Siegfried Hörmann

Fluctuation analysis of dependent random processes

DISSERTATION

zur Erlangung des akademischen Grades
Doktor der technischen Wissenschaften

Graz University of Technology

Technische Universität Graz

Betreuer:
Univ.-Prof. Dr. István Berkes
Institut für Statistik

Graz, Oktober 2006

Contents

Abstract

Probability theory is based on the concept of independence, but independent stochastic models are unsuitable in many applications and starting with Markov's studies of one-step dependence models, the investigation of dependent processes began early in the theory. Beside Markov chains, the most classical dependent structures are martingales and stationary processes, whose intensive study started in the 1930's and whose structure and analytic properties are fairly well known. Still, many important problems in modern probability theory, statistics and econometrics lead to asymptotic problems not covered by the classical theory and the purpose of our dissertation is to give a detailed study of the refined asymptotic and fluctuation properties of these models. Among others, we will extend the classical fluctuation theory of martingales, proving upper-lower class results under optimal conditions and closing the gap between the independent and martingale theory. We will also prove several asymptotic results for ARCH type processes, a class of stationary processes playing an important role in modern econometrics. Finally, we will extend and sharpen a number of basic results in almost sure central limit theory, another recent, much investigated field in probability limit theory, exhibiting unusual dependence behavior. Specifically, we will describe the critical behavior of the ASCLT and determine the exact asymptotics of generalized moments in the theory. The basis for our proofs will be strong approximation, a method allowing one to reduce the asymptotic properties of dependent processes to those of a 'limiting' process, typically Brownian motion. This idea is due to Strassen [101] [103] and has been extended to wide classes of dependent processes by Philipp and Stout [90]. The new element in our approach is that instead of martingale approximation and Skorohod embedding used in the classical theory, we will give a direct a.s. approximation of nonlinear functionals of the studied processes by independent random variables, a method applying under very general conditions and leading to optimal results.

Acknowledgement

I am indebted to my PhD supervisor Prof. István Berkes for his guidance and for introducing me to the scientific world. I benefited extremely from numerous valuable discussions with him and from his many great thought-provoking impulses. I would also like to thank my colleagues at the Department of Statistics for the friendly and stimulating atmosphere. Many thanks go to my family who were behind me throughout my years of study. Last but not least I thank Tina for accompanying me to Graz and giving me a strong moral support. I want to dedicate this thesis to her.

Introduction

Probability theory grew out of the concept of independence and its most complete results concern sequences of independent random variables. However, independent sequences represent only a fraction of interesting random processes in practice and from the earliest days of probability theory, an intensive study of dependent processes has also begun. The simplest and best understood dependent structure is the class of Markov processes: Markovity is generally easy to verify and the linear character of the equations describing the time development of Markov processes (given by the Kolmogorov differential equations) enables one to give a rather complete description of the structure and asymptotics of such processes. A similarly well understood structure is the class of martingales, which grew out of the study of games of chance and has, like Markov processes, widespread applications in natural sciences, economics and finance. The martingale relation $E[X_n|X_1, \ldots, X_{n-1}] = X_{n-1}$ seems to be less informative than Markovity, but its consequences on the structure of the process are equally strong. Assume e.g. $EX_n^2 < \infty$ and let $s_n^2 = \sum_{k=1}^n E[(X_k - X_{k-1})^2|X_1, \ldots, X_{k-1}]$. Then, under $s_n^2 \to \infty$ a.s. the asymptotic behavior of $\{X_n, \, n \to \infty\}$ is the same as that of $\{W(s_n^2), \, n \to \infty\}$, where W is a Wiener process. This fact and its continuous time analogue connect martingale behavior to Wiener processes, a connection playing a crucial role in applications.

A third classical and widely studied class of dependent processes is the class of stationary sequences. Similarly to Markov processes and martingales, stationary processes have a rather transparent structure: by a classical result of Rosenblatt [92] (verifying a conjecture of Wiener), every strictly stationary process $\{X_n, \, n \in \mathbb{Z}\}$ admits, under mild additional conditions, the representation

$$X_n = f(\ldots, \varepsilon_{n-1}, \varepsilon_n) \qquad n \in \mathbb{Z}, \tag{1}$$

where $\{\varepsilon_n,\, n \in \mathbb{Z}\}$ is an i.i.d. sequence and $f : \mathbb{R}^\infty \to \mathbb{R}$ is a measurable function. Simple examples are linear processes

$$X_n = \sum_{k=0}^{\infty} c_k \varepsilon_{n-k}, \qquad n \in \mathbb{Z} \tag{2}$$

or Volterra series

$$X_n = \sum_{p=1}^{\infty} \sum_{i_1, i_2, \cdots, i_p = 1}^{\infty} c_{i_1} \cdots c_{i_p} \varepsilon_{n-i_1} \varepsilon_{n-i_1-i_2} \cdots \varepsilon_{n-i_1-\cdots-i_p}$$

appearing in the representation of nonlinear time series in econometrics. A further important example is the class of stationary sequences satisfying the nonlinear dynamics

$$X_n = g(X_{n-1}, \ldots, X_{n-p}, \varepsilon_n, \ldots, \varepsilon_{n-q})$$

typical for economical and financial phenomena. Despite the simplicity of (1), its infinite dimensional character presents substantial difficulties, even if f is linear. The first general results on the asymptotic properties of the process (1) were obtained by Ibragimov [62] and Billingsley [19] and extended later by several authors. As it turned out, such processes have two basic types. If the dependence of $f(\ldots, x_{-1}, x_0)$ on the variables $x_{-n}, x_{-n-1}, \ldots$ is weak enough for large n ("short memory" case), then (X_n) behaves like a sequence of independent random variables. If the dependence of $f(\ldots, x_{-1}, x_0)$ on its tail variables decreases slowly ("long memory" case), (X_n) has completely different asymptotic properties. For example, the linear process (2), where $E\varepsilon_0 = 0$, $E\varepsilon_0^2 = 1$ and $c_k = k^{-\alpha}$, $\alpha > 1/2$, behaves like an independent sequence for $\alpha > 1$, while for $\alpha < 1$ the situation is different: in this case, under suitable moment conditions, $\sum_{k=1}^{[nt]} f(X_k)$ converges weakly, after suitable normalization, to a nongaussian process. (Cf. Surgailis [104], Avram and Taqqu [6]). A similar change of behavior holds for processes with general nonlinear f, but the exact borderline between weak and strong dependence is generally difficult to find. Typically, the short memory case is easier and results in the long memory case lie considerably deeper.

The above processes are the simplest, oldest and best understood dependent structures in probability theory and there is a huge literature dealing with their properties. Still, their asymptotic theory is much less complete than the theory of independent random variables, and several important problems in theoretical and applied probability, statistics and econometrics lead to asymptotic problems lying outside of the reach of classical theory. The purpose of our dissertation is to develop an asymptotic theory for processes of this type, providing the answer to several open problems. In Chapter 5 we deal with the fluctuation theory of martingales. Starting with the classical paper of Strassen [103], several authors studied refined path properties of martingales, such as the LIL and the corresponding upper-lower class tests. Still, the results are much less complete than the classical LIL theory for independent random variables, developed by Feller [44] [45]. Our main result in Chapter 5, Theorem 5.1, provides a general upper-lower class test for martingales, which not only contains and extends most earlier results in the field, but it is essentially optimal and closes the gap between martingale and independent results. In particular, we find an optimal condition for the classical Kolmogorov-Erdős-Feller-Petrovski integral test for stationary ergodic martingale difference sequences, a long open problem in martingale LIL theory.

In Chapters 6 and 7, we prove several asymptotic results for nonlinear time series. Inspired by the seminal papers of Engle [39] and Bollerslev [21], ARCH and GARCH models got to the center of attention in econometrics in the past two decades, as the first models giving a realistic description of the volatility of financial processes. Despite extensive empirical statistical work in the field (see Bollerslev-Chou-Kroner [22] for a survey of the first 10 years of the theory), we know relatively little on the refined path and asymptotic properties of ARCH and related processes and in Chapters 6 and 7 we give a detailed asymptotic study of this model. All such models satisfy the representation (1), but the existing theory (see e.g. Carrasco and Chen [26]) is based on Markov methods and works only under restrictive conditions required by the Markovity of such processes.

A further important application of our method is almost sure central limit theory, a new and much studied field of probability theory growing out of the almost sure central limit theorem (Brosamler [25], Schatte [96], Lacey and Philipp [69]), a striking a.s. version of the classical CLT involving logarithmic measure. This result reveals a new side of the fluctuations of independent random variables and has led, in the past two decades, to an extensive literature featuring several new and remarkable limit theorems. Asymptotic results in this field fit into a slightly modified version of the stationary model (1), and our methods lead to a considerable sharpening of the theory. Among others, we will determine the critical behavior of the model and describe completely the convergence of generalized moments in the a.s. central limit theorem, two important open problems of the theory.

The basic method of our thesis is strong approximation, a powerful method developed by Strassen [101] [103] for the study of sums of independent random variables and martingales. Strassen's method was later extended for large classes of weakly dependent random variables, see e.g. Philipp and Stout [90]. The essential new element in our approach is that, instead of martingale approximation of block sums of the considered processes, we approximate functionals of these blocks directly by independent random variables, utilizing an idea first employed by Berkes and Horváth [15] in their studies of GARCH models. This approach leads to considerably sharper results and applies also for nonlinear limit theorems, opening the way for refined asymptotics of models of the type (1) and their generalizations.

We now summarize our main results.

(A) Almost sure central limit theory

Let $X_1, X_2, \dots$ be i.i.d. random variables with $EX_1 = 0$, $EX_1^2 = 1$ and let $S_n = \sum_{k=1}^n X_k$. Then $P(S_n/\sqrt{n} \le x) \to \Phi(x)$ for all $x \in \mathbb{R}$, but the asymptotic frequency of the integers $\{n : S_n/\sqrt{n} \le x\}$ is not $\Phi(x)$: the dependence of the random variables $S_n/\sqrt{n}$ is too strong for the law of large numbers to apply for the events

$\{S_n/\sqrt{n} \le x\}$. However, as Brosamler [25] and Schatte [96] proved under the existence of higher moments and Lacey and Philipp [69] without additional assumptions, the 'logarithmic' limit theorem

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k=1}^{N} \frac{1}{k} I\left\{\frac{S_k}{\sqrt{k}} \le x\right\} = \Phi(x) \quad \text{a.s.} \tag{3}$$

holds for any fixed $x \in \mathbb{R}$. This remarkable result (called almost sure central limit theorem) became the starting point of a new and much studied area of probability theory dealing with 'pathwise' versions of distributional limit theorems. As it turned out, the a.s. convergence relation (3) not only admits substantial refinements (precise convergence rates, LIL type results, etc.) but any weak limit theorem of the form $f_n(X_1, \dots, X_n) \xrightarrow{d} G$ of independent random variables has, under mild technical conditions on the functionals f_n, a logarithmic a.s. version similar to (3). (See Berkes and Csáki [9].) Relation (3) has also been extended, along with its nonlinear analogues, for various dependent structures such as Gaussian processes, martingales, mixing processes, etc. For a survey of the theory see Berkes [8] and Atlagh and Weber [4]. While in many directions essentially optimal results have been obtained, some basic problems of the theory remain open, in fact almost untouched. Most results in almost sure central limit theory, similarly to (3), involve logarithmic averaging and it is natural to ask if this averaging is the only possible one, or if other weight sequences can be used, too. Why logarithmic averages work in the theory is best seen from the Wiener analogue of (3), i.e.

$$\lim_{N\to\infty} \frac{1}{\log N} \int_1^N \frac{1}{t} I\left\{\frac{W(t)}{\sqrt{t}} \le x\right\} dt = (2\pi)^{-1/2} \int_{-\infty}^{x} e^{-t^2/2}\, dt \qquad \text{a.s. for all } x$$

which reduces, after the transformation $t = e^u$, to the ergodic theorem for the Ornstein-Uhlenbeck process $e^{-u/2}W(e^u)$. Despite the simplicity of this argument, the relation

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k=1}^{N} d_k I\left\{\frac{S_k}{\sqrt{k}} \le x\right\} = \Phi(x) \quad \text{a.s.} \tag{4}$$

where $D_N = \sum_{k=1}^{N} d_k$, actually holds for many other weight sequences (d_k) as well, as was first observed by Peligrad and Révész [86]. By standard results of general summation theory (cf. [27]), the larger the weight sequence (d_k) is, the stronger the a.s. limit theorem (4) becomes. Thus the strongest, optimal form of the ASCLT is the one with the largest weights d_k. This critical (d_k) marks the dividing line between weak and strong dependence of $d_k I\{S_k/\sqrt{k} \le x\}$, just like the weights $c_k \sim k^{-1}$ in (2) mark the boundary line between weak and strong dependence of the linear process $\{X_n,\, n \in \mathbb{Z}\}$. In Chapter 2 of our dissertation we will determine this critical weight sequence (d_k) and corresponding normalizing sequence D_k; our result is

$$D_N = \exp\left(\log N / (\log\log N)^\alpha\right) \tag{5}$$

for some $1 \le \alpha \le 3$. The surprising feature of this result is that the averaging method determined by (5) lies much closer to ordinary averaging defined by $D_N = N = \exp(\log N)$ than to logarithmic averaging, putting almost sure central limit theory in a new light: despite its prominent role in the theory, logarithmic averaging is of secondary importance and much stronger results can be obtained by using summation methods near ordinary (Cesàro) summation. The results of Chapter 2 stem from our paper Hörmann [56].

In Chapter 3 we extend these investigations to general (nonlinear) analogues of the a.s. central limit theorem. Our results here are less precise than for the linear CLT and an explicit determination of the critical weight sequence (d_k) remains open. However, the surprising phenomenon observed in the case of the linear CLT remains valid in the nonlinear case as well: the standard logarithmic averaging methods provided by the universal ASCLT of Berkes and Csáki [9] are stronger than necessary and all results remain valid with averaging methods much closer to the Cesàro summation method. In fact, we will construct examples where even Cesáro summation works in a.s. limit theorems, a most surprising consequence of our results. The material in Chapter 3 is published in Hörmann [57].

Another basic open problem of a.s. central limit theory concerns moment behavior. Let $X_1, X_2, \ldots$ be independent random variables with mean 0 and finite variances and put $S_n = \sum_{k=1}^n X_k$, $s_n^2 = \sum_{k=1}^n EX_k^2$. By a slight variant of the ASCLT (3) (implicit in Lacey and Philipp [69]), we have

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k=1}^{N} \frac{1}{k} f\left(\frac{S_k}{s_k}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-x^2/2}\, dx \quad \text{a.s.} \tag{6}$$

for any bounded continuous function f on $\mathbb{R}$. The validity of (6) for unbounded f is of major interest in the theory: for $f(x) = x^p$ this expresses convergence of moments in (3) and for this reason, the left hand side of (6) is called a *generalized moment*. A complete solution in the i.i.d. case was given by Ibragimov and Lifshits [63], who proved that in this case relation (6) holds if the integral on the right hand side is finite and the integrand is nonincreasing for $|x| \to \infty$; this monotonicity condition cannot be dropped. This result lies much deeper than the ordinary ASCLT (3); its proof depends on delicate fluctuation properties of i.i.d. random variables. The argument fails for general independent sequences (X_n) and generalized moment behavior in a.s. central limit theory remains open. In Chapter 4 of our dissertation we give a complete solution of this problem in the case of bounded r.v.'s X_n, revealing a striking connection between the ASCLT and LIL. In fact, we will show that a sufficient criterion for the validity of (6) for an independent sequence (X_n) and for all $|f(x)| \leq \text{const} \cdot e^{\gamma x^2}$, $\gamma < 1/2$ is

$$|X_n| = o(s_n/(\log\log s_n)^{1/2}). \tag{7}$$

Moreover, we show that this result is sharp, in the sense that replacing o by O in (7), the so obtained condition will not imply the ASCLT (6). Further, a sufficient condition for (6) to hold for all $f(x) = e^{x^2/2}\varphi(x)$ with $\varphi \in L_1(\mathbb{R})$ and φ uniformly continuous is

$$|X_n| = o(s_n/(\log\log s_n)^{3/2}). \tag{8}$$

Conditions (7) and (8) play a prominent role in fluctuation theory: (7) is Kolmogorov's classical (and optimal) criterion for the law of the iterated logarithm and (8) is a precise criterion for the upper-lower class refinement of the LIL due to Feller [44]. This shows the surprising fact that the validity of relation (6) is basically equivalent to the law of the iterated logarithm for (X_n), revealing a new side of almost sure central limit theory. The results of Chapter 4 stem from Berkes and Hörmann [12].

(B) Nonlinear time series.

Asymptotic properties of financial time series (asset returns, exchange rates, inflation data, etc.) have been studied intensively in the econometric literature in the past decades and several models for such processes have been proposed. A characteristic property of many financial processes $\{x_t\}$ is conditional heteroscedasticity, i.e. the fact that the conditional variance $\sigma_t^2 = \mathrm{Var}[x_t|x_0, \ldots, x_{t-1}]$ changes with the time t. The old classical models such as geometric Brownian motion (underlying, e.g., the Black-Scholes formula) and ARMA processes have constant volatility and thus they are, in many respects, unrealistic. The first suitable model, the ARCH (autoregressive conditional heteroscedastic) process defined by

$$x_t = \sigma_t \varepsilon_t, \qquad \sigma_t^2 = \omega + \sum_{j=1}^{p} \alpha_j x_{t-j}^2$$

was introduced by Engle [39] and became an instant success in the theory. In subsequent years several extensions and refinements of this model, taking into account various special properties of financial processes, were introduced and studied. The latest and most general of these models, the so called augmented GARCH process was introduced by Duan [37]; this model describes most of the characteristic features (asymmetry, threshold behavior) of volatilities very precisely. While considerable statistical work on these processes has been done, many basic structural and asymptotic properties of the augmented GARCH model remained unexplored. When stationary, such processes have a representation (1) and under restrictive conditions on the i.i.d. sequence $\{\varepsilon_t\}$, Markov methods lead to important information on $\{x_t\}$ (Carrasco and Chen [26]). However, the path and asymptotic properties of such sequences remain unknown even in the simplest case of binomial ε_t. In Chapters 6 and 7 we prove several asymptotic results for augmented GARCH sequences under essentially optimal conditions. For example, we obtain an a.s. invariance principle for the empirical process of augmented GARCH sequences under assuming $E(\log|\varepsilon_0|)^\gamma < \infty$ for some $\gamma > 0$, a condition only slightly stronger than the condition for the stationarity of $\{x_t\}$. Further, we give necessary and sufficient conditions for the validity of the functional CLT for such processes and deduce Berry-Esseen bounds in the CLT. Our results have important consequences for the statistics of such processes, extending the applicability of Dickey-Fuller and CUSUM type procedures for detecting unit roots and change of parameters.

Our results were published in the papers Hörmann [59], Hörmann [55], Berkes, Hörmann and Horváth [13].

(C) Fluctuation theory of martingales

Let $X_1, X_2, \dots$ be independent r.v.'s with mean 0 and finite variances and let $s_n^2 = \sum_{k=1}^n EX_k^2$. Let $\{S(t),\, t \geq 0\}$ denote the function which is linear in the intervals $[s_k^2, s_{k+1}^2]$ and $S(s_k^2) = X_1 + \dots + X_k$ $(k = 0, 1, \dots)$. By Kolmogorov's classical LIL, under (7) we have

$$\limsup_{t\to\infty} \frac{S(t)}{(2t \log\log t)^{1/2}} = 1 \quad \text{a.s.} \tag{9}$$

A much more refined result was proved by Feller [44] who showed that if (8) holds then for any positive nondecreasing function $\varphi(t)$, $t > 0$, we have

$$P\left\{S(t) \leq \sqrt{t}\varphi(t) \quad \text{eventually}\right\} = 1 \text{ or } 0 \tag{10}$$

according as

$$\int_1^\infty \frac{\varphi(t)}{t} \exp(-\varphi^2(t)/2)\, dt < \infty \quad \text{or } = \infty. \tag{11}$$

As noted earlier, the boundedness conditions (7) and (8) in the above results are sharp. If the X_n are i.i.d., then the LIL (9) holds provided $EX_1 = 0$, $EX_1^2 = 1$ (Hartman and Wintner [54]), but the situation with the integral test (10)–(11) is different. As Feller [45] showed, for an i.i.d. sequence (X_n) the test (10)–(11) holds iff $EX_1 = 0$, $EX_1^2 = 1$ and

$$EX_1^2 I\{|X_1| \geq t\} = O\left((\log\log t)^{-1}\right) \quad (t \to \infty). \tag{12}$$

In particular, this is the case if

$$EX_1^2 (\log\log |X_1|) < \infty. \tag{13}$$

In terms of moment conditions, the last condition is also best possible.

Feller's results lie very deep and the proofs make an essential use of the independence of (X_n). Using an almost sure invariance principle, Strassen [103] was the first to prove upper-lower class results for martingales. Specifically, he proved that if (X_n) is a martingale difference sequence with finite variances and

$$s_n^2 = \sum_{k=1}^{n} E[X_k^2 | X_1, \ldots, X_{k-1}] \to \infty \quad \text{a.s.} \tag{14}$$

then the test (10)–(11) holds provided

$$\sum_{k=1}^{\infty} s_k^{-2} (\log s_k)^5 E[X_k^2 I\{|X_k| \geq s_k (\log s_k)^{-5}\} | X_1, \ldots, X_{k-1}] < \infty \quad \text{a.s.}$$

In particular, this is the case if

$$|X_n| = o(s_n/(\log s_n)^5) \qquad \text{a.s.}$$

(Note that in the martingale case s_n^2 is defined differently than in the case of independent r.v.'s: it means the sum in (14), the "clock" of the process.) Strassen's conditions are far from optimal and his results were improved gradually by Jain, Jogdeo and Stout [65], Philipp and Stout [91] and Einmahl and Mason [38]. Specifically, Einmahl and Mason proved that the test (10)–(11) holds under Feller's condition (8), which is therefore an optimal condition. Much less is known for unbounded martingale difference sequences. Various criteria are given in Jain, Jogdeo and Stout [65] and Philipp and Stout [91], but they are substantially more restrictive than Feller's classical conditions in the independent case. In Chapter 5 of our dissertation we will give a set of sufficient conditions for (10)–(11) which not only improves earlier results in the field, but it is optimal. In particular, our criteria imply that in the stationary ergodic case Feller's condition (13) suffices for the test (10)–(11), establishing a long open conjecture for martingales. As we will show in a subsequent paper, our results lead to optimal upper-lower class results for weighted i.i.d. sequences as well.

Chapter 1

Some classical results of fluctuation theory

For the convenience of the reader we shall review a number of important limit theorems for partial sums of independent random variables. Some of them have been already mentioned in the introduction and the purpose of the present chapter is to give a compact overview.

1.1 The law of the iterated logarithm and its refinements

Let $X_1, X_2, \dots$ be independent random variables and $S_n = X_1 + \cdots + X_n$. If the X_k are i.i.d. with zero expectation, we get by the law of large numbers that $n^{-1} S_n \to 0$. The last relation is far from optimal if we assume the existence of higher moments. If $E|X_1|^p < \infty$ with $p \in (0, 2)$, it follows from the Marcinkiewicz-Zygmund law of large numbers (cf. [29, p. 122]), that $n^{-1/p} S_n \to 0$ (provided $EX_1 = 0$ for $p \geq 1$). The *law of the iterated logarithm (LIL)* gives the precise speed of growth of the partial sums process (S_n).

Theorem A1. (Hartman and Wintner [54]). *Assume that* (X_k) *is an i.i.d. sequence* *with* $EX_1 = 0$ *and* $EX_1^2 = 1$. *Then*

$$\limsup_{n\to\infty} (2n \log\log n)^{-1/2} S_n = 1 \quad \text{a.s.}$$

Conversely (Strassen [102]), *if* $EX_1^2 = \infty$ *then* $\limsup (2n \log\log n)^{-1/2} S_n = \infty$ a.s.

Assume now that X_k are not necessarily identically distributed and set $s_n^2 = \operatorname{Var} S_n$, where $S_n = X_1 + \cdots + X_n$. The following theorem for bounded random variables is due to Kolmogorov.

Theorem A2. (Kolmogorov's law of the iterated logarithm [67]). *Let* (X_k) *be a sequence of independent random variables with zero mean and finite variances such that* $s_n^2 \to \infty$. *If we assume that* $|X_k| \le M_k$ a.s. *and that*

$$M_k = o(s_k/(\log\log s_k^2)^{1/2}), \tag{1.1}$$

then

$$\limsup_{n} (2s_n^2 \log\log s_n^2)^{-1/2} \sum_{k=1}^{n} X_k = 1 \quad \text{a.s.} \tag{1.2}$$

Conversely (Marzinkiewicz and Zygmund [79]), *the LIL may fail if* o *in (1.1) is replaced by* O.

Note that the LIL is equivalent to

$$P(S_n > s_n(\alpha \log\log s_n)^{1/2} \text{ i.o.}) = \begin{cases} 0 & \text{if } \alpha > 2; \\ 1 & \text{if } \alpha < 2. \end{cases}$$

Using a terminology introduced by P. Lévy, a function $\varphi(t)$ belongs to the *upper class* $\mathcal{U}$ if $S_n > s_n\varphi(s_n)$ occurs only for finitely many n with probability one and it belongs to the *lower class* $\mathcal{L}$ if $S_n > s_n\varphi(s_n)$ occurs for infinitely many n a.s. By the 0-1 law, every function φ belongs to the upper or lower class. By the LIL (1.2), $\phi_\alpha(t) = (\alpha \log\log t)^{1/2} \in \mathcal{U}$ if $\alpha > 2$ and $\phi_\alpha \in \mathcal{L}$ if $\alpha < 2$. Whether ϕ_2 belongs to $\mathcal{U}$ or $\mathcal{L}$ is not determined by (1.2) and thus an upper-lower class test is a refinement of the LIL. In case of Rademacher r.v.'s a precise characterization of the upper and lower classes has been given by Erdős [41] in form of an integral test. He showed that a non-decreasing function φ belongs $\mathcal{U}$ (resp. $\mathcal{L}$) iff

$$I(\varphi) = \int_1^\infty \frac{\varphi(t)}{t} \exp(-\varphi^2(t)/2)\, dt < \infty \quad (= \infty).$$

This famous criterion is called the *Kolmogorov-Erdős-Feller-Petrovsky (KEFP) test.* The following classical results are due to W. Feller.

Theorem A3. (Feller [45]). *Let (X_k) be an i.i.d. sequence with $EX_1 = 0$ and $EX_1^2 = 1$. Then the KEFP integral test holds provided*

$$EX_1^2 I\left\{|X_1| \geq t\right\} = O\left((\log\log t)^{-1}\right) \quad (t \to \infty).$$

The statement becomes false if $EX_1 = 0$ and $EX_1^2 = 1$ and

$$(\log\log t)\, EX_1^2 I\left\{|X_1| \geq t\right\} \to \infty.$$

Theorem A4. (Feller [44]). *Let (X_k) be a sequence of independent r.v.'s with zero mean and finite variances, such that $s_n \to \infty$ and*

$$|X_k| \leq K_n s_n / (\log\log s_n^2)^{3/2} \quad \textit{with} \quad K_n = O(1). \tag{1.3}$$

Then the KEFP test holds. Condition (1.3) is sharp in the sense that if $K_n \to \infty$ then in general the KEFP test is no longer true.

1.2 Weak invariance principles

Erdős and Kac [42] developed a new method to study the distributional behavior of the process $\{S_n,\, n \geq 1\}$, where $S_n = X_1 + \cdots + X_n$ is a sum of independent r.v's. In order to obtain the asymptotics of certain functionals of the path $\{S_n,\, n \geq 1\}$ (like $\max\{S_k : 1 \leq k \leq n\}$ or $\sum_{k=1}^n S_k^2$), they used an *invariance principle.* They first derived the desired limit relations for a special choice of the underlying process (X_k) and showed afterwards that this limit does not depend on the law of X_k, i.e. it is invariant under changing the law of X_k.

Donsker [36] gave a more general form of the Erdős-Kac invariance principle. To state his theorem, we need some notations and we recall the basic definition of weak convergence in a metric space. We denote by $(\Omega, \mathcal{A}, P)$ the probability space on which the process (X_n) is defined. The space of continuous functions on the unit interval equipped with the uniform metric will be denoted by $C[0,1]$. If $\mathcal{C}$ denotes the Borel σ-algebra on $C[0,1]$, then the functions $W_n : \Omega \to C[0,1]$ defined by

$$W_n(t) := \frac{1}{\sqrt{n}} \left(S_{[nt]} + (nt - [nt]) X_{[nt]+1} \right), \tag{1.4}$$

are $(\mathcal{A}, \mathcal{C})$ measurable, i.e. they are random elements in C. Convergence in distribution of W_n is characterized via the weak convergence of the induced probability measures $P \circ W_n^{-1}$. We write $W_n \xrightarrow{d} W$. Remember that weak convergence of a sequence of probability measures (Q_n) on the Borel sets $\mathcal{S}$ of some metric space S is defined as

$$\int_S f \, dQ_n \to \int_S f \, dQ \quad (n \to \infty),$$

for all bounded and continuous functions f.

Theorem A5. (Donsker's invariance principle [36]). *Let $X_1, X_2, \ldots$ be i.i.d. random variables with $EX_1 = 0$ and $EX_1^2 = 1$ and let $S_n = X_1 + \cdots + X_n$. Define $W_n(t)$ as in (1.4). Then we have*

$$W_n \xrightarrow{d} W, \tag{1.5}$$

where $W = \{W(t), t \in [0,1]\}$ is a standard Brownian motion process.

Relation (1.5) is also called *functional central limit theorem (FCLT).* For the proof of Theorem A5 and for a detailed discussion on the weak convergence of random elements in metric spaces we refer to Billingsley [19].

The assumption of the identical distribution of the X_n in Theorem A5 can be easily dropped. Let (X_n) be a sequence of r.v.'s with $EX_k = 0$ and $EX_k^2 =: \sigma_k^2 < \infty$. Set $S_n = X_1 + \cdots + X_n$ and $s_n^2 = ES_n^2$. Define, with a natural modification of (1.4),

$$W_n(t) = \frac{1}{s_n} (S_{[nt]} + (nt - [nt]) X_{[nt]+1}).$$

Then we have

Threorem A6. (Prohorov). *Assume that* $X_1, X_2, \ldots$ *are independent random variables with zero mean and finite variances. If*

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{\{|X_k|>\varepsilon s_k\}} X_k^2 \, dP = 0, \tag{1.6}$$

then $W_n \xrightarrow{d} W$.

Condition (1.6) is the classical *Lindeberg* condition. It is satisfied e.g. if $E|X_k|^{2+\delta} < \infty$ $(k = 1, 2, \ldots)$ for some $\delta > 0$ and

$$\frac{1}{s_n^{2+\delta}} \sum_{k=1}^{n} E|X_k|^{2+\delta} \to 0,$$

i.e. under *Ljapunov's* condition.

Note that a sequence of random elements (R_n) on some metric space M converges in distribution to some random element R if and only if $h(R_n) \xrightarrow{d} h(R)$ for every real valued and continuous function h on M (cf. [47, Theorem 8.2.3]). Since the function $\pi_1 : C[0,1] \to \mathbb{R}$, $x \mapsto x(1)$ is continuous, we have under the conditions of Theorems A5–A6, $\pi_1(W_n) = S_n/s_n \xrightarrow{d} \pi_1(W) = W(1)$, i.e. the central limit theorem holds.

Sometimes it is more convenient to consider

$$W_n'(t) := S_{[nt]}/s_n$$

instead of $W_n(t)$. In this case W_n' is an element of $D[0,1]$, the space of right-continuous functions on $[0,1]$ which have a left-hand limit and which is equipped with the Skorokhod metric. Under the conditions of Theorems A5–A6 we have $W_n' \xrightarrow{d} W$. Especially for the investigation of empirical distribution functions, it is more convenient to consider the space $D[0,1]$ which allows jumps.

Theorem A7. (Invariance principle for the empirical process). *Let* $U_1, U_2, \ldots$ *be i.i.d. uniformly distributed r.v.'s on* $(0,1)$ *and consider the empirical distribution functions*

$$R_n(s) = n^{-1/2} \sum_{k=1}^{n} (I\{U_k \le s\} - s).$$

Then (R_n) *is a sequence of random elements in* $D[0,1]$ *and* $R_n \xrightarrow{d} B$, *where* $\{B(s), 0 \le s \le 1\}$ *is a Brownian Bridge* $(B(t) = W(t) - tW(1))$.

For the proof see e.g. [47, p. 386].

1.3 Strong invariance principles

A completely different approach to the study of the fluctuations of independent random variables was developed by Strassen [101]. His idea was to redefine the partial sum process $\{S_k,\, k \geq 0\}$ on a new probability space together with a Wiener process $\{W(t),\, t \geq 0\}$, such that

$$|S_n - W(n)| = o(g(n)) \quad \text{a.s.} \quad (n \to \infty) \tag{1.7}$$

for a suitable $g(n)$, depending on the moment behavior of the underlying sequence (X_n). Specifically, he proved the following theorem:

Theorem A8. (Strassen's invariance principle [101]). *Let $X_1, X_2, \dots$ be i.i.d. random variables with $EX_1 = 0$, $EX_1^2 = 1$. Then we can construct an i.i.d. sequence $\tilde{X}_1, \tilde{X}_2, \dots$ with $\tilde{X}_1 \stackrel{d}{=} X_1$ on a new probability space together with a Wiener process $\{W(t), t \geq 0\}$ such that*

$$|\tilde{S}_n - W(n)| = o(\sqrt{n \log\log n}) \quad \text{a.s.}, \tag{1.8}$$

where $\tilde{S}_n = \tilde{X}_1 + \cdots + \tilde{X}_n$.

The basic tool Strassen used to prove Theorem A8 is *Skorokhod embedding*. By a remarkable result of Skorokhod, under the assumptions of Theorem A8 there exists, on a suitable probability space, a Wiener process $\{W(t), t \geq 0\}$ and a sequence of stopping times $T_0 = 0, T_1, T_2, \dots$ such that $ET_n = n$, $T_n - T_{n-1}$ are i.i.d. and $\{W(T_n), n \geq 1\} \stackrel{d}{=} \{S_n, n \geq 1\}$. Thus setting $\tilde{S}_n = W(T_n)$ and observing that by the law of large numbers $T_n \sim n$ a.s., $\tilde{S}_n$ will be close, in some sense, to $W(n)$. It is important to note that Theorem A8 does not contain Donsker's invariance principle, due to the $\log\log$ factor in (1.8). However, if we could show that

$$|\tilde{S}_n - W(n)| = o(\sqrt{n}) \quad \text{a.s.} \tag{1.9}$$

then Donsker's theorem would follow. Assuming only the existence of finite second moments, Major [75] showed that the rate $o(\sqrt{n \log\log n})$ in (1.7) cannot be improved.

A better rate in (1.8) can be obtained if we assume more than the existence of finite second moments. Profound results in this direction were proved by Komlós, Major and Tusnády. In particular, Major [76] proved that under $E|X_1|^p < +\infty$, $p > 2$ the approximation (1.7) holds with $g(n) = n^{1/p}$. On the other hand Breiman [24] showed that this rate is best possible. We refer to Csörgő and Révész [31] for an elaborate treatment of further strong approximation results for sums of i.i.d. random variables.

Theorem A9. (Philipp and Stout [91]). *Let (X_k) be a sequence of independent, bounded random variables with zero mean. Put $S_k = X_1 + \cdots + X_k$, $s_k^2 = \operatorname{Var} S_k$ and define the function $\{S(t),\, t \geq 0\}$ by*

$$S(t) = S_k \quad \textit{if} \quad s_k^2 \leq t < s_{k+1}^2 \qquad (k = 0, 1, \ldots).$$

Further let f be a non-increasing differentiable function such that for all $x \geq x_0$

$$1/\log x \leq f(x) \leq 10^{-3}$$
$$f(x)x(\log\log x)^{-1/2} \nearrow, \quad g(x) := \log x / f(x) \nearrow \infty \quad \textit{with } xg'(x) \textit{ bounded.}$$

If we assume that

$$|X_n| \leq f(s_n)s_n(\log\log s_n)^{-1/2},$$

then we can redefine the sequence $X_1, X_2, \ldots$ on a new probability space together with a Wiener process $\{W(t), t \geq 0\}$ such that

$$|S(t) - W(t)| \leq 10^3(f(t)t\log\log t)^{1/2} \quad \textit{for all } t \geq t_0.$$

In contrast to Donsker's invariance principle providing information on the distributional behavior of the process $\{S_k,\, k \geq 0\}$, Strassen's invariance principle can be used to derive strong limit theorems for $\{S_k,\, k \geq 0\}$ as well. For example, it is relatively easy to prove the LIL

$$\limsup_{n\to\infty} \frac{1}{\sqrt{2n\log\log n}} W(n) = 1 \quad \text{a.s.}$$

for the Wiener process. Using the last relation and Theorem A8, the Hartman-Wintner LIL follows immediately. Furthermore, by Theorem A9, Kolmogorov's condition (1.1) implies an error term $o((t\log\log t)^{1/2})$ in the Wiener approximation.

This implies immediately Theorem A2. Actually, even the more refined KEFP integral test of the preceding section can be deduced directly from Wiener fluctuation theory using strong approximation results. For example, it is well known that

$$|S(t) - W(t)| \le K t^{1/2} (\log\log t)^{-1/2} \quad \text{a.s.} \quad (t \to \infty)$$

for some constant K implies that the KEFP integral test holds for the sequence S_n. Later on we shall need the following result due to Sakhanenko, which is a partial improvement of Theorem A9. In view of the previous remarks, it implies that the KEFP integral test holds for bounded sequences satisfying (1.10) below.

Theorem A10. (Sakhanenko [95]). *Assume that $X_1, X_2, \ldots$ satisfy the conditions of Theorem A9 and*

$$|X_n| \le \varepsilon_n s_n (\log\log s_n)^{-3/2} \tag{1.10}$$

with some real numbers $\varepsilon_n \to 0$. Then the conclusion of Theorem A9 holds with an error term $o(t^{1/2}(\log\log t)^{-1/2})$.

Chapter 2

Critical behavior in almost sure central limit theory

2.1 Introduction and results

Let $X_1, X_2, \dots$ be i.i.d. random variables with $EX_1 = 0$, $EX_1^2 = 1$ and let $S_k = X_1 + \cdots + X_k$. The simplest version of the almost sure central limit theorem (ASCLT) states that

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k=1}^{N} \frac{1}{k} I\left\{\frac{S_k}{\sqrt{k}} \le x\right\} = (2\pi)^{-1/2} \int_{-\infty}^{x} e^{-t^2/2}\, dt \qquad \text{a.s.} \tag{2.1}$$

for every fixed $x \in \mathbb{R}$. This result was proved by Brosamler [25] and Schatte [96] under some additional moment conditions and by Fisher [46] and Lacey and Philipp [69] assuming only finite variances. (Actually, (2.1) was known to Lévy [71, p. 270] but he did not specify conditions and gave no proof.) In recent years, many authors investigated limit theorems of this type and several variants and extensions of (2.1) have been obtained. We refer to Atlagh and Weber [4] and Berkes [8] for surveys of the field.

A characteristic feature of the theory is the use of logarithmic averages in (2.1), and from the arc sine law it follows that with ordinary averages relation (2.1) fails even for $x = 0$. Why logarithmic averages work here is best seen from the Wiener analogue of (2.1), i.e.

$$\lim_{N\to\infty}\frac{1}{\log N}\int_1^N \frac{1}{t} I\left\{\frac{W(t)}{\sqrt{t}}\le x\right\} dt = (2\pi)^{-1/2}\int_{-\infty}^x e^{-t^2/2}\, dt \qquad \text{a.s. for all } x.$$

After the transformation $t = e^u$ this reduces to the ergodic theorem for the Ornstein-Uhlenbeck process $e^{-u/2}W(e^u)$. Via a strong approximation argument, this also proves the a.s. central limit theorem (2.1) under moment conditions only slightly stronger than $EX_1^2 < \infty$. Despite the simplicity of this argument, it is important to note that logarithmic summation is not the only possible summation that leads to a.s. convergence to $\Phi(x)$ in (2.1). Peligrad and Révész [86] showed that

$$\lim_{N\to\infty}\frac{1}{D_N}\sum_{k=1}^N d_k I\left\{\frac{S_k}{\sqrt{k}}\le x\right\} = \Phi(x) \quad \text{a.s.,} \tag{2.2}$$

holds if

$$d_k = \frac{(\log k)^\alpha}{k}, \quad D_n = \sum_{k=1}^n d_k \qquad (\alpha > -1) \tag{2.3}$$

and Berkes and Csáki [9] showed that (2.2) holds even if

$$d_k = \frac{\exp((\log k)^\alpha)}{k} \qquad (0\le \alpha < 1/2). \tag{2.4}$$

For summation methods in a.s. limit theory which are different from log-summation we also refer to Becker-Kern [7] and Weber [109]. In Hörmann [60] we proved that the summation defined in (2.4) works for any $\alpha \in [0, 1)$. In [9] it is also observed that (2.2) is valid for any $d_k \le 1/k$ with $\sum d_k = \infty$. Thus the weight sequence $d_k = 1/k$ plays no special role in (2.2); it is one in a large class of possible weight sequences, and several smaller and larger sequences work equally well. To understand this phenomenon better, let us recall some results from classical summability theory. Given a positive sequence $\mathbf{D} = (d_k)$ with $D_n = \sum_{k=1}^n d_k \to \infty$, we say that a sequence (x_n) is $\mathbf{D}$-summable to x if

$$\lim_{n\to\infty} D_n^{-1}\sum_{k=1}^n d_k x_k = x.$$

By a result of Hardy (see [27, p. 35]), if $\mathbf{D}$ and $\mathbf{D}^*$ are summation procedures with $D_n^* = O(D_n)$, then under minor technical assumptions, the summation $\mathbf{D}^*$ is stronger than **D**, i.e. if a sequence (x_n) is **D**-summable to x, then it is also $\mathbf{D}^*$-summable to x. Moreover, by a result of Zygmund (see [27, p. 35]) if $D_n^\alpha \le D_n^* \le D_n^\beta \quad (n \ge n_0)$ for some $\alpha > 0, \beta > 0$, then **D** and $\mathbf{D}^*$ are equivalent, and if $D_n^* = O(D_n^\varepsilon)$ for every $\varepsilon > 0$, then $\mathbf{D}^*$ is strictly stronger than **D**. For example, logarithmic summation, defined by $d_n = 1/n$ is stronger than ordinary (Cesàro) summation defined by $d_n = 1$ and weaker than loglog summation defined by $d_n = 1/(n \log n)$. On the other hand, all summation methods defined by

$$d_n = (\log n)^\alpha / n, \qquad \alpha > -1$$

are equivalent to logarithmic summation and all summation methods defined by

$$d_n = n^\alpha, \quad \alpha > -1$$

are equivalent to Cesàro summation. These remarks show that relation (2.2) with the sequence in (2.3) is, despite their formal difference, actually equivalent to the case $d_k = 1/k$ and show also that the sequences in (2.4) define summation procedures which are pairwise nonequivalent and also nonequivalent with logarithmic averaging. The result of Hardy also shows that by increasing the weight sequence (d_n) in (2.2), the result becomes stronger. Thus the strongest, "true" form of the a.s. central limit theorem is the one with the largest weight sequence (d_k). This weight sequence (d_k) is unknown and its determination will be the objective of the present chapter. We will also study the optimal weight sequences (d_k) in refined versions of (2.2), for example in the CLT and LIL corresponding to the strong law (2.2). Our main result will show that, under certain regularity conditions on (d_k), relation (2.2) and the corresponding CLT and LIL hold provided

$$d_n = O\left(\frac{D_n}{n(\log\log n)^\alpha}\right) \tag{2.5}$$

for $\alpha > 3$ and this becomes false if $\alpha < 1$. Here and in the sequel we write $\log\log x$ for $\log(\max\{\log x, e\})$. Thus the optimal weight condition for (2.2) and the corresponding CLT and LIL is relation (2.5) with some $1 \le \alpha \le 3$, whose value remains unknown. Condition (2.5) is an asymptotic negligibility condition resembling Kolmogorov's classical condition for the LIL, except the factor n in the denominator on the right hand side, which is due to the strong dependence of the sequence $(S_n/\sqrt{n})$ and which forces the summation D_n to grow considerably slower than the norming sequence in the classical LIL for independent r.v.'s. (Note that a similar effect of strong dependence leads to unusual coefficients in the LIL for lacunary trigonometric series, see Takahashi [105].) In terms of the norming factor D_n, our results show that (2.2) and the corresponding CLT and LIL are valid if

$$D_n = \exp(\log n/(\log\log n)^\alpha)$$

for $\alpha > 3$, but not if $\alpha < 1$. Recalling that $D_n = \log n$ resp. $D_n = n = \exp(\log n)$ correspond to logarithmic, resp. Cesàro averaging, the last relation shows the surprising fact that the critical weight sequence in the ASCLT is, in some sense, much closer to Cesàro than to logarithmic averaging. Thus, despite the prominent role log averaging plays in a.s. central limit theory, its true significance is secondary.

Incidentally, the methods of our paper will also lead to optimal conditions for the "stochastic" version of the ASCLT, i.e. when we require relation (2.2) in probability. Theorem 2.5 at the end of this section will show that the stochastic ASCLT holds if

$$d_n = o\left(D_n/n\right). \tag{2.6}$$

Replacing o by O in (2.6), the result becomes false, as the example $d_n = 1$ shows. Relation (2.6) holds e.g. if

$$D_n = \exp(\log n/\omega(n)) \tag{2.7}$$

where $\omega(n) \to \infty$, $\log n/\omega(n) \nearrow \infty$ and $\omega(n)$ satisfies mild regularity conditions. By Zygmund's theorem quoted above, the summation method defined by (2.7) is stronger than Cesàro summation (corresponding to $\omega(n) = 1$) and conversely, any summation method stronger than Cesàro summation and satisfying suitable regularity conditions has the above representation. Thus if we are interested in convergence in probability in (2.2), the summation method can be pushed arbitrary close to Cesàro summation.

We turn now to formulating our results in detail. In order to specify regularity conditions for the summation procedure $\mathbf{D}$, recall first that for each $\varepsilon > 0$, the sequence $D_n = n^\varepsilon$ defines a summation equivalent to Cesàro summation, and hence this D_n is already too large for (2.2). Thus without losing much generality, we can assume that $D_n = O(n^\varepsilon)$ for each $\varepsilon > 0$. All slowly varying sequences D_n satisfy this condition and confining our attention to this class will not put additional restrictions on the speed of growth of D_n, since for every sequence D_n satisfying $D_n = O(n^\varepsilon)$ for all $\varepsilon > 0$ there exists a slowly varying D_n^* such that $D_n = o(D_n^*)$ for $n \to \infty$ (cf. Bingham et al. [10, Theorem 2.3.6]). Hence in searching for the largest possible norming sequence D_n in ASCLT theory, we may assume that D_n is slowly varying. By the theory of regular variation, D_n can be represented in the form

$$D_n = c_n \exp\left(\int_A^n \varepsilon(u)/u\, du\right) \quad (n \ge A), \tag{2.8}$$

where $A > 0$, $c_n \to c \in (0, \infty)$, and $\varepsilon(x) \to 0$ for $x \to \infty$. Our final technical assumption on D_n will require that $c_n = 1$, ε is non-increasing, slowly varying and obeys the condition

$$\varepsilon(x)/\varepsilon(x^2) = O(1) \qquad (x \to \infty). \tag{2.9}$$

These conditions are stronger than necessary and could be easily weakened, but they will simplify our calculations considerably, and, as before, put no extra restrictions on the speed of growth of D_n. Regarding (2.9), note that $\varepsilon(x)$ must tend to 0 very slowly in order that $D_n \to \infty$. E.g., $\varepsilon(x) = (\log x)^{-(1+\eta)}$, $\eta > 0$ implies already that the exponent in (2.8) is bounded. However, this $\varepsilon(x)$ still satisfies (2.9). For the same reason, the assumption of slow variation of ε imposes only a regularity condition for $\varepsilon(x)$, but it puts no restriction on its speed of decrease.

The previously discussed technical conditions are summarized in the following

Definition. A summation method $\mathbf{D}$ belongs to the class $\mathcal{W}$ if $D_n \to \infty$ and

$$D_n = \exp\left(\int_A^n \varepsilon(u)/u\, du\right) \quad (n \ge A), \tag{2.10}$$

where $\varepsilon(x)$ is non-increasing, slowly varying, tends to 0 for $x \to \infty$, and satisfies (2.9).

From the mean value theorem it follows that

$$d_n \sim D_n \frac{\varepsilon(n)}{n} \quad (n \to \infty)$$

($a_n \sim b_n$ means that $a_n/b_n \to 1$) and thus $d_n = L(n)/n$ where L is slowly varying. We mention a few examples.

(a) $D_N = (\log N)^\gamma \quad \gamma > 0$;

(b) $D_N = \exp((\log N)^\beta)) \quad 0 < \beta < 1$;

(c) $D_N = \exp(\log N/(\log\log N)^\alpha) \quad \alpha > 0$.

Let $\mathcal{L}$ denote the class of bounded Lipschitz 1 functions on $\mathbb{R}$. By a standard observation in a.s. central limit theory (see e.g. Lacey and Philipp [69]), relation (2.2) follows if

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k=1}^N d_k f\left(\frac{S_k}{\sqrt{k}}\right) = \int_{-\infty}^{\infty} f(t) d\Phi(t) \quad \text{a.s.} \tag{2.11}$$

for every $f \in \mathcal{L}$. For this reason, we will work in our paper with the version of the ASCLT of the type (2.11).

We are ready now to formulate our results. In the sequel $\lambda(B)$ denotes the Lebesgue measure of some Borel set B. Finally we allude once more to the notation $\log\log x = \log(\max\{\log x, e\})$.

Theorem 2.1. *Let $X_1, X_2, \dots$ be i.i.d. random variables satisfying $EX_1 = 0$ and $EX_1^2 = 1$ and put $S_n = X_1 + \cdots + X_n$. Assume that $\mathbf{D} \in \mathcal{W}$, the relation*

$$d_k = O\left(\frac{D_k}{k(\log\log k)^\alpha}\right) \tag{2.12}$$

holds for some $\alpha > 3$ and

$$kd_k \text{ is non-decreasing.} \tag{2.13}$$

Then we have for every $f \in \mathcal{L}$ or f an indicator function of a Borel set A with $\lambda(\partial A) = 0$

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k=1}^N d_k \left(f\left(\frac{S_k}{\sqrt{k}}\right) - Ef\left(\frac{S_k}{\sqrt{k}}\right)\right) = 0 \quad \text{a.s.} \tag{2.14}$$

Theorem 2.2. *Let $X_1, X_2, \dots$ be i.i.d. random variables satisfying $EX_1 = 0$ and $EX_1^2 = 1$ and put $S_n = X_1 + \cdots + X_n$. Assume that $\mathbf{D} \in \mathcal{W}$, (2.12) holds for some $\alpha > 1$ and*

$$\liminf_{k\to\infty} kd_k > 0. \tag{2.15}$$

Then we have for every non-constant $f \in \mathcal{L}$

$$\lambda_N^{-1/2} \sum_{k=1}^N d_k \left(f\left(\frac{S_k}{\sqrt{k}}\right) - Ef\left(\frac{S_k}{\sqrt{k}}\right)\right) \xrightarrow{d} \mathcal{N}, \tag{2.16}$$

where

$$\lambda_N := \mathrm{Var}\left(\sum_{k=1}^N d_k f\left(\frac{S_k}{\sqrt{k}}\right)\right) \qquad (N \geq 1), \tag{2.17}$$

and $\mathcal{N}$ is a standard normal r.v.

Both relations (2.13) and (2.15) imply that $d_k \geq C/k$. In Theorem 2.1 this is indeed the interesting case, since relation (2.14) is known to hold for all $d_k \leq 1/k$ with $\sum d_k = \infty$, see Berkes and Csáki [9]. Regarding Theorem 2.2, we will show in Lemma 2.4 and Lemma 2.5 that if $\mathbf{D} \in \mathcal{W}$, then the order of magnitude of λ_N is $\sum_{k=1}^N kd_k^2$. Hence (2.15) implies that $\lambda_N \to \infty$. Without (2.15), Theorem 2.2 fails: for example, if $d_k = k^{-1}(\log k)^{-a}$ with $1/2 < a \leq 1$ (which is in $\mathcal{W}$), then $\limsup_N \lambda_N < \infty$ and thus the sum in (2.16) remains bounded in probability.

In the case $D_N = \log N$ Theorem 2.2 was proved by Berkes and Horváth [14]. Moreover, the conditions of Theorem 2.2 are satisfied in Example **(a)** if $\gamma \geq 1$, in Example **(b)** for all $0 < \beta < 1$ and in Example **(c)** if $\alpha > 1$.

Theorem 2.3. *Let $X_1, X_2, \dots$ be i.i.d. random variables satisfying $EX_1 = 0$ and $EX_1^2 = 1$ and put $S_n = X_1 + \cdots + X_n$. Let λ_N $(N \geq 1)$ be defined as in (2.17). Assume that $\mathbf{D} \in \mathcal{W}$ satisfies (2.15) and relation (2.12) holds for some $\alpha > 3$. Then we have for every non-constant $f \in \mathcal{L}$*

$$\limsup_{N\to\infty} (2\lambda_N \log\log \lambda_N)^{-1/2} \sum_{k=1}^N d_k \left(f\left(\frac{S_k}{\sqrt{k}}\right) - Ef\left(\frac{S_k}{\sqrt{k}}\right)\right) = 1 \quad \text{a.s.} \tag{2.18}$$

The conditions of Theorem 2.3 are satisfied in Example **(a)** if $\gamma \geq 1$, in Example **(b)** for all $0 < \beta < 1$ and in Example **(c)** if $\alpha > 3$.

Our theorems show that under regularity assumptions on $\mathbf{D}$, the strong law (2.14), the CLT (2.16) and the LIL (2.18) are all valid provided the Kolmogorov type condition (2.12) holds for $\alpha > 3$. The next theorem shows that except for the numerical value of α, condition (2.12) is sharp.

Theorem 2.4. *For every $0 < \alpha < 1$ there exists a summation procedure $\mathbf{D} \in \mathcal{W}$ satisfying (2.12) and (2.13) such that the LIL (2.18) fails.*

In fact, this is the case if $\mathbf{D}$ is defined by

$$D_N = \exp(\log N/(\log\log N)^\alpha). \tag{2.19}$$

Thus the sequence (2.19) is critical in the theory: for $\alpha > 3$ it implies all of (2.14), (2.16), (2.18) and this becomes false if $\alpha < 1$. What happens for $1 \leq \alpha \leq 3$ remains open.

By Hardy's minoration principle, if (2.2) holds with a weight sequence (d_k), then, under certain regularity conditions, it will also hold for all smaller weight sequences (d_k^*). Hardy assumed that $D_n^* = \psi(D_n)$, where $\psi(x) \leq x$ is an elementary function composed of rational, exponential and logarithmic functions. A much larger class of ψ's was constructed by Hirst [27, p. 37]. It seems likely that for sequences $d_k \geq 1/k$ an analogous minoration principle holds for the CLT (2.16) and the LIL (2.18), but this remains open. As the remarks after Theorem 2.2 show, without $d_k \geq 1/k$ this minoration principle is not valid.

Our final result gives a sharp condition for relation (2.14) to hold in probability.

Theorem 2.5. *Let $X_1, X_2, \dots$ be i.i.d. random variables satisfying $EX_1 = 0$ and $EX_1^2 = 1$ and put $S_n = X_1 + \cdots + X_n$. Assume that*

$$d_n = o\left(D_n/n\right). \tag{2.20}$$

Then we have for every $f \in \mathcal{L}$ or f an indicator function of a Borel set A with

$\lambda(\partial A) = 0$

$$\frac{1}{D_N}\sum_{k=1}^{N} d_k \left(f\left(\frac{S_k}{\sqrt{k}}\right) - Ef\left(\frac{S_k}{\sqrt{k}}\right)\right) \to 0 \quad \text{in probability.}$$

If o in (2.20) is replaced with O the proposition fails.

In comparison to (2.12) we can omit the the extra factor $(\log\log n)^{\alpha}$ in the denominator for the stochastic version of the ASCLT. As the example $d_n = 1$ (Cesáro summation) shows, we cannot replace o with O in (2.20). For example relation (2.20) holds for

$$D_n = \exp(\log n/\omega(n))$$

if $w(x)$ is some differentiable function satisfying $w(x) \to \infty$ (in order that D_n is not equivalent to Cesáro averaging) and $\log x/w(x) \nearrow \infty$ (in order that D_n defines a summation method) and

$$x\omega'(x)/\omega(x) = O\left((\log x)^{-1}\right). \tag{2.21}$$

The last condition is satisfied e.g. if $\omega(x) = \log_k x \quad (k = 2, 3, \ldots)$ where $\log_k$ denotes k times iterated logarithm. It is also easy to see that (2.21) permits arbitrary slow increase of ω.

2.2 Auxiliary lemmas

Lyapunov's classical CLT condition or Kolmogorov's condition for the LIL provide the corresponding limit theorems in terms of specific moment assumptions. Via a blocking technique the proofs of our theorems will make use of these results. Hence an accurate study of the variances respectively higher moments of the processes under investigation is important. The core of this section are Lemma 2.4 and Lemma 2.5. Lemma 2.3 and Lemma 2.4 will be needed in a more general form in Chapter 3. Therefore the proofs will follow in this Chapter. In what follows, let f be a bounded Lipschitz 1 function on $\mathbb{R}$; without loss of generality we assume $|f| \le 1$. All the constants occurring in the following lemmas may depend on f, $\mathbf{D}$ and the sequence $X_1, X_2, \ldots$; we will make no mention of this fact in the sequel. Constants like C_p, $C(\varepsilon)$, etc. may depend also on the parameters indicated. The relation $a_n \ll b_n$ will mean $|a_n/b_n| = O(1)$.

Lemma 2.1. *Assume $f \in \mathcal{L}$. Then there is a constant C such that for all $1 \le k \le l$*

$$\left|\operatorname{Cov}\left(f\left(\frac{S_k}{\sqrt{k}}\right), f\left(\frac{S_l}{\sqrt{l}}\right)\right)\right| \le C\left(\frac{k}{l}\right)^{1/2}.$$

Lemma 2.1 is a standard tool in a.s. central limit theory. For completeness, we give the short proof. Clearly

$$\operatorname{Cov}\left(f\left(\frac{S_k}{\sqrt{k}}\right), f\left(\frac{S_l}{\sqrt{l}}\right)\right) = \operatorname{Cov}\left(f\left(\frac{S_k}{\sqrt{k}}\right), f\left(\frac{S_l}{\sqrt{l}}\right) - f\left(\frac{S_l - S_k}{\sqrt{l}}\right)\right).$$

Hence the the Lipschitz continuity of f, $|f| \le 1$ and the Cauchy-Schwarz inequality give

$$\begin{aligned}&\left|\operatorname{Cov}\left(f\left(\frac{S_k}{\sqrt{k}}\right), f\left(\frac{S_l}{\sqrt{l}}\right)\right)\right| \\ &\qquad \le 2E\left|f\left(\frac{S_l}{\sqrt{l}}\right) - f\left(\frac{S_l - S_k}{\sqrt{l}}\right)\right| \le CE\left|\frac{S_k}{\sqrt{l}}\right| \le C\sqrt{\frac{k}{l}}. \qquad (2.22)\end{aligned}$$

Lemma 2.2. *Let $f \in \mathcal{L}$, f non-constant. Then there exist an integer $m \ge 1$, a real $c > 0$ and for every $\varepsilon > 0$ an $A = A(\varepsilon)$ such that*

$$\operatorname{Cov}\left(f\left(\frac{S_k}{\sqrt{k}}\right), f\left(\frac{S_l}{\sqrt{l}}\right)\right) \ge c\left(\frac{k}{l}\right)^{m/2} \quad \textit{for} \quad A \le k < l, \quad k/l \ge \varepsilon^{2/m}.$$

Proof. Consider a Wiener process $\{W_t, t \ge 0\}$. From Rozanov [93, 182 f.] we get

$$\operatorname{Cov}\left(f\left(\frac{W_k}{\sqrt{k}}\right), f\left(\frac{W_l}{\sqrt{l}}\right)\right) = \sum_{\nu=1}^{\infty} \frac{\rho^\nu}{\nu\,!}\,\alpha_\nu^2 \qquad (1 \le k \le l), \qquad (2.23)$$

where α_ν are the coefficients of the Hermite expansion of $g := f - Ef(W_1)$, i.e.

$$g(x) = \sum_{\nu=1}^{\infty} \frac{\alpha_\nu}{\nu\,!} H_\nu(x),$$

$$H_\nu(x) = (-1)^\nu e^{x^2/2} \frac{d^\nu}{dx^\nu} e^{-x^2/2}$$

and $\rho = \sqrt{k/l}$ is the correlation between $W_k/\sqrt{k}$ and $W_l/\sqrt{l}$ for $k \le l$. Since we exclude the trivial case where f is constant, there is a $\nu \ge 1$ such that $\alpha_\nu^2 > 0$. Let m be the smallest of these integers. By (2.23) there exists a $c_1 > 0$ such that

$$\operatorname{Cov}\left(f\left(\frac{W_k}{\sqrt{k}}\right), f\left(\frac{W_l}{\sqrt{l}}\right)\right) \ge c_1 \left(\frac{k}{l}\right)^{m/2} \qquad (1 \le k \le l). \tag{2.24}$$

To estimate the covariance in the general case we use an invariance principle of Major [77] which implies that we can define $X_1, X_2, \dots$ on a new probability space together with a Wiener process $\{W_t,\, t \ge 0\}$ such that

$$(S_n - W_n)/\sqrt{n} \xrightarrow{P} 0.$$

Set

$$c_{k,l} := \operatorname{Cov}\left(f\left(\frac{S_k}{\sqrt{k}}\right), f\left(\frac{S_l}{\sqrt{l}}\right)\right), \quad c_{k,l}^* := \operatorname{Cov}\left(f\left(\frac{W_k}{\sqrt{k}}\right), f\left(\frac{W_l}{\sqrt{l}}\right)\right).$$

Since f is a bounded Lipschitz function, it is easy to see that

$$|c_{k,l}^* - c_{k,l}| \to 0 \quad \text{for} \quad \min\{k,l\} \to \infty.$$

Hence for every $\varepsilon > 0$ there is an $A = A(\varepsilon)$ such that

$$|c_{k,l} - c_{k,l}^*| \le \varepsilon^2 \quad \text{for} \quad \min\{k,l\} \ge A. \tag{2.25}$$

By (2.24) we have

$$c_{k,l}^* \ge c_1 \varepsilon \quad \text{if} \quad k/l \ge \varepsilon^{2/m}, \quad 1 \le k \le l,$$

and together with (2.25) and again (2.24) this yields

$$c_{k,l} \ge c_{k,l}^* \left(1 - \frac{\varepsilon}{c_1}\right) \ge (c_1 - \varepsilon)\left(\frac{k}{l}\right)^{m/2} \quad \text{if} \quad k/l \ge \varepsilon^{2/m}, \quad A \le k \le l.$$

Since we may assume that $\varepsilon < c_1/2$ this proves the assertion.

To simplify the notation, in the sequel we will write for $1 \le k \le l$

$$\xi_l := f\left(\frac{S_l}{\sqrt{l}}\right) - Ef\left(\frac{S_l}{\sqrt{l}}\right) \quad \text{and} \quad \xi_{k,l} := f\left(\frac{S_l - S_k}{\sqrt{l}}\right) - Ef\left(\frac{S_l - S_k}{\sqrt{l}}\right).$$

$\square$

Lemma 2.3. *If $k \le m \le n$ and (d_l) is an arbitrary sequence of positive numbers, then we have for every $p \in \mathbb{N}$*

$$E\left|\sum_{l=m}^{n} d_l(\xi_l - \xi_{k,l})\right|^p \le E_p(k/m)^{1/2}\left(\sum_{l=m}^{n} l d_l^2\right)^{p/2},$$

where

$$E_p = \text{const} \cdot 4^p(2p)^{p/2}.$$

Proof. See the proof of Lemma 3.1. □

Lemma 2.4. *Let $\mathbf{D} = (d_k)$ be a summation method with $d_k = L(k)/k, \ \ k \ge 1$, where $L(k) \gg 1$ and $L(k)$ is slowly varying at infinity. Then for every $f \in \mathcal{L}$ and every $p \in \mathbb{N}$*

$$E\left|\sum_{k=m}^{n} d_k\left(f\left(\frac{S_k}{\sqrt{k}}\right) - Ef\left(\frac{S_k}{\sqrt{k}}\right)\right)\right|^p \le C_p\left(\sum_{k=m}^{n} k d_k^2\right)^{p/2}, \qquad (2.26)$$

where $C_p > 0$ is a constant.

Proof. See the proof of Lemma 3.2. □

Lemma 2.5. *Let $\mathbf{D} = (d_k)$ be a summation method with $d_k = L(k)/k, \ \ k \ge 1$, where $L(k) \gg 1$ and $L(k)$ is slowly varying at infinity. Then for every non-constant $f \in \mathcal{L}$ we have*

$$\operatorname{Var}\left(\sum_{k=1}^{N} d_k f\left(\frac{S_k}{\sqrt{k}}\right)\right) \gg \sum_{k=1}^{N} k d_k^2. \qquad (2.27)$$

Proof. Let $0 < \varepsilon < 1$ to be chosen later and $A(\varepsilon)$ and m the same as in Lemma 2.2. Set $\delta = \varepsilon^{2/m}$. Clearly it suffices to prove Lemma 2.5 with the summations in (2.27) started with $k = A$ instead of $k = 1$. Now

$$\begin{aligned}
&\operatorname{Var}\left(\sum_{k=A}^{N} d_k f\left(\frac{S_k}{\sqrt{k}}\right)\right)\\
&= \sum_{A \le k \le N} d_k^2 E\xi_k^2 + 2\sum_{\substack{A \le i < k \le N \\ i/k < \delta}} d_i d_k E\xi_i\xi_k + 2\sum_{\substack{A \le i < k \le N \\ i/k \ge \delta}} d_i d_k E\xi_i\xi_k\\
&=: S^{(1)} + S^{(2)} + S^{(3)}.
\end{aligned}$$

Clearly $\sum_{k=1}^{\infty} d_k^2 < \infty$, and since the ξ_k are uniformly bounded, it follows that $S^{(1)} = O(1)$. Note that for a slowly varying function L we have

$$\sum_{k=1}^{N} L(k)k^{\rho} \sim \frac{1}{\rho+1} L(N)N^{\rho+1} \quad \text{if} \quad \rho > -1, \tag{2.28}$$

(see e.g. [20, Corollary 1.7.3]). From $d_k = L(k)/k$ with a slowly varying L we get by (2.28) and the definition of slow variation that for every $\gamma > 0$

$$\sum_{1 \le i < \delta k} d_i i^{\gamma} \sim \frac{1}{\gamma} L(\delta k)(\delta k)^{\gamma} \sim \frac{1}{\gamma} \delta^{\gamma} L(k) k^{\gamma} \quad (k \to \infty).$$

Hence for $k \ge k_0(\delta)$ we have

$$\delta^{\gamma} \frac{1}{2\gamma} k^{\gamma+1} d_k \le \sum_{1 \le i < \delta k} d_i i^{\gamma} \le \delta^{\gamma} \frac{2}{\gamma} k^{\gamma+1} d_k. \tag{2.29}$$

By Lemma 2.1 and (2.29) we get

$$\begin{aligned} |S^{(2)}| &\le 2C \sum_{\substack{1 \le i < k \le N \\ i/k < \delta}} d_i d_k \left(\frac{i}{k}\right)^{1/2} = 2C \sum_{1 \le k \le N} d_k k^{-1/2} \sum_{1 \le i < \delta k} d_i i^{1/2} \\ &\le 2C \sum_{1 \le k \le k_0} d_k k^{-1/2} \sum_{1 \le i < \delta k} d_i i^{1/2} + 8C\delta^{1/2} \sum_{1 \le k \le N} k d_k^2 \\ &= 8C\delta^{1/2} \sum_{1 \le k \le N} k d_k^2 + R_1, \end{aligned}$$

where $R_1 = R_1(\delta)$. Similarly we get from (2.29)

$$\sum_{\substack{1 \le i < k \le N \\ i/k < \delta}} d_i d_k \left(\frac{i}{k}\right)^{m/2} \le \frac{4}{m} \delta^{m/2} \sum_{1 \le k \le N} k d_k^2 + R_2 \tag{2.30}$$

and

$$\sum_{A \le i < k \le N} d_i d_k \left(\frac{i}{k}\right)^{m/2} \ge \frac{1}{m} \sum_{k=1}^{N} k d_k^2 - R_3, \tag{2.31}$$

where $R_2, R_3 > 0$ depend on δ. Now Lemma 2.2 gives

$$\begin{aligned} S^{(3)} &\ge 2c \sum_{\substack{A \le i < k \le N \\ i/k \ge \delta}} d_i d_k \left(\frac{i}{k}\right)^{m/2} \\ &\ge 2c \left(\sum_{A \le i < k \le N} d_i d_k \left(\frac{i}{k}\right)^{m/2} - \sum_{\substack{1 \le i < k \le N \\ i/k < \delta}} d_i d_k \left(\frac{i}{k}\right)^{m/2} \right). \end{aligned}$$

Note finally that $L(k) \gg 1$ implies $\sum_{k\geq 1} kd_k^2 = \infty$. Choosing ε small will also make δ small, and thus combining the estimate for $S^{(3)}$ with (2.30)-(2.31) and using the estimates for $S^{(1)}$, $S^{(2)}$ proves the lemma. □

Lemma 2.6. *Let* $\mathbf{D}$ *be a summation method with* $d_k = O(1)$. *If for some* $\alpha > 0$

$$d_N = O\left(\frac{D_N}{N(\log\log N)^\alpha}\right),$$

it follows that

$$\log\frac{D_N}{D_M} \ll (\log\log M)^{-\alpha}\log\frac{N}{M} \qquad (M_0 \leq M < N).$$

Proof. From $D_k \to \infty$ and $d_k = O(1)$ we conclude that $D_{k+1}/D_k \to 1$ and

$$\log\frac{D_{k+1}}{D_k} \ll \frac{d_{k+1}}{D_k} \ll \frac{1}{k(\log\log k)^\alpha}.$$

Hence $\sum_{k=M}^{N-1}\log\frac{D_{k+1}}{D_k} \ll (\log\log M)^{-\alpha}\sum_{k=M}^{N-1}\frac{1}{k}$. □

2.3 Proofs

The method to prove our theorems is based on a blocking technique. We partition $\mathbb{N}$ into disjoint blocks: $\mathbb{N} = A_1 \cup B_1 \cup A_2 \cup B_2 \cup \ldots$, where

$$A_j = \{2^{p_j'}+1,\ldots,2^{q_j'}\} \quad \text{and} \quad B_j = \{2^{p_j}+1,\ldots,2^{q_j}\} \quad (j \geq 1).$$

We set $q_0 = 1$ and for some $r > 0$ (to be chosen later) define the exponents as

$$p_j' = q_{j-1}, \quad q_j' = p_j' + [12\log j], \quad p_j = q_j', \quad q_j = p_j + [12(\log j)^{1+r}].$$

Obviously the length of the blocks B_j will grow much faster than the length of A_j and the block A_j precedes B_j on the real line. Set

$$M_j = 2^{p_j}, \quad N_j = 2^{q_j}, \quad M_j' = 2^{p_j'}, \quad N_j' = 2^{q_j'} \qquad (j \geq 1).$$

Recalling the definition of ξ_k and $\xi_{k,l}$ we define now

$$Z_j := \sum_{k\in B_j} d_k\xi_k, \qquad Z_j^* := \sum_{k\in B_j} d_k\xi_{N_{j-1},k} \quad (j \geq 1),$$

$$R_j := \sum_{k\in A_j} d_k\xi_k, \qquad R_j^* := \sum_{k\in A_j} d_k\xi_{N'_{j-1},k} \quad (j\geq 1).$$

To visualize this, the sum Z_j is

$$\sum_{k=M_j+1}^{N_j} d_k\left(f\left(\frac{S_k}{\sqrt{k}}\right) - Ef\left(\frac{S_k}{\sqrt{k}}\right)\right),$$

and Z_j^* is obtained from Z_j by replacing S_k in the previous sum by $S_k - S_{N_{j-1}}$, so that Z_j^* involves only X_k's with $N_{j-1} < k \leq N_j$. It follows that the Z_j^* are independent random variables which we will use as approximations for the original random variables Z_j. Similarly, the R_j^* are independent r.v.'s. We will call the Z_j and Z_j^* 'long block sums', R_j and R_j^* 'short block sums'. We stress again that the short block sum R_j precedes the long block sum Z_j on the real line. In a first step we derive limit theorems for the sequences (Z_j^*) and (R_j^*) via classical theorems for independent random variables. We will see that the contribution of the R_j and the error we make by replacing Z_j by Z_j^* will be small, so that our results carry over to the sequence $(Z_j + R_j)$. Observing that

$$\sum_{j=1}^{n}(Z_j + R_j) = \sum_{k=1}^{N_n} d_k\left(f\left(\frac{S_k}{\sqrt{k}}\right) - Ef\left(\frac{S_k}{\sqrt{k}}\right)\right) \tag{2.32}$$

shows that the desired results hold along the subsequence (N_j). In order to permit fast growing summation methods D_k and to provide that the limit relations hold along the subsequence (N_j), we had to choose "large" blocks. However, in a final step we have to control the fluctuation between the subsequence (N_j), which forces us to use "smaller" blocks. Hence in this stage of the proof the key for optimal results lies in an optimal choice of the block lengthes.

In order to give upper bounds for the moments of Z_j, Z_j^*, R_j and R_j^* we introduce the notation

$$V_j = V_j(\mathbf{D}) := \sum_{k\in B_j} kd_k^2 \quad \text{and} \quad U_j = U_j(\mathbf{D}) := \sum_{k\in A_j} kd_k^2\,. \tag{2.33}$$

For $p\in\mathbb{N}$ we get immediately from Lemma 2.3 and Lemma 2.4

$$\begin{aligned} E|Z_j - Z_j^*|^p &\leq 2^{(q_{j-1}-p_j)/2} E_p V_j^{p/2}, \\ E|R_j - R_j^*|^p &\leq 2^{(q'_{j-1}-p'_j)/2} E_p U_j^{p/2}, \end{aligned} \tag{2.34}$$

and

$$E|Z_j|^p \leq C_p V_j^{p/2} \quad \text{and} \quad E|R_j|^p \leq C_p U_j^{p/2}. \tag{2.35}$$

Proof of Theorem 2.2. From Lyapunov's theorem (cf. Petrov [88, Theorem 4.9]) it follows that the sequence (Z_n^*) obeys the CLT if we show that

$$\frac{\sum_{j=1}^n E|Z_j^*|^4}{\operatorname{Var}^2\left(\sum_{j=1}^n Z_j^*\right)} \to 0. \tag{2.36}$$

Hence we need estimates for the expectation of the perturbed random variables $(\sum_{j=1}^n Z_j^*)^2$ and $\sum_{j=1}^n (Z_j^*)^4$. A look at (2.34) and (2.35) makes it evident that we have to study the order of magnitude of U_j and V_j given in (2.33). In order to handle this issue we use the the representation of D_n in (2.10). It implies that

$$\sum_{k=m}^n k d_k^2 \asymp \int_m^n \exp\left(2\int_A^x \varepsilon(u)/u\,du\right) \varepsilon^2(x)/x\,dx \quad (m, n \to \infty). \tag{2.37}$$

(For numerical sequences (a_n) and (b_n) let $a_n \asymp b_n$ mean $a_n \ll b_n$ and $b_n \ll a_n$.) The monotonicity of the terms in the integrand of (2.37) and condition (2.9) yield

$$U_j \ll \int_{M'_j}^{N'_j} \exp\left(2\int_A^x \varepsilon(u)/u\,du\right) \varepsilon^2(x)/x\,dx \ll \varepsilon^2(M'_j) D^2_{N'_j} \log j,$$

where we used

$$\int_{M'_j}^{N'_j} 1/x\,dx \sim 12 \log j.$$

Similarly we get

$$V_j \gg \varepsilon^2(N_j) D^2_{M_j} (\log j)^{1+r}.$$

A simple calculation shows that

$$p_j \sim q_j \sim p'_j \sim q'_j \sim 12 j (\log j)^{1+r}, \tag{2.38}$$

and thus condition (2.9) implies that the ratio $\varepsilon(M_j')/\varepsilon(N_j)$ remains bounded for $j \to \infty$ and since $N_j' = M_j$ this shows that $U_j/V_j \to 0$. By condition (2.15) we clearly have $\sum_{j=1}^n V_j \to \infty$ and hence we obtain from the Minkowski inequality, Lemma 2.4, the independence of R_j^* and (2.34)-(2.35)

$$\begin{aligned}
\left| \Big\| \sum_{j=1}^n Z_j^* \Big\|_2 - \Big\| \sum_{j=1}^n (Z_j + R_j) \Big\|_2 \right| &\le \Big\| \sum_{j=1}^n (Z_j^* - (Z_j + R_j)) \Big\|_2 \\
&\le \Big\| \sum_{j=1}^n R_j^* \Big\|_2 + \sum_{j=1}^n \Big\| Z_j - Z_j^* \Big\|_2 + \sum_{j=1}^n \Big\| R_j - R_j^* \Big\|_2 \\
&\ll \left(\sum_{j=1}^n U_j \right)^{1/2} + \sum_{j=1}^n j^{-3\log 2} V_j^{1/2} = o\left(\Big(\sum_{j=1}^n V_j \Big)^{1/2} \right).
\end{aligned}$$

(Here $\| \cdot \|_2$ denotes the L_2 norm). We can now compare the variances λ_N defined in (2.17) to $\mathrm{Var}(\sum_{j=1}^n Z_j^*)$. Combining the latter estimate with Lemma 2.4 and Lemma 2.5 shows that

$$\mathrm{Var}\left(\sum_{j=1}^n Z_j^* \right) \sim \lambda_{N_n} \quad \text{and} \quad \lambda_{N_n} \asymp \sum_{k=1}^{N_n} k d_k^2 \asymp \sum_{j=1}^n V_j. \tag{2.39}$$

From (2.34) and (2.35) we get that

$$\sum_{j=1}^n E|Z_j^*|^4 \ll \sum_{j=1}^n V_j^2 \le \max\{V_1, \cdots, V_n\} \sum_{j=1}^n V_j \qquad (n \to \infty).$$

On the other hand (2.39) gives $\mathrm{Var}(\sum_{j=1}^n Z_j^*) \gg \sum_{j=1}^n V_j$. Thus

$$\frac{\sum_{j=1}^n E|Z_j^*|^4}{\mathrm{Var}^2\left(\sum_{j=1}^n Z_j^*\right)} \ll \frac{\max_{1\le j\le n} V_j}{\sum_{j=1}^n V_j}. \tag{2.40}$$

From (2.37) we get further

$$\sum_{k=1}^N k d_k^2 \gg \varepsilon(N) D_N^2 \quad \text{and} \quad V_n \ll \varepsilon(M_n) D_{N_n}^2 \int_{M_n}^{N_n} \varepsilon(x)/x \, dx. \tag{2.41}$$

Using again (2.39), (2.41), (2.9), (2.10), the explicit formulas for M_n, N_n (see (2.38)) and (2.12) combined with Lemma 2.6 we derive

$$\frac{V_n}{\sum_{j=1}^n V_j} \ll V_n \left(\sum_{k=1}^{N_n} k d_k^2\right)^{-1} \ll \frac{\varepsilon(M_n)}{\varepsilon(N_n)} \int_{M_n}^{N_n} \varepsilon(x)/x\,dx$$
$$\ll \int_{M_n}^{N_n} \varepsilon(x)/x\,dx \ll \log \frac{D_{N_n}}{D_{M_n}} \ll (\log n)^{(1+r-\alpha)}\,.$$

Since $\alpha > 1$ in (2.12), we can chose $0 < r < \alpha - 1$, which shows

$$V_n/(V_1 + \cdots + V_n) \to 0 \qquad (n \to \infty). \tag{2.42}$$

Together with (2.40) this proves (2.36), i.e. the central limit theorem holds for the random variables (Z_j^*).

In the next step we show that $\sum_{j=1}^n (Z_j + R_j)$ and $\sum_{j=1}^n Z_j^*$ are "close" to each other. This shows that the CLT is valid for the sequence $(Z_j + R_j)$ as well. Observe that by (2.34)

$$\sum_{j=1}^{\infty} \left(\frac{E(Z_j - Z_j^*)^2}{\sum_{1\le l\le j} V_l}\right)^{1/2} < \infty,$$

and consequently

$$\sum_{j=1}^{\infty} \frac{|Z_j - Z_j^*|}{(\sum_{1\le l\le j} V_l)^{1/2}} < \infty \quad \text{a.s.}$$

Thus by the Kronecker lemma and (2.39) we have

$$\frac{1}{\lambda_{N_n}^{1/2}} \sum_{j=1}^n (Z_j - Z_j^*) \to 0 \quad \text{a.s.} \tag{2.43}$$

By the same arguments it follows easily that the last relation holds with $(R_j - R_j^*)$ instead of $(Z_j - Z_j^*)$. Further (2.34)-(2.35), the independence of R_j^* and the Cauchy-Schwarz inequality give

$$E\left|\sum_{j=1}^n R_j^*\right| \ll \left(\sum_{j=1}^n U_j\right)^{1/2}.$$

We have already shown that $U_j/V_j \to 0$. Hence the Markov inequality and (2.39) yield

$$P\left(\left|\sum_{j=1}^n R_j^*\right| > \varepsilon\,\lambda_{N_n}^{1/2}\right) = o(1) \qquad \text{for every } \varepsilon > 0.$$

Using $Z_j + R_j = Z_j^* + R_j^* + (Z_j - Z_j^*) + (R_j - R_j^*)$ and recalling (2.32) we proved that

$$\lambda_{N_n}^{-1/2} \sum_{k=1}^{N_n} d_k \xi_k \xrightarrow{d} \mathcal{N}.$$

In order to finish the proof of Theorem 2.2 it suffices now to show that

$$\lim_j \sup_{N \in (N_{j-1}, N_j]} E|T_N - T_{N_j}| = 0, \tag{2.44}$$

where

$$T_N = \lambda_N^{-1/2} \sum_{k=1}^{N} d_k \xi_k \,.$$

By Minkowski's inequality and Lemma 2.4 we get for $N_{j-1} \le N \le N_j$

$$|\lambda_{N_j}^{1/2} - \lambda_N^{1/2}| \le \mathrm{Var}^{1/2} \left(\sum_{k=N+1}^{N_j} d_k \xi_k \right) \le (4\gamma)^2 (U_j + V_j)^{1/2}. \tag{2.45}$$

This shows that

$$\begin{aligned} E|T_N - T_{N_j}| &\le |\lambda_N^{-1/2} - \lambda_{N_j}^{-1/2}| \, E \left| \sum_{k=1}^{N} d_k \xi_k \right| + \lambda_{N_j}^{-1/2} \, E \left| \sum_{k=N+1}^{N_j} d_k \xi_k \right| \\ &\le \lambda_{N_j}^{-1/2} \left(\left| \lambda_{N_j}^{1/2} - \lambda_N^{1/2} \right| + \mathrm{Var}^{1/2} \left(\sum_{k=N+1}^{N_j} d_k \xi_k \right) \right) \\ &\le 2(4\gamma)^2 \lambda_{N_j}^{-1/2} (U_j + V_j)^{1/2}. \end{aligned}$$

Applying (2.39) and (2.42) we can show (2.44). This finishes the proof of Theorem 2.2.
□

Proof of Theorem 2.3. Trivially $|Z_j^*| \le 2(D_{N_j} - D_{M_j})$. In order to apply Kolmogorov's law of the iterated logarithm (cf. Petrov [88, p. 239] to the sequence (Z_j^*), it suffices to verify that

$$D_{N_j} - D_{M_j} = o\left(\left(\frac{s_j^2}{\log\log s_j^2} \right)^{1/2} \right),$$

where $s_j^2 = \mathrm{Var}(\sum_{l=1}^{j} Z_l^*)$. First note that by the representation (2.10)

$$\begin{aligned} D_{N_j} - D_{M_j} &= D_{N_j}\left(1 - \exp\left(-\int_{M_j}^{N_j} \epsilon(u)/u\,du\right)\right) \\ &\le D_{N_j}\int_{M_j}^{N_j} \epsilon(u)/u\,du. \end{aligned} \tag{2.46}$$

Relation (2.39) and the first statement of (2.41) show that

$$D_{N_j}^2 \varepsilon(N_j) \ll s_j^2 \ll D_{N_j}^2 \ll N_j. \tag{2.47}$$

Using (2.9), (2.46), (2.47) and the explicit formulas for M_n, N_n we derive

$$\begin{aligned} &(D_{N_j} - D_{M_j})\left(\frac{s_j^2}{\log\log s_j^2}\right)^{-1/2} \\ &\ll \frac{D_{N_j} - D_{M_j}}{D_{N_j}}\left(\frac{\log\log N_j}{\varepsilon(N_j)}\right)^{1/2} \ll \left(\frac{\log\log N_j}{\varepsilon(N_j)}\right)^{1/2}\int_{M_j}^{N_j}\varepsilon(u)/u\,du \\ &\ll (\log\log N_j)^{1/2}\int_{M_j}^{N_j}\varepsilon^{1/2}(u)/u\,du. \end{aligned} \tag{2.48}$$

From Lemma 2.6, (2.10) and the Cauchy-Schwarz inequality it follows that

$$\begin{aligned} &\int_{M_j}^{N_j}\varepsilon^{1/2}(u)/u\,du \le \left(\int_{M_j}^{N_j}\varepsilon(u)/u\,du\right)^{1/2}\left(\int_{M_j}^{N_j}1/u\,du\right)^{1/2} \\ &\quad \ll \left(\log\frac{D_{N_j}}{D_{M_j}}\right)^{1/2}\left(\log\frac{N_j}{M_j}\right)^{1/2} \ll \left(\log\frac{N_j}{M_j}\right)(\log\log M_j)^{-\alpha/2}. \end{aligned} \tag{2.49}$$

By the definition of M_j and N_j we see that the last expression of (2.48) is bounded by const $\cdot\, (\log j)^{3/2-\alpha/2+r}$. This tends to zero if we choose $0 < r < (\alpha - 3)/2$. Remember that by (2.39) $\lambda_{N_j} \sim s_j^2$. Thus setting $L_N = (2\lambda_N \log\log\lambda_N)^{1/2}$ we get from Kolmogorov's law of the iterated logarithm that

$$\limsup_{j\to\infty} L_{N_j}^{-1} \sum_{l=1}^{j} Z_l^* = 1 \quad \text{a.s.}$$

Next we observe that by (2.43) we have also

$$\limsup_{j\to\infty} L_{N_j}^{-1} \sum_{l=1}^{j} Z_l = 1 \quad \text{a.s.}$$

Similar arguments show that for the short block sums R_l we get

$$L_{N_j}^{-1} \sum_{l=1}^{j} R_l \to 0 \quad \text{a.s.}$$

Thus by (2.32) the LIL (2.18) is true along the subsequence (N_j).

Finally we are confronted with the task of controlling the maximal fluctuation between the subsequence (N_j). Now for $N_{j-1} \le N < N_j$ a trivial estimate (using $|\xi_k| \le 2$) shows that

$$\begin{aligned} &\left| L_N^{-1} \sum_{k=1}^{N} d_k \xi_k - L_{N_{j-1}}^{-1} \sum_{k=1}^{N_{j-1}} d_k \xi_k \right| \\ &\qquad \le \left| \frac{L_{N_{j-1}}}{L_N} - 1 \right| L_{N_{j-1}}^{-1} \left| \sum_{k=1}^{N_{j-1}} d_k \xi_k \right| + 2 L_N^{-1} (D_{N_j} - D_{N_{j-1}}). \end{aligned}$$

We have already proved that $L_{N_{j-1}}^{-1} \left| \sum_{k=1}^{N_{j-1}} d_k \xi_k \right| = O(1)$ a.s. Hence Theorem 2.3 will be proved if we show that

$$\lim_{j\to\infty} \sup_{N_{j-1} \le N < N_j} \left| \frac{L_{N_{j-1}}}{L_N} - 1 \right| = 0 \tag{2.50}$$

and

$$\limsup_{j\to\infty} \sup_{N_{j-1} \le N < N_j} L_N^{-1} (D_{N_j} - D_{N_{j-1}}) = 0. \tag{2.51}$$

Relations (2.39), (2.42) and (2.45) imply that

$$\sup_{N_{j-1} \le N < N_j} \left| \lambda_{N_j}^{1/2} - \lambda_N^{1/2} \right| = o(\lambda_{N_j}^{1/2}) \quad \text{as } j \to \infty \tag{2.52}$$

which immediately yields (2.50). To show (2.51) we first observe that by (2.39), (2.41), (2.52) and (2.9) we have for $N_{j-1} \le N < N_j$

$$L_N^{-1}(D_{N_j} - D_{N_{j-1}}) \le \lambda_N^{-1/2}(D_{N_j} - D_{N_{j-1}}) \ll D_{N_{j-1}}^{-1} \varepsilon^{-1/2}(N_j)(D_{N_j} - D_{N_{j-1}}).$$

By using the same argument as in (2.46) we get an upper bound for $D_{N_j} - D_{N_{j-1}}$. Thus by (2.9) and observing that $N_j \le N_{j-1}^2$ for large enough j we obtain

$$L_N^{-1}(D_{N_j} - D_{N_{j-1}}) \ll \frac{\varepsilon^{-1/2}(N_j) D_{N_j}}{D_{N_{j-1}}} \int_{N_{j-1}}^{N_j} \varepsilon(u)/u\, du \ll \frac{D_{N_j}}{D_{N_{j-1}}} \int_{N_{j-1}}^{N_j} \varepsilon^{1/2}(u)/u\, du.$$

An analogue of (2.49), Lemma 2.6 and (2.38) can be used to derive

$$\begin{aligned}
&L_N^{-1}(D_{N_j} - D_{N_{j-1}}) \\
&\quad \ll \frac{D_{N_j}}{D_{N_{j-1}}} \left(\log \frac{N_j}{N_{j-1}} \right) (\log\log N_{j-1})^{-\alpha/2} \\
&\quad \ll \left(\frac{N_j}{N_{j-1}} \right)^{(\log\log N_{j-1})^{-\alpha}} \left(\log \frac{N_j}{N_{j-1}} \right) (\log\log N_{j-1})^{-\alpha/2} \\
&\quad \ll \exp\left(\text{const} \cdot (\log j)^{1+r-\alpha} \right) (\log j)^{1+r-\alpha/2}.
\end{aligned}$$

This tends to zero if we choose $0 < r < \alpha/2 - 1$. □

Proof of Theorem 2.1. Lemma 2.4 shows (actually under weaker conditions than we assumed in Theorem 2.1) that

$$\lambda_N \ll \sum_{k=1}^N k d_k^2.$$

If we assume that $k d_k$ is non-decreasing, we get

$$\sum_{k=1}^N k d_k^2 \le N d_N D_N, \qquad N = 1, 2, \ldots$$

Hence by relation (2.12) we conclude that

$$\begin{aligned}
(\lambda_N \log\log \lambda_N)^{1/2} &\ll (N d_N D_N \log\log N)^{1/2} \\
&\ll D_N (\log\log N)^{(1-\alpha)/2}.
\end{aligned}$$

This shows that for $\alpha > 1$

$$(\lambda_N \log\log \lambda_N)^{1/2} = o(D_N),$$

and thus for $f \in \mathcal{L}$ Theorem 2.1 follows from Theorem 2.3.

The proof for indicator functions follows from routine approximation arguments as e.g. those in [69]. $\square$

Proof of Theorem 2.5. We will show that

$$\operatorname{Var}\left(\frac{1}{D_N}\sum_{k=1}^{N} d_k\left(f\left(\frac{S_k}{\sqrt{k}}\right) - Ef\left(\frac{S_k}{\sqrt{k}}\right)\right)\right) = \lambda_N/D_N^2 \to 0.$$

By (2.20) we have for every $\varepsilon > 0$ an $N_0(\varepsilon)$, such that for all $N \geq N_0$

$$\sup_{k\leq N} \frac{kd_k}{D_N} < \varepsilon. \tag{2.53}$$

Remember the definition of ξ_k right before Lemma 2.3. By Lemma 2.1 we get

$$E\left(\sum_{k=m}^{n} d_k\xi_k\right)^2 \leq 2\sum_{m\leq k\leq l\leq n} d_k d_l\,|E\xi_k\xi_l| \leq 2C\sum_{m\leq l\leq n} d_l l^{-1/2}\sum_{1\leq k\leq l} d_k k^{1/2},$$

and hence by (2.53) we have for $N \geq N_0$

$$\lambda_N/D_N^2 \ll \frac{1}{D_N}\sum_{1\leq l\leq N} d_l l^{-1/2}\sum_{1\leq k\leq l}\frac{kd_k}{D_N}k^{-1/2} \leq 2\varepsilon.$$

$\square$

Consider the summation $D_N = \log N$. By Lemma 2.4 and Lemma 2.5 we have $\lambda_N \asymp D_N$. On the other hand if $D_N = N$, then it's possible to show (using Lemma 2.1 and Lemma 2.2) that $\lambda_N \asymp D_N^2$. Trivially, a necessary condition for the LIL is $\lambda_N \log\log\lambda_N \leq D_N^2$. If the numerical value α of the weight sequence (D_N) defined in (2.19) falls below the critical value 1 the last condition is violated. The proof of Theorem 2.4 is based on this simple observation.

Proof of Theorem 2.4. Let

$$D_N = \exp(\log N/(\log\log N)^\alpha) \quad \text{where } 0 < \alpha < 1. \tag{2.54}$$

A simple calculation shows that the representation (2.10) holds with $c = 1$ and

$$\varepsilon(x) = (\log\log x)^{-\alpha} - \alpha(\log\log x)^{-\alpha-1} \sim (\log\log x)^{-\alpha} \tag{2.55}$$

and thus this $\mathbf{D}$ is in $\mathcal{W}$. Also, by (2.54) and the mean value theorem

$$d_N = D_N - D_{N-1} \sim \frac{1}{N}\exp(\log N/(\log\log N)^\alpha)(\log\log N)^{-\alpha}$$

and thus relations (2.12) and (2.13) are valid. Finally, Lemma 5, (2.41) and (2.55) yield

$$\lambda_N \gg \sum_{k=1}^{N} k d_k^2 \gg \varepsilon(N) D_N^2 \gg D_N^2 (\log\log N)^{-\alpha}$$

whence

$$(\lambda_N \log\log \lambda_N)^{1/2} \gg D_N (\log\log N)^{(1-\alpha)/2}.$$

Since $|\sum_{k=1}^N d_k \xi_k| \le 2D_N$, relation (2.18) cannot hold if $\alpha < 1$. □

2.4 A conjecture for the critical weights

We have proved that the critical weight sequence for the LIL (Theorem 2.3) is given by (2.19) with some unknown $\alpha \in [1, 3]$. In contrast to Theorem 2.3, we cannot conclude that the ASCLT Theorem 2.1 fails for $\alpha < 1$. We formulate now a conjecture based on a heuristic argument which states that for both theorems the critical value is $\alpha = 1$. First we prove the following result.

Proposition 2.1. *Let $\xi_1, \xi_2, \dots$ be a sequence of independent random variables with $E\xi_k = 0$ and $E\xi_k^2 = \sigma_k^2$ and $|\xi_k| \le M_k < \infty$. Set $B_n^2 = \sum_{k=1}^n \sigma_k^2$ and assume that*

$$M_k = o\left(B_k (\log\log B_k)^{-1/2}\right). \tag{2.56}$$

Then

$$\frac{1}{\sum_{k=1}^n \sigma_k} \sum_{k=1}^n \xi_k \to 0 \quad \text{a.s.}$$

Proof. The Kolmogorov condition (2.56) implies that

$$B_n^2 \leq \max_{1\leq k\leq n} \sigma_k \sum_{k=1}^{n} \sigma_k = o(B_n(\log\log B_n)^{-1/2} \sum_{k=1}^{n} \sigma_k).$$

This gives $B_n(\log\log B_n)^{1/2} = o(\sum_{k=1}^n \sigma_k)$. Now the Proposition follows from Theorem A2. □

Proposition 2.1 is even optimal. Berkes and Csáki [9, p. 119] construct a counter-example which implies that o in the Kolmogorov condition cannot be replaced by O. Specifically, they consider $\xi_k = c(k)X_k$, where X_k are i.i.d. Bernoulli r.v.'s and $c(k) = \exp(k/\log k)$. It is easy to see that in this case (2.56) holds with O. Hence by the second part of Theorem A2 we have for a sequence of bounded random variables that (2.56) is sharp for both, the LIL and the LLN.

Conjecture. *Theorem 2.1 and Theorem 2.3 are valid if the Kolmogorov type condition (2.12) holds with $\alpha > 1$ and both theorems fail if $\alpha \leq 1$.*

This conjecture contains four assertions. One of them, namely the failure of the LIL in case of $\alpha \leq 1$, has been already shown (except in the case $\alpha = 1$). We shall give now a heuristic argument why also the ASCLT should fail if $\alpha \leq 1$. We assume that $X_1, X_2, \ldots$ are i.i.d. standard normal r.v.'s. Further we let f be a Lipschitz function with $|f| \leq 1$ and consider the specific averaging methods defined by

$$D_\alpha(t) = \exp(\log t(\log\log t)^{-\alpha})$$

at $t = 1, 2, \ldots$ Finally, we set $d_\alpha(t) = dD_\alpha(t)/dt$. It is natural to expect that

$$\frac{1}{D_\alpha(N)} \sum_{k=1}^{N} d_\alpha(k) \left(f(S_k/\sqrt{k}) - Ef(N) \right) \to 0 \quad \text{a.s.} \tag{2.57}$$

fails if $\alpha \leq 1$. The continuous version of (2.57) is

$$\frac{1}{D_\alpha(N)} \int_1^N d_\alpha(t) \left(f(W_t/\sqrt{t}) - Ef(N) \right) dt \to 0 \quad \text{a.s.},$$

where $\{W_t,\, t \geq 0\}$ is a standard Brownian motion process. By the parameter transformation $t \mapsto e^s$ the last relation holds if

$$\frac{1}{D_\alpha(e^n)} \int_0^n d_\alpha(e^t)e^t \left(f(U_t) - Ef(N)\right) dt \to 0 \quad \text{a.s.}, \tag{2.58}$$

where $U_t = W(e^t)e^{-t/2}$ is the (stationary) Ornstein-Uhlenbeck process. Let $C_\alpha(n) = D_\alpha(e^n)$ and $c_\alpha(n) = d_\alpha(e^n)e^n$. It's easy to see that

$$\sum_{k=1}^{n} c_\alpha(n) \sim C_\alpha(n) \quad \text{and} \quad c_\alpha(n)/c_\alpha(n+1) \to 1.$$

The fact that $c_\alpha(t)$ grows only subexponentially yields

$$\int_n^{n+1} c_\alpha(t) \left(f(U_t) - Ef(N)\right) dt \sim c_\alpha(n) \int_n^{n+1} \left(f(U_t) - Ef(N)\right) dt \quad (n \to \infty)$$

and thus (2.58) will follow if we show

$$\frac{1}{C_\alpha(n)} \sum_{k=1}^{n} c_\alpha(k)\eta_k \to 0 \quad \text{a.s.}, \tag{2.59}$$

where $\eta_k = \int_k^{k+1} \left(f(U_t) - Ef(N)\right) dt$. Note that the sequence (η_k) is stationary and $|\eta_k| \leq 1$. This reduces our problem to a weighted law of large numbers for a bounded, stationary sequence. If we replace the η_k in (2.59) with an i.i.d. sequence (η_k^*) satisfying $E\eta_1^* = 0$ and $|\eta_1^*| \leq 1$, we can apply Proposition 2.1. In fact, routine calculations show that Kolmogorov's condition (2.56) is satisfied if $\alpha > 1$ and it fails if $\alpha \leq 1$.

Whether Proposition 2.1 can be extended to our specific dependent sequence as well, remains open. However, the example of Berkes and Csáki [9] shows that not even an i.i.d. Bernoulli sequence is $C_1(n)$-summable (and consequently not $C_\alpha(n)$-summable if $\alpha \leq 1$). Thus the conjecture on the failure of the ASCLT with $\alpha \leq 1$ is evident. In any case this transformation argument exhibits a strong connection between Kolmogorov's condition (2.56) and the crucial assumption (2.12).

Chapter 3

A sharpening of the universal a.s. limit theorem

3.1 Introduction and results

Most results in the ASCLT literature concern partial sum behavior, i.e. they state relations of the type

$$D_n^{-1} \sum_{k=1}^{n} d_k I\{S_k/a_k \le x\} \to \Phi(x) \quad \text{a.s.} \tag{3.1}$$

There exist, however, a few extensions for nonlinear functionals of independent r.v.'s as well: Marcus and Rosen [80], Csáki and Földes [30] and Horváth and Khoshnevisan [61] obtained analogues of (3.1) for local times and Cheng et al. [28] and Fahrner and Stadtmüller [43] proved a similar result for extreme order statistics. In particular, in [28] and [43] it is shown that if $X_1, X_2, \dots$ are i.i.d. r.v.'s such that for some numerical sequences (a_k), (b_k) we have

$$(\max_{i \le k} X_i - a_k)/b_k \xrightarrow{d} G$$

with a nondegenerate distribution G, then

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k \le N} \frac{1}{k} I\left\{ \frac{\max_{i\le k} X_i - a_k}{b_k} \le x \right\} = G(x) \qquad \text{a.s. for any } x \in \mathbb{R}. \tag{3.2}$$

For further examples of nonlinear extensions of (3.1), see Antonini and Weber [1], Mörters [81], [82], Peng and Qi [87] and Hörmann [58]. In view of these results the question arises which distributional limit theorems have similar a.s. versions and Berkes and Csáki [9] proved the surprising result that *every* weak limit theorem for independent random variables, subject to minor technical assumptions, has an a.s. "logarithmic" version. Specifically, they proved the following result:

Theorem B. (Universal almost sure limit theorem). *Let $X_1, X_2, \ldots$ be independent random variables satisfying the weak limit theorem*

$$g_l(X_1, X_2, \ldots, X_l) \xrightarrow{d} G, \tag{3.3}$$

where $g_l : \mathbb{R}^l \to \mathbb{R}$ $(l = 1, 2 \ldots)$ are measurable functions and G is a distribution function. Assume that for each $1 \le k < l$ there exists a measurable function $g_{k,l} : \mathbb{R}^{l-k} \to \mathbb{R}$ such that

$$E\left(|g_l(X_1, \ldots, X_l) - g_{k,l}(X_{k+1}, \ldots, X_l)| \wedge 1\right) \le A(c_k/c_l) \tag{3.4}$$

with a constant $A > 0$ and a positive, nondecreasing sequence (c_k) satisfying $c_k \to \infty$, $c_{k+1}/c_k = O(1)$. Then we have

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k\le N} d_k I\left\{f_k(X_1, \ldots, X_k) < x\right\} = G(x) \qquad \text{a.s. for any } x \in C_G, \tag{3.5}$$

where

$$d_k = \log(c_{k+1}/c_k), \quad D_N = \log c_N \tag{3.6}$$

and C_G denotes the set of continuity points of G.

The simplest form (3.1) of the ASCLT is obtained by letting

$$g_l = (X_1 + \ldots + X_l)/\sqrt{l}, \qquad g_{k,l} = (X_{k+1} + \ldots + X_l)/\sqrt{l},$$

where (X_n) is an i.i.d. sequence with $EX_1 = 0$, $EX_1^2 = 1$. It is easy to see that in this case (3.4) is satisfied with $c_k = k^\alpha$ for some $\alpha > 0$ and thus (3.5) holds with $d_k \sim$ const$\cdot 1/k$. Similarly, the limit theorem (3.2) is obtained with

$$g_l(X_1, \ldots, X_l) = (\max_{1\le i\le l} X_i - a_l)/b_l, \quad g_{k,l}(X_{k+1}, \ldots, X_l) = (\max_{k+1\le i\le l} X_i - a_l)/b_l$$

and in this case (3.4) is satisfied with $c_k = k$, leading again to $d_k \sim \text{const}\cdot 1/k$. In both cases (3.4) means that changing the first $o(l)$ terms of the sequence $X_1, \dots, X_l$ has very little effect on the normed partial sum $(X_1 + \cdots + X_l)/\sqrt{l}$ respectively the extremal statistics

$$(\max_{1\le i\le l} X_i - a_l)/b_l.$$

An example for a limit theorem with a different c_k is the Darling-Erdős theorem stating

$$a_l \left(\max_{i\le l} \frac{S_i}{\sqrt{i}} - b_l \right) \xrightarrow{d} e^{-e^{-x}} \tag{3.7}$$

for suitable (a_l) and (b_l), where S_l are partial sums of independent r.v.'s with mean 0, variance 1 and uniformly bounded $(2+\delta)$-th moments. (See Darling and Erdős [33] for $\delta = 1$ and Shorack [100] for $\delta > 0$.) Here (3.4) holds with $c_k = \log k$, and (3.4) means that changing the first $l^{o(1)}$ terms of the sequence $X_1, \dots, X_l$ will change the left hand side of (3.7) only unessentially. Note that in this case the initial segment of $X_1, \dots, X_l$ not influencing the value of $g_l(X_1, \dots, X_l)$ is much shorter: the dependence of g_l on its initial variables became more sensitive. As a consequence, we get a different weight sequence in (3.5): instead of $d_k = 1/k$ we get now $d_k = 1/(k \log k)$, $D_N = \log\log N$, i.e. in the a.s. version of the Darling-Erdős limit theorem we have loglog averages. We thus see that the more sensitively the functional $g_l(x_1, \dots, x_l)$ depends on its initial variables, the smaller weight sequence (d_k) in (3.5) is obtained.

For a detailed discussion of Theorem B and several examples as well as more refined versions of the latter result we refer to Berkes and Csáki [9].

The purpose of the present chapter is to show that, similarly to the ordinary ASCLT discussed in the previous chapter, the universal almost sure limit theorem can also be substantially sharpened by replacing the "canonical" averaging method (given by (3.6)) by suitable, weaker averaging methods. Due to the general character of the limit theorem (3.3), however, the methods of the previous chapter are not applicable in the present case, preventing us from getting the same complete characterization as in the case of the ASCLT. However, we will show that a slightly modified version of the Kolmogorov condition (2.5) remains sufficient even for the most general limit theorem (3.3).

In Section 3.2 we will give several examples. Specifically, in Section 3.2.2 we give an example where the "canonical" weights (i.e. the weights provided by Theorem B) are $d_k = 1/k$, but the considered a.s. convergence result actually holds with Cesàro averaging, i.e. $d_k = 1$.

Theorem 3.1. *Let $X_1, X_2, \dots$ be independent r.v.'s such that for some measurable functions $g_l : \mathbb{R}^l \to \mathbb{R}$ the weak limit theorem*

$$g_l(X_1, \dots, X_l) \xrightarrow{d} G \tag{3.8}$$

holds with some distribution function G. Assume further that for some functions $g_{k,l} : \mathbb{R}^{l-k} \to \mathbb{R}$, $1 \le k < l$, we have

$$E\left(|g_l(X_1, \dots, X_l) - g_{k,l}(X_{k+1}, \dots, X_l)| \wedge 1\right) \le A(c_k/c_l)^\beta \tag{3.9}$$

for some A, $\beta > 0$ and some positive nondecreasing sequence (c_k) with $c_k \to \infty$. Finally set $d_k^* = \log(c_{k+1}/c_k)$ and assume that the Kolmogorov type condition

$$d_k = O\left(d_k^* \frac{D_k}{(\log D_k)^\rho}\right) \tag{3.10}$$

is satisfied for some $\rho > 0$ and in addition

$$d_k \gg d_k^*, \qquad d_k/(d_k^* c_k^\beta) \quad \textit{is nonincreasing} \tag{3.11}$$

and

$$d_k D_k/(\log D_k)^\rho = O(1) \tag{3.12}$$

hold. Then if f is a bounded Lipschitz 1 function on the real line or is the indicator function of a Borel set A with $G(\partial A) = 0$, we have

$$\lim_{N\to\infty} D_N^{-1} \sum_{k=1}^N d_k f\left(g_k(X_1, \dots, X_k)\right) = \int_{-\infty}^{\infty} f(x)\, dG(x) \quad \text{a.s.} \tag{3.13}$$

In view of the result of Hardy quoted in Section 2.1 the validity of relation (3.13) automatically extends to smaller weight sequences (d_k), and thus we are interested in finding large weight sequences such that (3.13) is valid. Since (3.13) holds for $d_k = d_k^*$, the first assumption $d_k \gg d_k^*$ in (3.11) is a natural one. The second condition of (3.11) and relation (3.12) clearly limit the speed of growth of D_k from above. In view of relation (3.27) below, we have $\sum_{k=1}^n d_k^* c_k^\beta \sim \text{const} \cdot c_k^\beta$, and thus the second condition in (3.11) implies $D_k \ll c_k^\beta$. On the time scale determined by the c_k, the weights d_k^* give logarithmic averaging, while $D_k = c_k^\beta$ is equivalent, by the theorem of Zygmund, with (rescaled) ordinary Cesàro averaging belonging to $D_k = c_k$. Thus the second relation of (3.11) limits D_k above by Cesàro averaging and the same holds with relation (3.12), which implies that the increments of $D_k^2(\log D_k)^{-\rho}$ are bounded and thus $D_k \ll k^{1/2}(\log k)^{\rho/2}$. Restricting the summation procedures in (3.13) above by Cesàro averaging is quite natural, since ordinary averaging is usually too large even in the simplest versions of the ASCLT. However, the example given in Subsection 3.2.2 will show that in certain non-i.i.d. situations one can actually use even the Cesàro weights in the general strong limit theorem (3.13).

The crucial condition in Theorem 3.1 is (3.10). It tells us how far we can move from the natural logarithmic weights (d_k^*) towards a larger weight sequence (d_k). As we will see (Section 3.2), in typical cases (3.10) permits choosing (d_k) much closer to the corresponding Cesàro averaging than to logarithmic weights.

The following theorem covers the case when the functionals g_k depend not on an independent sequence (X_k), but on a more general stochastic process $\{X_t, t \geq 0\}$ with independent increments.

Theorem 3.2. *Let $\{X_t, t \geq 0\}$ be a stochastic process with independent increments and assume $X_0 = 0$. Let η_l, $l = 1, 2, \ldots$ be r.v.'s such that η_l is measurable with respect to $\sigma\{X(t), t \leq l\}$ and assume that the weak limit theorem*

$$\eta_l \xrightarrow{d} G$$

holds with some distribution function G. Assume finally that for every $1 \leq k < l$ there exists a $\sigma\{X(t) - X(s), k \leq s \leq t \leq l\}$-measurable random variable $\eta_{k,l}$ such that

$$E\left(|\eta_l - \eta_{k,l}| \wedge 1\right) \leq A(c_k/c_l)^\beta \quad (1 \leq k < l), \tag{3.14}$$

with $A,\ \beta > 0$ and some nondecreasing sequence (c_k) with $c_k \to \infty$. If (3.10)-(3.12) are satisfied with $d_k^* = \log(c_{k+1}/c_k)$, then for any bounded Lipschitz 1 function f on the real line or for the indicator function f of any Borel set $A \subset \mathbb{R}$ with $G(\partial A) = 0$, we have

$$\lim_{N\to\infty} D_N^{-1} \sum_{k=1}^{N} d_k f(\eta_k) = \int_{-\infty}^{\infty} f(x)\, dG(x) \quad \text{a.s.} \tag{3.15}$$

3.2 Examples

In this section we will give some applications of Theorems 3.1 and 3.2. In each case we will specify the natural weights d_k^* and the corresponding best weight sequence our theorems provide. In what follows, f denotes a bounded Lipschitz 1 function or the indicator function of a Borel set $A \subset \mathbb{R}$ with $G(\partial A) = 0$, where G is some distribution function occurring as a weak limit of the form (3.8) in the given examples.

3.2.1 Partial sums of i.i.d. r.v.'s.

Let $X_1, X_2, \ldots$ be i.i.d. r.v.'s and let S_l denote the l-th partial sum. Assume that there exist numerical sequences (a_l) and (b_l) such that a weak limit theorem of the form

$$g_l(X_1, \ldots, X_l) := \frac{S_l}{a_l} - b_l \xrightarrow{d} G$$

holds with some (possibly degenerate) distribution function G. Assume further that $\sup_{l\geq 1} E|S_l/a_l - b_l|^\nu < \infty$ for some $\nu > 0$. We choose

$$g_{k,l}(X_{k+1}, \ldots, X_l) = \frac{S_l - S_k}{a_l} - b_l \quad (1 \leq k \leq l).$$

A standard argument in a.s. central limit theory shows that there exist positive constants C and β such that

$$E(|g_l(X_1, \ldots, X_l) - g_{k,l}(X_{k+1}, \ldots, X_l)| \wedge 1) \leq C(a_k/a_l)^\beta.$$

For example, if $EX_1 = 0$ and $EX_1^2 = 1$, then $a_l = \sqrt{l}$ and Theorem 3.1 applies with $d_k^* = 1/k$ and with G denoting the standard normal distribution function. If

X_k are i.i.d. r.v.'s belonging to the domain of attraction of a stable distribution G, then (a_l) is regularly varying with exponent $1/\alpha$ for some $0 < \alpha \leq 2$ and from the representation theorem for regularly varying functions (cf. Bingham et al. [20, p. 12]) we obtain easily that $a_k/a_l \leq C'(k/l)^{\beta'}$. Hence in this case the natural weights d_k^* are again the logarithmic ones. On the other hand, it is easily checked that for

$$D_N = \exp((\log N)^{\alpha}), \quad d_k = D_k - D_{k-1}, \quad 0 < \alpha < 1, \tag{3.16}$$

conditions (3.10)-(3.12) are satisfied, and thus Theorem 3.1 yields the stronger result

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k=1}^{N} d_k f\left(\frac{S_k}{a_k} - b_k\right) = \int_{-\infty}^{\infty} f(x)\, dG(x) \quad \text{a.s.}$$

3.2.2 Sums of not identically distributed r.v.'s.

Let X_k be independent r.v.'s with $EX_k = 0$ and $EX_k^2 = s_{k+1}^2 - s_k^2$, where $s_k^2 = e^{(\log k)^{1+\varepsilon}}$, $\varepsilon > 0$, and let S_k denote the k-th partial sum. Assume that the sequence (X_k) satisfies the Lindeberg condition and let g_l and $g_{k,l}$ be the same as in the last example with $a_l = s_l$ and $b_l = 0$. Then (3.9) is satisfied with $c_k = s_k$ and thus we get the natural weights $d_k^* = (\log k)^{\varepsilon} k^{-1}$. From a summability point of view, the summation method defined by (d_k^*) is equivalent to log summation. However, Theorem 3.1 shows the surprising fact that in this case, even Cesàro means work in (3.13). Indeed, define $d_k = 1/\sqrt{k}$. Again conditions (3.10)-(3.12) are easily checked. By Zygmund's theorem the summation defined by $d_k = 1/\sqrt{k}$ is equivalent to Cesàro summation. Hence in this example we get

$$\frac{1}{N} \sum_{k=1}^{N} f\left(\frac{S_k}{s_k}\right) \to \int_{-\infty}^{\infty} f(x)\, d\Phi(x) \quad \text{a.s.} \tag{3.17}$$

Clearly, the faster the sequence (s_k) grows, the more influence X_k has on the partial sum S_k and the smaller is the dependence between the $f(S_k/s_k)$. For example if $s_k^2 = \log k$ then the dependence between the $f(S_k/s_k)$ is much stronger than in the standard case $s_k^2 = k$ and Example 1 in Berkes and Dehling [11, p. 1649] shows that the ASCLT fails with $d_k = k^{-1}$. In this case the natural weights are $d_k^* = 1/(k \log k)$

and the norming sequence is $D_N^* = \log\log N$. But an application of Theorem 3.1 shows that the ASCLT actually holds in this example with the larger sequence $D_N = \exp((\log\log N)^\alpha)$ for any $0 < \alpha < 1$.

3.2.3 Subsequences

As an immediate consequence of the first example in 3.2.2 we get an almost sure central limit theorem for i.i.d. r.v.'s along subsequences, using Cesàro summation. Let X_k be i.i.d. r.v.'s with $EX_1 = 0$ and $EX_1^2 = 1$ and set $n_k = [s_k^2]$, where $s_k^2 = \exp((\log k)^{1+\varepsilon})$ with $\varepsilon > 0$. Define $Y_k = S_{n_k} - S_{n_{k-1}}$. Clearly Y_k are independent, $EY_k = 0$, $EY_k^2 = n_k - n_{k-1}$ and by the central limit theorem $S_{n_k}/\sqrt{n_k} \xrightarrow{d} N(0,1)$. Hence we have by (3.17)

$$\frac{1}{N}\sum_{k=1}^{N} f\left(\frac{S_{n_k}}{\sqrt{n_k}}\right) \to \int_{-\infty}^{\infty} f(x)\, d\Phi(x) \quad \text{a.s.} \tag{3.18}$$

Schatte [96] and Atlagh and Weber [3] showed a similar result for $n_k = [c^k]$ with $c > 1$. Of course, the faster n_k grows, the less is the dependence between the partial sums S_{n_k} and the weaker summation methods apply. Rychlik and Szuster [94] showed that (3.18) also holds for $n_k = [c^{k^\alpha}]$ with $c > 1$ and $\alpha > 0$. The last result shows that we can weaken the growth rates used in [96], [3] and [94] of (n_k) substantially. Note also that the growth condition for the subsequence (n_k) is sharp in some sense. Choosing $\varepsilon = 0$ gives $n_k = k$ and then (3.18) fails (cf. [96, Theorem 1]).

3.2.4 Sample extremes

Let $X_1, X_2, \ldots$ be i.i.d. r.v.'s. Assume further that there are numerical sequences (a_l) and (b_l) such that for some distribution function G a weak limit theorem of the form

$$g_l(X_1, \ldots, X_l) := \frac{\max_{i\le l} X_i - a_l}{b_l} \xrightarrow{d} G$$

is valid. Then setting

$$g_{k,l}(x_{k+1}, \ldots, x_l) = \frac{\max_{k<i\le l} x_i - a_l}{b_l}$$

we get (cf. [9, p. 122])

$$E(|g_l(X_1,\dots,X_l)-g_{k,l}(X_{k+1},\dots,X_l)|\wedge 1)\le \frac{k}{l}.$$

Thus Theorem A yields

$$\frac{1}{D_N^*}\sum_{k=1}^N d_k^* f\left(\frac{\max_{i\le l}X_i-a_l}{b_l}\right)\to\int_{-\infty}^{\infty}f(x)\,dG(x)\quad\text{a.s.,}\tag{3.19}$$

where the natural weights d_k^* are the logarithmic ones. As observed in Fahrner and Stadtmüller [43] and in Cheng et. al. [28], if we replace logarithmic means in (3.19) by the ordinary Cesàro means, the result fails. However, Theorem 3.1 shows that (3.19) is still true when the summation method (D_N^*) is replaced by (D_N) defined in (3.16).

3.2.5 The Darling-Erdős theorem

Let $\mathbb{W}$ be a Wiener process and

$$\eta_l=a_l\left(\sup_{1\le t\le l}\frac{\mathbb{W}(t)}{\sqrt{t}}-b_l\right)$$

where

$$a_l=(2\log\log l)^{1/2}\quad\text{and}\quad b_l=a_l+\frac{\log\log\log l-\log 4\pi}{2a_l}\quad(l\ge 3).$$

By the Darling-Erdős theorem (see [33]) we have $\eta_l\xrightarrow{d}G$ where $G(x)=e^{-e^{-x}}$. Let $c_l=\exp(\sqrt{\log\log l})$, $A_l=\exp(\log l/\exp(\sqrt{\log\log l}))$ and

$$\eta_{k,l}=\begin{cases}a_l\left(\sup_{A_l^2\le t\le l}\frac{\mathbb{W}(t)-\mathbb{W}(k)}{\sqrt{t}}-b_l\right) & \text{if } k\le A_l\\ 0 & \text{otherwise.}\end{cases}$$

In Berkes and Csáki [9] it is shown that

$$E(|\eta_l-\eta_{k,l}|\wedge 1)\le 4(c_k/c_l)^{1/2}\quad(k\le l).\tag{3.20}$$

Also, η_l is measurable with respect to $\sigma\{\mathbb{W}(t), t \le l\}$ and $\eta_{k,l}$ is measurable with respect to $\sigma\{\mathbb{W}(t') - \mathbb{W}(t) : k \le t \le t' \le l\}$. Thus Theorem 5 in Berkes and Csáki [9] implies

$$\lim_{N\to\infty} \frac{1}{D_N} \sum_{k=1}^{N} d_k f\left(a_k \left(\sup_{1\le t\le k} \frac{\mathbb{W}(k)}{\sqrt{k}} - b_k\right)\right) = \int_{-\infty}^{\infty} f(x)\, dG(x) \quad \text{a.s.,} \tag{3.21}$$

where

$$d_k = \frac{1}{k \log k \sqrt{\log\log k}}, \qquad D_N = \sqrt{\log\log N}. \tag{3.22}$$

By Zygmund's theorem, the averaging procedure determined by the weight sequence in (3.22) is equivalent to loglog averaging, i.e. (3.21) holds also with

$$d_k = \frac{1}{k \log k}, \qquad D_N = \log\log N.$$

On the other hand, using relation (3.20) with $c_l = \exp(\sqrt{\log\log l})$, Theorem 3.2 implies (3.21) with the considerably larger weights

$$d_k \sim \frac{1}{k \log k} e^{(\log\log k)^\alpha}, \quad D_N \sim e^{(\log\log N)^\alpha} \quad 0 < \alpha < 1/2.$$

More generally, consider a nondecreasing, unbounded function $h : \mathbb{R}^+ \to \mathbb{R}^+$ with $1 \le h(x) \le x$ and set

$$\eta_k^{(h)} = a_k^{(h)} \left(\sup_{k/h(k)\le t\le k} \frac{\mathbb{W}(t)}{\sqrt{t}} - b_k^{(h)} \right)$$

where

$$a_k^{(h)} = (2\log\log h(k))^{1/2} \quad \text{and} \quad b_k^{(h)} = a_k^{(h)} + \frac{\log\log\log h(k) - \log 4\pi}{2a_k^{(h)}}$$

Then a slightly more general form of the Darling-Erdős theorem yields

$$\eta_k^{(h)} \xrightarrow{d} G. \tag{3.23}$$

An a.s. version of the last theorem was obtained in Berkes and Weber [18] who showed that

$$\frac{1}{D_N} \sum_{k=1}^{N} d_k f(\eta_k^{(h)}) \to \int_{-\infty}^{\infty} f(x)\, dG(x) \quad a.s., \tag{3.24}$$

where the weights d_k depend on the function h: the slower h grows, the stronger summation method is required in (3.24). The dependence of the weights d_k on h is rather involved, so we consider a simple special case, e.g. $h(x) = e^{(\log x)^\alpha}$, $0 < \alpha < 1$. Letting

$$\eta_{k,l}^{(h)} = \begin{cases} a_l^{(h)} \left(\sup_{R_l \le t \le l} \frac{\mathbb{W}(t) - \mathbb{W}(k)}{\sqrt{t}} - b_l^{(h)} \right) & \text{if } k \le R_l \\ 0 & \text{otherwise,} \end{cases}$$

with $R_l = l/h(l)$, in Berkes and Weber [18] it is shown that the analogue of (3.20) for $\eta_l^{(h)}$, $\eta_{k,l}^{(h)}$ holds with $c_k = e^{2(\log k)^{1-\alpha}}$. We can now apply Theorem 3.2 and get (3.24) with the natural weights $d_k^* = \frac{1}{k(\log k)^\alpha}$. By Zygmund's theorem this summation is equivalent to log summation, which is exactly the result derived in [18]. But Theorem 3.2 shows that we can use also the summation defined by $D_N = e^{(\log N)^\gamma}$ with $0 < \gamma < 1 - \alpha$, giving a stronger result.

In our previous considerations we dealt with the Darling-Erdős theorem for the Wiener process, but with a simple invariance argument like in [9], [18], [100], we can extend the results for partial sums of independent random variables with mean 0, variance 1 and uniformly bounded $(2 + \delta)$-th moments.

3.3 Proofs

In what follows, we will give the proof of Theorem 3.1; the proof of Theorem 3.2 is very similar. We first prove some preparatory lemmas. Let f be a Lipschitz 1 function with $|f| \le 1$ and put for $1 \le k < l$

$$\begin{aligned} \xi_l &:= f\left(g_l(X_1, \ldots, X_l)\right) - Ef\left(g_l(X_1, \ldots, X_l)\right) \\ \xi_{k,l} &:= f\left(g_{k,l}(X_{k+1}, \ldots, X_l)\right) - Ef\left(g_{k,l}(X_{k+1}, \ldots, X_l)\right), \end{aligned} \tag{3.25}$$

where (X_k) is a sequence of independent random variables and $g_l : \mathbb{R}^l \to \mathbb{R}$ and $g_{k,l} : \mathbb{R}^{l-k} \to \mathbb{R}$ are measurable functions. Here, and in the sequel, the constants c, α, etc. depend only on the sequences (X_k), (g_l), $(g_{k,l})$ and f. Constants like C_p, E_p will depend on the parameter p as well.

Lemma 3.1. *Define ξ_l and $\xi_{k,l}$ as in (3.25) and assume that (3.9) holds. If (d_k) is a numerical sequence satisfying (3.11), then we have for any $k \le m \le n$ and $p \in \mathbb{N}$*

$$E\left|\sum_{l=m}^{n} d_l(\xi_l - \xi_{k,l})\right|^p \le E_p \left(\frac{c_k}{c_m}\right)^\beta \left(\sum_{l=m}^{n} d_l c_l^{-\beta} \left(\sum_{k=1}^{l} d_k c_k^\beta\right)\right)^{p/2},$$

where

$$E_p = K^p p^{p/2},$$

with some constant K.

Proof. We set $Q(l) = Q(k,l) = \xi_l - \xi_{k,l}$. Trivially $|Q(l)| \le 4$ and thus

$$\begin{aligned} E|Q(l)|^p &\le 4^{p-1} E|Q(l)| \le 2 \cdot 4^{p-1} E|f(g_l(X_1, \ldots, X_l)) - f(g_{k,l}(X_{k+1}, \ldots, X_l))| \\ &\le \text{const} \cdot 4^p E\left(|g_l(X_1, \ldots, X_l) - g_{k,l}(X_{k+1}, \ldots, X_l)| \wedge 1\right). \end{aligned}$$

By (3.9) we get thus

$$E|Q(l)|^p \le K_0 \cdot 4^p (c_k/c_l)^\beta,$$

with some $K_0 > 0$. Using the Hölder and the Cauchy-Schwarz inequality we get, setting $d_l^* = \log(c_{l+1}/c_l)$,

$$\begin{aligned} E\left|\sum_{l=m}^{n} d_l(\xi_l - \xi_{k,l})\right|^p &\le \sum_{l_1=m}^{n} \cdots \sum_{l_p=m}^{n} d_{l_1} \cdots d_{l_p} \left(E|Q(l_1)|^p \cdots E|Q(l_p)|^p\right)^{1/p} \\ &\le K_0 \cdot 4^p c_k^\beta \sum_{l_1=m}^{n} \cdots \sum_{l_p=m}^{n} d_{l_1} \cdots d_{l_p}\, c_{l_1}^{-\beta/p} \cdots c_{l_p}^{-\beta/p} \\ &= K_0 \cdot 4^p c_k^\beta \left(\sum_{l=m}^{n} d_l c_l^{-\beta/p}\right)^p \\ &\le K_0 \cdot 4^p c_k^\beta \left(\sum_{l=m}^{n} d_l^2 (d_l^*)^{-1}\right)^{p/2} \left(\sum_{l=m}^{n} c_l^{-\frac{2\beta}{p}} \log(c_{l+1}/c_l)\right)^{p/2}. \end{aligned}$$

Since (c_k) is nondecreasing, for any $\gamma > 0$ we have, setting $u_k = \log c_k$,

$$\sum_{k=1}^{n-1} e^{\gamma u_k}(u_{k+1} - u_k) \le \int_{u_1}^{u_n} e^{\gamma x}\, dx \le \sum_{k=1}^{n-1} e^{\gamma u_{k+1}}(u_{k+1} - u_k). \tag{3.26}$$

By (3.11) and (3.12) we have $d_k^* = \log(c_{k+1}/c_k) \to 0$, i.e. $c_{k+1}/c_k \to 1$ and thus (3.26) yields

$$\sum_{k=1}^{n} c_k^{\gamma} \log(c_{k+1}/c_k) \sim \frac{1}{\gamma} c_n^{\gamma} \quad (\gamma > 0,\, n \to \infty). \tag{3.27}$$

A similar argument yields for $\gamma < 0$

$$\sum_{k=n}^{\infty} c_k^{\gamma} \log(c_{k+1}/c_k) \sim \frac{1}{|\gamma|} c_n^{\gamma} \quad (\gamma < 0,\, n \to \infty). \tag{3.28}$$

Hence (3.11) and (3.27) imply that

$$\sum_{k=1}^{l} d_k c_k^{\beta} \gg \frac{d_l}{d_l^* c_l^{\beta}} \sum_{k=1}^{l} d_k^* c_k^{2\beta} \gg \frac{d_l}{d_l^*} c_l^{\beta}.$$

Thus there is some $K_1 > 0$ with

$$\sum_{l=m}^{n} d_l^2 (d_l^*)^{-1} \le K_1 \sum_{l=m}^{n} d_l c_l^{-\beta} \left(\sum_{k=1}^{l} d_k c_k^{\beta} \right).$$

On the other hand, (3.28) implies that

$$\sum_{l=m}^{n} c_l^{-\frac{2\beta}{p}} \log(c_{l+1}/c_l) \le K_2 \cdot \frac{p}{2\beta} \cdot c_m^{-\frac{2\beta}{p}},$$

with $K_2 > 0$. This completes the proof of Lemma 3.1.

The crucial step of the proof of Theorem 3.1 is the following moment inequality: $\square$

Lemma 3.2. *Assume that $X_1, X_2, \ldots$ are independent r.v.'s and assume that (3.9)-(3.12) hold. Then for every $p \in \mathbb{N}$ we have*

$$\begin{aligned} E\left| \sum_{k=1}^{N} d_k \left\{ f(g_k(X_1, \ldots, X_k)) - Ef\left(g_k(X_1, \ldots, X_k)\right) \right\} \right|^p \\ \le C_p \left(\sum_{1 \le k \le l \le N} d_k d_l \left(\frac{c_k}{c_l} \right)^{\beta} \right)^{p/2}, \end{aligned} \tag{3.29}$$

where $C_p > 0$ is a constant.

Proof. Again we use notation (3.25) and set

$$V_{m,n} := \sum_{l=m}^{n} d_l c_l^{-\beta} \left(\sum_{k=1}^{l} d_k c_k^{\beta} \right) \qquad (1 \le m \le n).$$

Further we put

$$C_p = (4\gamma)^{p^2}. \tag{3.30}$$

We show that if the number γ is chosen large enough, then

$$E \left| \sum_{k=m}^{n} d_k \xi_k \right|^p \le C_p (V_{m,n})^{p/2} \quad \text{for all} \quad 1 \le m \le n. \tag{3.31}$$

Since $V_{1,N}$ equals the double sum on the right hand side of (3.29), this will prove Lemma 3.2. We use induction on p.

First observe that ξ_k and $\xi_{k,l}$ are independent for $1 \le k < l$. Therefore from $|\xi_k| \le 2$ and the Lipschitz 1 continuity of f we get

$$\begin{aligned} |E\xi_k\xi_l| &= |E\xi_k(\xi_l - \xi_{k,l})| \le 2E|\xi_l - \xi_{k,l}| \\ &\le K_3 E(|g_l(X_1, \ldots, X_l) - g_{k,l}(X_{k+1}, \ldots, X_l)| \wedge 1), \end{aligned}$$

for some $K_3 > 0$. Together with (3.9) we get

$$\begin{aligned} E\left(\sum_{k=m}^{n} d_k \xi_k \right)^2 &\le 2 \sum_{m \le k \le l \le n} d_k d_l |E\xi_k\xi_l| \\ &\le 2K_3 A \sum_{m \le k \le l \le n} d_k d_l \left(\frac{c_k}{c_l} \right)^{\beta} \le 2K_3 A V_{m,n}. \end{aligned}$$

Hence if we choose γ so large that $(4\gamma)^4 \ge 2K_3 A$, then (3.31) holds for $p = 2$.

Assume now that (3.31) is true for $p - 1 \ge 2$. From $d_k \gg d_k^* = \log(c_{k+1}/c_k)$ and (3.27) it follows that there is a positive constant K_4 such that

$$\sum_{k=1}^{l} d_k c_k^{\beta} \ge K_4 c_l^{\beta}.$$

Now choose γ so large that the C_p defined in (3.30) satisfies $C_p > (2/K_4)^p\gamma^{p/2}$. Then using $|\xi_l| \le 2$ we get for $V_{m,n} \le \gamma$

$$\left|\sum_{l=m}^{n} d_l\xi_l\right| \le (2/K_4)\sum_{l=m}^{n} d_l c_l^{-\beta}\left(\sum_{k=1}^{l} d_k c_k^{\beta}\right) = (2/K_4)V_{m,n} \le (2/K_4)\,\gamma^{1/2}V_{m,n}^{1/2}.$$

Hence in the case $V_{m,n} \le \gamma$ relation (3.31) is valid. We now show that if $X \ge \gamma$ is arbitrary and (3.31) holds for $V_{m,n} \le X$, then it will also hold for $V_{m,n} \le 3X/2$. As the validity of (3.31) is already verified for $V_{m,n} \le \gamma$, this will show that (3.31) holds for any value of $V_{m,n}$, and this will complete the induction step.

Assume $V_{m,n} \le 3X/2$ and set

$$S_1 + S_2 := \sum_{k=m}^{w} d_k\xi_k + \sum_{k=w+1}^{n} d_k\xi_k \qquad (m \le w \le n).$$

Put further

$$T_2 := \sum_{k=w+1}^{n} d_k\xi_{w,k}.$$

For a fixed m and n we choose w in such a way that

$$V_{m,w} \le X, \qquad V_{w+1,n} \le X \qquad \text{and} \qquad \frac{V_{w+1,n}}{V_{m,w}} = \lambda \in [1/2, 1].$$

To see that this is possible, we have to show that for every choice of $1 \le m < n$ with $V_{m,n} \ge X$ there is some $w \in \{m+1, \ldots, n\}$ with

$$\frac{1}{2}V_{m,n} \le V_{m,w} \le \frac{2}{3}V_{m,n}. \tag{3.32}$$

We define $w := \min\{k \ge m : (1/2)V_{m,n} \le V_{m,k}\}$. Then (3.32) will follow if we show that the increment $V_{m,w} - V_{m,w-1} \le 1/6V_{m,n}$. But from (3.10) we get

$$\begin{aligned} V_{m,w} - V_{m,w-1} &= d_w c_w^{-\beta}\left(\sum_{k=1}^{w} d_k c_k^{\beta}\right) \ll d_w c_w^{-\beta}\frac{D_w}{(\log D_w)^{\rho}}\sum_{k=1}^{w} d_k^* c_k^{\beta} \\ &\ll d_w D_w/(\log D_w)^{\rho}, \end{aligned}$$

where we used again (3.27) in the last step. Since by assumption (3.12) the last term is bounded, and since $V_{m,n} \ge X \ge \gamma$ the result follows for sufficiently large γ.

Now we prove that

$$E|S_1 + S_2|^p \le C_p (V_{m,n})^{p/2}.$$

To do so, we need some simple inequalities.

From the mean value theorem we get

$$|S_2^j - T_2^j| \le j\,|S_2 - T_2|(|S_2|^{j-1} + |T_2|^{j-1}) \qquad (j \ge 1). \tag{3.33}$$

Using Lemma 3.1 and the assumption that (c_k) is nondecreasing we get for all $j \ge 1$

$$E|S_2 - T_2|^j \le E_j (V_{w+1,n})^{j/2}.$$

We also have

$$E|S_1|^j \le C_j (V_{m,w})^{j/2} \qquad (1 \le j \le p) \tag{3.34}$$

and

$$E|S_2|^j \le C_j (V_{w+1,n})^{j/2} \le C_j \lambda^{j/2} (V_{m,w})^{j/2} \qquad (1 \le j \le p). \tag{3.35}$$

For $1 \le j \le p-1$ the last two inequalities are valid by the induction hypothesis, and for $j = p$ they follow from the validity of (3.31) for $V_{m,n} \le X$. Hence Minkowski's inequality yields

$$E|T_2|^j \le 2^j C_j \lambda^{j/2} (V_{m,w})^{j/2} \qquad (1 \le j \le p). \tag{3.36}$$

Finally combining the Hölder inequality with the latter results shows for $1 \le j \le p-1$

$$\begin{aligned} E|S_1|^j |S_2 - T_2| |S_2|^{p-j-1} &\le (E|S_1|^p)^{j/p} \, (E|S_2 - T_2|^p)^{1/p} \, (E|S_2|^p)^{(p-j-1)/p} \\ &\le C_p^{(p-1)/p} E_p^{1/p} \lambda^{(p-j)/2} (V_{m,w})^{p/2}. \end{aligned} \tag{3.37}$$

The last inequality remains valid, with an extra factor 2^{p-j-1} on the right hand side, if $|S_2|^{p-j-1}$ on the left hand side is replaced by $|T_2|^{p-j-1}$. Since S_1 and T_2 are independent, we get by the binomial formula and the triangle inequality

$$\begin{aligned} E\,|S_1 + S_2|^p &\le E|S_1|^p + E|S_2|^p \\ &+ \sum_{j=1}^{p-1} \binom{p}{j} \left(E|S_1|^j |S_2^{p-j} - T_2^{p-j}| + E|S_1|^j \, E|T_2|^{p-j} \right). \end{aligned}$$

We substitute (3.33) and (3.34)–(3.37) (using also the analogue of (3.37) with $|T_2|^{p-j-1}$) in the above inequality and get

$$E|S_1+S_2|^p \le C_p(V_{m,w})^{p/2}\Bigg(1+\lambda^{p/2} + C_p^{-1/p}\,E_p^{1/p}\sum_{j=1}^{p-1}2^{p-j}\binom{p}{j}(p-j)\lambda^{(p-j)/2} + C_p^{-1}\sum_{j=1}^{p-1}2^{p-j}\lambda^{(p-j)/2}\binom{p}{j}C_j\,C_{p-j}\Bigg).$$

Now

$$C_p^{-1/p}E_p^{1/p} \le K\cdot p^{1/2}(4\gamma)^{-p}, \qquad C_jC_{p-j}/C_p \le (4\gamma)^{-p} \quad (1\le j\le p-1)$$

and thus by $\lambda \le 1$

$$C_p^{-1/p}\,E_p^{1/p}\sum_{j=1}^{p-1}2^{p-j}\binom{p}{j}(p-j)\lambda^{(p-j)/2} \le \text{const}\cdot p^{3/2}\gamma^{-p}$$

and

$$C_p^{-1}\sum_{j=1}^{p-1}2^{p-j}\lambda^{(p-j)/2}\binom{p}{j}C_j\,C_{p-j} \le \text{const}\cdot\gamma^{-p}.$$

Since $\lambda \ge 1/2$ we see that for a large enough γ the relation $E|S_1+S_2|^p \le C_p\,(1+\lambda)^{p/2}\,(V_{m,w})^{p/2} = C_p(V_{m,n})^{p/2}$ is true. Thus we proved the validity of (3.31) for $V_{m,n} \le 3X/2$ and the proof of Lemma 3.2 is completed. □

The following lemma estimates the double sum appearing on the right hand side of (3.29).

Lemma 3.3. *Assume that (3.10) holds, then for any $\beta>0$ and any $\eta<\rho$ we have*

$$\sum_{1\le k\le l\le N} d_kd_l\left(\frac{c_k}{c_l}\right)^{\beta} = O\left(\frac{D_N^2}{(\log D_N)^{\eta}}\right).$$

Proof. By the monotonicity of (c_k) we have

$$\sum_{1\le k\le l\le N} d_kd_l\left(\frac{c_k}{c_l}\right)^{\beta} \le \sum_{\substack{1\le l\le N\\ c_k\le c_l/(\log D_N)^{\rho/\beta}}} d_kd_l\left(\frac{c_k}{c_l}\right)^{\beta} + \sum_{\substack{1\le l\le N\\ c_l/(\log D_N)^{\rho/\beta}<c_k\le c_l}} d_kd_l =: \sigma_N+\tau_N.$$

Clearly $\sigma_N \le D_N^2 (\log D_N)^{-\rho}$ and by (3.10)

$$\tau_N \ll \sum_{1\le l\le N} d_l \sum_{c_l/(\log D_N)^{\rho/\beta} < c_k \le c_l} d_k^* \frac{D_k}{(\log D_k)^\rho}$$
$$\ll \frac{D_N}{(\log D_N)^\rho} \sum_{1\le l\le N} d_l \sum_{c_l/(\log D_N)^{\rho/\beta} < c_k \le c_l} \log(c_{k+1}/c_k) \ll \frac{D_N^2}{(\log D_N)^\rho} \log\log D_N.$$

□

Proof of Theorem 3.1. We use notation (3.25). From Lemmas 3.2–3.3 and the Markov inequality we derive for any $\varepsilon > 0$, $p \in \mathbb{N}$,

$$P\left(\left| \sum_{k=1}^N d_k \xi_k \right| > \varepsilon D_N \right) \le c(p,\varepsilon)(\log D_N)^{-p\eta/2} \quad \text{for} \quad N \ge N_0.$$

By (3.12) we get $d_k \to 0$ and consequently $D_{N+1}/D_N \to 1$. Thus we can choose an increasing sequence (N_j) of positive integers such that $D_{N_j} \sim \exp(\sqrt{j})$. Hence choosing $p > 4/\eta$ and using the Borel-Cantelli lemma we get

$$\lim_{j\to\infty} \frac{1}{D_{N_j}} \sum_{k=1}^{N_j} d_k \xi_k = 0 \quad \text{a.s.}$$

For $N_j \le N < N_{j+1}$ we have by $|\xi_k| \le 2$

$$\frac{1}{D_N} \left| \sum_{k=1}^N d_k \xi_k \right| \le \frac{1}{D_{N_j}} \left| \sum_{k=1}^{N_j} d_k \xi_k \right| + 2 \left(\frac{D_{N_{j+1}}}{D_{N_j}} - 1 \right).$$

Since $D_{N_{j+1}}/D_{N_j} \to 1$, the convergence of the subsequence implies that the whole sequence converges a.s.

We have proved Theorem 3.1 for all bounded Lipschitz functions f. The result for indicator functions follows by routine approximation arguments, similar e.g. to those in [69]. □

Chapter 4

Generalized moments in a.s. central limit theory

4.1 Introduction and results

Let $X_1, X_2, \ldots$ be i.i.d. random variables with $EX_1 = 0$, $EX_1^2 = 1$. By a slightly more general version of the ASCLT (implicit in Lacey and Philipp [69]) we have, letting $\phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$,

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{k=1}^{N} \frac{1}{k} f\left(\frac{S_k}{\sqrt{k}}\right) = \int_{-\infty}^{\infty} f(x)\phi(x)\,dx \qquad \text{a.s.} \tag{4.1}$$

for every bounded and almost everywhere continuous function f. The assumption that f is a.e. continuous cannot be replaced solely by the measurability of f, as an example of Schatte [96] shows. The validity of (4.1) for unbounded f is of major interest in the theory: for $f(x) = x^p$ this expresses convergence of moments in the ASCLT and for this reason, the left hand side of (4.1) is called a *generalized moment*. Clearly, a necessary condition for (4.1) is

$$\int_{\mathbb{R}} |f(x)\phi(x)|\,dx < \infty. \tag{4.2}$$

As Ibragimov and Lifshits [63, Example 3] pointed out, condition (4.2) is not sufficient: there is a continuous function f satisfying (4.2) such that the ASCLT fails even for a

Bernoulli sequence (X_n). The reason for this failure lies in the very irregular shape of the specific f. Ibragimov and Lifshits [63] also showed, improving earlier results of Schatte [97] and Berkes, Csáki and Horváth [10], that relation (4.1) holds if (4.2) is valid and f satisfies minor regularity conditions, e.g. if $f(|x|)e^{-cx^2}$ is nonincreasing for some $c > 0$. This result lies much deeper than the ordinary ASCLT; its proof depends on delicate fluctuation properties of i.i.d. random variables. The argument fails for general independent sequences (X_n) and generalized moment behavior in a.s. central limit theory remains open. The purpose of the present chapter is to give an essentially complete solution of this problem. Our results will show the surprising fact that for an independent sequence (X_n), the validity of a relation of type (4.1) for a sufficiently large class of f's is closely related to the LIL behavior of X_n, revealing a new, unexpected side of ASCLT theory.

Before we present our results, we formulate the following theorem of Atlagh [2] and Ibragimov and Lifshits [64] expressing the ASCLT for independent, not identically distributed random variables.

Theorem C. *Let $X_1, X_2, \dots$ be independent random variables with zero mean and finite variances. Set $S_n = X_1 + \dots + X_n$,$s_n^2 = ES_n^2$ and $\sigma_n^2 = EX_n^2$. If the Lindeberg condition (1.6) holds then*

$$\lim_{n\to\infty} \frac{1}{\log s_n^2} \sum_{k=1}^{N} \frac{\sigma_{k+1}^2}{s_k^2} I\left\{\frac{S_k}{s_k} \le z\right\} = \Phi(z) \qquad \text{a.s.}$$

As in the previous sections, we write $\log x$ for $\log(x \vee e)$ and $\log\log x$ for $\log(\log x \vee e)$. In the sequel S_n will denote the partial sums of a sequence (X_n) and $s_n^2 = \operatorname{Var} S_n$.

Theorem 4.1. *Let $X_1, X_2, \dots$ be independent random variables with zero mean and finite variances. Suppose that $s_n \to \infty$ and that (X_n) satisfies Kolmogorov's condition for the LIL, i.e.*

$$|X_n| \le \varepsilon_n s_n / (\log\log s_n)^{1/2} \tag{4.3}$$

with a positive numerical sequence $\varepsilon_n \to 0$. Then for every almost everywhere continuous function f satisfying

$$|f(x)| = O(\exp(\gamma x^2)) \quad \textit{for some } \gamma < 1/2 \quad (|x| \to \infty), \tag{4.4}$$

we have

$$\lim_{n\to\infty} \frac{1}{\log s_n^2} \sum_{k=1}^{n} \frac{\sigma_{k+1}^2}{s_k^2} f\left(\frac{S_k}{s_k}\right) = \int_{\mathbb{R}} f(x)\phi(x)\,dx. \tag{4.5}$$

Theorem 4.1 is sharp as our next result shows.

Theorem 4.2. *For every $\varepsilon > 0$ there exists a sequence (X_n) of symmetric, bounded random variables such that we have $s_n \to \infty$,*

$$|X_n| \le \varepsilon s_n / (\log\log s_n)^{1/2} \tag{4.6}$$

and for $f(x) = \exp(\gamma x^2)$ with some $0 < \gamma < 1/2$ we have

$$\limsup_{n\to\infty} \frac{1}{\log s_n^2} \sum_{k=1}^{n} \frac{\sigma_{k+1}^2}{s_k^2} f\left(\frac{S_k}{s_k}\right) = +\infty \quad \text{a.s.}$$

If we strengthen Kolmogorov's condition (4.3), the result of Theorem 4.1 will hold for a larger class of functions f.

Theorem 4.3. *Let $X_1, X_2, \ldots$ be independent random variables with zero mean and finite variances and assume that $s_n \to \infty$. Further assume that*

$$|X_k| \le \varepsilon_k s_k / (\log\log s_k)^{3/2}$$

with a positive numerical sequence $\varepsilon_k \to 0$. If $f(x)$ is an almost everywhere continuous function on $\mathbb{R}$ such that $|f(x)| \le e^{x^2/2} b(x)$ with some positive function b such that $\log b(x)$ is uniformly continuous on $\mathbb{R}$ and $\int_{\mathbb{R}} b(x)\,dx < \infty$, then (4.5) holds.

The assumption that $\log b(x)$ is uniformly continuous is not necessary and can be replaced by other conditions. In [10] it is shown that if $b(x) > 0$ for all $x \in \mathbb{R}$ and if $b(x)$ and $b(-x)$ are non-increasing for $x > 0$ and $\int_R b(x)\,dx < \infty$, then there is a function b^* on $\mathbb{R}$ such that $b(x) \le b^*(x)$ for all x, $\log b^*$ is uniformly continuous on $\mathbb{R}$ and $\int_R b^*(x)\,dx < \infty$.

The following theorems will give sharp conditions for the validity of (4.5) for unbounded random variables X_n.

Theorem 4.4. *Let $X_1, X_2, \dots$ be independent random variables with zero mean and finite variances. Set $\sigma_n^2 = \operatorname{Var} X_n$ and assume that*

$$s_n \to \infty \quad \textit{and} \quad \sigma_{n+1} = o\left(s_n/(\log\log s_n)^{1/2}\right). \tag{4.7}$$

Let $S(t) = S_n$ if $s_n^2 \le t < s_{n+1}^2$ $n = 0, 1, \dots$ and assume that there exists a Wiener process $\{W(t),\, t \ge 0\}$ on the same probability space such that

$$S(t) = W(t) + o(t^{1/2}) \qquad \text{a.s.} \quad (t \to \infty). \tag{4.8}$$

If f is an a.e. *continuous function satisfying (4.4), then relation (4.5) holds.*

Again, this result is sharp in the sense that condition (4.8) cannot be weakened to

$$S(t) = W(t) + o(t^{1/2}\psi(t)) \quad \text{a.s.}$$

for any function $\psi(t) \nearrow \infty$.

Theorem 4.5. *Let $X_1, X_2, \dots$ be independent random variables with zero mean and finite variances. Set $S(t) = S_n$ if $s_n^2 \le t < s_{n+1}^2$ $(n \ge 1)$. The* a.s. *approximation*

$$S(t) = W(t) + o\left(t^{1/2}\psi(t)\right) \quad \text{a.s.} \tag{4.9}$$

with some $\psi(t) \nearrow \infty$ and the additional regularity assumptions (4.7) and the central limit theorem

$$P(S_n/s_n \le x) \to \Phi(x) \tag{4.10}$$

do not imply Theorem 4.4 even for polynomial f.

By a theorem of Major [78], Kolmogorov's condition (4.3) implies that we can redefine $S(t)$ on a new probability space together with a Wiener process $\{W(t),\, t \ge 0\}$ such that

$$S(t) = W(t) + o(t \log\log t)^{1/2} \quad \text{a.s.} \tag{4.11}$$

Since the last condition is still sufficient for the validity of the LIL for (X_n), it is natural to ask if Kolmogorov's condition (4.3) in Theorem 4.1 can be replaced by (4.11). As Theorem 4.5 above shows, the answer is negative.

Strengthening the remainder term in (4.8) leads to a larger class of functions in (4.5).

Theorem 4.6. *Let $X_1, X_2, \ldots$ be independent random variables with zero mean and finite variances. Define $S(t) = S_n$ if $s_n^2 \le t < s_{n+1}^2$, $n = 1, 2, \ldots$. Assume that (4.7) holds and that there exists a standard Wiener process $\{W(t), t \ge 0\}$ on the same probability space such that*

$$S(t) = W(t) + o(t/\log\log t)^{1/2} \quad \text{a.s.} \quad (t \to \infty). \tag{4.12}$$

Then if $f(x)$ is an almost everywhere continuous function on $\mathbb{R}$ satisfying the conditions of Theorem 4.3, relation (4.5) holds.

We note that conditions (4.8) and (4.12) imply the ordinary CLT for the partial sums $S_n = X_1 + \cdots + X_n$. In our next theorem we assume that the CLT is valid. As a further assumption we require an exponential bound for large deviations of S_n, which is e.g. implied by Kolmogorov's LIL condition (4.3).

Theorem 4.7. *Let $X_1, X_2, \ldots$ be independent random variables with zero mean and finite variances. Set $s_n^2 = \operatorname{Var} S_n$ and assume that*

$$s_n \to \infty \quad \textit{and} \quad \sigma_n = o(s_n). \tag{4.13}$$

Assume further that (4.10) is valid and that

$$P(S_n \ge x) \le \exp\left(-\frac{x^2}{2s_n^2}(1-\epsilon_n)\right) \quad \textit{for} \quad 0 \le x \le K(s_n^2 \log\log s_n^2)^{1/2}, \tag{4.14}$$

where $\epsilon_n = o(1)$ and $K > 0$. Then (4.5) holds for any almost everywhere continuous function $f : \mathbb{R} \to \mathbb{R}$ satisfying (4.4).

4.2 Proofs

Lemma 4.1. *Assume (4.5) is valid for indicator functions of intervals and for a fixed a.e. continuous function f_0. Then (4.5) is also true for all a.e. continuous functions f with $|f(x)| \le |f_0(x)|$, $x \in \mathbb{R}$.*

Proof. We refer to Schatte [97, Section 2.3] where a similar result is proved. $\square$

Lemma 4.2. *Let $f(x)$ be a real valued and integrable function with $\int_R |f(x)|\, d\Phi(x) < \infty$. Then*

$$\lim_{N\to\infty} \frac{1}{\log N} \int_1^N \frac{1}{t} f\left(\frac{W(t)}{\sqrt{t}}\right) dt = \int_R f(x)\, d\Phi(x) \quad \text{a.s.}$$

Proof. This follows from the ergodic theorem by using the substitution $t = e^s$:

$$\frac{1}{\log N} \int_1^N \frac{1}{t} f\left(\frac{W(t)}{\sqrt{t}}\right) dt = \frac{1}{\log N} \int_0^{\log N} f\left(U(s)\right) ds \to Ef(U(0)) \quad \text{a.s.},$$

where $U(s) = W(e^s)\, e^{-s/2}$ is the Ornstein-Uhlenbeck process. $\square$

Lemma 4.3. *Let $f(x) = e^{x^2/2} b(x)$ with some positive function b such that $\log b(x)$ is uniformly continuous on $\mathbb{R}$ and $\int_{\mathbb{R}} b(x)\, dx < \infty$. Under the conditions of Theorem 4.6 we have*

$$\int_{s_k^2}^{s_{k+1}^2} \frac{1}{t} f\left(\frac{W(t)}{\sqrt{t}}\right) dt \sim \frac{\sigma_{k+1}^2}{s_k^2} f\left(\frac{S_k}{s_k}\right) \quad \text{a.s.} \quad (k \to \infty).$$

Proof. We show that almost surely

$$\int_{s_k^2}^{s_{k+1}^2} \frac{1}{t} f\left(\frac{W(t)}{\sqrt{t}}\right) dt \sim \int_{s_k^2}^{s_{k+1}^2} \frac{1}{t} f\left(\frac{S(t)}{\sqrt{t}}\right) dt \sim \int_{s_k^2}^{s_{k+1}^2} \frac{1}{s_k^2} f\left(\frac{S_k}{s_k}\right) dt. \tag{4.15}$$

From (4.12) we conclude that there is a function $\epsilon(t) = \epsilon(t, \omega) \geq 0$ with $\epsilon(t) \to 0$ a.s. for $t \to \infty$ such that

$$\left| \frac{S(t)}{\sqrt{t}} - \frac{W(t)}{\sqrt{t}} \right| = \epsilon(t)(\log\log t)^{-1/2}\,, \tag{4.16}$$

and hence

$$\left| \frac{S^2(t)}{t} - \frac{W^2(t)}{t} \right| \leq \epsilon(t)(\log\log t)^{-1/2} \left(\left| \frac{S(t)}{\sqrt{t}} \right| + \left| \frac{W(t)}{\sqrt{t}} \right| \right). \tag{4.17}$$

Note that (4.12) implies the LIL for $S(t)$, and thus the right hand side of (4.17) tends to zero almost surely as $t \to \infty$. Since we have

$$\frac{f\left(\frac{W(t)}{\sqrt{t}}\right)}{f\left(\frac{S(t)}{\sqrt{t}}\right)} = \exp\left(\log b\left(\frac{W(t)}{\sqrt{t}}\right) - \log b\left(\frac{S(t)}{\sqrt{t}}\right) \right) \exp\left(\frac{W^2(t)}{2t} - \frac{S^2(t)}{2t} \right),$$

we conclude from the uniform continuity of $\log b$ that

$$\lim_{k\to\infty} \sup_{t\in[s_k^2, s_{k+1}^2]} \frac{f\left(\frac{W(t)}{\sqrt{t}}\right)}{f\left(\frac{S(t)}{\sqrt{t}}\right)} = 1 \quad \text{a.s.} \tag{4.18}$$

Relation (4.18) remains valid if we replace the supremum by an infimum. This proves the first part in (4.15).

In order to obtain the second part of (4.15) we first note that by (4.7) we have $s_{k+1}^2 \sim s_k^2$. Hence it suffices to show relation (4.18) with $W(t)/\sqrt{t}$ replaced by S_k/s_k. Now the reasoning in the proof is similar to the previous arguments. □

Proof of Theorem 4.6. From (4.12) we get the central limit theorem (4.10) for the process $X_1, X_2, \ldots$, and thus Theorem C implies that the ASCLT (4.5) holds for indicator functions f. Thus using Lemma 4.1 it suffices to prove Theorem 4.6 for a function $f(x) = b(x)e^{x^2/2}$, where b satisfies the conditions in Theorem 4.6. From Lemma 4.2 and Lemma 4.3 we get a.s. for $N \to \infty$

$$\log s_N^2 \int_R f(x)\, d\Phi(x) \sim \sum_{k=1}^{N} \int_{s_k^2}^{s_{k+1}^2} \frac{1}{t} f\left(\frac{W(t)}{\sqrt{t}}\right) dt \sim \sum_{k=1}^{N} \frac{\sigma_{k+1}^2}{s_k^2} f\left(\frac{S_k}{s_k}\right),$$

which completes the proof. □

Proof Theorem 4.4. The strong approximation (4.8) yields an almost surly finite random variable $t_0(\omega, \epsilon)$ such that for $t \geq t_0$

$$\left|\frac{S(t)}{\sqrt{t}}\right| \leq \left|\frac{W(t)}{\sqrt{t}}\right| + \epsilon.$$

Let $f(x) = \exp(\gamma x^2)$, $\gamma < 1/2$. Obviously

$$\begin{aligned} &\frac{1}{\log N} \int_1^N \frac{1}{t} f\left(\frac{S(t)}{\sqrt{t}}\right) dt \\ &\quad \leq \frac{1}{\log N} \int_1^{t_0} \frac{1}{t} f\left(\frac{S(t)}{\sqrt{t}}\right) dt + \frac{1}{\log N} \int_1^N \frac{1}{t} f\left(\left|\frac{W(t)}{\sqrt{t}}\right| + \epsilon\right) dt. \end{aligned}$$

From Lemma (4.2) we conclude that

$$\limsup_N \frac{1}{\log N} \int_1^N \frac{1}{t} f\left(\frac{S(t)}{\sqrt{t}}\right) dt \leq \int_{\mathbb{R}} \exp(\gamma(|x| + \epsilon)^2)\, d\Phi(x) \quad \text{a.s.}$$

If $\epsilon \to 0$, the integral on the right hand side tends to $\int_{\mathbb{R}} \exp(\gamma x^2)\, d\Phi(x)$. A similar computation in the other direction gives

$$\frac{1}{\log N} \int_1^N \frac{1}{t} f\left(\frac{S(t)}{\sqrt{t}}\right) \to \int_{\mathbb{R}} \exp(\gamma x^2)\, d\Phi(x) \quad \text{a.s.}$$

It is easy to show that

$$(1-\delta_k)\frac{\sigma_{k+1}^2}{s_k^2} f\left(\frac{S_k}{s_{k+1}}\right) \le \int_{s_k^2}^{s_{k+1}^2} \frac{1}{t} f\left(\frac{S_k}{\sqrt{t}}\right) \le (1+\delta_k)\frac{\sigma_{k+1}^2}{s_k^2} f\left(\frac{S_k}{s_k}\right),$$

where $\delta_k = o(1)$. Using the explicit form of f the proof follows by simple calculations from (4.7) and the LIL. $\square$

Proof of Theorem 4.7. In the sequel we will put

$$\xi_k := \frac{S_k}{s_k}.$$

For some $\delta > 0$ we set $d_k = ((2+\delta)\log\log s_k^2)^{1/2}$ and define

$$\hat{\xi}_k := \operatorname{sign}(\xi_k)\{|\xi_k| \wedge d_k\}.$$

By the usual argument in this theory it suffices to prove the proposition of Theorem 4.7 for a fixed function $f(x) := C\exp(\gamma x^2)$, where $C > 0$ and $0 < \gamma < 1/2$. We put

$$I_n := \frac{1}{\log s_n^2} \sum_{k=1}^n \frac{\sigma_{k+1}^2}{s_k^2} f(\xi_k).$$

In an obvious way we define $\hat{I}_n$. Now we have

$$\hat{I}_n \le I_n \le \hat{I}_n + \frac{1}{\log s_n^2} \sum_{k=1}^n \frac{\sigma_{k+1}^2}{s_k^2} f(\xi_k) I\{|\xi_k| > d_k\}. \tag{4.19}$$

We first show that the event $\{|\xi_k| > d_k \quad \text{i.o.}\}$ has probability zero and thus the last term in (4.19) tends to zero a.s. For this purpose we define $\chi(n) = (2s_n^2\log\log s_n^2)^{1/2}$. By (4.14) we have

$$P(S_n \ge (1+\epsilon)^{1/2}\chi(n)) \le (\log s_n^2)^{-(1+\epsilon')},$$

for $0 < \epsilon' < \epsilon$ if n is large enough. It is easy to show that the assumption $\sigma_n^2/s_n^2 \to 0$ implies that for every $\tau > 0$ there is a subsequence (n_l) such that $s_{n_{l+1}}^2 \geq (1+\tau)^l > s_{n_l}^2$ and $s_{n_l}^2 \sim (1+\tau)^l$, $l = 1, 2, \cdots$. Hence

$$P(S_{n_l} \geq (1+\epsilon)^{1/2}\chi(n_l)) \leq 2(l\log(1+\tau))^{-(1+\epsilon')}.$$

By a well known maximal inequality of Kolmogorov we get for $0 < \eta' < \eta < \epsilon$ and l large enough

$$\begin{aligned} P(\max_{n \leq n_l} S_n \geq (1+\epsilon)^{1/2}\,\chi(n_l)) &\leq 2P(S_{n_l} \geq (1+\epsilon)^{1/2}\chi(n_l) - \sqrt{2s_n^2}) \\ &\leq 2P(S_{n_l} \geq (1+\eta)^{1/2}\chi(n_l)) \\ &\leq 4(l\log(1+\tau))^{-(1+\eta')}. \end{aligned}$$

Thus, by the Borel-Cantelli-lemma we have for all $l \geq l_0(\omega)$

$$\max_{n \leq n_l} S_n < (1+\epsilon)^{1/2}\chi(n_l),$$

where $l_0 < \infty$ a.s. For large enough l and a small τ we can ascertain that

$$\frac{\chi(n_{l+1})}{\chi(n_l)} \sim (1+\tau)^{1/2} < (1+\epsilon)^{1/2}.$$

Therefore we get for $n_l \leq m \leq n_{l+1}$

$$S_m \leq \max_{n \leq n_{l+1}} S_n < (1+\epsilon)^{1/2}\chi(n_{l+1}) \leq (1+\epsilon)\chi(n_l) \leq (1+\epsilon)\chi(m).$$

It suffices thus to show that

$$\lim_n \hat{I}_n = \int_{\mathbb{R}} f(x)\,d\Phi(x) \quad \text{a.s.}$$

First we prove the integrated version

$$\lim_n E\hat{I}_n = \int_{\mathbb{R}} f(x)\,d\Phi(x). \tag{4.20}$$

Observe that for some $0 < a \leq f(d_k)$

$$\int_{\{f(\hat{\xi}_k) > a\}} f(\hat{\xi}_k)\,dP = \int_{\{f(d_k) > f(\xi_k) > a\}} f(\xi_k)\,dP + f(d_k)P(|\xi_k| \geq d_k).$$

Clearly the integral above is zero if $a > f(d_k)$. By (4.14) we have for small enough δ

$$\begin{aligned} f(d_k)P(|\xi_k| \geq d_k) &\leq 2f(d_k)\exp(-d_k^2(1-\epsilon_k)/2) \\ &= 2(\log s_k^2)^{(2+\delta)(\gamma-1/2\,(1-\epsilon_k))} = o(1) \quad (k \to \infty). \end{aligned}$$

Further we have

$$\int_{\{f(d_k)>f(\xi_k)>a\}} f(\xi_k)\,dP = \left(\int_{(1/\gamma\log a)^{1/2}}^{d_k} + \int_{-d_k}^{-(1/\gamma\log a)^{1/2}}\right) e^{\gamma x^2}\,d[P(\xi_k \leq x) - 1].$$

Product integration yields

$$\begin{aligned} &\int_{(1/\gamma\log a)^{1/2}}^{d_k} e^{\gamma x^2}\,d[P(\xi_k \leq x) - 1] \\ &\qquad \leq aP(\xi_k \geq (1/\gamma\log a)^{1/2}) + 2\gamma\int_{(1/\gamma\log a)^{1/2}}^{d_k} xe^{\gamma x^2}P(\xi_k \geq x)\,dx. \end{aligned}$$

Now we can use again the tail estimate (4.14) and easy computations show that for some constants $c, \mu > 0$ we have for all $k \geq 1$

$$\int_{\{f(d_k)>f(\xi_k)>a\}} f(\xi_k)\,dP \;\leq\; ca^{-\mu}.$$

From the previous calculations we conclude that

$$\lim_{a\to\infty}\sup_k \int_{\{f(\hat{\xi}_k)>a\}} f(\hat{\xi}_k)\,dP = 0,$$

hence the sequence $(f(\hat{\xi}_k))$ is uniformly integrable. By (4.10) $\xi_k \xrightarrow{d} N$ and hence $\hat{\xi}_k \xrightarrow{d} N$, where N is a standard normal random variable. By the uniform integrability it follows that $Ef(\hat{\xi}_k) \to Ef(N)$ and this proves (4.20).

In the next step we estimate the variance of $\hat{I}_n$. For $1 \leq k \leq l$ we define

$$\eta_{k,\,l} := \frac{S_l - S_k}{s_l},$$

and

$$\hat{\eta}_{k,\,l} := \mathrm{sign}(\eta_{k,\,l})\{|\eta_{k,\,l}| \wedge d_l\} \quad \text{where} \quad d_l = ((2+\delta)\log\log s_l^2)^{1/2},$$

for some $\delta > 0$. It is clear that

$$|\hat{\xi}_l - \hat{\eta}_{k,\,l}| \leq |\,\xi_l - \eta_{k,\,l}| \leq |\xi_k|\,\frac{s_k}{s_l}.$$

Also it is obvious that $f(x)$ decreases as $|x|$ grows. Consequently by the independence of $\hat{\xi}_k$ and $\hat{\eta}_{k,l}$

$$\begin{aligned}
&Ef(\hat{\xi}_k)f(\hat{\xi}_l)I\{|\xi_k| \le (s_l/s_k)^{1/3}\} \le Ef(\hat{\xi}_k)f(|\hat{\eta}_{k,l}| + |\xi_k|s_k/s_l)I\{|\xi_k| \le (s_l/s_k)^{1/3}\} \\
&\quad \le Ef(\hat{\xi}_k)\, Ef(|\hat{\eta}_{k,l}| + (s_k/s_l)^{2/3}) \\
&\quad \le Ef(\hat{\xi}_k)\, Ef(|\hat{\eta}_{k,l}| + (s_k/s_l)^{2/3})I\{|\xi_k| \le (s_l/s_k)^{1/3}\} + f(d_k)f(d_l+1)\,(s_k/s_l)^{2/3} \\
&\quad \le Ef(\hat{\xi}_k)\, Ef(|\hat{\xi}_l| + 2(s_k/s_l)^{2/3}) + f(d_k)f(d_l+1)\,(s_k/s_l)^{2/3}
\end{aligned}$$

From the convexity of $f(x)$ we deduce

$$\begin{aligned}
&f(|\hat{\xi}_l| + 2(s_k/s_l)^{2/3}) - f(|\hat{\xi}_l|) \\
&\quad \le f(d_l + 2(s_k/s_l)^{2/3}) - f(d_l) \le 4\gamma(s_k/s_l)^{2/3}(d_l+2)\exp(\gamma(d_l+2)^2).
\end{aligned}$$

Some algebra shows that for sufficiently small δ there are constants c_1, $\nu_1 > 0$ such that

$$Ef(\hat{\xi}_k)f(\hat{\xi}_l)I\{|\xi_k| \le (s_l/s_k)^{1/3}\} \le Ef(\hat{\xi}_k)\, Ef(\hat{\xi}_l) + c_1\,(s_k/s_l)^{2/3}\,(\log s_l^2)^{2-\nu_1}. \tag{4.21}$$

Applying Tschebyschev's inequality gives

$$Ef(\hat{\xi}_k)f(\hat{\xi}_l)I\{|\xi_k| > (s_l/s_k)^{1/3}\} \quad \le \quad c_2\,(s_k/s_l)^{2/3}\,(\log s_l^2)^{2-\nu_2} \tag{4.22}$$

for some constants c_2, $\nu_2 > 0$. Combining (4.21) and (4.22) we have shown that

$$\operatorname{Cov}(f(\hat{\xi}_k), f(\hat{\xi}_l)) \le c_3(s_k/s_l)^{2/3}(\log s_l^2)^{2-\nu_3} \quad (c_3, \nu_3 > 0). \tag{4.23}$$

Hence

$$\begin{aligned}
\operatorname{Var}\hat{I}_n \quad &\le \quad \frac{2}{(\log s_n^2)^2} \sum_{\substack{1\le k\le l\le n \\ (s_k/s_l)\le(\log s_n^2)^{-3}}} \frac{\sigma_{k+1}^2}{s_k^2}\,\frac{\sigma_{l+1}^2}{s_l^2}\, c_3\,(s_k/s_l)^{2/3}\,(\log s_l^2)^{2-\nu_3} \\
&\quad + \frac{2}{(\log s_n^2)^2} \sum_{\substack{1\le k\le l\le n \\ (s_k/s_l)>(\log s_n^2)^{-3}}} \frac{\sigma_{k+1}^2}{s_k^2}\,\frac{\sigma_{l+1}^2}{s_l^2}\,(Ef^2(\hat{\xi}_k))^{1/2}(Ef^2(\hat{\xi}_l))^{1/2} \\
&=: \quad V_1 + V_2.
\end{aligned}$$

Using (4.13) we get $\sigma_{k+1}^2/s_k^2 \sim \log(1+\sigma_{k+1}^2/s_k^2) = \log s_{k+1}^2 - \log s_k^2$ and thus

$$V_1 \ll \frac{1}{(\log s_n^2)^{\nu_3+2}} \sum_{l=1}^{n} \frac{\sigma_{l+1}^2}{s_l^2} \sum_{k=1}^{l} \frac{\sigma_{k+1}^2}{s_k^2} \ll (\log s_n^2)^{-\nu_3} ,$$

where $a_n \ll b_n$ means that $\limsup_n |a_n/b_n| < \infty$. Using similar ideas as before, we see that if δ is sufficiently small then $Ef^2(\hat{\xi}_l) \ll (\log s_l^2)^{1-\nu_4}$ for some $\nu_4 > 0$. Hence

$$\begin{aligned} V_2 &\ll (\log s_n^2)^{-1-\nu_4} \sum_{1\le l\le n} \frac{\sigma_{l+1}^2}{s_l^2} \sum_{s_l(\log s_n^2)^{-3}<s_k\le s_l} (\log s_{k+1}^2 - \log s_k^2) \\ &\ll (\log s_n^2)^{-\nu_4} \log\log s_n^2. \end{aligned}$$

We have proved that there is a $\nu > 0$ with

$$\operatorname{Var} \hat{I}_n \le \text{const} \cdot (\log s_n^2)^{-\nu}.$$

Now choose a positive α with $\alpha\nu > 1$. Taking into account the growth conditions (4.13) for s_n^2 it is clear that there is a sequence (n_l) such that

$$l^\alpha \le \log s_{n_l}^2 < (l+1)^\alpha \quad (l \ge L), \tag{4.24}$$

for $\alpha > 1$. The Borel-Cantelli-lemma and (4.20) show

$$\lim_l \hat{I}_{n_l} = \int_{\mathbb{R}} f(x)\, d\Phi(x).$$

Since f is nonnegative, we can easily show that

$$|\hat{I}_{n_l} - \hat{I}_n| \le -\frac{\log s_{n_l}^2}{\log s_{n_{l+1}}^2}\, \hat{I}_{n_l} + \frac{\log s_{n_{l+1}}^2}{\log s_{n_l}^2}\, \hat{I}_{n_{l+1}} ,$$

for $n_l \le n < n_{l+1}$. Hence from (4.24) we infer

$$\lim_n \hat{I}_n = \int_{\mathbb{R}} f(x)\, d\Phi(x).$$

□

Proof Theorem 4.1. Kolmogorov's condition (4.3) trivially implies the Lindeberg condition (1.6) as well as (4.13). The Lindeberg condition implies the CLT (4.10). From Petrov [88, Lemma 7.1] we deduce the exponential bound estimate (4.14). Hence the conditions of Theorem 4.7 are valid. □

Proof Theorem 4.3. This follows directly from Sakhaneko's Theorem A10 combined with Theorem 4.6. □

Proof of Theorem 4.2. We use a classical counterexample of Weiss [110] concerning the law of the iterated logarithm. Let $X_1, X_2, \ldots$ be a sequence of independent random variables, where

$$X_k = \pm \exp\left(k/\log^2 k\right) \quad \text{with probability} \quad \frac{\alpha}{2(\log k + \alpha)}$$

and

$$X_k = 0 \quad \text{with probability} \quad 1 - \frac{\alpha}{\log k + \alpha},$$

where $\alpha > 0$. Thus

$$\sigma_k^2 \sim \exp(2k/\log^2 k)\frac{\alpha}{\log k}$$

and

$$s_k^2 \sim \frac{\alpha}{2}\,\sigma_k^2\,\log^2 k.$$

Therefore we have

$$|X_k| \leq \sqrt{2/\alpha}\,(1 + o(1))\,\frac{s_k}{(\log\log s_k^2)^{1/2}},$$

and thus by choosing α large enough we have (4.6). In [110, Theorem 1] it is shown that with probability 1

$$\limsup_n \frac{S_n}{(2s_n^2 \log\log s_n^2)^{1/2}} > 1.$$

Hence there is a $\delta > 0$ such that almost surely

$$\frac{S_n}{s_n} > ((2+\delta)\log\log s_n^2)^{1/2} \tag{4.25}$$

for infinitely many n. Now let $f(x) = \exp(\gamma x^2)$ where $\frac{1}{2+\delta} < \gamma < 1/2$. If now n is large enough and satisfies (4.25), then we get some $\epsilon > 0$ with

$$\frac{1}{\log s_n^2}\sum_{k=1}^{n}\frac{\sigma_{k+1}^2}{s_k^2}\exp\left(\gamma\,\frac{S_k^2}{s_k^2}\right) \geq \frac{1}{\log s_n^2}\,\frac{\sigma_{n+1}^2}{s_n^2}\exp\left(\gamma\,\frac{S_n^2}{s_n^2}\right) \geq n^{\epsilon} \to \infty.$$

□

Proof of Theorem 4.5. We construct a counterexample using the ideas of Lifshits [72]. We assume without loss of generality that $\psi(t) \leq (\log\log t)^{1/2}$ and define an integer sequence (z_n) with

$$z_n = [\psi^{-1}(n)] + 1 \quad (n \geq 1).$$

(As usual $[x]$ denotes the integer part of x). By the assumption for the growth rate of ψ is easy to see that $z_n \geq \exp\exp(n^2)$. Now let $N_1, N_2, \dots$ be i.i.d. standard normal random variables and let $Y_1, Y_2, \dots$ be independent random variables with $Y_i = 0$ for $i \neq z_k,\ \ k \geq 0$ and

$$Y_{z_k} = \begin{cases} \pm(z_k \log\log z_k)^{1/2} & \text{with probability} \quad \frac{1}{2}(\log\log z_k)^{-3}, \\ 0 & \text{else,} \end{cases}$$

such that the process (Y_i) is independent from (N_i). Finally we set $X_i := N_i + Y_i$. First we show that (X_i) matches the conditions of Theorem 4.5. Observe that

$$EX_n^2 = 1 + z_k(\log\log z_k)^{-2} I\{n = z_k\}, \tag{4.26}$$

and hence for $n \in [z_k, z_{k+1})$ we get by the definition of z_n that

$$s_n^2 = n + \sum_{i=1}^{k} z_i(\log\log z_i)^{-2}.$$

Since the z_k are growing super exponentially only the last summand carries weight and thus

$$s_n^2 = n(1 + \epsilon_n), \quad \text{with} \quad 0 < \epsilon_n \leq 2(\log\log n)^{-2}. \tag{4.27}$$

Consequently we have

$$\frac{\sigma_{n+1}^2}{s_n^2} \sim \begin{cases} \frac{1}{n} & \text{if} \quad n \neq z_k - 1; \\ (\log\log n)^{-2} & \text{if} \quad n = z_k - 1, \end{cases} \tag{4.28}$$

which shows (4.7). By the Markov inequality and by the definition of z_n we have

$$P(|\sum_{k=1}^{n} Y_{z_k}| \geq \varepsilon z_n^{1/2}) \ll (\varepsilon \log\log z_n)^{-2} \quad \forall \varepsilon > 0.$$

Since $Y_i = 0$ if $i \neq z_k$, $k \geq 1$ we infer that $|\sum_{i=1}^{n} Y_i| = o(n^{1/2})$ holds in probability. Thus assumption (4.10) also holds for the sequence (X_n). Finally we note that by the definition of Y_i and the super exponential growth rate of (z_k) there exists for every $h > 0$ some $n_0(h)$ which is not random, such that

$$\left|\sum_{i=1}^{n} Y_i\right| \leq (1+h)(n \log\log n)^{1/2}, \quad \text{for all } n \geq n_0. \tag{4.29}$$

Clearly we can define our random variables on a probability space with $\sum_{i=1}^{n} N_i = W(n)$ where $\{W(t), t \geq 0\}$ is a standard Wiener process. Hence we conclude that for $t \in [s_n^2, s_{n+1}^2)$

$$S(t) = \sum_{i=1}^{n} N_i + \sum_{i=1}^{n} Y_i = W(n) + o\left(t^{1/2}\psi(t)\right). \tag{4.30}$$

From well known properties of the fluctuation of a Wiener process (cf. Csörgő, M. and Révész, P. [31, Theorem 1.2.1.]) we derive that for all $n \geq n_0(\omega)$, with $n_0 < \infty$ a.s.

$$|W(n) - W(n(1+\epsilon_n))| \leq 2\left(n\epsilon_n\left(\log\frac{1}{\epsilon_n} + \log\log n\right)\right)^{1/2} = o\left(n^{1/2}\right) \tag{4.31}$$

Now that

$$\begin{aligned} &P(\sup_{t\in[0,\sigma_{n+1}^2]} |W(s_n^2) - W(s_n^2 + t)| > s_n\psi(n)^{1/2}) \\ &\quad \leq 2P(|W(s_n^2) - W(s_{n+1}^2)| > s_n\psi(n)^{1/2}) \\ &\quad \leq \text{const} \cdot \exp\left(-1/2\frac{s_n^2}{\sigma_{n+1}^2}\psi(n)^{1/2}\right), \end{aligned}$$

we get by (4.28), the assumption $\psi(z_n) \geq n$ and the Borel-Cantelli-lemma

$$\sup_{t\in[0,\sigma_{n+1}^2]} |W(s_n^2) - W(s_n^2 + t)| = o(n^{1/2}\psi(n)) \quad \text{a.s.} \tag{4.32}$$

Combining (4.31) and (4.32) this shows that

$$\begin{aligned} \sup_{t\in[s_n^2, s_{n+1}^2]} |W(n) - W(t)| &\leq |W(n) - W(s_n^2)| + \sup_{t\in[0,\sigma_{n+1}^2]} |W(s_n^2) - W(s_n^2 + t)| \\ &= o(n^{1/2}\psi(n)) \quad \text{a.s.} \end{aligned}$$

which proves (4.9). We have established that (X_n) meets the conditions of Theorem 4.2.

Now we show that the ASCLT fails for $X_1, X_2, \dots$ and $f(x) = x^m$. Set

$$\xi_n^{(1)} := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} N_i \quad \text{and} \quad \xi_n^{(2)} := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i.$$

and denote

$$\xi_n = \xi_n^{(1)} + \xi_n^{(2)} \quad \text{and} \quad \hat{\xi}_n = \operatorname{sign}(\xi_n)(|\xi_n| \wedge ((2+\delta)\log\log n)^{1/2}) \quad (\delta > 0)$$

and in an analogous way we define $\hat{\xi}_n^{(1)}$ and $\hat{\xi}_n^{(2)}$. Let $p > 0$ be an integer. Since

$$(\hat{\xi}_n)^{2p} \geq \frac{1}{2^{2p}} \left(\hat{\xi}_n^{(1)} + \hat{\xi}_n^{(2)} \right)^{2p},$$

we get by independence and symmetry of $\hat{\xi}_n^{(1)}$ and $\hat{\xi}_n^{(2)}$

$$E(\hat{\xi}_n)^{2p} \geq \frac{1}{2^{2p}} E(\hat{\xi}_n^{(2)})^{2p}.$$

By (4.29) $\hat{\xi}_n^{(2)} = \xi_n^{(2)}$, if n is large enough. From the symmetry of the Y_i we get

$$E(\xi_{z_k}^{(2)})^{2p} \geq \frac{1}{z_k^p} \sum_{i=1}^{k} E|Y_{z_i}|^{2p} \geq (\log\log z_k)^{p-3}$$

and hence for $p > 3$ this shows that $\limsup_n E(\hat{\xi}_n)^{2p} = \infty$. We use this result to prove that there is a sequence (n_l) such that

$$\frac{1}{\log s_{n_l}^2} \sum_{k=1}^{n_l} \frac{\sigma_{k+1}^2}{s_k^2} \left(\frac{S_k}{s_k} \right)^{2p} \to \infty.$$

By (4.27) this follows from

$$\frac{1}{\log s_{n_l}^2} \sum_{k=1}^{n_l} \frac{\sigma_{k+1}^2}{s_k^2} \hat{\xi}_k^{2p} \to \infty. \tag{4.33}$$

Similar estimates as in the proof of Theorem 4.7 show, that

$$\operatorname{Cov}(\hat{\xi}_k^{2p}, \hat{\xi}_l^{2p}) \leq \text{const} \cdot (\log\log s_l^2)^{2p} \left(\frac{s_k}{s_l} \right)^{2/3},$$

and consequently that

$$\operatorname{Var} \left(\frac{1}{\log s_n^2} \sum_{k=1}^{n} \frac{\sigma_{k+1}^2}{s_k^2} \hat{\xi}_k^{2p} \right) \to 0.$$

Hence

$$\left| \frac{1}{\log s_n^2} \sum_{k=1}^{n} \frac{\sigma_{k+1}^2}{s_k^2} \hat{\xi}_k^{2p} - \frac{1}{\log s_n^2} \sum_{k=1}^{n} \frac{\sigma_{k+1}^2}{s_k^2} E\hat{\xi}_k^{2p} \right| \xrightarrow{P} 0,$$

which assures almost sure convergence along a subsequence. Since $E\hat{\xi}_k^{2p} \to \infty$ we get (4.33). $\square$

Chapter 5

Upper-lower class tests for martingales

5.1 Introduction

Let $X_1, X_2, \ldots$ be independent random variables with mean 0 and finite variances and let $S_n = \sum_{k=1}^n X_k$, $s_n^2 = \sum_{k=1}^n EX_k^2$. By Kolmogorov's law of the iterated logarithm (Theorem A2), if

$$|X_n| \le \lambda_n s_n / (\log\log s_n^2)^{1/2}$$

with a positive numerical sequence $\lambda_n \to 0$, then

$$\limsup_{n\to\infty} (2s_n^2 \log\log s_n^2)^{-1/2} S_n = 1 \qquad \text{a.s.}$$

A much more refined result was proved by Feller [44], who showed that if

$$|X_n| \le K s_n / (\log\log s_n^2)^{3/2} \tag{5.1}$$

and $\varphi : \mathbb{R}^+ \to \mathbb{R}^+$ is a nondecreasing function, then

$$P\{S_n > s_n \varphi(s_n) \quad \text{i.o.}\} = 0 \quad \text{or} \quad 1 \tag{5.2}$$

according as

$$I(\varphi) := \int_1^\infty t^{-1}\varphi(t)e^{-\varphi(t)^2/2}dt < \infty \quad \text{or} \quad = \infty. \tag{5.3}$$

Condition (5.1) is best possible: replacing it by

$$|X_n| \le K_n s_n/(\log\log s_n^2)^{3/2}$$

with any fixed sequence $K_n \to \infty$, the test (5.2)–(5.3) becomes generally false.

Using truncation, the above result extends easily for sequences (X_n) of unbounded r.v.'s. For example, Feller showed that the test (5.2)–(5.3) remains valid if we replace (5.1) by

$$\sum_{k=1}^\infty s_k^{-2}(\log\log s_k)^3 EX_k^2 I\{|X_k| \ge Ms_k/(\log\log s_k)^{3/2}\} < \infty, \tag{5.4}$$

for some positive constant M. This condition is obviously satisfied if (5.1) holds and covers also a large class of unbounded sequences, but condition (5.4) is far from optimal. For example, in the case of i.i.d. sequences (X_n) relation (5.4) requires $EX_1^2(\log|X_1|)^\alpha < \infty$ for some $\alpha > 1$, which is too strong. Using a more delicate truncation argument, Feller [45] showed that if X_n are i.i.d. random variables with mean 0 and finite variance, then the test (5.2)–(5.3) is valid provided

$$EX_1^2 I\{|X_1| > t\} = O((\log\log t)^{-1},$$

and the last condition is best possible. In particular, the test holds if

$$EX_1^2 \log\log|X_1| < \infty. \tag{5.5}$$

The previous results give a fairly complete description of the upper-lower class behavior of independent random variables. Using strong approximation methods, Strassen [103] was the first to extend Feller's results for dependent random variables. Let $\{X_n, \mathcal{F}_n, n \ge 1\}$ be a martingale difference sequence with finite second moments and let $S_n = X_1 + \cdots + X_n$, $s_n^2 = \sum_{k=1}^n E[X_k^2|\mathcal{F}_{k-1}]$. Strassen proved that the test (5.2)–(5.3) remains valid if $s_n^2 \to \infty$ a.s. and

$$\sum_{k=1}^\infty s_k^{-2}(\log s_k)^5 E[X_k^2 I\{|X_k| > s_k/(\log s_k)^5\}|\mathcal{F}_{k-1}] < \infty \qquad \text{a.s.}$$

The last relation is similar to (5.4), but it is considerably more restrictive. Strassen's result was improved gradually by Jain, Jogdeo and Stout [65], Philipp and Stout [91] and Einmahl and Mason [38]. The last authors proved that the test (5.2)–(5.3) remains valid if condition (5.1) holds, a result which is obviously optimal. By truncation, this implies the test (5.2)–(5.3) under the conditional version of (5.4), i.e.

$$\sum_{k=1}^{\infty} s_k^{-2}(\log\log s_k)^3 E[X_k^2 I\{|X_k| \geq M s_k/(\log\log s_k)^{3/2}\}|\mathcal{F}_{k-1}] < \infty \qquad \text{a.s.}$$

with some $M > 0$. Similar criteria are given in Jain, Jogdeo and Stout [65] and Philipp and Stout [91], but they are all far from optimal. For example, Theorem 3.1 of Jain, Jogdeo and Stout [65] implies (see their Remark 2) that the test (5.2)–(5.3) holds for stationary, ergodic martingale difference sequences $\{X_n, \mathcal{F}_n, n \geq 1\}$ under

$$EX_1^2 \log|X_1|(\log\log|X_1|)^2 < \infty. \tag{5.6}$$

In analogy with the i.i.d. case, it is natural to expect that in the stationary case Feller's condition (5.5) suffices for the test (5.2)–(5.3), but this conjecture remained open until today.

The purpose of this chapter is to give an upper-lower class test for unbounded martingale difference sequences which not only improves earlier results in the field, but it is essentially optimal. In particular, we will prove that for stationary ergodic martingale difference sequences Feller's condition (5.5) implies the test (5.2)–(5.3).

Theorem 5.1. *Let $\{X_n, \mathcal{F}_n,\, n \geq 1\}$ be a martingale difference sequence with $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$, finite variances and assume that $s_n^2 := \sum_{k=1}^n E[X_k^2|\mathcal{F}_{k-1}] \to \infty$ a.s. Let $S_n = X_1 + \cdots + X_n$, $B_n^2 = \sum_{k=1}^n EX_k^2$, $f(k) = B_k/(\log\log B_k)^{1/2}$ and assume that the following conditions hold:*

(**a**) $\sum_{n=1}^\infty f(n)^{-4} EX_n^4 I\{|X_n| \leq \delta f(n)\} < \infty$ *for some $\delta > 0$;*

(**b**) $\sum_{n=1}^\infty f(n)^{-1} E|X_n| I\{|X_n| \geq \varepsilon f(n)\} < \infty$ *for all $\varepsilon > 0$;*

(**c**) $B_n^{-2}\sum_{k=1}^n X_k^2 \to 1$ a.s.

Then for any positive, nondecreasing function φ the test (5.2)–(5.3) holds.

Theorem 5.2. *Let $\{X_n, \mathcal{F}_n, n \geq 1\}$ be a stationary ergodic martingale difference sequence with*

$$EX_1^2 \log\log|X_1| < \infty.$$

Then for any positive, nondecreasing function φ the test (5.2)–(5.3) *holds.*

5.2 Preliminary lemmas

We start with defining a truncated MDS $\{X_k^*, \mathcal{F}_k^*, n \geq 1\}$ as follows:

$$X_k^* = X_k I\{|X_k| \leq \delta f(k)\} - E[X_k I\{|X_k| \leq \delta f(k)\}|\mathcal{F}_{k-1}^*],$$

where

$$f(k) = B_k/(\log\log B_k)^{1/2},$$

with the usual convention $\log\log x := \log(\log\min\{x, e^2\})$. Here the filtration $\mathcal{F}_k^* = \sigma(X_1^*, \dots, X_k^*)$ if $k \geq 1$ and $\mathcal{F}_0^* = \{\emptyset, \Omega\}$. Denote by S_n^* the partial sum $X_1^* + \cdots + X_n^*$ and similar to s_n^2 set

$$s_n^{*\,2} = \sum_{k=1}^n E[X_k^{*2}|\mathcal{F}_{k-1}^*].$$

Finally let $X_k^{**} = X_k - X_k^*$ and define $S_n^{**} = X_1^{**} + \cdots + X_n^{**}$. First we observe that

$$|E[X_k I\{|X_k| \leq \delta f(k)\}|\mathcal{F}_{k-1}^*]| = |E[X_k I\{|X_k| > \delta f(k)\}|\mathcal{F}_{k-1}^*]| \quad \text{a.s.} \tag{5.7}$$

Since

$$E[X_k|\mathcal{F}_{k-1}^*] = E[X_k I\{|X_k| \leq \delta f(k)\}|\mathcal{F}_{k-1}^*] + E[X_k I\{|X_k| > \delta f(k)\}|\mathcal{F}_{k-1}^*],$$

(5.7) will follow, once we show that

$$E[X_k|\mathcal{F}_{k-1}^*] = 0 \quad \text{a.s.}$$

Since (X_k) is an MDS, we know that

$$E[X_k|\mathcal{F}_{k-1}] = 0 \quad \text{a.s.} \tag{5.8}$$

By the definition of a conditional expected value we have

$$\int_A X_k\,dP = \int_A E[X_k|\mathcal{F}^*_{k-1}]\,dP \quad \text{for all } A \in \mathcal{F}^*_{k-1}.$$

Since $\mathcal{F}^*_k \subset \mathcal{F}_k$ it follows from (5.8) that

$$\int_A E[X_k|\mathcal{F}^*_{k-1}]\,dP = 0 \quad \text{for all } A \in \mathcal{F}^*_{k-1},$$

and consequently that $E[X_k|\mathcal{F}^*_{k-1}] = 0$ a.s.

From now on $a_n \sim b_n$ means $a_n/b_n \to 1$ for $n \to \infty$.

Lemma 5.1. *Under condition* (**b**) *we have*

$$f(n)^{-2}\sum_{k=1}^n X_k^2 I\{|X_k| > \delta f(k)\} \to 0 \quad \text{a.s.}$$

Proof. We have

$$\begin{aligned}\sum_{k=1}^n X_k^2 I\{|X_k| > \delta f(k)\} &\leq \sup_{k\leq n}|X_k| I\{|X_k| > \delta f(k)\}\sum_{k=1}^n |X_k| I\{|X_k| > \delta f(k)\} \\ &\leq \left(\sum_{k=1}^n |X_k| I\{|X_k| > \delta f(k)\}\right)^2,\end{aligned}$$

and thus the result follows from (**b**) and Kronecker's lemma. □

Lemma 5.2. *Under conditions* (**a**)–(**c**) *we have*

$$s_n^{*2} \sim B_n^2 \sim s_n^2 \quad \text{a.s.}$$

Proof. We only need to show that $s_n^{*2} \sim B_n^2$. The second relation is similar, in fact simpler. An easy calculation gives

$$\begin{aligned}&B_n^{-2}\sum_{k=1}^n E[X_k^{*2}|\mathcal{F}^*_{k-1}] \\ &= B_n^{-2}\sum_{k=1}^n E[X_k^2 I\{|X_k| \leq \delta f(k)\}|\mathcal{F}^*_{k-1}] - B_n^{-2}\sum_{k=1}^n (E[X_k I\{|X_k| \leq \delta f(k)\}|\mathcal{F}^*_{k-1}])^2.\end{aligned} \tag{5.9}$$

Now by (5.7) we obtain

$$\begin{aligned} &B_k^{-2}(E[X_k I\{|X_k| \le \delta f(k)\}|\mathcal{F}^*_{k-1}])^2 \\ &\qquad \le \delta(f(k)\log\log B_k)^{-1} E[|X_k| I\{|X_k| > \delta f(k)\}|\mathcal{F}^*_{k-1}], \end{aligned}$$

and thus by Kronecker's lemma and (**b**) the second term in (5.9) tends to zero. Set

$$Y_k := X_k^2 I\{|X_k| \le \delta f(k)\} - E[X_k^2 I\{|X_k| \le \delta f(k)\}|\mathcal{F}^*_{k-1}].$$

Clearly (Y_k) is an MDS and since by (**a**)

$$\sum_{k=1}^{\infty} B_k^{-4} E Y_k^2 \le \sum_{k=1}^{\infty} B_k^{-4} E X_k^4 I\{|X_k| \le \delta f(k)\} < \infty,$$

we get from the law of large numbers for martingales (c.f. [47, Corollary 6.7.2.])

$$B_n^{-2} \sum_{k=1}^{n} \left(X_k^2 I\{|X_k| \le \delta f(k)\} - E[X_k^2 I\{|X_k| \le \delta f(k)\}|\mathcal{F}^*_{k-1}]\right) \to 0 \quad \text{a.s.}$$

In view of (**c**), (5.9) and the last relation we have to show that

$$B_n^{-2} \sum_{k=1}^{n} X_k^2 I\{|X_k| > \delta f(k)\} \to 0 \quad \text{a.s.},$$

which follows from Lemma 5.1. □

Lemma 5.3. *Under conditions* (**a**)–(**b**) *we have*

$$f(n)^{-2} \sum_{k=1}^{n} E X_k^{*4} \to 0 \quad \text{a.s.}$$

Proof. We have

$$\begin{aligned} &\sum_{k=1}^{\infty} f(k)^{-4} E X_k^{*4} \\ &\quad = \sum_{k=1}^{\infty} f(k)^{-4} E(X_k I\{|X_k| \le \delta f(k)\} - E[X_k I\{|X_k| \le \delta f(k)\}|\mathcal{F}^*_{k-1}])^4 \\ &\quad \le \sum_{k=1}^{\infty} f(k)^{-4} \left(E X_k^4 I\{|X_k| \le \delta f(k)\} + 15\delta^3 f(k)^3 E|E[X_k I\{|X_k| \le \delta f(k)\}|\mathcal{F}^*_{k-1}]|\right) \\ &\quad \le \sum_{k=1}^{\infty} f(k)^{-4} \left(E X_k^4 I\{|X_k| \le \delta f(k)\} + 15\delta^3 f(k)^3 E|X_k| I\{|X_k| > \delta f(k)\}\right), \end{aligned}$$

where the last sum is finite by (**a**) and (**b**). Hence the proof follows from Kronecker's lemma. □

Lemma 5.4. *Under conditions* (**a**)–(**c**) *we have*

$$\frac{\log\log s_n}{s_n^2}(s_n^2 - s_n^{*\,2}) \to 0 \quad \text{a.s.}$$

Proof. According to Lemma 5.2 it suffices to show that

$$f(n)^{-2}(s_n^2 - s_n^{*\,2}) \to 0 \quad \text{a.s.}$$

We can write

$$\begin{aligned} s_n^2 - s_n^{*\,2} = \sum_{k=1}^{n} \big(E[X_k^2|\mathcal{F}_{k-1}] - E[X_k^{*2}|\mathcal{F}_{k-1}]\big) \\ + \sum_{k=1}^{n} \big(E[X_k^{*2}|\mathcal{F}_{k-1}] - E[X_k^{*2}|\mathcal{F}_{k-1}^{*}]\big) := S_n^1 + S_n^2 \quad \text{(say)}. \end{aligned}$$

A straightforward calculation shows that

$$\begin{aligned} &|E[X_k^2 - X_k^{*2}|\mathcal{F}_{k-1}]| \\ &\quad \le E[X_k^2 I\{|X_k| > \delta f(k)\}|\mathcal{F}_{k-1}] + (E[X_k I\{|X_k| \le \delta f(k)\}|\mathcal{F}_{k-1}^{*}])^2 \\ &\quad +2|E[X_k I\{|X_k| \le \delta f(k)\}|\mathcal{F}_{k-1}]||E[X_k I\{|X_k| \le \delta f(k)\}|\mathcal{F}_{k-1}^{*}]| \\ &\quad \le E[X_k^2 I\{|X_k| > \delta f(k)\}|\mathcal{F}_{k-1}] + 3\delta f(k) E[|X_k| I\{|X_k| > \delta f(k)\}|\mathcal{F}_{k-1}^{*}]. \end{aligned}$$

Thus Lemma 5.1 and (**b**) in connection with Kronecker's lemma imply that

$$f(n)^{-2} S_n^1 \to 0 \quad \text{a.s.}$$

It remains to prove

$$f(n)^{-2} S_n^2 \to 0 \quad \text{a.s.,}$$

which follows immediately from Lemma 5.3. □

By a standard argument in this theory we can assume that for some $0 < a \le 1 < 2 \le b < \infty$

$$a\sqrt{\log\log t} \le \varphi(t) \le b\sqrt{\log\log t}. \tag{5.10}$$

Hence we can assume without loss of generality that the function φ which occurs in Theorem 5.1 tends to ∞. If φ is some nondecreasing function with $\lim_{t\to\infty}\varphi(t)=\infty$ than we can write

$$\varphi(t)=\tilde{\varphi}(t)+\frac{1}{\tilde{\varphi}(t)} \quad \text{with} \quad \tilde{\varphi}(t)=\frac{\varphi(t)}{2}+\left(\frac{\varphi^2(t)}{4}-1\right)^{1/2}, \tag{5.11}$$

if $\varphi(t)\geq 2$. It is very easy to see that $I(\varphi)=\infty$ if and only if $I(\tilde{\varphi})=\infty$. The next lemma shows it is enough to consider only functions φ which are in some sense smooth.

Lemma 5.5. *Assume that $I(\varphi)=\infty(<\infty)$. Then there is some $\hat{\varphi}\geq\varphi(\leq\varphi)$ and some absolute constant A such that $I(\hat{\varphi})=\infty(<\infty)$ and*

$$|\hat{\varphi}(x)-\hat{\varphi}(y)|\leq A\cdot\frac{\hat{\varphi}(x)}{x}|x-y| \quad \textit{if} \quad [y,2y]\cap[x,2x]\neq\emptyset. \tag{5.12}$$

Proof. We assume that $I(\varphi)=\infty$. Define

$$\hat{\varphi}(x)=\frac{1}{x}\int_x^{2x}\varphi(t)\,dt \quad (x>x_0).$$

Since φ is monotone we have

$$\varphi(x)\leq\hat{\varphi}(x)\leq\varphi(2x).$$

We can also assume that φ is trapped as in (5.10) and therefore it follows that for some $c>0$

$$\frac{a}{2b}\leq\frac{\varphi(x)}{\varphi(cx)} \quad \text{for all} \quad x>x_0(c). \tag{5.13}$$

By simple analysis we get that $I(\hat{\varphi})=\infty$. Next we write

$$\hat{\varphi}(x)-\hat{\varphi}(y)=\left(\frac{1}{x}-\frac{1}{y}\right)\int_x^{2x}\varphi(t)\,dt+\frac{1}{y}\left(\int_x^{2x}\varphi(t)\,dt-\int_y^{2y}\varphi(t)\,dt\right).$$

Using that φ is monotone, (5.13) and the condition on x and y we get immediately that

$$|\hat{\varphi}(x)-\hat{\varphi}(y)|\leq 4\frac{|x-y|}{y}\varphi(4x)\leq\frac{16b}{a}\frac{|x-y|}{x}\hat{\varphi}(x).$$

The case $I(\varphi) < \infty$ can be treated similarly by defining

$$\hat{\varphi}(x) = \int_{x/2}^{x} \varphi(t)\, dt.$$

Assume that we can show (5.21) for functions that satisfy (5.12). Now if $I(\varphi) = \infty$ it follows that $I(\hat{\varphi}) = \infty$ and thus

$$1 = P(S_k^* > \hat{\varphi}(s_k^*)s_k^* \text{ i.o.}) = P(S_k > \hat{\varphi}(s_k)s_k \text{ i.o.}) \leq P(S_k > \varphi(s_k)s_k \text{ i.o.}).$$

An analogues result holds if $I(\varphi) < \infty$. □

5.3 Proofs

We first observe that

$$|X_n^*| \leq s_n^* K_n, \tag{5.14}$$

where $K_n = 2f(k)/s_k^*$, i.e. $K_n \in \mathcal{F}_{n-1}^*$, $K_n s_n^* \nearrow \infty$ and by Lemma 5.2 $K_n \to 0$. Following Einmahl and Mason [38] we can assume that the sequence (X_k^*) is defined on the probability space of some standard Wiener Process $\{W(t),\, t \geq 0\}$ such that

$$S_n^* = W_{T_n}, \quad \text{where} \quad T_n = \sum_{m=1}^{n} \tau_m, \tag{5.15}$$

where τ_n are non-negative and $\mathcal{F}_{n-1}^*$ measurable for each $n \geq 1$ and

$$E[\tau_n | \mathcal{F}_{n-1}^*] = E[X_n^{*2} | \mathcal{F}_{n-1}^*] \quad \text{a.s.} \tag{5.16}$$

Also we have for any $r \geq 1$

$$E[\tau_n^r | \mathcal{F}_{n-1}^*] \leq L_r E[X_n^{*2r} | \mathcal{F}_{n-1}^*] \quad \text{a.s.,} \tag{5.17}$$

where L_r is some constant which depends only on r and moreover, for $T_n \leq t \leq T_{n+1}$

$$|W(t) - W(T_n)| \leq s_{n+1}^* K_{n+1}. \tag{5.18}$$

The following important lemma is implicit in the proof of Theorem 1.1 of Einmahl and Mason [38].

Lemma 5.6. *Assume that* $\{X_n^*, \mathcal{F}_{n-1}^*\, n \geq 1\}$ *is an MDF with finite variances such that* $s_n^* := \sum_{k=1}^n E[X_k^{*2}|\mathcal{F}_{k-1}^*] \to \infty$. *Assume further that (5.14) holds with some* $K_n \sim \mathrm{const} \cdot (\log\log s_n^*)^{-1/2}$. *Finally let* T_n *be given as in (5.15). If there exists some positive constant* K *such that*

$$\limsup_{n\to\infty} \frac{\log\log s_n^*}{{s_n^*}^2} |T_n - s_n^*| \leq K \quad \text{a.s.,} \tag{5.19}$$

then for every positive and nondecreasing function φ

$$P(S_k^* > s_k^* \varphi(s_k^*) \text{ i.o.}) = \begin{cases} 1 & \text{if } I(\varphi) = \infty, \\ 0 & \text{if } I(\varphi) < \infty. \end{cases}$$

Proof. We note that we implicitly assumed that S_n^* is defined on the same space with some Wiener process $\{W(t),\, t \geq 0\}$ such that (5.15) holds. Of course we can use the Skorokhod embedding from above and hence this is no loss of generality. We note that our assumptions imply that $s_n^* \sim s_{n+1}^*$. Thus it follows from (5.18) and (5.22) below that for $T_n \leq t \leq T_{n+1}$ and for sufficiently large n

$$|W(t) - W(T_n)| \leq 2\sqrt{t}/(\log\log t)^{1/2} \quad \text{a.s.} \tag{5.20}$$

Now the proof of Theorem 1.1 of Einmahl and Mason [38] can be taken almost verbally, observing that the argument still goes through if their equation (2.6) is replaced by (5.20). □

Proof of Theorem 5.1. The proof will be divided into two steps. In the first step we will show that the integral test holds for the truncated MDS $\{X_k^*, \mathcal{F}_{k-1}^*\, k \geq 1\}$. Then we will show that

$$P(S_k > \varphi(s_k^2) s_k \text{ i.o.}) = P(S_k^* > \varphi({s_k^*}^2) s_k^* \text{ i.o.}). \tag{5.21}$$

Step 1. By Lemma 5.2 and Lemma 5.6 it suffices to show that

$$|T_n - {s_n^*}^2| = o\big(f(n)^2\big) \quad \text{a.s.} \tag{5.22}$$

Using the explicit expression of ${s_n^*}^2$ and (5.16) we have

$$T_n - {s_n^*}^2 = \sum_{k=1}^n (\tau_k - E[X_k^{*2}|\mathcal{F}_{k-1}^*]) = \sum_{k=1}^n (\tau_k - E[\tau_k|\mathcal{F}_{k-1}^*]) \quad \text{a.s.}$$

By (5.17) we get

$$\sum_{k=1}^{\infty} f(k)^{-4} E(\tau_k - E[\tau_k|\mathcal{F}_{k-1}^*])^2$$
$$\leq \sum_{k=1}^{\infty} f(k)^{-4} E\tau_k^2 \leq L_2 \sum_{k=1}^{\infty} f(k)^{-4} E X_k^{*4},$$

and hence by the arguments in the proof of Lemma 5.3 the last series is convergent, showing relation (5.22).

Step 2. Define $\tilde{\varphi}$ as in (5.11) and set

$$R_k = \frac{s_k}{\tilde{\varphi}(s_k^2)} + (\tilde{\varphi}(s_k^2)s_k - \tilde{\varphi}(s_k^{*2})s_k^*).$$

Then we have on the one hand

$$\begin{aligned} &P(S_k > \varphi(s_k^2)s_k \text{ i.o.}) \\ &\quad = P(S_k^* + S_k^{**} > \tilde{\varphi}(s_k^{*2})s_k^* + R_k \text{ i.o.}) \\ &\quad \leq P(S_k^* > \tilde{\varphi}(s_k^{*2})s_k^* \text{ i.o.}) + P(|S_k^{**}| > R_k \text{ i.o.}). \end{aligned}$$

and on the other hand

$$\begin{aligned} &P(S_k > \varphi(s_k^2)s_k \text{ i.o.}) \\ &\quad \geq P(S_k^* > \tilde{\varphi}(s_k^{*2})s_k^* + |R_k - S_k^{**}| \text{ i.o.}) \\ &\quad = P(S_k^* > \tilde{\tilde{\varphi}}(s_k^{*2})s_k^* + s_k^*\tilde{\tilde{\varphi}}(s_k^{*2})^{-1} + |R_k - S_k^{**}| \text{ i.o.}). \end{aligned}$$

We show now that

$$P(|S_k^{**}| > R_k \text{ i.o.}) = 0 \quad \text{and} \quad |R_k - S_k^{**}| \leq \frac{(\kappa-1)}{2} s_k^* \tilde{\tilde{\varphi}}(s_k^{*2})^{-1}(1+o(1)) \quad \text{a.s.,} \tag{5.23}$$

for some large enough κ, which implies in view of the forgoing estimates that

$$\begin{aligned} &P(S_k^* > \tilde{\tilde{\varphi}}(s_k^{*2})s_k^* + \kappa s_k^* \tilde{\tilde{\varphi}}(s_k^{*2})^{-1} \text{ i.o.}) \\ &\quad \leq P(S_k > \varphi(s_k^2)s_k \text{ i.o.}) \leq P(S_k^* > \tilde{\varphi}(s_k^{*2})s_k^* \text{ i.o.}). \end{aligned}$$

Since $I(\varphi) = \infty$ if and only if $I(\tilde{\varphi}) = \infty$ and $I(\tilde{\tilde{\varphi}} + \kappa/\tilde{\tilde{\varphi}}) = \infty$ we get (5.21). We have to prove (5.23) and start with showing that the dominating part in R_k is $s_k/\tilde{\varphi}(s_k^2)$, i.e.

$$\frac{\tilde{\varphi}(s_k^2)}{s_k}(\tilde{\varphi}(s_k^2)s_k - \tilde{\varphi}(s_k^{*2})s_k^*) \to 0 \quad \text{a.s.} \tag{5.24}$$

In view of Lemma 5.2 and

$$\tilde{\varphi}(s_k^2)s_k - \tilde{\varphi}(s_k^{*2})s_k^* = \frac{\tilde{\varphi}(s_k^2)(s_k^2 - s_k^{*2})}{s_k + s_k^*} + s_k^*(\tilde{\varphi}(s_k^2) - \tilde{\varphi}(s_k^{*2}))$$

this will follow if

$$\frac{\tilde{\varphi}(s_k^2)^2}{s_k^2}(s_k^2 - s_k^{*2}) \to 0 \quad \text{a.s.} \quad \text{and} \quad \tilde{\varphi}(s_k^2)(\tilde{\varphi}(s_k^2) - \tilde{\varphi}(s_k^{*2})) \to 0 \quad \text{a.s.} \tag{5.25}$$

Now the first part in (5.25) follows from Lemma 5.4 and (5.10). By Lemma 5.5 it suffices to prove (5.21) for the smoothed version of φ. Hence without loss of generality we can assume that $\tilde{\varphi}$ satisfies (5.12). Clearly, since $s_k^{*2} \sim s_k^2$ the intervals $[s_k^2, 2s_k^2]$ and $[s_k^{*2}, 2s_k^{*2}]$ will not be disjoint for any $k \geq k_0(\omega)$, where k_0 is almost surly finite. Thus we get from (5.12)

$$\tilde{\varphi}(s_k^2)(\tilde{\varphi}(s_k^2) - \tilde{\varphi}(s_k^{*2})) \leq \frac{16b}{a}\frac{\tilde{\varphi}(s_k^2)^2}{s_k^2}(s_k^2 - s_k^{*2}) \quad \text{for all } k \geq k_0,$$

where the righthand side tends to zero as we have already noted. In order to proof the first relation in (5.23) it suffices to show

$$|S_k^{**}| = o\left(\frac{s_k}{\tilde{\varphi}(s_k^2)}\right) \quad \text{a.s.} \tag{5.26}$$

which by Lemma 5.2 and (5.10) will follow if

$$|S_k^{**}| = o\left(f(k)\right) \quad \text{a.s.} \tag{5.27}$$

It easy to show that $E|X_k^{**}| \leq 2\,E|X_k|I\{|X_k| > \delta f(k)\}$ and hence (5.27) follows from (**b**) and Kronecker's lemma.

Since we proved that $|S_k^{**}| = o(R_k)$ a.s. we have in consideration of (5.10)

$$\begin{aligned} |R_k - S_k^{**}| &= |R_k|(1 + o(1)) \\ &= \frac{s_k}{\tilde{\varphi}(s_k^2)}(1 + o(1)) \leq \frac{b}{a}\frac{s_k^*}{\tilde{\tilde{\varphi}}(s_k^{*2})}(1 + o(1)) \quad \text{a.s.} \end{aligned}$$

This shows the second relation of (5.23). □

Proof of Theorem 5.2. It is shown in Jain et al. [65, Lemma 4.1.] that under the assumptions of Theorem 5.2 conditions (**a**)–(**c**) are satisfied. □

Chapter 6

The functional CLT for augmented GARCH sequences

6.1 Preliminaries

6.1.1 Definitions and existence conditions

The seminal work of Engle [39] gave a new impact to the theory of time series analysis. Engle introduced the ARCH (autoregressive conditionally heteroscedastic) process, which allows the conditional variance of the time series to change as a function of past observations. Since then, this model and its extensions has been widely used in econometrics to describe financial data with time varying volatility. One of the most popular models is the GARCH (p, q) process introduced by Bollerslev [21]. A sequence $\{y_k, -\infty < k < \infty\}$ is a GARCH (p, q) process if it satisfies the equations

$$y_k = \sigma_k \varepsilon_k \tag{6.1}$$

and

$$\sigma_k^2 = w + \sum_{1 \le i \le p} \beta_i \sigma_{k-i}^2 + \sum_{1 \le j \le q} \alpha_j y_{k-j}^2, \tag{6.2}$$

where

$$w > 0, \quad \beta_i \ge 0 \quad (1 \le i \le p), \quad \alpha_j \ge 0 \quad (1 \le j \le q), \tag{6.3}$$

and $\{\varepsilon_k, -\infty < k < \infty\}$ is an i.i.d. sequence of real r.v.'s. Bougerol and Picard [23] found necessary and sufficient conditions for the existence of a unique strictly stationary solution of (6.1) and (6.2).

Despite the wide applicability of the GARCH (p, q) model, the quadratic dependence of the volatility σ_k^2 on past values of the process is often unrealistic. In typical stock market situations, large negative y_k's can have a totally different effect on the volatility than large positive values; also, the dependence of the volatility on y_k can have a threshold character, and it can be markedly non-quadratic even for positive y_k's. Starting with the exponential GARCH (EGARCH) of Nelson and the asymmetric GARCH (AGARCH) process of Engle and Ng, in the '90s several extensions of the GARCH model were introduced to deal with 'special effects' of the above kind. See Section 6.1.2 below for examples. In 1997, Duan [37] unified the theory by introducing the so-called augmented GARCH process, containing most of the above mentioned models as a special case. A sequence of r.v.'s $\{y_k, -\infty < k < \infty\}$ is called an augmented GARCH $(1, 1)$ sequence if it satisfies

$$y_k = \sigma_k \varepsilon_k \tag{6.4}$$

and

$$\Lambda(\sigma_k^2) = c(\varepsilon_{k-1})\Lambda(\sigma_{k-1}^2) + g(\varepsilon_{k-1}), \tag{6.5}$$

where

$$\{\varepsilon_k, -\infty < k < \infty\} \text{ is an i.i.d. sequence,} \tag{6.6}$$

$\Lambda(x)$, $c(x)$ and $g(x)$ are real-valued and measurable functions and

$$\Lambda^{-1}(x) \quad \text{exists.} \tag{6.7}$$

Obviously the GARCH $(1, 1)$ model of Bollerslev [21] satisfies (6.4)-(6.7). Nelson [83] showed that if $p = q = 1$ and $E \log^+(\beta_1 + \alpha_1 \varepsilon_0^2) < \infty$, then a strictly stationary solution of (6.1) and (6.2) exists if and only if

$$E \log(\beta_1 + \alpha_1 \varepsilon_0^2) < 0.$$

Moreover, this solution is unique. The following result of Aue, Berkes, and Horváth [5] extends this result for augmented GARCH $(1, 1)$ sequences.

Theorem D. *Assume that (6.4)-(6.7) hold. If*

$$E \log^+ |g(\varepsilon_0)| < \infty \quad \textit{and} \quad E \log^+ |c(\varepsilon_0)| < \infty \tag{6.8}$$

and

$$E \log |c(\varepsilon_0)| < 0 \tag{6.9}$$

then the unique strictly stationary solution of (6.4) and (6.5) is given by

$$\Lambda(\sigma_k^2) = \sum_{i=1}^{\infty} g(\varepsilon_{k-i}) \prod_{1 \le j < i} c(\varepsilon_{k-j}), \tag{6.10}$$

where the series in (6.10) is almost surely convergent. If in addition

$$P(g(\varepsilon_0) = 0) < 1, \quad c(\varepsilon_0) \ge 0 \quad \textit{and} \quad g(\varepsilon_0) \ge 0 \tag{6.11}$$

hold, then (6.9) is also necessary for the existence of a strictly stationary and non-negative solution of (6.4) and (6.5).

Nelson [83] showed that for a GARCH $(1,1)$ process we have $E|y_0|^{2p} < \infty$ $(p > 0)$ if and only if $E(\beta_1 + \alpha_1 \varepsilon_0^2)^p < 1$. Aue et al. [5] provide conditions for the existence of moments of augmented GARCH sequences. We state their result below.

Theorem E. *Assume that (6.4)-(6.7) hold. If for some $\mu > 0$*

$$E|g(\varepsilon_0)|^{\mu} < \infty \tag{6.12}$$

and if

$$E|c(\varepsilon_0)|^{\mu} < 1, \tag{6.13}$$

then $E|\Lambda(\sigma_0^2)|^{\mu} < \infty$. On the other hand, if the the series in (6.10) is convergent, (6.11) holds and if $E|\Lambda(\sigma_0^2)|^{\mu} < \infty$, then (6.12) and (6.13) are satisfied.

Remark 6.1. *Note that from Jensen's inequality and (6.13) it follows that (6.9) holds. Hence from Theorem D we conclude that (6.12) and (6.13) imply the existence of a unique stationary solution of (6.4) and (6.5), where $\Lambda(\sigma_k^2)$ is given by (6.10).*

6.1.2 Examples

Duan [37] and Carrasco and Chen [26] gave several examples of augmented GARCH $(1,1)$ type processes appearing in the literature satisfying (6.4),(6.5) and (6.7). For the convenience of the reader we give a brief overview.

Example 6.1. The GARCH $(1,1)$ model introduced by Bollerslev [21] satisfies

$$\Lambda(x) = x, \quad c(x) = \beta + \alpha x^2 \quad \text{and} \quad g(x) = w,$$

with $w > 0$ and $\alpha, \beta \geq 0$.

Example 6.2. In the asymmetric power ARCH $(1,1)$ model (APARCH) by Ding, Granger and Engle [35] we have

$$\Lambda(x) = x^\delta, \quad c(x) = \beta \quad \text{and} \quad g(x) = w + \alpha(|x| - \mu x)^{2\delta},$$

where $w, \delta > 0$, $|\mu| \leq 1$ and $\alpha, \beta \geq 0$. If $\delta = 1$ this model is referred to as asymmetric GARCH (AGARCH).

Example 6.3. The threshold GARCH $(1,1)$ model (TGARCH) is defined by

$$\Lambda(x) = x^\delta, \quad c(x) = \beta + \alpha_1 (x^2)^\delta I_{\{x<0\}} + \alpha_2 (x^2)^\delta I_{\{x \geq 0\}} \quad \text{and} \quad g(x) = w,$$

where $w, \delta, \alpha_1 > 0$ and $\alpha_2, \beta \geq 0$. The special case $\delta = 1/2$ was proposed by Taylor [107] and Schwert [98] and includes the threshold model of Zakoian [111]. If $\delta = 1$ this is the GJR-ARCH model by Glosten, Jagannathan and Runkle [50].

Example 6.4. The quadratic GARCH $(1,1)$ (QGARCH) was considered by Engle and Ng [40] and Sentana [99]. It is given by

$$\Lambda(x) = x, \quad c(x) = \beta \quad \text{and} \quad g(x) = w + \alpha(x + \mu)^2,$$

where $w > 0$ and $\alpha, \beta \geq 0$ and μ is some constant.

Example 6.5. The nonlinear symmetric GARCH model (NGARCH) was introduced by Engle and Ng [40]. It satisfies

$$\Lambda(x) = x, \quad c(x) = \beta + \alpha(x + \mu)^2 \quad \text{and} \quad g(x) = w,$$

where $w > 0$ and $\alpha, \beta \geq 0$ and μ is some constant.

Example 6.6. Carrasco and Chen [26] introduced the power GARCH (PGARCH) model. The PGARCH $(1,1)$ model is defined by

$$\Lambda(x) = x^\delta, \quad c(x) = \beta + \alpha(x^2)^\delta \quad \text{and} \quad g(x) = w,$$

where $w > 0$ and $\alpha, \beta \geq 0$ and μ is some constant.

Example 6.7. The smooth transition ARCH $(1,1)$ model (STARCH) of Hagerud [51] is given by

$$\Lambda(x) = x, \quad c(x) = \beta \quad \text{and} \quad g(x) = w + (\alpha_1 + \alpha_2 F(x))\, x^2,$$

where $w > 0$, $\beta \geq 0$ and $F(x)$ is a so-called transition function. In all the examples given so far we have $\Lambda(x) = x^\delta$ and thus they are called polynomial GARCH. The last two examples are referred to as exponential GARCH.

Example 6.8. Geweke [48] defined the multiplicative GARCH (MGARCH):

$$\Lambda(x) = \log x, \quad c(x) = \beta + \alpha \quad \text{and} \quad g(x) = w + \alpha \log x^2.$$

Example 6.9. The exponential GARCH model (EGARCH) was defined by Nelson [84]. Here we have

$$\Lambda(x) = \log x, \quad c(x) = \beta \quad \text{and} \quad g(x) = w + \alpha_1 x + \alpha_2 |x|.$$

6.2 Results

Statistical inference based on GARCH sequences often requires the establishment of functional central limit theorems (FCLT's) for partial sum processes like

$$S_n^f(t) = \frac{1}{n^{1/2}} \sum_{1 \leq i \leq nt} (f(y_i) - Ef(y_0)),$$

or

$$Z_n^f(t) = \frac{1}{n^{1/2}} \sum_{1 \leq i \leq nt} (f(\sigma_i) - Ef(\sigma_0)).$$

For example, to derive the asymptotic distribution of the CUSUM or MOSUM statistics in the theory of change-point detection, a functional CLT is needed. For ordinary GARCH processes, several authors obtained FCLT's for $S_n^f(t)$ with $f(x) = x$, $f(x) = |x|$ or $f(x) = x^2$ under various conditions. Berkes et al. [16] observed that in the special case $f(x) = x$ and $E\varepsilon_0 = 0$, the desired result follows from a FCLT for ergodic martingale difference sequences (cf. Billingsley [19, Theorem 23.1]) under the optimal condition $Ey_0^2 < \infty$. However, if $Ey_0 \neq 0$ or if we are e.g. interested in the squared GARCH sequence, the martingale structure does not apply. Without this special property, all existing results in the literature impose unnecessarily stringent conditions on ε_0 and y_0. For example, Hansen [52] used the NED (*near-epoch dependence*) property of GARCH (1,1) processes to derive a FCLT and later Davidson [34] used a similar approach to get an FCLT for GARCH (p,q) processes. Both authors require $Ey_0^4 < \infty$. If (y_k) is a GARCH (1,1) sequence with $E\varepsilon_0^2 < \infty$, then the y_k^2 satisfy another known weak dependence criterion, the so called $(\theta, \mathcal{L}, \psi)$ dependence. (See Nze and Doukhan [85] for details.) Assuming that $E|y_0|^\kappa < \infty$ for $\kappa > 8$, this implies a FCLT for the squared GARCH sequence (y_k^2), see [16, Theorem 2.9].

Technically, the existence of moments of y_0 is a restriction on the parameters in the GARCH model. But with the exception of $p = q = 1$ and for moments of even integer order, in the GARCH (p,q) model no explicit formula exists for the moments of y_0 in terms of the the values of α_i and β_j (cf. Ling and McAleer [74]). Also, in many important practical applications, no moments of y_0 beyond the second exist. For example, in the GARCH $(1,1)$ case it occurs frequently in practice that the estimates for the parameters α and β fall close to $\alpha + \beta = 1$, in which case we already have $Ey_0^2 = \infty$.

A possible way to weaken the moment assumptions is to verify uniform mixing properties of y_k and to utilize FCLT's for mixing sequences. For example, Carrasco and Chen [26] verified β-mixing with exponential decay for augmented GARCH sequences under assuming only finite second moments. Together with Theorem 5.1 in Nze and Doukhan [85] this implies the FCLT if $E|y_0|^{2+\delta} < \infty$. Apart from the non-expressibility of this moment condition in terms of the parameters α_i and β_j of the

GARCH process, this approach also requires the existence of a continuous positive density of ε_k on $\mathbb{R}$.

The purpose of this chapter is to prove the FCLT for augmented GARCH $(1,1)$ and GARCH (p,q) sequences under optimal conditions. Specifically, we get an FCLT for S_n^f under $Ef(y_0)^2 < \infty$, a condition which is also necessary. Our proof also provides the FCLT for $Z_n^f(t)$. Further, will give an almost optimal Berry-Esseen type bound for the rate of convergence in the ordinary CLT under finite third moments. The choice of f will comprise power functions, i.e.

$$f(x) = x^\nu \quad \text{or} \quad f(x) = |x|^\nu \quad (\nu > 0). \tag{6.14}$$

We will only consider augmented GARCH sequences with $\Lambda(x) = x^\delta \quad (\delta > 0)$ and $\Lambda(x) = \log x$. This special form of Λ is motivated by the Box-Cox transformation of the observations and it covers all the examples given in the last section. We forgo a more general setting, which is possible, but which would force us to make more restrictive moment assumptions.

Theorem 6.1. *Assume that (6.4)-(6.7) and (6.10)-(6.11) hold. Assume further that* $\Lambda(x) = x^\delta$ $(\delta > 0)$ *and that (6.14) holds with some* $\nu > 0$*. If* $Ef(y_0)^2 < \infty$*, then*

$$\tau^2 = \operatorname{Var} f(y_0) + 2 \sum_{1 \le k < \infty} \operatorname{Cov}(f(y_0), f(y_k)) \tag{6.15}$$

is convergent and

$$S_n^f(t) \xrightarrow{d} \tau W(t)$$

where $\{W(t),\ 0 \le t \le 1\}$ *is a Brownian motion.*

Here $\xrightarrow{d}$ means convergence with respect to the Skorokhod metric in the space $D[0,1]$. In case of polynomial GARCH processes the left hand side of (6.5) is always positive, and thus it is natural to assume (6.11). Note however, that $c(\varepsilon_0) \ge 0$ and $g(\varepsilon_0) \ge 0$ are not necessary in order to get $\Lambda(\sigma_k^2) \ge 0$. A non-trivial example is $g(x) = 1$ and the distribution of $c(\varepsilon_0)$ is concentrated on the interval $[-1/2, 0]$.
It follows from the definition of f and Λ that $E|f(y_0)|^2 = E|y_0|^{2\nu} < \infty$ if and only

if $E|\varepsilon_0|^{2\nu} < \infty$ and $E|\Lambda(\sigma_0^2)|^{\nu/\delta} < \infty$. Since we assume (6.10) and (6.11) we get by Theorem E that

$$E|c(\varepsilon_0)|^{\nu/\delta} < 1 \quad \text{and} \quad E|g(\varepsilon_0)|^{\nu/\delta} < \infty \tag{6.16}$$

are necessary to assure $E|\Lambda(\sigma_0^2)|^{\nu/\delta} < \infty$. In particular when the y_k are GARCH (1,1) variables, we obtain FCLT's

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{nt} (y_k - Ey_k) \xrightarrow{d} \tau_1 W(t)$$

or

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{nt} (y_k^2 - Ey_k^2) \xrightarrow{d} \tau_2 W(t) \tag{6.17}$$

(here τ_1^2, τ_2^2 are the corresponding variances arising from (6.24)) under the necessary conditions $Ey_k^2 < \infty$ respectively $Ey_k^4 < \infty$. As we noted above, a FCLT (6.17) was obtained by Berkes et al. [16, Theorem 2.9] under the condition $E|y_0|^\kappa < \infty$ with some $\kappa > 8$.

Theorem 6.2. *Assume that (6.4)-(6.7) hold. Assume further that* $\Lambda(x) = \log x$ *and that (6.14) holds with some* $\nu > 0$. *If* $E|\varepsilon_0|^{2\nu} < \infty$,

$$|c(\varepsilon_0)| \leq c < 1 \tag{6.18}$$

and if the moment generating function

$$M(t) = E\exp(t|g(\varepsilon_0)|) \quad \textit{exists on the interval } [0, \mu], \tag{6.19}$$

with some $\mu > \nu$ *then the proposition of Theorem 6.1 holds.*

It follows from Theorem D that (6.18) and (6.19) imply the existence of a unique strictly stationary solution of (6.4) and (6.5), where $\Lambda(\sigma_k^2)$ is given by (6.10). In order to connect the assumption of Theorem 6.2 with specific moment conditions for the GARCH variables we formulate the following result.

Proposition 6.1. *Assume that (6.4)-(6.7) hold and that* $\Lambda(x) = \log x$. *If* $E|\varepsilon_0|^{2\mu} < \infty$ *and (6.18)-(6.19) are satisfied, then*

$$E|y_0|^{2\mu} < \infty. \tag{6.20}$$

On the other hand if (6.10) holds and if $g(\varepsilon_0) \geq 0$, $c(\varepsilon_0) \geq 0$ *then (6.19),*

$$P\{c(\varepsilon_0) \leq 1\} = 1 \tag{6.21}$$

and $E|\varepsilon_0|^{2\mu} < \infty$ *are necessary to assure (6.20).*

Proof. By the definition of Λ we have

$$E|y_0|^{2\mu} = E|\varepsilon_0|^{2\mu} E\exp(\mu\Lambda(\sigma_0^2)).$$

Form (6.18) we get

$$\Lambda(\sigma_0^2) \leq \sum_{1 \leq i < \infty} c^{i-1}|g(\varepsilon_{-i})|.$$

This shows that

$$\begin{aligned} E\exp(\mu\Lambda(\sigma_0^2)) &\leq \prod_{1 \leq i < \infty} E\exp(\mu c^{i-1}|g(\varepsilon_{-i})|) \\ &\leq i_0 E\exp(\mu|g(\varepsilon_{-i})|) + \prod_{i_0 < i < \infty} E\exp(\mu c^{i-1}|g(\varepsilon_{-i})|), \end{aligned} \tag{6.22}$$

where the individual terms in the product above are finite by (6.19). If $t < \mu/2$ we get from Taylor's formula and the Cauchy Schwarz inequality that

$$\begin{aligned} E\exp(t|g(\varepsilon_0)|) &\leq 1 + tE\big(|g(\varepsilon_0)|\exp(\mu/2|g(\varepsilon_0)|)\big) \\ &\leq 1 + tE(|g(\varepsilon_0)|^2)^{1/2}\big(E\exp(\mu|g(\varepsilon_0)|\big)^{1/2} \\ &= 1 + tA, \end{aligned}$$

with some positive constant A. Now choose i_0 such that $\mu c^{i_0} \leq \mu/2$. Then it follows that

$$\begin{aligned} &\prod_{i_0 < i < \infty} E\exp(\mu c^{i-1}|g(\varepsilon_{-i})|) \\ &\qquad \leq \prod_{0 \leq i < \infty} E\exp(\frac{\mu}{2}\, c^i|g(\varepsilon_0)|) \leq \exp\left(\sum_{i=0}^{\infty} \log(1 + A\frac{\mu}{2}\, c^i)\right) < \infty. \end{aligned}$$

To prove the other direction we first note that without $E|\varepsilon_0|^{2\nu} < \infty$ (6.20) trivially fails. By (6.10) and the nonnegativity of $c(\varepsilon_0)$ and $g(\varepsilon_0)$ it follows that

$$E|y_0|^{2\mu} = E|\varepsilon_0|^{2\mu} E\exp(\mu\Lambda(\sigma_0^2)) \geq E|\varepsilon_0|^{2\mu} E\exp(\mu g(\varepsilon_0)), \tag{6.23}$$

and hence (6.19) is necessary. Further it is clear from (6.23) that $\Lambda(\sigma_0^2)$ must have moments of all orders. If we assume that (6.21) does not hold, then there is some $\delta > 0$ and some $\alpha > 0$ such that $P\{c(\varepsilon_0) \geq 1 + \delta\} = \alpha$. Consequently $Ec(\varepsilon_0)^p \geq \alpha(1+\delta)^p > 1$ for sufficiently large p. By (6.10) and the assumption $c(\varepsilon_0), g(\varepsilon_0) \geq 0$ we have

$$E\Lambda(\sigma_0^2)^p \geq \sum_{i=1}^{\infty} Eg(\varepsilon_0)^p \big(Ec(\varepsilon_0)^p\big)^{i-1} = \infty$$

which contradicts to $E\Lambda(\sigma_0^2)^p < \infty$ for any $p > 0$. $\square$

Theorem 6.3. *Assume that* (6.3), (6.6) *hold. Assume further that* (6.1) *and* (6.2) *have a strictly stationary solution with* $Ey_0^2 < \infty$. *In this case the sum*

$$\tau^2 = \operatorname{Var} y_0 + 2 \sum_{1 \leq k < \infty} \operatorname{Cov}(y_0, y_k) \tag{6.24}$$

is convergent and

$$S_n(t) \xrightarrow{d} \tau W(t)$$

where $\{W(t),\ 0 \leq t \leq 1\}$ *is a Brownian motion.*

If follows from Bollerslev [21] and Bougerol and Picard [23, p. 122] that a strictly stationary solution satisfying $Ey_0^2 < \infty$ exists if and only if

$$E\varepsilon_0^2 < \infty \tag{6.25}$$

and

$$(\alpha_1 + \cdots + \alpha_p)E\varepsilon_0^2 + \beta_1 + \cdots + \beta_q < 1. \tag{6.26}$$

Moreover, in this case this stationary solution is unique, ergodic and nonanticipating.

Having established the central limit theorem for the partial sums of various GARCH models, it is natural to ask for the normal approximation error. We will obtain the rate of convergence of

$$S_n(f) = f(y_1) + \ldots + f(y_n) - nEf(y_1)$$

to the normal distribution provided that $m_3(f) = E|f(y_1)|^3 < \infty$. For notational convenience we will write S_n instead of $S_n(f)$ and we set $B_n^2 = B_n(f)^2 = \operatorname{var} S_n(f)$ and

$$\sigma_n^2 = \sigma_n^2(f) = \operatorname{var} f(y_0) + 2\sum_{j=1}^{n-1}(1 - j/n)\operatorname{cov}(f(y_1), f(y_{j+1})),$$

i.e. $\sigma_n^2 = B_n^2/n$.

Theorem 6.4. *Assume that (6.4)-(6.7) hold and that $f(x)$ is defined as in (6.14). If we assume that*
(I) $\Lambda(x) = x^\delta$, (6.10)-(6.11) hold and $E|f(y_0)|^3 < \infty$ or if
(II) $\Lambda(x) = \log x$, $E|\varepsilon_0|^{3\nu} < \infty$, (6.18) holds and (6.19) is valid on the interval $[0, 3\mu/2]$ with some $\mu > \nu$, then

$$\lim \sigma_n =: \sigma \quad \textit{exists} \tag{6.27}$$

and if $\sigma > 0$ then there is some $C > 0$ such that

$$|P\{S_n < xB_n\} - \Phi(x)| \le C\frac{(\log n)^2}{\sqrt{n}} \quad \textit{for all } n \ge 2 \textit{ and } x \in \mathbb{R}.$$

The constant C may depend on f, Λ, c, g and the law of ε_0.

The additional assumptions in Theorem 6.4 compared with Theorems 6.1–6.2 arise from the requirement that $E|f(y_1)|^3 < \infty$ which is the classical assumption in the context of Berry-Esseen bounds. The existence of the limit in (6.27) follows from Theorems 6.1–6.2. The rate $(\log n)^2 n^{-1/2}$ coincides with that obtained by Tihomirov [108, Theorem 2] for sequences which are β-mixing with geometric rate. The proof of Theorem 6.4 is based on a Berry-Esseen bound for m-dependent sequences also due to Tihomirov [108, Theorem 5]. Again it becomes clear that m-dependence rather than β-mixing is the crucial structural property required in order to study the asymptotics of augmented GARCH variables.

6.3 Applications

In this section we shall give three important applications of our results.

Example 6.10. The CUSUM (cumulative sum) statistics defined by

$$C_n = \max_{1\le k\le n}\left|\sum_{1\le i\le k} y_i - \frac{k}{n}\sum_{1\le i\le n} y_i\right|$$

is one of the most often used statistics to test for the stability of $\{y_i,\, 1\le i\le n\}$. Under the assumptions of Theorems 6.1– 6.3

$$\frac{C_n}{\tau n^{1/2}} \xrightarrow{\mathcal{D}} \sup_{0\le t\le 1}|B(t)|, \tag{6.28}$$

where $\{B(t),\ 0\le t\le 1\}$ is a Brownian bridge and τ is defined by (6.24). The MOSUM (moving sum) version of C_n is

$$M_n = \max_{nh<k\le n}\left|\sum_{k-nh\le i\le k} y_i - h\sum_{1\le i\le n} y_i\right|,$$

where $0<h<1$. Under the conditions of Theorems 6.1– 6.3

$$\frac{M_n}{\tau n^{1/2}} \xrightarrow{\mathcal{D}} \sup_{h\le t\le 1}|B(t)-B(t-h)|. \tag{6.29}$$

In order to use (6.28) and (6.29) we need to estimate τ. One could use, for example, the Bartlett estimator (cf. Giraitis et al. [49]).

For a review on CUSUM and MOSUM we refer to Csörgö and Horváth [32]. Zeilis et al. [112] provides a comparison between CUSUM and MOSUM.

Example 6.11. CUSUM as well as MOSUM require the estimation of τ. The estimation of τ is not needed, however, if ratio based statistics are used. Following Kim [66] and Taylor [106] we define

$$K_n(t) = \left\{\frac{1}{(n(1-t))^{1/2}}\max_{nt\le k\le n}\left|\sum_{nt<i\le k} y_i - \frac{k-nt}{n(1-t)}\sum_{nt\le i\le n} y_i\right|\right\}$$
$$\Big/\left\{\frac{1}{(nt)^{1/2}}\max_{1\le k\le nt}\left|\sum_{1\le i\le k} y_i - \frac{k}{nt}\sum_{1\le i\le nt} y_i\right|\right\}.$$

If the conditions of Theorem 6.1– 6.3 are satisfied, then for any $0<\delta<1/2$

$$K_n(t) \xrightarrow{\mathcal{D}[\delta,1-\delta]} \left(\frac{t}{1-t}\right)^{1/2} \sup_{t\le s\le 1}\left|W(s)-W(t)-\frac{s-t}{1-t}(W(1)-W(t))\right|$$
$$\Big/ \sup_{0\le s\le t}\left|W(s)-\frac{s}{t}W(t)\right|.$$

Functionals of $K_n(t)$ can be considered as ratio based versions of the Kwiatowski et al. [68] test.

Example 6.12. Starting with $x_0 = 0$ we define

$$x_k = \varrho x_{k-1} + y_k \quad 1 \le k < \infty.$$

We assume that $E\varepsilon_0 = 0$. The least square estimator for ϱ is given by

$$\widehat{\varrho}_n = \sum_{1\le k\le n} x_k x_{k-1} \Big/ \sum_{1\le k\le n} x_{k-1}^2.$$

If $\varrho = 1$ (unit root), then under the conditions of Theorem 6.1– 6.3

$$n(\widehat{\varrho}_n - 1) \xrightarrow{\mathcal{D}} \left(\int_0^1 W(s)dW(s) \right) \Big/ \int_0^1 W^2(s)\, ds. \tag{6.30}$$

The result in (6.30) is the asymptotics for the Dickey–Fuller test with augmented GARCH or GARCH (p, q) errors. The same result was obtained by Ling Li and McAleer [73] assuming GARCH $(1, 1)$ errors, $E\varepsilon_0^4 < \infty$ and the existence of a symmetric density of ε_0.

6.4 Proofs

Since

$$\sigma_k^2 = \Lambda^{-1}\left(\sum_{1\le i<\infty} g(\varepsilon_{k-i}) \prod_{1\le j<i} c(\varepsilon_{k-j}) \right)$$

should always be a non-negative real number, we assume either that $g(\varepsilon_0)$, $c(\varepsilon_0) \ge 0$ and $\Lambda^{-1} : \mathbb{R}^+ \to \mathbb{R}^+$ or that $\Lambda^{-1} : \mathbb{R} \to \mathbb{R}^+$. We may define now

$$y_{km} = \varepsilon_k \sigma_{km}, \tag{6.31}$$

where σ_{km}^2 is the solution of

$$\Lambda(\sigma_{km}^2) = \sum_{1\le i\le m} g(\varepsilon_{k-i}) \prod_{1\le j<i} c(\varepsilon_{k-j}). \tag{6.32}$$

It follows that y_{km} defines an m-dependent sequence. Setting $\eta_k = f(y_k) - Ef(y_0)$ and $\eta_{km} = f(y_{km}) - Ef(y_0)$ we get

$$\sum_{m=1}^{\infty} \|\eta_0 - \eta_{0m}\|_2 < \infty. \tag{6.33}$$

Hence Theorem 6.1 and Theorem 6.2 will follow immediately from Billingsley [19, Theorem 21.1]. The same approach we will use for the proof of Theorem 6.3. The difficulty we were facing is that for GARCH (p, q) models there seemed to exist only a series representation in terms of matrices (cf. Bougerol and Picard [23]) and no explicit recursion formula as we have at the augmented GARCH $(1, 1)$ model. Employing the matrix representation we were not able to obtain (6.33) under assuming only $Ey_0^2 < \infty$. However, we show that it is possible to give an explicit recursion formula for the GARCH (p, q) variables as well and via this representation we will obtain (6.33) assuming only the existence of second moments of y_k. It is also worth noting that our approach is similar to the NED approach by Hansen [52] or Davidson [34] for ordinary GARCH sequences. Here the approximating functions η_{0m} are chosen to be conditional expectations

$$\eta_{0m} = E[y_0|\sigma(\varepsilon_0, \varepsilon_{-1}, \ldots, \varepsilon_{-m})].$$

In order to prove Theorem 6.4 we use a result due to Tihomirov [108, Theorem 5.], which provides a Berry-Esseen bound for m-dependent sequences.

6.4.1 Perturbation error

Lemma 6.1. *Assume that the conditions of Theorem 6.1 hold. Set*

$$\eta_k = f(y_k) - Ef(y_0) \quad \textit{and} \quad \eta_{km} = f(y_{km}) - Ef(y_0). \tag{6.34}$$

Then there are constants $C_1 > 0$ and $0 < \varrho < 1$ such that

$$E|\eta_k - \eta_{km}|^2 \le C_1 \varrho^m \quad (m \ge 1).$$

Proof. We assume without loss of generality that $k=0$. Remember that $\Lambda(x)=x^\delta$ and $f(x)=x^\nu$. By our assumption (6.11) $\Lambda(\sigma_{km}^2)$ is non-negative and thus we have

$$\begin{aligned}
|\eta_0-\eta_{0m}|^2 &= |\varepsilon_0|^{2\nu}\Big|\Big(\Lambda^{-1}\circ\Lambda(\sigma_0^2)\Big)^{\nu/2}-\Big(\Lambda^{-1}\circ\Lambda(\sigma_{0m}^2)\Big)^{\nu/2}\Big|^2 \\
&\le |\varepsilon_0|^{2\nu}\Big|\Big(\Lambda^{-1}\circ\Lambda(\sigma_0^2)\Big)^{\nu}-\Big(\Lambda^{-1}\circ\Lambda(\sigma_{0m}^2)\Big)^{\nu}\Big| &(6.35)\\
&= |\varepsilon_0|^{2\nu}\Big|\Lambda(\sigma_0^2)^{\nu/\delta}-\Lambda(\sigma_{0m}^2)^{\nu/\delta}\Big|. &(6.36)
\end{aligned}$$

Let us first consider the case $\nu/\delta\le 1$. From (6.10) and from Minkowski's inequality (Hardy et al. [53]) we infer

$$\begin{aligned}
\big|\Lambda(\sigma_0^2)^{\nu/\delta}-\Lambda(\sigma_{0m}^2)^{\nu/\delta}\big| &\le \left(\sum_{i=m+1}^{\infty} g(\varepsilon_{-i})\prod_{1\le j<i} c(\varepsilon_{-j})\right)^{\nu/\delta}\\
&\le \sum_{i=m+1}^{\infty} g(\varepsilon_{-i})^{\nu/\delta}\prod_{1\le j<i} c(\varepsilon_{-j})^{\nu/\delta}.
\end{aligned}$$

Hence it follows from (6.25) that

$$E|\eta_0-\eta_{0m}|^2 \le E|\varepsilon_0|^{2\nu}\sum_{i=m+1}^{\infty} Eg(\varepsilon_{-i})^{\nu/\delta}\prod_{1\le j<i} Ec(\varepsilon_{-j})^{\nu/\delta} \le c_1\varrho_1^m,$$

with some constant $c_1>0$ and $\varrho_1=Ec(\varepsilon_0)^{\nu/\delta}<1$.

If $\nu/\delta>1$ then by the mean value theorem (6.36) is bounded by

$$|\varepsilon_0|^{2\nu}\frac{\nu}{\delta}\big|\Lambda(\sigma_0^2)\big|^{\nu/\delta-1}\big|\Lambda(\sigma_0^2)-\Lambda(\sigma_{0m}^2)\big|.$$

From this we get by the Hölder and the Minkowski inequality that

$$\begin{aligned}
&E|\eta_0-\eta_{0m}|^2\\
&\le \frac{\nu}{\delta}\,E|\varepsilon_0|^{2\nu}\Big(E|\Lambda(\sigma_0^2)|^{\nu/\delta}\Big)^{\frac{\nu-\delta}{\nu}}\left(E\left(\sum_{i=m+1}^{\infty} g(\varepsilon_{-i})\prod_{1\le j<i} c(\varepsilon_{-j})\right)^{\nu/\delta}\right)^{\delta/\nu}\\
&\le c_2\sum_{i=m+1}^{\infty}\Big(Eg(\varepsilon_{-i})^{\nu/\delta}\prod_{1\le j<i} Ec(\varepsilon_{-j})^{\nu/\delta}\Big)^{\delta/\nu} \le c_3(\varrho_1^{\delta/\nu})^m.
\end{aligned}$$

□

Lemma 6.2. *Assume that the conditions of Theorem 6.2 hold and let η_k and η_{km} be defined as in (6.34). Then there is some $C_2 > 0$ such that*

$$E|\eta_k - \eta_{km}|^2 \le C_2 c^m \quad (m \ge 1). \tag{6.37}$$

Proof. Assume again that $k = 0$. We can formally derive (6.35) which reads here as

$$|\eta_0 - \eta_{0m}|^2 \le |\varepsilon_0|^{2\nu} |e^{\nu \log \sigma_0^2} - e^{\nu \log \sigma_{0m}^2}|. \tag{6.38}$$

By the mean value theorem we get

$$|e^{\nu \log \sigma_0^2} - e^{\nu \log \sigma_{0m}^2}| \le \nu(e^{\nu \log \sigma_0^2} + e^{\nu \log \sigma_{0m}^2}) |\log \sigma_0^2 - \log \sigma_{0m}^2|. \tag{6.39}$$

Since $\nu < \mu$ we can find some $\zeta > 0$ such that $\nu(1+\zeta) < \mu$. It follows from Proposition 6.1 that

$$Ee^{\nu(1+\zeta)\log \sigma_0^2} < \infty,$$

and similarly we get

$$Ee^{\nu(1+\zeta)\log \sigma_{0m}^2} < \infty.$$

Thus the Hölder inequality and the Minkowski inequality give

$$\begin{aligned}
&Ee^{\nu \log \sigma_0^2} |\log \sigma_0^2 - \log \sigma_{0m}^2| \\
&\quad \le \left(Ee^{\nu(1+\zeta)\log \sigma_0^2} \right)^{1/(1+\zeta)} \left(E|\log \sigma_0^2 - \log \sigma_{0m}^2|^{(1+\zeta)/\zeta} \right)^{\zeta/(1+\zeta)} \\
&\quad \le \left(Ee^{\nu(1+\zeta)\log \sigma_0^2} \right)^{1/(1+\zeta)} \left(E|g(\varepsilon_0)|^{(1+\zeta)/\zeta} \right)^{\zeta/(1+\zeta)} \sum_{i=m+1}^{\infty} c^{i-1} \qquad (6.40) \\
&\quad \le c_1 c^m,
\end{aligned}$$

where we used (6.18) to get (6.40). The analogue result can be obtained if we replace $\exp(\nu \log \sigma_0^2)$ by $\exp(\nu \log \sigma_{0m}^2)$. Consequently we get from (6.38) and (6.39) relation (6.37). □

As we mentioned in the introduction for the proof Theorem 6.3 we need an explicit series representation of for the GARCH (p, q) variable. This can be obtained in following way. We can assume without loss of generality that $p = q$. Applying (6.1) and (6.2) we get that

$$\sigma_k^2 = \omega + \sum_{1 \le i \le p} \big(\beta_i + \alpha_i \varepsilon_{k-i}^2\big)\sigma_{k-i}^2. \tag{6.41}$$

Repeated application of (6.41) yields

$$\begin{aligned}\sigma_k^2 &= \omega + \omega \sum_{1 \le i_1 \le p} \big(\beta_{i_1} + \alpha_{i_1}\varepsilon_{k-i_1}^2\big) \\ &\quad + \sum_{1 \le i_1, i_2 \le p} \big(\beta_{i_1} + \alpha_{i_1}\varepsilon_{k-i_1}^2\big)\big(\beta_{i_2} + \alpha_{i_2}\varepsilon_{k-i_1-i_2}^2\big)\sigma_{k-i_1-i_2}^2\end{aligned}$$

and more generally for $n \ge 2$

$$\begin{aligned}\sigma_k^2 &= \omega\left(1 + \sum_{m=1}^{n-1} \sum_{1 \le i_1, \ldots, i_m \le p} \prod_{j=1}^{m} \big(\beta_{i_j} + \alpha_{i_j}\varepsilon_{k-i_1-i_2\cdots-i_j}^2\big)\right) \\ &\quad + \sum_{1 \le i_1, \ldots, i_n \le p} \prod_{j=1}^{n} \big(\beta_{i_j} + \alpha_{i_j}\varepsilon_{k-i_1-i_2\cdots-i_j}^2\big)\sigma_{k-i_1-i_2-\cdots-i_j}^2.\end{aligned} \tag{6.42}$$

This suggests as solution of (6.1) and (6.2)

$$\sigma_k^2 = \omega\left(1 + \sum_{1 \le m < \infty} \sum_{1 \le i_1, \ldots, i_m \le p} \prod_{j=1}^{m} \big(\beta_{i_j} + \alpha_{i_j}\varepsilon_{k-i_1-i_2\cdots-i_j}^2\big)\right). \tag{6.43}$$

Since

$$E\left(\sum_{1 \le i_1, \ldots, i_m \le p} \prod_{j=1}^{m} \big(\beta_{i_j} + \alpha_{i_j}\varepsilon_{k-i_1-i_2-\cdots-i_j}^2\big)\right) = \gamma^m$$

with $0 \le \gamma = (\alpha_1 + \cdots + \alpha_p)E\varepsilon_0^2 + \beta_1 + \cdots + \beta_q$, we get by (6.26) that the random variable on the right hand side of (6.43) is finite with probability one. It is easy to see that the right hand side of (6.43) defines a stationary and ergodic solution of the GARCH equations (6.1) and (6.2) and by Bougerol and Picard [23, Theorem 1.3] this solution is unique. (It is simple to prove uniqueness anew by using (6.43).)

Every positive integer m can be written as

$$m = pK + r, \tag{6.44}$$

where K, r are integers satisfying $K \geq 0$ and $0 \leq r \leq p-1$. Next we define

$$\sigma_{km}^2 = \omega\Bigg(1 + \sum_{1\leq l\leq K}\ \sum_{1\leq i_1,\ldots,i_l\leq p}\ \prod_{1\leq j\leq l}\big(\beta_{i_j} + \alpha_{i_j}\varepsilon^2_{k-i_1-i_2-\cdots-i_j}\big)\Bigg)$$

where K is defined in (6.44). (We define $\sigma^2_{km} = \omega$ if $K = 0$, i.e. $0 \leq m \leq p-1$.)

Lemma 6.3. *Assume that the conditions of Theorem 6.3 hold and set $\eta_k = y_k - Ey_k$ and $\eta_{km} = \varepsilon_k\sigma_{km} - Ey_k$. Then there is a constant C_3 such that*

$$E|\eta_k - \eta_{km}|^2 \leq C_3\gamma^m \quad (m \geq 1),$$

with $\gamma = (\alpha_1 + \cdots + \alpha_p)E\varepsilon_0^2 + \beta_1 + \cdots + \beta_q$.

Proof. Again we may assume that $k = 0$. We have

$$\begin{aligned}|\eta_0 - \eta_{0m}| &= |\varepsilon_0|\,|\sigma_0 - \sigma_{0m}| \leq |\varepsilon_0|\,(\sigma_0^2 - \sigma_{0m}^2)^{1/2} \leq \\ &\leq \omega^{1/2}|\varepsilon_0|\Bigg(\sum_{K+1\leq l<\infty}\ \sum_{1\leq i_1,\ldots,i_l\leq p}\ \prod_{1\leq j\leq l}\big(\beta_{i_j} + \alpha_{i_j}\varepsilon^2_{-i_1-i_2-\cdots-i_j}\big)\Bigg)^{1/2}.\end{aligned}$$

By (6.6) and the independence of ε_0 and σ_{0m} it follows that

$$\begin{aligned}&E(\eta_0 - \eta_{0m})^2 \\ &\leq \omega E\varepsilon_0^2 E\Bigg(\sum_{K+1\leq l<\infty}\ \sum_{1\leq i_1,\ldots,i_l\leq p}\ \prod_{1\leq j\leq l}\big(\beta_{i_j} + \alpha_{i_j}\varepsilon^2_{-i_1-i_2-\cdots-i_j}\big)\Bigg) \\ &= \frac{\omega}{1-\gamma}\gamma^{K+1}E\varepsilon_0^2.\end{aligned}$$

□

6.4.2 Proof of Theorems 6.1 – 6.4

According to Billingsley [19, Theorem 21.1], for the proof of Theorems 6.1– 6.3 it is enough to find measurable mappings g_m from $\mathbb{R}^m$ into $\mathbb{R}$ such that

$$\sum_{1\leq m<\infty}\big(E(\eta_0 - \xi_{0m})^2\big)^{1/2} < \infty, \tag{6.45}$$

where $\xi_{0m} = g_m(\varepsilon_{-m+1}, \varepsilon_{-m+2}, \ldots, \varepsilon_0)$. Clearly the r.v.'s η_{0m} defined in the lemmas of the last subsection meet this demand and thus setting $\xi_{0m} = \eta_{0m}$ shows Theorems 6.1–6.3.

Our main ingredient for the proof of Theorem 6.4 will be the following lemma due to Tihomirov [108, Theorem 5.]:

Lemma 6.4. *Let $X_1, X_2, \ldots$ be a strictly stationary sequence of m-dependent random variables with $EX_1 = 0$ such that*

$$E|X_1|^3 < \infty.$$

Let $B_n^2 = \mathrm{var} S_n$ and $\sigma^2 = EX_1^2 + 2\sum_{k=2}^m EX_1X_k$. If $\sigma > 0$ then absolute constants C_3 and C_4 exist such that

$$\sup_{x\in\mathbb{R}} |P\{S_n \le B_n x\} - \Phi(x)| \le C_3 \frac{b_m^2 E^{1/3}|X_1|^3}{\sigma^3\sqrt{n}} + C_4 \frac{m b_m E^{1/3}|X_1|^3 \log n}{\sigma^2 n},$$

where $b_m = \max_{1\le p\le m+1} E^{1/3}|\sum_{v=1}^p X_v|^3$.

Suppose that the conditions of Theorem 6.4 hold. Let η_k be the same as in (6.34). Now we want that the approximating process has expected value zero and thus we redefine η_{km} by

$$\eta_{km} = f(\varepsilon_k \sigma_{km}) - Ef(\varepsilon_k \sigma_{km}).$$

By a routine argument we can extend the proof of Lemma 6.1-6.2 to show

$$E|\eta_k - \eta_{km}|^2 \le C_5 \varrho^m \quad \text{for all } m \ge 1 \text{ and } k \in \mathbb{Z}. \tag{6.46}$$

We choose some n which we assume to be fixed for the moment and set $S_n' = \sum_{k=1}^n \eta_{km}$ and $(B_n')^2 = \mathrm{var} S_n'$ with $m = [t \log n]$. The specific value of t will be specified later. Then clearly (η_{km}) defines some strictly stationary and m-dependent sequence. Let (δ_k) be a sequence of positive reals. A simple estimate gives

$$P\{S_n > xB_n\} \le P\{S_n' > (x - \delta_n)B_n\} + P\{|S_n - S_n'| > \delta_n B_n\}$$

and

$$P\{S_n' > (x + \delta_n)B_n\} \le P\{S_n > xB_n\} + P\{|S_n - S_n'| > \delta_n B_n\}.$$

A repeated application of the triangular inequality together with the latter inequalities shows after a moments reflection that

$$\sup_{x\in\mathbb{R}} |P\{S_n \le B_n x\} - \Phi(x)| \le R_n^{(1)} + R_n^{(2)} + R_n^{(3)}, \tag{6.47}$$

where

$$\begin{aligned} R_n^{(1)} =& \sup_{x\in\mathbb{R}} |P\{S_n' \le B_n(x+\delta_n)\} - \Phi(x+\delta_n)|, \\ R_n^{(2)} =& \sup_{x\in\mathbb{R}} |\Phi(x+\delta_n) - \Phi(x)|, \\ R_n^{(3)} =& P\{|S_n - S_n'| > \delta_n B_n\}. \end{aligned}$$

In the sequel we shall estimate $R_n^{(i)}$, $i = 1, 2, 3$. We set $\delta_n = n^{-1/2}$ and get by the mean-value theorem

$$R_n^{(2)} \le n^{-1/2}/(2\pi)^{1/2}. \tag{6.48}$$

By (6.27) we conclude that $2B_n^2 > \sigma^2 n$ for all $n \ge n_0$. Since we assume that $\sigma > 0$ we infer from the Markov and the Minkowski inequality and (6.46)

$$\begin{aligned} R_n^{(3)} =& P\{|S_n - S_n'| > \delta_n B_n\} \le nE|S_n - S_n'|^2/B_n^2 \\ \le& n\left(\sum_{k=1}^n E^{1/2}|y_k - y_{km}|^2\right)^2 \Big/ B_n^2 \\ \le& c_1 n^2 \varrho^m, \end{aligned}$$

where c_1 can be chosen such that it is not depending on n. (The following constants c_i occurring in the proof are independent of n as well.) Hence if $t \ge -5/\log\varrho$ we have

$$R_n^{(3)} = O(n^{-1/2}). \tag{6.49}$$

In order to estimate $R_n^{(1)}$ we will use Theorem 6.4. For this purpose we need the difference between B_n and B_n'. By the Cauchy Schwarz and again Minkowski's inequality we derive

$$\begin{aligned} |B_n^2 - (B_n')^2| \le& E|S_n - S_n'||S_n + S_n'| \le c_2(E|S_n - S_n'|^2)^{1/2} B_n \\ &\le c_3 n^{1/2} \sum_{k=1}^n E^{1/2}|y_k - y_{km}|^2 \le c_4 n^{3/2} \varrho^{m/2} \le c_5 n^{-1}, \end{aligned}$$

where the last inequality follows from the choice $t \geq -5/\log \varrho$. Using again $2B_n^2 > \sigma^2 n$ if $n \geq n_0$ we can reformulate the latter estimate to

$$(1 + c_5 n^{-2})^{-1/2} \leq \frac{B_n}{B_n'} \leq (1 - c_5 n^{-2})^{-1/2}, \tag{6.50}$$

provided that $c_5 n^{-2} < 1$. By routine arguments (cf. [88, Lemma 5.2]) it follows that

$$\sup_{y \in \mathbb{R}} |\Phi(yp) - \Phi(y)| \leq \begin{cases} (p-1)/(2\pi e)^{1/2} & \text{if } p \geq 1 \\ (p^{-1} - 1)/(2\pi e)^{1/2} & \text{if } 0 < p < 1. \end{cases} \tag{6.51}$$

Trivially we have

$$\begin{aligned} R_n^{(1)} \leq & |P\{S_n' \leq B_n(x + \delta_n)\} - \Phi(B_n/B_n'(x + \delta_n))| \\ & + |\Phi(B_n/B_n'(x + \delta_n)) - \Phi(x + \delta_n)| =: R_n^{(11)} + R_n^{(12)}. \end{aligned}$$

From (6.50) and (6.51) we get

$$R_n^{(12)} = O(n^{-2}). \tag{6.52}$$

It remains to show that

$$R_n^{(11)} = O(n^{-1/2} \log^2 n). \tag{6.53}$$

Observe that $E|\eta_{1m}|^3 \leq c_6 E|f(y_1)|^3$ and that as consequence of (6.50) we have $\sigma' \sim \sigma$, where

$$\sigma' = \Big(\operatorname{var} \eta_{1m} + 2 \sum_{j=2}^{[t \log n]} \operatorname{cov}(\eta_{1m}, \eta_{jm})\Big)^{1/2}.$$

Hence (6.53) follows immediately from Lemma 6.4. Finally (6.47) and (6.48)-(6.49) plus (6.52)-(6.53) complete the proof of Theorem 6.4.

Chapter 7

Strong approximation of the empirical process of augmented GARCH sequences

7.1 Introduction and results

This chapter is devoted to the study of the empirical process of augmented GARCH sequences. For this purpose we define

$$R(s,t) = \sum_{1 \le k \le t} (I\{y_k \le s\} - P\{y_k \le s\}).$$

We will derive an almost sure approximation theorem for $R(s,t)$ by a two-parameter Gaussian process $K(s,t)$ under assuming only the existence of logarithmic moments of the ε_k and $c(\varepsilon_k)$, $g(\varepsilon_k)$. Our result extends Theorem 1.1 in Berkes and Horváth [15] to augmented GARCH sequences. In particular, our result will yield a weak invariance principle for the empirical process

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} (I\{y_k \le s\} - P\{y_k \le s\}) \xrightarrow{d} B(s),$$

where $B(s)$ is a Gaussian process whose covariance structure will be given below. In order to formulate our main theorem, we introduce some regularity conditions which we refer to in the sequel.

$$
\begin{array}{ll}
(i) & E(\log^+|g(\varepsilon_0)|)^\mu < \infty; \\
(ii) & E(\log^+|c(\varepsilon_0)|)^\mu < \infty; \\
(iii) & E(\log^+|\varepsilon_0|)^{(\mu-2)/2} < \infty; \\
(iv) & E|g(\varepsilon_0)|^\mu < \infty \quad \text{and} \quad E|c(\varepsilon_0)|^\mu < 1;
\end{array}
$$

We denote by $\Lambda(\tilde{\sigma}_k^2)$ the truncated expansion of (6.10) at $[k^\rho]$ where $0 < \rho < 1$ (see (7.5)) and we assume

$$
(v) \qquad \Lambda(\tilde{\sigma}_k^2) \geq \omega \quad \text{with some} \quad \omega > 0;
$$

Set $h(x) = \sqrt{\Lambda^{-1}(x)}$. By (6.7) this definition makes sense. The last condition we give imposes that h' exists and there are constants C, γ such that

$$
(vi) \qquad \left|h'(x)\right| \leq C|x|^\gamma.
$$

It will be convenient to define $Y_k(s) = I\{y_k \leq s\} - F(s)$, where $F(s) = P\{y_k \leq s\}$.

Theorem 7.1. *Assume that (6.4)-(6.7) and (6.9) hold. Assume further that the distribution function $H(x) = P\{\varepsilon_0 \leq x\}$ is Lipschitz continuous of order $\theta > 0$ and*

$$
E\sigma_0^{-\theta} < \infty. \tag{7.1}
$$

Assume finally that
(I) $(i) - (iii)$ hold with some $\mu > 10$ and $(v) - (vi)$ are satisfied
or
(II) $\Lambda(x) = \log x$ and $(iii) - (iv)$ hold with some $\mu > 10$.
Then the series

$$
\Gamma(s, s') = \sum_{-\infty<k<\infty} EY_0(s)Y_k(s') \tag{7.2}
$$

converges absolutely for every choice of parameters $-\infty \leq s, s' \leq \infty$. Moreover, there exists a two-parameter Gaussian process $K(s,t)$ such that $EK(s,t) = 0$ and $EK(s,t)K(s',t') = (t \wedge t')\,\Gamma(s,s')$ and that for some $\alpha > 0$

$$
\sup_{0\leq t\leq T}\sup_{s\in\mathbb{R}} |R(s,t) - K(s,t)| = o\big(T^{1/2}(\log T)^{-\alpha}\big) \quad \text{a.s.}
$$

Since ε_0 and σ_0 are independent, we get from the Lipschitz continuity of H and (7.1) that

$$\begin{aligned}|P\{y_0 \le x_2\} - P\{y_0 \le x_1\}| &= |P\{\varepsilon_0 \le x_2/\sigma_0\} - P\{\varepsilon_0 \le x_1/\sigma_0\}| \\ &\le E|H(x_2/\sigma_0) - H(x_1/\sigma_0)| \\ &\le LE\sigma_0^{-\theta}|x_2 - x_1|^{\theta}.\end{aligned}$$

We summarize this in the following

Remark 7.1. *Under the conditions of Theorem 7.1 the distribution function $F(x) = P\{y_0 \le x\}$ is Lipschitz continuous of order θ. Consequently, $F(y_k)$ is uniformly distributed on the unit interval.*

Note that assumption (v) is satisfied in all the given examples for polynomial GARCH (Section 6.1.2). It is not requested for exponential GARCH and the proof will show that it can be omitted if γ in (vi) is non-negative. Conditions $(i) - (ii)$ with $\mu \ge 1$ as well as (iv) imply (6.8). Hence by the assumptions of Theorem 7.1 the unique nonnegative and strictly stationary solution of (6.4) and (6.5) is given by (6.10). A sufficient condition for (7.1) is $\sigma_0 \ge \delta > 0$ which is satisfied in all our examples for polynomial GARCH processes. In case of exponential GARCH the conditions

$$|c(\varepsilon_0)| \le c < 1 \quad \text{and} \quad E\exp(\theta/2\,|g(\varepsilon_0)|) < \infty$$

suffice to assure (7.1). To see this note that

$$E\sigma_0^{-\theta} \le E\exp(\theta/2\,|\Lambda(\sigma_0^2)|).$$

Hence we use the same arguments as in the proof of Proposition 6.1 to show that the last term is finite.

An immediate consequence of Theorem 7.1 is the weak convergence of the empirical process of $y_1, \ldots, y_n$. Let $B(s)$ be a Gaussian process with $EB(s) = 0$ and $EB(s)B(s') = \Gamma(s, s')$. Then

$$n^{1/2}\left(\frac{1}{n}\sum_{1\le k\le n}(I\{y_k \le s\} - F(s))\right) \xrightarrow{d} B(s) \quad (n \to \infty).$$

As we pointed out in the introduction, our approach yields sharper results than the theory of mixing. For example, Carrasco and Chen [26] verified β-mixing with exponential decay for y_k, an approach requiring the existence of a continuous positive density of ε_0 and $E\varepsilon_0 = 0$, $E\varepsilon_0^2 = 1$ and

$$|c(0)| < 1, \quad E|c(\varepsilon_0)| < 1, \quad E|g(\varepsilon_0)| < \infty. \tag{7.3}$$

Together with Theorem 2 in Philipp and Pinzur [89] this yields the proposition of Theorem 7.1. Clearly (7.3) is more restrictive than (i)-(iii). Also, our Theorem 7.1 does not require a positive and continuous density. In the literature special attention has been paid to the IGARCH (1,1) process, i.e. GARCH (1,1) with $E(\beta + \alpha\varepsilon_0^2) = 1$. For an IGARCH process (7.3) does not hold, but Theorem 7.1 applies. It is worth pointing out that the moment conditions (7.3) for $c(\varepsilon_0)$ and $g(\varepsilon_0)$ are milder than those of Theorem 7.1 in case of exponential GARCH.

7.2 Proofs

In the previous section we defined $Y_k(s) = I\{y_k \leq s\} - F(s)$, $s \in \mathbb{R}$. By Remark 7.1 $F(y_k)$ is uniformly distributed and thus it will be comfortable to define $Y_k^\star(s) := I\{F(y_k) \leq s\} - s$, $s \in [0,1]$. Note that since F is monotone we have $Y_k^\star(F(s)) = Y_k(s)$. Similarly we may define $R^\star(s,t)$, $\Gamma^\star(s,t)$ etc. Once we prove the analogue of Theorem 7.1 for $R^\star(s,t)$ this will immediately show Theorem 7.1, having the advantage that we only need to study the empirical process of uniformly distributed r.v.'s. In order to make the notation not too cumbersome we shall agree to omit the "$\star$" in the sequel and write $Y_k(s)$ when we actually mean $Y_k^\star(s)$ etc.

This section is organized as follows. In Subsection 7.2.1 we will derive several estimates on the difference between y_k and y_{km}, where y_{km} is defined as in (6.31). We use these estimates in order to obtain moment inequalities as well as probability inequalities for the empirical process.

Now the rest of the proof will be structured as the proof of Theorem 1 in Berkes and Philipp [17]. In Subsection 7.2.2 we derive upper bounds for the maximal fluctuation of $R(s,t)$ and $K(s,t)$ in some rectangle $[s_0, s_1] \times [t_0, t_1]$. Finally in Subsection 7.2.3 we construct the approximating Gaussian process on a special grid in $[0,1] \times [0,\infty)$. We will show that the processes are 'near' enough on this grid. Together with the afore derived bounds for the maximal fluctuation of $R(s,t)$ and $K(s,t)$ Theorem 7.1 will follow.

We agree upon denoting 'local' constants (within a proof) with $c_1, c_2, \cdots$ and 'global' constants (within a subsection) with capital letters $C_1, C_2, \cdots$.

7.2.1 Probability inequalities for the perturbation error

Lemma 7.1. *Assume that (6.4)-(6.7) and (6.9) hold. Assume further that (i) and (ii) hold for some $\mu > 2$. Then for any $t \geq t_0$ we have*

$$P\{|\Lambda(\sigma_0^2)| > t\} \leq C_0 (\log t)^{-(\mu-2)/2}.$$

Proof. We set $\gamma_1 = E|c(\varepsilon_0)|$ and $\gamma_2 = E\log|g(\varepsilon_0)|$. Let $0 \leq \varrho < 1$ such that $\log\varrho - \gamma_1 = a > 0$. From (6.10) and the Markov inequality we get

$$\begin{aligned}
&P\{|\Lambda(\sigma_0^2)| > t\} \\
&\leq \sum_{i=1}^{\infty} P\Big\{|g(\varepsilon_{-i})| \prod_{1\leq j<i} |c(\varepsilon_{-j})| > t(1-\varrho)\varrho^{i-1}\Big\} \\
&= \sum_{i=1}^{\infty} P\Big\{\log|g(\varepsilon_{-i})| + \sum_{1\leq j<i} (\log|c(\varepsilon_{-j})| - \gamma_1) > \log t - c_1 + i(\log\varrho - \gamma_1)\Big\} \\
&\leq \sum_{i=1}^{\infty} P\Big\{\log|g(\varepsilon_{-i})| - \gamma_2 + \sum_{1\leq j<i} (\log|c(\varepsilon_{-j})| - \gamma_1) > \frac{\log t}{2} + ia\Big\} \\
&\leq \sum_{i=1}^{\infty} E\Big|\log|g(\varepsilon_{-i})| - \gamma_2 + \sum_{1\leq j<i} (\log|c(\varepsilon_{-j})| - \gamma_1)\Big|^{\mu} \Big(\frac{\log t}{2} + ia\Big)^{-\mu}
\end{aligned}$$

if t is large enough. Hence if (i) and (ii) hold for $\mu > 2$ we get by the Rosenthal inequality [88, p. 59] and some routine analysis

$$
\begin{aligned}
P\{|\Lambda(\sigma_0^2)| > t\} &\le c_2 \sum_{i=1}^{\infty} i^{\mu/2} \left(\frac{\log t}{2} + ia\right)^{-\mu} \\
&\le c_3 \sum_{i=1}^{\infty} \left(\frac{\log t}{2} + ia\right)^{-\mu/2} \le C_0 (\log t)^{-(\mu-2)/2}.
\end{aligned}
$$

□

Lemma 7.2. *Assume that (6.4)-(6.7) and (6.9) hold. Assume further that (iii) holds for some $\mu > 0$. Then for any $t \ge 0$ we have*

$$
P\{|\Lambda(\sigma_0^2)| > t\} \le C_0 t^{-\mu}.
$$

Proof. Let $0 \le \varrho < 1$ such that $\varrho^\mu > E|c(\varepsilon_0)|^\mu$. From (6.10) and the Markov inequality we get

$$
\begin{aligned}
P\{|\Lambda(\sigma_0^2)| > t\} &\le \sum_{i=1}^{\infty} P\Big\{|g(\varepsilon_{-i})| \prod_{1 \le j < i} |c(\varepsilon_{-j})| > t(1-\varrho)\varrho^{i-1}\Big\} \\
&\le c_1 t^{-\mu} E|g(\varepsilon_0)|^\mu \sum_{i=1}^{\infty} \left(\frac{\varrho^\mu}{E|c(\varepsilon_0)|^\mu}\right)^{(-i+1)}.
\end{aligned}
$$

□

From now on we define

$$
y_k' := \varepsilon_k \tilde{\sigma}_k, \tag{7.4}
$$

where $\tilde{\sigma}_k$ is the solution of

$$
\Lambda(\tilde{\sigma}_k^2) = \sum_{i=1}^{[k^\rho]} g(\varepsilon_{k-i}) \prod_{1 \le j < i} c(\varepsilon_{k-j}) \tag{7.5}
$$

with some $0 < \rho < 1$.

Lemma 7.3. *Assume that (6.4)-(6.7) and (6.9) hold. Assume also that (i) and (ii) hold for some $\mu > 2$. Then there are positive constants C_1 and C_2 such that*

$$
P\Big\{|\Lambda(\sigma_k^2) - \Lambda(\tilde{\sigma}_k^2)| > \exp(-C_1 k^\rho)\Big\} \le C_2 k^{-\rho(\mu-2)/2}. \tag{7.6}
$$

Proof. Observe that

$$|\Lambda(\sigma_k^2) - \Lambda(\tilde{\sigma}_k^2)| = \prod_{i=1}^{[k^\rho]} |c(\varepsilon_{k-i})||\Lambda(\sigma_{k-[k^\rho]}^2)|.$$

Choose $c_1 > 0$ such that $-c_1 - \gamma_1 > 0$. From the stationarity of $\Lambda(\sigma_k^2)$ we get

$$\begin{aligned}
&P\Big\{|\Lambda(\sigma_k^2) - \Lambda(\tilde{\sigma}_k^2)| > \exp(-c_1/2\, k^\rho)\Big\} \\
&= P\left\{\prod_{i=1}^{[k^\rho]} |c(\varepsilon_{k-i})|\, |\Lambda(\sigma_{k-[k^\rho]})| > \exp(-c_1/2\, k^\rho)\right\} \\
&\le P\left\{\sum_{i=1}^{[k^\rho]} (\log|c(\varepsilon_{k-i})| - \gamma_1) > (-c_1 - \gamma_1) k^\rho\right\} + P\Big\{|\Lambda(\sigma_0)| > \exp(c_1/2\, k^\rho)\Big\}
\end{aligned}$$

and by the Markov inequality and again the Rosenthal inequality [88, p. 59] and Lemma 7.1

$$\begin{aligned}
&\le c_2 E\left|\sum_{i=1}^{[k^\rho]} (\log|c(\varepsilon_{k-i})| - \gamma_1)\right|^\mu k^{-\rho\mu} + P\Big\{|\Lambda(\sigma_0)| > \exp(c_1/2\, k^\rho)\Big\} \\
&\le c_3 k^{-\rho\mu/2} + c_4 k^{-\rho(\mu-2)/2}.
\end{aligned}$$

□

Lemma 7.4. *Assume that (6.4)-(6.7) and (6.9) hold. Assume also that we have $(i)-(iii)$ for some $\mu > 2$ and that $(v)-(vi)$ hold. Then there are constants $C_3, C_4 > 0$ such that*

$$P\Big\{|y_k - y_k'| > \exp(-C_3 k^\rho)\Big\} \le C_4 k^{-\rho(\mu-2)/2}. \tag{7.7}$$

Proof. The mean-value theorem gives

$$|y_k - y_k'| = |\varepsilon_k|\big|h(\Lambda(\sigma_k^2)) - h(\Lambda(\tilde{\sigma}_k^2))\big| = |\varepsilon_k||\Lambda(\sigma_k^2) - \Lambda(\tilde{\sigma}_k^2)|\big|h'(\xi)\big|$$

where ξ is between $\Lambda(\sigma_k^2)$ and $\Lambda(\tilde{\sigma}_k^2)$. Consequently (vi) implies

$$|y_k - y_k'| \le C|\varepsilon_k||\Lambda(\sigma_k^2) - \Lambda(\tilde{\sigma}_k^2)|\Big(|\Lambda(\sigma_k^2)|^\gamma + |\Lambda(\tilde{\sigma}_k^2)|^\gamma\Big).$$

It follows that

$$\begin{aligned}
&P\{|y_k - y_k'| > \exp(-c_1/3\, k^\rho)\}\\
&\quad \le P\Big\{2C|\varepsilon_k||\Lambda(\sigma_k^2) - \Lambda(\tilde\sigma_k^2)||\Lambda(\sigma_k^2)|^\gamma > \exp(-c_1/3\, k^\rho)\Big\}\\
&\quad + P\Big\{2C|\varepsilon_k||\Lambda(\sigma_k^2) - \Lambda(\tilde\sigma_k^2)||\Lambda(\tilde\sigma_k^2)|^\gamma > \exp(-c_1/3\, k^\rho)\Big\}.
\end{aligned} \tag{7.8}$$

Assuming that $\gamma > 0$ we have

$$\begin{aligned}
&P\Big\{2C|\varepsilon_k||\Lambda(\sigma_k^2) - \Lambda(\tilde\sigma_k^2)||\Lambda(\sigma_k^2)|^\gamma > \exp(-c_1/3\, k^\rho)\Big\}\\
&\quad \le P\Big\{\log^+ 2C|\varepsilon_k| > c_1/3\, k^\rho\Big\} + P\Big\{|\Lambda(\sigma_k^2) - \Lambda(\tilde\sigma_k^2)| > \exp(-c_1\, k^\rho)\Big\}\\
&\quad + P\Big\{|\Lambda(\sigma_k^2)| > \exp(c_1/(3\gamma)\, k^\rho)\Big\}.
\end{aligned} \tag{7.9}$$

Since $|\Lambda(\tilde\sigma_k^2)| \le |\Lambda(\sigma_k^2)| + |\Lambda(\sigma_k^2) - \Lambda(\tilde\sigma_k^2)|$ we can use similar methods to estimate (7.8). For an upper bound of the first probability in (7.9) we use the Markov inequality and (iii). Thus if $\gamma > 0$ the result follows easily from Lemma 7.1 and Lemma 7.3. If $\gamma \le 0$ we can use (v) to carry out the proof in a similar way. $\square$

Lemma 7.5. *Assume that (6.4)-(6.7) and (6.9) hold. Let* $\Lambda(x) = \log x$. *Assume further that* $(iii) - (iv)$ *hold for some* $\mu > 2$. *Then there are constants* $C_5, C_6 > 0$ *such that*

$$P\Big\{|y_k - y_k'| > \exp(-C_5 k^\rho)\Big\} \le C_6 k^{-\rho(\mu-2)/2}. \tag{7.10}$$

Proof. Observe that (iv) implies $(i) - (ii)$.

$$\begin{aligned}
|\sigma_k - \tilde\sigma_k| &= \big|\exp(1/2\, \Lambda(\sigma_k^2)) - \exp(1/2\, \Lambda(\tilde\sigma_k^2))\big|\\
&\le \big|\Lambda(\sigma_k^2) - \Lambda(\tilde\sigma_k^2)\big| \exp\Big(1/2\, (|\Lambda(\sigma_k^2)| + |\Lambda(\tilde\sigma_k^2)|)\Big)\\
&\le \big|\Lambda(\sigma_k^2) - \Lambda(\tilde\sigma_k^2)\big| \exp\Big(|\Lambda(\sigma_k^2)| + 1/2\, |\Lambda(\sigma_k^2) - \Lambda(\tilde\sigma_k^2)|\Big).
\end{aligned}$$

Thus

$$\begin{aligned}
&P\big\{|y_k - y_k'| > \exp(-c_1/2\, k^\rho)\big\}\\
&\quad \le P\big\{|\varepsilon_k| > \exp(c_1/2\, k^\rho)\big\} + P\big\{|\sigma_k - \tilde\sigma_k| > \exp(-c_1 k^\rho)\big\}\\
&\quad \le P\big\{\log^+ |\varepsilon_k| > c_1/2\, k^\rho\big\} + P\big\{|\Lambda(\sigma_k^2) - \Lambda(\tilde\sigma_k^2)| > \exp(-2c_1 k^\rho)\big\}\\
&\quad + P\big\{|\Lambda(\sigma_k^2)| + 1/2\, |\Lambda(\sigma_k^2) - \Lambda(\tilde\sigma_k^2)| > c_1 k^\rho\big\}.
\end{aligned}$$

Finally, applying the Markov inequality in connection with (iii) and (iv) and Lemma 7.2 and Lemma 7.3 finishes the proof. $\square$

Similar to $Y_n(s)$ we define

$$Y_n'(s) = I\{F(y_n') \le s\} - s.$$

Lemma 7.6. *Under the conditions of Theorem 7.1 we have*

$$|EY_0(s)Y_k(s')| \le C_7 k^{-\rho(\mu-2)/2}, \tag{7.11}$$

uniformly for $0 \le s, s' \le 1$.

Proof. Since $|Y_k(s)| \le 1$ and $Y_0(s)$ is independent from $Y_k'(s')$ if $k \ge 2$ we have

$$\begin{aligned} |EY_0(s)Y_k(s')| &= |EY_0(s)(Y_k(s') - Y_k'(s'))| \\ &\le E|Y_k(s') - Y_k'(s')| = P\{Y_k(s') \ne Y_k'(s')\}. \end{aligned}$$

The event $\{Y_k(s') \ne Y_k'(s')\}$ implies that $F(y_k)$ and $F(y_k')$ are on different sides of s'. Remember that $F(y_k)$ is uniformly distributed on the unit interval. It follows from Remark 7.1 that for some $c_1 > 0$

$$\begin{aligned} P\{Y_k(s') \ne Y_k'(s')\} \le& P\{F(y_k) \in (s' - \exp(-c_1 k^\rho), s' + \exp(-c_1 k^\rho))\} \\ &+ P\{|F(y_k) - F(y_k')| > \exp(-c_1 k^\rho)\} \\ \le& 2\exp(-c_1 k^\rho) + P\{|y_k - y_k'| > c_2 \exp(-c_1/\theta\, k^\rho)\}. \end{aligned}$$

Now the result follows from Lemmas 7.4 and 7.5. $\square$

Remark 7.2. *The previous lemma shows that the series in (7.2) converges absolutely if* $\rho(\mu-2) > 2$.

Remark 7.3. *The factor* $k^{-\rho}$ *in (7.11) stems from the definition of* $\Lambda(\tilde\sigma_k)$ *in (6.32), where we expanded the occuring series up to* $[k^\rho]$. *Assume now we go* m *steps into the past, i.e. we replace* $[k^\rho]$ *in (6.32) by* m. *An* m *step approximation of* y_k *in the spirit of (7.4) will be denoted by* y_k^{m+1} *and similar* $Y_k^{m+1}(s)$. *The same arguments as before show for* $1 \le m < k$

$$|EY_0(s)Y_k(s')| \le E|Y_k(s') - Y_k^{m+1}(s')| \le C_7 m^{-(\mu-2)/2}.$$

7.2.2 Increments of the empirical process

We define for $0 \le s \le s' \le 1$

$$\begin{aligned} \bar{Y}_k(s,s') &= I\{s < F(y_k) \le s'\} - (s'-s) \\ \bar{Y}'_k(s,s') &= I\{s < F(y'_k) \le s'\} - (s'-s). \end{aligned}$$

The next lemmas are devoted to the estimation of the increments

$$R(s',t') - R(s,t) = \sum_{1\le k\le t} \bar{Y}_k(s,s') + \sum_{t<k\le t'} Y_k(s') \qquad (t'>t). \tag{7.12}$$

Lemma 7.7. *Assume that the conditions of Theorem 7.1 are satisfied. Then for $0 \le s \le s' \le 1$ there are constants $C, \tau > 0$ such that*

$$E\left|\sum_{1\le k\le N} \bar{Y}_k(s,s')\right|^2 \le CN(s-s')^\tau,$$

where C, τ do not depend on N, s, s'.

Proof. From the stationarity of the underlying augmented GARCH process it follows that $E\bar{Y}_k(s,s')\bar{Y}_l(s,s') = E\bar{Y}_1(s,s')\bar{Y}_{l-k+1}(s,s')$. Hence one may easily derive

$$E\left|\sum_{1\le k\le N} \bar{Y}_k\right|^2 = N\left\{E\bar{Y}_1^2 + 2\sum_{k=2}^{N} E\bar{Y}_1\bar{Y}_k - \frac{2}{N}\sum_{k=2}^{N}(k-1)E\bar{Y}_1\bar{Y}_k\right\}, \tag{7.13}$$

where we shortened for notational reasons $\bar{Y}_k = \bar{Y}_k(s,s')$. The same arguments as in Lemma 7.6 show that there is a constant $c_1 > 0$ such that

$$|E\bar{Y}_1(s,s')\bar{Y}_k(s,s')| \le c_1 k^{-\rho(\mu-2)/2} \quad \forall\, 0 \le s \le s' \le 1. \tag{7.14}$$

On the other hand the Cauchy-Schwartz inequality gives

$$|E\bar{Y}_1(s,s')\bar{Y}_k(s,s')| \le EY_1^2(s,s') = (s'-s)(1-(s'-s)) \le (s'-s). \tag{7.15}$$

Putting together (7.14) and (7.15) we have some $C > 0$ with

$$|E\bar{Y}_1(s,s')\bar{Y}_k(s,s')| \le Ck^{-\rho(\mu-2)/2(1-\tau)}(s-s')^\tau. \tag{7.16}$$

Now choose $\tau > 0$ and $\rho < 1$ such that $\rho(\mu-2)(1-\tau) > 2$. Then by some standard analysis the proof follows from (7.13) and (7.16). $\square$

Lemma 7.8. *Assume that the conditions of Theorem 7.1 are satisfied. Let* $\rho < 1/2$. *There are constants* $C_1, C_2, C_3, \eta > 0$ *such that for all* $x > 1$ *and for any choice of parameters* $0 \le s \le s' \le 1$

$$P\left\{\left|\sum_{1\le k\le N} \bar{Y}_k(s,s')\right| > x\right\}$$
$$\le C_1\Big[\exp(-C_2x^2/(N(s'-s)^\eta)) + \exp(-C_3x/N^\rho) + x^{-(2+\eta)}\Big].$$

Proof. We define

$$S_N = \sum_{1\le k\le N} \bar{Y}_k(s,s') \quad \text{and} \quad S_N' = \sum_{1\le k\le N} \bar{Y}_k'(s,s').$$

Again we shorten notation and set $\bar{Y}_k = \bar{Y}_k(s,s')$ and $\bar{Y}_k' = \bar{Y}_k'(s,s')$. Then we get by the Markov and the Minkowski inequality

$$\begin{aligned} P(|S_N - S_N'| > x) &\le x^{-\kappa}\, E|S_N - S_N'|^\kappa \\ &\le x^{-\kappa}\left[\sum_{k=1}^N \Big(E|\bar{Y}_k - \bar{Y}_k'|^\kappa\Big)^{1/\kappa}\right]^\kappa. \end{aligned}$$

Note that $|\bar{Y}_k - \bar{Y}_k'| \in \{0,1\}$. Thus we get by similar arguments as in the proof of Lemma 7.6

$$E|\bar{Y}_k - \bar{Y}_k'|^\kappa = E|\bar{Y}_k - \bar{Y}_k'| = P\{\bar{Y}_k \ne \bar{Y}_k'\} \le c_1 k^{-\rho(\mu-2)/2}. \tag{7.17}$$

Since $\mu > 10$ we can choose some $\eta > 0$ and $\rho < 1/2$ such that $\rho(\mu-2)/2 > 2+\eta$. This shows that

$$P\{|S_N - S_N'| > x\} \le c_2 x^{-(2+\eta)}. \tag{7.18}$$

Observe that by definition the variables $\bar{Y}_k'$, $k = 1,\cdots,N$, are $[N^\rho]$-dependent. We define

$$Z_l = \sum_{k=2l[N^\rho]+1}^{(2l+1)[N^\rho]\wedge N} \bar{Y}_k' \qquad 0 \le l \le m,$$

where m is the smallest integer such that $2m[N^\rho] < N$. Consequently the Z_l are independent. In an analogous way we let $\hat{Z}_l$ be the sums over the sets $\{(2l+1)[N^\rho]+1,\cdots,(2l+2)[N^\rho]\}$.

We choose ρ close enough to $1/2$ in order to get $\delta = \rho(\mu-2)/4 - 1 > 0$. Then we define X_l similar to Z_l only with $\bar{Y}_k$ instead of $\bar{Y}'_k$. We get from (7.17) with some simple analysis

$$\begin{aligned}(E|Z_l - X_l|^2)^{1/2} &\le \sum_{k=2l[N^\rho]+1}^{(2l+1)[N^\rho]\wedge N} (E|\bar{Y}'_k - \bar{Y}_k|^2)^{1/2} \\ &\le \sum_{k=2l[N^\rho]+1}^{(2l+1)[N^\rho]\wedge N} c_1^{1/2} k^{-\rho(\mu-2)/4} \le c_3(2l)^{-(1+\delta)} N^{-\rho\delta}.\end{aligned}$$

Now we derive with Minkowski's inequality and Lemma 7.7

$$\begin{aligned}E|Z_l|^2 &\le \Big[\Big(E|X_l|^2\Big)^{1/2} + \Big(E|Z_l - X_l|^2\Big)^{1/2}\Big]^2 \\ &\le \Big[(CN^\rho(s'-s)^\tau)^{1/2} + c_3(2l)^{-(1+\delta)}N^{-\rho\delta}\Big]^2.\end{aligned}$$

Consequently it follows from $m \sim 1/2N^{1-\rho}$ that

$$\begin{aligned}\sum_{l=1}^m E|Z_l|^2 &\le c_4\big[N(s'-s)^\tau + N^{-2\rho\delta} + N^{\rho/2-\rho\delta}(s'-s)^{\tau/2}\big] \\ &\le c_5 N\big[(s'-s)^{\tau/2} + N^{-(1+\varepsilon)}\big], \qquad (7.19)\end{aligned}$$

where $\varepsilon = 2\rho\delta$. Obviously

$$|Z_l| \le N^\rho \quad 0 \le l \le m. \qquad (7.20)$$

Taking into account (7.19) and (7.20) we may apply Kolmogorov's exponential bound [88, Lemma 7.1] to get

$$P\Big\{\sum_{l=1}^m Z_l > x\Big\} \le \exp\Big(-c_6x^2/\big(N(s'-s)^{\tau/2} + N^{-\varepsilon}\big)\Big) + \exp(-c_7x/N^\rho).$$

An analogue inequality can be shown for $\sum_{k=1}^m \hat{Z}_l$. Since $S'_N = \sum_{l=1}^m (Z_l + \hat{Z}_l)$ we have proved that

$$P\{S'_N > x\} \le 2\exp\Big(-c_6x^2/\big(N(s'-s)^{\tau/2} + N^{-\varepsilon}\big)\Big) + 2\exp(-c_7x/N^\rho). \qquad (7.21)$$

If $N(s'-s)^{\tau/2} \ge N^{-\varepsilon}$ the result follows from (7.18). Otherwise the first term on the righthand side of (7.21) is dominated by $2\,x^{-(2+\eta)}$ which completes the proof of Lemma 7.8. □

Lemma 7.9. *Under the conditions of Theorem 7.1 we have for any* $0 \le z_0 < z \le 1$, $T > 1$ *and* $\lambda \ge \min\{(z - z_0)^{\eta/2}, (\log T)^{-1}\}$ *positive constants* C_4, C_5, α *such that*

$$P\left\{ \sup_{\substack{z_0 \le s \le z \\ 0 \le t \le T}} \left| \sum_{k \le t} \bar{Y}_k(z_0, s) \right| \ge \lambda T^{1/2} \right\}$$
$$\le C_4 \Big[\exp(-C_5 \lambda^2/(z - z_0)^\eta) + T^{-\alpha} \Big],$$

where η *stems from Lemma 7.8.*

Proof. The proof uses a dyadic chaining argument. We assume without loss of generality that $z_0 = 0$. Assume (t, s) is an element in the rectangle $X = [0, T] \times [0, z]$. Then we can represent

$$s = z \sum_{i=1}^{\infty} \epsilon_i 2^{-i} \quad \epsilon_i \in \{0, 1\} \quad \text{and define} \quad s_v = z \sum_{i=1}^{v} \epsilon_i 2^{-i},$$

and similarly let

$$t = T \sum_{i=1}^{\infty} \eta_i 2^{-i} \quad \eta_i \in \{0, 1\} \quad \text{and} \quad t_u = T \sum_{i=1}^{u} \eta_i 2^{-i}.$$

(We set $t_0 = s_0 = 0$). Observe that we have

$$(t_u, t_{u+1}] \times (s_v, s_{v+1}] \subset (Ti2^{-u}, T(i+1)2^{-u}] \times (zj2^{-v}, z(j+1)2^{-v}],$$

where $(i, j) \in \{0, \cdots, 2^u - 1\} \times \{0, \cdots, 2^v - 1\}$ depend on (t, s). Thus if

$$M_{u,v} = \max_{\substack{0 \le i \le 2^u - 1 \\ 0 \le j \le 2^v - 1}} \left| \sum_{k \in (Ti2^{-u}, T(i+1)2^{-u}]} \bar{Y}_k(zj2^{-v}, z(j+1)2^{-v}) \right|,$$

then it follows that

$$\begin{aligned}
&\left| \sum_{k \le t} \bar{Y}_k(0, s) \right| \\
&= \left| \sum_{u,v=1}^{m} \sum_{t_{u-1} < k \le t_u} \bar{Y}_k(s_v, s_{v+1}) + \sum_{t_m < k \le t} \bar{Y}_k(0, s) + \sum_{0 < k \le t_m} \bar{Y}_k(s_m, s) \right| \\
&\le \sum_{u,v=1}^{m} M_{u,v} + \left| \sum_{t_m < k \le t} \bar{Y}_k(0, s) \right| + \left| \sum_{0 < k \le t_m} \bar{Y}_k(s_m, s) \right| \\
&\le \sum_{u,v=1}^{m} M_{u,v} + \frac{T}{2^m} + \left| \sum_{0 < k \le t_m} \bar{Y}_k(s_m, s) \right|.
\end{aligned}$$

For any $x \geq 0$ we have $\bar{Y}_k(s,s') \leq \bar{Y}_k(s,s'+x)+x$ and since $s_m \leq z_m = z - z2^{-m}$ we get

$$\begin{aligned}\Big|\sum_{0<k\leq t_m} \bar{Y}_k(s_m,s)\Big| &\leq \sum_{u=1}^{m}\left(\Big|\sum_{t_{u-1}<k\leq t_u} \bar{Y}_k(s_m,s_m+z2^{-m})\Big| + \frac{t_u-t_{u-1}}{2^m}\right)\\ &\leq \sum_{u=1}^{m} M_{u,m} + \frac{T}{2^m}.\end{aligned}$$

This gives

$$\Big|\sum_{k\leq t} \bar{Y}_k(0,s)\Big| \leq 2\left(\sum_{u,v=0}^{m-1} M_{u,v} + \frac{T}{2^m}\right).$$

For $T>1$ and $\lambda>0$ we can choose an $m=m(T,\lambda)\in\mathbb{N}$ such that

$$\min\{1,\lambda/2\}2^{m-1} \leq T^{1/2} \leq \lambda 2^{m-1}. \tag{7.22}$$

Define $x_\beta := \sum_{u,v=1}^{\infty} 2^{-\beta(u+v)}$ where $\beta>0$. Then we get

$$\begin{aligned}P\Big\{\sum_{u,v=1}^{m} M_{u,v}+2^{-m}T > \lambda T^{1/2}\Big\} &\leq P\Big\{\sum_{u,v=1}^{m} M_{u,v} > \frac{\lambda}{2}T^{1/2}\Big\}\\ &\leq P\Big\{\sum_{u,v=1}^{m} M_{u,v} > \frac{\lambda}{2x_\beta}T^{1/2}\sum_{u,v=1}^{m}2^{-\beta(u+v)}\Big\}\\ &\leq \sum_{u,v=1}^{m} P\Big\{M_{u,v} > \frac{\lambda}{2x_\beta}T^{1/2}2^{-\beta(u+v)}\Big\}.\end{aligned}$$

By Lemma 7.8 we have

$$\begin{aligned}&P\Big\{M_{u,v} > \frac{\lambda}{2x_\beta}T^{1/2}2^{-\beta(u+v)}\Big\}\\ &\leq C_1\,2^{u+v}\Big[\exp(-c_1\lambda^2 2^{-2\beta(u+v)+u+v\eta}z^{-\eta})\\ &\quad+\exp(-c_2\lambda 2^{-\beta(u+v)+u\rho}T^{1/2-\rho}) + (\lambda T^{1/2}2^{-\beta(u+v)}/(2x_\beta))^{-(2+\eta)}\Big]\\ &=:C_1\,2^{u+v}[s_1(u,v)+s_2(u,v)+s_3(u,v)].\end{aligned}$$

We fix a small β. Then if $2\beta < \eta$ we have an $\epsilon_1 > 0$ such that

$$\begin{aligned}
\sum_{u,v=1}^{m} 2^{u+v} s_1(u,v) &\le \sum_{u,v=1}^{m} 2^{u+v} \exp(-c_3\lambda^2(2^{\epsilon_1 u} + 2^{\epsilon_1 v}) z^{-\eta}) \\
&\le \left(\sum_{u=1}^{m} 2^u \exp(-c_3\lambda^2 2^{\epsilon_1 u} z^{-\eta}) \right)^2 \qquad (7.23) \\
&\le c_4 \exp(-c_5\lambda^2 z^{-\eta}).
\end{aligned}$$

Here we used the relation $\lambda^2 z^{-\eta} \ge 1$ to assure that c_4, c_5 do no longer depend on λ and z.

Choose β and n such that $n(\beta - \rho) < -1$. Since $e^{-x} \le c(n)x^{-n}$ for any $x > 0$ we derive

$$\begin{aligned}
\sum_{u,v=1}^{m} 2^{u+v} s_2(u,v) &\le \sum_{u,v=1}^{m} c_6 \lambda^{-n} 2^{n(\beta(u+v)-u\rho)+u+v} T^{-n(1/2-\rho)} \\
&\le c_6 \lambda^{-n} 2^{m(\beta n+1)} T^{-n(1/2-\rho)} m \sum_{u \ge 1} 2^{u(n\beta - n\rho + 1)} \\
&\le c_7 \lambda^{-n} 2^{m(\beta n+1)} T^{-n(1/2-\rho)} m.
\end{aligned}$$

By (7.22) and $\lambda \ge (\log T)^{-1}$ we get $2^m \le 4T^{1/2}(\log T + 1)$. Choosing β and n such that $(\beta/2 + \rho - 1/2)n < -1/2$ gives $\sum_{u,v=1}^{m} 2^{u+v} s_2(u,v) \le c_8 T^{-\epsilon_2}$ for some positive ϵ_2. Using similar arguments it is now very easy to show that $\sum_{u,v=1}^{m} 2^{u+v} s_3(u,v) \le c_9 T^{-\epsilon_3}$ with some $\epsilon_3 > 0$. Subsuming our results completes the proof. $\square$

Now we proceed in estimating the increments of the approximating Gaussian process.

Lemma 7.10. *Assume the process $K(s,t)$ is defined as in Theorem 7.1. Then there is a constant $C_6 > 0$ such that for all $x \ge x_0$ and any $0 \le z_0 \le z \le 1$ and $0 \le T_0 \le T$*

$$P\left\{ \sup_{(s,t)\in I} |K(s,t) - K(z_0,T_0)| \ge x(T(z-z_0)^\tau + |T - T_0|)^{1/2} \right\} \le C_6\, e^{-C_7 x^2},$$

where $I = [z_0, z] \times [T_0, T]$ and τ stems from Lemma 7.7.

Proof. We define

$$Z(s,t) := K(z_0 + s(z - z_0), T_0 + t(T - T_0)) - K(z_0, T_0), \quad (s,t) \in [0,1]^2.$$

Then clearly $\sup_{(s,t)\in I} |K(s,t) - K(z_0,T_0)| = \sup_{(s,t)\in[0,1]^2} |Z(s,t)|$. We note that $\Gamma(s,s') = \Gamma(s',s)$ which implies

$$\begin{aligned} E|K(s,t) - K(s',t)|^2 &= t(\Gamma(s,s) + \Gamma(s',s') - 2\Gamma(s,s')) \\ &= t\sum_{k\in\mathbb{Z}} E\bar{Y}_0(s,s')\bar{Y}_k(s,s'). \end{aligned}$$

Using (7.16) we obtain from the previous relation

$$E|K(s,t) - K(s,t')|^2 \leq c_1 t|s - s'|^\tau.$$

It follows from the definition of $Z(s,t)$ that

$$E|Z(s,t) - Z(s',t)|^2 \leq c_2 T|s - s'|^\tau (z - z_0)^\tau. \tag{7.24}$$

Lemma 7.6 shows that $\Gamma(s,s')$ is uniformly bounded. Thus

$$E|Z(s,t) - Z(s,t')|^2 \leq c_3(T - T_0)|t - t'|. \tag{7.25}$$

Next observe that by the Minkowski inequality

$$E|Z(s,t)|^2 \leq \Big(E^{1/2}|Z(s,t) - Z(0,t)|^2 + E^{1/2}|Z(0,t) - Z(0,0)|^2\Big)^2.$$

Together with (7.24) and (7.25) this yields

$$\sup_{(s,t)\in[0,1]^2} E|Z(s,t)|^2 \leq c_4(T(z - z_0)^\tau + (T - T_0)). \tag{7.26}$$

Combining (7.24)-(7.26) with Lemma 2 in [70] completes the proof. □

We partition the set $[0,1]\times[0,\infty)$ in rectangles $[s_{k_i}, s_{k_{i+1}}] \times [t_k, t_{k+1}]$, with

$$s_{k_i} = i2^{-[\log k/(2\log 2)]} \quad (0 \leq i \leq d_k) \quad \text{and} \quad t_k = \exp(k^{1-\epsilon}) \quad (k \geq 1),$$

where $d_k = 2^{[\log k/(2\log 2)]}$ and ϵ is some positive constant to be specified later. Additionally we set $t_0 = 0$. Let

$$\mathcal{G} = \bigcup_{k\geq 0}\{(s_{k_i}, t_k)|0 \leq i \leq d_k\}$$

denote the grid defined by this partition. Lemma 7.11 below shows, that in order to prove Theorem 7.1 it suffices to construct a Gaussian process $K(s,t)$ with $EK(s,t) = 0$ and $EK(s,t)K(s',t') = (t \wedge t')\Gamma(s,s')$ which satisfies for some $\gamma_1 > 0$

$$\max_{0\leq i\leq d_k-1} |R(s_{k_i}, t_k) - K(s_{k_i}, t_k)| = O\Big(t_k^{1/2}(\log t_k)^{-\gamma_1}\Big) \quad \text{a.s.} \tag{7.27}$$

I.e. it suffices to show that $K(s,t)$ and $R(s,t)$ are near on the grid $\mathcal{G}$.

Lemma 7.11. *Let $\hat{R}(i,k)$ denote the maximal fluctuation of $R(s,t)$ over the rectangle $[s_{k_i}, s_{k_{i+1}}] \times [t_k, t_{k+1}]$. Similarly define for $K(s,t)$ the random variables $\hat{K}(i,k)$. Then there is a $\gamma_0 > 0$ such that*

$$\max_{0\leq i\leq d_k-1} \hat{R}(i,k) = O\Big(t_k^{1/2}(\log t_k)^{-\gamma_0}\Big) \quad \text{a.s.}$$

A similar result is true for $\hat{K}(i,k)$.

Proof. Note that by (7.12)

$$\begin{aligned}\max_{0\leq i\leq d_k-1} \hat{R}(i,k) &\leq 2 \max_{0\leq i\leq d_k-1} \sup_{\substack{t_k\leq t\leq t_{k+1}\\ s_{k_i}\leq s\leq s_{k_{i+1}}}} |R(s,t) - R(s_{k_i}, t_k)| \\ &\leq 2 \max_{0\leq i\leq d_k-1} \sup_{s_{k_i}\leq s\leq s_{k_{i+1}}} \Big|\sum_{1\leq l\leq t_k} \bar{Y}_l(s_{k_i}, s)\Big| + 2 \sup_{\substack{t_k\leq t\leq t_{k+1}\\ s\in[0,1]}} \Big|\sum_{t_k<l\leq t} \bar{Y}_l(0,s)\Big|.\end{aligned}$$

By Lemma 7.9 we get

$$\begin{aligned}&P\Big\{\max_{0\leq i\leq d_k-1} \sup_{s_{k_i}\leq s\leq s_{k_{i+1}}} \Big|\sum_{1\leq l\leq t_k} \bar{Y}_l(s_{k_i}, s)\Big| \geq t_k^{1/2}(\log t_k)^{-\eta/8}\Big\} \\ &\leq c_1 k^{1/2}\Big[\exp(-c_2(\log t_k)^{-\eta/4}k^{\eta/2}) + t_k^{-\alpha}\Big] \\ &\leq c_3 k^{1/2}\Big[\exp(-c_2 k^{\eta/4}) + \exp(-\alpha k^{1-\epsilon})\Big] \leq c_4 k^{-2},\end{aligned} \tag{7.28}$$

where we used $d_k \sim k^{1/2}$. Observing that

$$t_{k+1} - t_k \sim (1-\epsilon) t_k (\log t_k)^{-\epsilon/(1-\epsilon)} \tag{7.29}$$

and again by Lemma 7.9 we have

$$\begin{aligned}
&P\Bigg\{ \sup_{\substack{t_k \le t \le t_{k+1} \\ s\in[0,1]}} \Big| \sum_{t_k < l \le t} \bar{Y}_l(0,s) \Big| \ge t_k^{1/2} (\log t_k)^{-\epsilon/4} \Bigg\} \\
&\quad \le P\Bigg\{ \sup_{\substack{t_k \le t \le t_{k+1} \\ s\in[0,1]}} \Big| \sum_{t_k < l \le t} \bar{Y}_l(0,s) \Big| \ge (t_{k+1}-t_k)^{1/2} (\log t_k)^{\epsilon/4} \Bigg\} \\
&\quad \le c_5 \Big[\exp(-c_6 (\log t_k)^{\epsilon/2}) + (t_{k+1}-t_k)^{-\alpha} \Big] \\
&\quad \le c_7 \Big[\exp(-c_8 k^{(1-\epsilon)\epsilon/2}) + \exp(-\alpha k^{1-\epsilon}) k^{\epsilon\alpha} \Big] \le c_9 k^{-2}.
\end{aligned} \tag{7.30}$$

The Borel-Cantelli lemma implies the first proposition. Using Lemma 7.10 one can show easily a similar result for the fluctuation of $K(s,t)$. □

7.2.3 Construction of the approximating Gaussian process

We define

$$\begin{aligned}
\Delta_\ell^{(j)} &= R(s_{\ell_j}, t_{\ell+1}) - R(s_{\ell_j}, t_\ell) \qquad \ell \ge 0 \\
B_\ell^{(j)} &= R(s_{\ell_j}, t_\ell) - R(s_{\ell_m}, t_\ell) \qquad \ell \ge 1,\ m = \max\{j-1, 0\}.
\end{aligned}$$

If (s, t_k) is an element of the grid $\mathcal{G}$ we can represent $R(s, t_k)$ as sum of horizontal and vertical increments $\Delta_\ell^{(j)}$ and $B_\ell^{(j)}$. Depending on s there are constants m_ℓ and j_ℓ such that

$$R(s, t_k) = \sum_{1 \le \ell \le k} \Big(\delta_\ell B_\ell^{(j_\ell)} + \Delta_\ell^{(m_\ell)} \Big), \tag{7.31}$$

where δ_ℓ is either zero or one, depending on s as well. A similar representation holds for $K(s, t_k)$.

$$K(s, t_k) = \sum_{1 \le \ell \le k} \Big(\delta_\ell \hat{B}_\ell^{(j_\ell)} + \hat{\Delta}_\ell^{(m_\ell)} \Big), \tag{7.32}$$

where the definition of $\hat{B}_\ell^{(j_\ell)}$ and $\hat{\Delta}_\ell^{(m_\ell)}$ is obvious. Choosing ϵ in the definition of t_k smaller than $\eta/8$ we get by (7.28) and the Borel-Cantelli lemma some $\gamma_2 > 0$ such that for $k \to \infty$

$$\begin{aligned}\left|\sum_{1\le l\le k} \delta_\ell B_\ell^{(j_\ell)}\right| &\le \sum_{1\le l\le k} \max_{0\le i\le d_l-1} \sup_{s_{l_i}\le s\le s_{l_{i+1}}} \left|\sum_{1\le j\le t_l} \bar{Y}_j(s_i,s)\right| \\ &\ll \sum_{l=1}^{k} t_l^{1/2}(\log t_l)^{-\eta/8} \quad \text{a.s.} \\ &\ll t_k^{1/2}(\log t_k)^{-\gamma_2}.\end{aligned}$$

Here $a_k \ll b_k$ means $|a_k/b_k| = O(1)$. By similar arguments we get an analogous result for the process $K(s,t)$. Hence, in view of (7.27) and (7.31)-(7.32) Theorem 7.1 will be proved if we succeed in constructing the approximating Gaussian process such that for any $s = s_{k_i}$, $i = 1,\dots,d_k$ the sum of vertical increments

$$\left|\sum_{1\le\ell\le k} (\Delta_\ell^{(m_\ell)} - \hat{\Delta}_\ell^{(m_\ell)})\right|$$

is not too large. Specifically Theorem 7.1 follows from

$$\begin{aligned}\sum_{\ell=1}^{k} \max_{0\le i\le d_\ell-1} &\left|R(s_{\ell_i},t_\ell) - R(s_{\ell_i},t_{\ell-1}) - (K(s_{\ell_i},t_\ell) - K(s_{\ell_i},t_{\ell-1}))\right| \\ &\ll t_k^{1/2}(\log t_k)^{-\gamma_3} \quad \text{a.s.}\end{aligned} \tag{7.33}$$

for some $\gamma_3 > 0$ and for $k \to \infty$.

Lemma 7.12. *Let $\{\mathbf{X}_l, l \ge 1\}$ be a sequence of independent $\mathbb{R}^{d_l}$, $d_l \ge 1$, valued random variables with characteristic functions $f_l(\mathbf{u})$, $\mathbf{u} \in \mathbb{R}^{d_l}$, and let $\{G_l, l \ge 1\}$ be a sequence of probability distributions on $\mathbb{R}^{d_l}$ with characteristic functions $g_l(\mathbf{u})$, $\mathbf{u} \in \mathbb{R}^{d_l}$. Suppose that for some nonnegative numbers λ_l, δ_l and $W_l \ge 10^8 d_l$*

$$|f_l(\mathbf{u}) - g_l(\mathbf{u})| \le \lambda_l$$

for all $\mathbf{u}$ with $\|\mathbf{u}\| \le W_l$ and

$$G_l\{\mathbf{u} : \|\mathbf{u}\| > W_l/4\} \le \delta_l.$$

Then without changing its distribution we can redefine the sequence $\{\mathbf{X}_l, l \geq 1\}$ *on a richer probability space together with a sequence* $\{\mathbf{Y}_l, l \geq 1\}$ *of independent random variables such that* $\mathbf{Y}_l \stackrel{d}{=} G_l$ *and*

$$P\{\|\mathbf{X}_l - \mathbf{Y}_l\| \geq \alpha_l\} \leq \alpha_l \qquad (l \in \mathbb{N})$$

where $\alpha_1 = 1$ *and*

$$\alpha_l = 16 d_l W_l^{-1} \log W_l + 4\lambda_l^{1/2} W_l^{d_l} + \delta_l \qquad (l \geq 2).$$

Similar to (7.4) we define for $j \in \{t_{l-1}+1, \cdots, t_l\}$ the random variables $\hat{y}_j = \varepsilon_j \hat{\sigma}_j$ where $\hat{\sigma}_j$ is the solution of

$$\Lambda(\hat{\sigma}_j^2) = \sum_{1 \leq i \leq [t_l^\rho]} g(\varepsilon_{j-i}) \prod_{1 \leq m < i} c(\varepsilon_{j-m}) \quad \text{for some} \quad 0 < \rho < 1/2.$$

Next we set

$$\hat{Y}_j(s) = I\{F(\hat{y}_j) \leq s\} - P\{F(\hat{y}_j) \leq s\}$$

and for $p_{l-1} = [t_l^\rho]$ we divide the interval $I_l = \{t_{l-1} + p_{l-1} + 1, \cdots, t_l\}$ into blocks $I_{l_1}, J_{l_1}, I_{l_2}, J_{l_2} \cdots\ I_{l_n}, J_{l_n}$ where $|I_{l_k}| = [|I_l|^{\rho^*}]$, $\rho < \rho^* < 1/2$, and $|J_{l_k}| = [t_l^\rho]$. This implies that $|J_{l_k}|/|I_{l_k}| \ll t_l^{-\delta}$, for $0 < \delta < \rho^* - \rho$. The last blocks may be incomplete and of course $n = n(l)$. Then

$$\sum_{j=t_{l-1}+p_{l-1}+1}^{t_l} \hat{Y}_j(s) = \sum_{k=1}^{n} \sum_{j \in I_{l_k}} \hat{Y}_j(s) + \sum_{k=1}^{n} \sum_{j \in J_{l_k}} \hat{Y}_j(s) =: \sum_{k=1}^{n} T_{l_k}(s) + \sum_{k=1}^{n} T'_{l_k}(s).$$

We set

$$\mathbf{T}_{l_k} := \big(T_{l_k}(s_{l_0}), \cdots, T_{l_k}(s_{l_{d_l}})\big).$$

Clearly n is proportional to $|I_l|^{1-\rho^*}$ and by definition $\{\mathbf{T}_{l_1}, \cdots, \mathbf{T}_{l_n}\}$ is an $\mathbb{R}^{d_l}$ valued i.i.d. sequence with $E\mathbf{T}_{l_1} = \mathbf{0}$. Next we introduce the random vectors

$$\boldsymbol{\xi}_{l_k} = |I_{l_1}|^{-1/2}\, \mathbf{T}_{l_k} \qquad (1 \leq k \leq n).$$

Lemma 7.13. *Set* $\operatorname{var} \boldsymbol{\xi}_{l_1} = \boldsymbol{\Sigma}_l = ((\Sigma_l(s_{l_i}, s_{l_j})))_{i,j=0}^{d_l}$. *Under the conditions of Theorem 7.1 there is a constant* C_5 *such that*

$$\sup_{0 \leq i,j \leq d_l} |\Sigma_l(s_{l_i}, s_{l_j}) - \Gamma(s_{l_i}, s_{l_j})| \leq C_5 |I_{l_1}|^{-1} \qquad (l \geq 1).$$

Proof. Using the stationarity of $\{Y_k(s), k \in \mathbb{Z}\}$ little algebra shows that

$$\frac{1}{N-M} E\left(\sum_{M<k,m\leq N} Y_k(s)Y_m(s') \right) = \sum_{|k|<(N-M)} EY_0(s)Y_k(s')$$
$$-\frac{1}{N-M} \sum_{k=1}^{(N-M)-1} k(EY_0(s)Y_k(s') + EY_k(s)Y_0(s')) \qquad (M < N).$$

Hence we may write

$$\begin{aligned}\Gamma(s,s') &= \frac{1}{|I_{l_1}|} E\left(\sum_{k,m\in I_{l_1}} Y_k(s)Y_m(s') \right) \\ &\quad + \frac{1}{|I_{l_1}|} \sum_{k=1}^{|I_{l_1}|-1} k(EY_0(s)Y_k(s') + EY_k(s)Y_0(s')) + \sum_{|k|=|I_{l_1}|}^{\infty} EY_0(s)Y_k(s') \\ &= \frac{1}{|I_{l_1}|} E\left(\sum_{k,m\in I_{l_1}} Y_k(s)Y_m(s') \right) + O\big(|I_{l_1}|^{-1}\big) \quad (l \to \infty), \end{aligned} \tag{7.34}$$

where (7.34) follows from Lemma 7.6. (Note that O is uniformly in $0 \leq s, s' \leq 1$.) Consequently we have

$$\begin{aligned}&|\Sigma_l(s_{l_i}, s_{l_j}) - \Gamma(s_{l_i}, s_{l_j})| \\ &\quad \leq \frac{1}{|I_{l_1}|} \sum_{k,m\in I_{l_1}} E|\hat{Y}_k(s_{l_i})\hat{Y}_m(s_{l_j}) - Y_k(s_{l_i})Y_m(s_{l_j})| + O\big(|I_{l_1}|^{-1}\big).\end{aligned}$$

By Remark 7.3 we infer for $k, m \in I_{l_1}$

$$\begin{aligned}E|\hat{Y}_k(s)\hat{Y}_m(s') - Y_k(s)Y_m(s')| &\leq E|\hat{Y}_m(s') - Y_m(s')| + E|\hat{Y}_k(s) - Y_k(s)| \\ &\leq c_1 t_l^{-\rho(\mu-2)/2}.\end{aligned}$$

By (7.29) we get $|I_{l_1}| = O\big(t_l^{\rho^*} l^{-\epsilon\rho^*}\big)$. Since $\rho^* < 1/2$ it suffices e.g. to choose $\rho < 1/2$ such that $\rho(\mu-2)/2 \geq 1$ in order to finish the proof. □

We set $\mathbf{\Gamma}_l = ((\Gamma(s_{l_i}, s_{l_j})))_{i,j=0}^{d_l}$ and denote $\|A\|_\infty = \sup_{i,j} |a_{ij}|$ for some matrix $A = ((a_{ij}))$. Since $\Gamma(s,s')$ is a bounded function we infer by the last lemma that $\sup_l \|\mathbf{\Sigma}_l\|_\infty < \infty$.

Set

$$\mathbf{X}_l = (n-1)^{-1/2} \sum_{k=1}^{n-1} \boldsymbol{\xi}_{l_k}$$

and denote by $\langle \cdot | \cdot \rangle$ the inner product of real vectors.

Lemma 7.14. *Let $\|\mathbf{u}\| \leq K \exp(l^{1/2})$ for some absolute number K. Then there are constants C_6, C_7 such that*

$$\left| E \exp\left(i \langle \mathbf{u} | \mathbf{X}_l \rangle\right) - \exp\left(-1/2 \langle \mathbf{u} | \boldsymbol{\Gamma}_l \mathbf{u} \rangle\right) \right| \leq C_6 \exp(-C_7 l^{1-\epsilon}) \|\mathbf{u}\|^2,$$

where ϵ stems from the definition of t_l.

Proof. Let A be a $d \times d$ matrix. Some elementary estimates show that for $\mathbf{u} \in \mathbb{R}^d$

$$|\langle \mathbf{u} | A \mathbf{u} \rangle| \leq d \, \|A\|_\infty \|\mathbf{u}\|^2.$$

Thus by Lemma 7.13 and the Cauchy Schwarz inequality we derive

$$\begin{aligned} &\left| \exp\left(-1/2 \langle \mathbf{u} | \boldsymbol{\Gamma}_l \mathbf{u} \rangle\right) - \exp\left(-1/2 \langle \mathbf{u} | \boldsymbol{\Sigma}_l \mathbf{u} \rangle\right) \right| \leq \\ &\quad \leq |\langle \mathbf{u} | \boldsymbol{\Gamma}_l \mathbf{u} \rangle - \langle \mathbf{u} | \boldsymbol{\Sigma}_l \mathbf{u} \rangle| \leq C_5 |I_{l_1}|^{-1} d_l \|\mathbf{u}\|^2 \leq c_4 \exp(-c_5 l^{1-\epsilon}) \|\mathbf{u}\|^2. \end{aligned} \tag{7.35}$$

Since the the vectors $\boldsymbol{\xi}_{l_k} = (\xi_{l_k}(s_{l_0}), \ldots, \xi_{l_k}(s_{l_{d_l}}))$, $1 \leq k \leq n-1$, are i.i.d. we obtain

$$E \exp\left(i \langle \mathbf{u} | \mathbf{X}_l \rangle\right) = \left(E \exp\left(i \, n^{-1/2} \sum_{j=0}^{d_l} u_j \xi_{l_k}(s_{l_j})\right) \right)^n.$$

Some routine analysis shows $|e^{ix} - (1 + ix - x^2/2)| \leq |x|^3/6$, hence there is some complex Θ (which may depend on $\mathbf{u}$ and l) with $|\Theta| \leq 1$ such that

$$E \exp\left(i \, n^{-1/2} \sum_{j=0}^{d_l} u_j \xi_{l_k}(s_{l_j})\right) = 1 - \frac{1}{2n} \langle \mathbf{u} | \boldsymbol{\Sigma}_l \mathbf{u} \rangle + \frac{\Theta}{6 n^{3/2}} E \left| \sum_{j=0}^{d_l} u_j \xi_{l_k}(s_{l_j}) \right|^3.$$

From $|\xi_{l_k}(s)| \leq |I_{l_1}|^{1/2}$ we infer by the Cauchy Schwartz inequality

$$\begin{aligned} E \left| \sum_{j=0}^{d_l} u_j \xi_{l_k} \right|^3 &\leq \left\| \sum_{j=0}^{d_l} u_j \xi_{l_k}(s_{l_j}) \right\|_\infty E \left| \sum_{j=0}^{d_l} u_j \xi_{l_k}(s_{l_j}) \right|^2 \\ &\leq d_l^{1/2} |I_{l_1}|^{1/2} \|\mathbf{u}\| \langle \mathbf{u} | \boldsymbol{\Sigma}_l \mathbf{u} \rangle \leq c |I_l|^{\rho^*/2} d_l^{3/2} \|\mathbf{u}\|^3. \end{aligned}$$

Since $n \sim |I_l|^{1-\rho^*}$ we infer that there is some Θ' within the complex unit circle (which may depend on $\mathbf{u}$ and l) such that

$$\begin{aligned} E\exp\big(i\, n^{-1/2}\sum_{j=0}^{d_l} u_j \xi_{l_k}(s_{l_j})\big) &= 1 - \frac{1}{2n}\big\langle \mathbf{u}|\mathbf{\Sigma}_l\mathbf{u}\big\rangle + \frac{c\Theta'}{6}|I_l|^{2\rho^*-3/2} d_l^{3/2}\|\mathbf{u}\|^3 \\ &=: 1 - \frac{1}{2n}\big\langle \mathbf{u}|\mathbf{\Sigma}_l\mathbf{u}\big\rangle + r(l,\mathbf{u}). \end{aligned}$$

The relation $|(1-t)^r - e^{-rt}| \le t/2$, which holds for $0 \le t \le 1$ and every $r > 0$ implies for $\big\langle \mathbf{u}|\mathbf{\Sigma}_l\mathbf{u}\big\rangle \le 2n$

$$|\exp(-1/2\big\langle \mathbf{u}|\mathbf{\Sigma}_l\mathbf{u}\big\rangle) - (1 - \frac{1}{2n}\big\langle \mathbf{u}|\mathbf{\Sigma}_l\mathbf{u}\big\rangle)^n| \le \frac{1}{4n}\big\langle \mathbf{u}|\mathbf{\Sigma}_l\mathbf{u}\big\rangle. \tag{7.36}$$

Again assuming $\big\langle \mathbf{u}|\mathbf{\Sigma}_l\mathbf{u}\big\rangle \le 2n$ we get by some basic analysis

$$|(1 - \frac{1}{2n}\big\langle \mathbf{u}|\mathbf{\Sigma}_l\mathbf{u}\big\rangle)^n - (1 - \frac{1}{2n}\big\langle \mathbf{u}|\mathbf{\Sigma}_l\mathbf{u}\big\rangle + r(l,\mathbf{u}))^n| \le n|r(l,\mathbf{u})|. \tag{7.37}$$

(Note that the absolute value of both functions occurring on the left hand side of (7.37) are within the complex unit circle; the first by assumption and the second since it is a characteristic function.) Combining (7.35)-(7.37) with the appropriate value for n and the restriction for $\|\mathbf{u}\|$ will yield the proof. □

Let $f_l(\mathbf{u})$ be the characteristic function of $\mathbf{X}_l$ and let $g_l(\mathbf{u})$ be the characteristic function of a d_l-dimensional Gaussian vector $\mathbf{G}_l = (G_l(1), \cdots, G_l(d_l))$ with covariance matrix $\mathbf{\Gamma}$. Since $\Gamma(s,s')$ is bounded we get by choosing $W_l = \exp(c_6 l^\epsilon)$

$$\begin{aligned} P\{\|\mathbf{G}_l\| > W_l/4\} &\le P\{\max_{1\le i\le d_l} |G_l(i)| > (W_l/d_l)/4\} \\ &\le c_7 d_l \exp(-c_8(W_l/d_l)^2) \\ &\le c_9\exp(-c_{10} l^\epsilon). \end{aligned} \tag{7.38}$$

By Lemma 7.12, Lemma 7.14 and (7.38) we can redefine the sequence $\{\mathbf{X}_l\}$ on a richer probability space together with a sequence of independent Gaussian vectors $\{\mathbf{Y}_l\}$ with covariance matrix $\mathbf{\Gamma}_l$ such that

$$P\{\|\mathbf{X}_l - \mathbf{Y}_l\| \ge c_{11}\exp(-c_{12}l^\epsilon)\} \le c_{11}\exp(-c_{12}l^\epsilon).$$

Next we define

$$\mathbf{Z}_l = (t_l - t_{l-1})^{-1/2}\big(R(s_{l_i}, t_l) - R(s_{l_i}, t_{l-1})\big)_{i=0}^{d_l},$$

and

$$\mathbf{V}_l = (t_l - t_{l-1})^{-1/2}\big(K(s_{l_i}, t_l) - K(s_{l_i}, t_{l-1})\big)_{i=0}^{d_l},$$

Yet again applying similar estimates as in our consideration so far we can easily derive

$$P\{\|\mathbf{Z}_l - \mathbf{X}_l\| \geq \exp(-c_{13}l^{\epsilon})\} \leq c_{14}l^{-2}.$$

Hence by the Borel-Cantelli lemma we have constants $c_{15}, c_{16} > 0$ such that for all $l \geq l_0(\omega)$

$$\|\mathbf{Z}_l - \mathbf{Y}_l\| \leq c_{15}\exp(-c_{16}l^{\epsilon}).$$

By the definition of $\mathbf{V}_l$ we have

$$\{\mathbf{Y}_l, l \geq 1\} \stackrel{d}{=} \{\mathbf{V}_l, l \geq 1\}.$$

We can enlarge the probability space such that

$$\{\mathbf{Y}_l, l \geq 1\} = \{\mathbf{V}_l, l \geq 1\}.$$

Summing up our results shows that

$$\begin{aligned}\max_{0\leq i\leq d_l}\Big|(R(s_{l_i}, t_l) - R(s_{l_i}, t_{l-1})) - (K(s_{l_i}, t_l) - K(s_{l_i}, t_{l-1}))\Big| \\ \leq c_{17}(t_l - t_{l-1})^{1/2}\exp(-c_{18}l^{\epsilon}) \quad \text{a.s.} \quad (l \to \infty).\end{aligned}$$

This shows (7.33) and completes the proof of Theorem 7.1.

Bibliography

[1] R. G. Antonini and M. Weber. The intersective ASCLT. *Stochastic Anal. Appl.*, 22:1009–1025, 2004.

[2] M. Atlagh. Théorèm central limite presque sûr et loi du logarithme itéré pour de sommes de variables alétoires indépendantes. *C.R. Acad. Sci. Paris Sér. I Math.*, 316:929–933, 1993.

[3] M. Atlagh and M. Weber. Un théorème central limite presque sûr relatif à des sous-suites. *C. R. Acad. Sci. Paris Sér. I Math.*, 315:203–206, 1992.

[4] M. Atlagh and M. Weber. Le théorème central limite presque sûr. *Expo. Math.*, 18:97–126, 2000.

[5] A. Aue, I. Berkes, and L. Horváth. Strong approximations for the sums of squares of augmented GARCH sequences. *Bernoulli*, 12:583–608, 2006.

[6] F. Avram and M. S. Taqqu. Noncentral limit theorems and Appell polynomials. *Ann. Probab.*, 15:767–775, 1987.

[7] P. Becker-Kern. Almost sure limit theorems of mantissa type for semistable domains of attraction. *Acta Math. Hungar.* To appear.

[8] I. Berkes. Results and problems related to the pointwise central limit theorem. In *Asymptotic Methods in Probability and Statistics (Ottawa, ON, 1997)*, pages 59–96. North-Holland, Amsterdam, 1998.

[9] I. Berkes and E. Csáki. A universal result in almost sure central limit theory. *Stochastic Process. Appl.*, 94:105–134, 2001.

[10] I. Berkes, E. Csáki, and L. Horváth. Almost sure central limit theorem under minimal conditions. *Statist. Probab. Lett.*, 37:67–76, 1998.

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

[12] I. Berkes and S. Hörmann. On the equivalence of LIL and pathwise CLT behavior. Preprint.

[13] I. Berkes, S. Hörmann, and L. Horváth. The functional central limit theorem for sums of GARCH observations with applications. Submitted to Economics Letters.

[14] I. Berkes and L. Horváth. Almost sure invariance principles for logarithmic averages. *Studia Sci. Math. Hungar.*, 33:1–24, 1997.

[15] I. Berkes and L. Horváth. Strong approximation of the empirical process of GARCH sequences. *Ann. Appl. Probab.*, 11:789–809, 2001.

[16] I. Berkes, L. Horváth, and P. Kokoszka. Probabilistic and statistical properties of GARCH processes. In *Asymptotic methods in stochastics*, volume 44 of *Fields Inst. Commun.*, pages 409–429. Amer. Math. Soc., Providence, RI, 2004.

[17] I. Berkes and W. Philipp. An almost sure invariance principle for the empirical distribution function of mixing random variables. *Z. Wahrsch. Verw. Gebiete*, 41:115–137, 1977/78.

[18] I. Berkes and M. Weber. Almost sure versions of the Darling-Erdős theorem. *Stat. Prob. Lett.*, 76:280–290, 2006.

[19] P. Billingsley. *Convergence of Probability Measures.* John Wiley & Sons Inc., New York, 1968.

[20] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation*, volume 27. Encyclopedia of Mathematics and its Applications, Cambridge, 1987.

[21] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *J. Econometrics*, 31:307–327, 1986.

[22] T. Bollerslev, R. Y. Chou, and K. F. Kroner. ARCH modeling in finance. A review of the theory and empirical evidence. *J. Econom.*, 52:5–59, 1992.

[23] P. Bougerol and N. Picard. Stationarity of GARCH processes and of some nonnegative time series. *J. Econometrics*, 52:115–127, 1992.

[24] L. Breiman. On the tail behavior of sums of independent random variables. *Z. Wahrsch. Verw. Gebiete*, 9:20–25, 1967.

[25] G. A. Brosamler. An almost everywhere central limit theorem. *Math. Proc. Cambridge Philos. Soc.*, 104:561–574, 1988.

[26] M. Carrasco and X. Chen. Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, 18:17–39, 2002.

[27] K. Chandrasekharan and S. Minakshisundaram. *Typical Means.* Oxford University Press, 1952.

[28] S. Cheng, L. Peng, and Y. Qi. Almost sure convergence in extreme value theory. *Math. Nachr.*, 190, 1998.

[29] Y. S. Chow and H. Teicher. *Probability Theory.* Springer-Verlag, New York, 1978.

[30] E. Csáki and A. Földes. On the logarithmic average of additive functionals. *Statist. Probab. Lett.*, 22:261–268, 1995.

[31] M. Csörgő and P. Révész. *Strong Approximations in Probability and Statistics.* Academic Press, Budapest, 1981.

[32] M. Csörgő and L. Horváth. *Limit Theorems in Change-Point Analysis.* John Wiley, New York, 1997.

[33] D. A. Darling and P. Erdös. A limit theorem for the maximum of normalized sums of independent random variables. *Duke Math. J.*, 23:143–155, 1956.

[34] J. Davidson. Establishing conditions for the functional central limit theorem in nonlinear and semiparametric time series processes. *J. Econometrics*, 106:243–269, 2002.

[35] Z. Ding, C. W. J. Granger, and R. F. Engle. A long memory property of stock market returns and a new model. *J. Empirical Finance*, 1:83–106, 1993.

[36] M. D. Donsker. An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.*, (6), 1951.

[37] J-C. Duan. Augmented GARCH (p, q) process and its diffusion limit. *J. Econometrics*, 79:97–127, 1997.

[38] U. Einmahl and D. M. Mason. Some results on the almost sure behavior of martingales. In *Limit Theorems in Probability and Statistics (Pécs, 1989)*, volume 57 of *Colloq. Math. Soc. János Bolyai*, pages 185–195. North-Holland, Amsterdam, 1990.

[39] R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50:987–1007, 1982.

[40] R. F. Engle and V. K. Ng. Measuring and testing the impact of news on volatility. *J. Finance*, 48:1749–1778, 1993.

[41] P. Erdős. On the law of the iterated logarithm. *Ann. Math.*, 43:419–436, 1942.

[42] P. Erdös and M. Kac. On certain limit theorems of the theory of probability. *Bull. Amer. Math. Soc.*, 52:292–302, 1946.

[43] I. Fahrner and U. Stadtmüller. On almost sure max-limit theorems. *Statist. Probab. Lett.*, 37:229–236, 1998.

[44] W. Feller. The general form of the so-called law of the iterated logarithm. *Trans. Am. Math. Soc.*, 54:373–402, 1943.

[45] W. Feller. The law of the iterated logarithm for identically distributed random variables. *Ann. of Math. (2)*, 47:631–638, 1946.

[46] A. Fisher. A pathwise central limit theorem for random walks. 1989. Preprint.

[47] P. Gänssler and W. Stute. *Wahrscheinlichkeitstheorie*. Springer-Verlag, Berlin, 1977.

[48] J. Geweke. Modeling the persistence of conditional variances: a comment. *Econometric Rev.*, 5:57–61, 1986.

[49] L. Giraitis, P. Kokoszka, R. Leipus, and G. Teyssière. Rescaled variance and related tests for long memory in volatility and levels. *J. Econometrics*, 112:265–294, 2003.

[50] L. R. Glosten, J. Ravi, and D. E. Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *J. Finance*, 48:1779–1801, 1993.

[51] G. E. Hagerud. A smooth transition ARCH model for asset returns. Working Paper Series in Economics and Finance 162, 1997. `http://ideas.repec.org/p/hhs/hastef/0162.html`.

[52] B. E. Hansen. GARCH $(1,1)$ processes are near epoch dependent. *Econom. Lett.*, 36:181–186, 1991.

[53] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge University Press, Cambridge, 1978.

[54] P. Hartman and A. Wintner. On the law of the iterated logarithm. *Amer. J. Math.*, 63:169–176, 1941.

[55] S. Hörmann. Augmented GARCH sequences: dependence structure and asymptotics. Submitted to Bernoulli.

[56] S. Hörmann. Critical behavior in almost sure central limit theory. *J. Theor. Probab.* To appear.

[57] S. Hörmann. On the universal a.s. central limit theorem. *Act. Math. Hungar.* To appear.

[58] S. Hörmann. A note on the almost sure convergence of central order statistics. *Probab. Math. Stat.*, 25:317–329, 2005.

[59] S. Hörmann. Asymptotic properties of augmented GARCH (1,1) sequences. *Proceedings of Prague Stochastics Conference*, pages 407–417, 2006.

[60] S. Hörmann. An extension of the almost sure central limit theorem. *Statist. Probab. Lett.*, 76:191–202, 2006.

[61] L. Horváth and D. Khoshnevisan. Weight functions and pathwise local central limit theorems. *Stochastic Process. Appl.*, 59:105–123, 1995.

[62] I. A. Ibragimov. Some limit theorems for stationary processes. *Teor. Verojatnost. i Primenen.*, 7:361–392, 1962.

[63] I. A. Ibragimov and M. A. Lifshits. On the convergence of generalized moments in almost sure central limit theorem. *Statist. Probab. Lett*, 40:343–351, 1998.

[64] I. A. Ibragimov and M. A. Lifshits. On limit theorems of "almost sure" type. *Teor. Veroyatnost. i Primenen.*, 44:328–350, 1999.

[65] N. C. Jain, K. Jogdeo, and W. F. Stout. Upper and lower functions for martingales and mixing processes. *Ann. Probab.*, 3:119–145, 1975.

[66] J-Y. Kim. Detection of change in persistence of a linear time series. *J. Econometrics*, 95:97–116, 2000.

[67] A. N. Kolmogorov. Über das Gesetz des iterierten Logarithmus. *Math. Ann.*, 101:126–135, 1929.

[68] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Y. Shin. Testing the null hypothesis of stationarity against the alternative of a unit root : How sure are we that economic time series have a unit root? *J. Econometrics*, 54:159–178, 1992.

[69] M. T. Lacey and W. Philipp. A note on the almost sure central limit theorem. *Statist. Probab. Lett.*, 9:201–205, 1990.

[70] T. L. Lai. Reproducing kernel Hilbert spaces and the law of the iterated logarithm for Gaussian processes. *Z. Wahrsch. Verw. Gebiete*, 29:7–19, 1974.

[71] P. Lévy. *Théorie de l'addition des variables alétoires.* Jacques Gabay, 12 edition, 2003.

[72] M. Lifshits. On the difference between CLT and ASCLT. *Zapiski Seminarov POMI*, 260:186–201, 2000.

[73] S. Ling, W. K. Li, and M. McAleer. Estimation and testing for unit root processes with GARCH $(1,1)$ errors: theory and Monte Carlo evidence. *Econometric Rev.*, 22:179–202, 2003.

[74] S. Ling and M. McAleer. Necessary and sufficient moment conditions for the GARCH (r,s) and asymmetric power GARCH (r,s) models. *Econometric Theory*, 18:722–729, 2002.

[75] P. Major. Approximation of partial sums of i.i.d. r.v.s when the summands have only two moments. *Z. Wahrsch. Verw. Gebiete*, 35:221–229, 1976.

[76] P. Major. The approximation of partial sums of independent RV's. *Z. Wahrsch. Verw. Gebiete*, 35:213–220, 1976.

[77] P. Major. An improvement of Strassen's invariance principle. *Ann. Probab.*, 7:55–61, 1979.

[78] P. Major. A note on Kolmogorov's law of iterated logarithm. *Studia Sci. Math. Hungar.*, 12:161–167, 1980.

[79] J. Marcinkiewicz and A. Zygmund. Remarque sûr la loi du logarithme itéré. *Fundam. Math.*, 29:215–222, 1937.

[80] M. Marcus and J. Rosen. Logarithmic averages for the local times of recurrent random walks and Lévy processes. *Stochastic Process. Appl.*, 59:175–184, 1995.

[81] P. Mörters. Almost sure Kallianpur-Robbins laws for Brownian motion in the plane. *Probab. Theory Relat. Fields*, 118:49–64, 2000.

[82] P. Mörters. A pathwise version of Spitzer's law. In *Limit Theorems in Probability and Statistics, Vol. II (Balatonlelle, 1999)*, pages 427–436. János Bolyai Math. Soc., Budapest, 2002.

[83] D. Nelson. Stationarity and persistence in the GARCH $(1, 1)$ model. *Econometric Theory*, 6:318–334, 1990.

[84] D. Nelson. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 59:347–370, 1991.

[85] P. A. Nze and P. Doukhan. Weak dependence: models and applications to econometrics. *Econometric Theory*, 20:995–1045, 2004.

[86] M. Peligrad and P. Révész. On the almost sure central limit theorem. In *Almost Everywhere Convergence*, volume 2, pages 209–225, 1991.

[87] L. Peng and Y. Qi. Almost sure convergence of the distributional limit theorem for order statistics. *Probab. Math. Statist.*, 23:217–228, 2003.

[88] V. Petrov. *Limit Theorems of Probability Theory.* Oxford Science Publications, Oxford, 1995.

[89] W. Philipp and L. Pinzur. Almost sure approximation theorems for the multivariate empirical process. *Z. Wahrsch. Verw. Gebiete*, 54:1–13, 1980.

[90] W. Philipp and W. Stout. Almost sure invariance principles for partial sums of weakly dependent random variables. *Mem. Amer. Math. Soc. 2*, 1975.

[91] W. Philipp and W. Stout. Invariance principles for martingales and sums of independent random variables. *Math. Z.*, 192:253–264, 1986.

[92] M. Rosenblatt. Stationary processes as shifts of functions of independent random variables. *J. Math. Mech.*, 8:665–681, 1959.

[93] Y. A. Rozanov. *Stationary Random Processes.* Holden-Day, 1967.

[94] Z. Rychlik and K. Szuster. Some remarks on the almost sure central limit theorem for independent random variables. *Probab. Math. Statist.*, 23:241–249, 2003.

[95] A. I. Sakhanenko. On the convergence rate in the invariance principle almost surely. *Theor. Prob. Appl.*, 29:424, 1985.

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

[97] P. Schatte. On the central limit theorem with almost sure convergence. *Probab. Math. Statist.*, 11:237–246, 1991.

[98] G. W. Schwert. Why does stock market volatility change over time. *J. Finance*, 44:1115–1153, 1989.

[99] E. Sentana. Quadratic ARCH models. *Rev. Econom. Stud.*, 62:639–661, 1995.

[100] G. R. Shorack. Extension of the Darling and Erdős theorem on the maximum of normalized sums. *Ann. Probab.*, 7:1092–1096, 1979.

[101] V. Strassen. An invariance principle for the law of the iterated logarithm. *Z. Wahrsch. Verw. Gebiete*, 3:211–226, 1964.

[102] V. Strassen. A converse to the law of the iterated logarithm. *Z. Wahrsch. Verw. Gebiete*, 4:265–268, 1965/1966.

[103] V. Strassen. Almost sure behavior of sums of independent random variables and martingales. In *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability (Berkeley, Calif., 1965/66)*, pages Vol. II: pp. 315–343. Berkeley, Calif., 1967.

[104] D. Surgailis. Domains of attraction of self-similar multiple integrals. *Litovsk. Mat. Sb.*, 22:185–201, 1982.

[105] S. Takahashi. On the law of the iterated logarithm for lacunary trigonometric series. *Tohoku Math. J.*, 24:319–329, 1972.

[106] A. M. R. Taylor. Fluctuation tests for a change in persistence. *Oxford Bull. Econ. Stat.*, 67:207–230, 2005.

[107] S. J. Taylor. *Modelling Financial Time Series.* John Wiley & Sons Inc., New York, 1986.

[108] A. N. Tihomirov. On the convergence rate in the central limit theorem for weakly dependent random variables. *Theory Probab. Appl.*, 25:791–809, 1980.

[109] M. Weber. Un théorème central limite presque sûr à moments généralisés pour les rotations irrationnelles. *Manuscripta Math.*, 101:175–190, 2000.

[110] M. Weiss. On the law of iterated logarithm. *J. Math. Mech.*, 8:121–132, 1959.

[111] J. Zakoian. Threshold heteroskedastic models. *J. Econ. Dynamics and Control*, 18:931–955, 1994.

[112] A. Zeileis, F. Leisch, C. Kleiber, and K. Hornik. Monitoring structural change in dynamic econometric models. *J. Appl. Econometrics*, 20:99–121, 2005.