arXiv:2603.04685v1 [math.ST] 5 Mar 2026

序贯多重检验：二阶渐近分析

Jingyu Liu[a] 与 Yanglei Song[b]

数学与统计系，女王大学，金斯顿，加拿大a*jingyu.liu@queensu.ca*，
b*yanglei.song@queensu.ca*

我们研究具有独立数据流的序贯多重检验，其目标是在控制常用误差度量的同时识别未知的信号子集，这些误差度量包括广义族错误率以及错误发现率与未发现率。对于这些问题，已知存在一阶最优的程序，即当误差容忍水平趋于零时，其期望样本量（ESS）与最小可达到期望样本量的比值收敛于1。在本文中，我们建立了二阶渐近最优性的统一理论。我们给出一般充分条件，在这些条件下，对于广泛类别的序贯检验程序，二阶贝叶斯最优性蕴含二阶频率学最优性。由此，若干先前已知为一阶最优的程序被证明是二阶最优的：对每一种信号配置，当误差容忍水平趋于零时，其期望样本量与最小可达到期望样本量之间的差值保持一致有界。此外，我们推导了最小可达到期望样本量的二阶渐近展开，通过识别由多维随机游走的越界问题所产生的二阶修正项，改进了经典的一阶近似。我们将这一结果应用于若干常用的误差度量。

关键词：渐近最优性；期望样本量；非线性更新理论；二阶渐近；序贯多重检验

1. 引言

多重检验关注对 K 个假设进行同时评估，是统计学中的核心议题。经典表述假设样本量固定，并侧重于控制预先指定的第一类错误准则。典型例子包括族错误率（FWER），它限制至少出现一次错误拒绝的概率 [16,17]；广义族错误率，它控制错误拒绝至少 $k \geq 1$ 个的概率 [21]；以及错误发现率（FDR），其定义为所有发现中错误拒绝所占比例的期望值 [5]。

在许多应用中，例如工业过程控制、临床试验和多通道信号检测，数据是按序收集的，决策必须根据累积的观测进行调整[34]。序贯程序可以大幅减少抽样工作量，同时维持规定的第一类和第二类错误保证。该框架源于 Wald 关于序贯概率比检验（SPRT）的开创性工作 [35]。后续研究将 Wald 的思想扩展到通过多条数据流观测的多重假设，由此形成两种主要范式：各数据流的抽样可在不同时间终止 [1,3,25,38],，也可在所有数据流中同时终止 [6–8,15,31,32,37]。本文聚焦于同时停止范式。

具体而言，我们考虑 K 个独立数据流，其中 $\{\mathbf{X}_1^k$、$\mathbf{X}_2^k$，.....$\}$ 表示第 k 个数据流的观测，其中 $k \in [K] := \{1$，.....，$K\}$。对于每个数据流 k，我们检验两个简单假设 H_0^k 与 H_1^k。令 $A \subset [K]$ 表示（未知的）信号集合，即备择假设成立的数据流。我们允许 A 取自一个预先指定的集合 A $\subset 2^{[K]}$，其中 $2^{[K]}$ 是 $[K]$ 的幂集。如果 A 是 $2^{[K]}$ 的真子集，那么它编码了有关信号的结构信息

子集。例如，若 $\mathcal{A} = \{A \subset [K] : \ |A| = m\}$，那么信号数量已知为 m。本文聚焦于独立数据流的情形，并请参阅 [6]，了解向依赖情形的推广。

在每个时间 $t \geq 1$，观测样本 $\{\mathbf{X}_t^1, \ldots, \mathbf{X}_t^K\}$。抽样持续进行，直到停止规则确定已经收集到足够的证据，此时便对未知信号子集作出决策。形式上，序贯多重检验规则由一个停止时间 T 和一个终端决策规则 $D \in \mathrm{A}$ 组成，其中 T 表示收集到的观测总数，而 D 则根据截至时间 T所观测的数据对 A进行估计。目标是构造 (T, D)，在控制预先规定的错误标准的同时，使每个 $A \in \mathrm{A}$ 的期望样本量（ESS）$\mathbb{E}_A[T]$ 尽可能小。

与经典固定样本量表述不同，序贯程序通常旨在同时控制第一类错误和第二类错误。现有程序可用于控制族错误率（FWER）[2,3,7,8], 广义族错误率 [1,9], 广义误分类率 [22,25],以及错误发现率与未发现率 [4,15,19]；精确定义请参见小节2.1。对于这些标准，已有一阶渐近最优性结果；在某些情况下，还需要额外的结构信息，例如信号数量已知 [15,31,32]。

具体而言，用 $\Delta(\boldsymbol{\theta})$ 表示由 $\boldsymbol{\theta} \in (0,\ 1)^r$ 索引的一般序贯程序类，其中$r \geq 1$。在具体应用中，$\Delta(\boldsymbol{\theta})$ 可能表示满足显著性水平为 $\boldsymbol{\theta}$ 的指定错误准则的程序集合，此时 r 通常为 1 或 2。给定 $\Delta(\boldsymbol{\theta})$ 和 $A \in \mathrm{A}$，定义

$$T_A^{\min}(\Delta(\boldsymbol{\theta})) := \inf_{(T,D)\in\Delta(\boldsymbol{\theta})} \mathbb{E}_A[T], \tag{1}$$

在信号配置 A 下，$\Delta(\boldsymbol{\theta})$ 中可达到的最小期望样本量（ESS）。

设 $\{\delta(\boldsymbol{\theta}) = (T(\boldsymbol{\theta}),\ D(\boldsymbol{\theta})) : \boldsymbol{\theta} \in (0,\ 1)^r\}$ 为一族程序，满足对每个 $\boldsymbol{\theta}$，$\delta(\boldsymbol{\theta}) \in \Delta(\boldsymbol{\theta})$。若对每个 $A \in \mathrm{A}$，均有

$$\mathbb{E}_A[T(\boldsymbol{\theta})] = T_A^{\min}(\Delta(\boldsymbol{\theta}))(1 + o(1)) \quad \text{as } \boldsymbol{\theta} \to \mathbf{0},$$

也就是说，程序的期望样本量与可达到的最小期望样本量之比在 $\boldsymbol{\theta}$ 趋近于零向量$\mathbf{0}$时收敛到 1。此外，对于一组实数 $\{Q_1(\boldsymbol{\theta}) : \boldsymbol{\theta} \in (0,\ 1)^r\}$，若 $Q_1(\boldsymbol{\theta})$ 为 $T_A^{\min}(\Delta(\boldsymbol{\theta}))$ 提供一阶渐近近似，则

$$T_A^{\min}(\Delta(\boldsymbol{\theta})) = Q_1(\boldsymbol{\theta})(1 + o(1)).$$

对于上述误差控制准则，已有研究建立了一阶最优程序，并给出了最小可达到期望样本量的相应一阶近似 [15,31,32]。详情请参见第4节。这里，为了具体说明，我们考虑一类序贯程序$\Delta_{m_0}^{\mathrm{gmr}}(\alpha)$：在每种信号配置 $A \subset [K]$下，无论是第一类错误还是第二类错误，该程序控制至少犯 m_0 个错误的概率，使其不超过 $\alpha \in (0, 1)$，即 $\mathrm{A} = 2^{[K]}$。在 [32], 中，提出了和交集规则 $\{\delta_S(b_\alpha) = (T_S(b_\alpha),\ D_S(b_\alpha))\}$，并证明其为一阶渐近最优，其中 b_α是一个阈值，使得 $\delta_S(b_\alpha) \in \Delta_{m_0}^{\mathrm{gmr}}(\alpha)$（见第 4.1 小节）。此外，最小可达到期望样本量已得到一阶刻画。也就是说，对于每个 $A \subset [K]$，当 $\alpha \to 0$时，

$$\mathbb{E}_A[T_S(b_\alpha)] = T_A^{\min}(\Delta_{m_0}^{\mathrm{gmr}}(\alpha))(1 + o(1)), \quad \text{and} \quad T_A^{\min}(\Delta_{m_0}^{\mathrm{gmr}}(\alpha)) = \frac{|\log(\alpha)|}{\kappa_A}(1 + o(1)),$$

其中 κ_A 是一个与 α 无关的正常数（见第 4.1 小节中的定义）。

在图 1中，我们展示了关于 $K = 20$、$m_0 = 1$ 和 $A = \emptyset$ 的数值研究；详情请参见第 5 节。左侧面板用方形标记显示和交规则的期望样本量，$\mathbb{E}_A[T_S(b_\alpha^*)]$, 其中$b_\alpha^*$ 是一个非保守阈值。图中还用三角形标记显示一阶近似$|\log(\alpha)|/\kappa_A$。这两个量均以 $|\log(\alpha)|$ 为横轴绘制，且 $\alpha \to 0$。中间和右侧面板显示

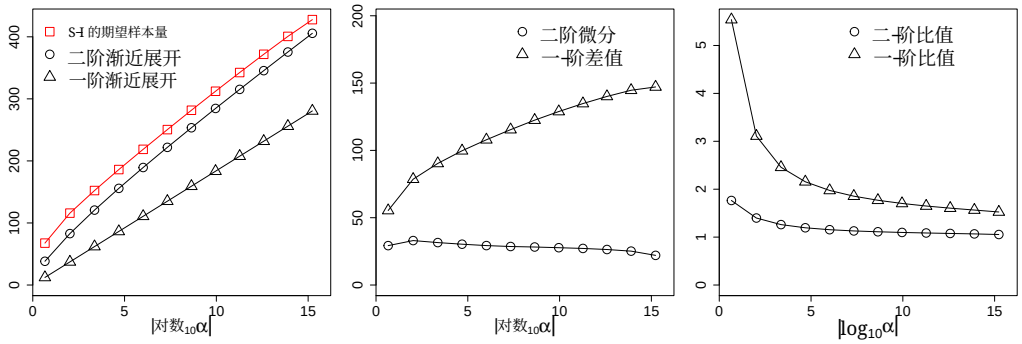


图1：对称情形： $K = 20$， $m_0 = 1$。所有面板的x轴均为 $|\log_{10}\alpha|$。在最左侧的图中，“S-I”表示和交规则（见（26）），“AP”表示可实现的最小 ESS 的渐近近似。在中间的图中，“Diff.”表示S-I规则的期望样本量与两种近似之间的差值。在最右侧的图中，“Ratio”表示S-I规则的期望样本量与两种近似各自的比值。

差值 $\mathbb{E}_A[T_S(b^*_\alpha)] - |\log(\alpha)|/\kappa_A$，以及比值 $\mathbb{E}_A[T_S(b^*_\alpha)]/(|\log(\alpha)|/\kappa_A)$。正如一阶渐近理论所预测的那样，当 $\alpha \to 0$ 时，该比值收敛于 1。然而，中间面板表明，差值并非有界；相反，当 $\alpha \to 0$ 时，它会发散。

这引出了这样一个问题：和交规则是否具有更高阶的最优性，即当 $\alpha \to 0$ 时，差值 $\mathbb{E}_A[T_S(b_\alpha)] - T_A^{\min}(\Delta_{m_0}^{\mathrm{gmr}}(\alpha))$ 是否仍保持有界。这里的比较是与最小可达到期望样本量 $T_A^{\min}(\Delta_{m_0}^{\mathrm{gmr}}(\alpha))$ 进行的，而不是与其近似值进行比较。同样重要的是，如何获得 $T_A^{\min}(\Delta_{m_0}^{\mathrm{gmr}}(\alpha))$ 超越主导对数项的更精确渐近刻画。更一般地说，对于其他误差度量和序贯程序类别，也会产生这些问题。

我们的贡献。 本文为若干序贯多重检验程序建立了二阶频率学最优性，并在广泛的误差度量下，对最小可达到期望样本量进行了二阶刻画。这些误差度量包括广义误分类率、广义族错误率、错误发现/非发现率，以及包含信号配置结构信息的情形。

首先，我们在定理 1 中建立了一个通用框架，为一族规则成为二阶渐近最优提供了明确且可验证的充分条件。回顾一下，对于每个 $\boldsymbol{\theta} \in (0,1)^r$， $\Delta(\boldsymbol{\theta})$ 表示一个序贯程序类。若一族 $\{\delta(\boldsymbol{\theta}) = (T(\boldsymbol{\theta}), D(\boldsymbol{\theta})) \in \Delta(\boldsymbol{\theta})\}$ 满足：对于每个 $A \in \mathrm{A}$，关于 $\{\Delta(\boldsymbol{\theta})\}$ 而言，它被称为二阶渐近最优，

$$\mathbb{E}_A[T(\boldsymbol{\theta})] = T_A^{\min}(\Delta(\boldsymbol{\theta})) + O(1) \quad \text{as } \boldsymbol{\theta} \to \mathbf{0}.$$

也就是说，当误差容限水平趋于零时，超额期望样本量始终保持有界。

我们的框架建立在 [20,23,24] 提出的贝叶斯最优性理论之上。在相应的贝叶斯表述中，信号配置 A 赋有先验分布，其性能通过积分风险来衡量。积分风险由抽样代价和决策损失构成，并通过针对类别 $\Delta(\boldsymbol{\theta})$ 精心选取的代价参数和损失函数来定义。我们证明，若对于每个 $\boldsymbol{\theta} \in (0,1)^r$，停止时间 $T(\boldsymbol{\theta})$ 几乎必然不超过适当构造的贝叶斯规则所对应的停止时间，且由错误决策导致的积分损失在相关渐近尺度下于 $\Delta(\boldsymbol{\theta})$ 上一致受控，那么该程序在频率学派意义下是二阶渐近最优的。

其次，我们在定理 2 中推导出 $T_A^{\min}(\Delta(\boldsymbol{\theta}))$的二阶精确渐近展开，该展开刻画了此类别中的基本效率极限。在适当条件下，我们证明

$$T_A^{\min}(\Delta(\boldsymbol{\theta})) = \frac{\log(c_{\boldsymbol{\theta}}^{-1})}{\kappa_A} + \kappa'_A\sqrt{\log(c_{\boldsymbol{\theta}}^{-1})} + O\left((\log(c_{\boldsymbol{\theta}}^{-1}))^{1/4+\epsilon}\right), \quad \text{as } \boldsymbol{\theta} \to \mathbf{0},$$

对于任意固定的 $\epsilon \in (0,\ 1/4)$，其中代价参数 $c_{\boldsymbol{\theta}} > 0$ 是 $\boldsymbol{\theta}$的函数，而 κ_A 和 κ'_A 是与 $\boldsymbol{\theta}$无关的常数。这改进了现有的一阶近似

$$T_A^{\min}(\Delta(\boldsymbol{\theta})) = \frac{\log(c_{\boldsymbol{\theta}}^{-1})}{\kappa_A}(1 + o(1)),$$

通过确定显式的二阶修正项 $\kappa'_A\sqrt{\log(c_{\boldsymbol{\theta}}^{-1})}$。该项源于多维随机游走的边界穿越问题。我们的分析依赖于非线性更新理论[27], ，并且据我们所知，这一理论在当前背景下尚属首次应用；参见备注 8。

最后，我们将上述框架应用于若干序贯多重检验程序。这些程序在各种误差度量和信息结构下已被证明是一阶渐近最优的，我们进一步确立了它们的二阶最优性。此外，我们还针对相应类别的程序，具体给出了最小可达到期望样本量的二阶渐近近似。

为便于说明，考虑上文提到的类 $\Delta_{m_0}^{\mathrm{gmr}}(\alpha)$ 和和交规则 $\delta_S(b_\alpha)$。图 1 左侧面板中的圆形标记表示 $A = \emptyset$ 时 $T_A^{\min}(\Delta_{m_0}^{\mathrm{gmr}}(\alpha))$ 的二阶渐近近似，其形式为 $|\log(\alpha)|/\kappa_A + \kappa'_A\sqrt{|\log(\alpha)|}$，其中 $\kappa'_A > 0$ 与 α无关。中间和右侧面板分别展示了 $\mathbb{E}_A[T_S(b_\alpha^*)]$与该二阶近似之间的差值和比值。中间面板显示，随着 $\alpha \to 0$，超额期望样本量有界，这证实了本文所确立的 $\{\delta_S(b_\alpha)\}$ 的二阶最优性。

结构安排。第2节介绍序贯多重检验框架，包括误差度量、渐近最优性的概念，以及最小可达到期望样本量的渐近刻画。第3节给出主要理论结果：我们阐述相关的贝叶斯问题，确立贝叶斯二阶最优性与频率学派最优性之间的联系，并推导最小期望样本量的二阶展开。第4 节将这些结果应用于若干问题实例和序贯程序类别。第5 节给出了广义误分类率下的数值研究，第6 节总结全文。附录中给出了省略的证明，并补充了讨论和数值结果。

2. 频率学派表述、最优性与最小ESS

考虑由 $K \geq 2$ 个独立数据流组成的随机向量序列 $\{\mathbf{X}^k\colon\ k \in [K]\}$，其中 $[K] = \{1, \ldots, K\}$。对于每个 $k \in [K]$，$\mathbf{X}^k = \{\mathbf{X}_t^k \in \mathrm{X}^k\colon\ t = 1,2,\ldots\}$ 是一个独立同分布（i.i.d.）的 X^k 值随机向量序列，其中$\mathbf{X}_t^k$ 表示时间t来自第 k 个数据流的观测。在每个时间 $t \geq 1$，定义

$$\mathbf{X}_t^{[K]} := (\mathbf{X}_t^1, \ldots, \mathbf{X}_t^K)^\top \in \mathcal{X}^1 \times \cdots \times \mathcal{X}^K$$

表示时间 t 各数据流上的观测向量，并令 $\mathrm{F}_t := \sigma(\mathbf{X}_1^{[K]}, \ldots, \mathbf{X}_t^{[K]})$ 表示由截至时间 t 的观测生成的 σ-域。

对于每个 $k \in [K]$，令 μ^k 为一个 σ-有限测度，并令 f^k 表示 X_1^k 的边缘分布相对于 μ^k 的密度。我们考虑以下简单假设对

$$H_0^k : f^k = f_0^k \quad \text{versus} \quad H_1^k : f^k = f_1^k. \tag{2}$$

我们施加如下假设，它要求 f_0^k与f_1^k之间的Kullback–Leibler散度为正，并且对数似然比在零假设和备择假设下均具有有限的二阶矩。

假设 1. 对每个数据流 $k \in [K]$，我们假设对于$(i, i') \in \{(0,1),(1,0)\}$，

$$\int \log\left(f_i^k / f_{i'}^k\right) f_i^k \, d\mu^k > 0, \quad \text{and} \quad \int \left[\log\left(f_i^k / f_{i'}^k\right)\right]^2 f_i^k \, d\mu^k < \infty.$$

此外，对每个 $k \in [K]$，令 P_0^k与 P_1^k 分别表示数据流$\mathbf{X}^k$ 在 H_0^k 与 H_1^k下的相应分布。对任意子集 $A \subset [K]$，将 A 解释为备择假设为真的指标集合，称为信号子集，并令 $\mathbb{P}_A$ 表示该配置下 $\{\mathbf{X}^k : k \in [K]\}$的联合分布。特别地，由各数据流之间的独立性，

$$\mathbb{P}_A = \bigotimes_{k \in [K]} P^k, \quad \text{where} \quad P^k = \begin{cases} P_1^k, & k \in A, \\ P_0^k, & k \in A^c, \end{cases}$$

其中 $\bigotimes$表示乘积测度。用 $\mathbb{E}_A$ 表示关于测度 $\mathbb{P}_A$的期望。在 $\mathbb{P}_A$下，$\mathbf{X}_1^{[K]}$的分布关于 $\mu^1 \otimes \cdots \otimes \mu^K$具有密度，由下式给出

$$f_A(x^1, \ldots, x^K) = \prod_{i \in A} f_1^i(x^i) \prod_{j \in A^c} f_0^j(x^j), \quad \text{for } x^k \in \mathcal{X}^k,\ k \in [K]. \tag{3}$$

令 $\mathrm{A} \subset 2^{[K]}$表示所有可能信号子集的集合，其中 $2^{[K]}$是 $[K]$的幂集。$\mathrm{A} = 2^{[K]}$ 对应于对信号配置没有任何先验信息的情形，而 $\mathrm{A} \subsetneq 2^{[K]}$ 则反映了关于信号结构的先验知识。

序贯多重检验程序 δ定义为一对 (T, D)，其中 T是一个$\{\mathrm{F}_t\}$-停止时间，在该时间点所有数据流的抽样均终止；$D = (D^1, \ldots\ldots, D^K)^\top$是一个 F_T-可测的决策规则。每个分量 D^k、$k \in [K]$都是伯努利随机变量，其中 $D^k = 0$ 对应于选择 H_0^k，而 $D^k = 1$ 对应于选择 H_1^k。因此，停止时根据截至时间 T可获得的信息，若 $D^k = 1$（若 $D^k = 0$），我们就判定第 k 个数据流中存在信号（分别为噪声）。稍微滥用一下记号，我们也写作

$$D := \{k \in [K] : D^k = 1\}$$

表示被判定为包含信号的数据流子集。我们将注意力限制在满足 $D \in \mathrm{A}$ 的决策规则上，并用 Δ 表示所有序贯检验程序的类别。

我们的目标是在所有 $A \in \mathrm{A}$ 下最小化期望样本量（ESS）$\mathbb{E}_A[T]$，同时控制与决策规则 D相关的预设错误度量。

2.1. 多重检验错误率

在本小节中，我们定义了若干类序贯检验程序，它们分别对应于不同的错误度量和信息结构。

给定决策 $D \in \mathrm{A}$ 和真实信号子集 $A \in \mathrm{A}$，对于每个数据流 $k \in [K]$，有两种自然的错误类型：如果第 k 个数据流的真实假设是 H_0^k 但 $D^k = 1$ （即假阳性），则会发生第一类错误；而如果真实假设是 H_1^k 但$D^k = 0$ （即假阴性），则会发生第二类错误。

我们首先回顾文献中的三个程序子类，它们不假设任何先验信息，即 $\mathrm{A} = 2^{[K]}$，并控制不同的多重检验错误准则。

第1类。给定用户指定的整数 $1 \le m_0 \le K$ 和容忍水平 $\alpha \in (0, 1)$，我们将控制广义误分类率（GMR） [22,25,32] 的序贯程序类定义为

$$\mathcal{A} = 2^{[K]}, \quad \text{and} \quad \Delta^{\mathrm{gmr}}_{m_0}(\alpha) = \left\{(T, D) \in \Delta : \max_{A \in \mathcal{A}} \mathbb{P}_A\big(|D \triangle A| \ge m_0\big) \le \alpha\right\}.$$

类别2。给定两个用户指定的整数 $1 \le m_1, m_2 \le K$ 使得 $m_1 + m_2 \le K$，以及容忍水平$\alpha, \beta \in (0, 1)$，我们将控制广义族错误率（GFWER） [1,9,32], 的序贯程序类定义为 $\mathrm{A} = 2^{[K]}$，并且

$$\Delta^{\mathrm{gfr}}_{m_1,m_2}(\alpha, \beta) = \left\{(T, D) \in \Delta : \max_{A \in \mathcal{A}} \mathbb{P}_A\big(|D \setminus A| \ge m_1\big) \le \alpha, \max_{A \in \mathcal{A}} \mathbb{P}_A\big(|A \setminus D| \ge m_2\big) \le \beta\right\}.$$

第3类。给定容忍水平 $\alpha, \beta \in (0, 1)$，我们将控制错误发现率（FDR）和错误非发现率（FNR） [4,15,19] 的序贯程序类定义为 $\mathrm{A} = 2^{[K]}$，并且

$$\Delta^{\mathrm{fdr}}(\alpha, \beta) = \left\{(T, D) \in \Delta : \max_{A \in \mathcal{A}} \mathbb{E}_A\left[\frac{|D \setminus A|}{\max\{|D|, 1\}}\right] \le \alpha, \max_{A \in \mathcal{A}} \mathbb{E}_A\left[\frac{|A \setminus D|}{\max\{|D^c|, 1\}}\right] \le \beta\right\}.$$

需要说明几点。首先，对于上述三个类别中的每一类，都要求对所有可能的信号子集 $A \subset [K]$ 进行统一的误差控制。其次，与第1类不同，GFWER 和 FDR/FNR 标准会区分第一类错误与第二类错误。第三，对于 GFWER，当 $m_1 = m_2 = 1$时，该标准便简化为经典族系错误率 [2,3,7,8]；更一般地说，GFWER 允许最多出现 $m_1 - 1$ 个第一类错误和 $m_2 - 1$ 个第二类错误。类似的讨论也适用于广义误分类率。最后，第3类侧重于控制第一类错误和第二类错误的期望比例。

在某些应用中，可能会有关于信号子集结构的先验知识。例如，信号的确切数量可能已知，也可能会给出信号数量的上下界。接下来，我们回顾文献中针对 $\mathrm{A} \subsetneq 2^{[K]}$ [15,31] 情形设计的一类序贯程序。

第 4 类。假设已知信号数量为 $m \in [1, K)$。对于容忍水平 $\alpha \in (0,1)$，我们将控制错误决策概率的序贯程序类定义为

$$\mathcal{A} = \{A \subset [K] : |A| = m\}, \qquad \Delta^{\mathrm{kns}}_{m}(\alpha) = \left\{(T, D) \in \Delta : \max_{A \in \mathcal{A}} \mathbb{P}_A\big(D \ne A\big) \le \alpha\right\}. \tag{4}$$

我们在附录D.2中回顾序贯程序的其他子类。对于给定的子类，目标是在适当的渐近意义下，确定该类中使期望样本量（ESS）最小的程序。在下一小节中，我们将回顾渐近最优性的几种概念。

2.2. 渐近最优性

令 $\Delta(\boldsymbol{\theta}) \subset \Delta$ 表示由误差容忍度向量 $\boldsymbol{\theta} \in (0,1)^r$索引的一般程序子类，其中 $r \in \{1,2\}$。例如， $\Delta(\boldsymbol{\theta})$ 可以对应于上文定义的四类程序之一，其中 $r=1$ 表示第 1 类， $r=2$ 表示类别 2。根据（1）可知， $T_A^{\min}(\Delta(\boldsymbol{\theta}))$ 表示在真实信号子集 $A \in \mathsf{A}$ 下，类 $\Delta(\boldsymbol{\theta})$ 上可实现的最小期望样本量（ESS）。

对于许多类序贯程序，例如第 1–4 类，可达到的最小期望样本量会随着错误概率要求 $\boldsymbol{\theta}$ 变得更加严格而趋于无穷大，即

$$T_A^{\min}(\Delta(\boldsymbol{\theta})) \to \infty \quad \text{as} \quad \boldsymbol{\theta} \to \mathbf{0}, \quad \text{for each } A \in \mathcal{A}, \tag{5}$$

其中$\mathbf{0}$ 表示 $\mathbb{R}^r$ 中的零向量。现在我们正式定义渐近最优性的两种概念。

定义 1（渐近最优性）。设 $\{\delta(\boldsymbol{\theta}) = (T(\boldsymbol{\theta}),\ D(\boldsymbol{\theta})) : \boldsymbol{\theta} \in (0,\ 1)^r\}$ 是一组序贯多重检验程序，且对所有 $\boldsymbol{\theta} \in (0,\ 1)^r$，均有 $\delta(\boldsymbol{\theta}) \in \Delta(\boldsymbol{\theta})$。那么：

（a）$\{\delta(\boldsymbol{\theta})\}$ 称为一阶渐近最优于$\{\Delta(\boldsymbol{\theta})\}$，若对每个 $A \in \mathcal{A}$,

$$\lim_{\boldsymbol{\theta} \to \mathbf{0}} \frac{\mathbb{E}_A[T(\boldsymbol{\theta})]}{T_A^{\min}(\Delta(\boldsymbol{\theta}))} = 1.$$

(b) $\{\delta(\boldsymbol{\theta})\}$ 称为 二阶渐近最优 于 $\{\Delta(\boldsymbol{\theta})\}$，若对于每个 $A \in \mathcal{A}$,

$$\limsup_{\boldsymbol{\theta} \to \mathbf{0}} \Big(\mathbb{E}_A[T(\boldsymbol{\theta})] - T_A^{\min}(\Delta(\boldsymbol{\theta})) \Big) < \infty.$$

我们强调，渐近最优性（AO）的定义要求所述极限对所有 $A \in \mathcal{A}$ 都成立。虽然一阶 AO 确保期望样本量在相对意义上渐近最优，但这并不排除这样的可能性：对于某些 $A \in \mathcal{A}$, 超额量 $\mathbb{E}_A[T(\boldsymbol{\theta})] - T_A^{\min}(\Delta(\boldsymbol{\theta}))$ 会随着 $\boldsymbol{\theta} \to \mathbf{0}$ 发散；图 1 对此作了示意。当（5）成立时，二阶 AO 的概念通过以加法意义控制超额量，提出了更严格的要求。关于高阶最优性的进一步讨论，请参见第 4.4 小节。

对于第 1–4 类以及附录D.2中的其他类别，现有文献提出了检验程序，并证明了其一阶渐近最优性 [15,31,32]；见第4 节和附录D.1。然而，这些一阶最优程序是否也具有更强意义上的二阶最优性，仍是一个未解问题。本文对这一问题给出了 肯定的 回答。

上述表述本质上属于频率学派，旨在对 $A \in \mathcal{A}$上的相关错误概率进行一致控制，同时控制每个 $A \in \mathcal{A}$的期望样本量（ESS）。为确立现有程序的二阶最优性，我们的策略是利用相关贝叶斯表述中的最优性结果，即根据$\mathcal{A}$上的先验分布对错误和期望样本量（ESS）进行平均，然后将这些结果转换为相应的频率学派结论。为此，在第3.1节中，我们介绍贝叶斯表述，并建立联系贝叶斯最优性与频率学派最优性的框架，随后将其应用于序贯程序的一般子类。

2.3. 最小ESS表征

二阶最优性结果使我们能够推导出比现有文献中更精确的最小ESS渐近近似。接下来，我们定义两种渐近近似的概念。

定义 2（渐近近似）。设 $A \in \mathrm{A}$，且 $\boldsymbol{\Theta} \subset (0,1)^r$。设 $\{Q_1(\boldsymbol{\theta}), Q_2(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ 是一组由 $\boldsymbol{\theta}$ 索引的正数。

(a) $\{Q_1(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ 称为 一阶精确逼近，对应于 $T_A^{\min}(\Delta(\boldsymbol{\theta}))$，如果当 $\boldsymbol{\theta} \to \mathbf{0}$ 且 $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ 时，

$$T_A^{\min}(\Delta(\boldsymbol{\theta})) = Q_1(\boldsymbol{\theta})\big(1 + o(1)\big).$$

(b) $\{Q_1(\boldsymbol{\theta}), Q_2(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ 称为 二阶精确逼近，对应于 $T_A^{\min}(\Delta(\boldsymbol{\theta}))$，如果当 $\boldsymbol{\theta} \to \mathbf{0}$ 且 $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ 时，

$$T_A^{\min}(\Delta(\boldsymbol{\theta})) = Q_1(\boldsymbol{\theta}) + Q_2(\boldsymbol{\theta})\big(1 + o(1)\big), \qquad \text{and} \qquad Q_2(\boldsymbol{\theta}) = Q_1(\boldsymbol{\theta})o(1).$$

注 1. 在后续应用中，当 $r = 1$ 时，我们令 $\boldsymbol{\Theta} = (0,1)$。当 $r = 2$ 时，我们可以对 $\boldsymbol{\theta}$ 的两个分量趋于零的方式施加条件。因此，我们在上述定义中引入 $\boldsymbol{\Theta}$。

根据上述定义，如果 $\{Q_1(\boldsymbol{\theta})\}$ 是一阶精确逼近，那么 $T_A^{\min}(\Delta(\boldsymbol{\theta}))$ 的特征可确定到一个误差项，该误差项相对于 $Q_1(\boldsymbol{\theta})$ 可忽略。因此，$Q_1(\boldsymbol{\theta})$ 是渐近近似中的一阶主导项。

另一方面，如果 $\{Q_1(\boldsymbol{\theta}), Q_2(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ 是二阶准确的，那么近似 $Q_1(\boldsymbol{\theta}) + Q_2(\boldsymbol{\theta})$ 的错误相较于 $Q_2(\boldsymbol{\theta})$ 可以忽略不计；由于 $Q_2(\boldsymbol{\theta}) = Q_1(\boldsymbol{\theta})o(1)$，这比 $o(Q_1(\boldsymbol{\theta}))$ 阶错误更精确。这也意味着，$Q_2(\boldsymbol{\theta})$ 是渐近近似中的二阶主导项，从而说明了这一术语的合理性。

如第 4 节所述，文献中针对各类序贯程序已有一阶准确的近似，我们将其推广到二阶。先简要预览一下，暂且假设 $r = 1$ 并将 $\boldsymbol{\theta}$ 记为 α。对于若干类情况，已有的 $T_A^{\min}(\Delta(\alpha))$ 一阶近似形式为 $Q_1(\alpha) = |\log(\alpha)|/\kappa_A$，其中 $\kappa_A > 0$ 是一个常数。在本文中，我们证明，当 $\alpha \to 0$，

$$T_A^{\min}(\Delta(\alpha)) = \frac{|\log(\alpha)|}{\kappa_A} + \kappa'_A\sqrt{|\log(\alpha)|} + O\left(|\log(\alpha)|^{\epsilon_0}\right),$$

其中 $\kappa'_A \geq 0$ 是另一个常数，且 $\epsilon_0 \in (1/4, 1/2)$ 是一个固定常数。

注意，当 $\kappa'_A > 0$ 时，一阶近似与二阶近似具有不同的形式。此外，粗略地说，现有结果的精度可达 $o(|\log(\alpha)|)$，而我们的结果精度可达 $O(|\log(\alpha)|^{\epsilon_0})$，其中在适当的矩条件下，$\epsilon_0$ 可以任意接近 $1/4$。最后，当 $\kappa'_A = 0$ 时，近似 $|\log(\alpha)|/\kappa_A$ 实际上达到了 $O(1)$ 错误。

3. 通过贝叶斯表述实现频率学派最优性的框架

本节首先介绍序贯多重检验问题的一般贝叶斯表述，并回顾一个在贝叶斯意义下二阶最优的程序（见第3.1小节）。随后，我们建立充分条件，在这些条件下，通过与该二阶贝叶斯最优程序进行比较，可以证明某程序在频率学派意义下二阶最优（见第3.2小节）。最后，我们推导频率学派表述下最小可达期望样本量（ESS）的二阶精确渐近近似（见第3.3小节）。

3.1. 贝叶斯表述与贝叶斯最优性

回顾一下，$\mathrm{A} \subset 2^{[K]}$ 表示所有可能信号子集的集合，并令 π_0 为A上的均匀分布，即对于每个 $A \in \mathrm{A}$，$\pi_0(A) = |\mathrm{A}|^{-1}$。在贝叶斯表述中，真实信号子集 A 被视为一个从先验分布 π_0中抽取的随机变量。

设 $\delta = (T, D) \in \Delta$ 为一个序贯程序，并回顾决策规则 D 取值于 A。我们将 δ 的积分风险定义为抽样代价与错误决策所致损失之和。具体而言，令 $c > 0$ 表示每增加一组观测所产生的代价。积分抽样代价定义为

$$\mathrm{ESS}^{\mathrm{int}}(\delta; c) := c\,\mathbb{E}[T] = c \sum_{A \in \mathcal{A}} \pi_0(A)\,\mathbb{E}_A[T], \tag{6}$$

其中，期望 $\mathbb{E}$ 是针对 A 的先验分布和观测的随机性共同计算的。此外，令

$$\mathcal{W}(\cdot \mid \cdot) : \mathcal{A} \times \mathcal{A} \to [0, \infty)$$

为一个非负损失函数，其中 $\mathrm{W}(D \mid A)$ 表示在真实信号子集为 A时，宣告 D 所产生的损失。我们假设，对所有 $A \in \mathrm{A}$，$\mathrm{W}(A \mid A) = 0$。于是，与错误决策相关的积分误差损失定义为

$$\mathrm{ERR}^{\mathrm{int}}(\delta; \mathcal{W}) := \mathbb{E}\left[\mathcal{W}(D \mid A)\right] = \sum_{A \in \mathcal{A}} \pi_0(A) \sum_{d \in \mathcal{A}} \mathbb{P}_A(D = d)\,\mathcal{W}(d \mid A). \tag{7}$$

因此，给定 $c > 0$ 和 W，积分代价定义为

$$r^{\mathrm{int}}(\delta; c, \mathcal{W}) := \mathrm{ESS}^{\mathrm{int}}(\delta; c) + \mathrm{ERR}^{\mathrm{int}}(\delta; \mathcal{W}). \tag{8}$$

注 2. 要将频率学派最优性与贝叶斯最优性联系起来，必须指定代价参数 c 以及损失函数W。

接下来，我们回顾贝叶斯意义下渐近最优性的概念。令

$$r^{\mathrm{int}}_{\min}(c, \mathcal{W}) := \inf_{\delta \in \Delta} r^{\mathrm{int}}(\delta; c, \mathcal{W}) \tag{9}$$

表示可达到的最小积分风险。

定义 3（二阶贝叶斯最优性）。设 $\{\delta(c) : c > 0\} \subset \Delta$ 为一组序贯程序。若存在常数 $M > 0$ 使得，则称族 $\{\delta(c)\}$ 关于 W二阶贝叶斯最优。

$$r^{\mathrm{int}}(\delta(c); c, \mathcal{W}) \le r^{\mathrm{int}}_{\min}(c, \mathcal{W}) + Mc, \quad \text{for all } c > 0.$$

关于此定义的讨论，我们留待注 3。

最后，我们回顾文献中的一个二阶贝叶斯最优程序。回想式（ $f_A(\cdot)$ （3）中的密度的定义。对于 $t \ge 1$，后验分布 π_t： 给定 $\to [0, 1]$ 的信号子集 A关于 σ-代数 $\mathcal{F}_t$ 定义如下：

$$\pi_t(A) := \frac{f_{A,[t]}}{\sum_{B \in \mathcal{A}} f_{B,[t]}}, \quad \text{for each } A \in \mathcal{A}, \qquad \text{where } f_{A,[t]} := \prod_{s=1}^{t} f_A(\mathbf{X}_s^{[K]}). \tag{10}$$

给定 $c>0$ 和损失函数 $\mathcal{W}$，以下程序 $\delta_{\mathrm{Ld}}(c,\mathcal{W})=(T_{\mathrm{Ld}}(c,\mathcal{W}),\, D_{\mathrm{Ld}}(c,\mathcal{W}))$，其研究见 [20,23,29], , 定义如下：

$$T_{\mathrm{Ld}}=\min\Big\{t\geq 1:\min_{D\in\mathcal{A}}\sum_{A\in\mathcal{A}}\pi_t(A)\,\mathcal{W}(D\mid A)<c\Big\},\quad D_{\mathrm{Ld}}\in\operatorname*{argmin}_{D\in\mathcal{A}}\sum_{A\in\mathcal{A}}\pi_{T_{\mathrm{Ld}}}(A)\,\mathcal{W}(D\mid A),\quad (11)$$

其中平局时任意打破，并省略了对 (c,W) 的依赖以简化记号。换言之，该程序在首次出现某个决策 $D\in\mathrm{A}$ 的后验期望损失低于阈值 c 的时间 t 时停止；停止后，选择使后验期望损失最小的决策。Lorden 的论文 [23] 表明，族 $\{\delta_{\mathrm{Ld}}(c,\mathrm{W})\}$ 是二阶贝叶斯最优的，这也是下标“Ld”的由来。

注 3. 文献 [20], 表明，$r^{\mathrm{int}}_{\mathrm{min}}(c,\mathrm{W})$ 的阶为 $c|\log c|$（当 $c\to 0$ 时），并且族 $\{\delta_{\mathrm{Ld}}(c,\mathrm{W})\}$ 在如下意义上渐近一阶最优：

$$\lim_{c\to 0}\frac{r^{\mathrm{int}}(\delta_{\mathrm{Ld}}(c,\mathcal{W});c,\mathcal{W})}{r^{\mathrm{int}}_{\mathrm{min}}(c,\mathcal{W})}=1.$$

因此，$\{\delta_{\mathrm{Ld}}(c,\mathcal{W})\}$的二阶最优性（定义见 3 ，并在 [23], 中得到确立）是一个严格更强的最优性概念。

3.2. 通过贝叶斯最优性实现二阶频率学派最优性

在本小节中，我们给出一项主要结果，该结果提供了一种技术方案，可借助贝叶斯最优性来确立二阶频率学最优性。具体而言，对于每个 $\boldsymbol{\theta}\in(0,1)^r$且满足 $r\in\{1,2\}$，令 $\Delta(\boldsymbol{\theta})\subset\Delta$ 为序贯多重检验程序的一个子类。回顾 $T^{\mathrm{min}}_A(\Delta(\boldsymbol{\theta}))$ 在（1）中的定义，它是该类程序可达到的最小期望样本量。

现在，令 $\delta_0(\boldsymbol{\theta})=(T_0(\boldsymbol{\theta}),\, D_0(\boldsymbol{\theta}))$，$\boldsymbol{\theta}\in(0,1)^r$ 为一组程序，使得对每个 $\boldsymbol{\theta}$，$\delta_0(\boldsymbol{\theta})\in\Delta(\boldsymbol{\theta})$。我们的首个结果给出了充分条件，在这些条件下，$\{\delta_0(\boldsymbol{\theta})\}$ 相对于 $\{\Delta(\boldsymbol{\theta})\}$ 具有二阶频率学最优性。证明见附录 A。

定理 1. 假设条件 *1* 成立。假设对每个 $c_{\boldsymbol{\theta}}>0$ 都存在一个代价 $\boldsymbol{\theta}\in(0,1)^r$、一个损失函数 $\mathcal{W}:\mathcal{A}\times\mathcal{A}\to[0,\infty)$以及一个常数 $L>0$，使得以下条件成立：

$$\max_{A\in\mathcal{A}}\mathcal{W}(D\mid A)>0,\quad \textit{for each } D\in\mathcal{A};\qquad (12)$$

$$T_0(\boldsymbol{\theta})\leq T_{\mathrm{Ld}}(c_{\boldsymbol{\theta}},\mathcal{W})\ \ \mathbb{P}_A\textit{-almost surely, for each } A\in\mathcal{A}\textit{ and }\boldsymbol{\theta}\in(0,1)^r;\qquad (13)$$

$$\mathrm{ERR}^{\mathrm{int}}(\delta;\mathcal{W})\leq Lc_{\boldsymbol{\theta}},\quad \textit{for any procedure } \delta\in\Delta(\boldsymbol{\theta})\textit{ and each }\boldsymbol{\theta}\in(0,1)^r.\qquad (14)$$

那么存在一个常数 $M>0$，它仅取决于 $\mathcal{W}$，而不取决于 $\boldsymbol{\theta}$，使得

$$\mathbb{E}_A[T_0(\boldsymbol{\theta})]-T^{\mathrm{min}}_A(\Delta(\boldsymbol{\theta}))\leq(2L+M)\,|\mathcal{A}|,\quad \textit{for each } A\in\mathcal{A}\textit{ and }\boldsymbol{\theta}\in(0,1)^r.$$

注意，定理 1 是非渐近的。尽管如此，令 $\boldsymbol{\theta}\to\mathbf{0}$ 即可立即推出 $\{\delta_0(\boldsymbol{\theta})\}$ 相对于 $\{\Delta(\boldsymbol{\theta})\}$的二阶频率学派最优性，依据定义 1。

应用定理 1 的关键在于指定一个代价 $c_{\boldsymbol{\theta}}$，该代价依赖于 $\boldsymbol{\theta}$，以及一个损失函数 $\mathcal{W}(\cdot\mid\cdot)$，该函数与 $\boldsymbol{\theta}$无关，从而可以验证其中的三个条件。特别地，条件（12）要求对于每个决策 $D\in\mathcal{A}$，存在一个信号子集 $A\in\mathcal{A}$，在该子集下 D 是

不正确，其含义是 $\mathrm{W}(D \mid A) > 0$。条件（13）要求所考虑的停止时间 $T_0(\boldsymbol{\theta})$ 不晚于（11）中定义的 $T_{\mathrm{Ld}}(c_{\boldsymbol{\theta}}, \mathrm{W})$。最后，条件（14）要求 $\Delta(\boldsymbol{\theta})$ 中任何程序的积分误差损失都以 $Lc_{\boldsymbol{\theta}}$ 为上界。在第4节和附录 D.1 中，我们将定理 1 应用于第 1–4 类，以证明文献中现有的程序是二阶频率学派最优的。

备注4。 通过贝叶斯最优性推导频率学派最优性的思想，此前已出现在序贯分析文献中。特别是，定理 1的证明中的一些论证，与 [24]中定理 4 的证明在思想上相似。然而，现有结果针对的是特定的程序和设定。我们的贡献在于提出明确且可验证的充分条件，并在定理 1中加以陈述；在这些条件下，二阶贝叶斯最优性可推出二阶频率学最优性。该定理以足够一般的层次表述，从而涵盖多种误差度量，包括第 1 类和第 2类，同时不要求假设 $\mathbb{P}_A(D(\boldsymbol{\theta}) \neq A) \to 0$对于 $\boldsymbol{\theta} \to 0$ 所有 $A \in \mathcal{A}$均成立，而 [24]中采用了这一假设。

3.3. 频率学派表述下最小ESS的二阶刻画

在本小节中，我们将 $T_A^{\min}(\Delta(\boldsymbol{\theta}))$ 刻画至二阶，即对于由 $\Delta(\boldsymbol{\theta}) \subset \Delta$ 索引、由 $\boldsymbol{\theta} \in (0,\ 1)^r$ 参数化的子类族，将其表示为 $\boldsymbol{\theta} \to \mathbf{0}$ 。我们考虑一般的损失函数 $\mathrm{W}\colon\ \mathrm{A} \times \mathrm{A} \to [0, \infty)$，并引入一些记号和条件。

对于每个 $k \in [K]$，定义

$$\mathcal{I}_1^k := \int \log\left(f_1^k / f_0^k\right) f_1^k \, d\mu^k, \quad \mathcal{I}_0^k := \int \log\left(f_0^k / f_1^k\right) f_0^k \, d\mu^k, \tag{15}$$

它们是密度函数 f_1^k与 f_0^k之间的 Kullback–Leibler（KL）散度。此外，对于$A,\ C \in \mathcal{A}$ 且给定真实信号子集为 A时，我们将密度函数f_A 与 f_C 之间的 KL 散度（见（3））记为

$$\mathrm{KL}(f_A \mid f_C) := \mathbb{E}_A\left[\log\left(\frac{f_A}{f_C}(\mathbf{X}_1^{[K]})\right)\right] = \sum_{k \in A \setminus C} \mathcal{I}_1^k + \sum_{k \in C \setminus A} \mathcal{I}_0^k. \tag{16}$$

对于每个 $D \in \mathrm{A}$，将使 D 相对于损失函数 W 不正确的信号子集集合记为

$$\mathcal{H}_D^{\mathcal{W}} := \{C \in \mathcal{A} : \mathcal{W}(D \mid C) > 0\}. \tag{17}$$

此外，对于 $A,\ D \in \mathrm{A}$，定义

$$\mathrm{KL}_{A,D}^{\mathcal{W}} := \min_{C \in \mathcal{H}_D^{\mathcal{W}}} \mathrm{KL}(f_A \mid f_C), \quad \text{and} \quad \mathrm{KL}_{A,*}^{\mathcal{W}} := \max_{B \in \mathcal{A}} \mathrm{KL}_{A,B}^{\mathcal{W}}. \tag{18}$$

也就是说，给定真实信号子集 A，$\mathrm{KL}_{A,\,D}^{\mathcal{W}}$ 是所有会导致决策 D 错误的信号子集 C 中，f_A 与 f_C 之间的最小KL散度；而 $\mathrm{KL}_{A,\,*}^{\mathcal{W}}$ 则是在所有 $D \in \mathrm{A}$ 上的 $\mathrm{KL}_{A,\,D}^{\mathcal{W}}$ 中的最大值。

定义 4。设 $A \in \mathsf{A}$。若 $\mathrm{KL}_{A,*}^{\mathcal{W}}$ 在唯一的子集 $D_A^{\mathcal{W}} \in \mathsf{A}$ 处取得，我们称 A 关于 W 具有唯一最有利子集，即：

$$\mathrm{KL}_{A,D_A^{\mathcal{W}}}^{\mathcal{W}} = \mathrm{KL}_{A,*}^{\mathcal{W}} > \mathrm{KL}_{A,D}^{\mathcal{W}}, \quad \text{for all } D \in \mathcal{A} \setminus \{D_A^{\mathcal{W}}\}.$$

注 5. 我们称 $\mathcal{W}$ 为零一损失，若 $\mathcal{W}(D \mid A) > 0$ 对所有 $D \neq A$ 满足 $A,\ D \in \mathcal{A}$；也就是说，损失为正，除非 $D = A$。若 $\mathcal{W}$ 为零一损失，则显然 $\mathcal{H}_D^{\mathcal{W}} = \mathcal{A} \setminus \{D\}$。现固定 $A \in \mathcal{A}$。在假设1下，我们有

$$\mathrm{KL}_{A,A}^{\mathcal{W}} > 0, \quad \mathrm{KL}_{A,D}^{\mathcal{W}} \leq \mathrm{KL}(f_A \mid f_A) = 0, \qquad \text{for all } D \neq A.$$

因此，在零一损失 W 下，每个 $A \in \mathsf{A}$ 都容许一个唯一最有利子集（关于 W），即 $D_A^{\mathcal{W}} = A$。

设 $A \in \mathsf{A}$ 容许一个唯一最有利子集（关于 W）。我们记

$$C_A^{\mathcal{W}} := \left\{ C \in \mathcal{H}_{D_A^{\mathcal{W}}}^{\mathcal{W}} : \mathrm{KL}(f_A \mid f_C) = \mathrm{KL}_{A,*}^{\mathcal{W}} \right\} \quad \text{and} \quad r_A^{\mathcal{W}} := \left| C_A^{\mathcal{W}} \right| \tag{19}$$

在 $\mathrm{H}_{D_A^{\mathcal{W}}}^{\mathcal{W}}$ 上达到最小 KL散度 的信号子集的集合及其基数。在 $\mathbb{P}_A$ 下所得的渐近行为会因 $r_A^{\mathcal{W}} = 1$ 还是 $r_A^{\mathcal{W}} \geq 2$ 而在性质上有所不同。我们将前者称为非对称 情形，将后者称为对称 情形。

现在，令 $\left(C_{A,1}^{\mathcal{W}}, \ldots\ldots, C_{A,r_A^{\mathcal{W}}}^{\mathcal{W}} \right)$ 是 $\mathsf{C}_A^{\mathcal{W}}$ 中各元素的有序列表。我们定义一个 $r_A^{\mathcal{W}}$ 维随机向量 $\boldsymbol{R}_A^{\mathcal{W}}$，其中对每个 $1 \leq j \leq r_A^{\mathcal{W}}$，

$$\boldsymbol{R}_{A,j}^{\mathcal{W}} := \log f_A\left(\mathbf{X}_1^{[K]}\right) - \log f_{C_{A,j}^{\mathcal{W}}}\left(\mathbf{X}_1^{[K]}\right) - \mathrm{KL}_{A,*}^{\mathcal{W}}.$$

注意到 $\mathbb{E}_A(\boldsymbol{R}_A^{\mathcal{W}}) = \mathbf{0}$，并记

$$\boldsymbol{\Sigma}_A^{\mathcal{W}} := \mathbb{E}_A\left[\boldsymbol{R}_A^{\mathcal{W}} (\boldsymbol{R}_A^{\mathcal{W}})^\top\right], \tag{20}$$

$r_{\mathcal{W}_A} \times r_A^{\mathcal{W}}$ 协方差矩阵，表示当 A 为真实信号子集时 $\boldsymbol{R}_A^{\mathcal{W}}$ 的协方差矩阵。最后，令 $\boldsymbol{Y}_{\mathcal{W}A} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_A^{\mathcal{W}})$，即均值向量为 $\mathbf{0}$、协方差矩阵为 $\boldsymbol{\Sigma}_A^{\mathcal{W}}$ 的 $r_A^{\mathcal{W}}$ 维正态分布。记其最大分量的期望值为

$$h_A^{\mathcal{W}} := \mathbb{E}_A\left[\max\left\{\boldsymbol{Y}_{A,1}^{\mathcal{W}}, \ldots, \boldsymbol{Y}_{A,r_A^{\mathcal{W}}}^{\mathcal{W}}\right\}\right]. \tag{21}$$

备注 6。当 $r_{\mathcal{W}_A} = 1$ 时，$h_A^{\mathcal{W}} = 0$。对于对称情形 $r_A^{\mathcal{W}} \geq 2$，我们有 $h_A^{\mathcal{W}} \geq 0$，可通过蒙特卡洛方法从 $\mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_A^{\mathcal{W}})$ 中抽取随机向量，并对每个样本中的最大分量取平均值来估计。

下面，我们刻画 $\Delta(\boldsymbol{\theta}),\ \boldsymbol{\theta} \in (0,\ 1)^r$ 的最小可达期望样本量（ESS）的二阶近似。对于稍后将确定的某个 $q \geq 2$，我们假设

$$\int \left|\log(f_i^k / f_{i'}^k)\right|^q f_i^k \, d\mu^k < \infty, \quad \text{for } (i, i') \in \{(0,1),(1,0)\} \text{ and } k \in [K]. \tag{22}$$

此外，我们考虑这样一种情形：对于每个 $\boldsymbol{\theta} \in (0,\ 1)^r$，一个程序是否属于 $\Delta(\boldsymbol{\theta})$ 仅取决于其决策规则，而不取决于其停止规则。

假设 2. 令 $\delta = (T, D) \in \Delta$ 和 $\delta' = (T', D') \in \Delta$为任意序贯程序。若对于某个 $\boldsymbol{\theta} \in (0,1)^r$ 和 $D = D'$，有 $\delta \in \Delta(\boldsymbol{\theta})$，则 $\delta' \in \Delta(\boldsymbol{\theta})$。

这一假设适用于若干类序贯程序，例如 Class1–4，这些程序是根据错误决策所导致的错误概率定义的。下述定理的证明见附录 B。

定理*2.* 假设假设*1*和*2*成立。设 $\delta_0(\boldsymbol{\theta}) = (T_0(\boldsymbol{\theta}),\ D_0(\boldsymbol{\theta}))$，$\boldsymbol{\theta} \in (0,\ 1)^r$，为一组程序，使得对每个 $\boldsymbol{\theta}$，均有 $\delta_0(\boldsymbol{\theta}) \in \Delta(\boldsymbol{\theta})$。假设存在对每个$\boldsymbol{\theta} \in (0,\ 1)^r$而言的代价 $c_{\boldsymbol{\theta}} > 0$ 、损失函数 $W : A \times A \to [0, \infty)$，以及常数 $L > 0$，使得 *(12)*、*(13)* 和 *(14)* 成立。此外，假设当 $\boldsymbol{\theta} \to \mathbf{0}$ 时，$c_{\boldsymbol{\theta}} \to 0$ 。

令 $A \in \mathcal{A}$，并假设 A 相对于 $\mathcal{W}$ 存在唯一最有利子集。

(a) 考虑情形 $r_A^{\mathcal{W}} = 1$。假设 (22) 对 $q = 3$成立。那么，当 $\boldsymbol{\theta} \to \mathbf{0}$，

$$T_A^{\min}(\Delta(\boldsymbol{\theta})) = \frac{\log(c_{\boldsymbol{\theta}}^{-1})}{\mathrm{KL}_{A,*}^{\mathcal{W}}} + O(1).$$

(b) 考虑情形 $r_A^{\mathcal{W}} \geq 2$。对于 $\epsilon \in (0, 1/2)$，假设 (22) 对某个满足以下条件的 q 成立

$$q > \epsilon^{-1} \quad \text{and} \quad q \geq 3. \tag{23}$$

那么，当 $\boldsymbol{\theta} \to \mathbf{0}$，

$$T_A^{\min}(\Delta(\boldsymbol{\theta})) = \frac{\log(c_{\boldsymbol{\theta}}^{-1})}{\mathrm{KL}_{A,*}^{\mathcal{W}}} + \frac{h_A^{\mathcal{W}} \sqrt{\log(c_{\boldsymbol{\theta}}^{-1})}}{(\mathrm{KL}_{A,*}^{\mathcal{W}})^{3/2}} + O\left((\log(c_{\boldsymbol{\theta}}^{-1}))^{1/4+\epsilon/2}\right).$$

这里有几点值得说明。首先，假设 1 和 (22) 对对数似然比施加了矩条件。特别地，当 $q \geq 2$ 时，假设 1 的第二部分可由 (22) 推出。其次，条件 $c_{\boldsymbol{\theta}} \to 0$ 当 $\boldsymbol{\theta} \to 0$ 以及假设 2 都是自然的，并且对大多数程序类别都成立，尤其是本文讨论的类别 1–7。第三，在应用定理 2之前，我们通常会考虑一组程序 $\delta_0(\boldsymbol{\theta}) \in \Delta(\boldsymbol{\theta})$，其中 $\boldsymbol{\theta} \in (0,\ 1)^r$，并通过验证对于适当选择的 $c_{\boldsymbol{\theta}}$、W 和 L，(12)、(13) 和 (14) 成立，进而利用定理 1 确立它们的二阶最优性。由于定理 1 中已经完成了这项验证，因此在此阶段应用定理2无需额外工作。

因此，要应用定理2，关键步骤是验证给定的 $A \in \mathcal{A}$ 相对于$\mathcal{W}$存在唯一最有利子集，而W是针对某个感兴趣的错误度量而设定的。在第4 节和附录D.1中，我们证明，对于类别1、3和4，每个 $A \in \mathcal{A}$ 相对于相应的$\mathcal{W}$都存在唯一最有利子集。然而，对于类别2，我们通过一个反例证明这一性质未必成立，并进一步推导出确保该性质成立的充分条件。

注7。 在定理2的证明中，我们定义了一个程序 $\delta_1(\boldsymbol{\theta})$，它与 $\delta_{\mathrm{Ld}}(c_{\boldsymbol{\theta}}, \mathcal{W})$ 采用相同的停止规则，并与 $\delta_0(\boldsymbol{\theta})$采用相同的决策规则。这有两个目的：(i) 采用相同的决策规则，使我们无需验证 $\delta_{\mathrm{Ld}}(c_{\boldsymbol{\theta}}, \mathcal{W}) \in \Delta(\boldsymbol{\theta})$；(ii) 采用相同的停止规则可确保，根据

定理 1，在 $\Delta(\boldsymbol{\theta})$ 上的最小ESS与 $\mathbb{E}_A[T_{\mathrm{Ld}}(c_{\boldsymbol{\theta}},\mathcal{W})]$ 相差至多一个 $O(1)$ 项。因此，问题归结为推导特定规则 $\delta_{\mathrm{Ld}}(c_{\boldsymbol{\theta}},\mathcal{W})$的期望样本量二阶近似。

此外，由于 A 存在唯一最有利子集 $D_A^{\mathcal{W}}$，我们证明 $\mathbb{E}_A[T_{\mathrm{Ld}}(c_{\boldsymbol{\theta}},\ \mathrm{W})]$ =$\mathbb{E}_A[\tau_A^{\mathcal{W}}(c_{\boldsymbol{\theta}})] + O(1)$，其中对于 $c>0$,

$$\tau_A^{\mathcal{W}}(c)=\min\left\{t\geq 1:\boldsymbol{R}_{t,G}\geq\log(c^{-1})\ \text{for all}\ G\in\mathcal{H}_{D_A^{\mathcal{W}}}^{\mathcal{W}}\right\},$$

且 $\boldsymbol{R}_{t,\ G}:=\sum_{s=1}^{t}\log\left(f_A(\mathrm{X}_s^{[K]})/\ f_G(\mathrm{X}_s^{[K]})\right)$。因此，$\tau_A^{\mathcal{W}}(c)$ 是多维随机游走 $\{\boldsymbol{R}_t:t\geq 1\}$ 的所有分量首次越过水平 $\log(c^{-1})$ 的时间。最后，我们应用 [27] 中发展的非线性更新理论来近似 $\mathbb{E}_A[\tau_A^{\mathcal{W}}(c)]$ （见附录 C）。

注 8。对于非对称情形 $r_A^{\mathcal{W}}=1$，[11] 中的定理 3.1 给出了 $\mathbb{E}_A[\tau_A^{\mathcal{W}}(c)]$ 的渐近近似，其精度可达 $o(1)$ 项；该结果假设对于每个 G，增量 $\boldsymbol{R}_{1,\ G}$ 是 $\mathbb{P}_A$-非格点的。然而，在本研究中，由于我们关注二阶最优性，只需将误差项刻画到 $O(1)$ 的精度，因此不需要非格点条件。

更重要的是，在对称情形 $r_A^{\mathcal{W}}\geq 2$中，定理 3.3 在 [11] 中的适用性似乎存疑。其中的论证经由 [39]的定理 3 调用了非线性更新理论。然而，如附录C.3所解释的，[39] 中所要求的一个关键正则性条件在当前设定下并不满足。据我们所知，$\mathbb{E}_A[\tau_A^{\mathcal{W}}(c)]$ 在对称情形下的严格二阶展开尚未建立，而获得 $O(1)$精度的展开仍然是一个未决问题。

4. 不同错误度量下的应用

在本节中，我们将主要结果应用于第1–3类，它们对应于不同的误差度量，并且不假定先验知识，即$\mathcal{A}=2^{[K]}$。在附录D.1中，我们将这些结果应用于第4类，该类假定信号数量已知。其他类别的序贯程序在附录D.2中加以讨论。

对于每一类，我们首先回顾文献中的一个程序，该程序提供显式的误差控制，并且已被证明能够达到一阶频率学派意义上的渐近最优性（见定义1）。随后，我们证明其二阶最优性，方法是应用定理1。此外，利用定理2，我们刻画了每个信号子集下的二阶最小期望样本量（ESS）。

回顾在（$\{f_0^k,\ f_1^k:\ k\in[K]\}$ 2）中定义的密度函数。对于每个数据流 $k\in[K]$ 和时间 $t\geq 1$，定义截至时间 t 的对数似然比（LLR）统计量（即，限制在$\mathcal{F}_t$内）：

$$\lambda_t^k:=\log\frac{dP_1^k}{dP_0^k}\bigg|_{\mathcal{F}_t}=\sum_{s=1}^{t}\log\frac{f_1^k(\mathbf{X}_s^k)}{f_0^k(\mathbf{X}_s^k)}. \tag{24}$$

4.1. 广义误分类率

在本小节中，我们关注第 1 类 $\Delta_{m_0}^{\mathrm{gmr}}$（$\alpha$），其将广义误分类率控制在水平 $\alpha\in$（0，1），且不假设拥有先验信息，即 $\mathcal{A}=2^{[K]}$。我们首先回顾和交规则，记作 δ_S（b）=（T_S（b），D_S（b））， 该规则由 [32],提出，其中 $b>0$ 是一个阈值参数。对于每个 $t\geq 1$，记

$$0<\tilde{\lambda}_t^{(1)}\leq\cdots\leq\tilde{\lambda}_t^{(K)} \tag{25}$$

对数似然比绝对值 $\{|\lambda_t^k|: k \in [K]\}$ 的次序统计量。注意，如果 λ_t^k的绝对值较小，则它在时间 t不能为零假设或备择假设提供有力证据。于是，对于 $b > 0$,

$$T_S(b) := \min\left\{t \geq 1 : \sum_{i=1}^{m_0} \tilde{\lambda}_t^{(i)} \geq b\right\}, \quad D_S(b) := \left\{k \in [K] : \lambda_{T_S(b)}^k > 0\right\}. \tag{26}$$

也就是说，程序在最小的 m_0 个对数似然比之和超过 b 时首次停止；停止后，所有对数似然比为正的数据流中的原假设均被拒绝。

定理 3.1 证明了 [32] 对于 $\alpha > 0$，若阈值选为

$$b_\alpha = |\log \alpha| + \log \binom{K}{m_0}, \tag{27}$$

则 δ_S（b_α）$\in \Delta_{m_0}^{\mathrm{gmr}}$（$\alpha$）。采用这一阈值后，族$\{\delta_S$（$b_\alpha$）$\}$在 $\alpha \to 0$ 时一阶渐近最优（见 [32]中的定理 3.3）。下一个定理证明$\{\delta_S$（b_α）的二阶最优性。这里给出证明，以说明定理1的应用。

定理 3。假设条件*1* 成立。对于每个 $\alpha > 0$，将阈值选为 b_α 在(27)中的值。则和交规则$\{\delta_S$（b_α）在$\{\Delta_{m_0}^{\mathrm{gmr}}$（$\alpha$）中二阶渐近最优，当$\alpha \to 0$时；也就是说，对于每个 $A \subset [K]$,

$$\limsup_{\alpha \to 0}\left(\mathbb{E}_A[T_S(b_\alpha)] - T_A^{\min}\left(\Delta_{m_0}^{\mathrm{gmr}}(\alpha)\right)\right) < \infty.$$

证明。令 π_0为 $2^{[K]}$上的均匀分布，即对于 $A \subset [K]$，有 $\pi_0(A) = 1/2^K$。此外，我们定义以下损失函数：对于 A, $D \subset [K]$,

$$\mathcal{W}(D \mid A) = \begin{cases} 1 & \text{if } |D \triangle A| \geq m_0 \\ 0 & \text{else} \end{cases}. \tag{28}$$

也就是说，当且仅当决策 D与 A至少在 m_0个分量上不同时，$W(D|A) = 1$。此外，我们定义

$$c_\alpha := (2^K e^{b_\alpha})^{-1} = 2^{-K}\binom{K}{m_0}^{-1}\alpha, \quad L := 2^K \binom{K}{m_0}. \tag{29}$$

一旦验证上述损失函数 W、代价 c_α 和常数 L满足条件（12）–（14），结论便由定理 1 立即得出。

我们先验证条件（12）。任取 $D \subset [K]$。令 $A = D^c$，则 $D \triangle A = [K]$。由于$m_0 \leq K$（见定义 1），有 $\mathcal{W}(D|A) = 1$。因此，条件（12）成立。

验证条件（13）需要证明 $T_S(b_\alpha)$的停止时间不晚于 $T_{\mathrm{Ld}}(c_\alpha; \mathcal{W})$。根据定义，对于每个 $\alpha \in (0, 1)$，有 $b_\alpha = \log(2^{-K} c_\alpha^{-1})$且 $c_\alpha < 2^{-K}$。于是，条件（13）直接由附录 E.2中的引理 15得出。

最后，我们验证条件 (14)。固定任意程序 $\delta = (T, D) \in \Delta_{m_0}^{\mathrm{gmr}}(\alpha)$。根据类别 1 的定义，有 $\mathbb{P}_A(|A \triangle D| \geq m_0) \leq \alpha$ 对每个 $A \subset [K]$ 均成立。然后，根据 $\mathcal{W}$ 的定义，

$$\mathrm{ERR}^{\mathrm{int}}(\delta; \mathcal{W}) = \sum_{A \in \mathcal{A}} \frac{1}{2^K} \mathbb{P}_A(|A \triangle D| \geq m_0) \leq \sum_{A \in \mathcal{A}} \frac{\alpha}{2^K} = \alpha = L c_\alpha.$$

因此，条件（14）成立。证明完毕。 □

接下来，我们考虑每个 $T_A^{\min}\left(\Delta_{m_0}^{\mathrm{gmr}}(\alpha)\right)$的可实现的最小 ESS 的渐近近似，其中 $A \subset [K]$ 为 $\alpha \to 0$。具体而言， [32] 的定理 3.3 表明，对于每个 $A \subset [K]$，其一阶性能满足

$$T_A^{\min}\left(\Delta_{m_0}^{\mathrm{gmr}}(\alpha)\right) = \frac{|\log \alpha|}{\mathrm{KL}_{A,A}^{\mathcal{W}}}(1 + o(1)), \tag{30}$$

其中 KL散度$_{A}^{\mathcal{W}}$，A 定义见（18），其中 $\mathcal{W}$ 见（28），即

$$\mathrm{KL}_{A,A}^{\mathcal{W}} = \min\{\mathrm{KL}(f_A \mid f_C) : C \subset [K],\ |A \triangle C| \geq m_0\}.$$

在本文中，我们通过应用定理2，推导出更精确的二阶近似。关键步骤是验证每个 $A \subset [K]$ 相对于$\mathcal{W}$（定义见（28））都具有唯一最有利子集，具体如下引理所述。

注 9。 注意，当 $m_0 \geq 2$时，$\mathcal{W}$ 在（28）中不是注5中定义的零一损失。

Recall the definition of $\mathrm{KL}_{A,D}^{\mathcal{W}}$ and $\mathrm{KL}_{A,*}^{\mathcal{W}}$ in (18).

引理 1。 假设*1* 成立。令 $m_0 \in [1, K]$ 为一个整数，并令$\mathcal{W}$ 定义于（28）。对于每个 $A \subset [K]$,

$$\mathrm{KL}_{A,A}^{\mathcal{W}} > \mathrm{KL}_{A,D}^{\mathcal{W}}, \quad \textit{for any } D \subset [K] \textit{ such that } D \neq A.$$

因此，每个 $A \subset [K]$ 相对于$\mathcal{W}$ 且满足 $D_A^{\mathcal{W}} = A$时，都具有唯一最有利子集，并且KL散度$_{A}^{\mathcal{W}}$，$_{*}$ =KL散度$_{A}^{\mathcal{W}}$，$_{A}$。

证明。 见附录E.2。 □

定理 4. 假设 *1* 成立。令 $\mathcal{W}$ 定义见 (28)，并令 $A \subset [K]$ 为任意信号子集。

(a) 考虑非对称情形 $r_A^{\mathcal{W}} = 1$。假设 (22) 对 $q = 3$成立。那么，当 $\alpha \to 0$,

$$T_A^{\min}\left(\Delta_{m_0}^{\mathrm{gmr}}(\alpha)\right) = \frac{|\log \alpha|}{\mathrm{KL}_{A,A}^{\mathcal{W}}} + O(1).$$

(b) 考虑对称情形 $r_A^{\mathcal{W}} \geq 2$。令 $\epsilon \in (0,\ 1/2)$，并假设 (22) 对某个 q满足 (23)。回忆一下，$h_A^{\mathcal{W}}$定义见 (21)。那么，当 $\alpha \to 0$,

$$T_A^{\min}\left(\Delta_{m_0}^{\mathrm{gmr}}(\alpha)\right) = \frac{|\log \alpha|}{\mathrm{KL}_{A,A}^{\mathcal{W}}} + \frac{h_A^{\mathcal{W}}\sqrt{|\log \alpha|}}{(\mathrm{KL}_{A,A}^{\mathcal{W}})^{3/2}} + O\left((\log \alpha)^{1/4+\epsilon/2}\right).$$

证明。正如上文所述，对于每个 $\alpha \in (0, 1)$，有 $b_\alpha$in (27)、$\delta_S(b_\alpha) \in \Delta_{m_0}^{\mathrm{gmr}}(\alpha)$。此外，我们已在定理 3 的证明中证明，使用式 (28) 中的损失函数 W、式 (29) 中的代价c_α以及常数 L 时，条件 (12)–(14) 成立。显然，当 $\alpha \to 0$ 时， $c_\alpha \to 0$，且假设 2 成立。

最后，引理 1 表明，每个 A 都存在唯一的有利子集，满足 $D_A^{\mathcal{W}} = A$ 且 KL$_{A}^{\mathcal{W}}$，$_{*}$ =KL$_{A}^{\mathcal{W}}$，$_{A}$。随后由定理2完成证明。 □

4.2. 广义族错误率

在本小节中，我们关注第 2 类，$\Delta^{\mathrm{gfr}}_{m_1,\,m_2}(\alpha,\ \beta)$，该类控制广义家族错误率中的假阳性率和假阴性率分别低于 α 和 β，并假设没有先验信息，即 $\mathrm{A} = 2^{[K]}$。我们首先回顾 Leap规则，记作 $\delta_L(a,\ b) = (T_L(a,\ b),\ D_L(a,\ b))$，该规则由 [32], 提出，其中 $a,\ b > 0$ 是阈值参数。

回顾式（$\{\lambda_t^k\colon\ k \in [K],\ t = 1,\ 2,\ \ldots\}$ 24）中定义的对数似然比。对于每个时间 $t \geq 1$，记

$$0 < \hat{\lambda}_t^{(1)} \leq \cdots \leq \hat{\lambda}_t^{(p(t))}$$

正对数似然比的次序统计量 $\{\lambda_t^k\colon \lambda_t^k > 0,\ \ k \in [K]\}$，其中 $p(t)$ 表示时间 t时严格为正的对数似然比的数量。同样，记

$$0 < \check{\lambda}_t^{(1)} \leq \cdots \leq \check{\lambda}_t^{(K-p(t))}$$

非正对数似然比绝对值的次序统计量 $\{|\lambda_t^k|\colon \lambda_t^k \leq 0,\ \ k \in [K]\}$。

我们还将相应的数据流索引记为 $\{\hat{l}_1(t), \ldots, \hat{l}_{p(t)}(t)\}$ 以及由 $\{\check{l}_1(t), \ldots, \check{l}_{K-p(t)}(t)\}$ 所构成的集合，满足 $|\lambda_{\hat{l}_j(t)t}| = \hat{\lambda}_t^{(j)}$ 对于 $j \in [p(t)]$ 以及 $|\lambda_{\check{l}_j(t)t}| = \check{\lambda}_t^{(j)}$ 对于 $j \in [K - p(t)]$。对于统计量 $\hat{\lambda}_t^{(i)}$，较小的索引 i 表示拒绝第 k个原假设的证据较弱（因此第一类错误的风险较高）。类似地，对于 $\check{\lambda}_t^{(i)}$，较小的索引 i 表示接受第k个原假设的证据较弱（因此第二类错误的风险较高）。

对于阈值 $a,\ b > 0$，我们首先定义如下规则集合：对于 $i = 0,\ \ldots\ldots,\ m_1 - 1$,

$$\hat{\tau}_i := \min\left\{ t \geq 1 : \sum_{j=1}^{m_1 - i} \hat{\lambda}_t^{(j)} \geq b,\ \sum_{j=i+1}^{i+m_2} \check{\lambda}_t^{(j)} \geq a \right\},\ \ \hat{D}_i := \{k \in [K] : \lambda^k_{\hat{\tau}_i} > 0\} \ \cup\ \{\check{\ell}_1(\hat{\tau}_i), \ldots, \check{\ell}_i(\hat{\tau}_i)\},$$

对于 $i = 1,\ \ldots\ldots,\ m_2 - 1$,

$$\check{\tau}_i := \min\left\{ t \geq 1 : \sum_{j=i+1}^{i+m_1} \hat{\lambda}_t^{(j)} \geq b,\ \sum_{j=1}^{m_2 - i} \check{\lambda}_t^{(j)} \geq a \right\},\ \ \check{D}_i := \{k \in [K] : \lambda^k_{\check{\tau}_i} > 0\} \ \setminus\ \{\hat{\ell}_1(\check{\tau}_i), \ldots, \hat{\ell}_i(\check{\tau}_i)\}.$$

The 为简洁起见，省略了对 $(a,\ b)$ 的依赖。此外，我们采用如下约定 th at $\hat{\lambda}_t^k = \infty$ 对于 $k > p(t)$ 和 $\check{\lambda}_t^k = \infty$ 对于 $k > K - p(t)$。

接下来，参数为 $a,\ b > 0$ 的 Leap规则 $\delta_L(a,\ b)$ 定义如下：

$$T_L(a,b) := \min\{\hat{\tau}_0, \cdots, \hat{\tau}_{m_1-1}, \check{\tau}_1, \ldots, \check{\tau}_{m_2-1}\},\quad D_L(a,b) = \begin{cases} \hat{D}_i, & \text{if } \hat{\tau}_i = T_L(a,b), \\ \check{D}_i, & \text{if } \check{\tau}_i = T_L(a,b), \end{cases} \tag{31}$$

其中，如果最小值由多个停止规则同时取得，则 D_L 定义为相应决策的并集。关于 Leap规则 的动机和讨论，请感兴趣的读者参阅 [32] 的第4节。

定理 4.1 在 [32] 中证明，对于任意 $\alpha,\ \beta \in (0,\ 1)$，若阈值选为

$$a_{\alpha,\beta} = |\log\beta| + \log\left[2^{m_2}\binom{K}{m_2}\right],\quad b_{\alpha,\beta} = |\log\alpha| + \log\left[2^{m_1}\binom{K}{m_1}\right], \tag{32}$$

则 $\delta_L(a_{\alpha,\beta}, b_{\alpha,\beta})$属于类 $\Delta^{\mathrm{gfr}}_{m_1,m_2}(\alpha,\beta)$。此外，采用这组阈值时，族 $\{\delta_L(a_{\alpha,\beta}, b_{\alpha,\beta})\}$在 $\alpha\vee\beta\to 0$(参见 [32]中的定理 4.3)。

下一个定理证明，当 α 和 β 以相同速率趋于零时，该方法具有二阶最优性；具体而言，我们假设

$$C^{-1}\alpha\le\beta\le C\alpha,\quad \text{for some absolute constant } C>0. \tag{33}$$

定理 5. 假设条件 *1* 成立，且 $\alpha,\beta\in(0,1)$ 满足条件 (33)。那么，*Leap*规则$\{\delta_L(a_{\alpha,\beta}, b_{\alpha,\beta})\}$在 $\{\Delta^{\mathrm{gfr}}_{m_1,m_2}(\alpha,\beta)\}$ 中是二阶渐近最优的，其中 $\alpha,\beta\to 0$，也就是说，对每个 $A\subset[K]$,

$$\limsup_{\alpha,\beta\to 0}\left(\mathbb{E}_A[T_L(a_{\alpha,\beta},b_{\alpha,\beta})]-T_A^{\min}\left(\Delta^{\mathrm{gfr}}_{m_1,m_2}(\alpha,\beta)\right)\right)<\infty.$$

证明。 证明见附录E.3。 □

注 10。 对于一阶渐近最优性， α 和 β 可以以任意相对速率趋于零；然而，在上述定理中，它们必须以相同速率趋于零。

在定理 5的证明中，我们将定理 1 应用于以下损失函数：对于 A， $D\subset[K]$,

$$\mathcal{W}(D\mid A)=\begin{cases}1, & \text{if } |D\setminus A|\ge m_1 \text{ or } |A\setminus D|\ge m_2,\\ 0, & \text{otherwise.}\end{cases} \tag{34}$$

当且仅当真实信号子集为 A时，决策 D 至少会导致 m_1 个第一类错误或至少 m_2 个第二类错误（或两者兼有），此时该值等于 1。

接下来，我们考虑可实现的最小 ESS 的渐近近似， $T_A^{\min}\left(\Delta^{\mathrm{gfr}}_{m_1,m_2}(\alpha,\beta)\right)$，对于每个 $A\subset[K]$ 满足 α， $\beta\to 0$ 且满足条件（33）的情形。具体而言，在满足条件（33）时，[32] 的定理 4.3 表明，对于每个 $A\subset[K]$，一阶性能满足

$$T_A^{\min}\left(\Delta^{\mathrm{gfr}}_{m_1,m_2}(\alpha,\beta)\right)=\frac{|\log\alpha|}{\mathrm{KL}^{\mathcal{W}}_{A,*}}(1+o(1)),$$

其中 $\mathrm{KL}^{\mathcal{W}}_{A,*}$ 在（18）中定义，其中 $\mathcal{W}$ 由（34）给出，即

$$\mathrm{KL}^{\mathcal{W}}_{A,*}=\max_{D\subset[K]}\mathrm{KL}^{\mathcal{W}}_{A,D},\ \text{with}\ \mathrm{KL}^{\mathcal{W}}_{A,D}=\min\{\mathrm{KL}(f_A|f_C): C\subset[K],\ |D\setminus C|\ge m_1 \text{ or } |C\setminus D|\ge m_2\}.$$

本文旨在通过应用定理 2，推导出更精确的二阶近似。

定理 6。 假设假设 *1* 成立，且 α， $\beta\in(0,\ 1)$ 满足条件 (33)。令 $A\in\mathcal{A}$，并假设 A 相对于（34）中的 W 有唯一最有利子集。

(a) 考虑非对称情形 $r_A^{\mathcal{W}}=1$。假设 (22) 对 $q=3$成立。那么，当 $\alpha\to 0$,

$$T_A^{\min}\left(\Delta^{\mathrm{gfr}}_{m_1,m_2}(\alpha,\beta)\right)=\frac{|\log(\alpha)|}{\mathrm{KL}^{\mathcal{W}}_{A,*}}+O(1).$$

(b) 考虑对称情形 $r_A^{\mathcal{W}} \geq 2$。令 $\epsilon \in (0,\ 1/2)$，并假设 (22) 对某个 q 满足 (23)。那么，当 $\alpha \to 0$,

$$T_A^{\min}\left(\Delta_{m_1,m_2}^{\mathrm{gfr}}(\alpha,\beta)\right) = \frac{|\log(\alpha)|}{\mathrm{KL}_{A,*}^{\mathcal{W}}} + \frac{h_{A,*}^{\mathcal{W}}\sqrt{|\log(\alpha)|}}{(\mathrm{KL}_{A,*}^{\mathcal{W}})^{3/2}} + O\left((\log(\alpha))^{1/4+\epsilon/2}\right).$$

证明。正如上文所述，对于 (32) 中的 $a_{\alpha,\beta}, b_{\alpha,\beta}$，每个 $\alpha,\beta \in (0,\ 1)$ 都有 $\delta_L(a_{\alpha,\beta}, b_{\alpha,\beta}) \in \Delta_{m_1,m_2}^{\mathrm{gfr}}(\alpha,\beta)$。此外，我们在定理 5 的证明中表明，对于 (34) 中的损失函数 W，以及 (81) 中定义的代价 $c_{\alpha,\beta}$和常数 L，条件 (12)–(14) 成立。显然，$c_{\alpha,\beta} \to 0$随着 $\alpha,\beta \to 0$，并且假设 2 成立。由于假设 A 相对于 W 存在唯一最有利子集，定理2即完成证明。 □

与第1类不同，第 1 类控制广义误分类率（见引理1），对于$\mathcal{W}$（34）中定义的 W，不能保证每个 $A \subset [K]$ 都按照定义4的含义存在唯一最有利子集，以下反例对此作了说明。

例 1. 设 $K = 3$ 与 $m_1 = m_2 = 2$。考虑完全齐次情形 $\mathcal{I}_i^j = \mathcal{I} > 0$，对于$i \in \{0,\ 1\}$且 $j \in \{1,\ 2,\ 3\}$，其中信息数定义于（15）。现考虑 $A = \emptyset$。根据（18）中的定义以及$\mathcal{W}$（34），

$$\mathrm{KL}_{A,*}^{\mathcal{W}} = 3\mathcal{I} = \mathrm{KL}_{A,\{1\}}^{\mathcal{W}} = \mathrm{KL}_{A,\{2\}}^{\mathcal{W}} = \mathrm{KL}_{A,\{3\}}^{\mathcal{W}}.$$

因此，三个子集 {1}、{2} 与 {3} 在 $\max_{D\subset[K]}$ KL散度$_A^{\mathcal{W}}$，D 中取得最大值。

接下来，我们给出充分条件，在这些条件下 $A \subset [K]$ 关于 $\mathcal{W}$（定义于（34））具有唯一最有利子集。

引理 2. 假定假设 *1* 成立。则 $A \subset [K]$ 关于$\mathcal{W}$（见（34））具有唯一最有利子集，如果下列条件之一成立：

(a) $m_1 = m_2 = 1$。

(b) $m_1 = m_2 := m^*$, $\mathcal{I}_0^k = \mathcal{I}_1^k := \mathcal{I}^*$ for $k \in [K]$, and $\min\{|A|, |A^c|\} \geq m^*$.

在任一情形下，$D_A^{\mathcal{W}} = A$，且 KL散度$_A^{\mathcal{W}}$，$_*$ = KL散度$_A^{\mathcal{W}}$，$_A$。

证明。 证明见附录E.3。 □

备注 11。 当 $m_1 = m_2 = 1$时，对应于经典的家族错误率，定理6因此提供了二阶精确的渐近近似。然而，对于广义族错误率，这通常仍是一个尚未解决的问题。

4.3. 错误发现率与未发现率

在本小节中，我们考虑第3类，$\Delta^{\mathrm{fdr}}(\alpha,\beta)$，它控制错误发现率与错误非发现率，并假设 $\mathcal{A} = 2^{[K]}$。我们首先回顾交集规则，记为 $\delta_I(a,b) = (T_I(a,b),\ D_I(a,b))$，该规则在 [7], 中提出，其中 a、$b > 0$ 为阈值参数。具体而言，

$$T_I(a,b) = \min\{t \geq 1 : \lambda_t^k \notin (-a,b) \text{ for all } k \in [K]\}, \qquad D_I(a,b) = \{k \in [K] : \lambda_{T_I(a,b)}^k > 0\}. \quad (35)$$

在 [15] 的推论 4.1 中表明，对于 α， $\beta \in (0,1)$，如果

$$a_{\alpha,\beta} = |\log(\beta)| + \log(K), \quad b_{\alpha,\beta} = |\log(\alpha)| + \log(K), \tag{36}$$

那么 $\delta_I(a_{\alpha,\beta}, b_{\alpha,\beta}) \in \Delta^{\mathrm{fdr}}(\alpha, \beta)$。此外，根据同一推论，采用这组阈值时，族 $\{\delta_L(a_{\alpha,\beta}, b_{\alpha,\beta})\}$ 在 $\alpha \vee \beta \to 0$ 下是一阶渐近最优的。

下一个定理证明，当 α 与 β 以相同速率趋于零时，即当条件（33）成立时，它是二阶最优的。

定理 7。假设条件 *1* 成立，且 $\alpha, \beta \in (0, 1)$ 满足条件 (33)。那么，交集规则 $\{\delta_I(a_{\alpha,\beta}, b_{\alpha,\beta})\}$ 在 $\{\Delta^{\mathrm{fdr}}(\alpha,\beta)\}$ 中，当 $\alpha, \beta \to 0$ 时是二阶渐近最优的；也就是说，对每个 $A \subset [K]$,

$$\limsup_{\alpha,\beta\to 0} \left(\mathbb{E}_A[T_I(a_{\alpha,\beta}, b_{\alpha,\beta})] - T_A^{\min}\left(\Delta^{\mathrm{fdr}}(\alpha,\beta)\right) \right) < \infty.$$

证明。证明见附录E.4。 □

注释 12。在定理 7的证明中，我们使用以下类别 3与类别2 之间的关系，其中 $m_1 = m_2 = 1$，该关系最早由 [15]提出：

$$\Delta^{\mathrm{gfr}}_{1,1}(\alpha,\beta) \subset \Delta^{\mathrm{fdr}}(\alpha,\beta) \subset \Delta^{\mathrm{gfr}}_{1,1}(K\alpha, K\beta).$$

此外，当 $m_1 = m_2 = 1$ 时，(35) 中的交集规则 $\delta_I(a,b)$ 与 (31) 中的 Leap规则 $\delta_L(a,b)$ 相同，并且 (36) 中的阈值 $(a_{\alpha,\beta}, b_{\alpha,\beta})$ 与 (32) 中的 $(a_{\alpha,\beta}, b_{\alpha,\beta})$ 相同。

接下来，我们考虑可实现的最小期望样本量的渐近近似， $T_A^{\min}\left(\Delta^{\mathrm{fdr}}(\alpha, \beta)\right)$，对于每个 $A \subset [K]$，当 α， $\beta \to 0$ 以使得 (33) 成立。具体而言，Corollary 4.1表明，在 (33) 条件下，对于每个 $A \subset [K]$, 一阶性能可表示为：

$$T_A^{\min}\left(\Delta^{\mathrm{fdr}}(\alpha,\beta)\right) = \frac{|\log\alpha|}{\mathrm{KL}^{\mathcal{W}}_{A,A}}(1 + o(1)),$$

其中，KL散度$^{\mathcal{W}}_{A,A}$ 在 (18) 中定义，其中 $\mathcal{W}$ 在 (34) 中给出，且 $m_1 = m_2 = 1$，即：

$$\mathrm{KL}^{\mathcal{W}}_{A,A} = \min\{\mathrm{KL}(f_A|f_C) : C \subset [K],\ |A \setminus C| \geq 1 \text{ or } |C \setminus A| \geq 1\}.$$

本文首先利用备注 12中描述的关系，然后应用定理6，并取 $m_1 = m_2 = 1$。注意，当 $m_1 = m_2 = 1$ 时，每个 $A \subset [K]$在$\mathcal{W}$ （定义见（34））下均存在唯一最有利子集。

定理 8. 假设条件 *1* 成立，且 α、$\beta \in (0,1)$ 满足条件 (33)。令 $\mathcal{W}$ 由 (34)定义，并令 $A \subset [K]$ 为任意信号子集。

(a) 考虑非对称情形 $r^{\mathcal{W}}_A = 1$。假设 (22) 对 $q = 3$成立。则当 $\alpha \to 0$时，

$$T_A^{\min}\left(\Delta^{\mathrm{fdr}}(\alpha,\beta)\right) = \frac{|\log(\alpha)|}{\mathrm{KL}^{\mathcal{W}}_{A,A}} + O(1).$$

(b) 考虑对称情形 $r_A^{\mathcal{W}} \geq 2$。令 $\epsilon \in (0, 1/2)$，并假设 (22) 对某个 q成立，且其满足 (23)。则当 $\alpha \to 0$时，

$$T_A^{\min}\left(\Delta^{\mathrm{fdr}}(\alpha,\beta)\right) = \frac{|\log(\alpha)|}{\mathrm{KL}_{A,A}^{\mathcal{W}}} + \frac{h_{A,A}^{\mathcal{W}}\sqrt{|\log(\alpha)|}}{(\mathrm{KL}_{A,A}^{\mathcal{W}})^{3/2}} + O\left((\log(\alpha))^{1/4+\epsilon/2}\right).$$

证明。 证明如下，承接前面的讨论。 □

备注 13。 我们强调，定理8适用于每个信号子集 $A \subset [K]$。然而， α 和 β需要以相同的速率趋于零。

4.4. 最优性讨论

It is instructive t为了比较我们的二阶 d-ord结果 lts with classical optimalit结果。lts in related setti

序列。在单数据流情形（$K = 1$）下，Wald 的序贯概率比检验（SPRT）[35],，该检验与（35）中的交集规则相一致，在类别 2 中对于 $m_1 = m_2 = 1$（即 $\Delta_{1\,1}^{\mathrm{gfr}}$，（$\alpha$，$\beta$））是恰好最优的。当选择阈值（$a$，$b$）使第一类和第二类错误概率等于（$\alpha$，$\beta$）时，在所有满足相同错误约束的序贯检验中，SPRT 同时最小化 $\mathbb{E}_0[T]$ 和 $\mathbb{E}_1[T]$；例如参见 [34,36]。因此，在这一单数据流情形下，最小可达到期望样本量恰好由 SPRT 实现。

接下来，我们回顾经典的多假设检验问题 [11,12,24],，该问题与本文所研究的多流情形密切相关。在这一表述中，观测到独立同分布数据 $\{X_t\}_{t\geq 1}$，其共同分布属于有限个相互竞争的 μ-密度 $\{f_i : i \in [M]\}$ 所构成的集合。一个序贯程序由停止时间 T 和终止决策 $D \in [M]$组成。误差控制通常通过错误概率族 $\{\mathbb{P}_i(D = j) : i \neq j \in [M]\}$ 或以下形式的加权风险 $\{\sum_{j\in[M]} \omega_{i,\,j}\mathbb{P}_i(D = j) : i \in [M]\}$ 来表示，其中 $\omega_{i,\,j} \geq 0$ 是固定权重， $\mathbb{P}_i$ 表示假设 i 下的分布。

与单一假设情形不同，一旦 $M \geq 3$，通常就无法实现精确最优性。Lorden [24] 提出了矩阵序贯概率比检验，这是 Wald 序贯概率比检验的多维扩展，并证明了它在每个 $\mathbb{P}_i, i \in [M]$ 下具有三阶 渐近最优性。更准确地说，在适当选择阈值后，矩阵序贯概率比检验的期望样本量与满足错误概率不大于矩阵序贯概率比检验的程序中的最小可达到期望样本量相差 $o(1)$，无论是按完整集合 $\{\mathbb{P}_i(D = j)\}$ 还是按加权风险衡量。

需要指出的是，这一最优性结果仅适用于满足以下条件的受限程序比较类： $\mathbb{P}_i(D = j) \to 0$ 对所有 $i \neq j$ 均成立，且误差控制日趋严格。本文考察的若干多重检验类别并不具备这一性质，包括当 1 类满足 $m_0 > 1$时，2 类满足 $m_1 + m_2 > 2$时，以及3类。前几小节研究的程序虽然计算效率高且具有二阶最优性，但预计无法在各自类别中达到三阶最优性。

5. 广义误分类率下的数值结果

在本节中，我们针对类 1， $\Delta_{m_0}^{\mathrm{gmr}}(\alpha)$给出一项数值研究，它在水平 $\alpha \in (0,1)$ 上控制广义误分类率，且不假设先验信息。目标是印证第 4.1 节中的结果。

具体而言，回顾（26）中的和交规则：若阈值选取为 b_α （见（27）），则 $\delta_S(b_\alpha) \in \Delta_{m_0}^{\mathrm{gmr}}(\alpha)$ 对每个 $\alpha \in (0,1)$成立，并且由定理 3可知，族 $\{\delta_S(b_\alpha)\}$ 在 $\{\Delta_{m_0}^{\mathrm{gmr}}(\alpha)\}$ 中当 $\alpha \to 0$ 时是二阶最优的。

如 [32], 所述，阈值 b_α 为和交规则提供了强有力的误差控制理论保证，但在实践中往往较为保守。在数值研究中，对于给定的 $\alpha \in (0,1)$，我们首先通过模拟确定一个阈值 b_α^*，使最大错误概率等于 α，即 $\max_{B\subset[K]} \mathbb{P}_B(|D_S(b_\alpha^*) \triangle B| \geq m_0) = \alpha$。随后，我们再次通过模拟计算空信号子集下的期望样本量（ESS），即 $\mathbb{E}_\emptyset[T_S(b_\alpha^*)]$。注意， b_α^* 是一个非保守临界值。为估计错误概率，我们采用重要性抽样；参见 [30,33]。

回顾（$\{f_0^k, f_1^k\colon k \in [K]\}$ 2）中的密度函数，以及（$\{I_0^k, I_1^k\colon k \in [K]\}$ 15）中的信息数。我们考虑正态均值检验问题，其中对于第 k个数据流，

$$H_0^k:\ f_0^k = \mathcal{N}(0, \sigma_k^2) \qquad \text{versus} \qquad H_1^k:\ f_1^k = \mathcal{N}(\mu_k, \sigma_k^2).$$

对于每个 $k \in [K]$, $I_1^k = I_0^k = \mu_k^2/(2\sigma_k^2)$。

我们考虑式（28）中定义的损失函数 W 。根据引理 1，每个 $A \subset [K]$ 相对于 $\mathcal{W}$，都存在唯一最有利子集，其中 $D_A^{\mathcal{W}} = A$ 且 $\mathrm{KL}_{A,*}^{\mathcal{W}} = \mathrm{KL}_{A,A}^{\mathcal{W}}$。此外，回顾式（$C_A^{\mathcal{W}}$和 $r_A^{\mathcal{W}} = |\mathcal{C}_A^{\mathcal{W}}|$）中的定义。这里我们考虑 19）所定义的对称情形，其中 $r_{\mathcal{W}_A} \geq 2$；附录 F则介绍 $r_A^{\mathcal{W}} = 1$ 所对应的非对称情形。

具体而言，我们考虑以下设定

$$\mu_1 = \cdots = \mu_K = \mu, \quad \text{and} \quad \sigma_1^2 = \cdots = \sigma_K^2 = \sigma^2,$$

其中 $\mu \in \mathbb{R}$ 且 $\sigma^2 > 0$。根据定义（18）和（19），

$$\mathrm{KL}_{\emptyset,*} = m_0 \frac{\mu^2}{2\sigma^2}, \quad \left\{C_{\emptyset,j}^{\mathcal{W}}:\ 1 \leq j \leq r_\emptyset^{\mathcal{W}}\right\} = \{B \subset [K]:\ |B| = m_0\}, \quad \text{and} \quad r_\emptyset^{\mathcal{W}} = \binom{K}{m_0}.$$

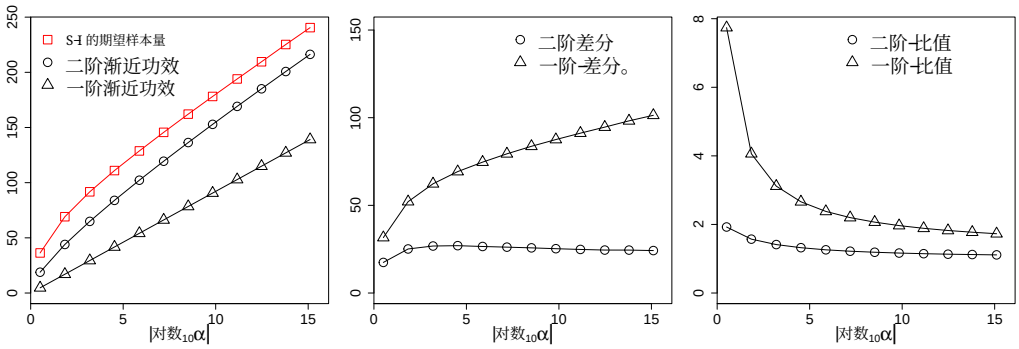
回顾式（ $\mathbf{\Sigma}_\emptyset^{\mathcal{W}}$中 20）给出的定义。对于 $1 \leq i,\ j \leq r_\emptyset^{\mathcal{W}}$，经初等计算可得，

$$(\mathbf{\Sigma}_\emptyset^{\mathcal{W}})_{ij} = \mathrm{Cov}_\emptyset\left(\sum_{k \in C_{\emptyset,i}^{\mathcal{W}}} \log\frac{f_1^k(\mathbf{X}_1^k)}{f_0^k(\mathbf{X}_1^k)}, \sum_{k \in C_{\emptyset,j}^{\mathcal{W}}} \log\frac{f_1^k(\mathbf{X}_1^k)}{f_0^k(\mathbf{X}_1^k)},\right) = \left|C_{\emptyset,i}^{\mathcal{W}} \cap C_{\emptyset,j}^{\mathcal{W}}\right| \frac{\mu^2}{\sigma^2},$$

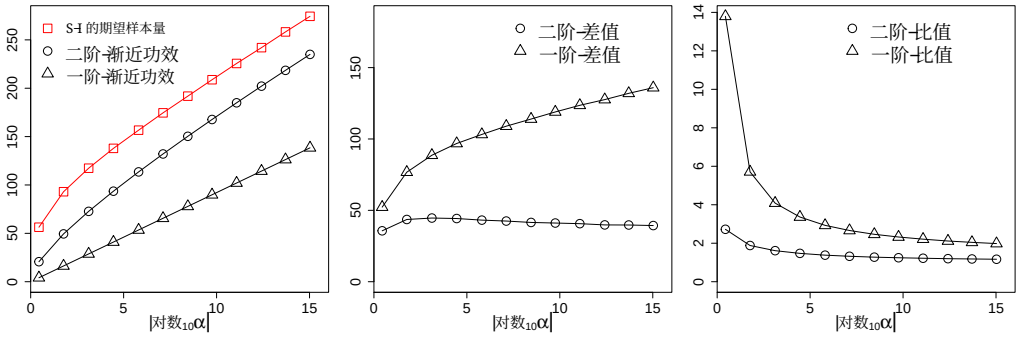
其中，$\mathrm{Cov}_\emptyset$ 表示相对于概率 $\mathbb{P}_\emptyset$ 的协方差。因此，式（$h_\emptyset^{\mathcal{W}}$21）中定义的 是分布为 $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_\emptyset^{\mathcal{W}})$ 的高斯随机向量的期望最大值，可通过蒙特卡罗模拟计算。由于它通过 m_0 依赖于 $\mathcal{W}$，我们将其记为 $h^{(m_0)}$, 而非 $\emptyset$。根据式（30）和定理 4（b），1 类可实现的最小 ESS 的一阶和二阶渐近近似分别为1 ：

$$\mathrm{FO}_\alpha := \frac{2\sigma^2}{m_0\mu^2}|\log\alpha|, \quad \text{and} \quad \mathrm{SO}_\alpha := \frac{2\sigma^2}{m_0\mu^2}|\log\alpha| + \left(\frac{2\sigma^2}{m_0\mu^2}\right)^{3/2} h_\emptyset^{(m_0)}\sqrt{|\log\alpha|}.$$

在图 1 和图 2 中，我们考虑三种情形：$(K, m_0) \in \{(20,1), (20,2), (50,2)\}$。每种情形均展示三幅图。最左侧的图显示期望样本量 ESS $\mathbb{E}_\emptyset[T_S(b_\alpha^*)]$（方形标记），以及一阶近似 FO_α （三角形标记）和二阶近似 SO_α（圆形标记），它们均为 $|\log_{10}(\alpha)|$ 的函数，其中 $\alpha \in$ 在 (0, 1) 内变化。中间的图显示差值



(a) 对称情形： $K = 20,\ \ m_0 = 2$



(b) 对称情形： $K = 50, m_0 = 2$

图2：所有面板的x轴均为 $|\log_{10}\alpha|$。最左侧的图中，“S-I”表示和交规则，“AP”表示可实现的最小期望样本量的渐近近似。中间的图中，“Diff.”表示S-I规则的期望样本量与两种近似之间的差值。最右侧的图中，“Ratio”表示S-I规则的期望样本量与两种近似各自的比值。

$\mathbb{E}_\emptyset[T_S(b^*_\alpha)] - \mathrm{FO}\alpha$ 与 $\mathbb{E}_\emptyset[T_S(b^*_\alpha)] - \mathrm{SO}\alpha$，它们是 $|\log_{10}(\alpha)|$的函数。最右侧的图显示比值 $\mathbb{E}_\emptyset[T_S(b^*_\alpha)]/\mathrm{FO}\alpha$ 和 $\mathbb{E}_\emptyset[T_S(b^*_\alpha)]/\mathrm{SO}\alpha$，它们是 $|\log_{10}(\alpha)|$的函数。

如定理 3 所示，采用非保守阈值 b^*_α的和交规则是二阶最优的。因此，正如定理 4(b) 所预测且图中所示，二阶近似 $\mathrm{SO}\alpha$ 比一阶近似 $\mathrm{FO}\alpha$准确得多。具体而言，差值 $\mathbb{E}_\emptyset[T_S(b^*_\alpha)] - \mathrm{FO}\alpha$ 随 $\alpha \to 0$增大而发散。此外，尽管定理 3 和定理 4(b)仅能推出，对于任意固定的 $\epsilon \in (0, 1/4)$，$\mathbb{E}_\emptyset[T_S(b^*_\alpha)] - \mathrm{SO}_\alpha = O(|\log(\alpha)|^{1/4+\epsilon})$，但中间的面板表明该差值仍然有界，即为 $O(1)$。若能严谨地证明这一更精细的结论，将会很有意义。此外，在我们的分析中，K 是固定的。从图 2 可以看出，对于较小的 K，二阶近似似乎更加准确。K 的影响有待今后系统研究。

6. 结论与未来工作

我们为独立数据流下同时停止的序贯多重检验建立了统一的二阶渐近理论。通过采用适当的

代价参数和损失函数所对应的贝叶斯表述，我们给出了充分条件，在这些条件下，二阶贝叶斯最优性意味着二阶频率学派最优性。在这些条件下，一些此前已知在各种误差度量和信息结构下达到一阶最优的程序，实际上是二阶最优的：对于每一种信号配置，当误差容限趋于零时，它们相对于可达到的最小期望样本量的超额期望样本量始终保持有界。

我们还得到了最小可达到期望样本量的二阶渐近展开。在满足适当矩条件，且由损失函数诱导的最有利备择满足唯一性条件时，我们确定了主导对数项的二阶修正项。其形式取决于相应的越界问题是非对称还是对称的。这些结果拓展了经典的一阶理论。

仍有若干方向有待探索。首先，对于本文考察的所有类别（第 2 类除外），每个信号子集 $A \in \mathcal{A}$ 在适当选择损失函数 $\mathcal{W}$后，都有唯一的最有利子集。对于第 2 类，我们给出了一个反例，表明这种唯一性可能不成立，并给出了唯一性成立的充分条件。第2 类的最小可达到期望样本量的完整二阶刻画仍有待研究。

其次，在对称情形下，期望样本量展开中的余项阶为 $O((\text{对数}(c_{\theta}^{-1}))^{\epsilon})$，其中任意 $\epsilon \in (1/4,1/2)$均成立。数值证据表明，近似误差实际上可能有界，这意味着更精细的分析或许能改进余项界。第三，对于包括第2 类和第3类在内的若干类别，一阶渐近最优性并不要求误差容限（例如 α 和 β）以相同速率趋于零（见条件（33））。将二阶理论推广到此类更一般的情形值得进一步研究。最后，将理论推广到复合假设以及涉及相关数据流的设定（如 [6]）也是一个自然的研究方向。

参考文献

[1] BARTROFF, J.（2018）。控制序贯数据广义错误率的多重假设检验。《Statistica Sinica》363–398。[2] BARTROFF, J. 和 LAI, T. L.（2010）。多重假设的多阶段检验。《Communications in Statistics—Theory and Methods》39 1597–1607。[3] BARTROFF, J. 和 SONG, J.（2014）。控制 I 类和 II 类族错误率的多重假设序贯检验。《Journal of Statistical Planning and Inference》153 100–114。[4] BARTROFF, J. 和 SONG, J.（2020）。控制错误发现率和未发现率的多重假设序贯检验。《Sequential Analysis》39 65–91。[5] BENJAMINI, Y. 和 HOCHBERG, Y.（1995）。控制错误发现率：一种实用且强大的多重检验方法。《Journal of the Royal Statistical Society: Series B (Methodological)》57 289–300。[6] CHAUDHURI, A. 和 FELLOURIS, G.（2024）。相关数据流的联合序贯检测与隔离。《The Annals of Statistics》52 1899–1926。[7] DE, S. K. 和 BARON, M.（2012a）。用于多重假设检验且对家族错误率 I 和 II 进行强控制的序贯Bonferroni方法。《Sequential Analysis》31 238–262。[8] DE, S. K. 和 BARON, M.（2012b）。序贯试验中检验多重假设的逐步向上和逐步向下方法。《Journal of Statistical Planning and Inference》142 2059–2070。[9] DE, S. K. 和 BARON, M.（2015）。控制广义族错误率的序贯检验。《Statistical Methodology》23 88–102。[10] DE LA PENA, V. 和 GINÉ, E.（2012）。去耦合：从相关性到独立性。施普林格科学与商业媒体。[11] DRAGALIN, V. P.、TARTAKOVSKY, A. G. 和 VEERAVALLI, V. V.（2000）。多假设序贯概率比检验。II. 期望样本量的精确渐近展开。《IEEE Transactions on Information Theory》46 1366–1383。[12] DRAGLIA, V.、TARTAKOVSKY, A. G. 和 VEERAVALLI, V. V.（1999）。多假设序贯概率比检验。I. 渐近最优性。《IEEE Transactions on Information Theory》45 2448–2461。

[13] DURRETT, R.（2019）。概率：理论与例子 49。剑桥大学出版社。[14] GUT, A.（2009）。停止随机游走。施普林格。[15] HE, X. 和 BARTROFF, J.（2021）。具有或不具有信号数量先验信息的渐近最优序贯错误发现率和假发现率控制。《Journal of Statistical Planning and Inference》210 87 99–。[16] HOLM, S.（1979）。一种简单的序贯拒绝式多重检验程序。《斯堪的纳维亚统计学杂志》65–70。[17] HOMMEL, G.（1988）。基于修正的 Bonferroni 检验的分阶段拒绝式多重检验程序。《生物计量学》383–386。[18] JANSON, S.（1986）。具有正漂移的随机游走的首次通过时间和最后离开时间、最小值及相关量的矩。《应用概率进展》18 865–879。[19] JAVANMARD, A. 和 MONTANARI, A.（2018）。控制错误发现率和错误发现超限率的在线规则。《The Annals of Statistics》46 526–554。[20] KIEFER, J. 和 SACKS, J.（1963）。渐近最优的序贯推断与设计。《数学统计年刊》705–750。[21] LEHMANN, E. L. 和 ROMANO, J. P.（2005）。族错误率的推广。《Annals of Statistics》33 1138–1154。[22] LI, Y.、NITINAWARAT, S. 和 VEERAVALLI, V. V.（2014）。普适异常值假设检验。《IEEE Transactions on Information Theory》60 4066–4082。[23] LORDEN, G.（1967）。渐近贝叶斯序贯检验的积分风险。《数学统计年刊》38 1399–1422。[24] LORDEN, G.（1977）。针对有限个参数值的近似最优序贯检验。《The Annals of Statistics》5 1–21。[25] MALLOY, M. L. 和 NOWAK, R. D.（2014）。稀疏恢复的序贯检验。《IEEE Transactions on Information Theory》60 7862–7873。[26] MEI, Y.（2008）。传感器网络中分布式序贯假设检验的渐近最优性理论。《IEEE Transactions on Information Theory》54 2072–2089。[27] NAGAI, K. 和 ZHANG, C.-H.（2006）。具有中间阶扰动的随机游走的非线性更新定理。《讲义与专著丛书》164–175。[28] RAI ˇC, M.（2018）。关于 Lipschitz 和光滑检验函数的多元中心极限定理。arXiv 预印本 arXiv:1812.08268。[29] SCHWARZ, G.（1962）。贝叶斯序贯检验区域的渐近形状。《数学统计年刊》224–236。[30] SONG, Y. 和 FELLOURIS, G.（2016）。序贯多重检验中误分类概率的对数效率模拟。载于《2016 年冬季仿真会议（WSC）》314–325。IEEE。[31] SONG, Y. 和 FELLOURIS, G.（2017）。具有信号数量先验信息的渐近最优序贯多重检验程序。《电子统计学杂志》11 338–363。[32] SONG, Y. 和 FELLOURIS, G.（2019）。具有广义错误控制的序贯多重检验。《The Annals of Statistics》47 1776–1803。[33] SONG, Y. 和 FELLOURIS, G.（2025）。组合数量级的稀有区域上的错误退出概率的高效重要性抽样。arXiv 预印本 arXiv:2509.14596。[34] TARTAKOVSKY, A.、NIKIFOROV, I. 和 BASSEVILLE, M.（2014）。序贯分析：假设检验与变点检测。CRC 出版社。[35] WALD, A.（1945 16 117 186）。统计假设的序贯检验。《数学统计年刊》–。[36] WALD, A. 和 WOLFOWITZ, J.（1948）。序贯概率比检验的最优性。《数学统计年刊》326–339。[37] XING, Y. 和 FELLOURIS, G.（2023）。采用多阶段检验且不受稀疏性约束的信号恢复。《IEEE Transactions on Information Theory》69 7220–7245。[38] XING, Y. 和 FELLOURIS, G.（2025）。具有异步决策的渐近最优序贯多重检验。《伯努利》31 271–294。[39] ZHANG, C.-H.（1988）。非线性更新理论。《概率年刊》793–824。

附录 A：定理1的证明

回想一下， π_0 是 A 上的均匀分布，而 $\mathrm{ERR}^{\mathrm{int}}(\cdot;\cdot)$ 表示式（7）中的积分误差损失。回想一下式（11）中 $\delta_{\mathrm{Ld}}(c,\mathrm{W}) = (T_{\mathrm{Ld}}(c,\mathrm{W}), D_{\mathrm{Ld}}(c,\mathrm{W}))$、式（8）中 $r^{\mathrm{int}}(\delta; c, \mathrm{W})$ 以及式（9）中$r^{\mathrm{int}}_{\min}(c,\mathrm{W})$的定义。

定理 1的证明。固定 $\boldsymbol{\theta} \in (0,1)^r$。令 $\delta(\boldsymbol{\theta}) = (T(\boldsymbol{\theta}), D(\boldsymbol{\theta})) \in \Delta(\boldsymbol{\theta})$为 $\Delta(\boldsymbol{\theta})$中的任意程序。

根据下文引理 3 的第（b）部分，存在一个不依赖于 $M > 0$ 的常数 $\boldsymbol{\theta}$，使得对于任意信号子集 $A \in \mathcal{A}$,

$$\mathbb{E}_A[T_0(\boldsymbol{\theta})] - \mathbb{E}_A[T(\boldsymbol{\theta})] \le \left(\frac{\mathrm{ERR}^{\mathrm{int}}(\delta_0(\boldsymbol{\theta}); \mathcal{W}) + \mathrm{ERR}^{\mathrm{int}}(\delta(\boldsymbol{\theta}); \mathcal{W})}{c_{\boldsymbol{\theta}}} + M \right) |\mathcal{A}|. \tag{37}$$

由于 $\delta_0(\boldsymbol{\theta})$, $\delta(\boldsymbol{\theta}) \in \Delta(\boldsymbol{\theta})$，根据条件（14），可得

$$\mathbb{E}_A[T_0(\boldsymbol{\theta})] - \mathbb{E}_A[T(\boldsymbol{\theta})] \le (2L + M)|\mathcal{A}|.$$

由于 $\delta(\boldsymbol{\theta})$ 是 $\Delta(\boldsymbol{\theta})$ 中的任意程序，因此

$$\mathbb{E}_A[T_0(\boldsymbol{\theta})] - T_A^{\min}(\Delta(\boldsymbol{\theta})) \le (2L + M)|\mathcal{A}|.$$

证明完毕。 □

下面的引理是非渐近的。

引理 3. 假设假设条件 *1* 成立。设 $c > 0$ 为代价，且 $\mathcal{W}(\cdot\,|\,\cdot) : \mathcal{A} \times \mathcal{A} \mapsto [0, \infty)$ 为满足条件 (12)的损失函数。若程序 $\delta_0 = (T_0, D_0) \in \Delta$ 满足

$$T_0 \le T_{\mathrm{Ld}}(c, \mathcal{W}) \textit{ (i.e. } \delta_0 \textit{ stops earlier than } \delta_{\mathrm{Ld}}),$$

则存在一个不依赖于 c 的常数 $M > 0$，使得

(a) 我们有

$$r^{\mathrm{int}}(\delta_0; c, \mathcal{W}) - r^{\mathrm{int}}_{\min}(c, \mathcal{W}) \le \mathrm{ERR}^{\mathrm{int}}(\delta_0; \mathcal{W}) + Mc;$$

(b) 对于任意序贯程序 $\delta = (T, D) \in \Delta$，我们有

$$\mathbb{E}_A(T_0 - T) \le \left(\frac{\mathrm{ERR}^{\mathrm{int}}(\delta_0; \mathcal{W}) + \mathrm{ERR}^{\mathrm{int}}(\delta; \mathcal{W})}{c} + M \right) |\mathcal{A}|, \quad \textit{for any } A \in \mathcal{A}.$$

证明。 首先，我们证明第（a）部分。回忆 ESS 的定义$^{\mathrm{int}}$，见（6）以及 ERR 的定义$^{\mathrm{int}}$，见（7）。由于 $T_0 \le T_{\mathrm{Ld}}(c, \mathcal{W})$，我们有

$$\mathrm{ESS}^{\mathrm{int}}(\delta_0; c) \le \mathrm{ESS}^{\mathrm{int}}(\delta_{\mathrm{Ld}}(c, \mathcal{W}); c).$$

由于 W 非负，我们有

$$\begin{aligned} &r^{\mathrm{int}}(\delta_0; c, \mathcal{W}) - r^{\mathrm{int}}(\delta_{\mathrm{Ld}}(c, \mathcal{W}); c, \mathcal{W}) \\ &= \mathrm{ESS}^{\mathrm{int}}(\delta_0; c) - \mathrm{ESS}^{\mathrm{int}}(\delta_{\mathrm{Ld}}; c) + \mathrm{ERR}^{\mathrm{int}}(\delta_0; \mathcal{W}) - \mathrm{ERR}^{\mathrm{int}}(\delta_{\mathrm{Ld}}(c, \mathcal{W}); \mathcal{W}) \le \mathrm{ERR}^{\mathrm{int}}(\delta_0; \mathcal{W}). \end{aligned}$$

根据 [23], , 结合假设1 和条件（12）， δ_{Ld} 是二阶贝叶斯最优的，也就是说，存在一个常数 $M>0$ 不依赖于 c，使得

$$r^{\mathrm{int}}(\delta_{\mathrm{Ld}}(c,\mathcal{W});c,\mathcal{W})-r^{\mathrm{int}}_{\mathrm{min}}(c,\mathcal{W})\le Mc.$$

将上述两个不等式相加，得到

$$r^{\mathrm{int}}(\delta_0;c,\mathcal{W})-r^{\mathrm{int}}_{\mathrm{min}}(c,\mathcal{W})\le \mathrm{ERR}^{\mathrm{int}}(\delta_0;\mathcal{W})+Mc,$$

这就完成了第（a）部分的证明。

接下来，我们证明第（b）部分。任取程序 $\delta=(T,\ D)\in\Delta$。我们定义一个辅助程序 $\tilde{\delta}=(\tilde{T},\ \tilde{D})$，如下所示

$$\tilde{T}=\min\{T,T_0\},\quad \tilde{D}=\begin{cases}D & \text{if } T\le T_0\\ D_0 & \text{otherwise}\end{cases}.$$

显然， $\tilde{D}$ 是$\mathcal{F}_{\tilde{T}}$ 可测的。因此， $\tilde{\delta}$ 是一个有效的序贯程序，即， $\tilde{\delta}\in\Delta$。根据（7）中的定义，

$$\begin{aligned}\mathrm{ERR}^{\mathrm{int}}(\tilde{\delta};\mathcal{W})&=\sum_{A\in\mathcal{A}}\pi_0(A)\mathbb{E}_A[\mathcal{W}(D_0\mid A)\mathbb{1}\{T_0<T\}+\mathcal{W}(D\mid A)\mathbb{1}\{T\le T_0\}]\\&\le \mathrm{ERR}^{\mathrm{int}}(\delta_0;\mathcal{W})+\mathrm{ERR}^{\mathrm{int}}(\delta;\mathcal{W}).\end{aligned}\tag{38}$$

此外，根据定义9，我们有 $r^{\mathrm{int}}_{\mathrm{min}}(c,\ \mathcal{W})\le r^{\mathrm{int}}(\tilde{\delta};\ c,\ \mathcal{W})$，由此可根据第（a）部分得到

$$r^{\mathrm{int}}(\tilde{\delta};c,\mathcal{W})\ge r^{\mathrm{int}}(\delta_0;c,\mathcal{W})-[\mathrm{ERR}^{\mathrm{int}}(\delta_0;\mathcal{W})+Mc].\tag{39}$$

因此，由（38）和（39），我们有

$$\begin{aligned}\mathrm{ESS}^{\mathrm{int}}(\delta_0;c)-\mathrm{ESS}^{\mathrm{int}}(\tilde{\delta};c)&=[r^{\mathrm{int}}(\delta_0;c,\mathcal{W})-\mathrm{ERR}^{\mathrm{int}}(\delta_0;\mathcal{W})]-[r^{\mathrm{int}}(\tilde{\delta};c,\mathcal{W})-\mathrm{ERR}^{\mathrm{int}}(\tilde{\delta};\mathcal{W})]\\&\le \mathrm{ERR}^{\mathrm{int}}(\delta;\mathcal{W})+\mathrm{ERR}^{\mathrm{int}}(\delta_0;\mathcal{W})+Mc,\end{aligned}$$

根据定义（6），这意味着

$$\frac{c}{|\mathcal{A}|}\sum_{A\in\mathcal{A}}\mathbb{E}_A(T_0-\tilde{T})\le \mathrm{ERR}^{\mathrm{int}}(\delta;\mathcal{W})+\mathrm{ERR}^{\mathrm{int}}(\delta_0;\mathcal{W})+Mc.$$

~ ~ ~根据 T 的定义，我们有 $T\le T$和 $T\le T_0$。因此，对于任意 $A\in\mathrm{A}$，我们有

$$\frac{c}{|\mathcal{A}|}\mathbb{E}_A(T_0-T)\le\frac{c}{|\mathcal{A}|}\mathbb{E}_A(T_0-\tilde{T})\le \mathrm{ERR}^{\mathrm{int}}(\delta;\mathcal{W})+\mathrm{ERR}^{\mathrm{int}}(\delta_0;\mathcal{W})+Mc,$$

由此完成了第（b）部分的证明。 □

附录 B：定理 2 的证明

定理2的证明。回顾式（ $\delta_{\mathrm{Ld}}(c_{\theta},\mathrm{W})$ （11）中的程序。**步骤 1：我们证明** $T_A^{\mathbf{min}}(\Delta(\boldsymbol{\theta}))=\mathbb{E}_A[T_{\mathrm{Ld}}(c_{\boldsymbol{\theta}},\ \mathcal{W})]+O(1)$ **当** $\boldsymbol{\theta}\to\mathbf{0}$**时**。

我们通过修改 $\delta_{\mathrm{Ld}}(c_{\boldsymbol{\theta}})$ 和 $\delta_0(\boldsymbol{\theta})$，构造一个程序，记为 $\delta_1(\boldsymbol{\theta}) = (T_1(\boldsymbol{\theta}), D_1(\boldsymbol{\theta}))$:

$$T_1(\boldsymbol{\theta}) := T_{\mathrm{Ld}}(c_{\boldsymbol{\theta}}, \mathcal{W}), \quad \text{and} \quad D_1(\boldsymbol{\theta}) := D_0(\boldsymbol{\theta}).$$

根据定义以及（13），

$$D_1(\boldsymbol{\theta}) = D_0(\boldsymbol{\theta}) \in \mathcal{F}_{T_0(\boldsymbol{\theta})} \subset \mathcal{F}_{T_{\mathrm{Ld}}(c_{\boldsymbol{\theta}}, \mathcal{W})} = \mathcal{F}_{T_1(\boldsymbol{\theta})}.$$

因此，$\delta_1(\boldsymbol{\theta})$ 是一个有效的序贯程序，也就是说，对于每个 $\boldsymbol{\theta} \in (0,1)^r$，$\delta_1(\boldsymbol{\theta}) \in \Delta$。

由于 $\delta_0(\boldsymbol{\theta}) \in \Delta(\boldsymbol{\theta})$ 和 $\delta_1(\boldsymbol{\theta})$ 采用相同的决策规则，根据假设 2，对于每个 $\boldsymbol{\theta} \in (0,1)^r$，有 $\delta_1(\boldsymbol{\theta}) \in \Delta(\boldsymbol{\theta})$。

注意，我们假设（12）和（14）对 $\mathcal{W}$ 和 $\Delta(\boldsymbol{\theta})$成立。此外，$T_1(\boldsymbol{\theta}) = T_{\mathrm{Ld}}(c_{\boldsymbol{\theta}}, \mathcal{W})$ 根据定义满足条件（13）。因此，根据定理 1，我们有 $\{\delta_1(\boldsymbol{\theta})\}$ 在 $\{\Delta(\boldsymbol{\theta})\}$中是二阶最优的，这尤其意味着当 $\boldsymbol{\theta} \to \mathbf{0}$,

$$\mathbb{E}_A[T_{\mathrm{Ld}}(c_{\boldsymbol{\theta}}, \mathcal{W})] - T_A^{\min}(\Delta(\boldsymbol{\theta})) = \mathbb{E}_A[T_1(\boldsymbol{\theta})] - T_A^{\min}(\Delta(\boldsymbol{\theta})) = O(1).$$

步骤 2：我们建立 $\mathbb{E}_A[T_{\mathrm{Ld}}(c_{\boldsymbol{\theta}}, \mathcal{W})]$ 的二阶近似。

De记 $m_{\mathcal{W}}$ 为矩阵 W (·|·) 的最小非零元素，记 $M_{\mathcal{W}}$ 为最大元素，即 ,

$$m_{\mathcal{W}} := \min\{\mathrm{W}(D|A) : A, D \in \mathrm{A} \text{且} \mathrm{W}(D|A) > 0\}, \qquad M_{\mathcal{W}} := \max\{\mathrm{W}(D|A) : A, D \in \mathrm{A}\}. \tag{40}$$

由于A是有限的，我们有 $0 < m_{\mathcal{W}} \le M_{\mathcal{W}} < \infty$。此外，对每个 $c > 0$，定义

$$\begin{aligned} \tilde{T}(c, \mathcal{W}) &:= \min_{B, D \in \mathcal{A}} \tilde{T}_{B,D}(c, \mathcal{W}), \quad \text{where} \\ \tilde{T}_{B,D}(c, \mathcal{W}) &:= \min\left\{ t \ge 1 : f_{G,[t]} < c f_{B,[t]}, \text{ for each } G \in \mathcal{H}_D^{\mathcal{W}} \right\}. \end{aligned} \tag{41}$$

也就是说，$\tilde{T}(c, \mathcal{W})$停止于首次时间 t，此时存在 B_0，$D_0 \in \mathcal{A}$使得$f_{B_0,\,[t]} / f_{G,\,[t]} > c^{-1}$对所有 $G \in \mathcal{H}_{D_0}^{\mathcal{W}}$，其中回顾$\mathcal{H}_{\mathcal{W}}^{D_0}$定义于（17）。

由引理4，可得

$$\tilde{T}\left(c_{\boldsymbol{\theta}} \frac{|\mathcal{A}|}{m_{\mathcal{W}}}, \mathcal{W}\right) \le T_{\mathrm{Ld}}(c_{\boldsymbol{\theta}}, \mathcal{W}) \le \tilde{T}\left(c_{\boldsymbol{\theta}} \frac{1}{|\mathcal{A}| M_{\mathcal{W}}}, \mathcal{W}\right).$$

此外，引理5提供了关于 $\mathbb{E}_A[\tilde{T}(c, \mathcal{W})]$ 在 $c \to \infty$下的渐近近似，借助非线性更新理论。由于 $|\mathcal{A}|$，$m_{\mathcal{W}}$ 与 $M_{\mathcal{W}}$ 均为常数，从而证明完成。 □

对于代价 $c > 0$和损失函数 $\mathcal{W} : \mathcal{A} \times \mathcal{A} \to [0, \infty)$，停止时间 $\tilde{T}(c, \mathcal{W})$定义于 (41) 中，而 $m_{\mathcal{W}}$, $M_{\mathcal{W}}$ 定义于 (40) 中。

引理4。 设 $\mathcal{W} : \mathcal{A} \times \mathcal{A} \to [0, \infty)$ 是一个损失函数，且 $c > 0$。则

$$T_1 := \tilde{T}\left(\frac{c|\mathcal{A}|}{m_{\mathcal{W}}}, \mathcal{W}\right) \le T_{\mathrm{Ld}}(c, \mathcal{W}) \le \tilde{T}\left(\frac{c}{|\mathcal{A}| M_{\mathcal{W}}}, \mathcal{W}\right) := T_2.$$

证明。 我们先证明第二个不等式。在时间 T_2，根据（41）中的定义，存在B_0，$D_0 \in \mathcal{A}$ 使得

$$\frac{c}{|\mathcal{A}| M_{\mathcal{W}}} \ge \frac{\max_{G \in \mathcal{H}_{D_0}^{\mathcal{W}}} f_{G,[T_2]}}{f_{B_0,[T_2]}},$$

其中 $D_{D_0}^{\mathcal{W}}$ 如（17）中所定义。另一方面，根据 $M_{\mathcal{W}}$的定义，

$$\max_{G\in\mathcal{H}_{D_0}^{\mathcal{W}}} f_{G,[T_2]} \geq \frac{1}{|\mathcal{A}|}\sum_{G\in\mathcal{H}_{D_0}^{\mathcal{W}}} f_{G,[T_2]} \geq \frac{1}{|\mathcal{A}|M_{\mathcal{W}}}\sum_{A\in\mathcal{H}_{D_0}^{\mathcal{W}}} f_{G,[T_2]}W(D|G).$$

另一方面，我们有

$$f_{B_0,[T_2]} \leq \sum_{B\in\mathcal{A}} f_{B,[T_2]}.$$

然后，根据 π_t 在（10）中的定义，上述三个不等式意味着

$$c > \sum_{A\in\mathcal{H}_{D_0}^{\mathcal{W}}} \pi_{T_2}(G)W(D|G).$$

因此， $T_{\mathrm{Ld}}(c,\ \mathrm{W})$ 的停止准则在时间 T_2 得到满足，这意味着 $T_{\mathrm{Ld}}(c,\ \mathrm{W}) \leq T_2$。

接下来，我们关注第一个不等式。由于在本证明中 $\mathcal{W}$ 是固定的，我们将 T_{Ld}（c）记作 T_{Ld}（c, W）。在时间 T_{Ld}（c），根据（10）和（11）中的定义，存在 $D_0 \in \mathcal{A}$ 使得

$$\sum_{G\in\mathcal{A}} f_{G,[T_{\mathrm{Ld}}(c)]}\mathcal{W}(D_0 \mid G) < c\sum_{B\in\mathcal{A}} f_{B,[T_{\mathrm{Ld}}(c)]}.$$

令 $B_0 := \mathrm{argmax}_{B\in\mathcal{A}} f_{B,\ [T_{\mathrm{Ld}}\ (c)\ ]}$。那么，根据 $m_{\mathcal{W}}$ 的定义，可得

$$c|\mathcal{A}|f_{B_0,[T_{\mathrm{Ld}}(c)]} > \sum_{G\in\mathcal{H}_{D_0}^{\mathcal{W}}} f_{G,[T_{\mathrm{Ld}}(c)]}\mathcal{W}(D_0 \mid G) \geq m_{\mathcal{W}} \max_{G\in\mathcal{H}_{D_0}^{\mathcal{W}}} f_{G,[T_{\mathrm{Ld}}(c)]}.$$

因此， T_1 的停止准则在时间 $T_{\mathrm{Ld}}(c)$得到满足。从而我们有 $T_1 \leq T_{\mathrm{Ld}}(c,\mathrm{W})$。证明完毕。

□

接下来，我们给出对 $\mathbb{E}_A[\tilde{T}(c,\ \mathcal{W})]$的渐近近似，当 $c \to \infty$时，其中 A出现在定理2中。回顾$h_A^{\mathcal{W}}$在(21)中的定义以及停止时间 $\tilde{T}(c,\ \mathcal{W})$ 在(41)中的定义。定义

$$L_c := \log(c^{-1}), \quad \text{for } c \in (0,1). \tag{42}$$

引理 5. 假设假设*1* 成立。设 $A \in \mathcal{A}$，并假设 A 关于$\mathcal{W}$具有唯一最有利子集。

(a) 考虑情形 $r_A^{\mathcal{W}} = 1$。假设(22) 对 $q = 3$成立。则当 $c \to 0$

$$\mathbb{E}_A\left[\tilde{T}(c,\mathcal{W})\right] = \frac{L_c}{\mathrm{KL}_{A,*}^{\mathcal{W}}} + O(1).$$

(b) 考虑情形 $r_A^{\mathcal{W}} \geq 2$。对于 $\epsilon \in (0,1/2)$，假设(22) 对某个 q 成立且其满足(23)。则当 $c \to 0$

$$\mathbb{E}_A\left[\tilde{T}(c,\mathcal{W})\right] = \frac{L_c}{\mathrm{KL}_{A,*}^{\mathcal{W}}} + \frac{h_A^{\mathcal{W}}\sqrt{L_c}}{(\mathrm{KL}_{A,*}^{\mathcal{W}})^{3/2}} + O(L_c^{1/4+\epsilon/2}).$$

证明。回顾 $D_A^{\mathcal{W}}$，其定义见(17)。为简便起见，记

$$\tilde{T}_{A,*} := \tilde{T}_{A,D_A^{\mathcal{W}}} \quad \text{and} \quad \mathcal{H}_{A,*}^{\mathcal{W}} := \mathcal{H}_{D_A^{\mathcal{W}}}^{\mathcal{W}}.$$

由于 W 在整个证明过程中固定不变，我们在记号中省略其依赖关系，并将 $\tilde{T}_{A,D}(c)$ 记为 $\tilde{T}_{A,D}(c,\mathsf{W})$，将 $\tilde{T}_{A,*}(c)$ 记为 $\tilde{T}_{A,*}(c,\mathsf{W})$，并将 $\tilde{T}(c)$ 记为 $\tilde{T}(c,\mathsf{W})$

我们注意到以下分解

$$\mathbb{E}_A\left[\tilde{T}(c)\right] = \mathbb{E}_A\left[\tilde{T}_{A,*}(c)\right] - \mathbb{E}_A\left[(\tilde{T}_{A,*}(c) - \tilde{T}(c))\mathbb{1}\{\tilde{T}(c) < \tilde{T}_{A,*}(c)\}\right]. \tag{43}$$

根据定义（41），对每个 $c \in (0,1)$，$0 < \tilde{T}(c) \le \tilde{T}_{A,*}(c)$。由霍尔德不等式可得

$$\begin{aligned} 0 &\le \mathbb{E}_A\left[(\tilde{T}_{A,*}(c) - \tilde{T}(c))\mathbb{1}\{\tilde{T}(c) < \tilde{T}_{A,*}(c)\}\right] \\ &\le \mathbb{E}_A\left[\tilde{T}_{A,*}(c)\mathbb{1}\{\tilde{T}(c) < \tilde{T}_{A,*}(c)\}\right] \le \left(\mathbb{E}_A\left[(\tilde{T}_{A,*}(c))^3\right]\right)^{1/3}\left(\mathbb{P}_A\left(\tilde{T}(c) < \tilde{T}_{A,*}(c)\right)\right)^{2/3}. \end{aligned}$$

[12] 中的定理 4.1 表明

$$\left(\mathbb{E}_A\left[(\tilde{T}_{A,*}(c))^3\right]\right)^{1/3} = O(L_c).$$

在引理 6 中，我们证明了对于任意 $(B,D) \neq (A, D_A^{\mathcal{W}})$，$\mathbb{P}_A(\tilde{T}_{B,D}(c) < \tilde{T}_{A,*}(c)) = O\left((L_c)^{-3/2}\right)$。然后，由于 A 是有限集，根据并集界可得

$$\mathbb{P}_A\left(\tilde{T}(c) < \tilde{T}_{A,*}(c)\right) \le \sum_{(B,D)\in\mathcal{A}^2\setminus\{(A,D_A^{\mathcal{W}})\}} \mathbb{P}_A\left(\tilde{T}_{B,D}(c) < \tilde{T}_{A,*}(c)\right) = O\left((L_c)^{-3/2}\right).$$

因此，我们有

$$\left|\mathbb{E}_A\left[(\tilde{T}_{A,*}(c) - \tilde{T}(c))\mathbb{1}\{\tilde{T}(c) < \tilde{T}_{A,*}(c)\}\right]\right| = O(1),$$

根据（43），这意味着

$$\mathbb{E}_A\left[\tilde{T}(c)\right] = \mathbb{E}_A\left[\tilde{T}_{A,*}(c)\right] + O(1).$$

最后，注意到根据定义，

$$\tilde{T}_{A,*}(c) = \min\left\{t \ge 1 : \min_{C\in\mathcal{H}_{A,*}^{\mathcal{W}}}\left(\sum_{s=1}^{t}\log\frac{f_A(\mathbf{X}_s^{[K]})}{f_C(\mathbf{X}_s^{[K]})}\right) \ge L_c\right\}.$$

此外，回忆（18）可知，$\mathrm{KL}_{A,*}^{\mathcal{W}} = \min_{C\in\mathcal{H}_{A,*}^{\mathcal{W}}} \mathrm{KL}(f_A|f_C)$。

第（a）部分 $r_A = 1$。由定理 9，当 $c \to 0$ 我们有

$$\mathbb{E}_A\left[\tilde{T}_{A,*}(c)\right] = \frac{L_c}{\mathrm{KL}_{A,*}^{\mathcal{W}}} + O(1).$$

第（b）部分 $r_A \ge 2$。根据定理 10，当 $c \to 0$ 时，我们有

$$\mathbb{E}_A(\tilde{T}_{A,*}) = \frac{L_c}{\mathrm{KL}_{A,*}^{\mathcal{W}}} + \frac{h_A^{\mathcal{W}}\sqrt{L_c}}{(\mathrm{KL}_{A,*}^{\mathcal{W}})^{3/2}} + O(L_c^{1/4+\epsilon/2}).$$

证明完毕。 □

B.1. 辅助引理

回顾 $L_c := \log(c^{-1})$ 在（42）中，以及停止时间 $\tilde{T}_{B,D}(c, \mathrm{W})$ 在（41）中。对于 $B, G \in \mathcal{A}$，我们定义

$$\widehat{T}_{B,G}(c) := \min\left\{t \geq 1 : f_{G,[t]} \leq c f_{B,[t]}\right\}. \tag{44}$$

对于 B，G，$D \in \mathsf{A}$，我们定义如下量

$$\mathrm{Div}_A(f_B, f_G) := \mathbb{E}_A\left[\log\left(\frac{f_B(\mathbf{X}_1^{[K]})}{f_G(\mathbf{X}_1^{[K]})}\right)\right] = \int \log(f_B/f_G) f_A \, d(\prod_{k\in[K]} \mu_k),$$

并进一步记为

$$\mathrm{Div}_A^{\mathcal{W}}(B, D) := \min_{G \in \mathcal{H}_D^{\mathcal{W}}} \mathrm{Div}_A(f_B, f_G). \tag{45}$$

引理 6。设假设 1 成立。令 $A \in \mathcal{A}$，并假设 A 关于 $\mathcal{W}$ 具有唯一最有利子集。令 $(B, D) \in \mathcal{A} \times \mathcal{A}$，使得 $(B, D) \neq (A, D_{\mathcal{W}A})$，其中 $D_A^{\mathcal{W}}$ 出现在定义 4 中。假设 (22) 对 $q = 3$ 成立。则

$$\mathbb{P}_A\left(\tilde{T}_{B,D}(c, \mathcal{W}) < \tilde{T}_{A,D_A^{\mathcal{W}}}(c, \mathcal{W})\right) = O\left((L_c)^{-3/2}\right).$$

证明。 若 $B \neq A$，根据定义及假设 1，对任意 $G \in \mathcal{A}$，

$$\mathrm{Div}_A(f_B, f_G) = \mathrm{KL}(f_A \mid f_G) - \mathrm{KL}(f_A \mid f_B) < \mathrm{Div}_A(f_A, f_G) = \mathrm{KL}(f_A | f_G).$$

由于 $\mathcal{A}$ 是有限的，且根据（18），我们有

$$\mathrm{Div}_A^{\mathcal{W}}(B, D) < \min_{G \in \mathcal{H}_D^{\mathcal{W}}} \mathrm{KL}(f_A | f_G) = \mathrm{KL}_{A,D}^{\mathcal{W}} \leq \mathrm{KL}_{A,*}^{\mathcal{W}},$$

若 $B = A$，由于 $(B, D) \neq (A, D_A^{\mathcal{W}})$，我们必有 $D \neq D_{\mathcal{W}A}$。又由于 A 关于 W 存在唯一最有利子集，我们有

$$\mathrm{Div}_A^{\mathcal{W}}(B, D) = \mathrm{Div}_A^{\mathcal{W}}(A, D) = \mathrm{KL}_{A,D}^{\mathcal{W}} < \mathrm{KL}_{A,*}^{\mathcal{W}},$$

其中我们利用了这样的事实：$D_{\mathcal{W}A}$ 是 $\max_{G \in \mathcal{A}} \mathrm{KL}_A^{\mathcal{W},G}$ 的唯一最大化者。

因此，在两种情况下，$\mathrm{KL}_{A,*}^{\mathcal{W}} > \mathrm{Div}_A^{\mathcal{W}}(B, D)$，这意味着存在 $\mu^* > 0$，使得

$$\mathrm{KL}_{A,*}^{\mathcal{W}} > \mu^* > \mathrm{Div}_A^{\mathcal{W}}(B, D).$$

由于并集界，我们有

$$\mathbb{P}_A\left(\tilde{T}_{B,D}(c) < \tilde{T}_{A,D_A^{\mathcal{W}}}(c)\right) \leq \mathbb{P}_A\left(\tilde{T}_{B,D}(c) \leq \frac{L_c}{\mu^*}\right) + \mathbb{P}_A\left(\tilde{T}_{A,D_A^{\mathcal{W}}}(c) \geq \frac{L_c}{\mu^*}\right).$$

我们从上述上界中的第一项开始。根据定义（45），存在某个 $G^* \in \mathcal{H}_D^{\mathcal{W}}$使得

$$\mathrm{Div}_A(f_B, f_{G^*}) = \mathrm{Div}_A^{\mathcal{W}}(B, D) < \mu^*. \tag{46}$$

根据定义（44）和（41），$\tilde{T}_{B,\,D}(c) \geq \hat{T}_{B,\,G^*}(c)$，因此

$$\mathbb{P}_A\left(\tilde{T}_{B,D}(c) \leq \frac{L_c}{\mu^*}\right) \leq \mathbb{P}_A\left(\hat{T}_{B,G^*}(c) \leq \frac{L_c}{\mu^*}\right)$$

由（46）以及引理 7(a)，可得 $\mathbb{P}_A\left(\tilde{T}_{B,D}(c) \leq \frac{L_c}{\mu^*}\right) = O(L_c^{-3/2})$。

最后，由引理 8，$\mathbb{P}_A\left(\tilde{T}_{A,D_A^{\mathcal{W}}}(c) \geq \frac{L_c}{\mu^*}\right) = O\left((L_c)^{-3/2}\right)$。证明至此完成。 □

回顾 $\hat{T}_{B,\,G}$ 在（44）中的定义。

引理 7. 假设假设条件 *1* 成立，且 (22) 对 $q = 3$成立。设 A，B，$G \in \mathcal{A}$ 满足$B \neq G$。

(a) 对于每个 $x > \max\{0、\mathrm{Div}_A(f_B, f_G)\}$，当 $c \to 0$ 时，

$$\mathbb{P}_A\left(\widehat{T}_{B,G}(c) \leq \frac{L_c}{x}\right) = O\left((L_c)^{-3/2}\right).$$

（b）假设 $G \neq A$。那么对每个 $x \in$（0,KL散度（$f_A|f_G$），当 $c \to 0$ 时，

$$\mathbb{P}_A\left(\widehat{T}_{A,G}(c) \geq \frac{L_c}{x}\right) = O\left((L_c)^{-3/2}\right).$$

证明。 注意，$\{\mathbf{X}_t^{[K]}: t \geq 1\}$ 是由 $i_。$ $i_。$ $d_。$ 随机向量构成的集合，且

$$\mathbb{E}_A\left[\log \frac{f_B(\mathbf{X}_1^{[K]})}{f_G(\mathbf{X}_1^{[K]})}\right] = \mathrm{Div}_A(f_B, f_G)$$

并且由于（22）对 $q = 3$成立，因此有 $\mathbb{E}_A\left|\text{对数}\left(f_B\left(\mathbf{X}_1^{[K]}\right) / f_G\left(\mathbf{X}_1^{[K]}\right)\right)\right|^3 < \infty_。$

第（a）部分。 令 θ 为任意常数，使得 $\max\{0, \mathrm{Div}_A(f_B, f_G)\} < \theta < x_。$ 此外，定义

$$\tau_{B,G}(c) := \min\left\{t \geq 1 : \sum_{s=1}^{t}\left(\log \frac{f_B(\mathbf{X}_s^{[K]})}{f_G(\mathbf{X}_s^{[K]})} + \theta - \mathrm{Div}_A(f_B, f_G)\right) \geq L_c\right\}.$$

根据 θ 的定义，显然 $\tau_{B,\,G}(c) \leq T_{B,\,G}(x)$，并且

$$\mathbb{E}_A\left[\log \frac{f_B(\mathbf{X}_1^{[K]})}{f_G(\mathbf{X}_1^{[K]})} + \theta - \mathrm{Div}_A(f_B, f_G)\right] = \theta > 0.$$

因此，根据 [14], 第 3 章中的定理 8.4作为 $c \to 0$,

$$\mathbb{E}_A\left|\tau_{B,G} - \frac{L_c}{\theta}\right|^3 = O(L_c^{3/2}).$$

最后，注意到

$$\mathbb{P}_A\left(\widehat{T}_{B,G} \le \frac{L_c}{x}\right) \le \mathbb{P}_A\left(\tau_{B,G} \le \frac{L_c}{x}\right) \le \mathbb{P}_A\left(\left|\tau_{B,G} - \frac{L_c}{\theta}\right| \ge L_c\left(\frac{1}{\theta} - \frac{1}{x}\right)\right).$$

由于 $0 < \theta < x$，$1/\theta - 1/x > 0$。根据马尔可夫不等式，

$$\mathbb{P}_A\left(\left|\tau_{B,G} - \frac{L_c}{\theta}\right| \ge L_c\left(\frac{1}{\theta} - \frac{1}{x}\right)\right) = O((L_c)^{-3/2}),$$

由此完成第（a）部分的证明。

第（b）部分。注意，$\mathrm{Div}_A(f_A, f_G) = \mathrm{KL}(f_A \mid f_G) > 0$。因此，该论证与第（a）部分相似且更简单，故省略细节。 □

回顾停止时间 $\tilde{T}_{B,D}(c, \mathcal{W})$（41）以及 $\hat{T}_{B,G}$（44）中的定义。

引理 8。假设条件 *1* 成立，且对于 $q = 3$，(22) 成立。设 $A \in \mathcal{A}$，并假设相对于 $\mathcal{W}$，A 有唯一最有利子集。那么，对于每个 $0 < \mu^* <$KL散度$^{\mathcal{W}}_{A,*}$，当$c \to 0$时，

$$\mathbb{P}_A\left(\tilde{T}_{A,D_A^{\mathcal{W}}}(c) \ge \frac{L_c}{\mu^*}\right) = O((L_c)^{-3/2}).$$

证明。证明过程与 [26] 中定理2的证明类似。回顾定义 4 中的 $D_A^{\mathcal{W}}$。为简单起见，我们定义 $\mathrm{H}^{\mathcal{W}}_{A,*} := \mathrm{H}_{\mathcal{W}A, D_A^{\mathcal{W}}}$。

首先，对于每个整数 $s_0 \ge 0$ 和 $G \in \mathrm{A}$，定义

$$\tau_G(s_0) := \max\left\{t \ge 1 : \sum_{s=s_0+1}^{s_0+t} \log\frac{f_A(\mathbf{X}_s^{[K]})}{f_G(\mathbf{X}_s^{[K]})} \le 0\right\}.$$

对于每个 $G \in \mathrm{H}^{\mathcal{W}}_{A,*}$，根据定义，

$$\sum_{s=1}^{t} \log\frac{f_A(\mathbf{X}_s^{[K]})}{f_G(\mathbf{X}_s^{[K]})} \le L_c, \text{ for all } t \ge \widehat{T}_{A,G}(c) + \tau_G\left(\widehat{T}_{A,G}(c)\right) + 1.$$

因此，

$$\tilde{T}_{A,D_A^{\mathcal{W}}}(c) \le \max_{G \in \mathcal{H}^{\mathcal{W}}_{A,*}} \left\{\widehat{T}_{A,G}(c) + \tau_G\left(\widehat{T}_{A,G}(c)\right) + 1\right\}.$$

令 $\theta \in \mathbb{R}$ 满足 $\mu^* < \theta <$KL散度$^{\mathcal{W}}_{A,*}$。由于并集界，

$$\mathbb{P}_A\left(\tilde{T}_{A,D_A^{\mathcal{W}}}(c) \ge \frac{L_c}{\mu^*}\right) \le \sum_{G \in \mathcal{H}^{\mathcal{W}}_{A,*}} \mathbb{P}_A\left(\widehat{T}_{A,G}(c) \ge \frac{L_c}{\theta}\right) + \sum_{G \in \mathcal{H}^{\mathcal{W}}_{A,*}} \mathbb{P}_A\left(\tau_G(\widehat{T}_{A,G}(c)) + 1 \ge \frac{L_c}{\mu^*} - \frac{L_c}{\theta}\right). \tag{47}$$

首先，我们考虑式（47）右侧的第一项。根据定义（18），

$$\theta < \mathrm{KL}^{\mathcal{W}}_{A,*} = \mathrm{KL}^{\mathcal{W}}_{A,D^{\mathcal{W}}_A} \le \mathrm{KL}(f_A|f_G), \quad \text{for each } G \in \mathcal{H}^{\mathcal{W}}_{A,*},$$

根据引理 7(b)，这意味着对每个 $G \in \mathrm{H}^{\mathcal{W}}_{A*}$，当 $c \to 0$ 时，

$$\mathbb{P}_A\left(\widehat{T}_{A,G}(c) \ge \frac{L_c}{\theta}\right) = O(L_c^{-3/2}).$$

接下来，我们考虑式（47）右侧的第二部分。由于 $\{\mathbf{X}_t^{[K]} : t \ge 1\}$ 是一组 *i.i.d.* 随机变量，对每个 $G \in \mathrm{H}_{\mathcal{W}A,*}$，$\tau_G(\widehat{T}_{A,G}(c))$ 与 $\tau_G(0)$ 在 $\mathbb{P}_A$ 下具有相同的分布。由于

$$\mathbb{E}_A\left[\log \frac{f_A(\mathbf{X}_s^{[K]})}{f_G(\mathbf{X}_s^{[K]})}\right] = \mathrm{KL}(f_A|f_G) > 0,$$

根据 [18], 中的定理 1，当 $c \to 0$ 时，我们有 $\mathbb{E}_A\left[(\tau_G(0))^2\right] = O(1)$。最后，由马尔可夫不等式，且由于 $1/\mu^* - 1/\theta > 0$，对每个 $G \in \mathrm{H}^{\mathcal{W}}_{A*}$，当 $c \to 0$ 时，

$$\mathbb{P}_A\left(\tau_G(\widehat{T}_{A,G}(c)) \ge \frac{L_c}{\mu^*} - \frac{L_c}{\theta} - 1\right) = O(L_c^{-2}).$$

证明完毕。 □

附录 C：与多维随机游走相关的停止时刻的渐近有效样本量

令 $\boldsymbol{V}_1, \boldsymbol{V}_2, \ldots$ 为一列 d 维独立同分布随机向量，并记 $\boldsymbol{R}_n = \boldsymbol{V}_1 + \boldsymbol{V}_2 + \cdots + \boldsymbol{V}_n$ 为 $n \ge 1$ 所对应的随机游走。对于每个 $j \in [d]$，记 $\mu_j := \mathbb{E}[\boldsymbol{V}_{1,j}]$，其中 $\boldsymbol{V}_{1,j}$ 表示 $\boldsymbol{V}_1$ 的第 j 个分量。此外，对于 $t \ge 1$，记 $\boldsymbol{V}_{t,j} := \boldsymbol{V}_{t,j} - \mu_j$，于是

$$\boldsymbol{R}_{n,j} - n\mu_j = \sum_{t=1}^{n} \tilde{\boldsymbol{V}}_{t,j}, \quad \text{for } n \ge 1. \tag{48}$$

我们考虑停止时间

$$T_b := \inf\left\{n \ge 1 : \min_{j \in [d]} \boldsymbol{R}_{n,j} \ge b\right\}, \tag{49}$$

也就是说，这是随机游走的所有分量都超过阈值的首次时间。在本节中，我们假设

$$0 < \mu_1 \le \mu_2 \le \cdots \le \mu_d.$$

令 $1 \le r_* \le d$ 为满足以下条件的唯一整数

$$\mu_1 = \mu_2 = \ldots = \mu_{r_*} < \mu_{r_*+1}, \tag{50}$$

其中 $\mu_{d+1} := +\infty$。我们将 $r_* = 1$ 称为非对称设定，将 $r_* \neq 1$ 称为对称设定。

在本节中，对于某个待指定的 $q \geq 2$，我们假设

$$\mathbb{E}\left(|\tilde{\boldsymbol{V}}_{1,j}|^q\right) < \infty \text{ for all } j \in [d]. \tag{51}$$

定理 9. 考虑非对称设定，即 $r_* = 1$。假设(51) 对某个 $q > 2\sqrt{2}$成立。则当 $b \to \infty$时，

$$\mathbb{E}(T_b) = \frac{b}{\mu_1} + O(1).$$

证明。 我们应用 [27]中的定理 4.2。具体而言，令 $Z_n := \min_{j\in[d]} \boldsymbol{R}_{n,\,j}$（针对 $n \geq 1$），并将$\{Z_n : n \geq 1\}$改写为扰动随机游走：

$$Z_n = S_n + \xi_n, \quad \text{where} \quad S_n := \boldsymbol{R}_{n,1}, \quad \text{and} \quad \xi_n := \min\{0, \boldsymbol{R}_{n,2} - \boldsymbol{R}_{n,1}, \ldots, \boldsymbol{R}_{n,d} - \boldsymbol{R}_{n,1}\}. \tag{52}$$

则 T_b 在(49)中可写为

$$T_b = \inf\{n \geq 1 : Z_n \geq b\} = \inf\{n \geq 1 : S_n + \xi_n \geq b\}.$$

此外，由于 $q > 2\sqrt{2}$，存在某个 $\alpha^* \in (1/2, 1]$，使得

$$\frac{2}{q} \leq \alpha^* < \frac{q}{4}.$$

引理 9 表明过程 $\{\xi_n : n \geq 1\}$ 是 1-正则的，其参数为 $\mu = \mu_1, \alpha = \alpha^*$，以及 $p = 1$（见定义5），常数如下：

$$\delta_0 = \frac{1}{2}, \ \theta = \frac{\mu}{2}, \ \omega_0 = \frac{\mu}{4}, \ K = 3, \ \theta^* = \frac{\mu}{2}.$$

对 [11] 中 (A.16) 的验证表明，[27] 中的条件 (4.2) 对 $\delta_0 = 1/2$ 成立。此外，由于 $\alpha^* \geq 2/q$，我们有 $2/\alpha^* \leq q$，因此 $\mathbb{E}[|\tilde{\boldsymbol{V}}_{1,1}|^{2/\alpha^*}] < \infty$。

因此，根据 [27], 中的定理 4.2

$$\mu_1 \mathbb{E}(T_b) = b - \mathbb{E}(\zeta_{n_b}) + O(1),$$

其中 $n_b := \lfloor b/\mu_1 \rfloor$ 且 $\zeta_n := (\xi_n \wedge \theta n^\alpha) \vee (-\theta^* n^\alpha)$ 对 $n \geq 1$。此外，引理 9 表明$\mathbb{E}(\zeta_{n_b}) = o(1)$ 当 $b \to \infty$时。证明完毕。 □

接下来，我们转向对称设定，即 $r_* \geq 2$。定义

$$\tilde{\boldsymbol{V}}_{1,[r_*]} := \left(\tilde{\boldsymbol{V}}_{1,1}, \ldots, \tilde{\boldsymbol{V}}_{1,r_*}\right)^\top, \qquad \boldsymbol{\Sigma}_{[r_*]} := \mathbb{E}\left(\tilde{\boldsymbol{V}}_{1,[r_*]} \tilde{\boldsymbol{V}}_{1,[r_*]}^\top\right) \in \mathbb{R}^{r_* \times r_*},$$

使得 $\boldsymbol{\Sigma}_{[r_*]}$是 $\tilde{\boldsymbol{V}}_{1\,\text{的}[r_*]}$协方差矩阵。令 $(Y_1, \ldots, Y_{r_*})^\top$ 为协方差矩阵为 $\boldsymbol{\Sigma}_{[r_*]}$ 的零均值高斯随机向量。最后，定义

$$h_{[r_*]} := \mathbb{E}\left[\max\{Y_1, \ldots, Y_{r_*}\}\right]. \tag{53}$$

定理 10。考虑对称设定，即 $r_* \geq 2$。设 $\epsilon \in (0,\ 1/2)$ 为一个常数，并假设(51) 对某个 q 满足条件

$$q > \frac{1}{\epsilon} \quad and \quad q \geq 3. \tag{54}$$

于是，当 $b \to \infty$ 时，

$$\mathbb{E}(T_b) = \frac{b}{\mu_1} + \frac{h_{[r_*]}\sqrt{b}}{\mu_1^{3/2}} + O(b^{1/4+\epsilon/2}).$$

证明。我们在 [27] 中应用定理 4.2。记 $Z_n = \min_{j\in[d]} \boldsymbol{R}_{n,\ j}$，其中 $n \geq 1$，并将 $\{Z_n\colon n \geq 1\}$ 重写为一个扰动随机游走：

$$Z_n = S_n + \xi_n,\ \text{where } S_n := n\mu_1,\ \xi_n := \min\{\boldsymbol{R}_{n,1} - n\mu_1, \cdots, \boldsymbol{R}_{n,d} - n\mu_1\} \tag{55}$$

那么 T_b 在 (49) 中可写为

$$T_b = \inf\{n \geq 1 : Z_n \geq b\} = \inf\{n \geq 1 : S_n + \xi_n \geq b\}.$$

引理 10 表明，过程 $\{\xi_n\colon n \geq 1\}$ 是 $n^{\alpha/2}$-正则的，其参数为 $\mu = \mu_1$，$\alpha = 1/2 + \epsilon$，以及$p = 1$，常数为

$$\delta_0 = \frac{1}{2},\ \theta = \frac{\mu}{2},\ \omega_0 = \frac{\mu}{4},\ K = 3,\ \theta^* = \frac{\mu}{2}.$$

参见定义 5。在 [11] 中对 (A.16) 的验证表明，[27] 中的 (4.2) 在 $\delta_0 = 1/2$ 时成立。此外，由于 $S_n = n\mu_1$，这是一个增量为 μ_1 的确定性随机游走。

因此，根据 [27], 中的定理 4.2，我们有

$$\mu_1\mathbb{E}(T_b) = b - \mathbb{E}(\zeta_{n_b}) + O(b^{\alpha/2}),$$

其中 $n_b :=\ \lfloor b/\mu_1 \rfloor$ 且 $\zeta_n :=\ (\xi_n \wedge \theta n^\alpha) \vee (-\theta^* n^\alpha)$ 对于 $n \geq 1$。此外，引理 10 表明$\mathbb{E}(\zeta_{n_b})$ $= -h_{[r_*]}\sqrt{n_b} + O(1)$ 当 $b \to \infty$时。证明完毕。 □

C.1. 辅助引理

我们考虑一个函数 $\rho(\cdot) : \{1, 2, \cdots\} \mapsto [0, \infty)$，使得

$$1 \leq \rho(x) = o(x) \text{ as } x \to \infty, \quad \sup_{x\geq 1}\sup_{x\leq t\leq 3x} \rho(t)/\rho(2x) < \infty.$$

我们回顾随机过程 $\{\xi_n \in \mathbb{R}\colon n \geq 1\}$ 的 ρ-正则性定义，见 [27]。

定义 5. $\{\xi_n \in \mathbb{R}\colon n \geq 1\}$ 称为具有参数 $\mu > 0$、$1/2 < \alpha \leq 1$ 和 $p \geq 1$ 以及常数的 ρ-正则

$$\delta_0, \theta > 0,\ \theta < \mu \ \text{ for } \alpha = 1, K > 0, \omega_0 > 0, 0 < \theta^* < K\mu,$$

若以下条件成立：

$$\mathbb{P}(\max_{\delta_0 n < k \leq n} \xi_k > \theta n^\alpha) = o(\rho(n)/n)^p,$$

$$\sum_{k\geq n+Kn^{\alpha}} k^{p-1}\mathbb{P}(\xi_k \leq -(k-n)\mu + \omega_0 k^{\alpha}) = o(\rho(n))^p,$$

$$\left\{ \max_{0\leq k\leq Mn^{\alpha}} \left(\frac{|\zeta_n - \xi_{n+k}| \wedge n^{\alpha}}{\rho(n)} \right)^p : n \geq 1 \right\} \text{ is uniformly integrable for all } M > 0,$$

其中，我们定义

$$\zeta_n := (\xi_n \wedge \theta n^{\alpha}) \vee (-\theta^* n^{\alpha}) = \begin{cases} \theta n^{\alpha} & \text{for } \xi_n > \theta n^{\alpha} \\ \xi_n & \text{for } -\theta^* n^{\alpha} \leq \xi_n \leq \theta n^{\alpha} \\ -\theta^* n^{\alpha} & \text{for } \xi_n < -\theta^* n^{\alpha} \end{cases}. \tag{56}$$

注 14. 对于第二个条件，由于 $(k-n)/k^{\alpha}$ 关于 k 在 $k \geq n + Kn^{\alpha}$ 和 $1/2 < \alpha \leq 1$上递增，我们有

$$-(k-n)\mu + \omega_0 k^{\alpha} \leq \left(-\frac{K\mu}{(K+1)^{\alpha}} + \omega_0 \right) k^{\alpha}. \tag{57}$$

对于 $p > 0$，我们将随机变量 $X \in \mathbb{R}$的l^p范数记为 $\|X\|_p = [\mathbb{E}(|X|^p)]^{1/p}$。对于 $a, b \in \mathbb{R}$，令 $a \vee b = \max\{a, b\}$。

引理 9. 考虑非对称设定，即 $r_* = 1$。假设 (51) 对某个 $q > 2$成立，并令 α^*为任意常数，使得 $\alpha^* \in (1/2, q/4)$。对每个整数 $n \geq 1$，定义

$$\rho(n) = 1.$$

则下列结论对于 $\{\xi_n : n \geq 1\}$ （定义于 (52)）以及 $\{\zeta_n : n \geq 1\}$ （见 (56)）成立：

1. $\{\xi_n : n \geq 1\}$ 是具有参数 $\mu = \mu_1$、$\alpha = \alpha^*$ 和 $p = 1$ 的 ρ-正则，且常数

$$\delta_0 = \frac{1}{2},\ \theta = \frac{\mu}{2},\ \omega_0 = \frac{\mu}{4},\ K = 3,\ \theta^* = \frac{\mu}{2}. \tag{58}$$

2. $\mathbb{E}(\zeta_n) = o(1)$ 当 $n \to \infty$。

证明。 在此证明中，我们用 C 表示一个不依赖于 n、但可能依赖于 q 和 $\{\mu_j,\ \|\tilde{\boldsymbol{V}}_{1,\ j}\|_q : 1 \leq j \leq d\}$ 的常数。

尾概率界。 根据定义（52），$\xi_n \leq 0$ 对于 $n \geq 1$。此外，由于并集界和推论 1，存在一个常数 $C > 0$，使得对于 $n \geq 1$,

$$\mathbb{P}(\xi_n \neq 0) = \mathbb{P}(\xi_n < 0) \leq \sum_{2\leq j\leq d} \mathbb{P}(\boldsymbol{R}_{n,j} - \boldsymbol{R}_{n,1} < 0) \leq Cdn^{-q/2}. \tag{59}$$

首先 定义 5 中的条件。 根据定义（52），我们有 最大值$_{1\leq k\leq n}\xi_k \leq 0$，这意味着 th at

$$\mathbb{P}\left(\max_{\delta_0 n\leq k\leq n} \xi_k \geq \theta n^{\alpha} \right) = 0.$$

第二个条件（定义5）。由于（57）和（58），我们有

$$-(k-n)\mu + \omega_0 k^{\alpha} \leq -\frac{\mu}{2} k^{\alpha} < 0.$$

则由（59），存在常数 $C>0$，使得对每个 $k \geq n+Kn^{\alpha}$，我们有，

$$\mathbb{P}(\xi_k \leq -(k-n)\mu + \omega_0 k^{\alpha}) \leq \mathbb{P}(\xi_k < 0) \leq Cdk^{-q/2}.$$

因此，对于 $p=1$，我们有

$$\sum_{k=n+Kn^{\alpha}}^{\infty} k^{p-1}\mathbb{P}(\xi_k \leq -(k-n)\mu + \omega_0 k^{\alpha}) \leq \sum_{k=n+Kn^{\alpha}} Cdk^{-q/2} = o(1),$$

其中，最后一个不等式中 $q>2$。

定义中的第三个条件5。首先，根据定义（52），注意到

$$|\zeta_n - \xi_{n+k}| \leq |\zeta_n - \xi_n| + |\xi_n - \xi_{n+k}| \leq |\zeta_n - \xi_n|\mathbb{1}\{\xi_n \neq 0\} + |\xi_n - \xi_{n+k}|\mathbb{1}\left\{\bigcup_{k=1}^{Mn^{\alpha}} \{\xi_{n+k} \neq 0\}\right\}.$$

因此，我们有

$$|\zeta_n - \xi_{n+k}| \wedge n^{\alpha} \leq a_n + b_n, \text{ where we define}$$

$$a_n := n^{\alpha}\mathbb{1}\{\xi_n \neq 0\}, \quad b_n = n^{\alpha}\mathbb{1}\left\{\bigcup_{k=1}^{Mn^{\alpha}} \{\xi_{n+k} \neq 0\}\right\}.$$

我们证明 $\{a_n, b_n : n \geq 1\}$ 是一致可积的。根据 [13], 中的定理 4.6.3，只需证明 $\mathbb{E}[a_n] \to 0$ 和 $\mathbb{E}[b_n] \to 0$。

由（59）可知，存在一个常数 $C>0$ 使得

$$\mathbb{E}(a_n) \leq n^{\alpha}\mathbb{P}(\xi_n \neq 0) \leq Cdn^{\alpha - q/2} \to 0 \text{ as } n \to \infty,$$

$$\mathbb{E}(b_n) \leq n^{\alpha} \sum_{k=1}^{Mn^{\alpha}} \mathbb{P}(\xi_{n+k} \neq 0) \leq C(M+1)dn^{2\alpha - q/2} \to 0 \text{ as } n \to \infty,$$

这里我们用到了 $\alpha = \alpha^* < q/4$ 这一事实。因此，第 1 部分的证明完成。

第 2 部分。期望 ζ_n。根据定义（52）和（56），我们有

$$|\zeta_n| \leq |\xi_n| \wedge \frac{\mu}{2}n^{\alpha} \leq \frac{\mu}{2}n^{\alpha}\mathbb{1}\{\xi_n \neq 0\}.$$

因此，由（59），存在一个常数 $C>0$，使得

$$\mathbb{E}(|\zeta_n|) \leq \frac{\mu}{2}n^{\alpha}\mathbb{P}(\xi_n \neq 0) \leq Cdn^{\alpha - q/2} = o(1),$$

其中　对于最后一个不等式，我们用到了 $\alpha = \alpha^* < q/4 < q/2$ 这一事实。于是证明完成。 □

引理 10。考虑对称情形，即 $r_* \geq 2$。设 $\epsilon \in (0,1/2)$ 为一个常数，并假设 (51)对某个满足 q (54)的量成立。定义

$$\alpha^* := 1/2 + \epsilon, \quad \textit{and} \quad \rho(n) := n^{\alpha^*/2} \textit{ for integer } n \geq 1.$$

那么，对于以下在 $\{\xi_n : \ n \geq 1\}$ 中定义的 (55) 以及 ζ_n 在 (56)中：

1. $\{\xi_n：\ n\geq 1\}$ 是 ρ-正则的，其参数为 $\mu=\mu_1$、$\alpha=\alpha^*$ 和 $p=1$，常数

$$\delta_0=\frac{1}{2},\ \theta=\frac{\mu}{2},\ \omega_0=\frac{\mu}{4},\ K=3,\ \theta^*=\frac{\mu}{2}. \tag{60}$$

2. $\mathbb{E}(\zeta_n)=-h_{[r_*]}\sqrt{n}+O(1)$ 为 $n\to\infty$，其中 $h_{[r_*]}$ 定义见 (53)。

证明。在本证明中，我们用 C 表示一个不依赖于 n，但可能依赖于 q，$\{\mu_j,\ \|\tilde{\boldsymbol{V}}_{1,\ j}\|_q：1\leq j\leq d\}$的常数。由于 $\epsilon\in(0,\ 1/2)$，根据（54）中的定义，我们有

$$q\geq 3,\qquad q>\frac{1}{\epsilon},\qquad q>\frac{3}{4\epsilon}-\frac{1}{2},\qquad q>\frac{1}{2\epsilon}+\frac{1}{2},\quad \text{and}\quad q>\frac{3}{2}+3\epsilon=3\alpha. \tag{61}$$

尾概率界。对于每个 $n\geq 1$，我们定义

$$\hat{\xi}_n:=\min\{\boldsymbol{R}_{n,1}-n\mu_1,\cdots,\boldsymbol{R}_{n,r_*}-n\mu_1\},\quad \check{\xi}_n:=\min\{\boldsymbol{R}_{n,r_*+1}-n\mu_1,\cdots,\boldsymbol{R}_{n,d}-n\mu_1\}, \tag{62}$$

其中我们采用约定 $\min\{\emptyset\}：=+\infty$。然后，根据定义（55），我们有 $\xi_n=\min\left\{\hat{\xi}_n,\ \check{\xi}_n\right\}$。由并集界和推论 1，存在一个常数 $C>0$，使得对于 $n\geq 1$，有

$$\mathbb{P}(\check{\xi}_n<\hat{\xi}_n)=\sum_{j=r_*+1}^{d}\sum_{i=1}^{r_*}\mathbb{P}\left(\boldsymbol{R}_{n,j}-\boldsymbol{R}_{n,i}<0\right)\leq Cr_*(d-r_*)n^{-q/2}. \tag{63}$$

由于 $\mu_1=\cdots=\mu_{r_*}$ 在定义（62）中，对于每个 $1\leq k\leq n$,

$$|\hat{\xi}_k|\leq\max_{j\in[r_*]}|\boldsymbol{R}_{k,j}-k\mu_j|\leq\sum_{j\in[r_*]}\left(\max_{1\leq k\leq n}|\boldsymbol{R}_{k,j}-k\mu_j|\right).$$

因此，根据三角不等式和引理 11，存在一个常数 $C>0$，使得

$$\left\|\max_{1\leq k\leq n}|\hat{\xi}_k|\right\|_q\leq\sum_{j\in[r_*]}\left\|\max_{1\leq k\leq n}|\boldsymbol{R}_{k,j}-k\mu_j|\right\|_q\leq Cr_*n^{1/2}, \tag{64}$$

并且根据马尔可夫不等式，对于任意 $\delta>0$，有

$$\mathbb{P}\left(\max_{1\leq k\leq n}|\hat{\xi}_k|\geq\delta\right)\leq\left\|\max_{1\leq k\leq n}|\hat{\xi}_k|\right\|_q^q/\delta^q\leq Cr_*^qn^{q/2}/\delta^q. \tag{65}$$

定义 5 中的第一个条件。由于 $\xi_n=\min\left\{\hat{\xi}_n,\ \check{\xi}_n\right\}$ 并且根据（65），存在一个常数 $C>0$使得

$$n^{-(\alpha/2-1)}\mathbb{P}\left(\max_{\delta_0n\leq k\leq n}\xi_k\geq\theta n^\alpha\right)\leq n^{-(\alpha/2-1)}\mathbb{P}\left(\max_{1\leq k\leq n}|\hat{\xi}_k|\geq\theta n^\alpha\right)\leq Cr_*^qn^{-(\alpha/2-1)-(\alpha-1/2)q}.$$

由于 $q>（1-\alpha/2）/（\alpha-1/2）=3/（4\epsilon）-1/2$ 根据（61），最后一项为 $o（1）$，因此

$$\mathbb{P}\left(\max_{\delta_0n\leq k\leq n}\xi_k\geq\theta n^\alpha\right)=o((n^{\alpha/2}/n)^p),$$

其中我们回顾一下，$p=1$。

定义中的第二个条件5。 根据（57）和（60），我们有

$$-(k-n)\mu+\omega_0 k^\alpha \le -\frac{\mu}{2}k^\alpha.$$

此外，由于并集界，（63）和（65），存在常数 $C>0$，使得对于 $k\ge n+Kn^\alpha$，

$$\begin{aligned}&\mathbb{P}(\xi_k \le -(k-n)\mu+\omega_0 k^\alpha) \le \mathbb{P}(\xi_k \le -\frac{\mu}{2}k^\alpha) \le \mathbb{P}(|\hat{\xi}_k| \ge \frac{\mu}{2}k^\alpha) + \mathbb{P}(\check{\xi}_k < \hat{\xi}_k)\\ &\le Cr_*^q k^{-(\alpha-1/2)q} + Cr_*(d-r_*)k^{-q/2} \le Cr_*^q(d-r_*)k^{-(\alpha-1/2)q},\end{aligned}$$

由于 $p=1$，由此可得

$$\sum_{k=n+Kn^\alpha}^{\infty} k^{p-1}\mathbb{P}(\xi_k \le -(k-n)\mu+\omega_0 k^\alpha) \le Cr_*^q(d-r_*)\sum_{k=n+Kn^\alpha}^{\infty} k^{-(\alpha-1/2)q}.$$

由于 $(\alpha-1/2)q=\epsilon q>\epsilon(1/\epsilon)=1$ 根据（61），我们有

$$\sum_{k=n+Kn^\alpha}^{\infty} k^{p-1}\mathbb{P}(\xi_k \le -(k-n)\mu+\omega_0 k^\alpha) = o(1) = o((n^{\alpha/2})^p).$$

定义 5 中的第三个条件。注意，在事件 $\{\check{\xi}_n \ge \hat{\xi}_n\}\cap\{\check{\xi}_{n+k} \ge \hat{\xi}_{n+k}\}$ 上，我们有 $\xi_{n+k}=\hat{\xi}_{n+k}$和 $\zeta_n=\hat{\zeta}_n$，其中我们记 $\hat{\zeta}_n := (\hat{\xi}_n \wedge \theta n^\alpha) \vee (-\theta^* n^\alpha)$。

根据三角不等式和并集界，可得

$$|\zeta_n-\xi_{n+k}| \le |\hat{\zeta}_n-\hat{\xi}_n| + |\hat{\xi}_n-\hat{\xi}_{n+k}| + (|\zeta_n-\xi_{n+k}|)\left(\mathbb{1}\{\check{\xi}_n<\hat{\xi}_n\}+\mathbb{1}\{\check{\xi}_{n+k}<\hat{\xi}_{n+k}\}\right),$$

因此，对于任意 $M>0$，

$$\max_{1\le k\le Mn^\alpha}|\zeta_n-\xi_{n+k}|\wedge n^\alpha \le a_n+b_n+c_n,$$

其中我们定义

$$a_n := n^\alpha \mathbb{1}\{|\hat{\xi}_n| \ge (\mu/2)n^\alpha\}, \qquad b_n := \max_{1\le k\le Mn^\alpha,\ j\in[r_*]}|\boldsymbol{R}_{n+k,j}-(n+k)\mu_j-(\boldsymbol{R}_{n,j}-n\mu_j)|,$$

$$c_n = \max_{1\le k\le Mn^\alpha}\left\{n^\alpha\left(\mathbb{1}\{\check{\xi}_n<\hat{\xi}_n\}+\mathbb{1}\{\check{\xi}_{n+k}<\hat{\xi}_{n+k}\}\right)\right\}$$

我们首先证明 $\{a_n/n^{\alpha/2}:\ n\ge 1\}$ 是一致可积的。根据 [13], 中的定理 4.6.3，只需证明 $\mathbb{E}(a_n/n^{\alpha/2}) \to 0$ 当 $n\to\infty$ 时成立。由 (65)，存在常数 $C>0$，使得

$$\mathbb{E}(|a_n/n^{\alpha/2}|) = n^{\alpha/2}\mathbb{P}\left(|\hat{\xi}_n| \ge (\mu/2)n^\alpha\right) \le n^{\alpha/2}\frac{Cr_*^q n^{q/2}}{((\mu/2)n^\alpha)^q} = Cr_*^q n^{\alpha/2+q/2-\alpha q}.$$

由于 $q>(\alpha/2)/(\alpha-1/2)=1/(4\epsilon)+1/2$ 根据 (61)，我们有 $\mathbb{E}(a_n/n^{\alpha/2})=o(1)$。

接下来，我们证明 $\{b_n/n^{\alpha/2}:\ n\ge 1\}$ 一致可积。由于增量是独立同分布的，对每个$n\ge 1$

$$b_n \text{ has the same distribution as } \max_{1\le k\le Mn^\alpha,\ j\in[r_*]}|\boldsymbol{R}_{n,j}-n\mu_j|.$$

然后，由于（64），存在一个常数 $C>0$，使得对每个 $n\geq 1$,

$$\mathbb{E}[|b_n/n^{\alpha/2}|^q]=\left\|\max_{1\leq k\leq Mn^{\alpha},\, j\in[r_*]}|\boldsymbol{R}_{k,j}-k\mu_j|\right\|_q^q/n^{\alpha q/2}\leq Cr_*^qM^{q/2},$$

由此可得 $\sup_{n\geq 1}\mathbb{E}\left[\left(|b_n/n^{\alpha/2}|\right)^q\right]<\infty$。随后，根据 [13], 中的定理 4.6.2，$\{b_n/n^{\alpha/2}:\, n\geq 1\}$ 是一致可积的。

最后，我们证明$\{c_n/n^{\alpha/2}:\, n\geq 1\}$ 是一致可积的。根据定理 4.6.3 在 [13], 中，只需证明 $\mathbb{E}\left(c_n/n^{\alpha/2}\right)\to 0$。根据（63）和并集界，存在常数 $C>0$，使得

$$\mathbb{E}(c_n/n^{\alpha/2})\leq\sum_{k=1}^{Mn^{\alpha}}n^{\alpha/2}\left[\mathbb{P}\left(\check{\xi}_n<\hat{\xi}_n\right)+\mathbb{P}\left(\check{\xi}_{n+k}<\hat{\xi}_{n+k}\right)\right]\leq Cr_*(d-r_*)Mn^{3\alpha/2-q/2}=o(1),$$

其中，对于最后一个不等式，我们利用了如下事实：$q>3\alpha$，这是由于（61）。第 1 部分的证明完成。

第 2 部分。关于 ζ_n的期望。根据上文对$\hat{}\,\zeta_n$ 的定义，并且由于（63）成立，存在常数$C>0$，使得

$$\left|\mathbb{E}[\zeta_n]-\mathbb{E}[\hat{\zeta}_n]\right|\leq\mathbb{E}\left[\mu n^{\alpha}\mathbb{1}\{\check{\xi}_n<\hat{\xi}_n\}\right]\leq Cr_*(d-r_*)n^{\alpha-q/2}.$$

由于 $q\geq 2\alpha$ 根据（61），我们有 $\mathbb{E}[\zeta_n]=\mathbb{E}[\hat{}\,\zeta_n]+O$（1）。

此外，根据（64），存在一个常数 $C>0$，使得

$$\left|\mathbb{E}(\hat{\xi}_n-\hat{\xi}_n)\right|\leq\mathbb{E}\left[|\hat{\xi}_n|\mathbb{1}\left\{|\hat{\xi}_n|\geq(\mu/2)n^{\alpha}\right\}\right]\leq\frac{\mathbb{E}|\hat{\xi}_n|^q}{[(\mu/2)n^{\alpha}]^{q-1}}\leq C(r_*)^qn^{-(\alpha-1/2)q+\alpha}.$$

由于 $q\geq\alpha/$（$\alpha-1/2$）$=1/$（2ϵ）$+1$ 根据（61），我们有 $\mathbb{E}$（$\hat{}\,\zeta_n$）$=\mathbb{E}$（$\hat{}\,\xi_n$）$+O$（1）。

最后，回顾$h_{[r_*]}$ 见（53）。由引理12,

$$\left|\mathbb{E}(\hat{\xi}_n)-(-h_{[r_*]}\sqrt{n})\right|=O(1),$$

这意味着 $\mathbb{E}[\zeta_n]=-h_{[r_*]}\sqrt{n}+O(1)$。证明完毕。 □

C.2 基本结果

引理 11. 假设 (51) 对某个 $q\geq 2$ 成立。则存在一个绝对常数 $C>0$，使得对每个 $j\in[d]$，我们有

$$\left\|\max_{1\leq k\leq n}|\boldsymbol{R}_{k,j}-k\mu_j|\right\|_q\leq Cq\left\|\tilde{\boldsymbol{V}}_{1,j}\right\|_q n^{1/2}.$$

证明。根据罗森塔尔不等式（例如参见 [10]中的定理 1.5.13），存在一个绝对常数$C>0$，使得对于每个 $j\in[d]$ 和 $q\geq 2$，我们有

$$\left\|\max_{1\leq k\leq n}\sum_{t=1}^{k}\tilde{\boldsymbol{V}}_{t,j}\right\|_q\vee\left\|\max_{1\leq k\leq n}\sum_{t=1}^{k}(-\tilde{\boldsymbol{V}}_{t,j})\right\|_q\leq Cq\left(\left\|\sum_{t=1}^{n}\tilde{\boldsymbol{V}}_{t,j}\right\|_2+\left\|\max_{1\leq t\leq n}|\tilde{\boldsymbol{V}}_{t,j}|\right\|_q\right).$$

根据独立性，且由于 $q \geq 2$，我们有

$$\left\|\sum_{t=1}^{n} \tilde{\boldsymbol{V}}_{t,j}\right\|_2 = \left\|\tilde{\boldsymbol{V}}_{1,j}\right\|_2 n^{1/2} \leq \left\|\tilde{\boldsymbol{V}}_{1,j}\right\|_q n^{1/2}, \text{ and } \left\|\max_{1\leq t\leq n} |\tilde{\boldsymbol{V}}_{t,j}|\right\|_q \leq \left[\sum_{t=1}^{n} \mathbb{E}(|\tilde{\boldsymbol{V}}_{t,j}|^q)\right]^{1/q} \leq \left\|\tilde{\boldsymbol{V}}_{1,j}\right\|_q n^{1/2}.$$

注意，对于任意 $1 \leq k \leq n$ 和 $j \in [d]$，我们都有

$$|\boldsymbol{R}_{k,j} - k\mu_j| = \max\left\{\sum_{t=1}^{k} \tilde{\boldsymbol{V}}_{t,j}, -\sum_{t=1}^{k} \tilde{\boldsymbol{V}}_{t,j}\right\},$$

这意味着，对于每个 $j \in [d]$,

$$\left\|\max_{1\leq k\leq n} |\boldsymbol{R}_{k,j} - k\mu_j|\right\|_q \leq 2\left\|\max_{1\leq k\leq n} \sum_{t=1}^{k} \tilde{\boldsymbol{V}}_{t,j}\right\|_q \vee \left\|\max_{1\leq k\leq n} \sum_{t=1}^{k} (-\tilde{\boldsymbol{V}}_{t,j})\right\|_q \leq 4Cq \left\|\tilde{\boldsymbol{V}}_{1,j}\right\|_q n^{1/2}.$$

于是证明完成。 □

推论 1。 假设 (51) 对某个 $q \geq 2$成立。令 $i < j \in [d]$ 满足 $\mu_i < \mu_j$。那么，存在一个绝对常数 $C > 0$ 使得

$$\mathbb{P}(\boldsymbol{R}_{n,j} - \boldsymbol{R}_{n,i} \leq 0) \leq \left(\frac{Cq}{\mu_j - \mu_i}\left(\left\|\tilde{\boldsymbol{V}}_{1,j}\right\|_q + \left\|\tilde{\boldsymbol{V}}_{1,i}\right\|_q\right)\right)^q n^{-q/2}.$$

证明。 注意到

$$\mathbb{P}(\boldsymbol{R}_{n,j} - \boldsymbol{R}_{n,i} \leq 0) = \mathbb{P}\left(\sum_{t=1}^{n} (\tilde{\boldsymbol{V}}_{t,j} - \tilde{\boldsymbol{V}}_{t,i}) \leq -n(\mu_j - \mu_i)\right).$$

根据马尔可夫不等式和罗森塔尔不等式（见引理 11 的证明），存在一个绝对常数 $C > 0$ 使得

$$\mathbb{P}(\boldsymbol{R}_{n,j} - \boldsymbol{R}_{n,i} \leq 0) \leq \frac{\left[Cq\left(\left\|\tilde{\boldsymbol{V}}_{1,j}\right\|_q + \left\|\tilde{\boldsymbol{V}}_{1,i}\right\|_q\right) n^{1/2}\right]^q}{\left[n(\mu_j - \mu_i)\right]^q}.$$

那么证明就完成了。 □

回顾式（48）中随机游走的定义，以及 r_* 式（50）中的定义。此外，回顾 $\boldsymbol{\Sigma}_{[r_*]}$ 是˜的协方差矩阵$\boldsymbol{V}_{1,\,[r_*]}$，并且$h_{[r_*]}$在式（53）中定义。

引理 12。 假设(51) 对 $q = 3$成立。那么存在一个有限常数 $C > 0$，它仅取决于 $\boldsymbol{\Sigma}_{[r_*]}$的正特征值 r_* 以及{$\mathbb{E}[|\tilde{\boldsymbol{V}}_{1,j}|^3]$、 $j \in [r_*]$}，使得

$$\left|\mathbb{E}\left[\min_{j\in[r_*]} (\boldsymbol{R}_{n,j} - n\mu_j)\right] - (-h_{[r_*]})\sqrt{n}\right| \leq C.$$

证明。 考虑其谱分解

$$\boldsymbol{\Sigma}_{[r_*]} = U\Lambda U^\top,$$

其中 $\Lambda := \mathrm{diag}(\lambda_1, \lambda_2, \ldots\ldots, \lambda_{r_*})$ 是一个对角矩阵，$\lambda_1 \geq \lambda_2 \geq \ldots\ldots \geq \lambda_{r_*} \geq 0$，且 $U \in \mathbb{R}^{r_* \times r_*}$ 是一个正交矩阵，即 $UU^\top$ 等于单位矩阵 $\boldsymbol{I}_{r_* \times r_*}$。

记 $s = \mathrm{rank}(\boldsymbol{\Sigma}_{[r_*]})$ 为协方差矩阵 $\boldsymbol{\Sigma}_{[r_*]}$ 的秩。根据定义，$s \leq r_*, \lambda_1 \geq \cdots \geq \lambda_s > 0$ 和 $\lambda_{s+1} = \cdots = \lambda_{r_*} = 0$。令 $U_1 \in \mathbb{R}^{r_* \times s}$ 为 U 的前 s 列，$U_2 \in \mathbb{R}^{r_* \times (s - r_*)}$ 为 U 的最后 $s - r_*$ 列。

根据定义，

$$\mathrm{Cov}(U_2^\top \tilde{\boldsymbol{V}}_{1,[r_*]}) = \mathbb{E}\left[U_2^\top \tilde{\boldsymbol{V}}_{1,[r_*]} \left(U_2^\top \tilde{\boldsymbol{V}}_{1,[r_*]}\right)^\top\right] = U_2^\top \boldsymbol{\Sigma}_{[r_*]} U_2 = \boldsymbol{0}_{(r_* - s) \times (r_* - s)},$$

这意味着

$$\mathbb{P}\left(\tilde{\boldsymbol{V}}_{1,[r_*]} \in \mathrm{column}(U_1)\right) = 1, \tag{66}$$

其中，$\mathrm{column}(U_1)$ 是 U_1 的列空间。

接下来，令 $\Lambda_1 = \mathrm{diag}(\lambda_1, \ldots, \lambda_s) \in \mathbb{R}^{s \times s}$ 表示所有正特征值构成的对角矩阵，我们定义

$$\mathcal{B}_0 := {\Lambda_1}^{-1/2} U_1^\top \in \mathbb{R}^{s \times r_*}, \quad \mathcal{B}_1 := U_1 {\Lambda_1}^{1/2} \in \mathbb{R}^{r_* \times s}.$$

注意，通过初等计算，

$$\mathcal{B}_0 \mathcal{B}_1 = \boldsymbol{I}_{s \times s}, \quad \mathcal{B}_0 \boldsymbol{\Sigma}_{[r_*]} \mathcal{B}_0^\top = \boldsymbol{I}_{s \times s}, \quad \mathcal{B}_1 \mathcal{B}_0 u = u \text{ for any } \ u \in \mathrm{column}(U_1). \tag{67}$$

此外，对于 $t \geq 1$，我们定义

$$\bar{\boldsymbol{V}}_t := \mathcal{B}_0(-\tilde{\boldsymbol{V}}_{t,[r_*]}) \in \mathbb{R}^s, \quad \bar{\boldsymbol{R}}_n := \sum_{t=1}^n \bar{\boldsymbol{V}}_t / \sqrt{n} \in \mathbb{R}^s, \quad \boldsymbol{Y} := \mathcal{B}_1 \boldsymbol{Z} \in \mathbb{R}^{r_*}, \text{ where } \boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{0}_s, \boldsymbol{I}_{s \times s}).$$

那么，根据定义和（67），对于 $j \in [r_*]$,

$$\mathbb{E}\bar{\boldsymbol{V}}_1 = \boldsymbol{0}_s, \quad \mathrm{Cov}(\bar{\boldsymbol{V}}_1) = \boldsymbol{I}_{s \times s}, \quad \bar{\boldsymbol{R}}_{n,j} = \mathcal{B}_0\left[-(\boldsymbol{R}_{n,j} - n\mu_j)/\sqrt{n}\right], \quad \boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{0}_d, \boldsymbol{\Sigma}_{[r_*]}). \tag{68}$$

那么，由于（66）和（67），对于每个 $j \in [r_*]$,

$$(\boldsymbol{R}_{n,j} - n\mu_j)/\sqrt{n} = \mathcal{B}_1 \mathcal{B}_0 (\boldsymbol{R}_{n,j} - n\mu_j)/\sqrt{n} = -\mathcal{B}_1 \bar{\boldsymbol{R}}_{n,j}.$$

我们定义函数 $g(u) := \max_{j \in [r_*]} (\mathrm{B}_1 u)_j$，对任意 $u \in \mathbb{R}^s$ 均成立。于是

$$\min_{j \in [r_*]} \left((\boldsymbol{R}_{n,j} - n\mu_j)/\sqrt{n}\right) = -g(\bar{R}_n).$$

此外，鉴于 $h_{[r_*]}$ 在（53）中定义，

$$h_{[r_*]} = \mathbb{E}[\max\{Y_1, \ldots, Y_{r_*}\}] = \mathbb{E}[\max\{(\mathcal{B}_1 \boldsymbol{Z})_1, \ldots, (\mathcal{B}_1 \boldsymbol{Z})_{r_*}\}] = \mathbb{E}[g(\boldsymbol{Z})],$$

这意味着

$$\left|\mathbb{E}\left[\min_{j \in [r_*]} \left((\boldsymbol{R}_{n,j} - n\mu_j)/\sqrt{n}\right)\right] - (-h_{[r_*]})\right| = \left|\mathbb{E}[g(\bar{\boldsymbol{R}}_n)] - \mathbb{E}[g(\boldsymbol{Z})]\right|.$$

最后，我们应用正态近似来界定上述项。记 $\|u\|$ $u \in \mathbb{R}^s$ 的欧几里得范数为 $|$。由于对于任意 u，$v \in \mathbb{R}^s$，$|\max_{j \in [r_*]} u_j - \max_{j \in [r_*]} v_j| \leq \|u - v\|_{\max}$，因此我们有

$$\frac{|g(u) - g(v)|}{\|u - v\|} \leq \frac{\|\mathcal{B}_1 u - \mathcal{B}_1 v\|}{\|u - v\|} \leq \lambda_{\max}(\mathcal{B}_1 \mathcal{B}_1') = \lambda_{\max}(\boldsymbol{\Sigma}) := M_1(g) \text{ for any } u, v \in \mathbb{R}^s.$$

现在，由于 [28]中的 (3.5) 和 (68),

$$
\begin{aligned}
&|\mathbb{E}g(\bar{\boldsymbol{R}}_n) - \mathbb{E}g(\boldsymbol{Z})| \le M_1(g) \sum_{t=1}^{n} \mathbb{E}\left(\|\bar{\boldsymbol{V}}_t/\sqrt{n}\|^2 \min\{4.5, (11.1 + 0.83\log(s))\|\bar{\boldsymbol{V}}_t/\sqrt{n}\|\} \right) \\
&\le \log(r_*)\lambda_{\max}(\boldsymbol{\Sigma})\mathbb{E}(\|\bar{\boldsymbol{V}}_1\|^3)n^{-1/2} \\
&\le \log(r_*)\lambda_{\max}(\boldsymbol{\Sigma}) \left(\lambda_{\max}({\Lambda_1}^{-1})\right)^{3/2} \mathbb{E}(\|\tilde{\boldsymbol{V}}_{1,[r_*]}\|^3)n^{-1/2},
\end{aligned}
$$

其中，最后一个不等式由 $\|\bar{\boldsymbol{V}}_1\|^2 \le \lambda_{\max}({\Lambda_1}^{-1})\|\tilde{\boldsymbol{V}}_{1,\,[r_*]}\|^2$ 得出。至此，证明完成 . □

C.3. [11] 中定理 3.3 的有效性

我们注意到， [11] 中定理 3.3 的证明似乎有误。该文在 (A.14) 和 (A.15) 中定义了

$$
\xi(n) = \sqrt{n}(\zeta(n) - h_{r,i}), \quad \text{with} \quad \zeta(n) := \frac{1}{\sqrt{n}} \max_{1\le k\le r} [\lambda_{k,i} + Z_{\langle M-r-1+k\rangle}(n) - n\mu_{i[M-1]}],
$$

其中， i 是一个固定假设，$\{\lambda_{k,\,i},\ 1 \le k \le r\}$ 是常数，$\{Z_{\langle M-r-1+k\rangle}(n) :\ 1 \le k \le r\}$ 是一个 r 维随机游走，且对于 $1 \le k \le r$， $\mathbb{E}[Z_{\langle M-r-1+k\rangle}(1)] = \mu_{i[M-1]}$，并且

$$
h_{r,i} := \mathbb{E}\left[\max\{Y_1, \ldots, Y_r\}\right], \quad \text{where} \quad (Y_1, \ldots, Y_r)^\top \sim \mathcal{N}(0, \Sigma),
$$

并且， Σ 是随机向量 $\{Z_{\langle M-r-1+k\rangle}(1),\ 1 \le k \le r\}$ 的协方差矩阵。

他们基于 [39], 中的定理 3 使用了非线性更新理论，其中 [11] 中的条件 (A.19) 必须成立，该条件陈述如下:

$$
\{\max_{1\le i\le n} |\xi(n+i)| : n \ge 1\} \text{ is uniform integrable.} \tag{69}
$$

然而，该条件并不成立。具体而言，根据中心极限定理和连续映射定理，当 $n \to \infty$ 时，

$$
\zeta_n - h_{r,i} \ \text{ converges in distribution to } \ \max\{Y_1, \ldots, Y_r\} - h_{r,i}.
$$

由于该极限是非退化的， $\xi(n) = \sqrt{n}(\zeta(n) - h_{r,\,i})$ 不是一致可积的。

在本文中，我们不通过验证（69）来应用 [39]中的定理 3，而是调用 [27]中的定理 4.2，并验证下式的一致可积性

$$
\left\{ \max_{0\le i\le Mn^\alpha} \left(\frac{|\xi(n+i) - \xi(n)|}{\rho(n)} \right) : n \ge 1 \right\},
$$

其中 $\rho(n) = n^{\alpha/2}$ 适用于恰当的 $\alpha \in (1/2, 1]$。详情请参见前两个小节。

附录 D：其他类别的序贯程序与讨论

D.1. 已知信号数量

在本小节中，我们考虑类别4， $\Delta_m^{\mathrm{kns}}(\alpha)$，其中已知信号数量为 m，并且经典误分类率受到控制。我们首先回顾间隙规则，记作$\delta_G(b) = (T_G(b),\ D_G(b))$，由 [31],提出，其中 $b>0$是一个阈值。

回顾在$\{\lambda_t^k\colon\ k\in[K]$、 $t=1$、2，.....$\}$中定义的24对数似然比。对于每个时间 $t\geq 1$，记

$$\bar{\lambda}_t^{(1)} \leq \cdots \leq \bar{\lambda}_t^{(K)} \tag{70}$$

对数似然比的次序统计量 $\{\lambda_t^k\colon\ k\in[K]\}$（不取绝对值）。此外，记相应的数据流索引为 $\{\ \bar{j}_1(t)$、...、$\bar{j}_K(t)\}$，使得 $\lambda^{\bar{j}_i}\left(\ {}_t\ {}^t = \bar{\lambda}_t^{(i)}\right)$ 对 $i\in[K]$成立。于是，对于 $b>0$，间隙规则 $\delta_G(b)$ 定义如下：

$$T_G(b) := \min\{t\geq 1 : \bar{\lambda}_t^{(K-m+1)} - \bar{\lambda}_t^{(K-m)} \geq b\}, \qquad D_G(b) := \{\bar{j}_{K-m+1}(t), \cdots, \bar{j}_K(t)\}. \tag{71}$$

也就是说，当第（$K-m+1$）个与第（$K-m$）个次序统计量之间的间隙大于阈值 b 时， $\delta_G(b)$ 停止；停止时，拒绝具有前 m 大对数似然比的数据流所对应的原假设。

定理 3.1 [31] 指出，如果我们选择

$$b_\alpha = |\log\alpha| + \log[m(K-m)], \tag{72}$$

那么 $\delta_G(b_\alpha)\in\Delta_m^{\mathrm{kns}}(\alpha)$。此外，根据定理 5.3 [31],，采用这一阈值时，族 $\{\delta_G(b_\alpha)\}$ 在 $\alpha\to 0$ 下是一阶渐近最优的。下一个定理将这一最优性加强到二阶。

定理 11. 假设 *1* 成立。那么，间隙规则 $\{\delta_G(b_\alpha)\}$ 在 $\{\Delta_m^{\mathrm{kns}}(\alpha)\}$ 下是二阶渐近最优的，当 $\alpha\to 0$时；也就是说，对每个 $A\subset[K]$,

$$\limsup_{\alpha\to 0}\left(\mathbb{E}_A[T_G(b_\alpha)] - T_A^{\min}\left(\Delta_m^{\mathrm{kns}}(\alpha)\right)\right) < \infty.$$

证明。 证明见附录E.5。 □

注 15。 在上述定理的证明中，我们应用定理1 并取

$$\mathcal{W}(D\mid A) = \begin{cases} 1 & \text{if } D\neq A \\ 0 & \text{else} \end{cases}, \quad \text{for } A,D\subset[K] \text{ with } |A|=|D|=m. \tag{73}$$

接下来，我们考虑每个 $A\subset[K]$ 在 $\alpha\to 0$时可实现的最小 ESS， $T_A^{\min}\left(\Delta_m^{\mathrm{kns}}(\alpha)\right)$ 的渐近近似。 [31] 中的定理 5.3 表明，对于每个 $A\subset[K]$，一阶渐近性能满足

$$T_A^{\min}\left(\Delta_m^{\mathrm{kns}}(\alpha)\right) = \frac{|\log(\alpha)|}{\mathrm{KL}^{\mathcal{W}}_{A,A}}(1+o(1)),$$

其中 KL散度$_{A}^{\mathcal{W}}$，$_{A}$ 定义见式（18），其中 $\mathcal{W}$ 见式（73），即

$$\mathrm{KL}_{A,A}^{\mathcal{W}} = \min\{\mathrm{KL}(f_A|f_C) : C \in \mathcal{A} \text{ and } C \neq A\}.$$

在本文中，我们给出了二阶精确近似。具体而言，由于（73）中的 W 是零一损失，根据备注 5，按照定义 4，每个 $A \in \mathrm{A}$ 关于 W 都有一个唯一最有利子集，其中 $D_A^{\mathcal{W}} = A$ 和 $\mathrm{KL}_{A}^{\mathcal{W}}{}_{,*} = \mathrm{KL}_{A}^{\mathcal{W}}{}_{,A}$。

定理 12. 假设 *1* 中的假设成立。令 $\mathcal{W}$ 由 (73)定义，并令 $A \in \mathcal{A}$ 为任意信号子集。

(a) 考虑非对称情形 $r_A^{\mathcal{W}} = 1$。假设 (22) 对 $q = 3$成立。那么，当 $\alpha \to 0$时，

$$T_A^{\min}\left(\Delta_m^{\mathrm{kns}}(\alpha)\right) = \frac{|\log\alpha|}{\mathrm{KL}_{A,A}^{\mathcal{W}}} + O(1).$$

(b) 考虑对称情形 $r_A^{\mathcal{W}} \geq 2$。令 $\epsilon \in (0,\ 1/2)$，并假设 (22) 对某个 q成立，且该 (23)满足条件。那么，当 $\alpha \to 0$时，

$$T_A^{\min}\left(\Delta_m^{\mathrm{kns}}(\alpha)\right) = \frac{|\log\alpha|}{\mathrm{KL}_{A,A}^{\mathcal{W}}} + \frac{h_{A,A}^{\mathcal{W}}\sqrt{|\log\alpha|}}{(\mathrm{KL}_{A,A}^{\mathcal{W}})^{3/2}} + O\left((\log\alpha)^{1/4+\epsilon/2}\right).$$

证明。 证明与定理 4 的证明类似，因此略去。 □

D.2. 序贯程序的其他子类

本小节回顾序贯程序的其他子类。对于这些子类，我们可以应用定理 1 和定理 2，将文献中已有的一阶结果加强到二阶。由于论证与第 1–4 类的论证类似，因此省略细节。

首先，假设信号数量介于l 和 u之间，并控制经典族系错误率 [15,31]。

第 5 类。 设l, u为整数，且满足 $0 \leq l < u \leq K$。对于容忍水平 $\alpha, \beta \in (0,1)$，我们将此设定下控制家族错误率的一类程序定义如下：

$$\mathcal{A} = \{A \subset [K] : \ell \leq |A| \leq u\},$$

$$\Delta_{\ell,u}^{\mathrm{bns}}(\alpha,\beta) = \left\{(T,D) \in \Delta : \max_{A\in\mathcal{A}} \mathbb{P}_A\left(|D \setminus A| \geq 1\right) \leq \alpha,\ \max_{A\in\mathcal{A}} \mathbb{P}_A\left(|A \setminus D| \geq 1\right) \leq \beta\right\}.$$

注 16。 当 $l = 0$ 且 $u = K$时，类 $\Delta_{\ell\ u,}^{\mathrm{bns}}(\alpha,\beta)$ 与 $\Delta_{1\ 1,}^{\mathrm{gfr}}(\alpha,\beta)$一致；参见第 2 类。

在 [31], 中提出了一种称为间隙交集规则的程序，记为 $\delta_{GI}(a,\ b)$，其中$a,\ b > 0$为阈值。 [31] 中证明，只要适当选择 $a,\ b,\ \delta_{GI}(a,\ b)$ 就属于$\Delta_{\ell\ u,}^{\mathrm{bns}}(\alpha,\ \beta)$。此外，当 $\alpha \vee \beta \to 0$ 时，族 $\{\delta_{GI}(a,\ b)\}$ 是一阶最优的，并且针对每个 $A \in \mathrm{A}$ 刻画了一阶最小 ESS。

现在，定义如下零一损失（见备注 5）：

$$\mathcal{W}(D \mid A) = \begin{cases} 1 & \text{if } D \neq A \\ 0 & \text{else} \end{cases}, \quad \text{for } A, D \subset [K] \text{ with } \ell \leq |A|, |D| \leq u,$$

应用定理 1和定理2 以及损失函数 $\mathcal{W}$，我们可以确立 $\{\delta_{GI}(a,b)\}$ 的二阶最优性，并刻画 $T_A^{\min}(\Delta^{\text{bns}}_{\ell\, u,}(\alpha,\beta))$ 对每个 $A \in \mathcal{A}$，当 $\alpha \vee \beta \to 0$ 满足（33）时。

其次，我们控制错误发现率（FDR）和错误非发现率（FNR），并假设信号数量已知为 m，或介于l 和 u [15]之间。回顾第$\Delta^{\text{fdr}}(\alpha,\beta)$ 类中3的定义。

第6类。 给定容忍水平 $\alpha, \beta \in (0,1)$ 和一个整数 $m \in [1, K)$，定义

$$\mathcal{A} := \{A \subset [K] : |A| = m\}, \quad \text{and} \quad \Delta^{\text{fdr}}_m(\alpha,\beta) := \Delta^{\text{fdr}}(\alpha,\beta).$$

第7类。 给定容忍水平 $\alpha, \beta \in (0,1)$ 和整数 $0 \leq l < u \leq K$，定义

$$\mathcal{A} = \{A \subset [K] : \ell \leq |A| \leq u\}, \quad \text{and} \quad \Delta^{\text{fdr}}_{\ell,u}(\alpha,\beta) = \Delta^{\text{fdr}}(\alpha,\beta).$$

在 [15], 中证明了，经过适当设计阈值后，（71）中的间隙规则（分别地，上文提到的间隙交集规则）对于类 $\Delta^{\text{fdr}}_m(\alpha,\beta)$（分别地，$\Delta^{\text{fdr}}_{\ell\,,u}(\alpha,\beta)$）是一阶渐近最优的。此外，最小期望样本量的一阶刻画也已给出。通过与小节 4.3中类似的论证，我们可以应用定理 1 确立相应的二阶最优性，并应用定理 2 得到最小期望样本量的二阶精确逼近。

最后，我们指出， [15] 也考虑在已知信号数量先验知识的情况下控制边际错误发现率和边际错误非发现率。感兴趣的读者可参阅该文，了解相关定义和一阶最优程序；并请注意，我们的结果同样适用于相应的序贯程序类别。

附录E：特定误差度量与信息结构下的证明

E.1. 关于Lorden规则的结果

在本小节中，令 $\mathcal{W} : \mathcal{A} \times \mathcal{A} \to [0,\infty)$ 为一般损失函数，并令 π_0 为 $\mathcal{A}$上的均匀分布。我们给出关于规则 $\delta_{\text{Ld}}(c,\mathcal{W}) = (T_{\text{Ld}}(c,\mathcal{W}), D_{\text{Ld}}(c,\mathcal{W}))$ 在（11）中的结果，后文将用到这些结果。

引理 13。 固定代价 $c > 0$。在事件 $\{T_{\text{Ld}} < \infty\}$ 上，对每个 $A_0 \in \mathcal{A}$,

$$\sum_{i \in D_0} \lambda^i_{T_{\text{Ld}}} - \sum_{j \in A_0} \lambda^j_{T_{\text{Ld}}} > \log\left[c^{-1}\pi_0(A_0)\mathcal{W}(D_{\text{Ld}} \mid A_0)\right],$$

其中 $D_0 := \{k \in [K] : \lambda^k_{T_{\text{Ld}}} \geq 0\}$ 是停止时具有非负对数似然比的数据流索引集合，$\log(0) := -\infty$ 且 $(T_{\text{Ld}}, D_{\text{Ld}})$ 是 $(T_{\text{Ld}}(c,\mathcal{W}),\ D_{\text{Ld}}(c,\mathcal{W}))$ 的简写。

证明。 回顾对数似然比 $\{\lambda_t^k\colon\ k\in[K]\}$ 见（24）。我们关注事件 $\{T_{\mathrm{Ld}}(c,\mathcal{W})<\infty\}$。对每个数据流 $k\in[K]$ 和子集 $A\subset[K]$，我们定义

$$\Lambda_t^k:=\exp\{\lambda_t^k\}=\prod_{s=1}^{t}\frac{f_1^k(\mathbf{X}_s^k)}{f_0^k(\mathbf{X}_s^k)}.,\quad\text{and}\quad\Lambda_t^A:=\prod_{k\in A}\Lambda_t^k.$$

根据 D_0 的定义，我们有

$$\Lambda_{T_{\mathrm{Ld}}}^A\le\Lambda_{T_{\mathrm{Ld}}}^{D_0},\quad\text{for each } A\in\mathcal{A}. \tag{74}$$

根据 $\delta_{\mathrm{Ld}}(c,\mathcal{W})$在 (11) 中的定义，我们有 $\sum_{A\in\mathcal{A}}\pi_{T_{\mathrm{Ld}}}(A)\mathcal{W}(D_{\mathrm{Ld}}\mid A)<c$，根据 $\pi_t(A)$在 (10) 中的定义，这等价于

$$\sum_{A\in\mathcal{A}}f_{A,[T_{\mathrm{Ld}}]}(A)\mathcal{W}(D_{\mathrm{Ld}}\mid A)<c\sum_{B\in\mathcal{A}}f_{B,[T_{\mathrm{Ld}}]}.$$

此外，将两边同时除以 $\prod_{s=1}^{T_{\mathrm{Ld}}}\prod_{k\in[K]}f_0^k(\mathbf{X}_s^k)$，得到

$$\sum_{A\in\mathcal{A}}\Lambda_{T_{\mathrm{Ld}}}^A\mathcal{W}(D_{\mathrm{Ld}}\mid A)<c\sum_{B\in\mathcal{A}}\Lambda_{T_{\mathrm{Ld}}}^B.$$

任取一个 $A_0\in$A。一方面，

$$\sum_{A\in\mathcal{A}}\Lambda_{T_{\mathrm{Ld}}}^A\mathcal{W}(D_{\mathrm{Ld}}\mid A)\ge\Lambda_{T_{\mathrm{Ld}}}^{A_0}\mathcal{W}(D_{\mathrm{Ld}}\mid A_0).$$

另一方面，由于 (74)，我们有

$$\sum_{B\in\mathcal{A}}\Lambda_{T_{\mathrm{Ld}}}^B\le\sum_{B\in\mathcal{A}}\Lambda_{T_{\mathrm{Ld}}}^{D_0}=|\mathcal{A}|\Lambda_{T_{\mathrm{Ld}}}^{D_0}=(\pi_0(A))^{-1}\Lambda_{T_{\mathrm{Ld}}}^{D_0}.$$

因此，结合上述三个不等式，得到

$$\frac{\Lambda_{T_{\mathrm{Ld}}}^{D_0}}{\Lambda_{T_{\mathrm{Ld}}}^{A_0}}>c^{-1}\pi_0(A_0)\mathcal{W}(D_{\mathrm{Ld}}\mid A_0),$$

两边取对数即可完成证明。 □

回顾 $\bar{j}_k$（t）在附录 D.1 中的定义。

引理 14。 固定一个代价 $c>0$。假设 $\mathcal{A}=\{A\subset[K]\colon\ |A|=m\}$，其中 $m\in[1,K)$ 是一个整数。在事件 $\{T_{\mathrm{Ld}}<\infty\}$发生的情况下，对于每个 $A_0\in\mathcal{A}$，我们有

$$\sum_{i\in D_0}\lambda_{T_{\mathrm{Ld}}}^i-\sum_{j\in A_0}\lambda_{T_{\mathrm{Ld}}}^j>\log[c^{-1}\pi_0(A_0)\mathcal{W}(D_{\mathrm{Ld}}\mid A_0)],$$

其中 $D_0\colon=\{\bar{j}_{K-m+1}(T_{\mathrm{Ld}}),\ldots,\bar{j}_K(T_{\mathrm{Ld}})\}$ 是具有 m 个最大对数似然比的T_{Ld}处的数据流索引，$\log(0)\colon=-\infty$，而 $(T_{\mathrm{Ld}},D_{\mathrm{Ld}})$ 是 $(T_{\mathrm{Ld}}(c,\mathcal{W}),\ D_{\mathrm{Ld}}(c,\mathcal{W}))$ 的简写。

证明。根据 D_0 的定义，我们有

$$\Lambda_{T_{\mathrm{Ld}}}^{A} \le \Lambda_{T_{\mathrm{Ld}}}^{D_0}, \quad \text{for each } A \in \mathcal{A}.$$

其余论证与引理 13 的证明类似，因此略去。 □

E.2. 广义误分类率

本小节给出小节 4.1 中结果的证明。回顾停止规则 $T_S(b)$，见式 (26)；损失函数 $\mathcal{W}$，见式 (28)；以及相应的规则 $T_{\mathrm{Ld}}(c,\mathcal{W})$，见式 (11)。此外，$\pi_0$ 表示 $\mathcal{A} = 2^{[K]}$上的均匀分布。

引理 15。令代价 $c \in (0, 2^{-K})$ 且 $b = \log(2^{-K}c^{-1})$。则，

$$T_S(b) \le T_{\mathrm{Ld}}(c,\mathcal{W}).$$

证明。由于 c、b 和 W 均为固定值，我们省略它们在 $T_S(b)$ 和 $T_{\mathrm{Ld}}(c,\mathrm{W})$ 中的依赖关系。设 $T_{\mathrm{Ld}} < \infty$，否则结论显然成立。我们证明 T_S 的停止准则在时间 T_{Ld} 满足，即

$$\sum_{i=1}^{m_0} \tilde{\lambda}_{T_{\mathrm{Ld}}}^{(i)} \ge b = \log(2^{-K}c^{-1}), \tag{75}$$

其中我们回顾次序统计量 $\{\tilde{\lambda}_t^{(i)} : i \in [K]\}$ 在（25）。这立即意味着 $T_S \le T_{\mathrm{Ld}}$。

对于每个 $i \in [K]$，令 l_i 表示达到 $\tilde{\lambda}_t^{(i)}$ 的数据流的索引，即，

$$|\lambda_t^{\ell_i}| = \tilde{\lambda}_t^{(i)},$$

其中平局被任意打破。进一步，定义 $B_0 := \{l_1, \ldots, l_{m_0}\}$为具有最小对数似然比绝对值的 m_0 个数据流的集合。令 $D_0 := \{k \in [K] : \lambda_{T_{\mathrm{Ld}}}^{k} \ge 0\}$表示在时间 T_{Ld}具有非负对数似然比的数据流集合。

我们定义如下子集

$$A_0 := (B_0 \cap D_{\mathrm{Ld}}^{c}) \cup (B_0^{c} \cap D_0),$$

如图 3所示。我们断言

$$|A_0 \triangle D_{\mathrm{Ld}}| \ge m_0 \quad \text{and} \quad A_0 \cap B_0^{c} = D_0 \cap B_0^{c}. \tag{76}$$

上述第二部分直接由德摩根定律得出。对于第一部分，只需证明 $A_0 \triangle D_{\mathrm{Ld}} \supseteq B_0$。为此，固定任意的 $k \in B_0$。根据 A_0的定义，若 $k \in A_0$，则 $k \in D_{\mathrm{Ld}}^{c}$，即 $k \in A_0 \setminus D_{\mathrm{Ld}}$；若 $k \notin A_0$，则 $k \notin D_{\mathrm{Ld}}^{c}$，即 $k \in D_{\mathrm{Ld}} \setminus A_0$。因此，$k \in A_0 \triangle D_{\mathrm{Ld}}$。

接下来，根据引理13，我们有

$$\sum_{i \in D_0} \lambda_{T_{\mathrm{Ld}}}^{i} - \sum_{j \in A_0} \lambda_{T_{\mathrm{Ld}}}^{j} > \log[c^{-1}\pi_0(A_0)\mathcal{W}(D_{\mathrm{Ld}} \mid A_0)].$$

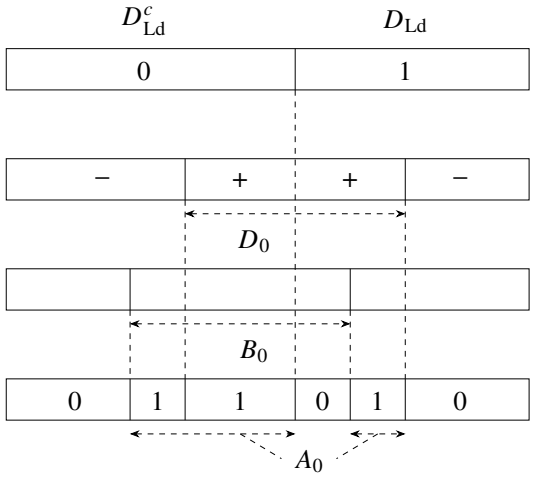


图 3：和交规则二阶最优性证明中涉及的子集的可视化。

由于（76）中的第二部分，我们有

$$\sum_{i\in D_0}\lambda^i_{T_{\mathrm{Ld}}}-\sum_{j\in A_0}\lambda^j_{T_{\mathrm{Ld}}}=\sum_{i\in D_0\cap B_0}\lambda^i_{T_{\mathrm{Ld}}}-\sum_{j\in A_0\cap B_0}\lambda^j_{T_{\mathrm{Ld}}}\le\sum_{k\in B_0}|\lambda^k_{T_{\mathrm{Ld}}}|=\sum_{i=1}^{m_0}\tilde{\lambda}^{(i)}_{T_{\mathrm{Ld}}}.$$

根据 (76) 的第一部分以及 ($\mathcal{W}$28) 中 W 的定义，我们有 $\mathrm{W}(D_{\mathrm{Ld}}\mid A_0)=1$。这证明了 (75)，并完成了证明，因为 $\pi_0(A_0)=2^{-K}$。 □

回顾 (16) 中 $\mathrm{KL}(f_A\mid f_C)$ 的定义、(17) 中 $\mathcal{H}\mathcal{W}_D$ 的定义，以及 (18) 中 $\mathrm{KL}^{\mathcal{W}}_{A,D}$ 的定义。

引理 1 的证明。 任取 $A\subset[K]$。令 $\tilde{C}\in\mathrm{H}^{\mathcal{W}}_A$（即 $|\tilde{C}\triangle A|\ge m_0$）满足 $\mathrm{KL}(f_A\mid f_{\tilde{C}})=\min_{D\neq A,\,C^*\in\mathcal{W}_{A,D}}\mathrm{KL}(f_A\mid f_{C^*})$。只需证明，对于任意 $|C^*\triangle D|\ge m_0$，存在一个子集 $\mathrm{H}\in\mathcal{W}$（即 $\triangle$），使得

$$\mathrm{KL}(f_A\mid f_{C^*})<\mathrm{KL}(f_A\mid f_{\tilde{C}}). \tag{77}$$

我们记作

$$L_1=A\setminus D,\qquad L_2=A\cap D,\qquad L_3=D\setminus A,\qquad L_4=D^c\setminus A,$$
$$\Gamma_1=L_1\setminus\tilde{C},\qquad \Gamma_2=L_2\setminus\tilde{C},\qquad \Gamma_3=L_3\cap\tilde{C},\qquad \Gamma_4=L_4\cap\tilde{C},$$

如图所示4。由此可得

$$A\setminus\tilde{C}=\Gamma_1\cup\Gamma_2,\qquad \tilde{C}\setminus A=\Gamma_3\cup\Gamma_4,\qquad \tilde{C}\triangle A=\Gamma_1\cup\Gamma_2\cup\Gamma_3\cup\Gamma_4,\qquad D\triangle A=L_1\cup L_3.$$

由于 $D\neq A$，且 $|C\triangle A|\ge m_0$，我们有

$$|L_1|+|L_3|\ge 1,\qquad |\Gamma_1|+|\Gamma_2|+|\Gamma_3|+|\Gamma_4|\ge m_0. \tag{78}$$

情形1. $\Gamma_1 \cup \Gamma_3 \neq \emptyset$。在这种情况下，我们定义

$$C^* = L_1 \cup (L_2 \setminus \Gamma_2) \cup \Gamma_4,$$

由此可得

$$A \setminus C^* = \Gamma_2, \qquad C^* \setminus A = \Gamma_4, \qquad C^* \triangle D = L_1 \cup \Gamma_2 \cup L_3 \cup \Gamma_4.$$

另一方面，由于 L_1, L_2, L_3, L_4 互不相交，因此我们有

$$|C^* \triangle D| = |L_1| + |\Gamma_2| + |L_3| + |\Gamma_4| \geq |\Gamma_1| + |\Gamma_2| + |\Gamma_3| + |\Gamma_4| \geq m_0,$$

其中，最后一个不等式由 (78) 得出。另一方面，根据 (16),

$$\begin{aligned} \mathrm{KL}(f_A \mid f_{\tilde{C}}) - \mathrm{KL}(f_A \mid f_{C^*}) &= \sum_{k \in A \setminus \tilde{C}} \mathcal{I}_k^1 + \sum_{k \in \tilde{C} \setminus A} \mathcal{I}_k^0 - \Big(\sum_{k \in A \setminus C^*} \mathcal{I}_k^1 + \sum_{k \in C^* \setminus A} \mathcal{I}_k^0 \Big) \\ &= \sum_{k \in \Gamma_1} \mathcal{I}_k^1 + \sum_{k \in \Gamma_3} \mathcal{I}_k^0 > 0, \end{aligned}$$

where 最后一个不等式由 $\Gamma_1 \cup \Gamma_3 \neq \emptyset$ 和假设 1 得出。因此，(77) 在情形1中得证 .

情形2。 $\Gamma_1 \cup \Gamma_3 = \emptyset$。在此情形下，由于 (78)，我们有

$$|\Gamma_2| + |\Gamma_4| \geq m_0. \tag{79}$$

如果 $|L_1| + |L_3| \geq |\Gamma_2| + |\Gamma_4|$，我们定义 $\tilde{\Gamma}_2 = \tilde{\Gamma}_4 = \emptyset$。否则，令 $\tilde{\Gamma}_2 \subset \Gamma_2$ 和 $\tilde{\Gamma}_4 \subset \Gamma_4$ 满足 $|\tilde{\Gamma}_2| + |\tilde{\Gamma}_4| = |\Gamma_2| + |\Gamma_4| - |L_1| - |L_3|$。根据（79）和（78），我们总有

$$\big(\Gamma_2 \setminus \tilde{\Gamma}_2\big) \cup \big(\Gamma_4 \setminus \tilde{\Gamma}_4\big) \neq \emptyset. \tag{80}$$

现在，我们定义

$$C^* = L_1 \cup (L_2 \setminus \tilde{\Gamma}_2) \cup \tilde{\Gamma}_4,$$

由此可得

$$A \setminus C^* = \tilde{\Gamma}_2, \qquad C^* \setminus A = \tilde{\Gamma}_4, \qquad C^* \triangle D = L_1 \cup \tilde{\Gamma}_2 \cup L_3 \cup \tilde{\Gamma}_4.$$

一方面，由于 L_1， L_2， L_3， L_4 互不相交，且根据（79），我们有

$$|C^* \triangle D| = |L_1| + |\tilde{\Gamma}_2| + |L_3| + |\tilde{\Gamma}_4| \geq |\Gamma_2| + |\Gamma_4| \geq m_0,$$

另一方面，根据（16），

$$\begin{aligned} \mathrm{KL}(f_A \mid f_{\tilde{C}}) - \mathrm{KL}(f_A \mid f_{C^*}) &= \sum_{k \in A \setminus \tilde{C}} \mathcal{I}_k^1 + \sum_{k \in \tilde{C} \setminus A} \mathcal{I}_k^0 - \Big(\sum_{k \in A \setminus C^*} \mathcal{I}_k^1 + \sum_{k \in C^* \setminus A} \mathcal{I}_k^0 \Big) \\ &= \sum_{k \in \Gamma_2 \setminus \tilde{\Gamma}_2} \mathcal{I}_k^1 + \sum_{k \in \Gamma_4 \setminus \tilde{\Gamma}_4} \mathcal{I}_k^0 > 0, \end{aligned}$$

其中，最后一个不等式由 (80) 和假设 1 得出。因此，情形2中的 (77) 得证，证明也就完成了。 □

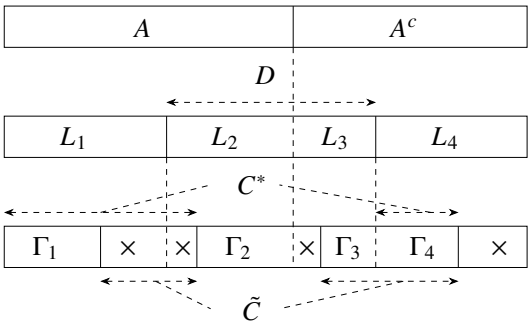


图 4：广义误分类率下 ESS 近似证明中出现的子集的可视化。

E.3. 广义族错误率

本小节给出小节 4.2 中各项结果的证明。回顾停止规则 $T_L(a,b)$（见 (31)）、损失函数 $\mathcal{W}$（见 (34)）以及相应规则 $T_{\mathrm{Ld}}(c,\mathcal{W})$（见 (11)）的定义。此外，$\pi_0$ 表示 $\mathcal{A}=2^{[K]}$ 上的均匀分布。

定理 5 的证明。 令 $\boldsymbol{\theta}:=(\alpha,\beta)$，并定义

$$
\begin{aligned}
c_{\boldsymbol{\theta}} &:= 2^{-K} e^{-(a_{\boldsymbol{\theta}} \vee b_{\boldsymbol{\theta}})} = 2^{-K} \min\left\{ 2^{-m_2} \binom{K}{m_2}^{-1} \beta, 2^{-m_1} \binom{K}{m_1}^{-1} \alpha \right\}, \\
L &:= (1+C) 2^K \max\left\{ 2^{m_2} \binom{K}{m_2}, 2^{m_1} \binom{K}{m_1} \right\},
\end{aligned}
\tag{81}
$$

回忆一下，$a_{\boldsymbol{\theta}}$，$b_{\boldsymbol{\theta}}$ 定义于式（32）中，常数 C 定义于式（33）中。于是，只需验证在式（1 2）–（12）–（14）中，条件在损失函数 $\mathcal{W}$ 定义于式（34））、代价 $c_{\boldsymbol{\theta}}$ 和常数 L成立，结论便可直接由定理 1 得出。

先考虑条件（12）。对每个 $D\subset[K]$，令 $A=D^c$。则 $D\setminus A=D$ 且 $A\setminus D=D^c$。由于$|D|+|D^c|=K\geq m_1+m_2$，可知 $|D\setminus A|\geq m_1$或 $|A\setminus D|\geq m_2$。根据式（$\mathcal{W}$中34）中 $\mathcal{W}(D\mid A)>0$。因此，条件（12）成立。

接下来，我们验证条件（13）。根据 $c_{\boldsymbol{\theta}}$ 在（81）中的定义，我们有

$$
\max\{a_{\boldsymbol{\theta}}, b_{\boldsymbol{\theta}}\} = \log\left(2^{-K} c_{\boldsymbol{\theta}}^{-1}\right).
$$

然后，根据引理 16，$T_L(a_{\boldsymbol{\theta}}, b_{\boldsymbol{\theta}}) \leq T_{\mathrm{Ld}}(c_{\boldsymbol{\theta}}, \mathcal{W})$，这验证了条件（13）。

最后，我们关注条件（14）。固定任意一个程序 $\delta=(T, D)\in\Delta^{\mathrm{gfr}}_{m_1,m_2}(\boldsymbol{\theta})$ 属于（2）。根据其定义以及（34）中 W 的定义，

$$
\begin{aligned}
\mathrm{ERR}^{\mathrm{int}}(\delta;\mathcal{W}) &\leq \sum_{A\in\mathcal{A}} \pi_0(A)\left(\mathbb{P}_A(|D\setminus A|\geq m_1)+\mathbb{P}_A(|A\setminus D|\geq m_2)\right) \\
&\leq (\alpha+\beta)\sum_{A\in\mathcal{A}}\pi_0(A) = (\alpha+\beta).
\end{aligned}
$$

注意，由于条件（33）

$$\alpha+\beta\le(1+C)\min\{\alpha,\beta\}\le Lc_{\boldsymbol{\theta}},$$

其中，最后一个不等式是根据 $c_{\boldsymbol{\theta}}$ 和 L 在（81）中的定义得到的。因此，条件（14）得到验证，证明完毕。 □

引理 16。设代价为 $c\in(0,2^{-K})$，并假设阈值为 $0<a,b\le\log(2^{-K}c^{-1})$。则

$$T_L(a,b)\le T_{\mathrm{Ld}}(c,\mathcal{W}).$$

证明。由于 $c,\ a,b$ 和 W 是固定的，我们在 $T_L(a,b)$ 和 $T_{\mathrm{Ld}}(c,\mathrm{W})$ 中省略它们的依赖关系。假设 $T_{\mathrm{Ld}}<\infty$，否则结论显然成立。我们证明 T_L 的停止准则在时间 T_{Ld} 满足，即以下条件之一成立：要么 (i) 对某个 $0\le i<m_1$,

$$\sum_{j=1}^{m_1-i}\hat{\lambda}^{(j)}_{T_{\mathrm{Ld}}}\ge\log(2^{-K}c^{-1}),\quad\text{and}\quad\sum_{j=i+1}^{i+m_2}\check{\lambda}^{(j)}_{T_{\mathrm{Ld}}}\ge\log(2^{-K}c^{-1}),\tag{82}$$

要么 (ii) 对某个 $0\le i<m_2$,

$$\sum_{j=i+1}^{i+m_1}\hat{\lambda}^{(j)}_{T_{\mathrm{Ld}}}\ge\log(2^{-K}c^{-1}),\quad\text{and}\quad\sum_{j=1}^{m_2-i}\check{\lambda}^{(j)}_{T_{\mathrm{Ld}}}\ge\log(2^{-K}c^{-1})$$

这里回顾第 $\{\hat{\lambda}^{(i)}_t\}$ 和 $\{\check{\lambda}^{(i)}_t\}$ 节中定义的次序统计量及其对应的数据流索引 $\{\hat{l}_i(t),\check{l}_i(t)\}$。这立即推出4.2。 $T_L\le T_{\mathrm{Ld}}$。

回顾对数似然比 $\{\lambda^k_t:\ k\in[K]\}$ 在（24）中，并令 $D_0:=\{k\in[K]:\ \lambda^k_{T_{\mathrm{Ld}}}\ge0\}$ 表示在时间 T_{Ld} 具有非负对数似然比的数据流集合。我们定义

$$\begin{aligned}&L_1:=\{k\in D^c_{\mathrm{Ld}}:\lambda^k_{T_{\mathrm{Ld}}}<0\}=D^c_{\mathrm{Ld}}\cap D^c_0,\quad L_2:=\{k\in D^c_{\mathrm{Ld}}:\lambda^k_{T_{\mathrm{Ld}}}\ge0\}=D^c_{\mathrm{Ld}}\cap D_0,\\&L_3:=\{k\in D_{\mathrm{Ld}}:\lambda^k_{T_{\mathrm{Ld}}}\ge0\}=D_{\mathrm{Ld}}\cap D_0,\quad L_4:=\{k\in D_{\mathrm{Ld}}:\lambda^k_{T_{\mathrm{Ld}}}<0\}=D_{\mathrm{Ld}}\cap D^c_0.\end{aligned}\tag{83}$$

我们着重讨论情形 $|L_2|\le|L_4|$；情形 $|L_2|>|L_4|$ 的证明类似，故略去。记

$$i^*:=\min\{|L_4|-|L_2|,\ m_1-1\}.$$

我们用 $B^+_0=\{\hat{l}_1(T_{\mathrm{Ld}}),\ \ldots\ldots,\ \hat{l}_{m_1-i^*}(T_{\mathrm{Ld}})\}$ 表示与正对数似然比的前m_1-i^*个次序统计量相对应的、在时间 T_{Ld} 的数据流索引集合，并用 $B^-_0=\{\check{l}_1(T_{\mathrm{Ld}}),\ \ldots\ldots,\ \check{l}_{m_2+i^*}(T_{\mathrm{Ld}})\}$ 表示与非正对数似然比绝对值的前 m_2+i^* 个次序统计量相对应的、在时间 T_{Ld}的集合。我们定义如下子集：

$$\begin{aligned}&A^+_0:=L_2\cup(L_3\setminus B^+_0),\quad\text{and}\quad A^-_0:=\tilde{B}^-\cup L_2\cup L_3,\quad\text{where}\\&\tilde{B}^-=\begin{cases}B^-_0\cap L_1 & \text{if }\ |B^-_0\cap L_1|\le m_2\\ \text{any subset of } B^-_0\cap L_1 \text{ with cardinality } m_2 & \text{otherwise}\end{cases},\end{aligned}\tag{84}$$

如图所示 5。根据定义，我们有

$$A^+_0\subset D_0\subset A^-_0,\quad D_0\setminus A^+_0=B^+_0\cap L_3,\quad\text{and}\quad A^-_0\setminus D_0=\tilde{B}^-.\tag{85}$$

此外，根据引理 17，

$$|D_{\mathrm{Ld}} \setminus A_0^+| \geq m_1, \quad \text{and} \quad |A_0^- \setminus D_{\mathrm{Ld}}| \geq m_2. \tag{86}$$

现在，由于引理 13，我们有

$$\sum_{i \in D_0} \lambda^i_{T_{\mathrm{Ld}}} - \sum_{j \in A_0} \lambda^j_{T_{\mathrm{Ld}}} > \log[c^{-1}\pi_0(A_0)\mathcal{W}(D_{\mathrm{Ld}} \mid A_0)], \text{ for } A_0 \in \{A_0^+, A_0^-\}.$$

一方面，根据 $\mathcal{W}$ 在（34）和（86）中的定义，可得

$$\mathcal{W}(D_{\mathrm{Ld}} \mid A_0^+) = 1, \quad \text{and} \quad \mathcal{W}(D_{\mathrm{Ld}} \mid A_0^-) = 1.$$

另一方面，由于（85），可得

$$\sum_{i \in D_0} \lambda^i_{T_{\mathrm{Ld}}} - \sum_{j \in A_0} \lambda^j_{T_{\mathrm{Ld}}} = \begin{cases} \sum\limits_{k \in B_0^+ \cap L_3} \lambda^k_{T_{\mathrm{Ld}}} & \text{for } A_0 = A_0^+ \\ \sum\limits_{k \in \tilde{B}^-} |\lambda^k_{T_{\mathrm{Ld}}}| & \text{for } A_0 = A_0^- \end{cases}.$$

因此，根据 B_0^+的定义 $\tilde{B}^-$，我们有

$$\sum_{j=1}^{m_1 - i^*} \hat{\lambda}^{(j)}_{T_{\mathrm{Ld}}} = \sum_{k \in B_0^+} \lambda^k_{T_{\mathrm{Ld}}} \geq \sum_{k \in B_0^+ \cap L_3} \lambda^k_{T_{\mathrm{Ld}}} > \log[c^{-1}\pi_0(A_0^+)],$$

$$\sum_{j=1+i^*}^{m_2 + i^*} |\check{\lambda}^{(j)}_{T_{\mathrm{Ld}}}| \geq \sum_{j \in \tilde{B}^-} |\check{\lambda}^{(j)}_{T_{\mathrm{Ld}}}| > \log[c^{-1}\pi_0(A_0^-)].$$

Si因此 π（A）$= 2^{-K}$对每个 $A \subset [K]$，这意味着（82）成立，其中 $i = i^*$。证明完毕 . □

引理17。 考虑 L_1、L_2、L_3、L_4、D_0在（83）中的子集，以及 A_0^+、A_0^-、B_0^+、B_0^-、$\tilde{B}^-$在（84）中的子集。我们假设$|L_2| \leq |L_4|$并记i^*：$= |L_4| - |L_2|$。那么

$$|D_{\mathrm{Ld}} \setminus A_0^+| \geq m_1, \quad \textit{and} \quad |A_0^- \setminus D_{\mathrm{Ld}}| \geq m_2.$$

证明。 由于 $B_0^+ \subset L_2 \cup L_3$、$B_0^- \subset L_1 \cup L_4$ 和 L_1，L_2，L_3，L_4 两两不相交，因此我们有 $B_0^+ \cap L_3 = B_0^+ \setminus L_2$ 和 $B_0^- \cap L_1 = B_0^- \setminus L_4$，这意味着

$$|B_0^+ \cap L_3| \geq |B_0^+| - |L_2| = m_1 - i^* - |L_2| = m_1 - |L_4|,$$

$$|B_0^- \cap L_1| \geq |B_0^-| - |L_4| = m_2 + i^* - |L_4| = m_2 - |L_2|.$$

此外，根据 $\tilde{B}^-$ 的定义，如果 $|B_0^- \cap L_1| \leq m_2$，则有 $\tilde{B}^- = B_0^- \cap L_1$；否则，有$|\tilde{B}^-| = m_2$。

接下来，由于 $D_{\mathrm{Ld}} = L_3 \cup L_4$, $A_0^+ = L_2 \cup (L_3 \setminus B_0^+)$ 和 $A_0^- = \tilde{B}^- \cup L_2 \cup L_3$，可得

$$D_{\mathrm{Ld}} \setminus A_0^+ = (B_0^+ \cap L_3) \cup L_4, \quad \text{and} \quad A_0^- \setminus D_{\mathrm{Ld}} = \tilde{B}^- \cup L_2.$$

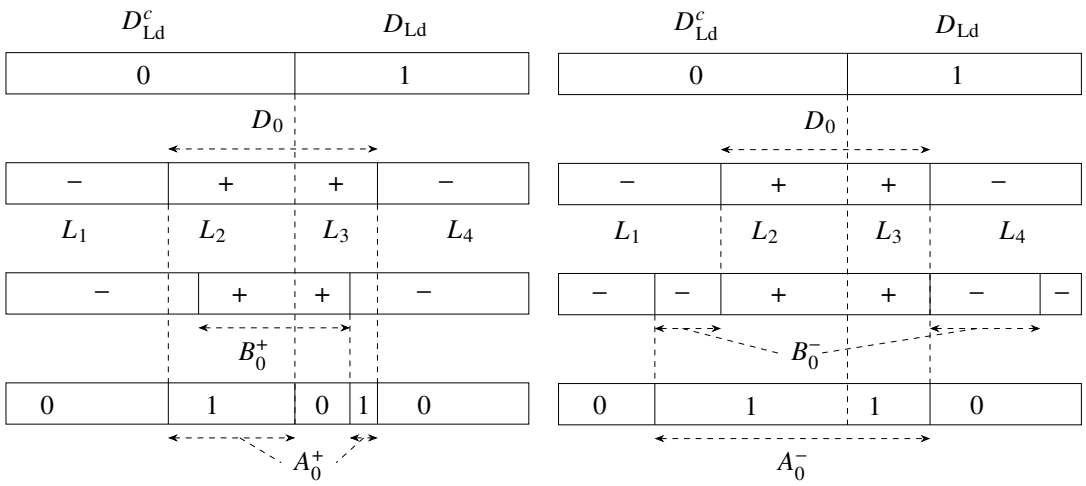


图5：Leap规则二阶最优性证明中出现的子集的可视化。

因此，我们有

$$|D_{\mathrm{Ld}} \setminus A_0^+| = |B_0^+ \cap L_3| + |L_4| \geq m_1, \quad \text{and} \quad |A_0^- \setminus D_{\mathrm{Ld}}| = |\tilde{B}^-| + |L_2| \geq m_2.$$

证明至此完成。 □

引理 2 的证明。对于第（a）部分，注意到当 $m_1 = m_2 = 1$时，损失函数$\mathcal{W}$ 在（34）中是备注 5所定义的零一损失。因此，结论可直接由备注 5得出。

接下来，我们关注第（b）部分，其中 $\mathrm{I}_0^k = \mathrm{I}_1^k = \mathrm{I}^*$ 对于 $k \in [K]$ 和 $m_1 = m_2 = m^*$。设 $A \subset [K]$ 满足 $\min\{|A|, |A^c|\} \geq m^*$。回顾（34）中 W 的定义，以及（18）中 $\mathrm{KL}^{\mathcal{W}}_{A,D}$的定义。显然，$\mathrm{KL}^{\mathcal{W}}_{A,A} = m^* \mathrm{I}^*$。因此，只需证明对于任意 $D \neq A$，存在一个子集 $C^* \in \mathrm{H}^{\mathcal{W}}_D$，使得

$$\mathrm{KL}(f_A \mid f_{C^*}) < m^* \mathcal{I}^*.$$

我们记

$$L_1 = A \setminus D,\ L_2 = A \cap D,\ L_3 = D \setminus A,\ L_4 = D^c \setminus A.$$

由于 $\min\{|A|, |A^c|\} \geq m^*$，我们有 $\sum_{i=1}^4 |L_i| = |A| + |A^c| \geq 2m^*$，这意味着 $|L_2| + |L_3| \geq m^*$ 或 $|L_1| + |L_4| \geq m^*$。我们给出情形 $|L_2| + |L_3| \geq m^*$ 的证明；另一种情形的证明类似。**情形1.** $L_3 \neq \emptyset$。在这种情形下，我们定义

$$C^* = L_1 \cup (L_2 \setminus \Gamma_2), \text{ with } \Gamma_2 \subset L_2 \text{ such that } |\Gamma_2| = \max\{m^* - |L_3|, 0\},$$

这里我们注意到，这样的 Γ_2 总是存在，因为 $|L_2| \geq m^* - |L_3|$。注意，

$$D \setminus C^* = \Gamma_2 \cup L_3 \quad A \setminus C^* = \Gamma_2, \quad \text{and} \quad C^* \setminus A = \emptyset.$$

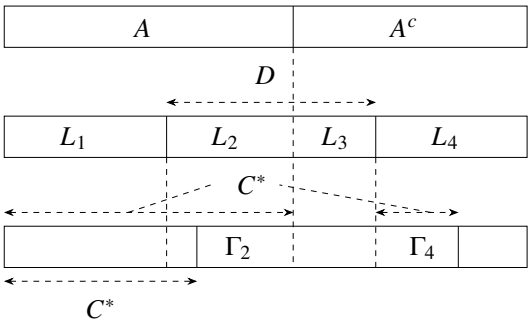


图 6：广义族错误率下 ESS 近似证明中出现的子集的可视化。

一方面，这意味着 $|D\backslash C^*| = |\Gamma_2| + |L_3| \geq m^*$，从而 $C^* \in \mathrm{H}_{\mathcal{W}_D}$。另一方面，由于在这种情况下 $|\Gamma_2| < m^*$，

$$\mathrm{KL}(f_A \mid f_{C^*}) = |\Gamma_2| \mathcal{I}^* < m^* \mathcal{I}^*.$$

情形1下的证明已完成。

情形2。 $L_3 = \emptyset$。在这种情况下，由于 $D \neq A$，我们必有 $L_1 \neq \emptyset$。此外，我们有 $|L_2| \geq m^*$ 且 $|L_4| = |A^c| \geq m^*$。我们定义

$$C^* = L_1 \cup L_2 \cup \Gamma_4, \text{ with } \Gamma_4 \subset L_4 \text{ such that } |\Gamma_4| = \max\{m^* - |L_1|, 0\},$$

其中我们注意到，由于 $|L_4| \geq m^*$，这样的 Γ_4 总是存在。注意

$$C^* \setminus D = L_1 \cup \Gamma_4, \quad C^* \setminus A = \Gamma_4, \quad \text{and} \quad A \setminus C^* = \emptyset$$

一方面，这意味着 $|C^* \setminus D| = |L_1| + |\Gamma_4| \geq m^*$，从而 $C^* \in \mathrm{H}_D^{\mathcal{W}}$。另一方面，由于$|L_1| > 0$，我们有 $|\Gamma_4| < m^*$，从而

$$\mathrm{KL}(f_A \mid f_{C^*}) = |\Gamma_4| \mathcal{I}^* < m^* \mathcal{I}^*.$$

情形2下的证明已经完成，由此也完成了第（b）部分的证明。 □

E.4. 错误发现率与非发现率

定理的证明 7。 如 [15], 中第2.2小节所示，对于任意决策规则 D，以下不等式成立：对于每个 $A \subset [K]$,

$$\frac{1}{K}\mathbb{P}_A(|D \setminus A| \geq 1) \leq \mathbb{E}_A\left[\frac{|D \setminus A|}{|D| \vee 1}\right] \leq \mathbb{P}_A(|D \setminus A| \geq 1),$$

$$\frac{1}{K}\mathbb{P}_A(|A \setminus D| \geq 1) \leq \mathbb{E}_A\left[\frac{|A \setminus D|}{|D| \vee 1}\right] \leq \mathbb{P}_A(|A \setminus D| \geq 1),$$

这意味着类别3 与类别2 之间存在以下关系 $m_1 = m_2 = 1$:

$$\Delta_{1,1}^{\mathrm{gfr}}(\alpha, \beta) \subset \Delta^{\mathrm{fdr}}(\alpha, \beta) \subset \Delta_{1,1}^{\mathrm{gfr}}(K\alpha, K\beta).$$

然后，根据定义可知，对于 α， $\beta \in (0,\ 1)$ 以及每个 $A \subset [K]$,

$$T_A^{\min}\left(\Delta_{1,1}^{\mathrm{gfr}}(\alpha,\beta)\right) \geq T_A^{\min}\left(\Delta^{\mathrm{fdr}}(\alpha,\beta)\right) \geq T_A^{\min}\left(\Delta_{1,1}^{\mathrm{gfr}}(K\alpha,K\beta)\right). \tag{87}$$

此外，当 $m_1 = m_2 = 1$ 时，式 (35) 中的交集规则 $\delta_I(a,b)$ 与式 (31) 中的 Leap规则$\delta_L(a,b)$ 一致，并且式 (36) 中的阈值 $(a_{\alpha,\beta}, b_{\alpha,\beta})$ 与式 (32) 中的 $(a_{\alpha,\beta}, b_{\alpha,\beta})$ 相同。由定理 5 可知，存在一个常数 $C' > 0$，它不依赖于 α,β，使得对于每个 $A \subset [K]$，以及满足式 (33) 的$\alpha,\beta \in (0,1)$，我们有

$$\mathbb{E}_A[T_I(a_{K\alpha,K\beta}, b_{K\alpha,K\beta})] - T_A^{\min}\left(\Delta_{1,1}^{\mathrm{gfr}}(K\alpha,K\beta)\right) \leq C'.$$

于是，由于（87），我们立即得到

$$\begin{aligned}
&\mathbb{E}_A[T_I(a_{\alpha,\beta}, b_{\alpha,\beta})] - T_A^{\min}\left(\Delta^{\mathrm{fdr}}(\alpha,\beta)\right) \leq \mathbb{E}_A[T_I(a_{\alpha,\beta}, b_{\alpha,\beta})] - T_A^{\min}\left(\Delta_{1,1}^{\mathrm{gfr}}(K\alpha,K\beta)\right) \\
&\leq \mathbb{E}_A[T_I(a_{\alpha,\beta}, b_{\alpha,\beta})] - \mathbb{E}_A[T_I(a_{K\alpha,K\beta}, b_{K\alpha,K\beta})] + C'
\end{aligned}$$

根据引理 18，证明至此完成。 □

记 $\boldsymbol{\theta}_{\alpha,\ \beta} = (a_{\alpha,\ \beta},\ b_{\alpha,\ \beta})$。考虑式（35）中定义的交集规则。

引理 18. 假设假设条件 *1* 成立，且 α， $\beta \in (0,\ 1)$ 满足 (33)。那么，存在一个常数$L' > 0$，它不依赖于 α， β，使得对于每个 $A \subset [K]$ 以及任意 α， $\beta \in (0,\ 1)$

$$\mathbb{E}_A[T_I(\boldsymbol{\theta}_{\alpha,\beta})] - \mathbb{E}_A[T_I(\boldsymbol{\theta}_{K\alpha,K\beta})] \leq L'.$$

证明。 不失一般性，假设 $A = [K]$，且 $\alpha,\beta \leq (K)^{-1}$。那么

$$\boldsymbol{\theta}_{K\alpha,K\beta} = (|\log(\beta)|, |\log(\alpha)|), \quad \text{and} \quad \boldsymbol{\theta}_{\alpha,\beta} = (|\log(\beta)| + \log(K), |\log(\alpha)| + \log(K)). \tag{88}$$

对于 $a,b > 0$，定义

$$\tilde{T}_I(a,b) := \inf\left\{t \geq 1 : \lambda_t^k \geq b,\ \text{for } k \in [K]\right\}.$$

根据式（35）中交集规则的定义，

$$T_I(\boldsymbol{\theta}_{\alpha,\beta}) \leq \tilde{T}_I(\boldsymbol{\theta}_{\alpha,\beta}), \quad T_I(\boldsymbol{\theta}_{K\alpha,K\beta}) \leq \tilde{T}_I(\boldsymbol{\theta}_{K\alpha,K\beta}).$$

此外，按照定理 2 中相同的论证，当 $\alpha \vee \beta \to 0$ 满足式（33）时，有

$$\mathbb{E}_{[K]}\left[T_I(\boldsymbol{\theta}_{K\alpha,K\beta})\right] = \mathbb{E}_{[K]}\left[\tilde{T}_I(\boldsymbol{\theta}_{K\alpha,K\beta})\right] + O(1),$$

这意味着

$$\mathbb{E}_{[K]}[T_I(\boldsymbol{\theta}_{\alpha,\beta})] - \mathbb{E}_{[K]}[T_I(\boldsymbol{\theta}_{K\alpha,K\beta})] \leq \mathbb{E}_{[K]}\left[\tilde{T}_I(\boldsymbol{\theta}_{\alpha,\beta})\right] - \mathbb{E}_{[K]}\left[\tilde{T}_I(\boldsymbol{\theta}_{K\alpha,K\beta})\right] + O(1).$$

最后，我们证明当 $\alpha \vee \beta \to 0, \mathbb{E}_{[K]}\left[\tilde{T}_I(\boldsymbol{\theta}_{\alpha,\beta})\right] - \mathbb{E}_{[K]}\left[\tilde{T}_I(\boldsymbol{\theta}_{K\alpha,K\beta})\right] = O(1)$。我们定义

e

$$\tau^* := \tilde{T}_I(\log(K), \log(K)) = \inf\left\{t \geq 1 : \lambda_t^k \geq \log(K),\ \text{for } k \in [K]\right\}.$$

显然， $\tilde{L} := \mathbb{E}_{[K]}[\tau^*] < \infty$，这是一个常数。此外，根据定义，

$$\tilde{T}_I(\boldsymbol{\theta}_{\alpha,\beta}) \le \tau^* + \widehat{T}_I(\boldsymbol{\theta}_{K\alpha,K\beta}),$$

根据88，我们定义

$$\widehat{T}_I(\boldsymbol{\theta}_{K\alpha,K\beta}) := \inf\left\{t \ge \tau^* : \lambda_t^k - \lambda_{\tau^*}^k \ge |\log(\alpha)| \text{ for } k \in [K]\right\} - \tau^*.$$

由于 $\{\mathbf{X}_t^{[K]} : t \ge 1\}$ 是独立同分布的，因此在 $\mathbb{P}_{[K]}$ 下， $\widehat{T}_I(\boldsymbol{\theta}_{K\alpha,K\beta})$ 与 $\tilde{T}_I(\boldsymbol{\theta}_{K\alpha,K\beta})$ 具有相同的分布。因此，

$$\mathbb{E}_{[K]}\left[\tilde{T}_I(\boldsymbol{\theta}_{\alpha,\beta})\right] \le \mathbb{E}_{[K]}[\tau^*] + \mathbb{E}_{[K]}\left[\widehat{T}_I(\boldsymbol{\theta}_{K\alpha,K\beta})\right] \le \tilde{L} + \mathbb{E}_{[K]}\left[\tilde{T}_I(\boldsymbol{\theta}_{K\alpha,K\beta})\right],$$

由此完成证明。 □

E.5. 已知信号数量

在本小节中，我们给出附录D.1中各项结果的证明。我们回顾停止规则 $T_G(b)$ 在（71）中的定义、损失函数 $\mathcal{W}$ 在（73）中的定义，以及相应的规则 $T_{\mathrm{Ld}}(c, \mathcal{W})$ 在（11）中的定义。此外， π_0 表示定义在 $\mathcal{A} = \{A \subset [K] : |A| = m\}$ 上的均匀分布。

定理11的证明。我们定义

$$c_\alpha := \binom{K}{m}^{-1} e^{-b_\alpha} = \binom{K}{m}^{-1} [m(K-m)]^{-1}\alpha, \qquad L := \binom{K}{m} m(K-m). \tag{89}$$

其中我们回顾 b_α 定义于（72）。于是结论立即由定理 1 得出，只要我们验证条件（12）–（14）对损失函数 $\mathcal{W}$ （见（73））、代价 c_α 以及常数 L 成立。

我们从条件（12）开始。由于 $1 \le m < K$，对每个 $D \in \mathcal{A}$，存在 $i \in D$ 和 $i' \in D^c$。现在，令 $A := (D \setminus \{i\}) \cup \{i'\}$。于是显然 $|A| = m$，从而 $A \in \mathcal{A}$。然后由定义（73），我们有 $W(D|A) = 1$，这验证了条件（12）。

接下来，我们考虑条件（13）。由于 $b_\alpha = \log\left[\left(\binom{K}{m} c_\alpha\right)^{-1}\right]$，由引理 19，我们有

$$T_G(b_\alpha) \le T_{\mathrm{Ld}}(c_\alpha, \mathcal{W}).$$

最后，我们关注条件（14）。对任意程序 $\delta = (T, D) \in \Delta_m^{\mathrm{kns}}(\alpha)$，由（4）、（73）和（89）中的定义，我们有

$$\mathrm{ERR}^{\mathrm{int}}(\delta; \mathcal{W}) = \sum_{|A|=m} \pi_0(A)\mathbb{P}_A(|D \triangle A| \ge 1) \le \alpha \sum_{|A|=m} \pi_0(A) = \alpha = Lc_\alpha.$$

证明完毕。 □

回顾式（73）中的损失函数 W 、式（ $\bar{\lambda}_t^{(k)}$ 70）中 $\bar{j}_k$的定义，以及附录 t（D.1）中 $\bar{j}_k$ 的定义。

引理 19。假设代价为 $c \le \binom{K}{m}^{-1}$。那么，对于阈值 $b = \log\left[\left(\binom{K}{m} c\right)^{-1}\right]$，有

$$T_G(b) \le T_{\mathrm{Ld}}(c, \mathcal{W}).$$

证明。由于 c、b 和 W 均为固定值，我们在 $T_G(b)$ 和 $T_{\mathrm{Ld}}(c,\ \mathrm{W})$ 中省略它们的依赖关系。假设 $T_{\mathrm{Ld}} < \infty$，否则结论显然成立。我们证明 T_G 的停止准则在时间 T_{Ld} 满足，即

$$\bar{\lambda}^{(K-m+1)}_{T_{\mathrm{Ld}}} - \bar{\lambda}^{(K-m)}_{T_{\mathrm{Ld}}} \ge b. \tag{90}$$

这立即意味着 $T_G \le T_{\mathrm{Ld}}$。

令 $D_0 := \{\bar{j}_{K-m+1}(T_{\mathrm{Ld}}),\ \ldots\ldots,\ \bar{j}_K(T_{\mathrm{Ld}})\}$ 表示在 T_{Ld} 时具有最大 m 个对数似然比的数据流索引集合。我们定义

$$A_0 := (D_0 \setminus \{\bar{j}_{K-m+1}(T_{\mathrm{Ld}})\}) \cup \{\bar{j}_{K-m}(T_{\mathrm{Ld}})\},$$

这意味着

$$D_0 \setminus A_0 = \{\bar{j}_{K-m+1}(T_{\mathrm{Ld}})\}, \qquad A_0 \setminus D_0 = \{\bar{j}_{K-m}(T_{\mathrm{Ld}})\}.$$

根据引理14，我们有

$$\sum_{i \in D_0} \lambda^i_{T_{\mathrm{Ld}}} - \sum_{j \in A_0} \lambda^j_{T_{\mathrm{Ld}}} > \log[c^{-1}\pi_0(A_0)\mathcal{W}(D_{\mathrm{Ld}} \mid A_0)].$$

一方面，我们有

$$\sum_{i \in D_0} \lambda^i_{T_{\mathrm{Ld}}} - \sum_{j \in A_0} \lambda^j_{T_{\mathrm{Ld}}} = \bar{\lambda}^{(K-m+1)}_{T_{\mathrm{Ld}}} - \bar{\lambda}^{(K-m)}_{T_{\mathrm{Ld}}}.$$

另一方面，由于 $D_{\mathrm{Ld}} \ne A_0$，根据 $\mathcal{W}$ 在（73）中的定义，$\mathcal{W}(D_{\mathrm{Ld}} \mid A_0) = 1$。由于 π_0（A_0）$= \left(\binom{K}{m}\right)^{-1}$，因此（90）成立，证明完毕。 □

附录 F：非对称情形的数值结果

在本附录中，我们给出一般化误分类率下的其他数值结果。回顾第 5 节中的设定。考虑一种非齐次且非对称的设定，其中

$$K = 20, \qquad \mu_1 = 1/2,\ \mu_2 = \cdots = \mu_K = 1, \qquad \sigma_1^2 = \cdots = \sigma_K^2 = 1,$$

因此，第一个假设比其他假设难解决得多。我们取 $m_0 = 1$ 并考虑空信号子集 $A = \emptyset$。在这种情况下，

$$C^{\mathcal{W}}_{\emptyset} = \{\{1\}\}, \qquad r^{\mathcal{W}}_{\emptyset} = 1, \quad \text{and} \quad \mathrm{KL}^{\mathcal{W}}_{\emptyset,*} = \mathrm{KL}^{\mathcal{W}}_{\emptyset,\emptyset} = 1/8.$$

由（30）和定理 4(a)，一阶近似和二阶近似相同，且

$$T^{\min}_{\emptyset}\left(\Delta^{\mathrm{gmr}}_{m_0}(\alpha)\right) = \mathrm{SO}_\alpha + O(1), \quad \text{where} \quad \mathrm{SO}_\alpha := 8|\log\alpha|.$$

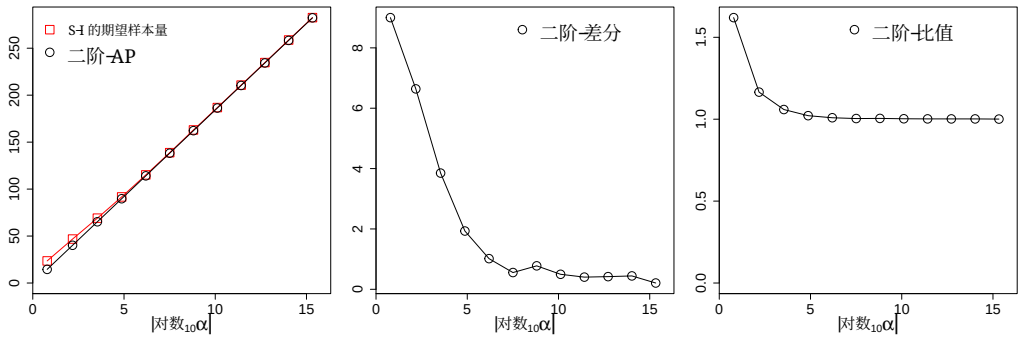


图7：非对称情形： $K = 20$， $m_0 = 1$。所有面板的x轴均为 $|\text{对数}_{10}\,\alpha|$。最左侧的图中，“S-I”表示和交规则，“AP”表示对可实现的最小 ESS 的渐近近似。中间的图中，“Diff.”表示S-I规则的期望样本量与两种近似值之间的差值。最右侧的图中，“Ratio”表示S-I规则的期望样本量与两种近似值各自的比值。

我们在图7中展示数值结果。最左侧的图绘出了期望样本量 $\mathbb{E}_\emptyset[T_S(b^*_\alpha)]$（方形标记）以及二阶近似 $\mathrm{SO}\alpha$（圆形标记），它们均为 $\alpha \in (0, 1)$ 变化时 $|\log_{10}(\alpha)|$ 的函数。中间和最右侧的图分别绘出了差值 $\mathbb{E}_\emptyset[T_S(b^*_\alpha)] - \mathrm{SO}\alpha$ 和比值 $\mathbb{E}_\emptyset[T_S(b^*_\alpha)]/\mathrm{SO}\alpha$，它们均为 $|\log_{10}(\alpha)|$ 的函数。尽管定理 3 和 4(a) 仅能推出 $\mathbb{E}_\emptyset[T_S(b^*_\alpha)] - \mathrm{SO}_\alpha = O(1)$，中间的图表明该差值会随着 $\alpha \to 0$ 趋于零。若能严格证明这一改进结果，将很有意义。