arXiv:2302.00235v1 [math.ST] 1 Feb 2023

最小化变点估计误差

作者：ChanHock Peng，新加坡国立大学

摘要

本文研究多个序列中的变点，目标是利用其他序列的信息，最小化单个序列的估计误差。这与近期对多个序列中变点的研究有所不同，后者关注的是检测共同变点。我们首先研究具有恒定变点强度的单个序列这一典型情形。我们考察变点算法的两项指标：第一是无误差地估计变点的概率；第二是真实变点与估计变点之间的期望距离。我们给出无误差概率的理论上界和期望距离的下界，所有算法都必须满足这些界限。我们提出一种扫描-CUSUM算法，该算法达到无误差概率的上界，并接近期望距离的下界。接下来，我们研究强度非恒定的情形，并确定估计误差趋于零的精确条件。我们针对非恒定强度函数提出扫描-CUSUM算法的扩展，并证明该算法在零误差区域的边界处渐近实现零误差。我们通过一个应用示例展示扫描-CUSUM算法如何用于共享未知非恒定强度函数的多个序列。我们利用所有序列的变点剖面似然估计强度函数，并在估计得到的强度函数上应用扫描-CUSUM算法。

1 引言

本文采用贝叶斯框架对变点进行建模，并研究变点估计。有关变点估计贝叶斯模型的代表性论文包括 Yao（1984）、Barry and Hartigan（1993）、Chib（1998）、Lai and Xing（2011）、Ko、Chong and Ghosh（2015）以及 Du、Kao and Kou（2016）。Yao（1984）针对正态观测下的变点提出了隐马尔可夫模型（HMM），并展示了如何使用前向—后向公式计算后验变点概率。Lai and Xing（2011）将 HMM 方法扩展到多参数指数族，并提出有界复杂度混合模型以限制计算复杂度。Ko、Chong and Ghosh（2015）使用贝塔先验估计变点强度，而不是假设

这些强度已知。Du、Kao and Kou（2016）表明，可以利用 HMM 的边际似然一致地估计变点。

近期变点检测领域的一个研究热点，是探讨这样的可检测性问题：数据由多个序列组成，且变点共同出现在其中一小部分稀疏序列中。Zhang、Siegmund、Ji 和 Li（2010）受到多个拷贝数变异 DNA序列中存在共同变点的启发，提出了利用卡方统计量之和汇聚信号的扫描与分割算法。Mei（2010）提出了用于多个数据流中变点在线检测的 CUSUM 检验统计量之和，并证明当发生分布变化的序列比例被正确指定时，该方法能够实现最优检测。Xie 和 Siegmund（2013）提出了一种广义似然检验，该检验对于发生变化的序列比例具有最优性和自适应性。

Cho 和 Fryzlewicz（2015）利用野二分割（WBS）研究了高维时间序列中的变点检测，并对各个序列的 CUSUM 得分进行阈值化处理。Jeng、Cai 和 Li（2010）刻画了在何种情形下，可以汇聚来自少量序列的弱信号以实现成功检测，并表明高阶批判检验统计量能够在这类情形的边界处实现检测。Chan 和 Walther（2015）强调了序列信号比如何影响能够实现检测的稀疏性程度。针对稀疏变点的最优检测理论，包括时间序列中的稀疏变点以及具有空间依赖的情形，已由 Horv´áth and Huskova（2014）、Jirak（2015）、Cho（2016）、Wang 和 Samworth（2018）、Enikeeva 和 Harchaoui（2019）、Pilliart、Carpentier 和 Verzelen（2020）以及 Liu、Gao 和 Samworth（2021）建立。

1.1 本文的贡献

我们的研究涉及多个序列，但与早期研究不同，我们着重于利用其他序列中的信息来尽可能降低估计误差。该方法的关键在于提出共同变点强度函数，并利用变点剖面似然对该函数进行估计。

我们在恒定强度假设下，给出了变点算法无误差概率的上界，并提出了一种达到该上界的扫描-CUSUM算法。我们还给出了真实变点与估计变点之间期望距离的下界，并证明我们的算法所得期望距离在数值上接近该下界。对于非恒定强度函数，我们确立了

能够实现渐近零估计误差的精确条件，并提出了扫描-CUSUM算法的扩展版本，使其在零误差区域的边界处达到渐近零误差。扫描-CUSUM算法为利用序列间的信息共享来改进多个序列中变点的估计奠定了基础。

1.2 论文结构

第2节介绍变点模型，并在变点强度为常数时，给出无误差估计变点的概率上界，以及真实变点与估计变点之间期望距离的下界。第3节提出扫描-CUSUM算法，该算法达到无误差概率上界，且数值结果接近距离下界。算法首先利用扫描统计量定位变点，再应用 WBS 的 CUSUM检验统计量，最终确定变点估计。第4节将讨论扩展到非恒定强度，并证明在零误差区域之外，渐近零估计误差是不可能的。我们将扫描-CUSUM算法扩展到处理随位置变化的变点强度，并证明该算法在零误差区域内可实现渐近零误差。第5节展示如何应用扫描-CUSUM算法，在多个序列的变点由共同强度函数生成时降低估计误差。第6节通过实验比较逐个序列估计变点与利用所有可用信息估计变点的效果。第7节证明理论结果。附录提供额外的技术论证和计算。

1.3 记号

令 $\#A$ 为集合 A 中的元素个数，令 $\emptyset$ 为空集。令 $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$ 为密度函数，$\Phi(z) = \int_{-\infty}^{z} \phi(x)dx$ 为标准正态分布的分布函数。令 $\lfloor \cdot \rfloor$ 为下取整函数，$\lceil \cdot \rceil$ 为上取整函数。对于数 t 和集合 A，令 $d(t, A) = \min_{u \in A} |u - t|$。令 $\mathbb{1}$ 为示性函数，$\mathbb{Z}$ 为全体整数构成的集合。令 $a = o(b)$ if $\frac{a}{b} \to 0$。

2 变点模型与估计误差界

令 $\mathbf{X} = (X(1) \ldots, X(T))$ 为一个序列的观测值，令 $\mu(t)$ 为 $X(t)$ 的均值。

Change-point model。假设存在随机变量 $\Delta(t)$ 和独立的伯努利随机变量 $Y(t)$，使得

$$\mu(t+1) = \mu(t) + Y(t)\Delta(t). \tag{2.1}$$

我们假设，对于所有 t，$\Delta(t)|Y(t) = 1$ 都服从共同的分布 F，且该分布在零点没有点质量。变点为 $\boldsymbol{\tau} = \{t : Y(t) = 1\}$。令 $a(t) = EY(t)$ 表示 t 处的变点强度。本节中我们考虑

$$a(1) = \cdots = a(T-1) = q \text{ for some } q > 0. \tag{2.2}$$

我们假设，在给定 $\mu(t)$ 的条件下，观测值 $X(t)$ 相互独立，且

$$X(t) = \mu(t) + \epsilon(t), \quad \epsilon(t) \sim \mathrm{N}(0, \sigma_X^2). \tag{2.3}$$

Example。Yao（1984）考虑了一个隐马尔可夫模型，其中

$$\mu(t+1) = (1 - Y(t))\mu(t) + Y(t)\xi(t), \tag{2.4}$$

其中 $\mu(1)$ $\xi(1) \ldots, \xi(T-1)$ 服从 $\mathrm{N}(0, \sigma_\xi^2)$，且与所有 $Y(t)$ 相互独立。在给定 $Y(t) = 1$ 的条件下，$\Delta(t) = \xi(t) - \mu(t)$ 服从 $\mathrm{N}(0, 2\sigma_\xi^2)$。但请注意，$\Delta(1) \ldots, \Delta(T-1)$ 并不独立。Yao（1984）证明，在满足 (2.3) 和 (2.4) 的条件下，可以利用前向后向公式高效地计算$P(Y(t) = 1|\mathbf{X})$。

~Remarks。令 $\hat{\boldsymbol{\tau}} = \{t : Y(t) = 1\}$ 为变点算法估计得到的变点。假设变点模型能够高效地计算 $P(Y(t) = 1|\mathbf{X})$，我们通过令 $\hat{\tau} = \{t : P(Y(t) = 1|\mathbf{X}) \geq \frac{1}{2}\}$，来最大化 $\sum_{t=1}^{T-1} P(\widehat{Y}(t) = Y(t))$。然而，这个估计量过于保守，因为与完全不估计变点相比，如果估计的变点位置有误，我们会受到双重惩罚。此外，它没有考虑估计的变点与真实变点之间的距离。

Notations。

1. 令 $J = \#\boldsymbol{\tau}$ 为变点的数量。将 $\boldsymbol{\tau}$ 中的变点排列如下

$$\tau_1 < \cdots < \tau_J.$$

为完整起见，定义 $\tau_0 = 0$ 和 $\tau_{J+1} = T$。

^2. 设 $J = \#\widehat{\boldsymbol{\tau}}$ 为算法估计出的变点数量。

在下文对 α_T、β_T 和 $\gamma_T(\Delta_0)$ 的定义中，τ指在以 $J > 0$ 为条件时，从 J 个变点中随机选择一个。

DEFINITIONS。

1. 对每个 j，令

$$\kappa(\tau_j) = \#\{\widehat{\tau} \in \widehat{\boldsymbol{\tau}} : |\tau_j - \widehat{\tau}| < \tfrac{1}{2}\min(\tau_{j+1} - \tau_j, \tau_j - \tau_{j-1})\}. \qquad (2.5)$$

2. 令 $\alpha_T = P(\kappa(\tau) \neq 1 | J \geq 1)$。

3. 无误差地估计变点的概率为

$$\beta_T = P(\widehat{Y}(\tau) = 1 \text{ and } \kappa(\tau) = 1 | J \geq 1).$$

4. 令 $\Delta_0 > 0$。期望位置误差为

$$\gamma_T(\Delta_0) = E(d(\tau, \widehat{\boldsymbol{\tau}}) | J \geq 1, \kappa(\tau) = 1, |\Delta(\tau)| \geq \Delta_0).$$

REMARKS。

1. 在 (2.5) 中，我们将每个 $\widehat{\tau}$ 至多分配给一个 τ_j，而 $\kappa(\tau_j)$ 是分配给 τ_j 的 $\widehat{\tau}$ 的数量。我们在评估变点算法时纳入 $\kappa(\tau) = 1$，包括在 β_T 和 $\gamma_T(\Delta_0)$ 中，以避免因估计出过多变点，或不估计具有弱信号的变点而造成虚假的改进。

2. 在 $\gamma_T(\Delta_0)$ 的定义中，我们以 $|\Delta(\tau)| \geq \Delta_0$ 为条件，因为 $d(\tau, \widehat{\boldsymbol{\tau}})$ 的量级为 $[\Delta(\tau)]^{-2}$，因此若没有这一限制，预期的 L_1 位置误差将是无穷大。我们在定理 1 中证明，在 (2.1)–(2.3) 条件下，任何满足约束 $\alpha_T \to 0$ 且以 $T \to \infty$ 为条件的算法都满足

$$\liminf_{T\to\infty} \gamma_T(\Delta_0) \geq \gamma_{\text{lower}}(\tfrac{\Delta_0}{\sigma_X})$$

对于某个正的 $\gamma_{\text{lower}}(\frac{\Delta_0}{\sigma_X})$，我们使用以下术语对其进行刻画。

DEFINITIONS。

1. 对于非负函数 $\mathbf{r} = (r(i) : i \in \mathbf{Z})$，若 $s = \sum_{i=-\infty}^{\infty} r(i)$ 为正且有限，则 $\mathrm{med}(\mathbf{r})$ 是满足以下条件的整数 u

$$\sum_{i=-\infty}^{u-1} r(i) < \tfrac{s}{2} \le \sum_{i=-\infty}^{u} r(i),$$

且 $\mathrm{mode}(\mathbf{r})$ 是满足 $r(u) = \max_{i\in\mathbf{Z}} r(i)$ 的最小整数 u。

2. 设 $Z^+(u)$ 和 $Z^-(u)$ 为相互独立的标准正态随机变量，且 $\Delta \sim F$ 与 $Z^+(u)$ 和 $Z^-(u)$ 相互独立。设 $S^+(i) = \sum_{u=1}^{i} Z^+(u)$，$S^-(i) = \sum_{u=1}^{i} Z^-(u)$，并令 $\mathbf{p} = (p(i) : i \in \mathbf{Z})$，使得对于 $i > 0$,

$$\begin{aligned} p(i) &= \exp(\Delta S^+(i) - \tfrac{i\Delta^2}{2}), \\ p(-i) &= \exp(\Delta S^-(i) - \tfrac{i\Delta^2}{2}), \\ p(0) &= 1. \end{aligned} \tag{2.6}$$

定义

$$\gamma_{\mathrm{lower}}(\Delta_0) = E\Big(|\mathrm{med}(\mathbf{p})| \Big| |\Delta| \ge \Delta_0\Big). \tag{2.7}$$

第7节中的引理1证明了，对于所有 $\Delta_0 > 0$，$\gamma_{\mathrm{lower}}(\Delta_0) < \infty$。

3. 设 $\Delta \sim F$。定义

$$\beta_{\mathrm{upper}} = \tfrac{1}{2} E(\Delta^2 \nu(\Delta)), \tag{2.8}$$

其中 $\nu(\Delta) = 2\Delta^{-2}\exp(-2\sum_{i=1}^{\infty} i^{-1}\Phi(-\frac{\sqrt{i}}{2}|\Delta|))$ 是超越函数，参见 Siegmund (1985) 中的 (4.37)。超越函数的上界为 1，且 $\nu(\Delta) \to 1$ as $\Delta \to 0$。

定理 1. 考虑 $q \to 0$as $T \to \infty$。在 (2.1)–(2.3) 条件下，所有变点算法均满足

$$\limsup_{T\to\infty} \beta_T \le \beta_{\mathrm{upper}}.$$

设 $\Delta_0 > 0$。若某个变点算法满足 $\alpha_T \to 0$，则

$$\liminf_{T\to\infty} \gamma_T(\Delta_0) \ge \gamma_{\mathrm{lower}}(\tfrac{\Delta_0}{\sigma_X}).$$

3 常强度函数的最优性

当变点强度 $a(t)$ 为常数时，扫描-CUSUM算法可达到 β_{upper}。该算法扫描观测值窗口以定位变点，并在每个标出的窗口内应用 CUSUM得分来估计变点的精确位置。

Niu和Zhang（2012）的筛选与排序算法采用扫描统计量来定位变点。Fryzlewicz（2014）的WBS算法采用CUSUM统计量来估计变点。scan-CUSUM中使用多个窗口长度的扫描统计量，其灵感来自Arias-Castro、Donoho和Huo（2005，2006）的多尺度方法。

DEFINITIONS。令 $S(t) = \sum_{u=1}^{t} X(u)$。

1. 扫描统计量。对于给定的窗口长度 2ℓ，定义

$$Z_\ell(t) = \sqrt{\tfrac{1}{2\ell\sigma_X^2}}[S(t+\ell) + S(t-\ell) - 2S(t)],$$

对于 $\ell \le t \le T - \ell$。

2. CUSUM统计量。对于给定区间 $I = (u, v)$，定义

$$Z_I(t) = \sqrt{\tfrac{(v-t)(t-u)}{(v-u)\sigma_X^2}}\left(\tfrac{S(v)-S(t)}{v-t} - \tfrac{S(t)-S(u)}{t-u}\right), \tag{3.1}$$

对于 $u < t < v$。

3. 剖面似然。对于给定的区间 $I = (u, v)$，定义 $\mathrm{L}(I) = (L_I(t) : t \in \mathrm{Z})$，其中

$$L_I(t) = \exp(\tfrac{1}{2}Z_I^2(t)) \text{ for } t \in I, \tag{3.2}$$

且 $L_I(t) = 0$ 对于 $t \notin I$。验证对于 $t \in I$，

$$L_I(t) = C_I\Big[\sup_{\mu}\prod_{r=u+1}^{t}\phi(\tfrac{X_r-\mu}{\sigma_X})\Big]\Big[\sup_{\nu}\prod_{r=t+1}^{v}\phi(\tfrac{X_r-\nu}{\sigma_X})\Big],$$

where $C_I = (2\pi)^{\frac{v-u}{2}}\exp[\tfrac{1}{2\sigma_X^2}\sum_{r=u+1}^{v}(X_r-\bar{X}_I)^2]$, $\bar{X}_I = \tfrac{1}{v-u}\sum_{r=u+1}^{v}X_r$.

REMARKS。扫描-CUSUM算法将剖面似然视为真实似然；一旦确定变点位于区间 I 内，就选择 $t = \widehat{\tau}$，使$P(\tau = t|\tau \in I)$最大化。当

变点强度为常数时，这等价于选择 t，使剖面似然 $L_I(t)$ 最大化。不过，无需提及剖面似然，因为我们只是在选择 t，使 $|Z_I(t)|$ 最大化。

在以下情况下需要使用剖面似然函数。

1. 当强度函数为常数，且目标是最小化 τ 与 $\hat{\tau}$ 之间的期望距离时。在此情况下，最优估计量选择 $\hat{\tau}$ 为剖面似然的中位数。然而，由于扫描-CUSUM 的期望距离接近通用下界，本文不再对此进行探讨。

2. 当强度函数不为常数时，通过选择 $t = \hat{\tau}$ 使 $a(t)L_I(t)$ 最大化，即可实现 $P(\tau = t|\tau \in I)$ 的最大化。我们将在第 4 节对此进行研究。

ALGORITHM。令 $1 < \rho \leq 2$ 并定义 $\ell_b = \lceil \rho^b \rceil$。将 $Z_b(t)$ 写作 $Z_{\ell_b}(t)$。令 b_T 为满足 $2\ell_b \leq T-1$ 的最大 b。

算法 1 扫描累积和

输入：$c_{\text{scan}} > 0$

初始化：$\widehat{Y}(t) = 0$ 对于 $1 \leq t \leq T-1$

对于 $b = 0, \ldots, b_T$:

$A = \{t : \ell_b \leq t \leq T - \ell_b, \sum_{u=t-\ell_b+1}^{t+\ell_b-1} \widehat{Y}(u) = 0\}$

当 $(\max_{t\in A} |Z_b(t)| \geq c_{\text{scan}})$

$t^* =$使$_{t\in A}|Z_b(t)|$取得最大值

$I = (t^* - \ell_b, t^* + \ell_b)$

$\hat{\tau} =$使$_{t\in I}|Z_I(t)|$取得最大值

更新 $\widehat{Y}(\hat{\tau}) = 1$

根据变化更新 A $\widehat{Y}(\hat{\tau})$

输出：$\hat{\boldsymbol{\tau}} = \{t : \widehat{Y}(t) = 1\}$

定理 2. 将 $q = o(1/\log T)$ 视为 $T \to \infty$。在满足 (2.1)–(2.3) 的条件下，扫描-CUSUM算法采用 $c_{\text{scan}} = \sqrt{2\log(T/\log T)}$，满足 $\alpha_T \to 0$ 且

$$\lim_{T\to\infty} \beta_T = \beta_{\text{upper}}. \tag{3.3}$$

REMARKS。

1. 阈值 $c_{\text{scan}} = \sqrt{2\log(T\log T)}$ 是一个理论值，用于突出 scan-CUSUM 的渐近行为。在实践中，我们可能希望设定阈值以满足其他标准，例如满足第一类错误概率约束。

2. 令 p 如 (2.6) 中所定义，并令 $\Delta_0 > 0$。根据用于证明 (3.3) 的计算，详见第7.2.3节，

$$\lim_{T\to\infty} \gamma_T(\Delta_0) = \gamma_{\text{scan}}(\tfrac{\Delta_0}{\sigma_X}), \tag{3.4}$$

其中

$$\gamma_{\text{scan}}(\Delta_0) = E\Big(|\text{mode}(\mathbf{p})|\Big||\Delta| \geq \Delta_0\Big).$$

然而，下面的数值研究表明，尽管 $\gamma_{\text{scan}}(\Delta_0) > \gamma_{\text{lower}}(\Delta_0)$，但 $\gamma_{\text{scan}}(\Delta_0)$ 与 $\gamma_{\text{lower}}(\Delta_0)$ 之间的差异很小，因此 scan-CUSUM 在最小化 L_1 位置误差方面表现良好。

NUMERICAL COMPARISON。表示为

$$\begin{aligned}
\gamma_{\text{lower}}(\Delta_0) &= [1 - F(\Delta_0) + F(-\Delta_0)]^{-1} \int_{|\Delta|\geq\Delta_0} g_{\text{lower}}(\Delta)F(d\Delta), \\
\gamma_{\text{scan}}(\Delta_0) &= [1 - F(\Delta_0) + F(-\Delta_0)]^{-1} \int_{|\Delta|\geq\Delta_0} g_{\text{scan}}(\Delta)F(d\Delta),
\end{aligned}$$

其中

$$g_{\text{lower}}(\Delta) = E\Big(|\text{med}(\mathbf{p})|\Big|\Delta\Big) \text{ and } g_{\text{scan}}(\Delta) = E\Big(|\text{mode}(\mathbf{p})|\Big|\Delta\Big).$$

表1基于生成的10,000个正态序列显示，$g_{\text{scan}}(\Delta)$比 $g_{\text{lower}}(\Delta)$大约高7–9%。因此，扫描-CUSUM算法的 L_1 定位误差接近下界。

Δ	$\Delta^2 g_{\text{lower}}(\Delta)$	$\Delta^2 g_{\text{scan}}(\Delta)$
1.0	2.71±0.04	2.91±0.04
0.3	2.75±0.04	2.95±0.04
0.1	2.78±0.04	2.98±0.04
0.03	2.82±0.04	3.06±0.04

表1： $g_{\text{lower}}(\Delta)$与 $g_{\text{scan}}(\Delta)$的比较。

4 非恒定强度函数的最优性

Change-Point Intensity Model。假设存在一个分布 G_q，其均值为 q，且支撑集为 $[0,1],$，使得

$$a(t) \sim_{\text{i.i.d.}} G_q \text{ for } 1 \le t \le T-1. \tag{4.1}$$

在给定 $\{a(t)\}_{t=1}^{T-1}$ 的条件下，$Y(t)$ 是独立的伯努利($a(t)$) 随机变量。

Remarks。定理 3 给出了在何种条件下，即使已知强度函数并据此应用变点算法，也无法实现渐近为零的估计误差。这些条件是紧的，因为当这些条件不成立时，我们能够实现渐近为零的估计误差。

定理 3。假设 (2.1)、(2.3) 和 (4.1) 成立，且 $q \to 0$ 为 $T \to \infty$。若

$$\limsup_{q\to 0} \tfrac{1}{q} \int_0^1 [1 - G_q(x)]^2 dx > 0, \tag{4.2}$$

则任何变点算法都无法达到 $\beta_T \to 1$。

Examples。

1. 如果 $G_q(yq) = G_1(y)$，且 G_1 为非负且有界、均值为 1，则通过变量替换 $y = x/q$,

$$\tfrac{1}{q} \int_0^1 [1 - G_q(x)]^2 dx = \int_0^{1/q} [1 - G_1(y)]^2 dy$$

并且（4.2）成立。

2. 如果 G_q 是贝塔分布（$\frac{q}{1-q}$, 1），则 $G_q(x) = x^{\frac{q}{1-q}}$，且当 $q \to 0$时,

$$\tfrac{1}{q} \int_0^1 [1 - G_q(x)]^2 dx = \tfrac{2q}{1+q} \to 0,$$

因此（4.2）不成立。

Remarks。下面我们定义一种扩展扫描累积和算法，该算法选择 $t = \widehat{\tau}$ 以最大化

$$a(t) L_I(t) \Big[= a(t) \exp(\tfrac{1}{2} Z_I^2(t)) \Big],$$

其中，L(*I*) 是剖面似然（3.2）。相比之下，第3节中针对常强度函数的扫描-CUSUM算法选择 $t = \hat{\tau}$ 以最大化 CUSUM得分 $|Z_I(t)|$。当（4.2）不成立时，扩展扫描-CUSUM可实现$\beta_T \to 1$ 。

Algorithm. 扩展扫描-CUSUM。按照扫描-CUSUM算法进行。将算法第 8 行中的“$\hat{\tau} = \arg\max_{t\in I}|Z_I(t)|$”替换为“$\hat{\tau} = \arg\max_{t\in I} a(t)L_I(t)$”。

定理 4。设 $q = o(\frac{1}{\log T})$ 为 $T \to \infty$。在条件 (2.1)、(2.3) 和 (4.1) 下，若

$$\frac{1}{q}\int_0^\infty [1 - G_q(x)]^2 dx \to 0 \ \ as \ q \to 0,$$

则扩展扫描累积和算法以 $c_{\text{scan}} = \sqrt{2\log(T\log T)}$ 实现 $\beta_T \to 1$ as $T \to \infty$。

5 多个序列的估计

这里我们展示如何在存在多个序列时应用扫描-CUSUM来降低变点估计误差。

Model and notations。

1. 设 $\mathbf{X}^n = (X^n(1) \ldots, X^n(T))$ 为第 n 个观测序列。假设存在未知的强度 $a(t)$，使得

$$Y^n(t) \sim_{\text{indep.}} \text{Bernoulli}(a(t)),$$

对于 $1 \le t \le T-1$ 和 $1 \le n \le N$。

2. 设 $\boldsymbol{\tau}^n = \{t : \ Y^n(t) = 1\}$ 为第 n 个序列的变点，并设 $J^n = \#\boldsymbol{\tau}^n$ 为变点的数量。将变点按 $\boldsymbol{\tau}^n$ 排列如下

$$\tau_1^n < \cdots < \tau_{J^n}^n.$$

3. 设 $\mu^n(t)$ 为 $X^n(t)$ 的均值。假设对于每个 n,

$$\mu^n(t+1) = \mu^n(t) + \Delta^n(t)Y^n(t),$$

其中 $\Delta^n(t)|Y^n(t) = 1$ 非零。

4. 假设在给定 $\mu^n(t)$ 的条件下，对于 $1 \le t \le T$ 和 $1 \le n \le N$,

$$X^n(t) = \mu^n(t) + \epsilon^n(t),$$

其中 $\epsilon^n(t) \sim_{\text{i.i.d.}}$ 服从N$(0, \sigma_X^2)$。

Estimation procedure for multiple sequences。

1. 令 $\widehat{\boldsymbol{\tau}}^n$ 为第 n 个序列的估计变点，这些变点是通过对 X^{n}应用第3节中的扫描-CUSUM算法得到的。

2. 令 $\widehat{J}^n = \#\widehat{\boldsymbol{\tau}}^n$ 为估计变点的数量，并令$\widehat{I}^n(1), \ldots, \widehat{I}^n(\widehat{J}^n)$为扫描-CUSUM算法识别出的包含变点的区间。令 L$^n(\widehat{I}^n(1)), \ldots,$ L$^n(\widehat{I}^n(\widehat{J}^n))$ 为相应的剖面似然。为简化记号，将这些似然表示为 L$^n(1), \ldots,$ L$^n(\widehat{J}^n)$。

3. 令 $\mathbf{a} = (a(t) : 1 \le t \le t-1)$ 为真实的变点强度函数，并令 $\widehat{a}(t)$, $1 \le t \le T-1$, 为最大似然估计

$$\ell(\mathbf{a}) = \sum_{n=1}^{N} \sum_{j=1}^{\hat{J}^n} \log\Big(\sum_{t \in I^n(j)} a(t) L_j^n(t) \Big) - N \sum_{t=1}^{T-1} a(t). \qquad (5.1)$$

4. 令 $\widehat{\boldsymbol{\tau}}^n = \{\widehat{\tau}_j^n : 1 \le j \le \widehat{J}^n\}$ 为基于所有序列信息得到的变点估计，其中

$$\widehat{\tau}_j^n = \arg\max_{t \in I^n(j)} \widehat{a}(t) L_j^n(t).$$

强度似然（5.1）通过不完整数据论证推导而得，其中变点作为潜变量。这引出了用于计算 $\widehat{a}(t)$ 的以下 EM算法。技术细节见附录 A。

EM procedure to compute $\widehat{a}(t)$。我们采用以下步骤迭代估计 $\widehat{a}(t)$。

1. Initialize with $a_0(t) = \frac{1}{N(T-1)} \sum_{n=1}^{N} \widehat{J}^n$ for all t.

2. 对于 $k = 0, \ldots, K-1$: 令

$$a_{k+1}(t) = \frac{1}{N} \sum_{(n,j): t \in I^n(j)} \Big(\frac{a_k(t) L_j^n(t)}{\sum_{u \in I^n(j)} a_k(u) L_j^n(u)} \Big).$$

3. 令 $\widehat{a}(t) = a_K(t)$。

6 数值比较

我们在此比较两种应用扫描-CUSUM的方法。

1. 不进行信息共享。对每个序列分别应用扫描-CUSUM，不估计 $a(t)$。这些估计值是估计过程第5节步骤1中的 $\widehat{\boldsymbol{\tau}}^n$ 。 2. 进行信息共享。使用所有序列估计强度函数，并将估计得到的强度用于变点估计。这些估计值是估计过程第5节步骤5中的 $\widetilde{\boldsymbol{\tau}}^n$ 。

我们针对 $N = 100$、 $T = 10{,}000$ 和$q = 0.0001$进行了三组实验，每组使用不同的强度函数生成器 G_q。

1. G_q=在 q 处的概率质量为 1。序列之间的变点相互独立。应用估计得到的强度时，预计会有一定的精度损失。

2. G_q=在 $100q$ 处的概率质量为 0.01，在 0 处的概率质量为 0.99。

$$(\#\{n : Y^n(t) = 1\}|a(t) > 0) \sim_{\text{approx}} \text{Poisson}(100Nq = 1).$$

由于 $a(t)$ 的变异性，变点之间存在一定程度的对齐，但这种对齐还不足以使估计误差接近于零。

3. G_q=贝塔分布($\frac{q}{1-q}$,1)。

$$(a(t)|Y^n(t) = 1) \sim \text{Beta}(\tfrac{1}{1-q}, 1) \doteq \text{Uniform}(0, 1).$$

变点的对齐程度足够高，因此估计误差接近于零。

在每组实验中，我们从G_q生成 $a(t)$、 $1 \le t \le T - 1$，且它们独立同分布。对于每个 $1 \le n \le N$，我们使用隐马尔可夫模型 (2.3) 和 (2.4) 生成 $X^n(t)$，其中 $\sigma_X = \sigma_\xi = 1$。我们为扫描 CUSUM 算法选择阈值 $c_{\text{scan}} = 5.05$ 。该阈值基于 0.05 的第一类错误概率，即$\backsim P(J^n > 0|J^n = 0) = 0.05$。

对于具有 n $J^n > 0$ 的序列，我们计算

$$\begin{aligned} \widehat{\alpha}_T^n &= \#\{j : \kappa(\tau_j^n) \neq 1\}/J^n, \\ \widehat{\beta}_T^n &= \#\{j : \widehat{Y}(\tau_j^n) = 1 \text{ and } \kappa(\tau_j^n) = 1\}/J^n. \end{aligned}$$

^我们沿序列对它们取平均，得到 $\widehat{\alpha}_T$ 和 β_T。

实验重复进行100次，并计算标准误。表2显示，在对齐的变点上应用信息共享可显著改善结果。

G_q		不共享信息	共享信息
在 q处的概率质量为1	α_T	0.069±0.003	0.062±0.002
	β_T	0.308±0.006	0.271±0.005
$100q$ 处的概率质量为 0.01	α_T	0.069±0.003	0.061±0.002
以及 0 处的 0.99	β_T	0.303±0.005	0.403±0.006
贝塔分布($\frac{q}{1-q}$,1)	α_T	0.064±0.005	0.056±0.005
	β_T	0.305±0.010	0.835±0.011

表2：无信息共享与信息共享的估计准确度比较。

7 证明

考虑 $^{X}{}_{(t)}\,{}_{\sigma X}$ 而非 $X(t)$，我们可以不失一般性地假设 $\sigma_X = 1$。以下引理在附录B中证明。引理1确保 $\gamma_{\text{lower}}(\Delta_0) < \infty$。引理 2 表明，似然函数的收敛会导致其中位数收敛。

DEFINITION. 给定一个非负函数 $\mathbf{r} = (r(i) : -\infty < i < \infty)$，且满足 $r(0) > 0$，定义

$$W(\mathbf{r}) = \sum_{i=-\infty}^{\infty} |i| \min(1, \tfrac{r(i)}{r(0)}).$$

引理1. (a) 对于任意满足 $r(0) > 0$ 且 $s = \sum_{i=-\infty}^{\infty} r(i) < \infty$的非负函数 $\mathbf{r}$,

$$|\text{med}(\mathbf{r})| \le 2W(\mathbf{r}).$$

(b) 对于 (2.6) 中定义的 $\mathbf{p}$ 以及 $\Delta_0 > 0$, $E(W(\mathbf{p})\mathbf{1}_{\{|\Delta| \ge \Delta_0\}}) < \infty$。

引理 2。对于非负函数 $\mathbf{r}$ 和 $\mathbf{r}_T$，若 $s = \sum_{t=-\infty}^{\infty} r(t) < \infty$ 且 $\sum_{t=-\infty}^{\infty} |r_T(t) - r(t)| \to 0$，当 $T \to \infty$ 时，

$$\sum_{t=-\infty}^{\text{med}(\mathbf{r})-1} r(t) < \tfrac{s}{2} < \sum_{t=-\infty}^{\text{med}(\mathbf{r})} r(t), \tag{7.1}$$

则当 T 足够大时，$\text{med}(r_T) = \text{med}(r)$。

NOTATIONS。若 $\sup_T |\frac{a_T}{b_T}| < \infty$，则令 $a_T = O(b_T)$。若$P(|\frac{X_T}{Y_T}| \geq \epsilon) \to 0$对所有 $\epsilon > 0$成立，则令 $X_T = o_p(Y_T)$。若对所有 $\epsilon > 0$都存在常数 $M_\epsilon > 0$，使得 $P(|\frac{X_T}{Y_T}| > M_\epsilon) \leq \epsilon$，则令 $X_T = O_p(Y_T)$。当 $u > v$时，约定$\prod_{t=u}^{v} \cdot = 1$ 。

7.1 定理 1 的证明

PROOF OF 上极限 $\beta_T \leq \beta_{\text{upper}}$。令 $\tau = \tau_j$为随机选取的变点。令 $U > 0$，并令 k 满足 $\tau \in I_k := \{(k-1)U + 1, \ldots\ldots, kU\}$。令 $K = \lfloor \frac{T-1}{U} \rfloor$。令 J_k为 I_k中的变点数量。

对于 $t \in I_k$，令似然

$$L_k(t) = \prod_{u=(k-1)U+2}^{t} \phi(X_u - \mu(\tau)) \prod_{u=t+1}^{kU} \phi(X_u - \mu(\tau+1)). \tag{7.2}$$

考虑一个神谕估计量

$$\widehat{\tau}_j = \begin{cases} \tau & \text{if } J_{k-1} + J_k + J_{k+1} > 1 \text{ or } k > K, \\ \arg\max_{t \in I_k} L_k(t) & \text{otherwise.} \end{cases} \tag{7.3}$$

由于 $q \to 0$, (7.3) 中的第一种情况以概率 $\to 0$ 发生。以下分析中，我们以第二种情况为条件。

令 $Z(r) = X(r) - \mu(\tau)$，其中 $r \leq \tau$，$Z(r) = \mu(\tau+1) - X(r)$，其中 $r > \tau$，且 $\Delta = \Delta(\tau) = \mu(\tau+1) - \mu(\tau)$。令 $S^+(i) = \sum_{u=\tau+1}^{\tau+i} Z(u)$，且 $S^-(i) = \sum_{u=\tau-i+1}^{\tau} Z(u)$。

对于 $t \in I_k$,

$$L_k(t) = L_k(\tau) \times \begin{cases} \exp(\Delta S^+(t-\tau) - \frac{(t-\tau)\Delta^2}{2}) & \text{if } t > \tau, \\ \exp(\Delta S^-(\tau-t) - \frac{(\tau-t)\Delta^2}{2}) & \text{if } t < \tau. \end{cases} \tag{7.4}$$

因此，$\widehat{\tau}_j = \tau$ 为真的概率是

$$\begin{aligned} & P(\Delta S^+(i) < \tfrac{i\Delta^2}{2} \text{ for } 1 \leq i \leq kU - \tau) \\ \times \quad & P(\Delta S^-(i) < \tfrac{i\Delta^2}{2} \text{ for } 1 \leq i \leq \tau - (k-1)U - 1). \end{aligned}$$

变点估计量不可能比神谕估计量具有更小的 β_T，因此

$$\limsup \beta_T \leq P\Big(\max_{1 \leq i \leq V-1} (\Delta S^+(i) - \tfrac{i\Delta^2}{2}) < 0, \max_{1 \leq i \leq U-V} (\Delta S^-(i) - \tfrac{i\Delta^2}{2}) < 0 \Big), \tag{7.5}$$

其中 V在 $\{1,\ldots,U\}$上服从均匀分布，且独立于 $S^+(i)$和 $S^-(i)$。

由于 $V\uparrow\infty$ and $(U-V)\uparrow\infty$ 在概率意义下，当 $U\to\infty$ 时，由 (7.5),

$$\begin{aligned}\limsup\beta_T &\le P^2\Big(\sup_{i\ge1}(\Delta S^+(i)-\tfrac{i\Delta^2}{2})<0\Big)\\ &= E\Big[\exp\Big(-2\sum_{i=1}^{\infty}i^{-1}\Phi(-\tfrac{\sqrt{i}|\Delta|}{2})\Big)\Big]=\beta_{\rm upper},\end{aligned}$$

第一个等式参见 Siegmund (1985) 的推论 8.44。 □

Proof of 下极限 $\gamma_T(\Delta_0)\ge\gamma_{\rm lower}(\Delta_0)$。令 $\tau=\tau_j$ 为随机选取的一个变点。令 $\Delta=\Delta(\tau)$。由于 $\alpha_T\to0$，只需证明对于任意$U>0$，存在 $\gamma_{{\rm lower},\ U}(\Delta_0)$，使得

$$\begin{aligned}&\liminf[\gamma_T(\Delta_0)P(|\Delta|\ge\Delta_0)+U\alpha_T] \qquad (7.6)\\ =\ &\liminf[E(d(\tau,\widehat{\boldsymbol\tau})\mathbf{1}_{\{\kappa(\tau)=1\text{ and }|\Delta|\ge\Delta_0\}})+U\alpha_T]\\ \ge\ &\gamma_{{\rm lower},U}(\Delta_0)P(|\Delta|\ge\Delta_0),\end{aligned}$$

并且 $\gamma_{{\rm lower},\ U}(\Delta)\to\gamma_{\rm lower}(\Delta_0)$，当 $U\to\infty$ 时。

令 k满足 $\tau\in I_k:=\{(k-1)U+1,\ldots,kU\}$。令 $K=\lfloor\frac{T-1}{U}\rfloor$。令 J_k 为 I_k中的变点个数。考虑一个神谕估计量

$$\widehat{\tau}_j=\begin{cases}\tau & \text{if } J_{k-1}+J_k+J_{k+1}>1 \text{ or } k>K,\\ \mathrm{med}(\mathbf{L}_k) & \text{otherwise,}\end{cases}$$

其中，$\mathbf{L}_k=(L_k(t):\ -\infty<t<\infty)$，其中 $L_k(t)$ 是在$t\in I_k$ 时的似然（7.2），而在其他情况下为 $L_k(t)=0$。

根据（7.4），

$$\mathrm{med}(\mathbf{L}_k)-\tau=\mathrm{med}(\mathbf{p}_{(k-1)U+1-\tau,kU-\tau}),$$

其中 $p_{a,\ b}(i)=p(i)\mathbf{1}_{\{a\le i\le b\}}$ 对所有 i 均成立。

变点估计量不可能有更小的

$$E(d(\tau,\widehat{\boldsymbol\tau})\mathbf{1}_{\{\kappa(\tau)=1\text{ and }|\Delta|\ge\Delta_0\}})+U\alpha_T$$

值（相较于神谕估计量），因此（7.6）中的不等式成立，且

$$\gamma_{{\rm lower},U}(\Delta_0)=E\Big(|\mathrm{med}(\mathbf{p}_{V-U,V-1})|\Big||\Delta|\ge\Delta_0\Big), \qquad (7.7)$$

其中 V 在 $\{1,\ldots,U\}$ 上均匀分布，且与 $\mathbf{p}$ 独立。

由于 $V\uparrow\infty$ 和 $(V-U)\uparrow\infty$ 依概率收敛于 $U\to\infty$，根据引理 1、2 和支配收敛定理，

$$\gamma_{{\rm lower},U}(\Delta_0)\to E\Big(|\mathrm{med}(\mathbf{p})|\Big||\Delta|\ge\Delta_0\Big)=\gamma_{\rm lower}(\Delta_0)$$

作为 $U\to\infty$。 □

7.2 定理 2 和 (3.4) 的证明

7.2.1 $\alpha_T \to 0$ 的证明

将 $q = \frac{\omega^3}{\text{对数 } T}$表示为，其中 $\omega \to 0$作为 $T \to \infty$。从 $\boldsymbol{\tau}$中随机选择 τ_j。概率为 $\to 1$,

$$|\Delta(\tau_j)| \geq \omega,\ |\Delta(\tau_{j-1})| \geq \omega \text{ (if } j > 1) \text{ and } |\Delta(\tau_{j+1})| \geq \omega \text{ (if } j < J). \quad (7.8)$$

由于 $q = o(\frac{\omega^2}{\text{对数 } T})$，以概率$\to 1$,

$$\begin{aligned} \tau_{j-1} &\leq \tau_j - 48\omega^{-2}\log T, \\ \tau_{j+1} &\geq \tau_j + 48\omega^{-2}\log T. \end{aligned} \quad (7.9)$$

在 (7.8) 下，若 $6\omega^{-2}$ 对数 $T \leq \ell_b < 12\omega^{-2}$ 对数 T，则对于 $i = j-1$, j和$j+1$, $Z_b(\tau_i) \sim \mathsf{N}(\nu_i, 1)$，且

$$|\nu_i| = \sqrt{\tfrac{\ell_b}{2}}|\Delta(\tau_j)| \geq \sqrt{3\log T}.$$

因此，以概率$\to 1$，$|Z_b(\tau_i)| \geq c_{\text{scan}}$，这意味着 $|\tau_i - \widehat{\tau}_i| \leq 12\omega^{-2}$ 对数 T。根据 (7.9)，我们得出

$$\begin{aligned} \tau_j - \widehat{\tau}_{j-1} &> \tfrac{1}{2}(\tau_j - \tau_{j-1}), \\ \widehat{\tau}_{j+1} - \tau_j &> \tfrac{1}{2}(\tau_{j+1} - \tau_j), \\ |\widehat{\tau}_j - \tau_j| &< \tfrac{1}{2}\min(\tau_{j+1} - \tau_j, \tau_j - \tau_{j-1}). \end{aligned} \quad (7.10)$$

以概率$\to 1$，$\boldsymbol{\tau}$ 中的所有 $\widehat{\tau}$ 都具有 $\widehat{\tau}_j$的形式，因为

$$\sum_{b=0}^{b_T}\sum_{t=\ell_b}^{T-\ell_b} P(|Z_b(t)| \geq c_{\text{scan}}, Y(u) = 0 \text{ for } |u-t| < \ell_b) \leq (b_T+1)\Phi(-c_{\text{scan}}) \to 0.$$

因此，由 (7.10)，$P(\kappa(\tau_j) = 1) \to 1$。

7.2.2 $\beta_T \to \beta_{\text{upper}}$的证明

从 $\boldsymbol{\tau}$ 中随机选取 $\tau = \tau_j$，并令 $\Delta = \Delta(\tau)$。令 $\delta > 0$，并令

$$A_T = \{\delta \leq |\Delta| \leq \delta^{-1}, Y(t) = 0 \text{ for } 1 \leq |t - \tau| \leq 12\delta^{-2}\log T\}. \quad (7.11)$$

由于 $q = o(\frac{1}{\log T})$,

$$P(A_T) \to P(\delta \leq |\Delta| \leq \delta^{-1}) \text{ as } T \to \infty. \quad (7.12)$$

令 $\mathcal{I}_T$ 为所有满足以下条件的区间集合 $I = (t^* - \ell_b, t^* + \ell_b)[= (u^*, v^*)]$

$$
\begin{aligned}
|t^* - \tau| &\leq (\log T)^{\frac{1}{4}}, && (7.13)\\
2\delta^2 \log T < \ell_b &< 12\delta^{-2} \log T. && (7.14)
\end{aligned}
$$

根据附录 C 中的引理 6，在 A_T 条件下，以 $\to 1$ 的概率，对于满足 $\tau \in I$ 的区间 I （扫描-CUSUM算法的第 7 行），有 $\mathcal{I}_T$。

令 $S^+(i) = \sum_{u=1}^{i} Z^+(u)$，且令 $S^-(i) = \sum_{u=1}^{i} Z^-(u)$，其中 $Z^+(u) = \mu(\tau+1) - X(\tau+u)$，以及 $Z^-(u) = X(\tau-u+1) - \mu(\tau)$。根据 Siegmund（1985）的推论 8.44,

$$
\tfrac{1}{2}\Delta^2 \nu(\Delta) = P\Big(\sup_{i\geq 1}(S^+(i) - \tfrac{i|\Delta|}{2}) < 0, \sup_{i\geq 1}(S^-(i) - \tfrac{i|\Delta|}{2}) < 0 \Big| \Delta \Big). \quad (7.15)
$$

下面证明

$$
\begin{aligned}
\sup \Big| \sqrt{2\ell_b}[Z_I(\tau+i) - Z_I(\tau)] - (2S^+(i) - i\Delta) \Big| \mathbf{1}_{A_T} &\xrightarrow{p} 0, && (7.16)\\
\sup \Big| \sqrt{2\ell_b}[Z_I(\tau-i) - Z_I(\tau)] - (2S^-(i) - i\Delta) \Big| \mathbf{1}_{A_T} &\xrightarrow{p} 0, && (7.17)
\end{aligned}
$$

其中上确界取遍 $I \in \mathcal{I}_T$ 和 $1 \leq i \leq (\log\ T)^{\frac{1}{4}}$。我们在附录 C 的引理 7 中证明

$$
\sum_{I\in\mathcal{I}_T} \sum_{u\in I: |u-\tau| \geq (\log T)^{\frac{1}{4}}} P(|Z_I(u)| \geq |Z_I(\tau)|, A_T) \to 0. \quad (7.18)
$$

根据 (7.16)–(7.18)，$\widehat{\tau} = \tau$ 的概率渐近等于 $\sup_{i\geq 1}(S^+(i) - \frac{i|\Delta|}{2}) < 0$ 和 $\sup_{i\geq 1}(S^-(i) - \frac{i|\Delta|}{2}) < 0$ 的概率。由 (7.15) 并令 $\delta \to 0$，我们得出 $\beta_T \to \beta_{\text{upper}}$。 □

Proofs of (7.16) 和 (7.17)。考虑 $t = \tau + i$，其中 $1 \leq i \leq (\log\ T)^{\frac{1}{4}}$。由于 $v^* - u^* = 2\ell_b$，我们可以表示为

$$
\begin{aligned}
\sqrt{2\ell_b} Z_I(\tau) &= \lambda(\tau)(\widehat{\nu} - \widehat{\mu}), && (7.19)\\
\text{where } \widehat{\nu} &= \tfrac{S(v^*) - S(\tau)}{v^* - \tau},\\
\widehat{\mu} &= \tfrac{S(\tau) - S(u^*)}{\tau - u^*},\\
\lambda(t) &= \sqrt{(v^* - t)(t - u^*)}.
\end{aligned}
$$

由于 $|t^*-\tau| \le (\log\ T)^{\frac{1}{4}}$,

$$\begin{aligned}
\lambda(\tau+i) &= \sqrt{\ell_b^2 - |\tau+i-t^*|^2} && (7.20)\\
&= \ell_b[1+O((\log T)^{-\frac{3}{2}})],\\
\tfrac{v^*-\tau}{v^*-\tau-i} &= (1-\tfrac{i}{v^*-\tau})^{-1} && (7.21)\\
&= 1+\tfrac{i}{\ell_b}+O((\log T)^{-\frac{3}{2}}),\\
\tfrac{\tau-u^*}{\tau+i-u^*} &= 1-\tfrac{i_b}{\ell_b}+O((\log T)^{-\frac{3}{2}}). && (7.22)
\end{aligned}$$

根据迭代对数律，或者根据标准正态分布的尾概率界,

$$\begin{aligned}
|\widehat{\nu}-\mu(\tau+1)| &= O_p((\log T)^{-\frac{1}{3}}), && (7.23)\\
|\widehat{\mu}-\mu(\tau)| &= O_p((\log T)^{-\frac{1}{3}}). && (7.24)
\end{aligned}$$

由 (7.19)–(7.24),

$$\begin{aligned}
&\sqrt{2\ell_b}[Z_I(\tau+i)-Z_I(\tau)]\\
=\ &\lambda(\tau+i)[(\tfrac{v^*-\tau}{v^*-\tau-i})\widehat{\nu}-(\tfrac{\tau-u^*}{\tau+i-u^*})\widehat{\mu}-\tfrac{2\ell_b}{\lambda^2(\tau+i)}(S(\tau+i)-S(\tau))]-\lambda(\tau)(\widehat{\nu}-\widehat{\mu})\\
=\ &(\mathrm{A})+(\mathrm{B})+(\mathrm{C}),
\end{aligned}$$

其中

$$\begin{aligned}
(\mathrm{A}) &= -\tfrac{2\ell_b}{\lambda(\tau+i)}(S(\tau+i)-S(\tau))\\
&= [1+O((\log T)^{-\frac{3}{2}})](2S^+(i)-2i\Delta)\\
&= 2S^+(i)-2i\Delta+(\log T)^{-\frac{3}{2}}O_p\Big(\max_{1\le i\le(\log T)^{\frac{1}{4}}}|S^+(i)-i\Delta|\Big)\\
&= 2S^+(i)-2i\Delta+O_p((\log T)^{-\frac{5}{4}}),\\
(\mathrm{B}) &= [\lambda(\tau+i)(\tfrac{v^*-\tau}{v^*-\tau-i})-\lambda(\tau)]\widehat{\nu}\\
&= \tfrac{i\lambda(\tau+i)}{\ell_b}\widehat{\nu}+O((\log T)^{-\frac{1}{2}})\\
&= i\mu(\tau+1)+O_p((\log T)^{-\frac{1}{12}}),\\
(\mathrm{C}) &= -[\lambda(\tau+i)(\tfrac{\tau-\mu^*}{\tau+i-\mu^*})-\lambda(\tau)]\widehat{\mu}\\
&= -i\mu(\tau)+O_p((\log T)^{-\frac{1}{12}}).
\end{aligned}$$

根据上述展开式和 $-2i\Delta+i(\mu(\tau+1)\ -\mu(\tau))\ =\ -i\Delta$，我们得出 (7.16)。对于带有 $-(\log\ T)^{\frac{1}{4}} \le i \le -1$ 的 (7.17)，计算过程类似。 □

7.2.3 (3.4) 的证明

从 $\boldsymbol{\tau}$ 中随机选择 $\tau = \tau_j$，并令 $\Delta = \Delta(\tau)$。令 A_T 按照 (7.11) 定义，并令 $\mathcal{I}_T$ 为满足 (7.13) 和 (7.14) 的所有区间 I。根据附录 C 中的引理 6，在 A_T 下，以概率 $\to 1$，扫描 -CUSUM 算法中满足 $\tau \in I$ 的区间 I 属于 $\mathcal{I}_T$。

令 $\widehat{\tau}$ 为 $Z_I(t)$ 的最大化点。根据 (7.19)–(7.24)，$i^* = \widehat{\tau} - \tau$ 渐近上是 p $= (p(i) : i \in \mathrm{Z})$ 的众数，其中

$$\begin{aligned} p(i) &= \exp(\Delta S^+(i) - \tfrac{i\Delta^2}{2}), \\ p(-i) &= \exp(\Delta S^-(i) - \tfrac{i\Delta^2}{2}), \\ p(0) &= 1. \end{aligned} \tag{7.25}$$

令 $\delta \to 0$，我们得出 $\gamma_T(\Delta_0) \to \gamma_{\mathrm{scan}}(\Delta_0)$。

7.3 定理 3 的证明

令 $\tau = \tau_j$ 为随机选取的变点，并令 k 满足 $\tau \in I_k := \{2k-1, 2k\}$。令 $K = \lfloor \frac{T-1}{U} \rfloor$。令 J_k 为 I_k 中变点的数量。令 $\Delta = \Delta(\tau)$。

考虑一个神谕估计量

$$\widehat{\tau}_j = \begin{cases} \tau & \text{if } J_{k-1} + J_k + J_{k+1} > 1 \text{ or } k > K, \\ \arg\max_{t \in I_k} a(t) L_k(t) & \text{otherwise,} \end{cases} \tag{7.26}$$

其中，$L_k(t)$ 是式（7.2）中定义的似然，且 $U = 2$。由于 $q \to 0$，式（7.26）的第一种情况以概率 $\to 0$ 发生。在下面的分析中，我们以第二种情况为条件。

由于

$$(a(t) | Y(t) = 1) \sim G_q^* \text{ where } dG_q^*(a) = \tfrac{a}{q} dG_q(a),$$

根据式（7.4），$\widehat{\tau}_j \neq \tau$ 的概率渐近地为

$$\begin{aligned} & q^{-1} E_q(a(t) \mathbf{1}_{\{a(t) < a(t+1) \exp(\Delta S^+(1) - \frac{\Delta^2}{2})\}}) \\ \geq\ & q^{-1} P(\Delta S^+(1) - \tfrac{\Delta^2}{2} > 0) E_q(a(t) \mathbf{1}_{\{a(t) < a(t+1)\}}), \end{aligned}$$

其中，E_q 表示在该条件下的期望，$a(t)$ 和 $a(t+1)$ 是独立同分布的 G_q。

变点估计量不可能比神谕估计量具有更小的 β_T，因此 $E_q(a(t) \mathbf{1}_{\{a(t) < a(t+1)\}}) = \frac{1}{2} E_q[\min(a(t)\ a(t+1))]$，

$$\limsup_{T \to \infty} (1 - \beta_T) \geq 1 - \tfrac{1}{2} P(Z > \tfrac{|\Delta|}{2}) \Big\{ \limsup_{q \to 0} q^{-1} E_q[\min(a(t), a(t+1))] \Big\},$$

其中 $Z \sim$ 服从 N(0, 1)，且 $\Delta \sim F$ 相互独立。

由 (4.2),

$$\limsup_{q\to 0} q^{-1} E_q[\min(a(t), a(t+1))] = \limsup_{q\to 0} q^{-1} \int_0^1 [1 - G_q(x)]^2 dx > 0, \tag{7.27}$$

因此我们得出 $\limsup_{T\to\infty}(1 - \beta_T) > 0$。 □

7.4 定理 4 的证明

定理 4 可由引理 3–5 推出。引理 3 和引理 4 在第 7.4.1 节和第 7.4.2 节中证明。引理 5 在附录 D 中证明。

设 E_q 为期望，其中 a_0、$a^+(i)$ $a^-(i) \sim G_q$ 对于 $i > 0$、$Z^+(u)$ $Z^-(u) \sim$ N(0, 1)，对于 $u > 0$，且 $\Delta \sim F$ 均相互独立。设 $S^+(i) = \sum_{u=1}^i Z^+(u)$，且 $S^-(i) = \sum_{u=1}^i Z^-(u)$。设

$$\begin{aligned} p(i) &= \exp(\Delta S^+(i) - \tfrac{i\Delta^2}{2}), \\ p(-i) &= \exp(\Delta S^-(i) - \tfrac{i\Delta^2}{2}), \\ B^+ &= \{a_0 \le 2a^+(i)p(i) \text{ for some } i > 0\}, \\ B^- &= \{a_0 \le 2a^-(i)p(-i) \text{ for some } i > 0\}. \end{aligned}$$

从 $\boldsymbol{\tau}$ 中随机选择 $\tau = \tau_j$，并令 $\Delta = \Delta(\tau)$。令 A_T 如 (7.11) 中所定义，并令 $\mathcal{I}_T$ 为满足 (7.13) 和 (7.14) 的所有区间。

引理 3. 若 $q^{-1}\int_0^1 [1 - G_q(x)]^2 dx \to 0$ 则 $E_q(a_0 \mathbf{1}_{B^+\cup B^-}) \to 0$。

引理 4。

$$\sup \left| \tfrac{1}{2}[Z_I^2(\tau + i) - Z_I^2(\tau)] - (\Delta S^+(i) - \tfrac{i\Delta^2}{2}) \right| \mathbf{1}_{A_T} \xrightarrow{p} 0, \tag{7.28}$$

$$\sup \left| \tfrac{1}{2}[Z_I^2(\tau - i) - Z_I^2(\tau)] - (\Delta S^-(i) - \tfrac{i\Delta^2}{2}) \right| \mathbf{1}_{A_T} \xrightarrow{p} 0, \tag{7.29}$$

其中，上确界取遍 $I \in \mathcal{I}_T$ 和 $1 \le i \le (\log T)^{\frac{1}{4}}$。

引理 5。

$$\sum_{I\in\mathcal{I}_T} \sum_{u\in I: |u-\tau| \ge (\log T)^{\frac{1}{4}}} P(a(u) e^{\frac{1}{2}Z_I^2(u)} \ge a(\tau) e^{\frac{1}{2}Z_I^2(\tau)}, A_T) \to 0.$$

Proof of Theorem 4。由于当 $a(\tau+i)e^{\frac{1}{2}Z_I^2(\tau+i)} < a(\tau)e^{\frac{1}{2}Z_I^2(\tau)}$ 时，$\widehat{\tau} = \tau$ 对所有满足 $i \neq 0$ 的 $\tau + i \in I$ 均成立，根据 (7.12)、引理 4–6 以及 (7.25) 中 $p(i)$ 的展开式，

上极限 $(1-\beta_T) \le$ 上极限 $P(a(\tau) \le 2a(\tau+i)p(i)$，对某个 $1 \le |i| \le (\log T)^{\frac{1}{4}})$ 成立。 (7.30)

令 $a^+(i) = a(\tau + i)$，且 $a^-(i) = a(\tau - 1)$。由于

$$a(\tau) \sim G_q^* \text{ with } dG_q^*(a) = \tfrac{a}{q} dG_q(a),$$

由 (7.30) 可得

$$\limsup(1 - \beta_T) \le q^{-1} E_q(a_0 \mathbf{1}_{\{B^+ \cup B^-\}}),$$

且 $\beta_T \to 1$ 由引理 3 可得，其中 $a_0[= a(\tau)] \sim G_q$ 满足 E_q。 □

7.4.1 引理 3 的证明

由于 $E_q(a_0 \mathbf{1}_{B^+}) = E_q(a_0 \mathbf{1}_{B^-})$，只需证明

$$q^{-1} E_q(a_0 \mathbf{1}_{B^+} | \Delta) \to 0 \text{ for all } \Delta \neq 0. \tag{7.31}$$

令 $B^+(i) = \{a_0 \le 2a^+(i)p(i)\}$，并验证

$$\begin{aligned}
& q^{-1} E_q(a_0 \mathbf{1}_{B^+(i)} | \Delta) \\
= \; & q^{-1} E_q(a_0 \mathbf{1}_{\{a_0 \le 2a^+(i)p(i), \Delta S^+(i) \le \frac{i\Delta^2}{4}\}} | \Delta) \\
& \quad + q^{-1} E_q(a_0 \mathbf{1}_{\{a_0 \le 2a^+(i)p(i), \Delta S^+(i) > \frac{i\Delta^2}{4}\}} | \Delta) \\
\le \; & q^{-1} E_q(2a^+(i) e^{-\frac{i\Delta^2}{4}} | \Delta) + q^{-1} E_q(a_0 \mathbf{1}_{\{\Delta S^+(i) > \frac{i\Delta^2}{4}\}} | \Delta) \\
= \; & 2e^{-\frac{\Delta^2}{4}} + \Phi(-\tfrac{\sqrt{i}|\Delta|}{4}).
\end{aligned}$$

另一个有用的界为

$$\begin{aligned}
q^{-1} E_q(a_0 \mathbf{1}_{B^+(i)} | \Delta) &\le q^{-1} E_q[\min(a_0, 2a^+(i)p(i)) | \Delta] \\
&\le q^{-1} E_q[\min(a_0, a^+(i))(1 + 2p(i)) | \Delta] \\
&= 3q^{-1} E_q[\min(a_0, a^+(i))].
\end{aligned}$$

对于任意 $i_0 > 0$,

$$\begin{aligned}
& q^{-1} E_q(a_0 \mathbf{1}_{B^+} | \Delta) \\
\le \; & \sum_{i=1}^{i_0} q^{-1} E_q(a_0 \mathbf{1}_{B^+(i)} | \Delta) + \sum_{i=i_0+1}^{\infty} q^{-1} E_q(a_0 \mathbf{1}_{B^+(i)} | \Delta) \\
\le \; & 3i_0 q^{-1} E_q[\min(a(1), a(2))] + \sum_{i=i_0+1}^{\infty} [2e^{-\frac{i\Delta^2}{4}} + \Phi(-\tfrac{\sqrt{i}\Delta}{4})].
\end{aligned}$$

由于 $q^{-1} E_q[\min(a(1)\ a(2))] = q^{-1} \int_0^1 [1 - G_q(x)]^2 dx \to 0$，因此只要将 i_0 取得足够大，就能使 $\limsup_{q \to 0} q^{-1} E_q(a_0 \mathbf{1}_{B^+} | \Delta)$ 任意小。

7.4.2 引理4的证明

根据 (7.16) 和 (7.17)，只需证明

$$\sup\left|Z_I(\tau+i)+Z_I(\tau)-\Delta\sqrt{2\ell_b}\right|\left|Z_I(\tau+i)-Z_I(\tau)\right|\mathbf{1}_{A_T}\xrightarrow{p}0. \tag{7.32}$$

回忆 $I=(t^*-\ell_b,t^*+\ell_b)=(u^*,v^*)$，$\lambda(t)=\sqrt{(v^*-t)(t-u^*)}$，以及$|\tau-t^*|\le(\log T)^{\frac{1}{4}}$for $I\in\mathcal{I}_T$。由于 $v^*-u^*=2\ell_b$，根据 (3.1)，$Z_I(\tau+i)\sim$N(ν_i,1)，其中

$$\begin{aligned}\nu_i &= \begin{cases}\frac{\lambda(\tau+i)}{\sqrt{2\ell_b}}(1-\frac{i}{\tau+i-u^*})\Delta & \text{if } i>0,\\ \frac{\lambda(\tau+i)}{\sqrt{2\ell_b}}(1-\frac{|i|}{v^*-\tau+|i|})\Delta & \text{if } i<0,\end{cases}\\ &= [1+O((\log T)^{-\frac{3}{4}})]\sqrt{\tfrac{\ell_b}{2}}\Delta.\end{aligned} \tag{7.33}$$

由于 $\mathcal{I}_T$ 中有 $O(\log T)$ 个区间，因此可得 $\sup|Z_I(\tau+i)-\nu_i|=O_p(\log\log T)$，从而根据 (7.33)，

$$\sup|Z_I(\tau)+Z_I(\tau+i)-\Delta\sqrt{2\ell_b}|\mathbf{1}_{A_T}=O_p((\log T)^{-\frac{1}{4}}). \tag{7.34}$$

根据 (7.16) 和 (7.17)，

$$\sup|Z_I(\tau+i)-Z_I(\tau)|\mathbf{1}_{A_T}=O_p((\log T)^{-\frac{1}{4}}). \tag{7.35}$$

由 (7.34) 和 (7.35) 可得 (7.32)。

$\ell(a)$的推导与EM算法

给定 $\mathbf{X}^n$, $1\le n\le N$，定义潜变量 $\boldsymbol{\tau}=(\tau_j^n:\ 1\le n\le N, 1\le j\le J^n)$，其中

$$P(\tau_j^n=t|\mathbf{X}^n)=\frac{a(t)L_j^n(t)}{\sum_{u\in I^n(j)}a(u)L_j^n(u)}\ \text{for}\ t\in I_j^n.$$

基于对 J^n 的 Poisson($\sum_{t=1}^{T-1}a(t)$) 近似，不完全数据似然函数为

$$L(\mathbf{a})=e^{\ell(a)}=e^{-N\sum_{t=1}^{T-1}a(t)}\prod_{n=1}^{N}\prod_{j=1}^{\hat{J}^n}\Big(\sum_{t\in I^n(j)}a(t)L_j^n(t)\Big).$$

完全数据似然函数为

$$L(\mathbf{a};\boldsymbol{\tau})=e^{-N\sum_{t=1}^{T-1}a(t)}\prod_{n=1}^{N}\prod_{j=1}^{\hat{J}^n}[a(t)L_j^n(t)]^{\mathbf{1}_{\{\tau_j^n=t\}}}.$$

给定参数的当前估计值 $\mathbf{a}_k = (a_k(t): 1 \le t \le T-1)$，$\tau_j^n = t$ 的条件概率为

$$T_{jk}^n(t) = \frac{a_k(t)L_j^n(t)}{\sum_{u\in I_j^n} a_k(u)L_j^n(u)} \text{ for } t \in I_j^n.$$

因此，E步，

$$\begin{aligned} Q(\mathbf{a}|\mathbf{a}_k) &= E_{\boldsymbol{\tau}|\mathbf{a}_k}[\log L(\mathbf{a};\boldsymbol{\tau})] \\ &= -N\sum_{t=1}^{T-1} a(t) + \sum_{n=1}^{N}\sum_{j=1}^{\hat{J}^n}\sum_{t\in I^n(j)} T_{jk}^n(t)\log(a(t)L_j^n(t)). \end{aligned}$$

$Q(\mathbf{a}|\mathbf{a}_k)$ 的最大化发生在

$$-N + \sum_{(n,j):t\in I^n(j)} \frac{T_{jk}^n(t)}{a(t)} = 0 \Rightarrow a_{k+1}(t) = \frac{1}{N}\sum_{(n,j):t\in I^n(j)} T_{jk}^n(t).$$

B 引理1和2的证明

Proof of Lemma 1。(a) 设 $s = \sum_{i=-\infty}^{\infty} r(i)$。

$$\begin{aligned} |\text{med}(\mathbf{r})| &= \frac{1}{s}\sum_{i=-\infty}^{\infty} |\text{med}(\mathbf{r})|r(i) \\ &\le \frac{1}{s}\sum_{i=-\infty}^{\infty} |i-\text{med}(\mathbf{r})|r(i) + \frac{1}{s}\sum_{i=-\infty}^{\infty} |i|r(i). \end{aligned}$$

由于 $M(\mathbf{r})$ =中位数(r) 在所有 $M(\mathbf{r})$ 中（包括 $M(\mathbf{r}) = 0$）使 $\sum_{i=-\infty}^{\infty} |i - M(\mathbf{r})|r(i)$ 最小，

$$|\text{med}(\mathbf{r})| \le \frac{2}{s}\sum_{i=-\infty}^{\infty} |i|r(i).$$

引理1(a) 可由 $s \ge \max(r(0)\ r(i))$ 得出。

(b) 由于 $p(0) = 1$,

$$W(\mathbf{p}) = \sum_{i=-\infty}^{\infty} |i|\min(1, p(i)), \tag{B.1}$$

因此

$$E(W(\mathbf{p})\mathbf{1}_{\{|\Delta|\ge\Delta_0\}}) = 2\sum_{i=1}^{\infty} iE[\min(1, e^{\Delta S^+(i) - \frac{i\Delta^2}{2}})\mathbf{1}_{\{|\Delta|\ge\Delta_0\}}].$$

由于

$$E[\min(1, e^{\Delta S^+(i) - \frac{i\Delta^2}{2}})|\Delta] = \Phi(-\frac{\sqrt{i}|\Delta|}{2}) + \int_{-\infty}^{\frac{i|\Delta|}{2}} \frac{1}{\sqrt{2\pi i}} e^{-\frac{z^2}{2i} + \Delta z - \frac{i\Delta^2}{2}} dz$$
$$= 2\Phi(-\frac{\sqrt{i}|\Delta|}{2}),$$

我们得出结论

$$E(W(\mathbf{p})\mathbf{1}_{\{|\Delta| \geq \Delta_0\}}) \leq 4 \sum_{i=1}^{\infty} i\Phi(-\frac{\sqrt{i}\Delta_0}{2}) < \infty. \square$$

PROOF OF LEMMA 2。令 $s_T = \sum_{i=-\infty}^{\infty} r_T(i)$、$s = \sum_{i=-\infty}^{\infty} r(i)$ 和 $\delta_T = \sum_{i=-\infty}^{\infty} |r_T(i) - r(i)|$。根据三角不等式，$|s_T - s| \leq \delta_T$。令

$$\epsilon = \min\Big(\sum_{t \leq \mathrm{med}(\mathbf{r})} r(t) - \frac{s}{2}, \frac{s}{2} - \sum_{t < \mathrm{med}(\mathbf{r})} r(t) \Big),$$

由 (7.1) 可知其为正。对于足够大的 T，使得 $\delta_T \leq \frac{\epsilon}{2}$，

$$\min\Big(\sum_{t \leq \mathrm{med}(\mathbf{r})} r_T(t) - \frac{s_T}{2}, \frac{s_T}{2} - \sum_{t < \mathrm{med}(\mathbf{r})} r_T(t) \Big) \geq \epsilon - \frac{3}{2}\delta_T > 0,$$

因此 $\mathrm{med}(\mathbf{r}_T) =$ 中位数$(\mathbf{r})$。$\square$

C 定理 2 证明中的概率界

引理 6 和引理 7 中的概率界用于证明 $\beta_T \to \beta_{\mathrm{upper}}$ 和 $\gamma_T(\Delta_0) \to \gamma_{\mathrm{scan}}(\Delta_0)$。回顾

$$A_T = \{\delta \leq |\Delta| \leq \delta^{-1},\ Y(t) = 0 \text{ for } 1 \leq |t - \tau| \leq 12\delta^{-2} \log T\},$$

其中 $\Delta = \Delta(\tau)$。

引理 6。(a) 对于 $6\delta^{-2} \log\, T \leq \ell_b < 12\delta^{-2} \log\, T$,

$$P(|Z_b(\tau)| < c_{\mathrm{scan}}, A_T) = o(\tfrac{1}{\log T}).$$

(b) 对于 $18\delta^{-2} \log\, T \leq \ell_b < 36\delta^{-2} \log\, T$,

$$P(|Z_b(\tau)| < c_{\mathrm{scan}}, A_T) = o(\tfrac{1}{T}).$$

(c)

$$\sum_{b:\ell_b\le 2\delta^2\log T}\ \sum_{t:|t-\tau|<\ell_b} P(|Z_b(t)|\ge c_{\rm scan}, A_T) = o(\tfrac{1}{\log T}).$$

(d)

$$\sum_{b:2\delta^2\log T\le\ell_b\le 12\delta^{-2}\log T}\ \sum_{t:(\log T)^{\frac14}<|t-\tau|<\ell_b} P(|Z_b(t)|\ge |Z_b(\tau)|, A_T) = o(\tfrac{1}{\log T}).$$

Proof of Lemma 6。(a) 在 A_T 条件下，对于 $\ell_b \ge 6\delta^{-2}\log T$，$Z_b(\tau) \sim N(\nu_b,1)$，其中

$$\nu_b = \sqrt{\tfrac{\ell_b}{2}}|\Delta| \ge \sqrt{3\log T},$$

因此 $P(|Z_b(\tau)| < c_{\rm scan}, A_T) \le \Phi(-\sqrt{3\log T} + c_{\rm scan}) = o(\frac{1}{\text{对数 } T})$。

(b) 在 A_T 下，对于 $\ell_b \ge 18\delta^{-2}\log T$, $Z_b(\tau) \sim N(\nu_b, 1)$，其中

$$\nu_b = \sqrt{\tfrac{\ell_b}{2}}|\Delta| \ge 3\sqrt{\log T},$$

因此 $P(|Z_b(\tau)| < c_{\rm scan}, A_T) \le \Phi(-3\sqrt{\log T} + c_{\rm scan}) = o(\frac{1}{T})$。

(c) 在 A_T 下，对于 $\ell_b \le 2\delta^2\log T$和 $\tau - \ell_b < t < \tau + \ell_b$, $Z_b(t)$ 是单位方差正态分布，其中

$$|EZ_b(t)| \le |EZ_b(\tau)| \le \sqrt{\log T},$$

因此，引理 6(c) 中概率之和为 $O((\log T)^2)\Phi(-c_{\rm scan} + \sqrt{\log T}) = o(\frac{1}{\text{对数 } T})$。

(d) 在 A_T 下，$Z_b(t) - Z_b(\tau) \sim N(-\frac{|t-\tau|}{\sqrt{2\ell_b}}\Delta, \frac{|t-\tau|}{\ell_b})$ 且 $Z_b(\tau) \sim N(\Delta\sqrt{\frac{\ell_b}{2}}, 1)$，因此引理 6(d) 中的概率之和为

$$O(\log T)\Big[\sum_{i>(\log T)^{\frac14}} \Phi\Big(-\delta\sqrt{\tfrac{i}{2}}\Big) + \Phi(-\delta^2\log T)\Big] = o(\tfrac{1}{\log T}).\ \square$$

回顾

$$\mathcal{I}_T = \{(t^*-\ell_b, t^*+\ell_b) : |t^*-\tau| \le (\log T)^{\frac14}, 2\delta^2\log T < \ell_b < 12\delta^{-2}\log T\}.$$

引理 7。

$$\sum_{I\in\mathcal{I}_T}\ \sum_{t\in I:|t-\tau|\ge(\log T)^{\frac14}} |t-\tau|P(|Z_I(t)|\ge|Z_I(\tau)|, A_T) \to 0.$$

PROOF。设 $I=(u^*,v^*)$，并回顾 $\lambda(t)=\sqrt{(v^*-t)(t-u^*)}=\sqrt{\ell_b^2-|t^*-\tau|^2}$。考虑 $\Delta>0$。在 A_T 条件下，对于满足 $t>\tau$ 的 $t\in I$，

$$\begin{aligned} E(Z_I(t)-Z_I(\tau)) &= [\tfrac{\lambda(t)}{\sqrt{2\ell_b}}(1-\tfrac{t-\tau}{t-u^*})-\tfrac{\lambda(\tau)}{\sqrt{2\ell_b}}]\Delta \\ &= -\tfrac{\lambda(\tau)\Delta}{\sqrt{2\ell_b}}\Big[1-\sqrt{\tfrac{(v^*-t)(\tau-u^*)}{(v^*-\tau)(t-u^*)}}\Big]. \end{aligned} \tag{C.1}$$

此外

$$\begin{aligned} \mathrm{Cov}(Z_I(t),Z_I(\tau)) &= \tfrac{\lambda(t)\lambda(\tau)}{2\ell_b}(\tfrac{1}{t-u^*}+\tfrac{1}{v^*-\tau}-\tfrac{t-\tau}{(t-u^*)(v^*-\tau)}) \\ &= \sqrt{\tfrac{(v^*-t)(\tau-u^*)}{(v^*-\tau)(t-u^*)}}, \end{aligned}$$

因此

$$\mathrm{Var}(Z_I(t)-Z_I(\tau))=2\Big[1-\sqrt{\tfrac{(v^*-t)(\tau-u^*)}{(v^*-\tau)(t-u^*)}}\Big]. \tag{C.2}$$

对于 $|t-\tau|\geq(\log T)^{\frac{1}{4}}$以及 $|t^*-\tau|\leq(\log T)^{\frac{1}{4}}$，

$$\begin{aligned} \sqrt{\tfrac{(v^*-t)(\tau-u^*)}{(v^*-\tau)(t-u^*)}} &= (1-\tfrac{t-\tau}{v^*-\tau})^{\frac12}(1-\tfrac{t-\tau}{t-u^*})^{\frac12} \\ &\leq (1-[1+o(1)]\tfrac{(\log T)^{\frac14}}{2\ell_b})^2 \\ &= 1-[1+o(1)]\tfrac{(\log T)^{\frac14}}{\ell_b}. \end{aligned} \tag{C.3}$$

根据 (C.1)–(C.3)，

$$\begin{aligned} \tfrac{E[Z_I(t)-Z_I(\tau)]}{\sqrt{\mathrm{Var}(Z_I(t)-Z_I(\tau))}} &= -\tfrac{\lambda(\tau)\Delta}{2\sqrt{\ell_b}}\Big[1-\sqrt{\tfrac{(v^*-t)(\tau-u^*)}{(v^*-\tau)(t-u^*)}}\Big]^{\frac12} \\ &\leq -[\Delta+o(1)]\tfrac{\lambda(\tau)(\log T)^{\frac18}}{2\ell_b} \\ &\leq -[\tfrac{\Delta}{2}+o(1)](\log T)^{\frac18}. \end{aligned} \tag{C.4}$$

对于 $\Delta<0$，计算过程类似。

因此，当 T 足够大时，

$$P(|Z_I(t)|\geq|Z_I(\tau)|,A_T)\leq\Phi(-\tfrac{\delta(\log T)^{\frac18}}{3})=o(\tfrac{1}{(\log T)^3}),$$

而引理 7 可由 $\#\mathcal{I}_T=O(\log T)$ 推出。 □

D 引理 5 的证明

设 $u \in I$，其中 $I \in \mathcal{I}_T$，且 $|u-\tau| \geq (\log\ T)^{\frac{1}{4}}$。设

$$W_I(u,\tau) = \frac{Z_I(u)-Z_I(\tau)-E[Z_I(u)-Z_I(\tau)]}{\sqrt{\mathrm{Var}(Z_I(u)-Z_I(\tau))}} (\sim \mathrm{N}(0,1)).$$

表示

$$\begin{aligned} & \tfrac{1}{2}(Z_I^2(u) - Z_I^2(\tau)) \\ = \ & \{\tfrac{E[Z_I(u)-Z_I(t)]}{\sqrt{\mathrm{Var}(Z_I(u)-Z_I(t))}} + W_I(u,\tau)\} \\ & \quad \times \tfrac{1}{2}[Z_I(u)+Z_I(\tau)]\sqrt{\mathrm{Var}(Z_I(u)-Z_I(\tau))}. \end{aligned} \tag{D.1}$$

考虑 $\Delta > 0$。由于 $I \in \mathcal{I}_T$, $|\tau - t^*| \leq (\log\ T)^{\frac{1}{4}}$ and $\ell_b > 2\delta^2 \log\ T$, 见 (7.13) 和 (7.14),

$$\lambda(\tau) \sim \sqrt{\ell_b^2 - |\tau - t^*|^2} \sim \ell_b. \tag{D.2}$$

令 $W_I(u) = Z_I(u) - E(Z_I(u))(\sim \mathrm{N}(0,1))$。根据 (D.2) 和 $E[Z_I(u)] > 0$, 对于$|\tau - t^*| \leq (\log\ T)^{\frac{1}{4}}$,

$$\begin{aligned} \tfrac{1}{2}[Z_I(u)+Z_I(\tau)] &> \tfrac{1}{2}E(Z_I(\tau)) + W_I(u) + W_I(\tau) \\ &= \tfrac{\lambda(\tau)\Delta}{2\sqrt{2\ell_b}} + W_I(u) + W_I(\tau) \\ &= [\Delta + o(1)]\sqrt{\tfrac{\ell_b}{8}} + W_I(u) + W_I(\tau). \end{aligned} \tag{D.3}$$

根据 (C.2) 和 (C.3),

$$\sqrt{\mathrm{Var}(Z_I(u) - Z_I(\tau))} \sim (\log T)^{\frac{1}{8}} \sqrt{\tfrac{2}{\ell_b}}. \tag{D.4}$$

将 (C.4)、(D.3) 和 (D.4) 代入 (D.1)，得到

$$\begin{aligned} \tfrac{1}{2}(Z_I^2(u) - Z_I^2(\tau)) \ \leq \ & -[\tfrac{\Delta^2}{4} + o(1)](\log T)^{\frac{1}{4}} \\ & + [W_I(u,\tau) - W_I(u) - W_I(\tau)][\tfrac{\Delta}{2} + o(1)](\log T)^{\frac{1}{8}}. \end{aligned}$$

对 $\Delta < 0$ 的计算类似。因此

$$P(\tfrac{1}{2}[Z_I^2(u) - Z_I^2(\tau)] \geq -\tfrac{\delta^2}{5}(\log T)^{\frac{1}{4}},\ |\Delta| \geq \delta) = o(\tfrac{1}{(\log T)^3}).$$

因此，只需证明

$$P(a(u) \geq a(\tau)\exp(\tfrac{\delta^2}{5}(\log T)^{\frac{1}{4}}),\ |\Delta| \geq \delta) = o(\tfrac{1}{(\log T)^3}). \tag{D.5}$$

事实上，由于 $a(\tau) \sim G_q^*$（其中 $dG_q^*(a) = \frac{a}{q} dG_q(a)$）且 $a(u) \sim G_q$, (D.5) 中的概率上界为

$$\begin{aligned} & q^{-1} E_q(a(\tau) \mathbf{1}_{\{a(u) \geq a(\tau) \exp(\frac{\delta^2}{5} (\log T)^{\frac{1}{4}})\}}) \\ \leq \; & q^{-1} \exp(-\tfrac{\delta^2}{5} (\log T)^{\frac{1}{4}}) E_q(a(u)) \\ = \; & \exp(-\tfrac{\delta^2}{5} (\log T)^{\frac{1}{4}}) = o(\tfrac{1}{\log T}). \end{aligned}$$

参考文献

[1] Arias-Castro, E., Donoho, D. 和 Huo X.(2005). 通过快速多尺度方法近乎最优地检测几何对象。IEEE Trans. Inf. Theory 51 2402–2425. [2] Arias-Castro, E., Donoho, D. 和 Huo X. (2006). 在均匀噪声背景下自适应多尺度检测丝状结构。Ann. Statist. 34 326–349.[3] Barry, D. 和 Hartigan, J.A.(1993). 变点模型的贝叶斯分析。J. Amer. Statist. Assoc. 88 309–319.[4] Chan, H.P. 和 Walther, G. (2015). 多样本对齐稀疏信号的最优检测。Ann. Statist. 43 1865–1895[5] Chib, S.(1998). 多重变点模型的估计与比较。J. Econometrics 86 221–241.[6] Cho, H. (2016). 通过双重累积和统计量检测面板数据中的变点。Electron. J. Statist. 10 2000–2038.[7] Cho, H. 和 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。J. Roy. Statist. Soc. B 77 475–507.[8] Du, C., Kao, C.L. 和 Kou, S.C. (2016). 通过边际似然逐步提取信号。J. Amer. Statist. Assoc. 111 314–330.[9] Enikeeva, F.和 Harchaoui, Z. (2019). 稀疏备择下的高维变点检测。Ann. Statist. 47 2051–2079.[10] Fryzlewicz, P. (2014). 用于多个变点检测的野生二分分割。Ann. Statist. 42 2243–2281.

[11] Horváth L. 和 Hušková M. (2014). 面板数据中的变点检测。J. Time Ser. Anal. 23 631–648.[12] Jeng, X. J., Cai, T. T. 和 Li, H. (2013). 稀有和常见片段变异的同步发现。Biometrika 100 157–172.
[13] Jirak, M. (2015). 高维一致变点检验。Ann. Statist. 43 2451–2483.
[14] Ko, S., Chong, T. 和 Ghosh, P. (2015). 狄利克雷过程隐马尔可夫多变点模型。Bayesian Anal. 10 275–296.[15] Lai, T.L. 和 Xing, H. (2011). 一种简单的多变点贝叶斯方法。Statist. Sinica 21 539–569.
[16] Liu, H., Gao, C. 和 Samworth, R. (2021). 稀疏高维变点检测的极小极大速率。Ann. Statist. 49 1081–1112.[17] Mei, Y.(2010). 用于监控大量数据流的高效可扩展方案。Biometrika 97 419–433.[18] Niu, Y.S. 和 Zhang, H. (2012). 用于检测DNA拷贝数变异的筛选与排序算法。Ann. Appl. Statist. 6 1306–1326.[19] Pilliat, E., Carpentier. A. 和 Verzelen, N. (2020). 高维数据的最优多变点检测。arXiv预印本 arXiv:2011.07818.[20] Siegmund, D.(1985). 《序贯分析：检验与置信区间》。Springer，纽约。[21] Wang, T. 和 Samworth, R. (2018). 通过稀疏投影进行高维变点估计。J. Roy. Statist. Soc. B 80 57–83.[22] Xie, Y. 和 Siegmund, D. (2013). 序贯多传感器变点检测。Ann. Statist. 41 670–692.[23] Yao, Y.C.(1984). 含噪离散时间阶梯函数的估计：贝叶斯方法与经验贝叶斯方法。Ann. Statist. 12 1434–1447.[24] Zhang, N.R. 和 Siegmund, D. (2007). 一种改进的贝叶斯信息准则及其在比较基因组杂交分析中的应用。Biometrics 63 22–52.

[25] Zhang, N.R., Siegmund, D., Ji, H. 和 Li, J.Z.(2010). 检测多个序列中的同时变点。Biometrika 97 631– 645.