

ESTIMATING THE CORRELATION DIMENSION FROM A CHAOTIC SYSTEM WITH DYNAMIC NOISE

Atsushi Kawaguchi*, Koji Yonemoto** and Takashi Yanagawa*

This paper proposes an estimator of the correlation dimension of the skeleton for chaotic dynamical system with dynamic noise and prove the consistency of the estimator under some assumptions.

Key words and phrases: Chaotic dynamical system, correlation dimension, Nadaraya-Watson kernel type estimator, U-statistics.

1. Introduction

Consider dynamical system $Y_t = F(Y_{t-1}, Y_{t-2}, \dots, Y_{t-d})$. Putting $\mathbf{Y}_t = {}^t(Y_t, Y_{t-1}, \dots, Y_{t-d+1})$, the system is represented by

$$(1.1) \quad \mathbf{Y}_t = \mathbf{F}(\mathbf{Y}_{t-1}),$$

where $\mathbf{F}(\mathbf{x}) = {}^t(F(\mathbf{x}), x_1, \dots, x_{d-1})$ for $\mathbf{x} = {}^t(x_1, x_2, \dots, x_d)$. Assuming $\mathbf{Y}_t \in \Omega$ for closed $\Omega \subseteq \mathbf{R}^d$ and also assuming the ergodicity of $\{\mathbf{Y}_t\}$, we formulate the system as $(\Omega, \mathcal{F}, \mu, \mathbf{F})$, where \mathcal{F} is the completion of the Borel σ -field with respect to μ , and μ is an invariant measure, i.e. $\mu(\mathbf{F}^{-1}A) = \mu(A)$ for $A \in \mathcal{F}$. See Carlsson (2002) for a description of the sufficient conditions for the existence of the unique invariant measure.

This paper is concerned with the correlation dimension which is defined by Grassberger and Procaccia (1983a, b) as follows. Putting

$$C(r) = \iint_{\Omega \times \Omega} I(\|\mathbf{y}_1 - \mathbf{y}_2\| \leq r) d\mu(\mathbf{y}_1) d\mu(\mathbf{y}_2),$$

where I denotes an indicator function and $\|\cdot\|$ is a norm, the correlation dimension is defined as

$$(1.2) \quad \nu = \lim_{r \rightarrow 0} \frac{\log C(r)}{\log r}$$

if the limit exists.

Estimators of the correlation dimension have been proposed by Takens (1985), Cutler (1990), Smith (1992), Judd (1992), and Kawaguchi (2002). Also, Schreiber (1993), Diks (1996), Kugiumtzis (1997), and Oltmans and Verheijen (1997) have discussed the estimators of the correlation dimension taking into

Received April 1, 2004. Revised June 28, 2004. Accepted May 16, 2005.

*Biostatistics Center, Kurume University, 67 Asahi-machi, Kurume 830-0001, Japan.

**Graduate School of Medical Sciences, Kyushu University, Health C&C Center, Kubara 1822-1, Hisayama Town, Kasuya-gun, Fukuoka 811-2501, Japan.

account the effect of observational noise, i.e. they considered trajectory $\{X_t\}$ generated from $X_t = W_t + \varepsilon_t$ where $\{W_t\}$ is underlying system and ε_t is observational noise.

Consider trajectory $\{X_t\}_{t=1,2,\dots,N}$ generated from

$$(1.3) \quad X_t = F(X_{t-\tau}, X_{t-2\tau}, \dots, X_{t-d\tau}) + \varepsilon_t,$$

where $F : \mathbf{R}^d \rightarrow \mathbf{R}$ is unknown non-linear function, d and τ are unknown positive integers called the embedding dimension and delay time, respectively, and $\{\varepsilon_t\}$ is called dynamic noise. We assume ε_t 's are random variables on the probability space $(\Omega', \mathcal{F}', \mathcal{E})$ such that $E[\varepsilon_t \mid \mathcal{A}_1^{t-1}(X)] = 0$ and $E[\varepsilon_t^2 \mid \mathcal{A}_1^{t-1}(X)] = \sigma^2$ ($\sigma > 0$), almost surely, where $\mathcal{A}_s^t(X)$ denotes the sigma algebra generated by (X_s, \dots, X_t) . In addition, we assume

ASSUMPTION 1.1. There exists a compact set $G \subset \mathbf{R}^d$ such that for any $\mathbf{x} \in G^\circ$,

$$\mathbf{F}(\mathbf{x}) + \mathbf{e} \in G^\circ \quad \text{a.e.},$$

where G° denotes an interior of G , $\mathbf{e} = {}^t(\varepsilon, 0, \dots, 0)$, and ε is identically distributed as ε_t for every t .

Remark 1. When F is satisfied Assumption 1.1, it follows from theorem of An and Huang (1996) that $\{X_t\}$ is geometrically ergodic. Then, an invariant measure for $\{X_t\}$ exists, and if it is taken to be the distribution of X_1 then $\{X_t\}$ is strictly stationary.

Model (1.3) may be represented as

$$(1.4) \quad \mathbf{X}_t = \mathbf{F}(\mathbf{X}_{t-\tau}) + \mathbf{e}_t,$$

where $\mathbf{X}_t = {}^t(X_t, X_{t-\tau}, \dots, X_{t-(d-1)\tau})$, and $\mathbf{e}_t = {}^t(\varepsilon_t, 0, \dots, 0)$. We call model (1.1) the skeleton of model (1.4). Our goal is to estimate the correlation dimension of the skeleton by observing $\{X_t\}_{t=1,2,\dots,N}$. Smith (1992) proposed an estimator of the correlation dimension that seems to function well in the presence of observational noise. Unfortunately, his estimator does not work adequately in the case of dynamic noise. Further research on the effect of dynamic noise on the estimator of the correlation dimension is warranted.

In this paper, we propose a method for estimating the correlation dimension after filtering out the noise by the Nadaraya-Watson kernel type estimator and show consistency of the estimator under some assumptions. In order to filter out the noise, we take the same approach as Yonemoto and Yanagawa (2004), who proposed a method for estimating the Lyapunov exponent from nonlinear time series with dynamic noise and showed its consistency. Kawaguchi (2003) proved the consistency of the estimator of the correlation dimension that was proposed in Kawaguchi (2002) for data from deterministic system (1.1), employing the same approach as Serinko (1994).

This paper is organized as follows. We propose a method for estimating the correlation dimension in Section 2. In Section 3, the consistency of the estimator is proven.

2. The proposed procedure and its consistency

Suppose that $\{X_t\}_{t=1,2,\dots,N}$ are observed data from (1.3). We propose the following procedure for estimating the correlation dimension of the skeleton of model (1.4) from $\{X_t\}_{t=1,2,\dots,N}$.

Step 1. Estimate the embedding dimension d and delay time τ from $\{X_t\}_{t=1,2,\dots,N}$ using the procedure proposed by Yonemoto and Yanagawa (2001). Denote the estimated embedding dimension and delay time by \hat{d} and $\hat{\tau}$, respectively.

Step 2. Estimate the skeleton from $\{X_t\}_{t=1,2,\dots,N}$ by the Nadaraya-Watson kernel type estimator (Nadaraya (1964); Watson (1964)) using \hat{d} and $\hat{\tau}$. Then, generate $\{\hat{Y}_{N,t}\}_{t=1,2,\dots,K}$ from the estimated skeleton by giving an initial value.

Step 3. Estimate the correlation dimension from $\{\hat{Y}_{N,t}\}_{t=1,2,\dots,K}$. The details of each steps are as follows.

(Step 1 of the procedure)

For a positive integer $L \geq d\tau$, putting

$$CV(d, \tau) = \frac{1}{N - L + 1} \sum_{t=L}^N (X_t - \hat{F}_{\setminus t}(X_{t-\tau}, \dots, X_{t-d\tau}))^2$$

where $\hat{F}_{\setminus t}$ denotes the Nadaraya-Watson kernel type estimated regression function (Nadaraya (1964), Watson (1964)) with the t -th point deleted, that is,

$$\hat{F}_{\setminus t}(z) = \frac{1}{N - L} \sum_{s=L, s \neq t}^N X_s K(z - (X_{s-\tau}, \dots, X_{s-d\tau})) / \hat{f}_{\setminus t}(z)$$

for $z = (z_1, z_2, \dots, z_d)$, where the summation over s omit t in each cases, and

$$\hat{f}_{\setminus t} = \frac{1}{N - L} \sum_{s=L, s \neq t}^N K(z - (X_{s-\tau}, \dots, X_{s-d\tau})),$$

and

$$K(z) = \frac{1}{h_N^d} \prod_{i=1}^d \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z_i^2}{2h_N^2}\right),$$

where $h_N \in \mathbf{R}_+$ which is called the bandwidth. Then, the embedding dimension and delay time are estimated as follows, see Yonemoto and Yanagawa (2001) for more details.

- (i) Give sufficiently large integers $D(\geq d)$ and $T(\geq \tau)$, and set $L = DT$. Set $h_N = h_N(c) = c \times N^{-1/(2d+1)}$.
- (ii) For each $d \in \{1, 2, \dots, D\}$ and $\tau \in \{1, 2, \dots, T\}$ minimize $CV(d, \tau)|_{h_N=h_N(c)}$ with respect to $c \in C$ and put

$$\hat{C}V(d, \tau) = \min_{c \in C} CV(d, \tau) |_{h_N=h_N(c)},$$

where C denotes a parameter set.

- (iii) Then select $d \in \{1, 2, \dots, D\}$ and $\tau \in \{1, 2, \dots, T\}$ which attain the ‘minimum’ value of $\{\hat{C}V(d, \tau) : d \in \{1, 2, \dots, D\}, \tau \in \{1, 2, \dots, T\}\}$ as estimators of embedding dimension \hat{d} and delay time $\hat{\tau}$.

(Step 2 of the procedure)

First, we estimate $F(\mathbf{x})$ by

$$\hat{F}_N(\mathbf{x}) = \frac{\sum_{i=(\hat{d}-1)\hat{\tau}+1}^{N-\hat{\tau}} K_{h_N}(\mathbf{x} - \mathbf{X}_i) X_{i+\hat{\tau}}}{\sum_{i=(\hat{d}-1)\hat{\tau}+1}^{N-\hat{\tau}} K_{h_N}(\mathbf{x} - \mathbf{X}_i)},$$

where $\mathbf{x} = {}^t(x_1, x_2, \dots, x_{\hat{d}})$,

$$K_{h_N}(\mathbf{x} - \mathbf{X}_i) = \frac{1}{h_N^{\hat{d}}} K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_N}\right),$$

and $K(\mathbf{x}) = \prod_{i=1}^{\hat{d}} \frac{1}{\sqrt{2\pi}} \exp(-\frac{x_i^2}{2})$. The same value of h_N is employed as that used for estimating the embedding dimension and delay time. Next, put $\hat{\mathbf{F}}_N(\mathbf{x}) = {}^t(\hat{F}_N(\mathbf{x}), x_1, \dots, x_{\hat{d}-1})$. Giving an initial vector $\hat{\mathbf{Y}}_{N,0}$ selected randomly from data $\{\mathbf{X}_t\}_{t=1}^N$, we generate $\{\hat{\mathbf{Y}}_{N,t}\}_{t=1,2,\dots,K}$ by

$$(2.1) \quad \hat{\mathbf{Y}}_{N,t} = \hat{\mathbf{F}}_N(\hat{\mathbf{Y}}_{N,t-1}), \quad t = 1, 2, \dots, K,$$

where $\hat{\mathbf{Y}}_{N,t} = {}^t(\hat{Y}_{N,t}, \hat{Y}_{N,t-1}, \dots, \hat{Y}_{N,t-\hat{d}+1})$.

(Step 3 of the procedure)

Let

$$C_K(r, \hat{\mathbf{Y}}_N) = \binom{K}{2}^{-1} \sum_{i < j} I(\|\hat{\mathbf{Y}}_{N,i} - \hat{\mathbf{Y}}_{N,j}\| \leq r),$$

$$C_{K2}(r, \hat{\mathbf{Y}}_N) = \binom{K}{3}^{-1} \sum_{i \neq j, i \neq k, j \neq k} I(\|\hat{\mathbf{Y}}_{N,i} - \hat{\mathbf{Y}}_{N,j}\| \leq r, \|\hat{\mathbf{Y}}_{N,i} - \hat{\mathbf{Y}}_{N,k}\| \leq r),$$

and $r_j^{(M_{N,K})} = r_0 s^{M_{N,K} - [M_{N,K}/2] + j}$, ($j = 0, 1, \dots, L_{N,K} = \lfloor \frac{M_{N,K}}{2} \rfloor$) where some given $0 < s < 1$, $r_0 > 1$, and

$$(2.2) \quad M_{N,K} = \max\{m \in \mathbf{N}; C_{K2}(r_m, \hat{\mathbf{Y}}_N) \neq 0 \text{ for } r_m = r_0 s^m\}.$$

We propose an estimator of the correlation dimension as follows.

$$(2.3) \quad \hat{\nu}_{N,K} = \sum_{j=0}^{L_{N,K}} (u_j - \bar{u}) \log C_K(r_j^{(M_{N,K})}, \hat{\mathbf{Y}}_N) \bigg/ \sum_{j=0}^{L_{N,K}} (u_j - \bar{u})^2,$$

where $u_j = \log r_j^{(M_{N,K})}$ and $\bar{u} = (L_{N,K} + 1)^{-1} \sum_{j=0}^{L_{N,K}} u_j$.

Remark 2. The selection of $M_{N,K}$ is important in practice. Cutler (1990)'s selection of $M_{N,K}$ provides unstable estimates (Kawaguchi and Yanagawa (2001)). To improve the stability, we proposed in Kawaguchi (2002) to select $M_{N,K}$ in (2.2) with relation to $C_{K2}(r_m, \hat{Y}_N)$. It is shown there that the selection provides substantial stability to the estimates. Note that the stability and consistency of the estimator which will be proved below are irrelevant. In particular, Theorem 3.3 and Theorem 3.7 hold true for any $M_{N,K}$ such that $\lim_{K \rightarrow \infty} \lim_{N \rightarrow \infty} M_{N,K} = \infty$ a.e.

Our goal is to prove the consistency of the proposed estimator of the correlation dimension. More precisely, letting $\mathcal{E} \times \mu$ a product measure of a measure for dynamic noise and an invariant measure for dynamical system, we prove the following theorem.

THEOREM 2.1. *Under Assumption 1.1 and the assumptions that are given in the following section, it follows that for any $\varepsilon > 0$,*

$$\lim_{K \rightarrow \infty} \lim_{N \rightarrow \infty} \mathcal{E} \times \mu(|\hat{\nu}_{N,K} - \nu| > \varepsilon) = 0.$$

3. Proof of Theorem 2.1

In this section, we give a proof of Theorem 2.1. To begin with, we note the following lemma.

LEMMA 3.1. *We may assume d and τ are known in proving the consistency of $\hat{\nu}_{N,K}$.*

PROOF. Fueda and Yanagawa (2001) showed the consistency of the estimator of the embedding dimension and delay time. Therefore, the proof is trivial and we omit it. \square

We note also that the estimator is decomposed as follows.

LEMMA 3.2. (Serinko (1994))

$$\hat{\nu}_{N,K} = \nu + d_{N,K} + e_{N,K},$$

where

$$d_{N,K} = \frac{1}{S_{uu}} \sum_{j=0}^{L_{N,K}} \{\log C(r_j^{(M_{N,K})}) - \nu \log r_j^{(M_{N,K})}\} (u_j - \bar{u}),$$

$$e_{N,K} = \frac{1}{S_{uu}} \sum_{j=0}^{L_{N,K}} \{\log C_K(r_j^{(M_{N,K})}, \hat{Y}_N) - \log C(r_j^{(M_{N,K})})\} (u_j - \bar{u}), \quad \text{and}$$

$$S_{uu} = \sum_{j=0}^{L_{N,K}} (u_j - \bar{u})^2.$$

Next, we prove the convergence of $d_{N,K}$ in probability. We assume the following assumption.

ASSUMPTION 3.1. F is C^1 class on G .

Select an initial vector \mathbf{Y}_1 randomly from G° with uniform probability, and set $\hat{\mathbf{Y}}_{N,1} = \mathbf{Y}_1$.

THEOREM 3.1. Under Assumptions 1.1 and 3.1, it follows that for any $\varepsilon > 0$ and $t \in \{1, 2, \dots, K\}$,

$$\lim_{N \rightarrow \infty} \mathcal{E} \times \mu(\|\hat{\mathbf{Y}}_{N,t} - \mathbf{Y}_t\| > \varepsilon) = 0.$$

PROOF. The proof is given in a preprint by Yonemoto and Yanagawa (2004), but since it is not well circulated we will give a brief proof here by a mathematical induction. First, we prove the formula in the case of $t = 2$. Collomb (1984) proved that the Nadaraya-Watson kernel type estimator \hat{F}_N satisfies

$$(3.1) \quad \lim_{N \rightarrow \infty} \mathcal{E} \times \mu \left(\sup_{\mathbf{x} \in G} |\hat{F}_N(\mathbf{x}) - F(\mathbf{x})| > \varepsilon \right) = 0.$$

Hence, for any $\varepsilon > 0$,

$$\begin{aligned} \mathcal{E} \times \mu(|\hat{Y}_{N,2} - Y_2| > \varepsilon) &= \mathcal{E} \times \mu(|\hat{F}_N(\hat{\mathbf{Y}}_{N,1}) - F(\mathbf{Y}_1)| > \varepsilon) \\ &\leq \mathcal{E} \times \mu \left(\sup_{\mathbf{x} \in G} |\hat{F}_N(\mathbf{x}) - F(\mathbf{x})| > \varepsilon \right) \\ &\rightarrow 0 \quad (N \rightarrow \infty). \end{aligned}$$

This implies that $\mathcal{E} \times \mu(\|\hat{\mathbf{Y}}_{N,2} - \mathbf{Y}_2\| > \varepsilon) \rightarrow 0$ as $N \rightarrow \infty$.

Next, suppose that the formula holds at the case of $t = n - 1$, that is, $\hat{\mathbf{Y}}_{N,n-1} \rightarrow \mathbf{Y}_{n-1}$ in probability as $N \rightarrow \infty$. It holds that

$$\begin{aligned} \mathcal{E} \times \mu(|\hat{Y}_{N,n} - Y_n| > \varepsilon) &\leq \mathcal{E} \times \mu(|F(\hat{\mathbf{Y}}_{N,n-1}) - F(\mathbf{Y}_{n-1})| > \varepsilon/2) \\ &\quad + \mathcal{E} \times \mu(|\hat{F}_N(\hat{\mathbf{Y}}_{N,n-1}) - F(\hat{\mathbf{Y}}_{N,n-1})| > \varepsilon/2, \hat{\mathbf{Y}}_{N,n-1} \in G^\circ) \\ &\quad + \mathcal{E} \times \mu(\hat{\mathbf{Y}}_{N,n-1} \notin G^\circ). \end{aligned}$$

Since F is continuous on the compact set G , the first term of the righthand side of the inequality goes to zero as $N \rightarrow \infty$. The second term goes to zero as $N \rightarrow \infty$ from (3.1), and the third term goes to zero as $N \rightarrow \infty$ from Assumption 1.1 and $\mathbf{Y}_1 \in G^\circ$. Therefore, $\mathcal{E} \times \mu(\|\hat{\mathbf{Y}}_{N,n} - \mathbf{Y}_n\| > \varepsilon) \rightarrow 0$ as $N \rightarrow \infty$. The proof is completed. \square

Let $C_{K2}(r, \mathbf{Y}) = \binom{K}{3}^{-1} \sum_{i \neq j, i \neq k, j \neq k} I(\|\mathbf{Y}_i - \mathbf{Y}_j\| \leq r, \|\mathbf{Y}_i - \mathbf{Y}_k\| \leq r)$. We assume the following assumption.

ASSUMPTION 3.2. For any $r > 0$,

$$\lim_{K \rightarrow \infty} \mu(C_{K2}(r, \mathbf{Y}) > 0) = 1.$$

THEOREM 3.2. Under Assumptions 1.1, 3.1, and 3.2, for $M_{N,K}$ in (2.2) it follows that

$$\mathcal{E} \times \mu \left(\lim_{K \rightarrow \infty} \lim_{N \rightarrow \infty} M_{N,K} = \infty \right) = 1.$$

PROOF. For any integers m, N_0 and K_0 , let

$$\begin{aligned} A(m, N_0, K_0) = \{(\omega, \omega') \in \Omega \times \Omega'; C_{K2}(r_l, \hat{\mathbf{Y}}_N) > 0, \\ \text{for all } N \geq N_0, K \geq K_0, \text{ and } l = 0, 1, \dots, m, \\ \text{where } r_l = r_0 s^l\}. \end{aligned}$$

Then, $A(m, N_0, K_0) \subset \{M_{N,K} \geq m \text{ for } N \geq N_0 \text{ and } K \geq K_0\}$, and therefore for any m

$$\begin{aligned} \mathcal{E} \times \mu \left(\lim_{K \rightarrow \infty} \lim_{N \rightarrow \infty} M_{N,K} \geq m \right) &\geq \mathcal{E} \times \mu \left(\bigcap_{K_0=1}^{\infty} \bigcap_{N_0=1}^{\infty} A(m, N_0, K_0) \right) \\ &= \lim_{K_0 \rightarrow \infty} \lim_{N_0 \rightarrow \infty} \mathcal{E} \times \mu(A(m, N_0, K_0)). \end{aligned}$$

Thus, in order to prove the theorem, we must show for any m that

$$\lim_{K_0 \rightarrow \infty} \lim_{N_0 \rightarrow \infty} \mathcal{E} \times \mu(A(m, N_0, K_0)) = 1.$$

For any $r > 0$ and $t \in \{1, 2, \dots, K\}$, it holds that

$$\begin{aligned} \mathcal{E} \times \mu(C_{K2}(r, \hat{\mathbf{Y}}_N) > 0) \\ = \mathcal{E} \times \mu(C_{K2}(r, \hat{\mathbf{Y}}_N) > 0 \text{ and } \|\hat{\mathbf{Y}}_{N,t} - \mathbf{Y}_t\| > r/3) \\ + \mathcal{E} \times \mu(C_{K2}(r, \hat{\mathbf{Y}}_N) > 0 \text{ and } \|\hat{\mathbf{Y}}_{N,t} - \mathbf{Y}_t\| \leq r/3). \end{aligned}$$

From Theorem 3.1,

$$\begin{aligned} \mathcal{E} \times \mu(C_{K2}(r, \hat{\mathbf{Y}}_N) > 0 \text{ and } \|\hat{\mathbf{Y}}_{N,t} - \mathbf{Y}_t\| > r/3) &\leq \mathcal{E} \times \mu(\|\hat{\mathbf{Y}}_{N,t} - \mathbf{Y}_t\| > r/3) \\ &\rightarrow 0 \quad (N \rightarrow \infty). \end{aligned}$$

Furthermore,

$$\begin{aligned} \mathcal{E} \times \mu(C_{K2}(r, \hat{\mathbf{Y}}_N) > 0 \text{ and } \|\hat{\mathbf{Y}}_{N,t} - \mathbf{Y}_t\| \leq r/3) \\ \geq \mathcal{E} \times \mu \left(\binom{K}{3}^{-1} \sum_{i \neq j, i \neq k, j \neq k}^K \right. \\ \left. \times I(\|\hat{\mathbf{Y}}_{N,i} - \mathbf{Y}_i\| + \|\mathbf{Y}_i - \mathbf{Y}_j\| + \|\hat{\mathbf{Y}}_{N,j} - \mathbf{Y}_j\| \leq r, \right. \end{aligned}$$

$$\begin{aligned}
 & \left(\|\hat{\mathbf{Y}}_{N,i} - \mathbf{Y}_i\| + \|\mathbf{Y}_i - \mathbf{Y}_k\| + \|\hat{\mathbf{Y}}_{N,k} - \mathbf{Y}_k\| \leq r \right) > 0 \\
 & \qquad \qquad \qquad \text{and } \|\hat{\mathbf{Y}}_{N,t} - \mathbf{Y}_t\| \leq r/3 \Big) \\
 & \geq \mathcal{E} \times \mu \left(\binom{K}{3}^{-1} \sum_{i \neq j, i \neq k, j \neq k}^K I(\|\mathbf{Y}_i - \mathbf{Y}_j\| \leq r/3, \|\mathbf{Y}_i - \mathbf{Y}_k\| \leq r/3) > 0 \right. \\
 & \qquad \qquad \qquad \left. \text{and } \|\hat{\mathbf{Y}}_{N,t} - \mathbf{Y}_t\| \leq r/3 \right) \\
 & \geq \mathcal{E} \times \mu(C_{K2}(r/3, \mathbf{Y}) > 0) + \mathcal{E} \times \mu(\|\hat{\mathbf{Y}}_{N,t} - \mathbf{Y}_t\| \leq r/3) - 1.
 \end{aligned}$$

From Theorem 3.1 and Assumption 3.2,

$$\begin{aligned}
 & \mathcal{E} \times \mu(C_{K2}(r/3, \mathbf{Y}) > 0) + \mathcal{E} \times \mu(\|\hat{\mathbf{Y}}_{N,t} - \mathbf{Y}_t\| \leq r/3) - 1 \\
 & \quad \rightarrow \mathcal{E} \times \mu(C_{K2}(r/3, \mathbf{Y}) > 0) \quad (N \rightarrow \infty) \\
 & \quad \rightarrow 1 \quad (K \rightarrow \infty).
 \end{aligned}$$

Hence, for any m ,

$$\lim_{K \rightarrow \infty} \lim_{N \rightarrow \infty} \mathcal{E} \times \mu(C_{K2}(r_m, \hat{\mathbf{Y}}_N) > 0) = 1.$$

From the definition of $A(m, N_0, K_0)$, the proof is completed. \square

THEOREM 3.3. *Under Assumptions 1.1, 3.1, and 3.2, it follows that for any $\varepsilon > 0$*

$$\lim_{K \rightarrow \infty} \lim_{N \rightarrow \infty} \mathcal{E} \times \mu(|d_{N,K}| > \varepsilon) = 0.$$

PROOF. First, from the definition of the correlation dimension,

$$\frac{\log C(r) - \nu \log r}{\log r} \rightarrow 0 \quad (r \rightarrow 0).$$

Thus, there exists a positive real $A(r) \rightarrow 0$ as $r \rightarrow 0$ such that for sufficiently small $r > 0$,

$$|\log C(r) - \nu \log r| \leq A(r) |\log r|.$$

Next, it follows by the Cauchy-Schwarz inequality that

$$\begin{aligned}
 |d_{N,K}| & \leq \frac{1}{S_{uu}} \sqrt{\sum_{j=0}^{L_{N,K}} \{\log C(r_j^{(M_{N,K})}) - \nu \log r_j^{(M_{N,K})}\}^2} \sqrt{\sum_{j=0}^{L_{N,K}} (u_j - \bar{u})^2} \\
 & \leq \frac{1}{\sqrt{S_{uu}}} \sqrt{\sum_{j=0}^{L_{N,K}} \{A(r_j^{(M_{N,K})}) \log r_j^{(M_{N,K})}\}^2} \\
 & \leq \frac{1}{\sqrt{S_{uu}}} \times \max_{0 \leq j \leq L_{N,K}} A(r_j^{(M_{N,K})}) \times \sqrt{\sum_{j=0}^{L_{N,K}} \{\log r_j^{(M_{N,K})}\}^2}.
 \end{aligned}$$

It follows that

$$S_{uu} = \frac{1}{12} L_{N,K} (L_{N,K} + 1) (L_{N,K} + 2) (\log s)^2.$$

Thus, from Theorem 3.2 and the definition of $L_{N,K}$,

$$\sqrt{S_{uu}} = O_p(M_{N,K}^{3/2}).$$

On the other hand, from Theorem 3.2,

$$\sqrt{\sum_{j=0}^{L_{N,K}} \{\log r_j^{(M_{N,K})}\}^2} = O_p(M_{N,K}^{3/2}),$$

and for any $j \in \{0, 1, \dots, L_{N,K}\}$,

$$\mathcal{E} \times \mu \left(\lim_{K \rightarrow \infty} \lim_{N \rightarrow \infty} A(r_j^{(M_{N,K})}) = 0 \right) = 1.$$

Therefore,

$$\mathcal{E} \times \mu \left(\lim_{K \rightarrow \infty} \lim_{N \rightarrow \infty} \left\{ \frac{1}{\sqrt{S_{uu}}} \max_{0 \leq j \leq L_{N,K}} A(r_j^{(M_{N,K})}) \times \sqrt{\sum_{j=0}^{L_{N,K}} \{\log r_j^{(M_{N,K})}\}^2} \right\} = 0 \right) = 1$$

and

$$\mathcal{E} \times \mu \left(\lim_{K \rightarrow \infty} \lim_{N \rightarrow \infty} |d_{N,K}| = 0 \right) = 1.$$

Convergence with probability 1 implies convergence in probability, therefore the proof is completed. \square

Next, we prove the convergence of $e_{N,K}$ in probability. For the purpose, we use Theorem 3.4 below proved by Serinko (1994) under the following notation and assumptions. Let $\alpha = \{A_1, A_2, \dots, A_m\}$ and $\beta = \{B_1, B_2, \dots, B_n\}$ be finite measurable partitions of Ω . From these one may construct the following partitions:

- (i) $\alpha \vee \beta = \{A \cap B; A \in \alpha, B \in \beta\}$
- (ii) $\mathbf{F}^{-1}\alpha = \{\mathbf{F}^{-1}A; A \in \alpha\}$
- (iii) $\alpha_r^s = \mathbf{F}^{(-r)}\alpha \vee \mathbf{F}^{(-r-1)}\alpha \vee \dots \vee \mathbf{F}^{(-s+1)}\alpha \vee \mathbf{F}^{(-s)}\alpha$ ($r, s \in \mathbf{N}$, s.t. $r < s$), where $\mathbf{F}^{(k)}$ denotes k times convolution of \mathbf{F} .

Let \mathcal{F}_r^s denotes the σ -algebra generated by α_r^s ($r, s \in \mathbf{N}$, s.t. $r < s$) and \mathcal{F}_0^∞ denotes the smallest σ -algebra which contains all of the \mathcal{F}_r^s ($r, s \in \mathbf{N}$, s.t. $r < s$).

DEFINITION 3.1. (Generator)

$$\alpha \text{ is generator} \Leftrightarrow \mathcal{F}_0^\infty = \mathcal{F}.$$

DEFINITION 3.2. (Weak Bernoulli) A measurable partition α is said to be weak Bernoulli for dynamical system $(\Omega, \mathcal{F}, \mu, \mathbf{F})$ if

$$\beta_k = \sup_{r,s \in \mathbb{N}} \sum_{A \in \alpha_0^r} \sum_{B \in \alpha_{r+k}^{r+s+k}} |\mu(A \cap B) - \mu(A)\mu(B)|$$

goes to zero as $k \rightarrow \infty$. The β_k 's are called the mixing coefficients.

We assume the following assumptions for the dynamical system $(\Omega, \mathcal{F}, \mu, \mathbf{F})$.

ASSUMPTION 3.3. $(\Omega, \mathcal{F}, \mu, \mathbf{F})$ has the measurable partition α which is weak Bernoulli and generator.

ASSUMPTION 3.4. $(\Omega, \mathcal{F}, \mu, \mathbf{F})$ is such that the mixing coefficients satisfy

$$\beta_k^{\delta/(2+\delta)} = O(k^{-(1+\varepsilon)})$$

for some $\delta > 0$ and $0 < \varepsilon < 1$.

Let $\eta^{(l)}(r) = \|\mu(\bar{B}_r(\mathbf{Y}_j)) - \mu(\bar{B}_r(\mathbf{Y}_j^{(l)}))\|_2^2$, where $\bar{B}_r(\mathbf{y}) = \{\mathbf{x}; \|\mathbf{x} - \mathbf{y}\| \leq r\}$ and $\mathbf{Y}_j^{(l)} = E[\mathbf{Y}_j | \alpha_j^{j+l}]$, $j = 1, 2, \dots$. Let $\bar{\eta}^{(l)} = \sup_r \eta^{(l)}(r)$.

ASSUMPTION 3.5. $(\Omega, \mathcal{F}, \mu, \mathbf{F})$ is such that

$$\bar{\eta}^{(l)1/2} = o(l^{-(1+\gamma)})$$

for some $\gamma > 0$, $\gamma/(1 + \gamma) > \varepsilon$.

Let $\psi_k^{(l)}(r) = \|I(\|\mathbf{Y}_i - \mathbf{Y}_j\| \leq r) - I(\|\mathbf{Y}_i^{(l)} - \mathbf{Y}_j^{(l)}\| \leq r)\|_2^2$ for $i, j = 1, 2, \dots$ and $k = |i - j|$, and $\bar{\psi}_k^{(l)} = \sup_r \psi_k^{(l)}(r)$.

ASSUMPTION 3.6. $(\Omega, \mathcal{F}, \mu, \mathbf{F})$ is such that for any sequence of reals $\{c_n\}_{n=0}^\infty$ satisfying $\lim_{n \rightarrow \infty} c_n = \infty$ and $c_n = o(n^{1/2})$, one has

$$\sum_{k=0}^{n-1} \bar{\psi}_k^{(c_n)1/2} = o(n^{1/2}).$$

Let $b_K = (\frac{1}{K})^{1/2(\nu+\varepsilon_0)}$ where $\varepsilon_0 > 0$, and

$$C_K(r, \mathbf{Y}) = \binom{K}{2}^{-1} \sum_{i < j}^K I(\|\mathbf{Y}_i - \mathbf{Y}_j\| \leq r).$$

THEOREM 3.4. (Serinko (1994)) *If $(\Omega, \mathcal{F}, \mu, \mathbf{F})$ satisfies Assumptions 3.3, 3.4, 3.5, and 3.6, then whenever ν exists, the condition*

$$\lim_{K \rightarrow \infty} r_K = 0 \quad \text{and} \quad \limsup_{K \rightarrow \infty} \frac{b_K}{r_K} < \infty$$

imply

$$\lim_{K \rightarrow \infty} \mu \left(\left| \frac{C_K(r_K, \mathbf{Y}) - C(r_K)}{C(r_K)} \right| > \varepsilon \right) = 0.$$

Let $M_K = \lim_{N \rightarrow \infty} M_{N,K}$. We may prove that M_K is given by $M_K = \max\{m \in \mathbf{N}; C_{K2}(r_m, \mathbf{Y}) \neq 0, \text{ for } r_m = r_0 s^m\}$ following the proof of Theorem 3.6. We assume the following assumption for M_K .

ASSUMPTION 3.7. For some $\varepsilon_0 > 0$ and some $\delta < \frac{1}{d+\varepsilon_0}$, there exists $K_0 \in \mathbf{N}$ such that for any $K > K_0$,

$$M_K > \frac{\delta \log K}{2|\log s|} \quad \text{a.e.}$$

LEMMA 3.3. *Under Assumption 3.7, let $r_K = r_0 s^{M_K}$, then*

$$\lim_{K \rightarrow \infty} r_K = 0 \quad \text{and} \quad \limsup_{K \rightarrow \infty} \frac{b_K}{r_K} < \infty \quad \text{a.e.}$$

PROOF. The first part of lemma is obtained immediately from Theorem 3.2. Setting $r'_K = K^{-1/2(d+\varepsilon_0)}$, Serinko (1994) proved $\limsup_{K \rightarrow \infty} \frac{b_K}{r'_K} < \infty$. Thus, the second part of lemma may be proved if we prove $r_K > r'_K$ a.e. for any $K > K_0$.

From Assumption 3.7, we have

$$\begin{aligned} \log \frac{r_K}{r'_K} &= \log \frac{r_0 s^{M_K}}{K^{-1/2(d+\varepsilon_0)}} \\ &> \log r_0 + \left(-\delta + \frac{1}{d+\varepsilon_0} \right) \frac{1}{2} \log K > 0. \end{aligned} \quad \square$$

THEOREM 3.5. *Under the assumptions of Theorem 3.1, it follows that for a fixed K , any $\varepsilon > 0$ and $r > 0$*

$$\lim_{N \rightarrow \infty} \mathcal{E} \times \mu(|C_K(r, \hat{\mathbf{Y}}_N) - C_K(r, \mathbf{Y})| > \varepsilon) = 0.$$

PROOF. It follows by the Markov inequality that

$$\begin{aligned} & \mathcal{E} \times \mu(|C_K(r, \hat{\mathbf{Y}}_N) - C_K(r, \mathbf{Y})| > \varepsilon) \\ & \leq \frac{1}{\varepsilon} E[|C_K(r, \hat{\mathbf{Y}}_N) - C_K(r, \mathbf{Y})|] \\ & \leq \frac{1}{\varepsilon} \binom{K}{2}^{-1} \sum_{i < j}^K E[|I(\|\hat{\mathbf{Y}}_{N,i} - \hat{\mathbf{Y}}_{N,j}\| \leq r) - I(\|\mathbf{Y}_i - \mathbf{Y}_j\| \leq r)|]. \end{aligned}$$

It follows by the triangle inequality that for $i, j \in \{1, 2, \dots, K\}$ ($i < j$)

$$\begin{aligned} (3.2) \quad & E[|I(\|\hat{\mathbf{Y}}_{N,i} - \hat{\mathbf{Y}}_{N,j}\| \leq r) - I(\|\mathbf{Y}_i - \mathbf{Y}_j\| \leq r)|] \\ & \leq E[|I(\|\hat{\mathbf{Y}}_{N,i} - \hat{\mathbf{Y}}_{N,j}\| \leq r) - \phi_\delta(\hat{\mathbf{Y}}_{N,i} - \hat{\mathbf{Y}}_{N,j})|] \\ & \quad + E[|\phi_\delta(\mathbf{Y}_i - \mathbf{Y}_j) - I(\|\mathbf{Y}_i - \mathbf{Y}_j\| \leq r)|] \\ & \quad + E[|\phi_\delta(\hat{\mathbf{Y}}_{N,i} - \hat{\mathbf{Y}}_{N,j}) - \phi_\delta(\mathbf{Y}_i - \mathbf{Y}_j)|] \end{aligned}$$

where for any $\delta > 0$,

$$\phi_\delta(\mathbf{x}) = \begin{cases} 0 & \text{if } \|\mathbf{x}\| \geq r \\ 1 & \text{if } \|\mathbf{x}\| \leq r - \delta \end{cases},$$

$0 \leq \phi_\delta(\mathbf{x}) \leq 1$ and $\phi_\delta \in C^1$.

The first term of the righthand side of the inequality (3.2) is

$$\begin{aligned} & \iint (I(\|\mathbf{x} - \mathbf{y}\| \leq r) - \phi_\delta(\mathbf{x} - \mathbf{y})) d(\mathcal{E} \times \mu)(\mathbf{x}) d(\mathcal{E} \times \mu)(\mathbf{y}) \\ & \leq \iint_{r-\delta < \|\mathbf{x}-\mathbf{y}\| < r} d(\mathcal{E} \times \mu)(\mathbf{x}) d(\mathcal{E} \times \mu)(\mathbf{y}) \\ & = \mathcal{E} \times \mu(r - \delta < \|\hat{\mathbf{Y}}_{N,i} - \hat{\mathbf{Y}}_{N,j}\| < r). \end{aligned}$$

Thus, for any $\varepsilon_1 > 0$, there exists $\delta_1 \in \mathbf{N}$ such that for any $\delta < \delta_1$

$$E[|I(\|\hat{\mathbf{Y}}_{N,i} - \hat{\mathbf{Y}}_{N,j}\| \leq r) - \phi_\delta(\hat{\mathbf{Y}}_{N,i} - \hat{\mathbf{Y}}_{N,j})|] < \varepsilon_1.$$

By the same way, for any $\varepsilon_2 > 0$, there exists $\delta_2 \in \mathbf{N}$ such that for any $\delta < \delta_2$

$$E[|\phi_\delta(\mathbf{Y}_i - \mathbf{Y}_j) - I(\|\mathbf{Y}_i - \mathbf{Y}_j\| \leq r)|] < \varepsilon_2.$$

Therefore, from the continuity of ϕ_δ and Theorem 3.1, for any $\delta < \min\{\delta_1, \delta_2\}$,

$$\begin{aligned} & E[|I(\|\hat{\mathbf{Y}}_{N,i} - \hat{\mathbf{Y}}_{N,j}\| \leq r) - I(\|\mathbf{Y}_i - \mathbf{Y}_j\| \leq r)|] \\ & < \varepsilon_1 + \varepsilon_2 + E[|\phi_\delta(\hat{\mathbf{Y}}_{N,i} - \hat{\mathbf{Y}}_{N,j}) - \phi_\delta(\mathbf{Y}_i - \mathbf{Y}_j)|] \\ & \rightarrow \varepsilon_1 + \varepsilon_2 \quad (N \rightarrow \infty). \end{aligned}$$

Consequently, $E[|I(\|\hat{\mathbf{Y}}_{N,i} - \hat{\mathbf{Y}}_{N,j}\| \leq r) - I(\|\mathbf{Y}_i - \mathbf{Y}_j\| \leq r)|] \rightarrow 0$ as $N \rightarrow \infty$ for $i, j \in \{1, 2, \dots, K\}$ ($i < j$). Hence, the proof is completed. \square

THEOREM 3.6. *Under the assumptions of Theorem 3.1, and the condition and assumptions of Theorem 3.4, it follows that for any $\varepsilon > 0$,*

$$\lim_{K \rightarrow \infty} \lim_{N \rightarrow \infty} \mathcal{E} \times \mu(|\log C_K(r_K, \hat{\mathbf{Y}}_N) - \log C(r_K)| > \varepsilon) = 0.$$

PROOF.

$$\begin{aligned} & \mathcal{E} \times \mu \left(\left| \log \frac{C_K(r_K, \hat{\mathbf{Y}}_N)}{C(r_K)} \right| > \varepsilon \right) \\ &= \mathcal{E} \times \mu \left(\frac{C_K(r_K, \hat{\mathbf{Y}}_N) - C(r_K)}{C(r_K)} > e^\varepsilon - 1 \right) \\ & \quad + \mathcal{E} \times \mu \left(\frac{-C_K(r_K, \hat{\mathbf{Y}}_N) + C(r_K)}{C(r_K)} > 1 - e^{-\varepsilon} \right). \end{aligned}$$

For any $\varepsilon' > 0$,

$$\begin{aligned} & \mathcal{E} \times \mu \left(\left| \frac{C_K(r_K, \hat{\mathbf{Y}}_N) - C(r_K)}{C(r_K)} \right| > \varepsilon' \right) \\ & \leq \mathcal{E} \times \mu \left(\left| \frac{C_K(r_K, \hat{\mathbf{Y}}_N) - C_K(r_K, \mathbf{Y})}{C(r_K)} \right| > \varepsilon'/2 \right) \\ & \quad + \mathcal{E} \times \mu \left(\left| \frac{C_K(r_K, \mathbf{Y}) - C(r_K)}{C(r_K)} \right| > \varepsilon'/2 \right). \end{aligned}$$

From Theorem 3.5, the first term goes to zero as $N \rightarrow \infty$ for a fixed K . From Theorem 3.4, the second term goes to zero as $K \rightarrow \infty$. \square

THEOREM 3.7. *Under the assumption of Theorem 3.2, Theorem 3.6, and Assumption 3.7, it follows that for any $\varepsilon > 0$,*

$$\lim_{K \rightarrow \infty} \lim_{N \rightarrow \infty} \mathcal{E} \times \mu(|e_{N,K}| > \varepsilon) = 0.$$

PROOF. It follows by the Cauchy-Schwarz inequality that

$$\begin{aligned} |e_{N,K}| & \leq \frac{1}{S_{uu}} \sqrt{\sum_{j=0}^{L_{N,K}} \{\log C_K(r_j^{(M_{N,K})}, \hat{\mathbf{Y}}_N) - \log C(r_j^{(M_{N,K})})\}^2} \sqrt{\sum_{j=0}^{L_{N,K}} (u_j - \bar{u})^2} \\ & \leq \frac{1}{\sqrt{S_{uu}}} \sqrt{L_{N,K} + 1} \max_{0 \leq j \leq L_{N,K}} |\log C_K(r_j^{(M_{N,K})}, \hat{\mathbf{Y}}_N) - \log C(r_j^{(M_{N,K})})| \\ & = \left\{ \frac{1}{12} (\log s)^2 L_{N,K} (L_{N,K} + 2) \right\}^{-1/2} \\ & \quad \times \max_{0 \leq j \leq L_{N,K}} |\log C_K(r_j^{(M_{N,K})}, \hat{\mathbf{Y}}_N) - \log C(r_j^{(M_{N,K})})|. \end{aligned}$$

Moreover,

$$\begin{aligned} & \mathcal{E} \times \mu \left(\left\{ \frac{1}{12} (\log s)^2 L_{N,K} (L_{N,K} + 2) \right\}^{-1/2} \right. \\ & \quad \times \left. \max_{0 \leq j \leq L_{N,K}} |\log C_K(r_j^{(M_{N,K})}, \hat{\mathbf{Y}}_N) - \log C(r_j^{(M_{N,K})})| > \varepsilon \right) \\ & \leq \mathcal{E} \times \mu \left(\left\{ \frac{1}{12} (\log s)^2 L_{N,K} (L_{N,K} + 2) \right\}^{-1/2} > 1 \right) \\ & \quad + \mathcal{E} \times \mu \left(\max_{0 \leq j \leq L_{N,K}} |\log C_K(r_j^{(M_{N,K})}, \hat{\mathbf{Y}}_N) - \log C(r_j^{(M_{N,K})})| > \varepsilon \right). \end{aligned}$$

From Theorem 3.2,

$$\lim_{K \rightarrow \infty} \lim_{N \rightarrow \infty} \mathcal{E} \times \mu \left(\left\{ \frac{1}{12} (\log s)^2 L_{N,K} (L_{N,K} + 2) \right\}^{-1/2} > 1 \right) = 0.$$

For any $j \in \{0, 1, \dots, L_{N,K}\}$

$$\begin{aligned} & \mathcal{E} \times \mu(|\log C_K(r_j^{(M_{N,K})}, \hat{\mathbf{Y}}_N) - \log C(r_j^{(M_{N,K})})| > \varepsilon) \\ & \leq \mathcal{E} \times \mu(|\log C_K(r_j^{(M_{N,K})}, \hat{\mathbf{Y}}_N) - \log C_K(r_j^{(M_K)}, \hat{\mathbf{Y}}_N)| > \varepsilon/3) \\ & \quad + \mathcal{E} \times \mu(|\log C(r_j^{(M_{N,K})}) - \log C(r_j^{(M_K)})| > \varepsilon/3) \\ & \quad + \mathcal{E} \times \mu(|\log C_K(r_j^{(M_K)}, \hat{\mathbf{Y}}_N) - \log C(r_j^{(M_K)})| > \varepsilon/3). \end{aligned}$$

From right continuity of $C_K(r, Y)$, for any $\varepsilon > 0$, $\varepsilon_1 > 0$, and a fixed K

$$\begin{aligned} & \mathcal{E} \times \mu \left(\exists N_1 \in \mathbf{N}, \forall N > N_1, \right. \\ & \quad \left. \sup_y |\log C_K(r_j^{(M_{N,K})}, y) - \log C_K(r_j^{(M_K)}, y)| > \varepsilon \right) < \varepsilon_1. \end{aligned}$$

Thus,

$$\begin{aligned} & \mathcal{E} \times \mu(\exists N_1 \in \mathbf{N}, \forall N > N_1, \\ & \quad |\log C_K(r_j^{(M_{N,K})}, \hat{\mathbf{Y}}_N) - \log C_K(r_j^{(M_K)}, \hat{\mathbf{Y}}_N)| > \varepsilon) < \varepsilon_1. \end{aligned}$$

Therefore, there exists $N_1 \in \mathbf{N}$ such that for any $N > N_1$ and a fixed K ,

$$\mathcal{E} \times \mu(|\log C_K(r_j^{(M_{N,K})}, \hat{\mathbf{Y}}_N) - \log C_K(r_j^{(M_K)}, \hat{\mathbf{Y}}_N)| > \varepsilon) < \varepsilon_1.$$

By the same way, from right continuity of $C(r)$, for any $\varepsilon > 0$ and $\varepsilon_2 > 0$, there exists $N_2 \in \mathbf{N}$ such that for any $N > N_2$ and a fixed K

$$\mathcal{E} \times \mu(|\log C(r_j^{(M_{N,K})}) - \log C(r_j^{(M_K)})| > \varepsilon) < \varepsilon_2.$$

On the other hand, from Lemma 3.3 and Theorem 3.6, for any $\varepsilon > 0$ and $\varepsilon_3 > 0$, there exists $N_3 \in \mathbf{N}$ and $K_0 \in \mathbf{N}$ such that for any $N > N_3$ and $K > K_0$,

$$\mathcal{E} \times \mu(|\log C_K(r_j^{(M_K)}, \hat{\mathbf{Y}}_N) - \log C(r_j^{(M_K)})| > \varepsilon) < \varepsilon_3.$$

Thus, for any $N > \max\{N_1, N_2, N_3\}$ and $K > K_0$

$$\mathcal{E} \times \mu(|\log C_K(r_j^{(M_{N,K})}, \hat{\mathbf{Y}}_N) - \log C(r_j^{(M_{N,K})})| > \varepsilon) \leq \varepsilon_1 + \varepsilon_2 + \varepsilon_3.$$

Consequently,

$$\lim_{K \rightarrow \infty} \lim_{N \rightarrow \infty} \mathcal{E} \times \mu \left(\max_{0 \leq j \leq L_{N,K}} |\log C_K(r_j^{(M_{N,K})}, \hat{\mathbf{Y}}_N) - \log C(r_j^{(M_{N,K})})| > \varepsilon \right) = 0.$$

Hence, the proof is completed. \square

PROOF OF THEOREM 2.1. From Lemma 3.2, for any $\varepsilon > 0$,

$$\mathcal{E} \times \mu(|\hat{\nu}_{N,K} - \nu| > \varepsilon) \leq \mathcal{E} \times \mu(|d_{N,K}| > \varepsilon/2) + \mathcal{E} \times \mu(|e_{N,K}| > \varepsilon/2).$$

Hence from Theorem 3.3 and Theorem 3.7,

$$\lim_{K \rightarrow \infty} \lim_{N \rightarrow \infty} \mathcal{E} \times \mu(|\hat{\nu}_{N,K} - \nu| > \varepsilon) = 0.$$

The proof of Theorem 2.1 is completed. \square

4. Discussion

One referee commented on Assumption 1.1. Many data generated from chaotic systems are scattered in a compact space, so we expect that this assumption is satisfied. More study is clearly needed to construct a practical example. We would like to undertake this task in a follow-up paper.

Acknowledgements

The authors express their gratitude to the referee for careful reading and many helpful comments and suggestions.

REFERENCES

- An, H. Z. and Huang, F. C. (1996). The geometrical ergodicity of nonlinear autoregressive models, *Statistica Sinica*, **6**, 943–956.
- Carlsson, N. (2002). A contractivity condition for iterated function systems, *Journal of Theoretical Probability*, **15**, 613–630.
- Collomb, G. (1984). Propriétés de convergence presque complète du prédicteur à noyau, *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **66**, 441–460.
- Cutler, C. D. (1990). Some results on the behavior and estimation of the fractal dimensions of distributions on attractor, *Journal of Statistical Physics*, **62**, 651–708.
- Diks, C. (1996). Estimating invariants of noisy attractors, *Physical review E*, **53**, 4263–4266.
- Fueda, K. and Yanagawa, T. (2001). Estimating the embedding dimension and delay time from chaotic time series with dynamic noise, *J. Japan Statist. Soc.*, **31**(1), 27–38.

- Grassberger, P. and Procaccia, I. (1983a). Characterization for strange attractors, *Physical Review Letters*, **50**(5), 346–349.
- Grassberger, P. and Procaccia, I. (1983b). Measuring the strangeness of strange attractors, *Physica D*, **9**(5), 189–208.
- Judd, K. (1992). An improved estimator of dimension and some comments on providing confidence intervals, *Physica D*, **56**, 216–228.
- Kawaguchi, A. (2002). Estimating the correlation dimension from chaotic dynamical systems by U-statistics, *Bulletin of Informatics and Cybernetics*, **34**(2), 143–150.
- Kawaguchi, A. (2003). Consistency of the correlation dimension estimator from a chaotic system, *Bulletin of Informatics and Cybernetics*, **35**(1–2), 27–34.
- Kawaguchi, A. and Yanagawa, T. (2001). Estimating correlation dimension in chaotic time series, *Bulletin of Informatics and Cybernetics*, **33**(1–2), 63–71.
- Kugiumtzis, D. (1997). Correction of the correlation dimension for noisy time series, *International Journal of Bifurcation and Chaos*, **7**(6), 1283–1294.
- Nadaraya, E. A. (1964). On estimating regression, *Theory of Probability and Its Applications*, **9**, 141–142.
- Oltmans, H. and Verheijen, P. J. T. (1997). The influence of noise on power law scaling functions and an algorithm for dimension estimations, *Physical Review E*, **56**, 1160.
- Schreiber, T. (1993). Determination of the noise level of chaotic time series, *Physical Review E*, **48**(1), 13–16.
- Serinko, R. J. (1994). A consistent approach to least squares estimation of correlation dimension in weak Bernoulli dynamical systems, *The Annals of Applied Probability*, **4**(4), 1234–1254.
- Smith, R. L. (1992). Estimation dimension in noisy chaotic time series, *Journal of the Royal Statistical Society B*, **54**(2), 329–351.
- Takens, F. (1985). On the numerical determination of the dimension of an attractor, *Lecture Notes in Mathematics*, **1125**, 99–106.
- Watson, G. S. (1964). Smooth regression analysis, *Sankhyā, Ser. A*, **26**, 359–372.
- Yonemoto, K. and Yanagawa, T. (2001). Estimating the embedding dimension and delay time of chaotic time series by an autoregressive model, *Bulletin of Informatics and Cybernetics*, **33**(1–2), 53–62.
- Yonemoto, K. and Yanagawa, T. (2004). Estimating the Lyapunov exponent from chaotic time series with dynamic noise, *MHF Preprint Series*, 2004–1, Kyushu University.