

时间序列确定性的指标检验方法

张 硕, 惠晓峰

(哈尔滨工业大学经济及管理学院, 黑龙江 哈尔滨 150001)

摘 要: 对于时间序列的生成机理是否服从于确定规律性的判断将直接影响到对其进行研究的理论框架的选择。提出了一种时间序列确定性成分的指标检验方法。该方法首先将待检验的时间序列重构于高维相空间中, 寻找其主要几何特征在低维欧式空间中的投影, 再将投影反射回高维相空间。通过比较原高维相空间与反射空间之间的偏差, 来判断该序列的确定性。仿真实验证明了该方法具有可靠性。

关键词: 时间序列; 确定性; 几何特征; 随机性; 奇异值分解

中图分类号: N 94

文献标志码: A

DOI: 10.3969/j.issn.1001-506X.2012.03.19

Index method to detect determinism in time series

ZHANG Shuo, HUI Xiao-feng

(Economics and Management School, Harbin Institute of Technology, Harbin 150001, China)

Abstract: The determinism of time series has a direct impact on the choice of the theoretical framework. A method to detect the determinism of nonlinear time series is proposed. This method first reconstructs the time series into a high dimensional space. Then the low-dimensional projector of the high-dimensional manifolds is found in a low-dimensional Euclidean space. When the projector is reflected back to the high-dimensional space, there are differences between the projector and the high-dimensional manifolds. By comparing the high-dimensional manifolds and the projector reflected back to the high-dimensional space, the determinism of the time series can be detected. The simulation results show that the method is reliable.

Keywords: time series; determinism; geometry; randomness; singular value decomposition (SVD)

0 引 言

对复杂系统产生时间序列数据的潜在规律性判断将直接影响到对该系统所采取的研究范式的选择。其中, 对于时间序列是否受制于某种确定规则的研究一直是时间序列研究的一个重要问题。传统确定性检验的方法主要包括吸引子相关维数^[1]、Kolmogorov 熵^[2]、Lyapunov 指数^[3]等。这些方法需要的数据量较大, 且运算量大, 计算繁琐。同时, 这些方法对采样周期十分敏感^[4]。在系统噪声含量较高的情况下, 相关维数、Kolmogorov 熵、Lyapunov 指数的值将会受到干扰。文献[5]通过实验检验证明, 当信噪比小于 10 dB 时, 通过这些方法对序列的确定性规律判断将受到影响。奇异谱方法也广泛应用于确定性检验^[6]。该方法主要基于 Takens 吸引子重构定理在高维相空间恢复原始时间序列的轨道矩阵 \mathbf{X} , 由于随机序列数据之间不存在相关性, 所以 $\mathbf{X}\mathbf{X}^T$ 的奇异值接近, 归一化奇异谱分布比较平坦。由于确定性序列数据之间存在相关性, 所以 $\mathbf{X}\mathbf{X}^T$ 的奇异值较

为陡峭。然而, 文献[7]指出奇异谱方法是基于线性变化所得出的方法, 在研究非线性奇异吸引子的问题上将可能得出错误的结论。

由于对时间序列确定性检验的重要性, 有很多学者提出了一系列比较成功的新方法, 主要有递归图法^[8]、辛几何谱法^[9-11]及延迟向量差异 (delay vector variance, DVV) 方法^[12]等。递归图法通过把高维相空间的轨迹映射到可以观测的 2-3 维空间中, 通过观察低维空间的图像是否具有某种规律性来判断系统是否存在可预测性, 从而定性地判断是否为确定性系统。辛几何谱方法与奇异谱较类似, 主要是通过原数据与替代数据进行辛相似变换, 获得它们的辛几何谱, 在二维平面上对得到的原数据与替代数据的辛几何谱进行对比, 来检验时间序列的确定性。DVV 方法通过检验目标方差的变化来检验系统的确定性。它指出如果数据的确定性较强, 那么邻近相似的轨迹点演化的目标值也应该相似。因此, 通过分割轨迹来计算目标方差, 在分割区间逐渐变大的过程中, 目标方差的取值变化不应过大。

收稿日期: 2011-05-13; 修回日期: 2011-09-30。

基金项目: 国家自然科学基金(71173060)资助课题

作者简介: 张硕(1978-), 男, 博士研究生, 主要研究方向为系统工程。E-mail: surezhang4@163.com

递归图法和辛几何谱法主要是通过直观的形式对时间序列的确定性进行了检验。然而,实验数据在获取的同时往往无法摆脱环境噪声的影响。由于观察噪声对实验数据的污染,所以直观的检验形式往往受到比较严重的干扰,而影响到检验结果的准确性。DVV 方法虽然立足于统计手段对数据的确定性进行检验,但是当噪声非常大时,也就是数据只具有弱确定性的时候,仍然会产生误判。同时,这些方法在对实验数据检验的时候需要生成与被检验数据具有相同统计特性的替代数据,这就增加了检验过程的复杂程度。

本文基于确定性时间序列与随机时间序列在高维重构相空间中的局部拓扑特征之间的区别,设计了一种可以在高噪声背景下检验时间序列确定性的检验方法。此方法通过指标的形式对时间序列的确定性进行检验。

1 方法的推导与设计

1.1 方法的原理

确定性时间序列产生的系统可以描述如下

$$\mathbf{X}_{i+1} = f(\mathbf{X}_i) \quad (1)$$

通常观测所获得的确定性时间序列往往是被噪声污染过的,则描述如下:

$$\mathbf{S}_{i+1} = f(\mathbf{S}_i) + \boldsymbol{\varepsilon}_i \quad (2)$$

式中, $\mathbf{X}_{i+1}, \mathbf{X}_i, \mathbf{S}_{i+1}, \mathbf{S}_i \in \mathbf{R}^d, \mathbf{R}^d$ 表示 d 维欧氏空间; $\boldsymbol{\varepsilon}_i$ 为噪声; f 为 \mathbf{C}^r 映射, \mathbf{C}^r 表示 r 阶导数存在。

该系统在一维空间的映射按照时间顺序排列,得到的即为确定性时间序列。

由 Takens 定理^[13]可知一个维数 d_e 的流形可以完全展开(无交点)地嵌入到一个 $d \geq 2d_e + 1$ 的欧氏空间中去。所以,如果将一段确定性的一维时间序列重构到一个 d 维的相空间中去,从而得到一个 d 维流形,只要 $d \geq 2d_e + 1$,则在重构的相空间中可以恢复原系统的所有动力学特征。由 d 维流形定义可知任何一个 d 维的 Hausdorff 拓扑空间,其上每一个点的一个邻域可以通过一个维的欧氏空间描述其全部拓扑信息。所以当通过相空间重构方法,把如式(2)的 d 维确定性系统映射到 m 维空间得到一个新的空间 \mathbf{M}_m , 该系统的局部拓扑特征仍然可以通过一个 d 维的欧氏空间描述。而系统存在的观察噪声则在重构空间的每个维度上表现为相似分布。如果把 d 维嵌入空间从 m 维空间识别出来,则该 d 维空间将不包含与其垂直的 $m-d$ 维度上的噪声分量。 m 越大,则分离出来的 d 维空间含的噪声就越少。所以,含噪的确定性系统在维度为 m 的重构相空间 \mathbf{M}_m 中,当 $m \gg d$, 则存在一个 d 维嵌入可以近似描述确定性系统的全部拓扑信息;而随机过程在重构相空间 \mathbf{M}_m 中,不具有一个 d 维嵌入可以描述其全部拓扑信息。

所以,对于随机过程和一个 d 维确定性系统,把它们都重构到一个高维相空间 \mathbf{M}_m 中,提取 d 维嵌入空间之后,它们在高维相空间 \mathbf{M}_m 中的剩余信息将会有所不同。可以此为判据来研究一段时间序列的确定性问题。

基于前面的论述,对系统同确定性是别的问题转化为两个问题:①把 d 维嵌入空间从 m 维空间识别出来;② 剩余信息的表征量。

首先描述把 d 维嵌入空间从 m 维空间识别出来的方法。一个微分流形的局部信息可以通过该流形的局部切空间来描述。所以,可以把一个 d 维流形的局部子空间仿射到其局部切空间,而用这个仿射空间表示其局部拓扑信息。文献[14]证明了局部切空间可以用局部样本协方差矩阵的特征向量来表示,因此可以把求取表示局部信息的样本点在局部切空间中的投影坐标的问题转换为局部奇异值分解(singular value decomposition, SVD)问题。当一个由式(2)表述的 d 维的确定性系统通过相空间重构到一个 m 维的重构相空间 $\mathbf{M}_m (m > d)$, 其某一个足够小的子空间可以看作空间 \mathbf{M}_m 中的一个空间椭球。该椭球的产生可看作是由一个 \mathbf{M}_m 空间中的单位球映射生成的,而该子空间的奇异值则表征了对单位球半径通过怎样地拉伸或压缩而得到了新的空间椭球。对于一个由连续随机分布构成的 \mathbf{M}_m 空间中的一个子流形,它的一个足够小的子空间仍然是在 \mathbf{M}_m 的各维度上近似相同分布的,所以该子空间的奇异值近似相等。而对于具有确定空间结构的确定性系统,该空间椭球将在 d 维空间中被扭曲,所以必有 d 个奇异值远大于其他奇异值。

设给定的一个流形空间的样本点集为 $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, $\mathbf{x}_i \in \mathbf{R}^m$ 。 \mathbf{x}_i 点的邻域空间可以用其 k 个邻近点表示。 $\mathbf{X}_i^{(k)} = \{\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{ik}\}$, 设 $\bar{\mathbf{x}}_i$ 为该邻域的中心点,则该邻域在切空间的仿射坐标可以表示为

$$\boldsymbol{\Theta}_k = \{\boldsymbol{\theta}_1^{(i)}, \boldsymbol{\theta}_2^{(i)}, \dots, \boldsymbol{\theta}_k^{(i)}\} = \mathbf{Q}_i^T (\mathbf{X}_i^{(k)} - \bar{\mathbf{x}}_i \mathbf{I}_k^T) \quad (3)$$

式中, $\mathbf{Q}_i, \boldsymbol{\Sigma}_i, \mathbf{V}_i^T$ 为 $\mathbf{X}_i^{(k)} - \bar{\mathbf{x}}_i \mathbf{I}_k^T$ 的奇异值分解, \mathbf{I}_k 为 k 维全 1 列向量。当只对空间前 d 维最大奇异值感兴趣的时候,可以取奇异值分解 $\mathbf{Q}_i^* \boldsymbol{\Sigma}_i^* \mathbf{V}_i^{*T}$, $\mathbf{Q}_i^*, \boldsymbol{\Sigma}_i^*, \mathbf{V}_i^*$ 分别为 $\mathbf{Q}_i, \boldsymbol{\Sigma}_i, \mathbf{V}_i$ 的前 d 行。这样就可以在得到的仿射空间中保留该邻域最突出的 d 维拓扑信息。则式(3)转化为

$$\boldsymbol{\Theta}_k = \{\boldsymbol{\theta}_1^{(i)}, \boldsymbol{\theta}_2^{(i)}, \dots, \boldsymbol{\theta}_k^{(i)}\} = \mathbf{Q}_i^{*T} (\mathbf{X}_i^{(k)} - \bar{\mathbf{x}}_i \mathbf{I}_k^T) \quad (4)$$

$$\mathbf{X}_i^{(k)*} = \mathbf{Q}_i^* \boldsymbol{\Theta}_k + \bar{\mathbf{x}}_i \mathbf{I}_k^T \quad (5)$$

$$\mathbf{E} = \mathbf{X}_i^{(k)*} - \mathbf{X}_i^{(k)*} \quad (6)$$

式中, \mathbf{Q}_i^* 为正交矩阵,所以,对于一个由式(1)描述的 d 维确定性系统, $\mathbf{E} \approx \mathbf{0}$ 。而对于一个由式(2)描述的 d 维确定性系统, \mathbf{E} 表现了由于噪声影响而造成的新生成拓扑空间与原拓扑空间之间的拓扑信息的差异。而对于一个由随机序列生成的空间,这种拓扑信息的差异将会更为明显。

这个表征拓扑信息差异的量可以选择 \mathbf{E} 的谱范数。

接下来证明对于确定性系统和随机系统,在由式(4)和式(5)所引导的空间变化之后,其谱范数将会出现差异。

为了使问题更加清晰,其数学描述如下:

设有两个 2 阶以上的实矩阵 \mathbf{A} 和 \mathbf{B} 。 $\mathbf{A} = \mathbf{X}_{m \times k} + \boldsymbol{\varepsilon}_{m \times k}$, $\mathbf{B} = \boldsymbol{\eta}_{m \times l} (k, l \geq m)$, $\mathbf{X}_{m \times k}$ 的非零奇异值的个数 d 小于 m , $\boldsymbol{\varepsilon}_{m \times k}$ 和 $\boldsymbol{\eta}_{m \times l}$ 为随机分布噪声。 $\|\mathbf{A}\|_2 = a$, $\|\mathbf{B}\|_2 = b$ 。欲证明: $\|\mathbf{E}_A\|_2$ 与 $\|\mathbf{E}_B\|_2$ 之间存在固定的大小关系, \mathbf{E}_A 为矩

2 实例检验

接下来通过仿真实验来验证本方法的有效性。选取数据组 I 为标准正态分布;数据组 II 为均值 μ 为 10, 方差 σ^2 为 100 的正态分布;数据 III 为 μ_1 为 1, μ_2 为 -1, 方差 σ_1^2 为 2, σ_2^2 为 4 的均匀混合高斯分布;数据 IV 为 Lorenz 方程;数据 V 为 Rössler 方程;数据 VI 为 Chens 方程。样本数据各取 1 000 点, m 取 40, d 取 3, j 取 960, 计算 E_j^* 如表 1 所示。

表 1 不同类型数据的 E_j^*

	I	II	III	IV	V	VI
E_j^*	651.72	682.92	638.97	3.068	0.293 9	5.320

由表 1 可见,正如前文论述,随机时间序列的 E_j^* 数量级基本近似,而确定性系统的 E_j^* 数量级均远小于随机系统的 E_j^* 。

接下来研究 m 的取值对 E_j^* 的影响,如表 2 所示。

表 2 m 对 E^* 的影响

m	I	II	III	IV
$m=20$	479.07	523.28	480.46	1.549 0
$m=30$	568.21	614.47	557.64	1.796 8
$m=40$	651.74	682.92	638.97	3.068 2

设正态分布数据的 E^* 为 E^{**} , 计算数据 m 取不同值时各数据组 E^* 对 E^{**} 的比值 D_e 如表 3 所示。

表 3 m 取不同值时各数据组 E^* 对正态分布 E^{**} 的比值

m	II	III	IV
$m=20$	1.092 3	1.002 9	0.002 7
$m=30$	1.081 4	0.981 4	0.003 2
$m=40$	1.047 8	0.980 4	0.004 7

由表 3 可见, m 对各数据的 E_j^* 与正态分布数据的 E_j^* 值的比值影响不大。

由于观测噪声的存在对系统具有较大的影响,在系统有噪声污染的情形下对方法的有效性进行研究,取数据组 IV, 分别加入信噪比为 0 dB 的噪声生成数据 VII 和 -10 dB 的噪声生成数据 VIII, 在 m 取 40, d 取 3, j 取 960, 计算 D_e 如表 4 所示。

表 4 噪声污染的情况下的 D_e^*

数据类型	VII	VIII
D_e	0.034 1	0.132 4

由表 4 可见,在噪声污染的情况下,数据组的 E_j^* 受到了影响,信噪比越小则影响越严重,但是 E_j^* 值仍然远小于随机分布的 E_j^{**} , 说明该方法的抗噪能力比较强,同时,以上的实验说明数据的 E_j^* 的值越大,其运动规律越接近于

随机运动。

接下来验证该方法对高维混沌检验的可行性,选取文献[15]中的模拟脑电图(electroencephalogram, EEG)非线性耦合方程产生的信号,取 1 000 个样本,设为数据组 IX, 如图 1 所示。

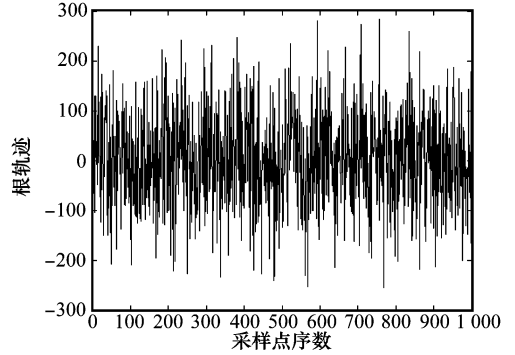


图 1 EEG 非线性耦合信号

该信号的维度约为 9, m 取 40, 对 d 分别取 1~7, 对该信号求取 D_e , 如表 5 所示。

表 5 d 变化下 EEG 非线性耦合信号与标准正态分布的 E_j^* 的比值 D_e

d	1	2	3	4	5	6	7
D_e	0.14	0.13	0.14	0.14	0.15	0.15	0.15

由表 5 可见,该信号有明显的确定性特征, d 的变化对比值的影响不大。

由上述实验结果可知,在对任意一段时间序列进行确定性检验时,只需计算此时间序列的 E_j^* 与在相同的 m, d, j 条件下的标准正态分布的 E_j^* 的比值 D_e , 只要 D_e 远小于 1, 此时间序列即为确定性系统产生的,而不需要生成此时间序列的同分布替代数据。

3 结论

为了尽量削弱高强度观测噪声对确定性检验结果的影响,本文设计了一种通过指标的形式对时间序列的确定性进行检验的方法。通过一些典型的混沌吸引子和随机系统采用仿真的方法进行了检验,验证了算法的有效性。还对上述时间序列分别添加不同信噪比的噪声检验了方法的抗噪声性能。结果表明该方法在强噪声环境下仍能可靠地辨识时间序列的确定性。该方法还同时适用于高维混沌与随机噪声的判别。

参考文献:

[1] Grassberger P, Procaccia I. Characterization of strange attractors[J]. *Physical Review Letters*, 1985, 50(5): 346 - 349.
 [2] Cohen A, Procaccia I. Computing the Kolmogorov entropy from time signals of dissipative conservative dynamical system[J].

- Physical Review A*, 1985, 31(3): 1872 - 3890.
- [3] Sano S, Sawada Y. Measurement of the Lyapunov spectrum from a chaotic time series[J]. *Physical Review Letters*, 1985, 55(10): 1082 - 1085.
- [4] 雷敏, 孟光, 冯正进. 连续动力系统时间序列的非线性检验[J]. 物理学报, 2005, 54(3): 1059 - 1063. (Lei M, Meng G, Feng Z J. Detecting the nonlinearity for time series sampled from continuous dynamic systems[J]. *Chinese Physics*, 2005, 54(3): 1059 - 1063.)
- [5] 高思莉, 邱红梅, 刘春霞, 等. 噪声对离散混沌系统非线性特征量的影响[J]. 东北师大学报(自然科学版), 2003, 35(1): 35 - 43. (Gao S L, Qiu H M, Liu C X, et al. Effects of noise to the nonlinear values of chaotic discrete system[J]. *Journal of North-east Normal University (Natural Science Edition)*, 2003, 35(1): 35 - 43.)
- [6] Bhattacharya J, Kanjilal P P. On the detection of determinism in a time series[J]. *Physica D*, 1999, 132(1-2): 100 - 110.
- [7] Palus M, Dvorak I. Singular-value decomposition in attractor reconstruction: Pitfalls and precautions[J]. *Physica D*, 1992, 55(1-2): 221 - 234.
- [8] 卢山, 王海燕. 证券市场的非线性和确定性检验[J]. 系统工程理论方法应用, 2005, 14(3): 235 - 238. (Lu S, Wang H Y. Testing for nonlinearity and determinism in stock market[J]. *Systems Engineering Theory Methodology Applications*, 2005, 14(3): 235 - 238.)
- [9] 吴延东, 谢洪波. 一种新的时间序列确定性辨识方法[J]. 物理学报, 2007, 56(11): 6295 - 6311. (Wu Y D, Xie H B. A new method to recognize determinism in time series[J]. *Chinese Physics*, 2007, 56(11): 6295 - 6311.)
- [10] Xie H B, Wang Z Z, Huang H. Identification determinism in time series based on symplectic geometry spectra[J]. *Physics Letters A*, 2005, 342(1-2): 156 - 161.
- [11] Xie H B, Guo J Y, Zheng Y P. Using the modified sample entropy to detect determinism[J]. *Physics Letters A*, 2010, 374(38): 3926 - 3931.
- [12] Gautama T, Mandic D P, Van Hulle M M. The delay vector variance method for detecting determinism and nonlinearity in time series[J]. *Physica D*, 2004, 190(3-4): 167 - 189.
- [13] Takens F. Detecting strange attractors in fluid turbulence[J]. *Lecture Notes in Mathematics*, 1981, 898(1): 366 - 381.
- [14] Min W L, Lu L, He X F. Locality pursuit embedding[J]. *Pattern Recognition*, 2004, 37(4): 781 - 788.
- [15] Jeong J, Kim M S, Kim S Y. Test for low-dimensional determinism in electroencephalograms [J]. *Physical Review E*, 1999, 60(1): 831 - 837.