SOME LIMIT THEOREMS ON CONDITIONAL *U*-STATISTICS AND CENSORED DATA NON-PARAMETRIC REGRESSION

Arusharka Sen

DISSERTATION SUBMITTED TO THE INDIAN STATISTICAL INSTITUTE IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF DOCTOR OF PHILOSOPHY

CALCUTTA

1993

SOME LIMIT THEOREMS ON CONDITIONAL U-STATISTICS AND CENSORED DATA NON-PARAMETRIC REGRESSION

Arusharka Sen

Dissertation submitted to the Indian Statistical Institute
in partial fulfilment of the requirements for the award of
the degree of Doctor of Philosophy
CALCUTTA
1993

TO :

THE INTERESTED READER

ACKNOWLEDGEMENTS

I would like to begin by recording my debt to two persons whose contribution to my career as a researcher has been the most significant : Professor B.L.S. Prakasa Rao, my supervisor, who introduced me to my research topic and spent long hours listening to my ideas and putting me on the right track; Professor T. Parthasarathy who recommended me for a research fellowship and restored my confidence – at a time when I needed it most – in my ability to do research.

I am indebted to my teachers at I.S.I., Calcutta and Delhi, who exposed me to the frontiers of modern Probability and Statistics, and to whom I have turned many a times for help and clarifications. I also thank the staff at these two centres for being most cooperative and Mr. S.P. Jakhmola of the Delhi centre for typing a part of this thesis.

I thank National Board for Higher Mathematics, Bombay, for providing me with a Research Award.

The city of Delhi and my 6-year stay at the I.S.I. Delhi hostel have taught me a great deal. I am grateful to my friends R. Sridhar, Saikat Datta, Anilesh Mohari, Rajarama Bhat, Abhay Bhatt, Anish Sarkar and Tirthankar Bhattacharyya who were always ready to help and would often share with me the ups and downs of their own research, thereby giving me an added stimulus. I must also mention Shashi Brahmankar, Abhijit Datta, Pinaki Das, Deepanwita Chattopadhyay, Arbind Lal, Uma Datta, Sanjib Pohit, Anindita Mukhopadhyay, Arup Pal, Samir Neogy and many others who make up a large part of my Delhi years.

I am grateful to my parents and my brother for their encouragement and unqualified support in whatever choices I have made regarding my career so far.

And finally, ... to Udayā — for many things ...

– Arusharka Sen

Contents

1 **Introduction** 3

2 **Uniform Strong Consistency of Conditional U-statistics** 10

2.1 Introduction 10

2.2 Assumptions and preliminary results 11

2.3 The main result 15

2.4 Supplementary remarks 23

2.5 Appendix : rudiments of empirical processes 24

3 **Limit Distributions of Conditional U-statistics** 28

3.1 Introduction 28

3.2 Symmetric tensor products and multiple Wiener integrals 30

3.3 Limit distributions of $U_n(h, \mathbf{t})$ 34

3.4 Limit laws of $U_n^h(\mathbf{t})$ 61

3.5 An example and some remarks 67

4 **Censored Data and the 'Martingale Methods'** 69

4.1 Introduction 69

4.2 The compensator of $P_n(t)$ and related predictable variation processes 72

4.3 Weak uniform consistency and asymptotic normality of $m_n^0(t)$ 78

4.4 Some remarks on the extension of the 'martingale methods' to conditional U-statistics . . . 86

5 Problems for Further Research **95**

5.1 A law of the iterated logarithm for $U_n^h(\mathbf{t})$ . . . 95

5.2 Limit distribution of the uniform absolute deviation . . . 97

6 Bibliography **99**

Chapter 1

Introduction

In Statistics, a classical problem is that of estimating the *regression function* which is defined as

$$m(t) := E(Y|X = t),\ t \in I\!R,$$

for two random variables X and Y such that $E|Y| < \infty$. The estimators are constructed based on a sample $\{(X_i, Y_i)\}, 1 \leq i \leq n, n \geq 1$, from the distribution of (X, Y). Throughout this thesis, we assume X and Y to be real-valued for the sake of convenience.

The classical approach to this problem is to assume a parametrized, polynomial form for $m(\cdot)$, i.e., $m(t) := \beta_0 + \sum_{j=1}^{p} \beta_j t^j,\ p \geq 1$, and obtain estimates of the unknown parameters $\beta_0, \beta_j,\ 1 \leq j \leq p$. Later, with the development of techniques for ***non-parametric density estimation***, it was sought to extend these techniques to regression estimation. Heuristically, the two problems can be seen to be related as follows : let $f_1(\cdot)$ be the marginal density of X and note that

$$E1(X \leq x) = \int_{-\infty}^{x} f_1(t)dt,\ x \in I\!R, \tag{1.0.1}$$

whereas

$$EY1(X \leq x) = \int_{-\infty}^{x} m(t)f_1(t)dt,\ x \in I\!R. \tag{1.0.2}$$

In other words, (1.0.1) can be looked upon as a special case of (1.0.2), with $Y \equiv 1$. (This similarity, as we shall see later on, has been the underlying theme in Chapters 2 and 4 of the present work.) The following *non-parametric regression estimator* was proposed independently by Nadaraya (1964) and Watson (1964):

$$m_n^{NW}(t) := m_n(Y,t)/m_n(1,t),\ t \in \mathbb{R}, \tag{1.0.3}$$

where

$$\left.\begin{array}{rcl} m_n(Y,t) & = & (na_n)^{-1}\sum_{i=1}^n Y_i K((t-X_i)/a_n), \\ m_n(1,t) & = & (na_n)^{-1}\sum_{i=1}^n K((t-X_i)/a_n). \end{array}\right\} \tag{1.0.4}$$

Here $K(\cdot)$, the so-called *kernel function*, is chosen to satisfy various analytical conditions (typically, $K(\cdot)$ is taken to be a density function), and $a_n \downarrow 0$ are the *bandwidths* which go to zero sufficiently slowly (e.g., $na_n \to \infty$ as $n \to \infty$) in order to ensure consistency of the estimator $m_n^{NW}(\cdot)$. The intuition behind such an estimator is that $m_n(Y,\cdot)$ is an estimator of $m(\cdot)f_1(\cdot)$ while $m_n(1,\cdot)$ estimates the density $f_1(\cdot)$. See Prakasa Rao (1983), Chapters 1–4, for an introduction to non-parametric density and regression estimation.

Now, $m(t)$ is a functional of the conditional distribution of Y, given $X = t$. A natural generalisation of the regression estimation problem seems to be the estimation of the following functionals:

$$m^h(t_1,\ldots,t_k) := E\{h(Y_1,\ldots,Y_k) \mid X_1 = t_1,\ldots,X_k = t_k\}, (t_1,\ldots,t_k) \in \mathbb{R}^k, k \geq 1, \tag{1.0.5}$$

where $h : \mathbb{R}^k \to \mathbb{R}$ is such that $E|h(Y_1,\ldots,Y_k)| < \infty$. A similar generalisation led Hoeffding (1948) from the sample mean to the theory of so-called *U-statistics*, in the *unconditional* set-up. The estimation of (1.0.5) were considered, for the first time in published form, in Stute (1991) where the following **conditional** U**-statistics** were proposed as estimators:

$$U_n^h(\mathbf{t}) := U_n(h,\mathbf{t})/U_n(1,\mathbf{t}),\ \mathbf{t} = (t_1,\ldots,t_k) \in \mathbb{R}^k, \tag{1.0.6}$$

where

$$U_n(h,\mathbf{t}) := (n)_k^{-1} a_n^{-k} \sum_{\beta(n,k)} h(Y_{\beta(1)},\ldots,Y_{\beta(k)}) \prod_{j=1}^k K((t_j - X_{\beta(j)})/a_n), n \geq k,$$

and $U_n(1, \mathbf{t})$ is obtained by putting $h \equiv 1$ in $U_n(h, \mathbf{t})$ as before. Here

$$\begin{aligned} \beta(n,k) &:= \{(\beta(1),\ldots,\beta(k)) : 1 \le \beta(i) \le n, \beta(i) \neq \beta(j), \forall i \neq j\}, \\ (n)_k &:= \text{card}(\beta(n,k)) = n!/(n-k)!. \end{aligned}$$

Earlier, however, in the thesis of Bochynek (1987), a student of Stute, *nearest neighbour* conditional U-statistics were considered. Those were as follows :

$$\tilde{U}_n^h(\mathbf{t}) := (n)_k^{-1} a_n^{-k} \sum_{\beta(n,k)} h(Y_{\beta(1)}, \ldots, Y_{\beta(k)}) \prod_{j=1}^{k} K((F_n(t_j) - F_n(X_{\beta(j)}))/a_n), n \ge k,$$

where $F_n(\cdot) := n^{-1} \sum_{i=1}^n \mathbf{1}(X_i \le \cdot)$ denotes the *empirical distribution function* (e.d.f). Bochynek discussed the asymptotic normality of conditional U- and V-statistics and performed simulation studies on them. Stute (1991) established weak and strong *pointwise* consistency and asymptotic normality of $U_n^h(\mathbf{t})$. Liero (1991) studied *uniform* strong consistency of conditional U-statistics and established asymptotic normality of the *integrated squared error* (ISE) statistic:

$$\int_A (U_n^h(\mathbf{t}) - m(\mathbf{t}))^2 w(\mathbf{t}) d\mathbf{t}$$

for suitable $A \subset \mathbb{R}^k$ and weight function $w(\cdot)$.

We quote the following examples to illustrate the possible use of conditional U-statistics. See Stute (1991) and Bochynek (1987) for other examples. Throughout this thesis, our set-up will be as follows: $\{(X_n, Y_n)\}_{n \ge 1}$ is a bi-variate i.i.d sequence, with (X_1, Y_1) having joint density $f(\cdot,\cdot)$ and X_1 having marginal density $f_1(\cdot)$. Consequently,

$$m^h(\mathbf{t}) = \int_{\mathbb{R}^k} h(y_1, \ldots, y_k) \left(\prod_{j=1}^{k} f(t_j, y_j) \right) \left(\prod_{j=1}^{k} f_1(t_j) \right)^{-1} dy_1 \ldots dy_k. \qquad (1.0.7)$$

In the sequel, however, we drop 'h' from the superscript of $m^h(\cdot)$.

Example 1.1. (Stute (1991)). Suppose we want to estimate $\text{var}(Y_1 | X_1 = t)$. Let

$$h(y_1, y_2) = (y_1 - y_2)^2/2.$$

Then

$$m(t,t) = E(h(Y_1,Y_2)|X_1 = t, X_2 = t) = \mathrm{var}(Y_1|X_1 = t),$$

and the corresponding $U_n^h(\mathbf{t})$ can be used as an estimator.

Example 1.2. (Bochynek (1987)). Let

$$h(y_1,y_2,y_3) = \mathbf{1}(y_1 - y_2 - y_3 > 0).$$

(Here, and elsewhere in this work, $\mathbf{1}(\cdot)$ denotes the indicator function.) Then

$$m(t,t,t) := P(Y_1 > Y_2 + Y_3|X_1 = t, X_2 = t, X_3 = t),$$

and the corresponding $U_n^h(t)$ can be looked upon as a conditional analogue of the Hollander-Proschan test-statistic (Hollander and Proschan (1972)). It may be used to test the hypothesis that the conditional distribution of Y_1, given $X_1 = t$, is exponential, against the alternative that it is of the NBU (New -Better than-Used) type.

Stute (1991) remarks, "Generally speaking, we may take for h any function which has been found interesting in the unconditional set-up; ..."

In Chapter 2 of this work, we establish *uniform strong consistency* of $U_n^h(\mathbf{t})$, i.e., almost sure convergence to zero of

$$\sup_{\mathbf{t}\in C} \left|U_n^h(\mathbf{t}) - m(\mathbf{t})\right|,$$

where C belongs to a class of compact subsets of $\mathbb{R}^k$. We take the following approach: just as $m_n(1,t)$ in (1.0.4) can be expressed as

$$m_n(1,t) = \int_{\mathbb{R}} a_n^{-1} K((t-x)/a_n)dF_n(x),$$

where $F_n(\cdot)$ is the e.d.f., we can similarly write

$$U_n(h,\mathbf{t}) = \int_{\mathbb{R}^k} a_n^{-k} \prod_{j=1}^{k} K((t_j - x_j)/a_n)\mu_n(\mathbf{dx}|h),$$

where $\mu_n(\mathbf{dx}|h)$ is the empirical measure defined in Section 2.2, Chapter 2. We then obtain an exponential tail-probability bound for $\mu_n(\cdot|h)$, similar to the DKW bound for $F_n(\cdot)$ (see

Dvoretzky, Kiefer and Wolfowitz (1956)). We apply this bound to

$$|U_n(h,\mathbf{t}) - EU_n(h,\mathbf{t})|$$

via an integration by parts. As a tool, we employ a probability inequality for empirical processes, due to Alexander (1984). In order to introduce and verify the conditions of his result, we present a brief review of the Vapnik-Chervonenkis theory of empirical processes on general spaces, as well as a simple lemma, in the Appendix of Chapter 2.

In Chapter 3, we study the limit distributions of $U_n^h(\mathbf{t})$ for a fixed $\mathbf{t}$. Stute (1991) obtained conditions for asymptotic normality of $U_n^h(\mathbf{t})$. We may term this the 'first order' case. In Observation 2 in the introduction of Chapter 3, we show by an example (in fact, Example 4.1 of Stute (1991)) that 'degeneracy' (i.e., the variance of the limiting normal distribution being zero) *can* occur in the case of conditional U-statistics also. So, following Dynkin and Mandelbaum (1983) who considered classical U-statistics, we give a general description of the limit distributions in terms of *multiple Wiener integrals* upto order k (order 1 corresponds to asymptotic normality). While we use the classical multiple Wiener integral when $\mathbf{t} = (t_1, \ldots, t_k)$ has all its components identical ($t_1 = \cdots = t_k = t$, say), we have to use a generalized version of the integral when the components are not identical. Through the use of Wiener integral we get, in particular, a better insight into the expression for limiting variance in Theorem 1 of Stute (1991). It turns out to be the second moment of the Wiener integral (of order 1) of an appropriate function (cf. Theorem 3.4.1 in Chapter 3).

In Chapter 4, we consider the situation where the data (i.e., the sequence $\{(X_n, Y_n)\}_{n\geq 1}$) is subject to bivariate *right random censoring*. Examples of bivariate censoring include: times for two types of non-catastrophic failure in a complex system (Cuzick (1982)), life-times of relatives and response times in two successive courses of treatment for the same patient (Mielniczuk (1991)). The censoring mechanism is explained in the introduction to Chapter 4. This chapter is different in spirit from the previous two, in that it deals primarily with *regression estimation* – which is a special case of (1.0.5) (with $k = 1$) – and contains basically a *negative* statement on the case $k \geq 2$. The regression problem was also studied by Mielniczuk (1991). Here, we employ 'martingale methods'. In the regression case, we get hold of an empirical *submartingale* $P_n(t), t \geq 0$, whose *compensator* $A_n(t),\ t \geq 0$, is the integral of the product of $m(\cdot)$, the so-called *hazard function* and a certain *predictable* process. Motivated by this fact, we construct a kernel estimator for $m(\cdot)$, along the lines of Ramlau-Hansen (1983) who proposed a kernel estimator for the *intensity* of a counting process. Using Doob's maximal inequality and a central limit theorem for semi-martingales, due to Liptser and Shiryayev (1980), we establish *weak uniform consistency* and asymptotic normality of our estimator. However, the unknown survival function of the *censoring* distribution occurs in the definition of our basic sub-martingale, and it has to replaced by its Kaplan-Meier estimator. This creates a problem for our analysis, and we are forced to use the unnatural-looking conditions (4.3.5) and (4.3.18).

For the general case, we first restrict our attention to $k = 2$. In this case, the natural analogues of the basic sub-martingale above and the compensator are two-parameter processes. As a further simplification, we consider only those h (cf. (1.0.5)) which have a product structure, i.e., $h(y_1, y_2) = \varphi(y_1)\varphi(y_2)$. But even in this simple case, the 'sub-martingale' minus the 'compensator' (i.e., the process $L_n(t_1, t_2|\varphi)$ in Section 4.4) fails to satisfy even the weakest form of two-parameter martingale property (the so-called *weak martingales* – see Definition 4.4.2 in Chapter 4). The problem is caused by what seems to be the natural choice for an appropriate two-parameter *filtration*. Hence there does not seem to be much hope for an extension of the 'martingale methods' to the case $k \geq 2$. As a by-product of our efforts, however, we obtain a Wong and Zakai (1976)-type decomposition of '$L_{ij}(t_1, t_2|\varphi)$, where

$$L_n(t_1, t_2|\varphi) = \sum_{1 \leq i \neq j \leq n} L_{ij}(t_1, t_2|\varphi).$$

The decomposition gives a version of Doob's inequality for $L_{ij}(t_1, t_2|\varphi)$, but evidently, pooling the bounds together to obtain a similar inequality for $L_n(t_1, t_2|\varphi)$ leads to too large a bound.

Finally, in Chapter 5, we describe two problems for further investigation. The first is the problem of obtaining a *law of the iterated logarithm* (LIL) for $U_n^h(\mathbf{t})$ for a fixed $\mathbf{t}$. We outline a tentative program for attacking the problem. The second problem is that of deriving the limit distribution of the *uniform absolute deviation* criterion

$$\sup_{\mathbf{t}\in C}\left|U_n^h(\mathbf{t}) - m(\mathbf{t})\right|$$

for a suitable compact set C, with suitable normalization. Here we suggest an use of strong approximation of the empirical process $\mu_n((-\infty,\mathbf{x}]|h) - \mu((-\infty,\mathbf{x}]|h)$, $\mathbf{x} \in \mathbb{R}^k$, by an appropriate Gaussian random field.

Chapter 2

Uniform Strong Consistency of Conditional U-statistics

2.1 Introduction

In this chapter, we derive uniform strong consistency rates for $U_n^h(\mathbf{t})$ under suitable assumptions. Our proofs are based almost entirely on the results of Alexander(1984). We make use of a certain 'randomly weighted' empirical process on $\mathbb{R}^k$, which also has a *U-statistic structure*. Empirical processes of the U-statistic structure have been studied, among others, by Janssen(1988) and Schneemeier(1989). See the former for an account of statistical applications of such processes.

In Section 2.2, we state the basic assumptions and give some preliminary results. Section 2.3 contains the main result. It is in this section that we make use of the aforementioned empirical process. A discussion of some elementary concepts from the theory of empirical processes on general spaces may be found in the Appendix. The Appendix also contains a simple lemma which is rather crucial for our proofs. Finally, a few supplementary remarks are included in Section 2.4.

2.2 Assumptions and preliminary results

In order to prove uniform strong consistency results for $U_n^h(\mathbf{t}), \mathbf{t} \in \mathbb{R}^k$, we shall consider the numerator and denominator of $U_n^h(\mathbf{t})$ separately. Recall

$$U_n(h,\mathbf{t}) = (n)_k^{-1} a_n^{-k} \sum_{\beta(n,k)} h(Y_{\beta(1)}, \ldots, Y_{\beta(k)}) \prod_{j=1}^{k} K((t_j - X_{\beta(j)})/a_n). \tag{2.2.1}$$

and

$$U_n^h(\mathbf{t}) = U_n(h,\mathbf{t})/U_n(1,\mathbf{t}),$$

i.e., in the denominator $h \equiv 1$. Consider the following decomposition:

$$U_n^h(\mathbf{t}) - m(\mathbf{t}) = R_{1n}(\mathbf{t}) + R_{2n}(\mathbf{t}), \text{ say,} \tag{2.2.2}$$

where

$$\begin{aligned} R_{1n}(\mathbf{t}) &= \frac{U_n(h,\mathbf{t}) - m(\mathbf{t})\prod_{j=1}^{k} f_1(t_j)}{U_n(1,\mathbf{t})} \\ R_{2n}(\mathbf{t}) &= -m(\mathbf{t})\frac{U_n(1,\mathbf{t}) - \prod_{j=1}^{k} f_1(t_j)}{U_n(1,\mathbf{t})} \end{aligned}$$

We shall obtain rates of a.s. uniform convergence of $R_{in}(\mathbf{t}), i = 1, 2$, to zero, which will establish our main result.

To begin with, we shall assume that $m(\cdot) \in L_\infty(\mathbb{R}^k)$, $f_1(\cdot) \in L_\infty(\mathbb{R})$, which imply that both $m(\cdot)\prod_{j=1}^{k} f_1(\cdot)$ and $\prod_{j=1}^{k} f_1(\cdot)$ belong to $L_\infty(\mathbb{R}^k)$.

Further, without loss of generality, we may take h, and consequently m, to be non-negative. Otherwise, let $h = h^+ - h^-$ and $m = m^+ - m^-$. This will not change the consistency results, as is evident from (2.2.2).

Now consider

$$\begin{aligned} U_n(h,\mathbf{t}) - m(\mathbf{t})\prod_{j=1}^{k} f_1(t_j) &= U_n(h,\mathbf{t}) - EU_n(h,\mathbf{t}) \\ &\quad + EU_n(h,\mathbf{t}) - m(\mathbf{t})\prod_{j=1}^{k} f_1(t_j). \end{aligned} \tag{2.2.3}$$

Note that

$$\begin{aligned} EU_n(h,\mathbf{t}) &= a_n^{-k} E h(Y_1,\ldots,Y_k) \prod_{j=1}^{k} K((t_j - X_j)/a_n) \\ &= \int_{\mathbb{R}^k} a_n^{-k} m(x_1,\ldots,x_k) \left[\prod_{j=1}^{k} K((t_j - x_j)/a_n) f_1(x_j) \right] dx_1 \ldots dx_k. \end{aligned}$$

We can write

$$U_n(h,\mathbf{t}) = \int_{\mathbb{R}^k} a_n^{-k} \prod_{j=1}^{k} K((t_j - x_j)/a_n) \mu_n(\mathbf{dx} \mid h), \ \mathbf{dx} = dx_1 \ldots dx_k,$$

where $\mu_n(\mathbf{dx} \mid h)$ is an empirical measure on $\mathbb{R}^k$ given by

$$\begin{aligned} &\mu_n((-\infty,\mathbf{x}] \mid h) \\ = \ &(n)_k^{-1} \sum_{\beta(n,k)} h(Y_{\beta(1)},\ldots,Y_{\beta(k)}) \mathbf{1}_{(-\infty,\mathbf{x}]}(X_{\beta(1)},\ldots,X_{\beta(k)}), \ \mathbf{x} \in \mathbb{R}^k. \end{aligned}$$

Consequently, $EU_n(h,\mathbf{t})$ can be written as

$$EU_n(h,\mathbf{t}) = \int_{\mathbb{R}^k} a_n^{-k} \prod_{j=1}^{k} K((t_j - x_j)/a_n) \mu(\mathbf{dx} \mid h),$$

where

$$\begin{aligned} \mu((-\infty,\mathbf{x}]|h) &= E\mu_n((-\infty,\mathbf{x}] \mid h) \\ &= \int_{(-\infty,\mathbf{x}]} m(v_1,\ldots,v_k) \prod_{j=1}^{k} f_1(v_j) dv_1 \ldots dv_k. \end{aligned}$$

We shall establish a.s. uniform convergence to zero of $U_n(h,\mathbf{t}) - EU_n(h,\mathbf{t})$, the stochastic part in (2.2.3), by obtaining an exponential bound for

$$\Pr\left\{ \sup_{\mathbf{x} \in \mathbb{R}^k} |\mu_n((-\infty,\mathbf{x}]|h) - \mu((-\infty,\mathbf{x}]|h)| > \epsilon \right\}, \epsilon > 0.$$

The deterministic part $EU_n(h,\mathbf{t}) - m(\mathbf{t}) \prod_{j=1}^{k} f_1(t_j)$ will be taken care of by certain standard analytic assumptions. The convergence of $U_n(1,\mathbf{t}) - \prod_{j=1}^{k} f_1(t_j)$ will then follow immediately as a special case, with $h \equiv 1$. Finally, the convergence of $R_{in}(\mathbf{t}), i = 1,2$, in (2.2.2) will be dealt with using the assumptions referred to above, and some more.

We shall now state our basic assumptions. These will be followed by a lemma and an estimate, which will conclude the present section. We shall invoke other assumptions as and when they are needed.

A1.

$$m(\cdot) \in L_\infty(\mathbb{R}^k),\ f_1(\cdot) \in L_\infty(\mathbb{R}).$$

A2.

$$\begin{aligned} |m(\mathbf{x}_1) - m(\mathbf{x}_2)| &\leq C_1\|\mathbf{x}_1 - \mathbf{x}_2\|, \mathbf{x}_i \in \mathbb{R}^k, i = 1, 2; \\ |f_1(y_1) - f_1(y_2)| &\leq C_2|y_1 - y_2|, y_i \in \mathbb{R}, i = 1, 2; \end{aligned}$$

here $C_i > 0, i = 1, 2$, and $\|\mathbf{x}\| := |x_1| + \cdots + |x_k|$ for $\mathbf{x} \in \mathbb{R}^k$.

Note that by **A2** and **A1**,

$$|m(\mathbf{x}) \prod_{j=1}^{k} f_1(x_j) - m(\mathbf{y}) \prod_{j=1}^{k} f_1(y_j)| \leq C\|\mathbf{x} - \mathbf{y}\|,\ \mathbf{x}, \mathbf{y} \in \mathbb{R}^k,$$

for some $C \equiv C(m, g, k) > 0$.

A3.

$K(\cdot) \geq 0$, $\int_{\mathbb{R}} K(u)du = 1$, $\int_{\mathbb{R}} |u|K(u)du < \infty$. Further, $K(\cdot)$ is of bounded variation with $\int_{-\infty}^{\infty} |dK(x)| = V$, say.

By **A1** - **A3**, we immediately obtain the following bound:

Lemma 2.2.1.

$$\sup_{\mathbf{t}\in\mathbb{R}^k} |EU_n(h, \mathbf{t}) - m(\mathbf{t}) \prod_{j=1}^{k} f_1(t_j)| = O(a_n).$$

Proof: Observe that (use $\int_{\mathbb{R}^k} \prod_{j=1}^{k} K(x_j)d\mathbf{x} = 1$)

$$\begin{aligned} &\sup_{\mathbf{t}\in\mathbb{R}^k} |EU_n(h, \mathbf{t}) - m(\mathbf{t}) \prod_{j=1}^{k} f_1(t_j)| \\ = &\sup_{\mathbf{t}\in\mathbb{R}^k} \left| \int_{\mathbb{R}^k} m(t_1 - a_n y_1, \ldots, t_k - a_n y_k) \prod_{j=1}^{k} f_1(t_j - a_n y_j) K(y_j) d\mathbf{y} \right. \\ &\left. - m(\mathbf{t}) \prod_{j=1}^{k} f_1(t_j) \right|. \end{aligned}$$

$$
\begin{aligned}
&\le \sup_{\mathbf{t}\in \mathbb{R}^k} \int_{\mathbb{R}^k} \Bigg| m(t_1 - a_n y_1, \ldots, t_k - a_n y_k) \prod_{j=1}^{k} f_1(t_j - a_n y_j) \\
&\qquad - m(\mathbf{t}) \prod_{j=1}^{k} f_1(t_j) \Bigg| \prod_{j=1}^{k} K(y_j) d\mathbf{y} \\
&\le C a_n \int_{\mathbb{R}^k} (|y_1| + \cdots + |y_k|) \prod_{j=1}^{k} K(y_j) d\mathbf{y}, \text{ from } \mathbf{A2} \text{ and } \mathbf{A1}, \\
&= C a_n k \int_{\mathbb{R}} |y| K(y) dy \\
&= O(a_n), \text{ from } \mathbf{A3}. \square
\end{aligned}
$$

Further, we obtain the following key estimate. Consider

$$
\begin{aligned}
& U_n(h, \mathbf{t}) - EU_n(h, \mathbf{t}) \\
= & \int_{\mathbb{R}^k} a_n^{-k} \prod_{j=1}^{k} K((t_j - x_j)/a_n) \mu_n(d\mathbf{x}|h) - \int_{\mathbb{R}^k} a_n^{-k} \prod_{j=1}^{k} K((t_j - x_j)/a_n) \mu(d\mathbf{x}|h) \\
= & T_1 - T_2, \text{ say.}
\end{aligned}
$$

Then

$$
\begin{aligned}
T_1 &= \int_{\mathbb{R}^{k-1}} a_n^{-k} \prod_{j=2}^{k} K((t_j - x_j)/a_n) \left[\int_{\mathbb{R}} K((t_1 - x_1)/a_n) \mu_n(dx_1 \ldots dx_k|h) \right] \\
&= \int_{\mathcal{R}^{k-1}} a_n^{-k} \prod_{j=2}^{k} K((t_j - x_j)/a_n) \Big[K((t_1 - x_1)/a_n) \mu_n(x_1 dx_2 \ldots dx_k|h)|_{-\infty}^{\infty} \\
&\quad - \int_{\mathbb{R}} dK((t_1 - x_1)/a_n) \mu_n(x_1 dx_2 \ldots dx_k|h) \Big]
\end{aligned}
$$

by applying *integration by parts* to the integral inside the square-brackets, using **A3**. Note that

$$
\mu_n(x_1 dx_2 \ldots dx_k|h) = \int_{(-\infty, x_1]} \mu_n(dz dx_2 \ldots dx_k|h)
$$

is an empirical measure on $\mathbb{R}^{k-1}$ for each $x_1 \in \mathbb{R}$. Now, our assumptions on $K(\cdot)$ imply that $\lim_{x \to \pm\infty} K(x) = 0$. Further,

$$
\begin{aligned}
\lim_{x_1 \to -\infty} \mu_n(x_1 dx_2 \ldots dx_k|h) &= 0 \\
\lim_{x_1 \to \infty} \mu_n(x_1 dx_2 \ldots dx_k|h) &= (n)_k^{-1} \textstyle\sum_{\beta(n,k)} h(Y_{\beta(1)}, \ldots, Y_{\beta(k)}). \\
&\quad 1(X_{\beta(2)} \in dx_2, \ldots, X_{\beta(k)} \in dx_k)
\end{aligned}
$$

These arguments imply that

$$K((t_1 - x_1)/a_n)\mu_n(x_1 dx_2 \ldots dx_k|h)|_{-\infty}^{\infty} = 0.$$

Repeating the above arguments,

$$T_1 = (-1)^k a_n^{-k} \int_{\mathbb{R}^k} \mu_n((-\infty, \mathbf{x}]|h) \prod_{j=1}^{k} dK((t_j - x_j)/a_n).$$

Similarly,

$$T_2 = (-1)^k a_n^{-k} \int_{\mathbb{R}^k} \mu((-\infty, \mathbf{x}]|h) \prod_{j=1}^{k} dK((t_j - x_j)/a_n).$$

Thus

$$\begin{aligned} & \sup_{\mathbf{t}\in\mathbb{R}^k} |U_n(h,\mathbf{t}) - EU_n(h,\mathbf{t})| \\ \leq\ & a_n^{-k} \sup_{\mathbf{x}\in\mathbb{R}^k} |\mu_n((-\infty,\mathbf{x}]|h) - \mu((-\infty,\mathbf{x}]|h)| \int_{\mathbb{R}^k} \prod_{j=1}^{k} |dK((t_j - x_j)/a_n)| \\ =\ & V^k a_n^{-k} D_n \end{aligned} \quad (2.2.4)$$

where

$$D_n := \sup_{\mathbf{x}\in\mathbb{R}^k} |\mu_n((-\infty,\mathbf{x}]|h) - \mu((-\infty,\mathbf{x}]|h)|.$$

Hence, in view of estimate (2.2.4), we shall now proceed to obtain an almost sure order bound for D_n.

In doing so, our main tool will be a sharp probability inequality proved by Alexander(1984). We apply his result in the next section. A few elementary definitions and facts from the *Vapnik-Chervonenkis* theory of empirical processes on general spaces, which are needed for an application of his result, are presented in the Appendix for the sake of completeness.

2.3 The main result

We begin with the following lemma, which is an easy consequence of Theorem 2.11 of Alexander(1984). In this section, some elementary concepts from the theory of empirical processes on general spaces have been used. See the Appendix for a discussion, and for the meaning of the technical terms (emphasised) used in this section.

Lemma 2.3.1. *Let $(\mathcal{X}, \mathcal{S})$ be a measurable space. Let $(X_i, U_i)_{i=1}^n$ be a sample of independent, identically distributed bivariate random vectors such that $X_i \in \mathcal{X}, 0 \leq U_i \leq M_n, 1 \leq i \leq n$ a.s., where $0 < M_n \uparrow$ as $n \to \infty$. Let $\mathcal{A} \subseteq \mathcal{S}$ be a* V-C class *such that $v(\mathcal{A}) = v_0$ and the class of functions*

$$\varphi_A(x, u) = \frac{u}{M_n} \mathbf{1}_A(x), \ A \in \mathcal{A}$$

is deviation measurable. *Then*

(i)

$$\Pr\{\sup_{A\in\mathcal{A}} |n^{-1} \textstyle\sum_{i=1}^n U_i \mathbf{1}_A(X_i) - EU_1 \mathbf{1}_A(X_1)| > \epsilon\}$$
$$\leq \ 16(nM_n^{-2}\epsilon^2)^{v_0 2^{11}} \exp(-2nM_n^{-2}\epsilon^2), \text{ for } \epsilon \geq 8M_n/\sqrt{n}.$$

(ii) Let $G_n = \sup_{A\in\mathcal{A}} |n^{-1} \sum_{i=1}^n U_i \mathbf{1}_A(X_i) - EU_1 \mathbf{1}_A(X_1)|$; then $\forall\ 0 < s < 2$,

$$E \exp((2-s)nM_n^{-2}G_n^2) \leq K_0(s, v_0)$$

where $K_0(s, v_0)$ is a constant depending only on s and v_0.

Proof:

(i) Note that

$$\Pr\left\{\sup_{A\in\mathcal{A}} |n^{-1} \sum_{i=1}^n U_i \mathbf{1}_A(X_i) - EU_1 \mathbf{1}_A(X_1)| > \epsilon\right\}$$
$$= \ \Pr\left\{\sup_{A\in\mathcal{A}} \left|n^{-1} \sum_{i=1}^n \frac{U_i}{M_n} \mathbf{1}_A(X_i) - E\frac{U_1}{M_n} \mathbf{1}_A(X_1)\right| > \epsilon M_n^{-1}\right\}. \ \ .$$

Now the class of functions $\{\varphi_A : A \in \mathcal{A}\}$ satisfies $0 \leq \varphi_A \leq 1 \ \forall A \in \mathcal{A}$ and is a *V-C graph class* by Lemma 2-A of Appendix, with the *graph region class* having *V-C number* v_0. Hence the inequality follows by Theorem 2.11 of Alexander (1984).

(ii) Using (i), we get

$$\int_1^\infty \Pr\{\exp((2-s)nM_n^{-2}G_n^2) > \theta\} d\theta$$

$$
\begin{aligned}
&= \int_1^\infty \Pr\{G_n > (M_n/\sqrt{n})\sqrt{\ln\theta/(2-s)}\}d\theta \\
&= \int_{\{\sqrt{\ln\theta/(2-s)}\le 8\}}(\ldots) + \int_{\{\sqrt{\ln\theta/(2-s)}>8\}}(\ldots) \\
&\le e^{64(2-s)} + \int_{e^{64(2-s)}}^\infty 16(\ln\theta/(2-s))^{v_0 2^{11}} \exp(-2(2-s)^{-1}\ln\theta)d\theta;
\end{aligned}
$$

the result now follows from the last inequality. Note that the exponential probability bound in (i) is valid for the range $\epsilon > 8M_n/\sqrt{n}$ only, hence the splitting of the integral at the second equality above.□

Observe that

$$\mu_n((-\infty,\mathbf{x}]|h), \mathbf{x} \in \mathbb{R}^k,$$

is an empirical process of 'U-statistic structure'. In order to be able to use the above bounds, we take recourse to the so-called 'Hoeffding decomposition' which expresses $\mu_n((-\infty,\mathbf{x}]|h)$ as an average of empirical processes of iid structure.

Let $\Sigma_n = \{$ all permutations of $\{1,2,\ldots,n\}\}$. For $\sigma \in \Sigma_n$, define

$$
\begin{aligned}
&\mu_n^\sigma((-\infty,\mathbf{x}]|h) = \\
&\quad [n/k]^{-1}\sum_{j=0}^{[n/k]-1} h(Y_{\sigma(kj+1)},\ldots,Y_{\sigma(kj+k)})\mathbf{1}_{(-\infty,\mathbf{x}]}(X_{\sigma(kj+1)},\ldots,X_{\sigma(kj+k)}).
\end{aligned}
$$

Here $[x] :=$ the greatest integer $\le x, x \in \mathbb{R}$. Let

$$D_n^\sigma = \sup_{\mathbf{x}\in\mathbb{R}^k} |\mu_n^\sigma((-\infty,\mathbf{x}]|h) - \mu((-\infty,\mathbf{x}]|h)|.$$

Then D_n^σ has a structure similar to G_n in Lemma 2.3.1. Recall

$$D_n = \sup_{\mathbf{x}\in\mathbb{R}^k} |\mu_n((-\infty,\mathbf{x}]|h) - \mu((-\infty,\mathbf{x}]|h)|.$$

Then we have

Lemma 2.3.2.

(i) The following holds:

$$\mu_n((-\infty,\mathbf{x}]|h) = (n!)^{-1}\sum_{\sigma\in\Sigma_n}\mu_n^\sigma((-\infty,\mathbf{x}]|h).$$

(ii) For all $\theta > 0$,

$$E\exp(\theta D_n^2) \leq E\exp(\theta(D_n^\sigma)^2),\ \forall\sigma\in\Sigma_n.$$

Proof:

(i) This is just the Hoeffding decomposition. For a proof see, for example, Serfling 1980), p. 180.

(ii) From (i), by monotonicity and convexity of the functions y^2 and $\exp(\theta y)$ for $y > 0, \theta > 0$, and Jensen's inequality,

$$\exp(\theta D_n^2) \leq (n!)^{-1}\sum_{\Sigma_n}\exp(\theta(D_n^\sigma)^2).$$

The result follows by taking expectations on both sides. Note that the measurability of D_n and D_n^σ can be easily established, by considering, for each $\mathbf{x} \in I\!R^k$, rationals $\mathbf{r}^{(m)}$ such that $r_i^{(m)} \downarrow x_i, 1 \leq i \leq k$, as $m \to \infty$.□

Lemma 2.3.3.

(i) Suppose $\exists M > 0$ such that $0 \leq h \leq M$ a.s. and $na_n^{2k}(\log n)^{-1} \to \infty$ as $n \to \infty$. Then

$$\sup_{\mathbf{t}\in I\!R^k}|U_n(h,\mathbf{t}) - EU_n(h,\mathbf{t})| = O(a_n^{-k}[n/k]^{-1/2}(\log n)^{1/2}),\ a.s.,$$

as $n \to \infty$. In particular,

$$\sup_{\mathbf{t}\in I\!R^k}|U_n(1,\mathbf{t}) - EU_n(1,\mathbf{t})| = O(a_n^{-k}[n/k]^{-1/2}(\log n)^{1/2}),\ a.s.$$

(ii) Suppose $Eh^2 < \infty$, and $na_n^{4k}(\log n)^{-1} \to \infty$ as $n \to \infty$. Then

$$\sup_{\mathbf{t}\in I\!R^k}|U_n(h,\mathbf{t}) - EU_n(h,\mathbf{t})| = O(a_n^{-k}[n/k]^{-1/4}(\log n)^{1/4})\ a.s.,$$

as $n \to \infty$.

Proof:

(i) Consider the class of functions

$$\mu_n^{\sigma,M}((-\infty,\mathbf{x}]|h) =$$
$$[n/k]^{-1}\sum_{j=0}^{[n/k]-1} M^{-1}h(Y_{\sigma(kj+1)},\ldots,Y_{\sigma(kj+k)})1_{(-\infty,\mathbf{x}]}(X_{\sigma(kj+1)},\ldots,X_{\sigma(kj+k)}),\mathbf{x}\in\mathbb{R}^k,$$

where $\sigma \in \Sigma_n$ as in Lemma 2.3.2.

Here (refer to Appendix) $\mathcal{X} = \mathbb{R}^k$, $\mathcal{A} = \{(-\infty,\mathbf{x}] : \mathbf{x} \in \mathbb{R}^k\}$, and for each $\sigma \in \Sigma_n$, this class satisfies the hypotheses of Lemma 2.3.1, with $v_0 = k+1$ (cf. Example 2-A.1 in the Appendix) and $M_n = M$. The *deviation measurability* is easy to establish, by considering the rationals as in Lemma 2.3.2. Hence by Lemma 2.3.1(ii) and Lemma 2.3.2(ii), with $M_n = M\ \forall n \geq 1$, we have $\forall\ 0 < s < 2$,

$$\begin{aligned} E\exp((2-s)[n/k]M^{-2}D_n^2) &\leq E\exp((2-s)[n/k]M^{-2}(D_n^\sigma)^2) \\ &\leq K_0,\ n \geq k, \end{aligned}$$

where K_0 depends only on k and s.

Now by an application of Markov's inequality, for all $B > 0$, $0 < s < 2$.

$$\begin{aligned} &\Pr\{D_n > B[n/k]^{-1/2}(\log n)^{1/2}\} \\ \leq\ &K_0\exp(-B^2(2-s)M^{-2}\log n). \end{aligned}$$

Now fixing $0 < s < 2$ and choosing $B > 0$ large enough, and by Borel-Cantelli Lemma (see the proof of (ii) below),

$$D_n = O([n/k]^{-1/2}(\log n)^{1/2})\ \text{a.s.},\ n \to \infty.$$

The result then follows from (2.2.4), from which

$$\sup_{\mathbf{t}\in\mathbb{R}^k}|U_n(h,\mathbf{t}) - EU_n(h,\mathbf{t})| = O(a_n^{-k}D_n)\ \text{a.s.}$$

(ii) This involves a little more work. The idea is adapted from Härdle et al. (1988). Let $M_n = ([n/k]/\log n)^{1/4}$. With

$$\begin{aligned} D_n(M_n) &= \sup_{\mathbf{x}\in \mathbb{R}^k} \Big| (n)_k^{-1} \sum_{\beta(n,k)} h\mathbf{1}(h \le M_n)\mathbf{1}_{(-\infty,\mathbf{x}]}(X_{\beta(1)},\ldots,X_{\beta(k)}) \\ &\quad - Eh\mathbf{1}(h \le M_n)\mathbf{1}_{(-\infty,\mathbf{x}]}(X_1,\ldots,X_k)\Big|, \end{aligned}$$

we have

$$\begin{aligned} D_n &= \sup_{\mathbf{x}\in \mathbb{R}^k} |\mu_n((-\infty,\mathbf{x}]|h) - \mu((-\infty,\mathbf{x}]|h)| \\ &\le D_n(M_n) + \sup_{\mathbf{x}\in \mathbb{R}^k} \Big| (n)_k^{-1} \sum_{\beta(n,k)} h(Y_{\beta(1)},\ldots,Y_{\beta(k)})\mathbf{1}(h > M_n)\mathbf{1}_{(-\infty,\mathbf{x}]}(X_{\beta(1)},\ldots,X_{\beta(k)}) \\ &\quad - Eh\mathbf{1}(h > M_n)\mathbf{1}_{(-\infty,\mathbf{x}]}(X_1,\ldots,X_k)\Big| \\ &\le D_n(M_n) + (n)_k^{-1} \sum_{\beta(n,k)} h(Y_{\beta(1)},\ldots,Y_{\beta(k)})\mathbf{1}(h > M_n) + Eh\mathbf{1}(h > M_n) \end{aligned} \quad (2.3.1)$$

Now,

$$\begin{aligned} &\Big[(n)_k^{-1} \sum_{\beta(n,k)} h(Y_{\beta(1)},\ldots,Y_{\beta(k)})\mathbf{1}(h > M_n) + Eh\mathbf{1}(h > M_n)\Big] \\ &= (\log n/[n/k])^{1/4} \left[(n)_k^{-1} \sum_{\beta(n,k)} hM_n\mathbf{1}(h > M_n) + EhM_n\mathbf{1}(h > M_n)\right] \\ &\le (\log n/[n/k])^{1/4} \left[(n)_k^{-1} \sum_{\beta(n,k)} h^2\mathbf{1}(h > M_n) + Eh^2\mathbf{1}(h > M_n)\right]. \end{aligned} \quad (2.3.2)$$

Then, obviously,

$$Eh^2\mathbf{1}(h > M_n) = o(1) \text{ as } n \to \infty,$$

since $M_n \to \infty$ and $Eh^2 < \infty$. Further, for fixed $M > 0$ and n large enough,

$$\begin{aligned} &0 \le (n)_k^{-1} \textstyle\sum_{\beta(n,k)} h^2\mathbf{1}(h > M_n) \le (n)_k^{-1} \sum_{\beta(n,k)} h^2\mathbf{1}(h > M) \\ \Rightarrow\ &0 \le \limsup_n (n)_k^{-1} \textstyle\sum_{\beta(n,k)} h^2\mathbf{1}(h > M_n) \le \limsup_n (n)_k^{-1} \sum_{\beta(n,k)} h^2\mathbf{1}(h > M). \end{aligned} \quad (2.3.3)$$

Since $Eh^2 < \infty$, we have by the SLLN for U-statistics (see, e.g., Serfling, 1980, p. 180)

$$(n)_k^{-1} \sum_{\beta(n,k)} h^2\mathbf{1}(h > M) \stackrel{a.s.}{\to} Eh^2\mathbf{1}(h > M) \text{ as } n \to \infty. \quad (2.3.4)$$

Thus from (2.3.3) and (2.3.4),

$$\limsup_n \ (n)_k^{-1} \sum_{\beta(n,k)} h^2 \mathbf{1}(h > M_n) \leq Eh^2 \mathbf{1}(h > M) \text{ a.s.}$$

Letting $M \to \infty$, we have

$$\lim_{n\to\infty} \ (n)_k^{-1} \sum_{\beta(n,k)} h^2(Y_{\beta(1)}, \ldots, Y_{\beta(k)}) \mathbf{1}(h > M_n) \stackrel{\text{a.s.}}{=} 0. \tag{2.3.5}$$

Thus from (2.3.1), (2.3.2) and (2.3.5),

$$D_n \leq D_n(M_n) + O([n/k]^{-1/4}(\log n)^{1/4}) \text{ a.s.} \tag{2.3.6}$$

Since $0 \leq \frac{h}{M_n}\mathbf{1}(h \leq M_n) \leq 1$, we apply Lemma 2.3.1(ii) and Lemma 2.3.2(ii) again, to get : $\forall \ 0 < s < 2$,

$$E\exp((2-s)[n/k]M_n^{-2}D_n^2(M_n)) \leq K_0, \ n \geq k,$$

where K_0 depends only on s and k.

Fix $0 < s < 2$. For $B > 0$,

$$\begin{aligned} & \Pr\{D_n(M_n) > BM_n^{-1}\} \\ \leq \ & K_0 \exp(-B^2(2-s)[n/k]M_n^{-4}) \text{ by Markov's inequality} \\ \leq \ & K_0 n^{-B^2(2-s)}. \end{aligned}$$

But $\sum_{n\geq k} K_0 n^{-B^2(2-s)} < \infty$, provided $B > (2-s)^{-1/2}$. Thus,

$$D_n(M_n) = O([n/k]^{-1/4}(\log n)^{1/4}) \text{ a.s.}, n \to \infty.$$

Hence from (2.3.6), $D_n = O([n/k]^{-1/4}(\log n)^{1/4})$ a.s. , $n \to \infty$, and from (2.2.4),

$$\sup_{\mathbf{t}\in\mathbb{R}^k} |U_n(h,\mathbf{t}) - EU_n(h,\mathbf{t})| = O(a_n^{-k}[n/k]^{-1/4}(\log n)^{1/4}) \text{ a.s.}$$

as $n \to \infty$.□

Combining the above results, we finally have :

Theorem 2.3.1. *Assume* **A1** *to* **A3**. *Define*

$$\mathcal{K} = \{C \subset \mathbb{R}^k : C \text{ compact, and } \inf_{\mathbf{t}\in C} \prod_{j=1}^{k} f_1(t_j) > 0\}.$$

Further, let $b_n(p) := a_n^{-k}[n/k]^{-1/p}(\log n)^{1/p}, p = 2, 4$. *Then*

(i) *if the conditions of Lemma 2.3.3(i) hold,*

$$\sup_{\mathbf{t}\in C} |U_n^h(\mathbf{t}) - m(\mathbf{t})| = O(\max\{b_n(2), a_n\}) \text{ a.s. }, n \to \infty;$$

(ii) *if the conditions of Lemma 2.3.3(ii) hold,*

$$\sup_{\mathbf{t}\in C} |U_n^h(\mathbf{t}) - m(\mathbf{t})| = O(\max\{b_n(4), a_n\}) \text{ a.s. }, n \to \infty,$$

for each $C \in \mathcal{K}$.

Proof:

(i) By Lemma 2.2.1,

$$\sup_{\mathbf{t}\in \mathbb{R}^k} |EU_n(h, \mathbf{t}) - m(\mathbf{t}) \prod_{j=1}^{k} f_1(t_j)| = O(a_n).$$

By Lemma 2.3.3(i),

$$\sup_{\mathbf{t}\in \mathbb{R}^k} |U_n(h, \mathbf{t}) - EU_n(h, \mathbf{t})| = O(b_n(2)) \text{ a.s.}$$

Hence,

$$\sup_{\mathbf{t}\in \mathbb{R}^k} |U_n(h, \mathbf{t}) - m(\mathbf{t}) \prod_{j=1}^{k} f_1(t_j)| = O(\max\{b_n(2), a_n\}) \text{ a.s.}$$

In particular,

$$\sup_{\mathbf{t}\in \mathbb{R}^k} |U_n(1, \mathbf{t}) - \prod_{j=1}^{k} f_1(t_j)| = O(\max\{b_n(2), a_n\}) \text{ a.s.}$$

This implies that, a.s. for large n, and $\forall \mathbf{t} \in \mathbb{R}^k$,

$$\begin{aligned} & U_n(1, \mathbf{t}) \geq \textstyle\prod_{j=1}^{k} f_1(t_j) - o(1) \\ \Rightarrow \quad & \inf_{\mathbf{t}\in C} U_n(1, \mathbf{t}) > 0 \text{ a.s. for large } n, \end{aligned} \tag{2.3.7}$$

for each $C \in \mathcal{K}$, by our hypothesis.

Consider $R_{1n}(\mathbf{t})$ and $R_{2n}(\mathbf{t})$ of decomposition (2.2.2). By the above observations and **A1**, for each $C \in \mathcal{K}$,

$$\sup_{\mathbf{t}\in C} |R_{in}(\mathbf{t})| = O(\max\{b_n(2), a_n\}) \text{ a.s. }, i = 1, 2, n \to \infty.$$

Hence the result.

(ii) By Lemma 2.3.3(ii),

$$\sup_{\mathbf{t}\in \mathbb{R}^k} |U_n(h, \mathbf{t}) - EU_n(h, \mathbf{t})| = O(b_n(4)) \text{ a.s. }, n \to \infty.$$

The result now follows exactly as in (i) above, using (2.3.7) in particular.□

2.4 Supplementary remarks

Remark 2.4.1. Note that, in Theorem 2.3.1, we have considered a.s. uniform convergence over a class of *compact* sets. Also, by **A2**, the functions $m(\cdot)$ and $f_1(\cdot)$ are continuous. In view of these facts, the assumption **A1** is actually redundant, once we replace '$\sup_{\mathbf{t}\in \mathbb{R}^k}$' by '$\sup_{\mathbf{t}\in C}$' in Lemma 2.2.1. Here of course, $C \in \mathcal{K}$ as in Theorem 2.3.1.

Remark 2.4.2. For the case $k = 1$, Härdle et al. (1988) obtained uniform convergence rates which are stronger than ours, viz. $O((na_n)^{-1/2}(\log n)^{1/2})$. But they had to make certain stronger assumptions, e.g. $E|h|^p < \infty$, for some $p > 2$, and $\sup_{t\in \mathbb{R}} |E(h^2(Y_1) \mid X_1 = t)| < \infty$, besides using a special, 'discrete', sequence of kernel functions. They, however, dealt with certain other types of conditional functionals as well. Recently, we came to know that Liero (1991) had obtained *optimal* rates for conditional U-statistics, whereas our rates are not optimal even in the case when h is bounded. (Our modest aim was a simple proof of uniform strong consistency through the use of empirical processes, rather than optimal rates.) Liero's conditions on h appear to be similar to those of Härdle et al.(1988).

2.5 Appendix : rudiments of empirical processes

Let $(\mathcal{X}, \mathcal{S})$ be any measurable space. A class $\mathcal{C} \subseteq \mathcal{S}$ is called a **Vapnik-Chervonenkis class** (a V-C class) if there exists $n \geq 1$ such that $m_{\mathcal{C}}(n) < 2^n$ where

$$m_{\mathcal{C}}(n) := \sup\{ \operatorname{card}(F \cap C : C \in \mathcal{C}) \mid F \subseteq \mathcal{X}, \operatorname{card}(F) = n\},$$

i.e., $m_{\mathcal{C}}(n)$ is the maximum number of subsets that can be cut out by $\mathcal{C}$ from a finite set of cardinality n. Clearly, $m_{\mathcal{C}}(n) \leq 2^n \forall n \geq 1$. Further,

$$v(\mathcal{C}) := \inf\{n \geq 1 : m_{\mathcal{C}}(n) < 2^n\} < \infty$$

is called the **V-C number** of the class $\mathcal{C}$.

Example 2-A.1. Let $\mathcal{X} = \mathbb{R}^k$. If $\mathcal{C} = \{(-\infty, \mathbf{x}] : \mathbf{x} \in \mathbb{R}^k\}$, then $v(\mathcal{C}) = k+1$. If $\mathcal{C} = \{(\mathbf{a}, \mathbf{b}) : \mathbf{a}, \mathbf{b} \in \mathbb{R}^k, \mathbf{a} \leq \mathbf{b}\}$, then $v(\mathcal{C}) = 2k+1$. Note that $\mathbf{a} \leq \mathbf{b}$ iff $a_i \leq b_i, 1 \leq i \leq k$.

Example 2-A.2. Let $\mathcal{X} = \mathbb{R}^k$. If $\mathcal{C} = \{$ all closed balls in $\mathbb{R}^k\}$, then $v(\mathcal{C}) = k+2$. (See Dudley(1979).)

Given a sequence $X_1, X_2, \ldots,$ of independent $\mathcal{X}$-valued random variables with $P_{(i)}$ the law of X_i, and a class $\mathcal{F}$ of real-valued functions on $\mathcal{X}$, we define

$$\nu_n(f) := n^{-1} \sum_{i=1}^{n} [f(X_i) - Ef(X_i)], \; f \in \mathcal{F},$$

which can be looked upon as an empirical process indexed by $f \in \mathcal{F}$. Then the usual empirical process $\nu_n(C) = n^{-1} \sum_{i=1}^{n} [1_C(X_i) - P_{(i)}(C)], \; C \in \mathcal{C}$, where $\mathcal{C}$ is some class (typically, a V-C class) of subsets of $\mathcal{X}$, becomes a special case with $\mathcal{F} = \{1_C : C \in \mathcal{C}\}$.

For a class of functions $\mathcal{F}$ and a probability law P on $\mathcal{X}$, define

$$\begin{aligned} N_p(\varepsilon, \mathcal{F}, P) \;=\; & \inf\{r \geq 1 : \exists f_1, f_2, \ldots, f_r \in \mathcal{F} \\ & \text{such that } \forall f \in \mathcal{F}, \inf_{1 \leq j \leq r} \|f - f_j\|_p < \varepsilon\} \end{aligned}$$

for $\varepsilon > 0$, $p > 0$, and $\|\cdot\|_p$ being the L_p-norm with respect to P. Then $H_p(\varepsilon, \mathcal{F}, P) := \ln N_p(\varepsilon, \mathcal{F}, P)$ is called the **metric entropy** of $\mathcal{F}$ in $L_p(P)$.

Now if $\mathcal{F} = \{\mathbf{1}_C : C \in \mathcal{C}\}$, where $\mathcal{C}$ is a V-C class, then one can obtain good bounds for $N_p(\varepsilon, \mathcal{F}, P)$, which are uniform in $\mathcal{P}$, the class of all probability laws on $\mathcal{X}$. (See Lemma 2.7, Alexander(1984).) Such a control of entropy, Alexander demonstrates, is very useful in obtaining sharp exponential bounds for the tail probability for $\sup_{f\in\mathcal{F}} |\nu_n(f)|$.

In view of this fact, Alexander proves his results for a class $\mathcal{F}$ which satisfies $0 \le f \le 1$ for all $f \in \mathcal{F}$ and $\mathcal{C}(\mathcal{F}) := \{C_f : f \in \mathcal{F}\}$ is a V-C class, where

$$C_f := \{(x,t) : 0 \le t \le f(x)\}$$

is the region under the graph of f. Note that $C_f \subseteq \mathcal{X} \times [0,1]\ \forall f \in \mathcal{F}$. In case $\mathcal{C}(\mathcal{F})$ is a V-C class, $\mathcal{F}$ is called a **V-C graph class** and $\mathcal{C}(\mathcal{F})$ its **graph region class.** It is easy to see that $\{\mathbf{1}_C : C \in \mathcal{C}\}$ is a V-C graph class if $\mathcal{C}$ is a V-C class.

Now,for a V-C graph class $\mathcal{F}$,

$$N_1(\varepsilon, \mathcal{F}, P) = N_1(\varepsilon, \mathcal{C}(\mathcal{F}), P \times \lambda), 0 < \varepsilon \le 1,$$

where λ is the Lebesgue measure on [0,1]. Thus one can bound the function on the left, as shown in Lemma 2.7 of Alexander(1984), by using the bound for the right hand side which involves *indicator functions of sets from a V-C class.*

Further, since $\sup_{f\in\mathcal{F}} |\nu_n(f)|$ need not always be measurable, Alexander imposes a measurability restriction. Let $\{X_i\}_{i\ge 1}$ be an independent and identically distributed sequence. A class $\mathcal{F}$ of functions is **n-supremum measurable** if $\sup_{f\in\mathcal{F}} \left(\sum_{i=1}^n \sum_{j=1}^n c_{ij} f(X_i) f(X_j) + \sum_{i=1}^n c_i f(X_i)\right)$ is measurable for all real $c_{ij}, c_i, 1 \le i, j \le n$. A class $\mathcal{F}$ is **n-deviation measurable** if both $\mathcal{F}$ and $\{(f-g) : f, g \in \mathcal{F}, \operatorname{var}(\nu_n(f-g)) \le \theta\}$, where 'var' denotes variance, are n-supremum measurable for all $\theta > 0$. These two conditions ensure the measurability of $\sup_{\mathcal{F}} |\nu_n(f)|$ and certain 'sample variances' which figure in Alexander's proofs. The 'n'- is dropped from the definition if the conditions hold for all $n \ge 1$.

Now consider the space $\mathcal{X} \times [0,1]$. Let $\mathcal{A} \subseteq \mathcal{S}$ be a V-C class on $\mathcal{X}$ with V-C number v_0. Denote by z a typical element of $\mathcal{X} \times [0,1]$, i.e., $z = (x,u), x \in \mathcal{X}, 0 \le u \le 1$.

Lemma 2-A. *The class of functions* $\mathcal{F}(\mathcal{A}) = \{f_A : A \in \mathcal{A}\}$, *where*

$$f_A(z) = f_A(x,u) := u\mathbf{1}_A(x),$$

is a V-C graph class. If $\Gamma(\mathcal{A})$ *denotes its graph region class, then* $v(\Gamma(\mathcal{A})) = v_0$.

Proof: Note that $\Gamma(\mathcal{A}) = \{C_A : A \in \mathcal{A}\}$, where $C_A = \{(x,u,t) : 0 \leq t \leq u\mathbf{1}_A(x)\} \subseteq \mathcal{X} \times [0,1] \times [0,1]$. That is,

$$C_A = (A \times \{(u,t) : 0 \leq t \leq u \leq 1\}) \cup (A^c \times [0,1] \times \{0\}).$$

We have to show that (recall the definition of a V-C number)

$$\begin{aligned}&\sup\{n \geq 1 : \exists \bar{z}_1, \ldots, \bar{z}_n \in \mathcal{X} \times [0,1] \times [0,1] \text{ such that}\\ &\text{card}(C_A \cap \{\bar{z}_1, \ldots, \bar{z}_n\} : A \in \mathcal{A}) = 2^n\} = v_0 - 1 < \infty.\end{aligned}$$

Now let $\{\bar{z}_1, \ldots, \bar{z}_n\}$ be a set of any n points on $\mathcal{X} \times [0,1] \times [0,1]$, $\bar{z}_i = (x_i, u_i, t_i), 1 \leq i \leq n$. Denote

$$\begin{aligned}&m_{\mathcal{A}}(\bar{z}_i : 1 \leq i \leq n)\\ =\ &\text{card}(C_A \cap \{\bar{z}_1, \ldots, \bar{z}_n\} : A \in \mathcal{A})\\ =\ &\text{card}(\{(x_i,u_i,t_i) : x_i \in A, 0 \leq t_i \leq u_i \leq 1\}\\ &\cup \{(x_i,u_i,t_i) : x_i \in A^c, t_i = 0\} : A \in \mathcal{A}).\end{aligned}$$

Suppose first that $\{i : t_i > u_i\} \neq \emptyset$. Then $\nexists A \in \mathcal{A}$ for which $C_A \cap \{\bar{z}_1, \ldots, \bar{z}_n\} = \{\bar{z}_i : t_i > u_i\}$. So, obviously, $m_{\mathcal{A}}(\bar{z}_i : 1 \leq i \leq n) < 2^n$. Hence, assume $t_i \leq u_i, 1 \leq i \leq n$.

If now $\{i : t_i = 0\} = \{1, 2, \ldots, n\}$, then

$$C_A \cap \{\bar{z}_i : 1 \leq i \leq n\} = \{\bar{z}_i : 1 \leq i \leq n\} \forall A \in \mathcal{A},$$

so $m_{\mathcal{A}}(\bar{z}_i : 1 \leq i \leq n) = 1 < 2^n$.

If $\{1, 2, \ldots, n\} \neq \{i : t_i = 0\} \neq \emptyset$, then $\{i : t_i > 0\} \neq \emptyset$. Now, if $\exists A \in \mathcal{A}$ for which $C_A \cap \{\bar{z}_i : 1 \leq i \leq n\} = \{\bar{z}_i : t_i > 0\}$, i.e.,

$$\{(x_i,u_i,t_i) : x_i \in A\} \cup \{(x_i,u_i,t_i) : x_i \in A^c, t_i = 0\} = \{(x_i,u_i,t_i) : t_i > 0\},$$

then we must have

$$\{\bar{z}_i : x_i \in A\} = \{\bar{z}_i : t_i > 0\}, \text{ and hence } \{\bar{z}_i : x_i \in A^c\} = \{\bar{z}_i : t_i = 0\}.$$

This means that

$$\{\bar{z}_i : t_i > 0\} = C_A \cap \{\bar{z}_i : 1 \le i \le n\} = \{\bar{z}_i : 1 \le i \le n\},$$

which is impossible since $\{i : t_i = 0\} \neq \emptyset$. Thus the set $\{z_i : t_i > 0\}$ cannot be cut out by any $A \in \mathcal{A}$. Hence again, $m_{\mathcal{A}}(\bar{z}_i : 1 \le i \le n) < 2^n$.

If $\{i : t_i = 0\} = \emptyset$, then

$$C_A \cap \{\bar{z}_i : 1 \le i \le n\} = \{(x_i, u_i, t_i) : x_i \in A\}, \ \forall A \in \mathcal{A}.$$

Hence only in this case, and for $n = 1, 2, \ldots, v_0 - 1, \exists \{\bar{z}_1, \ldots, \bar{z}_n\}$ s.t. $m_{\mathcal{A}}(\bar{z}_i : 1 \le i \le n) = 2^n$, since $v(\mathcal{A}) = v_0$. Thus $v(\Gamma(\mathcal{A})) = v_0$.□

Chapter 3

Limit Distributions of Conditional U-statistics

3.1 Introduction

In this chapter, we study the limit distributions of $U_n^h(\mathbf{t})$ for a fixed $\mathbf{t} \in I\!R^k$. Stute (1991) has obtained, among other things, conditions for asymptotic normality of $U_n^h(\mathbf{t})$, suitably normalised and centered. We make the following observations:

1. Stute has remarked that though h is assumed to be symmetric in the theory of classical U-statistics, such an assumption here would involve both h and $\prod K_n$ ($\equiv a_n^{-k} \prod_{j=1}^{k} K((t_j - \cdot)/a_n)$). Since the roles of h and $\prod K_n$ are different, little, he says, will be gained from a symmetrisation. We point out that h can, in most examples, be taken to be symmetric; further, in case $t_1 = t_2 = \cdots = t_k$ in $\mathbf{t}$ (the 'diagonal' case), $\prod K_n$ is also symmetric. In any case, it follows from (1.0.7) that m is symmetric in its arguments provided h is. These facts, as we shall see later on, will lead to a convenient description of the possible limit distributions of $U_n^h(\mathbf{t})$.

2. Consider Example 4.1 in Stute (1991). Here $k = 2$, $h(y_1, y_2) = y_1 y_2$. Hence

$$m(t_1, t_2) = E(Y_1 Y_2 | X_1 = t_1, X_2 = t_2) = \bar{m}(t_1)\bar{m}(t_2),$$

where $\bar{m}(t) = E(Y|X = t)$. By Theorem 1 of the above paper,

$$\sqrt{na_n}(U_n^h(\mathbf{t}) - EU_n(h, \mathbf{t})/EU_n(1, \mathbf{t})) \xrightarrow{d} N(0, \rho^2) \text{ as } n \to \infty,$$

where

$$\rho^2 = \begin{cases} 4D_1\bar{m}^2(t_1)/f_1(t_1), & \text{if } t_1 = t_2 \\ D_1\bar{m}^2(t_2)/f_1(t_1) + D_2\bar{m}^2(t_1)/f_1(t_2), & \text{if } t_1 \neq t_2 \end{cases}$$

and

$$D_i = \operatorname{var}(Y|X = t_i) \int_{I\!R} K^2(u)du, \ i = 1, 2.$$

Now if, for example, $\bar{m}(t_1) = 0$ when $t_1 = t_2$, or if $\bar{m}(t_1) = \bar{m}(t_2) = 0$ when $t_1 \neq t_2$, then obviously what we have is an example of so-called 'degeneracy'.

Observation 2 led us to consider higher-order limiting distributions. Following Dynkin and Mandelbaum (1983), we describe the limit distributions in terms of *multiple Wiener integrals*. We use the classical multiple Wiener integral in case $t_1 = t_2 = \cdots = t_k = t$ (say), and a modified version thereof in case t_i, $1 \leq i \leq k$, are not necessarily equal. The requisite definitions and facts are given in Section 3.2.

In Section 3.3, we present the limit distributions of $U_n(h, \mathbf{t})$. As a tool in our proofs of weak convergence, we use the combinatorial lemma and the main theorem of Rubin and Vitale (1980), as well as the techniques of Schuster (1972). We then discuss the possible limit distributions of $U_n^h(\mathbf{t})$ in Section 3.4. Here, unlike in the classical U-statistic set-up, the 'degenerate' case (of which the example in Observation 2 is a special case) is a bit messy to deal with. However, we have succeeded in simplifying the things somewhat when h has a product structure. Finally, in Section 3.5, an example and a couple of remarks are given.

3.2 Symmetric tensor products and multiple Wiener integrals

Let P be a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then there exists a probability space which supports the Gaussian process $\{I_1(\varphi) \mid \varphi \in L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)\}$, given by

$$\left.\begin{aligned} E_0 I_1(\varphi) &= 0, \\ E_0 I_1(\varphi) I_1(\psi) &= E_P(\varphi\psi), \end{aligned}\right\} \varphi, \psi \in L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P), \tag{3.2.1}$$

where E_0 denotes expectation in the latter probability space. Note that $\varphi \mapsto I_1(\varphi)$ is a linear map, by (3.2.1). It is thus a linear isometry between $L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$ and the Hilbert space

$$\mathcal{H} := \{I_1(\varphi) \mid \varphi \in L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)\}.$$

Call $I_1(\varphi)$ the *Wiener integral* of φ. It follows that there exists a linear isometry between $\sigma(\otimes^k L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P))$ and $\sigma(\otimes^k \mathcal{H})$, $k \geq 1$, where

$$\sigma(\otimes^k L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)) := k\text{-fold } \textit{symmetric tensor product} \text{ of } L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P).$$

and $\sigma(\otimes^k \mathcal{H})$ is defined similarly. For a definition and discussion of symmetric tensor products of Hilbert spaces, see, e.g., Kallianpur (1980), pp.139-155, or Parthasarathy (1992), Section 17, pp.105 - 111. The following facts are well-known (see, for example, Kallianpur (1980), Equation (6.6.31), pp.153, in the proof of Theorem 6.6.4 and Lemma 6.4.1, pp. 139). Here we denote an isomorphism between vector spaces by '$\simeq$':

Lemma 3.2.1.

(i)

$$\sigma(\otimes^k L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)) \simeq L_2^{\text{sym}}(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), \underbrace{P \times \cdots \times P}_{k\text{-fold}}),$$

where

$$L_2^{sym}(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), \underbrace{P \times \cdots \times P}_{k\text{-fold}}) := \{f : \mathbb{R}^k \to \mathbb{R} \mid f \text{ symmetric}, E_P f^2 < \infty\}.$$

(ii) Functions of the form

$$\underbrace{\varphi(\cdot)\ldots\varphi(\cdot)}_{k\text{-fold}} := \varphi^{\otimes k},\ \varphi \in L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P),$$

and vectors of the form

$$\underbrace{I_1(\psi)\otimes\cdots\otimes I_1(\psi)}_{k\text{-fold}},\ \psi \in L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P).$$

are total in $\sigma(\otimes^k L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P))$ *and* $\sigma(\otimes^k \mathcal{H})$, *respectively.*

In view of Lemma 3.2.1(i), we shall identify the two Hilbert spaces there, from now on. Note that, for $\varphi, \psi \in L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$,

$$\begin{aligned}
& < I_1(\varphi)\otimes\cdots\otimes I_1(\varphi), I_1(\psi)\otimes\cdots\otimes I_1(\psi) >_{\sigma(\otimes^k \mathcal{H})} \\
= & < \varphi^{\otimes k}, \psi^{\otimes k} >_{\sigma(\otimes^k L_2)} \\
= & \{E_P \varphi\psi\}^k \\
= & \{E_0 I_1(\varphi) I_1(\psi)\}^k \\
= & E_0 \left\{ \|\varphi\|^k (k!)^{-1/2} H_k(I_1(\varphi/\|\varphi\|)) . \|\psi\|^k (k!)^{-1/2} H_k(I_1(\psi/\|\psi\|)) \right\},
\end{aligned} \tag{3.2.2}$$

where $< \cdot, \cdot >_{\mathcal{S}}$ denotes the inner-product in a Hilbert space $\mathcal{S}$, $\|\cdot\|$ denotes the L_2-norm in $L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$ and $H_k(\cdot)$ is the *k-th Hermite polynomial.* Define

$$\mathcal{W}_k := \text{ closed linear span of } \{\|\varphi\|^k (k!)^{-1/2} H_k(I_1(\varphi/\|\varphi\|)) \mid \varphi \in L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)\}.$$

It follows from (3.2.2) and Lemma 3.2.1 that there exists a linear isometry I_k : $L_2^{\rm sym}(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), P^k) \to \mathcal{W}_k$, $\varphi^{\otimes k} \mapsto \|\varphi\|^k (k!)^{-1/2} H_k(I_1(\varphi/\|\varphi\|))$, where P^k denotes the k-fold product of P. Call $I_k(f)$ the k-th order **multiple Wiener integral** of f, $f \in L_2^{\rm sym}$. It has the following orthogonality property:

Lemma 3.2.2. *Let* $f \in L_2^{sym}(\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m), P^m)$, $g \in L_2^{sym}(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), P^k)$. *Then*

$$E_0 I_m(f).I_k(g) = \delta_{mk} E_P(fg), \; m, k \geq 1.$$

Proof: Follows easily from the fact that, for $\varphi, \psi \in L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$,

$$\begin{aligned} & E_0 I_m(\varphi^{\otimes m}) I_k(\psi^{\otimes k}) \\ = \; & E_0\{\|\varphi\|^m (m!)^{-1/2} H_m(I_1(\varphi/\|\varphi\|)).\|\psi\|^k (k!)^{-1/2} H_k(I_1(\psi/\|\psi\|))\} \\ = \; & \delta_{mk}\{E_P \varphi\psi\}^k, \end{aligned}$$

the functions $\varphi^{\otimes k}$ are total in $L_2^{\rm sym}(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), P^k)$ and I_k is an isometry.□

We now proceed to define a modified, 'non-homogeneous' version of the multiple Wiener integral. Let $P_1, \ldots, P_m$, $m \geq 1$, be distinct probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then, on some probability space, the following *independent* Gaussian processes exist:

$$\{I_1^{(i)}(\varphi) \mid \varphi \in L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_i)\}, \; 1 \leq i \leq m,$$

such that

$$\left.\begin{aligned} & E_0 I_1^{(i)}(\varphi) = 0 \\ & E_0 I_1^{(i)}(\varphi) I_1^{(j)}(\psi) = \delta_{ij} E_{P_i}(\varphi\psi) \end{aligned}\right\} \forall\, 1 \leq i, j \leq m$$

and $\varphi \in L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_i), \psi \in L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_j)$. Define, as before,

$$\begin{aligned} \mathcal{H}_i &= \{I_1^{(i)}(\varphi) \mid \varphi \in L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_i)\} \\ \mathcal{W}_{r_i} &= \text{closed linear span of } \left\{\|\varphi\|^{r_i}(r_i!)^{-1/2} H_{r_i}(I_1(\varphi/\|\varphi\|)) \mid \varphi \in L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_i)\right\}, \end{aligned}$$

where $r_1 \geq 1$ is an integer and $\| \cdot \|_i$ is the norm in $L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_i)$, $1 \leq i \leq m$. It is easy to see that the following chain of linear isometries can be etablished as before:

$$\begin{aligned} & \otimes_{i=1}^m(\sigma(\otimes^{r_i} L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_i))) \\ & \xrightarrow{I_{r_1,\ldots,r_m}} \otimes_{i=1}^m(\sigma(\otimes^{r_i}\mathcal{H}_i)) \xrightarrow{J_{r_1,\ldots,r_m}} \otimes_{i=1}^m \mathcal{W}_{r_i} \end{aligned} \tag{3.2.3}$$

Now take, any two of the $\mathcal{W}_{r_i}$'s, say $\mathcal{W}_{r_1}$ and $\mathcal{W}_{r_2}$. Denote, for the sake of simplicity,

$$\frac{\|\varphi\|_i^{r_i}}{\sqrt{r_i!}} H_{r_i}(I_1^{(i)}(\varphi/\|\varphi\|_i)); = \bar{H}_{r_i}(\varphi_i),\ i = 1,2$$

Then, for $\varphi, \varphi' \in L_2(P_1)$ and $\psi, \psi' \in L_2(P_2)$,

$$\begin{aligned} & < \bar{H}_{r_1}(\varphi)\otimes\bar{H}_{r_2}(\psi), \bar{H}_{r_1}(\varphi')\otimes\bar{H}_{r_2}(\psi') >_{\mathcal{W}_{r_1}\otimes\mathcal{W}_{r_2}} \\ = & < \bar{H}_{r_1}(\varphi), \bar{H}_{r_1}(\varphi') >_{\mathcal{W}_{r_1}} . < \bar{H}_{r_2}(\psi), \bar{H}_{r_2}(\psi') >_{\mathcal{W}_{r_2}} \\ = & E_0\bar{H}_{r_1}(\varphi).\bar{H}_{r_1}(\varphi').E_0 H_{r_2}(\psi).\bar{H}_{r_2}(\psi') \\ = & E_0\{\bar{H}_{r_1}(\varphi).\bar{H}_{r_2}(\psi).H_{r_1}(\varphi').\bar{H}_{r_2}(\psi')\}, \end{aligned}$$

by independence of the processes $\{I_1^{(i)}(\varphi) : \varphi \in L_2(P_i)\}, i = 1, 2.$ (3.2.4)

It is clear from (3.2.4) that one can identify (in the sense of linear isometry) the Hilbert space $\mathcal{W}_{r_1}\otimes\mathcal{W}_{r_2}$ with

$$\begin{aligned} \mathcal{W}_{r_1,r_2} := & \text{ closed linear span of } \{\bar{H}_{r_1}(\varphi).H_{r_2}(\psi) : \\ & \varphi \in L_2(I\!R, \mathcal{B}(I\!R), P_1), \psi \in L_2(I\!R, \mathcal{B}(I\!R), P_2)\}. \end{aligned}$$

Thus, by (3.2.3) and (3.2.4), there exists the following isometry:

$$I_{r_1,r_2,\ldots,r_m} : \otimes_{i=1}^m(\sigma(\otimes^{r_i} L_2(I\!R, \mathcal{B}(I\!R), P_i))) \rightarrow \mathcal{W}_{r_1,r_2,\ldots,r_m}$$

such that

$$I_{r_1,r_2,\ldots,r_m}(\varphi_1^{\otimes r_1}\otimes\varphi_2^{\otimes r_2}\cdots\otimes\varphi_m^{\otimes r_m}) = \bar{H}_{r_1}(\varphi_1)\bar{H}_{r_2}(\varphi_2)\ldots\bar{H}_{r_m}(\varphi_m)$$

for $\varphi_i \in L_2(I\!R, \mathcal{B}(I\!R), P_i), 1 \le i \le m$. Here $\mathcal{W}_{r_1,r_2,\ldots,r_m}$ is defined in an obvious way as $\mathcal{W}_{r_1,r_2}$, and

$$\varphi_1^{\otimes r_1}\otimes\varphi_2^{\otimes r_2}\otimes\cdots\otimes\varphi_m^{\otimes r_m} = \underbrace{\varphi_1(\cdot)\ldots\varphi_1(\cdot)}_{r_1\text{-fold}}\underbrace{\varphi_2(\cdot)\ldots\varphi_2(\cdot)}_{r_2\text{-fold}}\cdots\underbrace{\varphi_m(\cdot)\ldots\varphi_m(\cdot)}_{r_m\text{-fold}}.$$

Further, with $r = r_1 + r_2 + \cdots + r_m$,

$$\begin{aligned} & \otimes_{i=1}^m(\sigma(\otimes^{r_i} L_2(I\!R, \mathcal{B}(I\!R), P_i))) \\ \simeq & \{f : I\!R^r \rightarrow I\!R \mid E_{P_1P_2\ldots P_m}f^2 < \infty, f(t_1, t_2, \ldots, t_r) \text{ is invariant under} \\ & \text{the permutation of 1st } r_1, \text{ 2nd } r_2, \ldots, \text{ last } r_m \text{ co-ordinates }\} \end{aligned}$$

We note the following approximation result for later use:

Lemma 3.2.3. *Let $\mathcal{S}_i = \{u : u$ a measurable, simple function on $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_i)\}, 1 \leq i \leq m$, $m \geq 1$. Then the product functions of the form*

$$u_1^{\otimes r_1} \otimes u_2^{\otimes r_2} \otimes \ldots \otimes u_m^{\otimes r_m}, \quad u_i \in \mathcal{S}_i, 1 \leq i \leq m$$

and random variables of the form

$$\bar{H}_{r_1}(u_1)\bar{H}_{r_2}(u_2)\ldots\bar{H}_{r_m}(u_m)$$

are total in $\otimes_{i=1}^m(\sigma(\otimes^{r_i} L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_i)))$ and $\mathcal{W}_{r_1,r_2,\ldots,r_m}$ respectively, where $r_i, 1 \leq i \leq m$, and $\bar{H}_{r_i}(\cdot)$ are as before.

Proof: Denote

$$\otimes_{i=1}^m(\sigma(\otimes^{r_i} L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_i))) := \sigma(r_1, r_2, \ldots, r_m)$$

Then $\varphi_1^{\otimes r_1} \otimes \varphi_2^{\otimes r_2} \otimes \cdots \otimes \varphi_m^{\otimes r_m}$ and $\bar{H}_{r_1}(\varphi_1)\bar{H}_{r_2}(\varphi_2)\ldots\bar{H}_{r_m}(\varphi_m), \varphi_i \in L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_i), 1 \leq i \leq m$, are total in $\sigma(r_1, r_2, \ldots, r_m)$ and $\mathcal{W}_{r_1,r_2,\ldots,r_m}$ respectively (in fact, we used this implicitly while defining $I_{r_1,r_2,\ldots,r_m}$). Further,

$$\begin{aligned} &\| \textstyle\sum_{j=1}^k c_j \varphi_{1j}^{\otimes r_1} \otimes \ldots \otimes \varphi_{mj}^{\otimes r_m} - \sum_{j=1}^k c_j u_{1j}^{\otimes r_1} \otimes \ldots \otimes u_{mj}^{\otimes r_m} \| \\ = \;& \textstyle\sum_{l=1}^k \sum_{j=1}^k c_l c_j \{\prod_{i=1}^m < \varphi_{il} \varphi_{ij} >^{r_i} - 2\prod_{i=1}^m < \varphi_{il}, u_{ij} >^{r_i} + \prod_{i=1}^m < u_{il}, u_{ij} >^{r_i}\}. \end{aligned}$$

The result now follows from the fact that $\mathcal{S}_i$ is *dense* in $L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_i), 1 \leq i \leq m$ and that $I_{r_1,r_2,\ldots,r_m}$ is an isometry.□

3.3 Limit distributions of $U_n(h, \mathbf{t})$

This section contains our main result. For the sake of convenience, we shall discuss the cases '$t_1 = \ldots = t_k = t$' (say) and 't_i's are not all equal, $1 \leq i \leq k$', in two separate subsections.

3.3.1 The diagonal case: $t_1 = t_2 = \ldots = t_k = t$

In this case we have the following triangular array of U-statistics, with the 'U-kernel' symmetric and depending on n:

$$U_n(h, t.\mathbf{1}) = \binom{n}{k}^{-1} \sum_{1 \le \alpha(1) < \ldots < \alpha(k) \le n} (h.\prod K_n)(\mathbf{Z}_{\alpha(1)}, \ldots, \mathbf{Z}_{\alpha(k)}), n \ge k$$

where

$$\mathbf{1} = (1, 1, \ldots, 1)'_{1\times k},$$

$$(h.\prod K_n)(\mathbf{z}_1, \ldots, \mathbf{z}_k) = h(y_1, y_2, \ldots, y_k)a_n^{-k} \prod_{j=1}^{k} K((t - x_j)/a_n),$$

$$\mathbf{Z}_i = (X_i, Y_i), \mathbf{z}_i = (x_i, y_i), 1 \le i \le n. \tag{3.3.1}$$

Following the notation of Dynkin and Mandelbaum (1983), we decompose $(h.\prod K_n)$ into 'canonical' parts as follows:

$$(h.\prod K_n) \equiv [(I - Q_1) + Q_1] \ldots [(I - Q_k) + Q_k](h.\prod K_n) \tag{3.3.2}$$

where I denote the identity operator and Q_i denotes the following conditional expectation operator:

$$Q_i := E(\cdot|\mathbf{Z}_1, \ldots, \mathbf{Z}_{i-1}, \mathbf{Z}_{i+1}, \ldots, \mathbf{Z}_k), 1 \le i \le k$$

(with the obvious modification when $i = k$). In other words, Q_i means integrating out $\mathbf{Z}_i$, holding $\mathbf{Z}_j, j \ne i$, fixed.

Expanding (3.3.2), we have

$$\begin{aligned}(h.\prod K_n)(\mathbf{z}_1, \ldots, \mathbf{z}_k) \equiv\ & Eh.\prod K_n + \sum_{i=1}^{k} h_n^{(1)}(\mathbf{z}_i) + \sum_{1 \le i_1 < i_2 \le k} h_n^{(2)}(\mathbf{z}_{i_1}, \mathbf{z}_{i_2}) \\ & + \ldots + \sum_{1 \le i_1 < i_2 < \ldots < i_r \le k} h_n^{(r)}(\mathbf{z}_{i_1}, \mathbf{z}_{i_2}, \ldots, \mathbf{z}_{i_r}) \\ & + \ldots + h_n^{(k)}(\mathbf{z}_1, \ldots, \mathbf{z}_k),\end{aligned} \tag{3.3.3}$$

where

$$\begin{aligned}h_n^{(r)}(\mathbf{z}_1, \ldots, \mathbf{z}_r) = \ & [(I - Q_1)(I - Q_2) \ldots (I - Q_r)Q_{r+1} \ldots Q_k] \\ & (h.\prod K_n)(\mathbf{z}_1, \ldots, \mathbf{z}_k), 1 \le r \le k.\end{aligned}$$

Hence,

$$U_n(h, t.1) = Eh.\prod K_n + \binom{k}{1} U_n(h_n^{(1)}) + \ldots + \binom{k}{r} U_n(h_n^{(r)}) + \ldots + U_n(h_n^{(k)}).$$

For example, for $k = 2$ we have

$$h_n^{(1)}(\mathbf{z}) = a_n^{-1} K_n(t, x_1) Eh(y_1, Y_2) a_n^{-1} K_n(t, X_2) - Eh.\prod K_n$$

$$\begin{aligned} h_n^{(2)}(\mathbf{z}_1, \mathbf{z}_2) &= h(y_1, y_2) a_n^{-2} \prod_{j=1}^{2} K_n(t, x_j) \\ &- a_n^{-1} K_n(t, x_1).Eh(y_1, Y_2) a_n^{-1} K_n(t, X_2) \\ &- a_n^{-1} K_n(t, x_2) Eh(Y_1, y_2) a_n^{-1} K_n(t, X_1) + Eh\prod K_n, \end{aligned}$$

where $K_n(t, x) = K((t - x)/a_n)$. The $h_n^{(r)}$, $1 \leq r \leq k$, have the property (termed 'canonical' by Dynkin and Mandelbaum) that

$$Eh_n^{(r)}(\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{r-1}, \mathbf{Z}_r) = 0 \text{ for all } (\mathbf{z}_1, \ldots, \mathbf{z}_{r-1}) \in I\!R^{2(r-1)}. \tag{3.3.4}$$

Though the decomposition (3.3.2) (or, equivalently, (3.3.3)) was introduced by Hoeffding (1961), we shall henceforth call it the *Dynkin - Mandelbaum (D-M) decomposition*, since 'Hoeffding decomposition' means something else in the theory of U-statistics.

Further, consider the probability measure P_t on $(I\!R, B(I\!R))$ given by the following conditional density:

$$f(y|t) = \frac{f(t, y)}{f_1(t)}, y \in I\!R$$

Denote by E_t the expectation with respect to P_t. We shall express the limiting random variables as multiple Wiener integrals with respect to P_t (refer to Section 3.2). Let us state our main assumptions now. Other assumptions will be stated as and when they are required.

A1. $f(y|\cdot)$ is continuous at t for all $y \in I\!R$.

A2. $f_1(\cdot)$ is continuous at t and $f_1(t) > 0$.

A3. **Either,** $E|h|^{2+\delta} < \infty$ for some $\delta > 0$, h is continuous and

$$E\{|h(y_1, \ldots, y_k)|^{2+\delta} \mid X_1 = x_1, \ldots, X_k = x_k\}$$

is bounded in a neighbourhood of $t.1$,

or, $h(\cdot)$ is bounded.

A4. $K(\cdot)$ is a density on $I\!R$ such that $\lim_{|u|\to\infty} |uK(u)| = 0$ and $\sup_{u\in I\!R} K(u) < \infty$.

A5. $a_n \to 0, na_n^3 \to \infty$ as $n \to \infty$.

Lemma 3.3.1 *Under A1 - A5, we have*

$$\text{(i) } var\ (U_n(h, t.\mathbf{1}')) = \sum_{r=1}^{k} \binom{k}{r}^2 var\ (U_n(h_n^{(r)})) = \sum_{r=1}^{k} \binom{k}{r}^2 r! E(h_n^{(r)})^2 . O(1/n^r)$$

$$\text{(ii) } a_n^r E(h_n^{(r)})^2 \to E_t(m_r(Y_1, \ldots, Y_r))^2 (f_1(t) \int_{I\!R} K^2(u) du)^r . (f_1(t))^{2(k-r)}, 0 \le r \le k,$$

as $n \to \infty$, *where* $h_n^{(0)} = Eh. \prod K_n$,

$m_r(y_1, \ldots, y_r) = E_t h(y_1, \ldots, y_r, Y_{r+1}, \ldots, Y_k), m_0 \equiv m(t)$.

Proof: (i) By the canonicity property (3.3.4), we have

$$Eh_n^{(l)}(\mathbf{Z}_{i_1}, \ldots, \mathbf{Z}_{i_l}).h_n^{(r)}(\mathbf{Z}_{j_1}, \ldots, \mathbf{Z}_{j_r}) = \begin{cases} E(h_n^{(r)})^2 & \text{if } r = l \text{ and} \\ & \{i_1, \ldots, i_l\} = \{j_1, \ldots, j_r\} \\ 0 & \text{otherwise.} \end{cases}$$

Further, $\binom{n}{r}^{-1} = O(1/n^r), 1 \le r \le k$. The result follows.

(ii) Note that

$h_n^{(r)}(\mathbf{z}_1, \ldots, \mathbf{z}_r)$

$= [(I - Q_1)(I - Q_2) \ldots (I - Q_r) Q_{r+1} \ldots Q_k] h(y_1 \ldots y_k) a_n^{-k} \prod_{j=1}^{k} K_n(t, x_j)$

$= a_n^{-r} \prod_{j=1}^{r} K_n(t, x_j) . Eh(y_1 \ldots, y_r, Y_{r+1}, \ldots, Y_k) a_n^{-(k-r)} \prod_{i=r+1}^{k} K_n(t, X_i)$

$+$(remaining terms),

where the remaining terms are of the form

$$a_n^{-l}\prod_{j=1}^{l}K_n(t,x_j).Eh(y_1,\dots,y_l,Y_{l+1},\dots,Y_k)a_n^{-(k-l)}\prod_{i=l+1}^{k}K_n(t,X_i),$$

with $0\le l\le r-1$. Denote

$$h_{nj}(y_1,\dots,y_j)=Eh(y_1,\dots,y_j,Y_{j+1},\dots,Y_k)a_n^{-(k-j)}\prod_{i=j+1}^{k}K_n(t,X_i),$$

$0\le j\le r$, where $h_{n0}\equiv Eh.\prod K_n$.

Hence,

$$a_n^r E(h_n^{(r)}(\mathbf{Z}_1,\dots,\mathbf{Z}_r))^2$$

$$=a_n^{-r}E\prod_{j=1}^{r}K_n^2(t,X_j)h_{nr}^2(Y_1,\dots,Y_r)+(\text{remaining terms}) \tag{3.3.5}$$

where, in this case, the remaining terms look like either

$$a_n^r a_n^{-2l}E\prod_{j=1}^{l}K_n^2(t,X_j)h_{nl}^2(Y_1,\dots,Y_l),0\le l\le r-1 \tag{3.3.6}$$

or,

$$\begin{aligned}a_n^r a_n^{-l}a_n^{-p}E\Bigg\{\prod_{j=1}^{l}K_n(t,X_{\alpha(j)})h_{nl}(Y_{\alpha(1)},\dots,Y_{\alpha(l)})\prod_{i=1}^{p}K_n(t,X_{\beta(i)})\\ h_{np}(Y_{\beta(1)},\dots,Y_{\beta(p)})\Bigg\},0\le l\le r-1,l\le p\le r\\ 1\le\alpha(1)<\dots<\alpha(l)\le r,1\le\beta(1)<\dots<\beta(p)\le r\\ \{\alpha(1),\dots,\alpha(l)\}\ne\{\beta(1),\dots,\beta(p)\}.\end{aligned} \tag{3.3.7}$$

It is easy to see that the first term on the right of (3.3.5) can be written as

$$\begin{aligned}\int_{\mathbb{R}^r}\int_{\mathbb{R}^{(k-r)}}\int_{\mathbb{R}^{(k-r)}}\Bigg\{M_r(t-a_nx_1,\dots,t-a_nx_r,t-a_nv_1,\dots,t-a_nv_{k-r},\\ t-a_nv_1',\dots t-a_nv_{k-r}')\\ \prod_{j=1}^{r}K^2(x_j)f_1(t-a_nx_j)\prod_{i=1}^{k-r}K(v_i)f_1(t-a_nv_i)K(v_i')f_1(t-a_nv_i')\Bigg\}d\mathbf{x}d\mathbf{v}d\mathbf{v}'\end{aligned} \tag{3.3.8}$$

where

$$M_r(x_1,\dots,x_r,v_1,\dots,v_{k-r},v_1',\dots,v_{k-r}')$$

$$:= E\{h(Y_1,\ldots,Y_r,U_1,\ldots,U_{k-r})h(Y_1,\ldots,Y_r,U_1',\ldots,U_{k-r}')|X_j = x_j, V_i = v_i,$$

$$V_i' = v_i', 1 \le j \le r, 1 \le i \le k-r\}, \tag{3.3.9}$$

and $(X_j, Y_j)_{j=1}^r, (U_i, V_i)_{i=1}^{k-r}, (U_i', V_i')_{i=1}^{k-r}$ are i.i.d. Now, we can apply Bochner's Theorem (cf Prakasa Rao (1983), Ch.2, pp.35) and conclude that the integral in (3.3.8) converges to the limiting expression given in (ii) of Lemma 3.3.1 provided we establish that $M_r(\cdot)$ is continuous at $t.1$. Note that $f_1(\cdot)$ is already assumed to be continuous at t, by A2. Now, if $h(\cdot)$ is bounded, then by A1 we immediately get the continuity of $M_r(\cdot)$ at $t.\mathbf{1}$, applying Scheffe's Lemma to the family of density functions $f(\cdot|t+\delta), \delta \downarrow 0$. Also, by the same Lemma, the family of product-measures given by

$$\prod_{j=1}^{r} f(\cdot|x_j) \prod_{i=1}^{k-r} f(\cdot|v_i) f(\cdot|v_i')$$

on $\mathbb{R}^r \times \mathbb{R}^{k-r} \times \mathbb{R}^{k-r}$ converge *weakly* to the i.i.d product measure $\prod f(\cdot|t)$ as $x_j \to t, v_i \to t, v_i' \to t$. By (3.3.9), continuity of $M_r(\cdot)$ is equivalent to the convergence of the expectation in (3.3.9) with respect to these product measures. Therefore if $h(\cdot)$ is continuous (but not necessarily bounded), it suffices to show that, for some $\varepsilon > 0$,

$$E\{|h(Y_1,\ldots,Y_r,U_1,\ldots,U_{k-r})h(Y_1,\ldots,Y_r,U_1',\ldots,U_{k-r}')|^{1+\varepsilon} \mid X_j = x_j, V_i = v_i,$$

$$V_i' = v_i', 1 \le j \le r, 1 \le i \le k-r\} < \infty$$

uniformly for all x_j, v_i, v_i' near t. But for $\varepsilon = \delta/2, \delta$ as in A3,

$$\begin{aligned} & E\{|h(\mathbf{Y},\mathbf{U})h(\mathbf{Y},\mathbf{U}')|^{1+\delta/2}|\mathbf{X} = \mathbf{x}, \mathbf{V} = \mathbf{v}, \mathbf{V}' = \mathbf{v}'\} \\ \le \; & (E\{|h(\mathbf{Y},\mathbf{U})|^{2+\delta}|\mathbf{X} = \mathbf{x}, \mathbf{V} = \mathbf{v}\})^{1/2}(E\{|h(\mathbf{Y},\mathbf{U}')|^{2+\delta}|\mathbf{X} = \mathbf{x}, \mathbf{V}' = \mathbf{v}'\})^{1/2}, \end{aligned}$$

using the Cauchy - Schwartz inequality and the independence of the conditional expectations involved. But this final product is bounded for $(\mathbf{x}, \mathbf{v}, \mathbf{v}')$ in a neighbourhood of $(t.\mathbf{1}_{r\times 1}, t.\mathbf{1}_{(k-r)\times 1}, t.\mathbf{1}_{(k-r)\times 1})$, by A3. Thus (3.3.8) indeed converges to the limiting quantity in (ii) of this lemma.

By a similar argument, it can be shown that the terms of the form (3.3.6) are of the order $a_n^{r-l}.O(1), l \leq r-1$, and those of the form (3.3.7) are of the order $a_n^{r-m}.O(1), 0 \leq m \leq \min\{l, p\}$, where

$$m = \text{card } (\{\alpha(1), \ldots, \alpha(l)\} \cap \{\beta(1), \ldots, \beta(p)\})$$

Hence the terms of both type go to zero as $n \rightarrow \infty$, and our proof is complete.□

Corollary 3.3.1

Let $h, \tilde{h} : I\!R^k \rightarrow I\!R$ satisfy A3, $h, \tilde{h}$ being symmetric. Then under A1 - A5,

$$(na_n)^r var(U_n(h_n^{(r)}) - U_n(\tilde{h}_n^{(r)}))$$

$$= E((na_n)^{r/2}\{U_n(h_n^{(r)}) - U_n(\tilde{h}_n^{(r)})\})^2$$

$$\rightarrow E_t\{m_r(Y_1, \ldots, Y_r) - \tilde{m}_r(Y_1, \ldots, Y_r)\}^2 (f_1(t) \int_{I\!R} K^2(u)du)^r$$

$$(f_1(t))^{2(k-r)} r!, 1 \leq r \leq k,$$

as $n \rightarrow \infty$.

Proof: Note, simply, that

$$h_n^{(r)} - \tilde{h}_n^{(r)} = [(I - Q_1)(I - Q_2) \ldots (I - Q_r)Q_{r+1}, \ldots, Q_k]$$

$$(h(y_1, \ldots, y_k) - \tilde{h}(y_1, \ldots, y_k))a_n^{-k} \prod_{j=1}^{k} K_n(t, x_j)$$

The result now follows immediately from Lemma 3.3.1. □

Let $\varphi_i, 1 \leq i \leq l$, be simple functions on $(I\!R, \mathcal{B}(I\!R))$. Denote

$$U_n(\varphi_i^{(r)}, t) := \binom{n}{r}^{-1} \sum_{1 \leq \alpha(1) < \ldots < \alpha(r) \leq n} \prod_{j=1}^{r} [\varphi_i(Y_{\alpha(j)})a_n^{-1} K_n(t, X_{\alpha(j)}) - E\varphi_i(Y_1)a_n^{-1} K_n(t, X_1)], \; r \geq 1, \; 1 \leq i \leq l.$$

Then we have the following result:

Lemma 3.3.2

Let the assumptions A1 - A5 hold. Then, for any $c_i \in I\!R, 1 \le i \le l$,

$$\sum_{i=1}^{l} c_i (na_n)^{r/2} U_n(\varphi_i^{(n)}, t) \xrightarrow{d} \sum_{i=1}^{l} c_i (\|\varphi_i\|_t)^r H_r(I_1^{(t)}(\frac{\varphi_i}{\|\varphi_i\|_t}))$$

$$(f_1(t) \int_{I\!R} K^2(u)du)^{r/2}$$

as $n \to \infty$, where $(\xrightarrow{d})$ denotes convergence in distribution,
$\|\varphi_i\|_t = (E_t \varphi_i^2(Y_1))^{1/2}$, *$H_r(\cdot)$ is the r-th Hermite polynomial as in Section 3.2 and*

$$(I_1^{(t)}(\frac{\varphi_1}{\|\varphi_1\|_t}), \ldots, I_1^{(t)}(\frac{\varphi_l}{\|\varphi_l\|_t})) \sim N_l(\mathbf{0}, ((\sigma_{ij}))_{l \times l}),$$

where

$$\sigma_{ij} = (E_t \varphi_i(Y_1)\varphi_j(Y_1)) \|\varphi_i\|_t^{-1} \|\varphi_j\|_t^{-1}, \; 1 \le i, j \le l.$$

Proof: First, denote

$$\varphi_i^n(\mathbf{z}) := \varphi_i(y) a_n^{-1} K_n(t, x) - E\varphi_i(Y_1) a_n^{-1} K_n(t, X_1), 1 \le i \le l$$

where $\mathbf{z} = (x, y)$.

Then,

$$\begin{aligned}
& (na_n)^{r/2} U_n(\varphi_i^{(r)}, t) \\
= \; & \frac{(n-r)! n^{r/2}}{n!} r! \sum_{1 \le \alpha(1) < \cdots < \alpha(r) \le n} \sqrt{a_n}\varphi_i^n(\mathbf{Z}_{\alpha(1)}) \cdots \sqrt{a_n}\varphi_i^n(\mathbf{Z}_{\alpha(r)}) \\
= \; & b_n n^{-r/2} \sum_{1 \le \alpha(1) \ne \cdots \ne \alpha(r) \le n} \sqrt{a_n}\varphi_i^n(\mathbf{Z}_{\alpha(1)} \cdots \sqrt{a_n}\varphi_i^n(\mathbf{Z}_{\alpha(r)}) \qquad . \qquad (3.3.10)
\end{aligned}$$

where $b_n \to 1$ as $n \to \infty$.

Further,

$$\sum_{i=1}^{l} c_i (n^{-1/2} \sum_{\alpha=1}^{n} \sqrt{a_n}\varphi_i^n(\mathbf{Z}_\alpha))$$

$$\xrightarrow{d} (f_1(t) \int_{I\!R} K^2(u)du)^{1/2} \sum_{i=1}^{l} c_i \|\varphi_i\|_t I_1^{(t)}(\varphi_i / \|\varphi_i\|_t) \qquad (3.3.11)$$

as $n \to \infty$. This follows by applying the Berry-Esseen Theorem (cf. Chung (1974), Theorem 7.4.1, p.225) to

$$n^{-1/2} \sum_{\alpha=1}^{n} (\sqrt{a_n} \sum_{i=1}^{l} c_i \varphi_i^n(\mathbf{Z}_\alpha))$$

as in Lemma 1 of Schuster (1972). Note that the φ_i's are bounded (hence has a finite third moment each) $1 \leq i \leq l$, and the continuity of the functions

$$\left. \begin{array}{c} E(\varphi_i(Y_1)|X_1 = x) \\ E(\varphi_i(Y_1)\varphi_j(Y_1)|X_1 = x) \end{array} \right\} 1 \leq i, j \leq l,$$

at $X_1 = t$ follows by Scheffe's Lemma, as shown in the proof of Lemma 3.3.1. Hence the proof of Lemma 1 of Schuster (1972) can be easily adapted and (3.3.11) holds. In particular, we have

$$(V_{n1}, \ldots, V_{nl})' \xrightarrow{d} (I_1^{(t)}(\frac{\varphi_1}{\|\varphi_1\|_t}), \ldots, I_1^{(t)}(\frac{\varphi_l}{\|\varphi_l\|_t}))'$$

as $n \to \infty$, where

$$V_{ni} = (f_1(t) \int_{I\!R} K^2(u) du)^{-1/2} \|\varphi_i\|_t^{-1} n^{-1/2} \sum_{\alpha=1}^{n} \sqrt{a_n} \varphi_i^n(\mathbf{Z}_\alpha)$$

It follows that

$$\sum_{i=1}^{l} c_i H_r(V_{ni}) \xrightarrow{d} \sum_{i=1}^{l} H_r(I_1^{(t)}(\frac{\varphi_i}{\|\varphi_i\|_t})),$$

as $n \to \infty$. Note that each $V_{ni}, 1 \leq i \leq l$, has the following structure:

$$V_{ni} = \sum_{\alpha=1}^{n} v_\alpha^{(ni)} \tag{3.3.12}$$

where $v_\alpha^{(ni)}, 1 \leq \alpha \leq n$, has an obvious definition. Further

$$(na_n)^{r/2} U_n(\varphi_i^{(r)}, t)$$

$$= b_n(f_1(t) \int_{I\!R} K^2(u) du)^{r/2} (\|\varphi_i\|_t)^r \sum_{1 \leq \alpha(1) \neq \cdots \neq \alpha(r) \leq n} v_{\alpha(1)}^{(ni)} \cdots v_{\alpha(r)}^{(ni)},$$

by (3.3.10). Hence, to prove the lemma, it suffices to show that

$$\sum_{1\le\alpha(1)\ne\cdots\ne\alpha(r)\le n} v_{\alpha(1)}^{(ni)}\cdots v_{\alpha(r)}^{(ni)} \rightarrow H_r(\sum_{\alpha=1}^{n} v_\alpha^{(ni)})$$

in probability as $n\rightarrow\infty$, for each $1\le i\le l$. Fix $i=1$. Now, we have

$$\sum_{\alpha=1}^{n} v_\alpha^{(n1)} \xrightarrow{d} N(0,1), \max_{1\le\alpha\le n} P\{|v_\alpha^{(n1)}|>\varepsilon\}\rightarrow 0,$$

for every $\varepsilon>0$, as $n\rightarrow\infty$. These two facts imply that (see, for example, Laha and Rohatgi (1979), Proposition 5.3.2, p.312)

$$\max_{1\le\alpha\le n} |v_\alpha^{(n1)}|\rightarrow 0, \sum_{\alpha=1}^{n}(v_\alpha^{(n1)})^2\rightarrow 1 \tag{3.3.13}$$

in probability, as $n\rightarrow\infty$. Now, by the identity in the Appendix of Rubin and Vitale (1980)

$$\sum_{1\le\alpha(1)\ne\cdots\ne\alpha(r)\le n} v_{\alpha(1)}^{(n1)}\cdots v_{\alpha(r)}^{(n1)}$$

$$=\sum_{\mathcal{A}}\prod_{A\in\mathcal{A}}\{(-1)^{|A|-1}(|A|-1)!\sum_{\alpha=1}^{n}(v_\alpha^{(n1)})^{|A|}\}, \tag{3.3.14}$$

where $\mathcal{A}$ runs over the class of all partitions of $\{1,2,\ldots,r\}$ into non-empty disjoint sets, and $|A|$ denotes cardinality of $A\subseteq\{1,2,\ldots,r\}$. Hence, combining (3.3.13) and (3.3.14), as in the proof of the Theorem in Rubin and Vitale (1980), we get the required convergence in probability.□

Thus, we have shown that

$$(na_n)^{r/2}U_n(\varphi_i^{(r)},t) \xrightarrow{d} (f_1(t)\int_{I\!R} K^2(u)du)^{r/2}\sqrt{r!}I_r^{(t)}(\varphi_i^{\otimes r})$$

where $I_r^{(t)}$ denotes the r-th order multiple Wiener integral with respect to P_t (refer to Section 3.2). In the next lemma, we establish the result for $U_n(h_n^{(r)}), 0\le r\le k$, where the $h_n^{(r)}, 0\le r\le k$, are as in (3.3.3).

Lemma 3.3.3 *Let A1 - A5 hold. Then*

$$\sum_{r=0}^{k}(na_n)^{r/2}U_n(h_n^{(r)}) \xrightarrow{d} \sum_{r=0}^{k}(f_1(t)\int_{I\!R} K^2(u)du)^{r/2}(f_1(t))^{k-r}\sqrt{r!}I_r^{(t)}(m_r),$$

as $n \to \infty$ where $m_r, 0 \le r \le k$, are as defined in Lemma 3.3.1 (ii).

Proof: First, note that it is enough to consider $\sum_{r=1}^{k}(na_n)^{r/2}U_n(h_n^{(r)})$, because $U_n(h_n^{(0)}) \equiv Eh.\prod K_n$, a constant $\forall n \ge k$, and

$$Eh.\prod K_n \to m(\mathbf{t})(f_1(t))^k,$$

as $n \to \infty$, by Bochner's Theorem. Here $\mathbf{t}$ denotes $t.\mathbf{1}$. Next, denote

$$f_1(t)\int_{\mathbb{R}} K^2(u)du := G, f_1(t) := g$$

for the rest of the proof. Now fix an arbitrary $\varepsilon > 0$, and choose $c_{rp} \in \mathbb{R}, \varphi_{rp}$ simple functions on $(\mathbb{R}, \mathcal{B}(\mathbb{R})), 1 \le p \le l_r, 1 \le r \le k$, such that

$$\sum_{r=1}^{k} G^r g^{2(k-r)} r! E_t\{m_r(Y_1, \dots, Y_r) - \sum_{p=1}^{l_r} c_{rp}\varphi_{rp}(Y_1)\dots\varphi_{rp}(Y_r)\}^2 < \varepsilon. \tag{3.3.15}$$

Such a choice is possible, by Lemma 3.2.3. Note that

$$m_r \in L_2^{\text{sym}}(\mathbb{R}^r, \mathcal{B}(\mathbb{R}^r), \underbrace{P_t \times \dots \times P_t}_{\text{r-fold}}), r \ge 1.$$

As before, let us denote by E_0 the expectation on the probability space that supports the Gaussian process

$$\{I_1^{(t)}(\varphi) : \varphi \in L_2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_t)\}.$$

We shall show that

$$\begin{aligned} & E\exp(i\theta \textstyle\sum_{r=1}^{k}(na_n)^{r/2}U_n(h_n^{(r)})) \\ \to \quad & E_0\exp(i\theta \textstyle\sum_{r=1}^{k} G^{r/2}g^{(k-r)}\sqrt{r!}I_r^{(t)}(m_r)), \forall\theta \in \mathbb{R} \end{aligned} \tag{3.3.16}$$

as $n \to \infty$. Fix $\theta \in \mathbb{R}$. We shall use the fact that, for any two square-integrable random variables S, T on some probability space,

$$|E\exp(i\theta S) - E\exp(i\theta T)| \le |\theta|(E(S-T)^2)^{1/2}.$$

Now, it follows from Lemma 3.3.2 that

$$\sum_{r=1}^{k}\sum_{p=1}^{l_r} c_{rp}(na_n)^{r/2}U_n(\varphi_{rp}^{(r)},t)$$

$$= \sum_{r=1}^{k}(na_n)^{r/2}U_n(\sum_{p=1}^{l_r} c_{rp}\varphi_{rp}^{(r)},t)$$

$$\xrightarrow{d} \sum_{r=1}^{k} G^{r/2}\sum_{p=1}^{l_r} c_{rp}\sqrt{r!}I_r^{(t)}(\varphi_{rp}^{\bigotimes r}) \tag{3.3.17}$$

as $n \to \infty$. After a little modification of Corollary 3.3.1, it follows that

$$E\{\sum_{r=1}^{k}(na_n)^{r/2}(U_n(h_n^{(r)}) - \sum_{p=1}^{l_r} c_{rp}U_n(\varphi_{rp}^{(r)},t)g^{k-r})\}^2$$

$$= \sum_{r=1}^{k}(na_n)^r E\{U_n(h_n^{(r)}) - \sum_{p=1}^{l_r} c_{rp}U_n(\varphi_{rp}^{(r)},t).g^{k-r}\}^2$$

$$\to \sum_{r=1}^{k} G^r.g^{2(k-r)}r!E_t\{m_r - \sum_{p=1}^{l_r} c_{rp}\varphi_{rp}^{\bigotimes r}\}^2 \tag{3.3.18}$$

as $n \to \infty$. Also,

$$E_0(\sum_{r=1}^{k} G^{r/2}g^{(k-r)}\sqrt{r!}\{I_r^{(t)}(m_r) - \sum_{p=1}^{l_r} c_{rp}I_r^{(t)}(\varphi_{rp}^{\bigotimes r})\})^2$$

$$= \sum_{r=1}^{k} G^r g^{2(k-r)}r!E_t\{m_r - \sum_{p=1}^{l_r} c_{rp}\varphi_{rp}^{\bigotimes r}\}^2, \tag{3.3.19}$$

by the isometry in the definition of $I_r^{(t)}(\cdot)$ and Lemma 3.2.2. Put

$$S_n := \sum_{r=1}^{k}(na_n)^{r/2}U_n(h_n^{(r)})$$

$$T_n := \sum_{r=1}^{k}(na_n)^{r/2}g^{k-r}\sum_{p=1}^{l_r} c_{rp}U_n(\varphi_{rp}^{(r)},t),$$

$$S := \sum_{r=1}^{k} G^{r/2}g^{k-r}\sqrt{r!}I_r^{(t)}(m_r)$$

$$T := \sum_{r=1}^{k} G^{r/2}g^{k-r}\sqrt{r!}(\sum_{p=1}^{l_r} c_{rp}I_r^{(t)}(\varphi_{rp}^{\bigotimes r}))$$

Then

$$|E\exp(i\theta S_n) - E_0\exp(i\theta S)|$$

$$\leq |\theta|\{E(S_n - T_n)^2\}^{1/2} + |E\exp(i\theta T_n) - E_0\exp(i\theta T)| + |\theta|\{E_0(T - S)^2\}^{1/2}$$

Hence

$$\overline{\lim}_{n\to\infty}|E\exp(i\theta S_n) - E_0\exp(i\theta S)| \leq |\theta|\sqrt{\varepsilon} + 0 + |\theta|.\sqrt{\varepsilon} = 2|\theta|\sqrt{\varepsilon},$$

combining, (3.3.15), (3.3.17), (3.3.18) and (3.3.19). Since $\varepsilon > 0$ is arbitrary, (3.3.16) follows.□

The lemmas proved above give the following theorem which is the main result of this subsection:

Theorem 3.3.1 *Let the assumptions A1 - A5 hold. Define*

$$r_0 = \inf\{r \geq 1 : m_r(y_1, \ldots, y_r) \not\equiv 0\}$$

Then

$$(na_n)^{r_0/2}\{U_n(h, t.1) - \sum_{j=0}^{r_0-1}\binom{k}{j}U_n(h_n^{(j)})\}$$

$$\xrightarrow{d} \binom{k}{r_0}(f_1(t)\int_{\mathbb{R}} K^2(u)du)^{r_0/2}(f_1(t))^{k-r_0}\sqrt{r_0!}I_{r_0}^{(t)}(m_{r_0})$$

as $n\to\infty$.

Proof: Note that

$$(na_n)^{r_0/2}\{U_n(h, t.1) - \sum_{j=0}^{r_0-1}\binom{k}{j}U_n(h_n^{(j)})\}$$

$$= \binom{k}{r_o}(na_n)^{r_0/2}U_n(h_n^{(r_0)}) + (na_n)^{r_0/2}[\sum_{j=r_0+1}^{k}\binom{k}{j}U_n(h_n^{(j)})] \qquad (3.3.20)$$

Now, by Lemma 3.3.3,

$$\binom{k}{r_0}(na_n)^{r_0/2}U_n(h_n^{(r_0)}) \xrightarrow{d} \binom{k}{r_0}(f_1(t)\int_{\mathbb{R}} K^2(u)du)^{r_0/2}(f_1(t))^{k-r_0}\sqrt{r_0!}I_{r_0}^{(t)}(m_{r_0})$$

Further,

$$
\begin{aligned}
& E((na_n)^{r_0/2}[\sum_{j=r_0+1}^{k}\binom{k}{j}U_n(h_n^{(j)})])^2 \\
= \; & (na_n)^{r_0}\sum_{j=r_0+1}^{k}\binom{k}{j}^2 j!O(\frac{1}{n^j})E(h_n^{(j)})^2, \text{ by Lemma 3.3.1 (i),} \\
= \; & \sum_{j=r_0+1}^{k} a_n^j E(h_n^{(j)})^2 O((na_n)^{-(j-r_0)}) \\
= \; & \sum_{j=r_0+1}^{k} O((na_n)^{-(j-r_0)}), \text{ by Lemma 3.3.1 (ii),} \\
& \to 0
\end{aligned}
$$

as $n \to \infty$, by A5. This completes the proof.□

This brings us to the end of Subsection 3.3.1. In the next subsection we deal with the non-diagonal case.

3.3.2 The non-diagonal case: $t_i, 1 \leq i \leq k$, are not necessarily equal

Let a point $\mathbf{t}^0 \in \mathbb{R}^k$ be given. Then the co-ordinates of $\mathbf{t}^0$ can be partitioned into m equivalence classes, $1 \leq m \leq k$, such that k_1 of them are equal to t_1, k_2 of them are equal to $t_2, t_1 \neq t_2$, and so on. Here, $k_i \geq 1, 1 \leq i \leq m, \sum_{i=1}^{m} k_i = k$. Further, since $h(\cdot)$ is symmetric, $m(\mathbf{t}^0)$ is also symmetric, and we may write, WOLG

$$
(t_1^0, t_2^0, \ldots, t_k^0)' = (\underbrace{t_1, \ldots, t_1}_{k_1\text{-times}}, \underbrace{t_2, \ldots, t_2}_{k_2\text{-times}}, \ldots, \underbrace{t_m, \ldots, t_m}_{k_m\text{-times}}).
$$

Now, in this case

$$
(h\prod K_n)(\mathbf{z}_1, \ldots, \mathbf{z}_k) = h(y_1, \ldots, y_k)a_n^{-k_1}\prod_{j=1}^{k_1} K_n(t_1, x_j) \ldots
$$

$$
a_n^{-k_m} \prod_{j=k_1+\cdots+k_{m-1}+1}^{k} K_n(t_m, x_j).
$$

Also,

$$U_n(h, \mathbf{t}^0) = \frac{(n-k)!}{n!} k_1! \ldots k_m! \sum_{\beta(n,\mathbf{k})} h(Y_{\beta(1)}, \ldots, Y_{\beta(k)}) a_n^{-k} \prod_{j=1}^{k} K_n(t_j^0, X_{\beta(j)}),$$

where

$$\begin{aligned} & \beta(n, \mathbf{k}) \\ := \; & \{(\beta(1), \ldots, \beta(k) : 1 \leq \beta(1) < \ldots < \beta(k_1) \neq \beta(k_1+1) < \ldots < \beta(k_1+k_2) \neq \\ & \beta(k_1+k_2+1) < \ldots < \beta(k) \leq n\} \end{aligned}$$

Now, we shall apply the 'D-M decomposition' to $(h. \prod K_n)$ as in (3.3.2). In order to keep track of the t_i^0's, we shall denote $Q_j \equiv Q_j(t_i), k_1 + \ldots + k_{i-1} + 1 \leq j \leq k_1 + \ldots + k_i, 1 \leq i \leq m$, with $k_0 = 0$. Thus

$$\begin{aligned} & (h \prod K_n)(z_1, \ldots, z_k) \\ = \; & \prod_{j=1}^{k_1} [(I - Q_j(t_1)) + Q_j(t_1)] \prod_{j=k_1+1}^{k_1+k_2} [(I - Q_j(t_2) + Q_j(t_2)] \ldots \\ & \prod_{j=k_1+\ldots+k_{m-1}+1}^{k} [(I - Q_j(t_m)) + Q_j(t_m)](h. \prod K_n)(\mathbf{z}_1, \ldots, \mathbf{z}_k) \\ = \; & \sum_{r=0}^{k} [\sum_{\mathbf{r}(k,m)} \sum_{\mathbf{i}(r,k,m)} h_n^{(r_1,\ldots,r_m)}(\mathbf{z}_{i(1)}, \ldots, \mathbf{z}_{i(r_1)}; \ldots; \mathbf{z}_{i(r_1+\ldots+r_{m-1}+1)}, \ldots, \mathbf{z}_{i(r)})]. \end{aligned} \quad (3.3.21)$$

where

$$\mathbf{r}(k, m) := \{(r_1, \ldots, r_m) : 0 \leq r_i \leq k_i, 1 \leq i \leq m, \sum_{i=1}^{m} r_i = r\},$$

$$\begin{aligned} & \mathbf{i}(r, k, m) \\ := \; & \{(i(1), \ldots, i(r)) : \\ & 1 \leq i(1) < \ldots < i(r_1) \leq k_1; k_1 + 1 \leq i(r_1+1) < \ldots < i(r_1+r_2) \\ & \leq k_1 + k_2; \ldots; k_1 + \ldots + k_{m-1} + 1 \leq i(r_1 + \ldots + r_{m-1} + 1) < \\ & \ldots < i(r) \leq k\}, \end{aligned}$$

and

$$h_n^{(r_1,\ldots,r_m)}(\mathbf{z}_1, \ldots, \mathbf{z}_{r_1}; \mathbf{z}_{k_1+1}, \ldots, \mathbf{z}_{k_1+r_2}; \ldots; \mathbf{z}_{k_1+\cdots+k_{m-1}+1}, \ldots, \mathbf{z}_{k_1+\ldots+k_{m-1}+r_m})$$

$$
\begin{aligned}
:= & [\prod_{j=1}^{r_1}(I-Q_j(t_1)) \prod_{j=r_1+1}^{r_1} Q_j(t_1) \prod_{j=k_1+1}^{k_1+r_2}(I-Q_j(t_2)) \prod_{j=k_1+r_2+1}^{k_1+k_2} Q_j(t_2)\dots \\
& \prod_{j=k_1+\dots+k_{m-1}+1}^{k_1+\dots+k_{m-1}+r_m}(I-Q_j(t_m)) \prod_{j=k_1+\dots+k_{m-1}+r_m+1}^{k} Q_j(t_m)](h.\prod K_n)(\mathbf{z}_1,\dots,\mathbf{z}_k).
\end{aligned}
$$

Note that, by the symmetry of $h(\cdot)$, the function $h_n^{(r_1,\dots,r_m)} : I\!R^r \to I\!R$ is the same as the function

$$
\begin{aligned}
& \tilde{h}_n^{(r_1,\dots,r_m)}(\mathbf{z}_1,\dots,\mathbf{z}_{r_1};\mathbf{z}_{r_1+1},\dots,\mathbf{z}_{r_1+r_2};\dots;\mathbf{z}_{r_1+\dots+r_{m-1}+1},\dots,\mathbf{z}_r) \\
:= & \left[\prod_{j=1}^{r_1}(I-Q_j(t_1)) \prod_{j=r_1+1}^{r_1+r_2}(I-Q_j(t_2))\dots \prod_{j=r_1+\dots+r_{m-1}+1}^{r}(I-Q_j(t_m)) \prod_{j=r+1}^{r+(k_1-r_1)} Q_j(t_1)\right. \\
& \left.\prod_{j=r+(k_1-r_1)+1}^{r+(k_1-r_1)+(k_2-r_2)} Q_j(t_2)\dots \prod_{j=r_m+(k-k_m)+1}^{k} Q_j(t_m)\right](h.\prod \tilde{K}_n)(\mathbf{z}_1,\dots,\mathbf{z}_k), \qquad (3.3.22)
\end{aligned}
$$

where $\tilde{K}_n$ is the corresponding permutation in $\prod_{j=1}^{k} K_n(t_j^0, x_j)$. The advantage is that, in $\tilde{h}_n^{(r_1,\dots,r_m)}$ the subscripts of the $\mathbf{z}_i'$s are in a consecutive sequence 1 to r, which was not the cae with $h_n^{(r_1,\dots,r_m)}$. Note that, the argument sequence in $\tilde{h}_n^{(r_1,\dots,r_m)}$ actually begins with $\mathbf{z}_1,\dots,\mathbf{z}_{r_{i_0}})$ rather than $(\mathbf{z}_1,\dots,\mathbf{z}_{r_1})$, where

$$i_0 = \inf\{1 \le i \le m : r_i > 0\}.$$

For example, let $k=4, m=2$ and $r=2$. Then we have, for $r_1=0, r_2=2$,

$$
\begin{aligned}
& h_n^{(0,2)}(\mathbf{z}_3,\mathbf{z}_4) \\
= & Q_1(t_1)Q_2(t_1)(I-Q_3(t_2))(I-Q_4(t_2))h(y_1,y_2,y_3,y_4) \\
& a_n^{-2}\prod_{j=1}^{2} K_n(t_1,x_j)a_n^{-2}\prod_{j=3}^{4} K_n(t_2,x_j),
\end{aligned}
$$

whereas

$$
\begin{aligned}
& \tilde{h}_n^{(0,2)}(\mathbf{z}_1,\mathbf{z}_2) \\
= & (I-Q_1(t_2))(I-Q_2(t_2))Q_3(t_1)Q_4(t_1)h(y_1,y_2,y_3,y_4) \\
& a_n^{-2}\prod_{j=1}^{2} K_n(t_2,x_j)a_n^{-2}\prod_{j=3}^{4} K_n(t_1,x_j),
\end{aligned}
$$

and it is clear that the two functions are the same. Hence, in the sequel, by $h_n^{(r_1,\dots,r_m)}$ we shall mean $\tilde{h}_n^{(r_1,\dots,r_m)}$. Finally, we have

$$U_n(h,\mathbf{t}^0) = Eh\prod K_n + \sum_{r=1}^{k}\left[\sum_{\mathbf{r}(k,m)}\binom{k_1}{r_1}\cdots\binom{k_m}{r_m}U_n(h_n^{(r_1,\dots,r_m)},\mathbf{t}^0)\right] \tag{3.3.23}$$

Note that, when $m=1$ and $k_1=k$, all the expressions above reduce to those in the previous subsection. Now, let A1′, A2′ and A3′ be the assumptions A1.A2 and A3 respectively, modified by substituting $\mathbf{t}^0$ for $t.\mathbf{1}$, e.g.,

A1′. $f(y|\cdot)$ is continuous at t_i^0 for all $y\in I\!R, 1\le i\le k$.

Then we have the following analogue of Lemma 3.3.1:

Lemma 3.3.4. *Let A1′ - A3′, A4 and A5 hold. Then*

(i)

$$\begin{aligned}
&\mathrm{var}(U_n(h,\mathbf{t}^0))\\
&= \sum_{k=1}^{r} E[\sum_{\mathbf{r}(k,m)}\binom{k_1}{r_1}\cdots\binom{k_m}{r_m}U_n(h_n^{(r_1,\dots,r_m)},\mathbf{t}^0)]^2\\
&= \sum_{r=1}^{k}\frac{(n-r)!}{n!}[\sum_{\mathbf{r}(k,m)}\sum_{\mathbf{r}(k,m)}\binom{\mathbf{k}_m}{\mathbf{r}_m}\binom{\mathbf{k}_m}{\mathbf{r}'_m}\frac{\mathbf{r}_m!\mathbf{r}'_m!}{r!}\\
&\quad\{\sum_{\beta(r,r)}\sum_{\beta(r',r)} Eh_n^{(\mathbf{r}_m)}(\mathbf{Z}_{\beta(1)},\dots,\mathbf{Z}_{\beta(r)}).h_n^{(\mathbf{r}'_m)}(\mathbf{Z}_{\beta'(1)},\dots,\mathbf{Z}_{\beta'(r)})\}],
\end{aligned}$$

where

$$\binom{\mathbf{k}_m}{\mathbf{r}_m} = \prod_{i=1}^{m}\binom{k_i}{r_i},\ \mathbf{r}_m=\prod_{i=1}^{m} r_i!,\ h_n^{(\mathbf{r}_m)}\equiv h_n^{(r_1,\dots,r_m)},$$

and $\beta(r,r) := \{(\beta(1),\dots,\beta(r)) : 1\le\beta(1),\dots,\beta(r_1)\ne\beta(r_1+1)<\dots<\beta(r_1+r_2)\ne\dots\ne\beta(r_1+\dots+r_{m-1}+1)<\dots<\beta(r)\le r\}$, *and* $\binom{\mathbf{k}_m}{\mathbf{r}'_m}$ *etc. are defined analogously. (Note that* card $(\beta(r,r)) = \frac{r!}{r_1!\dots r_m!}$.)

(ii) Let

$$\begin{aligned}
&S_n(\mathbf{r}_m,\mathbf{r}'_m)\\
&:= \sum_{\beta(r,r)}\sum_{\beta(r',r')} Eh_n^{(\mathbf{r}_m)}(\mathbf{Z}_{\beta(1)},\dots,\mathbf{Z}_{\beta(r)})h_n^{(\mathbf{r}'_m)}(\mathbf{Z}_{\beta'(1)},\dots,\mathbf{Z}_{\beta'(r)})
\end{aligned}$$

Then $\lim_{n\to\infty} a_n^r S_n(\mathbf{r}_m, \mathbf{r}'_m) = 0$ *if* $(r_1, \ldots, r_m) \neq (r'_1, \ldots, r'_m)$, *and*

$$\begin{aligned}
&\lim_{n\to\infty} a_n^r S_n(\mathbf{r}_m, \mathbf{r}_m) \\
&= \frac{r!}{r_1! \ldots r_m!} \lim_{n\to\infty} a_n^r E(h_n^{(\mathbf{r}_m)}(\mathbf{Z}_1, \ldots, \mathbf{Z}_r))^2 \\
&= \frac{r!}{r_1! \ldots r_m!} \prod_{i=1}^{m} (f_1(t_i) \int_{I\!R} K^2(u)du)^{r_i} \prod_{i=1}^{m} (f_1(t_i))^{2(k_i - r_i)} \\
&\quad E_{t_1,\ldots,t_m}(m_{r_1,\ldots,r_m}(Y_1, \ldots, Y_r))^2,
\end{aligned} \tag{3.3.24}$$

where $E_{t_1,\ldots,t_m}(m_{r_1,\ldots,r_m}(Y_1, \ldots, Y_r))^2$

$$= \int_{I\!R^r} \{m_{r_1,\ldots,r_m}(y_1, \ldots, y_r)\}^2 \prod_{i=1}^{r_1} f(y_i|t_1) \ldots \prod_{i=r_1+\ldots+r_{m-1}+1}^{r} f(y_i|t_m) dy, \ldots dy_r,$$

and

$$m_{r_1,\ldots,r_m}(y_1, \ldots, y_r)$$

$$= E\{h(y_1, \ldots, y_r, Y_{r+1}, \ldots, Y_k) | X_{r+j} = t_1, 1 \le j \le k_1 - r_1; \ldots; X_{r+j} = t_m,$$

$$k - (k_m - r_m) + 1 \le j \le k\}.$$

Proof: (i) The first equality follows from the fact that the functions

$$\{h_n^{(r_1,\ldots,r_m)}, 1 \le r \le\}$$

are 'canonical' (cf. (3.3.4)). For the second equality, note that

$$EU_n(h_n^{(\mathbf{r}_m)}, \mathbf{t}^0) U_n(h_n^{(\mathbf{r}'_m)}, \mathbf{t}^0)$$

$$= \{\frac{(n-r)!}{n!}\}^2 \mathbf{r}_m! \mathbf{r}'_m! E(\sum_{\beta(n,r)} h_n^{(\mathbf{r}'_m)})(\sum_{\beta(n,r')} h_n^{(\mathbf{r}'_m)}) \tag{3.3.25}$$

Now, terms in both $\sum_{\beta(n,r)} h_n^{(\mathbf{r}_m)}$ and $\sum_{\beta(n,r')} h_n^{(\mathbf{r}'_m)}$ are based on $\binom{n}{r}$ choices of r random variables out of $(\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$. But in the former, each choice is permuted card $(\beta(r,r))$ times, whereas in the latter, it is permuted card $(\beta(r', r'))$ times. Since, by the canonicity property, a product of terms corresponding to distinct choices has zero expectation, the right-hand side of (3.3.25) boils down to the following

$$\{\frac{(n-r)!}{r!}\}^2 \mathbf{r}_m!\mathbf{r}'_m!\binom{n}{r}\sum_{\beta(r,r)}\sum_{\beta(r',r')} Eh_n^{(\mathbf{r}_m)}h_n^{(\mathbf{r}'_m)}.$$

Hence the proof.

(ii) By (3.3.22),

$$\begin{aligned}
& h_n^{(\mathbf{r}_m)}(\mathbf{z}_1,\ldots,\mathbf{z}_r) \\
= \ & a_n^{-r_1}\prod_{j=1}^{r_1} K_h(t_1,x_j)\ldots a_n^{-r_m}\prod_{j=r_1+\ldots+r_{m-1}+1}^{r} K_n(t_m,x_j) \\
& .Eh(y_1,\ldots,y_r,Y_{r+1},\ldots,Y_k)a_n^{-(k_1-r_1)}\prod_{j=r+1}^{r+(k_1-r_1)} K_n(t_1,X_j)\ldots \\
& a_n^{-(k_m-r_m)}\prod_j K_n(t_m,X_j) + \text{ (remaining terms)},
\end{aligned}$$

where the 'remaining terms' are analogous to those in the proof of Lemma 3.3.1 (ii). Thus

$$\begin{aligned}
& a_n^r E(h_n^{(r_1,\ldots,r_m)}(\mathbf{Z}_{\beta(1)},\ldots,\mathbf{Z}_{\beta(r)}))^2, \beta\in\beta(r,r) \\
= \ & a_n^r E(h_n^{(\mathbf{r}_m)}(\mathbf{Z}_1,\ldots,\mathbf{Z}_r))^2 \\
= \ & E\{a_n^{-r_1}\prod_{j=1}^{r_1} K_n^2(t_1,X_j)\ldots a_n^{-r_m}\prod_{j=r_1+\ldots+r_{m-1}+1}^{r} K_n^2(t_m,X_j) \\
& h_{nr}^2(Y_1,\ldots,Y_r)\} + \text{(remaining terms)}.
\end{aligned}$$

where

$$\begin{aligned}
h_{nr}(y_1,\ldots,y_r) = \ & Eh(y_1,\ldots,y_r,Y_{r+1},\ldots,Y_k)a^{-k_1-r_1}\textstyle\prod_j K_n(t_1,X_j) \\
\ldots \ & a_n^{-(k_m-r_m)}\textstyle\prod_j K_n(t_m,X_j).
\end{aligned}$$

Now, in the same way as in the proof of Lemma 3.3.1 (ii), the 'leading term' above can be shown to converge, as $n\to\infty$, to the expression in (3.3.24) (without the factor $\frac{r!}{r_1!\ldots r_m!}$). The remaining terms are of the order $a_n^p, 1\le p\le r$, hence they all converge to zero as $r\to\infty$. Further, since

$$\text{card}(\beta(r,r)) = \frac{r!}{r_1!\ldots r_m!},$$

the total contribution from the square terms of the above kind will be exactly (3.3.24). Hence, our proof will be complete , if we show that

$$\lim_{n\to\infty} a_n^r E h_n^{(r_m)}(\mathbf{Z}_{\beta(1)},\dots,\mathbf{Z}_{\beta(r)}) h_n^{(r'_m)}(\mathbf{Z}_{\beta'(1)},\dots,\mathbf{Z}_{\beta'(r)}) = 0,$$

where

either : (A) $(r_1,\dots,r_m) = (r'_1,\dots,r'_m)$ but $\beta \neq \beta', \beta, \beta' \in \beta(r,r)$,

or : (B) $(r_1,\dots,r_m) \neq (r'_1,\dots,r'_m)$.

In either case (i.e., (A) or (B)), there exists $I := \{i_1,\dots,i_p\}$, $J := \{j_1,\dots,j_p\}, I, J \subseteq \{1,\dots,r\}$, $1 \leq p \leq r$, such that

$$\begin{aligned} & a_n^r E h_n^{(r_m)}(\mathbf{Z}_{\beta(1)},\dots,\mathbf{Z}_{\beta(r)}).h_n^{(r'_m)}(\mathbf{Z}_{\beta'(1)},\dots,\mathbf{Z}_{\beta'(r)}) \\ = \;& E[\prod_{l=1}^{p}\{a_n^{-1}K_n(t_{i_l}^0)K_n(t_{j_l}^0)\}\prod_{j\notin J}\{a_n^{-1}K_n^2(t_j^0, X_{\beta(j)})\} \\ & h_{nr}(Y_1,\dots,Y_r)h'_{nr}(Y_1,\dots,Y_r)] + \text{(remaining terms)}, \qquad (3.3.26)\end{aligned}$$

and $\beta(i_l) = \beta'(j_l)$ but $t_{i_l}^0 \neq t_{j_l}^0, 1 \leq l \leq p$, while for $i \notin I$, $j \notin J$, both $\beta(i) = \beta'(j)$ and $t_i = t_j$. Here $h'_{nr}(\cdot)$ denotes the function $h_{nr}(\cdot)$ corresponding to $(r'_1,\dots,r'_m)$. Now as $n \to \infty$, the 'remaining terms' go to zero, as before. Consider the 'leading term' in (3.3.26). Assume, for the sake of simplicity, $p = 2$. Let us denote $\max\{t_{i_l}^0, t_{j_l}^0\}$ by $t_{i_l}^0$. Put $\delta_l = t_{i_l}^0 - t_{j_l}^0, l = 1,2$. Denote

$$\begin{aligned} & M_{nr}(x_1,\dots,x_r) \\ := \;& E\{|h_{nr}(Y_1,\dots,Y_r)h'_{nr}(Y_1,\dots,Y_r)| \mid X_1 = x_1,\dots,X_r = x_r\}.\end{aligned}$$

Then the 'leading term' in (3.3.26) is bounded by

$$\begin{aligned} & \int_{\mathbb{R}^r}\{M_{nr}(t_1^0 - a_n u_1,\dots,t_r^0 - a_n u_r)\prod_{j\notin J} f_1(t_j^0 - a_n u_j)K^2(u_j) \\ & \prod_{l=1}^{2} f_1(t_{i_l}^0 - a_n u_{i_l})K(u_{i_l})K(a_n^{-1}\delta_l + u_{i_l})\}du_1\dots du_r \\ = \;& \int_{\{|u_{i_1}|\leq a_n^{-1}\delta_1, |u_{i_2}|\leq a_n^{-1}\delta_2\}}(\dots) + \int_{\{|u_{i_1}|> a_n^{-1}\delta_1, |u_{i_2}|\leq a_n^{-1}\delta_2\}}(\dots)\end{aligned}$$

$$
\begin{aligned}
&+\int_{\{|u_{i_1}|\le a_n^{-1}\delta_1, |u_{i_2}|>a_n^{-1}\delta_2\}}(\ldots)+\int_{\{|u_{i_1}|>a_n^{-1}\delta_1, |u_{i_2}|>a_n^{-1}\delta_2\}}(\ldots) \\
&= T_n^{(1)}+T_n^{(2)}+T_n^{(3)}+T_n^{(4)}, \text{ say.}
\end{aligned}
$$

Now,

$$
\begin{aligned}
T_n^{(1)} &\le \sup_{|u_{i_1}|\le a_n^{-1}\delta_1} \sup_{|u_{i_2}|\le a_n^{-1}\delta_2} K(a_n^{-1}\delta_1+u_{i_1})K(a_n^{-1}\delta_2+u_{i_2}). \\
&\quad \int_{I\!R^r}\{M_{nr}(\ldots)\prod_{j\notin J}(f_1K^2(u_j))\prod_{l=1}^{2} f_1(t_{i_l}^0-a_nu_{i_l})K(u_{i_2})\} \\
&\quad du_1\ldots du_r \\
&\le \sup_{|u_{i_1}|>a_n^{-1}\delta_1} \sup_{|u_{i_2}|>a_n^{-1}\delta_2} K(u_{i_1})K(u_{i_2}).O(1)
\end{aligned}
$$

by the arguments in the proof of Lemma 3.3.1 (ii) and Bochner's Theorem;

$$
\begin{aligned}
T_n^{(2)} &\le \sup_{|u_{i_1}|>a_n^{-1}\delta_1} \sup_{|u_{i_2}|\le a_n^{-1}\delta_2} K(u_{i_1})K(a_n^{-1}\delta_2+u_{i_2}). \\
&\quad \int_{I\!R^r}\{M_{nr}(\ldots)\prod_{j\notin J}(f_1.K^2(u_j))\prod_{l=1}^{2} f_1(t_{i_l}^0-a_nu_{i_l}) \\
&\quad K(a_n^{-1}\delta_1+u_{i_1})K(u_{i_2})\}du_1\ldots du_r \\
&\le \sup_{|u_{i_1}|>a_n^{-1}\delta_1} \sup_{|u_{i_2}|>a_n^{-1}\delta_2} K(u_{i_1})K(u_{i_2}). \\
&\quad \int_{I\!R^r}\{M_{nr}(t_1^0-a_nu_1,\ldots,t_{j_1}^0-a_nu_{i_1},\ldots,t_r^0-a_nu_r)\prod_{j\notin J}(f_1.K^2(u_j)). \\
&\quad f_1(t_{j_1}^0-a_nu_{i_1})f_1(t_{i_2}^0-a_nu_{i_2})K(u_{i_1})K(u_{i_2})\}du_1\ldots du_r \\
&= \sup_{|u_{i_1}|>a_n^{-1}\delta_1} \sup_{|u_{i_2}|>a_n^{-1}\delta_2} K(u_{i_1}).K(u_{i_2}).O(1);
\end{aligned}
$$

similarly,

$$
T_n^{(3)} \text{ and } T_n^{(4)} \le \sup_{|u_{i_1}|>a_n^{-1}\delta_1} \sup_{|u_{i_2}|>a_n^{-1}\delta_2} K(u_{i_1})K(u_{i_2}).O(1).
$$

(Note that the above is similar to the proof of (4), Lemma 1, in Schuster (1972).)

Thus, we have

the 'leading term' in (3.3.26)

$$\leq \sum_{j=1}^{4} T_n^{(j)}$$

$$\leq 4.\prod_{l=1}^{2} \sup_{|u_{i_l}|>a_n^{-1}\delta_l} K(u_{i_l}).O(1)$$

$$\leq 4.\prod_{l=1}^{2} \sup_{|u_{i_l}|>a_n^{-1}\delta_l} a_n \frac{|u_{i_l}|}{\delta_l} K(u_{i_l}).O(1) = O(a_n^2).$$

This completes the proof.□

Combining Lemma 3.3.4 (i) and (ii), we get

$$\lim_{n\to\infty}(na_n)^r E[\sum_{\mathbf{r}(k,m)} \binom{k_1}{r_1}\cdots\binom{k_m}{r_m} U_n(h_n^{(r_1,\ldots,r_m)},\mathbf{t}^0)]^2$$

$$= \sum_{\mathbf{r}(k,m)} \{\binom{k_1}{r_1}^2 \cdots \binom{k_m}{r_m}^2 r_1!\ldots r_m! \prod_{i=1}^{m}(f_1(t_i)\int_{\mathbb{R}} K^2(u)du)^{r_i}(f_1(t_i))^{2(k_i-r_i)}$$

$$E_{t_1,\ldots,t_m}(m_{r_1,\ldots,r_m}(Y_1,\ldots,Y_r))^2\} \qquad (3.3.27)$$

Let P_{t_i} be the probability measure on on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$, given by $f(\cdot|t_i), 1 \leq i \leq m$. Note that $E_{t_1,\ldots,t_m}$ is the expectation with respect to $P_{t_1}\times\cdots\times P_{t_m}$. We shall make use of the modified multiple Wiener integral

$$I_{r_1,\ldots,r_m}^{t_1,\ldots,t_m} : \otimes_{i=1}^m(\sigma(\otimes^{r_i}L_2(\mathbb{R},B(\mathbb{R}),P_{t_i})))\rightarrow\mathcal{W}_{r_1,\ldots,r_m}^{t_1,\ldots,t_m},$$

as defined in (3.2.3) (the superscript $(t_1,\ldots,t_m)$ merely shows the dependence on $t_i, 1 \leq i \leq m$).

Let $\varphi_i, 1 \leq i \leq m$, be simple functions on $(\mathbb{R},B(\mathbb{R}))$. Define

$$\varphi_i^n(\mathbf{z},t_i) := \varphi_i(y)a_n^{-1}K_n(t_i,x) - E\varphi_i(Y_1)a_n^{-1}K_n(t_i,X_1),$$

$1 \leq i \leq m$. Let $U_n(\varphi_1^{(r_1)},\ldots,\varphi_m^{(r_m)},\mathbf{t}^0)$ be the U-statistic based on the function

$$(\varphi_1^n)^{\otimes r_1}\otimes\ldots\otimes(\varphi_m^n)^{\otimes r_m}, r_i \geq 0, 1 \leq i \leq m.$$

Lemma 3.3.5

Let $h : \mathbb{R}^k \to \mathbb{R}$ satisfy A3′. Let also A1′, A2′, A4′ and A5′ hold. Then for every $1 \le r \le k$ and $(r_1, \dots, r_m)$ such that $\sum_{i=1}^m r_i = r, 0 \le r_i \le k_i, 1 \le i \le m$, we have

$$\begin{aligned}
&\lim_{n\to\infty}(na_n)^r E(U_n(h_n^{(r_1,\dots,r_m)}, \mathbf{t}^0) - G(\mathbf{t}^0, \mathbf{r}_m)U_n(\varphi_1^{(r_1)}, \dots, \varphi_m^{(r_m)}, \mathbf{t}^0))^2 \\
= \;& r_1! \dots r_m! \prod_{i=1}^m (f_1(t_i) \int_{\mathbb{R}} K^2(u)du)^{r_i} (f_1(t_i))^{2(k_i - r_i)} \\
& E_{t_1,\dots,t_m}\{m_{r_1,\dots,r_m}(Y_1, \dots, Y_r) - \varphi_1^{\otimes r_1} \otimes \dots \otimes \varphi_m^{\otimes r_m}(Y_1, \dots, Y_r)\}^2,
\end{aligned}$$

where $G(\mathbf{t}^0, \mathbf{r}_m) = \prod_{i=1}^m (f_1(t_i))^{k_i - r_i}$ and $h_n^{(r_1,\dots,r_m)}$ is as defined in (3.3.22).

Proof: Note that $(\varphi_1^n)^{\otimes r_1} \otimes \dots \otimes (\varphi_m^n)^{\otimes r_m}$ satisfies the 'canonicity' property (3.3.4) and $\varphi_1^{\otimes r_1} \otimes \dots \otimes \varphi_m^{\otimes r_m}$ satisfies the assumption A3′. Based on these observations, the result is just a corollary to Lemma 3.3.4. Put

$$R(f_1, K, \mathbf{t}^0) := \prod_{i=1}^m r_i!(f_1(t_i) \int_{\mathbb{R}} K^2(u)du)^{r_i}$$

Then we have to show that

(a)

$$\begin{aligned}
\lim_{n\to\infty}(na_n)^r E(U_n(h_n^{(r_1,\dots,r_m)}, \mathbf{t}^0))^2 = R(f_1, K, \mathbf{t}^0) G^2(\mathbf{t}^0, \mathbf{r}_m) \\
E_{t_1,\dots,t_m}\{m_{r_1,\dots,r_m}(Y_1, \dots, Y_r)\}^2,
\end{aligned}$$

(b)

$$\begin{aligned}
&\lim_{n\to\infty}(na_n)^r G(\mathbf{t}^0, \mathbf{r}_m) E(U_n(h_n^{(r_1,\dots,r_m)}, \mathbf{t}^0) U_n(\varphi_1^{(r_1)}, \dots \varphi_m^{(r_m)}, \mathbf{t}^0)) \\
&= R(f_1, K, \mathbf{t}^0) G^2(\mathbf{t}^0, \mathbf{r}_m) E_{t_1\dots t_m}(m_{r_1,\dots,r_m} \varphi_1^{\otimes r_1} \otimes \dots \otimes \varphi_m^{\otimes r_m}),
\end{aligned}$$

(c)

$$\begin{aligned}
&\lim_{n\to\infty}(na_n)^r G^2(\mathbf{t}^0, \mathbf{r}_m) E(U_n(\varphi_1^{(r_1)}, \dots, \varphi_m^{(r_m)}, \mathbf{t}^0))^2 \\
&= R(f_1, K, \mathbf{t}^0) G^2(\mathbf{t}^0, \mathbf{r}_m) E_{t_1,\dots,t_m}\{\varphi_1^{\otimes r_1} \otimes \dots \otimes \varphi_m^{\otimes r_m}\}^2
\end{aligned}$$

Now, (a) follows from (3.3.27) directly. The proofs of (b) and (c) are similar to that of (3.3.24). □

Lemma 3.3.6

Let $\varphi_i, 1 \le i \le m$, be simple functions as before. Then, as $n \to \infty$,

$$(na_n)^{r/2} U_n(\varphi_1^{(r_1)}, \ldots, \varphi_m^{(r_m)}, \mathbf{t}^0)$$
$$\xrightarrow{d} \prod_{i=1}^{m} \sqrt{r_i!}(f_1(t_i) \int_{I\!R} K^2(u)du)^{r_i/2} I_{r_1,\ldots,r_m}^{t_1,\ldots,t_m}(\varphi_1^{\otimes r_1} \otimes \ldots \otimes \varphi_m^{\otimes r_m})$$

Proof: Note that

$$I_{r_1,\ldots,r_m}^{t_1,\ldots,t_m}(\varphi_1^{\otimes r_1} \otimes \ldots \otimes \varphi_m^{\otimes r_m}) = \bar{H}_{r_1}^{(t_1)}(\varphi_1) \ldots \bar{H}_{r_m}^{(t_m)}(\varphi_m),$$

where

$$\bar{H}_{r_i}^{(t_i)}(\varphi_i) = \frac{\|\varphi_i\|_{t_i}^{r_i}}{\sqrt{r_i!}} H_{r_i}(I_1^{(t_i)}(\frac{\varphi_i}{\|\varphi_i\|_{t_i}})), 1 \le i \le m,$$

and

$$\{I_1^{(t_i)}(\varphi) : \varphi \in L_2(I\!R, B(I\!R), P_{t_i})\}, 1 \le i \le m,$$

are independent Gaussian processes on some probability space (with expectation E_0) such that

$$E_0 I_1^{(t_i)}(\varphi) I_1^{(t_j)}(\psi) = \delta_{ij} E_{t_i}(\varphi\psi), \forall 1 \le i,j \le m$$

Denote, as before,

$$v_\alpha^{(ni)} = (f_1(t_i) \int_{I\!R} K^2(u)du)^{-1/2} \|\varphi_i\|_{t_i}^{-1} n^{-1/2} \sqrt{a_n} \varphi_i^n(\mathbf{Z}_\alpha, t_i), 1 \le \alpha \le n,$$

for $1 \le i \le m$. Then, as in (3.3.10),

$$(na_n)^{r/2} U_n(\varphi_1^{(r_i)}, \ldots, \varphi_m^{(r_m)}, \mathbf{t}^0)$$
$$= b_n \prod_{i=1}^{m} (f_1(t_i) \int_{I\!R} K^2(u)du)^{r_i/2} (\|\varphi_i\|_{t_i})^{r_i}$$
$$\sum_{1 \le \alpha(1) \ne \ldots \ne \alpha(r) \le n} v_{\alpha(1)}^{(n1)} \ldots v_{\alpha(r_1)}^{(n1)} \ldots v_{\alpha(r_1+\ldots+r_{m-1}+1)}^{(nm)} \ldots v_{\alpha(r)}^{(nm)}, \tag{3.3.28}$$

where $b_n \to 1$ as $n \to \infty$. Further, as in the proof of Lemma 3.3.2, we adapt the method of Schuster (1972) (Lemma 1), and obtain that

$$(\sum_{\alpha=1}^{n} v_\alpha^{(n1)}, \ldots, \sum_{\alpha=1}^{n} v_\alpha^{(nm)})'$$

$$\xrightarrow{d} (I_1^{(t_1)}(\frac{\varphi_1}{\|\varphi_1\|_{t_1}}), \dots, I_1^{(t_m)}(\frac{\varphi_m}{\|\varphi_m\|_{t_m}}))'$$

as $n \to \infty$, since $\varphi_i(\cdot), 1 \le i \le m$, are bounded and, for any scalars $c_i, 1 \le i \le m$, we have

$$E(\sum_{i=1}^{m} c_i \sum_{\alpha=1}^{n} v_\alpha^{(ni)})^2 = \sum_{i=1}^{m} c_i^2 E(\sum_{\alpha=1}^{n} v_\alpha^{(ni)})^2 + \sum_{i \ne j} c_i c_j E(\sum_{\alpha=1}^{n} v_\alpha^{(ni)})(\sum_{\alpha=1}^{n} v_\alpha^{(nj)}),$$

and

$$E(\sum_{\alpha=1}^{n} v_\alpha^{(ni)})^2 = (f_1(t_i) \int_{\mathbb{R}} K^2(u)du)^{-1} \|\varphi_i\|_{t_i}^{-2} E a_n [\varphi_i(Y_1) a_n^{-1} K_n(t_i, X_1) - E\varphi_i(Y_1) a_n^{-1} K_n(t_i, X_1)]^2,$$

whereas

$$E(\sum_{\alpha=1}^{n} v_\alpha^{(ni)})(\sum_{\alpha=1}^{n} v_\alpha^{(nj)}),\ i \ne j$$

$$= \text{constt.} E\{a_n[\varphi_i(Y_1) a_n^{-1} K_n(t_i, X_1) - E\varphi_i(Y_1) a_n^{-1} K_n(t_i, X_1)]$$

$$[\varphi_j(Y_1) a_n^{-1} K_n(t_j, X_1) - E\varphi_j(Y_1) a_n^{-1} K_n(t_j, X_1)]\}.$$

Thus, by the methods of proofs of Lemma 3.3.1 (ii) and 3.3.4 (ii),

$$\lim_{n\to\infty} E(\sum_{\alpha=1}^{n} v_\alpha^{(ni)})^2 = 1, \lim_{n\to\infty} E(\sum_{\alpha=1}^{n} v_\alpha^{(ni)})(\sum_{\alpha=1}^{n} v_\alpha^{(nj)}) = 0.$$

for all $i, j, \dots, i \ne j$. This establishes the asymptotic normality claimed above. Consequently, as $n \to \infty$,

$$\prod_{i=1}^{m} H_{r_i}(\sum_{\alpha=1}^{n} v_\alpha^{(ni)}) \xrightarrow{d} \prod_{i=1}^{m} H_{r_i}(I_1^{(t_i)}(\frac{\varphi_i}{\|\varphi_i\|_{t_i}}))$$

$$= (\prod_{i=1}^{m} \frac{\sqrt{r_i!}}{\|\varphi_i\|_{t_i}^{r_i}}) I_{r_1,\dots,r_m}^{t_1,\dots,t_m}(\varphi_1^{\otimes r_1} \otimes \dots \otimes \varphi_m^{\otimes r_m}).$$

Hence, in view of (3.3.28), it is enough to show that, as $n \to \infty$,

$$\sum_{1 \le \alpha(1) \ne \dots \ne \alpha(r) \le n} v_{\alpha(1)}^{(n1)} \dots v_{\alpha(r_1)}^{(n1)} \dots v_{\alpha(r_1+\dots+r_{m-1}+1)}^{(nm)} \dots v_{\alpha(r)}^{(nm)}$$

$$\to H_{r_1}(\sum_{\alpha=1}^{n} v_\alpha^{(n1)}) \dots H_{r_m}(\sum_{\alpha=1}^{n} v_\alpha^{(nm)}) \tag{3.3.29}$$

in probability. Now, as in Lemma 3.3.2, we get

$$\left.\begin{array}{rcl}\lim_{n\to\infty}\sum_{\alpha=1}^{n}(v_\alpha^{(ni)})^2 & = & 1\\ \lim_{n\to\infty}\sum_{\alpha=1}^{n}v_\alpha^{(ni)}v_\alpha^{(nj)} & = & 0\\ \lim_{n\to\infty}\max_{1\le\alpha\le n}|v_\alpha^{(ni)}| & = & 0\end{array}\right\}\text{in probability}, 1\le i\ne j\le m.$$

Hence, as shown in the proof of the Theorem of Rubin and Vitale (1980), using the identity in the Appendix of the same paper, (3.3.29) follows.□

Now we come to the end of Subsection 3.3.2 with the following theorem which is the counterpart of Theorem 3.3.1.

Theorem 3.3.2

Let $h : \mathbb{R}^k \to \mathbb{R}$ be symmetric and the assumptions A1′ - A5′ hold. Let

$$r_0 := \inf\{1 \le r \le k : \exists (r_1, \ldots, r_m) \in \mathrm{r}(k,m) \text{ such that}$$

$$m_{r_1,\ldots,r_m}(y_1,\ldots,y_r) \not\equiv 0\}$$

and

$$\mathbf{R}_0(k,m) := \{(r_1,\ldots,r_m) \in \mathrm{r}_0(k,m) : m_{r_1,\ldots,r_m}(\cdot) \not\equiv 0\}.$$

Then

$$(na_n)^{r_0/2}\{U_n(h,\mathrm{t}^0) - \sum_{j=0}^{r_0-1}\sum_{\mathrm{j}(k,m)}\binom{k_1}{j_1}\cdots\binom{k_m}{j_m}U_n(h_n^{(j_1,\ldots,j_m)},\mathrm{t}^0)\}$$
$$\xrightarrow{d} \sum_{\mathbf{R}_0(k,m)}\binom{k_1}{r_1}\cdots\binom{k_m}{r_m}(\prod_{i=1}^{m}\sqrt{r_i!}(f_1(t_i)\int_{\mathbb{R}}K^2(u)du)^{r_i/2}(f_1(t_i))^{k_i-r_i})$$
$$I_{r_1,\ldots,r_m}^{t_1,\ldots,t_m}(m_{r_1,\ldots,r_m}(\cdot))\,, \qquad (3.3.30)$$

as $n\to\infty$.

Proof: Denote

$$\mathrm{r}_m := (r_1,\ldots,r_m)$$
$$C(K,f_1,\mathrm{t}^0,\mathrm{r}_m) := \prod_{i=1}^{m}\sqrt{r_i!}(f_1(t_i)\int_{\mathbb{R}}K^2(u)du)^{r_i/2}(f_1(t_i))^{k_i-r_i}$$

Now as in the proof of Lemma 3.3.3, we can get, for given $\varepsilon > 0$, simple functions $\varphi_{i,p,\mathbf{r}_m}(\cdot)$, and $a_{p,\mathbf{r}_m} \in I\!R, 1 \le i \le m, 1 \le p \le l(\mathbf{r}_m), \mathbf{r}_m \in \mathbf{R}_0(k,m)$, such that

$$\sum_{\mathbf{r}_m \in \mathbf{R}_0(k,m)} \prod_{i=1}^{m} \binom{k_i}{r_i}^2 C^2(K, f_1, \mathbf{t}^0, \mathbf{r}_m) E_{t_1,\dots,t_m}\{m_{r_1,\dots,r_m}(Y_1,\dots,Y_r) - \sum_{p=1}^{l(\mathbf{r}_m)} a_{p,\mathbf{r}_m}\varphi_{1,p,\mathbf{r}_m}(Y_1)\cdots\varphi_{1,p,\mathbf{r}_m}(Y_{r_1})\cdots\varphi_{m,p,\mathbf{r}_m}(Y_r)\}^2 < \varepsilon \tag{3.3.31}$$

Further, define $U_n(\varphi_{1,p,\mathbf{r}_m}^{(r_1)},\dots,\varphi_{m,p,\mathbf{r}_m}^{(r_m)},\mathbf{t}^0)$ as in Lemma 3.3.5. Denote $G(\mathbf{t}^0,\mathbf{r}_m) = \prod_{i=1}^m (f_1(t_i))^{k_i - r_i}$. Then, as in the proof of Lemma 3.3.5, one can show that

$$\begin{aligned}
&\lim_{n\to\infty} (na_n)^{r_0} E(\sum_{\mathbf{r}_m \in \mathbf{R}_0(k,m)} \prod_{i=1}^{m}\binom{k_i}{r_i}\{U_n(h_n^{(r_1,\dots,r_m)},\mathbf{t}^0) \\
&-G(\mathbf{t}^0,\mathbf{r}_m)\sum_p a_{p,\mathbf{r}_m} U_n(\varphi_{1,p,\mathbf{r}_m}^{(r_1)},\dots,\varphi_{m,p,\mathbf{r}_m}^{(r_m)},\mathbf{t}^0)\})^2 \\
=&\ \text{left-hand side of (3.3.31)}
\end{aligned}$$

Now, by Lemma 3.3.6,

$$\sum_{\mathbf{R}_0(k,m)} G(\mathbf{t}^0,\mathbf{r}_m) \sum_{p=1}^{l(\mathbf{r}_m)} a_{p,\mathbf{r}_m}(na_n)^{r_0/2} U_n(\varphi_{1,p,\mathbf{r}_m}^{(r_1)},\dots,\varphi_{m,p,\mathbf{r}_m}^{(r_m)},\mathbf{t}^0)$$

$$\xrightarrow{d} \sum_{\mathbf{R}_0(k,m)} \sum^{l(\mathbf{r}_m)} a_{p,\mathbf{r}_m} C(K,f_1,\mathbf{t}^0,\mathbf{r}_m) I_{r_1,\dots,r_m}^{t_1,\dots,t_m}(\varphi_{1,p,\mathbf{r}_m}^{\otimes r_1}\otimes,\dots,\otimes\varphi_{m,p,\mathbf{r}_m}^{\otimes r_m}).$$

as $n\to\infty$. Here (3.3.30) follows from the two limit results above, along the lines of Lemma 3.3.3, since

$$\begin{aligned}
&(na_n)^{r_0}\sum_{j=r_0+1}^{k} E[\sum_{\mathbf{j}(k,m)}\binom{k_1}{j_1}\cdots\binom{k_m}{j_m}U_n(h_n^{(\mathbf{r}_m)},\mathbf{t}^0)]^2 \\
=&\ \sum_{j=r_0+1}^{k} O((na_n)^{-(j-r_0)}) \text{ (by (3.3.27))} \\
&\to 0
\end{aligned}$$

as $n\to\infty$, exactly as in Theorem 3.3.1.□

3.4 Limit laws of $U_n^h(\mathbf{t})$

In this section we discuss the limit distributions of $U_n^h(\mathbf{t}) = U_n(h,\mathbf{t})/\mathbf{U_n}(\mathbf{1},\mathbf{t})$. This section is fairly technical in nature, in that some more analytical assumptions will be imposed on $m(\cdot)$ and $K(\cdot)$, in addition to those in A1 - A5 (or A1′ - A3′, A4, A5). Further, we shall retain the results and notation of Subsection 3.3.2 only, because this subsection contains the results of Subsection 3.3.1.

Let us consider the case of asymptotic normality first. This is equivalent, in the notation of Theorem 3.3.2, to $r_0 = 1$. Hence $r_i = 0$ or $1, 1 \le i \le m$. Denote

$$I_{0,\dots,1,\dots,0}^{t_1,\dots,t_m}(m_{0,\dots,1,\dots,0}(\cdot)) := I_1^{t_i}(m_i(\cdot)), 1 \le i \le m,$$

where, on the left hand side, the '1' in the subscript occurs at the i-th place. Then

$$\mathbf{R}_0(k,m) = \{1 \le i \le m : m_i(y_1) \not\equiv 0\},$$

and we have

$$(na_n)^{1/2}(U_n(h,\mathbf{t}^0) - EU_n(h,\mathbf{t}^0))$$
$$\xrightarrow{d} \sum_{i:m_i(\cdot)\not\equiv 0} k_i(f_1(t_i)\int_{I\!R} K^2(u)du)^{1/2}(f_1(t_i))^{k_i-1}(\prod_{j\ne i}(f_1(t_j))^{k_j})I_1^{t_i}(m_i(\cdot)). \quad (3.4.1)$$

Further, consider $U_n(1,\mathbf{t}^0)$. Here $h \equiv 1$ and

$$\begin{aligned} EU_n(1,\mathbf{t}^0) &= a_n^{-k}\prod_{j=1}^{k} EK((t_j^0 - X_1)/a_n) \\ &= \prod_{i=1}^{m} a_n^{-k_i}\{EK((t_i - X_1)/a_n)\}^{k_i}. \end{aligned}$$

Hence,

$$(na_n)^{1/2}(U_n(a,\mathbf{t}^0) - a_n^{-k}\prod_{j=1}^{k} EK((t_j^0 - X_1)/a_n))$$
$$\xrightarrow{d} \sum_{i=1}^{n} k_i(f_1(t_i)\int_{I\!R} K^2(u)du)^{1/2}(f_1(t_i))^{k_i-1}(\prod_{j\ne i}(f_1(t_j)^{k_j})I_1^{t_i}(1)), \quad (3.4.2)$$

as $n\to\infty$. Note that $\{I_1^{t_i}(m_i(\cdot)), I_1^{t_i}(1)\}, 1 \le i \le m$, are m independent random vectors, by their definitions.

Now we modify the previous assumptions as follows:

B1. Same as A1′.

B2. $f_1(\cdot)$ is continuous at t_i and $f_1(t_i) > 0, 1 \leq i \leq m$; further $f_1(x)$ is twice continuously differentiable at each $t_i, 1 \leq i \leq m$.

B3. Same as A3′.

B4. $K(\cdot) \geq 0, \int_{\mathbb{R}} K(u)du = 1, \lim_{|u|\to\infty} |uK(u)| = 0$ and $\sup_{u\in\mathbb{R}} K(u) < \infty$; further $K(\cdot)$ is symmetric and $\int_{\mathbb{R}} u^2 K(u)du < \infty$.

B5. $a_n \to 0, na_n^3 \to \infty$, but $na_n^5 \to 0$ as $n \to \infty$.

Moreover, we impose the new assumption:

B6. $m(\mathbf{x}) = E\{h(Y_1, \ldots, Y_k)|X_1 = x_1, \ldots, X_k = x_k\}$ is twice continuously differentiable in a neighbourhood of $\mathbf{t}^0$.

Then we have the following result:

Theorem 3.4.1 *(i) Under A1′ - A3′, A4 and A5,*

$$(na_n)^{1/2}(U_n^h(\mathbf{t}^0) - EU_n(h, \mathbf{t}^0)/EU_n(1, \mathbf{t}^0))$$

$$\xrightarrow{d} \sum_{i=1}^{m} k_i (f_1(t_i))^{-1/2} \left(\int_{\mathbb{R}} K^2(u)du\right)^{1/2} I_1^{t_i}(m_i(\cdot) - m(\mathbf{t}^0)),$$

as $n \to \infty$.

(ii) Under B1 - B6

$$(na_n)^{1/2}(U_n^h(\mathbf{t}^0) - m(\mathbf{t}^0)) \xrightarrow{d} Z(h, \mathbf{t}^0) \text{ as } n \to \infty,$$

where $Z(h, \mathbf{t}^0)$ is the limiting random variable given in (i) above.

Proof: (i) This follows from (3.4.1) and (3.4.2), by considering the function $(u, v) \mapsto uv^{-1}$, as shown in the proof of Theorem 1 of Stute (1991), and the fact that $EU_n(h, \mathbf{t}^0) \to m(\mathbf{t}^0) \prod_{j=1}^k f_1(t_j^0)$ and $EU_n(1, \mathbf{t}^0) \to \prod_{j=1}^k f_1(t_j^0)$, as A1′ - A3′, A4 and A5.

(ii) Note that

$$
\begin{aligned}
& (na_n)^{1/2}(EU_n(h,\mathbf{t}^0)/EU_n(1,\mathbf{t}^0) - m(\mathbf{t}^0)) \\
= \;& \{EU_n(1,\mathbf{t}^0)\}^{-1}\left\{(na_n)^{1/2}(EU_n(h,\mathbf{t}^0) - m(\mathbf{t}^0)\prod_{j=1}^{k} f_1(t_j^0))\right. \\
& \left. -m(\mathbf{t}^0)(na_n)^{1/2}(EU_n(1,\mathbf{t}^0) - \prod_{j=1}^{k} f_1(t_j^0))\right\} \\
= \;& o(1) + o(1),
\end{aligned}
$$

as $n \to \infty$, as in Corollary 2.4 of Stute (1991), using B2,B6, Taylor's expansion for $m(\mathbf{x})\prod_{j=1}^{k} f_1(x_j)$ and $\prod_{j=1}^{k} f_1(x_j)$ upto order 2, B4 and B5. The result now follows from (i) above.□

In case $r_0 \geq 2$, (i.e., the so-called 'degenerate' case), we get the following result directly:

Theorem 3.4.2

Let A1' - A3, A4 and A5 hold. If

$$\sum_{n=1}^{\infty} n^{-3/2} a_n^{-2} < \infty\,, \tag{3.4.3}$$

then

$$
\begin{aligned}
& \{U_n(1,\mathbf{t}^0)\}^{-1}(na_n)^{r_0/2}\{U_n(h,\mathbf{t}^0) - \sum_{j=0}^{r_0-1}\sum_{\mathbf{j}(k,m)} \binom{k_1}{j_1}\cdots\binom{k_m}{j_m} U_n(h_n^{(j_1,\ldots,j_m)},\mathbf{t}^0)\} \\
& \xrightarrow{d} \sum_{\mathbf{R}_0(k,m)} \binom{k_1}{r_1}\cdots\binom{k_m}{r_m}\left(\prod_{i=1}^{m}\sqrt{r_i!}\left(\int_{\mathbb{R}} K^2(u)du/f_1(t_i)\right)^{r_i/2}\right) I^{t_1,\ldots,t_m}_{r_1,\ldots,r_m}(m_{r_1,\ldots,r_m}(\cdot))
\end{aligned}
\tag{3.4.4}
$$

as $n \to \infty$.

Proof: The result is immediate from Theorem 3.3.2, provided we can show that

$$U_n(1,\mathbf{t}^0) \to \prod_{j=1}^{k} f_1(t_j^0) \text{a.s.} \tag{3.4.5}$$

as $n \to \infty$, since $\prod_{j=1}^{k} f_1(t_j^0) = \prod_{i=1}^{m}(f_1(t_i))^{k_i}$. Now we have

$$U_n(1,\mathbf{t}^0) - \prod_{j=1}^{k} f_1(t_j^0) = U_n(1,\mathbf{t}) - EU_n(1,\mathbf{t}^0) + EU_n(1,\mathbf{t}^0) - \prod_{j=1}^{k} f_1(t_j^0).$$

By A2′, A4, A5,

$$\lim_{n\to\infty} EU_n(1, \mathbf{t}^0) = \prod_{j=1}^{k} f_1(t_j^0).$$

By the 'D-M decomposition' (cf. (3.3.23)),

$$U_n(1,\mathbf{t}^0) - EU_n(1,\mathbf{t}^0)$$

$$= \sum_{i=1}^{m} k_i a_n^{-(k_i-1)}\{EK_n(t_i,X_1)\}^{k_i-1}\prod_{j\neq i} a_n^{-k_j}\{EK_n(t_j,X_1)\}^{k_j}\cdot$$

$$\left(\frac{1}{n}\sum_{l=1}^{n}[a_n^{-1}K_n(t_i,X_l)]\right) + (\text{remainder})$$

where $E(\text{remainder})^2$ is $O((na_n)^{-2})$, by (3.3.27), and $\forall 1 \le i \le m$,

$$E\left|\frac{1}{n}\sum_{l=1}^{n}[a_n^{-1}K_n(t_i,X_l) - a_n^{-1}EK_n(t_i,X_1)]\right|^3 = O(n^{-3/2}a_n^{-2})$$

by the Marcinkiewicz-Zygmund inequality, as in the proof of Theorem 3 in Stute (1991). Thus by Borel-Cantelli Lemma, and (3.4.3),

$$U_n(1,\mathbf{t}^0) - EU_n(1,\mathbf{t}^0) \to 0 \text{ a.s.},$$

as $n\to\infty$, and (3.4.5) follows.□

Note that the sequence of weakly convergent random variables in (3.4.4) involve the U-statistics $U_n(h_n^{(j_1,\ldots,j_m)},\mathbf{t}^0)$ which depend on the unknown distribution of (X_1,Y_1), hence cannot be calculated from the data $(X_i,Y_i), 1\le i\le n$. Ideally, one could hope either to estimate then from data, or to show that

$$(na_n)^{r_0/2}U_n(h_n^{(j_1,\ldots,j_m)},\mathbf{t}^0) = o_p(1),\ 0\le j\le r_0-1, \tag{3.4.6}$$

as $n\to\infty$.

We shall now show that (3.4.6) holds when $h(\cdot)$ has a product structure, under the assumptions B1 - B6. *The general case remains as yet unsettled.*

Let $h_\varphi(y_1,\ldots,y_k) = \varphi(y_1)\ldots\varphi(y_k)$, where $\varphi(\cdot)$ is such that $h_\varphi(\cdot)$ satisfies B3. Denote $m_\varphi(x) := E\{\varphi(Y_1)|X_1 = x\}$, and assume $m_\varphi(\cdot)$ satisfies B6. Now, suppose that

$$m_\varphi(t_i) = 0 \text{ for } 1\le i\le l, \tag{3.4.7}$$

for some $l, 1 \leq l \leq m$. This implies that, $\forall 0 \leq r \leq \sum_{i=1}^{l} k_i - 1$,

$$\begin{aligned} & m_{r_1,\ldots,r_m}(y_1,\ldots,y_r) \\ = \ & E\{h_\varphi(y_1,\ldots,y_r,Y_{r+1},\ldots,Y_k)|X_{r+j}=t_1, 1 \leq j \leq k_1 - r_1,\ldots, \\ & X_{r+j} = t_m, k-(k_m-r_m)+1 \leq j \leq k\} \\ = \ & \varphi(y_1)\ldots\varphi(y_r)\prod_{i=1}^{l}\{m_\varphi(t_i)\}^{k_i-r_i}\prod_{i=l+1}^{m}\{m_\varphi(t_i)\}^{k_i-r_i} \\ \equiv \ & 0, \end{aligned}$$

since $k_i - r_i > 0$ for at least one $1 \leq i \leq l$. (Otherwise, $k_i = r_i, 1 \leq i \leq l$, and $r = \sum_{i=1}^{m} r_i \geq \sum_{i=1}^{l} r_i = \sum_{i=1}^{l} k_i > \sum_{i=1}^{l} k_i - 1$). Hence, in this case, $r_0 = \sum_{i=1}^{l} k_i$. We are now ready to establish (3.4.6)

Theorem 3.4.3

(i) Let $\varphi(\cdot)$ be such that (3.4.7) holds. Then, under B1 - B6,

$$(na_n)^{r_0/2}U_n(h_{\varphi,n}^{(r_1,\ldots,r_m)},\mathbf{t}^0) = o_p(1), \forall\, 0 \leq r < r_0 = \sum_{i=1}^{l} k_i,$$

as $n \to \infty$, where $r = r_1 + \cdots + r_m$.

(ii) Under the conditions of (i), and if $\sum_{n\geq 1} n^{-3/2}a_n^{-2} < \infty$,

$$\begin{aligned} & (na_n)^{r_0/2}U_n^{h_\varphi}(\mathbf{t}^0) \\ \xrightarrow{d} & \sum_{\mathbf{R}_0(k,m)}\binom{k_1}{r_1}\cdots\binom{k_m}{r_m}\Big(\prod_{i=1}^{m}\sqrt{r_i!}\Big(\int_{\mathbb{R}}K^2(u)du/f_1(t_i)\Big)^{r_i/2}\bar{H}_{r_i}(\varphi)\Big) \\ & \prod_{i=l+1}^{m}\{m_\varphi(t_i)\}^{k_i} \end{aligned}$$

as $n \to \infty$, $\bar{H}_{r_i}(\varphi)$ as in Lemma 3.3.6.

Proof:

(i) Consider the case $r = 0$ first. Then

$$(na_n)^{r_0/2}EU_n(h_\varphi,\mathbf{t}^0)$$

$$= (na_n)^{r_0/2} \prod_{i=1}^{l} a_n^{-k_i} \{E\varphi(Y_1)K_n(t_i, X_1)\}^{k_i} \prod_{i=l+1}^{m} a_n^{-k_i} \{E\varphi K_n(t_i)\}^{k_i}$$

$$= \prod_{i=1}^{l} \{a_n^{-1}(na_n)^{1/2} E\varphi(Y_1)K_n(t_i, X_1)\}^{k_i} \prod_{i=l+1}^{m} \{a_n^{-1} E\varphi(Y_1)K_n(t_i, X_1)\}^{k_i}. \quad (3.4.8)$$

The second product is convergent, hence bounded, in n by our assumptions. The first product in (3.4.8) equals

$$\prod_{i=1}^{l} \{a_n^{-1}(na_n)^{1/2} \int_{\mathbb{R}} [m\varphi(t_i - a_n x_1) f_1(t_i - a_n x_1) - m\varphi(t_i) f_1(t_i)] K(x_1) dx_1\}^{k_i}$$
$$= o(1), \text{as} n \to \infty \quad (3.4.9)$$

by Taylor's expansion and our assumptions.

For $1 \le r \le r_0 - 1 = \sum_{i=1}^{l} k_i - 1$, we have, with $\mathbf{z}_i = (x_i, y_i), 1 \le i \le r$,

$$h_{\varphi,n}^{(r_1,\ldots,r_m)}(\mathbf{z}_1, \ldots, \mathbf{z}_r)$$
$$= \prod_{i=1}^{m} \prod_{\alpha=1}^{r_i} [\varphi(y_\alpha) a_n^{-1} K_n(t_i, x_\alpha) - a_n^{-1} E\varphi(Y_1) K_n(t_i, X_1)]$$
$$\prod_{i=1}^{m} [a_n^{-1} E\varphi(Y_1) K_n(t_i, X_1)]^{k_i - r_i} \quad (3.4.10)$$

Also.

$$E\{(na_n)^{r_0/2} U_n(h_{\varphi,n}^{(r_1,\ldots,r_m)}, \mathbf{t}^0)\}^2$$
$$= E\{(na_n)^{r/2} U_n((na_n)^{(r_0-r)/2}.h_{\varphi,n}^{(r_1,\ldots,r_m)}, \mathbf{t}^0)\}^2 \quad (3.4.11)$$

Now, $r_0 - r = \sum_{i=1}^{l} k_i - \sum_{i=1}^{m} r_i \le \sum_{i=1}^{l} k_i - \sum_{i=1}^{l} r_i$, hence we can get $0 \le \delta_i \le k_i - r_i, 1 \le i \le l$, such that $\sum_{i=1}^{l} \delta_i = r_0 - r$. Thus, from (3.4.10),

$$(na_n)^{(r_0-r)/2} h_{\varphi,n}^{(r_1,\ldots,r_m)}(\mathbf{z}_1, \ldots, \mathbf{z}_r)$$
$$= \prod_{i=1}^{m} \prod_{\alpha=1}^{r_i} [\varphi(y_\alpha) a_n^{-1} K_n(t_i, x_\alpha) - E\varphi(Y_1) a_n^{-1} K_n(t_i, X_1)].$$
$$\prod_{i=1}^{l} [(na_n)^{1/2} a_n^{-1} E\varphi(Y_1) K_n(t_i, X_1)]^{\delta_i} [\ldots]_1^{k_i - r_i - \delta_i} \prod_{i=l+1}^{m} [\ldots]_2$$
$$= (\prod_{i=1}^{m} \prod_{\alpha=1}^{r_i} [\varphi(y_\alpha) a_n^{-1} K_n(t_i, x_\alpha) - E\varphi(Y_1) a_n^{-1} K_n(t_i, X_1)]).o(1) \quad (3.4.12)$$

in the same way as in (3.4.9). Thus, by (3.4.11), (3.4.12) and (3.3.27),

$$E\{(na_n)^{r_0/2}U_n(h_{\varphi,n}^{(r_1,\ldots,r_m)},\mathbf{t}^0)\}^2 = O(1).o(1) = o(1),$$

for $\forall\, r_1 + \ldots + r_m = r$, $1 \le r \le r_0 - 1$. This completes the proof.

(ii) This follows immediately from Theorem 3.4.2 and (i) above.□

We close Section 3.4 here. In the next, concluding, section, we give an example to illustrate the 'degenerate' case and a few explanatory remarks.

3.5 An example and some remarks

We continue with Example 4.1 of Stute (1991), which was mentioned first in the Introduction. See the above paper for other examples of the *non-degenerate case* , i.e., where asymptotic normality holds.

Example 3.5.1

Let $k = 2$ and $h(y_1, y_2) = y_1y_2$. Then $m(t_1^0, t_2^0) = \bar{m}(t_1^0)\bar{m}(t_2^0)$, where $\bar{m}(x) = E\{Y_1|X_1 = x\}$. Suppose our conditions B1 - B6 are satisfied.

Consider, first, the case $t_1^0 = t_2^0 = t$, say, and $\bar{m}(t) = 0$. Here, $m = 1$, $k_1 = k = 2$, $l = 1$ (in the notation of Theorem 3.4.3). Further $r_0 = 2$. Thus we have, by Theorem 3.4.3.

$$(na_n)U_n^h(t.\mathbf{1}) \xrightarrow{d} (\int_{\mathbb{R}} K^2(u)du/f_1(t))E\{Y_1^2|X_1 = t\}(Z^2 - 1),$$

as $n\to\infty$, where Z is a standard normal random variable. Next, consider the case $t_1^0 \ne t_2^0$, but $\bar{m}(t_1^0) = \bar{m}(t_2^0) = 0$. Here, $m = 2, k_1 = k_2 = 1, l = 2$ and $r_0 = 2$. Again, by Theorem 3.4.3, as $n\to\infty$.

$$(na_n)U_n^h(\mathbf{t}^0) \xrightarrow{d} (\int_{\mathbb{R}} K^2(u)du)(f_1(t_1^0)f_1(t_2^0))^{-1/2}\prod_{i=1}^{2}\{E(Y_1^2|X_1 = t_i^0)\}^{1/2}Z_1Z_2,$$

where $Z_i \sim N(0,1)$ are independent, $i = 1, 2$.

If however, $\bar{m}(t_1^0) = 0$, but $\bar{m}(t_2^0) \neq 0$. then

$$\left.\begin{array}{rcl} m_1(y_1) & = & y_1\bar{m}(t_2^0) \not\equiv 0 \\ m_2(y_1) & = & y_1\bar{m}(t_1^0) \equiv 0 \end{array}\right\} \Rightarrow m(t_1^0, t_2^0) = 0$$

in the notation of Theorem 3.4.1. Hence by the same theorem, as $n \rightarrow \infty$,

$$(na_n)^{1/2} U_n^h(\mathbf{t}^0) \xrightarrow{d} \left(\int_{\mathbb{R}} K^2(u)du / f_1(t_1^0)\right)^{1/2} I_1^{t_1^0}(m_1(\cdot)),$$

where of course, $I_1^{t_1^0}(m_1(\cdot)) \sim N(0, E\{Y_1^2|X_1 = t_1^0\}(\bar{m}(t_2^0))^2)$.

Remark 3.5.1

Note that the assumptions A1′ and A3′ together imply that $m(\mathbf{x})$ is continuous at $\mathbf{x} = \mathbf{t}^0$. This fact has been used implicitly throughout the paper, while proving

$$\lim_{n\rightarrow\infty} EU_n(h, \mathbf{t}^0) = m(\mathbf{t}^0) \prod_{j=1}^{k} f_1(t_j^0),$$

via Bochner's Theorem. Here, $(t_1^0, \ldots, t_k^0) = \mathbf{t}^0$.

Remark 3.5.2

In retrospect, it may be worthwhile to compare our conditions with those of Stute (1991). For example, our A1′ and A3′ imply Stute's Conditions (iii), (v) and (vii), as is clear from the proof of Lemma 3.3.1 (ii). Further, we do not need the functions $m_{jlm}(\cdot;\cdot;\cdot)$ in Stute's Condition (vi), because our method is different.

Remark 3.5.3

In treating the 'degenerate' case in Theorem 3.4.3, we should have considered product functions of the form

$$\prod_{j=1}^{p_1} \varphi^{(1)}(y_j) \ldots \prod_{j=p_1+\ldots+p_{r-1}+1}^{k} \varphi^{(r)}(y_j), \ \sum_{j=1}^{r} p_j = k, r \geq 1,$$

for a greater generality. But then we would have to deal with the $r \times m$ quantities

$$E\{\varphi^{(i)}(Y_1)|X_1 = t_j\}, 1 \leq i \leq r, 1 \leq j \leq m.$$

This would mean more complicated expressions, while the result would essentially be the same as in Theorem 3.4.3.

Chapter 4

Censored Data and the 'Martingale Methods'

4.1 Introduction

Let, as before, $\{(X_i, Y_i)\}_{i\geq 1}$ be a bi-variate i.i.d sequence. Further, suppose X_i and Y_i are both non-negative (i.e., they represent life-time or survival data), $i \geq 1$, and subject to *right random censoring*. That means there exist two other i.i.d., non-negative, sequences $\{X_i'\}_{i\geq 1}$ and $\{Y_i'\}_{i\geq 1}$, independent of each other and of $\{(X_i, Y_i)\}_{i\geq 1}$, such that we can only observe

$$\left.\begin{array}{ll} \tilde{X}_i = \min\{X_i, X_i'\}, & \delta_i = \mathbf{1}(X_i \leq X_i') \\ \tilde{Y}_i = \min\{Y_i, Y_i'\}, & \eta_i = \mathbf{1}(Y_i \leq Y_i') \end{array}\right\} \tag{4.1.1}$$

Let X_i, Y_i have joint density $f(x, y)$, marginal densities $f_1(x)$ and $f_2(y)$ and marginal distribution functions $F_1(x)$ and $F_2(y)$ respectively, $i \geq 1$. Let X_i' and Y_i' have distribution functions $G_1(x)$ and $G_2(y)$ respectively, $i \geq 1$.

Assume that $E|Y_1| < \infty$. The problem of estimating the regression function

$$m(t) := E(Y_1 \mid X_1 = t), \ t \geq 0, \tag{4.1.2}$$

was studied by Mielniczuk (1991) under the censoring scheme described above. He introduced a class of *kernel* and *nearest-neighbour* type estimators and established their pointwise consistency and asymptotic normality.

Under the same set-up, we propose a new class of estimators for (4.1.2), following a different approach. The idea is borrowed from Ramlau-Hansen (1983). It is well known (see Gill (1980), Theorem 3.1.1 and Corollary 3.1.1)that the counting process

$$N_n(t) := \sum_{i=1}^{n} \mathbf{1}(\tilde{X}_i \le t, \delta_i = 1), t \ge 0, \tag{4.1.3}$$

has the *compensator*

$$\Lambda_n(t) := \int_0^t V_n(u)\alpha_1(u)du,$$

i.e., $l_n(t) := N_n(t) - \Lambda_n(t)$ is a martingale, where $\alpha_1(\cdot) = f_1(\cdot)/[1 - F_1(\cdot)]$ and

$$V_n(t) := \sum_{i=1}^{n} \mathbf{1}(\tilde{X}_i \ge t).$$

In other words, $N_n(\cdot)$ comes under the *multiplicative intensity* model introduced by Aalen (1978). Ramlau-Hansen (1983) (to be called R-H hereafter) proposed the following kernel estimator for $\alpha_1(\cdot)$:

$$\hat{\alpha}_n(t) = a_n^{-1} \int_0^T K((t-u)/a_n)J_n(u)(V_n(u))^{-1}dN_n(u), \tag{4.1.4}$$

where $0 \le t < T$, $J_n(u) := \mathbf{1}(V_n(u) > 0)$, and it is assumed that $F_1(T) < 1$, $G_1(T) < 1$ and $G_1(\cdot)$ is continuous. We have observed that the following increasing process, with $\bar{G}_2 := 1 - G_2$,

$$P_n(t) := \sum_{i=1}^{n} \eta_i \tilde{Y}_i(\bar{G}_2(\tilde{Y}_i))^{-1}\mathbf{1}(\tilde{X}_i \le t, \delta_i = 1),\ t \ge 0, \tag{4.1.5}$$

has the compensator

$$A_n(t) := \int_0^t V_n(u)m(u)f_1(u)(1 - F_1(u))^{-1}du,\ t \ge 0. \tag{4.1.6}$$

This fact is established in Section 4.2. In other words, $A_n(t)$, too, has some sort of a 'multiplicative intensity' structure, where the 'intensity' is given by $V_n(\cdot)m(\cdot)\alpha_1(\cdot)$. Motivated by this fact, we propose the following estimator for $m(\cdot)$:

$$m_n^0(t) := \hat{\alpha}_n{}^0(t|m)/\hat{\alpha}_n(t),\ 0 \le t < T, \tag{4.1.7}$$

where $\hat{\alpha}_n(t)$ is given by (4.1.4),

$$\hat{\alpha}_n{}^0(t|m) := a_n^{-1} \int_0^T K((t-u)/a_n) J_n(u)(V_n(u))^{-1} dP_n^0(u),$$

and finally $P_n^0(t)$ is obtained from $P_n(t)$ by replacing $\bar{G}_2(\tilde{Y}_i)$ by $\bar{G}_{2n}(\tilde{Y}_i)$, the well-known Kaplan-Meier (K-M) product-limit estimator for $\bar{G}_2(\cdot)$. Both in (4.1.4) and (4.1.7) T is chosen large enough, but it is assumed that $T < \min\{T_{G_1}, T_{F_1}\}$, where, for any distribution function H,

$$T_H := \sup\{t \ : \ H(t) < 1\}.$$

However, to facilitate the use of 'martingale methods', we shall work with

$$\hat{\alpha}_n(t|m) := a_n^{-1} \int_0^T K((t-u)/a_n) J_n(u)(V_n(u))^{-1} dP_n(u), \tag{4.1.8}$$

and show in our proofs that the difference $|\hat{\alpha}_n{}^0(t|m) - \hat{\alpha}_n(t|m)|$ is negligible.

Exploiting the martingale structure, we establish *weak uniform consistency* (i.e..

$$\sup_{t \in C} |m_n^0(t) - m(t)| \rightarrow 0$$

in probability, as $n \rightarrow \infty$, where C is a compact subset of $[0, T]$) and asymptotic normality of our estimator, under suitable conditions. Regarding asymptotic normality, we obtain the same expression for the limiting variance as in the case of Mielniczuk's (1991) estimator, although our method is different. The results are given in Section 4.3. We have mostly adapted the techniques of R-H and Mielniczuk.

In Section 4.4, we consider the more general estimation problem (1.0.5). Our goal was to construct an analogue of $U_n(h, \mathbf{t})$ via the theory of two-parameter martingales. But, unfortunately, this approach does not seem to work here. Restricting our attention to the case $k = 2$ and functions $h(\cdot, \cdot)$ with a product structure, we show, in some detail, how far the said theory explains the structure of the analogues of $P_n(\cdot)$ and $A_n(\cdot)$ – and where it fails.

4.2 The compensator of $P_n(t)$ and related predictable variation processes

In this section, we show that $L_n(t) := P_n(t) - A_n(t)$ is a martingale with respect to a suitable filtration. We also compute the predictable variation process $< L_n, L_n > (t)$ and the process $< L_n, l_n > (t)$. See any standard text (e.g., Metivier (1982)) for the definition of such processes and the elementary results on martingales used in this article.

We shall deal with $P_n(t)$ first. Consider the following filtrations:

$$\mathcal{G}_i(t) := \sigma\{1(\eta_i = 1, \tilde{Y}_i \in B)1(\tilde{X}_i \leq s, \delta_i = 1), 1(\tilde{X}_i \leq s) : 0 \leq s \leq t, B \in \mathcal{B}(I\!R_+)\},\ 1 \leq i \leq n, 0 \leq t \leq T, \quad (4.2.1)$$

where $\mathcal{B}(I\!R_+)$ denotes the Borel σ-field on $I\!R_+$. Further, let

$$\mathcal{F}_n(t) := \bigvee_{i=1}^{n} \mathcal{G}_i(t),\ 0 \leq t \leq T.$$

We also make the assumption:

A1. $T_{F(\cdot|x)} \leq T_{G_2}\ \forall x \geq 0$, where $F(\cdot|x)$ denotes the conditional distribution function of Y_1, given $X_1 = x$.

Then we have the following result:

Lemma 4.2.1. *$P_n(t)$ is a sub-martingale with respect to $\mathcal{F}_n(t)$. It has compensator $A_n(t)$, i.e., $A_n(t)$ is the unique, increasing, predictable process such that*

$$L_n(t) = P_n(t) - A_n(t) \quad (4.2.2)$$

is an $\mathcal{F}_n(t)$-martingale, $0 \leq t \leq T$.

Proof: It is clear from (4.1.5) that $P_n(t)$ is adapted to $\mathcal{F}_n(t)$, $EP_n(t) \leq nEY_1 < \infty$, $0 \leq t \leq T$, and it has increasing paths. Hence obviously, $P_n(t)$ is a sub-martingale. Note that

$$A_n(t) = \sum_{i=1}^{n} \int_0^{t\wedge\tilde{X}_i} m(u)\alpha_1(u)du,\ 0 \leq t \leq T. \quad (4.2.3)$$

It is clear from (4.2.3) that $A_n(t)$ is increasing in t (since $m(\cdot)$ is non-negative), it is $\mathcal{F}_n(t)$-measurable,

$$EA_n(t) \le EA_n(T) \le nEY_1(1-F_1(T))^{-1} < \infty,$$

and the continuity of the maps

$$t \mapsto t \wedge \tilde{X}_i, \quad t \mapsto \int_0^t m(u)\alpha_1(u)du \tag{4.2.4}$$

implies that $A_n(t)$ has continuous paths, hence it is predictable. Thus it remains to show that

$$E(P_n(t+s) - P_n(t) \mid \mathcal{F}_n(t)) = E(A_n(t+s) - A_n(t) \mid \mathcal{F}_n(t)) \tag{4.2.5}$$

for all $0 \le t < T$ and $s \ge 0$. Fix $1 \le i \le n$. Put

$$\left.\begin{array}{ll} Q_i(t) & = \eta_i \tilde{Y}_i(\bar{G}_2(\tilde{Y}_i))^{-1}\mathbf{1}(\tilde{X}_i \le t, \delta_i = 1) \\ B_i(t) & = \int_0^t \mathbf{1}(\tilde{X}_i \ge u)m(u)\alpha_1(u)du. \end{array}\right\} \tag{4.2.6}$$

It suffices to show that

$$E(Q_i(t+s) - Q_i(t) \mid \mathcal{G}_i(t)) = E(B_i(t+s) - B_i(t) \mid \mathcal{G}_i(t)), \tag{4.2.7}$$

since $P_n(t) - A_n(t) = \sum_{i=1}^n (Q_i(t) - B_i(t))$ and $\mathcal{F}_n(t) = \bigvee_{i=1}^n \mathcal{G}_i(t)$.

Fix $0 \le t < T$ and $s \ge 0$. Denote

$$\mathbf{1}(\eta_i = 1, \tilde{Y}_i \in B)\mathbf{1}(\tilde{X}_i \le s, \delta_i = 1) := Z_i(B, s)$$

and $\mathbf{1}(\tilde{X}_i \le s) := W_i(s)$. Then $\mathcal{G}_i(t)$ is generated by functions of the form

$$\prod_{j=1}^{p}\prod_{l=1}^{m} \mathbf{1}(Z_i(B_l, s_j) = \epsilon_{jl}) \cdot \prod_{j=1}^{p} \mathbf{1}(W_i(s_j) = \epsilon_j), \tag{4.2.8}$$

where $p \ge 1$, $m \ge 1$, $0 \le s_1 < \cdots < s_p \le t$, $B_l \in \mathcal{B}(\mathbb{R}_+)$ and $\epsilon_{jl}, \epsilon_j$ take values 0 or 1, $1 \le j \le p$, $1 \le l \le m$. Now

$$\begin{aligned} & E(Q_i(t+s) - Q_i(t)).\mathbf{1}(Z_i(B_l, s_j) = \epsilon_{jl}, W_i(s_j) = \epsilon_j, 1 \le j \le p, 1 \le l \le m) \\ = \; & E\left[\eta_i \tilde{Y}_i(\bar{G}_2(\tilde{Y}_i))^{-1}\mathbf{1}(t < \tilde{X}_i \le t+s, \delta_i = 1)\times\right. \end{aligned}$$

$$\mathbf{1}(Z_i(B_l,s_j)=\epsilon_{jl},W_i(s_j)=\epsilon_j,1\le j\le p,\ 1\le l\le m)]$$

$$= \begin{cases} 0 \text{ if } \epsilon_j = 1 \text{ for some } 1\le j\le p \\ E\eta_i\tilde{Y}_i(\bar{G}_2(\tilde{Y}_i))^{-1}\mathbf{1}(t<\tilde{X}_i\le t+s,\delta_i=1)\mathbf{1}(W_i(s_p)=0) \text{ otherwise} \end{cases}$$

$$= \begin{cases} 0 \text{ if } \epsilon_j = 1 \text{ for some } 1\le j\le p \\ E\eta_i\tilde{Y}_i(\bar{G}_2(\tilde{Y}_i))^{-1}\mathbf{1}(t<\tilde{X}_i\le t+s,\delta_i=1) \text{ otherwise,} \end{cases} \tag{4.2.9}$$

since $\mathbf{1}(W_i(s_p)=0)=\mathbf{1}(\tilde{X}_i>s_p)$ and $s_p\le t$. Similarly,

$$E(B_i(t+s)-B_i(t)).\mathbf{1}(Z_i(B_l,s_j)=\epsilon_{jl},W_i(s_j)=\epsilon_j,1\le j\le p,\ 1\le l\le m)$$

$$= \begin{cases} 0 \text{ if } \epsilon_j = 1 \text{ for some } 1\le j\le p \\ E\mathbf{1}(\tilde{X}_i>s_p)\int_t^{t+s}\mathbf{1}(\tilde{X}_i\ge u)m(u)\alpha_1(u)du \text{ otherwise.} \end{cases} \tag{4.2.10}$$

By comparing (4.2.9) and (4.2.10), it can be seen that they are equal,since, by A1,

$$E(\eta_1\tilde{Y}_1(\bar{G}_2(\tilde{Y}_1))^{-1}\mid X_1=u)=m(u).$$

Hence (4.2.7) follows.□

We now compute the predictable variation process of the martingale $L_n(t)$. But in order to ensure square-integrability of $L_n(t)$, $0\le t\le T$, we impose the following, somewhat restrictive, assumption:

A2. $0<T_{F_2}<T_{G_2}\le\infty$.

A2 makes Y_1 essentially bounded. We then have the following result:

Lemma 4.2.2. *Under A1 and A2, $L_n(t)$, $0\le t\le T$, is a square-integrable $\{\mathcal{F}_n(t)\ :0\le t\le T\}$ martingale. The predictable variation process of $L_n(t)$ is given by* ,

$$<L_n,L_n>(t)=\int_0^t V_n(u)m_2(u)\alpha_1(u)du, \tag{4.2.11}$$

where $m_2(u):=E(Y_1^2(\bar{G}_2(Y_1))^{-1}\mid X_1=u)$.

Proof: Note that

$$L_n^2(t)\le 2(P_n^2(t)+A_n^2(t)),$$

and

$$
\begin{aligned}
EP_n^2(t) &\leq nE\left[\eta_1\tilde{Y}_1(\bar{G}_2(\tilde{Y}_1))^{-1}\mathbf{1}(\tilde{X}_1 \leq t, \delta = 1)\right]^2 \\
&= n.EY_1^2(\bar{G}_2(Y_1))^{-1}\mathbf{1}(\tilde{X}_1 \leq t, \delta = 1) \\
&\leq n\int_0^{T_{F_2}} y^2(\bar{G}_2(y))^{-1}f_2(y)dy \\
&\leq n(\bar{G}_2(T_{F_2}))^{-1}EY_1^2 < \infty, \\
EA_n^2(t) &\leq E\left(\int_0^t V_n(u)m_2(u)\alpha_1(u)du\right)\left(\int_0^t V_n(u)\alpha_1(u)du\right), \\
&\qquad \text{by Cauchy-Schwartz inequality and the fact that } (m(u))^2 \leq m_2(u), \\
&\leq n^2\left(\int_0^t m_2(u)\alpha_1(u)du\right)(-\log(1 - F_1(T))), \text{ since } V_n(u) \leq n, \\
&\leq n^2(1 - F_1(T))^{-1}(\bar{G}_2(T_{F_2}))^{-1}EY_1^2(-\log(1 - F_1(T))) < \infty,
\end{aligned}
$$

for $0 \leq t \leq T$. Thus $L_n(t)$ is square-integrable. Put $\bar{M}_i(t) = Q_i(t) - B_i(t), 0 \leq t \leq T, 1 \leq i \leq n$, where $Q_i(\cdot)$ and $B_i(\cdot)$ are as in the proof of Lemma 4.2.1. Then we have

$$
\begin{aligned}
&< L_n, L_n > (t) \\
&= < \sum_{i=1}^n \bar{M}_i, \sum_{i=1}^n \bar{M}_i > (t) \\
&= \sum_{i=1}^n < \bar{M}_i, \bar{M}_i > (t) + \sum_{1 \leq i \neq j \leq n} < \bar{M}_i, \bar{M}_j > (t) \qquad (4.2.12)
\end{aligned}
$$

Fix $1 \leq i, j \leq n, i \neq j$. Then, for $s \geq 0$,

$$
\begin{aligned}
&E(\bar{M}_i(t+s)\bar{M}_j(t+s) \mid \mathcal{F}_n(t)) \\
&= E(\bar{M}_i(t+s)\bar{M}_j(t+s) \mid \mathcal{G}_i(t), \mathcal{G}_j(t)), \\
&\qquad \text{by the independence of } \mathcal{G}_i(t), 1 \leq i \leq n, \\
&= E(\bar{M}_i(t+s) \mid \mathcal{G}_i(t))E(\bar{M}_j(t+s) \mid \mathcal{G}_j(t)) \\
&= \bar{M}_i(t)\bar{M}_j(t)
\end{aligned}
$$

It is clear from above that $< \bar{M}_i, \bar{M}_j > \equiv 0$ for all $i \neq j$. Now note that, for each i, $\bar{M}_i(\cdot)$ is a function of bounded variation, since both $Q_i(\cdot)$ and $B_i(\cdot)$ are increasing, and is right-continuous. Thus we can write

$$\bar{M}_i^2(t) = 2\int_0^t \bar{M}_i(s-)d\bar{M}_i(s) + \sum_{0\le s\le t} (\triangle \bar{M}_i(s))^2, \tag{4.2.13}$$

where $\triangle \bar{M}_i(s) = \bar{M}_i(s) - \bar{M}_i(s-)$. (See, for example, Liptser and Shiryayev (1978), Equation (18.72), Lemma 18.12, p.269.) Now $\triangle \bar{M}_i(s) = \triangle Q_i(s) - \triangle B_i(s) = \triangle Q_i(s)$, since $B_i(\cdot)$ is continuous. Thus

$$\begin{aligned} & \sum_{0\le s\le t} (\triangle \bar{M}_i(s))^2 \\ = \;& \eta_1^2 \tilde{Y}_1^2 (\bar{G}_2(\tilde{Y}_1))^{-2} \mathbf{1}(\tilde{X}_1 \le t, \delta = 1) \\ = \;& \bar{M}_{2i}(t) + B_{2i}(t), \end{aligned} \tag{4.2.14}$$

where

$$\begin{aligned} \bar{M}_{2i}(t) &= \eta_1^2 \tilde{Y}_1^2 (\bar{G}_2(\tilde{Y}_1))^{-2} \mathbf{1}(\tilde{X}_1 \le t, \delta = 1) - \int_0^t \mathbf{1}(\tilde{X}_i \ge u) m_2(u)\alpha_1(u)du \\ B_{2i}(t) &= \int_0^t \mathbf{1}(\tilde{X}_i \ge u) m_2(u)\alpha_1(u)du \end{aligned}$$

Using the arguments of Lemma 4.2.1, it is clear that $\bar{M}_{2i}(\cdot)$ is a martingale and $B_{2i}(\cdot)$ is a predictable, increasing process. We also use the fact that

$$E(\eta_1^2 \tilde{Y}_1^2 (\bar{G}_2(\tilde{Y}_1))^{-2} \mid X_1 = u) = m_2(u),$$

by A1. Combining (4.2.13) and (4.2.14),

$$M_i^2(t) = 2\int_0^t \bar{M}_i(s-)d\bar{M}_i(s) + \bar{M}_{2i}(t) + B_{2i}(t).$$

Note that the first two terms above are martingales. Hence, by the uniqueness of the Doob-Meyer decomposition, we have

$$< \bar{M}_i, \bar{M}_i >= B_{2i}.$$

The result now follows from (4.2.12).□

By virtue of Lemma 4.2.2 and Doob's Maximal Inequality (see Metivier (1982), Theorems 10.12 and 19.1), we have, for any $\mathcal{F}_n(t)$-predictable process $Z(t)$,

$$\begin{aligned} E\left\{\sup_{0\le t\le T}\left|\int_0^t Z(s)dL_n(s)\right|^2\right\} &\le 4E\left(\int_0^T Z(s)dL_n(s)\right)^2 \\ &= 4E\int_0^T Z^2(s)V_n(s)m_2(s)\alpha_1(s)ds. \end{aligned}$$

We shall make use of this fact in Theorems 4.3.1 and 4.3.2 of Section 4.3.

Consider also the simpler counting process martingale $l_n(t) = N_n(t) - \Lambda_n(t)$. Let us modify the filtrations $\mathcal{G}_i(t)$, $1 \le i \le n$, as follows:

$$\begin{aligned} \bar{\mathcal{G}}_i(t) &:= \sigma\left\{1(\eta_i = a, \tilde{Y}_i \in B)1(\tilde{X}_i \le s, \delta_i = 1), 1(\tilde{X}_i \le s) :\right. \\ &\qquad \left. 0 \le s \le t, B \in \mathcal{B}(\mathbb{R}_+), a = 0 \text{ or } 1\right\},\ 0 \le t \le T, \end{aligned}$$

and also

$$\bar{\mathcal{F}}_n(t) = \bigvee_{i=1}^n \bar{\mathcal{G}}_i(t),\ 0 \le t \le T.$$

Then it is easy to see that $l_n(t)$ and $L_n(t)$ are both $\bar{\mathcal{F}}_n(t)$-martingales. Write $l_n(t) = \sum_{i=1}^n \bar{n}_i(t)$, where

$$\left.\begin{aligned} \bar{n}_i(t) &:= q_i(t) - b_i(t), \\ q_i(t) &:= 1(\tilde{X}_i \le t, \delta_i = 1), \\ b_i(t) &:= \textstyle\int_0^t 1(\tilde{X}_i \ge u)\alpha_1(u)du, \end{aligned}\right\} \tag{4.2.15}$$

$1 \le i \le n$. Then the following lemma shows that $< L_n, l_n >= A_n$. We shall use this in Theorem 4.3.2.

Lemma 4.2.3. *Under A1 and A2, we have*

$$< L_n, l_n > (t) = A_n(t),\ 0 \le t \le T,$$

i.e., $L_n(t)l_n(t) - A_n(t)$ *is an* $\bar{\mathcal{F}}_n(t)$*-martingale.*

Proof: Note that

$$\begin{aligned} &< L_n, l_n > (t) \\ = &\sum_{i=1}^n < \bar{M}_i, \bar{n}_i > (t) + \sum_{i\ne j} < \bar{M}_i, \bar{n}_j > (t). \end{aligned} \tag{4.2.16}$$

Now by the same arguments as the ones used to show that $< \bar{M}_i, \bar{M}_j >\equiv 0, i \neq j$ (in the proof of Lemma 4.2.2), we get that $< \bar{M}_i, \bar{n}_j >\equiv 0$ for $i \neq j$. Next, observe that we can write

$$\begin{aligned}\bar{M}_i(t)\bar{n}_i(t) &= \int_0^t \bar{M}_i(s-)d\bar{n}_i(s) + \int_0^t \bar{n}_i(s-)d\bar{M}_i(s) \\ &\quad + \sum_{0\leq s\leq t} [\bar{M}_i(s) - \bar{M}_i(s-)][\bar{n}_i(s) - \bar{n}_i(s-)], \qquad (4.2.17)\end{aligned}$$

as in Liptser and Shiryayev (1978), Lemma 18.7, Equation (18.40). Further,

$$\begin{aligned}&\sum_{0\leq s\leq t} [\bar{M}_i(s) - \bar{M}_i(s-)][\bar{n}_i(s) - \bar{n}_i(s-)] \\ &= \eta_i(\bar{G}_2(\tilde{Y}_i))^{-1}\tilde{Y}_i\mathbf{1}(\tilde{X}_i \leq t, \delta_i) \\ &= Q_i(t) \qquad (4.2.18)\end{aligned}$$

The result now follows from (4.2.18), Lemma 4.2.1 (Equation (4.2.7)) and the fact that the first two terms on the right-hand side of (4.2.17) are martingales.□

We close Section 4.2 here.

4.3 Weak uniform consistency and asymptotic normality of $m_n^0(t)$

Recall, from (4.1.7), that

$$m_n^0(t) = \hat{\alpha}_n{}^0(t|m)/\hat{\alpha}_n(t),\ 0 \leq t \leq T,$$

where $\hat{\alpha}_n(t)$ is given by (4.1.4) and

$$\begin{aligned}\hat{\alpha}_n{}^0(t|m) &= a_n^{-1}\int_0^T K((t-u)/a_n)J_n(u)(V_n(u))^{-1}dP_n^0(u) \\ &= \sum_{j=1}^n \eta_j\tilde{Y}_j(\bar{G}_{2n}(\tilde{Y}_j))^{-1}a_n^{-1}K((t-\tilde{X}_j)/a_n)\delta_j/V_n(\tilde{X}_j),\end{aligned}$$

where $\bar{G}_{2n}(\cdot)$ is the well-known Kaplan-Meier estimator for $\bar{G}_2(\cdot)$:

$$\bar{G}_{2n}(u) = \begin{cases} \prod_{j=1}^n (1-(V_n(\tilde{Y}_{(j)}))^{-1})^{1-\eta_{(j)}}\mathbf{1}(\tilde{Y}_{(j)} \leq u), & \text{if } u < \tilde{Y}_{(n)}, \\ 0 & \text{otherwise,}\end{cases}$$

$\tilde{Y}_{(j)}$, $1 \leq j \leq n$, being the ordered $\tilde{Y}_j$'s and $\eta_{(j)}$'s the corresponding censoring indicators. For our proofs, however, we shall make use of

$$\hat{\alpha}_n(t|m) = a_n^{-1} \int_0^T K((t-u)/a_n) J_n(u)(V_n(u))^{-1} dP_n(u)$$

in view of the martingale results obtained in Section 4.2. Define

$$\hat{\alpha}_n^*(t|m) = a_n^{-1} \int_0^T K((t-u)/a_n) J_n(u) m(u) \alpha_1(u) du. \tag{4.3.1}$$

Note that, by Lemma 4.2.1, $Z_n(s) - Z_n^*(s)$, $0 \leq s \leq T$, is an $\mathcal{F}_n(s)$-martingale where

$$\left.\begin{array}{ll} Z_n(s) & = \int_0^s J_n(u)(V_n(u))^{-1} dP_n(u) \\ Z_n^*(s) & = \int_0^s J_n(u)(V_n(u))^{-1} dA_n(u) = \int_0^s J_n(u) m(u) \alpha_1(u) du \end{array}\right\} \tag{4.3.2}$$

Further,

$$\hat{\alpha}_n(t|m) - \hat{\alpha}_n^*(t|m) = a_n^{-1} \int_0^T K((t-u)/a_n) d(Z_n(u) - Z_n^*(u)),$$

and

$$\begin{aligned} & < Z_n - Z_n^*, Z_n - Z_n^* > (s) \\ = & < \int_0^{\cdot} J_n(u)(V_n(u))^{-1} dL_n(u), \int_0^{\cdot} J_n(u)(V_n(u))^{-1} dL_n(u) > (s) \\ = & \int_0^s J_n(u)(V_n(u))^{-2} V_n(u) m_2(u) \alpha_1(u) du \\ = & \int_0^s J_n(u)(V_n(u))^{-1} m_2(u) \alpha_1(u) du, \end{aligned} \tag{4.3.3}$$

by Lemma 4.2.2.

We are now ready to give our main results. We make the following assumption on the *kernel* $K(\cdot)$:

A3. $K(\cdot)$ is a bounded, continuous and symmetric density function on $[-1, 1]$. Further, it is of bounded variation with total variation $v(K) < \infty$.

Theorem 4.3.1. *Let A1 - A3 hold and* $na_n^2/\log n \to \infty$ *as* $n \to \infty$. *Suppose* $m(\cdot)$ *and* $\alpha_1(\cdot)$ *are continuous on* $[0, T]$, *and* $[t_0, t_1] \subseteq [0, T]$ *is such that*

$$\inf_{t_0 \leq t \leq t_1} f_1(t)(1 - G_1(t)) \geq c > 0. \tag{4.3.4}$$

Further, suppose there exists $\epsilon > 0$ such that

$$\Pr\{Y_1 < T_{F_2} - \epsilon \mid t_0 - r < X_1 < t_1 + r\} = 1 \tag{4.3.5}$$

for $r > 0$ small enough. Then

$$\sup_{t_0 \le t \le t_1} \left| m_n^0(t) - m(t) \right| \to 0, \text{ as } n \to \infty,$$

in probability.

Proof: First, consider the following decomposition:

$$\begin{aligned} & (m_n^0(t) - m(t)) \\ = \ & (\hat{\alpha}_n(t))^{-1}(\hat{\alpha}_n{}^0(t|m) - \hat{\alpha}_n(t|m)) + (\hat{\alpha}_n(t))^{-1}(\hat{\alpha}_n(t|m) - m(t)\alpha_1(t)) \\ & -m(t)(\hat{\alpha}_n(t))^{-1}(\hat{\alpha}_n(t) - \alpha_1(t)). \end{aligned} \tag{4.3.6}$$

Now,

$$\begin{aligned} & (\hat{\alpha}_n(t))^{-1}|\hat{\alpha}_n{}^0(t|m) - \hat{\alpha}_n(t|m)| \\ \le \ & \sum_{i=1}^{n} \eta_i \tilde{Y}_i |(\bar{G}_{2n}(\tilde{Y}_i))^{-1} - (\bar{G}_2(\tilde{Y}_i))^{-1}| c_{ni}(t) / (\sum_{i=1}^{n} c_{ni}(t)). \end{aligned} \tag{4.3.7}$$

where

$$c_{ni}(t) := a_n^{-1} K((t - \tilde{X}_i)/a_n)\delta_i(V_n(\tilde{X}_i))^{-1}.$$

Hence, by A2, the right-hand side of (4.3.7) is bounded above by

$$T_{F_2} \sup_{1 \le i \le n} D_n(Y_i) 1(|X_i - t| < a_n),$$

with $D_n(Y_i) := |\bar{G}_{2n}(Y_i) - \bar{G}_2(Y_i)|(\bar{G}_{2n}(Y_i)\bar{G}_2(Y_i))^{-1}$, because $K(\cdot)$ is supported on $[-1, 1]$. Hence,

$$\begin{aligned} & \sup_{t_0 \le t \le t_1} |(\hat{\alpha}_n(t))^{-1}(\hat{\alpha}_n{}^0(t|m) - \hat{\alpha}_n(t|m))| \\ \le \ & T_{F_2} \sup_{1 \le i \le n} D_n(Y_i) 1(t_0 - a_n < X_i < t_1 + a_n) \\ \le \ & O(n^{-1/2}(\log\log n)^{1/2}), \text{ a.s., for } n \text{ large enough,} \end{aligned} \tag{4.3.8}$$

using (4.3.5) and Corollary 1 of Földes and Rejtő (1981). Next, we have

$$\begin{aligned}\hat{\alpha}_n(t) &= \sum_{i=1}^{n} a_n^{-1} K((t-\tilde{X}_i)/a_n)\delta_i (V_n(\tilde{X}_i))^{-1}\mathbf{1}(\tilde{X}_i \le T) \\ &\ge \sum_{i=1}^{n} n^{-1} a_n^{-1} K((t-\tilde{X}_i)/a_n)\delta_i \mathbf{1}(\tilde{X}_i \le T) \text{ a.s.},\end{aligned}$$

since $V_n(s) \le n,\ 0 \le s \le T$, a.s. Now, since (4.3.4) holds and $na_n^2/\log n \to \infty$ as $n \to \infty$, it follows by (2.3.7) in the proof of Theorem 2.3.1 of Chapter 2 that

$$\inf_{t_0 \le t \le t_1} \hat{\alpha}_n(t) \ge c - o(1) \text{ a.s. as } n \to \infty. \tag{4.3.9}$$

Further,

$$\begin{aligned} &\sup_{t_0 \le t \le t_1} |\hat{\alpha}_n(t|m) - m(t)\alpha_1(t)| \\ \le\ &\sup_{t_0 \le t \le t_1} |\hat{\alpha}_n(t|m) - \hat{\alpha}_n^*(t|m)| + \sup_{t_0 \le t \le t_1} |\hat{\alpha}_n^*(t|m) - m(t)\alpha_1(t)| \end{aligned} \tag{4.3.10}$$

Now it is easy to see that

$$\left.\begin{aligned} &\sup\nolimits_{0\le s\le T} |J_n(s) - 1| \xrightarrow{P} 0 \text{ as } n \to \infty, \\ &\sup\nolimits_{n \ge 1} \sup\nolimits_{0 \le s \le T} E\{nJ_n(s)(V_n(s))^{-1}\} < \infty. \end{aligned}\right\} \tag{4.3.11}$$

(See R-H, Example 3.2.3). It follows that

$$\left.\begin{aligned} &E\left\{\sup\nolimits_{t_0\le t\le t_1} |\hat{\alpha}_n(t|m) - \hat{\alpha}_n^*(t|m)|^2\right\} \to 0 \\ &\sup\nolimits_{t_0 \le t \le t_1} |\hat{\alpha}_n^*(t|m) - m(t)\alpha_1(t)| \xrightarrow{P} 0 \end{aligned}\right\} \tag{4.3.12}$$

as $n \to \infty$, since by (4.3.11)

$$\begin{aligned} &\sup_{t_0 \le t \le t_1} |\hat{\alpha}_n^*(t|m) - m(t)\alpha_1(t)| \\ \le\ &\sup_{t_0 \le t \le t_1} \int_{-1}^{1} K(u)|m(t-a_nu)\alpha_1(t-a_nu) - m(t)\alpha_1(t)|du \\ &+ \sup_{t_0 \le t \le t_1} m(t)\alpha_1(t) \int_{-1}^{1} K(u)(1 - J_n(t-a_nu))du \\ \to\ &0, \text{ in probability as } n \to \infty, \end{aligned}$$

and

$$\begin{aligned}
& E\left\{\sup_{t_0\le t\le t_1} |\hat{\alpha}_n(t|m) - \hat{\alpha}_n{}^*(t|m)|^2\right\} \\
= \; & E\left\{\sup_{t_0\le t\le t_1} \left|a_n^{-1}\int_0^T (Z_n(u) - Z_n^*(u))dK((t-u)/a_n)\right|^2\right\} \\
\le \; & a_n^{-2}(v(K))^2 E\left\{\sup_{0\le s\le T} |Z_n(s) - Z_n^*(s)|^2\right\} \\
\le \; & a_n^{-2}(v(K))^2 4E < Z_n - Z_n^*, Z_n - Z_n^* > (T) \\
= \; & 4(na_n^2)^{-1}\int_0^T nE\{J_n(s)(V_n(s))^{-1}\}m_2(s)\alpha_1(s)ds \\
\to \; & 0 \text{ as } n\to\infty
\end{aligned}$$

by **A3**, integration by parts and Doob's Inequality.

Thus from (4.3.12) and (4.3.10),

$$\sup_{t_0\le t\le t_1} |\hat{\alpha}_n(t|m) - m(t)\alpha_1(t)| \xrightarrow{P} 0 \text{ as } n\to\infty. \tag{4.3.13}$$

By similar arguments,

$$\sup_{t_0\le t\le t_1} |\hat{\alpha}_n(t) - \alpha_1(t)| \xrightarrow{P} 0 \text{ as } n\to\infty. \tag{4.3.14}$$

The result now follows from (4.3.6), (4.3.8), (4.3.9), (4.3.13) and (4.3.14).□

Now we consider asymptotic normality of $m_n^0(t)$ at a fixed $t, 0\le t\le T$. Note that, since $K(\cdot)$ is continuous (hence $\bar{\mathcal{F}}_n(t)$-predictable), for any two real numbers c_1, c_2,

$$U_n(s|t, c_1, c_2) := \int_0^s a_n^{-1}K((t-u)/a_n)J_n(u)(V_n(u))^{-1}(c_1 dL_n(u) + c_2 dl_n(u))$$

is an $\bar{\mathcal{F}}_n(s)$-martingale. Of course, as is clear from (4.3.2),

$$\begin{aligned}
& U_n(T|t, c_1, c_2) \\
= \; & c_1(\hat{\alpha}_n(t|m) - \hat{\alpha}_n{}^*(t|m)) + c_2(\hat{\alpha}_n(t) - \hat{\alpha}_n{}^*(t)),
\end{aligned} \tag{4.3.15}$$

where

$$\begin{aligned}
& \hat{\alpha}_n{}^*(t) \\
= \; & \int_0^T a_n^{-1}K((t-u)/a_n)J_n(u)(V_n(u))^{-1}d\Lambda_n(u) \\
= \; & \int_0^T a_n^{-1}K((t-u)/a_n)J_n(u)\alpha_1(u)du.
\end{aligned}$$

In order to establish asymptotic normality, we need the following lemma.

Lemma 4.3.1. *Suppose $na_n^5 \to 0$ as $n \to \infty$, and $\alpha_1(\cdot)$ and $m(\cdot)$ are twice continuously differentiable in a neighbourhood of t. Then, under A3,*

$$\left.\begin{aligned} (na_n)^{1/2}(\hat{\alpha}_n{}^*(t|m) - m(t)\alpha_1(t)) &\xrightarrow{L_1} 0, \\ (na_n)^{1/2}(\hat{\alpha}_n{}^*(t) - \alpha_1(t)) &\xrightarrow{L_1} 0, \end{aligned}\right\} \tag{4.3.16}$$

as $n \to \infty$.

Proof: We show only the first of the two statements in (4.3.16). The proof for the second one is similar.

We have, for $0 \le t \le T$;

$$\begin{aligned} & (na_n)^{1/2}|\hat{\alpha}_n{}^*(t|m) - m(t)\alpha_1(t)| \\ = \; & (na_n)^{1/2}\left|\int_{-1}^{1} K(u)\{m(t-a_nu)\alpha_1(t-a_nu)J_n(t-a_nu) - m(t)\alpha_1(t)\}\,du\right| \\ \le \; & \int_{-1}^{1} K(u)(na_n)^{1/2}|m(t-a_nu)\alpha_1(t-a_nu) - m(t)\alpha_1(t)|\,du \\ & + m(t)\alpha_1(t)\int_{-1}^{1} K(u)(na_n)^{1/2}(1 - J_n(t-a_nu))du \end{aligned} \tag{4.3.17}$$

Thus, from (4.3.17) and our assumptions, we get

$$\begin{aligned} & (na_n)^{1/2}E|\hat{\alpha}_n{}^*(t|m) - m(t)\alpha_1(t)| \\ \le \; & o(1) + m(t)\alpha_1(t)\int_{-1}^{1} K(u)(na_n)^{1/2}\{E(1 - J_n(t-a_nu))\}\,du \\ \le \; & o(1) + (m(t)\alpha_1(t)\int_{-1}^{1} K(u)du)(na_n)^{1/2}E(1 - J_n(T)) \\ = \; & o(1) + o(1), \text{ as } n \to \infty, \end{aligned}$$

since $E(1 - J_n(T)) = (1 - p(T))^n$, where $p(T) = (1 - F_1(T))(1 - G_1(T))$, and $0 < p(T) < 1$. Hence the result.□

Theorem 4.3.2. *Let A1 - A3 hold, $m_2(u) = E(Y_1^2(\bar{G}_2(Y_1))^{-1} \mid X_1 = u)$ be continuous at t and $m(\cdot)$ and $\alpha_1(\cdot)$ be twice continuously differentiable in a neighbourhood of t. Suppose there exists $\epsilon_0 > 0$ such that*

$$\Pr\{Y_1 < T_{F_2} - \epsilon_0 \mid |X_1 - t| < r_0\} = 1 \tag{4.3.18}$$

for some $r_0 > 0$ small enough. Then, if $na_n^5 \to 0$ and $na_n \to \infty$ as $n \to \infty$, we have

$$(na_n)^{1/2}(m_n^0(t) - m(t)) \xrightarrow{d} N\left(0, \frac{D_2(t)}{f_1(t)(1 - G_1(t))} \int_{-1}^{1} K^2(u)du\right)$$

as $n \to \infty$, where $D_2(u) = m_2(u) - (m(u))^2$.

Proof: Consider the decomposition in (4.3.6) again. Using (4.3.18) (which is the same as Assumption (4) of Mielniczuk (1991)), the right-hand side of (4.3.7) and Corollary 1 of Földes and Rejtő (1981), we get, as in the proof of Theorem 3.1(i) of Mielniczuk (1991),

$$(na_n)^{1/2}(\hat{\alpha}_n(t))^{-1}(\hat{\alpha}_n{}^0(t|m) - \hat{\alpha}_n(t|m)) = O((a_n \log\log n)^{1/2}) \text{ a.s.} \tag{4.3.19}$$

Since $na_n^5 \to 0$ as $n \to \infty$, and in view of (4.3.19), (4.3.6), Lemma 4.3.1 and (4.3.15), it is enough to show that, for every pair of real numbers (c_1, c_2)

$$(na_n)^{1/2} U_n(T|t, c_1, c_2) \xrightarrow{d} N(0, Q(t, c_1, c_2)), \text{ as } n \to \infty, \tag{4.3.20}$$

where

$$\begin{aligned} & Q(t, c_1, c_2) \\ = \; & [c_1^2 m_2(t) + 2c_1c_2 m(t) + c_2^2]\alpha_1(t)(1 - F_1(t))^{-1}(1 - G_1(t))^{-1} \int_{-1}^{1} K^2(u)du, \end{aligned}$$

because we already have, by Theorem 4.3.1,

$$\hat{\alpha}_n(t) \xrightarrow{P} \alpha_1(t) = f_1(t)/(1 - F_1(t))$$

as $n \to \infty$. Fix $(c_1, c_2) \in \mathbb{R}^2$. Put

$$H_n(s) = (na_n)^{1/2} a_n^{-1} K((t - s)/a_n) J_n(s)/V_n(s).$$

Then, as noted in Proposition 4.2.1 of R-H, it is enough to show that, as $n \to \infty$,

(i) $\forall \epsilon > 0, \int_0^T H_n^2(s) 1(|H_n(s)| > \epsilon)(c_1 dA_n(s) + c_2 d\Lambda_n(s)) \xrightarrow{P} 0$,

(ii) $\int_0^T H_n^2(s) d < c_1 L_n + c_2 l_n, c_1 L_n + c_2 l_n > (s) \xrightarrow{P} Q(t, c_1, c_2)$.

(Cf. Liptser and Shiryayev (1980), Corollary 2 and Remark 1.) We shall show (ii) first. By Lemma 4.2.2 and 4.2.3,

$$
\begin{aligned}
&\int_0^T H_n^2(s)d < c_1L_n + c_2l_n, c_1L_n + c_2l_n > (s) \\
= \;& c_1^2 \int_{-1}^1 K^2(s)nJ_n(t-a_ns)(V_n(t-a_ns))^{-1}m_2(t-a_ns)\alpha_1(t-a_ns)ds \\
&+2c_1c_2 \int_{-1}^1 K^2(s)nJ_n(t-a_ns)(V_n(t-a_ns))^{-1}m(t-a_ns)\alpha_1(t-a_ns)ds \\
&+c_2^2 \int_{-1}^1 K^2(s)nJ_n(t-a_ns)(V_n(t-a_ns))^{-1}\alpha_1(t-a_ns)ds \qquad (4.3.21)
\end{aligned}
$$

Thus (ii) will follow from (4.3.21), by the continuity of the functions $m_2(\cdot), m(\cdot)$ and $\alpha_1(\cdot)$, if we can show that

$$\sup_{t-h\leq s\leq t+h} |nJ_n(s)(V_n(s))^{-1} - (p(s))^{-1}| \xrightarrow{P} 0, \qquad (4.3.22)$$

as $n\to\infty$, for some $h > 0$ such that $0 \leq t-h < t+h \leq T$, where $p(s) = (1-F_1(s))(1-G_1(s))$. Since $0 < p(T) < 1$, it is clear that there exist $h > 0$ and $a > 0$ such that $[t-h, t+h] \subseteq [0, T]$ and $\inf_{t-h\leq s\leq t+h} p(s) \geq a > 0$. Now for this $h > 0$,

$$
\begin{aligned}
&\sup_{t-h\leq s\leq t+h} \left|nJ_n(s)(V_n(s))^{-1} - (p(s))^{-1}\right| \\
\leq\;& \sup_{t-h\leq s\leq t+h} nJ_n(s)(V_n(s))^{-1}(p(s))^{-1}\left|n^{-1}V_n(s) - p(s)\right| + \sup_{t-h\leq s\leq t+h} (p(s))^{-1}|1 - J_n(s)| \\
\leq\;& a^{-1}\left[\sup_{t-h\leq s\leq t+h} nJ_n(s)(V_n(s))^{-1}\left|n^{-1}V_n(s) - p(s)\right| + \sup_{t-h\leq s\leq t+h} |1 - J_n(s)|\right]. \qquad (4.3.23)
\end{aligned}
$$

Further, for any $\delta > 0$,

$$
\begin{aligned}
&\Pr\left\{\sup_{t-h\leq s\leq t+h} nJ_n(s)(V_n(s))^{-1}\left|n^{-1}V_n(s) - p(s)\right| > \delta\right\} \\
\leq\;& \Pr\left\{\sup_{t-h\leq s\leq t+h} nJ_n(s)(V_n(s))^{-1}\left|n^{-1}V_n(s) - p(s)\right| > \delta, n^{-1}V_n(t+h) \geq p(t+h) - \delta_0\right\} \\
&+ \Pr\{n^{-1}V_n(t+h) < p(t+h) - \delta_0\} \\
\leq\;& \Pr\left\{\sup_{t-h\leq s\leq t+h} \left|n^{-1}V_n(s) - p(s)\right| > \delta(p(t+h) - \delta_0)^{-1}\right\} \\
&+ \Pr\{|n^{-1}V_n(t+h) - p(t+h)| > \delta_0\}, \qquad (4.3.24)
\end{aligned}
$$

where $\delta_0 > 0$ is such that $p(t+h) - \delta_0 > 0$. But by the Glivenko-Cantelli Theorem,

$$\sup_{0\leq s\leq T} |n^{-1}V_n(s) - p(s)| \to 0 \text{ almost surely} \qquad (4.3.25)$$

as $n \to \infty$. Thus (4.3.22) follows from (4.3.23) and (4.3.24), using (4.3.11) and (4.3.25).

For (i), note that

$$\begin{aligned} & \mathbf{1}(|H_n(s)| > \epsilon) \\ = \; & \mathbf{1}(K((t-s)/a_n) n J_n(s)/V_n(s) > \epsilon(na_n)^{1/2}). \end{aligned}$$

Since $na_n \to \infty$ and $K(\cdot)$ is bounded, we get by (4.3.22) that

$$\sup_{t-h \le s \le t+h} \mathbf{1}(|H_n(s)| > \epsilon) \xrightarrow{P} 0 \text{ as } n \to \infty. \tag{4.3.26}$$

But

$$\begin{aligned} & \int_0^T H_n^2(s)\mathbf{1}(|H_n(s)| > \epsilon)(c_1 dA_n(s) + c_2 d\Lambda_n(s)) \\ = \; & \int_{-1}^1 K^2(s) n J_n(t - a_n s)(V_n(t - a_n s))^{-1}\mathbf{1}(|H_n(t - a_n s)| > \epsilon) \\ & [c_1 m(t - a_n s)\alpha_1(t - a_n s) + c_2 \alpha_1(t - a_n s)]ds, \end{aligned}$$

and (i) follows from (4.3.26) and (4.3.22).□

4.4 Some remarks on the extension of the 'martingale methods' to conditional U-statistics

In this section, we make a few observations which led us to believe that the martingale methods are probably not extendable to the more general problem in (1.0.5). For the sake of convenience, we focus on conditional U-statistics of degree two (i.e., $k = 2$) and their weak uniform consistency alone. The difficulty stems from the fact that, in this case, we have to deal with two-parameter processes and filtrations. Though we can obtain a version of Doob's Maximal Inequality for the two-parameter analogue of $\bar{M}_i(\cdot)$ (cf. 2.12), we are unable to do so for the analogue of $L_n(\cdot)$ (cf. Lemma 4.2.1). In other words, while the result is available for each of the summand processes, it is not available for the sum. In fact, we cannot establish even the weakest form of (two-parameter) martingale property for the said analogue of $L_n(\cdot)$.

To keep our calculations at a simple level, we take $h(\cdot,\cdot)$ to be a product function, i.e.,

$$h(y_1, y_2) = \varphi(y_1)\varphi(y_2),$$

where $\varphi(\cdot)$ is such that $E|\varphi(Y_1)| < \infty$. Further, assume that $\varphi(\cdot)$ is non-negative. Let the set-up be as described in the Introduction. Define, in analogy with (4.2.6),

$$\left.\begin{array}{rcl} Q_i(t|\varphi) & = & \eta_i\varphi(\tilde{Y}_i)(\bar{G}_2(\tilde{Y}_i))^{-1}\mathbf{1}(\tilde{X}_i \leq t, \delta_i = 1), \\ B_i(t|\varphi) & = & \int_0^t \mathbf{1}(\tilde{X}_i \geq u)m(u|\varphi)\alpha_1(u)du, \end{array}\right\} \quad (4.4.1)$$

$0 \leq t \leq T$, $1 \leq i \leq n$, where $m(u|\varphi) := E(\varphi(Y_1) \mid X_1 = u)$. Put

$$L_{ij}(\cdot,\cdot|\varphi) := Q_i(\cdot|\varphi)Q_j(\cdot|\varphi) - B_i(\cdot|\varphi)B_j(\cdot|\varphi).$$

Note that $L_{ij}(\cdot,\cdot|\varphi)$ is the two-parameter analogue of $\bar{M}_i(\cdot)$ defined in Lemma 4.2.2.

Now, with $\bar{M}_i(\cdot|\varphi) := Q_i(\cdot|\varphi) - B_i(\cdot|\varphi)$, we have the following decomposition:

$$\begin{aligned} & Q_i(t_1|\varphi)Q_j(t_2|\varphi) - B_i(t_1|\varphi)B_j(t_2|\varphi) \\ = \; & \bar{M}_i(t_1|\varphi)B_j(t_2|\varphi) + B_i(t_1|\varphi)\bar{M}_j(t_2|\varphi) + \bar{M}_j(t_1|\varphi)\bar{M}_j(t_2|\varphi). \end{aligned} \quad (4.4.2)$$

By arguments similar to the proof of Equation (4.2.7), we can see that $\bar{M}_i(\cdot|\varphi)$, $0 \leq t \leq T$, is a $\{\mathcal{G}_i(t) \,:\, 0 \leq t \leq T\}$-martingale, $1 \leq i \leq n$. Further, note that $B_i(\cdot|\varphi), 1 \leq i \leq n$, has sample paths of bounded variation (increasing, in fact). Thus $\bar{M}_i(\cdot|\varphi)B_j(\cdot|\varphi)$ is a *proper 1-martingale*, i.e., a martingale in the 1st co-ordinate and a process of bounded variation in the 2nd. Similarly, $B_i(\cdot|\varphi)\bar{M}_j(\cdot|\varphi)$ is a *proper 2-martingale*. We make these concepts precise in the following.

Let $\mathcal{G}_i(t)$ be as defined in (4.2.1) and consider the two-parameter filtration, with $1 \leq i \neq j \leq n$,

$$\mathcal{G}_{ij}(t_1, t_2) = \mathcal{G}_i(t_1) \otimes \mathcal{G}_i(t_2), 0 \leq t_r \leq T, r = 1, 2. \quad (4.4.3)$$

We give below the definitions of three types of two-parameter martingales. For more details on these and related concepts, see Cairoli and Walsh (1975), Wong and Zakai (1976) and Merzbach and Zakai (1980).

In the following, it is assumed that $E|X(t_1,t_2)| < \infty$ and $X(t_1,t_2)$ is $\mathcal{F}(t_1,t_2)$-measurable, where $\{\mathcal{F}(t_1,t_2) : (t_1,t_2) \in [0,T]^2\}$ is a filtration on some probability space, and $[0,T]^2 := [0,T] \times [0,T]$.

Definition 4.4.1. A process $X(\cdot,\cdot)$ is said to be a *'martingale'* if

$$E\{X(t_1+s_1,t_2+s_2) - X(t_1,t_2) \mid \mathcal{F}(t_1,t_2)\} = 0$$

$\forall (t_1,t_2) \in [0,T]^2, s_r \geq 0, r = 1,2.$

Definition 4.4.2. A process $X(\cdot,\cdot)$ is said to be a *weak martingale* if

$$E\{\triangle(s_1,s_2)X(t_1,t_2) \mid \mathcal{F}(t_1,t_2)\} = 0,$$

where

$$\triangle(s_1,s_2)X(t_1,t_2) := X(t_1+s_1,t_2+s_2) - X(t_1,t_2+s_2) - X(t_1+s_1,t_2) + X(t_1,t_2),$$

$\forall (t_1,t_2) \in [0,T]^2, s_r \geq 0, r = 1,2.$

Definition 4.4.3.

(a) A process $X(\cdot,\cdot)$ is said to be a *1-martingale* if, for each fixed $0 \leq t_2 \leq T$, $X(\cdot,t_2)$ is a martingale with respect to $\{\mathcal{F}(t,t_2) : 0 \leq t \leq T\}$.

(b) A 1-martingale $X(\cdot,\cdot)$ is said to be *proper* if, for each fixed $0 \leq t_1 \leq T$, the function $X(t_1,\cdot)$ is of bounded variation on $[0,T]$ and $Ev(X(t_1,\cdot)) < \infty$, where $v(X(t_1,\cdot))$ denotes the total variation of $X(t_1,\cdot)$ on $[0,T]$.

A *proper 2-martingale* is defined similarly.

Remark 4.4.1. The definition of a 1-martingale here is taken from Wong and Zakai (1976). It is slightly different from that of Cairoli and Walsh (1975) and other authors. Note that a 'martingale' is necessarily an i-martingale, $i = 1,2$, which is necessarily a weak martingale.

We then have the following result:

Proposition 4.4.1. Let the assumption A1 hold. Then, with respect to $\{\mathcal{G}_{ij}(t_1,t_2) : (t_1,t_2) \in [0,T]^2\}$,

(i) $\bar{M}_i(\cdot|\varphi)\bar{M}_j(\cdot|\varphi)$ is a 'martingale',

(ii) $L_{ij}(\cdot,\cdot|\varphi)$ is a weak martingale.

(iii) If, in addition, A2 holds and $E\varphi^2 < \infty$, then

$$M_i^2(\cdot|\varphi)M_j^2(\cdot|\varphi) - B_{2i}(\cdot|\varphi)B_{2j}(\cdot|\varphi)$$

is a weak martingale, where

$$\begin{aligned} B_{2i}(t|\varphi) &:= <\bar{M}_i(\cdot|\varphi), \bar{M}_i(\cdot|\varphi)> (t) \\ &= \int_0^t 1(\tilde{X}_i \geq u) m_2(u|\varphi)\alpha_1(u)du, \end{aligned}$$

$0 \leq t \leq T$, and $m_2(u|\varphi) := E\{\varphi^2(Y_1)(\bar{G}_2(Y_1))^{-1} \mid X_1 = u\}$ (see (4.2.14)).

Proof: First, note that the adaptedness and the integrability requirements are satisfied in (i), (ii) and (iii). Next,

(i) for $0 \leq t_r < T,\ s_r \geq 0, r = 1,2,$

$$\begin{aligned} & E\{\bar{M}_i(t_1+s_1|\varphi)\bar{M}_j(t_2+s_2|\varphi) - \bar{M}_i(t_1|\varphi)\bar{M}_j(t_2|\varphi)\,|\mathcal{G}_{ij}(t_1,t_2)\} \\ = \;& E\{\bar{M}_i(t_1+s_1|\varphi)\bar{M}_j(t_2+s_2|\varphi) - \bar{M}_i(t_1|\varphi)\bar{M}_j(t_2+s_2|\varphi)\,|\mathcal{G}_i(t_1)\otimes\mathcal{G}_j(t_2)\} \\ & + E\{\bar{M}_i(t_1|\varphi)\bar{M}_j(t_2+s_2|\varphi) - \bar{M}_i(t_1|\varphi)\bar{M}_j(t_2|\varphi)\,|\mathcal{G}_i(t_1)\otimes\mathcal{G}_j(t_2)\} \\ = \;& E\{\bar{M}_i(t_1+s_1|\varphi) - \bar{M}_i(t_1|\varphi)\,\big|\mathcal{G}_i(t_1)\}E\{\bar{M}_j(t_2+s_2|\varphi)\big|\,\mathcal{G}_j(t_2)\} \\ & + E\{\bar{M}_i(t_1|\varphi)\,\big|\mathcal{G}_i(t_1)\}E\{\bar{M}_j(t_2+s_2|\varphi) - \bar{M}_j(t_2|\varphi)\big|\,\mathcal{G}_j(t_2)\} \\ = \;& 0, \end{aligned}$$

by the martingale property of $\bar{M}_i(\cdot|\varphi)$ and the independence of $\mathcal{G}_i(t_1)$ and $\mathcal{G}_j(t_2)$.

(ii) Recall

$$L_{ij}(\cdot,\cdot|\varphi) := Q_i(\cdot|\varphi)Q_j(\cdot|\varphi) - B_i(\cdot|\varphi)B_j(\cdot|\varphi).$$

Then

$$\begin{aligned}&\Delta(s_1,s_2)L_{ij}(t_1,t_2)\\ =\;&(Q_i(t_1+s_1|\varphi)-Q_i(t_1|\varphi))(Q_j(t_2+s_2|\varphi)-Q_j(t_2|\varphi))\\ &-(B_i(t_1+s_1|\varphi)-B_i(t_1|\varphi))(B_j(t_2+s_2|\varphi)-B_j(t_2|\varphi)).\end{aligned} \tag{4.4.4}$$

Now by the fact that $Q_i(\cdot|\varphi)-B_i(\cdot|\varphi)$ is a martingale, and using the independence of $\mathcal{G}_i(t_1)$ and $\mathcal{G}_j(t_2)$, as in (i) above, we get the result.

(iii) Note that A2 and $E\varphi^2(Y_1)<\infty$ together imply that $\bar{M}_i(\cdot|\varphi)$ is a square-integrable martingale and that

$$<\bar{M}_i(\cdot|\varphi),\bar{M}_i(\cdot|\varphi)>=B_{2i}(\cdot|\varphi),$$

in the same way as in the proof of Lemma 4.2.2 (see the arguments following Equation (4.2.14)). In other words, $\bar{M}_i^2(\cdot|\varphi)-B_{2i}(\cdot|\varphi)$ is a $\{\mathcal{G}_i(t):0\le t\le T\}$-martingale. Putting

$$L'_{ij}(\cdot,\cdot|\varphi):=\bar{M}_i^2(\cdot|\varphi)\bar{M}_j^2(\cdot|\varphi)-B_{2i}(\cdot|\varphi)B_{2j}(\cdot|\varphi),$$

and comparing $L'_{ij}(\cdot,\cdot|\varphi)$ with $L_{ij}(\cdot,\cdot|\varphi)$ in (ii) above, it is clear that the result follows exactly as in (ii).□

Remark 4.4.2. By the discussion following Equation (4.4.2), it is already seen that $\bar{M}_i(\cdot|\varphi)B_j(\cdot|\varphi)$ and $B_i(\cdot|\varphi)\bar{M}_j(\cdot|\varphi)$ are a proper 1-martingale and a proper 2-martingale respectively. Combining Equation (4.4.2) and Proposition 4.4.1, it is thus seen that Equation (4.4.2) serves as an example for the general decomposition result (Theorem 2.4) of Wong and Zakai (1976). Further, Proposition 4.4.1(iii) above is an example of a two-parameter analogue of the well-known *Doob-Meyer decomposition* (see Cairoli and Walsh (1975), Theorem 1.5, for a general result).

Now, we can obtain a version Doob's Inequality for $L_{ij}(\cdot,\cdot|\varphi)$ (cf. Proposition 4.4.1(ii) above) using (4.4.2). First, since $\bar{M}_i(\cdot|\varphi)$ is a martingale and $B_i(\cdot|\varphi)$ is non-negative and increasing in t, we get

$$E\left\{\sup_{(t_1,t_2)\in[0,T]^2}|\bar{M}_i(t_1|\varphi)B_j(t_2|\varphi)|^2\right\}$$

$$
\begin{aligned}
&\leq E\left\{\sup_{0\leq t_1\leq T}|\tilde{M}_i(t_1|\varphi)|^2\right\}EB_j^2(T|\varphi),\\
&\qquad\text{by the independence of } \tilde{M}_i(\cdot|\varphi) \text{ and } B_j(\cdot|\varphi),\\
&\leq 4EB_{2i}(T|\varphi)B_{2j}(T|\varphi).\{-\log(1-F_1(T))\},
\end{aligned}
\tag{4.4.5}
$$

since, by Cauchy-Schwartz Inequality and the fact that $\{m(u|\varphi)\}^2 \leq m_2(u|\varphi)$,

$$
\begin{aligned}
B_j^2(T|\varphi) &\leq \left\{\int_0^T \mathbf{1}(\tilde{X}_j\geq u)\{m(u|\varphi)\}^2\alpha_1(u)du\right\}\left\{\int_0^T \mathbf{1}(\tilde{X}_j\geq u)\alpha_1(u)du\right\}\\
&\leq \left\{\int_0^T \mathbf{1}(\tilde{X}_j\geq u)m_2(u|\varphi)\alpha_1(u)du\right\}\left\{\int_0^T \alpha_1(u)du\right\}\\
&= B_{2j}(T|\varphi).\{-\log(1-F_1(T))\}.
\end{aligned}
$$

Next, by similar arguments,

$$
\begin{aligned}
&E\left\{\sup_{(t_1,t_2)\in[0,T]^2}|\tilde{M}_i(t_1|\varphi)\tilde{M}_j(t_2|\varphi)|^2\right\}\\
&\leq 16EB_{2i}(T|\varphi)B_{2j}(T|\varphi).
\end{aligned}
\tag{4.4.6}
$$

Combining (4.4.2), (4.4.5) and (4.4.6), we have

$$
\begin{aligned}
&E\left\{\sup_{(t_1,t_2)\in[0,T]^2}|L_{ij}(t_1,t_2|\varphi)|^2\right\}\\
&\leq 3\,\{4EB_{2i}(T|\varphi)B_{2j}(T|\varphi)[-\log(1-F_1(T))]+4EB_{2i}(T|\varphi)B_{2j}(T|\varphi)[-\log(1-F_1(T))]\\
&\quad +16EB_{2i}(T|\varphi)B_{2j}(T|\varphi)\}\\
&\leq 48EB_{2i}(T|\varphi)B_{2j}(T|\varphi)[1-\log(1-F_1(T))].
\end{aligned}
\tag{4.4.7}
$$

Now, some version of Doob's Inequality is useful in proving weak uniform consistency, as is evident from the proof of Theorem 4.3.1. *But, unfortunately, our method of analysis fails beyond this point.* To see this, note that the natural analogue of $L_n(t)$ is

$$
L_n(t_1,t_2|\varphi) := \sum_{1\leq i\neq j\leq n} L_{ij}(t_1,t_2|\varphi),
$$

but we cannot even establish that $L_n(\cdot,\cdot|\varphi)$ is a weak martingale, following our method. The *problem* is that of choosing a suitable analogue of the filtration $\{\mathcal{F}_n(t): 0\leq t\leq T\}$.

By direct analogy, we can define

$$\mathcal{F}_n(t_1,t_2) := \bigvee_{1\le i\ne j\le n} \mathcal{G}_{ij}(t_1,t_2), \ (t_1,t_2)\in[0,T]^2.$$

Using the notation of (4.2.8), we can write

$$\begin{aligned}\mathcal{G}_i(t_1)\otimes\mathcal{G}_j(t_2) &= \sigma\{\ Z_i(B^{(i)},s^{(i)}), W_i(s^{(i)}), Z_j(B^{(j)},s^{(j)}), W_j(s^{(j)}) : \\ &\qquad 0\le s^{(i)}\le t_1, 0\le s^{(j)}\le t_2, B^{(i)}, B^{(j)}\in\mathcal{B}(\mathbb{R}_+)\}. \qquad (4.4.8)\end{aligned}$$

It is clear from (4.4.8) that

$$\begin{aligned}\mathcal{F}_n(t_1,t_2) &= \bigvee_{1\le i\ne j\le n} \mathcal{G}_i(t_1)\otimes\mathcal{G}_j(t_2) \\ &= \bigvee_{1\le i\ne j\le n} \mathcal{G}_i(t_1\vee t_2)\otimes\mathcal{G}_j(t_1\vee t_2) \\ &= \mathcal{F}_n(t_1\vee t_2, t_1\vee t_2). \qquad (4.4.9)\end{aligned}$$

We get '$t_1\vee t_2$' in (4.4.9) because, for each pair (i,j), both the σ-fields $\mathcal{G}_i(t_1)\otimes\mathcal{G}_j(t_2)$ and $\mathcal{G}_j(t_1)\otimes\mathcal{G}_i(t_2)$ are pooled into $\mathcal{F}_n(t_1,t_2)$. It is because of (4.4.9) that our method breaks down. Suppose, for example, we claim that

$$\sum_{1\le i\ne j\le n} \bar{M}_i(\cdot|\varphi)\bar{M}_j(\cdot|\varphi)$$

is an $\{\mathcal{F}_n(t_1,t_2):(t_1,t_2)\in[0,T]^2\}$-'martingale'. Then we must show that, for all $(t_1,t_2)\in[0,T]^2, s_r\ge 0, r=1,2,$

$$E\{\sum_{1\le i\ne j\le n}\left[\bar{M}_i(t_1+s_1|\varphi)\bar{M}_j(t_2+s_2|\varphi)-\bar{M}_i(t_1|\varphi)\bar{M}_j(t_2|\varphi)\right]|\mathcal{F}_n(t_1,t_2)\} = 0. \qquad (4.4.10)$$

But, assuming WOLG $t_2\ge t_1$, we have by (4.4.9) and the independence of $\bar{M}_i(\cdot|\varphi)\bar{M}_j(\cdot|\varphi)$ and $\mathcal{G}_{lm}(t_1,t_2)$ whenever $\{i,j\}\cap\{l,m\}=\emptyset$,

the left-hand side of (4.4.10)

$$= \sum_{1\le i\ne j\le n} E\left\{\left[\bar{M}_i(t_1+s_1|\varphi)\bar{M}_j(t_2+s_2|\varphi)-\bar{M}_i(t_1|\varphi)\bar{M}_j(t_2|\varphi)\right]|\mathcal{G}_i(t_2)\otimes\mathcal{G}_j(t_2)\right\}$$

$$
\begin{aligned}
&= \sum_{1\le i\ne j\le n} \Big[E\{\bar{M}_i(t_1+s_1|\varphi) - \bar{M}_i(t_1|\varphi)\,\big|\mathcal{G}_i(t_2)\}E\{\bar{M}_j(t_2+s_2|\varphi)\big|\,\mathcal{G}_j(t_2)\} \\
&\qquad + E\{\bar{M}_i(t_1|\varphi)\,\big|\mathcal{G}_i(t_2)\}E\{\bar{M}_j(t_2+s_2|\varphi) - \bar{M}_j(t_2|\varphi)\big|\,\mathcal{G}_j(t_2)\}\Big],
\end{aligned}
$$

as in the proof of Proposition 4.4.1(i),

$$
= \begin{cases} \sum_{1\le i\ne j\le n}[\bar{M}_i(t_1+s_1|\varphi) - \bar{M}_i(t_1|\varphi)]\bar{M}_j(t_2|\varphi), & \text{if } t_1+s_1 < t_2 \\ \sum_{1\le i\ne j\le n}[\bar{M}_i(t_2|\varphi) - \bar{M}_i(t_1|\varphi)]\bar{M}_j(t_2|\varphi), & \text{if } t_1+s_1 \ge t_2 \end{cases}. \tag{4.4.11}
$$

It is clear from (4.4.11) that for $\sum_{i\ne j}\bar{M}_i(\cdot|\varphi)\bar{M}_j(\cdot|\varphi)$ to be a 'martingale' for $n\ge 2$, we must have

$$
\sum_{i\ne j}\left[\bar{M}_i(t_2|\varphi) - \bar{M}_i(t_1|\varphi)\right]\bar{M}_j(t_2|\varphi) = 0 \ \forall T\ge t_2\ge t_1\ge 0,\ n\ge 2. \tag{4.4.12}
$$

Specialize n to $n=2$, i.e., suppose that there are just two data-points with indices i and j, say. Then (4.4.12) gives

$$
\left[\bar{M}_i(t_2|\varphi) - \bar{M}_i(t_1|\varphi)\right]\bar{M}_j(t_2|\varphi) = -\left[\bar{M}_j(t_2|\varphi) - \bar{M}_j(t_1|\varphi)\right]\bar{M}_i(t_2|\varphi). \tag{4.4.13}
$$

Since we have Doob's inequality in mind, we may further assume that $\bar{M}_i(\cdot|\varphi)$ is square-integrable. Then by (4.4.13) and using the martingale property we have

$$
\begin{aligned}
& E\left[\bar{M}_i(t_2|\varphi) - \bar{M}_i(t_1|\varphi)\right]^2\bar{M}_j^2(t_2|\varphi) \\
= & -E\left[\bar{M}_i(t_2|\varphi) - \bar{M}_i(t_1|\varphi)\right]\bar{M}_i(t_2|\varphi)E\left[\bar{M}_j(t_2|\varphi) - \bar{M}_j(t_1|\varphi)\right]\bar{M}_j(t_2|\varphi) \\
= & -E\left[\bar{M}_i(t_2|\varphi) - \bar{M}_i(t_1|\varphi)\right]^2E\left[\bar{M}_j(t_2|\varphi) - \bar{M}_j(t_1|\varphi)\right]^2,
\end{aligned}
$$

which can hold if and only if

$$
E\left[\bar{M}_i(t_2|\varphi) - \bar{M}_i(t_1|\varphi)\right]^2 = 0,
$$

i.e., if and only if either $t_2 = t_1$ or $\bar{M}_i(\cdot|\varphi)$ has a constant path almost surely. Similarly one runs into trouble trying to show that $\sum_{i\ne j}L_{ij}(\cdot,\cdot|\varphi)$ is a weak martingale.

Note further that using the term-wise bounds in (4.4.7), we end up with *too large* a bound, viz.,

$$E\left\{\sup_{(t_1,t_2)\in[0,T]^2}\left|\sum_{i\neq j}L_{ij}(t_1,t_2|\varphi)\right|^2\right\}$$
$$\leq \binom{n}{2}\sum_{i\neq j}E\left\{\sup_{(t_1,t_2)\in[0,T]^2}L_{ij}^2(t_1,t_2|\varphi)\right\}.$$

These observations demonstrate that the martingale methods are probably non-applicable to the study of conditional U-statistics.

Chapter 5

Problems for Further Research

5.1 A law of the iterated logarithm for $U_n^h(\mathbf{t})$

In this section, we outline a possible approach towards proving a *law of the iterated logarithm* (LIL) for $U_n^h(\mathbf{t})$ for a fixed $\mathbf{t} \in I\!R^k$. We try to adapt the technique of an almost-sure representation of classical U-statistics as a mean of i.i.d random variables, as given in Serfling (1980) (Section 5.3.3, p.189). Let, as before, $h : I\!R^k \to I\!R$ be a symmetric function and let $U_n(h)$ be the U-statistic based on $h(\cdot)$ and an i.i.d. sample $X_1, \ldots, X_n$, $n \geq k$. Further, assume $Eh^2 < \infty$. Let

$$h^{(1)}(\cdot) := Eh(\cdot, X_2, \ldots, X_k) - Eh \tag{5.1.1}$$

and $U_n(h^{(1)})$ be the U-statistic based on $h^{(1)}(\cdot)$, i.e.,

$$U_n(h^{(1)}) = n^{-1} \sum_{i=1}^{n} h^{(1)}(X_i). \tag{5.1.2}$$

Thus $U_n(h^{(1)})$ has the structure of a mean of i.i.d's. It is the 1st-order term in the *Dynkin-Mandelbaum decomposition* (cf. Section 3.3 of this thesis) of $U_n(h)$. Now the following almost-sure representation is a special case of Theorem 5.3.3 of Serfling (1980):

Theorem 5.1.1. *Put*

$$R_n(h) = U_n(h) - Eh - kU_n(h^{(1)}).$$

Then, for any $\delta > 1/2$, with probability 1,

$$R_n(h) = o(n^{-1}(\log n)^{\delta}), \quad as\ n \to \infty.$$

Now, an LIL holds for $U_n(h^{(1)})$ by virtue of its i.i.d. mean structure and the square-integrability of $h(\cdot)$. It follows from Theorem 5.1.1 that the LIL holds also for $U_n(h)$, since the remainder $R_n(h)$ is sufficiently small almost surely (Problem 5.P.15, p.208, Serfling (1980)).

Now consider $U_n(h, \mathbf{t})$. Here the 1st-order term in the D-M decomposition is based on

$$h_{nj}^{(1)}(\cdot) := h_n^{(0,\ldots,1,\ldots,0)}(\cdot),\ 1 \leq j \leq m, \tag{5.1.3}$$

where $h_n^{(0,\ldots,1,\ldots,0)}(\cdot)$ is as defined in (3.3.22) of Section 3.3, and the vector $(0, \ldots, 1, \ldots, 0)$ has m co-ordinates with '1' in the j-th co-ordinate (cf. Subsection 3.3.2). Since $U_n(h_{nj}^{(1)}, \mathbf{t})$, $1 \leq j \leq m$, has the structure of a triangular array of i.i.d. means, one may attempt to obtain an LIL for $U_n(h, \mathbf{t})$ via the approach suggested by Theorem 5.1.1: (1) obtain an LIL for $U_n(h_{nj}^{(1)}, \mathbf{t})$; (2) obtain a suitable almost-sure order bound for

$$R_n(h, \mathbf{t}) := U_n(h, \mathbf{t}) - Eh. \prod K_n - \sum_{j=1}^{m} k_j U_n(h_{nj}^{(1)}, \mathbf{t}). \tag{5.1.4}$$

An LIL for $U_n^h(\mathbf{t})$ will then follow from the decomposition (2.2.2).

As for the method of proof, note that two facts are made use of in the proof of Theorem 5.1.1: first, $E(R_n(h))^2 = O(n^{-2})$ (Theorem 5.3.2, p.188, of Serfling (1980)), and second,

$$\Pr\left\{\sup_{j \geq n} |R_j(h)| > \epsilon\right\} \leq \epsilon^{-2} E(R_n(h))^2,\ \epsilon > 0, \tag{5.1.5}$$

by Doob's maximal inequality, since $\{R_n(h), n \geq k\}$ is a *reverse martingale* with respect to the 'symmetric' filtration. Now by the assumptions of Section 3.3 (cf. Equation (3.3.27)), we have

$$E(R_n(h, \mathbf{t}))^2 = O((na_n)^{-2}). \tag{5.1.6}$$

By the Borel-Cantelli Lemma, (5.1.6) implies $R_n(h,\mathbf{t}) \to 0$ almost surely as $n \to \infty$, provided $\sum_{n\geq 1}(na_n)^{-2} < \infty$. However, the sequence $\{R_n(h,\mathbf{t})\}_{n\geq k}$ does not have any martingale property. Instead, one could think of using a result of Burkholder (1964) who obtains conditions under which almost sure convergence implies a maximal inequality. Unfortunately, our attempt to use Theorem 2 of Burkholder (1964) did not succeed. We are not able to verify his Condition C_p for our problem. Probably some other technique would be effective here.

5.2 Limit distribution of the uniform absolute deviation

Another possible line of investigation could be a study of the limiting behaviour of the (normalized) *uniform absolute deviation* criterion for $U_n^h(\mathbf{t})$:

$$\sup_{\mathbf{t}\in C} (U_n(1,\mathbf{t}))^{1/2} \left|U_n^h(\mathbf{t}) - m(\mathbf{t})\right|, \tag{5.2.1}$$

where C is a suitable compact set. One may try to extend the results of Csörgő and Révész (1981) (Section 6.3, p.237, in particular Theorem 6.3.3) where a *strong approximation* of partial-sum processes by Wiener processes is made use of. Recall that, in the notation of Section 2.2 (of the present work),

$$U_n^h(\mathbf{t}) - m(\mathbf{t}) = R_{1n}(\mathbf{t}) + R_{2n}(\mathbf{t}), \text{ say,} \tag{5.2.2}$$

where

$$(U_n(1,\mathbf{t}))R_{1n}(\mathbf{t}) = U_n(h,\mathbf{t}) - EU_n(h,\mathbf{t}) + EU_n(h,\mathbf{t}) - m(\mathbf{t})\prod_{j=1}^{k} f_1(t_j) \tag{5.2.3}$$

$$(U_n(1,\mathbf{t}))R_{2n}(\mathbf{t}) = (-m(\mathbf{t}))\left(U_n(1,\mathbf{t}) - EU_n(1,\mathbf{t}) + EU_n(1,\mathbf{t}) - \prod_{j=1}^{k} f_1(t_j)\right) \tag{5.2.4}$$

Further, the stochastic part in (5.2.3) can be written as

$$U_n(h,\mathbf{t}) - EU_n(h,\mathbf{t}) = (-1)^k a_n^{-k} \int_{\mathbb{R}^k} [\mu_n((-\infty,\mathbf{x}]|h) - \mu((-\infty,\mathbf{x}]|h)] \prod_{j=1}^{k} dK((t_j - x_j)/a_n). \tag{5.2.5}$$

In Chapter 2, we obtained exponential tail-probability bounds for the integrand in (5.2.5):

$$E_n(\mathbf{x}|h) := (\mu_n((-\infty,\mathbf{x}]|h) - \mu((-\infty,\mathbf{x}]|h)),\ \mathbf{x} \in \mathbb{R}^k.$$

Regarding the problem mentioned above, one may instead look for a strong approximation of the centered empirical process $E_n(\cdot|h)$ by a suitable Gaussian random field. Then the limiting behaviour of (5.2.1) will be related to the distribution of the supremum of that random field. By looking at the structure of $\mu_n((-\infty,\mathbf{x}]|h)$, it seems that the appropriate candidate would be the mean-zero Gaussian random field with covariance

$$R^h(\mathbf{x},\mathbf{y}) = Eh^2(Y_1,\ldots,Y_k)\mathbf{1}_{(-\infty,\mathbf{x}\wedge\mathbf{y}]}(X_1,\ldots,X_k) - \mu((-\infty,\mathbf{x}]|h)\mu((-\infty,\mathbf{y}]|h),$$

for $\mathbf{x},\mathbf{y} \in \mathbb{R}^k$, where $\mathbf{x}\wedge\mathbf{y} := (x_1 \wedge y_1, \ldots, x_k \wedge y_k)$.

Further, as is clear from (5.2.4), we shall have to consider a strong approximation of the process $E_n(\cdot|1)$ (obtained by putting $h \equiv 1$ in $E_n(\cdot|h)$) as well, in order to handle $R_{2n}(\mathbf{t})$.

The approach, however, appears to be quite difficult, given the complicated expression for $R^h(\mathbf{x},\mathbf{y})$ above.

BIBLIOGRAPHY

1. Aalen, O.O. (1978). Non-parametric inference for a family of counting processes. *Ann. Statist.* **6**, 701 - 726.

2. Alexander, K.S. (1984). Probability inequalities for empirical processes and a law of the iterated logarithm. *Ann. Probab.* **12**, 1041 - 1067.

3. Bochynek, J. (1987). *Asymptotische Normalität bedingter U-statistiken*, Ph.D. dissertation, Justus Liebig Universität, Giessen.

4. Burkholder, D.L. (1964). Maximal inequalities as necessary conditions for almost everywhere convergence. *Z. Wahrsch. verw. Gebiete* **3**, 75 - 88.

5. Cairoli, R. and Walsh, J.B. (1975). Stochastic integrals in the plane. *Acta Math.* **134**, 111 - 183.

6. Chung, K.L. (1974). *A course in probability theory.* New York: Academic Press.

7. Csörgő, M. and Révész, P. (1981). *Strong approximations in probability and statistics.* New York: Academic Press.

8. Cuzick, J. (1982). Rank tests for association with right-censored data. *Biometrika* **69**, 351 - 364.

9. Dudley, R.M. (1979). Balls in $\mathbb{R}^k$ do not cut all subsets of $k+2$ points. *Adv. Math.* **31**, 306 - 308.

10. Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution functions and the classical multinomial estimator. *Ann. Math. Statist.* **27**, 642 - 669.

11. Dynkin, E.B. and Mandelbaum, A. (1983). Symmetric statistics, Poisson point processes and multiple Wiener integrals. *Ann. Statist.* **11**, 739-745.

12. Földes, A. and Rejtő, L. (1981). A LIL type result for the product-limit estimator. *Z. Wahrsch. verw. Gebiete* **56**, 75 - 86.

13. Gill, R.D. (1980). *Censoring and stochastic integrals.* Mathematical Centre Tracts **124**. Mathematisch Centrum, Amsterdam.

14. Härdle, W. Janssen, P. and Serfling, R. (1988). Strong uniform consistency rates for estimators of conditional functionals. *Ann. Statist.* **16**, 1428 - 1449.

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

16. Hoeffding, W. (1961). *The strong law of large numbers for U-statistics.* Institute of Statistics, Univ. of N. Carolina, Mimeograph Series **302**.

17. Hollander, M. and Proschan, F. (1972). Testing whether new is better than used. *Ann. Math. Statist.* **43**, 1136 - 1146.

18. Janssen, P. (1988). *Generalized empirical distribution functions with statistical applications.* Limburgs Universitaire Centrum, Diepenbeek.

19. Kallianpur, G. (1980). *Stochastic filtering theory.* New York: Springer.

20. Laha, R.G. and Rohatgi, V.K. (1979) *Probability theory.* New York: Wiley.

21. Liero, H. (1991). Asymptotic behavior of conditional U-statistics. *Proc. Satellite Meeting of the ISI Congress on Order Statistics and Nonparametrics,' Theory and Applic.* Alexandria, Egypt. (to appear)

22. Liptser, R. Sh. and Shiryayev, A.N. (1978). *Statistics of random processes* (vol. 2). New York : Springer.

23. Liptser, R. Sh. and Shiryayev, A.N. (1980). A functional central limit theorem for semimartingales. *Theor. Probab. Appl.* **25**, 4, 667 -688.

24. Merzbach, E. and Zakai, M. (1980). Predictable and dual predictable projections of two-parameter stochastic processes. *Z. Wahrsch. verw. Gebiete* **53**, 263 -269.

25. Metivier, M. (1982). *Semimartingales : a course on stochastic processes.* Berlin : Walter de Gruyter.

26. Mielniczuk, J. (1991). Some asymptotic properties of non-parametric regression estimators in case of randomly censored data. *Statistics* **22**, 85 - 93.

27. Nadaraya, E. (1964). On estimating regression. *Theory Probab. Appl.* **9**, 141 - 142.

28. Parthasarathy, K.R. (1992). *An introduction to quantum stochastic calculus.* Basel: Birkhäuser.

29. Prakasa Rao, B.L.S. (1983). *Non-parametric functional estimation.* Orlando: Academic Press.

30. Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11**, 453 - 466.

31. Rubin, H. and Vitale, R.A. (1980). Asymptotic distribution of symmetric statistics. *Ann. Statist.* **8**, 165-170.

32. Schneemeier, W. (1989). Weak convergence and Glivenko-Cantelli results for empirical processes of U-statistic structure. *Stoch. Processes and their Appl.* **33**, 325 - 334.

33. Schuster, E.F. (1972). Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. *Ann. Math. Stat.* **43**, 84-88.

34. Serfling, R.J. (1980). *Approximation theorems of mathematical statistics.* New York: Wiley.

35. Stute, W. (1991). Conditional U-statistics, *Ann. Probab.* **19**, 812 - 825.

36. Watson, G.S. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26**, 359 - 372.

37. Wong, E. and Zakai, M. (1976). Weak martingales and stochastic integrals in the plane. *Ann. Probab.* **4**, 570 - 586.