arXiv:2601.07503v1 [math.ST] 12 Jan 2026

金标准过程的马尔可夫式污染：一种半参数方法

Claire Lacour 和 Pierre Vandekerkhove

摘要

本文研究一种随机过程：两个独立的随机性来源随时间按照一个未观测到的平稳马尔可夫选择过程进行混合：i）一个分布已知的平稳过程（金标准）；ii）一个分布未知的纯独立同分布序列（污染过程）。在此框架下，我们提出两种不同的方法，在最小条件（可识别性、平稳性和时间依赖性）下，估计隐马尔可夫选择过程的转移机制，以及未知独立同分布序列的分布；我们不假设该分布属于任何参数族。我们证明，这两种估计量都能一致地估计欧几里得转移参数，并进一步证明，其中一种估计量具有 $\sqrt{n}$-一致性，可用于建立关于未知污染序列累积分布函数的函数中心极限定理。我们通过若干具有挑战性的实例展示了这些估计量的数值性能。

关键词。 污染、隐马尔可夫、时间顺序混合、半参数、绕过源、模型几何。

1 引言

过去二十年来，半参数潜在模型得到了广泛研究，并被证明在某些情形下是适用的工具：此时观察到多个未标记的随机性来源，而输出特征的参数假设又不易确定。这一研究领域始于 Hall and Zhou (2003) 的论文，他们考虑了 $\mathbb{R}^d$、 $d \geq 3$中两个分布的混合，每个分布都具有相互独立的分量。在这项开创性工作中，作者首次证明了该模型具有可识别性，并且无需对混合分布作出任何参数假设即可进行一致估计；关于非参数二阶边缘分布无张量化条件，参见 Hall and Zhou (2003) 中的定理4.1。之后，Bordes 等（2006b）和 Hunter 等（2007）分别采用两种不同方法，共同研究了一个单变量半参数双成分混合模型，其中两个对称成分仅相差一个位移参数。作者证明了该模型的可识别性，并据此提出了一致的估计方法；不过，这些论文尚未讨论$\sqrt{n}$一致性。近期，Butucea and Vandekerkhove (2014) 以及 Butucea 等（2017）研究了该模型一类新型基于傅里叶变换的估计量的一致性和收敛速度，其中包括渐近正态性。沿用类似方法，Werner 等（2020）针对具有对称噪声的半参数回归混合模型，提出了上确界范数下的自适应估计。Bordes 等（2007）还提出了广义半参数版本的期望最大化（EM）算法，以处理对称且经位移的 m成分混合模型；多变量情形下 Levine 等（2011）提出的 MM算法也可供参考。

在Bordes 等（2006a）中，作者考虑了经典的污染模式 l:

$$F(x) = pF_0(x) + (1-p)F_1(x), \quad x \in \mathbb{R}, \tag{1}$$

其中， $p \in]0\,1[$ 是一个比例参数， F 表示独立同分布（i.i.d.）观测值的累积分布函数（cdf）， F_0 表示用于建模金标准的已知累积分布函数，而 F_1 是一个未知（污染）但对称的累积分布函数，用于建模偏离 F_0的情况，且不对其作任何参数假设。过去几十年来，这一模型得到了广泛研究，并与各种应用相关；有关这一主题的全面综述，参见 Shen 等（2018）。在考察受意外事件影响的一般情形时，该模型尤其值得关注：i）大流行病对死亡率的影响，参见 Milhaud 等（2024b）；ii）微阵列分析中病变组织的存在，参见 McLachlan 等（2006）；相关的多重检验问题则参见 Benjamini 和 Hochberg（1995）、Donoho 和 Jin（2004）；iii）银河系背景下恒星的金属丰度和视向速度等变量观测，参见 Walker 等（2009）；iv）存在额外品种时的树木直径建模，参见 Podlaski 和 Roesch（2014）。从半参数视角来看，值得一提的是 Patra 和 Sen（2016）：他们为模型（1）给出了一个有意义且适定的可识别性定义，并提出了相应的估计方法和收敛速度。Pommeret 和 Vandekerkhove（2019）提出了一种检验程序，用于考察式（1）中涉及的未知成分 F_1是否属于某个特定参数族。Milhaud 等（2024b）和 Milhaud 等（2024a）分别考察了关于未知 F_1成分等式的双样本和 K样本检验问题。Milhaud 等（2024a）还提出了一种基于模型的聚类程序，用于归集污染成分相同（组内）的样本组。受流式细胞术中稀有细胞检测问题的启发，Gaucher 等（2025）研究了基于分布分别为 F_0、 F_1、 F的三个样本，对 $\{F = F_0\}$与 $\{F = pF_0 +（1-p）F_1\}$进行检验。

混合模型按时间顺序扩展后，便产生了所谓的隐马尔可夫模型（HMM）；这类模型本质上是由潜在马尔可夫链选取的随机性的混合。对于该模型在有限离散情形下的可识别性，最早由Gilbert（1959进行了研究，而最大似然估计量（MLE）的渐近行为则在Baumand Petrie （1966年）中得到了完整处理。20世纪90年代初，Leroux （1992）重新研究了这一模型：其中潜在马尔可夫链具有有限状态空间，但隐马尔可夫模型的发射分布位于欧几里得空间中（不一定是有限的）。在Leroux （1992）中，作者在最小条件下证明了最大似然估计量的一致性，而该模型的渐近正态性随后由Bickeletal.（1998年）建立。此后，隐马尔可夫模型在参数情形下的推断得到了广泛研究，Capp´e et al. （2005年）还就这一主题提出了一部优秀的专著；关于近期的发展与应用，另请参见Bouguilaetal. （2022）。

近年来，Gassiat 和 Rousseau（2016年）提出了非参数有限平移隐马尔可夫模型。作者建立了可识别性条件，并提出一种基于傅里叶方法的一致估计量，用于估计总体数量、平移参数以及两个连续潜变量的联合分布；他们证明，在温和的依赖性假设下，该估计量渐近正态。作者还提出了一种未知平移后密度的非参数估计量。当潜变量构成马尔可夫链时，作者证明其估计量在密度正则性类上具有极小极大自适应性。在Alexandrovich 等人（2016年）中，作者研究了有限状态隐马尔可夫模型的非参数识别与极大似然估计。在De Castro 等人（2016年）中，作者针对采用非参数发射建模的一般隐马尔可夫模型，探讨了其估计问题

分布。他们提出了一种新的惩罚最小二乘估计量：该估计量基于将发射分布投影到复杂度递增的嵌套子空间上，用于估计发射分布，具有统计最优性且在实践中易于处理。他们还证明，其发射分布非参数估计量满足一个非渐近预言机不等式。因此，这种新估计量的收敛速率在相差一个对数项的意义下具有极小极大自适应性。若想更全面地了解半参数／非参数混合模型或隐马尔可夫模型，我们推荐阅读Gassiat （2019年）和Xiang 等人（2019年）

本文旨在研究一种模型的半参数估计，该模型具有若干与 Vandekerkhove (2005年提出的隐马尔可夫马尔可夫模型混合模型（H4M）相同的特征。这些潜在模型基本上基于 K 独立马尔可夫过程（各自具有自身的动态特性），其观测值由一个取值于（标签）状态空间的潜在马尔可夫链随机选取 $\{1, \ldots, K\}$，简要介绍见第 2.1 节，相关内容出自 $K = 2$。这类模型能够描述具有以下特征的离散时间序列，因此颇具价值：(i) 潜变量发生状态变化时出现突变；(ii) 潜在马尔可夫链保持在同一状态期间呈现局部平稳性；(iii) 混合结构产生多峰边缘分布；以及 (iv) 相位型反馈效应（轨迹从过去延续而来）。在 Vandekerkhove (2005年) 中，作者提出了一种通用且易于处理的方法，用于对这类模型进行参数估计（要求平稳分布可参数化且模型具有可识别性），并详细验证了对于含高斯噪声的两个线性 AR(1) 模型的马尔可夫混合模型而言，这些假设均完全成立。此外，当独立马尔可夫过程的不变概率密度没有解析表达式时，还提出了一种蒙特卡洛方法，用于计算 H4M 的分割数据似然。

本文将研究的半参数模型与 H4M 具有一些共同特征，可由三个相互独立的过程定义：i）一个分布已知的平稳随机过程（金标准）；ii）一个分布未知且不假定属于任何参数族的独立同分布随机变量序列（污染过程）；iii）一个二状态马尔可夫链，按时间从前述两个随机源中选取观测值，也就是说，标签 0 表示金标准过程，标签 1 表示独立同分布序列。尽管该模型看起来比一般的 H4M 简单，但仍包含经典隐马尔可夫模型的复杂性：在部分时段（即潜在马尔可夫链处于状态 0 的阶段），数据具有已知的随机动力学；而在其余时段，数据则呈现未知的独立同分布模式，这与标准隐马尔可夫模型完全相同（以潜在马尔可夫链处于状态 1 为条件）。该模型通过马尔可夫状态过程，将随机动力学（可能具有马尔可夫性）与独立同分布模式连接起来；正如接下来的两段所述，这一点在网络安全领域可能尤其值得关注。

在Dass 等人（2021）中，作者对网络安全领域中隐马尔可夫模型的应用作了有趣的介绍。简言之，如今，隐马尔可夫模型因易于调整模型以预测未知/隐藏状态序列，而应用于自然语言处理和机器学习等多个领域。预测依据的是各个状态发出的特征或观测值。因此，隐马尔可夫模型在网络安全领域非常有用，因为网络攻击通常分多个阶段或步骤实施，而这些步骤未必总是显而易见，因为攻击者往往会试图掩盖其活动。不过，隐马尔可夫模型可以帮助识别数据中的模式，例如随时间变化的网络踪迹，并显然有助于判定攻击，参见 Holgado 等人（2020）。

另一方面，正如Korczynski 和 Duda(2014年) 中所述，基于训练数据对应用程序进行马尔可夫指纹识别，是一种广受欢迎的互联网流量活动分类方法。正如Korczynski 和 Duda(2014年) 所解释的，过去有关

流量分析和分类的研究表明，只要能够基于数据包或消息载荷（例如 HTTP 请求头）生成唯一特征签名，就可以高准确率地对应用程序进行分类。遗憾的是，这类方法在面对加密流量时会失效。在他们的研究中，作者提出了一种基于载荷的方法，用于识别通过安全套接层/传输层安全协议（SSL/TLS）加密的应用程序流。SSL/TLS 是一种基础密码协议套件，支持通过互联网进行安全通信。指纹可以是任何有助于识别特定流量类别的显著特征。在 Korczynski 和 Duda (2014年) 中，指纹对应于一个反映 SSL/TLS 会话动态的一阶齐次马尔可夫链。它们还可用于揭示试图利用 SSL/TLS 协议、通过与服务器建立异常通信来实施入侵（安全漏洞）的行为。

总而言之，我们认为，我们的模型或许有助于在单一模型中分析黑客的入侵行为：黑客的马尔可夫计划和秘密活动未知（隐马尔可夫模型类型），而其攻击的系统具有已知的马尔可夫指纹。在黑客攻击期间，可以认为指纹的性质发生了改变（偏离通常观测到的状态），并被另一个假定为独立同分布的信号所取代（出于技术方面的考虑）。本文不讨论这类问题的实际处理方法，但我们欢迎从业者在适用时使用自己的数据测试我们的方法；具体实现请参见第6 节和公式（22），该实现基于来自指纹源（金标准）的训练数据。

在这一设定下，我们研究模型的可识别性：值得注意的是，金标准过程随时间的依赖性对于确保可识别性至关重要。我们提出一种半参数最小对比方法，基于观测过程比较中适当的一阶和二阶经验累积分布函数，并证明在非常温和的技术假设下，我们的估计量强相合。此外，我们证明，在满足某些混合性质时，对潜在马尔可夫链转移矩阵以及未知污染序列累积分布函数的估计具有$\sqrt{n}$-一致性。模型详见第2节；可识别性以及参数部分（潜在马尔可夫链的转移）的相合性结果见第3节。渐近正态性的证明见第4节，未知独立同分布序列的非参数累积分布函数估计见第5节。最后，第6节展示了我们的方法在一组具有挑战性的子模型上的数值实现。结论见第7 节，证明汇集于第8节。

2 随机过程的时间序列混合

本节将依次介绍 H4M 和所谓的金标准过程毒化模型（即我们关注的模型）。据我们所知，这两个模型都具有一个独特特点：它们按照一个未观测到的平稳马尔可夫选择过程，在时间上混合多个相互独立的随机过程片段。例如，隐马尔可夫模型（HMM）只混合在给定潜在马尔可夫链的条件下相互独立的观测值。这种特定结构使得解决 HMM 的三个标准问题更为容易，即似然计算、解码和参数学习；借助前向、Viterbi 和前向-后向算法，分别都能在线性计算时间内完成。更多详情请参见Capp´等人（2005年）。遗憾的是，对于马尔可夫过程的时间序列混合，无法推导出基于滤波和平滑方程的此类估计策略。这正是这一新型时间序列隐变量模型类别面临的主要挑战。

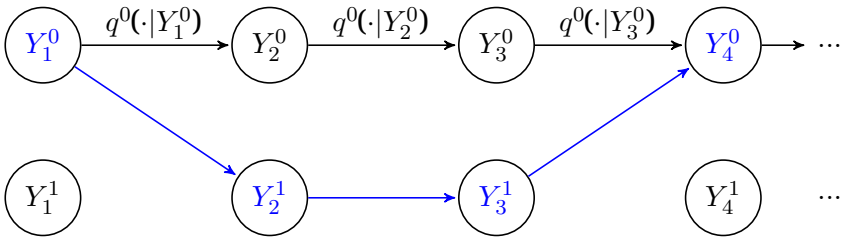


图1：马尔可夫过程与独立同分布序列的马尔可夫混合示意图。这里 $X_1 = 0$， $X_2 = 1$， $X_3 = 1$， $X_4 = 0$。

2.1 马尔可夫模型的隐马尔可夫混合（H4M）

在Vandekerkhove （2005年中，作者提出了一类新的缺失数据过程，称为马尔可夫过程的隐马尔可夫混合（H4M）。该模型最简单的版本由三个相互独立的平稳马尔可夫过程定义，分别记为 $X = (X_i)_{i\geq 1}$， $Y^0 = (Y_i^0)_{i\geq 1}$ 和 $Y^1 = (Y_i^1)_{i\geq 1}$。马尔可夫链 X 取值于 $\{0,1\}$，其未知转移矩阵为 Π，并采用自然参数化

$$\Pi_\theta = \begin{bmatrix} \Pi_\theta(0,0) & \Pi_\theta(0,1) \\ \Pi_\theta(1,0) & \Pi_\theta(1,1) \end{bmatrix} = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}, \tag{2}$$

参数化为

$$\theta = (\alpha, \beta) \in [\delta, 1-\delta]^2, \quad \text{with } 0 < \delta < 1/2.$$

参数 θ^* 的真实值满足 $\Pi_{\theta^*} = \Pi$。与 Π_θ 相关联的不变概率向量记为 $(\pi_\theta(0)\ \pi_\theta(1)) = (\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta})$。马尔可夫过程 Y^0 取值于可测状态空间 $(E, \mathcal{E})$，该空间配备有限支配测度 λ；相对于此参考测度，其转移核密度为 $q^0(\cdot|\cdot)$，通常未知（但在污染建模的视角下，可以假定它已知，详见下一节）。而马尔可夫过程 Y^1也取值于 $(E, \mathcal{E})$，其转移核 $q^1(\cdot|\cdot)$ 未知（如下一节所述，该核可简化为一个简单的非条件密度函数 f^1）。换言之， $\mathbb{P}(Y_{n+1}^j \in dv | Y_n^j = u) = q^j(v|u)d\lambda(v)$， $j = 0,1$。在这一设定中，观测过程 $Z = (Z_i)_{i\geq 1}$ 定义如下：

$$Z_i = \mathbb{1}_{\{X_i=0\}} Y_i^0 + \mathbb{1}_{\{X_i=1\}} Y_i^1, \qquad i \geq 1. \tag{3}$$

我们在图 1 中展示了一个简单情形：马尔可夫过程 Y^0 在时刻 $t = 1$ 和 4 被观测到，并在时间 Y^1 期间被独立同分布过程 $t = 2$ 和 3 绕过。

正如 Vandekerkhove（2005，引理1的证明）所研究的，只要过程 Y^0 和 Y^1 本身是几何α-混合的，模型（3）就是平稳的，并且是几何 α-混合的。基于来自 Z 的一条 n-轨迹（Z_1，...， Z_n），基本的统计挑战是在较弱的技术条件下，提出并研究一种统计方法，以恢复真实的转移矩阵 Π_{θ^*} 以及未知的转移核 $q^j(\cdot|\cdot)$、 $j = 0$， 1，其中假定这些转移核属于参数族。鉴于 H4M 的完全似然难以处理，参见 Vandekerkhove（2005）中的表达式（8），作者考虑了 Ryd´en（1994）针对隐马尔可夫模型提出的一种思想，即基于更易处理的 m-分割似然，其中 m 为分割长度。Ryd´en（1994）证明，在标准技术

条件（可识别性、遍历性、正则性等）。在 Vandekerkhove(2005年，第2节）中，作者给出了一组条件（C1–7），在这些条件下，适用于 H4M 的最小距离最小似然估计同样是强一致且渐近正态的，参见 Vandekerkhove(2005年，定理1和定理2）。在该论文的第4节中，还提出了一种蒙特卡罗方法，用于计算一般转移密度核 $f(z)$ 的平稳概率密度函数 $q(\cdot|\cdot)$。请注意，为了计算并最大化涉及 m-分割似然（其中包含 $f^j(z)$ 的似然项）至关重要的是完成这一步；这些似然项通常在解析上是未知的（高斯 AR(1) 过程除外），并且出现在每个局部 m-似然的开头。

2.2 金标准过程污染

为保持一般性，本节不一定假设 Y^0 和 Y^1过程相对于某个参考测度具有概率密度函数（pdf）。我们将改用过程的累积分布函数来描述其随机特征。因此，从现在起，假设空间 E 是 $\mathbb{R}$的子集。说明这一点后，下面定义我们的金标准过程污染模型。我们可以通过指出它与 H4M 的两个显著差异来简单描述该模型：i）所谓的金标准过程 Y^0不再是马尔可夫过程，而更一般地说，是一个一阶和二阶分布均可完全获取的平稳混合过程；ii）所谓的污染过程 Y^1 不再是马尔可夫的，而是由一组具有未知共同累积分布函数的独立同分布随机变量组成，这组变量称为污染序列 F^1。作为示例，可参见图1 ，并假设 Y^0 过程的分布已知，但不再由马尔可夫动态驱动（省略转移 $q^0(\cdot|\cdot)$ ）。我们的目标是根据上述模型的样本（Z_1，...， Z_n），估计底层马尔可夫链 Π_θ 的转移 X ，以及 F^1 的未知累积分布函数 Y^1。我们的统计方法主要基于 Y^0 和 Y^1 过程的一阶和二阶累积分布函数。因此，我们将考虑以下分布函数

$$F^j(x) = \mathbb{P}(Y_i^j \leq x), \quad j = 0, 1 \quad \text{and} \quad G^0(x, y) = \mathbb{P}(Y_i^0 \leq x, Y_{i+1}^0 \leq y),$$

对所有 $i \geq 1$ 以及 $(x, y) \in E^2$。现在，我们定义观测过程 Z （3）中按上述设定给出的其一阶和二阶分布。

一阶分布。记 $F(\cdot)$ 为过程 Z 的平稳累积分布函数，其定义为

$$\begin{aligned} F(x) &= \pi(0)F^0(x) + \pi(1)F^1(x) \\ &= \frac{\beta^*}{\alpha^* + \beta^*}F^0(x) + \frac{\alpha^*}{\alpha^* + \beta^*}F^1(x), \end{aligned} \tag{4}$$

对所有 $x \in E$。为简化后续计算，记

$$p^* = \pi(0) = \frac{\beta^*}{\alpha^* + \beta^*}, \quad \text{and} \quad r^* = \pi(1) = \frac{\alpha^*}{\alpha^* + \beta^*} = 1 - p^*.$$

因此， X 的平稳分布向量为（p^*， r^*）^{T} 且

$$F(x) = p^*F^0(x) + r^*F^1(x),$$

对于所有 $x \in E$，其中 F 是 Z 的累积分布函数。由（4）我们可以根据对 F（·）的了解（可观测性）构建一个伪参数累积分布函数族，即：

$$\mathcal{F} = \left\{F_\theta^1,\ \theta = (\alpha, \beta) \in \Theta\right\}, \tag{5}$$

其中

$$F_\theta^1(x) = \frac{1}{r}\left(F(x) - pF^0(x)\right), \quad p = \frac{\beta}{\alpha+\beta}, \quad r = \frac{\alpha}{\alpha+\beta} = 1-p, \tag{6}$$

参数空间记为并定义为 $\Theta := [\delta, 1-\delta]^2$，其中 $\delta \in]0,\ 1/2[$。由（4）可知，如果我们在上述参数化函数（6）中考虑 $\theta = \theta^*$，便可直接得到 $F_{\theta^*}^1 = F^1$。

二阶分布。记 $G(\cdot,\ \cdot)$为过程Z 的二阶平稳累积分布函数，其定义为

$$\begin{aligned} G(x,y) \ &= \ \pi(0)\Pi(0,0)G^0(x,y) + \pi(0)\Pi(0,1)F^0(x)F^1(y) \\ &\quad +\pi(1)\Pi(1,0)F^1(x)F^0(y) + \pi(1)\Pi(1,1)F^1(x)F^1(y), \end{aligned}$$

对于所有（$x,\ y$）$\in E^2$。现在记

$$\lambda_1 = \frac{\beta(1-\alpha)}{\alpha+\beta}, \quad \lambda_2 = \lambda_3 = \frac{\alpha\beta}{\alpha+\beta}, \quad \lambda_4 = 1-\lambda_1-2\lambda_2 = \frac{\alpha(1-\beta)}{\alpha+\beta},$$

这自然促使我们考虑这些参数所对应的真实版本（当$\theta = \theta^*$时）

$$\lambda_1^* = \pi(0)\Pi(0,0), \quad \lambda_2^* = \pi(0)\Pi(0,1), \quad \lambda_3^* = \pi(1)\Pi(1,0), \quad \lambda_4^* = \pi(1)\Pi(1,1).$$

注意，$\lambda_1^* + \lambda_2^* = p^*$ 且 $\lambda_3^* + \lambda_4^* = r^*$。通过这一重新参数化，我们得到过程 Z 二阶累积分布函数一个更易解释且更易处理的表示：

$$G = \lambda_1^* G^0 + \lambda_2^* F^0 \otimes F^1 + \lambda_3^* F^1 \otimes F^0 + \lambda_4^* F^1 \otimes F^1 \tag{7}$$

采用记号（$\phi \otimes \psi$）（$x,\ y$）$= \phi$（x）ψ（y），对于所有（$x,\ y$）$\in E^2$。

3 可识别性与参数估计

现在，我们可以提出两个差异函数，分别记为 $\mathbf{d}(\cdot)$ 和 $\mathbf{s}(\cdot)$，用于估计问题中的参数部分。事实上，考虑所谓的（θ^*, θ）-偏差量，它将 Z 的真实二阶分布与在参数 θ 下、使用其自然的（7）重构进行比较，其中 F_θ^1 取自参数族 $\mathcal{F}$，见由下式定义的表达式（5）：

$$\Delta(\theta,x,y) = G(x,y) - \underbrace{\left[\lambda_1 G^0(x,y) + \lambda_2 F^0(x)F_\theta^1(y) + \lambda_3 F_\theta^1(x)F^0(y) + \lambda_4 F_\theta^1(x)F_\theta^1(y)\right]}_{\text{parametric } \theta\text{-reconstruction of } G}, \tag{8}$$

对于给定的有限权重测度 dH，考虑

$$\mathbf{d}(\theta) = \iint \Delta^2(\theta,x,y)dH(x,y), \quad \mathbf{s}(\theta) = \sup_{(x,y)\in E^2} |\Delta(\theta,x,y)|, \tag{9}$$

其中，$\iint$ 符号表示 $\int_{E\times E}$。为简单起见且不失一般性，我们假设 $\iint dH(x,y) = 1$。

从现在起，我们将重点研究差异函数$\mathbf{d}(\cdot)$，但$\mathbf{s}(\cdot)$基本上会继承相同的一致性性质，除了 $\sqrt{n}$-收敛性之外；后者要求具备一些光滑性，而这些光滑性遗憾地并不适用于$\mathbf{s}(\cdot)$。我们可以观察到，$\mathbf{d}(\cdot)$是一个满足以下性质的非负函数：

若 $\theta = \theta^*$ 则 $\Delta(\theta, x, y) = 0$ 对于几乎所有的$(x, y) \in E^2$都成立，并且 $\mathbf{d}(\theta^*) = 0 = \mathbf{s}(\theta^*)$.

该 f 以下命题保证了在一定条件下逆命题成立 .

命题 1. 命题该(θ^*, θ)-偏差量定义于式（*8*）中，可表示如下

$$\Delta(\theta, \cdot) = c_1(\theta^*, \theta)(G^0 - F^0 \otimes F^0) + c_2(\theta^*, \theta)(F^1 - F^0) \otimes (F^1 - F^0) \tag{10}$$

其中 c_1 和 c_2 仅取决于 θ 和 θ^*，闭式表达式见式（*26*）。此外，我们有如下推论

$$\{c_1(\theta^*, \theta) = 0 \quad and \quad c_2(\theta^*, \theta) = 0\} \Longrightarrow \theta = \theta^*.$$

该命题在第8.1节中证明。这引出了基本可识别性条件。

LinInd 假设：族 $\{G^0 - F^0 \otimes F^0, (F^1 - F^0) \otimes (F^1 - F^0)\}$ 在以下意义下线性无关：对于所有满足 $\iint_\Gamma dH = 1$ 的集合 Γ，限制在 Γ 上的两个函数线性无关。

在混合模型 $pF^0 + (1-p)F^1$ 中，自然可以假设 $F^1 \neq F^0$ 以确保可识别性。$G^0 = F^0 \otimes F^0$ 对应于独立过程 Y^0 的情况（此时 Z 是隐马尔可夫链）。请注意，本研究不考虑这种独立情形。在 **LinInd** 假设下，参数值 θ^* 是 **d**(·) 的唯一极小值，即 *i.e.* 我们有：

$$\mathbf{d}(\theta) = 0 \Longleftrightarrow \theta = \theta^*. \tag{11}$$

考虑上确界差异 **s**(·) 时，只需假设其在基本意义下线性无关。我们直接得到

$$\mathbf{s}(\theta) = 0 \Longleftrightarrow \theta = \theta^*.$$

对于积分差异 **d**(·)，我们必须确保 $\iint \Delta^2 dH = 0$ 蕴含 $\Delta = 0$，因此我们的假设需要在 H-测度为 1 的集合上成立。或者，我们可以假设其满足基本意义下的线性独立性，并额外假设 F^1 和 G^0 是连续的，且 dH 相对于勒贝格测度具有连续密度。

由于 Δ 仅取决于已知量 F^0，G^0 以及观测值的分布，等价关系 $\Delta(\theta, \cdot) = 0 \Longleftrightarrow \theta = \theta^*$ 使我们能够写出以下结果。

命题 2。 在 **LinInd** 条件下，θ^* 以及随后 F^1 是可辨识的。

一个已经观察到的现象是，潜变量之间的依赖性能够使一些在依赖情形下不可识别的混合模型具有可识别性，参见 Gassiat (2019年)。非正式地说，与独立性相比，依赖性能够存储更多信息（隐式可辨识性约束）。我们的研究与这些结果一致，而这些结果同样需要线性独立性假设。

现在可以提出对比函数**d** 和**s** 的经验版本

$$\mathbf{d}_n(\theta) = \iint \Delta_n^2(\theta, x, y) dH(x, y), \quad \mathbf{s}_n(\theta) = \sup_{(x,y) \in E^2} |\Delta_n(\theta, x, y)|, \tag{12}$$

wh对于所有（x， y）$\in E^2$，经验（θ^*， θ）偏差版本的（8）定义为 y

$$\Delta_n(\theta,x,y)=\hat{G}_n(x,y)-\left[\lambda_1 G^0(x,y)-\lambda_2 F^0(x)\hat{F}^1_{n,\theta}(y)-\lambda_3\hat{F}^1_{n,\theta}(x)F^0(y)-\lambda_4\hat{F}^1_{n,\theta}(x)\hat{F}^1_{n,\theta}(y)\right],$$

其中

$$\hat{F}^1_{n,\theta}(x)=\frac{1}{r}\left(\hat{F}_n(x)-pF^0(x)\right),\quad x\in E, \tag{13}$$

使用标准的一阶和二阶经验累积分布函数

$$\hat{F}_n(x)=\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{Z_i\le x\}},\quad \hat{G}_n(x,y)=\frac{1}{n-1}\sum_{i=1}^{n-1}\mathbb{1}_{\{Z_i\le x,Z_{i+1}\le y\}}, \tag{14}$$

对所有 $(x,\ y)\in\mathbb{R}^2$。最后，我们考虑 θ 的两种可能的参数估计量：

$$\hat{\theta}_n=\underset{\theta\in\Theta}{\operatorname{argmin}}\,\mathbf{d}_n(\theta),\quad \text{and}\quad \tilde{\theta}_n=\underset{\theta\in\Theta}{\operatorname{argmin}}\,\mathbf{s}_n(\theta). \tag{15}$$

定理 1。假设序列 $(Z_i)_{i\ge1}$ 是平稳的，且假设 **LinInd** 成立，则我们在*(15)*中定义的两个估计量都是强一致的，即

$$\hat{\theta}_n\xrightarrow{a.s.}\theta^*\quad and\quad \tilde{\theta}_n\xrightarrow{a.s.}\theta^*,\qquad as\qquad n\to+\infty.$$

ˆ 证明。定理 1 中 θ_n 的收敛性利用最小对比估计量理论中的经典结果得证，参见 van der Vaart（1998年）或 Dacunha-Castelle 和 Duflo（1986年）；这里使用了等价关系（11）、引理 1（d(·) 和 $\mathrm{d}_n(\cdot)$ 的正则性，见第 8.2 节）以及命题 3 和 4（$\mathrm{d}_n(\cdot)$ 一致收敛于 d(·)，见第 8.2 节）。对于 θ_n，同样的˜类型的证明也成立。 □

注意，我们也可以用特征函数而非分布函数来定义对比函数。代数计算以及命题 1 都相同。然而，要证明一致性需要格利文科-坎泰利定理，而特征函数并不存在以同样方式成立的该定理。记 c_n 为经验特征函数， $c=\mathbb{E}(c_n)$ 为真实特征函数。关于 c_n 趋向 c 的收敛性并不在整个实数轴上成立，参见 Feuerverger 和 Mureika (1977)。此外，过程 $\sqrt{n}(c_n-c)$ 仅在特定假设下才收敛，参见 Csorgo (1981)；而要证明渐近正态性，必须依赖这一收敛性，我们将在下一节对此进行研究。

4 渐近正态性

ˆ 为了建立与$\mathbf{d}(\cdot)$ 相关的最小对比估计量 θ_n 的 $\sqrt{n}$-一致性，我们需要对随机过程 X 和 Y^0引入一个额外的平稳性/混合条件；这两个过程在模型中均由独立同分布随机变量构成。˜ 对于任意平稳过程 Y 以及所有 $(t,k)\in\mathbb{N}^*\times\mathbb{N}$，定义与随机过程 Y 相关的α-混合系数序列如下：

$$\alpha^{\tilde{Y}}(k)=\sup_{A\in\mathcal{F}^{t}_{\tilde{Y},1},B\in\mathcal{F}^{\infty}_{\tilde{Y},t+k+1}}|\mathbb{P}(A\cap B)-\mathbb{P}(A)\mathbb{P}(B)|,\quad k\ge1, \tag{16}$$

其中 $\mathcal{F}^{t_2}_{\tilde{Y},\ t_1}$ 表示对所有 $t_2>t_1\ge1$，由（$\tilde{Y}_{t_1}$，.....，$\tilde{Y}_{t_2}$）生成的 σ-代数。

混合假设：随机过程 X 和 Y^0 是严格平稳的，且 Y^0 是α-混合的，其 $\alpha^{Y^0}(k)=O(k^{-a})$ 对于 $a>1$成立。

定理 2. 假设 F 连续。另假设 **LinInd** 和 **混合条件**成立。假设 θ^*是 $\Theta = [\delta, 1-\delta]^2$的内点。那么我们得到以下中心极限定理行为

$$\sqrt{n}\left(\hat{\theta}_n - \theta^*\right) \xrightarrow{d} \mathcal{N}(0, \Sigma), \quad as \quad n \to +\infty,$$

其中协方差矩阵 Σ 详见第 *8.3.3*节。

该定理证明中的一个关键步骤，是对过程 Z 使用经验中心极限定理，并结合二元过程（Z_i， Z_{i+1}）$_{i\geq 1}$。假设 F 连续，以便使用 Rio (2017）的中心极限定理结果；该结果还要求满足 α-混合条件。注意， α-混合条件是最弱的混合条件，因为其他所有混合条件（β， ρ，ϕ，等等）都蕴含 α-混合条件。此外，所要求的 α-混合率下降得相当缓慢（关于 n呈多项式级）。

这里提一下弱依赖这一较新的概念，它涵盖了比混合条件更弱的依赖条件。针对弱依赖过程，已有研究证明了经验中心极限定理，通常涉及 α、 β 和 θ 弱依赖；参见 Dedecker 等（2007），其中还给出了此类弱依赖序列的例子。然而， Y^0 弱依赖 $\Rightarrow$（Z_i、 Z_{i+1}）$_{i\geq 1}$ 弱依赖之间的蕴含关系很难详尽阐述，尤其是˜对于“未来两点”系数 $\tilde{\alpha}_2$、 β_2、 θ_2而言。此外，衰减条件限制更严格：根据系数不同，要求其为 n^{-a} ，且满足 $a > 4$ 或 $a > 8$ （Dedecker（2010）提出的、条件更优的结果仅针对单变量序列，且要求 $a > 1$ ）。

5 非参数估计

模型的参数部分估计完成后，可以F^1 使用经典的反演公式恢复非参数部分。事实上，根据表达式（13），我们可以考虑以下代入估计量来估计非参数部分

$$\hat{F}_n^1(x) = \hat{F}^1_{n,\hat{\theta}_n}(x) = \frac{1}{\hat{r}_n}\left(\hat{F}_n(x) - \hat{p}_n F^0(x)\right), \quad x \in E, \tag{17}$$

ˆ ˆ，其中 ˆ$r_n = 1 - \hat{p}_n = \hat{\alpha}_n/(\hat{\alpha}_n + \beta_n)$。由于 $\theta_n \to \theta^*$几乎必然成立（当 $n \to +\infty$时），我们可以预期 ˆ $F_n^1 = \hat{F}^1_{n,\hat{\theta}_n}$将以经典收敛速率依分布收敛到 $F^1_{\theta^*} = F^1$。这一点由以下两个定理表述。

定理 3。 假设 F 连续。还假设 **LinInd** 和 **混合条件** 成立，且 θ^* 是 Θ的内点。那么在空间 $B(\mathbb{R}^2)$ 中，我们有关于 $\mathbb{R}^2$上实值有界函数的依分布收敛

$$\sqrt{n}\left(\hat{F}_n^1 - F^1\right) \rightsquigarrow \mathcal{G}, \quad as \quad n \to +\infty, \tag{18}$$

其中 $\mathcal{G}$ 是零均值高斯过程。

定理4。 假设 F 连续。还假设满足 **LinInd** 和 **混合条件**，其中 $a > 4$，且 θ^* 是 Θ的内点。那么存在一个正的常数 C， 它取决于θ^*、 G^0、 F^1、 δ，使得

$$\mathbb{E}\|\hat{\theta}_n - \theta^*\| \leq \frac{C}{\sqrt{n}},$$

其中 $\|\cdot\|$ 为欧几里得范数，并且

$$\mathbb{E}\|\hat{F}_n^1 - F^1\|_\infty \leq \frac{C}{\sqrt{n}}.$$

这一结果并不是前一个结果的直接推论，而是需要进行特定处理，参见第 8.5 节中的证明。利用 Dedecker等人（2014年）的结果，混合条件可以替换为： Y^0 是具有 $\beta_n = O(n^{-a})$ 的 β-混合过程，其中 $a > 1$。我们还可以证明，对于所有整数 $q \geq 1$， $\mathbb{E}(\|\hat{}\theta_n - \theta^*\|^q) \leq C(q)n^{-q/2}$ 以及 $\mathbb{E}(\|\hat{} F_n^1 - F^1\|_\infty^q) \leq C(q)n^{-q/2}$。

一个值得探索、但超出本文范围的研究方向，是对 Y^1 序列密度 f^1进行非参数估计。估计 f^1的自然候选方法是以下插件半参数估计量

$$\hat{f}_n^1(x) = \frac{1}{\hat{r}_n}\left(\hat{f}_n(x) - \hat{p}_n f_0(x)\right), \quad \text{where} \quad \hat{p}_n = \frac{\hat{\beta}_n}{\hat{\alpha}_n + \hat{\beta}_n} \quad \text{and} \quad \hat{r}_n = \frac{\hat{\alpha}_n}{\hat{\alpha}_n + \hat{\beta}_n} \tag{19}$$

以及$\hat{} f_n(x) = (nh_n)^{-1}\sum_{i=1}^n K_{h_n}(x - Z_i)$，其中 $K(\cdot)$是一个概率核， $K_h(\cdot) = h^{-1}K(\cdot/h)$，对所有 $h > 0$均成立； h_n 是带宽参数，当 n 趋于无穷时趋于零。在这种设定下，我们确实可以研究局部解码问题，也就是预测观测值 Z_i 是否受到污染（并非由金标准生成）。事实上，对于 $i \geq 1$，与观测到的二元组(Z_{2i}, Z_{2i+1})相关联的马尔可夫潜在二元组(X_{2i}, X_{2i+1})的条件分布，对所有$(k, l) \in \{0,1\}^2$ 以及所有$(z, z') \in E^2$，由下式给出

$$\mathbb{P}_{\theta,f}(X_{2i} = k, X_{2i+1} = l | Z_{2i} = z, Z_{2i+1} = z') := \frac{\eta_{\theta,f^1}(k, l; z, z')}{\mathbb{P}_{\theta,f}(Z_{2i} = z, Z_{2i+1} = z')},$$

其中

$$\mathbb{P}_{\theta,f}(Z_{2i} = z, Z_{2i} = z') = \lambda_1 g^0(z, z') + \lambda_2 f^0(z) f(z') + \lambda_3 f(z) f^0(z') + \lambda_4 f(z) f(z'),$$

and

$$\begin{aligned}
\eta_{\theta,f}(0,0; z, z') &= \lambda_1 g^0(z, z'), \quad \eta_{\theta,f}(0,1; z, z') = \lambda_2 f^0(z) f(z'), \\
\eta_{\theta,f}(1,0; z, z') &= \lambda_3 f(z) f^0(z'), \quad \eta_{\theta,f}(1,1; z, z') = \lambda_4 f(z) f(z').
\end{aligned}$$

$\hat{}$ 假设现在我们运行估计方法，得到 θ_n 的估计量 θ^* ，以及 f^1 的密度估计量 $\hat{} f_n^1$，那么对于所有 $(k, l) \in \{0,1\}^2$ 和所有 $(z, z') \in E^2$，我们可以轻松估计真实的解码概率 $\mathbb{P}_{\theta^*, f^1}(X_{2i} = k, X_{2i+1} = l | Z_{2i} = z, Z_{2i+1} = z')$，方法如下：

$$\text{predict}_i((k,l)|z_1^n) := \mathbb{P}_{\hat{\theta}_n, \hat{f}_n^1}(X_{2i} = k, X_{2i+1} = l | Z_{2i} = z, Z_{2i+1} = z'). \tag{20}$$

请注意，我们也可以通过推导相应的条件概率来预测长度大于二的序列，其思路与 (20) 及后续内容相同（显然，计算量更大）。

6 数值性能

本节旨在考察我们针对模型（3）开发的半参数估计方法在不同情形下的数值性能，这些情形对相关过程的可观测性和混合性质的影响程度各不相同。为此，我们考虑以下模型

$$Z_j = (1 - X_j)Y_j^0 + X_j Y_j^1, \quad j = 1, \ldots, n \tag{21}$$

其中， $k = 1, \ldots\ldots, n-1$ 以及一个高斯噪声独立同分布序列（ε_k）$_{k\geq 1}$，该序列服从$\mathcal{N}$（m_0， v_0^2）分布：

$$\begin{array}{rcl} X_{1:n} & \sim & \text{Markov}(n, \Pi_\theta, \pi_\theta): \ X_1 \sim \pi_\theta, \mathbb{P}_\theta(X_{k+1} = \cdot | X_k = \cdot) = \Pi_\theta(\cdot, \cdot) \\ Y^0_{1:n} & \sim & \text{AR}(1)(n, \varphi, m_0, v_0): Y^0_1 \sim \mathcal{N}(\mu_0, \text{var}_0), \ Y^0_{k+1} = \varphi Y^0_k + \varepsilon_k \\ Y^1_{1:n} & \sim & \mathcal{N}^{\otimes n}(m, v^2). \end{array}$$

参数 $\varphi \in]0,1[$, 是 AR(1)过程的回归系数，其中 $(\mu_0, \text{var}_0) = \left(\frac{m_0}{1-\varphi}, \frac{v_0^2}{1-\varphi^2}\right)$ 分别是该过程的平稳（边际）均值和方差Y_0。为简便起见，在所有模拟中我们均取 $(m_0, v_0^2) = (1,1)$。为了实际估计/计算已知金标准过程对应的一阶和二阶累积分布函数 $F^0(\cdot)$ 和 $G^0(\cdot,\cdot)$（当没有闭式表达式时，我们就需要这样做），我们生成两个大小为 N 的独立同分布样本

$$\begin{array}{l} Y^0[i] \sim \mathcal{N}(\mu_0, \text{var}_0) \\ (Y^0_1, Y^0_2)[i] \quad \text{where} \quad Y^0_1[i] \sim \mathcal{N}(\mu_0, \text{var}_0) \quad \text{and } Y^0_2[i] = \varphi Y^0_1[i] + \varepsilon[i], \end{array}$$

对于 $i = 1, \ldots, N$，其中 $(\varepsilon[i])_{1\leq i\leq N}$是从 $\mathcal{N}(m_0, v_0^2)$ 分布中抽取的独立同分布序列。根据这些样本，我们计算

$$\hat{F}^0_N(x) = \frac{1}{N}\sum_{i=1}^N \mathbb{I}_{\{Y^0[i]\leq x\}}, \quad \hat{G}^0_N(x,y) = \frac{1}{N}\sum_{i=1}^N \mathbb{I}_{\{(Y^0_1, Y^0_2)[i]\leq (x,y)\}}, \quad (x,y) \in E^2, \tag{22}$$

它们是一阶自回归模型的一阶和二阶累积分布函数的一致强相合估计量 Y^0。请注意，对于 N 相对于 n 足够大的值，我们能够将 n经验对比量（定义见（12））中涉及的随机波动控制在令人满意的精度范围内。在模拟中，我们采用了 $N = 2n$ ，计算时间和准确性方面的结果都令人满意。模拟中考虑的 H 分布为 $\mathcal{U}^{\otimes 2}$，其中 $\mathcal{U}$表示在 [最小值$_{1\leq i\leq n}(Z_i)$与最大值$_{1\leq i\leq n}(Z_i)$] 上的均匀分布，以尽可能采用最中性的权函数。请注意，在 $\mathbb{R}$ 上使用任何重尾分布也能达到同样的效果。

6.1 两种强可观测性情形

我们首先考虑两种设定：污染输出具有强可观测性，其期望为 $2.5 \times \mu_0$，标准差为 $0.8 \times \sqrt{\text{var}_0}$（远高于 AR(1) 过程的边际期望，且标准差更小）；潜在马尔可夫链则具有弱或强的定性混合性质。

- (S0)$_{strong}$: $\theta = (0.7,\ 0.8)$, $\varphi = 0.7$和（m, v）$= (2.5 \times \mu_0,\ 0.8 \times \sqrt{\text{var}_0})$。
- (S0)$_{weak}$: $\theta = (0.3,\ 0.2)$, $\varphi = 0.7$和（m, v）$= (2.5 \times \mu_0,\ 0.8 \times \sqrt{\text{var}_0})$。

为便于说明，我们在图2 和3 中分别用红色显示对应于模型 (S0) 的轨迹$_{strong}$ 和 (S0)$_{weak}$。请注意，我们还用蓝色显示了一个额外的信息性虚拟过程，即 $X1 = X \times \mathbb{E}(Y^1)$，其中 X 表示未观测到的二状态马尔可夫链。请注意， $X1$ 不得与 X_1 马尔可夫过程的第一次观测值X混淆。这样，当 $\{X1 = 0\}$ 我们获知所观测的是已知的一阶自回归模型 Y^0时；相反，当 $\{X1 = m \neq 0\}$ 我们知道所观测的是以 Y^1 为中心的独立同分布过程时，

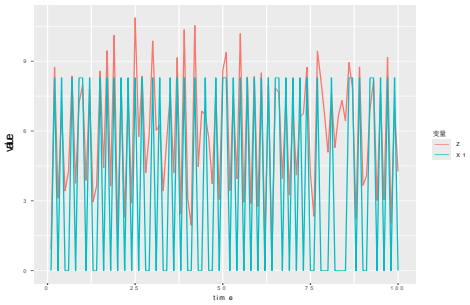


图2： (S0)$_{strong}$ 轨迹

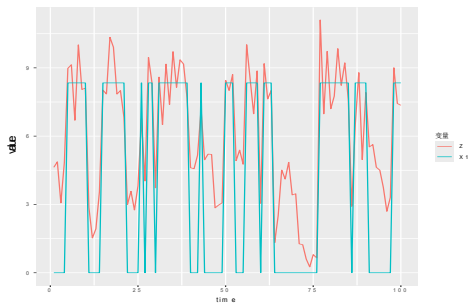


图3： (S0)$_{weak}$ 轨迹

$m = \mathbb{E}(Y^1)$。

在表 1中，我们给出了基于100次重复蒙特卡洛方案收集的偏差和标准差性能结果（除非明确 ˆ ˆ˜ ˜ 说明，本节中均一致采用该方案），这些结果对应于我们的估计量 $\theta_n = (\hat{\alpha}_n, \beta_n)$ 和 $\theta_n = (\tilde{\alpha}_n, \beta_n$，分别对应于差异度 $\mathbf{d}(\cdot)$（基于积分）和$\mathbf{s}(\cdot)$（基于上确界）。

最后，为了说明我们的估计量的双变量渐近高斯行为 $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ 我们在图 4 中展示了通过对模拟的 (S0)$_{strong}$ 模型进行 $\ell = 10{,}000$ 次估计所得到的以 $\sqrt{n}$ 为中心的估计量样本，即 $(\sqrt{n}[\hat{\theta}_{i,\,n} - \frac{1}{n}\sum_{j=1}^{\ell} \hat{\theta}_{j,\,n}])_{1 \le i \le \ell}$ 其中 $\hat{\theta}_{i,\,n}$ 表示基于 n 个观测值的第 i 次估计输出， $i = 1, \ldots, \ell$，观测值数量分别为 $n = 1{,}000$、$3{,}000$ 和 $5{,}000$。为简便起见，我们的估计统计量（偏差、方差或标准差）仅保留 4 位数字。

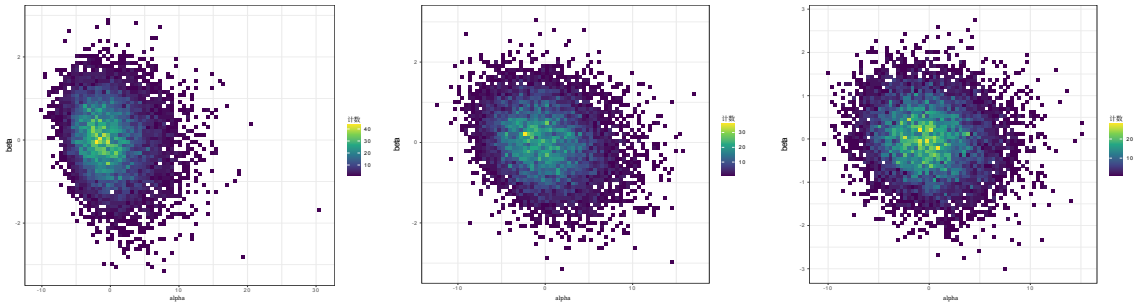
基于 $N = 10{,}000$ 次重复得到的估计协方差矩阵为：

$$\hat{\Sigma}_{1,000} = \begin{bmatrix} 14.111 & -0.543 \\ -0.543 & 0.623 \end{bmatrix}, \quad \hat{\Sigma}_{3,000} = \begin{bmatrix} 14.736 & -0.698 \\ -0.698 & 0.607 \end{bmatrix}, \quad \hat{\Sigma}_{5,000} = \begin{bmatrix} 11.709 & -0.334 \\ -0.334 & 0.536 \end{bmatrix}.$$

关于 (S0) 的评论$_{strong}$ **和 (S0)**$_{weak}$。首先请注意， α 和 β 越高，马尔可夫链从一个状态切换到另一个状态的频率就越高。因此，如图2和3所示，与 (S0)$_{weak}$ 模型相比，(S0)$_{strong}$ 模型中隐马尔可夫状态几乎可以通过视觉判断的时间段更长；相反，(S0)模型表现出高度不稳定，使得隐马尔可夫状态难以辨认（因此可能也难以估计）。我们在性能表1 中展示，这一直观

$(S0)_{strong}$	$\hat{\alpha}_n$	$\hat{\beta}_n$	$\tilde{\alpha}_n$	$\tilde{\beta}_n$
$n = 1{,}000$	(0.036,0.118)	(-0.004,0.023)	(0.017,0.127)	(-0.007,0.035)
$n = 3{,}000$	(0.034,0.056)	(-0.010,0.013)	(0.015,0.083)	(-0.005,0.019)
$n = 5,\ 000$	(-0.007,0.052)	(0.001,0.010)	(0.028,0.078)	(-0.002,0.013)
$(S0)_{weak}$	$\hat{\alpha}_n$	$\hat{\beta}_n$	$\tilde{\alpha}_n$	$\tilde{\beta}_n$
$n = 1{,}000$	(-0.218,0.116)	(-0.134,0.096)	(-0.048,0.057)	(-0.024,0.111)
$n = 3{,}000$	(-0.051,0.103)	(-0.010,0.114)	(-0.012,0.043)	(0.011,0.069)
$n = 5{,}000$	(-0.051,0.101)	(-0.011,0.112)	(0.002,0.033)	(0.015,0.065)

表1：模型 (S0)$_{strong}$ 和 (S0)$_{weak}$ 下，估计量 $\theta_n = (\hat{\alpha}_n, \beta_n)$ 和 $\theta_n = (\tilde{\alpha}_n, \beta_n)$ 的偏差及基于标准差的性能，分别对应差异度 $\mathbf{d}(\cdot)$（基于积分）和 $\mathbf{s}(\cdot)$（基于上确界），在 $n = 1{,}000$、 $n = 3{,}000$ 和 $n = 5{,}000$（按行排列）时，分别展示于最后两列。



ˆ 图4：模型 (S0)$_{strong}$ 下 $\sqrt{n}$ 居中的 θ_n 估计量的 N 个样本，其中重复 $N = 10,000$ 次，观测值数量分别为 $n = 1{,}000$、$3,000$ 和 $5,000$。

认为模型 (S0)$_{weak}$ 比 (S0)$_{strong}$ 更容易估计的观点其实并不完全准确。事实上，我们可以清楚地看到，就偏差和标准 ˆ ˜差而言，在 θ_n 和 θ_n上，整体而言 (S0)$_{strong}$ 下的表现优于 (S0)$_{weak}$ ，尤其是在样本量较大时。若进一步具体考察 ˆ ˜(S0)$_{strong}$ 下的结果，会发现估计量 θ_n 的表现优于 θ_n ；而在模型 (S0)$_{weak}$下，情况恰恰相反。为了深入分析估计量的渐近性质，我们在图11和图12中分别展示了附录部分所述、在 (S0)$_{strong}$和 (S0)$_{weak}$下估计量的经验 $\sqrt{n}$-分布（经中心化和标准化）。显然，在 (S0)$_{strong}$ 下，两种估计量大约从 $n = 1{,}000$ 开始达到中心极限定理的“钟形区间”；而在模型 (S0)$_{weak}$ 下，即使考虑到 $n = 5{,}000$，仍远未达到这一状态，尽管 θ 看起来稍微更接近高斯分布。

6.2 可观测性较低的挑战性情形

本节旨在进一步了解 **LinInd** 条件的敏感程度。事实上，根据 **LinInd** 条件可以看出，最不利的情形是：已知随机过程 Y^0 的平稳分布接近独立同分布序列 Y^1 的分布，且 Y^0 的时间依赖性较弱，即 接近独立设定（在一阶自回归模型情形下为 $\varphi \simeq 0$ ）。为此，我们定义了另外四种简单设定，以此对我们的估计方法进行挑战（同时仍能取得合理的渐近性能结果）：

- (S1)： $\theta = (0.2, 0.4)$， $\varphi = 0.7$ 和 $(m, v) = (1.5 \times \mu_0, 0.8 \times \sqrt{\text{var}_0})$。
- (S2)： $\theta = (0.2, 0.4)$， $\varphi = 0.7$ 以及 $(m, v) = (2 \times \mu_0, 0.8 \times \sqrt{\text{方差}_0})$ 。
- (S3)： $\theta = (0.6, 0.3)$， $\varphi = 0.7$ 以及 $(m, v) = (1.5 \times \mu_0, 0.8 \times \sqrt{\text{方差}_0})$ 。
- (S4)： $\theta = (0.6, 0.3)$， $\varphi = 0.5$ 以及 $(m, v) = (2 \times \mu_0, 0.8 \times \sqrt{\text{方差}_0})$ 。

为便于说明，我们在图5 至8中展示分别对应于模型（S1）和（S4）的轨迹。

关于模型（S1–4）的评论。由于模型（S1）和模型（S2）的转移值相同且取值不太大，我们可以从图 5 和6中看到，有时源 Y^0 和噪声序列 Y^1 会持续很长时间处于可观测状态，同时源的可分离性 Y^0 和 Y^1有时也较好。请注意，我们提高了模型（S2）的可观测性

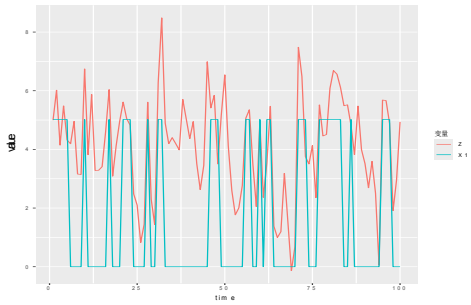


图5：（S1）轨迹

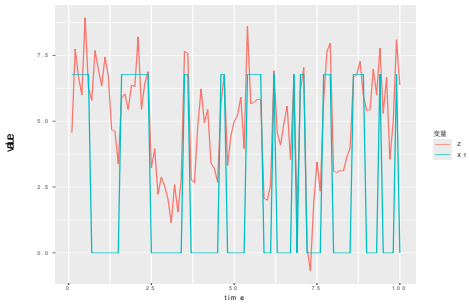


图6：（S2）轨迹

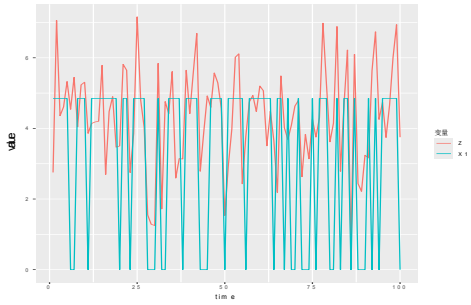


图7：（S3）轨迹

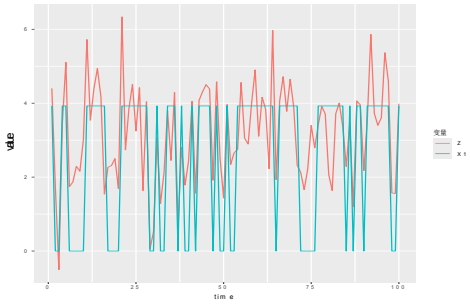


图8：（S4）轨迹

相较于（S1），通过将 Y^1 的期望从 5 提高到约 6.66。具体而言，在（S1）下，我们有 $(\mu_0, m) \sim (3.33, 5)$；而在（S2）下，我们有 $(\mu_0, m) \sim (3.33, 6.66)$，期望差距的正向跳跃 $m - \mu_0$ 从约 1.66 增至 2.33。在模型（S3）中，我们仅通过取更高的转移值来提高跳跃频率，同时保持与模型（S1）完全相同的 AR(1)动态和可观测性。我们可以在图7中清楚地看到这一变化的影响，图中所示的模式明显更加不规则。在模型（S4）中，我们保持与（S3）相同的转移值，但将回归系数从 0.7 降至 0.5，这也使噪声过程 Y^1 的期望从 5 降至 4。需要强调的是，在（S3）下，我们有 $(\mu_0, m) \sim (3.33, 5)$；而在（S4）下，我们有 $(\mu_0, m) \sim (2, 4)$，期望差距从约 1.66 跳升至 2，直觉上这应该会使估计更容易（我们将通过性能模拟结果表明，事实未必如此）。以上是一些初步评论，下面我们来讨论表2中汇总的性能结果。

正如预期，在模型（S1）和（S2）之间提高跳跃的可观测性，对我们的估计量产生了显著影响：与（S1）相比，（S2）下的方差几乎减半。与研究初步模型（S0）$strong$ 和（S0）$weak$时的情况类似，我们也可以预期（S3）的表现会明显优于（S1）。对于较大的 n，确实如此（当 $n = 5{,}000$时，这一点并不明显）；在考虑ˆ ˜估计量 θ时尤其如此，而基于上确界的估计量 θ 则较难在（S3）下取得优于（S1）的表现。最后，乍看之下，模型（S4）似乎具有有利的可观测性差距，但使用 θ_n,˜时估计效果相当差，这一点十分令人惊讶；而估计量θ_n 的性能则与（S3）中的结果更加接近。这或许可以解释为，（S4）下的回归系数 φ 较弱，使得时间上相邻两个数据点之间的联系不那么明显（从 AR-ˆ ˜动态转变为完全独立）。 θ 与 θ 之间的行为差异，可能源于变化过于细微（不确定性的处理反映在对比函数中），具体体现在

项 $\Delta_n(\theta, x, y)$ 中；该项参与经验对比函数（12），并且 θ 在 Θ上移动时， $\mathbb{R}^2$ 也会随之变化。我们认为， θ^* 中的某些奇异性（轻微的全局极值）或许更容易由局部的（面向奇异性的） θ 类型估计量˜捕捉到，而非通过平滑但全局的基于上确界ˆΔ_n对比函数，例如 θ。为了进一步深入分析估计量的渐近性质，我们还在附录部分的图13–16中展示了经$\sqrt{n}$标准化的估计量ˆ ˜ θ 和 θ的分布。结合估计量在（S1）下性能较差这一情况，我们可以从图13观察到，这也意味着当样本量为 $n = 5,000$或$10,000$时，估计量尚未达到渐近正态性阶段。不过，当样本量为 $n = 20,000$时，我们可以观察到基于积分的估计量 θ 呈现钟形分布，这表明即使在具有挑战性的设置下，定理 3所确立的中心极限定理仍然成立。相较之下，ˆ ˜经$\sqrt{n}$标准化的 θ 和 θ 估计量的分布，在模型（S2）下从 $n = 5,000$（该研究考察的较小样本量）开始，在模型（S3）下从 $n = 10,000$开始，均呈现钟形曲线。最后，同样结合估计量在（S4）下性能较差这一情况，对于$n = 5,000$或$10,000$的样本量，显然尚未达到渐近正态性阶段，而且不同估计量的表现大相径庭： θ 对应的˜呈现双峰性（这是对比函数存在伪极小值的迹象），而 θ 则呈现平坦性（精度不足）。ˆ 不过，当样本量为 $n = 20,000$时，结果稍显乐观，尤其是 θ; 事实证明，它具有渐近正态性，见定理 3。

6.3 函数型估计量的行为

本节旨在展示我们在（17）中定义的基于代入反演的函数估计量的渐近行为。为此，我们在图9中展示了10个逆累积分布函数的部分图形，这些函数是通过初步参数估计步骤推导得到的，其中 $(\hat{r}_n, \hat{p}_n) = \left(\hat{\alpha}_n\hat{\alpha}_n + \hat{\beta}_n, \hat{\beta}_n\hat{\alpha}_n + \hat{\beta}_n\right)$。图10还提供了一条有趣的信息，其中我们展示了一个

模型 (S1)	$\hat{\alpha}_n$	$\hat{\beta}_n$	$\tilde{\alpha}_n$	$\tilde{\beta}_n$
$n = 5,000$	(-0.099,0.075)	(-0.186,0.208)	(-0.034, 0.044)	(-0.016,0.186)
$n = 10,000$	(-0.037,0.065)	(-0.072,0.183)	(-0.057, 0.044)	(-0.154,0.124)
$n = 20,000$	(-0.001,0.018)	(0.053,0.098)	(-0.022,0.026)	(-0.058,0.106)
模型（S2）	$\hat{\alpha}_n$	$\hat{\beta}_n$	$\tilde{\alpha}_n$	$\tilde{\beta}_n$
$n = 5,000$	(-0.011,0.022)	(0.033,0.105)	(-0.037,0.029)	(-0.125,0.112)
$n = 10,000$	(-0.010,0.016)	(-0.041,0.084)	(-0.013,0.010)	(-0.044,0.085)
$n = 20,000$	(-0.005,0.006)	(-0.012,0.053)	(-0.020,0.012)	(-0.089,0.067)
模型（S3）	$\hat{\alpha}_n$	$\hat{\beta}_n$	$\tilde{\alpha}_n$	$\tilde{\beta}_n$
$n = 5,000$	(-0.051,0.126)	(-0.022,0.102)	(-0.032,0.050)	(0.010,0.130)
$n = 10,000$	(-0.002,0.020)	(0.010,0.061)	(-0.025,0.039)	(-0.010,0.098)
$n = 20,000$	(-0.000,0.016)	(-0.003,0.049)	(-0.007,0.022)	(0.017,0.082)
模型（S4）	$\hat{\alpha}_n$	$\hat{\beta}_n$	$\tilde{\alpha}_n$	$\tilde{\beta}_n$
$n = 5,000$	(-0.186,0.259)	(-0.124,0.145)	(-0.096,0.083)	(-0.052,0.150)
$n = 10,000$	(-0.169,0.233)	(-0.099,0.147)	(-0.025,0.041)	(0.044,0.146)
$n = 20,000$	(-0.033,0.109)	(-0.020,0.103)	(-0.027,0.038)	(0.021,0.120)

表2：模型 (S1–4) 下 $n = 10,000$、$n = 20,0000$ 和 $n = 30,0000$（按行排列）的偏差和标准差性能，以及差异度 d(·)（基于积分）和 s(·)（基于上确界）

.

观测过程 Z 的10条经验累积分布函数曲线（我们的方法以此为基础），并将其与真实的 Z-累积分布函数 F 进行比较。请注意，图9和图10是独立生成的（曲线之间没有颜色对应关系）。该图尤其有助于直观展示输入到我们的半参数估计方法中的函数估计质量，以及经过参数估计步骤与代入反演步骤后（见表达式（17））质量如何下降。为清晰起见并便于解读，我们建议在模型 (S0)^{strong} 上运行基于半参数反演的方法；该模型被认为易于估计，因为对于诸如 $n = 1,000$ 和 $5,000$ 这样的样本量，可以获得相当可靠的参数估计量。请注意，我们可以对估计量（17）进行截断/正则化，只保留逆曲线的正部（见表达式（17）），这样就能得到令人满意得多、类似累积分布函数的曲线。在图9中，我们选择保留估计量的原始形式，以清楚展示样本量 n 如何因定理3和定理4所述的收敛结果而影响基于代入反演的函数估计量（17）的正则性/累积分布函数相符性。事实上，从图9可以看出，当样本量较小（$n = 1,000$）时，目标 F^1 曲线的左侧通常估计得相当差（这显然会影响整条曲线）；当 n 变大（$n = 5,000$）时，无需任何技巧，这一缺陷几乎就能消除。要解释左侧表现不佳的原因，可以回看以下 F^1-误差分解：

$$
\begin{aligned}
F^1(x) - \hat{F}_n^1(x) &= \frac{1}{r^*}\left(F(x) - p^* F^0(x)\right) - \frac{1}{\hat{r}_n}\left(\hat{F}_n(x) - \hat{p}_n F^0(x)\right) \\
&= \frac{1}{r^*}\left(F(x) - \hat{F}_n(x)\right) + \left(\frac{1}{r^*} - \frac{1}{\hat{r}_n}\right)\hat{F}_n(x) - \left(\frac{p^*}{r^*} - \frac{\hat{p}_n}{\hat{r}_n}\right)F^0(x) \\
&\simeq \frac{1}{r^*}F(x) - \left(\frac{p^*}{r^*} - \frac{\hat{p}_n}{\hat{r}_n}\right)F^0(x), \quad \text{in the left-tail of } F \qquad (23) \\
&\simeq \left(\frac{1}{r^*} - \frac{1}{\hat{r}_n}\right)\hat{F}_n(x) - \left(\frac{p^*}{r^*} - \frac{\hat{p}_n}{\hat{r}_n}\right)F^0(x), \quad \text{in the right-tail of } F. \qquad (24)
\end{aligned}
$$

近似式（23）基于这样一个事实：在 F 的左尾，尤其是当 n (非常) ˆ 小时，观测值很少，因此与 F 和 F_0（精确量）相比，F_n 可以忽略不计。当我们仔细查看图10中设计值介于 -2 和 3 之间的部分时，对于 $n = 1,000$，这种现象仍然依稀可见（负偏差）。注意，在 (S0)^{strong} 下，我们一阶模型 (4) 的左侧分量是 F^0（远离地位于 F^1 的左侧），其权重为 $\beta^*/(\alpha^* + \beta^*) \simeq 0.533$。因此，当 $n = 1,000$ 时，F 的左尾实际上是基于仅略多于 500 个样本进行估计的。另一方面，近似式（24）基于这样一个事实：在 F 的右尾，差值 $F - F_n$ 非常 ˆ 接近 0，即使 n 的值很小也是如此，因为我们将 $F_n(x)$ 中的所有数据汇总起来，以近似 ˆ$F(x)$（ F 和 F_n 都非常接近 1）。由此可见，在 F 的左尾，较差的 F 估计与参数估计值的拟合质量共同产生了某种混合影响；而在 F 的右尾，偏差主要取决于参数估计值的拟合。据我们看来，这解释了为什么图9所示曲线在右尾的拟合效果远好于左尾，尤其是当 n 增大且欧几里得参数拟合在 $\sqrt{n}$ 混合条件下有所改善时，参见定理 3。

7 总结

本文介绍了一类新的半参数时序混合模型：所谓的金标准马尔可夫中毒模型。该模型是以下模型的扩展：

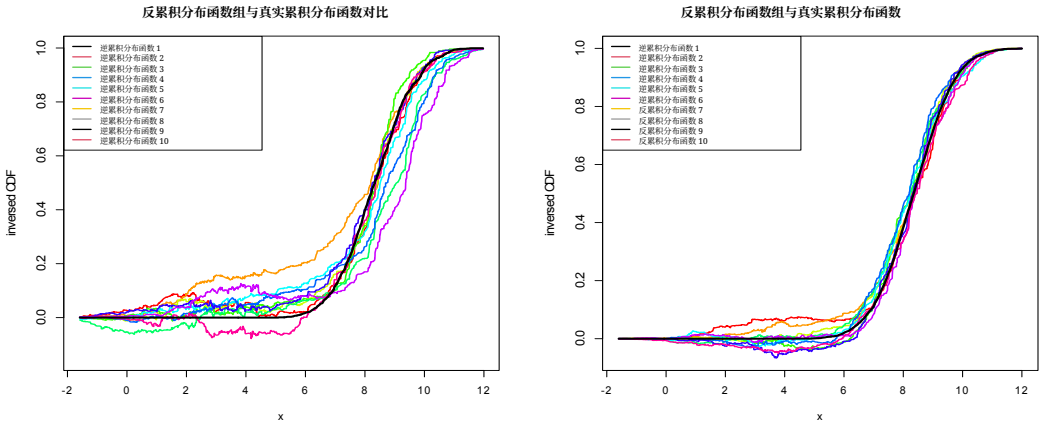


图9：10个基于代入反演的函数型估计量的各面板（17），对应于 F^1 模型（S0）下的（以黑色粗体显示）$strong$，分别对应于 $n = 1$，000和5，000。

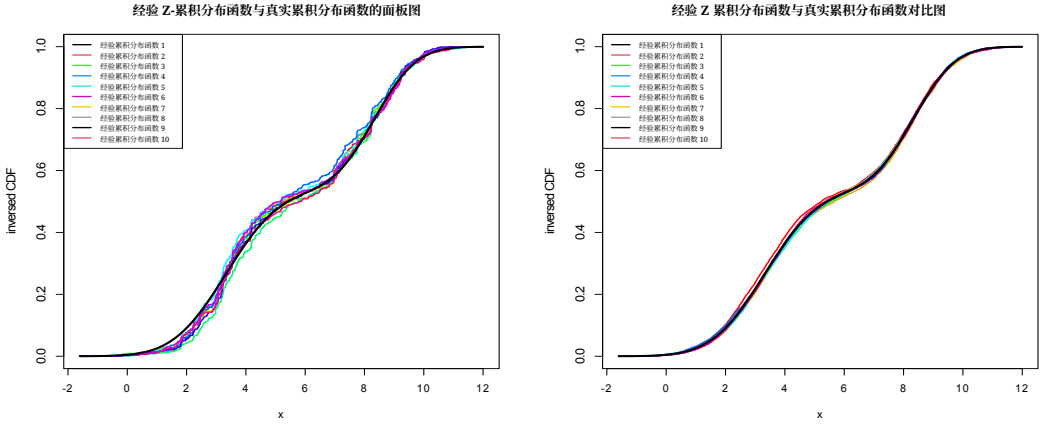


图10：模型（S0）下10个 Z-累积分布函数的各面板（以黑色粗体显示）$strong$，分别对应于 $n = 1$，000和5，000。

近二十年来得到广泛研究的经典半参数污染模型，例如Patra 和 Sen (2016)以及Milhaud 等人(2024a)及其中的参考文献。我们尤其研究了该模型的半参数估计，并提出了两个潜在马尔可夫链转移矩阵的最小对比估计量；我们证明，其中一个具有 $\sqrt{n}$-一致性，另一个则具有强相合性。我们还给出了函数中心极限定理，并为 F^1 的代入估计量（即未知的独立同分布污染序列累积分布函数）提供了“有限样本量”收敛速率。我们考察的模拟设置尤其凸显了分布 Y^1 与 Y^0 之间距离的作用（见假设**LinInd**），以及已知随机过程 Y^0所携带的时间依赖程度对估计性能的影响。理论结果的证明（见第8.3节）也表明以下因素发挥着关键作用：

- $T_1 = G^0 - F^0 \otimes F^0$，它随时间刻画了金标准过程 Y^0的偏离独立性，
- $T_2 = r^{*2}(F^1 - F^0) \otimes (F^1 - F^0)$，用于检测混合过程之间的偏离

Y^1 和 Y^0 平稳分布。

这些量尤其出现在对比函数的海森矩阵中，进而出现在我们的估计量 θ_n 的 $\sqrt{n}$-极限协方差矩阵 Σ 中，参见定理 3 及其证明（第 8.3.3 节）。证明的主要工具是恰当的变量变换（使 T_1 和 T_2 能够显现），参见第 8.3 节中 g 和 h 同胚的定义，以及在非独立观测情况下经验过程 $\mathbb{F}_n := \sqrt{n}(F_n - F)$ 和 $\mathbb{G}_n := \sqrt{n}(G_n - G)$ 的联合收敛。我们还使用基弗过程以及混合序列 $\mathbb{R}^2$ 的经验分布函数的强逼近，为参数和非参数估计量证明了一个非渐近界，参见第 8.5 节中引理 3 的证明。我们的方法可以扩展到金标准过程的分布未知、但可单独观测（训练数据）的情形。事实上，此时我们可以将 $G^0(x, y)$ 替换为

$$\tilde{G}^0_N(x, y) = \frac{1}{N-1} \sum_{k=1}^{N-1} \mathbb{1}_{\{Y^0[k] \le x, Y^0[k+1] \le y\}}, \quad (x, y) \in E^2,$$

假设我们可以获得额外的观测值 $(Y^0[k])_{1 \le k \le N}$，代价是增加一个阶为 $1/\sqrt{N}$ 的误差项（当 N 相对于 n较大时，该项可忽略）。拓展到时变污染过程，即 Y^1 假定其为平稳混合或马尔可夫的过程，是下一项挑战。这需要使用（Z_i, Z_{i+1}, Z_{i+2}）的三阶分布。初步计算表明，似乎可以应用我们先前的方法，但技术细节会相当繁琐（涉及许多复杂项）。最后，我们在第5节末讨论了一种解码策略，见表达式（20），用于恢复模型中涉及的潜在抽样马尔可夫链的状态。该方法基于对 Y^1 概率密度函数的逆核密度估计，见表达式（19），这将在后续工作中加以研究。

8 证明

8.1 命题1

回想一下，由于 $\alpha \neq 0$，我们有 $r \neq 0$，其中 $r = \alpha/(\alpha + \beta)$。现在记为

$$G_\theta = \lambda_1 G^0 + \lambda_2 F^0 \otimes F^1_\theta + \lambda_3 F^1_\theta \otimes F^0 + \lambda_4 F^1_\theta \otimes F^1_\theta.$$

将 F^1_θ 替换为其值 $(F - pF^0)/r$ 并使用 $\lambda_3 = \lambda_2$，我们得到

$$G_\theta = \lambda_1 G^0 + \left(\frac{\lambda_2}{r} - \frac{p\lambda_4}{r^2}\right)(F^0 \otimes F + F \otimes F^0) + \frac{\lambda_4}{r^2} F \otimes F + \left(-2\frac{\lambda_2 p}{r} + \frac{p^2\lambda_4}{r^2}\right) F^0 \otimes F^0.$$

然后我们将 λ_i 替换为其值

$$\lambda_1 = p(1-\alpha), \quad \lambda_2 = \lambda_3 = p\alpha = r\beta, \quad \lambda_4 = r(1-\beta),$$

并使用记号 $b = 1 - \beta$ 得到以下分解

$$G_\theta = p(1-\alpha)G^0 + \left(\beta - \frac{pb}{r}\right)(F^0 \otimes F + F \otimes F^0) + \frac{b}{r} F \otimes F + \left(-2\beta p + \frac{p^2 b}{r}\right) F^0 \otimes F^0.$$

接下来，我们只保留变量 r 和 b，得到

$$\begin{aligned} G_\theta &= (1-2r+rb)G^0 + \left(1-\frac{b}{r}\right)(F^0\otimes F + F\otimes F^0) + \frac{b}{r}F\otimes F + \left(-2+2r-rb+\frac{b}{r}\right)F^0\otimes F^0 \\ &= (1-2r+rb)\left(G^0 - F^0\otimes F^0\right) + \left(1-\frac{b}{r}\right)(F^0\otimes F + F\otimes F^0 - F^0\otimes F^0) + \frac{b}{r}F\otimes F \\ &= (1-2r+rb)\left(G^0 - F^0\otimes F^0\right) + \left(\frac{b}{r}-1\right)(F-F^0)\otimes(F-F^0) + F\otimes F. \end{aligned} \tag{25}$$

由于 $F - F^0 = r^*(F^1 - F^0)$，我们也可以写成

$$G_\theta = (1-2r+rb)\left(G^0 - F^0\otimes F^0\right) + \frac{r^{*2}}{r^2}\left(rb-r^2\right)(F^1-F^0)\otimes(F^1-F^0) + F\otimes F.$$

现在，由于 $G = G_{\theta^*}$，我们可以写出以下 Δ 分解

$$\begin{aligned} \Delta(\theta,\cdot) &= G_{\theta^*} - G_\theta = [(1-2r^*+r^*b^*) - (1-2r+rb)]\left(G^0 - F^0\otimes F^0\right) \\ &\quad + [(r^*b^* - r^{*2}) - \frac{r^{*2}}{r^2}(rb-r^2)](F^1-F^0)\otimes(F^1-F^0). \end{aligned}$$

等式 (10) 现在通过取值证明

$$c_1(\theta^*,\theta) = -2r^* + r^*b^* + 2r - rb, \quad \text{and} \quad c_2(\theta^*,\theta) = \frac{r^*}{r}(b^*r - r^*b). \tag{26}$$

此外，如果 $c_1(\theta^*, \theta) = 0$ 且 $c_2(\theta^*, \theta) = 0$，那么

$$0 = c_1(\theta^*,\theta) - c_2(\theta^*,\theta)r/r^* = -2r^* + r^*b^* + 2r - rb - b^*r + r^*b = (r^*-r)(-2+b^*+b).$$

但请回想，b和 b^*属于 $[\delta, 1-\delta]$，因此 $b+b^* < 2$且必然有 $r = r^*$。再次使用$c_1(\theta^*, \theta) = 0$, 可得 $b = b^*$，最后有 $(\alpha, \beta) = (\alpha^*, \beta^*)$。 □

8.2 一致性证明

引理 1. 对于所有 $(\theta, \theta') \in \Theta^2$，其中 $\Theta = [\delta, 1-\delta]$ 满足 $\delta \in]0, 1/2[$，我们有

$$|\mathbf{d}(\theta) - \mathbf{d}(\theta')| \le C(\delta)\|\theta - \theta'\|_1, \quad \textit{and} \quad |\mathbf{d}_n(\theta) - \mathbf{d}_n(\theta')| \le C(\delta)\|\theta - \theta'\|_1,$$

其中 $\|\cdot\|_1$ 是为任意一般向量 $u = (u_1, u_2)^T \in \mathbb{R}^2$ 按 $\|u\|_1 = |u_1| + |u_2|$定义的范数。因此，$\mathbf{d}(\cdot)$ 和 $\mathbf{d}_n(\cdot)$ 是 Θ上的一致连续映射。对于 $\mathbf{s}(\cdot)$和 $\mathbf{s}_n(\cdot)$，同样的结论也成立。

证明。记

$$\hat{G}_{n,\theta} = \lambda_1 G^0 + \lambda_2 F^0\otimes\hat{F}^1_{n,\theta} + \lambda_3\hat{F}^1_{n,\theta}\otimes F^0 + \lambda_4\hat{F}^1_{n,\theta}\otimes\hat{F}^1_{n,\theta},$$

其中我们回顾 $F^1_{n,\theta} = (\hat{F}_n - pF^0)/r$. 与命题 1 的证明类似，类似于等式（25），我们有：

$$\hat{G}_{n,\theta} = (1-2r+rb)\left(G^0 - F^0\otimes F^0\right) + \left(\frac{b}{r}-1\right)(\hat{F}_n - F^0)\otimes(\hat{F}_n - F^0) + \hat{F}_n\otimes\hat{F}_n.$$

此外，$1-2r+rb=pa$ 其中我们记 $a=1-\alpha$。现在利用 Δ_n 的定义，见下方等式（12），可得：

$$\Delta_n(\theta,\cdot)-\Delta_n(\theta',\cdot)=(p'a'-pa)\left(G^0-F^0\otimes F^0\right)+\left(\frac{b'}{r'}-\frac{b}{r}\right)(\hat{F}_n-F^0)\otimes(\hat{F}_n-F^0).$$

现在注意到 G^0，F^0 和 $\hat{} F_n$ 均以 1 为界。因此可以得到以下上界

$$\begin{aligned}|\Delta_n(\theta,\cdot)-\Delta_n(\theta',\cdot)| &\leq |p'a'-pa|+\left|\frac{b'}{r'}-\frac{b}{r}\right| \\ &\leq |p'(a'-a)+a(p'-p)|+\left|\frac{b'(r-r')+r'(b'-b)}{rr'}\right|.\end{aligned}$$

接下来，利用 α 和 β 属于 $[\delta,1-\delta]$ 这一事实，再由 $1/r=1+\beta/\alpha\in[1,1/\delta]$ 可得

$$|\Delta_n(\theta,\cdot)-\Delta_n(\theta',\cdot)|\leq|a'-a|+|p'-p|+\frac{1}{\delta^2}|r-r'|+\frac{1}{\delta}|b-b'|.$$

此外，$(r'-r)(\alpha'+\beta')=p(\alpha'-\alpha)+r(\beta-\beta')$，由此可得

$$|r'-r|\leq\frac{1}{2\delta}\left(|\alpha'-\alpha|+|\beta'-\beta|\right).$$

因此，由于 $p=1-r$、$a=1-\alpha$ 和 $b=1-\beta$，我们有

$$|\Delta_n(\theta,\cdot)-\Delta_n(\theta',\cdot)|\leq\left(\frac{1}{2\delta^3}+\frac{1}{2\delta}+1\right)|\alpha'-\alpha|+\left(\frac{1}{2\delta^3}+\frac{3}{2\delta}\right)|\beta'-\beta|,$$

并利用 $\delta\leq1$，最终得到

$$|\Delta_n(\theta,\cdot)-\Delta_n(\theta',\cdot)|\leq\frac{1}{\delta^3}\left(2|\alpha'-\alpha|+2|\beta'-\beta|\right)\leq2\delta^{-3}\|\theta-\theta'\|_1.$$

此外，由于 $|\Delta_n(\theta,\cdot)|\leq1+|pa|+\left|\frac{b}{r}=1\right|\leq(3+\delta^{-1})\leq4\delta^{-1}$，我们得到

$$|\Delta_n(\theta,\cdot)-\Delta_n(\theta',\cdot)|^2\leq(4\delta^{-1}+4\delta^{-1})2\delta^{-3}\|\theta-\theta'\|_1,$$

以及 $|\mathbf{d}_n(\theta)-\mathbf{d}_n(\theta')|\leq16\delta^{-4}\|\theta-\theta'\|_1$. 同理，我们有 $|\mathbf{d}(\theta)-\mathbf{d}(\theta')|\leq16\delta^{-4}\|\theta-\theta'\|_1$. □

命题 3. 若 $\Theta=[\delta,1-\delta]^2$，且 $\delta\in]0,1/2[$，则有：

$$\sup_{\theta\in\Theta}|\mathbf{d}_n(\theta)-\mathbf{d}(\theta)|\leq(6+2\delta^{-1})\left(\|\hat{G}_n-G\|_\infty+(2+4\delta^{-1})\|\hat{F}_n-F\|_\infty\right).$$

将 $\mathbf{d}_n(\cdot)-\mathbf{d}(\cdot)$ 替换为 $\mathbf{s}_n(\cdot)-\mathbf{s}(\cdot)$，同样的界也成立。

证明。回顾前面的证明，我们有如下分解

$$\Delta_n(\theta,\cdot)=\left(\hat{G}_n-\hat{F}_n\otimes\hat{F}_n\right)-pa\left(G^0-F^0\otimes F^0\right)-\left(\frac{b}{r}-1\right)(\hat{F}_n-F^0)\otimes(\hat{F}_n-F^0),$$

由此得到 $|\Delta_n(\theta,\cdot)|\leq1+|pa|+\left|\frac{b}{r}=1\right|\leq3+\delta^{-1}$。同理，$|\Delta(\theta,\cdot)|\leq3+\delta^{-1}$。因此，我们得到以下主要化结果：

$$\left|\Delta_n^2-\Delta^2\right|\leq|\Delta_n-\Delta|(|\Delta_n|+|\Delta|)\leq(6+2\delta^{-2})|\Delta_n-\Delta|,$$

以及

$$|\mathbf{d}_n(\theta) - \mathbf{d}(\theta)| \leq (6 + 2\delta^{-1}) \iint |\Delta_n - \Delta|(\theta, x, y) dH(x, y).$$

Now

$$\begin{aligned} \Delta_n(\theta, \cdot) - \Delta(\theta, \cdot) &= \left(\hat{G}_n - G\right) - \lambda_2 F^0 \otimes \left(\hat{F}^1_{n,\theta} - F^1_\theta\right) - \lambda_3 \left(\hat{F}^1_{n,\theta} - F^1_\theta\right) \otimes F^0 \\ &\quad -\lambda_4 \left(\hat{F}^1_{n,\theta} \otimes \hat{F}^1_{n,\theta} - F^1_\theta \otimes F^1_\theta\right), \end{aligned}$$

并利用 F^1_θ和F^1_n的定义，θ分别见式（6）和式（13），可得

$$\begin{aligned} \Delta_n(\theta, \cdot) - \Delta(\theta, \cdot) &= \left(\hat{G}_n - G\right) - \frac{\lambda_2}{r} F^0 \otimes \left(\hat{F}_n - F\right) - \frac{\lambda_3}{r} \left(\hat{F}_n - F\right) \otimes F^0 \\ &\quad -\frac{\lambda_4}{r^2} \left(\left(\hat{F}_n - pF^0\right) \otimes \left(\hat{F}_n - pF^0\right) - \left(F - pF^0\right) \otimes \left(F - pF^0\right)\right), \end{aligned}$$

其中 $\lambda_2/r = \lambda_3/r = \beta$和 $\lambda_4/r = 1 - \beta = b$。于是

$$\begin{aligned} \Delta_n(\theta, \cdot) - \Delta(\theta, \cdot) &= \left(\hat{G}_n - G\right) + \left(\frac{b}{r} - 1\right) \left[F^0 \otimes \left(\hat{F}_n - F\right) + \left(\hat{F}_n - F\right) \otimes F^0\right] \\ &\quad -\frac{b}{r} \left(\hat{F}_n \otimes \hat{F}_n - F \otimes F\right). \end{aligned}$$

关于最后一项，我们可以注意到

$$\begin{aligned} |\hat{F}_n(x)\hat{F}_n(y) - F(x)F(y)| &= |\hat{F}_n(x)(\hat{F}_n(y) - F(y)) + F(y)(\hat{F}_n(x) - F(x))| \\ &\leq |\hat{F}_n(y) - F(y)| + |\hat{F}_n(x) - F(x)|, \end{aligned}$$

对所有的（x，y）$\in E^2$ 均成立。这便得到

$$|\Delta_n(\theta, x, y) - \Delta(\theta, x, y)| \leq |\hat{G}_n - G|(x, y) + \left(1 + \frac{2b}{r}\right) \left[|\hat{F}_n - F|(x) + |\hat{F}_n - F|(y)\right],$$

其中 $\sup_{\theta \in \Theta}(1 + 2b/r) \leq 1 + 2\delta^{-1}$。最终得到

$$\sup_{\theta \in \Theta} |\mathbf{d}_n(\theta) - \mathbf{d}(\theta)| \leq (6 + 2\delta^{-1}) \left(\|\hat{G}_n - G\|_\infty + (2 + 4\delta^{-1}) \|\hat{F}_n - F\|_\infty\right),$$

回想一下，我们假设了 $\iint dH = 1$。

□

命题4。 如果过程$(Z_i)_{i \geq 1}$ 是平稳的，则有

$$\|\hat{G}_n - G\|_\infty \xrightarrow{a.s} 0, \quad and \quad \|\hat{F}_n - F\|_\infty \xrightarrow{a.s} 0, \quad as \quad n \to +\infty.$$

^证明。记 $R_i = (Z_i, Z_{i+1})$，$i \geq 1$，则 G_n 恰为 $R =(R_i)_{i \geq 1}$ 的经验分布函数。另请注意，Z 和 R 均为平稳序列。于是，利用Athreya 和 Roy (2016) 的定理 2 和注3，可得 Glivenko-Cantelli 结果 ^ ^ 对 F_n成立，对 G_n也成立，证明完毕。

□

8.3 渐近正态性证明

请回顾表达式（25）可知，

$$\Delta(\theta,\cdot)=G-F\otimes F-(1-2r+rb)\left(G^0-F^0\otimes F^0\right)-\left(\frac{b}{r}-1\right)(F-F^0)\otimes\left(F-F^0\right),$$

其中我们回顾一下，$b=1-\beta$, $p=\beta/(\alpha+\beta)$，以及 $r=\alpha/(\alpha+\beta)$。然后我们记

$$v_1:=-1+2r-rb,\quad \text{and}\quad v_2:=-\frac{b}{r}+1.$$

我们还使用一组新的记号：

$$\begin{array}{rcl} T_1 &:=& G^0-F^0\otimes F^0\\ T_2 &:=& (F-F^0)\otimes(F-F^0)\\ T_3 &:=& G-F\otimes F\\ \hat{T}_2 &:=& (\hat{F}_n-F^0)\otimes(\hat{F}_n-F^0)\\ \hat{T}_3 &:=& \hat{G}_n-\hat{F}_n\otimes\hat{F}_n. \end{array}$$

因此我们用（25）再次进行变换，可得到以下关于 Δ 和 Δ_n的重新整理后的分解

$$\begin{array}{rcl} \Delta(\theta,\cdot) &=& T_3-(1-2r+rb)T_1-(b/r-1)T_2=v_1T_1+v_2T_2+T_3,\\ \Delta_n(\theta,\cdot) &=& \hat{T}_3-(1-2r+rb)T_1-(b/r-1)\hat{T}_2=v_1T_1+v_2\hat{T}_2+\hat{T}_3. \end{array}$$

现在，我们利用以下同胚映射 h 和 $g=h^{-1}$，将自然参数 $\theta=(\alpha,\ \beta)$ 变换为新的参数 $v=(v_1,\ v_2)$：

$$h:\ \begin{array}{rcl} \Theta &\to& h(\Theta)=V\\ \theta &\mapsto& v=(-1+2r-rb,-\frac{b}{r}+1)=\left(-\frac{\beta(1-\alpha)}{\alpha+\beta},\frac{\beta}{\alpha}(\alpha+\beta-1)\right), \end{array}$$

and

$$g:\ \begin{array}{rcl} V &\to& \Theta\\ v &\mapsto& \theta=\left(\frac{\beta(1-\beta)}{\beta-v_2},\beta\right),\quad \text{where } \beta=\sqrt{v_1v_2+v_2-v_1}. \end{array}$$

令 $v^*=h(\theta^*)=g^{-1}(\theta^*)$ 和 $\hat{}v_n=g^{-1}(\hat{}\theta_n)$。事实上，由于 θ^* 是 $\mathbf{d}$ 在 Θ上的（唯一）极小点，新的真实参数 v^* 也是 $\mathbf{d}\circ g$ 在 V上的（唯一）极小点。同理，$\hat{}v_n$ 是 $\mathbf{d}_n\circ g$的极小点。以下记

$$\mathtt{d}:=\mathbf{d}\circ g,\quad \text{and}\quad \mathtt{d}_n:=\mathbf{d}_n\circ g.$$

因此，由于 $\mathbf{d}=\iint\Delta^2dH$，我们特别有

$$\mathtt{d}(v)=\iint(v_1T_1+v_2T_2+T_3)^2dH,\quad \text{and}\quad \mathtt{d}_n(v)=\iint(v_1T_1+v_2\hat{T}_2+\hat{T}_3)^2dH.$$

注意，这些函数仅为二次函数。采用此记号，$v^*=\mathsf{argmin}_v\mathtt{d}$ 和 $\hat{}v_n=\mathsf{argmin}_v\mathtt{d}_n$。我们假设 θ^* 属于 Θ的内部，记作 $\mathsf{Int}(\Theta)$。$\hat{}$ 这意味着，当 n 足够大时，θ_n 几乎必然属于 Θ 的内部。由于$h(\mathsf{Int}(\Theta))\subset\mathsf{Int}(h(\Theta))$，$v^*$ 且当 n 足够大时，$\hat{}v_n$ 也属于 $\mathsf{Int}(V)$。这确保我们同时拥有

$$\dot{\mathtt{d}}(v^*)=0,\quad \text{and}\quad \dot{\mathtt{d}}_n(\hat{v}_n)=0,\tag{27}$$

其中，d 和 d_n分别表示 d 和 d_n的梯度。为方便起见，我们将 $[\mathbf{x}]_i$ 记作向量 $\mathbf{x}$ 的第 i个坐标，并将$[M]_{i,j}$ 记作矩阵 M的第 (i,j) 个分量。现在，如果我们能够证明

$$\sqrt{n}\begin{pmatrix}[\hat{v}_n]_1 - v_1^* \\ [\hat{v}_n]_2 - v_2^*\end{pmatrix} \xrightarrow{d} W \sim \mathcal{N}(0,\Gamma), \quad \text{as} \quad n \to +\infty,$$

那么，δ方法可直接得到

$$\sqrt{n}\left(\hat{\theta}_n - \theta^*\right) = \sqrt{n}\left(g(\hat{v}_n) - g(v^*)\right) \xrightarrow{d} Dg(v^*)W, \quad \text{as} \quad n \to +\infty,$$

其中，$Dg(v)$ 是函数 $g\colon (v_1,v_2) \mapsto (\alpha,\beta)$ 在点 $v=(v_1,v_2)$ 处的微分。由于$g = h^{-1}$ 我们有 $Dg(h(\theta))Dh(\theta) = D(g\circ h)(\theta) = I$，由此得到

$$Dh(\theta) = \begin{pmatrix} \frac{\beta^2+\beta}{(\alpha+\beta)^2} & \frac{\alpha^2-\alpha}{(\alpha+\beta)^2} \\ \frac{\beta-\beta^2}{\alpha^2} & \frac{\alpha+2\beta-1}{\alpha}\end{pmatrix}, \quad \text{and} \quad Dg(v) = [Dh(\theta)]^{-1} = \begin{pmatrix} \frac{(\alpha+\beta)(\alpha+2\beta-1)}{2\beta^2} & \frac{\alpha^2(1-\alpha)}{2\beta^2(\alpha+\beta)} \\ \frac{(\alpha+\beta)(\beta-1)}{2\alpha\beta} & \frac{\alpha(\beta+1)}{2\beta(\alpha+\beta)}\end{pmatrix}. \tag{28}$$

记 $D^* = Dg(v^*) = [Dh(\theta^*)]^{-1}$，则有 $\Sigma = D^*\Gamma(D^*)^T$。现在我们需要研究当 $n \to +\infty$ 时 $\sqrt{n}(\hat{v}_n - v^*)$ 的极限。函数 d_n 是海森矩阵为常数的二次函数

$$\ddot{\mathrm{d}}_n = 2\begin{pmatrix} \iint T_1^2 dH & \iint T_1\hat{T}_2 dH \\ \iint T_1\hat{T}_2 dH & \iint \hat{T}_2^2 dH\end{pmatrix},$$

并且对于任意 $v \in V$，我们有：

$$\dot{\mathrm{d}}_n(v) = \dot{\mathrm{d}}_n(v^*) + \ddot{\mathrm{d}}_n(v - v^*).$$

然后，在点 $v = \hat{v}_n$ 处应用此结果，并利用 (27)，得到

$$\sqrt{n}\ddot{\mathrm{d}}_n(\hat{v}_n - v^*) = -\sqrt{n}\dot{\mathrm{d}}_n(v^*) = \sqrt{n}(\dot{\mathrm{d}} - \dot{\mathrm{d}}_n)(v^*). \tag{29}$$

剩下需要证明的是，$\sqrt{n}(\dot{\mathrm{d}} - \dot{\mathrm{d}}_n)(v^*)$ 收敛于一个中心高斯变量（见第 8.3.2 节），并且 $\ddot{\mathrm{d}}_n$ 收敛于一个可逆矩阵（见第 8.3.1 节），当 $n \to +\infty$ 时。为此，我们需要以下引理。

引理 2. 对所有 $n \geq 1$, $\mathbb{F}_n := \sqrt{n}(F_n - F)$和 $\mathbb{G}_n := \sqrt{n}(G_n - G)$，记为如下形式，我们有以下渐近行为：

i) 经验过程 $(\mathbb{F}_n)_{n\geq 1}$ 依分布收敛到一个高斯过程 $\mathcal{B}_F$，其协方差函数为

$$cov(\mathcal{B}_F(x), \mathcal{B}_F(y)) = \sum_{k\in\mathbb{Z}} cov(\mathbb{1}_{\{Z_0\leq x}, \mathbb{1}_{Z_k\leq y\}}).$$

ii) 经验过程$(\mathbb{G}_n)_{n\geq 1}$ 依分布收敛到一个高斯过程 $\mathcal{B}_G$，其协方差函数为

$$cov(\mathcal{B}_G(x,y), \mathcal{B}_G(z,t)) = \sum_{k\in\mathbb{Z}} cov(\mathbb{1}_{\{Z_0\leq x, Z_1\leq y\}}, \mathbb{1}_{\{Z_k\leq z, Z_{k+1}\leq t\}}).$$

iii) 联合过程 $(\mathbb{F}_n, \mathbb{G}_n)_{n\geq 1}$ 依分布收敛到高斯过程 $(\mathcal{B}_F, \mathcal{B}_G)$。

证明。 i) 注意，马尔可夫链 X 是几何 α-混合的。事实上，转移矩阵以速率 $(1-\alpha-\beta)^n$ 收敛到平稳分布矩阵，其中$|1-\alpha-\beta|<1-2\delta<1$。根据我们的混合假设，过程 $(Z_i)_{i\geq 1}$ 是 α-混合的，且 $\alpha^Z(n)\leq Cn^{-a}$：参见（Vandekerkhove，2005年）引理 1(i) 及其证明，或 Bradley(2005年）定理 5.2。 $(\mathbb{F}_n)_{n\geq 1}$ 的收敛性随后可由Rio (2017）定理 7.2 得出。

ii) 令 $R_i := (Z_i, Z_{i+1})$，对所有 $i\geq 1$ 均成立。现在可以观察到

$$\hat{G}_n(x,y)=\frac{1}{n-1}\sum_{i=1}^{n-1}\mathbb{1}_{\{Z_i\leq x, Z_{i+1}\leq y\}}=\frac{1}{n-1}\sum_{i=1}^{n-1}\mathbb{1}_{\{R_i\leq (x,y)\}}$$

因此， G_n 实际上是 $(R_i)_{i\geq 1}$ 的经验分布函数。于是有

$$\alpha^R(n)\leq \alpha^Z(n-1)\leq C(n/2)^{-a},$$

其中 $a>1$，其中 α-混合系数定义见（16）。记 B（T）为定义在 T 上的实值有界函数空间。利用 Rio （2017）的定理 7.3 以及 F 的连续性，可得 $\mathbb{G}_n$ 在分布上收敛于高斯过程 $\mathcal{B}_G$，且收敛发生在空间 B（$\mathbb{R}^2$）中。

iii) 现在暂且假设 Y^0, Y^1，并且 Z 支持于 $[0,1]$ 上。记作$\tilde{\mathbb{G}}_n=\sqrt{(n-1)/n}\mathbb{G}_n$。那么，对所有 $x\in E$, $G(x,1)=F(x)$，并且

$$\begin{aligned}\tilde{\mathbb{G}}_n(x,1) &= \sqrt{n-1}\left(\frac{1}{n-1}\sum_{i=1}^{n-1}\mathbb{1}_{\{Z_i\leq x, Z_{i+1}\leq 1\}}-G(x,1)\right)\\ &= \sqrt{n-1}\left(\hat{F}_{n-1}(x)-F(x)\right)=\mathbb{F}_{n-1}(x).\end{aligned}$$

Let

$$\phi: \begin{array}{ccc} B(\mathbb{R}^2) & \to & B(\mathbb{R})\times B(\mathbb{R}^2)\\ w & \mapsto & (w(.,1),w).\end{array}$$

那么 ϕ 是线性的，且 $\|w(.,1)\|_\infty\leq\|w\|_\infty$，因此 ϕ 是连续线性算子。我们还有 $\phi(\tilde{\mathbb{G}}_n)=(\mathbb{F}_{n-1},\mathbb{G}_n)$。由于 $\mathbb{G}_n$ 在空间 $B(\mathbb{R}^2)$ 中依分布收敛于高斯过程 $\mathcal{B}$，连续映射定理意味着 $(\mathbb{F}_{n-1},\tilde{\mathbb{G}}_n)$（进而 $(\mathbb{F}_n,\mathbb{G}_n)$）依分布收敛于高斯过程 $\phi(\mathcal{B})$。

如果现在不再假设 Z 支持于 $[0,1]$,，只需考虑$U_i=F(Z_i)$。众所周知， U_i 服从均匀分布（回想一下， F 是连续的）。我们用 $\hat{F}_n^U$ 表示 $(U_i)_{1\leq i\leq n}$ 的经验分布函数，用 $\hat{G}_n^U$ 表示 $(U_i, U_{i+1})_{1\leq i\leq n-1}$ 的经验分布函数。类似地，我们用记号 F^U, G^U, $\mathbb{F}_n^U$, $\mathbb{G}_n^U$ 表示真实分布函数和经验过程。注意，对于所有 $x\in E$，都有

$$\hat{F}_n(x)=\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{Z_i\leq x\}}=\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{F(Z_i)\leq F(x)\}}=\hat{F}_n^U(F(x)).$$

同理， $F(x)=F^U(F(x))$, $\hat{G}_n(x,y)=\hat{G}_n^U(F(x)\ F(y))$, $G(x,y)=G^U(F(x)\ F(y))$, 最后

$$\mathbb{F}_n(x)=\mathbb{F}_n^U(F(x)),\qquad \mathbb{G}_n(x,y)=\mathbb{G}_n^U(F(x),F(y)).$$

由于 U的支集位于 $[0,1]$ 上，我们可以将前面的推理应用于 $(\mathbb{F}_n^U, \mathbb{G}_n^U)_{n\geq 1}$。这确保了 $(\mathbb{F}_n, \mathbb{G}_n)_{n\geq 1}$依分布收敛。

□

8.3.1 海森收敛

¨ ¨我们将证明，当 $n \to +\infty$ 时，d_n 依概率收敛到 d。函数 d 是二次函数，其海森矩阵为常数

$$\ddot{\mathrm{d}} = 2 \begin{pmatrix} \iint T_1^2 dH & \iint T_1 T_2 dH \\ \iint T_1 T_2 dH & \iint T_2^2 dH \end{pmatrix}. \tag{30}$$

此外，¨d 可逆，当且仅当 $(\iint T_1 T_2 dH)^2 \neq \iint T_1^2 dH \iint T_2^2 dH$。但柯西-施瓦茨不等式取等号的情形是 T_1 与 T_2 成比例。假设 **LinInd** 确保 T_1 与 $(F^1 - F^0) \otimes (F^1 - F^0) = T_2 / r^{*2}$ 不成比例。因此，¨d 可逆。接下来只需证明，当 $n \to +\infty$ 时，(¨d_n − ¨d) 依概率收敛至 $\to 0$。首先，$[\ddot{\mathrm{d}}_n - \ddot{\mathrm{d}}]_{1,\ 1} = 0$. 此外，我们有

$$[\ddot{\mathrm{d}}_n - \ddot{\mathrm{d}}]_{1,2} = 2 \iint T_1 (\hat{T}_2 - T_2) dH = \frac{2}{\sqrt{n}} \iint T_1 \sqrt{n} (\hat{T}_2 - T_2) dH,$$

and

$$\begin{aligned} [\ddot{\mathrm{d}}_n - \ddot{\mathrm{d}}]_{2,2} &= 2 \iint (\hat{T}_2^2 - T_2^2) dH = 2 \iint (\hat{T}_2 - T_2)^2 dH + 4 \iint T_2 (\hat{T}_2 - T_2) dH \\ &= \frac{2}{n} \iint \left(\sqrt{n} (\hat{T}_2 - T_2) \right)^2 dH + \frac{4}{\sqrt{n}} \iint T_2 \sqrt{n} (\hat{T}_2 - T_2) dH. \end{aligned}$$

ˆ 注意，对于任意函数 ϕ 和 ϕ，我们都有如下恒等式

$$\phi \otimes \phi - \hat{\phi} \otimes \hat{\phi} = \phi \otimes (\phi - \hat{\phi}) + (\phi - \hat{\phi}) \otimes \hat{\phi},$$

由此可得

$$T_2 - \hat{T}_2 = (F - F^0) \otimes (F - \hat{F}_n) + (F - \hat{F}_n) \otimes (\hat{F}_n - F^0).$$

ˆ 回顾记号 $\mathbb{F}_n := \sqrt{n}(F - \bar{F})$，可得到如下分解

$$\sqrt{n}(T_2 - \hat{T}_2) = (F^0 - F) \otimes \mathbb{F}_n + \mathbb{F}_n \otimes (F^0 - F) - \frac{1}{\sqrt{n}} \mathbb{F}_n \otimes \mathbb{F}_n.$$

那么，$\mathbb{F}_n$ 的收敛性意味着 $\iint T_1 \sqrt{n}(\hat{}T_2 - T_2) dH$、$\iint T_2 \sqrt{n}(\hat{}T_2 - T_2) dH$ 和 $\iint \left(\sqrt{n}(\hat{}T_2 - T_2) \right)^2 dH$ 均收敛；为此，我们利用 F，F^0，T_1，T_2 有界且 $\iint dH < \infty$。最后再应用斯卢茨基定理，可得 $[\ddot{\mathrm{d}}_n - \ddot{\mathrm{d}}]_{1,2}$ 和 $[\ddot{\mathrm{d}}_n - \ddot{\mathrm{d}}]_{2,2}$ 依概率收敛至 0。因此，¨d_n 依概率收敛至 ¨d，而后者可逆。

8.3.2 $\sqrt{n}(\dot{\mathrm{d}} - \dot{\mathrm{d}}_n)(v^*)$的收敛性

首先回顾

$$\mathrm{d}(v) = \iint (v_1 T_1 + v_2 T_2 + T_3)^2 dH, \quad \text{and} \quad \mathrm{d}_n(v) = \iint (v_1 T_1 + v_2 \hat{T}_2 + \hat{T}_3)^2 dH.$$

那么梯度为

$$\dot{\mathrm{d}}(v) = 2 \begin{pmatrix} \iint T_1 (v_1 T_1 + v_2 T_2 + T_3) \\ \iint T_2 (v_1 T_1 + v_2 T_2 + T_3) \end{pmatrix}, \quad \text{and} \quad \dot{\mathrm{d}}_n(v) = 2 \begin{pmatrix} \iint T_1 (v_1 T_1 + v_2 \hat{T}_2 + \hat{T}_3) \\ \iint \hat{T}_2 (v_1 T_1 + v_2 \hat{T}_2 + \hat{T}_3) \end{pmatrix},$$

然后

$$[\dot{\mathtt{d}}-\dot{\mathtt{d}}_n]_1(v)=2v_2\iint T_1(T_2-\hat{T}_2)dH+2\iint T_1(T_3-\hat{T}_3)dH,$$

and

$$\begin{aligned}[\dot{\mathtt{d}}-\dot{\mathtt{d}}_n]_2(v) &= 2\iint(v_1T_1+v_2T_2+T_3)(T_2-\hat{T}_2)dH+2v_2\iint\hat{T}_2(T_2-\hat{T}_2)dH\\ &\quad+2\iint\hat{T}_2(T_3-\hat{T}_3)dH.\end{aligned}$$

^^^^^^我们再次利用，对于任意函数 ϕ 和 ϕ，$\phi\otimes\phi-\phi\otimes\phi=\phi\otimes(\phi-\phi)+(\phi-\phi)\otimes\phi$。因此

$$\begin{aligned}T_2-\hat{T}_2 &= (F-F^0)\otimes(F-\hat{F}_n)+(F-\hat{F}_n)\otimes(\hat{F}_n-F^0)\\ T_3-\hat{T}_3 &= (G-\hat{G}_n)-F\otimes(F-\hat{F}_n)-(F-\hat{F}_n)\otimes\hat{F}_n,\end{aligned}$$

进而

$$\sqrt{n}(T_2-\hat{T}_2) = (F^0-F)\otimes\mathbb{F}_n+\mathbb{F}_n\otimes(F^0-F)-\frac{1}{\sqrt{n}}\mathbb{F}_n\otimes\mathbb{F}_n \tag{31}$$

$$\sqrt{n}(T_3-\hat{T}_3) = -\mathbb{G}_n+F\otimes\mathbb{F}_n+\mathbb{F}_n\otimes F+\frac{1}{\sqrt{n}}\mathbb{F}_n\otimes\mathbb{F}_n. \tag{32}$$

我们有

$$\begin{aligned}\sqrt{n}[\dot{\mathtt{d}}-\dot{\mathtt{d}}_n]_1(v^*) &= 2v_2^*\iint T_1\left((F^0-F)\otimes\mathbb{F}_n+\mathbb{F}_n\otimes(F^0-F)-\frac{1}{\sqrt{n}}\mathbb{F}_n\otimes\mathbb{F}_n\right)dH\\ &\quad+2\iint T_1\left(-\mathbb{G}_n+F\otimes\mathbb{F}_n+\mathbb{F}_n\otimes F+\frac{1}{\sqrt{n}}\mathbb{F}_n\otimes\mathbb{F}_n\right)dH\\ &= -2\iint T_1\mathbb{G}_ndH+2\iint T_1(A\otimes\mathbb{F}_n+\mathbb{F}_n\otimes A)dH\\ &\quad+\frac{2(1-v_2^*)}{\sqrt{n}}\iint T_1(\mathbb{F}_n\otimes\mathbb{F}_n)dH,\end{aligned}$$

其中 $A=v_2^*(F^0-F)+F=(1-b^*)F^0+b^*F^1$。换言之

$$\sqrt{n}[\dot{\mathtt{d}}_1-\dot{\mathtt{d}}_n]_1(v^*)=2\iint\left(-T_1\mathbb{G}_n+T_1\overline{A\otimes\mathbb{F}_n}\right)dH+\frac{1}{\sqrt{n}}\chi_1(\mathbb{F}_n)$$

其中 $\chi_1(\phi)=2(1-v_2^*)\iint T_1(\phi\otimes\phi)dH$ 且 $a\otimes\underline{b}\equiv a\otimes b+b\otimes a$。同理，我们可以计算（参见第 A.1节）

$$\sqrt{n}[\dot{\mathtt{d}}-\dot{\mathtt{d}}_n]_2(v^*)=2\iint\left(-T_2\mathbb{G}_n+B\overline{(F^0-F)\otimes\mathbb{F}_n}+T_2\overline{F\otimes\mathbb{F}_n}\right)dH+\frac{1}{\sqrt{n}}\chi_2(\mathbb{F}_n,\mathbb{G}_n),$$

其中 $B=v_1^*T_1+2v_2^*T_2+T_3$ 和 χ_2 是连续映射。我们记

$$\Psi(\mathbb{F}_n,\mathbb{G}_n)=2\begin{pmatrix}\iint\left(-T_1\mathbb{G}_n+T_1\overline{A\otimes\mathbb{F}_n}\right)dH\\ \iint\left(-T_2\mathbb{G}_n+B\overline{(F^0-F)\otimes\mathbb{F}_n}+T_2\overline{F\otimes\mathbb{F}_n}dH\right)\end{pmatrix}, \tag{33}$$

因此

$$\sqrt{n}(\dot{\mathtt{d}}-\dot{\mathtt{d}}_n)(v^*)=\Psi(\mathbb{F}_n,\mathbb{G}_n)+\frac{1}{\sqrt{n}}\chi(\mathbb{F}_n,\mathbb{G}_n), \tag{34}$$

其中， Ψ 和 χ 是连续映射。从 $B(\mathbb{R}^2) \times B(\mathbb{R})$到 $\mathbb{R}^2$ 的连续性源于 F^0， F^1， T_i和 $\iint dH < \infty$的有界性。根据斯卢茨基定理，我们有

$$\frac{1}{\sqrt{n}}\chi(\mathbb{F}_n, \mathbb{G}_n) \xrightarrow{\mathbb{P}} 0, \quad \text{as} \quad n \to +\infty,$$

最后，引理2给出

$$\sqrt{n}(\dot{\mathrm{d}} - \dot{\mathrm{d}}_n)(v^*) \xrightarrow{d} \Psi(\mathcal{B}_F, \mathcal{B}_G), \quad \text{as} \quad n \to +\infty.$$

注意， Ψ 是一个依赖于 F^0， F^1， G^0和 θ^*的线性映射。于是， $\Psi(\mathcal{B}_F$， $\mathcal{B}_G)$是一个均值为 $\mathcal{N}(0$， $\Lambda)$的中心高斯变量。

8.3.3 协方差矩阵

¨在前几节中，我们已经证明， d_n 依概率收敛到可逆矩阵d，并且

$$\sqrt{n}(\dot{\mathrm{d}} - \dot{\mathrm{d}}_n)(v^*) \xrightarrow{d} \Psi(\mathcal{B}_F, \mathcal{B}_G) = \mathcal{N}(0, \Lambda), \quad \text{as} \quad n \to +\infty.$$

将矩阵 Λ显式写出。利用（33），我们可以写成

$$\Psi(\mathcal{B}_F, \mathcal{B}_G) = \iint M(x,y) \begin{pmatrix} \mathcal{B}_F(x) \\ \mathcal{B}_F(y) \\ \mathcal{B}_G(x,y) \end{pmatrix} dH(x,y),$$

with

$$M(x,y) = 2\begin{pmatrix} T_1(x,y)A(y) & T_1(x,y)A(x) & -T_1(x,y) \\ T_2(x,y)F(y) + B(x,y)(F^0 - F)(y) & T_2(x,y)F(x) + B(x,y)(F^0 - F)(x) & -T_2(x,y) \end{pmatrix},$$

其中 $A = v_2^*(F^0 - F) + F$和 $B = v_1^* T_1 + 2v_2^* T_2 + T_3$。记

$$C(x,y,z,t) := \mathrm{cov}(\mathcal{B}_G(x,y), \mathcal{B}_G(z,t)) = \sum_{k \in \mathbb{Z}} \mathrm{cov}\left(\mathbb{1}_{\{Z_0 \le x, Z_1 \le y\}}, \mathbb{1}_{\{Z_k \le z, Z_{k+1} \le t\}}\right),$$

并注意到 $\mathcal{B}_F(x) = \lim_{y \to \infty} \mathcal{B}_G(x, y)$。因此，向量$\mathcal{T}(x, y)$的协方差矩阵 $:= (\mathcal{B}_F(x)\ \mathcal{B}_F(y)\ \mathcal{B}_G(x, y))^T$为

$$C_{\mathcal{T}}(x,y) := \begin{pmatrix} C(x,\infty,x,\infty) & C(x,\infty,y,\infty) & C(x,\infty,x,y) \\ C(x,\infty,y,\infty) & C(y,\infty,y,\infty) & C(y,\infty,x,y) \\ C(x,\infty,x,y) & C(y,\infty,x,y) & C(x,y,x,y) \end{pmatrix}.$$

于是我们得到 Λ 的如下表示:

$$\Lambda = \iint M(x,y) C_{\mathcal{T}}(x,y) M^T(x,y) dH(x,y).$$

此外，当 n 足够大时，

$$\sqrt{n}(\hat{v}_n - v^*) = \ddot{\mathrm{d}}_n^{-1}\left[\sqrt{n}(\dot{\mathrm{d}} - \dot{\mathrm{d}}_n)(v^*)\right].$$

因此 $\sqrt{n}(\hat{v}_n - v^*) \xrightarrow{d} \mathcal{N}(0, \Gamma)$ 作为 $n \to +\infty$，其渐近协方差矩阵为

$$\Gamma = \ddot{\mathrm{d}}^{-1} \Lambda \ddot{\mathrm{d}}^{-1}.$$

由（30），可写为

$$\ddot{\mathrm{d}}^{-1} = \frac{1/2}{\iint T_1^2 dH \iint T_2^2 dH - \left(\iint T_1 T_2 dH\right)^2} \begin{pmatrix} \iint T_2^2 dH & -\iint T_1 T_2 dH \\ -\iint T_1 T_2 dH & \iint T_1^2 dH \end{pmatrix},$$

这里回顾一下

$$\begin{aligned} T_1 &= G^0 - F^0 \otimes F^0, \\ T_2 &= (F - F^0) \otimes (F - F^0) = r^{*2}(F^1 - F^0) \otimes (F^1 - F^0), \\ T_3 &= G - F \otimes F = -v_1^* T_1 - v_2^* T_2. \end{aligned}$$

最后，记 $D^* = Dg(v^*) = [Dh(\theta^*)]^{-1}$，我们有 $\Sigma = D^* \Gamma (D^*)^T$，因此

$$\Sigma = \iint \left(D^* \ddot{\mathrm{d}}^{-1} M(x,y)\right) C_{\mathcal{T}}(x,y) \left(D^* \ddot{\mathrm{d}}^{-1} M(x,y)\right)^T dH(x,y),$$

其中 $\ddot{\mathrm{d}}^{-1}$，M，$C_{\mathcal{T}}$已在本节上文定义，且 D^* 在（28）中详细说明。 □

8.4 定理3的证明

回顾一下

$$\hat{F}_n^1(x) = \frac{1}{\hat{r}_n}\left(\hat{F}_n(x) - \hat{p}_n F^0(x)\right), \quad x \in E.$$

那么，为简洁起见，省略对 x 的依赖关系，可得

$$\begin{aligned} \hat{F}_n^1 - F^1 &= \frac{1}{r^*}(\hat{F}_n - F) + \left(\frac{1}{\hat{r}_n} - \frac{1}{r^*}\right)\hat{F}_n + \left(\frac{p^*}{r^*} - \frac{\hat{p}_n}{\hat{r}_n}\right) F^0 \\ &= \frac{1}{r^*}(\hat{F}_n - F) + \left(\frac{\hat{\beta}_n}{\hat{\alpha}_n} - \frac{\beta^*}{\alpha^*}\right)(\hat{F}_n - F^0), \end{aligned}$$

因为 $1/r = 1 + p/r = 1 + \beta/\alpha$。因此，回想一下 $\mathbb{F}_n = \sqrt{n}\left(\hat{F}_n - F\right)$，

$$\sqrt{n}\left(\hat{F}_n^1 - F^1\right) = \frac{1}{r^*}\mathbb{F}_n + \sqrt{n}\left(\frac{\hat{\beta}_n}{\hat{\alpha}_n} - \frac{\beta^*}{\alpha^*}\right)\left(F - F^0 + \frac{\mathbb{F}_n}{\sqrt{n}}\right). \tag{35}$$

我们将使用证明渐近正态性时已经采用的变量变换（见第 8.3节）。回想一下，参数 $\theta = (\alpha, \beta)$ 通过以下同胚映射变换为新参数 $v = (v_1, v_2)$：h 和 $g = h^{-1}$：

$$h: \begin{array}{ccl} \Theta & \to & h(\Theta) = V \\ \theta & \mapsto & v = (-1 + 2r - rb, -\frac{b}{r} + 1) = \left(-\frac{\beta(1-\alpha)}{\alpha+\beta}, \frac{\beta}{\alpha}(\alpha + \beta - 1)\right), \end{array}$$

and

$$g: \begin{array}{ccl} V & \to & \Theta \\ v & \mapsto & \theta = \left(\frac{\beta(1-\beta)}{\beta - v_2}, \beta\right), \quad \text{where } \beta = \sqrt{v_1 v_2 + v_2 - v_1}. \end{array}$$

现在定义比值函数 $\varphi_1: \Theta \to \mathbb{R}^{+*}$ 它执行 $(\alpha, \beta) \mapsto \frac{\beta}{\alpha}$ 映射。另定义$\varphi_2 = \varphi_1 \circ g: V \to \mathbb{R}^{+*}$。采用这一记号，我们可以写成

$$\frac{\beta}{\alpha} = \varphi_1(\theta) = \varphi_1 \circ g(v) = \varphi_2(v).$$

回到（35），我们得到

$$\sqrt{n}(\hat{F}_n^1 - F^1) = \frac{1}{r^*}\mathbb{F}_n + \left[\sqrt{n}\left(\varphi_2(\hat{v}_n) - \varphi_2(v^*)\right)\right]\left(F - F^0 + \frac{\mathbb{F}_n}{\sqrt{n}}\right). \tag{36}$$

我们知道（$\mathbb{F}_n$，$\mathbb{G}_n$）在 $n \to +\infty$时依分布收敛于（$\mathcal{B}_F$，$\mathcal{B}_G$），并且我们已经在第8.3 节证明了，对于足够大的 n （只要 $\ddot{\mathtt{d}}_n$ 可逆），

$$\begin{aligned}\sqrt{n}(\hat{v}_n - v^*) &= (\ddot{\mathtt{d}}_n^{-1})\left[\sqrt{n}(\dot{\mathtt{d}} - \dot{\mathtt{d}}_n)(v^*)\right] \\ &= (\ddot{\mathtt{d}}_n^{-1})\left[\Psi(\mathbb{F}_n, \mathbb{G}_n) + \frac{1}{\sqrt{n}}\chi(\mathbb{F}_n, \mathbb{G}_n)\right],\end{aligned}$$

其在 $n \to +\infty$时依分布收敛于 $\ddot{\mathtt{d}}^{-1}\Psi(\mathcal{B}_F, \mathcal{B}_G)$。由于 v^*是V的内点，当 n 足够大时，线段 $[\hat{v}_n, v^*]$ 属于 V，而实值函数的泰勒定理给出：

$$\varphi_2(\hat{v}_n) - \varphi_2(v^*) = D\varphi_2(\tilde{v}_n)(\hat{v}_n - v^*),$$

对于某个 $\tilde{v}_n \in [\hat{v}_n, v^*]$成立。此外，$\varphi_2$ 的光滑性（连续性）意味着 $D\varphi_2(\tilde{v}_n)$几乎处处收敛于 $D\varphi_2(v^*)$。因此，当 n 足够大时，我们有

$$\sqrt{n}(\varphi_2(\hat{v}_n) - \varphi_2(v^*)) = (D\varphi_2(\tilde{v}_n)\ddot{\mathtt{d}}_n^{-1})\left[\Psi(\mathbb{F}_n, \mathbb{G}_n) + \frac{1}{\sqrt{n}}\chi(\mathbb{F}_n, \mathbb{G}_n)\right],$$

其在 $n \to +\infty$时依分布收敛于 $D\varphi_2(v^*)\ddot{\mathtt{d}}^{-1}\Psi(\mathcal{B}_F, \mathcal{B}_G)$。然后，根据斯卢茨基定理，我们有

$$R_n := \frac{1}{\sqrt{n}}\left[\sqrt{n}\left(\varphi_2(\hat{v}_n) - \varphi_2(v^*)\right)\right]\mathbb{F}_n \xrightarrow{d} 0, \quad \text{as} \quad n \to +\infty.$$

因此，表达式（36）变为

$$\begin{aligned}\sqrt{n}(\hat{F}_n^1 - F^1) &= \frac{1}{r^*}\mathbb{F}_n + \left[\sqrt{n}\left(\varphi_2(\hat{v}_n) - \varphi_2(v^*)\right)\right](F - F^0) + R_n \\ &= \frac{1}{r^*}\mathbb{F}_n + (D\varphi_2(\tilde{v}_n)\ddot{\mathtt{d}}_n^{-1})\left[\Psi(\mathbb{F}_n, \mathbb{G}_n) + \frac{1}{\sqrt{n}}\chi(\mathbb{F}_n, \mathbb{G}_n)\right](F - F^0) + R_n,\end{aligned}$$

由此得到所需的依分布收敛

$$\sqrt{n}(\hat{F}_n^1 - F^1) \xrightarrow{d} \underbrace{\frac{1}{r^*}\mathcal{B}_F + D\varphi_2(v^*)\ddot{\mathtt{d}}^{-1}\Psi(\mathcal{B}_F, \mathcal{B}_G)(F - F^0)}_{\mathcal{G}}, \quad \text{as} \quad n \to +\infty.$$

□

8.5 定理4的证明

• 定理的证明 4 关于 θ。首先，我们来估计 $\|\theta_n - \theta^*\|$，其上界为 $\|\hat{v}_n - v^*\|$，相差一个乘法常数。由于 $v^* \in \mathrm{Int}(V)$，存在一个开球 B，其中心为 v^*，使得对所有 $v \in B$，线段 $[v, v^*]$ 都包含于 V。于是由平均值不等式可得

$$\|g(v) - g(v^*)\| \le \|Dg\|_\infty \|v - v^*\| \le \frac{2}{\delta^2}\|v - v^*\|.$$

此外，在紧集 $V\backslash B$ 上，函数 $\|g(v)-g(v^*)\|/\|v-v^*\|$ 连续，因此有界。最后，存在一个常数 $C(\delta)>0$，使得对所有 $v\in V$

$$\|g(v)-g(v^*)\|\le C(\delta)\|v-v^*\|.$$

因此 $\|\theta_n-\theta^*\|=\|g(\hat{v}_n)-g(v^*)\|\le C(\delta)\|\hat{v}_n-v^*\|$ 只需对序列 $(\sqrt{n}\mathbb{E}\|\hat{v}_n-v^*\|)_{n\ge1}$ 进行界定即可。但 (29) 给出

$$\ddot{\mathrm{d}}_n(\hat{v}_n-v^*)=(\dot{\mathrm{d}}-\dot{\mathrm{d}}_n)(v^*),$$

并且由 (34) 可知

$$\sqrt{n}(\dot{\mathrm{d}}-\dot{\mathrm{d}}_n)(v^*)=\Psi(\mathbb{F}_n,\mathbb{G}_n)+\frac{1}{\sqrt{n}}\chi(\mathbb{F}_n,\mathbb{G}_n).$$

现在我们使用以下关于经验过程 $\mathbb{F}_n$ 和 $\mathbb{G}_n$的引理。

引理 3。回顾 $\mathbb{F}_n:=\sqrt{n}(F_n-F)$ 和 $\mathbb{G}_n:=\sqrt{n}(G_n-G)$。假设 $(Z_i)_{i\ge0}$ 和 $(Z_i,Z_{i+1})_{i\ge0}$ 满足 α-混合条件，且 $\alpha(n)=O(n^{-a})$ 对于 $a>4$ 成立（系数定义见 (16) 中的 α）。则有以下性质：

i) 以概率 *1*：$\forall n\ge1,\quad \|\mathbb{F}_n-\mathcal{B}\|_\infty\le C(\log n)^{-\lambda}$，其中 $\lambda>0$ 为某个常数，$\mathcal{B}$ 是布朗桥。

ii）对所有正整数 q，存在 $C_F(q)>0$ 使得：$\forall n\ge2,\ \mathbb{E}\left(\|\mathbb{F}_n\|_\infty^q\right)\le C_F(q)$.

iii）对所有正整数 q，存在 $C_G(q)>0$ 使得：$\forall n\ge2,\ \mathbb{E}\|\mathbb{G}_n\|_\infty\le C_G(q)$。

证明。 i）记 $\mathbb{K}(.,.)$ 为基弗过程。Dhompongsa Dhompongsa (1984) 的定理 3 保证，以概率 1，对所有 $n\ge2$

$$\sup_{s\in\mathbb{R}}|\sqrt{n}\mathbb{F}_n(s)-\mathbb{K}(s,n)|\le C\sqrt{n}(\log n)^{-\lambda},$$

对某个 $\lambda>0$ 成立。要得出结论，只需记住 $\mathcal{B}:=\mathbb{K}(.,n)/\sqrt{n}$ 是布朗桥，其分布不依赖于 n。

ii）已知 $\|\mathcal{B}\|_\infty$ 服从柯尔莫哥洛夫-斯米尔诺夫分布，其累积分布函数为 $1-2\sum_{k=1}^{\infty}(-1)^{k-1}\exp(-2k^2x^2)$（例如，参见 Shorack 和 Wellner (2009)）。该分布具有有限矩：$\mathbb{E}(\|\mathcal{B}\|_\infty^q)=q2^{-q/2}\Gamma(q/2)\eta(q)<\infty$。于是，对所有 $n\ge2$,

$$\begin{aligned}\mathbb{E}(\|\mathbb{F}_n\|_\infty^q) &\le 2^{q-1}\mathbb{E}(\|\mathcal{B}\|_\infty^q)+2^{q-1}C^q(\log n)^{-\lambda q}\\ &\le \underbrace{q2^{q/2-1}\Gamma(q/2)\eta(q)+2^{q-1}C^q(\log 2)^{-\lambda q}}_{C_F(q)}.\end{aligned}$$

iii）考虑过程 $R_i=(Z_i,Z_{i+1})$ 时，证明相同。事实上，我们使用了一个适用于取值于 $\mathbb{R}^d$ 的过程的定理，只要混合率满足 $a>d+2$。二维基弗过程的定义可见 Philipp and Pinzur（1980）。同样，$\mathcal{B}:=\mathbb{K}(.,n)/\sqrt{n}$ 的分布不依赖于 n，且是二维布朗桥。二维布朗桥上确界的分布不如一维情形那样广为人知，但 Adler and Brown（1986）证明了 $\mathbb{P}(\|\mathcal{B}\|_\infty>\lambda)\le C\lambda^2e^{-2\lambda^2}$，这足以确保有限矩，即 $\mathbb{E}(\|\mathcal{B}\|_\infty^q)<\infty$.

□

现在，我们可以对 $\Psi_1(\mathbb{F}_n, \mathbb{G}_n)$ 进行界定。我们计算

$$\begin{aligned}\mathbb{E}|\Psi_1(\mathbb{F}_n)| &= 2\mathbb{E}\left|\iint \left(-T_1\mathbb{G}_n + T_1\overline{A\otimes \mathbb{F}_n}\right) dH\right| \\ &\leq 2\mathbb{E}\|\mathbb{G}_n\|_\infty \iint |T_1|dH + 2\mathbb{E}\|\mathbb{F}_n\|_\infty \iint |T_1(x,y)|(|A(x)|+|A(y)|)dH(x,y) \\ &\leq 2C_G(1) + 8C_F(1).\end{aligned}$$

对 Ψ_2 和 χ_1, χ_2采用相同的方法，我们得到

$$\mathbb{E}\|\sqrt{n}(\dot{\mathtt{d}} - \dot{\mathtt{d}}_n)(v^*)\| \leq C,$$

然后

$$\sqrt{n}\mathbb{E}\|\ddot{\mathtt{d}}_n(\hat{v}_n - v^*)\| \leq C. \tag{37}$$

现在 $\hat{v}_n - v^* = \ddot{\mathtt{d}}^{-1}\ddot{\mathtt{d}}(\hat{v}_n - v^*)$，因为根据 **LinInd**，$\ddot{\mathtt{d}}$ 可逆。为简单起见，我们用 x_n 表示随机序列 $x_n = (\hat{v}_n - v^*)$。于是得到

$$\hat{v}_n - v^* = x_n = \ddot{\mathtt{d}}^{-1}\ddot{\mathtt{d}}x_n = \ddot{\mathtt{d}}^{-1}\left(\ddot{\mathtt{d}}_n x_n + (\ddot{\mathtt{d}} - \ddot{\mathtt{d}}_n)x_n\right),$$

and

$$\mathbb{E}\|\hat{v}_n - v^*\| \leq \|\ddot{\mathtt{d}}^{-1}\|\left(\mathbb{E}\|\ddot{\mathtt{d}}_n x_n\| + \mathbb{E}(\|\ddot{\mathtt{d}} - \ddot{\mathtt{d}}_n\|\|x_n\|)\right),$$

其中，我们用 $\|M\|$表示任意矩阵 M的谱范数。注意，对所有 $\theta \in \Theta$, $v = h(\theta) \in [-1,0]\times[-\frac{1}{\delta}, 1]$, 以概率 1，对所有 $n \geq 1$: $|[x_n]_1| \leq 1$ 以及 $|[x_n]_2| \leq 1+\delta^{-1}$.利用这一点和（37），我们得到

$$\mathbb{E}\|\hat{v}_n - v^*\| \leq \|\ddot{\mathtt{d}}^{-1}\|\left(\frac{C}{\sqrt{n}} + c\delta^{-1}\mathbb{E}(\|\ddot{\mathtt{d}} - \ddot{\mathtt{d}}_n\|)\right). \tag{38}$$

根据第 8.3.1 节，我们知道 $\ddot{\mathtt{d}}_n - \ddot{\mathtt{d}} = \frac{1}{\sqrt{n}}\Phi(\mathbb{F}_n)$，其中 $[\Phi(\mathbb{F}_n)]_{1,1} = 0$ 并且

$$\begin{aligned}[\Phi(\mathbb{F}_n)]_{1,2} &= [\Phi(\mathbb{F}_n)]_{2,1} = 2\iint T_1\sqrt{n}(\hat{T}_2 - T_2)dH, \\ [\Phi(\mathbb{F}_n)]_{2,2} &= \frac{2}{\sqrt{n}}\iint \left(\sqrt{n}(\hat{T}_2 - T_2)\right)^2 dH + 4\iint T_2\sqrt{n}(\hat{T}_2 - T_2)dH,\end{aligned}$$

其中 $\sqrt{n}(T_2 - \hat{T}_2) = \overline{\mathbb{F}_n \otimes (F^0 - F)} - \frac{1}{\sqrt{n}}\mathbb{F}_n \otimes \mathbb{F}_n$. 因此

$$\begin{aligned}[|\Phi(\mathbb{F}_n)|]_{1,2} &\leq 2\iint |T_1|\left|\overline{\mathbb{F}_n \otimes (F^0 - F)} - \frac{1}{\sqrt{n}}\mathbb{F}_n \otimes \mathbb{F}_n\right| dH \\ &\leq 2\|\mathbb{F}_n\|_\infty + \frac{2}{\sqrt{n}}\|\mathbb{F}_n\|_\infty^2.\end{aligned}$$

从而 $\mathbb{E}[|\Phi(\mathbb{F}_n)|]_{1,\,2} \leq 2C_F(1) + 2C_F(2)$，参见引理 3。同理，我们可以证明序列 $\mathbb{E}[|\Phi(\mathbb{F}_n)|]_{2,\,2}$ 有一个正的有限常数作为上界。这保证了$\mathbb{E}(\|\ddot{\mathtt{d}} - \ddot{\mathtt{d}}_n\|) \leq C'/\sqrt{n}$ 并且，由式(38)，最终得到 $\mathbb{E}\|\hat{v}_n - v^*\| = O(1/\sqrt{n})$。

- 定理证明4 关于 F^1。我们知道，对于所有 $n \geq 1$

$$\hat{F}_n^1 - F^1 = \frac{1}{r^*}(\hat{F}_n - F) + \left(\frac{\hat{\beta}_n}{\hat{\alpha}_n} - \frac{\beta^*}{\alpha^*}\right)(\hat{F}_n - F^0).$$

利用 $\mathbb{F}_n = \sqrt{n}(\hat{F}_n - F)$和 $\|\hat{F}_n - F^0\|_\infty \leq 1$，可得

$$\begin{aligned}\|\hat{F}_n^1 - F^1\|_\infty &\leq \frac{1}{r^*}\frac{\|\mathbb{F}_n\|_\infty}{\sqrt{n}} + \left|\frac{\hat{\beta}_n}{\hat{\alpha}_n} - \frac{\beta^*}{\alpha^*}\right| \\ &\leq \frac{1}{r^*}\frac{\|\mathbb{F}_n\|_\infty}{\sqrt{n}} + \frac{\sqrt{2}}{\delta^2}\|\hat{\theta}_n - \theta\|.\end{aligned}$$

因此，再次使用引理3，可得

$$\begin{aligned}\mathbb{E}\|\hat{F}_n^1 - F^1\|_\infty &\leq \frac{1}{r^*}\frac{1}{\sqrt{n}}\mathbb{E}\|\mathbb{F}_n\|_\infty + \frac{\sqrt{2}}{\delta^2}\mathbb{E}\|\hat{\theta}_n - \theta\| \\ &\leq \left(\frac{C_F(1)}{r^*} + \frac{C''}{\delta^2}\right)\frac{1}{\sqrt{n}},\end{aligned}$$

证明至此完成。 □

致谢

作者谨向 Florence Merlevede 教授和 J´erˆome Dedecker 教授表示诚挚感谢，感谢他们就 $(\mathbb{F}_n, \mathbb{G}_n)$ 的收敛性以及过程的弱依赖性提出了富有洞见的意见。

资助与拨款

Claire Lacour 和 Pierre Vandekerkhove 谨此说明，本研究得到 CNES 在 BIOSWOT-AdAC 项目和 MIO Axes Transverses 框架下的支持。Pierre Vandekerkhove 还感谢 ACTIONS 研究讲席提供的支持；该讲席由 RiskFoundation 主持发起，RiskFoundation 是 BNP Paribas Cardif 与法国精算师协会共同倡议的项目。

参考文献

R. J. Adler 和 L. D. Brown。经验过程上确界的尾部行为。《概率年刊》，第1–30页，1986年。G. Alexandrovich、H. Holzmann 和 A. Leister。隐马尔可夫模型的参数识别与极大似然估计。《生物统计学》，103(2):423–434，2016。K. B. Athreya 和 V. Roy。一般格里文科–坎特利定理。《Stat》，5(1):306–311，2016。L. E. Baum 和 T. Petrie。有限状态马尔可夫链的概率函数的统计推断。《数学统计年鉴》，37:1554–1563，1966年。Y. Benjamini 和 Y. Hochberg。控制假发现率：一种实用且有效的多重检验方法。《英国皇家统计学会杂志：B辑（方法论）》，57(1):289–300，1995年。P. J. Bickel、Y. Ritov 和 R. Tobias。一般隐马尔可夫模型极大似然估计量的渐近正态性。《统计年刊》，26(4):1614–1635，1998年。

L. Bordes、C. Delmas 和 P. Vandekerkhove。一个成分已知的双成分混合模型的半参数估计。《斯堪的纳维亚统计学杂志》，33(4):733–752，2006a。L. Bordes、S. Mottelet 和 P. Vandekerkhove。双成分混合模型的半参数估计。《统计年刊》，34(3):1204–1232，2006b。L. Bordes、D. Chauveau 和 P. Vandekerkhove。用于半参数混合模型的随机 EM算法。《计算统计与数据分析》，51(11):5429–5443，2007。《混合模型进展》。N. Bouguila、W. Fan 和 M. Manar。《隐马尔可夫模型及其应用》。施普林格，2022年。R. C. Bradley。强混合条件的基本性质：综述与若干开放问题。《概率调查》，2005年。C. Butucea 和 P. Vandekerkhove。对称分布的半参数混合模型。《斯堪的纳维亚统计学杂志》，41(1):227–239，2014年。C. Butucea、R. Nguyepe Tzoumpe 和 P. Vandekerkhove。半参数拓扑混合模型。《伯努利》，23(2):825–862，2017年。O. Capp´e、E. Moulines 和 T. Ryd´en。《隐马尔可夫模型中的推断》。施普林格统计学丛书，2005年。S. Csorgo。经验特征函数的极限行为。《概率年刊》，9(1):130–144，1981年。D. Dacunha-Castelle 和 M. Duflo。《概率与统计》。第二卷。施普林格出版社，纽约，1986年。David McHale 从法文译出。S. Dass、P. Datta 和 S. Namim, Akbar。使用隐马尔可夫模型进行攻击预测。载于《2021年IEEE第45届年度计算机、软件与应用会议》（COMPSAC），第1695–1702页，2021年。Y. De Castro、E. Gassiat 和 C. Lacour。非参数隐马尔可夫模型的极小极大自适应估计。《机器学习研究杂志》，17(1):3842–3884，2016。J. Dedecker。间歇映射的经验中心极限定理。《概率论及其相关领域》，148(1):177–195，2010年。J. Dedecker、P. Doukhan、G. Lang、L. R. Jos´e Rafael、S. Louhichi 和 C. Prieur。弱依赖。载于《弱依赖：附示例与应用》，第9–20页。施普林格，2007年。J. Dedecker、E. Rio 和 F. Merlevede。对 $\mathbb{R}^d$中绝对正则序列的经验分布函数进行强逼近。《电子概率学杂志》，19:1–56，2014年。S. Dhompongsa。关于弱依赖随机向量经验过程几乎必然逼近的注记。《横滨数学杂志》，32(1&2):113–121，1984年。

D. Donoho 和 J. Jin。用于检测稀疏异质混合分布的高阶检验。《统计年刊》，32(3)：962 – 994，2004 年。A. Feuerverger 和 R. A. Mureika。经验特征函数及其应用。《统计年刊》，第88–97页，1977年。´E. Gassiat。非参数成分的混合与隐马尔可夫模型。载于《混合分析手册》——编者：S. Fruhwirth-Schnatter、G. Celeux、C.P. Robert——CRC出版社，2019年。´E. Gassiat 和 J. Rousseau。非参数有限平移隐马尔可夫模型及其扩展。《伯努利》，22(1)：193–2012，2016。S. Gaucher、G. Blanchard 和 F. Chazal。监督式污染检测及其在流式细胞术中的应用。《生物统计学》，2025年。E. J. Gilbert。关于有限马尔可夫链函数的可识别性问题。《数学统计年鉴》，30：688–697，1959年。P. Hall 和 X.-H. Zhou。多元混合分布中成分分布的非参数估计。《统计年刊》，31(1)：201–224，2003年。P. Holgado、V. Villagra 和 L. Vazquez。基于隐马尔可夫模型的实时多步攻击预测。IEEE《可靠与安全计算汇刊》，17，2020年。D. R. Hunter、S. Wang 和 T. P. Hettmansperger。对称分布混合模型的推断。《统计年刊》，第224–251页，2007年。M. Korczynski 和 A. Duda。利用马尔可夫链指纹识别对加密流量进行分类。载于IEEE国际计算机通信会议（INFOCOM'14），第1–9页，2014年。B. G. Leroux。隐马尔可夫模型的最大似然估计。Sochastic Processes and their Applications，40：127–143，1992年。M. Levine、D. R. Hunter 和 D. Chauveau。多元混合分布的最大平滑似然。《生物统计学》，98(2)：403–416，2011年。G. J. McLachlan、R. W. Bean 和 L. B.-T. Jones。用于多类别微阵列差异基因表达的正态混合方法的简单实现。《生物信息学》，22(13)：1608–1615，2006年。X. Milhaud、D. Pommeret、Y. Salhi 和 P. Vandekerkhove。基于污染的 *k*样本聚类。《机器学习研究杂志》，2024年a。X. Milhaud、D. Pommeret、Y. Salhi 和 P. Vandekerkhove。双样本污染模型检验。《伯努利》，30(1)：170 – 197，2024年b。R. K. Patra 和 B. Sen。双成分混合模型的估计及其在多重检验中的应用。《皇家统计学会杂志：B辑：统计方法学》，第869–893页，2016。

W. Philipp 和 L. Pinzur。多元经验过程的几乎必然逼近定理。Zeitschrift f¨ur Wahrscheinlichkeitstheorie und verwandte Gebiete，54(1)：1–13，1980年。R. Podlaski 和 F. A. Roesch。不同优势种比例的双龄组林分直径分布建模：双成分混合模型方法。Mathematical biosciences，249：60–74，2014年。D. Pommeret 和 P. Vandekerkhove。污染模型中的半参数密度检验。Electronic Journal of Statistics，13(2)：4743 – 4793，2019年。E. Rio。弱相依随机过程的渐近理论，第80卷。施普林格，2017年。T. Ryd´en。隐马尔可夫模型的一致且渐近正态的参数估计值。《统计年刊》，22(4)：1884–1895，1994年。Z. Shen、M. Levine 和 Z. Shang。利用MM算法估计具有已知成分的双成分半参数密度混合模型。Electronic journal of statistics，12：1181–1209，2018年。G. R. Shorack 和 J. A. Wellner。应用于统计学的经验过程。SIAM，2009年。A. W. van der Vaart。渐近统计，第3卷《剑桥统计与概率数学丛书》。剑桥大学出版社，剑桥，1998年。P. Vandekerkhove。隐马尔可夫模型的混合模型的一致且渐近正态分布估计。《伯努利》，11(1)：103–129，2005年。M. G. Walker、M. Mateo、E. W. Olszewski、B. Sen 和 M. Woodroofe。矮椭球星系中的洁净运动学样本：在存在污染时评估成员资格并估计分布参数的算法。《The Astronomical Journal》，137(2)：3109，2009年。H. Werner、H. Holzmann 和 P. Vandekerkhove。半参数回归混合模型上确界范数下的自适应估计。Electronic Journal of Statistics，14(1)：1816 – 1871，2020年。S. Xiang、W. Yao 和 G. Yang。有限混合模型半参数扩展概述。Statistical Science，34(3)：391 – 404，2019年。

A 附录

A.1 $\dot{\mathtt{d}} - \dot{\mathtt{d}}_n$ 的计算详情

我们从

$$[\dot{\mathtt{d}} - \dot{\mathtt{d}}_n]_2(v) = 2\iint (v_1 T_1 + v_2 T_2 + T_3)(T_2 - \hat{T}_2) dH + 2v_2 \iint \hat{T}_2 (T_2 - \hat{T}_2) dH + 2\iint \hat{T}_2 (T_3 - \hat{T}_3) dH.$$

我们可以写成

$$\begin{aligned}
[\dot{\mathrm{d}}-\dot{\mathrm{d}}_n]_2(v^*) &= 2\iint \Delta(\theta^*,.)(T_2-\hat{T}_2)dH + 2v_2^*\iint (\hat{T}_2-T_2+T_2)(T_2-\hat{T}_2)dH \\
&\quad +2\iint (\hat{T}_2-T_2+T_2)(T_3-\hat{T}_3)dH \\
&= 2\iint \left[\Delta(\theta^*,.)+v_2^*T_2\right](T_2-\hat{T}_2)dH + 2\iint T_2(T_3-\hat{T}_3)dH \\
&\quad +2v_2^*\iint (\hat{T}_2-T_2)(T_2-\hat{T}_2)dH + 2\iint (\hat{T}_2-T_2)(T_3-\hat{T}_3)dH \\
&=: 2I_1+2I_2+2I_3+2I_4.
\end{aligned}$$

记$a\otimes b=\overline{a\otimes b}+b\otimes a$，等式（31）和（32）变为

$$\sqrt{n}(T_2-\hat{T}_2) = \overline{(F^0-F)\otimes\mathbb{F}_n}-\frac{1}{\sqrt{n}}\mathbb{F}_n\otimes\mathbb{F}_n, \tag{39}$$

$$\sqrt{n}(T_3-\hat{T}_3) = -\mathbb{G}_n+\overline{F\otimes\mathbb{F}_n}+\frac{1}{\sqrt{n}}\mathbb{F}_n\otimes\mathbb{F}_n. \tag{40}$$

根据（39）可得

$$\begin{aligned}
\sqrt{n}I_1 &= \iint \left[\Delta(\theta^*,.)+v_2^*T_2\right](T_2-\hat{T}_2)dH \\
&= \iint \left[\Delta(\theta^*,.)+v_2^*T_2\right]\left(\overline{(F^0-F)\otimes\mathbb{F}_n}-\frac{1}{\sqrt{n}}\mathbb{F}_n\otimes\mathbb{F}_n\right)dH \\
&= \iint B\left(\overline{(F^0-F)\otimes\mathbb{F}_n}-\frac{1}{\sqrt{n}}\mathbb{F}_n\otimes\mathbb{F}_n\right)dH,
\end{aligned}$$

其中 $B=v_1^*T_1+2v_2^*T_2+T_3$。现在根据（40）

$$\sqrt{n}I_2=\iint T_2(T_3-\hat{T}_3)dH=\iint T_2\left(-\mathbb{G}_n+\overline{F\otimes\mathbb{F}_n}+\frac{1}{\sqrt{n}}\mathbb{F}_n\otimes\mathbb{F}_n\right)dH.$$

对于第三项，利用（39）可写为

$$\sqrt{n}I_3=-v_2^*\iint (T_2-\hat{T}_2)^2dH=-\frac{v_2^*}{\sqrt{n}}\iint\left(\overline{(F^0-F)\otimes\mathbb{F}_n}-\frac{1}{\sqrt{n}}\mathbb{F}_n\otimes\mathbb{F}_n\right)^2dH.$$

最后，根据（40）可得

$$\begin{aligned}
\sqrt{n}I_4 &= \iint (\hat{T}_2-T_2)(T_3-\hat{T}_3)dH \\
&= \frac{1}{\sqrt{n}}\iint\left(\overline{(F^0-F)\otimes\mathbb{F}_n}-\frac{1}{\sqrt{n}}\mathbb{F}_n\otimes\mathbb{F}_n\right)\left(\mathbb{G}_n-\overline{F\otimes\mathbb{F}_n}-\frac{1}{\sqrt{n}}\mathbb{F}_n\otimes\mathbb{F}_n\right)dH.
\end{aligned}$$

合并所有项，得到

$$\begin{aligned}
\frac{\sqrt{n}}{2}[\dot{\mathrm{d}}-\dot{\mathrm{d}}_n]_2(v^*) &= \iint B\overline{(F^0-F)\otimes\mathbb{F}_n}dH+\iint T_2\left(-\mathbb{G}_n+\overline{F\otimes\mathbb{F}_n}\right)dH+\frac{1}{\sqrt{n}}\chi_2(\mathbb{F}_n,\mathbb{G}_n) \\
&= \iint\left(-T_2\mathbb{G}_n+B\overline{(F^0-F)\otimes\mathbb{F}_n}+T_2\overline{F\otimes\mathbb{F}_n}\right)dH+\frac{1}{\sqrt{n}}\chi_2(\mathbb{F}_n,\mathbb{G}_n),
\end{aligned}$$

由此结束我们的 $\dot{\mathrm{d}}-\dot{\mathrm{d}}_n$ 分析。

B 附录

B.1 模型 (S0)$_{strong}$ 和 (S0)$_{weak}$的 $\hat{\theta}_n$ 与 $\tilde{\theta}_n$ 的类贝尔态区制

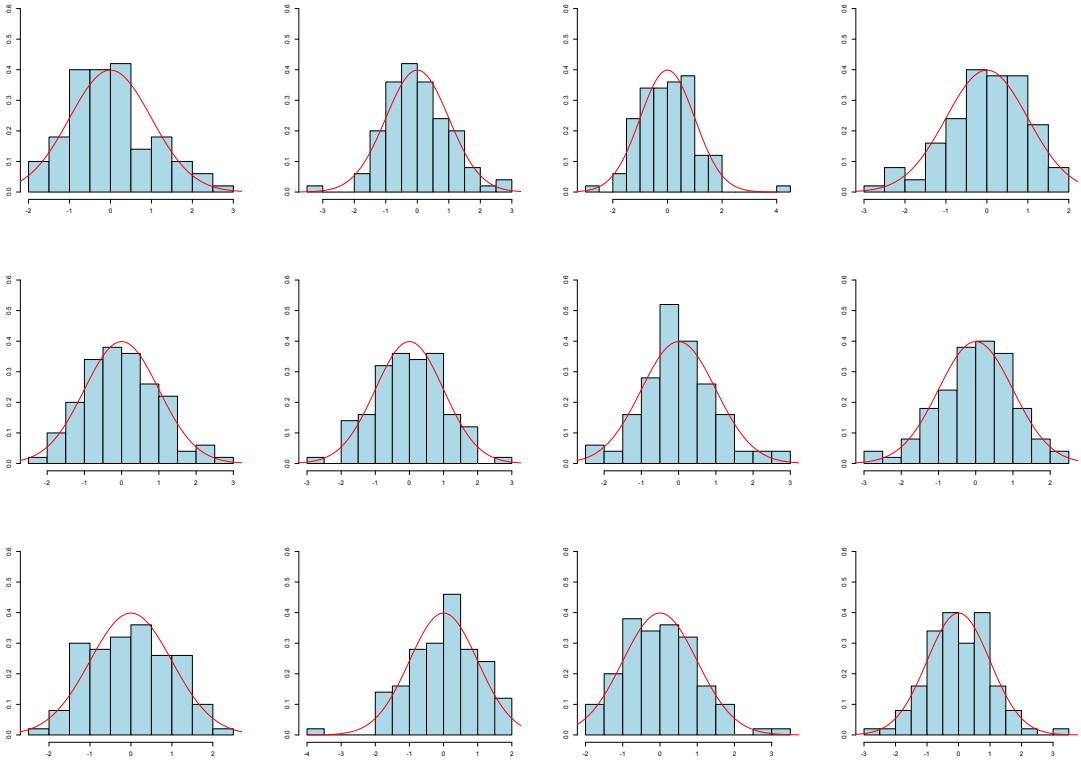


图11：模型 (S0)$_{strong}$下 $(\hat{\alpha},\ \hat{\beta})$ $\bar{\sqrt{n}}$归一化分布的直方图。第一行 $n = 1,000$，第二行 $n = 3,000$，第三行 $n = 5,000$。第一列 $\hat{\alpha}_n$，第二列 $\hat{\beta}_n$，第三列 $\tilde{\alpha}_n$，第四列 $\tilde{\beta}_n$。

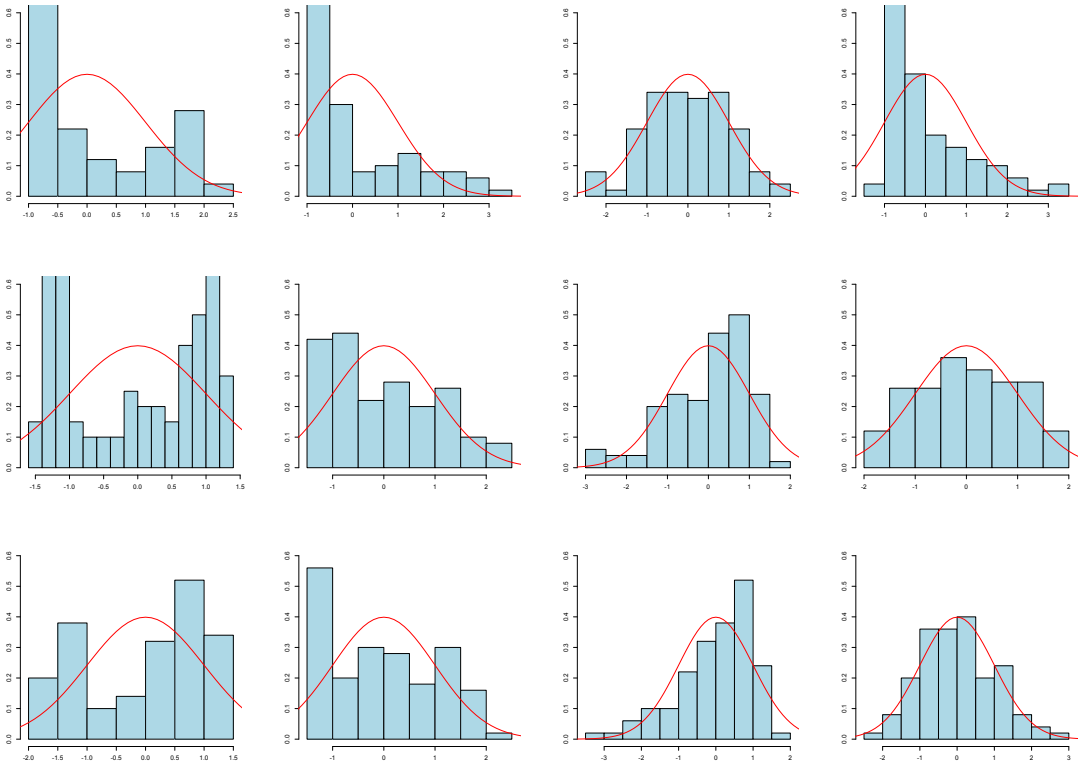


图12：模型 (S0)$_{weak}$ 下 $(\alpha,\ \beta)$ $\sqrt{n}$-标准化估计量的分布直方图。第一行 $n = 1,000$，第二行 $n = 3,000$，第三行 $n = 5,000$。第一列 $\hat{\alpha}_n$，第二 $\hat{\ }$~列 β_n，第三列 $\tilde{\alpha}_n$，第四列 β_n。

B.2 $\hat{\theta}_n$ 和 $\tilde{\theta}_n$ 模型（S1-4）的类贝尔态区制

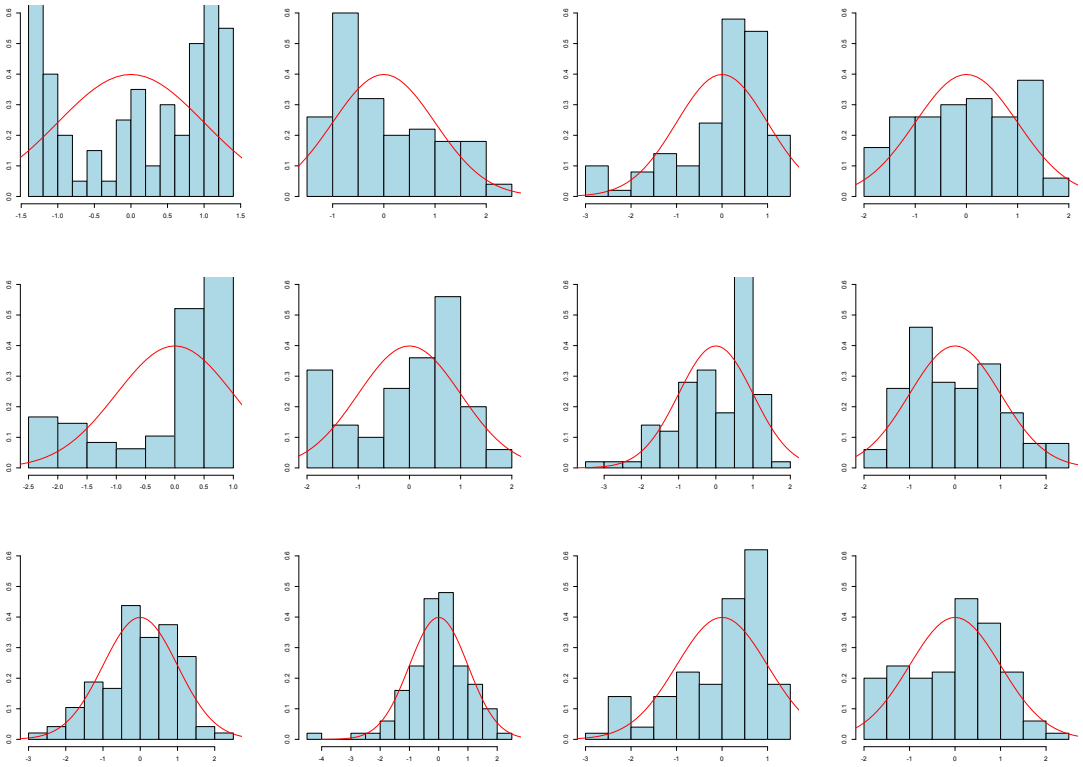


图13：（S1）模型下（α，β）$\sqrt{n}$标准化估计量的分布直方图。第一行 $n = 5,000$，第二行 $n = 10,000$，第三行 $n = 20,000$。第一列 $\hat{\alpha}_n$，第二 $\hat{\beta}_n$，第三列 $\tilde{\alpha}_n$，第四列 $\tilde{\beta}_n$。

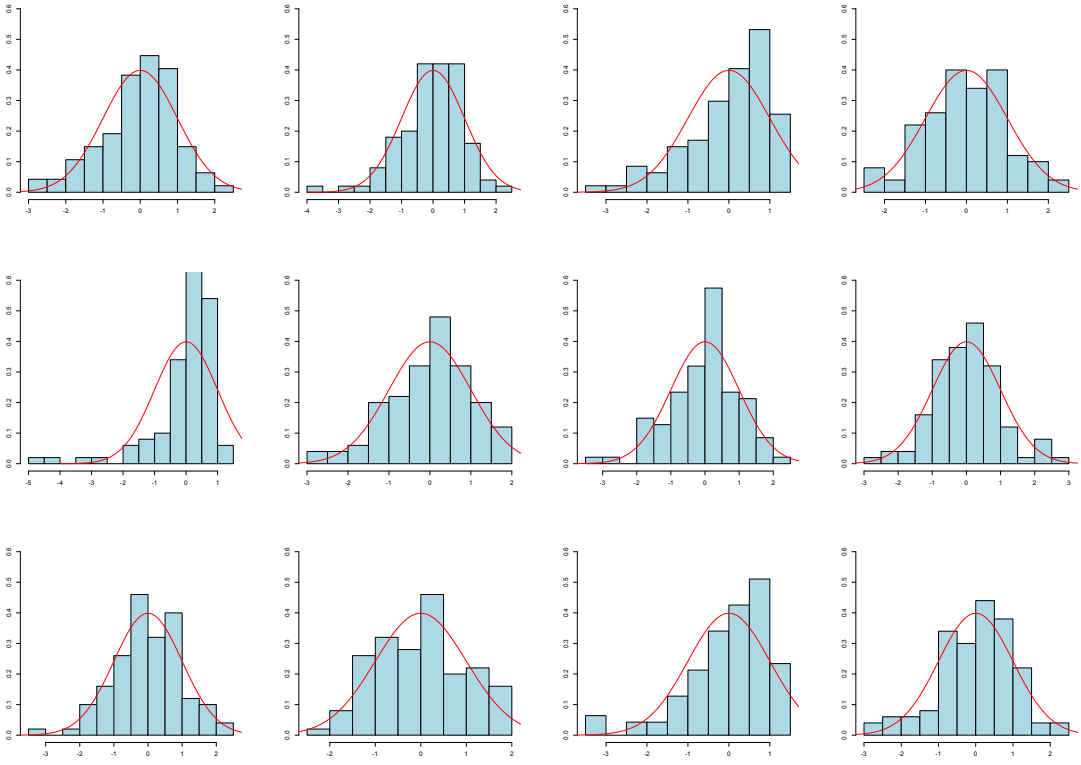


图14：模型（S2）下（α，β）$\sqrt{n}$-标准化估计量的分布直方图。第一行 $n = 5,000$，第二行 $n = 10,000$，第三行 $n = 20,000$。第一列 $\hat{}\alpha_n$，第二列 $\hat{}\tilde{}$列 β_n，第三列 $\tilde{\alpha}_n$，第四列 β_n。

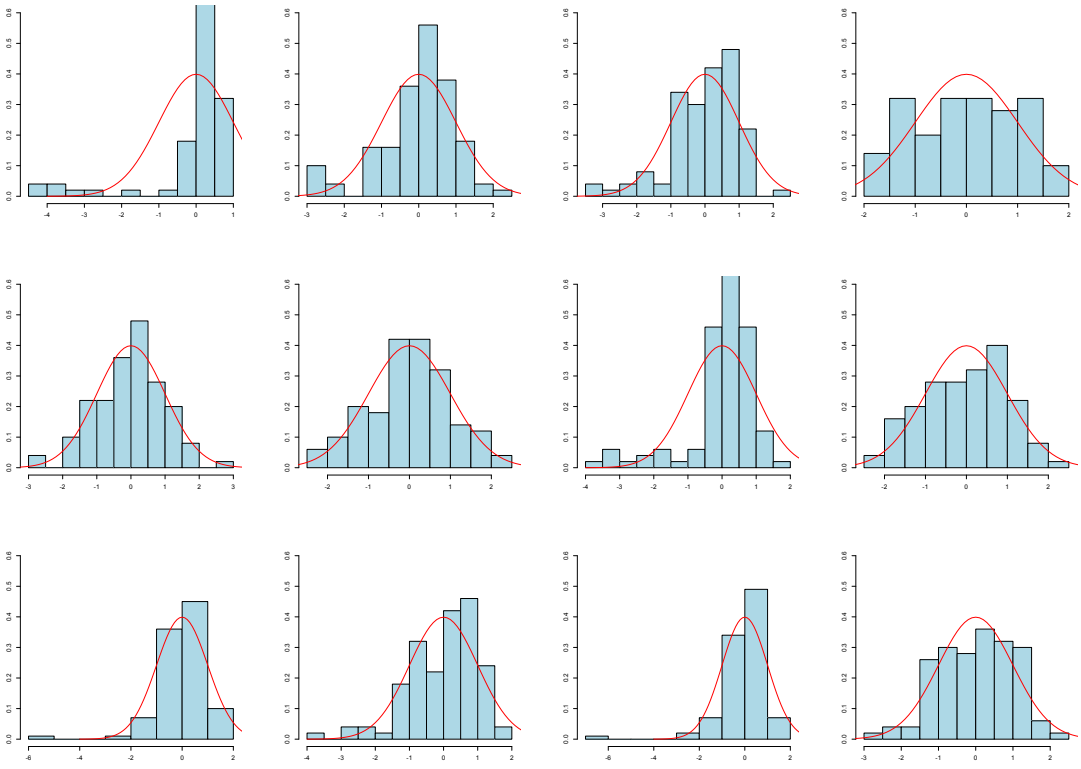


图15：模型（S3）下 $(\alpha,\ \beta)$ $\sqrt{n}$-标准化估计量的分布直方图。第一行 $n = 5,000$，第二行 $n = 10,000$，第三行 $n = 20,000$。第一列 $\hat{}\alpha_n$，第二列 $\hat{}\sim$列 β_n，第三列 $\tilde{\alpha}_n$，第四列 β_n。

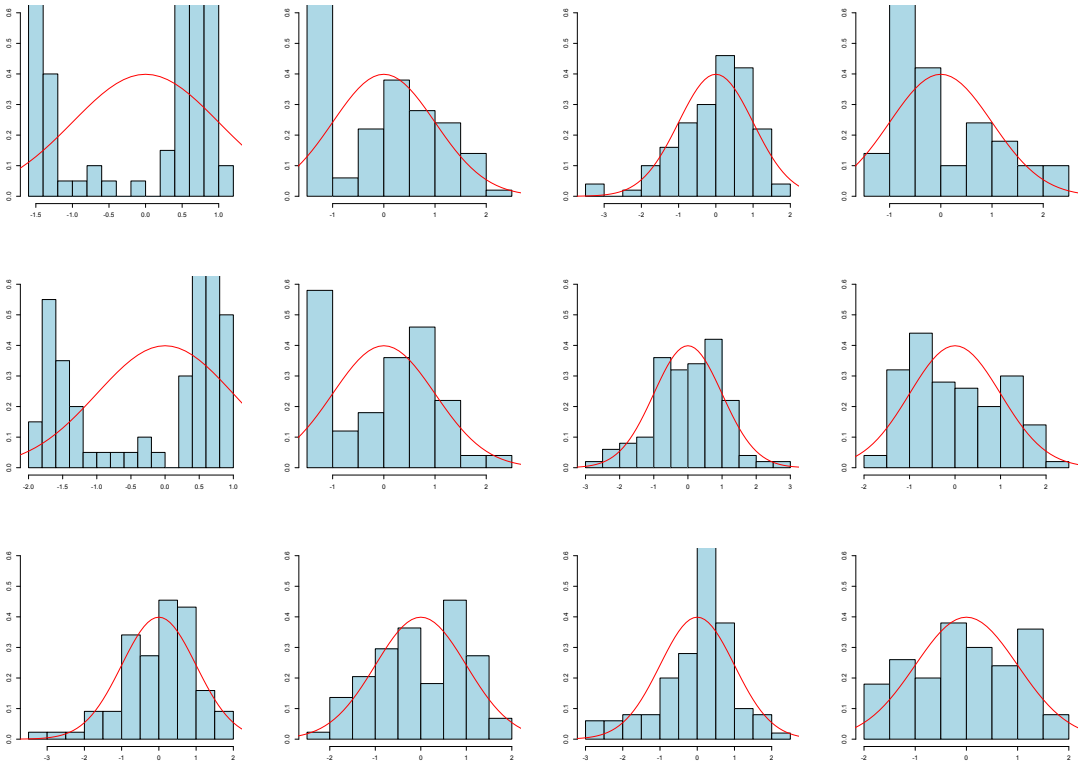


图16：模型（S4）下(α,β) $\sqrt{n}$-归一化估计量的分布直方图。第一行 $n=5{,}000$，第二行 $n=10{,}000$，第三行 $n=20{,}000$。第一列 $\hat{}\alpha_n$，第二列 $\hat{}\tilde{}$列 β_n，第三列 $\tilde{\alpha}_n$，第四列 β_n。