COMPARATIVE STUDY ON RATE OF CONVERGENCE TO NORMAL DISTRIBUTION FOR U-STATISTICS AND METHOD OF MOMENTS ESTIMATOR FOR POPULATION VARIANCE.

BY
JIMMY HAMBULO

A dissertation prepared in fulfilment of the requirements for the degree of Master of Science in Statistics

THE UNIVERSITY OF ZAMBIA

LUSAKA

2019

ABSTRACT

In this dissertation, we examine the small and large sample properties of the U-statistics (U_n) for an estimable parameter γ. For small sample, U_n is found to be MVUE and Sufficient. Large sample properties of U_n examined are Consistency and Asymptotic normality. We also compare the rate of convergence to normal distribution between the U-statistics and the method of moments estimator for the population variance. Berry-Esseen bound for U-statistics of order 2 is given and the results are used to find the rate of convergence to normal for the U-statistics for the population variance. Berry-Esseen bound for S^2 the method of moments estimator for the population variance is also given. Method of moments estimator is found to converge faster to normality than the U-statistics. Finally, we review the different proofs of Hoeffding's one sample U-statistics theorem as given in the period 1948-2012.

Keywords: U-statistics, method of moments estimator, Berry-Esseen bound, convergence rate, asymptotic normality.

COPYRIGHT

DECLARATION

The work described in this Master of Science (MSc) dissertation was carried out under the supervision of Dr. A. M Ngwengwe, Department of Mathematics and statistics, University of Zambia, Lusaka.

The MSc dissertation represents original work by the author and has not otherwise been submitted in any form for any degree or diploma to any other University. Where use has been made of the work of others it is duly acknowledged in the text.

Signed:

....

Jimmy Hambulo (Student)

....

Dr. A. M. Ngwengwe (Supervisor)

APPROVAL

This dissertation of Jimmy Hambulo has been approved as fulfilling the requirements or partial fulfillment of the requirements for the award of Master of Science in Statistics by the University of Zambia.

Examiner 1...................... Signature .... Date ....

Examiner 2....................... Signature ....Date ....

Examiner 3....................... Signature .... Date ....

Chairperson Board of Examiners............................ Signature .... Date ....

Supervisor......................... Signature .... Date ....

DEDICATION

I dedicate this work to my mother Racheal Siakachite, my wife Serah, my children Joseph,Rabecca and Twalumba my siblings Linah, Milimo, Salome, Grace and Mwansa, my cousins Lynette and Mwala for the support and words of courage they have rendered.

ACKNOWLEDGMENTS

First of all, I thank God Almight for this precious opportunity.

Secondly, I thank my supervisor **Dr. A.M. Ngwengwe** for the wonderful guidance, tireless support and patience he has shown during my entire study.

Thirdly, I thank the University Of Zambia for the financial support.

Finally, I thank the Department of Mathematics and Statistics at the University of Zambia for the wonderful support I have received through out my study. Also I thank the HOD for Mathematics Department at Kabulonga Girls Secondary School Mr Banda O.B. for his support during my research.

TABLE OF CONTENTS

Index of Notation

Below is a list of symbols that will be frequently used and a brief indication of their meaning.

$E(X)$	Expected value of random variable X
$Var(X)$	Variance of random variable X
$Cov(X,Y)$	Covariance of random variables X and Y
sup	Supremum (least upper bound)
$\bar{X}$	Mean of random sample $X_1, ..., X_n$
μ	Population mean
iid	Independent and identically distributed
iff	If and only if
S^2	Sample variance
σ^2	Population variance
inf	Infimum (greatest lower bound)
A^c	Complement of a set A
μ_k	The k^{th} central moment
M_k	The k^{th} sample moment
$X_{(m)}$	M^{th} order statistic
CLT	Central Limit Theorem
$MVUE$	Minimum Variance Unbiased Estimator
$\mathbb{R}$	The set of real numbers

CHAPTER 1 : INTRODUCTION

This chapter is divided into nine sections. In section one we have the introduction, in section two we have the statement of the problem, aim of the study is in section three, objectives of the study in section four and in section five we have the research questions. Significance of the study is in section six, Literature review in section seven, Methodology in section eight and the organisation of the study in section nine.

1.1. Introduction

Assume that we have designed and conducted a suitable experiment and collected data $X_1, ..., X_n$ where n, the sample size, is fixed and known. Statistical inference looks at the methods of obtaining information on the true value of the parameter from data underlying the experiment. The main topics in statistical inference are estimation and hypothesis testing.

In estimation we use a random sample $X_1, ..., X_n$ to find an estimator for some population numerical characteristic, the parameter. An estimator is a formula that tells us how to calculate an estimate based on the information in the sample. An estimator of the parameter (measure of the population) is said to be unbiased if the expectation of the estimator is equal to the parameter.

There is a class of unbiased estimators called U-statistics which are non-parametric and were mainly developed by a statistician called Wassily Hoeffding (1948). The letter U in U-statistics stands for unbiased. In 1948, Hoeffding stated and proved a theorem known as Hoeffding's one sample U-statistics theorem.

In 1990, Lee gave an alternative proof of the Hoeffding's one sample U-statistics theorem. Ferguson(1996), gave another alternative proof of Hoeffding's one sample U-statistics theorem. In 2012, Beutner and Zahle gave yet another alternative proof of Hoeffding's one sample U-statistics theorem. Carlet and Janssen (1978), discussed the rate of convergence to normal for the U-statistics for an estimable parameter of degree 2. The rate of convergence to the normal distribution of the U-statistics and the method of moments estimator for the population variance have not being compared.

1.2. Statement of the Problem

Since the proof of one sample U-statistics theorem by Hoeffding (1948), some researchers' have used different approaches in proving the theorem. Is there anything common in the different proofs of the theorem? We want to understand the basis of the asymptotic normality of the U- statistics. The theory of U- statistics can be used to study the efficiency of many estimators including the sample variance (Ahmad, 1979). However, the rate of convergence to normality of the U-statistics and the method of moments estimator for the population variance have not been compared. We also investigate if the U-statistics for an estimable parameter γ have qualities of a good estimator.

1.3. Aim of the Study

The aim is to find a common thread in the many proofs of the Hoeffding's one sample U-statistics theorem and explore its characteristics with respect to asymptotic normality and rate of convergence when compared with the method of moments estimator.

1.4. Research Objectives

The following were the objectives in the study.

i Examine the small and large sample properties of the U- statistics for an estimable parameter γ of degree k.

ii Compare the rate of convergence to normal between the U-statistics for the population variance and the method of moments estimator for the population variance.

iii Review the different proofs of Hoeffding's one sample U- statistics theorem.

1.5. Research Questions

The following were the research questions in the study.

i Which estimator is more efficient between the U- statistics for the population variance and the methods of moments estimator for the population variance?

ii What common attributes do the different proofs of Hoeffding's one sample U- statistics theorem share?

1.6. Significance of the study

The study provided a method of calculating the rate of convergence to the normal distribution for the U-statistics and method of moments estimator for the population variance. This study provides ideas on how to construct an alternative proof of Hoeffding's one sample U-statistic theorem.

1.7. Literature Review

Since the pioneering work of Hoeffding in 1948, the U- statistics have been an active research field in statistics due to their wide range of application. Hoeffding (1948) established some fundamental properties of the U-statistics for some estimable parameter of degree k. Ahmad (1981) discussed the problem of rate of convergence in the central limit theorem for the U- statistics for an estimable parameter of degree 2. Van Zwet (1984) investigated the rate of convergence to normal for a symmetric statistics and extended the results to the U-statistics with bounded kernel. Lee (1990) discussed some small properties of U-statistics and gave an alternative proof of Hoeffding's one sample U-statistic theorem. Ferguson (1996) gave another alternative proof of the theorem. Lehmann (1999) showed that the method of moments estimator for the population variance is asymptotically normally distributed. Bentkus, Jing and Zhou (2009) showed that the rate of convergence to normality of the U-statistics for an estimable parameter of degree k can simply be expressed as the rate of convergence to normal for the linear part plus a correction term in the decomposition of the definition of U- statistics. Beutner and Zahle (2012) gave yet another alternative proof of the theorem.

1.8. Methodology

To examine the small and large sample properties of the U-statistics for an estimable parameter γ of degree k, we discussed its unbiasedness, sufficiency, efficiency, consistency and asymptotic normality by reviewing a book by Fraser (1957) and articles done by Imbens (1992), Bentkus, Jing and Zhou (2009). To compare the rate of convergence to normality of the U-statistics for the population variance(σ^2) and the method of moments estimator(S^2) for the population variance(σ^2), we calculated the Berry-Esseen bounds for both estimators by reviewing articles done by Ahmad (1981) , Callaert and Janssen (1978), Grams and Serfling (1973), Van Zwet (1984), Hsing and Hu (2004) and Borovskikh and Koroljuk (1989) and books by Feller (1971), Lehmann (1999) and Berger and Casella (2002). To review the different proofs of Hoeffding's one sample U-statistics theorem, we examined alternative proofs of Hoeffding's one sample U-statistics theorem as given in the period 1948-2012 by reviewing articles done by Beutner and Zahle (2012), Hoeffding (1948) and books by Ferguson (1996) and Lee (1990).

1.9. Organisation of the Study

The dissertation is organised as follows: In chapter one we have the introduction. In chapter 2 we present results that will be used in examining the properties of U-statistics for an estimable parameter γ of degree k, and in comparing the rate of convergence to normal between the U-statistics and the method of moments estimator for the population variance. We also present some results used in comparing the four different proofs of Hoeffding's one sample U-statistics theorem as given in the period 1948-2012. In chapter 3 we examine the small and large sample properties of the U-statistics for an estimable parameter γ. In chapter 4, we examine the rate of convergence to normal for the U-statistics for the population variance and the method of moments estimator for the population variance. In chapter 5, we examine the different proofs of Hoeffding's one sample U-statistics theorem as given Hoeffding (1948), Lee (1990), Ferguson (1996) and Beutner and Zahle (2012). The discussion and conclusion is made in chapter 6.

CHAPTER 2 : PROPERTIES OF ESTIMATORS AND VARIANCE OF U-STATISTICS

Here, we present some concepts regarding some properties of random variables that are useful in this dissertation. We will also give the definition of a U-statistic and develop the general expression of a variance of U-statistics for an estimable parameter γ of degree k. Proofs of well known results which can easily be accessed from standard literature will generally be omitted. Only those proofs for which the reader may not have easy access to the appropriate literature will be given.

2.1. Some properties of estimators

We begin by giving a definition for a minimum variance unbiased estimator.

Definition 2.1.1. An estimator $T(x)$ is said to be a minimum variance unbiased estimator (MVUE) of the parameter γ if

(i) It is an unbiased estimator of γ

(ii) Among all unbiased estimators of γ, it has the smallest variance.

Definition 2.1.2. Let $X = (X_1, ..., X_n)$ be a random sample. A statistic $T(X)$ is a sufficient statistic for a parameter γ if the conditional distribution of the sample X given the value of $T(X)$ does not depend on γ.

Corollary 2.1.3 (Casella (2002))**.** *Let $X_1, ..., X_n$ be a random sample. If $T(X)$ is a sufficient statistic for θ then so is any one - one function of $T(X)$.*

Definition 2.1.4. Let $T(X)$ be a statistic for a family of distributions $f(t|\theta)$. If for every function $h(.)$, $E_\theta[h(T)] = 0$ for all θ implies that $P_\theta(h(T) = 0) = 1$ for all θ. Then $T(X)$ is called a complete statistic.

Theorem 2.1.5 (Lehmann-Scheffe)**.** If $T(X)$ is a complete sufficient statistic for the model $\{f_\gamma(x); \gamma \in \mathcal{F}\}$ and $U(x)$ is unbiased estimator of γ, then $E(U|T)$ is the MVUE of γ.

Definition 2.1.6.

(i) A sequence of random variables $\{Y_n\}_{n=1}^{\infty}$ is said to converge in probability to a constant c, if for any $\epsilon > 0$

$$\lim_{n\to\infty} P\{|Y_n - c| < \epsilon\} = 1.$$

(ii) If $\{Y_n\}_{n=1}^{\infty}$ is a sequence of estimators of a parameter γ, then Y_n is termed a consistent estimator of γ if Y_n converges in probability to γ, for every value of γ in the parameter space.

Result 2.1.7. *(Fraser (1957)). Let $T_n(X)$ be an estimator for γ based on a random sample $X_1, X_2, ..., X_n$. Then $T_n(X)$ is a consistent estimator for γ if*

(i) $\lim_{n\to\infty} E(T_n(X)) > 0$

(ii) $\lim_{n\to\infty} Var(T_n(X)) = 0$

Definition 2.1.8. If $\{Y_n\}_{n=1}^{\infty}$ is a sequence of random variables with corresponding sequence of cumulative distribution (cdf) $F_n(y)$, then the sequence $\{Y_n\}_{n=1}^{\infty}$ converges in distribution to a random variable Y with cdf $F(y)$ denoted by $Y_n \to^d Y$ if

$$\lim_{n\to\infty} F_n(y) = F(y)$$

Definition 2.1.9. The sequence $\{Y_n\}_{n=1}^{\infty}$ of random variables is said to converge in quadratic mean to a constant c if

$$\lim_{n\to\infty} E\{(Y_n - c)^2\} = 0$$

.

Lemma 2.1.10 (Chebyshev's inequality)**.** *Let X be a random variable with finite expected value μ and finite non-zero variance σ^2. Then for any $\epsilon > 0$*

$$P(|X - \mu| \geq \epsilon) \leq \frac{var(X)}{\epsilon^2}$$

PROOF

The result holds quiet generally, but we shall prove only for the case that the distribution of $Z = X - \mu$ has a density $f(z)$. Then

$$\begin{aligned} Var(X) = Var(X-\mu) = Var(Z) &= E(Z^2) \\ &= \int z^2 f(z)dz \\ &= \int_{|z|\geq\epsilon} z^2 f(z)dz + \int_{|z|<\epsilon} z^2 f(z)dz \\ &\geq \int_{|z|\geq\epsilon} z^2 f(z)dz \geq \epsilon^2 \int_{|z|\geq\epsilon} f(z)dz \\ &= \epsilon^2 P(|Z| \geq \epsilon). \end{aligned}$$

If the distribution of Z is discrete, the integrals are replaced by sums.

Theorem 2.1.11. Convergence in quadratic mean implies convergence in probability.

PROOF

Assume that $\{X_n\}$ converges in quadratic mean to c, then by Chebyshev's inequality

$$P\{|X_n - c| \geq \epsilon\} \leq \frac{E\{(X_n - c)^2\}}{\epsilon^2}$$

Taking the limit as $n \to \infty$ on both sides we get

$$\lim_{n\to\infty} P\{|X_n - c| \geq \epsilon\} \leq \lim_{n\to\infty} \frac{E\{(X_n - c)^2\}}{\epsilon^2} = \frac{1}{\epsilon^2} \lim_{n\to\infty} E\{(X_n - c)^2\} = 0$$

Since $\{X_n\}$ converges in quadratic mean to c. Therefore

$$\lim_{n\to\infty} P\{|X_n - c| \geq \epsilon\} = \lim_{n\to\infty} \{1 - P\{|X_n - c| < \epsilon\} = 0$$

$$\Rightarrow \lim_{n\to\infty} P\{|X_n - c| < \epsilon\} = 1$$

Theorem 2.1.12 (Weak law of large numbers)**.** Let $X_1, ..., X_n$ be random variables with mean μ and variance $\sigma^2 < \infty$. Then the average $\bar{X} = (X_1 + ... + X_n)/n$ converges in probability to μ.

PROOF

We have that

$$E(\bar{X} - \mu)^2 = Var(\bar{X}) = \frac{\sigma^2}{n} \to 0 \quad \text{as} \quad n \to \infty$$

thus $\bar{X}$ converges in quadratic mean to μ, hence $\bar{X}$ converges in probability to μ since convergence in quadratic mean implies convergence in probability.

Definition 2.1.13. Let X be a random variable with probability density function f. The characteristic function of X is defined by

$$\varphi_X(t) = E(e^{itX}) = \int_{-\infty}^{+\infty} e^{itX} f(x)dx$$

Result 2.1.14. *For $n = 1, 2, ...,$ and $t > 0$*

$$\left| e^{it} - 1 - \frac{it}{1} - ... - \frac{(it)^{n-1}}{(n-1)!} \right| \leq \frac{t^n}{n!} \tag{2.1}$$

PROOF

Denote the expression within the absolute value sign by $\rho_n(t)$. Then

$$\rho_1(t) = e^{it} - 1 = i \int_0^t e^{ix} dx$$

so that $|\rho_1(t)| \leq t$

$$\rho_2(t) = e^{it} - 1 - it = i \int_0^t (e^{ix} - 1) dx = i \int_0^t \rho_1(x) dx$$

Therefore for $n > 1$

$$\rho_n(t) = i \int_0^t \rho_{n-1}(x) dx$$

and (2.1.1) now follows by mathematical induction.

Theorem 2.1.15. (Central Limit Theorem) Let $X_1, X_2, ...X_n$ be iid with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 < \infty$. Then

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \rightarrow^d N(0, 1)$$

or equivalently

$$\sqrt{n}(\bar{X} - \mu) \rightarrow^d N(0, \sigma^2)$$

Theorem 2.1.16 (Slutsky's Theorem)**.** Let $\{W_n\}$ be a sequence of random variables with limiting distribution $F(w)$. Let $\{V_n\}$ denote a sequence of random variables that converges in probability to a constant c, then

i $W_n + V_n$ and $W_n + c$ have the same limiting distribution.

ii $V_n W_n$ and cW_n have the same limiting distribution.

iii $\frac{W_n}{V_n}$ and $\frac{W_n}{c}$ have the same limiting distribution provided $c \neq 0$

where W_n and V_n need not be independent.

PROOF

First we show that $W_n + V_n \to^d W_n + c$, For any $\epsilon > 0$

$$P(W_n + V_n \le x) \le P(\{W_n + V_n \le x\} \cap \{|V_n - c| \le \epsilon\} \cup |V_n - c| > \epsilon))$$

$$\le P(W_n \le x - c + \epsilon) + P(|V_n - c| > \epsilon)$$

$$\Rightarrow lim_n sup F_{W_n+V_n}(x) \le lim_n sup F_{W_n}(x - c + \epsilon) \le F_{W_n}(x - c + \epsilon)$$

On the other hand,

$$P(W_n+V_n > x) = P(W_n+V_n > x, |V_n-c| \le \epsilon)+P(|V_n-c| > \epsilon) \le P(W_n > x-c-\epsilon)+P(|V_n-c| > \epsilon)$$

.

$$\Rightarrow lim_n sup(1 - F_{W_n+V_n}(x)) \le lim_n sup P(W_n > x - c - \epsilon)$$

$$\le lim_n sup P(W_n \ge x - c - 2\epsilon) \le (1 - F_W(x - c - 2\epsilon))$$

.

$$\Rightarrow F_{W_n}(x - c - 2\epsilon) \le lim_n inf F_{W_n+V_n}(x) \le limsup_n F_{W_n+V_n}(x) \le F_{W_n}(x + c + \epsilon)$$

$$\Rightarrow F_{W_n+c}(x) \le lim_n inf F_{W_n+V_n}(x) \le lim_n sup F_{W_n+V_n}(x) \le F_{W_n+c}(x)$$

We now prove (ii)

$$P(|(V_n - c)W_n| > \epsilon) \le P(|V_n - c| > \epsilon^2) + P(|V_n - c| \le \epsilon^2, |W_n| > \frac{1}{\epsilon})$$

$$\Rightarrow limsup_n P(|(V_n-c)W_n| > \epsilon) \le lim_n sup P(|V_n-c| > \epsilon^2)+lim_n sup P(|W_n| \ge \frac{1}{2\epsilon}) \to P(|W| \ge \frac{1}{2\epsilon})$$

which implies that $(V_n - c)W_n \to^p 0$. So that $cW_n \to^d cW \Rightarrow V_n W_n \to^d cW$

proof of (iii) follows from the proof of (ii) if we let $Z_n = \frac{1}{V_n}$ so that $W_n Z_n = \frac{W_n}{V_n}$.

Theorem 2.1.17. If the sequence of random variables $\{V_n\}$ has asymptotic distribution with cdf $F(v)$ and if $\{W_n\}$ is a sequence of random variables such that $\{W_n - V_n\}$ converges in probability to zero, then the limiting distribution of $\{W_n\}$ is also given by the cdf $F(v)$.

PROOF

Let $G(w)$ be the distribution function of W_n [see Lehmann (1999)], now we define the following sets

$$A_n = \{z : |W_n(z) - V_n(z)| < \epsilon\}$$

$$B_n = \{z : W_n(z) < v, v \in R\}$$

$$C_n = \{z : V_n(z) < v + \epsilon\}$$

$$D_n = \{z : V_n(z) \geq v - \epsilon\}$$

then $G(v) = P(B_n) = P(B_n \cap A_n) + P(B_n \cap A_n^c)$
$1 - G(v) = P(B_n^c) = P(B_n^c \cap A_n) + P(B_n^c \cap A_n^c)$ now

$$B_n \cap A_n = \{z : W_n(z) < v, W_n(z) - \epsilon < V_n(z) < W_n(z) + \epsilon\} \subset C_n$$

$$B_n^c \cap A_n = \{z : W_n(z) \geq v, W_n(z) - \epsilon < V_n(z) < W_n(z) + \epsilon\} \subset D_n$$

thus

$$F(v) \leq P(C_n) + P(A_n^c) = F(v + \epsilon) + P(A_n^c)$$

so that

$$F(v) \leq F(v + \epsilon) + P(A_n^c)$$

$$1 - F(v) \leq P(D_n) + P(A_n^c) = 1 - G(v - \epsilon) + P(A_n^c)$$

so that

$$1 - F(v) \leq 1 - G(v - \epsilon) + P(A_n^c)$$

$$\Rightarrow G(v - \epsilon) - P(A_n^c) \leq F(v)$$

$$\Rightarrow G(v - \epsilon) - P(A_n^c) \leq F(v) \leq G(v + \epsilon) + P(A_n^c)$$

now since $\{W_n - V_n\}$ converges in probability to zero then

$$\lim_{n \to \infty} P(A_n^c) = 0$$

since convergence in quadratic mean implies convergence in probability. Thus

$$G(v - \epsilon) \leq \liminf_{n \to \infty} F(v) \leq \limsup_{n \to \infty} F(v) \leq G(v + \epsilon)$$

for all continuity points of F. Thus $\{W_n\}$ converges in distribution to a variable with cdf $F(v)$

We shall now give a definition of a U-statistics for an estimable parameter γ of degree k, but first we define an estimable parameter.

Definition 2.1.18. A parameter γ is said to be an estimable parameter of degree k within $\mathcal{F}$, a family of probability measure on an arbitrary measurable space, if k is the smallest sample size for which there exists a function $h(X_1, X_2, ..., X_k)$ such that $E\left(h(X_1, X_2, ..., X_k)\right) = \gamma$ for all distributions in $\mathcal{F}$, where $X_1, X_2, ..., X_k$ are iid from $\mathcal{F}$ and $h(.)$ is a statistic and does not depend on $\mathcal{F}$.

Note 2.1.19. The function $h(.)$ is called the kernel of the parameter γ. Without loss of generality, we can assume that the kernel $h(.)$ is symmetric in its arguments. ie $h(X_1, ..., X_k) = h(X_{\alpha_1}, ..., X_{\alpha_k})$ where $(\alpha_1, ..., \alpha_k)$ is any permutation of 1,2,...,k.

For any kernel $h(.)$ we can create one that is symmetric in its argument by forming

$$h^*(X_1, X_2, ..., X_k) = \frac{1}{k!} \sum_{i \in A} h(X_{i_1}, X_{i_2}, ..., X_{i_k})$$

where the sum is over A, a set of all the k! permutations of the integers $\{1, 2, ..., k\}$

Definition 2.1.20. Let $X_1, X_2, ..., X_n$ be a random sample from a distribution $\mathcal{F}$, and let $h(X_1, X_2, ..., X_k)$, where $k \leq n$ be a real- valued measurable function. Then a U-statistic for an estimable parameter γ with a symmetric kernel $h(X_{i_1}, X_{i_2}, ..., X_{i_k})$ is given by

$$U(X_1, X_2, ..., X_n) = U_n = \frac{1}{\binom{n}{k}} \sum_{i \in I} h(X_{i_1}, X_{i_2}, ..., X_{i_k})$$

where I is the set of $\binom{n}{k}$ unordered subsets of k different integers $1 \leq i_1 < i_2 < ... < i_k \leq n$. chosen without replacement from the set $\{1, 2, ..., n\}$.

Example 2.1.21. Let $X_1, X_2, ..., X_n$ be a random sample from a distribution with finite variance σ^2, where $\sigma^2 = E(X_1^2 - X_1X_2)$. Thus the variance is an estimable parameter of degree 2 with kernel $h(X_1, X_2) = X_1^2 - X_1X_2$ so that the symmetric kernel is

$$\begin{aligned} h^* &= \frac{1}{2}(X_1 - X_2)^2 \\ &= \frac{1}{2}(X_1^2 - 2X_1X_2 + X_2^2) \\ &= \frac{1}{2}(\sum_{i=1}^{2} X_i^2 - 2X_1X_2) \end{aligned}$$

The U-statistics for σ^2 is thus given by

$$U_2(X_1, ..., X_n) = \frac{1}{\binom{n}{2}} \sum_{\beta \in B} \frac{1}{2} (\sum_{i=1}^{2} X_{\beta_i}^2 - 2X_{\beta_i} X_{\beta_j}) \quad \text{for} \quad i < j \quad \text{where} \quad \beta = (\beta_i, \beta_j)$$
$$= \frac{2}{n(n-1)} . \frac{1}{2} \sum_{\beta \in B} (\sum_{i=1}^{2} X_{\beta_i}^2 - 2X_{\beta_i} X_{\beta_j}) \quad \text{for} \quad i < j$$
$$= \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 = S^2 \tag{2.2}$$

2.2. Variance of U-Statistics

In this section we develop the general expression of the variance of a U-statistics U_n.

Let $X_1, ..., X_n$ be iid random variables and

$$U_n = U(X_1, X_2, ..., X_n) = \frac{1}{\binom{n}{k}} \sum_{i \in I} h(X_{i_1}, X_{i_2}, ..., X_{i_k})$$

where $h(x_1, ..., x_k)$ is symmetric in its arguments and the sum is over all permutations $(i_1, ..., i_k)$ of k different integers $1 \leq i_1 < i_2 < ... < i_k \leq n$. We have that $E\{h(X_1, ...X_k)\} = \gamma$, now if $(i_1, ..., i_k)$ and $(j_1, ..., j_k)$ are two sets of k different integers where $1 \leq i_\alpha, j_\alpha \leq n$ for $\alpha = 1, 2, ..., k$ and c is the number of integers common to the two sets. Then by symmetry of h(.) the covariance of $h(X_{i_1}, ..., X_{i_k})$ and $h(X_{j_1}, ..., X_{j_k})$ is

$$\xi_c = E\{h(X_{i_1}, ..., X_{i_k}) h(X_{j_1}, ..., X_{j_k})\} - E\{h(X_{i_1}, ..., X_{i_k})\} E\{h(X_{j_1}, ..., X_{j_k})\}$$
$$\xi_c = E\{h(X_{i_1}, ..., X_{i_k}) h(X_{j_1}, ..., X_{j_k})\} - \gamma^2 \tag{2.3}$$

Now

$$
\begin{aligned}
Var(U_n) &= Var\left[\binom{n}{k}^{-1}\sum_{i\in I} h(X_{i_1}, X_{i_2}, ..., X_{i_k})\right] \\
&= E\left\{\left[\frac{1}{\binom{n}{k}}\sum_{i\in I}(h(X_{i_1}, ..., X_{i_k})) - \gamma\right]^2\right\} \\
&= \frac{1}{[\binom{n}{k}]^2}\sum_{i\in I,}\sum_{j\in J} E\{h(X_{i_1}, ..., X_{i_k}) - \gamma)(h(X_{j_1}, ..., X_{j_k}) - \gamma)\} \\
&= \binom{n}{k}^{-2}\sum_{i\in I,}\sum_{j\in J} Cov\left[h(X_{i_1}, X_{i_2}, ..., X_{i_k}), h(X_{j_1}, X_{j_2}, ..., X_{j_k})\right]
\end{aligned}
$$

If there are no common variables in $X_{i_1}, X_{i_2}, ..., X_{i_k}$ and $X_{j_1}, X_{j_2}, ..., X_{j_k}$, then $Cov\left[h(X_{i_1}, X_{i_2}, ..., X_{i_k}), h(X_{j_1}, X_{j_2}, ..., X_{j_k})\right] = 0$. The number of pairs of k-tuples $(i_1, ..., i_k)$ and $(j_1, ..., j_k)$ having exactly c element in common is

$$
\binom{n}{k}\binom{k}{c}\binom{n-k}{k-c}
$$

because there are $\binom{n}{k}$ ways of choosing $i_1, ..., i_k$ and then $\binom{k}{c}$ ways of choosing a subset of size c from them to put in $j_1, j_2, ..., j_k$, and finally $\binom{n-k}{k-c}$ ways of choosing the remaining $k-c$ elements of $j_1, ..., j_k$ from the remaining numbers. Therefore,

$$
Var(U_n) = \binom{n}{k}^{-2}\sum_{c=0}^{k}\binom{k}{c}\binom{n}{k}\binom{n-k}{k-c}\xi_c
$$

$$
= \binom{n}{k}^{-1}\sum_{c=0}^{k}\binom{k}{c}\binom{n-k}{k-c}\xi_c
$$

Note that c is the number of elements common to the pairs of k-tuples $(i_1, \dots, i_k)$ and $(j_1, \dots, j_k)$. So c can take values 0,1,2,. . . ,k. If there are no common elements between the pairs of k-tuples $(i_1, \dots, i_k)$ and $(j_1, \dots, j_k)$ then $\xi_0 = 0$.

Example 2.2.1. Consider $U_1(X_1, ..., X_n)$ as a U-statistic for the mean, where h(x)=x and k=1

$$Var(U_1(X_1, ..., X_n)) = \frac{1}{\binom{n}{1}} \sum_{c=0}^{1} \binom{1}{c} \binom{n-1}{1-c} \xi_c = \frac{1}{n} \binom{1}{1} \binom{n-1}{0} \xi_1 = \frac{\xi_1}{n}.$$

Now $\xi_1 = E(h(X_{\beta_i})h(X_{\beta_j})) - \gamma^2$, since k=1 and c=1 we have $E(X^2) - \gamma^2 = Var(X)$, therefore

$$Var(U_1(X_1, ..., X_n)) = \frac{\sigma^2}{n}$$

CHAPTER 3 : PROPERTIES OF U-STATISTICS

In this chapter we examine the small and large sample properties of the U-statistics for an estimable parameter γ. Small and large sample properties examined are; Unbiasedness, Sufficiency, Efficiency, Consistency and Asymptotic normality.

3.1. Small sample Properties of U-Statistics

Small sample properties to be examined are unbiasedness, sufficiency and efficiency. Fortunately by construction, the U-statistics for an estimable parameter γ is unbiased.

3.1.1 Sufficiency of U-statistics

Suppose $X_1, X_2, ..., X_n$ are iid random variables, let $Y = (X_{(1)}, ..., X_{(n)})$ be the set of order statistic. Fraser (1957) established that $Y = (X_{(1)}, ..., X_{(n)})$ is sufficient for a given parameter γ. Fraser (1957) stated that if T is a symmetric function then it can be written as a function of the order statistic Y. Now since U_n is a symmetric function, then it can be written as a function of the order statistic $Y = (X_{(1)}, ..., X_{(n)})$, since the kernel $h(X_1, ..., X_k)$ is unbiased for γ then the conditional expectation of $h(X_1, ..., X_k)$ given $Y = (X_{(1)}, ..., X_{(n)})$ is given by

$$E[h(X_1, .., X_k)|Y] = \frac{1}{\binom{n}{k}} \sum_{i \in I} h(X_{i_1}, ..., X_{i_k})$$

since given Y there are $\binom{n}{k}$ equally likely choices of $(i_1, i_2, ..., i_k)$ for the set of integers $\{1 \leq i_1 \leq i_2 \leq ... \leq i_k \leq n\}$ each with equal probability $\frac{1}{\binom{n}{k}}$. Thus the conditional expectation is the sum of all such possible values divided by $\binom{n}{k}$ which is nothing but the U-statistics given in definition 2.1.20. Thus the U-statistics U_n is a function of the sufficient statistics, therefore, U_n is sufficient for γ by corollary 2.1.3.

3.1.2 Efficiency of U-statistics

Here we show that the U-statistics (U_n) is a minimum variance unbiased estimator of γ. Let $X_1, ..., X_n$ be iid random variables with continuous distribution function F. Suppose $\mathcal{F}$ is the class of all continuous distribution functions, we have already shown that the full set of order statistics Y is sufficient for γ_F where $F \in \mathcal{F}$. According to Fraser (1956), Y is a complete sufficient statistics for $\mathcal{F}$. Therefore, since U_n is unbiased and is based on a complete sufficient statistic Y, by Lehmann-Scheffe's theorem (theorem 2.1.5), U_n is the minimum variance unbiased estimator(MVUE) of γ.

3.2. Large sample Properties of U-statistics

The large sample (asymptotic) properties of estimators deal with the behavior of the estimator as the sample size gets large, in this sense, an estimator which is calculated for different sample sizes can be understood as a sequence of random variables indexed by the sample size(for example U_n). Two relevant aspects to analyze in this sequence are convergence in probability and convergence in distribution.

Large sample properties are analyzed by consistency and asymptotic normality. In consistency we try to find the limiting value of a statistic. Consistency gives a limiting value of the statistic. A consistent estimator having asymptotic normality is known as Consistent Asymptotic Normal (CAN). Here we show that the U-statistics U_n is a CAN estimator.

Theorem 3.2.1. If U_n is a U-statistics for an estimable parameter γ of degree k and $0 < \xi_1 < \infty$, where $\xi_1 = E\{h(X_1, X_2, ..., X_k)h(X_1, X_{k+1}, ..., X_{2k-1})\} - \gamma^2$. Then

$$\lim_{n\to\infty} Var(\sqrt{n}U_n) = k^2\xi_1$$

PROOF

$$\begin{aligned}
Var(\sqrt{n}U_n) &= nVar(U_n) \\
&= n\binom{n}{k}^{-1}\sum_{c=1}^{k}\binom{k}{c}\binom{n-k}{k-c}\xi_c \\
&= \sum_{c=1}^{k}\frac{n\binom{k}{c}\binom{n-k}{k-c}\xi_c}{\binom{n}{k}} \\
&= \sum_{c=1}^{k}\frac{nk!(n-k)!k!(n-k)!\xi_c}{n!(k-c)!c!(k-c)!(n-2k+c)!}
\end{aligned} \tag{3.1}$$

Let $L_c = \left(\frac{k!}{(k-c)!}\right)^2$ then

$$
\begin{aligned}
Var(\sqrt{n}U_n) &= \sum_{c=1}^{k} \frac{n(n-k)!(n-k)!L_c\xi_c}{n!c!(n-2k+c)!} \\
&= \sum_{c=1}^{k} \frac{(n-k)!L_c\xi_c}{c!(n-1)(n-2)....(n-k+1)(n-2k+c)!} \\
&= \sum_{c=1}^{k} \frac{(n-k)(n-k-1)...(n-2k+c+1)(n-2k+c)!L_c\xi_c}{c!(n-1)(n-2)....(n-k+1)(n-2k+c)!} \\
&= \sum_{c=1}^{k} \frac{(n-k)(n-k-1)...(n-2k+c+1)L_c\xi_c}{c!(n-1)(n-2)....(n-k+1)} \qquad (3.2)
\end{aligned}
$$

Now, from (3.2.1) we have that $n\binom{n-k}{k-c} = \frac{n(n-k)(n-k-1)...(n-2k+c+1)}{(k-c)!}$ and $\binom{n}{k} = \frac{n(n-1)(n-2)...(n-k+1)}{k!}$. Thus there are $k-c+1$ terms involving n in the numerator and k such terms in the denominator in (3.2.2). If $c=1$, then

$$\lim_{n\to\infty} \frac{(n-k)(n-k-1)...(n-2k+c+1)}{c!(n-1)(n-2)....(n-k+1)} \to 1.$$

If $c > 1$, then

$$\lim_{n\to\infty} \frac{(n-k)(n-k-1)...(n-2k+c+1)}{c!(n-1)(n-2)....(n-k+1)} \to 0.$$

Thus we get

$$
\begin{aligned}
\lim_{n\to\infty} Var(\sqrt{n}U_n) &= L_1\xi_1 \\
&= \frac{k!k!\xi_1}{(k-1)!(k-1)!} \\
&= k^2\xi_1
\end{aligned}
$$

Result 3.2.2. *The U_n estimator of γ is consistent.*

PROOF

Recall that two sufficient conditions for consistency is that expectation converges to a non zero finite quantity and variance converges to zero (result 2.1.7). Thus we need to check whether the expected value of U_n approaches a finite quantity and variance of U_n tends to zero. Now by its construction U_n is unbiased for γ thus the first condition is satisfied.

Coming to the variance condition, if we consider the representation

$$Var(U_n) = \frac{1}{n} Var(\sqrt{n} U_n),$$

by theorem 3.2.1 as $n \to \infty$, $Var(\sqrt{n}U_n)$ goes to $k^2\xi_1$, and $\frac{1}{n}$ tends to zero as $n \to \infty$. Thus $Var(U_n)$ is a product of two converging sequences one converging to a finite quantity and the other to zero. Thus $Var(U_n) \to 0$ as $n \to \infty$. Therefore, U_n is a consistent estimator for γ.

Next we examine the asymptotic normality of the U-statistics U_n for an estimable parameter γ of degree k. We want to show that under certain conditions a U-statistic has a limiting normal distribution. Given any U-statistic we can obtain a related random variable that has an established normal limiting distribution, and is asymptotically equivalent to the U-statistic in the sense that their difference converges in probability to zero. First some definitions and results needed to assess the asymptotic normality of the U-statistics will be stated.

Let $X_1, ..., X_n$ be iid random variables with cdf $F(.)$. Let $W = w(X_1, ..., X_n)$ and have a limiting normal distribution where $w(.)$ is symmetric in its arguments. Let $W^* = w - E(w)$, so that $E(w^*) = 0$. Now consider a class of random variables, each member of which is a sum of iid random variables i.e $\mathcal{V} = \{v : v = \sum_{i=1}^{n} k(X_i)$ where $k(.)$ is some real valued function$\}$

.

Definition 3.2.3. The projection of W^* on $\mathcal{V}$ is given by

$$v^* = \sum_{i=1}^{n} k^*(X_i)$$

where $k^*(x_i) = E(W^*|X_i = x)$

Example 3.2.4. Consider the U-statistic for the population variance σ^2 of degree 2 with kernel $h(X_1, X_2) = X_1^2 - X_1X_2$. Then

$$U_2(X_1, ..., X_n) = \frac{1}{n(n-1)}\{(n-1)\sum_{i=1}^{n} X_i^2 - 2\sum\sum_{i<j} X_iX_j\}$$

if we now fix $X_i = x$ and take the expected value of $U_2(X_1, ..., X_n)$ with respect to the remaining n-1 X_i's we have

$$E\{U_2(X_1,...,X_n)-\gamma|X_i=x\} = \frac{1}{n(n-1)}\{(n-1)[(\sigma^2+\mu^2)(n-1)+x^2]-2\binom{n-1}{2}\mu^2-2(n-1)\mu x\}-\sigma^2$$

$$= \frac{1}{n(n-1)}\{(n-1)^2\sigma^2 + (n-1)^2\mu^2 + (n-1)x^2 - (n-1)(n-2)\mu^2 - 2(n-1)\mu x\} - \sigma^2$$

$$= \frac{1}{n}\{(n-1)\sigma^2 + (n-1)\mu^2 + x^2 - (n-2)\mu^2 - 2\mu x\} - \sigma^2$$

$$= \frac{1}{n}\{(n-1)\sigma^2 + x^2 + \mu^2 - 2\mu x\} - \sigma^2$$

$$= \frac{1}{n}\{(x-\mu)^2 - \sigma^2\}$$

where μ is the mean of the distribution $F(.)$. Hence the projection of $U_2(X_1, ..., X_n)$ on $\mathcal{V}$ is given by

$$v^* = \sum_{i=1}^{n} \frac{1}{n}((x_i-\mu)^2 - \sigma^2)$$

$$= \frac{1}{n}\sum_{i=1}^{n}((x_i-\mu)^2 - \sigma^2)$$

Note 3.2.5. The asymptotic distribution of v^* is easily determined from the usual central limit theorem, since it is a sum of iid terms $Y_i = (X_i - \mu)^2$. Where $Y_i = (X_i - \mu)^2 \rightarrow^d N(0, \mu_4 - \sigma^2)$, details on the proof of this distribution is given in chapter 4, section 4.3.

Lemma 3.2.6. *If $U(\cdot)$ is a U-statistic for the estimable parameter γ with symmetric kernel $h(\cdot)$, the projection of $U(\cdot) - \gamma$ on $\mathcal{V}$ is given by*

$$V^* = \frac{k}{n}\sum_{i=1}^{n}\{h_1(X_i) - \gamma\}$$

where $h_1(x) = E\{h(X_1, ..., X_k)|X_1 = x\}$

PROOF

The projection of $U(.) - \gamma$ on $\mathcal{V}$ is given by

$$
\begin{aligned}
v^* &= \sum_{i=1}^{n} \{E(U(.) - \gamma | X_i = x)\} \\
&= \sum_{i=1}^{n} E[\frac{1}{\binom{n}{k}} \sum_{\beta \in B} h(X_{\beta_1}, ..., X_{\beta_k}) - \gamma | X_i = x] \\
&= \sum_{i=1}^{n} \frac{1}{\binom{n}{k}} \sum_{\beta \in B} E[h(X_{\beta_1}, ..., X_{\beta_k}) - \gamma | X_i = x]
\end{aligned}
$$

where B is the set of all $\binom{n}{k}$ combinations of k integers $\beta_1 < \beta_2 < ... < \beta_k$ chosen from $\{1, 2, ..., n\}$. But

$$
E[h(X_{\beta_1}, ..., X_{\beta_k}) - \gamma / X_i = x] = \begin{cases} h_1(X_i) - \gamma & , \quad \beta \in B \\ \quad\quad 0 & , \quad otherwise \end{cases}
$$

now for each X_i there are $\binom{n-1}{k-1}$ of the $\binom{n}{k}$ subset that contains β. Thus,

$$
\begin{aligned}
v^* &= \sum_{i=1}^{n} \frac{k!(n-1)!(n-k)!}{n!(n-k)!(k-1)!} (h_1(x_i) - \gamma) \\
&= \sum_{i=1}^{n} \frac{k}{n} (h_1(x_i) - \gamma) \\
&= \frac{k}{n} \sum_{i=1}^{n} (h_1(x_i) - \gamma)
\end{aligned}
$$

hence the result.

Theorem 3.2.7 (Hoeffding's one sample U-statistics Theorem)**.** Let

$$
U(X_1, X_2, ..., X_n) = U_n = \frac{1}{\binom{n}{k}} \sum_{i \in I} h(X_{i_1}, X_{i_2}, ..., X_{i_k})
$$

where the summation is over the set I of all $\binom{n}{k}$ combinations of k integers $i_1 < i_2 < ... < i_k$ chosen from $\{1, 2, ..., n\}$. If $E[h^2(X_1, X_2, ..., X_k)] < \infty$, then $\sqrt{n}(U_n - \gamma)$ has a limiting normal distribution with mean zero and variance $k^2\xi_1$, provided

$$
\xi_1 = E\{h(X_1, X_2, ..., X_k) h(X_1, X_{k+1}, ..., X_{2k-1})\} - \gamma^2 < \infty.
$$

Details of the proof for the above theorem will be given in chapter 5.

Example 3.2.8. (Application of one sample U-statistics theorem)

Consider the estimator of variance σ^2 in example 3.2.4. The symmetric kernel is given by $h(X_1, X_2) = \frac{(X_1 - X_2)^2}{2}$

$$\begin{aligned}
\xi_1 &= Cov(h(X_1, X_2), h(X_1, X_3)) \\
&= Var(E[h(X_1, X_2)|X_1]) + E[Cov(h(X_1, X_2), h(X_1, X_3))|X_1] \\
&= Var(E[h(X_1, X_2)|X_1]) + 0 \\
&= Var(E[\frac{(X_1 - X_2)^2}{2}/X_1]) \\
&= Var(E[\frac{1}{2}(X_1 - \mu + \mu - X_2)^2/X_1]) \\
&= Var(\frac{1}{2}(E(X_1 - \mu)^2|X_1 + E(X_2 - \mu)^2/X_1 - 2E(X_1 - \mu)(X_2 - \mu)|X_1) \\
&= \frac{1}{4}Var((X_1 - \mu)^2 + E(X_2 - \mu)^2 - 2(X_1 - \mu)E(X_2 - \mu) \\
&= \frac{1}{4}Var((X_1 - \mu)^2 + \sigma^2) \\
&= \frac{1}{4}[E(X_1 - \mu)^4 - (E(X_1 - \mu)^2)^2] \\
&= \frac{1}{4}(\mu_4 - \sigma^4)
\end{aligned}$$

implying $\xi_1 < \infty$
where $\mu_4 = E((X_1 - \mu)^4)$ is the 4^{th} central moment. So $nVar(U) \to \mu_4 - \sigma^4$, hence $\sqrt{n}(U - \sigma^2) \to N(0, \mu_4 - \sigma^4)$

CHAPTER 4 : CONVERGENCE RATES OF U-STATISTICS AND S^2, TWO ESTIMATORS OF σ^2

We have seen in chapter 3 that the U-statistics for any estimable parameter γ has an asymptotic normal distribution. In this chapter, we compare the rate of convergence to the normal distribution of the U-statistics and the method of moments estimator for the population variance σ^2. The chapter is divided into three sections, in section one we discuss some results needed to assess the rate of convergence to normal of the U-statistics and the method of moments estimator for the population variance. In section two we discuss the rate of convergence to normal for the U-statistics for the parameter σ^2. Section three will focus on the rate of convergence to normal for the method of moments estimator for the population variance σ^2.

4.1. Introduction

Here we present some results necessary to assess the rate of convergence to normal for the U-statistics for σ^2 and the method of moments estimator for the population variance. We begin by giving some results on order relations with respect to convergence rate.

Two sequences $\{a_n\}$ and $\{b_n\}$ are defined to be asymptotically equivalent denoted by

$$a_n \sim b_n$$

if

$$\frac{a_n}{b_n} \to 1 \quad \text{as } n \longrightarrow \infty.$$

Here we consider three other relationships between sequences denoted by o, $\asymp$ and O, which corresponds to a_n being of smaller order than b_n, a_n being of order equal to b_n and a_n being of order less or equal to b_n, respectively.

We begin by the relationship $a_n = o(b_n)$, which states that for larger n, a_n is an order of magnitude smaller than b_n.

Definition 4.1.1. (a_n) is said to be of order smaller than (b_n) denoted by $a_n = o(b_n)$ if

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 0.$$

Example 4.1.2. Suppose that X has a binomial distribution b(p,n) corresponding to n independent trials with success probability p. The standard estimator $\frac{X}{n}$ for p is unbiased since

$$E(\frac{X}{n}) = \frac{np}{n} = p$$

and has variance $\frac{p(1-p)}{n}$. An interesting class of competitors is the set of estimators

$$\delta(X) = \frac{a+X}{a+b+n}.$$

Now

$$E(\delta(X)) = \frac{a+np}{a+b+n}$$

the bias of $\delta(X)$ is thus

$$E(\delta(X)) - p = \frac{a+np}{a+b+n} - p = \frac{aq-bp}{a+b+n}$$

and $Var(\delta(X)) = \frac{npq}{(a+b+n)^2}$. The accuracy of an estimator $\delta(X)$ of the parameter $g(\theta)$ is most commonly measured by the mean squared error $E[\delta(X) - g(\theta)]^2 = (bias(\delta(X)))^2 + Variance(\delta(X))$. For the binomial distribution we have

$$\frac{(bias(\delta(X)))^2}{Var(\delta(X))} = \frac{(aq-bp)^2/(a+b+n)^2}{npq/(a+b+n)^2} \to 0$$

and hence $(bias(\delta(X)))^2 = o[Var(\delta(X))]$. Both terms tend to zero, but the square of the bias tends to zero much faster at the rate $\frac{1}{n^2}$ compared to the rate $\frac{1}{n}$ for the variance. The bias therefore contributes relatively little to the mean squared error.

Note 4.1.3. It is not always true that the bias tends to zero much faster than the variance for a given estimator.

Definition 4.1.4. $\{a_n\}$ is said to be of order equal to $\{b_n\}$ denoted by $a_n \asymp b_n$ if there exists constants $0 < m < M < \infty$ and an integer n_0 such that $m < |\frac{a_n}{b_n}| < M$ for all $n > n_0$

Example 4.1.5. Let $a_n = 2n + 3, b_n = n$. Then $\frac{a_n}{b_n} < 3$ holds when $2n + 3 < 3n$ i.e for all $n > 3$ and $\frac{a_n}{b_n} > 2$ holds for all n. If we set $m = 2, M = 3$ and $n_0 = 3$. then $m < |\frac{a_n}{b_n}| < M$ is satisfied.

In addition to $\{a_n\}$ being of smaller order than $\{b_n\}$ or the same order as $\{b_n\}$, it is useful to have a notation also for $\{a_n\}$ being of order less than or equal to that of $\{b_n\}$. This relationship is denoted by $a_n = O(b_n)$. Now $a_n = O(b_n)$ if $|\frac{a_n}{b_n}|$ is bounded i.e if there exist M and n_0 such that $|\frac{a_n}{b_n}| < M$ for all $n > n_0$.

Definition 4.1.6. Suppose we have $\lim_{n\to\infty} f_n = f_o$, if we have that $|f_n - f_o| \leq Mg(n)$, where for large n $g_n \to 0$. We say that the rate of convergence of f_n to f_o is of the order $g(n)$ and we write $|f_n - f_o| = Og(n)$.

Example 4.1.7. We know that $\lim_{n\to\infty} \frac{1}{n^2-1} = 0$, $\frac{1}{n^2-1} = \frac{1}{n^2}\left(\frac{1}{1-\frac{1}{n^2}}\right)$ but $\left(\frac{1}{1-\frac{1}{n^2}}\right) \leq 2$ for $n \geq 2$ so that $\frac{1}{n^2-1} \leq 2\left(\frac{1}{n^2}\right)$. Thus we write $\frac{1}{n^2-1} = O\left(\frac{1}{n^2}\right)$

We shall now state without proof the following lemma which we shall use to prove Berry-Esseen theorem.

Lemma 4.1.8 (Smoothness Theorem (Feller (1971)). *Let X be a random variable with cdf F, expectation 0 and characteristic function φ. Suppose that G has a derivative g such that $|g| \leq m$. Finally, suppose that g has a continously differentiable Fourier transform ϑ such that $\vartheta(0) = 1$ and $\vartheta'(0) = 0$. Then*

$$|F(x) - G(x)| \leq \frac{1}{\pi}\int_{-T}^{T} |\frac{\varphi(\zeta) - \vartheta(\zeta)}{\zeta}|d\zeta + \frac{24m}{\pi T} \quad \textit{for all} \quad T > 0$$

Theorem 4.1.9 (Berry-Esseen Theorem). (Lehmann (1999)) Let $X_1, X_2, ..., X_n$ be a random sample from distribution F, such that $E(X_i) = 0$, $E(X_i^2) = \sigma^2$, $E(|X_i|^3) = \rho < \infty$ for $i = 1, 2, ..., n$. Then there exists a constant C (independent of F) such that for all $x \in \mathbb{R}$

$$|F_n(x) - \Phi(x)| \leq \frac{C}{\sqrt{n}}\frac{\rho}{\sigma^3}$$

where

$$F_n(x) = P\left(\frac{\sqrt{n}(\bar{X})}{\sigma} \leq x\right)$$

and Φ is the cdf of the standard normal distribution

PROOF We give the proof of Berry-Esseen theorem as given by Feller (1971).
With $F = F_n$ and $G = \Phi$, choosing $T = \frac{4}{3}.\frac{\sigma^3\sqrt{n}}{\rho} \leq \frac{4}{3}\sqrt{n}$. Since the standard normal density has a maximum $m < 2/5$ we get by Smoothness theorem that

$$\pi|F_n(x) - \Phi(x)| \leq \int_{-T}^{T} |\varphi^n(\zeta/\sigma\sqrt{n}) - e^{\frac{-\zeta^2}{2}}|\frac{d\zeta}{|\zeta|} + \frac{9.6}{T}. \tag{4.1}$$

To appraise the integrand in (4.1.1) we note that the expansion for $\alpha^n - \beta^n$ leads to the inequality

$$|\alpha^n - \beta^n| \leq n|\alpha - \beta|.\vartheta^{n-1} \quad \text{if} \quad |\alpha| \leq \vartheta, |\beta| \leq \vartheta. \tag{4.2}$$

We use this with $\alpha = \varphi(\frac{\zeta}{\sigma\sqrt{n}})$ and $\beta = e^{-\frac{1}{2}\zeta^2/n}$. From the inequality for e^{it} in result 2.1.14, we have

$$\left|\varphi(t) - 1 + \frac{1}{2}\sigma^2 t^2\right| = \left|\int_{-\infty}^{\infty} (e^{itx} - 1 - itx + \frac{1}{2}t^2x^2) F\{dx\}\right| \leq \frac{1}{6}\rho|t|^3 \tag{4.3}$$

and hence

$$|\varphi(t)| \leq 1 - \frac{1}{2}\sigma^2 t^2 + \frac{1}{6}\rho|t|^3 \quad \text{if} \quad \frac{1}{2}\sigma^2 t^2 \leq 1.$$

We conclude that for $|\zeta| \leq T$.

$$|\varphi(\frac{\zeta}{\sigma\sqrt{n}})| \leq 1 - \frac{\zeta^2}{2n} + \frac{\rho|\zeta|^3}{6\sigma^3 n^{3/2}} \leq 1 - \frac{5}{18n}\zeta^2 \leq e^{-\frac{5}{18}\zeta^2/n}$$

since $\sigma^3 < \rho$ the assertion of the theorem is trivially true for $\sqrt{n} \leq 3$ and hence we may assume $n \geq 10$. Then

$$|\varphi(\frac{\zeta}{\sigma\sqrt{n}})|^{n-1} \leq e^{-\frac{1}{4}\zeta^2} \tag{4.4}$$

and the right hand side may serve for the bound ϑ^{n-1} in (4.1.2) . Noting that $e^{-x} - 1 + x \leq \frac{x^2}{2}$ for $x > 0$ we get from (4.1.3) that

$$n|\varphi(\frac{\zeta}{\sigma\sqrt{n}}) - e^{-\zeta^2/2n}| \leq n|\varphi(\frac{\zeta}{\sigma\sqrt{n}}) - 1 + \frac{\zeta^2}{2n}| +$$

$$n|1 - \frac{\zeta^2}{2n} - e^{-\zeta^2/2n}| \leq \frac{\zeta}{6\sigma^3\sqrt{n}}|\zeta|^3 + \frac{\zeta^4}{8n}.$$

Since $\sqrt{n} \geq 3$ it follows from (4.1.2) and (4.1.4) that the integrand in (4.1.1) is $\leq \frac{1}{T}(\frac{2}{9}\zeta^2 + \frac{1}{18}|\zeta|^3)e^{-\frac{1}{4}\zeta^2}$. This function is integrable over $-\infty < \zeta < \infty$, and integration by parts now shows that

$$\pi T|F_n(x) - \Phi(x)| \leq \frac{8\sqrt{\pi}}{9} + \frac{8}{9} + 10. \tag{4.5}$$

Since $\sqrt{\pi} < 9/5$, the right hand side of (4.1.5) is $< 113/9 < 4\pi$, and the proof is complete.

4.2. Convergence rate of U-statistics for σ^2

Hoeffding (1948) proved that the distribution function of $\frac{U_n - \gamma}{\sigma_n}$, where U_n is a U-statistics for an estimable parameter γ of degree 2 and $\sigma_n = Var(U_n)$, converges to the standard normal distribution function Φ under the sole condition of the existence of $E(h^2(X_1, X_2))$. A study on the rate of this convergence started with Grams and Serfling (1973). They showed that $Sup|P(\frac{U_n - \gamma}{\sigma_n} \leq x) - \Phi(x)|$ is $O(n^{-r/(2r+1)}), n \to \infty$, when $E(h^{2r}) < \infty$, leading to $O(n^{-(1/2)+\epsilon}), \epsilon > 0$, when h has finite moments of all orders. An order bound of exactly $O(n^{-1/2})$ was found by Bickel (1974) assuming U-statistics with bounded kernels $h(X_1, X_2)$. Chan and Wierman (1977) succeeded in weakening considerably the assumptions of the previous theorem obtaining the order bound of $O(n^{-1/2})$ when the fourth moments of h(.) exists. Callaert and Janseen (1977) proved that $O(n^{-1/2})$ can be attained requiring only the existence of the third absolute moment of h(.). Bentkus, Gotze and Zitikis (1994) showed that the moments conditions of

$$E|E(h(X_1, X_2)|X_1)|^3 < \infty \tag{4.6}$$

$$E|h(X_1, X_2)|^{5/3} < \infty \tag{4.7}$$

are the weakest in achieving the strongest results on the Berry-Esseen bound for U-statistic for an estimable parameter of degree 2, i.e the rate of convergence in the CLT is of the order $O(n^{-1/2})$ provided that

$$E|E(h(X_1, X_2)/X_1)|^3 < \infty,$$

and

$$E|h(X_1, X_2)|^{5/3} < \infty$$

Where h(.) is a symmetric kernel corresponding to the U-statistic for an estimable parameter γ of degree 2. Bentkus, Jing and Zhou (2009) extended the optimal results on Berry-Esseen bounds for U-statistics of order 2 to those of higher order.

In this section we discuss the rate of convergence to normal for U_n the U-statistics for the population variance σ^2. We shall use the results of Ahmad (1981) in generality. Define the U-statistics for an estimable parameter γ of degree 2 by

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

If we write

$$\Lambda_n = Sup_x \left| P\left(\frac{U_n - \gamma}{\sigma_n} \le x\right) - \Phi(x)\right|,$$

$$\hat{\Lambda}_n = Sup_x \left| P\left(\frac{\sqrt{n}(U_n - \gamma)}{\sigma_g} \le x\right) - \Phi(x)\right|.$$

where $\sigma_n^2 = Var(U_n)$ and $\sigma_g^2 = Var[g(X_1)]$ where $g(X_1) = E[h(X_1, X_2)|X_1 = x] - \gamma$, assume that $\gamma = 0$, define the projection of U_n as

$$\hat{U}_n = \binom{n}{2}^{-1} \sum_{1\le i<j\le n} (g(X_i) + g(X_j)) = \binom{n}{2}^{-1} 2(n-1) \sum_{i=1}^{n} g(X_i) = \frac{4}{n} \sum_{i=1}^{n} g(X_i)$$

so that $E(\hat{U}_n) = 0$ and put $\hat{\sigma}_n^2 = \frac{4\sigma_g^2}{n}$. Set $\kappa_n = \frac{\hat{U}_n}{\hat{\sigma}_n}$, and

$$\begin{aligned} R_n &= \hat{\sigma}_n^{-1} \binom{n}{2}^{-1} \sum_{1\le i<j\le n} [h(X_i, X_j) - g(X_i) - g(X_j)] \\ &= \hat{\sigma}_n^{-1} \binom{n}{2}^{-1} \sum_{1\le i<j\le n} y_{ij}, \quad \text{say.} \end{aligned} \tag{4.8}$$

Let

$$R'_n = \hat{\sigma}_n^{-1} \binom{n}{2}^{-1} \sum_{1\le i<j\le c_n} y_{ij}$$

and

$$R''_n = \hat{\sigma}_n^{-1} \binom{n}{2}^{-1} \sum_{c_{n+1}\le i<j\le n} y_{ij}$$

for some integer c_n. Thus $R_n = R'_n + R''_n$.

Lemma 4.2.1. *For any $r > 2$, $R''_n = \hat{\sigma}_n^{-1} \binom{n}{2}^{-1} \sum_{c_{n+1}\le i<j\le n} y_{ij}$,*

$V_r = E|g(X_1)|^r$ and $\sigma_g^r = E(g^r(X_1))$

$$E|R''_n|^r \le C(V_r/\sigma_g^r)(n - c_n)^{r/2n-r}$$

where c_n is some integer and C is a an absolute constant.

Theorem 4.2.2. Let $U_n = \binom{n}{2}^{-1} \sum_{1\le i<j\le n} h(X_i, X_j)$ be a U-statistics for an estimable parameter γ of degree 2 such that $g(X_1) = E(h(X_1, X_2)|X_1)$ has positive variance σ_g^2. If

$V_3 = E|h(X_1, X_2)|^3 < \infty$ then there exists an absolute constant C such that for all $n \geq 2$

$$\Lambda_n = Sup_x \left| P\left(\frac{U_n - \gamma}{\sigma_n} \leq x\right) - \Phi(x) \right| \leq CV_3\sigma_g^{-3} n^{-\frac{1}{2}} \tag{4.9}$$

PROOF

We have for some $\epsilon_n > 0$, that

$$\hat{\Lambda}_n \leq Sup_x|[\kappa_n + R'_n \leq x] - \Phi(x)| + P[|R''_n| > \epsilon_n] + O(\epsilon_n) \tag{4.10}$$

But by Lemma 4.2.1 we get

$$P[|R''_n| \geq \epsilon_n] \leq \epsilon_n^{-3} E|R''_n|^3 \leq C(\sigma_g^{-3} V_3)(n - c_n)^{3/2} n^{-3} \epsilon_n^{-3} \tag{4.11}$$

Choosing $\epsilon_n = (n - c_n)^{3/8} n^{-3/4}$ and then choosing c_n such that $\epsilon_n \leq c_n^{-1/2}$ we get the last two terms of the RHS of (4.2.5) being bounded by $C(V_3/\sigma_n^3)n^{-1/2}$. It remains to evaluate the first term. Since $V_3 < \infty$, then $E|g(X_1)|^3 < \infty$ so that

$$\phi_{\kappa_n}(t) = exp[-(t^2/2)(1 + 0(|t/\sqrt{n}|))], \tag{4.12}$$

where $\phi_x(.)$ denote the characteristic function of X. Therefore, it follows that $|\phi_{s_n}(t) - e^{-t^2/2}| \leq C\frac{t^2}{n^{1/2}} e^{-t^2/4}$. Thus if $T = \alpha\sqrt{n}$ we get

$$\int_0^T (|t|^{-1}\phi_{s_n}(t) - e^{-t^2/2}|)dt \leq \frac{C}{n^{1/2}} \int_0^T |t|e^{-t^2/4} dt \leq C_1 n^{-1/2}(V_3/\sigma_g^3).$$

Next note that

$$|\phi_{s_n}(t) - \phi_{s_n + R'_n}(t)| \leq |Ee^{its_n}(itR'_n)| + \frac{t^2}{2} ER'^2_n.$$

But,

$$|Ee^{its_n}(itR'_n)| \leq \frac{1}{2}\sigma_g^{-3} n^{-1/2} t^2 e^{-t^2/4} E|g(X_1)g(X_2)y_{12}|,$$

while

$$E|g(X_1)g(X_2)y_{12}| \leq C\sigma_g^3(V_3/\sigma_g^3) = CV_3.$$

Hence we have

$$|Ee^{its_n}(itR'_n)| \leq C(V_3/\sigma_g^3) n^{-1/2} t^2 e^{-t^2/6}.$$

Also

$$ER'^2_n \leq \frac{E(f(X_1, X_2))^2}{n\sigma_g^2} \leq C(V_3/\sigma_g^2) n^{-1}.$$

Therefore from (4.2.5),

$$\int_0^T |t|^{-1}|\phi_{s_n}(t) - e^{-t^2/2}|dt$$

$$\leq \int_0^{d_n} |t|^{-1}|Ee^{its_n}(itR'_n)|dt$$

$$+ \int_0^{d_n} \frac{|t|}{2}ER'^2_n dt$$

$$+ \int_{d_n}^T |t|^{-1}|\phi_{s_n}(t) - \phi_{s_n+R'_n}(t)|dt$$

$$= I_{1n} + I_{2n} + I_{3n}, \quad \text{say.} \tag{4.13}$$

But

$$I_{3n} \leq C(\frac{V_3}{\sigma_g^3})\frac{d_n^2}{n},$$

thus choosing $d_n = n^{1/4}$ we get

$$I_{3n} \leq Cn^{-1/2}(V_3/\sigma_g^3).$$

Also with $d_n = n^{1/4}$,

$$I_{2n} \leq Cn^{-1/2}(V_3/\sigma_g^3).$$

Finally let η denote the characteristic function of $g(X_1)$. Then we have

$$|Ee^{its_n}(1 - e^{itR'_n})| \leq |t|E|R'_n||\eta(\frac{t}{\sqrt{n}\sigma_g})|^{n-c_n}$$

$$\leq |t|n^{-1/2}(V_3/\sigma_g^3) \times exp(-(t^2/2)(\frac{n-c_n}{n})).$$

Thus

$$I_{1n} \leq (V_3/\sigma_g^3)n^{-1/2}\int_{d_n}^T exp(-(t^2/2)(\frac{n-c_n}{n}))dt$$

$$\leq Cn^{-1/2}(V_3/\sigma_g^3).$$

From (4.2.6), (4.2.7) and (4.2.8) we conclude that

$$\hat{\Lambda}_n \leq C(\frac{V_3}{\sigma_g^3})n^{-1/2},$$

But since

$$|\frac{\hat{\sigma}_n}{\sigma_n} - 1| \leq (n-1)^{-1}(1 + \frac{E(f^2(x_1, x_2))}{\sigma_g^2}) \leq Cn^{-1/2}(V_3/\sigma_g^3),$$

if follows that

$$\Lambda_n \leq (\frac{CV_3}{\sigma_g^3})n^{-1/2}.$$

Hence

$$Sup_x \left| P\left(\frac{U_n - \gamma}{\sigma_n} \leq x\right) - \Phi(x) \right| \leq CV_3\sigma_g^{-3}n^{-1/2}$$

■

Since the U-statistics for the population variance is of order 2 (example 2.1.21), then by Callaert and Janseen (1978) [Theorem 4.2.2] the rate of convergence to normal distribution for U_n the U-statistics for the population variance is given by $CV_3\sigma_g^{-3}n^{-1/2}$ provided $E|h(X_1, X_2)|^{\frac{5}{3}} < \infty$

4.3. Convergence rate of S^2

Let $X_1, ..., X_n$ be a sample from a distribution with pdf $f(x|\theta_1, ..., \theta_l)$. Method of moments estimates are found by equating the first k sample moments to the corresponding k population moments, and solving the resulting system of simultaneous equations. More precisely, define

$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

$$\mu_k^{'} = E(X^k)$$

The population moment $\mu_k^{'}$ will typically be a function of $\theta_1, ..., \theta_l$. The method of moments estimator $(\widehat{\theta}_1, ..., \widehat{\theta}_l)$ of $(\theta_1, ..., \theta_l)$ is obtained by solving the following system of equations for $(\theta_1, ..., \theta_l)$ in terms of $(m_1, ..., m_l)$

$$m_k = \mu_k^{'}$$

for $k = 1, ..., l$

Next we discuss the asymptotic normality of the method of moments estimator for the population variance, $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

Let $X_1, X_2, ..., X_n$ be iid with $E(X_i) = \mu$, $Var(X_i) = \sigma^2$ and $Var(X_i - \mu)^2 = \tau^2 < \infty$. Then

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \to^p \sigma^2.$$

To prove this note that

$$\begin{aligned} S^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}((X_i - \mu) - (\bar{X} - \mu))^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - \frac{2}{n}(\bar{X} - \mu)\sum_{i=1}^{n}(X_i - \mu) + (\bar{X} - \mu)^2 \end{aligned} \tag{4.14}$$

now $(\bar{X} - \mu) \to^p 0$ and $\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 \to^p E(X_i - \mu)^2$ by the law of large numbers, thus all terms in (4.3.1) tends in probability to zero except the first term, which tends to $E(X_i - \mu)^2 = Var(X_i) = \sigma^2$ thus $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \to^p \sigma^2$.

■

Now let us consider the limit behavior of $\sqrt{n}(S^2 - \sigma^2)$. We have that

$$\begin{aligned} \sqrt{n}(S^2 - \sigma^2) &= \sqrt{n}(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 - \sigma^2) \\ &= \sqrt{n}(\frac{1}{n}\Sigma_{i=1}^{n}(X_i - \mu + \mu - \bar{X})^2 - \sigma^2) \\ &= \sqrt{n}(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - \sigma^2 - (\bar{X} - \mu)^2) \\ &= \sqrt{n}(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - \sigma^2) - \sqrt{n}(\bar{X} - \mu)^2 \end{aligned}$$

Since $\sqrt{n}(\bar{X} - \mu)^2 = \sqrt{n}(\bar{X} - \mu)(\bar{X} - \mu)$, and $\sqrt{n}(\bar{X} - \mu) \to^d N(0, \sigma^2)$ and $(\bar{X} - \mu) \to^p 0$ while $\sqrt{n}(\bar{X} - \mu)^2 \to^d N(0, \sigma^2)$ by Slutsky's theorem, then

$$\lim_{n\to\infty} \sqrt{n}(S^2 - \sigma^2) = \lim_{n\to\infty}\{\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - \sigma^2)\}$$

Now $E(X_i - \mu)^2 = \sigma^2$ and $Var(X_i - \mu)^2 = E[(X_i - \mu)^4 - \sigma^4] = \mu_4 - \sigma^4 = \tau^2$
Hence $(X_i - \mu)^2 - \sigma^2 \to^d N(0, \mu_4 - \sigma^4)$ so that $\sqrt{n}(S^2 - \sigma^2) \to^d N(0, \mu_4 - \sigma^4)$
i.e $\sqrt{n}(S^2 - \sigma^2) \to^d N(0, \tau^2)$.

We have shown that the method of moments estimator for the population variance σ^2 has a limiting normal distribution. To find how fast this convergence takes place we use the following results obtained by Van Zwet (1984). Let $X_1, X_2, ..., X_n$ be a random sample and let $t : \mathbb{R}^n \to \mathbb{R}$ be a symmetric function in its n arguments. Define $T = t(X_1, X_2, ..., X_n)$ and assume

$$E(T) = 0, E(T^2) = 1.$$

Lemma 4.3.1. *Suppose that $E(T) = 0, E(T^2) = 1$ hold and that positive numbers a and b exists such that*

$$E|E(T|X_1)|^3 \leq an^{-3/2}$$
$$1 + E\{E(T|X_1, ..., X_{n-2})\}^2 - 2E\{E(T|X_1, ..., X_{n-1})\}^2 \leq bn^{-3}$$
$$\text{Then} \qquad Sup_x|P(T \leq x) - \Phi(x)| \leq C(a+b)n^{-1/2}$$

where C denotes a universal constant.

A possible extension of lemma 4.3.1 is to relax the moments conditions and obtain the following result.

Corollary 4.3.2. *Let $T = \hat{T} + R$, if we have a Berry-Esseen bound for $\hat{T}$ given by*

(i)

$$Sup_x|P(\hat{T} \leq x) - \Phi(x)| \leq Cn^{-1/2}$$

and R satisfies

(ii)

$$P(|R| \geq an^{-1/2}) \leq bn^{-1/2}$$

where a and b are positive integers, then we have a Berry-Esseen bound for T as follows:

(iii)

$$Sup_x|P(T \leq x) - \Phi(x)| \leq (a+b+C)n^{-1/2}.$$

Now since $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is symmetric in its arguments $X_1, X_2, ..., X_n$, then we can write S^2 as $S^2 = L + \Delta$ where L is asymptotically normal. If we have that

$L = (X_1 + X_2 + ... + X_n)/\sqrt{n}$, then L have a Berry-Esseen bound (i) by Berry-Esseen theorem. By Chebyshev's inequality Δ satisfies (ii). We thus have a Berry-Esseen bound for S^2 given by (iii). Therefore, the rate of convergence to normal distribution for the method of moments estimator for the population variance is given by $(a + b + C)n^{-1/2}$.

4.4. Comparison of the two rates of convergence

It is clear from theorem 4.2.2 that the rate of convergence to the normal distribution for the U-statistics for the population variance depends on the moments condition $E|h(X_1, X_2)|^3$. By Bentkus, Gotze and Zitikis (1994) we can conclude that $CV_3\sigma_g^{-3}n^{-1/2}$ is the strongest rate of convergence to the normal distribution for the U-statistics. Now the rate of convergence to the normal distribution for the method of moments estimator for the population variance is given by $(a + b + C)n^{-1/2}$ as can be seen from Corollary 4.3.2 (iii). It is clear from Corollary 4.3.2 that the rate of convergence to the normal distribution of the method of moments estimator does not depend on the moments of S^2. Also we notice that both rates of convergence to normal for U-statistics and method of moments estimators are of order $n^{-1/2}$. Now $(a + b + C)n^{-1/2} < CV_3\sigma_g^{-3}n^{-1/2}$ since $(a + b + C) < CV_3\sigma_g^{-3}$, we thus conclude that the method of moments converges at a faster rate to normal distribution as compared to the U-statistics.

CHAPTER 5 : REVIEW OF PROOFS OF HOEFFDING'S ONE SAMPLE U-STATISTICS THEOREM

In this chapter we examine the different proofs of Hoeffding's one sample U-statistics theorem as given in the period 1948-2012. The chapter is organized as follows; in section one we state the Hoeffding's one sample U-statistics theorem. In sections two,three, four and five we give the different proofs of Hoeffding's one sample U-statistics theorem as given by Hoeffding (1948), Lee (1990), Ferguson (1996) and Beutner and Zahle (2012) respectively.

5.1. Introduction

Theorem 5.1.1 (Hoeffding's one sample U-statistics Theorem)**.** Let $X_1, ..., X_n$ be be a random sample from a population with parameter γ

$$U(X_1, X_2, ..., X_n) = U_n = \frac{1}{\binom{n}{k}} \sum_{i \in I} h(X_{i_1}, X_{i_2}, ..., X_{i_k})$$

where the summation is over the set I of all $\binom{n}{k}$ combinations of k integers $i_1 < i_2 < ... < i_k$ chosen from $\{1, 2, ..., n\}$. If $E[h^2(X_1, X_2, ..., X_k)] < \infty$, then $\sqrt{n}(U_n - \gamma)$ has a limiting normal distribution with mean zero and variance $k^2\xi_1$, provided

$$\xi_1 = E\{h(X_1, X_2, ..., X_k)h(X_1, X_{k+1}, ..., X_{2k-1})\} - \gamma^2 < \infty.$$

The Hoeffding's one sample U-statistics theorem states that the distribution of a U-statistics tends, under the conditions that $E(h^2(X_1, ..., X_k)) < \infty$, to the normal form. For k=1, U is the sum of n independent random variables, and in this case the theorem reduces to the central limit theorem for such sum. For $k > 1$, U is a sum of random variables which, in general, are not independent. Under certain assumptions about the function $h(x_1, ..., x_k)$ the asymptotic normality of U can be inferred from the well known results such as Slutsky's theorem and the Projection theorem. In the case of independent and identically distributed X_is the existence of $E(h^2(X_1, ..., X_k))$ is sufficient for the asymptotic normality of U. No regularity conditions are imposed on the function h.

$E(h^2(X_1, ..., X_k))$ is a linear combination of terms of the form $E(h(X_{i_1}, ..., X_{h_k})h(X_{j_1}, ..., X_{j_k}))$, whose existence follows from that of $E[h^2(X_1, ..., X_k)]$ by Schwarz's inequality.

We shall now look at the proof of Hoeffding one sample U-statistics theorem as given by Hoeffding (1948).

5.2. Proof of Hoeffding's One Sample U-statistics Theorem as given by Hoeffding(1948)

We have that

$$U_n = \frac{1}{\binom{n}{k}} \sum_{i \in I} h(X_{i_1}, ..., X_{i_k})$$

and that the projection of $U_n - \gamma$ on the space $\mathcal{U}$ consisting of all variables of the form $\sum_{i=1}^{n} g(X_i)$ where g(.) is some real valued function, is given by

$$U_n^* = \frac{k}{n} \sum_{j=1}^{n} \{h_1(X_j) - \gamma\}$$

If we consider the quantity

$$Y_n = \sqrt{n} U_n^* = \frac{k}{\sqrt{n}} \sum_{j=1}^{n} (h_1(X_j) - \gamma)$$

where $h_1(x_1)$ is defined by

$$h_1(x_1) = E(h(x_1, X_2, ..., X_k) | X_1 = x_1)$$

and

$$h_1(x) = E(h(X_{j_1}, ..., X_{j_k}) | X_j = x) \quad \text{whenever } j \in \{j_1, ..., j_k\}$$

Y_n is the sum of n independent random variables with mean zero and variance $k^2\xi_1$. By the central limit theorem Y_n tends to normal with mean zero and variance $k^2\xi_1$. The theorem will be proved by showing that the random variable $Z_n = \sqrt{n}(U_n - \gamma)$ has the same limiting distribution as Y_n. According to Slutsky's theorem we need to show that $Z_n - Y_n$ converges to zero in probability. It is sufficient to show that $\lim_{n \to \infty} E(Z_n - Y_n)^2 = 0$. Now we write

$$E(Z_n - Y_n)^2 = E(Z_n)^2 + E(Y_n)^2 - 2E(Z_n Y_n)$$

but $E(Z_n)^2 = nVar(U_n) = k^2\xi_1$ (Theorem 3.2.1) and $E(Y_n)^2 = k^2\xi_1$. Therefore, it is enough to show that $E(Z_n Y_n) = k^2\xi_1$. Since

$$U_n = \frac{1}{\binom{n}{k}} \sum_{i \in I} h(X_{i_1}, ..., X_{i_k})$$

then we can write Z_n as

$$Z_n = \sqrt{n}\binom{n}{k}^{-1} \sum_{i \in I}(h(X_{i_1}, ..., X_{i_k}) - \gamma)$$

hence

$$E(Z_n Y_n) = \frac{k}{\binom{n}{k}} \sum_{j=1}^{n} \sum_{i \in I} E(h_1(X_j) - \gamma)h(X_{i_1}, ..., X_{i_k}) - \gamma).$$

Now from the development of the general expression of the variance of a U-statistics in section 2.2, we have that

$$E(Z_n Y_n) = \frac{k}{\binom{n}{k}} n \binom{n-1}{k-1} \xi_1 = k^2 \xi_1$$

so that $E(Z_n - Y_n)^2 = k^2\xi_1 + k^2\xi_1 - 2k^2\xi_1 = 0$ thus $\lim_{n\to\infty} E(Z_n - Y_n)^2 = 0$. Hence Z_n and Y_n have the same limiting distribution and since Y_n is asymptotically normally distributed with mean zero and variance $k^2\xi_1$, we conclude that $Z_n \sim N(0, k^2\xi_1)$. This completes the proof.

5.3. Proof of Hoeffding's One Sample U-statistics Theorem as given by Lee(1990)

Next we look at the proof of Hoeffding's one sample U-statistic theorem as given by Lee(1990). But first we look at some results neccessary for the proof.

Let U_n be the U-statistic as given in definition 2.1.20. Recall that the general expression for the variance of U_n is given by

$$Var(U_n) = \binom{n}{k}^{-1} \sum_{c=1}^{k} \binom{k}{c}\binom{n-k}{k-c} \xi_c$$

so that

$$\lim_{n\to\infty} nVar(U_n) = \lim_{n\to\infty} \frac{nk!(n-k)!k(n-k)!}{n!(k-1)!(n-2k+1)!}\xi_1 = k^2\xi_1$$

Recall also that the projection of $U_n(.) - \gamma$ on a linear space $\mathcal{U}$ is given by

$$U_n^* = \frac{k}{n}\sum_{i=1}^{n}\{h_1(x_i) - \gamma\}$$

where $h_1(x) = E(h(x, X_2, ..., X_k))$

Definition 5.3.1 (Ha'jek Projection)**.** Suppose that $X_1, ..., X_n$ are random variables, for a random variable Y, the Ha'jek projection of Y on a space $\mathcal{U} = \{X_1, ..., X_n\}$ denoted by Y_p is given by

$$Y_p = \sum_{i=1}^{n} g_i(X_i)$$

where $g_i(.)$ are measurable functions such that $E(g_i(X_i))^2 < \infty$

Theorem 5.3.2 (Projection Theorem)**.** $\hat{V}$ is a projection of Y on $\mathcal{U}$ if and only if $\hat{V} \in \mathcal{U}$ and, for all $V \in \mathcal{U}$

$$E(Y - \hat{V})V = 0$$

that is the error $Y - \hat{V}$ is orthogonal to $\mathcal{U}$.

Theorem 5.3.3 (Ha'jek Projection Principle)**.** The Ha'jek projection of an arbitrary random variable $T = T(X_1, X_2, ..., X_n)$ with finite second moment onto $\mathcal{U}$ is given by

$$\hat{V} = \sum_{i=1}^{n} E(T|X_i) - (n-1)E(T)$$

.

PROOF

From the projection theorem,we need to check that $T - \hat{V}$ is orthogonal to each $g_i(X_i)$. Where $g_i(X_i)$ is as given in definition 5.3.1. It suffices if $E(T|X_i) = E(\hat{V}|X_i)$

$$E(T - \hat{V})g_i(X_i) = E[E(T - \hat{V})|X_i]g_i(X_i)$$

But

$$E(\hat{V}|X_i) = E\left[\sum_{j=1}^{n} E[T|X_j] - (n-1)E(T|X_i)\right]$$

$$= E(T|X_i) + \sum_{j \neq i} E[E[T/X_j]/X_i] - (n-1)E(T) = E(T|X_i)$$

because the X_i's are independent, so $T - \hat{V}$ is orthogonal to $\mathcal{U}$.

■

Now we give the proof of Hoeffding's one sample U-statistics theorem as given by Lee (1990).

Recall that the U-statistics for any estimable parameter γ is given by

$$U_n = \frac{1}{\binom{n}{k}} \sum_{i \in I} h(X_{i_1}, ..., X_{i_k})$$

Then by the Ha'jek projection principle, the projection of $U_n - \gamma$ is

$$\hat{U}_n = \frac{k}{n} \sum_{i=1}^{n} E(U_n - \gamma | X_i = x)$$

$$= \sum_{i=1}^{n} E[\frac{1}{\binom{n}{k}} \sum_{j \in I} h(X_{j_1}, ..., X_{j_k}) - \gamma | X_i = x]$$

$$= \sum_{i=1}^{n} \frac{1}{\binom{n}{k}} \sum_{j \in I} E[h(Xj_1, ..., X_{j_k}) - \gamma | X_i = x]$$

But

$$E[h(Xj_1, ..., X_{j_k}) - \gamma | X_i] = \begin{cases} h_1(X_i) & , \quad i \in j \\ 0 & , \quad otherwise \end{cases} \tag{5.1}$$

where $h_1(x) = E(h(x, X_2, ..., X_k))$

For each X_i, there are $\binom{n-1}{k-1}$ of the $\binom{n}{k}$ subsets that contain i. Thus

$$\hat{U}_n = \sum_{i=1}^{n} \frac{k!(n-k)!(n-1)!}{n!(k-1)!(n-k)!} h_1(X_i)$$

$$\hat{U}_n = \frac{k}{n} \sum_{i=1}^{n} h_1(X_i)$$

To see that $\hat{U}_n$ has the same asymptotic distribution as $U_n - \gamma$, notice that

$$E(\hat{U}_n) = \frac{k}{n}\sum_{i=1}^{n} E(h_1(X_i)) = 0$$

since the $h_1(X_i)'s$ are iid, then $E(h_1(X_i)) = 0$, and so the variance of $\hat{U}_n$ is asymptotically the same as that of U_n.

$$\begin{aligned} Var(\hat{U}_n) &= \frac{k^2}{n^2}\sum_{i=1}^{n} Var(h_1(X_i) \\ &= \frac{k^2}{n^2} n\xi_1 \\ &= \frac{k^2}{n}\xi_1 \end{aligned} \quad (5.2)$$

CLT and the fact that variance of $\hat{U}_n$ is $\frac{k^2}{n}\xi_1$ implies that $\sqrt{n}(\hat{U}_n) \to^d N(0, k^2\xi_1)$.

Also recall that $nVar(U_n) \to k^2\xi_1$ as $n \to \infty$. Thus

$$\lim_{n\to\infty} \frac{var(\hat{U}_n)}{var(U_n)} = 1.$$

So that

$$\frac{U_n - \gamma}{\sqrt{var(U_n)}} - \frac{\hat{U}_n}{\sqrt{var(\hat{U}_n)}} \to^p 0$$

which implies $\sqrt{n}(U_n - \gamma - \hat{U}_n) \to^p 0$, as $n \to \infty$
and hence

$$\sqrt{n}(U_n - \gamma) \to^d N(0, k^2\xi_1)$$

this concludes the proof.

5.4. Proof of Hoeffding's One Sample U-statistics Theorem as given by Ferguson(1996)

For a given estimable parameter, γ, and corresponding symmetric kernel $h(x_1, x_2, ..., x_k)$, we take $\mathcal{F}$ to be the class of distributions for which $Var(h(X_1, X_2, ..., X_k)) < \infty$. Recall that the U-statistic for an estimable parameter γ of degree k is given by

$$U_n = \frac{1}{\binom{n}{k}} \sum_{i \in I} h(X_{i_1}, ..., X_{i_k})$$

and that the variance of the general expression of U_n is given by

$$Var(U_n) = \binom{n}{k}^{-1} \sum_{c=1}^{k} \binom{k}{c} \binom{n-k}{k-c} \xi_c$$

For $c = 1$ the coefficient of ξ_1 is

$$\frac{k\binom{n-k}{k-1}}{\binom{n}{k}}$$

We need to show that, if $Var[h(X_1, X_2, ..., X_k)] < \infty$, then

$$\sqrt{n}(U_n - \gamma) \longrightarrow N(0, k^2\xi_1) \quad \text{in distribution as } n \to \infty$$

.

Let

$$U_n^* = \frac{k}{n} \sum_{i=1}^{n} (h_1(X_i) - \gamma)$$

since $h_1(X_i)'s$ are iid with mean zero and variance ξ_1 (5.3.2), the central limit theorem implies that $\sqrt{n}U_n^* \longrightarrow N(0, k^2\xi_1)$, we complete the proof by showing that $\sqrt{n}(U_n - \gamma)$ and $\sqrt{n}U_n^*$ are asymptotically equivalent and so have the same limiting distribution. For this it suffices to show that the difference $\sqrt{n}(U_n - \gamma - U_n^*)$ converges in probability to zero. Since convergence in quadratic mean implies convergence in probability, we therefore need to show that

$$\lim_{n \to \infty} E(\sqrt{n}(U_n^* - (U_n - \gamma))^2 = 0 \tag{5.3}$$

but

$$nE(U_n^* - (U_n - \gamma))^2 = nVar(U_n^*) - 2nCov(U_n^*, U_n) + nVar(U_n) \tag{5.4}$$

now

$$\begin{aligned} nVar(U_n^*) &= nVar(\frac{k}{n}\sum_{i=1}^{n}(h_1(X_i)-\gamma)) \\ &= n\frac{k^2}{n^2}\sum_{i=1}^{n}Var(h_1(X_i)) \\ &= \frac{nk^2}{n^2}.n.\xi_1 \\ &= k^2\xi_1 \end{aligned}$$

Thus the first and last terms on the RHS of (5.4.2) are equal to $k^2\xi_1$,we therefore need to show that $nCov(U_n^*, U_n)$ is equal to $k^2\xi_1$.

$$nCov(U_n^*, U_n) = \frac{k}{\binom{n}{k}}\sum_{m=1}^{n}\sum_{j\in I}Cov(h_1(X_m), h(X_{j_1}, ..., X_{j_k}))$$

Now $Cov(h_1(X_m), h(X_{j_1}, ..., X_{j_k})$ is equal to zero if m is not equal to one of the $j_i's$, and it is ξ_1 otherwise (From the development of the general expression of the variance of U-statistics for an estimable parameter γ). For fixed m the number of sets $\{i_1, ..., i_k\}$ containing m is $\binom{n-1}{k-1}$ and since there are n such m

$$ncov(U_n^*, U_n) = \frac{k}{\binom{n}{k}}n\binom{n-1}{k-1}\xi_1 = k^2\xi_1$$

so that (5.4.1) is satisfied. Therefore by Slutsky's theorem

$$\sqrt{n}(U_n - \gamma) \longrightarrow N(0, k^2\xi_1)$$

and this completes the proof.

5.5. Proof of Hoeffding's One Sample U-statistics Theorem as given by Beutner and Zahle(2012)

Define the U-statistics for an estimable parameter γ of degree k by

$$U_n = \frac{1}{\binom{n}{k}}\sum_{i\in I}h(X_{i_1}, ..., X_{i_k})$$

Recall that the projection of $U_n - \gamma$ on the space $\mathcal{U}$ is given by

$$U_n^* = \frac{k}{n}\sum_{j=1}^{n}\{h_1(X_j) - \gamma\}$$

now we have that

$$E(U_n - \gamma | X_j) = \frac{1}{\binom{n}{k}}\sum_{i \in I} E[h(X_{i_1}, ..., X_{i_k}) - \gamma | X_j]$$

now from (5.3.1) $E[h(X_{i_1}, ..., X_{i_k}) - \gamma | X_j]$ equals $h_1(X_j)$ whenever j is among $\{i_1, ..., i_k\}$ and 0 otherwise. The number of ways to choose $\{i_1, ..., i_k\}$ so that j is among them is $\binom{n-1}{k-1,}$ so we obtain

$$\begin{aligned} E(U_n - \gamma | X_j) &= \frac{\binom{n-1}{k-1}}{\binom{n}{k}}(h_1(X_j) - \gamma) \\ &= \frac{k}{n}(h_1(X_j) - \gamma) \end{aligned}$$

Thus $\sqrt{n}U_n^* = \frac{k}{\sqrt{n}}\sum_{j=1}^{n}\{h_1(X_j) - \gamma\}$. We thus have that

$$\begin{aligned} E[U_n^*(U_n - \gamma)] &= \frac{k}{n}\sum_{j=1}^{n} E[(h_1(X_j) - \gamma)(U_n - \gamma)] \\ &= \frac{k}{n}\sum_{j=1}^{n} E\{E[(h_1(X_j) - \gamma)(U_n - \gamma)/X_j]\} \\ &= \frac{k}{n}\sum_{j=1}^{n} E[(h_1(X_j) - \gamma)\frac{k}{n}(h_1(X_j) - \gamma)] \\ &= \frac{k^2}{n^2}\sum_{j=1}^{n} E(h_1(X_j) - \gamma)^2 \\ &= \frac{k^2}{n^2}\sum_{j=1}^{n} Var(h_1(X_j) - \gamma) \\ &= \frac{k^2}{n}\xi_1 \end{aligned}$$

Now if we consider the representation

$$\sqrt{n}(U_n - \gamma) = \sqrt{n}U_n^* + \epsilon_n = \frac{k}{\sqrt{n}}\sum_{j=1}^{n}\{h_1(X_j) - \gamma\} + \epsilon_n$$

where $h_1(X_j)$ is as given in (5.3.1) and $\epsilon_n = \sqrt{n}(U_n - \gamma) - \frac{k}{\sqrt{n}}\sum_{j=1}^{n}\{h_1(X_j) - \gamma\}$ where $Var(h_1(X_j)) = \xi_1$. Now $h_1(X_j)'s$ are iid random variables, and hence by the central limit theorem (CLT)

$$\frac{k}{\sqrt{n}}\sum_{j=1}^{n}\{h_1(X_j) - \gamma\} \longrightarrow N(0, k^2\xi_1) \quad \text{in distribution.}$$

Then the proof will be complete if we can establish that $\epsilon_n \longrightarrow 0$ in probability as $n \longrightarrow \infty$. Now by Definition 2.1.6 and Theorem 2.1.12, ϵ_n converges in probability to zero if $Var(\epsilon_n) \to 0$ as $n \to \infty$, but $Var(\epsilon_n) = E(\epsilon_n^2) - [E(\epsilon_n)]^2$. Since $E(\epsilon_n) = 0$, it will be enough to show that $E(\epsilon_n^2) \longrightarrow 0$.

Now

$$\begin{aligned} E(\epsilon_n)^2 = Var\{\sqrt{n}(U_n - \gamma)\} + Var\left[\frac{k}{\sqrt{n}}\sum_{j=1}^{n}\{h_1(X_j) - \gamma\}\right] \\ -2Cov\left(\sqrt{n}(U_n - \gamma), \frac{k}{\sqrt{n}}\sum_{j=1}^{n}\{h_1(X_j) - \gamma)\}\right). \end{aligned} \tag{5.5}$$

But

$$\begin{aligned} Var\left[\frac{k}{\sqrt{n}}\sum_{j=1}^{n}\{h_1(X_j) - \gamma\}\right] &= \frac{k^2}{n}\sum_{j=1}^{n} Var(h_1(X_j)) \\ &= \frac{k^2}{n}.n.\xi_1 \\ &= k^2\xi_1 \end{aligned}$$

The first term on the right hand side of (5.5.1) converges in probability to $k^2\xi_1$ by Theorem 3.2.1, the next term is equal to $k^2\xi_1$.
We therefore need to show that $Cov\left(\sqrt{n}(U_n - \gamma), \frac{k}{\sqrt{n}}\sum_{j=1}^{n}\{h_1(X_j) - \gamma)\}\right) = k^2\xi_1$. Now

$$Cov\left(\sqrt{n}(U_n - \gamma), \frac{k}{\sqrt{n}}\sum_{j=1}^{n}\{h_1(X_j) - \gamma)\}\right) = \frac{k}{\binom{n}{k}}\sum_{j=1}^{n} Cov\{h(X_{i_1}, ..., X_{i_k}), h_1(X_j)\} \tag{5.6}$$

but $Cov\{h(X_1, ..., X_k), h_1(X_j)\} = E(h(X_{i_1}, ..., X_{i_k})h_1(X_j)) = \xi_1$ if $i_1 = i$ or ... or $i_k = i$, and zero otherwise. For a fixed i, the number of sets $\{i_1, ..., i_k\}$ such that $1 \leq i_1 < ... < i_k \leq n$ is $\binom{n-1}{k-1}$, since there are n such co-variances terms with value ξ_1 we thus have the RHS of (5.5.2) equal to $k^2\xi_1$. Thus $E(\epsilon_n^2) \longrightarrow 0$ as $n \longrightarrow \infty$. Therefore $\epsilon_n \longrightarrow 0$ in probability. We therefore conclude that $\sqrt{n}(U_n - \gamma)$ and $\frac{k}{\sqrt{n}}\sum_{j=1}^{n}\{h_1(X_j) - \gamma)$ have the same asymptotic distribution. This concludes the proof.

CHAPTER 6 : DISCUSSION AND CONCLUSION

The small sample properties of the U-statistics for an estimable parameter γ examined shows that apart from being sufficient and efficient, the U-statistics for an estimable parameter γ is also a minimum variance Unbiased estimator (MVUE) of γ. The large sample properties examined showed that the U-statistics for an estimable parameter γ is a Consistent Asymptotic Normal (CAN) estimator. We therefore conclude that the U-statistics for an estimable parameter γ has the qualities of a good estimator. With reference to construction of test statistics, the large sample properties of the U-statistics for an estimable parameter γ can be used to construct the test statistics when testing for symmetry as the following example suggests. Let $X_1, X_2, ...X_n$ be a random sample from the continuous distribution and suppose we wish to test if the distribution is symmetric about zero. The Wilcoxon one-sample rank test statistic used to test if the distribution is symmetric about zero is given by

$$T^+ = \sum_{i=1}^{n} R_i I[X_i > 0]$$

where R_i is the rank(position when $|X_1|, ..., |X_n|$ are arranged in ascending order). Since T^+ is the sum of U-statistics (Ferguson, 1996), then we can use the asymptotic properties of U-statistics to construct the critical region for the test.

If the parameter to be estimated is $\gamma = E(h(X_1, X_2)) = P(X_1 + X_2 > 0)$ then we have that $\xi_1 = Cov(h(X_1, X_2)h(X_1, X_3)) = P(X_1 + X_2 > 0, X_1 + X_3 > 0) - \gamma^2$. Under the null hypothesis that the distribution is continuous and symmetric about 0, we have $\gamma = \frac{1}{2}$ and $P(X_1 + X_2 > 0, X_1 + X_3 > 0) = P(X_1 > -X_2, X_1 > -X_3) = P(X_1 > X_2, X_1 > X_3) = \frac{1}{3}$, since this is the probability that of three identically distributed random variables, the first random variable(X_1) is the largest. Therefore, under the null hypothesis, $\xi_1 = \frac{1}{3} - (\frac{1}{2})^2 = \frac{1}{12}$ applying Hoeffdings one sample U-statistics theorem we have that $\sqrt{n}(U_n - 1/2) \to N(0, 1/3)$

Clearly, the order bound for U-statistics for an estimable parameter γ depends on the moments of the kernel h, and the underlying distribution of the random sample $X_1, X_2, ..., X_n$. Bentkus, Gotze and Zitikis (1994) established that the sharpest Bessy-Esseen bound of order $O(n^{-1/2})$ for U-statistics for an estimable parameter of order 2 depends on the moment conditions $E|E(h(X_1, X_2)|X_1)|^3 < \infty$ and $E|h(X_1, X_2)|^{5/3} < \infty$. They established that the weaker the moment conditions $E|E(h(X_1, X_2)|X_1)|^3 < \infty$ and $E|h(X_1, X_2)|^{5/3} < \infty$, the sharper the Berry-Esseen bound of order $O(n^{-1/2})$ for U-statistics of an estimable parameter of degree 2. Now the Berry-Esseen bound for U-statistics for the population variance σ^2 discussed in section 4.3 clearly depends on the two moment conditions $E|E(h(X_1, X_2)|X_1)|^3 < \infty$ and $E|h(X_1, X_2)|^{5/3} < \infty$. We can therefore conclude that the Berry-Esseen bound of order $O(n^{-1/2})$ for the U-statistics for the population variance discussed in Theorem 4.2.2 is the sharpest with respect to the moment conditions $E|E(h(X_1, X_2)|X_1)|^3 < \infty$ and $E|h(X_1, X_2)|^{5/3} < \infty$.

Now coming to the rate of convergence to normal for the method of moments estimator(S^2) for the population variance (σ^2). We notice that the order bound for the method of moments estimator for the population variance does not depend on the moments of S^2 as can be seen from Corollary 4.3.2. The rate of convergence to the normal distribution of the U-statistics for the population variance is given by $CV_3\sigma_g^{-3}n^{-1/2}$ as can be seen from 4.2.4. The rate of convergence to the normal distribution of the method of moments estimator for the population variance is given by $(a+b+C)n^{-1/2}$ as can be seen from Corollary 4.3.2 (ii). Comparing the two rates of convergence we note that they are both of order $n^{-1/2}$ and that $(a+b+C)n^{-1/2}$ is smaller than $CV_3\sigma_g^{-3}n^{-1/2}$ since $(a+b+)C < CV_3\sigma_g^{-3}$. Even though the U-statistics and the method of moments estimator for the population variance are asymptotically equivalent, the method of moments estimator for the population variance converges at a faster rate to normal as compared to the U-statistics for the population variance.

Next we review the different proofs of Hoeffdings' one sample U-statistics theorem as given in the period 1948-2012. Since U_n is not a sum of iid random variables the large sample distribution cannot be obtained by a straight application of central limit theorem. Therefore, the authors for the different proofs of Hoeffdings, one sample U-statistics theorem used different approaches in making sure that the U_n is expressed in terms of the sum of iid random variables which are asymptotically normally distributed. The different proofs of Hoeffdings' one sample U-statistics theorem discussed above suggest that for one to develop an alternative proof of the theorem, one needs to find a variable with a limiting normal distribution and is asymptotically equivalent to $\sqrt{n}(U_n-\gamma)$. Slutsky's theorem is then used in the proof. Hoeffding used the random variable $Y_n = \frac{k}{\sqrt{n}}\sum_{j=1}^{n}(h_1(X_j)-\gamma)$ which is asymptotically similar to U_n in his proof, Lee used $\hat{U} = \frac{k}{n}\sum_{i=1}^{n}h_1(X_i)$, Ferguson used $U_n^* = \frac{k}{n}\sum_{j=1}^{n}(h_1(X_j)-\gamma)$ and finally Beutner and Zahle used $\frac{k}{\sqrt{n}}\sum_{j=1}^{n}(h_1(X_j)-\gamma)+\epsilon_n$.

In reviewing the different proofs of Hoeffding's one sample U-statistics theorem, the following can be said; All the proofs are based on the projection principle approach after its inventor Hoeffding. The approach Ferguson (1996) used in showing that $\sqrt{n}(U_n-\gamma-\hat{U}_n) \to^p 0$ is different from that used by Lee (1990). Ferguson (1996) used the fact that convergence in quadratic mean implies convergence in probability while Lee (1990) used the fact that $\lim_{n\to\infty}\frac{Var(\hat{U_n})}{Var(U_n)} = 1$ so that $\frac{U_n-\gamma}{\sqrt{var(U_n)}} - \frac{\hat{U_n}}{\sqrt{var(\hat{U_n})}} \to^p 0$. The approach Beutner and Zahle (2012) used to prove Hoeffdings' one sample U-statistics theorem seems to be developed from the one by Hoeffding(1948). Hoeffding used the random variable $Y_n = \frac{k}{\sqrt{n}}\sum_{j=1}^{n}(h_1(X_j)-\gamma)$ while for Beutner and Zahle (2012) there is an extra term ϵ_n. The ϵ_n term in the proof by Beutner and Zahle (2012) can be substituted by any random variable that is $o_p(1)$, since if A_n is the sequence of random variables then $A_n = o_p(1)$ iff $A_n \to^p 0$. Although Beutner and Zahle and Hoeffding used different notations the approaches they used in proving the Hoeffding's one sample U-statistics theorem are similar. That is the proofs by Beutner and Zahle and Hoeffding are the same. All the different proofs use the asymptotic variance of U_n. Now it is seen in $Var(U_n)$ that the actual variance of U_n involves not only ξ_1 but also the variances $\xi_2, ..., \xi_k$. If any of these latter variances is infinite but $\xi_1 < \infty$, we have a case where the actual variance of $\sqrt{n}U_n$ is infinite and therefore its limit is infinite while the asymptotic variance is finite. Therefore, if the actual variance of U_n was used in the Hoeffding's one sample U-statistics theorem then it will not be enough to require that ξ_1 is finite.

In conclusion, the small and large sample properties of the U-statistics for any estimable parameter γ were examined. It was noted that the U-statistics U_n for an estimable parameter γ is the minimum variance unbiased estimator and is also a CAN estimator. Thus U_n has qualities of a good estimator. The rate of convergence to normal between the U-statistics for the population variance and the method of moments estimator for the population variance were compared. It was noted that the method of moments estimator for the population variance converges at a faster rate to normal as compared to the U-statistics for the population variance. Slutsky's theorem was used to derive the asymptotic normality of U_n. All the different proofs of Hoeffding's one sample U-statistics theorem were proved using Slutsky's theorem and the projection principle approach after its inventor Hoeffding (1948).

REFERENCES

[1] Ahmad, I. A.(1981). On Some Asymptotic Properties of U-statistics. Scandinavian Journal of statistics, **8**, 175-182.

[2] Bentkus, V. Gotze, F. and Zitikis, R. (1994). Lower estimates of the convergence rate for U-statistics. The Annals of Probability, **22**, 1707-1714.

[3] Bentkus, V. Jing, B. and Zhou, W. (2009). On normal approximations to U-statistics. The Annals of Probability. **37**, 2174-2199.

[4] Borovskikh, Y. V. and Koroljuk, V. S. (1989). Approximations of nondegenarate U-statistics. Theory of Probability and its application. **30**, 439-450.

[5] Berger, R.L. and Casella, G. (2002). Statistical Inference. New York. Duxbury.

[6] Beutner, E. and Zahle, H. (2012). Deriving the asymptotic distribution of U- and V-statistics of dependent data using Weighted empirical processes. Bernoulli, **18**, 803-822.

[7] Bickel, P. J.(1974). Edgeworth expansions in nonparametric statistics. The Annals of Statistics, **2**, 1-20.

[8] Bose, A. and Daguptu, R. (1994). On some asymptotic properties of U-statistics and one-sided estimates. The Annals of Probability, **22**, 1715-1724.

[9] Callaert, H. and Janssen, P. (1978). The Berry-Esseen theorem for U-statistic. The Annals of Statistics, **6**, 417-421.

[10] Chan, Y. and Wierman, J. (1977). On the Berry-Essen theorem for U-statistics. The Annals of Probability, **5**, 136-139.

[11] Feller, W. (1971). An Introduction to Probability Theory and Its Applications, Volume II. New York. Wiley.

[12] Ferguson, T.S. (1996). A course in large sample Theory, Chapman-Hall, New York

[13] Fraser, D.A.S. (1957). Nonparametric Methods in Statistics. John Wiley and Sons, New York.

[14] Grams, W. F. and Serfling, R. J.(1973). Convergence rates for U-statistics and related statistics. The Annals of Statistics, **1**, 153-160.

[15] Hoeffding, W.(1948). A class of Statistics with asymptotically normal distribution. The Annals of Mathematics and Statistics, **19**, 293-325.

[16] Hsing, T. and Wu, W. B. (2004). On weighted U-statistics for stationary processes. The Annals of Probability, **32**, 1600-1631.

[17] Imbens, G. W. (1992). An efficient method of moments estimator for discrete choice models with choice-based sampling. Econometrica, **60**, 1187-1214.

[18] Lee A, J. (1990). U-statistics: Theory and Practice. Marcel Dekker. New York.

[19] Lehmann, E.L. (1999). Elements of Large-Sample Theory. New York. Springer-Verlag.

[20] Nomachi, T. and Yamato, H. (2001). Asymptotic comparisons of U-statistics, V-statistics and Limits of Bayes estimates by Deficiencies. Journal of the Japan Statistical Society. **31**, 571-579.

[21] Serfling, R, J.(1971). The law of iterated logarithm for U-statistics and related Von Mises statistics(Abstract). The Annals of Mathematics and Statistics, **42**, 1974.

[22] Van Zwet, W. R. (1984). A Berry-Esseen bound for symmetric statistics. Probability Theory and Related Fields-Springer, **66**, 425-440.