arXiv:2301.03542v1 [math.ST] 9 Jan 2023

对数凹性的序贯检验

Aditya Gangrade[1], Alessandro Rinaldo[1] 和Aaditya Ramdas[12]
agangra2@andrew.cmu.edu, arinaldo@cmu.edu, aramdas@cmu.edu

[1]卡内基梅隆大学统计与数据科学系[2]卡内基梅隆大学机器学习系

摘要

观察到一列分布为 P、定义在 $\mathbb{R}^d$ 上的独立同分布数据后，我们探讨如何检验零假设：P 具有对数凹密度。本文给出了一个有趣的否定性结果和一个肯定性结果：检验（超）鞅不存在，而通用推断具有相合性。具体而言，对数凹分布集合 $\mathcal{L}$ 是一个非参数类，其中包含所有均值和协方差任意的高斯分布集合 $\mathcal{G}$。我们进一步发展近期提出的叉凸性这一几何概念，首先证明，对于 $\mathcal{G}$（即对 $\mathcal{G}$ 中的每个分布而言，同时为非负超鞅的过程），不存在任何非平凡检验鞅或检验超鞅，因此对于其超集 $\mathcal{L}$ 也不存在。鉴于这一否定性结果，我们转而构造 e 过程——在 $\mathcal{L}$ 中的任意分布下，该过程在任意停止时刻的期望均不超过一——只需在 $1/\alpha$ 处进行阈值化，即可得到一个水平为 α 的检验。我们采用通用推断的方法，通过将备择假设下的非预见似然与原假设下的最大似然取比，避免了棘手的似然渐近理论。尽管该方法较为保守，我们证明所得检验是相合的（检验功效为一），并推导了其对海灵格备择假设的检验功效。据我们所知，目前不存在其他针对 $\mathcal{L}$ 的 e 过程或序贯检验。

1 引言

在现代形状约束非参数方法研究中，对数凹性是一项重要且普遍采用的建模假设 [Sam18]。对数凹分布涵盖许多常见的密度族，包括正态分布、指数分布、极值分布和逻辑斯谛分布；此外，在诸多应用领域中，这一假设也常有充分依据，例如经济学、工程学中的可靠性理论与滤波，以及医学中的生存分析 [BB06]。与此同时，这一分布族在技术上易于处理，并存在唯一的最大似然估计，且具有完善的极小极大理论和计算高效的估计量 [CS10；CDSS18；KDR19；Axe+19；DR11；RS19；CSS10]。因此，对数凹密度为实践者提供了一种适用范围广且实用的结构。

鉴于在对数凹族中进行估计具有诸多优点，检验分布是否属于该族是一项重要且必要的研究方向。我们指出，除了上述应用之外，此类检验也具有理论意义；例如，在许多计算学习理论研究中，只有当协变量按照对数凹分布抽样时，才已知存在高效的学习算法 [例如KKMS08]。过去十年或二十年间，对数凹密度估计取得了显著进展（例如，参见上述引文以及 Samworth 的综述 [Sam18]），但对数凹性的检验发展相对滞后。事实上，2021 年之前，除某些受限的一维情形外，无论在理论上还是实践中，都没有有效且检验功效强的相关检验。最近，Dunn 等人 [DGWR21] 基于 Wasserman 等人提出的通用推断策略，开发了这样一种检验 [WRB20]。

我们的研究关注在序贯设置下检验对数凹性。具体而言，我们假设可以流式访问序列 $\{X_t\}$，其中各样本独立同分布地取自某个 d 维密度 p，并希望检验 p 是否属于对数凹密度族。此类序贯检验可由停止时刻 τ 表示；停止即表示拒绝原假设

，若原假设成立，则当 $\tau < \infty$ 时其概率不超过 $\alpha.$，该检验称为 α-有效。此类序贯检验的主要优势在于其适应性：检验无需事先固定样本数量，而是可以根据具体问题的难度进行调整，在较容易的情形下更早拒绝原假设，并在需要检测与原假设之间细微偏离时允许使用更多样本。

下面，我们先介绍一些记号，然后说明本研究的背景，并简要概述本文的贡献。

1.1 问题设置与背景

我们首先介绍讨论所需的记号、所研究的检验问题，以及检验鞅和 e-过程的基本概念。我们将在 §2 中给出进一步的定义和细节，并在正文后续部分根据具体语境加以说明。

空间与测度。令 $\{X_t\} = (X_1, X_2, \dots)$ 表示一个由 d 维随机向量构成的序列，其分量以 t 为索引；这些随机向量是从 $\Omega := (\mathbb{R}^d)^{\mathbb{N}}$ 到 $\mathbb{R}^d$ 的可测映射，并赋有柱状 Borel σ-代数 $\mathscr{B}(\mathbb{R}^d)^{\mathbb{N}}$。我们用打字机字体（例如 P）表示随机过程的概率律（即 $(\Omega\ \mathscr{B}(\mathbb{R}^d)^{\mathbb{N}})$ 上的概率测度），用标准字体（例如 P ）表示 $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ 上的概率律。我们用 $\mathscr{F} = \{\mathscr{F}_t\}$ 表示过程 $\{X_t\}$ 的自然滤过，其中对每个 t， $\mathscr{F}_t := \sigma(X_1, \dots, X_t)$。对于 $\mathbb{R}^d$ 上的 Borel 概率测度 P，我们用 P^∞ 表示按 P 抽取的独立同分布过程的概率律。我们用 $\mathcal{D}$ 表示具有勒贝格密度的 $\mathbb{R}^d$ 上概率测度的集合，并且 $\mathcal{D}^\infty = \{P^\infty : P \in \mathcal{D}\}$。对于 $P \in \mathcal{D}$，我们用 p 表示其勒贝格密度。为方便技术处理，我们定义 $\mathcal{D}_1 := \{P \in \mathcal{D} : \mathbb{E}[\max(0, \log p(X))] < \infty, \mathbb{E}[\|X\|] < \infty\}$。如果一组概率律 $\mathcal{P}$ 满足对所有 $P, Q \in \mathcal{P}, P \ll Q \ll P$，则称其相互绝对连续（m.a.c.）。最后，我们常用 $X_1^t := (X_1, \dots, X_t)$ 表示 $\{X_t\}$ 的有限前缀。
对数凹测度。 若存在凹函数g 使得 $f = e^g$，则称函数 $f : \mathbb{R}^d \to \mathbb{R}$ 为对数凹函数。若 f 进一步是相对于勒贝格测度的密度，则称其为对数凹密度。我们将具有对数凹密度的测度集合记为 $\mathcal{L}$，并用 $\mathcal{L}^\infty$ 表示欧几里得序列上的独立同分布对数凹测度集合，即 $\mathcal{L}^\infty := \{P^\infty : P \in \mathcal{L}\}$.

对数凹性序贯检验。 我们关注的检验问题表述如下：设 $X_t \overset{\text{i.i.}}{\sim} P^{\mathrm{d}}$，其中 $P \in \mathcal{D}$. 未知。我们希望检验零假设 $H_0 : P \in \mathcal{L}$。

序贯检验对应于一个 $\{\mathscr{F}_t\}$-适应停止时刻，表示检验停止并拒绝零假设的时间（可能为无限）。我们将这个停止时刻称为该序贯检验的拒绝时间。若某检验的拒绝时间 τ 满足

$$\sup_{P \in \mathcal{L}} P^\infty(\tau < \infty) \le \alpha,$$

这意味着，在原假设下，曾经拒绝原假设（即犯第一类错误）的概率至多为 α。类似地，如果在备择分布 $\mathcal{Q} \subset \mathcal{D} \setminus \mathcal{L}$ 中的任意分布下，未能拒绝原假设的概率（也称第二类错误）均由 β 一致地界定，则称检验相对于 $\mathcal{Q} \subset \mathcal{D} \setminus \mathcal{L}$ 渐近具有 $(1 - \beta)$ 检验功效：

$$\inf_{Q \in \mathcal{Q}} Q^\infty(\tau = \infty) \le \beta.$$

如果某检验相对于 $\mathcal{Q}$ 渐近具有 1 检验功效，则称该检验相对于 $\mathcal{Q}$ 相合。请注意，这类检验在相合时通常称为“功效为一检验”（沿用 Robbins 的说法），以区别于传统的 Wald 式序贯检验范式；在该范式中，停止并不意味着拒绝原假设。

检验鞅、检验超鞅和 e-过程。 我们简要综述本讨论所依据的关键概念，即检验鞅和 e-过程；相关细节分别见第3节和第4 节。

定义过程 $\{M_t\}$ 关于滤过 $\{\mathscr{F}_t\}$ 和概率律 P是非负超鞅 (NSM)，如果它是适应的、非负的，且 $\mathbb{E}_{\mathrm{P}}[M_t | \mathscr{F}_{t-1}] \le M_{t-1}$ 对每个 t 都成立。如果该不等式在每个 t 处进一步取等号，那么 $\{M_t\}$ 就是非负鞅 (NM)。简洁地说，这样的过程分别是 P-*NSM* 或 P-*NM*。

显然，每个P-NM 也是一个 P-NSM。Ville 的一个重要基本不等式 [Vil39] 控制了 NSM 的尾部行为：如果 $\{M_t\}$ 是一个 P-NSM，且 $M_0 = 1$，那么对每个 $\alpha \in (0, 1]$,

$$\mathrm{P}(\exists t \geq 1 : M_t \geq 1/\alpha) \leq \alpha.$$

上述结果是马尔可夫不等式的序贯（时间一致）类比。等价地，我们可以在任意停止时作出断言：对所有停止时刻 τ, $\mathrm{P}(M_\tau \geq 1/\alpha) \leq \alpha$。这一点可通过对 NSM 应用可选停止定理[Mey66，第 V 章，定理 28] 以及马尔可夫不等式得到。

现在，我们将上述概念推广到序贯概率律的复合族。本文始终将滤过取为数据的自然滤过，并在下文的定义中省略不提。

定义对于一组序贯概率律 $\mathfrak{P}$，我们称过程 $\{M_t\}$ 为 $\mathfrak{P}$-NSM，如果对每个 $\mathrm{P} \in \mathfrak{P}$，$\{M_t\}$ 都是一个 P-*NSM*。类似地，如果对每个 $\mathrm{P} \in \mathfrak{P}.$，$\{M_t\}$ 都是一个 P-*NM*，则称 $\{M_t\}$ 为 $\mathfrak{P}$-*NM*。满足 $M_0 = 1$ 的 $\mathfrak{P}$-*NSM* 称为 $\mathfrak{P}$的检验超鞅；满足 $M_0 = 1$ 的 $\mathfrak{P}$-*NM* 称为 $\mathfrak{P}$的检验鞅。

注意，对于每个 $\mathrm{P} \in \mathfrak{P}$，检验超鞅都满足 Ville不等式；也就是说，如果 $\{M_t\}$ 是针对 $\mathfrak{P}$的检验超鞅，那么对每个 $\alpha \in (0, 1]$,

$$\forall \mathrm{P} \in \mathfrak{P},\ \mathrm{P}(\exists t \geq 1 : M_t \geq 1/\alpha) \leq \alpha. \tag{1}$$

检验超鞅之所以得名，是因为它们构成了对复合假设进行序贯检验的标准路径，而这一点完全体现在上述关系中：只有当检验超鞅越过阈值时才拒绝假设，即可构造有效检验。它们在非参数情形中特别有用；例如，可以用它们对有界随机变量的均值进行序贯检验 [WR23],用于检验对称性 [RRLK20], 用于双样本检验 [SR21], 独立性检验 [PBKR22], 以及检验校准 [AHZ21], 仅举几个有趣的序贯非参数问题为例。本文将深入讨论检验超鞅。

E过程 [RGVS22；RRLK20；GHK19；HRMS20] 是近期定义的一类过程，也将在本文中发挥核心作用。

定义若过程 $\{E_t\}$ 非负，且对于任意停止时刻 τ，均有 $\mathbb{E}_\mathrm{P}[E_\tau] \leq 1.$，则称其为关于序贯概率律 P 的 *e*过程。类似地，若 $\{E_t\}$ 是关于每个 $\mathrm{P} \in \mathfrak{P}$的 *e*过程，则称其为序贯概率律类 $\mathfrak{P}$的 *e*过程。

e过程有多种等价定义 [RRLK20，引理 6]。特别地，只需该过程满足 $\mathbb{E}_\mathrm{P}[E_\tau] \leq 1$ 对有界停止时刻成立。

根据可选停止定理（对于非负超鞅，该定理对停止时刻不作限制也成立），注意到，类 $\mathfrak{P}$ 上的每个检验超鞅也是该类的一个 e过程。因此，e过程推广了检验超鞅的概念。我们注意到，e过程也满足 Ville 型关系，这直接源于马尔可夫不等式：若 E_t 是关于 $\mathfrak{P}$的 e过程，则对于每个 $\alpha \in (0, 1]$,

$$\forall \mathrm{P} \in \mathfrak{P},\ \text{stopping times } \tau,\ \mathrm{P}(E_\tau \geq 1/\alpha) \leq \alpha \mathbb{E}[E_\tau] \leq \alpha. \tag{2}$$

正如 Ville不等式在类 (1) 上体现了检验超鞅与序贯检验的相关性，上述不等式也体现了 e-过程与序贯检验的相关性。由于这一关键性质，以及 e-过程存在于许多复合和非参数检验问题中，而这些问题中并不存在检验超鞅，因此 e-过程这一概念连同其非序贯对应概念 e 值，近年来在统计学研究中日益受到关注（例如，参见 Ramdas 等人最近的综述 [RGVS22]）。本文也将遇到这种情况。

需要注意的是，检验超鞅或 e过程可以直接解释为反对零假设的证据：由于我们预期它们在零假设下小于 1，因此其观测值越大，就越能说明零假设不成立。所以，没有必要明确地在 $1/\alpha$ 处设置阈值，或预先指定某个 α；也可以只报告实验最终停止时刻（该时刻本身可以任意选择）对应的最终值。尽管如此，本文仍以显著性水平为 α 的检验来阐述，因为这种表述方式更为普遍。关于 e过程的进一步讨论，感兴趣的读者可参阅前述文献。

1.2 Inad检验（超）鞅的适用性，以及 E 过程的检验功效 s

复合假设序贯检验背后一个重要（但有时不易察觉）的原理，是使用非负鞅（NM）或非负超鞅（NSM）。具体而言，为检验复合假设 $\mathrm{P} \in \mathfrak{P}$，人们会尝试构造一个 $\mathfrak{P}$-检验超鞅 $\{M_t\}$，其定义已在前文给出。根据 Ville不等式（1），在 M_t 曾经超过 $1/\alpha$ 的概率，在任意零假设分布下都不超过 α。因此，这些检验超鞅可直接构成有效检验：当 $M_t \geq 1/\alpha$. 时拒绝零假设。相应的拒绝时间自然就是 M_t的 $1/\alpha$首次到达时。这类检验具有博弈论解释，因为非负（超）鞅代表投注博弈中的财富过程 [RGVS22]。例如，$\mathfrak{P}$-检验鞅是赌徒的财富过程；这位赌徒下注反对这样的假设：序列 $\{X_t\}$是根据 $\mathfrak{P}$中的某个概率律抽取的。游戏的设计使得在零假设成立时，赌徒无法指望稳定地（从期望上看）赚钱；这一点通过以下限制来保证：对于$\mathfrak{P}$中的任意概率律，每一轮的预期财富倍数都不得超过 1。

然而，对于足够丰富的类 $\mathfrak{P}$，这样的博弈会让赌徒无计可施；上述限制对赌徒的约束如此之大，以至于他们唯一的选择就是完全不下注（或白白损失钱财）。Ramdas 等人在研究离散时间二元过程中的可交换性检验时，首次观察到了这一现象。[RRLK22], 他们证明，任何对于所有可交换二元概率律都是 NSM 的过程 $\{M_t\}$，几乎必然是严格递减过程（财富从 1 开始，此后只可能减少）。因此，任何基于对此类过程进行阈值化的检验，对于任何备择分布都不具检验功效。我们的首项技术贡献揭示了对数凹分布情形下的类似现象。具体而言，我们证明，较小的独立同分布高斯过程类无法用非负鞅（或 NSM）进行检验，因为所有此类过程都是平凡的，即几乎必然为常数（或递减）。下文概述了这一结论，其中 $\mathcal{G}^\infty$ 表示所有独立同分布高斯概率律的集合（均值和方差任意）。

定理。（非正式）不存在非平凡的 $\mathcal{G}^\infty$-*NSM* 或 $\mathcal{G}^\infty$-*NM*。更进一步，也不存在非平凡的 $\mathcal{L}^\infty$-*NSM* 或 $\mathcal{L}^\infty$-*NM*。

因此，对数凹　　密度代表了一类自然的分布，而这类分布无法通过鞅进行检验　　.

通过 E 过程进行检验。既然无法使用非负（超）鞅检验对数凹性（实际上也无法检验高斯性），我们便陷入了一种境地：序贯检验中占主导地位的设计范式失去了效用。对此可以采取两种截然不同的思路。

第一种思路是设计一个受限滤过 $\mathscr{G}_t$，使其区别于自然滤过，并且在该滤过下可能存在非平凡的检验超鞅。Ramdas 等人 [RRLK22；RGVS22] 指出了一个令人瞩目的事实：当原始滤过中不存在非平凡的（复合）检验鞅时，缩小滤过却可能引入新的非平凡（复合）检验鞅。Vovk 等人[VNG03；FGNV12] 就曾采用这种策略来构建可交换性的序贯检验；如上所述，在数据滤过下不存在非平凡的检验超鞅。这种方法有两个主要缺点。首先，这类检验鞅只能针对一组受限的停止时产生 e 过程（即受限滤过下的停止时）。从应用角度看，使用这类 e 过程需要实践者遵守纪律——他们不能查看原始数据来决定何时自适应地停止（像 $1/\alpha$ 的首次到达时这样的预定义停止规则是可以的，但它可能永远不会发生；在这种情况下，我们可能仍希望在停止时呈现已获得的证据）。其次，从设计角度看，构造合适的滤过本身就是一项微妙的任务，并且高度依赖具体问题，因此设计这类检验更像是一门艺术，而非科学。特别是，对于序贯对数凹性检验，目前还没有已知或显而易见的此类构造。

相比之下，我们采用通过 e 过程进行检验的备择策略。回想一下，若对每个停止时刻 $\{E_t\}$ 以及每个$\mathrm{P} \in \mathfrak{P}$，$\mathbb{E}_\mathrm{P}[E_\tau] \leq 1$. 过程 $\mathfrak{P}$ 满足以下条件，则称其为一组序贯概率律 τ的 e 过程。此类过程与前述检验鞅有着深刻的联系。事实上，有人指出，（可容许的）e 过程必定具有如下形式 $\inf_{\mathrm{P}\in\mathfrak{P}} M_t^\mathrm{P}$，其中每个 $\{M_t^\mathrm{P}\}$ 都是一个 P-非负鞅 [RRLK20]。同一观察也赋予 e 过程一种赌博解释：它是赌徒与一个「博弈族」对赌时的财富过程；赌徒同时参与针对每个 $\mathrm{P} \in \mathfrak{P}$的博弈，其财富取这些博弈中财富的最小值。只有当每个博弈都赢钱时，赌徒才能赢钱；也就是说，当 $\forall \mathrm{P} \in \mathfrak{P}$，$M_t^\mathrm{P}$增长至无界时，这就表明每个 $\mathrm{P} \in \mathfrak{P}$ 都可以被拒绝。

e-过程提供了一种 与前面讨论的检验超鞅类似的检验方法，具体如下

根据不等式（2）。确实，给定一个适用于 $\{E_t\}$ 的 $\mathfrak{P}$，我们可以构造一个 α-有效的成员资格检验，检验其是否属于 $\mathfrak{P}$，且仅在 $E_t \geq 1/\alpha$时才拒绝。确实，在这种情况下，拒绝时间为

$$\tau_\alpha := \inf\{t \geq 1 : E_t \geq 1/\alpha\},$$

并利用不等式（2），我们可以得出结论：

$$\forall \mathsf{P} \in \mathfrak{P}, \mathsf{P}(\tau_\alpha < \infty) \leq \alpha,$$

也就是说，对于复合原假设 $\mathfrak{P}$，该检验是有效的。还需注意，其有效性适用范围更广：设 σ 是关于数据自然滤过的任意其他停止时刻。我们还有 $\mathsf{P}(E_\sigma \geq 1/\alpha) \leq \alpha$，因此，只要仅在 $E_\sigma \geq 1/\alpha$ 时拒绝，任何额外的停止准则都不会影响该检验的有效性。

近期关于序贯分析的文献中，e-过程理论及其应用取得了长足进展（同时也发展了更基础的概念，即批处理情形下的 e变量）。这一概念因其灵活性和简洁性而颇具吸引力（尽管它是对非负鞅的推广），但构造具有高检验功效的 e-过程，既是一门科学，也是一门艺术 [RGVS22]。在复合检验中，e-过程至关重要，因为它们不会遇到 NSM 和非负鞅所面临的同类陷阱；事实上，即使在不存在此类 NSM 的类中，也确实存在非平凡的 e-过程。实际上，从某种意义上说，可以证明 e-过程处于序贯复合检验的核心 [RLKR22]。

1.3 使用通用似然比进行检验：一个简单的 e过程

通用推断策略 [WRB20] 在最大似然估计量易于计算时，可简单且通用地构造 e-过程。

为对这一方法提供背景说明，我们首先考虑点原假设和备择分布 P^∞ 以及 Q^∞的情形。在这种情况下，经典序贯检验理论认为，序贯似然比

$$L_t = \prod_{s=1}^{t} \frac{q(X_s)}{p(X_s)}$$

在 $1/\alpha$处进行阈值化后，可得到一个有效且具有较高检验功效的检验。事实上，在原假设下，$\{L_t\}$ 是一个 e过程，因为它是一个非负鞅。

对于简单原假设但复合备择假设的情形，通常会对上述似然比进行调整，以考虑各种可能的备择假设。一种方法是将上述分子替换为估计量 $\hat{}q_s(X_s)$。重要的是，只要该 $\hat{}q_s$ 是非预见的，即为 $\mathscr{F}_{t-1}$-可测的（仅取决于前 $t-1$ 个数据点），鞅性质就会继续成立。为突出这种非预见性，我们将这些估计量记为 $\hat{}q_{s-1}$。第二种方法是对备择分布进行混合，或许可以使用某种非信息性“先验”；但本文将采用第一种方法，因为我们处理的是高度非参数的备择假设（本质上是所有对数凹分布的补集，或其中希望对其具有检验功效的未指定子集）——可以很容易地使用核密度估计来构造 $\hat{}q_s$，但要对这样一个定义十分宽泛的非参数备择假设进行混合则不那么容易。

序贯通用似然比统计量（ULR）将上述方法推广到可计算最大似然估计量（MLE）的复合原假设。具体而言，该统计量如下：令 $\hat{}q_{t-1}$ 为任意可预测概率密度，即 $\hat{}q_{t-1}$ 可以表示为仅依赖于 $\{X_1, \ldots, X_{t-1}\}$ 和额外独立随机性的函数。与之前一样，我们应将 $\hat{}q$ 视为对底层分布p的估计。令 $\hat{}p_t$ 为基于数据 X_1^t在原假设类 $\mathcal{L}$ 上的最大似然估计量，即

$$\hat{p}_t = \arg\max_{\hat{p} \in \mathcal{L}} \sum_{s \leq t} \log \hat{p}(X_s).$$

注意，与 $\hat{}q_{t-1}$ 不同，最大似然估计 $\hat{}p_t$ 使用了 X_t。序贯 ULR 统计量是如下过程

$$R_t := \prod_{s \leq t} \frac{\hat{q}_{s-1}(X_s)}{\hat{p}_t(X_s)}.$$

（当然，如果分子只是 $\prod_{s\le t}\hat{q}_t(X_s)$，其中 $\hat{q}_t$ 是在更大的类上、使用 $\{X_1,\dots,X_t\}$ 计算得到的最大似然估计，那么我们就会得到通常的广义似然比过程。然而，我们将处理非常丰富的非参数备择假设，实际上几乎不可能在其中计算最大似然估计；此外，对于对数凹分布这类非正规模型，这种广义似然比表现极不稳定，也尚未得到充分理解。）

R_t 的效用主要源于它是一个 e过程。事实上，对于任意 $P\in\mathcal{L}$ 和 t，R_t 都被 $F_t(P)=\prod_{s\le t}\hat{q}_{s-1}(X_s)/p(X_s)$ 所支配。此外，由于 $\hat{q}$ 的可预测性，$\{F_t(P)\}$ 是一个以 1 为初值的 P^∞-鞅，因此对于任意停止时刻 τ，

$$\mathbb{E}_{P^\infty}[R_\tau]\le\mathbb{E}_{P^\infty}[F_\tau(P)]\le 1.$$

请注意，上述论证中，虽然 e过程被一个 P^∞-鞅所支配，但它本身并不是鞅。事实上，即使不存在这样的检验鞅，这一性质对于非平凡 e过程的存在也至关重要。我们指出，对于每个 P^∞，都被一个 -非负鞅支配 $P\in\mathcal{L}$ (或者一般而言，P-非负鞅支配$\mathrm{P}\in\mathfrak{P}$)这一性质，与 e过程性质本身等价，并可作为同一性质的另一种定义[RRLK20，引理 6]。

根据上述观察，在 $1/\alpha$. 处进行阈值化后，ULR e-过程便给出了一个有效检验。任何此类检验的检验功效取决于 $\hat{p}_t$ 和 $\{\hat{q}_s\}_{s\le t}$ 对底层分布 p的估计效果。事实上，我们将在 §4 中论证，若 $p\notin\mathcal{L}$，则 $\prod_{s\le t}p(X_s)/\hat{p}_t(X_s)$ 必定随 t呈指数增长。因此，只要序贯估计 $\hat{q}_t$ 在累积遗憾意义下能够很好地逼近 p，上述过程就必定相合。具体而言，将使用 进行预测的遗憾定义为 $\{\hat{q}_t\}$：

$$\rho_t(\hat{q};P):=\sum_{s\le t}(-\log\hat{q}_{s-1}(X_s))-\sum_{s\le t}(-\log p(X_s)),$$

这样，估计结果越好，遗憾就越低；并定义“易于估计”类

$$\mathcal{Q}(\hat{q}):=\{P:\rho_t(\hat{q};P)/t\to 0\ P\text{-a.s. as } t\nearrow\infty\}.$$

在第 4 节中，我们证明如下结果：

定理。（非正式表述）令 R_t 表示采用序贯估计量 $\mathscr{E}.$ 的 *ULR e*-过程。那么，当 $R_t\ge 1/\alpha$ 时拒绝的检验是 α-有效的，并且对 $\mathcal{Q}(\hat{q})$相合。

事实上，在 §4中，我们展示了上述结论的一个更精细版本，该版本允许 ρ_t 线性增长，但其增长速率受 p 与对数凹性的距离所限制。无论如何，我们指出，$\mathcal{Q}(\cdot)$ 这一类相当丰富。例如，使用筛估计量，可以在上述对数损失意义下，对诸如紧区间上具有光滑有界密度的概率律等非参数类实现低遗憾估计。因此，尽管没有任何检验鞅能够实现这些性质，ULR e-过程仍能针对丰富的备择集合，对数凹性进行有力检验。由此，我们的工作进一步阐明了对丰富复合原假设进行序贯检验的特点，以及在现代研究中 e-过程所应具有的首要地位。

除了上述渐近一致性结果外，我们还通过将 ULR e-过程的典型拒绝时间与备择概率律偏离对数凹性的海林格距离联系起来，推导出有限样本拒绝率界。具体而言，我们给出了针对单位盒上的利普希茨概率律和有界概率律的典型拒绝时间的显式界。我们还通过对一个由双高斯混合构成的简单参数族进行模拟研究，补充上述理论探究，以实证评估该检验的有效性和检验功效。我们发现，在低维情形 $d\le 3$ 下，该检验有效性极佳，且检验功效也较为理想。我们还利用这项模拟研究，突出说明估计量 $\hat{q}_t$ 的质量对检验功效的影响。

贡献概述 总而言之，本文关注对数凹性序贯检验的理论与方法学方面。我们首先给出一个否定性结果，表明构造检验（超）鞅的方法无法检验这一类概率律；同时，我们也由此简要刻画了独立同分布序贯概率律的叉凸包。从积极的一面来说，在不存在检验鞅的情况下，我们提出使用基于通用推断的 e过程来检验对数凹性。我们从理论上证明了由此得到的序贯检验具有一致性，并给出了在一大类备择分布下典型拒绝时间的具体自适应界，还通过模拟研究加以展示。

2 Defini分布及对数凹分布背景

我们先介绍对数凹分布的基本背景和必要的记号。有关更多细节，请参阅 Samuard 和 Wellner 的综述 [SW14]。

对数凹概率律。 若分布 P 在 $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d)$ 上，则称其为对数凹的（以下简称对数凹），当且仅当对于每一对紧集 A、B 以及 $\lambda \in (0,1)$,

$$P(\lambda A + (1-\lambda)B) \geq P(A)^{\lambda} P(B)^{1-\lambda},$$

$\lambda A + (1-\lambda)B$ 是闵可夫斯基和 $\{\lambda x + (1-\lambda)y : \ x \in A、\ y \in B\}$. 众所周知，对于相对于勒贝格测度具有密度的分布，当且仅当 $P(\mathrm{d}x) = e^{g(x)}\mathrm{d}x$ 由某个凹函数 g给出时，该分布是对数凹的。回顾一下， $\mathcal{L}$ 表示在 $\mathbb{R}^d$ 上具有密度的对数凹分布类，而$\mathcal{L}^{\infty}$ 表示 P^{∞} 的独立同分布序贯概率律集合， $P \in \mathcal{L}$。

对数凹 M 投影。 回顾一下， $\mathcal{D}$ 表示在 $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ 上、相对于勒贝格测度具有密度的概率律集合，并且

$$\mathcal{D}_1 := \{P \in \mathcal{D} : \mathbb{E}[\max(0, \log p(X))] < \infty, \mathbb{E}[\|X\|] < \infty\},$$

其中 $p(\cdot)$ 是 $P.$ 的密度。对于每个 $P \in \mathcal{D}_1$，都存在唯一的概率律

$$\mathfrak{L}_P := \arg\min_{L \in \mathcal{L}} \mathrm{KL}(P \| L),$$

其中 $\mathrm{KL}(\cdot\|\cdot)$ 是 KL 散度，称为其对数凹 M 投影。我们将稍作符号滥用，用 $\mathfrak{L}_p$ 表示 $\mathfrak{L}_P$ 的勒贝格密度（只要满足 $P \in \mathcal{D}_1$，就允许这样的密度存在）。对于一组点 $\{x_1^t\}, t \geq d+1$，对数凹最大似然估计量（最大似然估计）是经验分布 $\hat{}P_t = \sum_{s \leq t} \delta_{x_s}/t$ 的对数凹 M 投影，记为 P_t。我们最常提及的是其勒贝格密度 $\hat{}p_t$，它也可等价地定义为

$$\hat{p}_t := \underset{\substack{\log f \text{ is a concave function} \\ f \geq 0, \int f = 1}}{\arg\max} \sum_{s \leq t} \log f(x_s).$$

当 $X_1^t \overset{\text{i.i.d.}}{\sim} P$ d服从某个 $P \in \mathcal{D}_1$ 时，对数凹最大似然估计量具有极佳的理论性质。例如，$\hat{}p_t \to \mathfrak{L}_p$ 在强意义下满足 $\exists a > 0 : \ \int e^{a\|x\|} |\hat{}p_t(x) - \mathfrak{L}_p(x)| \to 0$ 几乎必然 [CS10]。

局部绝对连续序贯测度。 令 Γ 表示 $\mathbb{R}^d$ 上的标准高斯概率律，γ 表示其密度，并令 $\Gamma = \Gamma^{\infty}$. 。注意， Γ 是白噪声的概率律。若序贯概率律 P 相对于 Γ局部绝对连续 (l.a.c.)，记作 $\mathsf{P} \ll_{\text{loc.}} \Gamma$，则称 P 为局部绝对连续；这意味着对于所有 t，有限前缀的概率律 $\mathsf{P}|_t(\cdot) := \mathsf{P}(X_1^t \in \cdot)$ 相对于 $\Gamma|_t$绝对连续。此类局部绝对连续概率律存在一个密度过程，记作

$$Z_t^{\mathsf{P}} := \frac{\mathrm{d}\mathsf{P}|_t}{\mathrm{d}\Gamma|_t}.$$

例如，若对于某个具有勒贝格密度 p 的概率律 P，有 $\mathsf{P} = P^{\infty}$，则 $\mathsf{P} \ll_{\text{loc.}} \Gamma$，且 $Z_t^{\mathsf{P}} = \prod_{s \leq t} p(X_s)/\gamma(X_s)$。当然，我们可以通过指定密度过程来指定序贯概率律（它们是 $\ll_{\text{loc.}} \Gamma$）。注意，Z_t^{P} 是相对于 Γ 的似然比过程，因此是一个 Γ-鞅。此后，我们将使用 Γ 作为序贯概率律的参考测度，并且几乎完全在 $\ll_{\text{loc.}} \Gamma$的概率律下进行研究。

注意，如果 $Z_{t-1}^{\mathsf{P}} > 0$，那么对于 $\{X_t\} \sim \mathsf{P}$ $Z_t^{\mathsf{P}}/Z_{t-1}^{\mathsf{P}}$，它是给定 X_1^{t-1}. 时 X_t 的 Γ-条件密度。进一步，$Z_{t-1}^{\mathsf{P}} = 0 \implies Z_t^{\mathsf{P}} = 0$. 因此，对于任意适应过程 $\{M_t\}$，我们可以写成

$$Z_{t-1}^{\mathsf{P}} \mathbb{E}_{\mathsf{P}}[M_t(X_1^t)|\mathscr{F}_{t-1}] = Z_{t-1}^{\mathsf{P}} \mathbb{E}_{\mathsf{P}}[M_t(X_1^t)\mathbb{1}\{Z_t^{\mathsf{P}} > 0\}|\mathscr{F}_{t-1}] = \mathbb{E}_{\Gamma}[M_t Z_t^{\mathsf{P}} \mathbb{1}\{Z_t^{\mathsf{P}} > 0\}|\mathscr{F}_{t-1}] = \mathbb{E}_{\Gamma}[M_t Z_t^{\mathsf{P}}|\mathscr{F}_{t-1}].$$

由此可见，过程 $\{M_t\}$ 是 P-非负超鞅，当且仅当 $\{Z_t^{\mathsf{P}} M_t\}$ 是一个 Γ-NSM。事实上，若前者成立，则由上文可得 $\mathbb{E}_{\Gamma}[M_t Z_t^{\mathsf{P}}] \leq Z_{t-1}^{\mathsf{P}} M_{t-1}$；若后者成立，则可得

$Z_{t-1}^{\mathsf{P}} \mathbb{E}_{\mathsf{P}}[M_t|\mathscr{F}_{t-1}] \leq Z_{t-1}^{\mathsf{P}} M_{t-1} \iff \mathbb{1}\{Z_{t-1}^{\mathsf{P}} > 0\} \mathbb{E}_{\mathsf{P}}[M_t|\mathscr{F}_{t-1}] \leq \mathbb{1}\{Z_{t-1}^{\mathsf{P}} > 0\} M_{t-1}.$。由于 $\mathbb{1}\{Z_{t-1}^{\mathsf{P}} > 0\}$ 几乎处处在 P 下成立，因此 $\mathbb{E}_{\mathsf{P}}[M_t|\mathscr{F}_{t-1}] \leq M_{t-1}$ 几乎处处在 P 下成立，所以 M_t 是一个 P-非负超鞅。在上述分析中保持等式不变，也可得到关于非负鞅的类似结论。这些事实对我们后续研究叉凸包非常有用。

3 存在 对数凹性不存在非平凡检验超鞅

我们首先定义一个自然的平凡性概念。

定义 若一个 *NSM* $\{M_t\}$ 满足以下条件，则称其为平凡的 $\Gamma(\exists t: M_t > M_{t-1}) = 0$. 若一个 *NM* $\{M_t\}$ 满足$\Gamma(\exists t: M_t \neq M_{t-1}) = 0.$，则称其为平凡的

换言之，一个 NSM（NM）若几乎必然是一个非递增（常数）过程，则称其为平凡的。本节余下部分中，我们令 $\{\mathscr{F}_t\}$ 为自然滤过。我们回顾针对一类概率律 $\mathfrak{P}$的检验超鞅概念，并将其简称为非负超鞅。

定义 对于一组序贯概率律 $\mathfrak{P}$，若过程 $\{M_t\}$满足：对每个$\mathrm{P} \in \mathfrak{P}$， $\{M_t\}$ 都是一个 P-*NSM*，则称其为 $\mathfrak{P}$-NSM 。类似地，若对每个 $\mathrm{P} \in \mathfrak{P}.$， $\{M_t\}$ 都是一个 P-*NM*，则称 $\{M_t\}$ 为 $\mathfrak{P}$-*NM*。

有了这些定义，我们给出本节的主要结果，其证明留待 §3.3。

定理 1。 在自然滤过下，不存在非平凡的 $\mathcal{G}^\infty$-*NSM* 或 $\mathcal{G}^\infty$-非负鞅；由此更可知，在自然滤过下，不存在非平凡的 $\mathcal{L}^\infty$-*NSM* 或 $\mathcal{L}^\infty$- 非负鞅。

正如 Ramdas 等人所讨论的 [RRLK22], 上述结果意味着，任何基于阈值化 α -NSM 或 $\mathcal{L}^\infty$-非负鞅的、有效水平为 $\mathcal{L}^\infty$ 的对数凹性序贯检验都必然没有检验功效。事实上，在前一种情况下，对于任何相对于 Γ 局部绝对连续的概率律，这种检验几乎必然永远不会超过其初始值，因此几乎必然永远不会拒绝。

证明思路。 该结果源于反证法。为说明这一点，假设 $\{M_t\}$ 是一个 $\mathcal{G}^\infty$-NSM，并且对于某个时刻 t，给定 $\mathscr{F}_{t-1}$，它在事件 $\{X_t \in O\}$ 上递增，其中 $\{X_t \in O\}$ 是某个开球 O，即在给定 $X_1^{t-1} = x_1^{t-1}$的条件下，$\{X_t \in O\} \subset \{M_t > M_{t-1}\}$。注意，由于非负性，它在球外最坏也可能为零。现在，考虑一个方差足够小的高斯分布 G_O，使得 $G_O(O) \approx 1.$。通过调整这个方差，可以确保给定历史时， $M_t > M_{t-1}$ 发生的概率任意接近 1；又由于球外的 M_t 降幅有界，这就确保了 M_t 的条件期望严格增加。由于这违反了相对于 $G_B^\infty \in \mathcal{G}^\infty$的超鞅性质，我们必须得出结论：不存在这样的球 O。

当然， $\{M_t\}$ 所增加的集合未必包含任何球，但其测度仍可能非平凡，更不用说该集合还可能以复杂的方式随历史变化。我们通过利用叉凸性的概念来弥补这些缺口 [RRLK22] 它是凸性的序贯类比，尤其与（超）鞅性质密切相关，并将在下一节中讨论。特别地，任意过程$\{M_t\}$ 若为 $\mathcal{G}^\infty$-NSM（或 NM），则对于 $\mathcal{G}^\infty$的“叉凸包”中的任意序贯概率律，它也同样是 NSM（或 NM）。随后，主要论证将证明， $\mathcal{G}^\infty$ 的叉凸包极其丰富，并包含具有密度的任意独立过程的概率律（即，具有联合独立的过程 $\{X_t\}$，满足 $X_t \sim p_t \in \mathcal{D}$）。这一庞大的概率律集合完全阻碍了 NSM（或 NM）性质以任何非平凡方式成立，其本质上采用了前述直观示例的稳健版本。概括而言，我们将沿以下路径建立这一结果，其中每个蕴含关系的正向方向都利用叉凸组合（反向方向则是显然的）。

$$\mathcal{G}^\infty\text{-NSM} \Longleftrightarrow \mathcal{G}_*^\infty\text{-NSM} \Longleftrightarrow (\otimes \mathcal{G}_*)\text{-NSM} \Longleftrightarrow (\otimes \overline{\mathcal{G}_*})\text{-NSM} \Longleftrightarrow (\otimes \mathcal{D})\text{-NSM} \Longleftrightarrow \text{Trivial}$$

图1：论证的示意图。 $\mathcal{G}_*$ 是所有有限高斯混合的集合，而 $\overline{\mathcal{G}_*}$ 表示其L_1 闭包。对于任意集合 $\mathcal{P}$，类 $\bigotimes \mathcal{P}$ 由边缘分布属于 $\mathcal{P}$的独立序贯分布律组成。定义见§3.2。

3.1 分叉凸组合

从算法意义上说，对于两个概率律 P、Q， α-凸组合 $R = \alpha P + (1-\alpha)Q$ 是以下过程的输出所服从的概率律：独立采样 $U \sim P$ 和 $V \sim Q$，然后根据独立抛掷一枚 α 硬币的结果，输出 $X = U$ 或 V。叉凸组合是这种过程的自然序贯扩展

。具体而言，我们采样两条轨迹 $\{U_t\} \sim \mathrm{P}$ 和 $\{V_t\} \sim \mathrm{Q}$，释放 $X_t = U_t$，持续 $t \le s$ 一段时间 s，然后抛掷一枚 h 硬币（其中 h 可以取决于历史），以决定后续尾部是 $X_t = U_t$ 还是 $X_t = V_t$，持续 $t > s$。请注意，这比凸组合的概念丰富得多：首先，只需针对输出序列的尾部决定释放 U_t 还是 V_t；其次，混合比例可以取决于历史。形式化定义如下。

定义 ([RRLK22]) 设 $\mathtt{P}$、$\mathtt{Q} \ll_{\mathrm{loc.}} \Gamma$ 为序贯概率律。设 $s \in \mathbb{N}$，并设 $h \in [0,1]$ 为一个 $\mathscr{F}_s$-可测随机变量，使得 $\Gamma(h < 1,\ Z_s^{\mathtt{Q}} = 0) = 0$. 那么，$(s,\ h)$- 叉凸组合将 $\mathtt{P}$ 与 $\mathtt{Q}$ 组合成序贯概率律 $\mathtt{R}$，其密度过程为

$$Z_t^{\mathtt{R}} := Z_t^{\mathtt{P}} \mathbb{1}\{t \le s\} + \left(h Z_t^{\mathtt{P}} + (1-h) Z_t^{\mathtt{Q}} \frac{Z_s^{\mathtt{P}}}{Z_s^{\mathtt{Q}}} \right) \mathbb{1}\{t > s\}.$$

我们将其简洁地记作 $\mathtt{R} = \left[\mathtt{P} \xrightarrow{s\ h}, \mathtt{Q}\right]$.

注意，叉凸组合在概率意义上允许从概率律 $\mathtt{P}$ 切换到 $\mathtt{Q}$，且至多发生一次数据依赖变点，即“切换”。比率 $Z_s^{\mathtt{P}}/Z_s^{\mathtt{Q}}$ 反映了这样一个事实：发生切换时，直到时间 s 为止的前缀是根据 $\mathtt{P}$ 抽取的；而对 h 的条件则确保切换到 $\mathtt{Q}$ 时 $Z_s^{\mathtt{Q}} \neq 0$（非正式地说，即初始数据段在 $\mathtt{Q}$ 下并非不可能）。

上述定义的重要性在于，叉凸组合保留（超）鞅性质。回顾 §2 可知，$\{M_t\}$ 是一个 $\mathtt{P}$-NSM，当且仅当 $\{Z_t^{\mathtt{P}} M_t\}$ 是一个 Γ-NSM。现在假设 $\{M_t\}$ 同时是 $\mathtt{P}$-NSM 和 $\mathtt{Q}$-NSM，令 $\mathtt{R}$ 为 $(s,\ h)$- 叉凸组合，由 $\mathtt{P}$ 和 $\mathtt{Q}$ 组成。对于 $t \ge s+1$，我们有

$$\mathbb{E}_\Gamma[Z_t^{\mathtt{R}} M_t | \mathscr{F}_{t-1}] = h \mathbb{E}_\Gamma[Z_t^{\mathtt{P}} M_t | \mathscr{F}_{t-1}] + (1-h) \frac{Z_s^{\mathtt{P}}}{Z_s^{\mathtt{Q}}} \mathbb{E}_\Gamma[Z_t^{\mathtt{Q}} M_t] \le h Z_{t-1}^{\mathtt{P}} M_{t-1} + (1-h) \frac{Z_s^{\mathtt{P}}}{Z_s^{\mathtt{Q}}} Z_{t-1}^{\mathtt{Q}} M_{t-1} = Z_{t-1}^{\mathtt{R}} M_{t-1},$$

这里我们利用了这样一个事实：h 和 $Z_s^{\cdot}$ 是 $\mathscr{F}_{t-1}$-可测的。对于 $t \le s$，同样的计算显然成立，鞅性质也可类似地推出。该性质远不止适用于有限组合，还可推广到闭叉凸包；闭叉凸包是集合的标准闭凸包概念的推广。

定义 ([RRLK22])如果一个集合包含其元素的所有叉凸组合，就称其为叉凸集。设 $\mathfrak{P}$ 是一组相对于 Γ局部绝对连续的序贯概率律。 $\mathfrak{P}$ 的叉凸包，记作 $\text{f-conv}(\mathfrak{P})$， 是所有包含 $\mathfrak{P}$ 的叉凸集的交集。 $\mathfrak{P}$ 的闭叉凸包，记作 $\text{f-conv}(\mathfrak{P})$， 是其叉凸包~~在每个~~固定时刻 $L_1(\Gamma)$ 似然比过程收敛意义下的闭包 t。

具体而言，定义中的闭包包含所有满足以下条件的过程 $\mathtt{Q}$：存在一个序列 $\mathtt{Q}_n$，其密度过程为 $\{Z_t^{\mathtt{Q}_n}\}$，且 $\forall t,\ Z_t^{\mathtt{Q}_n} \to Z_t^{\mathtt{Q}}$ 在 $L_1(\Gamma)$中。我们将其称为局部 $L_1(\Gamma)$ 闭包。该闭包赋予闭叉凸包极大的灵活性，使这一概念成为强有力的工具，这源于 Ramdas 等人观察到的以下现象：[RRLK22，定理 13]，其论证将在下文重现。

命题 2. 对于一组序贯概率律 $\mathfrak{P}$，一个过程是 $\mathfrak{P}$-*NSM*，当且仅当它是 f-凸包($\mathfrak{P}$)-*NSM*。

证明。 叉凸包的情形可由前面的两点计算直接得出。下面说明这一结论如何推广到闭包。设 $\{M_t\}$ 为所讨论的过程，并假设按照上述意义，对于 $\mathtt{P}_n \in$ f-凸包($\mathfrak{P}$) 中的 $\mathtt{P}$，有 $\mathtt{P}_n \to \mathtt{P}$。令 $Z_t^n := Z_t^{\mathtt{P}_n}$ 且 $Z_t := Z_t^{\mathtt{P}}$.。我们知道，对于每个 t，$Z_t^n \to Z_t$ 都属于 $L_1(\Gamma)$。我们需要证明 $Z_t M_t$ 是一个 Γ-NSM。为此，固定一个 t，并通过转到一个子空间，假设 $Z_t^n \to Z_t$ 和$Z_{t-1}^n \to Z_{t-1}$ 几乎处处逐点成立。现在，由于 $Z_t^n M_t$ 是一个 Γ-鞅，利用法图引理可得

$$\mathbb{E}_\Gamma[Z_t M_t | \mathscr{F}_{t-1}] = \mathbb{E}_\Gamma[\liminf Z_t^n M_t | \mathscr{F}_{t-1}] \le \liminf \mathbb{E}_\Gamma[Z_t^n M_t | \mathscr{F}_{t-1}] \le \liminf Z_{t-1}^n M_{t-1} = Z_{t-1} M_{t-1}. \quad \square$$

值得注意的是，尽管上文中的 NSM 性质在闭包下保持不变，但鞅性质未必如此，这是因为上述证明在处理闭包时用到了法图引理。尽管如此，非负鞅（事实上，即使不借助非负性，鞅性质也）在叉凸包下仍然成立，无需取闭包，因此我们得到以下刻画。

命题 3。F 对于一组序贯概率律 $\mathfrak{P}$，一个过程是 $\mathfrak{P}$-非负鞅，当且仅当它是 f-凸包($\mathfrak{P}$)-非负鞅 .

3.2 独立序贯概率律的叉凸包

命题 2 为证明 $\mathcal{G}^\infty$-非负超鞅是平凡的提供了一种具体的攻击方法：我们将证明，该集合的叉凸包过于丰富，因而不可能存在非平凡非负超鞅。我们的论证主体将建立独立序贯概率律的叉凸包的简单结构刻画。本节将通过三个性质来描述这种刻画，证明留待 §A.2。我们先给出一个关键定义，为集合中的“独立序贯概率律”设定记号。

定义设 $\mathcal{P}$ 是 $\mathbb{R}^d$ 上的一个分布集合。对于分布序列 $\{P_t\}_{t\in\mathbb{N}}$，我们将 $\bigotimes\{P_t\}$ 定义为随机过程 $\{X_t\}_{t\in\mathbb{N}}$ 的序贯分布，其中所有 X_t 均联合独立，且对每个 $t\in\mathbb{N}$， $X_t\sim P_t$. 我们进一步定义 $\bigotimes\mathcal{P}:=\{\bigotimes\{P_t\}:\ P_t\in\mathcal{P}\ \forall t\}$，即各时刻的概率律均属于 $\mathcal{P}$. 的独立随机过程的概率律集合

注意， $\bigotimes\mathcal{P}$ 是一个比独立同分布序贯概率律丰富得多的集合，我们将后者记为 $\mathcal{P}^\infty:=\{P^\infty:\ P\in\mathcal{P}\}$. 鉴于此，以下结果展示了叉凸包的丰富性。回顾一下，如果一个概率律集合中的任意两个概率律都相互绝对连续，则称该集合相互绝对连续（相对于测度绝对连续）。

引理 4。 *Let $\mathcal{P}\subset\mathcal{D}$ 设 $\mathbb{R}^d$ 上具有密度的概率律构成一个最大绝对连续集。那么，* $\overline{\text{f-凸包}}(\mathcal{P}^\infty)\supset\bigotimes\mathcal{P}\supset\mathcal{P}^\infty$.

为概述上述论证的依据，固定任意 $\mathrm{P}=\bigotimes\{P_t\}$. 。只需证明存在一个概率律序列 $\{\mathrm{R}^T\}_{T\in\mathbb{N}}$，其中每个概率律均由 $\mathcal{P}^\infty$-概率律的有限叉凸组合（及其叉凸组合）生成，并且对于 $t\le T$， R^T 与 P 的密度过程一致。由于 $\mathrm{R}^T\to\mathrm{P}$ 在适当意义下成立，根据闭包即可得到结论。上述引理的具体见证序列如下

$$\mathtt{R}^1:=P_1^\infty,\mathtt{R}^T=\left[\mathtt{R}^{T-1}\overset{T-1,0}{\longrightarrow}P_T^\infty\right],$$

其中每个叉凸组合都是有效的，因为 $\mathcal{P}$ 是最大绝对连续的。本质上，这利用了叉凸组合允许我们在选定的某个时刻之后在不同概率律之间进行切换这一事实。详见 A.2 。

接下来，我们利用叉凸组合的凸组合性质来证明：独立同分布概率律的叉凸组合还包括关于混合的独立同分布乘积。为此，我们定义如下混合类。

定义 令 $\mathcal{P}$ 为定义在 $\mathbb{R}^d$上的分布集合。对于 $k\in\mathbb{N}$，令 $\mathcal{P}_k$ 为由 $\mathcal{P}$中的概率律进行 k重混合所形成的概率律类，并将 $\mathcal{P}_*=\bigcup_{k\in\mathbb{N}}\mathcal{P}_k$ 记为由 $\mathcal{P}$中的概率律进行有限混合所形成的概率律类。

注意， $\mathcal{P}_*$ 是良定义的，因为 $\mathcal{P}_k$ 构成一个递增集合。§A.2 中展示的第二个关键结果是

引理 5。 令 $\mathcal{P}\subset\mathcal{D}$ 为定义在 $\mathbb{R}^d$上的相对于测度绝对连续集。则 $\overline{\text{f-凸包}}(\mathcal{P}^\infty)\supset\mathcal{P}_*^\infty$。

上述结果所依据的关键观察，在证明 $\text{f-凸包}(\mathcal{P}^\infty)\supset\mathcal{P}_2^\infty$. 时已经得到说明。为此，固定任意的 P、$Q\in\mathcal{P}$和 $\alpha\in[0,1]$。我们需要构造一个通过反复进行叉凸组合得到的概率律序列 R^T， 使其与 R 的密度过程相匹配： $=(\alpha P+(1-\alpha)Q)^\infty$ 对于不超过T的时间均成立。具体构造如下：

$$\mathtt{R}^0:=P^\infty,\mathtt{S}^T:=\left[\mathtt{R}^{T-1}\overset{T-1,\alpha}{\longrightarrow}Q^\infty\right],\mathtt{R}^T:=\left[\mathtt{S}^T\overset{T,0}{\longrightarrow}P^\infty\right].$$

上文中，通过适当地混合 R^{T-1} （其尾部表现为 P^∞）与 Q^∞， S^T 的密度过程在时刻 T 之前与 R 相匹配。随后， R^T 将 S^T 的尾部切换为表现为 P^∞， 以便进行递归。通过对 k 归纳，此论证可推广到任意 k 的 $\mathcal{P}_k^\infty$（这是可行的，因为 $\mathcal{P}_k$ 中的任一元素都是一个 $\mathcal{P}_{k-1}$ 概率律与一个 $\mathcal{P}$ 概率律的混合）。由于 k 是任意的，这就立即推广到 $\mathcal{P}_*$.

最后，我们利用叉凸包在 $L_1(\Gamma)$ 下的闭包性质，将一组上的乘积测度所生成的叉凸包推广到该集合闭包上的乘积测度。

引理 6. 设 $\mathcal{P}$ 是 $\mathbb{R}^d$ 上具有密度的分布集合。那么 $\text{f-凸包}(\bigotimes\mathcal{P})\supset\bigotimes\overline{\mathcal{P}}$，其中 $\overline{\mathcal{P}}$ 是 $L_1(\Gamma)$-闭包 $\mathcal{P}$。

上述部分 上述引理是根据 § 中详述的闭包性质直接得出的A .2.

3.3 非平凡检验鞅不存在的证明

上一节表明，取闭叉凸包可以显著扩展序列上的乘积分布。本节利用这些性质证明 $\mathcal{G}^{\infty}$-NSM 的平凡性。其背后的关键观察是关于高斯混合丰富性的以下标准事实。回想一下，$\overline{\mathcal{P}}$ 表示 $\mathcal{P}$ 的 $L_1(\Gamma)$-闭包。

L引理 7。 $\mathcal{G}_*$ 在所有具有密度的分布集合中是 $L_1(\Gamma)$-稠密的，即 $\overline{\mathcal{G}_*} = \mathcal{D}$。

高斯混合在 $L_1(\mathrm{Leb})$中的稠密性是一个经典结果；例如参见 Alspach 和 Sorenson 的工作 $\mathcal{D}$ [AS72] ，或 Lo 的工作 [Lo72]。最近，Bacharoglu 给出了一个适用范围宽得多的结果 [Bac10], ，证明高斯混合在非负简单函数中同时按 anL_1 和 an L_∞ 意义稠密。这也足以满足我们的需要，因为非负简单函数本身在非负可积函数中是 L_1- 稠密的。由于 $L_1(\Gamma)$具有关于勒贝格测度一致有界的密度，因此它在 Γ 中的稠密性也随之成立。

我们注意到，我们的论证可推广到任何这样的集合，也就是说，推广到任何满足 $\overline{\mathcal{P}} = \mathcal{D}$ 的 $\mathcal{P}$。高斯分布在 $\mathcal{L}$ 中可作为一个方便的例子，并且该性质在此成立。有了这一点，我们便按如下方式继续。

定理 1 的证明。设 $\{M_t\}$ 是一个 $\mathcal{G}^{\infty}$-NSM。首先，由引理 5 和命题 2 可知，因此 $\{M_t\}$ 也是一个 $\mathcal{G}_*^{\infty}$-NSM。接着，由引理 4 和命题 2 可知，它还是一个 $(\bigotimes \mathcal{G}_*)$-NSM。同理，由引理 6 和命题 2 可得， $\{M_t\}$ 也是一个 $(\bigotimes \overline{\mathcal{G}_*})$-NSM。最后，由引理 7 可得， $\{M_t\}$ 是一个 $(\bigotimes \mathcal{D})$-NSM。[1]

现在我们来论证， $\bigotimes \mathcal{D}$ 过于丰富，无法容纳非平凡非负超鞅。我们采用反证法——假设对于某个 t ，$M_t > M_{t-1}$ 以非零概率成立，并据此构造一个属于 $\bigotimes \mathcal{D}$ 、违反 NSM 性质的概率律。论证反复利用 $(\mathbb{R}^d)^t$ 与 $\mathbb{R}^{dt}$ 在乘积拓扑和度量拓扑下的拓扑等价。我们将 m 维空间中的勒贝格测度记作 Leb_m，并注意到 $(\mathbb{R}^d)^t$ 上的乘积勒贝格测度与 Leb_{dt} 相同；我们用后者表示前者。

接下来继续论证。对于自然数 t，定义事件 $\mathsf{A}_t := \{M_t > M_{t-1}, M_{t-1} < \infty\}$，也就是说， $\{M_t\}$ 在时刻 t 增加。只需证明无论 t如何，A_t 的测度都为零，因为由于 M_{t-1}的可积性，$\Gamma(M_{t-1} = \infty)$ 必须为零。为导出矛盾，假设 $\Gamma(\mathsf{A}_t) > 0$. 对于 $n \in \mathbb{N}$，定义近似集 $\mathsf{A}_t^n := \{M_t \geq M_{t-1} + 1/n, M_{t-1} \leq n\}$. 。集合 A_t^n构成一个递增的序列，并收敛到 $\{M_t > M_{t-1}, M_{t-1} < \infty\} = \mathsf{A}_t$.

现在，由于 $\Gamma(\mathsf{A}_t) > 0$ 且 $\mathsf{A}_t \in \mathscr{F}_t$，根据高斯分布与欧几里得空间上的勒贝格测度彼此绝对连续，可得 $\mathrm{Leb}_{dt}(\mathsf{A}_t) > 0$ 。不失一般性，我们可以假设 $\mathrm{Leb}_{dt}(\mathsf{A}_t) < \infty$ （否则，利用勒贝格测度的 σ-有限性，可以取 A_t 的一个子集，使其测度为正且有限，并在该子集上进行论证）。由于 $\mathsf{A}_t^n \nearrow \mathsf{A}_t$，根据测度的正则性，有 $\mathrm{Leb}_{dt}(\mathsf{A}_t^n) \to \mathrm{Leb}_{dt}(\mathsf{A}_t)$，特别地，存在一个 n ，使得 $\mathrm{Leb}_{dt}(\mathsf{A}_t^n) \in (0, \infty)$。在余下的论证中固定这个 n 。

回忆一下，在 $\mathbb{R}^m$ 中，开矩形是开区间的笛卡儿积，即形如$\bigtimes_{i=1}^m (a_i, b_i)$ 的集合，其中 $a_i < b_i$. 类似地，我们称 R 为 $(\mathbb{R}^d)^t$ 中的开矩形， $\mathbb{R}^d$-矩形$S_1 \dots S_t$ 满足 $R = \bigtimes_{s=1}^t S_s$。下述结论是欧几里得空间的基本拓扑性质和测度论性质的推论，我们将在 §A.3 中证明。

引理 8。 设 $E \subset (\mathbb{R}^d)^t$ 满足 $\mathrm{Leb}_{dt}(E) > 0$. 。对于每个自然数 $m \in \mathbb{N}$，都存在一个位于 $(\mathbb{R}^d)^t$ 中的开矩形 R ，使得

$$\mathrm{Leb}_{dt}(R) > 0 \quad \textit{and} \quad \mathrm{Leb}_{dt}(R \cap E) \geq \frac{m}{m+1} \mathrm{Leb}_{dt}(R).$$

利用上述结果，我们可以构造一列矩形，其位于 $(\mathbb{R}^d)^t, \{R_m\}_{m \in \mathbb{N}}$ ，且每个矩形都具有正测度，使得

$$\frac{\mathrm{Leb}_{dt}(\mathsf{A}_t^n \cap R_m)}{\mathrm{Leb}_{dt}(R_m)} \geq \frac{m}{m+1}.$$

[1]我们也可以更直接地论证：注意到，取闭叉凸包是一个幂等运算，即f-凸包(f-凸包(𝔓)) = f-凸包(𝔓) (这可由以下事实推出：闭叉凸包是叉凸的，且闭集的闭包保持不变)。因此，利用 §3.2 中的一系列引理，f-凸包($\mathcal{G}^{\infty}$) $\supset \bigotimes \overline{\mathcal{G}}_*$，因此 $\{M_t\}$ 是一个 $(\bigotimes \mathcal{D})$-NSM。

现在，由于每个 R_m 都是一个矩形，因此存在一个概率律 $\mathsf{D}_m \in \bigotimes \mathcal{D}$，使得前缀限制 $\mathsf{D}_m|_t$ =均匀分布(R_m)。事实上，如果 $R_m = \bigtimes_{s=1}^{t} S_s^m$，那么 $\mathsf{D}_m = \bigotimes\{D_s^m\}$，其中 D_s^m = 均匀分布(S_s^m)，对于 $s \le t$，且 $D_s^m = \Gamma$ 对于$s > t$。我们声称，当 m 足够大时，D_m 可以证明 $\{M_t\}$ 不满足 NSM 性质。我们通过以下过程来说明这一点：$\{N_t\} := \{\min(M_t, n+1)\}$。

注意，如果 $\{M_t\}$ 是一个 P-非负超鞅，那么 $\{N_t\}$ 也是，因为

$$\mathbb{E}[N_t|\mathscr{F}_{t-1}] \le \min(\mathbb{E}[M_t|\mathscr{F}_{t-1}], \mathbb{E}[n+1|\mathscr{F}_{t-1}]) = \min(M_{t-1}, n+1) = N_{t-1},$$

由于 M_t 和 $n+1$ 都是非负的，因此非负性成立。此外，由于在 A_t^n 上有 $M_{t-1} \le n$，所以在 A_t^n 上同样有 $N_t \ge N_{t-1} + 1/n$，因为 $n + 1/n \le n + 1$.

因此，我们有

$$\begin{aligned}\mathbb{E}_{\mathsf{D}_m}[N_t] &\ge \mathbb{E}_{\mathsf{D}_m}\left[(N_{t-1} + 1/n)\mathbb{1}\{X_1^t \in \mathsf{A}_t^n\}\right] + 0 \\ &= \mathbb{E}_{\mathsf{D}_m}\left[N_{t-1}\mathbb{1}\{X_1^t \in \mathsf{A}_t^n\}\right] + \frac{\mathsf{D}_m(\mathsf{A}_t^n)}{n} \ge \mathbb{E}_{\mathsf{D}_m}\left[N_{t-1}\mathbb{1}\{X_1^t \in \mathsf{A}_t^n\}\right] + \frac{m}{n(m+1)},\end{aligned}$$

其中，最后一个不等式利用了以下事实：R_m 的测度中至少有 $m/(m+1)$ 的比例位于 A_t^n 中；同时我们使用了 N_t. 的非负性

然而，由于 N_{t-1} 的上界为 $n+1$，我们注意到

$$0 \le \mathbb{E}_{\mathsf{D}_m}[N_{t-1}\mathbb{1}\{X_1^t \in (\mathsf{A}_t^n)^c\}] \le (n+1)\mathsf{D}_m((\mathsf{A}_t^n)^c) \le \frac{n+1}{m+1},$$

因此

$$\mathbb{E}_{\mathsf{D}_m}[N_{t-1}\mathbb{1}\{X_1^t \in \mathsf{A}_t^n\}] = \mathbb{E}_{\mathsf{D}_m}[N_{t-1}] - \mathbb{E}_{\mathsf{D}^m}[N_{t-1}\mathbb{1}\{X_1^t \in (\mathsf{A}_t^n)^c] \ge \mathbb{E}_{\mathsf{D}_m}[N_{t-1}] - \frac{n+1}{(m+1)}.$$

但现在，我们得出

$$\mathbb{E}_{\mathsf{D}_m}[N_t] \ge \mathbb{E}_{\mathsf{D}_m}[N_{t-1}] + \frac{(m/n) - (n+1)}{m+1}.$$

取 $m > 3n^2$，并利用 $n \ge 1$，可得 $\mathbb{E}_{\mathsf{D}_m}[N_t] > \mathbb{E}_{\mathsf{D}_m}[N_{t-1}]$，这与 $\{N_t\}$ 在 D_m 下的超鞅性质相矛盾（因为超鞅的均值序列必定非增）。由此我们得出，$\Gamma(\mathsf{A}_t)$ 不可能成立，即 $M_t \le M_{t-1}$ Γ-几乎必然成立。

但是，由于 t 是任意的，我们立即得出

$$\Gamma(\exists t \ge 2 : M_t > M_{t-1}) \le \sum_{t \ge 2} \Gamma(M_t > M_{t-1}) = 0.$$

关于非负鞅的论证也可由此得出。若 $\{M_t\}$ 是一个 $\bigotimes\mathcal{D}$-非负鞅，那么它也是一个 NSM，因此几乎必然不会增加。但这意味着 $M_1 - M_t \ge 0$ 也是一个非负超鞅，因而不会增加，这就表明 M_t 几乎必然也不会减少。 □

注。 或许可以构造另一种论证，而不必显式地借助叉凸包这一概念。也许可以直接使用上文的 A_s^n，并用足够窄的高斯分布 G_m 替换 D_m a，使得 $\mathsf{G}_m(A_s^n \cap R) \approx \mathsf{G}_m(R) \approx 1$。不过，仍有足够多的技术细节需要厘清，因此这种方法未必更简短或更清晰。更重要的是，我们上述选择的论述路径能够更丰富地刻画具有密度的独立同分布过程的叉凸包，并进一步直接展示这种刻画的实用价值。因此，它加深了我们对叉凸性这一重要几何概念的理解。

4 序贯通用似然比 E-过程 S

我们首先回顾引言中给出的 e-过程定义。

定义 *A* n $\{\mathscr{F}_t\}$-适应过程 $\{E_t\}$ 被称为一组序贯概率律 $\mathfrak{P}$ 的 *e*-过程 *if*

$$\sup_{\mathrm{P}\in\mathfrak{P}}\sup_{\tau}\mathbb{E}[E_\tau]\le 1,$$

其中第二个上确界取遍所有停止时。进一步，如果对于某个 $n\ge 1$，相对于所有 $\mathrm{P}\in\mathfrak{P}$ 几乎处处都有 $E_1=E_2=\cdots=E_{n-1}=1$，那么我们称 E_t 是从时刻 n 开始、针对 $\mathfrak{P}$ 的一个 *e*过程。

接下来，我们定义通用似然比（ULR）过程 [WRB20], ，它是本节关注的主要对象。

定义 令 $\mathscr{E}$ 表示一列估计量 $\{\mathscr{E}_t\}_{t\ge 0}$，使得每个 $\mathscr{E}_t:(\mathbb{R}^d)^t\to\mathcal{D}$。在任意 t，记 $\hat{}q_t=\mathscr{E}_t(X_1,\dots,X_t)$。最后，令 $\hat{}p_t$ 表示基于数据 $X_1,\dots,X_t$ 的对数凹最大似然估计（当 $t>d$ 时该估计存在）。*ULR* 过程是统计量

$$R_t(X_1^t;\mathscr{E}):=\mathbb{1}\{t\le d\}+\mathbb{1}\{t>d\}\prod_{d+1\le s\le t}\frac{\hat{q}_{s-1}(X_s)}{\hat{p}_t(X_s)}.$$

我们通常省略 R_t 对 X_1^t 和 $\mathscr{E}$ 的依赖关系。将 $R_t=1$ 针对 $t\le d$ 的初始设定如此规定，是为了考虑到这样一个事实：只有在至少有 $d+1$ 个样本可用时，对数凹最大似然估计才已知存在。

如引言中所述，由于 $\hat{}q_{t-1}$ 的可预测性，且它们是概率密度，$\{R_t\}$ 构成一个 e过程。我们将 R_t 的有效性正式表述为一个命题。

命题 9. 对于任意 $\mathscr{E}$，过程 $\{R_t\}$ 是一个从时间 $d+1$开始的 $\mathcal{L}^\infty$ 的 *e*过程。因此，当 $R_t\ge 1/\alpha$ 时拒绝零假设，即得到一个对数凹性 α-有效检验。

关于精确的最大似然估计。上述测度 $\hat{}p_t$ 不一定能使似然比达到精确最大值。事实上，如果我们不使用精确的对数凹最大似然估计 $\hat{}p_t$，而是使用一个估计量 $\tilde{p}_t$，使得

$$\sum_{s\le t}\log\tilde{p}_t(X_s)\ge-\log(1/\varepsilon)+\sum_{s\le t}\log\hat{p}_t(X_s),$$

那么 $\varepsilon R_t\cdot\prod_{\overline{\hat{p}_t(X_s)\,\tilde{p}_t(X_s)}}$ 就是一个 e过程，并且像之前一样，可以用阈值 $1/\alpha$ 进行判定。这一点很重要，因为计算数据集的对数凹最大似然估计的实际方法并不精确，只能将解近似到对数似然目标函数中一个（由用户指定的）加性间隙以内，而且所需计算量随该加性间隙倒数的增大呈多项式增长。

在本节余下部分，我们将概率律 $P\in\mathcal{D}_1$ 与其密度等同，记作 p。

4.1 用于检验对数凹性的 ULR e-过程的一致性 y

ULR e-过程的一致性很大程度上取决于底层估计量 $\mathscr{E}$. 。事实上，考虑一个极端例子：令 $\hat{}q_t(X_t)=\mathbb{1}\{X_t=X_1\}$，此时只要概率律 P 是连续的，对于任意时刻 $t\ge d+1$，所得的 R_t 几乎必然为 0，因此该检验无法检验此类概率律。

由此可知，ULR e-过程只能对由估计量 $\mathscr{E}$所决定的一组概率律具有检验功效。具体而言，我们将针对以下一组“易于估计”的概率律提出同样的论证。下文中，d_H表示海林格距离。

定义 对于一个序贯估计量 $\mathscr{E}$，以及一个密度 $p\in\mathcal{D}_1$，将序列$\{X_t\}$ 的预测遗憾定义为

$$\rho_t(\mathscr{E};p):=\sum_{s\le t}\log p(X_s)-\log\hat{q}_{s-1}(X_s).$$

此外，令 $\mathfrak{L}_p$ 表示 p的对数凹 M 投影。我们将相对于对数凹性而言可由 $\mathscr{E}$ 良好估计的分布类定义为

$$\mathcal{Q}(\mathscr{E};c):=\left\{p\in\mathcal{D}_1:P^\infty\left(\limsup_{t\to\infty}\frac{\rho_t(\mathscr{E};p)}{td_H^2(p,\mathfrak{L}_p)}\le c\right)=1\right\}.$$

本节的主要结果是，基于 ULR 的检验对上述可良好估计的概率律具有检验功效；这一点将在本节后文中证明。

定理 10。 存在一个常数 $c > 1/25$，使得如果 $p \in \mathcal{Q}(\mathscr{E}; c) \setminus \mathcal{L}$，那么 $P^\infty(R_t \to \infty) = 1$.因此，*ULR e*-过程能够针对来自 $\mathcal{Q}(\mathscr{E}; c)$ 中任意分布的独立同分布抽样，给出一个相合检验。

上述可良好估计条件实质上要求能够在对数损失意义下良好地估计该分布。对于独立同分布的分布，人们预期，对于合理的 $\mathscr{E}$，估计量 $\hat{q}_t$ 会收敛到某个 $\hat{q}$，因此，当 t 较大时，遗憾会增长为 $\rho_t \approx t\mathrm{KL}(p\|\hat{q})$（如果 t 满足 $\mathrm{KL}(p\|\hat{q}_t)$，则该值可能关于 $\to 0$子线性增长，但不要求后一个收敛）。因此，类 $\mathcal{Q}$ 大致由能够以 KL 散度良好估计的分布组成。一般而言，这种估计可能颇具挑战性，因为 KL 散度对分布尾部的失配相当敏感。然而，在支撑集紧致性和光滑性等温和限制条件下，这种可估计性相当容易实现。事实上，我们给出如下陈述来阐明这一点。该陈述的证明见 §B.3。

推论 11。 令 $\mathcal{D}_{\text{Box、Lip和B}}$ 表示支撑于单位盒 $[-1,1]^d$ 且介于 $[1/B, B]$. 之间的 1-利普希茨密度的集合。存在一列筛最大似然估计量 $\mathscr{E}$，使得对于每个 $c > 0$、$\mathcal{D}_{\text{Box和Lip, B}} \subset \mathcal{Q}(\mathscr{E}; c)$，即 *ULR e*-过程对来自此类分布的独立同分布抽样给出相合检验。

进一步有趣的是，该检验的一致性并不要求遗憾 $\rho_t/t \to 0$足够小，而只要求相对于 p 与其对数凹 M 投影$\mathfrak{L}_p$之间的海林格距离足够小。这表明，在底层分布尚无法被估计之前，偏离对数凹性就可能已被检测到；从理论上看，这是相当有利的，不过其实际效果在很大程度上取决于在定理 c 中可以取多大的10。

定理 *10*的证明。 我们首先定义 $\sigma_t(p) = \sum_{s\le t} \log p(X_s) - \log \hat{p}_t(X_s)$。 注意到

$$\log R_t = \sigma_t(p) - \rho_t(\mathscr{E}; P).$$

此外，b 根据假设，我们有 $p \in \mathcal{Q}(\mathscr{E}, c)$，其中某个 c，因此对于任意 $\zeta > 0$，我们有 th at

$$\rho_t \le (1+\zeta)ctd_H^2(p, \mathfrak{L}_p),$$

当 e 足够 大时 t。因此，要证明 $R_t \to \infty$，只需证明 P^∞ 几乎必然 ly,

$$\text{下极限}_{t\to\infty} \frac{\sigma_t}{td_H^2(p, \mathfrak{L}_p)} \ge (1+2\zeta)c. \tag{3}$$

此时，以下引理便派上了用场，其证明留给 §B.1 .

引理 12。 对于任意 $p \in \mathcal{D}_1$，均有

$$P^\infty\left(\liminf_{t\to\infty} \frac{\sigma_t(P)}{td_H^2(P, \mathfrak{L}_P)} \ge \frac{1}{25}\right) = 1.$$

因此，只要满足 (3)，结论便成立，其中 $(1+2\zeta)c \le 1/25$，并且由于 $\zeta > 0$ 可以任意小，因此我们可以取任意 $c < 1/25$. 。我们注意到，此论证中的常数取值较宽松，非正式计算表明，或许可以将 c 提高至约 $1/6$。

引理 12 的证明依赖于对数凹最大似然估计 $\hat{p}_t$ 收敛至对数凹 M 投影 $\mathfrak{L}_p$ 的强收敛性质。回想一下，对于一对函数 $u \le v$，括号 $[u, v]$ 是指处处介于 u 和 v 之间的所有函数所构成的集合。利用 Cule 和 Samworth [CS10], Dunn 等人 [DGWR21 对数凹最大似然估计收敛性质的刻画，Lem. 1] 表明存在一个很小的括号，与 P 有显著间隔，使得 $\hat{p}_t$ 最终落在该括号内。以此事件为条件，我们接着利用 Wong 和 Shen [WS95] 的经典结果，该结果表明 $\sigma_t(p)$ 以至少 $1 - \exp(-\Omega(t))$ 的条件概率线性增长；由此，利用博雷尔-坎特利引理即可得证。如前所述，完整的引理 12 证明见 §B.1。

□

4.2 用于对数凹性检验的 ULR E-过程的检验功效

定理背后的论证 10 同样可以用于推导速率，但需要对底层分布 P作进一步限制。与上一节一样，我们通过分解 $\log R_t = \sigma_t(p) - \rho_t(\mathscr{E};\ p)$。

4.2.1 挑战与对数凹最大似然估计理论的背景

采用上述方法，论证分为两部分。首先，我们假设使用了一个足够好的估计量 $\mathscr{E}$，使得 ρ_t 以较高概率不会过大。对于我们采用的方法而言，这一假设是必要的；不过原则上，也可以用另一种分解方式来分析该检验，在这种情况下，或许可以放宽这一假设。无论如何，我们注意到，对于具体的备择假设，例如具有 Lipschitz 密度且支撑于单位超立方体上的概率律，ρ_t 确实可以得到适当控制。但值得注意的是，所得速率界在很大程度上取决于 ρ_t的表现，因而也取决于所考虑的估计量，这限制了后续结果的检验功效。

论证的第二部分要求我们证明 σ_t 很大，也就是说，当底层概率律不是对数凹时，对数凹最大似然估计无法很好地表示该概率律。尽管这一论断很自然，但要证明它并不容易，因为这需要我们理解对数凹最大似然估计在“模型之外”的表现，也就是数据本身并非来自对数凹分布时的表现。除了 Barber 和 Samworth [BS 21], 之外，文献中尚无人研究这一问题；大多数研究都聚焦于模型内的极小极大速率界 [KS16；KDR19；Han21；CDSS18]。下面我们对此进行一些详细讨论。

对模型内对数凹估计问题的精细分析，根本上依赖于一种微妙的化简：将对数凹最大似然估计的收敛速率问题，化为控制凸集上的经验过程偏差，即在来自 P 的数据下控制 $\sup_C |P(C) - P_t(C)|$ 的问题。其中，P_t 是经验概率律，上确界取遍有界区域 [CDSS18] 内的凸集。基于这一观察并对这些偏差作了更精细的研究，Kur 等人 [KDR19] 最近证明了形如 $d_H(\hat{}p_t, p) = O(n^{-1/(d+1)})$ 的紧确对模型内估计速率，其中 $p \in \mathcal{L}$ 且 $d \geq 3$（Han 证明了类似结果，并将其推广至 s-凹密度 [Han21]）。尽管这项研究中的重要部分可以扩展到分析模型外行为，但其分析最终无法应用于我们的情形。差异在于，他们的论证只能给出以下量的上界

$$\theta_t := \mathbb{E}_{X\sim P}[\log \mathfrak{L}_p(X)] - \mathbb{E}_{X\sim P}[\log \tilde{p}_t(X)],$$

其中，对于较小的常数 c，$\tilde{p}_t \propto \max(c, p_t)$ 是对数凹最大似然估计的一种轻微变体。当 $p = \mathfrak{L}_p$ 时，这个量是一个 KL 散度，因此有下界。然而，当 $p \neq \mathfrak{L}_p$ （也就是说，P 不是对数凹的）时，这个量完全可能为负。请注意，这恰恰是一个问题，因为 $\mathbb{E}[\sigma_t]/t \approx \theta_t + \mathbb{E}_{X\sim p}[\log p(X) - \log \mathfrak{L}_p(X)]$. 当 $p \notin \mathcal{L}$ 时，第二项确实可以证明很大，但第一项缺乏下界，限制了这类结果的适用性。我们还注意到，在模型内分析中相对简单的论证的其他方面（例如，证明测度 p 在形如 $\{\mathfrak{L}_p(x) < \gamma\}$ 的集合上赋予的测度很小），在模型外分析中也会失效。

当然，原则上我们也可以利用 Barber 和 Samworth 的结果。不过，这些结果给出的收敛速率相当差。粗略地说，他们论文中的定理5 [BS21] 表明，在模型外，$d_H(\hat{}p_t, \mathfrak{L}_p) \lesssim t^{-1/4d}$, 因此，任何利用这一结果的分析都无法指望证明 σ_t 对于$t \ll d_H(p, \mathcal{L})^{-4d}$的检验功效很大。 $4d$ 这种检验功效源于以下原因：[BS21] 的分析先将问题归结为经验分布在 Wasserstein距离下的收敛（这带来相对温和的因子 d），随后又相对于这一收敛遭受了$1/4$次方的速度减缓（这既不可避免，也导致了 $4d$ 指数）。

我们的分析通过基于有界对数凹概率律在紧支撑上的类的括号熵（见 §B.1）界来控制 σ_t 的增长，从而避开了这些问题。下面给出的界对所有d 都成立，但对于 $d \geq 4$似乎是新的。事实上，我们证明了以下结论。

引理 13。 令 $\mathcal{L}_{d,\ B}$ 表示具有定义在 $[-1,1]^d$ 上且一致有界于常数 B的对数凹密度的概率律集合。存在一个仅依赖于维数的常数 C_d，使得

$$\mathcal{H}_{[]}(\mathcal{L}_{d,B}, \zeta) = C_d \widetilde{\Theta}\left((B/\zeta)^{\max(d/2, d-1)}\right),$$

~其中 Θ 隐藏了随 ζ 或 B呈多对数级增长的项。

我们注意到，引理 13 的陈述中隐含的括号熵下界已由 Kim & Samworth [KS16，定理 8], 证明；他们还进一步给出了 $d \leq 3.$ 时相应的上界。虽然这并非本文的核心内容，但我们推导了 $d \geq 4$时相同情形的上界。关于上述界，有两个值得注意的技术要点。首先，注意对于 $d \geq 4$，熵界处于非 Donsker 区域，即 Dudley 积分 $\int_0^{\varepsilon} \sqrt{\mathcal{H}_{[]}(\mathcal{L}_d, B}$， $\zeta \mathrm{d}\zeta$ 由于在 $\zeta = 0$附近发散而不收敛，这通常~~（但并非总是）~~意味着通过熵积分所能证明的收敛速率会变慢。其次，对于 $d \geq 3$，该界按 $\zeta^{-(d-1)}$ 增长，而非按 $\zeta^{-d/2}$增长。后者与凸集的括号熵增长率相近，因此值得关注；凸集的括号熵增长率为 $\zeta^{-(d-1)/2}$；参见 §B.2.3。这一事实体现了此前讨论的简化方法的作用：将对数凹最大似然估计收敛率的分析，归结为控制凸集上经验过程的偏差；而凸集的熵增长较慢。

引理 13 在 §B.2.3 中得到证明。该结果的增长界保证，对于 $t \gtrsim d_H(p, \mathcal{L})^{2(d-1)}$, σ_t 相对于 t 呈线性增长，即使底层分布并非对数凹也是如此。这个 $2(d-1)$ 指数应与前文提到的 $4d$次幂尺度相比较；根据上文讨论的、基于 Wasserstein 连续性的方法，人们预期会得到这种尺度。当然，对 d 的依赖程度或许还能进一步改善。例如，如果模型内分析确实能够扩展到模型外，那么可以合理预期依赖形式为 $d+1$，而非 $2d-2.$。不过，这仍是未来工作中一项颇具挑战的问题。

值得注意的是，尽管文献中已有用于推导引理 13 中界的技术，但这些界本身似乎尚未被明确发现。我们认为，这可能是因为已有研究指出，由于这种熵存在下界，从熵的考量中得到的增长率界，对于对数凹最大似然估计的速率分析而言可能并非最优，至少在模型内情形下如此。还应指出，上述界明确针对紧支撑的对数凹分布（这是一项限制，但相对温和，因为所有对数凹密度都具有指数衰减尾部）。另请注意，我们为此情形构造的括号是“不适当的”，也就是说，括号函数本身并非对数凹；这可能会限制其在直接分析 $\hat{}p_t$ 与 $\mathfrak{L}_p$ 之间差异时的效用，但足以用于研究 σ_t 的行为。

4.2.2 界定 ULR 检验的典型拒绝时间

如前所述，我们对 σ_t 的分析需要借助有界且紧支撑的对数凹分布的括号熵界。要使这类界有效，我们需要确保对数凹最大似然估计 $\hat{}p_t$ 本身有界。为此，我们引入量 Δ_P，对概率律 P 定义为

$$\Delta_P := \min_{v:\|v\|=1} \mathbb{E}_P[|\langle v, X - \mathbb{E}_P[X]\rangle|],$$

该量由 Barber 和 Samworth 识别出来 [BS21] 作为对数凹投影的协方差下界的一个依据 P，进而可用于给 $\mathfrak{L}_p$ 以及（间接地）$\hat{}p_t$ 的上确界提供上界。注意， Δ_P 大致对应于 P 的协方差矩阵的最小特征值；事实上，最好将其视为该特征值的一种稳健版本。

有了这些准备，我们现在可以陈述主要结果，其证明见 §B.2.2。回顾一下， $d_H(p, \mathcal{L}) =$ 下确界$_{q\in\mathcal{L}}\, d_H(p, q)$ 且 $\tau_\alpha :=$ 下确界$\{t : R_t \geq 1/\alpha\}$为拒绝时间。

定理 14。 假设 p 的支撑集为 $[-1,1]^d$， 并令 $\pi_t \to 0$ 为一个序列，使得对每个 t,

$$P^\infty\left(\frac{\rho_t(\mathscr{C}, p)}{t d_H^2(p, \mathcal{L})} \geq \frac{1}{25}\right) \leq \pi_t.$$

那么存在一个常数 $c \in [1/600, 1]$ 和一个自然数 T_0 使得对于任意 $t \geq T_0 + \log(1/\alpha)\, c d_H^2(p,\mathcal{L})$,

$$P^\infty(\tau_\alpha > t) \leq \pi_t + \frac{1}{c}\exp\left(-ct d_H^2(p, \mathcal{L})\right) + \frac{1}{c}\exp\left(-ct\Delta_P^2/d^2\right),$$

and

$$T_0 = C_d \cdot \widetilde{O}\left(\Delta_P^{-\max(d^2/2, d^2-d)} d_H(p, \mathcal{L})^{-\max((d+4)/2, 2(d-1))} + d^2\Delta_P^{-2}\right),$$

where C_d depe ~仅取决于 d，且 O 隐藏了以多对数方式依赖于 $d_H(p, \mathcal{L})$ Δ_P 的项。

注意到 fr　　由上述结论可知，平均拒绝时间有界于 a　　s

$$\mathbb{E}[\tau_\alpha] = \sum_t P^\infty(\tau_\alpha > t) \le \sum_t \pi_t + O(T_0).$$

这里，第一项取决于使用估计量 $\mathscr{E}$ 对 p 的可预测性，而第二项则取决于我们对模型外情形下对数凹密度估计的噪声尺度所做的分析。

在典型情形下，这些项中的前一项将主导最终的界，因为典型的备择类通常远大于对数凹分布类。例如，利用关于一致有下界的Lipschitz密度估计的结果 [WS95，例如]，，我们得到关于集合 $\mathcal{D}_{\text{Box, Lip, B}}$ （在推论 11 中）所引入的集合。

推论 15。 对于任意常数 $B > 0$，存在一列筛最大似然估计量 $\mathscr{E}$，使得若 $p \in \mathcal{D}_{\text{Box, Lip, B}}$，则 *ULRT*拒绝时间 τ_α 的期望有界为 $\mathbb{E}[\tau_\alpha] = \widetilde{O}(d_H(p, \mathcal{L})^{-2(d+3)})$。

证明见 §B.3。我们指出，上述速率会根据底层分布p偏离对数凹性的程度而自适应，因为拒绝的时间尺度取决于 $d_H(p, \mathcal{L})$。 事实上，与批量检验相比，序贯检验的一项重要优势在于，它既能保持有效性，又能保证在自适应的时间尺度上检测到偏离。

关于紧致性。我们注意到，定理14以及尤其是推论15 中的指数对于对数凹性检验问题而言可能偏松。这是分析方法造成的；例如，推论15 中的缓慢速率主要由在单位盒上估计利普希茨分布的速率要求所决定，而这又是由于定理 π_t 中存在14项所致。这一方面或许可以改进，因为并没有什么理由要求我们使用能够很好刻画底层密度 p 的估计量。

事实上，我们不必针对 p 本身分析预测遗憾，而可以将 $R = \sigma_t(q) - \rho_t(\mathscr{E}; q)$ 分解为关于某个其他概率律 q 的项；该概率律或许属于一个比 p 可能对应的密度类 $\mathcal{Q}$ 更小的类。只要 (i) $\mathscr{E}$ 在对数损失下的预测效果不逊于 $\mathcal{Q}$ 中的任何概率律，并且 (ii) 无论 $p \notin \mathcal{L}$ 是什么， $\mathcal{Q}$ 中都存在一个比 $\mathcal{L}$ 中任何概率律都更“接近” p 的概率律，就应当能够进行类似的分析，尽管这可能需要对 $\mathscr{E}$ 的行为进行或许相当微妙的模型外控制，并仔细选择 $\mathcal{Q}$ 本身，以控制诸如 $d_H(p, \mathcal{Q})$ 和 $d_H(p, \mathcal{L})$ 这类距离的相对取值。对于对数凹分布，一种看起来颇有希望的方法是利用 s-凹密度来扮演 $\mathcal{Q}$ 的角色。它们尤其具有吸引力，因为它们构成了对数凹分布类的丰富扩展，同时却享有与对数凹分布相同的极小极大MLE收敛速率 [HW16; Han21]。

5 算法提案与模拟研究

现在，我们将以算法形式描述基于 ULR e-过程的对数凹性检验，并考察该检验在一个简单参数族上的具体实现表现。

5.1 计算方面、批处理与具体检验算法

在指定序贯估计量 $\mathscr{E}$ 并给出拟合对数凹最大似然估计的方法后，统计量R_t 便可显式计算，因此自然可以据此实现。尽管 e过程对范围广泛的备择假设类具有很强的检验功效，但其实现面临一个根本性的计算问题：每一轮都要重新计算 $\hat{}q_t$ 和 $\hat{}p_t$ 。由于必须对整个数据流计算分母 $\prod \hat{} p_t(X_s)$，这项开销会随 t 超线性增长；而估计该 $\hat{}p_t$ 本身的计算开销也会随样本数 t 超线性增长。数据维度 d 增加时，还会出现第二个问题，因为估计 $\hat{}p_t$ 的计算开销会随维度迅速增长。尽管存在关于 d 的多项式时间算法来计算对数凹最大似然估计 [Axe+19]，但在处理适中的 $d = 5$ 中的 ~ 100 个点时，现有最快方法通常也比处理规模相同的 $d = 1$ [RS19] 数据集所需的时间慢数百倍。我们通过利用批处理来减少计算负担，从而使计算在中等规模的 $d \le 4$上变得可行。其思路是等待

累积 e 个新样本后再重新计算 R_t，而不是每一轮都更新它。[2]

需要指出的是，只要 $\hat{q}_{t-1}$ 在整个批次中保持非预见性，这类分批更新仍保留 e过程性质，因此也保持有效性。具体而言，我们可以设定一个时间表，用递增的时间序列 $\mathcal{T} = \{t_k\}$ 表示，并计算统计量

$$R_t(\mathcal{T}) = R_{t-1}(\mathcal{T})\mathbb{1}\{t \notin \mathcal{T}\} + \sum \mathbb{1}\{t \in \mathcal{T}\} \prod_{j \le k(t)} \prod_{s=t_{j-1}}^{t_j} \frac{\hat{q}_{t_{j-1}}(X_s)}{\hat{p}_{t_k}(X_s)},$$

其中 $k(t) = \max\{k : t_k \le t\}$. 换言之，该时间表将数据流划分为一系列大小为 $t_k - t_{k-1}$ 的批次；每当累积出一个新批次，我们就基于先前的批次计算新的估计量 $\hat{q}$，并对已观测到的全部数据重新计算对数凹最大似然估计。这个过程仍由 $F_t(P)$ 的分批版本支配；在 P^∞ 下，它保留鞅性质，因此保证有效性。最简单可行的时间表是将 $t_k = kI$ 设为常数“批处理间隔” I。这实际上归结为检验 $p^{\otimes I}$ 的对数凹性；由于对数凹分布的张量积仍为对数凹分布，因此该检验有效。

请注意，这种批处理可能会降低检验功效。例如，只有在时间 t_k 才能拒绝检验；此外，由于批次其余部分的表现较为温和，信号较强的数据点可能会被“冲淡”，从而使统计量降低。尽管如此，我们在模拟研究中发现，检验功效的下降微乎其微，却换来了运行时间的显著改善。

有了这些内容，我们就可以在下面给出检验的明确算法描述。

算法 1 对数凹通用似然比检验

1: 输入：批处理时间表 $\{t_k\}_{k=1}^{\infty}$，其中 $t_1 \ge d+1$，估计量 $\mathscr{E}$，显著性水平 α。 2: 初始化： $R_t \leftarrow 1$，对于 $t \le t_1$、 $K \leftarrow 1$、 $N_1 = 1$、 $t \leftarrow 1$. 3: 当 $R_t < 1/\alpha$ 时执行 4: 如果 $t = t_K$，则 5: $\hat{q} = \mathscr{E}(X_1^{t-t_{K-1}})$。 6: $N_t \leftarrow N_{t-1} \cdot \prod_{s=t_{K-1}+1}^{t_K} \hat{q}(X_s)$。 7: $\hat{p} \leftarrow \mathscr{L}(X_1^{t_K})$。 8: $R_t \leftarrow N_t \cdot \left(\prod_{s=1}^{t_K} \hat{p}(X_s)\right)^{-1}$ .9: $K \leftarrow K+1$.10: 否则 11: $R_t \leftarrow R_{t-1}$。 12: $N_t \leftarrow N_{t-1}$。 13: $t \leftarrow t+1$。

5.2 评估 ULR E-过程检验

我们研究算法1 中的检验在以下简单测试分布族上的表现，其中$\mathbf{e}_d = \mathbf{1}_d/\sqrt{d}$ 是沿全 1 方向的单位向量，位于 $\mathbb{R}^d$，。

$$p(x;\mu,d) := \frac{1}{2(2\pi)^{d/2}} \left(\exp\left(-\|x - \tfrac{\mu}{2}\mathbf{e}_d\|^2/2\right) + \exp\left(-\|x + \tfrac{\mu}{2}\mathbf{e}_d\|^2/2\right)\right),$$

即，均值为$\pm\frac{\mu}{2}\mathbf{e}_d$的平衡双成分高斯混合分布，且协方差矩阵为单位协方差矩阵。均值差的范数恰为 μ，不失一般性，我们假设其非负。Dunn 等人提出的非序贯检验也曾采用这一分布族的一个小幅修改版本作为测试分布族。[DGWR21]。

这些概率律对于对数凹性检验的概念验证研究极为方便。事实上，可以看到，经过一次旋转后， d 维概率律是由一维概率律张量化得到的，其中包含一个

[2]当然需要注意的是，在我们的模拟研究中，为研究检验功效和检验尺寸而进行重复模拟，意味着我们只能将完全非参数检验用于不超过 $d = 4$ 的情形。尽管如此，在 $d = 6$ 上运行仍是可行的；在该情形下，时域为 100 步的单次运行耗时约 20 秒。

对数凹分布（具体而言，是 $d-1$维的标准高斯分布）。由于对数凹性在旋转下保持不变，且乘积分布的对数凹 M 投影是各边缘对数凹 M 投影的乘积[SW14], ，因此可以非常简单地刻画该概率律的对数凹性。具体而言，与对数凹性的距离完全取决于均值差的范数μ，且 $p(x;\ \mu,\ d)$ 是对数凹的，当且仅当 $\mu > 2.$ 因此，这些概率律为我们提供了一种简单的方法，既可检验统计量的检验尺寸和检验功效，也可研究维度增加对检验功效的影响。

最后，我们详细说明模拟过程。所有报告的数据都是每项实验运行 100 次所得的均值。所有模拟最多运行 100 个时间步，这主要是出于计算便利性考虑。因此请注意，我们的检验大小估计系统性地低于真实检验尺寸（时域无限长时的检验尺寸）。对于计算成本最低的 $d=1$ 情形，我们将时域延长至 500 个时间步，以说明在这种情况下变化不大，至少我们的检验在经验有效性方面如此。对于 $d=1$、2、3，检验每隔 $I=20$ 分批进行；而出于计算便利性考虑，对于 $d=4$，检验每隔 $I=25$ 分批进行。这些间隔占所研究时域的比例相当大，但并未显著降低检验功效，至少对于 $d=1$ 情形如此，明确的模拟结果证实了这一点。

所有代码均使用 R 实现。用于 $\mathscr{E}$ 的非参数估计量是通过 ks 软件包 [CD18]，实现的核密度估计，而所用的对数凹最大似然估计则来自 logConDens 软件包 [DR11]（用于 $d=1$ ）或 fmlcd 软件包（$d=2$、3、4） [RS19]。需要指出的是，后者并不保证返回对数凹最大似然估计，因为它优化的是定义该问题的非凸近似。不过，我们发现，与 logConcDEAD 软件包 [CSS10]，等备择方案相比，fmlcd 的实现保持了相近的有效性和检验功效，运行速度却显著更快。我们还研究了使用参数化高斯混合模型拟合的情形，以说明 $\mathscr{E}$ 中的低效率对检验功效的影响；为此，我们使用了通过 mclust 软件包 [SFMR16] 实现的期望最大化算法。所有模拟均在 AMD Ryzen 5650U 处理器上运行；这是一款用于笔记本电脑的中端中央处理器。

5.2.1 完全非参数检验

图2展示了我们采用完全非参数方法实例化该算法后的表现：对于 $d=1$，使用核密度估计量作为 $\mathscr{E}$over $p(\cdot;\ \mu, d)$，令 μ变化，在检验尺寸为 $\alpha=0.1$、$I=20$ 对应 $d\in\{1,2,3\}$ 且 $I=25$ 对应 $d=4.$ 的条件下运行。我们绘制了五条曲线，分别记录在 100 次独立运行中，检验在小于 20、40、60、80 和 100 个时步时拒绝零假设的运行比例；100 是检验运行的时域。

有三个主要观察结果。首先，我们观察到该检验具有出色的有效性。事实上，对于 $\mu\le 2$，零假设成立，而在这一情形下，两种情况下检验拒绝原假设的比例均不超过 0.02。其次，我们观察到，至少对于足够大的 μ，所有检验都会在 100 步内拒绝原假设。最后，我们注意到，随着 d 增加，检验功效急剧下降。具体而言，令 $\mu_*(d)$ 定义为使 $\mathbb{P}_{X\sim p(\cdot;\mu,d)}(\tau_{0.1}\le 100) = 0.9$的 μ 最小值。图2中的曲线给出了 $\mu_*(d)$ 的估计量；该值随 d急剧增大——从 $d=1$ 中的约 6 增至 $d=4$中的 1000 以上。[3]

鉴于非参数估计速率随着 d而显著下降，检验功效降低或许在意料之中。尽管如此，我们仍可以探究上述检验功效衰减在多大程度上是由拟合对数凹最大似然估计的低效所致，又有多少是核密度估计效率低下造成的。我们将在 §5.2.2 中通过研究预言机检验来考察这一影响。

更长时间运行以适用于 $d=1$。为展示有效性在更长时域内仍然成立，我们采用完全非参数方法，对500 个时间步进行检验，适用于 $d=1$，使用 $I=50.$ 。请看图3 ：即使在时间尺度增大的情况下，原假设 $\mu\le 2$ 下的拒绝仍得到了良好控制；随着时域增长，拒绝率稳步提高，不过时域从200增至500时，提升幅度较小。

[3]通过对 $d=5$进行初步模拟，我们观察到 $\mu_*(5)\approx 1500.$ 。需要指出的是，由于在更高维度下拟合最大似然估计的成本增加，而且拒绝率会随着维度增加而降低，因此需要在整个时域内执行更多次运行，从而延长实验的总成本；仅从耗时来看，这些模拟已经难以完整应用于 $d=5$。我们希望能够利用更强大的计算资源，将该方法应用于此类中等规模的 d。

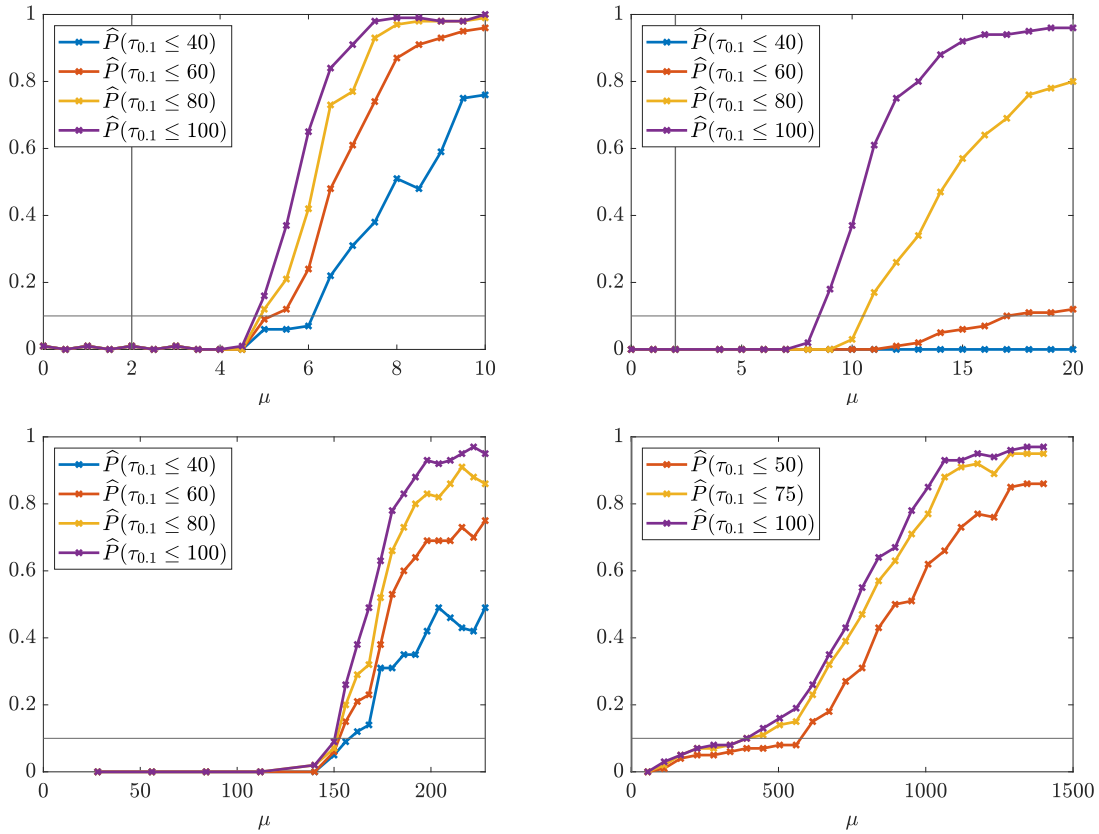


图2：**完全非参数检验的性能**。 在四种情形下，对完全非参数检验实现进行 $\alpha = 0.1$ 检验时，拒绝率与均值差 μ 的关系（基于 100 次模拟的经验拒绝率）： d =1, $I = 20$（左上）； $d = 2, I = 20$（右上）； $d = 3$, $I = 20$（左下）； $d = 4, I = 25$（右下）。细水平线表示水平 $\alpha = 0.1$，垂直线表示 $\mu = 2$，因为 $p(\cdot;\ \mu,\ d) \in \mathcal{L} \iff \mu \le 2$。请注意，所有图中的有效性性质都很强，同时随着维度增加，检验功效有所下降，这一点从 X 轴刻度范围的急剧扩大可以看出。

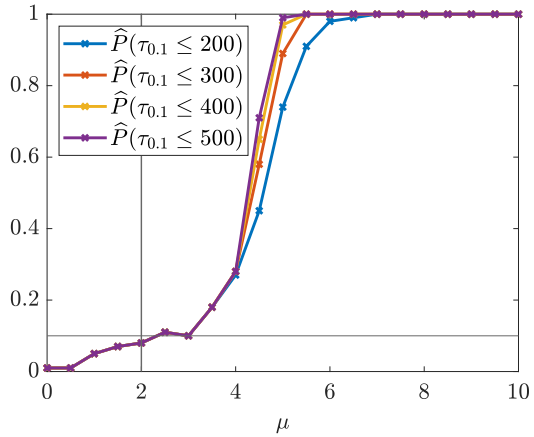


图3： **长时域下完全非参数检验的性能**。 在图2 所示的设定下进行（100次模拟）的经验拒绝率，其中 $d = 1$，在长度为 500 的时域内运行，且 $I = 50$。请注意，有效性在更长时域下仍然保持，而且对于 $\mu > 2$.

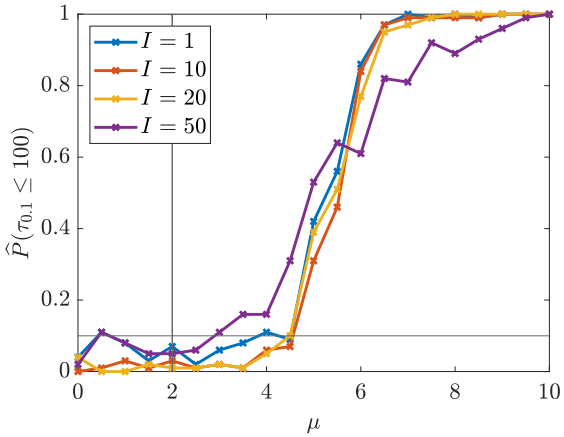


图 4： **对 I 完全非参数检验的影响**。 经验拒绝率（基于100 次模拟），采用图 2 中的设置， $d = 1$ 并改变$I \in \{1, 10, 20, 50\}$。请注意， $I = 10$、20 的拒绝率与 $I = 1$大致相同，而 $I = 50$ 则遭受了较大损失。

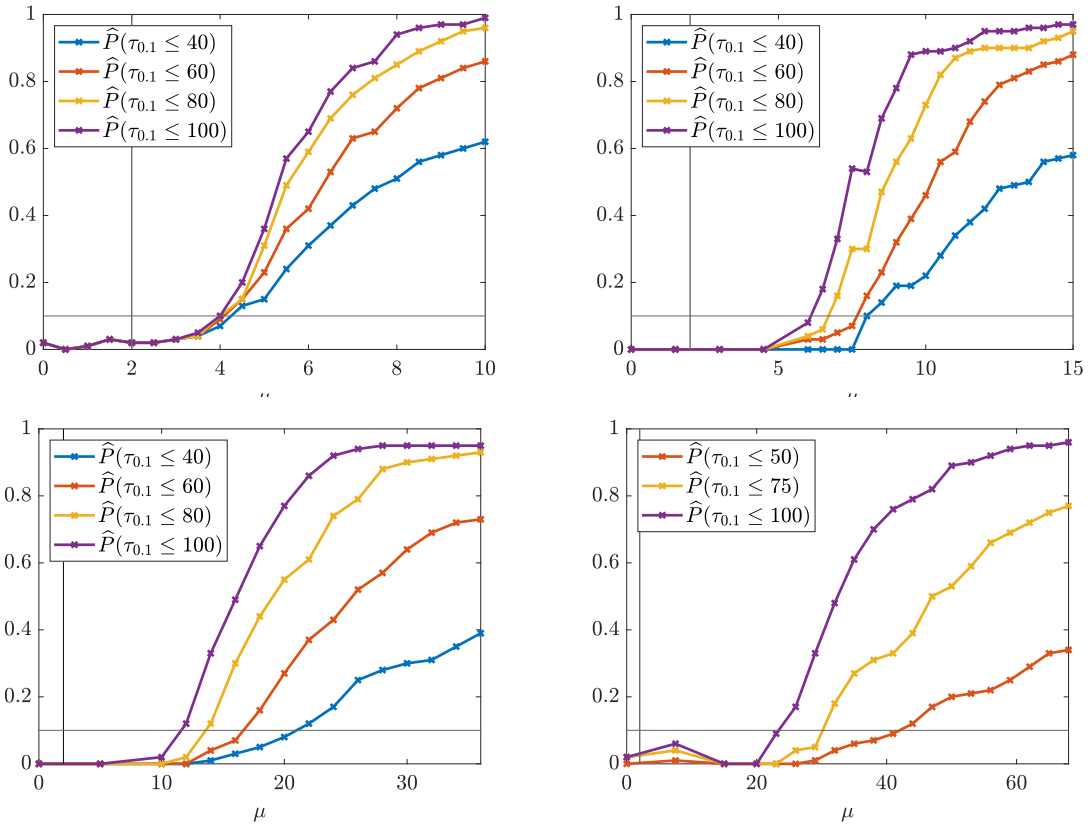


图 5：**部分预言机检验的性能**。经验拒绝率（基于100 次$\alpha = 0.1$ 模拟）随均值差 μ 变化的情况，展示了部分预言机检验实现在四种情形下的表现： $d = 1$， $I = 20$（左上）； $d = 2$， $I = 20$ （右上）； $d = 3$， $I = 20$ （左下）； $d = 4$， $I = 25$ （右下）。可以看到，在每幅图中，与完全非参数检验（图2）相比，检验功效均显著提高，这一点可从 X 轴刻度范围大幅收窄看出，尤其是在维度较高时。

批处理间隔的影响。如图4所示，批次大小为 $I = 10$ 和20 时，在备择设定（$\mu \ge 2$）下的拒绝率与直接 $I = 1$ 相比变化不大。有趣的是，注意到$I = 20$ 在中等 $I = 10$ 条件下的表现略好于 μ （范围 $4 - 6$），但在较大的中间 μ 条件下（范围 $6 - 8$）检验功效略有下降。另一方面，无批处理设定，即 $I = 1$，其检验尺寸（$\mu < 2$）有所恶化，但仍处于可接受水平。

大批处理大小 $I = 50$ 存在与 $I = 1$相同的有效性问题，但对于较小的非原假设值 μ (2-5)，其表现甚至优于 $I = 1$。当 μ 较大时（5-10），检验功效会显著下降。虽然目前尚不清楚这在多大程度上是由于时域长度仅为 100 所造成的假象，又在多大程度上是较大的批处理间隔直接导致的，但 $I = 10$,20 的良好表现表明，只要批处理间隔在时域长度中所占比例相对较小，检验功效的损失就不会太严重。

5.2.2 预言机检验及 $\mathscr{E}$质量的影响

预言机检验。 为了探究核密度估计的有损性对完全非参数检验功效的影响，我们运行“部分预言机”和“完全预言机”检验，对 $\mathscr{E}$ 进行调整，以利用有关底层分布 $p(\cdot;\, \mu, d)$ 的具体信息。在部分预言机检验中，我们调整 $\mathscr{E}$，使其估计双成分高斯混合模型，而不是核密度估计；在完全预言机情形下，我们直接将 $\hat{}q_{t-1}(\cdot) = p(\cdot;\, \mu, d)$，即对密度进行精确求值。

我们预期，对于从 $p(\cdot;\, \mu, d)$ 抽取的数据，这些检验比上文讨论的完全非参数检验具有更高的检验功效，因为在部分预言机的情形下，由于估计类的复杂度降低，遗憾 $\rho_t(\mathscr{E}; p)$ 会减小

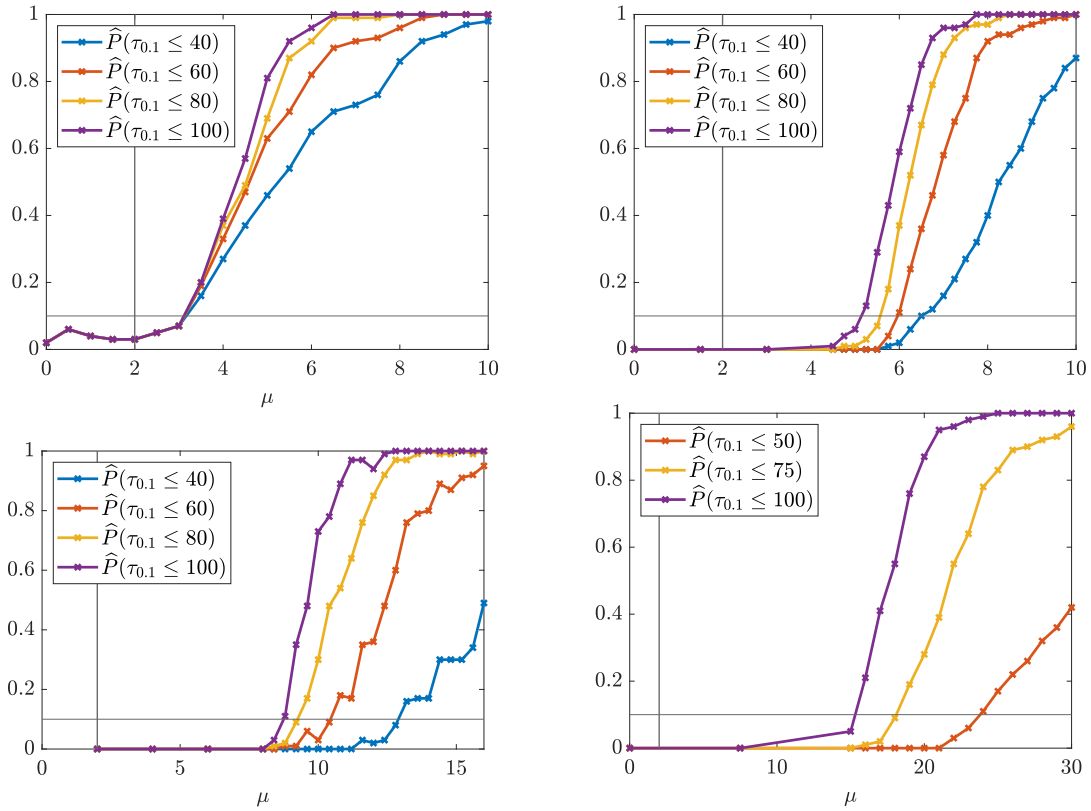


图 6：**完全预言机检验的表现**。经验拒绝率（基于100 次模拟），在 $\alpha = 0.1$下随均值差 μ 变化的完全预言机检验实现，涵盖四种情形： $d = 1$， $I = 20$ （左上）；$d = 2$， $I = 20$（右上）； $d = 3$， $I = 20$（左下）； $d = 4$， $I = 25$（右下）。与图2相比，检验功效显著提高，在高维情形下尤为明显，这可从 X 轴尺度大幅收缩看出。与图5相比，检验功效也有所提升：曲线达到高检验功效所需的 μ 约为部分预言机检验所需的一半。

；当然，在完全预言机的情形下，遗憾会恰好降至 0 。无论哪种情形，这实际上都会提高 R_t。因此，这些预言机检验使我们能够探究：在固定的 μ （即与对数凹性的距离固定）下，由对数凹最大似然估计收敛速率下降所导致的检验功效损失有多大。具体而言，完全预言机检验恰好捕捉了这一效应，而部分预言机检验则以较为渐进的方式逼近这一效应。图 5 展示了部分预言机检验的表现，图 6 则展示了完全预言机检验在 $d \in \{1, 2, 3, 4\}$.

比较图2 和5可见，对于 $d > 1$，采用部分预言机检验可显著提高检验功效。这一点在 $d = 2$ 中很明显：观察紫色曲线（100 个时间步内的总体拒绝率），可见其上升得更高，且在更小的 μ值处便不为零；同时，典型拒绝时间也显著缩短（例如，在完全非参数情形下，时间步60 之前从未出现拒绝，而在部分预言机检验下，较高的 μ值处便相当常见）。在 $d = 3$，4 中，这种效果更加显著——请注意，图的尺度完全改变了，在 $d = 3$上的量级从数百变为数十。这表明，由于参数高斯混合估计量的方差尺度较小，使用该估计量相比非参数核密度估计，能显著提升表现。

在完全预言机检验的情形下，上述效应表现得更加明显，每条拒绝率曲线都进一步改善（图 6）。例如，与部分预言机相比，我们对 $\mu_*(d)$ （满足以下条件的最小 μ ，使得$\mathbb{P}_{p(\cdot;\mu,d)}(\tau_{0.1} \le 100) = 0.9$)）的估计量在完全预言机情形下约减半；与完全非参数检验相比，也有了大幅改进。

的质量有显著影响。这些来自预言机检验的观察表明，$\mathscr{E}$所提供的估计量质量对检验的总体检验效力非常重要。在这些预言机示例中，通过降低估计量的方差尺度，同时保持偏差为 0，估计量质量得到了提升（因为概率律 $p(\cdot;\mu,d)$ 可以由每个估计量输出表示）。

当然，在实践中，我们无法总是期望在保持偏差为零的同时降低估计量的方差尺度。不过，这里隐含着二者之间的权衡。事实上，正如我们在 §4中简要讨论的那样，在检验中可以使用有偏的 $\mathscr{E}$，即不严格估计 p，只要$\mathscr{E}$的输出对 p 的表征优于对数凹最大似然估计。检验功效对 $\mathscr{E}$ 的强烈依赖表明，亟须探究这种设计自由度，并研究如何平衡方差（以 $\hat{}q_t$的收敛速率衡量）与偏差（即 $\lim\hat{}q_t$ 与 p之间的距离），从而优化检验功效。

6 讨论

我们的工作表明，对数凹性的序贯检验面临着有趣的挑战：在这一具有实际意义的场景中，普遍采用的检验鞅范式无法得到有效应用。在此过程中，我们刻画了连续空间上独立序贯概率律的闭叉凸包，从而为这一新工具的理论作出了贡献；该工具刻画了非负超鞅性质。随后，我们证明，通用似然 e-过程能够为对数凹性提供有力的检验。具体而言，我们展示了这些检验对于大类非参数备择分布律均相合，并且具有非平凡的收敛速率；为此，我们还对对数凹最大似然估计收敛的模型外分析以及通用检验功效分析的一般理论作出了贡献。通过在一个简单的参数分布族上运行检验，我们验证了这些性质，并进一步证明序贯估计量 $\mathscr{E}$ 对检验功效起着关键作用。从更广泛的视角来看，以上工作也可视为对新兴的 e-过程文献的贡献，尤其为这样一种观点增添了证据：对序贯检验的广泛研究必须利用这一强大而简单的工具。

这一有趣领域中，理论和方法方面仍有许多问题尚待解决，下面我们将讨论其中几个。

关于叉凸性，§3 和 §A 中对独立同分布高斯分布的闭叉凸包所作的刻画或许还可以进一步拓展，而精确了解哪些概率律属于这一集合将会非常有趣。此外，注意到，对独立同分布过程本身进行序贯高斯性检验是一项基本问题，但同样无法使用鞅进行检验（至少无法相对于数据的自然滤过进行检验）。构造并分析此类序贯高斯性检验，是一个自然而有趣的研究方向。当然，通用推断再次成为此类问题的一种自然方法，但也可以利用零假设（所有高斯分布）的平移不变性和旋转不变性，采用 [PLHG22] 中提出的方法。

关于基于 ULR e-过程的对数凹性检验，§5 表明，完全非参数检验的检验功效可能相当有限，尤其是在数据维度增大时。非序贯设置下， [DGWR21], 也提出了这一观察，并建议采用随机一维投影作为一种颇有意思的改进方法。在这种检验中，不再计算完整的 d 维核和对数凹估计，而是将数据投影到许多一维子空间上，并对每个投影数据集进行一维检验，由此得到的 e 值（在原假设下期望不超过 1 的非负检验统计量）取平均值。这种方法不仅得益于一维密度估计方法的速度，具有计算优势，而且在 §5 所述情形下也具有统计优势：检验功效随维度增加而衰减的程度大大减小。在序贯设置下当然也可以进行这类投影检验；无论从方法还是理论性质来看，这都是一个值得研究的自然方向。

从更广泛的角度来看，理论界和模拟都表明，估计量 $\mathscr{E}$ 的质量至关重要：它不仅直接影响对数凹性检验的检验功效，也更普遍地影响通用似然比 e过程的使用。鉴于此，并回想 $\mathscr{E}$ 中所讨论的隐含的“偏差—方差”权衡，见第4 节和第5节，探究 $\mathscr{E}$ 相对于原假设类

是一个值得探讨的问题，既涉及实际的方法学考量，也涉及研究基于 e 过程的检验之检验功效的理论问题。

致谢

作者感谢 Martin Larsson 就叉凸性进行了富有见地的讨论，并感谢 Robin Dunn 提供了对数凹性批处理通用检验的实现，该实现构成了我们模拟所用代码的基础。A. Rinaldo 和 A.G. 部分得到了美国国家科学基金会 DMS-EPSRC 2015489 号资助。

参考文献

[AHZ21] Sebastian Arnold、Alexander Henzi 和 Johanna F Ziegel。“用于预测校准的序贯有效检验”。载于：arXiv预印本 arXiv:2109.11761（2021）（见第 3 页）。[AS72] Daniel Alspach 和 Harold Sorenson。“使用高斯和近似的非线性贝叶斯估计”。载于：IEEE 自动控制汇刊 17.4（1972），第 439–448 页（见第 11 页）。[Axe+19] Brian Axelrod、Ilias Diakonikolas、Alistair Stewart、Anastasios Sidiropoulos 和 Gregory Valiant。“通过局部指数族实现对数凹最大似然估计的多项式时间算法”。载于：神经信息处理系统进展 32（2019）（见第 1、17 页）。[Bac10] Athanassia Bacharoglou。“用高斯测度的凸混合逼近概率分布”。载于：美国数学会会刊 138.7（2010），第 2619–2628 页（见第 11 页）。[BB06] Mark Bagnoli 和 Ted Bergstrom。“对数凹概率及其应用”。载于：理性与均衡。施普林格，2006，第 217–241 页（见第 1 页）。[Bro76] Efim Mikhailovich Bronshtein。“凸集和凸函数的ε熵”。载于：西伯利亚数学杂志 17.3（1976），第 393–398 页（见第 36 页）。[BS21] Rina Foygel Barber 和 Richard J Samworth。“对数凹投影的局部连续性及其在模型错设下估计中的应用”。载于：伯努利 27.4（2021），第 2437–2472 页（见第 15、16、35 页）。[CD18] José E Chacón 和 Tarn Duong。《多元核平滑及其应用》。Chapman and Hall/CRC，2018（见第 19 页）。[CDSS18] Timothy Carpenter、Ilias Diakonikolas、Anastasios Sidiropoulos 和 Alistair Stewart。“多元对数凹密度最大似然估计的近最优样本复杂度界”。载于：学习理论会议。PMLR，2018，第 1234–1262 页（见第 1、15 页）。[CS10] Madeleine Cule 和 Richard Samworth。“多维密度对数凹最大似然估计量的理论性质”。载于：电子统计学杂志 4（2010），第 254–270 页（见第 1、7、14、33 页）。[CSS10] Madeleine Cule、Richard Samworth 和 Michael Stewart。“多维对数凹密度的最大似然估计”。载于：英国皇家统计学会杂志：B辑（统计方法学）72.5（2010），第 545–607 页（见第 1、19 页）。[DGWR21] Robin Dunn、Aditya Gangrade、Larry Wasserman 和 Aaditya Ramdas。“通用推断与随机投影相结合：一种可扩展的对数凹性检验”。载于：arXiv预印本 arXiv:2111.09254（2021）（见第 1、14、18、23、33 页）。[DR11] Lutz Dümbgen 和 Kaspar Rufibach。“logcondens：与单变量对数凹密度估计相关的计算”。载于：统计软件杂志 39（2011），第 1–28 页（见第 1、19 页）。[FGNV12] Valentina Fedorova、Alex Gammerman、Ilia Nouretdinov 和 Vladimir Vovk。“用于在线检验可交换性的插件鞅”。载于：arXiv预印本 arXiv:1204.3251（2012）（见第 4 页）。

[GHK19] Peter Grünwald、Rianne de Heide 和 Wouter M Koolen。“安全检验”。载于：arXiv预印本 arXiv:1906.07801（2019）（见第 3 页）。[GW17] Fuchang Gao 和 Jon A Wellner。“ $\mathbb{R}^d$上凸函数的熵”。载于：构造性逼近 46.3（2017），第 565–592 页（见第 36 页）。[Han21] Qiyang Han。“集合结构全局经验风险最小化器在一般维数下具有速率最优性”。载于：统计年鉴 49.5（2021），第 2642–2671 页（见第 15、17 页）。[HRMS20] Steven R Howard、Aaditya Ramdas、Jon McAuliffe 和 Jasjeet Sekhon。“基于非负超鞅的时间一致切尔诺夫界”。载于：概率调查 17（2020），第 257–317 页（见第 3 页）。[HW16] Qiyang Han 和 Jon A Wellner。“通过 Rényi 散度逼近和估计 s-凹密度”。载于：统计年鉴 44.3（2016），第 1332 页（见第 17 页）。[KDR19] Gil Kur、Yuval Dagan 和 Alexander Rakhlin。“对数凹密度估计和有界凸回归中最大似然的最优性”。载于：arXiv预印本 arXiv:1903.05315（2019）（见第 1、15、36 页）。[KKMS08] Adam Tauman Kalai、Adam R Klivans、Yishay Mansour 和 Rocco A Servedio。“半空间的无先验学习”。载于：SIAM计算杂志 37.6（2008），第 1777–1805 页（见第 1 页）。[KS16] Arlene KH Kim 和 Richard J Samworth。“对数凹密度估计的全局收敛速率”。载于：统计年鉴 44.6（2016），第 2756–2779 页（见第 15、16、35、36 页）。[Lo72] J Lo。“有限维传感器轨道与最优非线性滤波”。载于：IEEE信息论汇刊 18.5（1972），第 583–588 页（见第 11 页）。[Mey66] P.A. Meyer。《概率与势》。科学与工业新闻。Blaisdell Publishing Company，1966（见第 3 页）。[PBKR22] Aleksandr Podkopaev、Patrick Blöbaum、Shiva Prasad Kasiviswanathan 和 Aaditya Ramdas。“序贯核化独立性检验”。载于：arXiv预印本 arXiv:2212.07383（2022）（见第 3 页）。[PLHG22] Muriel Felipe Pérez-Ortiz、Tyron Lardy、Rianne de Heide 和 Peter Grünwald。“E-统计量、群不变性与任意时刻有效检验”。载于：arXiv预印本 arXiv:2208.07610（2022）（见第 23 页）。[RGVS22] Aaditya Ramdas、Peter Grünwald、Vladimir Vovk 和 Glenn Shafer。“博弈论统计学与安全的任意时刻有效推断”。载于：arXiv预印本 arXiv:2210.01948（2022）（见第 3–5 页）。[RLKR22] Johannes Ruf、Martin Larsson、Wouter M Koolen 和 Aaditya Ramdas。“Ville 鞅定理的复合推广”。载于：arXiv预印本 arXiv:2203.04485（2022）（见第 5 页）。[RRLK20] Aaditya Ramdas、Johannes Ruf、Martin Larsson 和 Wouter Koolen。“可容许的任意时刻有效序贯推断必须依赖非负鞅”。载于：arXiv预印本 arXiv:2009.03167（2020）（见第 3、4、6 页）。[RRLK22] Aaditya Ramdas、Johannes Ruf、Martin Larsson 和 Wouter M Koolen。“可交换性检验：叉凸性、超鞅与 e-过程”。载于：国际近似推理杂志 141（2022），第 83–109 页（见第 4、8、9 页）。[RS19] Fabian Rathke 和 Christoph Schnörr。“快速多元对数凹密度估计”。载于：计算统计与数据分析 140（2019），第 41–58 页（见第 1、17、19 页）。[Sam18] Richard J Samworth。“对数凹密度估计的最新进展”。载于：统计科学 33.4（2018），第 493–509 页（见第 1 页）。[SFMR16] Luca Scrucca、Michael Fop、T Brendan Murphy 和 Adrian E Raftery。“mclust 5：使用高斯有限混合模型进行聚类、分类和密度估计”。载于：R杂志 8.1（2016），第 289 页（见第 19 页）。[SR21] Shubhanshu Shekhar 和 Aaditya Ramdas。“通过下注进行非参数双样本检验”。载于：arXiv预印本 arXiv:2112.09162（2021）（见第 3 页）。

[SW14] Adrien Saumard 和 Jon A Wellner。“对数凹性与强对数凹性：综述”。载于：统计调查 8 (2014)，第 45 页（引文见第 7、19 页）。[Vaa94] Aad van der Vaart。“光滑函数括号化”。载于：随机过程及其应用 52.1 (1994)，第 93–105 页（引文见第 38 页）。[Vil39] Jean Ville。“集体概念的批判性研究”。载于：《美国数学会通报》45.11 (1939)，第 824 页（引文见第 3 页）。[VNG03] Vladimir Vovk、Ilia Nouretdinov 和 Alexander Gammerman。“在线检验可交换性”。载于：第20届国际机器学习会议（ICML-03）论文集。2003，第 768–775 页（引文见第 4 页）。[VW96] Aad W Vaart 和 Jon A Wellner。“弱收敛”。载于：《弱收敛与经验过程》。施普林格，1996，第 16–28 页（引文见第 33 页）。[WR23] Ian Waudby-Smith 和 Aaditya Ramdas。“通过下注估计有界随机变量的均值”。载于：皇家统计学会杂志（B 辑），即将发表并附讨论（2023）（引文见第 3 页）。[WRB20] Larry Wasserman、Aaditya Ramdas 和 Sivaraman Balakrishnan。“通用推断”。载于：《美国国家科学院院刊》117.29 (2020)，第 16880–16890 页（引文见第 1、5、13 页）。[WS95] Wing Hung Wong 和 Xiaotong Shen。“似然比的概率不等式与筛最大似然估计的收敛速率”。载于：统计年鉴 (1995)，第 339–362 页（引文见第 14、17、33、37 页）。

A 叉凸包的平凡性证明与性质

本附录旨在证明关于叉凸包的结构性引理，并讨论我们论证中的技术细节。

A.1 关于局部 $L_1(\Gamma)$ 闭包的细节

首先，我们来明确说明闭叉凸组合中隐含的收敛概念。

回顾一下，f-$\overline{\text{conv}}(\mathfrak{P})$ 是 f-conv($\mathfrak{P}$) 关于似然比过程在每个固定时刻 t 的 $L_1(\Gamma)$-收敛意义下的闭包。我们用简单的方式来拆解这句话。设 $\mathtt{P}^n$ 是 f-conv($\mathfrak{P}$) 中密度比为 $Z_t^n := Z_t^{\mathtt{P}^n}$ 的某个序列。我们称 $\mathtt{P}^n \to \mathtt{P}$，当且仅当对于每个 t，都有 $Z_t^n \to Z_t$ 属于 $L_1(\Gamma)$。由于 Z_t 和 Z_t^n 是 $\mathscr{F}_t$ 可测对象，这种收敛就是在 $L_1(\Gamma|_t)$ 中的收敛。说收敛需要在每个固定时刻 t 发生，意味着这种收敛不必在 t 上一致：例如，Z_{100}^n 比 Z_1^n 收敛得更慢也无妨。这种收敛概念可以通过以下方式度量

$$\Delta(\mathtt{P},\mathtt{Q}) := \sum_{t\in\mathbb{N}} 2^{-t}\|Z_t^{\mathtt{P}} - Z_t^{\mathtt{Q}}\|_{L_1(\Gamma)}.$$

注意到 Δ 有界，因为

$$\|Z_t^{\mathtt{P}} - Z_t^{\mathtt{Q}}\|_{L_1(\Gamma)} = \int \left|\mathtt{P}|_t(\mathrm{d}x_1^t) - \mathtt{Q}|_t(\mathrm{d}x_1^t)\right| \le \int \mathtt{P}|_t(\mathrm{d}x_1^t) + \int \mathtt{Q}|_t(\mathrm{d}x_1^t) = 2.$$

有了这一点，我们先证明下面这个反复用到的辅助结论 .

引理 16。 设 $\mathfrak{P}$ 是一个序贯概率律集合，且 $\mathtt{R}$ 是任意一个序贯概率律。假设存在一个序贯概率律序列 $\{\mathtt{R}^T\}$，使得每个 $\mathtt{R}^T \in$ f-凸包($\mathfrak{P}$)，且对所有 $t \le T$，$Z_t^{\mathtt{R}^T} = Z_t^{\mathtt{R}}$. 均成立。那么 $\mathtt{R} \in$ f-凸包($\mathfrak{P}$)。

证明。 我们断言 $\mathtt{R}^T \to \mathtt{R}$。事实上，由于 $Z_t^{\mathtt{R}^T} = Z_t^{\mathtt{R}}$ 对所有 $t \le T$ 均成立，

$$\Delta(\mathtt{R}^T,\mathtt{R}) \le \sum_{t>T} 2^{-t}\cdot 2 = 2^{-(T-1)}.$$

因此，$\lim_{T\to\infty}\Delta(\mathtt{R}^T,\mathtt{R}) = 0$，也就是说 $\mathtt{R}^T \to \mathtt{R}$。根据定义，$\mathfrak{P}$ 的闭叉凸包包含此类极限，故断言得证。 □

上述结论为我们的论证带来了很大便利，因为它使我们只需构造过程，使其在有限时间内与叉凸包中某个声称的元素相匹配即可。通常，在下面的论证中，仅用有限叉凸组合就很容易做到这一点。

A.2 关于 独立序贯概率律的叉凸包的证明 s

现在我们可以继续证明在 §3 中省略的引理。

引理 4 的证明。 如正文所述，通过重复进行叉凸组合，可得$\mathtt{R}^T \in$f-凸包($\mathcal{P}^\infty$)， 其中

$$\mathtt{R}^1 := P_1^\infty, \mathtt{R}^T := \left[\mathtt{R}^{T-1} \overset{T-1,0}{\longrightarrow} P_T^\infty\right],$$

混合权重 0 的有效性利用了 $\mathcal{P}$ 中概率律的相互绝对连续性。由引理 16 得证。 □

引理 5 的证明。只需证明对所有有限的 $k, \mathcal{P}_k^\infty \subset$ f-凸包($\mathcal{P}^\infty$)，因为 $\mathcal{P}_* = \bigcup_k \mathcal{P}_k$，且 $\mathcal{P}_k \subset \mathcal{P}_{k+1}$ 对所有 k 均成立。对于 $T \in \mathbb{N}$ 以及定义在 $\mathbb{R}^d$ 上的两个概率律 P, Q，将序贯概率律 $\mathtt{R}_{P,Q,T}$ 定义为独立序列 $\{X_t\}$ 的概率律，使得 $X_t \sim P$对 $t \le T$成立，且 $X_t \sim Q$对 $t > T$成立，即 $\mathtt{R}_{P,Q,T} = \left[P^\infty \overset{T}{,\longrightarrow}\overset{0}{} Q^\infty\right]$。对于 $T \in \mathbb{N}$，定义 $\mathfrak{P}_{k,T}$ 为如下形式的序贯概率律集合：$\mathtt{R}_{P,Q,T}$，其中 $P \in \mathcal{P}_k$ 且 $Q \in \mathcal{P}$。

我们首先声称 $\mathfrak{P}_{k,\ T} \subset$ f-凸包(P^∞)。我们对 k 使用归纳法证明这一点。任取 T，注意到显然 $\mathfrak{P}_{1,\ T}$属于这个叉凸包。对于 $k \geq 2$，我们可以将每个 $P \in \mathcal{P}_k$表示为 $P = \alpha P^1 + (1-\alpha)P^2$，其中某些 $\alpha \in [0,1]$，$P^1 \in \mathcal{P}_{k-1}$和 $P^2 \in \mathcal{P}$。我们需要证明，对于任意这样的 P以及任意 $Q \in \mathcal{P}$，$\mathrm{R}_{P,\ Q,\ T}$属于 $\mathcal{P}^\infty$的叉凸包。根据归纳假设，$\mathrm{R}_{P^1,\ Q,\ T} \in$属于 f-凸包$(\mathcal{P}^\infty)$，且 $\mathrm{R}_{P^2,\ Q,\ T} \in$属于 f-凸包$(\mathcal{P}^\infty)$。但接下来定义这些概率律

$$\mathrm{S}^0 := \mathrm{R}_{P^1,Q,T}, \widetilde{\mathrm{S}}^\tau := \left[\mathrm{S}^{\tau-1} \overset{\tau-1,\alpha}{\longrightarrow} \mathrm{R}_{P^2,Q,T}\right], \mathrm{S}^\tau := \left[\widetilde{\mathrm{S}}^\tau \overset{\tau,0}{\longrightarrow} \mathrm{R}_{P^1,Q,T}\right].$$

我们注意到，上述每个叉凸组合的权重都是有效的，因为 $\mathcal{P}$ 相对于测度绝对连续，因此密度过程从不为 0。我们声称 $\mathrm{S}^T = \mathrm{R}_{P,\ Q,\ T}$。

事实上，令 p^1、p^2、q分别表示 P^1、P^2和 Q相对于标准高斯分布的密度，并令 Z_t^1和 Z_t^2分别表示 $\mathrm{R}_{P^1,\ Q,\ T}$和 $\mathrm{R}_{P^2,\ Q,\ T}$的密度过程。它们可以显式计算为

$$Z_t^i = \prod_{s \leq \min(t,T)} p^i(X_s) \cdot \prod_{s=\min(t,T+1)}^{t} q(X_s),$$

其中 $i \in \{1,2\}$，并且我们注意到，对于 $u < v$，$\prod_{s=v}^{u} \cdot = 1$。注意，对于每个 i以及任意 $t_1 < T$和 $t > t_1$，我们有

$$\frac{Z_t^i}{Z_{t_1}^i} = \prod_{s=\min(t_1,T)+1}^{\min(t,T)} p^i(X_s) \cdot \prod_{s=\min(t,T+1)}^{t} q(X_s).$$

我们将归纳地声称，对于每个 τ，S^τ 的密度过程满足

$$Z_t^{\mathrm{S}^\tau} = \prod_{s \leq \min(t,\tau)} (\alpha p^1(X_s) + (1-\alpha)p^2(X_s)) \cdot \prod_{s=\min(t,\tau+1)}^{\min(t,T)} p^1(X_s) \cdot \prod_{s=\min(t,T+1)}^{t} q(X_s).$$

事实上，基础情形显然成立，因为对于 $\tau = 0$，$\mathrm{S}^0 = \mathrm{R}_{P^1,\ Q,\ T}$。假设对$\tau$归纳假设成立，我们注意到，由于 $\tilde{S}^{\tau+1}$在时刻 τ是 S^τ与 $\mathrm{R}_{P^2,\ Q,\ T}$的叉凸组合，因此它在时刻 τ之前与 S^τ 共享密度过程，而在此之后，其密度是两个密度过程的混合，于是得到

$$\begin{aligned} Z_t^{\mathrm{S}^{\tau+1}} = & \prod_{s \leq \min(t,\tau)} (\alpha p^1(X_s) + (1-\alpha)p^2(X_s)) \\ & \times \left(\alpha \left\{\prod_{s=\min(t,\tau+1)}^{\min(t,T)} p^1(X_s) \cdot \prod_{s=\min(t,T+1)}^{t} q(X_s)\right\} + (1-\alpha)\left\{\prod_{s=\min(t,\tau+1)}^{\min(t,T)} p^2(X_s) \cdot \prod_{s=\min(t,T+1)}^{t} q(X_s)\right\}\right), \end{aligned}$$

其中我们使用了上面对 Z_t^i/Z_τ^i 在 $t \geq \tau+1$中的行为。

最后，$\mathrm{S}^{\tau+1}$以混合权重 0 在时刻 $\tau+1$ 将上述概率律与 $\mathrm{R}_{P^1,\ Q,\ T}$进行混合。这意味着，$\mathrm{S}^{\tau+1}$在时刻 $\tau+2$之后的后缀概率律恰好等于 $\mathrm{R}_{P^1,\ Q,\ T}$的概率律，而截至时刻$\tau+1$ 的前缀保持不变。换句话说，

$$\begin{aligned} Z_t^{\mathrm{S}^\tau} = & \prod_{s \leq \min(t,\tau)} (\alpha p^1(X_s) + (1-\alpha)p^2(X_s)) \cdot \prod_{s=\min(t,\tau+1)}^{\tau+1} (\alpha p^1(X_s) + (1-\alpha)p^2(X_s)) \\ & \times \prod_{s=\min(t,\tau+2)}^{\min(t,T)} p^1(X_s) \cdot \prod_{s=\min(t,T+1)}^{T} q(X_s). \end{aligned}$$

注意到前两个乘积可以合并为 $\prod_{s \leq \min(t,\ \tau+1)}(\alpha p^1(X_s) + (1-\alpha)p^2(X_s)$，即可得到该结论。

有了这一点，论证就很直接了。对于任意元素$\mathrm{P} \in \mathcal{P}_k^\infty$，我们注意到，$\mathfrak{P}_{k\text{中存在某个元素},\ T}$，记为$\mathrm{P}^T$，使得$\mathrm{P}^T$的密度过程与$\mathrm{P}$ 在时间 T之前相同。应用引理16 即可得到所需结论。

□

引理 6 的证明。令 $\mathrm{P} = \bigotimes\{P_t\}$ 为任意一个 $P_t \in \overline{\mathcal{P}}$序列。我们需要证明 $\mathrm{P} \in$是$\bigotimes \mathcal{P}$的 f-凸包。然而，由于 $P_t \in \mathcal{P}$ 对于每个 t，对每个 t 还存在序列 $\{P_t^n\}_{n\in\mathbb{N}}$，其中每个$P_t^n \in \mathcal{P}$都满足 $P_t^n \to P_t$属于 $L_1(\Gamma)$。令 $\mathcal{Q} := \{P_t^n\colon\ t, n \in \mathbb{N}\}$。我们注意到 $\bigotimes \mathcal{Q} \subset \bigotimes \mathcal{P} \implies$是$\bigotimes \mathcal{Q}$的 f-凸包 $\subset$是$\bigotimes \mathcal{P}$的 f-凸包。令 $\mathfrak{Q} :=$为$\bigotimes \mathcal{Q}$的 f-凸包。我们将论证 $\mathrm{P} \in \mathfrak{Q}$.

令 P^T 为具有密度过程的序贯概率律

$$Z_t^{\mathrm{P}^T} = \begin{cases} \prod_{s\le t} p_s(X_s) & \text{if } t \le T \\ \prod_{s\le T} p_s(X_s) \cdot \prod_{T<s\le t} p_s^1(X_s) & \text{if } t > T \end{cases}.$$

若我们能证明对每个 T，$\mathrm{P}^T \in \mathfrak{Q}$，则结论成立，因为 $\mathrm{P}^T \to \bigotimes\{P_t\}$ 正如引理 16 的证明中所述，并且 $\mathfrak{Q}$ 在相关的收敛概念下是闭的。

我们将用归纳法证明这一点。令 $\mathrm{P}^{1,\ n}$ 为密度为 $Z^{1,\ n}_{\ t} := p_1^n(X_1)\cdot\prod_{s>\min(1,\ t)} p_s^1(X_s)$ 的序贯概率律。注意，对每个 n，$\mathrm{P}^{1,\ n} \in \bigotimes \mathcal{Q} \subset \mathfrak{Q}$。此外，

$$\Delta(\mathrm{P}^{1,n}, \mathrm{P}^1) \le \|Z_1^{1,n} - Z_1^{\mathrm{P}^1}\|_{L_1(\Gamma)} \to 0.$$

因此 $\mathrm{P}^1 \in \mathfrak{Q}$.

现在假设　　对某个 $T \ge 2.$，$\mathrm{P}^{T-1} \in \mathfrak{Q}$ 成立。对于 T，$n \in \mathbb{N}$，定义 Q^n 为密度比为

$$Z_t^{\mathrm{Q}^n} := \begin{cases} \prod_{s<t} p_s^1(X_s) & t < T \\ Z_{T-1}^{\mathrm{Q}^n} \cdot p_T^n(X_T) & t = T \\ Z_T^{\mathrm{Q}^n} \cdot \prod_{s=T+1}^t p_s^1(X_s) & t > T \end{cases}.$$

显然，对于所有 n，$\mathrm{Q}^n \in \bigotimes \mathcal{Q} \subset \mathfrak{Q}$。现在，定义

$$\mathrm{P}^{T,n} = \left[\mathrm{P}^{T-1} \overset{T-1,0}{\longrightarrow} \mathrm{Q}^n\right],$$

这是成立的，因为每个 P_t^n和 P_t 都相互绝对连续。但对于$t \le T-1$，$Z^{\mathrm{P}^T}{}_{,t}{}^{\ n} = Z_t^{\mathrm{P}^{T-1}} = Z_t^{\mathrm{P}^T}$，而对于 $t \ge T$,

$$Z_t^{\mathrm{P}^{T,n}} - Z_t^{\mathrm{P}^T} = Z_{T-1}^{\mathrm{P}^T} \cdot (p_T^n(X_T) - p_T(X_t)) \cdot \prod_{s=T+1}^t p_s(X_s).$$

由此可得

$$\|Z_t^{\mathrm{P}^{T,n}} - Z_t^{\mathrm{P}^T}\| = \begin{cases} 0 & t < T \\ \|P_T^n - P_T\|_{L_1(\Gamma)} & t \ge T \end{cases},$$

因此，$\Delta(\mathrm{P}^{T,n}, \mathrm{P}^T) \le \|P_T^n - P_T\|_{L_1(\Gamma)} \to 0.$由 $\mathfrak{Q}$ 的接近性，我们得出 $\mathrm{P}^T \in \mathfrak{Q}$。 □

A.3 引理8的证明

引理8的证明。固定一个 $m \in \mathbb{N}$。由于 E 具有正测度且可测，存在一个开集$O \in (\mathbb{R}^d)^t$ 使得 $O \supset E$ 且 $\mathrm{Leb}_{dt}(O) \le (1 + 1/m)\mathrm{Leb}_{dt}(E)$。注意，此处 的“大部分”测度 O都位于 E内。

由于 O 是开集，　存在一列互不相交的开矩形 R_i，位于 $(\mathbb{R}^d)^t$中，使得 $\bigsqcup R_i \subset O \subset \bigcup \overline{R}_i$,
and

$$\mathrm{Leb}_{dt}\left(\bigsqcup R_i\right) = \sum \mathrm{Leb}_{dt}(R_i) = \mathrm{Leb}_{dt}(O).$$

此外，由于 O 的大部分测度集中在 E中，我们得出结论：至少存在一个 i 使得

$$\mathrm{Leb}_{dt}(R_i) > 0 \quad and \quad \mathrm{Leb}_{dt}(E \cap R_i) \ge \frac{m}{m+1}\mathrm{Leb}_{dt}(R_i).$$

事实上，否则我们就会有

$$\mathrm{Leb}_{dt}(E) = \mathrm{Leb}_{dt}(E \cap O) = \sum \mathrm{Leb}_{dt}(E \cap R_i) < \frac{m}{m+1}\sum \mathrm{Leb}_{dt}(R_i) \le \frac{m}{m+1} \cdot \frac{m+1}{m}\mathrm{Leb}_{dt}(E),$$

这是不可能的。 □

A.4 Tec叉凸包的技术方面与我们的平凡性论证 nt

我们对证明非平凡非负超鞅不存在这一论证中的一些技术问题作出评论。具体而言，我们讨论了非平凡性的定义是否必要、论证中反复使用的相对于测度绝对连续条件、如何将论证扩展到有界集合上的对数凹分布，以及试图放宽叉凸组合的定义以处理支撑集不匹配时会出现的问题。

超越几乎必然平凡性。 正文将平凡非负超鞅（以及非负鞅）定义为Γ-几乎必然非递增（分别为常数）的过程。能否改为证明不存在非平凡的$\mathcal{L}^\infty$-非负超鞅，并采用更强的意义，即这类过程必须是非递增的（而不只是几乎必然非递增）？事实证明这是不可能的，以下过程便是一个例子

$$M_t := \frac{1}{1 - \mathbb{1}\{\exists (t_1,t_2,t_3,t_4) \in [1:t]^4 : X_{t_1} = X_{t_2}, X_{t_3} = X_{t_4}, X_{t_1} \neq X_{t_3}\}}.$$

由于对数凹测度至多只有一个原子（这是单峰性的结果），因此 $\{M_t\}$ 是一个$\mathcal{L}^\infty$-鞅（事实上，正如定理所述，它几乎必然恒等于常数 1）。然而， M_t 确实会发散至 ∞；对于任何至少含有两个原子的独立同分布序贯概率律，例如抛硬币过程，这种情况几乎必然发生。这意味着，虽然可能无法使用检验鞅拒绝具有勒贝格密度的过程，但却可以拒绝原子过程。从结构上看，这个例子与以下事实有关：无法用具有密度的测度以 L_1 意义逼近点质量。因此，尽管对于具有密度的独立过程， $\mathcal{L}^\infty$-非负超鞅也必须是非负超鞅，但这一结论并不适用于从含原子的分布中抽取的序列。从另一个角度看，这个问题与下文讨论的、在支撑集不匹配的概率律的叉凸组合扩展下 NSM 性质丧失的问题相同，因为各自仅含不同单个原子的两个概率律，其部分相互奇异。

相互绝对连续条件在 $\mathcal{P}$中的作用。两个概率律$\mathtt{P}$和$\mathtt{Q}$在时刻 s的叉凸组合定义涉及密度过程之比 $Z_s^{\mathtt{P}}/Z_s^{\mathtt{Q}}$。正如§3 的算法视角所示，这个比值确实不可或缺，因为若$\mathtt{R}$ 是叉凸组合，那么其前缀分布$\mathtt{R}|_s = \mathtt{P}|_s$. 然而，如果 $Z_s^{\mathtt{Q}} = 0$，也就是说，如果对于 $\{X_t\} \sim \mathtt{R}$，前缀 X_1^s 落在一个在$\mathtt{Q}$下几乎不可能发生的集合中，那么上述比值就没有意义。这一观察构成了如下条件的依据：若 $Z_s^{\mathtt{Q}} = 0$，则混合权重 h 必须恰好等于1。

我们的论证最终表明，任何形如 $\bigotimes\{P_t\}$ 的概率律都属于 f-凸包($\mathcal{P}^\infty$)。然而，为了证明这一点，我们的构造依赖于设定 $h = 0$ ，以便在 $\mathcal{P}$ 中不同的概率律之间产生切换。我们假设相互绝对连续，正是为了能够灵活地做到这一点，同时避免遇到上一段讨论的问题。

高斯分布在我们论证中的作用。由于我们利用高斯分布的密度证明了 $\mathcal{L}^\infty$-NSM 也必定是 $\bigotimes\mathcal{D}$-NSM，因此有必要探讨 $\mathcal{L} \supset \mathcal{G}$ 对于结果的核心结论究竟有多重要。[4] 在论证中，高斯分布发挥着两种作用：首先，由于所有高斯分布的支撑集都是整个定义域，因此这一类相对于测度绝对连续，我们可以灵活地取叉凸组合。其次，高斯非负超鞅的平凡性源于高斯混合在实数上的密度集合中 L_1稠密。只要 $\mathcal{L}$ 的任何子集满足这两个性质，就足以达到我们的目的。

将该论证扩展至 $\mathbb{R}^d$ 上的对数凹分布。最后，我们注意到，该论证可以直接推广到实数受限子集上的对数凹分布：对于有界凸集 K，定义 $\mathcal{L}_K$ 为支撑于 K上的对数凹密度。于是，所有 $\mathcal{L}_K^\infty$-NSM 都是平凡的，即相对于参考测度（均匀分布(K)）几乎处处不增。$^\infty$. 这是因为截断高斯分布仍然稠密，并且支撑于整个定义域 K 上。

[4]尽管这一结果对于高斯分布本身也颇有意义，因为它告诉我们，存在一个简单且十分自然的参数族，无法通过非负超鞅进行检验。

为说明这一点，首先注意，如果 $\gamma := \sum \alpha_i \phi_i$ 是高斯混合，那么对于任意具有非零勒贝格测度的 K，截断后的 $\gamma|_K$ 同样是截断高斯分布的混合。事实上，定义 $\theta_i = \int_K \phi_i$. 则有

$$\gamma|_K(x) = \sum \frac{\alpha_i}{\sum \alpha_i \theta_i} \phi_i(x) \cdot \mathbf{1}\{x \in K\} = \sum \frac{\alpha_i \theta_i}{\sum \alpha_i \theta_i} \phi_i|_K(x).$$

现在，令 p 为支撑在 K 上的任意密度，并令 $\gamma_n \to p$ 为一个高斯混合序列，收敛至 $d_n := \int |p - \gamma_n| \to 0$. 。于是，定义 $\pi_n = \int_{K^c} \gamma_n$ 后，我们有

$$\int |p - \gamma_n|_K| = \int_K \frac{|p(1-\pi_n) - \gamma_n|}{1-\pi_n} \le \int_K \frac{|p - \gamma_n|}{1-\pi_n} + \int_K \frac{\pi_n p}{1-\pi_n} \le \frac{\pi_n + \int |p - \gamma_n|}{1-\pi_n}.$$

此外，由于 p 支撑在 K 上，$\pi_n = \int_{K^c} \gamma_n \le \int_{K^c} \gamma_n + \int_K |p - \gamma_n| = d_n$。因此，

$$\mathrm{TV}(p, \gamma_n|_K) \le \frac{2 d_n}{1 - d_n} \to 0.$$

但这意味着，我们可以完整地重复 §3 中的论证，只需将高斯分布截断在 K，并得出相同的结论。

我们能否将非平凡的叉凸组合推广到所有概率律？ 如上所述，由于叉凸组合定义中的“$Z_T^{\mathsf{Q}} = 0 \implies h = 1$”条件，无法在序贯概率律之间任意进行叉凸组合。以极端情形为例，对于支撑不相交的 $\mathsf{P} = P^\infty$ 和 $\mathsf{Q} = Q^\infty$（P、Q），唯一可能的叉凸组合是形如 $\alpha P^\infty + (1-\alpha) Q^\infty$ 的混合。尽管这一技术问题并未对本文造成严重影响，但总体而言，这种情况相当令人不满。毕竟，叉凸组合的算法视角十分自然，而且很容易推广到支撑不相交的情形。

可以利用条件密度将这一算法图景形式化。对于一列（适当可测的）映射 $k_t^{\mathsf{P}} \colon (\mathbb{R}^d)^{t-1} \times \mathbb{R}^d \to \mathbb{R}_{\ge 0}$，记作 $k_t^{\mathsf{P}}(x_t | x_1^{t-1})$，如果对每个 x_1^{t-1}，$k_t(\cdot | x_1^{t-1})$ 都是相对于 Γ 的密度，且对于任意 t，$A \in \mathscr{F}_t$，则称 $\{k_t^{\mathsf{P}}\}$ 是 P 的条件密度过程。

$$\mathsf{P}(X_1^t \in A) = \int_A \prod_{s \le t} k_s(x_s | x_1^{s-1}) \Gamma(\mathrm{d}x_1^t).$$

更一般地，我们可以通过马尔可夫核定义类似的概念。根据定义，如果 P 存在条件密度过程，那么对于任意 t，以及 Γ-几乎处处的 x_1^t，都有

$$Z_t^{\mathsf{P}}(x_1^t) = \prod_{s \le t} k_s(x_s | x_1^{s-1}).$$

利用上述刻画，我们可以给出叉凸组合的以下自然扩展定义：对于具有条件密度过程 $\{k_t^{\mathsf{P}}\}$、$\{k_t^{\mathsf{Q}}\}$ 的两个序贯概率律 P、Q，如果概率律 R 在时刻 T 是 P 和 Q 的叉凸组合，且权重 h 是 $\mathscr{F}_T$-可测的，则

$$Z_t^{\mathsf{R}} = \begin{cases} \prod_{s \le t} k_s^{\mathsf{P}}(x_s | x_1^{s-1}) & t \le T \\ \prod_{s \le T} k_s^{\mathsf{P}}(x_s | x_1^{s-1}) \cdot \left(h \prod_{s=T+1}^t k_s^{\mathsf{P}}(x_s | x_1^{s-1}) + (1-h) \prod_{s=T+1}^t k_s^{\mathsf{Q}}(x_s | x_1^{s-1}) \right) & t > T \end{cases}, \tag{4}$$

区别在于，我们现在不施加限制，即 $h = 1 \text{ if } Z_T^{\mathsf{Q}} = 0$. 简单来说，这是可行的，因为我们从未除以可能为原假设的 Z_T^{Q}；更严格地说，这是在考虑形式比值 $Z_t^{\mathsf{Q}} / Z_T^{\mathsf{Q}}$，并将其按自然方式解释为 $\prod k_s^{\mathsf{Q}}(x_s | x_1^{s-1})$。上述扩展定义概括了我们先前对叉凸组合的定义，我们也可以将其同样扩展到叉凸包及其闭包。

尽管上述密度过程是一个完全严谨的数学对象，但由于在一般情况下，这种扩展无法在这些扩展叉凸组合下保持 NSM 性质，因此这种扩展并无实际意义。

为了说明上述扩展定义为何无法保持 NSM 性质（不同于论文中使用的受限定义），请看以下示例。

例1. P = （均匀分布(0, 1)）$^\infty$ 且 Q = （均匀分布(1, 2)）$^\infty$，以及过程

$$M_t := \begin{cases} 2 & \exists s_1, s_2 \le t : X_{s_1} \in (0,1), X_{s_2} \in (1,2) \\ 1 & otherwise \end{cases}.$$

在P、Q下，该过程都是 *NSM*（事实上是鞅）。然而，在这两种概率律的任何非平凡凹凸组合下，该过程都必须从1开始，并以正概率增长到2，但此后永不下降，因此不可能是超鞅。

究其根本，上述例子中的问题源于这样一个事实：在扩展定义下，对于 $t \ge T+1$，$\{Z_t^{\mathtt{R}} > 0\} = \{Z_t^{\mathtt{P}} > 0\} \cup \{Z_T^{\mathtt{P}} > 0, \prod_{T+1}^{t} k_s(X_s | X_1^{s-1}) > 0\}$，但 $\{M_t\}$ 在 P 或 Q 下的 NSM 性质只控制 $M_t Z_t^{\mathtt{P}} \mathbb{1}\{Z_t^{\mathtt{P}} > 0\}$ 和 $M_t Z_t^{\mathtt{Q}} \mathbb{1}\{Z_t^{\mathtt{Q}} > 0\}$ 在 Γ 下的条件期望；因此，当 R 对某个概率律下为零的事件赋予正测度时，$M_t Z_t^{\mathtt{R}}$ 的条件行为便不受控制。

$\widetilde{\ }$需要指出的是，在上面的例子中，过程 $\{M_t\}$ 存在一个过程版本，即一个过程 $\{M_t\}$ $\widetilde{\ }$，使得 P($\forall t, M_t = M_t$) = Q($\forall t, M_t = M_t$) = 1，但即使在扩展的$\widetilde{\ }$叉凸组合下，$\{M_t\}$ 仍是一个鞅。具体而言，这个过程就是 $M_t = 1.$。因此，我们可能会想知道这一现象是否在更一般的情形下成立：如果 $\{M_t\}$ 在 P 和 Q 下都是 NSM，那么是否存在它的一个过程版本 $\widetilde{\ }\{M_t\}$（在 P 和 Q 下），使得 $\{M_t\}$ 对 P 和 Q 的任意扩展叉凸组合都是 NSM，而不受“$Z_T^{\mathtt{Q}} = 0 \implies h = 1$”这一限制？以下例子表明，一般而言这同样是不可能的。

例2。令 P = 均匀分布(0,1)$^\infty$，且 Q = 均匀分布(0, 1/2)$^\infty$。定义 $\rho_t = \mathbb{1}\{X_t \in (0, 1/2)\}$，其中 $t \ge 1$，且 $\rho_0 = 0$。令 $\{N_t\}$ 为一个适应过程，使得

$$N_t = \begin{cases} 1 & \rho_{t-1} = 1 \\ 3/2 & \rho_{t-1} = 0, \rho_t = 1 \\ 1/2 & \rho_{t-1} = 0, \rho_t = 0 \end{cases}.$$

最后定义 $M_t = \prod_{s \le t} N_t$。很容易检查出，在P 和Q下，$M_t$ 都是一个非负鞅。

现在假设R 是在时间 $T \ge 1$ 、混合权重为 $h < 1.$ 时，由P、Q 进行的扩展叉凸组合。这意味着，以概率 $1-h$，对所有 $t \ge T+1$，都有 $X_t \in (0,\ 1/2)$ 必然成立。因此，我们可以明确计算出

$$\mathbb{E}[N_{T+1} | \mathscr{F}_T] = \rho_T + (1-\rho_T)\left((1-h)\cdot 3/2 + h(1/2 \cdot 3/2 + 1/2 \cdot 3/2)\right) = \begin{cases} 1 & \rho_T = 1 \\ 1 + (1-h)/2 & \rho_T = 0 \end{cases},$$

因此，只要 $h < 1$，$\mathbb{E}[N_{T+1} | \mathscr{F}_T] > 1$ 若 $\rho_T = 0$，那么 $\{M_t\}$ 就在时间 $T+1$下违反了R的 *NSM* 性质。这里需要注意的是，由于P的概率律支配Q的概率律，因此很难构造出上述过程的任何非平凡不同版本。

鉴于上述讨论，叉凸组合的广义定义与维持这些组合的 NSM 性质相冲突。当然，由于我们使用叉凸性的目的是断言 NSM 在大类序贯概率律上的平凡性，因此，对于此类统计应用而言，维持后一性质至关重要。同时，尽管原始的受限定义确实维持了 NSM 性质，但其中包含的限制并不令人满意，而且与这些组合理念背后的算法直觉相冲突。寻找一种适当的叉凸组合广义定义，在不强加这些支撑条件的情况下，仍能在 NSM 性质下保留 NSM 闭包，是一个有趣且富有挑战性、留待未来工作的问题。

B 一致性证明与功效分析

回顾记号 $\sigma_t(P) := \sum_{s \le t} \log\ p(X_s)\ - \log \hat{}p_t(X_s)$。本节的主要论证控制 $\sigma_t(P)$ 的行为，特别是论证了如果 P 与对数凹性的海林格距离较大，那么

$\sigma_t(P)$ 最终必然呈线性增长。我们分别在第 §B.1 节和第 §B.2节中展示了这一点，分别针对渐近情形和非渐近情形。

推论 11 和推论 15 都还依赖于对以下情形下 $\rho_t(\mathscr{E}; p) = \sum_{s\le t} \log p(X_s) - \log \hat{q}_{s-1}(X_s)$ 行为的进一步控制，其中 p 是单位盒上的有界利普希茨概率律。此论证留待 §B.3。

B.1 一致性证明

我们的论证依赖于以下括号尾部估计，该估计由 Wong 和 Shen 提出，用于分析基于筛子的极大似然估计的行为 [WS95，定理 1]。该估计涉及一个概率律类 $\mathcal{Q}$ 在海灵格度量下的括号熵。关于 Van der Vaart 和 Wellner [VW96] 的著作可为读者提供全面介绍，下面我们简要说明。

括号 $[u, v]$ 由两个函数 $u(x)$ $v(x)$定义，满足对所有 x， $u(x) \le v(x)$，并且由所有满足对任意 x， f 使得 $u(x) \le f(x) \le v(x)$ 的函数组成。由于我们只关注密度函数，可以将注意力限制在非负函数上。此类括号的海灵格大小$[u, v]$ 定义为 $|[u, v]| = \|\sqrt{u} - \sqrt{v}\|_2/2.$。若对所有 $Q \in \mathcal{Q}$，都存在一个 i 使得 $q \in [u_i, v_i]$，则称分布类 $\mathcal{Q}$ 被 $\{[u_i, v_i]\}_{i=1}^N$括住；回顾一下，对于分布 Q，我们用 q表示其密度。注意，这种括号通常是“不适当的”，即 u_i, v_i 通常不属于 $\mathcal{Q}$ （因为q 的积分为一，因此其下界括号 u 的积分小于一，而上界括号 v 的积分大于一）。 $\mathcal{Q}$ 在尺度 ζ 下的海灵格括号熵，是用大小至多为 ζ的括号对 $\mathcal{Q}$ 进行最简括住时所需括号数量的对数，即

$$\mathcal{H}_{[]}(\mathcal{Q}, \zeta) := \inf\{\log N : \mathcal{Q} \text{ has an } N\text{-sized Hellinger bracketing at scale } \zeta\}$$

注意，括号熵关于 ζ 是非增的。

引理 17. (对[WS95，定理 1]的简化）对于分布类 $\mathcal{Q}$ 和自然数 t，定义ε_t 为满足以下条件的较小数 ε ：

$$\int_{\varepsilon^2/2^8}^{\sqrt{2}\varepsilon} \sqrt{\mathcal{H}_{[]}(\mathcal{Q}, \zeta/10)}\mathrm{d}\zeta \le 2^{-11}\sqrt{t}\varepsilon^2.$$

对于每个 t 和 $\varepsilon \ge \varepsilon_t$，对任意满足 $d_H(P, \mathcal{Q}) \ge \varepsilon$的概率律 P，均有

$$P^{\otimes t}\left(\inf_{q\in\mathcal{Q}} \sum_{s\le t} \log p(X_s) - \log q(X_s) \le t\varepsilon^2/24\right) \le 4\exp\left(-Ct\varepsilon^2\right),$$

其中 $C > 2^{-14}$ 是一个常数。

非正式地说，如果 P 与 $\mathcal{Q}$ 在海灵格度量下相距足够远（这里足够远的程度由 $\mathcal{Q}$的括号熵决定），那么在样本量意义下， $\mathcal{Q}$ 下的最大对数似然与 P下的对数似然呈线性接近的可能性就呈指数级地小。要在我们的情形中利用这一观察结果，需要论证对数凹最大似然估计$\hat{p}_t$ 最终必然落在一个熵较小的集合中。为此，我们采用 Dunn 等人的以下结果，该结果扩展了 Cule 和 Samworth[CS10]的收敛分析。

引理 18. [DGWR21，引理 1] 考虑任意分布 $P \in \mathcal{D}_1$，不一定是对数凹的。对于任意$\eta > 0$，存在一个检验尺寸至多为 $[u_\eta, v_\eta]$ 的括号 η，其中包含对数凹投影 $\mathfrak{L}_P$，并且最终也包含对数凹最大似然估计$\hat{p}_t$： $P^\infty(\exists t_0 : \forall t \ge t_0, \hat{p}_t \in [u_\eta, v_\eta]) = 1.$

换言之，该引理指出，当 t足够大时，对数凹最大似然估计$\hat{p}_t$ 必定落在真实分布 $\mathfrak{L}_P$ 的对数凹投影 P附近一个非常小的括号内。有了这一结果，我们便可以证明引理12，它是定理10证明中的核心结论。

引理的证明 *12*。 令 $\varepsilon := d_H(P, \mathfrak{L}_P) > 0.$定义 $\eta_\varepsilon = \varepsilon^2/2^{11}.$利用引理 18，我们知道存在一个括号 $[u^*, v^*]$ 使得 $|[u^*, v^*]| \le \eta_\varepsilon$并且，几乎必然地，$\hat{p}_t \in [u^*, v^*]$ 对于所有充分大的 t。但注意到

由于 $\mathcal{H}_{[]}([u^*, v^*], \varepsilon^2/2^{11}) = 0$，因为 $[u^*, v^*]$ 的检验尺寸已经是 η_ε。此外，由于 $\mathfrak{L}_P \in [u^*, v^*]$，根据三角不等式，

$$d_H(P, [u^*, v^*]) = \inf_{Q \in [u_\eta, v_\eta]} d_H(P, Q) \geq d_H(P, \mathfrak{L}_P) - |[u^*, v^*]| \geq \varepsilon(1 - 2^{-11}) \geq \varepsilon \cdot \sqrt{24/25}.$$

定义

$$\widetilde{\sigma}_t(P) := \inf_{q \in [u^*, v^*]} \sum_{s \leq t} \log p(X_s) - \log q(X_s).$$

利用上述观察，引理 17 可得，对任意 t，

$$P^\infty \left(\widetilde{\sigma}_t(P) \leq t\varepsilon^2/25 \right) \leq 4 \exp\left(-Ct\varepsilon^2\right).$$

进一步注意，若 $\hat{}p_t \in [u^*, v^*]$，由于 $\hat{}p_t$ 是最大似然估计量，因此必有$\sigma_t(P) = \widetilde{\sigma}_t(P)$。令 $\mathsf{E}_s := \{\forall t \geq s, \hat{}p_t \in [u^*, v^*]\}$ 表示 $\hat{}p_t$ 在时间 s 之后落入小括号内的事件，并令 $\mathsf{A}_t := \{\sigma_t(P)/t\varepsilon^2 \geq 1/25\}$ 表示 $\sigma_t(P)$ 大于 $t\varepsilon^2/25$ 的事件。

由引理 17，对每个固定时刻 s 和 $t \geq s$，

$$P^\infty(\mathsf{A}_t^c \cap \mathsf{E}_s) \leq 4 \exp\left(-Ct\varepsilon^2\right),$$

由于此上界可求和，根据 Borel-Cantelli 引理

$$0 = P^\infty\left(\limsup_t (\mathsf{A}_t^c \cap \mathsf{E}_s)\right) = P^\infty\left((\limsup_t \mathsf{A}_t^c) \cap \mathsf{E}_s\right),$$

因此，对于任意时刻 s，

$$P^\infty(\limsup_t \mathsf{A}_t^c) \leq P^\infty(\limsup_t \mathsf{A}_t^c \cap \mathsf{E}_s) + P^\infty(\mathsf{E}_s^c) = P^\infty(\mathsf{E}_s^c).$$

根据引理18，$\hat{}p_t$ 最终几乎必然落入 $[u^*, v^*]$，$\lim_{s\to\infty} P^\infty(\mathsf{E}_s^c) \to 0.$。进一步注意到

$$\limsup_t \mathsf{A}_t^c = \{\sigma_t(P)/t\varepsilon^2 < 1/25 \text{ infinitely often}\} = \{\liminf \sigma_t(P)/t\varepsilon^2 < 1/25\}.$$

综合这些观察，令 $s \to \infty$ 趋于无穷后，我们得出

$$P^\infty(\liminf \sigma_t(P)/t\varepsilon^2 < 1/25) \leq \lim_{s\to\infty} P^\infty(\mathsf{E}_s^c) = 0. \qquad \square$$

B.2 检验功效分析的证明

我们先陈述论证所依据的关键引理。这些引理利用引理13 中的括号熵控制，并结合文献中的结果——这些结果对对数凹密度所能达到的最大值给出了界——以使该控制发挥同等效用。随后，我们证明主要结果，最后证明引理13。

B.2.1 控制对数凹最大似然估计所能达到的最大值

该速率分析定量地利用了引理 17。为此，我们首先需要对数凹分布的括号熵界，而这正是引理 13的主题。我们回顾一下，该引理控制了类 $\mathcal{L}_{d,\ B}$ 中的对数凹分布的括号熵；这些分布的密度支撑集位于 $[-1, 1]^d$ 上，且其上界为 B，从而表明

$$\mathcal{H}_{[]}(\mathcal{L}_{d,B}, \zeta) = \widetilde{O}((B/\zeta)^{\max(d/2,(d-1))}).$$

只要该常数不随相关参数缩放，B 在上述结论中的作用在定量上并不重要。对于协方差接近单位矩阵的对数凹分布，这一点可以得到保证。直观而言，由于协方差在所有方向上都有下界，这些分布就不会过度集中，因此众数处的密度值不会过大。Kim & Samworth 的工作直接蕴含了以下结果，概括了这一观察。

引理 19. [KS16，推论 6] 令 $\mathcal{L}_d^\gamma$ 表示定义在 $[-1,1]^d$ 上的对数凹分布集合，其协方差矩阵在半正定序下以下界 γI.那么存在一个依赖于维度的常数 C_d使得对任意 $f \in \mathcal{L}_d^\gamma$，

$$\max_{x\in[-1,1]^d} f(x) \le \gamma^{-d/2} C_d.$$

当然，定理 14 中的界取决于 Δ_P，粗略地说，它只约束底层分布 P的协方差。这个量之所以重要，源于 Barber 和 Samworth 提出的以下观察。

引理 20。[BS21，推论 8] 设 $P \in \mathcal{D}_1$ 是一个支撑于 $[-1,1]^d$ 上的概率律，且满足

$$\Delta_P := \min_{v:\|v\|=1} \mathbb{E}_p[|\langle v, X - \mathbb{E}_p[X]\rangle|] > 0.$$

那么存在一个依赖于维数的常数 c_d，使得 $\mathrm{Cov}(\mathfrak{L}_P) \succeq c_d \Delta_P^2 I$. 此外，存在一个与维数无关的常数 C，使得对于任意 $t \ge 2Cd^3/\Delta_P^2$，以至少 $1 - 2\exp\left(-Ct\Delta_P^2/d^2\right)$ 的概率，有 $\mathrm{Cov}(\hat{}p_t) \succeq \frac{c_d\Delta_P^2}{4} I$，其中对数凹最大似然估计为 $\hat{}p_t$。

引理 *20* 的证明。 第一个观察是对 Barber 和 Samworth 的引理 7 的直接重述。第二个结论源于这样一个事实：在 v： $\|v\| = 1$， $v \mapsto \langle v, X - \mathbb{E}_P[X]\rangle$ 有界于 $2\sqrt{d}$，且显然关于 v连续。因此，利用单位球上的标准次高斯集中结果可知，以至少 $1-2\exp\left(-Ct\Delta_P^2/d^2\right)$ 的概率，经验分布 $p_t = \frac{1}{t}\sum_{s\le t}\delta_{X_s}$满足

$$\min_{v:\|v\|=1} \mathbb{E}_{p_t}[|\langle v, X - \mathbb{E}_{p_t}[X]\rangle|] \ge \Delta_P/2.$$

但请注意，$\hat{}p_t = \mathfrak{L}_{p_t}$，由此可得出该结论。 □

合并引理19 和20 可立即得到以下观察结果，它为引理 B 13中所需采用的尺度提供了一个具体界。

引理 21。存在一个常数 C_d，它仅取决于 d，使得对任意 $t \ge 2C_d d^3/\Delta_P^2$，以下结论以至少 $1 - 2\exp\left(-C_d t\Delta_P^2/d^2\right)$ 的概率成立：

$$\max_{x\in[-1,1]^d} \hat{p}_t(x) \le C_d \Delta_P^{-d}.$$

证明。应用引理19，我们观察到 $\{\max \hat{}p_t \le (c_d\Delta_P^2/4)^{-d/2}\} \subset \{\mathrm{Cov}(\hat{}p_t) \preceq c_d\Delta_P^2/4 I\}$，且后者以至少 $1-2\exp\left(-t(C\Delta_P^2/d^2)\right)$ 的概率成立，其中$t \ge 2Cd^3/\Delta_P^2$。取 $C_d = \max(C, c_d^{-d/2})$。 □

B.2.2 拒绝时间界的证明

有了上述结果，我们可以继续进行主要论证。
定理 *14* 的证明。回忆定义 $\sigma_t := \sum_{s\le t}\log p(X_s) - \log \hat{}p_t(X_s)$。我们首先证明 σ_t以高概率有下界。

令 B 为一个稍后选定的量。令 ε_t 表示引理 17 中不动点方程的解，其中代入的是 $\mathcal{L}_{d,B}$ 的括号熵。此外，定义事件

$$\mathsf{E}_t := \{\hat{p}_t \in \mathcal{L}_{d,B}\}.$$

对于任意 t，只要满足 $\varepsilon_t \le d_H(p,\mathcal{L})$ 且 $\hat{}p_t \in \mathcal{L}_{d,\ B}$，引理 17 可得

$$\sigma_t \ge \inf_{q\in\mathcal{L}_{d,B}: d_H(p,q)\ge d_H(p,\mathcal{L})} \log \prod_{s\le t}\frac{p(X_s)}{q(X_s)} \ge \frac{t d_H^2(p,\mathcal{L})}{24} \tag{5}$$

以至少 $1 - \exp\left(-Ctd_H^2(p,\mathcal{L})\right) - P^\infty(\mathsf{E}_t^c)$ 的概率。

在证明的其余部分，我们将确定 t 的取值范围，使 ε_t 足够小，从而确保条件 $\varepsilon_t \le d_H(p,\mathcal{L})$ 得到满足，同时控制 $P^\infty(\mathsf{E}_t^c)$。为此，我们应用引理 13。首先注意到，对于 $d \ge 3$ 以及任意正常数 c 和 C

$$\int_{c\varepsilon^2}^{C\varepsilon} \sqrt{\tilde{O}(B^{d-1}\zeta^{-(d-1)})}\mathrm{d}\zeta = \widetilde{O}(B^{(d-1)/2}\varepsilon^{-(d-3)}).$$

注意，多对数项不会影响积分的主要增长。因此，求解不动点方程

$$\widetilde{O}(B^{(d-1)/2}\varepsilon^{-(d-3)}) = \varepsilon^2 t^{1/2},$$

我们得到，对于 $d \ge 3$

$$\varepsilon_t(B) = \widetilde{O}(B^{1/2}t^{-1/2(d-1)}),$$

其中我们强调了对尚未确定的量 B 的依赖。

利用熵界 $\zeta^{-d/2}$ 进行类似的论证，可得对于 $d \in \{1,2\}$，$\varepsilon_t(B) = \widetilde{O}(B^{d/(d+4)}t^{-2/(d+4)})$。现在定义

$$T_1(B) = \inf\{t : \varepsilon_t(B) \le d_H(p,\mathcal{L})\}$$

并注意到

$$T_1(B) = \begin{cases} \widetilde{O}(B^{(d-1)}(d_H(p,\mathcal{L}))^{-2(d-1)}) & d \ge 3 \\ \widetilde{O}(B^{d/2}(d_H(p,\mathcal{L})^{-(4+d)/2})) & d \in \{1,2\} \end{cases}.$$

最后，由引理 21，对于 $B \ge C_d\Delta_P^{-d}$ 和 $t \ge T_2 := C\Delta_P^2/d^2$ 事件 E_t 的概率至少为$1 - 2\exp\left(-tC\Delta_P^2/d^2\right)$. 令 $B_* = C_d\Delta_P^{-d}$ 并令 $T_0 := \max(T_1(B_*)\ T_2)$。我们得到下界

$$\sigma_t \ge \frac{td_H^2(p,\mathcal{L})}{24}$$

以至少 $1-C\exp\left(-tcd_H^2(p,\mathcal{L})\right)-C\exp\left(-tc\Delta_P^2/d^2\right)$ 的概率成立，其中 $t \ge T_0$。现在注意，在任意时刻 $t \ge \max(T_0, {}^{600}\log(1/\alpha)\, d_H^2(p,\mathcal{L}))$，它以至少 $1-\pi_t-C\exp\left(-td_H^2(p,\mathcal{L})\right)-C\exp\left(-Ct\Delta_P^2/d^2\right)$ 的概率成立，即

$$\log R_t = \sigma_t - \rho_t \ge \frac{td_H^2(p,\mathcal{L})}{600} \ge \log(1/\alpha),$$

因此，拒绝时间 $\tau_\alpha := \inf\{t : R_t \ge 1/\alpha\}$ 超过上述界限的概率至多为 $\pi_t + C\exp\left(-td_H^2(p,\mathcal{L})\right) + C\exp\left(-Ct\Delta_P^2/d^2\right)$。 □

B.2.3 对数凹分布上括号熵界的证明

接下来我们证明引理 13。我们注意到，$d \le 3$ 的上界已由 Kim 和 Samworth 证明[KS16]。下面我们重点讨论 $d \ge 4$。我们将利用文献中关于凸集和函数的两个已有结果。第一个结果主要归功于 Bronshtein（另见[KDR19，引理 3]）。

引理 22。 [Bro76] 令 $\mathcal{K}_d$ 表示 $[-1,1]^d$中凸集的集合。对于任意 $\zeta > 0$，存在一组凸集对 $\mathcal{K}_{d,\zeta} \subset \mathcal{K}_d \times \mathcal{K}_d$ 满足 $\log|\mathcal{K}_{d,\zeta}| = O(\zeta^{-(d-1)/2})$ 使得

- 每个$(\underline{K}, \overline{K}) \in \mathcal{K}_{d,\zeta}$都满足 $\mathrm{Leb}_d(\overline{K} \setminus \underline{K}) \le \zeta$.
- 对于每个 $K \in \mathcal{K}_d$，都存在$(\underline{K}, \overline{K}) \in \mathcal{K}_{d,\zeta}$ 满足 $\underline{K} \subset K \subset \overline{K}$.

换言之，凸集在集合差度量下的括号熵以 $(d-1)/2$ 的速率受控。值得注意的是，上述构造的括号是适当的。Gao 和 Wellner 给出的关于凸函数的以下括号熵界，可视为这一结果的推广。

引理 23。 [GW17，定理 1。5] 设 K 是 $[-1,1]^d$ 中的一个凸集，且设 $\mathcal{C}_{K,B}$ 是在 K 上被 B 上界所限制的凸函数集合。那么，$\mathcal{C}_{K,B}$ 在尺度 ζ 下的 $L_2(K)$ 括号熵有如下上界：$O((B/\zeta)^{(d-1)})$。

[5]这可以通过反复应用关系式 $\int x^n \log^m x = {}^{x^{n+1}}\log_{n+1}{}^{m}\, x - \frac{m}{n+1}\int x^n \log^{m-1}(x)$ 得出。

上文中，$L_2(K)$ 度量是通常的 L_2 距离 $\|f-g\|_{L_2(K)} = (\int_K (f-g)^2 \mathrm{d}x)^{1/2}$，而 $L_2(K)$ 括号熵是括号大小为 $[u,v]$ $|[u,v]| = \|u-v\|_{L_2(K)}$ 时的括号熵。

有了上述结果，我们现在可以继续证明。

引理 13 的证明。对于任意对数凹分布 f，令 $S := \{x \in \mathbb{R}^d : f(x) \geq \zeta^3\} = \{x \in \mathbb{R}^d : \log f(x) \geq 3\log \zeta\}$. 由于 f 是对数凹的，集合 S 是凸集。因此，根据引理 22，存在某个凸集 $S \in \mathcal{K}_{d,\zeta^2/B}$，使得 $\mathrm{Leb}(S \setminus \tilde{S}) \leq \zeta^2/B$ 且 $\tilde{S} \subset S$。令 $\tilde{\mathcal{C}}_{\tilde{S},\zeta,B}$ 表示在 S 上由 B 界定的凸函数的一个 ζ-括号。由于在 S 上，函数 $-\log f$ 是凸的，且上界为 $-\log B$，根据引理 23，存在一个括号 $[-u,-l] \in \tilde{\mathcal{C}}_{\tilde{S},\zeta/B,\log B/\zeta^3}$，使得在 $\tilde{S}$ 上，$l \leq \log f \leq u$，并且 $\int_{\tilde{S}} (u(x)-l(x))^2 \mathrm{d}x \leq \zeta^2/B^2$。注意，在 S 上，f 有下界 $-3\log\zeta$；并且不失一般性地，我们可以假设 $\sup_{x\in\tilde{S}} u(x) \leq \log B$，因为这已经是 f. 的逐点上界

接下来，我们构造这些函数

$$x \in [-1,1]^d \mapsto U(x) := \begin{cases} e^{u(x)} & x \in \tilde{S} \\ B & x \in S \setminus \tilde{S} \\ \zeta^3 & x \in [-1,1]^d \setminus S \end{cases},$$

$$x \in [-1,1]^d \mapsto L(x) := \begin{cases} e^{l(x)} & x \in \tilde{S} \\ \zeta^3 & x \in S \setminus \tilde{S} \\ 0 & x \in [-1,1]^d \setminus S \end{cases}.$$

注意，在 $[-1,1]^d$ 上有 $U \geq f \geq L$。此外，对于 $\zeta < 2^{-d}$，

$$\begin{aligned} \int (\sqrt{U}-\sqrt{L})^2 \mathrm{d}x &= \int_{\tilde{S}} (e^{u(x)/2} - e^{l(x)/2})^2 \mathrm{d}x + \int_{S\setminus\tilde{S}} B \mathrm{d}x + \int_{[-1,1]^d\setminus S} \zeta^3 \mathrm{d}x \\ &\leq \int_{\tilde{S}} e^{u(x)} (1 - e^{u(x)-l(x)/2})^2 \mathrm{d}x + B\cdot\zeta^2/B + 2^d\zeta^3 \\ &\leq \int_{\tilde{S}} B^2 (u(x)-l(x))^2/4 \,\mathrm{d}x + 2\zeta^2 \\ &\leq B^2\cdot\zeta^2/B^2 + 2\zeta^2 = 3\zeta^2, \end{aligned}$$

这里我们利用了这样一个事实：$z \mapsto e^{z/2}$ 在 $[-\infty, \log C]$ 上是利普希茨的，且导数有界于 $e^{(\log B)/2}/2 = \sqrt{B}/2$，从而得出 $e^{(u(x)-l(x))/2} - e^0 \leq \sqrt{B}|u(x)-l(x)-0|/2$。

由于对于任意 f，都可以进行这一构造，因此我们可以构造一个 $\mathcal{L}_{d,B}$ 尺度为 $O(\zeta)$ 的括号覆盖，它是各较小集合上凸函数括号覆盖的并集，这些较小集合见 $\mathcal{K}_{d,\zeta^2/B}$。根据引理 22 和 23，该覆盖的检验尺寸为

$$\exp\left(O((B/\zeta)^{d-1})\right) \cdot \exp\left(O((\log B/\zeta^3)(B)/\zeta)^{d-1}\right) = \exp\left(\widetilde{O}((B/\zeta)^{d-1})\right),$$

由此便得证。我们再次注意到，所得覆盖是不适当的，因为映射 $U(x)$ 和 $L(x)$ 不是对数凹的。

□

B.3 单位区间上有界利普希茨概率分布的遗憾控制

本小节将说明，推论15 和11 都依赖于这样一个论证：$\mathcal{D}_{\mathrm{Box,Lip,B}}$ 中的概率律可以通过在线方式以低遗憾进行估计。我们通过利用以下结果来论证这一点，该结果是对 Wong 和 Shen 关于筛估计量的结果进行简化后得到的。

引理 24。（改编自 [WS95，推论 1 和定理 6]）对于每个 $P \in \mathcal{D}_{\mathrm{Box,Lip,B}}$ 和 $t \geq 1$，存在一个筛最大似然估计 $\hat{q}(\cdot) = \hat{q}(\cdot; X_1^t)$ 以及一个常数 $A > 1$，该常数仅取决于 B，使得对于每个 $\zeta \geq \zeta_t$，

$$P^\infty\left(\mathrm{KL}(p\|\hat{q}) > \frac{1}{A}\zeta^2\log(1/\zeta)\right) \leq A\exp\left(-t\frac{\zeta^2}{A\log(1/\zeta)}\right),$$

其中 $\zeta_t = \widetilde{O}\left(t^{-1/2(d+2)}\right)$。

证明。 Wong 和 Shen 的相关结果适用，因为 $\mathcal{D}_{\text{Box、 Lip、 B}}$ 中的概率律的密度均有统一的上界。这直接得出该陈述的全部内容，但 ζ_t 的尺度界除外。该尺度由引理 17 中出现的同一个熵积分不动点方程确定；对于此情形，利用以下标准事实即可推导出该界：在尺度 η 下，盒上的 Lipschitz函数的 Hellinger括号熵满足 $O(\eta^{-(d+1)})$ [Vaa94]。 □

该结果中的筛估计量可以相当宽松地选取。具体而言，一种明确的选择方法是：对每个 $\zeta > 0$，在尺度 ζ 下构造类 $\mathcal{D}_{\text{Box, Lip, B}}$ 的一个括号，并从该类的每个括号中选取一个代表性密度。随后，筛最大似然估计需要在每个时刻选取一个 ζ，并在上述代表性密度中选择似然最大的密度来估计概率律。对我们而言，重要的是，这些括区中的下括界可以取为一致地大于 $1/B$，而上括界可以取为小于 B，因为 $p \in [1/B, B]$，因此筛估计量一致地介于 $1/B$和 B之间。

下面我们首先利用上述结果证明推论15 ，然后说明推论11 可由这一论证直接推出。

推论*15*的证明。如正文所述，期望拒绝时间有如下界： $\mathbb{E}[\tau] \le \sum \pi_t + O(T_0)$，其中 $T_0 = o(d_H(p, \mathcal{L})^{-2(d+3)})$。因此，我们只需证明存在一个序列 π_t 使得 $\sum \pi_t$ 足够小，并且对于任意 t,

$$P^\infty\left(\frac{\rho_t(\mathscr{E}; p)}{t d_H^2(p, \mathcal{L})} \ge \frac{1}{25}\right) \le \pi_t,$$

其中 $p \in \mathcal{D}_{\text{Box, Lip, B}}$ 和 $\mathscr{E}$ 是筛估计量。下面我们将继续进行。

为简洁起见，我们定义 $\varepsilon = d_H(p, \mathcal{L})$。令 A 为引理 24中的常数，并设

$$T_1 := \min\{t : \zeta_t^2 \log(1/\zeta_t)/A < \varepsilon^2/200, \zeta_t < 1/\sqrt{e}\}.$$

进一步令

$$\zeta(\varepsilon) := \max\{\zeta \in [0, 1/\sqrt{e}] : \zeta^2 \log(1/\zeta) \le A\varepsilon^2/200\}.$$

在接下来的证明中，我们将对 $\zeta = \zeta(\varepsilon) \ge \zeta_{T_1}$ 使用引理 24。为此，我们注意到，如果 $\zeta(\varepsilon) < 1/\sqrt{e} \iff A\varepsilon^2/200 < 1/2e$，则此时等式 $\zeta(\varepsilon)^2 \log(1/\zeta(\varepsilon)) = A\varepsilon^2/200$ 成立。由此，我们可以推出[6] 对于某个足够小的常数 a，有 $\zeta(\varepsilon)^2 > a\varepsilon^2/\log(1/\varepsilon)$，因此引理 24 上界中的指数为

$$\frac{\zeta(\varepsilon)^2}{A \log(1/\zeta(\varepsilon))} = \frac{400\zeta^4}{A^2\varepsilon^2} \ge \frac{\varepsilon^2}{A' \log(1/\varepsilon)}$$

对于某个足够大的常数 A'. 。我们还将假设 $A' \ge \max(1, A)$。

令 $\mathscr{E}$为一组筛估计量，使得对每个 $t, x \in [-1, 1]^d$，$\hat{q}_{t-1}(x) \in [1/B, B]$，根据上文的讨论，这一点可以保证。注意，由于数据 $\{X_t\}$ 的数据独立性，对于任意 t,

$$\mathbb{E}[\log p(X_t)/\hat{q}_{t-1}(X_t) | \mathscr{F}_{t-1}] = \mathrm{KL}(p \| \hat{q}_{t-1}).$$

令 $\theta \in (0, 1)$ 和 $M \ge 0$ 为两个稍后将设定的论证参数。考虑某个 $\tau \ge MT_1$ 时 $t = T_1 + \tau$ 的情形。

由于对每个 $\tau > 0$，$\zeta_{T_1+\theta\tau} \le \zeta_{T_1} \le \zeta(\varepsilon)$，引理 24 的界在每个时刻 $s \in [T_1 + \theta\tau : T_1 + \tau]$ 且 $\zeta = \zeta(\varepsilon)$ 时均有效。因此，将引理 24 应用于此范围内的每个 s，并利用上文确立的 $\zeta(\varepsilon)^2$ 的性质，

$$P^\infty(\mathrm{KL}(p \| \hat{q}_{s-1}) > \varepsilon^2/200) \le A' \exp\left(-s \frac{\varepsilon^2}{A' \log(1/\varepsilon)}\right).$$

[6]该方程等价于当 $x = \zeta(\varepsilon)^2$，$y = A\varepsilon^2/100$ 处于范围 $0 < x < 1/e$ 内时，$x \log x = y$。注意映射 $x \mapsto x \log(1/x)$ 在 [0,1/e], 上单调递增，并验证对于 $y \in [0, 1/2e]$, 2 $\frac{y}{\log(1/y)} \cdot \log(2 \log(1/y)/y) < y.$，即可得到该结论。事实上，这个不等式等价于证明 $\log(2 \log(1/y)) < \log(1/y) \iff y \log(1/y) < 1/2$，而这成立是因为 $y \mapsto y \log(1/y)$ 的最大值为 $1/e < 1/2$。

接下来，根据应用对 $s \in [T_1+\theta\tau$ 应用并集界：将上面的结果 $T_1+\tau]$ 代入，我们得出

$$\begin{aligned}
&P^\infty\left(\exists s \in [T_1+\theta\tau : T_1+\tau] : \mathrm{KL}(p\|\hat{q}_{s-1}) > \frac{\varepsilon^2}{200}\right) \\
&\quad\le \sum_{s=T_1+\theta\tau}^{T_1+\tau} A' \exp\left(-s\frac{\varepsilon^2}{A'\log(1/\varepsilon)}\right) \\
&\quad= A'\exp\left(-(T_1+\theta\tau)\frac{\varepsilon^2}{A'\log(1/\varepsilon)}\right)\cdot\frac{1}{1-\exp\left(-\varepsilon^2/(A'\log(1/\varepsilon)\right)}, \\
&\quad\le \frac{A'^2\log(1/\varepsilon)}{\varepsilon^2}\exp\left(-\theta\tau\frac{\varepsilon^2}{A'\log(1/\varepsilon)}\right).
\end{aligned}$$

其中等号是对几何级数求和所得，最后一个不等式利用了 $T_1 \ge 0$ ，并且对于$u<1$， $1/\left(1-e^{-u}\right) \le \frac{2}{u}$。

接下来，注意到对于任意 $x \in [-1,1]^d$， $\frac{1}{B^2} \le \frac{p(x)}{\hat{q}_{t-1}(x)} \le B^2$，我们有界 $|\log(p(X_t)/\hat{q}_{t-1}(X_t))| \le 2\log B.$。因此，Azuma-Hoeffding 不等式适用，并可得对每个 $\tau \ge 1$， $\delta > 0$

$$P^\infty\left(\sum_{s=T_1+\theta\tau}^{T_1+\tau}\log\frac{p(X_s)}{\hat{q}_{s-1}(X_s)} > \sum_{s=T_1+\theta\tau}^{T_1+\tau}\mathrm{KL}(p\|\hat{q}_{s-1}) + (\tau-\theta\tau)\delta\right) \le \exp\left(-(\tau-\theta\tau)\delta^2/8\log^2 B\right).$$

我们通过在上述结果中令 $\delta = \varepsilon^2/200$ 并应用并集界，得出存在一个常数 C 使得

$$P^\infty\left(\sum_{s=T_1+\theta\tau}^{T_1+\tau}\log\frac{p(X_s)}{\hat{q}_{s-1}(X_s)} > \frac{(1-\theta)\tau\varepsilon^2}{100}\right) \le \exp\left(-\frac{(1-\theta)\tau\varepsilon^4}{C\log^2 B}\right) + \frac{C\log(1/\varepsilon)}{\varepsilon^2}\exp\left(-\theta\tau\frac{\varepsilon^2}{C\log(1/\varepsilon)}\right). \quad (6)$$

我们将（6）的右侧记为 $\pi(\tau, \theta)$。根据 ρ_t的定义，以及对每个 s，$\log \frac{p(x)}{\hat{q}_{s-1}(x)}$ 的有界性可知，以至少 $1-\pi(\tau, \theta)$的概率，

$$\begin{aligned}
\rho_{T_1+\tau}(\mathscr{E};p) &= \sum_{s=1}^{T_1+\tau}\log\frac{p(X_s)}{\hat{q}_{s-1}(X_s)} \\
&\le 2(T_1+\theta\tau)\log B + \frac{(1-\theta)\tau\varepsilon^2}{100}.
\end{aligned}$$

只要我们能够选择 θ， M使得上述上界小于 $(\tau+T_1)\varepsilon^2/25$，不等式（6）便会限制 $\rho_{T_1+\tau} > \varepsilon^2\left(T_1+\tau\right)/25$ 的概率，这正是我们的目标。但请注意，当 $(\theta+1/M) \le 3\varepsilon^2/(200\log B)$ 时，确实满足这一条件，因为此时

$$\begin{aligned}
2(T_1+\theta\tau)\log B + \frac{(1-\theta)\tau\varepsilon^2}{100} &\le \tau\left(2(1/M+\theta)\log B + \frac{\varepsilon^2}{100}\right) \\
&\le \tau\left(\frac{3\varepsilon^2}{200\log B}\cdot 2\log B + \frac{\varepsilon^2}{100}\right) \\
&= \frac{\tau\varepsilon^2}{25} \le \frac{(T_1+\tau)\varepsilon^2}{25}.
\end{aligned}$$

因此，我们可以设 $\theta = \min(1/2, \varepsilon^2/(100\log B))$，并令 $M = \max(1, (200\log B)/\varepsilon^2)$，从而得出，对于任意 $\tau \ge MT_1$，都有

$$P^\infty(\rho_{T_1+\tau}(\mathscr{E};p)/t\varepsilon^2 > 1/25) \le \pi(\tau),$$

其中，对于常数 C'，

$$\pi(\tau) = \exp\left(-\frac{\tau\varepsilon^4}{C'\log^2 B}\right) + \frac{C'\log(1/\varepsilon)}{\varepsilon^2}\exp\left(-\frac{\tau\varepsilon^2}{C'\log(1/\varepsilon)\cdot\log B}\right).$$

注意，按照定理 14， $\pi_t = \pi(t - T_1)$ 对于 $t \geq (M+1)T_1$。当然，我们始终可以给出平凡的界 $\pi_t \leq 1$ 对于 $t < (M+1)T_1$. 接下来需要计算由此得到的期望拒绝时间界。为此，将相应的几何级数求和可得：

$$\begin{aligned}\sum_{t\geq 1}\pi_t &\leq (M+1)T_1 + \sum_{\tau=MT_1}^{\infty}\pi(\tau)\\ &\leq (M+1)T_1 + \frac{1}{1-\exp\left(-\varepsilon^4/C'\log^2 B\right)} + \frac{C'\log(1/\varepsilon)}{\varepsilon^2(1-\exp\left(-\varepsilon^2/C'\log(1/\varepsilon)\cdot\log B\right)}\\ &\leq O\left(\frac{1}{\varepsilon^2}\right)T_1 + \widetilde{O}\left(\frac{1}{\varepsilon^4}\right),\end{aligned}$$

其中， O 界如 $\varepsilon \to 0$所示，并且我们省略了对 B 和 $\log(1/\varepsilon)$ 的依赖。不过，由于引理 24 中， $\zeta_t = \widetilde{O}(t^{-1/2(d+2)})$，且 T_1 是满足 $\zeta_t^2\log^2(1/\zeta_t) \leq A\varepsilon^2/200$的首次时刻，我们可以得出 $T_1 = \widetilde{O}(\varepsilon^{-2(d+2)})$。注意到 $O(\varepsilon^{-2}) \cdot T_1 = \widetilde{O}(\varepsilon^{-2(d+3)})$，并回想起$\varepsilon = d_H(p, \mathcal{L})$，即可得到所述结论。 □

最后，我们简要证明推论11 ，并利用上述论证中得到的界。

推论 *11*的证明。 只需说明，使用推论 15证明中的估计量，对于任意$P \in \mathcal{D}_{\text{Box, Lip, B}}$,

$$P^{\infty}\left(\limsup\frac{\rho_t(\mathscr{E};p)}{td_H^2(p,\mathcal{L})} \leq \frac{1}{25}\right) = 1.$$

这是因为对于每个 t,

$$P^{\infty}\left(\frac{\rho_t(\mathscr{E};p)}{td_H^2(p,\mathcal{L})} > \frac{1}{25}\right) \leq \pi_t,$$

以及 $\sum\pi_t < \infty$，由博雷尔-坎特利引理可得上述关系。 a. □