

Identification of Chaos Phenomenon using the Classical Nonparametric Tests¹⁾

Young-Sun Park²⁾ · Hang-Suk Choi³⁾ · Eun-Sun Choi⁴⁾
Kyung-Joon Cha⁵⁾ · Moon-Il Park⁶⁾ · Jae-Eung Oh⁷⁾

Abstract

The data resulting from a deterministic dynamic system may often appear to be random. However, it is important to distinguish a deterministic and a random processes for statistical analysis. In this paper, we propose a nonparametric test procedure to distinguish a noisy chaos from i.i.d. random process. The proposed procedure can be easily implemented by computer. We notice that the test is very effective to identify a low dimensional chaos process in some cases.

Keywords : Chaos, Deterministic dynamic system, Nonparametric test

1. 서론

카오스(chaos)는 ‘혼돈(混沌)’이라는 뜻으로서 보통은 질서가 없는 무작위 상태를 뜻하며, 수학이나 물리학에 의해 원리적으로 확정되어 있음에도 불구하고 장래의 예측이 불가능한 현상을 의미하기도 한다.

카오스연구는 어원적인 개념에서 출발하여 무질서(disorder)와 비규칙성(irregularity)의 상태를 표현하는 것으로 발전하게 되었으며 또한, Lorenz(1963) 이후로 지난 40여 년 동안 비선형 동력학에서 포괄적으로 연구되어 왔다. 그러나 카오스

1) 본 연구는 한국과학재단 특정기초연구(No. R01-2005-000-10866-0)의 지원을 받아 수행되었음.

2) 제1저자 : 서울특별시 성동구 행당동 17번지 한양대학교 자연과학대학 수학과 강의전담교수
E-mail : pppppys@hanyang.ac.kr.

3) 서울특별시 성동구 행당동 17번지 한양대학교 자연과학대학 수학과 박사수료

4) 서울특별시 성동구 행당동 17번지 한양대학교 자연과학대학 수학과 석사수료

5) 교신저자 : 서울특별시 성동구 행당동 17번지 한양대학교 자연과학대학 수학과 교수
E-mail : kjcha@hanyang.ac.kr.

6) 서울특별시 성동구 행당동 17번지 한양대학교 의과대학 산부인과 교수

7) 서울특별시 성동구 행당동 17번지 한양대학교 공과대학 기계공학과 교수

의 예측 불가능한 현상은 겨우 몇 가지 요인의 작용 때문에 일어나기도 한다는 것이 밝혀졌다. 이것은 언뜻 보기에는 무질서하다고 생각되는 복잡한 현상의 배후에 정연한 질서가 숨어 있다는 것을 의미하며, 그 알려지지 않은 법칙의 진모를 탐색하는 일이 카오스를 연구하는 최대 목적이다. 실제로 몇몇 관심 있는 과학영역들 중에서 특히, 비선형 결정론적 동력학계(nonlinear deterministic dynamic system)로부터 생성한 자료는 무질서한 형태를 띠고 있다. 이러한 자료가 여러 학자들에 의해 카오스 현상을 보이는 것으로 입증되었다(Berliner, 1992; Chan & Tong, 1994).

본 연구에서 카오스 현상은 동력학계로부터 생성한 비규칙적이고 무질서한 자료를 무작위성(randomness)으로 정의할 것이다. 무작위한 측정 오차(error)와 결합된 카오틱 자료를 'noisy chaotic data'라고 하는데, 무질서한 자료로부터 카오스현상을 구별하는 많은 연구들이 여러 학자들에 의해 이루어지고 있다.

이러한 연구의 초기단계로서 단순한 수치적인 방법인 동력학계의 'chaoticity'를 정량화하는 두 개의 대표적인 방법은 프랙탈 차원(fractal dimensions)을 추정하는 것(Chatterjee & Yilmaz, 1992a)과 관측된 시계열로부터 리야프노프 지수(Lyapunov exponents)를 추정하는 것(Lai & Chen, 1995) 등이 있다. 프랙탈 차원과 리야프노프 지수가 좋은 추정량이 되기 위해서는, 수 천개 또는 더 많은 정확한 자료가 요구된다. 이런 종류의 자료는 물리학에서 활용 가능할지 모르지만 다른 분야 특히, 생물학, 전염병학, 경제학에서의 관측치 수열은 프랙탈 차원과 리야프노프 지수의 좋은 추정량을 계산하는데 충분할 만큼 길지도 정확하지도 않다. 게다가, 이들 자료는 종종 카오틱과 무질서한 개념 둘 다를 포함한다. 이러한 종류의 잡음이 포함된 카오스 자료를 분석하는데 주된 문제점은 'noisy' 환경에서 프랙탈 차원과 리야프노프 지수를 추정하기 위한 방법들을 개선하고 더 많은 이해를 하는데 초점을 맞추어 연구되어 왔다(Chan & Tong, 1994; Lele, 1994).

이에 본 연구에서는 규칙성이 없는 무질서한(random) 자료로부터 저차원의 카오스를 식별하는 고전적인 비선형 수치검정법에 대하여 논하고, 위상공간에서 재구성(reconstruction) 방법을 포함한 모의실험을 실시하였다.

2. 잡음이 섞인 카오스 현상과 검정기법

다음 형식을 갖는 동력학계에서

$$x_{t+1} = f(X_t)$$

여기서, $X_t = (x_t, x_{t-1}, \dots, x_{t-d+1})$, $t = 1, 2, \dots$ 은 시간 지수(index)이고 d 는 동력학계의 'lag dimension'을 가리키는 정수이다. 여러 경우에서, 그러한 동력학계는 무질서한 것 같지만 사실은 카오스 현상을 포함하는 자료를 생성할 수 있다. 그러나 고전적인 수치적 검정은 무질서한 수열로부터 이들 자료를 구별할 수 없었다.

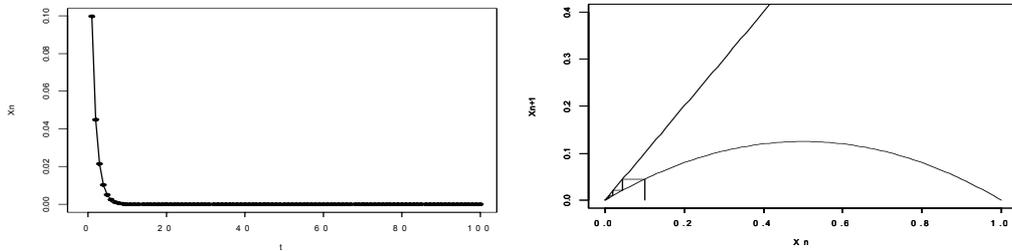
다음은 대부분 많은 연구자들이 공통적으로 다루는 로지스틱 사상(logistic map)이다(Tong, 1990).

예제 1. 로지스틱 사상은 다음과 같이 정의되는 함수이다.

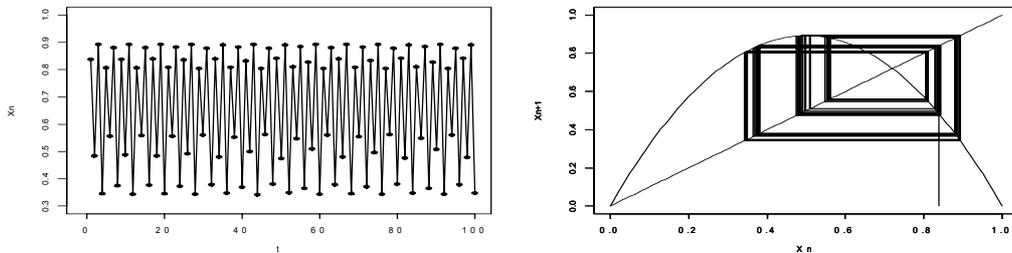
$$x_{t+1} = a * x_t(1 - x_t)$$

여기서, $t(=0, 1, 2, \dots)$ 는 시간을 의미하며, 초기값 $x_0 \in [0, 1]$, 계수는 $a \in [0, 4]$ 로서 제한하자. 특히, $a \geq 3.56994$ 에 대해 x_t 는 외관상으로 무질서한 현상을 나타낸다.

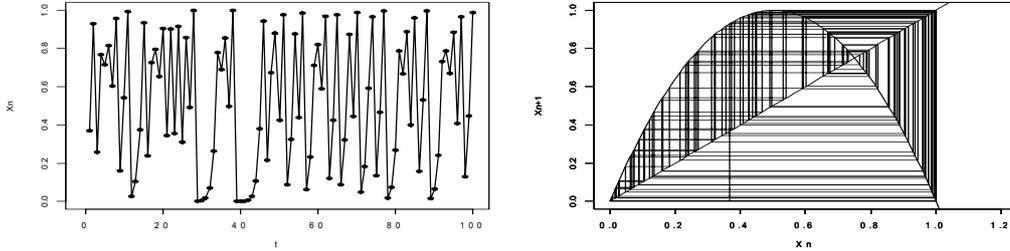
실제로, 그림 1은 $a=0.500$ 이고 초기값이 $x_0 = 0.013$ 일 때의 로지스틱 사상에 의한 시계열 자료와 x_{n+1} 대 x_n 그림으로서 리턴 맵을 이용한 로지스틱 방정식의 해의 거동을 보여 주고 있다. 이러한 자료에 비해 $a=3.56994$ 이고 초기값이 $x_0 = 0.841$ 일 때의 로지스틱



<그림 1> $a=0.500$ ($x_0 = 0.013$) 일 때의 로지스틱 사상에 의한 시계열 자료(Left)와 리턴 맵(x_{n+1} vs. x_n)을 이용한 로지스틱 방정식의 해의 거동(Right)



<그림 2> $a=3.56994$ ($x_0 = 0.841$) 일 때의 로지스틱 사상에 의한 시계열 자료(Left)와 리턴 맵(x_{n+1} vs. x_n)을 이용한 로지스틱 방정식의 해의 거동(Right)



<그림 3> $a=4.000$ ($x_0 = 0.382$) 일 때의 로지스틱사상에 의한 시계열 자료(Left)와 리턴 맵(x_{n+1} vs. x_n)을 이용한 로지스틱 방정식의 해의 거동(Right)

사상은 그림 2에서 보는 바와 같이 비 패턴적인 현상이 보이며 특히, 그림 3과 같이 $a=4.000$ ($x_0 = 0.382$) 일 때의 로지스틱사상은 카오스 현상이 두드러진다.

본 연구에서의 모의실험(simulation)은 $a=4.000$ 그리고 초기값 x_0 를 균일분포 $[0, 1]$ 로부터 생성한 카오스 자료를 이용하여 분석을 시행하였는데, 이것과 다른 단순한 동력학계들은 생물학, 전염병학 그리고 경제학 자료를 분석하는데 폭넓게 사용되고 있다(May, 1987).

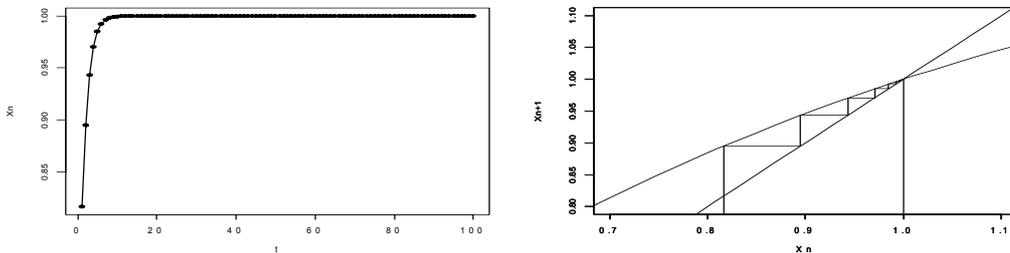
또 하나의 카오스현상을 다루는데 유용한 것으로서, 지수 사상(exponential map)이 있다.

예제 2. 지수 사상(exponential map)은 다음과 같이 정의 한다.

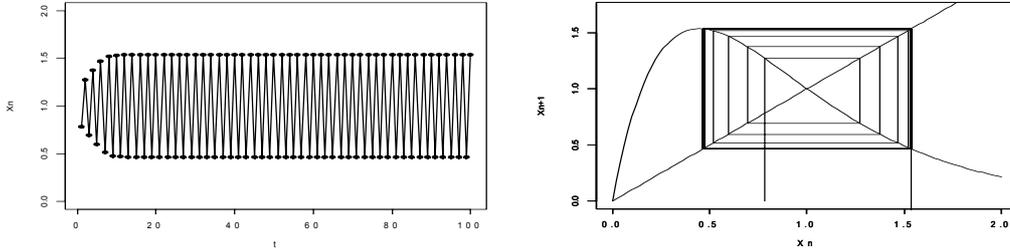
$$x_{t+1} = x_t e^{r(1-x_t)}$$

여기서, $t=0, 1, 2, \dots$, $x_0 \in [0, \infty)$, $r \in [0, 4]$ 이고 특히, $r > 2.6924$ 일 때 관측치는 카오스 현상을 보이게 된다.

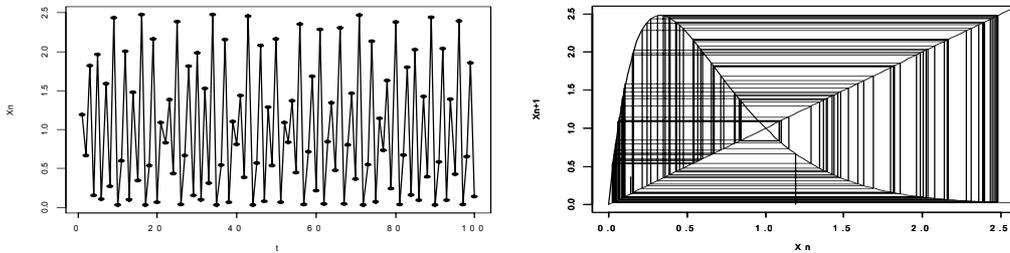
실제로 그림 4는 $r=0.500$ 이고 초기값이 $x_0 = 0.7040118$ 일 때의 지수사상(exponential map)에 의한 시계열 자료와 x_{n+1} 대 x_n 그림으로서 리턴 맵을 이용한 지수 방정식의 해의



<그림 4> $r=0.500$ ($x_0 = 0.7040118$) 일 때의 지수 사상에 의한 시계열 자료(Left)와 리턴 맵(x_{n+1} vs. x_n)을 이용한 지수 방정식의 해의 거동(Right)



<그림 5> $r=2.231$ ($x_0=1.187076$) 일 때의 지수 사상에 의한 시계열 자료(Left)와 리턴 맵(x_{n+1} vs. x_n)을 이용한 지수 방정식의 해의 거동(Right)



<그림 6> $r=3.011$ ($x_0=0.9095736$) 일 때의 지수 사상에 의한 시계열 자료(Left)와 리턴 맵(x_{n+1} vs. x_n)을 이용한 지수 방정식의 해의 거동(Right)

거동을 보여 주고 있다. 또한, 그림 5는 $r=2.231$ 이고 초기값이 $x_0=1.187076$ 일 때의 지수사상을 도식한 것인데, 그림에서 보는 바와 같이 비 패턴적인 현상이 보이며 특히, 그림 6 ($r=3.011$, $x_0=0.9095736$)에서의 지수 사상은 카오스 현상이 본격화 된다.

본 연구에서의 지수사상에 대한 모의실험은 $r=3.731$ 을 고정하고 초기값 x_0 를 균일분포 $[0, 2]$ 로부터 생성하여 실시하였으며, 계(system)로부터 일반적으로 관측되는 측정오차 ϵ_t 는 여러 종류가 있을 수 있으나 본 연구에서는 단지, $\{y_t=x_t+\epsilon_t\}$ 또는 $\{y_t=x_t\epsilon_t\}$ 의 오차만을 고려하기로 한다.

전통적으로 무질서 또는 무작위(randomness) 식별에 관한문제는 오랜 역사를 갖고 있으며 무질서를 입증하기 위해 많은 수치적 기법들과 통계량들이 제안되어왔다. 그들 중 대표적인 6가지에 대해서 기술하여 보자.

Sign test

표본의 크기 N 이 작고 모집단의 분포가 정규분포를 따른다는 확신이 없다면 모수적 방법은 재고해야한다. 이런 경우에 사용되는 비모수적(nonparametric) 방법에서는 모수는 모집단의 중앙값이고 검정통계량도 표본의 중앙값을 사용하지만 표본중앙값의 분포는 알려져 있지 않다.

Sign test는 가장 오래된 비모수검정방법으로 측정치들의 값을 부호(sign)로 바꾼 후 그것을 이용하여 모집단의 중앙값에 대한 가설검정을 하게 된다.

관측치 Y_1, Y_2, \dots, Y_N 은 모집단 분포 $f(y; M)$ 에서 추출한 확률표본이다. 이 때 귀무가설은 $M = M_0$ (단, M 은 모집단의 중앙값이며 M_0 는 고정된 값)이고 검정통계량은 다음과 같이 구한다(Gibbons & Chakraborti, 1992).

먼저, 각각의 관측치에 대하여 $(y_i - M_0)$ 의 부호를 계산한다. 이 때 M_0 는 귀무가설에 설정된 중앙값이다. 또한 $y_i = M_0$ 이기 때문에 부호가 결정되지 않는 것은 표본에서 제외한다. 만약 귀무가설이 참이면, 관측치로부터 얻은 '+'와 '-' 부호의 수는 같을 것이다. 만약 '+' 부호수가 상대적으로 많거나 또는 '-' 부호수가 많다면 귀무가설은 기각될 것이다. 실제로 $(y_i - M_0)$ 의 부호는 '+', '-' 두 개만 가지므로 베르누이(Bernoulli) 시행이 되며, 만일 $y_i = M_0$ 인 경우가 l 개이면 $N^* = N - l$ 에 대해서 '+' 부호의 개수 (K)는 이항분포 $B(N^*, p)$ 를 따르며, 이 때 귀무가설은 $K \sim B(N^*, 0.5)$ 에 부합된다. 일반적으로 표본크기가 크면 이항분포의 정규분포로의 근사이론이 적용되어 검정 통계량은 다음과 같이 식 (1)로 결정된다.

$$Z = \frac{(K+0.5) - 0.5N^*}{\sqrt{(0.5)^2 N^*}} \quad (1)$$

Runs test

자료수가 N 인 임의의 표본 Y_1, Y_2, \dots, Y_N 에 대해 식 (2)를 이용하여 0과 1로 정의되는 새로운 자료를 생성할 수 있다.

$$w_i = \begin{cases} 0 & \text{if } Y_i > \bar{Y} \\ 1 & \text{if } Y_i < \bar{Y} \end{cases} \quad (2)$$

여기서 \bar{Y} 는 자료의 평균값을 의미한다. 예를 들어, $N=15$ 인 경우 식 (2)에 의해 다음과 같은 임의의 값을 갖는 자료가 있다고 하자.

1 1 0 1 1 1 0 0 0 1 0 0 0 1 1

여기서 'run'은 연속적으로 0이거나 연속적으로 1인 경우를 말하며, 위 예에서 보면 0인 'run'은 3개이며, 1인 'run'은 4개 총 7개의 'run'으로 구성되어 있다.

대상 자료가 무작위성(독립성)을 갖는 경우 전체 'run'의 개수 (U)에 대해서 식 (3)과 같이 평균 $E[U]$ 와 분산 $Var[U]$ 을 갖는 정규분포가 된다(Keeping, 1966).

$$E[U] = \frac{2N_1N_0}{N_1+N_0} + 1, \quad Var[U] = \frac{2N_1N_0(2N_1N_0 - N_1 - N_0)}{(N_1+N_0)^2(N_1+N_0-1)} \quad (3)$$

여기서 N_0 와 N_1 은 각각 0과 1의 개수이다. 한편, 표준정규분포를 가지는 검정 통계량을 다음 식 (4)와 같이 나타낼 수 있다.

$$Z_U = \frac{U - E[U]}{\sqrt{\text{Var}(U)}} \quad (4)$$

여기서 $z_{1-\alpha/2}$ 가 표준정규분포의 $1-\alpha/2$ 분위수일 때, $|Z_U| < z_{1-\alpha/2}$ 라면 유의수준 α 에서 무작위성을 나타내는 귀무가설은 기각할 수 없으므로 이 경우 무작위성을 가지고 있다고 할 수 있다.

즉, 이 검정 방법은 순서적으로 양분된 수열에 대하여, 'run'을 정의함으로써 검정 통계량을 유도하는 방법인데(Bradley, 1968 참조), 식 (4)의 귀무가설은 주어진 수열이 무질서하다는 가정 하에 근사적으로 표준정규분포를 따른다(Kotz & Johnson, 1986).

Turning point test

주어진 계열 Y_i ($i=1, 2, \dots, N$, N 은 자료 수)에 대하여 'peak'와 'trough'를 각각 다음 식 (5)와 같이 정의한다.

$$\text{peak: } Y_{i-1} < Y_i > Y_{i+1}, \quad \text{trough: } Y_{i-1} > Y_i < Y_{i+1} \quad (5)$$

여기서 'peak'와 'trough'의 총 수 M 은 'turning points'의 개수이다. 또한, 이것은 근사적으로 정규분포를 따르며, 이 때 평균 $E(M)$ 와 분산 $\text{Var}(M)$ 는 다음 식 (6)과 같다(Clarke, 1973).

$$E(M) = \frac{2(N-2)}{3}, \quad \text{Var}(M) = \frac{16N-29}{90} \quad (6)$$

또한, 다음 Z_M 의 귀무가설 분포는 Kotz & Johnson(1986)에 의해 근사적으로 표준정규분포를 따른다는 사실이 증명되었으며, 이는 다음 식 (7)과 같다.

$$Z_M = \frac{M - E(M)}{\sqrt{\text{Var}(M)}} \quad (7)$$

따라서 $z_{1-\alpha/2}$ 가 표준정규분포의 $1-\alpha/2$ 분위수일 때 $|Z_M| \leq z_{1-\alpha/2}$ 라면 유의수준 α 에서 무작위성을 나타내는 귀무가설은 기각할 수 없다.

Kolmogorov-Smirnov test

카이제곱(Chi-square)검정은 연속형 변수에도 사용할 수 있는데, 일반적으로 사용하는 방법이 Kolmogorov-Smirnov test이다 (Conover, 1998). 이는 누적분포함수(cumulative distribution) 개념을 이용하여 두 확률변수의 분포함수가 동일하다면 확률밀도함수도 동일하다는 정리가 사용된다.

표본 Y_1, Y_2, \dots, Y_N 이 $F(y)$ 에서의 확률표본일 때, 이 표본이 균일분포 $F_0(y)$ 를 따르는지 검정하기 위해 먼저, 경험 분포함수(empirical distribution function)를 생성된 자료로부터 다음 식 (8)과 같이 정의 하자.

$$F_N(y) = \frac{1}{N} \sum_{i=1}^N I_{(-\infty, y)}(y_i) \quad (8)$$

여기서 $I_{(-\infty, y)}(y_i) = \begin{cases} 1, & y_i \leq y \\ 0, & y_i > y \end{cases}$ 이다. 즉, $F_N(y)$ 는 y 값 이하가 되는 y_i 의 개수를 N 으로 나눈 값이며 계단함수의 형태를 띠게 된다. 이때, 귀무가설 $H_0: F_N(y) = F_0(y)$ 이 참이라는 가정에서, 임의의 y 값에 대하여 식 (8)은 이항분포 $B(N, F_0(y))$ 를 따른다. 그러나 우리는 모든 y 에 대하여 위의 관계가 성립함을 검정하는데 관심이 있으므로 글리벤코-칸텔리 정리(Glivenko-Cantelli lemma)를 이용한다(Billingsley, 1986 참조).

임의의 양의 실수 ϵ 에 대하여 다음이 성립 한다.

$$\lim_{N \rightarrow \infty} P(\max_y |F_N(y) - F_0(y)| > \epsilon) = 0$$

즉, 귀무가설이 참이라는 가정 하에서, N 이 충분히 큰 경우에 모든 y 값에 대하여 주어진 $|F_N(y) - F_0(y)|$ 의 값이 매우 작을 확률이 1에 가까워진다는 의미이다. Kolmogorov-Smirnov검정은 이 성질을 이용하여 적합도 검정을 하는 것으로 다음 식 (9)와 같은 검정통계량을 이용한다.

$$D_N = \max_y |F_N(y) - F_0(y)| \quad (9)$$

실제로 우리가 검정하고자하는 것은 주어진 확률표본이 균일분포 $F_0(y)$ 를 따르는지를 검정하는 것이기 때문에 귀무가설에서 $F_0(y) = y$ 가 되며, 의사결정은 'Kolmogorov-table'를 이용하여 검정한다.

Anderson correlogram test

일반적으로 자료 수 N 이 증가하면, lag-k 표본 자기상관계수 r_k 는 평균이 0이고 분산이 $1/N$ 인 정규분포를 따른다(Anderson, 1942). 이러한 경우 귀무가설 $\rho_k = 0$ ($k = 1, 2, \dots$)은 아래 식 (10)과 같은 양측검정 구간을 사용하여 검정할 수 있다.

$$\left[\frac{-z_{1-\alpha/2}}{\sqrt{N}}, \frac{-z_{1-\alpha/2}}{\sqrt{N}} \right] \quad (10)$$

여기서, $z_{1-\alpha/2}$ 는 표준정규분포의 $1-\alpha/2$ 분위수이며, α 는 유의수준, N 은 자료수이다. 따라서 r_k 가 식 (10)의 구간 내에 존재하게 되면, 유의수준 α 에서 귀무가설을 기각할 수 없다. 한편 Anderson(1942)은 아래 식 (11)과 같이 r_1 (lag=1인 표본자기상관계수)에 대한 기대값 $E(r_1)$ 과 분산 $Var(r_1)$ 을 제시하였으며, 이 식들을 사용하여 정규분포라는 가정 하에 귀무가설 $\rho_1 = 0$ 을 검정할 수 있음을 보였다.

$$E(r_1) = -\frac{1}{N-1}, \quad Var(r_1) = \frac{(N-2)}{(N-1)^2} \quad (11)$$

또한, Yevjevich(1972)는 대립가설 $\rho_k \neq 0$ 에 대한 귀무가설 $\rho_k = 0$ 을 검정하기 위

하여 식 (11)에서 N 대신 $N-k+1$ 을 대입하여 수정된 식 (12)를 제안하였다.

$$E(r_k) = -\frac{1}{N-k}, \quad \text{Var}(r_k) = \frac{(N-k-1)}{(N-k)^2} \quad (12)$$

따라서 귀무가설을 검정하기 위한 $r=1-\alpha$ 구간은 다음과 같다.

$$\left[\frac{-1 - z_{1-\alpha/2} \sqrt{N-k-1}}{N-k}, \frac{-1 + z_{1-\alpha/2} \sqrt{N-k-1}}{N-k} \right] \quad (13)$$

그러므로 귀무가설 $\rho_k=0$ ($k=1, 2, \dots, M$, M 은 총 lag 수)에 대한 가설검정은 표본 자기상관계수 r_k 의 값이 식 (13)의 구간을 벗어나는 총수가 $(1-r)M$ 보다 크면 귀무가설이 기각된다. 일반적으로 M 의 값은 $N/4$ 을 이용한다.

Kendall's τ test

관측된 시계열 $\{y_t\}$ ($i=1, 2, \dots, N$, N 은 자료 수)에 대하여, 모든 $s < t$ 에 대한 $y_t - y_s$ 을 고려할 때, 만약 $y_t - y_s > 0$ 이라면 일치하는 쌍(concordant pair)을 갖고, $y_t - y_s < 0$ 이라면 불일치하는 쌍을 갖는다. c 는 일치된 쌍의 개수, d 는 불일치된 쌍의 개수 그리고 $N = c + d$ 라고 하자. 이때 Kendall의 τ 는 $\tau = c - d$ 로서 정의된다. 그리고 아래 식 (14)의 귀무가설 분포는 Kotz & Johnson(1983)에 의해 표준정규 분포임이 증명되었다.

$$Z = \frac{\tau}{\sqrt{2(2N+5)/[9N(N-1)]}} \quad (14)$$

3. 무질서(randomness)와 카오스의 식별

위에서 정의된 sign test, runs test, turning point test, Kolmogorov-Smirnov test, Anderson correlogram test 그리고 Kendall의 τ test 등 6가지의 비모수적 검정법은 무질서 수열과 로지스틱 사상 또는 지수 사상으로 부터 생성한 카오스 수열을 구별하는데 한계를 가지고 있다.

Chatterjee & Yilmaz(1992b)는 이 통계량들을 효율성을 높이기 위해서 다음과 같이 관측된 수열 위에 단순한 위상공간에서의 재구성(reconstruction)을 함으로서 개선될 수 있음을 제안하였다.

y_1, y_2, \dots, y_n 을 어떤 관측된 시계열로 표시하자. 수열이 무질서과정(random process)을 독립적으로 따른다는 귀무가설을 검정하고자 할 때, 한 배열로부터 lag를 1로 하여 다음과 같이 나열하자.

$$\begin{pmatrix} y_1, y_2, \dots, y_{n-1} \\ y_2, y_3, \dots, y_n \end{pmatrix} \quad (15)$$

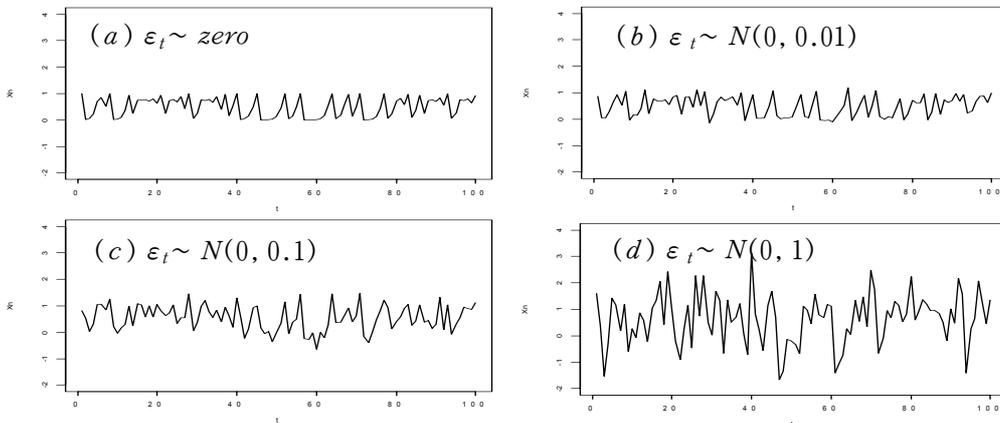
이제, 위식 (15)에서 첫 번째 행 y_1, y_2, \dots, y_{n-1} 의 크기를 기준으로 열을 재배열 하자. i 번째 가장 작은 관측치를 표시하기 위해 $y_{(i)}$ 를 사용하고 식 (15)의 두 번째 행에서 j 번째 원소가 i 번째로 작은 값을 나타내기 위해 y_{j_i} 을 사용해서 식 (16)과 같이 나타내자. 아래 식 (16)은 'lag=1' 위상공간에서의 재구성(reconstruction)을 표시한다.

$$\begin{pmatrix} y_{(1)}, y_{(2)}, \dots, y_{(n-1)} \\ y_{i_1}, y_{i_2}, \dots, y_{i_{n-1}} \end{pmatrix} \quad (16)$$

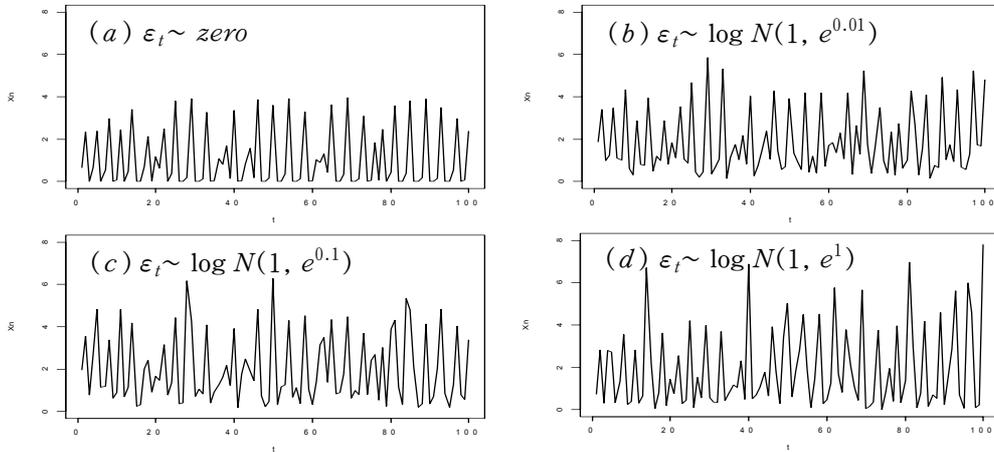
관측치 y_1, y_2, \dots, y_n 이 독립적인(i.i.d.) 확률변수의 실측값이고 n 이 클 때, 관측치 $y_{i_1}, y_{i_2}, \dots, y_{i_{n-1}}$ (식 (16)의 두 번째 행)이 근사적으로 독립적인 확률변수의 실측값으로 가정 할 수 있다. 이 가정은 실제로 표 1에서 보는 바와 같이 동일한 확률변수를 갖는 수열로서 판단할 수 있음을 보여주었다. 따라서 수열의 무질서 가설을 검정하기 위해서, 식 (16)의 두 번째 행에 상기한 기법 즉, 위상공간에 재구성을 하여 6개의 통계량에 각각 적용 하였다.

이는 다음의 모의실험 결과로부터 확인해 볼 수 있다.

먼저, 그림 7은 본래의 로지스틱 사상($a=4.000, x_0=0.4691803$)에 4가지 실험을 각각 시행하였는데, 첫째, 오차 ε_t 에 대해 잡음을 주지 않은 순수한 자료: (a) zero, 그리고 잡음을 정규분포 (b) $N(0, 0.01)$ 로 준 경우, 그리고 각각 (c) $N(0, 0.1)$, (d) $N(0, 1)$ 를 따르는 시계열 자료를 각각 도시한 것이다. 첫 번째 그림에서 볼 수 있듯이 그리 복잡하지 않은, 즉 저차원의 카오스 현상을 확인 할 수 있다.



<그림 7> 로지스틱 사상($a=4.000, x_0=0.4691803$)에 4종류의 오차 ε_t ((a) $\varepsilon_t \sim zero$, (b) $N(0, 0.01)$, (c) $N(0, 0.1)$, (d) $N(0, 1)$) 시계열자료.



<그림 8> 지수 사상($r=3.731$, $x_0=1.146091$)에서 오차 ϵ_t 에 대해 log-normal 분포 (a) zero, (b) $\log N(1, e^{0.01})$, (c) $\log N(1, e^{0.1})$, (d) $\log N(1, e^1)$ 의 시계열자료.

마찬가지로 그림 8은 지수사상 ($r=3.731$, $x_0=1.146091$)에 오차 ϵ_t 를 주지 않은 자료: (a) zero (*multiplier* = 1), 그리고 잡음을 log-normal 분포 (b) $\log N(0, e^{0.01})$ 로 준 경우, 그리고 (c) $\log N(1, e^{0.1})$, (d) $\log N(1, e^1)$ 를 따르는 시계열 자료를 표현한 것이다.

본 연구에서는 이러한 과정을 1,000회 반복 실험을 통하여 최종 모의실험을 하였다. 즉, 로지스틱 및 지수 사상 각각에 대해서 $N=100$ 개의 자료를 갖는 수열을 1,000회 반복하여 모의 수열을 생성하였다.

이 절차를 수행하기에 앞서 먼저, 길이가 100인 각각 독립적으로 균일분포 $[0, 1]$ 를 따르는 수열과 표준정규분포 $N(0, 1^2)$ 를 따르는 시계열자료를 각각 1,000개(회)를 생성하여, 각각 6개의 통계량에 대해서 본래의 수열과 또한 재구성한 수열에 적용하였다. 이는 각각 표준균일분포와 표준정규분포를 따르는 원래의 수열에 대해서 이를 재구성한다고 해도 동일한 분포를 따르며, 고유한 자료의 정보가 손실되지 않는다는 것을 보여주기 위함이다.

그 결과, 표 1은 독립적인 분포를 따르는(i.i.d.) 가설 하에 유의수준 $\alpha=0.05$ 에서 귀무가설을 기각하지 못한 비율(%)을 나타낸 것이다. 여기서 표준정규분포를 따르는 지에 대한 검정에 적합하지 않은 sign test는 제외하고, 나머지 검정 결과를 살펴보면, 두 기법에 대해 구별할 수 없을 정도의 유사한 기대비율을 보여 어떠한 검정 통계량도 독립적인 분포를 따르는(i.i.d.) 수열로부터 재구성 되어진 수열과 다르지 않다고 할 수 있다. 따라서 각각의 검

<표 1> 표준균일분포와 표준정규분포를 따르는지(*i.i.d.*)에 대한 결과

Test Statistics	Uniform	Reconstruct. Uniform	Normal	Reconstruct. Normal
Sign test	100.0	100.0	-	-
Runs test	94.8	95.4	94.2	96.2
Turning points test	93.8	96.0	93.4	98.8
Kolmogorov-Smirnov test	97.8	97.2	96.0	96.4
Anderson correlogram test	93.6	97.6	98.0	92.8
Kendall's τ test	94.6	96.4	97.8	93.2

정기법의 적용은 본래의 수열을 이용하는 재구성한 수열을 대상으로 하든 동일한 정보를 갖기 때문에 동일한 조건 하에 검정을 수행한다고 볼 수 있다.

따라서 6가지의 비모수검정기법은 독립적으로 $N(0, \sigma^2)$ 의 분포를 따르는 다른 크기의 측정오차를 갖는 잡음자료에 대해서 각각 로지스틱 및 지수 사상에 대해 순수한 카오스 자료와 이를 위상공간에서 재구성하여 생성한 자료에 각각 적용하였다.

본래의 로지스틱에서 도출된 수열과 지수사상에서의 수열에 대한 분석 결과는 표 2와 표 3과 같다. 6종류의 검정기법 실험 자료는 각각 4종류의 대응하는 측정오차 ε_t 가 *zero*, $N(0, 0.01)$, $N(0, 0.1)$, $N(0, 1)$ 를 따를 때 적용된 것이다.

먼저 로지스틱사상에서(표 2), 첫 번째 실험 $\varepsilon_t \sim \text{zero}$ 에 대해 유의 수준 $\alpha=0.05$ 에서 무질서 또는 무작위성(randomness)을 갖는다는 귀무가설을 기각하지 못한 비율은 각각 sign test는 100.0 %, runs test는 92.4 %, turning point test는 90.2 %, Kolmogorov-Smirnov test는 25.4 %, Anderson correlogram test는 98.2 % 그리고 Kendall의 τ test는 95.6 % 이었다. 그리고 대응하는 오차 ε_t 가 $N(0, 0.01)$, $N(0, 0.1)$, $N(0, 1)$ 일 때 적용된 경우에는 먼저, sign test서는 모두 100.0 %으로 나타났으며, Kolmogorov-Smirnov test에서는 각각 34.5 %, 90.1 %, 21.7 %로 나타났다. 그리고 나머지 검정기법에서는 $\varepsilon_t \sim \text{zero}$ 에서의 결과와 비슷한 수준을 보였다.

또한, 로지스틱 사상처럼 지수사상에 대해서도 유사한 결과를 볼 수 있는데(표 3), 각각의 100개씩 1,000회의 모의자료에 대하여, 실험 $\varepsilon_t \sim \text{zero}$ 의 경우에 대해 $\alpha=0.05$ 에서 귀무가설

<표 2> 로지스틱사상 ($\alpha=4.000$)에서의 무작위성(randomness) 검정 결과(%)

Test Statistics	$\varepsilon_t \sim \text{zero}$	$N(0, 0.01)$	$N(0, 0.1)$	$N(0, 1)$
Sign test	100.0	100.0	100.0	100.0
Runs test	92.4	90.2	92.0	95.0
Turning points test	90.2	93.2	94.6	92.6
Kolmogorov-Smirnov test	25.4	34.5	90.1	21.7
Anderson correlogram test	98.2	96.6	97.4	98.4
Kendall's τ test	95.6	96.4	96.2	97.8

<표 3> 지수사상 ($r=3.731$) 에서의 무작위성(randomness) 검정 결과(%)

Test Statistics	$\varepsilon_t \sim zero$	$\log N(1, e^{0.01})$	$\log N(1, e^{0.1})$	$\log N(1, e^1)$
Sign test	100.0	100.0	100.0	100.0
Runs test	96.2	99.0	92.0	96.8
Turning points test	96.6	99.8	98.6	95.2
Kolmogorov-Smirnov test	0.0	0.0	0.0	0.0
Anderson correlogram test	94.0	100.0	100.0	100.0
Kendall's τ test	95.4	100.0	100.0	100.0

을 기각하지 못한 비율은 각각 sign test는 100.0 %, runs test는 96.2 %, turning point test는 96.6 %, Kolmogorov-Smirnov test는 0.0 %, Anderson's γ 는 94.0 % 그리고 Kendall의 τ test는 95.4 % 이었다. 그리고 대응하는 측정오차 ε_t 가 $\log N(1, e^{0.01})$, $\log N(1, e^{0.1})$, $\log N(1, e^1)$ 일 때 적용된 경우도 비슷한 수준의 채택률(%)을 보였다.

위 실험결과는 Kolmogorov-Smirnov test를 제외한 5가지 검정법 sign test, runs test, turning point test, Anderson correlogram test 그리고 Kendall's τ test 등의 기법이 무질서로부터 결정론적 카오스를 구별하는데 유용하지 못함을 의미한다. 왜냐하면, 순수한 1,000회 모의실험 카오스 자료(측정오차 $\varepsilon_t \sim zero$)와 각각 오차 잡음 $N(0, 0.01)$, $N(0, 0.1)$, $N(0, 1)$ 을 더해준 자료에 대해서, Kolmogorov-Smirnov test를 제외한 각각의 5개의 검정기법들이 유의수준 $\alpha=0.05$ 에서 무질서(randomness)하다는 귀무가설을 기각하지 못한 비율이 기대 비율인 95 %정도로 나타났기 때문이다. 더욱이 sign test는 100.0 %의 결과를 나타냈다. 그러나 Kolmogorov-Smirnov test는 로지스틱사상에 대해, 저차원 카오스현상을 나타내는 잡음이 없는($\varepsilon_t \sim zero$) 자료에서 잡음이 커짐에 따라 점차 채택비율이 증가하다가 $N(0, 1)$ 에서는 그 비율이 다시 감소하여 잡음이 포함된 저차원 카오스현상을 어느 정도 설명하는 것으로 나타났다.

표 2와 3에서의 결과를 정리해 보면, Kolmogorov-Smirnov test를 제외한 각각의 5가지 검정기법이 직접적으로 카오스와 잡음현상을 식별할 수 없다는 것을 설명한다고 하겠다.

위 경우, 위상공간에 재구성(reconstruction)한 자료에서 검정 통계량을 적용해보면 (표 4, 5), 먼저, sign test는 잡음과 무관하게 원래자료와 동일하게 두 사상 모두에서 100 %으로 나타났으며, 반면에 Kolmogorov-Smirnov test에서는 잡음에 무관하게 주어진 수열이 랜덤하다는 귀무가설을 기각하지 못한 비율이 0.0 %로 나타났다.

그리고 runs test, turning point test 등의 기법에 대해서는 관측된 수열이 독립적으로 균일분포 즉, 무작위한 분포를 따르지 않는다는 것을 뜻하고 있다. 왜냐하면 잡음(noise)이 없거나 크지 않은($\varepsilon_t \sim zero$ 또는 $\varepsilon_t \sim N(0, 0.01)$) 카오스자료에서는 표 4에서 보는바와 같이 무질서(randomness)하다는 귀무가설에 대해 유의수준 $\alpha=0.05$ 에서 기각비율이 95 % 기대 비율과 큰 차이를 보이기 때문이다.

또한, Anderson's γ 그리고 Kendall's τ 에서는 그 비율이 각각 63.8, 75.8 %로 원 자료에 비해 다소 개선된 효과가 있어 보인다. 특히, 잡음(noise)이 충분히 크지 않은

경우에는 ($N(0, 0.01)$, $N(0, 0.1)$) 기대 비율 95 % 에 미치지 못하여 무질서한 현상과 카오스 현상을 어느 정도 식별할 수 있음을 보여 주고 있다.

또한, turning points test는 수열의 비 무질서(non-randomness)를 감지해 낼 수는 있으나, 매우 작은 잡음일지라도 잡음이 존재한다면 이 기법은 오류가 발생 할 수 있음을 보여주고 있다. 그리고 자료의 추세를 인지하는데 Anderson correlogram test와 Kendall's τ 가 일반적으로 많이 활용되지만, 위 경우에는 무질서로부터 카오스를 식별하는 데에는 유용하지 못하다는 것을 나타내고 있다.

다음으로 지수사상에 대한 모의실험 결과는 표 5에서 제시된 바와 같다. 로지스틱 사상의 경우와 유사하게 4가지 모의실험 자료는 각각 오차 ϵ_t 가 $\epsilon_t \sim zero$ ($multiplier = 1$) 그리고 log-normal 분포 $\log N(1, e^{0.01})$, $\log N(1, e^{0.1})$, $\log N(1, e^1)$ 를 따르는 시계열 자료를 대상

<표 4> 로지스틱사상 ($a=4.000$) 에서 재구성된 자료의 무작위성 검정 결과(%)

Test Statistics	$\epsilon_t \sim zero$	$N(0, 0.01)$	$N(0, 0.1)$	$N(0, 1)$
Sign test	100.0	100.0	100.0	100.0
Runs test	0.0	0.0	0.0	94.8
Turning points test	0.0	16.0	90.8	91.4
Kolmogorov-Smirnov test	26.1	35.3	91.6	21.3
Anderson correlogram test	63.8	80.2	91.4	95.4
Kendall's τ test	75.8	83.8	93.4	95.2

<표 5> 지수사상 ($r=3.731$) 에서 재구성된 자료의 무작위성 검정 결과(%)

Test Statistics	$\epsilon_t \sim zero$	$\log N(1, e^{0.01})$	$\log N(1, e^{0.1})$	$\log N(1, e^1)$
Sign test	100.0	100.0	100.0	100.0
Runs test	0.0	0.0	0.0	2.6
Turning points test	0.0	0.0	13.4	97.4
Kolmogorov-Smirnov test	0.0	0.0	0.0	0.0
Anderson correlogram test	54.8	62.6	92.4	100.0
Kendall's τ test	94.8	88.6	93.8	100.0

으로 한 것이다.

그 결과는 로지스틱사상의 경우와 유사하게 나타났는데, 6가지 검정법 중에서 먼저, sign test는 원래 자료와 동일하게 재구성효과가 전혀 없이 카오스현상을 설명하지 못하였으며, Kolmogorov-Smirnov test는 잡음에 무관하게 그 현상을 설명하고 있는 것으로 나타났다.

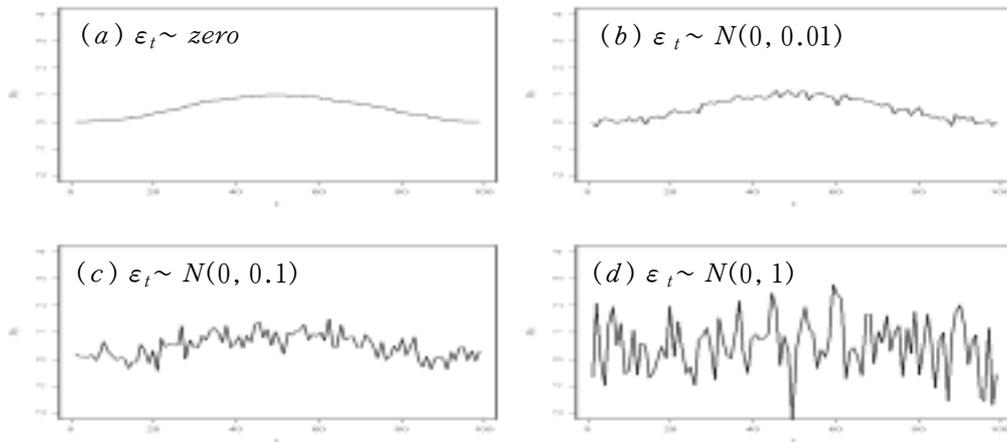
그리고 runs test, turning points test 등은 작은 잡음에 대해서도 식별가능성을 보여주었으나, 상대적으로 큰 잡음에 대해서는 무질서와 카오스현상을 식별하는 데에 문제가 생길 수 있다는 것을 나타내었다. 또한, Anderson correlogram 과 Kendall's τ test 는 재구성 효과가 다소 있으나 역시, 잡음이 포함된 카오스현상을 식별하지

못하고 있음을 말해준다.

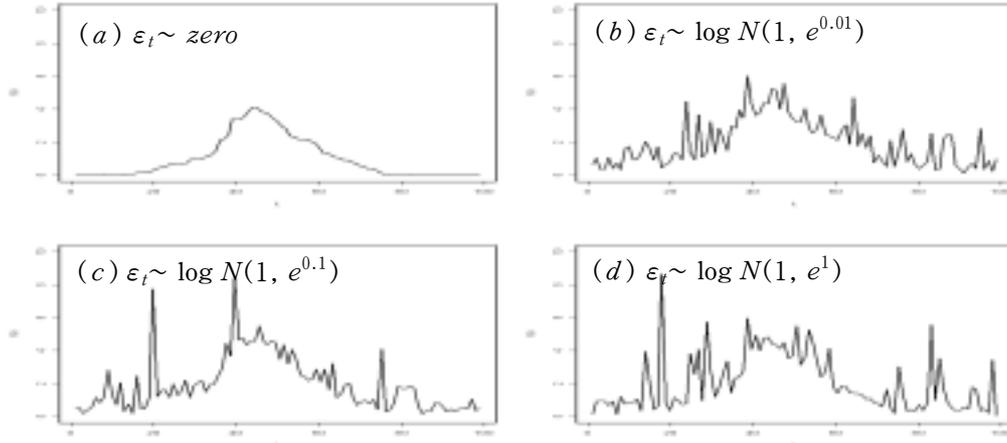
표 4와 표 5에서의 결과로부터, sign test 와 Kolmogorov-Smirnov test는 재구성 (reconstruction) 기법의 효과가 없으며, 반면에 나머지 기법에서는 다소 있었는데 특히, runs test는 4가지 검정 방법 중에서 보다 큰 분산이 있는 경우에 유용하다는 것을 알 수 있으며, 상관성에 기초한 Anderson correlogram test 와 Kendall's τ test 는 무질서와 카오스현상을 식별하는 데에는 한계가 있음을 나타내고 있다.

이러한 재구성(reconstruction) 기법에 의한 비모수적 검정의 개선 효과는 아마도, 원래의 자료(raw data)에서의 변화는 무질서해보이지만 (그림 7, 8), 재구성된 자료 (reconstruction data)의 경우엔 그림 9 와 그림 10에서 보는 바와 같이 그 변화가 규칙성을 갖는 패턴으로의 변화과정으로 설명될 수 있다. 여기서 그림 9는 로지스틱사상에 대해 그림 7에서 나타난 원 자료를 재구성하여 나타낸 것이고 그림 10은 지수사상에 대응하는 그림 8을 재구성하여 도시한 것이다.

그리고 이러한 효과에 무관한 검정 기법인 sign test 와 Kolmogorov-Smirnov test 는 본 연구의 경우처럼 저차원 카오스현상을 설명하는데에는 적합하지 못한 검정법으로 사료된다. 먼저, sign test 는 상술한 바와 같이 전혀 카오스현상을 설명할 수 없음을 실험을 통해 알 수 있었다. 왜냐하면 카오스현상은 일견 불규칙하게 보이지만 그 이면에는 질서 정연한



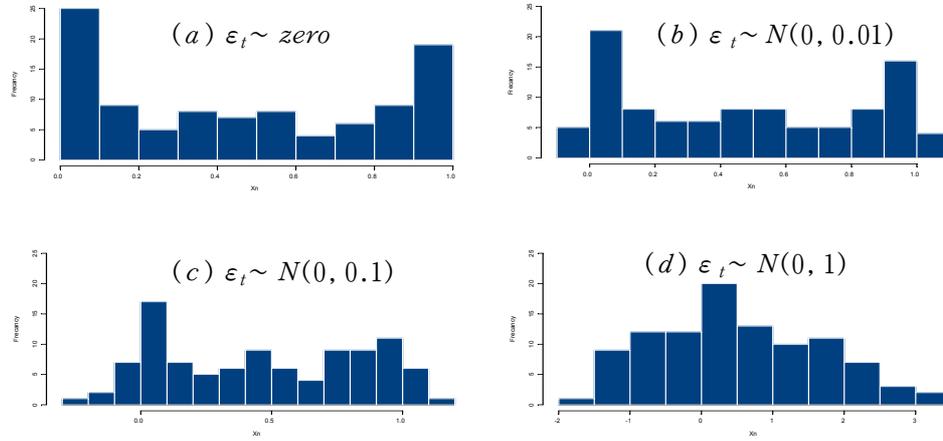
<그림 9> 로지스틱 사상($a=4.000, x_0=0.4691803$)에 4종류의 오차 ϵ_t ((a) $\epsilon_t \sim zero$, (b) $N(0, 0.01)$, (c) $N(0, 0.1)$, (d) $N(0, 1)$) 시계열자료.



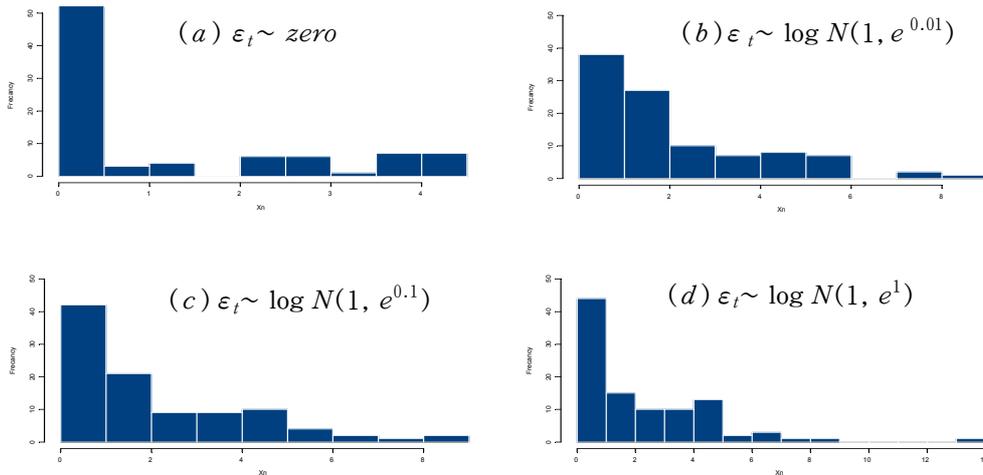
<그림 10> 지수 사상($r=3.731$, $x_0=1.146091$)에서 오차 ε_t 에 대해 log-normal 분포 (a) zero, (b) $\log N(1, e^{0.01})$, (c) $\log N(1, e^{0.1})$, (d) $\log N(1, e^1)$ 의 시계열자료.

패턴을 갖고 있기 때문에 표준균일분포를 따르는 귀무가설을 기각 해야만 한다. 그리고 Kolmogorov-Smirnov test의 경우, 로지스틱사상에서는 잡음 $N(0, 0.1)$ 에서 90.1%의 다소 높은 채택률을 보였으나(표 2), 지수사상에서는 잡음이 없거나 큰 경우 모두 0.0%로 나타나 그 결과의 일관성이 없는 것으로 나타났다(표 3). 이러한 결과는 저차원 카오스함수인 로지스틱함수와 지수함수의 특성에 기인한 것으로 추측된다. 그림 11과 그림 12는 각각 그림 7과 그림 8에 대응하는 자료 각각에 대해 히스토그램으로 나타낸 것으로서, 그림 11의 로지스틱사상에서는 원 자료에서 잡음의 증가에 따라 분포형태의 변화가 있으나, 이에 비해 그림 12의 지수사상에서는 오른쪽으로 기울어진(right skew) 원 자료의 모양이 크게 변하지 않았다. 이러한 현상은 주어진 수열이 무작위성(randomness)을 나타내는 균일분포를 따른다는 귀무가설을 기각하는 결과를 초래한 것으로 분석된다. 따라서 Kolmogorov-Smirnov test는 저차원카오스 현상과 잡음을 식별하는 데에는 주의를 기울일 필요가 있다고 하겠다.

이상의 결과로 부터 위상공간에서의 재구성(reconstruction) 기법은 낮은 차원의 카오스와 잡음을 식별하는 비모수적검정의 효율을 높이고 또한, 무질서한 수열을 생성하는 절차를 검정하는 여러 기법 중에 하나로 활용가능하다고 할 수 있다.



<그림 11> 로지스틱 사상($a=4.000$, $x_0=0.4691803$)에서 오차 ε_t 에 따른 ((a) $\varepsilon_t \sim zero$, (b) $N(0, 0.01)$, (c) $N(0, 0.1)$, (d) $N(0, 1)$) 히스토그램.



<그림 12> 지수 사상($r=3.731$, $x_0=1.146091$)에서 오차 ε_t 에 따른 (a) $zero$, (b) $\log N(1, e^{0.01})$, (c) $\log N(1, e^{0.1})$, (d) $\log N(1, e^1)$ 히스토그램

4. 결론

최근까지 카오스 현상 특히, 비선형 동력학시스템의 성질을 규명하는 데에 여러 수학기법들이 제안되고 있는데, 여러 제약조건 즉, 자료의 측정오차(error)와 내재된 잡음(noise) 등으로 인하여 실증적 적용에 있어서 신뢰할 만한 결론을 얻는데 어려움을

안고 있다.

이에 본 연구에서는 저차원의 카오스 현상을 식별하는데 새로운 수치적인 접근법으로서의 위상공간에서 재구성(reconstruction) 기법을 소개하였으며 이는 좀 더 확장된 기법으로서 발전 가능성을 시사한다고 하겠다.

참고문헌

1. Anderson, R. L. (1942). Distribution of the Serial Correlation Coefficient, *Ann. Math. State.*, 13, 34-45.
2. Berliner, L. M. (1992). Statistics, probability and chaos, *Statistical Science*, 7, 69-90.
3. Billingsley, P. (1986). Probability and Measure. Wiley and Sons, New York, second edition.
4. Bradley, J. (1968). *Distribution-free statistical tests*, Englewood Cliffs, NJ: Prentice-Hall.
5. Chan K. S. and Tong H. (1994), A note on noisy chaos, *Journal of the Royal Statistical Society, Ser., B.*, 56, 301-311.
6. Chatterjee S. and Yilmaz M. (1992a), Chaos, fractals and statistics, *Statistical Sciences*, 7, 49-68.
7. Chatterjee S. and Yilmaz M. (1992b), Use of estimated fractal dimension in model identification for time series, *Journal of Statistical Computation and Simulation*, 41, 129-141.
8. Clarke, R. T. (1973), *Mathematical Models in Hydrology, Irrigation and Drainage Paper*, Vol. 19, Food and Agriculture Organization of the United Nations, Rome, Italy.
9. Conover, W. J. (1998), *Practical Nonparametric Statistics*, Wiley, 3rd edition
10. Gibbons J. D. and Chakraborti S. (1992), *Nonparametric statistical inference*, New York: Marcel Dekker.
11. Keeping, E. S. (1966), *Distribution Free Methods in Statistics.*, in *Proceedings of Hydrology Symposium*, No. 5, McGill University, Canada.
12. Kotz S. and Johnson N. L. (1983), *Encyclopedia of Statistical Sciences*, Vol. 4, Wiley, New York, 367-369.
13. Kotz S. and Johnson N. L. (1986), *Encyclopedia of Statistical Sciences*. Vol. 7, Wiley, New York, 555-562.
14. Lai D. and Chen G. (1995), Computing the distribution of the Lyapunov exponent from time series: the one-dimensional case study, *International Journal of Bifurcations and Chaos*, 5, 1721-1726.
15. Lele S. (1994), Estimating functions in chaotic systems, *Journal of the American Statistical Association*, 89, 512-516.

16. Lorenz E. N. (1963), Deterministic nonperiodic flows, *Journal of Atmospheric Science*, 20, 282-293.
17. May R. M. (1987), Chaos and the dynamics of biological populations, *Proceedings of the Royal Society, London, A*, 413, 27-44.
18. Tong H. (1990), *Non-linear Time Series: A Dynamical System Approach*, Oxford University Press, New York.
19. Yevjevich, V. (1972), *Stochastic Processes in Hydrology*, Water Resources Pub., Fort Collins, Colorado.

[2005년 10월 접수, 2006년 2월 채택]