6 Estimating the variance of the sample correlation integral

The estimation of the correlation dimension of a chaotic attractor from the corresponding time series is a useful tool for understanding the complexity of the underlying dynamics and for discriminating between different types of time series (see Chapter 1). Moreover, the obtained estimate of the correlation dimension can be used in time series prediction as an estimate for the autoregression parameter. For such practical purposes the question of reliability of the estimator is important: the estimate itself is usually not enough, one needs the confidence bounds or another tool to express statistical uncertainty. This problem is essential when the dimension estimate is used for the discrimination between different time series. Difference in estimates of correlation dimensions can be explained either by different nature of time series or by sampling error: only a finite sequence of observations is used to compute the estimates. It is important to separate between these two cases. Here the estimation of the variance enters as the main tool.

As we have seen in Chapter 2, the correlation dimension can be estimated in many different ways. The most widely used one is the Grassberger-Procaccia method (or some variation of it), where the correlation dimension is estimated by first estimating the correlation integrals (2.3). Recall that in that case the variance of the estimator for the correlation dimension can be computed via the variance of the sample correlation integral (2.5).

In this chapter we want to study different procedures for estimating the variance of the sample correlation integral and compare them for the examples of a chaotic time series and of a data set consisting of independent observations. We shall carry out the Monte-Carlo study, apply the bootstrap method and consider the method based on the Hoeffding decom-
position of $U$-statistics. For the bootstrap method we shall also provide some theoretical results.

The first step in estimating the correlation integral and the correlation dimension from a time series is the reconstruction technique. From the original univariate time series we form the sequence of the reconstruction vectors $(X_1, ..., X_n)$, $X_t \in \mathbb{R}^k$, which we take to be our data set, where $k \geq 1$ is the embedding dimension.

The method of Monte-Carlo is an empirical way to estimate variability of a certain estimator. It amounts to computing the estimate a number of times on the basis of independent replications of the data set, and taking the empirical variance as the estimate for the variance of the estimator (the bias is ignored in these discussions). More precisely, let $(X^{(s)}_1, ..., X^{(s)}_n)$, $s = 1, 2, ..., m$, be $m$ independent replications of the sample from the unknown distribution $F$, and let $\theta(F)$ be the parameter that we want to estimate. Then on the basis of each sample the estimate of the parameter is obtained: $\hat{\theta}^{(s)}_n$, $s = 1, ..., m$, and its mean and the variance are given by

$$\bar{\theta}_n = \frac{1}{m} \sum_{s=1}^m \hat{\theta}^{(s)}_n, \quad (6.1)$$

$$\sigma^2_n = \frac{1}{m-1} \sum_{s=1}^m (\hat{\theta}^{(s)}_n - \bar{\theta}_n)^2. \quad (6.2)$$

Hence, the variance of the mean $\bar{\theta}_n$ is $\sigma^2_n/m$.

This method is reliable if the estimator is unbiased and enough independent replications of the sample are available. In most practical applications this is not the case. If the estimator is unbiased but the number of replications is small, then the estimate of the variance may be highly unreliable. This can be seen by the following argument. Suppose that the unknown true variance of $\hat{\theta}_n$ is $\sigma^2_n$. If the asymptotic distribution of $\hat{\theta}^{(s)}_n$ is normal, then the distribution of $\frac{n^{-1}}{\sigma^2} \hat{\theta}^2_n$ is asymptotically chi-square with $(m - 1)$ degrees of freedom. If the number of independent replications of the sample is, for instance, $m = 100$, then the 2.5% and 97.5% quantiles of the chi-square distribution with 99 degrees of freedom are $q_{0.025} = 0.75$ and $q_{0.975} = 1.29$, so that with probability 95% we have

$$0.75 \sigma^2_n \leq \hat{\sigma}^2_n \leq 1.29 \sigma^2_n. \quad (6.3)$$

In order to speed up the Monte-Carlo study, which is computationally
quite involved, one might be tempted to take a much smaller number of replications, like \( m = 10 \). The same arguments as above show that now \( 9\sigma_n^2 / \sigma_n^2 \) is \( \chi_2^2 \)-distributed. This distribution has 2.5% and 97.5% quantiles \( q_{0.025} = 0.30 \) and \( q_{0.975} = 2.11 \), yielding the following 95% interval for \( \hat{\sigma}_n^2 \):

\[
0.30\sigma^2_n \leq \hat{\sigma}_n^2 \leq 2.11\sigma^2_n.
\]

Comparing (6.4) and (6.3) one can see that the estimates for the variance obtained from 10 replications are much less reliable than those obtained from 100 replications.

When estimating the correlation dimension from a time series, we often have the outcome of a (univariate) time series of a finite length and independent replications cannot be obtained. Then one possibility is to split the time series into a number of long enough segments and carry on the Monte-Carlo study in the way described above, treating segments as if they were independent. This assumption is reasonable for a time series which is a realisation of a mixing stochastic process since, due to mixing, segments of a time series become almost independent as they are separated in time.

If the time series is relatively short, then this method does not apply. Still, the sampling distribution of the estimator used is very much needed. The method of bootstrap, first introduced by Efron [31], overcomes the problem of absence of independent replications of the sample and allows to consistently estimate the sample distribution and the variance of a statistic.

### 6.1 Bootstrap for the data from dynamical systems

#### 6.1.1 General background of the bootstrap

The bootstrap method was suggested in [31] by Efron as a method to approximate the distribution of an estimator, which is a functional of the sample and an unknown distribution function, by the so-called bootstrap distribution. This distribution is obtained by replacing the unknown distribution function by the empirical distribution of the sample. The bootstrap distribution is easily approximated by performing the Monte-Carlo study, i.e. in this case by resampling the data.
Let $X_1, \ldots, X_n$ be a sample of observations coming from an unknown distribution $F$, and let $\theta(F)$ be the parameter of interest. Suppose that the estimate of $\theta(F)$ is \( \hat{\theta}_n \). The question arises: what is the distribution of $\hat{\theta}_n - \theta(F)$ (properly scaled)? The answer is quite difficult, since this distribution will almost always depend on the unknown $F$, and, even if $F$ is known, it still might be impossible to compute precisely the distribution of $\hat{\theta}_n - \theta(F)$.

For some statistics $\hat{\theta}_n$, such as the sample mean, some other linear statistics and functions of them, $U$- and $V$-statistics, the asymptotic distribution of $\hat{\theta}_n - \theta(F)$, properly scaled, is known (in most cases it is normal), and sometimes the variance of a statistic can be computed directly from the data. For instance, if $\overline{X}_n$ is a sample mean and $\mu$ is the true mean, then the variance of $\overline{X}_n$ is $\sigma^2_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2$, and the asymptotic distribution of $\sqrt{n}(\overline{X}_n - \mu)/\sigma_n^2$ is $N(0,1)$. However, such an example is more an exception than the rule. Usually, the variance of an estimator cannot be computed directly because its distribution or its parameters are unknown. This is where the bootstrap method enters the game.

Let $F_n$ be the empirical distribution of $X_1, \ldots, X_n$, putting mass $\frac{1}{n}$ on each $X_i$. The essence of the bootstrap method is in replacing the unknown distribution $F$ by the empirical distribution $F_n$, which simply amounts to resampling the data. Let $X_1^*, \ldots, X_n^*$ be conditionally independent, with common distribution $F_n$ (i.e. what we call a bootstrap sample), and compute $\theta^*_n$ in the same way as $\hat{\theta}_n$, but using $X_1^*, \ldots, X_n^*$. Now the original $n$ data points $X_1, \ldots, X_n$ represent a finite population with distribution function $F_n$. Under the distribution $F_n$ the original estimator $\hat{\theta}_n = \theta(F_n)$ is the “true” value of the parameter to be estimated, and $\theta^*_n$ as its estimator. The distribution of $\theta^*_n - \theta_n$ is taken to be an approximation (called also the bootstrap distribution) of the distribution of $\hat{\theta}_n - \theta(F)$. Since $F_n$ is discrete, the distribution of $\theta^*_n - \hat{\theta}_n$ can be easily simulated.

The idea now is that $\mathcal{L}(\theta^*_n - \theta_n)$ is essentially the same as $\mathcal{L}(\hat{\theta}_n - \theta(F))$ (by $\mathcal{L}$ we mean the distribution). For stationary and ergodic sequences, as $n \rightarrow \infty$, $F_n$ tends to $F$ by the Glivenko-Cantelli theorem. Then, if $\mathcal{L}(\hat{\theta}_n - \theta)$ is in some sense a “uniformly continuous” function of $F$, then $\mathcal{L}(\theta^*_n - \hat{\theta}_n)$ converges to $\mathcal{L}(\theta_n - \theta)$. This “uniform continuity” condition is, however, very difficult to check. In the i.i.d. case for some statistics, such as the sample mean, $U$-statistics and some other, it was shown by Bickel
6.1. Bootstrap for the data from dynamical systems

and Freedman ([16]).

When working with time series, the assumption of independence is not satisfied. Moreover, the dependence structure of a time series will be destroyed by resampling individual observations. For dependent sequences of observations Kunsch suggested a different resampling scheme ([48], see also Carlstein [17]), the so-called moving block bootstrap. For this procedure one resamples not the original individual observations, but blocks of length \( l \) of observed data.

The motivation here is that for mixing sequences, if blocks are sufficiently long, the observations in different blocks become almost independent, while the dependence inside the blocks is preserved. The choice of the block length is essential here. On the one hand, choosing the blocks that are too long results into a small number of them, which does not allow for good approximation of the underlying distribution by the empirical distribution. On the other hand, choosing blocks that are too short destroys the short-range dependence, which still can be relatively high (e.g. high values of autocovariances). In practice, these two factors should be balanced. The sample size and the sample autocorrelation function can be good indicators for choosing a suitable block length.

Asymptotic results, such as consistency of the moving block bootstrap for certain statistics, are considered, among others, by Kunsch [48], Buhlmann [16], etc. However, in all these results some mixing condition on the sequence of observation is imposed, such as the absolute regularity or the strong mixing. As we noted above, a sequence of data arising from a dynamical system in general does not satisfy any of these mixing conditions, but is a functional of a mixing stochastic process. We are not aware of any theoretical results concerning the bootstrap for the data of this type. So, in the following section we shall provide some theoretical justification for the bootstrap in this case.

6.1.2 Consistency of the bootstrap for the functionals of mixing processes

Here we want to show that the moving block bootstrap provides a consistent procedure for estimating the distribution of statistics such as the sample mean and \( U \)-statistics, when the sequence of data is a functional of an
Estimating the variance of the sample correlation integral

absolutely regular process. For this we shall employ the notation and the main results of Chapter 4.

By consistency of the bootstrap we mean that the bootstrap distribution and the sample distribution are close in some sense. We shall prove here that for the sample mean the bootstrap distribution is close to the sample distribution in the 2-Mallows metric (Mallows [50]). Our proof uses the results of Bickel and Freedman [6] on consistency of the bootstrap for the i.i.d. case. The analogous result for \( U \)-statistics follows then in a rather straightforward way.

The 2-Mallows metric is defined on the space \( \mathcal{F}_2 \) of distributions with finite variance: \( \mathcal{F}_2 = \{ F : \int x^2 dF < \infty \} \) by

\[
d_2(F, G) = \inf \{(E[X - Y]^2)^{1/2} : \mathcal{L}(X) = F, \mathcal{L}(Y) = G \}, \quad F, G \in \mathcal{F}_2,
\]

where the infimum is taken over all joint distributions of the pair of random variables \( X \) and \( Y \) with fixed marginal distributions \( F \) and \( G \) respectively. The following two statements are equivalent:

(i) \( F_n \) converges to \( G \) in the metric \( d_2 \);

(ii) \( F_n \) converges to \( G \) weakly and \( \int x^2 dF_n(x) \rightarrow \int x^2 dG(x) \).

Thus, convergence in the 2-Mallows metric is equivalent to weak convergence together with convergence of the second moments. (For proof see Mallows [50]).

Another property of this metric we shall use here is the the following inequality, which holds whenever the sequences \( X_1, ..., X_n \) and \( Y_1, ..., Y_n \) are i.i.d.:

\[
d_2^5 \left[ \mathcal{L} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mathbb{E}X_i) \right), \mathcal{L} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i - \mathbb{E}Y_i) \right) \right] \leq d_2^5 (\mathcal{L}(X_1), \mathcal{L}(Y_1)).
\]

(6.5)

For proof see Bickel and Freedman [6].

The first theorem here concerns consistency of the bootstrap for the sample mean.

Let \( X_1, ..., X_n \), \( X_i \in \mathbb{R} \), be a sample of \( F \)-distributed random variables, which we take to be the first \( n \) observations of an infinite sequence, and let \( F_n \) be the empirical distribution. From now on we shall assume that the stationary sequence \( \{X_n\}_{n \in \mathbb{N}} \) is a functional of an absolutely regular process, satisfying for some \( r \geq 1 \) the \( r \)-approximation condition.
We consider the parameter \( \theta(F) = \mathbb{E}_F X_1 \) and its sample analogue \( \hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n X_i \), which is the sample mean. We want to apply the moving block bootstrap to estimate the distribution of \( \sqrt{n}(\hat{\theta}_n - \theta(F)) \). For simplicity, we treat here only the 1-dimensional case \( (k = 1, X_i \in \mathbb{R}) \).

The following resampling scheme is considered: for some integers \( N, m \) \((N = N_n, m = m_n \text{ shall be specified later})\), we resample long blocks of length \( N \) of \( X_1, \ldots, X_n \): \( B_1, \ldots, B_p \), separated by short blocks of length \( m \) (which are disregarded), where \( p = \lceil \frac{n}{Nm} \rceil \), and recompute the value of the sample mean on the basis of a bootstrap sample obtained in this way. In terms of distributions this bootstrap procedure can be described as follows.

Let \( G \) the theoretical distribution of the \( N \)-block \( B_1 \), and let \( G_p \) be the empirical distribution of \( B_1, \ldots, B_p \), putting the mass \( \frac{1}{p} \) on each \( B_i \). Then, let the bootstrap sample be all \( X_i^s \) in blocks \( B_1^s, \ldots, B_p^s \), which are independently chosen from the common distribution \( G_p \). The bootstrap analogue of \( \theta_n \) is now

\[
\theta_p^s = \frac{1}{pN} \sum_{s=1}^p \sum_{X_i^s \in B_s^s} X_i^s,
\]

and its expected value under the distribution \( G_p \) is

\[
\theta(G_p) = \frac{1}{pN} \sum_{s=1}^p \sum_{i \in I_s} X_i.
\]

where \( I_s \) denotes the set of indices in the block \( B_s \). (Note that in this context the bootstrap procedure can be regarded as resampling the blocks sums \( \sum_{i \in I_s} X_i \) rather than the blocks themselves).

The following theorem states that the bootstrap distribution \( \mathcal{L} \left( \sqrt{n}(\theta_p^s - \theta(G_p)) \right) \) gives a consistent estimate of the true distribution \( \mathcal{L} \left( \sqrt{n}(\theta_n - \theta(F)) \right) \), under the specified conditions.

**Theorem 6.1** Let \( \{X_n\}_{n \in \mathbb{N}} \) be a functional of an absolutely regular process \( \{Z_n\}_{n \in \mathbb{Z}} \) with mixing coefficients \( (\beta_k) \) satisfying

\[
\sum_{k=1}^{\infty} k^2 \beta_k < \infty.
\]

Suppose that the 1-approximation condition holds and \( (\alpha_l) \) are such that

\[
\sum_{l=1}^{\infty} l^2 \alpha_l < \infty.
\]
Suppose, moreover, that all $X_i$ are bounded. Then
\[
d_2 \left[ \mathcal{L} \left( \sqrt{n}(\theta_{p_n} - \theta(G_{p_n})) \right), \mathcal{L} \left( \sqrt{n}(\theta_n - \theta(F)) \right) \right] \to 0 \tag{6.6}
\]
as $n \to \infty$, for almost all realisations of $\{X_n\}$, whenever the block length $N$ satisfies $N_n = O(n^{1/2+\delta})$ for some $\delta \in (0, \frac{1}{2})$.

**Proof** By Theorem 4.1 the process $\{X_n\}_{n \in \mathbb{N}}$ is nearly regular, which implies that for all integers $N,K,L$ there exists a sequence of independent $N$-dimensional vectors $\{\tilde{B}_s\}_{s \geq 1}$, such that for $(N, K + 2L)$-blockings $B_1, B_2, \ldots$ of $\{X_n\}$
\[
P(\|B_s - \tilde{B}_s\| \geq \alpha_L) \leq 2\alpha_L + \beta_K
\]
for all $s$, and $\tilde{B}_1, \tilde{B}_2, \ldots$ have the same distribution as $B_1$, i.e. $G$. Let $K, L, N$ be such that the following relationships hold:
\[
\frac{n}{N + K + 2L} \to \infty; \quad \frac{K + 2L}{N + K + 2L} \to 0
\]
and let $m = K + 2L$. If the sample size is $n$, then the number of blocks in the sample is $p = p_n = \lfloor \frac{n}{Nm} \rfloor$. Let $G_p$ and $\tilde{G}_p$ denote the empirical distribution of $B_1, \ldots, B_p$ and $\tilde{B}_1, \ldots, \tilde{B}_p$ respectively.

Define a new function $H : \mathbb{R}^N \to \mathbb{R}$ by
\[
H(x) = \sum_{i=1}^{N} x_i, \quad x = (x_1, \ldots, x_N).
\]
Then $H(B_s) = \sum_{i \in L} X_i$ is a block $B_s$ sum. Note that
\[
\mathbf{E}_G H(B_1) = \mathbf{E}_G H(\tilde{B}_1) = N \mathbf{E}_F X_1.
\]
Let $\tilde{B}_1^s, \ldots, \tilde{B}_p^s$ be the bootstrap sample from the empirical distribution $\tilde{G}_p$.
Denote
\[
\tilde{H}_p = \frac{1}{p} \sum_{s=1}^{p} H(\tilde{B}_s)
\]
the expected value of $H(\tilde{B}_1)$ under the distribution $\tilde{G}_p$, and
\[
H_p = \frac{1}{p} \sum_{s=1}^{p} H(B_s)
\]
the expected value of $H(B_1)$ under the distribution $G_p$.

Let $F^H_p$ and $\bar{F}^H_p$ be the empirical distribution functions of block sums $H(B_1), \ldots, H(B_p)$ and $H(\bar{B}_1), \ldots, H(\bar{B}_p)$ respectively, and $F^H$ the distribution function of $H(B_1)$. We can regard the bootstrap samples $B^*_1, \ldots, B^*_p$ and $\bar{B}^*_1, \ldots, \bar{B}^*_p$ as coming from the empirical distributions $G_p$ and $\bar{G}_p$ resp., or, equivalently, the bootstrap samples $H(B^*_1), \ldots, H(B^*_p)$ and $H(\bar{B}^*_1), \ldots, H(\bar{B}^*_p)$ as coming from resp. $F^H_p$ and $\bar{F}^H_p$.

For $\bar{B}_1, \bar{B}_2, \ldots$ independent and fixed block length $N$, the theorem of Bickel and Freedman [6] implies that

$$d_2 \left[ \mathcal{L} \left( \frac{1}{\sqrt{pN}} \sum_{s=1}^p (H(\bar{B}_s) - \bar{H}_p) \right), \mathcal{L} \left( \frac{1}{\sqrt{pN}} \sum_{s=1}^p (H(B_s) - \mathbf{E}_GH(B_1)) \right) \right] \leq d_2 [\mathcal{L}(H(\bar{B}^*_1)), \mathcal{L}(H(\bar{B}_1))] = d_2[F^H_p, F^H] \rightarrow 0 \quad (6.7)$$

as $p \rightarrow \infty$, for almost all realisations of $\bar{B}_1, \bar{B}_2, \ldots$, i.e. that the (regular) bootstrap works for the sample mean of $H(\bar{B}_1), \ldots, H(\bar{B}_p)$. Here the block length is not fixed, but in that case there exists a subsequence $p_{ni} = p_{ni}(N) \rightarrow \infty$ as $n \rightarrow \infty$ for which the convergence (6.7) holds for all $N$. Hence, also

$$d_2 \left[ \mathcal{L} \left( \frac{1}{\sqrt{pN}} \sum_{s=1}^p (H(\bar{B}_s) - \bar{H}_p) \right), \mathcal{L} \left( \frac{1}{\sqrt{pN}} \sum_{s=1}^p (H(B_s) - \mathbf{E}_GH(B_1)) \right) \right] \rightarrow 0. \quad (6.8)$$

What we are interested in, is not exactly the distributions in (6.8), but those in (6.6). However, in what follows we shall show that the respective distributions in (6.6) and (6.8) are close to each other.

Define the following random variables:

$$A := \sqrt{n}(\theta^*_p - \theta(G_p)) = \frac{\sqrt{n}}{pN} \sum_{s=1}^p (H(B^*_s) - H_p),$$

$$A' := \frac{1}{\sqrt{pN}} \sum_{s=1}^p (H(\bar{B}^*_s) - \bar{H}_p),$$

$$B := \sqrt{n}(\theta_n - \theta(F)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mathbf{E}F X_1),$$

$$B' := \frac{1}{\sqrt{pN}} \sum_{s=1}^p (H(\bar{B}_s) - \mathbf{E}_G H(B_1)).$$
Then, in this notation, we have that $d_2[\mathcal{L}(A'),\mathcal{L}(B')] \to 0$, due to (6.8), and the theorem then requires that $d_2[\mathcal{L}(A),\mathcal{L}(B)] \to 0$. Since $d_2$ is a metric, the triangle inequality implies that

$$d_2[\mathcal{L}(A),\mathcal{L}(B)] \leq d_2[\mathcal{L}(A),\mathcal{L}(A')] + d_2[\mathcal{L}(A'),\mathcal{L}(B')] + d_2[\mathcal{L}(B'),\mathcal{L}(B)].$$

If we can show that $d_2[\mathcal{L}(A),\mathcal{L}(A')]$ and $d_2[\mathcal{L}(B'),\mathcal{L}(B)]$ both tend to 0 as $n \to \infty$, the statement of the theorem will follow.

First we show that $|B - B'| \to 0$ in probability as $n \to \infty$, which implies that also $d_2[\mathcal{L}(B'),\mathcal{L}(B)] \to 0$ as $n \to \infty$.

Denote $g(X) = X - \mathbf{E}_F X$, $I^s$ the set of indices between $st$th and $(s+1)$st blocks of length $N$, and $i_p$ the last index in the $p$th block. We have

$$\frac{1}{\sqrt{n}} \left[ \sum_{i=1}^{n} (X_i - \mathbf{E}_F X_1) - \sum_{s=1}^{p} (H(B_s) - \mathbf{E}_G H(B_1)) \right]$$

$$= \frac{1}{\sqrt{n}} \left[ \sum_{i=1}^{n} g(X_i) - \sum_{s=1}^{p} \sum_{i \in I_s} g(X_i) \right]$$

$$= \frac{1}{\sqrt{n}} \left[ \sum_{s=1}^{p} \sum_{i \in I_s} (g(X_i) - g(\bar{X}_i)) + \sum_{s=1}^{p} \sum_{i \notin I_s} g(X_i) + \sum_{i=i_p}^{n} g(X_i) \right]. \quad (6.9)$$

To show that all terms in (6.9) are small in probability, we proceed in the same way as in Theorem 4.2. For the first term in (6.9) note that

$$\left| \sum_{i \in I_s} (g(X_i) - g(\bar{X}_i)) \right| \leq \sum_{i \in I_s} |g(X_i) - g(\bar{X}_i)| = \| B_s - \bar{B}_s \|,$$

and then

$$\mathbf{P} \left( \frac{1}{\sqrt{n}} \sum_{s=1}^{p} \sum_{i \in I_s} (g(X_i) - g(\bar{X}_i)) \geq \frac{\rho_0L}{\sqrt{n}} \right)$$

$$\leq \sum_{s=1}^{p} \mathbf{P} \left( \left| \sum_{i \in I_s} (g(X_i) - g(\bar{X}_i)) \right| \geq \alpha_L \right)$$

$$\leq p(\beta_K + 2\alpha_L) \quad \quad (6.10)$$

by the near regularity condition (4.24). If $K, L, N$ are such that $p(2\alpha_L + \beta_K) \to 0$ as $n \to \infty$ (which implies $\frac{\rho_0}{\sqrt{n}} \to 0$), yields

$$\frac{1}{\sqrt{n}} \sum_{s=1}^{p} \sum_{i \in I_s} (g(X_i) - g(\bar{X}_i)) \to 0 \text{ in probability,}$$
as \( n \to \infty \). Since \( N = N_n = O(n^{1/2+\delta}) \), we can take \( L = L_n = K = K_n = O(n^{1/2+\delta/2}) \). Then the condition on \((\alpha_k)\) (4.22) implies

\[
L_n \sim n^{1/2}, \quad \sum_{l=1}^{\infty} l^2 \alpha_l < \infty \Rightarrow \sum_{n=1}^{\infty} n^{1/2-\delta} \alpha L_n < \infty \Rightarrow n^{1/2-\delta} \alpha L_n \to 0,
\]

and so, \( p \alpha L_n \sim \frac{n}{N} \alpha L_n \sim n^{1/2-\delta} \alpha L_n \to 0 \). Also \( p \beta K_n \to 0 \) by the same reasoning.

To estimate the last two terms in (6.9) we can use Lemma 4.9. Since \( g(X_i) \) are centered and bounded, it implies

\[
E \left( \frac{1}{\sqrt{n}} \sum_{s=1}^{p} \sum_{i=i_s-1+N+1}^{i_s} g(X_i) \right)^2 \leq \frac{C p (K + 2L)}{n}
= \frac{C (K + 2L)}{N + K + 2L} \to 0
\]
as \( n \to \infty \), by the choice of \( K, N, L \), and so,

\[
\frac{1}{\sqrt{n}} \sum_{s=1}^{p} \sum_{i=i_s-1+N+1}^{i_s} g(X_i) \to 0 \text{ in probability, as } n \to \infty.
\]

Also

\[
E \left( \frac{1}{\sqrt{n}} \sum_{i=i_p+1}^{n} g(X_i) \right)^2 \leq \frac{C N}{n} \to 0 \text{ as } n \to \infty,
\]

which implies

\[
\frac{1}{\sqrt{n}} \sum_{i=i_p+1}^{n} g(X_i) \to 0 \text{ in probability, as } n \to \infty.
\]

Hence, since \( \frac{1}{\sqrt{n}} \to 1 \),

\[
|B - B'| = \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - E[X_i]) - \frac{1}{\sqrt{pN}} \sum_{s=1}^{p} (H(\bar{B}_s) - E[H(\bar{B}_s)]) \right| \to 0
\]
in probability, as \( n \to \infty \).

Next, note that since \( B_p^s, \ldots, B_p^s \) and \( \bar{B}_1^s, \ldots, \bar{B}_p^s \) are (conditionally) independent, (6.5) implies that

\[
d_2^2 \left[ \mathcal{L} \left( \frac{1}{\sqrt{p}} \sum_{s=1}^{p} (H(B_p^s) - H_p) \right), \mathcal{L} \left( \frac{1}{\sqrt{p}} \sum_{s=1}^{p} (H(\bar{B}_s) - \bar{H}_p) \right) \right] \leq d_2^2[F_p^H, \bar{F}_p^H].
\]
Again, for a fixed block length, the triangle inequality, the Glivenko-Cantelli theorem and the ergodic theorem together imply

\[ d_2^2[F_p^H, \bar{F}_p^H] \leq d_2^2[F^H_p, \bar{F}_p^H] + d_2^2[F^H_p, \bar{F}_p^H] \to 0 \quad (6.11) \]

as \( p \to \infty \). Although the block length is not fixed, still we can find, as above, the subsequence \( p_{n_i} \to \infty \) as \( n \to \infty \) for which the convergence in (6.11) holds for all \( N \). Therefore, as \( n \to \infty \), also

\[ d_2[\mathcal{L}(A), \mathcal{L}(A')] \to 0, \]

which concludes the proof of the theorem. \( \Box \).

### 6.1.3 Hoeffding decomposition method and bootstrap for U-statistics

Above we considered the moving block bootstrap method for the sample mean. However, in our application we want to estimate the variance of the sample correlation integral, which has the form of a \( U \)-statistic. Here we shall see how its variance can be estimated directly, using the Hoeffding decomposition, or by the bootstrap method, alone or in combination with the Hoeffding decomposition.

Let again \( X_1, \ldots, X_n, X_i \in \mathbb{R}^k \), be \( F \)-distributed random vectors (reconstruction vectors, in our case). Recall that, according to the Hoeffding decomposition, a \( U \)-statistic can be represented as

\[ U_n = \theta(F) + \frac{2}{n} \sum_{i=1}^{n} [h_1(X_i - \theta(F))] + R_n, \quad (6.12) \]

where

\[ h_1(x) = \int_{\mathbb{R}^k} h(x, y) dF(y) \]

and the remainder \( R_n \), defined by the relation (6.12), is in our case negligible comparing to the first term in (6.12), which dominates the behaviour of a \( U \)-statistic (see Chapter 5). As a consequence, the asymptotic variance of a \( U \)-statistic is the asymptotic variance of the leading term in the Hoeffding decomposition, and it is given by

\[ \sigma^2 = 4 \text{Var}[h_1(X_1)] + 8 \sum_{k=1}^{\infty} \text{Cov}[h_1(X_1)h_1(X_{k+1})], \quad (6.13) \]
while the exact variance of the leading term in (6.12) is

\[ \sigma_n^2 = \mathbb{E} \left[ \left( \frac{2}{n} \sum_{i=1}^{n} [h_1(X_i) - \theta(F)] \right)^2 \right]. \]

Note that

\[ \sigma^2 = \lim_{n \to \infty} n^{-1} \sigma_n^2. \]

In case of the sample correlation integral the function \( h_1(X) \) has the form

\[ h_1(X) = \int_{\mathbb{R}^d} 1\{\| X - y \| \leq \epsilon \} dF(y) = \mathbb{P}_y(\| X - y \| \leq \epsilon), \quad (6.14) \]

so \( h_1(X) \) is the “mass” of a ball of radius \( \epsilon \) around \( X \).

The above expressions for the variance of \( U \)-statistics can be used to estimate the variance of the sample correlation integral \( C_n(\epsilon) \). First \( h_1(X_i), i = 1, \ldots, n \), are estimated by

\[ \hat{h}_1(X_i) = \frac{1}{n-1} \sum_{j \neq i}^{n} 1\{\| X_i - X_j \| \leq \epsilon \} \quad (6.15) \]

and obtain a real-valued sequence \( \hat{h}_1(X_1), \ldots, \hat{h}_1(X_n) \). Then the sample correlation integral is computed:

\[ C_n(\epsilon) = \frac{1}{n} \sum_{i=1}^{n} \hat{h}_1(X_i) \]

and the variance of \( C_n(\epsilon) \) is estimated by

\[ \hat{\sigma}^2_n = \frac{4}{n} \left\{ \text{Var}[h_1(X_1)] + 2 \sum_{k>1} \text{Cov}[h_1(X_1)h_1(X_{k+1})] \right\}, \quad (6.16) \]

with

\[ \text{Var}[h_1(X_1)] = \frac{1}{n-1} \sum_{i=1}^{n} [\hat{h}_1(X_i) - C_n(\epsilon)]^2 \]

and

\[ \text{Cov}[h_1(X_1)h_1(X_{k+1})] = \frac{1}{n-k} \sum_{i=1}^{n-k} [(\hat{h}_1(X_i) - C_n(\epsilon))(\hat{h}_1(X_{i+k}) - C_n(\epsilon))]. \quad (6.17) \]
Although in the theoretical expression for the variance we have infinite sum of covariances, in practical situations we have to truncate this sum by some maximal considered lag $M << n$, since for higher lags the estimates for covariances (6.17) become unreliable and, moreover, in most practical situations covariances on higher lags are negligible.

The application of the bootstrap to $U$-statistics is closely related to the Hoeffding decomposition. The asymptotic variance of a $U$-statistic is given by the asymptotic variance of the first, linear, term of the Hoeffding decomposition (6.13). We can consistently estimate this variance by applying the moving block bootstrap. Consistency in this case follows from Theorem 6.1, provided the conditions of the theorem are satisfied for the sequence $\{h_1(X_n)\}_{n \in \mathbb{N}}$. The remainder term is negligible, which can be seen using the same arguments as in Chapter 5 (Lemmas 5.3 and 5.8).

In practice, the application of the moving block bootstrap to $U$-statistics can be carried out in two different ways. One way is to resample blocks of the sequence of data, as it was described above, and then recompute the value of the $U$-statistic on the basis of each new bootstrap sample. Another way is to use the fact that the variance of a $U$-statistic is determined by the variance of the leading term of the Hoeffding decomposition, which can be estimated by applying the moving block bootstrap to the sequence $\{h_1(X_i)\}_{i \in \mathbb{N}}$, or rather to the estimated sequence $\{\hat{h}_1(X_i)\}_{i \in \mathbb{N}}$. We shall test both these procedures on the example of a chaotic time series in the next section.

6.2 Application to a chaotic time series

The underlying time series is the first component of the Lorenz dynamical system, given by the following differential equations:

\[
\begin{align*}
\dot{x} &= \sigma(y - x) \\
\dot{y} &= \rho x - y - xz \\
\dot{z} &= -\beta z + xy.
\end{align*}
\]  

(6.18)

The parameter values were used $(\sigma, \rho, \beta) = (10, 28, 8/3)$, where the Lorenz system is known to exhibit chaotic behaviour. The system was integrated numerically and then sampled at the time distance $\Delta t = 0.08$, yielding the
discrete time series
\[ x_n = x(t_0 + n \Delta t). \] (6.19)

Fig. 6.1 gives the plot of a part of the time series \( \{x_n\} \) and Fig. 6.2 - its autocorrelation function. For this time series the sample of reconstruction vectors was generated: \( \{X_i\} = \{(x_i, \ldots, x_{i+k-1})\} \in \mathbb{R}^k, \quad i = 1, \ldots, n \), and further taken as the data set on which all estimates are based. The embedding dimension was chosen \( k = 5 \), since the chaotic attractor of the Lorenz system is known to be of fractal dimension close to 2. For the value of \( k = 5 \) the condition on the embedding dimension appearing in the Takens theorem, which assures successful reconstruction of the attractor, is satisfied.

To compute the sample correlation integral we have chosen for a slight modification of (2.3), taking into account only non-overlapping reconstruction vectors. The resulting estimator is then
\[
C_n(\epsilon) = \frac{2}{(n - 4)(n - 5)} \sum_{1 \leq i \neq j \leq n}^{1 \leq i \leq n} 1\{\|X_i - X_j\| \leq \epsilon\}. \quad (6.20)
\]

For the norm \( \| \cdot \| \) we take the maximum norm. The total size of the Lorenz attractor is around 30, and, since the correlation integral should be estimated for small values of \( \epsilon \), it was chosen approximately 1.5% of the total size of the attractor, i.e. \( \epsilon = 0.5 \). We compare methods for estimating
the variance of the sample correlation integral $C_n(\epsilon)$, based on samples of size $n = 1000$.

### 6.2.1 Monte-Carlo method

To have the estimate of the variance which is some kind of a “benchmark” for the further comparison, i.e. which is as close as possible (in terms of computer abilities) to the “real” value, we first performed a rather large Monte-Carlo study. We generated $m = 100$ independent replications of the sample $X_{1}^{(s)}, \ldots, X_{n}^{(s)}$, $s = 1, \ldots, 100$; $n = 1000$ (by choosing different initial values $(x_0, y_0, z_0)$ of the Lorenz dynamical system), and for each replication the sample correlation integral was computed: $C_{n}^{(s)}(\epsilon)$, $s = 1, \ldots, m$. Fig. 6.3 shows the histogram of the values of $C_{n}^{(s)}(\epsilon)$. For the estimate of $C(\epsilon)$ the mean of $C_{n}^{(s)}(\epsilon)$ was taken, and the variance of $C_{n}(\epsilon)$ was computed as in (6.2). The obtained numerical values are:

\[
\bar{C}_n(\epsilon) = 1.19 \cdot 10^{-3} \quad \text{and} \quad \tilde{\sigma}_n = 1.75 \cdot 10^{-4}.
\]

The standard deviation of the mean $\bar{C}_n(\epsilon)$ is $\tilde{\sigma}_n / \sqrt{m} = 1.75 \cdot 10^{-5} = 0.0175 \cdot 10^{-3}$, yielding the following 95% confidence interval for $C(\epsilon)$:

\[
1.16 \cdot 10^{-3} \leq C(\epsilon) \leq 1.22 \cdot 10^{-3}.
\]

In Fig. 6.3 we have compared the histogram of the values of $C_{n}^{(s)}$, $s = 1, \ldots, 100$ with the density of the normal distribution with mean $\bar{C}_n$ and variance $\tilde{\sigma}_n^2$. Good fit is confirmed by a Q-Q-plot (Fig. 6.4) of the empirical distribution of the $C_{n}^{(s)}$s versus a normal distribution. To test the performance of the Monte-Carlo study with smaller number of replications than $m = 100$ and to verify our theoretical considerations of the section 6.1, and, in particular, (6.4), we repeated 30 times the Monte-Carlo study with $m = 10$. The obtained 30 estimates of the variance $\tilde{\sigma}_n^2$ are indeed well fitted by a chi-square density, as Figures 6.5 and 6.6 (Q-Q-plot) show.

### 6.2.2 Application of the bootstrap

Here we want to apply the moving block bootstrap to our data set, i.e. the sequence of the reconstruction vectors $X_1, \ldots, X_{1000}$. For practical purposes
we use a resampling procedure which slightly differs from that suggested in Section 6.2. We split the original data set into blocks of specified length and do not leave out observations between blocks. Then we construct a bootstrap sample by drawing blocks with replacement and “gluing” them together. This brings certain boundary effect into the estimate. There are different ways to overcome this problem, such as tapering, weighting the observations close to the ends of blocks, selecting only non-overlapping blocks, etc. For instance, Kunsch [48], Carlstein [17] considered these techniques and the influence of the boundary effect on the estimates. We, however, shall not go into details of this problem and will select only non-overlapping blocks to minimise the effect of gluing the blocks together.

The important remaining question is the choice of the block length. On the one hand, we would like to have blocks of the length as short as possible to assure good approximation of the theoretical distribution by the empirical distribution, and, on the other hand, not too short to destroy significant short-range dependence. One of the indicators of the optimal block length can be the autocorrelation function of the original time series (Fig. 6.2).

We see that the autocorrelations for lags $\geq 15$ are almost zero. They will be disregarded in any case. The steepest decay of the autocorrelations is observed for the lag 6. That is why we take 6 to be the smallest possible block length. Whether taking longer blocks would make our estimates more
reliable, i.e. whether the autocorrelations of lags 7-12 are important for our estimates, is not yet clear. To explain this question, we shall compare the estimates obtained by the moving blocks bootstrap with the block lengths 6 and 10. The number of bootstrap resampling iterations was chosen to be $B = 200$ (in literature it is noted that already for $B \geq 50$ a reasonable approximation of the empirical distribution can be obtained).

First, we perform the moving block bootstrap with the block length $l = 6$. For each bootstrap sample the value of the sample correlation integral was computed: $C_n^{(b)}(\epsilon)$, $b = 1, \ldots, 200$, and Fig. 6.7 gives a histogram of the obtained values. We compute the mean of $C_n^{(b)}(\epsilon)$ and the bootstrap estimate of the standard deviation of $C_n(\epsilon)$ as

$$\bar{C}_n = \frac{1}{B} \sum_{b=1}^{B} C_n^{(b)}(\epsilon)$$

$$\hat{\sigma}_n^* = \left\{ \frac{1}{B - 1} \sum_{b=1}^{B} (C_n^{(b)}(\epsilon) - \bar{C}_n^*)^2 \right\}^{1/2},$$

where $B = 200$, $n = 1000$. The following numerical values were obtained:

$$\bar{C}_n^* = 2.11 \cdot 10^{-3},$$

$$\hat{\sigma}_n^* = 1.97 \cdot 10^{-4}. \quad (6.24)$$

Then the standard deviation of the mean $\bar{C}_n^*$ is $\hat{\sigma}_n^*/\sqrt{B} = 1.4 \cdot 10^{-5} = 0.014 \cdot 10^{-3}$, leading to the following bootstrap 95% confidence interval for
6.2. Application to a chaotic time series

\[ C(\epsilon) : \quad 2.08 \cdot 10^{-3} \leq C(\epsilon) \leq 2.14 \cdot 10^{-3}. \quad (6.25) \]

In Fig. 6.7 the histogram of the bootstrap distribution of \( C_n^{(b)}(\epsilon) \), \( b = 1, \ldots, 200 \), is compared with the normal density with mean \( \bar{C}_n^{(b)} \) and variance \( \sigma_n^{(b)}^2 \), and in Fig. 6.8 the Q-Q-plot vs. normal distribution is shown.

Comparing the result in (6.25) with (6.23), one can see that the bootstrap estimate for \( C(\epsilon) \) is biased by a constant factor, which is the result of repetition of some blocks in a bootstrap sample, what artificially increases the number of \( \epsilon \)-close reconstruction vectors in the sequence. The computer simulation gives the estimate of the average bias

\[ \text{bias} = 0.92 \cdot 10^{-3}. \quad (6.26) \]

Then bias-corrected confidence interval (6.25) becomes

\[ 1.16 \cdot 10^{-3} \leq C(\epsilon) \leq 1.22 \cdot 10^{-3} \quad (6.27) \]

Comparison of (6.27) with (6.23), and (6.24) with (6.22) shows that the moving block bootstrap with the block length \( l = 6 \) gives a very close value of the estimate of the standard deviation as that obtained by the Monte-Carlo study.

The moving block bootstrap with the block length \( l = 10 \) gives the following estimates:

\[ \bar{C}_n^{(b)} = 2.08 \cdot 10^{-3} \]
Estimating the variance of the sample correlation integral

\[ \hat{\sigma}_n^2 = 2.98 \cdot 10^{-4}. \]  (6.28)

The estimate of the standard deviation becomes much worse in this case, it is significantly (almost by factor 2) higher than the estimate in (6.23). This shows that choosing the blocks too long indeed does not allow for good distribution approximation and decreases precision of the variance estimate.

6.2.3 Hoeffding decomposition method

We estimated \( \hat{h}_1(X_1), \ldots, \hat{h}_1(X_n) \) from 1000 reconstruction vectors as in (6.15), computed \( C_n(\epsilon) \) as the mean of \( \hat{h}_1 \)'s and the variance estimate by (6.16). In (6.17) we truncate the sum of covariances at \( M = 20 \), since for higher lags the covariances are relatively small and their estimates become unreliable. For our data set we have

\[ C_n(\epsilon) = 1.15 \cdot 10^{-3} \]
\[ \hat{\sigma}_n = 1.01e-04. \]

The estimate for \( \sigma_n \) obtained by this method is smaller than the estimate (6.22) obtained by the Monte-Carlo study. This is possibly the result of truncating the sum of covariances in (6.16) at a too low lag. Indeed, the plot of the autocorrelation function of \( \{\hat{h}_1(X_i)\}_{i=1,\ldots,n} \) (Fig. 6.9) shows that the covariances at higher lags do not decay rapidly to 0, but remain sort of periodic.

![Figure 6.9: Autocorrelation function of \( \{\hat{h}_1(X_i)\} \)](image-url)
6.2.4 Application of the bootstrap to \( \{ h_i(X_i) \} \)

Here we shall apply the moving block bootstrap not to the original data set \( \{ X_i \}_{i=1,...,n} \), but to the sequence \( \hat{h}_1(X_1), ..., \hat{h}_1(X_n) \).

As it was shown in Section 6.2, in application of the moving block bootstrap to \( U \)-statistics resampling the original data sequence is (asymptotically) equivalent to resampling the members of the leading term of the Hoeffding decomposition of a \( U \)-statistic, i.e. \( h_1(X_i) \)’s. This is also frequently suggested in literature on bootstrapping \( U \)-statistics (see, for example, Atrey 

As it was shown in Section 6.2, in application of the moving block bootstrap to \( U \)-statistics resampling the original data sequence is (asymptotically) equivalent to resampling the members of the leading term of the Hoeffding decomposition of a \( U \)-statistic, i.e. \( h_1(X_i) \)’s. This is also frequently suggested in literature on bootstrapping \( U \)-statistics (see, for example, Atrey [2]). This gives us the motivation to test this method in practice.

The block length was again taken \( l = 6 \) and the number of bootstrap iterations \( B = 200 \). For each bootstrap sample \( \hat{h}_{11}^{(b)}, ..., \hat{h}_{1n}^{(b)}, b = 1, ..., 200 \) we compute the sample correlation integral by

\[
C_n^{(b)}(\epsilon) = \hat{C}_n(\epsilon) + \frac{2}{n} \sum_{j=1}^{n} [\hat{h}_{ij}^{(b)} - \hat{C}_n(\epsilon)]
\]

as in (6.12), disregarding the remainder term, with

\[
\hat{C}_n(\epsilon) = \frac{1}{n} \sum_{j=1}^{n} \hat{h}_1(X_j).
\]

Then the bootstrap estimate of the standard deviation \( \hat{\sigma}_n^* \) is

\[
\hat{\sigma}_n^* = \left[ \frac{1}{B-1} \sum_{b=1}^{B} (C_n^{(b)}(\epsilon) - \hat{C}_n^*(\epsilon))^2 \right]^{1/2},
\]

where

\[
\hat{C}_n^*(\epsilon) = \frac{1}{B} \sum_{b=1}^{B} C_n^{(b)}(\epsilon).
\]

Numerical values for our data set are

\[
\hat{C}_n^*(\epsilon) = 1.18 \cdot 10^{-3}, \\
\hat{\sigma}_n^* = 1.60 \cdot 10^{-4},
\]

and the bootstrap-estimated 95\% confidence interval for \( C(\epsilon) \) is

\[
1.16 \cdot 10^{-3} \leq C(\epsilon) \leq 1.20 \cdot 10^{-3}.
\]
The estimated value of the standard deviation $\hat{\sigma}_n^2 = 1.60 \cdot 10^{-4}$ is lower, but again reasonably close to that obtained by the Monte-Carlo study (6.22), indicating the good performance of the moving block bootstrap applied to $\hat{h}_1(X_1),...,\hat{h}_1(X_n)$.

6.3 Comparison with a data set of independent observations

To compare the bootstrap method for dependent and independent observations, we repeat the above analysis for the sequence of independent observations taken from the uniform distribution on the Cantor set on $[0,1]$.

First, proceeding as in Section 6.2, we perform the Monte-Carlo study to estimate the correlation integral for $\epsilon=0.002$, which is 2% of the size of the unit interval, and the standard deviation $\sigma_n$ of $C_n(\epsilon)$, $n = 1000$. We again took $m = 100$ independent replications of our data set of size $n = 1000$, and obtained the following numerical values:

\[
\bar{C}_n = 1.951 \cdot 10^{-2},
\hat{\sigma}_n = 3.16 \cdot 10^{-4}.
\] (6.29)

Hence, the standard deviation of the mean $\bar{C}_n$ is $\hat{\sigma}_n / \sqrt{m} = 3.16 \cdot 10^{-5} = 0.00316 \cdot 10^{-2}$, yielding the following 95% confidence interval for $C(\epsilon)$:

\[
1.947 \cdot 10^{-2} \leq C(\epsilon) \leq 1.954 \cdot 10^{-2}.
\] (6.30)

Here we see that for independent data set the standard deviation is by one order smaller relative to the value of $\bar{C}_n$, than in the case of dependent data from the Lorenz system (6.22). It shows that dependence brings significant contribution into the variance.

Now we apply the bootstrap procedure to the set of independent data, and compare the results with the values above, which we again consider as “true” values, as well as with the performance of the bootstrap in the dependent case. The moving block bootstrap with block length $l = 6$ and number of bootstrap iterations $B = 200$ gives the following numerical values:

\[
\bar{C}^*_n = 2.017 \cdot 10^{-2},
\hat{\sigma}^*_n = 3.26 \cdot 10^{-4}.
\]
6.3. *Comparison with a data set of independent observations*

Then the standard deviation of the mean \( \tilde{C}_n^* \) is \( \tilde{\sigma}_n^*/\sqrt{B} = 2.3 \cdot 10^{-5} = 0.00023 \cdot 10^{-2} \), and the 95% confidence interval for \( C(\epsilon) \):

\[
2.013 \cdot 10^{-2} \leq C(\epsilon) \leq 2.021 \cdot 10^{-2}.
\]

As above, we observe bias in the estimates of the correlation integral. The estimate for the standard deviation obtained by the bootstrap and the “true” value in (6.29) differ even less in this case than the corresponding results for dependent data set.

As in the previous section, we also performed moving block bootstrap for longer blocks \( (l = 10) \), and the following numerical results were obtained:

\[
\tilde{C}_n^* = 2.018 \cdot 10^{-2}, \quad \tilde{\sigma}_n^* = 4.02 \cdot 10^{-4}.
\]

As in (6.28), the bootstrap longer blocks gives the estimate for the standard deviation which is further away from the “true” value.

This comparison shows that the moving block bootstrap works just as good for dependent as for independent data, but gives rather significant deviation from the true value of standard deviation if too long blocks are taken. The results improve considerably as we decrease the block length.