arXiv:2302.02544v1 [math.ST] 6 Feb 2023

基于后向置信序列的序贯变化检测

Shubhanshu Shekhar *[1] 和 Aaditya Ramdas †[1,2]

[1]统计与数据科学系，卡内基梅隆大学[2]机器学习系，卡内基梅隆大学

摘要

我们提出了一种从序贯估计到序贯变点检测（SCD）的简单归约。简言之，假设我们希望检测底层分布的某个参数或泛函 θ 是否发生变点。我们证明，如果能够为 θ构造置信序列（CS），那么也能成功地对 θ进行序贯变点检测。这可以通过检查两条置信序列——一条向前构造，另一条向后构造——是否曾经不相交来实现。近年来，关于置信序列的研究迅速发展，因此本文提出的归约能够立即解决若干新旧变化检测问题。此外，我们通过反转时间构造的"后向置信序列"是全新的，也可能具有独立的研究价值。我们为误报频率和检测延迟提供了有力的非渐近保证，并通过若干问题展示了该方法在数值上的有效性。

目录

1 引言 **2**

2 预备知识 **3**

3 预热：通过前向置信序列进行变化检测 **4**

4 通过后向置信序列进行变化检测 5

5 `BCS-Detector` 的实例化 **6**

5.1 参数型均值变化检测.7 5.2 非参数化均值变化检测.8 5.3 检测 CDF 变化.9 5.4 检测两条数据流的齐性变化.9 5.5 检测有害分布偏移.10 5.6 其他变化检测任务.11

6 结论 **11**

A 补充背景 15 B 命题8 的证明 16 C 定理 13 的证明 17 D 第 5 节中的延后证明 19D.1 推论 17、推论 22、推论 23 的证明.19 D.2 命题 19 的证明.20

*shubhan2@andrew.cmu.edu
†aramdas@stat.cmu.edu

E 实验细节 **21**

F 重复序贯检验的解读 **24**

1 引言

我们研究序贯变点检测问题（SCD）：其目标是在控制虚警率不超过指定水平的同时，快速检测观测流所生成的分布是否发生变化。形式化地，设某个（可能是无限维的）指标集为 Θ，令 $\{P_\theta\}_{\theta\in\Theta}$ 表示观测空间 $\mathcal{X}$ 上的一个分布类。假设对于某个 $T \geq 1$，观测值 $\{X_t\colon 1 \leq t \leq T\}$独立同分布地取自 P_{θ_0}，且 $\theta_0 \in \Theta$；而观测值 $\{X_t\colon t > T\}$ 对于某个 $\theta_1 \in \Theta$独立同分布地取自 P_{θ_1}，且$\theta_1 \neq \theta_0$。那么，序贯变点检测问题就是要在零假设 $H_0\colon \{T = \infty\}$（即没有发生变化）和备择假设 $H_1 = \cup_{i\in\mathbb{N}}\{T = i\}$之间作出判断。

由于观测值是按顺序到达的，我们的任务是设计一个随机停止时间 τ，使其适应自然滤过 $\{\mathcal{F}_t\colon t \geq 1\}$ 其中 $\mathcal{F}_t = \sigma(X_1, \ldots, X_t)$，并在此时拒绝零假设。一个好的停止规则 τ 在零假设下（即当 $T = \infty$ 时）取较大的值，同时尽量缩短在备择假设下检测到变化所需的时间（即当 $T < \infty$ 时）。形式上，当 $T = \infty$ 时，对于给定的 $\alpha \in (0, 1],$，我们要求平均运行长度 ***(ARL)*** $\mathbb{E}_\infty[\tau]$ 的下界为 $1/\alpha$，同时确保检测延迟 $\mathbb{E}_T[(\tau - T)^+]$较小，其中 $T < \infty$。通俗地说，这意味着大约每隔 $1/\alpha$ 步就会出现一次虚警，因此读者可以将 $\alpha = 10^{-3}$ 作为粗略的参考。（我们还会简要讨论如何将至少发生一次虚警的概率控制在 α 以下，但虚警保证与检测延迟之间存在权衡；要快速检测到真实变化，就必须容忍偶尔发生的虚警。）

序贯变点检测领域的文献浩如烟海，因为这一问题出现在若干重要的现实应用中，例如质量控制（Shewhart，1930）、电力网络监测（Chen 等，2015）、基因组分析（Chen 等，2011；Shen 和 Zhang，2012）以及流行病检测（Baron 等，2004；Yu 等，2013）。该领域的一些早期研究（Shewhart，1925；Page，1954；Shiryaev，1963）假设变化前后分布具有已知的密度函数 f_0 和 f_1（相对于某个共同参考测度）。这些方法使用包含似然比的统计量，这些统计量可以高效地逐步计算，而且已证明具有良好的最优性。基于似然的方法背后的思想也已扩展到分布的（有限维）参数族情形，例如指数族；详见 Tartakovsky 等人（2014）的讨论。然而，这些参数假设往往过于严格，难以适用于许多实际应用，因为数据分布可能属于更广泛的非参数类。针对参数情形的大多数思想，尤其是基于似然的方法，都不适用于非参数情形。除后文讨论的一些例外情况外，构建非参数变点检测方案的一般原则寥寥无几。本文通过提出一种概念简单的“元算法”来解决这一问题，该算法可以将任意置信序列构造转化为一种强大的变点检测方法。因此，我们可以立即利用置信序列构造方面的最新进展，建立新的变点检测方法。

备注1。我们注意到，SCD问题通常在两种设定下进行研究，这两种设定之间的差异十分微妙。在第一种（也是更常见的）设定中，假设变化前和变化后的分布（P_{θ_0}和 P_{θ_1}）分别属于两个不同且通常彼此相距甚远的分布类别。用我们的记号来说，这等价于假设存在两个已知的不相交集合 Θ_0 和 Θ_1，使得$\theta_i \in \Theta_i$，对于 $i = 0, 1$。第二种设定是本文研究的对象，它所需的信息较少。也就是说，假设 θ_0 和 θ_1 都属于某个共同的指标集（用我们的记号表示为Θ），唯一的条件是 $\theta_0 \neq \theta_1$。因此，从某种意义上说，第一种设定“更容易”，因为可以利用关于 Θ_0 和Θ_1（其大小及彼此间的距离）的额外知识来设计合适的SCD方案。尽管已有一些研究针对第二种设定提出了SCD方法，但这些方案往往依赖于所研究问题的特定结构。本文通过提出一种用于设计第二种设定下SCD方案的一般性原则，来解决这一问题。

2 预备知识

我们策略中采用的主要技术工具是置信集的时间一致版本，称为置信序列 （CSs）；它们最早由Darling and Robbins（1967）在统计学文献中提出。下面给出一个针对我们问题调整后的定义。

定义2 （置信序列）。 假设 $X_1, X_2, \ldots$ 是从 P_θ 独立同分布抽取的，其中 $\theta \in \Theta$。那么，对于任意 $\alpha \in$ （0,1），一个水平为 $(1-\alpha)$ 的置信序列（CS），记作 $\{C_t\colon t \geq 1\}$，是一族子集 $C_t \subset \Theta$，满足 (i) C_t 是 $\sigma(X_1, \ldots, X_t)$-可测，且 (ii) $\mathbb{P}\big(\forall t \geq 1 : \theta \in C_t\big) \geq 1-\alpha$。

注3。 由于置信序列的定义具有时间一致性，我们可以将置信集 C_t 替换为较小的集合 $C_t := \cap_{s \leq t} C_s$。新的置信序列 $\{C_t : t \geq 1\}$ 由嵌套的置信集组成；也就是说，对于 $s < t$， $C_t \subset C_s$。

备注4。 定义置信序列并不要求数据是独立同分布的。上述定义可以轻松推广到独立随机变量的情形，其中 $X_t \sim P_{\vartheta_t}$，且 $\vartheta_t \in \Theta$ （见附录 A ）。不过，这要求 Θ 具备加法和标量乘法的概念（一个充分条件是 Θ 是向量空间），在需要时我们将对此作隐式假设。

现在我们引入置信集 C_t“大小”的概念，以反映不确定性的程度。

定义5（置信序列宽度）。设 Θ 赋有距离度量 d，并令 $\{C_t\colon t \geq 1\}$ 表示一个在观测值 X_1、X_2、.....上构造的置信水平为 $1-\alpha$ 的置信序列，这些观测值独立同分布地取自 P_θ。如果对于所有 $t \in \mathbb{N}$ 和 $\theta \in \Theta$，都有 $\sup_{\theta' \in C_t} d(\theta, \theta') \leq w(t, \theta, \alpha)$，则函数 $w(t, \theta, \alpha)$ 表示该置信序列的逐点宽度（界）。我们将 Θ上的一致宽度定义为 $w(t, \Theta, \alpha) = \sup_{\theta \in \Theta} w(t, \theta, \alpha)$。

正如我们将在第 5 节看到的，大多数非平凡的置信序列，其逐点宽度（通常一致宽度也是如此）都会随着观测值数量的增加而趋于 0。

例6。 考虑独立随机变量 $\{X_t\colon t \geq 1\}$，其中 $X_t \sim N(\theta, 1)$ 且对所有 $t \geq 1$ 都有 $\theta \in \mathbb{R}$。在这种情况下，参数集为 $\Theta = \mathbb{R}$。对于这一过程，我们可以如下定义置信序列 $\{C_t\colon t \geq 1\}$，其中 $\bar{X}_t = (1/t)\sum_{i=1}^t X_i$，且 $w_t = 3.4\sqrt{(\log\log(2t) + 0.72\log(10.4/\alpha))/t}$。因此，若赋予参数空间 Θ 度量 $d(\theta, \theta') = |\theta - \theta'|$，我们会发现， C_t 的一致宽度（记为 $w(t, \Theta, \alpha) = w_t$）趋于 0。

相关工作。 如前所述，现有序贯变化检测文献中有很大一部分关注参数型设定。有关详情，请参阅近期综述，例如 Veeravalli 和 Banerjee（2014）；Xie 等人（2021），以及 Tartakovsky 等人（2014）的教科书。本节将讨论一些与本文工作更为相关的非参数序贯变化检测方法的研究结果。

Shin 等人（2022）通过引入*e*-检测器，提出了一种新的变点检测框架；这些检测器通过组合一系列针对变化前分布类统一定义的 e-过程而得到。他们证明，在非常一般的条件下，使用 e-检测器的策略能够控制平均运行长度，并且还证明了他们的方法在某些情况下具有最优性（就最坏情况下的检测延迟而言）。然而，正如我们在备注1中提到的，他们的框架主要适用于变化前后分布已知属于不同类别的情形。我们在第3节和第4节中介绍的技术解决了这一问题。

Maillard（2019年b）研究了如何检测一列相互独立的单变量次高斯随机变量的均值变化，并通过为广义似然比方案（Lai 和 Xing，2010）相关的扫描统计量推导一种新的双重时间一致置信序列，提出了一种序贯变化检测方法。这一集中不等式的原始证明（Maillard，2019年b，定理 4）并不完整，作者后来在（Maillard，2019年a，第 3 章，第 4.1 节）中给出了修正版本（增加了一个 log log 项）。对于由此得到的方案，Maillard（2019年a）给出了虚警概率和检测延迟的界，并证明了该方案在某些情形下（例如高斯观测）具有最优性。与 Maillard（2019年a）不同，我们的序贯变化检测框架适用于远超单变量均值检验的广泛问题。不过，当应用于单变量次高斯观测时，我们的方案可达到相同的检测延迟界，同时还能控制平均运行长度（而非虚警概率）。

Puchkin 和 Shcherbakova（2022年）研究了序贯变点检测问题，假设变化前后分布相对于共同参考测度均存在密度函数，并提出了一种基于学习判别器来估计密度比的策略。他们证明，该策略能够将平均运行长度控制在所需水平，并且还根据 Jensen-Shannon（JS）散度以及密度函数之差的 L_2 范数，得到了检测延迟的高概率上界。与他们的方法不同，我们的框架不要求密度函数存在，而且适用于在多种度量意义下相互分离的分布，而不仅限于 JS 散度。

另一类用于序贯变化检测的非参数型方案基于核 MMD 度量。Gretton 等人（2012年）最早将其用于设计高效的非参数双样本检验。Li 等人（2019年）提出了一种序贯变化检测方案，该方案利用观测值和一组变化前数据计算块MMD统计量的一种变体（Zaremba 等人，2013年）。最近，Flynn 和 Yoo（2019年）以及 Wei 和 Xie（2022年）提出了新的序贯变化检测方案，利用线性统计量和分块 MMD 统计量定义 Page（1954年）CuSum检验的非参数对应形式。然而，这些方案在检测延迟方面缺乏有力的理论保证。此外，与 Puchkin 和 Shcherbakova（2022年）的情况类似，这些序贯变化检测方案的设计策略专用于核 MMD 度量，且显然无法将其扩展到 Kolmogorov-Smirnov 度量等其他常用度量。我们的工作旨在解决这些问题。

类似地，现有的其他大多数序贯变化检测研究都针对特定的问题设定。因此，方案的设计及其分析都与所研究问题的具体细节紧密相关。例如，针对有限字母表上分布的经验似然方法（Lau 等，2018年）、适用于多元或非欧氏数据的最近邻技术（Chen，2019年），以及针对图值数据的谱扫描统计量（Sharpnack 等，2013年）。然而，本文的目标有所不同：我们不针对某项具体任务开发一种强大的序贯变化检测方案，而是提出一个用于设计序贯变化检测方案的抽象统一模板，并可将其应用于范围广泛的新旧序贯变点检测问题。

我们的贡献。在第3节中，我们首先介绍（作为热身）一种使用单个 α 级前向置信序列的变点检测方案。我们的策略是一旦置信序列变得“不一致”就立即停止；也就是说，置信序列包含了此前曾将其排除的点。我们在命题8中证明，这一简单策略能够将虚警概率控制在 α水平，并且还得到了检测延迟的高概率上界。然而，该方案过于保守，因为其平均运行长度为无穷大；在实践中，这可能导致较大的检测延迟，尤其是在 T 较大时。

在第4节中，我们介绍主要策略：在每个时间 t 检查前向置信序列与后向置信序列（一种新概念）是否一致，并在检测到不一致时停止。在定理 13 中，我们证明该方案的平均运行长度至少为 $1/(2\alpha)$，并刻画其在一般条件下的期望检测延迟。

最后，在第 5 节中，我们通过将所提出的方案应用于五种不同的置信序列，展示了该方案的效力和通用性。定理 13 中得到的检测延迟一般界很容易转化为所有这些情形下针对具体问题的上界；我们还通过一些简单的数值模拟，在经验上验证了理论预测。

3 预备：通过前向置信序列进行变化检测

在下一节介绍一般方案之前，我们先介绍一种更简单的序贯变化检测方法，它只使用单个前向置信序列。我们将此方案称为 FCS-Detector。该方案的基本思路是：如果生成观测值 $\{X_t$ 的分布在 $t \geq 1\}$ 处发生变化，那么置信序列的交集最终将变为空集。具体而言，我们按照定义2中的介绍，为未知的 θ_0 构造一个水平为 $(1-\alpha)$ 的置信序列（CS），记作 $\{C_t:\ t \geq 1\}$。若不存在变点，则该置信序列满足 $\mathbb{P}_\infty(\forall t \in \mathbb{N}:\ \theta_0 \in C_t) \geq 1-\alpha$。然而，若在某个时间 T 出现变点，我们预计置信集 C_t 会偏离用于 $t > T$ 的置信集 C_T。经过足够多的变化后观测值后，置信序列最终将变得不一致、自相矛盾。也就是说，在某个满足 $t - T$ 足够大的时间 t，我们预计 $\cap_{s=1}^{t} C_s = \emptyset$。因此，我们将停止时间 τ 定义为观测到上述不一致的最小 t。

定义7 (FCS-Detector)。给定观测值 $X_1,\ X_2, \ldots$，我们为变化前参数 θ_0 构造一个置信序列（置信序列），记为 $\{C_t:\ t \geq 1\}$。我们在时间 $\tau := \min\{n \geq 1:\ \exists t < n,\ \ C_t \cap C_n = \emptyset\}$ 时停止。

该策略满足以下性质。

命题8。 考虑一个变点检测问题，观测值 $X_1, X_2, \ldots$ 在 $t \le T$ 时从 P_{θ_0} 独立同分布地抽取，在 $t > T$ 时从 P_{θ_1} 独立同分布地抽取，其中 T 位于 $\mathbb{N} \cup \{\infty\}$，且 $\theta_0, \theta_1 \in \Theta$。假设对于任意 $\theta \in \Theta$，我们都能构造置信序列 $\{C_t : \ t \ge 1\}$，其一致宽度为 $w(t) \equiv w(\cdot, \Theta, \alpha)$。那么，我们有以下结论：

(i) 当 $T = \infty$ 时， *FCS-Detector* 将虚警概率（*PFA*）控制在水平 α。也就是说，$\mathbb{P}_\infty(\tau < \infty) \le \alpha$。

(ii) 假设 T 足够大，从而保证 $w(T) < d(\theta_0, \theta_1)$。对于 $t > T$，定义 $\lambda_t := T/t$，以及 $\theta_t := \lambda_t \theta_0 + (1 - \lambda_t)\theta_1$。那么，以至少 $1 - \alpha$ 的概率，有 $\tau - T \le \min\left\{t - T : \ w(t) + w(T) \le d(\theta_0, \widetilde{\theta}_t)\right\}$。

备注 9. 如果 d 由 Θ上的某个范数诱导，我们可以在事件 $\mathcal{E}$ 下将检测延迟界定为 $(\tau - T)^+ \le \min\left\{t - T : \ w(t) + w(T) \le \frac{t-T}{t} d(\theta_0, \theta_1)\right\}$。在许多情况下，忽略对数因子后，我们有 $w(t) \approx 1/\sqrt{t}$。因此，上述表达式意味着，当 T较"小"时，延迟与 T呈线性关系；而当 T较大时，延迟的表现近似为 $\sqrt{T}$。这些计算的详细过程，以及延迟随 T 变化的图表，见附录 B。

备注10. 上文为得到检测延迟界所使用的关于 T 的条件是必要的，因为我们并未假设已知变化前分布（即参数 θ_0）。因此，为了检测分布变化，我们必须有足够多来自变化前分布的观测值，以便足够准确地估计 θ_0，并与变化幅度 $d(\theta_0, \theta_1)$进行比较。一个极端例子是，如果 $T = 1$，那么实际上没有任何方法能够检测到变化发生了（因为从统计上看，所有数据都可能只是独立同分布的，根本没有发生变化）。换言之， $T = 0$ 和 $T = \infty$ 在信息论意义上是等价的；要实现实际可检测性，我们必须与这些极端情况保持足够距离。

当 $T = \infty$ 时，FCS-Detector 以至少 $1 - \alpha$ 的概率继续采样而不停机。因此，其平均运行长度为无穷大。这使得它在检测变点时过于保守——当 $T < \infty$ 时，我们无法给出期望检测延迟的上界，只能刻画概率为 $1 - \alpha$ 的事件下的延迟（见附录 B 中的图 5）。接下来，我们将提出一种方案，通过一系列后向置信序列扩展前向置信序列，以解决这一问题。

4 通过后向置信序列进行变化检测

现在，我们介绍主要的序贯变化检测策略，以解决上一节讨论的较简单策略的两个缺陷。简而言之，我们策略的核心思想如下：在每一轮 $t \ge 2$ 中，我们构造常规的前向置信序列，以及一个新的后向置信序列（使用逆序观测，见定义11）。我们将此方案称为 BCS-Detector。如果发生了变点，我们预计前向和后向置信序列会分别集中在 Θ 的不同区域（即分别集中在 θ_0 和 θ_1 附近）。因此，一旦它们变得不一致，我们就停止。附录 A 中的图 4 提供了直观示意，附录F则从重复序贯检验的角度解释了我们的方案。

现在我们给出后向置信序列的定义。

定义11 （后向置信序列）。 设 $X_1, X_2, \ldots$ 是从 P_θ 中独立同分布抽取的，其中 $\theta \in \Theta$。对于任意 $n \ge 1$，若一列集合 $\{B_t^{(n)}\}_{1 \le t \le n} \subseteq \Theta$ 满足以下条件，我们就称其为后向置信序列：**(i)** $B_t^{(n)}$ 是关于 $X_t, \ldots, X_n$ 的 σ 可测的；且 **(ii)** $\mathbb{P}\left(\forall t \in [n] : \ \theta \in B_t^{(n)}\right) \ge 1 - \alpha$。

注意，对于 $n > 1$，前向置信序列 C_t 不满足第一个条件，因为 C_t 是利用 $X_1, \ldots, X_t$ 构造的，而 $B_t^{(n)}$ 只能使用 $X_t, \ldots, X_n$。不过，任意 n 处的后向置信序列都可以解释为通常的前向置信序列（定义2中引入），它是基于从 n 到 1 按逆序观察到的观测值构造的；详情见附录 A。

不失一般性地，我们假设任何置信序列都由一组嵌套的集合构成，正如注3中所述。换句话说，在给定时间 n， C_n 是 $\{C_t : \ t \in [n]\}$ 中最小的集合，而 $B_1^{(n)}$ 是 $\{B_t^{(n)} : \ t \in [n]\}$ 中最小的集合。我们称两个置信序列 $\{C_t : \ t \ge 1\}$ 和 $\{B_t^{(n)} : \ t \in [n]\}$

"在时间 n 不相交"，如果对于某个 $1 \le r, s \le n$，有 $C_r \cap B_s^{(n)} = \emptyset$。容易验证，两个嵌套的置信序列不相交，当且仅当它们的最小集合（即 C_n 和 $B_1^{(n)}$）不相交。

当 $T = \infty$ 时，在每个时间 n，两个置信序列 $\{C_t : t \ge 1\}$ 和 $\{B_t^{(n)} : t \in [n]\}$ 以至少 $1 - 2\alpha$ 的概率包含 θ_0，因此它们以相同的概率相交。这促使我们在它们首次不相交的时间 n 停止，并宣布出现变点。

定义 12 (BCS-Detector)。给定观测值 $X_1, X_2, \ldots$，假设我们构造了一个前向置信序列 $\{C_t : t \ge 1\}$，并对每个 $n \ge 1$ 构造新的后向置信序列 $\{B_t^{(n)} : t \in [n]\}$。假设所有构造出的置信序列均为嵌套的。于是，我们将停止时间 τ 定义为前向和后向置信序列首次不相交的时间：$\tau := \inf\{n \ge 1 : C_n \cap B_1^{(n)} = \emptyset\}$。

现在，我们给出本节的主要结果。

定理 13。考虑一个序贯变化检测问题，观测值为 $X_1, X_2, \ldots$，在 $t \le T$ 时从 P_{θ_0} 独立同分布地抽取，在 $t > T$ 时从 P_{θ_1} 抽取，其中 T 取值于 $\mathbb{N} \cup \{\infty\}$，且 $\theta_0, \theta_1 \in \Theta$。假设对于任意 $\theta \in \Theta$，我们可以构造置信序列 $\{C_t : t \ge 1\}$，其逐点宽度为 $w(\cdot, \theta, \alpha)$。那么，我们有如下结论：*(i)* 当不存在变点（$T = \infty$）时，*BCS-Detector* 满足 $\mathbb{E}_\infty[\tau] \ge \frac{1}{2\alpha} - \frac{3}{2}$。

(ii) 假设 $T < \infty$，且变化前参数 θ_0 未知。引入事件 $\mathcal{E} = \{\theta_0 \in C_t : 1 \le t < T\}$，并注意到，根据构造，$\mathbb{P}_T(\mathcal{E}) \ge 1 - \alpha$。于是，对于 $\alpha \in (0, 0.5)$，有 $\mathbb{E}_T[(\tau - T)^+ | \mathcal{E}] \le 3 w(\theta_0, \theta_1, T)/(1 - \alpha)$，其中 $u_0 := \min\{t \ge 1 : w(t, \theta_1, \alpha) + w(T, \theta_0, \alpha) < d(\theta_1, \theta_0)\}$。

回顾定义 5，$w(t, \theta_1, \alpha)$ 表示当真实参数为 θ_1 时，具有 t 个观测值的置信水平为 α 的置信集的宽度。该定理的证明见附录 C。在许多问题实例中，变化前参数 θ_0 是已知的，因为它代表所观测过程的"自然状态"。因此，我们可以将上述结果专门应用于这种情况，得到如下结论。

推论14。假设变化前参数 θ_0 已知，且 $T < \infty$。那么，我们有 $\mathbb{E}_T[(\tau - T)^+] \le (3/(1 - \alpha)) t_0(\theta_0, \theta_1, T)$，其中 $t_0 \equiv t_0(\theta_0, \theta_1) :=$ 最小值$\{t \ge 1 : \ w(t, \theta_1, \alpha) < d(\theta_1, \theta_0)\}$。

注释15。这些结果表明，通过在设计策略中谨慎地融入后向置信序列的思想，可以克服 FCS-Detector（第3节中介绍）的缺点。具体而言，这种称为 BCS-Detector 的新方案较为不保守，并且在零假设下具有有限的平均运行长度下界。更重要的是，当 $T < \infty$ 时，BCS-Detector 的期望检测延迟也是有限的，而且与变点 T 的取值无关。这与 FCS-Detector 的检测延迟（高概率）界形成对比；对于后者，随着变点 $T \to \infty$，检测延迟近似按 $\sqrt{T}$ 增长。

注释16（其他估计值）。我们还可以将 T定义为变点的一个估计值，即 C_t和 $B_t^{(\tau)}$距离最远时所对应的时间：$\widehat{T} \equiv \widehat{T}(\tau) := \max \operatorname{argmax}_{1 \le t \le \tau} d\big(C_t, B_t^{(\tau)}\big) = \max\{s \le \tau : d(C_s, B_s^{(\tau)}) = \max_{1 \le t \le \tau} d(C_t, B_t^{(\tau)})\}$。此外，我们将变化幅度的估计值定义为 $\epsilon := d(\theta_0, \theta_1)$，即 $C_{\widehat{T}}$和 $B_{\widehat{T}}^{(\tau)}$之间的距离，并用距离度量 d:$\widehat{\epsilon} \equiv \widehat{\epsilon}(\tau) := \max_{\theta \in C_{\widehat{T}}, \theta' \in B_{\widehat{T}}^{(\tau)}} d(\theta, \theta')$来衡量。尽管我们没有获得理论保证，下一节的一些经验结果表明，对于 BCS-Detector 的若干种实例化，这些估计值是准确的。

BCS-Detector的 5 种实例化

在上一节中，我们介绍了一种概念简单的方法，可以将任意置信序列构造转化为强大的序贯变点检测器。借助置信序列构造方面的最新进展，我们可以将通用的变点检测元算法应用于各种情形。本节将通过介绍多种参数型和非参数型序贯变点检测问题来说明这一点。

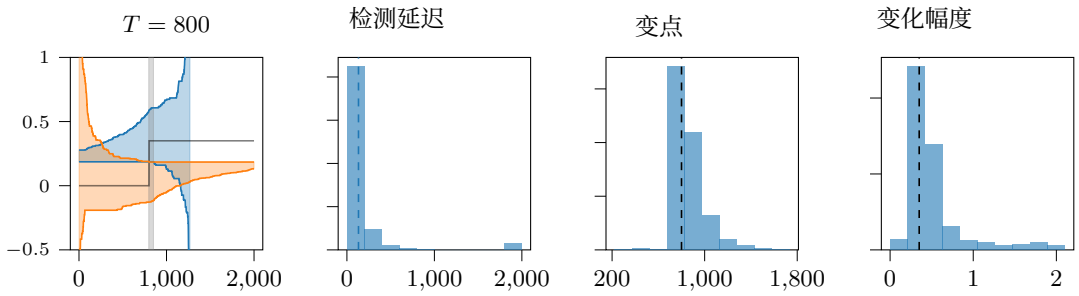

图1：图中展示了我们的变点检测方案 BCS-Detector 的性能。对于单变量高斯观测，其均值在时间 $T = 800$ 从 0 变为 0.4。第一幅图展示了其中一次试验在检测时刻（$\tau = 1264$）的前向和后向置信序列，灰色阴影区域表示两个置信序列不一致的点。接下来的三幅图展示了在 250 次重复试验中，检测延迟、估计的变点位置和估计的变点幅度的经验分布。

5.1 参数化均值变化检测

我们首先考虑单变量高斯均值变点检测这一简单情形（不过，同样的结果也适用于更广泛的次高斯随机变量类别）。在这个问题中，我们观测到$\{X_t : \ t \geq 1\}$，它们是从分布 $N(\mu_t, 1)$中独立同分布抽取的，其中 $\mu_t = \theta_0$对于 $t \leq T$成立，而 $\mu_t = \theta_1$对于$t > T$成立。注意，在这个问题中，我们有 $\Theta = \mathbb{R}$，并且可以将距离度量 d设为差的绝对值。我们将使用 Howard 等（2021）最近推导出的高斯均值置信序列，该序列此前已在例6中介绍。此外，对于任意时间 n，我们都可以使用同一表达式构造后向置信序列，记作 $\{B_t^{(n)} : \ t \geq 1\}$，但需按第4节所述反转观测值的顺序。以下结果表明，在这种参数型设定下，我们的变点检测方案能够达到阶最优检测延迟（即除多对数因子外最优）：

推论17。 假设 $\{X_t : \ t \geq 1\}$独立同分布地取自 $P_{\theta_0} = N(\theta_0, 1)$，其中 $t \leq T$；而$P_{\theta_1} = N(\theta_1, 1)$，其中 $t > T$。注意，在这种情况下，我们有 $d_{KL}(P_{\theta_1}, P_{\theta_0}) = (\theta_1 - \theta_0)^2/2$。那么，如果$T = \Omega(\log(1/d_{KL}(P_{\theta_1}, P_{\theta_0}))/d_{KL}(P_{\theta_1}, P_{\theta_0}))$，则有

$$\mathbb{E}_T[(\tau - T)^+ | \mathcal{E}] = \mathcal{O}\left(\frac{\log\log(\frac{1}{d_{KL}(P_{\theta_1}, P_{\theta_0})}) + \log(\frac{1}{\alpha})}{d_{KL}(P_{\theta_1}, P_{\theta_0})}\right),$$

其中 $\mathcal{E}$ 是定理 *13* 中的（$1 - \alpha$）概率事件。

注 18. 如果已知变点前均值（θ_0），那么上述上界对最坏情况下的检测延迟也成立，无需进行条件化。该延迟定义为 $J_L(\tau) := \sup_{T>0} \operatorname{esssup} \mathbb{E}[(\tau - T)^+ | \mathcal{F}_T]$。在假设 θ_1 也已知的情况下，Lorden（1971）证明了该量的以下普适下界：$\inf_{\tau'} J_L(\tau') = \log(1/\alpha)\, d_{\mathrm{KL}}(P_{\theta_1}, P_{\theta_0})\,(1 + o(1))$，其中 $\alpha \to 0$。因此，即使不知道变点后参数，BCS-Detector 仍能达到这一最优性能（相差对数因子）。这不同于一些现有的高斯均值变化检测方案，例如 Pollak 和 Siegmund（1991）：这些方案虽然能达到同阶最优的检测延迟，但需要额外假设（已知变化的下界，且在 $T \to \infty$ 的极限下）。

实证验证。 现在，我们使用来自单位方差正态分布的观测值来验证所提出的变点检测方案的理论结论，其中变化前的均值为 $\theta_0 = 0$，变化后的均值为 $\theta_1 = \Delta$。图1展示了变化发生在 $T = 800$ 时 $\Delta = 0.4$ 的情况。图中显示了某次试验中检测到变化时的前向和后向置信序列，以及250次试验中检测延迟、估计变点和估计变化幅度的分布。图中结果表明，BCS-Detector 能够快速、准确地检测变化。

在图 2 中，我们研究了变化幅度 Δ变化时，我们的变点检测方案的平均检测延迟如何变化。经验结果验证了理论结果所声称的平均检测延迟与 $1/\Delta^2$ 成正比。

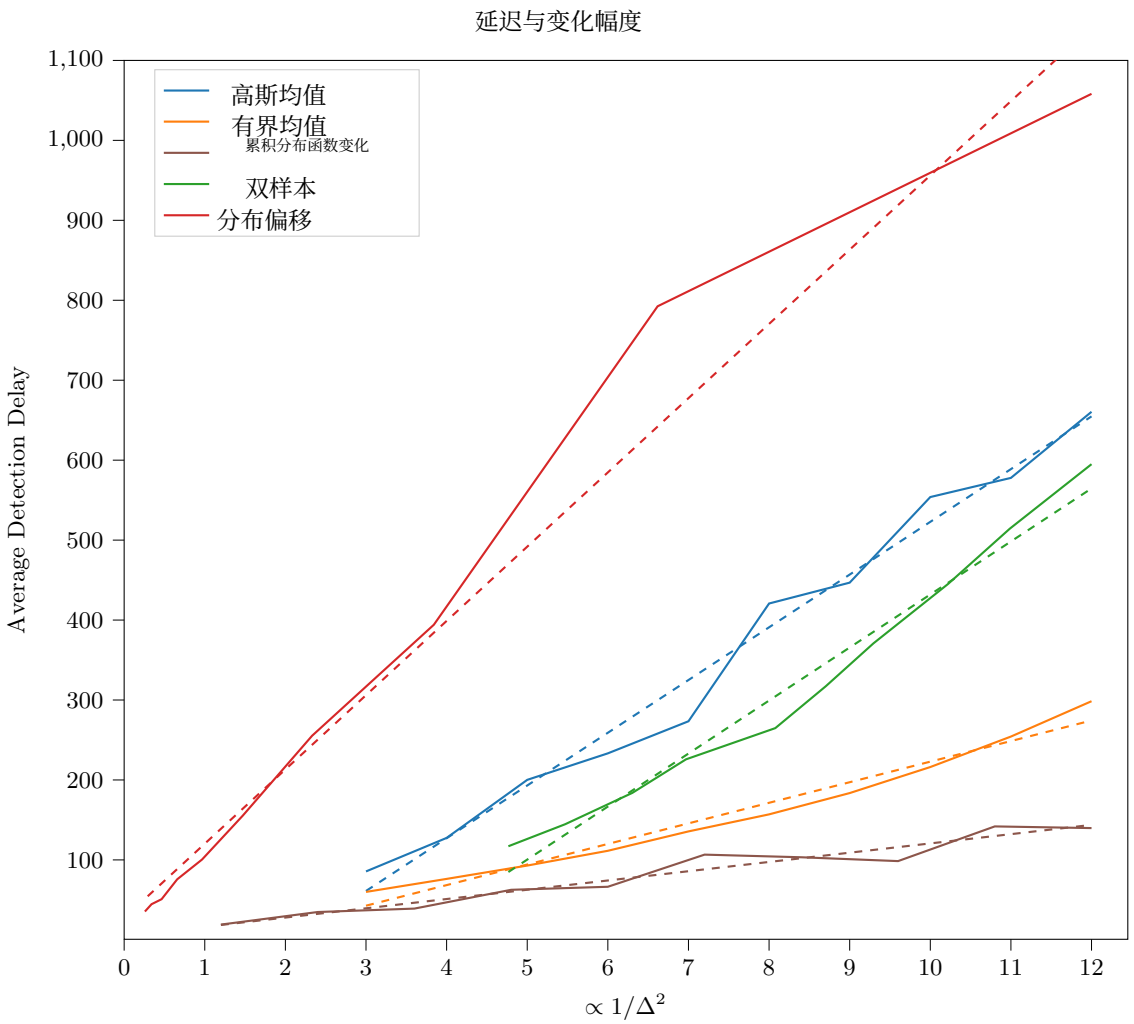

图 2：在第 5 节中 BCS-Detector 的所有实例中，a$\sqrt{}$处置信序列的宽度为 (log log(t) $+$log($1/\alpha$)$/t$ rate。因此，根据定理 13，我们预计检测延迟与变化幅度 (Δ) 呈反二次依赖。上图通过实证验证了这一结论。

5.2 非参数型均值变化检测

现在，我们考虑上一节均值变化检测问题的非参数型对应问题。这里假设 $X_1, X_2, \ldots$ 是取值于有界区间 $\mathcal{X} \subset \mathbb{R}$ 内的独立随机变量；不失一般性，我们令该区间为 $[0,1]$。在变点 T 之前，假设观测值的均值为 $\theta_0 \in \Theta = [0,1]$；而在时间 T 之后，均值变为 $\theta_1 \neq \theta_0$。在这种情况下，我们可以使用 Waudby-Smith 和 Ramdas（2023年）提出的经验 Bernstein（EB）置信区间。为给出 EB 置信序列的闭式表达式，我们首先需要引入以下术语：$\widehat{\mu}_t = \frac{\frac{1}{2}+\sum_{i=1}^{t} X_i}{t+1}$, $\widehat{\sigma}_t^2 = \frac{\frac{1}{4}+\sum_{i=1}^{t}(X_i-\widehat{\mu}_t)^2}{t+1}$, $\lambda_t = \sqrt{\frac{2\log(2/\alpha)}{\widehat{\sigma}_t^2 t \log(t+1)}} \wedge \frac{1}{2}$, $\widehat{\theta}_t = \frac{\sum_{i=1}^{t}\lambda_i X_i}{\sum_{i=1}^{t}\lambda_i}$, $v_t = (X_t-\widehat{\mu}_{t-1})^2$, 以及 $\Psi_E(x) = \log(1-x) - x$，对所有 $x \in [0,1$ 均成立。借助这些术语，我们现在可以将 Waudby-Smith 和 Ramdas（2023年）推导出的 EB-CS 表述如下：$C_t = [\widehat{\theta}_t + w_t/2, \widehat{\theta}_t - w_t/2]$，其中 $w_t = \frac{\log(2/\alpha)+\sum_{i=1}^{t} v_i \Psi_E(\lambda_i)}{\sum_{i=1}^{t}\lambda_i}$。同样，应用一般结果定理 13，可得使用上述 EB-CS 的序贯变化检测方案的期望检测延迟上界。

命题 19. 假设 $X_1, X_2, \ldots, X_T$ 是从 $\mathcal{X} = [0,1]$ 上的某个分布中独立同分布抽取的，其均值为 $\theta_0 \in \Theta = [0,1]$；而 $X_{T+1}, \ldots$ 是从 P_{θ_1} 中抽取的，其均值为 $\theta_1 \neq \theta_0$。那么，假设已知 θ_0，此实例

BCS-Detector的 (定义 12) 满足以下条件（其中 $\Delta := |\theta_0 - \theta_1|$，且 $\sigma_1^2 = \mathbb{E}_{P_{\theta_1}}[(X-\theta_1)^2]$）：

$$\mathbb{E}_T[(\tau - T)^+] = \mathcal{O}\left(\sigma_1^2 \frac{\log(1/\Delta) + \log(1/\alpha)}{\Delta^2}\right).$$

注释 20。 尽管上述结果是在变点前和变点后的观测值分别独立同分布于 P_{θ_0} 和 P_{θ_1} 的假设下得出的，但我们注意到，当随机变量仅相互独立且具有固定的变点前后均值时，也能得到类似结果。此外，也可以不再要求已知 θ_0，代价是需要对“良好”事件 $\mathcal{E}$进行条件化。

注释 21。 在本小节中，我们考察了均值变化检测这一任务，其所处的非参数设置或许是最简单的（但并非平凡）。不过，这里提出的相同思路也可以轻松扩展到其他有趣的情形，例如利用 Howard 等（2021）推导的置信序列处理次高斯族，或处理重尾分布（Wang 和 Ramdas，2022）。

实证验证。 为验证我们的理论结论，我们考虑观测值取值于 $\mathcal{X} = [0,1]$的独立观测流，其变点前后参数分别为 θ_0 和 θ_1。我们按照附录 E 中的描述，通过对均匀分布进行适当混合，定义具有指定均值的分布。对于固定的问题实例（θ_0，θ_1，T）=（0.4，0.6，800），我们在图 6（附录 E）中绘制了变点检测方案的性能。图 2 验证了检测延迟呈预测的反二次依赖关系。

5.3 检测累积分布函数的变化

继续讨论实值观测（或更一般地，全序空间上的观测值），现在我们考虑一个更一般的问题：检测生成观测值的分布是否发生了变化。由于实值随机变量完全由其累积分布函数（CDF）刻画，因此可以将此任务表述为检测 CDF 是否发生变化。更正式地说，假设我们得到一个观测流，X_1，X_2，……，这些观测值按照分布 $\theta_0 = F_0$ for $t < T$生成，并按照分布 $\theta_1 = F_1$ for $t \geq T$生成。因此，在这种情况下，Θ是所有定义在 $\mathbb{R}$ 上的可行 CDF 组成的无限维空间，我们赋予它 Kolmogorov-Smirnov （KS）度量 d_{KS}，其定义为 $d_{\mathrm{KS}}(F, G) = \sup_{x\in\mathbb{R}} |F(x) - G(x)|$。

为实例化我们的 SCD 方案，我们将采用 Howard 和 Ramdas（2022）最近推导出的、基于 KS 度量的 CDF 的 level-α 置信序列 $\widehat{C}_t = \{\theta \in \Theta : d_{\mathrm{KS}}(\theta, \theta_t) \leq w_t/2\}$，其中 $w_t = 1.7\sqrt{\log\log(et) + 0.8\log(1612/\alpha)}/t$。根据定理 13 的一般结果，我们可以得到该方案的如下性能保证。

推论 22. 假设对于某个 $T < \infty$，观测值 X_1、……、X_T 来自一个已知分布F_0，而对于 $t \geq 1$，观测值 X_{T+1}、X_{T+2}、……来自一个未知的 F_1。那么，采用上述置信序列实例化的 *SCD* 方案满足（其中 $\Delta := d_{KS}(F_1, F_0)$）：

$$\mathbb{E}_T[(\tau - T)^+] = \mathcal{O}\left(\frac{\log\log(1/\Delta) + \log(1/\alpha)}{\Delta^2}\right).$$

实证验证。 我们测试了所提出方案在自由度为 3 的 t分布上的性能。在附录 E 的图 7 中，我们展示了所提出方案在一个固定问题实例中的性能，其中变更前和变更后的累积分布函数满足 $\Delta = d_{\mathrm{KS}}(F_0, F_1) \approx 0.4$。图 2 绘出了平均检测延迟随 Δ 取值变化的情况，呈现出预期的反二次依赖。

5.4 检测两条数据流的齐次性变化

假设我们有一个位于乘积空间 $\mathcal{X} = \mathcal{U} \times \mathcal{U}$ 中的观测流，且参数集由 $\mathcal{U} \times \mathcal{U}$ 上的所有乘积分布组成；即，$\Theta = \{P \times Q : P, Q \in \mathcal{P}(\mathcal{U})\}$。在变点之前，我们假设观测值 X_1，X_2，……，X_T 满足 $X_t = (U_t, V_t) \in \mathcal{U} \times \mathcal{U}$，并且来自 $\theta_0 = P_U \times P_V$；而变化后观测值 X_{T+1}，X_{T+2}，……则假设来自某个其他乘积

分布 $Q_U \times Q_V$。给定某种统计距离度量 ρ： $\mathcal{P}(\mathcal{U}) \times \mathcal{P}(\mathcal{U}) \to \mathbb{R}$，我们假设 $\rho(P_U, P_V) \neq \rho(Q_U, Q_V)$。受双样本检验问题启发，该问题的一个有趣特例是 $P_U = P_V$，且 $Q_U \neq Q_V$。

若 ρ 是概率度量，我们可以据此在参数空间 Θ 上定义距离度量 d，如下所示： $d(\theta_0, \theta_1) = |\rho(P_U, P_V) - \rho(Q_U, Q_V)$，其中 $\theta_0 = P_U \times P_V$和 $\theta_1 = Q_U \times Q_V$。接下来，为了得到变点检测方案，我们可以对统计距离 ρ使用置信序列。我们以核 k对应的核 MMD 度量（定义见附录 A）来具体实现这一策略，记为 $d_{\mathrm{MMD}}(\cdot, \cdot)$。对于两个分布之间的核 MMD 距离，Manole 和 Ramdas（2021）推导出的置信序列如下（假设 $\sup_{u,u'} k(u, u') \leq 1$）： $C_t = \{(P, Q): d_{\mathrm{MMD}}(P_t, Q_t) - \gamma_t \leq d_{\mathrm{MMD}}(P, Q) \leq d_{\mathrm{MMD}}(\widehat{P}_t, \widehat{Q}_t) + 2\kappa_t\}$ 其中 $\kappa_t = \sqrt{\left(\log((1 \vee \log_2(t))^2 \pi^2/6) + \log(4/\alpha)\right)/t}$，且 $\gamma_t = (4\sqrt{2}/\sqrt{t})\left(1 + \sqrt{\log\left(3.54e(1 \vee \log_2 t)^3\right)} + \sqrt{\log(2/\alpha)}\right)$. 对于这一具体实现，定理 13 推出以下结论。

推论 23. 令 X_t 表示 $\mathcal{X} \times \mathcal{X}$ 上的配对(U_t, V_t)。假设 $X_1, \ldots, X_T$独立同分布地取自$\theta_0 = P_U \times P_V$，而 $X_{T+1}, X_{T+2}, \ldots$ 独立同分布地取自分布 $\theta_1 = Q_U \times Q_V$。那么，令 $\Delta > 0$表示 $d(\theta_0, \theta_1) = |d_{MMD}(P_U, P_V) - d_{MMD}(Q_U, Q_V)|$，我们可以得到基于上述置信序列的 ***BCS-detector*** 的期望检测延迟的以下上界：

$$\mathbb{E}\left[(\tau - T)^+ | \mathcal{E}\right] = \mathcal{O}\left(\frac{\log(1/\alpha) + \log\log(1/\Delta)}{\Delta^2}\right),$$

其中，$\mathcal{E}$表示与前向置信序列 $\{C_t : t \geq 1\}$相关联的“良好”事件 $\{d_{MMD}(P_U, P_V) \in C_t : t \leq T - 1\}$。

备注24. 考虑前面提到的特殊情形，其中 $P_U = P_V = Q_U = P$ 对于某个分布P成立，而 $Q_Y = Q \neq P$ 对于另一个分布成立。此外，假设已知在变点 $P_U = P_V$ 之前（也就是说，事件 $\mathcal{E}$ 是一个概率为 1 的事件）。那么，在这种情形下，上述结果给出了期望检测延迟 $\mathcal{O}\left(\frac{\log(1/\alpha) + \log\log(1/\Delta)}{\Delta^2}\right)$的上界，其中 $\Delta = d_{\mathrm{MMD}}(P, Q)$。这与现有结果相符，例如 Wei 和 Xie（2022年）的核 CuSum 方案。

备注25。 我们主要关注核 MMD 度量，原因在于其具有普适性（它适用于定义了正定核的任意空间上的分布）。然而，通过采用 Manole 和 Ramdas（2021）基于反向次鞅的置信序列构造方法，同样的思想也适用于关于其参数具有凸性的任何统计距离度量。这一类度量包括所有常用的统计距离，例如 Wasserstein 度量、 f-散度以及一般的积分概率度量。

备注26。 我们的方案总体计算成本为 $\mathcal{O}(\tau^3)$，因为我们的方案需要在每一轮构造一个新的后向 CS，其成本为 $\mathcal{O}(t^2)$。在实践中，可以通过使用线性统计量或分块 MMD 统计量，和/或降低构造新的后向 CS 的频率（而不是每轮都进行构造），来降低这一复杂度。

实证验证。 我们研究了方案在 $p = 5$ 维成对多元高斯观测流上的性能。变点前分布 P_U 和 P_V 均具有零均值和单位协方差；而对于变化后分布，我们有 $Q_U = P_U$，且 Q_Y 的均值为 $\delta\mathbf{1}$，协方差矩阵为取值随机的对角协方差矩阵。在附录 E 的图 8 中，我们绘制了固定问题实例下变点检测方案的性能，其中 $\Delta = d_{\mathrm{MMD}}(Q_U, Q_V) \approx 0.33$，且为 $T = 800$；同时，平均检测延迟关于 Δ 的反二次依赖关系在图 2 中得到了验证。

5.5 检测有害分布偏移

作为最后一个应用，我们来考虑检测机器学习（ML）模型训练分布与测试分布之间“有害”变化这一任务。遵循 Podkopaev 和 Ramdas（2021）的做法，我们关注的是检测

那些会导致训练好的 ML 模型风险（即预期损失）显著增加的分布变化。

形式上，假设一个记作 h 的机器学习模型是在从某个源分布 P_S 中独立同分布抽取的数据集上训练的，该分布取值于某个空间 $\mathcal{X}$。对于某个有界损失函数 ϕ，令 θ_0 表示该模型的期望训练损失；$\theta_0 = \mathbb{E}_{P_S}[\phi(X, h)]$。接下来，假设该模型 h 部署在由测试数据构成的数据流上，记为 X_1，X_2，.....；对于 $t < T$，这些数据来自源（或训练）分布P_S，而对于 $P_T \neq P_S$，则来自其他某个分布 $P_T \neq P_S$。我们的目标是检测出对训练好的模型有“危害”的变点后分布 P_T；也就是说，这些分布会导致期望损失增加：$\theta_1 := \mathbb{E}_{P_T}[\phi(h, X)] > \theta_0$（见附录 E 中的图 10）。

对于有界损失函数 ϕ，这个问题属于第5.2节的非参数化均值变化检测框架。由于我们只关注单侧变化，因此可以调整第5.2节中的策略，在正向过程中仅使用上置信序列，在反向过程中使用下置信序列。与命题 19 中的结果类似，我们可以证明，对于这种策略，该方案的期望检测延迟与目标分布的“危害性”（即差距 $\theta_1 - \theta_0$）成反比。

实证验证。为了说明上述讨论的思想，我们考虑一个简单的二元分类问题，其中使用线性分类器和二维特征（详情见附录 E）。我们在附录 E 的图 9 中展示了方案在一个特定问题上的性能（$\Delta \approx 0.16$），并在图 2 中验证了平均检测延迟对 Δ 的反二次依赖关系。

5.6 其他变化检测任务

本节通过将 BCS-Detector 策略应用于五种不同场景，展示了该策略的通用性。为简单起见，我们主要关注单变量观测（第 5.4 节除外）。不过，我们注意到，先前应用中使用的相同思路也很容易推广到更一般的观测情形，或满足额外的鲁棒性或隐私约束。下面列举一些示例，不再详述其分析或实际实现：

（i）指数族。在这种情况下，观测值 X_1，X_2，.....属于 $\mathcal{X} = \mathbb{R}^p$，且变化前后的分布（$P_{\theta_0}$ 和 P_{θ_1}）取自某个有限维指数族，其中 $\Theta = \mathbb{R}^m$，对某个 $m < \infty$ 成立。这里，我们可以使用 BCS-Detector，并采用 Chowdhury等人（2022）推导的置信序列。

（ii）协方差矩阵。同样，我们假设 $\mathcal{X} = \mathbb{R}^p$，但现在假设在变点 T 处，观测值的协方差矩阵从 $\theta_0 \in \mathbb{R}^{p \times p}$ 变为某个 $\theta_1 \neq \theta_0$。对于这一问题，我们可以使用 Howard 等人（2021 年，第 4.3 节）推导的协方差矩阵置信序列来实例化 BCS-Detector。

（iii）非参数回归。设 U_1，U_2，..... 表示来自 $\mathcal{U} = [0,1]^p$ 的独立同分布的均匀抽样，其中某个 $p \geq 1$。令 Θ 表示一个从 $\mathcal{U}$ 到 $\mathbb{R}$ 的函数再生核希尔伯特空间（核函数为 k）。于是，对于任意 $\theta \in \Theta$，定义随机变量 $Y_t \equiv Y_t(\theta) = \theta(U_t) + \eta_t$，其中 $\{\eta_t : t \geq 1\}$ 是一个独立同分布的 1-次次高斯噪声序列。显然，(U_t, Y_t) 的联合分布由 θ 参数化。考虑序贯变点检测问题，其中在变点 T 之前 $\theta = \theta_0$，之后为 $\theta = \theta_1$，且 $\|\theta_0 - \theta_1\|_k > 0$。我们的 BCS-Detector 策略很容易应用于这一情形，只需使用 Chowdhury 和 Gopalan（2017）构造的置信序列，并采用无限维索引集 Θ。

(iv) 稳健序贯变化检测。序贯变点检测问题的一个有趣变体，是在对抗性污染下检测变点（Li 和 Yu，2021年）。借助近期的稳健置信序列构造（例如 Wang 和 Ramdas（2023年）提出的方法），我们的 BCS-Detector 策略可以轻松扩展到此类场景。

(v) 隐私序贯变点检测。隐私在许多应用中都是一项重要考量，尤其是在涉及个人数据时；通常的做法是仅向分析人员披露实际数据的随机化版本。这为常规的序贯变化检测任务（Cummings 等人，2018年）增添了另一层复杂性。不过，我们的 BCS-Detector 框架可以轻松应对这一问题，并借助隐私置信序列构造方面的最新进展（Waudby-Smith 等人，2022年）实现这一点。

6 结论

我们提出了一种通用策略（BCS-Detector），通过精心结合置信序列（CSs）与后向置信序列——这是本文提出的一种新型置信序列变体——来设计序贯变点检测（SCD）方案。在对置信序列提出非常宽松且自然的要求下，我们证明了 BCS-Detector 能够严格控制平均运行长度和检测延迟。借助近期在构造置信序列方面取得的进展

，我们针对范围广泛的序贯变化检测问题（包括参数型和非参数型）实例化了这一策略，并通过一些小规模数值实验验证了理论结论。我们的工作为未来研究开辟了多个方向：

(i)每轮都构造一个新的后向置信序列，计算成本可能很高，而且在大多数情况下会导致总体复杂度达到二次（甚至三次）。一个值得探索的方向是研究能否通过减少更新后向置信序列的次数，实现同样的性能。
(ii) 在备注16中，我们定义了变点的估计量（T）和变化幅度（Δ）；如图1和附录 E所示，这些估计量在实证中表现良好。为它们建立理论保证，是未来研究的另一个有趣问题。

参考文献

M. Baron、V. Antonov、C. Huber、M. Nikulin 和 V. Polischook。《将疫情早期检测作为序贯变点问题》。可靠性、公共卫生、医学和生物学中的寿命、老化与退化模型，*LAD*，第7–9页，2004年。H. Chen。《基于近邻的序贯变点检测》。*The Annals of Statistics*，47(3)：1381–1407，2019年。J. Chen、A. Yi˘giter 和 K.-C. Chang。《贝叶斯方法：推断变点模型及其在DNA拷贝数实验数据中的应用》。《应用统计学杂志》，38(9)：1899–1913，2011年。Y. C. Chen、T. Banerjee、A. D. Dominguez-Garcia 和 V. V. Veeravalli。《线路停运的快速检测与识别》。《*IEEE*电力系统汇刊》，31(1)：749–758，2015年。S. R. Chowdhury 和 A. Gopalan。《论核化多臂老虎机》。载于国际机器学习会议，第844–853页。PMLR，2017年。S. R. Chowdhury、P. Saux、O.-A. Maillard 和 A. Gopalan。《一般指数族的布雷格曼偏差》。*arXiv*预印本 *arXiv:2201.07306*，2022年。R. Cummings、S. Krehbiel、Y. Mei、R. Tuo 和 W. Zhang。《差分隐私变点检测》。《神经信息处理系统进展》，31，2018年。D. A. Darling 和 H. Robbins。《均值、方差和中位数的置信序列》。《美国国家科学院院刊》，58(1)：66–68，1967年。T. Flynn 和 S. Yoo。《使用核累积和算法进行变化检测》。载于*2019*年*IEEE*第*58*届决策与控制会议（*CDC*），第6092–6099页。IEEE，2019年。A. Gretton、K. M. Borgwardt、M. J. Rasch、B. Sch¨olkopf 和 A. Smola。《核两样本检验》。《机器学习研究杂志》，13(1)：723–773，2012年。S. R. Howard 和 A. Ramdas。《分位数的序贯估计及其在A/B测试和最佳臂识别中的应用》。《伯努利》，28(3)：1704–1728，2022年。S. R. Howard、A. Ramdas、J. McAuliffe 和 J. Sekhon。《时间一致的非参数非渐近置信序列》。*The Annals of Statistics*，49(2)：1055–1080，2021年。T. L. Lai 和 H. Xing。《当变化前后参数未知时的序贯变点检测》。*Sequential Analysis*，29(2)：162–175，2010年。T. S. Lau、W. P. Tay 和 V. V. Veeravalli。《利用分箱方法对未知变化后分布进行最快变化检测》。《*IEEE*信号处理汇刊》，67(3)：609–621，2018年。M. Li 和 Y. Yu。《对抗鲁棒变点检测》。《神经信息处理系统进展》，34：22955–22967，2021年。S. Li、Y. Xie、H. Dai 和 L. Song。《用于核变点检测的扫描B统计量》。*Sequential Analysis*，38(4)：503–544，2019年。G. Lorden。《对分布变化作出响应的程序》。《数理统计年刊》，第1897–1908页，1971年。O.-A. Maillard。《统计序贯决策的数学》。2019a。O.-A. Maillard。《序贯变点检测：扫描统计量的拉普拉斯集中与非渐近延迟界》。载于算法学习理论，第610–632页。PMLR，2019b。T. Manole 和 A. Ramdas。《用于凸泛函及散度序贯估计的鞅方法》。*arXiv*预印本 *arXiv:2103.09267*，2021年。

F. Orabona. 在线学习的现代导论。*arXiv*预印本 *arXiv:1912.13213*，2019年。E. S. Page. 连续检验方案。*Biometrika*，41(1/2):100–115，1954年。A. Podkopaev 和 A. Ramdas. 跟踪已部署模型的风险并检测有害分布偏移。载于国际学习表征会议，2021年。M. Pollak 和 D. Siegmund. 初始值未知时正态均值变化的序贯检测。*The Annals of Statistics*，19(1):394–416，1991年。N. Puchkin 和 V. Shcherbakova. 在线变点检测的对比方法。*arXiv*预印本 *arXiv:2206.10143*，2022年。J. Sharpnack、A. Singh 和 A. Rinaldo. 基于谱扫描统计量的图上变点检测。载于人工智能与统计学，第545–553页。PMLR，2013年。J. J. Shen 和 N. R. Zhang. 非齐次泊松过程上的变点模型及其在新一代DNA测序拷贝数分析中的应用。*The Annals of Applied Statistics*，6(2):476–496，2012年。W. A. Shewhart. 将统计学应用于辅助维持制成品质量。*Journal of the American Statistical Association*，20(152):546–548，1925年。W. A. Shewhart. 制成品的经济质量控制。*Bell System Technical Journal*，9(2):364–389，1930年。J. Shin、A. Ramdas 和 A. Rinaldo. E-检测器：在线变点检测的非参数框架。*arXiv*预印本 *arXiv:2203.03532*，2022年。A. N. Shiryaev. 关于最快检测问题中的最优方法。*Theory of Probability & Its Applica-tions*，8(1):22–46，1963年。A. Tartakovsky、I. Nikiforov 和 M. Basseville. 序贯分析：假设检验与变点检测。CRC Press，2014年。V. V. Veeravalli 和 T. Banerjee. 最快变化检测。载于*Academic press library in signal processing*，第3卷，第209–255页。Elsevier，2014年。H. Wang 和 A. Ramdas. 重尾均值估计的 Catoni 型置信序列。*arXiv*预印本 *arXiv:2202.01250*，2022年。H. Wang 和 A. Ramdas. Huber鲁棒置信序列。*arXiv*预印本 *arXiv:2301.09573*，2023年。I. Waudby-Smith 和 A. Ramdas. 通过投注估计有界随机变量的均值。*Journal ofthe Royal Statistical Society B*，2023年。I. Waudby-Smith、Z. S. Wu 和 A. Ramdas. 局部隐私非参数置信区间与置信序列。*arXiv*预印本 *arXiv:2202.08728*，2022年。S. Wei 和 Y. Xie. 用于变点检测的在线核CUSUM。*arXiv*预印本 *arXiv:2211.15070*，2022年。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 Journal on Selected Areas in Information Theory*，2(2):494–514，2021年。X. Yu、M. Baron 和 P. K. Choudhary. 二项稀疏过程中的变点检测及其在流行病学中的应用。*Sequential Analysis*，32(3):350–367，2013年。W. Zaremba、A. Gretton 和 M. Blaschko. B检验：一种非参数、低方差的核两样本检验。《神经信息处理系统进展》，26，2013年。

A 补充背景

非独立同分布观测的置信序列。如备注4中所述，我们也可以如下定义非独立同分布观测的置信序列。

定义27. 设 $X_1, X_2, \ldots,$ 表示关于 $\mathcal{X}$ 的独立观测流，其中 $X_t \sim P_{\vartheta_t}$ 对于 $\vartheta_t \in \Theta$ 成立。假设 Θ 是一个向量空间，若 $\mathbb{P}\left(\{\forall t \geq 1 : (1/t)\sum i = 1^t \vartheta_t \in C_t\}\right) \geq 1-\alpha$，则称 $\{C_t \subset \Theta : t \geq 1\}$ 是参数运行平均值的水平为 $(1-\alpha)$ 的置信序列。

我们展示一个简单的 图3展示了例6中引入的、参数随时间变化时的置信序列示例 .

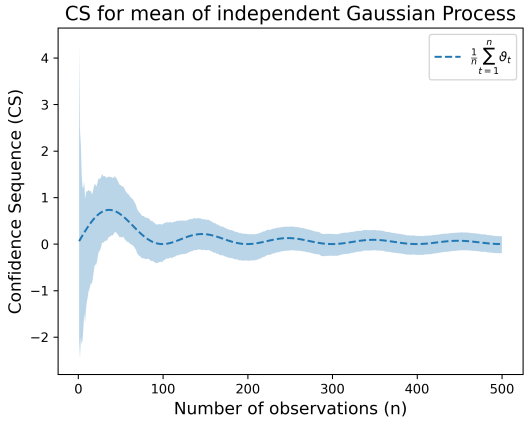

图3：例6中引入的置信序列示例，用于独立高斯过程的运行（条件）均值，其均值函数随时间变化，方差固定为1。

实现后向置信序列。 如果我们知道如何构造前向置信序列，就可以直接按照以下步骤构造后向置信序列：

- 在第 n 轮结束时，我们已经观测到 $X_1, \ldots, X_n$。引入观测值的时间反向版本 $Y_s = X_{n+1-s}$，其中 $s \in [n] :$ $= \{1, \ldots, n\}$。

- 利用 $\{Y_s$ 构造一个新的、水平为 $(1-\alpha)$ 的置信序列： $s \in [n]\}$，记为 $\{\bar{B}_{(n)s} : s \in [n]\}$。注意，对于所有 $s \in [n]$， $\bar{B}_{(n)s}$ 关于 $\sigma(Y_1, \ldots, Y_s)$ 和 $= \sigma(X_n, X_{n-1}, \ldots, X_{n-s+1})$ 都是可测的。

- 最后，我们再次反转置信序列的索引，得到 $\{B_{(n)\,t} : t \in [n]\}$，其中 $B_{(t}{}^{n} = \bar{B}_{)\,(n)\,n+1-s}$。注意，由于它是置信序列，我们有 $\mathbb{P}_\infty\left(\forall t \in [n] : \theta_0 \in B_{(t}{}^{n}\right) \geq 1-\alpha_{)}$。上标 (n) 用来提醒我们，前向置信序列只有一个，而每个时间 n 都会重新构造一个不同的后向置信序列。

示意图：BCS-Detector 策略。图 4 展示了我们在第 4 节中提出的通用 BCS-Detector 策略背后的直观思路，所用任务是检测有界观测值的均值变化（第5.2节）。图 4 中的三幅图突显了方案的以下几个方面：

- 在变点之前（或不存在变点时），前向和后向置信序列都集中在参数空间中的同一区域附近 Θ （本例中为 $[0, 1]$）。特别要注意，在这种情况下，对于 t的所有取值，一个置信区间（CI）完全包含在另一个置信序列对应的置信区间（CI）中。

- 随着来自变化后分布的观测值开始到来，我们预计前向和后向置信序列会开始逐渐分离。这一点在图 4 的第二幅图中有所展示。

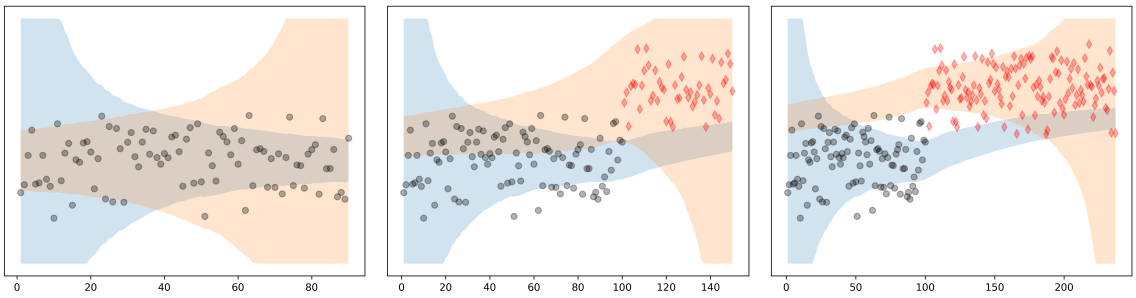

图 4：这些图展示了 BCS-Detector 策略背后的基本思想。在变化点之前（第一幅图），前向和后向置信序列有显著的重叠。在获得一些变化后观测值之后（中间图），后向置信序列开始偏离前向置信序列，尽管这种偏离尚不足以使两个置信序列变得不一致。最后一幅图展示了这样一种情形：数量足够多的变化后观测值已经到来，导致前向和后向置信序列产生不一致。当这种情况发生时，我们的方案停止并拒绝零假设。

- 最后，在收集到足够多的变点后观测值后，后向置信序列开始集中于变点后参数 θ_1 附近，因此有些后向置信区间会与前向置信区间不相交。我们的 BCS-Detector 方案将这种情况作为停止信号，并宣告检测到一个变点。

核 MMD 度量。在第 5.4 节中，我们构建了一种方案，用于检测生成成对观测的分布之间的成对核 MMD 距离是否发生变化。这里，我们回顾其定义。

我们假设 k： $\mathcal{X}\times\mathcal{X}\to\mathbb{R}$ 表示一个有界正定核，并令 $\mathcal{H}_k$ 表示与 k相关联的再生核希尔伯特空间（RKHS）。

定义 28. 给定一个正定核函数 k： $\mathcal{X}\times\mathcal{X}\to\mathbb{R}$，两个分布 P 和 Q 在 $\mathcal{X}$ 上的核最大均值差异距离定义为

$$d_{\mathrm{MMD}}(P,Q)=\sup_{g\in\mathcal{H}_k:\|g\|_k\le 1}\mathbb{E}_P[g(X)]-\mathbb{E}_Q[g(Y)].$$

上述定义的核最大均值差异距离是称为积分概率度量（*IPMs*）的一类统计距离的一个实例。对于一类称为特征核的核函数，已知 d_{MMD} 是概率分布空间上的距离度量。

B 命题 8 的证明

虚警概率上界的证明。回顾一下，停止时间定义为首次满足条件 $\cap_{t=1}^{T}C_t=\emptyset$的时间 τ。现在，考虑零假设下置信序列的“良好”事件：$\mathcal{E}=\cap_{t=1}^{\infty}\{\theta_0\in C_t\}$，根据定义，该事件满足 $\mathbb{P}_\infty(\mathcal{E})\ge 1-\alpha$。因此，在该事件 $\{\theta_0\}\subset\cap_{t=1}^{\infty}\neq\emptyset$下，这又意味着 $\mathbb{P}_\infty(\tau=\infty)\ge 1-\alpha$，正如所需。

检测延迟界的证明。令 $w(t)\equiv w(t,\Theta,\alpha)$ 表示经过 t 个观测值后的置信C_t 宽度。根据假设， T 足够大，能够确保在变点处，‘良好’事件 $\mathcal{E}=\cap_{t=1}^{\infty}\left\{\frac{1}{t}\vartheta_t\in C_t\right\}$发生时， $\theta_1\notin C_T$。注意，在 $\mathcal{E}$ 的定义中，我们有$\bar{\vartheta}_t=\theta_0\mathbf{1}_{t\le T}+\theta_1\mathbf{1}_{t>T}$。

在事件 $\mathcal{E}$ 下，我们知道 $\theta_0\in C_T$。对于任意 $t>T$，引入术语 $\lambda_t=T/t$和 $\lambda_t=1-\lambda_t$。于是，根据置信序列的定义，在事件 $\mathcal{E}$ 下，对于所有 $t>T$，我们有 $\lambda_t\theta_0+\lambda_t\theta_1\in C_t$。集合 C_t 在 $t>T$ 处的宽度不大于 $w(t)\equiv w(t,\Theta,\alpha)$。因此，若要确保在 $t>T$ 之前停止，只需满足 C_t 和 C_T 的宽度之和（即 $w(t)+w(T)$）小于 $d\left(\theta_0,\lambda_t\theta_0+\bar{\lambda}_t\theta_1\right)$。

备注 9 的非正式计算。如备注 9 所述，在许多情况下，停止时间τ满足： $\tau\approx\min\{t\ge T:\ 1/\sqrt{t}+1/\sqrt{T}\le((t-T)/t)\Delta\}$，其中 $\Delta=d(\theta_0,\ \theta_1)$。现在，我们考虑变点 T 处于两种不同状态时延迟 $\tau-T$ 的行为。

首先考虑 T“较小”的情况；即 $T \approx 1/\Delta^2$。为具体起见，假设 $T = 9/\Delta^2$。那么，很容易验证，当 $\tau \le 4T$ 时，由于对于 $t = 4\tau$

$$\frac{1}{\sqrt{t}} + \frac{1}{\sqrt{T}} = \frac{\Delta}{4} + \frac{\Delta}{8} \le \frac{3\Delta}{4} = \frac{t-T}{t}\Delta.$$

接下来，我们考虑 Δ 固定但 $T \to \infty$的情形。在这种情况下，我们有 $\tau - T = \mathcal{O}\left(\sqrt{T}\right)$。为此，考虑 $t = T + u$，其中 $u = o(T)$。于是，我们有

$$\frac{1}{\sqrt{t}} + \frac{1}{\sqrt{T}} \approx \frac{2}{\sqrt{T}}, \quad \text{and} \quad \frac{t-T}{t}\Delta \approx \frac{u}{T}\Delta.$$

因此，这意味着检测延迟的适当增长阶为 $u = \mathcal{O}(\sqrt{T}/\Delta)$，因为 $T \to \infty$且 Δ 固定。

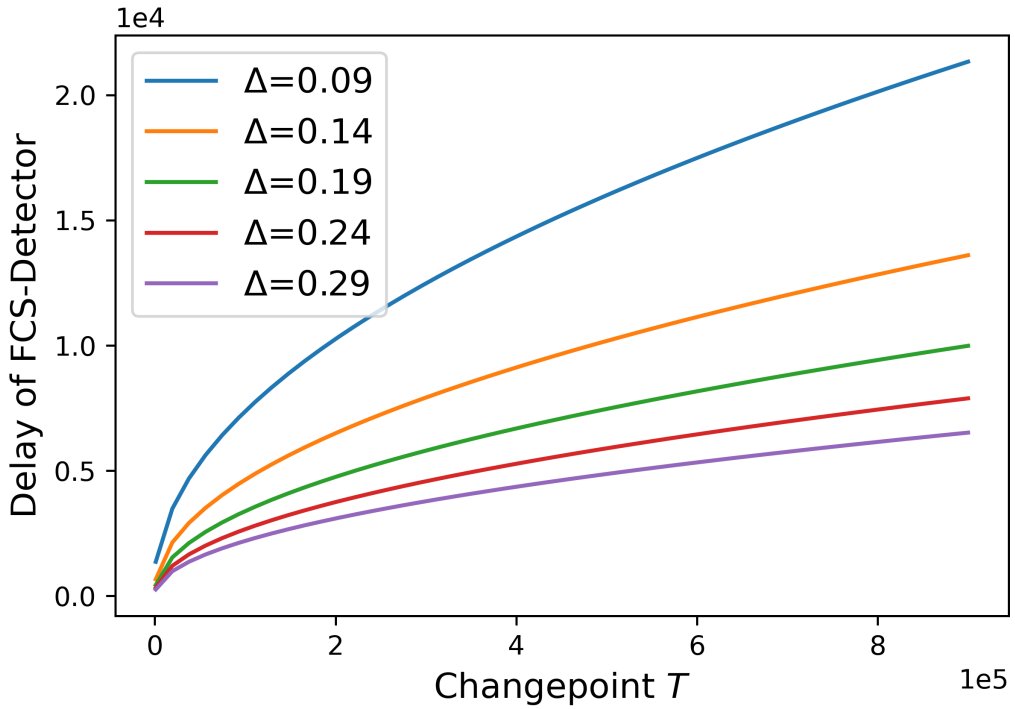

图 5：这些图展示了在备注 9 的条件下，第3节中介绍的 SCD 方案的延迟变化情况。当变点 $T \approx 1/\Delta^2$时，延迟与 T呈线性关系；而在 $T \to \infty$ 且 Δ 固定的情形下，延迟呈现 $\approx \sqrt{T}$的表现。

C 定理 13 的证明

平均运行长度控制的证明。 要证明这一结果，注意到在零假设成立时，我们若在某个时间 τ 停止，则前向置信序列或后向置信序列（或两者）必定未能覆盖并包含 θ_0。换言之，

$$\{\tau = N, T = \infty\} \implies \{C_N \text{ miscovers}\} \vee \{B_1^{(N)} \text{ miscovers}\}$$

$$\{\tau \le N, T = \infty\} \implies \{(C_t)_{t=1}^N \text{ miscovers}\} \vee \bigcup_{n=1}^{N} \{(B_s^{(n)})_{s=n}^1 \text{ miscovers}\}$$

由于每个都是 $(1-\alpha)$-置信序列，根据简单的并集界，对于任意固定时间 N，我们有 ，

$$\Pr_\infty(\tau \le N) \le \min\{(N+1)\alpha, 1\}.$$

换言之，我们有 $\mathbb{P}_\infty(\tau > N) \ge (1-\alpha(N+1)) \vee 0$，因此

$$\begin{aligned}
\mathbb{E}_\infty \tau = \sum_{N=1}^{\infty} \Pr_\infty(\tau > N) &\ge \sum_{N=1}^{1/\alpha-1} (1-\alpha-\alpha N) \\
&= (1/\alpha - 1)(1-\alpha) - \alpha \sum_{N=1}^{1/\alpha-1} N \\
&= 1/\alpha - 2 + \alpha - \alpha \frac{(1/\alpha-1)(1/\alpha)}{2} \\
&= \frac{1}{2\alpha} - 3/2 + \alpha.
\end{aligned}$$

正如我们在备注1中先前提到的，我们无需知道 θ_0 或 θ_1，也无需知道变点的"方向"。上述论证适用于任意带索引的分布族。

检测延迟界的证明。 接下来，我们将给出在事件 $\mathcal{E} := \{\theta_0 \in C_t\colon\ 1 \le t \le T\}$ 条件下平均检测延迟的上界，该结论陈述于定理 13 中。由于 $(\tau - T)^+ = \max\{0, \tau - T\}$ 是一个非负随机变量，我们有

$$\begin{aligned}
\mathbb{E}_T[(\tau-T)^+|\mathcal{E}] &= \sum_{t=0}^{\infty} \mathbb{P}_T\left((\tau-T)^+ \ge t|\mathcal{E}\right) \\
&\le \sum_{t=0}^{t'-1} 1 + \sum_{t\ge t'} \mathbb{P}_T\left((\tau-T)^+ \ge t'|\mathcal{E}\right) \\
&= \sum_{t=0}^{t'-1} 1 + \sum_{t\ge t'} \mathbb{P}_T\left(\tau > T + t'|\mathcal{E}\right) \\
&= t' + \sum_{t\ge t'} \mathbb{P}_T\left(\tau > T + t'|\mathcal{E}\right), \quad \text{for any } t' \ge 1. \qquad (1)
\end{aligned}$$

回想一下，$u_0 \equiv u_0(\theta_0, \theta_1, T)$ 在定理 13 的陈述中引入，它表示在 T 之后，后向置信序列必须停止与前向置信序列相交的最小时间的一个上界。对于某个整数 $i \ge 1$，考虑事件 $\{\tau > T + iu_0\}$，并注意到它满足以下包含关系：

$$\begin{aligned}
\{\tau > T + iu_0\} \cap \mathcal{E} &\subset \left(\cap_{j=1}^{i} \left\{\theta_0 \in B_T^{(T+ju_0)}\right\}\right) \cap \mathcal{E} \\
&\subset \left(\cap_{j=1}^{i} \left\{\theta_0 \in B_{T+(j-1)u_0}^{(T+ju_0)}\right\}\right) \cap \mathcal{E}. \qquad (2)
\end{aligned}$$

在上面的推导中，(2) 用到了这样一个事实：对于任意 $s' > s$，$B_s^{(n)} \subset B_{s'}^{(n)}$。接下来，我们提出以下两点观察：

(I) 对于 $j \ne j'$，事件 $\{\theta_0 \in B^{(T+ju_0)}{}_{T+(j-1)u_0}\}$ 和 $\{\theta_0 \in B^{(T+j'u_0)}{}_{T+(j'-1)u_0}\}$ 相互独立

(II) 对于所有 $1 \le j \le i$，事件 $\{\theta_0 \in B^{(T+ju_0)}{}_{T+(j-1)u_0}\}$ 与 $\mathcal{E}$ 相互独立。

陈述**(I)** 可由以下观察得出：事件 $\{\theta_0 \in B^{(T+ju_0)}{}_{T+(j-1)u_0}\}$ 属于 σ代数 $\sigma(\{X_k\colon\ T+(j-1)u_0 \le k < T+ju_0\})$，而 $\{\theta_0 \in B^{(T+j'u_0)}{}_{T+(j'-1)u_0}\}$ 属于$\sigma(\{X_k\colon\ T+(j'-1)u_0 \le k < T+j'u_0\})$；这两个事件相互独立。类似地，第二个陈述**(II)** 利用了以下事实：$\mathcal{E}$ 属于 $\sigma(\{X_k\colon\ 1 \le k < T\})$，而对于所有 $1 \le j \le i$，它与 $\sigma(\{X_k\colon\ T+(j-1)u_0 \le k < T+ju_0\})$ 相互独立。

基于以上观察，我们得出结论：

$$\begin{aligned}\mathbb{P}_T\left(\tau > T + iu_0|\mathcal{E}\right) &= \mathbb{P}_T\left(\{\tau > T + iu_0\} \cap \mathcal{E}\right)/\mathbb{P}_T(\mathcal{E}) \\ &\le \mathbb{P}_T\left(\left(\cap_{j=1}^{i}\{\theta_0 \in B_{T+(j-1)u_0}^{(T+ju_0)}\}\right) \cap \mathcal{E}\right)/\mathbb{P}_T(\mathcal{E}) \\ &= \mathbb{P}_T\left(\cap_{j=1}^{i}\{\theta_0 \in B_{T+(j-1)u_0}^{(T+ju_0)}\}\right) \qquad (3) \\ &\le \alpha^i, \quad \text{for all } i \ge 1. \qquad (4)\end{aligned}$$

在上面的推导中，(3)由(II)得出，(4)则用到了(I)。现在，我们回到(1)，并令 t' 取为 $i_0 \times u_0$，其中 i_0 是一个稍后指定的整数。接着注意到

$$\begin{aligned}\mathbb{E}_T[(\tau - T)^+|\mathcal{E}] &\le i_0u_0 + \sum_{i=i_0}^{\infty}\sum_{t=iu_0}^{(i+1)u_0-1}\mathbb{P}_T(\tau > T + t|\mathcal{E}) \\ &\le i_0u_0 + \sum_{i=i_0}^{\infty}u_0\mathbb{P}_T\left(\tau > T + iu_0|\mathcal{E}\right) \\ &\le i_0u_0 + u_0\frac{\alpha^{i_0}}{1-\alpha} = u_0\left(i_0 + \frac{\alpha^{i_0}}{1-\alpha}\right). \qquad (5)\end{aligned}$$

(5)中的不等式用到了这样一个事实：$\mathbb{P}_T(\tau > T + iu_0) \le \alpha^i$ 已在(4)中推导得出。然后，按照定理 13所述，通过选取 $i_0 = \lceil \log(1/1-\alpha)/\log(1/\alpha) \rceil$，即可得到最终结果。注意，当 $\alpha < 0.5$时，我们有$i_0 = 1$。

D 第5节的延后证明

D.1 推论17、推论 22、推论 23的证明

All 这三个结果可直接由以下命题得出 n.

命题29. 对于某个 $\Delta > 0$，定义时间 t_0 为

$$t_0 = \min\left\{t \ge 1 : c\sqrt{\frac{\log\log t + \log(1/\alpha)}{t}} \le \frac{\Delta}{2}\right\},$$

其中 $c > 0$ 是某个常数。于是，我们有

$$t_0 = \mathcal{O}\left(c^2\,\frac{\log\log(c/\Delta) + \log(1/\alpha)}{\Delta^2}\right).$$

证明。 不失一般性，我们假设 $c = 1$；或者等价地，可以用 Δ/c替换 Δ。现在注意到，我们可以给 $t_0 \le t_1 + t_2$ 一个上界，其中

$$t_1 = \min\{t \ge 1 : \sqrt{\log\log(t)/t} \le \Delta/4\}, \quad \text{and} \quad t_2 = \min\{t \ge 1 : \sqrt{\log(1/\alpha)/t} \le \Delta/4\}.$$

通过简单计算，我们可以得到 $t_2 = \mathcal{O}\left(\log(1/\alpha)/\Delta^2\right)$。因此，为了完成证明，我们将证明 $t_1 = \mathcal{O}\left(\log\log(1/\Delta)/\Delta^2\right)$。我们分两步进行：(i) 首先证明 $t_1 \le 32/\Delta^3$，(ii) 利用这一点，将结果进一步细化，以证明 $t_1 = \mathcal{O}\left(\log\log(1/\Delta)/\Delta^2\right)$。

令 $t_3 = 32/\Delta^3$，其中 $\Delta \le 1$。然后注意到

$$\left(\frac{4}{\Delta}\sqrt{\frac{\log\log(t_3)}{t_3}}\right)^2 = \frac{16}{\Delta^2} \times \frac{\Delta^3\,\log\log(32/\Delta^3)}{32} = \frac{1}{2}\frac{\log\log(32/\Delta^2)}{1/\Delta} \le 0.63 < 1.$$

最后一个不等式结合 t_1 的定义可知，$t_1 \le t_3$。

因此，$\log\log(t_1) \le \log\log(t_3) = \log\log(32/\Delta^3) \le 2.35 + \log\log(1/\Delta)$。于是，我们有

$$\frac{\log\log(t_1)}{t_1} \le \frac{2.35 + \log\log(1/\Delta)}{t_1},$$

这意味着 $t_1 \le \frac{16}{\Delta^2}(2.35 + \log\log(1/\Delta)) = \mathcal{O}\left(\log\log(1/\Delta)/\Delta^2\right)$。将 t_1 的这一上界与先前得到的 t_2 的上界结合起来，并利用 $t_0 \le t_1 + t_2$ 这一事实，即可得到所需结果。 □

D.2 命题 19 的证明

为了证明期望检测延迟的方差自适应上界，我们首先给出以下结果，说明 Waudby-Smith 和 Ramdas（2023年）所推导的置信序列的宽度。

命题 30. 我们可以修改第 *5.2* 节中用于实例化 *BCS-Detector* 的后向置信序列，从而得到一个水平为$(1-2\alpha)$的后向置信序列（对每个 $n \ge T$ 均成立），记为 $\{B_t^{(n)} : 1 \le t \le n\}$，使得 $B_t^{(n)}$ 的宽度满足

$$w(t, \theta_1, \alpha) = \mathcal{O}\left(\frac{\sigma_1}{\sqrt{t}}\left(\sqrt{\log(1/\alpha)} + \sqrt{\log t}\right)\right), \quad \textit{for all } T \le t \le n.$$

注31。 该结果的主要优势在于，它明确刻画了后向置信序列（对于$n \ge T$）的宽度，并将其表示为变化后分布 P_{θ_1} 的标准差（σ_1）的函数；相比之下，第5.2节所述的原始置信序列的宽度取决于 σ_1 的经验估计值。作为该结果的推论，我们首先应用推论14，然后重复推导（29）时所使用的计算，从而得到命题 19。

证明。 由于我们只关注置信序列宽度衰减的阶（而非精确的常数），因此在本证明中不会跟踪论证中的常数。特别地，我们将使用 $A \lesssim B$ 表示由 $A/B = \mathcal{O}(1)$ 得到，并使用 $A \approx B$ 表示 $A \lesssim B$ 和 $B \lesssim A$。

我们按以下步骤进行：

- 首先，我们证明可以基于来自变化后分布的样本，为经验方差构造一个水平为$1-\alpha$的置信序列。具体而言，令 $\widehat{\sigma}_n^2 = (\frac{1}{4} + \sum_{t=1}^n (X_t - \widehat{\mu}_t)^2)/(n+1)$，其中 $X_1, X_2, \ldots \sim P_{\theta_1}$ 独立同分布。于是，我们有如下结论：

$$\mathbb{P}(\mathcal{E}_1) \ge 1 - \alpha, \quad \text{where} \quad \mathcal{E}_1 := \cap_{n \ge 1}\left\{|\widehat{\sigma}_n^2 - \sigma_1^2| = \mathcal{O}\left(\sqrt{\frac{\log\log n + \log(1/\alpha)}{n}}\right)\right\}. \tag{6}$$

  因此，利用这一事件，对于任意 $n \ge T$，我们可以修改后向置信序列，得到一个 level-$(1-2\alpha)$ 后向置信序列，其中 $\widehat{\sigma}$ 被 σ_1（加上一个较小的近似误差项）替换，适用于 $T \le t \le n$。接下来的两步中，我们将刻画这些 level-$(1-2\alpha)$ 置信序列的宽度。

- 接下来，我们证明在事件 $\mathcal{E}_1$ 下，有

$$\sum_{i=1}^n v_i \psi_E(\lambda_i) \approx \log(\sigma_1^2 n). \tag{7}$$

- 在同一事件下，我们接着证明

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \lambda_i \approx \sqrt{\frac{2\log(2/\alpha)}{\sigma_1^2}}. \tag{8}$$

综合这些结果，我们得到置信序列的宽度为

$$w_n \approx \frac{1}{\sqrt{n}} \frac{\log(2/\alpha) + \sum_{t=1}^n v_t \psi_E(t)}{\frac{1}{\sqrt{n}}\sum_{t=1}^n \lambda_t} \approx \frac{\sigma_1}{\sqrt{n}}\left(\sqrt{\log(2/\alpha)} + \sqrt{\log(n)}\right),$$

所要求的阶。因此，接下来只需证明 (6)、(7) 和 (8)。

证明 (6)。 为证明这一点，我们首先引入常用的方差无偏估计： $\widetilde{\sigma}_n^2 = \frac{1}{(n-1)} \sum_{i=1}^n \sum_{j \neq i} \frac{(X_i - X_j)^2}{2}$。由于观测值 $X_1, X_2, \ldots \sim P_{\theta_1}$ 有界且取值于 $[0,1]$，容易验证 $|\widehat{\sigma}_n^2 - \widetilde{\sigma}_n^2| = \mathcal{O}(1/n)$，因此 $\widehat{\sigma}_n^2 \approx \widetilde{\sigma}_n^2$。

由于 $\widetilde{\sigma}_n^2$ 是 U 统计量的一个实例，因此过程 $\{\widetilde{\sigma}^2 : n \geq 1\}$ 是一个反向鞅，并且适应于可交换滤子。利用这一事实，以及随机变量 $\widetilde{\sigma}^2$ 对所有 $n \geq 1$都有界（因而是次高斯的），我们可以应用 Manole 和 Ramdas (2021, 推论 8)，得出以下时间一致浓缩结果： $\mathbb{P}\left(\forall n \geq 1 : |\widetilde{\sigma}^2 - \sigma_1^2| = \mathcal{O}(r_n)\right) \geq 1 - \alpha$，其中 $r_n = \sqrt{(\log\log n + \log(1/\alpha))/n}$。

证明 (7)。 为证明这一点，我们回顾以下事实： $\psi_E(\lambda)/(\lambda^2/8) \to 1$ 当 $\lambda \to 0$。因此，我们有：

$$
\begin{aligned}
\sum_{i=1}^n v_i \psi_E(\lambda_i) &= \frac{4}{\log(2/\alpha)} \sum_{i=1}^n (X_i - \widehat{\mu}_i)^2 \psi_E(\lambda_i) \approx \frac{4}{\log(2/\alpha)} \sum_{i=1}^n (X_i - \widehat{\mu}_i)^2 \frac{\lambda_i^2}{8} \\
&\lesssim \frac{1}{\log(2/\alpha)} \sum_{i=1}^n (X_i - \widehat{\mu}_i)^2 \frac{\log(2/\alpha)}{i \widehat{\sigma}_{i-1}^2} \lesssim \frac{1}{\log(2/\alpha)} \sum_{i=1}^n (X_i - \widehat{\mu}_i)^2 \frac{\log(2/\alpha)}{i \widehat{\sigma}_{i-1}^2} \\
&\approx \sum_{i=1}^n \frac{(X_i - \widehat{\mu}_{i-1})^2}{\sum_{j=1}^i (X_j - \widehat{\mu}_{j-1})^2} \\
&\leq \log\left(\sum_{i=1}^n (X_i - \widehat{\mu}_{i-1})^2\right) \qquad (9) \\
&\approx \log\left(n \widetilde{\sigma}_n^2\right) \approx \log\left(n \sigma_1^2\right).
\end{aligned}
$$

在上述推导中，(9) 是通过应用以下引理并取 $f(x) = 1/x$ 得到的。

引理 32 (Orabona (2019), 引理 4.13)。 设对所有 i， $a_i \geq 0$，且 $f : [0,\infty) \to [0,\infty)$ 是一个递增函数。那么

$$
\sum_{t=1}^T a_t f(a_0 + \sum_{i=1}^t a_i) \leq \int_{a_0}^{\sum_{t=0}^T a_t} f(x) dx.
$$

证毕。

证明 (8)。 我们按如下方式进行： $r_i = \sqrt{(\log\log i + \log(1/\alpha))/i}$

$$
\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{i=1}^n \lambda_i &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \sqrt{\frac{2\log(2/\alpha)}{\widehat{\sigma}_i^2 \, i}} \gtrsim \sqrt{\frac{2\log(2/\alpha)}{n}} \sum_{i=1}^n \frac{1}{\sqrt{i \sigma_1^2 (1 + r_i)}} \\
&\gtrsim \sqrt{\frac{2\log(2/\alpha)}{\sigma_1^2 \, n}} \left(\sum_{i=1}^n \frac{1}{\sqrt{i}} - \sum_{i=1}^n \frac{r_i}{\sqrt{i}}\right) \\
&\approx \sqrt{\frac{2\log(2/\alpha)}{\sigma_1^2 \, n}} \times \sqrt{n} = \sqrt{\frac{2\log(2/\alpha)}{\sigma_1^2}}.
\end{aligned}
$$

这就完成了命题 30 的证明。 □

E 实验详情

具有指定均值的有界源详情（第5.2节）。为了测试第5.2节所述 SCD 方案的性能，我们通过适当地混合两个均匀分布，构造了一个定义在 $\mathcal{X} = [0,1]$ 上、具有指定均值 $\theta \in [0,1]$ 的概率分布 P_θ。具体而言，我们定义 $P_\theta = (1-\theta)U_1 + \theta U_2$，其中 $U_1 \sim$ 均匀分布$([0,\theta])$， $U_2 \sim$ 均匀分布$([\theta,1])$。

累积分布函数变化检测实验的详细信息。（第5.3节）。在此实验中，我们使用自由度为 3 的单变量 t 分布。对于变化前分布，我们将均值设为 0，

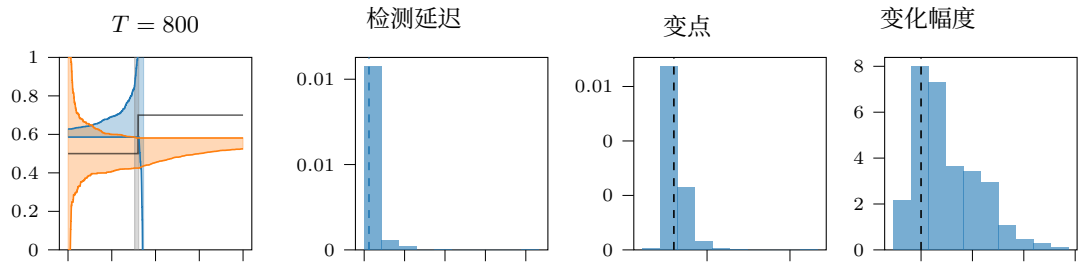

图 6：这些图展示了我们的变点检测方案在独立有界观测下的性能，其中均值在时间 $T = 800$ 由 $p_0 = 0.4$ 变为 $p_1 = 0.6$。第一幅图显示了某次试验中检测时刻（$\tau = 863$）的前向和后向置信序列，灰色阴影区域表示两个置信序列不一致的点。接下来的三幅图展示了在实验中取相同 $\Delta = 0.2$ 和 $\alpha = 0.01$ 值进行 250 次试验后，检测延迟、估计的变点位置和估计的变点幅度的经验分布。

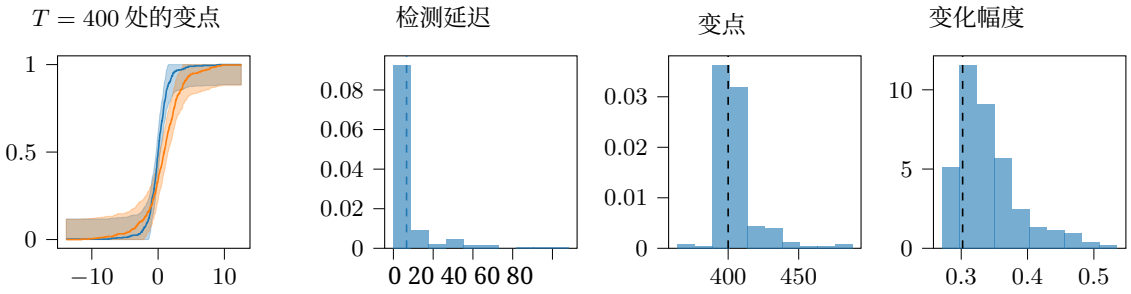

图 7：这些图展示了当观测值取自均值在时间 $T = 800$ 从 0 变为 1.0 的单变量 t 分布（3 个自由度）时，我们的变点检测方案的性能。第一幅图展示了某次试验中检测时刻经验累积分布函数周围的前向和后向置信序列。接下来的三幅图展示了该实验 250 次试验中检测延迟、估计的变点位置和估计的变点幅度的经验分布。

并将尺度参数设为 1。对于变化后分布，我们将均值设为某个值 $\delta > 0$，并将尺度参数设为 2。

二元分类源的详细信息（第5.5节）我们考虑特征-标签对 $(Z_i, L_i) \in \mathbb{R}^2 \times \{0,1\}$，其源分布为 $P_S = P_L \times P_{Z|L}$。我们假设标签 L 在集合 $\{0,1\}$ 上均匀抽取，而特征则根据标签条件服从二元正态分布：$P_{Z|L} = N(\mu_L, I_2)$，其中 $\mu_L = (2L-1)[1,0]^T \in \mathbb{R}^2$。对于这个问题，我们将考虑由权重向量 $w \in \mathbb{R}^2$ 参数化的线性分类器，其形式为 $h_w(z) = \mathbf{1}_{\langle w,z\rangle \geq 0}$。对于 0-1 损失函数 $\phi(z, l, h_w) = \mathbf{1}_{h_w(z) \neq l}$，容易验证，源分布的贝叶斯最优分类器为 $h^* \equiv h_{w^*}$，其中 $w^* \propto [1,0]^T$。对于变化后分布，我们将特征的均值旋转角度 γ；也就是说，$\mu_L = (2L-1)[\cos\gamma, \sin\gamma]^T$。

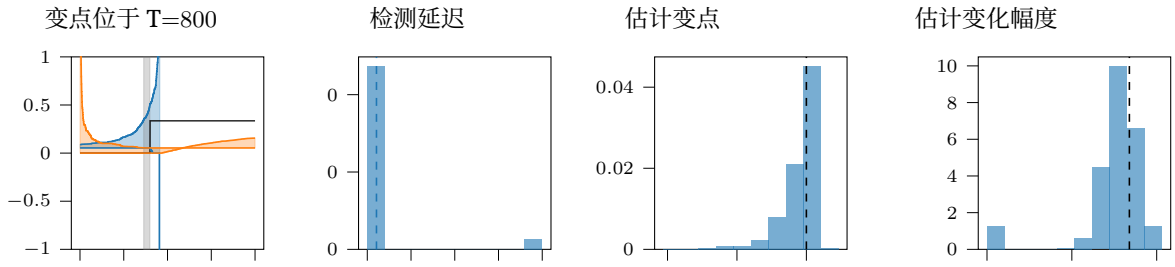

图 8：图中展示了我们的变点检测方案在独立配对多元高斯观测下的性能，其核最大均值差异距离在时间 $T = 800$ 从变化前取值 0 变为$\Delta \approx 0.33$ 。

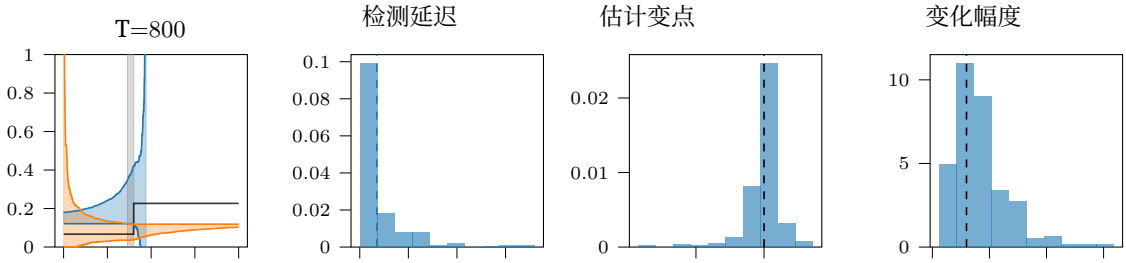

图 9：这些图展示了我们的方法在检测二维特征向量测试分布中的有害变化时的性能，具体如第5.5节所述。在这些图中，变化幅度为 $\Delta \approx 0.16$ ，发生时间为 $T = 800$。

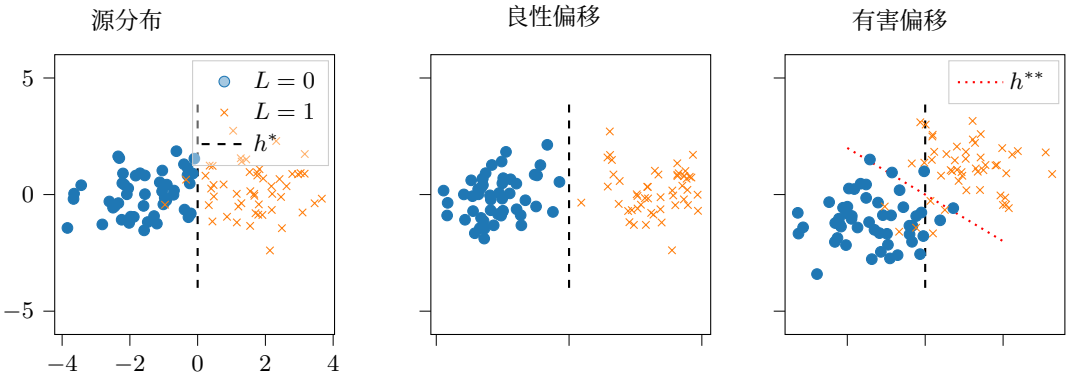

图10：第一幅图展示了对应两个标签（$L = 0$ 和 $L = 1$）的样本，以及此问题的最优线性分类器 h^* （黑色虚线）。第二幅图中，两个分布分离得更明显，同一个分类器 h^* 对于此问题也达到贝叶斯最优，且风险更低。最后，第三幅图展示了有害分布偏移的示例。在这种情况下，两个标签对应的特征分布均逆时针旋转了 45 度，因此 h^* 成为此问题的次优分类器。新的贝叶斯最优分类器以红色点线表示。

F 重复序贯检验的解释

将我们的方案视为重复序贯检验。 由于序贯假设检验与置信序列等价，我们也可以用序贯假设检验的语言来阐述我们的通用策略（定义 12）。具体而言，得益于置信序列的时间一致覆盖保证，我们的方法可以非正式地描述如下：在每一轮 $t \geq 2$，我们都会针对每个 $1 \leq s \leq t-1$ 运行一次新的序贯假设检验，以判定 $(X_1, \ldots, X_s)$ 和 $(X_{s+1}, \ldots, X_t)$ 是否来自同一分布。只要发现某个 t，使得观测值的某种划分之间存在足够显著的差异，我们就可以停止并宣告存在一个变点。正如我们在第4节所见，这一思路可以通过一种优雅的方式实现：将单个前向置信序列（类似于第3节）与一系列基于数据逆序版本构造的置信序列结合起来，我们将后者称为“后向置信序列”。

与 CuSum 的联系。本节最后指出，我们也可以将广受欢迎的参数化序贯变点检测方案 CuSum 解读为执行重复序贯检验。令 f_0 和 f_1 表示观测空间 $\mathcal{X}$ 上的两个概率密度函数。令 $\{X_t : t \geq 1\}$ 表示一列独立观测值，并考虑一个变点检测问题，其中 f_0 和 f_1 分别为变化前分布和变化后分布。

定义33（CuSum）。累积和（CuSum）方法如下：

$$\tau_c = \min\{n \geq 1 : W_n \geq b_\alpha\}, \quad \text{where}$$
$$W_1 = 0, \quad \text{and} \quad W_n = \max_{1 \leq t \leq n} \prod_{i=t+1}^{n} \frac{f_1(X_i)}{f_0(X_i)}, \text{ for } t \geq 2.$$

选择项 b_α，以确保对于给定的 $\alpha \in$ （0,1），平均运行长度至少为 $1/\alpha$.

请注意，CuSum 程序也可以解读为沿反向进行的一系列重复序贯（功效为 1）检验，用于检验零假设 f_0 对比备择假设 f_1。这些是功效为 1 的检验（而非常规的序贯概率比检验），因为它们只会在似然比过程较大时停止，而不会在其较小时停止。

现在我们介绍CuSum的另一种版本。在此定义中，我们约定空集上的下确界或最小值为无穷大，即$\inf\{x : x \in \emptyset\} = \infty$。

定义34 (CuSum-II)。 对于任意 $n \geq 1$和 $1 \leq t \leq n$，将反向序贯检验 τ_n^{back}定义如下：

$$\tau_n^{\text{back}} = \min\{1 \leq t \leq n : L_t^n \geq b_\alpha\}, \quad \text{where} \quad L_t^n = \prod_{s=n-t+1}^{n} \frac{f_1(X_s)}{f_0(X_s)}.$$

然后，我们可以将修正的CuSum停止时间定义为

$$\tau_c' := \inf\{n : \tau_n^{\text{back}} < \infty\}.$$

命题35。 定义*33*和定义*34*中定义的两个检验相同。

证明。 我们证明，对任意 $N \in \mathbb{N}$，集合 $\{\tau_c = N\}$和 $\{\tau_c' = N\}$相等。

$$\begin{aligned}
\{\tau_c = N\} &= \{W_N \geq b_\alpha\} \cap \left(\cap_{n=1}^{N-1} \{W_n < b_\alpha\}\right) \\
&= \{\exists t \in [N] : L_t^N \geq b_\alpha\} \cap \left(\cap_{n=1}^{N-1} \{L_t^n < b_\alpha,\ \forall t \in [n]\}\right) \\
&= \{\tau_N^{\text{back}} < \infty\} \cap \left(\cap_{n=1}^{N-1} \{\tau_n^{\text{back}} = \infty\}\right) \\
&= \{\tau_c' = N\}.
\end{aligned}$$

□