贝叶斯多重多元密度—密度回归

KhaiNguyen[1], Yang Ni[1], 和 Peter Mueller[1,2]1德克萨斯大学奥斯汀分校统计与数据科学系2德克萨斯大学奥斯汀分校数学系

2026年1月7日

arXiv:2601.02640v1 [stat.ME] 6 Jan 2026

摘要

我们提出了首个多重多元密度-密度回归（MDDR）方法，从而能够研究以多个多元密度值预测变量回归一个多元密度值响应。其核心思想是，利用预测变量推前分布的切片瓦瑟斯坦重心（SWB）定义拟合分布，并以切片瓦瑟斯坦（SW）距离量化拟合结果与观测响应之间的偏差。在广义贝叶斯框架下推断回归函数（将预测变量的支撑集映射到响应变量的支撑集）和重心权重，无需完全指定似然即可进行有原则的不确定性量化。该推断过程可视为逆向 SWB 问题的一个实例。我们建立了理论保证，包括 SWB 对边缘分布和重心权重扰动的稳定性、广义似然的样本复杂度以及后验一致性。在实际推断中，我们提出了 SWB 的可微近似，并采用平滑重参数化处理重心权重的单纯形约束，从而能够高效地进行基于梯度的马尔可夫链蒙特卡罗采样。我们通过一项群体规模单细胞数据推断应用展示了 MDDR。在该示例中，基于 MDDR 模型的后验分析包括推断多个源/发送细胞类型与一个靶/接收细胞类型之间的通信。所提出的方法能够提供准确的拟合、可靠的预测，以及可解释的重心权重后验估计；这些估计可用于构建稀疏的细胞间通信网络。

关键词：切片瓦瑟斯坦重心、最优传输、单细胞数据、广义贝叶斯。

1 引言

我们提出了多元密度—密度回归，针对多元密度值响应变量和若干多元密度值预测变量。所提出的方法不对预测变量和响应变量空间施加黎曼几何结构，而这正是现有密度-密度回归方法的一项关键要求。该方法具有良好的计算和统计可扩展性，并且能够处理任何形式的预测变量和响应变量，包括离散型和连续型。其主要特点是利用切片瓦瑟斯坦重心（Bonneel 等人，2015年）（SWB），根据多个预测变量的推前分布（边缘分布）定义拟合分布；并利用切片 Wasserstein 距离（Rabin 等，2012）（SW）度量拟合分布与观测到的响应分布之间的偏差。我们采用广义贝叶斯（Bissiri 等（2016）框架，对回归函数进行推断。回归函数将预测变量的支撑集映射到响应变量的支撑集，以定义推前分布和重心权重。广义贝叶斯设定无需完整指定似然，即可实现有原则的信念更新。尽管我们的首要目标是开发一种回归模型，但该方法也可被视为首个逆 *SWB* 问题实例，即在已知重心的含噪观测值时，推断其边缘分布及对应的重心权重。

除方法上的贡献外，我们还确立了所提方法的若干理论性质。首先，我们通过量化边缘分布、重心权重或两者受到扰动时 SWB 的变化，分析了 SWB 的稳定性。在此稳定性结果的基础上，我们推导出两个关键结果：广义似然的样本复杂度和后验一致性。具体而言，我们证明，即使只能通过样本观测分布，仍然可以可靠地估计广义似然；此外，我们还证明了回归函数和重心权重的贝叶斯估计量具有一致性。对于后验推断，我们提出了实用的后验模拟方案。首先，我们

使用迭代算法构造 SWB 的可微近似。其次，我们通过平滑且可逆的变量变换处理重心权重的单纯形约束。结合切片 Wasserstein 距离的可微性，这使得我们能够使用计算效率高的基于梯度的马尔可夫链蒙特卡罗（MCMC）采样器，例如 Metropolis 调整的 Langevin 算法（MALA）（Girolami & Calderhead 2011年，Ning 2025年）。

密度回归涵盖了使用分布值预测变量预测实值响应变量的情形（Szabó 等，2016年，Matabuena & Petersen，2023），也涵盖了使用实值预测变量对分布值响应变量建模的情形（Tokdar 等，2004年，Dunson 等，2007年，Shen & Ghosal，2016年）。密度-密度回归（DDR）将其推广到预测变量和响应变量均为分布的情形。例如，Zhao 等（2023）基于 Wasserstein-2 空间上的分布黎曼几何，研究了简单的单变量 DDR，其中预测变量和响应变量均为一维。将这一方法推广到多元分布的主要挑战在于，高维情形下黎曼几何工具在计算上难以处理。此外，该方法还要求分布连续，才能借助最优传输映射的存在来获取黎曼结构。为克服这些局限，Nguyen, Ni& Mueller （2025年）提出了一种基于广义贝叶斯框架（Bissiri 等（2016））的简单多元 DDR 模型。其核心建模策略是采用切片 Wasserstein（SW）距离（Rabin 等，2012）构造广义似然，从而实现可扩展的推断，而无需预测变量或响应变量具备任何黎曼结构

考虑多个预测变量便引出了多元回归问题。在 DDR 的背景下，Chen 等提出了多元单变量 DDR，见 Chen 等 (2024年)。同样，该方法需要分布连续，以便利用 Wasserstein 几何

进行计算，但这在多元情形下并不实用。我们将Nguyen, Ni & Mueller (2025年) 中的方法扩展到多元回归情形。首先，利用回归函数将预测变量分布转换为与响应分布相同的维数。随后，我们将转换后的预测变量分布的 SWB 作为拟合分布。然后，基于拟合分布与响应变量之间的切片 Wasserstein 距离，定义一个采用广义似然的推断模型。

使用 SWB 和切片 Wasserstein 距离解决了上述多元和有限样本挑战。接下来的讨论中有一项结果刻画了重心（预测）如何随边缘分布和重心权重的变化而变化。据此，我们证明，在分布具有紧支撑时，采用简单的插件估计，广义似然可达到近乎参数估计速率的 $\mathcal{O}(n^{-1/2})$（n 是所涉及分布的最小原子数）。此外，我们证明，在参数设定下，我们的模型能够一致地估计重心权重和回归函数的贝叶斯估计。对于切片 Wasserstein 文献和广义贝叶斯文献而言，这些理论研究都是新的。

我们的方法受到群体规模单细胞数据分析的启发：这类数据在大量从多个受试者采集的细胞中测量基因表达水平。由于可以利用成熟的细胞标记物可靠地识别细胞类型，这些数据为通过配体–受体对的表达揭示细胞类型之间的通信模式提供了独特机会。与独立推断成对细胞类型之间通信的 DDR 不同，多重 DDR（MDDR）能够同时对多个源细胞类型进行联合回归，以预测目标细胞类型。重要的是，对于每种细胞类型，我们只能观测到有限数量（尽管可能很多）的细胞，因此估计最优传输映射并进而采用Chen 等 (2024年)、Zhu & Müller (2023中所用的黎曼几何工具并不切实际。此外，我们的方法无需连续性

假设底层分布满足某些条件，并且能够超越欧几里得几何，在分布的原子上进行运算。

借助所提出的计算技术，我们能够对MDDR模型进行推断。我们表明，当预测变量所在空间的维数低于响应变量所在空间时（这在单细胞数据中十分常见），该模型尤为适用。该模型对模拟数据和真实单细胞数据集均能实现准确拟合与预测。此外，重心权重的后验分布能够用于构建加权的细胞间通信网络。通过在推断框架中明确考虑边纳入决策，可对其稍作扩展，从而获得稀疏网络。

本文其余部分安排如下。第2 节回顾 Wasserstein距离、切片 Wasserstein 距离和 SWB 的定义。第3节介绍贝叶斯MDDR框架，该框架使用切片 Wasserstein 距离和 SWB 定义广义似然，建立关键的理论性质，并通过 MALA采样器开展后验模拟。第4节展示模拟数据集上的结果，重点说明 MDR 与 DDR 下多元回归的根本差异。第5节将贝叶斯 MDDR 应用于单细胞数据，探讨模型拟合、预测性能以及细胞间通信网络的发现。第6 节总结全文，并讨论未来研究方向。技术性证明和更多实验结果见补充材料。

在记号方面，令 δ_x 表示位于 x处的狄拉克δ测度。对于任意 $d \geq 2$，定义单位超球面 $\mathbb{S}^{d-1} = \{\theta \in \mathbb{R}^d : \|\theta\|_2 = 1\}$。对于两个序列 a_n 和 b_n，如果存在一个普适常数 C，使得 $a_n \leq C\, b_n$ 对所有 $n \geq 1$均成立，我们记 $a_n = \mathrm{O}(b_n)$。给定两个可测空间 $(\mathrm{X1}, \Sigma_1)$ 和 $(\mathrm{X2}, \Sigma_2)$、一个可测函数 $f : \mathrm{X}_1 \to \mathrm{X}_2$, 以及 $(\mathrm{X1}, \Sigma_1)$ 上的测度 μ， μ经过 f的推前定义为 $f_\sharp\mu(B) = \mu(f^{-1}(B))$，对所有 $B \in \Sigma_2$.均成立。我们将 Δ^K 记为 K-单纯形，即 $(\pi_1, \ldots, \pi_K) \in \Delta^K$ 蕴含 $0 \leq \pi_k \leq 1, \forall k = 1, \ldots, K$,

以及 $\sum_{k=1}^{K} \pi_k = 1$。其他记号将根据需要引入。

2 背景

简要回顾 Wasserstein距离、切片 Wasserstein 距离和切片瓦瑟斯坦重心后，我们引入一些符号和定义。给定 $p \geq 1$，令G_1，$G_2 \in \mathcal{P}_p(\mathbb{R}^d)$ 其中 $\mathcal{P}_p(\mathbb{R}^d)$ 是所有支撑于 $\mathbb{R}^d$ 且具有有限p阶矩的分布集合。Wasserstein-p 距离（Villani 2009年，Peyré 和 Cuturi 2019年）定义 G_1 与G_2 之间的距离如下：

$$W_p^p(G_1, G_2) = \inf_{\pi \in \Pi(G_1, G_2)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|_p^p \, \mathrm{d}\pi(x, y), \tag{1}$$

其中

$$\Pi(G_1, G_2) = \left\{ \pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) \mid \pi(A, \mathbb{R}^d) = G_1(A),\ \pi(\mathbb{R}^d, B) = G_2(B)\ \forall A, B \subset \mathbb{R}^d \right\}$$

是所有运输计划/耦合构成的集合。在一维情形下，Wasserstein距离有如下闭式表达：

$$W_p^p(G_1, G_2) = \int_0^1 |Q_{G_1}(t) - Q_{G_2}(t)|^p \mathrm{d}t, \tag{2}$$

其中，Q_{G_1} 和 Q_{G_2} 分别是 G_1 和 G_2的分位数函数。

为利用 $d = 1$的闭式解，引入切片 Wasserstein（SW）距离（Rabin 等，2012，Nguyen 2025年）。具体而言，两个分布 G_1，$G_2 \in \mathcal{P}_p(\mathbb{R}^d)$ 之间的 SW 距离定义如下：

$$\mathrm{SW}_p^p(G_1, G_2) = \mathbb{E}_{\theta \sim \mathcal{U}(\mathbb{S}^{d-1})}[W_p^p(\theta \sharp G_1, \theta \sharp G_2)], \tag{3}$$

其中，$\mathcal{U}(\mathbb{S}^{d-1})$ 表示 d 维$\mathbb{S}^{d-1}$单位超球面上的均匀分布；而 $\theta\sharp G_1$ 和 $\theta\sharp G_2$ 分别表示 G_1 和 G_2 经投影$f_\theta(x)=\theta^\top x$得到的推前分布。SW 的计算依赖于对式（3）中期望的蒙特卡罗估计。例如，使用简单的蒙特卡罗估计：

$$\widehat{SW}_p^p(G_1,G_2;L)=\frac{1}{L}\sum_{l=1}^{L}W_p^p(\theta_l\sharp G_1,\theta_l\sharp G_2), \tag{4}$$

其中 $\theta_1,\ldots,\theta_L \overset{i.i.d.}{\sim}\mathcal{U}(\mathbb{S}^{d-1})$， L 为蒙特卡洛样本数或投影数。

与 Wasserstein距离相比，SW 的主要优势体现在分布由独立同分布样本表示时。具体而言，当观测到 $x_1,\ldots,x_{m_1}\overset{i.i.d.}{\sim}G_1$ 和$y_1,\ldots,y_{m_2}\overset{i.i.d.}{\sim}G_2$时，我们有 $\mathbb{E}\left[\left|W_p(\hat{}G_1,\hat{}G_2)-W_p(G_1,G_2)\right|\right]=\mathrm{O}(m_1^{-1/d}+m_2^{-1/d})$ (Fournier & Guillin 2015)，以及 $\mathbb{E}\left[\left|\mathrm{SW}_p(\hat{}G_1,\hat{}G_2)-\mathrm{SW}_p(G_1,G_2)\right|\right]=\mathrm{O}(m_1^{-1/2}+m_2^{-1/2})$ (Nadjahi et al. 2020, Nietert et al. 2022)，其中 $d>1$ 为维数，且 $\hat{}G_1=\frac{1}{m_1}\sum_{i=1}^{m_1}\delta_{x_i}$和 $\hat{}G_2=\frac{1}{m_2}\sum_{j=1}^{m_2}\delta_{y_j}$。除样本复杂度更优外，SW 的时间复杂度也更低，为 $\mathrm{O}\{(m_1+m_2)\log(m_1+m_2)\}$ (Peyré et al. 2019)；相比之下，Wasserstein 的时间复杂度为 $\mathrm{O}\{(m_1+m_2)^3\log(m_1+m_2)\}$ (Peyré et al. 2019)。 —

切片 Wasserstein 距离在分布空间 $\mathcal{P}_p(\mathbb{R}^d)$上诱导出一个度量。加权平均（Fréchet 均值）的概念被推广为 SW重心（Bonneel et al. 2015）（SWB）的概念。由 $K\geq 2$ 个边缘分布 $G_1,\ldots,G_K\in\mathcal{P}_p(\mathbb{R}^d)$ 及其边缘权重 $(\pi_1,\ldots,\pi_K)\in\Delta^K$ 定义的 SWB（Bonneel et al. 2015）为：

$$\mathrm{SWB}_p(G_1,\ldots,G_K,\pi_1,\ldots,\pi_K)=\mathrm{argmin}_{G\in\mathcal{P}_p(\mathbb{R}^d)}\sum_{k=1}^{K}\pi_k\mathrm{SW}_p^p(G,G_k). \tag{5}$$

我们将在后续章节讨论 SWB 的计算。

3 多重多元贝叶斯密度-密度回归

3.1 所提模型

我们考虑这样一个推断问题：将分布值响应 $G_i \in \mathcal{P}_p(\mathbb{R}^d)$ $(p \geq 1, d \geq 2)$ 对多个分布值预测变量 $F_{i1}, \ldots, F_{iK}$ $(K \geq 2)$ 进行回归，其中 $F_{ik} \in \mathcal{P}_p(\mathbb{R}^{h_k})$ 对于 $k = 1, \ldots, K$ 以及 $i = 1, \ldots, N$ $(N > 0)$。我们将数据记为 $\mathcal{S} = \{(F_{i1}, \ldots, F_{iK}, G_i)\}_{i=1}^N$。我们通过广义似然（Bissiri 等（2016））$\ell(f_1, \ldots, f_K; G_i, F_{i1}, \ldots, F_{iK})$ 引入 MDDR，该广义似然基于针对 $\widetilde{G}_i$ of G_i的拟合近似的损失函数：

$$\ell(f_1, \ldots, f_K, \pi_1, \ldots, \pi_K; G_i, F_{i1}, \ldots, F_{iK}) = \exp\left\{-w\, \mathrm{SW}_p^p[\widetilde{G}_i(f, \pi), G_i)]\right\}, \tag{6}$$

$$\widetilde{G}_i(f, \pi) = \mathrm{SWB}_p(f_1 \sharp F_{i1}, \ldots, f_K \sharp F_{iK}, \pi_1, \ldots, \pi_K), \tag{7}$$

其中，f_k 是一个可测函数，对于 $k = 1, \ldots, K$，$w > 0$，将 $\mathbb{R}^{h_k}$ 映射到 $\mathbb{R}^d$；$(\pi_1, \ldots, \pi_K) \in \Delta^K$ $(0 \leq \pi_k \leq 1, \forall k = 1, \ldots, K$，以及 $\sum_{k=1}^K \pi_k = 1)$，而 $f_k \sharp F_{ik}$ 表示 F_{ik} 经 f_k得到的推前测度。

~在（6）中，G_i 是表示响应变量（样本内）预测的拟合分布~分布 G_i 给定 F_{i1}、.....、F_{iK}。我们将 G_i 定义为随机分布的 SWB$f_1 \sharp F_{i1}$、.....、$f_K \sharp F_{iK}$，我们将其称为预测变量的推前分布。拟合分布$\widetilde{G}_i$ 可以看作预测变量的推前分布的 Fréchet 均值（广义加权平均）。随后，我们使用 SW 距离定义损失函数，将其作为广义（负对数）似然。广义贝叶斯框架（6）和（7）为多重多变量密度-密度回归提供了一种有原则的方法，同时在贝叶斯框架中自然地纳入不确定性（先验分布将在下一小节讨论）。

该模型可以视为一种新型推断范式的实例，我们将其称为逆切片瓦瑟斯坦重心。具体而言，我们将响应变量 G_i 视为一个含噪重心，并旨在推断生成它的潜在边缘分布和权重。在我们的表述中，重心权重 π_k 在 i之间共享，而边缘分布则通过共享的回归函数 f_k 构造，其输入为预测变量分布。为评估拟合优度，我们计算拟合响应分布与观测响应分布之间的残差误差（6）和（7）： PE（$\mathcal{S}$） $= \frac{1}{N}\sum_{i=1}^{N}$ SW$_p^p$（$\widetilde{G}_i$、G_i）。为校准 PE，我们引入一个代表最坏情况拟合的参考模型（稍后将介绍具体示例）： $\widetilde{G}'_i$（$f_{\prime}\pi'$） $=$ SWB$_p\Big(f_{\prime 1}\sharp F_{i1}$、……、$f_{\prime K}\sharp F_{iK}$； $1/K$、……、$1/K\Big)$，其中 $f_{\prime}$表示受到严格限制的回归函数。随后，我们将预测误差相对于该参考模型进行归一化，从而得到相对残差误差（RE）：

$$\mathrm{RE}(\mathcal{S}) = \frac{1}{N}\sum_{i=1}^{N}\frac{\mathrm{SW}_p^p(\widetilde{G}_i(f,\pi), G_i)}{\mathrm{SW}_p^p(\widetilde{G}'_i(f',\pi'), G_i)}. \tag{8}$$

这里，RE($\mathcal{S}$) $= 0$ 表示完美拟合，因为 SW 是分布之间的一种距离；而 RE($\mathcal{S}$) $= 1$则表示拟合较差。实践中，我们将按照讨论的内容，使用蒙特卡洛样本近似 SW距离。

3.2 似然样本复杂度与后验一致性

我们研究所提模型的两个理论方面。第一个结果刻画了以独立同分布样本表示分布时广义似然的样本复杂度。第二个结果确立了后验一致性。简而言之，在回归函数的参数设定下，随着样本量增加，后验会集中于真实参数附近。为准备这些结果，我们首先研究 SWB 的稳定性，即在预测变量和重心权重发生扰动时，重心在切片 Wasserstein 距离下会变化多少。

引理 1 (切片瓦瑟斯坦重心的稳定性)。 对于任意 F_1，……，F_K，$F\prime_1$，……，$F\prime_K \in \mathcal{P}_p(\mathbb{R}^d)$ 以及 $\pi = (\pi_1, \ldots\ldots, \pi_K) \in \Delta^K$，$\pi' = (\pi'_1, \ldots\ldots, \pi'_K) \in \Delta^K$，

(a) 设 $\bar{F} = SWB_p(F_1, \ldots, F_K, \pi)$ 和 $\bar{F}' = SWB_p(F\prime_1, \ldots, F\prime_K, \pi)$。则

$$SW_p^p(\bar{F}, \bar{F}') \leq \sum_{k=1}^{K} \pi_k SW_p^p(F_k, F'_k) \leq \max_{k \in \{1,\ldots,K\}} SW_p^p(F_k, F'_k). \tag{9}$$

(b) 设 $\bar{F} = SWB_p(F_1, \ldots, F_K, \pi)$ 和 $\bar{F}' = SWB_p(F\prime_1, \ldots, F\prime_K, \pi')$。则

$$SW_p^p(\bar{F}, \bar{F}') \leq 2^{p-1} \left(\sum_{k=1}^{K} \pi_k SW_p^p(F_k, F'_k) + M \|\pi - \pi'\|_p^p \right), \tag{10}$$

其中 M 是一个常数，它取决于 p 以及 $F\prime_1, \ldots, F'_K$ 的矩尺度 .

引理 1 的证明见补充材料 A.1。引理 1(a) 表明，当仅边缘分布发生扰动时，重心的变化幅度至多为最大的边缘分布变化量。引理 1(b) 指出，当边缘分布及其权重均发生扰动时，重心的变化幅度至多为最大边缘分布变化量与权重变化量共同作用的结果，该结果需乘以一个取决于边缘分布矩的常数。

借助 SWB 的稳定性，我们可以讨论该模型的一个实际方面。在实践中，我们通常通过样本观测预测变量和响应变量，并且必须对（广义）似然进行近似，见（6）。我们将证明，简单的代入估计量就足够了。

定理 1。 假设 F_1、……、F_K、$G \in P_p(\mathbb{R}^d)$ $(p \geq 1)$ 具有直径为 $R > 0$ 的紧支撑，且 F_1、……、F_K、G 是相应的经验分布，至少具有 n 个独立同分布的支撑点，$(\pi_1$、……、$\pi_K) \in \Delta^K$，并且 $\bar{F} = SWB_p(F_1$、……、F_K、π_1、……、$\pi_K)$ 以及 $\bar{F} =$

SWB^p（$\hat{}F_1$、……、$\hat{}F_K$、 π_1、……、 π_K）。以下不等式成立：

$$\mathbb{E}\left[\left|\exp\{-wSW_p^p(\bar{F},G)\}-\exp(-wSW_p^p\{\hat{\bar{F}},\hat{G})\}\right|\right]\leq C_{p,w,R}\frac{1}{\sqrt{n}}, \tag{11}$$

其中 $C_{p,\ w,\ R}>0$ 是一个常数，取决于 R、 p 和 w。

证明见补充材料 A.2 ，并用到了引理 1。定理1表明，广义似然的经验估计以 $\mathcal{O}(n^{-1/2})$ 的速率收敛，该速率为参数速率。请注意，该结果不要求总体分布 $F_1,\ldots,F_K,G$ 满足任何连续性假设；也就是说，它们可以是离散的，也可以是连续的。

接下来，我们讨论后验一致性。假设推前函数由参数 $\phi=(\phi_1,\ldots,\phi_K)\in\Phi:=\Phi_1\times\ldots,$ Φ_K 索引，形式为 $f_\phi=(f_{1,\ \phi_1},\ \ldots,\ f_{k,\ \phi_k})$，且研究者对 $p(\phi,\ \pi\mid\mathcal{S})$ 感兴趣。

将式（6）中的负对数似然视为损失函数，我们定义经验风险和总体风险

$$R_N(\phi,\pi)=\frac{1}{N}\sum_{i=1}^{n}\mathrm{SW}_p^p\{\tilde{G}_i(\phi,\pi),G_i\},\quad R(\phi,\pi)=\mathbb{E}_{(F_1,\ldots,F_K,G)\sim P}[\mathrm{SW}_p^p\{\tilde{G}(\phi,\pi),G\}]. \tag{12}$$

我们作出以下假设，以保证后验一致性。

令 P表示预测变量（F_1、……、 F_K）的真实生成模型，以及真实响应变量G。也就是说，P 是一个假想的真实过程，用于生成可能的实验。

假设1 (可识别性)。 存在$(\phi_0,\pi_0)\ \in\ \Phi\times\Delta^K$ ，使得 $R(\phi,\pi)$在(ϕ_0,π_0)处达到唯一最小值。此外，对于每个 $\epsilon>0$，$\Delta(\epsilon)$ =下确界$_{\{(\phi,\pi)\in\Phi\times\Delta^K:\ \|\phi-\phi_0\|_p+\|\pi-\pi_0\|_p\geq\epsilon\}}\left(R(\phi,\pi)-R(\phi_0,\pi_0)\right)>0.$

假设 2 (紧致性)。 参数空间 Φ是紧致的。

假设 3 (有界矩)。 对于所有 $(F_1、\ldots\ldots、 F_K、 G) \in \mathrm{supp}(P)$， $\mathbb{E}_{X\sim F_k}[\|X\|_p^p] \le C_k$、 $\mathbb{E}_{Y\sim G}[\|Y\|_p^p] \le C_G$，对于某些常数 C_k、 $C_G < \infty$ 以及 $k = 1、\ldots\ldots、 K$。

假设*4*（先验正性）。关于 p 的先验在 (ϕ_0, π_0) 的每个邻域上都赋予正质量（该邻域在假设*1*中定义）；也就是说，对任意 $\epsilon > 0$, $p(B_\epsilon(\phi_0, \pi_0)) > 0$，其中 $B_\epsilon(\phi_0, \pi_0) = \{(\phi, \pi) \in \Phi \times \Delta^K : \|\phi - \phi_0\|_p + \|\pi - \pi_0\|_p < \epsilon\}$。

假设*5*（回归函数的正则性）。对于任意 $k = 1, \ldots, K$，回归函数 f_{k,ϕ_k} 存在一个 ω_k : $\mathbb{R}_+ \to \mathbb{R}_+$ $(\lim_{t\to 0} \omega_k(t) = 0)$，作为连续模：对于所有 $(F_1, \ldots, F_K, G) \in supp(P)$ 以及 ϕ_k, $\phi_k' \in \Phi_k$, $\mathbb{E}_{X\sim F_k}\|f_{k,\phi_k}(X) - f_{k,\phi_k'}(X)\|_p^p \le \omega_k(\|\phi_k - \phi_k'\|_p^p)$。对于所有 $k = 1, \ldots, K$, $(F_1, \ldots, F_K, G) \in supp(P)$ 且 $\phi_k \in \Phi_k$, $\mathbb{E}_{X\sim F_k}[\|f_{k,\phi_k}(X)\|_p^p] \le C$，其中常数为 $C < \infty$。

定理 2。 U 在假设 *1–5*下，对于每个 $\epsilon > 0$，后验测度 p_N 满足

$$p_N\Big(\{(\phi, \pi) \in \Phi \times \Delta^K : \|\phi - \phi_0\|_p + \|\pi - \pi_0\|_p \ge \epsilon\}\Big) \xrightarrow{a.s} 0, \tag{13}$$

在 *i.i.d* 抽样下，*as* $N \to \infty$ 。

定理 2 的证明见补充材料A.3。特别地，引理 1对于为 $R_N(\phi, \pi)$建立一致大数定律至关重要，而这使我们能够在所述假设下证明一致性。

3.3 切片 Wasserstein 重心的计算

在介绍后验推断的细节之前，我们先讨论 SWB 的计算，这是评估（7）中似然所必需的。我们回顾一下定义具有边缘分布 $G_1、\ldots\ldots、 G_K \in \mathcal{P}_p(\mathbb{R}^d)$ 和 $\pi \in \Delta^K$的 SWB 的优化问题：

$$\min_{G\in\mathcal{P}_p(\mathbb{R}^d)} \sum_{k=1}^{K} \pi_k \mathrm{SW}_p^p(G, G_k) \tag{14}$$

求解 SWB 是一个优化问题，也仍是一个不断发展的研究方向（Bon- neel 等，2015，Nguyen、Nguyen 和 Ho，2025）。SWB 是分布空间上的一个凸问题，即映射 $G \to \sum_{k=1}^{K} \pi_k \mathrm{SW}_p^p(G,\ G_k)$ 是凸的（补充材料 1 中的命题 A.4）。然而在实践中，为了实现可处理优化，我们往往需要对重心进行参数化，这可能会破坏凸性。

在我们的情形中，我们对广义似然使用代入估计量，通过求解经验分布的 SWB $G_1 = \frac{1}{M_{G_1}} \sum_{i=1}^{M_{G_1}} \delta_{x_{i1}}$、……、$G_k = \frac{1}{M_{G_K}} \sum_{i=1}^{M_{G_K}} \delta_{x_{iK}}$。因此，自然也将重心限制为经验分布，即 $G = \frac{1}{M_G} \sum_{\ell=1}^{M_G} \delta_{z_\ell}$（也称为自由支撑重心（Cuturi 和 Doucet，2014），其中 z_1、……、$z_{M_G} \in \mathbb{R}^d$。我们使用以下梯度，通过迭代过程更新重心的支撑点：

$$\begin{aligned}
\nabla_{z_\ell} \sum_{k=1}^{K} \pi_k \mathrm{SW}_p^p(G, G_k) &= \sum_{k=1}^{K} \pi_k \nabla_{z_\ell} \mathrm{SW}_p^p(G, G_k) \\
&= \sum_{k=1}^{K} \pi_k \nabla_{z_\ell} \mathbb{E}_{\theta \sim \mathcal{U}(\mathbb{S}^{d-1})}[W_p^p(\theta \sharp G, \theta \sharp G_k)] \\
&= \sum_{k=1}^{K} \pi_k \mathbb{E}_{\theta \sim \mathcal{U}(\mathbb{S}^{d-1})}[\nabla_{z_\ell} W_p^p(\theta \sharp G, \theta \sharp G_k)], \qquad (15)
\end{aligned}$$

对于任意 $z_\ell \in \{z_1$、……、$z_{M_G}\}$。令 $\gamma_{k,\ \theta}^{\star}$ 为 $\theta \sharp G$ 与 $\theta \sharp G_k$ 之间的最优传输计划，这意味着

$$\begin{aligned}
\nabla_{z_\ell} W_p^p(\theta \sharp G, \theta \sharp G_k) &= \nabla_{z_\ell} \sum_{j=1}^{M_{G_k}} \gamma_{k,\theta,\ell j}^{\star} |\theta^\top (z_\ell - x_{jk})|^p \\
&= p\theta \sum_{j=1}^{M_{G_k}} \gamma_{k,\theta,\ell j}^{\star} \left|\theta^\top \left(z_\ell - x_{jk}\right)\right|^{p-2} \left(\theta^\top \left(z_\ell - x_{jk}\right)\right), \qquad (16)
\end{aligned}$$

当 $\theta^\top \left(z_\ell - x_{jk}\right) \neq 0$ 时。一般的更新形式如下：

$$z_\ell^{(t)} = g\left(z_\ell^{(t-1)}, \nabla_{z_\ell^{(t-1)}} \sum_{k=1}^{K} \pi_k \mathrm{SW}_p^p(G^{(t-1)}, G_k)\right), \qquad (17)$$

其中 $G^{(t)} = \frac{1}{M_G}\sum_{i=1}^{M_G}\delta_{z_\ell^{(t)}}$，$g$： $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$是更新规则，而 $z_1^{(0)}, \ldots, z_{M_G}^{(0)}$是随机初始化的。我们进行 $T > 0$ 次迭代，以获得 SWB 的近似值 $G^{(T)}$。总体而言，时间复杂度为 $\mathcal{O}(TM_{max}\log M_{max})$，其中$M_{max} = \max\{M_{G_1}, \ldots, M_{G_K}, M_G\}$。

在接下来的高效后验推断讨论中，我们将使用雅可比矩阵$\partial z_\ell^{(T)}(\phi)/\partial\phi$。具体而言，设 $f_\phi = (f_{1,\phi_1}, \ldots, f_{k,\phi_k})$为参数化回归函数，其中 $\phi = (\phi_1, \ldots, \phi_K)$，$G_1 = \frac{1}{M_{G_1}}\sum_{i=1}^{M_{G_1}}\delta_{f_{1,\phi_1}(x_{i1})}, \cdots, G_k = \frac{1}{M_{G_K}}\sum_{i=1}^{M_{G_k}}\delta_{f_{K,\phi_K}(x_{iK})}$。我们考虑雅可比矩阵$\partial z_\ell^{(T)}(\phi)/\partial\phi$，它是 ϕ的函数。设 $h_\ell^{(t-1)}(\phi) = \nabla_{z_\ell^{(t-1)}}\sum_{k=1}^{K}\pi_k \mathrm{SW}_p^p(G^{(t-1)}, G_k)$（见（15）–（16））。根据更新规则，我们有$z^{(T)}\ell(\phi) = g\left(z_\ell^{(T-1)}(\phi), h_\ell^{(T-1)}(\phi)\right)$。雅可比矩阵可写为：

$$\frac{\partial z_\ell^{(T)}(\phi)}{\partial\phi} = \frac{\partial z_\ell^{(T)}(\phi)}{\partial z_\ell^{(T-1)}(\phi)}\frac{\partial z_\ell^{(T-1)}(\phi)}{\partial\phi} + \frac{\partial z_\ell^{(T)}(\phi)}{\partial h_\ell^{(T-1)}(\phi)}\frac{\partial h_\ell^{(T-1)}(\phi)}{\partial\phi}, \tag{18}$$

其中$\partial z_\ell^{(T)}(\phi)/\partial z_\ell^{(T-1)}(\phi)$以及$\partial z_\ell^{(T)}(\phi)/\partial h_\ell^{(T-1)}(\phi)$取决于更新规则 g（针对特定 g 的确切形式将在后文给出），$\partial z_\ell^{(T-1)}(\phi)/\partial\phi$ 可以递归计算，并且

$$\begin{aligned}
\frac{\partial h_\ell^{(T-1)}(\phi)}{\partial\phi} &= \frac{\partial}{\partial\phi}\left(\sum_{k=1}^{K}\pi_k\mathbb{E}\left[p\theta\sum_{j=1}^{M_{G_k}}\gamma_{k,\theta,\ell j}^{\star(T-1)}\left|\theta^\top\left(z_\ell^{(T-1)}(\phi) - f_{k,\phi_k}(x_{jk})\right)\right|^{p-2}\right.\right.\\
&\qquad\left.\left.\left(\theta^\top\left(z_\ell^{(T-1)}(\phi) - f_{k,\phi_k}(x_{jk})\right)\right)\right]\right)\\
&= \sum_{k=1}^{K}\pi_k\mathbb{E}\left[p(p-1)\,\theta\sum_{j=1}^{M_{G_k}}\gamma_{k,\theta,\ell j}^{\star(T-1)}\left|\theta^\top\left(z_\ell^{(T-1)}(\phi) - f_{k,\phi_k}(x_{jk})\right)\right|^{p-2}\right.\\
&\qquad\left.\left(\theta^\top\left(\frac{\partial z_\ell^{(T-1)}(\phi)}{\partial\phi} - \frac{\partial f_{k,\phi_k}(x_{jk})}{\partial\phi}\right)\right)\right],
\end{aligned} \tag{19}$$

其中，$\gamma_{k\,\theta}^{\star(T-1)}$，是 $\theta\sharp G^{(T-1)}$ 与 $\theta\sharp G_k$之间的最优传输计划。在上述表达式中，期望是相对于 $\theta \sim \mathcal{U}(\mathbb{S}^{d-1})$而言的，$\gamma_\theta^{(T-1)}{}_{,\ell j}$是$\theta\sharp G^{(T-1)}$ 与 $\theta\sharp G_k$之间的最优传输计划，而$\partial f_{k,\phi_k}(x_{jk})/\partial\phi$取决于回归函数的参数化$f_{k,\phi_k}$ 后文将对此进行讨论。

类似地，我们还需要$\partial z_\ell^{(T)}(\pi)/\partial\pi$，其推导如下：

$$\frac{\partial z_\ell^{(T)}(\pi)}{\partial \pi} = \frac{\partial z_\ell^{(T)}(\pi)}{\partial z_\ell^{(T-1)}(\pi)} \frac{\partial z_\ell^{(T-1)}(\pi)}{\partial \pi} + \frac{\partial z_\ell^{(T)}(\pi)}{\partial h_\ell^{(T-1)}(\pi)} \frac{\partial h_\ell^{(T-1)}(\pi)}{\partial \pi}, \tag{20}$$

其中$\partial z_\ell^{(T)}(\pi)/\partial z_\ell^{(T-1)}(\pi)$以及$\partial z_\ell^{(T)}(\pi)/\partial h_\ell^{(T-1)}(\pi)$取决于更新规则$g$，$\partial z_\ell^{(T-1)}(\pi)/\partial\pi$可以递归计算，并且

$$\begin{aligned}\frac{\partial h_\ell^{(T-1)}(\pi)}{\partial \pi} = \frac{\partial}{\partial \pi}\Bigg(\sum_{k=1}^{K} \pi_k \mathbb{E}\Bigg[p\theta \sum_{j=1}^{M_{G_k}} \gamma_{k,\theta,\ell j}^{\star,(T-1)} \left|\theta^\top\left(z_\ell^{(T-1)}(\pi) - f_{k,\phi_k}(x_{jk})\right)\right|^{p-2} \\ \left(\theta^\top\left(z_\ell^{(T-1)}(\pi) - f_{k,\phi_k}(x_{jk})\right)\right)\Bigg]\Bigg) = (\mathcal{G}_1, \ldots, \mathcal{G}_K),\end{aligned} \tag{21}$$

with

$$\begin{aligned}\mathcal{G}_k = \mathbb{E}\left[p\theta \sum_{j=1}^{M_{G_k}} \gamma_{k,\theta,\ell j}^{\star,(T-1)} \left|\theta^\top\left(z_\ell^{(T-1)}(\pi) - f_{k,\phi_k}(x_{jk})\right)\right|^{p-2} \left(\theta^\top\left(z_\ell^{(T-1)}(\pi) - f_{k,\phi_k}(x_{jk})\right)\right)\right] \\ + \sum_{k=1}^{K} \pi_k \mathbb{E}\left[p(p-1) \sum_{j=1}^{M_{G_k}} \gamma_{k,\theta,\ell j}^{\star,(T-1)} \left|\theta^\top(z_\ell^{(T-1)}(\pi) - f_{k,\phi_k}(x_{jk}))\right|^{p-2} \theta\theta^\top\right] \frac{\partial z_\ell^{(T-1)}(\pi)}{\partial \pi_k}.\end{aligned} \tag{22}$$

在实践中，我们使用蒙特卡洛样本$\theta_1, \ldots, \theta_L \overset{i.i.d}{\sim} \mathcal{U}(\mathbb{S}^{d-1})$来近似梯度中的期望。对于更新规则$g$，我们使用 Adam（Kingma 和 Ba（2014年））， 其描述如下：

$$\begin{aligned}z_\ell^{(t)} &= g\left(z_\ell^{(t-1)}, h_\ell^{(t-1)}\right) = z_\ell^{(t-1)} - \eta \frac{\hat{m}_i^{(t)}}{\sqrt{\hat{v}_i^{(t)}} + \epsilon}, \\ \hat{m}_i^{(t)} &= \frac{m_i^{(t)}}{1-\beta_1^t}, \quad \hat{v}_i^{(t)} = \frac{v_i^{(t)}}{1-\beta_2^t}, \\ m_i^{(t)} &= \beta_1 m_i^{(t-1)} + (1-\beta_1) h_i^{(t-1)}, \quad v_i^{(t)} = \beta_2 v_i^{(t-1)} + (1-\beta_2)(h_i^{(t-1)})^2.\end{aligned} \tag{23}$$

其中$\eta > 0$是步长，$\beta_1 \in [0,1]$和$\beta_2 \in [0,1]$是动量参数，$\epsilon > 0$用于

避免分母为 0。对于此更新规则，我们有：

$$\frac{\partial z_\ell^{(t)}}{\partial z_\ell^{(t-1)}} = I, \quad \frac{\partial z_\ell^{(t)}}{\partial h_\ell^{(t-1)}} = -\eta \operatorname{diag}\left(\frac{\partial \hat{m}_i^{(t)}}{\partial h_\ell^{(t-1)}} \frac{1}{\sqrt{\hat{v}_i^{(t)}} + \epsilon} + \hat{m}_i^{(t)} \frac{\partial \left(\sqrt{\hat{v}_i^{(t)}} + \epsilon \right)^{-1}}{\partial h_\ell^{(t-1)}} \right), \tag{24}$$

其中

$$\frac{\partial \hat{m}_i^{(t)}}{\partial h_\ell^{(t-1)}} = \frac{1-\beta_1}{1-\beta_1^t} I, \quad \frac{\partial \left(\sqrt{\hat{v}_i^{(t)}} + \epsilon)^{-1} \right)}{\partial h_\ell^{(t-1)}} = -\frac{1}{\left(\sqrt{\hat{v}_i^{(t)}} + \epsilon \right)^2} \frac{1}{2\sqrt{\hat{v}_i^{(t)}}} \frac{\partial \hat{v}_i^{(t)}}{\partial h_\ell^{(t-1)}}, \tag{25}$$

$$\frac{\partial \hat{v}_i^{(t)}}{\partial h_\ell^{(t-1)}} = \frac{2(1-\beta_2)}{1-\beta_2^t} \operatorname{diag}(h_\ell^{(t-1)}). \tag{26}$$

我们注意到，更新规则仍有改进空间，例如可以采用黎曼 Silver 步长（Park 等人（2025年））。不过，我们会在保持良好实际性能的同时，尽可能简化更新规则。现在我们可以在下一节讨论后验推断。

3.4 后验推断

广义后验定义如下：

$$p(f, \pi \mid \mathcal{S}) \propto p(f, \pi) \prod_{i=1}^{N} \ell(f_1, \ldots, f_K, \pi_1, \ldots, \pi_K; G_i, F_{i1}, \ldots, F_{iK}). \tag{27}$$

我们使用带有提议分布的 Metropolis-Hastings (MH) 转移概率实现后验马尔可夫链蒙特卡罗 (MCMC) 模拟，以模拟 $p(f \mid \pi, \mathcal{S})$ 和 $p(\pi \mid f, \mathcal{S})$。具体而言，我们构造提议分布 $f^* \sim q(f^* \mid f)$ 和 $\pi^* \sim q(\pi^* \mid \pi)$，并以如下概率接受提议样本：

$$\min\left\{1, \frac{p(f^* \mid \pi, \mathcal{S}) q(f \mid f^*)}{p(f \mid \pi, \mathcal{S}) q(f^* \mid f)}\right\}, \quad \min\left\{1, \frac{p(\pi^* \mid f, \mathcal{S}) q(\pi \mid \pi^*)}{p(\pi \mid f, \mathcal{S}) q(\pi^* \mid \pi)}\right\}. \tag{28}$$

实践中，我们只能观测到来自分布的独立同分布样本。因此，我们需要近似广义似然，以计算接受-拒绝概率。利用经验分布，我们有：

$$\begin{aligned}&\ell(f_1,\ldots,f_K,\pi_1,\ldots,\pi_K;G_i,F_{i1},\ldots,F_{iK})\\&\approx \ell(f_1,\ldots,f_K,\pi_1,\ldots,\pi_K;\hat{G}_i,\hat{F}_{i1},\ldots,\hat{F}_{iK})=\exp\left\{-w\mathrm{SW}_p^p\left[\widehat{\widetilde{G}}_i(f,\pi),\hat{G}_i\right]\right\},\end{aligned}\tag{29}$$

其中 $\widetilde{G}_i(f,\pi)=\mathrm{SWB}^p(f_1\sharp\hat{F}_{i1},\ldots,f_K\sharp\hat{F}_{iK},\pi_1,\ldots,\pi_K)$，且 $\hat{G}_i,\hat{F}_{i1},\ldots,\hat{F}_{iK}$是 $G_i,F_{i1},\ldots,F_{iK}$的经验分布。如定理 1所述，这种近似具有良好的收敛性。接下来，我们按照前文（第3.3节）所述近似 SWB，即对 $T>0$，用 $\widehat{\widetilde{G}}_i^{(T)}(f,\pi)$ 近似 $\widehat{\widetilde{G}}_i(f,\pi)$。我们将近似的广义似然记为：

$$\hat{\ell}(f_1,\ldots,f_K,\pi_1,\ldots,\pi_K;G_i,F_{i1},\ldots,F_{iK},T)=\exp\left\{-w\mathrm{SW}_p^p\left[\widehat{\widetilde{G}}_i^{(T)}(f,\pi),\hat{G}_i\right]\right\}.\tag{30}$$

更新ϕ： 现在我们分别讨论 ϕ 和 π 的提议分布的构造。我们将回归函数 f_ϕ 参数化为 $f_\phi=(f_{1,\phi_1},\ldots,f_{K,\phi_K})$，其中 $\phi=(\phi_1,\ldots,\phi_K)$。对于固定的步长 $\eta_1>0$，我们使用 MALA 提议 $q(\phi^*\mid\phi)\propto\exp\left(-\frac{1}{4\eta_1}\|\phi^*-\phi-\eta_1\nabla_\phi\log\hat{p}(\phi\mid\pi,\mathcal{S})\|_2^2\right)$，并且

—

$$\begin{aligned}\nabla_\phi\log p(\phi\mid\pi,\mathcal{S})&\approx\nabla_\phi\log p(\phi)-w\sum_{i=1}^N\nabla_\phi\mathbb{E}\left\{W_p^p\left[\theta\sharp\widehat{\widetilde{G}}_i^{(T)}(\phi,\pi),\theta\sharp\hat{G}_i\right]\right\},\\&=\nabla_\phi\log p(\phi)-w\sum_{i=1}^N\mathbb{E}\left\{\nabla_\phi W_p^p\left[\theta\sharp\widehat{\widetilde{G}}_i^{(T)}(\phi,\pi),\theta\sharp\hat{G}_i\right]\right\}.\end{aligned}\tag{31}$$

令 $\psi_{i,\theta}^\star$ 为 $\theta\sharp\widehat{\widetilde{G}}_i^{(T)}(\phi,\pi)$ 与 $\theta\sharp\hat{G}_i$ 之间的最优传输计划，M_i 为

$\widetilde{G}_i^{(T)}(\phi,\ \pi)$ 的原子数，则有：

$$\nabla_\phi W_p^p(\theta\sharp\widehat{\widetilde{G}}_i^{(T)}(\phi,\pi),\theta\sharp\hat{G}_i)=\nabla_\phi\sum_{j=1}^{M_i}\sum_{j'=1}^{M_{G_i}}\psi^\star_{i,\theta,jj'}|\theta^\top(z_{i,j}^{(T)}(\phi,\pi)-y_{i,j'})|^p \tag{32}$$

$$=p\sum_{j=1}^{M_i}\sum_{j'=1}^{M_{G_i}}\psi^\star_{i,\theta,jj'}\left|\theta^\top(z_{i,j}^{(T)}(\phi,\pi)-y_{i,j'})\right|^{p-2}\left(\theta^\top(z_{i,j}^{(T)}(\phi,\pi)-y_{i,j'})\right)$$

$$\left(\frac{\partial z_{i,j}^{(T)}(\phi,\pi)}{\partial\phi}\right)^\top\theta, \tag{33}$$

其中 $\frac{\partial z_{i,j}^{(T)}(\phi,\pi)}{\partial\phi}$ 在第 3.3 节中讨论。我们对 $\phi^*\sim q(\phi^*\mid\phi)$ 进行采样，令 $\phi^*=\phi+\eta_1\nabla_\phi\log\,p(\phi\mid\pi,\mathcal{S})+\sqrt{2\eta_1}\epsilon_0$，其中 $\epsilon_0\sim\mathcal{N}(0,I)$是一个维数与参数 ϕ 相匹配的标准多元高斯分布。

更新 π： 为了为 π 构造基于梯度的转移概率，我们使用变量变换来消除 π 的单纯形约束：

$$\pi_k=\frac{e^{\varpi_k}}{1+\sum_{k'=1}^{K-1}e^{\varpi_{k'}}}\quad\text{for}\quad k=1,\dots,K-1,\quad\pi_K=\frac{1}{1+\sum_{k'=1}^{K-1}e^{\varpi_{k'}}}. \tag{34}$$

令 $D=\mathrm{diag}(\pi_1,\dots,\pi_{K-1})$。变换的雅可比矩阵为 $J=D-\pi\pi^\top$，其行列式为：$\det(J)=\det(D(1-\pi^\top D^{-1}\pi)=\prod_{k=1}^K\pi_k$，这意味着 $p(\varpi_k)=p(\pi)\prod_{k=1}^K\pi_k$。对于固定步长 $\eta_2>0$，我们使用提议分布 $q(\varpi^*\mid\varpi)\propto\exp\left(-\frac{1}{4\eta_2}\|\varpi^*-\varpi-\eta_2\nabla_\varpi\log\,p(\varpi\mid\phi,\mathcal{S})\|_2^2\right)$，并且

$$\nabla_\varpi\log p(\varpi\mid\phi,\mathcal{S})=\nabla_\phi\log p(\varpi)-w\sum_{i=1}^N\nabla_\varpi\mathbb{E}\left\{W_p^p\left[\theta\sharp\widehat{\widetilde{G}}_i^{(T)}(\phi,\varpi),\theta\sharp\hat{G}_i\right]\right\}, \tag{35}$$

可通过梯度推导得到$\frac{\partial z_{i,j}^{(T)}(\phi,\pi)}{\partial\pi}$ 正如第3.3节所述。由于 ϖ 和 π通过可逆确定性映射相互连接，因此对 ϖ 进行采样意味着对 π进行采样。

3.5 简约模型

我们指定一个简约模型，随后将其用于模拟和真实数据分析。首先，我们对回归函数采用线性参数化：

$$f_{k,\phi_k}(x) = f_{k,A_k,b_k}(x) = A_k x + b_k,\ \forall k = 1,\ldots,K, \tag{36}$$

其中 $A_k \in \mathbb{R}^{h\times d_k}$ 为系数矩阵，$b_k \in \mathbb{R}^h$ 为截距。如Nguyen, Ni & Mueller (2025年)所述，这类回归函数满足假设 5，即满足 Lipschitz 连续性。如第3.3 节和第3.4节所述，我们使用 $\partial f_{k,\ A_k,\ b_k(x)}\partial A_k := I_h \otimes x^\top$（应向量化，以匹配第3.3节中 ϕ 的向量形式）以及 $\partial f_{k,\ A_k,\ b_k(x)}\partial b_k = I_h$，用于 $k = 1,\ldots,K$ 的后验模拟。

我们通过为回归函数指定先验来补全模型：

$$A_k[ij] \sim Laplace(0,1),\ b_k[i] \sim \mathcal{N}(0,10^3),\ \forall i = 1,\ldots,d_k,\ \forall j = 1,\ldots,h, \tag{37}$$

以及重心权重：

$$(\pi_1,\ldots,\pi_K) \sim Dir(\alpha),\ \alpha = (0.01,\ldots,0.01), \tag{38}$$

这意味着以下先验适用于 ϖ：$p(\varpi) = {}_B{}^1{}_{(\alpha_0)} \prod_{k=1}^K \pi_k^{\alpha_k}$。如果需要或有必要，也可以使用其他可微函数族$f_{k,\ \phi_k}$ 例如深度神经网络（Wilson &Izmailov 2020）。我们采用线性形式，因为它具有可解释性、简约性，而且正如后文所示，足以满足我们的分析需求。对于相对误差（8），我们选择$f_k'(x) = b_k$，即截距模型。

模型	训练相对误差	训练相对误差95%最高可信区间	测试相对误差	测试相对误差95%最高可信区间
DDR	0.4097	0.4018-0.4168	0.4448	0.4370-0.4539
MDDR	**0.1331**	**0.078-0.1714**	**0.1330**	**0.094-0.1777**

表1：使用模拟数据比较单预测变量 DDR 模型与多预测变量 MDDR 模型的性能。报告了训练和测试相对误差（RE）及其 95%最高可信区间（HCI）。

4 模拟

我们通过模拟研究评估所提出的 MMDR。具体而言，我们着重说明使用 SWB 从低维预测变量生成多变量响应的优势（本例中，从具有一维结构的预测变量生成双变量响应）。

我们首先抽取 $A_1, A_2, A_3 \overset{i.i.d}{\sim} \mathcal{U}(SO(2))$，并设定 $(\pi_1, \pi_2, \pi_3) = (2/3, 1/6, 1/6)$。然后，我们通过以下步骤生成模拟真实值和假设数据：

1. 我们抽取 $v_1, v_2, v_3 \overset{i.i.d}{\sim} \mathcal{U}(\mathbb{S})$、$\mu_1, \mu_2, \mu_3 \overset{i.i.d}{\sim} \mathcal{N}(0, 9)$，以及 $\sigma_1^2, \sigma_2^2, \sigma_3^2 \overset{i.i.d}{\sim} IG(3, 1)$。

2. 我们定义预测变量 $F_1 = f_{v_1} \sharp \mathcal{N}(\mu_1, \sigma_1^2)$、$F_2 = f_{v_2} \sharp \mathcal{N}(\mu_2, \sigma_2^2)$、$F_3 = f_{v_3} \sharp \mathcal{N}(\mu_3, \sigma_3)$，其中 $f_v(x) = vx$；也就是说，预测变量是提升到 2 维的单变量高斯分布。$\hat{}$ $\hat{}$ $\hat{}$接着，我们将 F_1, F_2, F_3 分别定义为基于F_1, F_2, F_3的 100 个独立同分布样本构成的经验分布。

$\tilde{}$3. 我们定义 $G = \text{SWB}_2(f_{A_1} \sharp F_1, f_{A_2} \sharp F_2, f_{A_3} \sharp F_3, \pi_1, \pi_2, \pi_3)$，其中 $f_A(x) = Ax$，并定义$\tilde{}$响应变量 $G = G * \text{N}(0, 0.01)$。由于 G 难以求解，我们得到 G的经验版本：$\hat{}$ $G = \text{SWB}_2(f_{A_1} \sharp \hat{} F_1, f_{A_2} \sharp \hat{} F_2, f_{A_3} \sharp \hat{} F_3, \pi_1, \pi_2, \pi_3) * \text{N}(0, 0.01 I_2)$，其中按照第3.3节中的讨论，SWB 由 100 个原子近似表示。

$\hat{}$ $\hat{}$ $\hat{}$步骤 1–3 生成一个观测值（$F_{i,1}, \ldots, F_{i,K}, G_\ell$）。我们重复该过程 70 次，以获得一个数据集。我们选取 70% 的样本作为训练集 $\text{S} = \{(\hat{} F_{11}, \hat{} F_{12}, \hat{} F_{13}) \ldots, (\hat{} F_{N1}, \hat{} F_{N2}, \hat{} F_{N3})\}$（其中 $N = 49$），其余 30% 的样本作为测试集。我们将 MMDR 与 DDR（Nguyen, Ni & Mueller 2025）进行比较；

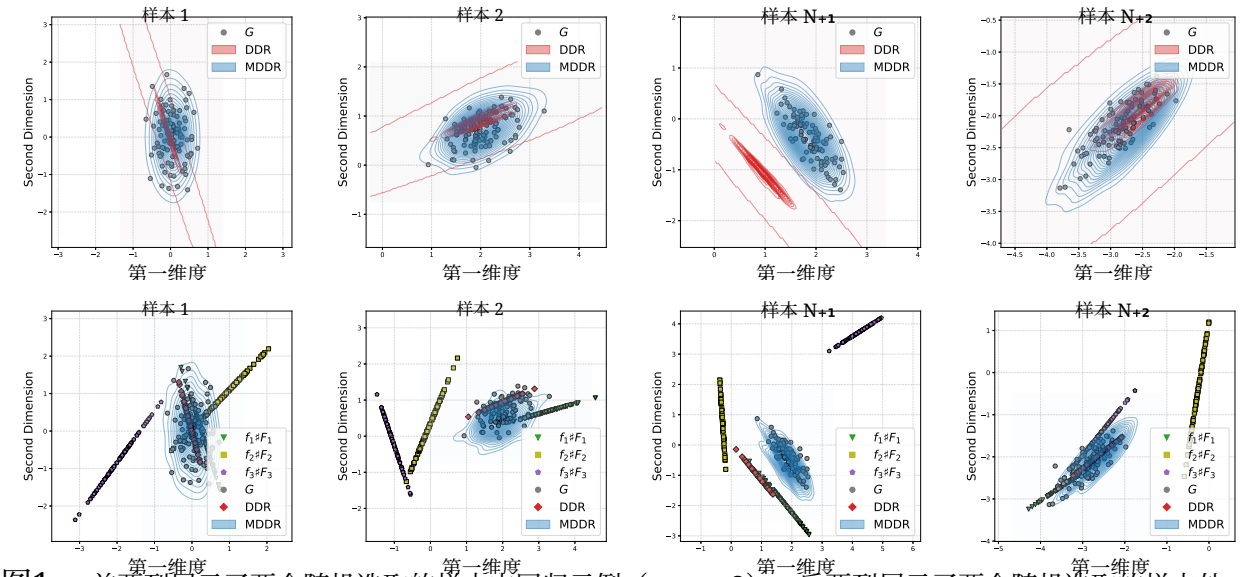

图1： 前两列展示了两个随机选取的样本内回归示例（$i = 1,\ 2$），后两列展示了两个随机选取的样本外示例（$i = N + 1,\ N + 2$）。第一行展示了观测响应变量 G_i，以及 DDR 和 MDDR 的拟合分布后验均值 $\mathbb{E}(G_i \mid data)$。

第二行展示了观测响应变量、链的最终后验样本下 DDR 和 MDDR 的拟合分布 G_i，以及基于该最终后验样本对预测变量进行的推前 $f_{\phi_k} \sharp F_\ell$。

对于后者，由于其不支持多个预测变量，我们使用线性回归函数对响应变量和第一个预测变量进行回归。对于这两个模型，我们使用 $w = 10$ 作为广义似然，并使用 MALA 获得 100 个马尔可夫链样本，MALA 将在第 3.4 节讨论，步长为 $\eta_1 = 0.001$ 和 $\eta_2 = 0.005$。我们使用 1000 次蒙特卡罗投影近似 SW距离。计算 SWB 时，我们使用 $T = 100,\ \ \eta = 0.1,$ 以及$(\beta_1,\ \ \beta_2) = (0.9,\ 0.999)$。

我们报告训练集的相对误差（8）、测试集的相对误差，以及相应的95%最高可信区间（最高可信区间），见表1。从表中可以看到，正如预期，MMDR 在样本内拟合和样本外预测方面的表现都显著优于 DDR。图 1展示了结果。从图的第一行可以看出，在样本内和样本外场景中，MMDR 得出的后验均值都更准确。我们展示了拟合分布的后验均值 $\mathbb{E}(G_i \mid F_i,\ \ data)$（为便于展示，实际呈现的是核密度估计（KDE））。MDDR 的表现

始终优于 DDR。图的第二行解释了 MMDR$\tilde{}$ $\hat{}$为何优于 DDR：它展示了来自马尔可夫链中最后一个后验样本的 G_i 和 $f_k\sharp F_{i,\ k}$。采用线性回归函数时，DDR 只能生成具有单变量结构的拟合分布。相比之下，尽管预测变量的推前分布也具有单变量结构，MMDR 仍能利用重心生成准确的拟合分布。

另一个值得关注的推断目标是重心权重，它们表示每个预测变量对响应变量的贡献。权重使我们能够量化预测变量与响应变量之间关联的强度。重心权重的后验均值为 $\bar{\pi}$ = (0.6482,0.1808,0.1710)，相应的 95% 后验可信区间为 (0.56, 0.1235, 0.1221)–(0.7677, 0.2629,0.2171)。后验分布在真实权重 $\pi = (2/3, 1/6, 1/6)$ 附近赋予了相当大的概率质量。在下一节中，我们将讨论如何利用 π 的后验分布，为单细胞数据构建细胞间通信网络。

5 细胞间通信

细胞间通信对于理解复杂的生物系统至关重要。细胞之间的相互作用是众多生理和病理过程的基础。一种自然的量化方法是对一个细胞类型的配体基因表达分布如何影响另一个细胞类型的受体基因表达分布进行建模。我们通过将所提出的 MDDR框架应用于群体规模单细胞数据集 OneK1K（Yazar等（2022）），展示该框架的实用性，推断一组作为发送者的细胞类型与一种作为接收者的细胞类型之间的通信。

我们关注四种主要细胞类型：B细胞、T细胞、单核细胞和NK（自然杀伤）细胞，并设置四项回归任务：(1) 预测变量：单核细胞、自然杀伤细胞和B细胞；响应变量：T细胞。(2) 预测变量：T细胞、单核细胞和自然杀伤细胞；响应变量：B细胞。(3)

预测变量	配体	响应变量	受体
单核细胞 自然杀伤细胞 B细胞	CD86、ICOSL、IL-23、TNF-α、IL-1β、白细胞介素-6、CD70 白细胞介素-15、IFN-γ、CD40配体，FasL CD40，CD86，ICOSL，IL-6，BAFF，增殖诱导配体	T细胞	CD3D、CD3E、CD3G、TRBC1、TRBC2、CD28、ICOS、白细胞介素-2受体、IFNGR、白细胞介素-21受体，PD-1，细胞毒性T淋巴细胞相关蛋白4，C-X-C趋化因子受体5，C-C趋化因子受体7、C-X-C趋化因子受体3、C-X-C趋化因子受体4、白细胞介素-12受体、白细胞介素-15R、IFN-γR、T细胞免疫球蛋白和黏蛋白结构域蛋白3
T细胞 单核细胞 自然杀伤细胞	CD40配体，IFN-γ CD40配体，白细胞介素-6，TNF-α，T细胞免疫球蛋白和黏蛋白结构域蛋白3，BAFF，增殖诱导配体 白细胞介素-15、IFN-γ、C-X-C基序趋化因子配体8，CD40配体	B细胞	免疫球蛋白重链恒定区μ、免疫球蛋白重链恒定区δ、免疫球蛋白重链恒定区γ、免疫球蛋白重链恒定区α、CD40，ICOSL，白细胞介素-21受体，白细胞介素-6受体，B细胞活化因子受体，C-X-C趋化因子受体5，CCR7，白细胞介素-10受体，C-X-C趋化因子受体4，PD-1，白细胞介素-2受体，白细胞介素-15受体，IFN-γR，C-C趋化因子受体1，C-X-C趋化因子受体2，ICOS，T细胞免疫球蛋白和黏蛋白结构域蛋白3
B细胞 T细胞 单核细胞	IL-6，BAFF，TNF-α，CD40配体 IFN-γ，CD86 CD40配体、TNF-α、白细胞介素-15，白细胞介素-18，T细胞免疫球蛋白和黏蛋白结构域蛋白3	自然杀伤细胞	白细胞介素-6受体，BAFF-R，TNFR，白细胞介素-10受体，CCR1，C-X-C趋化因子受体3，CD40，ICOS，PD-1，T细胞免疫球蛋白和黏蛋白结构域蛋白3，自然杀伤细胞受体NKG2D，白细胞介素-2受体，白细胞介素-4受体，白细胞介素-21R，IFN-γR，CD28，自然杀伤细胞受体NKp30，自然杀伤细胞受体NKp46，DNAM-1，自然杀伤细胞受体NKG2A，白细胞介素-15受体
自然杀伤细胞 B细胞 T细胞	白细胞介素-15、IFN-γ、Fas配体 CD40配体，CD70，白细胞介素-6，TNF-α，T细胞免疫球蛋白和黏蛋白结构域蛋白3，BAFF，增殖诱导配体 CD40配体、IFN-γ、CD70	单核细胞	白细胞介素-12受体、白细胞介素-15受体、IFN-γR、C-X-C趋化因子受体3，Fas受体、CD40、程序性死亡受体1、T细胞免疫球蛋白和黏蛋白结构域蛋白3、CD86，肿瘤坏死因子受体1，肿瘤坏死因子受体2，淋巴细胞活化基因3，白细胞介素-6受体，C-X-C趋化因子受体4

表2：4个回归问题中配体基因和受体基因的列表。(1) 预测变量：单核细胞、自然杀伤细胞、B细胞；响应变量：T细胞；(2) 预测变量：T细胞、单核细胞、自然杀伤细胞；响应变量：B细胞；(3) 预测变量：B细胞、T细胞、单核细胞；响应变量：自然杀伤细胞；(4) 预测变量：自然杀伤细胞、B细胞、T细胞；响应变量：单核细胞。

预测变量：B细胞、T细胞和单核细胞；响应变量：自然杀伤细胞。(4) 预测变量：自然杀伤细胞、B细胞和T细胞；响应变量：单核细胞。我们在表中汇总了所有 4 种关系涉及的配体基因和受体基因 2。从表中可以看出，配体的数量（预测变量的维数）小于受体的数量（响应变量的维数），这与第4节中的模拟设置类似。我们选取每种细胞类型至少有 90 个细胞的供体，共得到 75 名供体，并随机将其中 70% 分配至训练集，30% 分配至测试集。

我们再次比较所提出的 MDDR 与 DDR，两种模型均使用线性回归函数。由于 DDR 只能处理一个预测变量，因此我们针对每个响应变量分别设置三个 DDR 回归问题，每个问题使用不同的预测变量。在这两种

响应变量	模型	训练相对误差	训练相对误差相对误差95%最高密度区间	测试相对误差	测试相对误差 95%最高后验密度区间
T细胞	DDR（B细胞）	0.2515	0.2482-0.2574	0.2594	0.2533-0.2639
	DDR（单核细胞）	0.3552	0.3522-0.3587	0.3644	0.3565-0.3710
	DDR（NK细胞）	0.5085	0.5060-0.5113	0.5195	0.5164-0.5242
	MDDR	**0.1945**	**0.182-0.2174**	**0.2021**	**0.1879-0.2293**
B细胞	DDR（T细胞）	0.6131	0.6072-0.6175	0.6083	0.6011-0.6177
	DDR（单核细胞）	0.4515	0.4447-0.4576	0.4619	0.4552-0.4689
	DDR（NK细胞）	0.6210	0.6165-0.6256	0.6630	0.6512-0.669
	MDDR	**0.3685**	**0.3419-0.4076**	**0.379**	**0.3478-0.423**
单核细胞	DDR（T细胞）	0.5554	0.5422-0.5662	0.5275	0.5150-0.5389
	DDR（B细胞）	0.3844	0.3783-0.3905	0.4080	0.3974-0.4198
	DDR（NK细胞）	0.5629	0.5486-0.5716	0.5496	0.5380-0.5587
	MDDR	**0.3142**	**0.2749-0.3845**	**0.3162**	**0.2761-0.4023**
自然杀伤细胞	DDR（T细胞）	0.8000	0.7951-0.8066	0.7587	0.7518-0.7648
	DDR（B细胞）	0.4786	0.4734-0.4848	0.4959	0.4893-0.5021
	DDR（单核细胞）	0.4661	0.4625-0.4699	0.4537	0.4470-0.4615
	MDDR	**0.3604**	**0.3333-0.3817**	**0.3625**	**0.3401-0.3906**

表3：四种细胞类型中单预测变量 DDR 模型和多预测变量 MDDR 模型的平均残差误差（相对误差）。报告了训练和测试相对误差（相对误差）及其 95% 最高后验密度（HPD）区间。

方法中，我们设定 $w = 100$ 广义似然，并使用 MALA（见第 3.4 节）抽取 100 个后验样本，步长为 $\eta_1 = 0.001$ 和 $\eta_2 = 0.005$。推断过程中使用 100 个蒙特卡罗投影近似 SW距离，并在评估（相对误差计算）过程中使用 1000 个蒙特卡罗投影进行近似。对于 SWB 的计算，我们设定 $T = 100$，$\eta = 0.1$，以及 $(\beta_1, \beta_2) = (0.9, 0.999)$。

表 3总结了 MDDR 与单预测变量 DDR 模型在全部四种响应类型下的残差误差。所有情况下，MDDR 在训练集和测试集上的相对误差均最低，且 95% HCI范围显著更窄或与之相当，表明其估计稳定性和泛化能力均有所提升。这些结果与简单回归和多元（正态线性）回归的类似经验相呼应，并凸显了联合建模多个预测变量的优势。与任何基于单个预测变量的 DDR 模型相比，MDDR 能更有效地捕捉跨细胞类型关系，尤其是在响应变量的维数高于预测变量时。

我们展示了观测响应变量，以及拟合 DDR 和 MDDR 的后验均值

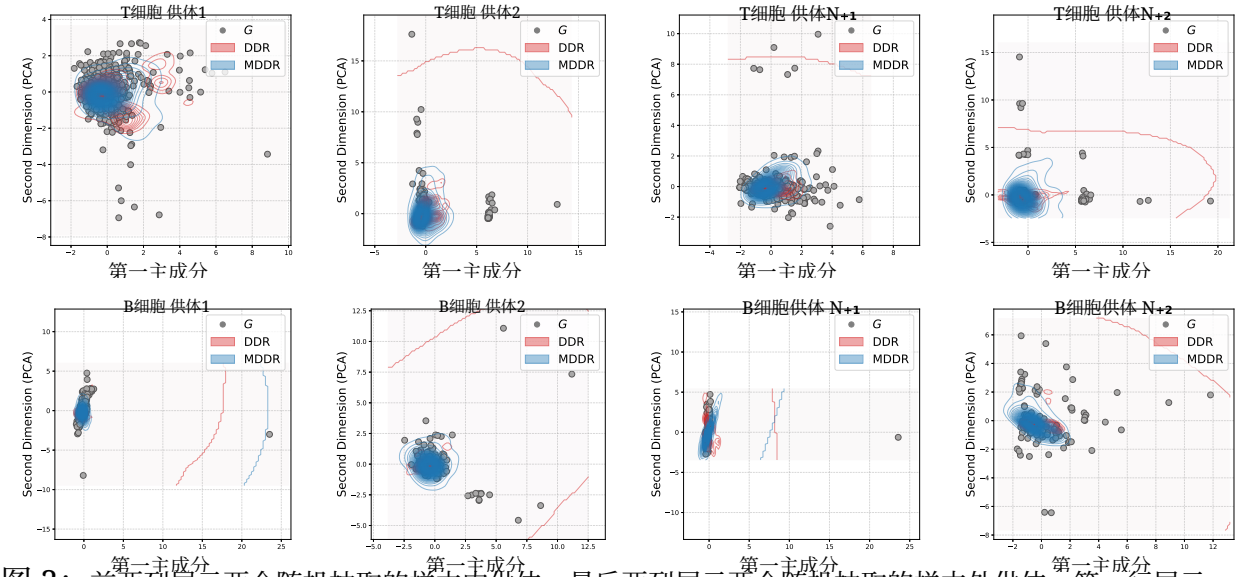

图 2：前两列展示两个随机抽取的样本内供体，最后两列展示两个随机抽取的样本外供体。第一行展示T细胞的结果，包括观测响应变量，以及 DDR 和 MDDR 拟合分布的核密度估计的后验均值。第二行展示B细胞的类似结果。我们使用主成分分析进行可视化。

T细胞和B细胞的分布（同样，为了展示而显示核密度估计）见图 2。对于训练供体和测试供体，MDDR 的拟合效果都明显优于 DDR。MDDR 的拟合分布能更准确地捕捉观测响应变量中的不确定性。虽然采用超越线性模型、更具表达能力的回归函数可以进一步改善拟合效果，但我们认为，在这种情形下，线性回归函数兼具简洁性与计算效率，已足够适用。补充材料 4 B图给出了单核细胞和自然杀伤细胞的类似结果，结论相同：MDDR 始终优于 DDR。总体而言，这些定性观察与表3中的定量汇总结果十分吻合。

最后，我们强调所提出的 MMDR 模型在推断方面的一项重要特征：它会自然地在预测变量与响应变量（此处为细胞类型）之间形成图结构。根据（6）和（7），MDDR 将拟合分布构造为预测变量的推前分布的重心。重心权重决定了每个预测变量对最终拟合分布的贡献程度。由于这些权重位于单纯形上，因此它们

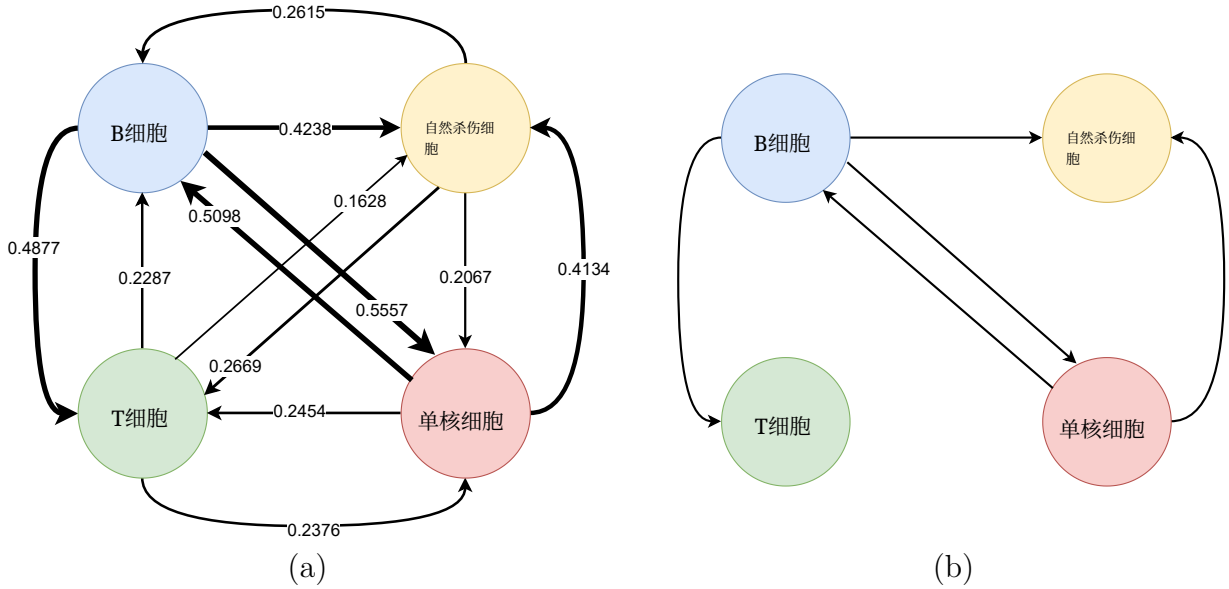

图3：(a) MMDR模型下采用重心权重后验均值（边的权重）构建的加权图。(b) 从加权图中仅保留权重大于 1/3 的边得到的稀疏图。

可以解释为预测变量与响应变量之间的关联程度（取值范围为 0 到 1）。针对对应四种细胞类型的四个回归问题，我们可以构建一个加权图，其顶点表示细胞类型。每个顶点都有三条来自其他顶点的入边，分别对应相关的 MDDR 问题。我们将每条边的权重设为相应重心权重的后验均值。所得加权图如图3（a）所示。最后，为了提高可解释性，我们去掉权重小于 1/3（狄利克雷先验的均值）的边，从而得到图 3（b）所示的稀疏图。从该图可以看出，有证据表明 B细胞与所有其他细胞类型均有相互作用，而单核细胞仅与 B细胞和自然杀伤细胞相互作用。相比之下，自然杀伤细胞和 T细胞均不与任何其他细胞类型相互作用。

6 结论

我们提出了贝叶斯多重多元密度-密度回归这一新颖框架，用于建模复杂的回归关系，其中多元随机分布

作为响应变量，多个随机分布则作为预测变量。

该框架仍存在若干局限。首先，由于依赖参数化回归函数 f_ϕ，框架可能对模型设定错误较为敏感。不过，在模拟和真实数据应用中，我们发现这些映射在实践中具有足够的灵活性。其次，我们在实现中使用的切片瓦瑟斯坦重心（SWB）近似，在计算效率或近似精度方面可能并非最优。不过，对于论文中考虑的数据集，目前的近似已足够有效。未来的工作将致力于解决这些局限。例如，可以将该框架扩展到多元密度—密度回归的贝叶斯非参数设定。此外，采用其他近似策略，例如收敛性更佳、效率更高的迭代方案，或对 SWB 进行确定性快速近似，也可以进一步提升性能。

参考文献

Bissiri, P. G., Holmes, C. C. & Walker, S. G.（2016年），《更新信念分布的一般框架》，英国皇家统计学会杂志*B*辑：统计方法**78**(5)，1103–1130。

Boedihardjo, M. T. (2025年), ‘最大切片瓦瑟斯坦距离的精确界’, 计算数学基础第1–32页。

Bonneel, N., Rabin, J., Peyré, G. & Pfister, H. (2015年), ‘测度的切片和 Radon Wasserstein 重心’, 数学成像与视觉杂志**1**(51), 22–45。

Chen, X., Fu, M., Huang, Y. & Deng, X. (2024年), ‘分布进分布出回归’,*arXiv*预印本 *arXiv:2405.11626*。

Cuturi, M. & Doucet, A. (2014年), Wasserstein 重心的快速计算，发表于‘机器学习国际会议’，PMLR，第685–693页。

Dunson, D. B., Pillai, N. & Park, J.-H. (2007年), ‘贝叶斯密度回归’, 英国皇家统计学会杂志*B*辑：统计方法**69**(2), 163–183。

Fournier, N. & Guillin, A.（2015年），《经验测度在Wasserstein距离中的收敛速率》，概率论及相关领域**162**(3)，707–738。

Girolami, M. & Calderhead, B.（2011年），《黎曼流形朗之万与哈密顿蒙特卡罗方法》，英国皇家统计学会杂志*B*辑：统计方法**73**(2)，123–214。

Goldfeld, Z., Kato, K., Rioux, G. & Sadhu, R.（2024年），《正则化最优传输的统计推断》，信息与推断：*IMA*期刊**13**(1)，iaad056。

Kingma, D. P. & Ba, J. (2014年),《Adam：一种随机优化方法》，*arXiv* 预印本 *arXiv:1412.6980*。

Manole,T.,Balakrishnan,S.& Wasserman,L.(2022),《切片 Wasserstein 距离的极小极大置信区间》，《电子统计学期刊》 16(1), 2252–2345。

Matabuena, M. & Petersen, A. (2023), '使用非参数调查回归模型对 NHANES 数据库中的加速度计数据进行分布数据分析', 英国皇家统计学会杂志 *C* 辑：应用统计 72(2), 294–313.

Nadjahi, K., Durmus, A., Chizat, L., Kolouri, S., Shahrampour, S. & Simsekli, U.(2020), '切片概率散度的统计与拓扑性质', 神经信息处理系统进展33, 20802–20812.

Nguyen, K. (2025), '切片最优传输导论：基础、进展、扩展与应用', 基础与趋势® 计算机图形学与视觉17(3-4), 171–406.

URL： *http://dx.doi.org/10.1561/0600000119*

Nguyen, K., Ho, N., Pham, T. & Bui, H. (2021), 分布切片 Wasserstein 及其在生成式建模中的应用, 载于 '国际学习表征会议'。

网址： *https://openreview.net/forum?id=QYjO70ACDK*

Nguyen, K., Nguyen, H. 和 Ho, N.（2025），《迈向边际公平性的切片瓦瑟斯坦重心》，发表于《第十三届国际学习表征会议》。**网址：** *https://openreview.net/forum?id=NQqJPPCesd*

Nguyen, K., Ni, Y. 和 Mueller, P.（2025），《贝叶斯多变量密度-密度回归》，*arXiv* 预印本 *arXiv:2504.12617*。

Nietert, S., Goldfeld, Z., Sadhu, R. & Kato, K.（2022），《切片 Wasserstein 距离的统计、鲁棒性与计算保证》，神经信息处理系统进展35，28179–28193。

Ning, N.（2025年），《Metropolis校正的次微分朗之万算法》，*arXiv*预印本 *arXiv:2507.06950* 。

Park, J., Roy, A., Siegel, J. W. & Bhattacharya, A.（2025年），《在黎曼流形上通过银色步长加速及其在Wasserstein空间中的应用》，神经信息处理系统进展。

Peyré, G. & Cuturi, M.（2019年），《计算最优传输：及其在数据科学中的应用》，机器学习基础与趋势 11(5-6)，355–607。

Peyré, G., Cuturi, M. 等（2019年），《计算最优传输：及其在数据科学中的应用》，机器学习基础与趋势 11(5-6)，355–607。

Rabin, J., Peyré, G., Delon, J. & Bernot, M.（2012年），Wasserstein 重心及其在纹理混合中的应用，载于 《计算机视觉中的尺度空间与变分方法：第三届国际会议，SSVM 2011，以色列 Ein-Gedi，2011年5月29日至6月2日，修订精选论文 3》，Springer，第435–446页。

Shen, W. & Ghosal, S.（2016年），‘面向高维数据的自适应贝叶斯密度回归’，*Bernoulli* 第396–420页。

Szabó, Z., Sriperumbudur, B. K., Póczos, B. & Gretton, A.（2016年），‘分布回归的学习理论’，机器学习研究杂志 17(152)，1–40。

Tokdar, S. T., Zhu, Y. M. & Ghosh, J. K.（2004），‘结合逻辑高斯过程和子空间投影的贝叶斯密度回归’，贝叶斯分析1(1)。

Villani, C.（2009），最优传输：旧与新，第338卷，Springer。

Wilson, A. G. & Izmailov, P.（2020），‘贝叶斯深度学习与泛化的概率视角’，神经信息处理系统进展 33，4697–4708。

Yazar, S., Alquicira-Hernandez, J., Wing, K., Senabouth, A., Gordon, M. G., Andersen,

S., Lu, Q., Rowson, A., Taylor, T. R., Clarke, L. 等（2022），《“单细胞 eQTL 定位揭示了自身免疫性疾病的细胞类型特异性遗传调控”》，《科学》**376**(6589), eabf3041。

Zhao, Y., Datta, A., Tang, B., Zipunnikov, V. & Caffo, B. S.（2023），‘密度对密度回归’，*arXiv* 预印本 *arXiv:2307.03642* 。

Zhu, C. & Müller, H.-G.（2023），《测地线最优传输回归》，*arXiv* 预印本 *arXiv:2312.15376* 。

《贝叶斯多重多元密度—密度回归》的补充材料

密度-密度回归》

A 证明

A.1 引理的证明1

我们首先陈述一个关于 SWB 投影的引理。特别是，我们证明 SWB 的投影是相应投影边缘分布的 Wasserstein 重心。

引理 2 （切片 Wasserstein 重心的投影）。 设 $F_1, \ldots, F_K \in \mathcal{P}_p(\mathbb{R}^d)$ 及其权重 $\bar{}$（$\pi_1, \ldots, \pi_K$） $\in \Delta$, $F = SWB_p$ （ $F_1, \ldots, F_K, \pi_1, \ldots, \pi_K$ ）。那么，对于几乎所有的 $\theta \in \mathbb{S}^{d-1}$, $\bar{F}$ 沿 θ 的投影 $\theta\sharp\bar{F}$

$$\theta\sharp\bar{F} = WB_p(\theta\sharp F_1, \ldots, \theta\sharp F_K, \pi_1, \ldots, \pi_K), \tag{39}$$

是投影测度 $\theta\sharp F_k$ 的一维 *Wasserstein* 重心。

证明。 定义 SW重心 的优化问题的目标函数可写为:

$$\mathcal{L}(\bar{F}) = \int_{\theta\in\mathbb{S}^{d-1}} L_\theta(\theta\sharp\bar{F})\,\mathrm{d}\sigma(\theta), \text{ where } \quad L_\theta(\theta\sharp\bar{F}) := \sum_{k=1}^{K} \pi_k W_p^p(\theta\sharp\bar{F}, \theta\sharp F_k), \tag{40}$$

且 $\sigma(\theta)$ 是 $\mathbb{S}^{d-1}$ 上的均匀测度。由于被积函数 $L(\theta\sharp\bar{F})$ 非负，当且仅当几乎处处对 θ,

$$\theta\sharp\bar{F} = \arg\min_{\nu\in\mathcal{P}_2(\mathbb{R})} L_\theta(\nu) = WB_p(\theta\sharp F_1, \ldots, \theta\sharp F_K, \pi_1, \ldots, \pi_K). \tag{41}$$

否则，可以在 θ 的任意正测度集合上，将投影替换为其一维 $^-$ 极小化解，从而减小积分，这与 F 的极小性矛盾。

□

(a) 根据 SW 的定义、引理 2 以及 $Q_{\theta\sharp\bar{F}} = \sum_{k=1}^{K} \pi_k Q_{\theta\sharp F_k}$ (Bonneelet al. 2015)，可得

$$
\begin{aligned}
\mathrm{SW}_p^p(\bar{F}, \bar{F}') &= \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}[W_p^p(\theta\sharp\bar{F}, \theta\sharp\bar{F}')] \\
&= \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[\int_0^1 |Q_{\theta\sharp\bar{F}}(t) - Q_{\theta\sharp\bar{F}'}(t)|^p \mathrm{d}t\right] \\
&= \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[\int_0^1 \left|\sum_{k=1}^{K} \pi_k Q_{\theta\sharp F_k}(t) - \sum_{k=1}^{K} \pi_k Q_{\theta\sharp F_k'}(t)\right|^p \mathrm{d}t\right] \\
&\leq \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[\int_0^1 \sum_{k=1}^{K} \pi_k |Q_{\theta\sharp F_k}(t) - Q_{\theta\sharp F_k'}(t)|^p \mathrm{d}t\right] \\
&= \sum_{k=1}^{K} \pi_k \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[\int_0^1 |Q_{\theta\sharp F_k}(t) - Q_{\theta\sharp F_k'}(t)|^p \mathrm{d}t\right] \\
&= \sum_{k=1}^{K} \pi_k \mathrm{SW}_p^p(F_k, F_k'), \qquad (42)
\end{aligned}
$$

其中，不等式由詹森不等式得出，下面的等式则由 Fubini 定理得出。

(b) 再次，根据 SW 的定义、引理 2 以及 $Q_{\theta\sharp\bar{F}} = \sum_{k=1}^{K} \pi_k Q_{\theta\sharp F_k}$ (Bon-

Neel 等人，2015年），我们有

$$
\begin{aligned}
\mathrm{SW}_p^p(\bar{F}, \bar{F}') &= \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}[W_p^p(\theta\sharp\bar{F}, \theta\sharp\bar{F}')] \\
&= \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[\int_0^1 |Q_{\theta\sharp\bar{F}}(t) - Q_{\theta\sharp\bar{F}'}(t)|^p \mathrm{d}t\right] \\
&= \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[\int_0^1 \left|\sum_{k=1}^K \pi_k Q_{\theta\sharp F_k}(t) - \sum_{k=1}^K \pi'_k Q_{\theta\sharp F'_k}(t)\right|^p \mathrm{d}t\right] \\
&\leq \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[\int_0^1 \left(\left|\sum_{k=1}^K \pi_k (Q_{\theta\sharp F_k}(t) - Q_{\theta\sharp F'_k}(t))\right|\right.\right. \\
&\qquad \left.\left. + \left|\sum_{k=1}^K (\pi'_k - \pi_k) Q_{\theta\sharp F'_k}(t)\right|\right)^p \mathrm{d}t\right] \\
&\leq 2^{p-1}\mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[\int_0^1 \left(\left|\sum_{k=1}^K \pi_k (Q_{\theta\sharp F_k}(t) - Q_{\theta\sharp F'_k}(t))\right|\right)^p \mathrm{d}t\right] \\
&\quad + \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[\int_0^1 \left(\left|\sum_{k=1}^K (\pi'_k - \pi_k) Q_{\theta\sharp F'_k}(t)\right|\right)^p \mathrm{d}t\right]. \qquad (43)
\end{aligned}
$$

根据引理1，我们知道

$$
\begin{aligned}
\mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[\int_0^1 \left(\left|\sum_{k=1}^K \pi_k (Q_{\theta\sharp F_k}(t) - Q_{\theta\sharp F'_k}(t))\right|\right)^p \mathrm{d}t\right] &\leq \sum_{k=1}^K \pi_k \mathrm{SW}_p^p(F_k, F'_k) \\
&\leq \max_{k\in\{1,\ldots,K\}} \mathrm{SW}_p^p(F_k, F'_k), \qquad (44)
\end{aligned}
$$

自 $\pi \in \Delta^K$ 以来。对于第二项，我们有

$$
\begin{aligned}
&\mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[\int_0^1 \left|\sum_{k=1}^K (\pi'_k - \pi_k)\, Q_{\theta\sharp F'_k}(t)\right|^p \mathrm{d}t\right] \\
&\leq \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[\int_0^1 \left(\sum_{k=1}^K |\pi'_k - \pi_k|\, |Q_{\theta\sharp F'_k}(t)|\right)^p \mathrm{d}t\right] \\
&\leq \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[\left(\sum_{k=1}^K |\pi'_k - \pi_k| \left(\int_0^1 |Q_{\theta\sharp F'_k}(t)|^p \mathrm{d}t\right)^{1/p}\right)^p\right] \\
&\leq \left(\sum_{k=1}^K |\pi'_k - \pi_k| \left(\mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})} \int_0^1 |Q_{\theta\sharp F'_k}(t)|^p \mathrm{d}t\right)^{1/p}\right)^p \\
&\leq \|\pi' - \pi\|_p^p\, M, \qquad (45)
\end{aligned}
$$

其中

$$M = \left(\sum_{k=1}^{K} \left(\mathbb{E}_{\theta \sim \mathcal{U}(\mathbb{S}^{d-1})} \int_0^1 |Q_{\theta \sharp F_k'}(t)|^p \, \mathrm{d}t \right)^{1/(p-1)} \right)^{(p-1)},$$

第二个不等式由闵可夫斯基不等式得出，第三个不等式由詹森不等式得出。由（43）、（44）和（45），我们得到所需的界，并完成证明。

A.2 定理1的证明

利用指数函数的中值定理，我们有

$$e^x - e^y = e^c(x - y), \tag{46}$$

对于任意 x， $y \in \mathbb{R}$ 以及某个 $c \in [x, y]$。因此，我们有

$$|e^x - e^y| \leq \max\{e^x, e^y\}|x - y|. \tag{47}$$

将上述不等式应用于我们的情形，可得：

$$\begin{aligned}
&\mathbb{E}\left[\left|\exp\{-w\mathrm{SW}_p^p(\bar{F}, G)\} - \exp\{-w\mathrm{SW}_p^p(\hat{\bar{F}}, \hat{G})\}\right|\right] \\
&\leq \mathbb{E}\left[\max\{\exp\{-w\mathrm{SW}_p^p(\bar{F}, G)\}, \exp\{-w\mathrm{SW}_p^p(\hat{\bar{F}}, \hat{G})\}\left|w(\mathrm{SW}_p^p(\bar{F}, G) - \mathrm{SW}_p^p(\hat{\bar{F}}, \hat{G}))\right|\right] \\
&\leq w\mathbb{E}\left[\left|\mathrm{SW}_p^p(\bar{F}, G) - \mathrm{SW}_p^p(\hat{\bar{F}}, \hat{G})\right|\right],
\end{aligned} \tag{48}$$

其中，最后一个不等式是基于以下事实：$\mathrm{SW}_p^p(\bar{F}, G) \geq 0$ 且 $\mathrm{SW}_p^p(\hat{\bar{F}}, \hat{G}) \geq 0$，这意味着 $\exp(-w\mathrm{SW}_p^p(\bar{F}, G)) \leq 1$ 且 $\exp(-w\mathrm{SW}_p^p(\hat{\bar{F}}, \hat{G})) \leq 1$。利用詹森不等式，

我们有：

$$
\begin{aligned}
& w\mathbb{E}\left[\left|\exp\{-w\mathrm{SW}_p^p(\bar{F}, G)\} - \exp\{-w\mathrm{SW}_p^p(\hat{\bar{F}}, \hat{G})\}\right|\right] \\
& \leq w\mathbb{E}\left[\left|\mathrm{SW}_p^p(\bar{F}, G) - \mathrm{SW}_p^p(\hat{\bar{F}}, \hat{G})\right|\right] \\
& = w\mathbb{E}\left[\left|\mathbb{E}_\theta\left[W_p^p(\theta\sharp\bar{F}, \theta\sharp G) - W_p^p(\theta\sharp\hat{\bar{F}}, \theta\sharp\hat{G})\right]\right|\right] \\
& \leq w\mathbb{E}\left[\mathbb{E}_\theta\left[\left|W_p^p(\theta\sharp\bar{F}, \theta\sharp G) - W_p^p(\theta\sharp\hat{\bar{F}}, \theta\sharp\hat{G})\right|\right]\right].
\end{aligned} \tag{49}
$$

利用 Goldfeld 等人（2024年）中关于直径为 $R > 0$ 的紧支撑引理 4，我们进一步得到：

$$
\begin{aligned}
& \mathbb{E}\left[\mathbb{E}_\theta\left[\left|W_p^p(\theta\sharp\bar{F}, \theta\sharp G) - W_p^p(\theta\sharp\hat{\bar{F}}, \theta\sharp\hat{G})\right|\right]\right] \\
& \leq C_{p,R}\mathbb{E}\left[\mathbb{E}_\theta\left[\left|W_1(\theta\sharp\bar{F}, \theta\sharp\hat{\bar{F}}) + W_1(\theta\sharp G, \theta\sharp\hat{G})\right|\right]\right] \\
& = C_{p,R}\mathbb{E}\left[\mathrm{SW}_1(\bar{F}, \hat{\bar{F}}) + \mathrm{SW}_1(G, \hat{G})\right]
\end{aligned} \tag{50}
$$

利用引理 1，我们进一步得到：

$$
\begin{aligned}
& C_{p,R}\mathbb{E}\left[\mathrm{SW}_1(\bar{F}, \hat{\bar{F}}) + \mathrm{SW}_1(G, \hat{G})\right] \\
& \leq C_{p,R}\mathbb{E}\left[\sum_{k=1}^K \pi_k \mathrm{SW}_1(F_k, \hat{F}_k) + \mathrm{SW}_1(G, \hat{G})\right] \\
& = C_{p,R}\sum_{k=1}^K \pi_k \mathbb{E}\left[\mathrm{SW}_1(F_k, \hat{F}_k)\right] + \mathbb{E}\left[\mathrm{SW}_1(G, \hat{G})\right]
\end{aligned} \tag{51}
$$

这转化为 SW 的样本复杂度$_1$。SW 的样本复杂度$_1$ 已在以下文献中得到研究：Nadjahi 等人(2020)、Nguyen 等人(2021)、Manole 等人(2022)、Nietert 等人(2022)、Boedihardjo (2025年）。我们基于紧支撑给出一个证明：

$$\begin{aligned}\mathbb{E}\left[\mathrm{SW}_1(G,\hat{G})\right] &= \mathbb{E}\left[\mathbb{E}_\theta\left[W_1(\theta\sharp G,\theta\sharp\hat{G})\right]\right]\\ &= \mathbb{E}\left[\mathbb{E}_\theta\left[\int_{\mathbb{R}}|F_{\theta\sharp G}(x)-F_{\theta\sharp\hat{G}}(x)|\mathrm{dx}\right]\right]\\ &\leq R\mathbb{E}\left[\mathbb{E}_\theta\left[\sup_x|F_{\theta\sharp G}(x)-F_{\theta\sharp\hat{G}}(x)|\right]\right]\\ &= R\mathbb{E}_\theta\left[\mathbb{E}\left[\sup_x|F_{\theta\sharp G}(x)-F_{\theta\sharp\hat{G}}(x)|\right]\right],\end{aligned}\tag{52}$$

根据 Fubini 的定理。根据 Dvoretzky–Kiefer–Wolfowitz 不等式，我们知道

$$\mathbb{P}(\sup_x|F_{\theta\sharp G}(x)-F_{\theta\sharp\hat{G}}(x)|>\epsilon)\leq 2\exp(-2n\epsilon^2),\tag{53}$$

对于任意 $\epsilon>0$ 和 $\theta\in\mathbb{S}^{d-1}$。因此,

$$\begin{aligned}\mathbb{E}\left[\sup_x|F_{\theta\sharp G}(x)-F_{\theta\sharp\hat{G}}(x)|\right] &= \int_0^\infty\mathbb{P}(\sup_x|F_{\theta\sharp G}(x)-F_{\theta\sharp\hat{G}}(x)|>\epsilon)\mathrm{d}\epsilon\\ &\leq \int_0^\infty 2\exp(-2n\epsilon^2)\mathrm{d}\epsilon\\ &= \sqrt{\frac{\pi}{2n}},\end{aligned}\tag{54}$$

由于高斯积分。根据 Fubini 定理,

$$\mathbb{E}_\theta\left[\mathbb{E}\left[\sup_x|F_{\theta\sharp G}(x)-F_{\theta\sharp\hat{G}}(x)|\right]\right]\leq\sqrt{\frac{\pi}{2n}}\tag{55}$$

由（48）、（50）、（51）、（52）和（55），我们得到

$$\mathbb{E}\left[\left|\exp(-w\mathrm{SW}_p^p(\bar{F},G))-\exp(-w\mathrm{SW}_p^p(\hat{\bar{F}},\hat{G}))\right|\right]\leq C_{p,R.w}\frac{1}{\sqrt{n}},\tag{56}$$

对于常数 $C_{p,\ R,\ w}$，其取决于 $p,\ w,\ R$，由此完成证明。

A.3 定理 2 的证明

我们首先为 $\left|\mathrm{SW}_p^p(\widetilde{G}(\phi,\ \pi)\ G)\ -\mathrm{SW}_p^p(\widetilde{G}(\phi',\ \pi')\ G)\right|$ 建立连续模，随后将其用于证明经验风险的一致大数定律。

引理 3. 在假设 3 和 5 下，我们有

$$\left|SW_p^p(\widetilde{G}(\phi,\pi),G)-SW_p^p(\widetilde{G}(\phi',\pi'),G)\right|\leq\gamma(\|\phi-\phi'\|_p^p+\|\pi-\pi'\|_p^p),\tag{57}$$

对于连续性模 $\gamma:\ \mathbb{R}_+\to\mathbb{R}_+$ 使得 $\lim_{t\to0}\gamma(t)\ =0$。

证明。 利用不等式 $|a^p-b^p|\leq p\mathrm{max}\{a^{p-1},b^{p-1}\}|a-b|$，我们有

$$\begin{aligned}&|\mathrm{SW}_p^p(\widetilde{G}(\phi,\pi),G)-\mathrm{SW}_p^p(\widetilde{G}(\phi',\pi'),G)|\\&\leq p\max\{\mathrm{SW}_p(\widetilde{G}(\phi,\pi),G)^{p-1},\mathrm{SW}_p(\widetilde{G}(\phi',\pi'),G)^{p-1}\}\\&\quad|\mathrm{SW}_p(\widetilde{G}(\phi,\pi),G)-\mathrm{SW}_p(\widetilde{G}(\phi',\pi'),G)|.\end{aligned}\tag{58}$$

我们首先给出 $\left|\mathrm{SW}_p(\widetilde{G}(\phi,\ \pi)\ G)\ -\mathrm{SW}_p\widetilde{G}(\phi',\ \pi')\ G)\right|$ 对于 $\phi,\ \phi'\in\Phi$ 的差值上界。根据 SW_p 的三角不等式，我们有：

$$\begin{aligned}&|\mathrm{SW}_p(\widetilde{G}(\phi,\pi),G)-\mathrm{SW}_p(\widetilde{G}(\phi',\pi'),G)|\\&\leq|\mathrm{SW}_p(\widetilde{G}(\phi,\pi),\widetilde{G}(\phi',\pi'))+\mathrm{SW}_p(\widetilde{G}(\phi',\pi'),G)-\mathrm{SW}_p(\widetilde{G}(\phi',\pi'),G)|\\&=\mathrm{SW}_p(\widetilde{G}(\phi,\pi),\widetilde{G}(\phi',\pi'))\\&\leq\left(2^{p-1}\max_{k\in\{1,\ldots,K\}}\mathrm{SW}_p^p(f_{k,\phi_k}\sharp F_k,f_{k,\phi_k'}\sharp F_k)+M\|\pi-\pi'\|_p^p\right)^{\frac{1}{p}},\end{aligned}\tag{59}$$

其中，最后一个不等式由引理 1 (b) 得出。我们进一步有：

$$
\begin{aligned}
&\mathrm{SW}_p^p(f_{k,\phi_k}\sharp F_k, f_{k,\phi_k'}\sharp F_k)\\
&= \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}[W_p^p(\theta\sharp f_{k,\phi_k}\sharp F_k, \theta\sharp f_{k,\phi_k'}\sharp F_k)]\\
&= \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[\inf_{\pi\in\Pi(F_k,F_k)}\mathbb{E}_{(X,Y)\sim\pi}[|\theta^\top f_{k,\phi_k}(X) - \theta^\top f_{k,\phi_k'}(Y)|^p]\right]\\
&\leq \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[\mathbb{E}_{(X,Y)\sim(Id,Id)\sharp F_k}[|\theta^\top f_{k,\phi_k}(X) - \theta^\top f_{k,\phi_k'}(Y)|^p]\right]\\
&= \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[\mathbb{E}_{X\sim F_k}[|\theta^\top f_{k,\phi_k}(X) - \theta^\top f_{k,\phi_k'}(X)|^p]\right]\\
&\leq \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[\mathbb{E}_{X\sim F_k}[\|f_{k,\phi_k}(X) - f_{k,\phi_k'}(X)\|_p^p]\right]\\
&= \mathbb{E}_{X\sim F_k}[\|f_{k,\phi_k}(X) - f_{k,\phi_k'}(X)\|_p^p \leq \omega_k(\|\phi_k - \phi_k'\|_p^p),
\end{aligned}
\tag{60}
$$

其中，第二个不等式由柯西–施瓦茨不等式以及 $\|\theta\|_2^2 = 1$得出，而最后一个不等式由假设 5得出。因此，我们有：

$$
\begin{aligned}
&|\mathrm{SW}_p(\widetilde{G}(\phi,\pi), G) - \mathrm{SW}_p(\widetilde{G}(\phi',\pi'), G)|\\
&\leq \left(2^{p-1}\max_{k\in\{1,\dots,K\}}\omega_k(\|\phi_k - \phi_k'\|_p^p) + M\|\pi - \pi'\|_p^p\right)^{\frac{1}{p}}\\
&\leq \left(2^{p-1}\max_{k\in\{1,\dots,K\}}\omega_k(\|\phi - \phi'\|_p^p) + M\|\pi - \pi'\|_p^p\right)^{\frac{1}{p}}\\
&\leq \left(2^{p-1}\max_{k\in\{1,\dots,K\}}\omega_k(\|\phi - \phi'\|_p^p + \|\pi - \pi'\|_p^p) + M(\|\phi - \phi'\|_p^p + \|\pi - \pi'\|_p^p)\right)^{\frac{1}{p}}
\end{aligned}
$$

其中 $\|\phi - \phi'\|_p^p = \sum_{k=1}^K \|\phi_k - \phi_k'\|_p^p$。

现在我们对 $\max\{\mathrm{SW}_p(\widetilde{G}(\phi,\pi),G)^{p-1},\mathrm{SW}_p(\widetilde{G}(\phi',\pi'),G)^{p-1}\}$ 进行界定。我们有

$$
\begin{aligned}
\mathrm{SW}_p^p(\widetilde{G}(\phi,\pi),G) &= \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[\int_0^1 |Q_{\theta\sharp\widetilde{G}(\phi,\pi)}(t)-Q_{\theta\sharp G}(t)|^p \mathrm{d}t\right] \\
&= \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[\int_0^1 \left|\sum_{k=1}^K \pi_k Q_{\theta\sharp f_{k,\phi_k}\sharp F_k}(t)-Q_{\theta\sharp G}(t)\right|^p \mathrm{d}t\right] \\
&= \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[\int_0^1 \left|\sum_{k=1}^K \pi_k Q_{\theta\sharp f_{k,\phi_k}\sharp F_k}(t)- Q_{\theta\sharp G}(t))\right|^p \mathrm{d}t\right] \\
&\leq \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[\int_0^1 \sum_{k=1}^K \pi_k \left|Q_{\theta\sharp f_{k,\phi_k}\sharp F_k}(t)- Q_{\theta\sharp G}(t))\right|^p \mathrm{d}t\right] \\
&= \sum_{k=1}^K \pi_k \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[\int_0^1 \left|Q_{\theta\sharp f_{k,\phi_k}\sharp F_k}(t)- Q_{\theta\sharp G}(t))\right|^p \mathrm{d}t\right] \\
&= \sum_{k=1}^K \pi_k \mathrm{SW}_p^p(f_{k,\phi_k}\sharp F_k, G) \\
&\leq \max_{k\in\{1,\ldots,K\}} \mathrm{SW}_p^p(f_{k,\phi_k}\sharp F_k, G) \\
&= \max_{k\in\{1,\ldots,K\}} \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}\left[\inf_{\pi\in\Pi(F_k,G)} \mathbb{E}_{(X,Y)\sim\pi}[\|\theta^\top f_{k,\phi_k}(X)-\theta^\top Y\|_p^p]\right] \\
&\leq \max_{k\in\{1,\ldots,K\}} \mathbb{E}_{(X,Y)\sim\pi}[\|f_{k,\phi_k}(X)\|_p^p + \|Y\|_p^p] \\
&\leq C + C_G. \qquad (61)
\end{aligned}
$$

通过类似的推理，$\mathrm{SW}_p(G(\phi',\ \pi'),G)$，我们得到

$$
\max\{\mathrm{SW}_p(\widetilde{G}(\phi,\pi),G)^{p-1},\mathrm{SW}_p(\widetilde{G}(\phi',\pi'),G)^{p-1}\} \leq (C+C_G)^{(p-1)/p}. \qquad (62)
$$

因此，我们得到

$$
\begin{aligned}
&|\mathrm{SW}_p^p(\widetilde{G}(\phi,\pi),G)-\mathrm{SW}_p^p(\widetilde{G}(\phi',\pi'),G)|\\
&\le p(C+C_G)^{(p-1)/p}\\
&\left(\max_{k\in\{1,\dots,K\}}\omega_k(\|\phi-\phi'\|_p^p+\|\pi-\pi'\|_p^p)+M(\|\phi-\phi'\|_p^p+\|\pi-\pi'\|_p^p)\right)^{\frac{1}{p}}\\
&=\gamma(\|\phi-\phi'\|_p^p+\|\pi-\pi'\|_p^p),
\end{aligned}
\tag{63}
$$

其中 $\gamma(t)=p(C+C_G)^{(p-1)/p}\left(\max_{k\in\{1,\dots,K\}}\omega_k(t)+Mt\right)^{\frac{1}{p}}$，且满足 $\lim_{t\to 0}\gamma(t)=0$。证明至此完成。

□

Lem命题 4（一致大数定律）。在假设 2、3和5下，

$$
\sup_{(\phi,\pi)\in\Phi\times\Delta^K}|R_N(\phi,\pi)-R(\phi,\pi)|\xrightarrow{a.s}0\quad as\ N\to\infty. \tag{64}
$$

证明。由假设 2 可知 Φ 是紧的，而单纯形 Δ^K 本身已经是紧的，因此存在一个 δ-网 $\{(\phi^{(1)},\pi^{(1)})\dots,(\phi^{(H)},\pi^{(H)})\}\subset\Phi\times\Delta^K$，使得对于任意 $(\phi,\pi)\in\Phi\times\Delta^K$，都存在 $(\phi^{(h)},\pi^{(h)})\in\{(\phi^{(1)},\pi^{(1)})\dots,(\phi^{(H)},\pi^{(H)})\}$，满足

$$
\|\phi-\phi^{(h)}\|_p^p+\|\pi-\pi^{(h)}\|_p^p\le\delta. \tag{65}
$$

根据引理 3，对于每个 $i=1,\dots\dots,N$：

$$
\left|\mathrm{SW}_p^p(\widetilde{G}_i(\phi,\pi),G_i)-\mathrm{SW}_p^p(\widetilde{G}_i(\phi^{(h)},\pi^{(h)}),G_i)\right|\le\gamma(\delta), \tag{66}
$$

其中 $\lim_{\delta\to 0}\gamma(\delta)=0$。对 i 取平均并求期望，可得

$$\frac{1}{N}\sum_{i=1}^{N}\left|\mathrm{SW}_p^p(\widetilde{G}_i(\phi,\pi),G_i)-\mathrm{SW}_p^p(\widetilde{G}_i(\phi^{(h)},\pi^{(h)}),G_i)\right|\le\gamma(\delta),\ \text{and} \tag{67}$$

$$\mathbb{E}\left[\left|\mathrm{SW}_p^p(\widetilde{G}(\phi,\pi),G)-\mathrm{SW}_p^p(\widetilde{G}(\phi^{(h)},\pi^{(h)}),G)\right|\right]\le\gamma(\delta) \tag{68}$$

根据詹森不等式，我们有：

$$\left|\frac{1}{N}\sum_{i=1}^{N}\mathrm{SW}_p^p(\widetilde{G}_i(\phi,\pi),G_i)-\frac{1}{N}\sum_{i=1}^{N}\mathrm{SW}_p^p(\widetilde{G}_i(\phi^{(h)},\pi^{(h)}),G_i)\right|\le\gamma(\delta),\ \text{and} \tag{69}$$

$$\left|\mathbb{E}\left[\mathrm{SW}_p^p(\widetilde{G}(\phi,\pi),G)-\mathrm{SW}_p^p(\widetilde{G}(\phi^{(h)},\pi^{(h)}),G)\right]\right|\le\gamma(\delta) \tag{70}$$

这等价于

$$|R_N(\phi,\pi)-R_N(\phi^{(h)},\pi^{(h)})|\le\gamma(\delta),\ \text{and}\quad |R(\phi,\pi)-R(\phi^{(h)},\pi^{(h)})|\le\gamma(\delta). \tag{71}$$

根据三角不等式，我们有：

$$\begin{aligned}
&\sup_{(\phi,\pi)\in\Phi\times\Delta^K}|R_N(\phi,\pi)-R(\phi,\pi)|\\
&\le\sup_{(\phi,\pi)\in\Phi\times\Delta^K}\left(|R_N(\phi,\pi)-R_N(\phi^{(h)},\pi^{(h)})|+|R_N(\phi^{(h)},\pi^{(h)})-R(\phi,\pi)|\right)\\
&\le\sup_{(\phi,\pi)\in\Phi\times\Delta^K}\Big(|R_N(\phi,\pi)-R_N(\phi^{(h)},\pi^{(h)})|+|R_N(\phi^{(h)},\pi^{(h)})-R(\phi^{(h)},\pi^{(h)})|\\
&\quad+|R(\phi^{(h)},\pi^{(h)})-R(\phi,\pi)|\Big)\\
&\le\max_{h=1,\dots,H}|R_N(\phi^{(h)},\pi^{(h)})-R(\phi^{(h)},\pi^{(h)})|+2\gamma(\delta).
\end{aligned} \tag{72}$$

令 $Z_N:=\sup_{(\phi,\pi)\in\Phi\times\Delta^K}|R_N(\phi,\pi)-R(\phi,\pi)|$ 且 $A_N^{\varepsilon}:=\{\,Z_N>\varepsilon\,\}\quad\varepsilon>0$. 几乎必然

收敛 $Z_N \xrightarrow{a.s} 0$ 等价于

$$\forall\, \varepsilon > 0: \quad \mathbb{P}\left(\limsup_{N\to\infty} A_N^{\varepsilon} \right) = 0, \tag{73}$$

其中上极限 $\limsup_{N\to\infty} A_N^{\varepsilon} := \bigcap_{m=1}^{\infty} \bigcup_{N\geq m} A_N^{\varepsilon}$。对于固定的 $\varepsilon > 0$，我们选择 $\delta > 0$ 足够小，使得$2\gamma(\delta) < \varepsilon/2$。那么 (72) 推出事件包含关系

$$A_N^{\varepsilon} \subseteq \bigcup_{h=1}^{H} \left\{ |R_N(\phi^{(h)}, \pi^{(h)}) - R(\phi^{(h)}, \pi^{(h)})| > \frac{\varepsilon}{2} \right\} =: \bigcup_{h=1}^{H} B_{N,h}^{\varepsilon/2}. \tag{74}$$

因此,

$$\limsup_{N\to\infty} A_N^{\varepsilon} \subseteq \bigcup_{h=1}^{H} \left(\limsup_{N\to\infty} B_{N,h}^{\varepsilon/2} \right). \tag{75}$$

对每个固定的 3，应用强大数定律（使用假设 $(\phi^{(h)},\ \pi^{(h)})$），

$$|R_N(\phi^{(h)}, \pi^{(h)}) - R(\phi^{(h)}, \pi^{(h)})| \xrightarrow{a.s} 0, \tag{76}$$

因此，对每个 h 和每个 $\eta > 0$,

$$\mathbb{P}\left(\limsup_{N\to\infty} \{\, |R_N(\phi^{(h)}, \pi^{(h)}) - R(\phi^{(h)}, \pi^{(h)})| > \eta \,\} \right) = 0. \tag{77}$$

取 $\eta = \varepsilon/2$ 并应用有限并集界，可得

$$\mathbb{P}\left(\limsup_{N\to\infty} A_N^{\varepsilon} \right) \leq \sum_{k=1}^{K} \mathbb{P}\left(\limsup_{N\to\infty} B_{N,k}^{\varepsilon/2} \right) = 0. \tag{78}$$

由于这对每个 $\varepsilon > 0$ 都成立，因此可得

$$\forall\, \varepsilon > 0: \quad \mathbb{P}\left(\limsup_{N\to\infty} A_N^{\varepsilon} \right) = 0, \tag{79}$$

这等价于 $Z_N \to 0$ 几乎必然成立，即，

$$\sup_{\phi\in\Phi} |R_N(\phi,\pi) - R(\phi,\pi)| \xrightarrow{a.s} 0, \tag{80}$$

这就完成了风险函数一致大数定律的证明。 □

有了已证明的一致大数定律，我们现在讨论后验一致性的证明。对于所有 $\epsilon > 0$，我们定义参数的“坏”集合：

$$S_\epsilon(\phi_0,\pi_0) := \{(\phi,\pi) \in \Phi \times \Delta^K : \|\phi - \phi_0\|_p + \|\pi - \pi_0\|_p \geq \epsilon\}, \tag{81}$$

其中 $(\phi_0,\ \pi_0)$ 定义见假设 1。我们的目标是证明

$$p_N(S_\epsilon(\phi_0,\pi_0)) \xrightarrow{a.s} 0, \tag{82}$$

对于所有 $\epsilon > 0$，p_N 是我们的后验测度。对于 $\varepsilon > 0$，令

$$A_N := \Big\{ p_N(S_\epsilon(\phi_0,\pi_0)) > \varepsilon \Big\}. \tag{83}$$

根据以事件的上极限来定义几乎必然收敛，只需证明

$$\mathbb{P}\Big(\limsup_{N\to\infty} A_N \Big) = 0, \tag{84}$$

其中 上极限$_{N\to\infty} A_N$：$= \bigcap_{m=1}^{\infty} \bigcup_{N\geq m} A_N$. 根据一致大数定律（引理 4），对于任意 $\eta > 0$ 定义事件

$$E_N := \Big\{ \sup_{(\phi,\pi)\in\Phi\times\Delta^K} |R_N(\phi,\pi) - R(\phi,\pi)| \leq \eta \Big\}. \tag{85}$$

那么 E_N 几乎必然最终发生。在 E_N 上：

- 对于（ϕ,π） $\in S_\epsilon$（ϕ_0,π_0），

$$R_N(\phi,\pi) \geq R(\phi,\pi) - \eta \geq R(\phi_0,\pi_0) + \Delta(\epsilon) - \eta, \tag{86}$$

  其中 $\Delta(\epsilon)$ $=$ 下确界$_{\{(\phi,\pi)\in\Phi\times\Delta^K:\ \|\phi-\phi_0\|_p+\|\pi-\pi_0\|_p\geq\epsilon\}}$ $\Big(R(\phi,\pi) - R(\phi_0,\pi_0)\Big) > 0$ (假设 1)。

- 对于 $(\phi,\pi) \in B_\delta(\phi_0,\pi_0) := \{(\phi,\pi) : \|\phi-\phi_0\|_p + \|\pi-\pi_0\|_p < \delta\}$ （“良好”集合），选择足够小的 δ，使得

$$\sup_{(\phi,\pi)\in B_\delta(\phi_0,\pi_0)} (R(\phi,\pi) - R(\phi_0,\pi_0)) \leq \eta, \tag{87}$$

  then

$$R_N(\phi,\pi) \leq R(\phi,\pi) + \eta \leq R(\phi_0,\pi_0) + 2\eta. \tag{88}$$

在 E_N 上，“不良”集合与“良好”集合的后验质量比

$$\begin{aligned}\frac{p_N(S_\epsilon(\phi_0,\pi_0))}{p_N(B_\delta(\phi_0,\pi_0))} &= \frac{\int_{S_\epsilon(\phi_0,\pi_0)} \exp(-wNR_N(\phi,\pi))\,\mathrm{d}p(\phi,\pi)}{\int_{B_\delta(\phi_0,\pi_0)} \exp(-wNR_N(\phi,\pi))\,\mathrm{d}p(\phi,\pi)} \\ &\le \frac{p(\Phi\times\Delta^K)\,\exp(-wN(R(\phi_0,\pi_0)+\Delta(\epsilon)-\eta))}{p(B_\delta(\phi_0,\pi_0))\,\exp(-wN(R(\phi_0,\pi_0)+2\eta))} \\ &= \frac{1}{p(B_\delta(\phi_0))}\exp(-wN(\Delta(\epsilon)-3\eta)),\end{aligned} \tag{89}$$

其中 $p(\phi,\pi)$ 是先验测度。因此，我们有

$$p_N(S_\epsilon(\phi_0,\pi_0)) \le C\exp(-wN(\Delta(\epsilon)-3\eta)), \tag{90}$$

其中 $C = {}_\pi \frac{1}{(B_\delta(\phi_0,\pi_0))} > 0$ 由假设4得出。选择 $\eta < \Delta(\epsilon)/4$，使得 $\Delta(\epsilon) - 3\eta > \Delta(\epsilon)/4 > 0$，那么右端将随 N 指数衰减，这意味着，对于几乎每一条样本路径 ω，都存在 $N_0(\omega)$，使得对所有 $N \ge N_0(\omega)$，

$$p_N(S_\epsilon(\phi_0,\pi_0)) < \varepsilon. \tag{91}$$

根据事件的上极限的定义，

$$\limsup_{N\to\infty} A_N = \bigcap_{m=1}^{\infty}\bigcup_{N\ge m}\{p_N(S_\epsilon(\phi_0,\pi_0)) > \varepsilon\} = \emptyset \quad \text{a.s.} \tag{92}$$

因此，

$$\mathbb{P}\Big(\limsup_{N\to\infty} A_N\Big) = 0. \tag{93}$$

由于 $\varepsilon > 0$ 的选取是任意的，我们得出结论

$$p_N(S_\epsilon(\phi_0, \pi_0)) \xrightarrow{\text{a.s.}} 0. \tag{94}$$

因此，对于任意 $\epsilon > 0$，后验都集中在 (ϕ_0, π_0) 附近，从而证明了贝叶斯后验一致性。

A.4 切片 Wasserstein 重心的凸性

我们将在下述命题中讨论分布空间中 SWB 的凸性。

命题 1 (切片 Wasserstein 重心的凸性). 给定 $F_1, \ldots, F_k, G \in \mathcal{P}_p(\mathbb{R}^d)$以及 $(\pi_1, \ldots, \pi_K) \in \Delta^K$，映射

$$G \to \sum_{k=1}^{K} \pi_k SW_p^p(G, F_k) \tag{95}$$

在 G 中是凸的，即

$$\sum_{k=1}^{K} \pi_k SW_p^p(G_t, F_k) \leq t \sum_{k=1}^{K} \pi_k SW_p^p(G_0, F_k) + (1-t) \sum_{k=1}^{K} \pi_k SW_p^p(G_1, F_k), \tag{96}$$

对于 $G_t = tG_0 + (1-t)G_1$，其中 $t \in [0,1]$，且对于任意一对测度 G_0、$G_1 \in \boldsymbol{P}_p(\mathbb{R}^d)$。

证明。我们首先证明 $W_p^p(G, F_k)$ 关于 G 是凸的。令 $G_t = tG_0 + (1-t)G_1$，对于 $t \in [0,1]$ 以及任意一对测度 (G_0, G_1)，我们将 γ_0 定义为 G_0 与 G_k 之间的最优耦合，将 γ_1 定义为 G_1 与 G_k 之间的最优耦合。令 $\gamma_t = t\gamma_0 + (1-t)\gamma_1$，我们

有 G_t 和 F_k，它们是 γ_t 的对应两个边缘分布。我们有：

$$\begin{aligned} W_p^p(G_t, F_k) &= \inf_{\gamma \in \Pi(G_t, G_k)} \mathbb{E}_{(x,y)\sim\gamma}[\|x-y\|_p^p] \\ &\leq \mathbb{E}_{(x,y)\sim\gamma_t}[\|x-y\|_p^p] \\ &= t\mathbb{E}_{(x,y)\sim\gamma_0}[\|x-y\|_p^p] + (1-t)\mathbb{E}_{(x,y)\sim\gamma_1}[\|x-y\|_p^p] \\ &= tW_p^p(G_0, F_k) + (1-t)W_p^p(G_1, F_k). \end{aligned} \tag{97}$$

接下来，我们证明 $\mathrm{SW}_p^p(G,\ F_k)$ 关于 G 也是凸的。利用已证明的 Wasserstein距离的凸性，我们有：

$$\begin{aligned} \mathrm{SW}_p^p(G_t, F_k) &= \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}[W_p^p(\theta\sharp G_t, \theta\sharp F_k)] \\ &\leq \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^{d-1})}[W_p^p(\theta\sharp tG_0, \theta\sharp F_k) + [W_p^p(\theta\sharp(1-t)G_1, \theta\sharp F_k)] \\ &= t\mathrm{SW}_p^p(F_0, G_k) + (1-t)\mathrm{SW}_p^p(F_1, G_k). \end{aligned} \tag{98}$$

由于 $\sum_{k=1}^K \pi_k \mathrm{SW}_p^p(G,\ F_k)$ 是 $\mathrm{SW}_p^p(G,\ F_k)$ 的正权重和，因此它关于 G 也是凸的。

□

B 附加结果

如正文所述，我们在图 4 中展示了自然杀伤细胞和单核细胞的观测响应，以及拟合的 DDR 和 MDDR 分布的核密度估计的后验均值。从这些图中可以看出，对于训练供体和测试供体，MDDR 的拟合效果再次优于 DDR。

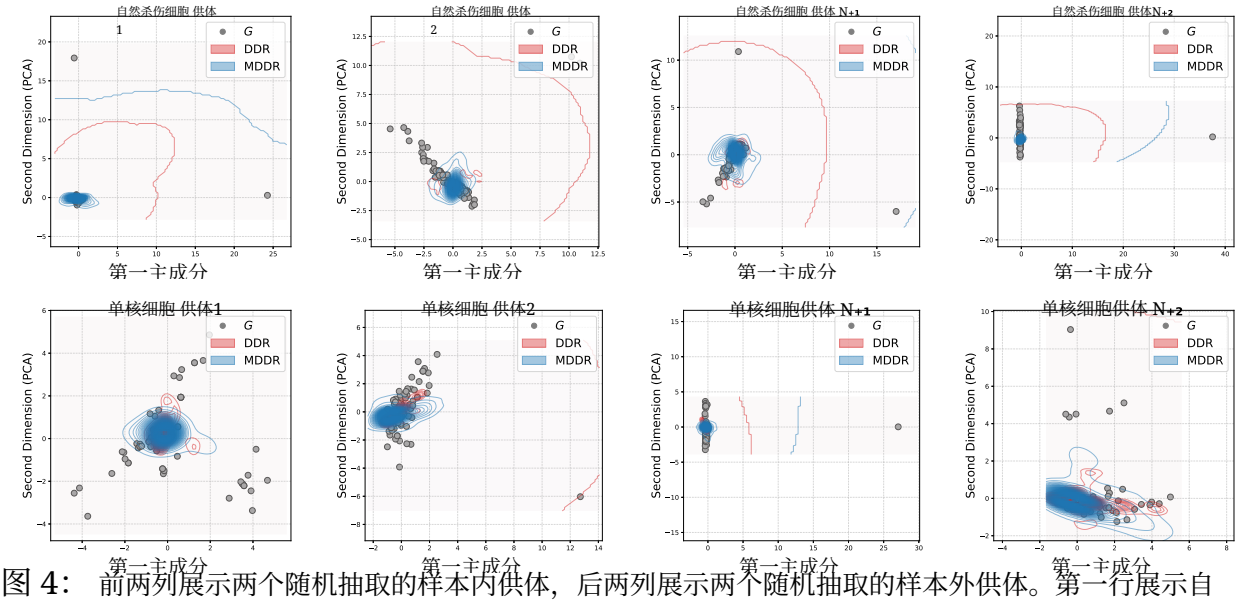

图 4: 前两列展示两个随机抽取的样本内供体，后两列展示两个随机抽取的样本外供体。第一行展示自然杀伤细胞的结果，包括观测响应变量，以及 DDR 和 MDDR 拟合分布的核密度估计的后验均值。第二行展示单核细胞的类似结果。