arXiv:2303.11868v1 [nlin.CD] 21 Mar 2023

符号时间序列的不可逆性：一个警示故事

Llu´ís Arola-Fern´andez* 和 Lucas Lacasa†
Instituto de Física Interdisciplinar y Sistemas Complejos IFISC (CSIC-UIB), Campus UIB, 07122 Palma de Mallorca, Spain（日期：2023年3月22日）

许多经验时间序列本质上是符号序列：例子包括网络科学中的链路激活模式、DNA 编码或神经科学中的放电模式，以及密码学或词组合学中的实例。在其他一些情境中，底层时间序列实际上是实值的，之后才进行符号化，例如混沌系统的符号动力学。在多种时间序列量化指标中，时间序列不可逆性——平稳时间序列中正向统计与反向统计之间的差异——具有重要意义。然而，符号化时间序列的不可逆特性并不总是等同于底层实值信号的不可逆特性，这会导致对可解释性的误解和困惑。对于二元时间序列，这种困惑尤为突出；二元时间序列是通过符号动力学对混沌轨迹进行编码的一种经典方式。本文旨在澄清一些常见误解，并为符号时间序列中时间不可逆性的实际分析及其解释提供理论基础。我们概述了平稳符号序列中由非回文对的频率不对称性导致的不可逆性来源，并对这些非回文对进行枚举；同时证明，基于长度为 $m < 4$ 的词，二元时间序列不会呈现任何不可逆性，并据此讨论其影响及困惑的来源。我们还研究了符号动力学中的不可逆性，并阐明为何即使底层动力系统不可逆，符号序列仍可能可逆，例如完全混沌的逻辑斯蒂映射这一情形。

I. 引言

不可逆性作为非平衡热力学的近义概念，其历史可追溯至 Prigogine 的开创性著作 [1]。简言之，统计时间不可逆性，或称时间不可逆性（TI），是时间序列（及其背后的生成过程）的一种性质：时间序列的统计特性与其时间反演后的统计特性存在显著差异（参见 [2] 及其中的参考文献）。非平稳信号[1] 或不对称周期信号中的时间不可逆性很容易观察到；然而，要判断一个（含噪）平稳信号是否可逆（在统计意义上），通常并非易事，图（1）对此作了示意。

时间不可逆性确实是一项重要指标，也是许多领域持续研究的重点。在非平衡热力学中，随机热力学研究单条轨迹[4], 的不可逆特征、第二定律表观违背的出现，以及通过涨落定理对其进行量化 [5]。在此背景下，时间不可逆性是一种表征熵的时间之矢存在的量，通常出现在偏离平衡态的过程中，并可与熵产生的度量相关联 [6]。

秉持类似的理念，时间反演也在（确定性）动力系统领域得到探讨，并与混沌系统中的相空间收缩和耗散等概念相关 [7]。与此同时，应用统计学领域也出现了关于时间反演的开创性理论研究，计量经济学和时间序列分析领域也有相关研究，其中包括关于线性随机过程可逆性的经典结果，以及“可逆性蕴含平稳性”这一格言可逆性蕴含平稳性[8]。更为独立的是，在词组合学领域 [9] ，时间可逆性的概念与抽象符号序列中回文“词”（符号块）的计数密切相关 [10]。

不同学科彼此缺乏交流，却围绕看似相同、定义略有不同的概念开展研究，这造成了一些混淆。例如，布朗运动是可逆的吗？这个问题的答案并不简单。有人可能会认为，处于热平衡状态的布朗粒子不会耗散，因此是可逆的；然而，支配布朗运动的方程（扩散方程）在时间反演下并不保持不变。与此同时，虽然布朗粒子的轨迹平均而言并不“产生熵”，应归类为可逆过程，但布朗运动本身是非平稳过程，因此根据[8] （不过，参见 [3]）它是不可逆的。问题的一部分可能在于，人们混淆了单条轨迹的可逆性与完整的（系综）过程。尽管如此，在时间反演这一多面概念尚待统一之际，时间反演的不同形式已广泛应用于金融经济学、神经科学和生物物理学等迥异领域，读者可参阅一篇有趣的综述文章 [2] 及其中的参考文献。

在实践中，生成时间序列的动力学方程（即哈密顿动力学、混沌映射或随机微分方程）鲜少可得，因此人们通常直接根据时间序列测量结果探究时间反演。尽管人们已从多个角度将实值时间序列（连续时间或离散时间）的时间反演测量操作化 [2], ，但本文旨在

* lluisarolaf@gmail.com
† lucas@ifisc.uib-csic.es

[1]不过， [3] 讨论了如何提取非平稳信号中有意义的时间反演度量。

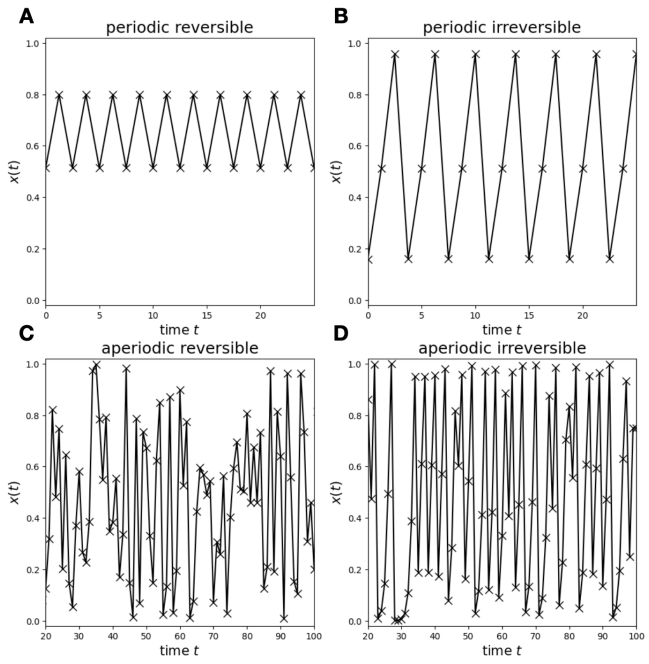

图 1. 可逆（左）和不可逆（右）信号的短时间窗口。（A面板）在 $r = 3.2$ 处由逻辑斯蒂映射 $x_{t+1} = rx_t(1 - x_t)$ 生成的周期性可逆轨迹。信号在两个值之间对称振荡。（B面板）逻辑斯蒂映射在 $r = 3.8284$ 处的周期性不可逆轨迹，对应于该映射的一个稳定岛（被混沌包围）；在此处，信号以非对称方式在三个值之间振荡，因此可以轻易检测出时间方向性。（C面板）由均匀独立同分布噪声生成的非周期性可逆轨迹。这是一个噪声过程的示例，在长序列极限下具有时间可逆性。（D面板）在 $r = 4$ 处由逻辑斯蒂映射生成的非周期性不可逆轨迹。该过程是混沌且耗散的，因此不可逆。为便于可视化，图中仅展示了短时间窗口内的轨迹，但要准确估计噪声信号的时间不可逆性，通常需要长得多的时间序列。

聚焦于这样一种情况：这些测量结果以完全离散的时间序列形式呈现，也就是说，数据序列中的时间和变量本身都只取离散值。如果这些值不存在自然的几何嵌入方式——例如，这些值属于类别型数据——那么我们处理的就是符号时间序列；时间序列中可能出现的值的集合称为字母表。符号时间序列确实广泛存在于各个学科中，直接根据符号时间序列测量时间反演，也常用于经济学 [11, 12], 、生理学 [13和遗传学 [16] 等领域。

撰写本文的主要动机，源于我们作为时间反演领域研究者和实践者的亲身经历。在处理这类符号序列的时间反演问题时，我们遇到了若干误解和困惑。例如，假设存在一个（潜在的）时间不可逆过程，它生成时间反演时间序列，而这些序列随后被符号化为由离散值或符号组成的序列。那么，后者的符号序列是否总是时间反演的？遗憾的是，情况并非总是如此，因此需要谨慎对待（相关讨论也可参见这篇有趣的论文 [17] ）。我们的目标是对一些常见误解作出温和的指引，并提供理论基础和一些新结果，为该领域的研究者和实践者提供指导。

本文其余部分安排如下：第二节介绍符号，给出基本定义，并简要讨论一些量化时间反演的标准方法。第三节阐明离散符号序列中不可逆性的两个潜在来源（我们称之为情景 1 和情景 2），并陈述、证明和讨论一条关于三点统计量的不可能性定理。该定理表明，仅通过考察一元、二元或三点统计量，我们无法判断二元时间序列是否存在时间反演对称性。第四节将超越三点统计量，详细讨论由确定性时间序列经符号化生成的二元时间序列中，情景 1 和情景 2 何时会出现、何时不会出现。这一分类表明，例如，典型的不可逆过程——如完全混沌的逻辑斯蒂映射——在使用（标准）生成划分 [18], 进行符号化时是时间可逆的；而采用非常规方式进行符号化时，则会显露其底层动力学真正的时间反演特征。我们还在此将情景 1 和情景 2 与拓扑熵、柯尔莫哥洛夫-辛奈熵等经典度量联系起来 [18]。我们的分析将这些度量与符号动力学中禁词的出现及其与词组合学领域的关系联系起来 [9]。第五节总结全文，并提出一些值得记取的要点。

II. 前言：符号、定义与方法综述

本文始终以一个（较长的）有序符号序列 $\mathcal{S} = (x_1, x_2, x_2, \ldots, x_n)$ 为研究对象，其中 x_i 是从含有 $|\mathcal{A}|$ 个符号的字母表 $\mathcal{A}$ 中提取的符号（对于二元时间序列这一特例，我们有 $\mathcal{A} = \{0, 1\}$、$|\mathcal{A}| = 2$）。我们还将长度为 m 的词 w 定义为 $m \geq 1$ 个符号的串接，例如，0111 是一个长度为 $m = 4$ 的词。

由于我们关注的是时间序列，不失一般性，假设 $\mathcal{S}$ 是一个时间序列，因此我们可以互换地说，x_t 是 $\mathcal{S}$ 中的第 t个条目，或是时间 t 处的条目。如果长度为 1 的词的数量随时间大致保持不变，则称 $\mathcal{S}$ 为平稳。我们将 $\mathcal{S}$ 定义为“沿时间正向”的序列，即 $\mathcal{S} \equiv \mathcal{S}_{\text{fwd}}$，其时间反转序列为$\mathcal{S}_{\text{bwd}} = (x_n, x_{n-1}, \ldots, x_2, x_1)$。当 n很大时，

[2] 需要为进行这项检验的时间窗口长度设定一个足够大的下限

归一化频率层级

$$p(x_1, x_2, \ldots, x_m) = \frac{\#\text{ of occurrences of the word } x_1 x_2 \ldots x_m}{n-m+1},$$

其中，分母考虑了序列中重叠词的数量，这在形式上定义了一个随机过程。一般而言，如果 $\mathcal{S}_{\text{fwd}}$ 和 $\mathcal{S}_{\text{bwd}}$ 的统计有所不同，则称平稳符号时间序列$\mathcal{S}$ 是时间不可逆的——或简称不可逆。我们可以进一步将其形式化如下

定义(m阶不可逆性）。 $\mathcal{S}$ 是 m 阶不可逆的（或简称为 m-不可逆），如果

$$p_{fwd}(w_m) \neq p_{bwd}(w_m), \tag{1}$$

其中， w_m 分别是正向和反向时间序列中所有长度为 m 的词的集合。

因此，如果存在有限的 $m < n$ 使得 $\mathcal{S}$ 是 m-不可逆的，我们就称 $\mathcal{S}$ 是不可逆的（注意， m-不可逆性意味着 $m+1$-不可逆性）。使 $\mathcal{S}$ 成为 m-不可逆的最低 m 值称为 $\mathcal{S}$的内在不可逆阶数。

为了将上述定义付诸实践，我们需要一种量化 m-不可逆性的方法。一种概念上直接的评估方法是使用分布之间的相似性度量 $\mathcal{D}$，从而可以计算 $\mathcal{D}(p_{\text{fwd}}(w_m)\ p_{\text{bwd}}(w_m))$。这种统计量需要满足无偏性：对于 m-可逆序列 $\mathcal{S}$，当n足够大时，应有 $\mathcal{D}(p_{\text{fwd}}(w_m)\ p_{\text{bwd}}(w_m)) \to 0$ 。对于有限的 n，要评估 $\mathcal{D}(p_{\text{fwd}}(w_m)\ p_{\text{bwd}}(w_m))$的显著性，一种标准方法是通过打乱[3] $\mathcal{S}$ 多次，获得与该统计量对应的 p值；在每次打乱所得的样本中计算相同的统计量，并检查测得的统计量大于或等于$\mathcal{S}$所得统计量的样本百分比。如果该 p值低于0.05，则可以有95%的把握认为 $\mathcal{S}$ 是 m-不可逆的，而m-不可逆性的程度由 $\mathcal{D}(p_{\text{fwd}}(w_m)\ p_{\text{bwd}}(w_m))$给出。

传统上，作者们会使用简单的度量来衡量$\mathcal{D}$ ，例如 L_2 距离 [19]。有趣的是，如果改用库尔贝克-莱布勒散度 [20], ，那么结果表明， $\mathcal{D}(p_{\text{fwd}}(w_m)\ p_{\text{bwd}}(w_m))$在 m 趋于无穷大的极限下，会收敛到一个熵产生度量 [4, 6, 21, 22]。此外，还有一种不同的方法，无需计算 m 不可逆性，而是直接尝试估计完整统计中的差异。这种方法基于比较 $\mathcal{S}_{\text{fwd}}$ 与$\mathcal{S}_{\text{bwd}}$ [6, 23]的压缩特性。其他用于估计时间反演（TI）的方法，在时间序列不一定是符号序列时，还包括多尺度不对称性 [24], 、可见性图 [25, 26], 或序数模式 [27] 等（详细综述请参见 [2] ）。

[3] 打乱序列会保留其边缘分布，并产生一个（渐近地）在统计意义上可逆的不相关随机序列

III. 符号序列中 m-不可逆性的来源

我们希望找出任意符号时间序列中可能导致不可逆性的词。考虑一个较长的符号序列 $\mathcal{S}$，其字母表大小为 $|\mathcal{A}|$，并令 w_m 表示长度为 m的词集合，原则上该集合有 $|\mathcal{A}|^m$ 个元素。

对于给定的 m，所有词都可以分为回文词或非回文词。回文词是在时间反演下保持不变的词，例如，101 或 01210 是回文词，而 001 或 012 是非回文词。非回文词的总数 NPW(m, $|\mathcal{A}|$) 可以通过从全部词的集合中减去回文词的数量来计算。由于回文词的数量等于词长一半所能构成的不同词的数量（如果 m 为偶数），或者等于词长一半减一所能构成的不同词的数量乘以$|\mathcal{A}|$（如果 m 为奇数），因此我们得到

$$\text{NPW}(m, |\mathcal{A}|) = \begin{cases} |\mathcal{A}|^m - |\mathcal{A}|^{(m-1)/2+1} & \text{如果 } m \text{ odd} \\ |\mathcal{A}|^m - |\mathcal{A}|^{m/2} & \text{如果 } m \text{ even} \end{cases} \tag{2}$$

例如，取 $|\mathcal{A}| = 2$ 和 $m = 3$。回文词有 000、111、010 和 101，非回文词有 001、100、011 和 110。有趣的是，所有非回文词都是成对出现的：一个词及其时间反演形式，例如前例中的 (001, 100) 和 (011, 110)。因此，在 $\mathcal{S}_{\text{fwd}}$ 中观察到某个特定的非回文词，等同于在$\mathcal{S}_{\text{bwd}}$中观察到该非回文对中的另一个词。根据定义，回文词在时间反演下保持不变，因此不可逆性的任何来源必然只会出现在非回文词中；具体而言，就是比较每个非回文对中两个词的频率。

此外，这些非回文词产生不可逆性来源的方式本质上有两种，即两种情形：

- 情景 1（S1）：给定一个非回文对，其中一个词出现在 $\mathcal{S}$ 中，而另一个词缺失（即所谓的禁用词）。

- 情景 2（S2）：给定一个非回文对，其中两个词都出现在 $\mathcal{S}$ 中，但频率不同。

实际上，S2 包含 S1 这一特例，但为方便起见，我们仍将这两种情形分开讨论。此外，S1 或 S2 的提出早于 m-不可逆性的实际估计；也就是说，实际估计 TI 有多种方法，但它们都以 S1 或 S2 的存在为前提。因此，本文其余部分将特别关注这两种情形。

A. 关于时间反演的不可行定理：在二元时间序列中，m-不可逆性在 $m < 4$ 时不可能

对于实值时间序列，仅通过考察 2-不可逆性[2]，就能发现不可逆性的来源。然而，我们将在下文表明，对于二元时间序列，情况并非如此，我们至少需要考察长度为 $m \geq 4$ 的词。在陈述并证明定理之前，我们需要先给出一些定义和一个有用的引理。

定义（平凡序列）。如果一个二元序列 $\mathcal{S}$最终变为常数，即存在一个 $q > 0$使得该序列 $\mathcal{S} = x_1 x_2 \ldots x_q 0_$ 或 $\mathcal{S} = x_1 x_2 \ldots x_q 1_$，其中 $_s = sss \ldots$，那么它就是平凡的。反之，不平凡的序列称为非平凡序列。

定义（平凡非回文对）。如果一个由长度为 m 的词组成的二元非回文对中，其中一个词是由单个符号 $\{0\}$ 与（$m-1$）个符号 $\{1\}$ 连接而成，或者是由符号 $\{1\}$ 与（$m-1$）个符号 $\{0\}$连接而成，那么这个二元非回文对就是平凡的。反之，不平凡的非回文对称为非平凡非回文对。

例如，(01，10)、(001，100) 和 (1000，0001) 分别是长度为 $m = 2$、$m = 3$ 和 $m = 4$ 的平凡非回文对。因此，不符合上述条件的非回文对是非平凡的，例如，当 $m = 4$ 时的 (0100，0010) 或 (1010，0101)，以及当 $m = 5$ 时的 (01001，10010)，这里只举几个例子。

下面的引理也有助于证明这一不可能性定理，因为它排除了那些不可能成为二元时间序列不可逆性来源的词（此外，按照定义，回文词是平凡的可逆源）。

引理（非回文可逆源）。设 $\mathcal{S}$ 是一个长度为 n的大型非平凡二元时间序列，且z_m 是所有长度为$m < n$的平凡非回文词组成的集合。那么

$$p_{\mathrm{fwd}}(z_m) = p_{\mathrm{bwd}}(z_m),\ \forall\ m$$

证明 – 首先注意，平凡非回文词总可以写成 $a_1 a_2 \ldots a_{m-1} b_m$或$a_1 b_2 b_3 \ldots b_m$的形式，其中 a_i是 $\{0\}$，而 b_i是 $\{1\}$（反之亦然），对任意 i 均成立（注意，此处的标记用于表示符号在词中的位置）。

让我们聚焦于第一种情况，并假设$a_1 a_2 \ldots a_{m-1} b_m$出现在$\mathcal{S}$中。若要它再次出现，我们需要从符号 b 切换到“符号” $a_1 a_2 \ldots a_{m-1}$。这种切换可以立即发生，例如 $a_1 \ldots a_{m-1} b_m a_{m+1} \ldots a_{2m-1} b_{2m}$，此时反向序列 $b_m a_{m+1} \ldots a_{2m-1} = b_1 a_2 \ldots a_m$已出现在序列中；也可以在间隔任意长度 k后发生，例如 $a_1 \ldots a_{m-1} b_m \overbrace{x_{m+1}} \ldots \overbrace{x_{m+k}}\, a_{m+k+1} \ldots a_{2m+k-1} b_{2m+k}$，这样反向序列便会出现在最后一个 k' 处，此时$x_{m+k'} = b$。无论如何，新的 $a_1 a_2 \ldots a_{m-1} b_m$要出现，就必然需要先出现一次 $b_1 a_2 a_3 \ldots a_m$，因此在序列长度趋于无穷时，这一对既不对 S1 有贡献，也不对 S2 有贡献。

对于 $a_1 b_2 b_3 \ldots b_m$也会出现类似的现象：假设我们在 $\mathcal{S}$ 中发现它，经过一段间隔后又再次发现它。这段间隔可以是（i）空的，从而得到 $a_1 b_2 b_3 \ldots b_m a_{m+1} b_{m+2} b_{m+3} \ldots b_{2m}$，即在中间找到反向序列；（ii）由 $a_1 x_2 x_3 \ldots x_k$构成，即 $a_1 b_2 \ldots b_m a_{m+1} x_{m+2} \ldots x_{m+k} a_{m+k+1} b_{m+k+2} \ldots b_{2m+k}$，这同样会导致 $b_1 b_2 \ldots b_{m-1} a_m$紧接在第一个 $a_1 b_2 b_3 \ldots b_m$之后出现，无论 $x_2 \ldots x_{k-1}$为何；或者（iii）由 $b_1 x_2 x_3 \ldots x_k$构成。在后一种情况下，只要 $x_i = a$中任意一个取值，反向序列 $b_1 b_2 \ldots b_{m-1} a_m$就必然会立即出现。如果它们都不取该值（即 $x_i = b\ \forall\ i$），那么在 b出现（$m-1$）次之后，$b_1 b_2 \ldots b_{m-1} a_m$就会出现；若 $k \geq$（$m-1$），则这些 b都出现在间隔中；若$k <$（$m-1$），则 $b_1 b_2 \ldots b_{m-1} a_m$来自初始词中的$(m-1-k)$个符号 b 与间隔中的 k 个符号 b 的拼接。

总之，在任意一个平凡非回文对中，这两个词的频率相同，因此$p_{\mathrm{fwd}}(z_m) = p_{\mathrm{bwd}}(z_m)$，证明至此完成。 □

现在我们可以陈述并证明二元时间序列的不可能性定理，该定理将不可逆性的检测限制在长度为 $m \geq 4$的词上。 $m-$

定理。设 $\mathcal{S}$ 为一个较大的非平凡二元时间序列，w_m 为长度为 m的词的集合。那么

$$p_{\mathrm{fwd}}(w_m) = p_{\mathrm{bwd}}(w_m),\ \text{for } m = 1, 2, 3.$$

证明——该证明是构造性的。

– $m = 1$: 时间反演不改变 $\mathcal{S}$ 的边缘分布，因此显然有$p_{\mathrm{fwd}}(w_1) = p_{\mathrm{bwd}}(w_1)$。

– $m = 2$: 在这种情况下，只有一对非回文对：(01,10)，且它是平凡的。利用引理，我们得到 $p_{\mathrm{fwd}}(01) = p_{\mathrm{fwd}}(10) = p_{\mathrm{bwd}}(01)$。于是有$p_{\mathrm{fwd}}(w_2) = p_{\mathrm{bwd}}(w_2)$。

– $m = 3$: 共有 NPW(3) 个非回文词，因此有两对非回文对：(001，100) 和 (011，110)，二者都是平凡的。再次利用引理可知，对于 $m = 3$，每个非回文对中的词出现频率相同，因此 $p_{\mathrm{fwd}}(w_3) = p_{\mathrm{bwd}}(w_3)$，证明得证。 □

[4]有趣的是，在评估时间反演时，若不以词频为依据，而是考察这些间歇块的复发时间，那么它们的大小分布实际上可能会产生可测量的影响 [28],不过，如果间歇块大小遵循复杂的非马尔可夫过程，请注意，这可能会首先破坏时间序列的平稳性。

该定理表明，对于二元时间序列，不能仅凭二阶和三阶不可逆性来判断其是否存在不可逆性；还需要考虑长度为 $m \geq 4$ 的词。为理解这一现象，请注意，对于 $m = 4$，有 12 个非回文词，但平凡非回文词仍只有 4 个：0001、1000、0111 和 1110（即两对）。因此，非平凡非回文词可能是不可逆性的来源，而且只有在 $m \geq 4$ 时才会出现。例如，考虑短序列 0010110010。词 0010 是非回文且非平凡的，它出现了两次，而其反转形式 0100 并未出现。相反，要使词 001 再次出现，我们会发现它的反转形式 100 必须在序列中出现一次。

因此，为了正确估计二元时间序列中的潜在不可逆性来源（$|\mathcal{A}| = 2$），我们应从非回文词集合中减去两对平凡非回文词（对于$m \geq 2$）。随后便可轻松计算二元时间序列中潜在不可逆性来源词所占的比例，$q(m, |\mathcal{A}| = 2)$，其中 $q(1, 2) = q(2, 2) = 0$ 且

$$q(m,2) = \begin{cases} 1 - 2^{(1-m)/2} - 2^{(2-m)} & \text{if } m \geq 3 \text{ } odd \\ 1 - 2^{-m/2} - 2^{(2-m)} & \text{if } m \geq 2 \text{ } even \end{cases} \quad (3)$$

由于对于 $m \geq 3$，平凡非回文词的数量始终为 4，因此对于较大的 m，这一修正可以忽略不计；但这正是无法检测 $m = 3$的m-不可逆性的原因。

B. 超越二元情形的符号时间序列

请注意，对于具有 $|\mathcal{A}| > 2$的符号时间序列，所有非回文词都可能对不可逆性有所贡献，因此不可能性定理并不适用。例如，只需有 $|\mathcal{A}| = 3$ 个符号，就可能出现 2-不可逆性；例如，在 01220120122220120 中，非回文词对（01，10）出现的频率不同（实际上 10 并未出现，因此属于情景 1）。在这种情况下，我们可以直接使用公式（2），并以 $|\mathcal{A}|$归一化，来估计潜在不可逆源所占的比例，从而得到

$$q(m, |\mathcal{A}| > 2) = \begin{cases} 1 - |\mathcal{A}|^{(1-m)/2} & \text{if } m \text{ } odd \\ 1 - |\mathcal{A}|^{-m/2} & \text{if } m \text{ } even \end{cases} \quad (4)$$

在图（2）中，我们绘出了不同字母表大小下，不可逆性来源所占比例随词长 m 的变化情况；该比例由式（3）和式（4）直接计算得出。我们观察到，对于 $|\mathcal{A}| = 2$，当 $m \geq 4$时，该比例为 $q > 0$；而对于$|\mathcal{A}| > 2$，即使 $m = 2$，我们也能检测到 m-不可逆性。

回顾起来，缺少不可能性定理或许（无意中）正是针对符号化时间序列提出的时间反演量化方法的核心问题，例如在 [13]中，作者直接提出了一种三元编码（$|\mathcal{A}| = 3$）；在 [28]中，作者则考虑了一个三状态

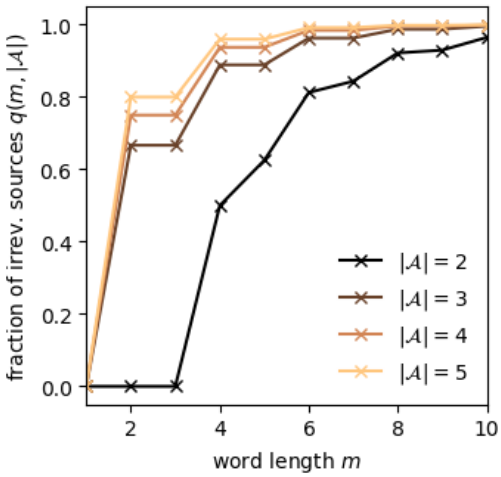

图 2. 不同字母表大小下，可作为不可逆性来源的词所占比例随词长的变化。请注意，代表二元字母表的黑色曲线在$m = 4$处从 $q = 0$ 跃迁至正值，这正如不可能性定理所预言的那样；而对于更大的字母表，任意 $m > 1$下均有 $q > 0$。

马尔可夫链并研究回返时间；又如在 [19],中，作者（未经进一步讨论）表示：“我们认为使用更大的字母表规模是可取的 [...], ，与通常使用的二元字母表形成对比 [...]”。顺带一提，在 [19], 的图（3）B面板中，作者似乎发现了 $|\mathcal{A}| = 2$的2-不可逆性和三阶不可逆性。这违反了上文证明的定理，因此我们推测该图某处可能有误（实际上，在 [19] 的图（3）A面板中，作者正确地发现 $|\mathcal{A}| = 2$ 不存在三阶不可逆性，这与B面板相矛盾）。

C. 补充说明

请注意，在实际中，当序列长度 n 不够大时，较大词长 m 对应的词频统计会很不充分。这些有限尺寸效应使不可能性定理在实际应用中更加棘手，因为要准确估计 m-不可逆性，所需的序列长度会随m呈指数增长。因此，在（二元）时间序列较短时，时间反演很难估计，而这通常正是实验观测中的情况。如果最终能够获得过程的额外信息（例如，过程以连续时间运行，且有关于状态转换之间耗时的元数据），那么已有研究提出了时间反演的替代量化方法 [29]。遗憾的是，许多经验时间序列纯粹是符号序列，并非某种具有潜在时间统计的电报过程的产物。有趣的是，此时或许可以改用回返时间统计 [28]。

例如，可以评估两个 001 事件之间符号的数量

并将这些统计数据与 100 的复发时间进行比较。当对间隔长度施加复杂的非马尔可夫过程时，实际得到的回返时间统计可能会有所不同，例如对于一个序列

$$\mathcal{S} = \mathbf{001}\underbrace{10101010}_{\tau_1}\mathbf{100}\underbrace{000\ldots00}_{\tau_2}\mathbf{001}\underbrace{10101101\ldots}_{\tau_3}$$
$$\mathbf{100}\underbrace{000\ldots00}_{\tau_4}\mathbf{001}\underbrace{11101101\ldots}_{\tau_5}\mathbf{100}\ldots, \qquad (5)$$

那么，

$$001 \quad \tau_1+\tau_2+3\ \tau_3+\tau_4+\text{的复}$$
$$3\ \tau_5+\tau_6+3\ \ldots\ =\ \tau_i+\tau_{i+1}+3\ \underset{i=1}{\ldots}$$

发时间为（,,,）（），而 100 的复发时间则为（$\tau_2+\tau_3+3$, $\tau_4+\tau_5+3, \tau_6+\tau_7+3, \ldots$）$= (\tau_i+\tau_{i+1}+3)\underset{i=2}{\ldots}$。因此，对于平稳间隔，统计上回返时间相同——此时可以省略索引 i——而对于随时间变化（即非平稳）的间隔，回返时间则不同：因此，发现非平凡模式可能会从一开始就损害信号的平稳性。

无论如何，不可能性定理阐明了为何在二元时间序列中量化不可逆性比在实值时间序列中更困难。对时间序列进行符号化时，实践者通常偏向于二元时间序列，因为它们更简单，而且只需较短的时间序列便能获得准确的词统计，因此可能会无意间落入不可能性定理的适用范围。将二元序列扩展为字母表更大的符号序列，确实可以让我们绕开不可能性定理。

此外，我们认为，研究人员也往往过度使用二元时间序列，部分原因在于信息论或符号动力学理论的声名（见下文）。例如，符号动力学优雅地表明，许多混沌性质可以从经过适当二值化的混沌轨迹时间序列中提取出来。不幸的是，时间不可逆性并不总是其中之一，因为正如我们接下来将讨论的那样，符号化相空间上的运动可能呈现严格随机的状态 [30] ——因而是时间可逆的——。

IV. 符号动力学中的不可逆性及其引发的不可逆性

二元时间序列通常见于数学领域的符号动力学 [18], ，其中区间映射 $x_{t+1} = f(x_t;\ r)$ $x \in [a,b]$ 通过定义区间 $[a,b] = \bigcup_{i=1}^{k}[a_i,b_i)$的划分来符号化。 $k = 2$ 个符号的情形受到了广泛关注，这主要是因为单峰映射——混沌运动的典型低维实例——通常具有所谓的生成划分 [18], 且包含两个符号：例如，混沌逻辑斯蒂映射$x_{t+1} = rx_t(1 - x_t)$ $x \in [0,1]$便是如此。它的二元划分 $[0,1] = [0,1/2) \cup [1/2,1]$ 会产生符号动力学——二元时间序列——从中可以获取重要性质 [31], 。例如，由初始条件 x 出发，经由 f 迭代后生成的时间序列，再经过符号化，便能唯一确定 x。

混沌时间序列与不可逆性的研究密切相关，因为混沌耗散映射往往会产生不可逆动力学；这一性质很容易从实值时间序列的统计特征中提取出来 [2]。但是，在符号动力学所提供的粗粒化描述中，不可逆性是否仍然保留？为了讨论这个问题，我们来考察两个重要的经典度量。所谓的拓扑熵$K(0)$统计长度为 m且实际出现在 $\mathcal{S}$中的可能词的数量，使得

$$K(0) = \lim_{m\to\infty}\frac{1}{m}\ln(|w_m|). \qquad (6)$$

（严格来说，只有当相空间划分是生成划分时， $K(0)$才确实收敛于拓扑熵[18, 32]。）不难证明，对于 $|\mathcal{A}| = 2$个符号，S1的出现直接意味着 $K(0) < \ln(2)$。

符号动力学理论的另一项重要结论是，当 f 是一个混沌映射时，可以利用所谓的块熵率估算其（正的）李雅普诺夫指数 λ

$$h_m = \frac{-1}{m}\sum_{w_m} p(w_m)\log p(w_m), \qquad (7)$$

通过所谓的佩辛恒等式 [18], ，该恒等式将 λ 与系统的柯尔莫哥洛夫-辛奈熵 h_{KS} 联系起来，其中

$$h_{KS} = \lim_{m\to\infty} h_m \qquad (8)$$

如果用于符号化轨迹的划分又是所谓的生成划分 （一般来说，需要对所有可能的划分进行适当优化）。请注意，式 (7) 超越了拓扑熵，因为它确实考察了 $\mathcal{S}$中各词的频率（即测度），而不仅仅是它们是否存在。从这个意义上说，对于 $|\mathcal{A}| = 2$ 个符号，很容易看出，情景 2 的出现意味着 $h_{KS} <$ln(2)。此处需要注意， h_{KS} 并非标准熵，因为它不是广延变量；实际上，它是熵率。事实上，式 (8) 也适用于并非必然是确定性轨迹符号化结果的符号序列（例如随机过程，以及一般而言任何类型的经验符号序列）。在这种情况下，式 (8) 收敛于序列 $\mathcal{S}$ 的熵率，或者等价地，收敛于生成 $\mathcal{S}$ [20] 的随机过程的熵率（不要将其与取值范围为 $\mathcal{A}$的离散随机变量的香农熵混淆；后者本质上是 h_1）。

众所周知，具体采用的划分以及映射的参数 r 会影响混沌映射产生的符号序列：它可以（i）不含禁词（其频率原则上可能相等——从而缺乏不可逆性，且 $h_m = \ln|\mathcal{A}|$ 为常数——也可能不同，从而导致情景 2），或（ii）具有一套层级式的禁词（即所谓的语法 [18, 30]），从而导致情景 1。

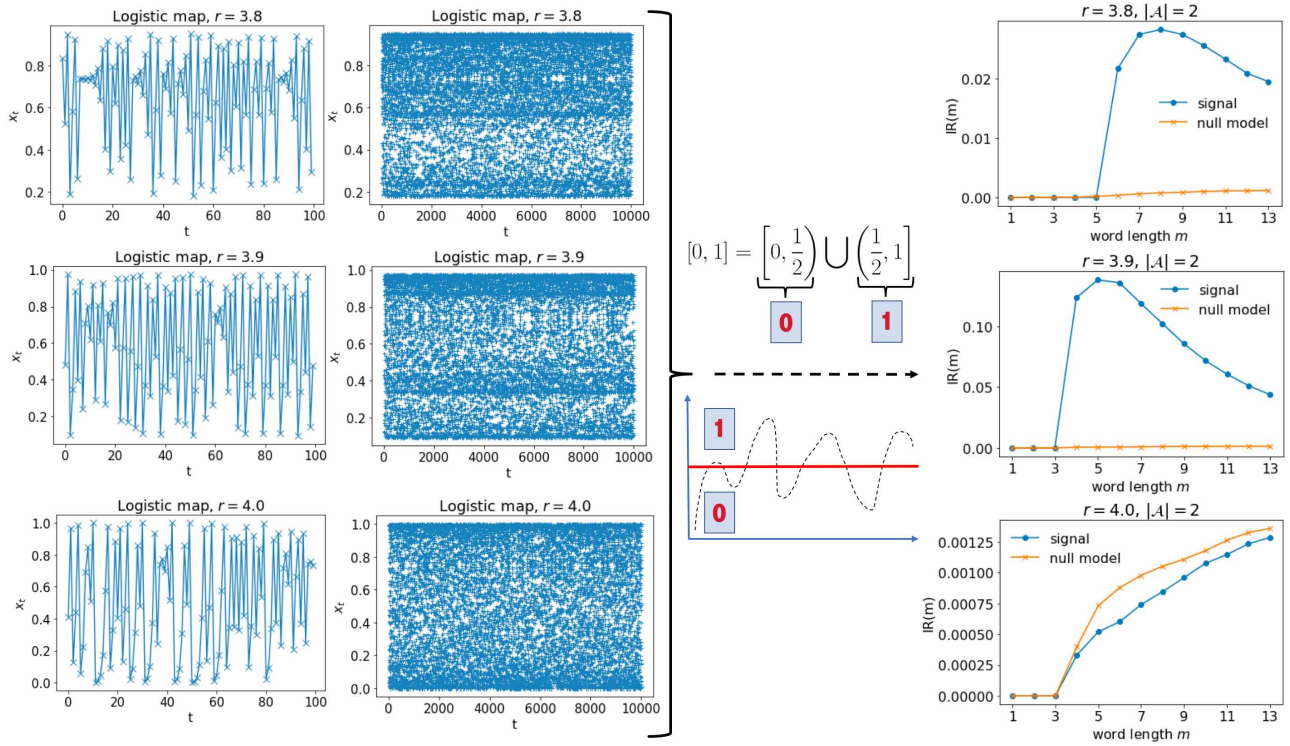

图 3. 展示由逻辑斯蒂映射 $x_{t+1} = rx_t(1-x_t)$ 生成的二元时间序列（含 10^5 个数据点）的 m-不可逆性（使用 L_2 范数计算）。图中给出了映射处于混沌状态时 r 的三个不同取值： $r = 3.8$ （Pomeau-Maneville 情景）、 $r = 3.9$ 和 $r = 4.0$ （完全混沌，且与伯努利移位拓扑共轭）。每种情形下，时间序列均通过二符号划分 $[0,1] = [0,1/2) \cup (1/2,0]$ 进行符号化。此类划分仅在 $r = 4.0$ 时是生成划分——此时诱导出的符号动力学是常规马尔可夫链，即时间可逆——而对于其他取值 $r \neq 4$，则预期会诱导出复杂的非马尔可夫过程，表现为某些非回文对的频率开始出现差异（通过情景 2 导致 m-不可逆性），以及禁用词（语法）开始出现，进而也通过情景 1 导致 m-不可逆性。最后一列同时展示了根据符号化时间序列计算的 m-不可逆性结果，以及将同一计算应用于零模型所得的结果；在零模型中，符号化时间序列经过了打乱（按构造，这种零模型对所有 m 都是时间可逆的，因此任何有限值都只是有限尺寸效应）。在所有情形下，我们都发现 $m < 4$ 不存在不可逆性，这与不可能性定理的预期一致。对于 $r = 4.0$ （最后一列的底部面板）， $m \geq 4$ 的结果与零模型没有差别，因此无法拒绝可逆性，这也符合预期。对于 $r = 3.8$ 和 $r = 3.9$，我们发现了 m-不可逆性的迹象： $r = 3.9$ 从 $m = 4$ 开始， $r = 3.8$则从 $m = 6$ 开始。这表明每种情形都有不同的语法；所诱导的非马尔可夫过程虽然不可逆，但不可逆程度各不相同。

不含禁词的序列。根据拓扑共轭关系，任何与伯努利移位共轭的映射，其符号动力学 [18] 中都不含禁符号。帐篷映射或乌拉姆映射 $x_{t+1} = 1 - 2x_t^2$ 就属于这种情况，它们等价于完全混沌的逻辑斯蒂映射（$x_{t+1} = rx_t(1-x_t)$，其中 $r = 4$）。对于所有这些情况，采用具有生成性的二元划分 $|\mathcal{A}| = 2$ 时，长度为 m 的可能（可容许）词的数量确实为 $|w_m| = 2^m$。因此，拓扑熵为 $K(0) = \ln(2)$：情景 1 不会出现。有趣的是，柯尔莫哥洛夫–辛奈熵也等于 $\ln 2$（即该映射的李雅普诺夫指数）。这进而表明，对于这类混沌映射， $p(w_m) \to 1/2^m$，也就是说，符号（二进制）序列可能是可逆的（即便很容易看出这些映射的实值轨迹是不可逆的，参见例如 [25]）。事实上，对于 $x_{t+1} = 4x_t(1-x_t)$，通常采用的划分不仅具有生成性，而且还是一种马尔可夫划分（它产生 $\mathcal{S}$ a

拓扑马尔可夫链）。不仅如此，这种划分使 $\mathcal{S}$ 成为一个传统的马尔可夫链，它显然是可逆的。这一结果与以下事实形成鲜明对比： $r = 4$ 处的实值映射实际上会生成不可逆轨迹。对于 r 的其他取值，该划分不会生成传统的马尔可夫链。

含禁词的序列。如果该划分不是生成划分，或者映射与伯努利映射不同构，那么上述结果就不成立。事实上，在逻辑斯蒂映射中，当参数取值超出 $r = 4$，并采用通常的划分$[0,1/2) \cup (1/2,\ 1]$（该划分同样是生成划分）时，会得到一个与 $\mathcal{S}$ 的统计特性相关联的随机过程，而该过程一般具有非马尔可夫性 [18]。有趣的是，对于 $r \neq 4$，会出现一整组禁词。这组禁词随 r 的取值而异，称为语法 [18, 30]。该语法中存在长度为 m 的禁词，它们会影响每个非回文对中的单个词（例如，将 0101 作为禁用词，而不将 1010 作为禁用词），这是根据情景 1 找到 -不可逆性的首个保证。

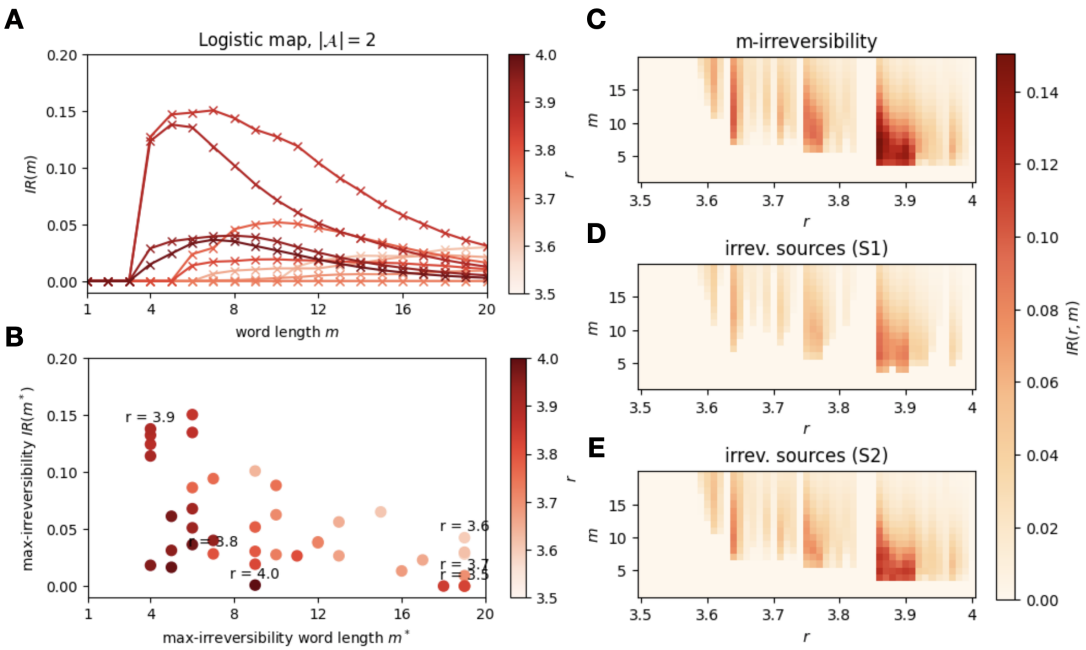

图 4。（A 面板）对于从不同参数值 r 下的逻辑斯蒂映射 $x_{t+1} = rx_t(1 - x_t)$ 中提取的、包含 $3 \cdot 10^6$ 个数据点的二元时间序列，绘制 m-不可逆性 IR(m) 曲线（通过 L_2 范数计算）。（B 面板） m-不可逆性 IR(m^*) 的最大值与出现该最大值时的词长 m^* 之间的散点图。可以看到，不可逆性曲线系统性地呈非单调变化，尽管峰值位置 m^* 会随映射参数 r 的不同取值而以非平凡方式变化；较大的不可逆性值通常对应于峰值出现在 $m^* \in [5, 10]$ 的 m-不可逆性曲线。（C 面板）逻辑斯蒂映射中 m-不可逆性值随 r 和 m 变化的热图。（D 面板）仅针对满足情景 1（S1）的词计算得到的 m-不可逆性热图。（E 面板）仅针对满足情景 2（S2）的词计算得到的 m-不可逆性热图；在方程（9）的求和中，忽略对 S1 有贡献的词。

0101 作为禁用词，但 1010 不作为禁用词）是根据情景 1 找到 m-不可逆性的首个保证。

为了获得一些直观认识，并进一步支持上述论断，我们现在进行一些数值模拟，考察若干情况下的 m-不可逆性（请注意，本文提供了用于检验 m-不可逆性的 Python 代码）。为便于说明，我们在此选择一种简单的方法，根据相应分布 [19] 之间的 L_2 距离来计算 m-不可逆性 IR(m)

$$\mathrm{IR}(m) = \sqrt{\sum_{w \in w_m} [p_{\mathrm{fwd}}(w) - p_{\mathrm{bwd}}(w)]^2}, \qquad (9)$$

不过可以预期，如果使用其他度量来衡量这些分布之间的可区分性，定性结果也会相同（例如库尔贝克-莱布勒散度或 Jensen-Shannon 距离 [6, 33]）。

在图（3）中，我们展示了逻辑斯蒂映射 $x_{t+1} = rx_t(1 - x_t)$ 在若干个 r 取值下的符号化时间序列的 IR(m)（使用生成划分得到 $|\mathcal{A}| = 2$ 个符号）；对于这些取值，我们已知该映射是混沌、耗散且时间不可逆的。我们发现，对于 $r = 4$，所得二元时间序列在所有 m 下确实都是 m-可逆的，正如预期；即使未符号化的时间序列显然不可逆 [25], ，这是因为用于将时间序列符号化的相空间划分，会在符号化序列中诱导出一个常规的马尔可夫链。对于其他 r取值，则会出现语法，并且有明显的 m-不可逆性迹象（根据不可能性定理，从$m > 3$ 开始）。请注意，禁词的语法以及每个非回文对中词的具体频率，都以非平凡的方式取决于 r的具体取值，从而导致时间不可逆性的净值 m差异很大（尽管这些值全都为正且显著，因为它们偏离了零模型的结果）：例如，对于 $r = 3.9$，IR(m) 从$m = 4$ 开始非零，并在 $m = 5$ 处达到峰值，数值约为 0.14；而对于 $r = 3.8$，IR(m) 从较大的 $m = 6$开始非零，峰值位于 $m = 8$，不可逆性的数值则小得多。

图（4）进一步展示了这些发现，其中我们扫描了从 3.5 到 4 的宽广 r 取值范围。该范围涵盖了逻辑斯蒂映射中出现混沌行为的区域。事实上，这里我们只关注混沌区域，因为我们的目标是检测含噪平稳信号中的不可逆性。因此，在图（4）中，我们忽略了在稳定岛上出现的纯周期非对称运动所产生的 m-不可逆性贡献，例如在 $1 + \sqrt{8} \approx 3.83$ [2]附近出现的稳定岛。在这个稳定岛上，不可逆性远大于在其他任何取值处测得的不可逆性

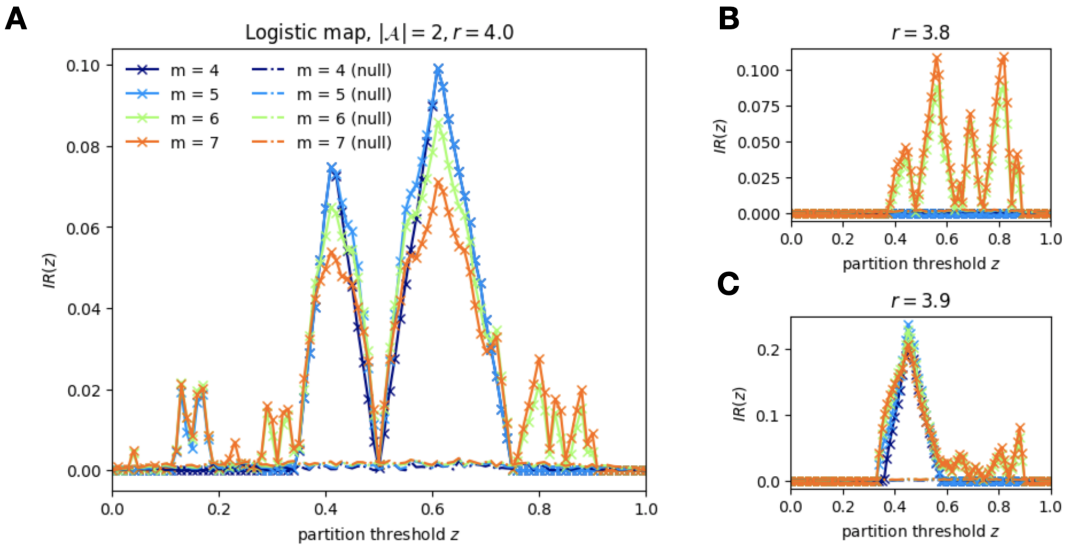

图 5. A 面板。使用方程（9）计算的 m-不可逆性，随逻辑斯蒂映射中二值化过程的划分阈值变化；映射参数为 $r = 4$，词长 m 取若干不同值，时间序列包含 $3 \cdot 10^6$ 个点。虚线表示零模型的不可逆性，其中符号经过随机打乱，因此检测不到不可逆性。B 面板（及 C 面板）展示了 $r = 3.8$ （及 $r = 3.9$）的结果。

r，但这一较大的幅值源自纯周期运动，其动力学表现为在 x的少数几个取值之间进行非对称振荡。周期性不可逆性很容易检测（更多详情亦见图（1）），因此这里我们只关注源自含噪（此处为混沌）平稳信号的不可逆性来源。

现在我们可以进一步详细讨论图（4）。在 A 面板和 B 面板中可以看到，IR(m) 通常呈非单调变化，其不可逆性最大值的位置 m^* 以非平凡的方式取决于 r；而对于那些 r 取值，符号时间序列的不可逆性始终较大，且不可逆性峰值出现在词长为 $m^* \in [5, 10]$时。在 C 面板中，我们用热图展示不可逆性度量如何同时取决于逻辑斯蒂映射中的 r 和 m，其中使用了二元划分 $[0,1] = [0,1/2) \cup [1/2,1]$。可以看到，在映射的混沌起始附近，约为 $r \approx 3.57$ 时，不可逆性从零上升到某个正值，并且在该范围内的大多数 r 取值下都持续存在。除了前述 $r \approx 3.83$附近的稳定岛外，我们还观察到其他不可逆性可以忽略不计的混沌区域，例如广为研究的 $r = 4$；这是因为在对称划分下，符号化动力学具有马尔可夫性质。

在图 C（D）中，我们绘制了方程（9）中对应于不可逆性来源 S1（S2）的各项贡献，以说明语法（即存在禁词）对二值化逻辑斯蒂映射的影响。我们观察到两种不可逆性来源的存在之间存在很强的相关性，这表明，具有禁词的动力学（S1）通常会使正向与反向序列中各词出现频率之间存在很强的不对称性，即使对于非禁用词（S2）也是如此。例如，在 $r \approx 3.9$ 附近，这一效应十分明显：在很低的词长值（$m^* = 4$）下，m-不可逆性较大。在这一区域，确实存在 S1，但 m-不可逆性的大部分来自 S2。总体而言，热图中的异质性凸显了映射的语法对所选 r和 m值的非平凡依赖性，即使我们采用的是对任意 r 值都已知为生成划分的对称划分 $[0,1] = [0,1/2) \cup [1/2,1]$。

在图（5）中，我们探究了逻辑斯蒂映射在非生成划分下的 m-不可逆性表现。我们改变划分阈值 z，使得 $[0,1] = [0,z) \cup [z,1]$ （只有 $z = 0.5$ 会产生生成划分）。在 A面板中，对于若干个 m 的取值，我们观察到当参数设为 $r = 4$ 时，m-不可逆性随 z 变化呈现非单调行为。在 $z = 0.5$ 处，IR(m) 为 $= 0$ 的谷两侧，是两个不可逆性较大的峰。特别是，不可逆性的最大值出现在 $z = 0.6$, 这出人意料地表明，由单位区间的非生成划分产生的二元动力学，恰恰使该映射的 m-不可逆性达到最大。然而，正如面板 B 和 C 所示，若映射的参数发生变化，整条曲线及最大值的位置也会随之改变，这表明过程的不可逆性不仅高度依赖于 r 的取值和词长 m，也高度依赖于所选的划分。

V. 讨论

处理（二元或符号）时间序列时，应谨慎看待关于时间反演的结果。特别是，正如我们所解释的，可能会出现假阴性。我们在此总结几条值得牢记的要点：

i) 如第三节所证明的不可能性定理所示，长度为 $m < 4$ 的词在二元时间序列及其时间反演序列中的统计结果相同。因此，不能仅依据这些统计结果得出时间不可逆性（TI）的结论，还需要考察更长词的统计结果。使用包含两个以上符号的字母表即可消除这一局限，因此在实践中，最好将二元时间序列改为三元（或更高元）时间序列，尽管字母表规模的增加意味着需要更长的时间序列，才能正确估计词统计。这一不可能性定理解释了为何当前用于评估符号时间序列中 TI 的算法方案会以启发式方法避开仅含两个符号的字母表，尽管研究人员往往偏好二元时间序列，这通常与信息论和符号动力学的盛名有关。

ii) 某些时间不可逆过程（例如混沌耗散过程）在经过符号化后可能会被误认为是时间可逆的。完全混沌的逻辑斯蒂映射 $x_{t+1} = 4x_t(1 - x_t)$ 就是一个例子：通过生成划分 $[0, 1] = [0, 1/2) \cup (1/2, \ 1]$（二元符号化）进行符号化后，它会在符号动力学中诱导出一个常规的马尔可夫链。非常规划分（例如非生成的二元划分）会在符号动力学中诱导出复杂的非马尔可夫随机过程，由此可以重新显现时间不可逆性，并对其进行正确估计（不过，仍需基于长度为 $m \geq 4$ 的词）。在缺乏明确的适当划分判据时，最好直接从实值时间序列评估时间反演，而否则，采用包含两个以上符号的划分是一种谨慎的选择。

iii) 符号动力学中时间不可逆性的出现（或缺失）（即针对实值确定性映射，其时间序列使用相空间的适当划分进行符号化）与禁用词（语法）的出现和/或每一非回文对中各词频率不对称性的出现有关。此类非回文对的数量随词长呈指数增长。语法高度依赖于底层动力系统的划分和参数，因此需要针对具体情况进行分析。

iv) 除了词的统计特性外，不可逆性也可能体现在词重复时间的统计特性中。对于连续时间、离散状态的统计，只需量化不可逆性的两个来源：不同非回文词的出现统计，以及等待时间统计 [29]。对于完全离散的时间序列，由于状态转移之间的时间信息不可得，这类信息直接编码在间隔的时长中，因此可以据此提取词之间的复发时间 [28]。然而，要从这些复发时间中找出不可逆性的来源，可能要求符号时间序列是非平稳的，而这首先就带来了概念上的问题。

致谢： 作者感谢 E. Rold´an 和 M. Zanin 提供的宝贵讨论。我们感谢西班牙国家研究署项目 DYNDEEP（EUR2021-122007）提供的资助。LL 另外感谢项目 MISLAND（PID2020-114324GB-C22）以及 Mar´ía de Maeztu 项目 CEX2021-001164-M 的资助。

代码可用性： 发表后，Python 代码m-irreversibility 将在 GitHub 上提供。

[1] I. Prigogine，《时间、结构与涨落》，《科学》**201**，777（1978）。[2] M. Zanin 和 D. Papo，《用于评估时间序列不可逆性的算法方法：综述与比较》，《熵》**23**，1474（2021）。[3] L. Lacasa 和 R. Flanagan，《通过非平稳过程的可视图研究时间可逆性》，《物理评论E》**92**，022817（2015）。[4] ´E. Rold´an 和 J. M. Parrondo，《从单条平稳轨迹估算耗散》，《物理评论快报》**105**，150607（2010）。[5] U. Seifert，《随机热力学、涨落定理与分子机器》，《物理学进展报告》**75**，126001（2012）。[6] E. Rold´an 和 J. M. R. Parrondo，《离散系统平稳轨迹之间的熵产生与库尔贝克-莱布勒散度》，《物理评论E》**85**，031129（2012）。[7] P. Gaspard，《混沌、散射与统计力学》，《混沌》（2005）。[8]G. Weiss，线性随机过程的时间可逆性，《应用概率学杂志》**12**，831–836（1975）。[9] M. Lothaire，*Applied combinatorics on words*，第105卷（剑桥大学出版社，2005）。[10] J.-P. Allouche、M. Baake、J. Cassaigne 和 D. Damanik，回文复杂度，《理论计算机科学》**292**，9（2003）。[11] J. Green Jr、C. S. Daw、J. S. Armfield、C. Finney、R. M. Wagner、J. Drallmeier、M. Kennel 和 P. Durbetaki，时间不可逆性与循环变异模型的比较，《SAE会刊》，355（1999）。[12] J. G. Brida，经济学中的符号时间序列分析，Documento de Trabajo/FCS-DE；10/00（2000）。[13] C. Cammarota 和 E. Rogora，时间反演、符号序列与人类心跳的不可逆性，《混沌、孤子与分形》**32**，1649（2007）。[14] W. Yao、W. Yao、J. Wang 和 J. Dai，量化

利用对称排列之间的概率差异来衡量时间不可逆性，《物理快报A》**383**，738（2019）。[15] W. Yao、J. Dai、M. Perc、J. Wang、D. Yao 和 D. Guo，癫痫脑电图中基于排列的时间不可逆性，《非线性动力学》**100**，907（2020）。[16] R. Salgado-Garc´ıa，随机长度时间序列的时间不可逆性检验：将匹配时间方法应用于 dna，《混沌：非线性科学跨学科期刊》**31**，123126（2021）。[17] G. Nicolis，熵产生的变换性质，《物理评论E》**83**，011112（2011）。[18] C. Beck 和 F. Sch ¨ *Thermodynamics of chaotic sys-* ögl, *tems* 1995（）。[19] S. Daw、C. Finney 和 M. Kennel，用于测量时间“不可逆性”的符号方法，Phys. Rev. E **62**，1912（2000）。[20] T. M. Cover， *Elements of information theory*（John Wiley & Sons，1999）。[21] R. Kawai、J. M. Parrondo 和 C. Van den Broeck，耗散：相空间视角，《物理评论快报》**98**，080602（2007）。[22] J. M. Parrondo、C. Van den Broeck 和 R. Kawai，熵产生与时间之矢，《新物理学杂志》**11**，073008（2009）。[23] M. B. Kennel，使用数据压缩字典检验时间序列中的时间对称性，《物理评论E》**69**，056208 2004（）。[24] M. Costa、A. L. Goldberger 和 C.-K. Peng，人类心跳的对称性破缺：衰老与疾病导致时间不可逆性的丧失，《物理评论快报》**95**，198102（2005）。[25] L. Lacasa, A. Nu˜ñez, E. Rold´án, J. M. Parrondo 和 B. Luque，《时间序列不可逆性：一种可视图方法》，《欧洲物理学杂志B》，217（2012）。[26] J. F. Donges、R. V. Donner 和 J. Kurths，《运用复杂网络方法检验时间序列不可逆性》，《欧洲物理快报》**102**，10004（2013）。[27] J. H. Mart´ınez、J. L. Herrera-Diestra 和 M. Chavez，《通过序数模式分析检测时间序列中的时间可逆性》，《混沌：非线性科学跨学科期刊》**28**，123111（2018）。[28] R. Salgado-Garcia 和 C. Maldonado，《从删失符号时间序列估计熵率：时间不可逆性检验》，《混沌：非线性科学跨学科期刊》**31**，013131（2021）。[29] P. E. Harunari、A. Dutta、M. Polettini 和 ´E. Rold´án，《从少量可见跃迁的统计中能学到什么？》，《物理评论X》**12**，041026（2022）。[30] P. Grassberger，《区间单峰映射的符号动力学》，Zeitschrift für Naturforschung A **43**，671（1988）。[31] R. G. James、K. Burke 和 J. P. Crutchfield，《混沌会遗忘也会记忆：信息创造、湮灭与存储的测量》，《物理快报A》**378**，2124（2014）。[32] L. Lacasa、B. Luque、I. G´ omez 和 O. Miramontes，《关于若干素数序列的动力学方法》，《熵》**20**，131（2018）。[33] L. Zunino、F. Olivares、H. V. Ribeiro 和 O. A. Rosso，《排列詹森-香农距离：用于复杂时间序列分析的灵活快速符号工具》，Phys. Rev. E **105**，045310（2022）。