arXiv:2603.08002v1 [math.ST] 9 Mar 2026

# 事后大样本统计推断

Ben Chugg*[1]、Etienne Gauthier*[2]、Michael I.Jordan[2,3]，
Aaditya Ramdas[1]，和 Ian Waudby-Smith[3]

[1]卡内基梅隆大学[2]法国国家信息与
自动化研究院和 ´巴黎高等师范学院 [3]
加利福尼亚大学伯克利分校

2026年3月

**摘要**

我们针对大样本量推导了推断程序，使其在显著性水平依赖于数据时仍然有效（即所谓的“事后有效推断”）。经典统计工具要求显著性水平——也就是“第一类错误率”——必须在查看或分析任何数据之前选定。这一限制会带来一些弊端。例如，如果分析者得到的置信区间无法得出明确结论，就不能通过提高显著性水平来重复这一过程——结果已成定局。近年来，e值应运而生，成为解决这一问题的工具；对于执行各种形式的事后推断而言，它们既是必要条件，也是充分条件。然而，迄今为止，这类结果都属于非渐近结果。因此，它们继承了非渐近推断程序的一些常见局限，例如要求较强的矩假设，且通常较为保守。本文在渐近情形下建立了事后推断理论，给出了渐近事后置信集和渐近事后p值；与非渐近对应物相比，这些方法所需的假设更弱，结果也更精确。

## 目录

**1 引言** **2**

1.1 主要贡献与文章结构.41.2 相关工作.51.3 符号与背景知识.6

**2 预备知识** **7**

2.1 事后置信区间与p值.72.2 渐近事后置信区间与p值.82.3 分布一致的渐近事后置信区间与p值.92.4 渐近 e 值的充分性与必要性.10

*贡献相同。联系邮箱：benchugg@cmu.edu

**3 构建　　渐近事后置信区间11**
3.1 iwr渐近e变量与aph-ci.123.2 选择 $\lambda$，选项I：事前锚定.143.3 选择 $\lambda$，选项II：混合方法.153.4 事件划分与 r-ws 渐近 e 变量.173.5 模拟.19

**4 事后　　序贯推断与渐近 e 过程21**
4.1 定义与期望性质.224.2 渐近事后混合置信序列.24

**5 结论** **24**

**A 关于事后置信区间的替代定义** **29**

**B 附加结果** **31**
B.1 两种替代渐近 e 变量.31B.2 基于 $E_n^{\mathrm{REG}}$.32B.3 APH-CI 的精确渐近第一类错误.33B.4 通过复合渐近 e 变量构造 APH-CI.35B.5 分布一致的期望收敛.37B.6 混合过程的广义充分条件.38B.7 截断高斯混合的更多细节.39

**C 省略的证明** **41**
C.1 命题 2.6 的证明.43 C.2 定理 3.1 的证明.44 C.3 定理 3.2 的证明.47 C.4 命题 3.3 的证明.47 C.5 定理3.6 的证明.48 C.6 定理 3.8 的证明.49 C.7 推论 3.9 的证明.50 C.8 命题 4.4 的证明.51 C.9 定理 B.1 的证明.52 C.10 定理 B.2 的证明.54 C.11 定理 B.3 的证明.57

**D 附录　　实验与仿真细节 58**
D.1 仿真细节.58D.2 影响和 $\kappa$ 对 $E^{\mathrm{MIX},\ n}$ $^{\mathrm{IWR}}$.59D.3 APH 置信区间：REG 与 IWR.60D.4 影响 $\eta$ 对 $\mathcal{H}_n^{\mathrm{REG}}$.61

# 1 引言

大多数应用统计方法都是渐近的，也就是说，只有当样本数量趋于无穷时，它们才具有有效性。渐近推断在实践中得到了广泛应用，因为它通常只要求较弱的矩假设。相比之下，非渐近推断必然依赖于对强矩假设的了解；有关缺少这些假设时非渐近推断如何失效的正式说明，请参见 Bahadur 和 Savage [2] 。

尽管适用范围广泛，实践中使用的渐近推断仍有以下缺点：显著性水平——通常记为 $\alpha \in (0,1)$——必须在进行任何数据分析之前设定。以置信区间（CI）为例。一旦计算出这样的区间，若使用不同的显著性参数，基于同一数据重新计算该区间，就会失去在显著性水平独立于数据时原本能够获得的统计保证。

在本文中，我们将说明渐近 e 值如何为这一问题提供解决方案——并且正如我们将要证明的，它是唯一的解决方案。我们提出了一个灵活的框架，用于构造渐近事后置信区间、p 值和假设检验，并允许 $\alpha$ 在看到数据后再进行选择。e 值在事后分析方面的潜力已在非渐近假设检验中得到研究 [26, 35, 34, 8] ，以及其他非渐近推断领域；例如，Gauthier 等人 [18, 19, 20]将其应用于保形预测。本文将这种潜力扩展到渐近情形。渐近方法适用范围广泛，这正是我们的研究动机；我们希望提供一种同样灵活且能够给出事后保证的替代方案。

为了更正式地引入问题，我们先聚焦于置信区间。假设数据$X_1$、.....、 $X_n$ 来自某个未知分布 $P$，且存在一个我们希望估计的实值参数$\theta \equiv \theta(P)$。给定显著性水平 $\alpha \in (0, 1)$，$\theta$ 的渐近置信区间是一个随机集合 $\mathcal{C}_n(\alpha)$，其覆盖 $\theta$ 的概率在极限下至多为 $\alpha$ 。也就是说：

$$\text{For all } \alpha \in (0,1), \quad \limsup_{n\to\infty} P(\theta \notin \mathcal{C}_n(\alpha)) \leq \alpha. \tag{1.1}$$

请注意， $\alpha$ 被假定为常数，因此必须在进行任何分析之前选定它。以下示例说明，在分析之前选择 $\alpha$ 如何导致数据被浪费在无法得出结论的结果上。

> 在流行病学和免疫学中，人们经常为某种疾病或疫苗的率比构建置信区间，即在适当的时间窗口内，不良病例数与某个比较组之比 [48, 53]。率比的估计值可能会影响采取不同的行动；例如，如果疫苗的率比过高，我们可能会停止推广。假设一位从业者给出的置信区间为 $\mathcal{C}_n(0。01)$，其设计目标是以99%的概率包含真实率比。如果区间过宽、难以解释，人们很容易想要计算 $\mathcal{C}_n(0。05)$， $\mathcal{C}_n(0。1)$，等等，以便提供某种形式的公众指导。不幸的是，目前大多数构建置信区间的方法都不允许作出这类由数据驱动的决策；以这种方式重新计算 $\mathcal{C}_n$将导致违反（1.1）中的保证。

经典置信区间无法处理事后显著性水平，这一点众所周知。根据数据选择 $\alpha$ 被称为“游移显著性水平问题” [22] ，并且已发现这一问题存在于多个学科领域，包括医学、心理学、商业和市场营销（参见29， 10， 30、 23）。

对于意识到游移显著性水平问题的分析人员，一种常见的解决方案是“$\alpha$支出” [14], ，它属于更广泛的一类方法，通常称为组序贯 方法。例如，新冠肺炎监测就是采用 $\alpha$ 支出进行的 [15]。其思路是设定一个全局 $\alpha$ ，并使用某个 $\alpha_1 < \alpha$进行初始分析。如有需要，再使用 $\alpha_2 < \alpha - \alpha_1$ 进行第二次分析，然后使用 $\alpha_3 < \alpha -$ 进行第三次分析（$\alpha_1 + \alpha_2$），依此类推。（这里

$\alpha_1$，$\alpha_2$，.....而且确实，$\alpha$必须预先确定。）然而，这种预算方式显然存在缺点。与使用原始 $\alpha$相比，每次分析的检验效能都显著降低，而且下游分析仍然只能局限于少数预先规划好的项目。请注意，$\alpha$支出可以视为对一组相互独立的统计分析进行Bonferroni校正的一种形式。

在这项工作中，我们研究并拓展了一种替代性的 $\alpha$ 支出方法。我们将介绍的方法允许进行任意数量的下游分析——无需对 $\alpha$ 进行划分。该方法与经典方法的主要区别在于，我们寻求的保证类型不同于（1.1）。我们不再界定错误概率，而是同时针对风险（待定义）的多个取值进行界定。$\alpha$ 这与近期事后假设检验 [26, 35, 34, 8] 和共形风险控制 [1]领域的趋势一致。

与错误概率相比，考虑风险控制既有优点，也有缺点。我们的目标并非主张孰优孰劣，而是为基于风险控制的事后渐近推断奠定理论基础，将其作为统计工具箱中的一种补充工具。

本文将研究逐点情形和分布一致情形下的渐近性，[37], 后者要求渐近结论对某一类分布一致有效。我们将在第2.3节更正式地介绍分布一致性，但我们先借助（1.1）中的保证，非正式地讨论这一概念。在分布逐点统计保证下，无论 $n$ 有多大，都可能存在某个分布 $P_n$，使得 $P_n$（$\theta \notin \mathcal{C}_n(\alpha)$）仍与 $\alpha$ 相距甚远（在某些情况下，甚至可以任意接近 1）。相比之下，分布一致保证要求渐近陈述对某一类 $P$ 中的所有 $\mathcal{P}$ 同时成立（从数学上说，在上极限内部对 $\mathcal{P}$ 取上确界；见定义2.5）。这种一致性排除了上述异常行为。

## 1.1 贡献与大纲

在第 APH-CI2.2 节中定义了渐近事后置信区间（APH-PVAL）和渐近事后p值（APH-PVAL）之后，我们的主要贡献如下：

1. 第 2.4 节给出了一种一般性构造，用于从渐近 e 变量推导出 APH-CI 和 APH-PVAL。事实上，命题 2.6 表明，任何 APH-CI（具有单调性和左连续性，这两种性质是置信区间通常要求的自然期望性质）都必须基于渐近 e 变量构造。我们还证明，APH-PVAL必须是渐近 e 变量的逆，从而将关于非渐近事后推断的基本结果 [34,定理 2]推广到了渐近情形。

2. 第3.1 节首先分析所谓的 IWR 由 Ignatiadis 等人提出的渐近 e 变量 [31]。我们放宽了它在逐点情形下成立的条件，并证明：在分布一致设定中，在一致三阶矩假设下，它仍然是渐近 e 变量；见定理3.1。我们采用两种不同的方法来选择 IWRe变量中的自由参数——事前锚定（第3.2节）和混合方法（第3.3节）——从而给出两个 APH-CI，基于 IWR。它们被陈述为定理3.2 和3.6。

3. 第3.4 节则提供了另一种方法，用以构造渐近e变量和 APH-CI。我们引入所谓的 R-WS 渐近 e 变量，其优点是在分布一致设定下、在 $2+\delta$矩假设下对任意 $\delta > 0$均成立。尽管所得 APH-CI 比基于IWR e变量的结果略为宽松，但我们表明，它实际上提供了更强的保证。具体而言，它是事后渐近置信序列（APH-CS）的一个实例，而 R-WSe变量是渐近*e* 过程。这些对象在第4节中被正式引入，可视为将渐近推断同时推广到事后设定与时间一致设定，从而把数理统计中两条近期的研究脉络结合起来。

第3.5 节包含对 APH-CI的实验比较，附录B 提供了额外的理论结果。特别地，附录B.1 和 B.2 包含额外的渐近e变量以及 APH-CI，这些并非本文此处论述的核心，但仍可能具有参考价值。

尽管本文的贡献主要是理论性的，我们的目标是改进执业统计学家的工作流程。我们希望为一种实用的推断模式奠定基础，使其比传统统计技术具有显著更高的灵活性。基于这一考虑，本文最后一节（第5节）给出了我们的建议，说明所引入的事后置信区间中哪些最适合在实践中使用。本文提出的所有方法的Python实现可在 https://github.com/bchugg/asymp-posthoc-confint获取。

## 1.2 相关工作

e值通过分析原假设下的期望而非尾概率而得到，为在重复分析中获得有效推断提供了另一种途径；参见Ramdas 和 Wang [41] 以及Ramdas等人 [42] 的全面介绍，以及 [49, 24, 60] 中阐述其与投注之联系的讨论论文。我们对e值的特别兴趣在于其事后有效性；正如我们将表明的那样，这允许使用数据依赖误差阈值。

e值与依赖数据的阈值之间的联系，是 Katsevich 和 Ramdas [32] 在错误发现率（FDR）控制的背景下建立的。这些作者研究了非负上鞅；值得注意的是，非负上鞅在停时处会产生 e值，这一事实可由 Doob 的可选停时定理推出。Wang 和 Ramdas [58, Lemma 1] 给出了 e值与依赖数据的阈值下 FDR 控制之间的正式联系。将假设数量设为一，就把多重检验简化为对单个假设进行检验，并将 FDR 化为标准的第一类错误。因此，他们的引理 1 首次证明了 e值阈值化所实现的事后误差控制。

[65]提出了置信区间领域中与错误发现率控制类似的结果。这项工作重点是使用*e*-置信区间控制错误覆盖率（FCR）。[1] 类似地，Gr¨unwald [26] 研究了事后假设检验，旨在推广 Neyman-Pearson范式，使其能够容许事后显著性

[1]请注意，如果所讨论的区间只有一个，他们关于 FCR 事后有效性的结果也适用于标准错误覆盖率；此外，将 FDR 特化为单个假设时也有类似结论，此时得到的是经典的第一类错误控制。

水平。这项工作特别证明了 e-置信区间具有事后有效性（我们将在下文正式说明）。

此后，许多作者指出，e值能够推动多种数据自适应决策。例如，Xu 等人 [66] 以及 Fischer 和 Sechidis [17]在多重检验中运用 e值进行选择后推断和错误发现率控制，从而使名义错误发现率水平能够依赖数据。尽管以往的研究强调 e值能够产生“事后 p 值”，Koning [35, 34]证明了反过来也成立：每个“事后 p 值”都是某个 e值的倒数。Gr¨unwald [25] 提出了“e 后验”，它利用 e值提供贝叶斯后验的频率学派替代方案，同时也允许采用事后损失函数。

Gauthier 等人 [19]在保形推断的背景下研究事后置信区间，提出自适应覆盖策略，通过策略性地改变显著性水平，获得在大小约束下实现优化的保形集合 [20, 18]。Chugg 等人 [8] 以 Gr¨unwald [26] 的研究为基础，探讨事后假设检验中的可容许性。

关于 e值事后检验效能的一般概述，请参见 Ramdas 和 Wang [41,第 4] 章。

### 1.3 符号与背景

固定一个可测空间 $(\Omega\ \mathcal{F})$。在全文中，我们用 $\mathcal{P}$ 表示实值随机变量上的一组分布，使得对于每个 $P \in \mathcal{P}$，$(\Omega\ \mathcal{F},\ P)$ 都是一个概率空间。对于 $(\Omega\ \mathcal{F})$ 上的随机变量 $X_1$ 以及某个实数 $q \geq 1$，我们用 $\mathcal{M}_q(\theta)$ 表示 $\mathbb{R}$ 上均值为 $\mathbb{E}_P[X_1] = \theta$ 且具有有限 $q$ 阶矩（即 $\mathbb{E}_P|X_1 - \theta|^q < \infty$）的分布集合。令 $\mathcal{M}_q = \cup_{\theta\in\mathbb{R}}\mathcal{M}_q(\theta)$ 表示所有具有有限 $q$ 阶矩的分布构成的集合。

回顾一下，若随机变量序列 $(X_n)_{n\geq 1}$ 定义在 $(\Omega\ \mathcal{F})$ 上，则称其关于 $\mathcal{P}$ 一致可积，当且仅当

$$\sup_{P\in\mathcal{P}} \lim_{K\to\infty} \sup_{n\in\mathbb{N}} \mathbb{E}_P[X_n \mathbf{1}\{X_n \geq K\}] = 0. \tag{1.2}$$

注意，这是一个逐点（关于 $\mathcal{P}$）的一致可积性概念。我们还将用到一个分布一致版本。具体而言，若 $(X_n)_{n\geq 1}$ $\mathcal{P}$ 分布一致可积，则

$$\lim_{K\to\infty} \sup_{n\in\mathbb{N}} \sup_{P\in\mathcal{P}} \mathbb{E}_P[X_n \mathbf{1}\{X_n \geq K\}] = 0. \tag{1.3}$$

在分布一致设定中，一致矩控制的概念也很重要。若随机变量 $X$ 的 $\mathcal{P}$ 阶矩一致有界，则称其具有 $q$，如果

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P|X|^q < \infty. \tag{1.4}$$

对于每个 $n \in \mathbb{N}$，令 $\mathcal{F}_n := \sigma(X_1, \ldots\ldots, X_n)$ 为由前 $n$ 个观测值生成的 $\sigma$ 域。我们通常将序列 $(A_n)_{n\geq 1}$ 简写为 $(A_n)$，并将前 $n$ 个元素 $(A_1, \ldots\ldots, A_n)$ 简写为 $A^n$。序列 $(A_n)$ 的指标集应始终视为自然数 $\mathbb{N}$。

观察到 $X^n \sim P$ 时，对于某个未知的 $P \in \mathcal{P}$，我们感兴趣的是估计某个实值泛函 $\theta$: $\mathcal{P} \to \mathbb{R}$（例如均值 $\theta(P) = \int X\mathrm{d}P$）的 $\theta = \theta(P)$。定义参数空间 $\Theta = \{\theta(P) :\ P \in \mathcal{P}\}$。

对于 $\mathcal{P}$，e变量是指满足 $E \sup_{P\in\mathcal{P}} \mathbb{E}_P[E] \le 1$ 的非负随机变量。e值是e变量的一次实现，尽管在本文中我们会交替使用这两个术语（正如通常交替使用p变量和p值一样）。我们的多数结果将依赖于渐近e值：在极限下为e值的非负随机变量序列。我们将在第 2.2 节更正式地介绍这些概念。

# 2 预备知识

## 2.1 事后置信区间和 p 值

在介绍渐近事后推断之前，我们先描述非渐近事后推断程序，以帮助阐明后文将要给出的渐近对应定义。

给定一组分布 $\mathcal{P}$ 和期望的第一类错误率 $\alpha \in (0, 1)$，若区间 $\mathcal{C}(\alpha)$ 满足下式，则称其为泛函 $\theta$ 的 $(1-\alpha)$-置信区间

$$\sup_{P\in\mathcal{P}} P(\theta(P) \notin \mathcal{C}(\alpha)) \le \alpha, \tag{2.1}$$

而 p 值 $Q$ 满足

$$\text{for all } \alpha \in (0,1), \quad \sup_{P\in\mathcal{P}} P(Q \le \alpha) \le \alpha. \tag{2.2}$$

转向事后置信区间和 p 值的定义，请注意，(2.1) 和 (2.2)中的条件等价于写成

$$\sup_{\alpha\in(0,1)} \sup_{P\in\mathcal{P}} \mathbb{E}_P\left[\frac{\mathbf{1}\{\theta(P) \notin \mathcal{C}_n(\alpha)\}}{\alpha}\right] \le 1 \quad \text{and} \quad \sup_{\alpha\in(0,1)} \sup_{P\in\mathcal{P}} \mathbb{E}_P\left[\frac{\mathbf{1}\{Q \le \alpha\}}{\alpha}\right] \le 1, \tag{2.3}$$

分别。重要的是，在（2.3）中，$\alpha$ 在取期望之前是固定的——它与数据无关。结合（2.3），可以直观地理解基于数据且依赖于 $\alpha$ 值的事后置信区间和p值的定义。这些定义已见于文献；参见 Xu 等人 [65, 推论 1], Gr¨unwald [26, 第 3]节，以及 Koning [35,定义 2]。

**定义 2.1**（事后置信区间和p值）。我们称一族随机集 $\mathcal{H} \equiv \{\mathcal{H}(\alpha)\}_{\alpha>0}$ 为 $\theta$ 和 $\mathcal{P}$ 的事后置信区间，如果

$$\sup_{P\in\mathcal{P}} \mathbf{R}(\mathcal{H};\theta,P) \le 1 \quad \text{where} \quad \mathbf{R}(\mathcal{H};\theta,P) := \mathbb{E}_P\left[\sup_{\alpha>0} \frac{\mathbf{1}\{\theta(P) \notin \mathcal{H}(\alpha)\}}{\alpha}\right]. \tag{2.4}$$

我们称非负随机变量 $Q$ 为 $\mathcal{P}$ 的事后*p*值，如果

$$\sup_{P\in\mathcal{P}} \mathbf{S}(Q;P) \le 1 \quad \text{where} \quad \mathbf{S}(Q;P) := \mathbb{E}_P\left[\sup_{\alpha>0} \frac{\mathbf{1}\{Q \le \alpha\}}{\alpha}\right] \le 1. \tag{2.5}$$

在上述定义及其前文中，我们使用符号“$\mathcal{H}$”和“$\mathcal{C}$”分别指代事后置信集和非事后置信集，并将在全文中沿用这一约定。我们将$\mathbf{R}$ 和$\mathbf{S}$ 都称为风险，并将

保证（2.4）和（2.5）称为风险控制。除非存在歧义，我们通常会在定义中省略“对于 $\theta$ 和 $\mathcal{P}$”这一短语。如果 $\mathcal{H}$ 是一个事后置信集，我们也称其为事后有效；事后有效 p 值也同样如此。如果根据上下文可以确定$\mathbf{R}(\mathcal{H}$，我们通常用它代替$\mathbf{R}(\mathcal{H};\ \theta,\ P)$，前提是 $\theta$和 $P$均可从上下文中理解。

**注 2.2**（关于 $\alpha$ 的取值范围可超过 1）。从（2.3）转到（2.4）时，我们不仅将 $\alpha$ 移入了期望之中，还允许它取大于 1 的值。这仅仅是为了后续推导中的数学便利，并不会实质性地改变保证。事实上，对于任意 $\alpha > 1$，$\mathbf{1}\{\theta(P) \notin \mathcal{H}(\alpha)\}/\alpha \leq 1$，因此这样的 $\alpha$绝不会破坏风险控制。

值得思考的是控制风险与控制错误概率之间的区别，尤其要注意，（2.4）和（2.5）中的不等式所提供的保证，与（2.1）和（2.2）中的保证有本质区别。以事后置信区间为例。如果我们基于 $k$ 次独立实验设计事后置信区间 $k$，就不应期望 $\theta$ 至少在 $1-\alpha\%$ 的实验中包含于我们的集合中，$k \to \infty$。相反，我们应当期望，对于任意 $\alpha$，所有试验中 $\mathbf{1}\{\theta(P) \notin \mathcal{H}_n(\alpha)\}/\alpha$ 的平均值至多为 1（因为 $k \to \infty$）。

**注 2.3**（从不依赖数据的 $\alpha$ 值的事后风险到第一类错误）。如果 $\alpha$ 不依赖数据，则（2.4）和（2.5）中的事后保证退化为（2.1）和（2.2）中通常的第一类错误保证。本文所研究的所有事后保证均属此种情形。

鉴于风险控制总体上不同于第一类错误控制，自然会问前者能为从业者提供何种效用。毕竟，从业者熟悉错误概率——我们能否期望他们转而使用风险？我们认为，在需要事后决策 且 频率学派式错误保证的情形中，这种转变是自然的。虽然贝叶斯决策理论允许事后决策（这是众所周知的，但可参见 Chugg 等人 [8, 附录 C] 中的正式陈述与证明），但它显然并不提供长期错误的界，而定义2.1正是为此而设计的。这一点，再加上事后风险控制理论实际上相当丰富，使我们相信，阐明（渐近）事后 p 值与事后置信区间的利弊是有价值的。

## 2.2 渐近事后置信区间和 p 值

有了事后 p 值和事后置信区间的定义后，现在来介绍它们的渐近形式。使用第1.3节中引入的记号，引言中关于渐近置信区间的保证可以更精确地写为

$$\text{For all } \alpha \in (0,1), \quad \sup_{P \in \mathcal{P}} \limsup_{n \to \infty} P(\theta(P) \notin \mathcal{C}_n(\alpha)) \leq \alpha. \tag{2.6}$$

沿用上一节的类似论证，我们将 (2.6) 改写为关于期望值的界，由此引出渐近事后置信区间和 p 值的以下定义。

**定义 2.4**（渐近事后置信区间和 p 值）。 如果序列 $\mathcal{H} \equiv (\mathcal{H}_n)_{n\geq 1}$ 满足以下条件，我们就称其为 $\theta$ 在 $\mathcal{P}$下的渐近事后置信区间 （APH-CI）：

$$\sup_{P\in\mathcal{P}} \limsup_{n\to\infty} \mathbf{R}(\mathcal{H}_n; \theta, P) \leq 1. \tag{2.7}$$

我们称随机变量序列 $(Q_n)_{n\geq 1}$ 为 $\mathcal{P}$ 定义了一个渐近事后$p$值(APH-PVAL)，如果

$$\sup_{P\in\mathcal{P}} \limsup_{n\to\infty} \mathbf{S}(Q_n; P) \leq 1. \tag{2.8}$$

换言之，正如渐近置信区间要求第一类错误在极限中有界， APH-CI也要求事后风险在极限中有界。

请注意，在 (2.7) 和 (2.8) 中，对 $\mathcal{P}$ 取上确界都位于极限之外。这意味着，对每个分布，极限上确界都逐点有上界。我们或许会希望极限上确界具有一致有界性，这相当于将对 $P \in \mathcal{P}$ 取上确界移到极限之内。接下来我们将介绍这类“分布一致”的 APH-CI及 APH-PVAL。

## 2.3 分布一致的渐近事后置信区间和p值

假设我们已经构造了一个 APH-CI $\mathcal{H}$，用于估计高斯分布 $P$ 的均值 $\mu(P)$。令 $P_\sigma$ 表示一个均值为零、方差为 $\sigma^2 > 0$ 的高斯随机变量的分布，并考虑分布族 $\mathcal{P} = \{P_\sigma : \sigma \in (0, \infty)\}$。举例来说，假设 **实数集**$(\mathcal{H}_n, P_\sigma) \asymp \sigma/n+1$。那么，对每个固定的 $\sigma$，$\limsup_n$ **实数集**$(\mathcal{H}_n, P_\sigma) = 1$，所以 $\mathcal{H}$ 确实是一个 APH-CI。然而，对任意 $n$，$\sup_{\sigma>0}$ **实数集**$(\mathcal{H}_n, P_\sigma) = \infty$，因此

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \mathbf{R}(\mathcal{H}_n, P) = \infty.$$

上述可能性促使我们将 APH-CIS和 APH-PVALS的定义扩展到分布一致设定，在该设定下，对于$\mathcal{P}$中的所有分布，极限同时成立。至少从 Chung [9], 开始，人们就一直在研究分布一致极限定理；但具体而言，分布一致渐近置信区间的研究始于 Li [37], ，他将其称为“诚实”置信区间。此后，分布一致方法已在各种各样的设定中得到研究，并且在过去十年中势头显著增强 [54, 44, 50, 61, 63, 64]。

从数学上讲，将 APH-CIS 的定义扩展到分布一致设定非常直接；我们只需在定义 2.4 中交换极限与上确界的位置。回顾第 **2.1** 节中 **R** 和 S 的定义。我们注意到，Ignatiadis 等人最近研究了分布一致 p 值（尽管不是事后 p 值） [31]。

**定义 2.5**（分布一致的 APH-CI与 APH-PVAL）。 固定一个分布族$\mathcal{P}$。对每个 $P \in \mathcal{P}$，令 $\theta(P) \in \mathbb{R}$ 为一个参数，可将其视为 $P$下 APH-CI 的目标估计量。我们称序列 $\mathcal{H} \equiv (\mathcal{H}_n)_{n\geq 1}$ 是 $\mathcal{P}$ 下对于 $\theta \equiv (\theta(P))_{P\in\mathcal{P}}$的分布一致 APH-CI（或 $\mathcal{P}$一致 APH-CI 对于 $\theta$），如果

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \mathbf{R}(\mathcal{H}_n; \theta(P), P) \leq 1. \tag{2.9}$$

我们称随机变量序列$(Q_n)_{n\geq 1}$是分布一致 APH-PVAL 对于$\mathcal{P}$（或等价地， $\mathcal{P}$一致 APH-PVAL），如果

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \mathbf{S}(Q_n; P) \leq 1. \tag{2.10}$$

我们通常会把“分布一致”简称为“一致”。分布 $\mathcal{P}$ 通常可从上下文中理解。如果我们没有明确说明某个 APH-CI 或 APH-PVAL是一致的，则应假定其仅逐点成立。在后文讨论 $\mathcal{P}$一致的 APH-CI时，我们可能直接将 $\theta$ 称为“估计目标”，但这应被理解为由 $\mathcal{P}$索引的估计目标$(\theta_P)_{P\in\mathcal{P}}$所构成的隐式集合。

接下来，我们研究如何利用一致渐近 e 变量构造一致的 APH-CI。

## 2.4 渐近 e 值的充分性与必要性

现在让我们回到如何构造 APH-CI的问题。正如事后置信区间可以用e值来设计——如Xu等人[65]与Gr¨unwald [26]所表明的那样——我们可以利用渐近 $e$ 值来设计渐近事后置信区间。遵循Ignatiadis等人中的定义 3.5与命题 3.6 [31], ，一个针对 $\mathcal{P}$ 的渐近e值是非负随机变量序列$(E_n)_{n\geq 1}$，其分量满足

$$\sup_{P\in\mathcal{P}} \limsup_{n\to\infty} \mathbb{E}_P[E_n] \leq 1. \tag{2.11}$$

（作者将这些称为强渐近 e 变量；为简洁起见，我们省略“强”一词。）如果不等式取等号，则称$(E_n)$ 为尖锐的渐近 e 变量。

渐近 e 变量的这一定义相对于 $\mathcal{P}$ 而言同样是一个逐点概念。为了构造分布一致的 APH-CIS 和 APH-PVALS，我们需要依赖分布一致的渐近 e 变量，即由非负随机变量序列 $(E_n)$ 构成，并满足

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \mathbb{E}_P[E_n] \leq 1. \tag{2.12}$$

我们通常将这样的序列称为 $\mathcal{P}$-一致渐近 e 变量，以明确分布集合 $\mathcal{P}$ 。这些对象也是由 Ignatiadis 等人 [31], 引入的，他们将其称为强一致渐近 e 变量。同样地，我们省略“强”字。

如果 (2.11) 或 (2.12) 取等号，则称这些（一致）渐近 e 变量为尖锐的。如果 $\mathcal{P}_\theta$ 是某个具有参数 $\theta$ 的类中的所有分布的集合，我们将称 $(E_n)$ 是针对 $\theta$ 的渐近 e 变量，当且仅当它是针对 $\mathcal{P}_\theta$ 的渐近 e 变量。类似地，如果它是一个 $\theta$ -一致渐近 e 变量，我们称其为针对 $_\theta$的一致渐近 e 变量。

下面的结果表明，我们可以通过对渐近 e 值的各个分量进行阈值化，得到一个 APH-CI 和一个 APH-PVAL 。事实上，假设这些集合具有标准行为，这是构造（渐近）事后置信集的唯一方法。证明见附录C.1 ，其灵感来自非渐近 e 值和事后推断的类似结果；参见 Gr¨unwald [26, 第 3] 节以及 Koning [35, 定理 2]。

**命题 2.6**。设 $\Theta$ 是一个可视为“参数空间”的集合，并设$(\mathcal{H}_n(\alpha))_{\alpha>0}$ 是 $\Theta$ 的一个子集族，该子集族关于 $\alpha$ 具有单调性和右连续性，即：

- $\mathcal{H}(\alpha_2) \subseteq \mathcal{H}(\alpha_1)$ *for all* $0 < \alpha_1 < \alpha_2$ *(monotonicity),*

- 对于每个 $\theta \in \Theta$，拒绝集 $\{\alpha > 0 : \ \theta \notin \mathcal{H}_n(\alpha)\}$ 是形如 $[\alpha_0, \infty)$ 的左闭区间，其中 $\alpha_0 > 0$ （右连续性）。

那么，序列$(\mathcal{H}_n)_{n\geq 1}$ 是关于 $\theta^\star$ 的（统一） APH-CI，当且仅当存在一个随机变量序列$(E_n(\theta))$，对每个 $\theta \in \Theta$，使得对所有 $\alpha > 0$， $H_n(\alpha)$ 具有如下形式

$$\mathcal{H}_n(\alpha) = \{\theta : E_n(\theta) < 1/\alpha\},$$

并且$(E_n(\theta^\star))$ 是一个（统一）渐近 *e* 变量。此外，随机变量序列$(Q_n)_{n\geq 1}$ 是 APH-PVAL，当且仅当存在某个（统一）渐近 *e* 变量$(E_n)_{n\geq 1}$，使得对每个 $n \in \mathbb{N}$， $Q_n = 1/E_n$ 。

单调性和右连续性是置信区间自然（且通常被假定具有）的期望性质。前者意味着 $\alpha$ 的值越小，区间也越小；后者则意味着区间具有嵌套性。也就是说，如果 $\theta$ 在一列趋于 $\alpha$ 且从上方逼近的水平下均被拒绝，那么在 $\alpha$处也会被拒绝。

命题2.6 中的“当且仅当”仅当 $\alpha$ 在定义 2.1中被允许取大于 1 的值时才成立。事实上，若 $\alpha$ 被限制在 (0,1) 内，则 $\{\theta : \ E_n(\theta) \ < 1/\alpha\}$ 是一个APH-CI，对于 $\theta^\star$，如果存在某个过程族 $\{(E_n(\theta)) : \theta \in \Theta\}$ 使得 $(E_n(\theta^\star))$ 是 $\theta^\star$的渐近 e 变量，但反之未必成立。

**备注 2.7.** 命题 2.6 隐含假定渐近 e 值不是 $\alpha$的函数。具有这种依赖性的 e值并不罕见。事实上，“可预测代入法”是构造序贯 e 值（即 e 过程）的常用方法，并且通常依赖于将参数选择为 $\alpha$ [60]的函数。对 $(0, \alpha)$ 中的参数积分而得到的 e值也已被提出 [16]。尽管 e值本身可能不依赖于 $\alpha$，由该 e值得到的置信区间 $\mathcal{H}_n(\alpha)$ 仍会依赖于它。

在实践中，仅使用单个 e变量构造 APH-CI 往往会得到单侧区间（除非使用混合方法；参见第3.3节）。为了获得双侧区间，我们将取由多个 e值所定义集合的交集。不过，我们可以放宽要求，不必通过各自有效的（渐近）e变量来构造这些集合，而是允许它们由（渐近）复合 *e* 变量 [31]构造。不过，这一结果并非本文的核心内容，因此留待附录B.4讨论。

证明随机变量序列是逐点渐近 e 变量，通常需要结合依分布收敛与一致可积性进行论证。根据 Billingsley 的一项结果， [5], 这意味着均值收敛（而非均值收敛）。在分布一致设定下，我们将阐述并运用附录B.5中的一个命题，可将其视为 Billingsley 结果的推广。此处提及该命题，是因为它或许也具有独立的参考价值。

## 3 构造渐近事后置信区间

根据命题2.6，构造 APH-CIS 依赖于找到渐近 e 变量。这类变量的研究不如其非渐近对应物充分。事实上，非平凡的渐近 e 值是否应该存在，并不显而易见。Ignatiadis 等人 [31, 定理 6。5] 消除了这一疑虑，并证明，对于独立同分布的 $X_1, \ldots, X_n$，其均值为 $\theta$ 且具有有限方差，对任意 $\lambda \in \mathbb{R}$

$$E_n^{\mathrm{IWR}}(\theta; \lambda) := \exp\left(\lambda \frac{S_n(\theta)}{V_n(\theta)} - \frac{\lambda^2}{2}\right) \tag{3.1}$$

是一个渐近 $e$ 值，其中

$$S_n(\theta) := \sum_{i\le n}(X_i - \theta) \quad \text{and} \quad V_n(\theta) := \left(\sum_{i\le n}(X_i-\theta)^2\right)^{1/2}. \tag{3.2}$$

上标 IWR指的是 Ignatiadis、Wang 和 Ramdas 的名字 [31]。在第3.1节中，我们将从几个方面扩展这一渐近 e 变量。首先，我们注意到，它不仅适用于所有有限方差分布，也适用于高斯吸引域中的分布（这是一个较弱的条件）。其次，我们将证明，在三阶矩一致有界的假设下，$(E_n^{\text{IWR}})$ 是一致渐近 e 变量。定理3.2随后给出了由 ( APH-CI ) 定义的$E_n^{\text{IWR}}(\theta;\ \lambda)$。

第3.2 节和第3.3 节探讨了选择 $\lambda$ 以优化 APH-CI宽度的不同方法。这比非渐近情形更复杂，一方面是因为 $\lambda$ 不能依赖$\alpha$ (注 释 2.7)，另一方面是因为需要满足一致可积性。尽管这两节尤其关注 $E_n^{\text{IWR}}$，其中讨论的原则也适用于更广泛的情形，例如附录 APH-CIB.1中介绍的另外两个。

第 3.4 节采用与 IWR e 变量完全不同的论证，给出了另一个渐近 e 变量。它基于一种截断技术，并结合了 Ruf 和 Waudby-Smith 最近提出的非渐近强大数定律 [47]。由此得到的 APH-CI 与 IWR APH-CI相比各有优缺点；我们将在接下来的讨论中更深入地探讨这些权衡。

### 3.1 IWR 渐近 e 变量与 APH-CI

在本节中，令 $\widehat{\sigma}_n^2 := (n-1)^{-1}\sum_{i\le n}(X_i - X_n)^2$ 表示无偏样本方差，并令 $\hat{s}_n^2 = n^{-1}\sum_{i\le n}(X_i - X_n)^2$ 表示有偏样本方差，其中 $X_n = n^{-1}\sum_{i\le n} X_i$ 为样本均值。 —

为陈述关于 $E_n^{\text{IWR}}$的结果，我们需要引入若干分布假设。回忆一下，对于 $q \ge 1$，集合 $\mathcal{M}_q(\theta)$表示在 $\mathbb{R}$ 上具有有限 $q$ 阶矩且均值为 $\theta$的那些分布，而 $\mathcal{M}_q$ 是所有具有有限 $q$ 阶矩的分布构成的集合。

我们称 $(X_n)$ 属于高斯吸引域，如果存在序列 $(a_n)$和 $(b_n)$，使得

$$\frac{\sum_{i\le n} X_i - b_n}{a_n} \xrightarrow{d} N(0,1). \tag{3.3}$$

如果分布具有有限方差 $\sigma^2$ ，那么3.3式成立，其中 $b_n = n\mathbb{E}[X_1]$ 且 $a_n = \sigma\sqrt{n}$。但对于不具有有限二阶矩的分布，3.3式也可能成立，例如形状参数 2 的帕累托分布。更多关于吸引域的细节，参见 Mikosch [39]。我们令 $\mathrm{DAN}(\theta)$ 表示所有均值为 $\theta$ 且属于某个高斯吸引域的分布。

我们的分布一致结果要求该 IWRe变量具有一致有界矩。具体而言，我们将使用一个 $\mathcal{P}$-有界的“偏度”，其含义是

$$\sup_{P\in\mathcal{P}} \frac{\mathbb{E}_P|X_1 - \mathbb{E}_P X_1|^3}{\sigma_P^3} < \infty, \tag{3.4}$$

其中， $\sigma_P^2$ 是 $X_1$ 在 $P$下的方差。要求矩一致有界（通常阶数高于二阶）是分布一致推断中的一项标准假设 [50, 61] ，并且

已知这是获得一致强收敛结果的必要条件 [63]。在我们的情形中，由于使用了 Berry-Esseen 界，我们将要求三阶矩/偏度一致有界 [3]。具体来说，证明依赖于确定性地给出 $S_n(\theta)/V_n(\theta)$的上界 $\sqrt{n}$ （使用柯西-施瓦茨不等式）。而这一上界会被 Berry-Esseen 界的 $1/\sqrt{n}$衰减速率抵消。这使我们能够得出结论：$\sup_P \mathbb{E}_P|S_n(\theta)/V_n(\theta)| < \infty$，这是证明的重要组成部分。

有了这些预备知识，我们可以陈述关于 IWR e变量的结果。在逐点情形下，借助 Gin´e 等人 [21], 的结果，证明相当直接；这些结果表明 $S_n(\theta)/V_n(\theta) \xrightarrow{d} N(0,\ 1)$，并且 $\exp(\lambda_n S_n(\theta)/V_n(\theta))$ 是一致可积的。分布一致情形则要复杂得多。证明的详细内容见附录C.2。

定理 *3.1.* 设 $(X_n)$ 是一个独立同分布随机变量序列。设 $(\lambda_n)$ 是一个收敛到某个 $\lambda \in \mathbb{R}$ 的非负随机变量序列，且几乎必然成立。若 $\sup_n |\lambda_n| < \infty$ 几乎必然成立，且 $(X_n)$ 取自 $\mathrm{DAN}(\theta)$ 中的某个分布，则过程 $(E_n^{\mathrm{IWR}}(\theta, \lambda_n))$ 是一个尖锐渐近 *e* 变量。此外，若 $(\lambda_n)$ 是一个收敛到 $\lambda$ 且满足 $X_1, \ldots, X_n$ 对某个 $\mathcal{P} \subseteq \mathrm{DAN}(\theta)$ 具有 $\mathcal{P}$- 一致有界偏度的确定性序列，则 $(E_n^{\mathrm{IWR}}(\theta; \lambda_n))$ 是一个 $\mathcal{P}$-一致渐近 *e* 变量。

我们对定理 3.1 的表述尽可能一般——渐近 e 变量是文献中相对较新的对象，因此新结果本身可能就很有趣。不过，接下来我们不再需要所有这些自由度。特别是，APH-CI 基于 $E_n^{\mathrm{IWR}}$不会选择 $\lambda$ 作为 $n$的函数。

现在我们给出 APH-CI 通过对 IWR 所蕴含的渐近 e 变量求逆而得到的，见命题 2.6。证明见附录 C.3。

**定理 3.2**。 设$(X_n)$ 是从 DAN中的某个分布抽取的独立同分布随机变量序列，满足$(\mu)$。对于任意 $\lambda > 0$ 和 $\alpha \in (0,1)$，令

$$A_{n,\alpha}(\lambda) = \frac{\log(2/\alpha) + \lambda^2/2}{\lambda\sqrt{n}}.$$

那么，对于任意与 $\alpha$独立的 $\lambda > 0$，$(\mathcal{H}_n^{\mathrm{IWR}}(\alpha;\ \lambda))$ 是关于 $\mu$的一个 APH-CI，其中

$$\mathcal{H}_n^{\mathrm{IWR}}(\alpha;\lambda) := (\overline{X}_n \pm W_n^{\mathrm{IWR}}(\alpha;\lambda)), \quad W_n^{\mathrm{IWR}}(\alpha;\lambda) := \frac{A_{n,\alpha}(\lambda)\hat{s}_n}{(1 - A_{n,\alpha}(\lambda)^2)^{1/2}}, \tag{3.5}$$

若 $n > (\log(2/\alpha)/\lambda + \lambda/2)^2$,，否则 $\mathcal{H}_n^{\mathrm{IWR}}(\lambda) = \mathbb{R}$。此外，若$(X_n)$ 具有某个 $\mathcal{P}$ 的 $\mathcal{P}$-一致有界偏度，则$(\mathcal{H}_n^{\mathrm{IWR}}(\alpha;\ \lambda))$ 是一个 $\mathcal{P}$-一致的 APH-CI。

我们该如何选择 $\lambda$ 中的参数 $\mathcal{H}_n^{\mathrm{IWR}}$? 对于任何固定的$\lambda$，宽度都会趋于零，因为 $A_{\alpha,n}(\lambda) \xrightarrow{n\to\infty} 0$。当 $n$很大时，$W_n^{\mathrm{IWR}}(\lambda) \approx A_{n,\alpha}(\lambda)\hat{s}_n$，因此 $\lambda$ 的理想选择是$\lambda_\alpha = \sqrt{2\log(2/\alpha)}$，这使 $\lambda \mapsto A_{n,\alpha}(\lambda)$ 最小。遗憾的是，正如定理陈述和备注 2.7 中都强调的，这种依赖于 $\alpha$的 $\lambda$选择是不合法的。

章节3.2 和3.3采用两种截然不同的策略来选择 $\lambda$。不过，首先，让我们通过研究 $\mathcal{H}_n^{\mathrm{IWR}}$ 的极限宽度如何随 $\sqrt{n}$变化，来更好地理解固定 $\lambda$ 的作用。假设（$X_n$）是具有有限方差的独立同分布$\sigma^2$。那么，将一次强大数定律的应用与连续映射定理相结合即可得到：

$$\sqrt{n}W_n^{\mathrm{IWR}}(\alpha;\lambda) \xrightarrow[n\to\infty]{a.s.} \sigma g(\lambda,\alpha) \quad \text{where} \quad g(\lambda,\alpha) := \frac{\log(2/\alpha)}{\lambda} + \frac{\lambda}{2}. \tag{3.6}$$

将该宽度与基于中心极限定理、针对有限（未知）方差的通常 Wald置信区间进行比较是很有益的（我们强调，这种区间不适用于事后分析）：

$$\mathcal{H}_n^{\mathrm{WALD}}(\alpha) := (\overline{X} \pm W_n^{\mathrm{WALD}}(\alpha)), \quad \text{where} \quad W_n^{\mathrm{WALD}}(\alpha) := z_{1-\alpha/2}\frac{\widehat{\sigma}_n}{\sqrt{n}}. \tag{3.7}$$

根据强大数定律，$\sqrt{n}W_n^{\mathrm{WALD}}(\alpha) \to z_{1-\alpha/2}\sigma$，其中 $z_{1-\alpha/2} = \Phi^{-1}(1-\alpha/2)$ 是标准正态分位数。当 $\alpha$较小时，$\Phi^{-1}(1-\alpha/2)$ 的表现类似于 $\sqrt{2\log(2/\alpha)}$（事实上，$\lim_{\alpha\to 0^+}\Phi^{-1}(1-\alpha/2)/\sqrt{2\log(2/\alpha)} = 1$）。因此，当 $n$ 较大且 $\alpha$较小时，$\sqrt{n}W_n^{\mathrm{WALD}}(\alpha) \approx \sqrt{2\sigma^2\log(2/\alpha)}$。

如上所述，$\lambda$ 的理想选择是 $\lambda_\alpha := \sqrt{2\log(2/\alpha)}$，使得 $g(\lambda_\alpha, \alpha) = \log(2/\alpha)/\lambda_\alpha + \lambda_\alpha/2 = \sqrt{2\log(2/\alpha)}$。在这种情况下，$\sqrt{n}W_n^{\mathrm{IWR}}(\alpha; \lambda_\alpha) \to \sqrt{2\sigma^2\log(2/\alpha)}$几乎必然成立，与 Wald 置信区间相同；当 $\alpha$ 较小时，也近似等同于 Wald 置信区间。

如果将（$\mathcal{H}_n^{\mathrm{IWR}}$）仅作为常规的渐近置信区间使用，并且不改变 $\alpha$，那么当然可以自由选择 $\lambda = \lambda_\alpha$。然而，在事后设定中（即 $\alpha$ 可能发生变化的设定），这是不允许的。接下来的两节将探讨两种不同的与数据无关的方法，用于在定理3.2中选择 $\lambda$。第一种方法直接取 $\lambda = \lambda_{\alpha_0}$，其中 $\alpha_0$ 是对 $\alpha$的一个与数据无关的“猜测”。第二种方法通过对 $\lambda$中的参数 $E_n^{\mathrm{IWR}}$ 进行积分消去，构造一个新的渐近 e 变量。

### 3.2 选择 $\lambda$，选项I：事前锚定

在这里，我们只需在看到数据之前预先固定某个名义水平 $\alpha_0$，并选择$\lambda = \lambda_{\alpha_0}$。这可以视为我们对 $\alpha$ 将会取何值的最佳猜测（或者说，我们对它取值的期望），并令 $\lambda_0 := \sqrt{2\log(2/\alpha_0)}$。在这种情况下，$g(\lambda_0, \alpha) = \log(2/\alpha)/\sqrt{2\log(2/\alpha_0)} + \sqrt{\log(2/\alpha_0)}/2$，当 $\alpha_0 \approx \alpha$ 时，它将接近$g(\lambda_\alpha, \alpha)$。

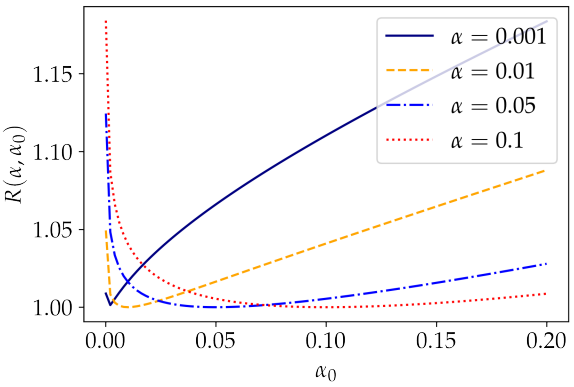



图1：（3.8）中比值的经验评估。这里 $\alpha_0$ 的取值范围为0.001到0.2，并绘制曲线 $\alpha_0 \mapsto R(\alpha, \alpha_0)$，其中 $\alpha$取若干选定值。对所有 $R(\alpha, \alpha_0)$ 以及 $\alpha$ 和 $\alpha_0$ 的取值而言，最大值为1.184。

这种策略以下称为事前锚定，与贝叶斯理论（尤其是 PAC-贝叶斯理论）中的“幸运性”概念有相似之处 [51], ：后验分布越接近先验，界就越紧，但该界对任意先验都有效；另请参见 [59, 图 3]，其中以贝叶斯方法为灵感的 e 值展示了这一现象。我们将 $\alpha_0$ 称为锚点。

正如我们将在模拟中看到的那样，Sec-锚定 3.5，事前锚定在实践中的表现出奇地好，即使对于 $\alpha \gg \alpha_0$也是如此。要了解其中的原因，请注意，选择 $\alpha$ 只会使这些APH-CI的渐近宽度之比相差一个对数因子的平方根。事实上，$g(\lambda_0, \alpha)$ 与 $g(\lambda_\alpha, \alpha)$（定义见（3.6））之比可写为

$$R(\alpha, \alpha_0) := \frac{g(\lambda_0, \alpha)}{g(\lambda_\alpha, \alpha)} = \frac{1}{2}\left(\sqrt{\frac{\log(2/\alpha)}{\log(2/\alpha_0)}} + \sqrt{\frac{\log(2/\alpha_0)}{\log(2/\alpha)}}\right). \tag{3.8}$$

对于合理的 $\alpha$ 和 $\alpha_0$取值，该比率相当小。图 1 绘出了$x \mapsto R(x,\ \alpha_0)$ 在若干 $\alpha_0$取值下的曲线。当 $x$ 介于 0.0001 和 0.2 之间时，该比率从未超过 1.19。因此，根据（3.6），对于此类 $\alpha_0$ 和 $\alpha$取值，$\mathcal{H}_n^{\text{IWR}}$的渐近宽度不会发生明显变化，即使 $|\alpha - \alpha_0| \gg 0$。

然而，人们可能会对 $\mathcal{H}_n^{\text{IWR}}$随 $\log(1/\alpha)$ 线性增长的渐近宽度感到不满意。下一节介绍的 APH-CI 将解决这个问题，使渐近宽度按 $\Theta(\sqrt{\sigma^2 \log(1/\alpha)})$ 的速度增长，与 Wald 置信区间类似。不过，尽管这一方法在理论上颇具吸引力，我们发现事前锚定在实践中效果最佳（第3.5节）。

### 3.3 选择 $\lambda$，选项II：混合方法

我们可以给出不依赖于 $\lambda$ 的渐近 e 值，所用方法是“混合方法”。这一技术由 Herbert Robbins 在现代形式下率先提出，最初用于超鞅的尾不等式 [45]。如今，它已成为任意时刻有效推断的基石，应用于单变量（例如，[33, 28, 62]）和多变量（例如，[12, 7]）情形。下面的引理给出了充分条件，在这些条件下，我们可以将混合方法应用于一族渐近 e 变量，从而得到另一个渐近 e 变量。证明见附录C.4。

**命题 3.3**。设 $\Lambda \subset \mathbb{R}$，并令 $\pi$ 为 $\Lambda$上的任意（与数据无关的）概率分布。假设对每个 $\lambda \in \Lambda$，序列 $(E_n(\lambda))$ 都是某个分布集 $\mathcal{P}$的渐近 *e*变量。如果对于所有 $P \in \mathcal{P}$，$(\sup_{\lambda\in\Lambda} E_n(\lambda))$ 一致可积，那么定义的过程$(E_n(\pi))$

$$E_n(\pi) := \int_\Lambda E_n(\lambda)\mathrm{d}\pi(\lambda), \tag{3.9}$$

是一个渐近 *e* 变量。此外，如果 $(\sup_{\lambda\in\Lambda} E_n(\lambda))$ 是 $\mathcal{P}$-一致可积的，那么 $(E_n(\pi))$是一个 $\mathcal{P}$-一致渐近 *e* 变量。

附录 B.6 给出了更一般的充分条件，在这些条件下，式（3.9）中的混合可以得到一个渐近 e 变量。参见命题B.10。不过，对于本节的目的而言，式（3.9）中的结果已经足够。

备注 3.4。在非渐近设定中，e 变量的任意混合仍然是 e 变量。这可由托内利定理直接推出；无需其他条件。因此我们或许会想知道，除了假设 $(E_n)$ 是渐近 e 变量之外，命题 3.3 是否确实还需要任何附加假设。答案是肯定的。考虑 $E_n(\lambda) = 1 + 2^n \mathbf{1}\left\{\lambda \in (2^{-(n+1)}, 2^{-n}]\right\}$对于 $\lambda \in [0,1]$。令 $\pi$为 $[0,1]$上的均匀分布。当 $n$ 足够大且$\lambda \in [0,1]$固定时，我们有 $E_n(\lambda) = 1$，因此 $\limsup_n \mathbb{E}_P[E_n(\lambda)] = 1$。另一方面，$\int E_n(\lambda)\mathrm{d}\pi(\lambda) = 1 + 1$，因此 $\limsup_n \mathbb{E}_P[E_n^\pi] = 2$。因此，若要使混合也成为渐近 e 变量，就需要满足渐近 e 变量这一条件之外的某些额外条件。

利用 $(E_n^{\text{IWR}})$ 在紧集上一致可积（引理 C.1），我们立即得到以下结果。

**推论 3.5.** 设 $\Lambda \subseteq \mathbb{R}$ 是一个紧集。设 $\pi$ 是 $\Lambda$上的任意（与数据无关的）概率分布。那么，对于任意固定的 $\eta > 0$，混合 $E_n^{\text{IWR}}(\theta;\pi) = \int_\Lambda E_n^{\text{IWR}}(\theta;\lambda)\mathrm{d}\pi$ 是 $\mathcal{M}_2(\theta)$的渐近 *e* 变量。

在非渐近设定中，如果有一族对所有 $E_n(\lambda)$ 均有效的 e 变量，$\lambda \in \mathbb{R}$一种强有力的方法是使用高斯混合（或其他支撑于整个 $\mathbb{R}$的分布）。事实上，这是 Robbins 最初采用的方法 [46]。遗憾的是，推论3.5排除了这种选择。看来，要应用高斯混合，就必须将分布类限制为矩母函数受到严格控制的分布。这不适用于我们的目的，因为我们最多只想假设存在有限个矩。

相反，我们将使用截断高斯分布在有限区域 $[-R, R]$ 上进行混合。此前已有研究采用这种方法构造非渐近置信序列 [6, 7]。为陈述结果，定义如下对象

$$I_R(y,u) := \sqrt{\frac{\pi}{u}} \exp\left\{\frac{y^2}{4u}\right\} (\Phi(g_1) - \Phi(g_2)), \tag{3.10}$$

其中 $\Phi$ 是标准正态分布的累积分布函数，而 $g_1 = R\sqrt{2u} - y/\sqrt{2u}$ 和 $g_2 = -(R\sqrt{2u} + y/\sqrt{2u})$。很容易看出，$I_R(y, u)$ 关于 $y = 0$ 对称。事实上，应将 $I_R(y, u)$ 视为一条开口向上的抛物线，其最小值位于 $y = 0$，其形状和（正）截距由 $u$ 和 $R$ 决定。

对 $E_n^{\text{IWR}}$截断高斯分布进行混合，可得到一个渐近 e 变量，其表达式涉及$I_R(y, u)$。该结果不太容易带来直观理解，因此我们将其放在附录 B.7 中，并立即转向由此得到的 APH-CI。下述定理的证明见附录 C.5。

**定理3.6。** 设 $(X_n)$ 是从 $\text{DAN}(\mu)$ 中某个分布抽取的独立同分布随机变量序列。固定任意的 $R, \kappa > 0$，并令 $Z_{R,\kappa} = \kappa\sqrt{2\pi}(\Phi(R/\kappa) - \Phi(-R/\kappa))$。则过程 $(\mathcal{H}^{\text{MIX}}_{,n}{}^{\text{IWR}}(\alpha; R, \kappa))$ 是 $\mu$ 的一个 APH-CI，其中 $\mathcal{H}^{\text{MIX}}_{,n}{}^{\text{IWR}}(\alpha; R, \kappa) = (X_n \pm W^{\text{MIX}}_{,n}{}^{\text{IWR}}(\alpha; R, \kappa))$，且

$$W_n^{\text{MIX,IWR}}(\alpha; R, \kappa) = \frac{\hat{s}_n y_\alpha^\star}{\sqrt{n - (y_\alpha^\star)^2}}, \tag{3.11}$$

其中 $y_\alpha^\star$ 满足 $I_R(y_\alpha^\star, (\kappa^2+1)/(2\kappa^2)) = Z_{R,\kappa}/\alpha$。如果 (3.11) 中的分母非正，则我们将置信区间取为整个 $\mathbb{R}$。

注意到，由于 $I_R$的形式优美，隐式方程 $I_R(y_\alpha^\star, (\kappa^2 + 1)/(2\kappa^2)) = Z_{R,\kappa}/\alpha$ 可以通过求根轻松地以任意精度近似。因此，定理 3.6 的计算上易于处理，也易于实现。实验结果见第节。

定理 3.6 并非分布一致结果。这是因为目前尚不清楚 $(\sup_{\lambda\in[-R,R]} E_n^{\text{IWR}}(\lambda))$ 是否 $\mathcal{P}$-一致可积，而根据引理 C.1，它在逐点意义下是一致可积的。要证明 $\mathcal{P}$-一致可积性（即使不考虑上确界）似乎也需要将 Gin´e 等人 [21] 关于学生t分布渐近性的研究扩展到分布一致设定。这本身就是一个有趣的问题，我们留待未来探究。

在有限二阶矩假设下，$W^{\text{MIX}}_{,n}{}^{\text{IWR}}$的渐近宽度尤其容易分析。事实上，如果 $\hat{s}_n^2 \xrightarrow{a.s.} \sigma^2$，那么

$$\sqrt{n} W_n^{\text{MIX,IWR}}(\alpha; R, \kappa) \xrightarrow{a.s.} \sigma y_\alpha^\star. \tag{3.12}$$

回顾第 3.1 节中的讨论，我们有 $W_n^{\mathrm{WALD}} \to \sigma z_{1-\alpha/2}$。通过实验，我们发现，对于 $R$ 的大多数合理取值以及 $\kappa$和 $R$ ，这些混合 e 变量的渐近宽度至多是 Wald 置信区间渐近宽度的两倍。此外，无论如何选择 $R$ 和 $\kappa$，这些界在实践中都非常稳定。从解析上看，附录 B.7 中的引理B.13 表明，对于 $\kappa = 1$， $y_\alpha^\star \to 2\sqrt{\log(\sqrt{2}/\alpha)}$ 当$R \to \infty$时。由此，我们得到 $\log(1/\alpha)$ 上所需的平方根行为，这正是第 3.2 节中所缺少的。

### 3.4 事件划分与 R-WS 渐近 e 变量

与其直接定义一个均值渐近受 1 限制的随机变量序列，构造渐近 e 变量的另一种方法是按如下方式进行事件划分。我们取一个在事件 $(A_n)$ 上表现良好的“良好”随机变量序列 $(G_n)$，并根据其表现不佳的概率截断该序列，即 $(A_n^c)$ 的概率。形式化地说，考虑以下观察。

**观察 3.7.** 设$(G_n)$ 是一个非负随机变量序列，且$(A_n)$是一个事件序列，使得$\limsup_n \mathbb{E}_P[G_n \mathbf{1}\{A_n\}] \le 1$ 对所有 $P \in \mathcal{P}$ 均成立。那么，对于任意标量序列$(T_n)$，只要满足 $T_n P(A_n^c) \to 0$ 对所有 $P \in \mathcal{P}$ 均成立， $E_n = G_n \wedge T_n$ 就定义了 $\mathcal{P}$ 的一个渐近 *e* 变量。另一方面，若 $\limsup_n \sup_{P\in\mathcal{P}} \mathbb{E}_P[G_n \mathbf{1}\{A_n\}] \le 1$ 且$T_n \sup_{P\in\mathcal{P}} P(A_n^c) \to 0$，则$(E_n)$ 是一个 $\mathcal{P}$-一致渐近 *e* 变量。

证明。 将 $\mathbb{E}_P[E_n] = \mathbb{E}_P[G_n \mathbf{1}\{A_n\}] + \mathbb{E}_P[T_n \mathbf{1}\{A_n^c\}] = \mathbb{E}_P[G_n \mathbf{1}\{A_n\}] + T_n P(A_n^c)$ 分解。根据假设，第一项随着 $n \to \infty$ 趋于 1，而第二项趋于 0。 □

我们将把观察结果3.7 应用于一个随机变量序列，该序列旨在逼近极限中的过程（$E_n^\star(\lambda)$ 对于任意 $\lambda \in \mathbb{R}$,

$$E_n^\star(\lambda) := \exp\left\{\lambda \sum_{i=1}^n (X_i - \theta) - \frac{\lambda^2}{2} U_n^{\otimes}\right\}, \quad U_n^{\otimes} := \frac{1}{3}\sum_{i=1}^n ((X_i-\theta)^2 + 2\sigma^2). \tag{3.13}$$

Howard 等人 [27, 表 3]表明，对于条件均值恒为 $\theta$ 且方差有限 $\sigma^2$的分布（$X_n$），（$E_n^\star(\lambda)$ 是初始值为 1 的超鞅。因此，在此类分布 $P$下，对于所有停时 $\tau$，它们都满足 $\mathbb{E}_P[E_\tau^\star(\lambda)] \le 1$。以下定理通过构造一个在极限中逼近 $E_n^\star$ 的过程，并对其失败概率进行界定。该定理依赖于 Ruf 和 Waudby-Smith [47]提出的非渐近强大数定律，因此得名 R-WS 渐近 e 变量。证明见附录 C.6。

**定理 3.8。** 设 $(X_n)$ 为一列独立同分布随机变量。固定 $\gamma \in (0, 1/2)$ 和 $\delta > 0$。设 $(T_n) \nearrow \infty$ 和 $(\epsilon_n) \searrow 0$ 为满足 $T_n = o(n^{\delta\gamma/2}\epsilon_n^{2+\delta/2})$ 的非负序列。那么 $(E_n^{\text{R-WS}}(\theta; \lambda, \epsilon_n, T_n))$ 是 $\mathcal{M}_{2+\delta}(\theta)$ 中所有分布的渐近 *e* 变量，其中

$$E_n^{\text{R-WS}}(\theta; \epsilon, T) := \int_{\lambda\in\mathbb{R}} \exp\left(\lambda S_n(\theta) - n(\hat{s}_n^2 + \epsilon_n)\frac{\lambda^2}{2}\right) \mathrm{d}F(\lambda) \wedge T_n, \tag{3.14}$$

且 $F$ 是 $\mathbb{R}$上的任意分布（包括点质量）。此外，若序列$(X_n)$对 $\mathcal{P} \subseteq \mathcal{M}_{2+\delta}(\theta)$ 具有 $\mathcal{P}$-一致有界的 $2+\delta$ 矩，则$(E_n^{\text{R-WS}}(\theta; \lambda, \epsilon_n, T_n))$ 是$\mathcal{P}$-一致渐近 *e* 变量。

正如定理3.8的陈述所表明的那样，截断方法的一个优点是它能够优雅地处理混合，同时避开一致可积性问题。事实上，若$(G_n(\lambda)\wedge T_n)$对所有 $\lambda \in \Lambda$都是渐近 e 变量，则$(\int G_n(\lambda)\,\mathrm{d}F\wedge T_n)$对任意分布 $F$ 在 $\Lambda$上亦然。

特别是，正如我们曾在第 3.3 节想要但未能做到的那样，我们可以取 $F$为上述定理中的中心高斯分布。具体取 $\epsilon_n = 1/\log(n)$ 和 $\gamma = 0.49$，即可得到下面的推论。计算过程是标准的，但为完整起见，仍在附录C.7中给出。

**推论 3.9**。令 $(X_n)$ 为独立同分布随机变量序列。固定 $C, \rho > 0$，并令$\hat{}u_n = n(\hat{}s_n^2 + \log^{-1} n)$。则序列 $(E_n^{\text{R-WS}}(\theta; v))$ 对于 $\mathcal{M}_{2+\delta}(\theta)$ 中的所有分布都是渐近 ***e*** 变量，其中

$$E_n^{\text{R-WS}}(\theta;\rho) = \frac{1}{\sqrt{\rho\hat{u}_n + 1}} \exp\left\{\frac{\rho S_n^2(\theta)}{2\rho\hat{u}_n + 2}\right\} \wedge C \cdot n^{\delta\cdot 0.24}. \tag{3.15}$$

此外，如果序列 $(X_n)$ 的 $2+\delta$ 矩对某个 $\mathcal{P} \subseteq \mathcal{M}_{2+\delta}(\theta)$一致有界，且上界为 $\mathcal{P}$，那么 $(E_n^{\text{R-WS}}(\theta;\rho))$ 是一个 $\mathcal{P}$-一致渐近 ***e*** 变量。

(3.14) 和 (3.15) 中的截断明确限制了 e变量的增长，并直接转化为对所得 APH-CI 大小的限制。事实上，对于截断序列 $(T_n)$，当所有满足 $T_n < 1/\alpha$ 的 $n$ 时，APH-CI 都将是空泛的。例如，在 (3.15) 中，若假设三阶矩 ($\delta = 1$) 存在，那么要使 $\alpha = 0.05$ 处的置信区间不空泛，我们需要 $Cn^{0.24} \geq 1/0.05$，也就是 $n \approx 2.64 \cdot 10^5/C^{1/0.24}$。当然，可以取极大的 $C$ 来弥补，但我们强调，$C$ 不应依赖于数据，且必须预先固定。此外，从证明中可以明显看出，$C$ 决定了渐近性何时开始显现。因此，较大的 $C$ 值会在 $n$取值不大时导致更多假阳性。我们建议将 $C = 100$ 设为标准。如表 1 所示，这一 $C$ 选择所得的经验风险远低于 1。

虽然 R-WS 渐近 e 变量有显式截断，但我们注意到 IWR 定理3.1中的 e 变量及其混合形式都采用隐式截断。具体而言，在事前调参的情形下（第3.2节），由于 $\lambda$ 不能依赖于 $\alpha$，随着 $\alpha$逐渐偏离锚点 $\alpha_0$，e 变量的效力会减弱。而在混合方法中（第3.3节），我们只能在有限范围 $[-R, R]$ 内进行混合，这一事实本身就限制了 e 变量的效力。因此，对于渐近 e 变量，似乎无法避开某种形式的截断，无论是隐式还是显式。我们认为，这一结论普遍成立，并非仅适用于本文提出的 e 变量。至于这一观察能否形式化，仍是一个开放问题。

下面陈述由推论 APH-CI3.9 中的渐近 e 变量得到的。证明沿用常见的公式：我们找出满足 $\theta$ 的那些 $E_n^{\text{R-WS}}(\theta;\ \rho) < 1/\alpha$，同时考虑到，当截断水平低于 $1/\alpha$时，置信区间是空的。由命题2.6即可得到结论。

**定理 3.10**。设$(X_n)$ 是从 $\mathcal{M}_{2+\delta}(\mu)$ 中的某个分布抽取的独立同分布随机变量序列，其中 $\delta > 0$。固定 $C, \rho > 0$，并令 $\hat{}u_n = n(\hat{}s_n^2 +$ 对数 $^{-1}\, n)$。则 $(\mathcal{H}_n^{\text{R-WS}}(\alpha;\rho))$ 是一个 APH-CI，适用于 $\mu$，其中

$$\mathcal{H}_n^{\text{R-WS}}(\alpha;\rho) = (\overline{X}_n \pm W_n^{\text{R-WS}}(\alpha;\rho)) \quad \textit{where} \quad W_n^{\text{R-WS}}(\alpha;\rho) := \sqrt{\frac{2\rho\hat{u}_n + 2}{n^2\rho} \log\left(\frac{\sqrt{\rho\hat{u}_n + 1}}{\alpha}\right)},$$

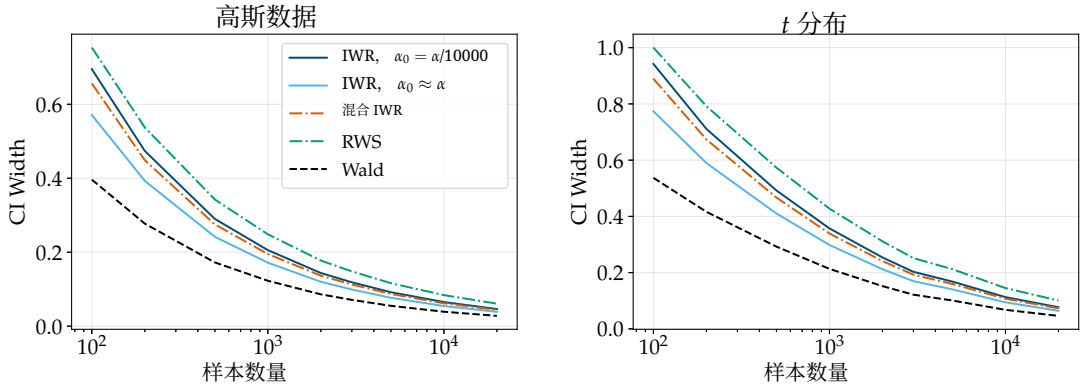



图 2：四种 APH-CI与高斯数据和来自自由度为三的 t 分布的重尾数据所对应的 Wald置信区间的宽度比较。“IWR”指的是定理 APH-CI 的 3.2 ，其中 $\lambda$ 通过事前锚定来选取。“混合 IWR”指的是定理 APH-CI 的 3.6。我们对 $\rho = 2$ 使用 $\mathcal{H}_n^{\text{R-WS}}$。更多细节见附录 D 。

只要 $Cn^{\delta \cdot 0.24} \geq 1/\alpha$，否则 $\mathcal{H}_n^{\text{R-WS}}(\alpha;\ \rho) = \mathbb{R}$。若$(X_n)$ 具有 $\mathcal{P}$-一致有界的$2+\delta$ 矩，则$(\mathcal{H}_n^{\text{R-WS}}(\rho))$ 是一个 $\mathcal{P}$-一致的 APH-CI。

在定理3.10 中， $\rho$ 取合理值时， $\rho = 2$是一个合理的选择，但我们在附录 D 中证明了，$\rho$ 的合理取值对宽度的影响不大。事实上，渐近宽度与 $\rho$无关，并按$\sqrt{\log(n)/n}$的比例变化。更具体地说，通过一些算术运算可得 $\sqrt{n}W_n^{\text{R-WS}}(\alpha;\ \rho) \xrightarrow[n\to\infty]{a.s.} \infty$，并且

$$\sqrt{\frac{n}{\log n}} W_n^{\text{R-WS}}(\alpha;\rho) \xrightarrow[n\to\infty]{a.s.} \sigma, \tag{3.16}$$

正如 Robbins 最初提出的混合方法（见 [6, 附录 C.3]）。请注意，这一速率 $\sqrt{\log(n)/n}$不同于先前的 APH-CI类所达到的速率 $1/\sqrt{n}$ 。在时间一致置信序列的文献中，这种速率很常见，而置信区间则不然。由于重对数律 [45]，置信序列无法按 $1/\sqrt{n}$ 的速率缩放；Robbins 的混合方法是构造此类对象的一种经典方法。这提示我们，$(\mathcal{H}_n^{\text{R-WS}}(\alpha;\rho))$ 或许满足了比目前为止给出的其他 APH-CI类更强的保证。事实确实如此。实际上，$(\mathcal{H}_n^{\text{R-WS}})$（经过轻微修改的版本）是事后渐近置信序列的一个例子，而 $(E_n^{\text{R-WS}})$ 是渐近 e 过程的一个例子。我们将在第4节介绍这些对象。

### 3.5 模拟

图2绘出了第 APH-CI 节讨论的三个 APH-CIS，其定义分别见定理 3、3.2、3.6 和 3.10，并将它们的宽度与式（3.7）中的 Wald 置信区间宽度进行比较。正如预期，Wald 置信区间的宽度始终更小（当然，它并非事后有效），且随着样本量增加，差距逐渐缩小。Wald 置信区间与我们的 APH-CIS 之间的宽度差通常小于 0.05，适用于$n = 10^4$。

对于 $\mathcal{H}_n^{\text{IWR}}$我们通过事前锚定（第 $\lambda$ 节）来选择 3.2。也就是说，我们固定 $\alpha_0$ 并取 $\lambda = \lambda_0 = \sqrt{2\log(2/\alpha_0)}$。我们针对锚点 $\alpha_0$检验两种情形：当 $\alpha_0 \ll \alpha$ （相差

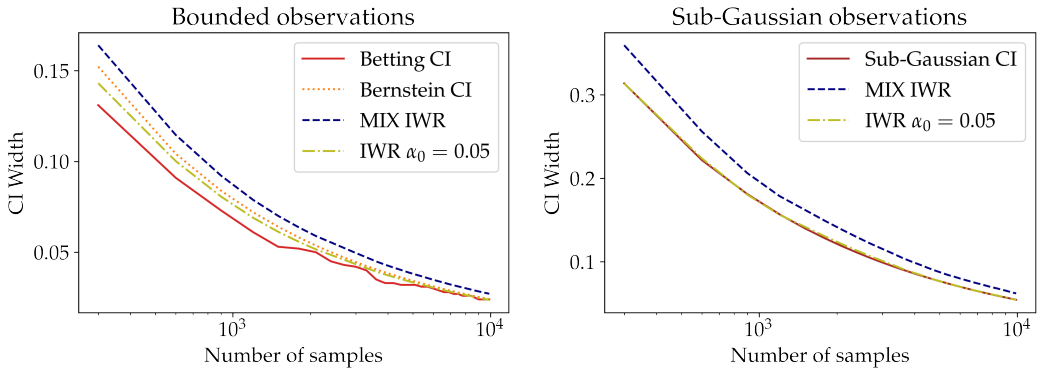



图 3：比较我们的一些 APH-CI与基于（非渐近）e 变量的非渐近置信区间的宽度。对于有界数据（我们将其设为独立同分布的伯努利分布（0.25）），我们将其与经典伯恩斯坦置信区间，以及 Waudby-Smith 和 Ramdas 提出的当前最先进的基于投注的经验伯恩斯坦置信区间进行比较 [60]。对于次高斯观测值（生成方式为 $N$（0，1）随机变量），我们将其与基于 Chernoff 方法的标准置信区间进行比较 —— 更多细节请参阅正文。我们在所有模拟中使用 $\alpha = 0.05$，并使用 $\mathcal{H}_n^{\text{IWR}}$，其中$\lambda = \sqrt{2\log(2/\alpha)}$ 。

$10^4$)以及当 $\alpha_0 \approx \alpha$时。当 $\alpha_0 \approx \alpha$时， $\mathcal{H}_n^{\text{IWR}}$比 $\mathcal{H}^{\text{MIX}}{}_{,n}$ IWR更紧。事实上，正如第3.2节所预示的，只有当 $\alpha$ 和 $\alpha_0$ 之间存在显著差异时， $\mathcal{H}^{\text{MIX}}{}_{,n}$ IWR才会更紧。因此，从业者需要在 $\mathcal{H}_n^{\text{IWR}}$和 $\mathcal{H}^{\text{MIX}}{}_{,n}$ IWR之间进行权衡：对于 $\alpha$ 和 $\alpha_0$的许多取值，前者会更紧，但后者的最坏情况表现更好。

基于 APH-CI 的 R-WS e 变量的 $\mathcal{H}_n^{\text{IWR}}$比 $\mathcal{H}^{\text{MIX}}{}_{,n}$和 IWR都更宽松 IWR。正如我们在第3.4 节中所暗示的，并将在第4节中进一步阐述，这主要是因为$\mathcal{H}_n^{\text{R-WS}}$提供了比另外两个 APH-CI更强的保证，实际上是一个事后渐近置信序列。

一个自然的问题是，依赖渐近 e 变量的大样本置信区间的宽度，与基于非渐近e 变量的有限样本置信区间相比如何。图3比较了基于 APH-CI的两个置信区间的宽度 IWR e变量与若干适用于有界观测值和次高斯观测值的知名有限样本置信区间。对于来自伯努利分布（0.25）的独立同分布有界观测值，我们将其与经典伯恩斯坦置信区间以及 Waudby-Smith 和 Ramdas 最近提出的界 [60]（投注置信区间）进行比较。后者是文献中最紧的经验伯恩斯坦界（即不需要知道方差），同时在理论上也近乎最优 [52]。

$\mathcal{H}_n^{\text{IWR}}$的宽度与预言机伯恩斯坦置信区间相当或更小，但比投注置信区间宽。值得强调的是，投注置信区间和预言机伯恩斯坦界都利用了观测值有界这一事实，而这两个 APH-CI 则没有这一要求，对所有具有有限二阶矩的分布均有效。

我们针对次高斯数据进行了类似的实验，将两种 APH-CI与基于 Chernoff 方法的次高斯置信区间进行比较，即 $(X_n \pm \sqrt{2\log(2/\alpha)/(n\sigma^2)})$。同样，$\mathcal{H}_n^{\text{IWR}}$的表现大致与有限样本置信区间相当，而 $\mathcal{H}^{\text{MIX}}{}_{,n}$ IWR则比两者都宽松。图3 的主要目的是考察渐近置信区间与非渐近置信区间的宽度，而非渐近置信区间的事后性质。因此，我们对$\mathcal{H}_n^{\text{IWR}}$采用事前锚定，并将锚点设为 $\alpha = \alpha_0$。

| 方法 | 实际风险 $(0 < \alpha \leq 1)$ | 理论风险 $(0 < \alpha)$ |
|---|---|---|
| 实数集$_n(\mathcal{H}_n^{\text{WALD}})$ | 7.089 | |
| 实数集$_n(\mathcal{H}_n^{\text{IWR}})$ | 0.532 | 0.620 |
| 实数集$_n(\mathcal{H}^{\text{MIX}}_{,\,n}{}^{\text{IWR}})$ | 0.367 | 0.976 |
| 实数集$_n(\mathcal{H}^{\text{MIX}}_{,\,n}{}^{\text{R-WS}})$ | 0.001 | 0.028 |

表 1：数据依赖选择 $\alpha$时的经验风险。“实际风险”列报告在范围 $0 < \alpha \leq 1$ 内搜索显著性时的风险。“理论风险”列将搜索范围扩展至 $\alpha_{\max}$ ，以确保拒绝原假设。对于 Wald 置信区间，两种情形下的风险相同，因为 $\alpha^\star$ 几乎必然为 $\leq 1$。

我们的第三个实验研究了我们的 APH-CI相较于 Wald 置信区间的风险，模拟分析师以依赖数据的方式选择 $\alpha$ 的情形。我们从均值为零、方差为四的高斯分布中生成 $n = 1000$ 个样本，并比较 Wald 置信区间$\mathcal{H}_n^{\text{WALD}}$与 APH-CI $\mathcal{H}_n^{\text{IWR}}(\lambda = \sqrt{2\log(2/0.01)})$，以及两个混合 APH-CI$\mathcal{H}^{\text{MIX}}_{,\,n}{}^{\text{IWR}}$（$R = 20$ 和 $\kappa = 1$）和 $\mathcal{H}^{\text{MIX}}_{,\,n}{}^{\text{R-WS}}$（$\rho = 2$）。为验证风险保证，我们模拟一种“显著性操纵”流程：分析师搜索能够拒绝原假设 $H_0 : \theta = 0$的最小显著性水平 $\alpha^\star$。具体而言，在每次试验中，我们确定：$\alpha^\star = \inf\{\alpha > 0 : 0 \notin \mathcal{H}_n(\alpha)\}$，通过在区间$[\alpha_{\min}, \alpha_{\max}]$ 上进行二分搜索，容差为 $10^{-5}$ （此处 $\alpha_{\min} = 10^{-5}$ 和 $\alpha_{\max} = 500$）。注意，这一流程在操作上等同于估计上确界风险 $\sup_\alpha \mathbf{1}\{0 \notin \mathcal{H}_n(\alpha)\}/\alpha$。我们在两种条件下评估风险：

1. **实用情形**（$0 < \alpha \leq 1$）：搜索范围限制在统计推断的可解释极限内。若对于所有 $\alpha \leq 1$，置信区间都包含零，则风险贡献为零。

2. **理论情形**（$0 < \alpha$）：在实验中，我们将搜索区间扩展至 $\alpha_{\max}$ =500。我们根据经验选取这一上界，因为它足够大，确保在全部 1000 次模拟试验中都拒绝原假设。虽然 $\alpha > 1$ 缺乏标准的统计解释，但强制该方法在每次试验中都拒绝原假设，使我们能够基准评估该界限的紧致程度。

表1 展示了结果。正如预期，Wald 置信区间得出的风险水平远高于 1，而我们的方法 APH-CI则维持了有效的风险控制。（$\mathcal{H}^{\text{MIX}}_{,\,n}{}^{\text{R-WS}}$）的风险在两种情形下都远低于 1，这反映出它比其他 APH-CI方法更保守。正如第 3.4 节所述，这是因为它提供了比（$\mathcal{H}_n^{\text{IWR}}$）或（$\mathcal{H}^{\text{MIX}}_{,\,n}{}^{\text{IWR}}$）更强的保证——见第 4 节。

请注意，在实际情形下，对于所有 APH-CIs，风险都严格低于 1。这种保守性是实现事后有效性所必须付出的代价：理论界必须考虑在任意 $\alpha > 0$ 处进行拒绝的可能性，包括 $\alpha > 1$ 在统计上毫无意义但在数学上可能的取值。

## 4 事后序贯推断与渐近 e 过程

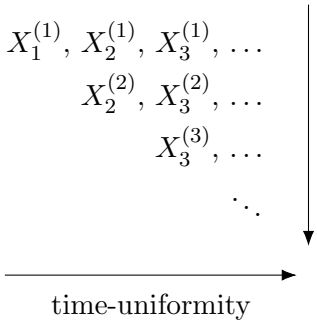
到目前为止，讨论重点一直是渐近置信区间、假设检验和 p 值的事后类比。关于统计推断程序的序贯 （而非事后）类比，相关文献日益增多，

其形式分别为置信序列、序贯假设检验和任意时刻 p 值。这些序贯程序源自 Wald [56, 57], 、Robbins 及其同事的研究 [45, 11, 36]，近年来相关研究日益活跃；参见 Ramdas 和 Wang [41], 的第 7 章、Ramdas 等人撰写的综述论文 [42], 以及其他相关论文 [28, 60, 40]。此外，Waudby-Smith 等人 [61]借助近年来分布一致强渐近理论 [63, 64] 和分布逐点任意时刻有效推断 [46, 62, 4]，发展出了这些程序的分布一致渐近版本。

然而，现有方法均未被证明具备本文所说的事后有效性，而且事实上，“事后序贯推断”究竟应如何定义也并不明确先验 。本节将定义事后渐近序贯推断，并推导出满足这些定义的明确程序。

### 4.1 定义与期望性质

在接下来的定义中，我们将频繁用到“上三角阵列”这一概念；该对象常见于时间一致渐近理论中 [61]。随机变量的上三角阵列是一个双重索引序列 $((X_k^{(m)})_{k\geq m})_{m\in\mathbb{N}}$，之所以这样称呼，是因为它可以直观地表示为




其中，“渐近性”是指 $m$ 趋于无穷，而“时间一致性”则是针对那些索引 $k \geq m$而言。了解了这些术语后，请看以下定义。

**定义 4.1**（渐近事后置信序列与任意时刻 p 值）。 设 $\mathcal{P}$为一个分布集合， $f\colon \mathbb{N}\to(0,1)$ 为一个非增函数。我们称由上三角阵列 $((\mathcal{H}_k^{(m)}(\alpha))_{k\geq m})_{m\in\mathbb{N}}$ 构成的集合（其中 $\alpha\in(0,1)$）是参数 $(\theta(P))_{P\in\mathcal{P}}$ 的分布一致渐近事后置信序列 (APH-CS)，并称其具有范围增长函数 $f$，如果对任意停时 $\tau$，均有

$$\limsup_{m\to\infty}\sup_{P\in\mathcal{P}}\mathbb{E}_P\left[\sup_{\alpha>f(m)}\frac{\mathbf{1}\left\{\theta(P)\notin\mathcal{H}_{\tau\vee m}^{(m)}(\alpha)\right\}}{\alpha}\right]\leq 1. \tag{4.1}$$

同样，对于分布子集 $\mathcal{P}_0\subseteq\mathcal{P}$，我们称上三角阵列$((p_k^{(m)})_{k\geq m})_{m\in\mathbb{N}}$ 是一个 $\mathcal{P}_0$-一致的渐近事后任意时刻$p$值，其范围增长函数为 $f$，如果对任意停时 $\tau$，

$$\limsup_{m\to\infty}\sup_{P\in\mathcal{P}_0}\mathbb{E}_P\left[\sup_{\alpha>f(m)}\frac{\mathbb{1}\{p_{\tau\vee m}^{(m)}\leq\alpha\}}{\alpha}\right]\leq 1. \tag{4.2}$$

让我们简要解析定义4.1，并将注意力转向展示式（4.1）。首先假设（4.1）中的保证对固定的 $m$成立，而不仅仅是在极限情形下。在这种情况下，集合$\mathcal{H}_\tau^{(m)}$在所有停时 $\tau$ 上都是事后置信区间，且这些停时的取值至少为 $m$ ，但仅针对那些 $\alpha$ 使得 $\alpha > f\left(m\right)$ 。也就是说，我们设想一位分析者承诺首先收集 $m$个观测值并进行分析。基于这些结果，他们可以继续抽样，且结果仍然有效（即，只要 $\alpha > f\left(m\right)$ ，风险就始终以一为上界）。换言之，$\mathcal{H}_\tau^{(m)}$在初始预热期 $m$之后满足可选延续。一个 APH-CS的渐近保证不过是当预热期 $m$ 趋于 $\infty$时成立的那种保证。

根据上述事后渐近序贯推断程序的定义，自然会产生这样一个问题：渐近 *e* 值的序贯类比是什么？正如 e 过程 [41, §7] 是 e 值的序贯类比，以下渐近 e 过程的定义也构成了渐近 e 值的序贯类比。尽管渐近 e 值和非渐近 e 过程已在文献中出现并得到研究，但以下定义似乎是新的。

**定义 4.2**（渐近 e 过程）。 设 $\mathcal{P}$ 为一个分布集合。若对于任意停时 $\tau$，均有如下结论，则称上三角阵列 $((E_k^{(m)})_{k\geq m})_{m\in\mathbb{N}}$ 为分布一致的渐近 $\mathcal{P}$-*e* 过程：

$$\limsup_{m\to\infty}\sup_{P\in\mathcal{P}}\mathbb{E}_P\left[E_{\tau\vee m}^{(m)}\right]\leq 1. \tag{4.3}$$

正如渐近 e 值是构造渐近事后置信区间和 p 值的充要条件，渐近 e 过程也是构造定义4.1中所述程序的充要条件。以下命题将上述陈述形式化，并作为命题 2.6的序贯类比。

**命题 4.3**（关于渐近 e 过程的必要性与充分性）。 假设$\mathcal{H}_k^{(m)}\left(\cdot\right)$ 构成针对 $\theta^\star$ 的 APH-CS 。那么存在一个函数族 $(E_k^{(m)}\left(\theta\right))_{\theta\in\mathbb{R}}$ ，使得对于任意 $\alpha\in(0,\ 1)$， $\mathcal{H}_k^{(m)}\left(\alpha\right)$ 可写为

$$\mathcal{H}_k^{(m)}(\alpha)=\left\{\theta\in\mathbb{R}:E_k^{(m)}(\theta)<1/\alpha\right\}, \tag{4.4}$$

且 $E_k^{(m)}\left(\theta^\star\right)$ 构成事后渐近 $\mathcal{P}$-*e* 过程。反之，若 $E_k^{(m)}\left(\theta^\star\right)$ 构成渐近 $\mathcal{P}$-*e* 过程，则由下式定义的集合 (4.4) 对任意 $\alpha\in(0,\ 1)$ 构成一个 APH-CS。

证明。 设 $\mathcal{H}_k^{(m)}\left(\cdot\right)$ 是针对 $\theta^\star$ 的 APH-CS 。对于任意 $\theta\in\mathbb{R}$ 以及任意满足 $k\geq m$ 的 $k$，$m\in\mathbb{N}$ ，定义

$$E_k^{(m)}(\theta):=\sup\left\{\frac{1}{\alpha}:\alpha>0\text{ and }\theta\notin\mathcal{H}_k^{(m)}(\alpha)\right\}=\sup_{\alpha>0}\frac{\mathbf{1}\left\{\theta\notin\mathcal{H}_k^{(m)}(\alpha)\right\}}{\alpha}. \tag{4.5}$$

根据 $\mathcal{H}_k^{(m)}$ （$\alpha$）构成一个 APH-CS （针对 $\theta^\star$）的性质，可知 $E_k^{(m)}$ （$\theta^\star$）是渐近e过程。证明的其余部分与命题2.6的证明中的逐情形分析完全相同。

□

请注意，本节所述的所有定义均处于分布一致设定中。当然，若有需要，可将分布集合 $\mathcal{P}$ 取为单点集，从而给出逐点类比；但我们将继续在一致设定下讨论，因为我们在下一节给出的 APH-CS 自然满足这一准则。

### 4.2 渐近事后混合置信序列

这里我们注意到， R-WS 第 3.4 节中的渐近 e 变量可用于定义一个渐近 e 过程。证明见附录C.8。

**命题 4.4**。 固定 $\delta > 0$ ，并假设 $X_n$ 是独立同分布的，且其 $\mathcal{P}$ 阶矩一致有界，其中 $2+\delta$ 阶矩对应某个 $\mathcal{P} \subseteq \mathcal{M}_{2+\delta}(\theta)$ 。对于 $\gamma \in (0, 1/2)$ ，假设 $T_n = o(n^{\delta\gamma/2}/log^{2+\delta/2} n)$ 。定义过程上三角阵列 $(E_k^{(m)}(\theta))_{k \geq m}$ 为

$$E_k^{(m)}(\theta) := \int_{\lambda \in \mathbb{R}} \exp\left\{\lambda S_k(\theta) - k(\hat{s}_k^2 + \log^{-1}(m))\frac{\lambda^2}{2}\right\} \mathrm{d}F(\lambda) \wedge T_m, \tag{4.6}$$

其中 $F$ 是一个概率测度，其支撑集包含于 $\mathbb{R}$ 的某个子集。则 $((E_k^{(m)}(\theta))_{k \geq m})_{m \in \mathbb{N}}$ 构成一个 $\mathcal{P}$-一致渐近 $e$ 过程。

现在，我们可以结合命题 4.3 和4.4 ，给出以下 APH-CS。与定理 3.10 中一样，我们取 $F$ 为中心高斯分布，并令 $\gamma = 0.49$。随后，混合项的计算方式与推论 3.9 证明中的方式完全相同。

**定理 4.5** （事后混合置信序列与任意时刻 p 值）。 设 $(X_n)$ 为一列独立同分布随机变量，对某个 $\delta > 0$ 具有 $\mathcal{P}$-一致有界的$(2+\delta)$阶矩，且 $\mathcal{P} \subseteq \mathcal{M}_{2+\delta}(\mu)$。固定 $C$、$\rho > 0$，并令 $\hat{u}_k^{(m)} = k(\hat{s}_k^2 + \log^{-1} m)$。则

$$\mathcal{H}_k^{(m)}(\alpha) := \left(\overline{X}_k \pm \sqrt{\frac{2\rho\hat{u}_k^{(m)} + 2}{k^2\rho} \log\left(\frac{\sqrt{\rho\hat{u}_k^{(m)} + 1}}{\alpha}\right)}\right), \tag{4.7}$$

构成 $\mu$ 的一个 $\mathcal{P}$-一致 APH-CS，其范围增长函数为 $f(m) = C \cdot m^{-0.24}$。

注意到（4.7）与 Waudby-Smith 等人的某些渐近置信序列在形式上基本相同 [62]。对定理4.5 的另一种解读是：在对方差估计稍作修改后，上述渐近置信序列也享有 $\mathcal{P}$-一致的事后保证。从这个角度看，尽管本文引入的某些方法可被视为为满足事后风险控制而专门构造的，定理 4.5则用于“升级”文献中已有结果的保证。

## 5 结论

我们将渐近推断的范式扩展到事后设定，从而允许根据数据选择显著性水平。我们的方法既有优点，也有缺点。尽管它比传统技术灵活得多，但我们

获得的是风险控制方面的保证，而非错误概率；错误概率是大多数推断方法通用的衡量标准。

我们的目标并非反对传统推断，而是为一种替代性的统计方法论奠定基础，同时保留严格的频率学派保证。我们并不需要以此取代标准的 p 值和置信区间，也可以在报告它们的同时，报告本文研究的事后变体。

尽管我们的重点是给出事后置信区间、p 值和假设检验的具体构造，但我们的结果依托于渐近 e 变量理论方面的创新。具体而言，我们提出了新的渐近 e 变量，介绍了构造这类变量的一般方法，并将现有方法的适用范围扩展到分布一致设定。

我们还提出了“事后渐近序贯推断”这一概念，将渐近 e 变量以及渐近事后置信区间和 p 值的概念扩展到序贯设定。具体而言，我们提出了渐近 e 过程和事后渐近置信序列，并各自给出了一个示例。事后推断和序贯推断各自都蕴含丰富的研究空间，因此我们预计，两者交集处还有许多值得探讨的问题。

对于有意采用渐近事后置信区间的实践科学家而言，所有APH-CI都各有其优点。这两个APH-CI基于 IWR 渐近e变量者表现最佳，其中（$\mathcal{H}_n^{\mathrm{IWR}}$）在以事前锚定实例化时是最紧的 APH-CI ，在合理范围内针对 $\alpha$ 以及锚点 $\alpha_0$，而（$\mathcal{H}^{\mathrm{MIX}}{}_{,\ n}{}^{\mathrm{IWR}}$）在最坏情形下更紧。 APH-CI 基于 R-WSe变量者则略松一些，但正如我们在第4 节中所见，它提供了更强的保证，并且不仅在事后 $\alpha$ 选择方面，也在样本量方面提供了灵活性。鉴于可选延续十分普遍，我们推荐后者作为最安全的 APH-CI实践方案。

总体而言，我们希望已卓有成效地将方兴未艾的事后推断领域拓展至渐近情形，为从业者提供更灵活的推断工具，从而支持数据驱动决策。

## 致谢

作者感谢 Anastasios Angelopoulos 和 Arun Kumar Kuchibhotla 提供的宝贵讨论。BC 和 AR 感谢美国国家科学基金会通过 IIS-2229881 和 DMS-2310718 资助。EG 感谢谷歌博士奖学金的支持。IW-S 感谢米勒基础科学研究所的支持。

本项目部分由欧盟（ERC-2022-SYG-OCEAN-101071601）资助。本文表达的观点和意见仅代表作者本人，不一定反映欧盟或欧洲研究理事会执行局的观点。欧盟和资助机构均不对此承担责任。

本出版物是“市场与学习”讲席项目的一部分，该项目由液化空气集团、法国巴黎银行资产管理欧洲、法国电力公司、Orange 和法国国家铁路公司资助；上述机构为法国国家信息与自动化研究所基金会的赞助方。

本研究还获得了法国政府的资助，该资助由法国国家研究署管理，属于法国2030计划，项目编号为“PR[AI]RIE- PSAI”（ANR-23-IACL-0008）。

# 参考文献

[1] Anastasios Nikolas Angelopoulos、Stephen Bates、Adam Fisch、Lihua Lei 和 Tal Schuster。共形风险控制。载于《国际学习表征会议》第十二届会议，2024年。4[2] Raghu R Bahadur 和 Leonard J Savage。非参数问题中某些统计程序的不存在性。《数学统计学年鉴》，27(4):1115–1122，1956年。2[3] Vidmantas Bentkus 和 Friedrich G¨otze。关于 Student 统计量的 Berry-Esseen 界。《概率年刊》，24(1):491–503，1996年。13, 45[4] Aurelien Bibaut、Nathan Kallus 和 Michael Lindon。针对可能存在依赖关系的观测值的近最优非参数序贯检验与置信序列。arXiv 预印本 arXiv:2212.14411，2022年。22[5] Patrick Billingsley。《概率与测度》。Wiley-Interscience 出版物。Wiley，纽约[等]，3版，1995年。11, 37, 44, 54[6] Ben Chugg 和 Aaditya Ramdas。标量和矩阵的闭式经验Bernstein置信序列。arXiv 预印本 arXiv:2512.21300，2025年。16, 19[7] Ben Chugg、Hongjian Wang 和 Aaditya Ramdas。随机向量均值的时间一致置信球。《机器学习研究汇刊》，2025年。15, 16, 52[8] Ben Chugg、Tyron Lardy、Aaditya Ramdas 和 Peter Gr¨unwald。论事后假设检验中的可容许性。《国际近似推理杂志》，第109634页，2026年。3, 4, 6, 8[9] 钟开莱。强大数定律。载于《第二届伯克利数理统计与概率研讨会论文集》，第2卷，第341–353页。加利福尼亚大学出版社，1951年。9[10] Reuven Dar、Ronald C Serlin 和 Haim Omer。三十年来心理治疗研究中对统计检验的误用。《咨询与临床心理学杂志》，62(1):75，1994年。3[11] Donald A. Darling 和 Herbert E. Robbins。均值、方差和中位数的置信序列。《美国国家科学院院刊》，58 1: 66–8，1967年。22[12] Victor H De la Pena、黎子良和邵启满。《自归一化过程：极限定理与统计应用》。施普林格，2009年。15, 45[13] Charles-Jean De La Vall´ee Poussin。论勒贝格积分。《美国数学学会会刊》，16(4):435–501，1915年。42[14] David L Demets 和 KK Gordon Lan。期中分析：α消耗函数方法。《医学统计学》，13(13-14):1341–1352，1994年。3[15] 美国食品药品监督管理局。新冠肺炎疫苗安全性监测：主动监测主方案。主方案，美国食品药品监督管理局，2021年2月。网址：https://bestinitiative.org/wp-content/uploads/2021/02/C19-Vaccine-Safety-Protocol-2021.pdf。CBER 监测计划；BEST计划。3[16] Lasse Fischer 和 Aaditya Ramdas。基于投注的序贯蒙特卡洛检验。《英国皇家统计学会会刊B辑：统计方法学》，第 qkaf014 页，2025年。11

[17] Lasse Fischer 和 Konstantinos Sechidis。《低维情形下的 Knockoff 方法：事后调整名义水平，在控制错误发现率的同时提高检验效能》。arXiv 预印本 arXiv:2511.11166，2025年。6[18] Etienne Gauthier、Francis Bach 和 Michael I Jordan。反向保形预测。载于《神经信息处理系统进展》，2025年。3, 6[19] Etienne Gauthier、Francis Bach 和 Michael I Jordan。e值拓展了保形预测的适用范围。arXiv 预印本 arXiv:2503.13050，2025年。3, 6[20] Etienne Gauthier、Francis Bach 和 Michael I Jordan。保形预测中的自适应覆盖策略。载于人工智能与统计国际会议，2026年。3, 6[21] Evarist Gin´e、Friedrich Goetze 和 David Mason。Student $t$-统计量何时渐近标准正态？《概率年刊》，25，1997年7月。doi: 10.1214/aop/1024404523。13, 16, 42, 44, 45[22] Steven N Goodman。P 值、假设检验与似然：一场被忽视的历史争论对流行病学的启示。《美国流行病学杂志》，137(5):485–496，1993年。3[23] Steven N Goodman。迈向循证医学统计。1：P 值谬误。《内科学年鉴》，130(12):995–1004，1999年。3[24] Peter Gr¨unwald、Rianne de Heide 和 Wouter Koolen。安全检验。《皇家统计学会杂志B辑：统计方法学》，86(5):1091–1128，2024年。5[25] Peter D Gr¨unwald。e 后验。皇家学会哲学汇刊 A 辑，381(2247):20220146，2023年。6[26] Peter D Gr¨unwald。超越 Neyman–Pearson：E 值使依赖数据的 alpha 假设检验成为可能。《美国国家科学院院刊》，121(39):e2302098121，2024年。3, 4, 5, 6, 7, 10[27] Steven R Howard、Aaditya Ramdas、Jon McAuliffe 和 Jasjeet Sekhon。基于非负上鞅的时间一致 Chernoff 界。《概率调查》，17:257–317，2020年。17, 31, 52[28] Steven R Howard、Aaditya Ramdas、Jon McAuliffe 和 Jasjeet Sekhon。时间一致的非参数非渐近置信序列。《统计学年鉴》，49(2):1055–1080，2021年。15, 22[29] Raymond Hubbard 和 M. J. Bayarri。P 值不是错误概率。工作论文 03-26，杜克大学统计与决策科学研究所，北卡罗来纳州达勒姆，2003年。3[30] Raymond Hubbard 和 Mar´ıa Jes´us Bayarri。经典统计检验中证据度量（p 值）与错误（$\alpha$）的混淆。《美国统计学家》，57(3):171–178，2003年。3[31] Nikolaos Ignatiadis、Ruodu Wang 和 Aaditya Ramdas。渐近与复合 e 值：多重检验与经验贝叶斯。arXiv 预印本 arXiv:2409.19812，2024年。4, 9, 10, 11, 12, 30, 35[32] Eugene Katsevich 和 Aaditya Ramdas。结构化、回归及在线场景下假发现比例的同时高概率界。《统计学年鉴》，48(6):3465–3487，2020年。5[33] Emilie Kaufmann 和 Wouter M Koolen。重新审视混合鞅及其在序贯检验和置信区间中的应用。《机器学习研究杂志》，22(246):1–44，2021年。15

[34]Nick W Koning。连续检验：统一检验与 e 值。arXiv 预印本 arXiv:2409.05654，2024年。3, 4, 6[35] Nick W Koning。事后 $\alpha$ 假设检验与事后 $p$值。arXiv 预印本 arXiv:2312.08040，2025年。3, 4, 6, 7, 10[36] Tze Leung Lai。关于置信序列。《统计学年鉴》，4(2):265–280，1976年。22[37] 李克昭。非参数回归的诚实置信域。《统计学年鉴》，17(3):1001–1008，1989年。4, 9[38] Paul Andr´e Meyer。《概率与位势》。Blaisdell 出版公司，马萨诸塞州沃尔瑟姆，1966年。42[39] Thomas Mikosch。正则变化、次指数性及其在概率论中的应用。EURANDOM 讲义，1999年。12[40] Francesco Orabona 和 Kwang-Sung Jun。由通用投资组合的遗憾导出的紧集中界与置信序列。《IEEE信息论汇刊》，70(1):436–455，2023年。22[41] Aaditya Ramdas 和 Ruodu Wang。使用 e 值进行假设检验。《统计学基础与趋势》，2025年。5, 6, 22, 23[42] Aaditya Ramdas、Peter Gr¨unwald、Vladimir Vovk 和 Glenn Shafer。博弈论统计与安全任意时刻有效推断。《统计科学》，38(4):576–601，2023年。5, 22[43] Zhimei Ren 和 Rina Foygel Barber。去随机化 knockoff：利用 e 值控制错误发现率。《皇家统计学会杂志B辑：统计方法学》，86(1):122–154，2024年。35[44] Alessandro Rinaldo、Larry Wasserman 和 Max G'Sell。用于高维、弱假设依赖推断的自助法与样本拆分。《统计学年鉴》，47(6)，2019年。9[45] Herbert Robbins。与重对数律相关的统计方法。《数学统计学年鉴》，41(5):1397–1409，1970年。15, 19, 22[46] Herbert Robbins 和 David Siegmund。维纳过程与样本和的越界概率。《数学统计学年鉴》，第1410–1429页，1970年。16, 22[47] Johannes Ruf 和 Ian Waudby-Smith。强大数律与重对数律的集中不等式。arXiv 预印本 arXiv:2511.00175，2025年。12, 17, 49, 51[48] Evert G Schouten、Jacqueline M Dekker、Frans J Kok、Saskia Le Cessie、Hans C Van Houwelingen、Jan Pool 和 Jan P Vandnbroucke。病例队列设计中的风险比与率比估计：高血压与心血管死亡率。《医学统计学》，12(18):1733–1745，1993年。3[49] Glenn Shafer。打赌式检验：一种统计与科学交流策略。《英国皇家统计学会会刊A辑：社会统计》，184(2):407–431，2021年。5[50] Rajen D Shah 和 Jonas Peters。条件独立性检验的难点及广义协方差测度。《统计学年鉴》，48(3):1514，2020年。9, 12[51] John Shawe-Taylor、Peter L Bartlett、Robert C Williamson 和 Martin Anthony。依赖数据的层级结构上的结构风险最小化。《IEEE信息论汇刊》，44(5):1926–1940，2002年。14

[52] Shubhanshu Shekhar 和 Aaditya Ramdas。关于有界均值投注置信集的近最优性。arXiv 预印本 arXiv:2310.01547，2023年。20[53] MJ Symons 和 DT Moore。风险率比与前瞻性流行病学研究。《临床流行病学杂志》，55(9):893–899，2002年。3[54] Ryan J Tibshirani、Alessandro Rinaldo、Rob Tibshirani 和 Larry Wasserman。模型选择后的统一渐近推断与自助法。《统计学年鉴》，46(3):1255，2018年。9[55] Bengt von Bahr 和 Carl-Gustav Esseen。随机变量和的第 r 阶绝对矩不等式，$1 \leq r \leq 2$。《数学统计学年鉴》，第299–303页，1965年。56[56] Abraham Wald。统计假设的序贯检验。《数学统计学年鉴》，16(2):117–186，1945年。22[57] Abraham Wald。《序贯分析》。Wiley，纽约，1947年。22[58] Ruodu Wang 和 Aaditya Ramdas。利用 e值控制错误发现率。《英国皇家统计学会会刊B辑：统计方法学》，84(3):822–852，2022年。5, 35[59] Ian Waudby-Smith 和 Aaditya Ramdas。不放回抽样的置信序列。《神经信息处理系统进展》，33:20204–20214，2020年。14[60] Ian Waudby-Smith 和 Aaditya Ramdas。通过投注估计有界随机变量的均值。《英国皇家统计学会会刊B辑：统计方法学》，86(1):1–27，2024年。5, 11, 20, 22[61] Ian Waudby-Smith、Edward H Kennedy 和 Aaditya Ramdas。分布一致的任意时刻有效序贯推断。arXiv 预印本 arXiv:2311.03343，2023年。9, 12, 22[62] Ian Waudby-Smith、David Arbour、Ritwik Sinha、Edward H Kennedy 和 Aaditya Ramdas。时间一致中心极限定理与渐近置信序列。《统计学年鉴》，52(6):2613–2640，2024年。15, 22, 24[63] Ian Waudby-Smith、Martin Larsson 和 Aaditya Ramdas。分布一致的大数强律。arXiv 预印本 arXiv:2402.00713，2024年。9, 13, 22[64] Ian Waudby-Smith、Martin Larsson 和 Aaditya Ramdas。非渐近且分布一致的 Koml´os-Major-Tusn´ady 逼近。arXiv 预印本 arXiv:2502.06188，2025年。9, 22[65] Ziyu Xu、Ruodu Wang 和 Aaditya Ramdas。基于 e值的置信区间的选择后推断。《电子统计学杂志》，18(1):2292–2338，2024年。5, 7, 10[66] Ziyu Xu、Aldo Solari、Lasse Fischer、Rianne de Heide、Aaditya Ramdas 和 Jelle Goeman。为错误发现率控制带来闭包：多重检验的一般原则。arXiv 预印本 arXiv:2509.02517，2025年。6

# A 关于事后置信区间的其他定义

命题 2.6 表明，通过对随机变量进行阈值化而构造的事后置信区间 $E$ 是有效的，当且仅当 $E$ 是一个 e变量。有人可能会问，放宽 e变量的定义是否会产生相应放宽的事后置信区间。

考虑以下 近似 e值和 p值的定义，最初由 Ignatiadis 等人给出 [31, 定义 3。1]。我们称 $Q$ 是 $\epsilon$， $\delta$)- 近似 p 值，用于 $\mathcal{P}$，如果

$$P(Q \leq \alpha) \leq (1+\epsilon)\alpha + \delta, \quad \text{for all } \alpha \in (0,1) \text{ and all } P \in \mathcal{P}. \tag{A.1}$$

我们称 $E$ 是 $\mathcal{P}$ 的（$\epsilon$， $\delta$）近似 e 值，当且仅当

$$\mathbb{E}_P[E \wedge m] \leq 1 + \epsilon + \delta m, \quad \text{for all } m > 0 \text{ and } P \in \mathcal{P}. \tag{A.2}$$

遵循 Ignatiadis 等人 [31]，的说法，若对所有 $P \in \mathcal{P}$， $Q_n$ 都是某个 $\epsilon_n$（$P$）$\to 0$、 $\delta_n$（$P$）$\to 0$ 的（$\epsilon_n$（$P$） $\delta_n$（$P$））近似 p 值，则称（$Q_n$）是 $\mathcal{P}$ 的 弱渐近 ***p*** 值 。类似地定义弱渐近 ***e*** 变量 。请注意，Ignatiadis 等人 [31]并未使用“弱渐近”这一术语。我们强调，这些渐近 p 变量和 e 变量的定义不同于正文中给出的定义。事实上，Ignatiadis 等人 [31,的命题 3.6] 表明，弱渐近 e 变量的概念比我们在正文中使用的渐近 e 变量更弱（因此得名）。

基于近似 p 变量和 e 变量，我们可以引入泛函 $\theta$ 的（$\epsilon$， $\delta$）近似事后置信区间的概念，将其定义为满足以下条件的集合 $\mathcal{H}$（$\alpha$）：

$$\mathbf{R}(\mathcal{H}; \epsilon, \delta, P) \leq 1, \quad \text{where } \mathbf{R}(\mathcal{H}; \epsilon, \delta, P) = \mathbb{E}_P\left[\sup_{\alpha>0} \frac{\mathbf{1}\{\theta(P) \notin \mathcal{H}(\alpha)\}}{(1+\epsilon)\alpha + \delta}\right]. \tag{A.3}$$

动机与第 2.1 节相同：我们从近似保证$P(\theta(P) \notin \mathcal{H}(\alpha)) \leq (1+\epsilon)\alpha + \delta$出发，并将其转换为事后保证。随后，我们可以将弱渐近事后置信区间定义为一个集合序列 $(\mathcal{H}_n)$，使得 $\mathcal{H}_n$ 是一个 $(\epsilon_n(P)$ $\delta_n(P))$-近似事后置信区间，并且对于所有 $P \in \mathcal{P}$， $\epsilon_n(P)$ $\delta_n(P) \xrightarrow{n\to\infty} 0$。

关于弱渐近/近似 e 变量是否能产生事后置信区间，可以提出几个问题。遗憾的是，我们将在下文证明，其中大多数问题的答案是否定的。在开始之前，需要注意的是，一个 $(\epsilon, \delta)$-近似事后置信区间的风险可以写成

$$\begin{aligned}\mathbb{E}_P\left[\sup_{\alpha>0} \frac{\mathbf{1}\{E^\theta \geq 1/\alpha\}}{(1+\epsilon)\alpha + \delta}\right] &= \mathbb{E}_P\left[\sup_{\alpha \geq 1/E^\theta} \frac{1}{(1+\epsilon)\alpha + \delta}\right] \\ &= \mathbb{E}_P\left[\frac{1}{(1+\epsilon)1/E^\theta + \delta}\right] = \mathbb{E}_P\left[\frac{E^\theta}{1+\epsilon+\delta E^\theta}\right],\end{aligned}$$

w此时事后置信区间的形式为 $\mathcal{H}(\alpha) = \{\theta : E^\theta < 1/\alpha\}$，正如命题 2。 6.

**1. 近似 e 变量是否会产生近似的事后置信区间?** 似乎不会。考虑由以下定义的（0， $\delta$）近似 e 变量 $E$：

$$P(E = 1) = 1 - \delta, \quad P(E = u) = \delta,$$

对于某个足够大的 $u > 0$。很容易验证， $E$ 是 (0， $\delta$)-近似的（因此对于任意 $\epsilon > 0$，也是 $(\epsilon$， $\delta$)-近似的）：

$$\mathbb{E}_P[E \wedge m] = (1 \wedge m)(1-\delta) + (u \wedge m)\delta \leq 1 - \delta + m\delta \leq 1 + \delta m.$$

然而， $E$ 导致的风险严格大于 1。事实上，

$$\mathbb{E}_P\left[\frac{E}{1+\epsilon+\delta E}\right]=\frac{1-\delta}{1+\epsilon+\delta}+\frac{u\delta}{1+\epsilon+u\delta}\xrightarrow{u\to\infty}\frac{1-\delta}{1+\epsilon+\delta}+1>1,$$

只要 $\delta>0$，该式即成立。

**2. 弱渐近 e 变量能否得到渐近事后置信区间?** 此处所谓 APH-CI是指定义2.4，即本文其余部分所研究的通常概念。同上，答案似乎是否定的。考虑随机变量序列（$E_n$），其中 $P(E_n=1)=1-\delta_n$且 $P(E_n=u_n)=\delta_n$，其中 $\delta_n\to 0$ 且 $u_n\gg 1/\delta_n$。于是同上，$\mathbb{E}_P[E_n\wedge m]\le 1+\delta_n m$对每个 $n$成立，因此（$E_n$）是弱渐近 e-变量。但是

$$\mathbb{E}_P\left[\sup_{\alpha>0}\frac{\mathbf{1}\{E_n\ge 1/\alpha\}}{\alpha}\right]=\mathbb{E}_P[E_n]=1-\delta_n+u_n\delta_n\to\infty,$$

其中第一个等式由命题2.6的证明得出。

**3. 弱渐近 e 变量能否得到弱渐近事后置信区间?** 答案同样是否定的。使用与上文相同的弱渐近 e 变量，我们有

$$\mathbb{E}_P\left[\frac{E_n}{1+\epsilon_n+\delta_n E_n}\right]=\frac{1-\delta_n}{1+\epsilon_n+\delta_n}+\frac{u_n\delta_n}{1+\epsilon_n+u_n\delta_n}\xrightarrow{n\to\infty}2>1.$$

# B 附加结果

## B.1 二元替代渐近 e-变量

下面介绍另外两个渐近 e 变量。对于参数 $\lambda$，$\eta$，以及序列 $(\lambda_n)$，定义以下对象

$$E_n^{\mathrm{REG}}(\theta;\lambda,\eta):=\exp\left(\frac{\lambda S_n(\theta)}{\sqrt{n}\widehat{\sigma}_n+\eta V_n(\theta)}-\frac{\lambda^2}{2}\right). \tag{B.1}$$

$$E_n^{\mathrm{SN}}(\theta;\lambda^n):=\exp\left(\sum_{i\le n}\lambda_i(X_i-\theta)-\frac{1}{6}\sum_{i\le n}\lambda_i^2[(X_i-\theta)^2+2\widehat{\sigma}_{i-1}^2]\right), \tag{B.2}$$

上标 SN 表示“自归一化”，之所以这样命名，是因为它基于 Howard 等人 [27, 表 3]中描述的自归一化过程，并通过该过程得到证明。请注意，它是用参数序列 $(\lambda_n)$ 而非单个值定义的，因此依赖于 $S_n(\theta)$ 和 $V_n(\theta)$ 的加权版本。我们出于对渐近 e 变量的独立兴趣而介绍 $E_n^{\mathrm{SN}}$。它不会产生有用的 APH-CI。

与此同时， REG 表示“正则化”，因为将项 $\eta V_n(\theta)$ 视为正则化项会很有帮助。事实上，我们可能会认为（或希望）可以在渐近 e 变量中用$\widehat{\sigma}_n$ 替换 $V_n(\theta)$。具体来说，我们可以考虑 $T_n:=\exp(\frac{\lambda}{\sqrt{n}}S_n(\theta)/\widehat{\sigma}_n-\lambda^2/2)$。然而，事实并非如此：对于高斯分布$X_1,\dots,X_n$，$\log(T_n)$服从 t 分布，因此没有矩母函数。因此， $\mathbb{E}[T_n]$ 不存在，更不用说收敛了。在分母中添加项 $\eta V_n(\theta)$，可以对 $(T_n)$ 的增长进行正则化，从而确保该序列一致可积，并成为渐近 e 变量。

现在我们来陈述这些对象在什么条件下构成渐近 e 变量。证明见附录 C.9。

**定理 B.1**。 设$(\lambda_n)$ 是一个非负标量序列，几乎必然收敛到某个有限的$\lambda \in \mathbb{R}$ 。那么：

*1.* 若 $\lambda_n$ 对每个 $n$都是 $\mathcal{F}_n$-可测的， $\eta_n \to \eta$ 几乎必然成立，且对所有 $n$， $\eta_n > 0$ 和 $|\lambda_n| \lesssim \eta_n$，则$(E_n^{\mathrm{REG}}(\theta; \lambda_n, \eta_n))$ 对于 $\mathcal{M}_2(\theta)$中的所有分布，都是渐近 *e* 变量。*2.* 若$(\lambda_n)$ 是可预测的（即，对每个 $n$， $\lambda_n$ 都是 $\mathcal{F}_{n-1}$-可测的），且满足 *(i)* $\sup_n \lambda_n < \infty$，*(ii)* $\sum_{i\le n} \lambda_i^2 \xrightarrow{n\to\infty} L < \infty$ 几乎必然成立，以及 *(iii)* $\lambda_i^2 / \sum_{j\le n} \lambda_j^2 \xrightarrow{n\to\infty} 0$ 几乎必然成立，则$(E_n^{\mathrm{SN}}(\theta; \lambda_n))$是 $\mathcal{M}_2(\theta)$的渐近 *e* 变量。

上述表述在将 $\widehat{\sigma}_n^2$ 替换为（有偏）样本方差 $s_n^2 = n^{-1}\sum_{i\le n}(X_i - X_n)^2$, 或任何其他几乎必然收敛于真实方差的统计量时仍然成立。

值得强调的是，对于 $E_n^{\mathrm{REG}}$，当我们说 $\eta > 0$ 是固定的时，我们的意思是不仅它不能依赖数据，而且也不能依赖样本量 $n$。这与非渐近推断不同，在非渐近推断中，固定参数可以根据样本量进行优化（例如，在 Chernoff 方法中）。

我们还可以将（$E_n^{\mathrm{REG}}$）扩展到分布一致设定，如下所示。证明见附录 C.10。

**定理 B.2**。 设 （$X_n$）为均值为 $\theta$ 且具有 $\mathcal{P}$一致有界偏度的独立同分布序列。设 （$\lambda_n$）为有界的确定性序列，并收敛于 $\lambda$。假设 $0 < \inf_{P\in\mathcal{P}} \sigma_P$，并设 （$\eta_n$）为满足 $\lambda_n \lesssim \eta_n$和 $\eta_n > 0$的确定性序列。则 （$E_n^{\mathrm{REG}}$（$\theta$； $\lambda_n, \eta$）） 是一个 $\mathcal{P}$-一致渐近 *e* 变量。

## B.2 基于 $E_n^{\mathrm{reg}}$ 的 AnAPH-CI

接下来，我们给出与渐近 e 变量（$E_n^{\mathrm{REG}}$）相对应的 APH-CI 。通过算术运算可以写出基于（$E_n^{\mathrm{SN}}$）的 APH-CI，但它是若干个互不相交的区间的并。此外，虽然其中一个区间包含样本均值 $X_n$ ，且其宽度与 $\mathcal{H}_n^{\mathrm{IWR}}$ 和 $\mathcal{H}_n^{\mathrm{REG}}$ 类似，但最外侧区间的表现很奇怪，会延伸至 $\pm\infty$，导致所得置信区间毫无意义。

以下结果的证明见附录C.11。

**定理 B.3**。 设 $X_1, \ldots, X_n$ 为独立同分布且具有有限方差、均值为 $\mu$ 的随机变量。对于任意 $\lambda > 0$ 和 $\alpha \in (0,1)$，令

$$A_{n,\alpha}(\lambda) = \frac{\log(2/\alpha) + \lambda^2/2}{\lambda\sqrt{n}}.$$

对于任意固定的 $\lambda, \eta > 0$，二者均独立于 $\alpha$，$(\mathcal{H}_n^{\mathrm{REG}}(\alpha; \lambda, \eta))$ 是一个适用于 $\mathbb{E}X_1$ 的 APH-CI，其 $e$

$$\mathcal{H}_n^{\mathrm{REG}}(\alpha; \lambda, \eta) := (\overline{X}_n \pm W_n^{\mathrm{REG}}(\alpha; \lambda, \eta)), \quad W_n^{\mathrm{REG}}(\alpha; \lambda, \eta) := \frac{A_{n,\alpha}(\lambda)\widehat{\sigma}_n(1 + \sqrt{B_{\alpha,\eta}(\lambda)})}{1 - A_{n,\alpha}^2(\lambda)\eta^2}, \tag{B.3}$$

和$B_{\alpha,\ \eta}(\lambda) = 1 - (1 - A_{n,\ \alpha}^2(\lambda)\eta^2)(1 - \eta^2\frac{n-1}{n})$。

虽然 $\lambda$ 和 $\eta$ 都可以随 $n$ 在定理B.1 中变化 $E_n^{\mathrm{REG}}$，但在定理 B.3中它们是固定的。这是因为我们希望$A_{n,\,\alpha}(\lambda)$中的两项以相同的速率衰减，这就要求 $\lambda$ 为常数。进而，由于定理 B.1 要求 $\lambda_n \lesssim \eta_n$，因此 $\eta$ 也必须为常数。

注意　根据强大数定律和连续映射定理各应用一次 ，

$$\sqrt{n}W_n^{\mathrm{REG}}(\alpha;\lambda,\eta) \xrightarrow[n\to\infty]{a.s.} \sigma g(\lambda,\alpha)(1+\eta),$$

这严格大于第 $\mathcal{H}_n^{\mathrm{IWR}}$ 节中给出的（3.1 $\eta$）的极限宽度，因为 $g$（$\lambda$，$\alpha$）$=\log(2/\alpha)/\lambda+\lambda/2$ 必须为正。（回想一下，$\mathcal{H}_n^{\mathrm{REG}}$）的性能将在附录D中通过实验进行探究，同时也会考察 $\eta$ 对界限宽度的影响。

在实践中，我们建议固定 $\eta = 10^{-3}$。尽管从理论上讲，$\eta$ 可以任意小，但小于 $10^{-3}$ 的取值对宽度没有明显影响。此外，下一小节中的命题 B.7表明，较小的 $\eta$ 会增大 $\mathcal{H}_n^{\mathrm{REG}}$的第一类错误。因此，$\eta$ 取值过小可能会导致错误率膨胀，却无法相应改善置信区间的宽度。

## B.3 APH 置信区间的精确渐近第一类错误

正文中提出的 APH-CIs 旨在将事后风险控制在 1 以下。一个自然的次要问题是评估其经典渐近第一类错误，即在固定（与数据无关的）显著性水平 $\alpha$ 下未覆盖的概率。

令 $Z \sim \mathcal{N}(0,1)$，其累积分布函数记为 $\Phi$。根据中心极限定理和大数定律，在任意分布 $P \in \mathcal{M}_2(\theta^\star)$ 下，自归一化比率依分布收敛于标准正态分布：

$$\frac{S_n(\theta^\star)}{V_n(\theta^\star)} = \frac{S_n(\theta^\star)/\sqrt{n}}{V_n(\theta^\star)/\sqrt{n}} \xrightarrow{d} Z.$$

我们首先研究由 IWR e变量定义的 APH-CI。

**命题 B.4。** *令 $\mathcal{H}_n^{\mathrm{IWR}}(\alpha;\ \lambda) = \{\theta :\ E_n^{\mathrm{IWR}}(\theta;\ \lambda) < 2/\alpha$ 和 $E_n^{\mathrm{IWR}}(\theta;\ -\lambda) < 2/\alpha\}$ 为针对任意 $\lambda > 0$、由 IWR 渐近 e 值诱导的 APH-CI。则：*

$$\lim_{n\to\infty} P\big(\theta^\star \notin \mathcal{H}_n^{\mathrm{IWR}}(\alpha;\lambda)\big) = 2\left(1-\Phi\left(\frac{\log(2/\alpha)}{\lambda}+\frac{\lambda}{2}\right)\right). \tag{B.4}$$

*证明。* 取对数后，未覆盖事件等价于：

$$\lambda\frac{S_n(\theta^\star)}{V_n(\theta^\star)} - \frac{\lambda^2}{2} \ge \log(2/\alpha) \quad \text{or} \quad -\lambda\frac{S_n(\theta^\star)}{V_n(\theta^\star)} - \frac{\lambda^2}{2} \ge \log(2/\alpha),$$

这等价于

$$\left|\frac{S_n(\theta^\star)}{V_n(\theta^\star)}\right| \ge \frac{\log(2/\alpha)}{\lambda}+\frac{\lambda}{2}.$$

根据 连续映射定理，代入极限 $S_n(\theta^\star)/V_n(\theta^\star) \xrightarrow{d} Z$ 可得：

$$\lim_{n\to\infty} P\big(\theta^\star \notin \mathcal{H}_n^{\mathrm{IWR}}(\alpha;\lambda)\big) = 2\left(1-\Phi\left(\frac{\log(2/\alpha)}{\lambda}+\frac{\lambda}{2}\right)\right),$$

完成论证。 □

注意，如果我们设定 $\lambda_\alpha = \sqrt{2\log(2/\alpha)}$，则渐近第一类错误为 $2(1-\Phi(\sqrt{2\log(2/\alpha)}))$，对于任意 $\alpha \in (0,1)$，该值都小于 $\alpha$，这反映了为实现事后有效性所付出的代价。

接下来我们研究由截断高斯分布上的 APH-CI 混合 IWR e变量所定义的对象；参见第3.3节。

**Pr命题 B.5。** 由 $n^{\text{IWR}}$ 诱导的渐近第一类错误为：

$$\lim_{n\to\infty} P\big(\theta^\star \notin \mathcal{H}_n^{\text{MIX,IWR}}(\alpha; R, \kappa)\big) = 2(1-\Phi(y_\alpha^\star)), \tag{B.5}$$

*wh*其中 $y_\alpha^\star$ 是隐式方程 $I_R(y_\alpha^\star, \frac{\kappa^2+1}{2\kappa^2}) = Z_{R,\kappa}/$的唯一正根 —— $\alpha$.

证明。 根据定义，

$$E_n^{\text{MIX,IWR}}(\theta^\star; R, \kappa) = Z_{R,\kappa}^{-1} I_R\left(\frac{S_n(\theta^\star)}{V_n(\theta^\star)}, \frac{\kappa^2+1}{2\kappa^2}\right).$$

由于函数 $y \mapsto I_R(y,u)$ 连续，我们可以应用连续映射定理，得到：

$$E_n^{\text{MIX,IWR}}(\theta^\star; R, \kappa) \xrightarrow{d} Z_{R,\kappa}^{-1} I_R\left(Z, \frac{\kappa^2+1}{2\kappa^2}\right).$$

因此，渐近第一类错误为：

$$\lim_{n\to\infty} P\big(\theta^\star \notin \mathcal{H}_n^{\text{MIX,IWR}}(\alpha; R, \kappa)\big) = P\left(I_R\left(Z, \frac{\kappa^2+1}{2\kappa^2}\right) \geq Z_{R,\kappa}/\alpha\right).$$

函数 $I_R(y,u)$ 关于 $y=0$ 对称，且关于 $|y|$ 严格单调递增，因此，当且仅当 $|Z| \geq y_\alpha^\star$ 时，阈值不等式成立，其中 $y_\alpha^\star$ 是隐式方程 $I_R(y_\alpha^\star, \frac{\kappa^2+1}{2\kappa^2}) = Z_{R,\kappa}/\alpha$ 的唯一正根。因此，渐近第一类错误为：

$$\lim_{n\to\infty} P\big(\theta^\star \notin \mathcal{H}_n^{\text{MIX,IWR}}(\alpha; R, \kappa)\big) = 2(1-\Phi(y_\alpha^\star)). \qquad \square$$

最后，我们研究由渐近 e 变量 R-WS 给出的 APH-CI 。

**命题B.6。** 在定理3.10的设定下，由 $\mathcal{H}_n^{\text{R-WS}}$ 引起的渐近第一类错误趋于零：

$$\lim_{n\to\infty} P\big(\theta^\star \notin \mathcal{H}_n^{\text{R-WS}}(\alpha; \rho)\big) = 0. \tag{B.6}$$

证明。 注意，当 $n$ 充分大时，阈值条件 $Cn^{\delta\cdot 0.24} \geq 1/\alpha$ 成立，且 $\mathcal{H}_n^{\text{R-WS}}(\alpha;\ \rho) = (\overline{X}_n \pm W_n^{\text{R-WS}}(\rho))$。因此，未覆盖事件等价于：

$$\left|\overline{X}_n - \theta^\star\right| > W_n^{\text{R-WS}}(\rho).$$

经 $\sqrt{n}/\sigma$ 缩放后，未覆盖事件可等价地写为：

$$\frac{\sqrt{n}\left|\overline{X}_n - \theta^\star\right|}{\sigma} > \frac{\sqrt{n}}{\sigma} W_n^{\text{R-WS}}(\rho).$$

根据中心极限定理和连续映射定理，左侧依分布收敛到 $|Z|$。然而，正如正文所述，整个缩放后的阈值趋于无穷：

$$\frac{\sqrt{n}}{\sigma} W_n^{\text{R-WS}}(\rho) \xrightarrow[n\to\infty]{a.s.} \infty.$$

令 $B > 0$ 任意。对于充分大的 $n$，我们几乎必然有 $\frac{\sqrt{n}}{\sigma} W_n^{\text{R-WS}}(\rho) \geq B$，由此可得：

$$P\left(\frac{\sqrt{n}\left|\overline{X}_n - \theta^\star\right|}{\sigma} > \frac{\sqrt{n}}{\sigma} W_n^{\text{R-WS}}(\rho)\right) \leq P\left(\frac{\sqrt{n}\left|\overline{X}_n - \theta^\star\right|}{\sigma} > B\right).$$

右侧收敛到 $2(1-\Phi(B))$，因此：

$$\limsup_{n\to\infty} P\left(\frac{\sqrt{n}\left|\overline{X}_n - \theta^\star\right|}{\sigma} > \frac{\sqrt{n}}{\sigma} W_n^{\text{R-WS}}(\rho)\right) \leq 2(1-\Phi(B)).$$

由于 $B > 0$ 任意，我们得出：

$$\lim_{n\to\infty} P\big(\theta^\star \notin \mathcal{H}_n^{\text{R-WS}}(\alpha;\rho)\big) = 0. \qquad \square$$

这种误差恰为零的结果凸显了渐近置信区间与置信序列之间根本的结构差异。由于置信序列必须扩大一个 $\sqrt{\log(n)}$ 因子，才能应对无限时域内的停时，其宽度仅以 $\sqrt{\log(n)/n}$ 的速率缩小，最终远大于实际的 $O_P(1/\sqrt{n})$ 抽样误差，从而使第一类错误降为零。

此外，我们研究了 REG APH-CI 在小节 B.1中引入的对象的渐近第一类错误。证明与命题B.4的证明基本相似，注意到

$$\frac{S_n(\theta^\star)}{\sqrt{n}\widehat{\sigma}_n + \eta V_n(\theta^\star)} \xrightarrow{d} \frac{1}{1+\eta} Z$$

这由斯卢茨基定理得出，因此我们略去证明。

**命题 B.7**。设 $\mathcal{H}_n^{\text{REG}}(\alpha;\lambda,\eta) = \{\theta\colon E_n^{\text{REG}}(\theta;\lambda,\eta) < 2/\alpha$ 和 $E_n^{\text{REG}}(\theta;-\lambda,\eta) < 2/\alpha\}$是由适用于任意 $\lambda,\eta > 0$ 的 REG 渐近 *e* 值所诱导的 *APH-CI*。那么：

$$\lim_{n\to\infty} P\big(\theta^\star \notin \mathcal{H}_n^{\text{REG}}(\alpha;\lambda,\eta)\big) = 2\left(1 - \Phi\left((1+\eta)\left(\frac{\log(2/\alpha)}{\lambda} + \frac{\lambda}{2}\right)\right)\right). \tag{B.7}$$

## B.4 通过复合渐近 e 变量构造 APH-CI

复合 e 变量自然产生于多重检验中；例如在 e-BH 程序 [58] 和去随机化敲除法 [43] 中；事实上，[31] 表明，复合 e 值是所有错误发现率控制程序的基础。它们指的是一组 $K$ 非负随机变量$E^j$、$j = 1, \ldots\ldots, K$，若 $\mathcal{P}$ 满足 $\sup_{P\in\mathcal{P}} \sum_j \mathbb{E}_P[E^j] \leq K$，则称其为复合 *e* 变量。因此，若 $\mathcal{P}$ 满足 if，我们将随机变量的集合（$E_n^j$）、$j = 1, \ldots\ldots, K$ 称为渐近复合 e 变量

$$\sup_{P\in\mathcal{P}} \limsup_{n\to\infty} \sum_{j\leq K} \mathbb{E}_P[E_n^j] \leq K. \tag{B.8}$$

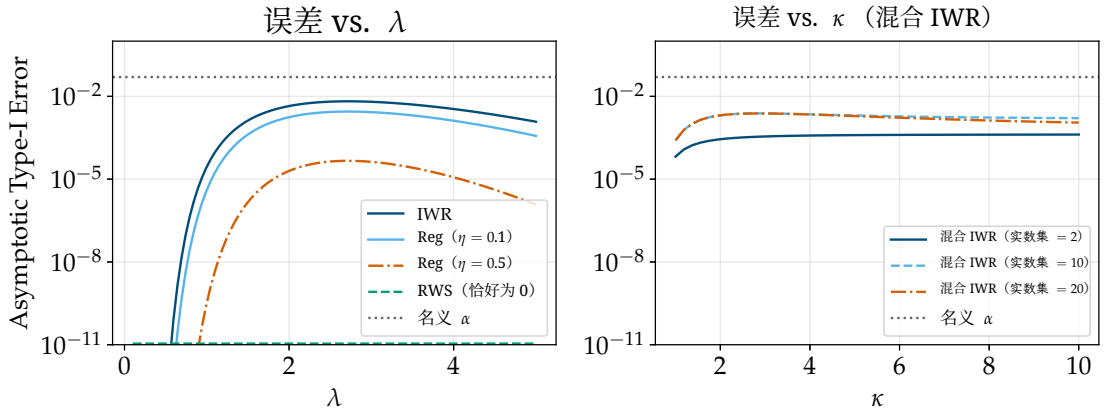



图 4：所提出的 APH-CIS 的渐近第一类错误随各自调优参数变化的情况。名义显著性水平固定为 $\alpha = 0.05$（黑色虚线）。由于真实的渐近错误远低于名义水平，这反映了确保事后有效性所付出的保守代价，因此 y 轴采用对数尺度。左图：IWR 和 REG（$\eta \in \{0.1, 0.5\}$）在不同 $\lambda$取值下的错误率。R-WS 错误率绘制在对数尺度图的底部，因为其理论渐近第一类错误恰好为零。右图：MIX、IWR 在不同截断半径（$R \in \{2, 10, 20\}$）下，随 $\kappa$ 变化的错误率。请注意，当截断界限较大时（$R \geq 10$），该分布与未截断高斯分布十分接近，因此 $R = 10$ 和 $R = 20$ 曲线几乎重合。

同样地，若 $\mathcal{P}$ 满足 if，则 $(E_n^j)$ 是 $\mathcal{P}$-一致的渐近复合 e 变量

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \sum_{j\leq K} \mathbb{E}_P[E_n^j] \leq K. \tag{B.9}$$

渐近复合 e 变量按如下方式产生 APH-CIS。

**命题 B.8**。对于 $j = 1, .., K$ 以及每个 $\theta \in \Theta$，令 $(E_n^j(\theta))$ 为一组随机变量。对于所有 $\alpha \in (0、1)$，令 $\mathcal{H}_n(\alpha) = \{\theta : \ E_n^j(\theta) < K/\alpha$ 对所有 $j \in [K]\}$均成立。则 $(\mathcal{H}_n)$ 是针对 $\theta^\star$ 的一个（一致）APH-CI，如果 $(E_n^j(\theta^\star))$ 是（一致的）渐近复合 *e* 变量。

证明。论证与命题 2.6类似。对于每个 $n$ 和 $\alpha$，若 $\cup_j\{E_n^j(\theta^\star) \geq K/\alpha\}$发生，则 $1/\alpha \leq E_n^j(\theta^\star)/K$ 对于某个 $j$成立，因此

$$\frac{\mathbf{1}\left\{\cup_j\{E_n^j(\theta^\star) \geq K/\alpha\}\right\}}{\alpha} \leq \max_j \frac{E_n^j(\theta^\star)}{K} \leq \frac{1}{K}\sum_j E_n^j(\theta^\star).$$

取期望并求极限，

$$\limsup_{n\to\infty} \mathbb{E}_P\left[\sup_\alpha \frac{\mathbf{1}\{\theta^\star \notin \mathcal{H}_n(\alpha)\}}{\alpha}\right] \leq \limsup_{n\to\infty} \frac{1}{K}\sum_j \mathbb{E}_P[E_n^j(\theta^\star)] \leq 1,$$

论证即告完成。在分布一致设定中，我们只需在求极限之前对分布取上确界。

□

在本工作中，我们通常在$(E_n^1)$与$(E_n^2)$仅为两个 e 变量时应用该结果（例如，在定理 3.2中）。

## B.5 期望中的分布一致收敛

在此，我们给出非负随机变量序列构成分布一致的渐近 e 变量的充分条件。在逐点情形下，为构造渐近e变量，我们依赖于 Billingsley [5, 定理 25。12] 该定理给出了随机变量序列依期望收敛的条件。特别地，他证明一致可积性与依分布收敛共同蕴含依期望收敛。由于随机变量序列（$X_n$）到 $X$的依分布收敛等价于 $\mathbb{E}[\min(X_n, K)] \to \mathbb{E}[\min(X, K)]$ 对所有 $K > 0$ 成立，下述结果可视为将 Billingsley 的结果推广到分布一致设定。

**命题 B.9.** 设 $g : \mathbb{R} \to \mathbb{R}_{\geq 0}$。设 $\mathcal{P}$ 为一组分布，对于每个 $P \in \mathcal{P}$，令$(T_{n,P})_{n\geq 1}$ 为依赖于 $P$ 的统计量序列。设 $Z$ 为一个分布为 $R$ 的随机变量，使得 $\mathbb{E}_R[g(Z)] < \infty$。进一步假设

1. $\lim_{K\to\infty} \limsup_{n\to\infty} \sup_P \mathbb{E}_P[g(T_{n,P})\mathbf{1}\{g(T_{n,P}) \geq K\}] = 0$, *and*
2. 对所有 $K > 0$,
$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \mathbb{E}_P[g(T_{n,P}) \wedge K] \leq \mathbb{E}_R[g(Z) \wedge K].$$

*Then*

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} \mathbb{E}_P[g(T_{n,P})] \leq \mathbb{E}_R[g(Z)]. \tag{B.10}$$

此外，若将条件 *(2)* 替换为对所有 $K > 0$，上极限$_{n\to\infty}$上确界$_{P\in\mathcal{P}}$ $|\mathbb{E}_P[g(T_{n,P})\wedge K]-\mathbb{E}_R[g(Z)\wedge K]| = 0$，则可得上极限$_{n\to\infty}$上确界$_{P\in\mathcal{P}}$ $|\mathbb{E}_P[g(T_{n,P})]-\mathbb{E}_R[g(Z)]|$ =0。若将所有的上极限$_n$ 替换为上确界$_n$，同样的结论依然成立。

证明。在证明过程中，我们用 $T_n$ 代替 $T_{n,P}$；$P$ 的含义由上下文确定。对于任意 $x$ 和 $K > 0$, 注意到我们可以写成 $g(x) = g(x) \wedge K + 0 \wedge (g(x) - K) \leq g(x) \wedge K + g(x)\mathbf{1}\{g(x) \geq K\}$。因此,

$$\mathbb{E}_P[g(T_n)] \leq \mathbb{E}_P[g(T_n) \wedge K] + \mathbb{E}_R[g(T_n)\mathbf{1}\{g(T_n) \geq K\}].$$

对所有 $P$取上确界，然后令 $n \to \infty$趋于极限，我们得到

$$\begin{aligned}\limsup_n \sup_P \mathbb{E}_P[g(T_n)] &\leq \limsup_n \sup_P \left(\mathbb{E}_P[g(T_n) \wedge K] + \mathbb{E}_R[g(T_n)\mathbf{1}\{g(T_n) \geq K\}]\right) \\ &\leq \mathbb{E}_R[g(Z) \wedge K] + \limsup_n \sup_P \mathbb{E}_R[g(T_n)\mathbf{1}\{g(T_n) \geq K\}]\end{aligned}$$

利用假设 (1)。令 $K \to \infty$上式右侧的第二项根据假设 (2) 趋于零，而第一项根据单调收敛定理收敛到 $\mathbb{E}_R[g(Z)]$ 。若将假设 (2) 替换为 $\limsup_n \sup_P |\mathbb{E}_P[g(T_n)\wedge K]-\mathbb{E}_R[g(Z)]| = 0$ 对每个 $K > 0$成立，则可采用类似的证明。根据三角不等式，写作

$$\begin{aligned}|\mathbb{E}_P[g(T_n) - \mathbb{E}_R[g(Z)]] \leq |\mathbb{E}_P[g(T_n) \wedge K] - \mathbb{E}_R[g(Z) \wedge K]| \\ + \mathbb{E}_P[g(T_n)\mathbf{1}\{g(T_n) \geq K\}] + \mathbb{E}_R[g(Z)\mathbf{1}\{g(Z) \geq K\}].\end{aligned}$$

因此，

$$\begin{aligned}&\limsup_n \sup_P |\mathbb{E}_P[g(T_n) - \mathbb{E}_R[g(Z)]]\\ &\le \limsup_n \sup_P \mathbb{E}_P[g(T_n)\mathbf{1}\{g(T_n) \ge K\}] + \mathbb{E}_R[g(Z)\mathbf{1}\{g(Z) \ge K\}].\end{aligned}$$

令 $K \to \infty$，不等式右侧的两项均趋于零——第一项由假设得出，第二项则由 $g(Z)$的可积性得出。至此，论证完毕。 □

## B.6 混合过程的广义充分条件

这里我们对命题 3.3作了一个温和的推广，放宽了混合过程成为渐近 e 变量所需的充分条件。陈述结果后，我们将进一步讨论这一推广的适用范围。

**命题 B.10**。令 $\Lambda \subset \mathbb{R}$，并令 $\pi$ 为 $\Lambda$上的任意（与数据无关的）概率分布。假设对于每个 $\lambda \in \Lambda$，序列 $(E_n(\lambda))$ 对于某个分布集合 $\mathcal{P}$而言是渐近 *e* 变量。假设对于每个 $P \in \mathcal{P}$，存在一个可测函数$h_P\colon \Lambda \to [0,\infty)$，使得：

- 上确界$_{n\ge1}\, \mathbb{E}_P[E_n(\lambda)] \le h_P(\lambda)$ 对于 $\pi$-几乎所有 $\lambda$ 成立（支配），
- $\int_\Lambda h_P(\lambda)\mathrm{d}\pi(\lambda) < \infty$ （可积性）。

于是，由下式定义的混合过程

$$E_n(\pi) := \int_\Lambda E_n(\lambda)\mathrm{d}\pi(\lambda), \tag{B.11}$$

是关于 $\mathcal{P}$ 的渐近 *e* 变量。此外，如果 $\sup_P h_P$ 可测且可积，即 $\int_\Lambda h_P(\lambda)\mathrm{d}\pi(\lambda) < \infty$，那么混合 $(E_n(\pi))$ 是一个 $\mathcal{P}$-一致渐近 *e* 变量。

证明。固定任意一个分布 $P \in \mathcal{P}$。根据托内利定理（由于 $E_n(\lambda)$ 非负），我们可以交换期望与混合积分：

$$\mathbb{E}_P[E_n(\pi)] = \mathbb{E}_P\left[\int_\Lambda E_n(\lambda)\mathrm{d}\pi(\lambda)\right] = \int_\Lambda \mathbb{E}_P[E_n(\lambda)]\mathrm{d}\pi(\lambda).$$

现在，我们可以将反向法图引理应用于函数序列 $\lambda \mapsto \mathbb{E}_P[E_n(\lambda)]$。该引理指出，如果这样的序列由一个可积函数控制，那么混合的上极限小于或等于其上极限的混合：

$$\limsup_{n\to\infty} \int_\Lambda \mathbb{E}_P[E_n(\lambda)]\mathrm{d}\pi(\lambda) \le \int_\Lambda \limsup_{n\to\infty} \mathbb{E}_P[E_n(\lambda)]\mathrm{d}\pi(\lambda).$$

由于 $(E_n(\lambda))$ 对每个 $\lambda$ 都是渐近 e 变量，因此对于所有 $\lambda$，均有 $\limsup_{n\to\infty} \mathbb{E}_P[E_n(\lambda)] \le 1$。因此：

$$\limsup_{n\to\infty} \int_\Lambda \mathbb{E}_P[E_n(\lambda)]\mathrm{d}\pi(\lambda) \le \int_\Lambda 1 \cdot \mathrm{d}\pi(\lambda) = 1.$$

由于 $P \in \mathcal{P}$ 是任意的，该结果对整个类 $\mathcal{P}$ 均成立。

在分布一致情形下，注意到以下事实后，证明基本不变：

$$\sup_P \mathbb{E}_P[E_n(\pi)] = \sup_P \int_\Lambda \mathbb{E}_P[E_n(\lambda)]\mathrm{d}\pi(\lambda) \le \int_\Lambda \sup_P \mathbb{E}_P[E_n(\lambda)]\mathrm{d}\pi(\lambda).$$

由于根据假设， $g(\lambda) = \sup_P h(\lambda)$是可测的，我们可以对 $\lambda \mapsto \sup_P \mathbb{E}_P[E_n(\lambda)] \le g(\lambda)$应用反向法图引理，得到

$$\limsup_n \int_\Lambda \sup_P \mathbb{E}_P[E_n(\lambda)]\mathrm{d}\pi(\lambda) \le \int_\Lambda \limsup_n \sup_P \mathbb{E}_P[E_n(\lambda)]\mathrm{d}\pi(\lambda) \le 1,$$

根据a $\mathcal{P}$-一致渐近 e 变量的定义。证毕。 □

命题3.3 中的支配和可积性这两个条件，严格弱于要求$(\sup_{\lambda\in\Lambda} E_n(\lambda))$一致可积（这是命题3.3中的要求）。事实上，考虑 $X \sim N(0,\ 1)$， $\Lambda = \mathbb{R}$， $\pi = N(0,\ 1)$，并令

$$E(\lambda) = \exp\left(\lambda X - \frac{\lambda^2}{2}\right).$$

那么 $\sup_\lambda E(\lambda) = \exp\left(X^2/2\right)$ 且 $\mathbb{E}\left[\exp\left(\frac{X^2}{2}\right)\right] = \infty$，因此它并非一致可积。然而， $\mathbb{E}[E(\lambda)] = 1$ ，所以我们只需在命题3.3 中选择 $h(\lambda) = 1$ ，即可得出$E_n(\pi)$是一个 e 变量。

## B.7 关于截断高斯混合的更多细节

首先，我们介绍通过混合 IWR e变量而得到的渐近 e 变量，依据的是推论 3.5。我们对附录 REG B.1 中介绍的 e 变量也采用同样的做法。

命题 *B.11.* 对任意固定的 $R, \kappa, \eta > 0$，定义 $Z_{R,\kappa} := \kappa\sqrt{2\pi}(\Phi(R/\kappa) - \Phi(-R/\kappa))$。那么，对于 $\mathcal{M}_2(\theta)$ 中的所有分布，$(E^{\mathrm{MIX}}{}_{,n}{}^{\mathrm{IWR}}(\theta; R, \kappa))$ 和 $(E^{\mathrm{MIX}}{}_{,n}{}^{\mathrm{REG}}(\theta; R, \kappa, \eta))$ 都是渐近 e 变量，其中

$$E_n^{\mathrm{MIX,IWR}}(\theta; R, \kappa) := Z_{R,\kappa}^{-1} I_R\left(\frac{S_n(\theta)}{V_n(\theta)}, \frac{\kappa^2+1}{2\kappa^2}\right), \tag{B.12}$$

*and*

$$E_n^{\mathrm{MIX,REG}}(\theta; R, \kappa, \eta) := Z_{R,\kappa}^{-1} I_R\left(\frac{S_n(\theta)}{\sqrt{n}\widehat{\sigma}_n + \eta V_n(\theta)}, \frac{\kappa^2+1}{2\kappa^2}\right). \tag{B.13}$$

证明。 $E_n^{\mathrm{IWR}}$和 $E_n^{\mathrm{REG}}$都可以写成 $E_n(\lambda) = \exp(\lambda M - \lambda^2/2)$，其中 $M$取某个值。令 $\rho$为 $[-R, R]$ 上均值为 0、方差为 $\kappa^2$ 的截断高斯分布的密度，即，

$$\rho(\xi) = Z_{R,\kappa}^{-1} \exp\left(\frac{-\xi^2}{2\kappa^2}\right) \quad \text{where} \quad Z_{R,\kappa} := \sqrt{2\pi\kappa^2}\left(\Phi\left(\frac{R}{\kappa}\right) - \Phi\left(\frac{-R}{\kappa}\right)\right), \tag{B.14}$$

其中 $\Phi$ 是标准正态分布的累积分布函数。令 $u = 1/2 + 1/(2\kappa^2)$，注意到

$$E_n(\rho) = \int_{-R}^{R} E_n(\lambda)\rho(\lambda)\mathrm{d}\lambda = Z_{R,\kappa}^{-1} \int_{-R}^{R} \exp\left(\lambda M - \lambda^2 u\right)\mathrm{d}\lambda.$$

配方法，写作

$$\exp(\lambda M-\lambda^2 u)=\exp\left(-u\left(\lambda-\frac{M}{2u}\right)^2\right)\exp\left(\frac{M^2}{4u}\right),$$

并令 $t=\sqrt{2u}(\lambda-M/(2u))$，使得 $\mathrm{d}t=\sqrt{2u}\mathrm{d}\lambda$。设 $t_2=\sqrt{2u}(R-M/(2u))$，且$t_1=-\sqrt{2u}(R+M/(2u))$。于是，

$$\begin{aligned}\int_{-R}^{R}\exp(\lambda M-\lambda^2 u)\mathrm{d}\lambda&=\frac{1}{\sqrt{2u}}\exp\left(\frac{M^2}{4u}\right)\int_{t_1}^{t_2}\exp(-t^2/2)\mathrm{d}t\\&=\frac{1}{\sqrt{2u}}\exp\left(\frac{M^2}{4u}\right)\left(\int_0^{t_2}\exp(-t^2/2)\mathrm{d}t-\int_0^{t_1}\exp(-t^2/2)\mathrm{d}t\right)\\&=\sqrt{\frac{\pi}{u}}\frac{1}{\sqrt{2u}}\exp\left(\frac{M^2}{4u}\right)(\Phi(t_2)-\Phi(t_1))\\&=I_R(M,u),\end{aligned}$$

利用 $\Phi(z)=\frac{1}{\sqrt{2\pi}}\int_0^z\exp(-t^2/2)\mathrm{d}t$。因此，$E_n(\rho)=Z_R^{-1},\kappa I_R(M,u)$。将$M=S_n(\theta)/V_n(\theta)$代入 $E_n^{\mathrm{IWR}}$，并将 $\overline{M}=S_n(\theta)/(\sqrt{n}\widehat{\sigma}_n+\eta V_n(\theta))$代入 $E_n^{\mathrm{REG}}$，即可完成证明。

□

对上述渐近 e 变量（$E^{\mathrm{MIX},}{}_{n}{}^{\mathrm{IWR}}$）求逆，得到定理 3.6。下面给出关于 ($E^{\mathrm{MIX},}{}_{n}{}^{\mathrm{REG}}$) 的等价结果。

**定理*B.12***。任取 $R,\kappa,\eta>0$，并令 $Z_{R,\kappa}=\kappa\sqrt{2\pi}(\Phi(R/\kappa)-\overline{\Phi}(-R/\kappa))$。$(\mathcal{H}^{\mathrm{MIX}},{}_{n}{}^{\mathrm{REG}}(R,\kappa,\eta))$ 是一个 APH-CI，其中 $\mathcal{H}^{\mathrm{MIX}},{}_{n}{}^{\mathrm{REG}}(R,\kappa,\eta)=(\overline{X_n\pm W^{\mathrm{MIX}},}{}_{n}{}^{\mathrm{REG}}(R,\kappa,\eta))$，且

$$W_n^{\mathrm{MIX,REG}}(R,\kappa,\eta)=\frac{\sqrt{n}\cdot s_n y_\alpha^\star(1+\eta D_{\alpha,\eta})}{n-(\eta y_\alpha^\star)^2},\tag{B.15}$$

其中 $D_{\alpha,\eta}=\sqrt{1+(y_\alpha^\star)^2(1-\eta^2)/n}$且 $y_\alpha^\star$ 满足 $I_R(y_\alpha^\star,(\kappa^2+1)/(2\kappa^2))=Z_{R,\kappa}/\alpha$。若 (B.15) 中的分母非正，则我们将置信区间取为整个 $\mathbb{R}$。

证明。现在我们转向 $E^{\mathrm{MIX},}{}_{n}{}^{\mathrm{REG}}$。我们使用一个 $E_n^{\mathrm{REG}}$以 $\hat{s}_n$ 替代分母中的 $\widehat{\sigma}_n$ 的版本，根据定理 B.3，该版本是成立的。令 $w_n=\sqrt{n}s_n$。方程 $|\overline{S_n}(\theta)/V_n(\theta)|<y^\star$ 于是变为

$$n|\Delta|-y^\star w_n<y^\star\eta\sqrt{w_n^2+n\Delta^2}.$$

若 $n|\Delta|<y^\star w_n$，则 $\theta\in\overline{(X_n\pm y^\star w_n/n)}$，因此它是我们解的一部分。若 $n\Delta\geq y^\star w_n$，则上面式子中的两项均非负，因此我们可以对两边平方，最终得到不等式

$$(n^2-(y^\star\eta)^2n)\Delta^2-2ny^\star w_n\Delta+(y^\star w_n)^2-(y^\star\eta)^2w_n^2<0.$$

经过大量算术运算，这个多项式的根为

$$\Delta_\pm=\frac{y^\star w_n\pm y^\star w_n\eta\sqrt{1+(y^\star)^2(1-\eta^2)/n}}{n-(y^\star\eta)^2}=\frac{y^\star w_n(1\pm\eta D)}{n-(y^\star\eta)^2},$$

其中 $D = D_{\alpha,\ \eta}$。保守起见，取正根，因为它定义的区间更大，我们得到

$$\theta \in \left( \overline{X}_n \pm \frac{y^\star w_n (1 \pm \eta D)}{n - (y^\star \eta)^2} \right),$$

它是 $(X_n \pm y^\star w_n / n)$ 的超集。 □

图 $\mathcal{H}^{\text{MIX}}{}_{,\ n}{}^{\text{REG}}$) 与 $(\mathcal{H}^{\text{MIX}}{}_{,\ n}{}^{\text{IWR}})$ 的性能对比见附录 D 中的图 6，此外还研究了 $\eta$ 的影响。

接下来，我们分析 $y^\star_\alpha$ 的极限行为，以便将 $W^{\text{MIX}}{}_{,\ n}{}^{\text{IWR}}$ 和 $W^{\text{MIX}}{}_{,\ n}{}^{\text{REG}}$ 的宽度与 Wald区间的宽度联系起来。

**引理 B.13**。对于 $\kappa = 1$，有 $y^\star_\alpha \xrightarrow{R\to\infty} 2\sqrt{\log(\sqrt{2}/\alpha)}$，~~其中 $y^\star$~~ 满足 $I_R(y^\star,\ 1) = Z_{R,\ 1}/\alpha$ 且 $Z_{R,\ \kappa} = \kappa\sqrt{2\pi}(\Phi(R/\kappa) - \Phi(-R/\kappa))$。

证明。令 $g_1(y) = \sqrt{2}R - y/\sqrt{2}$ 且 $g_2(y) = -\sqrt{2}R - y/\sqrt{2}$。那么

$$I_R(y, 1) = \sqrt{\pi} \exp\left(\frac{y^2}{4}\right) (\Phi(g_1(y)) - \Phi(g_2(y))).$$

令 $I_R(y,1) = Z_{R,1}/\alpha$，取对数后得到

$$\begin{aligned} y^\star_\alpha &= 2\sqrt{\log(Z_{R,1}/\alpha) - \log(\sqrt{\pi}) - \log(\Phi(g_1(y^\star_\alpha) - \Phi(g_2(y^\star_\alpha))} \\ &= 2\sqrt{\log(\sqrt{2}/\alpha) + \log(\Phi(R) - \Phi(-R)) - \log(\Phi(g_1(y^\star_\alpha) - \Phi(g_2(y^\star_\alpha))} \end{aligned}$$

现在，由于 $R \to \infty$, $\Phi(R) \to 1$ 且 $\Phi(-R) \to 0$，所以 $\log(\Phi(R) - \Phi(-R)) \to 0$。此外，$g_1(y^\star_\alpha) \to 1$，因为 $R \to \infty$ 且 $g_2(y^\star_\alpha) \to 0$。要看出这一点，请注意，随着 $R$ 增大，$y^\star_\alpha$ 要么趋于 $\infty$，要么趋于 0，要么保持有界（它不可能为负）。在每种情况下，我们在 $g_1$ 和 $g_2$ 上都能得到所需的行为。因此，$-\log(\Phi(g_1(y^\star_\alpha)) - \Phi(g_2(y^\star_\alpha))) \to 0$，由此得出

$$y^\star_\alpha \to 2\sqrt{\log(\sqrt{2}/\alpha)},$$

论证至此完成。 □

# C 省略的证明

**引理C.1**。设 $X_1, \ldots, X_n$ 相互独立同分布，均值为 $\theta$，且属于某个正态律吸引域。设 $S_n(\theta)$ 和 $V_n(\theta)$ 如 (3.2)所定义。任取 $\lambda_{\max} \geq 0$。则：

*(i)* 随机变量族 $(\sup^{\lambda \in [-\lambda_{\max}, \lambda_{\max}]} \exp(\lambda S_n(\theta)/V_n(\theta))$ 是一致可积的，并且

*(ii)* 对于一个固定的 $\eta > 0$，随机变量族 $(\sup_{\lambda \in [-\lambda_{\max}, \lambda_{\max}]} E_n^{\text{REG}}(\theta; \lambda, \eta)$，其中

$$E_n^{\text{REG}}(\theta; \lambda, \eta) = \exp\left( \lambda \frac{\sqrt{n}(\overline{X}_n - \theta)}{\widehat{\sigma}_n + \eta V_n(\theta)/\sqrt{n}} - \frac{\lambda^2}{2} \right),$$

是一致可积的。

*(iii)* 令 $\lambda_n \to \lambda$ 和 $\eta_n \to \eta$ 几乎必然，其中 $\eta_n > 0$ 和 $\eta \geq 0$。如果 $|\lambda_n| \lesssim \eta_n$，那么随机变量族 $(E_n^{\mathrm{REG}}(\theta; \lambda_n, \eta_n))$ 是一致可积的。

证明。 先考虑 (i)。由于 $X_i - \theta$ 是均值为零、独立同分布且属于高斯分布吸引域的随机变量，我们可以应用 [21] 中的定理 2.5，得知存在常数 $C > 0$ 使得：

$$\sup_n \mathbb{E}\left[\exp\left(t\left|\frac{S_n(\theta)}{V_n(\theta)}\right|\right)\right] \leq 2\exp\left(Ct^2\right) \quad \text{for all } t \geq 0.$$

事实上，

$$C = \left(1 + \frac{4e}{3}\right)^2 \max\left\{1, \sup_n \mathbb{E}\left|\frac{S_n(\theta)}{V_n(\theta)}\right|\right\}^2. \tag{C.1}$$

我们可以得出结论：$C < \infty$，因为 $S_n(\theta)/V_n(\theta) \to N$依分布收敛于 (0，1)（根据斯卢茨基定理），因此 $S_n(\theta)/V_n(\theta)$ 随机有界。因此，根据 Gin´e 等人 [21], 中的引理 2.4，$_n\,\mathbb{E}|S_n(\theta)/V_n(\theta)| < \infty$ 的上确界有界，这意味着 $C < \infty$。注意，$C$ 与 $\lambda$ 独立，但与 $X_i$ 的分布并不独立。这表明，对于每个 $n$ 和固定的 $\lambda$，$\exp(\lambda S_n(\theta)/V_n(\theta))$ 都是可积的。此外，

$$\sup_n \mathbb{E}\sup_\lambda \exp\left\{\lambda\frac{S_n(\theta)}{V_n(\theta)}\right\}^2 \leq \sup_n \mathbb{E}\exp\left(2\lambda_{\max}\left|\frac{S_n(\theta)}{V_n(\theta)}\right|\right) \leq 2\exp\left(4C\lambda_{\max}^2\right) < \infty,$$

因此，应用 de la Vall´ée Poussin 定理 [13, 38]，并取定义在 $\mathbb{R}_+$ 上的非负递增凸函数 $t \mapsto t^2$，我们可以得出结论：$\{\sup_\lambda \exp(\lambda S_n(\theta)/V_n(\theta)\}_n$ 一致可积。这就证明了 (i)。

对于 (ii)，注意我们可以写成

$$\begin{aligned}
\frac{\sqrt{n}(\overline{X}_n - \theta)}{\widehat{\sigma}_n + \eta V_n(\theta)/\sqrt{n}} &= \frac{\sqrt{n}(\frac{1}{n}\sum_{i\leq n} X_i^\theta)}{\sqrt{\sum_{i\leq n}(X_i - \overline{X}_n)^2/(n-1)} + \eta V_n(\theta)/\sqrt{n}} \\
&= \frac{\sqrt{n}(\frac{1}{n}\sum_{i\leq n} X_i^\theta)}{\sqrt{\sum_{i\leq n}(X_i^\theta - \frac{1}{n}\sum_{i\leq n} X_i^\theta)^2/(n-1)} + \eta V_n(\theta)/\sqrt{n}} \\
&= \frac{S_n(\theta)/V_n(\theta)}{\sqrt{(n - (S_n(\theta)/V_n(\theta))^2)/(n-1)} + \eta}.
\end{aligned}$$

那么：

$$\begin{aligned}
E_n^{\mathrm{REG}}(\theta; \lambda, \eta) &= \exp\left(\lambda\frac{\sqrt{n}(\overline{X}_n - \theta)}{\widehat{\sigma}_n + \eta V_n(\theta)/\sqrt{n}} - \frac{\lambda^2}{2}\right) \\
&= \exp\left(\lambda\frac{\sqrt{n}(\overline{X}_n - \theta)}{\widehat{\sigma}_n + \eta V_n(\theta)/\sqrt{n}}\right)\exp\left(-\frac{\lambda^2}{2}\right) \\
&= \exp\left(\lambda\frac{S_n(\theta)/V_n(\theta)}{\sqrt{(n - (S_n(\theta)/V_n(\theta))^2)/(n-1)} + \eta}\right)\exp\left(-\frac{\lambda^2}{2}\right) \\
&\leq \exp\left(\frac{|\lambda|}{\eta}\left|\frac{S_n(\theta)}{V_n(\theta)}\right|\right)\exp\left(-\frac{\lambda^2}{2}\right).
\end{aligned}$$

因此，

$$\sup_{\lambda} E_n^{\mathrm{REG}}(\theta;\lambda,\eta) \le \exp\left(\frac{\lambda_{\max}}{\eta}\left|\frac{S_n(\theta)}{V_n(\theta)}\right|\right).$$

现在我们可以借助 $\lambda_{\max} \leftarrow \lambda_{\max}/\eta$ 诉诸第(i)部分，以完成(ii)的证明。对于(iii)，令 $\infty > M \ge \sup_n |\lambda_n|/\eta_n$，（$M$ 保证为有限，因为 $|\lambda_n|/\eta_n \lesssim 1$）。然后，按上述方式进行，

$$\begin{aligned}\sup_n \mathbb{E}_P[E_n^{\mathrm{REG}}(\theta;\lambda_n,\eta_n)] &\le \sup_n \mathbb{E}_P\left[\exp\left(\frac{|\lambda_n|}{\eta_n}\left|\frac{S_n(\theta}{V_n(\theta)}\right|\right)\right]\exp\left(-\frac{\lambda_n^2}{2}\right)\\ &\le \sup_n \mathbb{E}_P\left[\exp\left(M\left|\frac{S_n(\theta}{V_n(\theta)}\right|\right)\right]\\ &\le 2\exp(CM^2).\end{aligned}$$

然后我们可以按上述方式应用 de la Vall´ee Poussin 定理，从而证明(iii)。 □

## C.1 命题的证明2.6

首先，假设 $\mathcal{H}_n(\alpha) = \{\theta : E_n(\theta) < 1/\alpha\}$，并且$(E_n(\theta^\star))$是 $\theta^\star$的渐近 e 变量。注意，对任意 $\alpha > 0$，通过对示性函数作情形分析，有 $\mathbf{1}\{E_n(\theta^\star) \ge 1/\alpha\}/\alpha \le E_n(\theta^\star)$。此外，取 $\alpha = 1/E_n(\theta^\star)$，我们看到 $\mathbf{1}\{E_n(\theta^\star) \ge 1/\alpha\}/\alpha \ge E_n(\theta^\star)$。因此，

$$\sup_{\alpha>0}\frac{\mathbf{1}\{\theta^\star \notin \mathcal{H}_n(\alpha)\}}{\alpha} = \sup_{\alpha>0}\frac{\mathbf{1}\{E_n(\theta^\star) \ge 1/\alpha\}}{\alpha} = E_n(\theta^\star). \tag{C.2}$$

因此，取期望值可知，$(E_n(\theta^\star))$ 是一个（一致）渐近 e 变量，这意味着 $(\mathcal{H}_n)$ 是一个（一致）APH-CI。

反过来，设 $\mathcal{H}_n(\cdot)$ 是 $\theta^\star$ 的一个 APH-CI 。对于任意 $\theta \in \Theta$，定义

$$E_n(\theta) := \sup\left\{\frac{1}{\alpha} : \alpha > 0 \text{ and } \theta \notin \mathcal{H}_n(\alpha)\right\} = \sup_{\alpha>0}\frac{\mathbb{1}\{\theta \notin \mathcal{H}_n(\alpha)\}}{\alpha}. \tag{C.3}$$

根据 $(\mathcal{H}_n)$ 构成 $\theta^\star$ 的一个 APH-CI 这一性质，可知 $E_n(\theta^\star)$ 是一个渐近 e 变量。在分布一致设定下也同样成立。

现在我们将证明，对于任意 $\alpha \in (0,1)$，$\mathcal{H}_n(\alpha) = \{\theta : E_n(\theta) < 1/\alpha\}$。事实上，任取 $\theta_0 \in \mathbb{R}$ 和 $\alpha_0 \in (0,1)$。由于 $\mathcal{H}_n(\cdot)$ 从下闭合，$E_n(\theta_0)$ 定义中的上确界可以取到，即

$$E_n(\theta_0) = \frac{1}{\alpha(\theta_0)} \quad \text{where} \quad \alpha(\theta_0) = \min\{\alpha > 0 : \mathbb{1}\{\theta_0 \notin \mathcal{H}_n(\alpha)\}\}, \tag{C.4}$$

若集合为空，则将上述最小值取为 $\infty$。

**情形 I：** $\alpha_0 = \alpha(\theta_0)$。显然，$E_n(\theta_0) \ge 1/\alpha_0$ 且 $\theta_0 \notin \mathcal{H}_n(\alpha_0)$。

**情况 II：** $\alpha_0 < \alpha(\theta_0)$。根据 $\alpha(\theta_0)$是使得 $\theta_0$ 不属于 $\mathcal{H}_n(\alpha)$ 的最小 $\alpha$ 的定义，我们有 $\theta_0 \in \mathcal{H}_n(\alpha_0)$。此外，根据定义，我们有 $E_n(\theta_0) = 1/\alpha(\theta_0) < 1/\alpha_0$。

**情况 III：** $\alpha_0 > \alpha(\theta_0)$。根据 $\mathcal{H}_n(\cdot)$ 的单调性，有 $\theta_0 \notin \mathcal{H}_n(\alpha_0)$。再次根据 $E_n(\theta_0)$ 的定义，我们有 $E_n(\theta_0) = 1/\alpha(\theta_0) > 1/\alpha_0$。

在这三种情形中，$E_n(\theta_0) \geq 1/\alpha_0$ 当且仅当 $\theta_0 \notin \mathcal{H}_n(\alpha_0)$。由于 $\theta_0$ 和$\alpha_0 \in (0, 1)$ 是任意的，证明至此完成。

## C.2 定理 3.1 的证明

我们分别处理逐点情形和分布一致情形。后者比前者复杂得多。

### C.2.1 逐点情形。

设 $(\lambda_n)$ 是一个有界序列，且几乎必然收敛到 $\lambda \in \mathbb{R}$。由引理 C.1 (i)，可知 $(E_n^{\mathrm{IWR}}(\theta); \lambda_n))_n$ 一致可积。此外，

$$\frac{S_n(\theta)}{V_n(\theta)} = \frac{\frac{1}{\sqrt{n}} \sum_{i \leq n} (X_i - \theta)}{\frac{1}{\sqrt{n}} \sqrt{\sum_i (X_i - \theta)^2}} \xrightarrow{d} N(0, 1),$$

根据 Gin´e 等人 [21, 的定理 3.3]。因此，由连续映射定理和再次应用斯卢茨基定理可得，$E_n^{\mathrm{IWR}}(\theta; \lambda_n) \xrightarrow{d} \exp(\lambda Z - \lambda^2/2)$，其中 $Z \sim N(0, 1)$。由于依分布收敛与一致可积性共同蕴含依期望收敛 [5, 定理 25.12], ，我们有 $\lim_{n\to\infty} \mathbb{E}[E_n^{\mathrm{IWR}}(\theta; \lambda_n)] = \mathbb{E}[\exp(\lambda Z - \lambda^2/2)] = 1$，论证至此完成。

### C.2.2 分布一致情形。

定义非负函数 $g(T; \lambda) = \exp(\lambda T - \lambda^2/2)$。令 $Z \sim N(0, 1) = R$为标准正态随机变量，因此 $\mathbb{E}_R[g(Z)] = 1$。对于固定的 $\lambda$ 以及 $T_n = S_n(\theta)/V_n(\theta)$，我们将证明 $g(T_n; \lambda)$ 满足命题 B.9 的条件。

**第一步。** 我们声称 $g(T_n; \lambda)$ 满足

$$\lim_{K\to\infty} \limsup_n \sup_P \mathbb{E}_P[g(T_n; \lambda) \mathbf{1}\{g(T_n; \lambda) \geq K\}. \tag{C.5}$$

注意到 $g(T_n; \lambda)\mathbf{1}\{g(T_n; \lambda) \geq K\} \leq \exp(\lambda T_n)\mathbf{1}\{\exp(\lambda T_n) \geq K\}$（因为 $\exp(-\lambda^2/2) \leq 1$）。此外，在 $\{\exp(\lambda T_n) \geq K\}$ 上，对于任意 $U > \lambda$，我们有 $\exp(T_n(U - \lambda)) \geq K^{(U-\lambda)/\lambda}$，即

$$1 \leq \exp(T_n(U - \lambda))K^{-(U-\lambda)/\lambda}.$$

将两边同时乘以 $\exp(\lambda T_n)\mathbf{1}\{\exp(\lambda T_n) \geq K\}$，得到

$$g(T_n; \lambda)\mathbf{1}\{g(T_n; \lambda) \geq K\} \leq \exp(\lambda T_n)\mathbf{1}\{\exp(\lambda T_n) \geq K\} \leq \frac{\exp(UT_n)}{K^{(U-\lambda)/\lambda}}. \tag{C.6}$$

因此，为了证明 (C.5) 成立，只需证明

$$\limsup_n \sup_P \mathbb{E}_P[\exp(UT_n)] < \infty. \tag{C.7}$$

与逐点情形一样，我们可以应用 [21] 中的定理 2.5，得出存在一个常数 $C > 0$ 使得：

$$\mathbb{E}_P[\exp(UT_n)] \leq 2\exp\left(C_P U^2\right),$$

其中

$$C_P = \left(1 + \frac{4e}{3}\right)^2 \max\left\{1, \max_{\ell \leq n} \mathbb{E}_P|T_\ell|\right\}^2. \tag{C.8}$$

因此，我们需要对所有 $P$ 一致地界定 $\mathbb{E}_P[|T_\ell|]$。为此，我们采用自归一化和的 Berry-Esseen 界（参见 Bentkus 和 G¨otze [3] 或 De la Pena 等人 [12,第 5 章，定理 5.9]），该界指出

$$\sup_{z \in \mathbb{R}} |P(T_n \leq z) - \Phi(z)| \leq \frac{25\mathbb{E}_P|X_i - \theta|^3}{\sqrt{n}\sigma_P^3} \leq \frac{25}{\sqrt{n}} M, \tag{C.9}$$

根据一致有界偏度的假设，存在有限的 $M$，且其中 $\Phi$ 是标准正态分布的累积分布函数。由此，注意到 $|T_n| \leq \sqrt{n}$，根据柯西-施瓦茨不等式。因此

$$\begin{aligned}
\mathbb{E}_P|T_n| &= \int_0^{\sqrt{n}} P(|T_n| > r)\mathrm{d}r = \int_0^{\sqrt{n}} P(T_n > r) + P(T_n < -r)\mathrm{d}r \\
&= \int_0^{\sqrt{n}} P(T_n > r)\mathrm{d}r + \int_{-\sqrt{n}}^0 P(T_n < r)\mathrm{d}r \\
&\leq \int_0^{\sqrt{n}} P(Z > r)\mathrm{d}r + \int_{-\sqrt{n}}^0 P(Z < r)\mathrm{d}r + 50M \\
&\leq \int_0^{\infty} P(|Z| \geq r)\mathrm{d}r + 50M < 1 + 50M < \infty.
\end{aligned}$$

因此，我们可以用与分布无关的常数 $C$ 对依赖于分布的常数 $C_P$ 给出上界：

$$C_P \leq \left(1 + \frac{4e}{3}\right)^2 (1 + 50M)^2 =: C,$$

在这种情况下

$$\sup_P \mathbb{E}_P[\exp(UT_n)] \leq 2\exp(CU^2).$$

回到（C.6），我们有

$$\begin{aligned}
\limsup_n \sup_P \mathbb{E}_P[g(T_n;\lambda)\mathbf{1}\{g(T_n;\lambda) \geq K\}] &\leq \limsup_n \sup_P \mathbb{E}_P\left[\frac{\exp(UT_n)}{K^{(U-\lambda)/\lambda}}\right] \\
&\leq \frac{2\exp(CU^2)}{K^{(U-\lambda)/\lambda}} \xrightarrow{K\to\infty} 0,
\end{aligned}$$

根据需要。

**第二步。** 接下来，我们要证明，对于任意 $K > 0$,

$$\limsup_n \sup_P |\mathbb{E}_P[g(T_n) \wedge K] - \mathbb{E}_R[g(Z) \wedge K]| = 0, \tag{C.10}$$

其中 $Z \sim N(0,\ 1) = R$。令 $h(x) = g(x) \wedge K$。利用黎曼–斯蒂尔杰斯积分的分部积分法，写作

$$\mathbb{E}_P[h(T_n)] = \int_{-\infty}^{\infty} h(x)\mathrm{d}P(x) = h(x)P(x)\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} P(x)\mathrm{d}h(x) = K - \int_{-\infty}^{\infty} P(x)\mathrm{d}h(x).$$

对 $\mathbb{E}_R[h(Z)]$, 采用类似的推理，我们有

$$\begin{aligned}
|\mathbb{E}_P[h(T_n)] - \mathbb{E}_R[h(Z)]| &= \left|\int_{-\infty}^{\infty} R(x)\mathrm{d}h(x) - \int_{-\infty}^{\infty} P(x)\mathrm{d}h(x)\right| \\
&= \left|\int_{-\infty}^{\infty} R(x) - P(x)\mathrm{d}h(x)\right| \\
&\le \sup_{z\in\mathbb{R}} |R(x) - P(x)| \int_{-\infty}^{\infty} |\mathrm{d}h(x)|,
\end{aligned}$$

其中，最后的积分 $\int |\mathrm{d}h(x)|$ 是 $h$ 的全变差。由于 $\lim_{x\downarrow-\infty} h(x) = 0$ 且 $\lim_{x\uparrow\infty} h(x) = K$, $h$ 的全变差为 $K$。因此，利用上述 Berry-Esseen 界，我们有

$$\sup_P |\mathbb{E}_P[h(T_n)] - \mathbb{E}_R[h(Z)]| \le \frac{25KM}{\sqrt{n}} \xrightarrow{n\to\infty} 0,$$

这就是所需结果。

**Step III.** 结合命题 B.9，前两个步骤已证明 t

$$\limsup_{n\to\infty} \sup_P |\mathbb{E}_P[g(T_n;\lambda) - 1| = 0. \tag{C.11}$$

还需要允许 $\lambda$ 随 $n$ 变化。根据三角不等式，写作

$$|\mathbb{E}_P[g(T_n;\lambda_n)] - 1| \le |\mathbb{E}_P[g(T_n;\lambda_n) - g(T_n;\lambda)]| + |\mathbb{E}_P[g(T_n;\lambda)] - \mathbb{E}_R[g(Z)]|. \tag{C.12}$$

由（C.6），我们已控制二阶差分。对于第一项，我们应用中值~定理于 $g$，从而得出存在某个 $\lambda$ 使得

$$\begin{aligned}
g(T_n;\lambda_n) - g(T_n;\lambda) &\le |\lambda_n - \lambda|\frac{\partial g(T_n;\widetilde{\lambda})}{\partial\lambda} = |\lambda_n - \lambda||T_n - \widetilde{\lambda}|g(T_n;\widetilde{\lambda}) \\
&\le |\lambda_n - \lambda||T_n + U|\exp(UT_n) \le U|\lambda_n - \lambda|\exp((U + 1/U)T_n),
\end{aligned}$$

其中我们用到了 $1 + x \le e^x$ 对 $x \ge 0$成立。因此，利用第一步的结果，我们有

$$\begin{aligned}
\sup_P |\mathbb{E}_P[g(T_n;\lambda_n) - g(T_n;\lambda)]| &\le U|\lambda_n - \lambda| \sup_P \mathbb{E}_P \exp((U + 1/U)T_n) \\
&\le 2U|\lambda_n - \lambda|\exp(C(U + 1/U)^2)
\end{aligned}$$

其在 $n \to \infty$ 时收敛到0，因为 $\lambda_n \to \lambda$。因此，结合（C.11）和（C.12），我们有

$$\limsup_{n\to\infty} \sup_P |\mathbb{E}_P[g(T_n;\lambda)] - 1| = 0,$$

由此完成对 $(E_n^{\mathrm{IWR}})$ 的证明。

## C.3 定理 3.2 的证明

在整个证明过程中，令 $A_n = A_{n,\lambda,\alpha}$。固定 $\lambda > 0$。我们将结合 B.8 号命题，取$K = 2$ 和 $E_n^{\mathrm{IWR}}(\theta;\ \lambda)$ 以及 $E_n^{\mathrm{IWR}}(\theta;\ -\lambda)$。也就是说，

$$\mathcal{H}_n^{\mathrm{IWR}}(\alpha) = \{\theta : E_n^{\mathrm{IWR}}(\theta;\lambda) < 2/\alpha \text{ and } E_n^{\mathrm{IWR}}(\theta;-\lambda) < 2/\alpha\}.$$

首先考虑满足 $\theta$ 使得 $E_n^{\mathrm{IWR}}(\theta;\lambda) < 2/\alpha$ 的情形，我们可以将其改写为

$$S_n(\theta) < \sqrt{n}V_n(\theta)A_n. \tag{C.13}$$

由于 $V_n(\theta)\ A_n \geq 0$，当且仅当 $S_n(\theta) < 0$ 时，该方程成立；也就是说，当且仅当 $\theta > X_n$。因此，方程 (C.13) 的解集的一部分为 $R_1 = (X_n, \infty)$。为了求出满足 (C.13) 的其余 $\theta$值，假设 $S_n(\theta) \geq 0$。然后，我们可以将两边平方，得到

$$S_n^2(\theta) < nV_n^2(\theta)A_n^2 \tag{C.14}$$

令 $y = X_n - \theta$ 并注意到 $S_n(\theta) = ny$ 且 $V_n(\theta) = \sqrt{\widehat{\sigma}_n^2(n-1) + ny^2}$。因此，(C.14) 可改写为 $n^2y^2 < n(\widehat{\sigma}_n^2(n-1) + ny^2)A_n^2$。若 $1 - A_n^2 > 0$，则有

$$y < \sqrt{\frac{(n-1)\widehat{\sigma}_n^2 A_n^2}{n(1-A_n^2)}} = \frac{\widehat{\sigma}_n A_n}{(1-A_n^2)^{1/2}}\sqrt{\frac{n-1}{n}} =: B_n. \tag{C.15}$$

这给出了区间 $R_2 = (X_n - B_n,\ \infty) \supset R_1$ 作为我们的解集。如果 $1 - A_n^2 \leq 0$，那么（C.14）的任意一个解为 $S_n(\theta) > 0$，即 $\theta < X_n$，在这种情况下，得到的置信区间为整个实数轴。为确保 $1 - A_n^2 > 0$，只需满足

$$n \geq \left(\frac{\log(2/\alpha)}{\lambda} + \frac{\lambda}{2}\right)^2,$$

这给出了定理陈述中对 $n$ 的约束。以类似方式求解 $E_n^{\mathrm{IWR}}(\theta;-\lambda) < 2/\alpha$ 中的 $\theta$，得到区间 $R_2' = (-\infty,\ X_n + B_n)$。取交集 $R_2 \cap R_2'$ 即得结果。

## C.4 命题 3.3 的证明

该结果可由附录 B.6 中陈述的更一般版本推出。不过，为清楚起见，我们直接给出证明。

令 $\epsilon > 0$ 任意。根据一致可积性的假设，存在某个 $T_\epsilon > 0$，使得

$$\sup_n \mathbb{E}_P\left[\sup_\lambda E_n(\lambda)\mathbf{1}\left\{\sup_\lambda E_n(\lambda) \geq T_\epsilon\right\}\right] \leq \epsilon,$$

其中 $\lambda$ 遍历所有 $\lambda \in \Lambda$。记

$$E_n(\pi) = \int E_n(\lambda)\mathbf{1}\left\{E_n(\lambda) < T_\epsilon\right\}\mathrm{d}\pi(\lambda) + \int E_n(\lambda)\mathbf{1}\left\{E_n(\lambda) \geq T_\epsilon\right\}\mathrm{d}\pi(\lambda).$$

根据托内利定理，

$$\begin{aligned}\mathbb{E}_P[E_n(\pi)] &= \int \mathbb{E}_P[E_n(\lambda)\mathbf{1}\{E_n(\lambda) < T_\epsilon\}]\mathrm{d}\pi(\lambda) + \int \mathbb{E}_P[E_n(\lambda)\mathbf{1}\{E_n(\lambda) \geq T_\epsilon\}]\mathrm{d}\pi(\lambda) \\ &\leq \int \mathbb{E}_P[E_n(\lambda)\mathbf{1}\{E_n(\lambda) < T_\epsilon\}]\mathrm{d}\pi(\lambda) + \epsilon.\end{aligned}$$

由于 $E_n(\lambda)\mathbf{1}\{E_n(\lambda) \geq T_\epsilon\}$ 有界，反向法图引理给出

$$\begin{aligned}\limsup_n \mathbb{E}_P[E_n(\lambda)] &\leq \int \limsup_n \mathbb{E}_P[E_n(\lambda)\mathbf{1}\{E_n(\lambda) < T_\epsilon\}]\mathrm{d}\pi(\lambda) + \epsilon \\ &\leq \int \limsup_n \mathbb{E}_P[E_n(\lambda)]\mathrm{d}\pi(\lambda) + \epsilon \leq 1 + \epsilon.\end{aligned}$$

由于 $\epsilon$ 是任意的，这在逐点情形下给出了所需结果。在分布一致情形下，论证类似。我们有如下保证

$$\sup_n \sup_P \mathbb{E}_P\left[\sup_\lambda E_n(\lambda)\mathbf{1}\left\{\sup_\lambda E_n(\lambda) \geq T_\epsilon\right\}\right] \leq \epsilon,$$

and，然后在应用托内利定理之后，我们对 $P$ 取上确界以得到 n

$$\sup_P \mathbb{E}_P[E_n(\pi)] = \int \sup_P \mathbb{E}_P[E_n(\lambda)\mathbf{1}\{E_n(\lambda) < T_\epsilon\}]\mathrm{d}\pi(\lambda) + \epsilon.$$

由此，论证几乎完全相同，只需作必要的改动。

## C.5 定理的证明 3.6

在开始之前，注意到我们可以写成

$$S_n(\theta) = n\Delta \quad \text{and} \quad V_n^2(\theta) = ns_n^2 + n\Delta^2,$$

其中 $\Delta = (\bar{X}_n - \theta)$，$s_n = n^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2 = (n-1)\hat{\sigma}_n^2$ 是有偏样本方差 .

我们先从 $E^{\text{MIX}}{}_{,n}$ IWR 开始。令 $u = (\kappa^2+1)/(2\kappa^2)$。由于 $I_R(y,u)$ 关于 $|y|$ 单调递增，我们要寻找满足 $\theta$ 使得 $|S_n(\theta)/V_n(\theta)| < y^\star$ 的值，其中 $I_R(y^\star,u) < Z_{R,\kappa}\alpha$。重写 $|S_n(\theta)/V_n(\theta)| < y^\star$ 可得

$$|n\Delta| < y^\star\sqrt{ns_n^2 + n\Delta^2},$$

对两边平方可得

$$|\Delta| < \frac{y^\star s_n}{\sqrt{n - (y^\star)^2}},$$

这正是由（3.11）定义的区间。

## 定理 3.8的证明

我们将使用 Ruf 和 Waudby-Smith 的以下引理 [47, 定理 2。3] （代入 $q = 1$）。

**引理 C.2** （非渐近强大数定律）。 令 $Z_1, Z_2, \ldots$ 为具有有限均值 $\theta$ 且分布为$P$的独立同分布随机变量。对于 $x \geq 0$，定义 $U(x) = \mathbb{E}_P[|Z_1 - \theta|\mathbf{1}\{|Z_1 - \theta| \geq x\}]$。则对于所有 $\epsilon \in$ （0、 6.13] 以及$\gamma \in$ （0、 1/2），

$$P\left[\sup_{k \geq m}\left|\frac{1}{k}\sum_{i=1}^{k} Z_i - \theta\right| \geq \epsilon\right] \leq c_\gamma \exp\{-m^{1-\gamma}\} + \frac{451}{\epsilon^2} U\left(\frac{\epsilon m^\gamma}{6.13}\right), \tag{C.16}$$

其中 $c_\gamma =$ $(3 - 4\gamma) / (2 - 4\gamma)$。

现在开始证明。回顾 $U_n^\otimes$ 在（3.13）中的定义，并注意到$\mathbb{E}[U_n^\otimes] = n\sigma^2$。进一步注意到

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \theta)^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 + (\overline{X}_n - \theta)^2,$$

对于所有 $n \geq 1$。因此

$$\hat{s}_n^2 - \frac{U_n^\otimes}{n} = \frac{2}{3}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \theta)^2 - \sigma^2\right) - (\overline{X}_n - \theta)^2.$$

取绝对值后，利用并集界可得

$$P\left(\left|\hat{s}_n^2 - \frac{U_n^\otimes}{n}\right| \geq \epsilon\right) \leq P\left(\frac{2}{3}\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - \theta)^2 - \sigma^2\right| \geq \frac{\epsilon}{2}\right) + P\left((\overline{X}_n - \theta)^2 \geq \frac{\epsilon}{2}\right).$$

我们可以将引理 C.2 应用于上述公式右侧的两项。由此可得

$$\begin{aligned}
&P\left(\frac{2}{3}\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - \theta)^2 - \sigma^2\right| \geq \frac{\epsilon}{2}\right) \\
&= P\left(\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - \theta)^2 - \sigma^2\right| \geq \frac{3\epsilon}{4}\right) \\
&\leq 2\exp(-n^{1-2\gamma}) + \frac{(4/3)^2 \cdot 451}{\epsilon^2}\mathbb{E}_P\left[|X_1 - \theta|^2\mathbf{1}\left\{|X_1 - \theta|^2 \geq \frac{3\epsilon n^\gamma}{4}\right\}\right] \\
&\leq 2\exp(-n^{1-2\gamma}) + \frac{A_\delta}{\epsilon^{2+\delta/2} n^{\delta\gamma/2}}\mathbb{E}_P[|X_1 - \theta|^{2+\delta}],
\end{aligned}$$

其中 $A_\delta = (4/3)^2 \cdot 451 \cdot (4 \cdot 38/3)^{\delta/2}$，最后一个不等式用到了：对于任意非负随机变量 $Z$（在 $\{Z \geq u\}$ 上），我们有 $Z^{2+\delta} = Z^2 Z^\delta \geq$， $\mathbb{E}[Z^2\mathbf{1}\{Z \geq u\}] \leq u^{-\delta}\mathbb{E}[Z^{2+\delta}]$

$Z^2u^\delta$）。请注意，这里我们用到了 $\delta > 0$。此外，

$$\begin{aligned}
&P\left((\overline{X}_n - \theta)^2 \geq \frac{\epsilon}{2}\right) \\
&= P\left(|\overline{X}_n - \theta| \geq \sqrt{\frac{\epsilon}{2}}\right) \\
&\leq 2\exp(-n^{1-2\gamma}) + \frac{902}{\epsilon}\mathbb{E}_P\left[|X_1 - \theta|\mathbf{1}\left\{|X_1 - \theta| \geq \frac{\sqrt{\epsilon/2}\cdot n^\gamma}{38}\right\}\right] \\
&\leq 2\exp(-n^{1-2\gamma}) + \frac{902}{\epsilon}\left(\frac{38\sqrt{2}}{\sqrt{\epsilon}n^\gamma}\right)^{1+\delta}\mathbb{E}_P\left[|X_1 - \theta|^{2+\delta}\right].
\end{aligned}$$

因此，结合上述两个式子，

$$P\left(\left|\hat{s}_n^2 - \frac{U_n^\otimes}{n}\right| \geq \epsilon\right) \leq 4\exp(-n^{1-2\gamma}) + \left(\frac{A_\delta}{n^{\delta\gamma/2}\epsilon^{2+\delta/2}} + \frac{B_\delta}{\epsilon^{(3+\delta)/2}n^{(1+\delta)\gamma}}\right)\mathbb{E}_P|X_1 - \theta|^{2+\delta},$$

对于常数 $A_\delta, B_\delta$。定义“好”事件，

$$A_n := \left\{\left|\hat{s}_n^2 - \frac{U_n^\otimes}{n}\right| \leq \epsilon_n\right\}.$$

在 $A_n$ 上，我们有 $n(\hat{s}_n^2 + \epsilon_n) \geq U_n^\otimes$，因此

$$\begin{aligned}
\mathbb{E}_P[E_n^{\text{R-WS}}(\theta; \epsilon_n, T_n)\mathbf{1}\{A_n\}] &\leq \mathbb{E}_P\left[\left(\int E_n^\star(\lambda)\mathrm{d}F \wedge T_n\right)\mathbf{1}\{A_n\}\right] \\
&\leq \mathbb{E}_P\left[\int E_n^\star(\lambda)\mathrm{d}F\right] = \int \mathbb{E}_P[E_n^\star(\lambda)]\mathrm{d}F \leq 1,
\end{aligned}$$

其中最后的等式由托内利定理得到，最后的不等式则来自$(E_n^\star(\lambda))$是初值为1的非负超鞅这一事实。因此，为简洁起见，令$E_n = E_n^{\text{R-WS}}(\theta; \epsilon_n, T_n)$，

$$\begin{aligned}
\mathbb{E}_P[E_n] &= \mathbb{E}_P[E_n\mathbf{1}\{A_n\}] + \mathbb{E}_P[E_n\mathbf{1}\{A_n^c\}] \leq 1 + T_nP(A_n^c) \\
&\leq 1 + \left(4\exp(-n^{1-2\gamma}) + \left(\frac{A_\delta}{\epsilon_n^{2+\delta/2}n^{\delta\gamma/2}} + \frac{B_\delta}{\epsilon_n^{(3+\delta)/2}n^{(1+\delta)\gamma}}\right)\mathbb{E}\left[|X_1 - \theta|^{2+\delta}\right]\right)\cdot T_n.
\end{aligned}$$

也就是说，

$$\mathbb{E}_P[E_n] \lesssim 1 + T_n\exp\{-n^{1-2\gamma}\} + \frac{T_n}{\epsilon_n^{2+\delta/2}n^{\delta\gamma/2}} + \frac{T_n}{\epsilon_n^{(3+\delta)/2}n^{(1+\delta)\gamma}},$$

只要 $T_n\epsilon_n^{-2-\delta/2} = o(n^{\delta\gamma/2})$，它便收敛到1。注意，只要$\sup_P \mathbb{E}_P|X_1 - \theta|^{2+\delta} < \infty$，完全相同的论证在分布一致情形中同样成立。

### C.7 推论的证明 3.9

取 $\epsilon_n = 1/\log n$和 $\gamma = 0.49$后，对于任意常数 $C$，有 $C\cdot n^{\delta 0.24} = o(n^{\delta\gamma/2}/\log^{2+\delta/2} n)$。因此，仍需计算该积分。设 $F$ 是均值为零、具有

方差 $v^2$ 的高斯分布。令 $U(X^n) = n(\hat{s}_n^2 + \epsilon_n)$。计算

$$\begin{aligned} G_n &:= \int \exp\left\{\lambda S_n(\theta) - \frac{\lambda^2}{2} U(X^n)\right\} \mathrm{d}F \\ &= \frac{1}{\sqrt{2\pi v^2}} \int \exp\left\{\lambda S_n(\theta) - \frac{\lambda^2}{2} U(X^n) - \frac{\lambda^2}{2v^2}\right\} \mathrm{d}\lambda \\ &= \frac{1}{\sqrt{2\pi v^2}} \int \exp\left\{-\frac{\lambda^2(v^2 U(X^n) + 1) - 2v^2 \lambda S_n(\theta)}{2v^2}\right\} \mathrm{d}\lambda. \end{aligned}$$

令 $h_n = v^2 U(X^n) + 1$，在这种情况下，我们可以将上式改写为

$$\begin{aligned} G_n &= \frac{1}{\sqrt{2\pi v^2}} \int \exp\left\{-\frac{\lambda^2 - 2v^2\lambda S_n(\theta)/h_n}{2v^2/h_n}\right\} \mathrm{d}\lambda \\ &= \frac{1}{\sqrt{2\pi v^2}} \int \exp\left\{-\frac{(\lambda - v^2 S_n(\theta)/h_n)^2 - v^4 S_n^2(\theta)/h_n^2}{2v^2/h_n}\right\} \mathrm{d}\lambda \\ &= \exp\left\{\frac{v^2 S_n^2(\theta)}{2h_n}\right\} \frac{1}{\sqrt{2\pi v^2}} \int \exp\left\{-\frac{(\lambda - v^2 S_n(\theta)/h_n)^2}{2v^2/h_n}\right\} \mathrm{d}\lambda. \end{aligned}$$

注意到最后的被积函数是方差为$v^2/h_n$的未归一化高斯分布的密度，我们有

$$G_n = \exp\left\{\frac{v^2 S_n^2(\theta)}{2h_n}\right\} \frac{1}{\sqrt{h_n}} = \exp\left\{\frac{v^2 \varphi^2(X^n)}{2h_n} - \frac{1}{2}\log(h_n)\right\}.$$

令 $\rho = v^2$，证明完成。

**命题 4.4 的证明**

固定 $m \in \mathbb{N}$，并令 $\tau$ 为任意停时。定义 $T_m = m^{\delta \cdot 0.24}$，并考虑事件

$$A_m = \left\{\sup_{k \geq m} \left|\hat{s}_k^2 - \frac{U_k^{\otimes}}{k}\right| \leq \frac{1}{\log(m)}\right\}.$$

在定理3.8 的证明中，我们证明了

$$P(A_m^c) \leq 4\exp(-m^{1-2\gamma}) + \left(\frac{A_\delta}{m^{\delta\gamma/2}\epsilon^{2+\delta/2}} + \frac{B_\delta}{\epsilon^{(3+\delta)/2} m^{(1+\delta)\gamma}}\right) \mathbb{E}_P |X_1 - \theta|^{2+\delta}, \tag{C.17}$$

请注意，在该证明中，我们没有在事件 $k \geq m$ 中对 $A_m$ 取上确界，因为这并非必需。不过，由于 Ruf 和 Waudby-Smith 的非渐近强大数定律 [47], 在引理 C.2中给出的界具有一致性，因此加入这个上确界后，界仍然相同。由此，记

$$\begin{aligned} \mathbb{E}_P\left[E_{\tau \vee m}^{(m)}\right] &= \mathbb{E}_P\left[E_{\tau \vee m}^{(m)} \mathbf{1}\{A_m\}\right] + \mathbb{E}_P\left[E_{\tau \vee m}^{(m)} \mathbf{1}\{A_m^c\}\right] \\ &\leq \mathbb{E}_P\left[E_{\tau \vee m}^{(m)} \mathbf{1}\{A_m\}\right] + T_m P(A_m^c). \end{aligned}$$

正如我们在定理 3.8的证明中所论证的， $T_m P(A_m^c) \xrightarrow{m\to\infty} 0$。此外，在 $A_m$上，对于所有 $s_k -$ 对数$^{-1}(m) \geq U_k^{\otimes}/k$ 均成立， $k \geq m$包括 $k = \tau \vee m$。因此，

$$\begin{aligned}\mathbb{E}_P\left[E_{\tau\vee m}^{(m)}\mathbf{1}\{A_m\}\right] &= \mathbb{E}_P\left[\left\{\int_{\lambda\in\mathbb{R}}\exp\left\{\lambda\sum_{i=1}^{\tau\vee m}(X_i-\theta)-U_{\tau\vee m}^{\otimes}\frac{\lambda^2}{2}\right\}dF(\lambda)\wedge T_m\right\}\mathbf{1}\{A_m\}\right]\\ &\leq \mathbb{E}_P\left[\int_{\lambda\in\mathbb{R}}\exp\left\{\lambda\sum_{i=1}^{\tau\vee m}(X_i-\theta)-U_{\tau\vee m}^{\otimes}\frac{\lambda^2}{2}\right\}dF(\lambda)\right]\leq 1,\end{aligned}$$

其中最后一个不等式由 Howard 等人 [27] 的结果得出（特别地，上述过程是一个初值为一的非负超鞅，因此在所有停时处的期望至多为 1）。至此证明完成。

## C.9 定理的证明 B.1

我们依次考虑每个渐近 e 变量。

### C.9.1 $(E_n^{\mathrm{sn}})$ 是一个渐近 e 值。

该渐近 e 值基于 Howard 等人 [27, 引理 3f]中的如下结果。另见 Chugg 等人 [7, 引理 3.1]，其陈述与我们此处的表述更为接近（尽管是在多元情形下），并附有直接证明。

**引理 C.3.** 对于任意取值于 $\mathbb{R}$ 的可预测序列 $(\lambda_n)$ （即 $\lambda_n$ 是 $\mathcal{F}_{n-1}$ $=\sigma(X_1,\ldots,X_{n-1})$-可测的），过程 $(E_n^\star)_n$ 由下式给出

$$E_n^\star(\lambda) := \exp\left\{\sum_{i=1}^n \lambda_i(X_i-\theta)-\frac{1}{2}V_n^{\otimes}\right\}, \tag{C.18}$$

是一个检验超鞅，其中 $V_n^{\otimes}$由下式给出

$$V_n^{\otimes} := \frac{1}{3}\sum_{i=1}^n \lambda_i^2((X_i-\theta)^2+2\sigma^2). \tag{C.19}$$

特别地， $E_\tau^\star$ 在任意停时 $\tau$都是一个 *e*值。

回想一下，我们试图证明

$$E_n^{\mathrm{SN}}(\theta;\lambda^n) = \mathbb{E}_P\exp\left\{\sum_{i\leq n}\lambda_i(X_i-\theta)-\sum_{i\leq n}\frac{\lambda_i^2}{6}[(X_i-\theta)^2+2\widehat{\sigma}_{i-1}^2]\right\},$$

定义了一个渐近 e 值。记

$$\lambda_i(X_i-\theta)-\frac{1}{6}\lambda_i^2[(X_i-\theta)^2+2\widehat{\sigma}_{i-1}^2] = \lambda_i(X_i-\theta)-\frac{1}{6}\lambda_i^2[(X_i-\theta)^2+2\sigma^2]+\frac{\lambda_i^2}{3}(\sigma^2-\widehat{\sigma}_{i-1}^2),$$

so

$$\mathbb{E}_P\exp\left\{\lambda_i(X_i-\theta)-\frac{\lambda_i^2}{6}[(X_i-\theta)^2+2\widehat{\sigma}_{i-1}^2]|\mathcal{F}_{i-1}\right\} \leq \mathbb{E}_P\exp\left\{\frac{\lambda_i^2}{3}(\sigma^2-\widehat{\sigma}_{i-1}^2)\right\}, \tag{C.20}$$

因此，全期望定律给出

$$\mathbb{E}_P \exp\left\{\sum_{i\le n}\lambda_i(X_i-\theta)-\sum_{i\le n}\frac{\lambda_i^2}{6}[(X_i-\theta)^2+2\widehat{\sigma}_{i-1}^2]\right\}\le \mathbb{E}_P\exp\left\{\sum_{i\le n}\frac{\lambda_i^2}{3}(\sigma^2-\widehat{\sigma}_{i-1}^2)\right\}.$$

因此，剩下要证明的是右侧的项的极限至多为 1。这可通过以下引理证明。

**引理 C.4**。 假设 *(i)* 上确界$_n\,\lambda_n<\infty$，*(ii)* $\sum_{i\le n}\lambda_i^2\xrightarrow{n\to\infty}L<\infty$ 几乎必然且 *(iii)* $\lambda_i^2/\sum_{j\le n}\lambda_j^2\xrightarrow{n\to\infty}0$ 几乎必然。那么

$$\lim_n \mathbb{E}_P\exp\left(\frac{1}{3}\sum_{i\le n}\lambda_i^2(\sigma^2-\widehat{\sigma}_{i-1}^2)\right)=1. \tag{C.21}$$

证明。 令 $A_n=\sum_{i\le n}\lambda_i^2$。我们声称，项 $A_n^{-1}\sum_{i\le n}\lambda_i^2\widehat{\sigma}_{i-1}^2$ 几乎必然趋于 $\sigma^2$。若如此，则由连续映射定理,

$$\sum_{i\le n}\lambda_i^2(\sigma^2-\widehat{\sigma}_{i-1}^2)=A_n\left(\sigma^2-\frac{\sum_{i\le n}\lambda_i^2\widehat{\sigma}_{i-1}^2}{A_n}\right)\xrightarrow{a.s.}L(\sigma^2-\sigma^2)=0.$$

由于根据假设 $\sup_n\sum_{i\le n}(\sigma^2-\widehat{\sigma}_{i-1}^2)\le\sup_n\sum_{i\le n}\lambda_i^2\sigma^2<\infty$，我们可以应用控制收敛定理，得出

$$\lim_n \mathbb{E}_P\exp\left(\frac{1}{3}\sum_{i\le n}\lambda_i^2(\sigma^2-\widehat{\sigma}_{i-1}^2)\right)=\mathbb{E}_P\exp\left(\frac{1}{3}\lim_n\sum_{i\le n}\lambda_i^2(\sigma^2-\widehat{\sigma}_{i-1}^2)\right)=1.$$

现在只需证明 $\sum_{i\le n}w_{i,\ n}\widehat{\sigma}_{i-1}^2\xrightarrow{a.s.}\sigma^2$，其中 $w_{i,\ n}=\lambda_i^2/A_n$。固定 $\epsilon>0$，并令$\omega$ 为一条满足 $\widehat{\sigma}_{i-1}^2\to\sigma^2$ 且 $w_{i,\ n}\to 0$的样本路径。于是存在某个 $N=N_{\epsilon,\ \omega}$，使得对所有 $i>N$，都有 $|\widehat{\sigma}_{i-1}^2(\omega)-\sigma^2|\le\epsilon$。因此，注意到 $\sum_{i\le n}w_{i,\ n}=1$,

$$\left|\sum_{i\le n}w_{i,n}\widehat{\sigma}_{i-1}^2(\omega)-\sigma^2\right|\le\sum_{i\le n}w_{i,n}|\widehat{\sigma}_{i-1}^2(\omega)-\sigma^2|\le\sum_{i\le N}w_{i,n}|\widehat{\sigma}_{i-1}^2(\omega)-\sigma^2|+\epsilon\sum_{N<i\le n}w_{i,n}.$$

最后一个求和至多为 $\epsilon$，因为 $\sum_{N<i\le n}w_{i,\ n}\le\sum_{i\le n}w_{i,\ n}=1$。令 $W_N=\max_{i\le N}|\widehat{\sigma}_{i-1}^2(\omega)-\sigma^2|$，倒数第二个求和可估计为

$$\sum_{i\le N}w_{i,n}|\widehat{\sigma}_{i-1}^2(\omega)-\sigma^2|\le N\cdot W_N\cdot\max_{i\le N}w_{i,n}\xrightarrow{n\to\infty},$$

由于$w_{i,n}=\lambda_i^2/A_n\to 0$ 对每个固定的 $i$ 依假设成立。由于 $\omega$ 落在一个以概率1出现的集合中，我们得出结论：$\sum_{i\le n}w_{i,n}\widehat{\sigma}_{i-1}^2\to\sigma$ 以概率1成立。 □

### C.9.2 ($E_n^{\mathrm{reg}}$) 是一个渐近 e 值

设 $\lambda_n \to \lambda$ 几乎必然。首先假设 $\eta > 0$ 是固定的。记

$$E_n^{\mathrm{REG}}(\theta;\lambda_n,\eta) = \exp\left(\lambda_n \frac{\sqrt{n}\,(\overline{X}_n - \theta)}{\widehat{\sigma}_n + \eta\, V_n(\theta)/\sqrt{n}} - \frac{\lambda_n^2}{2}\right)$$
$$= \exp\left(\lambda_n \frac{\sqrt{n}(\overline{X}_n - \theta)}{\sigma} \cdot \frac{\sigma}{\widehat{\sigma}_n + \eta\, V_n(\theta)/\sqrt{n}} - \frac{\lambda_n^2}{2}\right).$$

根据强大数定律，我们有

$$\widehat{\sigma}_n \xrightarrow{\text{a.s.}} \sigma \quad \text{and} \quad \frac{V_n(\theta)}{\sqrt{n}} \xrightarrow{\text{a.s.}} \sigma.$$

此外，根据中心极限定理，

$$\frac{\sqrt{n}(\overline{X}_n - \theta)}{\sigma} \xrightarrow{d} Z \quad \text{with} \quad Z \sim N(0,1).$$

应用斯卢茨基定理，可得

$$\exp\left(\lambda_n \frac{\sqrt{n}(\overline{X}_n - \theta)}{\widehat{\sigma}_n + \eta\, V_n(\theta)/\sqrt{n}} - \frac{\lambda_n^2}{2}\right) \xrightarrow{d} \exp\left(\lambda \frac{Z}{1+\eta} - \frac{\lambda^2}{2}\right).$$

引理 C.1 第 (ii) 部分保证了 $\{E_n^{\mathrm{REG}}(\theta;\ \lambda_n,\ \eta)\}_n$的一致可积性，因此[5] 中的定理25.12表明，依分布收敛可推出

$$\limsup_{n\to\infty} \mathbb{E}_P\left[E_n^{\mathrm{REG}}(\theta;\lambda_n,\eta)\right] = \mathbb{E}_P\left[\exp\left(\lambda \frac{Z}{1+\eta} - \frac{\lambda^2}{2}\right)\right].$$

由于 $Z \sim N$ 属于 (0，1)，我们可以计算高斯矩母函数，得到

$$\mathbb{E}_P\left[\exp\left(\lambda \frac{Z}{1+\eta} - \frac{\lambda^2}{2}\right)\right] = \exp\left(-\frac{\lambda^2}{2}\left[1 - \frac{1}{(1+\eta)^2}\right]\right) \le 1,$$

当且仅当 $\lambda = 0$时取等。这证明了 $E_n^{\mathrm{REG}}(\theta;\lambda_n,\eta)$ 是渐近 e 值。若 $\eta_n \to \eta$且 $\lambda_n \lesssim \eta_n$，则引理C.1 第 (iii) 部分表明 $\{E_n^{\mathrm{REG}}(\theta;\lambda_n,\eta_n)\}$ 是一致可积的。再次由斯卢茨基定理，我们有

$$\exp\left(\lambda_n \frac{\sqrt{n}(\overline{X}_n - \theta)}{\widehat{\sigma}_n + \eta_n\, V_n(\theta)/\sqrt{n}} - \frac{\lambda_n^2}{2}\right) \xrightarrow{d} \exp\left(\lambda \frac{Z}{1+\eta} - \frac{\lambda^2}{2}\right),$$

且其余证明与针对 IWR e变量的证明保持不变。

## C.10 定理B.2的证明

在整个证明中，当我们写 $\sup_P$ 时，该上确界应理解为在$Q(\theta)$上取。

设 $A_n = \lambda_n B_n - \lambda_n^2/2$, $B_n = \lambda_n S_n(\theta)/(n\widehat{\sigma}_n^2 + V_n(\theta)\eta)$，且 $g(y) = \exp(y)$。设 $Z = -\lambda^2/2$以概率 1 成立（其中 $\lambda_n \to \lambda$），于是 $g(Z) = \mathbb{E}_P[g(Z)] = \exp(-\lambda^2/2)$。我们要证明 $g(A_n) = E_n^{\mathrm{REG}}(\theta;\lambda_n,\eta_n)$ 满足命题 B.9 的两个条件。

首先，回顾一下在引理C.1 的证明中，我们已经证明

$$E_n^{\mathrm{REG}}(\theta;\lambda,\eta) \le \exp\left(\frac{\lambda}{\eta}|T_n|\right),$$

其中 $T_n = S_n(\theta)/V_n(\theta)$ 与通常一样。由于根据假设 $\lambda_n \lesssim \eta_n$，我们可以选取某个 $U$，使得

$$\max\left\{\sup_n \frac{\lambda_n}{\eta_n}, \lambda\right\} < U < \infty,$$

由此可得

$$E_n^{\mathrm{REG}}(\theta;\lambda_n,\eta_n) \le \exp(UT_n).$$

然后，采用与 $E_n^{\mathrm{IWR}}$ 情形类似的论证，写作

$$g(A_n)\mathbf{1}\left\{g(A_n) \ge K\right\} \le \frac{\exp(\lambda_n T_n/\eta_n)}{K^{(U-\lambda_n)/\lambda_n}} \le \frac{\exp(UT_n)}{K^{(U-\lambda_n)/\lambda_n}}.$$

由于 $n \to \infty$, $K^{(U-\lambda_n)/\lambda_n} \to K^{U/\lambda-1} > K$，因此 $1/K^{(U-\lambda_n)/\lambda_n} \xrightarrow{n\to\infty} 0$。（若 $\lambda = 0$，我们将 $U/0$ 解释为 $\infty$，并将 $1/\infty$ 解释为 $0$。）因此，要证明

$$\lim_{K\to\infty} \limsup_{n\to\infty} \sup_P \mathbb{E}_P[E_n^{\mathrm{REG}}(\theta;\lambda_n,\eta_n)\mathbf{1}\left\{E_n^{\mathrm{REG}}(\theta;\lambda_n,\eta_n) \ge K\right\}] = 0,$$

只需证明 $\sup_P \exp(UT_n)$ 有一个有限上界，且该上界与 $n$ 和 $k$ 无关。这与证明 $E_n^{\mathrm{IWR}}$ 时完全相同。

接下来，我们要证明 $\limsup_n \sup_P \mathbb{E}_P[g(A_n)\wedge K] \le \exp(-\lambda^2/2)\wedge K$ 对固定的 $K > 0$ 成立，这就是命题 B.9 的第二个条件。事实上，我们将证明 $\sup_P \mathbb{E}_P[g(A_n)\wedge K] \to 0$。记

$$\begin{aligned}
\mathbb{E}_P[g(A_n)\wedge K] &= \int_0^K P(g(A_n) > u)\mathrm{d}u \\
&= \int_0^K P\left(B_n > \frac{\log(u)}{\lambda_n} + \frac{\lambda_n}{2}\right)\mathrm{d}u \\
&\le \int_0^K P\Bigg(\frac{|S_n(\theta)|}{n\widehat{\sigma}_n^2} > \underbrace{\frac{\log(u)}{\lambda_n} + \frac{\lambda_n}{2}}_{:=\epsilon_n}\Bigg)\mathrm{d}u.
\end{aligned} \tag{C.22}$$

证明的其余部分将着重说明，对于所有 $P$，被积函数都以 $1/\mathrm{poly}(n)$ 的速率衰减，从而完成证明。尽管细节相当棘手，但直觉很简单：我们预期 $|S_n(\theta)|/(n\widehat{\sigma}_n^2)$ 会收敛到 $0$，因为 $S_n(\theta)/n \to 0$ 且 $\widehat{\sigma}_n^2 \to \sigma$。与此同时，$\epsilon_n$ 会收敛到一个常数或 $+\infty$。

令 $\sigma_P^2$ 为 $P$的方差，由于偏度有限，因此方差有限。注意，如果$|S_n(\theta)|/(n\widehat{\sigma}_n^2) > \epsilon_n$，那么要么 $|S_n(\theta)| > \epsilon n\sigma_P^2/2$，要么 $\widehat{\sigma}_n^2 < \sigma_P^2/2$。（否则，$|S_n(\theta)| \le \epsilon_n n\sigma^2/2 \le \epsilon_n n\widehat{\sigma}_n^2$。）因此，

$$P\left(\frac{|S_n(\theta)|}{n\widehat{\sigma}_n^2} > \epsilon_n\right) \le P\left(|S_n(\theta)| > \frac{\epsilon_n n\sigma_P^2}{2}\right) + P\left(\widehat{\sigma}_n^2 < \frac{\sigma_P^2}{2}\right). \tag{C.23}$$

对于右侧的第一项（C.23），我们应用马尔可夫不等式：

$$P\left(|S_n(\theta)| > \frac{\epsilon_n n\sigma_P^2}{2}\right) \leq \frac{4\mathbb{V}(\sum_i X_i)}{n^2\epsilon_n^2\sigma_P^4} = \frac{4}{n\epsilon_n^2\sigma_P^2} \leq \frac{4C}{n\epsilon_n^2}, \tag{C.24}$$

根据一致有下界的方差这一假设，某个 $C > 0$。注意，由于 $n\epsilon_n^2 \to \infty$

$$n\epsilon_n^2 \asymp \frac{n}{\lambda_n^2} + n\lambda_n + n\lambda_n^2.$$

这些项中至少有一项趋于 $+\infty$：如果 $n\lambda_n^2 \asymp c$ 那么 $\lambda \asymp 1/\sqrt{n}$ 所以 $n\lambda_n^2 \asymp n^2 \to \infty$。现在来看式（C.23）中的下一项。令 $Y_i = X_i - \theta$，$\overline{Y} = \frac{1}{n}\sum_y Y_i$ 且 $\overline{Y^2} = \frac{1}{n}\sum_i Y_i^2$。回顾标准恒等式

$$\widehat{\sigma}_n^2 = \frac{n-1}{n}\left(\overline{Y^2} - \overline{Y}^2\right).$$

注意，$\widehat{\sigma}_n^2 \geq \sigma_P^2/2$ 当且仅当 $\overline{Y^2} - \overline{Y}^2 \leq \frac{n-1}{2n}\sigma_P^2$，这意味着 $\overline{Y}^2 > \sigma_P^2/4$ 或 $\overline{Y^2} \leq \frac{n-1}{2n}\sigma_P^2 + \frac{1}{4}\sigma_P^2 \leq \frac{3}{4}\sigma_P^2$。由并集界可得

$$P\left(\widehat{\sigma}_n^2 \geq \frac{\sigma_P^2}{2}\right) \leq P\left(\overline{Y}^2 > \frac{\sigma_P^2}{4}\right) + P\left(\overline{Y^2} \leq \frac{3}{4}\sigma_P^2\right). \tag{C.25}$$

在 $P$ 条件下，$\overline{Y^2}$ 是 $n$ 个均值为 $\sigma_P^2$ 的独立同分布随机变量的平均值。因此，根据马尔可夫不等式，

$$\begin{aligned} P\left(\overline{Y^2} \leq \frac{3}{4}\sigma_P^2\right) &= P\left(\sigma^2 - \overline{Y^2} \geq \frac{1}{4}\sigma_P^2\right) \\ &\leq P\left(|\sigma^2 - \overline{Y^2}| \geq \frac{1}{4}\sigma_P^2\right) \leq \frac{\mathbb{E}_P|n\sigma_P^2 - \sum_i Y_i^2|^{3/2}}{(n\sigma_P^2/4)^{3/2}}. \end{aligned}$$

接下来，我们使用冯·巴尔和埃森的不等式 [55]，该不等式指出，对于 $Z_i$ 个零均值的独立同分布随机变量，以及任意满足 $\mathbb{E}|Z_1|^p < \infty$ 的 $1 \leq p \leq 2$，都有

$$\mathbb{E}\left|\sum_{i\leq n} Z_i\right|^p \leq C_p \sum_{i\leq n} \mathbb{E}|Z_i|^p,$$

其中 $C_p$ 是一个依赖于 $p$ 的常数（与分布无关）。将其用于上式，并取 $p = 3/2$，可得

$$\frac{\mathbb{E}_P|n\sigma_P^2 - \sum_i Y_i^2|^{3/2}}{(n\sigma_P^2/4)^{3/2}} \leq \frac{C_p n\mathbb{E}_P|\sigma_P^2 - Y_1^2|^{3/2}}{(n\sigma_P^2/4)^{3/2}} \leq \frac{D_p(\sigma_P^3 + \mathbb{E}_P|Y_1|^3}{\sqrt{n}\sigma_P^3},$$

其中 $D_p = \sqrt{2}\cdot 4^{3/2}C_p$。这里我们使用了不等式 $(a+b)^p \leq 2^{p-1}(a^p + b^p)$。根据一致三阶矩的假设，我们有

$$\sup_P P\left(\overline{Y^2} \leq \frac{3}{4}\sigma_P^2\right) \leq \frac{D_p}{\sqrt{n}}\left(1 + \sup_P \frac{\mathbb{E}_P|Y_1|^3}{\sigma_P^3}\right) \leq \frac{D_p}{\sqrt{n}}(1+M). \tag{C.26}$$

接下来我们需要处理 $P\left(Y^2 > \sigma_P^4/4\right)$ 在（C.25）中的项，这可以通过切比雪夫不等式完成。记

$$P\left(\overline{Y}^2 > \frac{\sigma_P^2}{4}\right) = P\left(\left|\sum_{i\le n}(X_i-\theta)\right| > \frac{n\sigma_P}{2}\right) \le \frac{4\mathbb{V}(\sum_i X_i)}{n^2\sigma_P^2} = \frac{4}{n}.$$

回到（C.25），我们已经证明

$$P\left(\widehat{\sigma}_n^2 \ge \frac{\sigma_P^2}{2}\right) \le \frac{D_p}{\sqrt{n}}(1+M)+\frac{4}{n},$$

将其与（C.24）、（C.23）以及（C.22）相结合，可得

$$\begin{aligned}\limsup_{n\to\infty}\sup_P \mathbb{E}_P[g(A_n)\wedge K] &\le \limsup_{n\to\infty}\int_0^K\left(\frac{4C}{n\epsilon_n^2}+\frac{D_p}{\sqrt{n}}(1+M)+\frac{4}{n}\right)\mathrm{d}u\\ &\le \int_0^K \limsup_{n\to\infty}\left(\frac{4C}{n\epsilon_n^2}+\frac{D_p}{\sqrt{n}}(1+M)+\frac{4}{n}\right)\mathrm{d}u = 0,\end{aligned}$$

其中，我们使用反向法图引理交换了上极限与积分，此外还用到了上文所论证的 $n\epsilon_n^2\to\infty$ 这一事实。

由此我们已证明 $g(A_n)$ 满足命题B.9的两个条件，从而完成了证明。

## C.11 定理B.3的证明

现在我们转向 $\mathcal{H}_n^{\mathrm{REG}}$。按通常做法，将 $E_n^{\mathrm{REG}}(\theta;\lambda,\eta) < 2/\alpha$ 改写为

$$y - \widehat{\sigma}_n A_n < \eta\frac{V_n(\theta)}{\sqrt{n}}A_n,$$

这里仍有，$y = \overline{X_n-\theta} = S_n(\theta)/n$。由于 $\eta>0$ 且 $V_n(\theta)\ge 0$，因此当 $y<\widehat{\sigma}_n A_n$ 时，这个不等式显然成立，即 $\theta > X_n - \widehat{\sigma}_n A_n$。于是 $\overline{R_1} = (X_n-\widehat{\sigma}_n A_n,\infty) \subset \{\theta : \overline{E_n^{\mathrm{REG}}(\theta;\lambda,\eta)} < 2/\alpha\}$。为求出该集合的其余部分，假设 $y\ge\widehat{\sigma}_n A_n$。那么，上述显示式两边均为正，我们可以将两边平方，得到不等式

$$y^2 - 2y\widehat{\sigma}_n A_n + \widehat{\sigma}_n^2 A_n^2 < \frac{\eta^2}{n}(\widehat{\sigma}_n^2(n-1)+ny^2)A_n^2. \tag{C.27}$$

合并各项后，将其改写为多项式方程

$$Q(y) := y^2(1-\eta^2A_n^2) - 2y\widehat{\sigma}_nA_n + \widehat{\sigma}_n^2A_n^2\left(1-\eta^2\left(\frac{n-1}{n}\right)\right) < 0. \tag{C.28}$$

对二次方程进行相关算术运算后，$Q(y)$ 的根为

$$y_\pm = \frac{\widehat{\sigma}_nA_n\left(1\pm\sqrt{1-(1-\eta^2A_n^2)(1-\eta^2\frac{n-1}{n})}\right)}{1-\eta^2A_n^2} \tag{C.29}$$

对于 $1-\eta^2 A_n^2 > 0$，我们断言 $y_- < \widehat{\sigma}_n A_n$，因此将其排除在考虑范围之外（因为我们考虑的是满足 $y \geq \widehat{\sigma}_n A_n$ 的 $y$）。事实上，令 $a = \eta^2 A_n^2$ 和 $b = \eta^2(n-1)/n$，并注意到

$$\sqrt{1-(1-a)(1-b)} \geq 1-(1-a)(1-b) = b - ab + a > a,$$

由于 $a,\ b < 1$。因此，

$$y_- = \frac{\widehat{\sigma}_n A_n (1-\sqrt{1-(1-a)(1-b)}}{1-a} > \frac{\widehat{\sigma}_n A_n (1-a)}{1-a} = \widehat{\sigma}_n A_n.$$

因此，我们考虑的解是 $y_+$，这给出了半区间 $\theta > X_n - y_+$。该区间包含先前的半区间，因为 $y_+ \geq \widehat{\sigma}_n A_n$（这一点很容易看出：$y_+$在分子中乘以某个 $\widehat{\sigma}_n A_n > 1$，并在分母中除以某个 $< 1$）。求解 $E_n^{\mathrm{REG}}(\theta; -\lambda,\ \eta) < 2/\alpha$ 可得到另一个半区间 $\theta < X_n - y_+$，从而完成证明。

# D 补充实验与仿真细节

## D.1 仿真细节

用于复现实验以及独立使用这些方法的代码可在以下网址找到：
https://github.com/bchugg/asymp-posthoc-confint。

**图 2。** 我们使用随机种子 0，从一个独立同分布的高斯模型中生成一个长度为 $N = 20000$ 的样本。对于每个 $t \in \{100, 200, 500, 1000, 2000, 3000, 5000, 10000, 20000\}$，我们对这一实现的共同前缀 $X_{1:t}$ 应用每种方法，并记录所得双侧区间的宽度 $U_t - L_t$。我们固定 $\alpha = 0.05$。对于事前锚定，我们同时考虑 $\alpha_0 = \alpha$ 和 $\alpha_0 = \alpha/10000$。对于 $(\mathcal{H}_n^{\text{R-WS}})$，我们使用 $\rho = 2$ 和 $\delta = 0.1$；对于 $(\mathcal{H}^{\text{MIX}}{}_n{}^{\text{IWR}})$，我们使用 $R = 20$ 和 $\kappa = 5$。

图 3。我们使用 confseq 软件包（https://github.com/gostevehoward/confseq/tree/master/src/confseq）实现投注置信区间。我们生成一个长度为 $N = 10,000$ 的单一数据流，其中包含独立同分布的伯努利分布观测值 $X_1, \ldots, X_N \sim \text{Bernoulli}(p)$，且 $p = 0.25$（随机种子固定为 0）。对于每个 $t \in \{300, 600, \ldots, 9900\}$，所有方法均应用于共同前缀 $X^t$，并记录所得双侧区间的宽度 $U_t - L_t$。伯恩斯坦区间基于 $X^t$ 和已知方差$p(1-p) = 0.25(1-0.25)$计算。对于混合 IWR 方法，我们使用 $R = 20$ 和 $\kappa = 5$。对于（$\mathcal{H}_n^{\text{IWR}}$），我们采用事前锚定，并使用 $\alpha_0 = \alpha$（如正文所述），即设置 $\lambda = \sqrt{2\log(2/\alpha_0)}$，

## D.2 $R$ 和 $\kappa$ 对 $E^{\mathbf{mix}}$的影响，$_{n}$ iwr

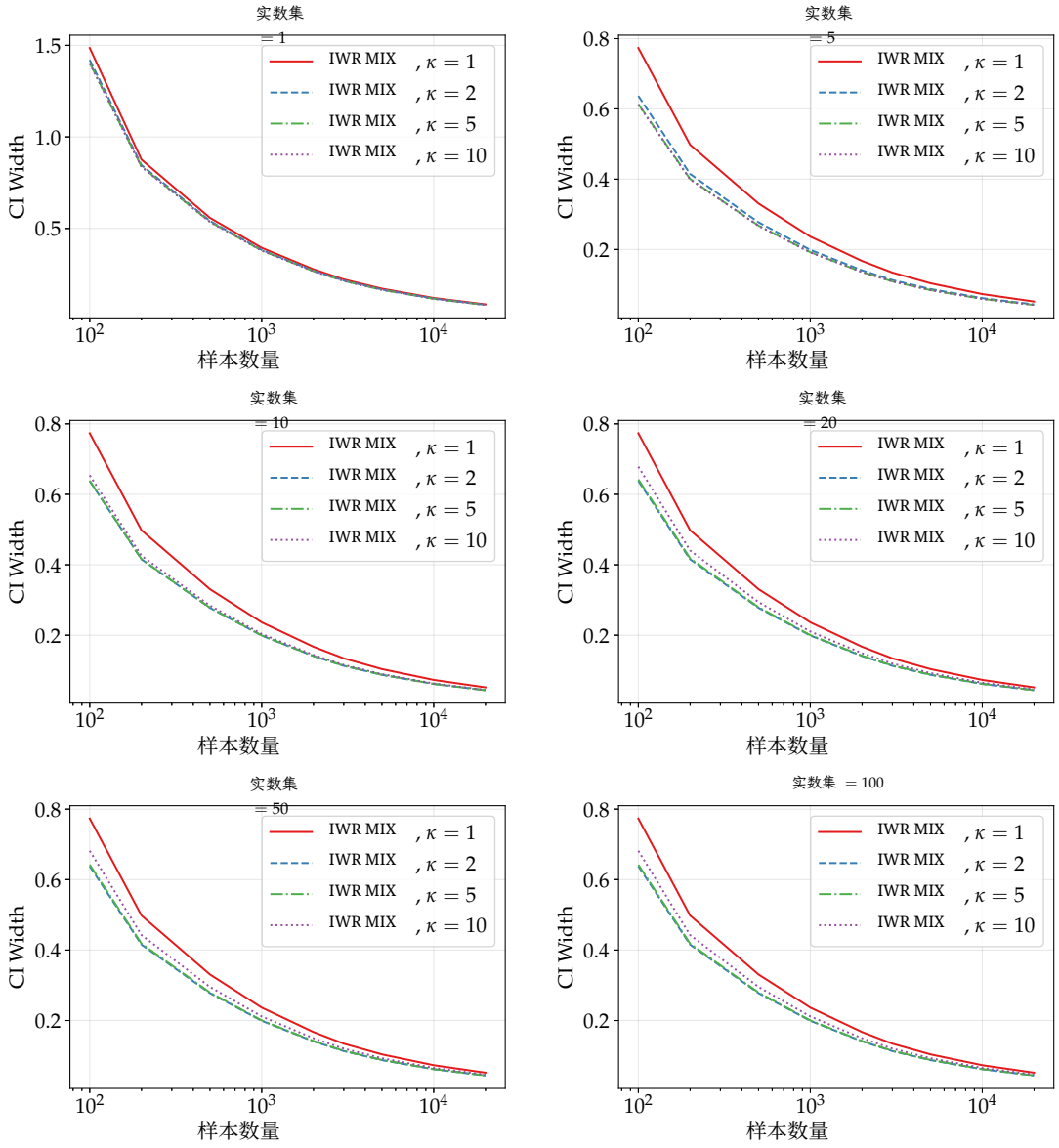



图5：基于截断高斯混合的 $R$ 和 $\kappa$ 对 APH-CI宽度的影响。观测值独立同分布地取自方差为一的中心高斯分布。我们在所有图中使用$\alpha = 0.05$ 。

图5 展示了参数 $R$ （截断高斯分布的范围）以及$\kappa$ （其方差）对（$\mathcal{H}^{\text{MIX}}$、$_{n}$ $^{\text{IWR}}$）宽度的影响。结果针对高斯分布绘制，但对于 t 分布，结果类似。对于（$\mathcal{H}^{\text{MIX}}$、$_{n}$ $^{\text{REG}}$），结果也类似，相关结果见附录 B.7。

### D.3 APH 置信区间：REG 与 IWR

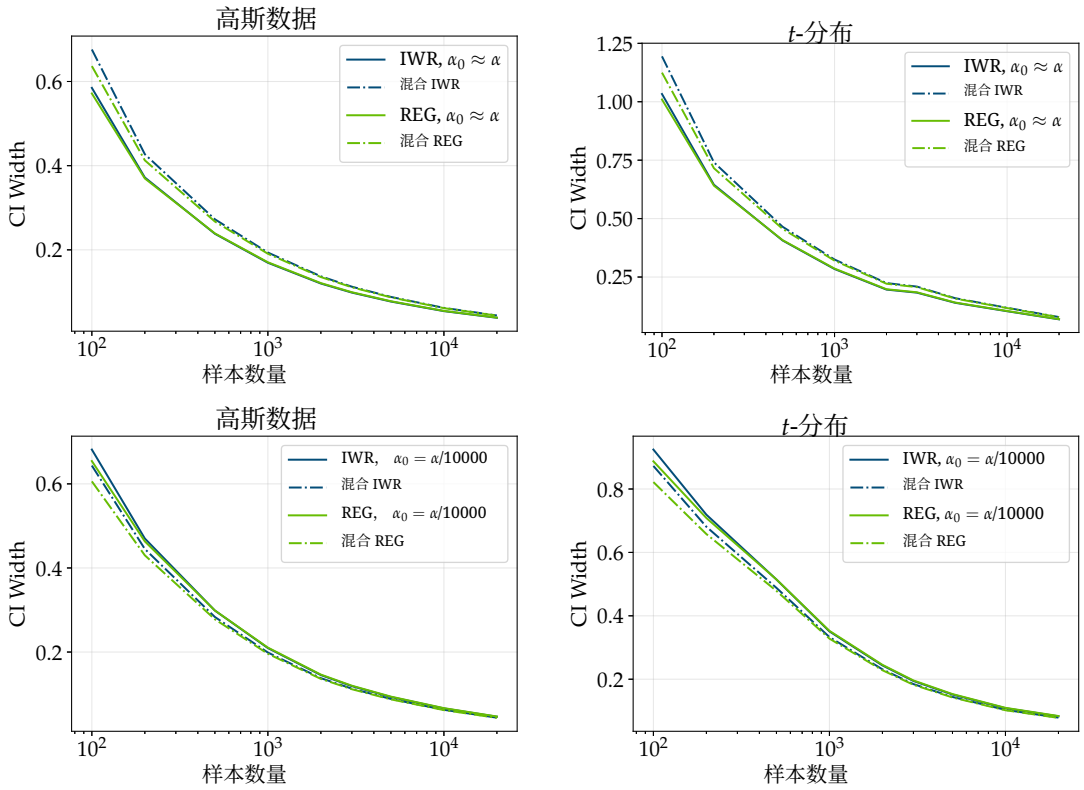



图6：基于 APH-CI 的两个 IWR 渐近 e 变量与基于 APH-CI 的两个 REG e 变量（附录 B.2 和B.7）。我们从高斯分布或自由度为三的 t 分布生成数据。与图 2 类似， IWR和 REG 均采用事前锚定，并研究两种情形：当 $\alpha_0 \approx \alpha$ 以及当$\alpha/\alpha_0 \gg 0$。

图6研究了以下两种 APH-CI，它们由 IWR 渐近 e 变量得出，即定理3.1 和3.6；以及以下两种，它们由 REG 渐近 e 变量得出，即定理B.3和B.12。我们通过事前锚定来实例化（$\mathcal{H}_n^{\mathrm{IWR}}$）和（$\mathcal{H}_n^{\mathrm{REG}}$）。其余两种基于混合方法。我们按照主论文中的设定选择 $R = 20$ 和 $\kappa = 5$ 。我们取 $\eta = 0.01$ （关于 $\eta$影响的研究见下一节）。总体而言，我们发现两种方法的表现极为相似。基于 APH-CI的 REGe变量有时在样本量较小时略微更紧，但这种影响很快就会消失。鉴于基于 IWR的 APH-CI所需调优参数较少，我们建议在实践中使用它们。

## D.4 $\eta$ 对 $\mathcal{H}_n^{\text{reg}}$的影响

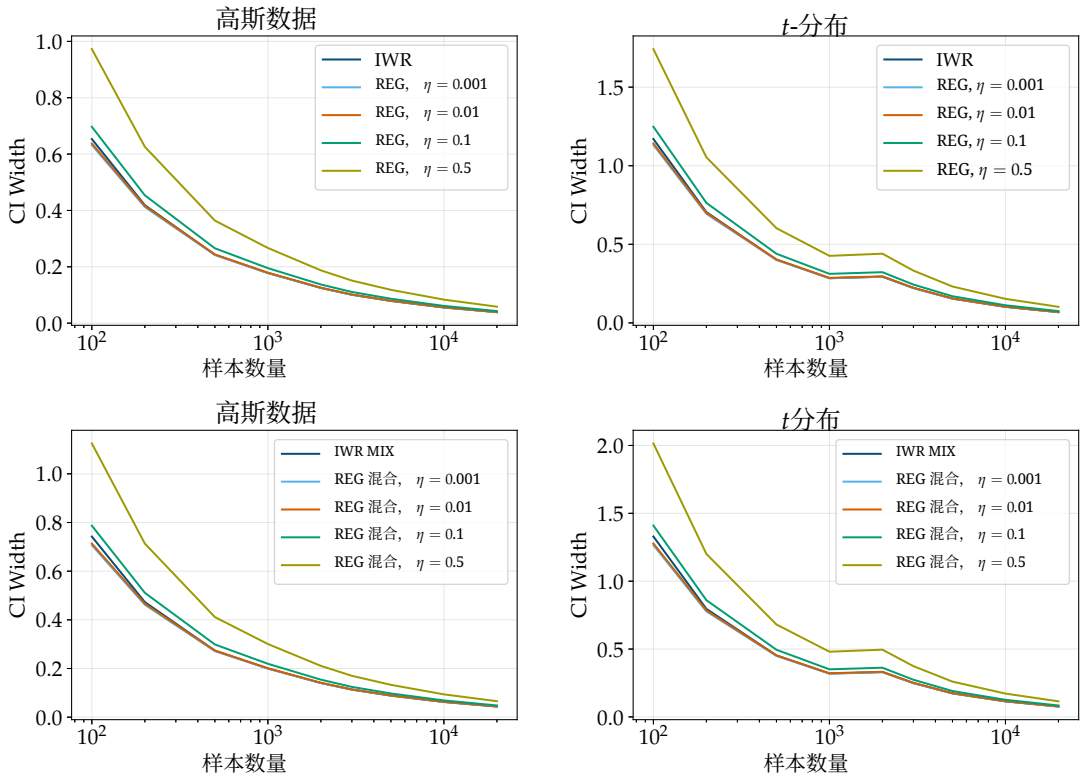



图7：基于 REG渐近 e 变量， $\eta$ 对两个 APH-CI的影响。此处$\alpha = 0.05$。顶行图中的 APH-CI 均采用事前锚定，并且 $\alpha_0 = 0.01$。底行的混合 APH-CI均采用$R = 20$ 和 $\kappa = 5$。

图7 研究了自由参数 $\eta$ 对（$\mathcal{H}_n^{\text{REG}}$）和（$\mathcal{H}^{\text{MIX}}$, $_n$ $^{\text{REG}}$）的影响。随着 $\eta \to 0$，宽度逐渐收敛。一般而言，当 $\eta$较小时， REGAPH-CI与 IWR APH-CI之间的差异可以忽略不计。