非归一化统计模型的最快速变化检测

Suya Wu、Enmao Diao、Taposh Banerjee、Jie Ding 和 Vahid Tarokh

arXiv:2302.00250v1 [stat.ML] 1 Feb 2023

摘要

经典的最快变化检测算法需要对变化前和变化后分布进行建模。对于各种机器学习模型而言，由于计算显式分布十分复杂，这种方法可能并不可行。此外，这些方法可能缺乏对模型失配和噪声的鲁棒性。本文提出了一种经典累积和（CUSUM）算法的新变体，用于最快变化检测。该变体基于 Fisher 散度和 Hyv¨arinen 评分函数，称为基于评分的 CUSUM（SCUSUM）算法。SCUSUM算法使变化检测能够应用于未归一化统计模型，即概率密度函数包含未知归一化常数的模型。本文通过推导平均检测延迟和平均虚警时间的表达式，研究了所提算法的渐近最优性。数值结果展示了所提算法的性能。

索引词

最快变化检测，CUSUM，Fisher 散度，得分匹配，未归一化模型

I. 引言

DETECTING 在线数据流底层统计特征的突变，是许多应用中常见的重要问题。例如，这一问题应用于传感器网络、信息物理系统、生物学和神经科学 [1]。在最快变化检测这一统计问题中，观测值被建模为随机过程的一次实现。该问题被表述为检测一系列随机变量的分布是否发生变化。变化点被定义为这种分布变化发生的时刻。最快变化检测算法能够以尽可能小的延迟尽快检测出变化点，同时满足对虚警率的约束 [1]。典型的最快变化检测算法是单阈值检验：随时间计算一系列统计量，并在该序列首次超过预设阈值时发出警报。阈值用于控制虚警率。

在最快变化检测领域，研究最为深入的设定是独立同分布（i.i.d.）设定。在这一设定中，假设变化发生前的随机变量独立同分布，且服从某个特定的概率密度函数（若无歧义，简称为密度）；变化发生后，这些随机变量也独立同分布，但服从另一种密度。在独立同分布设定下，主要的最优性结果见于 [2], [3], [4], [5]。在 [2], 中，研究表明，如果将变化点建模为几何分布随机变量，那么最优算法是在后验 概率首次超过固定阈值时停止。该算法也称为Shiryaev算法，并且已证明，在虚警概率受限的条件下，它能使平均检测延迟最小。在 [3],中，研究者提出了一种新的极小极大问题表述，并证明，随着平均虚警时间趋于无穷大， [6], 中提出的累积和（CUSUM）算法渐近最优。在 [4], 中，研究者考虑了另一种极小极大问题表述，并

本文的一个版本已被第26届国际人工智能与统计会议（AISTATS 2023）接收。Suya Wu、Enmao Diao 和 Vahid Tarokh 任职于美国北卡罗来纳州达勒姆 27708 杜克大学电气与计算机工程系。（电子邮件：suya.wu@duke.edu；enmao.diao@duke.edu；vahid.tarokh@duke.edu）Taposh Banerjee 任职于美国宾夕法尼亚州匹兹堡 15213 匹兹堡大学工业工程系。（电子邮件：taposh.banerjee@pitt.edu）Jie Ding 任职于美国明尼苏达州明尼阿波利斯 55455 明尼苏达大学双城分校统计学院。（电子邮件：dingj@umn.edu）Suya Wu 和 Vahid Tarokh 的部分研究经费由空军研究实验室通过编号为 FA-8750-20-2-0504 的拨款提供。Jie Ding 的部分研究经费由海军研究办公室通过编号为 N00014-21-1-2590 的拨款提供。Taposh Banerjee 的部分研究经费由美国陆军研究实验室通过编号为 W911NF2120295 的拨款提供。

证明，随着平均虚警时间趋于无穷大， [7], 中提出的Shiryaev-Roberts算法渐近最优。在 [5], 中，研究表明，对于 [3]中的问题表述，CUSUM算法恰好达到最优。在 [8], 中，研究表明，针对 [4]中研究的极小极大变体，CUSUM算法也渐近最优。经典的独立同分布设定已在 [8]和[9]中扩展至非独立同分布设置。有关这一经典设定下最新理论成果的更详细讨论，请参阅 [1, 10] 及其中的参考文献。

最快变化检测文献中所有最优算法的一个共同特点是，利用变化前和变化后的密度来计算观测值的似然比。随后使用这些似然比计算最优变化检测统计量 [2], [3], [4], [5] [8], [9]。在一些机器学习应用中，计算似然比可能会面临很大的计算挑战。

在一些机器学习应用中，数据模型可能具有高维特性，而且在某些情况下可能无法用显式分布表示。例如，基于能量的模型 [11] 根据观测变量和潜变量所关联的能量（未归一化的概率）来捕捉二者之间的依赖关系；基于得分的深度生成模型 [12] 则通过学习得分函数（对数密度函数的梯度）来生成高质量图像。将这些模型归一化为概率密度函数可能会带来很大的计算负担，因此，基于似然的变化检测算法在实现时计算开销很大。在第 II-C 小节中，我们通过两个例子说明这一难点。当无法获得变化前和变化后分布的完整信息时，必须利用现有数据对数据生成分布进行建模。在这种情况下，基于似然的检测算法无法达到预期效果。例如，Chen 和 Zhang [13] 通过数值结果指出，随着数据维度增加，基于广义似然比的算法会出现性能问题。对于图像数据集，Nalisnick 等人在 [14] 中证明，基于流的深度生成模型所学习的似然无法区分不同数据集之间的分布漂移。

受似然方法这些局限性的启发，我们提出了一种新颖的基于得分的最快变化检测算法，可应用于未归一化模型，即密度中包含未知归一化常数的统计模型。具体而言，我们使用 Hyv¨arinen 得分 [15] 提出 CUSUM 算法的一个基于得分的变体。下文将此变体称为 SCUSUM 算法。Hyv¨arinen 得分由 Hyv¨arinen 和 Dayan [15] 提出，旨在为未归一化模型建立一种经验估计方法。这种估计方法也称为得分匹配。最近，Wu 等人 [16]提出了一种基于得分的检验统计量，用作未归一化模型的对数似然比统计量的替代。其实验结果表明，在检验未归一化分布漂移时，该方法显著提升了性能并降低了计算复杂度。本文考虑的是 [16] 中所研究问题的序贯版本。

CUSUM算法的延迟与虚警分析采用鞅方法和更新理论方法 [8], [17]。鞅方法尤其利用了这样一个事实：CUSUM算法所用的似然比在变化前模型下构成鞅。SCUSUM算法是一种基于评分的算法，其累积评分不具备这样的鞅特征。为此，本文采用新颖的分析技术来分析SCUSUM算法。下面总结本文的主要贡献。

1）我们提出了 SCUSUM 算法，这是一种新的最快变化检测算法，适用于变化前后分布的未归一化模型。在该算法中，我们用 Hyv¨arinen 评分的一个倍数替换 CUSUM 中的负对数似然项。2）我们在独立同分布设定下，对 SCUSUM 算法进行延迟与虚警分析。我们考虑 Lorden 和 Pollak [3], [4] 的极小极大公式。我们证明，在无变化假设下，SCUSUM 的平均运行长度（ARL，即声明发生虚假变化所需的平均运行时间）随停止阈值呈指数增长（定理 3）。此外，如果发生变化，我们证明最坏情况检测延迟是停止阈值的线性函数（定理 4）。3）我们在合成数据上进行了大量数值实验，以展示 SCUSUM 的性能，并将其与基于似然比的 CUSUM [6], 、扫描B统计量 [18], 和 CALM-MMD [19]进行比较。在 ARL 约束相同的情况下，我们的方法在经验检测延迟方面与 CUSUM 表现相当。对于非正态分布数据，SCUSUM 的表现优于扫描B统计量和 CALM-MMD。我们的实验还进一步展示了 SCUSUM 在未归一化模型上相较于 CUSUM 的计算优势。

本文结构安排如下。在第 II 节中，我们对最快变化检测问题进行建模，并回顾 CUSUM 算法。我们还详细讨论基于似然比的程序的局限性。在第 III 节中，我们讨论适当评分规则的概念，并提出 SCUSUM 算法。在第 IV 节中，我们对 SCUSUM 算法进行延迟与虚警分析。在第五节中，我们展示数值实验，并将所提出的方法与基线方法进行比较。最后，我们在第 VI 节总结本文。

II. Classical Quickest Change Detection

A. 问题表述

令 $\{X_n\}_{n\geq 1}$ 表示定义在概率空间 $(\Omega\ \mathcal{F}, P_\nu)$ 上的一列独立随机变量。令$\mathcal{F}_n$ 为由随机变量 $X_1,\ X_2, \ldots,$ X_n 生成的 σ–代数，$\mathcal{F} = \sigma(\cup_{n\geq 1}\mathcal{F}_n)$为由子-$\sigma$代数的并集生成的 σ–代数。在 P_ν下，$X_1,\ X_2, \ldots,\ X_{\nu-1}$ 独立同分布，服从密度 p_∞；而$X_\nu,\ X_{\nu+1}, \ldots$ 独立同分布，服从密度 p_1。我们将 ν视为变化点，p_∞ 视为变化前密度，p_1 视为变化后密度。我们用 $\mathbb{E}_\nu$ 和 Var_ν 分别表示与测度 P_ν相对应的期望和方差。我们用 P_∞ 表示无变化时所对应的测度，并用 $\mathbb{E}_\infty$ 表示相应的期望。变化检测算法是相对于数据流$\{X_n\}_{n\geq 1}$的停止时刻 T：

$$\{T \leq n\} \in \mathcal{F}_n.$$

如果 $T \geq \nu$，我们就发生了延迟检测；否则，就发生了虚警 。直观来说，检测延迟与虚警之间存在权衡。我们考虑两种极小极大问题表述，以寻找最佳停止规则。

在 [3], 以下 极小极大指标——最坏情形平均检测延迟（WADD）——定义为 d:

$$\mathcal{L}_{\mathtt{WADD}}(T) \triangleq \sup_{\nu\geq 1} \operatorname{ess\,sup} \mathbb{E}_\nu[(T-\nu+1)^+|\mathcal{F}_\nu], \tag{1}$$

其中 $(y)^+ \triangleq \max(y,0)$ 对任意 $y \in \mathbb{R}$ 均成立。这引出了极小极大优化问题

$$\min_T\ \mathcal{L}_{\mathtt{WADD}}(T) \text{ subject to } \mathbb{E}_\infty[T] \geq \gamma. \tag{2}$$

我们还关注 Pollak [4], 提出的另一种极小极大指标，即最坏条件平均检测延迟（CADD）：

$$\mathcal{L}_{\mathtt{CADD}}(T) \triangleq \sup_{\nu\geq 1} \mathbb{E}_\nu[T-\nu|T\geq\nu]. \tag{3}$$

优化问题变为

$$\min_T\ \mathcal{L}_{\mathtt{CADD}}(T) \text{ subject to } \mathbb{E}_\infty[T] \geq \gamma. \tag{4}$$

B. 基于似然比的*CUSUM*算法

给定数据流 $\{X_n \qquad\}_{n\geq 1}$，基于似然比的CUSUM算法的停止规则定义为 by

$$T_{\mathtt{CUSUM}} \triangleq \inf\left\{n \geq 1 : \max_{1\leq k\leq n} \sum_{i=k}^{n} \log\frac{p_1(X_i)}{p_\infty(X_i)} \geq \tau\right\},$$

其中，空集的下确界定义为 $+\infty$，$\tau > 0$ 称为停止阈值。该阈值的取值显然与检测延迟和虚警之间的权衡有关。已知 [8]，$T_{\mathtt{CUSUM}}$ 可以写为

$$T_{\mathtt{CUSUM}} = \inf\{n \geq 1 : \Lambda(n) \geq \tau\},$$

其中 $\Lambda(n)$ 由以下递推定义

$$\begin{aligned}
&\Lambda(0) = 0,\\
&\Lambda(n) \triangleq \left(\Lambda(n-1) + \log\frac{p_1(X_n)}{p_\infty(X_n)}\right)^+, \forall n \geq 1,
\end{aligned} \tag{5}$$

这导出了一个计算上便捷的停止方案。

在 [5], 中证明，对于Lorden问题 (1)，CUSUM算法在每个固定约束 γ下都是精确最优的。正如 [8],中所指出的，该算法对于Pollak问题 (3) 也具有渐近最优性。 [3]和[8], 还刻画了CUSUM算法的渐近性能。具体而言，文中证明

$$\mathcal{L}_{\mathtt{WADD}}(T_{\mathtt{CUSUM}}) \sim \mathcal{L}_{\mathtt{CADD}}(T_{\mathtt{CUSUM}}) \sim \frac{\log \gamma}{\mathbb{D}_{\mathtt{KL}}(P_1 \| P_\infty)}, \text{ as } \gamma \to \infty. \tag{6}$$

这里， $\mathbb{D}_{\mathtt{KL}}(P_1 \| P_\infty)$ 表示变化后分布 P_1 （其密度为 p_1）与变化前分布 P_∞ （其密度为 p_∞）之间的Kullback-Leibler散度：

$$\mathbb{D}_{\mathtt{KL}}(P_1 \| P_\infty) = \int_x p_1(x) \log \frac{p_1(x)}{p_\infty(x)} dx,$$

并且，记号 $g(c) \sim h(c)$ as $c \to c_0$ 表示，对于任意两个函数 $c \mapsto g(c)$ 和$c \mapsto h(c)$，有 $\frac{g(c)}{h(c)} \to 1$ as $c \to c_0$。

C. 基于似然比的*CUSUM*算法存在的问题

我们分别考虑变化前和变化后的密度， p_1 和 p_∞。我们假设这些密度可能仅在一个归一化常数的意义下已知，也就是说，我们处理的是未归一化模型。换言之，我们得到的不是 $p_1(x)$ 和$p_\infty(x)$，而是 $\tilde{p}_1(x)$和 $\tilde{p}_\infty(x)$，其中

$$p_i(x) = \frac{\tilde{p}_i(x)}{\int_{x \in \mathcal{X}} \tilde{p}_i(x) dx}, \ i = 1, \infty.$$

正如引言中所述，这类模型偶尔会出现在若干机器学习应用中。在许多情况下，当积分无法用闭式解析求解时，分母（也称为归一化常数或配分函数）的计算可能难以处理。对于低维情形，可以使用数值积分来近似该函数。然而，近似积分所需的点数可能会随着数据空间维度的增加而呈指数级增长。对于高维数据，这种近似的计算开销很大。因此，对于未归一化模型，实现基于似然比的 CUSUM 算法在计算上十分繁琐。接下来，我们给出两个例子来说明这一问题。

示例 1 （指数族）。 我们考虑属于成对交互图模型 [20] 的指数族子族。令 $X \in \mathbb{R}^d$ 为随机变量， p_τ 表示其密度，其形式如下

$$p_\tau(X) = \frac{1}{Z_\tau} \exp\left\{ -\tau \left(\sum_{i=1}^{d} x_i^4 + \sum_{1 \le i \le d, i \le j \le d} x_i^2 x_j^2 \right) \right\},$$

其中， $\tau \in \mathcal{T} \subset \mathbb{R}^+$ 是模型参数， Z_τ 是 $p_\tau(X)$ 的归一化常数。这里，

$$Z_\tau = \int_{x_1} \cdots \int_{x_d} \exp\left\{ -\tau \left(\sum_{i=1}^{d} x_i^4 + \sum_{1 \le i \le d, i \le j \le d} x_i^2 x_j^2 \right) \right\} dx_1 \cdots dx_d.$$

如上所示，该积分无法通过闭式形式计算，因此密度 p_τ 也无法通过闭式形式计算。此外，当 d 较大时，数值近似非常耗时。特别是，在第五节中，我们将展示，当$d = 4$时，基于似然比的 CUSUM 无法在合理的计算时间内实现。

例 2 （高斯-伯努利受限玻尔兹曼机）。 受限玻尔兹曼机（RBM） [11]是一种定义在由隐变量和可见变量构成的二部图上的生成式图模型。具体而言，我们考虑高斯-伯努利 RBM（GB-RBM），其隐变量取二值 $H = (h_1, \ldots, h_{d_h})^T \in \{0,1\}^{d_h}$，可见变量取实值 $X = (x_1, \ldots, x_{d_x})^T \in R^{d_x}$，联合概率密度为

$$p(X, H) = \frac{1}{Z} \exp\left\{ -\left(\frac{1}{2} \sum_{i=1}^{d_x} \sum_{j=1}^{d_h} \frac{x_i}{\sigma_i} W_{ij} h_j + \sum_{i=1}^{d_x} b_i x_i + \sum_{j=1}^{d_h} c_j h_j - \frac{1}{2} \sum_{i=1}^{d_x} \frac{x_i^2}{\sigma_i^2} \right) \right\},$$

其中，模型参数为 $\theta = (\mathbf{W}, \mathbf{b}, \mathbf{c})$，而 Z 是 $p(X, H)$的归一化常数。我们对所有$i = 1, \ldots, d_x$设定 $\sigma_i = 1$。

令 p_θ 表示可见变量 X的密度，可写为

$$p_\theta(X) = \sum_{h \in \{0,1\}^{d_h}} p_\theta(X, H) = \frac{1}{Z_\theta} \exp\{-F_\theta(X)\},$$

wh其中 Z_θ 是 $p_\theta(X)$的归一化常数，而 $F_\theta(X)$ 是由以下给出的自由能 y

$$F_\theta(X) = \frac{1}{2} \sum_{i=1}^{d_x} (x_i - b_i)^2 - \sum_{j=1}^{d_h} \text{Softplus}\left(\sum_{i=1}^{d_x} W_{ij} x_i + b_j \right).$$

Softplus函数定义为 $\text{Softplus}(y) \triangleq \log(1 + \exp(y))$，默认尺度参数为 $\beta = 1$。 $p_\theta(X)$也存在相同的计算难题，因此实际上可能无法准确计算 GB-RBM 数据的似然。

III. 基于评分的最快变化检测

本节提出一种基于评分的 CUSUM（SCUSUM）算法，以解决未归一化模型中基于似然比的 CUSUM 所面临的问题。按照 CUSUM 的方案，所提算法可递归应用，其在线实现所需的计算和内存资源并不高。为此，我们首先回顾适当评分规则的框架，该框架直观地比较了 CUSUM 与 SCUSUM。

A. 适当评分规则

设 X 是一个取值于 $\mathcal{X} \subseteq \mathbb{R}^d$ 的随机变量， $\mathcal{P}$ 是定义在 $\mathcal{X}$ 上的一个分布族。令 P 和 $Q \in \mathcal{P}$ 分别表示真实的数据生成分布和假定分布，并令 p 和 q 分别表示它们对应的密度。Gneiting 和 Raftery [21] 将适当评分规则作为统一框架加以研究，用于衡量假定模型对观测数据的拟合质量。

定义 1（适当评分规则）。评分规则是一个函数 $(X, Q) \mapsto \mathcal{S}(X, Q)$，用于衡量 Q 对由 X 表示的数据进行建模的质量。若对所有 $P \in \mathcal{P}$，期望分数 $\mathbb{E}_{X \sim P} \mathcal{S}[(X, Q)]$ 在 $Q = P$ 处取最小值（最小值是在所有 $Q \in \mathcal{P}$ 中取得的），则称该评分规则是适当的。此外，若对于任意 $Q \in \mathcal{P}$ 和 $Q \neq P$， $\mathbb{E}_{X \sim P}[\mathcal{S}(X, Q)] > \mathbb{E}_{X \sim P}[\mathcal{S}(X, P)]$，则称 $\mathcal{S}$ 关于 $\mathcal{P}$ 是严格适当的。

对数评分规则 [22] 是严格适当评分规则中一个广为人知且应用广泛的例子。

定义 2(对数评　分)。 对数评分规则（也称为对数评分）定义为

$$(X, Q) \mapsto \mathcal{S}_{\text{L}}(X, Q) \triangleq -\log q(X).$$

最小化对数评分与最大似然估计（MLE）和 Kullback-Leibler（KL）散度相关

$$\mathbb{D}_{\text{KL}}(P \| Q) \triangleq \mathbb{E}_{X \sim P} \left[\log p(X) - \log q(X) \right].$$

由于对于任意 $Q \neq P$， $\mathbb{D}_{\text{KL}}(P \| Q) > 0$，对数评分是严格适当的。基于对数似然比的累积和算法的检测评分（定义见公式（5））可改写为

$$\log \frac{p_1(X_n)}{p_\infty(X_n)} = \mathcal{S}_{\text{L}}(X_n, P_\infty) - \mathcal{S}_{\text{L}}(X_n, P_1).$$

B. Fisher 散度与 *Hyv¨arinen* 评分

Hyv¨arinen 和 Dayan [15] 提出了一种通过最小化从 P 到 Q的 Fisher 散度来估计未归一化统计模型的方法，其定义为

$$\mathbb{D}_{\mathrm{F}}(P\|Q) \triangleq \mathbb{E}_{X\sim P}\left[\|\nabla_{\mathbf{x}}\log p(X)-\nabla_{\mathbf{x}}\log q(X)\|_2^2\right],$$

其中，$\|\cdot\|_2$ 表示欧几里得范数。显然，如果相对于 X将 p 和 q 按任意正数缩放，$\nabla_{\mathbf{x}}\log p(X)$ 和 $\nabla_{\mathbf{x}}\log q(X)$ 均保持不变。因此，对于密度函数的任意常数缩放，Fisher 散度仍然是尺度变异的 。在对 p 和 q施加一些较弱的正则性条件时，Hyv¨arinen 和 Dayan [15] 证明了

$$\mathbb{D}_{\mathrm{F}}(P\|Q) = \mathbb{E}_{X\sim P}\left[\frac{1}{2}\|\nabla_{\mathbf{x}}\log p(X)\|_2^2 + \mathcal{S}_{\mathrm{H}}(X,Q)\right],$$

其中，$\mathcal{S}_{\mathrm{H}}(X,Q)$ 是一个尺度不变的 适当评分函数，在适当评分规则 [23]的框架下称为 Hyv¨arinen 评分。由于$\frac{1}{2}\|\nabla_{\mathbf{x}}\log\, p(X)\|_2^2$ 相对于 Q为常数，因此最小化 Fisher 散度等价于最小化 $\mathcal{S}_{\mathrm{H}}(X,Q)$。

定义 3(Hyv¨arinen 评分)。 Hyv¨arinen 评分是一个映射$(X,Q)\mapsto\mathcal{S}_{\mathrm{H}}(X,Q)$，由 b 给出y

$$\mathcal{S}_{\mathrm{H}}(X,Q) \triangleq \frac{1}{2}\|\nabla_X\log q(X)\|_2^2 + \Delta_X\log q(X) \tag{7}$$

只要它可以被良好定义。这里，∇_X和 $\Delta_X=\sum_{i=1}^d\frac{\partial^2}{\partial x_i^2}$分别表示作用于 $X=(x_1,\ \cdots,\ x_d)^\top$的梯度和拉普拉斯算子。

$\mathcal{S}_{\mathrm{H}}$ 具有尺度不变性，这一性质源自 Fisher 散度的尺度不变性。该性质避免了计算未归一化模型的归一化常数。具体来说，当我们对 Q 的了解仅限于$\tilde{q}(x)$ 时，

$$q(x)=\frac{\tilde{q}(x)}{\int_{x\in\mathcal{X}}\tilde{q}(x)dx},$$

很容易看出，将密度 q 替换为相应的未归一化项$\tilde{q}$后，$\mathcal{S}_{\mathrm{H}}(X,\ Q)$ 保持不变。此外，容易验证，对于 $Q\neq P$，$\mathbb{D}_{\mathrm{F}}(P\|Q)>0$，因此 Hyv¨arinen 评分是严格适当的。

C. 基于评分的 *CUSUM* 算法

根据第 III-A 和 III-B 小节中的讨论，Hyv¨arinen 得分函数可视为对数评分函数的替代形式。受此类比启发，我们考虑在基于对数似然比的累积和算法中用 Hyv¨arinen 评分替代对数评分。接下来，我们定义 SCUSUM 的检测评分，并给出停止方案。

令 X 表示定义在概率空间 $(\Omega\ \mathcal{F},\ P)$ 上的一个一般随机变量。 P 可以是变化前分布或变化后分布。我们将瞬时 SCUSUM 评分函数 $X\mapsto z_\lambda(X)$ 定义为

$$z_\lambda(X) \triangleq \lambda\big(\mathcal{S}_{\mathrm{H}}(X,P_\infty)-\mathcal{S}_{\mathrm{H}}(X,P_1)\big), \tag{8}$$

其中，$\lambda>0$ 是预先选定的乘数，$\mathcal{S}_{\mathrm{H}}(X,\ P_\infty)$ 和 $\mathcal{S}_{\mathrm{H}}(X,\ P_1)$ 分别是变化前后分布的 Hyv¨arinen 得分函数。在第 IV 节中，我们将详细讨论 λ 在 SCUSUM算法中的作用。然后，我们提出的停止规则为

$$T_{\mathrm{SCUSUM}} \triangleq \inf\left\{n\geq 1:\max_{1\leq k\leq n}\sum_{i=k}^n z_\lambda(X_i)\geq\tau\right\}, \tag{9}$$

其中，$\tau>0$ 是停止阈值，通常预先选定以控制虚警。与 CUSUM 的停止方案类似，SCUSUM 的停止规则可写为

$$T_{\mathrm{SCUSUM}}=\inf\{n\geq 1: Z(n)\geq\tau\},$$

其中， $Z(n)$ 可通过以下递归方式计算

$$Z(0) = 0,$$
$$Z(n) \triangleq (Z(n-1) + z_\lambda(X_n))^+, \ \forall n \geq 1.$$

Z（n）称为 SCUSUM 在时刻 n 的检测评分。SCUSUM算法总结于算法 1 中。

算法 1： SCUSUM 检测算法

输入：变化前后分布的 Hyvarinen 分数函数 $\mathcal{S}_{\mathrm{H}}(\cdot, P_\infty)$ 和 $\mathcal{S}_{\mathrm{H}}(\cdot, P_1)$。数据： m 先前观测值 $X_{[-m+1,0]}$ 以及在线数据流 $\{X_n\}_{n\geq 1}$初始化：当前时刻 $k = 0$、 $\lambda > 0$、 $\tau > 0$ 和 $Z(0) = 0$当 $Z(k) < \tau$ 时$k = k + 1$更新 $z_\lambda(X_k) = \lambda(\mathcal{S}_{\mathrm{H}}(X_k, P_\infty) - \mathcal{S}_{\mathrm{H}}(X_k, P_1))$ 更新 $Z(k) = \max(Z(k-1) + z_\lambda(X_k), 0)$ 将当前时刻 k 记录为停止时刻 T_{SCUSUM}输出： T_{SCUSUM}



IV. SCUSUM算法的延迟与虚警分析

本节沿用第 II 节和第 III 节中定义的记号和问题设定，对 SCUSUM算法进行延迟与虚警分析。我们引入两个假设：1) $P_1 \neq P_\infty$；2) 满足 Hyv¨arinen 和 Dayan [15] 提出的相同温和正则性条件[1] ，以确保 Hyv¨arinen 得分有明确定义。

我们首先概述本节的结果。引理 1 表明，与 CUSUM算法一样，SCUSUM算法在变化前的漂移为负，在变化后为正。CUSUM算法中的 KL 散度在 SCUSUM算法中由 Fisher 散度取代。我们的核心结果见定理 3 和定理 4。定理 3 给出了未发生变化时平均运行长度的下界。正如引言中所述，此处的挑战在于，无法使用经典的鞅技术（例如 [8] 中采用的技术）推导出该界。这是因为 SCUSUM算法基于得分，而非对数似然比。后者具有经典证明所利用的鞅性质。我们在引理 1 之后、引理 2 中以及定理 3 的证明中提出了一种新颖的证明技术。定理 4 给出了变化点发生在 $v = 1$ 时期望检测延迟的上界。由此可得 SCUSUM算法的 $\mathcal{L}_{\mathrm{WADD}}$ 上界（见 (1)）。命题 5 考虑了多元正态变化前后分布这一特殊情形，并讨论了我们的算法在此情形下的渐近最优性。具体而言，我们表明，在这种情形下，KL 散度与 Fisher 散度相同，且 SCUSUM算法具有与 CUSUM算法相同的最优性。

引理 1 (正向和负向漂移)**.** 考虑公式 *(8)* 中定义的瞬时 *SCUSUM* 评分函数 $X \mapsto z_\lambda(X)$ 。于是，

$$\mathbb{E}_\infty [z_\lambda(X)] = -\lambda \mathbb{D}_{\scriptscriptstyle F}(P_\infty \| P_1) < 0, \ \textit{and}$$
$$\mathbb{E}_1 [z_\lambda(X)] = \lambda \mathbb{D}_{\scriptscriptstyle F}(P_1 \| P_\infty) > 0.$$

*Pr*哎呀。 在一些温和的正则性条件下，Hyv¨arinen 和 Dayan [15]证明了

$$\mathbb{D}_{\mathrm{F}}(P\|Q) = \mathbb{E}_{X\sim P}\left[\frac{1}{2}\|\nabla_X \log p(X)\|_2^2 + \mathcal{S}_{\mathrm{H}}(X, Q)\right].$$

[1]详情请参见 [15]。

令 $C(P)$ 表示任意 $P \in \mathcal{P}$ 对应的 $\mathbb{E}_{X\sim P}\left[\frac{1}{2}\|\nabla_X \log p(X)\|_2^2\right]$，则

$$\mathbb{E}_\infty[\mathcal{S}_{\mathrm{H}}(X,P_\infty)-\mathcal{S}_{\mathrm{H}}(X,P_1)] = \mathbb{D}_{\mathrm{F}}(P_\infty\|P_\infty)-C(P_\infty)-\mathbb{D}_{\mathrm{F}}(P_\infty\|P_1)+C(P_\infty) = -\mathbb{D}_{\mathrm{F}}(P_\infty\|P_1),$$

and

$$\mathbb{E}_1[\mathcal{S}_{\mathrm{H}}(X,P_\infty)-\mathcal{S}_{\mathrm{H}}(X,P_1)] = \mathbb{D}_{\mathrm{F}}(P_1\|P_\infty)-C(P_1)-\mathbb{D}_{\mathrm{F}}(P_1\|P_1)+C(P_1) = \mathbb{D}_{\mathrm{F}}(P_1\|P_\infty).$$

由于 $\lambda>0$ 相对于 P_1 和 P_∞为常数，证明完毕。 □

引理 1 表明，在变化发生之前，根据随机观测的测度，瞬时 SCUSUM 分数 z_λ（X）的期望均值为负。因此，在变化发生之前的每个时刻 n，累积分数都呈负漂移。因此，变化点之前，SCUSUM 检测分数 Z（n）会趋向于零。这直观地说明虚警不太可能发生。相反，变化发生后，瞬时分数的均值为正，累积分数呈正漂移。因此，SCUSUM 检测分数会不断增大并趋于无穷大，进而触发变化检测事件。

接下来，我们讨论理论分析中的乘数 λ 取值。显然，在停止阈值固定时，λ 取值越大，检测延迟越短，因为 SCUSUM 检测分数的增量较大，因而更容易达到阈值。然而，λ 取值较大也会导致 SCUSUM 在未发生变化时过早停止，从而增大虚警概率。因此，除了一些退化情形（此时 Hyv¨arinen 得分函数 P_∞（S_{H}（X，P_1）$-S_{\mathrm{H}}$（X，P_∞）≤ 0）$=1$），λ 的取值不能任意大。它需要满足以下关键条件：

$$\mathbb{E}_\infty[\exp(z_\lambda(X))] \le 1. \tag{10}$$

我们将给出一个技 术引理，以保证存在这样的 λ，使其满足不等式（10）。

引理 2（适当的 λ的存在性）。存在 $\lambda>0$ 使得不等式（*10*）成立。此外，要么 *1*）存在 $\lambda^\star\in$（$0,\infty$）使得（*10*）中的等式成立；要么 *2*）对于所有 $\lambda>0$，（*10*）中的不等式均为严格不等式。

证明。定义函数 $\lambda:\mapsto h$（λ）如下：

$$h(\lambda)\triangleq \mathbb{E}_\infty[\exp(z_\lambda(X))]-1.$$

注意到

$$h'(\lambda)\triangleq \frac{dh}{d\lambda}(\lambda)=\mathbb{E}_\infty[(S_{\mathrm{H}}(X,P_\infty)-S_{\mathrm{H}}(X,P_1))\exp(z_\lambda(X))].$$

注意，h（0）$=0$，且 h'（0）$=-\mathbb{D}_{\mathrm{F}}$（$P_\infty\|P_1$）$<0$。因此，存在 $\lambda>0$ 使得 h（λ）<0，且不等式（10）成立。

接下来，我们证明以下两种情况之一成立：1）存在 $\lambda^\star\in(0,\infty)$ 使得 $h(\lambda^\star)=0$；或者 2）对所有 $\lambda>0$，都有 $h(\lambda)<1$。

注意到

$$h''(\lambda)\triangleq \frac{d^2h}{d\lambda}(\lambda)=\mathbb{E}_\infty[(S_{\mathrm{H}}(X,P_\infty)-S_{\mathrm{H}}(X,P_1))^2\exp(z_\lambda(X))]\ge 0.$$

我们断言 $h(\lambda)$ 是严格凸的，即对所有 $\lambda\in[0,\infty)$，都有 $h''(\lambda)>0$。假设对某个 $\lambda\ge 0$，$h''(\lambda)=0$，则我们必有 $S_{\mathrm{H}}(X,P_\infty)-S_{\mathrm{H}}(X,P_1)$ 几乎必然成立。这意味着 $\mathbb{E}_\infty[(S_{\mathrm{H}}(X,P_\infty)-S_{\mathrm{H}}(X,P_1))]=0$，进而得到 $-\mathbb{D}_{\mathrm{F}}(P_\infty\|P_1)=0$，且该关系几乎处处成立，这与假设 $P_\infty\neq P_1$ 矛盾。因此，$h(\lambda)$ 是严格凸的，并且 $h'(\lambda)$ 严格递增。

由此可知，以下两种情况必有其一：1）$h(\lambda)$ 在$(0,\infty)$内至多有一个全局最小值；或者 2）它在 $[0,\infty)$上严格递减。我们注意到这两种情况，并证明第二种情况是退化的，在实际中没有意义。

- 情况 1：如果 $h(\lambda)$的全局最小值在 $a\in(0,\infty)$处取得，那么 $h'(a)=0$。由于 $h'(0)<0$ 且$h(0)=0$，全局最小值 $h(a)<0$。由于 $h'(\lambda)$严格递增，我们可以选取 $b>a$，并得出对于所有 $\lambda>b$，$h'(\lambda)>h'(b)>h'(a)=0$。由此可知，$\lim_{\lambda\to\infty}h(\lambda)=+\infty$。结合 $h(\lambda)$ 的连续性，我们得出，对于某个 $\lambda^*\in(0,\infty)$，$h(\lambda^*)=0$，且 $\lambda\in(0,\lambda^*]$ 的任意取值都满足不等式（10）。

  注意，在这种情况下，对于某个 $c>0$，必有 $P_\infty(S_{\mathrm{H}}(X,P_\infty)-S_{\mathrm{H}}(X,P_1)\ge c)>0$。否则，就有 $P_\infty(S_{\mathrm{H}}(X,P_\infty)-S_{\mathrm{H}}(X,P_1)\le 0)=1$。这意味着 $P_\infty(z_\lambda(X)\le 0)=1$，或者等价地，$\mathbb{E}_\infty[\exp(z_\lambda(X))]<1$ 对所有 $\lambda>0$ 都成立，因此可推出情形 2：$h(\lambda)<0$ 对所有 $\lambda>0$ 都成立。这里，

$\mathbb{E}_\infty[\exp(z_\lambda(X))] \neq 1$，因为 $P_\infty(S_{\mathrm{H}}(X,P_\infty) - S_{\mathrm{H}}(X,P_1) = 0) < 1$；否则 $P_\infty(S_{\mathrm{H}}(X,P_\infty) - S_{\mathrm{H}}(X,P_1) = 0) = 1$，于是 $\mathbb{E}_\infty[S_{\mathrm{H}}(X,P_\infty) - S_{\mathrm{H}}(X,P_1)] = -\mathbb{D}_{\mathrm{F}}(P_\infty \| P_1) = 0$，从而导致与 $P_1 \neq P_\infty$ 相同的矛盾。

- 情形 2：若 $h(\lambda)$ 在 $(0,\infty)$ 上严格递减，则任意 $\lambda \in (0,\infty)$ 都满足不等式（10）。如前所述，在这种情况下，必有 $P_\infty(S\mathrm{H}(X,P_\infty) - S\mathrm{H}(X,P_1) \leq 0) = 1$。等价地，在变化前分布下，SCUSUM 检测分数的所有增量均为非正，且对所有 n，都有 $P_\infty(Z(n) = 0) = 1$。因此，$\mathbb{E}_\infty[T\mathrm{SCUSUM}] = +\infty$。当发生变化时（在测度 P_1 下），我们还观察到，只要将 λ 选得足够大，SCUSUM 就能近乎瞬时地检测到变化点。显然，这种情况在实践中没有意义。

□

从现在起，我们考虑一个满足不等式（10）的固定 $\lambda > 0$，以阐述我们的核心结果。在实践中，可以使用 m 过去样本$\mathbf{X}^{[-m+1,0]}$ 来确定 λ 的取值。具体而言，可以将 λ 选为函数 $\lambda \to h(\lambda)$ 的正根，该函数定义为

$$\tilde{h}(\lambda) \triangleq \frac{1}{m}\sum_{i=1}^{m}[\exp(z_\lambda(X_{i-m}))] - 1. \tag{11}$$

根据引理 2 及其相关技术讨论，当 m 足够大时，上述方程以较高概率存在一个大于零的根。如果 λ 未选择妥当，该算法仍可实现，但无法保证检测延迟的性能。我们将在注 2 中进一步讨论这种情况。

定理 3. 考虑方程（9）中定义的停止规则 T_{SCUSUM}。那么，对于任意 $\tau > 0$,

$$\mathbb{E}_\infty[T_{SCUSUM}] \geq e^{\tau}. \tag{12}$$

证明. 我们遵循 Lai [8, 定理 4] 的证明，得出定理 3 的结论。结合构造的鞅与 Doob 次鞅不等式 [24]，即可完成证明。

1) 我们首先构造一个非负鞅，使其在测度 P_∞ 下的均值为1 。定义新的瞬时得分函数 $X \mapsto \tilde{z}_\lambda(X)$ 如下:

$$\tilde{z}_\lambda(X) \triangleq z_\lambda(X) + \delta,$$

其中

$$\delta \triangleq -\log\Big(\mathbb{E}_\infty\left[\exp(z_\lambda(X))\right]\Big).$$

进一步定义序列

$$\tilde{G}_n \triangleq \exp\Big(\sum_{k=1}^{n}\tilde{z}_\lambda(X_k)\Big),\ \forall n \geq 1.$$

假设 X_1, X_2, ...按照 P_∞ 独立同分布（未发生变化）。那么,

$$\mathbb{E}_\infty\left[\tilde{G}_{n+1} \mid \mathcal{F}_n\right] = \tilde{G}_n\mathbb{E}_\infty[\exp(\tilde{z}_\lambda(X_{n+1}))] = \tilde{G}_n e^{\delta}\mathbb{E}_\infty[\exp(z_\lambda(X_{n+1}))] = \tilde{G}_n,$$

and

$$\mathbb{E}_\infty[\tilde{G}_n] = \mathbb{E}_\infty\left[\exp\left(\sum_{i=1}^{n}(z_\lambda(X_i) + \delta)\right)\right] = e^{n\delta}\prod_{i=1}^{n}\mathbb{E}_\infty[\exp(z_\lambda(X_i))] = 1.$$

~ ~因此，在测度 P_∞下， $\{G_n\}_{n\geq 1}$ 是一个非负鞅，其均值为 $\mathbb{E}_\infty[G_1] = 1$。2）接下来，我们考察新的停止规则

$$\tilde{T}_{\mathrm{SCUSUM}} = \inf\left\{n \geq 1 : \max_{1\leq k\leq n}\sum_{i=k}^{n}\tilde{z}_\lambda(X_i) \geq \tau\right\},$$

其中 $\tilde{z}_\lambda($ $X_i) = z_\lambda(X_i) + \delta$。根据不等式（10），我们观察到 $\delta \geq 0$。根据詹森不等式 y,

$$\mathbb{E}_\infty[\exp(z_\lambda(X))] \geq \exp\left(\mathbb{E}_\infty[z_\lambda(X)]\right), \tag{13}$$

等号成立当且仅当 $z_\lambda(X) = c$ 几乎必然成立，其中 c 是某个常数。假设公式（13）中的等号成立，那么

$$-\lambda\mathbb{D}_{\mathrm{F}}(P_1||P_\infty) = \mathbb{E}_\infty[z_\lambda(X)] = c = \mathbb{E}_1[z_\lambda(X)] = \lambda\mathbb{D}_{\mathrm{F}}(P_\infty||P_1).$$

由此可得 $0 \le \mathbb{D}_{\mathrm{F}}(P_\infty||P_1) = -\mathbb{D}_{\mathrm{F}}(P_1||P_\infty) \le 0$，这意味着 $P_\infty = P_1$ 几乎处处成立。这与假设 $P_\infty \neq P_1$ 矛盾。因此，公式（13）中的不等式是严格的，并且˜因此 $\delta < \lambda\mathbb{D}_{\mathrm{F}}(P_\infty||P_1)$。所以，$T_{\mathrm{SCUSUM}}$ 并非平凡的。

定义一个停止时刻序列：

$$\begin{aligned}
&\eta_0 = 0,\\
&\eta_1 = \inf\left\{t : \sum_{i=1}^{t} \tilde{z}_\lambda(X_i) < 0\right\},\\
&\eta_{k+1} = \inf\left\{t > \eta_k : \sum_{i=\eta_k+1}^{t} \tilde{z}_\lambda(X_i) < 0\right\}, \text{ for } k \ge 1.
\end{aligned}$$

˜根据前面的讨论，$\{G_n\}_{n\ge1}$ 在 P_∞ 下是一个非负鞅，其均值为 1。于是，对于任意 k，且在 $\{\eta_k < \infty\}$ 上，

$$P_\infty\left(\sum_{i=\eta_k+1}^{n} \tilde{z}_\lambda(X_i) \ge \tau \text{ for some } n > \eta_k \mid \mathcal{F}_{\eta_k}\right) \le e^{-\tau}, \tag{14}$$

根据 Doob 子鞅不等式 [24]。令

$$M \triangleq \inf\left\{k \ge 0 : \eta_k < \infty \text{ and } \sum_{i=\eta_k+1}^{n} \tilde{z}_\lambda(X_i) \ge \tau \text{ for some } n > \eta_k\right\}. \tag{15}$$

结合不等式（14）和定义（15），

$$P_\infty(M \ge k+1 \mid \mathcal{F}_{\eta_k}) = 1 - P_\infty\left(\sum_{i=\eta_k+1}^{n} \tilde{z}(X_i) \ge \tau \text{ for some } n > \eta_k \mid \mathcal{F}_{\eta_k}\right) \ge 1 - e^{-\tau}, \tag{16}$$

and

$$P_\infty(M > k) = \mathbb{E}_\infty[P_\infty(M \ge k+1 \mid \mathcal{F}_{\eta_k})\,\mathbb{I}_{\{M\ge k\}}] = \mathbb{E}_\infty[P_\infty(M \ge k+1 \mid \mathcal{F}_{\eta_k})]P_\infty(M > k-1)。 \tag{17}$$

结合公式（17）和（16），

$$\mathbb{E}_\infty[M] = \sum_{k=0}^{\infty} P_\infty(M > k) \ge \sum_{k=0}^{\infty}(1 - e^{-\tau})^k = e^{\tau}.$$

注意到

$$\tilde{T}_{\mathrm{SCUSUM}} = \inf\left\{n \ge 1 : \sum_{i=\eta_k+1}^{n} \tilde{z}_\lambda(X_i) \ge \tau \text{ for some } \eta_k < n\right\} \ge M,$$

˜ ˜和 $T_{\mathrm{SCUSUM}} \le T_{\mathrm{SCUSUMo}}$ 由此得出 $\mathbb{E}_\infty[T_{\mathrm{SCUSUM}}] \ge \mathbb{E}_\infty[T_{\mathrm{SCUSUM}}] \ge \mathbb{E}_\infty[M] \ge e^{\tau}$。

□

$\mathbb{E}_\infty[T_{\mathrm{SCUSUM}}]$ 也称为平均运行长度 (ARL) [6]。定理 3 表明，随着停止阈值的增大，ARL 至少呈指数增长。以下定理给出了在 ARL 控制下 SCUSUM 在检测延迟方面的渐近性能。

定理 4。 在满足 $\mathbb{E}_\infty[T_{SCUSUM}] \ge \gamma > 0$ 的条件下，停止规则 T_{SCUSUM} 满足

$$\mathcal{L}_{WADD}(T_{SCUSUM}) \sim \mathcal{L}_{CADD}(T_{SCUSUM}) \sim \mathbb{E}_1[T_{SCUSUM}] \sim \frac{\log\gamma}{\lambda\mathbb{D}_F(P_1\|P_\infty)}, \tag{18}$$

当 $\gamma \to \infty$ 时。

我们首先引入一个技术性定义，以便将 [17, 推论 2.2] 应用于定理 5 的证明。

定义 4. 若分布 P 在 $(-\infty, \infty)$ 的博雷尔集上当且仅当集中于形如 $\pm nd$的一组点上，则称其为算术型，其中 $d > 0$ 且 $n = 1, 2, \ldots$。

注 1. 任何关于勒贝格测度绝对连续的概率测度都是非算术型的。

证明。 考虑由下式定义的随机游走

$$Z'(n) = \sum_{i=1}^{n} z_\lambda(X_i), \text{ for } n \geq 1.$$

我们考察由下式给出的另一个停止时刻

$$T'_{\texttt{SCUSUM}} \triangleq \inf\{n \geq 1 : Z'(n) \geq \tau\}.$$

接下来，对任意 τ，在 $\{T'_{\texttt{SCUSUM}} < \infty\}$ 上定义 R_τ

$$R_\tau \triangleq Z'(T'_{\texttt{SCUSUM}}) - \tau.$$

R_τ 是在停止时刻 $T'_{\texttt{SCUSUM}}$，随机游走超过停止阈值 τ 的部分。假设变化点为 $\nu = 1$，那么 $X_1, X_2, \ldots$，是服从分布 P_1的独立同分布随机变量。令 μ和 σ^2 分别表示均值$\mathbb{E}_1[z_\lambda(X)]$ 和方差 $\text{Var}_1[z_\lambda(X)]$。注意

$$\mu = \mathbb{E}_1[z_\lambda(X)] = \lambda \mathbb{D}_{\texttt{F}}(P_1 \| P_\infty) > 0,$$

and

$$\sigma^2 = \text{Var}_1[z_\lambda(X)] = \mathbb{E}_1[z_\lambda(X)^2] - (\lambda \mathbb{D}_{\texttt{F}}(P_1 \| P_\infty))^2 .$$

在 Hyv¨arinen 和 Dayan [15],给出的温和正则性条件下

$$\mathbb{E}_1[\mathcal{S}_{\texttt{H}}(X, P_\infty)]^2 < \infty, \text{ and}$$
$$\mathbb{E}_1[\mathcal{S}_{\texttt{H}}(X, P_1)]^2 < \infty.$$

这意味着 $\mathbb{E}_1[z_\lambda(X)^2] < \infty$，前提是适当选择 λ，例如使 λ 满足不等式（10），且 λ 并非任意大。因此，根据 Lorden [25,定理 1],

$$\sup_{\tau \geq 0} \mathbb{E}_1[R_\tau] \leq \frac{\mathbb{E}_1[(z_\lambda(X)^+)^2]}{\mathbb{E}_1[z_\lambda(X)]} \leq \frac{\mu^2 + \sigma^2}{\mu},$$

其中 $z_\lambda(X)^+ = \max(z_\lambda(X), 0)$。此外，为使 Hyv¨arinen 得分有良好定义， P_1 必须是非算术型。因此，根据 Woodroofe [17,推论 2.2。],

$$\mathbb{E}_1[T'_{\texttt{SCUSUM}}] = \frac{\tau}{\mu} + \frac{\mathbb{E}_1[R_\tau]}{\mu} \leq \frac{\tau}{\mu} + \frac{\mu^2 + \sigma^2}{\mu^2}, \ \forall \tau \geq 0.$$

注意，对于任意 n, $Z'(n) \leq Z(n)$，因此 $T_{\texttt{SCUSUM}} \leq T'_{\texttt{SCUSUM}}$。于是，

$$\mathbb{E}_1[T_{\texttt{SCUSUM}}] \leq \mathbb{E}_1[T'_{\texttt{SCUSUM}}] \leq \frac{\tau}{\mu} + \frac{\mu^2 + \sigma^2}{\mu^2}, \ \forall \tau \geq 0. \tag{19}$$

根据定理 4，我们选择 $\tau = \log \gamma$ 以满足约束 $\mathbb{E}_\infty[T_{\texttt{SCUSUM}}] \geq \gamma > 0$。将其代回公式（19），我们得出，当 $\gamma \to \infty$ 时,

$$\mathbb{E}_1[T_{\texttt{SCUSUM}}] \sim \frac{\log \gamma}{\mu} = \frac{\log \gamma}{\lambda \mathbb{D}_{\texttt{F}}(P_1 || P_\infty)},$$

以完成证明。

由于 SCUSUM 的停止方案，期望时间 $\mathbb{E}_\nu[T_{\text{SCUSUM}} - \nu | T_{\text{SCUSUM}} \geq \nu]$ 与变化点 ν 无关（这一点显而易见，Xie 等人 [26]已证明 CUSUM 也具有相同的性质）。令 $\nu = 1$，则有

$$\mathcal{L}_{\text{CADD}}(T_{\text{SCUSUM}}) = \mathbb{E}_1[T_{\text{SCUSUM}}] - 1.$$

因此，我们得出结论：

$$\mathcal{L}_{\text{CADD}}(T_{\text{SCUSUM}}) \sim \frac{\log \gamma}{\lambda \mathbb{D}_{\text{F}}(P_1 \| P_\infty)}.$$

类似的论证也适用于 $\mathcal{L}_{\text{WADD}}(T_{\text{SCUSUM}})$。 □

$\mathbb{E}_1[T_{\text{SCUSUM}}]$ 在文献中也称为期望检测延迟 （EDD）。定理 3 和定理 4 表明，在满足 ARL 约束的条件下，EDD 随停止阈值 τ 的增加而线性增长。

注 2. 值得注意的是，尽管我们的核心结果适用于预先选定且满足不等式（10）的 λ，但选择任何其他 λ' 的效果相当于将 SCUSUM 的所有增量按常数因子 λ'/λ 进行缩放。这意味着，对结果作相应的缩放因子调整后，这些结果仍然成立。例如，定理 3 的结果可改写为

$$\mathbb{E}_\infty[T_{\text{SCUSUM}}] \geq \exp\left\{\frac{\lambda \tau}{\max(\lambda, \lambda')}\right\},$$

对任意 $\lambda' > 0$ 均成立。显然，这种缩放会相应改变定理 4 的表述，使其变为

$$\mathbb{E}_1[T_{\text{SCUSUM}}] \sim \frac{\max(\lambda, \lambda')}{\lambda} \frac{\log \gamma}{\lambda' \mathbb{D}_{\text{F}}(P_1 || P_\infty)},$$

当 $\gamma \to \infty$ 时。为了使定理 3 和定理 4 的结果最强，我们必须选择尽可能接近 λ^* 的 λ。

最后，我们考虑一种特殊情形，即变化前后分布均为多元正态分布。在这种情形下，SCUSUM 在 Pollak 指标和 Lorden 指标的意义下达到渐近最优性。

命题 5 （多元正态变化前后分布）。 假设 $X_1, \cdots, X_{\nu-1} \sim N(\boldsymbol{\theta}_0, \Sigma)$，且 $X_\nu, X_{\nu+1}, \cdots \sim N(\boldsymbol{\theta}_1, \Sigma)$。设 $\boldsymbol{\theta}_0, \boldsymbol{\theta}_1 \in \Theta \subset \mathbb{R}^d$ 和 $\Sigma \in \mathbb{R}^{d \times d}$ 已知，且 $\Sigma = \sigma_c \mathbf{I}_d$，其中标量 $\sigma_c > 0$。那么，当 $\gamma \to \infty$ 时，停止规则 T_{SCUSUM} 达到问题 ***(1)*** 和问题 ***(4)*** 的渐近最优性，即，当 $\gamma \to \infty$ 时，

$$\mathcal{L}_{WADD}(T_{SCUSUM}) \sim \mathcal{L}_{CADD}(T_{SCUSUM}) \sim \frac{\log \gamma}{\mathbb{D}_{KL}(N(\boldsymbol{\theta}_1, \Sigma) \| N(\boldsymbol{\theta}_0, \Sigma))}, \tag{20}$$

在约束 $\mathbb{E}_\infty[T_{SCUSUM}] \geq \gamma > 0$ 下。

证明。 通过直接计算可得

$$z_\lambda(X) = \lambda \left(-\frac{1}{2}(X - \boldsymbol{\theta}_0)^T \Sigma^{-2} (X - \boldsymbol{\theta}_0) + \frac{1}{2}(X - \boldsymbol{\theta}_1)^T \Sigma^{-2} (X - \boldsymbol{\theta}_1) \right),$$

其中 Σ^{-2} 是 $\Sigma^{-1} \cdot \Sigma^{-1}$的简写。接着

$$\mathbb{E}_\infty[\exp(z_\lambda(X))] = \int_{X \in \mathcal{X}} \frac{1}{\sqrt{2\pi} \det(\Sigma)} \exp\left(-\frac{1}{2}(X - \boldsymbol{\theta}_0)^T \Sigma^{-1} (X - \boldsymbol{\theta}_0) + \frac{\lambda}{2}(X - \boldsymbol{\theta}_0)^T \Sigma^{-2} (X - \boldsymbol{\theta}_0) \right.$$
$$\left. - \frac{\lambda}{2}(X - \boldsymbol{\theta}_1)^T \Sigma^{-2} (X - \boldsymbol{\theta}_1) \right) dX.$$

上述积分可计算为

$$\mathbb{E}_\infty[\exp(z_\lambda(X))] = \exp(-\lambda^2 (\boldsymbol{\theta}_0 - \boldsymbol{\theta}_1)^T \Sigma^{-3} (\boldsymbol{\theta}_0 - \boldsymbol{\theta}_1) + \lambda (\boldsymbol{\theta}_0 - \boldsymbol{\theta}_1)^T \Sigma^{-2} (\boldsymbol{\theta}_0 - \boldsymbol{\theta}_1)).$$

显然，$\mathbb{E}_\infty[\exp(z_\lambda(X))] = 1$ 如果

$$\lambda = \frac{(\boldsymbol{\theta}_0 - \boldsymbol{\theta}_1)^T \Sigma^{-2} (\boldsymbol{\theta}_0 - \boldsymbol{\theta}_1)}{(\boldsymbol{\theta}_0 - \boldsymbol{\theta}_1)^T \Sigma^{-3} (\boldsymbol{\theta}_0 - \boldsymbol{\theta}_1)}.$$

两个正态分布之间的 Fisher 散度和 KL 散度可通过以下公式计算:

$$\mathbb{D}_{\mathrm{F}}(\mathcal{N}(\boldsymbol{\theta}_1,\Sigma)||\mathcal{N}(\boldsymbol{\theta}_0,\Sigma))=(\boldsymbol{\theta}_0-\boldsymbol{\theta}_1)^T\Sigma^{-2}(\boldsymbol{\theta}_0-\boldsymbol{\theta}_1),$$

and

$$\mathbb{D}_{\mathrm{KL}}(\mathcal{N}(\boldsymbol{\theta}_1,\Sigma)||\mathcal{N}(\boldsymbol{\theta}_0,\Sigma))=(\boldsymbol{\theta}_0-\boldsymbol{\theta}_1)^T\Sigma^{-1}(\boldsymbol{\theta}_0-\boldsymbol{\theta}_1),$$

分别。因此

$$\frac{\lambda\mathbb{D}_{\mathrm{F}}(P_1||P_\infty)}{\mathbb{D}_{\mathrm{KL}}(P_1||P_\infty)}=\frac{[(\boldsymbol{\theta}_0-\boldsymbol{\theta}_1)^T\Sigma^{-2}(\boldsymbol{\theta}_0-\boldsymbol{\theta}_1)]^2}{[(\boldsymbol{\theta}_0-\boldsymbol{\theta}_1)^T\Sigma^{-3}(\boldsymbol{\theta}_0-\boldsymbol{\theta}_1)][(\boldsymbol{\theta}_0-\boldsymbol{\theta}_1)^T\Sigma^{-1}(\boldsymbol{\theta}_0-\boldsymbol{\theta}_1)]}.$$

令 $\{v_1,v_2,\cdots,v_d\}$ 表示 Σ 的一组特征向量正交归一基，其对应的特征值为 $\{\sigma_1,\sigma_2,\cdots,\sigma_d\}$。我们可将 $\boldsymbol{\theta}_0-\boldsymbol{\theta}_1$($\boldsymbol{\theta}_0-\boldsymbol{\theta}_1$) 在此正交归一基下写为

$$(\boldsymbol{\theta}_0-\boldsymbol{\theta}_1)=\sum_{k=1}^{d}c_kv_k.$$

通过直接计算可得

$$\frac{\lambda\mathbb{D}_{\mathrm{F}}(P_1||P_\infty)}{\mathbb{D}_{\mathrm{KL}}(P_1||P_\infty)}=\frac{(\sum_{k=1}^{d}\frac{c_k^2}{\sigma_k^2})^2}{(\sum_{k=1}^{d}\frac{c_k^2}{\sigma_k^3})(\sum_{k=1}^{d}\frac{c_k^2}{\sigma_k})}.$$

应用柯西-施瓦茨不等式，可得

$$\lambda\mathbb{D}_{\mathrm{F}}(P_1||P_\infty)\le\mathbb{D}_{\mathrm{KL}}(P_1||P_\infty),$$

当且仅当 $c_i\neq0$ 的所有特征值 $\sigma_i,\ i=1,2,\cdots,d$ 相等时，等号成立。特别地，当 Σ 为数量矩阵时，$\lambda\mathbb{D}_{\mathrm{F}}(P_1||P_\infty)=\mathbb{D}_{\mathrm{KL}}(P_1||P_\infty)$，因此 CUSUM 和 SCUSUM 的渐近性能相同。

□

注 3. 我们注意到，在上述高斯情形（其中概率密度已归一化）下，只要

$$\lambda\mathbb{D}_{\mathrm{F}}(P_1||P_\infty)<\mathbb{D}_{\mathrm{KL}}(P_1||P_\infty),$$

CUSUM 的性能优于 SCUSUM。然而，CUSUM 并不容易应用于未归一化模型。这是 SCUSUM 为发挥其计算优势所付出的一点代价。

V. Numerical Results

本节中，我们在合成数据上进行了大量数值实验，以比较我们的方法与各种变化检测算法的性能。

A. 实验设置

a) 数据集：我们模拟来自多元正态分布（MVN）、指数族（EXP）的一个子族 [20]以及高斯-伯努利受限玻尔兹曼机（GB-RBM） [11]的合成数据流。对于指数族，我们使用哈密顿蒙特卡洛（HMC）从未归一化模型中生成样本。我们通过数值积分计算归一化常数，从而基于对数似然执行 CUSUM。值得注意的是，当 EXP 的维度较高时，这一计算难以进行。我们使用吉布斯抽样从 GB-RBM 中抽取样本，并进行1000 次迭代以确保收敛。由于难以计算 GB-RBM 的精确对数似然，我们未提供 GB-RBM 的 CUSUM 结果。有关这些分布的更多详情，请参见 V-B 小节。

b) 基线：我们使用经验平均运行长度和经验最坏条件平均检测延迟来评估性能，其中平均运行长度和最坏条件平均检测延迟分别由 $\mathbb{E}_\infty[T]$ 和 $\mathbb{E}_\nu[T-\nu|T\geq\nu]$给出。当没有变化时，我们期望经验平均运行长度较大；当发生变化时，我们期望经验最坏条件平均检测延迟较小。所有经验最坏条件平均检测延迟和经验平均运行长度的结果均以对数尺度报告。在所有实验中，我们将变化点设为 $\nu=500$。为确保数据流足够长以便检测，我们将总长度固定为10000。平均运行长度的取值范围为500至20000。其理论性质已在第 III 节中讨论。

我们将 SCUSUM 与另外三种方法进行比较：CUSUM [6], 扫描B统计量 [18, 27], 和 CALM-MMD [19]。对于 CUSUM 和 SCUSUM，我们采用算法 1。对于给定的平均运行长度，停止阈值根据公式 (6) 选为 $\tau=\log($平均运行长度$)$。扫描B统计量算法的灵感来自 B统计量 [28]。它定义为数据流的滑动自助法数据块之间的核化最大均值差异（MMD）。Cobb等人 [19] 提出了一种动态选择阈值的算法，名为 CALM，适用于大多数类似双样本检验的变化检测方法。将 CALM 程序应用于核化双样本 MMD 统计量 [29]，即可得到 CALM-MMD 算法。我们使用 [19]发布的代码实现扫描B统计量和 CALM-MMD 算法。这两种方法都是基于核化 MMD 的方法，采用高斯径向基函数（RBF）核。它们的停止阈值根据历史观测值进行经验选择，这在实践中可能导致严重的校准偏差，正如 [19] 以及我们之后的数值结果所示。

除了评估平均运行长度与最坏条件平均检测延迟之间的权衡外，我们还研究了细微变化情况下的变化检测性能，即变化前后分布彼此非常接近的情况。接近程度由参数漂移的幅度衡量。在这里，我们固定变化前分布，并通过扰动其参数来构造变化后分布，从而开展实验。对于不同的分布族，我们考虑不同幅度的扰动。每项实验重复100 次试验。

B. 合成数据集

a) 多元正态分布（*MVN*）：我们考虑多元正态分布。令 $\boldsymbol{\mu}$ 和 Σ分别表示均值和协方差矩阵。相应的得分函数计算如下:

$$S_{\mathrm{H}}(X,P)=\frac{1}{2}(X-\boldsymbol{\mu})^T\Sigma^{-2}(X-\boldsymbol{\mu})-\mathrm{tr}(\Sigma^{-1}),$$

其中，算子$\mathrm{tr}(\cdot)$ 表示求矩阵的迹。

我们考虑均值为 $\boldsymbol{\mu}=(0,0)^T$、协方差矩阵为 $\Sigma=\begin{pmatrix}1 & 0.5\\ 0.5 & 1\end{pmatrix}$的变化前分布。对于变化后分布，我们首先研究均值变化的情形：固定协方差矩阵$\Sigma=\begin{pmatrix}1 & 0.5\\ 0.5 & 1\end{pmatrix}$，并将变化后均值设为 $\boldsymbol{\mu}=(0,0)^T+\epsilon_\mu$，其中 $+$表示逐元素相加，ϵ_μ表示对 $\boldsymbol{\mu}$的扰动。我们令ϵ_μ从0取到0.5，步长为0.05。接下来，我们考虑协方差变化的情形。在此情形下，我们将变化后均值固定为 $\boldsymbol{\mu}=(0,0)^T$，并将变化后协方差设为 $\Sigma=\begin{pmatrix}1 & 0.5\\ 0.5 & 1\end{pmatrix}\circ\exp(\epsilon_{\log(\sigma^2)})$，其中 $\circ$ 表示逐元素乘积，$\epsilon_{\log(\sigma^2)}$表示对协方差矩阵逐元素施加的扰动。为使扰动后的协方差矩阵正定，我们对协方差矩阵各分量的对数进行扰动。我们令$\epsilon_{\log(\sigma^2)}$从0.05变化到0.5，步长为0.05。

b) 指数族（指数分布）：如第 II-C 小节所述，我们考虑其概率密度函数如下的指数族:

$$p_\tau(X)=\frac{1}{Z_\tau}\exp\left\{-\tau\left(\sum_{i=1}^d x_i^4+\sum_{1\leq i\leq d,i\leq j\leq d}x_i^2x_j^2\right)\right\}.$$

相应的 Hyvärinen 评分函数计算如下:

$$S_{\mathrm{H}}(X,P_\tau)=\frac{1}{2}\sum_{i=1}^d\left(\frac{\partial}{\partial x_i}\log p_\tau(X)\right)^2+\sum_{i=1}^d\frac{\partial^2}{\partial x_i}\log p_\tau(X),$$

其中

$$\frac{\partial}{\partial x_i}\log p_\tau(X) = -\tau\left(4x_i^3 + 2\sum_{1\le i\le d, i\le j\le d} x_i x_j^2\right), \text{ and}$$

$$\frac{\partial^2}{\partial x_i}\log p_\tau(X) = -\tau\left(12x_i^2 + 2\sum_{1\le i\le d, i\le j\le d} x_j^2\right).$$

我们考虑具有 $\tau = 1$ 的变化前分布和具有 $\tau = 1+\epsilon_\tau$ 的变化后分布，其中 ϵ_τ 表示尺度参数 τ 的扰动。我们将 ϵ_τ 的取值范围设为从 0.1 到 2.0，步长为 0.1。

c) 高斯-伯努利受限玻尔兹曼机（*GB-RBM*）：如第 II-C 小节所述，我们考虑 GB-RBM 模型，其概率密度函数为 $p_\theta(X) = \sum_{h\in\{0,1\}^{d_h}} p_\theta(X,H) = \frac{1}{Z_\theta}\exp\{-F_\theta(X)\}$，其中 $F_\theta(X)$ 是自由能，定义为

$$F_\theta(X) = \frac{1}{2}\sum_{i=1}^{d_x}(x_i - b_i)^2 - \sum_{j=1}^{d_h}\text{Softplus}\left(\sum_{i=1}^{d_x} W_{ij}x_i + b_j\right).$$

我们以闭式形式计算相应的 Hyv¨arinen 分数

$$S_{\text{H}}(X, P_\theta) = \sum_{i=1}^{d_x}\left[\frac{1}{2}\left(x_i - b_i + \sum_{j=1}^{d_h} W_{ij}\phi_j\right)^2 + \sum_{j=1}^{d_h} W_{ij}^2\phi_j\left(1-\phi_j\right) - 1\right],$$

其中 $\phi_j \triangleq \text{Sigmoid}(\sum_{i=1}^{d_x} W_{ij}x_i + b_j)$。Sigmoid 函数定义为 $\text{Sigmoid}(y) \triangleq (1+\exp(-y))^{-1}$。

变化前分布的参数为 $\mathsf{W} = \mathsf{W}_0$, $\mathsf{b} = \mathsf{b}_0$, 和 $\mathsf{c} = \mathsf{c}_0$, ，其中 $\mathsf{W}_0, \mathsf{b}_0,$ 和 c_0 的每个分量均从标准正态分布 $\mathcal{N}(0, 1)$ 中随机抽取。对于变化后分布，我们设定参数 $\mathsf{W} = \mathsf{W}_0 + \epsilon_{\mathbf{W}}$、$\mathsf{b} = \mathsf{b}_0,$ 和 $\mathsf{c} = \mathsf{c}_0$。这里，我们仅考虑权重矩阵 W 的偏移，记为 $\epsilon_{\mathbf{W}}$。$\epsilon_{\mathbf{W}}$ 的每个分量均从 $\mathcal{N}(0, \sigma_\epsilon^2)$ 中抽取。我们令 σ_ϵ 的取值范围为 0.005 至 0.1，步长为 0.005。

C. 实验结果

a) 检测评分：我们在图1中展示了各时间步的即时检测评分。我们将平均运行长度控制为固定值2000。数据流由二元正态分布（MVN-ϵ_μ）生成，并在时间 $t = 500$发生均值漂移 $\epsilon_\mu = 0.3$。我们报告了以实线表示的平均检测评分，以及以阴影区间表示的标准误。如图1所示，在变化点处，CUSUM 和 SCUSUM 在变化发生后立即作出反应。相比之下，扫描B统计量和 CALM-MMD 的检测评分在0 和1之间波动。在这种情况下，这两种基于 MMD 的方法未能检测到变化。具体而言，CUSUM 和 SCUSUM 的检测评分在变化发生后单调增加。然而，扫描B统计量和 CALM-MMD 的检测评分在变化发生后保持稳定水平。因此，结果表明，即使经过足够多的时间步，扫描B统计量和 CALM-MMD 也可能无法达到阈值。

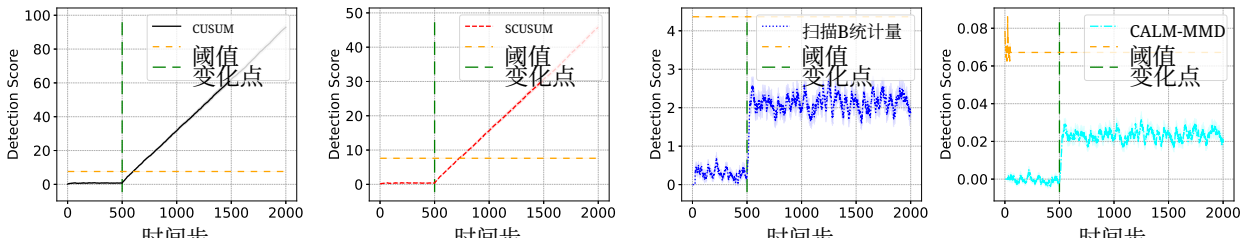


图 1. 使用多元正态分布（$\epsilon_\mu = 0.3$）和平均运行长度= 2000时的检测评分结果（变化前后）。

b) 经验条件平均检测延迟与平均运行长度：在图2中，我们分别展示了二元正态分布均值漂移（MVN-ϵ_μ）、二元正态分布协方差漂移（MVN-$\epsilon_{\log(\sigma^2)}$）、指数族的尺度参数漂移（EXP-$\epsilon_\tau$）以及 GB-RBM 的权重矩阵漂移（GB-RBM-σ_ϵ）情况下，经验条件平均检测延迟与平均运行长度之间的关系。符号 ϵ_μ、 $\epsilon_{\log(\sigma^2)}$、 ϵ_τ 和 σ_ϵ 分别表示 MVN 均值、MVN 协方差矩阵、EXP 尺度参数和 GB-RBM 权重矩阵的偏移幅度。结果表明，就经验条件平均检测延迟与平均运行长度而言，我们提出的 SCUSUM 与 CUSUM 具有相当的竞争力。具体而言，我们看到红线（SCUSUM）与黑线（CUSUM）相互平行，这意味着 SCUSUM 的经验条件平均检测延迟以与 CUSUM 相似的速率增加。此外，在图2(b、c)中，对于固定的平均运行长度，SCUSUM 也能优于 CUSUM。

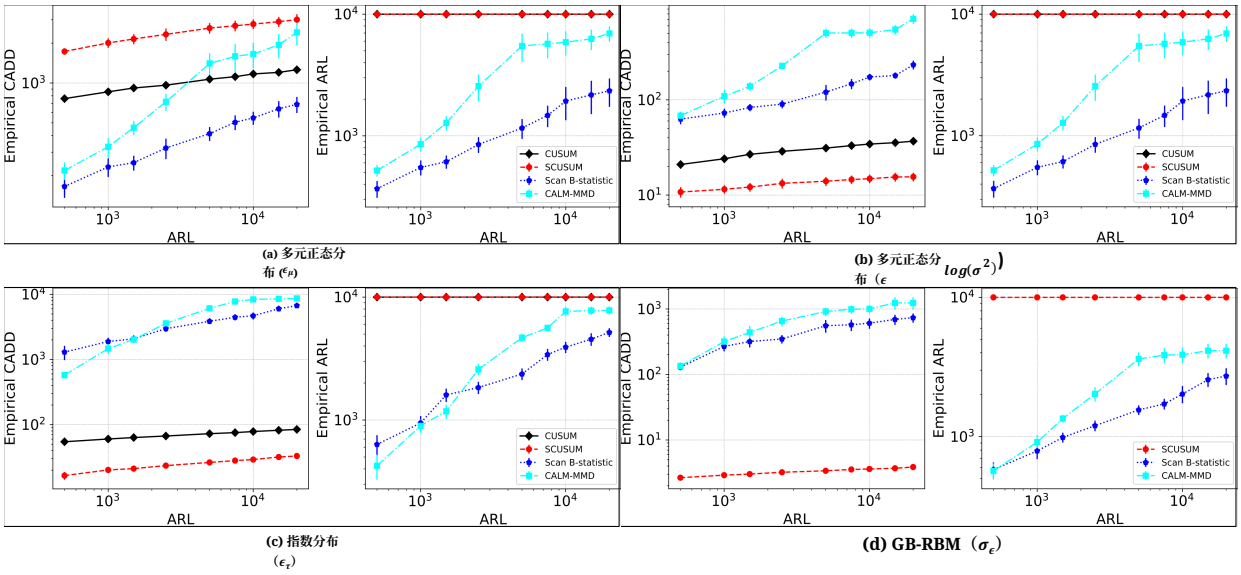
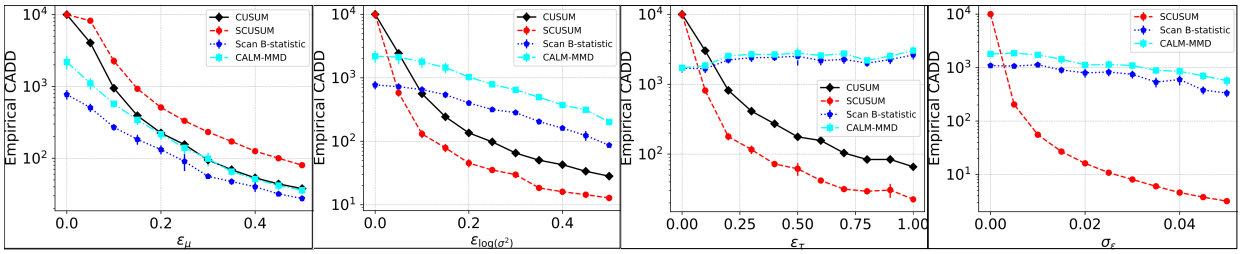


图 2. 多元正态分布（$\epsilon_\mu = 0.1$）、多元正态分布（$\epsilon_{\log(\sigma^2)} = 0.5$）、指数分布（$\epsilon_\tau = 1.0$）和 GB-RBM（$\sigma_\epsilon = 0.05$）的经验条件平均检测延迟与平均运行长度的关系，以及经验平均运行长度与平均运行长度的关系。

图2中每个子图的右列展示了在所有时间步内均未发生变化时，经验平均运行长度与平均运行长度的关系。结果表明，CUSUM 和 SCUSUM 能够成功控制虚警率，而基于 MMD 的方法则无法做到这一点。对于正态分布均值偏移，在平均运行长度较低时，扫描B统计量和 CALM-MMD 的表现优于 CUSUM 和 SCUSUM。然而，我们指出，这种优势是以虚警失控为代价的，如图2(a)的右列所示。此外，基于 MMD 的方法不仅无法控制虚警，其表现也劣于 CUSUM 和 SCUSUM，如图2(b-d)所示。

c) 经验条件平均检测延迟与变化的关系：我们在图3中研究了检测方法在微小变化情况下的性能，即变化前后分布彼此非常接近。在微小变化场景下，图3(b-d)显示，CUSUM 和 SCUSUM 的表现优于基于 MMD 的方法。特别是，随着变化幅度增大，CUSUM 和 SCUSUM 的经验条件平均检测延迟显著更小。尽管基于 MMD 的方法在图3(a)中的表现优于 CUSUM 和 SCUSUM，但需要注意的是，这是以虚警失控为代价的，如图2所示。

d) 计算成本：我们通过改变 EXP 数据集的维度，从计算成本方面将 SCUSUM 与其他基线方法进行比较。计算成本以算法在一轮在线数据流试验中检测到变化并停止所需的运行时间来衡量。请注意，精确似然的归一化常数是在已知变化前后分布的情况下离线计算的。这部分离线计算时间计入 CUSUM 的运行时间。表 I 展示了当维度从 $2D$ 增至 $4D$ 时，CUSUM 所需的运行时间显著增长。这是由于精确对数似然的计算需要进行数值积分。与此同时，SCUSUM 的运行时间略有增长，这是因为需要计算 Hyv¨arinen 分数。随着维度增加，基于 MMD 的方法的运行时间保持稳定。CALM-MMD 由于需要计算并选择候选阈值，运行时间长得多。



(a) 多元正态分布 (b) 多元正态分布 ($\epsilon_{log(\sigma^2)}$) (c) 指数分布 ((d) GB-RBM (σ_ϵ)

图 3. 扰动下的经验条件平均检测延迟与平均运行长度= 2000的关系

表 I 指数族变化检测中各次试验的运行时间（秒）（变化前后分布属于指数族）

检测算法	$d=1$	$d=2$	$d=3$	$d=4$
CUSUM	2.4	2.9	294.8	66409.2
SCUSUM	2.2	9.1	21.0	38.4
扫描B统计量	8.1	8.2	8.2	8.3
CALM-MMD	111.9	111.4	110.2	111.0

e) λ的选择：在实践中，我们将 λ 选为函数 $h(\lambda)$ 的正根，如第 IV 节公式 (11) 所定义。过去观测值的不同样本可能会得到不同的 λ 值，这可能导致 SCUSUM 的性能不一致。接下来，我们通过数值模拟研究这一问题。在图 4 (a) 至 (d) 中，数据流分别由具有 $\epsilon_\mu = 0.1$ 的多元正态分布、具有 $\epsilon_{\log(\sigma^2)} = 0.5$ 的多元正态分布、具有 $\epsilon_\tau = 1$ 的指数分布，以及具有 $\sigma_\epsilon = 0.05$ 的 GB-RBM 生成。图 4 第一列展示了确定的 λ 值如何随过去观测值的数量而变化。图 4 第二列（和第三列）报告了 SCUSUM 的经验条件平均检测延迟（以及经验平均运行长度）如何随过去观测值的数量而变化。所有数值均为 100 次随机运行的平均值，并附有误差线。

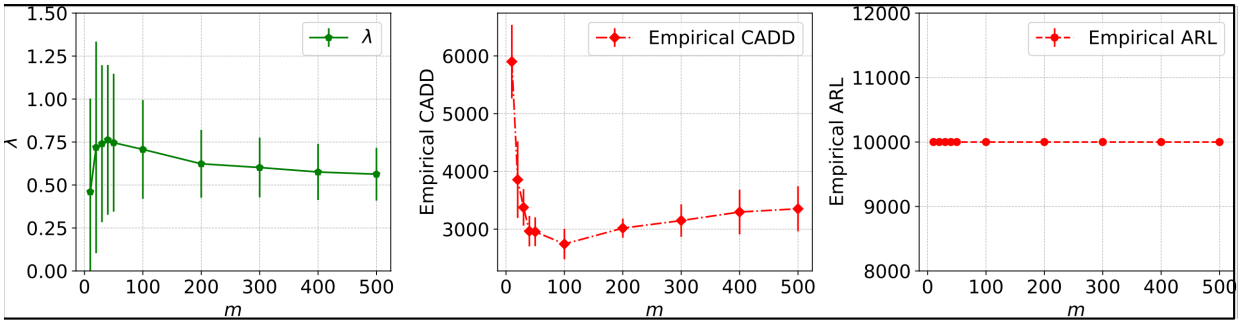
如图 4 所示，只要 m 足够大，λ 的值对不同样本就不会太敏感。具体而言，当 $m > 100$时，我们可以看到图 4(a)-(c)中的标准误较小。因此，SCUSUM 在经验最坏条件平均检测延迟方面的表现往往较为稳定。请注意，在 GB-RBM 的情况下（如图 4(d) 所示），当 $m < 300$时，我们取 $\lambda = 1$ 。这是因为在过去观测值数量较少时，我们无法通过数值方法求得公式 (11) 的正根。最后，如图 4 第 3 列所示，经验平均运行长度始终处于受控状态。

VI. 总结

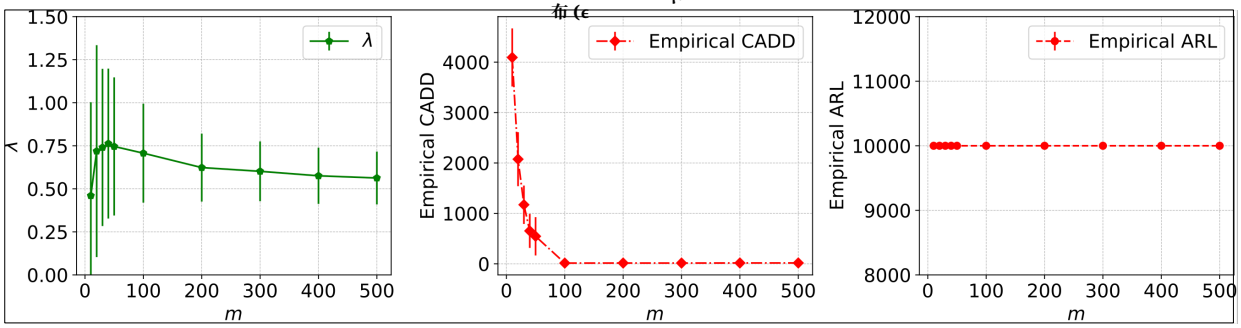
本文提出了用于检测未归一化模型变化的 SCUSUM 算法。我们的检测算法遵循经典的 CUSUM 检测算法，并具有递归实现带来的计算优势。我们从 Lorden 指标和 Pollak 指标的角度分析了 SCUSUM 的渐近性质。我们还提供了数值结果，展示了其在性能提升和降低计算复杂度方面的潜力。在未来的工作中，我们将研究放宽对已知变化后分布和数据独立性的假设所带来的影响。

参考文献

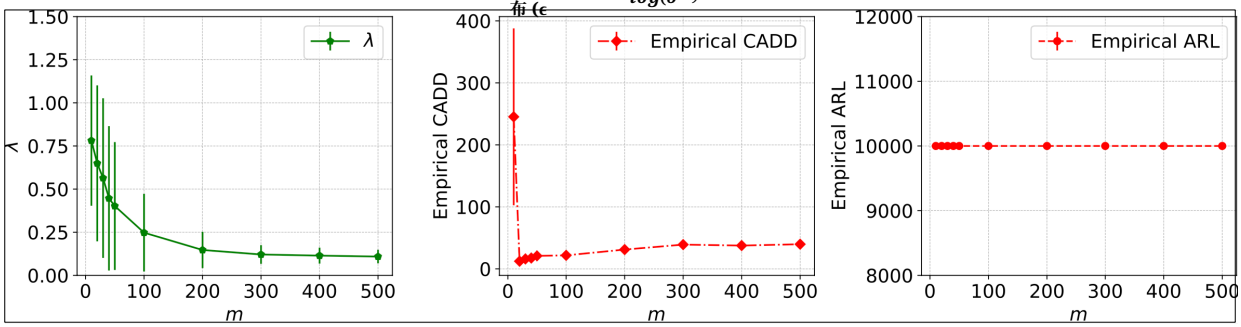
[1] V. V. Veeravalli 和 T. Banerjee，“最快变化检测”，载于*Academic press library in signal processing*。爱思唯尔，2014 年，第 3 卷，第 209–255 页。[2] A. N. Shiryaev，“关于最快检测问题中的最优方法”，概率论及其应用，第 8 卷，第 1 期，第 22–46 页，1963 年。[3] G. Lorden，“对分布变化作出反应的程序”，《数理统计年刊》，第 1897–1908 页，1971 年。[4] M. Pollak，“分布变化的最优检测”，《统计年刊》，第 206–227 页，1985 年。



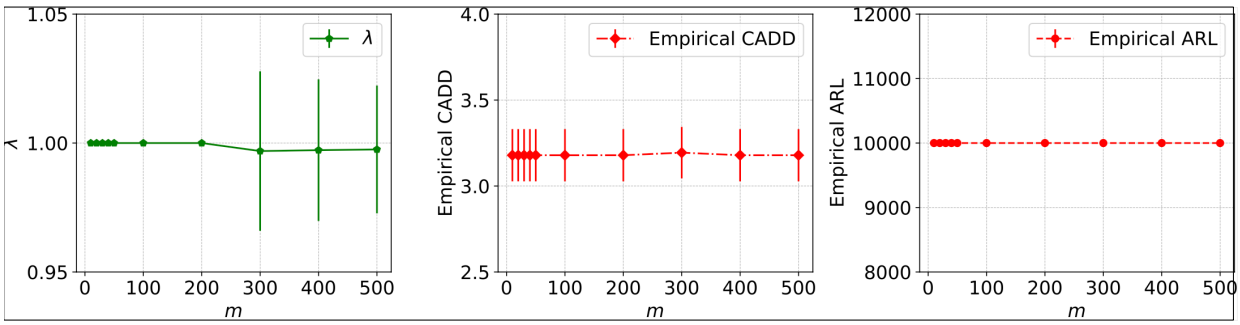
(a) 多元正态分布 ($\mu = 0.1$)



(b) 多元正态分布 ($log(\sigma^2) = 0.5$)



(c) 指数分布 ($\epsilon_\tau = 1$)



(d) GB-RBM（$\sigma_\epsilon = 0.05$）

图4。 第1列： λ 与 m；第2列：经验条件平均检测延迟与 m；第3列：经验平均运行长度与 m .

[5] G. V. Moustakides, 《检测分布变化的最优停止时间》, 《统计年刊》, 第 14 卷, 第 4 期, 第 1379–1387 页, 1986 年。[6] E. Page, 《对发生于未知时刻的参数变化进行检验》, 《生物统计学》, 第 42 卷, 第 3/4 期, 第 523–527 页, 1955 年。[7] S. Roberts, 《若干控制图程序的比较》, 《计量学》, 第 8 卷, 第 3 期, 第 411–430 页, 1966 年。[8] 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, 第 2917–2929 页, 1998 年。[9] A. G. Tartakovsky 和 V. V. Veeravalli, 《最快变化检测的一般渐近贝叶斯理论》, 《概率论及其应用》, 第 49 卷, 第 3 期, 第 458–497 页, 2005 年。[10] A. S. Polunchenko 和 A. G. Tartakovsky, 《序贯变点检测的最新进展》, 《方法论与计算应用》, 第 14 卷, 第 3 期, 第 649–684 页, 2012 年。[11] Y. LeCun、S. Chopra、R. Hadsell、M. Ranzato 和 F. Huang, 《基于能量的学习教程》, 载于《结构化数据预测》。麻省理工学院出版社, 2006 年, 第 1 卷。[12] 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评分的生成建模》, arXiv 预印本 arXiv:2011.13456, 2020 年。[13] H. Chen 和 N. Zhang, 《基于图的变点检测》, 《统计年刊》, 第 43 卷, 第 1 期, 第 139–176 页, 2015 年。[14] E. Nalisnick、A. Matsukawa、Y. W. Teh、D. Gorur 和 B. Lakshminarayanan, 《深度生成模型知道自己不知道什么吗?》, arXiv 预印本 arXiv:1810.09136, 2018 年。[15] A. Hyv¨arinen 和 P. Dayan, 《通过得分匹配估计未归一化统计模型》, 《机器学习研究期刊》, 第 6 卷, 第 4 期, 2005 年。[16] S. Wu、E. Diao、K. Elkhalil、J. Ding 和 V. Tarokh, 《针对未归一化模型的基于评分的假设检验》, 《IEEE Access》, 第 10 卷, 第 71 936–71950 页, 2022 年。[17] M. Woodroofe, 《序贯分析中的非线性更新理论》。美国工业与应用数学学会, 1982 年。[18] S. Li、Y. Xie、H. Dai 和 L. Song, 《用于核变点检测的扫描B统计量》, 《序贯分析》, 第 38 卷, 第 4 期, 第 503–544 页, 2019 年。[19] O. Cobb、A. Van Looveren 和 J. Klaise, 《具有校准且无记忆误检率的序贯多元变化检测》, 载于国际人工智能与统计学会议（AISTATS）。PMLR, 2022 年, 第 226–239 页。[20] M. Yu、M. Kolar 和 V. Gupta, 《使用得分匹配对成对图模型进行统计推断》, 《神经信息处理系统进展》（NeurIPS）, 第 29 卷, 2016 年。[21] T. Gneiting 和 A. E. Raftery, 《严格适当的评分规则、预测与估计》, 《美国统计学会杂志》, 第 102 卷, 第 477 期, 第 359–378 页, 2007 年。[22] I. J. Good, 《理性决策》, 载于《统计学突破》。施普林格, 1992 年, 第 365–377 页。[23] M. Parry、A. P. Dawid 和 S. Lauritzen, 《适当局部评分规则》, 《统计年刊》, 第 40 卷, 第 1 期, 第 561–592 页, 2012 年。[24] J. L. Doob, 《随机过程》。威利, 纽约, 1953 年, 第 7 卷。[25] G. Lorden, 《关于越界超额》, 《数理统计年刊》, 第 41 卷, 第 2 期, 第 520–527 页, 1970 年。[26] L. Xie、S. Zou、Y. Xie 和 V. V. Veeravalli, 《序贯（最快）变化检测：经典结果与新方向》, 《IEEE 信息论选题领域期刊》（JSAIT）, 第 2 卷, 第 2 期, 第 494–514 页, 2021 年。[27] S. Li、Y. Xie、H. Dai 和 L. Song, 《用于核变点检测的 M 统计量》, 《神经信息处理系统进展》（NeurIPS）, 第 28 卷, 2015 年。[28] W. Zaremba、A. Gretton 和 M. Blaschko, 《B 检验：一种非参数、低方差的核双样本检验》, 《神经信息处理系统进展》（NeurIPS）, 第 26 卷, 2013 年。[29] A. Gretton、K. M. Borgwardt、M. J. Rasch、B. Sch¨olkopf 和 A. Smola, 《核双样本检验》, 《机器学习研究期刊》, 第 13 卷, 第 1 期, 第 723–773 页, 2012 年。