Limit theorems for statistical functionals with applications to dimension estimation

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

vorgelegt von
Aleksey Min
aus
Koshkupir, Usbekistan

Göttingen 2004

D7

Referent: Prof. Dr. Manfred Denker

Korreferentin: Prof. Dr. Susanne Koch

Tag der mündlichen Prüfung: 23.06.2004

Contents

Introduction

In [27] HAUSDORFF laid the foundation of dimension theory of sets generalizing the classical notion. His definition can be extended via CARATHEODORY's construction (see MATTILA [41]). Investigations of this type were at least partly inspired by the problem of finding space filling curves. One of the most popular fractal set (the SIERPIŃSKI gasket) also originates in this context. Basic results were obtained in the 1920's by BESIKOVIČ [7] and JARNICK [34].
The theory being dormant for many years was revived about 30 years ago (mainly in physics), when new methods for computing dimension were introduced. This led to box dimension (KOLMOGOROV [36]) and packing dimension (TAYLOR and TRICOT [49], SULLIVAN [48]). All these notions are based on some distance functions and the exterior measure generated from it.
Starting with RÉNYI [45], some notions of dimensions were introduced for random variables and later by HENTSCHEL and PROCACCIA [29] for probability measures. This has been developed in mathematical terms by CUTLER and DAWSON [13], [14] and CUTLER [10]. These dimensions are called local dimension and information dimension.
It is one of the important problems in dimension theory to examine the relations between the different types of dimension. If we denote HAUSDORFF, packing, lower box and upper box dimensions by $dim_{haus}, dim_{pack}, dim^{-}_{box}$ and dim^{+}_{box}, respectively, then one has the following inequalities

$$dim_{haus}(E) \leq dim^{-}_{box}(E) \quad \text{and} \quad dim_{pack}(E) \leq dim^{+}_{box}(E)$$

for any bounded set $E \in \mathbb{R}^d$ and

$$dim_{haus}(F) \leq dim_{pack}(F)$$

for any set $F \subseteq \mathbb{R}^d$. Equality holds, for example, for many hyperbolic dynamical systems restricted to their attractors.

One of the main drawbacks in this theory is the fact that HAUSDORFF dimension may not be computable. As is clear from the definition that it becomes computable if it agrees with one of the "measure theoretic" notions of dimension. Since the early 1980's several methods have been introduced to estimate the dimension consistently. The basic assumption which has to be made is the equality of dimension to at least one of the other computable dimensions.

In case this is the correlation dimension the first paper is GRASBERGER and PROCACCIA [22], rigorously put into the framework of regression analysis by DENKER and KELLER [17], [18]. This method is based on estimating consistently the correlation integral

$$C(\varepsilon) = \int \mu(B(\mathbf{x}, \varepsilon))\mu(\mathbf{dx})$$

for some given sequence of radii $\varepsilon_1, \ldots, \varepsilon_m$ by the sample proportion $C_n(\varepsilon)$ of pairs of observations that are no more than ε apart. Here and in the sequel $B(\mathbf{x}, \varepsilon)$ denotes a ball of radius ε centered at $\mathbf{x}$ and μ denotes a probability measure of interest. Then the slope of the least square line through the data pairs $(\log \varepsilon_1, \log C_n(\varepsilon_1)), \ldots, (\log \varepsilon_m, \log C_n(\varepsilon_m))$ is taken as a point estimator for the correlation dimension.

In case of local dimension GUCKENHEIMER [23] has introduced the method of the nearest neighbors, again rigorously examined by CUTLER and DAWSON [13], [14]. This method is based on computing distances $\delta_j(\mathbf{x})$ between the point $\mathbf{x}$ where we want to estimate local dimension and its j-th nearest neighbors. Then, for some chosen m $(m < n)$, the reciprocal of the slope of the least squares line through the data pairs $(\log(1/n), \log \delta_1(\mathbf{x})), \ldots, (\log(m/n), \log \delta_m(\mathbf{x}))$ is taken as a point estimator for the local dimension at $\mathbf{x}$.

The first estimation method for information dimension on the basis of independent observations was developed by CUTLER [10]. It consists of combination of GUCKENHEIMER's method and the averaging over several basepoints and requires three independent samples of observations. Later this method was extended by HAMANN [24] to dependent observations.

KELLER [35] extended the method of GRASBERGER and PROCACCIA [22] to estimate information dimension by introducing some type of outlier analysis. Requiring continuity of a distribution function of $\mu(B(\mathbf{X}, \varepsilon))$ where $\mathbf{X}$ is distributed according to μ and using some known score function J on $[\delta, 1-\delta]$ $(0 < \delta < 1/2)$ such that $\int J(t)dt = 1$, he generalized the method of correlation dimension for estimating information dimension.

In all these cases the mathematical background is well understood. For the GRASSBERGER and PROCACCIA method one has asymptotic normality of each finite dimensional statistics $(C_n(\varepsilon_1), \ldots, C_n(\varepsilon_m))$. CUTLER and DAWSON [14] showed that the log minimum distance, when observations are sampled from measures belonging to a special family of fractal distributions, follows either the normal distribution or the extreme value distribution. CUTLER [10], HAMANN [24] and KELLER [35] proved asymptotic normality of the statistics arising in their methods.

CUTLER [10] reduced the general problem of dimension estimation to the problem for classes of measures (and not of dynamic origin). Here one can start with an independent identically distributed sample, which makes the analysis simple but still meaningful in the stationary case.

Each of two methods for information dimension has its merits from a numerical point of view. However they also have some drawbacks. The method of CUTLER [10] does not equally use three samples while in the method of KELLER [35] it is not clear how to select the score function J in general and what is the impact of the choice of J on the accuracy of the estimation is.

The first part of this thesis solves a problem originating from the work of KELLER [35]. Note that $\mu(B(\mathbf{x}, \varepsilon))$ can always be estimated by $\hat{\mu}(B(\mathbf{x}, \varepsilon))$, where $\hat{\mu}$ is the empirical probability measure of independent identically distributed random variables. It is evident that $\hat{\mu}(B(\mathbf{x}, \varepsilon)) = 0$ if no observation falls into $B(\mathbf{x}, \varepsilon)$ hence $\log \hat{\mu}(B(\mathbf{x}, \varepsilon))$ does not make sense at all. This was pointed by KELLER [35]. One of purposes of this work is to show how such procedure is meaningful if enough data is available to fall into $B(\mathbf{x}, \varepsilon)$.

The second part of the dissertation deals with a different but related problem. In the late 1980's BROSAMLER [8] and SCHATTE [46] independently proved a new type of limit theorems. This type of statements extends the classical central limit theorem to a pathwise version and is therefore called the *almost sure central limit* *theorem* (ASCLT). The first ASCLT for a sequence of *independent identically distributed* (i.i.d.) random variables $X_1, \ldots, X_n$ states that if $EX_1 = 0$, $Var(X_1) = 1$ and $E|X_1|^{2+\delta} < \infty$ for some $\delta > 0$ ($\delta = 1$ in SCHATTE [46]) then

$$\lim_{n\infty} \frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{1} \left\{ \frac{S_k}{\sqrt{k}} \leq x \right\} = \Phi_{(0,1)}(x) \qquad \text{a.s. for any} \quad x, \tag{1}$$

where $\mathbb{1}$ denotes the indicator function and $\Phi_{(0,1)}$ denotes the distribution function of the standard normal random variable. If (1) holds then we say that the sequence $S_n/\sqrt{n}$ satisfies the ASCLT. It should be mentioned that the first version of ASCLT for a kind of martingales has already been stated without proof by LÉVY [40], and a special case of statement (1), namely for $x = 0$, was established by ERDÖS and HUNT [19].

In the 1990's, a lot of theoretical investigations have been done to prove the ASCLT in different situations. First, FISHER [20] and LACEY and PHILIPP [38] proved the ASCLT under finite second moment for X_1. Moreover, LACEY and PHILIPP [38] gave a general condition for the validity of (1) so that a large class of dependent sequences satisfies the ASCLT. Later PELIGRAD and SHAO [44] proved (1) directly for associated, strongly mixing and ρ-mixing sequences under the same conditions that assure the usual central limit theorem.

Statements of type (1) with some non-normal limiting distribution function G are usually called *almost sure* (or pointwise) *limit theorems* (ASLT). The first result in this field belongs to PELIGRAD and RÉVÉSZ [43]. They showed that a weak convergence of properly normalized and centered partial sum of i.i.d. random variables to a limiting α-stable distribution G_α $(0 < \alpha < 2)$ implies the corresponding ASLT. Analogous result was proved by BERKES and DEHLING [5] for the normal limiting distribution. Thus for i.i.d. random variables, almost sure limit theorems are weaker results than corresponding classical limit theorems. Moreover, BERKES, DEHLING and MÓRI [6] provided counterexamples which show that the reverse is not valid. An excellent survey on this topic can be found in BERKES [3] as well as in ATLAGH and WEBER [1].

Recently BERKES and CSÁKI [4] obtained a general result in the almost sure limit theory. They used it to prove almost sure versions of several classical limit theorems. In particular they proved the ASLT for U-statistics under finite second moment of the kernel.

The second part of this thesis is devoted to the ASCLT and ASLT for U-statistics. We show that HOEFFDING's decomposition for U-statistics which plays an important role in deriving their weak limits is still significant in this context. It will be shown that a small modification of the standard technique for proving classical limit theorems for U-statistics allows us to refine and extend the result of BERKES and CSÁKI [4].

The thesis has the following structure.

In Chapter 1 we give the most popular measure dependent notions of dimension and discuss some relations between them. Then we briefly review estimation theory for local and information dimensions. Next we introduce a new estimator for the information dimension. Finally we give some preliminary results which will be used in the next chapter.

In Chapter 2 we establish consistency and asymptotical normality of the estimator introduced in Chapter 1. Then we give a consistent estimator for the variance arising in the central limit theorem. Finally, we prove the multivariate central limit theorem for a vector of statistics whose components are the introduced estimator constructed for a given finite sequence of radii.

In Chapter 3 we apply our theory to some fractal distributions on the unit cube. We construct confidence intervals for the information dimension when underlying probability measures are the Cantor distribution in $\mathbb{R}^2$ and the generalized Cantor distribution in $\mathbb{R}^3$.

In Chapter 4 we prove the ASCLT for non-degenerate U-statistics of a sequence of strongly mixing and absolutely regular random variables. Then we relax the moment condition of BERKES and CSÁKI's ASLT for U-statistics of i.i.d. random variables. Finally, we prove the ASLT with a stable limiting distribution for non-degenerate U-statistics of i.i.d. random variables $X_1, \dots, X_n$ (for the i.i.d. case see HOLZMANN, KOCH and MIN [32]).

We will index definitions, theorems and lemmas in the following way: the first number will refer to the chapter and the second number will refer to their order in the chapter. The same holds for numbering of equations and formulas.

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to my supervisor, Prof. Dr. Manfred Denker, for proposing the topic of my dissertation and his excellent supervision. I am also grateful to my coadvisor Prof. Dr. Susanne Koch for clarifying discussions about almost sure limit theorems, and agreeing to act as a Korreferentin.

I would like to thank all people from the Institute for Mathematical Stochastics for creating a warm atmosphere which above all supports scientific research. In particular, I gained new insights from discussions with Prof. Dr. Axel Munk, Hajo Holzmann, Steffi Burgemeister and Janis Valeinis.

My special thanks go to Prof. Dr. Urlich Parlitz for introducing me to the software TSTOOL which I used for my numerical simulations.

I would like to thank my family for supporting me with great patience and loving care throughout my life.

Finally, I acknowledge the financial and scientific support by the DFG's Graduiertenkolleg "Strömungsinstabilitäten und Turbulenz".

Chapter 1

Preliminaries on Dimension Theory

1.1 Measure dependent definitions of fractal dimension

In this section we give three different definitions of dimension for probability measures and some relations between them.
Let μ denote the probability measure defined on the Borel sets of $\mathbb{R}^d$ and $\mathbf{S} \subseteq \mathbb{R}^d$ denote its support. Further, let $B(\mathbf{x}, \varepsilon)$ be a closed ball of radius ε centered at $\mathbf{x} \in \mathbf{S}$.

Definition 1.1. *The spatial correlation integral $C(\varepsilon)$ is defined by*

$$C(\varepsilon) = \int_{\mathbf{S}} \mu(B(\mathbf{x}, \varepsilon))\mu(\mathbf{dx})$$

and the correlation dimension ν_μ of a probability measure μ is defined by

$$\nu_\mu = \lim_{\varepsilon \to 0} \frac{\log C(\varepsilon)}{\log \varepsilon} = \lim_{\varepsilon \to 0} \frac{1}{\log \varepsilon} \log \int_{\mathbf{S}} \mu(B(\mathbf{x}, \varepsilon))\mu(\mathbf{dx}) \ . \tag{1.1}$$

It is obvious that $C(\varepsilon) = E\mu((B(\mathbf{X}, \varepsilon)))$, where $\mathbf{X}$ is distributed according to μ. Thus, the spatial correlation integral measures the concentration of μ and describes the mean volume of a ball of radius ε.

The correlation dimension which is also often called the correlation exponent was initially introduced and first numerically studied by GRASSBERGER and PROCACCIA [22]. They found that the spatial correlation integral is proportional to ε^{ν_μ} for small ε. Moreover, they noticed that in many cases the correlation exponent ν_μ agrees with the dimension of the support $\mathbf{S}$ of a probability measure μ and so they suggested to estimate it.
It should be noted that the correlation dimension is the most popular dimension for experimentalists since it is relatively easy to estimate it. The natural choice for an estimator of the spatial correlation integral $C(\varepsilon)$ is the sample correlation integral $C_n(\varepsilon)$ which is defined as follows

$$C_n(\varepsilon) = \frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \mathbb{1}\{\|\mathbf{X}_i - \mathbf{X}_j\| \le \varepsilon\},$$

where $\mathbf{X}_1, \dots, \mathbf{X}_n$ is a sample drawn from a distribution μ and $\|.\|$ is some norm in Euclidean space $\mathbb{R}^d$. So, we can easily see that the sample correlation integral $C_n(\varepsilon)$ is, in fact, a U-statistic of degree 2 with kernel $\mathbb{1}\{\|\mathbf{x} - \mathbf{y}\| \le \varepsilon\}$.
The first rigorous results on dimension estimation of probability measures, namely for the correlation dimension, were obtained by DENKER and KELLER [18]. They studied asymptotical properties of the spatial correlation integral using the theory of U-statistics and proved its consistency and asymptotical normality.
However, it turns out that it is more important to consider the local dimension than the correlation dimension since in many examples the local dimension reflects a complexity of the support of a probability measure μ much better than the correlation dimension. We will illustrate this after the definition of the local dimension and a comment following it.

Definition 1.2. *The local (or pointwise) dimension $\alpha_\mu(\mathbf{x})$ of a probability measure μ at a point $\mathbf{x} \in \mathbf{S}$ is defined by*

$$\alpha_\mu(\mathbf{x}) = \lim_{\varepsilon \to 0} \frac{\log \mu(B(\mathbf{x}, \varepsilon))}{\log \varepsilon}. \tag{1.2}$$

It is not difficult to show that if $\mathbf{x} \in \mathbf{S}$ then $\alpha_\mu(\mathbf{x}) = 0$ μ-a.s. for all discrete distributions μ in $\mathbb{R}^d$ and $\alpha_\mu(\mathbf{x}) = d$ μ-a.s. for all absolutely continuous distributions μ in $\mathbb{R}^d$ (see e.g. CUTLER [12]).

Now we are ready to illustrate an example where the information dimension is preferred to the correlation dimension. For this purpose we cite CUTLER's example. Consider the absolutely continuous measure μ_γ on $(0,1)$ with the density function $f(x) = \gamma x^{\gamma-1}$, where $\gamma > 0$. Since μ_γ is absolutely continuous $\alpha_{\mu_\gamma}(x) = 1$ μ_γ-a.s., but it can be shown that $\nu_{\mu_\gamma} = 2\gamma$ for $0 < \gamma < 1/2$ (see CUTLER [12] for more information).
This example is specifically important since it shows that an observation of a fractional correlation dimension does not imply that a measure can be supported on a set of fractional dimension. By contrast, if a measure μ has a constant fractional local dimension α μ-a.s. then the support $\mathbf{S}$ of μ must be a set of fractional dimension α (see OTT, WITHERS and YORKE [42] for more details).
In many examples, the local dimension coincides with the information dimension. In general, there is a more deep connection between them. If $\alpha_\mu(\mathbf{x})$ is constant μ-a.s. and the support $\mathbf{S}$ of μ is a bounded subset of $\mathbb{R}^d$ then the local dimension coincides with the information dimension (see CUTLER [12]).

Definition 1.3. *The information dimension σ_μ of a probability measure μ is defined by*

$$\sigma_\mu = \lim_{\varepsilon\to 0} \frac{1}{\log \varepsilon} \int \log \mu(B(\mathbf{x},\varepsilon)) d\mu(x). \tag{1.3}$$

There is a simple relation between the correlation and information dimensions of a probability measure μ, namely

$$\nu_\mu \leq \sigma_\mu. \tag{1.4}$$

This inequality follows from (1.1), (1.3) and Jenssen's inequality. It should be mentioned that a strict inequality in (1.4) may occur (see CUTLER [12]).
Since smooth ergodic dynamical systems naturally give rise to an exact dimensional invariant measure which means that $\alpha_\mu(\mathbf{x})$ is constant μ-a.s., and they usually have bounded attracting sets, in this work, we will consider only those probability measures whose local dimension $\alpha_\mu(\mathbf{x})$ is equal to their information dimension σ_μ.

1.2 Estimation methods for the local dimension

In this section, we briefly describe two estimation methods for the local dimension, namely the least square method and the nearest neighbor method. These methods will be generalized in the next two sections devoted to the information dimension. We start the section with describing the least square method.

Equation (1.2) suggests a trivial method of estimating the local dimension $\alpha_\mu(\mathbf{x})$ at $\mathbf{x}$. First, one has to obtain an appropriate estimator $\hat{\mu}(B(\mathbf{x},\varepsilon))$ of $\mu(B(\mathbf{x},\varepsilon))$ for small ε. Since a measure μ typically is not given in analytical form but instead by a finite sample $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ of d-dimensional vectors drawn from it, the natural choice of estimators $\hat{\mu}(B(\mathbf{x},\varepsilon_k))$ is a sample proportion of observations falling within a distance ε to the point $\mathbf{x}$, i.e.

$$\hat{\mu}(B(\mathbf{x},\varepsilon)) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{\|\mathbf{X}_i - \mathbf{x}\| \leq \varepsilon\} \ . \tag{1.5}$$

It should be noted that we also do not have any other additional information about μ. Secondly, one has to take the ratio $\log\hat{\mu}(B(\mathbf{x},\varepsilon))/\log\varepsilon$ as an estimator for $\alpha_\mu(\mathbf{x})$.

This method usually does not work well even if we get a very accurate estimator $\hat{\mu}(B(\mathbf{x},\varepsilon))$ of $\mu(B(\mathbf{x},\varepsilon))$. The reason is that $\log\mu(B(\mathbf{x},\varepsilon))/\log\varepsilon$ typically converges to $\alpha_\mu(\mathbf{x})$ very slowly.

The solution here is to detect a linear relationship of the type $\log\mu(B(\mathbf{x},\varepsilon)) \approx C(x) + \alpha_\mu(\mathbf{x})\log(\varepsilon)$ for a sequence of radii $0 < \varepsilon_1 < \ldots < \varepsilon_m$ on the basis of the observations. In practice, one has to obtain estimators $\hat{\mu}(B(\mathbf{x},\varepsilon_k))$ $(k = 1,\ldots,m)$ for a sequence of radii $0 < \varepsilon_1 < \ldots < \varepsilon_m$ and take the slope of least square line through the data pairs $(\log\hat{\mu}(B(\mathbf{x},\varepsilon_1)), \log\varepsilon_1), \ldots, (\log\hat{\mu}(B(\mathbf{x},\varepsilon_m)), \log\varepsilon_m)$ as an estimator for the local dimension $\alpha_\mu(\mathbf{x})$.

The main advantage of using least squares analysis is that it eliminates the intercept effect over the employed ε-range. The least squares analysis also allows us to examine the fit of the data pairs $(\log\hat{\mu}(B(\mathbf{x},\varepsilon_1)), \log\varepsilon_1), \ldots, (\log\hat{\mu}(B(\mathbf{x},\varepsilon_m)), \log\varepsilon_m)$ to a straight line.

However, the error of estimators obtained from ordinary least squares will generally be wrong since the estimators $\hat{\mu}(B(\mathbf{x},\varepsilon_j))$, $j = 1,\ldots,k$ are always correlated with unequal variances. If a covariance matrix of a vector

$(\log \hat{\mu}(B(\mathbf{x}, \varepsilon_1)), \ldots, \log \hat{\mu}(B(\mathbf{x}, \varepsilon_m)))$ is available and we can consistently estimate its components from the data, then it is possible to perform generalized least squares analysis. As far as we know, there are no results for the local dimension which show a preference of using the generalized least square analysis instead of the least squares analysis.

The second approach of estimating the local dimension is the nearest neighbor method which is an opposite method to the least square method. In this method radii constitute statistics unlike the least square method, where they are fixed. In the language of regression analysis it means that the dependent and independent variables are reversed. As a result, we will expect that the slope of the least squares line in this method is an estimator of $1/\alpha_\mu(\mathbf{x})$.

Now let us describe this method. Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be an independent sample from the distribution μ. First calculate the distances $\delta_j(\mathbf{x}) = \|\mathbf{X}_j - \mathbf{x}\|$ for each $j = 1, 2, ..., n$ and write them in the ascending order. We do it by computing the order statistics $\delta_{1:n}(\mathbf{x}), \delta_{2:n}(\mathbf{x}), ..., \delta_{n:n}(\mathbf{x})$. Then we perform the least squares analysis of $\log \delta_{j:n}(\mathbf{x})$ vs. $\log(j/n)$, $j = 1, ..., k$ for some chosen integer k and take reciprocal of the slope of the resulting least square line as an estimate of $\alpha_\mu(\mathbf{x})$. Note that $\delta_{1:n}(\mathbf{x}), \delta_{2:n}(\mathbf{x}), ..., \delta_{n:n}(\mathbf{x})$ are in fact the distances from $\mathbf{x}$ to its k nearest neighbors in the sample. This method was originally proposed and numerically studied by Guckenheimer [23].

The validity of the nearest neighbor method procedure has been shown by Cutler and Dawson [13] and Cutler [11]. They showed that if the actual pointwise dimension $\alpha_\mu(\mathbf{x})$ exists then

$$\lim_{n\to\infty} \frac{\log \delta_{1:n}(\mathbf{x})}{\log(1/n)} = \frac{1}{\alpha_\mu(\mathbf{x})} \quad \text{w.p. } 1.$$

Moreover, the asymptotic behavior of the statistic $\log \delta_{1:n}(\mathbf{x})/\log(1/n)$ has also been investigated by them.

As noted previously, we usually deal with probability measures whose pointwise dimension is constant μ-a.s. and coincides with the information dimension. So, it is very natural to develop statistical methods for estimating the information dimension because an application of these two last methods is always accompanied with a local effect of point $\mathbf{x} \in \mathbf{S}$.

1.3 Estimation methods for the information dimension

In this section we outline two estimation methods for the information dimension. The first method for estimating the information dimension has been proposed and investigated by CUTLER [10]. This method is a generalization of the nearest neighbor method for the local dimension. We start the section with it.

Let $\mathcal{B} = \{\mathbf{X}_1, \dots, \mathbf{X}_k\}$, $\mathcal{S}_1 = \{\mathbf{Y}_{1,1}, \dots, \mathbf{Y}_{1,n}\}$ and $\mathcal{S}_2 = \{\mathbf{Y}_{2,1}, \dots, \mathbf{Y}_{2,m\cdot n}\}$ be three independent samples from a distribution μ of interest. Further, the first sample $\mathcal{B}$ will be called the basepoint sample. Then the minimum distances from each basepoint $\mathbf{X}_j$ to each of the two samples $\mathbf{S}_1$ and $\mathbf{S}_2$ are computed as follows

$$\delta_{1,n}(\mathbf{X}_j) = \min_{1\le i\le n} \|\mathbf{X}_j - \mathbf{Y}_{1,i}\| \quad \text{and} \quad \delta_{2,m\cdot n}(\mathbf{X}_j) = \min_{1\le i\le m\cdot n} \|\mathbf{X}_j - \mathbf{Y}_{2,i}\| \ .$$

Furthermore, for each basepoint $\mathbf{X}_j$, the statistic

$$R_{m,n}(\mathbf{X}_j) = \frac{1}{\log m} \log\left(\frac{\delta_{1,n}(\mathbf{X}_j)}{\delta_{2,m\cdot n}(\mathbf{X}_j)}\right)$$

is computed and the reciprocal of the sample mean of these statistics $R_{m,n}(\mathbf{X}_j),\ j = 1, \dots, k$ is taken as an estimator for the σ_μ, i.e.

$$\left(\overline{R}_{m,n}\right)^{-1} = \left(\frac{1}{k}\sum_{j=1}^{k} R_{m,n}(\mathbf{X}_j)\right)^{-1} .$$

CUTLER [10] established asymptotical normality of the statistic $\overline{R}_{m,n}$ and constructed a confidence interval for $1/\sigma_\mu$ which can easily be transformed into a confidence interval for the information dimension σ_μ. Numerical results based on this method were also provided.

Since data from dynamical systems are correlated and the above theory is based on independent observations, it was desirable to extend this method for dependent observations. This was done by HAMANN [24] for a stationary sequence of random vectors which satisfies ψ-mixing condition.

The main disadvantage of this method, from our point of view, is that we do not utilize the information which is contained in the collective sample $\mathbf{X}_1, \dots, \mathbf{X}_k, \mathbf{Y}_{1,1}, \dots, \mathbf{Y}_{1,n}, \mathbf{Y}_{2,1}, \dots, \mathbf{Y}_{2,m\cdot n}$.

The second estimator for the information dimension has been proposed and investigated by Keller [35]. Note that the expression (1.3) in Definition 1.3 differs from the expression (1.1) in Definition 1.1 in the order of integral and logarithm which makes it difficult for finding a good estimator for the information dimension. Keller avoided this problem by finding an alternative definition for the information dimension which has the same order of integral and logarithm as in Definition 1.1 for the correlation dimension. The following theorem was proved by him.

Theorem 1.1. *Let μ be a dimension regular probability measure on $\mathbb{R}^d$ with bounded support, and F_ε denote the distribution function of $\mu(B(\mathbf{X}, \varepsilon))$, where $\mathbf{X}$ has a distribution μ. Suppose that F_ε is continuous. Consider a continuous functional $J : [0, 1] \to [0, \infty)$ with $\int_0^1 J(t)dt = 1$ and $J(t) = 0$ if $t \notin (\delta, 1 - \delta)$ for some $\delta > 0$. Then*

$$\sigma_\mu = \lim_{\varepsilon \to 0} \frac{1}{\log \varepsilon} \log \int \mu(B(\mathbf{x}, \varepsilon)) J(F_\varepsilon(\mu(B(\mathbf{x}, \varepsilon))) \mu(\mathbf{dx})) \ . \tag{1.6}$$

Denote the argument of the logarithm in expression (1.6) as $\overline{C}(\varepsilon)$, i.e.

$$\overline{C}(\varepsilon) = \int \mu(B(\mathbf{x}, \varepsilon)) J(F_\varepsilon(\mu(B(\mathbf{x}, \varepsilon))) \mu(\mathbf{dx})) \ .$$

Further, define the following location parameter $T(F_\varepsilon)$ of F_ε by

$$T(F_\varepsilon) = \int_0^1 F_\varepsilon^{-1}(s) J(s) ds \ . \tag{1.7}$$

Note that, if F_ε is continuous then $\overline{C}(\varepsilon) = T(F_\varepsilon)$ and hence, the problem of estimation of the information dimension reduces to a problem of estimating the statistical functional $T(F_\varepsilon)$.

From the theory of statistical functionals, a natural choice of an estimator for $T(F_\varepsilon)$ is $T(G_n)$, where G_n is the empirical distribution function of the sample proportions $\hat{\mu}(B(\mathbf{X}_1, \varepsilon)), \ldots, \hat{\mu}(B(\mathbf{X}_n, \varepsilon))$ which, in turn, are approximations of unobserved random variables $\mu(B(\mathbf{X}_1, \varepsilon)), \ldots, \mu(B(\mathbf{X}_n, \varepsilon))$. Now $T(G_n)$ can be easily written as a mixture between a U-statistic and a L-statistic and its asymptotic behavior can be studied through the well advanced theory of U- and L-statistics (see Keller [35] for more details).

The asymptotical normality of $\sqrt{n}\,(T(G_n) - T(F_\varepsilon))$ has been proven by Keller [35] for independent samples as well as for random vectors which are absolutely regular with mixing coefficients $\beta(n)$ decreasing at a suitable polynomial rate. He also provided numerical results for data produced by a cubic full-unimodal map and data from a Henon system.

Keller [35] also pointed out that the sample average of $\log \hat{\mu}(B(\mathbf{X}_1, \varepsilon)), \ldots, \log \hat{\mu}(B(\mathbf{X}_n, \varepsilon))$ is not always a meaningful estimator for $E \log \mu(B(\mathbf{X}, \varepsilon))$. However, if sufficient data is available, then from our point of view, this averaging procedure deserves attention and we will deal with it in the next section.

1.4 Statistical functionals of unobservables

The object of this section is to introduce a new estimator for the information dimension whose asymptotical behavior will be studied in Chapter 2. First we would like to discuss the main problem in estimating the information dimension. Let $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ be independent $\mathbb{R}^d$-valued observations from distribution μ. In order to estimate the information dimension σ_μ accurate, without any additional assumptions, we have to properly estimate $E \log \mu(B(\mathbf{X}_1, \varepsilon))$. However this can not be done in the standard way since we do not observe $\mu(B(\mathbf{X}_1, \varepsilon)), \ldots, \mu(B(\mathbf{X}_n, \varepsilon))$. This circumstance provides an explanation for us about the existence of two completely different methods developed by Cutler [10] and Keller [35].

Now we present a third method of solving this problem which can be generalized for other statistical problems and this generalization will be discussed at the end of this section. As noted before, the standard estimator for $\mu(B(\mathbf{x}, \varepsilon))$ is the sample proportion $\hat{\mu}(B(\mathbf{x}, \varepsilon))$ which was defined in (1.5). Therefore we replace unobservable sample $\log \mu(B(\mathbf{X}_1, \varepsilon)), \ldots, \log \mu(B(\mathbf{X}_n, \varepsilon))$ in the standard estimator sample mean for $E \log \mu(B(\mathbf{X}_1, \varepsilon))$ by observable $\log \hat{\mu}(B(\mathbf{X}_1, \varepsilon)), \ldots, \log \hat{\mu}(B(\mathbf{X}_n, \varepsilon))$ assuming that $\hat{\mu}(B(\mathbf{X}_j, \varepsilon)) > 0$ for all $j = 1, \ldots, n$. The last assumption holds, for example, if $\varepsilon > \max_j \delta(X_j)$, where

$\delta(X_j) = \min_{i:i\neq j} \|X_j - X_i\|$. Thus we obtain the following statistic

$$T_n(\varepsilon) = \frac{1}{n}\sum_{j=1}^{n} \log\left(\frac{1}{n-1}\sum_{i=1\ ;\ i\neq j}^{n} \mathbb{1}\{\|\mathbf{X}_i - \mathbf{X}_j\| \le \varepsilon\}\right) \tag{1.8}$$

as an estimator for $E\log\mu(B(\mathbf{X}_1,\varepsilon))$.

The statistic $T_n(\varepsilon)$ can be considered as a U-statistic whose kernel has a U-statistical structure. The advantage of this estimator is that it makes use of the information contained in the whole sample as much as possible. We also hope that its relative simplicity will make it popular for experimentalists. It should be mentioned that this estimator requires huge calculations, but that should not make any difficulty in the coming future.

In order to avoid an effect of the intercept, the least squares analysis of $T_n(\varepsilon_j)$ vs. $\log\varepsilon_j$ for some appropriate sequence of radii $0 < \varepsilon_1 < \varepsilon_2 < ... < \varepsilon_k$ should be performed. Then the slope of the least squares line can be considered as an estimator of σ_μ. Numerical results of Chapter 3 show that the errors of disregarding the dependence of $T_n(\varepsilon_j), j = 1,\dots,k$ in the least square analysis are usually very small.

Another heuristic justification for this approach can be made by observing the connection between the local and information dimensions. Recall that we only deal with probability measures whose information dimension coincides with their local dimension, i.e. $\sigma_\mu = \alpha_\mu(\mathbf{x})$ $\mu-$a.s. Assume now for a moment that "lim" and "$\int$" in (1.3) can be interchanged. Then it follows that $\sigma_\mu = E\alpha_\mu(\mathbf{X}_1)$. Choose some small ε and compute estimators of the local dimension $\hat{\alpha}_\mu(\mathbf{X}_j)$ for $j = 1,\dots,n$ based on the sample $\mathbf{X}_1,\dots,\mathbf{X}_{j-1},\mathbf{X}_{j+1},...,\mathbf{X}_n$. Since $\sigma_\mu = E\alpha_\mu(\mathbf{X}_1)$, the standard estimator $\hat{\sigma}_\mu$ for the information dimension σ_μ will be a sample mean of $\hat{\alpha}_\mu(\mathbf{X}_1),\dots,\hat{\alpha}_\mu(\mathbf{X}_n)$. Thus we find out that $\hat{\sigma}_\mu = T_n(\varepsilon)/\log\varepsilon$.

One of the main goals of this work is to investigate the asymptotic behavior of a slightly more general form of the statistic $T_n(\varepsilon)$ which is denoted by T_n and is given by

$$T_n = \begin{cases} \frac{1}{n}\sum\limits_{j=1}^{n}\log\left(\frac{1}{n-1}\sum\limits_{\substack{i=1\\ i\neq j}}^{n} h(X_i,X_j)\right) & \text{if } \sum\limits_{\substack{i=1\\ i\neq j}}^{n} h(X_i,X_j) > 0 \text{ for } j = 1,\dots,n \\ -\infty & \text{otherwise} \end{cases} \tag{1.9}$$

where $h : \mathbb{R}^{2d} \to \mathbb{R}$ is some measurable symmetric function. If $h(\mathbf{x}, \mathbf{y}) = \mathbb{1}\{\|\mathbf{x} - \mathbf{y}\| \leq \varepsilon\}$ then $T_n = T_n(\varepsilon)$.
Note that the statistic T_n can also be rewritten in a more general form

$$\tilde{T}_n = \frac{1}{n} \sum_{j=1}^{n} g\left(\frac{1}{n-1} \sum_{i=1:i \neq j}^{n} h(\mathbf{X}_i, \mathbf{X}_j)\right)$$

and considered as an estimator for the superposition of statistical functionals

$$\tilde{T} = \int g\left(\int h(\mathbf{x}, \mathbf{y}) dF(\mathbf{y})\right) dF(\mathbf{x})$$

assuming that $\tilde{T}_n$ and $\tilde{T}$ are well defined.
Such estimators naturally appear when statistics are based on an unobservable sample which has an observable approximation. The simplest example of the statistic $\tilde{T}_n$ is a wide class of statistics which can be written as U-statistics of degree 2, for instance, the sample correlation integral. Another example is the statistic $T_n(\varepsilon)$ defined in (1.8). The special form of the statistic $\tilde{T}_n$ also appears in the theory of nonparametric statistics in factorial designs (see Brunner and Denker [9]). It would be desirable to develop a theory for statistical functionals of unobservables analogously to the theory of U-statistics.
At the end of this section we would like to explain what we mean by "unobservables". First let us give the notion of a random variable by Halmos [25]:
"A random variable is a quantity whose values are determined by chance. ... Accordingly a random variable is a function: a function whose numerical values are determined by chance. This means in other words that a random variable is a function attached to an experiment-once the experiment has been performed the value of function is known. ..."
Thus, an "unobservable" is a random variable whose value is not known after performing an experiment.

1.5 Auxilary results

In this section we introduce U-statistics and give some auxilary results which will be used in the next chapters. We begin the section with U-statistics.

Many important statistical functionals may be represented as

$$\theta(F) = \int \cdots \int h(x_1, x_2, \ldots, x_m) dF(x_1) \cdot \ldots \cdot dF(x_m) \ , \tag{1.10}$$

where $m \in \mathbb{N}$, h is some measurable function, called the kernel and F is a distribution function from some given set of distribution functions. A minimal number $m \in \mathbb{N}$ is called a rank of a statistical functional $\theta(F)$ if there exists a kernel h with m arguments such that (1.10) holds . Without loss of generality, we can assume that h is symmetric. If it would be not the case then the following transformation

$$\frac{1}{m!} \sum_{1 \le i_1 \ne i_2 \ne \ldots \ne i_m \le n} h(x_{i_1}, \ldots, x_{i_m})$$

will give us a symmetric kernel for $\theta(F)$.
Statistical functionals of type (1.10) are called regular or parametric functionals. The simplest examples of regular functionals are the mean and the variance with the following kernels $h(x) = x$ and $h(x, y) = \frac{1}{2}(x - y)^2$, respectively.
Hoeffding [30], partly influenced by the early work of Halmos [26], introduced U-statistics as unbiased and asymptotical normal estimators for regular functionals $\theta(F)$. They also possess good consistency and optimality properties. In fact, they have a minimal variance among all unbiased estimators. And this property of optimality makes them a popular object for theoretical investigations of statisticians and probabilists.
Let $X_1, \ldots, X_n$ be i.i.d. random variables from some distribution F.

Definition 1.4. *A U-statistic with kernel h of degree m based on a sample $X_1, \ldots, X_n$ is a statistic*

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

It should be mentioned that U-statistics are closely connected to von Mises' functionals (see von Mises [50]).
The main achievement of Hoeffding [30] is that he developed an analytical method to investigate asymptotic properties of U-statistics. It is based on his decomposition theorem for U-statistics. Before stating it here, we need some additional notation.

We introduce the following auxilary symmetric functions

$$\begin{aligned}\tilde{h}_1(x_1) &= \int \cdots \int h(x_1,\ldots,x_m)dF(x_2)\cdots dF(x_m)\ , && (1.11)\\ \tilde{h}_2(x_1,x_2) &= \int \cdots \int h(x_1,x_2,\ldots,x_m)dF(x_3)\cdots dF(x_m)\ , && (1.12)\\ \vdots &= \vdots \\ \tilde{h}_m(x_1,\ldots,x_m) &= h(x_1,\ldots,x_m)\ . && (1.13)\end{aligned}$$

Further, define the following functions for $1 \le c \le m$

$$h_c(x_1,\ldots,x_c) = (-1)^c\theta(F) + \sum_{d=1}^{c}(-1)^{c-d} \sum_{1\le i_1<\ldots<i_d\le c} h_d(x_{i_1},\ldots,x_{i_d})\ . \quad (1.14)$$

It is not difficult to see that $h_c,\ c=1,\ldots,m$ are also symmetric and moreover, degenerate, i.e. the integral over one variable with respect to the distribution of any random variable X_i with $i \in \{1,\ldots,n\}$ vanishes (see for example Denker [16] or Koroljuk and Borovskikh [37]). A number r is called a rank of a U-statistic with kernel h if $h_1 \equiv \ldots \equiv h_{r-1} \equiv 0$ and $h_r \ne 0$ a.s.. It is obvious, that r takes values from $1,\ldots,m$. If $r=1$ then a U-statistic is called non-degenerate and otherwise, degenerate.
Now we are ready to state Hoeffding's decomposition theorem for U-statistics. The proof of this theorem can be found in any classical textbook on U-statistics as well as in Hoeffidng's original article [30].

Theorem 1.2. *If r is a rank of a statistic U_n then the following decomposition holds*

$$U_n - \theta(F) = \sum_{c=r}^{m}\binom{m}{c}U_{nc}\ , \quad (1.15)$$

where U_{nc} are U-statistics with degenerate kernels h_c, i.e.

$$U_{nc} = \binom{n}{c}^{-1}\sum_{1\le i_1<\ldots<i_c\le n} h_c(X_{i_1},\ldots,X_{i_c})\ . \quad (1.16)$$

The next theorem gives upper bounds for the second moments of degenerate U-statistics U_{nc} and shows that the first term in Hoeffding's decomposition (1.15) determines a limiting behavior of U_n.

Theorem 1.3. *If*

$$Eh^2(X_1, \dots, X_m) < \infty \tag{1.17}$$

then the following moment inequality holds for $1 \le c \le m$

$$EU_{nc}^2 \le K_1 n^c , \tag{1.18}$$

where K_1 *is some absolute constant depending on* $Eh^2(X_1, \dots, X_m)$*.*
Moreover, if r *is the rank of* U_n *then*

$$E\left(\binom{m}{r} U_{nr} - (U_n - \theta(F))\right)^2 \le K_2 n^{-r-1} , \tag{1.19}$$

where K_2 *is some absolute constant depending on* $Eh^2(X_1, \dots, X_m)$*.*

Hoeffding [30] originally considered a non-degenerate U-statistic and proved its asymptotical normality. But it should be mentioned that asymptotic distribution of a degenerate U-statistic completely differs from the normal distribution. In fact, if $r \ge 2$ is a rank of U_n then $n^{r/2}U_n$ converges weakly to a multiple Wiener integral whenever (1.17) holds (see Denker [16]).
Now we give Hoeffding's central limit theorem for non-degenerate U-statistics.

Theorem 1.4. *If* (1.17) *holds and*

$$\sigma^2 = Eh_1^2(X_1) > 0$$

then $\sqrt{n}(m\sigma)^{-1}(U_n - \theta(F))$ *is asymptotically normal with mean* 0 *and variance* 1*.*

Many limit theorems for sums of i.i.d. random variables have their analog for U-statistics. For example, the strong law of large numbers for U-statistics was established by Hoeffding [31] and Berk [2].

Theorem 1.5. *If*

$$E|h(X_1, \dots, X_m)| < \infty$$

then $U_n \to \theta(F)$ *a.s.*

The rest of this section is devoted to obtaining preliminary results for Chapter 2 and we start with the definition of complete convergence.

Definition 1.5. *A sequence* $(\xi_n)_{n=1}^{\infty}$ *of random variables is said to converge completely to* 0 *if for all* $\varepsilon > 0$*,*

$$\sum_{n=1}^{\infty} P(|\xi_n| > \varepsilon) < \infty.$$

The concept of complete convergence was introduced by Hsu and Robbins [33]. It is not difficult to see that complete convergence is stronger than almost sure convergence. It is also one of the main tools for proving an almost sure convergence of a sequence of random variables.

Now we are going to give another representation of the statistic T_n defined in (1.9) using the auxilary functions (1.11)–(1.14). For convenience, we rewrite them for a function h with two arguments.

$$\tilde{h}_1(\mathbf{x}) = E(h(\mathbf{X}_1, \mathbf{X}_2)/\mathbf{X}_1 = \mathbf{x}) , \tag{1.20}$$

$$h_1(\mathbf{x}) = \tilde{h}_1(\mathbf{x}) - Eh(\mathbf{X}_1, \mathbf{X}_2) , \tag{1.21}$$

$$h_2(\mathbf{x}, \mathbf{y}) = h(\mathbf{x}, \mathbf{y}) - Eh(\mathbf{X}_1, \mathbf{X}_2) - h_1(\mathbf{x}) - h_1(\mathbf{y}) . \tag{1.22}$$

Recall that functions h_1 and h_2 are degenerate with respect to the distribution of $\mathbf{X}_i,\ i = 1, \ldots, n$.

Now using (1.20) – (1.22), we find out that (1.9) is equivalent to

$$T_n = \frac{1}{n}\sum_{j=1}^{n} \log\left(\tilde{h}_1(\mathbf{X}_j) + \frac{1}{n-1}\sum_{i=1:i\neq j}^{n} h_1(\mathbf{X}_i) + \frac{1}{n-1}\sum_{i=1:i\neq j}^{n} h_2(\mathbf{X}_i, \mathbf{X}_j)\right)$$

whenever $\liminf_n T_n > -\infty$ μ-a.s.

Here and in the sequel, we will make use of the following notation $\eta_{j,n}$ for $j = 1, \ldots, n$ which denote random variables defined by

$$\eta_{j,n} = \frac{1}{n-1}\sum_{i=1:i\neq j}^{n} h_1(\mathbf{X}_i) + \frac{1}{n-1}\sum_{i=1:i\neq j}^{n} h_2(\mathbf{X}_i, \mathbf{X}_j) . \tag{1.23}$$

Now the statistic T_n can be rewritten in the following form

$$T_n = \frac{1}{n}\sum_{j=1}^{n} \log\left(\tilde{h}_1(\mathbf{X}_j) + \eta_{j,n}\right) . \tag{1.24}$$

The representation (1.24) will be used in Chapter 2 and plays a major role in investigating the asymptotic behavior of T_n.

The following lemma gives us some important properties of the sequence $\eta_{j,n}$ for fixed $j \in \{1, \ldots, n\}$.

Lemma 1.1. *Let $\eta_{j,n}$ for $j = 1, \ldots, n$ and $n \in \mathbb{N}$ be random variables defined in (1.23). Then $\eta_{j,n}$'s are identically distributed w.r. to j for every fixed n. Moreover, if $Eh^4(\mathbf{X}_1, \mathbf{X}_2) < \infty$ then for every fixed j*

$$E\eta_{j,n}^4 = O\left(\frac{1}{n^2}\right) \tag{1.25}$$

and consequently,

$$\eta_{j,n} \to 0 \quad \textit{completely as} \quad n \to \infty. \tag{1.26}$$

Proof. The first statement is obvious. The statement (1.26) follows from the relation (1.25) and from Definition 1.5.

Thus, it suffices to prove the relation (1.25). By c_r-inequality, it follows

$$\begin{aligned} E\eta_{j,n}^4 &= E\left(\frac{1}{n-1}\sum_{i=1:i\neq j}^{n} h_1(\mathbf{X}_i) + \frac{1}{n-1}\sum_{i=1:i\neq j}^{n} h_2(\mathbf{X}_i, \mathbf{X}_j)\right)^4 \\ &\leq 2^3 E\left(\frac{1}{n-1}\sum_{i=1:i\neq j}^{n} h_1(\mathbf{X}_i)\right)^4 + 2^3 E\left(\frac{1}{n-1}\sum_{i=1:i\neq j}^{n} h_2(\mathbf{X}_i, \mathbf{X}_j)\right)^4 \\ &= 8W_{r1} + 8W_{r2}\ . \end{aligned} \tag{1.27}$$

Consider W_{r1}. Using the degeneracy of h_1 and the assumption of the lemma, we have

$$\begin{aligned} W_{r1} &= \frac{1}{(n-1)^4}\sum_{i=1:i\neq j}^{n} Eh_1^4(\mathbf{X}_i) + \frac{3}{(n-1)^4}\sum_{i=1:i\neq j}^{n}\sum_{k=1:k\neq j}^{n} Eh_1^2(\mathbf{X}_i)Eh_1^2(\mathbf{X}_k) \\ &= O\left(\frac{1}{n^2}\right)\ . \end{aligned} \tag{1.28}$$

A similar argument yields the next bound for W_{r2}

$$\begin{aligned} W_{r2} &= \frac{1}{(n-1)^4}\sum_{i=1:i\neq j}^{n} Eh_2^4(\mathbf{X}_i, \mathbf{X}_j) \\ &+ \frac{3}{(n-1)^4}\sum_{i=1:i\neq j}^{n}\sum_{k=1:k\neq i:k\neq j}^{n} Eh_2^2(\mathbf{X}_i, \mathbf{X}_j)Eh_2^2(\mathbf{X}_k, \mathbf{X}_j) \\ &= O\left(\frac{1}{n^2}\right). \end{aligned} \tag{1.29}$$

Combining relations (1.27)–(1.29), we obtain the relation (1.25). □

Remark 1.1. *Statement (1.26) can be proved under the condition $Eh^2(\mathbf{X}_1, \mathbf{X}_2) < \infty$. It is enough for it to use the results of* Hsu *and* Robbins [33] *and* Dehling [15].

Chapter 2

Asymptotic properties of the statistic T_n

2.1 Consistency

In this section, we prove the consistency of the statistic T_n. Here and in the sequel, we will make use of notations (1.20)–(1.23) without any notice.

Theorem 2.1. *Let* $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ *be i.i.d.* $\mathbb{R}^d$*-valued random variables with a probability distribution* μ *and* T_n *be the statistic defined in* (1.9) *such that* $\liminf_n T_n > -\infty$ μ*-a.s. Suppose that*

$$Eh^4(\mathbf{X}_1, \mathbf{X}_2) < \infty \tag{2.1}$$

and

$$P\{\tilde{h}_1(\mathbf{X}_1) \geq A\} = 1 \tag{2.2}$$

for some constant $A > 0$*. Then* $T_n \to E \log \tilde{h}_1(\mathbf{X}_1)$ *in probability.*

Remark 2.1. *Before starting to prove the theorem, we would like to explain assumption* (2.2) *for the special kernel* $h(\mathbf{x}, \mathbf{y}) = \mathbb{1}(||\mathbf{x} - \mathbf{y}|| \leq \varepsilon)$*. Note that*

$$E(h(\mathbf{X}_1, \mathbf{X}_2)/\mathbf{X}_1) = E(\mathbb{1}\{||\mathbf{X}_1 - \mathbf{X}_2|| \leq \varepsilon\}/\mathbf{X}_1) = \mu(B(\mathbf{X}_1, \varepsilon)).$$

So, (2.2) *means that the measure of the ball* $B(\mathbf{X}_1, \varepsilon)$ *for given* ε *should not be less than given positive constant* A *which is quite natural to assume.*

Proof. First, we note that $0 < E\tilde{h}_1(\mathbf{X}_1) \leq E|h(\mathbf{X}_1,\mathbf{X}_2)|$ and $|\log u| \leq \max\{\log A, u\}$ for $u \geq A > 0$. From (2.1) it follows now that

$$E|\log \tilde{h}_1(\mathbf{X}_1)| \leq \infty\ . \tag{2.3}$$

For convenience, we introduce the following notation

$$a = E\log \tilde{h}_1(\mathbf{X}_1). \tag{2.4}$$

Thus, we need to show that $P(|T_n - a| > \varepsilon) \to 0$ as $n \to \infty$ for $\forall\ \varepsilon > 0$. The representation (1.24) of T_n and a simple argument now yield that

$$\begin{aligned}
P(|T_n - a| > \varepsilon) &= P\left(\left|\frac{1}{n}\sum_{j=1}^{n}(\log(\tilde{h}_1(\mathbf{X}_j) + \eta_{j,n}) - a)\right| > \varepsilon\right)\\
&= P\left(\frac{1}{n}\left|\sum_{j=1}^{n}(\log(\tilde{h}_1(\mathbf{X}_j) + \eta_{j,n}) - a)\right| > \varepsilon;\ \max_{k=1,\dots,n}|\eta_{k,n}| > \frac{A}{2}\right)\\
&+ P\left(\frac{1}{n}\left|\sum_{j=1}^{n}(\log(\tilde{h}_1(\mathbf{X}_j) + \eta_{j,n}) - a)\right| > \varepsilon;\ \max_{k=1,\dots,n}|\eta_{k,n}| \leq \frac{A}{2}\right)\\
&= W_{g1} + W_{g2}.
\end{aligned} \tag{2.5}$$

Consider W_{g1}. From (2.1) and Lemma 1.1, it follows that

$$\begin{aligned}
W_{g1} &\leq P\left(\max_{k=1,\dots,n}|\eta_{k,n}| > \frac{A}{2}\right)\\
&\leq \sum_{k=1}^{n} P\left(|\eta_{k,n}| > \frac{A}{2}\right)\\
&= nP\left(|\eta_{1,n}| > \frac{A}{2}\right) \to 0, \quad \text{as} \quad n \to \infty.
\end{aligned} \tag{2.6}$$

Consider W_{g2}. Using the Taylor expansion of $\log(b+x)$ with the remainder term in the Lagrange form, we find

$$\begin{aligned}
W_{g2} &= P\left(\left|\frac{1}{n}\sum_{j=1}^{n}\left[\log \tilde{h}_1(\mathbf{X}_j) - a + \frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}}\right]\right| \mathbb{1}\left\{\max_{k=1,\dots,n}|\eta_{k,n}| \leq \frac{A}{2}\right\} > \varepsilon\right)\\
&\leq P\left\{\left|\frac{1}{n}\sum_{j=1}^{n}\left[\log \tilde{h}_1(\mathbf{X}_j) - a\right]\right| \mathbb{1}\left\{\max_{k=1,\dots,n}|\eta_{k,n}| \leq \frac{A}{2}\right\} > \frac{\varepsilon}{2}\right\}\\
&+ P\left(\left|\frac{1}{n}\sum_{j=1}^{n}\frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}}\right| \mathbb{1}\left\{\max_{k=1,\dots,n}|\eta_{k,n}| \leq \frac{A}{2}\right\} > \frac{\varepsilon}{2}\right)
\end{aligned}$$

$$= \quad W_{h1} + W_{h2}, \tag{2.7}$$

where $\theta_{j,n}$, $j = 1, \dots, n$ are random variables depending on $\tilde{h}_1(\mathbf{X}_j)$ and $\eta_{j,n}$, moreover, $0 < \theta_{j,n} < 1$ μ-a.s.

Consider W_{h1}. From (2.3) and the law of large numbers for $\{\log \tilde{h}_1(\mathbf{X}_j)\}_{j\in\mathbb{N}}$, we deduce that

$$\begin{aligned} W_{h1} &= P\left(\left|\frac{1}{n}\sum_{j=1}^{n}\left[\log \tilde{h}_1(\mathbf{X}_j) - a\right]\right| \mathbb{1}\left\{\max_{k=1,\dots,n} |\eta_{k,n}| \le \frac{A}{2}\right) > \frac{\varepsilon}{2}\right\} \\ &\le P\left(\left|\frac{1}{n}\sum_{j=1}^{n}\left[\log \tilde{h}_1(\mathbf{X}_j) - a\right]\right| > \frac{\varepsilon}{2}\right) \to 0 \quad \text{as} \quad n \to \infty. \end{aligned} \tag{2.8}$$

Consider W_{h2}. Using Tchebychev and Cauchy-Schwarz inequalities, we have that

$$\begin{aligned} W_{h2} &= P\left(\left|\frac{1}{n}\sum_{j=1}^{n}\frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}}\right| \mathbb{1}\left\{\max_{k=1,\dots,n} |\eta_{k,n}| \le \frac{A}{2}\right\} > \frac{\varepsilon}{2}\right) \\ &\le \frac{4}{\varepsilon^2} E\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\eta_{j,n}\mathbb{1}\left(\max_{k=1,\dots,n}|\eta_{k,n}| \le \frac{A}{2}\right)}{\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}}\right)^2 \\ &\le \frac{4}{\varepsilon^2 n}\sum_{j=1}^{n} E\left(\frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}}\right)^2 \mathbb{1}\left(\max_{k=1,\dots,n}|\eta_{k,n}| < \frac{A}{2}\right). \end{aligned}$$

Further, note that if $|\eta_{jn}| \le A/2$ then

$$\left(\frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}}\right)^2 \le \frac{4}{A^2}\eta_{j,n}^2 \qquad \mu - a.s. \tag{2.9}$$

and hence,

$$\begin{aligned} W_{h2} &\le \frac{4}{\varepsilon^2 n}\sum_{j=1}^{n} E\left(\frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}}\right)^2 \mathbb{1}\left(\max_{k=1,\dots,n}|\eta_{k,n}| < \frac{A}{2}\right) \\ &\le \frac{16}{A^2\varepsilon^2 n}\sum_{j=1}^{n} E\eta_{j,n}^2 \\ &= \frac{16}{A^2\varepsilon^2}\left(\frac{Eh_1^2(\mathbf{X}_1)}{n-1} + \frac{Eh_2^2(\mathbf{X}_1, \mathbf{X}_2)}{n-1}\right) \to 0. \end{aligned} \tag{2.10}$$

In the relation (2.10) we have used the following equality

$$E\eta_{j,n}^2 \;=\; E\left(\frac{1}{n-1}\sum_{i=1:i\neq j}^{n} h_1(\mathbf{X}_i) + \frac{1}{n-1}\sum_{i=1:i\neq j}^{n} h_2(\mathbf{X}_i, \mathbf{X}_j)\right)^2$$

$$= \frac{Eh_1^2(\mathbf{X}_1)}{n-1} + \frac{Eh_2^2(\mathbf{X}_1, \mathbf{X}_2)}{n-1} \quad \text{for any} \quad j = 1, 2, \ldots, n. \tag{2.11}$$

Finally, the theorem is proved by putting together the relations (2.5), (2.6), (2.7), (2.8) and (2.10).

□

2.2 Asymptotic distribution

In this section, we examine the asymptotic distribution of the statistic T_n. First, we need to introduce some notation.

$$A_1 = E\left(\frac{1}{\tilde{h}_1(\mathbf{X}_1)}\right); \tag{2.12}$$

$$\Phi(\mathbf{x}, \mathbf{y}) = \frac{1}{2}\left(\frac{1}{\tilde{h}_1(\mathbf{x})} h_2(\mathbf{y}, \mathbf{x}) + \frac{1}{\tilde{h}_1(\mathbf{y})} h_2(\mathbf{x}, \mathbf{y})\right); \tag{2.13}$$

$$\psi(\mathbf{x}) = E(\Phi(\mathbf{X}_1, \mathbf{X}_2)/\mathbf{X}_1 = \mathbf{x}); \tag{2.14}$$

$$Z_j = \log \tilde{h}_1(\mathbf{X}_j) - a + A_1 h_1(\mathbf{X}_j) + 2\psi(\mathbf{X}_j); \tag{2.15}$$

$$\sigma^2 = Var(Z_1). \tag{2.16}$$

Theorem 2.2. *Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be i.i.d. $\mathbb{R}^d$-valued random variables with a probability distribution μ and T_n be the statistic defined in (1.9) such that $\liminf_n T_n > -\infty$ μ-a.s. Assume that*

$$P(\tilde{h}_1(\mathbf{X}_1) \geq A) = 1 \quad \text{for some} \quad A > 0 \text{ and} \tag{2.17}$$

$$Eh^4(\mathbf{X}_1, \mathbf{X}_2) < \infty. \tag{2.18}$$

If $\sigma > 0$ then $\sqrt{n}\sigma^{-1}(T_n - a)$ is asymptotically normal with mean 0 and variance 1, where a was defined in (2.4).

Remark 2.2. *Note that, using (1.20) – (1.22), Z_j's can also be written in the following form*

$$Z_j = \log(\tilde{h}_1(\mathbf{X}_j)) - a + \int \frac{h(\mathbf{X}_j, \mathbf{y})}{\tilde{h}_1(\mathbf{y})} \mu(\mathbf{dy}) - 1. \tag{2.19}$$

We will make use of this form of Z_j's for constructing a consistent estimator $\hat{\sigma}^2$ for σ^2 when $h(\mathbf{x}, \mathbf{y}) = \mathbb{1}\{\|\mathbf{x} - \mathbf{y}\| \leq \varepsilon\}$. First we will simulate random variables

$\hat{Z}_j$ by replacing $\mu(\mathbf{dy})$ by $\hat{\mu}(\mathbf{dy})$ and the expectation by the sample mean in (2.19), i.e.

$$\hat{Z}_j = \log(\hat{\mu}(B(\mathbf{X}_j,\varepsilon)) - \hat{a} + \frac{1}{n-1}\sum_{i=1:i\neq j}^{n} \frac{\mathbb{1}\{\|\mathbf{X}_j - \mathbf{X}_i\| \leq \varepsilon\}}{\hat{\mu}(B(\mathbf{X}_i,\varepsilon))} - 1,$$

where

$$\hat{\mu}(B(\mathbf{X}_j,\varepsilon) = \frac{1}{n-1}\sum_{i=1:i\neq j}^{n} \mathbb{1}\{\|\mathbf{X}_j - \mathbf{X}_i\| \leq \varepsilon\}$$

and

$$\hat{a} = \frac{1}{n}\sum_{j=1}^{n} \log\left(\hat{\mu}(B(\mathbf{X}_j,\varepsilon)\right).$$

Then we take the sample second moment of $\{\hat{Z}_j\}_{j=1}^n$ which we denote by $\hat{\sigma}^2$ as an estimator for σ^2 since $EZ_j = 0$ for $j = 1, \ldots, n$. The consistency of $\hat{\sigma}^2$ will be proved in the next section.

Proof. First we give some simple consequences of the assumptions (2.17) and (2.18) which will be used in the sequel:

$$E\left|\frac{h_1(\mathbf{X}_1)}{\tilde{h}_1(\mathbf{X}_1)}\right| < \infty \quad ; \quad Var\left(\frac{1}{\tilde{h}_1(\mathbf{X}_1)}\right) < \infty \quad ; \quad E\Phi^2(\mathbf{X}_1,\mathbf{X}_2) < \infty\,. \tag{2.20}$$

An analogous argumentation as at the beginning of the proof of Theorem 2.1 shows that

$$E\log(\tilde{h}_1(\mathbf{X}_1)) = a < \infty \quad \text{and} \quad \sigma^2 < \infty,$$

and consequently, σ^2 in (2.16) is well defined under the assumptions of the theorem.

Consider $\sqrt{n}\sigma^{-1}(T_n - a)$. Now relation (2.10) in the proof of Theorem 2.1 does not hold anymore and we need one more term in the Taylor expansion of the representation (1.24) of T_n, namely

$$\begin{aligned}
\sqrt{n}\sigma^{-1}(T_n - a) &= \frac{1}{\sqrt{n}\sigma}\sum_{j=1}^{n}[\log(\tilde{h}_1(\mathbf{X}_j) + \eta_{j,n}) - a] \\
&= \frac{1}{\sqrt{n}\sigma}\sum_{j=1}^{n}\left[\log(\tilde{h}_1(\mathbf{X}_j)) - a + \frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j)}\right] \\
&- \frac{1}{\sqrt{n}\sigma}\sum_{j=1}^{n}\frac{\eta_{j,n}^2}{\left(\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}\right)^2}
\end{aligned}$$

$$= \quad W_1 - W_2, \tag{2.21}$$

where $\theta_{j,n}$, $j = 1, \ldots, n$ are random variables depending on $\tilde{h}_1(\mathbf{X}_j)$ and $\eta_{j,n}$, moreover, $0 < \theta_{j,n} < 1$ μ-a.s.
The first step in the proof is to show that

$$W_2 \to 0 \quad \text{in probability.} \tag{2.22}$$

A simple argumentation yields

$$\begin{aligned}
W_2 &= P\left(\frac{1}{\sqrt{n}\sigma}\sum_{j=1}^{n}\frac{\eta_{j,n}^2}{\left(\tilde{h}_1(\mathbf{X}_j)+\theta_{j,n}\eta_{j,n}\right)^2} > \varepsilon\right) \\
&= P\left(\frac{1}{\sqrt{n}\sigma}\sum_{j=1}^{n}\frac{\eta_{j,n}^2}{\left(\tilde{h}_1(\mathbf{X}_j)+\theta_{j,n}\eta_{j,n}\right)^2} > \varepsilon\ ;\ \max_{k=1,\ldots,n}|\eta_{k,n}| \le \frac{A}{2}\right) \\
&+ P\left(\frac{1}{\sqrt{n}\sigma}\sum_{j=1}^{n}\frac{\eta_{j,n}^2}{\left(\tilde{h}_1(\mathbf{X}_j)+\theta_{j,n}\eta_{j,n}\right)^2} > \varepsilon\ ;\ \max_{k=1,\ldots,n}|\eta_{k,n}| > \frac{A}{2}\right) \\
&= W_{21} + W_{22}.
\end{aligned} \tag{2.23}$$

Consider W_{21}. By virtue of (2.9), (2.11), (2.17) and Chebyshev inequality, we have

$$\begin{aligned}
W_{21} &= P\left(\frac{1}{\sqrt{n}\sigma}\sum_{j=1}^{n}\frac{\eta_{j,n}^2}{\left(\tilde{h}_1(\mathbf{X}_j)+\theta_{j,n}\eta_{j,n}\right)^2} > \varepsilon\ ;\ \max_{k=1,\ldots,n}|\eta_{k,n}| \le \frac{A}{2}\right) \\
&\le P\left(\frac{1}{\sqrt{n}\sigma}\sum_{j=1}^{n}\frac{4}{A^2}\eta_{j,n}^2 > \varepsilon\ ;\ \max_{k=1,\ldots,n}|\eta_{k,n}| \le \frac{A}{2}\right) \\
&\le P\left(\frac{1}{\sqrt{n}\sigma}\sum_{j=1}^{n}\frac{4}{A^2}\eta_{j,n}^2 > \varepsilon\right) \\
&\le \frac{4}{\sqrt{n}\sigma\varepsilon A^2}\sum_{j=1}^{n}E\eta_{j,n}^2 \\
&= \frac{4n}{\sqrt{n}\sigma\varepsilon A^2}E\eta_{1,n}^2 \to 0.
\end{aligned} \tag{2.24}$$

Consider W_{22}. By Lemma 1.1 it follows

$$\begin{aligned} W_{22} &= P\left(\frac{1}{\sqrt{n}\sigma}\sum_{j=1}^{n}\frac{\eta_{j,n}^2}{\left(\tilde{h}_1(\mathbf{X}_j)+\theta_{j,n}\eta_{j,n}\right)^2} > \varepsilon\ ;\ \max_{k=1,\ldots,n}|\eta_{k,n}| > \frac{A}{2}\right) \\ &\leq P\left(\max_{k=1,\ldots,n}|\eta_{k,n}| > \frac{A}{2}\right) \\ &\leq nP\left(|\eta_{1,n}| > \frac{A}{2}\right) \to 0. \end{aligned} \tag{2.25}$$

Thus, the relations (2.23), (2.24) and (2.25) together prove (2.22).
The second step in the proof is to establish asymptotical equivalence of W_1 and S_n in distribution, where

$$S_n = \frac{1}{\sqrt{n}\sigma}\sum_{j=1}^{n} Z_j,$$

Z_j's and σ are defined in (2.15) and (2.16), respectively. To verify this, it is enough to show that

$$E(W_1 - S_n)^2 \to 0 \quad \text{as} \quad n \to \infty. \tag{2.26}$$

First, note that

$$\begin{aligned} \frac{1}{\sqrt{n}\sigma}\sum_{j=1}^{n}\frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j)} &= \frac{1}{\sqrt{n}\sigma}\sum_{j=1}^{n}\frac{1}{\tilde{h}_1(\mathbf{X}_j)}\frac{1}{n-1}\sum_{i=1:i\neq j}^{n} h_1(\mathbf{X}_i) \\ &+ \frac{1}{\sqrt{n}\sigma}\sum_{j=1}^{n}\frac{1}{\tilde{h}_1(\mathbf{X}_j)}\frac{1}{n-1}\sum_{i=1:i\neq j}^{n} h_2(\mathbf{X}_i,\mathbf{X}_j) \\ &= SS_1 + SS_2. \end{aligned}$$

Secondly, define the following random variables

$$SS = \frac{A_1}{\sqrt{n}\sigma}\sum_{i=1}^{n} h_1(\mathbf{X}_i)$$

and

$$\frac{2\sqrt{n}}{\sigma}U_1 = \frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n} 2\psi(\mathbf{X}_i),$$

where A_1 and $\psi(x)$ defined in (2.12) and (2.14), respectively.
Simple calculation yields that

$$E(W_1 - S_n)^2 = E\left(SS_1 + SS_2 - SS - \frac{2\sqrt{n}}{\sigma}U_1\right)^2. \tag{2.27}$$

By virtue of c_r-inequality and (2.27), (2.26) follows now from

$$E\left(SS_1 - SS\right)^2 \to 0 \tag{2.28}$$

and

$$E\left(SS_2 - \frac{2\sqrt{n}}{\sigma}U_1\right)^2 \to 0. \tag{2.29}$$

In order to prove the relation (2.28), rewrite SS_1 in the following form

$$SS_1 = \frac{1}{\sqrt{n}\sigma}\sum_{i=1}^{n} h_1(\mathbf{X}_i)\frac{1}{n-1}\sum_{j=1:j\neq i}^{n}\frac{1}{\tilde{h}_1(\mathbf{X}_j)}$$

and consequently, we have

$$\begin{aligned}
E(SS - SS_1)^2 &= \frac{1}{n\sigma^2}\sum_{i=1}^{n} E\left(h_1(\mathbf{X}_i)\frac{1}{n-1}\sum_{j=1:j\neq i}^{n}\left[\frac{1}{\tilde{h}_1(\mathbf{X}_j)} - A_1\right]\right)^2 \\
&+ \frac{1}{n\sigma^2}E\left(\sum_{i=1}^{n}\sum_{m=1:m\neq i}^{n} h_1(\mathbf{X}_i)\frac{1}{n-1}\left[\sum_{j=1:j\neq i}^{n}\left(\frac{1}{\tilde{h}_1(\mathbf{X}_j)} - A_1\right)\right]\right. \\
&\times \left. h_1(\mathbf{X}_m)\frac{1}{n-1}\left[\sum_{l=1:l\neq m}^{n}\left(\frac{1}{\tilde{h}_1(\mathbf{X}_l)} - A_1\right)\right]\right) \\
&= QQ_1 + QQ_2.
\end{aligned} \tag{2.30}$$

Consider QQ_1. It follows from (2.18) and (2.20) that

$$\begin{aligned}
QQ_1 &= \frac{1}{n\sigma^2}\sum_{i=1}^{n} E\left(h_1(\mathbf{X}_i)\frac{1}{n-1}\sum_{j=1:j\neq i}^{n}\left[\frac{1}{\tilde{h}_1(\mathbf{X}_j)} - A_1\right]\right)^2 \\
&= \frac{1}{n\sigma^2}\sum_{i=1}^{n} Eh_1^2(\mathbf{X}_i)E\left(\frac{1}{n-1}\sum_{j=1:j\neq i}^{n}\left[\frac{1}{\tilde{h}_1(\mathbf{X}_j)} - A_1\right]\right)^2 \\
&= \frac{1}{\sigma^2}\frac{1}{(n-1)^2}Eh_1^2(\mathbf{X}_1)E\left(\sum_{j=2}^{n}\left[\frac{1}{\tilde{h}_1(\mathbf{X}_j)} - A_1\right]\right)^2 \\
&= \frac{1}{\sigma^2}\frac{1}{(n-1)^2}Eh_1^2(\mathbf{X}_1)\sum_{j=2}^{n} E\left(\frac{1}{\tilde{h}_1(\mathbf{X}_j)} - A_1\right)^2 \\
&= \frac{1}{\sigma^2}\frac{1}{(n-1)}Eh_1^2(\mathbf{X}_1)E\left(\frac{1}{\tilde{h}_1(\mathbf{X}_2)} - A_1\right)^2 \to 0.
\end{aligned} \tag{2.31}$$

Consider QQ_2. Using (2.20) and the degeneracy of the function $h_1(\mathbf{x})$, we find

$$QQ_2 = \frac{1}{n\sigma^2}\sum_{i=1}^{n}\sum_{m=1:m\neq i}^{n} E\left(h_1(\mathbf{X}_i)\frac{1}{n-1}\left[\sum_{j=1:j\neq i}^{n}\left(\frac{1}{\tilde{h}_1(\mathbf{X}_j)} - A_1\right)\right]\right.$$

$$
\begin{aligned}
& \times \quad h_1(\mathbf{X}_m)\frac{1}{n-1}\left[\sum_{l=1:l\neq m}^{n}\left(\frac{1}{\tilde{h}_1(\mathbf{X}_l)}-A_1\right)\right]\Bigg) \\
&= \frac{1}{n(n-1)^2\sigma^2}\sum_{i=1}^{n}\sum_{m=1:m\neq i}^{n} E\left(h_1(\mathbf{X}_m)\left[\sum_{j=1:j\neq i}^{n}\left(\frac{1}{\tilde{h}_1(\mathbf{X}_j)}-A_1\right)\right]\right. \\
& \times \quad \mathbf{E_i}\; h_1(\mathbf{X}_i)\left[\sum_{l=1:l\neq m}^{n}\left(\frac{1}{\tilde{h}_1(\mathbf{X}_l)}-A_1\right)\right]\Bigg) \\
&= \frac{1}{n(n-1)^2\sigma^2}\sum_{i=1}^{n}\sum_{m=1:m\neq i}^{n} E\left(h_1(\mathbf{X}_m)\left[\sum_{j=1:j\neq i}^{n}\left(\frac{1}{\tilde{h}_1(\mathbf{X}_j)}-A_1\right)\right]\mathbf{E_i}\;\frac{h_1(\mathbf{X}_i)}{\tilde{h}_1(\mathbf{X}_i)}\right) \\
&= \frac{1}{n(n-1)^2\sigma^2}\sum_{i=1}^{n}\sum_{m=1:m\neq i}^{n} E\frac{h_1(\mathbf{X}_m)}{\tilde{h}_1(\mathbf{X}_m)}E\;\frac{h_1(\mathbf{X}_i)}{\tilde{h}_1(\mathbf{X}_i)} \\
&= \frac{1}{(n-1)\sigma^2}E\frac{h_1(\mathbf{X}_1)}{\tilde{h}_1(\mathbf{X}_1)}E\;\frac{h_1(\mathbf{X}_2)}{\tilde{h}_1(\mathbf{X}_2)} \to 0, \qquad (2.32)
\end{aligned}
$$

where $\mathbf{E_i}$ denotes the conditional expectation with respect to $\mathbf{X}_i$.
The relations (2.30),(2.31) and (2.32) together prove the relation (2.28).
To estabilsh (2.29), we rewrite SS_2 in the following form

$$
\begin{aligned}
SS_2 &= \frac{1}{\sqrt{n}\sigma}\sum_{j=1}^{n}\frac{1}{\tilde{h}_1(\mathbf{X}_j)}\frac{1}{n-1}\sum_{i=1:i\neq j}^{n} h_2(\mathbf{X}_i,\mathbf{X}_j) \\
&= \frac{1}{\sqrt{n}(n-1)\sigma}\sum_{1\leq i\neq j\leq n}^{n}\frac{1}{\tilde{h}_1(\mathbf{X}_j)}h_2(\mathbf{X}_i,\mathbf{X}_j) \\
&= \frac{1}{\sqrt{n}(n-1)\sigma}\sum_{1\leq i\neq j\leq n}^{n}\frac{1}{2}\left(\frac{1}{\tilde{h}_1(\mathbf{X}_j)}h_2(\mathbf{X}_i,\mathbf{X}_j)+\frac{1}{\tilde{h}_1(\mathbf{X}_i)}h_2(\mathbf{X}_j,\mathbf{X}_i)\right) \\
&= \frac{1}{\sqrt{n}(n-1)\sigma}\sum_{1\leq i\neq j\leq n}^{n}\frac{1}{2}\left(\frac{1}{\tilde{h}_1(\mathbf{X}_j)}h_2(\mathbf{X}_i,\mathbf{X}_j)+\frac{1}{\tilde{h}_1(\mathbf{X}_i)}h_2(\mathbf{X}_j,\mathbf{X}_i)\right) \\
&= \frac{1}{\sqrt{n}(n-1)\sigma}\sum_{1\leq i\neq j\leq n}^{n}\Phi(\mathbf{X}_i,\mathbf{X}_j),
\end{aligned}
$$

where $\Phi(x, y)$ was defined in (2.13).
Now we see that SS_2 can be represented through a U-statistic with the kernel $\Phi(\mathbf{x},\mathbf{y})$ in the following way

$$SS_2 = \frac{\sqrt{n}}{\sigma}U_n.$$

Noting that $E\Phi(\mathbf{X}_1,\mathbf{X}_2) = 0$, the relation (2.29) follows from (2.20) and (1.19).

Finally, the relations (2.21), (2.22), (2.26) and an application of the central limit theorem to S_n complete the proof.

□

2.3 Consistent estimator of the variance

In this section, we prove the consistency of $\hat{\sigma}^2$ mentioned before in Remark 2.2. Since we are going to apply Theorem 2.1 for $h(\mathbf{x},\mathbf{y}) = \mathbb{1}\{\|\mathbf{x}-\mathbf{y}\| \leq \varepsilon\}$, we assume further that

$$P(0 \leq h(\mathbf{X}_1,\mathbf{X}_2) \leq C_1) = 1. \tag{2.33}$$

Theorem 2.3. *Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be i.i.d. $\mathbb{R}^d$-valued random variables with a probability distribution μ and T_n be the statistic defined in (1.9) such that $\liminf_n T_n > -\infty$ μ-a.s. Assume that (2.2) and (2.33) hold. Then the following statistic*

$$\begin{aligned}
\hat{\sigma}_n^2 &= \frac{1}{n}\sum_{j=1}^{n}\log^2\left(\frac{1}{n-1}\sum_{i=1:i\neq j}^{n} h(\mathbf{X}_i,\mathbf{X}_j)\right) - T_n^2 \\
&+ \frac{1}{n^3}\sum_{s=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{h(\mathbf{X}_s,\mathbf{X}_i)}{\frac{1}{n-1}\sum_{u\neq i} h(\mathbf{X}_u,\mathbf{X}_i)}\cdot\frac{h(\mathbf{X}_s,\mathbf{X}_j)}{\frac{1}{n-1}\sum_{v\neq j} h(\mathbf{X}_v,\mathbf{X}_j)} + 1 \\
&- \frac{2}{n^2}\sum_{s=1}^{n}\sum_{i=1}^{n}\frac{h(\mathbf{X}_s,\mathbf{X}_i)}{\frac{1}{n-1}\sum_{u\neq i} h(\mathbf{X}_u,\mathbf{X}_i)} - 2T_n \\
&+ \frac{2}{n^2}\sum_{s=1}^{n}\sum_{j=1}^{n}\left[\log\left(\frac{1}{n-1}\sum_{u\neq s} h(\mathbf{X}_u,\mathbf{X}_s)\right)\right]\cdot\frac{h(\mathbf{X}_s,\mathbf{X}_j)}{\frac{1}{n-1}\sum_{v\neq j} h(\mathbf{X}_v,\mathbf{X}_j)}
\end{aligned} \tag{2.34}$$

is a consistent estimator for σ^2 defined in (2.16).

Proof. The proof of the theorem consists of several steps. First we will give an exact formula for σ^2. Further we give consistent estimators for each term of this formula.
Simple calculation shows that

$$\begin{aligned}
\sigma^2 &= Var(Z_1) = EZ_1^2 \\
&= E\left(\log(\tilde{h}_1(\mathbf{X}_1)) - a + \int \frac{h(\mathbf{X}_1,\mathbf{y})}{\tilde{h}_1(\mathbf{y})}\mu(\mathbf{dy}) - 1\right)^2
\end{aligned}$$

$$
\begin{aligned}
&= E\left(\log(\tilde{h}_1(\mathbf{X}_1)) - a\right)^2 + E\left(\int \frac{h(\mathbf{X}_1,\mathbf{y})}{\tilde{h}_1(\mathbf{y})}\mu(\mathbf{dy}) - 1\right)^2 \\
&+ 2E\left(\log(\tilde{h}_1(\mathbf{X}_1)) - a\right)\left(\int \frac{h(\mathbf{X}_1,\mathbf{y})}{\tilde{h}_1(\mathbf{y})}\mu(\mathbf{dy}) - 1\right) \\
&= E_1 + E_2 + 2E_3 \ , \qquad (2.35)
\end{aligned}
$$

where a was defined in (2.4).

Now we give consistent estimators for E_1, E_2 and E_3.

Consider E_1. It is enough to find a consistent estimator for $E\log^2(\tilde{h}_1(\mathbf{X}_1))$ since $E_1 = E\log^2(\tilde{h}_1(\mathbf{X}_1)) - a^2$ and a^2 can be consistently estimated by T_n^2 which is a consequence of Slutzky's Theorem and Theorem 2.1. The natural choice of the estimator for $E\log^2(\tilde{h}_1(\mathbf{X}_1))$ is

$$
\hat{E}_1 = \frac{1}{n}\sum_{j=1}^{n}\log^2\left(\frac{1}{n-1}\sum_{i=1:i\neq j}^{n} h(\mathbf{X}_i,\mathbf{X}_j)\right).
$$

Using the representation (1.24) and the Taylor expansion, we can rewrite $\hat{E}_1$ in the following form

$$
\begin{aligned}
\hat{E}_1 &= \frac{1}{n}\sum_{j=1}^{n}\left(\log\tilde{h}_1(\mathbf{X}_j) + \frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}}\right)^2 \\
&= \frac{1}{n}\sum_{j=1}^{n}\log^2\tilde{h}_1(\mathbf{X}_j) \\
&+ \frac{2}{n}\sum_{j=1}^{n}\log\tilde{h}_1(\mathbf{X}_j)\cdot\frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}} \\
&+ \frac{1}{n}\sum_{j=1}^{n}\left(\frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}}\right)^2 \\
&= S_{1n} + 2S_{2n} + S_{3n} \ , \qquad (2.36)
\end{aligned}
$$

where $\eta_{j,n}$ are defined in (1.23) and $\theta_{j,n}$, $j = 1,\ldots,n$ are $(0,1)$-valued random variables depending on $\eta_{j,n}$ and $\tilde{h}_1(\mathbf{X}_j)$.

By virtue of (2.2) and (2.33), it follows that $E\log^2\tilde{h}_1(\mathbf{X}_1) < \infty$ and therefore

$$
S_{1n} \xrightarrow{P} E\log^2\tilde{h}_1(\mathbf{X}_1) \qquad \text{as } n\to\infty \qquad (2.37)
$$

by the law of large numbers.

It remains now to prove that

$$S_{2n} \xrightarrow{P} 0 \qquad \text{as } n \to \infty \tag{2.38}$$

and

$$S_{3n} \xrightarrow{P} 0 \qquad \text{as } n \to \infty. \tag{2.39}$$

Consider S_{2n}. Thus, by the conditions imposed on $\tilde{h}_1$ and h, we find that $|\log \tilde{h}_1(X_1)| < C_2$ μ-a.s., where $C_2 = \max\{|\log A|, |\log C_1|\}$. Further, a simple argumentation yields that

$$\begin{aligned}
P(|S_{2n}| > \varepsilon) &\leq P\left(\sum_{j=1}^{n}\left|\frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}}\right| > \frac{\varepsilon \cdot n}{C_2}\right) \\
&\leq \sum_{j=1}^{n} P\left(\left|\frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}}\right| > \frac{\varepsilon}{C_2}\right) \\
&= \sum_{j=1}^{n} P\left(\left|\frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}}\right| > \frac{\varepsilon}{C_2}\ ;\ |\eta_{j,n}| \leq \frac{A}{2}\right) \\
&+ \sum_{j=1}^{n} P\left(\left|\frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}}\right| > \frac{\varepsilon}{C_2}\ ;\ |\eta_{j,n}| > \frac{A}{2}\right) \\
&= P_{2an} + P_{2bn}.
\end{aligned} \tag{2.40}$$

Consider P_{2bn}. By Lemma 1.1 we get that

$$\begin{aligned}
P_{2bn} &\leq \sum_{j=1}^{n} P\left(|\eta_{j,n}| > \frac{A}{2}\right) \\
&= nP(|\eta_{1,n}| > \frac{A}{2}) \\
&\to 0.
\end{aligned} \tag{2.41}$$

Consider P_{2an}. Using Lemma 1.1 and (2.2), we find that

$$\begin{aligned}
P_{2an} &= \sum_{j=1}^{n} P\left(\left|\frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}}\right| > \frac{\varepsilon}{C_2}\ ;\ |\eta_{j,n}| \leq \frac{A}{2}\right) \\
&\leq \sum_{j=1}^{n} P\left(\left|\frac{\eta_{j,n}}{A - \frac{A}{2}}\right| > \frac{\varepsilon}{C_2}\ ;\ |\eta_{j,n}| \leq \frac{A}{2}\right) \\
&\leq \sum_{j=1}^{n} P\left(|\eta_{j,n}| > \frac{A\varepsilon}{2C_2}\right)
\end{aligned}$$

$$
\begin{aligned}
&= nP\left(|\eta_{1,n}| > \frac{A\varepsilon}{2C_2}\right) \\
&\to 0\ . \qquad (2.42)
\end{aligned}
$$

Now the relation (2.38) follows from the relations (2.40), (2.41) and (2.42).
In order to prove the relation (2.39), we use the same technique as in (2.40) and get that

$$
\begin{aligned}
P(|S_{3n}| > \varepsilon) &= P\left(\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}}\right)^2 > \varepsilon\right) \\
&\leq \sum_{j=1}^{n} P\left(\left(\frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}}\right)^2 > \varepsilon\right) \\
&= \sum_{j=1}^{n} P\left(\left(\frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}}\right)^2 > \varepsilon\ ;\ |\eta_{j,n}| \leq \frac{A}{2}\right) \\
&+ \sum_{j=1}^{n} P\left(\left(\frac{\eta_{j,n}}{\tilde{h}_1(\mathbf{X}_j) + \theta_{j,n}\eta_{j,n}}\right)^2 > \varepsilon\ ;\ |\eta_{j,n}| > \frac{A}{2}\right) \\
&= P_{3an} + P_{3bn}\ . \qquad (2.43)
\end{aligned}
$$

In the same way as in the derivation of the relations (2.41) and (2.42), one can show that

$$
P_{3an} \to 0 \qquad \text{and} \qquad P_{3bn} \to 0\ , \qquad \text{as} \qquad n \to \infty. \qquad (2.44)
$$

Now the relation (2.39) follows from the relations (2.43) and (2.44).
Combining together (2.36), (2.37), (2.38) and (2.39), we obtain that $\hat{E}_1 - T_n^2$ is a consistent estimator for E_1.
Further, consider E_2. Simple calculation yields that

$$
\begin{aligned}
E_2 &= E\left(\int \frac{h(\mathbf{X}_1, \mathbf{y})}{\tilde{h}_1(\mathbf{y})}\mu(\mathbf{dy}) - 1\right)^2 \\
&= E\left(\int \frac{h(\mathbf{X}_1, \mathbf{y})}{\tilde{h}_1(\mathbf{y})}\mu(\mathbf{dy})\right)^2 + 1 - 2E\left(\int \frac{h(\mathbf{X}_1, \mathbf{y})}{\tilde{h}_1(\mathbf{y})}\mu(\mathbf{dy})\right) \\
&= E_{2a} + 1 - 2E_{2b} \qquad (2.45)
\end{aligned}
$$

and hence, it is enough to estimate consistently E_{2a} and E_{2b}.

Rewrite E_{2a} in the following form

$$\begin{aligned}
E_{2a} &= E\left(\int \frac{h(\mathbf{X}_1,\mathbf{y})}{\tilde{h}_1(\mathbf{y})}\mu(\mathbf{dy})\right)^2 \\
&= E\left(\int \frac{h(\mathbf{X}_1,\mathbf{x})}{\tilde{h}_1(\mathbf{x})}\mu(\mathbf{dx})\right)\cdot\left(\int \frac{h(\mathbf{X}_1,\mathbf{y})}{\tilde{h}_1(\mathbf{y})}\mu(\mathbf{dy})\right) \\
&= E\int\int \frac{h(\mathbf{X}_1,\mathbf{x})}{\tilde{h}_1(\mathbf{x})}\cdot\frac{h(\mathbf{X}_1,\mathbf{y})}{\tilde{h}_1(\mathbf{y})}\mu(\mathbf{dx})\mu(\mathbf{dy})
\end{aligned}$$

and we see that the natural estimator for the E_{2a} is

$$\hat{E}_{2a} = \frac{1}{n^3}\sum_{s=1}^n\sum_{i=1}^n\sum_{j=1}^n \frac{h(\mathbf{X}_s,\mathbf{X}_i)}{\frac{1}{n-1}\sum_{u\neq i}h(\mathbf{X}_u,\mathbf{X}_i)}\cdot\frac{h(\mathbf{X}_s,\mathbf{X}_j)}{\frac{1}{n-1}\sum_{v\neq j}h(\mathbf{X}_v,\mathbf{X}_j)}\ .$$

Using the same technique as for $\hat{E}_1$, we obtain that

$$\begin{aligned}
\hat{E}_{2a} &= \frac{1}{n^3}\sum_{s=1}^n\sum_{i=1}^n\sum_{j=1}^n \frac{h(\mathbf{X}_s,\mathbf{X}_i)}{\tilde{h}_1(\mathbf{X}_i)+\eta_{i,n}}\cdot\frac{h(\mathbf{X}_s,\mathbf{X}_j)}{\tilde{h}_1(\mathbf{X}_j)+\eta_{j,n}} \\
&= \frac{1}{n^3}\sum_{s=1}^n\sum_{i=1}^n\sum_{j=1}^n h(\mathbf{X}_s,\mathbf{X}_i)\cdot h(\mathbf{X}_s,\mathbf{X}_j) \\
&\times \left(\frac{1}{\tilde{h}_1(\mathbf{X}_i)}-\frac{\eta_{i,n}}{(\tilde{h}_1(\mathbf{X}_i)+\theta_{i,n}\eta_{i,n})^2}\right)\cdot\left(\frac{1}{\tilde{h}_1(\mathbf{X}_j)}-\frac{\eta_{j,n}}{(\tilde{h}_1(\mathbf{X}_j)+\theta_{j,n}\eta_{j,n})^2}\right) \\
&= \frac{1}{n^3}\sum_{s=1}^n\sum_{i=1}^n\sum_{j=1}^n \frac{h(\mathbf{X}_s,\mathbf{X}_i)}{\tilde{h}_1(\mathbf{X}_i)}\cdot\frac{h(\mathbf{X}_s,\mathbf{X}_j)}{\tilde{h}_1(\mathbf{X}_j)} \\
&+ \frac{1}{n^3}\sum_{s=1}^n\sum_{i=1}^n\sum_{j=1}^n \frac{h(\mathbf{X}_s,\mathbf{X}_i)\eta_{i,n}}{(\tilde{h}_1(\mathbf{X}_i)+\theta_{i,n}\eta_{i,n})^2}\cdot\frac{h(\mathbf{X}_s,\mathbf{X}_j)\eta_{j,n}}{(\tilde{h}_1(\mathbf{X}_j)+\theta_{j,n}\eta_{j,n})^2} \\
&- \frac{1}{n^3}\sum_{s=1}^n\sum_{i=1}^n\sum_{j=1}^n \frac{h(\mathbf{X}_s,\mathbf{X}_i)}{\tilde{h}_1(\mathbf{X}_i)}\cdot\frac{h(\mathbf{X}_s,\mathbf{X}_j)\eta_{j,n}}{(\tilde{h}_1(\mathbf{X}_j)+\theta_{j,n}\eta_{j,n})^2} \\
&- \frac{1}{n^3}\sum_{s=1}^n\sum_{i=1}^n\sum_{j=1}^n \frac{h(\mathbf{X}_s,\mathbf{X}_i)\eta_{i,n}}{(\tilde{h}_1(\mathbf{X}_i)+\theta_{i,n}\eta_{i,n})^2}\cdot\frac{h(\mathbf{X}_s,\mathbf{X}_j)}{\tilde{h}_1(\mathbf{X}_j)} \\
&= \hat{E}_{2a1}+\hat{E}_{2a2}+\hat{E}_{2a3}+\hat{E}_{2a4}\ , \qquad (2.46)
\end{aligned}$$

where $\theta_{j,n}$, $j = 1,\ldots,n$ are $(0,1)$-valued random variables depending on $\tilde{h}_1(\mathbf{X}_j)$ and $\eta_{j,n}$.

It is not difficult to see that

$$\hat{E}_{2a1} \xrightarrow{P} E_{2a} \qquad \text{as } n\to\infty \qquad (2.47)$$

by the law of large numbers for U-statistics and therefore, it remains to prove that

$$\hat{E}_{2a2}, \quad \hat{E}_{2a3} \text{ and } \hat{E}_{2a4} \xrightarrow{P} 0 \qquad \text{as } n \to \infty . \tag{2.48}$$

The relation (2.48) can be derived in the same way as the relations (2.38) and (2.39). Now from (2.46), (2.47) and (2.48), it follows that

$$\hat{E}_{2a} \xrightarrow{P} E_{2a} \qquad \text{as } n \to \infty. \tag{2.49}$$

Analogously, one can show that

$$\hat{E}_{2b} = \frac{1}{n^2} \sum_{s=1}^{n} \sum_{i=1}^{n} \frac{h(\mathbf{X}_s, \mathbf{X}_i)}{\frac{1}{n-1} \sum_{u \neq i} h(\mathbf{X}_u, \mathbf{X}_i)} . \tag{2.50}$$

is a consistent estimator for E_{2b}, i.e.

$$\hat{E}_{2b} \xrightarrow{P} E_{2b} \qquad \text{as} \qquad n \to \infty. \tag{2.51}$$

Combining (2.45)–(2.51), we obtain that $\hat{E}_{2a} + 1 - 2\hat{E}_{2b}$ is a consistent estimator for E_2.

Consider now the last term of (2.35), namely E_3. Note that E_3 can be written in the following form

$$\begin{aligned} E_3 &= E\left(\log(\tilde{h}_1(\mathbf{X}_1))\right) \cdot \int \frac{h(\mathbf{X}_1, \mathbf{y})}{\tilde{h}_1(\mathbf{y})} \mu(\mathbf{dy}) - a \\ &= E_{3a} - a. \end{aligned} \tag{2.52}$$

By Theorem 2.1, a can be consistently estimated by T_n and hence, it remains only to find a consistent estimator for E_{3a}. The same technique, used for proving consistency of $\hat{E}_1$ and $\hat{E}_{2a}$, will show that the estimator

$$\hat{E}_{3a} = \frac{1}{n^2} \sum_{s=1}^{n} \sum_{j=1}^{n} \left[\log\left(\frac{1}{n-1} \sum_{u \neq s} h(\mathbf{X}_u, \mathbf{X}_s) \right) \right] \cdot \frac{h(\mathbf{X}_s, \mathbf{X}_j)}{\frac{1}{n-1} \sum_{v \neq j} h(\mathbf{X}_v, \mathbf{X}_j)} .$$

is a consistent estimator for E_{3a}.

Finally, note that the following equality for $\hat{\sigma}_n^2$ defined in (2.34) holds

$$\hat{\sigma}_n^2 = \hat{E}_1 - T_n^2 + \hat{E}_{2a} + 1 - 2\hat{E}_{2b} + 2\hat{E}_3 - 2T_n$$

and therefore, $\hat{\sigma}_n^2$ consistently estimates σ^2.

□

2.4 Multivariate Central Limit Theorem

In this section, we prove the multivariate central limit theorem for the vector $\mathbf{T}_n = (T_n(\varepsilon_1), T_n(\varepsilon_2), \ldots, T_n(\varepsilon_m))$, where $0 < \varepsilon_1 < \varepsilon_2 < \ldots < \varepsilon_m < 1$ and $T_n(\varepsilon)$ was defined (1.8). Note that the statistic $T_n(\varepsilon)$ has a kernel depending on ε. We will indicate this by an upper index k enclosed in brackets which ranges from 1 to m. In the same way we modify notations (2.12)–(2.16), i.e.

$$\begin{aligned}
A_1^{(k)} &= E\left(\frac{1}{\tilde{h}_1^{(k)}(\mathbf{X}_1)}\right); \\
\Phi^{(k)}(\mathbf{x},\mathbf{y}) &= \frac{1}{2}\left(\frac{1}{\tilde{h}_1^{(k)}(\mathbf{x})}h_2^{(k)}(\mathbf{y},\mathbf{x}) + \frac{1}{\tilde{h}_1^{(k)}(\mathbf{y})}h_2^{(k)}(\mathbf{x},\mathbf{y})\right); \\
\psi^{(k)}(\mathbf{x}) &= E(\Phi^{(k)}(\mathbf{X}_1,\mathbf{X}_2)/\mathbf{X}_1 = \mathbf{x}); \\
Z_j^{(k)} &= \log \tilde{h}_1^{(k)}(\mathbf{X}_j) - a + A_1^{(k)} h_1^{(k)}(\mathbf{X}_j) + 2\psi^{(k)}(\mathbf{X}_j); \\
\left(\sigma^{(k)}\right)^2 &= Var(Z_1^{(k)}).
\end{aligned}$$

Finally, denote the vector of expectations $(E \log \tilde{h}_1^{(1)}(\mathbf{X}_1), \ldots, E \log \tilde{h}_1^{(m)}(\mathbf{X}_1))$ by $\mathbf{a}$.

Theorem 2.4. *Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be i.i.d. $\mathbb{R}^d$-valued random variables with a probability distribution μ. Assume that*

$$P(\tilde{h}_1^{(k)}(\mathbf{X}_1) \geq A) = 1 \quad \textit{for some} \quad A > 0 \quad \textit{and} \tag{2.53}$$

$$E\left(h^{(k)}(\mathbf{X}_1,\mathbf{X}_2)\right)^4 < \infty \tag{2.54}$$

for all $k = 1, \ldots, m$. Then the distribution of $\sqrt{n}(\mathbf{T}_n - \mathbf{a})$ converges weakly to m-dimensional normal distribution with zero expectation and the covariance matrix $\mathbf{V}$, whose entries are given by $v_{kl} = EZ_j^{(k)} \cdot Z_j^{(l)}$.

Proof. According to Cramer-Wald device, it is enough to prove that a distribution of a linear combination $\sqrt{n}\sum_{k=1}^m c_k T_n(\varepsilon_k)$ tends to the normal distribution with zero expectation and the variance $\sum_{k=1}^m \sum_{l=1}^m c_k v_{kl} c_l$, where $\mathbf{c} = (c_1, \ldots, c_m)$ is some arbitrary fixed vector of constants such that $\exists k_0 : c_{k_0} \neq 0$. Without loss of generality, we may assume that $\mathbf{c}$ has positive entries.

A simple argumentation yields that

$$\triangle_n(\mathbf{c}) = P\left(\left|\sum_{k=1}^m c_k \left[\sqrt{n}T_n(\varepsilon_k) - \frac{1}{\sqrt{n}}\sum_{j=1}^n Z_j^{(k)}\right]\right| > \varepsilon\right)$$

$$\leq \sum_{k=1}^{m} P\left(\left|\sqrt{n}T_n(\varepsilon_k) - \frac{1}{\sqrt{n}}\sum_{j=1}^{n} Z_j^{(k)}\right| > \frac{\varepsilon}{c_k m}\right) . \qquad (2.55)$$

It follows from (2.21), (2.22), (2.26), (2.53) and (2.54) that every term of the sum on the right hand side of (2.55) converges to 0 as $n \to \infty$. Hence, $\sum_{k=1}^{m} c_k \sqrt{n}T_n(\varepsilon_k)$ and $\sum_{k=1}^{m} c_k \frac{1}{\sqrt{n}} \sum_{j=1}^{n} Z_j^{(k)}$ have the same limit distribution. In order to investigate an asymptotic distribution of $\sum_{k=1}^{m} c_k \frac{1}{\sqrt{n}} \sum_{j=1}^{n} Z_j^{(k)}$, rewrite it in the following form

$$\begin{aligned}
\sum_{k=1}^{m} c_k \frac{1}{\sqrt{n}} \sum_{j=1}^{n} Z_j^{(k)} &= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \sum_{k=1}^{m} c_k Z_j^{(k)} \\
&= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} V_j(\mathbf{c}) .
\end{aligned}$$

Note that $V_1(\mathbf{c}), \ldots, V_n(\mathbf{c})$ is a sequence of i.i.d. random variables with zero expectation and variance τ^2 given by

$$\begin{aligned}
\tau^2 &= EV_1^2(\mathbf{c}) = E\left(\sum_{k=1}^{m} c_k Z_1^{(k)}\right)^2 \\
&= \sum_{k=1}^{m} \sum_{l=1}^{m} c_k c_l E Z_1^{(k)} Z_1^{(l)} = \sum_{k=1}^{m} \sum_{l=1}^{m} c_k v_{kl} c_l .
\end{aligned}$$

An application of the central limit theorem to $V_1(\mathbf{c}), \ldots, V_n(\mathbf{c})$ completes the proof.

□

Chapter 3

Numerical results

3.1 Theoretical background

In this chapter we will give applications of the theoretical results from the previous chapter. First, let us give some theoretical background of our practical algorithm. Assume that the information dimension

$$\sigma_\mu = \lim_{\varepsilon \to 0} \frac{1}{\log \varepsilon} \int \log \mu(B(\mathbf{x}, \varepsilon)) d\mu(x)$$

exists and note that if $\mathbf{X}$ has a distribution μ then $E \log \mu(B(\mathbf{X}, \varepsilon)) = \int \log \mu(B(\mathbf{x}, \varepsilon)) d\mu(x)$.

Further, we require $E \log \mu(B(\mathbf{X}, \varepsilon))$ to obey a linear law with respect to $\log \varepsilon$, i.e. $E \log \mu(B(\mathbf{X}, \varepsilon)) \approx K + \sigma_\mu \log \varepsilon$ $(\varepsilon \to 0)$ for some constants K and σ_μ. For a sequence of radii $\varepsilon_1 < \varepsilon_2 < \ldots < \varepsilon_m$ and $n \geq 2$ the vectors $\mathbf{T}_n = (T_n(\varepsilon_1), T_n(\varepsilon_2), \ldots, T_n(\varepsilon_m))$ provide consistent estimators for

$$E \log \mu(B(\mathbf{X}, \varepsilon_1)) \approx K + \sigma_\mu \log \varepsilon_1 \; ; \quad (3.1)$$

$$E \log \mu(B(\mathbf{X}, \varepsilon_2)) \approx K + \sigma_\mu \log \varepsilon_2 \; ; \quad (3.2)$$

$$\vdots \qquad \vdots$$

$$E \log \mu(B(\mathbf{X}, \varepsilon_m)) \approx K + \sigma_\mu \log \varepsilon_m \; . \quad (3.3)$$

Moreover, $\sqrt{n}\mathbf{T}_n$ converges weakly to the m-dimensional normal distribution with zero expectation and a covariance matrix $\mathbf{V}$.

Introduce the following vector $\mathbf{u} = (u_1, \ldots, u_m)$ with entries defined by

$$u_i = \frac{\log \varepsilon_i - \frac{1}{m}\sum_{k=1}^{m} \log \varepsilon_k}{\frac{1}{m}\sum_{k=1}^{m} \log^2 \varepsilon_k - \left(\frac{1}{m}\sum_{k=1}^{m} \log \varepsilon_k\right)^2} \ .$$

Now assume that a strict equality holds in (3.1)–(3.3). Then $\hat{\sigma}_\mu$ given by

$$\hat{\sigma}_\mu = \frac{1}{m}\sum_{i=1}^{m} u_i T_n(\varepsilon_i) \tag{3.4}$$

is a least squares estimator for σ_μ, i.e. $\hat{\sigma}_\mu$ and $\widehat{K} = K - \hat{\sigma}_\mu \cdot \frac{1}{m}\sum_{i=1}^{m} \log \varepsilon_i$ minimize the sum of squares of deviations $\sum_{i=1}^{m} \left(T_n(\varepsilon_i) - a - b \log \varepsilon_i\right)^2$ over all possible choices $a \in \mathbb{R}$ and $b \in \mathbb{R}$. Furthermore, $\sqrt{n}(\hat{\sigma}_\mu - \sigma_\mu)$ is asymptotically normal with expectation 0 and variance σ^2_{LS} defined by

$$\sigma^2_{LS} = \frac{1}{m}\sum_{i=1}^{m} \frac{1}{m}\sum_{k=1}^{m} u_i v_{ij} u_k \ .$$

Note that the entries of the covariance matrix $\mathbf{V}$ can be consistently estimated from the underlying data in the same way as it was done for σ^2 in Theorem 2.3. We will denote a consistent estimator of σ^2_{LS} by $\hat{\sigma}^2_{LS}$, i.e.

$$\hat{\sigma}^2_{LS} = \frac{1}{m}\sum_{i=1}^{m} \frac{1}{m}\sum_{k=1}^{m} u_i \hat{v}_{ij} u_k \ . \tag{3.5}$$

By virtue of (3.4), (3.5) and Theorem 2.4, a two sided 95%-confidence interval for σ_μ is given by

$$\left[\hat{\sigma}_\mu - \frac{1.96 \cdot \hat{\sigma}_{LS}}{\sqrt{n}} \ , \hat{\sigma}_\mu + \frac{1.96 \cdot \hat{\sigma}_{LS}}{\sqrt{n}}\right] .$$

In the next two sections we apply our theory to probability measures of fractional information dimension. Without loss of generality, we restrict ourselves by considering a special family of probability measures on the unit cube in $\mathbb{R}^d$, originally introduced by Cutler [10], since usually dynamical systems have a bounded attractor. Another reason for considering this special family of probability measures is that it is well studied while the true information dimension of many well-known dynamical systems is still unknown. We also hope that this family of probability measures exhibits dimension behavior of most distributions on the unit cube. The general construction of this family of probability measures

can be found in Cutler [12]. To illustrate the main idea, we will construct the generalized Cantor distribution on unit cube in $\mathbb{R}^3$ in Section 3.3.
In both numerical simulations of this chapter, the 95%-confidence intervals cover the true value of the information dimension in 97 out of 100 cases. In Section 3.3 we even obtained slightly shorter confidence intervals than Cutler [10]. Note that we construct confidence intervals on the basis of 5000 observations while she did them on the basis of 20400 observations. However, the main point of our investigations is to show how the whole sample can be used in such estimation problems without splitting into parts.
Simulations have been performed using a software TSTOOL based on the core platform MATLAB.

3.2 Cantor distribution in $\mathbb{R}^2$

In this section we present numerical results for a two-dimensional Cantor distribution. If we denote by C the Cantor set then the two-dimensional Cantor distribution is a uniform distribution on the Cartesian product of $C \times C$. It is known that the information dimension σ_μ of this distribution is approximately equal to 1.2619 (see Cutler [12]).
We have produced 100 simulations. In each simulation a sample of a size 5000 was randomly drawn from the two-dimensional Cantor distribution. $\varepsilon_k = 0.0021 + 0.0001 \cdot k$, $k = 0, \dots, 8$ were chosen as a radii for balls $B(X_i, \varepsilon)$, for $i = 1, \dots, 5000$. Then points which did not have any neighbor in their ε_1-neighborhood have been thrown out. Results of the simulations are shown in Tables 3.1-3.4. Intervals which do not cover the true value of the dimension are marked by asterisk "$*$".

Table 3.1: Cantor distribution in $\mathbb{R}^2$
($\sigma_\mu \approx 1.2619$, $\varepsilon = 0.0021 + 0.0001 \cdot k, k = 0, \ldots, 8$)

confidence intervals	**length of confidence intervals**	$\hat{\sigma}$	**sample size**
[1.1780 , 1.3344]	0.1564	2.5210	3993
[1.2007 , 1.3617]	0.1610	2.6053	4024
[1.1414 , 1.2987]	0.1573	2.5358	3995
[1.1276 , 1.2830]	0.1553	2.4845	3931
[1.1371 , 1.2898]	0.1528	2.4641	3997
[1.1679 , 1.3209]	0.1530	2.4677	3997
[1.1561 , 1.3139]	0.1577	2.5489	4012
[1.1489 , 1.3066]	0.1578	2.5376	3974
[1.1132 , 1.2672]	0.1539	2.4639	3936
[1.1474 , 1.3029]	0.1555	2.5083	3999
[1.1218 , 1.2779]	0.1561	2.5070	3962
[1.1339 , 1.2909]	0.1570	2.5311	3996
[1.1499 , 1.3071]	0.1573	2.5497	4038
[1.1491 , 1.3030]	0.1539	2.4809	3993
[1.1769 , 1.3342]	0.1573	2.5299	3974
[1.1204 , 1.2762]	0.1558	2.5008	3960
[1.1474 , 1.3060]	0.1585	2.5420	3950
[1.1435 , 1.2970]	0.1535	2.4688	3973
[1.2047 , 1.3641]	0.1594	2.5766	4016
[1.2030 , 1.3612]	0.1582	2.5423	3967
[1.1642 , 1.3252]	0.1610	2.5800	3944
[1.1416 , 1.3044]	0.1628	2.6213	3984
[1.1743 , 1.3307]	0.1564	2.5135	3970
[1.1439 , 1.2997]	0.1558	2.5201	4019
[1.1526 , 1.3064]	0.1539	2.4704	3960

Table 3.2: Cantor distribution in $\mathbb{R}^2$
($\sigma_\mu \approx 1.2619$, $\varepsilon = 0.0021 + 0.0001 \cdot k, k = 0, \ldots, 8$)

confidence intervals	**length of confidence intervals**	$\hat{\sigma}$	**sample size**
[1.1432 , 1.3012]	0.1580	2.5399	3972
[1.1185 , 1.2741]	0.1556	2.5120	4007
[1.1791 , 1.3383]	0.1592	2.5661	3992
[1.1281 , 1.2812]	0.1530	2.4684	3999
[1.1317 , 1.2896]	0.1579	2.5550	4021
[1.1726 , 1.3322]	0.1596	2.5769	4007
[1.1820 , 1.3404]	0.1583	2.5535	3996
[1.1351 , 1.2914]	0.1563	2.5053	3946
[1.2095 , 1.3676]	0.1581	2.5392	3965
[1.1408 , 1.2980]	0.1572	2.5214	3955
[1.1116 , 1.2626]	0.1511	2.4334	3987
[1.1885 , 1.3471]	0.1586	2.5346	3924
[1.1692 , 1.3285]	0.1593	2.5590	3966
[1.1372 , 1.2950]	0.1578	2.5415	3986
[1.2087 , 1.3654]	0.1567	2.5302	4007
[1.1819 , 1.3397]	0.1578	2.5332	3961
[1.1498 , 1.3028]	0.1530	2.4818	4043
[1.1290 , 1.2833]	0.1543	2.4728	3946
[1.1244 , 1.2805]	0.1561	2.5106	3974
[1.1699 , 1.3226]	0.1527	2.4800	4055
[1.1464 , 1.3001]	0.1537	2.4700	3968
[1.1312 , 1.2876]	0.1564	2.4997	3926
[1.1341 , 1.2882]	0.1541	2.4829	3989
[1.1489 , 1.3061]	0.1571	2.5285	3978
[1.1767 , 1.3373]	0.1605	2.5802	3970

Table 3.3: Cantor distribution in $\mathbb{R}^2$
($\sigma_\mu \approx 1.2619$, $\varepsilon = 0.0021 + 0.0001 \cdot k, k = 0, \ldots, 8$)

confidence intervals	**length of confidence intervals**	$\hat{\sigma}$	**sample size**
[1.1608 , 1.3198]	0.1590	2.5615	3990
[1.1322 , 1.2864]	0.1542	2.4914	4010
[1.1446 , 1.3012]	0.1565	2.5097	3950
[1.1275 , 1.2811]	0.1536	2.4732	3984
[1.1916 , 1.3520]	0.1605	2.5699	3941
[1.1510 , 1.3097]	0.1587	2.5563	3986
[1.1870 , 1.3449]	0.1579	2.5512	4009
[1.1679 , 1.3276]	0.1597	2.5730	3991
[1.1716 , 1.3257]	0.1540	2.4850	4000
[1.1509 , 1.3084]	0.1575	2.5304	3965
[1.1460 , 1.3040]	0.1580	2.5557	4021
[1.1072 , 1.2638]	0.1565	2.5145	3965
[1.1173 , 1.2708]	0.1535	2.4692	3976
[1.1728 , 1.3356]	0.1627	2.6113	3956
[1.1351 , 1.2882]	0.1532	2.4806	4031
[1.1548 , 1.3138]	0.1590	2.5515	3956
[1.1220 , 1.2787]	0.1567	2.5286	4002
[1.1142 , 1.2692]	0.1550	2.5150	4046
[1.1196 , 1.2760]	0.1564	2.5259	4007
[1.1754 , 1.3332]	0.1578	2.5391	3978
[1.1241 , 1.2780]	0.1540	2.4819	3993
[1.1797 , 1.3347]	0.1550	2.4977	3990
[1.1437 , 1.2964]	0.1527	2.4619	3994
[1.2072 , 1.3626]	0.1554	2.5181	4033
[1.1402 , 1.2991]	0.1589	2.5548	3970

Table 3.4: Cantor distribution in $\mathbb{R}^2$
($\sigma_\mu \approx 1.2619$, $\varepsilon = 0.0021 + 0.0001 \cdot k, k = 0, \ldots, 8$)

confidence intervals	**length of confidence intervals**	$\hat{\sigma}$	**sample size**
[1.1356 , 1.2873]	0.1518	2.4472	3995
[1.1267 , 1.2818]	0.1552	2.4980	3982
[1.1735 , 1.3331]	0.1595	2.5554	3942
[1.1481 , 1.3028]	0.1547	2.4848	3966
[1.1658 , 1.3233]	0.1575	2.5268	3954
[1.1195 , 1.2687]	0.1492	2.4054	3995
[**1.1055 , 1.2594**]*	0.1539	2.4786	3988
[**1.1067 , 1.2617**]*	0.1550	2.4845	3949
[**1.0639 , 1.2120**]*	0.1481	2.3881	3994
[1.1504 , 1.3089]	0.1585	2.5598	4009
[1.1364 , 1.2916]	0.1553	2.4992	3981
[1.1580 , 1.3138]	0.1558	2.5199	4018
[1.1636 , 1.3203]	0.1567	2.5226	3981
[1.1871 , 1.3427]	0.1555	2.5128	4011
[1.1320 , 1.2858]	0.1538	2.4761	3983
[1.1832 , 1.3457]	0.1625	2.6095	3963
[1.1806 , 1.3380]	0.1575	2.5552	4045
[1.1566 , 1.3131]	0.1565	2.5197	3985
[1.2390 , 1.3993]	0.1604	2.5791	3974
[1.1642 , 1.3169]	0.1527	2.4812	4059
[1.1900 , 1.3474]	0.1575	2.5435	4009
[1.1409 , 1.2981]	0.1572	2.5219	3956
[1.1288 , 1.2804]	0.1516	2.4569	4035
[1.1390 , 1.2949]	0.1559	2.5108	3984
[1.1320 , 1.2875]	0.1555	2.4968	3963

3.3 Generalized Cantor distribution in $\mathbb{R}^3$

In this section we consider a generalization of 3-dimensional Cantor distribution on the Cartesian product $C \times C \times C$. First we describe a construction of this measure by the following procedure.

Consider the unit cube U in $\mathbb{R}^3$ and divide it into 3^3 nonoverlapping equal cubes $U_1, \ldots, U_{27}$. We assign the order numbers to them in each vertical stratum from left to right, beginning with the leftmost bottom cube. Now define a vector of probabilities $\mathbf{p} = (p_1, \ldots, p_{27})$ whose entries are given by $p_1 = 0.8^3$, $p_3 = p_7 = p_{19} = 0.8^2 \cdot 0.2$, $p_9 = p_{21} = p_{25} = 0.8 \cdot 0.2^2$, $p_{27} = 0.2^3$ and $p_i = 0$ otherwise. Further, we assign the probability p_i to U_i.

Repeat this process for each U_i: divide U_i into 3^3 nonoverlapping equal subcubes $U_{i,1}, \ldots, U_{1,27}$ (maintaining the same pattern of labelling the cubes as used at the first stage) and assign probability $p_i p_j$ to $U_{i,j}$. Continuing this procedure, we obtain, for each positive integer n, a collection of 3^{3n} nonoverlapping equal cubes $U_{i_1,\ldots,i_n}$, $1 \leq i_k \leq 27$, with the product probability $p_{i_1} \cdot \ldots \cdot p_{i_n}$ assigned to $U_{i_1,\ldots,i_n}$.

Now there exists a unique probability measure $\mu_{\mathbf{p}}$ defined on the Borel sets of U such that $\mu_{\mathbf{p}}(U_{i_1,\ldots,i_n}) = p_{i_1} \cdot \ldots \cdot p_{i_n}$. This measure is considered here. It is known that its information dimension is approximately equal to 1.3665 (see Cutler [12]).

Note that if the nonzero entries of a vector of probabilities $\mathbf{p}$ would be given by $p_1 = p_3 = p_7 = p_9 = p_{19} = p_{21} = p_{25} = p_{27} = 0.5^3$ then we obtain the three-dimensional Cantor distribution.

We have produced 100 simulations. In each simulation sample of size 5000 was randomly drawn from the generalized three-dimensional Cantor distribution. $\varepsilon_k = 0.030 + 0.001 \cdot k$, $k = 0, \ldots, 8$ were chosen as a radii for balls $B(X_i, \varepsilon)$, for $i = 1, \ldots, 5000$. Then points which did not have any neighbor in their ε_1-neighborhood have been thrown out. Results of the simulations are shown in Tables 3.5-3.8. Intervals which do not cover the true value of dimension are marked by asterisk "$*$".

Table 3.5: Generalized Cantor distribution in $\mathbb{R}^3$
($\sigma_\mu \approx 1.3665$, $\varepsilon = 0.030 + 0.001 \cdot k, k = 0, \ldots, 8$)

confidence intervals	**length of confidence intervals**	$\hat{\sigma}$	**sample size**
[1.3102 , 1.4381]	0.1279	2.2786	4876
[1.2898 , 1.4174]	0.1276	2.2698	4862
[1.2539 , 1.3785]	0.1246	2.2180	4870
[1.3129 , 1.4417]	0.1288	2.2939	4874
[1.3165 , 1.4427]	0.1263	2.2446	4857
[1.3120 , 1.4392]	0.1273	2.2656	4870
[1.3295 , 1.4588]	0.1293	2.2997	4861
[1.2960 , 1.4231]	0.1271	2.2628	4868
[1.3258 , 1.4553]	0.1295	2.2980	4842
[1.2965 , 1.4224]	0.1259	2.2394	4865
[1.3397 , 1.4689]	0.1292	2.2974	4860
[1.3418 , 1.4712]	0.1293	2.3018	4866
[1.2694 , 1.3964]	0.1270	2.2577	4857
[1.3058 , 1.4308]	0.1250	2.2242	4861
[1.2702 , 1.3986]	0.1283	2.2788	4847
[1.2844 , 1.4121]	0.1277	2.2709	4861
[1.2658 , 1.3918]	0.1261	2.2358	4832
[1.2983 , 1.4213]	0.1230	2.1880	4861
[1.2823 , 1.4128]	0.1305	2.3249	4880
[1.3626 , 1.4921]	0.1295	2.3009	4852
[1.2527 , 1.3800]	0.1273	2.2596	4844
[1.2996 , 1.4294]	0.1298	2.3092	4863
[1.2899 , 1.4131]	0.1231	2.1890	4855
[1.3361 , 1.4627]	0.1266	2.2488	4851
[**1.2406 , 1.3656**]*	0.1249	2.2217	4858

Table 3.6: Generalized Cantor distribution in $\mathbb{R}^3$
($\sigma_\mu \approx 1.3665$, $\varepsilon = 0.030 + 0.001 \cdot k, k = 0, \ldots, 8$)

confidence intervals	**length of confidence intervals**	$\hat{\sigma}$	**sample size**
[1.3375 , 1.4634]	0.1259	2.2354	4847
[1.3268 , 1.4542]	0.1275	2.2680	4865
[1.3215 , 1.4473]	0.1258	2.2380	4863
[1.2984 , 1.4260]	0.1276	2.2722	4874
[1.2934 , 1.4171]	0.1237	2.1968	4845
[1.2990 , 1.4283]	0.1293	2.3012	4866
[1.3030 , 1.4306]	0.1276	2.2668	4849
[1.3145 , 1.4428]	0.1283	2.2796	4853
[1.3277 , 1.4566]	0.1288	2.2952	4876
[1.3301 , 1.4595]	0.1294	2.3036	4869
[1.3853 , 1.5122]*	0.1269	2.2593	4868
[1.2585 , 1.3852]	0.1266	2.2490	4848
[1.3436 , 1.4705]	0.1269	2.2548	4853
[1.3164 , 1.4457]	0.1294	2.2943	4834
[1.3551 , 1.4863]	0.1312	2.3322	4858
[1.3119 , 1.4413]	0.1293	2.3006	4862
[1.3038 , 1.4303]	0.1265	2.2526	4873
[1.3326 , 1.4549]	0.1223	2.1745	4859
[1.2902 , 1.4168]	0.1266	2.2547	4874
[1.3328 , 1.4599]	0.1271	2.2578	4849
[1.2788 , 1.4064]	0.1276	2.2714	4866
[1.2824 , 1.4079]	0.1255	2.2313	4855
[1.3084 , 1.4366]	0.1281	2.2834	4879
[1.3440 , 1.4699]	0.1259	2.2352	4847
[1.3132 , 1.4400]	0.1268	2.2565	4866

Table 3.7: Generalized Cantor distribution in $\mathbb{R}^3$
($\sigma_\mu \approx 1.3665$, $\varepsilon = 0.030 + 0.001 \cdot k, k = 0, \ldots, 8$)

confidence intervals	**length of confidence intervals**	$\hat{\sigma}$	**sample size**
[1.2637 , 1.3903]	0.1267	2.2520	4858
[1.2913 , 1.4189]	0.1276	2.2664	4846
[1.2819 , 1.4063]	0.1244	2.2164	4874
[1.2870 , 1.4150]	0.1280	2.2770	4863
[1.2675 , 1.3925]	0.1249	2.2226	4862
[1.3176 , 1.4440]	0.1264	2.2534	4883
[1.2394 , 1.3663]*	0.1270	2.2569	4856
[1.3378 , 1.4683]	0.1305	2.3211	4858
[1.2637 , 1.3886]	0.1250	2.2256	4873
[1.3256 , 1.4532]	0.1276	2.2689	4858
[1.3307 , 1.4586]	0.1279	2.2747	4860
[1.3108 , 1.4369]	0.1261	2.2435	4864
[1.2746 , 1.4029]	0.1283	2.2841	4867
[1.3005 , 1.4281]	0.1276	2.2725	4872
[1.3183 , 1.4474]	0.1291	2.2900	4835
[1.3035 , 1.4302]	0.1268	2.2536	4855
[1.3595 , 1.4908]	0.1313	2.3341	4857
[1.2658 , 1.3934]	0.1277	2.2697	4857
[1.2842 , 1.4083]	0.1241	2.2045	4851
[1.3016 , 1.4269]	0.1254	2.2277	4851
[1.3428 , 1.4735]	0.1306	2.3194	4846
[1.3090 , 1.4364]	0.1275	2.2664	4859
[1.3377 , 1.4680]	0.1303	2.3199	4874
[1.2905 , 1.4202]	0.1296	2.3041	4854
[1.3235 , 1.4506]	0.1271	2.2650	4882

Table 3.8: Generalized Cantor distribution in $\mathbb{R}^3$
($\sigma_\mu \approx 1.3665$, $\varepsilon = 0.030 + 0.001 \cdot k, k = 0, \ldots, 8$)

confidence intervals	length of confidence intervals	$\hat{\sigma}$	sample size
[1.3353 , 1.4645]	0.1292	2.2991	4863
[1.3349 , 1.4596]	0.1247	2.2254	4890
[1.3229 , 1.4519]	0.1290	2.2927	4850
[1.3632 , 1.4904]	0.1271	2.2579	4848
[1.3382 , 1.4675]	0.1293	2.2977	4854
[1.3004 , 1.4284]	0.1280	2.2755	4855
[1.2926 , 1.4226]	0.1300	2.3137	4867
[1.2507 , 1.3795]	0.1288	2.2885	4852
[1.3183 , 1.4489]	0.1306	2.3230	4864
[1.3438 , 1.4757]	0.1318	2.3425	4852
[1.3057 , 1.4318]	0.1262	2.2453	4867
[1.3479 , 1.4753]	0.1274	2.2663	4860
[1.3446 , 1.4714]	0.1267	2.2546	4862
[1.3498 , 1.4804]	0.1306	2.3260	4877
[1.3019 , 1.4284]	0.1265	2.2506	4863
[1.3369 , 1.4671]	0.1302	2.3113	4841
[1.2899 , 1.4193]	0.1294	2.3002	4857
[1.3099 , 1.4381]	0.1282	2.2758	4844
[1.2979 , 1.4244]	0.1266	2.2481	4846
[1.3259 , 1.4550]	0.1290	2.2939	4857
[1.3480 , 1.4757]	0.1277	2.2733	4870
[1.3161 , 1.4433]	0.1272	2.2594	4849
[1.2961 , 1.4221]	0.1260	2.2449	4877
[1.3177 , 1.4457]	0.1280	2.2766	4858
[1.2651 , 1.3908]	0.1257	2.2360	4860

Chapter 4

ASCLT and ASLT for U-statistics

4.1 U-statistics of dependent random variables

Here we give some preliminary results which will be used in the next two sections. First, we recall some definitions of mixing coefficients. Let

$$X_1, X_2, \ldots, X_n \tag{4.1}$$

be a strictly stationary sequence of random variables defined on probability space $(\Omega, \Im, P)$. Denote the distribution function of X_i by $F(x)$. Further, let $\Im_a^b$ denote a σ-algebra of events generated by $X_a, X_{a+1}, \ldots, X_b$, where $1 \le a \le b < \infty$, and $\Im_a^\infty$ denote a σ-algebra generated by $X_a, X_{a+1}, \ldots$. If σ-algebras $\Im_1^k$ and $\Im_{k+n}^\infty$ are independent then for all events $A \in \Im_1^k$ and $B \in \Im_{k+n}^\infty$ we have

$$P(A \cap B) - P(A)P(B) = 0. \tag{4.2}$$

In general, (4.2) doesn't hold, and the expression on the left hand side of (4.2) is taken as a basis of the measure of dependence between $\Im_1^k$ and $\Im_{k+n}^\infty$.

Definition 4.1. *The sequence (4.1) is called strongly mixing or α-mixing if*

$$\alpha(n) := \sup\{|P(A \cap B) - P(A)P(B)| : A \in \Im_1^k, B \in \Im_{k+n}^\infty, k = 1, 2, \ldots\}$$

tends to zero as $n \to \infty$.

Definition 4.2. *The sequence (4.1) is called absolutely regular if*

$$\beta(n) := \sup_{k \in N} E[\sup\{|P(A/\Im_1^k) - P(A)| : A \in \Im_{k+n}^\infty\}]$$

tends to zero as $n \to \infty$.

It is well known that

$$\alpha(n) \le \beta(n) \tag{4.3}$$

and therefore a class of all sequences of strongly mixing random variables contains a class of all sequences of absolutely regular random variables.
Consider a statistical functional $\theta(F)$ defined in (1.10) and take a non-degenerate U-statistic U_n based on the sample (4.1) as an estimator for $\theta(F)$. It should be mentioned that U_n is not an unbiased estimator for $\theta(F)$ anymore, but it is still a consistent and asymptotical normal estimator for a wide class of dependent variables (see SEN [47], YOSHIHARA [51], DENKER and KELLER [17], YOSHIHARA [52]).
In this chapter we make use of the functions $\tilde{h}_c$ and h_c defined in (1.11)–(1.14) for $c = 1, \ldots, m$. It is not difficult to see that the functions h_c, $c = 1, \ldots, m$ are degenerate even if (4.1) is not a sequence of i.i.d. random variables. Note that if U_n is based on the independent sequence (4.1) then the functions $\tilde{h}_c$ can also be written through conditional expectations, but in general, this cannot be done.
Further, HOEFFIDNG's decomposition is valid for U_n based on (4.1) and therefore, Theorem 1.2 holds in the context of this chapter. But the upper bounds of the second moment of U_{nc}, $c = 1, \ldots, m$ defined in (1.16) and based on (4.1), differ from (1.18) and depend on the type of the dependency of the sequence (4.1). We give them in the corresponding sections.
Next, we introduce the value

$$\sigma^2 = Eh_1^2(X_1) + 2\sum_{k=1}^{\infty} Eh_1(X_1)h_1(X_{k+1}) \ , \tag{4.4}$$

which is well defined under conditions given in the next two sections.
The main idea of the proofs in Section 4.2 and Section 4.3 is to approximate $\sqrt{n}(\sigma m)^{-1}(U_n - \theta(F))$ by a sequence of random variables W_n which satisfies the ASCLT and obtain the following bound for the probabilities

$$P\left(\left|\sqrt{n}(\sigma m)^{-1}(U_n - \theta(F)) - W_n\right| > \epsilon\right) = O\left(\frac{1}{(\log n)^{\tau}}\right) \tag{4.5}$$

for all $\epsilon > 0$ and some $\tau > 0$. Then $(\sqrt{n}(\sigma m)^{-1}(U_n - \theta(F))$ also satisfies the ASCLT by virtue of the next Lemma which is due to LESIGNE [39]. Here we cite a bit modified Preposition 1 from [39].

Lemma 4.1. *Let $\sqrt{n}(\sigma m)^{-1}(U_n - \theta(F))$ and W_n be two sequences of random variables. Assume that the sequence W_n satisfies the ASCLT. Further, let (4.5) holds. Then the sequence $\sqrt{n}(\sigma m)^{-1}(U_n - \theta(F))$ also satisfies the ASCLT.*

If U_n is a non-degenerate U-statistic then the natural choice of the approximating sequence W_n is $\sqrt{n}\sigma^{-1}U_{n1}$, where U_{n1} is the first term of HOEFFDING's decomposition (1.15). Since U_{n1} is a sample average of sequence $\{h_1(X_i)\}_{i=1}^n$ the next lemma which is due to PELIGRAD and SHAO [44] is used for showing that the approximating sequence W_n satisfies the ASCLT.

Lemma 4.2. *Let $\{X_n, n \geq 1\}$ be a stationary strongly mixing sequence with $EX_1 = 0$ and $E|X_1|^{2+\delta} < \infty$ for some $\delta > 0$. Assume*

$$\sum_{n=1}^{\infty} \alpha^{\delta/(2+\delta)}(n) < \infty.$$

Then $\sigma_1^2 = EX_1^2 + 2\sum_{k=2}^{\infty} EX_1X_k < \infty$. If in addition $\sigma_1^2 > 0$ then $(\sqrt{n}\sigma_1)^{-1}S_n$ satisfies the ASCLT, where $S_n = \sum_{i=1}^n X_i$.

4.2 ASCLT for U-statistics of absolutely regular random variables

In this section, we prove the ASCLT for U-statistics of absolutely regular random variables. Assume that sequence (4.1) is absolutely regular with mixing coefficients $\beta(n)$.

First, we give upper bounds for the second moment of degenerate statistics U_{nc}, $c = 2, \ldots, m$, which are due to YOSHIHARA [51].

Assume that for some $r > 2$

$$\mu_r = \int \cdots \int |h(x_1, \ldots, x_m)|^r dF(x_1) \cdots dF(x_m) \leq M_0 < \infty \tag{4.6}$$

and for all integers $i_1, i_2, \ldots, i_m$ $(i_1 < i_2 < \cdots < i_m)$

$$\nu_r = E|h(X_{i_1}, X_{i_2}, \ldots, X_{i_m})|^r \leq M_0 < \infty \ , \tag{4.7}$$

where M_0 is some absolute constant.

Lemma 4.3. *If there is a positive δ such that for $r = 2+\delta$ (4.6) and (4.7) hold, and for some δ' $(0 < \delta' < \delta)$ $\beta(n) = O(n^{-(2+\delta')/\delta'})$, then we have*

$$E(U_n(h_c))^2 = O(n^{-1-\gamma}) \quad (2 \le c \le m)$$

where $\gamma = \frac{2(\delta-\delta')}{\delta'(2+\delta)} > 0$.

Now we are ready to state our theorem and give its simple proof.

Theorem 4.1. *Let $\{X_n, n \ge 1\}$ be a stationary absolutely regular sequence of random variables. If there is a positive δ such that for $r = 2+\delta$ (4.6) and (4.7) hold, and for some δ' $(0 < \delta' < \delta)$ $\beta(n) = O(n^{-(2+\delta')/\delta'})$, then the series (4.4) converges absolutely; if $\sigma^2 > 0$ holds then $\sqrt{n}(\sigma m)^{-1}(U_n(h) - \theta)$ satisfies the ASCLT.*

Remark 4.1. *The assumptions of Theorem 4.1 ensure the central limit theorem for U-statistics of absolutely regular random variables (see* Yoshihara *[51]).*

Proof. The first statement of the theorem was proved in Theorem 1 of Yoshihara [51] and hence, σ^2 in (4.4) is well defined.
From the Hoeffding's decomposition (1.15) it follows that

$$\frac{n^{1/2}}{\sigma m}(U_n(h) - \theta) - \frac{n^{1/2}}{\sigma}U_n(h_1) = \frac{n^{1/2}}{\sigma m}\sum_{k=2}^{m}\binom{m}{k}U_n(h_k). \tag{4.8}$$

Note that if the sequence $\{X_n\}$ is absolutely regular (β-mixing) then it is strongly mixing (α-mixing) which follows from (4.3). Secondly, if $\{X_n\}$ is β-mixing then a sequence $\{f(X_n)\}$ is also β-mixing for any measurable function f with some mixing coefficient which is not greater than $\beta(n)$.
These remarks give us a possibility of applying Lemma 4.2 to the second term on the left hand side of (4.8) and therefore, $n^{1/2}\sigma^{-1}U_n(h_1)$ satisfies the ASCLT.
Lemma 4.3 yields

$$E\left(\frac{n^{1/2}}{\sigma m}U_n(h_c)\right)^2 = O(n^{-\gamma}) \tag{4.9}$$

for any $c = 2, \dots, m$, where $\gamma = \frac{2(\delta-\delta')}{\delta'(2+\delta)} > 0$.
Now (4.5) follows from (4.8) and (4.9) using Chebyshev and c_r- inequalities. Thus, the assumptions of Lemma 4.1 are fulfilled and its application to the random variables on the left hand side of (4.8) completes the proof.

□

4.3 ASCLT for U-statistics of strongly mixing random variables

In this section we prove the ASCLT for U-statistics of strongly mixing random variables. Assume that sequence (4.1) is strongly mixing with mixing coefficients $\alpha(n)$.

YOSHIHARA [52] investigated an asymptotic distribution of U_n based on (4.1) when all degenerate kernels $h_c,\ c = 2, \ldots, m$ can be represented as a Fourier sum by some orthogonal basis of $L^2(\mathbb{R}, F)$ defined below. We make use of his method for deriving the ASCLT for U_n.

Let $L^2(\mathbb{R}, F)$ be the Hilbert space of square integrable functions with respect to a distribution function F of X_1. Further, let $\{g_i\}_{i=0}^{\infty}$ be an orthogonal basis of $L^2(\mathbb{R}, F)$ such that $g_0 = 1$. For each c $(1 \le c \le m)$ put

$$g_{i_1,\ldots,i_c}(x_1, \ldots, x_c) = \prod_{j=1}^{c} g_{i_j}(x_j).$$

It is well known that for each c $(1 \le c \le m)$ the system $\{g_{i_1,\ldots,i_c} : 0 \le i_1 < i_2 < \ldots < i_c < \infty\}$ is a basis of the Hilbert space $L^2(\mathbb{R}^c, F^c)$. Let $\lambda_c(i_1, \ldots, i_c)$ be the Fourier coefficient of the function h_c, i.e.,

$$\lambda_c(i_1, \ldots, i_c) = \int \cdots \int h_c(x_1, \ldots, x_c) \prod_{j=1}^{c} g_{i_j}(x_j) \prod_{j'=1}^{c} dF(x_{j'}).$$

Then, for each c $(1 \le c \le m)$

$$h_c(x_1, \ldots, x_c) = \sum_{\{i_j\}} \lambda_r(i_1, \ldots, i_c) g_{i_1,\ldots,i_c}(x_1, \ldots, x_c) \tag{4.10}$$

in the L^2-sense. Moreover, it follows from the Parseval inequality that

$$\int \cdots \int h_c^2(x_1, \ldots, x_c) \prod_{j=1}^{c} dF(x_j) = \sum |\lambda_c(i_1, \ldots, i_c)|^2 < \infty.$$

Define

$$\lambda(i_1, \ldots, i_c, 0, \ldots, 0) = \lambda_c(i_1, \ldots, i_c) \quad (1 \le c \le m-1)$$

and

$$\lambda(i_1, \ldots, i_m) = \lambda_m(i_1, \ldots, i_m).$$

One of the main assumptions in this section is

$$\sum |\lambda(i_1, \ldots, i_m)| < \infty \ . \tag{4.11}$$

Let

$$\sigma_n^2 = E\left(\sum_{j=1}^{n} h_1(X_j)\right)^2 \ . \tag{4.12}$$

Under the conditions of Lemma 4.4 given below, the series in (4.4) converges absolutely and hence, σ^2 is well defined. Moreover,

$$\sigma_n^2 = n\sigma^2(1 + o(1)) \quad \text{as} \quad n \to \infty \ .$$

We use the following notation. For any random variables η and $r \geq 1$, let $\|\eta\|_r = (E|\eta|^r)^{1/r}$ if $E|\eta|^r < \infty$.
Now assume that for some $\delta > 0$

$$\|h_1(X_1)\|_{2+\delta/m} < \infty \tag{4.13}$$

and

$$\max_{2\leq c\leq m} \sup_{k\geq 1} \|g_{k,c}(X_1)\|_{2m+\delta} < \infty. \tag{4.14}$$

The next lemma gives the upper bound for the second moments of U_{nc}, $c = 2, \ldots, m$, which is due to YOSHIHARA [52].

Lemma 4.4. *Let $\{X_n, n \geq 1\}$ be a stationary sequence of strongly mixing random variables with mixing coefficients $\alpha(n)$. Suppose that (4.10) holds for all c $(2 \leq c \leq m)$ and (4.11) is satisfied. Assume that there exists a positive δ such that (4.13) and (4.14) hold. Further, assume that*

$$\sum_{n=1}^{\infty} n^m \alpha^{\delta/(2m+\delta)}(n) < \infty \ .$$

Then

$$E|U_{nc}|^2 \leq C(r)n^{-r} \quad r = 2, \ldots, m$$

where $C(2), \ldots, C(m)$ are absolute constants which do not depend on n.

Now we state the ASCLT for the U-statistic U_n.

Theorem 4.2. *If $\sigma^2 > 0$ then under the assumptions of Lemma 4.4, $\sqrt{n}\sigma^{-1}(U_n(h) - \theta)$ satisfies the ASCLT.*

Proof. First note that if X_n is a strongly mixing sequence with mixing coefficients $\alpha(n)$ then for any measurable function f, $f(X_n)$ is also a strongly mixing sequence with some mixing coefficient which is not greater than $\alpha(n)$.
Now from the Hoeffding decomposition, we have that

$$\frac{n^{1/2}}{\sigma}(U_n(h) - \theta) - \frac{n^{1/2}}{\sigma} m U_n(h_1) = \frac{n^{1/2}}{\sigma} \sum_{k=2}^{m} \binom{m}{k} U_n(h_k). \tag{4.15}$$

By Lemma 4.2 and the remark at the beginning of the proof, we can conclude that the second term on the left hand side of (4.15) satisfies the ASCLT.
Using Lemma 4.4 we find out that for any $c = 2, \ldots, m$

$$E\left(\frac{n^{1/2}}{\sigma} U_n(h_c)\right)^2 = O(n^{1-c}) \ .$$

Thus, the assumptions of Lemma 4.1 are fulfilled and its application to random variables on the left hand side of (4.15) completes the proof. □

4.4 Refinement of BERKES and CSÁKI's Theorem

In the rest of this chapter we present another technique, developed by HOLZMANN, KOCH and MIN [32], for proving the ASLT for U-statistics based on i.i.d. random variables $X_1, \ldots, X_n$. Here we describe this method and improve Theorem F of BERKES and CSÁKI [4].
Let $(Y_n)_{n\geq 1}$ be a sequence of random elements taking values in a Polish space $(\mathcal{S}, d)$ and let G be a probability measure on the Borel σ-field in $\mathcal{S}$. We say that $(Y_n)_{n\geq 1}$ satisfies the ASLT with a limiting distribution G if

$$(\log n)^{-1} \sum_{k=1}^{n} \delta_{Y_k}/k \Rightarrow G \quad \text{as } n \to \infty$$

with probability 1. Here δ_{Y_k} is the Dirac measure at Y_k and " $\Rightarrow$ " denotes weak convergence of measures.
The following lemma illustrates the main idea of the method.

Lemma 4.5. *Let $(Y_n)_{n\geq 1}$ be a sequence of $\mathcal{S}$-valued random elements which satisfies the ASLT with some limiting distribution G. Assume that Z_n is another sequence of $\mathcal{S}$-valued random elements on the same probability space such that almost surely, $d(Y_n, Z_n) \to 0$. Then Z_n also satisfies the ASLT with limiting distribution G.*

Proof. By a well known principle in almost sure limit theory (see e.g. Lacey and Philipp [38], $(Y_n)_{n\geq 1}$ satisfies the ASLT with limiting distribution G if and only if

$$(\log n)^{-1} \sum_{k=1}^{n} \frac{1}{k} \Psi\big(Y_k(\omega)\big) \to \int_{\mathcal{S}} \Psi(x)\, dG(x) \quad a.s.$$

for any bounded Lipschitz function Ψ. Using the Lipschitz property of Ψ and the assumption that $d(Y_n, Z_n) \to 0$ we conclude that

$$\frac{1}{\log n} \left| \sum_{k=1}^{n} \frac{1}{k} \Big[\Psi\big(Y_k(\omega)\big) - \Psi\big(Z_k(\omega)\big) \Big] \right| \leq \frac{C}{\log n} \sum_{k=1}^{n} \frac{1}{k} d\big(Y_k(\omega), Z_k(\omega)\big) \to 0 \quad a.s.,$$

where C is a Lipschitz constant for Ψ. This proves the Lemma. □

In the sequel we will make use of the following lemma which is a consequence of a more general result due to Gine and Zinn [21].

Lemma 4.6. *Let $h(x_1, \ldots, x_l)$ be measurable and degenerate. Let $q \in (\frac{l}{2}, l)$. If*

$$E|h(X_1, \ldots, X_l)|^{l/q} < \infty, \tag{4.16}$$

then with probability 1

$$n^{-q} \sum_{1 \leq i_1 < \ldots < i_l \leq n} h(X_{i_1}, \ldots, X_{i_l}) \to 0.$$

For the weak convergence of $n^{c/2} U_n(h)$ (where c denotes the rank of $U_n(h)$) Korolyuk and Borovskich [37] weakened the assumption (1.17) to

$$E|h_i(X_1, \ldots, X_i)|^{2k/(2k-c)} < \infty, \; i = c, \ldots, m. \tag{4.17}$$

Now we show that these conditions also imply the validity of the ASLT for U-statistics of i.i.d. random variables.

Theorem 4.3. *Let c be the rank of the U-statistic $U_n(h)$ based on i.i.d. random variables $X_1, \dots, X_n$. If (4.17) is satisfied then the following relation holds*

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{1}\left\{k^{r/2}(U_k(h) - \theta(F)) < x\right\} = G(x) \text{ a.s. for any } x \in C_G,$$

where G is the limit distribution of $n^{r/2}(U_n(h) - \theta(F))$, C_G denotes the set of continuity points of G and $\theta(F)$ defined in (1.10).

Remark 4.2. *Note that the moment assumptions of the Theorem imposed on degenerate functions h_i, $i = c, \dots, m$ are weaker than the moment assumption (1.17) of* Berkes *and* Csáki *[4].*

Proof. First of all note that, if c is the critical parameter of $U_n(h)$, then the functions $h_i(x_1, \dots, x_i) = 0$ a.s. for all $i = 1, \dots, c-1$, and so from Hoeffding's decomposition

$$n^{c/2}U_n(h) - n^{c/2}\binom{m}{c}U_n(h_c) = n^{c/2}\sum_{k=c+1}^{m}\binom{m}{k}U_n(h_k). \tag{4.18}$$

Using Theorem F of Berkes and Csáki [4], we conclude that $n^{c/2}\binom{m}{c}U_n(h_c)$ satisfies the ASLT.
By virtue of Lemma 4.6, the sum on the right-hand side of (4.18) converges to zero a.s. by letting $l = k$ and $q = k - c/2$ for $k = c+1, \dots, m$. An application of Lemma 4.5 to the random variables on the left hand side of (4.18) completes the proof. □

4.5 ASLT for U-statistics with limiting stable distribution

As we shall see in this section, under some mild moment conditions weak convergence of a sequence of non-degenerate U-statistics to a stable limit distribution implies the validity of the corresponding ASLT. Let G_α denote a stable law of order α, $0 < \alpha \leq 2$.

Theorem 4.4. *Let $U_n(h)$ be a U-statistic based on i.i.d. random variables $X_1, \dots, X_n$. Assume that for some $\alpha \in (1,2]$*

$$\frac{n^{1-\frac{1}{\alpha}}}{mL(n)} U_n(h) - A_n \Rightarrow G_\alpha, \tag{4.19}$$

where $L(n)$ is a slowly varying function for which $\liminf_{n\to\infty} L(n) > 0$. If

$$E|h_k(X_1, \dots, X_k)|^{\frac{\alpha k}{\alpha(k-1)+1}} < \infty, \quad k = 2, \dots, m, \tag{4.20}$$

then

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{1}_{\left\{\frac{k^{1-1/\alpha}}{mL(k)} U_k(h) - A_k < x\right\}} = G_\alpha(x) \quad a.s.$$

Remark 4.3. **(i)** *Assumption (4.19) is very common in almost sure limit theory, when one deduces an ASLT from the validity of the corresponding weak limit theorem (see e.g.* Berkes *and* Dehling *[5]).*
(ii) *One has weak convergence in (4.19) if the distribution function of $h_1(X_1)$ belongs to the domain of attraction of G_α and if the moment condition*

$$E|h(X_1, \dots, X_m)|^{\frac{2\alpha}{\alpha+1}} < \infty \tag{4.21}$$

holds (see Heinrich *and* Wolf *[28]).*
(iii) *It is not difficult to see that (4.21) implies (4.20).*
(iv) *Theorem 4.4 will be true for any slowly varying function $L(n)$, if $E|h_k(X_1, \dots, X_k)|^{p_k} < \infty$ for some $p_k > \frac{\alpha k}{\alpha(k-1)+1}$, $k = 2, \dots, m$.*

Proof. Let us start by showing that

$$\frac{n^{1-\frac{1}{\alpha}}}{mL(n)} \big(U_n(h) - mU_n(h_1)\big) \to 0 \quad a.s. \tag{4.22}$$

Indeed

$$\frac{n^{1-\frac{1}{\alpha}}}{mL(n)} \big(U_n(h) - mU_n(h_1)\big) = \sum_{k=2}^{m} \binom{m}{k} \frac{n^{1-\frac{1}{\alpha}}}{mL(n)} U_n(h_k), \tag{4.23}$$

and letting $q = k - 1 + \frac{1}{\alpha}$ and $l = k$ we can apply Lemma 4.6. This proves (4.22). Making use of (4.19) and (4.22) we conclude that also

$$\frac{n^{1-\frac{1}{\alpha}}}{L(n)} U_n(h_1) - A_n \Rightarrow G_\alpha. \tag{4.24}$$

It is known that weak convergence of normalized sums of real valued i.i.d. random variables to some stable law G_α implies the corresponding ASLT (see PELIGRAD and RÉVÉSZ [43]). Hence (4.24) implies

$$\lim_{n\to\infty} \frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} \mathbb{1}_{\{\frac{k^{1-\frac{1}{\alpha}}}{L(k)} U_k(h_1) - A_k < x\}} = G_\alpha(x) \quad a.s.$$

An application of Lemma 4.5 to the random variables on the left hand side of (4.23) completes the proof. □

Bibliography

[1] M. Atlagh, M. Weber: *Le théorème central limite presque sûr.* Expositiones Mathematicae **18** (2000), 97–126.

[2] R.H. Berk: *Limiting behavior of posterior distributions when the model is incorrect.* Ann. Math. Statist. **37** (1966), 51–57.

[3] I. Berkes: *Results and problems related to the pointwise central limit theorem.* In: Asymptotic Methods in Probability and Statistics. Ed.: B. Szyszkowicz, Elsevier, Amsterdam (1998) pp. 59–96.

[4] I. Berkes, F. Csáki: *A universal result in almost sure limit theory.* Stoch. Proc. Appl. **94** (2001), 105–134.

[5] I. Berkes, H. Dehling: *Some limit theorems in log density.* Ann. Probab. **21** (1993), 1640–1670.

[6] I. Berkes, H. Dehling, T.F. Móri: *Counterexamples related to the a.s. central limit theorem.* Studia Sci. Math. Hung. **26** (1991), 153–164.

[7] A. Besiokovič: *On the fundamental geometrical properties of linearly measurable plane sets of points.* Math. Ann. **98** (1928), 422–464.

[8] G.A. Brosamler: *An almost everywhere central limit theorem.* Math. Proc. Camb. Phil. Soc. **104** (1988), 561–574.

[9] E. Brunner, M. Denker: *Rank statistics under dependent observations and applications to factorial designs.* J. Statist. Plann. Inference **42 N. 3** (1994), 353–378.

[10] C. Cutler: *Some results on the behaivoir and estimation of the fractal dimensions of distributions on attractors.* J. Stat. Physics **62** (1991), 651–708. Errata and Addenda (1991), **65**, pp. 417–418.

[11] C. Cutler: k^{th} *nearest neighbors and the generalized logisitic distribution.* In: Handbook of the Logistic Distribution **28** (1992). Ed.: N. Balakrishnan, pp. 512–522.

[12] C. Cutler: *A review of the theory and estimation of fractal dimension.* In: Nonlinear Time Series and Chaos **1** (1993), Dimension Estimation and Models. Ed.: H. Tong, pp. 1–107.

[13] C. Cutler, D. Dawson: *Estimation of dimension for spatially-distributed data and related limits theorems.* J. Multiv. Anal. **28** (1989), 115–148.

[14] C. Cutler, D. Dawson: *Nearest neighbor analysis of a family of fractal distributions.* Ann. Probab. **18** (1990), 256–271.

[15] H. Dehling: *Complete convergence of triangular arrays and the law of the iterated logarithm for U-statistics.* Stat. Probab. Letters **7** (1989), 319-321.

[16] M. Denker: *Asymptotic theory in nonparametric statistics.* Vieweg & Sohn, Braunschweig (1985).

[17] M. Denker, G. Keller : *On U-statistics and von Mises' statistics for weakly dependent processes.* Z. Wahrscheinlichketstheorie verw. Gebiete **64** (1983), 505-522.

[18] M. Denker, G. Keller : *Rigorous Statistical Procedures for data from Dynamical systems.* J. Stat. Physics **44, N. 1/2** (1986), 67-93.

[19] P. Erdös, G.A. Hunt : *Changes of signs of sums of random variables.* Pacific J. Math. **3** (1953), 673-687.

[20] A. Fisher : *A pathwise central limit theorem for random walks.* Preprint (1989).

[21] E. Gine, J. Zinn : *Marchinkiewicz type laws of large numbers and convergence of moments for U-statistics.* In: Probability in Banach Spaces **8**. Ed.: R. M. Dudley, M. G. Hahn and J. Kuelsbs, Birkhäuser (1992), pp. 273–291.

[22] P. Grassberger, I. Procaccia : *Characterization of strange attractors.* Physical Review Letters **50, N. 5** (1983), 346–349.

[23] J. Guckenheimer : *Dimension estimates for attractors.* Contemporary Mathematics **28** (1984), 357–367.

[24] A. Hamann: *Methoden zur Stabilisierung von Dimensionsschätzungen.* PhD Thesis, Göttingen, (1998).

[25] P. Halmos: *The foundations of probability.* Amer. Math. Monthly **51, N 9** (1944), 493–510.

[26] P. Halmos: *The theory of unbiased estimation.* Ann. Math. Statist. **17** (1946), 34–43.

[27] F. Hausdorff: *Dimension und äusseres mass.* Math. Ann. **79** (1919), 157–179.

[28] L. Heinrich, W. Wolf: *On the convergence of U-statistics with stable limit distribution.* J. Multiv. Anal. **44** (1993), 226–278.

[29] H.G.E. Hentschel, I. Procaccia: *The infinite number of generalized dimensions of fractals and strange attractors.* Physica D **8** (1993), 435–444.

[30] W. Hoeffding: *A class of statistics with asymptotical normal distribution.* Ann. Math. Statist. **19** (1948), 293-325.

[31] W. Hoeffding: *The strong law of large numbers for U-statistics.* University of North Carolina Mimeo Report **N. 302,** (1961).

[32] H. Holzmann, S. Koch, A. Min: *Almost sure limit theorems for U-statistics.* Stat. Probab. Letters (2004), *to appear* .

[33] P. L. Hsu, H. Robbins: *Complete convergence and the law of large numbers.* Proc. Nat. Acad. Sci. U. S. A. **33, N. 2** (1947), 229-240.

[34] V. Jarnik: *Diophantische Approximationen und Hausdroff Maß* . Mathematicheskii Sbornik **36** (1929), 371–382.

[35] G. Keller: *A new estimator for the information dimension with standard errors and confiedence intervals.* Stoch. Proc. Appl. **71** (1997), 187–206.

[36] A. N. Kolmogorov: *On certain asymptotic characteristics of completely bounded metric spaces.* Dokl. Akad. Nauk SSSR **108** (1956), 385–385 (In Russian).

[37] V. Koroljuk, Yu. Borovskikh: *Theory of U-statistics.* Kluwer Academic Publishers Group, Dordrecht (1994).

[38] M. Lacey, W. Philipp: *A note on the almost sure central limit theorem.* Stat. Probab. Letters **9** (1990), 201-205.

[39] M. Lesigne: *Almost sure central limit theorem for strictly stationary processes.* Proc. of Amer. Math. Soc. **128 N.6** (1999), 1751-1759.

[40] P. Lévy: *Théorie de l'addition des variables aléatoires.* Gauthier-Villars, Paris (1937).

[41] P. Mattila: *Geometry of sets and measures in euclidean spaces.* Cambridge University Press, Cambridge (1995).

[42] E. Ott, W.D. Withers, J.A. Yorke : *Is the dimension of chaotic attractors invariant under coordinate changes?* Journal of Statistical Physics **36** (1984), 687-697.

[43] M. Peligrad, P. Révész: *On the almost sure limit theorem.* In: Almost everywhere convergence II. Ed: A.Bellow, R.Jones, Academic Press, Boston (1991), 209-225.

[44] M. Peligrad, Qi-Man Shao: *A note on the almost sure central limit theorem for weakly dependent random variables.* Stat. Probab. Letters **22** (1995), 131-136.

[45] A. Rényi: *On the dimension and entropy of probability distributions.* Acta. Math. Acad. Sci. Hungar. **10** (1959), 193-215.

[46] P. Schatte: *On strong versions of the central limit theorem.* Math. Nachr. **137** (1988), 249–256.

[47] P.K. Sen: *Limiting behavior of regular functionals of empirical distributions for stationary ∗-mixing processes.* Z. Wahrscheinlichketstheorie verw. Gebiete **25** (1972), 71–82.

[48] D. Sullivan: *Entropy, Hausdorff measures old and new, and limit sets of geometrically finite Kleinian groups.* Acta Math. **153** (1984), 259–277.

[49] S. Taylor, C. Tricot: *Packing measure, and its evalution for a Brownian path.* Trans. Amer. Math. Soc. **288** (1985), 679–699.

[50] R. von Mises: *On the asymptotic distribution of differentiable statistical functions.* Ann. Math. Statist. **18** (1947), 309–348.

[51] Ken-ichi Yoshihara: *Limiting behavior of U-statistics for stationary, absolutely regular processes.* Z. Wahrscheinlichkeitstheorie verw. Gebiete **35** (1976), 237-252.

[52] Ken-ichi Yoshihara: *Limiting behavior of U-statistics for strongly mixing sequences.* Yokohama Math. Journal **39** (1992), 107-113.

Curriculum Vitae

15.12.1972	born in Koshkupir district of Khorezm region, Uzbekistan
1979-1987	Middle School N1 of Koshkupir district, Khorezm region of Uzbekistan
1987-1989	Republican School-Internat of young Physicists and Mathematicians in Tashkent, Uzbekistan
1989-1994	Department of Mathematics of the Tashkent State University, Tashkent, Uzbekistan
02.09.1994	Diploma in Mathematics from Tashkent State University
1994-1995	Assistant Lecturer at the Physics and Mathematics Department of Urgench State University, Urgench, Uzbekistan
1995-1998	Postgraduate Study at the Mechanics and Mathematics Department of National University of Uzbekistan, Tashkent, Uzbekistan
1998-2001	Junior Scientific Researcher at the Institute of Mathematics, Tashkent, Uzbekistan
01.10.2001- 31.03.2002	Scholar of DAAD
since 01.04.2002	Ph.D. student of Prof. Dr. Manfred Denker, Institute for Mathematical Stochastics of the University of Göttingen, member of Graduate College "Strömungsinstabilitäten und Turbulenz"
since 01.10.2002	member of Interdisciplinary PhD Program "Applied Statistics and Empirical Methods" of the University of Göttingen
November, 2003- March, 2004	Scientific assistant at the Institute for Mathematical Stochastics of the University of Göttingen