arXiv:2602.09833v1 [math.ST] 10 Feb 2026

基于批次化断裂随机样本的密度估计

S

Hancheng Bi、Bernhard Schmitzer、Thilo D. Stier

2026年2月11日

摘要

这一问题最早由 DeGroot、Feder 和 Gole（1971年，《数学统计年刊》）提出：在每次观测（批次）中，从密度为 $p(x,y)$ 的联合分布中抽取 M 个独立同分布的点对 $((X_i,Y_i))_{i=1}^M$，但我们只能分别观测无序多重集 $(X_i)_{i=1}^M$ 和 $(Y_i)_{i=1}^M$；也就是说，配对信息丢失了。研究表明，当 M 较大时，几乎不可能仅凭单次观测推断出 p。本文提出一种基于伪对数似然的参数方法，用于从 N 个独立同分布的破损样本批次中估计 p；在温和的假设下，我们证明了该估计量具有关于 N 的快速收敛速率，且该速率对 M 是一致的。

1 引言

1.1 问题陈述与相关工作

问题陈述。设 $\mathbb{X}$、$\mathbb{Y}$ 为可测空间，$\pi \in \mathcal{P}(\mathbb{X}\times\mathbb{Y})$ 为联合概率测度。设 $(X_i,Y_i)_{i=1}^M$ 为服从分布律 π 的独立同分布（i.i.d.）随机变量对。我们假设无法获取实际的配对信息，即只能观测随机多重集 $\{X_1,\cdots,X_M\}$ 和 $\{Y_1,\cdots,Y_M\}$，而无法获知各点的排列顺序。这称为来自 π 的断裂随机样本，最初由 [7] 引入。关于从破损样本中提取信息，且其中 π 服从二元正态分布的先前研究包括 [8] 和 [5]。尤其是，[3]证明，在某些温和条件下，相关参数无法估计。本文假设，我们并非只观测单个断裂样本，而是可以获取 N 个独立同分布的破损样本观测，即有服从分布律 π 的独立同分布随机变量对 $((X_i^k,Y_i^k)_{i=1}^M)_{k=1}^N$，并观测多重集的集合

$$\Big(\{X_1^k,\cdots,X_M^k\},\{Y_1^k,\cdots,Y_M^k\}\Big)_{k=1}^N.$$

我们感兴趣的是在参数设定下，利用这些独立同分布样本估计 π，并研究估计量相对于 N 和 M 的表现。我们将每一对观测到的多重集 $(\{X_1^k,\cdots,X_M^k\},\ \{Y_1^k,\cdots,Y_M^k\})$ 称为一个批次，并将估计 π 的问题称为批量破损样本问题。

如果 $M=1$，我们就得到了利用 N 个独立同分布样本估计 π 的经典问题。然而，由于缺少配对信息，当 $M\to\infty$ 时，直观来看，各批次将近似为来自 π 的边缘分布的独立样本，这使人怀疑每个批次所携带的关于 π 的信息会随着 M 增大而降至零。下面我们将证明事实并非如此。

应用。批量破损样本问题可作为粒子共定位分析的理想化模型。图1 展示了免疫标记的人类细胞的超分辨率荧光显微镜图像的一部分。线粒体外膜中的 HA 标签与线粒体内膜中的蛋白质 Mic60 分别用绿色和紫色标记物标记。大多数蛋白质似乎位于另一类蛋白质的近距离空间邻近处。这是因为这些蛋白质形成结合对以实现其生物学功能，还是蛋白质位置彼此独立，而共定位仅仅是偶然的，源于粒子的高密度，且这些粒子处于一个受限

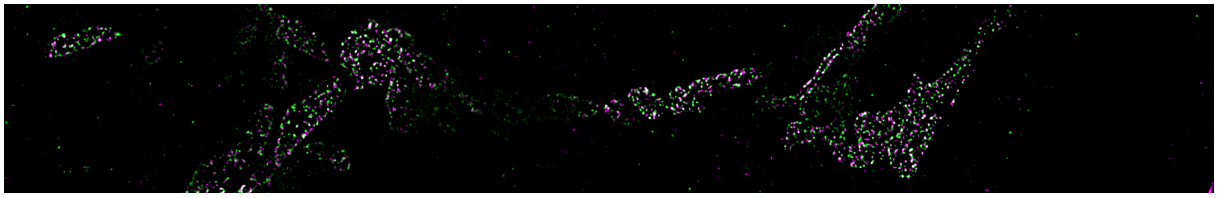

图 1：受激发射损耗（STED）显微镜图像，部分内容取自 [9, 图 37(A)]。细胞经 HA 标签（绿色）和 Mic60（紫色）染色。

区域? 若如上所述，将两类蛋白质的位置建模为随机变量 X_i 和 Y_i，则可根据 π 的分布律区分这两种情形：对于结合对， X_i 和 Y_i 将以高概率彼此接近；在偶然共定位的情况下， π 是一个独立乘积测度。当然，要从图像中现实地估计 π ，还必须考虑若干额外效应：并非所有粒子都会在成像过程中被实际捕获；即使粒子确实倾向于形成配对，也并非所有粒子在所有时刻都必然配对；并且不同图像中的粒子将遵循略有不同的分布律 π。然而，我们的简化模型确实表明，粒子密度的增加对于推断 π而言未必构成问题。

另一项应用是离散时间动力系统的分析。在这种情况下，我们将 X_i 和Y_i 解释为粒子在相邻两个时间步的位置；一项常见任务是根据 X_i推断 Y_i的条件分布。 $M = 1$ 对应于观测单个粒子的演化，而$M > 1$ 对应于观测多个无法区分的粒子的情形，例如粒子图像测速中的流体漂浮粒子：将小粒子播撒到流体中，并用脉冲激光器照射。每次激光脉冲时，都会记录这些粒子散射的光所形成的分布，从而生成一系列离散图像（例如参见 [1, 图 6]）。任务是求出流速。

当点密度相对较低或点移动较慢时，上述两个问题都可以采用基于最优传输的方法解决；分别参见 [11] 和 [1, 10], 。然而，最优传输往往会使配对偏向于偶然距离较近的点，而点密度越高，这种情况就越容易发生。本文提出一种参数方法，用于估计分布律 π的密度，并定量证明该方法在高密度情形下（即当 $M \to \infty$时）不受影响。 [4], 讨论了一种采用类似损失函数的非参数方法，并在适当意义下给出了当 $N \to \infty$ 时收敛的定性证明。

1.2 概述

本节余下部分将介绍假设和记号，详细陈述问题，并概述主要结果。第 2 节汇集或证明所需的辅助定理和引理。我们的主要结果将在第 3 节证明。第 4 节给出一些数值示例。

1.3 问题描述与主要结果

设 $\mu \in \mathcal{P}(\mathbb{X})$ 和 $\nu \in \mathcal{P}(\mathbb{Y})$ 是 π 的边缘分布。在 π 可由边缘分布确定的情况下，问题便简化为传统的密度估计。因此，在本文中，我们始终假设 μ， ν 已知，并关注相对于 $\mu \otimes \nu$ 估计 π 的密度，这与 [3, p. 529] 类似。在实践中，可以先利用 $M \cdot N$ 独立同分布随机变量 $(X_i^k)_{i,\ k}$ 和 $(Y_i^k)_{i,\ k}$，以极高精度估计 μ 和 ν。

假设 1.1. $\pi = p^* \cdot (\mu \otimes \nu)$ 对于某个 p^*： $\mathbb{X} \times \mathbb{Y} \to \mathbb{R}$。

我们还需 要定义一个候选集合，并从中选出估计量。这里我们使用一个参数化类 s.

定义 1.2. 设 Θ 是有限维欧几里得空间中的一个开子集，其紧闭包为$\bar{\Theta}$，令 $D := $上确界$_{\theta,\theta' \in \Theta} \|\theta - \theta'\|$ 为其直径。设 $(p^\theta)_{\theta \in \bar{\Theta}}$ 是一个满足以下性质的参数化密度类：

1. 函数集合 $\{\bar{\Theta} \ni \theta \mapsto p^\theta(x,y) \mid (x,y) \in \mathbb{X}\times\mathbb{Y}\}$ 限制在 Θ上时二阶连续可微，一致有界于某个常数 $0<U<\infty$，并且是一致利普希茨的，其 *Lipschitz* 常数为 $\mathcal{L}<\infty$；所有关于 $\theta\in\Theta$的二阶方向导数均以 $\mathcal{L}'<\infty$为界。

2. 对所有 $\theta\in\Theta$, $x\in\mathbb{X}$, $y\in\mathbb{Y}$，我们有 $\int_{\mathbb{X}} p^\theta(x',y)\mathrm{d}\mu(x') = \int_{\mathbb{Y}} p^\theta(x,y')\mathrm{d}\nu(y') = 1$。

注 1.3 （样本拆分）。为了“忘记”批次内的配对信息，*[4]* 的作者引入了一个辅助随机变量，即一个均匀采样的置换 $\sigma\sim\mathcal{U}(Sym(M))$，并假设对于每个批次，我们只能观察到 $((X_i,Y_{\sigma(i)}))_{i=1}^M$。正如 *[4],* 所讨论的，在该模型中，负对数似然为

$$\ell(\theta) = -\frac{1}{N}\sum_{k=1}^{N}\log\left(\frac{1}{M!}\sum_{\sigma\in S_M}\prod_{i=1}^{M} p^\theta\left(X_i^k, Y_{\sigma(i)}^k\right)\right). \tag{1}$$

虽然这是表述拆分样本问题的一种严谨方式，但它引入了一个额外的随机变量，并使概率空间变得复杂。

为简化起见，本文采用一种不同的方法：我们不显式地对拆分样本进行建模（例如，对这些点施加一个未知的均匀随机置换），而是对具有已知配对关系的数据进行建模，同时确保估计量不使用任何配对信息。具体而言，这意味着将估计量定义为一个损失函数的极小化子，而该损失函数在批次内对点进行置换时保持不变。

显然，方程（1）具有置换不变性，因此原则上可以用作损失函数。然而，除非 M 非常小，否则在实践中并不可行，因为计算 $\ell(\theta)$ 的复杂度通常是 NP难 [12]。受 [4], 的启发，我们通过负伪对数似然来定义损失函数

$$f_M^N(\theta) := -\frac{1}{N}\sum_{k=1}^{N}\sum_{i=1}^{M}\sum_{j=1}^{M}\log\left(\frac{1}{M}p^\theta(X_i^k,Y_j^k)+\frac{M-1}{M}\right) \quad \text{and its minimiser} \quad \theta_M^N\in\operatorname*{argmin}_{\theta\in\bar{\Theta}} f_M^N(\theta). \tag{2}$$

此外，我们还引入

$$f_M(\theta) := \mathbb{E}(f_M^N(\theta)), \qquad f_\infty(\theta) := \lim_{M\to\infty} f_M(\theta), \qquad \theta_M\in\operatorname*{argmin}_{\theta\in\bar{\Theta}} f_M(\theta), \qquad \theta_\infty\in\operatorname*{argmin}_{\theta\in\bar{\Theta}} f_\infty(\theta).$$

乍一看，人们可能会猜想 f_M^N随 $O(M^2$按 $M\to\infty$的尺度增长，因为对i 和 j求和两次。定理 1.6解决了这一疑虑，该定理表明，期望 $f_M(\theta)$ 在 $M\to\infty$时存在有限极限。

注释 *1.4*。直观地说，负伪对数似然（*2*）的由来如下。在某个批次 k 中，均匀随机选取一对点索引 *(i, j)* 时，有 $\frac{1}{M}$ 的概率使得 $X := X_i^k$ 和 $Y := Y_j^k$ 构成一个服从分布 π 的样本对；有 $\frac{M-1}{M}$ 的概率使它们相互独立，即 $(X,Y) \sim \left(\frac{1}{M}p^* + \frac{M-1}{M}\right)(\mu\otimes\nu)$。不同的样本对 (X_i^k, Y_j^k)（对于相同的 k，但 i, j 不同）通常彼此并不独立。然而，如果接受这一近似，（*2*）就可以表示为所有样本对的（负）对数似然之和。

注释 1.5 (与 [4]的比较)。在 *[4]* 以下伪似然

$$\theta\mapsto -\frac{1}{N}\sum_{k=1}^{N}\sum_{j=1}^{M}\log\left(\frac{1}{M}\sum_{i=1}^{M}p^\theta(X_i^k,Y_j^k)\right) \tag{3}$$

这里考虑的是另一种形式，而非 *(2)*。它同样具有置换不变性。对数中的项 $\frac{1}{M}\sum_{i=1}^{M}p^\theta(X_i^k,Y_j^k)$ 对应于在假设无法获得配对信息的情况下，给定元组 $(X_i^k)_{i=1}^M$ 时 Y_j^k 的条件概率。通过作出变量 $(Y_j^k)_{j=1}^M$ 在给定 $(X_i^k)_{i=1}^M$ 的条件下相互独立这一近似，并类比注释 *1.4* 进行论证，便可得到 *(3)*。运用 *[4]* 中的技术可以论证，*(3)* 比 *(2)* 是更准确的近似；不过，本文考虑后者，因为它更便于进行定量收敛分析。

现在我们可以陈述主要结果。第一个结果给出了当 $M \to \infty$ 时期望伪似然的渐近形式。

定理 1.6。

$$f_M(\theta) = f_\infty(\theta) + O(M^{-1})$$

其中

$$f_\infty(\theta) = \frac{1}{2}\Big(\|p^\theta - p^*\|^2_{L^2(\mu\otimes\nu)} - \|p^*\|^2_{L^2(\mu\otimes\nu)} + 1\Big).$$

注 1.7 (极小化点的行为 f_M)。 当 $M = 1$， f_M 就变成了标准的总体负对数似然。众所周知，最小化负对数似然等价于最小化真实分布与我们估计的分布之间的 KL 散度。对于任意 M，类似于[4, 命题 A。4], 我们有

$$f_M(\theta) = M^2 \cdot \mathrm{KL}\left((\tfrac{1}{M}p^* + \tfrac{M-1}{M}) \cdot \mu\otimes\nu \;\middle|\; (\tfrac{1}{M}p^\theta + \tfrac{M-1}{M}) \cdot \mu\otimes\nu \right) + \text{const.}$$

这由定理 *1.6* 的证明中的公式 *(5)* 得出。因此，若 $p^* \in (p^\theta)_{\theta\in\Theta}$，则可得出， f_M 的任意极小化子都几乎处处等于 p^* $\mu\otimes\nu$。若 $p^* \notin (p^\theta)_{\theta\in\Theta}$，那么对于 $M = 1$， f_M 的极小化点将是 p^* 在参数类上的 *KL* 投影。根据定理 *1.6*，当 $M \to \infty$ 时，f_M 的极小化点应收敛到 $L^2(\mu\otimes\nu)$-投影。定理 *1.9* 将对此作更具体的说明。

下一个结果说明了 f_M^N围绕 f_M的集中性，且对 θ一致成立。

定理 1.8。 设 D， U， $\mathcal{L}$ 如定义 *1.2*所述，令

$$J := \int_0^D \log\left(1 + N(\varepsilon; \Theta, \|\cdot\|)\right) \mathrm{d}\varepsilon < \infty,$$

其中 $N(\varepsilon;\ \Theta, \|\cdot\|)$ 是半径为 ε的$(\Theta, \|\cdot\|)$ 的覆盖数。注意，根据我们对 Θ的假设， J是有限的。因此，存在一个绝对常数 $c > 0$， 使得

$$\mathbb{P}\left(\sup_{\theta\in\Theta} \left| f_M^N(\theta) - f_M(\theta) \right| \geq 6t \right) \leq 6\exp\left(-c\min\left\{ \frac{t^2N}{U^4}, \frac{tN}{U} \right\} \right) + 6\exp\left(\frac{J}{D}\right)\exp\left(-\frac{ct\sqrt{N}}{D\mathcal{L}U} \right)$$

对所有 $t > 0$均成
立。

这表明，正如预期，较大的 N 有助于提升估计质量，而较大的 M不会造成不利影响。相反，结合定理 1.6，我们看到，当 M 和 N 较大时，可以预期 p^* 的估计量接近 p^* 在 $(p^\theta)_{\theta\in\bar{\Theta}}$ 上的 $L^2(\mu\otimes\nu)$- 投影。因此，若假设该投影具有唯一性，且 $\|p^\theta - p^*\|^2_{L^2(\mu\otimes\nu)}$ 在其附近具有严格正曲率，则对于较大的 M，可以得到 $f_M^N(\theta)$的极小化子收敛到 $f_M(\theta)$的极小化子的收敛速率下界：

定理 1.9。 假设 $\theta_\infty \in \Theta$ 是 f_∞的唯一极小化子，并且存在某些 τ、 $r > 0$， 使得对所有满足 $\|\theta - \theta_\infty\|_2 \leq r$的 $\theta \in \Theta$， $\nabla^2 f_\infty(\theta)$ 的最小特征值都大于 τ。那么存在某些 $M_0 \in \mathbb{N}$、 $t_0 > 0$， 使得对所有 $M \geq M_0$ 和 $t \leq t_0$， 我们都有

$$\mathbb{P}\left(\left\|\theta_M^N - \theta_M\right\|_2 \geq t \right) \leq C_1\exp\left(-C_2\min\left\{ t^2N, t\sqrt{N} \right\} \right),$$

其中经验极小化子 θ_M是唯一的，且 C_1、 $C_2 > 0$是独立于 M和 N的常数。

2 预备知识

定理 2.1 （参数积分的可微性 [2, 定理 3.18]）。 假设 Θ 是 $\mathbb{R}^n$中的开集，并且假设 $f\colon\ \mathbb{X}\times\Theta\to\mathbb{R}$ 满足

1. $f(\cdot,\theta)\in L^1_\mu(\mathbb{X})$ 对每个 $\theta\in\Theta$ 均成立；*2.* $f(x,\cdot)\in C^1(\Theta)$ 对 μ-几乎处处的 $x\in\mathbb{X}$ 成立；

3. 存在某个 $g\in L^1_\mu(\mathbb{X})$ 使得

$$\left|\frac{\partial}{\partial\theta_m}f(x,\theta)\right|\le g(x)\qquad \textit{for all } (x,\theta)\in\mathbb{X}\times\Theta\quad \textit{and}\quad m\in\{1,\dots,n\}.$$

那么， $\theta\mapsto\int_{\mathbb{X}}f(x,\theta)\mathrm{d}\mu$ 连续可微，且

$$\frac{\partial}{\partial\theta_m}\int_{\mathbb{X}}f(x,\theta)\mathrm{d}\mu(x)=\int_{\mathbb{X}}\frac{\partial}{\partial\theta_m}f(x,\theta)\mathrm{d}\mu(x)\qquad \textit{for all } \theta\in\Theta\quad \textit{and}\quad m\in\{1,\dots,n\}.$$

定义 2.2 (ψ_1-范数)。 *[13, 式 (2.21)]* 实值随机变量F 的次指数范数定义为

$$\|F\|_{\psi_1}=\inf\left\{t>0:\mathbb{E}\exp\left(\frac{|F|}{t}\right)\le 2\right\}.$$

引理 2.3。 给定一个随机变量 F，其满足 $\mathbb{E}(|F|^p)\le(Cp)^p$ 对所有 $p\in\mathbb{N}$ 均成立，则有 $\|F\|_{\psi_1}\le 2eC$。证明。 将 $t=2eC$ 代入以下方程

$$\mathbb{E}\exp\left(\frac{|F|}{t}\right)=\mathbb{E}\left(\sum_{p=0}^{\infty}\frac{|F|^p}{t^p p!}\right)\le\sum_{p=0}^{\infty}\frac{C^p p^p}{t^p p!}\le\sum_{p=0}^{\infty}\frac{C^p e^p}{t^p}=\frac{1}{1-\frac{Ce}{t}}$$

其中我们使用了 $(p/e)^p\le p!$ 以及几何级数的极限。 □

定理 2.4 (Bernstein，推论 2.8.3，见 [13])。 设 $F_1,\dots,F_N$ 为均值为零的独立同分布次指数随机变量，则对任意 $t\ge 0$，有

$$\mathbb{P}\left(\left|\frac{1}{N}\sum_{i=1}^{N}F_i\right|\ge t\right)\le 2\exp\left(-c\min\left\{\frac{t^2N}{\|F_1\|^2_{\psi_1}},\frac{tN}{\|F_1\|_{\psi_1}}\right\}\right)$$

其中 c是一个绝对常数。

引理 2.5。 设 $F_1,\dots,F_N$ 为均值为零的独立同分布次指数随机变量，则存在某个绝对常数 $c>0$ 使得

$$\left\|\frac{1}{N}\sum_{i=1}^{N}F_i\right\|_{\psi_1}\le\frac{c\|F_1\|_{\psi_1}}{\sqrt{N}}.$$

证明。 记 $\beta:=\|F_1\|_{\psi_1}$。对于任意 $p\ge 1$，由定理 2.4

$$\begin{aligned}
\mathbb{E}\left(\left|\frac{1}{N}\sum_{i=1}^{N}F_i\right|^p\right)&=\int_0^\infty\mathbb{P}\left(\left|\frac{1}{N}\sum_{i=1}^{N}F_i\right|^p\ge t\right)\mathrm{d}t\\
&=p\int_0^\infty\mathbb{P}\left(\left|\frac{1}{N}\sum_{i=1}^{N}F_i\right|\ge t\right)t^{p-1}\mathrm{d}t\\
&\le 2p\int_0^\infty t^{p-1}\exp\Big(-\frac{cNt^2}{\beta^2}\Big)\mathrm{d}t+2p\int_0^\infty t^{p-1}\exp\Big(-\frac{cNt}{\beta}\Big)\mathrm{d}t\\
&\le 2p\Big(\frac{cN}{\beta^2}\Big)^{-p/2}\Gamma\Big(\frac{p}{2}\Big)+2p\Big(\frac{cN}{\beta}\Big)^{-p}\Gamma(p)\\
&\le 6p\Gamma(p)\Big(\frac{\beta}{c\sqrt{N}}\Big)^p\le\Big(\frac{6\beta p}{c\sqrt{N}}\Big)^p.
\end{aligned}$$

在最后一行中，我们用到了定理 c 2.4 中的条件在不失一般性的情况下满足 $c \leq 1$，并且对于 $p \geq 1$，我们有$\Gamma(p/2) \leq 2\Gamma(p)$，以及 $p\Gamma(p) = p! \leq p^p$。应用引理 2.3 即可完成证明。 □

定义2.6 (ψ_1-过程，定义5.35，见 [14])。零均值随机过程 $\{F_\theta : \theta \in \Theta\}$ 若满足以下条件，则称其相对于范数 $\|\cdot\|_\Theta$ 是一个ψ_1 过程：

$$\|F_\theta - F_{\theta'}\|_{\psi_1} \leq \|\theta - \theta'\|_\Theta .$$

引理 2.7。考虑参数函数类 $g_\theta : \mathbb{X} \to \mathbb{R}$、概率测度 $\rho \in \mathcal{P}(\mathbb{X})$以及一些独立同分布的 $X_k \sim \rho$。若存在一个 $L > 0$，使得对于任意 $\theta, \theta' \in \Theta$，都有 $\left\|g^\theta - g^{\theta'}\right\|_\infty \leq L \|\theta - \theta'\|_\Theta$，那么

$$F_N(\theta) := \frac{1}{N}\sum_{k=1}^{N} g^\theta(X_k) - \mathbb{E}g^\theta(X)$$

对于某个绝对常数 $c > 0$，它相对于范数 $\frac{cL}{\sqrt{N}}\|\cdot\|_\Theta$ 是一个 ψ_1-过程。

证明。注意到 $\mathbb{E}(|F_1(\theta) - F_1(\theta')|^p) \leq (2L\|\theta - \theta'\|)^p$，应用引理 2.3 和引理 2.5。 □

Defi定义 2.8 （广义Dudley熵积分）。取 D 自定义 *1.2*。我们定义

$$\mathcal{J}(\delta, D) = \int_\delta^D \log\Big(1 + N(\varepsilon; \Theta, \|\cdot\|_\Theta)\Big)\mathrm{d}\varepsilon$$

其中 $N(\varepsilon; \Theta, \|\cdot\|_\Theta)$ 是$(\Theta, \|\cdot\|_\Theta)$的 ε-覆盖数。

定理2.9 （定理5.36，见 [14]）。设 $\{F_\theta : \theta \in \Theta\}$ 是关于 $\|\cdot\|_\Theta$ 的 ψ_1-过程，则存在普适常数 c_1，使得对所有 $t \in \mathbb{R}$，我们有

$$\mathbb{P}\left(\sup_{\theta,\theta' \in \Theta} |F_\theta - F_{\theta'}| \geq c_1(\mathcal{J}(0, D) + t)\right) \leq 2\exp\Big(-\frac{t}{D}\Big).$$

引理 2.10。取 D 自定义 *1.2*，并令 $\|\cdot\|$ 表示标准欧几里得范数。定义

$$J := \int_0^D \log\left(1 + N(\varepsilon; \Theta, \|\cdot\|)\right)\mathrm{d}\varepsilon.$$

设 $f(\theta)$ 对某个 $C > 0$ 是关于 $\frac{C}{\sqrt{N}}\|\cdot\|$ 的 ψ_1-过程。存在绝对常数 $c_1 > 0$，使得对所有 $t > 0$

$$\mathbb{P}\left(\sup_{\theta,\theta' \in \Theta} |f(\theta) - f(\theta')| \geq t\right) \leq 2\exp\left(\frac{J}{D}\right)\exp\left(-\frac{t\sqrt{N}}{c_1 DC}\right).$$

证明。首先注意到，由于 Θ 是有限维欧几里得空间的紧子空间，我们有$J < \infty$，由 [6, 命题 5]。令 $D' := \sup_{\theta,\, \theta' \in \Theta} \frac{C}{\sqrt{N}}\|\theta - \theta'\| = \frac{DC}{\sqrt{N}}$。由定义 2.8，对积分作换元，我们得到

$$\mathcal{J}(0, D') = \int_0^{D'} \log\left(1 + N\Big(\varepsilon; \Theta, \frac{C}{\sqrt{N}}\|\cdot\|\Big)\right)\mathrm{d}\varepsilon = \frac{CJ}{\sqrt{N}}. \tag{4}$$

根据定理2.9 存在某个绝对常数 $c_1 > 0$

$$\mathbb{P}\left(\sup_{\theta,\theta' \in \Theta} |f(\theta) - f(\theta')| \geq c_1\Big(\frac{CJ}{\sqrt{N}} + \tau\Big)\right) \leq 2\exp\Big(-\frac{\tau}{D'}\Big) = 2\exp\Big(-\frac{\tau\sqrt{N}}{DC}\Big) \qquad \forall \tau \in \mathbb{R}.$$

代入 $t = c_1\Big(\frac{CJ}{\sqrt{N}} + \tau\Big)$，证明完毕。 □

引理 2.11。 设 $\{F_\theta:\ \theta\in\Theta\}$ 为一个随机过程，并固定某个 $\theta_0\in\Theta$。则

$$\mathbb{P}\left(\sup_{\theta\in\Theta}|F_\theta|\geq 2t\right)\leq\mathbb{P}\left(\sup_{\theta,\theta'\in\Theta}|F_\theta-F_{\theta'}|\geq t\right)+\mathbb{P}\left(|F_{\theta_0}|\geq t\right).$$

证明。 应用三角不等式和并集界。 □

引理 2.12 （霍夫丁， [13]中的定理 2.2.6）。 设 $F_1,...,F_N$ 为独立同分布，假设 $|F_k|\leq C$，那么对任意 $t\geq 0$，有

$$\mathbb{P}\left(\left|\frac{1}{N}\sum_{k=1}^{N}F_k-\mathbb{E}(F_1)\right|\geq t\right)\leq 2\exp\left(-\frac{Nt^2}{2C^2}\right).$$

3 非渐近收敛结果

3.1 $f_M^N(\theta)$ 的集中

从现在起，我们假设 $M\geq 2$。本节中，我们将随机变量 $p^\theta(X_i^k,Y_j^k)$ 简记为$p_{ij}^k(\theta)$。类似地，对于点 $(x_i)_i\subset\mathbb{X}$、$(y_i)_i\subset\mathbb{Y}$，我们将 $p^*(x_i,y_i)$ 简记为 p_{ii}^*，将 $p^\theta(x_i,y_j)$ 简记为 $p_{ij}(\theta)$。我们还定义 $q_{ij}^k(\theta):=p_{ij}^k(\theta)-1$ 以及 $q_{ij}(\theta)=p_{ij}(\theta)-1$。注意， p_{ii}^* 和 p_{ij} 在两个边缘分布上的积分均为 1，即有

$$\int_{\mathbb{X}}p_{ii}^*\mathrm{d}\mu(x_i)=\int_{\mathbb{X}}p_{ii}^*\mathrm{d}\nu(y_i)=1$$

以及关于 $p_{ij}(\theta)$ 的相应陈述。类似地， $q_{ij}(\theta)$ 的积分为 0。我们用 $\mathrm{d}\pi_{ii}$ 表示对 $\mathrm{d}\pi(x_i,\ y_i)$ 积分，用 $\mathrm{d}\pi^{\otimes M}$ 表示对所有因子 $\Pi_{i=1}^M\mathrm{d}\pi_{ii}$ 的联合积分。对于乘积测度 $\mu\otimes\nu$ 下的积分，我们也采用类似的记号。

定理 （定理 1.6 的回顾）。 现在我们证明一个关于期望伪似然渐近形式的陈述，其中 $f_M(\theta)=\frac{1}{2}\left(\left\|p^\theta-p^*\right\|_{L^2(\mu\otimes\nu)}^2-\|p^*\|_{L^2(\mu\otimes\nu)}^2+1\right)+O(M^{-1}$ $M\to\infty$。

证明。 回忆 $q_{ij}(\theta):=p_{ij}(\theta)-1$。我们得到

$$\begin{aligned}
-f_M(\theta)&=\int\sum_{i,j=1}^{M}\log\left(\frac{1}{M}p_{ij}(\theta)+\frac{M-1}{M}\right)\mathrm{d}\pi^{\otimes M}\\
&=\int\sum_{i,j=1}^{M}\log\left(1+\frac{1}{M}q_{ij}(\theta)\right)\prod_{l=1}^{M}p_{ll}^*\ \mathrm{d}(\mu\otimes\nu)^{\otimes M}\\
&=\sum_{i=1}^{M}\int\log\left(1+\frac{1}{M}q_{ii}(\theta)\right)\prod_{l=1}^{M}p_{ll}^*\ \mathrm{d}(\mu\otimes\nu)^{\otimes M}\\
&\quad+\sum_{i\neq j}\int\log\left(1+\frac{1}{M}q_{ij}(\theta)\right)\prod_{l=1}^{M}p_{ll}^*\ \mathrm{d}(\mu\otimes\nu)^{\otimes M}.
\end{aligned}$$

对于第一部分，考虑求和项 $i=1$。它等于 $\int\log(1+q_{11}/M)p_{11}^*\prod_{l\neq 1}p_{ll}^*\mathrm{d}(\mu\otimes\nu)^{\otimes M}$，并且可以利用 $\int p_{ll}^*d\mu\otimes\nu=1$对所有 $l\neq 1$ 在$(x_l,\ y_l)$上积分。对 i的所有其他取值可以同样论证，从而得到 M倍的相同值。对于第二部分，类似地考虑求和项 $i=1$和 $j=2$。它等于 $\int\log(1+q_{12}/M)p_{11}^*p_{22}^*\prod_{l=3}^{M}p_{ll}^*\mathrm{d}(\mu\otimes\nu)^{\otimes M}$，我们可以再次对$l\geq 3$在$(x_l,\ y_l)$上积分，然后通过 $\int_{\mathbb{X}}p_{22}^*\mathrm{d}\mu(x_2)=\int_{\mathbb{Y}}p_{11}^*\mathrm{d}\nu(y_1)=1$对 y_1 和 x_2 积分。对 $i,\ j$的所有其他取值同样论证，可得该贡献的 $M(M-1)$倍。因此，

$$\begin{aligned}
-f_M(\theta)&=M\int\log\left(1+\frac{1}{M}q_{11}(\theta)\right)p_{11}^*\mathrm{d}\mu\otimes\nu+M(M-1)\int\log\left(1+\frac{1}{M}q_{11}(\theta)\right)\mathrm{d}\mu\otimes\nu\\
&=M^2\int\log\left(1+\frac{1}{M}q_{11}(\theta)\right)\left(\frac{1}{M}p_{11}^*+\frac{M-1}{M}\right)\mathrm{d}\mu\otimes\nu.
\end{aligned}\tag{5}$$

接下来，我们推导当 $M \to \infty$时的渐近行为。为此，我们利用泰勒展开$\log(1+x) = x - x^2/2 + O(x^3)$（基于 $q_{11}(\theta) \in [-1, U-1]$这一事实，从而使近似误差一致有界）以及 $M \geq 2$，得到

$$\begin{aligned}
-f_M(\theta) &= M^2 \int \Big(\frac{q_{11}(\theta)}{M} - \frac{(q_{11}(\theta))^2}{2M^2} + O(M^{-3})\Big)\Big(\frac{1}{M}p_{11}^* + \frac{M-1}{M}\Big)\mathrm{d}\mu \otimes \nu \\
&= M^2 \int \frac{q_{11}(\theta)}{M}\frac{1}{M}p_{11}^* + \frac{q_{11}(\theta)}{M}\frac{(M-1)}{M} - \frac{(q_{11}(\theta))^2}{2M^2}\Big(\frac{1}{M}p_{11}^* + \frac{(M-1)}{M}\Big)\mathrm{d}\mu \otimes \nu + O(M^{-1}).
\end{aligned}$$

接下来，我们通过 $\int q_{11}\mathrm{d}\mu \otimes \nu = 0$ 消去第二项，将更多项吸收到 $O(M^{-1})$ 中，并重新整理其余部分

$$\begin{aligned}
-f_M(\theta) &= \frac{1}{2}\int 2q_{11}(\theta)p_{11}^* - (q_{11}(\theta))^2\left(\frac{1}{M}p_{11}^* + \frac{(M-1)}{M}\right)\mathrm{d}\mu \otimes \nu + O(M^{-1}) \\
&= \frac{1}{2}\int 2q_{11}(\theta)p_{11}^* - (q_{11}(\theta))^2\mathrm{d}\mu \otimes \nu + O(M^{-1}) \\
&= -\frac{1}{2}\int \Big(q_{11}(\theta) - p_{11}^*\Big)^2\mathrm{d}\mu \otimes \nu + \frac{1}{2}\int (p_{11}^*)^2\mathrm{d}\mu \otimes \nu + O(M^{-1}).
\end{aligned}$$

现在我们重新代入 $q_{11} = p_{11} + 1$，以完成证明

$$\begin{aligned}
-f_M(\theta) &= -\frac{1}{2}\int \Big(p_{11}(\theta) - p_{11}^*\Big)^2 - 2\Big(p_{11}(\theta) - p_{11}^*\Big) + 1\ \mathrm{d}\mu \otimes \nu + \frac{1}{2}\int (p_{11}^*)^2\mathrm{d}\mu \otimes \nu + O(M^{-1}) \\
&= -\frac{1}{2}\int \Big(p_{11}(\theta) - p_{11}^*\Big)^2\mathrm{d}\mu \otimes \nu + \frac{1}{2}\int (p_{11}^*)^2\mathrm{d}\mu \otimes \nu - \frac{1}{2} + O(M^{-1}).
\end{aligned}$$

□

定理 (定理 1.8的回顾)。 设 J， D 如引理 *2.10*所述。存在一个绝对常数 $c > 0$ 使得

$$\mathbb{P}\left(\sup_{\theta\in\Theta}\left|f_M^N(\theta) - f_M(\theta)\right| \geq 6t\right) \leq 6\exp\left(-c\min\left\{\frac{t^2N}{U^4}, \frac{tN}{U}\right\}\right) + 6\exp\left(\frac{J}{D}\right)\exp\left(-\frac{ct\sqrt{N}}{D\mathcal{L}U}\right).$$

完整证明将在本节末尾给出。为控制 $\sup_{\theta\in\Theta}\left|f_M^N(\theta) - f_M(\theta)\right|$，我们将 $f_M^N(\theta)$ 分成三项，即

$$\begin{aligned}
f_M^N(\theta) &= \frac{1}{NM}\sum_{k=1}^{N}\sum_{i,j=1}^{M} q_{ij}^k(\theta) + f_M^N(\theta) - \frac{1}{NM}\sum_{k=1}^{N}\sum_{i,j=1}^{M} q_{ij}^k(\theta) \\
&= \underbrace{-\frac{1}{NM}\sum_{k=1}^{N}\sum_{i,j=1,i\neq j}^{M} q_{ij}^k(\theta)}_{f_a(\theta)} \underbrace{-\frac{1}{NM}\sum_{k=1}^{N}\sum_{i=1}^{M} q_{ii}^k(\theta)}_{f_b(\theta)} + \underbrace{f_M^N(\theta) + \frac{1}{NM}\sum_{k=1}^{N}\sum_{i,j=1}^{M} q_{ij}^k(\theta)}_{f_c(\theta)}
\end{aligned} \tag{6}$$

并应用三角不等式，得到

$$\sup_{\theta\in\Theta}\left|f_M^N(\theta) - f_M(\theta)\right| \leq \sup_{\theta\in\Theta}\left|f_a(\theta) - \mathbb{E}f_a(\theta)\right| + \sup_{\theta\in\Theta}\left|f_b(\theta) - \mathbb{E}f_b(\theta)\right| + \sup_{\theta\in\Theta}\left|f_c(\theta) - \mathbb{E}f_c(\theta)\right|. \tag{7}$$

我们的目标是获得一个与 M 无关的 f_M^N 集中速率。为便于表述，我们将 f_M^N 分成三部分。请注意，f_a 的集中比其他部分需要更多工作，因为从直观上看，该部分的规模似乎按 $O(M)$ 的速率增长。下面的引理将用于界定上述各个加项。

引理 3.1 （组合学插曲）。 设 $l \in \mathbb{N}$，并将 $[\![1, l]\!] := \{1, \ldots, l\}$ 记为前 l 个自然数。定义

$$S := \{(z_i)_{i=1}^{2l} \in [\![1, M]\!]^{2l} \mid \forall i\, \exists j \neq i : z_j = z_i\}.$$

这可以解释为字母表 $[\![1,\ M]\!]$ 上长度为 $2l$ 且没有任何字符恰好出现一次的字符串集合。我们有

$$|S| \le (2Ml)^l.$$

证明。令 $B := \{b := (b_i)_{i=1}^{2l} \in \{0,1\}^{2l} \mid \|b\|_1 = l\}$ 表示由长度为 $2l$ 且恰有l 个元素等于 1 的二进制字符串组成的集合。我们定义一个映射

$$\sigma : B \times [\![1, M]\!]^l \times [\![1, l]\!]^l \to S, \quad (b, c, k) \mapsto s$$

按照以下步骤进行:

- 对于 $i \in [\![1, l]\!]$，令 j 为 b 中等于 1 的第 i 个元素。设 $s_j := c_i$。
- 对于 $i \in [\![1, l]\!]$，令 j 为 b 中等于 0 的第 i 个元素。设 $s_j := c_{k_i}$。

接下来我们证明 σ 是满射，并且任意 $s \in S$ 的原像至少有 2^l 个元素。给定任意一个 $s \in S$，令 τ 为一个排列，使得 $(s_{\tau(i)})_{i=1}^{2l}$ 以 s 中每个字母各出现两次的一对开头，后面接其余字母，顺序任意。这样的排列存在，因为没有字母恰好出现一次。在下文中，不失一般性地假设 $\tau = \mathrm{id}$。我们按照以下步骤构造 $s \in S$ 在 σ 下的原像:

- For $i \in [\![1, l]\!]$ set $(b_{2i-1}, b_{2i}) \in \{(1,0),(0,1)\}$ arbitrarily.
- For $i \in [\![1, l]\!]$ set $c_i := \begin{cases} s_{2i-1} & \text{if } b_{2i-1} = 1, \\ s_{2i} & \text{if } b_{2i} = 1. \end{cases}$
- For $i \in [\![1, l]\!]$ set $k_i := \begin{cases} c^{-1}(s_{2i-1}) & \text{if } b_{2i-1} = 0, \\ c^{-1}(s_{2i}) & \text{if } b_{2i} = 0 \end{cases}$

  其中 $c^{-1}(a)$ 表示任意一个指标 $j \in [\![1,\ l]\!]$ 使得 $c_j = a$ （根据假设和构造，该指标存在）。

将定义 σ 的过程应用于此，很容易看出，构造出的值确实是原像；又由于（b_{2i-1}，b_i）分别从各自的两种可能性中独立且任意地选取，因此我们有 2^l 个不同的原像。因此

$$|S| \le \frac{|B|\, M^l l^l}{2^l} \le \frac{2^{2l} M^l l^l}{2^l} = (2Ml)^l. \qquad \square$$

引理 3.2。 $f_a(\theta)$ 是关于 $\frac{c\mathcal{L}}{\sqrt{N}}\|\cdot\|$ 的一个 ψ_1-过程，~~其~~中 $c > 0$ 是某个绝对常数，且$\|\cdot\|$ 是 Θ 上的原范数。

证明。令 $F_k(\theta) := \frac{1}{M}\sum_{i\neq j}^{M} q_{ij}^k(\theta)$，则 $f_a(\theta) = \frac{1}{N}\sum_{k=1}^{N} F_k(\theta)$。注意， $F_k(\theta)$ 的均值为

$$\mathbb{E}F_k(\theta) = \frac{1}{M}\sum_{i\neq j}^{M}\int q_{ij}\prod_{m=1}^{M} p_{mm}^* \mathrm{d}(\mu\otimes\nu)^{\otimes M} = \frac{1}{M}\sum_{i\neq j}^{M}\int q_{ij}\mathrm{d}(\mu\otimes\nu)^{\otimes M} = 0.$$

对于偶数 $l \in \mathbb{N}$，我们得到

$$\mathbb{E}\left(|F_1(\theta) - F_1(\theta')|^l\right) = \mathbb{E}\left((F_1(\theta) - F_1(\theta'))^l\right) = \frac{1}{M^l}\sum_{i_1\neq j_1}^{M}\cdots\sum_{i_l\neq j_l}^{M}\mathbb{E}\left(\prod_{\alpha=1}^{l}\left(q_{i_\alpha j_\alpha}^1(\theta) - q_{i_\alpha j_\alpha}^1(\theta')\right)\right).$$

我们考察期望项。根据定义，有

$$\mathbb{E}\left(\prod_{\alpha=1}^{l}\left(q_{i_\alpha j_\alpha}^1(\theta) - q_{i_\alpha j_\alpha}^1(\theta')\right)\right) = \int\prod_{\alpha=1}^{l}\left(q_{i_\alpha j_\alpha}(\theta) - q_{i_\alpha j_\alpha}(\theta')\right)\prod_{m=1}^{M} p_{mm}^*\ \mathrm{d}(\mu\otimes\nu)^{\otimes M}.$$

给定如上所述的一组指标，如果存在某个 i_β 是唯一的，即 $i_\beta \notin \{i_\alpha \mid \alpha : \quad \alpha \neq \beta\} \cup \{j_\alpha \mid \alpha\}$，我们可以先对 y_{i_β} 进行积分（它只出现在 p^* 项中，积分结果为 1），然后再对 x_{i_β} 进行积分（它只出现在一个 q 项中，积分结果为 0）。因此，该因子意味着整个积分为 0。类似地，如果存在某个唯一的 j_β，则积分为 0。我们得出结论：如果在（$i_1, \cdots, i_l, j_1, \cdots, j_l$）中存在任何只出现一次的值，则积分为零。结合引理 S、$\mathcal{L}$ 中定义的 3.1 以及定义 1.2，并利用上面的方程，我们得到

$$\mathbb{E}\left(\left|F_1(\theta) - F_1(\theta')\right|^l\right) \leq \frac{1}{M^l} |S| \sup_{x,y} \left|q^\theta(x,y) - q^{\theta'}(x,y)\right|^l \leq \frac{1}{M^l} |S| \mathcal{L}^l \|\theta - \theta'\|^l \leq (2\mathcal{L} \|\theta - \theta'\| l)^l .$$

现在，对于 $l \in \mathbb{N}$ 为奇数的情况，应用詹森不等式

$$\mathbb{E}\left(\left|F_1(\theta) - F_1(\theta')\right|^l\right) \leq \left(\mathbb{E}\left(\left|F_1(\theta) - F_1(\theta')\right|^{l+1}\right)\right)^{\frac{l}{l+1}} \leq (2\mathcal{L} \|\theta - \theta'\| (l+1))^l \leq (4\mathcal{L} \|\theta - \theta'\| l)^l .$$

根据引理 2.3 可得 $\|F_1(\theta) - F_1(\theta')\|_{\psi_1} \leq 8e\mathcal{L} \|\theta - \theta'\|$。由于 F_k 是独立同分布的，且 $f_a(\theta) = \frac{1}{N} \sum_{k=1}^N F_k$，由引理 2.5 可得 $\|f_a(\theta) - f_a(\theta')\|_{\psi_1} \leq \frac{c\mathcal{L}}{\sqrt{N}} \|\theta - \theta'\|$，其中某个绝对常数 $c > 0$。 □

Lemma 3.3. $f_b(\theta) - \mathbb{E}f_b(\theta)$ 是一个相对于 $\frac{c\mathcal{L}}{\sqrt{N}} \|\cdot\|$ 的 ψ_1-过程，其中存在某个绝对常数 $c > 0$.

证明。 注意 $\left|\frac{1}{M} \sum_{i=1}^M \left(q_{ii}^1(\theta) - q_{ii}^1(\theta')\right)\right| \leq \mathcal{L} \|\theta - \theta'\|$ 并应用引理 2.7。 □

引理 3.4. $f_c(\theta) - \mathbb{E}f_c(\theta)$ 是关于 $\frac{c\mathcal{L}U}{\sqrt{N}} \|\cdot\|$ 的 ψ_1-过程，其中存在某个绝对常数 $c > 0$。证明。我们有

$$\begin{aligned}\nabla_\theta \sum_{i,j=1}^M \log\left(\frac{1}{M} p_{ij}^1(\theta) + \frac{M-1}{M}\right) - \nabla_\theta \frac{1}{M} \sum_{i,j=1}^M q_{ij}^1(\theta) &= \sum_{i,j=1}^M \left(\frac{\frac{1}{M} \nabla_\theta p_{ij}^1(\theta)}{\frac{1}{M} p_{ij}^1(\theta) + \frac{M-1}{M}} - \frac{1}{M} \nabla_\theta p_{ij}^1(\theta)\right) \\ &= \frac{1}{M^2} \sum_{i,j=1}^M \frac{(1 - p_{ij}^1(\theta)) \nabla_\theta p_{ij}^1(\theta)}{\frac{1}{M} p_{ij}^1(\theta) + \frac{M-1}{M}}.\end{aligned}$$

利用假设 $M \geq 2$，可得

$$\left\| -\nabla_\theta \sum_{i,j=1}^M \log\left(\frac{1}{M} p_{ij}^1(\theta) + \frac{M-1}{M}\right) + \nabla_\theta \frac{1}{M} \sum_{i,j=1}^M q_{ij}^1(\theta) \right\|_2 \leq 2\mathcal{L}U.$$

应用引理 2.7即可完成证明。 □

引理 3.5. 对于固定的 $\theta_0 \in \Theta$，存在一个绝对常数 $c > 0$，使得

$$\mathbb{P}\left(|f_a(\theta_0)| \geq t\right) \leq 2 \exp\left(-c \min\left\{\frac{t^2 N}{U^2}, \frac{tN}{U}\right\}\right).$$

证明。 证明过程与引理 3.2 的证明非常相似。令 $F_k(\theta_0) = \frac{1}{M} \sum_{i \neq j} q_{ij}^k(\theta_0)$。对于偶数 $l \in \mathbb{N}$，我们得到

$$\mathbb{E}\left(\left|F_1(\theta_0)\right|^l\right) = \frac{1}{M^l} \sum_{i_1 \neq j_1} \cdots \sum_{i_l \neq j_l} \mathbb{E}\left(\prod_{\alpha=1}^l q_{i_\alpha j_\alpha}^1\right) \leq \left(2l \sup_{x,y,\theta} \left|p^\theta(x,y) - 1\right|\right)^l \leq (2lU)^l .$$

利用 Jensen 不等式得到奇数 l 的界，然后由引理 2.3 可得 $\|F_1\|_{\psi_1} \leq 8eU$。现在应用定理2.4 即可完成证明。 □

引理 3.6. 对于固定的 $\theta_0 \in \Theta$，我们有

$$\mathbb{P}\left(|f_b(\theta_0) - \mathbb{E}f_b(\theta_0)| \geq t\right) \leq 2 \exp\left(-\frac{t^2 N}{2U^2}\right).$$

证明。 注意 $\left|\frac{1}{M}\sum_{i=1}^{M} q_{ii}^1(\theta)\right| \leq$ 上确界$_{x,\ y,\ \theta}\left|p^\theta(x,\ \ y)-1\right| \leq U$。应用引理 2.12。 □

引理 3.7. 对于固定的 $\theta_0 \in \Theta$，我们有

$$\mathbb{P}\left(\left|f_c(\theta_0)-\mathbb{E}f_c(\theta_0)\right| \geq t\right) \leq 2\exp\left(-\frac{t^2 N}{8U^4}\right).$$

证明. 注意，对于任意 $x>0$，存在一个介于 1 和 $1+x$ 之间的 $\xi=\xi(x)$，使得 $\log(1+x)=x-\frac{x^2}{2\xi^2}$。因此

$$\left|\sum_{i,j=1}^{M}\left(\log\left(1+\frac{1}{M}q_{ij}^1(\theta)\right)-\frac{1}{M}q_{ij}^1(\theta)\right)\right| = \left|\frac{1}{2M^2}\sum_{i,j=1}^{M}\frac{(q_{ij}^1(\theta))^2}{\xi^2}\right| \leq 2\sup_{x,y,\theta}\left|p^\theta(x,y)-1\right|^2 \leq 2U^2$$

其中 ξ 介于 1 和 $1+\frac{1}{M}q_{ij}^1$之间，并且由于 $M\geq 2$，我们有 $\inf|\xi| \geq 1/2$。应用引理 2.12。 □

定理 *1.8*的证明. 根据引理 3.5 f_a 以高概率在某一点有界；根据引理 3.2，它是一个 ψ_1-过程。引理 2.10 随后以高概率界定两两差值。最后，引理 2.11以高概率一致地界定 f_a。对于某个绝对常数 c，我们得到

$$\mathbb{P}\left(\sup_\theta|f_a(\theta)|\geq 2t\right) \leq 2\exp\left(-c\min\left\{\frac{t^2N}{U^2},\frac{tN}{U}\right\}\right)+2\exp\left(\frac{J}{D}\right)\exp\left(-\frac{ct\sqrt{N}}{D\mathcal{L}}\right).$$

类似地，对于 $f_b(\theta)-\mathbb{E}f_b(\theta)$（使用引理 3.3 和 3.6）以及 $f_c(\theta)-\mathbb{E}f_c(\theta)$（使用引理 3.4 和3.7），我们得到

$$\mathbb{P}\left(\sup_\theta|f_b(\theta)-\mathbb{E}f_b(\theta)|\geq 2t\right) \leq 2\exp\left(-\frac{t^2N}{2U^2}\right)+2\exp\left(\frac{J}{D}\right)\exp\left(-\frac{ct\sqrt{N}}{D\mathcal{L}}\right),$$

$$\mathbb{P}\left(\sup_\theta|f_c(\theta)-\mathbb{E}f_c(\theta)|\geq 2t\right) \leq 2\exp\left(-\frac{t^2N}{8U^4}\right)+2\exp\left(\frac{J}{D}\right)\exp\left(-\frac{ct\sqrt{N}}{D\mathcal{L}U}\right).$$

结合并集界使用公式（7）即可完成证明。 □

3.2 $\nabla f_M^N(\theta)$ 的集中

在本小节中，我们沿用第 $_\theta$ $||\nabla f_M^N(\theta)-\nabla f_M(\theta)||_2$ 节中的相同策略，证明梯度的 sup3.1。由于 Θ 是有限维的，我们可以逐项控制其范数。我们将 $\frac{\partial}{\partial\theta_l}$ 简记为 ∂_l。与之前一样，我们首先将其分解为

$$\begin{aligned}\partial_l f_M^N(\theta) &= -\frac{1}{N}\sum_{k=1}^{N}\sum_{i,j=1}^{M}\frac{\frac{1}{M}\partial_l p_{ij}^k(\theta)}{\frac{1}{M}p_{ij}^k(\theta)+\frac{M-1}{M}}\left(1-\frac{1}{M}+\frac{1}{M}-\frac{1}{M}p_{ij}^k(\theta)+\frac{1}{M}p_{ij}^k(\theta)\right)\\ &= \underbrace{-\frac{1}{NM}\sum_{k=1}^{N}\sum_{i\neq j}^{M}\partial_l p_{ij}^k(\theta)}_{f_a'(\theta)}\underbrace{-\frac{1}{NM}\sum_{k=1}^{N}\sum_{i=1}^{M}\partial_l p_{ii}^k(\theta)}_{f_b'(\theta)}+\underbrace{\frac{1}{NM^2}\sum_{k=1}^{N}\sum_{i,j=1}^{M}\frac{(p_{ij}^k(\theta)-1)\partial_l p_{ij}^k(\theta)}{\frac{1}{M}p_{ij}^k(\theta)+\frac{M-1}{M}}}_{f_c'(\theta)}\end{aligned}$$

然后使用三角不等式，得到

$$\sup_{\theta\in\Theta}\left|\partial_l f_M^N(\theta)-\partial_l f_M(\theta)\right| \leq \sup_{\theta\in\Theta}|f_a'(\theta)-\mathbb{E}f_a'(\theta)|+\sup_{\theta\in\Theta}|f_b'(\theta)-\mathbb{E}f_b'(\theta)|+\sup_{\theta\in\Theta}|f_c'(\theta)-\mathbb{E}f_c'(\theta)|.$$

以下引理的证明，除非另有说明，与上一节相应引理的证明完全相同，只需将 $q_{ij}^k(\theta)$ 替换为 $\partial_l p_{ij}^k(\theta)$。注意，$\int\partial_l p_{ij}^k(\theta)\mathrm{d}\mu(x)=\partial_l\int p_{ij}^k\mathrm{d}\mu(x)$，$=\partial_l 1=0$ 另一边缘分布也同样如此。由于定理 2.1，我们可以交换积分与微分，因为根据定义 1.2，p_{ij}^k 和 $\partial_l p_{ij}^k$ 是一致有界的。

引理 3.8. $f_a'(\theta)$ 是相对于 $\frac{c\mathcal{L}'}{\sqrt{N}}\|\cdot\|$ 的一个 ψ_1-过程，其中 $c>0$ 是某个绝对常数，$\|\cdot\|$ 是 Θ 上的原范数。

引理 3.9. $f_b'(\theta)-\mathbb{E}f_b'(\theta)$ 是关于 $\frac{c\mathcal{L}'}{\sqrt{N}}\|\cdot\|$ 的 ψ_1-过程，对某个绝对常数 $c>0$.

引理 3.10. $f_c'(\theta)-\mathbb{E}f_c'(\theta)$ 是关于 $\frac{cU(\mathcal{L}^2+\mathcal{L}')}{\sqrt{N}}\|\cdot\|$ 的 ψ_1-过程，对某个绝对常数 $c>0$。

证明。 我们采用与引理 3.4 的证明类似的方法。首先我们给出 Lipschitz 常数的界

$$\begin{aligned}\left\|\nabla_\theta\frac{1}{M^2}\sum_{i,j=1}^{M}\frac{q_{ij}\partial_l p_{ij}(\theta)}{\frac{1}{M}p_{ij}(\theta)+\frac{M-1}{M}}\right\|_2 &\le \sup_{x_1,y_1}\left\|\nabla_\theta\frac{q_{11}(\theta)\partial_l p_{11}(\theta)}{\frac{1}{M}p_{11}(\theta)+\frac{M-1}{M}}\right\|_2\\ &\le \sup_{x_1,y_1}\left\|\frac{\nabla_\theta[q_{11}(\theta)\partial_l p_{11}(\theta)]}{\frac{1}{M}p_{11}(\theta)+\frac{M-1}{M}}-\frac{q_{11}(\theta)\partial_l p_{11}(\theta)\nabla_\theta(\frac{1}{M}p_{11}(\theta)+\frac{M-1}{M})}{(\frac{1}{M}p_{11}(\theta)+\frac{M-1}{M})^2}\right\|_2\\ &\le 2\mathcal{L}^2+2U\mathcal{L}'+\frac{4U\mathcal{L}^2}{M}\le 4U(\mathcal{L}^2+\mathcal{L}')\end{aligned}$$

然后我们应用引理2.7。 □

引理 3.11. 对固定的 $\theta_0\in\Theta$，存在绝对常数 $c>0$，使得

$$\mathbb{P}\left(|f_a'(\theta_0)|\ge t\right)\le 2\exp\left(-c\min\left\{\frac{t^2N}{\mathcal{L}^2},\frac{tN}{\mathcal{L}}\right\}\right).$$

引理 3.12. 对固定的 $\theta_0\in\Theta$，我们有

$$\mathbb{P}\left(|f_b'(\theta_0)-\mathbb{E}f_b'(\theta_0)|\ge t\right)\le 2\exp\left(-\frac{t^2N}{2\mathcal{L}^2}\right).$$

引理 3.13. 对于固定的 $\theta_0\in\Theta$，我们有

$$\mathbb{P}\left(|f_c'(\theta_0)-\mathbb{E}f_c'(\theta_0)|\ge t\right)\le 2\exp\left(-\frac{t^2N}{8U^2\mathcal{L}^2}\right).$$

证明. 对于 $|(p^\theta-1)\partial_l p^\theta|$ 的上界为 $U\mathcal{L}$，对于 $|\frac{1}{M}p^\theta+\frac{M-1}{M}|$ 的下界为 $1/2$，因为 $M\ge 2$。应用引理 2.12。 □

定理 3.14. 设 J，D 与引理 *2.10*中的相同。存在一个绝对常数 $c>0$ 使得

$$\mathbb{P}\left(\sup_{\theta\in\Theta}\left|\partial_l f_M^N(\theta)-\partial_l f_M(\theta)\right|\ge 6t\right)\le 6\exp\left(-c\min\left\{\frac{t^2N}{U^2\mathcal{L}^2},\frac{tN}{\mathcal{L}}\right\}\right)+6\exp\left(\frac{J}{D}\right)\exp\left(-\frac{ct\sqrt{N}}{D(\mathcal{L}^2+\mathcal{L}')U}\right).$$

定理 3.15. 存在与 M、N 无关的常数 C_1、$C_2>0$，使得

$$\mathbb{P}\left(\sup_{\theta\in\Theta}\left\|\nabla f_M^N(\theta)-\nabla f_M(\theta)\right\|_2\ge t\right)\le C_1\exp\left(-C_2\min\{t^2N,t\sqrt{N}\}\right)$$

证明. 我们有

$$\begin{aligned}\mathbb{P}\left(\sup_{\theta\in\Theta}\left\|\nabla f_M^N(\theta)-\nabla f_M(\theta)\right\|_2\ge t\right)&\le\mathbb{P}\left(\sup_{\theta\in\Theta}\left\|\nabla f_M^N(\theta)-\nabla f_M(\theta)\right\|_\infty\ge\frac{t}{\sqrt{\dim(\Theta)}}\right)\\ &\le\sum_{m=1}^{M}\mathbb{P}\left(\sup_{\theta\in\Theta}\left|\partial_l f_M^N(\theta)-\partial_l f_M(\theta)\right|\ge\frac{t}{\sqrt{\dim(\Theta)}}\right).\end{aligned}$$

应用定理3.14并整理化简。 □

3.3 估计量的收敛

在本小节中，我们给出估计量 θ_M^N 的显式收敛速率。

引理 3.16。 $\nabla^2 f_M(\theta) \to \frac{1}{2}\nabla^2 \left\|p^\theta - p^*\right\|^2_{L^2(\mu\otimes\nu)} = \nabla^2 f_\infty(\theta)$ 一致地，当 $M \to \infty$ 时。

证明。 根据公式（5），有 $f_M(\theta) = -M^2 \int \log(\frac{1}{M}p^\theta + \frac{M-1}{M})(\frac{1}{M}p^* + \frac{M-1}{M})\mathrm{d}\mu\otimes\nu$。注意，根据定义 1.2， p^θ的值以及其一阶和二阶导数均一致有界。简单计算可知，对于 $M \geq 2$，$\log(\frac{1}{M}p^\theta + \frac{M-1}{M})$ 也具有相同性质。由此，定理 2.1 允许我们交换微分与积分的顺序，因此

$$
\begin{aligned}
-\nabla^2 f_M(\theta) &= M^2 \int \left(\frac{\frac{1}{M}\nabla^2 p^\theta}{\frac{1}{M}p^\theta + \frac{M-1}{M}} - \frac{\frac{1}{M^2}(\nabla p^\theta)\otimes(\nabla p^\theta)}{\left(\frac{1}{M}p^\theta + \frac{M-1}{M}\right)^2} \right) \left(\frac{1}{M}p^* + \frac{M-1}{M} \right) \mathrm{d}\mu\otimes\nu \\
&= M^2 \int \left(\frac{\frac{1}{M}\nabla^2 p^\theta}{\frac{1}{M}p^\theta + \frac{M-1}{M}} \right) \left(\frac{1}{M}(p^* - p^\theta) + \left(\frac{1}{M}p^\theta + \frac{M-1}{M} \right) \right) \mathrm{d}\mu\otimes\nu \\
&\quad - M^2 \int \left(\frac{\frac{1}{M^2}(\nabla p^\theta)^{\otimes 2}}{\left(\frac{1}{M}p^\theta + \frac{M-1}{M}\right)^2} \right) \left(\frac{1}{M}(p^* - p^\theta) + \left(\frac{1}{M}p^\theta + \frac{M-1}{M} \right) \right) \mathrm{d}\mu\otimes\nu \\
&= \int \frac{(p^* - p^\theta)\nabla^2 p^\theta}{\frac{1}{M}p^\theta + \frac{M-1}{M}} + M\nabla^2 p^\theta - \frac{1}{M}\frac{(p^* - p^\theta)(\nabla p^\theta)^{\otimes 2}}{\left(\frac{1}{M}p^\theta + \frac{M-1}{M}\right)^2} - \frac{(\nabla p^\theta)^{\otimes 2}}{\frac{1}{M}p^\theta + \frac{M-1}{M}} \mathrm{d}\mu\otimes\nu.
\end{aligned}
$$

利用恒等式 $\frac{1}{\frac{1}{M}p^\theta + \frac{M-1}{M}} = 1 + \frac{1}{M}\frac{1-p^\theta}{\frac{1}{M}p^\theta + \frac{M-1}{M}}$ 和 $\int \nabla^2 p^\theta \mathrm{d}\mu\otimes\nu = 0$，我们得到

$$
\begin{aligned}
-\nabla^2 f_M(\theta) &= \int \left(p^* - p^\theta\right) \nabla^2 p^\theta \left(1 + \frac{1}{M}\left(\frac{1-p^\theta}{\frac{1}{M}p^\theta + \frac{M-1}{M}} \right) \right) \mathrm{d}\mu\otimes\nu \\
&\quad - \int (\nabla p^\theta)^{\otimes 2} \left(1 + \frac{1}{M}\left(\frac{1-p^\theta}{\frac{1}{M}p^\theta + \frac{M-1}{M}} + \frac{p^* - p^\theta}{\left(\frac{1}{M}p^\theta + \frac{M-1}{M}\right)^2} \right) \right) \mathrm{d}\mu\otimes\nu.
\end{aligned}
$$

注意到 $\frac{1}{2}\nabla^2 \left\|p^\theta - p^*\right\|^2_{L^2(\mu\otimes\nu)} = \int (\nabla p^\theta)^{\otimes 2} - (p^* - p^\theta)\nabla^2 p^\theta \mathrm{d}\mu\otimes\nu$，因此

$$
\nabla^2 f_M(\theta) = \frac{1}{2}\nabla^2 \left\|p^\theta - p^*\right\|^2_{L^2(\mu\otimes\nu)} + \frac{1}{M}A(\theta)
$$

其中 $A(\theta)$ 是某个一致有界（关于 θ）的算子。 □

那么，在以下假设下，我们能够控制 $f_M(\theta)$ 在 M 较大时的曲率。

假设 3.17。 假设 $\theta_\infty \in \Theta$ 是 Θ 上 f_∞ 的唯一极小化子，并且存在某些τ、 $r > 0$，使得对于所有满足 $\|\theta - \theta_\infty\| \leq r$ 的 $\theta \in \Theta$， $\nabla^2 f_\infty(\theta)$ 的最小特征值都大于 τ。

定理 (定理 1.9的回顾)。 存在一个 $M_0 \in \mathbb{N}$ 使得 $\forall M \geq M_0$ 以下结论成立。

1. θ_M 的 f_M 极小化子唯一。
2. 存在一些 C_1、 C_2、 $t_0 > 0$，使得对于所有 $t \leq t_0$以及任意 $\theta_M^N \in$极小值点集合$_{\bar{\Theta}}$ f_M^N，我们有

$$
\mathbb{P}\left(\left\|\theta_M^N - \theta_M\right\|_2 \geq t \right) \leq C_1 \exp\left(-C_2 \min\left\{ t^2 N, t\sqrt{N} \right\} \right).
$$

证明。 假设 argmin$_{\tilde{\bar{\Theta}}}$ f_M $\{\theta_\infty\}$ 0 and 之间的最大成对距离在 $M \to \infty$ 时不趋于 0，即存在一个序列 $(\theta_M)_M$，使得 $\theta_M \in$argmin$_{\tilde{\bar{\Theta}}}$ $f_M$and lim sup$_M$ $||\theta_\infty - \theta_M|| > 0$.根据 Θ 的紧性，存在 f_M 的极小化子的一个收敛子列，但它并不收敛

到 θ_∞。根据假设 3.17 后者是 f_∞ 的唯一极小值点。因此，这是一个矛盾，因为根据定理 1.6，我们有 f_M 一致收敛到 f_∞，当 $M \to \infty$ 时。
此外，根据假设 3.17 和引理 3.16，存在一个以 θ_∞ 为中心、半径为 r 的闭球 $B(\theta_\infty, r)$，以及某个 $M_1 \in \mathbb{N}$，使得对所有 $M > M_1$，f_M 在 $B(\theta_\infty, r)$ 上严格凸。由证明的第一部分可知，当 M 大于某个 $M_2 > 0$ 时，极小值点集合 $\operatorname{argmin} f_M$ 也包含在 $B(\theta_\infty, r)$ 中。因此，对于 $M > \max(M_1, M_2)$，f_M 存在唯一的极小化子，记为 θ_M。

由于连续函数 f_∞ 的极小化子 θ_∞ 唯一，且 Θ 具有紧性，对每个 $\epsilon > 0$，我们有 $\inf\{f_\infty(\theta) \mid \theta \in \Theta, \|\theta - \theta_\infty\| \geq \epsilon\} > f_\infty(\theta_\infty)$，特别地，$\forall \epsilon > 0\, \exists \delta > 0$：$\|\theta - \theta_\infty\| \geq \epsilon/2 \Rightarrow f_\infty(\theta) \geq f_\infty(\theta_\infty) + 2\delta$。回顾一下，我们有 $\theta_M \to \theta_\infty$，且当 $M \to \infty$ 时，$f_M \to f_\infty$ 一致收敛（定理 1.6），因此存在一个 $M_3 > 0$（取决于 ϵ 和 δ），使得对所有 $M > M_3$ 和 $\theta \in \Theta$，我们有

$$\|\theta - \theta_M\| \geq \epsilon \Rightarrow \|\theta - \theta_\infty\| \geq \epsilon/2 \Rightarrow f_\infty(\theta) \geq f_\infty(\theta_\infty) + 2\delta \Rightarrow f_M(\theta) \geq f_M(\theta_M) + \delta. \tag{8}$$

令 $\epsilon < r/2$。对于 $\delta' := \sup_\theta |f_M^N(\theta) - f_M(\theta)|$，对于任意 $\theta_M^N \in$极小值点集合$_\Theta f_M^N$，我们都有

$$f_M(\theta_M^N) - \delta' \leq f_M^N(\theta_M^N) \leq f_M^N(\theta_M) \leq f_M(\theta_M) + \delta'.$$

如果 $\delta' < \delta/2$这将导致 $f_M(\theta_M^N) \leq f_M(\theta_M) + 2\delta' < f_M(\theta_M) + \delta$，根据（8）的逆否命题，我们有 $\|\theta_M^N - \theta_M\| < \epsilon < r/2$。由引理 3.16和假设 3.17，存在某个 $M_4 > 0$，使得对于所有 $M \geq M_4$，$\nabla^2 f_M(\theta)$的最小特征值都大于 $\tau/2$，其中 $\theta \in B(\theta_M, r/2)$。考虑路径 $\gamma(t) = t\theta_M^N + (1-t)\theta_M$，其中 $t \in [0,1]$。我们得到

$$\begin{aligned}
\left\|\nabla f_M(\theta_M^N) - \nabla f_M^N(\theta_M^N)\right\|_2 &= \left\|\nabla f_M(\theta_M^N) - \nabla f_M(\theta_M)\right\|_2 \\
&= \|\nabla f_M(\gamma(1)) - \nabla f_M(\gamma(0))\|_2 \\
&= \left\|\int_0^1 \nabla^2 f_M(\gamma(t)).\dot{\gamma}(t)\mathrm{d}t\right\|_2 \\
&= \left\|\int_0^1 \nabla^2 f_M(\gamma(t)).(\theta_M^N - \theta_M)\mathrm{d}t\right\|_2 \\
&\geq \left\langle \frac{\theta_M^N - \theta_M}{\left\|\theta_M^N - \theta_M\right\|_2}, \int_0^1 \nabla^2 f_M(\gamma(t)).(\theta_M^N - \theta_M)\mathrm{d}t \right\rangle \\
&= \int_0^1 \left\langle \frac{\theta_M^N - \theta_M}{\left\|\theta_M^N - \theta_M\right\|_2}, \nabla^2 f_M(\gamma(t)).(\theta_M^N - \theta_M) \right\rangle \mathrm{d}t \\
&\geq \frac{\tau}{2}\left\|\theta_M^N - \theta_M\right\|_2.
\end{aligned}$$

综合以上内容，我们可以得出结论

$$\sup_\theta |f_M^N(\theta) - f_M(\theta)| < \frac{\delta}{2} \Rightarrow ||\theta_M^N - \theta_M||_2 \leq \frac{2}{\tau}\sup_\theta \left\|\nabla f_M(\theta) - \nabla f_M^N(\theta)\right\|_2.$$

应用定理 1.8和 3.15，经过一些直接的整理后，我们证明了这一结论。 □

4 数值示例

本节给出两个合成数值示例，以说明我们的估计量的关键性质，并与理论结果相符。[1]

4.1 二维环面上的点共定位

设 $\mathbb{X} = \mathbb{Y} = \mathbb{R}^2/\mathbb{Z}^2$ 为二维环面，π 定义为

$$(X, Y) \sim \pi \quad \Leftrightarrow \quad \begin{cases} X & \sim \mathcal{U}(\mathbb{R}^2/\mathbb{Z}^2) \\ Y|X = x & \sim \tilde{\mathcal{N}}(x, \sigma^2 I). \end{cases}$$

[1]代码可在 https://github.com/OTGroupGoe/BatchedBrokenSamples获取

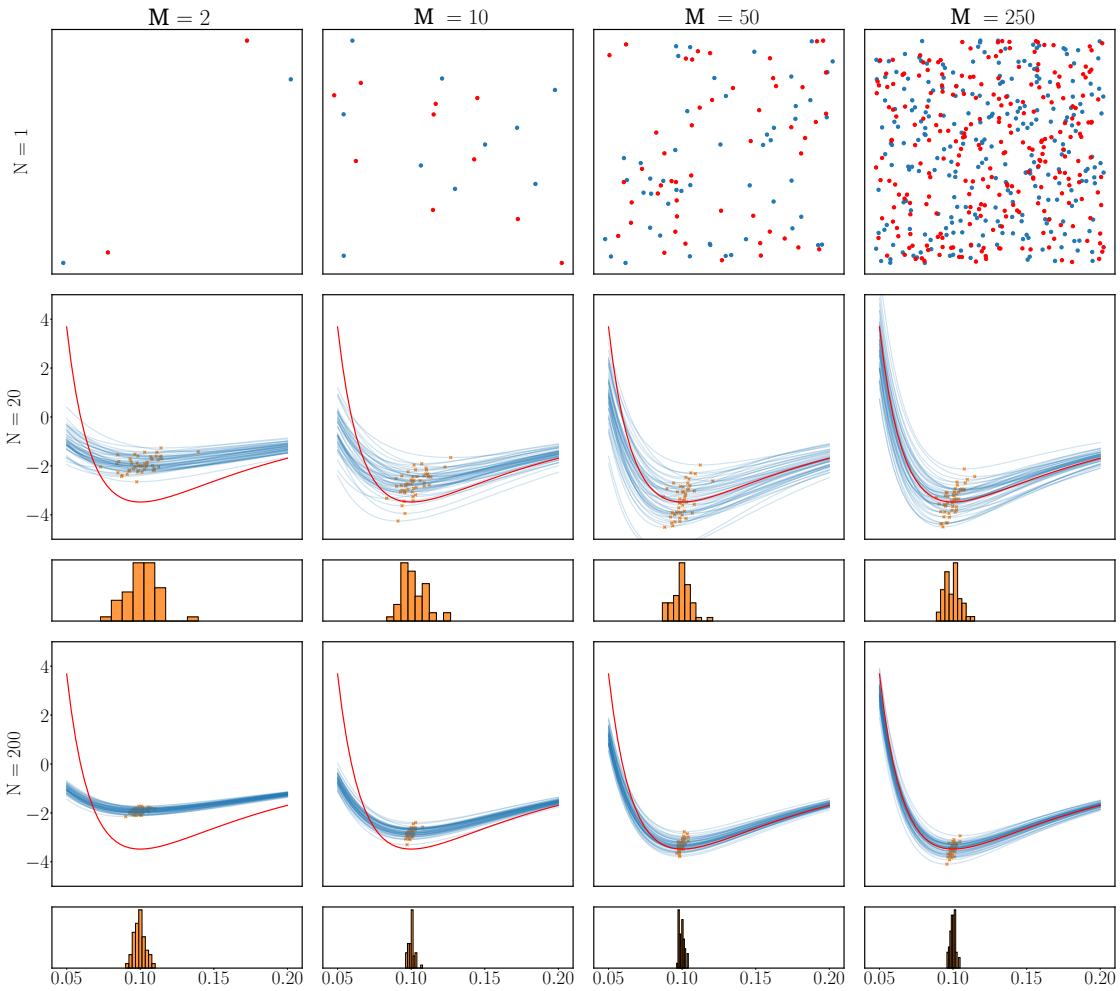

图 2：针对 $\sigma^* = 0.1$ 的数值实验，第一行展示了来自单个批次的示例点云。每条蓝线描绘了由样本计算得到的 $f_M^N(\sigma)$，其中 M 和 N 如相应列/行所示。每个绘图包含 50 个独立样本。橙色点表示极小值，下方的直方图展示了它们的分布。红线为 $f_\infty(\sigma) = \frac{1}{2}||p^\sigma - p^{\sigma^*}||^2_{L^2} + \frac{1}{2} - \frac{1}{2}||p^{\sigma^*}||^2_{L^2}$。

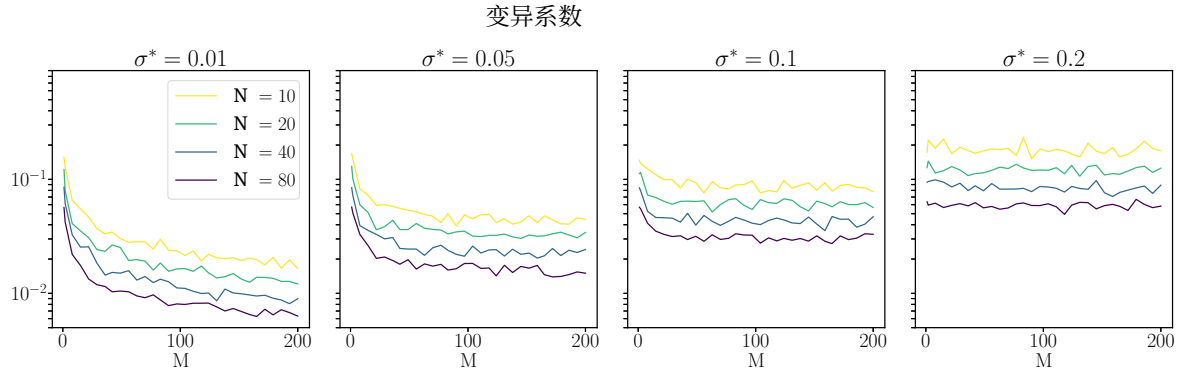

图 3：σ_M^N 的变异系数，由 100 次模拟计算得出，对应于 σ^*, N 和 M 的不同取值。

这里，$\mathcal{U}$和 $\tilde{\mathcal{N}}$（x，$\sigma^2 I$）分别表示二维环面上以 x 为中心的均匀分布和包裹的 $2D$-高斯分布，其协方差矩阵为 $\sigma^2 I$；σ 是我们想要估计的参数。请注意，在这种设定下，π 的两个边缘分布都是环面上的均匀分布。因此，密度为

$$p^{\sigma}(x,y)=\sum_{k\in\mathbb{Z}^2}\frac{1}{2\pi\sigma^2}\exp\Big(-\frac{\|x-y+k\|_{\mathbb{R}^2}^2}{2\sigma^2}\Big).$$

我们使用真实值 $\sigma^*=0.1$ 和 $M\in\{2,10,50,250\}$ 以及 $N\in\{20,200\}$生成样本。在图2的第一行中，我们展示了在不同 M取值下的一个破损样本批次。红色点（X）和蓝色点（Y）是成对生成的，但从这些样本中无法观察到配对信息。尽管 M 较小时可能猜出配对关系，但 M 较大时显然不可能做到。

图2的第二行展示了 $N=20$ 取值下、不同 M对应的模拟结果。红色曲线表示极限 $f_\infty(\sigma)=\frac{1}{2}\|p^{\sigma}-p^{\sigma^*}\|_{L^2}^2+\frac{1}{2}-\frac{1}{2}\|p^{\sigma^*}\|_{L^2}^2$，其在 $\bar{\sigma}=\bar{\sigma}^*(=0.1)$处达到最小值。对于每个（$N$，$M$），我们进行50次模拟，并将相应的 $f_M^N(\sigma)$）绘制为蓝色曲线。随着 M 增大，这些曲线逐渐聚集在红色曲线周围（定理1.6）。每次模拟得到的经验极小值点 σ_M^N 以橙色标出；第三行则给出了估计量经验分布的直方图。

第 4 行和第 5 行与第 2 行和第 3 行类似，其中 $N=200$。可以观察到，经验损失函数和估计量比 $N=20$对应的更为集中，这符合预期（定理 1.8和 1.9）。

在图 3中，我们展示了 σ_M^N的变异系数（CV），即经验标准差除以经验均值的结果，该结果由 100 次模拟计算得出，所用参数为 $\sigma^*\in\{0.01,0.05,0.1,0.2\}$、$N\in\{10,20,40,80\}$和 M，其取值从 1 变化到 200。

作为经过缩放的标准差，变异系数如预期那样以大约 $O(N^{-1/2})$ 的速率下降。对于$M=1$ 且固定 N的情况，它们起初都处于近似相同的水平；我们还可以观察到，当 σ^* 和 M 较小时，σ^* 会随 M急剧下降。直观地说，这是因为在这个低密度区域中，配对以较高概率能够被正确恢复，因此增加 M 的效果类似于增加 N。随着 M 和 σ^* 增大，提高 M 所带来的益处会逐渐减弱。

4.2 二元正态分布的协方差估计

断裂样本问题的另一个有趣情形是：在边缘分布已知的情况下，估计二元正态分布的相关系数（例如参见 [5, 第 4] 节）。令 $\mathbb{X}=\mathbb{Y}=\mathbb{R}$，并且

$$\begin{bmatrix}X\\Y\end{bmatrix}\sim\pi:=\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1&\rho^*\\\rho^*&1\end{bmatrix}\right)$$

对于某个相关参数 $\rho^*\in(-1,1)$。回顾一下，我们的目标是估计 $\mathrm{d}\pi\,\mathrm{d}(\mu\otimes\nu)$，其中 μ 和 ν 都是定义在 $\mathbb{R}$ 上的标准正态分布。令 $\varphi_{X,Y}^{\rho}$ 表示相关参数为 ρ 的 π 的勒贝格密度，并令 $\varphi_X=\varphi_Y$ 表示 $\mu=\nu=\mathcal{N}(0,1)$ 的勒贝格密度。于是，参数化密度类简单地取为：

$$p^{\rho}(x,y):=\frac{\varphi_{X,Y}^{\rho}(x,y)}{\varphi_X(x)\varphi_Y(y)}\qquad\forall\rho\in[-1+\tau,1-\tau].$$

这里，我们任取一个 $\tau>0$，使其满足定义 1.2中的条件。直接计算可得，对于任意 ρ，$\rho^*\in[-1+\tau,1-\tau]$:

$$\left\|p^{\rho^*}\right\|_{L^2(\mu\otimes\nu)}^2=\frac{1}{(1-\rho^{*2})},\qquad\left\|p^{\rho^*}-p^{\rho}\right\|_{L^2(\mu\otimes\nu)}^2=\frac{1}{(1-\rho^2)}+\frac{1}{(1-\rho^{*2})}-\frac{2}{|\rho\rho^*-1|}$$

我们选取真实的 $\rho^*=-0.5$，并使用 $M\in\{2,10,50,250\}$和 $N\in\{20,200\}$进行模拟。在图 4的第一行中，我们展示了不同 M下的一个破损样本批次，请注意这仅

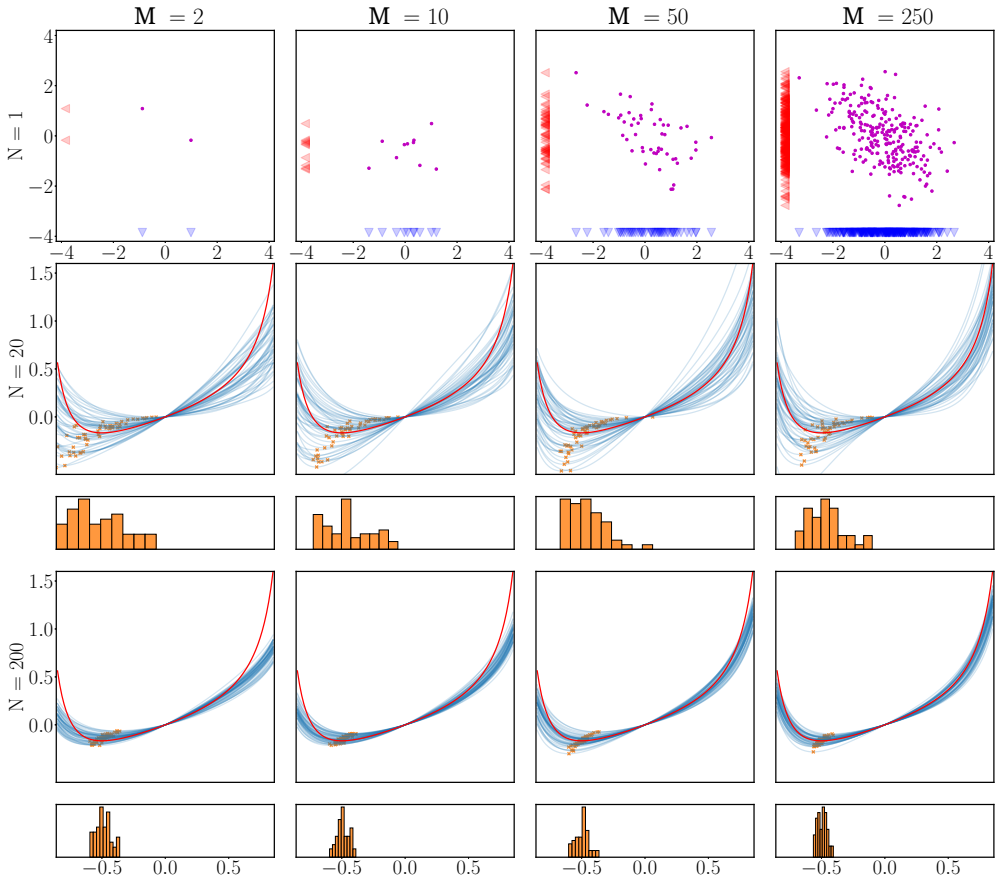

图 4：来自二元正态分布的破损随机样本数值实验，其中$\rho^* = -0.5$。第一行展示了来自单个批次的一个示例；紫色点是“未断裂”的（x，y）对，而在破损样本设置下，只能观测到坐标轴上的蓝色和红色边缘点。其余面板与图2类似，红色曲线为 $f_\infty(\rho) = \frac{1}{2}\|p^\rho - p^{\rho^*}\|^2_{L^2(\mu\otimes\nu)} + \frac{1}{2} - \frac{1}{2}\|p^{\rho^*}\|^2_{L^2(\mu\otimes\nu)}$。

包括坐标轴上的蓝色点和红色点；紫色点是未断裂点（x， y），因此在断裂样本设定下不可用。其余面板与环面示例中的图2类似，并呈现出非常相似的 f_M^N行为。容易验证， $p^0(x,\ y) = 1$对所有（x， y）$\in \mathbb{R}^2$均成立，这解释了曲线在 $\rho = 0$处的交点。

参考文献

[1] Martial Agueh、Boualem Khouider 和 Louis-Philippe Saumier。粒子图像测速的最优传输。数学科学通讯，13(1):269–296，2015年。[2] Herbert Amann 和 Joachim Escher。分析 *III*。施普林格，2009年。[3] Zhidong Bai 和 Tailen Hsing。断裂样本问题。概率论及相关领域，131:528–552，2005年。[4] Florian Beier、Hancheng Bi、Cl´ement Sarrazin、Bernhard Schmitzer 和 Gabriele Steidl。利用熵传输核，从批次未配对点构建传递算子。信息与推断：*IMA*期刊，14(2):iaaf005，2025年04月。[5] Hock-Peng Chan 和 Wei-Liem Loh。重新审视 DeGroot 和 Goel 的文件链接问题。中华统计学志，第1031–1045页，2001年。[6] Felipe Cucker 和 Steve Smale。论学习的数学基础。美国数学会通报，39(1):1–49，2002年。[7] Morris H DeGroot、Paul I Feder 和 Pre m K Goel。匹配。数学统计年刊，第578–593页，1971年。[8] Morris H. DeGroot 和 Prem K. Goel。从断裂随机样本估计相关系数。统计年刊，8(2):264–278，1980年。[9] Jan-Niklas Dohrke。*PUCK*：用于高通量超分辨率显微镜的 *CRISPR/Cas* 引物辅助敲入技术。博士论文，哥廷根大学，2024年。[10] Peter Koltai、Johannes von Lindheim、Sebastian Neumayer 和 Gabriele Steidl。利用最优传输计划进行相干集检测的传递算子。物理学 *D*，426:132980，2021年。[11] Carla Tameling、Stefan Stoldt、Till Stephan、Julia Naas、Stefan Jakobs 和 Axel Munk。通过最优传输实现超分辨率显微镜的共定位。自然·计算科学，1(3):199–211，2021年。[12] L.G. Valiant。计算矩阵永久式的复杂度。理论计算机科学，8(2):189–201，1979年。[13] Roman Vershynin。高维概率：导论及其在数据科学中的应用，第47卷。剑桥大学出版社，2018年。[14] Martin J Wainwright。高维统计：非渐近视角，第48卷。剑桥大学出版社，2019年。