时空点过程的基于评分的变点检测与区域定位

Wenbin Zhou、Liyan Xie 和 Shixiang Zhu

arXiv:2602.04798v1 [stat.ME] 4 Feb 2026

摘要——**我们研究时空点过程的序贯变点检测问题。在这类问题中，要使检测结果具有实际指导意义，不仅需要识别分布变化何时发生，还需要定位其在空间中的表现位置。经典最快变点检测方法能够为检测延迟和虚警率提供有力保证，但现有针对点过程数据的方法主要关注时间变化，并未明确推断受影响的空间区域。我们提出一种无需似然的基于评分的检测框架，在不假设变化前或变化后的动态过程具有参数知识的情况下，联合估计变化时间和连续时空中的变化区域。该方法利用局部化且经过条件加权的 Hyvärinen 评分来量化事件级别上相对于名义行为的偏离，并使用时空累积和型统计量，在预先指定的一类空间区域上汇总这些评分。该程序以序贯方式运行，同时输出一个停止时间和一个估计的变化区域，从而实现具有空间可解释性的实时检测。我们建立了关于虚警控制、检测延迟和空间定位精度的理论保证，并通过模拟和真实世界的时空事件数据展示了所提方法的有效性。**

索引词—**变点检测、时空点过程、分数模型、区域定位。**

I. 引言

变点检测是统计信号处理中的经典核心问题，关注对序贯观测概率结构中突变的实时检测 [1]。给定一个数据流，目标是在满足虚警约束的前提下，以最小检测延迟识别底层分布偏离其名义状态的时间。这类问题广泛存在于各种信号处理应用中，包括地震监测 [2], 工业过程控制 [3], 网络监控、以及供应链监测 [4]。在这些场景中，快速且可靠的检测对于及时干预和减轻后续影响至关重要。

近年来，许多现代信号处理系统会生成事件型时空数据，这些数据由同时以时间和空间位置索引的离散事件组成。此类数据可自然地用时空点过程[5], [6] 进行建模，并出现在地震目录[7], 、野火起火事件 [8], 、犯罪事件 [9], [10], 和交通事故事件 [11]等应用中。现有的变点检测方法针对点过程主要聚焦于时间上的变化，旨在检测事件强度或动态随时间发生的偏移，同时将空间视为固定的或辅助性的 [2], [12]。尽管这一视角对于检测全局分布变化十分有效，但它抽离了往往对解释与响应至关重要的空间结构。

然而，在许多现实应用中，变化的重要性与其发生的位置密不可分。例如，在地震监测中，升高的地震活动通常需要结合其空间集中性来解读——例如活动沿某一断层段分布，或集中在某个有限区域内——因为此类模式能够区分具有实际意义的地震序列（例如，震群样活动或前震）与分散的背景波动[13]。检测总体事件率的变化却不进行空间定位，只能提供有限的地球物理学洞见，也无法支持后续对断层行为或灾害演变的分析。在野火监测中也存在类似考量：只有当起火模式的变化能够在空间上得到解析时，才具有实际操作意义，从而使相关机构能够识别高风险区域、确定监测优先级并分配减灾资源 [14], [15]。这些例子表明，对于时空事件系统，具有实际行动价值的变化检测需要联合推断变化发生的时间及其表现的位置，而不能将变点检测视为纯粹的时间问题。

尽管具有重要的实际意义，能够同时进行序贯变点检测并定位受影响空间区域的方法，相较于纯时间方法仍发展不足，在时空点过程数据中尤为如此。这一局限主要源于两项建模与计算挑战：(*i*) 在时空点过程中，“变化”可能有多种并不等价的形式，例如，背景强度面的变化 [16], 事件活动出现空间局部性增长[5], 或交互结构或触发结构发生变化 [17],，如自激或互激过程中的情形。因此，变化后行为无法用有限个备选方案来描述；相反，它属于一个可能变化后情景构成的无限维族。相比之下，现有许多关于变化检测与定位的研究，都着重于从有限个候选方案中识别变化后情景 [18]。因此，这些方法并不能直接应用于我们的设定。(*ii*) 从推断角度来看，我们关注的对象是在连续时空中演化的潜在时空强度（或条件强度），而观测数据仅以稀疏且不规则的事件时间和位置形式出现。这使得完全基于似然的序贯程序难以处理：它们通常只有在强参数假设下才易于计算

W. Zhou 和 S. Zhu 就职于美国宾夕法尼亚州匹兹堡 15213 的卡内基梅隆大学信息系统与公共政策海因茨学院（电子邮件：wenbinz2@andrew.cmu.edu；shixiangzhu@cmu.edu）。L. Xie 就职于美国明尼苏达州明尼阿波利斯 55455 的明尼苏达大学工业系统与工程系（电子邮件：liyanxie@ umn.edu）。

（例如，具有指定结构的泊松过程/霍克斯过程），而当前后变化动态未知、高维或具有空间异质性时，处理起来会困难得多——而这些恰恰是区域识别最有价值的状态 [16]。

为应对这些挑战，我们提出一种无需似然、基于得分的时空变点检测框架，可同时识别变化时间和受影响的空间区域。我们的方法无需显式估计似然，也不假设预变化或后变化分布的先验知识。相反，它利用得分函数——即对数密度的梯度——该函数能够完整刻画底层分布，并且在高维设置下往往比似然本身更易于可靠估计 [19]。

我们方法的核心是局部条件加权海因里希得分，用于量化每个观测事件相对于变化前动态的异常程度。随后，我们使用时空CUSUM型统计量，在预先指定的空间区域类别上汇总这些事件级得分。该算法按序运行，搜索累积得分超过检测阈值的区域；一旦识别出此类区域，程序便输出停止时间和估计的变化区域，从而实现具有空间可解释性的实时检测。我们从三个方面为所提出的算法建立理论保证。首先，我们通过推导平均运行长度的下界来刻画其虚警行为。其次，我们分析检测效率并获得检测延迟的界，从而刻画算法在变化发生后响应的速度。第三，我们为变化区域定位准确率提供理论保证，并使用估计变化区域与真实变化区域之间的杰卡德指数进行量化。

我们的贡献主要有三点。(*i*) 我们为点过程提出了一种连续时空变化区域检测框架，其中变化点和受影响的空间区域均未知，从而弥合了经典纯时间累积和方法与临时性空间扫描方法之间的差距。(*ii*) 我们基于得分匹配，开发了一种可扩展的无似然检测方法，通过新颖的时空CUSUM型统计量和交替优化方案，联合检测变化点并定位受影响区域。(*iii*) 我们建立了将检测延迟与定位精度联系起来的渐近保证，并通过全面的模拟和两个现实应用展示了所提方法的有效性。

本文其余部分安排如下。第二节回顾相关工作；第三节介绍问题表述和预备知识；第四节详述所提方法；第五节给出算法的理论分析；第六节报告数值结果。

II. 相关工作

我们回顾与本研究最相关的三类既有工作：（*i*）经典的序贯变点检测与多流定位，（*ii*）时空点过程的变化检测，以及（*iii*）用于密度和变点建模的基于得分的方法。

a) 经典变化检测与定位：对于分布已知的独立同分布（i.i.d.）观测，经典的序贯变点检测已是一个发展成熟的领域。相关综述可参见 [1], [20], [21] 等著作或 [22], [23] 等论文。过去十年间，多数据流的序贯变化检测受到广泛关注，例如针对高维独立数据流 [24]–[27] 以及相关数据流 [28]–[31]的检测。与我们的研究部分相关的一类工作，研究的是多流场景中的变点检测与定位：变化发生在未知的数据流子集中，目标是同时检测变化点并识别受影响的数据流。在一系列研究中，Nikiforov 为随机动力系统提出了极小极大最优检测隔离算法 [32], ，开发了该算法的递归变体以提高计算效率 [33]，并在错误定位概率和虚警前平均时间受到约束的条件下，给出了平均检测延迟的渐近下界[18]。更详细的综述见 [21], [34], [35] 。值得一提的是，现有关于检测与定位的研究通常假设变化后情景构成有限集合，例如变化影响有限个数据流中的某一个。相比之下，我们研究连续空间变化区域，并旨在将该区域与变化检测联合识别。

b) 时空点过程及其变化检测： 点过程模型因能够直接刻画事件间等待时间 [36]，而被广泛用于对离散事件数据建模。在这一框架中，泊松过程假设事件间隔时间相互独立且服从指数分布，而霍克斯过程则允许条件强度依赖于过去事件的历史记录 [37]，从而引入时间依赖性。这种自激结构使霍克斯过程尤其适合捕捉时间依赖性和因果效应，应用范围涵盖金融市场 [38] 、地震建模 [39] 以及社交与信息网络 [40], [41]。多维霍克斯过程进一步拓展了这种建模能力，使其能够应用于网络化场景，表示高度相关的事件流以及信号在网络中的传播[5]。点过程模型中的变化检测在单流与多流场景中都持续受到关注。现有研究既包括对泊松过程的经典探讨 [42]–[44], ，也包括针对一维 [45], [46] 以及多维或基于网络的点过程 [16]，[47]的近期进展。值得注意的是， [47]提出了用于检测网络化事件流中变化的广义似然比型程序，而 [16] 则开发了一种惩罚动态规划方法，用于在离线场景下识别高维自激泊松过程中的参数变化。这些研究也与多源最快检测问题密切相关，此类问题通常假设各数据流之间相互独立；典型例子包括针对复合泊松模型中多个变点的最小值进行检测的最优贝叶斯程序 [48] ，以及最优

由伊藤过程 [49]所支配的耦合系统的规则。我们的工作与这些研究相关，因为我们假设事件动态服从时空霍克斯过程。然而，我们的检测问题有所不同：我们旨在联合检测时间和连续空间中的变化，而现有方法大多聚焦于离散化多流设置。

c) 基于得分的方法： 关于得分匹配的奠基性工作最早在 [50] 中提出，作为一种估计未归一化得分模型的有效方法。其核心思想是最小化得分模型与真实的（Stein）得分函数之间的均方误差 [50]，其中后者可以通过分部积分技术被消去。然而，这一技巧依赖于密度的支集为无约束域这一假设，从而限制了其在序列建模领域中的应用 [51]。为解决这一局限性，文献 [52]–[54],提出了加权得分匹配，其核心思想是构造满足相关假设的重加权密度。另一条研究去噪得分匹配的工作路线 [55]将估计问题重新表述为去噪任务：得分模型尝试根据扰动数据预测注入的噪声，并证明当注入噪声的方差为零时，该方法与原始得分匹配等价。后续研究表明，去噪得分匹配还能够提高数值稳定性和估计效率，尤其适用于高维数据或流形结构数据，从而推动了基于得分的扩散模型的一系列发展 [56]–[58]。近年来，基于得分的方法也被应用于点过程建模 [59]–[61], ，其中 [61] 提出了一种通用的基于去噪得分匹配的模型，用于建模点过程的条件概率密度函数。近期的另一条研究路线则探讨了基于得分的方法在变点检测问题中的应用[62]，通过调整Hyvärinen得分，将其用作检测统计量（SCUSUM）。目前已经发展出多种扩展方法，包括贝叶斯设置 [63]、鲁棒设置 [64]、基于扩散的设置 [65]、条件设置 [66]、去噪得分匹配估计 [19], 以及多流设置 [67]。我们的工作与这些关于基于得分的变化检测的研究最为密切相关，但关注的对象是点过程事件类型序列，而非时间序列。

III. 问题设置与预备知识

本节将在第三节-A中回顾时空点过程，在第三节-B中表述变点检测与区域定位问题，并在第三节-C中重新探讨经典的累积和检验，着重说明其在时空场景中的局限性。

A. 时空点过程

时空点过程（STPP）是一种用于描述在时间和空间中发生的离散事件的随机模型 [5]。有限时间范围 $[0,T)$ 内的一次实现由事件历史表示

$$\mathcal{H}_T = \{x_1, \dots, x_{N_T}\} \subseteq [0,T) \times \mathcal{S} =: \mathcal{X},$$

其中，每个事件 $x_j = (t_j,\ s_j)$ 由时间 $t_j \in [0,T)$ 和空间位置 $s_j \in \mathcal{S}$ 组成。我们分别用 $t(x)$ 和 $s(x)$ 表示任意 $x \in \mathcal{X}$ 的时间分量和空间分量，

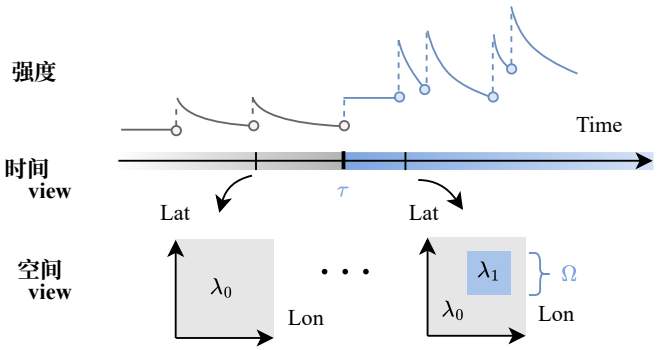

图 1. 问题设置示意图。经过一段时间 τ 后，事件生成的强度函数在有限空间区域 Ω 内从 λ_0 转变为 λ_1 。

而 N_T 表示时间 T 之前观测到的事件总数。

该过程还可以通过其计数测度 $\mathbb{N}$ 来表征，该测度对任意可测子集 $\mathcal{X}' \subseteq \mathcal{X}$ 定义如下：

$$\mathbb{N}(\mathcal{X}') = |\mathcal{H}_T \cap \mathcal{X}'|,$$

它统计落在 $\mathcal{X}'$ 中的事件数量。时空点过程的局部动态由其条件强度函数支配，该函数指定给定截至时间 t 的历史时，在 $x = (t,s)$ 处的瞬时事件率：

$$\lambda(x \mid \mathcal{H}_t) = \lim_{|\mathrm{d}x| \to 0} \frac{\mathbb{E}\left[\mathbb{N}(\mathrm{d}x) \mid \mathcal{H}_t\right]}{\mathrm{d}x}. \tag{1}$$

条件强度完全表征未来事件的分布。具体而言，令 t_n 表示 $t(x)$ 之前最近一次事件发生的时间，则下一事件在 x 发生的条件密度为

$$p(x \mid \mathcal{H}_{t(x)}) = \lambda(x \mid \mathcal{H}_{t(x)}) \cdot \exp\left(- \int_{[t_n, t(x)) \times \mathcal{S}} \lambda(u \mid \mathcal{H}_{t(u)}) \mathrm{d}u \right).$$

给定观测事件流 $\mathcal{H}_T$，强度为 λ 的模型的对数似然具有如下形式：

$$\ell(\mathcal{H}_T) = \int_{\mathcal{X}} \log \lambda(x | \mathcal{H}_{t(x)}) \mathrm{d}\mathbb{N}(x) - \int_{\mathcal{X}} \lambda(x | \mathcal{H}_{t(x)}) \mathrm{d}x, \tag{2}$$

其中，关于计数测度的积分对应于对观测事件求和，即，

$$\int_{\mathcal{X}'} \phi(x) \mathrm{d}\mathbb{N}(x) = \sum_{x_j \in \mathcal{H}_T \cap \mathcal{X}'} \phi(x_j) \quad \text{for any } \phi : \mathcal{X} \to \mathbb{R}.$$

B. 时空检测的问题设定

我们考虑由条件强度如式 (1) 所定义的时空点过程生成的事件流。我们的目标是确定底层强度是否以及何处发生结构变化。具体而言，我们旨在检测强度从变化前强度 $\lambda_0(\cdot)$ 转变为变化后强度 $\lambda_1(\cdot)$，该转变发生在时间上未知的变点 τ，并局限于未知的空间区域 Ω。

回顾一下，$\mathcal{X} := [0,T) \times \mathcal{S} \subseteq \mathbb{R}_+ \times \mathbb{R}^2$ 表示完整的时空域。若存在变化，假定它仅影响

变化后区域 $\mathcal{G}_{\text{post}} \coloneqq [\tau, T) \times \Omega$，而在其补集（称为变化前区域 $\mathcal{G}_{\text{pre}} \coloneqq \mathcal{X} \setminus \mathcal{G}_{\text{post}}$）上，该过程保持不变。在这一表述下，检测任务可转化为一个假设检验问题：

$$\begin{aligned} H_0: &\quad \lambda(x \mid \mathcal{H}_{t(x)}) = \lambda_0(x \mid \mathcal{H}_{t(x)}), && \forall x \in \mathcal{X}, \\ H_1: &\quad \lambda(x \mid \mathcal{H}_{t(x)}) = \lambda_0(x \mid \mathcal{H}_{t(x)}), && \forall x \in \mathcal{G}_{\text{pre}}, \\ &\quad \lambda(x \mid \mathcal{H}_{t(x)}) = \lambda_1(x \mid \mathcal{H}_{t(x)}), && \forall x \in \mathcal{G}_{\text{post}}. \end{aligned} \tag{3}$$

图1展示了这一时空变点情境。为简化记号，在本文余下部分的记号中，我们省略对历史 $\mathcal{H}_{t(x)}$ 的显式依赖。请注意，通过在时空域 $\mathcal{X}$ 中加入标记空间，问题设置自然地扩展到带标记时空点过程 [39], [68]。

(3)中的假设表述范围广泛，涵盖了许多常用的点过程模型。下面重点介绍两个具有代表性的例子。

例 1（齐次泊松过程）。 假设变化前后的动态过程均服从齐次泊松过程，其强度为常数 $\mu_0 \neq \mu_1$。在这种情况下，$\lambda_i(x) = \mu_i$ 对于 $i = 0,\ 1$。

例 2（自激霍克斯过程）。 霍克斯过程 *[68],* 提供了一个更灵活的类别，其变化前后的条件强度形式如下：

$$\lambda_i(x) = \mu_i + \alpha \sum_{x' \in \mathcal{H}_{t(x)}} \kappa(x, x'), \qquad i = 0, 1, \tag{4}$$

其中，$\mu_i \in \mathbb{R}_{\geq 0}$ 是基准强度，$\kappa:\ \mathcal{X} \times \mathcal{X} \to \mathbb{R}_{\geq 0}$是自激核，$\alpha$ 控制触发效应的强度。指数核是常见的选择，

$$\kappa(x, x') = \beta \exp\big(-\beta \|x - x'\|\big),$$

它能够捕捉局部的时空聚类效应；请注意，该核在 x上积分为一。

在例 2 中，尽管假设变化仅发生在基准强度（即，$\mu_0 \neq \mu_1$）上，自激结构仍会显著增加检测的复杂性，因为变化前事件可能通过累积历史持续影响未来强度，从而产生延迟、平滑或空间传播的特征，模糊变化真正发生的时间和空间范围。与此同时，这种对过去事件的依赖也凸显了霍克斯过程的实际重要性，因为自激是许多现实现象的内在特征——例如地震余震、 [39], 犯罪和社会互动, [10], 以及信息或疫情传播 [17]——因此，霍克斯过程是对具有传染性或级联效应的时空数据进行建模的一种经典且易于解释的方法。

C. 经典累积和（*CUSUM*）程序

累积和（CUSUM）方法最初提出于 [69], ，是用于序贯变点检测的经典且强大的方法。其核心原理是累积分布变化的证据，并在这些证据足够强时宣告变点。在离散时间设置下，对于观测流 $\{x_1, x_2, \ldots\}$，时间 t 处的CUSUM统计量定义为

$$W_t^{\text{CUSUM}} = \max_{1 \leq k \leq t} \sum_{j=k}^{t} \log \frac{p_1(x_j)}{p_0(x_j)}, \qquad t = 1, 2, \ldots,$$

其中，$p_0(\cdot)$ 和 $p_1(\cdot)$ 分别表示变化前和变化后的概率密度函数。当统计量超过预先设定的阈值 γ 时，就会触发警报，从而得到停止时间

$$\nu = \inf\big\{t \geq 1 : W_t^{\text{CUSUM}} \geq \gamma\big\}.$$

在时空点过程设置中，观测以连续时间到达，因此可以在任意 $t \in \mathbb{R}_+$ 评估式（2）中的似然。假设空间变化区域 Ω 已知且 $\Omega = \mathcal{S}$ 时，离散时间的累积和统计量可自然地扩展到连续时间，得到 [16]:

$$W_t^{\text{CUSUM}} = \sup_{0 \leq \tau \leq t} \big[\ell_{1,\tau}(\mathcal{H}_t) - \ell_0(\mathcal{H}_t)\big], \tag{5}$$

其中，$\ell_{1,\ \tau}(\mathcal{H}_t)$ 表示在备择假设（式（3））下、假设变点为 τ 时截至时间 t 的观测历史的对数似然，而 $\ell_0(\mathcal{H}_t)$ 表示无变化这一原假设下的对数似然。

尽管经典累积和（CUSUM）框架颇具吸引力，但在我们的时空点过程设置中，它面临两个根本挑战：（i）变化区域 Ω 未知，是 $\mathcal{S}$ 的子集，必须与检测过程联合推断，因此变化后似然 $\ell_{1,\ \tau}$无法事先获得，必须搜索候选区域。（ii）空间维度的引入大幅增加了计算特定区域对数似然的负担，因此有必要开发在统计上可靠且在计算上可行的新检测统计量。

IV. 方法

本节针对式（3）中的时空变点检测问题，基于 [19] 的序贯得分框架，提出一种基于得分的CUSUM型统计量。我们的方法将局部得分近似与交替最小化过程相结合，以联合检测变点并定位受影响的空间区域。

A. 基于得分的时空累积和过程 *dure*

对于每个观测事件 x，我们将 $\psi_i(x)$定义为一种特定于状态的分数，用于评估 x 与状态 $i \in \{0, 1\}$的动态特征吻合得有多好。分数基于变换后的事件 $(t(x) - t_n,\ s(x))$ 进行评估，该事件编码了局部事件间隔动态。差值 $\Delta(x) := \psi_0(x) - \psi_1(x)$ 作为一种局部异常度量，为那些与变化后状态比与变化前状态更为相符的事件赋予正权重。得分函数 $\psi_i(x)$ 的具体构造与计算方法详见第四节-B。

该差异定义了时间 t的检测统计量：

$$W_t = \sup_{0 \leq \tau \leq t} \sup_{\Omega \subseteq \mathcal{S}} \int_{[\tau, t) \times \Omega} \Delta(x) \mathrm{d}\mathbb{N}(x). \tag{6}$$

该统计量汇总所有候选变点 τ 及空间区域 Ω 上的局部异常证据。上确界定位了两种状态之间差异最为显著的时空区域。与经典的式 (5) 中的累积和统计量不同，后者仅对离散时间索引取最大值，而式 (6) 对连续时间与空间进行联合优化，因此能够在任意时空尺度上进行检测。

一旦统计量超过预先设定的阈值 γ，便会发出警报，从而得到停止时间

$$\nu = \inf\{t : W_t \geq \gamma\}. \tag{7}$$

在停止时间，达到式（6）中 W_ν 的最大化集合（$\hat{}\tau$，$\hat{}\Omega$）提供了变点和受影响的空间区域的估计。

B. 局部得分近似

我们使用加权 Hyvärinen 评分构造 $\psi_i(x)$。加权 Hyvärinen 评分是一种严格适当评分规则，仅依赖于得分函数 $\nabla_x \log p_i(x)$，而非似然本身 [50]。这使其特别适用于时空点过程：对于这类过程，基于似然的方法需要计算难以处理的时空积分，同时该方法还能够在无需计算归一化常数 [59] 的情况下，高效地对事件动态进行非归一化建模。

形式化地，我们定义

$$\psi_i(x) := \left\| \sqrt{w(x)} \odot f_i(x) \right\|_2^2 + 2\,\mathrm{div}\big[w(x) \odot f_i(x)\big], \tag{8}$$

其中，$f_i(\cdot)$ 是一个学习得到的得分模型，用于捕捉 $\nabla_x \log p_i(x)$，而 $w(\cdot)$ 是一个加权函数。我们选择

$$w(x) := \min_{x' \in \mathcal{X}^c} \|x' - x\|_\infty,$$

其会降低截断域 $[0, T$边界附近区域的权重；第 C 节还给出了更一般的构造。这种加权至关重要，因为事件间隔时间的支持集是单侧的，违反了无加权得分匹配和分部积分论证所要求的正则性条件 [54]。

估计得分函数 f_i 的一个关键挑战在于，其计算依赖于未知的变化区域 Ω。回顾一下，事件 x 的条件密度具有如下形式：

$$p(x) = \lambda(x) \exp\left(- \int_{[t_n, t(x)) \times \mathcal{S}} \lambda(u) \mathrm{d}u \right), \tag{9}$$

其中，积分是在整个空间域 $\mathcal{S}$上进行的。当变化局限于未知区域 Ω 时，强度 $\lambda(u)$ 在 Ω 和 $\mathcal{S} \setminus \Omega$ 之间不同，使其成为未知的强度混合。因此，在不了解 Ω 的情况下，无法计算 $p_i(x)$ 中的积分。

为应对这一挑战，我们提出一种局部得分近似，仅利用局部事件历史来捕捉条件强度动态，而不对式（9）所隐含的全局得分函数 f_i 进行建模。这种局部构造有两大优势：(i) 它聚焦于最相关的邻近事件，从而保留对局部强度起主导作用的因素；(ii) 通过将关注范围限制在一个较小的邻域内，它避免纳入可能处于不同状态的远处区域，从而得到局部稳定的近似，更有可能反映单一的底层强度。形式化保证见第五节。

具体而言，设 $\delta > 0$ 为固定的空间尺度。对于任意事件 x，将其 ℓ_∞ 空间邻域定义为：

$$\mathcal{B}_\delta(x) := \{s' \in \mathcal{S} : \|s' - s(x)\|_\infty \leq \delta\}. \tag{10}$$

随后，在给定其局部历史的情况下，我们将 x 在状态 i 下的条件密度定义为：

$$p_i(x; \delta) = \lambda_i(x) \exp\left(- \int_{[t_n, t(x)) \times \mathcal{B}_\delta(x)} \lambda_i(u)\, \mathrm{d}u \right), \tag{11}$$

并将对应的得分模型记为 $f_i(x; \delta) := \nabla_x \log p_i(x; \delta)$。(11) 的形式化推导见附录中的引理 8。在我们的实现中，每个得分模型 $f_i(\cdot; \delta)$ 都由混合神经网络架构参数化，该架构由循环组件（例如循环神经网络）和前馈组件组成。随后，将这一表示与当前事件的协变量拼接，并输入前馈网络，由其输出相应的得分值。模型架构的示意图见 G 节。每个得分模型 $f_i(\cdot; \delta)$ 都通过去噪得分匹配 [19], [61] 进行学习，$\mathcal{D}_0$ 使用分别从变化前和变化后状态中抽取的离线参考数据集 $\mathcal{D}_1$ 。具体而言，我们最小化去噪得分匹配损失

$$L(x, f) := \left\| f(x + \epsilon; \delta) + \frac{\epsilon}{\sigma^2} \right\|_2^2, \quad \epsilon \sim \mathcal{N}(0, \sigma^2 I), \tag{12}$$

其中，ϵ 为独立同分布采样的噪声数据，$\sigma > 0$ 是用户指定的噪声水平，用于控制扰动的尺度。由此得到的得分估计量为

$$f_i = \arg\min_{f \in \mathcal{F}} \sum_{x \in \mathcal{D}_i} L(x, f), \quad i = 0, 1.$$

备注 1 讨论了在无法获取变点后参考数据的情况下如何扩展到在线设置。

C. 提出的算法：*ST-Score*

为了实现式 (6) 中提出的检测统计量，我们开发了一种交替优化算法，称为 ST-Score，用于求解由此得到的两阶段最大化问题。其核心思想是将对变点和空间区域的联合优化分解为两个相互耦合的子问题，并在 K 个训练轮次中迭代求解。

在第 k 次迭代中，给定当前对变点 $\hat{}\tau^{(k-1)}$ 的估计，我们首先通过求解以下问题来更新空间变化区域：

$$\hat{\Omega}^{(k)} = \arg\max_{\Omega \subseteq \mathcal{S}} \int_{[\hat{\tau}^{(k-1)},\, t) \times \Omega} \Delta(x)\, \mathrm{d}\mathbb{N}(x). \tag{13}$$

固定更新后的区域 $\hat{}\Omega^{(k)}$ ，然后通过求解以下问题来更新变点估计值

$$\hat{\tau}^{(k)} = \arg\max_{0 \leq \tau \leq t} \int_{[\tau,\, t) \times \hat{\Omega}^{(k)}} \Delta(x)\, \mathrm{d}\mathbb{N}(x). \tag{14}$$

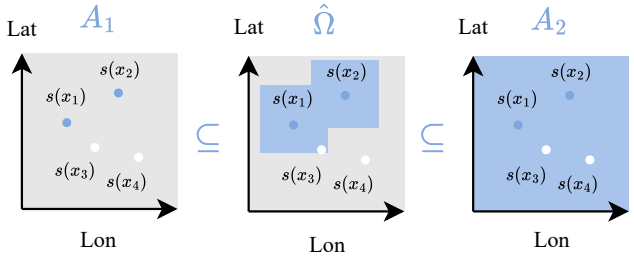

图2. 引理1的示意图。一个包含四个事件的数据集 $\{x_i\}_{i=1}^4$，其中 x_1 和 x_2 具有 $\Delta(x_i) > 0$，而 x_3 和 x_4 具有 $\Delta(x_i) < 0$。它们的空间坐标分别由蓝色和白色圆点表示。

经过 K 次迭代后，算法返回 $(\hat{\tau}^{(K)}, \hat{\Omega}^{(K)})$，作为变点和受影响的空间区域的最终估计值。接下来，我们将介绍求解 (13) 和 (14) 的高效过程，分别称为内层优化步骤（*I*步）和外层优化步骤 （*O* 步骤）。

*a) I*步：给定前一次迭代得到的$\hat{\tau}^{(k-1)}$，我们首先提出以下理论见解。

引理1. (13) 的解不唯一，可由以下集合族给出：

$$\hat{\Omega}^{(k)} \in \Big\{\Omega \subseteq \mathcal{S} : A_1 \subseteq \Omega \subseteq A_2\Big\},$$

其中 A_1、$A_2 \subset \mathcal{S}$ 定义如下

$$A_1 = \Big\{s(x) : x \in \mathcal{H}_t \setminus \mathcal{H}_{\hat{\tau}^{(k-1)}} \text{ and } \Delta(x) > 0\Big\},$$

$$A_2 = \mathcal{S} \setminus \Big\{s(x) : x \in \mathcal{H}_t \setminus \mathcal{H}_{\hat{\tau}^{(k-1)}} \text{ and } \Delta(x) < 0\Big\},$$

如图2所示，A_1 是一个有限的空间坐标集合，其对应的异常度量为正；A_2 则是整个空间 $\mathcal{S}$，但排除了一个有限的空间坐标集合，这些坐标对应的异常度量为负。尽管引理1指出(13)的解并不唯一，但在实践中，解的选择对数值稳定性有重要影响。为此，我们建议采用以下构造：

$$\hat{\Omega}^{(k)} := \bigcup_{s \in A_1} \{s' \in \mathcal{S} : \|s - s'\|_\infty \le \delta\} \setminus A_2^c, \qquad (15)$$

其中 δ 是用于计算$\hat{x}$的空间邻域的半径。注意，(15)易于计算，因为减号两侧均为具有有限元素的集合[1]。

b) O 步骤：给定$\hat{\Omega}^{(k)}$，外部问题(14)可简化为求解

$$\hat{\tau}^{(k)} = \underset{\tau \in \{t\} \cup \{t_j\}_{j=1}^{N(t)}}{\arg\max} \int_{[\tau, t) \times \hat{\Omega}^{(k)}} \Delta(x) \mathrm{d}\mathbb{N}(x), \qquad (16)$$

因为计数测度仅在有限点集处发生跳跃。这将问题简化为一个有限最大化问题，从而可以高效求解。完整算法总结于算法 1。

备注 1 （在线分数估计）。 所提出的分数估计可扩展到 $\mathcal{D}_1$ 不可用的情形，

[1]左侧是有限个球的并集，在几何上较为简单。

算法 1ST 分数检测过程

要求： 参考数据集 $\mathcal{D}_0$， $\mathcal{D}_1$；测试数据集流 $\mathcal{D}$；
优化轮数 K；阈值 γ。
1：训练分数模型 f_0 和 f_1。
2: **对于** $x_j \in \mathcal{D}$ **执行**
3: 根据 (8) 中的定义，计算 $\psi_0(x_j)$ 和 $\psi_1(x_j)$。
4: $\Delta(x_j) \leftarrow \psi_0(x_j) - \psi_1(x_j)$。
5: **结束循环**
6: 初始化 $t \leftarrow 0$ 和 $j = 1$。
7: **当** $W_t < \gamma$ **时执行**
8: $t \leftarrow t(x_j)$， $j \leftarrow j + 1$
9: 初始化 $\hat{\tau}^{(0)} \leftarrow 0$.
10: **对于** $k \in \{1, \ldots, K\}$，**执行**
11: $\hat{\Omega}^{(k)} \leftarrow$ 按式 (15) 计算变化区域。
12: $\hat{\tau}^{(k)} \leftarrow$ 按式 (16) 计算变点。
13: **结束 for 循环**
14: 设置 $\hat{\tau} \leftarrow \hat{\tau}^{(K)}$ 和 $\hat{\Omega} \leftarrow \hat{\Omega}^{(K)}$。
15：计算时空累积和统计量：

$$W_t \leftarrow \int_{[\hat{\tau}, t) \times \hat{\Omega}} \Delta(x) \mathrm{d}\mathbb{N}(x).$$

16: **结束 while 循环**
17：设置停止时间 $\nu \leftarrow t$。18：**返回** 估计（$\hat{\tau}$，$\hat{\Omega}$）；停止时间 ν。

需要对变后得分函数进行在线估计。具体而言，我们使用变前历史数据 $\mathcal{D}_0$ 同时初始化变前与变后分数模型。在检测阶段，对于每个新到达的数据 x_j，连同先前观测到的数据流，我们通过运行交替优化算法（第四节-*C*） K 个训练轮次，估计变点 $\hat{\tau}_j^{(K)}$ 以及变化区域 $\hat{\Omega}_j^{(K)}$。随后，通过单步（或多步）梯度下降更新变化后模型：

$$\theta \leftarrow \theta - \eta \cdot \nabla_\theta \int_{[\hat{\tau}_j^{(K)}, t) \times \hat{\Omega}_j^{(K)}} L(x, f_1) \mathrm{d}\mathbb{N}(x), \qquad (17)$$

其中 θ 表示变后分数模型f_1的参数， $\eta > 0$ 是学习率超参数，并且 $L(\cdot)$ 是在(12)中定义的去噪得分匹配损失。最后，利用更新后的变化后模型，为下一个到来的测试数据计算得分规则 $\psi_i(\cdot)$ 和异常度量 $\Delta(\cdot)$。伪代码见附录中的算法 *2*。

五、理论结果

本节从虚警性能、检测延迟和变化区域定位准确率三个方面，为所提出的算法建立理论保证。为简化起见，我们假设学得的得分模型是精确的，即， $f_i(x; \delta) = \nabla_x \log p_i(x; \delta)$。请注意，当得分模型类别具有足够的表达能力且使用充足的数据进行训练时，这一假设是可行的。我们将 $\mathbb{E}_\tau$ 记为变化发生在 τ时所对应分布下的期望。所有证明均置于附录中。

首先，我们分析所提出算法的虚警性能。我们采用平均运行长度（ARL）这一指标，它定义为变化前域 $\mathbb{E}_\infty[\nu]$ 下停止时间的期望值。在下面的引理 2 中，我们证明平均运行长度（ARL）的下界为检测阈值的指数函数乘以一个常数。因此，为满足给定的 ARL 下界要求，只需设置$\gamma \asymp \log(\mathrm{ARL}\cdot 2^{\mu_0\cdot \mathrm{T}})$。

引理 2（误报率）。考虑（7）中定义的停止规则 ν，则对于任意 $\gamma > 0$,

$$\mathbb{E}_\infty[\nu] \geq \min\left\{O\left(\frac{e^\gamma}{\mathbb{E}_\infty[2^{N(T)}]}\right), T\right\}.$$

为量化所提出算法的检测效率，我们首先推导检测统计量的漂移，该漂移刻画了变化后 ST-Score 统计量的增长率。我们的理论表明，该项类似于 Fisher 散度的加权局部化版本。

引理 3（漂移）。定义

$$D_F(f_0\|f_1) \coloneqq \mathbb{E}_0\left\|\sqrt{w(x)}\odot\left[f_0(x;\delta)-f_1(x;\delta)\right]\right\|_2^2,$$

$$D_F(f_1\|f_0) \coloneqq \mathbb{E}_\infty\left\|\sqrt{w(x)}\odot\left[f_1(x;\delta)-f_0(x;\delta)\right]\right\|_2^2,$$

然后，在关于[2] w的某些假设下（x），有

$$\mathbb{E}_\infty\left[\Delta(x)\right] = -D_F(f_1\|f_0), \quad \mathbb{E}_0\left[\Delta(x)\right] = D_F(f_0\|f_1).$$

由引理 3 可以看出，统计量的漂移在变化前状态下为负，而在变化后状态下为正。这证实了所提出的得分规则是恰当的，并且能够检测变化。此外，在对权重 w（x）和强度函数 λ（x）作出假设的情况下，引理 3 可用于推导漂移的进一步显式形式。例如，在霍克斯过程假设（4）下，我们有以下一般引理。

引理 4（霍克斯过程的评分差异）。假设λ（x）定义如下：（4），则

$$f_1(x;\delta)-f_0(x;\delta) = (\mu_0-\mu_1)\cdot\left[\frac{\alpha\boldsymbol{k}(x)}{\lambda_1(x)\lambda_0(x)}+\begin{pmatrix}4\delta^2\\ \mathbf{0}_2\end{pmatrix}\right].$$

其中 $\boldsymbol{k}(x)$ 是一个定义如下的向量：

$$\boldsymbol{k}(x) = \int_{[0,t_n)\times\mathcal{B}(x)}\left[\nabla_x\kappa(x,x')+\nabla_{x'}\kappa(x,x')\right]\mathrm{d}\mathbb{N}(x'),$$

引理 4 还使我们能够在核函数 $\kappa(\cdot,\cdot)$ 取特定形式时，对漂移项进行明确刻画。特别地，当 κ 为径向函数时，项 $\boldsymbol{k}(x)$ 恒等于零，从而得到简化的漂移表达式。我们总结了两种典型情形下的结果形式：(*i*) 齐次泊松过程（HPP；例 1）和 (*ii*) 具有指数核的 Hawkes过程（例 2）。

推论 1. 设 $w(x)\coloneqq x$。我们有如下结果。

1) 如果 $\lambda(x)$是泊松过程（例 *1*），

$$D_F(f_{1-i}\|f_i) = \frac{4(\mu_0-\mu_1)^2\delta^2}{\mu_i},$$

2) 如果 $\lambda(x)$ 是具有指数激励核的霍克斯过程（例 *2*），

$$D_F(f_{1-i}\|f_i) = \left[\frac{4(\mu_0-\mu_1)^2\delta^2}{\mu_i}\right](1-\alpha).$$

推论 1 有两点值得强调：(*i*) 对于齐次泊松过程，漂移与基准率之差$|\mu_0-\mu_1|$ 和邻域半径 δ成比例。这表明，变化前后的强度差异越明显、观测区域越大，变化的证据就越强。(*ii*) 对于具有指数核的霍克斯过程，与齐次泊松过程情形相比，漂移幅度会乘以因子 $1-\alpha$。这表明，分支比 α越大（对应于更强的自激和更大的时间变异性），变化的证据就越不明显，因为观测到的偏差可能会与内生突发事件或偏离平稳性的自发变化相混淆。

最后，我们利用引理 3 推导所提出算法的检测性能保证。我们考察两项指标：期望检测延迟（EDD） [16], [47] 定义为 $\mathbb{E}_\tau[\nu-\tau \mid \nu>\tau]$，用于评估估计的变点；以及估计的变化区域的杰卡德指数，定义为：$\hat{}\,\hat{}\,\hat{}J(\Omega,\ \Omega) = |\Omega\cap\Omega|/|\Omega\cup\Omega|$，用于评估恢复出的变化区域 $\Omega\hat{}$ 的准确性。

定理1。在满足某些正则性条件的情况下（*F* 节）：

1）对于给定的阈值 $\gamma>0$ 以及任意真实变点 $\tau\geq 0$，有：

$$\mathit{EDD} \lesssim \min\left\{\frac{\gamma|\mathcal{S}|}{\bar{\lambda}_1|\Omega|D_F(f_0\|f_1)}(1+o(1)),\ T\right\}$$

¯其中 λ_1 定义为变化后过程的平稳强度。

2）杰卡德指数渐近下界为：

$$\lim_{\gamma\to\infty} J(\hat{\Omega},\Omega) \geq \min\left\{\frac{|\Omega|}{|\Omega\oplus\mathcal{B}_\delta|},\ \frac{|\Omega\ominus\mathcal{B}_\delta|}{|\Omega|}\right\},$$

其中 $\oplus$ 和 $\ominus$ 分别表示集合膨胀算子和腐蚀算子（(25)和 (26)），$\mathcal{B}_\delta$ 的定义见(10)。

定理1的第一个陈述给出了 EDD 的渐近上界。与标准的变点检测结果不同，该上界与以下两者均呈负相关：¯平稳的变化后过程强度 λ_1 以及变化区域的大小。这种依赖关系源于统计量W_t 通过底层计数过程累积信息，因此检测延迟取决于观测到变化后事件的速率。定理1的第二个陈述表明，若留出充足的时间供统计量累积证据，估计的变化区域将与真实变化区域相差一个定位邻域 $\mathcal{B}_\delta$。产生这一差距的原因是，跨越 Ω 边界的局部区域包含混合的变化前和变化后过程，这会妨碍应用引理 3 来确立明确的

[2]参见附录中引理 7 的式 (20)。

收敛。

六、数值实验

我们在合成数据和两个真实世界数据集上评估所提出的算法 ST-Score 的性能：地震目录数据集 [70] 和野火事件数据集[71]。在所有实验中，ST-Score 使用如下参数化的得分模型实现：(*i*) 一个含有 32 个隐藏单元的单层 LSTM网络，用于编码事件历史；随后接入 (*ii*) 一个含有 512 个隐藏单元的单层前馈网络，其输入为历史嵌入与当前观测事件的拼接，输出分数。通过模拟确定检测阈值，以控制平均运行长度 [16], [19]。更多详情请参见 G 节。

A. 模拟结果

合成事件序列由定义在时间范围$[0,1)$和单位二维空间盒 $[0,1]\times[0,1]$上的时空霍克斯过程 (4) 生成。激发核 $\kappa(x,\ x')$被指定为指数核函数，默认参数为 $\beta = 0.1$。我们将真实变化点设为 $\tau = 0.5$ ，真实变化区域设为 $\Omega = [0.4,0.6]\times[0.4,0.6]$。变化前后的基准率分别设为 $\mu_0 = 1\times 10^2$ 和 $\mu_1 = 10\times\mu_1$。此配置既产生适度的时空溢出效应，又确保变化可被检测。 f_0 和 f_1 的训练数据分别由变化前过程和变化后过程各生成的一条数据流组成。

1) 与基线方法比较：我们将ST-Score的检测性能与五个基线进行比较：(*i*) CUSUM：第三节中讨论的标准CUSUM统计量 [69]；(*ii*) SCUSUM： [62]中提出的基于得分的CUSUM统计量；(*iii*) PP CUSUM：以(5)中定义的估计点过程对数似然比作为异常度量的CUSUM统计量；(*iv*) MinCUSUM(n)： [25]中提出的用于多流变化检测的CUSUM统计量，并以 $n\times n$ 均匀空间离散化作为其通道。除PP-CUSUM外，所有基线均基于由事件序列导出的预处理后的分箱计数数据进行运算，因为它们无法直接处理时空事件数据。MinCUSUM还要求通过人工划分将数据分为多个通道。需要注意的是，SCUSUM和PP CUSUM未纳入杰卡德指数的比较，因为它们估计的变化区域默认为整个空间区域，从而会退化为CUSUM给出的结果。纳入这些基线是为了反映在第三节-B所述问题设定下处理变点检测时通常采用的方法。

图3展示了 ST-Score 与上述基线方法相比的性能指标与平均运行长度（ARL）权衡图。报告的结果是对100次试验取平均后得到的。在所有 ARL 值和超参数设置下，与其他基线方法相比，ST-Score 始终表现出更低的期望检测延迟（EDD）和更高的杰卡德指数。这体现了所提出算法的优越性，其优势得益于检测机制、统计设计和优化过程的结合，我们在前文中介绍了这一点。顺便指出，与无时空溢出效应的设置（$\alpha = 0$）相比，在强时空溢出效应（$\alpha = 0.9$）下，所有模型的 EDD 始终更高，杰卡德指数也更低。这一观察结果与推论 1 一致，也为溢出效应会对检测性能造成负面影响提供了进一步证据。除了性能优越之外，ST-Score 还可直接处理原始事件序列，无需预处理；与大多数基线方法不同，这进一步凸显了它在时空变化检测任务中的适用性。

此外，图5展示了不同合成实验设置下估计变化区域的时间演变。累积和始终将变化区域估计为整个空间域，这并不意外，因为累积和（以及许多其他标准变点检测方法）仅用于发出时间警报，因此会将检测到的变化视为全局变化，而非空间定位的变化。与累积和相比，两个 MinCUSUM 变体能够定位更有意义的受影响区域，但其定位仍较粗略，因为估计的变化区域呈矩形，且与真实变化区域 Ω 不对齐。这一局限源于将时空变点检测问题转化为多流变点检测问题时依赖人工进行空间离散化；在缺乏变化区域几何形状的先验信息时，这可能导致显著的定位误差。相比之下，在所有设置中，我们的方法 ST-Score 都能在变化发生后准确描绘真实变化区域的轮廓，即使受影响区域不规则或不连通等具有挑战性的情形下也不例外，而且无需获取有关真实变化区域形状的先验信息。这解释了该方法在图3中的杰卡德指数表现为何更优，也凸显了 ST-Score 在变化区域连续、复杂且难以通过预定义空间分区准确表示的情形下的独特适用性。

2) 关于局部半径的消融研究 δ： 如图4所示，我们研究了局部半径 δ 对检测性能的影响。检测延迟曲线呈现出大致的凸趋势，而 Jaccard指数曲线则表现出大致的凹形模式。这反映了选择 δ 时存在一个基本权衡：较小的 δ 所形成的邻域过于稀疏，无法提供具有信息量的信号；而较大的邻域则可能将变化前状态和变化后状态中的事件混合在一起，从而污染得分函数和检测统计量，二者都会降低检测性能。在当前的合成设置下，根据图4， $\delta = 0.1$ 是较为均衡的半径选择，并用于我们的整个合成实验。

3) 迭代次数与基准率的消融研究：我们还针对不同优化轮数 K 和变化前基准率 μ_0 对检测性能的影响进行了消融研究。变化前基准率可作为输入算法的测试样本预期数量的替代指标。表 I 展示了在不同优化迭代次数 $K = 1$、3、5 和基准率 $\mu_0 = 10^1$、10_2、10_3下，基于 ST-Score 评估的性能指标。对于 μ_0的不同取值，EDD 和杰卡德指数都几乎保持不变，这表明尽管所提出的统计量仅在离散事件时刻更新，其检测性能

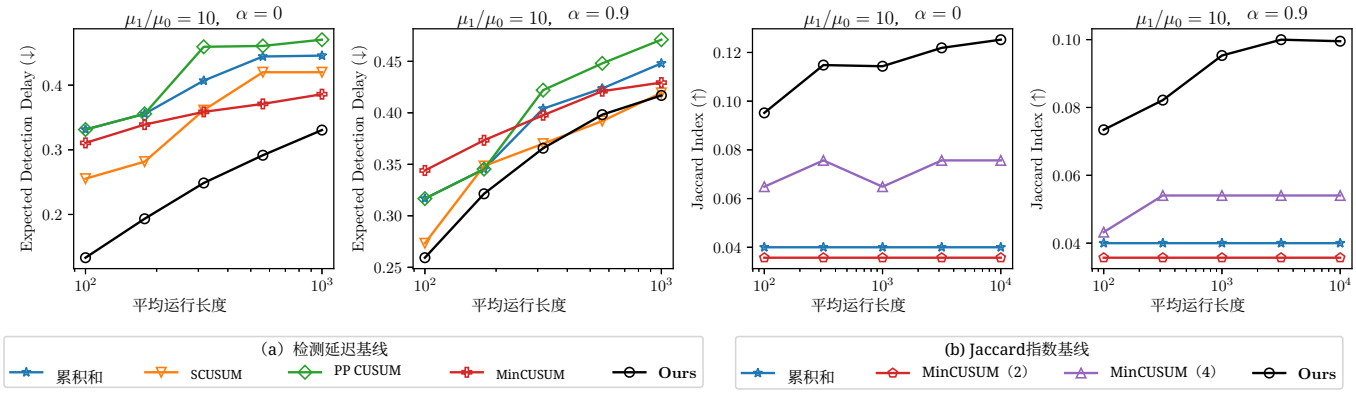

图 3. 不同设置下检测性能与 平均运行长度（ARL）的关系。前两幅图展示期望检测延迟与 ARL 的关系；后两幅图展示 Jaccard 指数与 ARL 的关系。在这些图中，我们将分支比 α 的取值设为 0 到 0.9。

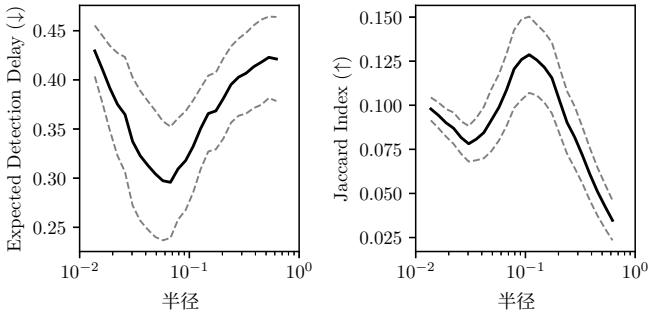

图 4. 不同邻域半径 (δ) 下的模型性能。实线表示均值；虚线表示均值标准误。

表 IST−SCORE 在不同 K 和 μ_0 规范下的性能。

		EDD	Jaccard 指数	运行时间（秒）
$K=1$	$\mu_0=10^1$	0.39 ± 0.05	0.04 ± 0.01	0.01 ± 0.00
	$\mu_0=10^2$	0.39 ± 0.05	0.04 ± 0.01	0.03 ± 0.00
	$\mu_0=10^3$	0.39 ± 0.05	0.04 ± 0.01	0.06 ± 0.00
$K=3$	$\mu_0=10^1$	0.38 ± 0.04	0.11 ± 0.02	0.03 ± 0.00
	$\mu_0=10^2$	0.38 ± 0.04	0.11 ± 0.02	0.08 ± 0.01
	$\mu_0=10^3$	0.38 ± 0.04	0.11 ± 0.02	0.14 ± 0.00
$K=5$	$\mu_0=10^1$	0.15 ± 0.04	0.09 ± 0.01	0.13 ± 0.01
	$\mu_0=10^2$	0.15 ± 0.04	0.09 ± 0.01	0.29 ± 0.01
	$\mu_0=10^3$	0.15 ± 0.04	0.09 ± 0.01	0.45 ± 0.01

对事件频率仍具有稳健性，这与我们对定理1中推导的期望检测延迟上界的观察不同。同时，增加 K 可持续改善 EDD 和杰卡德指数，说明为优化过程提供充足的迭代次数有助于其收敛。最后，尽管增加 K 或 μ_0 会提高计算成本，但每次试验的运行时间始终不超过一秒，表明该方法在大规模数据集上的计算可行性。这种高效率源于统计量的基于分数的设计，它避免了成本高昂的数值积分。

B. 真实数据实验

我们将 ST-Score 应用于两个真实世界数据集，以展示其实际效用。在这些实验中，我们不假设能够获得来自变化后状态的历史数据，而是采用备注 1 中详述的算法相应扩展，利用流数据按序估计并更新变后分数模型。

1) 地震预警： 我们使用日本气象厅（JMA）的日本能登地区地震目录 [70], , 涵盖 2018 年至 2024 年期间。在此期间，该地区发生了两次重大地震：2023 M_j 6.5 事件和 2024 M_j 7.6 事件。地震学研究已确定 2021—2023 年为群震期[72], , 其特征是地震活动异常增加，可能预示着地震风险升高。我们的目标是利用 ST-Score 自动识别群震期，以支持地震预警。

图 6（上）展示了所得的 ST-Score 检测统计量，该统计量从 2021 年末开始呈现显著的上升趋势，随后于 2022 年年中再次大幅上升。这些时间点可能标志着后续重大地震背后的两次地震活动，并表明 ST-Score 能够捕捉群震期的变化信号。图 6（下）展示了变化区域估计结果的两个代表性快照，分别对应变化前状态和变化后状态。可以看出，在进入群震期之前，ST-Score 无法指出主震区域。然而，群震期过后，ST-Score 清晰地定位了主震区域。这些结果表明，ST-Score 不仅可用作有效的地震预警和定位工具，其发现也为地震学研究中的地震前兆群震理论提供了有效证据。

2）野火检测：我们使用加利福尼亚州林业与消防局公开发布的损害检查数据集 [71]。该数据集记录了野火周边 100 米范围内受影响的建筑物，并包含事件时间、地理坐标和损害程度等属性。我们重点研究于 2025年1月7日在洛杉矶地区燃起的帕利塞兹野火，该火灾造成了重大

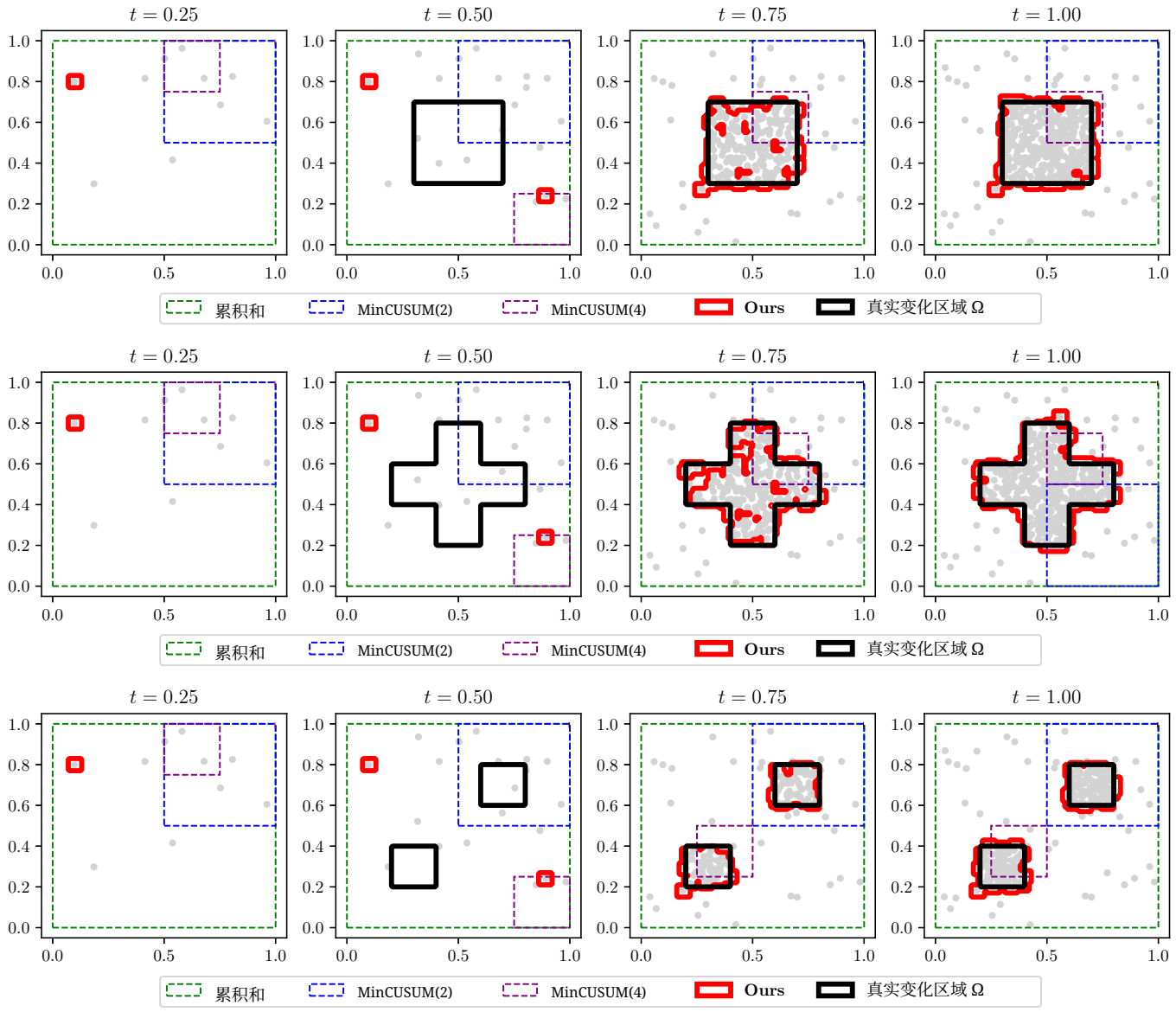

图5. 不同基线下估计的变化区域的时间演变，每个快照对应于将当前时间视为停止时间。我们假设真实变化点为 $\nu = 0.5$。第一行沿用现有设置，最后两行则将真实变化区域改为更复杂的形状（十字形和双峰）。灰点表示事件的空间坐标。

社会混乱，其广泛的破坏和负面社会影响给社会带来了严重冲击。与之前的地震场景类似，我们的目标是利用该数据集模拟帕利塞兹野火的自动化早期检测场景。

图7展示了野火案例研究的结果。尽管检测统计量在监测期末呈现明显的上升漂移，但变化点恰好出现在野火起燃当天，而非更早的时间。这表明，与地震场景中群震期可作为有意义的前兆不同，在野火场景中，ST-Score未能识别出可检测的前兆信号。这种表现符合野火起燃的特点：野火往往突然发生，甚至由人为引发，因此事件数据中几乎没有或完全没有预警信号。虽然起燃前时期观测到的野火热点源自另一场无关的野火事件，但ST-Score仍正确追踪了起燃后的估计变化区域。这一结果表明，野火发生后可部署ST-Score，以便快速进行起燃后定位并及时预警。

七、结论

本文提出一种基于得分的CUSUM统计量，用于时空点过程中的时空变化检测，其中假设变化发生在未知的连续时间点，并出现在连续空间区域内。通过结合去噪得分匹配、加权Hyvärinen得分和空间定位，我们构造了一个定义良好的

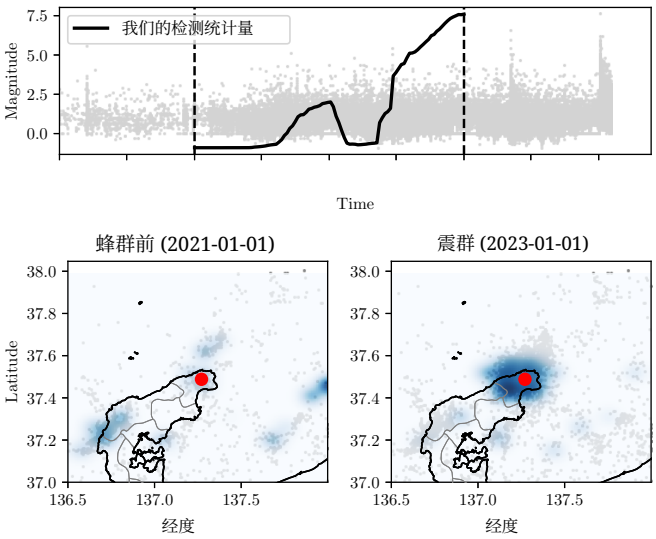

图 6. ST-Score 在地震数据集上的实现结果。灰点表示事件点。上图：黑线表示归一化检测统计量。下方两图：红点表示主震震中，蓝色阴影区域表示平均估计变化区域。

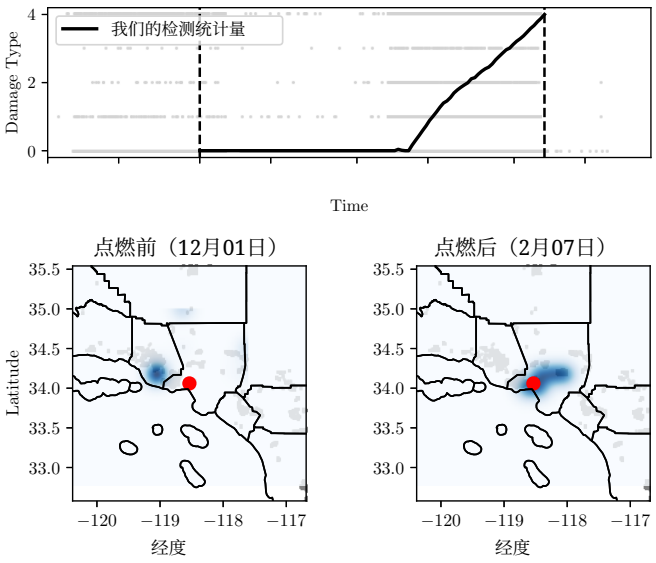

图 7. ST-Score 在野火数据集上的实现结果。灰点表示事件点。上图：黑线表示归一化检测统计量。下方两图：红点表示帕利塞兹火灾震中，蓝色阴影区域表示平均估计变化区域。

检测统计量，无需显式积分强度函数，从而相较于传统的基于似然的方法具有显著的计算优势。为高效计算该统计量，我们针对由此产生的嵌套最大化问题设计了一种替代优化方案，从而能够在连续时间与空间上进行可扩展的优化。在理论方面，我们建立了关于虚警率的性能保证，刻画了漂移的参数形式，推导了期望检测延迟的界，并通过杰卡德指数分析了估计变化区域的准确性。通过大量合成实验和两个真实案例研究，我们证明了所提方法优于现有的时空变化检测基线方法，并对其行为和部署提供了实践见解。

未来可以探索几个有前景的方向研究。（i）我们目前的理论分析刻画了在无核激发或采用指数核这两种假设下漂移的参数形式。专门的理论研究可以将这一分析扩展到更广泛的激发核类别，从而更全面地理解核类型对统计量检测性能的影响。（ii）根据经验观察，所提出算法的性能取决于定位半径 δ 和优化迭代次数 K。未来的工作可以通过研究这些超参数与检测性能之间的理论联系，探索一种有原则的超参数配置方法；也可以开发一种校准算法，在实际应用中自动完成超参数选择。

参考文献

[1] H. V. Poor 和 O. Hadjiliadis，《Quickest Detection》。剑桥大学出版社，2008年。 [2] L. Xie、Y. Xie 和 G. V. Moustakides，“地震震颤的异步多传感器变点检测”，载于《IEEE国际信息论研讨会》，第787–791页，2019年。 [3] J. Shi 和 S. Zhou，“多阶段系统的质量控制与改进：综述”，《IIE Transactions》，第41卷，第9期，第744–753页，2009年。 [4] K. Yamin、H. Wang、B. Montreuil 和 Y. Xie，“使用霍克斯过程的序贯变点检测在线检测供应链网络中断”，《arXiv预印本 arXiv:2211.12091》，2022年。 [5] A. Reinhart，“自激时空点过程及其应用综述”，《Statistical Science》，第33卷，318，2018 3，299第。页 – 。[6] D. J. Daley 和 D. Vere-Jones，《点过程理论导论：第一卷：基础理论与方法》。施普林格科学与商业媒体，2006年。 [7] T. Omi、Y. Ogata、Y. Hirata 和 K. Aihara，“根据早期余震序列估计ETAS模型”，《Geophysical Research Letters》，第41卷，第3期，第850–857页，2014年。 [8] C. Xu、Y. Xie、D. A. Z. Vazquez、R. Yao 和 F. Qiu，“利用多模态数据进行时空野火预测”，《IEEE信息论选题期刊》，第4卷，第302–313页，2023年。 [9] J. Park、F. P. Schoenberg、A. L. Bertozzi 和 P. J. Brantingham，“利用带协变量的时空点过程研究帮派相关暴力犯罪数据中的聚类与暴力干预”，《美国统计学会会刊》，第116卷，第536期，第1674–1687页，2021年。 [10] S. Zhu 和 Y. Xie，“用于犯罪关联检测的时空文本点过程”，《The Annals of Applied Statistics》，第16卷，第2期，第1151–1170页，2022年。 [11] S. Zhu、R. Ding、M. Zhang、P. Van Hentenryck 和 Y. Xie，“用于交通拥堵事件建模的注意力时空点过程”，《IEEE智能交通系统汇刊》，第23卷，第7期，第7298–7309页，2021年。 [12] K. Liu、Y. Mei 和 J. Shi，“用于在线高维过程监测的自适应采样策略”，《Technometrics》，第57卷，第3期，第305–319页，2015年。 [13] J. Horálek、T. Fischer、P. Einarsson 和 S. S. Jakobsdótir，“地震震群”，载于《地震工程百科全书》，第871–885页，施普林格，2015年。 [14] R. Romero-Calcerrada、C. Novillo、J. D. Millington 和 I. Gomez-Jimenez，“马德里西南部（西班牙中部）人为野火点燃风险空间模式的GIS分析”，《Landscape Ecology》，第23卷，第3期，第341–354页，2008年。 [15] M. A. Khan、A. Gupta、P. Sharma 和 A. Roy，“利用GIS集成AHP方法调查野火风险并绘制其分布图：印度中部霍尚阿巴德森林分区案例研究”，《Environment, Development and Sustainability》，第1–35页，2024年。 [16] H. Wang、L. Xie、Y. Xie、A. Cuozzo 和 S. Mak，“相互激发点过程的序贯变点检测”，《Technometrics》，第65卷，第1期，第44–56页，2023年。 [17] Z. Dong、S. Zhu、Y. Xie、J. Mateu 和 F. J. Rodríguez-Cortés，“高分辨率COVID-19数据的非平稳时空点过程建模”，《英国皇家统计学会会刊C辑：应用统计学》，第72卷，第2期，第368–386页，2023年。

[18] I. V. Nikiforov，“一类序贯检验中检测/隔离延迟的下界”，IEEE信息论汇刊，第49卷，第11期，第3037–3047页，2003年。 [19] W. Zhou、L. Xie、Z. Peng 和 S. Zhu，“通过去噪得分匹配进行序贯变点检测”，载于第61届Allerton通信、控制与计算会议论文集，2025年。 [20] D. Siegmund，《序贯分析：检验与置信区间》。Springer统计学丛书，Springer-Verlag，纽约，1985年。 [21] A. Tartakovsky、I. Nikiforov 和 M. Basseville，《序贯分析：假设检验与变点检测》。统计学与应用概率专著丛书第136卷。博卡拉顿、伦敦、纽约：Chapman & Hall/CRC Press，Taylor & Francis Group，2015年。 [22] T. L. Lai，“序贯分析：若干经典问题与新挑战”，《中华统计学志》，第11卷，第2期，第303–351页，2001年。 [23] L. Xie、S. Zou、Y. Xie 和 V. V. Veeravalli，“序贯（最快）变点检测：经典成果与新方向”，IEEE信息论选题期刊，第2卷，第2期，第494–514页，2021年。 [24] Y. Xie 和 D. Siegmund，“序贯多传感器变点检测”，《统计年鉴》，第41卷，第2期，第670–692页，2013年。 [25] Y. Mei，“用于监测大量数据流的高效可扩展方案”，《Biometrika》，第97卷，第2期，第419–433页，2010年。 [26] Y. Wang 和 Y. Mei，“通过收缩变点后估计进行大规模多流最快变点检测”，IEEE信息论汇刊，第61卷，第12期，第6926–6938页，2015年。 [27] H. P. Chan，“多流数据的最优序贯检测”，《The Annals of Statistics》，第45卷，第6期，第2736–2763页，2017年。 [28] L. Xie 、 Y. Xie 和 G. V. Moustakides，“ 序贯子空间变点检测 ”，《序贯分析》，第39卷，第3期，第307–335页，2020年。 [29] H. Yan、K. Paynabar 和 J. Shi，“通过时空平滑稀疏分解实时监测高维函数型数据流”，《Technometrics》，第60卷，第2期，第181–197页，2018年。 [30] H. Keshavarz 、 G. Michailidis 和 Y. Atchadé，“ 高维高斯图模型的序贯变点检测”，《机器学习研究杂志》，第21卷，第1期，第3125–3181页，2020年。 [31] Y. Chen、T. Wang 和 R. J. Samworth，“高维多尺度在线变点检测”，《英国皇家统计学会会刊B辑：统计方法学》，第84卷，第1期，第234–266页，2022年。 [32] I. V. Nikiforov，“广义变化检测问题”，IEEE信息论汇刊，第41卷，第1期，第171–187页，1995年。 [33] I. V. Nikiforov，“用于诊断随机信号突变的简单递归算法”，IEEE信息论汇刊，第46卷，第7期，第2740–2746页，2000年。 [34] A. G. Tartakovsky，“多决策最快变点检测：既有成果与未解问题”，《序贯分析》，第27卷，第2期，第201–231页，2008年。 [35] A. G. Tartakovsky，《序贯变点检测与假设检验：一般非独立同分布随机模型及渐近最优规则》。统计学与应用概率专著丛书第165卷。博卡拉顿、伦敦、纽约：Chapman & Hall/CRC Press，Taylor & Francis Group，2020年。 [36] D. J. Daley 和 D. Vere-Jones，《点过程理论导论：第二卷：一般理论与结构》。施普林格科学与商业媒体，2007年。 [37] M. Farajtabar、Y. Wang、M. Gomez-Rodriguez、S. Li、H. Zha 和 L. Song，“COEVOLVE：用于信息扩散和网络演化的联合点过程模型”，《机器学习研究杂志》，第18卷，第1期，第1305–1353页，2017年。 [38] I. M. Toke，“订单簿模型中的做市行为及其对买卖价差的影响”，arXiv:1003.3796，2010年。 [39] Y. Ogata，“地震发生的统计模型与点过程的残差分析”，《美国统计学会会刊》，第83卷，第401期，第9–27页，1988年。 [40] E. W. Fox、M. B. Short、F. P. Schoenberg、K. D. Coronges 和 A. L. Bertozzi，“利用自激点过程建模电子邮件网络并推断领导力”，《美国统计学会会刊》，第111卷，第514期，第564–584页，2016年。 [41] X. He、T. Rekatsinas、J. Foulds、L. Getoor 和 Y. Liu，“HawkesTopic：基于文本级联进行网络推断和主题建模的联合模型”，载于国际机器学习会议，2015年。 [42] J. J. Shen 和 N. R. Zhang，“非齐次泊松过程的变点模型及其在下一代DNA测序拷贝数分析中的应用”，《The Annals of Applied Statistics》，第6卷，第360期，第476–496页，2012年。 [43] N. R. Zhang、B. Yakir、C. L. Xia 和 D. O. Siegmund，“扫描泊松随机场以检测局部信号”，《The Annals of Applied Statistics》，第10卷，第74期，第726–755页，2016年。 [44] T. Herberts 和 U. Jensen，“不同观测方案下泊松过程中变点的最优检测”，《斯堪的纳维亚统计学杂志》，第 31 卷，第 3 期，第 347–366 页，2004 年。[45] M. Ludkovski，“基于观测-变点反馈的贝叶斯最快检测”，载于 IEEE年度决策与控制会议（CDC）论文集，第 166–171 页，2012 年。[46] J. C. L. Pinto、T. Chahed 和 E. Altman，“利用霍克斯过程进行社交网络趋势检测”，载于 IEEE/ACM社交网络分析与挖掘进展国际会议论文集，第 1441–1448 页，2015 年。[47] S. Li、Y. Xie、M. Farajtabar、A. Verma 和 L. Song，“检测网络动态事件中的变化”，《IEEE信号与信息处理网络汇刊》，第 3 卷，第 2 期，第 346–359 页，2017 年。[48] E. Bayraktar 和 H. V. Poor，“至少两个泊松失序时刻的最快检测”，《SIAM控制与优化杂志》，第 46 卷，第 1 期，第 308–331 页，2007 年。[49] H. Zhang、O. Hadjiliadis、T. Schäfer 和 H. V. Poor，“耦合系统中的最快检测”，《SIAM控制与优化杂志》，第 52 卷，第 3 期，第 1567–1596 页，2014 年。[50] A. Hyvärinen 和 P. Dayan，“通过得分匹配估计非归一化统计模型”，《机器学习研究杂志》，第 6 卷，第 4 期，2005 年。[51] C. Meng、L. Yu、Y. Song、J. Song 和 S. Ermon，“自回归得分匹配”，《神经信息处理系统进展》，第 33 卷，第 6673–6683 页，2020 年。[52] A. Hyvärinen，“得分匹配的一些扩展”，《计算统计学与数据分析》，第 51 卷，第 5 期，第 2499–2512 页，2007 年。[53] S. Yu、M. Drton 和 A. Shojaie，“针对非负数据的广义得分匹配”，《机器学习研究杂志》，第 20 卷，第 76 期，第 1–70 页，2019 年。[54] S. Liu、T. Kanamori 和 D. J. Williams，“利用得分匹配估计具有截断边界的密度模型”，《机器学习研究杂志》，第 23 卷，第 186 期，第 1–38 页，2022 年。[55] P. Vincent，“得分匹配与去噪自编码器之间的联系”，《神经计算》，第 23 卷，第 7 期，第 1661–1674 页，2011 年。[56] Y. Song 和 S. Ermon，“通过估计数据分布的梯度进行生成建模”，《神经信息处理系统进展》，第 32 卷，2019 年。[57] Y. Song、J. Sohl-Dickstein、D. P. Kingma、A. Kumar、S. Ermon 和 B. Poole，“通过随机微分方程进行基于得分的生成建模”，arXiv预印本 arXiv:2011.13456，2020 年。[58] J. Ho、A. Jain 和 P. Abbeel，“去噪扩散概率模型”，《神经信息处理系统进展》，第 33 卷，第 6840–6851 页，2020 年。[59] Z. Li、Y. Xu、S. Zuo、H. Jiang、C. Zhang、T. Zhao 和 H. Zha，“Smurf-thp：基于得分匹配的Transformer霍克斯过程不确定性量化”，载于国际机器学习会议，第 20210–20220 页，PMLR，2023 年。[60] H. Cao、Z. Meng、T. Ke 和 F. Zhou，“得分匹配适用于点过程估计吗？”，arXiv预印本 arXiv:2410.04037，2024 年。[61] Z. Dong、Z. Fan 和 S. Zhu，“高维带标记时间点过程的条件生成建模”，2025 年。[62] S. Wu、E. Diao、T. Banerjee、J. Ding 和 V. Tarokh，“非归一化统计模型的最快变点检测”，《IEEE信息论汇刊》，第 70 卷，第 2 期，第 1220–1232 页，2024 年。[63] T. Banerjee 和 V. Tarokh，“针对非归一化模型和基于得分模型的贝叶斯最快变点检测”，《序贯分析》，第 43 卷，第 3 期，第 359–378 页，2024 年。[64] S. Moushegian、S. Wu、E. Diao、J. Ding、T. Banerjee 和 V. Tarokh，“稳健的基于得分的最快变点检测”，《IEEE信息论汇刊》，2025 年。[65] S. Moushegian、T. Banerjee 和 V. Tarokh，“基于扩散的假设检验与变点检测”，arXiv预印本 arXiv:2506 16089，2025 年。。[66] W. Chen、T. Banerjee 和 V. Tarokh，“针对马尔可夫转移核的最快变点检测条件得分学习”，arXiv预印本 arXiv:2511.03953，2025 年。[67] W. Chen、S. Moushegian、V. Tarokh 和 T. Banerjee，“多流信号的基于得分的最快变点检测与故障识别”，arXiv预印本 arXiv:2511.03967，2025 年。[68] A. G. Hawkes 和 D. Oakes，“自激过程的簇过程表示”，《应用概率杂志》，第 11 卷，第 3 期，第 493–503 页，1974 年。[69] E. S. Page，“连续检查方案”，《Biometrika》，第 41 卷，第 1/2 期，第 100–115 页

，1954。

[70] 日本气象厅，《日本气象厅地震目录：能登半岛地区》。https://www.data.jma.go.jp/svd/eqev/data/bulletin/hypo. html，2024。访问时间：2026年1月。[71] 加利福尼亚州林业与消防局，《加州消防局损害检查（DINS）数据》。https://data.ca.gov/dataset/cal-fire-damage-inspection-dins-data，2026。访问时间：2026-01-24。[72] Z. Peng 等，《日本中部能登半岛下方震群与2024年能登半岛M7.6地震序列之间的演化过程》，《地震研究进展》，第5卷，第1期，页码100332，2025年。[73] J. E. Marsden 和 A. Tromba，《向量微积分》。麦克米伦，2003年。[74] T. L. Lai，《随机系统中参数变化的信息界与快速检测》，*IEEE*信息论汇刊，第44卷，第7期，页码2917–2929，1998年。

附录

在所有证明中，我们固定 δ，并为简洁起见省略对 δ 的依赖。

A. 引理1的证明

证明。 我们首先定义辅助函数：

$$L(s) \coloneqq \int_{[\hat{\tau}^{(k-1)},T)} \Delta(x)\mathrm{d}\mathbb{N}(t)$$

其中， $\mathbb{N}(t) \coloneqq \mathbb{N}([0,T)\times\mathcal{S})$ 表示时间维度的边缘计数过程。然后，(14) 可重写为：

$$\hat{\Omega}^{(k)} = \arg\max_{\Omega\subseteq\mathcal{S}} \int_{\Omega} L(s)\mathrm{d}\mathbb{N}(t). \tag{18}$$

当该集合映射到所有正（且非零）的被积函数时，积分达到最大值，

$$\hat{\Omega} \in \left\{\Omega \subseteq \mathcal{S} : L^{-1}(\mathbb{R}_+) \subseteq \Omega \subseteq L^{-1}(\mathbb{R}_+ \cup \{0\})\right\}, \tag{19}$$

其中 $L^{-1}(\cdot)$ 定义为 L 的原像集，即，对于任意$\mathcal{I} \subseteq \mathbb{R}$，存在

$$L^{-1}(\mathcal{I}) \coloneqq \{s \in \mathcal{S} : L(s) \in \mathcal{I}\}.$$

根据 $L(\cdot)$的定义，存在

$$L^{-1}(\mathbb{R}_+) = \left\{s(x) : x \in \mathcal{D}' \text{ and } \Delta(x) > 0\right\},$$
$$L^{-1}(\mathbb{R}_+ \cup \{0\}) = \mathcal{S} \setminus \left\{s(x) : x \in \mathcal{D}' \text{ and } \Delta(x) < 0\right\},$$

为简化记号，我们记 $\mathcal{D}' = \mathcal{H}_t \setminus \mathcal{H}_{\hat{\tau}^{(k-1)}}$。注意，满足 $\Delta(x)<0$ 和 $\Delta(x)>0$ 的两组空间坐标均为有限集，因为 $\mathcal{D}$ 是有限的。因此，证明得证。 □

B. 引理 2 的证明

我们首先陈述两个引理，它们证明了具有良好理论性质的有用代理过程的存在性。这些引理将作为推导该界的中间步骤。

引理 5. 记以下过程为

$$\tilde{W}^{\Omega}_{u:t} \coloneqq \int_{[u,t)\times\Omega} \left\{\zeta\cdot\Delta(x) - \log\left(\mathbb{E}_{\infty}[e^{\zeta\cdot\Delta(x)}]\right)\right\}\mathrm{d}\mathbb{N}(x).$$

那么，对于任意 $\zeta > 0$、 $u < t$ 和 $\Omega \subseteq \mathcal{S}$，$\exp(\tilde{W}^{\Omega}_{t})$ 是一个非负的 $\mathbb{P}_{\infty}$-鞅，且均值为 *1*。

引理 *5* 的证明。 在上下文明确时，我们将 $\tilde{W}^{\Omega}_{u:t}$ 简记为 $\tilde{W}_t$。对于任意 $\zeta>0$、 $u<t$ 和 $\Omega \subseteq \mathcal{S}$，其非负性显而易见。为证明其是一个 $\mathbb{P}_{\infty}$-鞅，注意对于任意 $\tau \in [0,T)$，

$$\begin{aligned}
&\mathbb{E}_{\infty}[e^{\tilde{W}_t}|\mathcal{H}_\tau] = \mathbb{E}_{\infty}[e^{\tilde{W}_{t'}}|\mathcal{H}_\tau] \\
&\times \mathbb{E}_{\infty}\left[\exp\left(\int_{[t',t)\times\Omega} \zeta\cdot\Delta(x)\mathrm{d}\mathbb{N}(x)\right)|\mathcal{H}_\tau\right] \\
&\div \mathbb{E}_{\infty}\left[\exp\left(\int_{[t',t)\times\Omega} \log\left(\mathbb{E}_{\infty}[e^{\zeta\cdot\Delta(x)}]\right)\mathrm{d}\mathbb{N}(x)\right)|\mathcal{H}_\tau\right] \\
&= \mathbb{E}_{\infty}[e^{\tilde{W}_{t'}}|\mathcal{H}_\tau] \times \mathbb{E}_{\infty}[e^{\zeta\cdot\Delta(x)}] \div \mathbb{E}_{\infty}[e^{\zeta\cdot\Delta(x)}] \\
&= \mathbb{E}_{\infty}[e^{\tilde{W}_{t'}}|\mathcal{H}_\tau],
\end{aligned}$$

其中 t' 表示 t之前的最后一个事件时间戳。我们可以˜接着对 $\mathbb{E}_{\infty}[e^{W_{t'}}|\mathcal{H}_\tau]$ 递归，每次展开一个最后事件，从而推导出

$$\mathbb{E}_{\infty}[e^{\tilde{W}_t}|\mathcal{H}_\tau] = \mathbb{E}_{\infty}[e^{\tilde{W}_{t'}}|\mathcal{H}_\tau] = \ldots = e^{\tilde{W}_\tau}.$$

这证明了它是一个 $\mathbb{P}_{\infty}$-鞅。要推导其均值，只需注意到

$$\mathbb{E}_{\infty}[e^{\tilde{W}_t}] = \mathbb{E}_{\infty}[e^{\tilde{W}_u}] = e^0 = 1$$

证明完毕。 □

引理 6. 令 $\tilde{W}^{\Omega}_{u:t}$ 如引理 *5* 中所定义。考虑以下过程：

$$M^{\Omega}_t = \int_0^t \left\{\exp(\tilde{W}^{\Omega}_{u:t}) - 1\right\}\mathrm{d}u.$$

在上下文明确时，我们将 M^{Ω}_t简记为 M_t。那么，对于任意 Ω， M_t 是一个 $\mathbb{P}_{\infty}$-鞅，均值为0。

证明。 注意，对于任意 $t'<t$，有如下递归关系：

$$\begin{aligned}
M_t &= \int_0^t \left\{\exp(\tilde{W}^{\Omega}_{u:t}) - 1\right\}\mathrm{d}u \\
&= \int_0^{t'} \exp(\tilde{W}^{\Omega}_{u:t})\mathrm{d}u + \int_{t'}^t \exp(\tilde{W}^{\Omega}_{u:t})\mathrm{d}u - t.
\end{aligned}$$

第一项可以进一步分解为

$$\begin{aligned}
\int_0^{t'} \exp(\tilde{W}^{\Omega}_{u:t})\mathrm{d}u &= \int_0^{t'} \exp(\tilde{W}^{\Omega}_{u:t'})\cdot\exp(\tilde{W}^{\Omega}_{t':t})\mathrm{d}u \\
&= M_{t'}\cdot\exp(\tilde{W}^{\Omega}_{t':t}).
\end{aligned}$$

因此，有

$$\begin{aligned}
\mathbb{E}_{\infty}[M_t \mid \mathcal{H}_{t'}] &= \int_0^{t'} \exp(\tilde{W}^{\Omega}_{u:t'})\mathrm{d}u\cdot\mathbb{E}_{\infty}[\exp(\tilde{W}^{\Omega}_{t':t}) \mid \mathcal{H}_{t'}] \\
&\quad + \int_{t'}^t \mathbb{E}_{\infty}[\exp(\tilde{W}^{\Omega}_{u:t}) \mid \mathcal{H}_{t'}]\mathrm{d}u - t \\
&= \int_0^{t'} \exp(\tilde{W}^{\Omega}_{u:t'})\mathrm{d}u - t' = M_{t'},
\end{aligned}$$

其中，第二个等号由引理 5 得出。因此，我们已经证明 M_t 是一个 $\mathbb{P}_{\infty}$-鞅。为求其均值，只需注意到

$$\mathbb{E}_{\infty}[M_t] = \mathbb{E}_{\infty}[M_0] = 0.$$

证明至此完成。 □

现在，我们给出引理 2 的完整证明。该证明改编自 [67] 中定理 V.2 的证明。

证明。 给定空间区域 $\Omega \subseteq$ 上的先验分布 π

$\mathcal{S}$，注意以下关系：

$$\begin{aligned}W_t &= \sup_{\Omega\subseteq\mathcal{S}}\sup_{0\le u<t}\int_{[u,t)\times\Omega}\Delta(x)\mathrm{d}\mathbb{N}(x)\\ &\le \sup_{\Omega\subseteq\mathcal{S}}\sup_{0\le u<t}\int_{[u,t)\times\Omega}\Delta(x)-\frac{1}{\zeta}\log\left(\mathbb{E}_\infty[e^{\zeta\cdot\Delta(x)}]\right)\mathrm{d}\mathbb{N}(x)\\ &= \sup_{\Omega\subseteq\mathcal{S}}\sup_{0\le u<t}\frac{1}{\zeta}\tilde{W}^{\Omega}_{u:t}\le\sum_{\Omega\in 2^{\mathcal{S}}}\int_{[0,t)}\frac{1}{\zeta}\tilde{W}^{\Omega}_{u:t}\mathrm{d}u.\end{aligned}$$

将与第 i 个事件关联的 Voronoi 单元记为

$$\mathcal{V}_i=\{s\in\mathcal{S}:\|s-s(x_i)\|_2\le\|s-s(x_j)\|_2,\quad\forall j\neq i\}.$$

接下来，我们定义由截至时间 t观察到的事件所能形成的所有 Voronoi 区域的集合：

$$\mathcal{V}(t)=\left\{\bigcup_{i\in\mathcal{I}}\mathcal{V}_i:\mathcal{I}\subseteq\{1,\dots,N(t)\}\right\}.$$

那么，引理1意味着以下等价关系：

$$\sum_{\Omega\in 2^{\mathcal{S}}}\int_{[0,t)}\frac{1}{\zeta}\tilde{W}^{\Omega}_{u:t}\mathrm{d}u=\sum_{\Omega\in\mathcal{V}(t)}\int_{[0,t)}\frac{1}{\zeta}\tilde{W}^{\Omega}_{u:t}\mathrm{d}u$$

定义停止时间

$$\tilde{\nu}\coloneqq\inf\left\{t\in[0,T):\sum_{\Omega\in\mathcal{V}(t)}\int_{[0,t)}\exp(\tilde{W}^{\Omega}_{u:t})\mathrm{d}u\ge e^{\gamma\zeta}\right\},$$

那么必有 $\nu\ge\tilde{\nu}$，因此我们只需对$\mathbb{E}_\infty[\tilde{\nu}]$给出下界。考虑以下过程：

$$\tilde{M}_t\coloneqq\sum_{\Omega\in\mathcal{V}(t)}\underbrace{\int_0^t\left\{\exp(\tilde{W}^{\Omega}_{u:t})-1\right\}\mathrm{d}u}_{=M_t}$$

˜根据引理 6，我们知道 M_t 和 M_t 都是均值为 0 的 $\mathbb{P}_\infty$鞅。对 M_t应用可选抽样定理˜，可得

$$0=\mathbb{E}_\infty\left[\tilde{M}_0\right]=\mathbb{E}_\infty\left[\tilde{M}_{\tilde{\nu}}\right]\ge e^{\gamma\zeta}-\mathbb{E}_\infty\left[\tilde{\nu}\cdot 2^{N(T)}\right],$$

其中，最后一个不等式由 $\tilde{\nu}$ 的定义得出。因此，重新整理各项可得

$$\mathbb{E}_\infty[\nu]\le\mathbb{E}_\infty[\tilde{\nu}]=\frac{e^{\gamma\zeta}}{\mathbb{E}_\infty[2^{N(T)}]}=O\left(\frac{e^{\gamma}}{\mathbb{E}_\infty[2^{N(T)}]}\right).$$

这里省略了对常数 ζ 和 C 的依赖关系。注意到，根据定义，平均运行长度不能大于时间范围 T，由此即可完成证明。□

C. 引理 3 的证明

为了证明引理 3，我们首先陈述一个改编自 [54] 的引理。条件（20）要求加权函数必须在边界项处平滑概率密度函数。请注意，这并不要求概率密度函数本身在其边界处趋于零，因此该条件比关于分部积分的常规假设 [50] 更弱。

引理 7 (定理 2 of [54])。 对于任意两个定义在 $\times\otimes_{k=2}^{d-1}[a_k,b_k]$ 上的连续概率密度函数$p(x)$ 和 $q(x)$，假设加权函数 $w(x)$ 满足

$$\begin{aligned}\lim_{x_1\to 0^+,x_1\to+\infty}w_k(x)p(x)\partial_k\log q(x)=0\\ \lim_{x_k\to a_k,x_k\to b_k}w_k(x)p(x)\partial_k\log q(x)=0,\quad\forall k\ge 2,\end{aligned}\tag{20}$$

那么， $\mathbb{E}_{x\sim p}$ 散度 [（$w\odot\nabla$对数 q）（x）] 可改写为

$$-\mathbb{E}_{x\sim p}\langle(w^{1/2}\odot\nabla\log q)(x),(w^{1/2}\odot\nabla\log p)(x)\rangle,$$

其中， ∇ 表示相对于 x 求得的梯度。

证明。记 $\mathcal{X}_k$ 为其第 k 个边缘化维度，记 $L(\cdot)$ $U(\cdot)$：$2^{\mathbb{R}}\to\mathbb{R}$ 为提取集合上界和下界的映射，并记 x_{-k} 为移除其第 k 个条目后的向量 x 。我们作如下展开：

$$\begin{aligned}&\mathbb{E}_{x\sim p}\operatorname{div}[(w\odot\log q)(x)]\\ &=\int_{\mathcal{X}}p(x)\sum_{k=1}^{d}\partial_k(w_k(x)\partial_k\log q(x))\mathrm{d}x\\ &=\sum_{k=1}^{d}\int_{\mathcal{X}_{-k}}\int_{\mathcal{X}_k}p(x)\partial_k(w_k(x)\partial_k\log q(x))\mathrm{d}x_k\mathrm{d}x_{-k}\\ &=\sum_{k=1}^{d}\int_{\mathcal{X}_{-k}}\left[p(x)w_k(x)\partial_k\log q(x)\Big|_{L(\mathcal{X}_k)}^{U(\mathcal{X}_k)}\right.\\ &\quad\left.-\int_{\mathcal{X}_k}\partial_k p(x)w_k(x)\partial_k\log q(x)\mathrm{d}x_k\right]\mathrm{d}x_{-k}\\ &=-\sum_{k=1}^{d}\int_{\mathcal{X}_{-k}}\int_{\mathcal{X}_k}\partial_k p(x)w_k(x)\partial_k\log q(x)\mathrm{d}x_k\mathrm{d}x_{-k}\\ &=-\int_{\mathcal{X}}\sum_{k=1}^{d}\partial_k p(x)w_k(x)\partial_k\log q(x)\mathrm{d}x\\ &=-\int_{\mathcal{X}}\sum_{k=1}^{d}\partial_k\log p(x)\cdot p(x)\cdot w_k(x)\partial_k\log q(x)\mathrm{d}x\\ &=-\mathbb{E}_{x\sim p}\left[\sum_{k=1}^{d}\partial_k\log p(x)w_k(x)\partial_k\log q(x)\right]\\ &=-\mathbb{E}_{x\sim p}\langle(w^{1/2}\odot\nabla\log q)(x),(w^{1/2}\odot\nabla\log p)(x)\rangle.\end{aligned}$$

其中，第一个等式根据散度算子的定义成立；第三个等式利用分部积分技巧得到；第四个等式由（20）得出；第六个等式是由于$\partial_k\log p(x)=\partial_k p(x)/p(x)$。证明完毕。□

引理 *3* 的证明。 我们仅证明 $\mathbb{E}_0[\Delta(x)]$, 的情形， $\mathbb{E}_1[\Delta(x)]$的证明类似。根据 $\Delta(x)$的定义，存在

$$\mathbb{E}_0[\Delta(x)]=\mathbb{E}_0[\psi_0(x)-\psi_1(x)].$$

根据（8），每一项 $\mathbb{E}_0[\psi_i(x)]$ 均等于：

$$\mathbb{E}_0\left\|\sqrt{w(x)}\odot f_i(x)\right\|_2^2+\underbrace{2\cdot\mathbb{E}_0\operatorname{div}\left[w(x)\odot f_i(x)\right]}_{(A)}$$

利用引理 7，取 $\nabla\log q\leftarrow f_i$和 $\nabla\log p\leftarrow f_0$，

$$(A)=-2\cdot\mathbb{E}_0\left\langle\underbrace{\sqrt{w(x)}\odot f_i(x)}_{(B)},\underbrace{\sqrt{w(x)}\odot f_0(x)}_{(C)}\right\rangle.$$

因此，存在

$$\begin{aligned}\mathbb{E}_0[\psi_i(x)] &= \mathbb{E}_0\left[\|(B)\|_2^2 - 2\langle(B),(C)\rangle + \|(C)\|_2^2 - \|(C)\|_2^2\right] \\ &= \mathbb{E}_0\left[\left\|\sqrt{w(x)}\odot[f_i(x)-f_0(x)]\right\|_2^2\right] - \mathbb{E}_0\|(C)\|_2^2.\end{aligned}$$

类似地，当 $i=0$ 时，第一项为零；由于对于这两个 $i=0,1$, $\mathbb{E}_0[\psi_i(x)]$ 共享第二项，因此该项会相互抵消。因此，有

$$\mathbb{E}_0[\Delta(x)] = \mathbb{E}_0\left[\left\|\sqrt{w(x)}\odot[f_1(x)-f_0(x)]\right\|_2^2\right].$$

与上述证明类似，也可以推导出

$$\mathbb{E}_1[\Delta(x)] = -\mathbb{E}_1\left[\left\|\sqrt{w(x)}\odot[f_1(x)-f_0(x)]\right\|_2^2\right].$$

按照引理 3 中的定义，令 $D_F(f_0\|f_1)$ 和 $D_F(f_1\|f_0)$，即可完成证明。 □

D. 引理 4 的证明

证明。 根据引理 3，我们先展开 $f_i(x)$:

$$f_i(x) = \underbrace{\nabla_x \log\lambda_i(x)}_{(A)} - \underbrace{\nabla_x \int_{[t_n,t)\times\mathcal{B}(x)} \lambda_i(x')\mathrm{d}x'}_{(B)}.$$

我们分别分析这两项。对于第一项:

$$(A) = \nabla_x\lambda_i(x)/\lambda_i(x) = \frac{\alpha\nabla_x\int_{[0,t_n)\times\mathcal{B}(x)}\kappa(x,x')\mathrm{d}\mathbb{N}(x')}{\lambda_i(x)},$$

这里用到了基准率为常数这一事实。由于$\mathcal{B}(x)$是由 x平移而得的半径为 δ 的 ℓ_p-范数球，因此可以利用雷诺输运定理 [73] 进一步展开右侧:

$$(A) = \frac{\alpha\int_{[0,t_n)\times\mathcal{B}(x)}\nabla_x\kappa(x,x') + \nabla_{x'}\kappa(x,x')\mathrm{d}\mathbb{N}(x')}{\lambda_i(x)}.$$

对于项（B），其第一个（时间）分量为:

$$\begin{aligned}[(B)]_1 &= \partial_t\int_{[t_n,t)\times\mathcal{B}(x)}\lambda_i(x')\mathrm{d}x' = \int_{\mathcal{B}(x)}\lambda_i(\tilde{x}')\mathrm{d}s' \\ &= \mu_i\cdot|\mathcal{B}_\delta| + \underbrace{\alpha\int_{\mathcal{B}(x)}\int_{[0,t_n)\times\mathcal{B}(x)}\kappa(\tilde{x}',x'')\mathrm{d}\mathbb{N}(x'')\mathrm{d}s'}_{(C)},\end{aligned}$$

为方便记号，我们记 $\tilde{x}' := (t,\ s')$。其余（空间）分量为:

$$\begin{aligned}[(B)]_{2:d} &= \nabla_s\int_{[t_n,t)\times\mathcal{B}(x)}\lambda_i(x')\mathrm{d}x' \\ &= \int_{[t_n,t)\times\mathcal{B}(x)}\nabla_{s'}\lambda_i(x')\mathrm{d}s' \\ &= \underbrace{\alpha\int_{[t_n,t)\times\mathcal{B}(x)}\int_{[0,t_n)\times\mathcal{B}(x)}\nabla_{s'}\kappa(x',x'')\mathrm{d}\mathbb{N}(x'')\mathrm{d}s'}_{(D)},\end{aligned}$$

其中，第二个等号也用到了雷诺输运定理。由于 (C) 和 (D) 与 i 无关（κ 是共有的），因此它们在 $f_1(x)-f_0(x)$ 中相互抵消。将 (A) 的分子记为

$$\boldsymbol{k}(x) = \int_{[0,t_n)\times\mathcal{B}(x)}\nabla_x\kappa(x,x') + \nabla_{x'}\kappa(x,x')\mathrm{d}\mathbb{N}(x'),$$

根据变化前过程与变化后过程具有相同形式的核 $\kappa(\cdot,\ \cdot)$这一事实，可得:

$$\frac{1}{\lambda_1(x)} - \frac{1}{\lambda_0(x)} = \frac{\mu_0-\mu_1}{\lambda_1(x)\lambda_0(x)},$$

因此，我们得出:

$$f_1(x) - f_0(x) = (\mu_0-\mu_1)\cdot\left[\frac{\alpha\boldsymbol{k}(x)}{\lambda_1(x)\lambda_0(x)} + \begin{pmatrix}4\delta^2\\ \mathbf{0}_2\end{pmatrix}\right].$$

□

E. 推论 1 的证明

我们为这两种情形分别给出证明。

齐次泊松过程情形的证明。 由于引理 4 中定义的向量项 $\boldsymbol{k}(x)$ 在齐次泊松过程假设下为零，因此，

$$[f_1(x)-f_0(x)]_1 = (\mu_0-\mu_1)\cdot|\mathcal{B}|,$$

该向量的所有其他元素也均为零。根据引理 3,

$$\begin{aligned}D_F(f_i\|f_{1-i}) &= \mathbb{E}_i\left\|\sqrt{w(x)}\odot(f_0(x)-f_1(x))\right\|_2^2 \\ &= \mathbb{E}_i[t-t_n]\cdot([f_0(x)-f_1(x)]_1)^2 \\ &= \frac{1}{\mu_i\cdot|\mathcal{B}|}\cdot(\mu_0-\mu_1)^2\cdot|\mathcal{B}|^2 \\ &= \frac{(\mu_0-\mu_1)^2\cdot|\mathcal{B}|}{\mu_i},\end{aligned}$$

其中第三个等式源于齐次泊松过程的到达间隔时间。因此，我们得出

$$\mathbb{E}_i\left[\Delta(x)\right] = \mathrm{sgn}\{i=1\}\cdot\frac{(\mu_0-\mu_1)^2|\mathcal{B}|}{\mu_i}.$$

□

核情形的证明。 我们首先注意到，在指数核下， $\boldsymbol{k}(x)$ 为零。事实上，对于所有平移不变的 核，这一性质更一般地成立。具体而言，回顾一下，若核 $\kappa:\ \mathcal{X}\times\mathcal{X}\to\mathbb{R}$ 满足

$$\kappa(x+a,x'+a) = \kappa(x,x') \quad \forall x,x',a\in\mathcal{X}. \tag{21}$$

等价地，存在一个函数 $\varphi:\mathcal{X}\to\mathbb{R}$ 使得

$$\kappa(x,x') = \varphi(x-x').$$

在这种表示下，应用链式法则可得

$$\nabla_x\kappa(x,x') = \nabla\varphi(x-x'),\quad \nabla_{x'}\kappa(x,x') = -\nabla\varphi(x-x').$$

因此，

$$\nabla_x\kappa(x,x') + \nabla_{x'}\kappa(x,x') \equiv 0,$$

由此证明了该论断。广泛的一类平移不变核都满足条件 (21)，包括指数核、高斯（径向）核，以及 ETAS 模型中使用的空间触发核。

回到主要证明，我们再次得到

$$[f_1(x) - f_0(x)]_1 = (\mu_0 - \mu_1) \cdot |\mathcal{B}|.$$

利用具有指数激励核的 Hawkes 过程事件间隔时间期望的著名结果,

$$\mathbb{E}_i[t - t_n] = \frac{1-\alpha}{\mu_i \cdot |\mathcal{B}|},$$

我们以类似于第一条结论证明的方法进行推导，得到

$$\mathbb{E}_i\left[\Delta(x)\right] = \mathrm{sgn}\{i=1\} \cdot \frac{(\mu_0 - \mu_1)^2 |\mathcal{B}|}{\mu_i} \cdot (1-\alpha)\,.$$

□

F. 定理*1*的证明

由于我们的设定涉及非独立同分布数据，我们调整定理4 [74] 的证明思路，以推导我们的 EDD 界。按照 [74], , 我们对统计量的信息率作出如下假设:

假设1。 在由 λ_1诱导的概率测度下,

$$\lim_{t\to+\infty} \operatorname*{esssup}_{\mathcal{H}_\tau} \frac{1}{N(t)} \int_{[\tau,t)\times\Omega} \Delta(x)\mathrm{d}\mathbb{N}(x) \xrightarrow{p} \mathbb{E}_1[\Delta(x)], \quad (22)$$

$$\textit{and} \quad \frac{N(t)}{t} \xrightarrow{p} \bar{\lambda}_1 \frac{|\Omega|}{|\mathcal{S}|}. \quad (23)$$

假设1指出，在变化后状态中，从渐近意义上看，每个变化后事件平均贡献大致相同数量的证据（22），且这些事件的数量大致随时间线性增长，平均增长率稳定（23）。这两点都可视为对计数过程的一种大数定律型集中要求；由于中心极限定理，它们对齐次泊松过程显然成立。也有研究 [68] 表明，在满足对其自激核的某些假设（如平稳性和轻尾性）时，霍克斯过程也具有类似性质。

期望检测延迟的证明。固定任意一个变化点 $\tau \in [0,T)$，并在由 λ_1 诱导的变化后概率测度下进行分析。为简化记号，记

$$I \coloneqq \mathbb{E}_1[\Delta(x)] = D_F(f_0 \| f_1) > 0, \quad r \coloneqq \bar{\lambda}_1 \frac{|\Omega|}{|\mathcal{S}|}.$$

步骤*1*：构造代理过程：我们定义一个（未经最大化的）累积证据过程

$$\tilde{W}_{0:t} \coloneqq \int_{[0,t)\times\Omega} \Delta(x)\, d\mathbb{N}(x),$$

及其阈值穿越时间

$$\tilde{\nu} \coloneqq \inf\big\{t \in [0,T) : \tilde{W}_{0:t} \geq \gamma\big\}.$$

根据 W_t 是对时间窗口和 ˜ 空间区域取上确界的定义，对于所有 t，均有 $W_t \geq W_{0:t}$，因此

$$\nu = \inf\{t : W_t \geq \gamma\} \leq \inf\{t : \tilde{W}_{0:t} \geq \gamma\} = \tilde{\nu},$$

因此只需给出 $\mathbb{E}_1[\tilde{\nu}]$ 的界，因为 $\mathbb{E}_1[\nu] \leq \mathbb{E}_1[\tilde{\nu}]$。

步骤 *2*：单块下溢概率。：固定 $\varepsilon \in (0, I)$并定义块长度

$$t' \coloneqq \frac{\gamma}{(I-\varepsilon)\, r}.$$

˜注意到 $\{\tilde{\nu} > t'\} \subseteq \{W_{0:t'} < \gamma\}$，因此

$$\begin{aligned}\mathbb{P}_1(\tilde{\nu} > t') &\leq \mathbb{P}_1\left(\tilde{W}_{0:t'} < \gamma\right) \\ &= \mathbb{P}_1\left(\frac{\tilde{W}_{0:t'}}{t'} < \frac{\gamma}{t'}\right) = \mathbb{P}_1\left(\frac{\tilde{W}_{0:t'}}{t'} < (I-\varepsilon)r\right).\end{aligned}$$

根据假设1，我们有如下依概率收敛关系

$$\begin{aligned}\frac{\tilde{W}_{0:t}}{t} &= \frac{1}{t}\int_{[0,t)\times\Omega} \Delta(x)\, d\mathbb{N}(x) \\ &= \Big(\frac{1}{N(t)}\int_{[0,t)\times\Omega} \Delta(x) d\mathbb{N}(x)\Big) \cdot \Big(\frac{N(t)}{t}\Big) \xrightarrow{p} Ir,\end{aligned}$$

其中，乘积的收敛由斯卢茨基定理得出。因此，

$$\mathbb{P}_1(\tilde{\nu} > t') = o(1) \qquad \text{as } \gamma \to \infty.$$

此外，由(22)–(23)中的esssup$_{\mathcal{H}_\tau}$ 形式，同一论证给出一致条件版本：存在$\gamma_0(\varepsilon)$，使得对所有 $\gamma \geq \gamma_0(\varepsilon)$,

$$c \coloneqq \sup_{u\geq 0} \operatorname*{esssup}_{\mathcal{H}_u} \mathbb{P}_1\left(\tilde{W}_{u:u+t'} < \gamma \;\middle|\; \mathcal{H}_u\right) < 1.$$

此外，当 $\gamma \to \infty$时还存在 $c = c(\gamma, \varepsilon) \to 0$ 。

步骤*3*：跨块的几何尾界：对于整数$j \geq 1$，利用塔式性质以及 c的定义，

$$\begin{aligned}\mathbb{P}_1(\tilde{\nu} > jt') &= \mathbb{E}_1\left[\mathbf{1}\{\tilde{\nu} > (j-1)t'\}\, \mathbb{P}_1\left(\tilde{\nu} > jt' \mid \mathcal{H}_{(j-1)t'}\right)\right] \\ &\leq \mathbb{E}_1\Big[\mathbf{1}\{\tilde{\nu} > (j-1)t'\} \times \\ &\qquad \mathbb{P}_1\left(\tilde{W}_{(j-1)t':jt'} < \gamma \;\middle|\; \mathcal{H}_{(j-1)t'}\right)\Big] \\ &\leq c\, \mathbb{P}_1(\tilde{\nu} > (j-1)t') \leq \ldots \leq c^j,\end{aligned}$$

其中最后一个不等式由递归得出。

步骤*4*：尾积分恒等式与期望界：利用尾积分恒等式并将时间划分为若干块,

$$\begin{aligned}\mathbb{E}_1[\tilde{\nu}] &= \int_0^\infty \mathbb{P}_1(\tilde{\nu} > t)\, dt = \sum_{j=0}^{\infty} \int_{jt'}^{(j+1)t'} \mathbb{P}_1(\tilde{\nu} > t)\, dt \\ &\leq \sum_{j=0}^{\infty} t'\, \mathbb{P}_1(\tilde{\nu} > jt') \leq t' \sum_{j=0}^{\infty} c^j = \frac{t'}{1-c}.\end{aligned}$$

由于 $c \to 0$ 当 $\gamma \to \infty$时，我们有 $(1-c)^{-1} = 1 + o(1)$，因此

$$\mathbb{E}_1[\tilde{\nu}] \leq \frac{\gamma}{(I-\varepsilon)\, r}\,(1 + o(1)).$$

令 $\varepsilon \downarrow 0$ ，得到

$$\mathbb{E}_1[\tilde{\nu}] \leq \frac{\gamma}{Ir}\,(1+o(1)) = \frac{\gamma\,|\mathcal{S}|}{\bar{\lambda}_1 |\Omega|\, D_F(f_0\|f_1)}\,(1+o(1)).$$

最后，由于 $\nu \leq \tilde{\nu}$，我们得出结论

$$\mathbb{E}_1[\nu] \leq \frac{\gamma\,|\mathcal{S}|}{\bar{\lambda}_1 |\Omega|\, D_F(f_0\|f_1)}\,(1+o(1)),$$

并注意到，EDD 必须始终不超过时间范围 T，由此得到所需的 EDD 界。 □

为证明杰卡德指数的相关结论，我们需要以下引理。我们将由 $\mathcal{B}_\delta(x)$ 形成的局部化过程称为 $\mathcal{B}_\delta(x)$-过程，并将该过程中发生的下一个事件称为下一个 $\mathcal{B}_\delta(x)$-事件。

引理 8。 给定定义在 $\mathcal{B}_\delta(x)$上的强度 $\lambda(x)$，下一个 $\mathcal{B}_\delta(x)$-事件的条件概率密度函数定义为：

$$p(x;\delta)=\lambda(x)\exp\left(-\int_{[t_n,t(x))\times\mathcal{B}_\delta(x)}\lambda(x')\mathrm{d}x'\right). \quad (24)$$

引理 *8* 的证明。 要使 x 成为下一个 $\mathcal{B}_\delta(x)$-事件，必须满足以下三个条件：

1）在 x处发生一个事件。2）在 $[t_n,t(x))\times\mathcal{B}_\delta(x)$ 内没有事件发生。3） $s(x)\in\mathcal{B}_\delta(x)$

依次将这三个事件记为 A、 B、 C，则有

$$p(x;\delta)=\lim_{\mathrm{d}x\to0}\frac{\mathbb{P}(A\cap B\cap C)}{\mathrm{d}x}=$$
$$\lim_{\mathrm{d}x\to0}\frac{\mathbb{P}(A|B)}{\mathrm{d}x}\cdot\mathbb{P}(B)\cdot\mathbb{1}\{s(x)\in\mathcal{B}_\delta(x)\},$$

其中第二个等式依据 C的定义成立。注意，由于 $B\subseteq\mathcal{H}_{t(x)}$，第一项为：

$$\lim_{\mathrm{d}x\to0}\frac{\mathbb{P}(A|B)}{\mathrm{d}x}=\lim_{\mathrm{d}x\to0}\frac{\mathbb{E}[\mathbb{N}(\mathrm{d}x)|\mathcal{H}_{t(x)}]}{\mathrm{d}x}=\lambda(x).$$

为求解第二项 $\mathbb{P}(B)$，我们将生存函数 $S(t)$定义为

$$S(t)\coloneqq\mathbb{P}(\text{No event in }[0,t)\times\mathcal{B}_\delta(x)),$$

$$\lim_{\mathrm{d}t\to0}\frac{\mathbb{P}(\text{No event in }[t,t+\mathrm{d}t)\times\mathcal{B}_\delta(x))}{\mathrm{d}t}$$
$$=1-\Lambda(t)\lim_{\mathrm{d}t\to0}\mathrm{d}t+\lim_{\mathrm{d}t\to0}o(\mathrm{d}t)),$$

其中 $\Lambda(t)$是定义为 $\Lambda(t)=\int_{\mathcal{B}_\delta(x)}\lambda(x)\mathrm{d}s$. 的累积风险。于是，根据定义，有

$$\lim_{\mathrm{d}t\to0}S(t+\mathrm{d}t)$$
$$=S(t)\cdot\lim_{\mathrm{d}t\to0}\frac{\mathbb{P}(\text{No event in }[t,t+\mathrm{d}t)\times\mathcal{B}_\delta(x))}{\mathrm{d}t}$$
$$=S(t)-\lim_{\mathrm{d}t\to0}\Lambda(t)\cdot S(t)\cdot\mathrm{d}t+\lim_{\mathrm{d}t\to0}o(\mathrm{d}t).$$

因此，存在如下常微分方程（ODE）：

$$\frac{\mathrm{d}}{\mathrm{d}t}S(t)=-\Lambda(t)\cdot S(t),\quad S(0)=1.$$

求解该方程，我们得到：

$$S(t)=\exp\left(-\int_0^t\Lambda(t)\mathrm{d}t\right)=\exp\left(-\int_{[0,t)\times\mathcal{B}_\delta(x)}\lambda(t)\mathrm{d}t\right).$$

因此，

$$\mathbb{P}(B)=S(t(x))-S(t_n)=\exp\left(-\int_{[t_n,t(x))\times\mathcal{B}_\delta(x)}\lambda(t)\mathrm{d}t\right),$$

其中， t_n 表示在 x之前观测到的最后一个 $\mathcal{B}$事件的时间。结合上述推导，我们得出：

$$p(x;\delta)=\lambda(x)\exp\left(-\int_{[t_n,t(x))\times\mathcal{B}_\delta(x)}\lambda(x')\mathrm{d}x'\right).$$

其中我们利用了 $s(x)\in\mathcal{B}_\delta(x)$恒成立这一事实。至此推导完成。 □

杰卡德指数的证明。 将集合膨胀与腐蚀算子定义为

$$\Omega\oplus\mathcal{B}\coloneqq\cup_{s\in\Omega}\{s':s'\in\mathcal{B}(s)\} \quad (25)$$
$$\Omega\ominus\mathcal{B}\coloneqq\cup_{s\in\mathcal{R}}\{s:\mathcal{B}(s)\subseteq\mathcal{R}\} \quad (26)$$

并定义以下集合族

$$\mathcal{I}=\{\Omega':(\Omega\ominus\mathcal{B})\subseteq\Omega'\subseteq(\Omega\oplus\mathcal{B})\}.$$

将确定的变化后和确定的变化前区域定义为

$$\Omega_{\text{in}}\coloneqq\Omega\ominus B_\delta=\{s\in S:B_\delta(s)\subseteq\Omega\},$$
$$\Omega_{\text{out}}\coloneqq(\Omega\oplus B_\delta)^c=\{s\in S:B_\delta(s)\subseteq\Omega^c\}.$$

根据构造，空间坐标位于 $\Omega_{\text{in}}(\Omega_{\text{out}})$ 中的事件，在其局部邻域内完全由变化后（变化前）状态生成。在得分统计量的适定性（引理3）成立的条件下，

$$\mathbb{E}_\tau[\Delta(x)\mid s(x)\in\Omega_{\text{in}}]>0,\quad\mathbb{E}_\tau[\Delta(x)\mid s(x)\in\Omega_{\text{out}}]<0.$$

根据变化后点过程的遍历性和大数定律（假设1），对于任意可测集合 $A\subseteq S$，当 $t\to\infty$时，

$$\frac{1}{t}\int_{[0,t)\times A}\Delta(x)\mathrm{d}\mathbb{N}(x)\xrightarrow{\text{a.s.}}\bar{\lambda}_1|A|\mathbb{E}_\tau[\Delta(x)|s(x)\in A], \quad (27)$$

ˆ反设最大化集合 Ω 包含一个子集 $A\subseteq\Omega_{\text{out}}$，且满足 $|A|>0$。那么，根据（27）， $[0,t)\times A$ 的贡献渐近为负；移除 A 会严格增加累积统计量ˆ，对所有充分大的 t 都成立，这与 Ω 的最优性矛盾。类似地，如果 ˆ Ω 不包含一个子集 $A\subseteq\Omega_{\text{in}}$，且满足 $|A|>0$，那么添加 A 会产生正漂移并严格增加统计量，这同样与最优性矛盾。因此，随着 $\gamma\to\infty$，其概率趋于一，

$$\Omega\ominus B_\delta\subseteq\hat{\Omega}\subseteq\Omega\oplus B_\delta,$$

ˆ 等价于 $\Omega\in\mathcal{I}$。因此，我们得出结论：

$$\lim_{\gamma\to+\infty}J(\hat{\Omega},\Omega)\geq\lim_{\gamma\to+\infty}\inf_{\Omega'\subseteq\mathcal{I}}J(\Omega',\Omega)$$
$$\geq\min\left\{\frac{|\Omega|}{|\Omega\oplus\mathcal{B}_\delta|},\frac{|\Omega\ominus\mathcal{B}_\delta|}{|\Omega|}\right\}.$$

□

G. 附加实验细节

1) 阈值仿真： 阈值 γ 通常根据目标进行选择，以权衡虚警概率与成功检测概率之间的取舍

算法 2在线算法

要求： 参考数据集 $\mathcal{D}_0$；测试数据集 $\mathcal{D}$；当前时间 t；优化轮数 K；阈值 γ；学习率 η。

1：训练 f_0。
2: $f_1 \leftarrow f_0$ 初始化变化后模型。
3: $\hat{\nu}^{(0)} \leftarrow 0$。
4: **对于** $x_j \in \mathcal{D}$，**执行**
5: 按照式 (8) 的定义，计算 $\psi_0(x_j)$ 和 $\psi_1(x_j)$。
6: $\Delta(x_j) \leftarrow \psi_0(x_j) - \psi_1(x_j)$.
7: $\hat{\nu}_j^{(0)} \leftarrow 0$。
8: **对** $k \in \{1, \ldots, K\}$ **执行**
9: $\hat{\Omega}_j^{(k)} \leftarrow$ 按式 (15) 计算变化区域。
10: $\hat{\tau}_j^{(k)} \leftarrow$ 按式 (16) 计算变点。
11: **结束循环**
12: 按照（17）通过梯度下降更新 f_1 。
13: **结束循环**
14：设置 $\hat{\tau} \leftarrow \hat{\tau}^{(k)}$ 和 $\hat{\Omega} \leftarrow \hat{\Omega}^{(k)}$。
15：计算时空 CUSUM 统计量：

$$W_t \leftarrow \int_{[\hat{\tau},t)\times\hat{\Omega}} \Delta(x)\mathrm{d}\mathbb{N}(x).$$

16: **如果** $W_t \geq \gamma$ **则** 17: $\nu \leftarrow t$ 设置停止时间。
18: **结束如果**^ 19: **返回** 估计 ($\hat{\tau}$，Ω)；停止时间 ν。

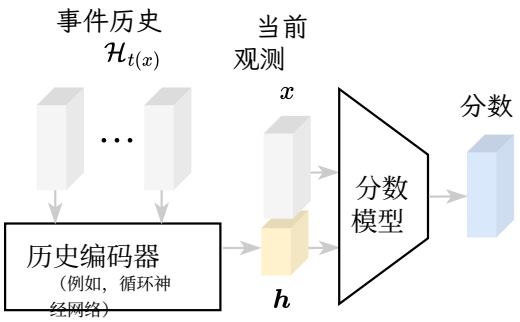

图 8。带条件历史信息的得分模型架构。历史编码器将给定的事件历史映射为潜向量 $\boldsymbol{h}$，该潜向量与当前观测事件 x拼接后，输入前馈神经网络，以生成最终的评分估计。

ARL 值。为减少确定大目标 ARL 值对应阈值时的计算量，我们采用一种高效的近似算法。当 ARL 较大时，该算法利用变化前状态下停止时间 ν 近似服从指数分布这一事实。此类近似方法广泛应用于在线变点检测。

该方法的总体思路是：我们不直接模拟 $T := \inf\{t : W_t \geq \tau\}$ 的分布均值，而是基于 N_1 次迭代，对 T 的累积分布函数进行估计，并据此估计均值。具体来说，在每次迭代中，我们模拟包含 N_2 个时间步的变化前轨迹，并计算这些N_2 个时间步上的检测统计量的最大值。这些通过 N_1 次迭代得到的最大值记为 $W_{1,\ \max}$，$W_{2,\ \max}$，…，$W_{N_1,\ \text{最大值}}$。对于所需的 ARL 值 γ，我们将停止时间T 近似为均值为 γ 的指数分布。因此，我们有 $P(W_{\max} < \gamma) = P(\nu > N_2) \approx e^{-N_2/\gamma}$。由此，相应的阈值 τ 可近似为集合 $\{W_{1,\max}, W_{2,\max}, \ldots, W_{N_1,\text{最大值}}\}$ 的 $e^{-N_2/\gamma}$分位数。

请注意，为适应点过程设定，我们在此做了一项额外修改：在确定分位数的显著性水平时，用观测事件的数量替换先前时间步的数量，如 [16]所建议。

a) 得分模型架构：图8展示了用于参数化局部得分模型f_i（·；δ）的架构。每个观测事件 $x = (t,\ s)$ 首先由其局部时空上下文表示，该上下文由事件间隔时间 $t - t_n$、空间坐标 s以及落入空间邻域 $B_\delta(x)$ 内的历史事件集合组成。这些局部先前事件由循环神经网络（RNN）依次编码，该网络将截断的事件历史汇总为固定维度的潜在表示，以捕捉局部时间依赖和自激效应。所得隐藏状态随后与当前事件 x 的特征拼接，并送入前馈神经网络，输出得分函数$\nabla_x \log\ p_i(x;\delta)$ 的估计。该设计遵循 [61]。在我们的实验中，我们将其参数化为：（i）一个具有32个隐藏单元的单层LSTM网络，用于编码事件历史；随后是（ii）一个具有512个隐藏单元的单层前馈网络，以历史嵌入与当前观测事件的拼接作为输入，并输出分数。