有界均值的渐近最优序贯变化检测

Ashwin Ram ARAM2@ANDREW.CMU.EDU
Aaditya Ramdas ARAMDAS@CMU.EDU
卡内基梅隆大学

arXiv:2602.05272v1 [math.ST] 5 Feb 2026

摘要

我们研究平均运行长度（ARL）约束下的最快变点检测问题，其中变化前分布和变化后分布分别属于复合族 $\mathcal{P}$ 和 $\mathcal{Q}$。在此类问题中，一个巨大的挑战是刻画可能达到的最佳检测延迟，因为 $\mathcal{P}$ 中“最难”的变化前分布取决于未知的变化后分布 $Q \in \mathcal{Q}$。而 Page-CUSUM 和 Shiryaev-Roberts 的典型简单假设似然比论证在此完全不适用。为此，我们在一般情形下，为任何经 ARL 校准的变点检测器推导出一个通用且精确的下界；该下界在第一类错误较低（$\gamma \to \infty$）的情形下，阶为 $\log(\gamma)/\mathrm{KL}_{\inf}(Q, \mathcal{P})$。我们证明，在重要的有界均值检测问题中，通过给出一个与之紧匹配的上界（具有相同的精确 $\log \gamma$ 常数），即可达到这一通用下界。此外，对于均值变化彼此分离的情形，我们还证明了这一可达性对所有备择假设均具有统一的极小极大保证。

关键词：最快变点检测、平均运行长度、检测延迟、序贯分析、有界均值检测。

1. 引言

考虑序贯分析中的以下典型情形：我们观测一个数据流，变化发生后检测到它，随后尝试构建检测器，以缩短这种“延迟”。在1954年的“连续检查”研究中（Page，1954），Page 将这一想法正式化。简而言之，我们不断累积证据，表明数据不再来自基准分布，并在证据越过阈值时停止。Shiryaev 的贝叶斯表述（Shiryaev，1961，1963）将这类停止规则构建为最优停止问题。另一方面，Lorden（Lorden，1971）和 Pollak（Pollak，1985）提出了非贝叶斯的最坏情形延迟准则，这些准则在当今的流式应用中仍占据重要地位（Basseville 和 Nikiforov，1993；Poor 和 Hadjiliadis，2008；Tartakovsky 等，2014）。在变化前分布 f 和已知的变化后分布 g 均已知的经典独立同分布（i.i.d.）模型下，Moustakides 证明，Page 的 CUSUM 程序对于 Lorden 准则实际上是精确最优的（Moustakides，1986）。Pollak 的 CADD 则促成了 Shiryaev-Roberts（SR）程序的改进版本，使其达到近乎极小极大性能（Pollak，1985；Pollak 和 Tartakovsky，2009；Polunchenko 和 Tartakovsky，2010）。

然而，在实践中，假设变化前分布和变化后分布都是单一分布，未免过于简化。例如，传感器漂移可能使变化前分布具有任意形式，只需满足某种性质（如有界均值 $\leq m$）；而异常则可能是违反该性质的“任何情况”（如均值 $> m$）。尤其是在如此宽泛的情形下，实践者仍可能希望采用一种停止规则 T，以保证特定的虚警率。这正是本文的核心动机，我们将在集合层面形式化这一设定。我们假设

© A. Ram & A. Ramdas。

无变化律为某个未知 $P \in \mathcal{P}$ 下的独立同分布 P；而给定在时刻 k 发生变化后，数据的分布将切换为独立同分布 $Q \in \mathcal{Q}$。我们将虚警约束定义为平均运行长度的下界

$$\inf_{P \in \mathcal{P}} \mathbb{E}_{P^\infty}[T] \geq \gamma, \tag{1}$$

而变化后的延迟则以 Pollak 条件平均检测延迟来衡量，

$$\mathcal{C}_Q(T) := \sup_{k \geq 1} \mathbb{E}_{k,Q}\left[(T-k+1)^+ \,\middle|\, T \geq k\right], \tag{2}$$

它始终不超过 Lorden 最坏情形条件延迟（Lorden，1971年；Pollak，1985年）。然而，在这些复合设定中衡量信息时，我们不再使用基本的 KL（$g\|f$），而是使用最不利分离度，

$$I(Q;\mathcal{P}) \;:=\; \inf_{P \in \mathcal{P}} \mathrm{KL}(Q\|P). \tag{3}$$

请注意，这种 $\mathrm{KL}_{\inf}$ 广泛应用于许多领域，从稳健假设检验到分布鲁棒序贯变点检测；在歧义集中，最不利的变化后分布律会使 KL（$\cdot\|P_0$）最小化 （Molloy 和 Ford，2017；Xie 等人，2024）。综上所述，我们的工作有几项关键贡献，使其有别于已有研究。首先，我们为复合变更前类别给出了完全一般且逐点成立的信息下界。也就是说，对于任意经平均运行长度校准的族 $\{T_\gamma\}$以及任意备择 Q，我们有：

$$\liminf_{\gamma \to \infty} \frac{\mathcal{C}_Q(T_\gamma)}{\log \gamma} \;\geq\; \frac{1}{I(Q;\mathcal{P})}. \tag{4}$$

简而言之，我们首先找出一个近似最不利变化前分布 P_δ，使其满足 $\mathrm{KL}(Q\|P_\delta) \leq I(Q;\mathcal{P}) + \delta$，以此证明该界。随后，我们运用分块论证（这是 (1) 的推论），保证至少能找到一个时间窗口，使得在该窗口内停止的条件零假设概率至多为 f/γ（其中 f 是每个分块的长度）。接着我们证明，在 Q 下，除非对数似然随机游走累积了异常多的证据，否则在这个特定“窗口”内停止的概率非常小。但正如我们将看到的，极大强大数定律排除了这种情况。关于这些内容，后文还会详述。其次，我们针对有界均值模型给出一个紧的上界，方法是构造一个有界均值检测器，使其达到 (4) 中的等式。也就是说，对于每个 $Q \in \mathcal{Q}_m$（令 Q_m 为均值为 $> m$ 的有界备择分布类，零假设则是均值为 $\leq m$ 的分布），我们有 $\lim_{\gamma\to\infty} \frac{\mathcal{C}_Q(T_\gamma^{\mathrm{BM}})}{\log \gamma} = \frac{1}{\inf \mathrm{KL}(Q;m)}$。这里，$T_{\mathrm{BM}\gamma}$ 是经 ARL 校准的有界均值检测器类，具体定义将在后文给出。最后，对于分离备择分布，我们得到一致极小极大常数 1，也表明该情形下的结果同样是尖锐的。

相关工作

变点检测通常被认为始于 Page 的 CUSUM（Page，1954年），它在满足 Lorden 最坏情形条件延迟准则（已知（f，g）时）下是最优的（Moustakides，1986年；Ritov，1990）。Lorden （1971年）提出了 $\sup_k \operatorname{esssup} \mathbb{E}_k[(T-k+1)^+ \mid \mathcal{F}_{k-1}]$ 准则。该准则实际上非常严格，即使变点与样本路径呈对抗性对齐，也会迫使检测器始终保持快速响应。 Pollak （1985年）提出了 CADD 准则（2），以存活至 k 为条件，并自然导出了基于 SR 的停止规则。此后，许多研究致力于通过 SR 初始加速和准平稳初始化来逼近极小极大性能

（并改善高阶渐近分析）（Pollak 和 Tartakovsky，2009年；Polunchenko 和 Tartakovsky，2010；Tartakovsky 等人，2012）。此外，还有大量研究共同构成了对这些结果的优美统一，并将其与马尔可夫扩展及更新理论联系起来（Basseville 和 Nikiforov，1993年；Poor 和 Hadjiliadis，2008年；Tartakovsky 等人，2014年）。

在复合分布和稳健变点检测中，当变点后参数未知时，通常（事实上也确实）需要使用广义似然比或混合检测器。Lai（1995）在窗口大小缓慢增长时证明了一阶最优性；Mei（2006）研究了变点前和变点后分布参数未知的情形，并在指数族中导出了渐近最优程序；Xie 和 Siegmund（2013）则在多传感器场景中使用混合程序，结合了各数据流的广义似然比。在分散式通信中，Hadjiliadis 等人（2009）表明，当每个传感器都运行累积和并只传输一次（即单次通信）时，相对于 Lorden 准则，这种方法可以达到渐近最优。即使在非独立同分布的情形下，对于复合变点后假设，也存在加权和混合 SR 程序（Tartakovsky 和 Veeravalli，2005；Pergamenchtchikov 和 Tartakovsky，2019）。有趣的是，Huber–Strassen 对 Neyman–Pearson 的基于容量的推广似乎是最不利分布（以及极小极大检验）的起源（Huber 和 Strassen，1973）。如今，对分布具有稳健性的序贯变点检测通常会涉及模糊集（例如，以变点后经验样本为中心的 Wasserstein 球）；在这类情形下，最不利的变点后分布是使 $\mathrm{KL}(\cdot\|P_0)$ 最小的分布，从而得到“稳健”累积和（Molloy 和 Ford，2017；Xie 等人，2024）。

将下注型非负超鞅视为序贯证据的思想，最早可追溯至 Ville（Ville，1939），后来由 Shafer 等人（2011）将其形式化。Howard 等人（2020）提出了通过非负超鞅实现时间一致边界穿越的思想，而 Ruf 等人（2023）将 Ville 定理推广至复合原假设，使我们能够进行非参数的任意时刻有效推断。关于 e 过程的研究也取得了大量进展，涵盖了利用 e 过程对通用推断操作进行综述、将序贯 e 值合并为单个 e 过程，以及提出增长率最优性等具体最优性概念来处理复合假设（Ramdas 和 Wang，2025；Vovk 和 Wang，2024；Gr¨unwald 等人，2024）。该领域的一项关键工作是提出了 E 检测器：它们是从连续时刻启动的 e 过程之和，能够为广泛的非参数问题提供非渐近平均运行长度控制，并实现近乎最优的检测延迟（Shin 等人，2024）。我们还见证了非常重要的归约，将序贯估计与序贯变点检测联系起来（Shekhar 和 Ramdas，2023）。尽管如此，我们是首个证明类似（4）的下界在变点场景中对任意复合 $\mathcal{P}$ 均成立、且不附加限制的工作。此外，我们还在有界均值场景中给出了精确且紧的 $\log\gamma$ 常数（证明了通用下界的可达性）。

本文其余部分安排如下。第2节介绍问题设置和通用下界，随后在第3节证明该下界在有界均值场景下可达到。最后，我们在附录A 和附录B 中给出所有完整证明。

2. CADD下界：将 γ 发送到 ∞

我们在过滤概率空间 $(\Omega\ \mathcal{F},\ (\mathcal{F}_n)_{n\geq 1})$ 上观察序列 $X_1,\ X_2,\ \ldots\ldots$，其中 $\mathcal{F}_n:=\sigma(X_1,\ X_2,\ \ldots\ldots,\ X_n)$。设 $\mathcal{P}$ 是可测空间 $(\mathcal{X},\ \mathcal{A})$ 上一个非空的变化前分布类。对于 $P\in\mathcal{P}$，记 P^∞ 为 $\mathcal{X}^{\mathbb{N}}$ 上的独立同分布乘积测度；换言之，这就是无变化律。现在，设 $\mathcal{Q}$ 是可测

空间 $(\mathcal{X},\ \mathcal{A})$ 上一个非空的变化后分布类。这个分布类完全可能是复合型/参数型（例如，$\mathcal{Q}=\{Q_\nu\colon\ \nu\in\mathcal{V}\}$）；不过，我们将直接从集合层面处理 $\mathcal{Q}$。对于变化时刻 $k\in\{1,2,\dots\}$、变化前分布 $P\in\mathcal{P}$ 和 $Q\in\mathcal{Q}$，记 $\mathbb{P}_k^{(P),\,Q}$，为如下分布： X_1，……， X_{k-1} 独立同分布地取自 P，而 X_k， X_{k+1}，……独立同分布地取自 Q。我们称检测规则为关于滤过 $(\mathcal{F}_n)$ 的停止时刻 T。现在，我们将对变化后分布类和信息机制提出以下说明。此处，我们通常在 $\gamma\to\infty$ 时研究渐近分析，因此考虑满足对每个 γ 均成立的约束的族 T_γ。

如前所述， $\mathcal{Q}$ 表示变化后分布类。每当我们调用涉及 $1/I(Q;\ \mathcal{P})$ 的界时，都会默认将关注范围限制在满足 $0<I(Q;\mathcal{P})\ <\infty$ 的那些 $Q\in\mathcal{Q}$ 上。此外，对于关于 Q 的一致最坏情况陈述，我们有时会假设存在正的信息间隙。也就是说， $I=\inf_{Q\in\mathcal{Q}} I(Q;\mathcal{P})\ >0$。换言之，直观地说，这意味着变化后分布类与 $\mathcal{P}$ 在 $\mathrm{KL}_{\inf}$ 意义下保持一致分离。本节中，在继续之前，我们先给出主定理，即一个完全一般的 CADD 下界。

定理 1 设 $\{T_\gamma\}_{\gamma>0}$ 满足平均运行长度约束 $\inf_{P\in\mathcal{P}}\mathbb{E}_{P^\infty}[T_\gamma]\ge\gamma$。那么，对于每个满足 $0<I(Q;\mathcal{P})\ <\infty$ 的变化后分布律 $Q\in\mathcal{Q}$，

$$\boxed{\liminf_{\gamma\to\infty}\ \frac{\mathcal{C}_Q(T_\gamma)}{\log\gamma}\ \ge\ \frac{1}{I(Q;\mathcal{P})}.}$$

推论 2 设 $\{T_\gamma\}_{\gamma>0}$ 满足平均运行长度约束，$\inf_{P\in\mathcal{P}}\mathbb{E}_{P^\infty}[T_\gamma]\ge\gamma$ 对所有 $\gamma>0$ 均成立。假设$\mathcal{Q}$ 至少包含一个满足 $0<I(Q;\mathcal{P})\ <\infty$ 的 Q。那么，

$$\boxed{\liminf_{\gamma\to\infty}\ \sup_{Q\in\mathcal{Q}}\ \frac{I(Q;\mathcal{P})\,\mathcal{C}_Q(T_\gamma)}{\log\gamma}\ \ge\ 1.}$$

请注意，定理 1 是 Q 中的逐点下界。另一方面，推论 2 是一个极小极大推论，因为它是在 Q 中的最坏情形。然而，它本身并不能推出 $I(Q;\mathcal{P})\mathcal{C}_Q(T_\gamma)\log\gamma\ \ \to 1$ 对所有 $Q\in\mathcal{Q}$ 一致成立。正如我们将看到的，要获得匹配的一致上界，必然需要一些额外的结构。特别是，我们将在第 3 节针对分离的有界均值备择分布给出这样的一致可达性结果。稍后再详述。此时，我们将陈述一系列帮助我们推导这些界的命题和引理。考虑一个具有 $I(Q,\mathcal{P})\ \in(0,\infty)$ 的特定 $Q\in\mathcal{Q}$。我们将阐明的第一个思路专门针对 KL，以便我们可以在某个特定分布律下进行分析。也就是说，对于每个 $\delta\in(0,1)$，都存在一个 $P_\delta\in\mathcal{P}$，使得 $0<\mathrm{KL}(Q\|P_\delta)\ \le I(Q;\mathcal{P})\ +\delta<\infty.$。特别地，$Q\ll P_\delta$。原因是，根据下确界的定义，存在一个序列 $(P_n)_{n\ge1}\subset\mathcal{P}$，满足 $\mathrm{KL}(Q\|P_n)\downarrow I(Q;\mathcal{P})$。现在，由于 $I(Q;\ \mathcal{P})\ \in(0,\infty)$，我们可以轻易地选取足够大的 n_δ，使得 $\mathrm{KL}(Q\|P_{n_\delta})\ \le I(Q;\ \mathcal{P})\ +\delta$。现在，令 $P_\delta:=P_{n_\delta}$。显然，由于 $\mathrm{KL}(Q\|P_\delta)\ <\infty$，必然有 $Q\ll P_\delta$ 成立。因此，$I(Q;\mathcal{P})\ >0$ 蕴含 $0<\mathrm{KL}(Q\|P_\delta)$。

从现在起，我们固定变化前分布律 $\mathbb{P}^{(P_\delta)}{}_{k,Q}$，并将 $\mathbb{P}_{k,Q}:=\mathbb{P}^{(P_\delta)}{}_{k,Q}$ 和 $\mathbb{P}_\infty:=\mathbb{P}_{P_\delta^\infty}$ 分别简记为……此外，定义 $\ell_i:=\log\frac{dQ}{dP_\delta}(X_i)$。按照 $L_{k:k-1}:=0$ 的约定，令$L_{k:n}:=\sum_{i=k}^n\ell_i$。另外，令 $\mu_\delta:=\mathrm{KL}(Q\|P_\delta)\ \in(0,\infty)$，且 $I_\delta:=I(Q;\mathcal{P})\ +\delta$。由此容易看出 $\mathbb{E}_Q[|\ell_1|]<\infty$；在 $\mathbb{P}_{k,Q}$ 条件下，变量 $(\ell_i)_{i\ge k}$ 独立同分布，且 $\mathbb{E}_{k,Q}[\ell_i]=\mu_\delta$ 和 $\mathbb{E}_{k,Q}[|\ell_i|]<\infty$ 对 $i\ge k$ 成立。下面说明原因。首先，写作 $\ell_1=\ell_1^+-\ell_1^-$，其中 $\ell_1^\pm\ge0$。根据 Lebesgue期望定义，$\mathbb{E}_Q[\ell_1]\in\mathbb{R}$ 当且仅当

$\mathbb{E}_Q[\ell_1^+]$ 和 $\mathbb{E}_Q[\ell_1^-]$ 均为有限值。这里，$\mathbb{E}_Q[\ell_1] = \mu_\delta \in (0, \infty)$，因此可得 $\mathbb{E}_Q[\ell_1^\pm] < \infty$ 和$\mathbb{E}_Q[|\ell_1|] = \mathbb{E}_Q[\ell_1^+] + \mathbb{E}_Q[\ell_1^-] < \infty$。在 $\mathbb{P}_{k,Q}$ 下，变化后片段 $(X_i)_{i\geq k}$ 独立同分布，且分布律为Q。因此，$(\ell_i)_{i\geq k}$ 独立同分布，其分布与 Q 下的 ℓ_1 相同，这意味着$\mathbb{E}_{k,Q}[\ell_i] = \mu_\delta$，且$\mathbb{E}_{k,Q}[|\ell_i|] = \mathbb{E}_Q[|\ell_1|] < \infty$ 对所有 $i \geq k$ 均成立。鉴于以上各点，我们还需要明确变化前和变化后分布律下的测度，这正是引入前缀分布律相等的原因。为此，前缀与变化后片段相互独立；我们也将在下面的引理中对此作出形式化说明。

引理 3 对任意 $k \geq 1$ 以及任意 $G \in \mathcal{F}_{k-1}$，$\mathbb{P}_{k,Q}(G) = \mathbb{P}_\infty(G)$。此外，在 $\mathbb{P}_{k,Q}$ 条件下，事件 $\{T \geq k\} \in \mathcal{F}_{k-1}$ 与 $\sigma(X_k, \ldots, X_{k+m})$ 独立，对每个 $m \geq 0$ 都成立。特别地，对于任意满足 $\mathbb{P}_{k,Q}(T \geq k) > 0$，$\mathbb{P}_{k,Q}(E \mid T \geq k) = \mathbb{P}_{k,Q}(E)$ 的事件 $E \in \sigma(X_k, \ldots, X_{k+m})$。

通常，求出任意特定时刻密度的精确表达式也很有用，这正是下面这个引理的动机；该引理给出了时刻 n 的精确密度。此外，我们的引理还阐述了重要的测度变换性质。

引理 4 对于每个 $n \geq 1$，我们有 $\mathbb{P}_{k,Q}|_{\mathcal{F}_n} \ll \mathbb{P}_\infty|_{\mathcal{F}_n}$。特别是，

$$\frac{d\mathbb{P}_{k,Q}|_{\mathcal{F}_n}}{d\mathbb{P}_\infty|_{\mathcal{F}_n}} = \begin{cases} 1, & n < k, \\ \exp(L_{k,n}), & n \geq k. \end{cases}$$

此外，若 $A \in \mathcal{F}_T$，则 $\mathbb{P}_{k,\,Q}(A) = \mathbb{E}_\infty\left[e^{L_{k,T}}\mathbb{1}_A\right]$。另外，取一个特定的 $k \geq 1$ $with \mathbb{P}_\infty(T \geq k) > 0$。于是可得，

$$\mathbb{P}_{k,Q}(A \mid T \geq k) \;=\; \mathbb{E}_\infty\left[e^{L_{k,T}}\mathbb{1}_A \;\middle|\; T \geq k\right].$$

我们还给出一个关于极大强大数定律（强大数定律）的简短引理，以及一个将在证明中用到的推论。

引理 5 设$(Y_i)_{i\geq 1}$ 是独立同分布的，具有 $\mathbb{E}[|Y_1|] < \infty$，且均值为 $\mu \geq 0$。令 $S_m := \sum_{i=1}^m Y_i$，

$$\limsup_{n\to\infty} \frac{1}{n} \max_{1\leq m\leq n} S_m \;= \mu \qquad a.s.$$

因此，若 b_n 满足某个 $\eta > 0$下的 $b_n/n \to \mu + \eta$，则

$$\mathbb{P}\left(\max_{1\leq m\leq n} S_m \;>\; b_n\right) \;\longrightarrow\; 0.$$

现在，我们已经具备了证明一般 CADD 下界所需的所有主要要素。为清晰起见，我们再次回顾一些记号和可测性条件。对于整数 $k \leq n$，令 $L_{k:n} := \sum_{i=k}^n \ell_i$，并作如下约定：$L_{k:k-1} = 0.$。对于 $n < k$，我们还令 $L_{k:n} := 0$，因为这是一个空和。对于固定整数 k 和 $f \in \mathbb{N}$，我们将变点 k 后接下来的 f 步内发生警报的事件记为 $A_k := \{k \leq T \leq k + f - 1\}.$。由于 T 是一个停止时刻，我们知道对于每个 n，都有 $\{T = n\} \in \mathcal{F}_n$，因此 $A_k \cap \{T = n\} = \{k \leq n \leq k + f - 1\} \cap \{T = n\} \in \mathcal{F}_n$，从而 $A_k \in \mathcal{F}_T$。类似地，对于任意 $c \in \mathbb{R}$，$\{L_{k,T} \leq c\} \cap \{T = n\} = \{L_{k:n} \leq c\} \cap \{T = n\} \in \mathcal{F}_n$，所以 $\{L_{k,T} \leq c\} \in \mathcal{F}_T$。最后，$\{T \geq k\} \in \mathcal{F}_{k-1}$，因为 $\{T \leq k-1\} \in \mathcal{F}_{k-1}$；同时 $\{T \geq k\} \in \mathcal{F}_T$，因为在 $\{T = n\}$ 上它等于 $\{n \geq k\} \cap \{T = n\} \in \mathcal{F}_n$。现在，我们将定义一个事实，将一个“块”的概率质量与其左端点处的存活联系起来。换言之，我们将证明，平均运行长度约束必然意味着至少存在一个条件零假设概率质量较小的块。

引理 6 令 T 为满足 $\mathbb{E}_\infty[T] \geq \gamma$ 的任意停止时刻。固定 $f \in \mathbb{N}$，并将 $\mathbb{N}$ 划分为互不相交的块 $C_r := \{(r-1)f+1、\ldots\ldots、rf\}$、$r \geq 1$。令 $x_r := \mathbb{P}_\infty(T \in C_r)$，以及 $y_r := \mathbb{P}_\infty(T \geq (r-1)f+1)$。则，

$$\sum_{r\geq 1} x_r = 1, \qquad \sum_{r\geq 1} y_r \;\geq\; \frac{\mathbb{E}_\infty[T]}{f} \;\geq\; \frac{\gamma}{f}.$$

此外，存在 $r^\star$，满足 $y_{r^\star} > 0$，使得：

$$\frac{x_{r^\star}}{y_{r^\star}} \leq \frac{f}{\gamma}.$$

等价地，存在 $k^\star := (r^\star - 1)f + 1$，使得对于 $A_{k^\star} := \{k^\star \leq T \leq k^\star + f - 1\}$，

$$\mathbb{P}_\infty(A_{k^\star} \mid T \geq k^\star) \;\leq\; \frac{f}{\gamma}.$$

最后，我们给出一个引理，以帮助理解参数的渐近分析。

引理 7 取一个特定但任意的 $\varepsilon, \delta \in (0,1)$；令 $I_\delta := I(Q;\ \mathcal{P}) + \delta$，并令 $\mu_\delta :=$KL 散度$(Q\|P_\delta) \leq I_\delta$。然后定义，

$$f_\gamma = \left\lfloor \frac{(1-\varepsilon)\log\gamma}{I_\delta} \right\rfloor, \qquad c_\gamma := b\log\gamma \quad \textit{for any } b \textit{ with } \frac{(1-\varepsilon)\mu_\delta}{I_\delta} \;<\; b \;<\; 1.$$

那么，当 $\gamma \to \infty$时，

$$\frac{f_\gamma}{\log\gamma} \to \frac{1-\varepsilon}{I_\delta}, \qquad \frac{c_\gamma}{f_\gamma} \to \frac{b}{1-\varepsilon} I_\delta > \mu_\delta, \qquad \frac{e^{c_\gamma} f_\gamma}{\gamma} \to 0.$$

特别地，存在 $\eta > 0$ 和 γ_0，使得对所有 $\gamma \geq \gamma_0$都有$\frac{c_\gamma}{f_\gamma} \geq \mu_\delta + \eta$。

我们现在可以证明定理1。

证明 [定理 1] 的证明取任意满足 $0 < I(Q;\mathcal{P}) < \infty$的 $Q \in \mathcal{Q}$。以下论证实际上关于 Q 是逐点的，因而适用于所有这样的 $Q \in \mathcal{Q}$。我们的证明将按如下方式进行。取 $\varepsilon \in (0,\ 1)$和 $\delta \in (0,\ 1)$；取 P_δ作为近极小化元，并简记$I_\delta := I(Q;\mathcal{P}) + \delta$和 $\mu_\delta :=$ KL 散度$(Q\|P_\delta) \leq I_\delta$。定义

$$f_\gamma := \left\lfloor \frac{(1-\varepsilon)\log\gamma}{I_\delta} \right\rfloor, \qquad c_\gamma := b\log\gamma \quad \text{for any } b \text{ with } \frac{(1-\varepsilon)\mu_\delta}{I_\delta} \;<\; b \;<\; 1.$$

然后，对每个 γ，我们将证明存在一个指标 k_γ，使得

$$\mathbb{E}_{k_\gamma,Q}\left[(T_\gamma - k_\gamma + 1)^+ \,\middle|\, T_\gamma \geq k_\gamma\right] \;\geq\; f_\gamma\,(1 - o(1)). \tag{5}$$

（本证明中以下所有条件概率都有定义，因为我们将证明 $\mathbb{P}_\infty(T_\gamma \geq k_\gamma) > 0$，显然$\mathbb{P}_{k_\gamma,\,Q}(T_\gamma \geq k_\gamma) = \mathbb{P}_\infty(T_\gamma \geq k_\gamma) > 0$）由引理3可知。）因此，这意味着

$$\liminf_{\gamma\to\infty} \frac{\mathcal{C}_Q(T_\gamma)}{\log\gamma} \;\geq\; \frac{1-\varepsilon}{I_\delta}.$$

现在，令 $\delta \downarrow 0$，然后 $\varepsilon \downarrow 0$，即可得到该定理。因此，现在只需证明（5）。首先，由引理7，必定存在一个 $\eta > 0$ 和 γ_0，使得 $\frac{c_\gamma}{f_\gamma} \geq \mu_\delta + \eta$ 对所有 $\gamma \geq \gamma_0$均成立。接下来，我们将选择一个条件零假设概率质量较小的区块。直观地说，我们将时间划分为长度为 f_γ的区块。注意，在变化前分布下，引理6 意味着，存在一个（特殊的）区块：在该区块起始时仍未触发警报的条件下，其内部发生误报的概率很小。这就是我们将要分析的时间窗口。

具体而言，应用引理6 于 T_γ，其中 $f = f_\gamma$，并令 $k_\gamma = (r^\star_\gamma - 1)f_\gamma + 1$ 为其返回的左端点，则有 $y_{r^\star_\gamma} = \mathbb{P}_\infty(T_\gamma \geq k_\gamma) > 0$，并且

$$\mathbb{P}_\infty(A_{k_\gamma} \mid T_\gamma \geq k_\gamma) \leq \frac{f_\gamma}{\gamma}, \qquad A_k := \{k \leq T_\gamma \leq k + f_\gamma - 1\}. \tag{6}$$

现在，我们将应用条件测度变换，将累积证据较低的警报事件与累积证据较高的警报事件区分开来。我们将证明这两种情况都很罕见。具体而言，应用引理4，令 $A = A_k$，并以 $\{T_\gamma \geq k\}$为条件，

$$\mathbb{P}_{k,Q}(A_k \mid T_\gamma \geq k) = \mathbb{E}_\infty\left[e^{L_{k,T_\gamma}} \mathbb{1}_{A_k} \,\middle|\, T_\gamma \geq k\right].$$

现在，我们可以按 $\{L_{k,T_\gamma} \leq c_\gamma\}$ 快速分类；显然它属于 $\mathcal{F}_{T_\gamma}$，

$$\begin{aligned}
\mathbb{P}_{k,Q}(A_k \mid T_\gamma \geq k) &= \mathbb{E}_\infty\left[e^{L_{k,T_\gamma}} \mathbb{1}_{A_k \cap \{L_{k,T_\gamma} \leq c_\gamma\}} \,\middle|\, T_\gamma \geq k\right] \\
&\quad + \mathbb{E}_\infty\left[e^{L_{k,T_\gamma}} \mathbb{1}_{A_k \cap \{L_{k,T_\gamma} > c_\gamma\}} \,\middle|\, T_\gamma \geq k\right] \\
&\leq e^{c_\gamma} \mathbb{P}_\infty(A_k \mid T_\gamma \geq k) + \mathbb{P}_{k,Q}\big(A_k \cap \{L_{k,T_\gamma} > c_\gamma\} \,\big|\, T_\gamma \geq k\big) \\
&=: \text{(I)} + \text{(II)},
\end{aligned}$$

第二项再次用到了引理4 。现在，我们利用平均运行长度约束快速地界定(I) 。由（6）和引理7，(I) $\leq e^{c_\gamma} f_\gamma/\gamma = f_\gamma \gamma^{b-1} \xrightarrow[\gamma\to\infty]{} 0$，这可直接由以下事实得出： $b < 1$ 且 $f_\gamma = O(\log \gamma)$，而$\log \gamma = o(\gamma^\alpha)$ 对任意 $\alpha > 0$均成立。接下来，利用最大强大数定律对(II) 进行条件界定，我们知道 $T_\gamma \leq k + f_\gamma - 1$ 因此可得:

$$A_k \cap \{L_{k,T_\gamma} > c_\gamma\} \subseteq \left\{\max_{1 \leq m \leq f_\gamma} L_{k:k+m-1} > c_\gamma\right\}.$$

这里，右侧的事件属于 $\sigma(X_k, \ldots, X_{k+f_\gamma-1})$ 由此，根据引理 3 以及 (II) 的定义，

$$\text{(II)} \leq \mathbb{P}_{k,Q}\left(\max_{1 \leq m \leq f_\gamma} L_{k:k+m-1} > c_\gamma \,\middle|\, T_\gamma \geq k\right) = \mathbb{P}_{k,Q}\left(\max_{1 \leq m \leq f_\gamma} L_{k:k+m-1} > c_\gamma\right).$$

在 $\mathbb{P}_{k,Q}$ 下，我们知道增量 $(\ell_i)^{i \geq k}$ 是独立同分布的，均值为 μ_δ 且 $\mathbb{E}_{k,Q}[|\ell_i|] < \infty$。并且，根据引理 7，可得 $\frac{c_\gamma}{f_\gamma} \geq \mu_\delta + \eta$ 对所有$\gamma \geq \gamma_0$均成立；因此，引理 5中的推论给出 (II) $\to 0$。现在，如果我们结合这两个结果，就得到

$$\mathbb{P}_{k_\gamma,Q}(T_\gamma \leq k_\gamma + f_\gamma - 1 \mid T_\gamma \geq k_\gamma) \to 0,$$

因此

$$\mathbb{P}_{k_\gamma,Q}(T_\gamma \geq k_\gamma + f_\gamma \mid T_\gamma \geq k_\gamma) \to 1.$$

根据条件马尔可夫不等式，可得：

$$\mathbb{E}_{k_\gamma,Q}\big[(\,T_\gamma - k_\gamma + 1)^+ \,\big|\, T_\gamma \geq k_\gamma\big] \;\geq\; f_\gamma\,\mathbb{P}_{k_\gamma,Q}(T_\gamma \geq k_\gamma + f_\gamma \,|\, T_\gamma \geq k_\gamma) \;=\; f_\gamma\,(1 - o(1)).$$

对 k 取上确界，得到 $\mathcal{C}_Q(T_\gamma) \geq f_\gamma(1 - o(1))$。最后，引理 7 给出 $f_\gamma/\log\gamma = \frac{1-\varepsilon}{I_\delta} + o(1)$，从而证明了 (5)，令 $\delta \downarrow 0$，则 $\varepsilon \downarrow 0$ 完成证明。由于上述论证对任意满足 $Q \in \mathcal{Q}$ 的 $0 < I(Q;\ \mathcal{P}) < \infty$ 均成立，因此该界实际上对所有 $Q \in \mathcal{Q}$ 都成立。 ■

请注意，在下一节中，以下关于 $\mathcal{C}_Q(T_\gamma^{\mathrm{BM}})$ 的所有界对每一个变化前分布 $P \in \mathcal{P}_m$ 都成立。因为，在 $\mathbb{P}_k^{(P)}{}_{,Q}$ 下，变化后片段是独立同分布的 Q，且与 $\mathcal{F}_{k-1}$ 独立。这意味着，我们将在下文给出的上界适用于上述一般下界中使用的任何最不利变化前分布。

3. 有界均值问题的渐近最优 E 检测器

本节中，我们将一般下界应用于经典的有界均值场景，为此增添一些趣味。通过这样做，我们将证明，在“纯 ARL$\to\infty$”情形下，该下界是紧的。也就是说，存在一族停止规则 $\{T_\gamma\}_{\gamma>0}$，它们满足平均运行长度约束，并且对于每个固定的变化后分布 Q，都能达到相匹配的 $\log\gamma$ 延迟常数。此外，在本节中，我们将取 $\mathcal{X} = [0,1]$，并考虑已知的基准 $m \in (0,1)$。记变化前分布类为 $\mathcal{P}_m := \Big\{\mathcal{P}[0,1]\colon\ \mathbb{E}_P[X_1] \leq m\Big\}$。再记变化后分布类为 $\mathcal{Q}_m := \Big\{\mathcal{Q}[0,1]\colon\ \mathbb{E}_Q[X_1] > m\Big\}$，显然它对应于均值高于基准均值的情形。对于这样的 $Q \in \mathcal{Q}_m$，将信息量定义为 $\mathrm{KL}_{\inf}(Q;\ m) := I(Q;\ \mathcal{P}_m) = \inf_{P\in\mathcal{P}_m}\mathrm{KL}(Q\|P)$。

回顾一下，根据我们的定理 1，任何满足平均运行长度约束的停止规则族 $\{T_\gamma\}$ $\inf_{P\in\mathcal{P}_m}\mathbb{E}_{P^\infty}[T_\gamma] \geq \gamma$，都必须满足检测延迟的通用下界：$\log\gamma/\mathrm{KL}_{\inf}(Q;\ m)$（当 $\gamma\to\infty$ 时）。接下来，我们将证明这个下界可以达到，因此它是紧的。由此，我们将确立渐近最优性（对于推论 2中的一致性概念也成立）。此外，我们将构造一个检测器，使其在 $\mathcal{Q}_m$ 的分离子类上达到匹配的一致上界。首先，我们定义构建 $\mathrm{KL}_{\inf}(Q;\ m)$的变分表示所需的一个概念。具体而言，对于 $\lambda \in (0,\ 1)$ 以及 $x \in [0,1]$，定义如下：

$$L_\lambda(x) := 1 + \lambda\Big(\frac{x}{m} - 1\Big) = (1-\lambda) + \frac{\lambda}{m}x, \tag{7}$$

这是一个 e 值，适用于 $\mathcal{P}_m$（e 值是指在每个 $P \in \mathcal{P}_m$ 下，期望不超过 1 的非负随机变量）。现在，我们定义具有平均运行长度（ARL）控制的混合 Shiryaev-Roberts（SR）型检测器。事实上，它与Shin等人 (2024) 中的 E 检测器完全相同。根据他们的结果，对 E 检测器设定阈值即可直接实现平均运行长度（ARL）控制；不过，为求完整起见，我们仍在下文的定理中给出证明。现在，考虑一个可数稠密集[1] $\Lambda \subset (0,1)$，并选取权重 $(w_\lambda)_{\lambda\in\Lambda}$，使得 $w_\lambda > 0$ 且 $\sum_{\lambda\in\Lambda} w_\lambda = 1$。那么，对每个 $\lambda \in \Lambda$，定义 SR 统计量如下：

$$R_n^{(\lambda)} := \sum_{k=1}^{n}\prod_{i=k}^{n} L_\lambda(X_i), \quad \text{where } R_0^{(\lambda)} := 0,$$

1. 例如，二进有理数。

且 L_λ 与式（7）中的定义相同。显然，通过提出 $L_\lambda(X_n)$，可以看出该统计量具有如下的一步递推关系：

$$R_n^{(\lambda)} = \big(1 + R_{n-1}^{(\lambda)}\big) L_\lambda(X_n). \tag{8}$$

现在，让我们构造混合统计量 $M_n := \sum_{\lambda\in\Lambda} w_\lambda R_n^{(\lambda)}$。接着，定义如下停止规则，作为我们的有界均值检测器：

$$T_\gamma^{\mathrm{BM}} := \inf\{n \geq 1 : M_n \geq \gamma\}. \tag{9}$$

有了这些准备，下面给出我们的可达性定理。

定理 8 取一个特定的 $m \in (0,1)$ 并定义 $\mathcal{P}_m$ 且令 T_γ^{BM}如 (9)所示。则：

1. 首先，我们有平均运行长度校准：$\inf_{P\in\mathcal{P}_m} \mathbb{E}_{P^\infty}[T_{\mathrm{BM}_\gamma} \ ] \geq \gamma$对所有 $\gamma > 0$ 均成立。

2. 其次，在此情形下我们实际上具有渐近最优性。也就是说，对每个 $Q \in \mathcal{Q}_m$,

$$\boxed{\lim_{\gamma\to\infty} \frac{\mathcal{C}_Q(T_\gamma^{\mathrm{BM}})}{\log\gamma} = \frac{1}{\mathrm{KL}_{\inf}(Q;m)}.}$$

3. 第 三，我们在分离类上具有一致极小极大最优性：对于任意 $\Delta \in (0, 1-m)$,

$$\boxed{\lim_{\gamma\to\infty} \sup_{Q\in\mathcal{Q}_{m+\Delta}} \frac{\mathrm{KL}_{\inf}(Q;m)\,\mathcal{C}_Q(T_\gamma^{\mathrm{BM}})}{\log\gamma} = 1.}$$

我们将证明该定理所需的关键要点概述如下。首先，我们需要验证 $\mathrm{KL}_{\inf}(Q;\, m)$ 的正性和有限性。此外，为了获得一致性，我们需要在均值分离条件下存在一致信息间隙。最后，我们采用 Honda 和 Takemura (2010) 的一项著名结论：在均值约束下，对所有分布 P进行的整体最小化，可简化为对 λ进行的一维最大化。我们用以下引理将这些结果形式化。

引理 *9* 假设 Q 支持集包含于 $[0,1]$ 和 $\mathbb{E}_Q[X_1] = q > m$。则有$0 < \mathit{KL}_{\inf}(Q;\, m) < \infty$。此外，$\mathit{KL}_{\inf}(Q;\, m) = \sup_{\lambda\in[0,1]} \mathbb{E}_Q[\log\ L_\lambda(X_1)]$。最后，对于任意的 $\Delta \in (0, 1-m)$，有 $\inf_{Q\in\mathcal{Q}_{m+\Delta}} \mathit{KL}_{\inf}(Q;\, m) \geq 2\Delta^2$。

有了目前为止的这些推导，在定理中证明渐近最优性的实际结果时，我们只需要利用上述引理第二个陈述的如下推论：对于每个 $\eta > 0$，都存在某个 $\lambda \in (0,\ 1)$，使得 $\mathbb{E}_Q[\log\ L_\lambda(X_1)] \geq \mathrm{KL}$ 散度$_{\inf}(Q;\ m) - \eta$。换言之，我们不需要任何闭式最优解 $\lambda^\star$。现在，在继续证明主定理之前，我们还需要一个关于正漂移随机游走的首达时间结论：也就是说，对于有界增量，我们将推导一些首次到达时间渐近结果。

引理 10 设$(Y_i)_{i\geq 1}$ 为独立同分布随机变量，且 $\mathbb{E}[Y_1] = \mu > 0$ ，并假设对某个有限的 $a < b$，$Y_1 \in [a, b]$ 几乎处处成立。定义 $S_n := \sum_{i=1}^n Y_i$ ，对于 $u > 0$， $\tau_u := \inf\{n \geq 1 : \ S_n \geq u\}$。那么，$\tau_u < \infty$ 几乎处处成立，且对每个 $u > 0$， $\mathbb{E}[\tau_u] < \infty$ 。此外，对每个 $u > 0$，

$$\frac{u}{\mu} \leq \mathbb{E}[\tau_u] \leq \frac{u+b}{\mu}. \tag{10}$$

此外，

$$\lim_{u\to\infty} \frac{\mathbb{E}[\tau_u]}{u} = \frac{1}{\mu}.$$

现在我们给出定理8中渐近最优性结论的证明。平均运行长度校准（定理的第一个结论）和一致极小极大证明（定理的第三个结论）将留待附录讨论。

证明 [定理 8中渐近最优性结论的证明。] 现在我们来证明第二个结论。我们先将检测器（如果可以这样说的话）简化为一个“良好”的 λ。为此，我们先考虑任意的 $\eta \in (0, \mathrm{KL}_{\inf}(Q; m))$。我们将证明

$$\limsup_{\gamma\to\infty} \frac{\mathcal{C}_Q(T_\gamma^{\mathrm{BM}})}{\log\gamma} \leq \frac{1}{\mathrm{KL}_{\inf}(Q; m) - \eta},$$

令 $\eta \downarrow 0$，即可得到该定理。

对于某个 $\lambda \in [0, 1)$，定义 $g(\lambda) := \mathbb{E}_Q[\log\ L_\lambda(X_1)]$，使得根据引理 9，$\sup_{\lambda\in[0,1]} g(\lambda) = \mathrm{KL}_{\inf}(Q; m)$。此外，由于对于每个 x， $\lambda \mapsto \log L_\lambda(x)$ 关于 λ 是凹函数，因此函数 $g(\lambda) = \mathbb{E}_Q[\log\ L_\lambda(X_1)]$ 在 $(0, 1)$ 上是凹函数。因此，它在 1 处必定存在左极限，该极限可能为 $-\infty$，故 $\sup_{\lambda\in[0,1]} g(\lambda) = \sup_{\lambda\in(0,1)} g(\lambda)$。现在我们知道，对于每个 x， L_λ 关于 λ 连续，并且对任意 $\delta > 0$，在 $\lambda \in [0, 1-\delta]$ 上一致地远离 0。因此，映射 $\lambda \mapsto g(\lambda)$ 确实在 $[0, 1)$ 上连续，并且若 $Q(\{0\}) > 0$，则它在 $\lambda = 1$ 处可能取值 $-\infty$ 。此外，由于 Λ 在 $(0, 1)$ 中稠密，我们很容易选择 $\lambda_\eta \in \Lambda$ ，使得，

$$d_\eta := g(\lambda_\eta) = \mathbb{E}_Q[\log L_{\lambda_\eta}(X_1)] \geq \mathrm{KL}_{\inf}(Q; m) - \eta > 0.$$

令 $w_\eta := w_{\lambda_\eta} > 0$为其混合权重。注意，对于所有 n，必然有 $M_n \geq w_\eta R_n^{(\lambda_\eta)}$。因此，我们当然可以得到，

$$\{R_n^{(\lambda_\eta)} \geq \gamma/w_\eta\} \subseteq \{M_n \geq \gamma\}.$$

这意味着，停止时刻 T_γ^{BM}始终不大于仅使用单个分量 λ_η 并采用阈值 γ/w_η所得到的停止时刻。为使这一点更加形式化，我们定义 $T^{(\lambda_\eta)\ \gamma/w_\eta} := \inf\{n \geq 1 : \ R^{(\lambda_\eta)n} \ \geq \gamma/w_\eta\}$。于是，我们有概率 1 地得到，

$$T_\gamma^{\mathrm{BM}} \leq T_{\gamma/w_\eta}^{(\lambda_\eta)}. \tag{11}$$

因此，我们已将这一想法形式化：混合分布的停止时间只能早于任何一个按其权重缩放的单独分量。既然如此，我们现在将论证，在变点之后，一个 SR项就足够了，即 $R^{(\lambda_\eta)n}$ 包含正确的乘积。因此，考虑一个特定的

变点 $k \geq 1$。对于任意变化前分布律 $P \in \mathcal{P}_m$，在 $\mathbb{P}_k^{(P)}$，Q 下，根据定义，我们知道变化后片段（X_k，X_{k+1}，.....）是独立同分布的，且独立于 $\mathcal{F}_{k-1}$。也就是说，事件 $\{T_\gamma^{\mathrm{BM}} \geq k\} \in \mathcal{F}_{k-1}$ 确实独立于变点后观测，因为它包含在由变点前观测生成的 σ代数中。现在注意，对于 $n \geq k$, $R^{(\lambda_\eta)n}$ 包含了 $j = k$ 加项；也就是说，

$$R_n^{(\lambda_\eta)} = \sum_{j=1}^{n} \prod_{i=j}^{n} L_{\lambda_\eta}(X_i) \geq \prod_{i=k}^{n} L_{\lambda_\eta}(X_i).$$

直观地说，我们上面说明的是，如果乘积从 k 开始确实达到 γ/w_η，$R^{(\lambda_\eta)n}$ 也必然如此。现在，取对数，令 $S_t := \sum_{i=1}^{t} Y_i$，并令 $Y_i := \log\ L_{\lambda_\eta}(X_{k+i-1})$，其中 $t \geq 1$。于是可得，对于 $t \geq 1$,

$$\prod_{i=k}^{k+t-1} L_{\lambda_\eta}(X_i) = \exp(S_t).$$

现在将变点后越界时间定义为 $\tau_u^{(\eta)} := \inf\{t \geq 1 :\ S_t \geq u\}$。令 $u_\gamma := \log(\gamma/w_\eta)$，我们可以得出结论：$S_{\tau_{u_\gamma}^{(\eta)}} \geq u_\gamma$。因此，

$$\prod_{i=k}^{k+\tau_{u_\gamma}^{(\eta)}-1} L_{\lambda_\eta}(X_i) \geq \gamma/w_\eta \quad \Longrightarrow R_{k+\tau_{u_\gamma}^{(\eta)}-1}^{(\lambda_\eta)} \geq \gamma/w_\eta.$$

因此，$T_{(\lambda_\eta)\gamma/w_\eta} \leq k + \tau_{(\eta)u_\gamma} - 1$。所以必然有$(T_{(\lambda_\eta)_{\gamma/w_\eta}} - k + 1)^+ \leq \tau_{u_\gamma}^{(\eta)}$。现在将此结果与（11）结合，可得路径意义下的不等式，该不等式表明检测延迟不超过此首次到达时间。也就是说，在 $\mathbb{P}_k^{(P)}$，Q 对于每个 $k \geq 1$,

$$(T_{\gamma/w_\eta}^{(\lambda_\eta)} - k + 1)^+ \leq \tau_{u_\gamma}^{(\eta)}. \tag{12}$$

现在我们来说明为什么可以去掉关于存活的条件，即去掉对 $\{T_\gamma^{\mathrm{BM}} \geq k\}$ 的条件。我们知道，$\{T_{\mathrm{BM}_\gamma} \geq k\} \in \mathcal{F}_{k-1}$ 和 $\tau_{(\eta)^{u_\gamma}}$ 仅取决于 $(X_k, X_{k+1}, \ldots)$，因此在 $\mathbb{P}_k^{(P)}$,Q 条件下，两者相互独立。因此，对 $\{T_{\mathrm{BM}_\gamma} \geq k\}$ 取条件不会改变 $\tau_{u_\gamma}^{(\eta)}$ 的分布。因此，

$$\mathbb{E}_{k,Q}^{(P)}[\tau_{u_\gamma}^{(\eta)} \,|\, T_\gamma^{\mathrm{BM}} \geq k] = \mathbb{E}_Q[\tau_{u_\gamma}^{(\eta)}].$$

此外，由（12）并利用条件期望的单调性，很容易看出，

$$\mathbb{E}_{k,Q}^{(P)}[(T_\gamma^{\mathrm{BM}} - k + 1)^+ \,|\, T_\gamma^{\mathrm{BM}} \geq k] \leq \mathbb{E}_Q[\tau_{u_\gamma}^{(\eta)}].$$

现在我们只需对 $k \geq 1$ 取上确界，即可得到，

$$\mathcal{C}_Q(T_\gamma^{\mathrm{BM}}) \leq \mathbb{E}_Q[\tau_{u_\gamma}^{(\eta)}]. \tag{13}$$

我们将通过推导越界时间的渐近分析来完成证明。在变化后分布 Q 下，我们知道 $Y_i = \log\ L_{\lambda_\eta}(X_i)$ 的增量是独立同分布且有界的。由于，$X_i \in [0,1]$意味着 $1 - \lambda_\eta \leq L_{\lambda_\eta}(X_i) \leq (1 - \lambda_\eta) + \frac{\lambda_\eta}{m}$，因此 $\log\ L_{\lambda_\eta}(X_i)$ 确实是有界的。此外，

我们知道 $\mathbb{E}_Q[Y_1] = d_\eta > 0$ 根据构造。因此，我们可以应用引理 10 ，结合 $\mu = d_\eta$得到 $\gamma \to \infty$:

$$\frac{\mathbb{E}_Q[\tau^{(\eta)}_{u_\gamma}]}{u_\gamma} \longrightarrow \frac{1}{d_\eta}.$$

我们知道 $u_\gamma = \log(\gamma/w_\eta) = \log\ \gamma + O(1)$，因此由此可得 $u_\gamma/\log\ \gamma \to 1$。于是，

$$\limsup_{\gamma\to\infty} \frac{\mathbb{E}_Q[\tau^{(\eta)}_{u_\gamma}]}{\log \gamma} = \frac{1}{d_\eta}.$$

最后，如果我们将（13）与此结合使用，就可以得到，

$$\limsup_{\gamma\to\infty} \frac{\mathcal{C}_Q(T^{\mathrm{BM}}_\gamma)}{\log \gamma} \le \frac{1}{d_\eta} \le \frac{1}{\mathrm{KL}_{\inf}(Q; m) - \eta}.$$

现在，我们只需令 $\eta \downarrow 0$，即可得到一个紧确的上界。此外，相应的下界可直接由定理 1 在 $\mathcal{P} = \mathcal{P}_m$ 的特例推出。因此，结合上下界便可得到紧确的极限，这恰好就是定理的第二个陈述。 ■

有一点显而易见，值得指出：如何将本节的结果推广到一般的有界区间。也就是说，若 $X_i \in [a, b]$ 几乎处处，当然可以通过变换 $X_i := (X_i - a)/(b - a) \in [0, 1]$，将问题化简为 $[0, 1]$ 情形 $\widetilde{}$；相应地，$\widetilde{m} := (m - a)/(b - a)$。为此，所得的过程和常数 $\mathrm{KL}_{\inf}(Q;\ m)$ 也将相应地变换。

4. 结论

在复合变点检测中，我们不得不面对这样一个事实：最难与特定 Q 区分的变点前分布事先并不为人所知。因此，正确的信息量确实是$\mathrm{KL}_{\inf}$ 投影 $I(Q; \mathcal{P})$。本文通过通用下界（4），证明了在平均运行长度约束下，这一特定量决定了一阶CADD渐近。随后，我们通过构造基于单步下注因子 L_λ 的混合SR停止规则，证明了有界均值模型中的紧致性（其对数增长与$\mathrm{KL}_{\inf}(Q;\ m)$相匹配）。当然，未来还有若干值得探索的方向。首先，一个直接的方向是运用同一框架，将研究拓展到有界均值检测之外。其次，如何在我们的独立同分布情形精确常数分析与非渐近非独立同分布e-检测器框架（Shin等人，2024；Ruf等人，2023年）之间建立衔接，也颇有意思。这样做或许能得到更优的停止规则：它们不仅在如此广泛的复合原假设下具有分布无关性，而且在具备额外结构（例如独立性）时，在对数 γ 渐近（或其他类似情形）下具有计算效率和常数最优性。

参考文献

Mich`ele Basseville 和 Igor V. Nikiforov。《突变检测：理论与应用》。Prentice Hall Information and System Sciences Series。Prentice Hall，恩格尔伍德克利夫斯，新泽西州，1993年。ISBN 978-0131267800。

Peter Gr¨unwald、Rianne de Heide 和 Wouter Koolen。安全检验。《Journal of the Royal Statis- tical Society: Series B (Statistical Methodology)》，86(5):1091–1128，2024。doi: 10.1093/jrsssb/ qkae011。

Olympia Hadjiliadis、Hongzhong Zhang 和 H. Vincent Poor。分布式最快变化检测的单次方案。*IEEE*信息论汇刊，55(7):3346–3359，2009年7月。doi: 10.1109/TIT.2009.2021311。URLhttps://doi.org/10.1109/TIT.2009.2021311。

Junya Honda 和 Akimichi Takemura。有界支撑模型的渐近最优赌博机算法。载于 Adam Tauman Kalai 和 Mehryar Mohri（编辑），第二十三届学习理论会议论文集（*COLT 2010*），第 67–79 页。Omnipress，2010。ISBN 978- 0-9822529-2-5。URLhttp://www.colt2010.org/papers/29honda.pdf。

Steven R. Howard、Aaditya Ramdas、Jon McAuliffe 和 Jasjeet Sekhon。基于非负超鞅的时间一致的切尔诺夫界。概率调查，17：257–317，2020。doi：10.1214/ 18-PS321。

Peter J. Huber 和 Volker Strassen。容量的极小极大检验与奈曼–皮尔逊引理。《统计年刊》，1(2)：251–263，1973。doi：10.1214/aos/1176342363。

Tze Leung Lai。质量控制和动力系统中的序贯变点检测（附讨论）。英国皇家统计学会会刊：*B*辑（方法论），57(4)：613–644，1995。doi：10.1111/j.2517-6161.1995.tb02052.x。

Gary Lorden。应对分布变化的程序。数理统计年鉴，42(6)：1897–1908，1971年。doi: 10.1214/aoms/1177693055。

Yajun Mei. 变化前分布中存在未知参数时的序贯变化点检测。《统计年刊》，34(1):92–122，2006年。doi: 10.1214/ 009053605000000859.

Timothy L. Molloy 和 Jason J. Ford。设定错误与渐近极小极大稳健最快变化检测。*IEEE*信号处理汇刊，65(21):5730–5742，2017年。doi: 10. 1109/TSP.2017.2740202.

George V. Moustakides。用于检测分布变化的最优停止时间。《统计年刊》，14(4):1379–1387，1986年。doi: 10.1214/aos/1176350164.

E. S. Page。连续检验方案。*Biometrika*，41(1–2):100–115，1954年06月。doi: 10.1093/biomet/41.1-2.100。网址https://doi.org/10.1093/biomet/41.1-2.100。

Serguei Pergamenchtchikov 和 Alexander G. Tartakovsky。针对具有复合变后假设的一般随机模型的渐近最优逐点及极小极大变点检测。多元分析杂志，174，2019 年 11 月。doi: 10.1016/j.jmva.2019.104541。

Moshe Pollak。分布变化的最优检测。《统计年刊》，13(1):206– 227，1985年。doi: 10.1214/aos/1176346587。

Moshe Pollak 和 Alexander G. Tartakovsky。Shiryaev–Roberts 程序的最优性性质。《中华统计学》，19(4):1729–1739，2009年。Aleksey S. Polunchenko 和 Alexander G. Tartakovsky。关于 Shiryaev–Roberts 程序在检测分布变化方面的最优性。《统计年刊》，38(6):3445–3457，2010。doi: 10.1214/09-AOS775。H. Vincent Poor 和 Olympia Hadjiliadis。最快检测。剑桥工程在线。剑桥大学出版社，剑桥，2008年。ISBN 9780521621045。doi: 10.1017/ CBO9780511754678。Aaditya Ramdas 和 Ruodu Wang。使用 e-值进行假设检验。《统计学基础与趋势》，1(1–2):1–390，2025年。doi: 10.1561/3600000002。Ya'acov Ritov。累积和程序的决策理论最优性。《统计年刊》，18(3):1464–1469，1990。doi: 10.1214/aos/1176347761。Johannes Ruf、Martin Larsson、Wouter M. Koolen 和 Aaditya Ramdas。利用 e 过程对 Ville 鞅定理进行复合推广。《电子概率学期刊》，28(127): 1–21，2023年。doi: 10.1214/23-EJP1019。Glenn Shafer、Alexander Shen、Nikolai Vereshchagin 和 Vladimir Vovk。检验鞅、贝叶斯因子和 p-值。《统计科学》，26(1):84–101，2011年。doi: 10.1214/10-STS347。Shubhanshu Shekhar 和 Aaditya Ramdas。通过反向置信序列进行序贯变点检测。载于第*40*届国际机器学习会议论文集，第202卷机器学习研究论文集，第30908–30930页，2023年。Jaehyeok Shin、Aaditya Ramdas 和 Alessandro Rinaldo。E检测器：一种用于序贯变化检测的非参数框架。《新英格兰数据科学统计学期刊》，2(2):229– 260，2024。doi: 10.53002/nejsds.59。A. N. Shiryaev。平稳过程中扰动最快检测问题。《苏联数学—报告》，2:795–799，1961年。Albert N. Shiryaev。关于最快检测问题中的最优方法。《概率论及其应用》，8(1):22–46，1963年。doi: 10.1137/1108002。Alexander Tartakovsky、Igor Nikiforov 和 Mich`ele Basseville。《序贯分析：假设检验与变点检测》，统计与应用概率专著第134卷。统计与应用概率专著。CRC出版社，2014年。ISBN 978-1-4398-3820-4。Alexander G. Tartakovsky 和 Venugopal V. Veeravalli。最快变化检测的一般渐近贝叶斯理论。《概率论及其应用》，49(3):458–497，2005年。doi: 10.1137/S0040585X97981202。Alexander G. Tartakovsky、Moshe Pollak 和 Aleksey S. Polunchenko。广义Shiryaev–Roberts变点检测程序的三阶渐近最优性。《概率论及其应用》，56(3):457–484，2012年。

Jean Ville。《集体》概念的批判性研究。*Gauthier-Villars*，1939。

Vladimir Vovk 和 Ruodu Wang。通过鞅合并序贯 e 值。统计学电子期刊，18(1):1185–1205，2024。doi: 10.1214/24-EJS2228。

Liyan Xie、Yuchen Liang 和 Venugopal V. Veeravalli。利用 Wasserstein不确定集进行分布鲁棒最快变化检测。载于第 *27* 届国际人工智能与统计学会议论文集，第 238 卷机器学习研究论文集，第 1063–1071 页。PMLR，2024。

Yao Xie 和 David Siegmund。序贯多传感器变点检测。《统计年刊》，41(2):670–692，2013。doi: 10.1214/13-AOS1094。

附录A. 第2节省略的证明

证明 [引理 3] 的证明。令 $\mathscr{C}$ 为 $\mathcal{F}_{k-1}$ 中形如 $C = \{(x_1, \ldots, x_{k-1}) \in A\}$ 且满足 $A \in \mathcal{A}^{\otimes(k-1)}$的柱集的集合。在 $\mathbb{P}_{k,Q}$ 和 $\mathbb{P}_\infty$下，$(X_1, \ldots, X_{k-1})$ 的有限维分布相同，且均等于 P_δ^{k-1}，因此对于所有 $C \in \mathscr{C}$，$\mathbb{P}_{k,Q}(C) = \mathbb{P}_\infty(C)$。由于 $\mathscr{C}$显然是生成 $\mathcal{F}_{k-1}$ 的 π-系，且两个测度在其上相等，根据 Dynkin π-λ定理，它们在整个 $\mathcal{F}_{k-1}$上必然相等。

还需证明引理的第二个断言。在 $\mathbb{P}_{k,Q}$ 下，向量 $(X_1, \ldots, X_{k-1})$ 的分布律为 P_δ^{k-1}，且独立于 $(X_k, X_{k+1}, \ldots)$，后者是独立同分布的 Q。由于 T是一个停止时刻，$\{T \geq k\} \in \mathcal{F}_{k-1}$因此仅是 $(X_1, \ldots, X_{k-1})$ 的函数。因此，$\{T \geq k\}$ 独立于 $\sigma(X_k, \ldots)$，从而证明了该断言。 ■

证明 [引理 4] 的证明我们先证明该引理的第一个结论。对于 $n < k$，在两种测度下都有 $(X_1, \ldots, X_n) \sim P_\delta^n$，因此密度等于 1。对于 $n \geq k$，我们利用独立性以及 Radon-Nikodym导数的链式法则。对于 $A \in \mathcal{F}_n$,

$$\mathbb{P}_{k,Q}(A) = \int \left(\prod_{i=1}^{k-1} dP_\delta(x_i)\right)\left(\prod_{i=k}^{n} dQ(x_i)\right)\mathbb{1}_A = \int \left(\prod_{i=1}^{n} dP_\delta(x_i)\right)\prod_{i=k}^{n}\frac{dQ}{dP_\delta}(x_i)\,\mathbb{1}_A.$$

因此，相对于 P_δ^n（即 $\mathbb{P}_\infty|_{\mathcal{F}_n}$）的密度等于 $\prod_{i=k}^n \frac{dQ}{dP_\delta}(X_i) = \exp(L_{k:n})$。由于$Q \ll P_\delta$，由此必然满足绝对连续性。现在我们证明引理的第二个结论。由于 T 是一个停止时刻，对于所有 n，都有 $\{T = n\} \in \mathcal{F}_n$ 和 $A \cap \{T = n\} \in \mathcal{F}_n$。此外，在事件 $\{T = n\}$ 上，我们有 $L_{k,\,T} = L_{k:n}$；并且 $e^{L_{k,\,T}}\mathbb{1}_{A\cap\{T=n\}}$ 是非负且$\mathcal{F}_n$-可测的。因此，通过分解可得，

$$1_A = \sum_{n\geq 1} \mathbb{1}_{A\cap\{T=n\}} \quad \text{(disjoint union)}.$$

它遵由引理的第一个结论可知，4 再由单调收敛定理可得，

$$\begin{aligned}\mathbb{P}_{k,Q}(A) &= \sum_{n\geq 1}\mathbb{P}_{k,Q}(A\cap\{T=n\}) = \sum_{n\geq 1}\mathbb{E}_\infty\left[\frac{d\mathbb{P}_{k,Q}|_{\mathcal{F}_n}}{d\mathbb{P}_\infty|_{\mathcal{F}_n}}\mathbb{1}_{A\cap\{T=n\}}\right]\\ &= \sum_{n\geq 1}\mathbb{E}_\infty\left[e^{L_{k:n}}\mathbb{1}_{A\cap\{T=n\}}\right] = \sum_{n\geq 1}\mathbb{E}_\infty\left[e^{L_{k:T}}\mathbb{1}_{A\cap\{T=n\}}\right] = \mathbb{E}_\infty\left[e^{L_{k:T}}\sum_{n\geq 1}\mathbb{1}_{A\cap\{T=n\}}\right]\\ &= \mathbb{E}_\infty\left[e^{L_{k:T}}\mathbb{1}_A\right].\end{aligned}$$

现在，我们通过证明引理的第三个断言来完成证明。根据引理 4 的第二个断言，对于任意 $B\in\mathcal{F}_T$，我们有，$\mathbb{P}_{k,\,Q}(B) = \mathbb{E}_\infty[e^{L_{k,T}}\mathbb{1}_B]$。将其应用于 $B = A\cap\{T\geq k\}$，然后除以 $\mathbb{P}_{k,\,Q}(T\geq k) = \mathbb{P}_\infty(T\geq k) > 0$ (这两个概率相等是由引理 3 得出的)，即可得到条件恒等式，证明完毕。 ■

证明 [引理 5] 的证明我们知道，根据强大数定律以及$\mathbb{E}[|Y_1|]<\infty$这一事实，存在一个几乎必然成立的事件Ω_0，在该事件上，

$$\lim_{m\to\infty}\frac{S_m(\omega)}{m} = \mu.$$

现在，固定某个 $\omega\in\Omega_0$ 和 $\varepsilon\in(0,1)$，那么存在 $N_\varepsilon(\omega)$，使得 $S_m(\omega)\leq(\mu+\varepsilon)m$ 对所有 $m\geq N_\varepsilon(\omega)$均成立。对于这个 $n\geq N_\varepsilon(\omega)$，

$$\max_{1\leq m\leq n}S_m(\omega) = \max\left\{\max_{1\leq m<N_\varepsilon(\omega)}S_m(\omega),\ \max_{N_\varepsilon(\omega)\leq m\leq n}S_m(\omega)\right\}\leq C_\varepsilon(\omega)+(\mu+\varepsilon)n,$$

其中有限和 $C_\varepsilon(\omega) := \max_{1\leq m<N_\varepsilon(\omega)}S_m(\omega) < \infty$。因此，除以 n 并取上极限可得，

$$\limsup_{n\to\infty}\frac{1}{n}\max_{1\leq m\leq n}S_m(\omega)\ \leq\ \mu+\varepsilon.$$

由于 $\varepsilon>0$ 是任意的，第一个断言必须几乎处处成立。为证明概率界，我们首先需要取一个 $\eta>0$ 并选择 $\varepsilon\in(0,\eta)$。现在，在同一个几乎处处成立的事件上，对于所有足够大的 n，

$$\max_{1\leq m\leq n}S_m\ \leq\ C_\varepsilon+(\mu+\varepsilon)n<(\mu+\eta)n\ \leq\ b_n,$$

由于 $b_n/n\to\mu+\eta$。这意味着事件 $\{\max_{1\leq m\leq n}S_m>b_n\}$ 几乎处处只会发生有限次。因此，它们的概率趋于 0。 ■

证明 [引理 6] 的证明 令 $s_n := \mathbb{P}_\infty(T\geq n)$，即为非递增函数。根据期望的定义，我们知道 $\mathbb{E}_\infty[T]=\sum_{n\geq 1}s_n$在 $[0,\infty]$中成立。因此，在每个区间块中，对于 $C_r := \{(r-1)f+1,\ldots,rf\}$，我们有，

$$s_{(r-1)f+1}\ \geq\ \frac{1}{f}\sum_{n\in C_r}s_n,$$

因此，

$$\sum_{r\geq 1}y_r = \sum_{r\geq 1}s_{(r-1)f+1}\ \geq\ \frac{1}{f}\sum_{r\geq 1}\sum_{n\in C_r}s_n = \frac{1}{f}\sum_{n\geq 1}s_n = \frac{\mathbb{E}_\infty[T]}{f}\ \geq\ \frac{\gamma}{f}.$$

此外，$\sum_{r\ge 1} x_r = 1$ 根据不相交性；定义 $r_r := x_r/y_r$ 对于 $y_r > 0$，并在 $y_r = 0$ 时令 $r_r := 0$。注意，由于 $\{T \in C_r\} \subseteq \{T \ge (r-1)f + 1\}$，所以当 $y_r = 0$ 时确实有 $x_r = 0$。于是，

$$\sum_{r\ge 1} y_r r_r = \sum_{r\ge 1} x_r = 1.$$

将平均值限制在满足 $y_r > 0$ 的指标上，可得，

$$\sum_{r:\, y_r>0} \underbrace{\frac{y_r}{\sum_{j:\, y_j>0} y_j}}_{\text{weights}} r_r \;=\; \frac{1}{\sum_{j:\, y_j>0} y_j} \;\le\; \frac{1}{\sum_{j\ge 1} y_j} \;\le\; \frac{f}{\gamma}.$$

因此 $\min_{r:\, y_r>0} r_r \le f/\gamma$。取达到该最小值的 $r^\star$；则有 $y_{r^\star} > 0$ 和 $x_{r^\star}/y_{r^\star} \le f/\gamma$。
■

证明 [引理的证明 7] 首先，写成，

$$x_\gamma := \frac{(1-\varepsilon)\log\gamma}{I_\delta}, \qquad f_\gamma = x_\gamma - \theta_\gamma \quad \text{with } \theta_\gamma \in [0, 1).$$

现在，由于 $\log\gamma \to \infty$ 当 $\gamma \to \infty$ 时，我们有，

$$\frac{f_\gamma}{\log\gamma} = \frac{x_\gamma - \theta_\gamma}{\log\gamma} = \frac{1-\varepsilon}{I_\delta} - \frac{\theta_\gamma}{\log\gamma} \xrightarrow[\gamma\to\infty]{} \frac{1-\varepsilon}{I_\delta},$$

这由以下事实可得：$0 \le \theta_\gamma < 1$ 意味着 $\theta_\gamma/\log\gamma \to 0$。应用上述结果，

$$\frac{c_\gamma}{f_\gamma} = \frac{b\log\gamma}{f_\gamma} = \frac{b}{f_\gamma/\log\gamma} \xrightarrow[\gamma\to\infty]{} \frac{b}{(1-\varepsilon)/I_\delta} = \frac{b}{1-\varepsilon} I_\delta.$$

我们假设 $b > (1-\varepsilon)\frac{\mu_\delta}{I_\delta}$，因此将两边乘以 $\frac{I_\delta}{1-\varepsilon}$，得到：

$$\frac{b}{1-\varepsilon} I_\delta \;>\; \mu_\delta.$$

令 $L := \frac{b}{1-\varepsilon} I_\delta$；则 $L > \mu_\delta$。设 $\eta := (L - \mu_\delta)/2 > 0$。正如我们刚才所证明的，$\frac{c_\gamma}{f_\gamma} \to L$ 显然成立，因此存在 γ_0，使得对所有 $\gamma \ge \gamma_0$，都有 $\left|\frac{c_\gamma}{f_\gamma} - L\right| < \eta$。由此可得：

$$\frac{c_\gamma}{f_\gamma} \;\ge\; L - \eta \;=\; \frac{L + \mu_\delta}{2} \;=\; \mu_\delta + \eta \qquad (\gamma \ge \gamma_0).$$

现在只需证明项 $e^{c_\gamma} f_\gamma/\gamma$ 趋于 0。由于 $c_\gamma = b\log\gamma, e^{c_\gamma} = \gamma^b$，因此，

$$\frac{e^{c_\gamma} f_\gamma}{\gamma} = f_\gamma\, \gamma^{\,b-1}.$$

利用对所有 γ 都有 $f_\gamma \le x_\gamma = \frac{1-\varepsilon}{I_\delta}\log\gamma$ 这一事实，

$$0 \;\le\; \frac{e^{c_\gamma} f_\gamma}{\gamma} \;\le\; \frac{1-\varepsilon}{I_\delta} (\log\gamma)\, \gamma^{\,b-1} = \frac{1-\varepsilon}{I_\delta} \frac{\log\gamma}{\gamma^{\,1-b}}.$$

由此显然可见，$\log \frac{\gamma}{\gamma^{1-b}} \to 0$，因此我们得到 $e^{c\gamma} f_\gamma/\gamma \to 0$，证明完毕。 ■

证明 [推论的证明 2] 任取 $Q \in \mathcal{Q}$ 满足 $0 < I(Q;\mathcal{P}) < \infty$。根据定理 1，我们有：

$$\liminf_{\gamma\to\infty} \frac{I(Q;\mathcal{P})\,\mathcal{C}_Q(T_\gamma)}{\log\gamma} \geq 1.$$

我们知道，对于每个 γ，在固定 Q 时，对 $Q \in \mathcal{Q}$ 取上确界所得的值不小于该值。因此ce,

$$\sup_{Q\in\mathcal{Q}} \frac{I(Q;\mathcal{P})\,\mathcal{C}_Q(T_\gamma)}{\log\gamma} \geq \frac{I(Q;\mathcal{P})\,\mathcal{C}_Q(T_\gamma)}{\log\gamma}.$$

现在我们只需取 $\liminf_{\gamma\to\infty}$，即可得到所需结论，证明完毕。 ■

附录 B. 第 3 节的遗漏证明3

定义 11 对于定义在$(\mathcal{X},\mathcal{A})$上的两个概率测度 P, Q，全变差距离定义为$\|P-Q\|_{\mathrm{TV}} :=$ 上确界$_{A\in\mathcal{A}} |P(A) - Q(A)|$。

证明 [引理 9] 的证明。我们首先证明 KL 散度有限。已知，由于 $q := \mathbb{E}_Q[X_1] > m$，根据定义，我们可以选择 $\varepsilon \in (0,1)$，使得 $(1-\varepsilon)q \leq m$。定义混合分布$P_\varepsilon := (1-\varepsilon)Q + \varepsilon\delta_0$。显然，根据我们选择 ε 的方式可知，$\mathbb{E}_{P_\varepsilon}[X_1] = (1-\varepsilon)\mathbb{E}_Q[X_1] \leq m$。因此，根据定义，混合分布 $P_\varepsilon \in \mathcal{P}_m$。此外，对于任意可测的 A，我们有 $P_\varepsilon(A) = (1-\varepsilon)Q(A) + \varepsilon\delta_0(A) \geq (1-\varepsilon)Q(A)$。因此，对于所有 A，$Q(A) \leq \frac{1}{1-\varepsilon} P_\varepsilon(A)$。所以，根据拉东-尼科迪姆定理，这个支配不等式实际上等价于 $Q \ll P_\varepsilon$，因而几乎处处相对于 $\frac{dQ}{dP_\varepsilon} \leq \frac{1}{1-\varepsilon}$ Q 成立。因此，我们得到，

$$\mathrm{KL}(Q\|P_\varepsilon) = \mathbb{E}_Q\left[\log\left(\frac{dQ}{dP_\varepsilon}\right)\right] \leq \log\left(\frac{1}{1-\varepsilon}\right) < \infty.$$

现在我们只需对所有 $P \in \mathcal{P}_m$ 取下确界，即可得到 $\mathrm{KL}_{\inf}(Q; m) \leq \mathrm{KL}(Q\|P_\varepsilon) < \infty$。接下来我们将证明其严格为正。用反证法证明。也就是说，反设 $\mathrm{KL}_{\inf}(Q; m) = 0$。这意味着存在一个序列 $(P_n)_{n\geq 1} \subset \mathcal{P}_m$，使得 $\mathrm{KL}(Q\|P_n) \to 0$。现在我们可以利用 Pinsker 不等式得到 $\|Q-P_n\|_{\mathrm{TV}} \leq \sqrt{\mathrm{KL}(Q\|P_n)}_2 \longrightarrow 0$。现在，由于 $X_1 \in [0,1]$，很容易看出 $|\mathbb{E}_Q[X_1] - \mathbb{E}_{P_n}[X_1]| \leq \|Q - P_n\|_{\mathrm{TV}} \longrightarrow 0$。因此，$\mathbb{E}_{P_n}[X_1] \to \mathbb{E}_Q[X_1] = q > m$，这与对所有 n 都有 $\mathbb{E}_{P_n}[X_1] \leq m$ 这一事实相矛盾。我们知道 $P_n \in \mathcal{P}_m$。所以，最初的假设必定是错误的！因此，确实有 $\mathrm{KL}_{\inf}(Q; m) > 0$。

现在我们来证明一致间隙陈述。取一个 $Q \in \mathcal{Q}_{m+\Delta}$ 和 $P \in \mathcal{P}_m$。那么根据定义，我们有 $\mathbb{E}_Q[X_1] \geq m+\Delta$和 $\mathbb{E}_P[X_1] \leq m$。因此，我们得到 $\mathbb{E}_Q[X_1] - \mathbb{E}_P[X_1] \geq \Delta$。显然，$\mathbb{E}_Q[X_1] - \mathbb{E}_P[X_1] \leq |\mathbb{E}_Q[X_1] - \mathbb{E}_P[X_1]| \leq \|Q-P\|_{\mathrm{TV}}$。于是 $\|Q-P\|_{\mathrm{TV}} \geq \Delta$。现在应用 Pinsker 不等式。这样可得 $\Delta \leq \|Q-P\|_{\mathrm{TV}} \leq \sqrt{\mathrm{KL}(Q\|P)}_2$。因此，我们得到 $\mathrm{KL}(Q\|P) \geq 2\Delta^2$。现在对 $P \in \mathcal{P}_m$ 取下确界，得到 $\mathrm{KL}_{\inf}(Q; m) \geq 2\Delta^2$。然后，对所有 $Q \in \mathcal{Q}_{m+\Delta}$ 取下确界，恰好得到引理中这部分的结论，因此证明完毕。

接下来还需证明 $\mathrm{KL}_{\inf}(Q;\, m) = \sup_{\lambda\in[0,1]} \mathbb{E}_Q[\log\, L_\lambda(X_1)]$。需要指出的是，这在文献中早已广为人知，Honda 和 Takemura（2010）对此作了详尽论述。不过，正如我们所提到的，这里仍给出一个自包含的证明。我们先来证明另一个事实。取某个 $\lambda \in (0, 1)$ 和 $P \in \mathcal{P}_m$，使得 $Q \ll P$。将似然比简记为 $r := \frac{dQ}{dP}$。由于 $X_1 \in [0,1]$ 和 $\mathbb{E}_P[X_1] \le m$，正如上文所述，显然必有 $\mathbb{E}_P[L_\lambda(X_1)] = 1 + \lambda\Big(\frac{\mathbb{E}_P[X_1]}{m} - 1\Big) \le 1$。现在我们当然知道，

$$\mathrm{KL}(Q\|P) - \mathbb{E}_Q[\log L_\lambda(X_1)] = \mathbb{E}_Q\left[\log\Big(\frac{r(X_1)}{L_\lambda(X_1)}\Big)\right] = \mathbb{E}_P\left[r(X_1)\log\Big(\frac{r(X_1)}{L_\lambda(X_1)}\Big)\right].$$

由此，我们可以直接应用对数和不等式，或者利用 KL 散度的非负性，得到：

$$\mathbb{E}_P\left[r\log\Big(\frac{r}{L_\lambda}\Big)\right] \ge \Big(\mathbb{E}_P[r]\Big)\log\Big(\frac{\mathbb{E}_P[r]}{\mathbb{E}_P[L_\lambda]}\Big) = 1\cdot\log\Big(\frac{1}{\mathbb{E}_P[L_\lambda]}\Big) \ge 0,$$

因为我们还知道 $\mathbb{E}_P[r] = \int \frac{dQ}{dP}dP = \int dQ = 1$ 和 $\mathbb{E}_P[L_\lambda] \le 1$。因此，综合以上结果，我们可以直接得出 $\mathrm{KL}(Q\|P) \ge \mathbb{E}_Q[\log\, L_\lambda(X_1)]$。现在回到要证明 $\mathrm{KL}_{\inf}(Q;\, m) = \sup_{\lambda\in[0,1]} \mathbb{E}_Q[\log\, L_\lambda(X_1)]$ 这一任务。我们的证明将分为两个不等式来完成。首先，我们证明 $\ge$ 方向；具体而言，先证明不包含边界的情形，再处理边界。首先，任取 $\lambda \in (0, 1)$。正如我们刚刚证明的，$\mathrm{KL}(Q\|P) \ge \mathbb{E}_Q[\log\, L_\lambda(X_1)]$。然后，对 $P \in \mathcal{P}_m$ 取下确界，得到 $\mathrm{KL}_{\inf}(Q;\, m) \ge \mathbb{E}_Q[\log\, L_\lambda(X_1)]$。接着，对所有 $\lambda \in (0, 1)$ 取上确界，得到 $\mathrm{KL}_{\inf}(Q;\, m) \ge \sup_{\lambda\in(0,1)} \mathbb{E}_Q[\log\, L_\lambda(X_1)]$。现在来看边界。由于 $L_0 \equiv 1$，我们得到 $\mathbb{E}_Q[\log L_0(X_1)] = 0$；又由于 0 是一个有限值，它确实可以由闭区间 $(0, 1)$ 上的上确界取到。对于另一种情况，在 $\lambda = 1$处，$\log L_1(X_1) = \log(X_1/m)$ 确实可能为 $-\infty$，原因有二：首先，因为 $Q\{0\} > 0$（即，Q 可以以非零概率取值 0）；或者因为 $\mathbb{E}_Q[\log X_1] = -\infty$。但无论哪种情况，它当然都不会使上确界增大。相反，如果 $\mathbb{E}_Q[\log L_1(X_1)] > -\infty$，那么根据控制收敛定理，沿着 $\lambda_n \uparrow 1$，我们得到 $\mathbb{E}_Q[\log L_{\lambda_n}(X_1)] \to \mathbb{E}_Q[\log L_1(X_1)]$,，因此端点处的值同样可以由 $(0, 1)$ 上的上确界取到。因此必有 $\mathrm{KL}_{\inf}(Q;\, m) \ge \sup_{\lambda\in[0,1]} \mathbb{E}_Q[\log\, L_\lambda(X_1)]$。

本证明的其余部分将致力于证明 $\le$方向，从而得到所需的等式。我们将通过凸对偶性来证明。也就是说，我们将把定义 KL 散度$_{\inf}(Q;\, m)$ 的最小化问题写成凸规划，然后计算其拉格朗日对偶。首先，我们将原问题化简为相对于 Q 的密度。考虑某个 $P \in \mathcal{P}_m$，使得 $\mathrm{KL}(Q\|P) < \infty$，那么我们必然知道 $Q \ll P$。现在，令 $P\mathrm{ac}$ 表示勒贝格分解 $P = P^{\mathrm{ac}} + P\mathrm{s}$ 中 P 相对于 Q 的绝对连续部分，其中此处$P\mathrm{s} \perp Q$（即，相对于 Q 奇异）。此外，定义简写 $p := \frac{dP^{\mathrm{ac}}}{dQ}$。由于 $P\mathrm{s} \perp Q$，根据奇异性的定义，必定存在一个可测集 N，使得 $Q(N) = 0$ 且$P\mathrm{s}(N^c) = 0$。现在，如果 $P\mathrm{ac}(A) = 0$，那么 $P(A\cap N^c) = P\mathrm{ac}(A\cap N^c) + P\mathrm{s}(A\cap N^c) = 0$。因此，由于 $Q \ll P$，我们得到 $Q(A\cap N^c) = 0$。现在请注意，我们可以将 $A = (A\cap N^c)\cup(A\cap N)$ 写成这样。我们刚刚证明了 $Q(A \cap N^c) = 0$。当然，集合 $A \cap N$ 是 N 的子集，因此 $Q(A \cap N)$ 必须为 0，因为 $Q(N) = 0$。由于 $Q(A)$ 只不过是这两个零测集并集的测度，因此 $Q(A)$ 也为零。因此，$Q \ll P^{\mathrm{ac}}$，因为我们已经证明，任何满足$P^{(\mathrm{ac})}(A) = 0$ 的集合 A，都蕴含 $Q(A) = 0$。由此可知，$p > 0$ Q 几乎处处成立。除此之外，在

Q-支撑上，由于奇异部分消失，我们有 $\frac{dQ}{dP} = \frac{dQ}{dP^{\mathrm{ac}}} = 1/p$。换言之，这些导数在 Q 有质量的地方相等。因此，我们得到：

$$\mathrm{KL}(Q\|P) = \mathbb{E}_Q\left[\log\Big(\frac{dQ}{dP}(X_1)\Big)\right] = \mathbb{E}_Q[-\log p(X_1)].$$

令 $c := \mathbb{E}_Q[p(X_1)] = P^{\mathrm{ac}}([0,1]) \le 1$。根据零假设的均值约束，我们知道：

$$\mathbb{E}_Q[X_1 p(X_1)] = \int x\, P^{\mathrm{ac}}(dx) \le \int x\, P(dx) = \mathbb{E}_P[X_1] \le m,$$

因为 P^{s} 对 $\mathbb{E}_P[X_1]$ 的贡献非负，因为请回忆 $X_1 \in [0,1]$。这意味着，每个可行零假设 P 都会导出一个函数 p，满足四个性质。首先，$p > 0$ Q- 几乎处处成立；其次，$\mathbb{E}_Q[p(X_1)] \le 1$；第三，$\mathbb{E}_Q[X_1 p(X_1)] \le m$；最后，$\mathrm{KL}(Q\|P) = \mathbb{E}_Q[-\log p(X_1)]$。

反过来，考虑某个固定的可测 $p > 0$，Q-几乎处处满足 $\mathbb{E}_Q[p(X_1)] \le 1$ 且 $\mathbb{E}_Q[X_1 p(X_1)] \le m$。为避免均值约束恰好取等时出现任何细微问题，取某个 $\delta \in (0,1)$，并定义 $p_\delta := (1-\delta)p$。那么我们知道 $\mathbb{E}_Q[p_\delta(X_1)] < 1$ 且 $\mathbb{E}_Q[X_1 p_\delta(X_1)] < (1-\delta)m < m$。现在引入“松弛”变量，以补足达到完整约束所需的剩余质量差和均值差：也就是说，令质量松弛量为 $s_\delta := 1 - \mathbb{E}_Q[p_\delta(X_1)] > 0$，均值松弛量为 $\mathrm{slack}_\delta := m - \mathbb{E}_Q[X_1 p_\delta(X_1)] > 0$。根据定义，$Q$ 是有限测度（任何概率测度都必须是有限测度），因此它只有可数多个原子。由于存在不可数多个无原子点，所以我们知道，对于每个 $\varepsilon > 0$，必定存在某个 $x \in (0,\varepsilon)$，使得 $Q(\{x\}) = 0$。为此，选择某个奇异点 $x_\delta > 0$，使得 $Q(\{x_\delta\}) = 0$ 且 $x_\delta \le \mathrm{slack}_{s_\delta}{}^{\delta}$。现在定义 $P_\delta(dx) := p_\delta(x)Q(dx) + s_\delta \delta_{x_\delta}(dx)$。显然，$P_\delta$ 确实是一个概率测度。此外，我们有：

$$\mathbb{E}_{P_\delta}[X_1] = \mathbb{E}_Q[X_1 p_\delta(X_1)] + s_\delta x_\delta \le \mathbb{E}_Q[X_1 p_\delta(X_1)] + \mathrm{slack}_\delta = m,$$

因此，当然 $P_\delta \in \mathcal{P}_m$。最后，由于 δ_{x_δ} 的质量集中在一个 Q-零测集上，因此它不会影响 Q 支撑集上的 $\frac{dQ}{dP_\delta}$。因此，我们有：

$$\mathrm{KL}(Q\|P_\delta) = \mathbb{E}_Q[-\log p_\delta(X_1)] = \mathbb{E}_Q[-\log p(X_1)] - \log(1-\delta).$$

现在，只要令 $\delta \downarrow 0$，就可以看出，对所有 $P \in \mathcal{P}_m$ 进行优化等价于对这类 p进行优化。因此，我们可以巧妙地将原始问题改写为：

$$\mathrm{KL}_{\inf}(Q;m) = \inf\left\{ \mathbb{E}_Q[-\log p(X_1)] : \begin{array}{l} p > 0 \ \ Q\text{-a.s.}, \\ \mathbb{E}_Q[p(X_1)] \le 1, \\ \mathbb{E}_Q[X_1 p(X_1)] \le m \end{array} \right\}. \tag{14}$$

显然，这是关于变量 p 的凸优化问题，因为 $p \mapsto -\log p$ 是凸的，且约束确实是线性的。此外，我们知道斯莱特条件的严格可行性成立。为什么？为了说明这一点，我们可以很容易地按如下方式进行。由于 $q = \mathbb{E}_Q[X_1] > m$，任取常数 $c \in (0,\ m/q)$。显然，$p \equiv c$ 将满足 $p > 0$、$\mathbb{E}_Q[p] = c < 1$ 和 $\mathbb{E}_Q[X_1 p] = cq < m$。因此，（14）确实是严格可行的。因此，不存在对偶间隙，且对偶最优解等于原问题最优解。因此，我们现在可以计算拉格朗日对偶来求解这个凸优化问题。首先，我们引入乘子 $\alpha \ge 0$ 和

$\beta \geq 0$ 分别用于约束 $\mathbb{E}_Q[p] \leq 1$ 和 $\mathbb{E}_Q[X_1 p] \leq m$。于是，拉格朗日函数变为，

$$\mathcal{L}(p;\alpha,\beta) := \mathbb{E}_Q[-\log p(X_1) + (\alpha+\beta X_1)p(X_1)] - \alpha - \beta m.$$

现在，对于某个特定的(α,β)对，我们在 $p>0$ 上逐点最小化拉格朗日函数 $\mathcal{L}$：首先，取某个 $x \in [0,1]$ 并令 $c(x) := \alpha + \beta x$。考虑 $\varphi_x(p) := -\log p + c(x)p$（其中 $p>0$如上所述）。如果事实上 $c(x) \leq 0$，则必然意味着$\inf_{p>0}\varphi_x(p) = -\infty$（令$p \to \infty$趋于无穷）。因此，我们将参数限制为使 $c(X_1) > 0$ Q-几乎必然成立；事实上，在我们的设定中，只需假设 $\alpha > 0$，或者更一般地假设 $\alpha + \beta x > 0$ Q-几乎必然成立。为此，当 $c(x) > 0$时，导数 $\varphi_x'(p) = -1/p + c(x)$将在 $p^\star(x) = 1/c(x)$处为零。因此，最小值就是 $\varphi_x(p^\star(x)) = \log c(x) + 1$。于是对偶函数变为，

$$\inf_{p>0} \mathcal{L}(p;\alpha,\beta) = 1 + \mathbb{E}_Q[\log(\alpha+\beta X_1)] - \alpha - \beta m,$$

相应的对偶问题很简单，即：

$$\sup_{\alpha\geq 0,\ \beta\geq 0:\ \alpha+\beta X_1>0\ Q\text{-a.s.}} \Big\{1 + \mathbb{E}_Q[\log(\alpha+\beta X_1)] - \alpha - \beta m\Big\}. \tag{15}$$

而且，正如我们已经解释过的，不存在对偶间隙，因此，(15) 的值等于 $\mathrm{KL}_{\inf}(Q;\, m)$ 的值。现在，我们将通过把 (15) 化简为单参数族 L_λ 来完成证明。也就是说，任取一个 (α,β)，满足 $\alpha > 0$ 且 $\beta \geq 0$，并考虑按 $t>0$ 缩放，$(\alpha,\beta) \mapsto (t\alpha, t\beta)$。于是，对偶目标函数变为：

$$1 + \mathbb{E}_Q[\log(t\alpha + t\beta X_1)] - t\alpha - t\beta m = 1 + \log t + \mathbb{E}_Q[\log(\alpha+\beta X_1)] - t(\alpha+\beta m).$$

我们知道，对于某个特定的 (α,β) 对，右侧只是关于 $t>0$ 的凹函数，因此在 $\frac{d}{dt}(\log t - t(\alpha+\beta m)) = \frac{1}{t} - (\alpha+\beta m) = 0$ 时达到最大值，或者更简单地说，在 $t^\star = \frac{1}{\alpha+\beta m}$ 时达到最大值。接着，将 $t^\star$代入即可得到最大值 $\mathbb{E}_Q[\log(\alpha+\beta X_1)] - \log(\alpha+\beta m)$。因此，在不损失最优性的前提下，我们可以将注意力限制在满足 $\alpha + \beta m = 1$ 的归一化参数上。在这种归一化下，对偶目标函数将化简为 $\mathbb{E}_Q[\log(\alpha+\beta X_1)]$。现在，我们用 $\lambda := \beta m \in [0,1]$ 对 $\alpha + \beta m = 1$ 进行参数化，于是 $\alpha = 1-\lambda$，且 $\beta = \lambda/m$。由此可得 $\alpha + \beta X_1 = (1-\lambda) + \frac{\lambda}{m} X_1 = L_\lambda(X_1)$。于是，对偶值变为 $\sup_{\lambda\in[0,1]} \mathbb{E}_Q[\log L_\lambda(X_1)]$。显然，由于对偶最优值等于原问题的最优值，我们得到：

$$\mathrm{KL}_{\inf}(Q;m) \leq \sup_{\lambda\in[0,1]} \mathbb{E}_Q[\log L_\lambda(X_1)],$$

如果将这一点与证明的第一部分结合起来，我们确实可以得到等式，从而完成证明。既然我们已经证明了该引理的所有部分，现在就大功告成了。 ■

证明 [引理 10] 的证明我们首先需要证明的是 $\tau_u < \infty$ 几乎处处（即，非亏格）。显然，我们知道 $\mathbb{E}[|Y_1|] < \infty$，这是由有界性得出的。此外，我们假设 $\mathbb{E}[Y_1] = \mu > 0$。根据强大数定律，我们知道 $S_n/n \to \mu > 0$ 几乎处处。因此，$S_n \to +\infty$ 几乎处处。因此，由于可以保证 S_n 总会在某个时刻超过阈值 u，我们知道对于每个固定的 $u>0$，$\tau_u < \infty$几乎处处。

现在我们需要证明 $\mathbb{E}[\tau_u] < \infty$。取任意 $u>0$，并定义有界停止规则 $\tau_{(N)u} := \tau_u \wedge N$。我们知道，过程 $M_n := S_n - \mu n$ 是关于

滤过 $\mathcal{F}_n := \sigma(Y_1,\ Y_2,\ \ldots\ldots,\ Y_n)$ 的鞅。根据有界时刻 $\tau_u^{(N)}$ 的可选停止定理，我们有 $\mathbb{E}[S_{\tau_u^{(N)}} - \mu\tau_u^{(N)}] = \mathbb{E}[M_{\tau_u^{(N)}}] = \mathbb{E}[M_0] = 0$。因此，$\mathbb{E}[S_{\tau_u^{(N)}}] = \mu\mathbb{E}[\tau(N)u\quad]$。现在我们来从上方界定 $S_{\tau_u^{(N)}}$。在 $\{\tau_u \le N\}$ 发生的事件上，我们知道 $S_{\tau_u - 1} < u$ 且 $Y_{\tau_u} \le b$，因此 $S_{\tau_u} = S_{\tau_u - 1} + Y_{\tau_u} < u + b$。而在 $\{\tau_u > N\}$ 发生的事件上，根据 τ_u 的定义，我们有 $S_N < u$。因此，$S_{\tau_u^{(N)}} = S_N < u < u + b$。无论如何，我们可以确定，$S_{\tau_u^{(N)}} < u + b$ 以概率 1 成立。迅速取期望可得 $\mathbb{E}[S_{\tau_u^{(N)}}] \le u + b$，这意味着对所有 N，

$$\mathbb{E}[\tau_u^{(N)}] = \frac{\mathbb{E}[S_{\tau_u^{(N)}}]}{\mu} \le \frac{u+b}{\mu}.$$

现在，若令 $N \to \infty$ 并对 $\tau_u^{(N)} \uparrow \tau_u$ 应用单调收敛定理，我们得到，

$$\mathbb{E}[\tau_u] = \lim_{N\to\infty} \mathbb{E}[\tau_u^{(N)}] \le \frac{u+b}{\mu} < \infty.$$

因此，我们已经证明期望确实有限：$\mathbb{E}[\tau_u] < \infty$。现在，我们将把这里证明的结果从有界停止时推广到实际的无界停止时刻τ_u。此外，我们知道 $M_n = S_n - \mu n$ 的鞅差是有界的。也就是说，$|M_n - M_{n-1}| = |Y_n - \mu| \le \max\{|a-\mu|, |b-\mu|\} := C < \infty$。因此，对每个 N，$|M_{\tau_u^{(N)}}| \le C\tau_u^{(N)}$几乎处处成立。因此，对每个 N，我们知道 $\sup_N \mathbb{E}[|M_{\tau_u^{(N)}}|] \le C\mathbb{E}[\tau_u] < \infty$。根据控制收敛定理，我们知道 $M_{\tau_u^{(N)}} \to M_{\tau_u}$ 在 L^1中收敛。因此，我们得到$\mathbb{E}[M_{\tau_u}] = \lim_{N\to\infty} \mathbb{E}[M_{\tau_u^{(N)}}] = 0$。等价地，

$$\mathbb{E}[S_{\tau_u}] = \mu\mathbb{E}[\tau_u]. \tag{16}$$

最后，我们知道超越量也是有界的。根据定义 $S_{\tau_u} \ge u$ 并且 $S_{\tau_u} < u + b$（使用上述相同论证即可立即得出）。因此可得 $u \le \mathbb{E}[S_{\tau_u}] \le u + b$。由（16）可得，

$$\frac{u}{\mu} \le \mathbb{E}[\tau_u] \le \frac{u+b}{\mu}.$$

现在我们只需除以 u 并令 $u \to \infty$，由夹逼定理可得$\mathbb{E}[\tau_u]/u \to 1/\mu$，因此证明完成。

■

证明 [定理的证明8：陈述 1和3。] 我们从第一个陈述开始。考虑任意选定的一个 $P \in \mathcal{P}_m$。首先，我们将证明，对于每个 $\lambda \in \Lambda$，过程$R_n^{(\lambda)} - n$ 在 P^∞下是一个上鞅。事实上，由（8）和期望的塔式性质可知，

$$\mathbb{E}_{P^\infty}\left[R_n^{(\lambda)} \mid \mathcal{F}_{n-1}\right] = \left(1 + R_{n-1}^{(\lambda)}\right)\mathbb{E}_{P^\infty}[L_\lambda(X_n) \mid \mathcal{F}_{n-1}].$$

我们知道，在 P^∞ 下，X_n 与 σ代数 $\mathcal{F}_{n-1}$ 独立，并且 $\mathbb{E}_P[X_1] \le m$，因此得到：

$$\mathbb{E}_{P^\infty}[L_\lambda(X_n) \mid \mathcal{F}_{n-1}] = \mathbb{E}_P[L_\lambda(X_1)] = 1 + \lambda\Big(\frac{\mathbb{E}_P[X_1]}{m} - 1\Big) \le 1.$$

由此可得 $\mathbb{E}_{P^\infty}[R_n^{(\lambda)} \mid \mathcal{F}_{n-1}] \le 1 + R_{n-1}^{(\lambda)}$。等价地，$\mathbb{E}_{P^\infty}[R_n^{\lambda} - n \mid \mathcal{F}_{n-1}] \le R_{n-1}^{(\lambda)} - (n-1)$。所以，我们知道 $R_n^{(\lambda)} - n$ 确实是一个上鞅。取凸组合可得 $M_n - n = \sum_{\lambda\in\Lambda} w_\lambda R_n^{(\lambda)} - n$，它在 P^∞ 下也仍是上鞅。现在令 $T := T_\gamma^{\mathrm{BM}} = \inf\{n \ge 1 : M_n \ge \gamma\}$。一个有缺陷的停止规则（即 $T = \infty$）会意味着 $\mathbb{E}_{P^\infty}[T] \ge \gamma$ 立即成立。因此，不失一般性，我们可以假设 $\mathbb{E}_{P^\infty}[T] < \infty$。特别地，假设 $T < \infty$ 几乎处处成立。对每个 $N \in \mathbb{N}$，若我们对有界停止规则 $T \wedge N$ 应用可选停止定理，可以看出 $\mathbb{E}_{P^\infty}[M_{T\wedge N} - (T \wedge N)] \le M_0 - 0 = 0$。因此，$\mathbb{E}_{P^\infty}[M_{T\wedge N}] \le \mathbb{E}_{P^\infty}[T \wedge N]$。由于 $T < \infty$ 几乎处处成立，我们有 $T \wedge N \uparrow T$ 几乎处处成立，$N \to \infty$ 也如此。此外，我们知道 $(M_n)_{n\ge 0}$ 是适应的，因此这意味着 $M_{T\wedge N} \to M_T$ 几乎处处成立，$N \to \infty$ 也如此。我们还知道 $M_{T\wedge N} \ge 0$。因此，可以应用法图引理得到：

$$\mathbb{E}_{P^\infty}[M_T] = \mathbb{E}_{P^\infty}\left[\lim_{N\to\infty} M_{T\wedge N}\right] \le \liminf_{N\to\infty} \mathbb{E}_{P^\infty}[M_{T\wedge N}] \le \liminf_{N\to\infty} \mathbb{E}_{P^\infty}[T \wedge N] = \mathbb{E}_{P^\infty}[T].$$

最后，根据 T 的定义，我们知道在 $\{T < \infty\}$ 上有 $M_T \ge \gamma$。因此可得 $\mathbb{E}_{P^\infty}[T] \ge \mathbb{E}_{P^\infty}[M_T] \ge \gamma$。注意，$P \in \mathcal{P}_m$ 是任意的。因此，我们确实可以得出 $\inf_{P\in\mathcal{P}_m} \mathbb{E}_{P^\infty}[T_{\mathrm{BM}_\gamma}] \ge \gamma$，这就完成了第一个陈述的证明。

现在还需证明该定理的最后一个陈述。我们来证明它。首先，取一个 $\Delta \in (0, 1-m)$，并考虑 $Q \in \mathcal{Q}_{m+\Delta}$。回顾一下，我们取 $I_Q = \mathrm{KL}_{\inf}(Q; m) \in (0, \infty)$。根据引理 9，我们必然有如下统一下界：$I(Q) \ge I_\Delta := 2\Delta^2$，对所有 $Q \in \mathcal{Q}_{m+\Delta}$ 均成立。现在，对于 $\lambda \in [0,1]$，我们定义函数 $g_Q(\lambda) := \mathbb{E}_Q[\log L_\lambda(X_1)]$。当然，根据引理 9，我们知道 $I_Q = \sup_{\lambda\in[0,1]} g_Q(\lambda)$，且 $g_Q(0) = 0$。现在任取 $\varepsilon \in (0, 1/2)$。我们首先要证明，确实存在一个有限的下注比例集合 $\Lambda_{\Delta,\varepsilon} \subset \Lambda$ 和一个常数 $w_{\Delta,\varepsilon} > 0$，使得对于每个 $Q \in \mathcal{Q}_{m+\Delta}$，都存在某个 $\lambda(Q) \in \Lambda_{\Delta,\varepsilon}$，其中，

$$g_Q(\lambda(Q)) \ge (1-\varepsilon) I_Q \quad \text{and} \quad w_{\lambda(Q)} \ge w_{\Delta,\varepsilon}. \tag{17}$$

现在我们来构造这个 $\Lambda_{\Delta,\varepsilon}$。首先定义 $\varepsilon' := \varepsilon/2 \in (0, 1/4)$。我们知道，对于 $u > -1$，$\log(1+u) \le u$。因此，对于任意 $\lambda \in [0,1]$ 和 $x \in [0,1]$，有 $\log L_\lambda(x) = \log\left(1 + \lambda\left(\frac{x}{m} - 1\right)\right) \le \lambda\left(\frac{x}{m} - 1\right) \le \lambda\left(\frac{1}{m} - 1\right)$。由此可得，对所有 $\lambda \in [0,1]$，

$$g_Q(\lambda) \le \lambda\left(\frac{1}{m} - 1\right). \tag{18}$$

现在取一个特定的 $Q \in \mathcal{Q}_{m+\Delta}$。由于 $I_Q = \sup_{\lambda\in[0,1]} g_Q(\lambda)$，我们可以测度地选取一点 $\lambda_Q \in (0, 1)$，使得

$$g_Q(\lambda_Q) \ge (1-\varepsilon') I_Q. \tag{19}$$

将（19）与（18）结合起来，我们得到下界：$(1-\varepsilon') I(Q) \le g_Q(\lambda_Q) \le \lambda_Q\left(\frac{1}{m} - 1\right)$。因此可得：

$$\lambda_Q \ge (1-\varepsilon')\frac{I_Q}{\frac{1}{m} - 1} \ge (1-\varepsilon')\frac{I_\Delta}{\frac{1}{m} - 1} =: \underline{\lambda}_{\Delta,\varepsilon} > 0.$$

现在令网格大小为 $\delta := \varepsilon' \underline{\lambda}_{\Delta,\varepsilon} / 4$。再令 $J := \lfloor 1/\delta \rfloor$，使得 $J\delta \le 1 < (J+1)\delta$。对于每个 $j \in \{1, \dots, J\}$，由于 Λ 在 $(0,1)$ 中稠密，我们可以轻松地选取一个 $\lambda_j \in \Lambda \cap \big(j\delta - \delta/2,\ j\delta\big]$。

此外，我们还要取并选取另一个靠近 1 的点，$\lambda_{J+1} \in \Lambda \cap (1-\delta/2, 1)$。完成这些后，我们来定义这个有限集，

$$\Lambda_{\Delta,\varepsilon} := \{\lambda_1, \ldots, \lambda_{J+1}\} \subset \Lambda, \qquad w_{\Delta,\varepsilon} := \min_{\lambda \in \Lambda_{\Delta,\varepsilon}} w_\lambda > 0.$$

此时，集合已构造完成，现在取一个特定但任意选定的 $Q \in \mathcal{Q}_{m+\Delta}$ 并考虑（λ_Q，它来自（19）。令 $j^\star := \lfloor \lambda_Q/\delta \rfloor$，于是可得 $j^\star \geq 1$，因为我们已知 $\lambda_Q \geq \lambda_{\Delta,\varepsilon} > \delta$。此外，$j^\star \leq J$ 也必然成立，因为 $\lambda_Q < 1$ 且 $J = \lfloor 1/\delta \rfloor$。进一步，根据构造可知 $\lambda_{j^\star} \leq j^\star\delta \leq \lambda_Q$ 且 $\lambda_{j^\star} \geq j^\star\delta - \frac{\delta}{2} \geq \lambda_Q - \delta - \frac{\delta}{2} = \lambda_Q - \frac{3\delta}{2}$。现在，我们知道$\delta = \varepsilon'\lambda_{\Delta,\varepsilon}/4 \leq \varepsilon'\lambda_Q/4$。由此可得 $\frac{3\delta}{2} \leq \varepsilon'\lambda_Q$。因此，我们得到 $\frac{\lambda_{j^\star}}{\lambda_Q} \geq 1 - \varepsilon'$。根据$g_Q$ 的凹性，并结合$g_Q(0) = 0$ 这一事实，对于任意 $t \in [0,1]$，都有$g_Q(t\lambda_Q) \geq t g_Q(\lambda_Q)$。考虑 $t = \lambda_{j^\star}/\lambda_Q \in [1-\varepsilon', 1]$ 可得，

$$g_Q(\lambda_{j^\star}) \geq \frac{\lambda_{j^\star}}{\lambda_Q} g_Q(\lambda_Q) \geq (1-\varepsilon')(1-\varepsilon') I_Q \geq (1-\varepsilon) I_Q,$$

其中，最后一个不等式成立是因为 $(1-\varepsilon')^2 \geq 1-\varepsilon$ whenever $\varepsilon' = \varepsilon/2$。当然，根据构造，我们也知道 $w_{\lambda_{j^\star}} \geq w_{\Delta,\varepsilon}$。因此，(17) 确实在 $\lambda(Q) := \lambda_{j^\star} \in \Lambda_{\Delta,\varepsilon}$时成立。现在，我们用这个良好分量，按照定理陈述对 CADD 作一致上界。与之前一样，考虑一个特定的 $Q \in \mathcal{Q}_{m+\Delta}$，并令 $\lambda(Q)$ 如上定义。我们知道$M_n \geq w_{\lambda(Q)} R_n^{\lambda(Q)} \geq w_{\Delta,\varepsilon} R_n^{(\lambda(Q))}$。因此有逐路径支配关系，

$$T_\gamma^{\mathrm{BM}} \leq \inf\left\{ n \geq 1 : R_n^{(\lambda(Q))} \geq \gamma / w_{\Delta,\varepsilon} \right\} =: T_{\gamma/w_{\Delta,\varepsilon}}^{(\lambda(Q))}.$$

现在，我们可以重复之前逐点论证中使用的变点后乘积论证。由此可知，对每个变化时刻 $k \geq 1$，$\mathcal{C}_Q(T\mathrm{BM}\gamma) \leq \mathbb{E}_Q\left[\tau_{u_\gamma}\right]$，且 $u_\gamma := \log(\gamma/w_{\Delta,\varepsilon})$。这里，$\tau_u := \inf\{t \geq 1 : \sum_{i=1}^t \log L_{\lambda(Q)}(X_i) \geq u\}$。在 Q 下，我们知道增量 Y_i 独立同分布，并且根据 (17) 满足 $\mathbb{E}_Q[Y_1] = g_Q(\lambda(Q)) \geq (1-\varepsilon) I(Q) > 0$。现在验证引理 10 所要求的有界增量假设。我们知道 $X_i \in [0,1]$，且$\lambda \in (0,1)$，因此对每个 $x \in [0,1]$，都有，

$$1 - \lambda \leq L_\lambda(x) = (1-\lambda) + \frac{\lambda}{m} x \leq (1-\lambda) + \frac{\lambda}{m} \leq \frac{1}{m}.$$

因此可得，$\log(1-\lambda) \leq \log L_\lambda(x) \leq \log(1/m)$。现在，由于 $\lambda(Q) \in \Lambda_{\Delta,\varepsilon}$ and$\Lambda_{\Delta,\varepsilon}$ 是一个有限集，下界确实必定一致有界。也就是说，$a_{\Delta,\varepsilon} := \min^{\lambda \in \Lambda_{\Delta,\varepsilon}} \log(1-\lambda) > -\infty$。并且，$b_m := \log(1/m) < \infty$。因此，在$Q$下以概率 1，$Y_1 \in [a_{\Delta,\varepsilon}, b_m]$。由此可见，引理 10 确实适用。这一点着实令人惊讶。因为，引理 10 中证明的界意味着，对于所有 γ，

$$\mathbb{E}_Q[\tau_{u_\gamma}] \leq \frac{u_\gamma + b_m}{\mathbb{E}_Q[Y_1]} \leq \frac{\log(\gamma/w_{\Delta,\varepsilon}) + b_m}{(1-\varepsilon) I_Q}.$$

将这一切乘以 I_Q，然后除以 $\log\gamma$，可得（回顾上文，我们已经证明 $C_Q(T_{\mathrm{BM}}\gamma) \leq \mathbb{E}_Q[\tau_{u_\gamma}]$），

$$\sup_{Q \in \mathcal{Q}_{m+\Delta}} \frac{I_Q \mathcal{C}_Q(T_\gamma^{\mathrm{BM}})}{\log\gamma} \leq \frac{\log\gamma + \log(1/w_{\Delta,\varepsilon}) + b_m}{(1-\varepsilon)\log\gamma}.$$

取 lim sup$_{\gamma\to\infty}$，因此可得，

$$\limsup_{\gamma\to\infty}\sup_{Q\in\mathcal{Q}_{m+\Delta}}\frac{\mathrm{KL}_{\inf}(Q;m)\mathcal{C}_Q(T_\gamma^{\mathrm{BM}})}{\log\gamma}\le\frac{1}{1-\varepsilon}.$$

现在，我们都知道 $\varepsilon\in(0,1/2)$ 是任意的。因此，我们可以令 $\varepsilon\downarrow 0$ 以得到所需的一致上界，≤ 1。但是！请注意，相应的一致下界 ≥ 1 同样可由推论2在 $\mathcal{P}=\mathcal{P}_m$ 与 $\mathcal{Q}=\mathcal{Q}_{m+\Delta}$下的应用得到。因此，

$$\lim_{\gamma\to\infty}\sup_{Q\in\mathcal{Q}_{m+\Delta}}\frac{\mathrm{KL}_{\inf}(Q;m)\mathcal{C}_Q(T_\gamma^{\mathrm{BM}})}{\log\gamma}=1.$$

这正是我们需要证明的第三项。因此，本定理的证明到此完成！

■