

动力系统实测数据的非线性混沌特性的判定^{*}

马军海^① 陈予恕^② 刘曾荣^③

(1996 年 9 月 20 日收到, 1998 年 3 月 1 日收到修改稿)

摘 要

本文利用相位随机化的替代数据方法, 给出了一个对动力系统实测时间序列数据的特性进行判定的方法。计算结果表明: 相位的充分随机化可提高判别的准确程度。把此判据用于随机时序与非线性混沌时序所得的判据值有明显的差异。

关键词 非线性混沌时序 替代数据 判据 随机时序

中图分类号 O175, O241

§ 1. 引 言

在现实生活中, 在工程界、自然科学界、社会科学界广泛地存在着各种各样的时序问题, 这些时序的形态, 往往呈现出复杂性和多样性。目前, 人们对时序问题一般采用两种不同方法处理: 一种是以概率统计学为基础, 用随机过程理论建立其线性模型。这种方法已比较成熟和完善, 如定阶准则、AR 模型、MA 模型、ARMA 模型及 ARMAV 模型都已经得到了广泛的应用^[1]。另一种是采用非线性动力学的方法对时序进行研究, 它又可分为对低自由度的混沌动力系统和高自由度的斑图动力系统的研究^[2]。对于反映低自由度混沌的时序问题已开展了十几年研究, 取得了不少结果^[3], 而对于反映高自由度斑图的时序问题的研究目前只有很少文献报道。

观测时序如果是随机序列, 就应该采用统计学的方法加以研究。如果序列是低自由度的确定性混沌, 则系统必是存在非线性机制的耗散系统, 在这种情况下不论多么高阶次的线性模型都不可能对该非线性系统的行为作出恰当的解释; 相反, 一个成功的低阶非线性模型或许能很好地刻画该动力学系统。即使对于一个确定系统产生的混沌时序, 由于测量仪器和测量过程中都不可避免要产生一些噪声, 到底序列中混入的噪声比例如何, 这是一个必须要解决的问题。综上所述, 自然要提出这样的问题: 如何判定时序是确定性的还是随机性的; 如果时序中既含有噪声又含有确定性因素, 那么它们到底是以确定性为主还是以随机性为主的。本文利用相位随机化的替代数据方法, 给出了一个对动力系统实测时序数据的特性进行判定的方法, 对是否是确定性的或如果含有随机因素是否以确定性因素为主的序列进行了探讨。计算结果

* 国家自然科学基金资助项目 (19672043)

① 东南大学管理学院系统工程研究所, 南京 210018

② 天津大学力学系, 天津 300072

③ 上海大学数学系, 上海 201800

表明:相位的充分随机化可以提高判别的准确程度,且随机时序与非线性混沌时序的判值有明显差异。

§ 2. 替代数据的生成和判据建立

类似于[4],设实验所得到的时间序列为 $X_0, X_1, X_2, \dots, X_{N-1}, X_i(t)$ 为第 T_i 时间采样所得,其中 $T_1, T_2, T_3, \dots, T_N$, 分别为: $0, \Delta T, 2\Delta T, \dots, (N-1)\Delta T$, ΔT 为采样时间间隔。应用离散的付里叶变换,变换算子为 \mathcal{F} 得到:

$$X(f) = \mathcal{F}\{x(t)\} = \sum_{n=0}^{N-1} x(t_n) e^{2\pi i f n \Delta t} \quad \text{方法1)} \quad (2.1)$$

(2.1)式还可写为:

$$X(f) = \mathcal{F}\{x(t) = A(f) e^{i\phi(f)}\} \quad \text{为系} \quad (2.2)$$

其中, $A(f)$ 为幅值, $\phi(f)$ 为相位, f 分别取: $-N\Delta f/2, \dots, -\Delta f, 0, \Delta f, \dots, N\Delta f/2$; $\Delta f = (1/N)\Delta T$ 。然后将 $\phi(f)$ 随机地旋转一相位角 $\psi(f)$, $\psi(f)$ 是由计算机在区间 $[0, 2\pi]$ 内任意随机所生成的随机数,这样 $\phi(f)$ 的取值便不带有任何人因素,是完全随机的,可以得到:

$$\hat{X}(f) = A(f) e^{i[\phi(f) + \psi(f)]} = X(f) e^{i\psi(f)} \quad (2.3)$$

再对(2.3)式进行付里叶逆变换可以得到:

$$\hat{x}(t) = \mathcal{F}^{-1}\{\hat{X}(f)\} = \frac{1}{N} \sum_{n=0}^{N-1} \hat{X}(n) e^{i2\pi nk/N} \quad (2.4)$$

这样所得数据 $\hat{x}_1, \hat{x}_2, \hat{x}_3, \dots, \hat{x}_N$ 称为原始数据的替代数据。这里要求得到的替代数据的虚部为零,若其虚部不为零将有可能使得到的替代数据不真,并最终导致混沌特性的判值失效。我们认为这是替代数据生成的一个重要问题。这一步的实现被称为相位的充分随机化。

现在以原始数据 $X_i(t)$ 为例建立判定数据特性的方法^[5]:

$$\text{令 } V_i = [x_i, x_{i+T}, x_{i+2T}, \dots, x_{i+(m-1)T}] \quad (2.5)$$

$$\text{其中 } V_i \in \mathbb{R}^m$$

定义:

$$\begin{aligned} C(r, N, w) &= \frac{1}{N^2} [\text{距离} | V_i - V_j | \leq r \text{ 的点的对数}] \\ &= \frac{2}{N^2} \sum_{n=w}^N \sum_{i=1}^{N-n} H[r - | V_{i+n} - V_i |] \end{aligned} \quad (2.6)$$

H 为 Heaviside 函数即:

$$H(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

由于实际问题中不可避免地要出现数据间的自相关性,以及 N 取值的限制。文献[6]中已证明(2.6)式中的 $C(r, N, w)$ 可采用如下式子:

$$C(r, N, w) = \frac{2}{(N-w+1)(N-w)} \sum_{n=w}^N \sum_{i=1}^{N-n} H[r - | V_{i+n} - V_i |] \quad (2.7)$$

这里要求 w 应满足:

$$w \geq \tau \left[\frac{2}{N} \right]^{2/m} \quad (2.8)$$

τ 为时间序列之间的时间间隔, m 为嵌入维数, 定义时序的关联维数为:

$$d = \lim_{r \rightarrow 0} \lim_{N \rightarrow \infty} \frac{\log C(r, N)}{\log r} = \lim_{r \rightarrow 0} \lim_{N \rightarrow \infty} \frac{d[\log_2 C(r, N, w)]/dr}{d(\log_2 r)/dr} \quad (2.9)$$

m 取值范围为 $d \leq m \leq 2d + 1$.

对原始数据, 取 $\langle D_{orig} \rangle = d$, 对替代数据的样本, 同样利用 d 可算得 $\langle D_{surr} \rangle = d_{surr}$, 该 $\langle D_{surr} \rangle$ 为替代数据 D_{surr} 的统计平均. 定义:

$$Z = | \langle D_{surr} \rangle - \langle D_{orig} \rangle | / \sigma_{surr} \quad (2.10)$$

σ_{surr} 为替代数据的均方差. Z 就是我们所建立的判据. 如果序列为随机的或以随机为主的, 相位随机化后 $| \langle D_{surr} \rangle - \langle D_{orig} \rangle |$ 较小, 判据 Z 就较小; 如果序列为确定性的或以确定性为主的, 相位随机化后 $| \langle D_{surr} \rangle - \langle D_{orig} \rangle |$ 的值较大, 自然判值 Z 就大. 根据概率论中的中心极限定理, 文献[7]指出: 若取显著性水平为 $\alpha = 0.05$, 则 $Z \geq 1.96$ 表明原始数据 $\langle D_{orig} \rangle$ 与替代数据 $\langle D_{surr} \rangle$ 有明显差别, 原时间序列以 95% 的置信水平为非线性混沌时序, 反之若 $Z \leq 1.96$ 则原时间序列为随机时序. 当然, 临界值 $Z = 1.96$ 并非最终的定论值, 只是目前暂被人们所接受.

§ 3. 计算结果

我们取如下 4 组数据对其随机特性或非线性混沌特性进行检验.

- ① 取 $[0, 2\pi]$ 内的随机数 7500 点 (其相图如图 7).
- ② 对 logistic map $x_{n+1} = \mu x_n(1 - x_n)$ 取 $\mu = 4$ 混沌的情况, 接下来前 1000 点作为暂态点去掉, 7500 点作为原始数据.
- ③ 对 logistic map 取 μ 略大于 $1 + \sqrt{8}$, 间歇混沌的情况, 同样取 7500 点作为原始数据.
- ④ 我们在实验室非线性机械振动模型上所得 3225 点作为实验数据, 其时间历程图如图 1.

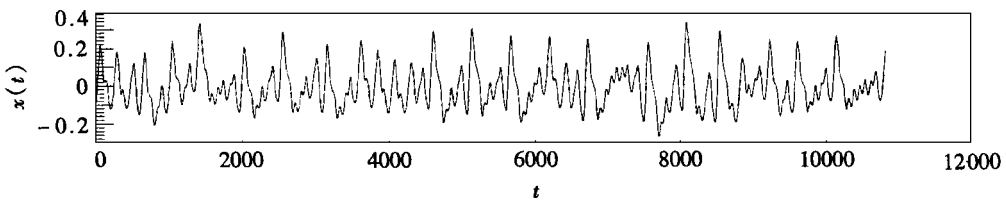


图 1 第(4)组数据的相图

对上述 4 组数据分别进行付里叶变换, 然后在其逆付氏变换的相位上, 分别加上用蒙特卡罗随机取余方法在 $[0, 2\pi]$ 内产生的均匀分布的随机数.

采样间隔 ΔT 的取法: 为了正确地获取连续信号中各种频率成份的信息, 在最高频率谐波的一个周期里至少应采样两次. 根据 Shannon 采样定理数据之间采样的时间间隔 $\Delta T \leq 1/(2f_{max})$. 这里 f_{max} 为连续信号中 $X_i(t)$ 中出现的最高频率成份. 为可靠地防止频混, 本文取 $\Delta T = 1/(2.5f_{max})$.

(2.1) 式中经过付里叶正变换后可得到 $N + 1$ 个付里叶正变换值. 除了 $f = 0$ 外, 剩余 N 个付氏正变换值, 其中有对应的 $N/2$ 个值实部分别两两相同. 这样我们产生 $N/2$ 个随机数,

与其对应的 $N/2$ 个随机数分别取成前 $N/2$ 个随机数的负值。再依其顺序分别加在其逆付氏变换的相位上, 然后再对 $f = 0$ 进行处理便可使付氏变换所生成复数的虚部为零。当然具体操作起来还非常复杂, 幅值、幅角、实部、虚部、相位, Σ 号下各分值的对应关系要搞清楚。

上述每组数据分别进行付氏变换后每组分别加上 3 组完全不同的随机数, 再进行付里叶逆变换, 每组得到 3 组不同的替代数据。其均值、方差和算得的维数均值如表 1~4。

表 1

	(1) 组原始数据	(1) 组替代数据 a	(1) 组替代数据 b	(1) 组替代数据 c
均值 x	0. 50152290	0. 50152291	0. 50152290	0. 50152291
均方差 S	0. 287420835463	0. 287420823566	0. 287420833872	0. 287420848483
$\langle D_{\text{orig}} \rangle$	2. 1768318976			
$\langle D_{\text{sur}} \rangle$		2. 207149932617	2. 129452247144	2. 226255894703

表 2

	(2) 组原始数据	(2) 组替代数据 a	(2) 组替代数据 b	(2) 组替代数据 c
均值 x	0. 50855721	0. 50855718	0. 50855721	0. 50855723
均方差 S	0. 350527896083	0. 350527884446	0. 350527870032	0. 350527882612
$\langle D_{\text{orig}} \rangle$	1. 339094249725			
$\langle D_{\text{sur}} \rangle$		2. 427022915208	2. 181922432328	2. 548350946926

表 3

	(3) 组原始数据	(3) 组替代数据 a	(3) 组替代数据 b	(3) 组替代数据 c
均值 x	0. 50988235	0. 50988232	0. 50988235	0. 50988236
均方差 S	0. 350182380613	0. 350182402727	0. 350182413633	0. 350182390892
$\langle D_{\text{orig}} \rangle$	1. 858021492790			
$\langle D_{\text{sur}} \rangle$		2. 991150511607	3. 403124555914	2. 811411329267

表 4

	(4) 组原始数据	(4) 组替代数据 a	(4) 组替代数据 b	(4) 组替代数据 c
均值 x	- 0. 00407823	- 0. 00407814	- 0. 00407814	- 0. 00407815
均方差 S	0. 073289616570	0. 073289625217	0. 073289596570	0. 073289610191
$\langle D_{\text{orig}} \rangle$	1. 456366767501			
$\langle D_{\text{sur}} \rangle$		1. 024648595272	1. 084153225743	1. 100250265527

应用(2. 10)式:

对(1)组 3 个替代数据可算得:

$$Z_a = 0. 0857, Z_b = 0. 1340, Z_c = 0. 1398 \quad (3. 1)$$

对(2)组 3 个替代数据可算得:

$$Z_a = 3. 1036, Z_b = 2. 4045, Z_c = 3. 4498 \quad (3. 2)$$

对(3)组 3 个替代数据可算得:

$$Z_a = 3.2364, Z_b = 4.4123, Z_c = 2.7226 \tag{3.3}$$

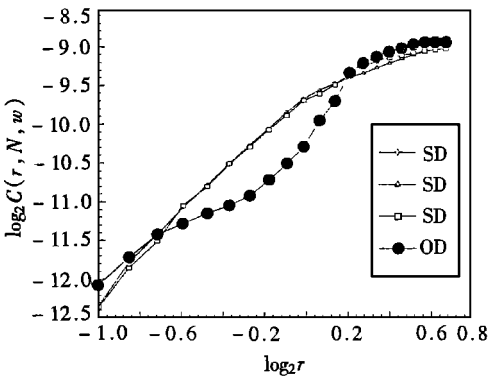
对(4)组 3 个替代数据可算得:

$$Z_a = 5.8906, Z_b = 5.0787, Z_c = 4.8590 \tag{3.4}$$

由前面所讨论的临界值不难看出: (1) 组数据为随机数据, (2), (3), (4) 组数据为非线性混沌数据。以上的计算结果与我们所采用的数据结果完全吻合。

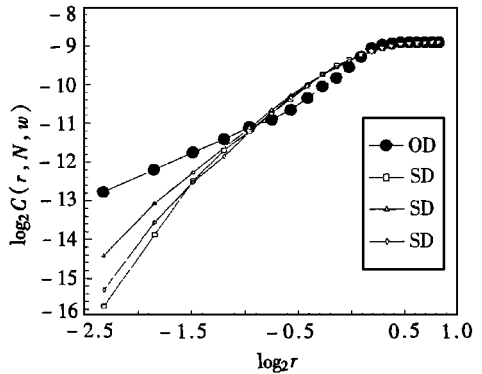
§ 4. 结 论

1. 计算结果表明: 通过正的付氏变换和扭转相位后的逆的付氏变换原始数据和各替代数据之间的均值和均方差均无多大变化。(文献[5]已证明原始数据和替代数据之间的自相关函数相同。)但 $\langle D_{orig} \rangle$ 和 $\langle D_{sur} \rangle$ 因数据的随机或混沌特性的不同, 其变化程度也各不相同, 对于混沌特性的数据判值 Z 变化较大。(2) 组数据和(4) 组数据均为混沌数据, 但(4) 组数据的判值比(2) 组数据的判值要大。在计算过程中, (2) 组数据及(4) 组数据的算法及扭转相位后的随机数取法也相同。其判值不同, 可能是由于(4) 组数据中混沌特征更多一些的缘故。式(3.1) ~ (3.4) 是对每组原始数据扭转相位时, 分别加上 3 组不同的随机数, 每组分别生成 3 组不同的替代数据。各 3 组不同的替代数据间的判值也稍有差别, 但不改变差别的本质。差别的细节还不清楚, 将在以后的文章中继续报道。



OD 为原始数据 SD 为替代数据

图 2 (1) 组数据 $\log_2 r \sim \log_2 C(r, N, w)$ 图



OD 为原始数据 SD 为替代数据

图 3 (2) 组数据 $\log_2 r \sim \log_2 C(r, N, w)$ 图

2. 图 2 至图 6 是第(1) 组至第(4) 组数据及其不同替代数据的 $\log_2 r \sim \log_2 C(r, N, w)$ 对照图。从图中可以看出其原始数据的图线与各替代数据的图线有明显不同, 但各替代数据之间的图线却变化较小。这说明替代数据不因相位上加入随机数的不同而有质的较大变化, 且随机数据与混沌数据之间的判值有明显的不同。这说明本文的算法是好的, 可行的和准确的。

3. 按照(2.1) 式~ (2.10) 式编制的本文的计算程序有以下特点:

a. 进行正的付氏变换时要分别求出原始数据的实部、虚部、幅值和幅角, 随机化其相位时要正确区分其相位所在的象限。(可同时根据其实部和虚部的正、负号来判定。)加入随机数

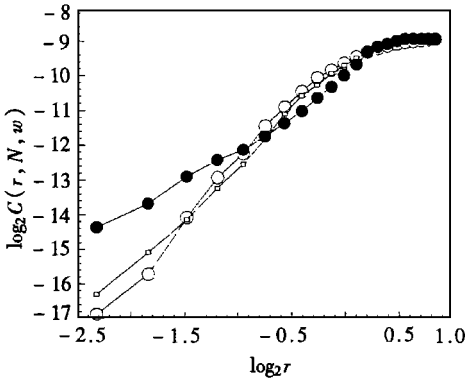
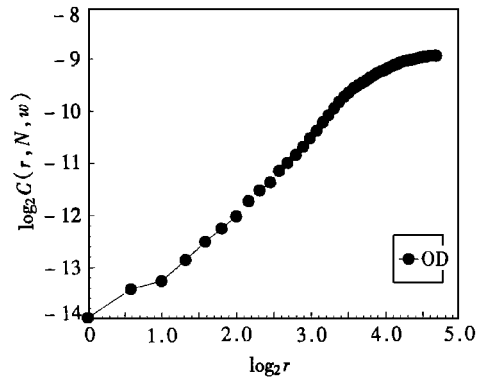
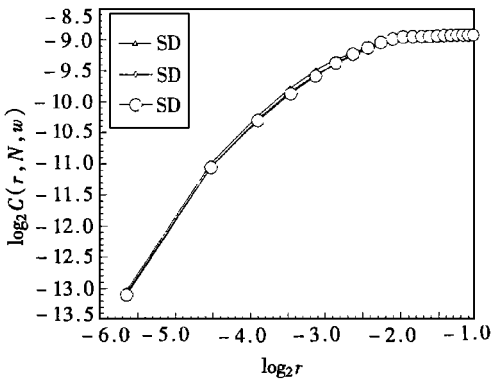


图4 (3)组数据 $\log_2 r \sim \log_2 C(r, N, w)$ 图



OD 为原始数据

图5 (4)组数据 $\log_2 r \sim \log_2 C(r, N, w)$ 图



SD 为替代数据

图6 第(4)组数据原始数据的三组替代数据的 $\log_2 r \sim \log_2 C(r, N, w)$ 图

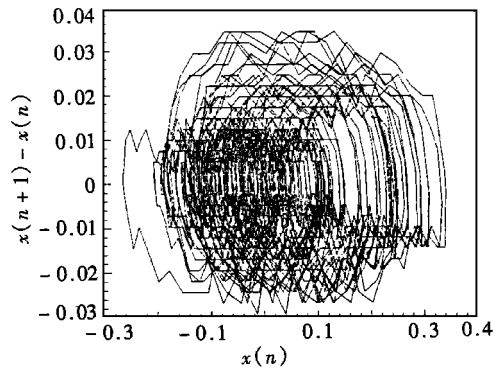


图7 (4)组数据原始数据的相图

前的相位判定错误会引起其计算结果的错误。

b. 由于实际问题中原始数据全为实数,即写成复数后其虚部为零。从坐标象限中来看原始数据全落在 x 轴上,经过正的付里叶变换和扭转相位后的逆的付里叶变换,若其虚部不为零则替代数据必落在整个复平面内。这样会造成原始数据中某些固有特性流失到虚部中。为了保证替代数据落在复平面中的 x 轴上,尽量减少由此而引起的替代数据中信息量的流失,增加判值 Z 的可靠程度。我们认为扭转相位后的付氏逆变换生成的替代数据虚部必须为零。

c. 逆付氏变换中相位加上不同的随机数,其相应的替代数据间只有细微差别而无质的不同。

4. 由于自己设计的机械振动实验是确定性的混沌过程,判值的结果也为确定性的。如何恢复它的确定性工作,将在以后的工作中继续报道。

参 考 文 献

- 1 Henry D. . Abarbanel, Prediction in chaotic nonlinear systems methods for timeseries with broad band Fourier spectra, *Phys. B*, **5** (1991), 1347—1375.
- 2 J. Luis Cabrera and F. Javier, Numerical analysis of transient behavior in the discrete random Logistic equation with delay, *Phys. Lett. A*, **197** (1995), 19—24.
- 3 Peter Grassberger, Finite sample corrections to entropy and dimension estimates, *Phys. Lett. A*, **125** (1988), 369—373.
- 4 James Theiler, Stephen Eubank, etc., Testing for nonlinearity in time series the method of surrogate data, *Physica D*, **58** (1992), 77—94.
- 5 Dean Prichard, The correlation dimension of difference data, *Phys. Lett. A*, **191** (1994), 245—250.
- 6 James Theiler, Spurious dimension from correlation algorithms applied to limited time series data, *Phys. Rev. A*, **34** (1986), 2427—2432.
- 7 S. Rombouts, R. Keunen, Investigation of nonlinear structure in multichannel EEG, *Phys. Lett. A*, **202** (1995), 352—358.
- 8 Matthew B. Kennel and Steven Sabelle, Method to distinguish possible chaos from colored noise and to determine embedding parameters, *Phys. Rev. Lett. A*, **46** (1992), 3111—3118.
- 9 P. E. Rapp and A. M. Albano, Filtered noise can mimic low dimensional chaotic attractors, *Phys. Rev. E*, **47** (1993), 2289—2297.
- 10 Dean Prichard Generating surrogate data for time series with several simultaneously measured variables, *Phys. Rev. Lett.*, **191** (1994), 230—245.
- 11 P. E. Rapp and A. M. Albano, Phase randomized surrogates can produce spurious identifications of non_random structure, *Phys. Lett. A*, **192** (1994), 27—33.
- 12 M. Casdagli and Alistair Mees, Modeling chaotic motions of a string from experimental data, *Phys. Rev. E*, **54** (1992), 303—328.
- 13 P. E. Rapp and A. M. Albano, Predicting chaotic time series, *Phys. Rev. E*, **47** (1993), 2289—2297.
- 14 Eric J. Kostelich, Problems in estimating dynamics from data, *Phys. D*, **58** (1992), 138—152.
- 15 S. J. Schiff and T. Chang, Information transport in temporal systems, *Phys. Rev. Lett. A*, **67** (1992), 378—393.
- 16 James Theiler, Some comments on the correlation dimension of noise, *Phys. Lett. A*, **155** (1991), 480—493.
- 17 J. Timonen and . Koskinen, An improved estimator of dimension and some comments on providing confidence intervals, *Geophys. Res. Lett.*, **20** (1993), 1527—1536.
- 18 D. Prichard and C. P. Price, Reconstructing attractors from scalar time series: A comparison of singular system and redundancy criteria, *Geophys. Res.*, **20** (1993), 2817—2825.

Threshold Value for Diagnosis of Chaotic Nature of the Data Obtained in Nonlinear Dynamic Analysis

Ma Junhai

(Institute of Systems Engineering, Southeast University, Nanjing 210018, P. R. China)

Chen Yushu

(Department of Mechanics, Tianjing University, Tianjing 300072, P. R. China)

Liu Zengrong

(Department of Mathematics, Shanghai University, Shanghai 201800, P. R. China)

Abstract

In this paper surrogate data method of phase-randomized is proposed to identify the random or chaotic nature of the data obtained in dynamic analysis. The calculating results validate the phase-randomized method to be useful as it can increase the extent of accuracy of the results. And the calculating results show that threshold values of the random timeseries and nonlinear chaotic timeseries have marked difference.

Key words chaotic timeseries, surrogate_data, threshold value, random timeseries