arXiv:2601.01497v1 [nlin.PS] 4 Jan 2026

关于 Rényi 转移熵的实际估计与解读

Zlata Tabachov´a,[1,2,*] Petr Jizba,[1,†] Hynek Lavi˘cka,[1,3,‡] 和 Milan Palu˘s[2,§]
[1] *FNSPE, Czech Technical University in Prague, Břehová 7, 11519, Prague, Czech Republic*
[2] *Department of Complex Systems, Institute of Computer Science of the Czech Academy of Sciences, Pod Vodárenskou věží 2, 182 00 Prague 8, Czech Republic*
[3] *Cerabyte GmbH, Rundfunkplatz 2, 80335 Munich, Germany*（日期：2026年1月6日）

Rényi 转移熵（RTE）是经典转移熵的一种推广，它以 Rényi 信息度量取代香农熵。这又引入了一个新的可调参数α，用于衡量对低概率或高概率事件的敏感度。尽管 RTE 在分析复杂非高斯系统中的因果关系方面展现出巨大潜力，但其实际应用受到限制，主要原因是准确估计和解读面临挑战。在处理有限、高维或异质数据集时，这些困难尤为突出。本文利用多种合成数据集，系统研究 k最近邻估计量对 Rényi 熵（RE）和 RTE 的表现；这些数据集的构造均包含明确的因果关系。我们在广泛的参数范围内测试该估计量，包括样本量、维数、记忆长度和 Rényi 阶数 α。具体而言，我们将该估计量应用于一组结构复杂性逐渐增加的模拟过程，涵盖从线性动力学到具有多源耦合的非线性系统。为应对 RE 和 RTE 值可能为负所带来的解读难题，我们提出三项可靠性条件，并制定了调整估计器参数的实用指南。我们表明，当满足可靠性条件并相应校准参数时，所得的有效RTE估计值能够在广泛情形下准确捕捉方向性信息流。结果表明，RTE 的解释能力对 Rényi 参数 α的选择十分敏感。这凸显了 RTE 框架在识别复杂系统中极端行为驱动因素方面的实用价值。

I. 引言

理解复杂系统中的因果机制，是神经科学、能源学、气候科学、金融、工程学和非线性动力系统等众多科学领域面临的一项根本性且持续存在的挑战 [1, 2]。这些系统通常具有高维、非平稳和非线性交互等特征，因此难以识别其组成部分之间的因果关系或有效关系。当所研究的过程随时间演变，并以连续时间点记录的测量值序列表示——通常排列成时间序列——便可运用定量方法，从不同数据集中推断变量之间的因果关系。时间序列分析中，因果关系研究的一种基础方法是格兰杰因果关系（GC）。该方法最初由 Granger 提出 [3]，其基础是 Wiener 早先提出的有关可预测性的思想 [4]。Granger 的表述以线性自回归模型为基础，因其在数学上易于处理且概念简单而得到广泛应用。

然而在实践中，许多现实世界系统呈现出复杂的非线性动力学特征，超出了线性模型的适用范围。为解决这一局限，文献中提出了多种 GC 的非线性扩展，其思路借鉴了非线性动力系统理论 [5–7], 数据压缩效率 [8],机器学习 [9], 和信息论 [10–13],等领域。在此背景下，信息论为通过信息流中的方向性依赖关系分析因果关系，提供了强大的工具和概念清晰的框架。具体而言，用于检测和量化时变过程之间方向性信息流的最广泛采用工具之一是传递熵(TE)，或等价的条件互信息(CMI)——这是一种由 Schreiber [14], 、Kantz 等人 [15],和 Palu˘s 等人 [16]提出的信息论度量。

从本质上说，TE 量化了从源过程到目标过程的有向（时间非对称）信息流量。具体而言，它通过测量在考虑源过程过去状态后，目标过程未来状态的不确定性减少量来实现这一点，且该减少量超出了目标过程自身历史已经能够解释的部分。不同于基于显式建模假设的方法，TE 独立于模型，不需要预先了解底层动力学。因此，TE 特别适合用于分析复杂系统，因为其中的相互作用可能具有高度非线性，或难以通过参数化描述。由此，TE 已成功应用于多个不同的科学领域 [17, 18], ，包括地球科学 [19, 20], 、神经科学 [21, 22] 以及金融市场 [11, 15, 23]。

尽管大多数 TE 应用都依赖于 Shannon 熵

* zlatatabachova@gmail.com；通讯作者。
† p.jizba@fjfi.cvut.cz
‡ hynek.lavicka@gmail.com§
mp@cs.cas.cz

框架，但近期研究开始探索 R´enyi 信息理论 [19, 24–26]。后者为传统 TE 提供了自然的推广，同时引入了一个新的可调参数，α该参数用于调节对稀有事件与频繁事件的敏感度。这种扩展称为 R´enyi 传递熵（RTE），具有多项优势，尤其适用于呈现重尾分布、偏度或存在异常值的系统，因为这些特征对系统动力学起着关键作用 [11, 12, 19, 27, 28]。RTE 框架最初在文献 [11] 中提出，旨在研究双变量金融时间序列之间的因果信息传递。其基本思想是，R´enyi 参数 α 的特定取值可以有选择地放大底层概率分布函数（PDF）的某些区域，同时大幅削弱其他区域。由于极端事件通常与概率分布函数的尾部相关，RTE 似乎特别适合识别此类稀有事件的因果驱动因素。多项研究 [11, 12, 19, 27–29]证实了这一预期。然而，尽管理论上颇具吸引力，RTE 的应用仍不广泛，主要原因是缺少系统性研究来探讨其估计方法与可解释性，尤其是在实证数据或高维数据的背景下。

本文对 Leonenko 等人 [30] 最初提出的基于 k-最近邻（k-NN）估计量进行了全面评估，用于估计 RE 和 RTE。我们的分析分为两个阶段。首先，我们系统考察了样本量、邻居数量 k、数据维度和记忆长度等关键参数如何影响广泛 R´enyi α 参数范围内的估计精度，尤其关注 $\alpha < 1$ 情形，因为该情形强调尾部事件。为探究不同的 PDF 行为，我们将该估计量与多维 t分布的 RE 进行基准比较；该分布涵盖正态分布和柯西分布这两种极限情形，其尾部分别呈指数衰减和多项式衰减。

在第二阶段，我们将该估计量应用于一组具有明确定义因果结构的模拟耦合随机过程。这一设置能够在受控条件下评估该估计量捕捉方向性信息流的性能。具体而言，我们对四对结构复杂性逐步增加的合成过程对进行 RTE 估计和分析，其中包括线性耦合、非线性耦合和多源耦合。

本文结构安排如下。在下一节中，我们介绍 RTE 的基础概念以及 k-NN 估计量。第III 节考察该估计量在正态分布和重尾分布中的关键参数下的表现，尤其是多元 Student t分布。第IV节中，我们对四种合成耦合随机过程进行 RTE 评估。第V节将详细讨论 Rényi 熵估计量和所采用的合成随机过程相关的结果。本节还提供实用指南，介绍参数调优的方法，总结 RTE 估计的可靠性条件，并探讨关键挑战——尤其是与 α取值较小相关的挑战。第VI 节总结主要研究发现，并指出未来研究的潜在方向。五个配套附录提供了更多技术细节和符号说明。

II. R´ENYI传递熵

A. RTE的一些基础知识

考虑三个随机过程

$$X = \{x_i,\}_{i=1}^{N}, \quad Y = \{y_i,\}_{i=1}^{N} \quad \text{and} \quad Z = \{z_i,\}_{i=1}^{N}, \tag{1}$$

它们的马尔可夫阶数分别为 r、l 和 s。我们假设这些过程之间存在方向依赖关系，具体而言，除自身过去之外，X 还受到 Y 和 Z的影响，即 $X = f(X, Y, Z)$。为了量化随机过程之间的方向性信息流，R´enyi 传递熵为标准的基于香农的信息传递提供了一种自然的扩展 [14, 15, 31]。具体来说，它量化目标过程 X 的未来状态在多大程度上由源过程Y 的过去决定（下文将纳入第三个过程 Z 的影响），同时以 X 自身的过去为条件。形式上，RTE 定义如下 [11]

$$\begin{aligned} T^{R}_{\alpha,Y\to X}(r,l) &= I_\alpha(x_{t+1} : y_t^{(l)} | x_t^{(r)}) \\ &= H_\alpha(x_{t+1}|x_t^{(r)}) - H_\alpha(x_{t+1}|x_t^{(r)}, y_t^{(l)}), \end{aligned} \tag{2}$$

其中，I_α 和 H_α 分别是 R´enyi 条件互信息和 R´enyi 熵。参数 $\alpha \in \mathbb{R}_0^+$ 指定熵的阶数，且$y_t^{(l)} = y_t$、y_{t-1}、……、y_{t-l+1}。关于 I_α和 H_α 的形式化定义，请参阅附录 B。

当 $\alpha = 1$时，RTE 会退化为 Schreiber [14] 最初提出的基于香农熵的传递熵。因此，RTE 是传统基于香农熵的传递熵的推广，并具有以下基本性质：

- 一般而言$T^{R}_{\alpha,\ Y\to X} \neq T^{R}_{\alpha,\ X\to Y}$，体现了其方向性。
- $T^{R}_{\alpha=1,\ Y\to X} \geq 0$，因为香农传递熵是通过 Kullback–Leibler散度定义的一个特例。
- $T^{R}_{\alpha=1,\ Y\to X} = 0$意味着 X 与 Y 之间具有统计独立性。
- 对于 $\alpha \neq 1$，RTE 可以取负值，因为额外的条件信息（例如通过过程 $y_{(l)\ t}$ 引入）可能会增加 Rényi 熵，即 $H_\alpha(x_{t+1}|x_{(r)\ t}, y_{(l)\ t}) > H_\alpha(x_{t+1}|x_{(r)\ t})$。

由于传递熵具有内在的方向性，因此被广泛用作信息流 [10], 的代理，作为因果关系的信息论指标。然而，区分预测能力与因果影响十分重要：TE 主要衡量了解源能在多大程度上减少对目标的不确定性，而信息流关注的是源的变化如何直接影响目标。如 [2], 在适当的条件化和合适的记忆参数下，完全传递熵 可以作为实际信息流的稳健近似。尽管在 Rényi 熵的背景下对信息流进行完整的理论阐述超出了本文的范围，我们仍强调其相关性，因为我们的实证结果显示二者存在显著的相似之处。

遵循 Prokopenko 和 Lizier [2], 提出的框架，他们区分了表观和完全香农转移熵，我们将这一差异扩展到 R´enyi 熵的情形。

当目标过程被描述为 $X = f(X, Y, Z)$ 时，若使用公式2计算 RTE，却未考虑所有影响源，则得到的只是从 Y 到 X的表观传递熵。要获得完整的传递熵，必须以所有影响目标的源为条件，其中包括 Z。因此，从 ´到的完整 Rényi 传递熵 Y 至 X 定义如下：

$$\begin{aligned} \bar{T}^{R}_{\alpha,Y\to X|Z}(r,l,s) &= I_\alpha(x_{t+1}:y_t^{(l)}|x_t^{(r)},z_t^{(s)}) \\ &= H_\alpha(x_{t+1}|x_t^{(r)},z_t^{(s)}) \\ &- H_\alpha(x_{t+1}|x_t^{(r)},z_t^{(s)},y_t^{(l)}). \end{aligned} \tag{3}$$

一般而言，当多个源直接影响目标过程 $Z_1, \ldots\ldots, Z_n$ 时，必须将它们全部纳入条件变量，才能准确刻画其背后的因果结构。同样重要的是，谨慎选择记忆参数——目标过程和源过程的记忆参数为 r 和 l，其他源的记忆参数为 $s_1, \ldots\ldots, s_n$。这些参数决定了依赖关系的时间深度，对于获得完整的 R´enyi 传递熵，而非仅仅是表观传递熵，至关重要。因此，准确识别所有相关源过程，并恰当地设定记忆长度，对于近似真实的信息流至关重要。

最后，当目标过程 X 仅受单一源 Y影响时，完全 RTE 与表观 RTE 等价，即 $\bar{T}^{R}_{\alpha,\,Y\to X} \equiv T^{R}_{\alpha,\,Y\to X}$。此外，我们注意到，定义 $\bar{T}^{R}_{\alpha}$对应于 [32]中引入的条件传递熵或因果熵。为清晰起见，我们将使用记法T^{CR}_{α}明确表示条件 Rényi 转移熵。如果条件不完整或记忆参数选择不当，所得的条件 Rényi 转移熵应解释为信息传递的表观度量，而非完全度量。因此，尽管 RTE、完全 Rényi 转移熵和条件 Rényi 转移熵的数学定义形式可能相同，但应将其解释为完全度量还是表观度量，取决于所考察过程的具体结构。

B. RTE 的估计

利用式（3）中的表达式，RTE 可以等价地改写为：

$$\begin{aligned} T^{R}_{\alpha,Y\to X|Z}(r,l,s) &= \underbrace{H_\alpha(x_{t+1},x_t^{(r)},z_t^{(s)})}_{\mathscr{I}_1} - \underbrace{H_\alpha(x_t^{(r)},z_t^{(s)})}_{\mathscr{I}_2} \\ &- \underbrace{H_\alpha(x_{t+1},x_t^{(r)},z_t^{(s)},y_t^{(l)})}_{\mathscr{I}_3} + \underbrace{H_\alpha(x_t^{(r)},z_t^{(s)},y_t^{(l)})}_{\mathscr{I}_4}. \end{aligned} \tag{4}$$

因此，计算 RTE 可归结为估计四个 Rényi熵，为简便起见，我们将其分别记为 $\mathscr{I}_1$、$\mathscr{I}_2$、$\mathscr{I}_3$ 和 $\mathscr{I}_4$。每个熵项的维数取决于分别与目标、源和附加源过程相关联的马尔可夫阶数 r、l 和 s。

为了估计这些熵项，我们采用 k-NN 微分 Rényi 熵估计量，即

$$H_\alpha(f_X) = \frac{1}{1-\alpha}\log_2\int_{\mathbb{R}^d} f^\alpha(x)d^dx, \tag{5}$$

由 Leonenko 等人提出，用于定义有界支撑上具有概率密度函数 f_X 的随机向量$X \in \mathbb{R}^d$ [30]，该函数相对于 d维勒贝格测度是良定义的。此估计量为：

$$\hat{H}_\alpha(N,k,\alpha,d) \;=\; \frac{1}{1-\alpha}\log_2\left(\frac{N-1}{N}B_d^{1-\alpha}C_k^{1-\alpha}\sum_{i=1}^{N}\frac{\rho_{k,i}^{d(1-\alpha)}}{(N-1)^\alpha}\right), \tag{6}$$

其中 $C_k = \left[\Gamma(k)\ (k+1-\alpha)\right]^{1/(1-\alpha)}$，且 $B_d = \Gamma^{\pi^{d/2}}$（$d/2+1$）表示 d维单位球的体积。项$\rho_{k,\ i}$表示 d维欧几里得距离，即第 i个点与其第 k个最近邻之间的距离 [33]。

香农熵的估计量定义（即 $\alpha = 1$时）也见于 [30]。不过，由于 Rényi 熵是 α的连续函数，因此也可以通过插值来近似其在 $\alpha = 1$ 处的值。

需要注意的是，Leonenko 估计量是针对微分 Rényi 熵推导的，即连续概率密度函数的 Rényi 熵。与离散情形不同，微分 Rényi 熵并不一定非负——甚至香农熵 [34, 35] 也不例外。对于退化分布或具有高度尖峰特征的满维分布，熵可能为负。

由于 Rényi 传递熵最初是针对离散过程定义的，因此有必要评估使用连续估计量会如何影响该度量的解释。Rényi 熵的离散形式与连续形式之间的关系由 [24, 36] 给出：

$$H_\alpha(f_X) \;=\; \lim_{M\to\infty}\left[H_\alpha(X) \;-\; d\log M\right]. \tag{7}$$

此处，M 表示用于镶嵌划分连续空间的 d维箱体数量。例如，在一维情况下，M 对应区间或分箱的数量。

关系式（7）表明，从离散情形过渡到连续情形会引入一个常数偏移量，该偏移量取决于空间的维数和离散化尺度。

III. 使用 k-近邻方法估计 Rényi 熵的参数

A. 样本量参数 N。

-近邻估计量的准确性对样本量高度敏感。较大的样本量能更有效地捕捉底层分布结构，从而得到更可靠的 R´enyi 熵（Rényi 熵）估计。图1展示了这一效应：随着 N 增大，三维正态分布的 Rényi 熵估计值逐渐收敛至解析值（柯西分布的情况也类似，如图12b所示）。

Rényi 熵估计的精度也与 R´enyi 参数 α密切相关。具体而言，对于较小的α，在图 1 中 $\alpha < 0.7$的示例里，即使相当大，估计值仍表现出更高的变异性和系统偏差

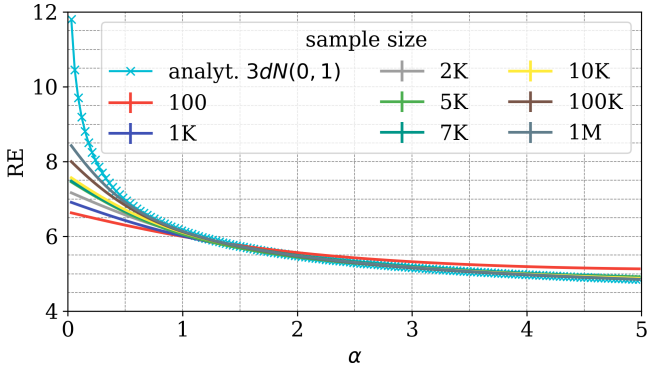


图 1： *Effect of Sample Size N on the Estimation of Rényi Entropy for a 3D Normal Distribution.* x轴表示 Rényi 参数 α，而 y 轴显示协方差矩阵为单位矩阵的三维正态分布的 Rényi 熵估计值。青色叉号表示解析值。实线表示使用 $k = 10$ 得到的估计值，样本量分别为 $N = 100$、10_3、2×10^3、5×10^3、7×10^3、10_4、10_5和 10^6, , 对应的颜色依次为红色、蓝色、浅灰色、绿色、青绿色、黄色、棕色和深灰色。对于 $\alpha > 1$，即使样本量适中（例如 $N = 10^3$），也能获得准确的估计值。相比之下，较小的 α——它会着重反映低概率事件——对 N高度敏感，只有在 $N = 10^6$时，$\alpha > 0.5$ 的估计值才会变得可靠。这种在较低 α下准确性下降的现象，部分归因于无界支撑上的密度估计所面临的困难。

样本。这在一定程度上是因为较小的α 值会加大低概率（尾部）事件的权重，而这类事件在有限样本中往往代表不足。本节末尾将详细总结估计$\alpha < 1$ 的 Rényi 熵时所面临的挑战。

在传递熵的语境下，可使用如下定义的有效 *R´enyi* 传递熵 （ERTE） [15], 来减轻样本量有限带来的不利影响：

$$T^{\text{eff}}_{\alpha,Y\to X}(r,l) \;\equiv\; T_{\alpha,Y\to X}(r,l) \;-\; T_{\alpha,Y_{\text{sh}}\to X}(r,l). \tag{8}$$

这里，源过程 Y_{sh} 的打乱版本保留了边缘统计量（例如均值、方差或功率谱），但会使用 Fisher–Yates 算法[37]等方法破坏时间依赖关系。因此，理论上，从 Y_{sh} 到 X 计算得到的传递熵应为零。该量的任何非零估计值都反映了有限样本效应引入的偏差。通过减去这一基线，ERTE 得以提供一种定向熵传递的偏差校正测度。请注意，也可以打乱目标过程，而非源过程。应根据对估计偏差主要来源的识别来决定打乱哪个过程；这些偏差通常源于高维性或重尾分布。

此外，有效传递熵这一概念使得 Leonenko 估计量能够应用于传递熵的评估，该传递熵最初是为离散系统定义的微分 Rényi 熵的

估计。当将公式（7）代入有效 Rényi 转移熵（8）的表达式时，与维数和分区有关的常数会相互抵消。因此，最终结果中仅有估计得到的微分 Rényi 熵项起作用。这种抵消证明了在有效传递熵框架内使用微分熵估计器的合理性，即使原始公式是离散形式的。

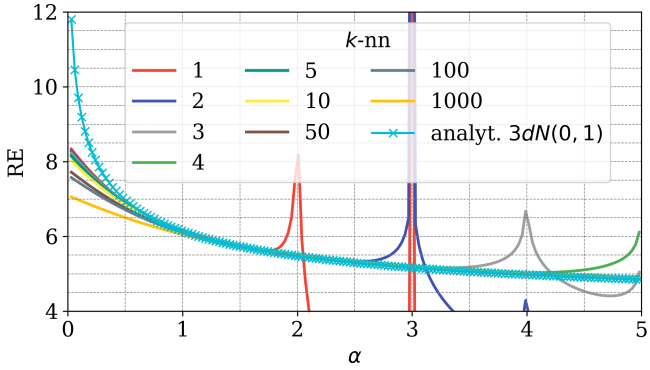


图 2 *Influence of the k-Nearest Neighbor Parameter on Estimation Accuracy for 3D Normal Distribution.* x α 轴表示 Rényi 熵参数，y 轴表示具有单位协方差矩阵的三维正态分布的 Rényi 熵。解析值以青色十字表示，k = 1 2 3 4 5 10 50 100 1000 、、、、、、、、、的估计值以实线表示。估计基于包含 $N = 10^5$ 个数据点的样本获得。由于估计量依赖于 $\Gamma(k+1-\alpha)$，当 α 和 k 的特定组合使伽马函数的自变量趋近于零并变为负值时，结果会发散。对于 $\alpha < 1$，估计值在 k 较大时发散，而 α 较高时会产生更稳定的结果。这表明，在不同的 α-区间内，k 的最优选择可能有所不同。具体而言，对于 $\alpha < 1$，使用 $k = 1$ 可获得最佳性能。

B. 近邻参数 k

在 Rényi 熵估计的关键参数中，近邻参数 k 的选择尤为微妙且具有挑战性。尽管其最优值自然取决于样本量 N，但一种广为引用的启发式方法建议设为 $k = \sqrt{N}$，见 [1]。然而，这一规则并未考虑数据分布在相空间中的局部结构。

直观而言，数据密集区域内有更多邻居分布在较小半径范围内，因此可以采用更大的 k 值。相反，在数据稀疏区域，较小的 k 更为合适，可避免高估局部概率密度。在这类区域使用过大的 k 可能会人为抬高估计熵，而在密集区域使用过小的 k 则可能导致距离极短$\rho_{k,\,i}$使得式中对数函数的自变量6 接近于零，从而产生负的熵估计值。

估计量对 k 的敏感性也会随 Rényi 参数 α 而变化。对于较小的 α，由于其赋予低概率（尾部）事件更大的权重，因此采用较小的 k 值更为合适。反之，较大的 α 值会强调高概率事件，因而适合使用较大的 k。

图2展示了这些依赖关系，其中给出了在一个包含 $N = 10^5$ 个样本的三维正态分布数据集上，使用一系列 k 值：$k = 1$、2、3、4、5、10、50、100 和 1000 得到的 Rényi 熵估计（柯西分布见图12a）。对于较小的 α，即 $\alpha < 1$，使用 $k = 1$ 得到的估计最准确，其估计值从 $\alpha \approx 0.5$ 开始收敛至理论值。增大样本量可以进一步提高精度。然而，正如 [30]，中所强调的，对于无界支撑的分布——例如正态分布——在极限 $\alpha \to 0$ 下会出现误差，因为无界性会因极端尾部行为的影响不断增强而违反收敛的前提条件。对于 $\alpha \geq 1$，较大的 k 能得到准确的估计。需要注意的是，估计量中存在函数 $\Gamma(k+1-\alpha)$，因此某些 k 与 α 的组合会受到限制。这就是图2中较小 k 的结果在 $\alpha \geq k+1$ 时发散的原因。

C. 样本维数参数 d。

如图3 和图12d所示，R´enyi 熵估计的准确性会随着数据维数的增加而下降。这种效应体现了众所周知的维数灾难 [38–40] , 它限制了非参数估计量（例如k-最近邻方法）在高维空间中的可靠性。

在 RTE 估计中，熵项 $\mathscr{I}_1$ through $\mathscr{I}_4$ 的维数由记忆参数（或马尔可夫参数） r、l 和 s 决定；这些参数分别定义了公式 4 中目标、源及附加过程的滞后状态向量。例如，当 $r = l = s = 1$ 时，熵项形式如下：$\mathscr{I}_1 = H_\alpha(x_{t+1}, x_t, z_t)$、$\mathscr{I}_2 = H_\alpha(x_t, z_t)$、$\mathscr{I}_3 = H_\alpha(x_{t+1}, x_t, z_t, y_t)$ 和 $\mathscr{I}_4 = H_\alpha(x_{t+1}, x_t, z_t, y_t)$，其维数分别为 3、2、4 和 3。

其中，$\mathscr{I}_3$ 的维数始终最高，因此通常对估计误差的贡献最大。随着维数增加，相空间中的点密度会降低，从而降低基于局部邻域的估计的可靠性。因此，谨慎选择记忆参数至关重要——这不仅关系到能否捕捉系统的动力学特性，也关系到能否确保估计量的稳定性和可解释性。

$\alpha < 1$ 的估计

图 1–3表明，当 R´enyi 熵（RE）的估计值趋于偏离解析值，且偏差更加显著时

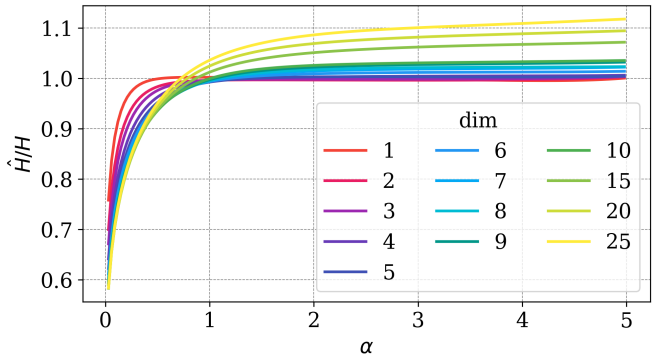


图 3 *Effect of Sample Dimensionality d on Estimation Accuracy.* x: α </style} 横轴表示 Rényi 参数，y 轴显示协方差矩阵为单位矩阵的 d 维正态分布的估计 Rényi 熵与解析 Rényi 熵之比。估计基于大小为 $N = 10^5$ 的样本计算。比值接近 1 表示估计精度较高。结果表明，维数增加会显著降低估计器性能，尤其是在 α 较小时，此时对低概率区域更为敏感。

当 α 降至 1 以下时，便会出现上述情况。在此范围内，样本量较大（N）、近邻值较小（k）且数据维数较低（d）时，估计结果最为准确。相反，高维性和有限的数据会显著降低估计精度。

对于 $\alpha < 1$，RE 会加大低概率事件的权重，同时降低高概率区域的权重。在极限情况下，当 $\alpha \to +\infty$ 时，熵由概率最高的事件主导；而当 $\alpha \to 0$ 时，所有事件都被视为概率相同，对应于均匀分布。然而，在无界域上并未定义均匀分布——这正是重尾分布的特征——因此在此类情形下进行 RE 估计颇具挑战。

我们使用多元 Student t-分布 $\mathcal{P}_{\mathcal{S}}$来说明这一行为，其中自由度参数 ν 控制分布尾部的厚重程度。附录 B 中给出了的 Rényi 熵的形式推导。只有满足条件 $\left(\alpha^{\frac{\nu+d}{2}} - \frac{d}{2}\right) > 0$时，Rényi 熵的积分表达式才能解析求解。此约束为 α 的允许取值设定了下限，该下限取决于分布的维数 d 和自由度 ν 。

图 4a 展示了尺度矩阵为单位矩阵、自由度各异的三维 t分布的 Rényi 熵（RE）解析估计值和经验估计值，其中自由度为（ν =1、2、3、4、5、10、50、100、200、1000）。解析值以叉号标示；经验估计值使用 $N = 10^6$ 个样本和 $k = 1$计算，其中 $k = 1$对于$\alpha < 1$ 是最优的。值得注意的是，当 $k = 1$时，该估计量仅在 $\alpha < 2$ 时有效（见公式 6），这决定了图的右边界。当 $\nu = 1$ 时（即柯西分布），由于尾部极重，RE 在大多数 $\alpha < 1$ 范围内未定义。随着 ν 增大，分布趋近于正态分布，有效 α 值的范围也随之扩大。

图 4b 展示了估计误差，即 RE 估计值与解析值之比，其中4b，图中x轴采用对数刻度，以突出显示 $\alpha \in (0, 1)$ 区间。结果表明，随着尾部厚度增加（ ν降低），估计值与解析值的偏差会在 α 区间 (0，1) 上逐渐出现在更高的位置。换言之，对于任意固定的 $\alpha < 1$，随着分布的重尾程度减轻（即 ν 增大），RE 估计的准确性会提高。对于 $\alpha > 1.5$ ，偏差同样是由于 $k = 1$的取值过小所致。

E。RTE估计的可靠性条件 tion

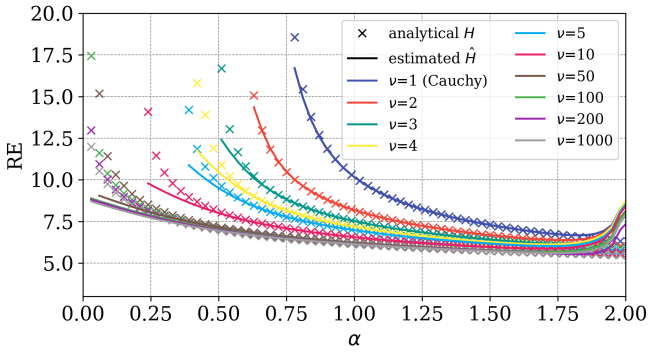
将 k-近邻估计量用于微分 Rényi 熵时，负值可能由两个主要原因造成。第一种原因来自估计量本身——在某些参数设置下，此类数值可能是有限样本效应、估计偏差、舍入误差或数值不稳定性造成的伪象。第二种可能是底层过程确实具有极低的熵。这两种情况都可能导致对有效RTE的错误评估，进而造成假阳性或假阴性。为支持稳健解读，我们建议在RTE估计过程中加入以下诊断检查，或采用式（可靠性条件（4）中定义的分量熵。

- 所有单个熵估计值均非负： $\mathscr{I}_1, \mathscr{I}_2, \mathscr{I}_3, \mathscr{I}_4 \geq 0$。
- 所有条件熵估计值均非负： $\mathscr{I}_{12} \equiv \mathscr{I}_1 - \mathscr{I}_2 \geq 0$和 $\mathscr{I}_{34} \equiv \mathscr{I}_3 - \mathscr{I}_4 \geq 0$。
- 香农转移熵的估计值（对应于 $\alpha = 1$）必须非负，即$T^R_{\alpha=1} \geq 0$ 。

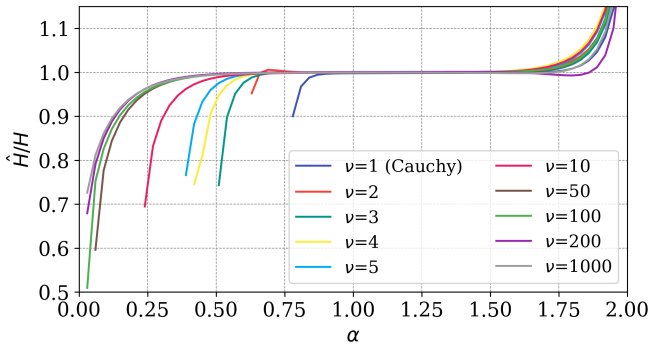
这些检查基于离散 Rényi 熵已知的理论界，具体为：$0 \leq H_\alpha(X) \leq \log_2 N$和 $0 \leq H_\alpha(X|Y) \leq \log_2 N$，其中 N表示样本量。若违反上述任一条件，则可能需要进一步检查或适当修正所得估计值，以确保其可解释性。

IV. 线性与非线性耦合随机系统中 R ´ényi 传递熵的估计

估计 RTE 的主要挑战在于需要处理高维、相关的联合概率分布。此外，底层过程通常是非平稳的，并表现出重尾行为。这些特征使得正确选择 k 最近邻和记忆参数时存在很大的不确定性。为了直观了解这些参数选择如何影响估计，



(a) *Estimated REs of 3-dim Multivariate t-distribution for various degrees of freedom* ν. x ´ , - 轴表示 Rényi 熵参数 α，而 y-轴表示以比特为单位的 R´enyi 熵（RE）。解析结果以 x 符号标出，实线表示使用 $N = 10^6$ 个样本和 $k = 1$得到的估计值，因为这种设置能为 $\alpha < 1$提供最准确的估计。对于$\nu = 1$，底层分布为柯西分布，其尾部最重。相反，当 $\nu \to +\infty$ 时，该分布收敛至尾部呈指数衰减的多元正态分布。对于 $\nu = 1000$，RE 在数值上已与三维正态分布的 RE 十分接近（参见图 1 进行比较）。解析值与估计值之间的偏差将在（b）图中讨论。



(b) *Estimate-to-truth ratios for Multivariate t-distribution.* RE 参数 $\alpha \in (0, 2)$ 位于对数刻度的 x 轴上，图 4a 中估计值与解析值的比值位于 y 轴上。为获得最佳精度，估计值是在样本量为 $N = 10^6$ 且 $k = 1$的条件下得到的；如图 1 和图 2 所示，这些设置可提高较小 α 时的准确度。尾部越重，$\alpha < 1$越早出现偏差。例如，当 $\nu = 1$（柯西分布）时，估计值在 $\alpha = 0.8$处偏差达到 3%；而对于 $\nu = 1000$，偏差仅为 0.001%。在 $\alpha \geq 0.9$时，比值等于 1，表明估计量在此区间内精度很高，但在达到 2 之前便开始偏离。这是因为对于 $k = 1$，当 α 较高时，该估计量的定义不够完善，因此此处应采用更高的 k。

图 4：尾部厚度变化时估计量的表现。

我们将该方法应用于四个具有说明性的玩具模型过程。首先从一个具有线性耦合和单一驱动过程的简单高斯系统开始，随后将分析扩展到具有非线性依赖的系统。通过引入额外的源并结合线性和非线性耦合，进一步增加系统的复杂性。此外，对于高斯系统，可以对 RTE 进行解析推导，从而能够将数值估计与精确的理论值进行验证。

A. 过程 1：与单个源的线性耦合

首先考虑的是一个带有高斯噪声的一阶线性自回归模型（AR(1)）。在此配置中，A 充当源（驱动或主导）过程，而 B^I 表示目标（或被驱动）过程。系统按照以下方程演化：

$$\begin{aligned} A_{t+1} &= aA_t + \varepsilon_t^A, \\ B_{t+1}^I &= bB_t^I + \eta A_t + \varepsilon_t^{B^I}, \end{aligned} \tag{9}$$

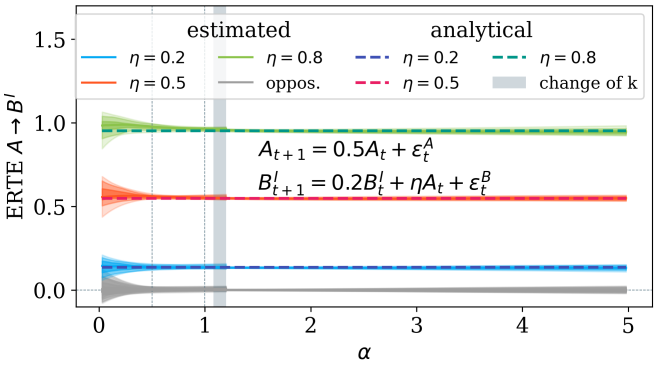
此处，耦合项 η 引入了目标对源的线性依赖。参数设为 $a = 0.5$, $b = 0.2$，其中 $\varepsilon_t^A \sim \mathcal{N}(0, 1)$，且 $\varepsilon_t^{B^I} \sim \mathcal{N}(0, 0.25)$。通过改变耦合系数 η，考察信息传递如何随相互作用强度的变化而响应。

图 5a 展示了两个方向上估计得到的有效 Rényi 转移熵（ERTE），即 $T\text{eff}_{\alpha, A\to B^I}(1, 1)$ 和 $T\text{eff}_{\alpha, B^I\to A}(1, 1)$。马尔可夫阶数 $r = 1$ 和 $l = 1$根据过程的定义确定。对于 $\eta = 0.2$ 的每个取值 0.5 和 0.8，系统均模拟 30 次，每次模拟包含 $N = 10^5$ 个时间˜步。每次模拟中，计算熵项 $\mathscr{I}_1$、ˆˆˆˆˆ$\mathscr{I}_2$、$\mathscr{I}_3$、$\mathscr{I}_4$、$\mathscr{I}_{12}$ 和 $\mathscr{I}_{34}$（见图 5b），随后用这些熵项推导 ERTE；最后两项还在打乱的源变量上进行计算。

为概括估计结果的变异性，我们使用基于百分位数的可视化方法汇总 30 次模拟的结果。对于每个 α 值，中心曲线表示 ERTE 中位数。围绕该中位数的半透明色带表示选定经验百分位数之间的区间：第 1–99、第 5–95 和第 25–75 百分位数。越靠近中位数的色带不透明度越高，反映出分布更集中，因此估计量的稳定性也更有保障。

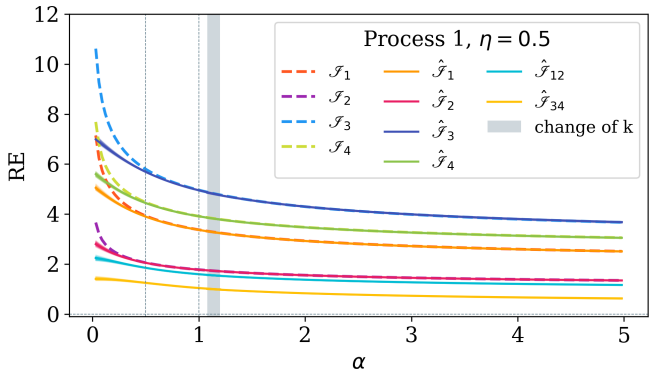
k-NN 参数在 $\alpha \in (0,\ 1.23)$时设为 $k = 3$，在 $\alpha \in (1.08,\ 5)$时则增至 $k = 50$，以提高在较大的 α 值会强调高概率事件的范围内的稳健性。灰色竖带表示这一瞬态区间。另两条竖直虚线分别位于 $\alpha = 0.5$ 和 $\alpha = 1$，后者对应香农情形。

蓝色、红色和绿色的分位数带（自下而上）分别对应耦合强度为 $\eta = 0.2$、0.5 和 0.8 时因果方向上的 ERTE 估计值。这些结果与虚线所示的理论值高度吻合。由于这两个系统



(a) *Effective Rényi transfer entropy in both directions between A and B^I.* *x* *´* *α*

横轴表示 Rényi 参数，而 y 轴展示两个过程之间估算得到的和解析的 R´ényi 传递熵值，单位为比特。 $k = 3$ 的取值用于 $\alpha < 1.1$，并用于区间其余部分的 $k = 50$；转变位置以灰色竖带标出。记忆参数 $r = l = 1$ 由过程定义决定。蓝色、红色和绿色分位数带——分别为第 1–99 、第 5–95 和第 25–75 百分位数，线条表示 30 次运行的中位数——（自下而上）对应因果方向上的估算有效传递熵，即在耦合强度为 $\eta = 0.2$、0.5 和 0.8 时，从 A 到 B^I 的传递熵。耦合强度越大，信息传递越强，估算值与解析结果吻合度很高。接近零的灰线表示反向的传递熵（从 B^I 到 A），此时预期不存在因果关系。由于 A 和 B^I 均服从正态分布，传递熵与 α 无关。



(b) *Rényi entropies $\mathscr{I}_1$ $\mathscr{I}_2$ $\mathscr{I}_3$ $\mathscr{I}_4$ $\mathscr{I}_{12}$ and $\mathscr{I}_{34}$ underlying the ,,,, RTE $A \to B^I$ for coupling parameter $\eta = 0.5$.* *x* *y*轴表示 R´ényi 参数 α，而 y轴显示以比特为单位测得的相应 R´ényi 熵的估计值和解析值。虚线表示 $\mathscr{I}_1 - \mathscr{I}_4$ 的解析结果，分位数带则显示 30 次运行中的估计结果。熵 $\mathscr{I}_1$、 $\mathscr{I}_2$、 $\mathscr{I}_3$ 和 $\mathscr{I}_4$（维数分别为 2、1、3、2）依次以橙色、洋红色、蓝色和绿色表示。估计值与解析值吻合良好，但在较小值处会出现偏差，尤其是当 $\alpha < 0.5$ 时。条件熵 $\mathscr{I}_{12} = \mathscr{I}_1 - \mathscr{I}_2$ 和 $\mathscr{I}_{34} = \mathscr{I}_3 - \mathscr{I}_4$ 分别以青色和黄色表示。估计值满足所有可靠性条件，即 $\hat{}\mathscr{I}_1$, $\hat{}\mathscr{I}_2$, $\hat{}\mathscr{I}_3$, $\hat{}\mathscr{I}_4$, $\hat{}\mathscr{I}_{12}$ 和 $\hat{}\mathscr{I}_{34}$ 均为正。此外，由于 $\alpha = 1$ 的香农转移熵满足 $\hat{}\mathscr{I}_{12} > \hat{}\mathscr{I}_{34}$，其值也为正。这意味着所得的 RTE 和有效 Rényi 转移熵具有可解释性。

图 5：过程 1: 线性单源耦合下 Rényi 熵与有效转移熵的估计值

高斯分布下，理论 RTE 与 α 无关，其结果取决于协方差矩阵和维数，详情见公式（B17）。在所有 η 取值范围内，反向的传递熵近似为零（三个灰色分位数带），这与模型的单向结构一致。

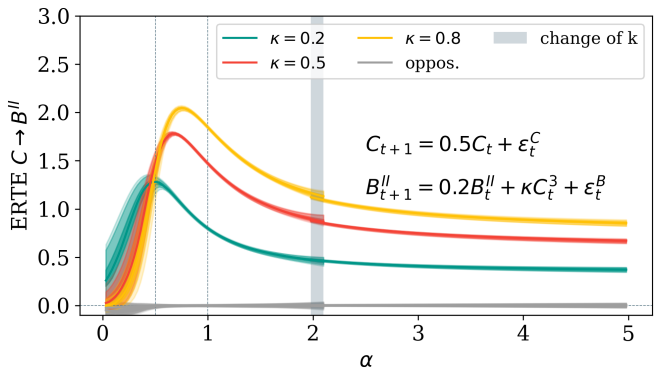
为验证第三节 E 中介绍的三个可靠性条件，图 5b 展示了构成从 A 到 B^I 的有效 Rényi 转移熵（ERTE）的熵项 $\mathscr{I}_1$、 $\mathscr{I}_2$、 $\mathscr{I}_3$ 和 $\mathscr{I}_4$ 的估计值，以及耦合为 $\eta = 0.5$ 时对应的解析值。图中还给出了估计值的分位数带；不过，偏差过小，在 y 轴的尺度下无法看清。灰色竖带表示估计器参数 k 从 3 切换至 50 的区间。对于 $\alpha < 1$，当 $k = 3$ 时，Rényi 熵（ RE ） $\mathscr{I}_1$、 $\mathscr{I}_2$、 $\mathscr{I}_3$ 和 $\mathscr{I}_4$（维数分别为 2、1、3、2）的估计值偏离了 $\alpha < 0.5$ 的解析值，这与图 1 和图 2 所示结果一致。尽管如此，熵估计值均为非负值；且面板 (a) 证实，Shannon 转移熵，即在 $\alpha = 1$ 时的值，也为非负值，因此三个可靠性条件均得到满足。相反方向的 Rényi熵也满足可靠性条件，见图 13a。

为了说明可靠性条件较难满足的情况，我们展示了附录中 Process1 的一个略作修改的版本C。该修改通过降低 ε^B的方差，增强了相关强度，并使源过程与目标过程之间形成近乎确定性的耦合。在这些条件下，先前使用的记忆和k-NN 参数不再满足可靠性标准。我们将说明，要纠正估计所得 Rényi 熵的负值，需要同时增加记忆长度和最近邻的数量。

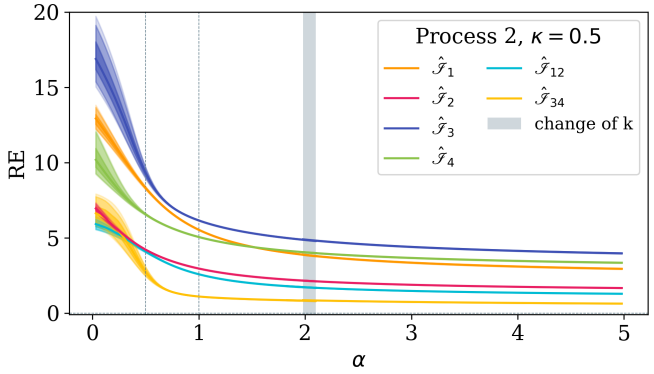
为说明偏差扣除的重要性，我们在图 14a 中给出了 $k = 50$ 的估计 RTE 值（未进行置乱校正），并在图 14b 中给出了 $k = 3$的估计值，二者均涵盖 α的完整范围。对于 $\alpha < 1$，由于有限样本效应，RTE 估计值偏离理论值。通过扣除从置乱源序列计算得到的 RTE，可以有效校正这一系统偏差，从而得到有效RTE；在整个 α 范围内，对于 k的两种取值，有效RTE均与解析预测高度吻合。值得注意的是，对于 $\alpha = 0.5$，RTE 估计值从 $k = 3$ 开始便已较为准确；而对于 $k = 50$，只有从约 $\alpha \approx 1$开始才能获得可靠的估计值。这一行为与先前的观察一致，如图 2所示。

B. 过程 2：与一个源的非线性耦合

第二个过程模型描述了一种非线性交互，其中目标变量 B^{II} 呈立方关系地依赖于



(a) *Effective Rényi transfer entropy in both directions between C and B^{II}.* *x* *α*



(b)

Rényi entropies $\mathscr{I}_1$ $\mathscr{I}_2$ $\mathscr{I}_3$ $\mathscr{I}_4$ $\mathscr{I}_{12}$ and $\mathscr{I}_{34}$ underlying the , RTE $C \to B^{II}$ for coupling parameter $\eta = 0.5$. *x*

横轴表示 Rényi 参数，而 *y*纵轴则以比特为单位，显示两个过程之间估计得到的 Rényi 传递熵。 $k = 3$ 取值用于 $\alpha < 2.1$，其余区间则使用 $k = 50$；转变处以灰色竖带标示。记忆参数 $r = l = 1$ 由过程定义确定。青色、红色和黄色分位数带——分别为第 1–99、第 5–95 和第 25–75 百分位数带，并以线条标出 30 次运行的中位数——从下到上分别对应因果方向上的估计有效传递熵，即在耦合强度为 $\kappa = 0.2$、0.5 和 0.8 时，从 C 到 B^{II} 的传递熵。耦合强度越大， $\alpha > 0.7$的信息传递越高；但在较小取值下，这种依赖关系则相反。这一微妙的转变凸显了低α Rényi 熵对尾部事件的敏感性，以及估计量在此区间内变异性增大；立方变换引入的分布重尾特性使这一现象更加明显。接近零的灰色线条表示反方向（从B^{II} 到 C）的传递熵，此处预计不存在因果关系。估计结果再次显示，在 α 取值较小时变异性增大，这主要是受重尾分布的影响。

, , , 横轴表示 Rényi 参数 α，纵轴显示以比特为单位测得的相应 Rényi 熵的估计值。分位数带展示了 30 次运行的估计结果。熵 $\mathscr{I}_1$、$\mathscr{I}_2$、 $\mathscr{I}_3$ 和 $\mathscr{I}_4$ （维数分别为 2、1、3、2）分别以橙色、洋红色、蓝色和绿色表示。当 α 较小时，估计值呈现出较大的变异性，且变异性会随维数增加而进一步增大。尤其是， $\mathscr{I}_3$ 在三维情况下计算得出，表现出最大的方差。条件熵 $\mathscr{I}_{12} = \mathscr{I}_1 - \mathscr{I}_2$ 和 $\mathscr{I}_{34} = \mathscr{I}_3 - \mathscr{I}_4$ 分别以青色和黄色表示。所有估计值均满足可靠性条件，即 $\hat{}\,\mathscr{I}_1$、$\hat{}\,\mathscr{I}_2$、$\hat{}\,\mathscr{I}_3$、$\hat{}\,\mathscr{I}_4$、$\hat{}\,\mathscr{I}_{12}$ 和 $\hat{}\,\mathscr{I}_{34}$ 均为正。此外，由于 $\hat{}\,\mathscr{I}_{12} > \hat{}\,\mathscr{I}_{34}$， $\alpha = 1$ 的 Shannon 转移熵也为正。这意味着所得的 RTE 和有效RTE 均具有可解释性。只有在 $\alpha < 0.3$ 时，传递熵才会变为负值，但这种由有限样本量导致的偏差可通过有效RTE有效校正，见图 15。

图6：过程 2：源过程 E 非线性单源耦合的 R´ényi 熵估计值和有效传递熵 ng

源过程 C。系统按照以下规律演化：

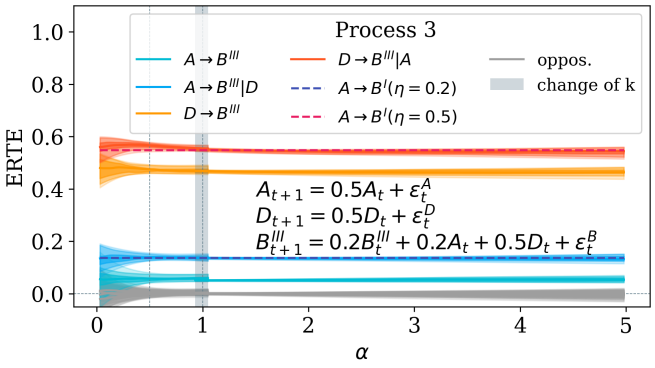
$$C_{t+1} = cC_t + \varepsilon_t^C,$$
$$B^{II}_{t+1} = bB_t^{II} + \kappa C_t^3 + \varepsilon_t^{B^{II}}, \qquad (10)$$

其中，三次项 κC_t^3 引入了从 C 到 B^{II}的非线性耦合。为确保模拟结果的一致性，所有参数均固定为： $c = 0.5$、$b = 0.2$、 $\varepsilon_t^C \sim \mathcal{N}(0, 1)$ 和 $\varepsilon_t^{B^{II}} \sim \mathcal{N}(0, 0.25)$。根据模型定义，记忆参数为 $l = r = 1$，且预期反向（即从 B^{II} 到 C）不存在信息传递。

图6a 展示了两个方向上的有效RTE， $T_{\mathrm{eff}\,\alpha,\, C\to B^{II}}(1, 1)$ 和 $T_{\mathrm{eff}\,\alpha,\, B^{II}\to C}(1, 1)$，对应的耦合强度为 $\kappa = 0.2$、0.5 和 0.8。 x轴对应 Rényi 参数 R´，而 y轴显示以比特为单位的 ERTE 估计值。对于每个耦合值，系统均模拟 30 次，每次运行包含 $N = 10^5$ 个样本。在每次运行中，计算估计值 $\mathscr{I}_1$、 $\mathscr{I}_2$、 $\mathscr{I}_3$、 $\mathscr{I}_4$、$\mathscr{I}_{12}$和 $\mathscr{I}_{34}$ ，随后据此推导 ERTE。

结果采用基于分位数的摘要进行可视化。对于每个 α，中央粗线表示 30 个实现值中的 ERTE 中位数。中位数周围的阴影区域表示离散程度逐渐增大的分位数区间：第 1–99、第 5–95 和第 25–75 百分位数；不透明度越高，表示分布越集中。这种方式能够清晰地呈现估计值的集中趋势和变异性。对于 $\alpha \in (0, 2。1)$，近邻参数设为 $k = 3$；对于 $\alpha \in (1。8, 5)$，则提高至 $k = 50$，以确保在更高的 α 范围内估计值的稳定性和准确性。

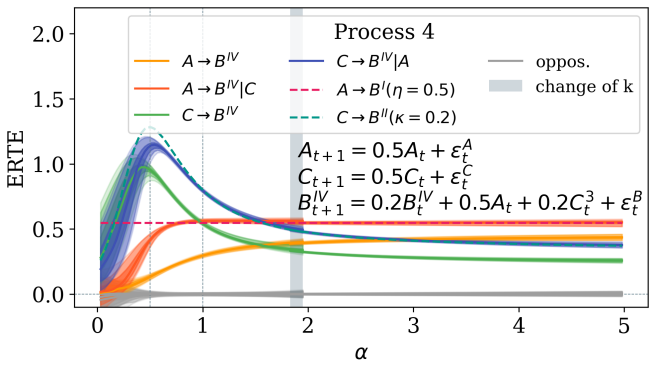
在图6a中，青绿色、红色和黄色曲线（从下到上）分别表示在耦合强度 C 到 B^{II}、 $\kappa = 0.2$、0.5 和 0.8 时，从 $\alpha > 0.7$到 α 的有效RTE（约在 0.6 以下）分别为 $\kappa = 0.2$和 $\kappa = 0.8$。如预期，估算的信息传递随耦合强度增大而增强，其中α R´ényi 测度的低值´对尾部事件中的非线性耦合尤为敏感。该区域对应于底层过程 B^{II} 的 Rényi 熵定义不明确。由于 C存在三次依赖， B^{II} 的分布呈现重尾——正如前文所示，这可能导



图

7 *Process 3: Effective Rényi transfer entropy in both directions between A or D and B^{III}.* ：

- 轴显示 Rényi 参数 α，而 y- 轴显示两个过程之间估计得到的和解析的 R´enyi 转移熵值，单位为比特。对于 $k = 3$，使用值 $\alpha < 1.08$，而在区间其余部分使用 $k = 50$；转变处以灰色竖带标示。记忆参数 $r = l = 1$ 由过程定义确定。分位数带（第 1–99、第 5–95 和第 25–75 百分位，线条表示 30 次运行的中位数）对应于因果方向上估计的有效传递熵。由于涉及两个源过程，还计算了条件 Rényi 转移熵。青色和橙色分位数带分别以约 0。05 和 0。47 比特为中心，表示来自 A（$\eta_1 = 0.2$）和 D（$\eta_2 = 0.5$）的表观 Rényi 转移熵。两种估计值均比图 5a 所示的理论值低约 0。08 比特。然而，以另一个源过程为条件时，估计值与预期结果高度吻合：浅蓝色和红色分位数带分别以 0。13 和 0。55 比特为中心，与虚线所示的理论预测重叠。正如预期，反向的传递熵估计为零。



图

8 *Process 4: Effective Rényi transfer entropy in both directions between A or C and B^{IV}.* ：

横轴表示 Rényi 参数 α，而 y 轴展示两个过程之间以比特为单位的 R´enyi 转移熵估计值。在区间其余部分，$\alpha < 1.98$ 和 $k = 50$ 采用 $k = 3$；其切换位置以灰色竖带标示。记忆参数 $r = l = 1$ 由过程定义确定。分位数带——第 1–99、第 5–95 和第 25–75 百分位数带，并以一条线标出 30 次运行的中位数——对应因果方向上的有效传递熵估计值。由于涉及两个源过程，还计算了条件 RTE。估计得到的条件有效 RTE 值 $T_{\text{eff}\alpha,\ A\to B^{IV}|C}$ 和 $T_{\alpha}^{\text{eff}}{}_{,\ C\to B^{IV}|A}$ 分别以红色和深蓝色表示，与根据先前过程推导的预期值吻合良好；相应的虚线参考线标出了 $\alpha > 0.7$。较小 α 值处出现的偏离，归因于重尾分布下 Rényi 熵定义不明确。相比之下，以橙色和绿色曲线表示的表观 RTE——$T_{\alpha}^{\text{eff}}{}_{,\ A\to B^{IV}}$ 和 $T_{\text{eff}\alpha,\ C\to B^{IV}}$——呈现出较低的信息传递水平，反映出未进行条件化。反向上的估计值仍接近于零，符合预期，但在较小 α 时观察到更大的变异性。这些发现表明，在存在多个源的情况下，可以可靠地提取 $\alpha > 0.7$ 对应的完全 RTE。

致小 α 时 RE 未定义。值得注意的是，三种情况下的有效 Rényi 转移熵都会在 $\alpha \to 0$ 时收敛至零，表明不存在信息传递

与第一个示例一样，此处观察到的大偏差主要是由于估计量中使用了较小的 k。比较图(a)、(c)和(e)（使用 $k = 50$）与图(b)、(d)和(f)（使用 $k = 3$）中的结果，见图15，证实较大的 k对应的方差明显更低。此处列出较小 k-NN 的结果，是因为如图2所示，较小的 k值在较小α时可提供更准确的熵估计值。图15 中的结果表明，在低α 区间进行偏差校正十分重要，这也有助于进一步降低方差。在此区间内，因果方向和非因果方向上的 RTE 均表现出显著偏差，且偏差随耦合增强而增大。为校正两个方向上的偏差，使用过程 A的打乱版本计算 ERTE，因为该过程对估计误差的贡献最大。

为进一步证明可靠性条件得到满足，图6b展示了估计熵的组成部分 $\mathscr{I}_1$、$\mathscr{I}_2$、$\mathscr{I}_3$、$\mathscr{I}_4$、$\mathscr{I}_{12}$和 $\mathscr{I}_{34}$，这些部分构成了从 C 到 B^{II} 的有效 Rényi 转移熵计算，其中 $\kappa = 0.5$。这些组成部分依次以橙色、洋红色、蓝色、绿色、青色和黄色表示。正如图中所示，所有必要的可靠性条件均已满足。

C. 过程 3：双源线性耦合

第三个过程通过引入一个额外的源变量 D 扩展了第一个过程，该变量通过第二个耦合系数 η_2 与目标过程 B^{III} 线性耦合，即

$$\begin{aligned} A_{t+1} &= aA_t + \varepsilon_t^A, \\ D_{t+1} &= aD_t + \varepsilon_t^D, \\ B_{t+1}^{III} &= bB_t^{III} + \eta_1 A_t + \eta_2 D_t + \varepsilon_t^{B^{III}}. \end{aligned} \tag{11}$$

所有模型参数（包括噪声项 ε_t^A，$\varepsilon_t^D \sim \mathcal{N}(0,\ 1)$，以及 $\varepsilon_t^{B^{III}} \sim \mathcal{N}(0,\ 0.25)$）均与先前实验保持一致。仅改变耦合强度 η_1 和 η_2。过程记忆参数按设计$r = l = 1$，因此预期反向不会发生信息传递。

以下是 $\eta_1 = 0.2$ 和 $\eta_2 = 0.5$ 的结果，见图7。采用与前述案例相同的模拟方案。鉴于存在多个源变量，我们还估计了条件 Rényi 转移熵´，T_{α}^{CR}以考虑

源之间的共享信息。从 A 到 B^{III}的条件传递熵（与 $\eta_1 = 0.2$, $T_{\alpha}^{\mathrm{eff}}{}_{,\ A\to B^{III}|D}$(1，1，1) 耦合）以浅蓝色线条及其分位数带表示。结果与第一个过程的结果非常接近，耦合强度相同 $\eta = 0.2$ ，以深蓝色虚线表示，为 0。13 比特。相比之下，青色线条所示的表观传递熵 $T_{\mathrm{eff}\alpha,\ A\to B^{III}}$(1，1) 低估了实际传递量，结果约为 0。05 比特；由于未对 D进行条件化，因果影响约少计了 0。08 比特。这表明，对于给定的耦合设置，从 A 到 B^{III} 的完全 Rényi 转移熵可由记忆参数为 $r = l = s = 1$ 的条件 Rényi 转移熵准确捕获，即 $\bar{T}_{\alpha\ ,\ A\to B^{III}}^{R} = T_{\alpha}^{CR}{}_{,\ A\to B^{III}|D}$(1，1，1)。

对第二个源也观察到了类似的效应。从 D 到 B^{III} 、在$\eta_2 = 0.5$, $T_{\alpha}^{\mathrm{eff}}{}_{,\ D\to B^{III}|A}$(1、1、1) 条件下的条件传递熵，以带有分位数带的红线表示，与过程 1 的理论结果一致： $\eta = 0.5$ 为 0.55 比特（红色虚线）。相应的表观传递熵 $T_{\mathrm{eff}\alpha,\ D\to B^{III}}$(1、1) 以橙色表示，其估计值再次比真实值低约 0.08 比特。这再次表明，从 D 到 B^{III} 的完全 Rényi 转移熵，能够由记忆参数为 $r = l = s = 1$ 的条件 Rényi 转移熵正确表示，即 $\bar{T}_{\alpha\ ,\ D\to B^{III}}^{R} = T_{\alpha}^{CR}{}_{,\ D\to B^{III}|A}$(1、1、1)。

由于该过程涉及线性动力学和高斯噪声， B^{III} 的联合分布仍为高斯分布。因此，从理论上说，真实传递熵在 α 的所有取值下均保持恒定，见式B17。 $\alpha < 0.5$ 的轻微偏差归因于根据打乱的时间序列计算基线时产生的估计噪声。尽管如此，中位数估计值仍可靠地收敛于预期的理论值。图16 展示了在整个 $k = 3$ 和 $k = 50$ 区间内计算得到的 RTE 和有效 Rényi 转移熵估计值。与第一个过程中的观察结果相同，使用 $k = 3$时偏差更为明显。

D. 过程 4：两个源的非线性耦合

最终过程结合了前述系统的特征，同时纳入来自两个源过程的线性和非线性耦合。具体而言，目标过程 B^{IV} 受 A 的线性影响以及 C的非线性（三次）影响，具体如下列方程所述：

$$\begin{aligned} A_{t+1} &= aA_t + \varepsilon_t^A, \\ C_{t+1} &= cC_t + \varepsilon_t^C, \\ B_{t+1}^{IV} &= bB_t^{IV} + \eta A_t + \kappa C_t^3 + \varepsilon_t^{B^{IV}}. \end{aligned} \tag{12}$$

所有模型系数和模拟方案均与先前实验所用的一致。该配置的结果见图8。从 A 到 A （线性耦合强度为 $\eta = 0.5$）以及从 C （立方耦合强度为 $\kappa = 0.2$）到 B^{IV}, $T_{\alpha}^{\mathrm{eff}}{}_{,\ A\to B^{IV}|C}$(1，1，1) 以及 $T_{\alpha}^{\mathrm{eff}}{}_{,\ C\to B^{IV}|A}$(1，1，1) 的条件有效 R $\alpha > 0.7$ 分别以红色和深蓝色分位数线表示。这些估计值与根据先前过程得到的预期值一致，正如相应的虚线参考线所示；尤其是对于 $\alpha > 0.7$而言。这再次表明，在这种耦合类型中，完全 Rényi 转移熵可以用条件 Rényi 转移熵近似，即 $\bar{T}_{\alpha\ ,\ A\to B^{IV}}^{R} = T_{\alpha}^{CR}{}_{,\ A\to B^{IV}|C}$(1，1，1) 以及 $\bar{T}_{\alpha\ ,\ C\to B^{IV}}^{R} = T_{\alpha}^{CR}{}_{,\ C\to B^{IV}|A}$(1，1，1)。较小 α 值下观察到的偏差，源于重尾分布存在时 Rényi 熵定义不明确。

在不进行条件化时，所得的表观转移熵 $T_{\alpha}^{\mathrm{eff}}{}_{,\ A\to B^{IV}}$(1、1) 和$T_{\alpha}^{\mathrm{eff}}{}_{,\ C\to B^{IV}}$(1、1)分别以橙色和绿色显示，低估了总信息传递。在反向传递方向上，ERTE——无论是否进行条件化——始终为零（见图 17）。

E. 相对新增解释

为了评估所得转移熵的解释能力，我们采用在 [15] 中提出的相对新增解释（REA）度量：

$$\mathcal{R}_{\alpha,Y\to X|Z}(r,l,s) \equiv \frac{T_{\alpha,Y\to X|Z}^{R}(r,l,s)}{H_{\alpha}(x_{t+1} \mid x_t^{(r)}, z_t^{(s)})}, \tag{13}$$

其中，分母等于我们也记作 $\mathscr{I}_{12}$的量。 $\mathcal{R}$ 量化了目标过程 X 中的不确定性比例：通过纳入来自源过程 Y的信息，在目标过程 X 的过去以及其他源Z（如有）已能解释的基础上，进一步减少了这部分不确定性。鉴于 $\mathscr{I}_{12}$ 和 $\mathscr{I}_{34}$ 均为非负值（符合可靠性条件），值 $\mathcal{R} \in [0,1]$ 表示该源降低了目标中的不确定性，其中 $\mathcal{R} \times 100\%$表示相对减少幅度。值 $\mathcal{R} < 0$则意味着纳入该源会增加不确定性。

对 $\mathcal{R}$ 的估计还需要处理有限样本量引入的偏差。虽然有效 Rényi 传递熵 $T_{\alpha}^{\mathrm{eff}}$在分子中考虑了此类偏差，但分母仍需进行校正——具体来说，是对项 $\hat{}\, H_{\alpha}($ $x_{t+1} \mid x_t^{(r)}$, $z_t^{(s)}$)进行校正。为估计这一偏差，我们假设 x_{t+1}与经过打乱的过去状态 $x^{(r)}{}_{,\ t}^{\ \mathrm{sh}}$ 和 $z^{(s)}{}_{,\ t}^{\ \mathrm{sh}}$ 在统计上相互独立。在这一假设下，条件熵应等于边缘熵，即 $\hat{}\, H_{\alpha}(x_{t+1} \mid x^{(r)}{}_{,\ t}^{\ \mathrm{sh}}, z^{(s)}{}_{,\ t}^{\ \mathrm{sh}}) = \hat{}H_{\alpha}($ $x_{t+1})$。任何偏离这一等式的情况都完全源于估计量偏差。我们将此偏差定义为：

$$H_{12}^{\mathrm{bias}} \equiv \hat{H}_{\alpha}\big(x_{t+1} \mid x_t^{(r),\mathrm{sh}}, z_t^{(s),\mathrm{sh}}\big) - \hat{H}_{\alpha}(x_{t+1}), \tag{14}$$

将此修正纳入相对解释增益后，得到：

$$\widehat{\mathcal{R}}_{\alpha,Y\to X|Z}(r,l,s) \equiv \frac{T^{\text{eff}}_{\alpha,Y\to X|Z}(r,l,s)}{\hat{H}_\alpha(x_{t+1}\mid x_t^{(r)},z_t^{(s)}) - H_{12}^{\text{bias}}}. \quad (15)$$

表 I 列出了沿耦合方向对所有考察过程计算得到的相对解释增益值，以百分比表示。对于每对过程，$\mathcal{R}$ 在 $\alpha = 0.2$、0.5、0.8、1、2、3、4 和 5 时使用 $k = 50$ 进行计算，结果取 30 次独立模拟运行的平均值。作为比较，采用 $k = 3$ 在较低 α 下得到的值列于括号中。

除一种耦合类型（$C \xrightarrow{\kappa=0.2} B$）外，表中可以看到一种显著的上升 $\mathcal{R}$趋势。这一观察很重要：尽管仅凭ERTE估计值可能会认为，分布尾部的信息传递更强或至少相当，但 $\mathcal{R}$ 表明这种印象并不准确。事实上，尾部的不确定性更高，从而抬高了 ERTE 的绝对值。 $\mathcal{R}$作为一种相对测度，对此进行了校正，并显示与均值附近的区域相比，尾部的信息传递实际上更弱。

对于较弱的非线性（三次）耦合，具体而言，在相互作用 $C \xrightarrow{\kappa=0.2} B$ 的情况下（见第 4、13 和 14 行），$\mathcal{R}$ 会增加至 $\alpha = 1$，之后略有下降或保持不变。这一行为表明，较弱的三次项 κC^3 的影响会受到噪声分量 ε^B 或过程 A 在较高 α 值下的削弱；此时占主导地位的高概率事件受到强调。不过，与较低 α时观察到的较高 $\mathcal{R}$ 相比，其贡献在尾部区域仍更为显著。总体而言，表I 中的观察趋势是实验设计的结果，在该设计中，耦合贡献在所有事件中都大致相当。也可以构建其他实验方案，让耦合强度主要影响低概率或高概率事件，例如在 [19] 中。

REA 值在共享相同源和耦合类型的过程之间也保持一致。表中的符号◇、 ⋆ 和 • 标记了不同过程的完整传递熵估计值一致的情形，这表明信息流得到了正确识别（在估计量可靠的区间内）。例如， ⋆符号表示由源过程A （或 $D \equiv A$）通过线性耦合 $\eta = 0.5$驱动的信息流。如图5a、7和8所示，这种配置产生的传递熵约为 0。55 比特。相应地，所有三个标记为$\mathcal{R}$ 的案例，其 ⋆值在 $\alpha \geq 0.8$处均一致，证实了绝对和相对信息传递度量的一致性。

V. 讨论

本文评估了 Leonenko 等人 [30], 最初提出的用于估计 Rényi 熵的 k-NN 估计量的性能，并将其应用扩展到 RTE。我们的分析分为两个阶段。首先，我们使用从多元正态分布和具有不同尾部厚度的多元 t-分布中抽取的合成数据，评估该估计量的准确性。其次，我们展示了该方法在一组模拟耦合随机过程中的适用性。结果表明，Rényi 熵估计的准确性取决于样本量 N、维数 d（通过记忆体现）以及 k-NN 的选择。估计性能也会随 α-参数而变化，具体取决于上述参数的组合。

对于多元正态分布——各维数的边缘分布相同——当 $\alpha > 1$ 为 N =2000 个样本时，估计值就已与解析值非常接近。随着 α 减小，需要更大的样本量： $\alpha \geq 0.75$ 为 $N > 10^4$时、 $0.5 \leq \alpha < 0.75$ 为 $N > 10^5$时，以及 $0.4 \leq \alpha < 0.5$ 为 $N = 10^6$时，都能获得可靠的估计。对于更小的 α，则需要更大的样本量才能提高准确性。然而，由于该估计量是针对概率分布函数定义的

表 I: *Relative Explanation Added for All Coupling Types.*表中数值表示每个过程进行 30 次模拟运行后得到的均值 $\mathcal{R}$ ，使用公式（15）并结合 $k = 50$计算，以百分比表示（对于 $k = 3$，小 α下的结果列于括号内）。各列对应不同的 Rényi 参数 R´，分别为 $\alpha = 0.2$、0.5、0.8、1、2、3、4 和 5。每行报告特定源–目标过程对的$\mathcal{R}$。例如，第一行显示，过程 A 使关于 B^I 的不确定性降低约 10%（当 $\alpha = 0.5$ 时）和 18%（当 $\alpha = 5$时）。符号◇、 ⋆ 和 • 表示完全传递熵估计值在不同过程中一致的情况，这表明信息流得到了正确提取。具有相同符号的 $\mathcal{R}$ 值在 1–3% 的平均误差界内一致。偏差主要出现在 α较小时，此时估计量的可靠性较低。不过，对于 $\alpha \geq 1$，在耦合情形相近的情况下，结果高度一致。

		0.2	0.5	0.8	1 2 3 4 5
$A \xrightarrow{\eta=0.2} B^I$	◇	8(8)	9(10)	11(11)	11 14 16 17 18
$A \xrightarrow{\eta=0.5} B^I$	⋆	26(26)	30(30)	33(33)	35 40 43 45 47
$A \xrightarrow{\eta=0.8} B^I$		37(38)	43(43)	46(47)	48 53 57 59 60
$C \xrightarrow{\kappa=0.2} B^{II}$	•	9(15)	32(43)	41(46)	41 36 36 37 38
$C \xrightarrow{\kappa=0.5} B^{II}$		2(4)	21(36)	46(56)	51 51 51 51 52
$C \xrightarrow{\kappa=0.8} B^{II}$		0(2)	15(30)	43(58)	52 56 56 57 58
$A \xrightarrow{\eta_1=0.2} B^{III}$		2(2)	3(3)	3(3)	3 4 4 4 4
$A \xrightarrow{\eta_1=0.2} B^{III}\|D$	◇	7(8)	9(10)	11(11)	12 14 16 17 18
$D \xrightarrow{\eta_2=0.5} B^{III}$		21(21)	25(25)	27(27)	29 33 35 37 38
$D \xrightarrow{\eta_2=0.5} B^{III}\|A$	⋆	25(25)	30(30)	33(33)	35 39 42 44 46
$A \xrightarrow{\eta=0.5} B^{IV}$		1(1)	4(4)	10(10)	14 23 27 29 31
$A \xrightarrow{\eta=0.5} B^{IV}\|C$	⋆	1(2)	13(19)	28(32)	32 39 43 45 47
$C \xrightarrow{\kappa=0.2} B^{IV}$		8(14)	22(30)	25(28)	23 19 18 18 18
$C \xrightarrow{\kappa=0.2} B^{IV}\|A$	•	4(11)	25(38)	37(44)	38 36 36 37 38

对于有界支撑的分布，无界支撑的正态分布在任何情况下都无法在极小的 α下达到足够的精度。这些发现表明，使用较小的 α 时，必须谨慎考虑样本量，因为数据有限会降低估计精度；同时也应为 α设定下限。不过，在RTE估计中，可以通过计算有效RTE，或采用替代数据方法 [41, 42] 来确定 RTE 估计值的统计显著性，从而减轻有限样本偏差的影响。

影响 RE 和 RTE 估计准确性的另一个关键因素是输入数据的维数 d。对于均匀分布的多元数据，随着 d 增大，估计量的精度会降低，不过当 $d < 6$ 时，结果仍然合理。在 RTE 中，每个熵项的维数（即从 $\mathscr{I}_1$ 到 $\mathscr{I}_4$ ）由记忆参数或马尔可夫阶数参数 r, l 和 s_i 决定；这些参数定义了估计所用时间窗口的长度。通常根据对系统的先验知识选择这些参数，或通过马尔可夫阶数检验进行经验确定。

需要注意的是，过大的记忆参数可能会因时间序列数据中存在较强的自相关而引入冗余，从而降低估计器性能。在这种情况下，子集化或时间平移等数据预处理技术可能会有所帮助。在适当情况下，可采用主成分分析（PCA）等降维方法提取信息量最高的特征，从而在保留所研究过程底层结构的同时降低复杂度[43, 44]。

影响估计过程的另一个关键参数是最近邻的数量 k，它充当估计量的参数。从概念上讲，该参数决定了如何估计相空间不同区域内的概率密度。在人口稀疏的区域，较小的 k 值能够带来更局部、因而更准确的估计；而在人口密集的区域，则更适合使用较大的 k 值，以避免对噪声过拟合。

在针对不同 R´enyi 参数 α 估计 Rényi 熵时，这种敏感性尤为重要。对于 $\alpha < 1$，Rényi 熵会更加重视罕见事件或低概率事件，从而有效放大尾部行为的重要性。在这种情况下，较小的 k 值更为合适，因为它们能够更好地捕捉稀疏区域中的局部变化。相反，对于 $\alpha \geq 1$ ，建议使用较大的k 值，以确保对在熵计算中占主导地位的高概率区域进行稳定估计。

此外， k 的选择受到估计量中所用伽马函数定义域所决定的数学约束。具体而言，只有当 $\Gamma(k+1-\alpha)$有限时，估计量才有明确定义。这对 k的允许取值设定了下限，尤其是在 α较大时；选择估计器参数时应仔细考虑这一点。

A. 耦合随机过程的洞见

为评估该估计量检测方向性信息传递的能力，我们将其应用于多种因果结构已知的模拟耦合随机过程。这些情形涵盖线性和非线性交互，以及不同的耦合强度。研究结果表明，当记忆参数设定正确时，即使存在多个源，完全 Rényi 转移熵也能可靠地捕捉信息流。表观 RTE（即未考虑额外因果源时的 RTE）往往会低估真实的信息传递，这证实了存在多个源时使用条件 Rényi 转移熵的必要性。

我们的研究结果表明，在大多数耦合类型中，相对新增解释 $\mathcal{R}$会随着 α 的增加而增加。然而，在非线性三次耦合示例中， κ =0.2时， $\mathcal{R}$ 在 $\alpha = 0.8$ 附近达到峰值，并随着 α 增加而进一步下降，这表明尾部事件可能携带比例更高的信息。这种转变说明，不同的 α 区间会突出不同的动力学特征，从而比单独使用基于香农熵的转移熵更全面地展现因果交互。

B. 重尾分布与较小 α带来的挑战

我们的结果还凸显了在重尾分布背景下对较小 α 进行 RTE 估计的局限性。由于 Rényi 熵在 $\alpha < 1$ 取值下会对低概率事件赋予越来越大的权重，估计量对尾部行为变得更加敏感。当底层分布具有无界支撑时，例如多元 Student's t 分布，这一问题尤为突出。随着尾部厚度增加（即自由度 ν 减小），Rényi 熵仍能保持良好定义的 α 范围会变窄。我们发现，对于重尾分布，偏离解析值的现象在接近 $\alpha = 1$时开始出现，并且随着 α 减小而变得更加明显。

尽管如此，即使对于柯西分布这一极端情形（对应于 $\nu = 1$），Rényi 熵对于 $\alpha > 0.75$ 仍然是定义良好的。这为尾部敏感信息度量（如 RTE）提供了一个可用区间，使其仍能得到有意义的应用。因此，尽管在重尾条件下解释小 α 样本量的 RTE 时需要谨慎，该估计量仍可适用于能够捕捉显著尾部行为的范围，尤其是在样本量足够大且k 参数经过适当调节时。

C. 可靠 RTE 估计的实用步骤

本研究完全基于模拟数据，其中基础分布和参数均已完全指定。相比之下，从经验联合分布中估计 RE 和 RTE——尤其是涉及异质相关边缘分布的情况——面临着重大挑战。

与按定义非负的香农转移熵不同，RTE 可以取负值。这一特性使我们必须区分两种情况：负值反映了所研究系统的有意义属性，还是源于估计偏差或方法局限。此外，负分量熵可能导致 RTE 假阳性。为解决这一歧义，我们提出三项可靠性条件；只有满足这些条件，所得 RTE 估计值才可视为具有可解释性。

这些条件要求所有估计的熵分量 $\mathscr{I}_1, \mathscr{I}_2, \mathscr{I}_3, \mathscr{I}_4, \mathscr{I}_{12}, \mathscr{I}_{34}$ 均为非负，且相应的香农转移熵也必须非负。这些标准基于离散 Rényi 熵、条件 Rényi 熵以及基于 Shannon 的传递熵的基本性质，可作为最低限度的诊断检查，以确保结果具有可解释性。

鉴于影响估计准确性的因素众多，我们提出一套实用步骤，帮助校准参数，从而提高满足可靠性条件的可能性。这些建议直接源自本文对模拟过程所做的系统分析；在这些过程中，真实值已知，且可以对估计量的表现进行全面评估。尽管这些步骤不能保证成功，但它们提供了一种结构化方法，用于指导参数选择，并评估结果是否可信，或是否需要进一步进行数据预处理。

- 从最小记忆开始：为所有其他源设置 $r = 1$、$l = 1$ 和 $s_i = 1$。
- 当 α 较小时，使用较小的 k（例如 k =1，2，3，4，5），以更好地捕捉 Rényi 熵所强调的尾部事件。
- 当 α 较大时，增大 k，以稳定高概率区域中的估计量。
- 如果 $\alpha \geq 1$ 对应的 Rényi 熵为负值，应显著增大 k，例如 $k = 100$，500，.....。这样可以降低极短距离的影响，因为极短距离可能使对数内部的参数小于 1，从而得到负值。
- 若 Rényi 熵为负，可考虑增大记忆参数，因为 Rényi 熵往往会随维数增加而增大（见附录 C）。改变源过程的记忆参数会影响 $\mathscr{I}_3$ 和 $\mathscr{I}_4$，而改变目标过程的记忆参数则会影响所有分量熵。
- 有限样本量导致的估计偏差在高维性或重尾分布的情况下尤为显著。计算有效RTE时，务必对导致这些偏差最大的过程进行随机打乱。
- 注意：如果上述步骤未能奏效，可考虑对数据进行预处理（例如滤波、归一化、时间序列移位），或纳入其他相关源，以便更好地考虑共享信息结构。

这些指南旨在帮助在实践中使用 RTE 的 k-NN 估计量，尤其是在模型参数和分布未知的实证应用中。它们还强调了诊断验证和谨慎进行参数调优的重要性，以便从复杂系统的信息论度量中得出有意义的结论。

VI. 结论

在本研究中，我们使用通过 k近邻方法估计的 RTE，研究了双变量信息流。首先，我们分析了三个关键因素——样本量 N、数据维度 d 和近邻参数 k ——如何影响一系列 R´ényi 熵参数值 α 下的 R´ényi 熵估计。随后，我们将该估计量应用于四个模拟的耦合随机过程，以量化其中逐步增强的影响复杂度所体现的信息流。

我们的分析表明，在 α取值较小时，RE估计最具挑战性。这种困难有两个原因。首先，较低的 α 值会使分布的尾部获得更大权重，而尾部反映了通常采样不足的极端事件，因此会降低估计器精度。其次，当基础分布具有无界支撑时，尽管这一影响在大多数实证应用中可以忽略不计。总体而言，我们的结果表明，在较小的 α 下，较小的 k会带来更可靠的估计；而对于 $\alpha > 1$，较大的 k 值则能提高准确性。正如预期的那样，有限样本偏差也会影响估计结果，但可以通过增加样本量 N 来减小；如果无法做到这一点，则可采用基于置乱的校正。为促进一致且可解释的 RTE估计，我们根据 RE 的基本性质提出了三个可靠性条件。

本研究迈出了探索此前未涉足领域的第一步：在 Rényi参数 α 的广阔连续范围内，估计模拟耦合过程的 R´ényi 传递熵。通过四个简单示例，我们观察到信息流对极端事件与频繁事件的贡献方式存在变异性。这种信息传递的特征取决于耦合强度以及相互作用的函数形式。此外，当一个目标过程受到

多个源过程的影响时，所产生的相互作用会变得更加复杂，因此必须在计算中纳入所有源。

我们的研究结果凸显了利用灵活的 k-NN 方法提取详细信息流所涉及的微妙之处。通过谨慎进行参数调优，在估计器精度与保留底层数据结构之间取得平衡，我们证明了 RTE 框架能够准确捕捉线性和非线性耦合系统中的因果交互。此外，该框架在稳健识别驱动极端事件的因果机制方面也展现出巨大潜力。

最后，所得结果还为未来研究指明了若干方向。其中一个方向是开发改进的估计方法，尤其适用于构型空间具有分形维数的退化过程。另一个方向是基于已知的底层动力学，构建具有多个源或复杂耦合机制的更复杂耦合系统。此类系统可以设计为有选择地影响分布的特定方面——例如尾部或均值——从而深入了解信息在复杂系统中的传递、共享或相互作用。最后，我们的研究结果为将基于 R´ényi 的信息论分析拓展至多个领域的实证应用奠定了基础，包括气候科学、医学和金融。

VII. 作者贡献

Z.T.、P.J.、H.L.：撰写初稿、方法学、概念构思。P.J.、M.P.：审阅与编辑、监督。Z.T.、H.L.：软件、调查、正式分析。

VIII. 关于在手稿准备过程中使用生成式 AI 和 AI 辅助技术的声明

在准备本研究成果期间，作者使用了 ChatGPT，以提升文本的清晰度和可读性、纠正语法问题，并协助查找相关文献。使用该工具/服务后，作者根据需要审阅并编辑了内容，并对已发表文章的内容承担全部责任。

IX. 代码和数据可用性

所有用于模拟和数据处理的代码均可在 GitHub 上获取：

https://github.com/zlatataa/OnPracticalEstimationofRenyiTransferEntropy

致谢

感谢 Stefan Thurner 提供有益的讨论。P.J.、H.L.、Z.T. 和 M.P. 得到了捷克科学基金会（GAˇCR）资助，项目编号为 25- 18105S。M.P. 还得到了扬·阿莫斯·夸美纽斯计划（P JAC）的支持，项目编号为 $CZ.02.01.01/00/22\text{-}008/0004605$，自然与人为地质灾害。

附录 A：术语表

- **通用符号**

  $\alpha \in \mathbb{R}_0^+$：Rényi 参数$r$，$l$，$s \in \mathbb{N}$：马尔可夫阶数（记忆参数）$N \in \mathbb{N}$：样本量$d \in \mathbb{N}$：维数$k \in \mathbb{N}$：近邻阶数 $\rho_{j,\,i} \in \mathbb{R}$：数据点之间的欧几里得距离 j 和 $i\mathcal{P}_{\mathcal{S}}$：学生 t 分布$\nu$：学生 t 分布的自由度

- **随机过程**

  $X \equiv \{x_1, x_2, \ldots, x_N\}$：目标过程
  $Y \equiv \{y_1, y_2, \ldots, y_N\}$：源过程
  $Z \equiv \{z_1, z_2, \ldots, z_N\}$：条件过程$y_t^{(l)}$：$l$ 个过去观测值的窗口，来自 Y，即y_t，y_{t-1}，……，$y_{t-l+1}Y_{\text{sh}}$：Y的打乱版本$x^{(r),\,t}{}^{\text{sh}}$：$X$ 的打乱历史$z^{(s),\,t}{}^{\text{sh}}$：$Z$ 的打乱历史

- **信息论量**

  $T^R_{\alpha,Y\to X}(r,l)$：从 Y 到 X 的 Rényi 传递熵 R´$T^{CR}_{\alpha,Y\to X|Z}(r,l,s)$：条件传递熵$\bar{T}^R_{\alpha,Y\to X}$：完全传递熵$I_\alpha(x_{t+1} : y_t^{(l)}|x_t^{(r)})$：互信息$H_\alpha(x_{t+1}|x_t^{(r)})$：条件 Rényi 熵 ˆ´$H_\alpha(N,k,\alpha,d)$：Rényi 熵估计值´$T$ eff$_{\alpha,Y\to X}(r,l)$：有效 Rényi 传递熵$\mathcal{R}_{\alpha,Y\to X|Z}(r,l,s)$：相对新增解释$H_{12}^{\text{bias}}$：ˆ $\mathscr{I}_{12}$ 的估计偏差

- **RTE 组成部分**

$\mathscr{I}_1$: $H_\alpha(x_{t+1}, x_t^{(r)}, z_t^{(s)})$
$\mathscr{I}_2$: $H_\alpha(x_t^{(r)}, z_t^{(s)})$
$\mathscr{I}_3$: $H_\alpha(x_{t+1}, x_t^{(r)}, z_t^{(s)}, y_t^{(l)})$
$\mathscr{I}_4$: $H_\alpha(x_t^{(r)}, z_t^{(s)}, y_t^{(l)})$
$\mathscr{I}_{12}$: $\mathscr{I}_1 - \mathscr{I}_2$
$\mathscr{I}_{34}$: $\mathscr{I}_3 - \mathscr{I}_4$

- **模拟过程与参数**

A, C, D: Source (driving) AR(1) processes
$B^I, B^{II}, B^{III}, B^{IV}$: Target (driven) processes
$\eta, \eta_1, \eta_2, \kappa \in \{0.2, 0.5, 0.8\}$: Coupling coefficients
$a = c = 0.5,\ b = 0.2$: AR parameters
$\varepsilon_t^A \sim \mathcal{N}(0,1)$: Noise in A
$\varepsilon_t^{B^I} \sim \mathcal{N}(0, 0.25)$: Noise in B^I
$\varepsilon_t^{B^{II}} \sim \mathcal{N}(0, 0.25)$: Noise in B^{II}
$\varepsilon_t^C \sim \mathcal{N}(0,1)$: Noise in C
$\varepsilon_t^D \sim \mathcal{N}(0,1)$: Noise in D
$\varepsilon_t^{B^{III}} \sim \mathcal{N}(0, 0.25)$: Noise in B^{III}

附录 B：信息论中的一些相关概念——香农熵

考虑一个离散随机变量 X，其取值属于有限集 $\mathfrak{X}$，概率分布函数（PDF）为 $\mathcal{P}_X$。为简便起见，对于实现值p（x），我们使用这一写法，而不采用更精确的记法 p_X（x）。香农熵 H（X）定义为 X： [45]:

$$H(X) \equiv H(\mathcal{P}_X) = -\sum_{x\in\mathfrak{X}} p(x)\log_2 p(x)\,. \tag{B1}$$

现在，考虑另一个随机变量 Y，其取值属于集合 Υ，边缘概率密度函数为 $p(y)$，定义在 $\mathfrak{X}\times\Upsilon$ 上的联合概率密度函数为 $p(x,\ y)$。联合熵 $H(X,\ Y)$定义为

$$H(X,Y) = -\sum_{x\in\mathfrak{X}}\sum_{y\in\Upsilon} p(x,y)\log_2 p(x,y)\,. \tag{B2}$$

此定义可自然推广到 n 个离散变量的联合熵。

条件熵 $H(Y|X)$量化了给定 X 后 Y 中剩余的不确定性，其定义为

$$H(Y|X) = -\sum_{x\in\mathfrak{X}}\sum_{y\in\Upsilon} p(x,y)\log_2 p(y|x)\,, \tag{B3}$$

其中，条件概率 $p(y|x) = p(x,\ y)/p(x)$，假设 $p(x)\neq 0$。

X 与 Y之间的共享信息量由互信息度量，定义为

$$\begin{aligned} I(X;Y) &= \sum_{x\in\mathfrak{X}}\sum_{y\in\Upsilon} p(x,y)\log_2\frac{p(x,y)}{p(x)p(y)} \\ &= H(X) + H(Y) - H(X,Y) \\ &= H(X) - H(X|Y)\,. \end{aligned} \tag{B4}$$

在此情形下，可以利用凹函数的詹森不等式 [，其中 $\log_2($ $\cdots)$]，证明 $I(X;Y)\geq 0$。后者也称为吉布斯不等式。

X 和Y 在给定第三个变量 Z 条件下的条件互信息 (CMI)为

$$I(X;Y|Z) = \sum_{x,y,z} p(x,y,z)\log_2\frac{p(x,y|z)}{p(x|z)p(y|z)}\,. \tag{B5}$$

从该定义可以清楚地看出，当 Z 与 X 和 Y均统计独立时，

$$I(X;Y|Z) = I(X;Y)\,. \tag{B6}$$

经过简单的代数变换，可得以下重要恒等式

$$\begin{aligned} I(X;Y|Z) &= H(X,Z) + H(Y,Z) \\ &- H(X,Y,Z) - H(Z)\,. \end{aligned} \tag{B7}$$

CMI 的恒等式（B7）可推广至*R´enyi* 熵框架。A. R´enyi 在其一系列开创性论文中引入了 Rényi 熵 [24, 25]。Rényi 熵是一种信息论度量，构成了一族单参数熵度量，并对经典香农熵进行了推广。Rényi 熵定义为

$$H_\alpha(X) \equiv H_\alpha(\mathcal{P}_X) = \frac{1}{1-\alpha}\log\left(\sum_{x\in\mathfrak{X}} p^\alpha(x)\right), \tag{B8}$$

其中 $\alpha\in\mathbb{R}_0^+$，当 $\alpha\to 1$时，该式退化为香农熵，即 $\lim_{\alpha\to 1} H_\alpha(X) = H(\{$v6)。

本文针对一维（标量）变量定义了 CMI 及其 Rényi 推广形式。我们在正文中给出的数值实验表明，这种形式足以揭示单向因果关系。关于 CMI 向高维形式的扩展及其在多元时间序列中的应用，参见例如 [11, 12]。

在此基础上，阶数为 α 的互信息和阶数为 α 的条件 Rényi《熵》定义为

分别为:

$$I_\alpha(X;Y) = \frac{1}{1-\alpha}\log_2\frac{\sum_{x\in\mathfrak{X},y\in\Upsilon}p^\alpha(x)p^\alpha(y)}{\sum_{x\in\mathfrak{X},y\in\Upsilon}p^\alpha(x,y)}, \quad \text{(B9)}$$

$$H_\alpha(X|Y) = \frac{1}{1-\alpha}\log_2\frac{\sum_{x\in\mathfrak{X},y\in\Upsilon}p^\alpha(x,y)}{\sum_{y\in\Upsilon}p^\alpha(y)}$$

$$= \frac{1}{1-\alpha}\log_2\frac{\sum_{x\in\mathfrak{X},y\in\Upsilon}p^\alpha(x|y)p^\alpha(y)}{\sum_{y\in\Upsilon}p^\alpha(y)}$$

$$\equiv \frac{1}{1-\alpha}\log_2\Big(\sum_{x\in\mathfrak{X},y\in\Upsilon}p^\alpha(x|y)\rho_\alpha(y)\Big), \quad \text{(B10)}$$

其中

$$\rho_\alpha(y) \equiv \frac{p^\alpha(y)}{\sum_{y'\in\Upsilon}p^\alpha(y')}, \quad \text{(B11)}$$

称为陪伴分布。这种分布在赋予基于 R´enyi 的信息度量（包括 R´enyi 转移熵）所谓的缩放性质方面发挥着关键作用。

上述定义的条件 Rényi 熵具有几个重要性质:

- $H_\alpha(X|Y) \geq 0$ 对所有 $\alpha \in \mathbb{R}_0^+$ 均成立;
- $H_\alpha(X|Y) = H_\alpha(X,Y) - H_\alpha(Y)$;
- 只有当 X 或 Y 服从均匀分布时，才能保证不等式 $H_\alpha(X|Y) \leq H_\alpha(X)$ 成立。

特别是，从最后一点可以看出， X 与 Y之间的 阶互信息，即

$$I_\alpha(X;Y) = H_\alpha(X) - H_\alpha(X|Y), \quad \text{(B12)}$$

在某些情况下可能为负。

1. 多元 t-分布的 Rényi 熵

下面我们推导多元 Student t-分布的 Rényi 熵:

$$\mathcal{P}_S(\mathbf{x}) = \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\Gamma(\frac{\nu}{2})(\nu\pi)^{\frac{n}{2}}\sqrt{|\Sigma|}}\left[1 + \frac{1}{\nu}(\mathbf{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]^{-\frac{\nu+n}{2}}. \quad \text{(B13)}$$

Σ 是一个正定的 $n\times n$ 缩放矩阵， $\boldsymbol{x}$、 $\boldsymbol{\mu}$ 是实 n维向量。令 $\boldsymbol{\mu}=\mathbf{0}$、 $a \equiv \frac{\nu+n}{2}$，为计算积分，我们使用*Schwinger* 技巧，该技巧广泛应用于量子场论中。它基于伽马函数的积分定义，可改写为

$$\frac{1}{A(z)^n} = \frac{1}{\Gamma(n)}\int_0^{+\infty} du\; u^{n-1}\exp(-uA(z)). \quad \text{(B14)}$$

因此，我们有

$$\int_{\mathbb{R}}\frac{d^nx}{\left(1+\frac{1}{\nu}\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{x}\right)^{\alpha a}} = \frac{1}{\Gamma(\alpha a)}\int_{\mathbb{R}}d^nx\int_{\mathbb{R}^+}du\; u^{\alpha a-1}\exp(-u[1+\nu^{-1}\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{x}])$$

$$= \frac{1}{\Gamma(\alpha a)}\int_{\mathbb{R}^+}du\; u^{\alpha a-1}e^{-u}\left(\frac{\nu\pi}{u}\right)^{\frac{n}{2}}\sqrt{|\Sigma|}$$

$$= \frac{\sqrt{|\Sigma|}}{\Gamma(\alpha a)}(\nu\pi)^{\frac{n}{2}}\int_{\mathbb{R}^+}du\; u^{\alpha a-1-\frac{n}{2}}e^{-u}$$

$$= \frac{\sqrt{|\Sigma|}}{\Gamma(\alpha a)}(\nu\pi)^{\frac{n}{2}}\Gamma(\alpha a - n/2), \quad \text{(B15)}$$

其中 ν和 α必须满足 $Re(\alpha\frac{\nu+n}{2}-\frac{n}{2}) > 0$。将此结果与（B13）结合，可得多元 t-分布的 Rényi 熵的显式定义

$$H_\alpha[\mathcal{P}_S] = \frac{1}{1-\alpha}\log_2\left[\frac{\Gamma^\alpha(\frac{\nu+n}{2})\Gamma(\alpha\frac{\nu+n}{2}-\frac{n}{2})}{\Gamma^\alpha(\frac{\nu}{2})\Gamma(\alpha\frac{\nu+n}{2})}|\Sigma|^{\frac{1}{2}(1-\alpha)}(\nu\pi)^{\frac{1}{2}n(1-\alpha)}\right]. \quad \text{(B16)}$$

当 $\nu=1$ 时，学生 t 分布等同于柯西分布；当 $\nu\to\infty$ 时，它收敛于多元正态分布，其 Rényi 熵定义为:

$$H_\alpha[\mathcal{P}_\mathcal{N}] = \log_2\left[(2\pi)^{\frac{n}{2}}|\tilde{\Sigma}|^{\frac{1}{2}}\right] + \frac{1}{1-\alpha}\log_2\left[\alpha^{-\frac{n}{2}}\right], \quad \text{(B17)}$$

˜其中， Σ 是正定协方差矩阵。可以看出，涉及 α 的项仅取决于维数 n。在计算 RTE 时，这些项会相互抵消，因此高斯过程的 RTE 与 α无关。

附录C：估计值的负性

这里，我们考察一种名为过程 1.b 的情形，其中源过程与目标过程强相关。如前所述，由于存在估计偏差和可解释性问题的风险，此类配置需要谨慎处理。该系统采用与过程 1 相同的结构：

$$\begin{aligned} A_{t+1} &= aA_t + \varepsilon_t^A, \\ B_{t+1}^{I.b} &= bB_t^{I.b} + \eta A_t + \varepsilon_t^{B^{I.b}}, \end{aligned} \tag{C1}$$

但噪声属性有所改变。具体而言，我们将目标过程中的噪声方差降低至$\varepsilon_t^{B^I} \sim \mathcal{N}(0,\ 0.0001)$，同时保持其余参数不变： $a = 0.5$， $b = 0.2$， $\varepsilon_t^A \sim \mathcal{N}(0,\ 1)$。我们关注耦合强度为 $\eta = 0.5$的情形。由于目标过程的噪声水平极低，B^I 几乎成为源的确定性函数，从而导致结构退化。因此，联合分布的协方差矩阵不再是正定的，使得解析解B17 不适用。

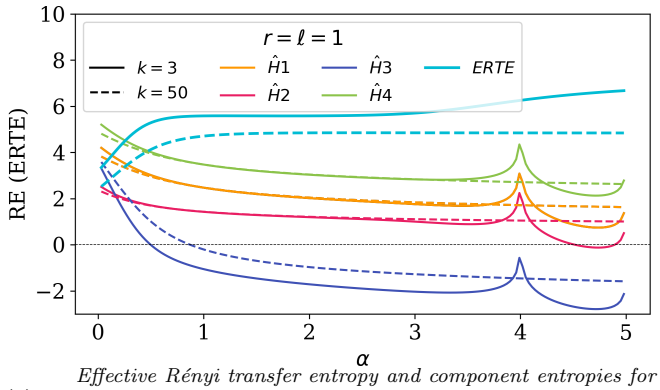
ˆ ˆ ˆ ˆ 图9a 展示了过程 1.B 的 Rényi 熵估计值 R´，以橙色、洋红色、深蓝色和浅绿色分别绘出 $\mathscr{I}_1$， $\mathscr{I}_2$， $\mathscr{I}_3$，以及 $\mathscr{I}_4$。这些估计基于 $N = 10^5$ 个样本，记忆参数为$r = l = 1$，并分别采用 k近邻参数 $k = 3$ （实线）和 $k = 50$ （虚线）计算。对应的因果方向上的有效 Rényi 转移熵（ERTE）以青色显示，并通过线型区分´的两个取值。 k。

ˆ 如图所示，并非所有可靠性条件都得到满足——最明显的是， $\mathscr{I}_3$ 为负。相比之下，反向（图9b）在 $r = l = 1$方面得出了更可靠的结果：所有可靠性条件均得到满足，而且正如预期，有效 Rényi 转移熵约为零，表明该方向上不存在因果影响。

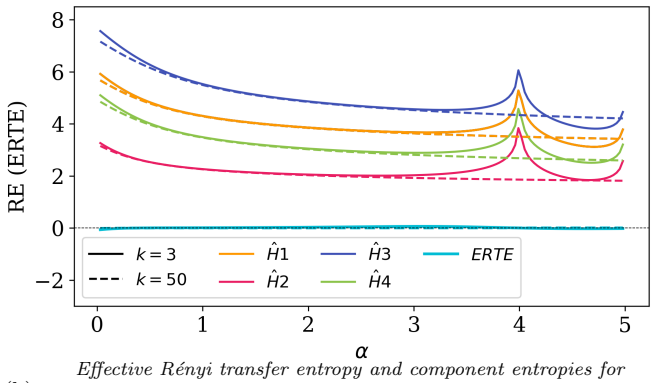
要在因果方向上满足可靠性条件，需要增大最近邻参数 k。源过程与目标过程之间的高相关性会导致相空间中的点聚集得十分紧密，使对数项中的参数值变小，进而产生负的熵估计值。此外，增大记忆参数 ˆ ˆ 也有助于降低这些相关性。图 10 展示了随着记忆参数 l （针对源过程 A）和 k 增大， $\mathscr{I}_3$ 和 $\mathscr{I}_4$ 的估计值如何变为正值。由于 $\mathscr{I}_1$ 和 $\mathscr{I}_2$ 与源过程相互独立，因此它们保持不变。作为对比，图中还包括 $k = 50$ 和 $l = 1$ 的估计值（如图 11a 所示），以最细的线条标出。将 k 增至 1000，同时保持 $l = 1$ 不变，可使 $\mathscr{I}_3$ 为正（蓝色虚线，自下而上第二条），但其仍低于 $\mathscr{I}_4$，因此 $\mathscr{I}_{34}$ 为负。要实现 $\mathscr{I}_{34} \geq 0$，还需要进一步增大 k 和 l 。最终，当 $k = 7000$ 和 $l = 5$ 取相应值时，该条件得到满足，所有可靠性标准也随之满足。

图 11a 展示了使用图中 Rényi熵计算得到的有效 Rényi 转移熵值。10。所有 ERTE估计值均为非负值，并且随着 k 和 l 的增加而趋于降低。相应的相对解释增益值见图11b。与 ERTE 不同，REA 指标可直接用于诊断假阳性，因为超过 1 的值（由水平虚线表示）违反了可解释性。因此，当 $k = 1000$ 和 $l = 4$ 时，在 $\alpha \leq 1$的情况下，以及在 $k = 7000$ 和 $l = 4$ 的整个 α 范围内，均可获得有效的 ERTE。

观察到的趋势表明，进一步增加 k 或 l 可能会继续降低 ERTE，使其趋近于零，甚至变为负值。这表明，对于特定的参数配置，有可能捕捉到暗示因果关系的定向熵传递。不过，我们将这些发现解释为表观传递熵的体现，其受估计量对参数选择的敏感性影响，而并非完整反映潜在的因果机制。若要准确捕捉完全传递熵——使其更贴近真实的信息流——则需要对该过程进行更深入的研究，包括考察其内在维度和结构；这超出了本研究的范围。



(a) *Effective Rényi transfer entropy and component entropies for the causal direction in Process 1.b.* $x \to y$ —轴表示 Rényi 参数 α，而 y—轴展示估计得到的各分量 Rényi 熵，以及因果方向 $A \to B^{I.b}$ 上由此得到的有效 Rényi 转移熵。估计基于 $N = 10^5$ 个样本，记忆参数为 $r = l = 1$，使用 $k = 3$ （实线）和 $k = 50$ （虚线）。从下到上，各线分别表示：$\hat{\mathscr{I}}_3$ （深蓝色）、$\hat{\mathscr{I}}_2$ （洋红色）、$\hat{\mathscr{I}}_1$ （橙色）、$\hat{\mathscr{I}}_4$ （浅绿色）和有效 Rényi 转移熵（青色）。图示的有效 Rényi 转移熵是假阳性，因为 $\hat{H}3$ 的值为负，违反了可靠性条件。



(b) *Effective Rényi transfer entropy and component entropies for the non-causal direction in Process 1.b.* $x \to y$ —轴表示 Rényi 参数 α，而 y—轴展示估计得到的各分量 Rényi 熵，以及非因果方向 $B^{I.b} \to A$ 上由此得到的有效 Rényi 转移熵。估计基于 $N = 10^5$ 个样本，记忆参数为 $r = l = 1$，使用 $k = 3$ （实线）和 $k = 50$ （虚线）。从下到上，各线分别表示：有效 Rényi 转移熵（接近零的青色线）、$\hat{\mathscr{I}}_2$ （洋红色）、$\hat{\mathscr{I}}_4$（浅绿色）、$\hat{\mathscr{I}}_1$ （橙色）和 $\hat{\mathscr{I}}_3$ （深蓝色）。所有可靠性条件均得到满足，而且与预期一致，有效 Rényi 转移熵为零。

图 9：过程 1.b：强相关性下线性单源耦合的 Rényi 熵估计值与有效传递熵

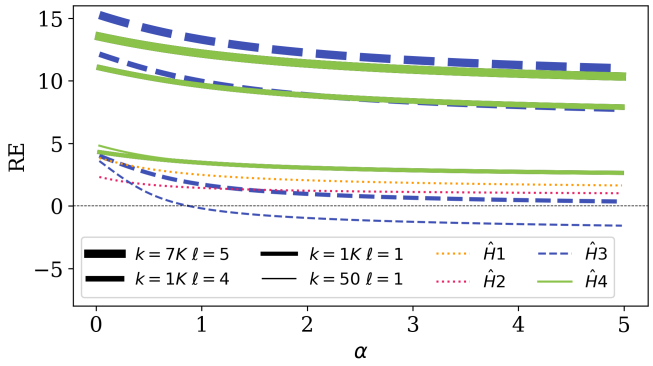
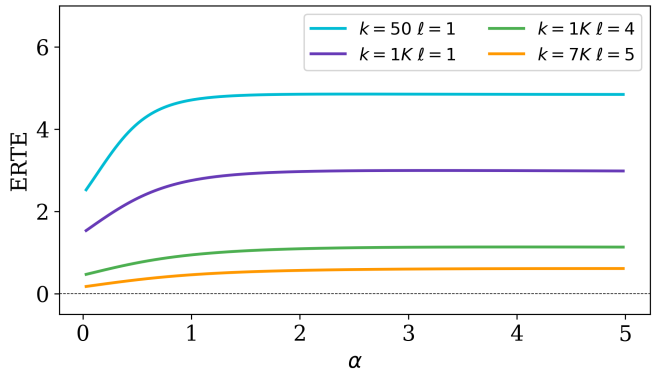
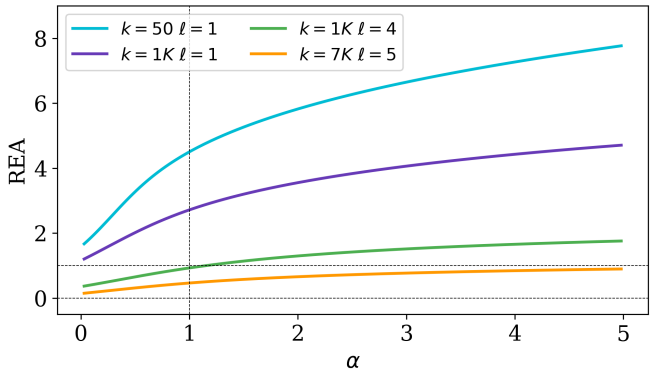


图 10 *Estimates of Rényi entropies $\mathscr{I}_3$ and $\mathscr{I}_4$ under varying k-NN and memory parameters for the causal direction in Process 1.b.* $x \to y$ ：横轴表示 Rényi 参数，纵轴表示因果方向 $A \to B^{I.b}$上从 α 到 y 的各分量 Rényi 熵估计值。所有估计均基于 $N = 10^5$ 个样本。目标记忆参数固定为 $r = 1$，同时改变源记忆参数 l 和 k近邻参数，以评估它们的影响。熵估计值如下所示：$\hat{\mathscr{I}}_1$ （橙色点线）、$\hat{\mathscr{I}}_2$ （洋红色点线）、$\hat{\mathscr{I}}_3$ （深蓝色虚线）和 $\hat{\mathscr{I}}_4$ （浅绿色实线）。随着 l 和 k 增大，$\hat{\mathscr{I}}_3$ 和 $\hat{\mathscr{I}}_4$因维数增加以及相空间中的距离增大而上升。为满足可靠性条件，$\hat{\mathscr{I}}_3$ 必须为正，且 $\hat{\mathscr{I}}_{34} = \hat{\mathscr{I}}_3 - \hat{\mathscr{I}}_4$ 也必须为非负——也就是说，浅绿色线（$\hat{\mathscr{I}}_4$）必须保持在深蓝色虚线（$\hat{\mathscr{I}}_3$）下方。对于 $\alpha \leq 1$，当 $k = 1000, l = 4$ 时，以及当 $k = 7000, l = 5$ 时，在 α 的整个取值范围内均满足这些条件。相较之下，$\hat{\mathscr{I}}_1$ 和 $\hat{\mathscr{I}}_2$ 保持相对稳定，因为它们不受 l 变化的影响，并且在 k较大时收敛，仅在 α较低时略有偏差。这些结果表明，特定参数组合能够满足可靠性条件；然而，由于此类组合并非唯一，且所得传递熵强烈依赖于参数调优，因此观测值应解释为 *apparent*，而非完全传递熵。



(a) *Effective Rényi transfer entropies under varying k-NN and memory parameters for the causal direction in Process 1.b.*



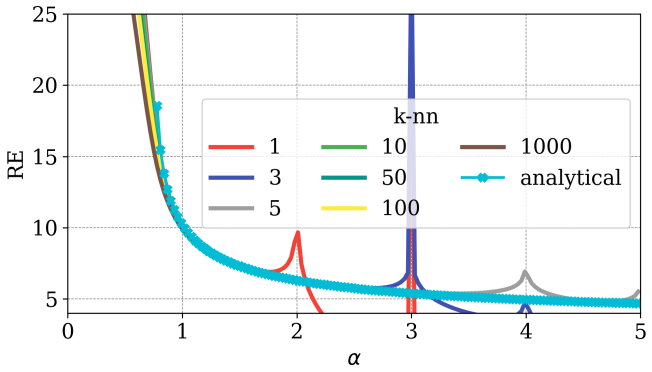
(b) *Relative explanation added (REA) for the causal direction in Process 1.b. under varying k-NN and memory parameters.*

横x轴表示 R´ényi 参数 α，而 y纵轴显示根据图´10中所示的分量熵推导出的有效 Rényi 转移熵估计值。所有估计均基于 $N = 10^5$ 个样本。虽然图中显示的所有有效 Rényi 转移熵值均为非负，但其中一些值源自违反可靠性条件的中间 Rényi 熵或条件 Rényi 熵估计值——例如单个分量为负——这削弱了其有效性。在整个 α 取值范围内满足所有可靠性标准的唯一有效 Rényi 转移熵曲线，对应于参数组合 $k = 7000$ 和 $l = 5$，由橙色线表示。

横x轴表示 R´ényi 参数 α，而 y纵轴显示相对新增解释（REA），其定义见公式15。与有效 Rényi 转移熵不同，REA 指标可直接用于诊断假阳性：超过 1 的值（由水平点线标出）表示违反可解释性。根据这一标准，当 $k = 1000$，$l = 4$ 且$\alpha \leq 1$，以及当 $k = 7000$，$l = 4$ 在所展示的整个 α 取值范围内，均可得到有效的有效 Rényi 转移熵。

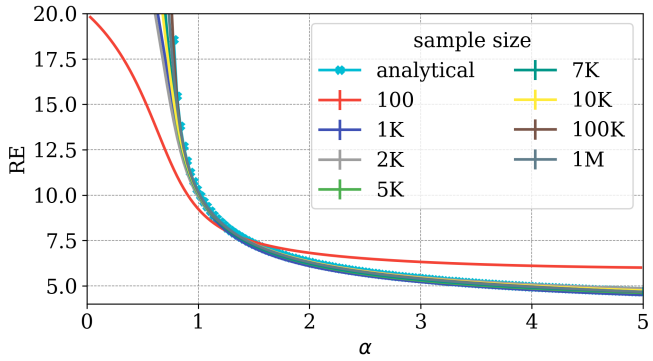
图 11： 过程 1.b：在不同的 k-近邻和记忆参数下估计的有效传递熵与相对解释增量

附录 D：柯西分布上的估计量性质



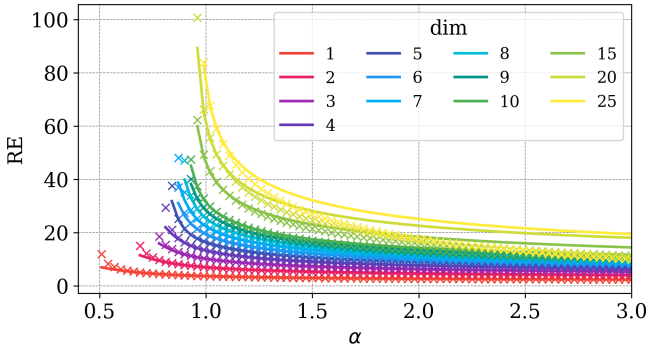
(a) *Influence of the k-Nearest Neighbor Parameter on Estimation Accuracy for the 3-Dimensional Cauchy Distribution.* *x*

横轴表示 Rényi 参数 α，y 轴显示了采用单位缩放矩阵的三维柯西分布的 Rényi 熵估计值。解析值以青色叉号标出，实线表示使用 $k = 1$、2、3、4、5、10、50、100 和 1000 得到的估计值。所有估计值均基于大小为 $N = 10^5$ 的样本。由于估计量依赖于 $\Gamma(k + 1 - \alpha)$，当伽马函数的自变量趋近于零或变为负数时，某些 k 与 α 的组合会导致发散。对于 $\alpha < 1$，最准确的估计值对应最小的邻域大小（$k = 1$）；而对于 $\alpha \geq 1$，较大的 k 能提供更稳定的估计值。这些发现与图2中观察到的趋势一致，进一步说明 k 的最优选择取决于 α。



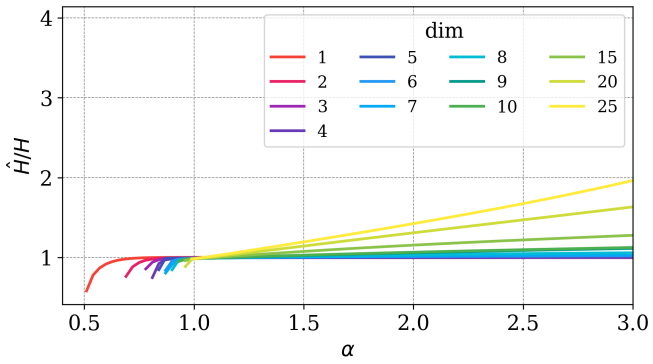
(b) *Effect of Sample Size N on the Estimation of Rényi Entropy for the Cauchy Distribution.* *x*

横轴表示 Rényi 参数 α，y 轴显示了单位尺度矩阵的三维柯西分布的估计 Rényi 熵。青色叉号表示解析值（注意，在此情况下，Rényi 熵对 $\alpha < 0.75$ 未定义；参见公式 B15）。实线表示使用 $k = 10$ 得到的估计值，对应样本量 $N = 100$、10^3、2×10^3、5×10^3、7×10^3、10^4、10^5 和 10^6，，颜色依次为红色、蓝色、浅灰色、绿色、青绿色、黄色、棕色和深灰色。对于 $\alpha > 1$，即使样本量适中（例如 $N = 10^3$），也能获得准确的估计。然而，范围为 $\alpha \in$（0.75,1）的估计值——该范围侧重于低概率事件——对样本量更为敏感，且收敛到理论值的速度较慢。在此范围内，$N \geq 10^5$ 时开始出现可靠的估计值。



(c) *Influence of Sample Dimensionality d on Estimation Accuracy for the Cauchy Distribution.* *x*

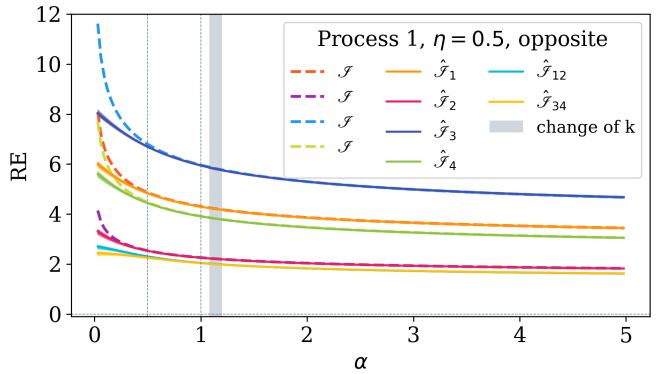
横轴表示 Rényi 参数 α，y轴显示 d 维柯西分布（采用单位缩放矩阵）的 Rényi 熵估计值（实线）和解析值（×标记）。随着维数增加，Rényi 熵的定义域逐渐变窄，尤其是在 α 值较小时。由于尾部更重且高维空间中的稀疏性增加，维数越高，估计准确性越低。如图（d）所示，通常在 $d < 10$时可以获得可靠的估计。



(d) *Estimation Errors on Cauchy Samples with Increasing Dimensionality* 横 *x*轴表示 R´ényi 参数 α，y轴显示采用单位缩放矩阵的 ´维柯西分布的估计 Rényi 熵与解析 Rényi 熵之比。样本 d量为 $N = 10^5$，$k = 10$。比值接近 1 表示估计精度较高。结果表明，维数增加会显著降低估计性能，这与图 3中正态分布所呈现的趋势一致。

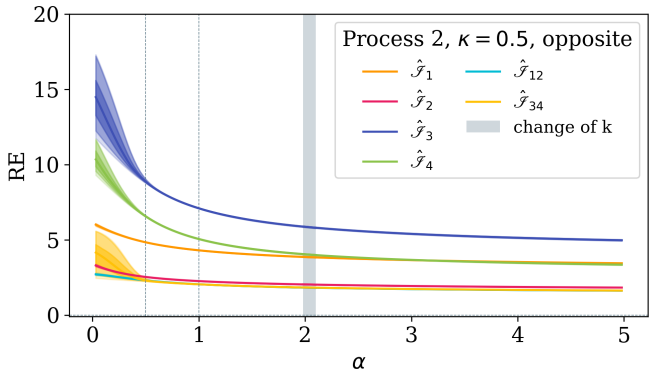
图 12：k-NN R´ényi《熵》估计量在柯西分布数据上的性能。

附录 E：过程 1–4 的详细可视化



(a) *Rényi entropies $\mathscr{I}_1$ $\mathscr{I}_2$ $\mathscr{I}_3$ $\mathscr{I}_4$ $\mathscr{I}_{12}$ and $\mathscr{I}_{34}$ underlying the RTE in non-causal direction i.e. $B^I \to A$ with coupling parameter $\eta = 0.5$.*

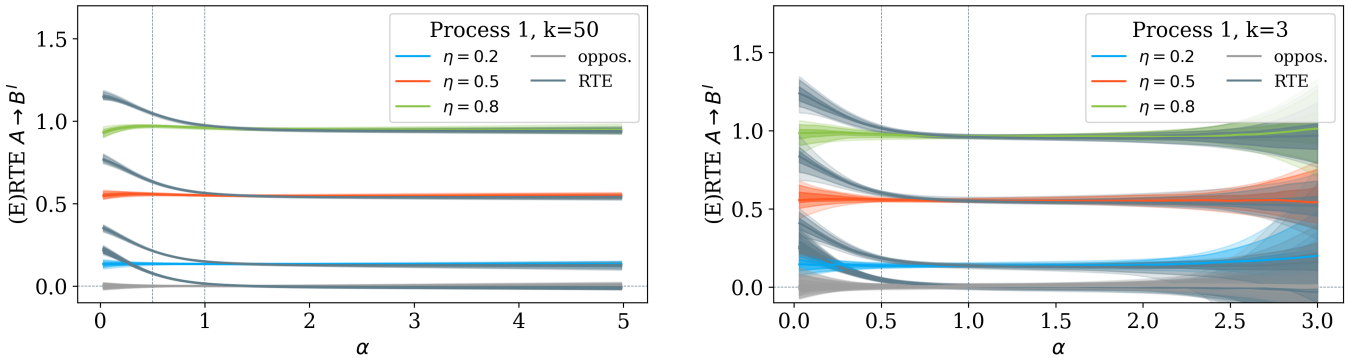
表示 Rényi 参数，y轴显示以比特为单位测得的相应 Rényi 熵的估计值（实线）和解析值（虚线）。分位数带展示了 30 次运行的估计结果。熵 $\mathscr{I}_1$、$\mathscr{I}_2$、$\mathscr{I}_3$和 $\mathscr{I}_4$（维数分别为 2、1、3、2）依次以橙色、洋红色、蓝色和绿色表示。对于较小的 α值（约小于 0.5），由于正态分布的支撑无界，估计值会偏离解析预测。条件熵 $\mathscr{I}_{12} = \mathscr{I}_1 - \mathscr{I}_2$ 和 $\mathscr{I}_{34} = \mathscr{I}_3 - \mathscr{I}_4$ 分别以青色和黄色表示。所有估计值均满足可靠性条件，即 $\hat{}\mathscr{I}_1$，$\hat{}\mathscr{I}_2$，$\hat{}\mathscr{I}_3$，$\hat{}\mathscr{I}_4$，$\hat{}\mathscr{I}_{12}$ 和 $\hat{}\mathscr{I}_{34}$均为正。此外，由于 $\hat{}\mathscr{I}_{12} \geq \hat{}\mathscr{I}_{34}$，$\alpha = 1$ 的 Shannon 传递熵也非负。这意味着所得 RTE 和有效 Rényi 转移熵具有可解释性。由于 $\hat{}\mathscr{I}_{12}$ 和 $\hat{}\mathscr{I}_{34}$ 在 α的整个范围内几乎相等，按照过程的定义，RTE 为零。$\alpha < 0.5$ 的微小偏差是由样本量有限造成的，并已在有效 Rényi 转移熵中得到校正。



(b) *Rényi entropies $\mathscr{I}_1$ $\mathscr{I}_2$ $\mathscr{I}_3$ $\mathscr{I}_4$ $\mathscr{I}_{12}$ and $\mathscr{I}_{34}$ underlying the RTE in non-causal direction i.e. $B^{II} \to C$ with coupling parameter $\eta = 0.5$.*

表示 Rényi 参数，y轴显示以比特为单位测得的相应 R´enyi 熵估计值。分位数带显示了 30 次运行的估计结果。熵 $\mathscr{I}_1$、$\mathscr{I}_2$、$\mathscr{I}_3$ 和 $\mathscr{I}_4$（维数分别为 2、1、3、2）依次以橙色、洋红色、蓝色和绿色表示。对于较小的 α 值，估计值呈现出较大的变异性，且变异性会随着维数增加而进一步增大。特别是，在三维中计算的 $\mathscr{I}_3$ 方差最大。条件熵 $\mathscr{I}_{12} = \mathscr{I}_1 - \mathscr{I}_2$ 和 $\mathscr{I}_{34} = \mathscr{I}_3 - \mathscr{I}_4$ 分别以青色和黄色表示。所有估计值均满足可靠性条件，即 $\hat{}\mathscr{I}_1$、$\hat{}\mathscr{I}_2$、$\hat{}\mathscr{I}_3$、$\hat{}\mathscr{I}_4$、$\hat{}\mathscr{I}_{12}$ 和 $\hat{}\mathscr{I}_{34}$ 均为正。此外，由于 $\hat{}\mathscr{I}_{12} > \hat{}\mathscr{I}_{34}$，$\alpha = 1$ 香农转移熵也为正。这意味着所得的 RTE 和有效RTE 均可解释。仅当 $\alpha < 0.5$ 时，传递熵才会变为负值；这种由有限样本量导致的偏差可通过有效RTE 部分校正，见图 15。

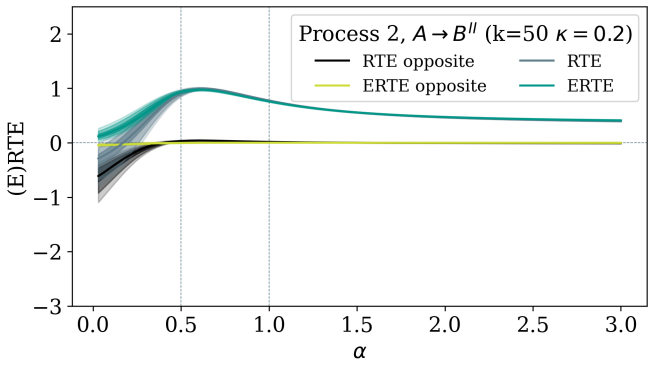
图 13：过程 1 和 2 中反方向的 Rényi 熵，$\mathscr{I}_1 - \mathscr{I}_4$，耦合参数为 $\eta = 0.5$ 和 $\kappa = 0.5$。



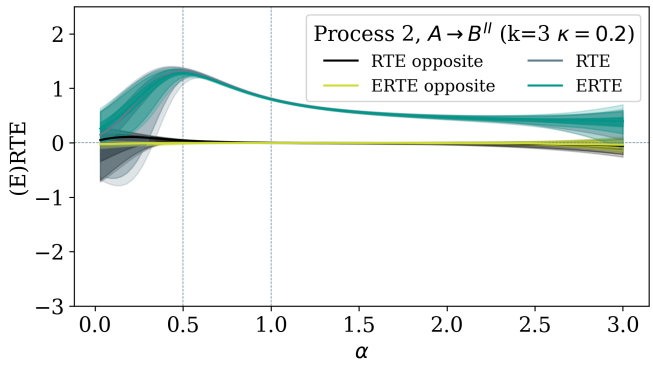
(a) $k = 50$两个方向上的 RTE 和有效 RTE。

(b) $k = 3$两个方向上的 RTE 和有效 RTE。

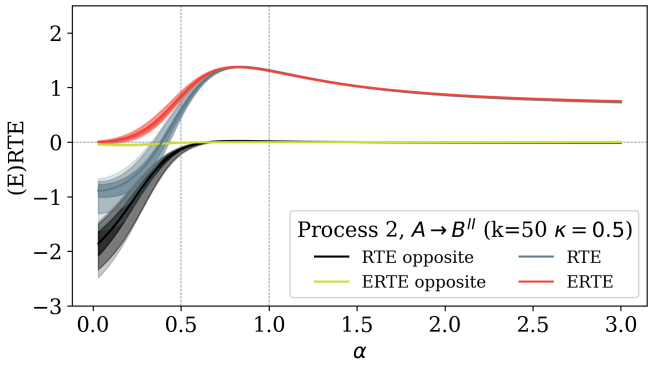
图14: *RTE and ERTE of Process 1 in both directions for* $k = 50$ *and* $k = 3$. 两个面板的 x 轴显示 Rényi 参数 α，y 轴显示估计的 RTE 和有效 RTE。记忆参数为 $r = l = 1$，样本量为 $N = 10^5$。分位数带——第 1–99、第 5–95 和第 25–75 百分位数带，并以一条线标示中位数——展示了 30 次运行所得的结果。蓝色、红色和绿色分位数带分别表示对应耦合方向上的有效 Rényi 转移熵。此外，以零为中心的灰色带对应非耦合方向上的有效 RTE，证实从 B 到 A 不存在因果影响。深灰色带表示耦合方向上的估计 RTE；对于 $\alpha < 1$，这些估计值表现出明显的估计偏差。这凸显了使用打乱的替代数据计算有效 RTE 以校正此类偏差的重要性。值得注意的是，尽管噪声较少，对于 $k = 3$，RTE 估计值从 $\alpha = 0.5$ 起就已准确；而对于 $k = 50$，只有从约 1 开始才能得到可靠的估计值。这再次说明，较小的 k 更适合对尾部进行估计。



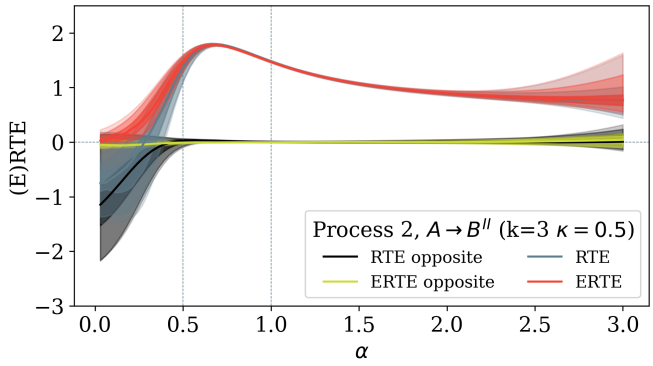
(a) $k = 50$ 和 $\kappa = 0.2$两个方向上的 RTE 和有效 Rényi 转移熵。



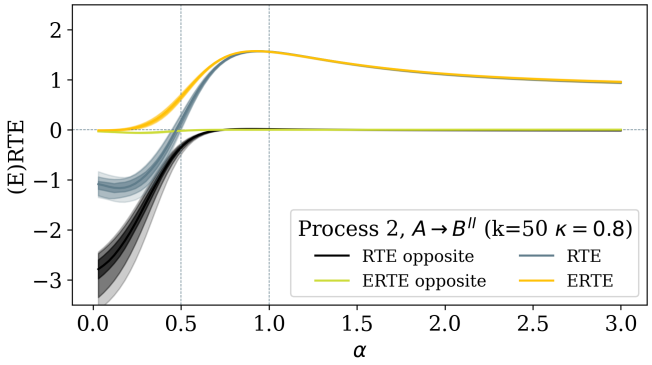
(b) $k = 3$ 和 $\kappa = 0.2$两个方向上的 RTE 和有效 Rényi 转移熵。



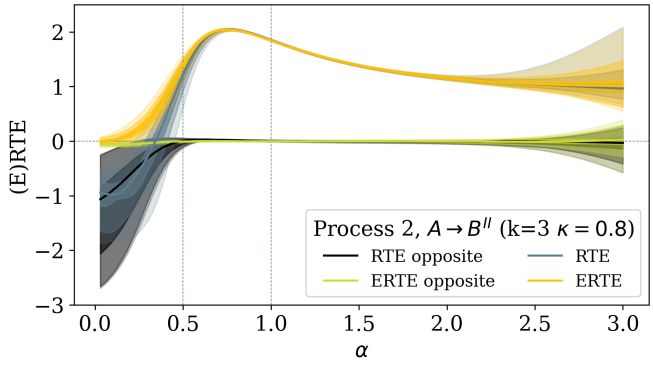
(c) $k = 50$ 和 $\kappa = 0.5$两个方向上的 RTE 和有效 Rényi 转移熵。



(d) $k = 3$ 和 $\kappa = 0.5$两个方向上的 RTE 和有效 Rényi 转移熵。

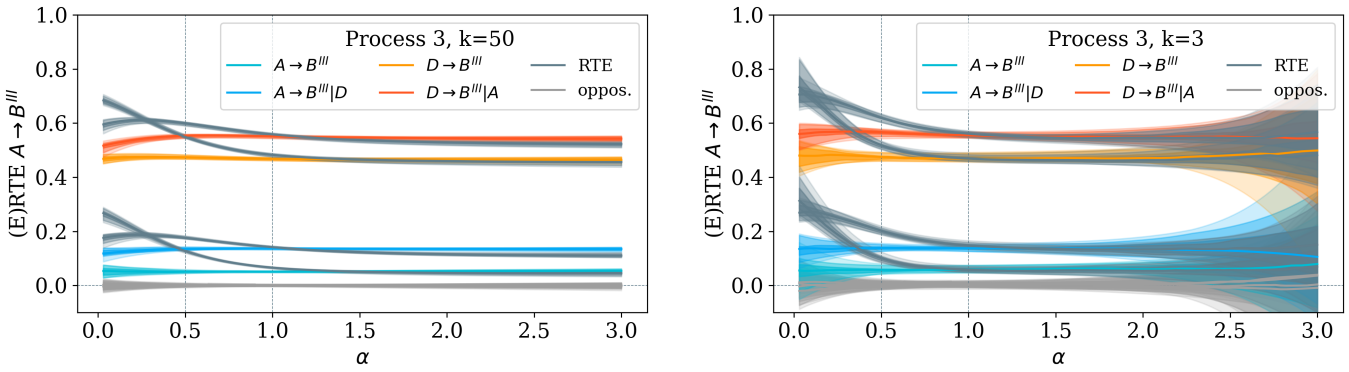


(e) $k = 50$ 和 $\kappa = 0.8$两个方向上的 RTE 和有效 Rényi 转移熵。



(f) $k = 3$ 和 $\kappa = 0.8$两个方向上的 RTE 和有效 Rényi 转移熵。

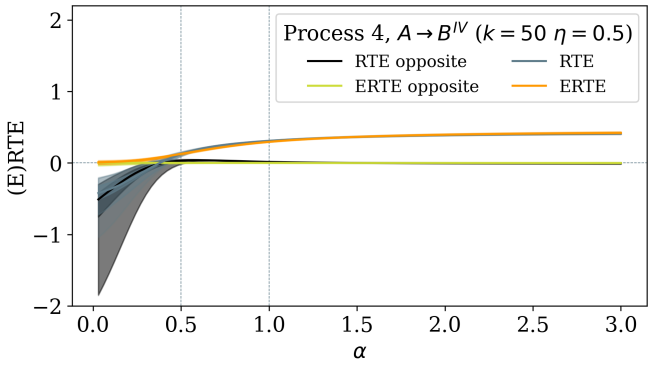
图 15： *RTE and ERTE of Process 2 in both directions for* $k = 50$ *and* $k = 3$. 所有面板中的 x 轴表示 R´enyi 参数 α， y 轴表示估计的 RTE 和有效 Rényi 转移熵。记忆参数为 $r = l = 1$，样本量为 $N = 10^5$。分位数带（第 1–99、第 5–95 和第 25–75 百分位数带，线条表示中位数）展示了 30 次运行的结果。在所有面板中，黑色带表示非因果方向上的 RTE，深灰色带表示因果方向上的 RTE，浅黄色带表示非因果方向上的有效 Rényi 转移熵，彩色带（黄色、红色和青绿色）表示因果方向上的有效 Rényi 转移熵，如图 6a所示。在每种情况下，当 α 取值较小时，两个方向上的 RTE 估计值（黑色和灰色）都存在显著偏差。这种偏差主要是由过程 B^{II} 的重尾分布造成的，通过减去基于 B^{II}的打乱版本计算出的 RTE，可以有效地校正这种偏差。



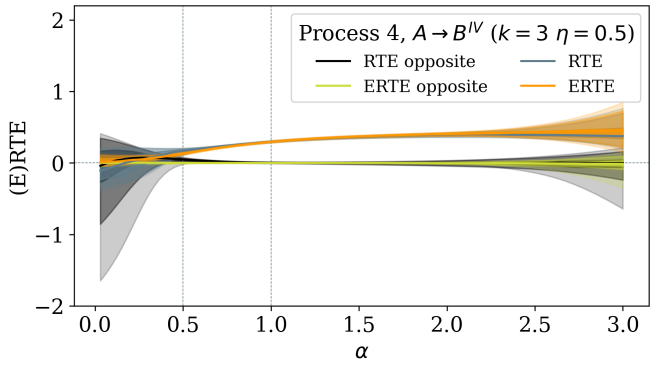
(a) $k = 50$两个方向上的 RTE 和有效RTE。

(b) $k = 3$两个方向上的 RTE 和有效RTE。

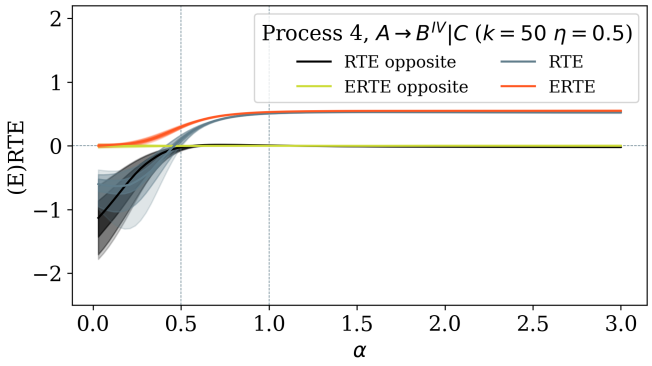
图16: *RTE and ERTE of Process 3 in both directions for* $k = 50$ *and* $k = 3$. 两个面板中的 x轴表示 R´ényi 参数 α， y轴表示估计的 RTE 和有效RTE。记忆参数为 $r = l = 1$，样本量为 $N = 10^5$。分位数带——第 1–99、第 5–95 和第 25–75 百分位数，线条表示中位数——展示了 30 次运行的结果。在两个面板中，深灰色带表示因果方向上估计的 RTE，浅灰色带（位于零处）表示非因果方向上的有效RTE，彩色带（青色、浅蓝色、橙色和红色）表示图7中因果方向上的有效RTE。对于 $\alpha < 1$，估计的 RTE 表现出明显的估计偏差。这凸显了使用打乱的替代数据计算有效RTE，以校正此类偏差的重要性。值得注意的是，尽管噪声较少，但对于 $\alpha \approx 0.7$，当 $k = 3$时，RTE 估计值已经相当准确；而对于 $k = 50$，只有从约 1 开始才能获得可靠的估计。这再次证实，较小的 k 值更适合对尾部进行估计。



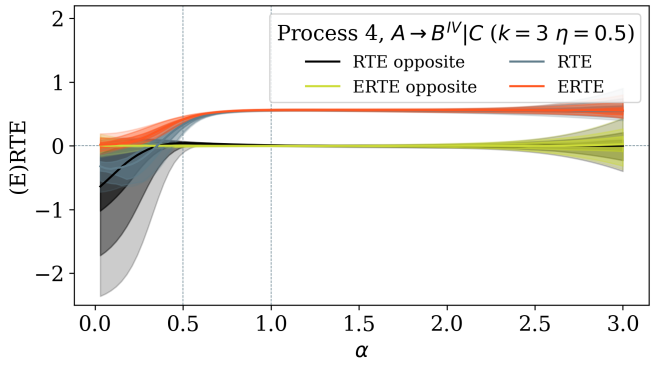
(a) $k = 50$ 和 $\eta = 0.5$两个方向上的 RTE 和有效 Rényi 转移熵。



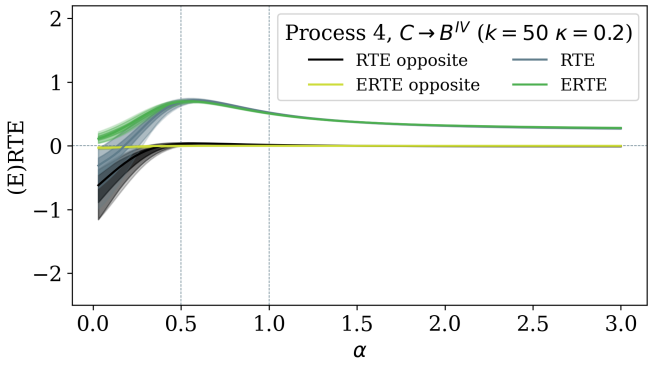
(b) $k = 3$ 和 $\eta = 0.5$两个方向上的 RTE 和有效 Rényi 转移熵。



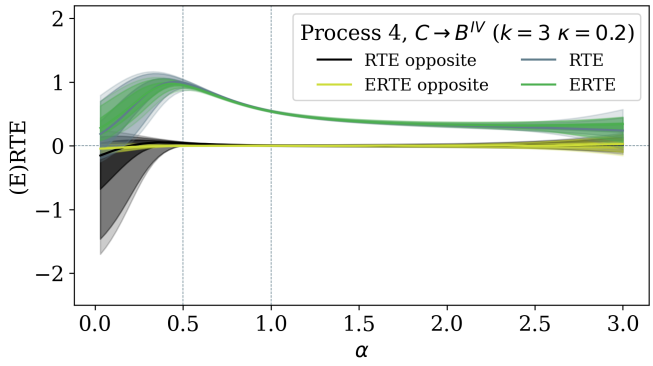
(c) $k = 50$ 和$\eta = 0.5$两个方向上的条件 Rényi 转移熵和有效 Rényi 转移熵。



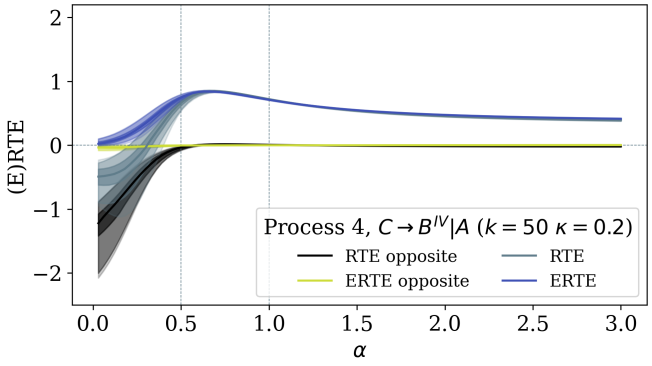
(d) $k = 3$ 和$\eta = 0.5$两个方向上的条件 Rényi 转移熵和有效 Rényi 转移熵。



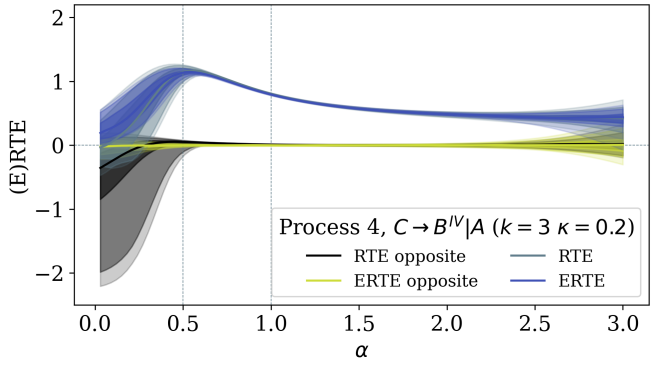
(e) $k = 50$ 和 $\kappa = 0.2$两个方向上的 RTE 和有效 Rényi 转移熵。



(f) $k = 3$ 和 $\kappa = 0.2$两个方向上的 RTE 和有效 Rényi 转移熵。



(g) $k = 50$ 和$\kappa = 0.2$双向的条件 Rényi 转移熵和有效 Rényi 转移熵。



(h) $k = 3$ 和$\kappa = 0.2$两个方向上的条件 Rényi 转移熵和有效 Rényi 转移熵。

图17: *RTE and ERTE of Process 4 in both directions for* $k = 50$ *and* $k = 3$. 所有面板中的 x 轴表示 R´enyi 参数 α， y 轴表示估计的 RTE 和有效 Rényi 转移熵。记忆参数为 $r = l = 1$，样本量为 $N = 10^5$。分位数带（第 1–99、第 5–95 和第 25–75 百分位数带，线条表示中位数）展示了 30 次运行所得的结果。在所有面板中，黑色带表示非因果方向上的 RTE，深灰色带表示因果方向上的 RTE，浅黄色带表示非因果方向上的有效 Rényi 转移熵，彩色带（橙色、红色、绿色和深蓝色）表示因果方向上的有效 Rényi 转移熵，如图8所示。在大多数情况下，当 α 取值较小时，两个方向上的 RTE 估计值（黑色和灰色）都存在显著偏差。这种偏差主要源于过程 B^{IV} 的重尾分布，通过减去基于 B^{IV}的打乱版本计算出的 RTE，可以有效地校正这一偏差。

[1] H. Kantz 和 T. Schreiber，《时间序列分析》（剑桥大学出版社，2003年）。[2] J. T. Lizier 和 M. Prokopenko，区分信息传递与因果效应，《欧洲物理学杂志B》73，605（2010）。[3] C. Granger，通过计量经济模型和交叉谱方法研究因果关系，《计量经济学》37，424（1969）。[4] N. Wiener，*Modern Mathematics for Engineers; Beckenbach E.F. Ed.*，《》（McGraw-Hill，纽约，美国，1956年）。[5] R. Q. Quiroga、J. Arnhold 和 P. Grassberger，从同步模式中学习驱动-响应关系，《物理评论E》61，5142（2000）。[6] H. Ye、E. R. Deyle、L. J. Gilarranz 和 G. Sugihara，使用收敛交叉映射区分时间延迟因果相互作用，《科学报告》5，14750（2015）。[7] M. Paluš、A. Krakovská、J. Jakubík 和 M. Chvostekov´a，因果关系、动力系统与时间之箭，《混沌》28，075307（2018）。[8] A. Kathpalia、P. Manshour 和 M. Paluš，利用序数模式的压缩复杂度对不规则采样时间序列进行稳健因果推断，《科学报告》12，14170（2022）。[9] Y. Huang、Z. Fu 和 C. L. E. Franzke，在机器学习框架下从时间序列中检测因果关系，《混沌》30，063116（2020）。[10] N. Ay 和 D. Polani，因果网络中的信息流，《复杂系统进展》11，17（2008）。[11] P. Jizba、H. Kleinert 和 M. Shefaat，金融时间序列之间的 Rényi 信息传递，《物理学A》391，2971（2012）。[12] P. Jizba、H. Lavička 和 Z. Tabachov´a，以 Rényi 转移熵表示时间序列中的因果推断，《熵》24（2022）。[13] K. Hlaváčkov´a-Schindler、M. Paluš、M. Vejmelka 和 J. Bhattacharya，基于信息论方法的时间序列分析中的因果关系检测，《物理学报告》441，1（2007）。[14] T. Schreiber，信息传递测量，《物理评论快报》85，461（2000）。[15] R. Marschinski 和 H. Kantz，分析金融时间序列之间的信息流：一种改进的传递熵估计器，《欧洲物理学杂志B—凝聚态物质与复杂系统》30，275（2002）。[16] M. Paluš、V. Komárek、Z. Hrnčíř 和 K. Šťerbová，将同步视为信息率的调节：从双变量时间序列中进行检测，《物理评论E》63，046211（2001）。[17] D. Gencaga，传递熵，《熵》20，288（2018）。[18] T. Bossomaier、L. Barnett、M. Harré 和 J. T. Lizier，*An Introduction to Transfer Entropy: Information Flow in Complex Systems*（施普林格，沙姆，瑞士，）。[19] M. Paluš、M. Chvostekov´a 和 P. Manshour，Rényi 信息传递揭示的极端事件成因，《科学进展》10，eadn1721（2024）。[20] N. Jajcay、S. Kravtsov、G. Sugihara 等，厄尔尼诺—南方涛动中跨时间尺度的同步与因果关系——ern 振荡，npj《气候与大气科学》1，33（2018）。[21] E. Pereda、R. Q. Quiroga 和 J. Bhattacharya，《神经生理信号的非线性多变量分析》，《神经生物学进展》77，1（2005）。[22] R. Vicente、M. Wibral、M. Lindner 等，《传递熵——一种用于神经科学的无模型有效连接性度量》，《计算神经科学杂志》30，45–67（2011）。[23] H. H. Permuter、Y.-H. Kim 和 T. Weissman，《定向信息在投资组合理论、数据压缩和假设检验中的解释》，IEEE《信息论汇刊》57，3248（2011）。[24] A. Rényi，*Selected Papers of Alfréd Rényi, 2nd Vol.*（Akademia Kiado，布达佩斯，1976）。[25] A. Rényi，《熵与信息的度量》，载于*Proceedings of the fourth Berkeley symposium on mathematical statistics and probability volume 1: contributions to the theory of statistics*，第4卷（加利福尼亚大学出版社，1961），第547–562页。[26] P. Jizba 和 T. Arimitsu，《Rényi视角下的世界：多重分形系统热力学》，《物理学年鉴》312，17（2004）。[27] J. Zhang、J. Cao、T. Wu、W. Huang、T. Ma 和 X. Zhou，《一种基于核密度估计的新型自适应多尺度 Rényi 传递熵》，《混沌、孤子与分形》175，113972（2023）。[28] S. Behrendt、T. Dimpfl、F. J. Peter 和 D. J. Zimmermann，《Rtransferentropy——使用有效传递熵量化不同时间序列之间的信息流》，《SoftwareX》10，100265（2019）。[29] J. Korbel、X. Jiang 和 B. Zheng，《复杂金融网络中社群间的传递熵》，《熵》21（2019）。[30] N. Leonenko、L. Pronzato 和 V. Savani，《一类用于多维密度的 Rényi 信息估计量》，《统计学年鉴》36，2153（2008）。[31] M. Paluš 和 A. Stefanovska，《从相互作用振荡器的相位确定耦合方向：一种信息论方法》，《物理评论E》67，055201(R)（2003）。[32] J. Sun 和 E. M. Bollt，《因果熵识别间接影响、邻居的主导作用及预期耦合》，《物理学D辑：非线性现象》267，49（2014）。[33] 该估计量默认使用欧几里得距离，但在适当情况下也可采用其他距离度量，例如马氏距离；球面分布数据尤其适用。[34] T. M. Cover，*Elements of information theory*（John Wiley & Sons，1999）。[35] H. L. Van Trees 和 K. L. Bell，*Detection estimation and modulation theory part I: detection estimation and filtering theory* 2013[36]（John Wiley & Sons，）。[36] P. Jizba 和 T. Arimitsu，《Rényi视角下的世界：多重分形系统热力学》，《物理学年鉴》312，17（2004）。[37] R. A. Fisher 和 F. Yates，*Statistical Tables for Bio-logical, Agricultural, and Medical Research*（Oliver and Boyd，1939）。[38] R. Bellman，*Adaptive Control Processes: A Guided Tour*

（普林斯顿大学出版社，普林斯顿，新泽西州，1961）。[39] C. M. Bishop， *Pattern Recognition and Machine Learning*（施普林格，纽约，2006）。[40] T. Hastie、R. Tibshirani 和 J. Friedman， *The Elements of Statistical Learning: Data Mining Inference and Prediction*，2，第2版（施普林格，纽约，纽约州，美国，2009）。[41] M. Paluš，《从非线性到因果性：复杂动力学背后物理机制的统计检验与推断》，《当代物理学》**48**，307（2007）。[42] G. Lancaster、D. Iatsenko、A. Pidde、V. Ticcinelli 和 A. Stefanovska，《用于假设检验的替代数据

物理系统，《物理学报告》**748**，1（2018）。[43] M. Vejmelka、L. Pokorn´a、J. Hlinka、D. Hartman、N. Jajcay 和 M. Paluˇs，《多变量气候数据中的非随机相关结构与降维》，《气候动力学》**44**，2663（2015）。[44] J. Runge、V. Petoukhov、J. F. Donges、J. Hlinka、N. Ja- jcay、M. Vejmelka、D. Hartman、N. Marwan、M. Paluˇs 和 J. Kurths，《识别复杂时空系统中的因果关口与中介因素》，《自然通讯》**6**，8502（2015）。[45] C. E. Shannon，《通信的数学理论》，《贝尔系统技术期刊》**27**，379（1948）。