arXiv:2302.04743v1 [stat.CO] 9 Feb 2023

一种适用于指数族模型在线变点检测的每次迭代时间恒定的似然比检验

Kes Ward
k.ward4@lancaster.ac.uk 数学与统计系，兰卡斯特大学，英国

Gaetano Romano
g.romano@lancaster.ac.uk 数学与统计系，兰卡斯特大学，英国

Idris Eckley
i.eckley@lancaster.ac.uk 数学与统计系，兰卡斯特大学，英国

Paul Fearnhead
p.fearnhead@lancaster.ac.uk 数学与统计系，兰卡斯特大学，英国

2023年2月10日

摘要

基于似然比检验的在线变点检测算法已被证明具有优异的统计性质。然而，简单的在线实现计算上不可行，因为在时刻 T，它需要考虑 $O(T)$ 个可能的变化位置。近期提出的 FOCuS 算法可检测高斯数据中的均值变化，并将每次迭代的开销降至 $O(\log T)$。这是通过运用剪枝思想实现的：在时刻 T，这些思想将需要考虑的变点位置集合缩减至约 $\log T$ 个。我们证明，对于另一种单参数指数族模型执行似然比检验时，可以使用完全相同的剪枝规则，并且在第 T 次迭代中，同样只需考虑约 $\log T$ 个位置。此外，我们还展示了如何自适应地执行算法中的最大化步骤，从而只需在这些可能位置的一个较小子集上对检验统计量求最大值。实证结果表明，由此得到的在线算法能够检测多种模型下的变化，且平均每次迭代的开销为常数。

1 引言

检测数据流中的变化是统计学和机器学习领域的一项重要挑战，广泛存在于气候记录（Beaulieu and Killick, 2018）、金融时间序列（Andreou and Ghysels, 2002）、虚拟机性能监控（Barrett et al., 2017）以及检测分类器输入中的概念漂移（Sakamoto et al., 2015）等应用中。在许多当代应用中，都需要在线检测变化。在这类场景下，我们会随时间顺序监控数据流，并力求尽快发出变化已经发生的警报。在线变化检测算法通常需要在计算资源有限的条件下运行。例如，Ward et al. (2022) 利用小型立方体卫星上的本地计算资源检测伽马射线暴；Varghese et al. (2016) 则研究传感器网络，其中计算需要由传感器在本地执行。此外，算法可能需要处理超高频数据（Iwata et al., 2018），或需要在大量独立数据流上并行运行。这些场景的共同特点是，对可行算法的计算复杂度有严格限制。

已有多种用于在线检测变化的方法，每种方法在统计效率与计算成本之间都涉及不同的权衡。例如，Yu 等人（2020）提出了一种似然比检验，具有优异的统计性质，但该方法的自然实现方式会使每次迭代的计算成本随时间线性增长。然而，对于在线应用，我们需要计算成本保持恒定。虽然存在每次迭代计算成本恒定的算法，但这些算法要么只能检测发生在预先指定时间之前的变化（例如 Eichinger and Kirch，2018；Ross 等人，2011；Ross 和 Adams，2012；Chen 和 Tian，2010），要么需要指定变化后数据的分布（例如 Page，1954；Lucas，1985）。如果实现这些算法时所作的选择不适用于实际希望检测的变化，就可能导致检验功效大幅下降。

最近，Romano 等人（2021a）提出了一种名为函数型在线累积和（FOCuS）的新算法。该算法能够执行似然比检验，且计算成本随时间仅呈对数增长。FOCuS 最初用于检测高斯数据中的均值变化，后来扩展至泊松数据（Ward 等人，2022）和二项分布数据（Romano 等人，2023）。FOCuS 包含两个部分：一部分用于剪枝，即排除未来无需考虑的过去变点时刻；另一部分是最大化步骤，用于考察所有尚未被剪枝的过去变点时刻。有趣的是，针对泊松数据和二项分布数据的剪枝步骤与针对高斯数据的剪枝步骤完全相同，只有最大化步骤有所不同。

本文表明，这种对应关系可以推广到其他单参数指数族模型。此外，我们还展示了如何大幅提升 FOCuS 的速度。在此前的实现中，剪枝步骤的平均计算成本在每次迭代中固定不变，而计算瓶颈在于最大化步骤：在时刻 T，该步骤平均需要考察 $O(\log T)$ 个可能的变点位置。我们展示了如何存储此前的计算结果，从而使最大化步骤能够考察更少的过去变点位置。实证结果表明，这使得最大化步骤每次迭代的计算成本达到 $O(1)$。据我们所知，这是首个能够精确执行似然比检验来检测变化、且平均每次迭代成本恒定的算法。

2 背景

2.1 问题陈述

假设我们观测到一个单变量时间序列信号 x_1，x_2……，并希望在线分析数据，尽可能快速地检测数据分布中的任何变化。我们用 T 表示

当前时间点。

解决这个问题的一种自然方法，是将数据建模为来自某个密度为 $f(x \mid \theta)$ 的参数族的独立实现。令 θ_0 为发生任何变化之前的密度参数。若发生变化，将变化时刻记为 τ，并将变化后的参数记为 θ_1。随后，我们可以使用似然比检验统计量来检验是否发生变化。

这类检验有两种情形。首先，我们可以假设变化前的分布已知，因此 θ_0 也已知（Eichinger and Kirch，2018）。当我们拥有大量来自变化前分布的训练数据，可据此估计 θ_0 时，通常会采用这一简化假设。另一种做法是令 θ_0 未知。我们将先关注变化前分布已知的情形，并在第 4 节说明如何将这些思路扩展到变化前分布未知的情形。

数据 $x_{1:T} = (x_1, \ldots, x_T)$ 的对数似然取决于变化前参数 θ_0、变化后参数 θ_1 以及变化的位置 τ，其表达式为

$$\ell(x_{1:T}|\theta_0, \theta_1, \tau) := \sum_{t=1}^{\tau} \log f(x_t|\theta_0) + \sum_{t=\tau+1}^{T} \log f(x_t|\theta_1).$$

因此，在 T 之前发生变化的对数似然比检验统计量为

$$LR_T := 2\left\{\max_{\theta_1,\tau} \ell(x_{1:T}|\theta_0, \theta_1, \tau) - \ell(x_{1:T}|\theta_0, \cdot, T)\right\}.$$

直接计算对数似然比统计量，需要在时刻 T 对一组包含 T 个项的集合取最大值。当 T 较大时，在在线场景下进行计算的成本高得难以承受。有两种简单的现有方法可以解决这个问题，使每次迭代的计算成本保持恒定。第一种是 MOSUM 方法（例如 Chu 等人，1995；Eichinger and Kirch，2018）：固定待检验的变点时刻数量 K，记作 say，并令这些时刻取如下形式：$\tau = T - h_i$，其中 h_1、……、h_K 的取值需适当选择。另一种方法是使用 Page 递推（Page，1954，1955），计算针对预先指定的变化后参数的似然比检验统计量。同样，我们可以使用一个包含 K 个可能变化后参数的网格。如果变点位置的选择（即 MOSUM 中的 h_i 值）或变化后参数的选择不适合所分析数据中的实际变化，这两种方法都会降低统计功效。

2.2 用于高斯数据的 FOCuS

作为 MOSUM 或 Page 递推的替代方法，Romano 等人 (2021a) 引入了 FOCuS 算法，可高效计算单变量高斯数据的对数似然比统计量，其中 θ 表示数据均值。

在这种情况下，很容易看出

$$\ell(x_{1:T}|\theta_0, \theta_1, \tau) - \ell(x_{1:T}|\theta_0, \cdot, T) = \sum_{t=\tau+1}^{T} \left\{\log f(x_t|\theta_1) - \log f(x_t|\theta_0)\right\}.$$

接下来，我们可以引入一个函数

$$Q_T(\theta_1) = \max_{\tau}\left\{\sum_{t=\tau+1}^{T}\Big(\log f(x_t|\theta_1) - \log f(x_t|\theta_0)\Big)\right\},$$

当变化后参数 θ_1 已知时，该函数即为对数似然比统计量。显然，$LR_T = \max_{\theta_1} 2Q_T(\theta_1)$。

对于均值 θ_0 和方差 σ^2 已知的高斯数据，我们可以对数据进行标准化，使变化前的均值为 0、方差为 1。在这种情况下，对数似然比统计量求和中的每一项都可简化为 $\theta_1(x_t - \theta_1/2)$，并且

$$Q_T(\theta_1) = \max_{\tau} \left\{ \sum_{t=\tau+1}^{T} \theta_1(x_t - \theta_1/2) \right\}. \tag{1}$$

这是 $T-1$ 个二次函数的逐点最大值。因此，我们只需存储这些二次函数的系数，即可存储 $Q_t(\theta_1)$。

FOCuS 的思想是递归计算 $Q_T(\theta_1)$。虽然我们将 $Q_T(\theta_1)$ 写成了关于 θ_1 的 $T-1$ 个二次函数的最大值，其中每个二次函数分别对应一个假定变化位置，但实际上，只有 $\approx \log T$ 个二次函数会对 Q_T 产生贡献（Romano 等人，2021a）。这意味着，如果我们能够识别出这组二次函数，就可以在 $O(\log T)$ 次运算内最大化 Q_T，进而计算检验统计量。此外，Romano 等人（2021a）表明，我们可以递归计算 Q_T 以及所需的最小二次函数集合，其平均成本为每次迭代 $O(1)$。

对于正向变化的情况，即$\theta_1 > \theta_0$，FOCuS 递推最容易描述。随后可对 $\theta_1 < \theta_0$ 应用相同的递推，并将结果合并，得到 Q_T。这种计算 Q_T 的方法采用了 Page（1954）的递推，

$$Q_T(\theta_1) = \max\{Q_{T-1}(\theta_1), 0\} + \theta_1(x_T - \theta_1/2).$$

为了说明如何高效地求解这一递推，引入一些记号会很有帮助。对于 $\tau_i < \tau_j$，定义

$$\mathcal{C}_{\tau_i}^{(\tau_j)}(\theta_1) = \sum_{t=\tau_i+1}^{\tau_j} \theta_1(x_t - \theta_1/2). \tag{2}$$

在时刻 $T-1$，设对 Q_{T-1}有贡献的二次函数（ $\theta_1 > \theta_0$）对应于在时刻 $\tau \in \mathcal{I}_{T-1}$发生的变化。那么

$$Q_{T-1}(\theta_1) = \max_{\tau \in \mathcal{I}_{T-1}} \left\{ \mathcal{C}_{\tau}^{(T-1)}(\theta_1) \right\}.$$

将其代入 Page 递推可得

$$Q_T(\theta_1) = \max\left\{ \max_{\tau \in \mathcal{I}_{T-1}} \left\{ \mathcal{C}_{\tau}^{(T)}(\theta_1) \right\}, \mathcal{C}_{T-1}^{T}(\theta_1) \right\},$$

由此可得 $\mathcal{I}_T \subseteq \mathcal{I}_{T-1} \cup \{T-1\}$.

现在的关键步骤是确定 $\mathcal{I}_{T-1} \cup \{T-1\}$ 中哪些变点位置不再对 Q_T有贡献。为与第 3 节提出的思路保持一致，我们将以略有不同于 Romano et al. (2021a) 的方式介绍 FOCuS 算法。假设 $\mathcal{I}_{T-1} = \{\tau_1, \ldots, \tau_n\}$，并将候选位置按 $\tau_1 < \tau_2 < \ldots < \tau_n$的顺序排列。现在可以将相邻二次函数之差定义为

$$\begin{aligned} \mathcal{C}_{\tau_i}^{(T)}(\theta_1) - \mathcal{C}_{\tau_{i+1}}^{(T)}(\theta_1) &= \mathcal{C}_{\tau_i}^{(T-1)}(\theta_1) - \mathcal{C}_{\tau_{i+1}}^{(T-1)}(\theta_1) \\ &= \mathcal{C}_{\tau_i}^{(\tau_{i+1})}(\theta_1). \end{aligned}$$

这些差异从时间 $T-1$ 到时间 T不会发生变化。

对于与 τ_i 和 τ_{i+1}处的变化相关联的二次函数之差，令 $l_i \geq 0$ 表示满足 $\mathcal{C}^{(\tau_{i+1})\tau_i}(\theta_1) \geq 0$ 的 θ_1 的最大值。根据定义， $\mathcal{C}^{(\tau_{i+1})\tau_i}(\theta_0) = 0$。因此容易证明

$$\mathcal{C}_{\tau_i}^{(T)}(\theta_1) \geq \mathcal{C}_{\tau_{i+1}}^{(T)}(\theta_1),$$

关于 $\theta \in [\theta_0, l_i]$。对于 $\theta_1 \geq l_i$，比较 $\mathcal{C}_{\tau_{i+1}}^{(T)}(\theta_1)$ 与 $\mathcal{C}^{(T)}{}_{T-1}(\theta_1)$。如果 $\mathcal{C}^{(T)}{}_{\tau_{i+1}}(\theta_1) \leq \mathcal{C}^{(T)}{}_{T-1}(\theta_1)$，那么

$$\begin{aligned} \mathcal{C}_{\tau_{i+1}}^{(T)}(\theta_1) - \mathcal{C}_{T-1}^{(T)}(\theta_1) &\leq 0 \\ \Leftrightarrow \mathcal{C}_{\tau_{i+1}}^{(T-1)}(\theta_1) &\leq 0. \end{aligned}$$

对于所有 $\theta_1 > l_i$，使 $\mathcal{C}_{\tau_{i+1}}^{(T-1)}(\theta_1) \leq 0$ 成立的一个充分条件是， $\mathcal{C}^{(T-1}{}_{\tau_{i+1}}(\theta_1)$ 的最大根小于 l_i。在这种情况下， $\mathcal{C}_{\tau_{i+1}}^{(T)}(\theta_1)$ 不会对 $Q_T(\cdot)$ 产生贡献，因此可以将其剪枝。

由此得到算法1。请注意，该算法的呈现方式与 Romano et al. (2021a) 中的算法不同，因为二次函数的存储方式不同。具体而言，我们在这里存储二次函数之差，而不是使用汇总统计量。输入仅为对 Q_{T-1} 有贡献的二次函数之差。算法的主循环只检查 $\mathcal{C}_{\tau_j}^{(T-1)}$ 的根是否小于 $\mathcal{C}^{(\tau_j)\tau_{j-1}}$ 的根；这是我们剪枝与 τ_j 相关联的二次函数的条件。如果不是，我们就停止进一步剪枝，并返回二次差函数集合以及二次函数 $\mathcal{C}_{T-1}^{(T)}$。如果是，则移除与 τ_j 相关联的二次函数，并更新与 τ_{j-1} 相关联的二次函数之差——方法是将与τ_j 相关联的二次函数之差加上去。然后，我们循环检查是否移除下一个二次函数（如果存在）。

算法的图示见图 1。很容易看出，该算法每次迭代的平均开销为 $O(1)$。这是因为，每次迭代中，while 循环的步数比被剪枝的二次函数数量多一步。由于每次迭代只会添加一个二次函数，且一个二次函数最多只能被移除一次，因此截至时间 T，while 循环的总步数将少于 $2T$。

3 指数族模型的 FOCuS

不同的参数族具有不同的似然函数和似然比统计量。不过，只要我们考虑的是单变量参数的变化，FOCuS 背后的思想仍可应用于这些情况，只需采用不同形式的曲线（见公式 2），从而得到不同的曲线根。尽管人们可能会认为，根的值不同会导致应用算法1时曲线的剪枝方式也不同，但 Ward et al. (2022) 和 Romano et al. (2023) 指出，对于泊松模型或二项模型，与高斯模型相比，剪枝结果（即与对 Q_T 有贡献的函数相关联的变点）相同；变化的仅是函数的形状。这里我们将证明，这是许多单参数指数族模型普遍具有的性质。

单参数指数族分布可以写为

$$f(x \mid \theta) = \exp\big[\, \alpha(\theta) \cdot \gamma(x) - \beta(\theta) + \delta(x) \,\big],$$

其中， $\alpha(\theta)$ $\beta(\theta)$ $\gamma(x)$ $\delta(x)$ 是依赖于具体分布的单参数函数。表 1 中给出的单参数指数族分布示例包括均值变化的高斯分布、方差变化的高斯分布、泊松分布、尺度变化的伽马分布以及二项分布；对于这些分布， $\alpha(\theta)$ 和 $\beta(\theta)$ 都是递增函数。 $\gamma(x)$ 是充分统计量

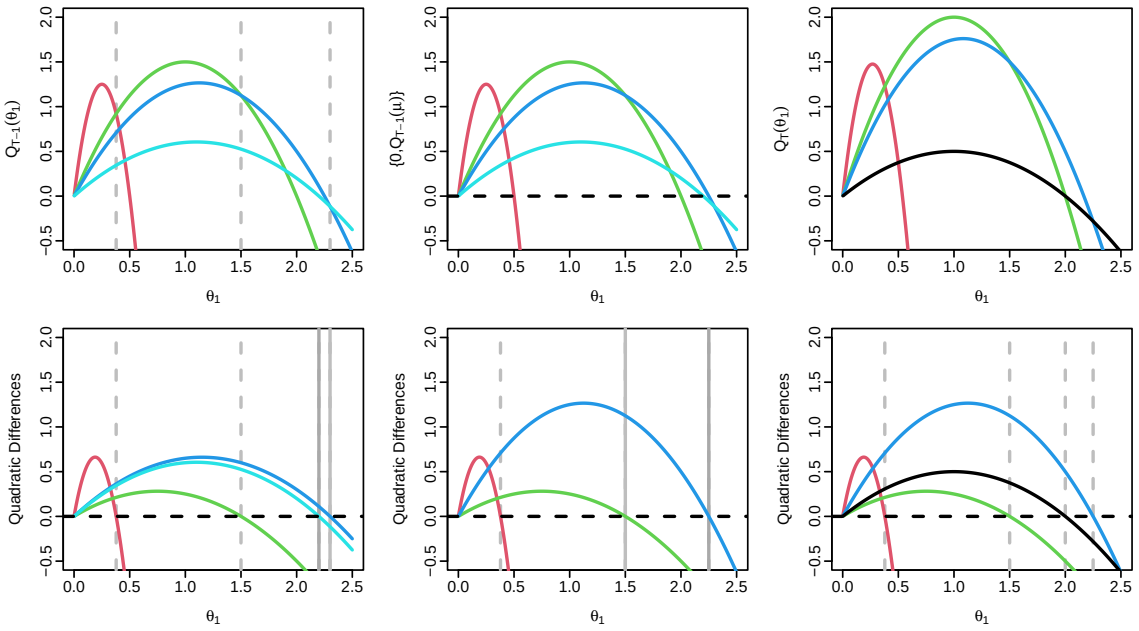


图 1：FOCuS 一次迭代的示例。顶行绘制了直接参与 $Q_T(\theta_1)$ 的二次函数 $C_{\tau_1}^{(T-1)}$（红色）、$C_{\tau_2}^{(T-1)}$（绿色）、$C_{\tau_3}^{(T-1)}$（蓝色）和 $C_{\tau_4}^{(T-1)}$（青色），以及每个二次函数处于最优状态的区间（以灰色竖线标示）。为进行剪枝，我们首先添加零线（黑色虚线），然后剪去 $C_{\tau_4}^{(T-1)}$，因为它对于任何 θ_1 都不再是最优的。接着，我们将 $\theta_1(x_T - \theta_1/2)$ 添加到所有二次函数上。左下图展示了算法1 中二次差函数的存储 $C^{(\tau_2)\tau_1}$（红色）、$C^{(\tau_3)\tau_2}$（绿色）、$C^{(\tau_4)\tau_3}$（蓝色）和 $C_{\tau_4}^{(T-1)}$（青色）。这些二次差函数的根以灰色竖线表示。前三个二次差函数的根标示了二次函数处于最优状态的区间。$C_{\tau_4}^{(T-1)}$ 的根显示了该曲线位于零线之上的区域。算法根据 $C_{\tau_4}^{(T-1)}$ 的根是否小于 $C^{(\tau_4)\tau_3}$ 的根，来判断是否剪去 τ_4。剪去 τ_4 时，将青色与蓝色合并为二次差函数 $C_{\tau_3}^{(T-1)}$（底部中图，蓝线）。随后，我们将 $C_{T-1}^{(T)}$（黑色）作为其自身的二次差函数添加进来（右下图）。我们无需遍历完整的二次函数列表，因为 $C^{(\tau_2)\tau_1}$ 和 $C^{(\tau_3)\tau_2}$ 保持不变。

算法1 FOCuS 在时刻 T 针对 $\theta_1 > \theta_0$ 和 $\theta_0 = 0$的更新。基于存储二次差函数的算法。

输入： 一组 n 二次差函数， $\mathcal{C}^{(\tau_{i+1})\tau_i}(\theta_1)$，其中 $i = 1, \dots, n$，且 $\tau_i < \tau_{i+1}$ 和 $\tau_{n+1} = T-1$满足

$$Q_{T-1}(\theta_1) = \max_i \{\mathcal{C}_{\tau_i}^{(\tau_{i+1})}\}.$$

最大根的集合 l_i，满足 $\mathcal{C}^{(\tau_{i+1})\tau_i}(l_i) = 0$，其中 $i = 1, \dots, n$。

$j \leftarrow n$
$l_0 \leftarrow \theta_0$
当 $j > 0$ **时执行**
  如果 $l_j \le l_{j-1}$ **，则**
    $C_{\tau_{j-1}}^{(T-1)}(\theta_1) \leftarrow C_{\tau_{j-1}}^{\tau_j}(\theta_1) + C^{(T-1}{}_{)\tau_j}(\theta_1)$
    重新计算 l_{j-1}， $C_{\tau_{j-1}}^{(T-1)}(\theta_1) = 0$ 的最大根
    $\tau_j \leftarrow T-1$
    $j \leftarrow j-1$
  结束 if
  中断
结束 while
$\mathcal{C}_{T-1}^{(T)}(\theta_1) \leftarrow \theta_1(x_T - \theta_1/2)$
$\tau_{j+1} \leftarrow T-1$ 和 $\tau_{j+2} \leftarrow T$
$l_{j+1} \leftarrow 2x_T$
$n \leftarrow j+1$

输出： n 二次差函数集合， $\mathcal{C}^{(\tau_i+1)\tau_i}(\theta_1)$ 以及 $i = 1, \dots, n$ 的根 l_i。

对于该模型，它通常是恒等函数。我们无需考虑 $\delta(x)$，因为它会在所有似然比中抵消。

在应用变化检测方法之前，可以通过各种简单的转换，将数据点从一种假定的指数族形式转换为另一种。例如，将指数分布数据分箱到不同的时间区间，从而得到泊松数据；当 n 较大且 θ 较小时，将 Binomal(n, θ) 数据近似为 Poisson$(n\theta)$；或者利用 $x \sim N(0,1)$ 时 $x^2 \sim$ Gamma$(1/2, 1/2)$ 这一事实，将高斯方差变化问题转化为伽马参数变化问题（相关示例见第6节）。不过，能够灵活处理所有可能的指数族情形，而无需进行数据预处理，也很有帮助。

第2.2节中的思路可用于检测单参数指数族参数的变化。主要变化在于对数似然的形式。对于算法1，我们需要存储针对变点位置的不同取值时对数似然中的差值 $C^{(\tau_j)\tau_i}(\theta_1)$ 。这变为

$$\begin{aligned}&\ell(x_{1:T}|\theta_0, \theta_1, \tau_i) - \ell(x_{1:T}|\theta_0, \theta_1, \tau_j) = \\ &\quad = [\alpha(\theta_1) - \alpha(\theta_0)] \sum_{t=\tau_i+1}^{\tau_j} \gamma(x_t) - [\beta(\theta_1) - \beta(\theta_0)](\tau_j - \tau_i).\end{aligned}$$

这些 c 曲线可用 $\alpha(\theta_1) - \alpha(\theta_0)$ 和 $\beta(\theta_1) - \beta(\theta_0)$ 的系数来概括，即

分布	$\alpha(\theta)$	$\beta(\theta)$	$\gamma(x)$
高斯（均值变化）	θ	θ^2	x
高斯（方差变化）	$-1/\theta^2$	$\log(\theta)$	x^2
泊松	$\log(\theta)$	θ	x
二项分布	$\log(\theta) - \log(1-\theta)$	$-n\log(1-\theta)$	x
伽马	$-1/\theta$	$k\log(\theta)$	x

表 1：单参数指数族及其对应的 $\alpha(\theta)$、$\beta(\theta)$ 和 $\gamma(x)$ 形式示例。高斯均值变化模型的方差为 1，高斯方差变化模型的均值为 0；二项模型假设试验次数为 n；伽马模型表示尺度参数发生变化，形状参数为 k。

$\sum_{t=\tau_i+1}^{\tau_j}\gamma(x_t)$ 和 $\tau_j - \tau_i$。Th

剪枝 log-likelihood

方法1基于比较曲线的根。将该算法应用于一般指数族模型的一大挑战在于，根通常无法像高斯模型那样通过解析方法求得，因此需要使用数值求根器。然而，剪枝只取决于根的大小顺序。下面的命题表明，通常无需计算根的具体值，就能确定两条曲线中哪一条的根较大。

定义

$$\bar{\gamma}_{\tau_i:\tau_j} = \frac{1}{\tau_j - \tau_i}\sum_{t=\tau_i+1}^{\tau_j}\gamma(x_t),$$

为 $t = \tau_i + 1, \ldots, \tau_j$ 时 $\gamma(x_t)$ 的平均值，并定义 $\theta_1^{\tau}(\neq \theta_0)$ 为下式的根：

$$\ell(x_{1:T}|\theta_0, \theta_1^{\tau}, \tau) - \ell(x_{1:T}|\theta_0, \cdot, T) = 0.$$

那么，下面的命题表明，根的大小顺序由 $\bar{\gamma}$ 值的大小顺序决定。

命题1 假设对于我们选取的 θ_0，函数

$$\theta_1 :\to \frac{\beta(\theta_1) - \beta(\theta_0)}{\alpha(\theta_1) - \alpha(\theta_0)}$$

严格递增。因此，$\bar{\gamma}_{\tau_i:\,\tau_j} - \bar{\gamma}_{\tau_j:\,T}$ 的符号与$\theta_1^{\tau_i} - \theta_1^{\tau_j}$ 的符号相同。

证明： 参见补充材料。

换言之，$\theta_1^{\tau_i} > \theta_1^{\tau_j}$ 当且仅当$\bar{\gamma}_{\tau_i:\,\tau_j} > \bar{\gamma}_{\tau_j:\,T}$。因此，我们可以将算法1中比较两条曲线的根的条件，改为比较它们的 $\bar{\gamma}$ 值的条件。或者，等价地，我们可以实现算法1，但使用$l_i = \bar{\gamma}_{\tau_i:\,\tau_{i+1}}$，而不是使用 $\mathcal{C}_{\tau_i}^{\tau_{i+1}} = 0$的根。

这一结果的一个直接推论是，满足命题1条件且 $\gamma(x)$取相同值的单参数指数族模型，将恰好剪枝相同的曲线集合。这就得到以下推论：考虑一组具有 $\gamma(x) = x$的指数族模型，该值与原始 FOCuS 算法的均值变化的高斯分布模型相同。

推论 2 高斯（均值变化）、泊松、二项分布和伽马版本的 *FOCuS* 算法具有相同的剪枝方式。

图 2 展示了该推论的一个图示示例。更一般地，我们有如下结论。

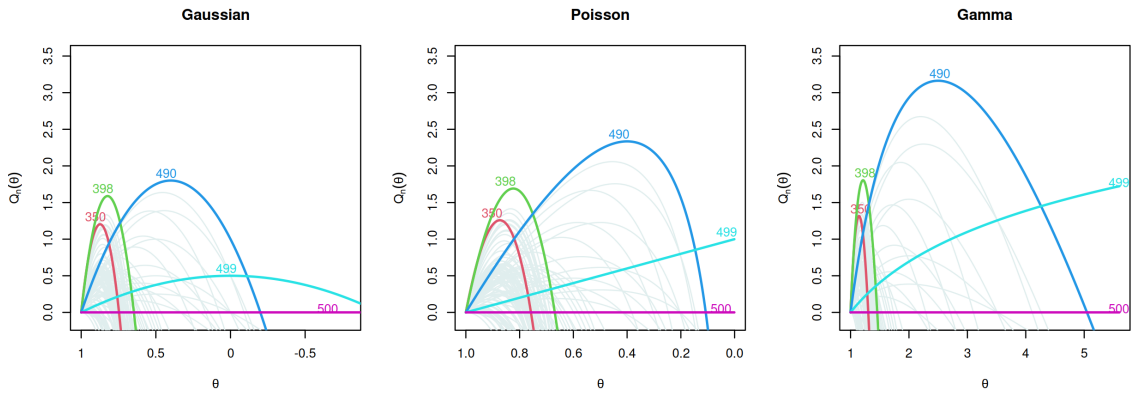


图 2：根据相同的实现值y_1，……，$y_{500} \sim$ Poi(1)计算得到的三种不同代价函数的比较。最左、中间和最右图分别展示了在假设损失函数为高斯、泊松或伽马损失时的代价函数 $Q_n(\theta)$。浮点数表示每条曲线被引入的时间步。灰色曲线表示不再最优、因此被剪枝的曲线。

推论3 如果某个指数族模型满足命题*1*的条件，那么在该模型下的剪枝将与针对高斯均值变化模型分析数据 $\gamma(x_t)$时 *FOCuS* 的剪枝完全相同。

例如，高斯方差变化模型的剪枝方式，与在数据 x_1^2，x_2^2，……上运行高斯均值变化模型时相同。

这一推论的一个结果是，关于原始 FOCuS 算法（Romano 等人，2021a）在时间 T 保留的曲线数量的强保证，同样适用于这些等价的指数族模型。FOCuS 所保留曲线数量的期望值相关结果对数据仅作出最低限度的假设，即观测值是可交换的。这些结果表明，迭代 T时保留的曲线数量的平均值为 $O(\log\ T)$。

4 未知的变化前参数

接下来，我们将探讨如何把该方法扩展到变化前参数和变化后参数均未知的情况。当 θ_0 未知时，对数似然比统计量 LR_T 满足

$$\begin{aligned}\frac{LR_T}{2} &= \max_{\theta_0,\theta_1,\tau}\left\{\sum_{t=1}^{\tau}\log f(x_t|\theta_0)+\sum_{t=\tau+1}^{T}\log f(x_t|\theta_1)\right\}\\ &\quad -\max_{\theta_0}\sum_{t=1}^{T}\log f(x_t|\theta_0).\end{aligned}$$

计算这一项的难点在于第一项。定义

$$Q_T^*(\theta_0,\theta_1)=\max_{\tau}\left\{\sum_{t=1}^{\tau}\log f(x_t|\theta_0)+\sum_{t=\tau+1}^{T}\log f(x_t|\theta_1)\right\}.$$

如果我们能计算出 θ_0 和 θ_1的这个函数，那么计算似然比统计量就很简单了。如果固定 θ_0 ，并将 Q_T^* 视为仅关于 θ_1 的函数，那么这就是我们在已知变化前参数时考虑过的函数$Q_T(\theta_1)$。

与之前一样，我们可以将 $Q_T^*(\theta_0, \theta_1)$ 写成一组曲线的最大值；现在这些曲线是关于两个变量 θ_0和 θ_1的，每个函数都对应于 τ的一个特定取值。与之前一样，如果我们能够轻松确定哪些取值对应的曲线会对最大值有所贡献，就可以删去其他函数，从而大幅加快 Q_T^*的计算。

为此，将 $Q_T^*(\theta_0, \theta_1)$ 视为仅关于 θ_1 的函数，并将其写作 $Q_{T,\theta_0}(\theta_1)$。算法1给出了在 $\theta_1 > \theta_0$ 时对该函数有贡献的曲线。这组曲线由各曲线根的顺序决定，即算法1中的 $i \geq 1$ 对应的 l_i 。现在，如果我们改变 θ_0，曲线的根会发生变化，但根据命题1，它们的顺序不会改变。唯一的区别在于 l_0 的定义。也就是说，随着我们减小 θ_0 ，由于 θ_1的取值范围变大，可能会有更多曲线对最大值有所贡献；但随着我们增大 θ_0 ，只能删去曲线。换言之，需要保留的曲线不会互换。因此，如果我们针对 $\theta_0 = -\infty$运行算法1，那么保留的曲线集合将正是对 $Q_T^*(\theta_0, \theta_1)$ 在 $\theta_1 > \theta_0$时有所贡献的曲线集合。

实际上，这意味着要在变化前参数未知的情况下实现 FOCuS 的剪枝，我们按照算法1进行，但在考虑变化 $\theta_1 > \theta_0$时设置 $l_0 = -\infty$ ，在考虑变化 $\theta_1 < \theta_0$时设置 $l_0 = \infty$ 。我们通过推论3证明了算法1在不同指数族模型之间的等价性，这一点也可立即得出。

5 自适应最大值检查

FOCuS 算法的主要计算成本来自每次迭代时对曲线求最大值。对于非高斯模型而言，这一点尤为明显，因为求一条曲线的最大值需要计算 $\max_{\theta_0, \theta_1} \ell(x_{1:T}|\theta_0, \theta_1, \tau)$，其中至少涉及一次对数运算（如泊松、二项分布和伽马数据的情形）。截至时间 T 所保留的曲线数量约为 $\log(T)$，因此计算所有最大值会带来一个随之（缓慢）增长的成本。不过，我们可以利用此前迭代中的信息来降低这一成本，从而只需对较少的曲线求最大值，就能判断 Q_T 是否高于或低于阈值。我们可以通过求得一个易于计算的Q_T 上界来做到这一点；如果该上界低于阈值，就无需计算Q_T。

下面的命题给出了所有曲线或部分曲线最大值的一个上界。首先，对于 $\tau_i < \tau_j$，我们定义信号在 τ_j 结束、在 τ_i 发生变化时的似然比统计量。该似然比统计量定义为

$$m_{\tau_i,\tau_j} = \max_{\substack{\theta_0 \in H_0, \\ \theta_1}} \ell(x_{1:\tau_j}|\theta_0, \theta_1, \tau_i) - \max_{\theta_0 \in H_0} \ell(x_{1:\tau_j}|\theta_0, \cdot, \tau_j),$$

其中， H_0 表示 θ_0 的可能取值集合。对于变化前参数已知情形， H_0 将只包含一个值；对于变化前参数未知情形， H_0 则为 $\mathbb{R}$ 。

命题4 对于任意 $\tau_1 < \tau_2 < \ldots < \tau_n < T$，有

$$\max_{i=1,\ldots,n} m_{\tau_i,T} \leq \sum_{i=1}^{n-1} m_{\tau_i,\tau_{i+1}} + m_{\tau_n,T}.$$

证明： 参见补充材料。图3也以图示方式展示了该结果。

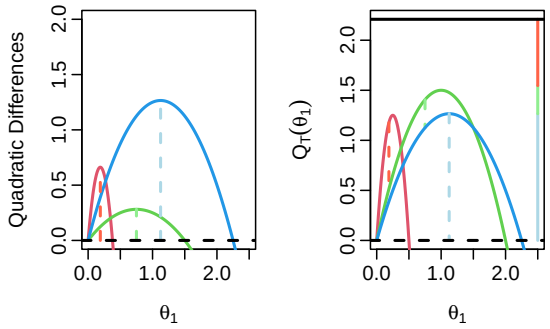


图3：变更前均值已知情形下命题1界限的示例。左图显示构成 $Q_T(\theta_1)$ 的三条曲线之间的差值。$m_{\tau_i:\tau_j}$ 值对应这些曲线的最大值（竖线）。右图显示 $Q_T(\theta)$、定义它的三条曲线，以及曲线之间的最大差值（竖直条）。该界限是最大差值之和（最右侧的堆叠线）。

算法2 在时间 T 对 $\theta_1 \geq \theta_0$ 执行 FOCuS 极大值检查。

输入：</style}一组 似然曲线及其对应的 (, M_{τ_k}) 值。

1: 设置 $k = n$2: **当** $k > 0$ **时执行**3: 计算$m_{\tau_k,}$
$_T$4: **如果** $m_{\tau_k,\ T} + M_{\tau_k} <$ 阈值 **则**5: **跳出**6: **否**
则7: **输出：** 返回在 $[\tau_k, T]$ 上发生变化 8: **结束**
如果9: $k \leftarrow k - 1$

10: **结束循环 输出：** 返回无
变化。

我们可以如下使用这一结果。和 $M_{\tau_k} := \sum_{i=1}^{k-1} m_{\tau_i,\tau_{i+1}}$ 可以作为 τ_k的似然曲线的一部分存储，最大值检查步骤则可按照算法2进行。其思路是，我们可以将 Q_T的上界限定为 $m_{\tau_k,T} + M_{\tau_k}$。因此，从 τ_k 值最大的曲线开始，我们检查 $m_{\tau_k,T} + M_{\tau_k}$ 是否低于阈值。如果是，我们就知道 Q_T 低于阈值，并且可以输出未检测到变化，而无需考虑其他曲线。如果不是，我们就检查 $m_{\tau_k,T}$，即在 τ_k 处发生变化的似然比检验统计量，是否高于阈值。如果是，我们就输出检测到变化。否则，我们转到 τ_k 值次大的曲线并重复上述过程。

实证结果表明，对于 $\tau_1 ... \tau_n \in \mathcal{I}_T$ 仅搜索向上变化（或类似地仅搜索向下变化）的情况，在底层数据没有变化的情形下，命题4中的上界相当紧，因为大多数$m_{\tau_i,\ \tau_{i+1}}$ 都非常小。此外，正如我们在第6节中所展示的，在大多数时间步，只需检查一条曲线，我们就能确定 Q_T小于我们的阈值。

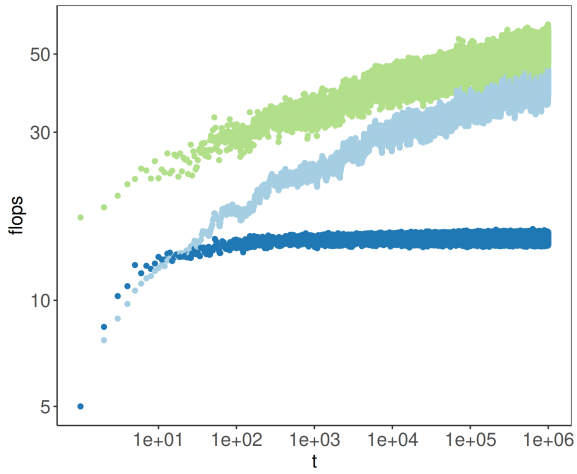


图4：三种 FOCuS 实现中每次迭代的浮点运算次数随时间的变化。绿色表示采用数值方法计算根 l_1、……、l_n 进行剪枝的 FOCuS 的浮点运算次数。浅蓝色表示采用平均值剪枝的 FOCuS。最后，蓝色表示采用平均值剪枝和自适应最大值检查的 FOCuS。两个坐标轴均为对数刻度。

6 数值示例

我们通过一些示例对 FOCuS 程序的计算复杂度进行实证评估，并将本文介绍的各种实现方式与文献中已有的实现方式进行比较。

图4展示了浮点运算次数随时间的变化。该图是对 50 个不同的、长度为 1×10^6 的序列所得结果取平均后绘制的。结果是在伯努利似然下得到的。在这种似然下，更新的开销可以忽略不计，因为更新只涉及整数运算，因此可以更好地比较剪枝和最大值检查的开销。我们比较了三种不同的 FOCuS 实现：(i) 采用有序根 l_1、……、l_n进行剪枝的 FOCuS，这些根通过牛顿-拉夫森过程以数值方法求得；(ii) 第 3 节介绍的采用平均值剪枝的 FOCuS；以及 (iii) 采用平均值剪枝和第 5 节介绍的自适应最大值检查的 FOCuS。

我们注意到，与牛顿-拉夫森方法相比，避免显式计算根可以降低计算开销。不过，加入自适应最大值检查程序后，性能最佳；在零假设下，我们发现每次迭代的计算成本恒定，约为每次迭代 15 次浮点运算。若不进行自适应最大值检查，最大化步骤就是 FOCuS 程序中计算量最大的步骤，因为每次迭代都需要评估 $\mathcal{O}(\log(T))$ 条曲线。

在图 5 中，我们将变化设置在时间点 1×10^5，并重点关注 FOCuS 存储的曲线数量，以及使用自适应最大值检查时需要评估的曲线数量。此外，为作比较，我们添加了一条曲线，表示直接计算 CUSUM 似然比检验的朴素开销。可以看到，在遇到变化之前，使用自适应最大值检查程序时，我们平均每次迭代只需对 1 条曲线求最大值，而标准 FOCuS 实现则约需对 7.4 条曲线求最大值。遇到变化之后，需要评估的曲线数量会增加，因为似然比统计量增大，满足命题4中的条件的可能性也随之提高

。从变化后出现的短暂尖峰可以看出，这种情况只会在检测到变化之前的一小段时间内发生。这一实证结果表明，FOCuS 的计算复杂度为每次迭代 $\mathcal{O}(1)$，而内存复杂度为 $\mathcal{O}(\log\ T)$，因为我们仍需在内存中平均存储 $\mathcal{O}(\log\ T)$ 条曲线。

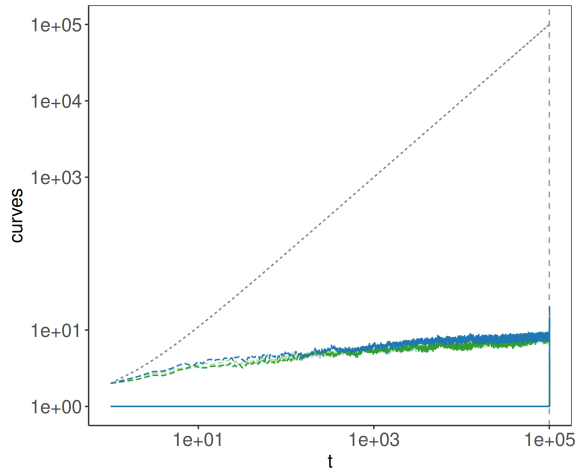


图 5：随时间变化的需存储曲线数量和每次迭代的评估次数。灰色点线表示计算 CUSUM 似然比检验的朴素成本。虚线表示 FOCuS 在某些高斯（浅绿色）、泊松（深绿色）、伯努利（浅蓝色）和伽马（深蓝色）实现值下存储的曲线数量。实线表示通过自适应最大值检查，在每次迭代中需要评估的曲线数量。两个坐标轴均采用对数刻度。

为说明针对正确的指数族模型运行 FOCuS 的优势，我们考虑在均值已知的高斯数据中检测方差变化。假设我们已对数据进行标准化，使其均值为零。检测方差变化的一种常见方法，是检测数据平方后的均值变化（Inclan 和 Tiao，1994），因此我们将比较应用于数据平方的高斯均值变化 FOCuS，与用于高斯方差变化模型的 FOCuS（如表 1 所示）。

对于零假设下服从均值为 0、方差为 $\theta_0 = 1$ 的正态分布的过程，我们针对 $\theta_1 = 0.75$、1。25、1。5、1。75 和 2 展示了 5 种模拟场景。每项实验包含 100 次重复。我们采用蒙特卡洛方法调整阈值，以按照 Chen 等人（2022 年，第 4.1 节）相同的方式，使零假设下的平均运行长度达到1×10^5。随后，我们在时间 1000 处引入变化，并通过检测延迟（检测时间与实际变化时间之差）来衡量性能。

在图 6 中，我们展示了这些情形，并以变化发生后 t 个观测值内的检测比例呈现结果。对于幅度足够大的正向变化，例如 $\theta_1 = 2$，与对数据平方后应用高斯均值变化模型相比，采用高斯方差变化模型的优势很小。然而，随着信噪比降低，变化变得更加细微，我们可以看到，使用正确的模型在缩短检测延迟方面的优势会逐渐增大。

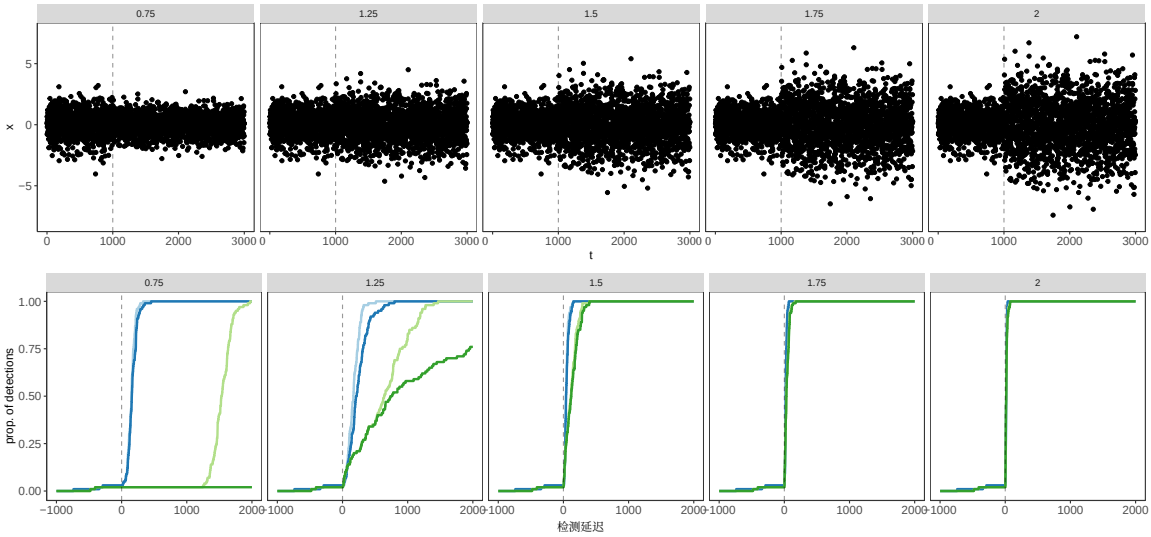


图 6：FOCuS 对高斯方差变化的实证评估。第一行：模拟场景中的示例序列，标签表示变化后参数 θ_1，垂直虚线表示变点位置 τ。第二行：对于变化前参数已知（浅蓝色）和未知（深蓝色）的高斯方差变化模型，以及将高斯均值变化应用于数据平方后、变化前参数已知（浅绿色）和未知（深绿色）的情形，检测比例随检测延迟变化的关系。这一次，垂直虚线表示变化的开始：变化发生后越快达到 1，效果越好。在垂直线之前，我们基本上是在统计假阳性。

7 讨论

我们提出了一种适用于单参数指数族模型的在线变点检测算法，该算法 (i) 在每次迭代时都能精确执行似然比检验；并且 (ii) 经实验证明，每次迭代的计算成本恒定。据我们所知，这是首个同时实现这两点的算法。

该算法只能检测单个参数的变化，因此只能分析单变量数据。不过，这可以作为分析多变量数据的基础。例如，Mei（2010）提出通过计算每条数据流的变化统计量，再将这些信息结合起来，实现对多条数据流的在线监测。关于如何高效地结合这类信息，已有大量文献探讨（例如 Cho 和 Fryzlewicz，2015；Enikeeva 和 Harchaoui，2019；Tickle 等，2021）。

另一项挑战是将该算法扩展到能够处理时间依赖数据。通常，即使数据存在自相关，只要提高检测变化的阈值，假设数据相互独立的方法也能很好地发挥作用（Lavielle 和 Moulines，2000）。如果自相关很强，这种简单的方法可能会损失一些检验功效；此时，对数据应用滤波器以消除自相关（Chakar 等人，2017），或改进 FOCuS 以对其建模（基于 Romano 等人，2021b；Cho 和 Fryzlewicz，2020；Hallgren 等人，2021 中的思路），或许会更好。

8 致谢

本研究获得了英国工程与自然科学研究委员会（EPSRC）资助，项目编号为 EP/N031938/1 和 EP/R004935/1；此外，BT 也通过“下一代融合数字基础设施”（NG-CDI）繁荣合作伙伴关系提供了支持。

参考文献

Andreou, E. and Ghysels, E.（2002）。检测金融市场波动动态中的多个结构突变。应用计量经济学期刊，17(5):579–600。Barrett, E., Bolz-Tereick, C. F., Killick, R., Mount, S., and Tratt, L.（2017）。虚拟机预热时冷时热。*ACM*编程语言论文集，1(OOPSLA):1–27。Beaulieu, C. and Killick, R.（2018）。区分气候数据中的趋势、变化与记忆效应。气候期刊，31(23):9519–9543。Chakar, S., Lebarbier, E., L´evy-Leduc, C., and Robin, S.（2017）。一种稳健的AR(1)过程均值变点估计方法。*Bernoulli*，23(2):1408–1447。Chen, Y., Wang, T., and Samworth, R. J.（2022）。高维、多尺度在线变点检测。英国皇家统计学会杂志（*B*辑），84:234–266。Chen, Z. and Tian, Z.（2010）。线性模型变点监测的修正程序。数学与计算机模拟，81(1):62–75。Cho, H. and Fryzlewicz, P.（2015）。通过稀疏化二分分割检测高维时间序列中的多个变点。英国皇家统计学会杂志：*B*辑（统计方法学），77(2):475–507。Cho, H. and Fryzlewicz, P.（2020）。序列依赖条件下的多变点检测：野性能量最大化与间隙施瓦茨准则。*arXiv*预印本 *arXiv:2011.13884*。Chu, C.-S. J., Hornik, K., and Kaun, C.-M.（1995）。参数恒定性的MOSUM检验。*Biometrika*，82(3):603–617。Eichinger, B. and Kirch, C.（2018）。一种用于估计多个随机变点的MOSUM程序。*Bernoulli*，24(1):526–564。Enikeeva, F. and Harchaoui, Z.（2019）。稀疏备择下的高维变点检测。统计年刊，47(4):2051–2079。Hallgren, K. L., Heard, N. A., and Adams, N. M.（2021）。非可交换数据中的变点检测。*arXiv*预印本 *arXiv:2111.05054*。Inclan, C. and Tiao, G. C.（1994）。利用平方累积和对方差变化进行回溯检测。美国统计学会会刊，89(427):913–923。Iwata, T., Nakamura, K., Tokusashi, Y., and Matsutani, H.（2018）。利用10 GbE FPGA网络接口卡加速在线变点检测算法。载于欧洲并行处理会议，第506–517页。Springer。Lavielle, M. and Moulines, E.（2000）。时间序列中未知数量变化的最小二乘估计。时间序列分析期刊，21(1):33–59。

Lucas, J. M.（1985）。计数数据累积和控制图。*Technometrics*，27(2):129–144。Mei, Y.（2010）。用于监测大量数据流的高效可扩展方案。*Biometrika*，97(2):419–433。Page, E.（1955）。针对未知时点发生的参数变化的检验。*Biometrika*，42(3/4):523–527。Page, E. S.（1954）。连续监控方案。*Biometrika*，41(1/2):100–115。Romano, G., Eckley, I., and Fearnhead, P.（2023）。一种基于函数剪枝的对数线性非参数在线变点检测算法。*arXiv*预印本 *arXiv:2302.02718*。Romano, G., Eckley, I., Fearnhead, P., and Rigaill, G.（2021a）。通过函数剪枝累积和统计量快速进行在线变点检测。arXiv.2110.08205。Romano, G., Rigaill, G., Runge, V., and Fearnhead, P.（2021b）。局部波动和自相关噪声存在时的突变检测。美国统计学会会刊，第1–16页。Ross, G. J. and Adams, N. M.（2012）。用于检测任意分布变化的两种非参数控制图。质量技术期刊，44(2):102–116。Ross, G. J., Tasoulis, D. K., and Adams, N. M.（2011）。对数据流进行非参数监测，以检测位置和尺度的变化。*Technometrics*，53(4):379–389。Sakamoto, Y., Fukui, K.-I., Gama, J., Nicklas, D., Moriyama, K., and Numao, M.（2015）。通过统计变化检测方法结合聚类检测概念漂移。载于*2015*年第七届国际知识与系统工程会议，第37–42页。IEEE。Tickle, S. O., Eckley, I., and Fearnhead, P.（2021）。一种计算高效的高维多变点检测程序及其在全球恐怖主义事件发生率中的应用。英国皇家统计学会杂志：*A*辑（社会统计学）。Varghese, B., Wang, N., Barbhuiya, S., Kilpatrick, P., and Nikolopoulos, D. S.（2016）。边缘计算中的挑战与机遇。载于*2016*年*IEEE*智能云国际会议（*SmartCloud*），第20–26页。Ward, K., Dilillo, G., Eckley, I., and Fearnhead, P.（2022）。Poisson-FOCuS：一种用于检测计数突发的高效在线方法及其在伽马射线暴检测中的应用。arXiv.2208.01494。Yu, Y., Padilla, O. H. M., Wang, D., and Rinaldo, A.（2020）。关于在线变点检测的简述。*arXiv*预印本 *arXiv:2006.03283*。

附录

A 证明

A.1 推导指数族似然比

对于如下形式的指数族模型

$$f(x \mid \theta) = \exp\big[\, \alpha(\theta) \cdot \gamma(x) - \beta(\theta) + \delta(x) \,\big],$$

我们的似然差具有如下形式

$$\begin{aligned}\ell(x_{1:T}|\theta_0, \theta_1, \tau_i) - \ell(x_{1:T}|\theta_0, \theta_1, \tau_j) = \\ [\alpha(\theta_1) - \alpha(\theta_0)] \sum_{t=\tau_i+1}^{\tau_j} \gamma(x_t) - [\beta(\theta_1) - \beta(\theta_0)](\tau_j - \tau_i).\end{aligned}$$

我们有

$$\ell(x_{1:T}|\theta_0, \theta_1, \tau) := \sum_{t=1}^{\tau} \log f(x_t|\theta_0) + \sum_{t=\tau+1}^{T} \log f(x_t|\theta_1).$$

因此，我们有

$$\begin{aligned}\ell(x_{1:T}|\theta_0, \theta_1, \tau_i) - \ell(x_{1:T}|\theta_0, \theta_1, \tau_j) = \\ = \sum_{t=\tau_i+1}^{\tau_j} \{\log f(x_t|\theta_1) - \log f(x_t|\theta_0)\}.\end{aligned}$$

代入

$$\begin{aligned}&\log f(x_t|\theta_1) - \log f(x_t|\theta_0) = \\ &= [\alpha(\theta_1) \cdot \gamma(x_t) - \beta(\theta_1) + \delta(x_t)] \\ &\quad - [\alpha(\theta_0) \cdot \gamma(x_t) - \beta(\theta_0) + \delta(x_t)] \\ &= [\alpha(\theta_1) - \alpha(\theta_0)]\gamma(x_t) - [\beta(\theta_1) - \beta(\theta_0)].\end{aligned}$$

即可得到所需结果。

A.2 根据 $\bar{\gamma}$ 值确定根的顺序

定义

$$\bar{\gamma}_{\tau_i:\tau_j} = \frac{1}{\tau_j - \tau_i} \sum_{t=\tau_i+1}^{\tau_j} \gamma(x_t)$$

作为 $t = \tau_i + 1, \ldots, \tau_j$ 的 $\gamma(x_t)$ 平均值，并将 $\theta_1^{\tau}(\neq \theta_0)$ 定义为下式的根

$$\ell(x_{1:T}|\theta_0, \theta_1^{\tau}, \tau) - \ell(x_{1:T}|\theta_0, \cdot, T) = 0.$$

命题 5 假设对于我们选择的 θ_0 ，函数

$$\theta_1 :\to \frac{\beta(\theta_1)-\beta(\theta_0)}{\alpha(\theta_1)-\alpha(\theta_0)}$$

严格递增。因此，$\bar{\gamma}_{\tau_i:\ \tau_j}-\bar{\gamma}_{\tau_j:\ T}$ 的符号与 $\theta_1^{\tau_i}-\theta_1^{\tau_j}$ 的符号相同。

我们有

$$[\alpha(\theta_1^{\tau})-\alpha(\theta_0)]\sum_{t=\tau+1}^{T}\gamma(x_t)-[\beta(\theta_1^{\tau})-\beta(\theta_0)](T-\tau)=0.$$

将其重新整理，得到

$$\frac{\beta(\theta_1^{\tau})-\beta(\theta_0)}{\alpha(\theta_1^{\tau})-\alpha(\theta_0)}=\bar{\gamma}_{\tau:T}.$$

根据单调性，我们有，θ_1^{τ} 是 $\bar{\gamma}_{\tau:T}$的递增函数。对于 $\tau_i<\tau_j<T$，我们还有

$$\bar{\gamma}_{\tau_i:T}=\frac{T-\tau_j}{T-\tau_i}\bar{\gamma}_{\tau_j:T}+\frac{\tau_j-\tau_i}{T-\tau_i}\bar{\gamma}_{\tau_i:\tau_j},$$

因此，$\bar{\gamma}_{\tau_i:\ \tau_j}-\bar{\gamma}_{\tau_j:\ T}$ 的符号与$\bar{\gamma}_{\tau_i:\ T}-\bar{\gamma}_{\tau_j:\ T}$ 的符号相同，因为$\bar{\gamma}_{\tau_i:\ T}$ 是$\bar{\gamma}_{\tau_i:\ \tau_j}$ 和$\bar{\gamma}_{\tau_j:\ T}$的凸组合。综合以上结果即可得证。

A.3 最大值检查界

定义

$$m_{\tau_i,\tau_j}=\max_{\substack{\theta_0\in H_0,\\ \theta_1}}\ell(x_{1:\tau_j}|\theta_0,\theta_1,\tau_i)-\max_{\theta_0\in H_0}\ell(x_{1:\tau_j}|\theta_0,\cdot,\tau_j),$$

其中，H_0 表示 θ_0的所有可能取值。 H_0 在变化前参数已知的情形下只包含一个值；在变化前参数未知的情形下则为 $\mathbb{R}$ 。

命题 6 对于任意 $\tau_1<\tau_2<...<\tau_n<T$，我们有

$$\max_{i=1,...,n}m_{\tau_i,T}\leq\sum_{i=1}^{n-1}m_{\tau_i,\tau_{i+1}}+m_{\tau_n,T}.$$

将 $\theta\in H_0$中$\sum_{t=1}^{\tau_i}\log\ f(x_t|\theta_0)$的最大化点记为$\hat{}\ \theta_0^{\tau_i}$。（注意，在变化前均值$\hat{}$已知的情形下，我们总是有 $\theta_0^{\tau_i}=\theta_{0\circ}$ ）

现在，考虑其形式

$$\begin{aligned}m_{\tau_i,\tau_j}=&\sum_{t=1}^{\tau_i}\log f(x_t|\hat{\theta}_0^{\tau_i})+\max_{\theta_1}\sum_{t=\tau_i+1}^{\tau_j}\log f(x_t|\theta_1)\\&-\sum_{t=1}^{\tau_j}\log f(x_t|\hat{\theta}_0^{\tau_j}).\end{aligned}$$

注意第一项和第三项的相似性，这使得在对 $m_{\tau_i,\tau_{i+1}}$ 求和时可以进行望远镜式消去。为方便起见，令 $\tau_{n+1} := T$，则对于任意 $1 \le k \le n$，

$$\begin{aligned}
&\sum_{i=1}^{n-1} m_{\tau_i,\tau_{i+1}} + m_{\tau_n,T} = \\
&= \left[\sum_{t=1}^{\tau_1} \log f(x_t|\hat{\theta}_0^{\tau_1}) + \sum_{i=1}^{k-1} \max_{\theta_1} \sum_{t=\tau_i+1}^{\tau_{i+1}} \log f(x_t|\theta_1)\right] \\
&\quad + \left[\sum_{i=k}^{n} \max_{\theta_1} \sum_{t=\tau_i+1}^{\tau_{i+1}} \log f(x_t|\theta_1)\right] - \sum_{t=1}^{T} \log f(x_t|\hat{\theta}_0^T).
\end{aligned}$$

我们可以将其与

$$\begin{aligned}
m_{\tau_k,T} = &\sum_{t=1}^{\tau_k} \log f(x_t|\hat{\theta}_0^{\tau_1}) + \max_{\theta_1} \sum_{t=\tau_k+1}^{T} \log f(x_t|\theta_1) \\
&- \sum_{t=1}^{T} \log f(x_t|\hat{\theta}_0^T).
\end{aligned}$$

注意到，由于在扩展后的假设集上对同一似然进行最大化，前两项满足不等式，而最后一项则相等。由此即可证明该结论。

构造 $\sum_{i=1}^{n-1} m_{\tau_i,\ \tau_{i+1}} + m_{\tau_n,\ T}$ 本质上是在每一个 τ_i处都拟合一个变点。我们将其与构造 $\max_{i=1,\ \ldots,\ n} m_{\tau_i,\ T}$进行比较，后者只在最有希望的 τ_i处拟合一个变点。

其中 $\{\tau_1, ..., \tau_n\} \in \mathcal{I}_T$和 因此按 $\bar{\gamma}_{\tau_i:\ \tau_{i+1}}$ 递增/递减排序，分别表示向上变化/向下变化。与只拟合最佳的一个相比，将所有 τ_i都作为变点进行拟合并不会带来太大收益。在没有变化的底层数据情形下，较早的 $m_{\tau_i,\ \tau_{i+1}}$ 会非常小，而 $m_{\tau_n,\ T}$ 的贡献最大，因为它捕捉了信号中近期事件的波动。