof. Let

$$S_{\mathbf{t}} = \sum_{i=1}^{t_1}\sum_{j=1}^{t_2} h(X_i;Y_j),\ \ S_{\mathbf{r}} = \sum_{i=1}^{r_1}\sum_{j=1}^{r_2} h(X_i;Y_j)$$

where $\mathbf{t} = (t_1, t_2)$ and $\mathbf{r} = (r_1, r_2)$ with $\mathbf{r} \leq \mathbf{t}$.

We can write

$$S_{\mathbf{t}} - S_{\mathbf{r}} = \sum_{i=r_1+1}^{t_1}\sum_{j=1}^{r_2} h(X_i;Y_j) + \sum_{i=1}^{r_1}\sum_{j=r_2+1}^{t_2} h(X_i;Y_j) + \sum_{i=r_1+1}^{t_1}\sum_{j=r_2+1}^{t_2} h(X_i;Y_j).$$

Then for any componentwise nondecreasing function f we have

$$\begin{aligned} E\left[(S_{\mathbf{t}} - S_{\mathbf{r}})\, f\left(S_{\mathbf{k}}, \mathbf{k} \leq \mathbf{r}\right)\right] &= \sum_{i=r_1+1}^{t_1}\sum_{j=1}^{r_2} E[h(X_i;Y_j) f\left(S_{\mathbf{k}}, \mathbf{k} \leq \mathbf{r}\right)] \\ &+ \sum_{i=1}^{r_1}\sum_{j=r_2+1}^{t_2} E[h(X_i;Y_j) f\left(S_{\mathbf{k}}, \mathbf{k} \leq \mathbf{r}\right)] + \sum_{i=r_1+1}^{t_1}\sum_{j=r_2+1}^{t_2} E[h(X_i;Y_j) f\left(S_{\mathbf{k}}, \mathbf{k} \leq \mathbf{r}\right)] \\ &= \sum_{i=r_1+1}^{t_1}\sum_{j=1}^{r_2} E[h(X_i;Y_j) g\left(X_1, X_2, ..., X_{r_1}; Y_1, Y_2, ..., Y_{r_2}\right)] \\ &+ \sum_{i=1}^{r_1}\sum_{j=r_2+1}^{t_2} E[h(X_i;Y_j) g\left(X_1, X_2, ..., X_{r_1}; Y_1, Y_2, ..., Y_{r_2}\right)] \\ &+ \sum_{i=r_1+1}^{t_1}\sum_{j=r_2+1}^{t_2} E[h(X_i;Y_j) g\left(X_1, X_2, ..., X_{r_1}; Y_1, Y_2, ..., Y_{r_2}\right)] \\ &\geq 0 \end{aligned}$$

where the function g defined as:

$$\begin{aligned} &g\left(X_1, X_2, ..., X_{r_1}; Y_1, Y_2, ..., Y_{r_2}\right) \\ &= f(h(x_1;y_1), \sum_{i=1}^{1}\sum_{j=1}^{2} h(x_i;y_j), \sum_{i=1}^{2}\sum_{j=1}^{1} h(x_i;y_j), ..., \sum_{i=1}^{r_1}\sum_{j=1}^{r_2} h(x_i;y_j)) \end{aligned}$$

is componentwise nondecreacing since h and f are componentwise nondecreacing, (see property 4 in Esary et al. (1967)). The last inequality follows from the nondecreasingness of the function g and the fact that the sequences $\{X_i, i \geq 1\}$ and $\{Y_j, j \geq 1\}$ are associated random variables, and the two sequences are independent. ■

3.3.1 A strong law in the case of nondecreasing kernels

Theorem 3.3.4. *Let U_{n_1,n_2} be a two-sample U-statistic based on two collections of associated random variables and on the kernel h. Assume that h is componentiwise nondecreasing and $E\left|U_{ij}\right|^p < \infty$ for $p \geq 1$ and for all $i \geq m_1,\ j \geq m_2$. Furthermore assume that*

$$\sum_{i=m_1}^{\infty} E\left|U_{iN}\right|^p < \infty \quad \forall \quad N \in \mathbb{N} \qquad \textit{and} \qquad \sum_{i=m_1}^{\infty}\sum_{j=m_2+1}^{\infty} j^{-1}E(\left|U_{ij}\right|^p) < \infty$$

or

$$\sum_{j=m_2}^{\infty} E\left|U_{Nj}\right|^p < \infty \quad \forall \quad N \in \mathbb{N} \qquad \textit{and} \qquad \sum_{i=m_1+1}^{\infty}\sum_{j=m_2}^{\infty} i^{-1}E(\left|U_{ij}\right|^p) < \infty.$$

Then

$$U_{n_1,n_2} - E\left(U_{n_1,n_2}\right) \xrightarrow{a.s.} 0, \qquad \textit{as} \qquad (n_1, n_2) \to \infty.$$

Proof. Let $S_{n_1,n_2} = \binom{n_1}{m_1}\binom{n_2}{m_2}U_{n_1,n_2}$ for $m_1 \leq n_1$ and $m_2 \leq n_2$. Since h is nondecreasing, S_{n_1,n_2} is a multidimensionally indexed demimartingale. Clearly

$$c_{n_1,n_2} = \binom{n_1}{m_1}^{-1}\binom{n_2}{m_2}^{-1}$$

is a decreasing sequence of positive numbers. By Lemma 2.1 of Christofides and Hadjikyriakou (2011) which obtain a strong law of large numbers for multidimensionally indexed demimartingales, where $g(x) = x$ and $d = 2$, we observe that

$$\binom{n_1}{m_1}^{-1}\binom{n_2}{m_2}^{-1}\left(S_{n_1,n_2} - E\left(S_{n_1,n_2}\right)\right) \xrightarrow{a.s.} 0, \quad \textit{as} \quad (n_1, n_2) \to \infty$$

provided that

$$\sum_{i=m_1}^{\infty} c_{iN}^p E\left|S_{iN}\right|^p < \infty$$

for all $N \in \mathbb{N}$, and

$$\sum_{i=m_1}^{\infty}\sum_{j=m_2+1}^{\infty} c_{ij}^p E(\left|S_{ij}\right|^p - \left|S_{ij-1}\right|^p) < \infty.$$

Observe that

$$\sum_{i=m_1}^{\infty} c_{iN}^p E\left|S_{iN}\right|^p = \sum_{i=m_1}^{\infty}\binom{i}{m_1}^{-p}\binom{N}{m_2}^{-p} E\left|\binom{i}{m_1}\binom{N}{m_2}U_{iN}\right|^p = \sum_{i=m_1}^{\infty} E\left|U_{iN}\right|^p < \infty$$

for all $N \in \mathbb{N}$, and

$$\begin{aligned}
&\sum_{i=m_1}^{\infty}\sum_{j=m_2+1}^{\infty} c_{ij}^{p} E(|S_{ij}|^p - |S_{ij-1}|^p) \\
&= \sum_{i=m_1}^{\infty}\sum_{j=m_2+1}^{\infty} \binom{i}{m_1}^{-p}\binom{j}{m_2}^{-p} E(|S_{ij}|^p - |S_{ij-1}|^p) \\
&= \sum_{i=m_1}^{\infty}\sum_{j=m_2+1}^{\infty} \left[\binom{i}{m_1}^{-p}\binom{j}{m_2}^{-p} - \binom{i}{m_1}^{-p}\binom{j+1}{m_2}^{-p}\right] E(|S_{ij}|^p) \\
&= \sum_{i=m_1}^{\infty}\sum_{j=m_2+1}^{\infty} \left[\binom{i}{m_1}^{-p}\binom{j}{m_2}^{-p} - \binom{i}{m_1}^{-p}\binom{j+1}{m_2}^{-p}\right] \binom{i}{m_1}^{p}\binom{j}{m_2}^{p} E(|U_{ij}|^p) \\
&= \sum_{i=m_1}^{\infty}\sum_{j=m_2+1}^{\infty} \left[\binom{j}{m_2}^{-p} - \binom{j+1}{m_2}^{-p}\right] \binom{j}{m_2}^{p} E(|U_{ij}|^p) \\
&< p\sum_{i=m_1}^{\infty}\sum_{j=m_2+1}^{\infty} \binom{j}{m_2}^{-p+1} \left[\binom{j}{m_2}^{-1} - \binom{j+1}{m_2}^{-1}\right] \binom{j}{m_2}^{p} E(|U_{ij}|^p) \\
&= pm_2\sum_{i=1}^{\infty}\sum_{j=m_2+1}^{\infty} (j+1)^{-1} E(|U_{ij}|^p) \\
&< \sum_{i=m_1}^{\infty}\sum_{j=m_2+1}^{\infty} j^{-1} E\,|U_{ij}|^p < \infty,
\end{aligned}$$

where the last equality follows from the inequality $x^r - y^r \leq r x^{r-1}(x-y)$ which is valid for $x, y > 0, r \geq 1$. ■

Chapter 4

Stochastic orders and distances for U-statistics

Limit theorems for U-statistics are usually considered for the Kolmogorov metric. Our goal is to depart from this approach and provide an alternative approach for the distance between a U-statistic and a normal random variable. In particular, in Section 4.1, we give the distance between a U-statistic U_n based on associated random variables and a U-statistic U_n^* based on i.i.d. random variables. Asymptotic normality for U-statistics based on associated random variables is presented in Section 4.2. In Section 4.3 we obtain the distance between a U-statistic based on i.i.d. random variables and a normal random variable by utilizing Zolotarev's ideal metric. This result also establishes a central limit theorem for U-statistics, with an alternative technique, using probability metrics. Corresponding results are also investigated for von Mises statistics. In Section 4.4 we also prove similar results for U-statistics based on negatively associated random variables also under Zolotarev's ideal metric.

4.1 Distance between U_n based on associated random variables and U_n^* based on i.i.d. random variables

Generally speaking, stochastic ordering tries to order random variables according to an appropriate criterion. In this section before we study our main object which is the distance between a U-statistic U_n based on associated random variables and a

U-statistic U_n^* based on i.i.d. random variables, we present some definitions and results that are connected with stochastic orderings.

4.1.1 Stochastic ordering of random variables

Definition 4.1.1. *(a) A random variable X is said to be smaller than a random variable Y in the convex order, denoted by $X \preccurlyeq_{cx} Y$, if $Ef(X) \leqslant Ef(Y)$ for all convex functions f for which the expectations exist.*

(b) A random variable X is smaller than a random variable Y in the increasing convex order, denoted by $X \preccurlyeq_{icx} Y$, if $Ef(X) \leqslant Ef(Y)$ for all increasing convex functions f for which the expectations exist.

Similarly, one may define the so-called concave and increasing concave orders denoted by $\preccurlyeq_{cv}$ and $\preccurlyeq_{icv}$ respectively.

Lemma 4.1.2. *(Shaked and Shanthikumar (1997), p.197). Let X and Y be a pair of random variables. If $X \preccurlyeq_{icx} Y$ and $E(X) = E(Y)$, then $X \preccurlyeq_{cx} Y$.*

Now we turn our attention to a multivariate stochastic order which is based on the notion of supermodularity. For any two points $\mathbf{x} = (x_1, x_2, ..., x_n)$ and $\mathbf{y} = (y_1, y_2, ..., y_n)$, $\mathbf{x},\ \mathbf{y} \in \Re^n$, define the componentwise maximum as

$$\mathbf{x} \vee \mathbf{y} := (\max\{x_1, y_1\}, \max\{x_2, y_2\}, ..., \max\{x_n, y_n\})$$

and the componentwise minimum as

$$\mathbf{x} \wedge \mathbf{y} := (\min\{x_1, y_1\}, \min\{x_2, y_2\}, ..., \min\{x_n, y_n\}).$$

Definition 4.1.3. *A function $f : \Re^n \to \Re$ is called supermodular (superadditive or L-superadditive) if*

$$f(\boldsymbol{x} \vee \boldsymbol{y}) + f(\boldsymbol{x} \wedge \boldsymbol{y}) \geq f(\boldsymbol{x}) + f(\boldsymbol{y}), \quad \forall\ \boldsymbol{x}, \boldsymbol{y} \in \Re^n.$$

On the other hand, a function is called submodular (subadditive or L-subadditive) if the reverse inequality holds true. A function f is supermodular if and only if -f is submodular.

Definition 4.1.4. *A random vector* $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ *is said to be smaller than a random vector* $\boldsymbol{Y} = (Y_1, Y_2, ..., Y_n)$ *in the supermodular order, denoted by* $X \preccurlyeq_{sm} Y$*, if* $Ef(\boldsymbol{X}) \leqslant Ef(\boldsymbol{Y})$ *for all supermodular functions f for which the expectations exist.*

Consider a collection of real-valued random variables $X_1, X_2, ..., X_n$. Then, throughout this chapter we shall use the notation $X_1^*, X_2^*, ..., X_n^*$ to denote independent random variables such that $X_i =_{st} X_i^*$ for all $i = 1, ..., n$ (where $=_{st}$ denotes equality in distribution).

Lemma 4.1.5. *(Christofides and Vaggelatou (2004)).*

(a) Let $X_1, X_2, ..., X_n$ *be a collection of weakly associated r.v.'s. Then*

$$\varphi(X_1, X_2, ..., X_n) \succcurlyeq_{icx} \varphi(X_1^*, X_2^*, ..., X_n^*)$$

for every φ *monotone and supermodular.*

(b) If $X_1, X_2, ..., X_n$ *is a collection of negatively associated r.v.'s, then*

$$\varphi(X_1, X_2, ..., X_n) \preccurlyeq_{icx} \varphi(X_1^*, X_2^*, ..., X_n^*)$$

for every φ *monotone and supermodular.*

Lemma 4.1.6. *If a function* $f : M \to R$ *defined on a nonempty subset* M *of* $\mathbb{R}^n$ *and taking real values is convex, then* $g(x) = f(cx)$ *is also convex, where c is a real number.*

Proof. For any $x, y \in M$ and every $\lambda \in [0, 1]$, we have that

$$g(\lambda x + (1 - \lambda)y) = f(c\lambda x + c(1 - \lambda)y).$$

Now from the convexity of f we have that function g is also convex,

$$f(c\lambda x + c(1 - \lambda)y) \leq \lambda f(cx) + (1 - \lambda)f(cy) \leq \lambda g(x) + (1 - \lambda)g(y). \quad \blacksquare$$

Lemma 4.1.7. *Let* X, Y *be random variables such that* $X \preccurlyeq_{cx} Y$*. For* $a > 0$*, then*

$$aX \preccurlyeq_{cx} aY.$$

Proof. From the definition of the convex order we have that $Ef(X) \leqslant Ef(Y)$ for all convex functions f for which the expectations exist. According to Lemma 4.1.6 we have that $Ef(aX) \leqslant Ef(aY)$ because $f(ax)$ is also convex. So we have the result. ■

Remark 4.1.8. *It is obvious that in the same manner we can prove that if f is an increasing convex function then $g(x) = f(cx)$ is also increasing convex, where $c > 0$.*

Lemma 4.1.9. *(Shaked and Shanthikumar (1997), p.123). Let X be a random variable with $E(X) = 0$. Then*

$$X \preccurlyeq_{cx} cX,$$

whenever $c \geq 1$.

Lemma 4.1.10. *Let X, Y be random variables such that $X \preccurlyeq_{cx} Y$, and $0 < a \leq b$. Then*

$$aX \preccurlyeq_{cx} bY$$

respectively.

Proof. Applying Lemma 4.1.7, for $a > 0$ we have that

$$X \preccurlyeq_{cx} Y \Longrightarrow aX \preccurlyeq_{cx} aY.$$

Since $\frac{b}{a} \geq 1$, from Lemma 4.1.9 we arrive at

$$aX \preccurlyeq_{cx} aY \Longrightarrow aX \preccurlyeq_{cx} a\frac{b}{a}Y \Longrightarrow aX \preccurlyeq_{cx} bY. \quad ■$$

Remark 4.1.11. *It is obvious that in the same manner we can prove that the previous result also applies in the case of $X \preccurlyeq_{icx} Y$.*

Lemma 4.1.12. *Let $\{X_n, n \geq 1\}$ be a finite collection of identically distributed associated stationary random variables with a continuous distribution F and $\{X_n^*, n \geq 1\}$ be a finite collection of identically distributed independent random variables with the same continuous distribution F. Moreover let U_n be the U-statistic based on the associated random variables and on the kernel h assumed to be monotone and supermodular. Let U_n^* be the U-statistic based on the independent random variables and on the kernel h. Assume that $E[h(X_1, X_2)] = E[h(X_1^*, X_2^*)] = 0$. If $\sigma_U^2 = \sigma_1^2 + 2\sum_{j=1}^{\infty} \sigma_{1j}^2$ where*

$\sigma_1^2 = \mathrm{Var}[h_1(X_1)]$ *and* $\sigma_{1j}^2 = \mathrm{cov}(h_1(X_1), h_1(X_{1+j}))$, *then*

$$\frac{n^{\frac{1}{2}} U_n^*}{2\sigma_U} \preccurlyeq_{cx} \frac{n^{\frac{1}{2}} U_n}{2\sigma_U}.$$

Proof. Applying Lemma 4.1.5 of Christofides and Vaggelatou (2004), if h is monotone and supermodular

$$\binom{n}{2} U_n^* \preccurlyeq_{icx} \binom{n}{2} U_n$$

because the summation of monotone and supermodular functions is monotone and supermodular. Furthermore from Lemma 4.1.2 we have that

$$\binom{n}{2} U_n^* \preccurlyeq_{cx} \binom{n}{2} U_n,$$

because $E[h(X_1, X_2)] = E[h(X_1^*, X_2^*)]$. Finally applying Lemma 4.1.10 we get that

$$\frac{n^{\frac{1}{2}} U_n^*}{2\sigma_U} \preccurlyeq_{cx} \frac{n^{\frac{1}{2}} U_n}{2\sigma_U}. \blacksquare$$

Proposition 4.1.13. *(Boutsikas and Vaggelatou (2002)). Let us denote by* $\yen_U^s(\Re)$, $U \subseteq \Re$, *the space of all random variables* X *defined on the probability space* $(\Omega, \mathcal{A}, \mathcal{P})$ *and taking values in* U *with* $E|X|^s < \infty$. *If* $X,\ Y \ \in \ \yen^2(\Re)$ *and* $X \preceq_{cx} Y$, *then*

$$\zeta_2(X, Y) = \frac{1}{2}\left[\mathrm{Var}(Y) - \mathrm{Var}(X)\right].$$

Proposition 4.1.14. *The variance of the U-statistic based on i.i.d. random variables is*

$$\mathrm{Var}(U_n) = \frac{4\sigma_1^2}{n} + o\left(\frac{1}{n}\right).$$

Proposition 4.1.15. *Let* U_n *be a U-statistic based on stationary associated (or negatively associated) random variables and* $\sigma_U^2 = \sigma_1^2 + 2\sum_{j=1}^{\infty} \sigma_{1j}^2$. *Then*

$$\mathrm{Var}(U_n) = \frac{4\sigma_U^2}{n} + o\left(\frac{1}{n}\right),$$

where $\sigma_1^2 = \mathrm{Var}[h_1(X_1)] < \infty$ *and* $\sigma_{1j}^2 = \mathrm{cov}(h_1(X_1), h_1(X_{1+j}))$.

4.1.2 Distance between U_n and U_n^*

Theorem 4.1.16. *Let $\{X_n, n \geq 1\}$ be a finite collection of identically distributed associated stationary random variables with a continuous distribution F and $\{X_n^*, n \geq 1\}$ be a finite collection of identically distributed independent random variables with the same continuous distribution F. Moreover let U_n be the U-statistic based on the associated random variables and on the kernel h assumed to be monotone and supermodular. Let U_n^* be the U-statistic based on the independent random variables and on the kernel h. Assume that $E[h(X_1, X_2)] = E[h(X_1^*, X_2^*)] = 0$. If $\sigma_U^2 = \sigma_1^2 + 2\sum_{j=1}^{\infty} \sigma_{1j}^2$ where $\sigma_1^2 = \mathrm{Var}[h_1(X_1)]$ and $\sigma_{1j}^2 = \mathrm{cov}(h_1(X_1), h_1(X_{1+j}))$. Then*

$$\zeta_2\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_U}, \frac{n^{\frac{1}{2}}U_n^*}{2\sigma_1}\right) = \left(1 - \frac{\sigma_1^2}{2\sigma_U^2}\right) + o(1).$$

Proof. Using the triangular inequality and applying Lemma 4.1.12 and Proposition 4.1.13 we have that

$$\zeta_2\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_U}, \frac{n^{\frac{1}{2}}U_n^*}{2\sigma_1}\right) \leq \zeta_2\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_U}, \frac{n^{\frac{1}{2}}U_n^*}{2\sigma_U}\right) + \zeta_2\left(\frac{n^{\frac{1}{2}}U_n^*}{2\sigma_U}, \frac{n^{\frac{1}{2}}U_n^*}{2\sigma_1}\right)$$

$$= \frac{1}{2}\left[\mathrm{Var}(\frac{n^{\frac{1}{2}}U_n}{2\sigma_U}) - \mathrm{Var}(\frac{n^{\frac{1}{2}}U_n^*}{2\sigma_U})\right] + \frac{1}{2}\left[\mathrm{Var}(\frac{n^{\frac{1}{2}}U_n^*}{2\sigma_1}) - \mathrm{Var}(\frac{n^{\frac{1}{2}}U_n^*}{2\sigma_U})\right]$$

$$= \frac{n}{8\sigma_U^2}\left(\frac{4\sigma_U^2}{n} + o\left(\frac{1}{n}\right)\right) - \frac{n}{8\sigma_U^2}\left(\frac{4\sigma_1^2}{n} + o\left(\frac{1}{n}\right)\right)$$

$$+ \frac{n}{8\sigma_1^2}\left(\frac{4\sigma_1^2}{n} + o\left(\frac{1}{n}\right)\right) - \frac{n}{8\sigma_U^2}\left(\frac{4\sigma_1^2}{n} + o\left(\frac{1}{n}\right)\right)$$

$$= \left(1 - \frac{\sigma_1^2}{2\sigma_U^2}\right) + o(1). \quad \blacksquare$$

4.2 Asymptotic normality for U-statistics based on associated random variables

In this section we investigate the distance between a U-statistic based on a collection of identically distributed associated random variables with distribution function F and a normal random variable under the Kolmogorov metric. We present an alternative way to the approach of Garg and Dewan (2015) to prove asymptotic normality for this type of U-statistics.

4.2.1 Definitions and related results

Below is a list of various metrics and related results that are needed for our exposition.

Proposition 4.2.1. *(Lyapunov's Inequality). Let* X *be a random variable, with* $E \mid X \mid^t < \infty$*. Then, for* $0 < s \leq t < \infty$,

$$\left[E\left|X\right|^s\right]^{\frac{1}{s}} \leq \left[E\left|X\right|^t\right]^{\frac{1}{t}}.$$

Definition 4.2.2. *Let* $\tau_s\left(X,Y\right) = E\left|X^{(s)} - Y^{(s)}\right|$*, where* $x^{(s)} = x\left|x\right|^{s-1}$ *for* $s \geq 1$*, and*

$$\widehat{\tau_s}\left(X,Y\right) = \inf \tau_s\left(X,Y\right),$$

where the infimum is taken over all joint distributions $P_{X,Y}$ *whose marginal distributions* $P\left(X < x\right)$ *and* $P\left(Y < y\right)$ *are fixed.*

Proposition 4.2.3. *(Sharakhmetov (2004)). The metric* $\kappa_s\left(X,Y\right)$ *is minimal for the metric* $\tau_s\left(X,Y\right)$*, that is,* $\kappa_s\left(X,Y\right) \leq \tau_s\left(X,Y\right)$ *and* $\widehat{\tau_s}\left(X,Y\right) = \kappa_s\left(X,Y\right)$*.*

Definition 4.2.4. *Let* X, Y *be random variables. Then* $\zeta_{m,p}$ *is defined as*

$$\zeta_{m,p}\left(X,Y\right) = \sup_f \left\{\left|Ef(X) - Ef(Y)\right| : \left\|f^{(m+1)}\right\|_q \leqslant 1\right\},$$

where $1/p + 1/q = 1$ *and* $m = 0, 1, 2, ...$ *By* $f^{(m+1)}$ *we denote the* $(m+1)$*th derivative of the density function* f *and* $||.||_q$ *denotes the* L^q *norm.*

Remark 4.2.5. *(Rachev (1991), p. 270). When* $m = 0$ *and* $p = 1$*, we have that*

$$\zeta_{0,1}\left(X,Y\right) \leq \kappa_1\left(X,Y\right).$$

Proposition 4.2.6. *(Rachev (1991), p. 303). Let* X, Y *random variables. Then*

$$L\left(X,Y\right) \leqslant \left[c_{m,p}\zeta_{m,p}\left(X,Y\right)\right]^{1/(r+1)},$$

where $L(X,Y)$ *is the Levy metric,* $r = m + 1/p$ *and*

$$c_{m,p} = \frac{(2m+2)!(2m+3)^{1/2}}{(m+1)!(3-2/p)^{1/2}}.$$

Applying Proposition 4.2.6 we have the next two remarks.

Remark 4.2.7. *(Rachev (1991), p.258).*

$$L(X,Y) \leqslant [4\zeta_2(X,Y)]^{1/3}.$$

Remark 4.2.8. *Combining Remark 4.2.5 and Proposition 4.2.6 we have that,*

$$L(X,Y) \leqslant \left[2\sqrt{3}\,\kappa_1(X,Y)\right]^{1/2}.$$

Proposition 4.2.9. *(Rachev (1991), p. 303). If* Y *has a bounded density* p_Y*, then*

$$\rho(X,Y) \leqslant \left(1 + \sup_{x\epsilon\Re} p_Y(x)\right) L(X,Y).$$

Proposition 4.2.10. *(Boutsikas and Vaggelatou (2002)). Let* $X_1, X_2, \ldots$ *be a strictly stationary sequence of associated random variables such that* $E(X_1) = 0$ *and* $0 < E(X_1^2) < \infty$*. If* $\sigma^2 := E(X_1^2) + 2\sum_{j=2}^{\infty} E(X_1 X_j) < \infty$*, then, for* $n = mk$*,*

$$\zeta_2\left(\frac{\sum_i^n X_i}{\sqrt{n}}, Y\right) \leq \frac{a_k}{k}\left[\sigma^2 - E\left(X_1^2\right)\right] + 2\left(1 - \frac{a_k}{k}\right) u(a_k) + c\frac{\rho_k}{m^{\frac{1}{2}}}$$

for some constant $c > 0$*, where* Y *is a standard normal random variable,* $\{a_k\}$ *is any sequence of positive integers such that* $a_k \leq k$*,* $a_k \to \infty$*,* $\frac{a_k}{k} \to 0$ *as* $k \to \infty$*,* $\rho_k := E\left|k^{-\frac{1}{2}}\sum_{i=1}^k X_i\right|^3$ *and* $u(i) := \sum_{j=i+1}^{\infty} E(X_1 X_j) \to 0$ *as* $i \to \infty$*.*

Lemma 4.2.11. *Let* U_n *be a U-statistic of dimension two based on stationary associated random variables and on the kernel* h*. Assume that* $E(h) = 0$ *and* $0 < E(h^2) < \infty$*. If* $\sigma_U^2 = \sigma_1^2 + 2\sum_{j=1}^{\infty}\sigma_{1j}^2$ *where* $\sigma_1^2 = \mathrm{Var}[h_1(X_1)]$ *and* $\sigma_{1j}^2 = \mathrm{cov}(h_1(X_1), h_1(X_{1+j})) < \infty$*, then, for* $n = mk$*,*

$$\zeta_2\left(n^{-\frac{1}{2}}\sum_{i=1}^n h^{(1)}(X_i), Y\right) \leq \frac{a_k}{k}\left[\sigma_U^2 - E\left(h^{(1)}(X_1)\right)^2\right] + 2\left(1 - \frac{a_k}{k}\right) u(a_k) + c\frac{\rho_k}{m^{\frac{1}{2}}}$$

for some constant $c > 0$*, where* Y *is a standard normal random variable,* $\{a_k\}$ *is any* *sequence of positive integers such that* $a_k \leq k$, $a_k \to \infty$, $\frac{a_k}{k} \to 0$ *as* $k \to \infty$, $\rho_k := E\left|k^{-\frac{1}{2}} \sum_{i=1}^{k} h^{(1)}(X_i)\right|^3$ *and* $u(i) := \sum_{j=i+1}^{\infty} E\left(h^{(1)}(X_1) h^{(1)}(X_j)\right) \to 0$ *as* $i \to \infty$.

Proof. Apply Lemma 4.2.10. ■

Remark 4.2.12. *Note that*

$$\zeta_2\left(n^{-\frac{1}{2}} \sum_{i=1}^{n} h^{(1)}(X_i), Y\right) = o(1), \quad \text{as } n \to \infty.$$

Definition 4.2.13. *(Garg and Dewan (2015)). The Vitali variation of a function* $f : [a,b] \to \mathbb{R}$, *where* $[a,b] = \{\mathbf{x} \in \mathbb{R}^k : a \leq \mathbf{x} \leq b\}$, $a, b \in \mathbb{R}^k, k \in \mathbb{N}$ *is defined as* $\|f\|_V = \sup \sum_{R \in \mathbb{A}} |\Delta_R f|$. *The supremum is taken over all finite collections of k-dimensional rectangles* $\mathbb{A} = \{R_i : 1 \leq i \leq m\}$ *such that* $\bigcup_{i=1}^{m} R_i = [a,b]$, *and the interiors of any two rectangles in* $\mathbb{A}$ *are disjoint. Here, if* $R = [c,d]$, *a k-dimensional rectangle contained in* $[a,b]$, *then,* $\Delta_R f = \sum_{I \subseteq \{1,2,\ldots,k\}} (-1)^{|I|} f(\boldsymbol{x}_I)$, *where,* $\boldsymbol{x}_I$ *is the vector in* $\mathbb{R}^k$ *whose ith element is given by* c_i *if* $i \in I$, *or by* d_i *if* $i \notin I$, $f_\emptyset = f(b)$. *For instance, if* $k = 2$ *and* $R = [c_1, d_1] \times [c_2, d_2]$ *then,* $\Delta_R f = f(d_1, d_2) - f(c_1, d_2) - f(d_1, c_2) + f(c_1, c_2)$.

Definition 4.2.14. *(Garg and Dewan (2015)). The Hardy-Krause variation of a function* $f : [a,b] \to \mathbb{R}$, *where* $[a,b] = \{\mathbf{x} \in \mathbb{R}^k : a \leq \mathbf{x} \leq b\}$, $a, b \in \mathbb{R}^k, k \in \mathbb{N}$ *is given by* $\|f\|_{HK} = \sum_{\emptyset \neq I \subseteq 1,\ldots,k]} \|f_I\|_V$. *Here, given a non-empty set* $I \subseteq \{1, 2, ..., k\}$, f_I *denotes the real valued function on* $\prod_{i \in I}[a_i, b_i]$ *obtained by setting the ith argument of* f *equal to* b_i *whenever* $i \notin I$.

Remark 4.2.15. *When* $k = 1$, *the Hardy-Krause variation is equivalent to the Vitali variation and hence the standard definition of total variation.*

Lemma 4.2.16. *(Garg and Dewan (2015)). Let* $\{X_n, \ n \geq 1\}$ *be a sequence of stationary associated random variables with* $|X_n| < C_1 < \infty$, *for all* $n \geq 1$. *Assume that the density function of* X_1, *denoted by* f, *is bounded. If* $h^{(2)}(x,y)$ *is a degenerate symmetric kernel of degree 2 which is of bounded Hardy-Krause variation and left continuous, then, under the condition*

$$\sum_{j=1}^{\infty} \mathrm{Cov}(X_1, X_j)^{\gamma} < \infty,$$

for some $0 < \gamma < \frac{1}{6}$,

$$\sum_{1 \leq i < j \leq n} \sum_{1 \leq k < l \leq n} \left| E\left(h^{(2)}\left(X_i, X_j\right) h^{(2)}\left(X_k, X_l\right)\right)\right| = o(n^3).$$

4.2.2 Asymptotic normality

Lemma 4.2.17. *Assume the conditions of Lemma* 4.2.16 *hold and let*

$$\Delta\left(i, j, k, l\right) = cov\left[h^{(2)}\left(X_i, X_j\right), h^{(2)}\left(X_k, X_l\right)\right].$$

Then

$$E\left[H_n^{(2)}\right]^2 = o(n^{-1}).$$

Proof.

$$E\left[H_n^{(2)}\right]^2 = \binom{n}{2}^{-2} \sum_{1 \leq i < j \leq n} \sum_{1 \leq k < l \leq n} \Delta\left(i, j, k, l\right)$$

$$\binom{n}{2}^{-2} \sum_{1 \leq i < j \leq n} \sum_{1 \leq k < l \leq n} \left| E\left(h^{(2)}\left(X_i, X_j\right) h^{(2)}\left(X_k, X_l\right)\right)\right|$$

$$= o(n^{-1}). \quad \blacksquare$$

Lemma 4.2.18. *Assume the conditions of Lemma* 4.2.16 *hold. Then*

$$E\left|\sum_{1 \leq i < j \leq n} h^{(2)}\left(X_i, X_j\right)\right| \leq \left[\sum_{1 \leq i < j \leq n} \sum_{1 \leq k < l \leq n} \Delta\left(i, j, k, l\right)\right]^{\frac{1}{2}}$$

$$= o\left(n^{\frac{3}{2}}\right).$$

Proof. Applying Lemma 4.2.17 and Proposition 4.2.1 we get

$$E\left|\sum_{1 \leq i < j \leq n} h^{(2)}\left(X_i, X_j\right)\right| = \binom{n}{2} E\left|H_n^{(2)}\right|$$

$$\leq \binom{n}{2}\left[E\left|H_n^{(2)}\right|^2\right]^{\frac{1}{2}}$$

$$= \binom{n}{2}\left[\binom{n}{2}^{-2} \sum_{1 \leq i < j \leq n} \sum_{1 \leq k < l \leq n} \Delta\left(i, j, k, l\right)\right]^{\frac{1}{2}}$$

$$= \left[\sum_{1 \le i < j \le n} \sum_{1 \le k < l \le n} \Delta(i,j,k,l) \right]^{\frac{1}{2}}$$

$$= o\left(n^{\frac{3}{2}}\right). \quad \blacksquare$$

Lemma 4.2.19. *Assume the conditions of Lemma* 4.2.16 *hold. Then*

$$\kappa_1 \left(\frac{n^{\frac{1}{2}} U_n}{2\sigma_U}, \frac{1}{n^{\frac{1}{2}}\sigma_U} \sum_{i=1}^{n} h^{(1)}(X_i) \right) = o(1).$$

Proof. Applying Proposition 4.2.3 we have that

$$\kappa_1 \left(\frac{n^{\frac{1}{2}} U_n}{2\sigma_U}, \frac{1}{n^{\frac{1}{2}}\sigma_U} \sum_{i=1}^{n} h^{(1)}(X_i) \right) \le \tau_1 \left(\frac{n^{\frac{1}{2}} U_n}{2\sigma_U}, \frac{1}{n^{\frac{1}{2}}\sigma_U} \sum_{i=1}^{n} h^{(1)}(X_i) \right)$$

$$= E \left| \frac{n^{\frac{1}{2}} U_n}{2\sigma_U} - \frac{1}{n^{\frac{1}{2}}\sigma_U} \sum_{i=1}^{n} h^{(1)}(X_i) \right|$$

$$= E \left| \frac{1}{n^{\frac{1}{2}}(n-1)\sigma_U} \sum_{1 \le i < j \le n} h^{(2)}(X_i, X_j) \right|$$

$$= \frac{1}{n^{\frac{1}{2}}(n-1)\sigma_U} E \left| \sum_{1 \le i < j \le n} h^{(2)}(X_i, X_j) \right|.$$

Finally, from Lemma 4.2.18 we have that

$$\kappa_1 \left(\frac{n^{\frac{1}{2}} U_n}{2\sigma_U}, \frac{1}{n^{\frac{1}{2}}\sigma_U} \sum_{i=1}^{n} h^{(1)}(X_i) \right) = o(1). \quad \blacksquare$$

Theorem 4.2.20. *Let* $\{X_n, n \ge 1\}$ *be a finite collection of identically distributed associated stationary random variables with a distribution function* F. *Assume the conditions of Lemma* 4.2.16 *hold. Moreover let* U_n *be a U-statistic of dimension two based on the kernel* h *and the given observations. Also assume that* $E(h) = 0$, $0 < E(h^2) < \infty$ *and* $\sum_{j=i+1}^{\infty} E(X_1 X_j) \to 0$ *as* $i \to \infty$. *Then*

$$\rho \left(\frac{n^{\frac{1}{2}} U_n}{2\sigma_U}, Y \right) \to 0 \quad as \quad n \to \infty,$$

where Y *is a standard normal random variable.*

Proof. Using the triangular inequality, Proposition 4.2.9 and the inequalities between the metrics (Remarks 4.2.7 and 4.2.8), we have that

$$\rho\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_U}, Y\right) \leq \rho\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_U}, \frac{1}{n^{\frac{1}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}\left(X_i\right)\right) + \rho\left(\frac{1}{n^{\frac{1}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}\left(X_i\right), Y\right).$$

$$\leq c_1\left[2\sqrt{3}\kappa_1\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_U}, \frac{1}{n^{\frac{1}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}\left(X_i\right)\right)\right]^{1/2}$$

$$+ c_2\left[4\zeta_2\left(\frac{1}{n^{\frac{1}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}\left(X_i\right), Y\right)\right]^{1/3},$$

where c_1, c_2 are positive constants. Finally combining Lemmas 4.2.11 and 4.2.19 we have that

$$\rho\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_U}, Y\right) = o\left(1\right). \quad \blacksquare$$

4.3 Distance between a U-statistic based on i.i.d. observations and a normal random variable

Limit theorems for U-statistics are usually considered for the Kolmogorov metric. Our goal is to depart from this approach and provide an alternative approach for the distance between a U-statistic and a normal random variable. In particular, we obtain the distance between a U-statistic and a normal random variable by utilizing Zolotarev's ideal metric. This result is also established as a central limit theorem for U-statistics, with an alternative technique, using probability metrics. Corresponding results are also investigated for von Mises statistics.

4.3.1 Definitions and notation

As mentioned before, our goal in this section is to calculate the distance between U-statistics based on i.i.d. random variables and a standard normal variable with an alternative method, using probability metrics. The following definitions and auxiliary results are essential for our exposition.

Proposition 4.3.1. *(Rachev (1991), p. 258). Let X, Y be random variables. Consider* *the mean metric κ_2 and the Zolotarev's ideal metric ζ_2. Then*

$$2\zeta_2(X,Y) \le \kappa_2(X,Y).$$

Proposition 4.3.2. *(Rachev (1991), p. 262). Let X, Y be random variables and $0 < \delta \le 1$. Then for any $N > 0$,*

$$\begin{aligned}\frac{1}{2}\kappa_2(X,Y) &= \int |t|\,|F_X(t) - F_Y(t)|\,dt \\ &\le N\kappa_1(X,Y) + N^{-\delta}\left[E|X|^{2+\delta} + E|Y|^{2+\delta}\right].\end{aligned}$$

Next, we present a closed-form expression for the variance and a moment bound for U-statistics and V-statistics.

Proposition 4.3.3. *The variance of the U-statistic with kernel based on i.i.d. random variables is*

$$\operatorname{Var}(U_n) = \frac{4\sigma_1^2}{n} + o\left(n^{-1}\right).$$

Proposition 4.3.4. *(Serfling (1980), p. 185). Let r be a real number, $r \ge 2$. Assume that $E|h|^r < \infty$ and $E(h) = 0$. Then*

$$E|U_n|^r = O\left(n^{-\frac{r}{2}}\right), \qquad n \to \infty.$$

Proposition 4.3.5. *(Serfling (1980), p. 206). Let r be a positive integer. Assume that $E|h|^r < \infty$. Then*

$$E|U_n - V_n|^r = O\left(n^{-r}\right), \qquad n \to \infty.$$

Lemma 4.3.6. *Let r be a positive integer, $r \ge 2$. Suppose that $E|h|^r < \infty$. Then*

$$E|V_n|^r = O\left(n^{-\frac{r}{2}}\right), \qquad n \to \infty.$$

Proof. Using Minkowski's inequality we have that

$$E|V_n|^r = E|V_n - U_n + U_n|^r \le 2^{r-1}\left[E|V_n - U_n|^r + E|U_n|^r\right].$$

Moreover, combining Propositions 4.3.4 and 4.3.5 we have that

$$E\left|V_n\right|^r = O(n^{-\frac{r}{2}}). \quad \blacksquare$$

4.3.2 Distance between a U-statistic and a normal random variable

Before we state and prove our main theorem of this section, we state and prove some useful auxiliary results. Lemma 4.3.8, is an application of the next proposition from Sharakhmetov (2004). This proposition provides an upper bound for the distance between the U-statistic and the first term of the decomposition in (1.5.1) under the mean metric κ_2.

Proposition 4.3.7. *(Sharakhmetov (2004)). Let Y be a standard normal random variable and $\xi_1, ..., \xi_n$ be independent identically distributed random variables such that $E(\xi_1) = 0$, $E(\xi_1^2) = \sigma^2$, and $E\left|\xi_1\right|^p < \infty$ for some $p \geq 2$. Then*

$$\kappa_s\left(\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n}\xi_i, Y\right) \leq cn^{-\frac{(p_0-2)}{2}}\frac{E\left|\xi_1\right|^{p_0}}{\sigma^{p_0}},$$

where $1 \leq s \leq p$, $p > 2$, $p_0 = \min(3, p)$ and the constant $c > 0$ depends only on s and p. Moreover,

$$\kappa_s\left(\frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n}\xi_i, Y\right) \to 0 \quad as \quad n \to \infty.$$

Lemma 4.3.8. *Let $\{X_1, X_2, ..., X_n\}$ be a collection of i.i.d random variables and Y be a standard normal random variable and $0 < \delta \leq 1$. If $E\left|h^{(1)}\left(X_1\right)\right|^{2+\delta} < \infty$, then*

$$\kappa_2\left(\frac{1}{n^{\frac{1}{2}}\sigma_1}\sum_{i=1}^{n}h^{(1)}\left(X_i\right), Y\right) \leq cn^{-\frac{\delta}{2}}\frac{E\left|h^{(1)}\left(X_1\right)\right|^{2+\delta}}{\sigma_1^{2+\delta}},$$

where c depends only on δ. Moreover,

$$\kappa_2\left(\frac{1}{n^{\frac{1}{2}}\sigma_1}\sum_{i=1}^{n}h^{(1)}\left(X_i\right), Y\right) \to 0 \quad as \quad n \to \infty.$$

Proof. Apply Proposition 4.3.7. $\blacksquare$

Lemma 4.3.10 that follows, is an application of Proposition 4.3.9 and provides a moment bound for the sum that appears in the second term of the decomposition in (1.5.1).

Proposition 4.3.9. *(Korolyuk and Borovskikh (1989), p. 72). Let $\{X_1, X_2, ..., X_n\}$ be a collection of independent identically distributed random variables. Let U_n be the U-statistic of dimension two based on the kernel h and the given observations. Assume that $E(h) = 0$ and $E \mid h \mid^p < \infty$ for some $1 \leq p \leq 2$. Then*

$$E \mid U_n \mid^p \leq \alpha_p^3 \binom{n}{2}^{1-p} E \left| h^{(2)}(X_1, X_2) \right|^p ,$$

where $\alpha_p \leq 2^{2-p}$.

Let us now denote with U_n^* the U-statistic of dimension 2 based on the degenerate kernel $h^{(2)}$. Then the Hoeffding decomposition for U_n^* is given by

$$U_n^* = \frac{2}{n} \sum_{i=1}^{n} \tilde{h}^{(1)} (X_i) + \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \tilde{h}^{(2)} (X_i, X_j) ,$$

where

$$\tilde{h}^{(1)} (x_i) = E \left(h^{(2)} (X_i, X_j) \,|X_i = x_i \right) ,$$

and

$$\tilde{h}^{(2)} (x_i, x_j) = h^{(2)} (x_i, x_j) - \tilde{h}^{(1)} (x_i) - \tilde{h}^{(1)} (x_j) .$$

Observe that

$$\begin{aligned} \tilde{h}^{(1)} (X_i) &= E \left[h^{(2)} (X_i, X_j) \,|X_i \right] \\ &= E \left[h (X_i, X_j) - h_1 (X_i) - h_1 (X_j) \,|X_i \right] \\ &= E \left[h (X_i, X_j) \,|X_i \right] - E \left[h_1 (X_i) \,|X_i \right] - E \left[h_1 (X_j) \,|X_i \right] \\ &= h_1 (X_i) - h_1 (X_i) - E \left[h_1 (X_j) \right] \\ &= 0. \end{aligned}$$

Similarly $\tilde{h}^{(1)} (X_j) = 0$. Then

$$\tilde{h}^{(2)} (X_i, X_j) = h^{(2)} (X_i, X_j) - \tilde{h}^{(1)} (X_i) - \tilde{h}^{(1)} (X_j) = h^{(2)} (X_i, X_j) .$$

Lemma 4.3.10. *Let $\{X_1, X_2, ..., X_n\}$ be a collection of independent identically distributed random variables. Let U_n^* be the U-statistic of dimension 2 based on the de-*

generate kernel $h^{(2)}$ *and the given observations. Assume that* $E \mid h^{(2)}(X_1, X_2) \mid^p < \infty$ *for some* $1 \leq p \leq 2$. *Then*

$$E\left|\sum_{1\leq i<j\leq n} h^{(2)}(X_i, X_j)\right|^p \leq n(n-1)2^{5-3p} E\left|h^{(2)}(X_1, X_2)\right|^p.$$

Proof. Applying Proposition 4.3.9 and the previous observation that $\tilde{h}^{(2)}(X_i, X_j) = h^{(2)}(X_i, X_j)$, we have that

$$E \mid U_n^* \mid^p \leq \binom{n}{2}^{1-p} 2^{3(2-p)} E\left|\tilde{h}^{(2)}(X_1, X_2)\right|^p$$

$$\Longleftrightarrow E\left|\binom{n}{2} U_n^*\right|^p \leq \binom{n}{2} 2^{3(2-p)} E\left|\tilde{h}^{(2)}(X_1, X_2)\right|^p$$

$$\Longleftrightarrow E\left|\sum_{1\leq i<j\leq n} h^{(2)}(X_i, X_j)\right|^p \leq n(n-1)2^{5-3p} E\left|h^{(2)}(X_1, X_2)\right|^p. \quad \blacksquare$$

The following Lemma 4.3.11 provides the distance between the U-statistic and the first term of the decomposition in (1.5.1) under the κ_1 metric.

Lemma 4.3.11. *Let* $\{X_1, X_2, ..., X_n\}$ *be a collection of i.i.d. random variables. Assume* $E \mid h^{(2)}(X_1, X_2) \mid^{2+\delta} < \infty$ *for* $0 < \delta \leq 1$. *Then*

$$\kappa_1\left(\frac{n^{\frac{1}{2}} U_n}{2\sigma_1}, \frac{1}{n^{\frac{1}{2}}\sigma_1}\sum_{i=1}^{n} h^{(1)}(X_i)\right) \leq n^{-\frac{\delta}{2}} \frac{2^{\frac{3-5\delta}{4}}}{\sigma_1}\left[E\left|h^{(2)}(X_1, X_2)\right|^{\frac{4}{3-\delta}}\right]^{\frac{3-\delta}{4}}.$$

Proof. Applying Proposition 4.2.3 we have that

$$\begin{aligned}
\kappa_1\left(\frac{n^{\frac{1}{2}} U_n}{2\sigma_1}, \frac{1}{n^{\frac{1}{2}}\sigma_1}\sum_{i=1}^{n} h^{(1)}(X_i)\right) &\leq \tau_1\left(\frac{n^{\frac{1}{2}} U_n}{2\sigma_1}, \frac{1}{n^{\frac{1}{2}}\sigma_1}\sum_{i=1}^{n} h^{(1)}(X_i)\right) \\
&= E\left|\frac{n^{\frac{1}{2}} U_n}{2\sigma_1} - \frac{1}{n^{\frac{1}{2}}\sigma_1}\sum_{i=1}^{n} h^{(1)}(X_i)\right| \\
&= E\left|\frac{1}{n^{\frac{1}{2}}(n-1)\sigma_1}\sum_{1\leq i<j\leq n} h^{(2)}(X_i, X_j)\right| \\
&= \frac{1}{n^{\frac{1}{2}}(n-1)\sigma_1} E\left|\sum_{1\leq i<j\leq n} h^{(2)}(X_i, X_j)\right|.
\end{aligned}$$

Furthermore, combining Proposition 4.2.1 and Lemma 4.3.10 we have that

$$\kappa_1\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_1},\frac{1}{n^{\frac{1}{2}}\sigma_1}\sum_{i=1}^{n}h^{(1)}\left(X_i\right)\right)\le\frac{1}{n^{\frac{1}{2}}(n-1)\sigma_1}\left[E\left|\sum_{1\le i<j\le n}h^{(2)}\left(X_i,X_j\right)\right|^{\frac{4}{3-\delta}}\right]^{\frac{3-\delta}{4}}$$

$$\le\frac{1}{n^{\frac{1}{2}}(n-1)\sigma_1}\left[n(n-1)2^{(5-3\frac{4}{3-\delta})}E\left|h^{(2)}\left(X_1,X_2\right)\right|^{\frac{4}{3-\delta}}\right]^{\frac{3-\delta}{4}}$$

$$\le n^{\frac{1-\delta}{4}}(n-1)^{-\frac{1+\delta}{4}}\frac{2^{\frac{3-5\delta}{4}}}{\sigma_1}\left[E\left|h^{(2)}\left(X_1,X_2\right)\right|^{\frac{4}{3-\delta}}\right]^{\frac{3-\delta}{4}}$$

$$\le n^{-\frac{\delta}{2}}\frac{2^{\frac{3-5\delta}{4}}}{\sigma_1}\left[E\left|h^{(2)}\left(X_1,X_2\right)\right|^{\frac{4}{3-\delta}}\right]^{\frac{3-\delta}{4}}. \quad\blacksquare$$

Now we are ready to prove the main result of this section.

Theorem 4.3.12. *Let $\{X_1,X_2,...,X_n\}$ be a collection of independent identically distributed random variables. Let U_n be the U-statistic of dimension two based on the kernel h and the given observations. Assume that $E(h)=0$, $E\mid h\mid^{2+\delta}<\infty$, $0<\delta\le 1$ and $0<\varepsilon<\frac{\delta}{2}$. Then*

$$\zeta_2\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_1},Y\right)\le n^{\varepsilon-\frac{\delta}{2}}\frac{2^{\frac{3-5\delta}{4}}}{\sigma_1}\left[E\left|h^{(2)}\left(X_1,X_2\right)\right|^{\frac{4}{3-\delta}}\right]^{\frac{3-\delta}{4}}+O\left(n^{-\varepsilon\delta}\right)$$

$$+n^{-\varepsilon\delta}\frac{c_1(\delta)}{\sigma_1^{2+\delta}}E\left|h^{(1)}\left(X_1\right)\right|^{2+\delta}+c_2(\delta)n^{-\frac{\delta}{2}}\frac{E\left|h^{(1)}\left(X_1\right)\right|^{2+\delta}}{\sigma_1^{2+\delta}},$$

where $c_1(\delta)$, $c_2(\delta)$ are positive constants depending only on δ and Y is a standard normal random variable.

Proof. Using the triangular inequality and Proposition 4.3.1 we have that

$$\zeta_2\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_1},Y\right)\le\zeta_2\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_1},\frac{1}{n^{\frac{1}{2}}\sigma_1}\sum_{i=1}^{n}h^{(1)}\left(X_i\right)\right)+\zeta_2\left(\frac{1}{n^{\frac{1}{2}}\sigma_1}\sum_{i=1}^{n}h^{(1)}\left(X_i\right),Y\right)$$

$$\le\frac{1}{2}\kappa_2\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_1},\frac{1}{n^{\frac{1}{2}}\sigma_1}\sum_{i=1}^{n}h^{(1)}\left(X_i\right)\right)+\frac{1}{2}\kappa_2\left(\frac{1}{n^{\frac{1}{2}}\sigma_1}\sum_{i=1}^{n}h^{(1)}\left(X_i\right),Y\right),$$

and furthermore applying Proposition 4.3.2 with $N = n^{\varepsilon}$, where $0 < \varepsilon < \frac{\delta}{2}$, we have that

$$\zeta_2\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_1}, Y\right) \leq n^{\varepsilon}\kappa_1\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_1}, \frac{1}{n^{\frac{1}{2}}\sigma_1}\sum_{i=1}^{n} h^{(1)}(X_i)\right)$$

$$+ n^{-\varepsilon\delta}\left[E\left|\frac{n^{\frac{1}{2}}U_n}{2\sigma_1}\right|^{2+\delta} + E\left|\frac{1}{n^{\frac{1}{2}}\sigma_1}\sum_{i=1}^{n} h^{(1)}(X_i)\right|^{2+\delta}\right]$$

$$+ \frac{1}{2}\kappa_2\left(\frac{1}{n^{\frac{1}{2}}\sigma_1}\sum_{i=1}^{n} h^{(1)}(X_i), Y\right).$$

Finally combining Propositions 4.3.4, 2.1.2 and Lemmas 4.3.8 and 4.3.11 we have that

$$\zeta_2\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_1}, Y\right) \leq n^{\varepsilon-\frac{\delta}{2}}\frac{2^{\frac{3-5\delta}{4}}}{\sigma_1}\left[E\left|h^{(2)}(X_1, X_2)\right|^{\frac{4}{3-\delta}}\right]^{\frac{3-\delta}{4}} + O\left(n^{-\varepsilon\delta}\right)$$

$$+ n^{-\varepsilon\delta}\frac{c_1(\delta)}{\sigma_1^{2+\delta}}E\left|h^{(1)}(X_1)\right|^{2+\delta} + c_2(\delta)n^{-\frac{\delta}{2}}\frac{E\left|h^{(1)}(X_1)\right|^{2+\delta}}{\sigma_1^{2+\delta}},$$

where $c_1(\delta)$, $c_2(\delta)$ are positive constants depending only on δ. ■

Corollary 4.3.13. *Under the assumptions of Theorem 4.4.6,*

$$\zeta_2\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_1}, Y\right) \to 0 \quad as \quad n \to \infty.$$

Proof. Follows easily from Theorem 4.4.6. ■

4.3.3 Distance between a V-statistic and a normal random variable

The corresponding theorem for V-statistics can be proved similarly.

Lemma 4.3.14. *Let $\{X_1, X_2, ..., X_n\}$ be a collection of independent identically distributed random variables and Y be a standard normal random variable. Furthermore assume that $E \mid h^{(1)}(X_1) \mid^{2+\delta} < \infty$ for $0 < \delta \leq 1$. Then*

$$\kappa_2\left(\frac{n-1}{n^{\frac{3}{2}}\sigma_1}\sum_{i=1}^{n} h^{(1)}(X_i), Y\right) \leq cn^{-\frac{\delta}{2}}\frac{E\left|h^{(1)}(X_1)\right|^{2+\delta}}{\sigma_1^{2+\delta}},$$

where c depends only on δ. Moreover,

$$\kappa_2\left(\frac{n-1}{n^{\frac{3}{2}}\sigma_1}\sum_{i=1}^{n}h^{(1)}\left(X_i\right),Y\right)\to 0 \quad as \quad n\to\infty.$$

Proof. Apply Proposition 4.3.7. ■

Lemma 4.3.15. *Let* $\{X_1, X_2, ..., X_n\}$ *be a collection of independent identically distributed random variables. Assume that* $0 < \delta \leq 1$ *and* $E \mid h^{(2)}\left(X_1, X_2\right) \mid^{2+\delta} < \infty$. *Then*

$$\kappa_1\left(\frac{n^{\frac{1}{2}}V_n}{2\sigma_1},\frac{n-1}{n^{\frac{3}{2}}\sigma_1}\sum_{i=1}^{n}h^{(1)}\left(X_i\right)\right)\leq n^{-\frac{\delta}{2}}\frac{2^{\frac{3-5\delta}{4}}}{\sigma_1}\left[E\left|h^{(2)}\left(X_1,X_2\right)\right|^{\frac{4}{3-\delta}}\right]^{\frac{3-\delta}{4}}.$$

Proof. Applying Proposition 4.2.3 we have that

$$\begin{aligned}\kappa_1\left(\frac{n^{\frac{1}{2}}V_n}{2\sigma_1},\frac{n-1}{n^{\frac{3}{2}}\sigma_1}\sum_{i=1}^{n}h^{(1)}\left(X_i\right)\right) &\leq \tau_1\left(\frac{n^{\frac{1}{2}}V_n}{2\sigma_1},\frac{n-1}{n^{\frac{3}{2}}\sigma_1}\sum_{i=1}^{n}h^{(1)}\left(X_i\right)\right)\\ &= E\left|\frac{n^{\frac{1}{2}}V_n}{2\sigma_1}-\frac{n-1}{n^{\frac{3}{2}}\sigma_1}\sum_{i=1}^{n}h^{(1)}\left(X_i\right)\right|\\ &= E\left|\frac{1}{2n^{\frac{3}{2}}\sigma_1}\sum_{i=1}^{n}h\left(X_i,X_i\right)+\frac{1}{n^{\frac{3}{2}}\sigma_1}\sum_{1\leq i<j\leq n}h^{(2)}\left(X_i,X_j\right)\right|\\ &\leq \frac{E\mid h\left(X_1,X_1\right)\mid}{2n^{\frac{1}{2}}\sigma_1}+\frac{1}{n^{\frac{3}{2}}\sigma_1}E\left|\sum_{1\leq i<j\leq n}h^{(2)}\left(X_i,X_j\right)\right|.\end{aligned}$$

Furthermore, combining Proposition 4.2.1 and Lemma 4.3.10 we have that

$$\begin{aligned}&\kappa_1\left(\frac{n^{\frac{1}{2}}V_n}{2\sigma_1},\frac{n-1}{n^{\frac{3}{2}}\sigma_1}\sum_{i=1}^{n}h^{(1)}\left(X_i\right)\right)\\ &\leq \frac{E\mid h\left(X_1,X_1\right)\mid}{2n^{\frac{1}{2}}\sigma_1}+\frac{1}{n^{\frac{3}{2}}\sigma_1}\left[E\left|\sum_{1\leq i<j\leq n}h^{(2)}\left(X_i,X_j\right)\right|^{\frac{4}{3-\delta}}\right]^{\frac{3-\delta}{4}}\\ &= \frac{E\mid h\left(X_1,X_1\right)\mid}{2n^{\frac{1}{2}}\sigma_1}+\frac{1}{n^{\frac{3}{2}}\sigma_1}\left[n(n-1)2^{5-3\frac{4}{3-\delta}}E\left|h^{(2)}\left(X_1,X_2\right)\right|^{\frac{4}{3-\delta}}\right]^{\frac{3-\delta}{4}}\\ &= \frac{E\mid h\left(X_1,X_1\right)\mid}{2n^{\frac{1}{2}}\sigma_1}+n^{-\frac{\delta}{2}}\frac{2^{\frac{3-5\delta}{4}}}{\sigma_1}\left[E\left|h^{(2)}\left(X_1,X_2\right)\right|^{\frac{4}{3-\delta}}\right]^{\frac{3-\delta}{4}}.\ \blacksquare\end{aligned}$$

Theorem 4.3.16. *Let $\{X_1, X_2, ..., X_n\}$ be a collection of independent identically distributed random variables. Let V_n be the von Mises statistic of dimension two based on the kernel h and the given observations. Assume that $E(h) = 0$, $E \mid h \mid^{2+\delta} < \infty$, $0 < \delta \leq 1$ and $0 < \varepsilon < \frac{\delta}{2}$. Then*

$$\zeta_2\left(\frac{n^{\frac{1}{2}}V_n}{2\sigma_1}, Y\right) \leq n^{\varepsilon-\frac{1}{2}}\frac{E \mid h\left(X_1, X_1\right) \mid}{2\sigma_1} + n^{\varepsilon-\frac{\delta}{2}}\frac{2^{\frac{3-5\delta}{4}}}{\sigma_1}\left[E\left|h^{(2)}\left(X_1, X_2\right)\right|^{\frac{4}{3-\delta}}\right]^{\frac{3-\delta}{4}}$$
$$+ O\left(n^{-\varepsilon\delta}\right) + n^{-\varepsilon\delta}\frac{c_1(\delta)}{\sigma_1^{2+\delta}}E\left|h^{(1)}\left(X_1\right)\right|^{2+\delta} + c_2(\delta)n^{-\frac{\delta}{2}}\frac{E\left|h^{(1)}\left(X_1\right)\right|^{2+\delta}}{\sigma_1^{2+\delta}},$$

where $c_1(\delta)$, $c_2(\delta)$ are positive constants depending only on δ and Y is a standard normal random variable.

Proof. Using the triangular inequality and Proposition 4.3.1 we have that

$$\zeta_2\left(\frac{n^{\frac{1}{2}}V_n}{2\sigma_1}, Y\right) \leq \zeta_2\left(\frac{n^{\frac{1}{2}}V_n}{2\sigma_1}, \frac{n-1}{n^{\frac{3}{2}}\sigma_1}\sum_{i=1}^{n}h^{(1)}\left(X_i\right)\right) + \zeta_2\left(\frac{n-1}{n^{\frac{3}{2}}\sigma_1}\sum_{i=1}^{n}h^{(1)}\left(X_i\right), Y\right)$$
$$\leq \frac{1}{2}\kappa_2\left(\frac{n^{\frac{1}{2}}V_n}{2\sigma_1}, \frac{n-1}{n^{\frac{3}{2}}\sigma_1}\sum_{i=1}^{n}h^{(1)}\left(X_i\right)\right) + \frac{1}{2}\kappa_2\left(\frac{n-1}{n^{\frac{3}{2}}\sigma_1}\sum_{i=1}^{n}h^{(1)}\left(X_i\right), Y\right),$$

and furthermore applying Proposition 4.3.2 with $N = n^{\varepsilon}$, where $0 < \varepsilon < \frac{\delta}{2}$, we have that

$$\zeta_2\left(\frac{n^{\frac{1}{2}}V_n}{2\sigma_1}, Y\right) \leq n^{\varepsilon}\kappa_1\left(\frac{n^{\frac{1}{2}}V_n}{2\sigma_1}, \frac{n-1}{n^{\frac{3}{2}}\sigma_1}\sum_{i=1}^{n}h^{(1)}\left(X_i\right)\right)$$
$$+ n^{-\varepsilon\delta}\left[E\left|\frac{n^{\frac{1}{2}}V_n}{2\sigma_1}\right|^{2+\delta} + E\left|\frac{n-1}{n^{\frac{3}{2}}\sigma_1}\sum_{i=1}^{n}h^{(1)}\left(X_i\right)\right|^{2+\delta}\right]$$
$$+ \frac{1}{2}\kappa_2\left(\frac{n-1}{n^{\frac{3}{2}}\sigma_1}\sum_{i=1}^{n}h^{(1)}\left(X_i\right), Y\right).$$

Finally combining Proposition 2.1.2 and Lemmas 4.3.6, 4.3.14 and 4.3.15 we have that

$$\zeta_2\left(\frac{n^{\frac{1}{2}}V_n}{2\sigma_1}, Y\right) \leq n^{\varepsilon-\frac{1}{2}}\frac{E \mid h\left(X_1, X_1\right) \mid}{2\sigma_1} + n^{\varepsilon-\frac{\delta}{2}}\frac{2^{\frac{3-5\delta}{4}}}{\sigma_1}\left[E\left|h^{(2)}\left(X_1, X_2\right)\right|^{\frac{4}{3-\delta}}\right]^{\frac{3-\delta}{4}}$$
$$+ O\left(n^{-\varepsilon\delta}\right) + n^{-\varepsilon\delta}\frac{c_1(\delta)}{\sigma_1^{2+\delta}}E\left|h^{(1)}\left(X_1\right)\right|^{2+\delta} + c_2(\delta)n^{-\frac{\delta}{2}}\frac{E\left|h^{(1)}\left(X_1\right)\right|^{2+\delta}}{\sigma_1^{2+\delta}},$$

where $c_1(\delta)$, $c_2(\delta)$ are positive constants depending only on δ. ■

Corollary 4.3.17. *Under the assumptions of Theorem 4.3.16,*

$$\zeta_2\left(\frac{n^{\frac{1}{2}}V_n}{2\sigma_1}, Y\right) \to 0 \quad as \quad n \to \infty.$$

Proof. Follows easily from Theorem 4.3.16. ■

4.4 Distance between a U-statistic based on negatively associated random variables and a normal random variable

Now let us turn our attention to collections of random variables that are dependent. In this section we study the distance of a U-statistic based on a collection of identically distributed negatively associated random variables and a normal random variable using again Zolotarev's ideal metric. In bibliography the asymptotic normality of U-statistics based on negatively associated random variables under the assumption of degenerate or non-degenerate kernel is investigated only in Huang and Zhang (2006).

4.4.1 Distance between a U-statistic and a normal random variable

Proposition 4.4.1. *(Christofides and Vaggelatou (2004)).*
If $X_1, X_2, ..., X_n$ is a collection of negatively associated r.v.'s and $X_1^*, X_2^*, ..., X_n^*$ are independent copies of $X_1, X_2, ..., X_n$. Then

$$(X_1, X_2, ..., X_n) \preccurlyeq_{sm} (X_1^*, X_2^*, ..., X_n^*).$$

Proposition 4.4.2. *(Boutsikas and Vaggelatou (2002)). Let $X_1, X_2, \ldots$ be a strictly stationary sequence of negatively associated random variables such that $E(X_1) = 0$ and $0 < E(X_1^2) < \infty$. If $\sigma^2 := E(X_1^2) + 2\sum_{j=2}^{\infty} E(X_1 X_j) > 0$, then, for $n = mk$,*

$$\zeta_2\left(\frac{\sum_i^n X_i}{\sigma\sqrt{n}}, Y\right) \leq \frac{a_k}{k}\left[E\left(X_1^2\right) - \sigma^2\right] - 2\left(1 - \frac{a_k}{k}\right) u\left(a_k\right) + c\frac{\rho_k + 1}{m^{\frac{1}{2}}}$$

for some constant $c > 0$, where Y is a standard normal random variable, $\{a_k\}$ is any sequence of positive integers such that $a_k \leq k$, $a_k \to \infty$, $\frac{a_k}{k} \to 0$ as $k \to \infty$,

$\rho_k := E\left|k^{-\frac{1}{2}}\sum_{i=1}^{k}X_i\right|^3$ *and* $u(i) := \sum_{j=i+1}^{\infty}E\left(X_1X_j\right) \to 0$ *as* $i \to \infty$.

Lemma 4.4.3. *Let* U_n *be a U-statistic of dimension two based on stationary negatively associated random variables and on the nondecreasing kernel* h. *Assume that* $h^{(1)}$ *is degenerate. Assume further that* $E(h^{(1)}) = 0$ *and* $0 < E[h^{(1)}]^2 < \infty$. *If* $\sigma_U^2 = \mathrm{Var}[h_1(X_1)] + 2\sum_{j=1}^{n}\mathrm{cov}[h_1(X_1), h_1(X_{1+j})] > 0$, *then, for* $n = mk$,

$$\zeta_2\left(\frac{n^{-\frac{1}{2}}}{\sigma_U}\sum_{i=1}^{n}h^{(1)}(X_i), Y\right) \leq \frac{a_k}{k}\left[E\left(h^{(1)}(X_1)\right)^2 - \sigma_U^2\right] - 2\left(1-\frac{a_k}{k}\right)u(a_k) + c\frac{\rho_k+1}{m^{\frac{1}{2}}}$$

for some constant $c > 0$, *where* Y *is a standard normal random variable,* $\{a_k\}$ *is any sequence of positive integers such that* $a_k \leq k$, $a_k \to \infty$, $\frac{a_k}{k} \to 0$ *as* $k \to \infty$, $\rho_k := E\left|k^{-\frac{1}{2}}\sum_{i=1}^{k}h^{(1)}(X_i)\right|^3$ *and* $u(i) := \sum_{j=i+1}^{\infty}E\left(h^{(1)}(X_1)h^{(1)}(X_j)\right) \to 0$ *as* $i \to \infty$.

Proof. Since nondecreasing functions of negatively associated random variables are negatively associated, we have that $\{h^{(1)}(X_i),\ i \geq 1\}$ are negatively associated. Applying now Lemma 4.4.2, we have the result. ■

Remark 4.4.4. *Note that if* $E\left|\sum_{i=1}^{\infty}h^{(1)}(X_i)\right|^3 < \infty$, *then*

$$\zeta_2\left(n^{-\frac{1}{2}}\sum_{i=1}^{n}h^{(1)}(X_i), Y\right) = o(1), \quad \text{as } n \to \infty.$$

Lemma 4.4.5. *Let* $\{X_1, X_2, ..., X_n\}$ *be a collection of identically distributed stationary negatively associated random variables. Let* U_n *be the U-statistic of dimension two based on the kernel* h *and the given observations and assume that the degenerate kernel* $h^{(2)}$ *is a componentwise nondecreasing function. Let now* $\{X_1^*, X_2^*, ..., X_n^*\}$ *be independent copies of* $\{X_1, X_2, ..., X_n\}$ *and assume that* $E(h(X_1^*, X_2^*)) = 0$. *Furthermore assume that the function* $g(x,y,z) = h^{(2)}(x,y)h^{(2)}(x,z)$ *is supermodular. If* $E\left|h^{(2)}(X_i, X_j)\right|^2 \leq c < \infty$ *for all* $1 \leq i < j \leq n$, *then*

$$\kappa_1\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_U}, \frac{1}{n^{\frac{1}{2}}\sigma_U}\sum_{i=1}^{n}h^{(1)}(X_i)\right) \leq \frac{(n-1)^{-\frac{1}{2}}}{\sigma_U}\left[\frac{1}{n(n-1)}\sum_{1\leq i<j\leq n}E\left[h^{(2)}(X_i, X_j)\right]^2\right]^{\frac{1}{2}},$$

where $\sigma_U^2 = \sigma_1^2 + 2\sum_{j=1}^{\infty}\sigma_{1j}^2 > 0$ *with* $\sigma_1^2 = \mathrm{Var}[h_1(X_1)]$ *and* $\sigma_{1j}^2 = \mathrm{cov}(h_1(X_1), h_1(X_{1+j}))$.

Proof. By applying Propositions 4.2.3 and 4.2.1 we may write

$$\begin{aligned}\kappa_1\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_U}, \frac{1}{n^{\frac{1}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}\left(X_i\right)\right) &\le \tau_1\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_U}, \frac{1}{n^{\frac{1}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}\left(X_i\right)\right)\\
&= E\left|\frac{n^{\frac{1}{2}}U_n}{2\sigma_U} - \frac{1}{n^{\frac{1}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}\left(X_i\right)\right|\\
&= E\left|\frac{1}{n^{\frac{1}{2}}(n-1)\sigma_U}\sum_{1\le i<j\le n} h^{(2)}\left(X_i, X_j\right)\right|\\
&= \frac{1}{n^{\frac{1}{2}}(n-1)\sigma_U}E\left|\sum_{1\le i<j\le n} h^{(2)}\left(X_i, X_j\right)\right|.\\
&\le \frac{1}{n^{\frac{1}{2}}(n-1)\sigma_U}\left[E\left|\sum_{1\le i<j\le n} h^{(2)}\left(X_i, X_j\right)\right|^2\right]^{\frac{1}{2}}.\end{aligned} \tag{4.4.1}$$

The expansion of the sum on the right-hand side of (4.4.1) gives

$$\begin{aligned}E\left[\sum_{1\le i<j\le n} h^{(2)}\left(X_i, X_j\right)\right]^2 &= \sum_{1\le i<j\le n} E\left[h^{(2)}\left(X_i, X_j\right)\right]^2\\
&+ 2\sum_{\substack{1\le i<j\le n\\(i,j)\ne(k,l)}}\sum_{1\le k<l\le n} E\left[h^{(2)}\left(X_i, X_j\right)h^{(2)}\left(X_k, X_l\right)\right].\end{aligned} \tag{4.4.2}$$

Let

$$B = \sum_{\substack{1\le i<j\le n\\(i,j)\ne(k,l)}}\sum_{1\le k<l\le n} E\left[h^{(2)}\left(X_i, X_j\right)h^{(2)}\left(X_k, X_l\right)\right].$$

It can be easily verified that the terms in expression B are of two kinds. In the first case indices $\{i, j, k, l\}$ are all different. Since the kernel $h^{(2)}$ is a componentwise nondecreasing function from the definition of negatively associated random variables it follows that

$$E\left[h^{(2)}\left(X_i, X_j\right)h^{(2)}\left(X_k, X_l\right)\right] \le 0. \tag{4.4.3}$$

Furthermore in the second case when indices $\{i, j, k, l\}$ are not all different, necessarily $\{i = k \text{ and } j \neq l\}$ or $\{i \neq k \text{ and } j = l\}$. Without loss of generality assume that $\{i = k \text{ and } j \neq l\}$. We will prove that in that case it derives also that

$$E\left[h^{(2)}\left(X_i, X_j\right) h^{(2)}\left(X_i, X_l\right)\right] \leq E\left[h^{(2)}\left(X_i^*, X_j^*\right) h^{(2)}\left(X_i^*, X_l^*\right)\right] = 0, \tag{4.4.4}$$

where X_i^*, X_j^*, X_l^* are independent copies of X_i, X_j, X_l.

Since g is supermodular from Proposition 4.4.1 we have that

$$E(g\left(X_i, X_j, X_l\right)) \leq E(g\left(X_i^*, X_j^*, X_l^*\right)).$$

Now working on $E(g\left(X_i^*, X_j^*, X_l^*\right))$, it follows that

$$\begin{aligned}
E(g\left(X_i^*, X_j^*, X_l^*\right)) &= E\left[h^{(2)}\left(X_i^*, X_j^*\right) h^{(2)}\left(X_i^*, X_l^*\right)\right] \\
&= E\left[E[h^{(2)}\left(X_i^*, X_j^*\right) h^{(2)}\left(X_i^*, X_l^*\right) | X_i^*]\right] \\
&= \int E\left[h^{(2)}(x, X_j^*) h^{(2)}(x, X_l^*)\right] dF(x) \\
&= \int E\left[h^{(2)}(x, X_j^*)\right] E\left[h^{(2)}(x, X_l^*)\right] dF(x) \qquad (4.4.5) \\
&= 0.
\end{aligned}$$

where (4.4.5) follows by independence and the last equality follows from the fact that

$$E\left[h^{(2)}(x, X_j^*)\right] = E\left[h(x, X_j^*) - h_1(x) - h_1(X_j^*) + \theta\right] = h_1(x) - h_1(x) - \theta + \theta = 0.$$

Combining the results in (4.4.1), (4.4.2), (4.4.3) and (4.4.4) we arrive at

$$\begin{aligned}
\kappa_1\left(\frac{n^{\frac{1}{2}} U_n}{2\sigma_U}, \frac{1}{n^{\frac{1}{2}}\sigma_U} \sum_{i=1}^{n} h^{(1)}\left(X_i\right)\right) &\leq \frac{1}{n^{\frac{1}{2}}(n-1)\sigma_U}\left[\sum_{1 \leq i < j \leq n} E\left[h^{(2)}\left(X_i, X_j\right)\right]^2\right]^{\frac{1}{2}} \\
&= \frac{(n-1)^{-\frac{1}{2}}}{\sigma_U}\left[\frac{1}{n(n-1)} \sum_{1 \leq i < j \leq n} E\left[h^{(2)}\left(X_i, X_j\right)\right]^2\right]^{\frac{1}{2}}. \quad \blacksquare
\end{aligned}$$

Now we are ready to prove the main result of this section.

Theorem 4.4.6. *Let $\{X_1, X_2, ..., X_n\}$ be a collection of identically distributed stationary negatively associated random variables. Let U_n be the U-statistic of dimension two based on the kernel h and the given observations. Let now $\{X_1^*, X_2^*, ..., X_n^*\}$ be independent copies of $\{X_1, X_2, ..., X_n\}$ and assume that $E(h(X_1^*, X_2^*)) = 0$. Further assume that h is componentwise nondecreasing function, $E|h|^{2+\delta} < \infty$, $0 < \delta \leq 1$ and $0 < \varepsilon < \frac{1}{2}$ with $\sigma_U^2 = \sigma_1^2 + 2\sum_{j=1}^{\infty} \sigma_{1j}^2 > 0$ where $\sigma_1^2 = \mathrm{Var}[h_1(X_1)]$ and $\sigma_{1j}^2 = \mathrm{cov}(h_1(X_1), h_1(X_{1+j}))$. Also assume that the degenerate kernels $h^{(1)}$, $h^{(2)}$ are nondecreasing functions and the function $g(x,y,z) = h^{(2)}(x,y)\, h^{(2)}(x,z)$ is supermodular. Then*

$$\zeta_2\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_U}, Y\right) \leq \frac{(n-1)^{\varepsilon-\frac{1}{2}}}{\sigma_U}\left[\frac{1}{n(n-1)}\sum_{1\leq i<j\leq n} E\left[h^{(2)}(X_i, X_j)\right]^2\right]^{\frac{1}{2}}$$

$$+ n^{-\varepsilon\delta}\frac{C_{2+\delta}}{\sigma_U^{2+\delta}}\left(\frac{3}{2}\right)^{1+\delta}\frac{1}{n(n-1)}\sum_{1\leq i<j\leq n} E\left|h(X_i, X_j)\right|^{2+\delta}$$

$$+ n^{-\varepsilon\delta}\frac{C_{2+\delta}}{\sigma_U^{2+\delta}}\frac{1}{n}\sum_{i=1}^{n} E|h^{(1)}(X_i)|^{2+\delta}$$

$$+ \frac{a_k}{k}\left[E\left(h^{(1)}(X_1)\right)^2 - \sigma_U^2\right] - 2\left(1-\frac{a_k}{k}\right)u(a_k) + c\frac{\rho_k+1}{m^{\frac{1}{2}}},$$

for some constant $c > 0$, where Y is a standard normal random variable, $\{a_k\}$ is any sequence of positive integers such that $a_k \leq k$, $a_k \to \infty$, $\frac{a_k}{k} \to 0$ as $k \to \infty$, $\rho_k := E\left|k^{-\frac{1}{2}}\sum_{i=1}^{k} h^{(1)}(X_i)\right|^3$ and $u(i) := \sum_{j=i+1}^{\infty} E\left(h^{(1)}(X_1)\, h^{(1)}(X_j)\right) \to 0$ as $i \to \infty$, $C_{2+\delta}$ is a positive constant depending only on δ.

Proof. The triangular inequality and Proposition 4.3.1 together give,

$$\begin{aligned}
\zeta_2\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_U}, Y\right) &\leq \zeta_2\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_U}, \frac{1}{n^{\frac{1}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}(X_i)\right) + \zeta_2\left(\frac{1}{n^{\frac{1}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}(X_i), Y\right) \\
&\leq \frac{1}{2}\kappa_2\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_U}, \frac{1}{n^{\frac{1}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}(X_i)\right) + \zeta_2\left(\frac{1}{n^{\frac{1}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}(X_i), Y\right),
\end{aligned}$$

and furthermore applying Proposition 4.3.2 with $N = n^{\varepsilon}$, where $0 < \varepsilon < \frac{1}{2}$, we have that

$$\zeta_2\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_U}, Y\right) \leq n^{\varepsilon}\kappa_1\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_U}, \frac{1}{n^{\frac{1}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}(X_i)\right)$$

$$+ n^{-\varepsilon\delta}\left[E\left|\frac{n^{\frac{1}{2}}U_n}{2\sigma_U}\right|^{2+\delta} + E\left|\frac{1}{n^{\frac{1}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}\left(X_i\right)\right|^{2+\delta}\right]$$
$$+ \zeta_2\left(\frac{1}{n^{\frac{1}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}\left(X_i\right), Y\right).$$

Finally combining Proposition 1.2.9 and Lemmas 2.1.3, 4.4.3 and 4.4.5 we arrive at

$$\zeta_2\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_U}, Y\right) \leq \frac{(n-1)^{\varepsilon-\frac{1}{2}}}{\sigma_U}\left[\frac{1}{n(n-1)}\sum_{1\leq i<j\leq n} E\left[h^{(2)}\left(X_i, X_j\right)\right]^2\right]^{\frac{1}{2}}$$
$$+ n^{-\varepsilon\delta}\frac{C_{2+\delta}}{\sigma_U^{2+\delta}}\left(\frac{3}{2}\right)^{1+\delta}\frac{1}{n\left(n-1\right)}\sum_{1\leq i<j\leq n} E\left|h\left(X_i, X_j\right)\right|^{2+\delta}$$
$$+ n^{-\varepsilon\delta}\frac{C_{2+\delta}}{\sigma_U^{2+\delta}}\frac{1}{n}\sum_{i=1}^{n} E|h^{(1)}\left(X_i\right)|^{2+\delta}$$
$$+ \frac{a_k}{k}\left[E\left(h^{(1)}\left(X_1\right)\right)^2 - \sigma_U^2\right] - 2\left(1-\frac{a_k}{k}\right)u\left(a_k\right) + c\frac{\rho_k+1}{m^{\frac{1}{2}}},$$

for some constant $c > 0$, where Y is a standard normal random variable, $\{a_k\}$ is any sequence of positive integers such that $a_k \leq k$, $a_k \to \infty$, $\frac{a_k}{k} \to 0$ as $k \to \infty$, $\rho_k := E\left|k^{-\frac{1}{2}}\sum_{i=1}^{k} h^{(1)}\left(X_i\right)\right|^3$ and $u(i) := \sum_{j=i+1}^{\infty} E\left(h^{(1)}\left(X_1\right)h^{(1)}\left(X_j\right)\right) \to 0$ as $i \to \infty$, $C_{2+\delta}$ is a positive constant depending only on δ. ■

Corollary 4.4.7. *Under the assumptions of Theorem 4.4.6 and* $E\left|\sum_{i=1}^{\infty} h^{(1)}\left(X_i\right)\right|^3 < \infty$*, then*

$$\zeta_2\left(\frac{n^{\frac{1}{2}}U_n}{2\sigma_U}, Y\right) \to 0 \quad as \quad n \to \infty.$$

Proof. Follows easily from Theorem 4.4.6. ■

Remark 4.4.8. *Note that the previous result, Theorem 4.4.6, is still valid for U-statistics based on a kernel h of bounded variation, with degenerate kernels $h^{(1)}, h^{(2)}$ of bounded variation also, because of the fact that a function of bounded variation can be written as the difference of two nondecreasing functions.*

Theorem 4.4.6 has an assumption related to the concept of supermodularity. Here, we should give some trivial examples of supermodular functions. The functions $f(x,y) = x+y$ and $f(x,y) = xy$ are supermodular on R^2. For various properties and applications concerning supermodular functions, see Topkis (1998). Furthermore, Example 4.4.9 below presents a U-statistic which satisfies all the kernel assumptions made in Theorem 4.4.6.

Example 4.4.9. *Let $\{X_1, X_2, ..., X_n\}$ be a collection of identically distributed stationary negatively associated random variables. Let U_n be the U-statistic of dimension two based on the kernel h and the given observations. Assume that $E(X_1) = 0$. Consider the estimation of $\theta = \sigma^2$. U_n is based on the kernels*

$$\begin{aligned} h(x_1, x_2) &= \frac{1}{2}(x_1^2 + x_2^2), \\ h^{(1)}(x_1) &= \frac{1}{2}(x_1^2 + \sigma^2), \\ h^{(1)}(x_2) &= \frac{1}{2}(x_2^2 + \sigma^2), \\ h^{(2)}(x_1, x_2) &= \frac{1}{2}(x_1^2 + x_2^2) - \frac{1}{2}(x_1^2 + \sigma^2) - \frac{1}{2}(x_2^2 + \sigma^2) + \sigma^2 = 0. \end{aligned}$$

One can verify that h, $h^{(1)}$ and $h^{(2)}$ are functions of bounded variation and trivially $g(x, y, z) = h^{(2)}(x, y)\, h^{(2)}(x, z)$ is supermodular.

4.4.2 Distance between a V-statistic and a normal random variable

Lemma 4.4.10. *Let $\{X_1, X_2, ..., X_n\}$ be a collection of identically distributed stationary negatively associated random variables. Let V_n be the von Mises statistic of dimension two based on the kernel h and the given observations. Let now $\{X_1^*, X_2^*, ..., X_n^*\}$ be independent copies of $\{X_1, X_2, ..., X_n\}$ and assume that $E(h(X_1^*, X_2^*)) = 0$. Furthermore assume that the degenerate kernel $h^{(2)}$ is a componentwise nondecreasing function and the function $g(x, y, z) = h^{(2)}(x, y)\, h^{(2)}(x, z)$ is supermodular. If $E\left|h^{(2)}(X_i, X_j)\right|^2 \leq c < \infty$ for all $1 \leq i < j \leq n$, then*

$$\begin{aligned} &\kappa_1\left(\frac{n^{\frac{1}{2}} V_n}{2\sigma_U}, \frac{n-1}{n^{\frac{3}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}(X_i)\right) \\ &< \frac{E \mid h(X_1, X_1) \mid}{2n^{\frac{1}{2}}\sigma_U} + \frac{n^{-\frac{1}{2}}}{\sigma_U}\left[\frac{1}{n(n-1)}\sum_{1\leq i<j\leq n} E\left|h^{(2)}(X_i, X_j)\right|^2\right]^{\frac{1}{2}}. \end{aligned}$$

Proof. Applying Proposition 4.2.3 we have that

$$\kappa_1\left(\frac{n^{\frac{1}{2}} V_n}{2\sigma_U}, \frac{n-1}{n^{\frac{3}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}(X_i)\right) \leq \tau_1\left(\frac{n^{\frac{1}{2}} V_n}{2\sigma_U}, \frac{n-1}{n^{\frac{3}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}(X_i)\right)$$

$$= E\left|\frac{n^{\frac{1}{2}}V_n}{2\sigma_U} - \frac{n-1}{n^{\frac{3}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}(X_i)\right|$$

$$= E\left|\frac{1}{2n^{\frac{3}{2}}\sigma_U}\sum_{i=1}^{n} h(X_i, X_i) + \frac{1}{n^{\frac{3}{2}}\sigma_U}\sum_{1\le i<j\le n} h^{(2)}(X_i, X_j)\right|$$

$$\le \frac{E\mid h(X_1, X_1)\mid}{2n^{\frac{1}{2}}\sigma_U} + \frac{1}{n^{\frac{3}{2}}\sigma_U}E\left|\sum_{1\le i<j\le n} h^{(2)}(X_i, X_j)\right|.$$

Furthermore, combining Proposition 4.2.1 and an inequality proved in Lemma 4.4.5 we have that

$$\kappa_1\left(\frac{n^{\frac{1}{2}}V_n}{2\sigma_U}, \frac{n-1}{n^{\frac{3}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}(X_i)\right) \le \frac{E\mid h(X_1, X_1)\mid}{2n^{\frac{1}{2}}\sigma_U} + \frac{1}{n^{\frac{3}{2}}\sigma_U}\left[E\left|\sum_{1\le i<j\le n} h^{(2)}(X_i, X_j)\right|^2\right]^{\frac{1}{2}}$$

$$< \frac{E\mid h(X_1, X_1)\mid}{2n^{\frac{1}{2}}\sigma_U} + \frac{1}{n^{\frac{3}{2}}\sigma_U}\left[\sum_{1\le i<j\le n} E\left|h^{(2)}(X_i, X_j)\right|^2\right]^{\frac{1}{2}}$$

$$< \frac{E\mid h(X_1, X_1)\mid}{2n^{\frac{1}{2}}\sigma_U} + \frac{n^{-\frac{1}{2}}}{\sigma_U}\left[\frac{1}{n(n-1)}\sum_{1\le i<j\le n} E\left|h^{(2)}(X_i, X_j)\right|^2\right]^{\frac{1}{2}}. \quad \blacksquare$$

Theorem 4.4.11. *Let $\{X_1, X_2, ..., X_n\}$ be a collection of identically distributed stationary negatively associated random variables. Let V_n be the von Mises statistic of dimension two based on the kernel h and the given observations. Let now $\{X_1^*, X_2^*, ..., X_n^*\}$ be independent copies of $\{X_1, X_2, ..., X_n\}$ and assume that $E(h(X_1^*, X_2^*)) = 0$. Further assume that h is componentwise nondecreasing function, $E|h|^{2+\delta} < \infty$, $0 < \delta \le 1$ and $0 < \varepsilon < \frac{1}{2}$ with $\sigma_U^2 = \sigma_1^2 + 2\sum_{j=1}^{\infty}\sigma_{1j}^2 > 0$ where $\sigma_1^2 = \mathrm{Var}[h_1(X_1)]$ and $\sigma_{1j}^2 = \mathrm{cov}(h_1(X_1), h_1(X_{1+j}))$. We also assume that the degenerate kernels $h^{(1)}$, $h^{(2)}$ are nondecreasing functions and the function $g(x,y,z) = h^{(2)}(x,y)\,h^{(2)}(x,z)$ is supermodular. Then*

$$\zeta_2\left(\frac{n^{\frac{1}{2}}V_n}{2\sigma_U}, Y\right) \le n^{\varepsilon-\frac{1}{2}}\frac{E\mid h(X_1, X_1)\mid}{2\sigma_U} + \frac{n^{\varepsilon-\frac{1}{2}}}{\sigma_U}\left[\frac{1}{n(n-1)}\sum_{1\le i<j\le n} E\left[h^{(2)}(X_i, X_j)\right]^2\right]^{\frac{1}{2}}$$

$$+ n^{-\varepsilon\delta}(n-1)^{-\frac{2+\delta}{2}} n^{\frac{2+\delta}{2}} \frac{3^{1+\delta}}{(2\sigma_U)^{2+\delta}} C_{2+\delta} \frac{2}{n(n-1)} \sum_{1\le i<j\le n} E\left|h\left(X_i, X_j\right)\right|^{2+\delta}$$

$$+ n^{-2-\delta(\varepsilon+1)} \frac{C_{2+\delta}}{(2\sigma_U)^{2+\delta}} \frac{1}{n} \sum_{i=1}^{n} E\left|h\left(X_i, X_i\right)\right|^{2+\delta}$$

$$+ n^{-\varepsilon\delta} \left(\frac{n-1}{n}\right)^{2+\delta} \frac{C_{2+\delta}}{\sigma_U^{2+\delta}} \frac{1}{n} \sum_{i=1}^{n} E|h^{(1)}\left(X_i\right)|^{2+\delta}$$

$$+ \frac{a_k}{k}\left[E\left(h^{(1)}\left(X_1\right)\right)^2 - \sigma_U^2\right] - 2\left(1 - \frac{a_k}{k}\right) u\left(a_k\right) + c\frac{\rho_k + 1}{m^{\frac{1}{2}}},$$

for some constant $c > 0$, where Y is a standard normal random variable, $\{a_k\}$ is any sequence of positive integers such that $a_k \le k$, $a_k \to \infty$, $\frac{a_k}{k} \to 0$ as $k \to \infty$, $\rho_k := E\left|k^{-\frac{1}{2}}\sum_{i=1}^{k} h^{(1)}\left(X_i\right)\right|^3$ and $u(i) := \sum_{j=i+1}^{\infty} E\left(h^{(1)}\left(X_1\right) h^{(1)}\left(X_j\right)\right) \to 0$ as $i \to \infty$, $C_{2+\delta}$ is a positive constant depending only on δ.

Proof. Using the triangular inequality and Proposition 4.3.1 we have that

$$\zeta_2\left(\frac{n^{\frac{1}{2}}V_n}{2\sigma_U}, Y\right) \le \zeta_2\left(\frac{n^{\frac{1}{2}}V_n}{2\sigma_U}, \frac{n-1}{n^{\frac{3}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}\left(X_i\right)\right) + \zeta_2\left(\frac{n-1}{n^{\frac{3}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}\left(X_i\right), Y\right)$$

$$\le \frac{1}{2}\kappa_2\left(\frac{n^{\frac{1}{2}}V_n}{2\sigma_U}, \frac{n-1}{n^{\frac{3}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}\left(X_i\right)\right) + \zeta_2\left(\frac{n-1}{n^{\frac{3}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}\left(X_i\right), Y\right),$$

and furthermore applying Proposition 4.3.2 with $N = n^{\varepsilon}$, where $0 < \varepsilon < \frac{\delta}{2}$, we have that

$$\zeta_2\left(\frac{n^{\frac{1}{2}}V_n}{2\sigma_U}, Y\right) \le n^{\varepsilon}\kappa_1\left(\frac{n^{\frac{1}{2}}V_n}{2\sigma_U}, \frac{n-1}{n^{\frac{3}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}\left(X_i\right)\right)$$

$$+ n^{-\varepsilon\delta}\left[E\left|\frac{n^{\frac{1}{2}}V_n}{2\sigma_U}\right|^{2+\delta} + E\left|\frac{n-1}{n^{\frac{3}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}\left(X_i\right)\right|^{2+\delta}\right]$$

$$+ \zeta_2\left(\frac{n-1}{n^{\frac{3}{2}}\sigma_U}\sum_{i=1}^{n} h^{(1)}\left(X_i\right), Y\right).$$

Finally combining Proposition 1.2.9 and Lemmas 4.4.3, 2.1.7, 4.4.10 and we have that

$$\zeta_2\left(\frac{n^{\frac{1}{2}}V_n}{2\sigma_U}, Y\right) \le n^{\varepsilon-\frac{1}{2}} \frac{E\mid h\left(X_1, X_1\right)\mid}{2\sigma_U} + \frac{n^{-\frac{1}{2}}}{\sigma_U}\left[\frac{1}{n(n-1)}\sum_{1\le i<j\le n} E\left|h^{(2)}\left(X_i, X_j\right)\right|^2\right]^{\frac{1}{2}}$$

$$+ n^{-\varepsilon\delta}(n-1)^{-\frac{2+\delta}{2}} n^{\frac{2+\delta}{2}} \frac{3^{1+\delta}}{(2\sigma_U)^{2+\delta}} C_{2+\delta} \frac{2}{n(n-1)} \sum_{1\leq i<j\leq n} E\left|h\left(X_i, X_j\right)\right|^{2+\delta}$$

$$+ n^{-2-\delta(\varepsilon+1)} \frac{C_{2+\delta}}{(2\sigma_U)^{2+\delta}} \frac{1}{n} \sum_{i=1}^{n} E\left|h\left(X_i, X_i\right)\right|^{2+\delta}$$

$$+ n^{-\varepsilon\delta} \left(\frac{n-1}{n}\right)^{2+\delta} \frac{C_{2+\delta}}{\sigma_U^{2+\delta}} \frac{1}{n} \sum_{i=1}^{n} E|h^{(1)}\left(X_i\right)|^{2+\delta}$$

$$+ \frac{a_k}{k}\left[E\left(h^{(1)}\left(X_1\right)\right)^2 - \sigma_U^2\right] - 2\left(1-\frac{a_k}{k}\right) u\left(a_k\right) + c\frac{\rho_k+1}{m^{\frac{1}{2}}},$$

for some constant $c > 0$, where Y is a standard normal random variable, $\{a_k\}$ is any sequence of positive integers such that $a_k \leq k$, $a_k \to \infty$, $\frac{a_k}{k} \to 0$ as $k \to \infty$, $\rho_k := E\left|k^{-\frac{1}{2}} \sum_{i=1}^{k} h^{(1)}\left(X_i\right)\right|^3$ and $u(i) := \sum_{j=i+1}^{\infty} E\left(h^{(1)}\left(X_1\right) h^{(1)}\left(X_j\right)\right) \to 0$ as $i \to \infty$, $C_{2+\delta}$ is a positive constant depending only on δ. ■

Corollary 4.4.12. *If* $E\left|\sum_{i=1}^{\infty} h^{(1)}\left(X_i\right)\right|^3 < \infty$, *under the assumptions of Theorem 4.4.11,*

$$\zeta_2\left(\frac{n^{\frac{1}{2}} V_n}{2\sigma_U}, Y\right) \to 0 \quad \text{as} \quad n \to \infty.$$

Proof. Follows easily from Theorem 4.4.11. ■

Remark 4.4.13. *Note that the previous result, Theorem 4.4.11, is still valid for V-statistics based on a kernel h of bounded variation, with degenerate kernels h^1, h^2 of bounded variation also, because of the fact that a function of bounded variation can be expressed as the difference of two nondecreasing functions.*

4.4.3 Statistical Applications

Estimators of mean and variance

Let $\{X_1, X_2, ..., X_n\}$ be a collection of identically negatively associated random variables from the distribution F with $E(X_1) = \mu$ and $Var(X_1) = \sigma^2$. Suppose that we want to estimate the parameters μ and σ^2. Consider the estimations of $\theta_1 = \mu$ and $\theta_2 = \sigma^2$. We obtain the following U-statistics respectively,

$$\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

$$\hat{\theta}_2 = \frac{2}{n(n-1)} \sum_{1\leq i<j\leq n} \frac{(X_i - X_j)^2}{2} = \frac{2}{n} \sum_{i=1}^{n} h^{(1)}\left(X_i\right) + \binom{n}{2}^{-1} \sum_{1\leq i_1<i_2\leq n} h^{(2)}\left(X_{i_1}, X_{i_2}\right),$$

where

$$h^{(1)}(x_1) = \frac{1}{2}x_1^2 - \mu x_1 + \frac{1}{2}(\mu^2 - \sigma^2),$$

$$h^{(2)}(x_1, x_2) = \mu(x_1 + x_2) - x_1 x_2 - \mu^2.$$

Note that the kernels h, $h^{(1)}$ and $h^{(2)}$ are functions of bounded variation.

Suppose now, that we have a large sample size and we want to construct asymptotic confidence intervals or to perform hypothesis tests for θ_1. Applying Corollary 4.4.7, we get that

$$\frac{\sqrt{n}\left(\hat{\theta}_1 - \theta_1\right)}{\sqrt{\hat{\theta}_2}} \xrightarrow{d} N\left(0, 1\right).$$

The $(1-a)\%$ asymptotic confidence interval for θ_1 is given by

$$\hat{\theta}_1 \pm z_{\frac{a}{2}}\sqrt{\frac{\hat{\theta}_2}{n}},$$

and for the test statistic to perform hypothesis tests we have that

$$\frac{\sqrt{n}\left(\hat{\theta}_1 - \theta_1\right)}{\sqrt{\hat{\theta}_2}} \xrightarrow{d} N\left(0, 1\right).$$

Wilcoxon's one sample rank statistic

Recall Wilcoxon's one sample test, which is used to test if a distribution F is symmetric about zero. Let $\vartheta = P(X_1 + X_2 > 0)$. Using the kernel $h(x_1, x_2) = \mathbb{I}_{\{x_1+x_2>0\}}$, we obtain the following U-statistic:

$$\hat{\vartheta} = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \mathbb{I}_{\{X_i + X_j > 0\}}$$

which is known as the one-sample Wilcoxon statistic. The Hoeffding-decomposition for $\hat{\vartheta}$ is based on kernels

$$h(x_1, x_2) = \mathbb{I}_{\{x_1+x_2>0\}},$$

$$h_1(x_1) = E[\mathbb{I}_{\{x_1+X_2>0\}}] = 1 - P(X_1 \le -x_1),$$

$$h^{(1)}(x_1) = 1 - P(X_1 \le -x_1) - \vartheta,$$

$$h^{(2)}\left(x_1, x_2\right) = h\left(x_1, x_2\right) - h_1\left(x_1\right) - h_1\left(x_2\right) + \vartheta.$$

Note that the kernels h, $h^{(1)}$ and $h^{(2)}$ are functions of bounded variation.

Suppose that we have a large sample size. Applying Corollary 4.4.7 we construct asymptotic confidence intervals and we have the asymptotic distribution to perform hypothesis tests for parameter ϑ.

Chapter 5

Jackknifing U-statistics based on associated and negatively associated random variables

In Chapter 5 we apply the jackknife technique on U-statistics based on associated and negatively associated random variables. The jackknife technique is a useful method of variance estimation. Chapter 5 is organized in two sections. In Section 5.1 we jackknife U-statistics based on associated random variables and in Section 5.2 we jackknife U-statistics based on negatively associated random variables.

5.1 Jackknifing U-statistics based on associated random variables

5.1.1 The jackknife estimate of variance for U-statistics

We consider the jackknife pseudovalues for U-statistics by Tukey

$$\hat{U}_i = nU_n - (n-1)U_n(-i) \quad for \quad i = 1, 2, ..., n,$$

where $U_n(-i)$ is the U-statistic computed on the sample of $n-1$ variables formed from the original data set by deleting the ith data value. Then the jackknife estimate is the average

$$\hat{U}_n = \frac{1}{n}\sum_{i=1}^{n} \hat{U}_i,$$

and the jackknife estimate of the variance is given by

$$\widehat{Var}(JACK) = \frac{n-1}{n} \sum_{i=1}^{n} [U_n(-i) - U_n]^2.$$

Now, using the Hoeffding decomposition we get that

$$U_n(-i) = \theta + 2H_n^{(1)}(-i) + H_n^{(2)}(-i)$$

where

$$H_n^{(1)}(-i) = \frac{1}{n-1} \sum_{\substack{k=1 \\ k \neq i}}^{n} h^{(1)}(X_k) \quad and \quad H_n^{(2)}(-i) = \binom{n-1}{2}^{-1} \sum_{\substack{1 \leq s < t \leq n \\ s,t \neq i}} h^{(2)}(X_s, X_t).$$

Applying the H-decomposition we rewrite the jackknife estimate of variance as

$$\begin{aligned} \widehat{Var}(JACK) &= \frac{n-1}{n} \sum_{i=1}^{n} \Big[2(H_n^{(1)}(-i) - H_n^{(1)}) + (H_n^{(2)}(-i) - H_n^{(2)}) \Big]^2 \\ &= \frac{n-1}{n} \Bigg\{ 4 \sum_{i=1}^{n} \big[H_n^{(1)}(-i) - H_n^{(1)} \big]^2 \\ &\quad + \sum_{i=1}^{n} \big[H_n^{(2)}(-i) - H_n^{(2)} \big]^2 \\ &\quad + 4 \sum_{i=1}^{n} [H_n^{(1)}(-i) - H_n^{(1)}][H_n^{(2)}(-i) - H_n^{(2)}] \Bigg\}. \end{aligned}$$

5.1.2 The bias of the jackknife estimate of the variance

Below we present some various results that are connected with associated random variables and with U-statistics based on this type of random variables that are needed for our exposition.

Lemma 5.1.1. *(Dewan and Prakasa Rao (2002)). Let* $\{X_n,\ n \geq 1\}$ *be a sequence of stationary associated random variables. Let* $\sigma_U^2 = \sigma_1^2 + 2\sum_{j=1}^{\infty} \sigma_{1j}^2 > 0$ *where* $\sigma_1^2 = \mathrm{Var}[h_1(X_1)]$ *and* $\sigma_{1j}^2 = \mathrm{cov}(h_1(X_1), h_1(X_{1+j}))$*. Assume that*

$$\sum_{j=1}^{\infty} \sigma_{1j}^2 < \infty.$$

Further suppose that for some non-negative function $r(k)$ *satisfying* $\sum_{k=0}^{\infty} r(k) < \infty$, *we have for all* (i, j, k, l)

$$|\Delta(i,j,k,l)| \leq r(\max[|i-k|, |j-l|]),$$

where

$$\Delta(i,j,k,l) = \text{Cov}(h^{(2)}(X_i, X_j), h^{(2)}(X_k, X_l)).$$

Then

$$\text{Var}(U_n) = \frac{4\sigma_U^2}{n} + o\left(\frac{1}{n}\right).$$

Proof. See the proof of Theorem 3.1 in Dewan and Prakasa Rao (2002). ■

Lemma 5.1.2. *Suppose that all conditions of Lemma 5.1.1 are satisfied. Then*

$$\text{Var}(H_n^{(1)}) = \frac{1}{n}\left(\sigma_1^2 + 2\sum_{j=1}^{\infty}\sigma_{1j}^2\right) + o\left(\frac{1}{n}\right),$$

$$\text{Var}(H_n^{(2)}) = o\left(\frac{1}{n}\right),$$

$$|\text{Cov}(H_n^{(1)}, H_n^{(2)})| \leq o\left(\frac{1}{n}\right).$$

Proof. See the proof of Theorem 3.1 from Dewan and Prakasa Rao (2002). ■

Lemma 5.1.3. *Let* $\{X_1, X_2, ..., X_n\}$ *be a collection of identically distributed stationary associated random variables. Let* U_n *be a U-statistic based on a symmetric kernel* $h(x, y)$ *of degree 2. Suppose that the conditions of Lemma 5.1.1 hold. Then*

$$\sqrt{n}(U_n - \theta) \xrightarrow{D} N(0, 4\sigma_U^2) \quad as \quad n \to \infty,$$

where $\theta = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(x,y)dF(x)dF(y)$.

Proof. See the proof of Theorem 3.2 in Dewan and Prakasa Rao (2002). ■

The next three Lemmas 5.1.4, 5.1.5, 5.1.6, give a simplified expression of $\widehat{Var}(JACK)$ which is crucial to exploit the bias of the jackknife estimate of the variance.

Lemma 5.1.4. *Considering the kernels* $H_n^{(1)}(-i)$ *and* $H_n^{(1)}$ *we get that*

$$4\sum_{i=1}^{n}\left[H_n^{(1)}(-i) - H_n^{(1)})\right]^2 = \frac{4}{(n-1)^2}\left[\sum_{i=1}^{n}[h^{(1)}(X_i)]^2 - n(H_n^{(1)})^2\right].$$

Proof. First we calculate the difference $H_n^{(1)}(-i) - H_n^{(1)}$,

$$H_n^{(1)}(-i) - H_n^{(1)} = \frac{1}{n-1}\sum_{\substack{k=1\\k\neq i}}^{n} h^{(1)}(X_k) - \frac{1}{n}\sum_{k=1}^{n} h^{(1)}(X_k)$$

$$= \frac{1}{n(n-1)}\left[n\sum_{\substack{k=1\\k\neq i}}^{n} h^{(1)}(X_k) - (n-1)\sum_{k=1}^{n} h^{(1)}(X_k)\right]$$

$$= \frac{1}{n(n-1)}\left[\sum_{k=1}^{n} h^{(1)}(X_k) - nh^{(1)}(X_i)\right]$$

$$= \frac{1}{(n-1)}\left[H_n^{(1)} - h^{(1)}(X_i)\right].$$

Using the previous result we arrive at

$$4\sum_{i=1}^{n}\left[H_n^{(1)}(-i) - H_n^{(1)})\right]^2 = \frac{4}{(n-1)^2}\sum_{i=1}^{n}\left[H_n^{(1)} - h^{(1)}(X_i)\right]^2$$

$$= \frac{4}{(n-1)^2}\left[n(H_n^{(1)})^2 - 2n(H_n^{(1)})^2 + \sum_{i=1}^{n}[h^{(1)}(X_i)]^2\right]$$

$$= \frac{4}{(n-1)^2}\left[\sum_{i=1}^{n}[h^{(1)}(X_i)]^2 - n(H_n^{(1)})^2\right]. \quad \blacksquare$$

Lemma 5.1.5. *Considering the kernels $H_n^{(2)}(-i)$ and $H_n^{(2)}$ it follows that*

$$\sum_{i=1}^{n}\left[H_n^{(2)}(-i) - H_n^{(2)}\right]^2 = \frac{4}{(n-1)^2(n-2)^2}\left\{2\sum_{1\leq i<j\leq n}\left(h^{(2)}(X_i, X_j)\right)^2\right.$$

$$\left. + 2\sum_{i=1}^{n}\sum_{\substack{1\leq s<t\leq n\\s,t\neq i}} h^{(2)}(X_i, X_s)\, h^{(2)}(X_i, X_t) - n(n-1)^2(H_n^{(2)})^2\right\}.$$

Proof. Starting with the difference $H_n^{(2)}(-i) - H_n^{(2)}$, we have that

$$H_n^{(2)}(-i) - H_n^{(2)} = \binom{n-1}{2}^{-1}\sum_{\substack{1\leq s<t\leq n\\s,t\neq i}} h^{(2)}(X_s, X_t) - \binom{n}{2}^{-1}\sum_{1\leq s<t\leq n} h^{(2)}(X_s, X_t)$$

$$= \frac{2}{n(n-1)(n-2)}\left[n\sum_{\substack{1\leq s<t\leq n\\s,t\neq i}} h^{(2)}(X_s, X_t) - (n-2)\sum_{1\leq s<t\leq n} h^{(2)}(X_s, X_t)\right]$$

$$= \frac{2}{n(n-1)(n-2)}\left[2\binom{n}{2}H_n^{(2)} - n\sum_{\substack{k=1\\k\neq i}}^{n} h^{(2)}\left(X_i, X_k\right)\right]$$

$$= \frac{2}{(n-1)(n-2)}\left[(n-1)H_n^{(2)} - \sum_{\substack{k=1\\k\neq i}}^{n} h^{(2)}\left(X_i, X_k\right)\right].$$

Using the previous result we can write

$$\sum_{i=1}^{n}\left[H_n^{(2)}(-i) - H_n^{(2)}\right]^2 = \frac{4}{(n-1)^2(n-2)^2}\sum_{i=1}^{n}\left[(n-1)H_n^{(2)} - \sum_{\substack{k=1\\k\neq i}}^{n} h^{(2)}\left(X_i, X_k\right)\right]^2$$

$$= \frac{4}{(n-1)^2(n-2)^2}\Bigg\{n(n-1)^2(H_n^{(2)})^2 - 2(n-1)H_n^{(2)}\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} h^{(2)}\left(X_i, X_k\right)$$

$$+\sum_{i=1}^{n}\left(\sum_{\substack{k=1\\k\neq i}}^{n} h^{(2)}\left(X_i, X_k\right)\right)^2\Bigg\}$$

$$= \frac{4}{(n-1)^2(n-2)^2}\Bigg\{n(n-1)^2(H_n^{(2)})^2 - 2(n-1)H_n^{(2)}2\binom{n}{2}H_n^{(2)} + 2\sum_{1\le i<j\le n}\left(h^{(2)}\left(X_i, X_j\right)\right)^2$$

$$+ 2\sum_{i=1}^{n}\sum_{\substack{1\le s<t\le n\\s,t\neq i}} h^{(2)}\left(X_i, X_s\right)h^{(2)}\left(X_i, X_t\right)\Bigg\}$$

$$= \frac{4}{(n-1)^2(n-2)^2}\Bigg\{2\sum_{1\le i<j\le n}\left(h^{(2)}\left(X_i, X_j\right)\right)^2 + 2\sum_{i=1}^{n}\sum_{\substack{1\le s<t\le n\\s,t\neq i}} h^{(2)}\left(X_i, X_s\right)h^{(2)}\left(X_i, X_t\right)$$

$$- n(n-1)^2(H_n^{(2)})^2\Bigg\}. \blacksquare$$

Lemma 5.1.6. *Considering the kernels $H_n^{(j)}(-i)$ and $H_n^{(j)}$ with $j = 1, 2$ it follows that*

$$4\sum_{i=1}^{n}[H_n^{(1)}(-i) - H_n^{(1)}][H_n^{(2)}(-i) - H_n^{(2)}]$$

$$= \frac{8}{(n-1)^2(n-2)}\left\{\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} h^{(1)}\left(X_i\right)h^{(2)}\left(X_i, X_k\right) - n(n-1)H_n^{(1)}H_n^{(2)}\right\}.$$

Proof. The multiplication of the differences $H_n^{(1)}(-i) - H_n^{(1)}$ and $H_n^{(2)}(-i) - H_n^{(2)}$ gives that

$$[H_n^{(1)}(-i) - H_n^{(1)}][H_n^{(2)}(-i) - H_n^{(2)}] =$$

$$= \frac{1}{(n-1)}\left[H_n^{(1)} - h^{(1)}(X_i)\right]\frac{2}{(n-1)(n-2)}\left[(n-1)H_n^{(2)} - \sum_{\substack{k=1\\k\neq i}}^{n} h^{(2)}(X_i, X_k)\right]$$

$$= \frac{2}{(n-1)^2(n-2)}\left\{(n-1)H_n^{(1)}H_n^{(2)} - H_n^{(1)}\sum_{\substack{k=1\\k\neq i}}^{n} h^{(2)}(X_i, X_k) - (n-1)h^{(1)}(X_i)\,H_n^{(2)}\right.$$

$$\left. + h^{(1)}(X_i)\sum_{\substack{k=1\\k\neq i}}^{n} h^{(2)}(X_i, X_k)\right\}.$$

Hence, from the previous calculations we have that

$$4\sum_{i=1}^{n}[H_n^{(1)}(-i) - H_n^{(1)}][H_n^{(2)}(-i) - H_n^{(2)}] =$$

$$= \frac{8}{(n-1)^2(n-2)}\left\{n(n-1)H_n^{(1)}H_n^{(2)} - H_n^{(1)}2\binom{n}{2}H_n^{(2)} - n(n-1)H_n^{(1)}H_n^{(2)}\right.$$

$$\left. + \sum_{i=1}^{n}\left[h^{(1)}(X_i)\sum_{\substack{k=1\\k\neq i}}^{n} h^{(2)}(X_i, X_k)\right]\right\}$$

$$= \frac{8}{(n-1)^2(n-2)}\left\{\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} h^{(1)}(X_i)\,h^{(2)}(X_i, X_k) - n(n-1)H_n^{(1)}H_n^{(2)}\right\}. \quad \blacksquare$$

The following Lemma 5.1.7 gives the simplified expression of $\widehat{Var}(JACK)$.

Lemma 5.1.7. *The jackknife estimator of variance, $\widehat{Var}(JACK)$ can be expressed as*

$$\widehat{Var}(JACK) = \frac{4}{(n-1)^2}\left[\sum_{i=1}^{n}[h^{(1)}(X_i)]^2 - n(H_n^{(1)})^2\right]$$

$$+ \frac{4}{(n-1)^2(n-2)^2}\left\{2\sum_{1\leq i<j\leq n}\left(h^{(2)}(X_i, X_j)\right)^2 + \right.$$

$$\left. 2\sum_{i=1}^{n}\sum_{\substack{1\leq s<t\leq n\\s,t\neq i}} h^{(2)}(X_i, X_s)\,h^{(2)}(X_i, X_t) - n(n-1)^2(H_n^{(2)})^2\right\}$$

$$+\frac{8}{(n-1)^2(n-2)}\left\{\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} h^{(1)}(X_i)\,h^{(2)}(X_i,X_k) - n(n-1)H_n^{(1)}H_n^{(2)}\right\}.$$

Proof. Apply Lemmas 5.1.4, 5.1.5, 5.1.6. ∎

Lemma 5.1.8. *The expectation of $\widehat{Var}(JACK)$ is given by*

$$E[\widehat{Var}(JACK)] = \frac{4}{(n-1)^2}\left[\sum_{i=1}^{n} E[h^{(1)}(X_i)]^2 - nE(H_n^{(1)})^2\right]$$

$$+\frac{4}{(n-1)^2(n-2)^2}\left\{2\sum_{1\leq i<j\leq n} E\Big(h^{(2)}(X_i,X_j)\Big)^2 + \right.$$

$$\left. 2\sum_{i=1}^{n}\sum_{\substack{1\leq s<t\leq n\\ s,t\neq i}} E[h^{(2)}(X_i,X_s)\,h^{(2)}(X_i,X_t)] - n(n-1)^2E(H_n^{(2)})^2\right\}$$

$$+\frac{8}{(n-1)^2(n-2)}\left\{\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} E[h^{(1)}(X_i)\,h^{(2)}(X_i,X_k)] - n(n-1)E[H_n^{(1)}H_n^{(2)}]\right\}.$$

Proof. Using Lemma 5.1.7 we easily get the result. ∎

Theorem 5.1.9. *Let $\{X_1, X_2, ..., X_n\}$ be a collection of identically distributed stationary associated random variables. Let U_n be the U-statistic of dimension two based on the kernel h and the given observations. Assume that h is a real valued function symmetric in its arguments with $E(h) = 0$. Define $\sigma_U^2 = \sigma_1^2 + 2\sum_{j=1}^{\infty}\sigma_{1j}^2 > 0$ where $\sigma_1^2 = \mathrm{Var}[h_1(X_1)]$ and $\sigma_{1j}^2 = \mathrm{cov}(h_1(X_1), h_1(X_{1+j}))$. Under the assumptions of Lemma 5.1.1 and furthermore let $E[h^{(1)}(X_i)]^2 \leq C_1 < \infty$ for all $1 \leq i \leq n$, $E[h^{(2)}(X_i,X_j)]^2 \leq C_2 < \infty$ for all $1 \leq i < j \leq n$, $E[h^{(2)}(X_i,X_s)\,h^{(2)}(X_i,X_t)] \leq C_3 < \infty$ for all $1 \leq i \leq n$ and $1 \leq s < t \leq n$ with $s,t \neq i$ and $E\left[h^{(1)}(X_i)\,h^{(2)}(X_i,X_k)\right] \leq C_4 < \infty$ for all $1 \leq i \leq n$, $1 \leq k \leq n$ with $k \neq i$, we have that*

$$E[\widehat{Var}(JACK)] \to 0 \quad as \quad n \to \infty,$$

where C_1, C_2, C_3, C_4 are positive constants.

Proof. Applying Lemmas 5.1.1, 5.1.2 and 5.1.8 we get the result. ∎

Corollary 5.1.10. *Under the assumptions of Theorem 5.1.9 we have that*

$$BIAS(JACK) = E[\widehat{Var}(JACK)] - Var(U_n) \to 0 \quad as \quad n \to \infty.$$

Proof. Applying Theorem 5.1.9 and Lemma 5.1.1 we get the result. ■

Remark 5.1.11. *Note that $E[\widehat{Var}(JACK)]$ and $BIAS(JACK)$ are of order n^{-1}.*

5.1.3 Jackknifing functions of U-statistics

Let θ be an unknown parameter and g be a real valued function. Consider using the function $g(U_n)$ of U_n to estimate $g(\theta)$. The jackknife estimate of $g(\theta)$ is given by

$$\widehat{g}(U_n) = \frac{1}{n}\sum_{i=1}^{n} \widehat{g}(U_i),$$

where

$$\widehat{g}(U_i) = ng(U_n) - (n-1)g(U_n(-i)) \quad for \quad i = 1, 2, ..., n, \tag{5.1.1}$$

are the jackknife pseudovalues for functions of U-statistics by Tukey. The estimate of $Var[g(U_n)]$ is

$$\widehat{Var}(JACK) = \frac{n-1}{n}\sum_{i=1}^{n} \left[g(U_n(-i)) - g(U_n)\right]^2.$$

Before we establish the theorem of jackknifing functions of U-statistics, we present some useful results.

Lemma 5.1.12. *(Taylor's Theorem) Let g be an $(n+1)$ times differentiable function on an open interval containing the points a and x. Then*

$$g(x) = g(a) + g'(a)(x-a) + \frac{g''(a)}{2!}(x-a)^2 + \cdots + \frac{g^{(n)}(a)}{n!}(x-a)^n + R_n(x)$$

where

$$R_n(x) = \frac{g^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}$$

for some number c between a and x.

Lemma 5.1.13. *Let* $X_n \xrightarrow{p} \theta$ *and* $\sqrt{n}(X_n - \theta) \xrightarrow{D} N(0, \sigma^2)$. *Suppose that a function* g *has a continuous derivative in* $(\theta - \varepsilon, \theta + \varepsilon)$, *with* $\varepsilon > 0$ *and* $g'(\theta) \neq 0$. *Then*

$$\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{D} N\left(0, \frac{\sigma^2}{[g'(\theta)]^2}\right).$$

We now obtain the theorem of jackknifing functions of U-statistics.

Theorem 5.1.14. *Suppose that the conditions of Lemma 5.1.3 hold. Let* g *the function have a bounded second derivative in a neighborhood of* θ. *Then*

$$\sqrt{n}(\widehat{g}(U_n) - g(\theta)) \xrightarrow{D} N\left(0, \frac{4\sigma_U^2}{[g'(\theta)]^2}\right)$$

provided that $g'(\theta) \neq 0$.

Proof. The function g has a bounded second derivative in a neighborhood of U_n. Applying Taylor's Theorem we may expand g about U_n and obtain

$$g(U_n(-i)) = g(U_n) + g'(U_n)(U_n(-i) - U_n) + \frac{g''(\xi_i)}{2!}(U_n(-i) - U_n)^2$$

where ξ_i lies between $U_n(-i)$ and U_n. Summing both side and dividing by n we have that

$$\frac{1}{n}\sum_{i=1}^{n} g(U_n(-i)) = g(U_n) + \frac{1}{2n}\sum_{i=1}^{n}(U_n(-i) - U_n)^2 g''(\xi_i) \tag{5.1.2}$$

according to the relation $U_n = \frac{1}{n}\sum_{i=1}^{n} U_n(-i)$.
Also summing both sides and dividing by n in (5.1.1), we have that

$$\frac{1}{n}\sum_{i=1}^{n} g(U_n(-i)) = \frac{ng(U_n) - \widehat{g}(U_n)}{n-1}. \tag{5.1.3}$$

Combining (5.1.2) and (5.1.3) we get

$$\widehat{g}(U_n) = g(U_n) - \frac{n-1}{2n}\sum_{i=1}^{n}(U_n(-i) - U_n)^2 g''(\xi_i),$$

and so

$$\sqrt{n}(\widehat{g}(U_n) - g(\theta)) = \sqrt{n}(g(U_n) - g(\theta)) - \frac{\sqrt{n}}{2}\frac{n-1}{n}\sum_{i=1}^{n}(U_n(-i) - U_n)^2 g''(\xi_i). \tag{5.1.4}$$

Since the second derivative of g is assumed to be bounded in a neighborhood of θ, combining Lemmas 5.1.3 and 5.1.13 the first term on the right hand side of (5.1.4) converges in distribution to $N\left(0, \frac{4\sigma_U^2}{[g'(\theta)]^2}\right)$. According to Slutsky's theorem we need to prove that

$$-\frac{\sqrt{n}}{2}\frac{n-1}{n}\sum_{i=1}^{n}(U_n(-i)-U_n)^2 g''(\xi_i) \xrightarrow{p} 0 \tag{5.1.5}$$

to have the result. Since $U_n(-i)$ and U_n are consistent estimators of θ, it suffices to show that $\sqrt{n}\,\widehat{Var}(JACK) \xrightarrow{D} 0$ which is obviously applying from Lemma 5.1.7. ■

5.2 Jackknifing U-statistics based on negatively associated random variables

Consider an orthonormal basis $\{e_k(x), k \geq 0\}$ with respect to the measure $dF(x)$, with $e_0(x) = 1$, such that

$$h(x,y) = \sum_{k=0}^{\infty} \lambda_k e_k(x) e_k(y).$$

Then

$$\int_{-\infty}^{+\infty} e_k(x) h(x,y) dF(x) = \lambda_k e_k(y). \tag{5.2.1}$$

Definition 5.2.1. *The U-statistic U_n and its kernel h are called degenerate if*

$$\int_{-\infty}^{+\infty} h(x,y) dF(y) = 0$$

for all x.

Lemma 5.2.2. *(Huang and Zhang (2006)). Let $\{X_1, X_2, ..., X_n\}$ be a collection of identically distributed stationary negatively associated random variables with $E(X_1) = 0$. Let U_n be a degenerate U-statistic where the kernel h satisfies*

$$\int_{-\infty}^{+\infty} h(x,y) dF(x) dF(y) < \infty.$$

Assume that the eigenfunctions $e_k(x)$ given by (5.2.1) are functions with bounded vari-

ation on any finite interval. Furthermore, assume that

$$sup_{k\geqslant 1}E\left[V_{e_k}^2\left(X_1\right)\right]<\infty$$

and

$$\sum_{i=1}^{\infty}|\lambda_k|<\infty.$$

Then

$$\sqrt{n}(U_n-E(U_n))\xrightarrow{D}N(0,\sigma^2)\quad as\quad n\to\infty,$$

where

$$\sigma^2=Var(\varphi(X_1))+2\sum_{j=2}^{\infty}Cov(\varphi(X_i),\varphi(X_j))$$

and

$$\varphi(x)=\int_{-\infty}^{+\infty}h(x,y)dF(y),$$

provided $\varphi(.)$ *is a function with bounded variation on any finite interval, which satisfies* $E\left[V_{\varphi}^2\left(X_1\right)\right]<\infty$, *where* $V_{\varphi}(x)$ *is the total variation function of* $\varphi(x)$, *i.e,* $V_{\varphi}(x)=V_0^x(\varphi)$ *for* $x\geq 0$, $V_{\varphi}(x)=-V_x^0(\varphi)$ *for* $x\leq 0$, *where*

$$V_a^b(\varphi)=\sup\sum_{k=1}^{n}\left|\varphi\left(x_k\right)-\varphi\left(x_{k-1}\right)\right|$$

denotes the total variation of $\varphi(x)$ *on* $[a,b]$. *The supremum is taken over all partitions of the interval* $[a,b]$.

Proof. See the proof of Theorem 2 from Huang and Zhang (2006). ■

Corollary 5.2.3. *Suppose, the conditions of Lemma 5.2.2 hold. Then*

$$\mathrm{Var}\left(U_n\right)=\frac{4\sigma^2}{n}\quad as\quad n\to\infty.$$

Lemma 5.2.4. *The jackknife estimator of variance,* $\widehat{Var}(JACK)$ *can be expressed as*

$$\widehat{Var}(JACK)=\frac{4}{(n-1)^2}\left[\sum_{i=1}^{n}[h^{(1)}\left(X_i\right)]^2-n(H_n^{(1)})^2\right]$$

$$+\frac{4}{(n-1)^2(n-2)^2}\Bigg\{2\sum_{1\le i<j\le n}\Big(h^{(2)}(X_i,X_j)\Big)^2+$$

$$2\sum_{i=1}^{n}\sum_{\substack{1\le s<t\le n\\ s,t\neq i}}h^{(2)}(X_i,X_s)\,h^{(2)}(X_i,X_t)-n(n-1)^2(H_n^{(2)})^2\Bigg\}$$

$$+\frac{8}{(n-1)^2(n-2)}\left\{\sum_{i=1}^{n}\sum_{\substack{k=1\\ k\neq i}}^{n}h^{(1)}(X_i)\,h^{(2)}(X_i,X_k)-n(n-1)H_n^{(1)}H_n^{(2)}\right\}.$$

Proof. Same as Lemma 5.1.7. ■

Lemma 5.2.5. *The expectation of $\widehat{Var}(JACK)$ is given by*

$$E[\widehat{Var}(JACK)]=\frac{4}{(n-1)^2}\left[\sum_{i=1}^{n}E[h^{(1)}(X_i)]^2-nE(H_n^{(1)})^2\right]$$

$$+\frac{4}{(n-1)^2(n-2)^2}\Bigg\{2\sum_{1\le i<j\le n}E\Big(h^{(2)}(X_i,X_j)\Big)^2+$$

$$2\sum_{i=1}^{n}\sum_{\substack{1\le s<t\le n\\ s,t\neq i}}E[h^{(2)}(X_i,X_s)\,h^{(2)}(X_i,X_t)]-n(n-1)^2E(H_n^{(2)})^2\Bigg\}$$

$$+\frac{8}{(n-1)^2(n-2)}\left\{\sum_{i=1}^{n}\sum_{\substack{k=1\\ k\neq i}}^{n}E[h^{(1)}(X_i)\,h^{(2)}(X_i,X_k)]-n(n-1)E[H_n^{(1)}H_n^{(2)}]\right\}.$$

Proof. Same as Lemma 5.1.8. ■

Theorem 5.2.6. *Let $\{X_1,X_2,...,X_n\}$ be a collection of identically distributed stationary negatively associated random variables. Under the assumptions of Lemma 5.2.2 and furthermore let $E[h^{(1)}(X_i)]^2\le C_1<\infty$ for all $1\le i\le n$, $E[h^{(2)}(X_i,X_j)]^2\le C_2<\infty$ for all $1\le i<j\le n$, $E[h^{(2)}(X_i,X_s)\,h^{(2)}(X_i,X_t)]\le C_3<\infty$ for all $1\le i\le n$ and $1\le s<t\le n$ with $s,t\neq i$ and $E\left[h^{(1)}(X_i)\,h^{(2)}(X_i,X_k)\right]\le C_4<\infty$ for all $1\le i\le n$, $1\le k\le n$ with $k\neq i$, we have that $E[h^{(2)}(X_s,X_t)]^2\le C_1<\infty$ for all $1\le s<t\le n$ and $E\left[h^{(1)}(X_i)\,h^{(2)}(X_i,X_k)\right]\le C_2<\infty$ for all $1\le i\le n$, $1\le k\le n$ and $k\neq i$, we have that*

$$E[\widehat{Var}(JACK)]\to 0\quad as\quad n\to\infty\quad and$$

$$BIAS(JACK)=E[\widehat{Var}(JACK)]-Var(U_n)\to 0\quad as\quad n\to\infty,$$

where C_1, C_2, C_3, C_4 are positive constants.

Proof. Applying Lemmas 5.2.5 and 5.2.3. ■

Remark 5.2.7. *Note that* $E[\widehat{Var}(JACK)]$ *and* $BIAS(JACK)$ *are of order* n^{-1}.

Theorem 5.2.8. *Suppose, the conditions of Lemma 5.2.2 hold. Let* g *the function have a bounded second derivative in a neighborhood of* $\theta = E(U_n)$. *Then*

$$\sqrt{n}(\widehat{g}(U_n) - g(\theta)) \xrightarrow{D} N\left(0, \frac{4\sigma_U^2}{[g'(\theta)]^2}\right)$$

provided that $g'(\theta) \neq 0$.

Proof. The function g has a bounded second derivative in a neighborhood of U_n. Applying Taylor's Theorem we may expand g about U_n and obtain

$$g(U_n(-i)) = g(U_n) + g'(U_n)(U_n(-i) - U_n) + \frac{g''(\xi_i)}{2!}(U_n(-i) - U_n)^2$$

where ξ_i lies between $U_n(-i)$ and U_n. Summing both side and dividing by n we have that

$$\frac{1}{n}\sum_{i=1}^{n} g(U_n(-i)) = g(U_n) + \frac{1}{2n}\sum_{i=1}^{n}(U_n(-i) - U_n)^2 g''(\xi_i) \tag{5.2.2}$$

according to the relation $U_n = \frac{1}{n}\sum_{i=1}^{n} U_n(-i)$.

Also summing both sides and dividing by n in (5.1.1), we have that

$$\frac{1}{n}\sum_{i=1}^{n} g(U_n(-i)) = \frac{ng(U_n) - \widehat{g}(U_n)}{n-1}. \tag{5.2.3}$$

Combining (5.2.2) and (5.2.3) we get

$$\widehat{g}(U_n) = g(U_n) - \frac{n-1}{2n}\sum_{i=1}^{n}(U_n(-i) - U_n)^2 g''(\xi_i),$$

and so

$$\sqrt{n}(\widehat{g}(U_n) - g(\theta)) = \sqrt{n}(g(U_n) - g(\theta)) - \frac{\sqrt{n}}{2}\frac{n-1}{n}\sum_{i=1}^{n}(U_n(-i) - U_n)^2 g''(\xi_i). \tag{5.2.4}$$

Since the second derivative of g is assumed to be bounded in a neighborhood of θ, combining Lemmas 5.2.2 and 5.1.13 the first term on the right hand side of (5.2.4) converges in distribution to $N\left(0, \frac{4\sigma_U^2}{[g'(\theta)]^2}\right)$. According to Slutsky's theorem we need to prove that

$$-\frac{\sqrt{n}}{2}\frac{n-1}{n}\sum_{i=1}^{n}(U_n(-i)-U_n)^2 g''(\xi_i) \xrightarrow{p} 0 \tag{5.2.5}$$

to have the result. Since $U_n(-i)$ and U_n are consistent estimators of θ, it suffices to show that $\sqrt{n}\ \widehat{Var}(JACK) \xrightarrow{D} 0$ which is obviously applying from Lemma 5.2.4. ■

Chapter 6

Future work

In this dissertation we studied the asymptotic behavior of U-statistics based on associated and negatively associated random variables. However in no way we have exhausted all possible directions of researching the specific asymptotic behavior. Some results that are presented in the thesis can be extended or can be a starting point for further research. In what follows we briefly describe some directions for future work.

6.1 Distance between a U-statistics based on associated random variables and a normal random variables

In this thesis we studied asymptotic results for U-statistics using the Zolotarev's ideal metric. In particular, we gave the distance between a U-statistic U_n based on associated random variables and a U-statistic U_n^* based on i.i.d. random variables. We obtained the distance between a U-statistic based on i.i.d. random variables and a normal random variable. The same result also established for U-statistics based on negatively associated random variables. In future work our goal is to obtain the distance between a U-statistic based on associated random variables and a normal random variable by utilizing the Zolotarev's ideal metric. This result will also provide a central limit theorem for this type of U-statistics. Corresponding results for V-statistics could also be a topic of investigation. For this aim existing results for partial sums of associated sequence will be very useful. It is also expected that results proved for the classical setup (i.e. for iid observations) which are not applicable for the case of associated observations will be modified or replaced altogether by results on associated random variables.

6.2 U-statistics based on $\mathcal{F}$-associated random variables

The type of dependence that we study in this thesis is association and negative association. Below we give the definition of an alternative concept of dependence called conditional association. The concept of condition association was introduced in Prakasa Rao (2009).

Definition 6.2.1. *Let X and Y be random variables defined on a probability space $(\Omega, \mathcal{A}, \mathcal{P})$ with $E(X^2) < \infty$ and $E(Y^2) < \infty$. Let $\mathcal{F}$ be a sub-σ-algebra of $\mathcal{A}$. We define the conditional covariance of X and Y given $\mathcal{F}$ or $\mathcal{F}$-covariance as*

$$Cov^{\mathcal{F}}(X,Y) = E^{\mathcal{F}}\left[(X - E^{\mathcal{F}}X)\left(Y - E^{\mathcal{F}}Y\right)\right],$$

where $E^{\mathcal{F}}(X) = E(X|\mathcal{F})$, (cf. Prakasa Rao (2009)). It is easy to see that the $\mathcal{F}$-covariance reduces to the ordinary concept of covariance when $\mathcal{F} = \{\emptyset, \Omega\}$. A set of random variables $\{X_k,\ 1 \leq k \leq n\}$ is said to be $\mathcal{F}$-associated if for any componentwise nondegreasing functions h, g defined on $\mathcal{R}^n$,

$$Cov^{\mathcal{F}}(h(X_1, ..., X_n), g(X_1, ..., X_n)) \geq 0 \quad a.s.$$

Remark 6.2.2. *A sequence of random variables $\{X_n,\ n \geq 1\}$ is said to be $\mathcal{F}$-associated if every finite subset of the sequences $\{X_n,\ n \geq 1\}$ is $\mathcal{F}$-associated.*

A relative concept is the concept of conditional demimartingales. Hadjikyriakou (2010) introduced the notion of conditional demimartingales and studied their properties.

Definition 6.2.3. *Let $\{S_n,\ n \geq 1\}$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{A}, \mathcal{P})$. Let $\mathcal{F}$ be a sub-σ-algebra of $\mathcal{A}$. The sequence $\{S_n,\ n \geq 1\}$ is called an $\mathcal{F}$-demimartingale if for every componentwise nondecreasing function*

$f : \mathbb{R}^j \to \mathbb{R}$,

$$E\left\{(S_j - S_i)\, f\,(S_1, S_2, ..., S_i)|\mathcal{F})\right\} \geq 0, \quad 1 \leq i < j < \infty.$$

If moreover, f is assumed to be nonnegative, then the sequence $\{S_n\}_{n\geq 1}$ is called an $\mathcal{F}$-demisubmartingale.

Remark 6.2.4. *From the property of conditional expectations that $E(E(Z|F)) = E(Z)$ for any random variable Z with $E|Z| < \infty$, it follows that any $\mathcal{F}$-demimartingale defined on a probability space $(\Omega, \mathcal{A}, \mathcal{P})$ is a demimartingale on the probability space $(\Omega, \mathcal{A}, \mathcal{P})$ and any $\mathcal{F}$-demisubmartingale defined on the probability space $(\Omega, \mathcal{A}, \mathcal{P})$ is a demisubmartingale on the probability space $(\Omega, \mathcal{A}, \mathcal{P})$. The converse cannot always be true due to Example 2.5.5 from Hadjikyriakou (2010).*

Moreover we present a trivial example of an $\mathcal{F}$-demimartingale sequence.

Example 6.2.5. *Let $\{X_n,\ n \geq 1\}$ be a sequence of $\mathcal{F}$-associated random variables such that $E^{\mathcal{F}}(X_n) = 0$ a.s., $n \geq 1$. Let*

$$S_n = \sum_{i=1}^{n} X_i, \quad n \geq 1.$$

Then it is easy to check that the sequence $\{S_n\}_{n\geq 1}$ is an $\mathcal{F}$-demimartingale.

In what follows we obtain the definition of U-statistics based on $\mathcal{F}$-associated random variables.

Definition 6.2.6. *Let $\{X_1, X_2, ..., X_n\}$ be a collection of identically distributed $\mathcal{F}$-associated random variables. Assume that m is a positive integer $m \leq n$ and h is a symmetric mapping from $\mathbb{R}^m$ to $\mathbb{R}$. Then*

$$U_n^{\mathcal{F}} = \binom{n}{m}^{-1} \sum_{1\leq i_1 < i_2 < \cdots < i_m \leq n} h\left(X_{i_1}, X_{i_2}, ..., X_{i_m}\right)$$

where $\sum_{1\leq i_1 < i_2 < \cdots < i_m \leq n}$ denotes summation over all $\binom{n}{m}$ combinations of the m distinct elements $\{i_1, i_2, ..., i_m\}$ from $\{1, ..., n\}$, is called a U-statistic based on $\mathcal{F}$-associated random variables.

Hadjikyriakou (2010) established a strong law of large numbers for $\mathcal{F}$-demimartingales. An interesting extension is to focus on the connection between U-statistics based on

$\mathcal{F}$-associated random variables and $\mathcal{F}$-demimartingales and provide strong laws for this type of U-statistics. One can also introduce U-statistics based on $\mathcal{F}$-associated multidimensionally indexed random variables and multisample U-statistics on collections of $\mathcal{F}$-associated random variables. One can turn his/her attention on their connection with multidimensionally indexed $\mathcal{F}$-demimartingales, and could establish strong laws for also those types of U-statistics. Based on the results presented in the thesis, one can also study the distance between a U-statistic based on $\mathcal{F}$-associated random variables and a normal random variable using probability metrics. That result will be useful to provide central limit theorems for this type of U-statistics.

6.3 U-statistics based on m-negatively associated random variables

The concept of m-negatively associated random variables are natural extensions from negatively associated random variables. Motivated by the definition of negatively associated random variables, Hu et al. (2009) introduced the concept of m-negatively associated random variables as follows.

Definition 6.3.1. *Let $m \geq 1$ be a fixed integer. A sequence of random variables $\{X_n, n \geq 1\}$ is said to be m-negatively associated if for any $n \geq 2$ and any $i_1, ..., i_k$ such that $|i_k - i_j| \geq m$ for all $1 \leq k \neq j \leq n$, $X_{i_1}, ..., X_{i_n}$ are negatively associated.*

Note that when $m = 1$, the concept of m-negatively associated random variables equals to the concept of negatively associated random variables.

Hu et al. (2009) studied the Kolmogorov exponential inequality for m-negatively associated random variables. The complete convergence and complete moment convergence for weighted sums of m-negatively associated random variables were proposed by Wu et al. (2015). The moment inequalities for m-negatively associated random variables were proved by Shen et al. (2017). Mengge et al. (2019) presented the Spitzer type law of large numbers for the maximum of partial sums of m-negatively associated random variables.

Similar properties we used and proved in this thesis for negatively associated random variables are still valid for m-negatively associated random variables. For more details see Shen et al. (2017).

There are no studies in bibliography yet for U-statistics based on m-negatively associated random variables. A final possible direction is to study the asymptotic behavior of U-statistics based on random variables with this alternative type of dependance. Our goal is to prove useful inequalities such as moment and exponential inequalities for this type of U-statistics. To study the asymptotic normality, to obtain strong laws of large numbers and central limit theorems using probability metrics for U-statistics based on m-negatively associated random variables.

References

Arvesen, J. N. (1969). Jackknifing U-statistics. *Ann. Math. Statist.* **40**, 2076-2100.

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

Bickel, P. (1974). Edgeworth expansion in nonparametric statistics. *Ann. Statist.* **2**, 1-20.

Birkel, T. (1988). Moment bounds for associated sequences. *Ann. Probab.* **16**, 1184-1193.

Birkel, T. (1989). A note on the strong law of large numbers for positively dependent random variables. *Statist. Probab. Lett.* **7**, 17-20.

Boutsikas, M.V. and Vaggelatou, E. (2002). On the distance between convex-ordered random variables, with applications. *Adv. in Applied Probab.* **34**, 349-375.

Budsaba, K., Chen, P., Panishkan, K. and Volodin, A. (2009). Strong laws for certain types of U-statistics based on negatively associated random variables. *Siberian Adv. Math.* **19**, 225-232.

Bulinski, A.V. (1994). Inequalities for the Moments of Sums of Associated Multi-Indexed Variables. *Theor. of Probab. and Its Appli.* **38**, 342-349.

Bulinski, A. (1996). On the convergence rates in the CLT for positively and negatively dependent random fields. *Probab. Theor. and Math. Statist.* **40**, 3-14.

Callaert, H. and Janssen, P. (1978). The Berry-Esseen theorem for U-statistics. *Ann. Statist.* **6**, 417-421.

Chan, Y.K. and Wierman, J. (1977). On the Berry-Esseen theorem for U-statistics. *Ann. Probab.* **5**, 136-139.

Chow, Y.S. (1960). A martingale inequality and the law of large numbers. *Proc. Amer. Math. Soc.* **11**, 107-111.

Christofides, T.C. (2000). Maximal inequalities for demimartingales and a strong law of large numbers. *Statist. Probab. Lett.* **50**, 357-363.

Christofides, T.C. (2003). Maximal Inequalities for N-demimartingales. *Archives of Inequalities and Applications.* **50**, 397-408.

Christofides, T.C. (2004). U-statistics on associated random variables. *Statist. Plann. Inference* **119**, 1-15.

Christofides, T.C. and Hadjikyriakou, M. (2009). Exponential inequalities for N-demimartingales and negatively associated random variables. *Statist. Probab. Lett.* **79**, 2060-2065.

Christofides, T.C. and Hadjikyriakou, M. (2011). Maximal inequalities for multidimensionally indexed demimartingales and the Hajek-Renyi inequality for associated random variables. *The Australian J. of Math. Analysis and Appl.* **7**, 1-9.

Christofides, T.C. and Vaggelatou, E. (2004). A connection between supermodular ordering and positive/negative association. *Journal of Multivariate Analysis* **88**, 138-151.

Cox, J.T. and Grimmett, G. (1984). Central limit theorems for associated random variables and the percolation model. *Ann. Probab.* **12**, 514-528.

Denker, M. (1985). *Asymptotic Distribution Theory in Nonparametric Statistics*, Fr. Vieweg & Sohn, Braunschweig, Wiesbaden.

Denker, M. and Keller, G. (1983). On U-statistics and von Mises statistics for weakly dependent processes. *Z. Wahrsch. Verw. Gebiete* **64**, 505-522.

Denuit, M., Dhaene, J., Goovaerts, M. and Kaas R. (2005). *Actuarial Theory for Dependent Risks: Measures, Orders and Models*, John Wiley & Sons, New York.

Dewan, I. and Prakasa Rao, B.L.S. (2001). Associated sequences and related inference problems. *In: Shanbag, D.N., Rao, C.R. (Eds.), Handbook of Statistics,* **19**, Stochastic Processes: Theory and Methods. North-Holland, Amsterdam, pp. 693-728.

Dewan, Isha and Prakasa Rao, B. L. S. (2002). Central limit theorem for U-statistics of associated random variables. *Statist. Probab. Lett.* **57**, 9-15.

Dewan, I. and Prakasa Rao (2015). Corrigendum to 'Central limit theorem for U-statistics of associated random variables.' *Statist. Probab. Lett.* **106**, 147-148.

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

Filippova, A.A. (1962). Mises's theorem of the asympototic behavior of functionals of empirical distribution functions and its statistical applications. *Theory Probab. Appl.* **7** , 24-57.

Fraser, D. A. S. (1957). *Nonparametric Methods in Statistics.* John Wiley & Sons, New York.

Friedrich, K.O. (1989). A Berry-Esseen bound for functions of independent random variables. *Ann. Statist.* **17**, 170-183.

Garg, M. and Dewan, I. (2015). On asymptotic behavior of U-statistics based on associated random variables. *Statist. Probab. Lett.* **105**, 209-220.

Garg, M. and Dewan, I. (2018). On limiting distribution of U-statistics based on associated random variables. *Statist. Probab. Lett.* **132**, 7-16.

Ghorpade, Sudhir R. and Limaye, Balmohan V. (2010). *A Course in Multivariable Calculus and Analysis.* Springer.

Grams, W.F. and Serfling, R. J. (1973). Convergence rates for U-statistics and related statistics. *Ann. Statist.* **1**, 153-160.

Hadjikyriakou, M. (2010), Probability and moment inequalities for demimartingales and associated random variables, Ph.D. Dissertation, Department of Mathematics and Statistics, University of Cyprus, Nicosia.

Halmos, P. (1946). The theory of unbiased estimation, *Ann. Math. Statist.* **17**, 34-43.

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

Hoeffding, W. (1961). The strong law of large numbers for U-statistics. *Institute of Statistics Mimeo Series* **302**. University of North Carolina.

Hu, T., Chiang, C. and Taylor, R. (2009). On complete convergence for arrays of rowwise m-negatively associated random variables. *Nonlinear Anal.* **71**, 1075-1081.

Huang, W. and Zhang, L. (2006). Asymptotic normality for U-statistics of negatively associated random variables. *Statist. Probab. Lett.* **76**, 1125-1131.

Ioannides, D. A. and Roussas, G. G. (1999). Exponential inequality for associated random variables. *Statist. Probab. Lett.* **42**, 423-431.

Joag-Dev, K. and Proschan, F. (1983). Negative association of random variables, with applications. *Ann. of Statist.* **11**, 286-295.

Kim, T. S. and Kim, H. C. (2007). On the exponential inequality for negative dependent sequence. *Comm. of the Korean Math. Soc.* **22**, 315-321.

Korolyuk, V. S. and Borovskikh, Yu. V. (1989). *Theory of U-statistics*, Naukova dumka, Kiev, 1989; English transl., Kluwer, Dordrecht, 1993.

Kowalski J. and Xin M. T. (2008). *Modern Applied U-statistics*, John Wiley & Sons, New York.

Krewski, D. (1978). Jackknifing U-stastics in finite populations. *Comm. in Statist.* **7**, 1-12.

Lee, A. J. (1990). *U-Statistics*, Marcel Dekker Inc. New York.

Lehmann, E. L. (1996). Some concepts of dependence. *Ann. Math. Statist.* **43**, 1137-1153.

Lehmann, E. L. (1999). *Elements of Large Sample Theory*, Springer.

Majumdar, H. and Kumar Sen, P. (1978). Invariance principles for jackknifing U-statistics for finite population sampling and some applications. *Comm. in Statist.* **7**,1007-1025.

Matula, P. (1998). On the almost sure central limit theorem for associated random variables. *Probab. Math. Statist.* **18**, 411-416.

Mengge W. and Xuejun W. (2019). Some convergence properties for the maximum of partial sums of m-negatively associated random variables. *Rev. R. Acad. Cienc. Exactas Fns. Nat. Ser. A Mat. RACSAM 113* **3**, 2345-2358.

Nandi, H.K., Sen, P.K. (1963). On the properties of U-statistics when the observations are not independent II. *Cal. Statist. Assoc. Bull.* **12**, 125-143.

Newman, C.M. and Wright, A.L. (1982). Associated random variables and martingale inequalities. *Z. Wahrsch. Verw. Gebiete.* **59**, 361-371.

Newman, C.M. (1984). Asymptotic independence and limit theorems for positively and negatively dependent random variables. Inequalities in probability (Lincoln, Neb., 1982), 127-140, IMS Lecture Notes-Monograph Series, Vol. 5. *Institute of Math. Statist.,* Hayward, CA.

Nooghabi, H. J., and Azarnoosh, H. A. (2009). Exponential inequality for negatively associated random variables. *Stat. Papers* **50**, 419-428.

Oliveira, P. E. (2005). An exponential inequality for associated variables. *Statist. Probab. Lett.* **73**, 189-197.

Oliveira, P. E. (2012). *Asymptotics for Associated Random Variables*, Springer.

Petrov, V.V. (1995). *Limit Theorems of Probability Theory.* Oxford University Press.

Prakasa Rao, B.L.S. (2004). On some inequalities for N-demimartingales. *Indian Soc. Agricultural Statist.* **57**, 208-216.

Prakasa Rao, B.L.S. (2007). On some maximal inequalities for demisubmartingales and N-demisupermartingales. *J. Inequal. Pure Appl. Math.* **8**, Article 112, 17pp.

Prakasa Rao, B.L.S. (2009). Conditional independence, conditional mixing and conditional association. *Ann. Math. Statist.* **61**, 441-460.

Prakasa Rao, B.L.S. (2012). *Associated Sequences, Demimartingales and Nonparametric Inference.* Birkhauser Basel AG, Basel.

Quenouille, M. H. (1949). Approximate test of correlation in time-series. *J. Roy. Statist. Soc. Ser. B.* **11**, 68-84.

Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika* **43**, 353-360.

Rachev, S.T. (1991). *Probability Metrics and the Stability of Stochastic Models.* John Wiley & Sons, New York.

Roussas, G. (1993). Curve estimation in random fields of associated processes. *J. Nonparametr. Statist.* **2**, 215-224.

Roussas, G. (1999). Positive and negative dependence with some statistical applications. *In: Ghosh, S. (Ed.), Nonparametr. and Time Series. Marcel Dekker,* New York, pp. 757-788.

Sharakhmetov, Sh. (2004). The convergence of U-statistics and von Misses functionals in the mean metric. *Theor. Probab. and Math. Statist.* **68**, 167-172.

Shen, A. T., Zhang, Y., Xiao, B. Q. and Volodin, A. (2017). Moment inequalities for m-negatively associated random variables and their applications. *Stat. Papers* **58**, 911-928.

Sen, P. K. (1960). On Some Convergence Properties of U-statistics. *Cal. Statist. Assoc. Bull.* **10**, 1-18.

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. (1968). The Wilcoxon two-sample statistic on strongly mixing processes. *Ann. Math. Statist.* **39**, 1202-1209.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, John Wiley & Sons, New York.

Shaked, M. and Shaunthikumar, J.G. (2007): *Stochastic Orders and their Applications.* Academic Press, New York.

Shao, Q. M. (2000). A comparison theorem on moment inequalities between negatively associated and independent random variables. *J. Theoret. Probab.* **13**, 343-356.

Su, C., Zhao L. C. and Wang Y.B. (1997). Moment inequalities and weak convergence for negatively associated sequences. *Sci. China Ser. A* **40**, 172-182.

Sung, S. H. (2009). An exponential inequality for negatively associated random variables. *J. Inequal. Appl.* Art. ID 649427, 7 p.

Topkis, D. (1998). *Submodularity and Complementarity.* Princeton University Press, Princeton.

Tukey, J. W. (1958). Bias and confidence in not quite large samples (abstract). *Ann. Math. Statist.* **29**, 614-623.

van Zwet,W.R. (1984). A Berry-Esseen bound for symmetric statistics. *Z. Wahrsch. Verw. Gebiete.* **66**, 425-440.

von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. *Ann. Math. Statist.* **18**, 301-348.

Wang, Jianfeng (2004). Maximal inequalities for associated random variables and demimartingales. *Statist. Probab. Lett.* **66**, 347-354.

Wang, XueJun; Hu, Shuhe (2009). Maximal inequalities for demimartingales and their applications. *Sci. China Ser. A* **52**, 2207-2217.

Wang, Xuejun; Hu, Shuhe, Zhao, Ting and Yang, Wenzhi (2010). Doob's type inequality and strong law of large numbers for demimartingales. *Inequal. Appl.*, Art. ID 838301, 11 pp.

Wang, Xuejun; Hu, Shuhe, Prakasa Rao, B. L. S. and Yang, Wenzhi (2011). Maximal inequalities for N-demimartingale and strong law of large numbers. *Statist. Probab. Lett.* **81**, 1348-1353.

Wu, Y. F., Hu, T. C. and Volodin, A. (2015). Complete convergence and complete moment convergence for weighted sums of m-NA random variables. *J. Inequ. Appl.* **58**, Article ID 200, 14.

Xing, G., Yang, S., Liu, A., and Wang, X. (2009). A remark on the exponential inequality for negatively associated random variables. *Korean Stat. Soc.* **38**, 53-57.

Xing, G., and Shanchao, Y. (2010). An exponential inequality for strictly stationary and negatively associated random variables. *Comm. in Statist.* **39**, 340-349

Yamato, H., Toda, K. and Nomachi T. (2007). Jackknifing a convex combination of one-sample U-statistics. *Comm. in Statist.* **36**, 2223-2234.

Zolotarev, V.M. (1983): Probability metrics. *Theory Prob. Appl.* **28**, 278-302.