第25卷 第2期
1999年3月

自 动 化 学 报
ACTA AUTOMATICA SINICA

Vol. 25, No. 2
Mar., 1999

基于分数维数的非线性相关度及其应用[1)]

樊重俊 王浣尘
(上海交通大学系统工程研究所 上海 200052)

韩崇昭 胡保生
(西安交通大学系统工程研究所 西安 710049)

摘 要 首先说明利用非线性动态系统的多维观测数据和一维观测数据对系统分数维数进行估计的一致性,然后基于这一思想,给出了一种推断两列观测数据是否来自同一非线性动态系统的方法,并引入了非线性相关度的概念,以度量两列数据的非线性相关程度. 该方法可用来解决非线性经济分析与预测中的变量选择问题. 数值结果说明该方法效果较好.

关键词 非线性动态系统,分数维数,时间序列,非线性相关,经济预测.

NONLINEAR DEPENDENCE COEFFICIENT BASED ON FRACTAL DIMENSION AND ITS APPLICATIONS

FAN Chongjun WANG Huanchen
(*Systems Engineering Institute, Shanghai Jiaotong University, Shanghai* 200052)

HAN Chongzhao HU Baosheng
(*Systems Engineering Institute, Xi'an Jiaotong University, Xi'an* 710049)

Abstract In this paper it is proved that the fractal dimension estimate of nonlinear dynamical system with its multivariate observation series is the same as that with its univariate observation series. Based on this result, an inference method is presented, and the Nonlinear Dependence Coefficient is defined. This method is designed for testing nonlinear dependence between time series, and can be used in economic analysis and economic forecasting. Numerical results show that the method is effective.

Key words Nonlinear dynamics, fractal dimension, time series, nonlinear dependence, economic forecasting.

1 引言

经济行为的复杂多变往往是由系统内部各种非线性关系的相互作用导致的,一些非

1) 国家自然科学基金与中国博士后科学基金资助.
收稿日期 1996-04-22 收修改稿日期 1998-02-20

线性方法以其解释复杂经济现象的能力而越来越引起人们的重视,其中浑沌现象的发现和浑沌理论所取得的成果尤为突出.然而,在非线性经济学研究方面所作的工作,大多是利用低阶非线性模型,在现有的经济学框架内讨论经济模型中产生分叉、浑沌等复杂动态的可能性,而涉及到对观测数据的解释与推断方法的研究工作并不多.考虑如下的离散时间非线性动态系统$(h,F,\boldsymbol{x}(0))$

$$\boldsymbol{x}(t)=F(\boldsymbol{x}(t-1)),\qquad \boldsymbol{x}(0)\text{ 给定} \tag{1a}$$

$$\boldsymbol{y}(t)=h(\boldsymbol{x}(t)),\qquad t=1,2,\cdots,n \tag{1b}$$

其中

$$\boldsymbol{x}(t)=(x_1(t),\cdots,x_r(t))'\in R^r,\boldsymbol{y}(t)=(y_1(t),\cdots,y_p(t))'\in R^p$$

分别为状态变量和观测变量,通常 $p=1$.当 $F(\cdot)$函数形式已知时,有很多文献对该系统的非线性统计特征进行了讨论.但往往对 $F(\cdot)$的函数形式一无所知,也不了解该动态系统应该由哪些状态变量组成,甚至连该系统的维数也不知道,所能得到的只是该动态系统一个观测变量的一部分观测数据.然而一维的时间序列做为多维系统的一个侧面,却包含着原经济系统所有变量的痕迹,可以根据它所提供的信息,通过相空间重构方法研究原来的经济系统的动力学特征.很多文献讨论了利用一维观测数据估计系统维数的方法[1-3].并且,基于相关积分的概念,Brock,Dechert 和 Scheinkman 给出了一种检验时间序列非线性相关性的方法(BDS 检验),这一方法得到了广泛的应用,尤其在非线性经济学中被用于检验时间序列的可预测性[4].而利用分数维数来推断两个观测序列非线性相关性的思想,并未见诸文献.在实际中用的较多的分数维数是相关维数.本文首先讨论利用多维观测数据和一维观测数据对系统相关维数进行估计的一致性,然后基于这一结果,引入了非线性相关度的概念,以评价两个观测序列的非线性相关性程度.数值结果说明该方法效果较好.

2 基于相关维数的非线性相关性推断思想

对于系统$(h,F,\boldsymbol{x}(0))$,定义

$$C_n(\boldsymbol{x},\varepsilon)=\frac{2}{n(n-1)}\cdot\sum_{1\leqslant s<t\leqslant n}\theta(\varepsilon-\|\boldsymbol{x}(t)-\boldsymbol{x}(s)\|), \tag{2}$$

其中

$$\theta(a)=\begin{cases}0, & a<0,\\ 1, & a\geqslant 0.\end{cases}$$

若下列极限存在

$$D(\boldsymbol{x})=\lim_{\varepsilon\to+0}\lim_{n\to\infty}\ln C_n(\boldsymbol{x},\varepsilon)/\ln\varepsilon, \tag{3}$$

则称 $D(\boldsymbol{x})$是系统$(h,F,\boldsymbol{x}(0))$的相关维数.通常范数取为

$$\|\boldsymbol{x}(t)-\boldsymbol{x}(s)\|=\max_{i=1,\cdots,r}|x_i(t)-x_i(s)|. \tag{4}$$

事实上,(2)式可看成是:系统轨道上任取两点,其距离小于 ε 的概率.

设已知 p 维观测向量 y 的观测数据为$\{\boldsymbol{y}(t):t=1,\cdots,n\}$,构造

$$\boldsymbol{y}^m(t)=(\boldsymbol{y}(t)',\cdots,\boldsymbol{y}(t+m-1)')',$$

其中 m 称为嵌入维数,定义相关积分

$$C_{m,n}(\boldsymbol{y},\varepsilon)=\frac{2}{N(N-1)}\cdot\sum_{1\leqslant s<t\leqslant N}\theta(\varepsilon-\|\boldsymbol{y}^m(t)-\boldsymbol{y}^m(s)\|),$$

其中 $N=n-m+1$,由此给出相关维数估计为

$$D_{m,n}(\boldsymbol{y},\varepsilon)=\ln C_{m,n}(\boldsymbol{y},\varepsilon)/\ln\varepsilon. \tag{5}$$

可以证明(见本文定理1,推论1),

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}D_{m,n}(\boldsymbol{y},\varepsilon)=D(\boldsymbol{x}),\qquad \text{当 } m>2\cdot D(\boldsymbol{x})+1, \tag{6}$$

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}D_{m,n}(\boldsymbol{z},\varepsilon)=D(\boldsymbol{x}),\boldsymbol{z}=(y_{i_1},\cdots y_{i_q})'. \tag{7}$$

特别的

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}D_{m,n}((y_1,y_2)'\varepsilon)=\lim_{\varepsilon\to 0}\lim_{n\to\infty}D_{m,n}((y_1,\varepsilon). \tag{8}$$

下面假设 $p=2$. 若 y_1,y_2来自不同的动态系统(即认为 y_1,y_2独立),则应有

$$P(\|\boldsymbol{y}^m(t)-\boldsymbol{y}^m(s)\|<\varepsilon)=P(\|\boldsymbol{y}_1^m(t)-\boldsymbol{y}_1^m(s)\|<\varepsilon,$$

$$\|\boldsymbol{y}_2^m(t)-\boldsymbol{y}_2^m(s)\|<\varepsilon)=P(\|\boldsymbol{y}_1^m(t)-\boldsymbol{y}_1^m(s)\|<\varepsilon)\cdot P(\|\boldsymbol{y}_2^m(t)-\boldsymbol{y}_2^m(s)\|<\varepsilon),$$

其中 P 表示概率,而由下式依概率收敛结果

$$C_{m,n}(\boldsymbol{y},\varepsilon)\xrightarrow{P}P(\|\boldsymbol{y}^m(t)-\boldsymbol{y}^m(s)\|<\varepsilon),n\to\infty.$$

因此得

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\ln(C_{m,n}((y_1,y_2)',\varepsilon))/\ln(\varepsilon)=$$

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}\ln(C_{m,n}(y_1,\varepsilon))/\ln(\varepsilon)+\lim_{\varepsilon\to 0}\lim_{n\to\infty}\ln(C_{m,n}(y_2,\varepsilon))/\ln(\varepsilon),$$

即

$$\lim_{\varepsilon\to 0}\lim_{n\to\infty}D_{m,n}((y_1,y_2)',\varepsilon)=\lim_{\varepsilon\to 0}\lim_{n\to\infty}D_{m,n}(y_1,\varepsilon)+\lim_{\varepsilon\to 0}\lim_{n\to\infty}D_{m,n}(y_2,\varepsilon). \tag{9}$$

注意到下式结果

$$D_{m,n}(c\boldsymbol{y},\varepsilon)=\frac{D_{m,n}(\boldsymbol{y},\varepsilon/c)}{1+\ln c/\ln(\varepsilon/c)},$$

$$\ln c/\ln(\varepsilon/c)\xrightarrow{\varepsilon\to 0}0, \tag{10}$$

其中 $c\boldsymbol{y}$ 表示序列$\{c\boldsymbol{y}(t),t=1,\cdots,n\}$,$c$ 是一常数. 由此知,当取 ε 充分小时估计相关维数的(5)受量纲影响不大. 然而,来自经济系统的观测数据一般长度不大,此时无法取 ε 充分小,(10)式无法满足,用于估计相关维数的(5)受量纲影响很大. 为此定义下式

$$D_{m,n}(\boldsymbol{y},\varepsilon_1,\varepsilon_2)=\ln(C_{m,n}(\boldsymbol{y},\varepsilon_1)/C_{m,n}(\boldsymbol{y},\varepsilon_2))/\ln(\varepsilon_1/\varepsilon_2). \tag{11}$$

(11)式不受量纲影响,并且有

$$\lim_{\varepsilon_1\to 0,\varepsilon_2\to 0}\lim_{n\to\infty}D_{m,n}(\boldsymbol{y},\varepsilon_1,\varepsilon_2)=\lim_{\varepsilon\to 0}\lim_{n\to\infty}D_{m,n}(\boldsymbol{y},\varepsilon).$$

3　基于相关维数的非线性相关度概念

可利用式(8),(9)的结果,对观测变量 y_1,y_2的非线性相关性进行推断:若 $D_{m,n}((y_1,y_2)',\varepsilon),D_{m,n}(y_1,\varepsilon),D_{m,n}(y_2,\varepsilon)$三项差别不大,则可认为观测变量 y_1,y_2来自同一动态系统,即 y_1,y_2具有很强的非线性相关性,预测时应做为一个总体来考虑;若 $D_{m,n}((y_1,y_2)',\varepsilon)$与 $D_{m,n}(y_1,\varepsilon),D_{m,n}(y_2,\varepsilon)$两项之和差别不大,则可认为 y_1和 y_2是完全没有关系的两个指标,即 y_1,y_2是非线性不相关的,预测时彼此不能提供信息. 然而,对于经济数据,大多数

情况则是介于两者之间. 事实上,在统计理论中,采用相关系数来反映两个变量的线性相关程度,为此,本文定义如下的两个序列的非线性相关度的概念.

定义1. 对于序列$\{y_1(t),t=1,\cdots,n\}$与$\{y_2(t),t=1,\cdots,n\}$定义

$$R_{m,n}(y_1,y_2,\varepsilon)=\frac{I_{m,n}(y_1,y_2,\varepsilon)}{D_{m,n}((y_1,y_2)',\varepsilon)},\tag{12}$$

其中

$$I_{m,n}(y_1,y_2,\varepsilon)=D_{m,n}(y_1,\varepsilon)+D_{m,n}(y_2,\varepsilon)-D_{m,n}((y_1,y_2)',\varepsilon),$$

则称式(12)为序列$\{y_1(t),t=1,\cdots,n\}$与$\{y_2(t),t=1,\cdots,n\}$的非线性相关度.

易知,当式(8)成立时非线性相关度为1,当式(9)成立时非线性相关度为0,其它情况则介于0,1之间,非线性相关度越大,非线性相关程度越强. 因此,利用非线性相关度来反映序列 $y_1(t),y_1(t)$的非线性相关程度是合理的. 另外有

$$R_{m,n}(cy_1,cy_2,\varepsilon)=R_{m,n}(y_1,y_2,\varepsilon/c),\tag{13}$$

其中 cy_1与 cy_2表示序列$\{cy_1(t),t=1,\cdots,n\}$与$\{cy_2(t),t=1,\cdots,n\}$,c 是一常数. 式(13)说明了定义1所给出的非线性相关度即使是在小样本的情况下,也不受量纲的影响,具有很好的稳定性.

4 一些理论结果

这里证明结果(6).

假设1. 对于系统$(h,F,\boldsymbol{x}(0))$

1) h 和 F 是 C^2光滑的. F 具有唯一的紧吸引子 Λ,且存在轨道在 Λ 上稠密. F 具有唯一的遍历不变测度 ρ,且 ρ 具有连续密度 $\rho(\mathrm{d}x)=j(x)\mathrm{d}x$;

2) 由 $\boldsymbol{x}(0)$所确定的轨道在 Λ 上稠密;

3) 最大 Lyapunov 指数为正.

引理1. 对于系统$(h,F,\boldsymbol{x}(0))$,在假设1和 $m>2\cdot D(\boldsymbol{x})+1$下,存在 K 使得

$$K\cdot\|\boldsymbol{x}(t)-\boldsymbol{x}(s)\|<\|\boldsymbol{y}_1^m(t)-\boldsymbol{y}_1^m(s)\|<K^{-1}\cdot\|\boldsymbol{x}(t)-\boldsymbol{x}(s)\|.$$

证明. 见文献[2]定理2.5的证明过程.

定理1. 对于系统$(h,F,\boldsymbol{x}(0))$,在假设1和范数(4)的意义下,结果(6)成立.

证明. 由引理1可得,存在$\{K_i,i=1,\cdots,p\}$,使得

$$K_i\cdot\|\boldsymbol{x}(t)-\boldsymbol{x}(s)\|<\|\boldsymbol{y}_i^m(t)-\boldsymbol{y}_i^m(s)\|<K_i^{-1}\cdot\|\boldsymbol{x}(t)-\boldsymbol{x}(s)\|,i=1,\cdots,p.$$

取

$$K=\min\{K_i,i=1,\cdots,p\}.$$

在范数(4)的意义下有

$$\|\boldsymbol{y}^m(t)-\boldsymbol{y}^m(s)\|=\max_{i=1,\cdots,p}\|\|\boldsymbol{y}_i^m(t)-\boldsymbol{y}_i^m(s)\|,$$

$$K\cdot\|\boldsymbol{x}(t)-\boldsymbol{x}(s)\|<\|\boldsymbol{y}^m(t)-\boldsymbol{y}^m(s)\|<K^{-1}\cdot\|\boldsymbol{x}(t)-\boldsymbol{x}(s)\|.$$

因此

$$C_n(\boldsymbol{x},K\varepsilon)<C_{m,n}(\boldsymbol{y},\varepsilon)<C_n(\boldsymbol{x},K^{-1}\varepsilon).$$

而

$$\lim_{\varepsilon\to+0}\lim_{n\to\infty}\ln C_n(\boldsymbol{x},K\varepsilon)/\ln\varepsilon=\lim_{\varepsilon\to+0}\lim_{n\to\infty}\ln C_n(\boldsymbol{x},K\varepsilon)/\ln K\varepsilon=$$

$$\lim_{\varepsilon\to+0}\lim_{n\to\infty}\ln C_n(\boldsymbol{x},\varepsilon)/\ln\varepsilon,$$

因此结果(6)成立.

推论1. 对于系统$(h,F,\boldsymbol{x}(0))$,任取$\boldsymbol{y}(t)$的q个分量,记为

$$z_i(t)=y_{k_i}(t)\in\{y_1(t),\cdots,y_r(t)\},i=1,\cdots,q,$$

$$\boldsymbol{z}(t)=(z_1(t),\cdots,z_q(t))'.$$

在定理1的意义下,通常有

$$\lim_{\varepsilon\to0}\lim_{n\to\infty}D_{m,n}(\boldsymbol{z},\varepsilon)=d.$$

推论1给出了利用任意多维观测数据估计系统维数的方法,并说明了利用多维观测数据和一维观测数据对系统维数进行估计的一致性.

推论2. 对于系统$(h,F,\boldsymbol{x}(0))$,在定理1的意义下,通常有

$$\lim_{\varepsilon\to0}\lim_{n\to\infty}R_{m,n}((y_1,y_2)',\varepsilon)=1.$$

推论2说明对于来自同一具有唯一紧吸引子的动态系统的两个观测变量,只要观测数据足够多,总能保证这两个观测变量数据序列的非线性相关度为1.

5　数值计算例子

通过两个数值计算实例来说明本文方法的用法及其有效性.

例1. 考虑如下的 Henon 映射[5]

$$x_1(t+1)=1-a\cdot x_1(t)^2+x_2(t),$$

$$x_2(t+1)=b\cdot x_1(t),$$

$$y_1(t+1)=0.2\cdot x_2(t+1)+0.2\cdot x_1(t+1),$$

$$y_2(t+1)=0.3\cdot x_2(t+1)-0.3\cdot x_1(t+1),$$

y_1,y_2是两个观测变量,另设观测数据$\{v(t):t=1,\cdots,n\}$来自如下的 Logistic 映射

$$u(t+1)=A\cdot u(t)\cdot(1-u(t)),$$

$$v(t+1)=u(t+1).$$

参数与初始值取

$$a=1.4,\quad b=0.3,\quad x_1(-10000)=x_2(-10000)=0,$$

$$A=4,\quad u(-10000)=0.1,$$

以保证观测数据$\{y_1(t):t=1,\cdots,n\}$,$\{y_2(t):t=1,\cdots,n\}$,$\{v(t):t=1,\cdots,n\}$稳定,$n=2\,000$. 基于公式(11)的数值结果如表1.

仿真结果说明,对于来自于同一动态系统的两个观测变量 y_1,y_2,即 y_1,y_2,具有很强的非线性关系,则 $D_{m,n}(y_1,y_2)$、$D_{m,n}(y_1)$、$D_{m,n}(y_2)$三项差别不大,此时 y_1,y_2的非线性相关度接近于1. 对于来自于不同动态系统的两个观测变量 y_1,v,即 y_1,v 之间不存在非线性关系,则近似的有

$$D_{m,n}(y_1,v)=D_{m,n}(y_1)+D_{m,n}(v),$$

此时 y_1,v 的非线性相关度接近于0.

表1 基于公式(11)的观测变量数值结果

m	4	4	5	5	6	6
ε_1	0.010	0.008	0.010	0.008	0.010	0.008
ε_2	0.005	0.004	0.005	0.004	0.005	0.004
$D_{m,n}(y_1,\varepsilon_1,\varepsilon_2)$	1.103 1	1.126 2	1.105 0	1.151 3	1.121 0	1.184 6
$D_{m,n}(y_2,\varepsilon_1,\varepsilon_2)$	1.212 5	1.181 4	1.212 6	1.144 8	1.193 9	1.131 3
$D_{m,n}(y_1,y_2,\varepsilon_1,\varepsilon_2)$	1.189 3	1.164 2	1.182 3	1.150 5	1.162 1	1.158 6
$R_{m,n}(y_1,y_2,\varepsilon_1,\varepsilon_2)$	0.947 0	0.982 2	0.960 2	0.995 7	0.991 9	0.998 8
$D_{m,n}(v,\varepsilon_1,\varepsilon_2)$	1.016 1	0.987 5	1.099 6	1.033 2	1.159 6	1.078 3
$D_{m,n}(y_1,v,\varepsilon_1,\varepsilon_2)$	2.032 5	2.013 0	2.044 0	2.163 2	2.136 9	2.153 5
$R_{m,n}(y_1,v,\varepsilon_1,\varepsilon_2)$	0.042 0	0.050 0	0.078 5	0.009 8	0.067 2	0.050 8

例2. 考虑伦敦有色金属市场镍价格，设 x,y 分别表示镍成交价、三个月期货价，选取从1994年10月到1996年1月的有关数据，并剔除趋势项，做规格化处理. 基于公式(11)的计算结果如表2. 从数据结果可以看出，镍成交价与三个月期货价具有很强的相关性，预测时应做为一个总体来考虑. 根据本文方法进行变量选择所建立的非线性模型如神经网络预测模型[6]，实际预测效果较好.

表2 伦敦有色金属市场镍价格数值结果

m	7	7	8	8	9	9
ε_1	0.08	0.07	0.09	0.08	0.10	0.09
ε_2	0.04	0.035	0.045	0.04	0.05	0.045
$D_{m,n}(x,\varepsilon_1,\varepsilon_2)$	1.586 8	1.856 1	1.595 3	1.822 1	1.571 8	1.782 1
$D_{m,n}(y,\varepsilon_1,\varepsilon_2)$	1.528 5	1.783 2	1.525 9	1.751 2	1.484 2	1.700 1
$D_{m,n}(x,y,\varepsilon_1,\varepsilon_2)$	1.652 5	1.957 5	1.651 1	1.892 3	1.620 9	1.844 0
$R_{m,n}(x,y,\varepsilon_1,\varepsilon_2)$	0.836 2	0.859 1	0.842 1	0.889 5	0.886 4	0.888 3

6 结语

本文给出了一种度量两列观测数据非线性相关程度的方法，以解决非线性经济预测中的变量选择问题. 式(6)是本文的中心思想，虽然其数学证明并不困难，但却为推断两列观测数据的非线性相关性提供了一种重要思路. 推论1还说明了利用多维观测数据和一维观测数据对系统分数维数进行估计的一致性. 数值结果说明，本文给出的推断方法效果较好.

本文提出的非线性相关度概念，还不很成熟，需要接受实际应用的考验，并在应用中不断对其理论进行完善. 混沌序列的奇异吸引子理论、分数维数概念与嵌入技术在复杂经济系统预测中应用，有广泛的研究前景，如何利用奇异吸引子来构造预测模型，本文作者已开始考虑这个问题. 相信这个问题比混沌序列的非线性相关更具有实际意义，对于传统

的预测模型来说也更具有挑战性.

参 考 文 献

1 Grassberger P, Procaccia I. Measuring the strangeness of strange attractors. *Physical Review*, *Ser. D*, 1983, **9**(1): 189－208

2 Brock W A. Distinguishing random and deterministic systems: abridged version. *Journal of Economic Theory*, 1986, **40**(2): 168－195

3 Takens F. Detecting nonlinearities in stationary time series. *International Journal of Bifurcation and Chaos*, 1993, **3**(2): 241－256

4 Scheinkman J, LeBaron B. Nonlinear dynamics and stock returns. *Journal of Business*, 1989, **62**(3): 311－337

5 Henon M. A two-dimensional mapping with a strange attractor. *Communacations in Mathematical Physics*, 1976, **50**(1): 69－77

6 Fan Chongjun, Han Chongzhao, Hu Baosheng. Forecasting the Behavior of Multivariate Time Series with Neural Networks in a DSS for Business Planning. In: Proceedings of the third China-Japan International Symposium on Industrial Management, Beijing: International Academic Publishers, 1996, 653－657

樊重俊　1963年生. 分别于1984, 1990, 1996年在复旦大学、武汉大学、西安交通大学获学士、硕士、博士学位. 现为上海交通大学系统工程所博士后、副教授. 研究领域为决策理论与决策支持系统, 计算机应用, 非线性分析, 经济分析与预测, 可持续发展等. 发表论文30余篇.

王浣尘　1933年生. 现为上海交通大学系统工程所所长、博士生导师. 研究领域为控制理论, 系统工程, 管理科学等.

韩崇昭　1943年生. 1968年毕业于西安交通大学, 1981年毕业于中国科学院研究生院. 现为西安交通大学电信学院副院长、博士生导师. 目前研究方向为非线性频谱分析, 决策支持系统等.

胡保生　1930年生. 1951年毕业于上海大同大学. 现为西安交通大学系统工程研究所所长、博士生导师. 目前研究方向为计算机集成制造系统, 并行控制算法等.