Mixe用于网络的潜在位置聚类模型

Chaoyi Lu[1,2],Riccardo Rastelli[1]，*[1]爱尔兰都柏林大学数学与统计学院[2]爱尔兰洞察数据分析研究中心，都柏林大学，爱尔兰

*riccardo.rastelli@ucd.ie

arXiv:2601.22380v1 [stat.ME] 29 Jan 2026

摘要

在过去二十年中，潜在位置模型（LPM）已成为获取基于模型的网络可视化结果的重要工具。然而，LPM 的几何结构本质上具有对称性，即假设传出边和传入边遵循相同的统计分布。因此，经典潜在位置模型框架并不适合分析有向网络。此外，边还可以通过加权来描述连接的持续时间或强度。这可能导致异配模式及其他无法由底层几何结构轻易捕捉的模体。为解决这些局限性，我们提出了 LPM 的一种新扩展，称为混合潜在位置聚类模型（MLPCM）。该模型能够处理非对称性和非欧几里得模式，同时为潜在空间提供新的解释。我们对网络中的有向边进行分解，从形式上将节点的行为方式与其他节点对它的感知方式区分开来。这形成了节点画像的双重表示，并识别出其“显性”和“隐性”社会位置。为高效估计模型参数，我们提出了一种变分贝叶斯方法，以近似后验分布。不同于许多现有的变分框架，我们的算法不需要任何额外的数值近似。我们通过引入一种新颖的部分积分完全似然准则来进行模型选择，该准则建立在惩罚似然方法相关文献的基础之上。我们使用合成数据集展示所提方法的准确性，并将其应用于国际武器转让数据集，以说明该方法的实际效用。

关键词：社交网络；网络分析；潜在空间模型；聚类；变分贝叶斯。

1 引言

潜在位置模型（LPM）是网络分析中一种重要的统计框架。自 Hoff 等（2002）的开创性研究以来，该模型已广泛应用于社会科学及其他众多研究领域。从方法论角度来看，研究者还提出了潜在位置模型的多种扩展和变体，包括 Sarkar 和 Moore（2005）、Sewell 和 Chen （ 2016 ）、 Kim 等（ 2018 ）、 Lu 等（ 2026 ）、 Jiang 等（ 2024 ）、 Rastelli 等（2024），以及 Rastelli 和 Corneli（2023）。关于潜在位置模型（LPM）的现有文献综述，可参阅 Kaur 等（2023）。

这种建模方法的一个关键特征是，将网络嵌入具有基本几何结构的潜在空间中，然后利用节点之间的距离来确定连接。

潜在空间的几何结构能够自然地支持同质性、聚类以及社交网络中常见的其他模体（Hoff 等（2002））。此外，这一潜在空间还可用于生成网络数据的清晰图形可视化，或推导基于模型的统计量，以概括网络复杂的拓扑。然而，潜在空间的几何属性也可能限制其对观测网络若干常见特征的刻画。事实上，行动者往往会根据其互动对象的不同，或更一般地说，根据互动所处的社会情境不同，呈现出不同的连接模式。这使得对边的刻画能够远远超越将节点的所有信息编码到单一潜在位置中的方式。

这方面的一个例子出现在包含有向边的网络中。由于潜在位置模型（LPM）是基于节点之间的成对距离构建的，因此生成的网络本质上是对称的。这意味着，对于有向网络而言，边的方向与相关统计分布的确定无关。然而，这一性质与观测到的网络拓扑结构相当不一致。节点发送边的模式与接收边的模式可能大不相同。例如，在电子邮件网络中，发送新闻简报的账户可能会发出大量电子邮件，却很少或完全不接收邮件。在存在核心-边缘结构时，可能难以找到一种潜在位置配置，使其与节点呈现的连接特征相符。

在这项工作中，我们着眼于潜在位置模型（LPM）的一种新表述，旨在利用潜在位置识别并区分节点可能呈现的多种社会特征。我们不再依赖单一潜在位置来完整描述节点的个体特征，而是为每个节点关联一个潜在位置分布。该分布用于拓宽节点可能呈现的社会特征范围，同时仍保持框架简约且易于解释。得益于这种新的模型结构，我们能够更灵活地刻画有向边，从而容纳非对称网络结构和异配模式。

我们的模型扩展与一些研究有向边情境下潜在位置模型（LPM）的文献相关，其中包括Sewell和Chen（2015）以及Sewell 和 Chen（2016）。与这两项研究类似，我们关注非负加权网络，因为与二元网络相比，这类网络中的边可以表现出程度更高的非对称性。不过，与这些研究的一个不同之处在于，我们通过潜在位置而非截距项来处理关系的非对称性。Krivitsky等人（2009）的研究也有类似之处，不过该研究是在节点上引入发送者与接收者随机效应项来捕捉非对称性。

本研究考虑的另一个方面是节点的聚类。我们所采用的表述紧密遵循 Handcock 等（2007）关于潜在位置聚类模型（LPCM）的开创性工作，其中对潜在位置使用有限高斯混合模型先验，以促进社区结构的出现。在这一聚类背景下产生的一个重要研究问题是最优组数的选择。为此，我们提出一种新的基于模型的准则，并将其命名为部分积分完全似然（PICL）。该准则受到惩罚似然方法文献的启发，其中包括 Akaike

信息准则（AIC）（Akaike 1974年）、Schwarz（1978年）的贝叶斯信息准则（BIC）、Biernacki 等人（2002）的积分完全似然（ICL），以及 Biernacki 等人（2010）和 Cˆome 与 Latouche（2015年）的精确 ICL（ICL_{ex}）。与 ICL_{ex} 类似，我们利用先验分布的共轭性，对完全似然关于模型参数进行边缘化。将模型参数积分掉的思想在文献中被广泛使用，既见于后验抽样，也见于优化情境。一些相关工作包括 Wyse 与 Friel（2012年）、Cˆome 与 Latouche（2015年）、Ryan 等人（2017年）、Legramanti 等人（2022年）以及 Lu 等人（2025年）。然而，在我们的情形中，与精确 ICL 框架不同，并非所有模型参数都能被积分掉，因此我们得到一种“部分坍缩”的模型选择准则，并用它来选择组数。

关于参数推断，潜在位置模型中最常见的参数推断方法，是从模型的后验分布中进行马尔可夫链蒙特卡罗（MCMC）抽样。本文推导了一种变分贝叶斯方法（Latouche 等人，2012年；Blei 等人，2017），此前已有多项研究在潜在位置模型的背景下验证了该方法，包括 Salter-Townshend 和 Murphy（2013）、Gollini 和 Murphy（2016）以及 Rastelli（2018）。与更常见的 MCMC 方法不同，变分贝叶斯框架旨在使用满足某些便于处理的独立性条件的另一种“变分”分布来近似后验分布。值得注意的是，在典型的变分贝叶斯方法中，通常需要进行各种数值近似，才能以解析形式写出变分后验分布。这些额外近似可能包括例如 Jensen 不等式步骤，Jordan 等人（1999）、Salter-Townshend 和 Murphy（2013）以及 Gollini 和 Murphy（2016）均采用了这一步骤。得益于模型假设，我们的框架能够绕过这一问题，无需进行任何进一步的数值近似，变分贝叶斯方法本身所隐含的近似除外。我们针对 MLPCM 实现的变分贝叶斯算法已在 GitHub代码库中公开提供*。

本文结构如下。第 2 节定义混合潜在位置聚类模型，并介绍相应的模拟算法。第 3 节和第 4 节详细介绍该模型的推断所用的变分贝叶斯程序，并概述我们定义的所有更新步骤及其相应方程。第 5 节介绍模型选择准则，并说明推断程序的一种初始设置方法。第 6 节包含两项模拟研究，分别针对由我们自己的模型生成的数据以及模型设定错误的情形，展示所提出的变分贝叶斯算法的性能。作为一种替代且具有竞争性的方法，我们采用 Lu 等（2026）的框架，并进行简要的敏感性分析，以验证结果在不同初始设置和先验分布下可能发生的变化。最后，我们在第 7 节将该方法应用于国际武器转让网络，并在第 8 节总结全文。

*https://github.com/Chaoyi-Lu/Mixed-Latent-Position-Cluster-Model.

2 模型

本文关注边具有非负离散权重的有向网络。我们用 $\boldsymbol{Y}:=\{y_{ij}\in\mathbb{N}:\ i,\ j=1,\ .....,\ N$ 表示观测到的交互矩阵，其中每个 y_{ij} 均为非负整数。不失一般性，我们可以假设该值表示从 i 到 j 的交互强度，强度为零意味着不存在边或交互。不允许出现自环，因此 $y_{ii}=0,\forall i=1,\ .....,\ N$。由于网络是有向的，$y_{ij}$ 对应从节点 i 到节点 j 的交互，而 y_{ji} 表示从节点 j 到节点 i 的交互，其取值可能不同于 y_{ij}。在这一框架下，一些常见的数据集包括记录电话呼叫次数、发送电子邮件数量或包裹投递数量的网络。

在真实有向加权网络中，我们通常会观察到y_{ij} 与 y_{ji}之间存在显著差异，例如，其中一个可能非常小甚至为零，而另一个却很大。这在呈现异配混合或核心—边缘结构的网络中并不少见，因为当出现星型结构时，这种情况就很典型；所谓星型结构，是指核心节点与许多孤立节点相连。在这些情景下，遵循基于距离的潜在位置模型原则的模型难以很好地拟合数据。事实上，潜在位置模型的几何构造偏向于由同质性（Hoff 等（2002））或聚类（Handcock 等，2007）产生的同配混合结构。更重要的是，由于经典潜在位置模型以节点之间的距离（这是一个对称算子）为基础，因此无法分别刻画 y_{ij} 与 y_{ji}。为了解决这一问题，我们将 LPCM 创新地扩展到非负离散加权有向网络，如下所示:

$$
\begin{aligned}
&y_{ij}\sim\mathrm{Pois}(\lambda_{ij}),\\
&\log(\lambda_{ij})=\beta-||\boldsymbol{u}_i-\boldsymbol{v}_{j\leftarrow i}||^2,\\
&\boldsymbol{u}_i|z_i=k\sim\mathrm{MVN}_d(\boldsymbol{\mu}_k,1/\tau_k\mathbb{I}_d),\\
&\boldsymbol{v}_{j\leftarrow i}\sim\mathrm{MVN}_d(\boldsymbol{u}_j,1/\gamma_j\mathbb{I}_d),\\
&z_i\sim\mathrm{categorical}(\pi_1,\ldots,\pi_K).
\end{aligned}
\tag{1}
$$

对于 $i,\ j=1,\ .....,\ N;\ \ i\neq j$。现在，我们从期望（1）的最末一行开始，向上说明其中的符号。每个节点都由一个聚类变量 z_i 表征，该变量源自混合比例 $\pi_1,\ .....,\ \pi_K$。这个聚类变量决定了潜在位置 $\boldsymbol{u}_i\in\mathbb{R}^d$ 的高斯有限混合先验分配，其中 $\boldsymbol{\mu}_k\in\mathbb{R}^d$ 和 $\tau_k>0$ 分别是与成分 k 相关的均值和精度。我们将 $\boldsymbol{u}_i$ 称为 i 的“隐性”位置:它是一种节点特征，决定了 i 的行为，因为它代表其真实、隐蔽的社会特征。这个隐性位置直接影响 i 如何发送边，而其他节点并不知道这一关于个体的信息。基于隐性位置 $\boldsymbol{u}_1,\ .....,\ \boldsymbol{u}_N$，我们为所有 i 和 j 创建若干“显性”位置 $\boldsymbol{v}_{j\leftarrow i}\in\mathbb{R}^d$。这些点散布在隐性位置附近，代表其他节点如何看待这些节点。例如，$\boldsymbol{v}_{j\leftarrow i}$ 表示 i 如何看待 j，这可能不同于 $\boldsymbol{v}_{j\leftarrow k}$，后者表示 k 如何看待同一个节点 j。然而，位置 $\boldsymbol{v}_{j\leftarrow i}$ 和 $\boldsymbol{v}_{j\leftarrow k}$ 预计都位于 $\boldsymbol{u}_j$ 附近，这意味着节点 j 真实、隐蔽的社会特征通常与 j 实际上对其的看法相似。数值 $\gamma_j>0$ 是用于调节

显性位置围绕相应隐性位置的离散程度。这个参数可以解释为节点行为真实性的度量，因为它描述了节点可能被看待的方式范围。具有较高 γ 的节点通常会被所有其他节点以非常相似的方式看待，其发送和接收边的模式通常也会一致。在这种情况下，该节点的行为可以被视为非常真实，其边也往往是对称的。作为一种特殊情况，如果每个节点都满足 $\gamma \to \infty$ ，那么我们就得到基本的 LPCM。相比之下，具有较低 γ 的节点会表现出较高的社会灵活性，即其他节点可能以截然不同的方式看待它，从而使原本极不可能出现的连接成为可能。

我们随后假设，边权重由截距 $\beta \in \mathbb{R}$以及发送方隐性位置与接收方显性位置之间的欧氏距离平方决定。换言之，交互强度由i真实且隐藏的社会特征，以及 i 如何感知 j 的社会特征所决定。相反，反向交互 y_{ji} 由个体 j的隐性潜在位置 $\boldsymbol{u}_j$ ，以及个体 j所感知的个体 i 的显性潜在位置 $\boldsymbol{v}_{i \leftarrow j}$ 决定。这使得我们可以分别刻画 y_{ij} 与 y_{ji}的两种不同分布，同时保持这两个变量之间的边际相依性，该相依性由每个节点的隐性位置与显性位置之间的关系所蕴含。

潜在空间为 d维，其中 $d \in \{1, 2, \dots\}$。尽管我们的记号保持一般性，并考虑任意维数的潜在空间，但在本文中我们仅关注情形 $d = 2$。我们注意到，对潜在维度数量的推断是一个关键且困难的研究问题，近期文献已对此进行了研究。由于这一研究方向并非我们工作的主要重点，为简便起见并获得更直观的可视化，我们选择 $d = 2$ 。但我们强调，我们的方法可以很容易地扩展到两个以上的维度，未来的工作可以针对给定数据集，研究为该模型选择理想维数的问题。

我们用一个 $d \times N$ 矩阵 $\boldsymbol{U} :=$($\boldsymbol{u}_1$、.....、 $\boldsymbol{u}_N$)表示所有隐性潜在位置的集合，其中每个 $\boldsymbol{u}_i$都是 $i = 1$、2、.....、N对应的 $d \times 1$向量。同样，所有显性潜在位置的集合表示为 $\boldsymbol{V} := \{\boldsymbol{v}_{j \leftarrow i} :\ i、\ j = 1$、2、.....、$N;\ i \neq j\}$，可将其理解为一个 $N \times N$ 矩阵，其中第 ij个条目为 $d \times 1$ 向量 $\boldsymbol{v}_{j \leftarrow i}$，也就是说， $\boldsymbol{V}$是一个 $d \times N \times N$数组。 $N \times 1$向量 $\boldsymbol{\gamma} :=$ (γ_1、.....、 γ_N)^{T}用于表示所有显性潜在位置的精度。沿用 Gollini 和 Murphy （ 2016 年）的方法，我们以平方距离代替简单的欧几里得距离，以增强数据分布对潜在空间中距离变化的敏感性，从而使数据可视化更加清晰。此外，正如我们在第 3 节所示，这一做法使推断算法的构建得以简化，无需在变分贝叶斯优化中进行额外的数值近似。

我们将由期望式 (1) 得到的新模型称为混合潜在位置聚类模型（MLPCM）。该模型的名称暗示了它与另一种著名网络方法——Airoldi等人（2008年）提出的混合成员随机块模型（MMSBM）之间的联系。与 MMSBM 一样，每个节点都由一组个体潜在特征刻画。从数据增广的角度来看，这些潜在

特征通过一个特定于边的潜在变量进行增广。在 MMSBM 中，该变量是用于生成特定边的簇成员身份。由于 MMSBM 会为每条边抽取新的簇成员身份，因此形成了“混合成员”框架。在我们的 MLPCM 模型中，这个增广变量是显性潜在位置 $\boldsymbol{v}_{\cdot}$。类似地，为 MLPCM 的每条边抽取新的显性位置，就形成了“混合潜在位置”结构，其中多个潜在位置可以同时刻画一个节点的特征。

MLPCM 也可以与（Hoff 2005年；Krivitsky 等人 2009 年）的随机效应网络模型联系起来。也就是说，期望式（1）中的模型等价于：

$$\begin{aligned} y_{ij} \sim \text{Pois}(\lambda_{ij}),\ \ \log(\lambda_{ij}) = \beta - ||\boldsymbol{u}_i - (\boldsymbol{u}_j + \boldsymbol{\epsilon}_{ij})||^2, \\ \boldsymbol{\epsilon}_{ij} \sim \text{MVN}_d(\mathbf{0}, 1/\gamma_j \mathbb{I}_d), \end{aligned} \tag{2}$$

其中，$\boldsymbol{\epsilon}_{ij}$ 可视为特定于边的随机效应，它可能包含发送者效应或接收者效应。然而，MLPCM 与随机效应模型的一个关键区别，在于将显性位置刻画为隐性位置的变体。这导致对潜在位置及其在定义节点社会特征中的作用有不同的解释。

使用期望式（1）为 MLPCM 生成虚拟数据十分简单，如算法 1 所示。为完整介绍我们的模型，我们展示两个模拟

算法 1 从 MLPCM 进行模拟。

输入：β，$\boldsymbol{\mu}$，$\boldsymbol{\tau}$，$\boldsymbol{\gamma}$ 以及 $\boldsymbol{z}$ 或 $\boldsymbol{\Pi}$。
如果 $\boldsymbol{z}$ 未提供**则**
  对 $i = 1, \ldots, N$ **do**
    1. $z_i \sim \text{cat}(\boldsymbol{\Pi})$.
  结束循环
结束条件判断

对 $j = 1, \ldots, N$**执行**
  2. $\boldsymbol{u}_j \sim$ 多元正态分布$_d(\boldsymbol{\mu}_{z_j}, 1/\tau_{z_j}\mathbb{I}_d)$.
  对 $i = 1, \ldots, N;\ \ i \neq j$**执行**
    3. $\boldsymbol{v}_{j \leftarrow i} \sim$ 多元正态分布$_d(\boldsymbol{u}_j, 1/\gamma_j \mathbb{I}_d)$.
  结束循环
结束循环
对 $i, j = 1, \ldots, N;\ \ i \neq j$**执行**
  4. $y_{ij} \sim$ 泊松分布 $[\exp(\beta - ||\boldsymbol{u}_i - \boldsymbol{v}_{j \leftarrow i}||^2)]$。
结束循环
输出：$\boldsymbol{Y}$，$\boldsymbol{U}$，$\boldsymbol{V}$，$\boldsymbol{z}$。

图1中的示例展示了不同二元组的隐性潜在位置和显性潜在位置之间的距离如何决定交互。不同颜色（红色、蓝色、绿色、紫色和橙色）对应网络中的不同个体，颜色较深的节点表示隐性潜在位置，颜色较浅的节点表示显性潜在位置。例如，深红色节点表示个体1的隐性潜在位置 $\boldsymbol{u}_1$，而围绕 $\boldsymbol{u}_1$ 的浅红色节点则是

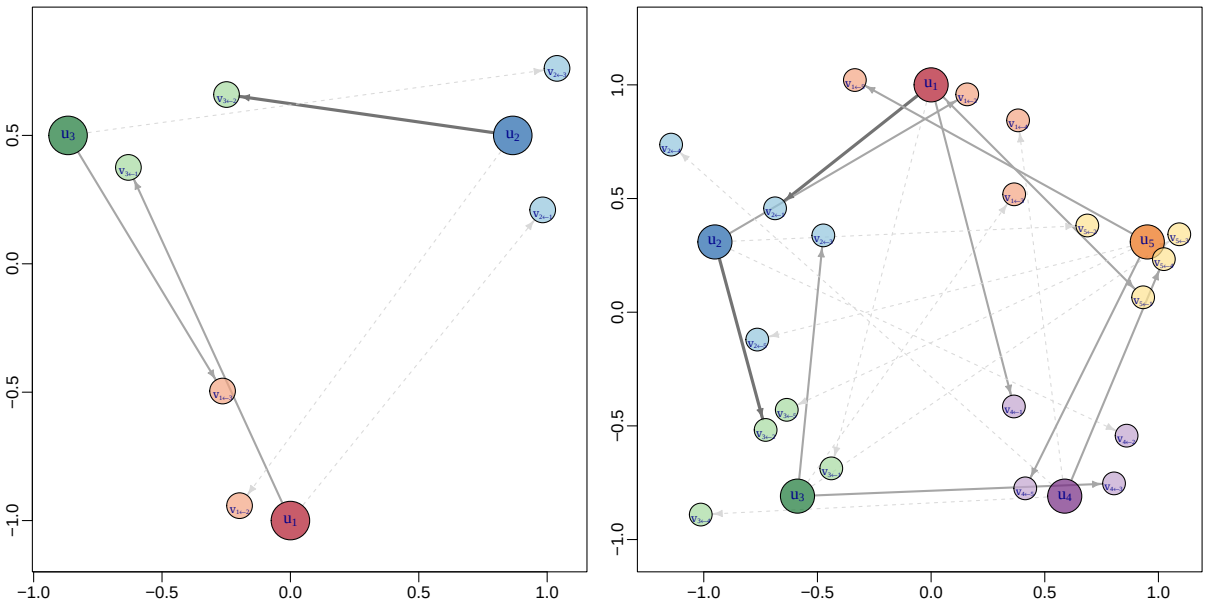

图1：由 MLPCM 生成的两个启发性网络示例。左侧网络有 3 个行动者，右侧网络有 5 个。每个深色节点表示各个体的隐性潜在位置，而对应的浅色节点则表示显性潜在位置。两个网络中的边均具有非负离散权重，取值范围为 0 到 2，边宽度表示其权重。节点之间的虚线表示边权重为零，实线则表示边权重非零。

从节点1指向其他节点的二元组所产生的显性潜在位置 $\{\boldsymbol{v}_{1\leftarrow 2}, \boldsymbol{v}_{1\leftarrow 3}, \dots\}$。

我们进一步通过引入层次结构扩展模型，并采用以下先验分布：

$$
\begin{aligned}
\beta &\sim \text{Normal}(\eta, \rho^2), \\
\boldsymbol{\mu}_k &\sim \text{MVN}_d(\mathbf{0}, \omega^2 \mathbb{I}_d), \quad \text{for } k = 1, \dots, K, \\
\tau_k &\sim \text{Gamma}(\xi, \psi), \quad \text{for } k = 1, \dots, K, \\
\gamma_i &\sim \text{Gamma}(a, b), \quad \text{for } i = 1, \dots, N, \\
\boldsymbol{\Pi} &\sim \text{Dirichlet}(\delta_1, \dots, \delta_K).
\end{aligned} \tag{3}
$$

因此，MLPCM 的后验分布如下：

$$
\begin{aligned}
&\pi(\boldsymbol{V}, \boldsymbol{U}, \boldsymbol{z}, \beta, \boldsymbol{\gamma}, \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\Pi} | \boldsymbol{Y}) \propto p(\boldsymbol{Y}|\boldsymbol{V}, \boldsymbol{U}, \beta) p(\boldsymbol{V}|\boldsymbol{U}, \boldsymbol{\gamma}) p(\boldsymbol{U}|\boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{z}) p(\boldsymbol{z}|\boldsymbol{\Pi}) \times \pi(\beta)\pi(\boldsymbol{\gamma})\pi(\boldsymbol{\mu})\pi(\boldsymbol{\tau})\pi(\boldsymbol{\Pi}) \\
&= \prod_{\substack{i,j=1,\\ i\neq j}}^{N} f_{\text{Pois}}\left(y_{ij} | \exp\left(\beta - ||\boldsymbol{u}_i - \boldsymbol{v}_{j\leftarrow i}||^2\right)\right) \prod_{\substack{i,j=1,\\ i\neq j}}^{N} f_{\text{MVN}_d}(\boldsymbol{v}_{j\leftarrow i} | \boldsymbol{u}_j, 1/\gamma_j \mathbb{I}_d) \prod_{i=1}^{N} f_{\text{MVN}_d}(\boldsymbol{u}_i | \boldsymbol{\mu}_{z_i}, 1/\tau_{z_i} \mathbb{I}_d) \prod_{i=1}^{N} f_{\text{cat}}(z_i | \boldsymbol{\Pi}) \\
&\quad \times f_{\text{Normal}}(\beta | \eta, \rho^2) \left[\prod_{i=1}^{N} f_{\text{Ga}}(\gamma_i | a, b) \prod_{k=1}^{K} f_{\text{MVN}_d}(\boldsymbol{\mu}_k | \mathbf{0}, \omega^2 \mathbb{I}_d) \prod_{k=1}^{K} f_{\text{Ga}}(\tau_k | \xi, \psi)\right] f_{\text{Dirichlet}}(\boldsymbol{\Pi} | \boldsymbol{\delta}).
\end{aligned}
$$

其中 $\boldsymbol{\delta} := (\delta_1, \dots, \delta_K)$ 是非负狄利克雷权重，a、b、ρ、ξ、$\psi > 0$ 和 $\eta \in \mathbb{R}$.

作为MLPCM的默认超参数，我们在所有应用中设置 $\eta = \rho^2 = \omega^2 = \xi = \psi = b = \delta_1 = \cdots = \delta_K = 1$和 $a = 10$。将 a 设为

较高值的动机在于，我们希望显性位置在其隐性位置周围具有较小的离散度。这在模型简约性的语境中是有意义的，因为我们试图惩罚一个可能过拟合的模型：在该模型中，显性位置并不被约束为接近其隐性位置，因而可能仅根据其所对应边的取值来加以识别。这可被视为一种过拟合情形，因为显性位置的数量随边的数量而增长，而这些参数并未受到任何有意义的约束。相比之下，对 $\boldsymbol{\gamma}$ 施加 Ga(10,1)先验能够确保显性位置被推向更接近其对应的隐性位置。从应用角度看，这一假设所描述的情形是：作为基线，节点的行为方式往往与其被感知的方式一致。选择不同的超参数完全可行，尤其是在可获得额外信息的情况下：我们将在第6.2节中对此作更详细的探讨。

3 变分近似

我们建立一种变分贝叶斯推断程序，用以近似 MLPCM 的后验分布。我们将模型参数的集合记为 $\boldsymbol{\Theta} := \{\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{z}, \beta, \boldsymbol{\mu}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \boldsymbol{\Pi}\}$。模型证据写为：

$$p(\boldsymbol{Y}) = p(\boldsymbol{Y})\frac{\pi(\boldsymbol{Y},\boldsymbol{\Theta})}{\pi(\boldsymbol{Y},\boldsymbol{\Theta})} = \frac{\pi(\boldsymbol{Y},\boldsymbol{\Theta})}{\pi(\boldsymbol{\Theta}|\boldsymbol{Y})} = \frac{\pi(\boldsymbol{Y},\boldsymbol{\Theta})}{q(\boldsymbol{\Theta})}\frac{q(\boldsymbol{\Theta})}{\pi(\boldsymbol{\Theta}|\boldsymbol{Y})}, \tag{4}$$

其中，$q(\boldsymbol{\Theta})$ 是所谓的变分分布，用于近似后验分布 $\pi(\boldsymbol{\Theta}|\boldsymbol{Y})$。一个常见假设是，$q(\cdot)$ 应属于一类形式比后验分布 $\pi(\boldsymbol{\Theta}|\boldsymbol{Y})$ 更简单的分布族。对式 (4) 作对数变换后，关于 $q(\boldsymbol{\Theta})$ 取期望，我们得到：

$$\log p(\boldsymbol{Y}) = \mathbb{E}_q[\log \pi(\boldsymbol{Y},\boldsymbol{\Theta})] - \mathbb{E}_q[\log q(\boldsymbol{\Theta})] + \mathbb{E}_q[\log q(\boldsymbol{\Theta})] - \mathbb{E}_q[\log \pi(\boldsymbol{\Theta}|\boldsymbol{Y})],$$

它对应于所谓的证据下界（ELBO）与 $q(\boldsymbol{\Theta})$ 和 $\pi(\boldsymbol{\Theta}|\boldsymbol{Y})$ 之间的 Kullback-Leibler（KL）散度之和。更明确地说：ELBO $=$期望$_q$[对数 $\pi(\boldsymbol{Y},\ \boldsymbol{\Theta})] -$期望$_q$[对数 $q(\boldsymbol{\Theta})]$,，而 KL$[q(\boldsymbol{\Theta})||\pi(\boldsymbol{\Theta}|\boldsymbol{Y})] =$期望$_q$[对数 $q(\boldsymbol{\Theta})] - \mathbb{E}_q$[对数 $\pi(\boldsymbol{\Theta}|\boldsymbol{Y})$]。显然，给定观测数据 $\boldsymbol{Y}$后，对数证据 $\log p(\boldsymbol{Y})$ 是固定的。因此，最小化 KL$[q(\boldsymbol{\Theta})||p(\boldsymbol{\Theta}|\boldsymbol{Y})]$ 等价于最大化上述 ELBO。

3.1 证据下界的推导

我们考虑满足均值场假设的 $q(\boldsymbol{\Theta})$ 的一组典型变分分布选择，如下方程所示：

$$
\begin{aligned}
q_{\boldsymbol{U}}(\boldsymbol{u}_i|\tilde{\boldsymbol{u}}_i,\tilde{\sigma}_i^2) &\sim \text{MVN}_d(\tilde{\boldsymbol{u}}_i,\tilde{\sigma}_i^2\mathbb{I}_d) \text{ for } i=1,\ldots,N,\\
q_{\boldsymbol{V}}(\boldsymbol{v}_{j\leftarrow i}|\tilde{\boldsymbol{v}}_{j\leftarrow i},\tilde{\varphi}_{ij}^2) &\sim \text{MVN}_d(\tilde{\boldsymbol{v}}_{j\leftarrow i},\tilde{\varphi}_{ij}^2\mathbb{I}_d) \text{ for } i,j=1,\ldots,N;\ i\neq j,\\
q_{\boldsymbol{z}}(z_i|\tilde{\pi}_{i1},\ldots,\tilde{\pi}_{iK}) &\sim \text{categorical}(\tilde{\pi}_{i1},\ldots,\tilde{\pi}_{iK}) \text{ for } i=1,\ldots,N,\\
q_{\beta}(\beta|\tilde{\eta},\tilde{\rho}^2) &\sim \text{Normal}(\tilde{\eta},\tilde{\rho}^2),\\
q_{\boldsymbol{\mu}}(\boldsymbol{\mu}_k|\tilde{\boldsymbol{\mu}}_k,\tilde{\omega}_k^2) &\sim \text{MVN}_d(\tilde{\boldsymbol{\mu}}_k,\tilde{\omega}_k^2\mathbb{I}_d) \text{ for } k=1,\ldots,K,\\
q_{\boldsymbol{\tau}}(\tau_k|\tilde{\xi}_k,\tilde{\psi}_k) &\sim \text{Gamma}(\tilde{\xi}_k,\tilde{\psi}_k) \text{ for } k=1,\ldots,K,\\
q_{\boldsymbol{\gamma}}(\gamma_j|\tilde{a}_j,\tilde{b}_j) &\sim \text{Gamma}(\tilde{a}_j,\tilde{b}_j) \text{ for } j=1,\ldots,N,\\
q_{\boldsymbol{\Pi}}(\pi_1,\ldots,\pi_K|\tilde{\delta}_1,\ldots,\tilde{\delta}_K) &\sim \text{Dirichlet}(\tilde{\delta}_1,\ldots,\tilde{\delta}_K).
\end{aligned}
\tag{5}
$$

我们用 $\tilde{\boldsymbol{\Theta}}:=\{\tilde{\boldsymbol{U}},\tilde{\boldsymbol{\sigma}}^{\boldsymbol{2}},\tilde{\boldsymbol{V}},\tilde{\boldsymbol{\varphi}}^{\boldsymbol{2}},\tilde{\boldsymbol{\Pi}},\tilde{\eta},\tilde{\rho}^2,\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\omega}}^{\boldsymbol{2}},\tilde{\boldsymbol{\xi}},\tilde{\boldsymbol{\psi}},\tilde{\boldsymbol{a}},\tilde{\boldsymbol{b}},\tilde{\boldsymbol{\delta}}\}$ 表示所有变分参数的集合，其中 $\tilde{\boldsymbol{U}}:=(\tilde{\boldsymbol{u}}_1,\ldots,\tilde{\boldsymbol{u}}_N)$ 是一个 $d\times N$ 矩阵，而 $\tilde{\boldsymbol{\sigma}}^{\boldsymbol{2}}:=(\tilde{\sigma}_1^2,\ldots,\tilde{\sigma}_N^2)^T$、$\tilde{\boldsymbol{a}}:=(\tilde{a}_1,\ldots,\tilde{a}_N)^T$ 和 $\tilde{\boldsymbol{b}}:=(\tilde{b}_1,\ldots,\tilde{b}_N)^T$ 是 $N\times 1$ 向量。~对象 $\tilde{\boldsymbol{V}}:=\{\tilde{\boldsymbol{v}}_{j\leftarrow i}:\ i,j=1,\ldots,N;\ i\neq j\}$ 由一个 $d\times N\times N$ 数组组成，而 $\tilde{\boldsymbol{\varphi}}^{\boldsymbol{2}}:=\{\tilde{\varphi}_{ij}^2:\ i,j=1,\ldots,N;\ i\neq j\}$ 是一个对角元素未定义的 $N\times N$ 矩阵。类似地，$\tilde{\boldsymbol{\Pi}}_i:=\begin{pmatrix}\tilde{\pi}_{i1} & \ldots & \tilde{\pi}_{iK}\end{pmatrix}$ $K\times 1$ $K\times N(,,)^T$ 是一个向量，构成~矩阵：$\tilde{\boldsymbol{\Pi}}:=(\boldsymbol{\Pi}_1,\ldots,\boldsymbol{\Pi}_N)$。变分参数集合 $\tilde{\boldsymbol{\mu}}:=(\tilde{\boldsymbol{\mu}}_1,\ldots,\tilde{\boldsymbol{\mu}}_K)$ 是一个 $d\times K$ 矩阵，而 $\tilde{\boldsymbol{\omega}}^{\boldsymbol{2}}:=(\tilde{\omega}_1^2,\ldots,\tilde{\omega}_K^2)^T$、$\tilde{\boldsymbol{\xi}}:=(\tilde{\xi}_1,\ldots,\tilde{\xi}_K)^T$、$\tilde{\boldsymbol{\psi}}:=(\tilde{\psi}_1,\ldots,\tilde{\psi}_K)^T$ 和 $\tilde{\boldsymbol{\delta}}:=(\tilde{\delta}_1,\ldots,\tilde{\delta}_K)^T$ 是 $K\times 1$ 向量。

~ ~令 $\mathcal{F}(\boldsymbol{\Theta})$ 表示证据下界，则 $\mathcal{F}(\boldsymbol{\Theta})$ 可分解为若干个可分别推导的项：

~证据下界 $=\mathcal{F}(\boldsymbol{\Theta})=\text{E}_q[\text{对数 }\pi(\boldsymbol{Y},\boldsymbol{\Theta})]-\text{E}_q[\text{对数 }q(\boldsymbol{\Theta})]=\text{E}_q[\text{对数 }p(\boldsymbol{Y}|\boldsymbol{U},\boldsymbol{V},\beta)]+\text{E}_q[\text{对数 }p(\boldsymbol{U}|\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{z})]+\text{E}_q[\text{对数 }p(\boldsymbol{V}|\boldsymbol{U},\boldsymbol{\gamma})]+\text{E}_q[\text{对数 }p(\boldsymbol{z}|\boldsymbol{\Pi})]+\text{E}_q[\text{对数 }\pi(\beta)]+\text{E}_q[\text{对数 }\pi(\boldsymbol{\mu})]+\text{E}_q[\text{对数 }\pi(\boldsymbol{\tau})]+\text{E}_q[\text{对数 }\pi(\boldsymbol{\gamma})]+\text{E}_q[\text{对数 }\pi(\boldsymbol{\Pi})]-\text{E}_q[\text{对数 }q_{\boldsymbol{U}}(\boldsymbol{U}|\tilde{\boldsymbol{U}},\tilde{\boldsymbol{\sigma}}^{\boldsymbol{2}})]-\text{E}_q[\text{对数 }q_{\boldsymbol{V}}(\boldsymbol{V}|\tilde{\boldsymbol{V}},\tilde{\boldsymbol{\varphi}}^{\boldsymbol{2}})]-\text{E}_q[\text{对数 }q_{\boldsymbol{z}}(\boldsymbol{z}|\tilde{\boldsymbol{\Pi}})]-\text{E}_q[\text{对数 }q_{\beta}(\beta|\tilde{\eta},\tilde{\rho}^2)]-\text{E}_q[\text{对数 }q_{\boldsymbol{\mu}}(\boldsymbol{\mu}|\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\omega}}^{\boldsymbol{2}})]-\text{E}_q[\text{对数 }q_{\boldsymbol{\tau}}(\boldsymbol{\tau}|\tilde{\boldsymbol{\xi}},\tilde{\boldsymbol{\psi}})]-\text{E}_q[\text{对数 }q_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}|\tilde{\boldsymbol{a}},\tilde{\boldsymbol{b}})]-\text{E}_q[\text{对数 }q_{\boldsymbol{\Pi}}(\boldsymbol{\Pi}|\tilde{\boldsymbol{\delta}})]$.

在许多框架中，无法通过解析方法获得证据下界，通常会利用詹森不等式近似来计算这一量。采用这种近似的相关研究包括 Salter-Townshend 和 Murphy（2013）、Gollini 和 Murphy（2016）以及 Jordan 等人（1999）。相比之下，MLPCM 建模框架的一个优势在于，可以利用以下命题精确计算证据下界。

命题1 在<*eq*>（*5*）</*eq*>的变分框架下，*MLPCM* 的证据下界中的第一项 $\mathbb{E}_q[\text{对数 }p(\boldsymbol{Y}|\boldsymbol{U},\boldsymbol{V},\beta)]$无需任何近似即可推导为：$\mathbb{E}_q[\text{对数 }p(\boldsymbol{Y}|\boldsymbol{U},\boldsymbol{V},\beta)]$

$$
=\sum_{\substack{i,j=1,\\ i\neq j}}^{N}\left\{y_{ij}\left[\tilde{\eta}-||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{v}}_{j\leftarrow i}||^2-d\left(\tilde{\sigma}_i^2+\tilde{\varphi}_{ij}^2\right)\right]-\frac{\exp\left(\tilde{\eta}+\frac{\tilde{\rho}^2}{2}-\frac{||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{v}}_{j\leftarrow i}||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2)^{d/2}}-\log(y_{ij}!)\right\}.
$$

命题2 在<*eq*>（*5*）</*eq*>的变分框架下，*MLPCM* 的证据下界可写为：

$$
\begin{aligned}
\mathcal{F}(\tilde{\boldsymbol{\Theta}}) = & \sum_{\substack{i,j=1,\\ i\neq j}}^{N} \left\{ y_{ij}\left[\tilde{\eta} - ||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j\leftarrow i}||^2 - d\left(\tilde{\sigma}_i^2 + \tilde{\varphi}_{ij}^2\right)\right] - \frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j\leftarrow i}||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2)^{d/2}} \right\} \\
& + \sum_{\substack{i,j=1,\\ i\neq j}}^{N} \left\{ \frac{d}{2}\left[\boldsymbol{\Psi}(\tilde{a}_j) - \log\left(\tilde{b}_j\right)\right] - \frac{1}{2}\frac{\tilde{a}_j}{\tilde{b}_j}[||\tilde{\boldsymbol{v}}_{j\leftarrow i} - \tilde{\boldsymbol{u}}_j||^2 + d(\tilde{\varphi}_{ij}^2 + \tilde{\sigma}_j^2)] + \frac{d}{2}\log\left(\tilde{\varphi}_{ij}^2\right) \right\} \\
& + \sum_{i=1}^{N}\sum_{k=1}^{K} \tilde{\pi}_{ik} \left\{ \frac{d}{2}\left[\boldsymbol{\Psi}(\tilde{\xi}_k) - \log\left(\tilde{\psi}_k\right)\right] - \frac{1}{2}\frac{\tilde{\xi}_k}{\tilde{\psi}_k}\left[||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{\mu}}_k||^2 + d(\tilde{\sigma}_i^2 + \tilde{\omega}_k^2)\right] \right\} \\
& + \sum_{i=1}^{N}\sum_{k=1}^{K} \tilde{\pi}_{ik} \left[\boldsymbol{\Psi}\left(\tilde{\delta}_k\right) - \boldsymbol{\Psi}\left(\sum_{g=1}^{K}\tilde{\delta}_g\right) - \log(\tilde{\pi}_{ik}) \right] \\
& + \sum_{k=1}^{K} \left\{ (\delta_k - \tilde{\delta}_k)\left[\boldsymbol{\Psi}\left(\tilde{\delta}_k\right) - \boldsymbol{\Psi}\left(\sum_{g=1}^{K}\tilde{\delta}_g\right)\right] + \log\Gamma(\tilde{\delta}_k) - \frac{1}{2\omega^2}\left(\tilde{\boldsymbol{\mu}}_k^T\tilde{\boldsymbol{\mu}}_k + d\tilde{\omega}_k^2\right) + \frac{d}{2}\log\left(\tilde{\omega}_k^2\right) \right\} \\
& + \sum_{k=1}^{K} \left\{ (\xi - \tilde{\xi}_k)\boldsymbol{\Psi}\left(\tilde{\xi}_k\right) - \xi\log\tilde{\psi}_k - \psi\frac{\tilde{\xi}_k}{\tilde{\psi}_k} + \log\Gamma(\tilde{\xi}_k) + \tilde{\xi}_k \right\} \\
& + \sum_{i=1}^{N} \left\{ (a - \tilde{a}_i)\boldsymbol{\Psi}\left(\tilde{a}_i\right) - a\log\tilde{b}_i - b\frac{\tilde{a}_i}{\tilde{b}_i} + \log\Gamma(\tilde{a}_i) + \tilde{a}_i + \frac{d}{2}\log\left(\tilde{\sigma}_i^2\right) \right\} \\
& - \frac{1}{2\rho^2}\left[(\tilde{\eta} - \eta)^2 + \tilde{\rho}^2\right] + \frac{1}{2}\log\left(\tilde{\rho}^2\right) - \log\Gamma\left(\sum_{k=1}^{K}\tilde{\delta}_k\right) + \texttt{const}.
\end{aligned}
\tag{6}
$$

ELBO 各项计算的详细证明见附录 A.1。

4 ELBO 的最大化

为最大化 ELBO，我们提出了一种结合闭式更新、标准梯度上升和自然梯度上升更新的算法框架。对于某些变分参数，可以使用闭式更新。对于其他所有参数，我们使用标准梯度上升或自然梯度上升（Amari 1998；Martens 2020）。

4.1 闭式更新

在下面的命题中，我们给出变分参数：$\{\boldsymbol{\Pi}, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\omega}}^2, \boldsymbol{\xi}, \boldsymbol{\psi}, \tilde{\boldsymbol{a}}, \boldsymbol{b}\}$的闭式更新步骤，它们分别对应于式（5）中所示的变分分布：$\{q_{\boldsymbol{z}}(z_i|\tilde{\boldsymbol{\Pi}}_i)\}$，$\{q_{\boldsymbol{\mu}}(\boldsymbol{\mu}_k|\tilde{\boldsymbol{\mu}}_k,\ \tilde{\omega}_k^2)\}$，$\{q_{\boldsymbol{\tau}}(\tau_k|\tilde{\xi}_k,\ \tilde{\psi}_k)\}$，$\{q_{\boldsymbol{\gamma}}(\gamma_j|\tilde{a}_j,\ \tilde{b}_j)\}$。详细证明见附录 A.2.1。

命题 3 对于 Θ的任意取值，

1. $\tilde{\pi}_{jk}$的闭式更新，对于 $j = 1$、2、.....、N以及对于 $k = 1$、2、.....、K，写作：$\tilde{\pi}^{\star}_{jk} = \exp\big(A_{jk}\big)/\sum_{g=1}^{K}\exp(A_{jg})$，其中

$$A_{jk} = \frac{d}{2}\left[\boldsymbol{\Psi}\left(\tilde{\xi}_k\right) - \log\left(\tilde{\psi}_k\right)\right] - \frac{1}{2}\frac{\tilde{\xi}_k}{\tilde{\psi}_k}\left[||\tilde{\boldsymbol{u}}_j - \tilde{\boldsymbol{\mu}}_k||^2 + d(\tilde{\sigma}_j^2 + \tilde{\omega}_k^2)\right] + \boldsymbol{\Psi}\left(\tilde{\delta}_k\right) - \boldsymbol{\Psi}\left(\sum_{g=1}^{K}\tilde{\delta}_g\right). \tag{7}$$

2. $\tilde{\boldsymbol{\mu}}_k$和 $\tilde{\omega}_k^2$的闭式更新，对于 $k = 1$、2、.....、K，分别写作：

$$\tilde{\boldsymbol{\mu}}_k^{\star} = \frac{\omega^2\tilde{\xi}_k\left(\sum_{i=1}^{N}\tilde{\pi}_{ik}\tilde{\boldsymbol{u}}_i\right)}{\omega^2\tilde{\xi}_k\left(\sum_{i=1}^{N}\tilde{\pi}_{ik}\right) + \tilde{\psi}_k}, \qquad \tilde{\omega}_k^{2^{\star}} = \frac{\omega^2\tilde{\psi}_k}{\tilde{\psi}_k + \omega^2\tilde{\xi}_k\sum_{i=1}^{N}\tilde{\pi}_{ik}}. \tag{8}$$

3. ξ_k和 ψ_k的闭式更新，对于 $k = 1$、2、.....、K，写作：

$$\tilde{\xi}_k^{\star} = \xi + \frac{d}{2}\sum_{i=1}^{N}\tilde{\pi}_{ik}, \qquad \tilde{\psi}_k^{\star} = \psi + \frac{1}{2}\sum_{i=1}^{N}\tilde{\pi}_{ik}\left[||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{\mu}}_k||^2 + d(\tilde{\sigma}_i^2 + \tilde{\omega}_k^2)\right]. \tag{9}$$

4. $\tilde{a}_j$和 b_j的闭式更新，对于 $j = 1$、2、.....、N，写作：

$$\tilde{a}_j^{\star} = a + \frac{d}{2}(N-1), \qquad \tilde{b}_j^{\star} = b + \frac{1}{2}\sum_{\substack{i=1,\\ i\neq j}}^{N}\left[||\tilde{\boldsymbol{v}}_{j\leftarrow i} - \tilde{\boldsymbol{u}}_j||^2 + d(\tilde{\varphi}_{ij}^2 + \tilde{\sigma}_j^2)\right]. \tag{10}$$

4.2 自然梯度上升与标准梯度上升更新

标准梯度上升的实现形式为 $\tilde{\boldsymbol{\theta}}^{\star} = \tilde{\boldsymbol{\theta}}^{\dagger} + \epsilon\nabla\mathcal{F}(\tilde{\boldsymbol{\theta}}^{\dagger})$。这里，$\tilde{\boldsymbol{\theta}}^{\dagger}$表示变分参数的当前状态，而 $\boldsymbol{\theta}^{\star}$ 表示其对应于更优证据下界值的新更新状态。自然梯度上升进一步纳入费 Fisher 信息矩阵 $\mathcal{I}(\boldsymbol{\theta})$，通过考虑参数空间的内在几何结构来修正梯度，从而实现更高效的优化。假设变分参数 $\boldsymbol{\theta}$ 由相应的变分分布 $q(\boldsymbol{\theta}|\boldsymbol{\theta})$ 按照期望（5）得到，则自然梯度上升的实现形式为 $\tilde{\boldsymbol{\theta}}^{\star} = \tilde{\boldsymbol{\theta}}^{\dagger} + \epsilon\mathcal{I}^{-1}(\tilde{\boldsymbol{\theta}}^{\dagger})\nabla\mathcal{F}(\tilde{\boldsymbol{\theta}}^{\dagger})$，其中

$$\mathcal{I}(\tilde{\boldsymbol{\theta}}) = -\mathbb{E}_{q(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}})}\left[\nabla^2_{\tilde{\boldsymbol{\theta}}}\log q(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}})\right] \equiv \mathbb{E}_{q(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}})}\left\{\left[\nabla_{\tilde{\boldsymbol{\theta}}}\log q(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}})\right]\left[\nabla_{\tilde{\boldsymbol{\theta}}}\log q(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}})\right]^T\right\}.$$

在我们的情形中，按照以下命题实现自然梯度步骤，以优化变分参数 $\{\tilde{\boldsymbol{U}}, \tilde{\boldsymbol{\sigma}}^2, \tilde{\boldsymbol{V}}, \tilde{\boldsymbol{\varphi}}^2, \tilde{\eta}, \tilde{\rho}^2\}$，它们分别对应于变分分布 $\{q_{\boldsymbol{U}}(\boldsymbol{u}_i|\tilde{\boldsymbol{u}}_i,\ \tilde{\sigma}_i^2)\}$，$\{q_{\boldsymbol{V}}(\boldsymbol{v}_{j\leftarrow i}|\tilde{\boldsymbol{v}}_{j\leftarrow i},\ \tilde{\varphi}_{ij}^2)\}$和 $q_{\beta}(\beta|\tilde{\eta},\ \tilde{\rho}^2)$。

~**命题4** 对于 $\boldsymbol{\Theta}$ 的任意取值以及足够小的 $\epsilon > 0$,

1. 对于 $i = 1$、2、.....、N，$(\tilde{\boldsymbol{u}}_i, \tilde{\sigma}_i^2)$ 的自然梯度上升更新不会降低 *MLPCM* 的证据下界，其定义为：

$$\begin{cases} \tilde{\boldsymbol{u}}_i^{\star} = \tilde{\boldsymbol{u}}_i^{\dagger} + \epsilon \tilde{\sigma}_i^{2\dagger} \frac{\partial \mathcal{F}}{\partial \tilde{\boldsymbol{u}}_i} \left(\tilde{\boldsymbol{u}}_i^{\dagger}, \tilde{\sigma}_i^{2\dagger}, \dots \right); \\ \tilde{\sigma}_i^{2\star} = \tilde{\sigma}_i^{2\dagger} \exp\left[\epsilon \frac{2}{d} \tilde{\sigma}_i^{2\dagger} \frac{\partial \mathcal{F}}{\partial \tilde{\sigma}_i^2} \left(\tilde{\boldsymbol{u}}_i^{\dagger}, \tilde{\sigma}_i^{2\dagger}, \dots \right) \right], \end{cases} \tag{11}$$

其中 $(\cdot)^{\star}$ 与 $(\cdot)^{\dagger}$ 分别表示变分参数的新状态与当前状态，并且，

$$\begin{aligned} \frac{\partial \mathcal{F}}{\partial \tilde{\boldsymbol{u}}_i} &= \sum_{\substack{j=1, \\ j \neq i}}^{N} \left\{ -2 \left[y_{ij} - \frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j \leftarrow i}||^2}{1 + 2\tilde{\sigma}_i^2 + 2\tilde{\varphi}_{ij}^2} \right)}{(1 + 2\tilde{\sigma}_i^2 + 2\tilde{\varphi}_{ij}^2)^{d/2+1}} \right] (\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j \leftarrow i}) + \frac{\tilde{a}_i}{\tilde{b}_i} (\tilde{\boldsymbol{v}}_{i \leftarrow j} - \tilde{\boldsymbol{u}}_i) \right\} - \sum_{k=1}^{K} \left[\tilde{\pi}_{ik} \frac{\tilde{\xi}_k}{\tilde{\psi}_k} (\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{\mu}}_k) \right], \\ \frac{\partial \mathcal{F}}{\partial \tilde{\sigma}_i^2} &= \frac{d}{2} \cdot \left[-2 \left(\sum_{\substack{j=1, \\ j \neq i}}^{N} y_{ij} \right) - (N-1) \frac{\tilde{a}_i}{\tilde{b}_i} - \left(\sum_{k=1}^{K} \tilde{\pi}_{ik} \frac{\tilde{\xi}_k}{\tilde{\psi}_k} \right) + \frac{1}{\tilde{\sigma}_i^2} \right] - \sum_{\substack{j=1, \\ j \neq i}}^{N} \frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j \leftarrow i}||^2}{1 + 2\tilde{\sigma}_i^2 + 2\tilde{\varphi}_{ij}^2} \right)}{(1 + 2\tilde{\sigma}_i^2 + 2\tilde{\varphi}_{ij}^2)^{d/2+1}} \left[2 \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j \leftarrow i}||^2}{1 + 2\tilde{\sigma}_i^2 + 2\tilde{\varphi}_{ij}^2} - d \right]. \end{aligned} \tag{12}$$

2. 对于 i、$j = 1$、2、.....、N 和 $i \neq j$，$(\tilde{\boldsymbol{v}}_{j \leftarrow i}, \tilde{\varphi}_{ij}^2)$ 的自然梯度上升更新不会降低 *MLPCM* 的证据下界，其定义为：

$$\begin{cases} \tilde{\boldsymbol{v}}_{j \leftarrow i}^{\star} = \tilde{\boldsymbol{v}}_{j \leftarrow i}^{\dagger} + \epsilon \tilde{\varphi}_{ij}^{2\dagger} \frac{\partial \mathcal{F}}{\partial \tilde{\boldsymbol{v}}_{j \leftarrow i}} \left(\tilde{\boldsymbol{v}}_{j \leftarrow i}^{\dagger}, \tilde{\varphi}_{ij}^{2\dagger}, \dots \right); \\ \tilde{\varphi}_{ij}^{2\star} = \tilde{\varphi}_{ij}^{2\dagger} \exp\left[\epsilon \frac{2}{d} \tilde{\varphi}_{ij}^{2\dagger} \frac{\partial \mathcal{F}}{\partial \tilde{\varphi}_{ij}^2} \left(\tilde{\boldsymbol{v}}_{j \leftarrow i}^{\dagger}, \tilde{\varphi}_{ij}^{2\dagger}, \dots \right) \right], \end{cases} \tag{13}$$

其中，

$$\begin{aligned} \frac{\partial \mathcal{F}}{\partial \tilde{\boldsymbol{v}}_{j \leftarrow i}} &= \left\{ 2 \left[y_{ij} - \frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j \leftarrow i}||^2}{1 + 2\tilde{\sigma}_i^2 + 2\tilde{\varphi}_{ij}^2} \right)}{(1 + 2\tilde{\sigma}_i^2 + 2\tilde{\varphi}_{ij}^2)^{d/2+1}} \right] (\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j \leftarrow i}) - \frac{\tilde{a}_j}{\tilde{b}_j} (\tilde{\boldsymbol{v}}_{j \leftarrow i} - \tilde{\boldsymbol{u}}_j) \right\}, \\ \frac{\partial \mathcal{F}}{\partial \tilde{\varphi}_{ij}^2} &= \frac{d}{2} \left(-2y_{ij} - \frac{\tilde{a}_j}{\tilde{b}_j} + \frac{1}{\tilde{\varphi}_{ij}^2} \right) - \frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j \leftarrow i}||^2}{1 + 2\tilde{\sigma}_i^2 + 2\tilde{\varphi}_{ij}^2} \right) \left[2 \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j \leftarrow i}||^2}{1 + 2\tilde{\sigma}_i^2 + 2\tilde{\varphi}_{ij}^2} - d \right]}{(1 + 2\tilde{\sigma}_i^2 + 2\tilde{\varphi}_{ij}^2)^{d/2+1}}. \end{aligned}$$

3. $(\tilde{\eta}, \tilde{\rho}^2)$ 的自然梯度上升更新不会降低 *MLPCM* 的证据下界，其定义为：

$$\begin{cases} \tilde{\eta}^{\star} = \tilde{\eta}^{\dagger} + \epsilon \tilde{\rho}^{2\dagger} \frac{\partial \mathcal{F}}{\partial \tilde{\eta}} \left(\tilde{\eta}^{\dagger}, \tilde{\rho}^{2\dagger}, \dots \right); \\ \tilde{\rho}^{2\star} = \tilde{\rho}^{2\dagger} \exp\left[\epsilon 2 \tilde{\rho}^{2\dagger} \frac{\partial \mathcal{F}}{\partial \tilde{\rho}^2} \left(\tilde{\eta}^{\dagger}, \tilde{\rho}^{2\dagger}, \dots \right) \right], \end{cases} \tag{14}$$

其中，

$$\begin{aligned} \frac{\partial \mathcal{F}}{\partial \tilde{\eta}} &= \left(\sum_{\substack{i,j=1, \\ i \neq j}}^{N} y_{ij} \right) - \frac{\tilde{\eta} - \eta}{\rho^2} - \sum_{\substack{i,j=1, \\ i \neq j}}^{N} \frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j \leftarrow i}||^2}{1 + 2\tilde{\sigma}_i^2 + 2\tilde{\varphi}_{ij}^2} \right)}{(1 + 2\tilde{\sigma}_i^2 + 2\tilde{\varphi}_{ij}^2)^{d/2}}, \\ \frac{\partial \mathcal{F}}{\partial \tilde{\rho}^2} &= -\frac{1}{2\rho^2} + \frac{1}{2\tilde{\rho}^2} - \frac{1}{2} \sum_{\substack{i,j=1, \\ i \neq j}}^{N} \frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j \leftarrow i}||^2}{1 + 2\tilde{\sigma}_i^2 + 2\tilde{\varphi}_{ij}^2} \right)}{(1 + 2\tilde{\sigma}_i^2 + 2\tilde{\varphi}_{ij}^2)^{d/2}}. \end{aligned}$$

上述命题的证明详见附录 A.2.2。请注意，自然梯度上升需要计算逆费希尔信息。如果费 Fisher 信息矩阵是对角矩阵，则很容易求得其逆，但事实并非总是如此。因此，计算逆矩阵会带来计算上的问题，因为这可能非常耗时。对于费希尔信息矩阵不是对角矩阵的更新，我们采用标准梯度上升进行优化。˜例如，与变分˜分布 $q_{\mathbf{\Pi}}(\mathbf{\Pi}|\boldsymbol{\delta})$ 相关的变分参数 $\boldsymbol{\delta}$ 就属于这种情况，我们将在下文的命题中专门讨论这一点（证明见附录 A.2.3）。

˜**命题5** 对于任意的 $\mathbf{\Theta}$，当 $\epsilon > 0$ 足够小时，*MLPCM* 的标准梯度˜ ˜ ˜上升更新不会降低证据下界，其定义如下：

$$\tilde{\delta}_k^{\star} = \tilde{\delta}_k^{\dagger} \exp\left[\epsilon \tilde{\delta}_k^{\dagger} \frac{\partial \mathcal{F}}{\partial \tilde{\delta}_k} \left(\tilde{\delta}_1^{\dagger}, \tilde{\delta}_2^{\dagger}, \ldots, \tilde{\delta}_K^{\dagger}, \ldots\right)\right], \; for \; k = 1, 2, \ldots, K, \tag{15}$$

其中，$(\cdot)^{\star}$ 和$(\cdot)^{\dagger}$ 分别表示变分参数的新状态和当前状态，并且，

$$\frac{\partial \mathcal{F}}{\partial \tilde{\delta}_k} = \mathbf{\Psi}'\left(\tilde{\delta}_k\right) \cdot \left(\delta_k - \tilde{\delta}_k + \sum_{i=1}^{N} \tilde{\pi}_{ik}\right) - \mathbf{\Psi}'\left(\sum_{g=1}^{K} \tilde{\delta}_g\right) \cdot \left[\sum_{g=1}^{K} \left(\delta_g - \tilde{\delta}_g + \sum_{i=1}^{N} \tilde{\pi}_{ig}\right)\right].$$

5 优化算法与模型选择

5.1 优化

由变分更新得到的变分贝叶斯优化算法如算法 2 所示。每个变分参数的学习率最初设为任意大的值。随后，每当自然梯度更新导致证据下界降低时，相应学习率就减半。这样，算法 2 中的过程可确保证据下界不会下降；如果将容差水平 `tol` 设得足够小，则可以获得证据下界的最大值。然而，这种最大化并不能保证达到证据下界的全局最大值，因此，使用参数的随机初始化进行多次运行，有助于验证输出。

5.2 模型选择

算法 2 以预先设定的簇数 K 作为输入。为了推断 K 的最佳取值，我们使用不同的 K 候选值运行该算法，然后比较各个解以选出最佳解。为了选择理想的组数，我们提出了一种用于模型选择的新型部分积分完全似然（PICL）得分。该准则结合了积分完全似然（ICL）的思想，以及 Biernacki 等人（2010）和 Cˆome 的精确 ICL（ICL$_{ex}$）思想，

算法 2 MLPCM 的变分贝叶斯推断算法。

输入：$\boldsymbol{Y}$，d，K，a，b，η，ρ^2，ω^2，ξ，ψ，$\boldsymbol{\delta}$，容差水平为 `tol`，梯度上升步长为 $\{\epsilon_{\boldsymbol{u}_i}\}_{i=1}^N$，$\{\epsilon_{\boldsymbol{v}_{j\leftarrow i}}\}_{i\ ,\ j=1,\ i\neq j}^N$，$\epsilon_\beta$，$\epsilon_{\boldsymbol{\Pi}}$，以及初始状态：

$$\tilde{\boldsymbol{\Theta}}^{(0)}=\{\tilde{\boldsymbol{U}}^{(0)},\tilde{\boldsymbol{\sigma}}^{\boldsymbol{2}(0)},\tilde{\boldsymbol{V}}^{(0)},\tilde{\boldsymbol{\varphi}}^{\boldsymbol{2}(0)},\tilde{\boldsymbol{\Pi}}^{(0)},\tilde{\eta}^{(0)},\tilde{\rho}^{2(0)},\tilde{\boldsymbol{\mu}}^{(0)},\tilde{\boldsymbol{\omega}}^{\boldsymbol{2}(0)},\tilde{\boldsymbol{\xi}}^{(0)},\tilde{\boldsymbol{\psi}}^{(0)},\tilde{\boldsymbol{a}}^{(0)},\tilde{\boldsymbol{b}}^{(0)},\tilde{\boldsymbol{\delta}}^{(0)}\}.$$

设置 `stop=FALSE` 和 $t=0$。计算初始状态下的证据下界：$\mathcal{F}(\tilde{\boldsymbol{\Theta}}^{(0)})$.

当 `stop` $\neq$ `TRUE` **时**

对于 $i=1,\ldots,N$ **do**

1. 根据期望 (11)，使用 $\epsilon=\epsilon_{\boldsymbol{u}_i}$ 更新 $(\tilde{\boldsymbol{u}}_i^{(t+1)},\ \tilde{\sigma}_i^{2(t+1)})$。

如果 $\mathcal{F}(\tilde{\boldsymbol{u}}_i^{(t+1)},\tilde{\sigma}_i^{2(t+1)},\ldots)<\mathcal{F}(\tilde{\boldsymbol{u}}_i^{(t)},\tilde{\sigma}_i^{2(t)},\ldots)$ **那么** $\epsilon_{\boldsymbol{u}_i}=\epsilon_{\boldsymbol{u}_i}/2$; 重新执行步骤 1.

结束如果

结束循环

对于 $i,j=1,\ldots,N$ 和 $i\neq j$ **执行**

2. 按照期望 (13) 用 $\epsilon=\epsilon_{\boldsymbol{v}}$ 更新 $(\tilde{\boldsymbol{v}}_{j\leftarrow i}^{(t+1)},\tilde{\varphi}_{ij}^{2(t+1)})$

如果 $\mathcal{F}(\tilde{\boldsymbol{v}}_{j\leftarrow i}^{(t+1)},\tilde{\varphi}_{ij}^{2(t+1)},\ldots)<\mathcal{F}(\tilde{\boldsymbol{v}}_{j\leftarrow i}^{(t)},\tilde{\varphi}_{ij}^{2(t)}\ldots)$ **然后** $\epsilon_{\boldsymbol{v}\ j\leftarrow i}=\epsilon_{\boldsymbol{v}_{j\leftarrow i}}/2$; 重新运行第 2步。

结束如果

结束循环

对于 $j=1,\ldots,N$ **执行**

对于 $k=1,\ldots,K$ **do**

3. 更新 $\tilde{\pi}_{jk}^{(t+1)}$，按照期望 (7)。

结束循环

4. 按照期望 (10) 更新 $(\tilde{a}_j^{(t+1)},\ \tilde{b}_j^{(t+1)})$。

结束 for

5. 按照期望 (14)，用 $\epsilon=\epsilon_\beta$ 更新 $(\tilde{\eta}^{(t+1)},\tilde{\rho}^{2(t+1)})$。

如果 $\mathcal{F}(\tilde{\eta}^{(t+1)},\tilde{\rho}^{2(t+1)}\ldots)<\mathcal{F}(\tilde{\eta}^{(t)},\tilde{\rho}^{2(t+1)},\ldots)$ **然后** $\epsilon_\beta=\epsilon_\beta/2$，步骤 5。

结束 if

对于 $k=1,\ldots,K$ **do**

6. 更新 $\tilde{\boldsymbol{\mu}}_k^{(t+1)}$ 和 $\tilde{\omega}_k^{2(t+1)}$，遵循期望公式 (8)。

7. 更新 $(\tilde{\xi}_k^{(t+1)},\ \tilde{\psi}_k^{(t+1)})$，遵循期望公式 (9)。

8. 更新 $\tilde{\delta}_k^{(t+1)}$，遵循期望公式 (15)，其中 $\epsilon=\epsilon_{\boldsymbol{\Pi}}$。

结束

如果 $\mathcal{F}(\tilde{\boldsymbol{\delta}}^{(t+1)},\ldots)<\mathcal{F}(\tilde{\boldsymbol{\delta}}^{(t)},\ldots)$ **则** $\epsilon_{\boldsymbol{\Pi}}=\epsilon_{\boldsymbol{\Pi}}/2$，重新执行步骤 8，针对 $k=1,\ \ldots,\ K$。

结束如果

如果 $\mathcal{F}(\tilde{\boldsymbol{\Theta}}^{(t+1)})-\mathcal{F}(\tilde{\boldsymbol{\Theta}}^{(t)})<$ `tol` **则** `stop=TRUE`

结束如果

设置 $t=t+1$。

结束 while

输出：$\{\tilde{\boldsymbol{\Theta}}^{(s)},\mathcal{F}(\tilde{\boldsymbol{\Theta}}^{(s)}):s=1,2,\ldots,t\}$.

Latouche（2015年）。当似然与先验分布的共轭性能够产生部分“折叠”的后验分布时，PICL 适用；在这种情况下，部分模型参数已被积分消去。

更具体地说，我们首先考虑 MLPCM 的 $\boldsymbol{Y}$、 $\boldsymbol{U}$、 $\boldsymbol{V}$、 $\boldsymbol{z}|K$积分对数似然，其形式如下:

$$\begin{aligned}
&\log p(\boldsymbol{Y},\boldsymbol{U},\boldsymbol{V},\boldsymbol{z}|K) = \log[p(\boldsymbol{Y}|\boldsymbol{U},\boldsymbol{V},\boldsymbol{z},K)p(\boldsymbol{U},\boldsymbol{V},\boldsymbol{z}|K)] \\
&= \log[p(\boldsymbol{Y}|\boldsymbol{U},\boldsymbol{V})p(\boldsymbol{V}|\boldsymbol{U},\boldsymbol{z},K)p(\boldsymbol{U},\boldsymbol{z}|K)] = \log[p(\boldsymbol{Y}|\boldsymbol{U},\boldsymbol{V})p(\boldsymbol{V}|\boldsymbol{U})p(\boldsymbol{U}|\boldsymbol{z})p(\boldsymbol{z}|K)] \\
&= \log \int_{\beta} p(\boldsymbol{Y}|\boldsymbol{U},\boldsymbol{V},\beta)\pi(\beta)\mathrm{d}\beta + \log \int_{\boldsymbol{\gamma}} p(\boldsymbol{V}|\boldsymbol{U},\boldsymbol{\gamma})\pi(\boldsymbol{\gamma})\mathrm{d}\boldsymbol{\gamma} \\
&\quad + \log \int_{\boldsymbol{\tau}} \int_{\boldsymbol{\mu}} p(\boldsymbol{U}|\boldsymbol{z},\boldsymbol{\mu},\boldsymbol{\tau})\pi(\boldsymbol{\mu})\pi(\boldsymbol{\tau})\mathrm{d}\boldsymbol{\mu}\mathrm{d}\boldsymbol{\tau} + \log \int_{\boldsymbol{\Pi}} p(\boldsymbol{z}|\boldsymbol{\Pi})\pi(\boldsymbol{\Pi}|K)\mathrm{d}\boldsymbol{\Pi}.
\end{aligned} \tag{16}$$

由期望 16 可知：（i）第一项中的参数 β 不易直接积分消去，因此我们采用贝叶斯信息准则（BIC）近似（Schwarz，1978年）来计算该项；（ii）由于存在共轭先验分布，第二项和第四项中的模型参数 $\boldsymbol{\gamma}$ 和 $\boldsymbol{\Pi}$ 可以很容易地积分消去；（iii） $\boldsymbol{\mu}$ 或 $\boldsymbol{\tau}$可以被积分消去，因此我们提出对参数 $\boldsymbol{\mu}$进行边缘化处理。该参数的维数高于 $\boldsymbol{\tau}$，随后我们可以对剩余的部分积分对数似然项应用 BIC近似。

对部分参数进行边缘化后，我们得到一个可用于模型选择的部分积分 ICL。该准则的方程如下方命题所示。

^ ^ 命题 6 给定 *MLPCM* 的模型参数点估计 $\boldsymbol{U},\boldsymbol{V}$，$\hat{}\boldsymbol{z}$，期望 (16) 的 *PICL* 准则等于:

$$\begin{aligned}
PICL = \texttt{const} &- \left(\sum_{\substack{i,j=1,\\ i\neq j}}^{N} y_{ij}\right) \log\left[\sum_{\substack{i,j=1,\\ i\neq j}}^{N} \exp\left(-||\hat{\boldsymbol{u}}_i - \hat{\boldsymbol{v}}_{j\leftarrow i}||^2\right)\right] - \sum_{\substack{i,j=1,\\ i\neq j}}^{N} \left(y_{ij}||\hat{\boldsymbol{u}}_i - \hat{\boldsymbol{v}}_{j\leftarrow i}||^2\right) \\
&- \left[a + \frac{d}{2}(N-1)\right] \sum_{j=1}^{N} \log\left(b + \frac{1}{2}\sum_{\substack{i=1,\\ i\neq j}}^{N} ||\hat{\boldsymbol{v}}_{j\leftarrow i} - \hat{\boldsymbol{u}}_j||^2\right) - \frac{1}{2}K\log(N) \\
&+ \max_{\boldsymbol{\tau}} \sum_{g=1}^{K} \left[\frac{d}{2}\hat{n}_g \log(\tau_g) - \frac{d}{2}\log\left(\tau_g \hat{n}_g \omega^2 + 1\right) + \frac{1}{2}\frac{\tau_g^2\omega^2}{\tau_g\hat{n}_g\omega^2+1}\left\|\sum_{\hat{z}_i=g}\hat{\boldsymbol{u}}_i\right\|^2 - \frac{\tau_g}{2}\left(\sum_{\hat{z}_i=g}||\hat{\boldsymbol{u}}_i||^2\right)\right] \\
&+ \left[\sum_{k=1}^{K}\log\Gamma(\hat{n}_k+\delta)\right] - \log\Gamma\left(N+K\delta\right) + \log\Gamma\left(K\delta\right) - K\log\Gamma(\delta),
\end{aligned}$$

其中，

$$\texttt{const} = \left(\sum_{\substack{i,j=1,\\ i\neq j}}^{N} y_{ij}\right)\left[\log\left(\sum_{\substack{i,j=1,\\ i\neq j}}^{N} y_{ij}\right) - 1\right] - \left[\sum_{\substack{i,j=1,\\ i\neq j}}^{N} \log(y_{ij}!)\right] - \frac{1}{2}\log[N(N-1)]$$

$$+ \sum_{j=1}^{N}\left[-\frac{d}{2}(N-1)\log(2\pi) + a\log(b) - \log\Gamma(a) + \log\Gamma\left(a + \frac{d}{2}(N-1)\right)\right] - \frac{d}{2}N\log(2\pi),$$

和　　由针对 $\{\tau_g$：　$g = 12$、.....、 K 进行的标准梯度上升得到最大值$\tau(\cdot)$ 项　　$\}$:

$$\tau_g^{\star} = \tau_g^{\dagger}\exp\left[\epsilon\tau_g^{\dagger}\frac{\partial h}{\partial \tau_g}\left(\tau_g^{\dagger}\right)\right].$$

这里，令 $\hat{}n_g := \sum_{i=1}^{N}\mathbb{1}(\hat{}z_i = g)$，则偏导数可写为

$$\frac{\partial h(\boldsymbol{\tau})}{\partial \tau_g} = \frac{d}{2}\hat{n}_g\left(\frac{1}{\tau_g} - \frac{\omega^2}{\tau_g\hat{n}_g\omega^2 + 1}\right) + \frac{\tau_g\omega^2\left(\tau_g\hat{n}_g\omega^2 + 2\right)}{2\left(\tau_g\hat{n}_g\omega^2 + 1\right)^2}\left\|\sum_{\hat{z}_i=g}\hat{\boldsymbol{u}}_i\right\|^2 - \frac{1}{2}\left(\sum_{\hat{z}_i=g}||\hat{\boldsymbol{u}}_i||^2\right).$$

根据 PICL 的定义，并且由于它与其他广泛使用的惩罚似然模型选择准则密切相关，我们选择使 PICL 最大化的值 K 作为最优值。

5.3 初始化

变分贝叶斯方法可能会遇到收敛问题，因为优化过程可能只会得到近似的局部后验众数。使用不同的随机起点多次重新运行算法，或许有助于解决收敛问题，但这会增加计算成本，因为算法必须运行多次。对于参数空间维数较高的复杂模型，这可能变得不切实际，因为随机化往往会提出较差的初始状态，使推断算法更容易收敛到某些局部后验众数。因此，良好的初始化方法对于有效探索解空间并获得良好的推断结果至关重要。本文提出以下默认初始化方法，该方法在模拟数据和真实数据应用中均表现良好。

˜通过对 $\boldsymbol{Y}$ 的测地距离矩阵应用 Gower（1966）的经典度量多维尺度分析（MDS），初始化变分参数 $\boldsymbol{U}$；而$\{\tilde{\boldsymbol{\upsilon}}_{j\leftarrow i}$：　$i = 1$，2，.....，N；$i \neq j\}$则初始化为与 $j = 1$，2，.....，N 对应的 $\tilde{\boldsymbol{u}}_j$保持一致。对初始 $\boldsymbol{U}$应用 k均值算法，并采用 1000 个随机初始点，根据个体到相应组中心的平方和最小化结果，得到初步分组；因此，变分参数 $\tilde{\boldsymbol{\mu}}$ 根据 $\boldsymbol{U}$初始化为相应的组中心。变分参数 $\{\tilde{\eta}, \tilde{\rho}^2, \tilde{\boldsymbol{\omega}}^{\boldsymbol{2}}, \tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\psi}}, \tilde{\boldsymbol{a}}, \tilde{\boldsymbol{b}}, \tilde{\boldsymbol{\delta}}\}$ 的初始˜ ˜　　˜ ˜值设定为与相应的先验设定 $\{\eta, \rho^2, \boldsymbol{\omega}^{\boldsymbol{2}}, \boldsymbol{\xi}, \boldsymbol{\psi}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\delta}\}$一致，而初始 $\{\tilde{\boldsymbol{\sigma}}^{\boldsymbol{2}}, \tilde{\boldsymbol{\varphi}}^{\boldsymbol{2}}\}$的各元素设为默认值 1。对于 $i = 1$，2，.....，N，在给定˜所有其他初始参数的条件下，通过应用算法 2 的步骤 3 来初始化变分参数 $\boldsymbol{\Pi}_i = \{\tilde{\pi}_{i1}, \tilde{\pi}_{i2}, \ldots, \tilde{\pi}_{i\bar{K}}\}$。

6 模拟研究

我们提出了两项模拟研究，以探究新提出的 MLPCM 以及变分贝叶斯推断方法的性能。在本节中，除了 MLPCM 外，我们还采用另一种潜在位置模型（LPM）：Lu 等（2026）提出的 PoisLPCM。该模型既用作比较工具，也可作为生成数据的备选模型。为了更公平地比较 MLPCM 和 PoisLPCM 的估计算法，我们为 PoisLPCM 推导了一种变分贝叶斯方法，以取代 Lu 等（2026）采用的马尔可夫链蒙特卡罗方法。PoisLPCM 的这一备选估计框架见附录 A.5。

第一个模拟研究设计了两种不同的情景：场景1中的合成网络由一个包含多个组的 MLPCM 随机生成，这些组共享相同的精度值 γ_j，对于每个 $j = 1$，2 、 ..... 、 N。相比之下，场景2中的网络由 PoisLPCM 随机生成，其模型设置与场景1中的设置类似，但不依赖混合潜在位置框架。第二个模拟研究则聚焦于一种更具挑战性的情况：我们从一个 MLPCM 随机生成一个人工网络，其中一些组的重叠程度各不相同，因此更难以将它们区分开来。此外，离散度参数 $\boldsymbol{\gamma}$ 设定为各个簇的值均不相同。

我们利用命题 6 和 9 中提出的新 PICL 准则，选择最佳的 K，并与用于生成这些人工网络的真实簇数 K^* 进行比较。本文中，默认情况下，所有变分贝叶斯推断的容差水平 `tol` 均设为 0.01。

6.1 模拟研究1

在所有模拟研究中，我们关注由 $N = 100$ 个节点组成的网络，并假设这些节点按分块顺序从节点1到节点100被均匀聚类为 $K^* = 4$ 个组。因此，真实聚类 $\boldsymbol{z}^*$ 对应的组大小为： $n_1 = n_2 = n_3 = n_4 = 25$。场景1中的合成网络是根据一个MLPCM随机模拟生成的，其设置为：$d = 2, \beta^* = 1, \boldsymbol{\tau}^* = (8, 6, 4, 2)^T$，并且对每个 $j = 1, 2, \ldots, N$均取 $\gamma_j^* = 10$。组中心设定为：

$$\boldsymbol{\mu}^* = \left[\begin{pmatrix} 1.25 \\ 1.25 \end{pmatrix}, \begin{pmatrix} 1.25 \\ -1.25 \end{pmatrix}, \begin{pmatrix} -1.25 \\ -1.25 \end{pmatrix}, \begin{pmatrix} -1.25 \\ 1.25 \end{pmatrix} \right].$$

从此以后，我们用$(\cdot)^*$ 表示用于生成网络的模型参数或变量的参考值，以便与我们从变分贝叶斯算法输出中得到的点估计进行比较。

同样，对于场景2，我们创建一个与上述设置相同的PoisLPCM网络，但不包含混合潜在位置的精度 $\boldsymbol{\gamma}^*$，因为该模型不包括这一特征。情景1和情景2的潜在空间表示与邻接矩阵分别如图 2 和图 3 所示。由MLPCM模拟得到的网络通常比在相似设置下由PoisLPCM模型模拟得到的网络更为稀疏。这是因为显在潜在位置是作为受扰动的隐性位置创建的，因此它们增加了额外一层离散度，进而转化为节点之间更大的距离。尽管并非一目了然，

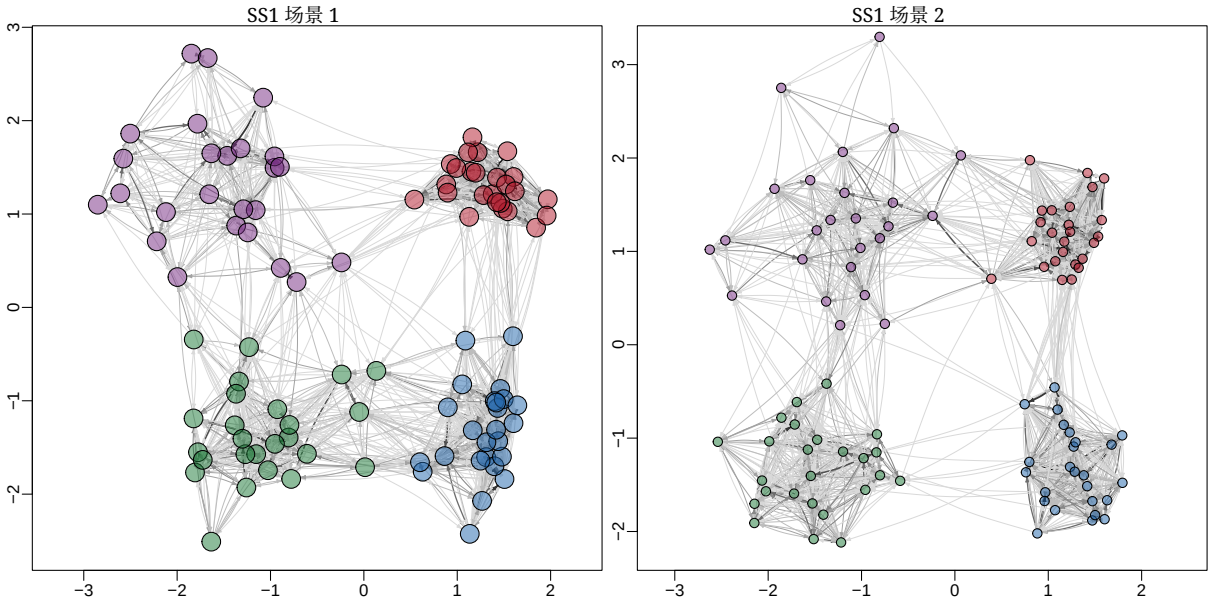

图 2：模拟研究1的合成网络。节点位置对应于用于模拟网络的潜在位置 $\boldsymbol{U}^*$。左图：通过 MLPCM 模拟的场景1，其中节点大小与 $\{1/\gamma_j^*\}$成正比。右图：通过 PoisLPCM 模拟的场景2，其中节点大小保持不变。边宽度和颜色与边权重成正比。

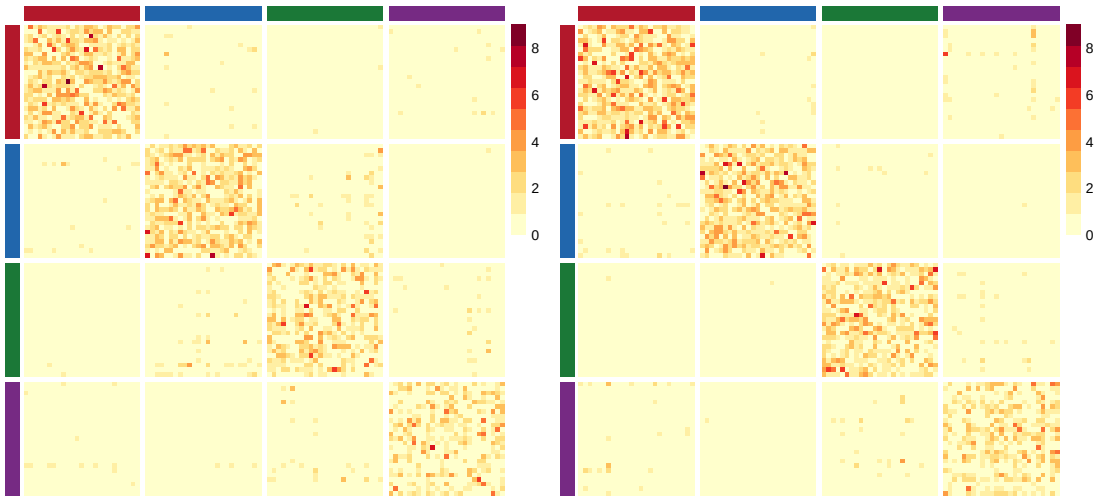

图 3：模拟研究1的邻接矩阵。颜色较深的方格对应较高的边权重。用于分隔区块的间隙以及侧边栏颜色展示了真实聚类 $\boldsymbol{z}^*$。左图：通过 MLPCM 模拟的场景1。右图：通过 PoisLPCM 模拟的场景2。

这一点实际上反映在图 2 和图 3 中：PoisLPCM网络的平均权重为 0.3734，而 MLPCM 的平均权重为 0.3403。

对于两种情景下变分贝叶斯算法的簇数候选值，我们考虑 $K = 2, 3, 4, 5, 6$。我们使用模型参数近似后验均值作为点估计。表1ˆ ˆ中显示的 PICL 值，是针对情景1和情景2中的每个模型，在每个固定的 K 下使用点估计 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 计算得到的。ˆ$\boldsymbol{z}$ 的估计对应于最大后验簇

表1：模拟研究1。组数从 $K = 2$ 到$K = 6$ 时的 PICL准则值。每行中的粗体值表示最高的 PICL值。

	K=2	K=3	K=4	K=5	K=6
场景1 MLPCM	13,457	14,705	**14,724**	14,722	14,720
场景1 PoisLPCM	-4197.2	-4161.3	**-4136.8**	-4141.1	-4146.5
场景2 MLPCM	14,750	14,782	**14,810**	14,806	14,802
场景2 PoisLPCM	-4452.6	-4167.5	**-4134.8**	-4140.3	-4226.2

成员归属。表1显示，PICL 准则在两种情景下都能成功恢复 MLPCM 和 PoisLPCM 的正确簇数$K^* = 4$ 。因此，从这里开始，我们只关注算法运行结果中满足 $K = 4$的输出，并在表2和图4中展示模型拟合结果。为了

表2：模拟研究1。MLPCM 和 PoisLPCM 在情景1和情景2中的表现。汇总统计量为：(i) $(\tilde{\eta},\ \tilde{\rho}^2)$：从 q_β 推断得到的变分参数；(ii) $\mathbb{E}(\{|\hat{d}_{ij}^2 - d_{ij}^{*2}|\})$[标准差]：成对距离平方的平均绝对误差及相应的标准差；(iii) $\mathbb{E}(\{\tilde{\sigma}_i^2\})$[标准差]：与隐性位置相关的变分方差的均值和标准差；(iv) $\mathbb{E}(\{\tilde{\varphi}_{ij}^2\})$[标准差]：与显性位置相关的变分方差的均值˜和标准差；(v) $\mathbb{E}(\{b_j/\tilde{a}_j\})$[标准差]：近似值˜的均值和标准差 $\{b_j/\tilde{a}_j\}$，对应于 $\{1/\gamma_j\}$；(vi) VI(ˆ$\boldsymbol{z}$，$\boldsymbol{z}^*$)：真实聚类与估计聚类之间的 VI距离。

	$(\tilde{\eta}, \tilde{\rho}^2)$	$\mathbb{E}(\{\lvert\hat{d}_{ij}^2 - d_{ij}^{*2}\rvert\})$[sd]	$\mathbb{E}(\{\tilde{\sigma}_i^2\})$[sd]	$\mathbb{E}(\{\tilde{\varphi}_{ij}\})$[sd]	$\mathbb{E}(\{\tilde{b}_j/\tilde{a}_j\})$[sd]	VI(ˆ$\boldsymbol{z}$，$\boldsymbol{z}^*$)
情景 1 MLPCM	(0.9655, 0.0003)	1.5760[1.6027]	0.0010[0.0001]	0.1011[0.0132]	0.1041[0.0094]	0.1217
情景 1 MLPCM tol=0.001	(0.9681, 0.0003)	1.5128[1.5261]	0.0010[0.0001]	0.1006[0.0132]	0.1036[0.0094]	0.1217
情景 1 PoisLPCM	(0.8211，0.0005)	1.8047[1.9058]	0.0143[0.0048]	–	–	0.1217
情景 2 MLPCM	(1.0505, 0.0003)	1.5332[1.7009]	0.0010[0.0001]	0.0996[0.0137]	0.1025[0.0104]	0.0000
情景 2 MLPCM a=20	(0.9800, 0.0003)	0.9939[1.1074]	0.0005[0.0001]	0.0507[0.0033]	0.0515[0.0017]	0.0000
情景 2 PoisLPCM	(0.9212, 0.0003)	1.0420[2.1206]	0.0126[0.0031]	–	–	0.0000

为了比较通过两种不同模型得到的估计潜在空间，我们采用 Procrustes变换（Hoff 等（2002）），使两种解与真实值相匹配。在两种情景中，MLPCM 和 PoisLPCM 在聚类和推断潜在位置 $\boldsymbol{U}$方面的表现相当ˆ，且都与用于生成网络的参考值吻合良好。

表2展示了两种情景和模型的各项汇总结果，包括截距参数的估计值、潜在成对距离估计的误差，以及 Wade 和 Ghahramani（2018）提出的用作真实聚类划分与估计聚类划分之间距离的变异信息度（VI）。总体而言，结果表明模型拟合效果相对较好，但无论数据是根据 LPCM（场景1）生成，还是根据 PoisLPCM（场景2）生成，MLPCM 的表现往往都更好。具体而言，表中显示 MLPCM

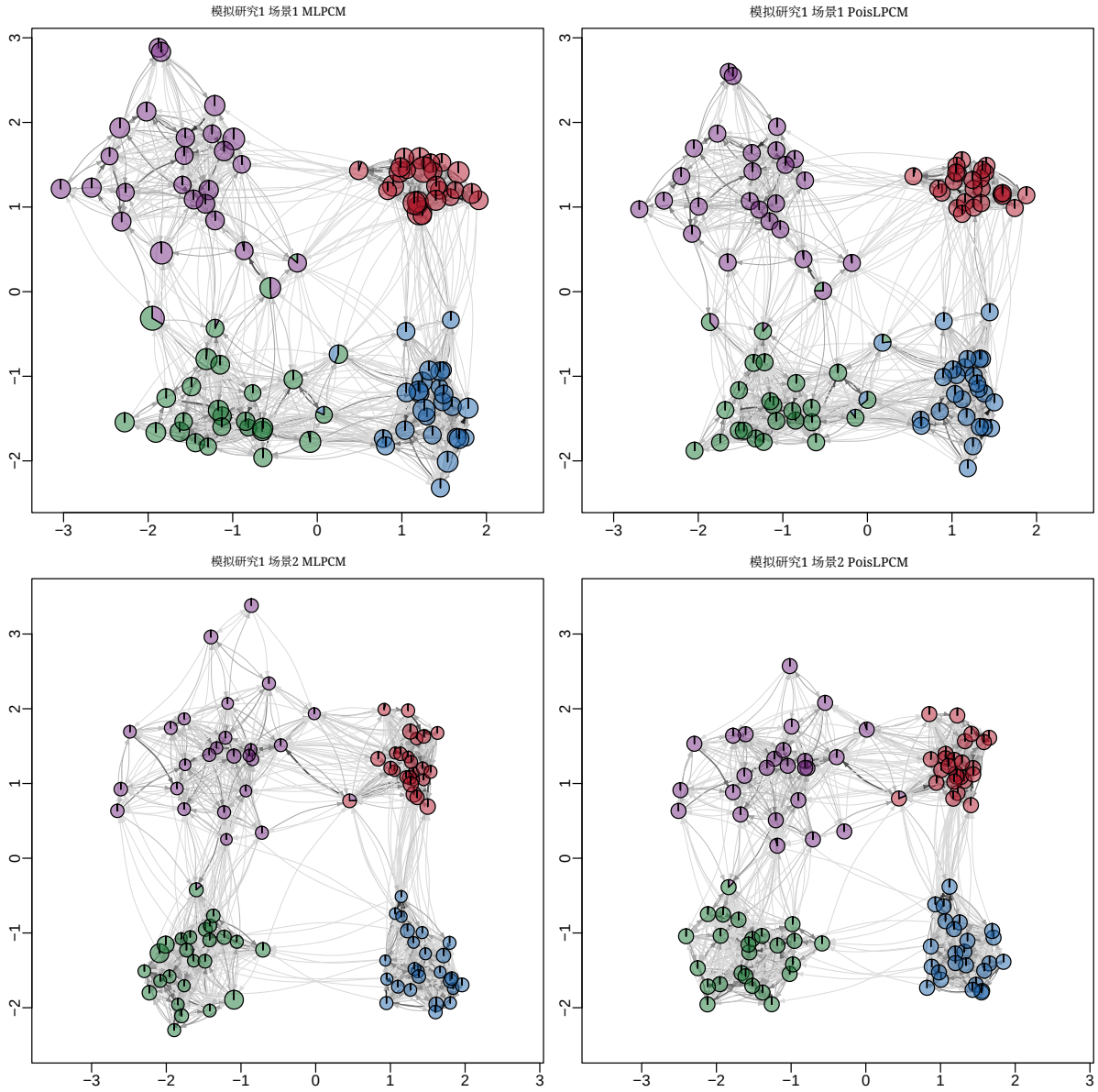

图4：模拟研究1。MLPCM（左）和 PoisLPCM（右）为场景1（上）和场景2（下）推断出的潜在空间。节点的位置根据 $\hat{\boldsymbol{U}}$确定，每个节点的饼图展示相应的群组分配变分概率推断值。对于 MLPCM$\tilde{}$图，节点大小与推断出的 $\{1/\hat{\gamma}_j = b_j/\tilde{a}_j\}$ 成正比，该值近似表示显性位置的离散度。PoisLPCM 的节点大小与一个常数成正比。边宽度和颜色与边权重成正比。

ˆ 对 β 给出了显著更好的估计，这一点可以通过将其变分参数 ˆη 与真实值 1 进行比较来确定。作为敏感性测试，我们还拟合了 MLPCM，将收敛容差从 $\texttt{tol} = 0.01$ 降至 $\texttt{tol} = 0.001$。这一 ˆ 表明，β 的估计还可以进一步改进，但在配备英特尔酷睿 i7 1.80GHz 中央处理器的笔记本电脑上，计算时间约会从 15 分钟增加到 25 分钟。我们针对参数 a 进行了第二项敏感性测试，在场景2的 MLPCM 中试用了数值 $a = 20$。这使显性位置能够更分散地分布在隐性位置周围。关于表2中该情形的结果，模型拟合有了明显改善，尤其是变分方差参数 $\tilde{\boldsymbol{\sigma}}^2$ 和 $\tilde{\boldsymbol{\varphi}}$ 更小。这表明潜在位置周围的后验不确定性大幅降低，说明拟合效果更好。该表结果显示，尽管 PoisLPCM 无法充分刻画生成的 MLPCM 的特征，但 MLPCM 方法足够灵活，能够刻画生成的 PoisLPCM 的所有特征。最后，我们注意到，MLPCM 和 PoisLPCM 对其他变分参数的推断表现相近，包括 $\boldsymbol{\delta}$，$\tilde{\boldsymbol{\omega}}^2$，$\boldsymbol{\xi}$，$\boldsymbol{\psi}$；在这两种情景下，这些参数的点估计通常都与相应的参考值吻合良好。

我们通过在 $K = 4$ 情况下运行 MLPCM 算法 10 次稳健性实验，结束本次模拟研究。10 次运行均采用框架的默认设置，但对各参数施加了微小扰动。输出结果见表 3。这些结果表明，结果是

表 3：模拟研究1。场景1中 MLPCM 在 10 次独立运行中的表现，随机初始化围绕默认设置进行。给出了多次实现的输出所对应汇总统计量的最大值、中位数、最小值和标准差。

	$(\tilde{\eta},\ \tilde{\rho}^2)$	$\mathbb{E}(\{\lvert\hat{d}^2_{ij} - d^{*2}_{ij}\rvert\})_{[\text{sd}]}$	$\mathbb{E}(\{\tilde{\sigma}^2_i\})_{[\text{sd}]}$	$\mathbb{E}(\{\tilde{\varphi}_{ij}\})_{[\text{sd}]}$	$\mathbb{E}(\{\tilde{b}_j/\tilde{a}_j\})_{[\text{sd}]}$	$\text{VI}(\hat{\boldsymbol{z}},\ \boldsymbol{z}^*)$
最大值	(1.0800, 0.0011)	$2.8866_{[2.9869]}$	$0.0012_{[0.0005]}$	$0.1277_{[0.0709]}$	$0.1347_{[0.0753]}$	0.2429
中位数	(0.8782, 0.0003)	$1.5985_{[1.6549]}$	$0.0010_{[0.0001]}$	$0.1012_{[0.0145]}$	$0.1042_{[0.0113]}$	0.1217
最小值	(0.8209, 0.0003)	$1.5191_{[1.5531]}$	$0.0010_{[0.0001]}$	$0.0996_{[0.0126]}$	$0.1024_{[0.0086]}$	0.1217
标准差	(0.0916, 0.0003)	$0.4587_{[0.4734]}$	$0.0001_{[0.0001]}$	$0.0088_{[0.0189]}$	$0.0102_{[0.0215]}$	0.0404

f在收敛时，各次运行的误差得分相近，因此具有相当的稳健性 .

6.2 模拟研究2

在本次模拟研究中，我们考虑一种更具挑战性的情形：不同组的混合潜在位置方差不再相等，并且四个组中有两个组的位置彼此显著更近。第二次模拟研究中的合成网络由一个 MLPCM 随机生成，其大多数模型设置与模拟研究1中的设置相似。主要区别如下：（i）每个簇中有五个节点的显性位置分散程度显著更大，也就是说，我们为这些特殊节点设定 $\gamma^*_j = 1$，而其余节点设定为 $\gamma^*_j = 10$；（ii）簇均值设定如下：

$$\boldsymbol{\mu}^* = \left[\begin{pmatrix}1.5\\1.5\end{pmatrix}, \begin{pmatrix}1.5\\-1.5\end{pmatrix}, \begin{pmatrix}-1.5\\-1.0\end{pmatrix}, \begin{pmatrix}-1.5\\1.0\end{pmatrix}\right],$$

随着组标签从 1 变为 4，隐性位置围绕其簇中心的离散度逐渐增大。随机模拟的网络如图 5 所示，

使用不同的颜色和节点大小来表示簇级别和节点级别的离散度。

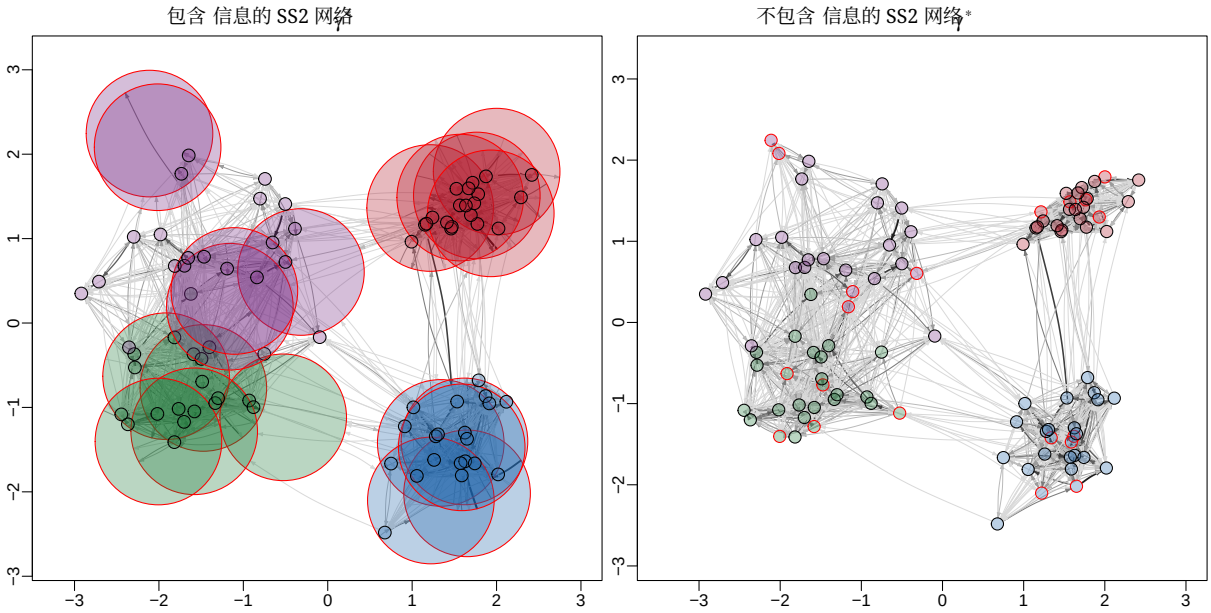

图 5：模拟研究2。节点位置对应于用于模拟网络的隐蔽潜在位置 $\boldsymbol{U}^*$ 。边宽度和颜色与边权重成正比。红色轮廓的节点，其显性位置的离散度较大。左图：节点大小与 $\{1/\gamma_j^*\}$ 成正比。右图：节点大小固定，以使图示更加清晰。

采用与第 6.1 节默认设置相同的实现设置，MLPCM 和 PoisLPCM 的候选值分别为 K $\{2,3,4,5,6\}$。模型选择结果见表 4。对于两者，PICL 均倾向于 $K = 3$

表 4：模拟研究2。组数从 $K = 2$ 到 $K = 6$ 时的 PICL 准则值。每行中的粗体数值表示最高的 PICL 值。

	K=2	K=3	K=4	K=5	K=6
MLPCM	13,709	**13,766**	**13,765**	13,735	13,738
PoisLPCM	-4575.0	**-4518.5**	**-4518.4**	-4524.0	-4529.7

然而，模型的 PICL 与 $K = 4$ 的差异很小。图6）展示了结果的图示，从中可以看出，两个几乎重叠的组在推断出的潜在空间中合并了。本图左侧的图展示了 $K = 3$ 和 $K = 4$的 MLPCM 结果。两种解几乎完全相同，区别仅在于两个重叠簇合并在了一起。图中以不同的轮廓颜色标出了那些被选中、以便更清楚地对比显性位置与隐性位置的特殊节点。对估计值 $\hat{\gamma}$进行排序后，我们发现，按 $\hat{\gamma}$排名前20的节点中，有17个属于显性离散度较高的特殊节点集合。模型能够很好地识别这些节点，但对隐性位置周围离散度的估计偏低。这或许与变分贝叶斯在近似后验方差方面的一些已知局限有关（Blei等人（2017））。至于 PoisLPCM 的结果，我们注意到，特殊节点的潜在位置明显

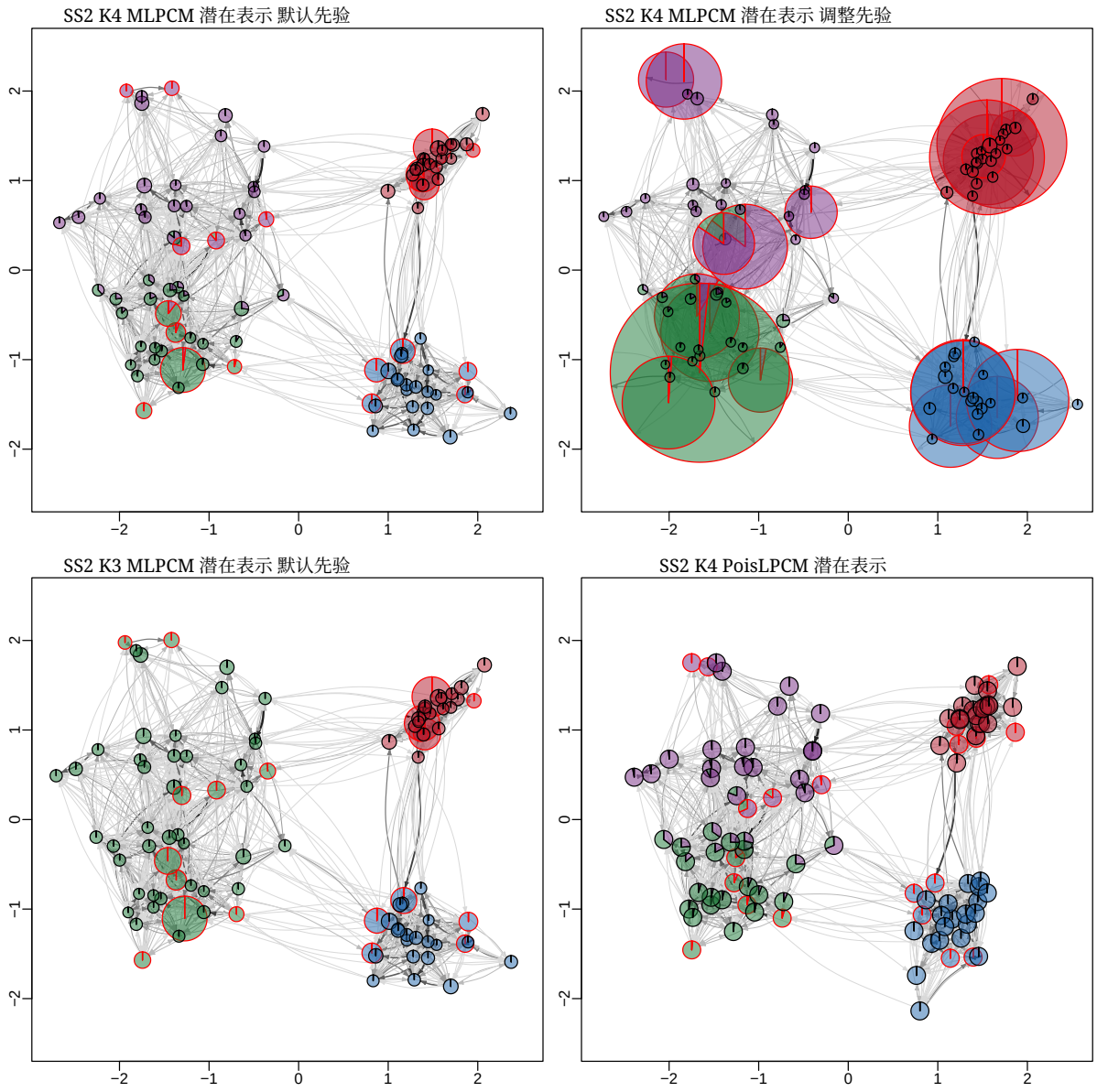

图6：模拟研究2。不同模型和设置下推断得到的潜在空间表示。节点的大小与其显性位置的离散度成正比。节点饼图展示群组分配的变分概率推断值。根据真实值，节点的轮廓颜色表示哪些节点的显性位置离散度较大。边宽度和颜色与边权重成正比。左上：采用 $K = 4$ 和默认先验分布的 MLPCM。右上：采用 $K = 4$ 和信息性先验的 MLPCM。左下：采用 $K = 3$ 和默认先验分布的 MLPCM。右下：PoisLPCM（$K = 4$）。

与 MLPCM 相比，PoisLPCM 的估计结果有所不同。这表明，数据集的复杂性使 PoisLPCM 的潜在空间估计产生了偏差。

举例来说，我们在一个信息更充分的设定下额外运行一次 MLPCM，并利用真实数据向模型提供一些额外信息。具体而言，我们调整 Ga(a, b) 先验，使其与 γ^* 提供的信息相匹配。由于通常无法获得这些信息，这种做法并不现实；但我们采用这一设定，演示不同的先验如何影响结果。图6显示，在采用这一信息更充分的先验后，我们得到的潜在空间估计往往相同，但显性位置的离散度估计更加准确。在这种情况下，模型能够清楚地识别具有不同发送者和接收者特征的节点，并能更准确地量化其特征中的非对称特征。我们可以得出结论：LPCM 能够以很高的准确度有效恢复潜在空间，通常也能识别具有非对称社会特征的节点。然而，与节点相关的显性位置能否得到准确估计，很大程度上取决于其先验分布，因此，从业者必须考虑各种先验设定及其对结果的影响。

为完整起见，我们在表5中展示了不同模型和先验分布的若干误差度量。我们强调，与模拟研究1相比，

表5：模拟研究2。各种算法和设置的性能。汇总统计量与表2中所用的相同

	$(\tilde{\eta},\ \tilde{\rho}^2)$	$\mathbb{E}(\{\lvert\hat{d}^2_{ij} - d^{*2}_{ij}\rvert\})_{[\text{sd}]}$	$\mathbb{E}(\{\tilde{\sigma}^2_i\})_{[\text{sd}]}$	$\mathbb{E}(\{\tilde{\varphi}_{ij}\})_{[\text{sd}]}$	$\mathbb{E}(\{\tilde{b}_j/\tilde{a}_j\})_{[\text{sd}]}$	$\text{VI}(\tilde{z},\ z^*)$
MLPCM K=4 默认先验	(0.7613, 0.0003)	$2.9426_{[3.5967]}$	$0.0011_{[0.0004]}$	$0.1161_{[0.0509]}$	$0.1190_{[0.0479]}$	0.2423
MLPCM K=4 调整先验	(0.8123, 0.0003)	$2.6054_{[3.1895]}$	$0.0021_{[0.0022]}$	$0.2509_{[0.3449]}$	$0.2630_{[0.3496]}$	0.2423
MLPCM K=3 默认先验	(0.7613, 0.0003)	$2.9393_{[3.5898]}$	$0.0011_{[0.0004]}$	$0.1158_{[0.0499]}$	$0.1186_{[0.0468]}$	0.5000
PoisLPCM K=4默认先验	(0.5965, 0.0012)	$3.5545_{[4.0102]}$	$0.0148_{[0.0052]}$	–	–	0.2423

PoisLPCM 的表现明显差于 MLPCM。这是意料之中的，因为当交互模式极不对称时，PoisLPCM 更难拟合网络。其表现不佳体现在表6.2所示的所有统计量均较差。

7 真实数据应用

作为真实数据应用，我们聚焦于斯德哥尔摩国际和平研究所（SIPRI）[†]公开提供的国际武器转让数据库。该数据库涵盖了从1950年至最近一个完整日历年度，各国、各地区、各次区域或非国家行为体之间的所有主要常规武器转让。我们聚焦于该数据集的一个子集，即美国在2024年的自我中心网络。在不失一般性的前提下，我们定义：如果国家 i从国家 j订购武器转让，则存在一条从国家 i 到国家 j 的有向加权边；其交互权重对应于2024年此类武器转让的规模，以百万SIPRI趋势指标值（TIVs）记录。换言之，边的方向遵循订购方向，因此与武器的流动方向相反。

[†]https://www.sipri.org/databases/armstransfers

SIPRITIV是斯德哥尔摩国际和平研究所开发的一种新型通用单位，旨在规范不同武器交付数据，并识别此类武器转让的总体趋势[‡]。

我们首先根据最新的斯德哥尔摩国际和平研究所军费数据库[§]列出 2024 年的所有国家。其中，163 个国家在 2001 年至 2023 年间有军费支出记录；在这 163 个国家中，有 110 个国家在 2024 年有武器转让记录。为了获取美国的自我中心网络，我们提取涉及美国的交互数据。随后，我们移除美国以及所有仅与美国进行武器转让的国家。剩余网络包含美国的邻居彼此之间的所有交互，不包括美国和任何孤立节点。这实际上对应于从美国视角所观察到的网络。该网络由 58 个节点组成，相应的交互矩阵如图7左侧的图所示。请注意，TIV 的原始记录值要么是非负整数，要么是 NULL，

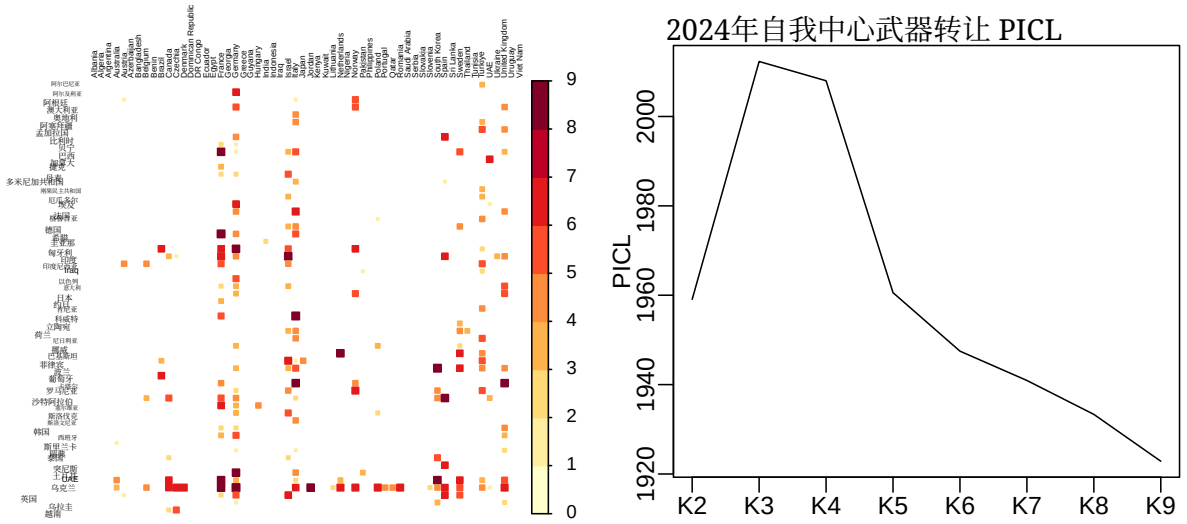

图7：2024年自我中心武器转让真实数据应用。左：观测到的包含58个国家的自我中心网络的交互矩阵 $\boldsymbol{Y}$ 。条目颜色越深，对应的转让量越大。该矩阵显示所下订单的价值，即武器从列节点流向行节点。右：使用MLPCM针对 $K = 2$，3，……，9得到的PICL值。

其中， NULL 对应于未观察到武器转让的情况。该网络非常稀疏，邻接矩阵中有 94.22% 的元素为 NULLs，其余 191 个交互的离散权重范围为 0 到 716。将 NULLs 视为真实零值（以区别于记录的 TIV 为零百万），所得网络交互权重分布极度右偏。因此，在实践中，我们通过对所有非 NULLs 的记录值应用变换 $f(x) = \lfloor \log_2(x+1) + 0.5 \rfloor$ 来处理原始数据，其中公式 $g(s) = \lfloor s + 0.5 \rfloor$ 表示一般的舍入方法，即将数值 s 舍入到最接近的整数。此变换有助于缩小不同交互权重之间的巨大差异，同时保留数据的总体模式。对于对应于零武器转让的那些NULLs，我们赋予其真实零值 $\log_2(1) = 0$ ，并将其纳入

[‡]更多详情见 https://www.sipri.org/databases/armstransfers/sources-and-methods[§]https://www.sipri.org/databases/milex

变换后的数据集。因此，处理后的数据仅包含非负整数，如图7左图所示。

为了将 MLPCM 拟合到这个真实的自我中心式武器转移网络，我们采用所提出的变分贝叶斯算法 2。在测试了各种超参数配置后，我们采用第5.3节提出的默认设置，但先验参数 $\{\boldsymbol{\omega^2}, \boldsymbol{\xi}, \boldsymbol{\delta}\}$ 设为 3，而不是 1。这进一步促进了潜在空间中的组内凝聚性，并使不同组之间更加分离。我们将 $K = 2$、3、.....、9 作为组数的候选值，并得到右图图7所示的 PICL结果。PICL准则选择$K = 3$ 作为最佳模型，相应的武器转移网络推断潜在空间可视化见图8。结果相当有力地支持该网络中存在三个不同的组，图8下方的图给出了推断聚类的点估计。蓝色组中有多个位置显著居中的国家，包括法国、德国、以色列、意大利、瑞典、波兰、英国、乌克兰和荷兰。其中一些节点具有显著的显性离散度，例如法国、德国、意大利、以色列和英国。较高的离散度意味着这些国家被认为具有多种不同特征，因此能够向潜在空间中的不同区域输送武器。换言之，这些国家的影响范围更广，连接模式多样，且连接模式更不对称。乌克兰在蓝色簇中也处于相当中心的位置，但显性离散度较低。这是意料之中的，因为该国主要是武器进口国，而这一点并未反映在潜在位置的离散度中。

在整个网络中，影响力最大的国家是土耳其，它似乎位于红色组的中心位置。由于离散度较高，土耳其向几乎所有红色组成员提供武器。不过，根据图8上方的图，土耳其仍有不可忽视的概率与蓝色组中的那些中心国家归为一簇。绿色组规模较小，联系也较少：所有绿色国家都有较大可能被归入蓝色组或红色组。挪威是这个绿色组的重要中心，其离散度相对较高，也更倾向于接收边。

在附录A.8中，我们还提供了 $K = 5$模型推断出的潜在空间可视化，因为它展现了一些有趣的模式。网络的整体面貌与图8中 $K = 3$ 所示的相似，但蓝色组进一步分为三个子组，其中中心节点构成一个枢纽集群。由此形成的网络呈现出典型的核心-边缘结构，最活跃的行动者都聚集在一起。

8 结论

本文提出了一种新的混合潜在位置聚类模型，该模型非常适用于具有加权边的有向社交网络。该模型能够识别表现出多种不同连接模式的节点，并能更好地处理有向网络中常见的不对称性。在聚类框架内，我们证明了变分贝叶斯程序能够对节点划分和潜在空间给出准确估计。我们的模型扩展了混合成员

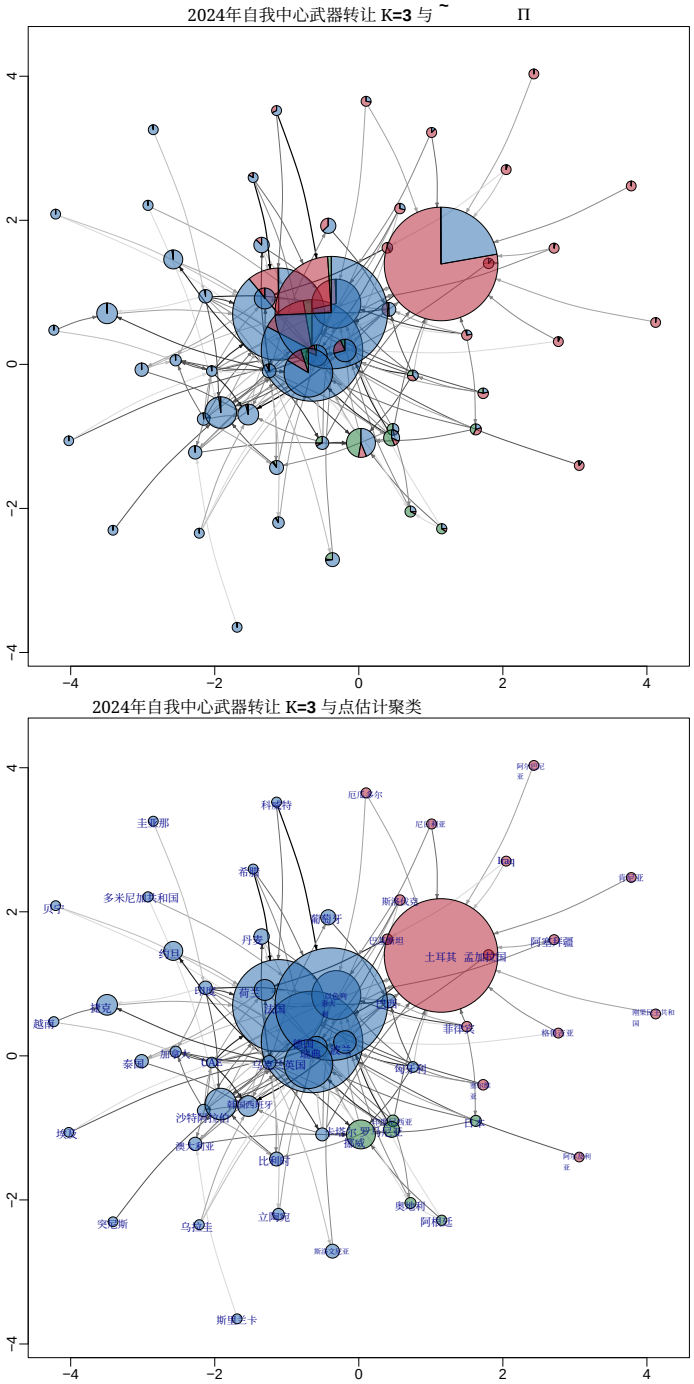

图8：2024年自我中心武器转让真实数据应用：针对 $K = 3$推断的潜在空间。对于两幅图，节点位置对应于$\hat{\boldsymbol{U}}$，节点大小与推断的 $\{1/\hat{\gamma}_j = \tilde{b}_j/\tilde{a}_j\}$成正比，而边宽度和颜色与边权重成正比。上图：每个节点饼图展示相应的推断群组分配变分概率。下图：节点颜色对应于最大后验聚类；每个节点标注相应的国家名称。

这一概念广泛应用于随机区块模型，现已延伸至潜在位置网络建模。这种方法模拟了随机效应框架，同时提供了更清晰的图形表示和解释。通过模拟数据和真实数据，我们展示了采用该方法获得的多种结果，评估了其恢复良好解的准确性，以及其对不同初始设置和先验分布的稳健性。

在后续工作方面，这项研究为未来的多个方向打开了可能性。就建模而言，可以考虑向模型中加入层次，使隐性位置和显性位置都由参数分布刻画。这样可以提高模型的灵活性，因为节点可能呈现更广泛的社会连接模式。在推断方面，LPM 通常依赖于从后验分布中进行马尔可夫链蒙特卡罗抽样。这可以成为我们变分贝叶斯方法的一个完全可行的替代方案，从而更详细地刻画模型的后验分布，但代价可能是算法速度较慢。在计算方面，一个有趣的方向是考虑能够自动选择簇数的方法，例如 Nobile 和 Fearnside （ 2007 年）、 Miller 和 Harrison （ 2018 年）以及 Legramanti 等人（2022年）所采用的方法。这样可以显著提升速度，因为变分贝叶斯不必针对组数的每个候选值分别运行。最后，还可以将该模型自然地扩展到时间网络数据或多视图网络数据中（Sarkar 和 Moore（2005年）、Sewell 和 Chen（2016年）以及 Kim 等（2018年）），使隐性位置和显性位置能够随时间或网络层次发生变化。

资助声明

爱尔兰洞察数据分析研究中心由爱尔兰科学基金会根据资助编号12/RC/2289 P2提供支持。

-

利益冲突

作者声明不存在利益冲突。

代码与数据

MLPCM软件可从我们的GitHub代码库 `https://github.com/Chaoyi-Lu/Mixed-Latent-Position-Cluster-Model`公开获取。应用于武器转让网络的数据可从 `https://www.sipri.org/databases/armstransfers`公开获取。

参考文献

Airoldi, E. M., D. Blei、S. Fienberg 和 E. Xing（2008年）。“混合成员随机区块模型”。载于：神经信息处理系统进展 21。Akaike, H.（1974年）。“对统计模型识别的新认识”。载于：*IEEE*自动控制汇刊 19.6，第716–723页。

Amari, S.-I.（1998）。“自然梯度在学习中的高效作用”。载于：*Neural Computation* 10.2，第251–276页。Biernacki, C.、G. Celeux 和 G. Govaert（2002）。“利用积分完备似然评估用于聚类的混合模型”。载于：*IEEE*模式分析与机器智能汇刊 22.7，第719–725页。—（2010）。“潜在类别模型的积分似然的精确计算与蒙特卡罗计算”。载于：统计规划与推断期刊 140.11，第2991–3002页。Blei, D. M.、A. Kucukelbir 和 J. D. McAuliffe（2017）。“变分推断：面向统计学家的综述”。载于：美国统计学会会刊 112.518，第859–877页。Cˆome, E. 和 P. Latouche（2015）。“基于精确积分完备数据似然的随机区块模型选择与聚类”。载于：*Statistical Modelling* 15.6，第564–589页。Gollini, I. 和 T. B. Murphy（2016）。“多个网络视图的联合建模”。载于：计算与图形统计学期刊 25.1，第246–265页。Gower, J. C.（1966年）。“多变量分析中所用潜根与向量方法的若干距离性质”。载于：*Biometrika* 53.3-4，第325–338页。Handcock, M. S.、A. E. Raftery 和 J. M. Tantrum（2007年）。“社交网络的基于模型的聚类”。载于：皇家统计学会会刊*A*辑：社会统计 170.2，第301–354页。Hoff, P. D.（2005年）。“二元组数据的双线性混合效应模型”。载于：美国统计学会会刊 100.469，第286–295页。Hoff, P. D.、A. E. Raftery 和 M. S. Handcock（2002）。“社交网络分析的潜在空间方法”。载于：美国统计学会会刊 97.460，第1090–1098页。Jiang, C.、D. La Vecchia 和 R. Rastelli（2024年）。“GLAMLE：存在潜变量时多视图网络数据的推断及其在大宗商品交易中的应用”。载于：计量经济学与统计学。Jordan, M. I.、Z. Ghahramani、T. S. Jaakkola 和 L. K. Saul（1999）。“图模型变分方法简介”。载于：*Machine Learning* 37，第183–233页。Kaur, H.、R. Rastelli 和 N. Friel（2023年）。“潜在位置网络模型”。载于：《*Sage*社交网络分析手册》。SAGE Publications。第36章，第526–541页。ISBN：9781529614671。Kim, B.、K. H. Lee、L. Xue 和 X. Niu（2018年）。“带潜变量的动态网络模型综述”。载于：*Statistics Surveys* 12，第105页。Krivitsky, P. N.、M. S. Handcock、A. E. Raftery 和 P. D. Hoff（2009）。“利用潜在聚类随机效应模型表示社交网络中的度分布、聚类和同质性”。载于：*Social Networks* 31.3，第204–213页。Latouche, P.、E. Birmele 和 C. Ambroise（2012年）。“随机区块模型的变分贝叶斯推断与复杂度控制”。载于：*Statistical Modelling* 12.1，第93–115页。Legramanti, S.、T. Rigon、D. Durante 和 D. B. Dunson（2022年）。“扩展随机区块模型及其在犯罪网络中的应用”。载于：应用统计学年刊 16.4，第2369页。

Lu, C., D. Durante 和 N. Friel（2025年3月）。“加权犯罪网络中效率—安全权衡的零膨胀随机区块建模”。载于：皇家统计学会会刊*A*辑：社会统计，qnaf029。Lu, C., R. Rastelli 和 N. Friel（2026）。“零膨胀泊松潜在位置聚类模型”。载于：网络科学，e2（即将发表）。Martens, J.（2020）。“关于自然梯度法的新见解与新视角”。载于：机器学习研究杂志 21.146，第1–76页。Miller, J. W. 和 M. T. Harrison（2018）。“对成分数量设定先验的混合模型”。载于：美国统计学会会刊 113.521，第340–356页。Nobile, A. 和 A. T. Fearnside（2007）。“成分数量未知的贝叶斯有限混合模型：分配采样器”。载于：统计学与计算 17，第147–162页。Rastelli, R.（2018）。“非负加权网络的稀疏潜在位置模型”。载于：*arXiv*预印本 *arXiv:1808.09262*。Rastelli, R. 和 M. Corneli （ 2023 ）。“ 瞬时交互的连续潜在位置模型 ”。载于：网络科学 11.4，第560–588页。Rastelli, R., F. Maire 和 N. Friel（2024）。“潜在位置网络模型的高效计算推断”。载于：电子统计学杂志 18.1，第2531–2570页。Ryan, C., J. Wyse 和 N. Friel（2017）。“社交网络潜在位置聚类模型的贝叶斯模型选择”。载于：网络科学 5.1，第70–91页。Salter-Townshend, M. 和 T. B. Murphy（2013）。“针对网络数据潜在位置聚类模型的变分贝叶斯推断”。载于：计算统计学与数据分析 57.1，第661–671页。Sarkar, P. 和 A. W. Moore（2005）。“使用潜在空间模型进行动态社会网络分析”。载于：*ACM SIGKDD*探索通讯 7.2，第31–40页。Schwarz, G.（1978）。“估计模型的维数”。载于：统计学年刊，第461–464页。Sewell, D. K. 和 Y. Chen（2015）。“动态网络的潜在空间模型”。载于：美国统计学会会刊 110.512，第1646–1657页。—（2016）。“具有加权边的动态网络潜在空间模型”。载于：社交网络 44，第105–116页。Wade, S. 和 Z. Ghahramani（2018）。“贝叶斯聚类分析：点估计与可信球（附讨论）”。载于。Wyse, J. 和 N. Friel（2012）。“采用折叠潜在区块模型进行区块聚类”。载于：统计学与计算 22，第415–428页。

A 附录

A.1 MLPCM 的证据下界推导

˜请注意，在 MLPCM 下，ELBO（ $\mathcal{F}$（$\boldsymbol{\Theta}$））可以分解为若干项，并且这些项可以分别推导：

$$
\begin{aligned}
\mathcal{F}(\tilde{\boldsymbol{\Theta}}) &= \mathbb{E}_q[\log p(\boldsymbol{Y}|\boldsymbol{U},\boldsymbol{V},\beta)] + \mathbb{E}_q[\log p(\boldsymbol{U}|\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{z})] + \mathbb{E}_q[\log p(\boldsymbol{V}|\boldsymbol{U},\boldsymbol{\gamma})] + \mathbb{E}_q[\log p(\boldsymbol{z}|\boldsymbol{\Pi})] \\
&+ \mathbb{E}_q[\log \pi(\beta)] + \mathbb{E}_q[\log \pi(\boldsymbol{\mu})] + \mathbb{E}_q[\log \pi(\boldsymbol{\tau})] + \mathbb{E}_q[\log \pi(\boldsymbol{\gamma})] + \mathbb{E}_q[\log \pi(\boldsymbol{\Pi})] \\
&- \mathbb{E}_q[\log q_{\boldsymbol{U}}(\boldsymbol{U}|\tilde{\boldsymbol{U}},\tilde{\boldsymbol{\sigma}}^{\boldsymbol{2}})] - \mathbb{E}_q[\log q_{\boldsymbol{V}}(\boldsymbol{V}|\tilde{\boldsymbol{V}},\tilde{\boldsymbol{\varphi}}^{\boldsymbol{2}})] - \mathbb{E}_q[\log q_{\boldsymbol{z}}(\boldsymbol{z}|\tilde{\boldsymbol{\Pi}})] - \mathbb{E}_q[\log q_{\beta}(\beta|\tilde{\eta},\tilde{\rho}^2)] \\
&- \mathbb{E}_q[\log q_{\boldsymbol{\mu}}(\boldsymbol{\mu}|\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\omega}}^{\boldsymbol{2}})] - \mathbb{E}_q[\log q_{\boldsymbol{\tau}}(\boldsymbol{\tau}|\tilde{\boldsymbol{\xi}},\tilde{\boldsymbol{\psi}})] - \mathbb{E}_q[\log q_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}|\tilde{\boldsymbol{a}},\tilde{\boldsymbol{b}})] - \mathbb{E}_q[\log q_{\boldsymbol{\Pi}}(\boldsymbol{\Pi}|\tilde{\boldsymbol{\delta}})].
\end{aligned}
$$

在以下各节中，我们将逐项推导上述证据下界项。

A.1.1 计算 $\mathbb{E}_q$[对数 $p(\boldsymbol{Y}|\boldsymbol{U},\ \boldsymbol{V},\ \beta)$]

$$
\begin{aligned}
\mathbb{E}_q[\log p(\boldsymbol{Y}|\boldsymbol{U},\boldsymbol{V},\beta)] &= \mathbb{E}_q\left[\log \prod_{\substack{i,j=1,\\ i\neq j}}^{N} f_{\text{Pois}}\left(y_{ij}\middle|\exp\left(\beta - ||\boldsymbol{u}_i - \boldsymbol{v}_{j\leftarrow i}||^2\right)\right)\right] \\
&= \mathbb{E}_q\left[\log \prod_{\substack{i,j=1,\\ i\neq j}}^{N} \frac{\left[\exp\left(\beta - ||\boldsymbol{u}_i - \boldsymbol{v}_{j\leftarrow i}||^2\right)\right]^{y_{ij}} \exp\left[-\exp\left(\beta - ||\boldsymbol{u}_i - \boldsymbol{v}_{j\leftarrow i}||^2\right)\right]}{y_{ij}!}\right] \\
&= \mathbb{E}_q\left[\sum_{\substack{i,j=1,\\ i\neq j}}^{N} \left[y_{ij}\left(\beta - ||\boldsymbol{u}_i - \boldsymbol{v}_{j\leftarrow i}||^2\right) - \exp\left(\beta - ||\boldsymbol{u}_i - \boldsymbol{v}_{j\leftarrow i}||^2\right) - \log(y_{ij}!)\right]\right] \\
&= \sum_{\substack{i,j=1,\\ i\neq j}}^{N} \left\{y_{ij}\left[\mathbb{E}_q(\beta) - \mathbb{E}_q\left(||\boldsymbol{u}_i - \boldsymbol{v}_{j\leftarrow i}||^2\right)\right] - \mathbb{E}_q\left[\exp\left(\beta - ||\boldsymbol{u}_i - \boldsymbol{v}_{j\leftarrow i}||^2\right)\right]\right\} - \sum_{\substack{i,j=1,\\ i\neq j}}^{N} \log(y_{ij}!) \\
&= \sum_{\substack{i,j=1,\\ i\neq j}}^{N} \left\{y_{ij}\left[\tilde{\eta} - ||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j\leftarrow i}||^2 - d\left(\tilde{\sigma}_i^2 + \tilde{\varphi}_{ij}^2\right)\right] - \frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j\leftarrow i}||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2)^{d/2}}\right\} + \texttt{const}.
\end{aligned}
$$

□

这里，const 表示常数项。上述三个期望的精确解析解均可推导得到。第一个期望 $\mathbb{E}_q(\beta)$ 可以很容易地得到，即 $\tilde{\eta}$。接下来给出 $\mathbb{E}_q\left(||\boldsymbol{u}_i - \boldsymbol{v}_{j\leftarrow i}||^2\right)$ 和 $\mathbb{E}_q\left[\exp\left(\beta - ||\boldsymbol{u}_i - \boldsymbol{v}_{j\leftarrow i}||^2\right)\right]$ 的详细证明：

- 这里我们首先证明$\mathbb{E}_q\left(||\boldsymbol{u}_i - \boldsymbol{v}_{j\leftarrow i}||^2\right)$的一般形式。*** 求 $\mathbb{E}(||\boldsymbol{x}||^2)$，其中 $\boldsymbol{x}$：**$=(x_1,\ \ldots\ldots,\ x_d)^T \sim$多元正态分布$_d(\boldsymbol{m},\ \boldsymbol{\Sigma})$：则存在一个矩阵$\boldsymbol{A}\in\mathbb{R}^{d\times d'}$，使得 $\boldsymbol{x}=\boldsymbol{A}\boldsymbol{w}+\boldsymbol{m}$，其中 $\boldsymbol{w}$：$=(w_1,\ \ldots\ldots,\ w_{d'})^T \sim$多元正态分布$_{d'}(\boldsymbol{0}_{d'},\ \mathbb{I}_{d'})$是维数为 d' 的标准多元正态分布，该维数不必等于 d。这里矩阵 $\boldsymbol{A}$ 实际上是 $\boldsymbol{\Sigma}$的平方根矩阵，即 $\boldsymbol{\Sigma}=\boldsymbol{A}\boldsymbol{A}^T$。

于是我们有

$$\mathbb{E}(||\boldsymbol{x}||^2) = \mathbb{E}(||\boldsymbol{A}\boldsymbol{w}+\boldsymbol{m}||^2) = \mathbb{E}[(\boldsymbol{A}\boldsymbol{w}+\boldsymbol{m})^T(\boldsymbol{A}\boldsymbol{w}+\boldsymbol{m})]$$
$$= \mathbb{E}(\boldsymbol{w}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{w} + \boldsymbol{w}^T\boldsymbol{A}^T\boldsymbol{m} + \boldsymbol{m}^T\boldsymbol{A}\boldsymbol{w} + \boldsymbol{m}^T\boldsymbol{m}) = \mathbb{E}(\boldsymbol{w}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{w}) + \boldsymbol{m}^T\boldsymbol{m}.$$

这里，由于$\mathbb{E}(\boldsymbol{w}^T) = \boldsymbol{0}_{d'}^T$，我们有$\mathbb{E}(\boldsymbol{w}^T\boldsymbol{A}^T\boldsymbol{m}) = \mathbb{E}(\boldsymbol{w}^T)\boldsymbol{A}^T\boldsymbol{m} = 0$。对$\mathbb{E}(\boldsymbol{m}^T\boldsymbol{A}\boldsymbol{w})$同理。然后我们将剩余的期望项分解为：

$$\mathbb{E}(\boldsymbol{w}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{w}) = \mathbb{E}\left[\sum_{i,j=1}^{d'}(\boldsymbol{A}^T\boldsymbol{A})_{ij}w_iw_j\right] = \sum_{i,j=1}^{d'}(\boldsymbol{A}^T\boldsymbol{A})_{ij}\mathbb{E}(w_iw_j),$$

其中，若 $i \neq j$，则$\mathbb{E}(w_iw_j) = 0$，而$\mathbb{E}(w_i^2) = \mathrm{Var}(w_i) + \mathbb{E}(w_i)^2 = 1$。因此我们有：

$$\mathbb{E}(||\boldsymbol{x}||^2) = \mathbb{E}(\boldsymbol{w}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{w}) + \boldsymbol{m}^T\boldsymbol{m} = \boldsymbol{m}^T\boldsymbol{m} + \sum_{i=1}^{d'}(\boldsymbol{A}^T\boldsymbol{A})_{ii} = ||\boldsymbol{m}||^2 + \mathrm{tr}(\boldsymbol{\Sigma}),$$

前提是 $\sum_{i=1}^{d'}(\boldsymbol{A}^T\boldsymbol{A})_{ii} = \mathrm{tr}(\boldsymbol{A}^T\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{A}\boldsymbol{A}^T) = \mathrm{tr}(\boldsymbol{\Sigma})$。 □

因此，仍需证明在 $q_{\boldsymbol{U}}(\boldsymbol{u}_i|\tilde{\boldsymbol{u}}_i, \tilde{\sigma}_i^2)$ ~多元正态分布$_d(\tilde{\boldsymbol{u}}_i, \tilde{\sigma}_i^2\mathbb{I}_d)$且 $q_{\boldsymbol{V}}(\boldsymbol{v}_{j\leftarrow i}|\tilde{\boldsymbol{v}}_{j\leftarrow i}, \tilde{\varphi}_{ij}^2)$ ~多元正态分布$_d(\tilde{\boldsymbol{v}}_{j\leftarrow i}, \tilde{\varphi}_{ij}^2\mathbb{I}_d)$的条件下，$(\boldsymbol{u}_i - \boldsymbol{v}_{j\leftarrow i})$的分布。我们通过多元正态分布的特征函数来证明这一点。

*** 求（$\boldsymbol{u}+\boldsymbol{v}$）的分布**：这里给出一般情形的证明，假设 $\boldsymbol{u} \sim \mathrm{MVN}_d(\boldsymbol{\mu}_{\boldsymbol{u}}, \boldsymbol{\Sigma}_{\boldsymbol{u}})$ 和 $\boldsymbol{v} \sim \mathrm{MVN}_d(\boldsymbol{\mu}_{\boldsymbol{v}}, \boldsymbol{\Sigma}_{\boldsymbol{v}})$，则

$$\begin{aligned}\boldsymbol{\phi}_{\boldsymbol{u}+\boldsymbol{v}}(\boldsymbol{t}) &= \boldsymbol{\phi}_{\boldsymbol{u}}(\boldsymbol{t})\boldsymbol{\phi}_{\boldsymbol{v}}(\boldsymbol{t}) = \mathbb{E}[\exp\big(i\boldsymbol{t}^T\boldsymbol{u}\big)]\mathbb{E}[\exp\big(i\boldsymbol{t}^T\boldsymbol{v}\big)] \\ &= \exp\big(i\boldsymbol{t}^T\boldsymbol{\mu}_{\boldsymbol{u}} - \boldsymbol{t}^T\boldsymbol{\Sigma}_{\boldsymbol{u}}\boldsymbol{t}\big)\exp\big(i\boldsymbol{t}^T\boldsymbol{\mu}_{\boldsymbol{v}} - \boldsymbol{t}^T\boldsymbol{\Sigma}_{\boldsymbol{v}}\boldsymbol{t}\big) \\ &= \exp\big(i\boldsymbol{t}^T(\boldsymbol{\mu}_{\boldsymbol{u}} + \boldsymbol{\mu}_{\boldsymbol{v}}) - \boldsymbol{t}^T(\boldsymbol{\Sigma}_{\boldsymbol{u}} + \boldsymbol{\Sigma}_{\boldsymbol{v}})\boldsymbol{t}\big),\end{aligned}$$

这是 $\mathrm{MVN}_d(\boldsymbol{\mu}_{\boldsymbol{u}}+\boldsymbol{\mu}_{\boldsymbol{v}}, \boldsymbol{\Sigma}_{\boldsymbol{u}}+\boldsymbol{\Sigma}_{\boldsymbol{v}})$ 的特征函数。因此 $\boldsymbol{u}+\boldsymbol{v} \sim \mathrm{MVN}_d(\boldsymbol{\mu}_{\boldsymbol{u}}+\boldsymbol{\mu}_{\boldsymbol{v}}, \boldsymbol{\Sigma}_{\boldsymbol{u}} + \boldsymbol{\Sigma}_{\boldsymbol{v}})$。考虑到 $-\boldsymbol{v} \sim \mathrm{MVN}_d(-\boldsymbol{\mu}_{\boldsymbol{v}}, \boldsymbol{\Sigma}_{\boldsymbol{v}})$，我们有 $\boldsymbol{u} - \boldsymbol{v} \sim \mathrm{MVN}_d(\boldsymbol{\mu}_{\boldsymbol{u}} - \boldsymbol{\mu}_{\boldsymbol{v}}, \boldsymbol{\Sigma}_{\boldsymbol{u}} + \boldsymbol{\Sigma}_{\boldsymbol{v}})$。□

结合上述两个证明，我们可以得到 $\boldsymbol{u}_i - \boldsymbol{v}_{j\leftarrow i} \sim \mathrm{MVN}_d(\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j\leftarrow i}, \tilde{\sigma}_i^2\mathbb{I}_d + \tilde{\varphi}_{ij}^2\mathbb{I}_d)$，因此 $\mathbb{E}_q\left(||\boldsymbol{u}_i - \boldsymbol{v}_{j\leftarrow i}||^2\right) = ||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j\leftarrow i}||^2 + \mathrm{tr}[(\tilde{\sigma}_i^2 + \tilde{\varphi}_{ij}^2)\mathbb{I}_d] = ||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j\leftarrow i}||^2 + d(\tilde{\sigma}_i^2 + \tilde{\varphi}_{ij}^2)$。□

• 注意，$\mathbb{E}_q\left[\exp\big(\beta - ||\boldsymbol{u}_i - \boldsymbol{v}_{j\leftarrow i}||^2\big)\right] = \mathbb{E}_{q(\beta)}[\exp(\beta)]\mathbb{E}_{q(\boldsymbol{U})\,q(\boldsymbol{V})}\left[\exp\big(-||\boldsymbol{u}_i - \boldsymbol{v}_{j\leftarrow i}||^2\big)\right]$。因此，我们先证明这里的第一个期望。

*** 求** $\mathbb{E}_q[\exp(\beta)]$, **其中** $q_\beta(\beta|\tilde{\eta}, \tilde{\rho}^2)$ ~ 正态分布$(\tilde{\eta}, \tilde{\rho}^2)$:

$$\mathbb{E}_q[\exp(\beta)] = \int_{-\infty}^{\infty}\exp(\beta)\frac{\exp\left[-\frac{(\beta-\tilde{\eta})^2}{2\tilde{\rho}^2}\right]}{\sqrt{2\pi\tilde{\rho}^2}}\mathrm{d}\beta = \int_{-\infty}^{\infty}\frac{\exp\left[-\frac{\beta^2}{2\tilde{\rho}^2} + \frac{\tilde{\eta}\beta}{\tilde{\rho}^2} - \frac{\tilde{\eta}^2}{2\tilde{\rho}^2} + \beta\right]}{\sqrt{2\pi\tilde{\rho}^2}}\mathrm{d}\beta$$
$$= \int_{-\infty}^{\infty}\frac{\exp\left[-\frac{\beta^2}{2\tilde{\rho}^2} + \frac{(\tilde{\eta}+\tilde{\rho}^2)\beta}{\tilde{\rho}^2} - \frac{\tilde{\eta}^2}{2\tilde{\rho}^2}\right]}{\sqrt{2\pi\tilde{\rho}^2}}\mathrm{d}\beta = \int_{-\infty}^{\infty}\frac{\exp\left[-\frac{[\beta-(\tilde{\eta}+\tilde{\rho}^2)]^2}{2\tilde{\rho}^2}\right]}{\sqrt{2\pi\tilde{\rho}^2}}\mathrm{d}\beta\cdot\exp\left[\frac{(\tilde{\eta}+\tilde{\rho}^2)^2}{2\tilde{\rho}^2} - \frac{\tilde{\eta}^2}{2\tilde{\rho}^2}\right]$$
$$= \exp\left[\frac{\tilde{\eta}^2 + 2\tilde{\eta}\tilde{\rho}^2 + \tilde{\rho}^4}{2\tilde{\rho}^2} - \frac{\tilde{\eta}^2}{2\tilde{\rho}^2}\right] = \exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2}\right).$$

□

接下来，我们将证明上面第二个期望的一般形式。

*** 求 $\mathbb{E}[\exp(-||\boldsymbol{x}||^2)]$，其中 $\boldsymbol{x} := (x_1, \dots, x_d)^T \sim$ 多元正态分布$_d(\boldsymbol{m}, \boldsymbol{\Sigma})$:**

$$
\begin{aligned}
&\mathbb{E}[\exp(-||\boldsymbol{x}||^2)] = \mathbb{E}[\exp(-\boldsymbol{x}^T\boldsymbol{x})] = \int_{\mathbb{R}^2} \exp(-\boldsymbol{x}^T\boldsymbol{x}) \frac{\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{m})\right]}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}|}} \mathrm{d}\boldsymbol{x} \\
&= \int_{\mathbb{R}^2} \frac{\exp\left[-\frac{1}{2}(\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} - \boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{m} - \boldsymbol{m}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} + \boldsymbol{m}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{m} + 2\boldsymbol{x}^T\boldsymbol{x})\right]}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}|}} \mathrm{d}\boldsymbol{x} \\
&= \int_{\mathbb{R}^2} \frac{\exp\left[-\frac{1}{2}\left(\boldsymbol{x}^T(\boldsymbol{\Sigma}^{-1} + 2\mathbb{I}_d)\boldsymbol{x} - 2\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{m}\right) - \frac{1}{2}\boldsymbol{m}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{m}\right]}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}|}} \mathrm{d}\boldsymbol{x} \\
&= \int_{\mathbb{R}^2} \frac{\exp\left[-\frac{1}{2}\left(\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}(\mathbb{I}_d + 2\boldsymbol{\Sigma})\boldsymbol{x} - 2\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}(\mathbb{I}_d + 2\boldsymbol{\Sigma})(\mathbb{I}_d + 2\boldsymbol{\Sigma})^{-1}\boldsymbol{m}\right) - \frac{1}{2}\boldsymbol{m}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{m}\right]}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}|}} \mathrm{d}\boldsymbol{x} \\
&= \int_{\mathbb{R}^2} \frac{\exp\left[-\frac{1}{2}(\boldsymbol{x} - (\mathbb{I}_d + 2\boldsymbol{\Sigma})^{-1}\boldsymbol{m})^T\boldsymbol{\Sigma}^{-1}(\mathbb{I}_d + 2\boldsymbol{\Sigma})(\boldsymbol{x} - (\mathbb{I}_d + 2\boldsymbol{\Sigma})^{-1}\boldsymbol{m})\right]}{\sqrt{(2\pi)^d|(\boldsymbol{\Sigma}^{-1}(\mathbb{I}_d + 2\boldsymbol{\Sigma}))^{-1}|}} \mathrm{d}\boldsymbol{x} \\
&\quad \times \frac{\sqrt{|(\boldsymbol{\Sigma}^{-1}(\mathbb{I}_d + 2\boldsymbol{\Sigma}))^{-1}|}}{\sqrt{|\boldsymbol{\Sigma}|}} \exp\left[\frac{1}{2}((\mathbb{I}_d + 2\boldsymbol{\Sigma})^{-1}\boldsymbol{m})^T\boldsymbol{\Sigma}^{-1}(\mathbb{I}_d + 2\boldsymbol{\Sigma})((\mathbb{I}_d + 2\boldsymbol{\Sigma})^{-1}\boldsymbol{m}) - \frac{1}{2}\boldsymbol{m}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{m}\right] \\
&= \frac{\sqrt{|(\mathbb{I}_d + 2\boldsymbol{\Sigma})^{-1}\boldsymbol{\Sigma}|}}{\sqrt{|\boldsymbol{\Sigma}|}} \exp\left[\frac{1}{2}((\mathbb{I}_d + 2\boldsymbol{\Sigma})^{-1}\boldsymbol{m})^T\boldsymbol{\Sigma}^{-1}\boldsymbol{m} - \frac{1}{2}\boldsymbol{m}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{m}\right] \\
&= \frac{\sqrt{|(\mathbb{I}_d + 2\boldsymbol{\Sigma})|^{-1}}\sqrt{|\boldsymbol{\Sigma}|}}{\sqrt{|\boldsymbol{\Sigma}|}} \exp\left[\frac{1}{2}\boldsymbol{m}^T(\mathbb{I}_d + 2\boldsymbol{\Sigma})^{-T}\boldsymbol{\Sigma}^{-1}\boldsymbol{m} - \frac{1}{2}\boldsymbol{m}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{m}\right] \\
&= \frac{\exp\left[-\frac{1}{2}\left[\boldsymbol{m}^T(\mathbb{I}_d - (\mathbb{I}_d + 2\boldsymbol{\Sigma})^{-T})\boldsymbol{\Sigma}^{-1}\boldsymbol{m}\right]\right]}{\sqrt{|\mathbb{I}_d + 2\boldsymbol{\Sigma}|}}.
\end{aligned}
$$

□

当 $\boldsymbol{\Sigma} = c\mathbb{I}_d$ 其中 c 是某个常数时，上述公式可简化为:

$$
\begin{aligned}
&= \frac{\exp\left[-\frac{1}{2}\left[\boldsymbol{m}^T(\mathbb{I}_d - (\mathbb{I}_d + 2c\mathbb{I}_d)^{-T})(c\mathbb{I}_d)^{-1}\boldsymbol{m}\right]\right]}{\sqrt{|\mathbb{I}_d + 2c\mathbb{I}_d|}} = \frac{\exp\left[-\frac{1}{2}\left[\boldsymbol{m}^T\left(\mathbb{I}_d - \frac{1}{1+2c}\mathbb{I}_d\right)\frac{1}{c}\mathbb{I}_d\boldsymbol{m}\right]\right]}{\sqrt{|(1+2c)\mathbb{I}_d|}} \\
&= \frac{\exp\left[-\frac{1}{2}\frac{2c}{1+2c}\frac{1}{c}\boldsymbol{m}^T\boldsymbol{m}\right]}{\sqrt{|(1+2c)\mathbb{I}_d|}} = \frac{\exp\left[-(1+2c)^{-1}||\boldsymbol{m}||^2\right]}{(1+2c)^{\frac{d}{2}}}.
\end{aligned}
$$

请注意，上述结果等价于以下形式:

$$
= \frac{\exp\left[-(1+2c)^{-1}\boldsymbol{m}^T\boldsymbol{m}\right]}{\sqrt{|(1+2c)\mathbb{I}_d|}} = \frac{\exp\left[-\boldsymbol{m}^T(\mathbb{I}_d + 2c\mathbb{I}_d)^{-1}\boldsymbol{m}\right]}{\sqrt{|\mathbb{I}_d + 2c\mathbb{I}_d|}} = \frac{\exp\left[-\boldsymbol{m}^T(\mathbb{I}_d + 2\boldsymbol{\Sigma})^{-1}\boldsymbol{m}\right]}{\sqrt{|\mathbb{I}_d + 2\boldsymbol{\Sigma}|}}.
$$

为了证明$\mathbb{E}_{\tilde{\boldsymbol{v}}_{j\leftarrow i}\ \tilde{\sigma}_i^2+\tilde{\varphi}_{ij}^2\ \mathbb{I}_d}{}^{q\,\boldsymbol{U}\ q\,\boldsymbol{V}}[\quad(-||\boldsymbol{u}_i - \boldsymbol{v}_{j\leftarrow i}||^2)]$　　$\boldsymbol{u}_i - \boldsymbol{v}_{j\leftarrow i} \sim$　　$_d\ \tilde{\boldsymbol{u}}_i - ()()\exp$，此处回顾多元正态分布（,()）。将此类信息纳入公式可得:

$$
\mathbb{E}[\exp(-||\boldsymbol{u}_i - \boldsymbol{v}_{j\leftarrow i}||^2)] = \frac{\exp\left[-(1+2(\tilde{\sigma}_i^2 + \tilde{\varphi}_{ij}^2))^{-1}||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j\leftarrow i}||^2\right]}{(1+2(\tilde{\sigma}_i^2 + \tilde{\varphi}_{ij}^2))^{\frac{d}{2}}}.
$$

□

A.1.2 计算 $\mathbb{E}_q$[对数 $p(\boldsymbol{U}|\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{z})$]

$$
\begin{aligned}
&\mathbb{E}_q[\log p(\boldsymbol{U}|\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{z})] = \mathbb{E}_q\left[\log\prod_{i=1}^{N} f_{\mathrm{MVN}_d}(\boldsymbol{u}_i|\boldsymbol{\mu}_{z_i}, 1/\tau_{z_i}\mathbb{I}_d)\right] \\
&= \mathbb{E}_q\left[\log\prod_{i=1}^{N}\frac{|\tau_{z_i}\mathbb{I}_d|^{\frac{1}{2}}}{(2\pi)^{\frac{d}{2}}}\exp\left[-\frac{1}{2}(\boldsymbol{u}_i-\boldsymbol{\mu}_{z_i})^T\tau_{z_i}\mathbb{I}_d(\boldsymbol{u}_i-\boldsymbol{\mu}_{z_i})\right]\right] \\
&= \mathbb{E}_q\left\{\sum_{i=1}^{N}\left[-\frac{d}{2}\log(2\pi)+\frac{d}{2}\log(\tau_{z_i})-\frac{1}{2}(\boldsymbol{u}_i-\boldsymbol{\mu}_{z_i})^T\tau_{z_i}\mathbb{I}_d(\boldsymbol{u}_i-\boldsymbol{\mu}_{z_i})\right]\right\} \\
&= -\left[\sum_{i=1}^{N}\frac{d}{2}\log(2\pi)\right]+\mathbb{E}_q\left\{\sum_{i=1}^{N}\left[\frac{d}{2}\log(\tau_{z_i})-\frac{1}{2}\tau_{z_i}||\boldsymbol{u}_i-\boldsymbol{\mu}_{z_i}||^2\right]\right\} \\
&\equiv -\left[\sum_{i=1}^{N}\frac{d}{2}\log(2\pi)\right]+\mathbb{E}_q\left\{\sum_{i=1}^{N}\sum_{k=1}^{K}z_{ik}\left[\frac{d}{2}\log(\tau_k)-\frac{1}{2}\tau_k||\boldsymbol{u}_i-\boldsymbol{\mu}_k||^2\right]\right\} \\
&= -\left[\sum_{i=1}^{N}\frac{d}{2}\log(2\pi)\right]+\sum_{i=1}^{N}\sum_{k=1}^{K}\mathbb{E}_q(z_{ik})\left[\frac{d}{2}\mathbb{E}_q[\log(\tau_k)]-\frac{1}{2}\mathbb{E}_q(\tau_k)\mathbb{E}_q\left(||\boldsymbol{u}_i-\boldsymbol{\mu}_k||^2\right)\right] \\
&= -\left[\sum_{i=1}^{N}\frac{d}{2}\log(2\pi)\right]+\sum_{i=1}^{N}\sum_{k=1}^{K}\tilde{\pi}_{ik}\left\{\frac{d}{2}\left[\boldsymbol{\Psi}(\tilde{\xi}_k)-\log\left(\tilde{\psi}_k\right)\right]-\frac{1}{2}\frac{\tilde{\xi}_k}{\tilde{\psi}_k}\left[||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{\mu}}_k||^2+d(\tilde{\sigma}_i^2+\tilde{\omega}_k^2)\right]\right\} \\
&= \mathtt{const}+\sum_{i=1}^{N}\sum_{k=1}^{K}\tilde{\pi}_{ik}\left\{\frac{d}{2}\left[\boldsymbol{\Psi}(\tilde{\xi}_k)-\log\left(\tilde{\psi}_k\right)\right]-\frac{1}{2}\frac{\tilde{\xi}_k}{\tilde{\psi}_k}\left[||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{\mu}}_k||^2+d(\tilde{\sigma}_i^2+\tilde{\omega}_k^2)\right]\right\},
\end{aligned}
$$

其中 $\boldsymbol{\Psi}(\theta)$：$=\frac{\partial}{\partial\theta}$对数$\Gamma(\theta)$，对于某个参数 θ。 □

这里我们使用一个技巧，将上文第 5 行中的所有 z_i 替换为等价的另一种记法 $\boldsymbol{z}_i=(z_{i1},\ldots,z_{iK})^T$，这样公式中的期望就能轻松分离。由于我们有 $q_{\boldsymbol{z}}(z_i|\tilde{\pi}_{i1},\ldots,\tilde{\pi}_{iK})\sim$ 分类分布$(\tilde{\pi}_{i1},\ldots,\tilde{\pi}_{iK})$，这意味着 $q_{\boldsymbol{Z}}(\boldsymbol{z}_i|\tilde{\pi}_{i1},\ldots,\tilde{\pi}_{iK})\sim$多项分布$(1;\tilde{\pi}_{i1},\ldots,\tilde{\pi}_{iK})$，因此上文最后一行中的 $\mathrm{E}_q(z_{ik})=\tilde{\pi}_{ik}$。若 $q_{\boldsymbol{\mu}}(\boldsymbol{\mu}_k|\tilde{\boldsymbol{\mu}}_k,\tilde{\omega}_k^2)\sim$多元正态分布$_d(\tilde{\boldsymbol{\mu}}_k,\tilde{\omega}_k^2\mathbb{I}_d\mathrm{I}_d+\tilde{\omega}_k^2\mathbb{I}_d)$，并结合附录 A.1.1 中的证明，我们有 $(\boldsymbol{u}_i-\boldsymbol{\mu}_k)\sim$多元正态分布$_d(\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{\mu}}_k,\tilde{\sigma}_i^2\mathrm{I}_d+\tilde{\omega}_k^2\mathbb{I}_d)$，因此 $\mathrm{E}_q\left(||\boldsymbol{u}_i-\boldsymbol{\mu}_k||^2\right)=||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{\mu}}_k||^2+d(\tilde{\sigma}_i^2+\tilde{\omega}_k^2)$。由于 $q_{\boldsymbol{\tau}}(\tau_k|\tilde{\xi}_k,\tilde{\psi}_k)\sim$伽马分布$(\tilde{\xi}_k,\tilde{\psi}_k)$，我们有 $\mathrm{E}_q(\tau_k)=\tilde{\xi}_k/\tilde{\psi}_k$。因此，剩下要证明的是下文所示的 $\mathrm{E}_q[$对数$(\tau_k)]$。这里我们证明一般形式：

*** 求**$\mathbb{E}$[对数**(**x**)**]**，其中** $x \sim$ 伽马分布**(**a, b**)**:

$$\mathbb{E}[\log(x)] = \int_0^\infty \log(x) \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx) \mathrm{d}x,$$

let $y = bx$ implying that $\dfrac{\partial y}{\partial x} = b$, and then,

$$= \int_0^\infty \log\left(\frac{y}{b}\right) \frac{b^a}{\Gamma(a)} \left(\frac{y}{b}\right)^{a-1} \exp(-y) \frac{\mathrm{d}y}{b} = \int_0^\infty [\log(y) - \log(b)] \frac{1}{\Gamma(a)} y^{a-1} \exp(-y) \mathrm{d}y$$

$$= \int_0^\infty \frac{1}{\Gamma(a)} \log(y) y^{a-1} \exp(-y) \mathrm{d}y - \int_0^\infty \frac{1}{\Gamma(a)} \log(b) y^{a-1} \exp(-y) \mathrm{d}y$$

$$= \frac{1}{\Gamma(a)} \int_0^\infty \frac{\partial}{\partial a} [y^{a-1} \exp(-y)] \mathrm{d}y - \frac{\log(b)}{\Gamma(a)} \int_0^\infty y^{a-1} \exp(-y) \mathrm{d}y$$

$$= \frac{1}{\Gamma(a)} \frac{\partial}{\partial a} \int_0^\infty y^{a-1} \exp(-y) \mathrm{d}y - \frac{\log(b)}{\Gamma(a)} \Gamma(a) = \frac{\frac{\partial}{\partial a} \Gamma(a)}{\Gamma(a)} - \log(b)$$

$$= \left[\frac{\partial}{\partial a} \log \Gamma(a)\right] - \log(b) = \boldsymbol{\Psi}(a) - \log(b).$$

因此 $\mathbb{E}_q$[对数**(**τ_k**)**] = $\boldsymbol{\Psi}$**(**$\tilde{\xi}_k$**)** − 对数 $\left(\tilde{\psi}_k\right)$。 □

A.1.3 计算 $\mathbb{E}_q$[对数 $p(\boldsymbol{V}|\boldsymbol{U},\ \boldsymbol{\gamma})$]

$$
\begin{aligned}
\mathbb{E}_q[\log p(\boldsymbol{V}|\boldsymbol{U},\boldsymbol{\gamma})] &= \mathbb{E}_q\left[\log \prod_{\substack{i,j=1,\\ i\neq j}}^{N} f_{\mathrm{MVN}_d}(\boldsymbol{v}_{j\leftarrow i}|\boldsymbol{u}_j, 1/\gamma_j \mathbb{I}_d)\right] \\
&= \mathbb{E}_q\left[\log \prod_{\substack{i,j=1,\\ i\neq j}}^{N} \frac{|\gamma_j \mathbb{I}_d|^{\frac{1}{2}}}{(2\pi)^{\frac{d}{2}}} \exp\left[-\frac{1}{2}(\boldsymbol{v}_{j\leftarrow i}-\boldsymbol{u}_j)^T \gamma_j \mathbb{I}_d (\boldsymbol{v}_{j\leftarrow i}-\boldsymbol{u}_j)\right]\right] \\
&= \mathbb{E}_q\left\{\sum_{\substack{i,j=1,\\ i\neq j}}^{N}\left[-\frac{d}{2}\log(2\pi)+\frac{d}{2}\log(\gamma_j)-\frac{1}{2}(\boldsymbol{v}_{j\leftarrow i}-\boldsymbol{u}_j)^T\gamma_j\mathbb{I}_d(\boldsymbol{v}_{j\leftarrow i}-\boldsymbol{u}_j)\right]\right\} \\
&= -\left[\sum_{\substack{i,j=1,\\ i\neq j}}^{N}\frac{d}{2}\log(2\pi)\right] + \mathbb{E}_q\left\{\sum_{i=1}^{N}\left[\frac{d}{2}\log(\gamma_j)-\frac{1}{2}\gamma_j||\boldsymbol{v}_{j\leftarrow i}-\boldsymbol{u}_j||^2\right]\right\} \\
&= -\left[\sum_{\substack{i,j=1,\\ i\neq j}}^{N}\frac{d}{2}\log(2\pi)\right] + \sum_{\substack{i,j=1,\\ i\neq j}}^{N}\left\{\frac{d}{2}\mathbb{E}_q[\log(\gamma_j)]-\frac{1}{2}\mathbb{E}_q(\gamma_j)\mathbb{E}_q(||\boldsymbol{v}_{j\leftarrow i}-\boldsymbol{u}_j||^2)\right\} \\
&= -\left[\sum_{\substack{i,j=1,\\ i\neq j}}^{N}\frac{d}{2}\log(2\pi)\right] + \sum_{\substack{i,j=1,\\ i\neq j}}^{N}\left\{\frac{d}{2}[\boldsymbol{\Psi}(\tilde{a}_j)-\log\left(\tilde{b}_j\right)]-\frac{1}{2}\frac{\tilde{a}_j}{\tilde{b}_j}[||\tilde{\boldsymbol{v}}_{j\leftarrow i}-\tilde{\boldsymbol{u}}_j||^2+d(\tilde{\varphi}_{ij}^2+\tilde{\sigma}_j^2)]\right\} \\
&= \texttt{const} + \sum_{\substack{i,j=1,\\ i\neq j}}^{N}\left\{\frac{d}{2}[\boldsymbol{\Psi}(\tilde{a}_j)-\log\left(\tilde{b}_j\right)]-\frac{1}{2}\frac{\tilde{a}_j}{\tilde{b}_j}[||\tilde{\boldsymbol{v}}_{j\leftarrow i}-\tilde{\boldsymbol{u}}_j||^2+d(\tilde{\varphi}_{ij}^2+\tilde{\sigma}_j^2)]\right\}.
\end{aligned}
$$

□

由于 $q_{\boldsymbol{\gamma}}(\gamma_j|\tilde{a}_j,\tilde{b}_j)$ ∼伽马分布$(\tilde{a}_j,\tilde{b}_j)$，且 $(\boldsymbol{v}_{j\leftarrow i}-\boldsymbol{u}_j)$ ∼多元正态分布$_d(\tilde{\boldsymbol{v}}_{j\leftarrow i}-\tilde{\boldsymbol{u}}_j,(\tilde{\varphi}_{ij}^2+\tilde{\sigma}_j^2)\mathbb{I}_d)$，上述公式的证明与附录 A.1.2 中所示的证明类似。

A.1.4 计算$\mathbb{E}_q$[对数 $p(\boldsymbol{z}|\boldsymbol{\Pi})$]

$$
\begin{aligned}
\mathbb{E}_q[\log p(\boldsymbol{z}|\boldsymbol{\Pi})] &= \mathbb{E}_q\left[\log\prod_{i=1}^{N}\prod_{k=1}^{K} f_{\mathrm{multinomial}}(\boldsymbol{z}_i|1;\boldsymbol{\Pi})\right] = \mathbb{E}_q\left(\log\prod_{i=1}^{N}\prod_{k=1}^{K}\pi_k^{z_{ik}}\right) \\
&= \mathbb{E}_q\left(\sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik}\log(\pi_k)\right) = \sum_{i=1}^{N}\sum_{k=1}^{K}\mathbb{E}_q(z_{ik})\mathbb{E}_q\left[\log(\pi_k)\right] \\
&= \sum_{i=1}^{N}\sum_{k=1}^{K}\tilde{\pi}_{ik}\left[\boldsymbol{\Psi}\left(\tilde{\delta}_k\right)-\boldsymbol{\Psi}\left(\sum_{g=1}^{K}\tilde{\delta}_g\right)\right].
\end{aligned}
$$

□

~ ~ ~ ~假设 $q_{\boldsymbol{\Pi}}(\pi_1,\ldots,\pi_K|\delta_1,\ldots,\delta_K)$ ∼狄利克雷分布$(\delta_1,\ldots,\delta_K)$，我们如下证明 $\mathbb{E}_q$ [对数(π_k)]。我们首先推导狄利克雷分布的边缘分布，即单个 π_k 的分布。不失一般性，我们从考虑 $\sum_{g=1}^{K}\pi_g = 1$ 开始，因此，一旦知道了 $(\pi_1,\ldots,\pi_{K-1})$，变量 $\pi_K = 1-\sum_{g=1}^{K-1}\pi_g$ 便可确定。这意味着狄利克雷分布$(\pi_1,\ldots,\pi_K|\delta_1,\ldots,\delta_K)$ 的概率密度函数（PDF）可写为:

$$f_{\text{Dirichlet}}(\pi_1,\ldots,\pi_{K-1}) \propto (1-\pi_1-\cdots-\pi_{K-1})^{\tilde{\delta}_K-1}\prod_{g=1}^{K-1}\pi_g^{\tilde{\delta}_g-1}.$$

*** 确定$(\pi_1,\ldots,\pi_{K-2})$的分布**：对 π_{K-1} 积分后，$(\pi_1,\ldots,\pi_{K-2})$ 的概率密度函数为:

$$\begin{aligned}
&f(\pi_1,\ldots,\pi_{K-2}) \propto \int_0^{1-\pi_1-\cdots-\pi_{K-2}}(1-\pi_1-\cdots-\pi_{K-2}-\pi_{K-1})^{\tilde{\delta}_K-1}\prod_{g=1}^{K-1}\pi_g^{\tilde{\delta}_g-1}\mathrm{d}\pi_{K-1}\\
&=\prod_{g=1}^{K-2}\pi_g^{\tilde{\delta}_g-1}\int_0^{1-\pi_1-\cdots-\pi_{K-2}}(1-\pi_1-\cdots-\pi_{K-2}-\pi_{K-1})^{\tilde{\delta}_K-1}\pi_{K-1}^{\tilde{\delta}_{K-1}-1}\mathrm{d}\pi_{K-1},\\
&\text{substituting } \pi_{K-1}=(1-\pi_1-\cdots-\pi_{K-2})x \text{ giving } \mathrm{d}\pi_{K-1}=(1-\pi_1-\cdots-\pi_{K-2})\mathrm{d}x, \text{ then,}\\
&=\prod_{g=1}^{K-2}\pi_g^{\tilde{\delta}_g-1}\int_0^1[(1-\pi_1-\cdots-\pi_{K-2})-(1-\pi_1-\cdots-\pi_{K-2})x]^{\tilde{\delta}_K-1}[(1-\pi_1-\cdots-\pi_{K-2})x]^{\tilde{\delta}_{K-1}-1}\\
&\times(1-\pi_1-\cdots-\pi_{K-2})\mathrm{d}x\\
&=\left(\prod_{g=1}^{K-2}\pi_g^{\tilde{\delta}_g-1}\right)(1-\pi_1-\cdots-\pi_{K-2})^{\tilde{\delta}_K+\tilde{\delta}_{K-1}-1}\int_0^1(1-x)^{\tilde{\delta}_K-1}x^{\tilde{\delta}_{K-1}-1}\mathrm{d}x\\
&\propto(1-\pi_1-\cdots-\pi_{K-2})^{\tilde{\delta}_K+\tilde{\delta}_{K-1}-1}\prod_{g=1}^{K-2}\pi_g^{\tilde{\delta}_g-1},
\end{aligned}$$

~ ~ ~ ~服从狄利克雷分布（$\delta_1,\ldots,\delta_{K-2},\delta_{K-1}+\delta_K$）。 □

*** 确定 π_k的分布**：通过反复应用上述方法，积分消去除 π_k 以外的所有变量，最终得到 $\left(\pi_k,\sum_{g=1,\ g\neq k}^{K}\pi_g\right)$ ∼狄利克雷分布 $\left(\tilde{\delta}_k,\sum_{g=1,\ g\neq k}^{K}\tilde{\delta}_g\right)$，这等价于 $\pi_k\sim$贝塔分布 $\left(\tilde{\delta}_k,\sum_{g=1,\ g\neq k}^{K}\tilde{\delta}_g\right)$。□

下一步是证明贝塔分布对数的期望。我们假设其一般形式如下:

*** 求** $\mathbb{E}$ [对数(x)]**,其中** $x\sim\mathrm{Beta}(a,b)$:

$$\begin{aligned}
\mathbb{E}[\log(x)] &= \int_0^1\log(x)\frac{x^{a-1}(1-x)^{b-1}}{\mathrm{B}(a,b)}\mathrm{d}x = \frac{1}{\mathrm{B}(a,b)}\int_0^1\frac{\partial}{\partial a}[x^{a-1}(1-x)^{b-1}]\mathrm{d}x\\
&=\frac{1}{\mathrm{B}(a,b)}\frac{\partial}{\partial a}\int_0^1x^{a-1}(1-x)^{b-1}\mathrm{d}x = \frac{1}{\mathrm{B}(a,b)}\frac{\partial}{\partial a}\mathrm{B}(a,b) = \frac{\partial}{\partial a}\log\mathrm{B}(a,b)\\
&=\frac{\partial}{\partial a}\log\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \frac{\partial}{\partial a}\log\Gamma(a)-\frac{\partial}{\partial a}\log\Gamma(a+b) = \boldsymbol{\Psi}(a)-\boldsymbol{\Psi}(a+b).
\end{aligned}$$

□

结合以上两个证明，我们得到：

$$\mathbb{E}_q\left[\log(\pi_k)\right] = \boldsymbol{\Psi}\left(\tilde{\delta}_k\right) - \boldsymbol{\Psi}\left(\tilde{\delta}_k + \sum_{\substack{g=1,\\ g\neq k}}^{K} \tilde{\delta}_g\right) = \boldsymbol{\Psi}\left(\tilde{\delta}_k\right) - \boldsymbol{\Psi}\left(\sum_{g=1}^{K} \tilde{\delta}_g\right).$$

□

A.1.5 计算 $\mathbb{E}_q$[对数 $\pi(\beta)$]

$$\begin{aligned}
&\mathbb{E}_q[\log \pi(\beta)] = \mathbb{E}_q\left[\log f_{\text{normal}}(\beta|\eta,\rho^2)\right] = \mathbb{E}_q \log\left[\frac{1}{\sqrt{2\pi\rho^2}}\exp\left[-\frac{(\beta-\eta)^2}{2\rho^2}\right]\right] \\
&= \mathbb{E}_q\left[-\frac{1}{2}\log 2\pi - \frac{1}{2}\log\rho^2 - \frac{(\beta-\eta)^2}{2\rho^2}\right] = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\rho^2 - \frac{1}{2\rho^2}\mathbb{E}_q\left[(\beta-\eta)^2\right] \\
&= -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\rho^2 - \frac{1}{2\rho^2}\left\{\left[\mathbb{E}_q(\beta-\eta)\right]^2 + \text{Var}_q(\beta-\eta)\right\} \\
&= -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\rho^2 - \frac{1}{2\rho^2}\left[(\tilde{\eta}-\eta)^2 + \tilde{\rho}^2\right] \\
&= \texttt{const} - \frac{1}{2\rho^2}\left[(\tilde{\eta}-\eta)^2 + \tilde{\rho}^2\right].
\end{aligned}$$

□

A.1.6 计算 $\mathbb{E}_q$[对数 $\pi(\boldsymbol{\mu})$]

$$\begin{aligned}
&\mathbb{E}_q[\log \pi(\boldsymbol{\mu})] = \mathbb{E}_q\left[\log \prod_{k=1}^{K} f_{\text{MVN}_d}(\boldsymbol{\mu}_k|\mathbf{0},\omega^2\mathbb{I}_d)\right] = \mathbb{E}_q\left[\log \prod_{k=1}^{K} \frac{\exp\left(-\frac{\boldsymbol{\mu}_k^T\boldsymbol{\mu}_k}{2\omega^2}\right)}{\sqrt{(2\pi)^d(\omega^2)^d}}\right] \\
&= \mathbb{E}_q\left\{\sum_{k=1}^{K}\left[-\frac{d}{2}\log(2\pi) - \frac{d}{2}\log(\omega^2) - \frac{\boldsymbol{\mu}_k^T\boldsymbol{\mu}_k}{2\omega^2}\right]\right\} = \sum_{k=1}^{K}\left[-\frac{d}{2}\log(2\pi) - \frac{d}{2}\log(\omega^2) - \frac{\mathbb{E}_q\left(\boldsymbol{\mu}_k^T\boldsymbol{\mu}_k\right)}{2\omega^2}\right] \\
&= \sum_{k=1}^{K}\left[-\frac{d}{2}\log(2\pi) - \frac{d}{2}\log(\omega^2)\right] - \sum_{k=1}^{K}\left[\frac{1}{2\omega^2}\left(\tilde{\boldsymbol{\mu}}_k^T\tilde{\boldsymbol{\mu}}_k + d\tilde{\omega}_k^2\right)\right] = \texttt{const} - \sum_{k=1}^{K}\left[\frac{1}{2\omega^2}\left(\tilde{\boldsymbol{\mu}}_k^T\tilde{\boldsymbol{\mu}}_k + d\tilde{\omega}_k^2\right)\right].
\end{aligned}$$

□

A.1.7 计算 $\mathbb{E}_q$[对数 $\pi(\boldsymbol{\tau})$]

$$
\begin{aligned}
&\mathbb{E}_q[\log \pi(\boldsymbol{\tau})] = \mathbb{E}_q\left[\log \prod_{k=1}^{K} f_{\text{gamma}}(\tau_k|\xi,\psi)\right] = \mathbb{E}_q\left[\log \prod_{k=1}^{K} \frac{\psi^{\xi}}{\Gamma(\xi)}\tau_k^{\xi-1}\exp(-\psi\tau_k)\right] \\
&= \mathbb{E}_q\left[\sum_{k=1}^{K}\left[\xi\log\psi - \log\Gamma(\xi) + (\xi-1)\log(\tau_k) - \psi\tau_k\right]\right] \\
&= \sum_{k=1}^{K}\left[\xi\log\psi - \log\Gamma(\xi)\right] + \sum_{k=1}^{K}\left\{(\xi-1)\mathbb{E}_q\left[\log(\tau_k)\right] - \psi\mathbb{E}_q\left(\tau_k\right)\right\} \\
&= \sum_{k=1}^{K}\left[\xi\log\psi - \log\Gamma(\xi)\right] + \sum_{k=1}^{K}\left\{(\xi-1)\left[\boldsymbol{\Psi}\left(\tilde{\xi}_k\right) - \log\tilde{\psi}_k\right] - \psi\frac{\tilde{\xi}_k}{\tilde{\psi}_k}\right\} \\
&= \texttt{const} + \sum_{k=1}^{K}\left\{(\xi-1)\left[\boldsymbol{\Psi}\left(\tilde{\xi}_k\right) - \log\tilde{\psi}_k\right] - \psi\frac{\tilde{\xi}_k}{\tilde{\psi}_k}\right\}.
\end{aligned}
$$

□

A.1.8 计算 $\mathbb{E}_q$[对数 $\pi(\boldsymbol{\gamma})$]

$$
\begin{aligned}
&\mathbb{E}_q[\log \pi(\boldsymbol{\gamma})] = \mathbb{E}_q\left[\log \prod_{i=1}^{N} f_{\text{gamma}}(\gamma_i|a,b)\right] = \mathbb{E}_q\left[\log \prod_{i=1}^{N} \frac{b^{a}}{\Gamma(a)}\gamma_i^{a-1}\exp(-b\gamma_i)\right] \\
&= \mathbb{E}_q\left[\sum_{i=1}^{N}\left[a\log b - \log\Gamma(a) + (a-1)\log(\gamma_i) - b\gamma_i\right]\right] \\
&= \sum_{i=1}^{N}\left[a\log b - \log\Gamma(a)\right] + \sum_{i=1}^{N}\left\{(a-1)\mathbb{E}_q\left[\log(\gamma_i)\right] - b\mathbb{E}_q\left(\gamma_i\right)\right\} \\
&= \sum_{i=1}^{N}\left[a\log b - \log\Gamma(a)\right] + \sum_{i=1}^{N}\left\{(a-1)\left[\boldsymbol{\Psi}\left(\tilde{a}_i\right) - \log\tilde{b}_i\right] - b\frac{\tilde{a}_i}{\tilde{b}_i}\right\} \\
&= \texttt{const} + \sum_{i=1}^{N}\left\{(a-1)\left[\boldsymbol{\Psi}\left(\tilde{a}_i\right) - \log\tilde{b}_i\right] - b\frac{\tilde{a}_i}{\tilde{b}_i}\right\}.
\end{aligned}
$$

□

A.1.9 计算 $\mathbb{E}_q$[对数 $\pi(\mathbf{\Pi})$]

$$\begin{aligned}
&\mathbb{E}_q[\log \pi(\mathbf{\Pi})] = \mathbb{E}_q\left[\log f_{\text{Dirichlet}}(\mathbf{\Pi}|\boldsymbol{\delta})\right] = \mathbb{E}_q \log\left[\frac{\Gamma\left(\sum_{k=1}^K \delta_k\right)}{\prod_{k=1}^K \Gamma(\delta_k)} \prod_{k=1}^K \pi_k^{\delta_k - 1}\right] \\
&= \mathbb{E}_q\left[\log\Gamma\left(\sum_{k=1}^K \delta_k\right) - \sum_{k=1}^K \log\Gamma(\delta_k) + \sum_{k=1}^K (\delta_k - 1)\log(\pi_k)\right] \\
&= \log\Gamma\left(\sum_{k=1}^K \delta_k\right) - \sum_{k=1}^K \log\Gamma(\delta_k) + \sum_{k=1}^K (\delta_k - 1)\mathbb{E}_q\left[\log(\pi_k)\right] \\
&= \log\Gamma\left(\sum_{k=1}^K \delta_k\right) - \sum_{k=1}^K \log\Gamma(\delta_k) + \sum_{k=1}^K (\delta_k - 1)\left[\mathbf{\Psi}\left(\tilde{\delta}_k\right) - \mathbf{\Psi}\left(\sum_{g=1}^K \tilde{\delta}_g\right)\right] \\
&= \texttt{const} + \sum_{k=1}^K (\delta_k - 1)\left[\mathbf{\Psi}\left(\tilde{\delta}_k\right) - \mathbf{\Psi}\left(\sum_{g=1}^K \tilde{\delta}_g\right)\right].
\end{aligned}$$

□

A.1.10 计算 $\mathbb{E}_q$[对数 $q_{\boldsymbol{U}}(\boldsymbol{U}|\tilde{\boldsymbol{U}}, \tilde{\boldsymbol{\sigma}}^{\mathbf{2}})$]

$$\begin{aligned}
&\mathbb{E}_q[\log q_{\boldsymbol{U}}(\boldsymbol{U}|\tilde{\boldsymbol{U}}, \tilde{\boldsymbol{\sigma}}^{\mathbf{2}})] = \mathbb{E}_q\left[\log \prod_{i=1}^N f_{\text{MVN}_d}(\boldsymbol{u}_i|\tilde{\boldsymbol{u}}_i, \tilde{\sigma}_i^2 \mathbb{I}_d)\right] \\
&= \mathbb{E}_q\left[\log \prod_{i=1}^N \frac{1}{(2\pi)^{\frac{d}{2}}|\tilde{\sigma}_i^2 \mathbb{I}_d|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\boldsymbol{u}_i - \tilde{\boldsymbol{u}}_i)^T \frac{1}{\tilde{\sigma}_i^2}\mathbb{I}_d(\boldsymbol{u}_i - \tilde{\boldsymbol{u}}_i)\right]\right] \\
&= \mathbb{E}_q\left\{\sum_{i=1}^N \left[-\frac{d}{2}\log(2\pi) - \frac{d}{2}\log\left(\tilde{\sigma}_i^2\right) - \frac{1}{2\tilde{\sigma}_i^2}(\boldsymbol{u}_i - \tilde{\boldsymbol{u}}_i)^T(\boldsymbol{u}_i - \tilde{\boldsymbol{u}}_i)\right]\right\} \\
&= \sum_{i=1}^N \left[-\frac{d}{2}\log(2\pi) - \frac{d}{2}\log\left(\tilde{\sigma}_i^2\right) - \frac{1}{2\tilde{\sigma}_i^2}\mathbb{E}_q\left(||\boldsymbol{u}_i - \tilde{\boldsymbol{u}}_i||^2\right)\right] \\
&= \sum_{i=1}^N \left[-\frac{d}{2}\log(2\pi) - \frac{d}{2}\log\left(\tilde{\sigma}_i^2\right) - \frac{1}{2\tilde{\sigma}_i^2}\left(||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_i||^2 + d\tilde{\sigma}_i^2\right)\right] \\
&= \sum_{i=1}^N \left[-\frac{d}{2}\log(2\pi) - \frac{d}{2}\log\left(\tilde{\sigma}_i^2\right) - \frac{d}{2}\right] = \sum_{i=1}^N \left[-\frac{d}{2}\log\left(\tilde{\sigma}_i^2\right)\right] - \sum_{i=1}^N \left[\frac{d}{2}\log(2\pi) + \frac{d}{2}\right] \\
&= \sum_{i=1}^N \left[-\frac{d}{2}\log\left(\tilde{\sigma}_i^2\right)\right] - \texttt{const}.
\end{aligned}$$

□

A.1.11 计算 $\mathbb{E}_q$[对数 $q_{\boldsymbol{V}}(\boldsymbol{V}|\tilde{\boldsymbol{V}}, \tilde{\boldsymbol{\varphi}}^{\mathbf{2}})$]

$$\mathbb{E}_q[\log q_{\boldsymbol{V}}(\boldsymbol{V}|\tilde{\boldsymbol{V}}, \tilde{\boldsymbol{\varphi}}^{\mathbf{2}})] = \mathbb{E}_q\left[\log \prod_{\substack{i,j=1,\\ i\neq j}}^{N} f_{\mathrm{MVN}_d}(\boldsymbol{v}_{j\leftarrow i}|\tilde{\boldsymbol{v}}_{j\leftarrow i}, \tilde{\varphi}_{ij}^2\mathbb{I}_d)\right]$$

$$= \mathbb{E}_q\left[\log \prod_{\substack{i,j=1,\\ i\neq j}}^{N} \frac{1}{(2\pi)^{\frac{d}{2}}|\tilde{\varphi}_{ij}^2\mathbb{I}_d|^{\frac{1}{2}}} \exp\left[-\frac{1}{2\tilde{\varphi}_{ij}^2}(\boldsymbol{v}_{j\leftarrow i} - \tilde{\boldsymbol{v}}_{j\leftarrow i})^T(\boldsymbol{v}_{j\leftarrow i} - \tilde{\boldsymbol{v}}_{j\leftarrow i})\right]\right]$$

$$= \sum_{\substack{i,j=1,\\ i\neq j}}^{N}\left[-\frac{d}{2}\log(2\pi) - \frac{d}{2}\log(\tilde{\varphi}_{ij}^2) - \frac{1}{2\tilde{\varphi}_{ij}^2}\mathbb{E}_q\left(||\boldsymbol{v}_{j\leftarrow i} - \tilde{\boldsymbol{v}}_{j\leftarrow i}||^2\right)\right]$$

$$= \sum_{\substack{i,j=1,\\ i\neq j}}^{N}\left[-\frac{d}{2}\log(2\pi) - \frac{d}{2}\log(\tilde{\varphi}_{ij}^2) - \frac{1}{2\tilde{\varphi}_{ij}^2}\left(||\tilde{\boldsymbol{v}}_{j\leftarrow i} - \tilde{\boldsymbol{v}}_{j\leftarrow i}||^2 + d\tilde{\varphi}_{ij}^2\right)\right]$$

$$= \sum_{\substack{i,j=1,\\ i\neq j}}^{N}\left[-\frac{d}{2}\log(2\pi) - \frac{d}{2}\log(\tilde{\varphi}_{ij}^2) - \frac{d}{2}\right] = \sum_{i=1}^{N}\left[-\frac{d}{2}\log(\tilde{\varphi}_{ij}^2)\right] - \sum_{i=1}^{N}\left[\frac{d}{2}\log(2\pi) + \frac{d}{2}\right]$$

$$= \sum_{\substack{i,j=1,\\ i\neq j}}^{N}\left[-\frac{d}{2}\log(\tilde{\varphi}_{ij}^2)\right] - \texttt{const}.$$

□

~A.1.12 计算 $\mathbb{E}_q$[对数 $q_{\boldsymbol{Z}}(\boldsymbol{Z}|\boldsymbol{\Pi})$]

$$\mathbb{E}_q\left[\log q_{\boldsymbol{Z}}(\boldsymbol{Z}|\tilde{\boldsymbol{\Pi}})\right] = \mathbb{E}_q\left[\log \prod_{i=1}^{N}\prod_{k=1}^{K} f_{\mathrm{multinomial}}(\boldsymbol{z}_i|1;\tilde{\boldsymbol{\Pi}})\right] = \mathbb{E}_q\left[\log \prod_{i=1}^{N}\prod_{k=1}^{K} \tilde{\pi}_{ik}^{z_{ik}}\right]$$

$$= \mathbb{E}_q\left[\sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik}\log(\tilde{\pi}_{ik})\right] = \sum_{i=1}^{N}\sum_{k=1}^{K}\mathbb{E}_q(z_{ik})\log(\tilde{\pi}_{ik}) = \sum_{i=1}^{N}\sum_{k=1}^{K}\tilde{\pi}_{ik}\log(\tilde{\pi}_{ik}).$$

□

A.1.13 计算$\mathbb{E}_q$[对数 $q_\beta(\beta|\tilde{\eta},\ \tilde{\rho}^2)$]

$$\mathbb{E}_q\left[\log q_\beta(\beta|\tilde{\eta},\tilde{\rho}^2)\right] = \mathbb{E}_q\left[\log f_{\text{normal}}(\beta|\tilde{\eta},\tilde{\rho}^2)\right] = \mathbb{E}_q\left\{\log\left[\frac{1}{\sqrt{2\pi\tilde{\rho}^2}}\exp\left(-\frac{(\beta-\tilde{\eta})^2}{2\tilde{\rho}^2}\right)\right]\right\}$$
$$= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\tilde{\rho}^2\right) - \frac{1}{2\tilde{\rho}^2}\mathbb{E}_q\left[(\beta-\tilde{\eta})^2\right]$$
$$= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\tilde{\rho}^2\right) - \frac{1}{2\tilde{\rho}^2}\left[\mathbb{E}_q^2(\beta-\tilde{\eta}) + \text{var}_q(\beta-\tilde{\eta})\right]$$
$$= -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\left(\tilde{\rho}^2\right) - \frac{1}{2\tilde{\rho}^2}\left(0+\tilde{\rho}^2\right)$$
$$= -\frac{1}{2}\log(2\pi) - \frac{1}{2} - \frac{1}{2}\log\left(\tilde{\rho}^2\right) = \texttt{const} - \frac{1}{2}\log\left(\tilde{\rho}^2\right).$$

□

A.1.14 计算 $\mathbb{E}_q$[对数 $q_{\boldsymbol{\mu}}(\boldsymbol{\mu}|\tilde{\boldsymbol{\mu}},\ \tilde{\boldsymbol{\omega}}^{\mathbf{2}})$]

$$\mathbb{E}_q[\log q_{\boldsymbol{\mu}}(\boldsymbol{\mu}|\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\omega}}^{\mathbf{2}})] = \mathbb{E}_q\left[\log\prod_{k=1}^{K} f_{\text{MVN}_d}(\boldsymbol{\mu}_k|\tilde{\boldsymbol{\mu}}_k,\tilde{\omega}_k^2\mathbb{I}_d)\right] = \mathbb{E}_q\left[\log\prod_{k=1}^{K}\frac{\exp\left(-\frac{(\boldsymbol{\mu}_k-\tilde{\boldsymbol{\mu}}_k)^T(\boldsymbol{\mu}_k-\tilde{\boldsymbol{\mu}}_k)}{2\tilde{\omega}_k^2}\right)}{\sqrt{(2\pi)^d(\tilde{\omega}_k^2)^d}}\right]$$
$$= \mathbb{E}_q\left\{\sum_{k=1}^{K}\left[-\frac{d}{2}\log(2\pi) - \frac{d}{2}\log\left(\tilde{\omega}_k^2\right) - \frac{(\boldsymbol{\mu}_k-\tilde{\boldsymbol{\mu}}_k)^T(\boldsymbol{\mu}_k-\tilde{\boldsymbol{\mu}}_k)}{2\tilde{\omega}_k^2}\right]\right\}$$
$$= \sum_{k=1}^{K}\left[-\frac{d}{2}\log(2\pi) - \frac{d}{2}\log\left(\tilde{\omega}_k^2\right) - \frac{\mathbb{E}_q\left[(\boldsymbol{\mu}_k-\tilde{\boldsymbol{\mu}}_k)^T(\boldsymbol{\mu}_k-\tilde{\boldsymbol{\mu}}_k)\right]}{2\tilde{\omega}_k^2}\right]$$
$$= \sum_{k=1}^{K}\left[-\frac{d}{2}\log(2\pi) - \frac{d}{2}\log\left(\tilde{\omega}_k^2\right) - \frac{1}{2\tilde{\omega}_k^2}\left[(\tilde{\boldsymbol{\mu}}_k-\tilde{\boldsymbol{\mu}}_k)^T(\tilde{\boldsymbol{\mu}}_k-\tilde{\boldsymbol{\mu}}_k) + d\tilde{\omega}_k^2\right]\right]$$
$$= \sum_{k=1}^{K}\left[-\frac{d}{2}\log\left(\tilde{\omega}_k^2\right)\right] - \sum_{k=1}^{K}\left[\frac{d}{2}\log(2\pi)+\frac{d}{2}\right] = \sum_{k=1}^{K}\left[-\frac{d}{2}\log\left(\tilde{\omega}_k^2\right)\right] - \texttt{const}.$$

□

˜ ˜A.1.15 计算 $\mathbb{E}_q$[对数 $q_{\boldsymbol{\tau}}(\boldsymbol{\tau}|\boldsymbol{\xi},\boldsymbol{\psi})$]

$$\mathbb{E}_q[\log q_{\boldsymbol{\tau}}(\boldsymbol{\tau}|\tilde{\boldsymbol{\xi}},\tilde{\boldsymbol{\psi}})] = \mathbb{E}_q\left[\log\prod_{k=1}^{K} f_{\text{gamma}}(\tau_k|\tilde{\xi}_k,\tilde{\psi}_k)\right] = \mathbb{E}_q\left\{\log\prod_{k=1}^{K}\frac{\tilde{\psi}_k^{\tilde{\xi}_k}}{\Gamma(\tilde{\xi}_k)}\tau_k^{\tilde{\xi}_k-1}\exp\Big(-\tilde{\psi}_k\tau_k\Big)\right]$$

$$= \mathbb{E}_q\left[\sum_{k=1}^{K}\left[\tilde{\xi}_k\log\tilde{\psi}_k - \log\Gamma(\tilde{\xi}_k) + (\tilde{\xi}_k-1)\log(\tau_k) - \tilde{\psi}_k\tau_k\right]\right\}$$

$$= \sum_{k=1}^{K}\left\{\tilde{\xi}_k\log\tilde{\psi}_k - \log\Gamma(\tilde{\xi}_k) + (\tilde{\xi}_k-1)\mathbb{E}_q\left[\log(\tau_k)\right] - \tilde{\psi}_k\mathbb{E}_q\left(\tau_k\right)\right\}$$

$$= \sum_{k=1}^{K}\left\{\tilde{\xi}_k\log\tilde{\psi}_k - \log\Gamma(\tilde{\xi}_k) + (\tilde{\xi}_k-1)\left[\boldsymbol{\Psi}\left(\tilde{\xi}_k\right) - \log\tilde{\psi}_k\right] - \tilde{\psi}_k\frac{\tilde{\xi}_k}{\tilde{\psi}_k}\right\}$$

$$= \sum_{k=1}^{K}\left\{-\log\Gamma(\tilde{\xi}_k) + (\tilde{\xi}_k-1)\boldsymbol{\Psi}\left(\tilde{\xi}_k\right) + \log\tilde{\psi}_k - \tilde{\xi}_k\right\}.$$

□

˜A.1.16 计算 $\mathbb{E}_q$[对数 $q_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}|\tilde{\boldsymbol{a}},\boldsymbol{b})$]

$$\mathbb{E}_q[\log q_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}|\tilde{\boldsymbol{a}},\tilde{\boldsymbol{b}})] = \mathbb{E}_q\left[\log\prod_{i=1}^{N} f_{\text{gamma}}(\gamma_i|\tilde{a}_i,\tilde{b}_i)\right] = \mathbb{E}_q\left[\log\prod_{i=1}^{N}\frac{\tilde{b}_i^{\tilde{a}_i}}{\Gamma(\tilde{a}_i)}\gamma_i^{\tilde{a}_i-1}\exp\Big(-\tilde{b}_i\gamma_i\Big)\right]$$

$$= \sum_{i=1}^{N}\left\{\tilde{a}_i\log\tilde{b}_i - \log\Gamma(\tilde{a}_i) + (\tilde{a}_i-1)\mathbb{E}_q\left[\log(\gamma_i)\right] - \tilde{b}_i\mathbb{E}_q\left(\gamma_i\right)\right\}$$

$$= \sum_{i=1}^{N}\left\{\tilde{a}_i\log\tilde{b}_i - \log\Gamma(\tilde{a}_i) + (\tilde{a}_i-1)\left[\boldsymbol{\Psi}\left(\tilde{a}_i\right) - \log\tilde{b}_i\right] - \tilde{b}_i\frac{\tilde{a}_i}{\tilde{b}_i}\right\}$$

$$= \sum_{i=1}^{N}\left\{-\log\Gamma(\tilde{a}_i) + (\tilde{a}_i-1)\boldsymbol{\Psi}\left(\tilde{a}_i\right) + \log\tilde{b}_i - \tilde{a}_i\right\}.$$

□

˜A.1.17 计算 $\mathbb{E}_q$[对数 $q_{\boldsymbol{\Pi}}(\boldsymbol{\Pi}|\boldsymbol{\delta})$]

$$\mathbb{E}_q\left[\log q_{\boldsymbol{\Pi}}(\boldsymbol{\Pi}|\tilde{\boldsymbol{\delta}})\right] = \mathbb{E}_q\left[\log f_{\text{Dirichlet}}(\boldsymbol{\Pi}|\tilde{\boldsymbol{\delta}})\right] = \mathbb{E}_q\log\left[\frac{\Gamma\left(\sum_{k=1}^{K}\tilde{\delta}_k\right)}{\prod_{k=1}^{K}\Gamma(\tilde{\delta}_k)}\prod_{k=1}^{K}\pi_k^{\tilde{\delta}_k-1}\right]$$

$$= \mathbb{E}_q\left[\log\Gamma\left(\sum_{k=1}^{K}\tilde{\delta}_k\right) - \sum_{k=1}^{K}\log\Gamma(\tilde{\delta}_k) + \sum_{k=1}^{K}(\tilde{\delta}_k-1)\log(\pi_k)\right]$$

$$= \log\Gamma\left(\sum_{k=1}^{K}\tilde{\delta}_k\right) - \sum_{k=1}^{K}\log\Gamma(\tilde{\delta}_k) + \sum_{k=1}^{K}(\tilde{\delta}_k-1)\mathbb{E}_q\left[\log(\pi_k)\right]$$

$$= \log\Gamma\left(\sum_{k=1}^{K}\tilde{\delta}_k\right) - \sum_{k=1}^{K}\log\Gamma(\tilde{\delta}_k) + \sum_{k=1}^{K}(\tilde{\delta}_k-1)\left[\boldsymbol{\Psi}\left(\tilde{\delta}_k\right) - \boldsymbol{\Psi}\left(\sum_{g=1}^{K}\tilde{\delta}_g\right)\right].$$

□

A.1.18 证据下界的最终公式

综合以上我们为证据下界计算得到的各项，最终可得：

$$
\begin{aligned}
\mathcal{F}(\tilde{\boldsymbol{\Theta}}) &= \mathbb{E}_q[\log p(\boldsymbol{Y}|\boldsymbol{U},\boldsymbol{V},\beta)] \\
&\quad + \mathbb{E}_q[\log p(\boldsymbol{U}|\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{z})] - \mathbb{E}_q[\log q_{\boldsymbol{U}}(\boldsymbol{U}|\tilde{\boldsymbol{U}},\tilde{\boldsymbol{\sigma}}^{\boldsymbol{2}})] \\
&\quad + \mathbb{E}_q[\log p(\boldsymbol{V}|\boldsymbol{U},\boldsymbol{\gamma})] - \mathbb{E}_q[\log q_{\boldsymbol{V}}(\boldsymbol{V}|\tilde{\boldsymbol{V}},\tilde{\boldsymbol{\varphi}}^{\boldsymbol{2}})] \\
&\quad + \mathbb{E}_q[\log p(\boldsymbol{z}|\boldsymbol{\Pi})] - \mathbb{E}_q[\log q_{\boldsymbol{z}}(\boldsymbol{z}|\tilde{\boldsymbol{\Pi}})] \\
&\quad + \mathbb{E}_q[\log \pi(\beta)] - \mathbb{E}_q[\log q_\beta(\beta|\tilde{\eta},\tilde{\rho}^2)] \\
&\quad + \mathbb{E}_q[\log \pi(\boldsymbol{\mu})] - \mathbb{E}_q[\log q_{\boldsymbol{\mu}}(\boldsymbol{\mu}|\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\omega}}^{\boldsymbol{2}})] \\
&\quad + \mathbb{E}_q[\log \pi(\boldsymbol{\tau})] - \mathbb{E}_q[\log q_{\boldsymbol{\tau}}(\boldsymbol{\tau}|\tilde{\boldsymbol{\xi}},\tilde{\boldsymbol{\psi}})] \\
&\quad + \mathbb{E}_q[\log \pi(\boldsymbol{\gamma})] - \mathbb{E}_q[\log q_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}|\tilde{\boldsymbol{a}},\tilde{\boldsymbol{b}})] \\
&\quad + \mathbb{E}_q[\log \pi(\boldsymbol{\Pi})] - \mathbb{E}_q[\log q_{\boldsymbol{\Pi}}(\boldsymbol{\Pi}|\tilde{\boldsymbol{\delta}})] \\
&= \sum_{\substack{i,j=1,\\ i\neq j}}^{N} \left\{ y_{ij}\left[\tilde{\eta} - ||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j\leftarrow i}||^2 - d\left(\tilde{\sigma}_i^2 + \tilde{\varphi}_{ij}^2\right)\right] - \frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j\leftarrow i}||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2)^{d/2}} \right\} \\
&\quad + \sum_{\substack{i,j=1,\\ i\neq j}}^{N} \left\{ \frac{d}{2}\left[\boldsymbol{\Psi}(\tilde{a}_j) - \log\left(\tilde{b}_j\right)\right] - \frac{1}{2}\frac{\tilde{a}_j}{\tilde{b}_j}[||\tilde{\boldsymbol{v}}_{j\leftarrow i} - \tilde{\boldsymbol{u}}_j||^2 + d(\tilde{\varphi}_{ij}^2 + \tilde{\sigma}_j^2)] + \frac{d}{2}\log\left(\tilde{\varphi}_{ij}^2\right) \right\} \\
&\quad + \sum_{i=1}^{N}\sum_{k=1}^{K} \tilde{\pi}_{ik} \left\{ \frac{d}{2}\left[\boldsymbol{\Psi}(\tilde{\xi}_k) - \log\left(\tilde{\psi}_k\right)\right] - \frac{1}{2}\frac{\tilde{\xi}_k}{\tilde{\psi}_k}\left[||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{\mu}}_k||^2 + d(\tilde{\sigma}_i^2 + \tilde{\omega}_k^2)\right] \right\} \\
&\quad + \sum_{i=1}^{N}\sum_{k=1}^{K} \tilde{\pi}_{ik} \left[\boldsymbol{\Psi}\left(\tilde{\delta}_k\right) - \boldsymbol{\Psi}\left(\sum_{g=1}^{K}\tilde{\delta}_g\right) - \log(\tilde{\pi}_{ik}) \right] \\
&\quad + \sum_{k=1}^{K} \left\{ (\delta_k - \tilde{\delta}_k)\left[\boldsymbol{\Psi}\left(\tilde{\delta}_k\right) - \boldsymbol{\Psi}\left(\sum_{g=1}^{K}\tilde{\delta}_g\right) \right] + \log\Gamma(\tilde{\delta}_k) - \frac{1}{2\omega^2}\left(\tilde{\boldsymbol{\mu}}_k^T\tilde{\boldsymbol{\mu}}_k + d\tilde{\omega}_k^2\right) + \frac{d}{2}\log\left(\tilde{\omega}_k^2\right) \right\} \\
&\quad + \sum_{k=1}^{K} \left\{ (\xi - \tilde{\xi}_k)\boldsymbol{\Psi}\left(\tilde{\xi}_k\right) - \xi\log\tilde{\psi}_k - \psi\frac{\tilde{\xi}_k}{\tilde{\psi}_k} + \log\Gamma(\tilde{\xi}_k) + \tilde{\xi}_k \right\} \\
&\quad + \sum_{i=1}^{N} \left\{ (a - \tilde{a}_i)\boldsymbol{\Psi}\left(\tilde{a}_i\right) - a\log\tilde{b}_i - b\frac{\tilde{a}_i}{\tilde{b}_i} + \log\Gamma(\tilde{a}_i) + \tilde{a}_i + \frac{d}{2}\log\left(\tilde{\sigma}_i^2\right) \right\} \\
&\quad - \frac{1}{2\rho^2}\left[(\tilde{\eta} - \eta)^2 + \tilde{\rho}^2\right] + \frac{1}{2}\log\left(\tilde{\rho}^2\right) - \log\Gamma\left(\sum_{k=1}^{K}\tilde{\delta}_k\right) + \texttt{const}.
\end{aligned}
$$

□

A.2 MLPCM 的证据下界最大化

A.2.1 命题 3 的证明

对于 $j=1$，2、.. . 、N以及 $k=1$，2、.. . 、K，证据下界关于 $\tilde{\pi}_{jk}$ 的一阶偏导数为:

$$\frac{\partial \mathcal{F}}{\partial \tilde{\pi}_{jk}}=\frac{\partial}{\partial \tilde{\pi}_{jk}}\left\{\tilde{\pi}_{jk}\left[\frac{d}{2}\left[\boldsymbol{\Psi}(\tilde{\xi}_k)-\log\left(\tilde{\psi}_k\right)\right]-\frac{1}{2}\frac{\tilde{\xi}_k}{\tilde{\psi}_k}\left[||\tilde{\boldsymbol{u}}_j-\tilde{\boldsymbol{\mu}}_k||^2+d(\tilde{\sigma}_j^2+\tilde{\omega}_k^2)\right]+\boldsymbol{\Psi}\left(\tilde{\delta}_k\right)-\boldsymbol{\Psi}\left(\sum_{g=1}^{K}\tilde{\delta}_g\right)\right]\right\}$$
$$-\frac{\partial}{\partial \tilde{\pi}_{jk}}\left[\tilde{\pi}_{jk}\log(\tilde{\pi}_{jk})\right]$$
$$=\frac{d}{2}\left[\boldsymbol{\Psi}(\tilde{\xi}_k)-\log\left(\tilde{\psi}_k\right)\right]-\frac{1}{2}\frac{\tilde{\xi}_k}{\tilde{\psi}_k}\left[||\tilde{\boldsymbol{u}}_j-\tilde{\boldsymbol{\mu}}_k||^2+d(\tilde{\sigma}_j^2+\tilde{\omega}_k^2)\right]+\boldsymbol{\Psi}\left(\tilde{\delta}_k\right)-\boldsymbol{\Psi}\left(\sum_{g=1}^{K}\tilde{\delta}_g\right)-\log(\tilde{\pi}_{jk})-1.$$

鉴于 $\sum_{k=1}^{K}\tilde{\pi}_{jk}=1$，我们应用拉格朗日乘子法，引入拉格朗日乘子 λ_j，以在约束 $\tilde{\pi}_{jk}$ 下优化证据下界，即令

$$A_{jk}=\frac{d}{2}\left[\boldsymbol{\Psi}(\tilde{\xi}_k)-\log\left(\tilde{\psi}_k\right)\right]-\frac{1}{2}\frac{\tilde{\xi}_k}{\tilde{\psi}_k}\left[||\tilde{\boldsymbol{u}}_j-\tilde{\boldsymbol{\mu}}_k||^2+d(\tilde{\sigma}_j^2+\tilde{\omega}_k^2)\right]+\boldsymbol{\Psi}\left(\tilde{\delta}_k\right)-\boldsymbol{\Psi}\left(\sum_{g=1}^{K}\tilde{\delta}_g\right),$$

我们求解以下方程:

$$\begin{cases}\frac{\partial}{\partial \tilde{\pi}_{jk}}\left[\mathcal{F}+\lambda_j\left(\sum_{g=1}^{K}\tilde{\pi}_{jg}-1\right)\right]=A_{jk}-\log(\tilde{\pi}_{jk})-1+\lambda_j;\\ \frac{\partial}{\partial \lambda_j}\left[\mathcal{F}+\lambda_i\left(\sum_{g=1}^{K}\tilde{\pi}_{jg}-1\right)\right]=\sum_{g=1}^{K}\tilde{\pi}_{jg}-1.\end{cases}$$

令上述两个方程均等于零，得到:

$$\begin{cases}\tilde{\pi}_{jk}=\exp(A_{jk}-1+\lambda_j);\\ \sum_{g=1}^{K}\exp(A_{jg}-1+\lambda_j)=1.\end{cases}\quad\Longrightarrow\quad\begin{cases}\lambda_j=-\log\left[\sum_{g=1}^{K}\exp(A_{jg}-1)\right];\\ \tilde{\pi}_{jk}=\frac{\exp(A_{jk}-1)}{\sum_{g=1}^{K}\exp(A_{jg}-1)}=\frac{\exp(A_{jk})}{\sum_{g=1}^{K}\exp(A_{jg})}.\end{cases}$$

□

证据下界关于 $\tilde{\boldsymbol{\mu}}_k$ for $k=1$，2、.....、K 的一阶偏导数写作:

$$\frac{\partial \mathcal{F}}{\partial \tilde{\boldsymbol{\mu}}_k}=\frac{\partial}{\partial \tilde{\boldsymbol{\mu}}_k}\sum_{i=1}^{N}\left(-\frac{\tilde{\pi}_{ik}}{2}\frac{\tilde{\xi}_k}{\tilde{\psi}_k}||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{\mu}}_k||^2\right)+\frac{\partial}{\partial \tilde{\boldsymbol{\mu}}_k}\left(-\frac{1}{2\omega^2}\tilde{\boldsymbol{\mu}}_k^T\tilde{\boldsymbol{\mu}}_k\right)=\left[\sum_{i=1}^{N}\tilde{\pi}_{ik}\frac{\tilde{\xi}_k}{\tilde{\psi}_k}(\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{\mu}}_k)\right]-\frac{1}{\omega^2}\tilde{\boldsymbol{\mu}}_k.$$

令这个一阶偏导数等于零，可得:

$$\left(\frac{\tilde{\xi}_k}{\tilde{\psi}_k}\sum_{i=1}^{N}\tilde{\pi}_{ik}\tilde{\boldsymbol{u}}_i\right)-\left(\frac{\tilde{\xi}_k}{\tilde{\psi}_k}\sum_{i=1}^{N}\tilde{\pi}_{ik}\right)\tilde{\boldsymbol{\mu}}_k-\frac{1}{\omega^2}\tilde{\boldsymbol{\mu}}_k=0$$
$$\Longrightarrow\ \tilde{\boldsymbol{\mu}}_k=\frac{\left(\frac{\tilde{\xi}_k}{\tilde{\psi}_k}\sum_{i=1}^{N}\tilde{\pi}_{ik}\tilde{\boldsymbol{u}}_i\right)}{\left(\frac{\tilde{\xi}_k}{\tilde{\psi}_k}\sum_{i=1}^{N}\tilde{\pi}_{ik}\right)+\frac{1}{\omega^2}}=\frac{\omega^2\tilde{\xi}_k\left(\sum_{i=1}^{N}\tilde{\pi}_{ik}\tilde{\boldsymbol{u}}_i\right)}{\omega^2\tilde{\xi}_k\left(\sum_{i=1}^{N}\tilde{\pi}_{ik}\right)+\tilde{\psi}_k}.$$

□

ELBO 关于 $\tilde{\omega}_k^2$for $k = 1,2,\ldots,K$ 的一阶偏导数写为:

$$\begin{aligned}\frac{\partial \mathcal{F}}{\partial \tilde{\omega}_k^2} &= \frac{\partial}{\partial \tilde{\omega}_k^2}\sum_{i=1}^{N}\left(-\frac{\tilde{\pi}_{ik}}{2}\frac{\tilde{\xi}_k}{\tilde{\psi}_k}d\cdot\tilde{\omega}_k^2\right)+\frac{\partial}{\partial \tilde{\omega}_k^2}\left[-\frac{1}{2\omega^2}d\cdot\tilde{\omega}_k^2+\frac{d}{2}\log(\tilde{\omega}_k^2)\right]\\ &=\left[-\frac{d}{2}\sum_{i=1}^{N}\left(\tilde{\pi}_{ik}\frac{\tilde{\xi}_k}{\tilde{\psi}_k}\right)\right]-\frac{d}{2\omega^2}+\frac{d}{2\tilde{\omega}_k^2}=\frac{d}{2}\left(\frac{1}{\tilde{\omega}_k^2}-\frac{1}{\omega^2}-\sum_{i=1}^{N}\tilde{\pi}_{ik}\frac{\tilde{\xi}_k}{\tilde{\psi}_k}\right).\end{aligned}$$

令这个一阶偏导数等于零，可得:

$$\frac{1}{\tilde{\omega}_k^2}=\frac{1}{\omega^2}+\frac{\tilde{\xi}_k}{\tilde{\psi}_k}\sum_{i=1}^{N}\tilde{\pi}_{ik}\implies\tilde{\omega}_k^2=\frac{1}{\frac{1}{\omega^2}+\frac{\tilde{\xi}_k}{\tilde{\psi}_k}\sum_{i=1}^{N}\tilde{\pi}_{ik}}=\frac{\omega^2\tilde{\psi}_k}{\tilde{\psi}_k+\omega^2\tilde{\xi}_k\sum_{i=1}^{N}\tilde{\pi}_{ik}}.$$

□

˜证据下界相对于 $k = 1$，2，.....，K的 ξ_k的一阶偏导数写作:

$$\begin{aligned}&\frac{\partial \mathcal{F}}{\partial \tilde{\xi}_k}=\frac{\partial}{\partial \tilde{\xi}_k}\sum_{i=1}^{N}\tilde{\pi}_{ik}\left\{\frac{d}{2}\boldsymbol{\Psi}(\tilde{\xi}_k)-\frac{1}{2}\frac{\tilde{\xi}_k}{\tilde{\psi}_k}\left[||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{\mu}}_k||^2+d(\tilde{\sigma}_i^2+\tilde{\omega}_k^2)\right]\right\}\\ &+\frac{\partial}{\partial \tilde{\xi}_k}\left[(\xi-\tilde{\xi}_k)\boldsymbol{\Psi}\left(\tilde{\xi}_k\right)-\psi\frac{\tilde{\xi}_k}{\tilde{\psi}_k}+\log\Gamma(\tilde{\xi}_k)+\tilde{\xi}_k\right]\\ &=\sum_{i=1}^{N}\tilde{\pi}_{ik}\left\{\frac{d}{2}\boldsymbol{\Psi}'(\tilde{\xi}_k)-\frac{1}{2\tilde{\psi}_k}\left[||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{\mu}}_k||^2+d(\tilde{\sigma}_i^2+\tilde{\omega}_k^2)\right]\right\}\\ &-\boldsymbol{\Psi}\left(\tilde{\xi}_k\right)+(\xi-\tilde{\xi}_k)\boldsymbol{\Psi}'\left(\tilde{\xi}_k\right)-\frac{\psi}{\tilde{\psi}_k}+\boldsymbol{\Psi}\left(\tilde{\xi}_k\right)+1\\ &=1+(\xi-\tilde{\xi}_k)\boldsymbol{\Psi}'\left(\tilde{\xi}_k\right)-\frac{\psi}{\tilde{\psi}_k}+\sum_{i=1}^{N}\tilde{\pi}_{ik}\left\{\frac{d}{2}\boldsymbol{\Psi}'(\tilde{\xi}_k)-\frac{1}{2\tilde{\psi}_k}\left[||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{\mu}}_k||^2+d(\tilde{\sigma}_i^2+\tilde{\omega}_k^2)\right]\right\}\\ &=1+\left(\xi-\tilde{\xi}_k+\frac{d}{2}\sum_{i=1}^{N}\tilde{\pi}_{ik}\right)\boldsymbol{\Psi}'\left(\tilde{\xi}_k\right)-\frac{1}{2\tilde{\psi}_k}\left\{2\psi+\sum_{i=1}^{N}\tilde{\pi}_{ik}\left[||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{\mu}}_k||^2+d(\tilde{\sigma}_i^2+\tilde{\omega}_k^2)\right]\right\}.\end{aligned}$$

˜证据下界关于 ψ_k 对于 $k = 1$，2，.. . ，K的一阶偏导数写作:

$$\begin{aligned}&\frac{\partial \mathcal{F}}{\partial \tilde{\psi}_k}=\frac{\partial}{\partial \tilde{\psi}_k}\sum_{i=1}^{N}\tilde{\pi}_{ik}\left\{-\frac{d}{2}\log\left(\tilde{\psi}_k\right)-\frac{1}{2}\frac{\tilde{\xi}_k}{\tilde{\psi}_k}\left[||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{\mu}}_k||^2+d(\tilde{\sigma}_i^2+\tilde{\omega}_k^2)\right]\right\}\\ &+\frac{\partial}{\partial \tilde{\psi}_k}\left[-\xi\log\left(\tilde{\psi}_k\right)-\psi\frac{\tilde{\xi}_k}{\tilde{\psi}_k}\right]\\ &=\left\{\sum_{i=1}^{N}\tilde{\pi}_{ik}\left[-\frac{d}{2\tilde{\psi}_k}+\frac{1}{2}\frac{\tilde{\xi}_k}{\tilde{\psi}_k^2}\left[||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{\mu}}_k||^2+d(\tilde{\sigma}_i^2+\tilde{\omega}_k^2)\right]\right]\right\}-\frac{\xi}{\tilde{\psi}_k}+\psi\frac{\tilde{\xi}_k}{\tilde{\psi}_k^2}\\ &=-\frac{1}{\tilde{\psi}_k}\left(\xi+\frac{d}{2}\sum_{i=1}^{N}\tilde{\pi}_{ik}\right)+\frac{\tilde{\xi}_k}{2\tilde{\psi}_k^2}\left\{2\psi+\sum_{i=1}^{N}\tilde{\pi}_{ik}\left[||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{\mu}}_k||^2+d(\tilde{\sigma}_i^2+\tilde{\omega}_k^2)\right]\right\}.\end{aligned}$$

因此，只要

$$\begin{cases} \frac{\partial \mathcal{F}}{\partial \tilde{\psi}_k} = -\frac{1}{\tilde{\psi}_k}\left(\xi + \frac{d}{2}\sum_{i=1}^{N}\tilde{\pi}_{ik}\right) + \frac{\tilde{\xi}_k}{2\tilde{\psi}_k^2}\left\{2\psi + \sum_{i=1}^{N}\tilde{\pi}_{ik}\left[||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{\mu}}_k||^2 + d(\tilde{\sigma}_i^2 + \tilde{\omega}_k^2)\right]\right\}; \\ \frac{\partial \mathcal{F}}{\partial \tilde{\xi}_k} = 1 + \left(\xi - \tilde{\xi}_k + \frac{d}{2}\sum_{i=1}^{N}\tilde{\pi}_{ik}\right)\boldsymbol{\Psi}'\left(\tilde{\xi}_k\right) - \frac{1}{2\tilde{\psi}_k}\left\{2\psi + \sum_{i=1}^{N}\tilde{\pi}_{ik}\left[||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{\mu}}_k||^2 + d(\tilde{\sigma}_i^2 + \tilde{\omega}_k^2)\right]\right\}, \end{cases}$$

并令上述两个一阶偏导数均为零，可得：

$$\begin{cases} 2\tilde{\psi}_k = \frac{\tilde{\xi}_k\left\{2\psi + \sum_{i=1}^{N}\tilde{\pi}_{ik}\left[||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{\mu}}_k||^2 + d(\tilde{\sigma}_i^2 + \tilde{\omega}_k^2)\right]\right\}}{\left(\xi + \frac{d}{2}\sum_{i=1}^{N}\tilde{\pi}_{ik}\right)}; \\ 1 + \left(\xi - \tilde{\xi}_k + \frac{d}{2}\sum_{i=1}^{N}\tilde{\pi}_{ik}\right)\boldsymbol{\Psi}'\left(\tilde{\xi}_k\right) - \frac{1}{2\tilde{\psi}_k}\left\{2\psi + \sum_{i=1}^{N}\tilde{\pi}_{ik}\left[||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{\mu}}_k||^2 + d(\tilde{\sigma}_i^2 + \tilde{\omega}_k^2)\right]\right\} = 0 \end{cases}$$

然后将第一个方程代入第二个方程，可得：

$$1 + \left(\xi - \tilde{\xi}_k + \frac{d}{2}\sum_{i=1}^{N}\tilde{\pi}_{ik}\right)\boldsymbol{\Psi}'\left(\tilde{\xi}_k\right) - \frac{\xi + \frac{d}{2}\sum_{i=1}^{N}\tilde{\pi}_{ik}}{\tilde{\xi}_k} = \left(\xi - \tilde{\xi}_k + \frac{d}{2}\sum_{i=1}^{N}\tilde{\pi}_{ik}\right)\left[\boldsymbol{\Psi}'\left(\tilde{\xi}_k\right) - \frac{1}{\tilde{\xi}_k}\right] = 0.$$

这里，我们可以通过 R 代码轻松验证 $\left[\boldsymbol{\Psi}'\left(\tilde{\xi}_k\right) - \frac{1}{\tilde{\xi}_k}\right]$ 项始终为正。因此，最终可得解：

$$\begin{cases} \tilde{\xi}_k = \xi + \frac{d}{2}\sum_{i=1}^{N}\tilde{\pi}_{ik}; \\ \tilde{\psi}_k = \psi + \frac{1}{2}\sum_{i=1}^{N}\tilde{\pi}_{ik}\left[||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{\mu}}_k||^2 + d(\tilde{\sigma}_i^2 + \tilde{\omega}_k^2)\right]. \end{cases}$$

□

证据下界关于 $\tilde{a}_j$（$j = 1, 2, \ldots\ldots, N$）的一阶偏导数写作：

$$\begin{aligned} \frac{\partial \mathcal{F}}{\partial \tilde{a}_j} &= \frac{\partial}{\partial \tilde{a}_j}\sum_{\substack{i=1,\\ i\neq j}}^{N}\left\{\frac{d}{2}\boldsymbol{\Psi}(\tilde{a}_j) - \frac{1}{2}\frac{\tilde{a}_j}{\tilde{b}_j}\left[||\tilde{\boldsymbol{v}}_{j\leftarrow i} - \tilde{\boldsymbol{u}}_j||^2 + d(\tilde{\varphi}_{ij}^2 + \tilde{\sigma}_j^2)\right]\right\} \\ &+ \frac{\partial}{\partial \tilde{a}_j}\left[(a - \tilde{a}_j)\boldsymbol{\Psi}(\tilde{a}_j) - b\frac{\tilde{a}_j}{\tilde{b}_j} + \log\Gamma(\tilde{a}_j) + \tilde{a}_j\right] \\ &= \sum_{\substack{i=1,\\ i\neq j}}^{N}\left\{\frac{d}{2}\boldsymbol{\Psi}'(\tilde{a}_j) - \frac{1}{2\tilde{b}_j}\left[||\tilde{\boldsymbol{v}}_{j\leftarrow i} - \tilde{\boldsymbol{u}}_j||^2 + d(\tilde{\varphi}_{ij}^2 + \tilde{\sigma}_j^2)\right]\right\} \\ &- \boldsymbol{\Psi}(\tilde{a}_j) + (a - \tilde{a}_j)\boldsymbol{\Psi}'(\tilde{a}_j) - \frac{b}{\tilde{b}_j} + \boldsymbol{\Psi}(\tilde{a}_j) + 1 \\ &= 1 - \frac{b}{\tilde{b}_j} + (a - \tilde{a}_j)\boldsymbol{\Psi}'(\tilde{a}_j) + (N-1)\frac{d}{2}\boldsymbol{\Psi}'(\tilde{a}_j) - \frac{1}{2\tilde{b}_j}\sum_{\substack{i=1,\\ i\neq j}}^{N}\left[||\tilde{\boldsymbol{v}}_{j\leftarrow i} - \tilde{\boldsymbol{u}}_j||^2 + d(\tilde{\varphi}_{ij}^2 + \tilde{\sigma}_j^2)\right] \\ &= 1 + \left[a - \tilde{a}_j + \frac{d}{2}(N-1)\right]\boldsymbol{\Psi}'(\tilde{a}_j) - \frac{1}{2\tilde{b}_j}\left\{2b + \sum_{\substack{i=1,\\ i\neq j}}^{N}\left[||\tilde{\boldsymbol{v}}_{j\leftarrow i} - \tilde{\boldsymbol{u}}_j||^2 + d(\tilde{\varphi}_{ij}^2 + \tilde{\sigma}_j^2)\right]\right\}. \end{aligned}$$

一阶　　证据下界关于b_j的˜阶偏导数对于 $j = 1, 2, \ldots\ldots, N$写为　　n

as:

$$\begin{aligned}
\frac{\partial \mathcal{F}}{\partial \tilde{b}_j} &= \frac{\partial}{\partial \tilde{b}_j} \sum_{\substack{i=1,\\ i\neq j}}^{N} \left\{ -\frac{d}{2}\log\left(\tilde{b}_j\right) - \frac{1}{2}\frac{\tilde{a}_j}{\tilde{b}_j}\left[||\tilde{\boldsymbol{v}}_{j\leftarrow i} - \tilde{\boldsymbol{u}}_j||^2 + d(\tilde{\varphi}_{ij}^2 + \tilde{\sigma}_j^2)\right] \right\} \\
&+ \frac{\partial}{\partial \tilde{a}_j}\left[-a\log\left(\tilde{b}_j\right) - b\frac{\tilde{a}_j}{\tilde{b}_j} \right] \\
&= \left\{ \sum_{\substack{i=1,\\ i\neq j}}^{N} \left\{ -\frac{d}{2\tilde{b}_j} + \frac{\tilde{a}_j}{2\tilde{b}_j^2}\left[||\tilde{\boldsymbol{v}}_{j\leftarrow i} - \tilde{\boldsymbol{u}}_j||^2 + d(\tilde{\varphi}_{ij}^2 + \tilde{\sigma}_j^2)\right] \right\} \right\} - \frac{a}{\tilde{b}_j} + b\frac{\tilde{a}_j}{\tilde{b}_j^2} \\
&= -\frac{1}{\tilde{b}_j}\left[\frac{d}{2}(N-1) + a\right] + \frac{\tilde{a}_j}{2\tilde{b}_j^2}\left\{ 2b + \sum_{\substack{i=1,\\ i\neq j}}^{N}\left[||\tilde{\boldsymbol{v}}_{j\leftarrow i} - \tilde{\boldsymbol{u}}_j||^2 + d(\tilde{\varphi}_{ij}^2 + \tilde{\sigma}_j^2)\right] \right\}.
\end{aligned}$$

倘若

$$\begin{cases}
\frac{\partial \mathcal{F}}{\partial \tilde{b}_j} = -\frac{1}{\tilde{b}_j}\left[\frac{d}{2}(N-1)+a\right] + \frac{\tilde{a}_j}{2\tilde{b}_j^2}\left\{ 2b + \sum_{\substack{i=1,\\ i\neq j}}^{N}\left[||\tilde{\boldsymbol{v}}_{j\leftarrow i} - \tilde{\boldsymbol{u}}_j||^2 + d(\tilde{\varphi}_{ij}^2 + \tilde{\sigma}_j^2)\right] \right\}; \\
\frac{\partial \mathcal{F}}{\partial \tilde{a}_j} = 1 + \left[a - \tilde{a}_j + \frac{d}{2}(N-1)\right]\boldsymbol{\Psi}'(\tilde{a}_j) - \frac{1}{2\tilde{b}_j}\left\{ 2b + \sum_{\substack{i=1,\\ i\neq j}}^{N}\left[||\tilde{\boldsymbol{v}}_{j\leftarrow i} - \tilde{\boldsymbol{u}}_j||^2 + d(\tilde{\varphi}_{ij}^2 + \tilde{\sigma}_j^2)\right] \right\},
\end{cases}$$

并令上述两个一阶偏导数均等于零，得到:

$$\begin{aligned}
&\begin{cases}
2\tilde{b}_j = \frac{\tilde{a}_j\left\{ 2b + \sum_{\substack{i=1,\\ i\neq j}}^{N}\left[||\tilde{\boldsymbol{v}}_{j\leftarrow i} - \tilde{\boldsymbol{u}}_j||^2 + d(\tilde{\varphi}_{ij}^2 + \tilde{\sigma}_j^2)\right] \right\}}{\left[\frac{d}{2}(N-1)+a\right]}; \\
1 + \left[a - \tilde{a}_j + \frac{d}{2}(N-1)\right]\boldsymbol{\Psi}'(\tilde{a}_j) - \frac{\left[\frac{d}{2}(N-1)+a\right]}{\tilde{a}_j} = 0.
\end{cases} \\
\Longrightarrow\ &\begin{cases}
\left[a - \tilde{a}_j + \frac{d}{2}(N-1)\right]\left[\boldsymbol{\Psi}'(\tilde{a}_j) - \frac{1}{\tilde{a}_j}\right] = 0 \implies \tilde{a}_j = a + \frac{d}{2}(N-1); \\
\tilde{b}_j = b + \frac{1}{2}\sum_{\substack{i=1,\\ i\neq j}}^{N}\left[||\tilde{\boldsymbol{v}}_{j\leftarrow i} - \tilde{\boldsymbol{u}}_j||^2 + d(\tilde{\varphi}_{ij}^2 + \tilde{\sigma}_j^2)\right].
\end{cases}
\end{aligned}$$

□

A.2.2 命题4的证明

证据下界关于 $\tilde{\boldsymbol{u}}_i$ for $i=1,2,\ldots\ldots,N$的一阶偏导数写作:

$$
\begin{aligned}
&\frac{\partial \mathcal{F}}{\partial \tilde{\boldsymbol{u}}_i}=\frac{\partial}{\partial \tilde{\boldsymbol{u}}_i}\sum_{\substack{j=1,\\ j\neq i}}^{N}\left\{-y_{ij}||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{v}}_{j\leftarrow i}||^2-\frac{\exp\left(\tilde{\eta}+\frac{\tilde{\rho}^2}{2}-\frac{||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{v}}_{j\leftarrow i}||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2)^{d/2}}-\frac{1}{2}\frac{\tilde{a}_i}{\tilde{b}_i}||\tilde{\boldsymbol{v}}_{i\leftarrow j}-\tilde{\boldsymbol{u}}_i||^2\right\}\\
&+\frac{\partial}{\partial \tilde{\boldsymbol{u}}_i}\sum_{k=1}^{K}\left[-\frac{\tilde{\pi}_{ik}}{2}\frac{\tilde{\xi}_k}{\tilde{\psi}_k}||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{\mu}}_k||^2\right]\\
&=\sum_{\substack{j=1,\\ j\neq i}}^{N}\left[-2y_{ij}(\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{v}}_{j\leftarrow i})-\frac{\exp\left(\tilde{\eta}+\frac{\tilde{\rho}^2}{2}-\frac{||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{v}}_{j\leftarrow i}||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2)^{d/2}}\frac{-2(\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{v}}_{j\leftarrow i})}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2}+\frac{\tilde{a}_i}{\tilde{b}_i}(\tilde{\boldsymbol{v}}_{i\leftarrow j}-\tilde{\boldsymbol{u}}_i)\right]\\
&-\sum_{k=1}^{K}\left[\tilde{\pi}_{ik}\frac{\tilde{\xi}_k}{\tilde{\psi}_k}(\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{\mu}}_k)\right]\\
&=\sum_{\substack{j=1,\\ j\neq i}}^{N}\left\{-2\left[y_{ij}-\frac{\exp\left(\tilde{\eta}+\frac{\tilde{\rho}^2}{2}-\frac{||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{v}}_{j\leftarrow i}||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2)^{d/2+1}}\right](\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{v}}_{j\leftarrow i})+\frac{\tilde{a}_i}{\tilde{b}_i}(\tilde{\boldsymbol{v}}_{i\leftarrow j}-\tilde{\boldsymbol{u}}_i)\right\}-\sum_{k=1}^{K}\left[\tilde{\pi}_{ik}\frac{\tilde{\xi}_k}{\tilde{\psi}_k}(\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{\mu}}_k)\right].
\end{aligned}
$$

ELBO 关于 $\tilde{\sigma}_i^2$（$i=1,2,\ldots\ldots,N$）的一阶偏导数写为:

$$
\begin{aligned}
&\frac{\partial \mathcal{F}}{\partial \tilde{\sigma}_i^2}=\frac{\partial}{\partial \tilde{\sigma}_i^2}\sum_{\substack{j=1,\\ j\neq i}}^{N}\left[-y_{ij}\cdot d\tilde{\sigma}_i^2-\frac{\exp\left(\tilde{\eta}+\frac{\tilde{\rho}^2}{2}-\frac{||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{v}}_{j\leftarrow i}||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2)^{d/2}}-\frac{1}{2}\frac{\tilde{a}_i}{\tilde{b}_i}d\tilde{\sigma}_i^2\right]\\
&-\frac{\partial}{\partial \tilde{\sigma}_i^2}\sum_{k=1}^{K}\left(\frac{\tilde{\pi}_{ik}}{2}\frac{\tilde{\xi}_k}{\tilde{\psi}_k}d\tilde{\sigma}_i^2\right)+\frac{\partial}{\partial \tilde{\sigma}_i^2}\left[\frac{d}{2}\log\left(\tilde{\sigma}_i^2\right)\right]\\
&=\sum_{\substack{j=1,\\ j\neq i}}^{N}\left(-y_{ij}d-\frac{1}{2}\frac{\tilde{a}_i}{\tilde{b}_i}d\right)-\sum_{k=1}^{K}\left(\frac{\tilde{\pi}_{ik}}{2}\frac{\tilde{\xi}_k}{\tilde{\psi}_k}d\right)+\frac{d}{2\tilde{\sigma}_i^2}\\
&-\sum_{\substack{j=1,\\ j\neq i}}^{N}\left[\frac{\exp\left(\tilde{\eta}+\frac{\tilde{\rho}^2}{2}-\frac{||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{v}}_{j\leftarrow i}||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2}\right)\cdot 2||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{v}}_{j\leftarrow i}||^2(1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2)^{-2}}{(1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2)^{d/2}}-\frac{d}{2}\frac{\exp\left(\tilde{\eta}+\frac{\tilde{\rho}^2}{2}-\frac{||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{v}}_{j\leftarrow i}||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2}\right)\cdot 2}{(1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2)^{d/2+1}}\right]\\
&=\frac{d}{2}\cdot\left[-2\left(\sum_{\substack{j=1,\\ j\neq i}}^{N}y_{ij}\right)-(N-1)\frac{\tilde{a}_i}{\tilde{b}_i}-\left(\sum_{k=1}^{K}\tilde{\pi}_{ik}\frac{\tilde{\xi}_k}{\tilde{\psi}_k}\right)+\frac{1}{\tilde{\sigma}_i^2}\right]\\
&-\sum_{\substack{j=1,\\ j\neq i}}^{N}\frac{\exp\left(\tilde{\eta}+\frac{\tilde{\rho}^2}{2}-\frac{||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{v}}_{j\leftarrow i}||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2)^{d/2+1}}\left[2\frac{||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{v}}_{j\leftarrow i}||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2}-d\right].
\end{aligned}
$$

为了得到（$\tilde{\boldsymbol{u}}_i,\tilde{\sigma}_i^2$）的自然梯度上升更新，首先注意到 $\tilde{\sigma}_i^2>0$。因此，我们应用变换 $\exp\left(\tilde{\zeta}_i\right)=\tilde{\sigma}_i^2$，并令

$$
\mathcal{H}\left(\tilde{\boldsymbol{u}}_i,\tilde{\zeta}_i,\ldots\right)=\mathcal{F}\left(\tilde{\boldsymbol{u}}_i,\exp\left(\tilde{\zeta}_i\right),\ldots\right)=\mathcal{F}\left(\tilde{\boldsymbol{u}}_i,\tilde{\sigma}_i^2,\ldots\right).
$$

也就是说，我们首先对 $\tilde{\sigma}_i^2$ 取对数，得到参数 $\tilde{\zeta}_i$，它位于（$-\infty, \infty$）内；随后，我们对 $\tilde{\zeta}_i$ 取指数，将其变换回 $\tilde{\sigma}_i^2$。这一变换意味着

$$\frac{\partial}{\partial \tilde{\zeta}_i}\mathcal{H}\left(\tilde{\boldsymbol{u}}_i, \tilde{\zeta}_i, \dots\right) = \frac{\partial}{\partial \tilde{\zeta}_i}\mathcal{F}\left(\tilde{\boldsymbol{u}}_i, \exp\left(\tilde{\zeta}_i\right), \dots\right) = \frac{\partial}{\partial \tilde{\sigma}_i^2}\mathcal{F}\left(\tilde{\boldsymbol{u}}_i, \tilde{\sigma}_i^2, \dots\right)\frac{\partial \tilde{\sigma}_i^2}{\partial \tilde{\zeta}_i} = \exp\left(\tilde{\zeta}_i\right)\frac{\partial}{\partial \tilde{\sigma}_i^2}\mathcal{F}\left(\tilde{\boldsymbol{u}}_i, \tilde{\sigma}_i^2, \dots\right).$$

因此，标准梯度上升可写为：

$$\begin{cases} \tilde{\boldsymbol{u}}_i^{\star} = \tilde{\boldsymbol{u}}_i^{\dagger} + \epsilon \frac{\partial \mathcal{F}}{\partial \tilde{\boldsymbol{u}}_i}\left(\tilde{\boldsymbol{u}}_i^{\dagger}, \exp\left(\tilde{\zeta}_i^{\dagger}\right), \dots\right); \\ \tilde{\zeta}_i^{\star} = \tilde{\zeta}_i^{\dagger} + \epsilon \exp\left(\tilde{\zeta}_i^{\dagger}\right)\frac{\partial \mathcal{F}}{\partial \tilde{\sigma}_i^2}\left(\tilde{\boldsymbol{u}}_i^{\dagger}, \tilde{\sigma}_i^{2^{\dagger}}, \dots\right). \end{cases}$$

接下来，我们计算与似然 $q_{\boldsymbol{U}}\left(\boldsymbol{u}_i|\tilde{\boldsymbol{u}}_i, \exp\left(\tilde{\zeta}_i\right)\right)$ 相关的费希尔信息 $(\tilde{\boldsymbol{u}}_i, \tilde{\zeta}_i)$。对数似然可写为：

$$\begin{aligned} &\log q_{\boldsymbol{U}}\left(\boldsymbol{u}_i|\tilde{\boldsymbol{u}}_i, \exp\left(\tilde{\zeta}_i\right)\right) = \log f_{\mathrm{MVN}_d}\left(\boldsymbol{u}_i|\tilde{\boldsymbol{u}}_i, \exp\left(\tilde{\zeta}_i\right)\mathbb{I}_d\right) \\ &= \log\left\{\frac{1}{(2\pi)^{\frac{d}{2}}\left|\exp\left(\tilde{\zeta}_i\right)\mathbb{I}_d\right|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\boldsymbol{u}_i - \tilde{\boldsymbol{u}}_i)^T\frac{1}{\exp\left(\tilde{\zeta}_i\right)}\mathbb{I}_d(\boldsymbol{u}_i - \tilde{\boldsymbol{u}}_i)\right]\right\} \\ &= -\frac{d}{2}\log(2\pi) - \frac{d}{2}\tilde{\zeta}_i - \frac{1}{2\exp\left(\tilde{\zeta}_i\right)}||\boldsymbol{u}_i - \tilde{\boldsymbol{u}}_i||^2. \end{aligned}$$

因此，上述对数似然的一阶偏导数可写为：

$$\begin{cases} \frac{\partial \log q_{\boldsymbol{U}}}{\partial \tilde{\boldsymbol{u}}_i} = \frac{1}{\exp\left(\tilde{\zeta}_i\right)}\left(\boldsymbol{u}_i - \tilde{\boldsymbol{u}}_i\right); \\ \frac{\partial \log q_{\boldsymbol{U}}}{\partial \tilde{\zeta}_i} = -\frac{d}{2} + \frac{1}{2\exp\left(\tilde{\zeta}_i\right)}||\boldsymbol{u}_i - \tilde{\boldsymbol{u}}_i||^2. \end{cases}$$

因此，二阶偏导数为：

$$\begin{cases} \frac{\partial^2 \log q_{\boldsymbol{U}}}{\partial \tilde{\boldsymbol{u}}_i^2} = -\frac{1}{\exp\left(\tilde{\zeta}_i\right)}; \\ \frac{\partial^2 \log q_{\boldsymbol{U}}}{\partial \tilde{\boldsymbol{u}}_i \partial \tilde{\zeta}_i} = -\frac{1}{\exp\left(\tilde{\zeta}_i\right)}\left(\boldsymbol{u}_i - \tilde{\boldsymbol{u}}_i\right); \\ \frac{\partial^2 \log q_{\boldsymbol{U}}}{\partial \tilde{\zeta}_i^2} = -\frac{1}{2\exp\left(\tilde{\zeta}_i\right)}||\boldsymbol{u}_i - \tilde{\boldsymbol{u}}_i||^2. \end{cases}$$

取负期望 $-\mathbb{E}_q$，得到：

$$\begin{cases} -\mathbb{E}_q\left(\frac{\partial^2 \log q_{\boldsymbol{U}}}{\partial \tilde{\boldsymbol{u}}_i^2}\right) = \frac{1}{\exp\left(\tilde{\zeta}_i\right)}; \\ -\mathbb{E}_q\left(\frac{\partial^2 \log q_{\boldsymbol{U}}}{\partial \tilde{\boldsymbol{u}}_i \partial \tilde{\zeta}_i}\right) = \frac{1}{\exp\left(\tilde{\zeta}_i\right)}\left(\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_i\right) = 0; \\ -\mathbb{E}_q\left(\frac{\partial^2 \log q_{\boldsymbol{U}}}{\partial \tilde{\zeta}_i^2}\right) = \frac{1}{2\exp\left(\tilde{\zeta}_i\right)}\left[||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_i||^2 + d\exp\left(\tilde{\zeta}_i\right)\right] = \frac{d}{2}. \end{cases}$$

这意味着费希尔信息 $\mathcal{I}(\tilde{\boldsymbol{u}}_i,\ \tilde{\zeta}_i)$ 和逆费希尔信息 $\mathcal{I}^{-1}(\tilde{\boldsymbol{u}}_i,\ \tilde{\zeta}_i)$ 分别可写为：

$$\mathcal{I}(\tilde{\boldsymbol{u}}_i, \tilde{\zeta}_i) = \begin{pmatrix} \frac{1}{\exp\left(\tilde{\zeta}_i\right)} & 0 \\ 0 & \frac{d}{2} \end{pmatrix}, \quad \mathcal{I}^{-1}(\tilde{\boldsymbol{u}}_i, \tilde{\zeta}_i) = \begin{pmatrix} \exp\left(\tilde{\zeta}_i\right) & 0 \\ 0 & \frac{2}{d} \end{pmatrix}.$$

因此，最终的自然梯度上升步如下：

$$\begin{cases} \tilde{\boldsymbol{u}}_i^{\star} = \tilde{\boldsymbol{u}}_i^{\dagger} + \epsilon \exp\left(\tilde{\zeta}_i^{\dagger}\right) \frac{\partial \mathcal{F}}{\partial \tilde{\boldsymbol{u}}_i} \left(\tilde{\boldsymbol{u}}_i^{\dagger}, \exp\left(\tilde{\zeta}_i^{\dagger}\right), \dots\right); \\ \tilde{\zeta}_i^{\star} = \tilde{\zeta}_i^{\dagger} + \epsilon \frac{2}{d} \exp\left(\tilde{\zeta}_i^{\dagger}\right) \frac{\partial \mathcal{F}}{\partial \tilde{\sigma}_i^2} \left(\tilde{\boldsymbol{u}}_i^{\dagger}, \tilde{\sigma}_i^{2^{\dagger}}, \dots\right). \end{cases}$$

对上文第2行取指数，并将 $\tilde{\zeta}_i$ 变换回 $\tilde{\sigma}_i^2$，得到：

$$\begin{aligned} &\begin{cases} \tilde{\boldsymbol{u}}_i^{\star} = \tilde{\boldsymbol{u}}_i^{\dagger} + \epsilon \exp\left(\tilde{\zeta}_i^{\dagger}\right) \frac{\partial \mathcal{F}}{\partial \tilde{\boldsymbol{u}}_i} \left(\tilde{\boldsymbol{u}}_i^{\dagger}, \exp\left(\tilde{\zeta}_i^{\dagger}\right), \dots\right); \\ \exp\left(\tilde{\zeta}_i^{\star}\right) = \exp\left(\tilde{\zeta}_i^{\dagger}\right) \exp\left[\epsilon \frac{2}{d} \exp\left(\tilde{\zeta}_i^{\dagger}\right) \frac{\partial \mathcal{F}}{\partial \tilde{\sigma}_i^2} \left(\tilde{\boldsymbol{u}}_i^{\dagger}, \tilde{\sigma}_i^{2^{\dagger}}, \dots\right)\right]. \end{cases} \\ &= \begin{cases} \tilde{\boldsymbol{u}}_i^{\star} = \tilde{\boldsymbol{u}}_i^{\dagger} + \epsilon \tilde{\sigma}_i^{2^{\dagger}} \frac{\partial \mathcal{F}}{\partial \tilde{\boldsymbol{u}}_i} \left(\tilde{\boldsymbol{u}}_i^{\dagger}, \tilde{\sigma}_i^{2^{\dagger}}, \dots\right); \\ \tilde{\sigma}_i^{2^{\star}} = \tilde{\sigma}_i^{2^{\dagger}} \exp\left[\epsilon \frac{2}{d} \tilde{\sigma}_i^{2^{\dagger}} \frac{\partial \mathcal{F}}{\partial \tilde{\sigma}_i^2} \left(\tilde{\boldsymbol{u}}_i^{\dagger}, \tilde{\sigma}_i^{2^{\dagger}}, \dots\right)\right]. \end{cases} \end{aligned}$$

□

证据下界关于 $\tilde{\boldsymbol{v}}_{j\leftarrow i}$ 的一阶偏导数，对于 $i, j = 1, 2, \dots, N$；$i \neq j$，写为：

$$\begin{aligned} \frac{\partial \mathcal{F}}{\partial \tilde{\boldsymbol{v}}_{j\leftarrow i}} &= \frac{\partial}{\partial \tilde{\boldsymbol{v}}_{j\leftarrow i}} \left\{ -y_{ij} ||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j\leftarrow i}||^2 - \frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j\leftarrow i}||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2)^{d/2}} - \frac{1}{2}\frac{\tilde{a}_j}{\tilde{b}_j} ||\tilde{\boldsymbol{v}}_{j\leftarrow i} - \tilde{\boldsymbol{u}}_j||^2 \right\} \\ &= \left[2y_{ij}(\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j\leftarrow i}) - \frac{\tilde{a}_j}{\tilde{b}_j}(\tilde{\boldsymbol{v}}_{j\leftarrow i} - \tilde{\boldsymbol{u}}_j) \right] - \frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j\leftarrow i}||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2)^{d/2}} \frac{2(\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j\leftarrow i})}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2} \\ &= \left\{ 2 \left[y_{ij} - \frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j\leftarrow i}||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2)^{d/2+1}} \right] (\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j\leftarrow i}) - \frac{\tilde{a}_j}{\tilde{b}_j}(\tilde{\boldsymbol{v}}_{j\leftarrow i} - \tilde{\boldsymbol{u}}_j) \right\}. \end{aligned}$$

证据下界关于 $\tilde{\varphi}_{ij}^2$ 的一阶偏导数，对于 $i, j = 1, 2, \dots, N$；$i \neq j$，写为：

$$\begin{aligned} \frac{\partial \mathcal{F}}{\partial \tilde{\varphi}_{ij}^2} &= \frac{\partial}{\partial \tilde{\varphi}_{ij}^2} \left\{ -y_{ij} \cdot d\tilde{\varphi}_{ij}^2 - \frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j\leftarrow i}||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2)^{d/2}} - \frac{1}{2}\frac{\tilde{a}_j}{\tilde{b}_j} d\tilde{\varphi}_{ij}^2 + \frac{d}{2}\log\tilde{\varphi}_{ij}^2 \right\} \\ &= -y_{ij}d - \frac{1}{2}\frac{\tilde{a}_j}{\tilde{b}_j}d + \frac{d}{2\tilde{\varphi}_{ij}^2} \\ &- \left[\frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j\leftarrow i}||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2}\right) \cdot 2||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j\leftarrow i}||^2 (1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2)^{-2}}{(1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2)^{d/2}} - \frac{d}{2}\frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j\leftarrow i}||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2}\right) \cdot 2}{(1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2)^{d/2+1}} \right] \\ &= \frac{d}{2}\left(-2y_{ij} - \frac{\tilde{a}_j}{\tilde{b}_j} + \frac{1}{\tilde{\varphi}_{ij}^2}\right) - \frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j\leftarrow i}||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2}\right) \left[2\frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{v}}_{j\leftarrow i}||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2} - d\right]}{(1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2)^{d/2+1}}. \end{aligned}$$

为了得到 $(\tilde{\boldsymbol{v}}_{j\leftarrow i}, \tilde{\varphi}_{ij}^2)$ 的自然梯度上升更新，我们采用与 $(\tilde{\boldsymbol{u}}_i, \tilde{\sigma}_i^2)$ 类似的步骤，令 $\exp(\tilde{\nu}_{ij}) = \tilde{\varphi}_{ij}^2$ 并令

$$\mathcal{H}\left(\tilde{\boldsymbol{u}}_i, \tilde{\nu}_{ij}, \dots\right) = \mathcal{F}\left(\tilde{\boldsymbol{u}}_i, \exp(\tilde{\nu}_{ij}), \dots\right) = \mathcal{F}\left(\tilde{\boldsymbol{u}}_i, \tilde{\varphi}_{ij}^2, \dots\right).$$

因此我们有 $\frac{\partial}{\partial\tilde{\nu}_{ij}}\mathcal{H}\left(\underline{\tilde{\boldsymbol{v}}_{j\leftarrow i}},\tilde{\nu}_{ij},\dots\right)=\exp(\tilde{\nu}_{ij})\frac{\partial}{\partial\tilde{\varphi}_{ij}^2}\mathcal{F}\left(\underline{\tilde{\boldsymbol{v}}_{j\leftarrow i}},\tilde{\varphi}_{ij}^2,\dots\right)$. 对数似然的形式为:

$$\log q_{\boldsymbol{V}}\left(\boldsymbol{v}_{j\leftarrow i}|\tilde{\boldsymbol{v}}_{j\leftarrow i},\exp(\tilde{\nu}_{ij})\right)=-\frac{d}{2}\log(2\pi)-\frac{d}{2}\tilde{\nu}_{ij}-\frac{1}{2\exp(\tilde{\nu}_{ij})}||\boldsymbol{v}_{j\leftarrow i}-\tilde{\boldsymbol{v}}_{j\leftarrow i}||^2.$$

二阶偏导数等于:

$$\frac{\partial^2\log q_{\boldsymbol{V}}}{\partial\tilde{\boldsymbol{v}}_{j\leftarrow i}^2}=-\frac{1}{\exp(\tilde{\nu}_{ij})},\quad\frac{\partial^2\log q_{\boldsymbol{V}}}{\partial\tilde{\boldsymbol{v}}_{j\leftarrow i}\partial\tilde{\nu}_{ij}}=-\frac{(\boldsymbol{v}_{j\leftarrow i}-\tilde{\boldsymbol{v}}_{j\leftarrow i})}{\exp(\tilde{\nu}_{ij})},\quad\frac{\partial^2\log q_{\boldsymbol{V}}}{\partial\tilde{\nu}_{ij}^2}=-\frac{||\boldsymbol{v}_{j\leftarrow i}-\tilde{\boldsymbol{v}}_{j\leftarrow i}||^2}{2\exp(\tilde{\nu}_{ij})}.$$

取负期望 $-\mathbb{E}_q$ 可得:

$$-\mathbb{E}_q\left(\frac{\partial^2\log q_{\boldsymbol{V}}}{\partial\tilde{\boldsymbol{v}}_{j\leftarrow i}^2}\right)=\frac{1}{\exp(\tilde{\nu}_{ij})},\quad-\mathbb{E}_q\left(\frac{\partial^2\log q_{\boldsymbol{V}}}{\partial\tilde{\boldsymbol{v}}_{j\leftarrow i}\partial\tilde{\nu}_{ij}}\right)=0,\quad-\mathbb{E}_q\left(\frac{\partial^2\log q_{\boldsymbol{V}}}{\partial\tilde{\nu}_{ij}^2}\right)=\frac{d}{2}.$$

因此，费希尔信息及其逆如下:

$$\mathcal{I}(\tilde{\boldsymbol{v}}_{j\leftarrow i},\tilde{\nu}_{ij})=\begin{pmatrix}\frac{1}{\exp(\tilde{\nu}_{ij})}&0\\0&\frac{d}{2}\end{pmatrix},\quad\mathcal{I}^{-1}(\tilde{\boldsymbol{v}}_{j\leftarrow i},\tilde{\nu}_{ij})=\begin{pmatrix}\exp(\tilde{\nu}_{ij})&0\\0&\frac{2}{d}\end{pmatrix}.$$

最后，自然梯度上升步的形式为:

$$\begin{cases}\tilde{\boldsymbol{v}}_{j\leftarrow i}^{\star}=\tilde{\boldsymbol{v}}_{j\leftarrow i}^{\dagger}+\epsilon\exp\left(\tilde{\nu}_{ij}^{\dagger}\right)\frac{\partial\mathcal{F}}{\partial\tilde{\boldsymbol{v}}_{j\leftarrow i}}\left(\tilde{\boldsymbol{v}}_{j\leftarrow i}^{\dagger},\exp\left(\tilde{\nu}_{ij}^{\dagger}\right),\dots\right);\\\exp\left(\tilde{\nu}_{ij}^{\star}\right)=\exp\left(\tilde{\nu}_{ij}^{\dagger}\right)\exp\left[\epsilon\frac{2}{d}\exp\left(\tilde{\nu}_{ij}^{\dagger}\right)\frac{\partial\mathcal{F}}{\partial\tilde{\varphi}_{ij}^2}\left(\tilde{\boldsymbol{v}}_{j\leftarrow i}^{\dagger},\tilde{\varphi}_{ij}^{2\dagger},\dots\right)\right].\end{cases}$$

$$=\begin{cases}\tilde{\boldsymbol{v}}_{j\leftarrow i}^{\star}=\tilde{\boldsymbol{v}}_{j\leftarrow i}^{\dagger}+\epsilon\tilde{\varphi}_{ij}^{2\dagger}\frac{\partial\mathcal{F}}{\partial\tilde{\boldsymbol{v}}_{j\leftarrow i}}\left(\tilde{\boldsymbol{v}}_{j\leftarrow i}^{\dagger},\tilde{\varphi}_{ij}^{2\dagger},\dots\right);\\\tilde{\varphi}_{ij}^{2\star}=\tilde{\varphi}_{ij}^{2\dagger}\exp\left[\epsilon\frac{2}{d}\tilde{\varphi}_{ij}^{2\dagger}\frac{\partial\mathcal{F}}{\partial\tilde{\varphi}_{ij}^2}\left(\tilde{\boldsymbol{v}}_{j\leftarrow i}^{\dagger},\tilde{\varphi}_{ij}^{2\dagger},\dots\right)\right].\end{cases}$$

□

关于 $\tilde{\eta}$ 的证据下界一阶偏导数写作:

$$\frac{\partial\mathcal{F}}{\partial\tilde{\eta}}=\frac{\partial}{\partial\tilde{\eta}}\sum_{\substack{i,j=1,\\i\neq j}}^{N}\left[y_{ij}\tilde{\eta}-\frac{\exp\left(\tilde{\eta}+\frac{\tilde{\rho}^2}{2}-\frac{||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{v}}_{j\leftarrow i}||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2)^{d/2}}\right]-\frac{\partial}{\partial\tilde{\eta}}\left[\frac{1}{2\rho^2}(\tilde{\eta}-\eta)^2\right]$$

$$=\left(\sum_{\substack{i,j=1,\\i\neq j}}^{N}y_{ij}\right)-\frac{\tilde{\eta}-\eta}{\rho^2}-\sum_{\substack{i,j=1,\\i\neq j}}^{N}\frac{\exp\left(\tilde{\eta}+\frac{\tilde{\rho}^2}{2}-\frac{||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{v}}_{j\leftarrow i}||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2)^{d/2}}.$$

证据下界关于 $\tilde{\rho}^2$ 的一阶偏导数写为:

$$\frac{\partial\mathcal{F}}{\partial\tilde{\rho}^2}=\frac{\partial}{\partial\tilde{\rho}^2}\sum_{\substack{i,j=1,\\i\neq j}}^{N}\left[-\frac{\exp\left(\tilde{\eta}+\frac{\tilde{\rho}^2}{2}-\frac{||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{v}}_{j\leftarrow i}||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2)^{d/2}}\right]+\frac{\partial}{\partial\tilde{\rho}^2}\left[-\frac{\tilde{\rho}^2}{2\rho^2}+\frac{1}{2}\log\tilde{\rho}^2\right]$$

$$=-\frac{1}{2\rho^2}+\frac{1}{2\tilde{\rho}^2}-\frac{1}{2}\sum_{\substack{i,j=1,\\i\neq j}}^{N}\frac{\exp\left(\tilde{\eta}+\frac{\tilde{\rho}^2}{2}-\frac{||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{v}}_{j\leftarrow i}||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\varphi}_{ij}^2)^{d/2}}.$$

为推导 $(\tilde{\eta},\ \tilde{\rho}^2)$ 的自然梯度上升更新，其中 $\tilde{\rho}^2>0$，令 $\exp(\tilde{\iota})=\tilde{\rho}^2$ 并令 $\mathcal{H}(\tilde{\eta},\ \tilde{\iota},\ \dots)=\mathcal{F}(\tilde{\eta},\ \exp(\tilde{\iota})\dots)=\mathcal{F}\left(\tilde{\eta},\ \tilde{\rho}^2,\ \dots\right)$，则

$$\frac{\partial}{\partial\tilde{\iota}}\mathcal{H}\left(\tilde{\eta},\tilde{\iota},\dots\right)=\frac{\partial}{\partial\tilde{\iota}}\mathcal{F}\left(\tilde{\eta},\exp(\tilde{\iota}),\dots\right)=\frac{\partial}{\partial\tilde{\rho}^2}\mathcal{F}\left(\tilde{\eta},\tilde{\rho}^2,\dots\right)\frac{\partial\tilde{\rho}^2}{\partial\tilde{\iota}}=\exp(\tilde{\iota})\frac{\partial}{\partial\tilde{\rho}^2}\mathcal{F}\left(\tilde{\eta},\tilde{\rho}^2,\dots\right).$$

回顾一下，$q_\beta(\beta|\tilde{\eta},\ \tilde{\rho}^2)\sim$ 正态分布$(\tilde{\eta},\ \tilde{\rho}^2)$，因此对数似然的形式为：

$$\begin{aligned}&\log q_\beta\left(\beta|\tilde{\eta},\exp(\tilde{\iota})\right)=\log f_{\text{normal}}\left(\beta|\tilde{\eta},\exp(\tilde{\iota})\right)\\&=\log\left\{\frac{1}{(2\pi)^{\frac{1}{2}}\exp(\tilde{\iota})^{\frac{1}{2}}}\exp\left[-\frac{1}{2\exp(\tilde{\iota})}(\beta-\tilde{\eta})^2\right]\right\}=-\frac{1}{2}\log(2\pi)-\frac{1}{2}\tilde{\iota}-\frac{1}{2\exp(\tilde{\iota})}(\beta-\tilde{\eta})^2.\end{aligned}$$

一阶偏导数为：

$$\frac{\partial\log q_\beta}{\partial\tilde{\eta}}=\frac{\beta-\tilde{\eta}}{\exp(\tilde{\iota})},\qquad\frac{\partial\log q_\beta}{\partial\tilde{\iota}}=-\frac{1}{2}+\frac{(\beta-\tilde{\eta})^2}{2\exp(\tilde{\iota})}.$$

二阶偏导数为：

$$\frac{\partial^2\log q_\beta}{\partial\tilde{\eta}^2}=-\frac{1}{\exp(\tilde{\iota})},\quad\frac{\partial^2\log q_\beta}{\partial\tilde{\eta}\partial\tilde{\iota}}=-\frac{(\beta-\tilde{\eta})}{\exp(\tilde{\iota})},\quad\frac{\partial^2\log q_\beta}{\partial\tilde{\iota}^2}=-\frac{(\beta-\tilde{\eta})^2}{2\exp(\tilde{\iota})}.$$

取负期望 $-\mathbb{E}_q$ 可得：

$$-\mathbb{E}_q\left(\frac{\partial^2\log q_\beta}{\partial\tilde{\eta}^2}\right)=\frac{1}{\exp(\tilde{\iota})},\quad-\mathbb{E}_q\left(\frac{\partial^2\log q_\beta}{\partial\tilde{\eta}\partial\tilde{\iota}}\right)=0,\quad-\mathbb{E}_q\left(\frac{\partial^2\log q_\beta}{\partial\tilde{\iota}^2}\right)=\frac{1}{2}.$$

因此，费希尔信息及其逆可写为：

$$\mathcal{I}(\tilde{\eta},\tilde{\iota})=\begin{pmatrix}\frac{1}{\exp(\tilde{\iota})}&0\\0&\frac{1}{2}\end{pmatrix},\quad\mathcal{I}^{-1}(\tilde{\eta},\tilde{\iota})=\begin{pmatrix}\exp(\tilde{\iota})&0\\0&2\end{pmatrix}.$$

由此得到如下形式的自然梯度上升步：

$$\begin{aligned}&\begin{cases}\tilde{\eta}^\star=\tilde{\eta}^\dagger+\epsilon\exp\left(\tilde{\iota}^\dagger\right)\frac{\partial\mathcal{F}}{\partial\tilde{\eta}}\left(\tilde{\eta}^\dagger,\exp\left(\tilde{\iota}^\dagger\right),\dots\right);\\\exp(\tilde{\iota}^\star)=\exp\left(\tilde{\iota}^\dagger\right)\exp\left[\epsilon2\exp\left(\tilde{\iota}^\dagger\right)\frac{\partial}{\partial\mathcal{F}\tilde{\rho}^2}\left(\tilde{\eta}^\dagger,\tilde{\rho}^{2^\dagger},\dots\right)\right].\end{cases}\\&=\begin{cases}\tilde{\eta}^\star=\tilde{\eta}^\dagger+\epsilon\tilde{\rho}^{2^\dagger}\frac{\partial\mathcal{F}}{\partial\tilde{\eta}}\left(\tilde{\eta}^\dagger,\tilde{\rho}^{2^\dagger},\dots\right);\\\tilde{\rho}^{2^\star}=\tilde{\rho}^{2^\dagger}\exp\left[\epsilon2\tilde{\rho}^{2^\dagger}\frac{\partial\mathcal{F}}{\partial\tilde{\rho}^2}\left(\tilde{\eta}^\dagger,\tilde{\rho}^{2^\dagger},\dots\right)\right].\end{cases}\end{aligned}$$

□

A.2.3 命题5的证明

˜证据下界对每个 δ_k 的一阶偏导数，对于 $k=1$，2，.....，K 可写为：

$$
\begin{aligned}
&\frac{\partial \mathcal{F}}{\partial \tilde{\delta}_k}=\frac{\partial}{\partial \tilde{\delta}_k} \sum_{i=1}^{N} \sum_{h=1}^{K} \tilde{\pi}_{ih}\left[\boldsymbol{\Psi}\left(\tilde{\delta}_h\right)-\boldsymbol{\Psi}\left(\sum_{g=1}^{K} \tilde{\delta}_g\right)\right] \\
&+\frac{\partial}{\partial \tilde{\delta}_k} \sum_{h=1}^{K}\left\{\left(\delta_h-\tilde{\delta}_h\right)\left[\boldsymbol{\Psi}\left(\tilde{\delta}_h\right)-\boldsymbol{\Psi}\left(\sum_{g=1}^{K} \tilde{\delta}_g\right)\right]+\log \Gamma\left(\tilde{\delta}_h\right)\right\}-\frac{\partial}{\partial \tilde{\delta}_k} \log \Gamma\left(\sum_{g=1}^{K} \tilde{\delta}_g\right) \\
&=\frac{\partial}{\partial \tilde{\delta}_k} \sum_{i=1}^{N} \sum_{h=1}^{K} \tilde{\pi}_{ih} \boldsymbol{\Psi}\left(\tilde{\delta}_h\right)-\frac{\partial}{\partial \tilde{\delta}_k}\left[\boldsymbol{\Psi}\left(\sum_{g=1}^{K} \tilde{\delta}_g\right) \cdot\left(\sum_{i=1}^{N} \sum_{h=1}^{K} \tilde{\pi}_{ih}\right)\right]+\frac{\partial}{\partial \tilde{\delta}_k} \sum_{h=1}^{K}\left(\delta_h-\tilde{\delta}_h\right) \boldsymbol{\Psi}\left(\tilde{\delta}_h\right) \\
&-\frac{\partial}{\partial \tilde{\delta}_k}\left\{\boldsymbol{\Psi}\left(\sum_{g=1}^{K} \tilde{\delta}_g\right) \cdot\left[\sum_{h=1}^{K}\left(\delta_h-\tilde{\delta}_h\right)\right]\right\}+\frac{\partial}{\partial \tilde{\delta}_k} \sum_{h=1}^{K} \log \Gamma\left(\tilde{\delta}_h\right)-\frac{\partial}{\partial \tilde{\delta}_k} \log \Gamma\left(\sum_{g=1}^{K} \tilde{\delta}_g\right) \\
&=\sum_{i=1}^{N} \tilde{\pi}_{ik} \frac{\partial}{\partial \tilde{\delta}_k} \boldsymbol{\Psi}\left(\tilde{\delta}_k\right)-\left[\frac{\partial}{\partial \tilde{\delta}_k} \boldsymbol{\Psi}\left(\sum_{g=1}^{K} \tilde{\delta}_g\right)\right] \cdot\left(\sum_{i=1}^{N} \sum_{h=1}^{K} \tilde{\pi}_{ih}\right)-\boldsymbol{\Psi}\left(\tilde{\delta}_k\right)+\left(\delta_k-\tilde{\delta}_k\right) \frac{\partial}{\partial \tilde{\delta}_k} \boldsymbol{\Psi}\left(\tilde{\delta}_k\right) \\
&-\left[\frac{\partial}{\partial \tilde{\delta}_k} \boldsymbol{\Psi}\left(\sum_{g=1}^{K} \tilde{\delta}_g\right)\right] \cdot\left[\sum_{h=1}^{K}\left(\delta_h-\tilde{\delta}_h\right)\right]+\boldsymbol{\Psi}\left(\sum_{g=1}^{K} \tilde{\delta}_g\right)+\boldsymbol{\Psi}\left(\tilde{\delta}_k\right)-\boldsymbol{\Psi}\left(\sum_{g=1}^{K} \tilde{\delta}_g\right) \\
&=\left[\frac{\partial}{\partial \tilde{\delta}_k} \boldsymbol{\Psi}\left(\tilde{\delta}_k\right)\right]\left(\delta_k-\tilde{\delta}_k+\sum_{i=1}^{N} \tilde{\pi}_{ik}\right)-\left[\frac{\partial}{\partial \tilde{\delta}_k} \boldsymbol{\Psi}\left(\sum_{g=1}^{K} \tilde{\delta}_g\right)\right] \cdot\left\{\left[\sum_{h=1}^{K}\left(\delta_h-\tilde{\delta}_h\right)\right]+\left(\sum_{i=1}^{N} \sum_{h=1}^{K} \tilde{\pi}_{ih}\right)\right\} \\
&=\boldsymbol{\Psi}'\left(\tilde{\delta}_k\right) \cdot\left(\delta_k-\tilde{\delta}_k+\sum_{i=1}^{N} \tilde{\pi}_{ik}\right)-\boldsymbol{\Psi}'\left(\sum_{g=1}^{K} \tilde{\delta}_g\right) \cdot\left[\sum_{g=1}^{K}\left(\delta_g-\tilde{\delta}_g+\sum_{i=1}^{N} \tilde{\pi}_{ig}\right)\right] .
\end{aligned}
$$

˜令 $\exp(\tilde{\alpha}_k)=\delta_k$ 且令

$$
\mathcal{H}\left(\tilde{\alpha}_1, \tilde{\alpha}_2, \ldots, \tilde{\alpha}_K, \ldots\right)=\mathcal{F}\left(\exp(\tilde{\alpha}_1), \exp(\tilde{\alpha}_2), \ldots, \exp(\tilde{\alpha}_K), \ldots\right)=\mathcal{F}\left(\tilde{\delta}_1, \tilde{\delta}_2, \ldots, \tilde{\delta}_K, \ldots\right),
$$

then

$$
\begin{aligned}
&\frac{\partial}{\partial \tilde{\alpha}_k} \mathcal{H}\left(\tilde{\alpha}_1, \tilde{\alpha}_2, \ldots, \tilde{\alpha}_K, \ldots\right)=\frac{\partial}{\partial \tilde{\alpha}_k} \mathcal{F}\left(\exp(\tilde{\alpha}_1), \exp(\tilde{\alpha}_2), \ldots, \exp(\tilde{\alpha}_K), \ldots\right) \\
&=\frac{\partial}{\partial \tilde{\delta}_k} \mathcal{F}\left(\tilde{\delta}_1, \tilde{\delta}_2, \ldots, \tilde{\delta}_K, \ldots\right) \frac{\partial \tilde{\delta}_k}{\partial \tilde{\alpha}_k}=\exp(\tilde{\alpha}_k) \frac{\partial}{\partial \tilde{\delta}_k} \mathcal{F}\left(\tilde{\delta}_1, \tilde{\delta}_2, \ldots, \tilde{\delta}_K, \ldots\right)
\end{aligned}
$$

现在我们首先证明，与似然相关的费希尔信息，

$$
f_{\mathrm{Dirichlet}}(\boldsymbol{\Pi} | \exp(\tilde{\alpha}_1), \exp(\tilde{\alpha}_2), \ldots, \exp(\tilde{\alpha}_K)),
$$

并非对角矩阵的形式。尽管它是对称矩阵，费舍尔信息矩阵的逆矩阵在计算上仍然开销很大。因此，我们最终改为考虑使用˜标准梯度上升来更新每个 δ_k，其中 $k=1$，2、.....、K。其对数为

上面的似然如下:

$$\log q_{\mathbf{\Pi}}\left(\mathbf{\Pi}|\exp(\tilde{\alpha}_1),\exp(\tilde{\alpha}_2),\ldots,\exp(\tilde{\alpha}_K)\right)=\log\frac{\Gamma\left(\sum_{k=1}^{K}\exp(\tilde{\alpha}_k)\right)\prod_{k=1}^{K}\pi_k^{\exp(\tilde{\alpha}_k)-1}}{\prod_{k=1}^{K}\Gamma\left(\exp(\tilde{\alpha}_k)\right)}$$

$$=\log\Gamma\left(\sum_{k=1}^{K}\exp(\tilde{\alpha}_k)\right)-\sum_{k=1}^{K}\log\Gamma\left(\exp(\tilde{\alpha}_k)\right)+\sum_{k=1}^{K}\left[\exp(\tilde{\alpha}_k)-1\right]\log(\pi_k).$$

˜关于 δ_g 的一阶偏导数，其中 $g=1$，2、.....、K，形式如下 :

$$\frac{\partial q_{\mathbf{\Pi}}}{\partial\tilde{\alpha}_g}=\Psi\left[\sum_{k=1}^{K}\exp(\tilde{\alpha}_k)\right]\exp(\tilde{\alpha}_g)-\Psi\left[\exp(\tilde{\alpha}_g)\right]\exp(\tilde{\alpha}_g)+\exp(\tilde{\alpha}_g)\log(\pi_g)$$

$$=\left\{\Psi\left[\sum_{k=1}^{K}\exp(\tilde{\alpha}_k)\right]-\Psi\left[\exp(\tilde{\alpha}_g)\right]+\log(\pi_g)\right\}\exp(\tilde{\alpha}_g).$$

$g=1$、2、.....、K的二阶偏导数形式如下:

$$\frac{\partial^2 q_{\mathbf{\Pi}}}{\partial\tilde{\alpha}_g^2}=\left\{\Psi'\left[\sum_{k=1}^{K}\exp(\tilde{\alpha}_k)\right]\exp(\tilde{\alpha}_g)-\Psi'\left[\exp(\tilde{\alpha}_g)\right]\exp(\tilde{\alpha}_g)\right\}\exp(\tilde{\alpha}_g)$$

$$+\left\{\Psi\left[\sum_{k=1}^{K}\exp(\tilde{\alpha}_k)\right]-\Psi\left[\exp(\tilde{\alpha}_g)\right]+\log(\pi_g)\right\}\exp(\tilde{\alpha}_g)$$

$$=\left\{\Psi'\left[\sum_{k=1}^{K}\exp(\tilde{\alpha}_k)\right]-\Psi'\left[\exp(\tilde{\alpha}_g)\right]\right\}\exp(\tilde{\alpha}_g)^2$$

$$+\left\{\Psi\left[\sum_{k=1}^{K}\exp(\tilde{\alpha}_k)\right]-\Psi\left[\exp(\tilde{\alpha}_g)\right]+\log(\pi_g)\right\}\exp(\tilde{\alpha}_g),$$

而对于 g、$h=1$、2、.....、K; $g\neq h$，我们有

$$\frac{\partial^2 q_{\mathbf{\Pi}}}{\partial\tilde{\alpha}_g\partial\tilde{\alpha}_h}=\Psi'\left[\sum_{k=1}^{K}\exp(\tilde{\alpha}_k)\right]\exp(\tilde{\alpha}_h)\exp(\tilde{\alpha}_g)$$

对上述两个方程取负期望 $-\mathbb{E}_q$，得到:

$$-\mathbb{E}_q\left(\frac{\partial^2 q_{\mathbf{\Pi}}}{\partial\tilde{\alpha}_g^2}\right)=-\left\{\Psi'\left[\sum_{k=1}^{K}\exp(\tilde{\alpha}_k)\right]-\Psi'\left[\exp(\tilde{\alpha}_g)\right]\right\}\exp(\tilde{\alpha}_g)^2$$

$$-\left\{\Psi\left[\sum_{k=1}^{K}\exp(\tilde{\alpha}_k)\right]-\Psi\left[\exp(\tilde{\alpha}_g)\right]+\mathbb{E}_q\left[\log(\pi_g)\right]\right\}\exp(\tilde{\alpha}_g)$$

$$=-\left\{\Psi'\left[\sum_{k=1}^{K}\exp(\tilde{\alpha}_k)\right]-\Psi'\left[\exp(\tilde{\alpha}_g)\right]\right\}\exp(\tilde{\alpha}_g)^2$$

$$-\left\{\Psi\left[\sum_{k=1}^{K}\exp(\tilde{\alpha}_k)\right]-\Psi\left[\exp(\tilde{\alpha}_g)\right]+\Psi\left[\exp(\tilde{\alpha}_g)\right]-\Psi\left[\sum_{k=1}^{K}\exp(\tilde{\alpha}_k)\right]\right\}\exp(\tilde{\alpha}_g)$$

$$=\left\{\Psi'\left[\exp(\tilde{\alpha}_g)\right]-\Psi'\left[\sum_{k=1}^{K}\exp(\tilde{\alpha}_k)\right]\right\}\exp(\tilde{\alpha}_g)^2.$$

同理，

$$-\mathbb{E}_q\left(\frac{\partial^2 q_{\boldsymbol{\Pi}}}{\partial\tilde{\alpha}_g\partial\tilde{\alpha}_h}\right)=-\Psi'\left[\sum_{k=1}^{K}\exp(\tilde{\alpha}_k)\right]\exp(\tilde{\alpha}_h)\exp(\tilde{\alpha}_g).$$

因此，费希尔信息 $\mathcal{I}(\tilde{\alpha}_1,\ \tilde{\alpha}_2,\ \ldots,\ \tilde{\alpha}_K)$ 是一个 $K\times K$矩阵，其每个对角元素均为 $-\mathbb{E}_q\left(\frac{\partial^2 q_{\boldsymbol{\Pi}}}{\partial\tilde{\alpha}_g^2}\right)$，适用于 $g=1,\ 2,\ \ldots,\ K$；每个非对角元素均为 $-\mathbb{E}_q\left(\frac{\partial^2 q_{\boldsymbol{\Pi}}}{\partial\tilde{\alpha}_g\partial\tilde{\alpha}_h}\right)$，适用于$g,\ h=1,\ 2,\ \ldots,\ K;\quad g\neq h$。可以预见，计算逆费希尔信息的计算量很大，因此我们改为考虑用于推断的标准梯度上升更新：

$$\tilde{\delta}_k^{\star}=\tilde{\delta}_k^{\dagger}\exp\left[\epsilon\tilde{\delta}_k^{\dagger}\frac{\partial\mathcal{F}}{\partial\tilde{\delta}_k}\left(\tilde{\delta}_1^{\dagger},\tilde{\delta}_2^{\dagger},\ldots,\tilde{\delta}_K^{\dagger},\ldots\right)\right],$$

适用于 $k=1,\ 2,\ \ldots,\ K$。 □

A.3 MLPCM 的 PICL（命题 6 的证明）

$\hat{}$ $\hat{}$请注意，给定参数 $\boldsymbol{U}$，$\boldsymbol{V}$，$\hat{}\boldsymbol{z}$ 的点估计后，PICL 准则基于积分对数似然：

$$\begin{aligned}
&\log p(\boldsymbol{Y},\hat{\boldsymbol{U}},\hat{\boldsymbol{V}},\hat{\boldsymbol{z}}|K)=\log p(\boldsymbol{Y}|\hat{\boldsymbol{U}},\hat{\boldsymbol{V}},\hat{\boldsymbol{z}},K)+\log p(\hat{\boldsymbol{U}},\hat{\boldsymbol{V}},\hat{\boldsymbol{z}}|K)\\
&=\log p(\boldsymbol{Y}|\hat{\boldsymbol{U}},\hat{\boldsymbol{V}})+\log p(\hat{\boldsymbol{V}}|\hat{\boldsymbol{U}},\hat{\boldsymbol{z}},K)+\log p(\hat{\boldsymbol{U}},\hat{\boldsymbol{z}}|K)\\
&=\log p(\boldsymbol{Y}|\hat{\boldsymbol{U}},\hat{\boldsymbol{V}})+\log p(\hat{\boldsymbol{V}}|\hat{\boldsymbol{U}})+\log p(\hat{\boldsymbol{U}}|\hat{\boldsymbol{z}})+\log p(\hat{\boldsymbol{z}}|K)\\
&=\log\int_{\beta}p(\boldsymbol{Y}|\hat{\boldsymbol{U}},\hat{\boldsymbol{V}},\beta)\pi(\beta)\mathrm{d}\beta+\log\int_{\boldsymbol{\gamma}}p(\hat{\boldsymbol{V}}|\hat{\boldsymbol{U}},\boldsymbol{\gamma})\pi(\boldsymbol{\gamma})\mathrm{d}\boldsymbol{\gamma}\\
&+\log\int_{\boldsymbol{\tau}}\int_{\boldsymbol{\mu}}p(\hat{\boldsymbol{U}}|\hat{\boldsymbol{z}},\boldsymbol{\mu},\boldsymbol{\tau})\pi(\boldsymbol{\mu})\pi(\boldsymbol{\tau})\mathrm{d}\boldsymbol{\mu}\mathrm{d}\boldsymbol{\tau}+\log\int_{\boldsymbol{\Pi}}p(\hat{\boldsymbol{z}}|\boldsymbol{\Pi})\pi(\boldsymbol{\Pi}|K)\mathrm{d}\boldsymbol{\Pi},
\end{aligned}$$

其中，我们对第 1 个对数 $p(\boldsymbol{Y}|\hat{}\boldsymbol{U},\hat{}\boldsymbol{V})$ 项应用 BIC近似，对第 2 个对数 $p(\hat{}\boldsymbol{V}|\hat{}\boldsymbol{U})$ 项和第 4 个对数 $p(\hat{}\boldsymbol{z}|K)$ 项应用直接边缘化。对第 3 个对数 $p(\hat{}\boldsymbol{U}|\hat{}\boldsymbol{z})$ 项应用分部积分与 BIC近似相结合的方法。

A.3.1 对数 $p(\boldsymbol{Y}|\hat{\boldsymbol{U}},\hat{\boldsymbol{V}})$的 BIC近似

$$\log p(\boldsymbol{Y}|\hat{\boldsymbol{U}},\hat{\boldsymbol{V}}) = \log \int_{\beta} p(\boldsymbol{Y}|\hat{\boldsymbol{U}},\hat{\boldsymbol{V}},\beta)\pi(\beta)\mathrm{d}\beta \overset{\text{BIC}}{\approx} \max_{\beta}\Big[\log p(\boldsymbol{Y}|\hat{\boldsymbol{U}},\hat{\boldsymbol{V}},\beta)\Big] - \frac{1}{2}\log[N(N-1)]$$

$$= \max_{\beta}\left[\log \prod_{\substack{i,j=1,\\ i\neq j}}^{N} f_{\text{Pois}}\left(y_{ij}|\exp\big(\beta - ||\hat{\boldsymbol{u}}_i - \hat{\boldsymbol{v}}_{j\leftarrow i}||^2\big)\right)\right] - \frac{1}{2}\log[N(N-1)]$$

$$= \max_{\beta}\left[\log \prod_{\substack{i,j=1,\\ i\neq j}}^{N} \frac{\big[\exp\big(\beta - ||\hat{\boldsymbol{u}}_i - \hat{\boldsymbol{v}}_{j\leftarrow i}||^2\big)\big]^{y_{ij}}\exp\big[-\exp\big(\beta - ||\hat{\boldsymbol{u}}_i - \hat{\boldsymbol{v}}_{j\leftarrow i}||^2\big)\big]}{y_{ij}!}\right] - \frac{1}{2}\log[N(N-1)]$$

$$= \max_{\beta}\sum_{\substack{i,j=1,\\ i\neq j}}^{N}\left[y_{ij}\left(\beta - ||\hat{\boldsymbol{u}}_i - \hat{\boldsymbol{v}}_{j\leftarrow i}||^2\right) - \exp\big(\beta - ||\hat{\boldsymbol{u}}_i - \hat{\boldsymbol{v}}_{j\leftarrow i}||^2\big) - \log(y_{ij}!)\right] - \frac{1}{2}\log[N(N-1)]$$

$$= \max_{\beta}\left[\beta\left(\sum_{\substack{i,j=1,\\ i\neq j}}^{N} y_{ij}\right) - \sum_{\substack{i,j=1,\\ i\neq j}}^{N}\exp\big(\beta - ||\hat{\boldsymbol{u}}_i - \hat{\boldsymbol{v}}_{j\leftarrow i}||^2\big)\right] - \sum_{\substack{i,j=1,\\ i\neq j}}^{N}\left[y_{ij}||\hat{\boldsymbol{u}}_i - \hat{\boldsymbol{v}}_{j\leftarrow i}||^2 + \log(y_{ij}!)\right]$$

$$- \frac{1}{2}\log[N(N-1)].$$

在此，定义

$$g(\beta) := \beta\left(\sum_{\substack{i,j=1,\\ i\neq j}}^{N} y_{ij}\right) - \sum_{\substack{i,j=1,\\ i\neq j}}^{N}\exp\big(\beta - ||\hat{\boldsymbol{u}}_i - \hat{\boldsymbol{v}}_{j\leftarrow i}||^2\big),$$

该 $g(\beta)$ 函数的最大化有唯一的解析解，其中对 β 求一阶导数可得:

$$\frac{\partial g(\beta)}{\partial \beta} = \left(\sum_{\substack{i,j=1,\\ i\neq j}}^{N} y_{ij}\right) - \exp(\beta)\sum_{\substack{i,j=1,\\ i\neq j}}^{N}\exp\big(-||\hat{\boldsymbol{u}}_i - \hat{\boldsymbol{v}}_{j\leftarrow i}||^2\big) := 0$$

$$\implies \beta = \log\left(\sum_{\substack{i,j=1,\\ i\neq j}}^{N} y_{ij}\right) - \log\left[\sum_{\substack{i,j=1,\\ i\neq j}}^{N}\exp\big(-||\hat{\boldsymbol{u}}_i - \hat{\boldsymbol{v}}_{j\leftarrow i}||^2\big)\right].$$

可以验证，相对于 β 的二阶导数为:

$$\frac{\partial^2 g(\beta)}{\partial \beta^2} = -\exp(\beta)\sum_{\substack{i,j=1,\\ i\neq j}}^{N}\exp\big(-||\hat{\boldsymbol{u}}_i - \hat{\boldsymbol{v}}_{j\leftarrow i}||^2\big) < 0.$$

这确保了在上述得到的 β 处取得全局最大值。代回 PICL 中的对数 $p(\boldsymbol{Y}|\hat{\boldsymbol{U}},\hat{\boldsymbol{V}})$ 项可得：

$$
\begin{aligned}
\log p(\boldsymbol{Y}|\hat{\boldsymbol{U}},\hat{\boldsymbol{V}}) &\overset{\text{BIC}}{\approx} \max_{\beta}\left[\beta\left(\sum_{\substack{i,j=1,\\ i\neq j}}^{N} y_{ij}\right) - \sum_{\substack{i,j=1,\\ i\neq j}}^{N}\exp\left(\beta - ||\hat{\boldsymbol{u}}_i - \hat{\boldsymbol{v}}_{j\leftarrow i}||^2\right)\right] \\
&- \sum_{\substack{i,j=1,\\ i\neq j}}^{N}\left[y_{ij}||\hat{\boldsymbol{u}}_i - \hat{\boldsymbol{v}}_{j\leftarrow i}||^2 + \log(y_{ij}!)\right] - \frac{1}{2}\log[N(N-1)] \\
&= \left\{\log\left(\sum_{\substack{i,j=1,\\ i\neq j}}^{N} y_{ij}\right) - \log\left[\sum_{\substack{i,j=1,\\ i\neq j}}^{N}\exp\left(-||\hat{\boldsymbol{u}}_i - \hat{\boldsymbol{v}}_{j\leftarrow i}||^2\right)\right]\right\}\left(\sum_{\substack{i,j=1,\\ i\neq j}}^{N} y_{ij}\right) \\
&- \frac{\sum_{\substack{i,j=1,\\ i\neq j}}^{N} y_{ij}}{\sum_{\substack{i,j=1,\\ i\neq j}}^{N}\exp(-||\hat{\boldsymbol{u}}_i - \hat{\boldsymbol{v}}_{j\leftarrow i}||^2)}\sum_{\substack{i,j=1,\\ i\neq j}}^{N}\exp\left(-||\hat{\boldsymbol{u}}_i - \hat{\boldsymbol{v}}_{j\leftarrow i}||^2\right) \\
&- \sum_{\substack{i,j=1,\\ i\neq j}}^{N}\left[y_{ij}||\hat{\boldsymbol{u}}_i - \hat{\boldsymbol{v}}_{j\leftarrow i}||^2 + \log(y_{ij}!)\right] - \frac{1}{2}\log[N(N-1)] \\
&= \left\{\log\left(\sum_{\substack{i,j=1,\\ i\neq j}}^{N} y_{ij}\right) - \log\left[\sum_{\substack{i,j=1,\\ i\neq j}}^{N}\exp\left(-||\hat{\boldsymbol{u}}_i - \hat{\boldsymbol{v}}_{j\leftarrow i}||^2\right)\right] - 1\right\}\left(\sum_{\substack{i,j=1,\\ i\neq j}}^{N} y_{ij}\right) \\
&- \sum_{\substack{i,j=1,\\ i\neq j}}^{N}\left[y_{ij}||\hat{\boldsymbol{u}}_i - \hat{\boldsymbol{v}}_{j\leftarrow i}||^2 + \log(y_{ij}!)\right] - \frac{1}{2}\log[N(N-1)] \\
&= \texttt{const} - \left(\sum_{\substack{i,j=1,\\ i\neq j}}^{N} y_{ij}\right)\log\left[\sum_{\substack{i,j=1,\\ i\neq j}}^{N}\exp\left(-||\hat{\boldsymbol{u}}_i - \hat{\boldsymbol{v}}_{j\leftarrow i}||^2\right)\right] \\
&- \sum_{\substack{i,j=1,\\ i\neq j}}^{N}\left(y_{ij}||\hat{\boldsymbol{u}}_i - \hat{\boldsymbol{v}}_{j\leftarrow i}||^2\right).
\end{aligned}
$$

□

A.3.2 对数 $p(\hat{\boldsymbol{V}}|\hat{\boldsymbol{U}})$积分项

$$
\begin{aligned}
&\log p(\hat{\boldsymbol{V}}|\hat{\boldsymbol{U}}) = \log \int_{\boldsymbol{\gamma}} p(\hat{\boldsymbol{V}}|\hat{\boldsymbol{U}}, \boldsymbol{\gamma}) \pi(\boldsymbol{\gamma}) \mathrm{d}\boldsymbol{\gamma} \\
&= \log \int_{\boldsymbol{\gamma}} \prod_{j=1}^{N} \left\{ \left[\prod_{\substack{i=1,\\ i\neq j}}^{N} f_{\mathrm{MVN}_d}\left(\hat{\boldsymbol{v}}_{j\leftarrow i} | \hat{\boldsymbol{u}}_j, 1/\gamma_j \mathbb{I}_d\right) \right] f_{\mathrm{Ga}}\left(\gamma_j | a, b\right) \right\} \mathrm{d}\boldsymbol{\gamma} \\
&= \log \int_{\boldsymbol{\gamma}} \prod_{j=1}^{N} \left\{ \left[\prod_{\substack{i=1,\\ i\neq j}}^{N} \frac{|\gamma_j \mathbb{I}_d|^{1/2}}{(2\pi)^{d/2}} \exp\left[-\frac{1}{2} \left(\hat{\boldsymbol{v}}_{j\leftarrow i} - \hat{\boldsymbol{u}}_j\right) \gamma_j \mathbb{I}_d \left(\hat{\boldsymbol{v}}_{j\leftarrow i} - \hat{\boldsymbol{u}}_j\right)^T \right] \right] f_{\mathrm{Ga}}\left(\gamma_j | a, b\right) \right\} \mathrm{d}\boldsymbol{\gamma} \\
&= \log \int_{\boldsymbol{\gamma}} \prod_{j=1}^{N} \left\{ (2\pi)^{-\frac{d}{2}(N-1)} \gamma_j^{\frac{d}{2}(N-1)} \exp\left[-\frac{1}{2} \gamma_j \sum_{\substack{i=1,\\ i\neq j}}^{N} ||\hat{\boldsymbol{v}}_{j\leftarrow i} - \hat{\boldsymbol{u}}_j||^2 \right] \frac{b^a}{\Gamma(a)} \gamma_j^{a-1} \exp(-b\gamma_j) \right\} \mathrm{d}\boldsymbol{\gamma} \\
&= \log \prod_{j=1}^{N} \left\{ (2\pi)^{-\frac{d}{2}(N-1)} \frac{b^a}{\Gamma(a)} \int_{\gamma_j} \gamma_j^{a+\frac{d}{2}(N-1)-1} \exp\left[-\left(b + \frac{1}{2} \sum_{\substack{i=1,\\ i\neq j}}^{N} ||\hat{\boldsymbol{v}}_{j\leftarrow i} - \hat{\boldsymbol{u}}_j||^2 \right) \gamma_j \right] \mathrm{d}\gamma_j \right\} \\
&= \log \prod_{j=1}^{N} \left[(2\pi)^{-\frac{d}{2}(N-1)} \frac{b^a}{\Gamma(a)} \frac{\Gamma\left(a + \frac{d}{2}(N-1)\right)}{\left(b + \frac{1}{2} \sum_{\substack{i=1,\\ i\neq j}}^{N} ||\hat{\boldsymbol{v}}_{j\leftarrow i} - \hat{\boldsymbol{u}}_j||^2 \right)^{a+\frac{d}{2}(N-1)}} \right] \\
&= \sum_{j=1}^{N} \left[-\frac{d}{2}(N-1)\log(2\pi) + a\log(b) - \log\Gamma(a) + \log\Gamma\left(a + \frac{d}{2}(N-1)\right) \right] \\
&- \sum_{j=1}^{N} \left[a + \frac{d}{2}(N-1) \right] \log\left(b + \frac{1}{2} \sum_{\substack{i=1,\\ i\neq j}}^{N} ||\hat{\boldsymbol{v}}_{j\leftarrow i} - \hat{\boldsymbol{u}}_j||^2 \right) \\
&= \texttt{const} - \left[a + \frac{d}{2}(N-1) \right] \sum_{j=1}^{N} \log\left(b + \frac{1}{2} \sum_{\substack{i=1,\\ i\neq j}}^{N} ||\hat{\boldsymbol{v}}_{j\leftarrow i} - \hat{\boldsymbol{u}}_j||^2 \right).
\end{aligned}
$$

□

A.3.3 + 对对数 $p(\hat{U}|\hat{z})$的 BIC近似分部积分

$$\log p(\hat{\boldsymbol{U}}|\hat{\boldsymbol{z}}) = \log \int_{\boldsymbol{\tau}} \int_{\boldsymbol{\mu}} p(\hat{\boldsymbol{U}}|\hat{\boldsymbol{z}}, \boldsymbol{\mu}, \boldsymbol{\tau}) \pi(\boldsymbol{\mu}) \mathrm{d}\boldsymbol{\mu}\ \pi(\boldsymbol{\tau}) \mathrm{d}\boldsymbol{\tau}$$

$$= \log \int_{\boldsymbol{\tau}} \int_{\boldsymbol{\mu}} \prod_{g=1}^{K} \left\{ \left[\prod_{\substack{i=1,\\ \hat{z}_i=g}}^{N} f_{\mathrm{MVN}_d}\left(\hat{\boldsymbol{u}}_i | \boldsymbol{\mu}_g, 1/\tau_g \mathbb{I}_d\right) \right] f_{\mathrm{MVN}_d}\left(\boldsymbol{\mu}_g | \mathbf{0}, \omega^2 \mathbb{I}_d\right) \right\} \mathrm{d}\boldsymbol{\mu}\ \pi(\boldsymbol{\tau}) \mathrm{d}\boldsymbol{\tau}$$

$$= \log \int_{\boldsymbol{\tau}} \left\{ \prod_{g=1}^{K} \int_{\boldsymbol{\mu}_g} \left[\prod_{\substack{i=1,\\ \hat{z}_i=g}}^{N} \frac{|\tau_g \mathbb{I}_d|^{1/2}}{(2\pi)^{d/2}} \exp\Big(-\frac{\tau_g}{2} ||\hat{\boldsymbol{u}}_i - \boldsymbol{\mu}_g||^2\Big) \right] f_{\mathrm{MVN}_d}\left(\boldsymbol{\mu}_g | \mathbf{0}, \omega^2 \mathbb{I}_d\right) \mathrm{d}\boldsymbol{\mu}_g \right\} \pi(\boldsymbol{\tau}) \mathrm{d}\boldsymbol{\tau}$$

$$= \log \int_{\boldsymbol{\tau}} \left\{ \prod_{g=1}^{K} \int_{\boldsymbol{\mu}_g} \left[\left(\frac{\tau_g}{2\pi}\right)^{\frac{d}{2}\hat{n}_g} \exp\left(-\frac{\tau_g}{2} \sum_{\substack{i=1,\\ \hat{z}_i=g}}^{N} ||\hat{\boldsymbol{u}}_i - \boldsymbol{\mu}_g||^2 \right) \right] \frac{\exp\big(-\frac{1}{2\omega^2}||\boldsymbol{\mu}_g||^2\big)}{(2\pi\omega^2)^{\frac{d}{2}}} \mathrm{d}\boldsymbol{\mu}_g \right\} \pi(\boldsymbol{\tau}) \mathrm{d}\boldsymbol{\tau}$$

$$= \log \int_{\boldsymbol{\tau}} \prod_{g=1}^{K} \frac{\int_{\boldsymbol{\mu}_g} \exp\left[-\frac{\tau_g}{2} \left(\left(\sum_{\substack{i=1,\\ \hat{z}_i=g}} ||\hat{\boldsymbol{u}}_i||^2 \right) - 2 \left(\sum_{\substack{i=1,\\ \hat{z}_i=g}}^{N} \hat{\boldsymbol{u}}_i^T \right) \boldsymbol{\mu}_g + \hat{n}_g ||\boldsymbol{\mu}_g||^2 \right) - \frac{1}{2\omega^2} ||\boldsymbol{\mu}_g||^2 \right] \mathrm{d}\boldsymbol{\mu}_g}{\left(\frac{\tau_g}{2\pi}\right)^{-\frac{d}{2}\hat{n}_g} (2\pi\omega^2)^{\frac{d}{2}}} \pi(\boldsymbol{\tau}) \mathrm{d}\boldsymbol{\tau}$$

$$= \log \int_{\boldsymbol{\tau}} \prod_{g=1}^{K} \frac{\int_{\boldsymbol{\mu}_g} \exp\left[-\frac{1}{2} \left(\tau_g \hat{n}_g + \frac{1}{\omega^2} \right) ||\boldsymbol{\mu}_g||^2 + 2\frac{\tau_g}{2} \left(\sum_{\substack{i=1,\\ \hat{z}_i=g}}^{N} \hat{\boldsymbol{u}}_i^T \right) \boldsymbol{\mu}_g - \frac{\tau_g}{2} \left(\sum_{\substack{i=1,\\ \hat{z}_i=g}}^{N} ||\hat{\boldsymbol{u}}_i||^2 \right) \right] \mathrm{d}\boldsymbol{\mu}_g}{\left(\frac{\tau_g}{2\pi}\right)^{-\frac{d}{2}\hat{n}_g} (2\pi\omega^2)^{\frac{d}{2}}} \pi(\boldsymbol{\tau}) \mathrm{d}\boldsymbol{\tau}$$

$$= \log \int_{\boldsymbol{\tau}} \prod_{g=1}^{K} \frac{\int_{\boldsymbol{\mu}_g} \exp\left[-\frac{1}{2} \left(\tau_g \hat{n}_g + \frac{1}{\omega^2} \right) \left\| \boldsymbol{\mu}_g - \frac{\tau_g}{\tau_g \hat{n}_g + \frac{1}{\omega^2}} \left(\sum_{\substack{i=1,\\ \hat{z}_i=g}}^{N} \hat{\boldsymbol{u}}_i \right) \right\|^2 \right] \mathrm{d}\boldsymbol{\mu}_g}{\left(\frac{\tau_g}{2\pi}\right)^{-\frac{d}{2}\hat{n}_g} (2\pi\omega^2)^{\frac{d}{2}}}$$

$$\prod_{g=1}^{K} \exp\left[\frac{1}{2} \left(\tau_g \hat{n}_g + \frac{1}{\omega^2} \right) \left(\frac{\tau_g}{\tau_g \hat{n}_g + \frac{1}{\omega^2}} \right)^2 \left\| \sum_{\hat{z}_i=g} \hat{\boldsymbol{u}}_i \right\|^2 - \frac{\tau_g}{2} \left(\sum_{\hat{z}_i=g} ||\hat{\boldsymbol{u}}_i||^2 \right) \right] \pi(\boldsymbol{\tau}) \mathrm{d}\boldsymbol{\tau}$$

$$= \log \int_{\boldsymbol{\tau}} \prod_{g=1}^{K} \frac{\exp\left[\frac{1}{2} \frac{\tau_g^2}{\tau_g \hat{n}_g + \frac{1}{\omega^2}} \left\| \sum_{\hat{z}_i=g} \hat{\boldsymbol{u}}_i \right\|^2 - \frac{\tau_g}{2} \left(\sum_{\hat{z}_i=g} ||\hat{\boldsymbol{u}}_i||^2 \right) \right]}{\left(\frac{\tau_g}{2\pi}\right)^{-\frac{d}{2}\hat{n}_g} (\tau_g \hat{n}_g \omega^2 + 1)^{\frac{d}{2}}} \pi(\boldsymbol{\tau}) \mathrm{d}\boldsymbol{\tau}$$

$$\overset{\text{BIC}}{\approx} \max_{\boldsymbol{\tau}} \left\{ \log \prod_{g=1}^{K} \frac{\exp\left[\frac{1}{2} \frac{\tau_g^2 \omega^2}{\tau_g \hat{n}_g \omega^2 + 1} \left\| \sum_{\hat{z}_i=g} \hat{\boldsymbol{u}}_i \right\|^2 - \frac{\tau_g}{2} \left(\sum_{\hat{z}_i=g} ||\hat{\boldsymbol{u}}_i||^2 \right) \right]}{\left(\frac{\tau_g}{2\pi}\right)^{-\frac{d}{2}\hat{n}_g} (\tau_g \hat{n}_g \omega^2 + 1)^{\frac{d}{2}}} \right\} - \frac{1}{2} K \log(N)$$

$$= \max_{\boldsymbol{\tau}} \sum_{g=1}^{K} \left[\frac{d}{2} \hat{n}_g \log(\tau_g) - \frac{d}{2} \log\left(\tau_g \hat{n}_g \omega^2 + 1\right) + \frac{1}{2} \frac{\tau_g^2 \omega^2}{\tau_g \hat{n}_g \omega^2 + 1} \left\| \sum_{\hat{z}_i=g} \hat{\boldsymbol{u}}_i \right\|^2 - \frac{\tau_g}{2} \left(\sum_{\hat{z}_i=g} ||\hat{\boldsymbol{u}}_i||^2 \right) \right]$$

$$- \frac{d}{2} N \log(2\pi) - \frac{1}{2} K \log(N),$$

其中 $\hat{n}_g := \sum_{i=1}^{N} \mathbb{1}(\hat{z}_i = g)$。我们记

$$h(\boldsymbol{\tau}) = \sum_{g=1}^{K} \left[\frac{d}{2} \hat{n}_g \log(\tau_g) - \frac{d}{2} \log\left(\tau_g \hat{n}_g \omega^2 + 1\right) + \frac{1}{2} \frac{\tau_g^2 \omega^2}{\tau_g \hat{n}_g \omega^2 + 1} \left\| \sum_{\hat{z}_i = g} \hat{\boldsymbol{u}}_i \right\|^2 - \frac{\tau_g}{2} \left(\sum_{\hat{z}_i = g} ||\hat{\boldsymbol{u}}_i||^2 \right) \right],$$

并采用标准梯度上升进行最大化。首先求关于 τ_g 的一阶导数:

$$\begin{aligned} \frac{\partial h(\boldsymbol{\tau})}{\partial \tau_g} &= \frac{d\hat{n}_g}{2\tau_g} - \frac{d\hat{n}_g \omega^2}{2\left(\tau_g \hat{n}_g \omega^2 + 1\right)} + \frac{1}{2} \left[\frac{2\tau_g \omega^2}{\tau_g \hat{n}_g \omega^2 + 1} - \frac{\tau_g^2 \omega^2 \hat{n}_g \omega^2}{\left(\tau_g \hat{n}_g \omega^2 + 1\right)^2} \right] \left\| \sum_{\hat{z}_i = g} \hat{\boldsymbol{u}}_i \right\|^2 - \frac{1}{2} \left(\sum_{\hat{z}_i = g} ||\hat{\boldsymbol{u}}_i||^2 \right) \\ &= \frac{d}{2} \hat{n}_g \left(\frac{1}{\tau_g} - \frac{\omega^2}{\tau_g \hat{n}_g \omega^2 + 1} \right) + \frac{\tau_g \omega^2 \left(\tau_g \hat{n}_g \omega^2 + 2\right)}{2\left(\tau_g \hat{n}_g \omega^2 + 1\right)^2} \left\| \sum_{\hat{z}_i = g} \hat{\boldsymbol{u}}_i \right\|^2 - \frac{1}{2} \left(\sum_{\hat{z}_i = g} ||\hat{\boldsymbol{u}}_i||^2 \right), \end{aligned}$$

它是 τ_g 的函数。接着，我们按照标准梯度上升方案求出使其达到最大值的 $\{\tau_g : \ g = 1、\ldots\ldots、\ K\}$。然而，由于 $\tau_g > 0$ 对于 $g = 1、\ldots\ldots、\ K$，我们令 $\exp(r_g) = \tau_g$，并令 $l(r_1、\ldots\ldots、\ r_K) = h(\exp(r_1)\ldots\ldots, \exp(r_K)) = h(\tau_1、\ldots\ldots、\ \tau_K)$，则

$$\frac{\partial}{\partial r_g} l\left(r_1, \ldots, r_K\right) = \frac{\partial}{\partial r_g} h(\exp(r_1), \ldots, \exp(r_K)) = \frac{\partial}{\partial \tau_g} h(\tau_1, \ldots, \tau_K) \frac{\partial \tau_g}{\partial r_g} = \exp(r_g) \frac{\partial}{\partial \tau_g} h(\tau_1, \ldots, \tau_K).$$

由此得到如下标准梯度上升方案:

$$r_g^{\star} = r_g^{\dagger} + \epsilon \exp\left(r_g^{\dagger}\right) \frac{\partial h}{\partial \tau_g} \left(\tau_g^{\dagger}\right),$$

对于 $g = 1、\ldots\ldots、\ K$。对两边取 $\exp(\cdot)$，得到最终的标准梯度上升方案步骤:

$$\begin{aligned} &\exp\left(r_g^{\star}\right) = \exp\left[r_g^{\dagger} + \epsilon \exp\left(r_g^{\dagger}\right) \frac{\partial h}{\partial \tau_g} \left(\tau_g^{\dagger}\right) \right] \\ &\Longrightarrow \ \exp\left(r_g^{\star}\right) = \exp\left(r_g^{\dagger}\right) \exp\left[\epsilon \exp\left(r_g^{\dagger}\right) \frac{\partial h}{\partial \tau_g} \left(\tau_g^{(t)}\right) \right] \\ &\Longrightarrow \ \tau_g^{\star} = \tau_g^{\dagger} \exp\left[\epsilon \tau_g^{\dagger} \frac{\partial h}{\partial \tau_g} \left(\tau_g^{\dagger}\right) \right]. \end{aligned}$$

最后，将最大化 $h(\boldsymbol{\tau})$ 的所得 $\{\tau_g : \ g = 1, \ldots, K\}$ 代回 $h(\boldsymbol{\tau})$，即可得到最终的 PICL 准则值。□

A.3.4 对数 $p(\hat{\boldsymbol{z}}|K)$的积分项

$$\log p(\hat{\boldsymbol{z}}|K) = \log \int_{\boldsymbol{\Pi}} p(\hat{\boldsymbol{z}}|\boldsymbol{\Pi})\pi(\boldsymbol{\Pi}|K)\mathrm{d}\boldsymbol{\Pi} = \log \int_{\boldsymbol{\Pi}} \left[\prod_{i=1}^{N} f_{\mathrm{Cat}}(\hat{z}_i|\boldsymbol{\Pi})\right] f_{\mathrm{Dirichlet}}(\boldsymbol{\Pi}|\delta_1,\ldots,\delta_K)\mathrm{d}\boldsymbol{\Pi}$$

$$= \log \int_{\boldsymbol{\Pi}} \left(\prod_{k=1}^{K} \pi_k^{\hat{n}_k}\right) \cdot \frac{\left(\prod_{k=1}^{K} \pi_k^{\delta_k-1}\right)}{\mathrm{B}(\delta_1,\ldots,\delta_K)}\mathrm{d}\boldsymbol{\Pi} = \log \int_{\boldsymbol{\Pi}} \frac{1}{\mathrm{B}(\delta_1,\ldots,\delta_K)} \cdot \left(\prod_{k=1}^{K} \pi_k^{\hat{n}_k+\delta_k-1}\right) \mathrm{d}\boldsymbol{\Pi}$$

$$= \log \frac{\mathrm{B}(\hat{n}_1+\delta_1,\ldots,\hat{n}_K+\delta_K)}{\mathrm{B}(\delta_1,\ldots,\delta_K)} = \log \left[\frac{\prod_{k=1}^{K}\Gamma(\hat{n}_k+\delta_k)}{\Gamma\left(N+\sum_{k=1}^{K}\delta_k\right)} \frac{\Gamma\left(\sum_{k=1}^{K}\delta_k\right)}{\prod_{k=1}^{K}\Gamma(\delta_k)}\right]$$

$$= \left[\sum_{k=1}^{K} \log\Gamma(\hat{n}_k+\delta_k)\right] - \log\Gamma\left(N+\sum_{k=1}^{K}\delta_k\right) + \log\Gamma\left(\sum_{k=1}^{K}\delta_k\right) - \sum_{k=1}^{K}\log\Gamma(\delta_k).$$

当 $\delta_k = \delta \text{ for } k = 1,..,K$ 时，上述公式简化为:

$$\log p(\hat{\boldsymbol{z}}|K) = \left[\sum_{k=1}^{K} \log\Gamma(\hat{n}_k+\delta)\right] - \log\Gamma\left(N+K\delta\right) + \log\Gamma\left(K\delta\right) - K\log\Gamma(\delta).$$

□

A.4 泊松分布LPCM的证据下界推导

PoisLPCM 的证据下界具有如下形式:

$$\begin{aligned}
\mathcal{F}(\tilde{\boldsymbol{\Theta}}) &= \mathbb{E}_q[\log p(\boldsymbol{Y}|\boldsymbol{U},\beta)] + \mathbb{E}_q[\log p(\boldsymbol{U}|\boldsymbol{\mu},\boldsymbol{\tau},\boldsymbol{z})] + \mathbb{E}_q[\log p(\boldsymbol{z}|\boldsymbol{\Pi})] \\
&+ \mathbb{E}_q[\log \pi(\beta)] + \mathbb{E}_q[\log \pi(\boldsymbol{\mu})] + \mathbb{E}_q[\log \pi(\boldsymbol{\tau})] + \mathbb{E}_q[\log \pi(\boldsymbol{\Pi})] \\
&- \mathbb{E}_q[\log q_{\boldsymbol{U}}(\boldsymbol{U}|\tilde{\boldsymbol{U}},\tilde{\boldsymbol{\sigma}}^{\boldsymbol{2}})] - \mathbb{E}_q[\log q_{\boldsymbol{z}}(\boldsymbol{z}|\tilde{\boldsymbol{\Pi}})] - \mathbb{E}_q[\log q_{\beta}(\beta|\tilde{\eta},\tilde{\rho}^2)] \\
&- \mathbb{E}_q[\log q_{\boldsymbol{\mu}}(\boldsymbol{\mu}|\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\omega}}^{\boldsymbol{2}})] - \mathbb{E}_q[\log q_{\boldsymbol{\tau}}(\boldsymbol{\tau}|\tilde{\boldsymbol{\xi}},\tilde{\boldsymbol{\psi}})] - \mathbb{E}_q[\log q_{\boldsymbol{\Pi}}(\boldsymbol{\Pi}|\tilde{\boldsymbol{\delta}})],
\end{aligned}$$

除第一项$\mathbb{E}_q$[对数 $p(\boldsymbol{Y}|\boldsymbol{U},\beta)$] 项外，其余各项均已在附录 A.1 中得到证明。因此，此处只需给出该项的推导。

$$\begin{aligned}
\mathbb{E}_q[\log p(\boldsymbol{Y}|\boldsymbol{U},\beta)] &= \mathbb{E}_q\left[\log \prod_{\substack{i,j=1,\\ i\neq j}}^{N} f_{\text{Pois}}\left(y_{ij}\middle|\exp\left(\beta-||\boldsymbol{u}_i-\boldsymbol{u}_j||^2\right)\right)\right] \\
&= \mathbb{E}_q\left[\log \prod_{\substack{i,j=1,\\ i\neq j}}^{N} \frac{\left[\exp\left(\beta-||\boldsymbol{u}_i-\boldsymbol{u}_j||^2\right)\right]^{y_{ij}}\exp\left[-\exp\left(\beta-||\boldsymbol{u}_i-\boldsymbol{u}_j||^2\right)\right]}{y_{ij}!}\right] \\
&= \mathbb{E}_q\left[\sum_{\substack{i,j=1,\\ i\neq j}}^{N}\left[y_{ij}\left(\beta-||\boldsymbol{u}_i-\boldsymbol{u}_j||^2\right)-\exp\left(\beta-||\boldsymbol{u}_i-\boldsymbol{u}_j||^2\right)-\log(y_{ij}!)\right]\right] \\
&= \sum_{\substack{i,j=1,\\ i\neq j}}^{N}\left\{y_{ij}\left[\mathbb{E}_q(\beta)-\mathbb{E}_q\left(||\boldsymbol{u}_i-\boldsymbol{u}_j||^2\right)\right]-\mathbb{E}_q\left[\exp\left(\beta-||\boldsymbol{u}_i-\boldsymbol{u}_j||^2\right)\right]\right\}-\sum_{\substack{i,j=1,\\ i\neq j}}^{N}\log(y_{ij}!) \\
&= \sum_{\substack{i,j=1,\\ i\neq j}}^{N}\left\{y_{ij}\left[\tilde{\eta}-||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{u}}_j||^2-d\left(\tilde{\sigma}_i^2+\tilde{\sigma}_j^2\right)\right]-\frac{\exp\left(\tilde{\eta}+\frac{\tilde{\rho}^2}{2}-\frac{||\tilde{\boldsymbol{u}}_i-\tilde{\boldsymbol{u}}_j||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2)^{d/2}}\right\}+\texttt{const},
\end{aligned}$$

这里的计算与附录 A.1.1 中所示的计算类似。 □

A.5 泊松潜在位置聚类模型

我们通过与泊松潜在位置聚类模型（PoisLPCM）的性能进行比较，展示我们提出的新型 MLPCM 的性能。我们遵循 Lu 等（2026）的做法，将 Handcock 等（2007）提出的二元潜在位置聚类模型扩展到非负离散加权网络：对于i，$j=1,\ldots,N$；$i\neq j$，有

$$y_{ij}\sim \text{Pois}(\lambda_{ij}),\ \log(\lambda_{ij})=\beta-||\boldsymbol{u}_i-\boldsymbol{u}_j||^2,$$

其中，$\boldsymbol{u}_i|z_i=g$和 z_i遵循式（1），适用于 $i=1$，$2,\ldots,N$。因此，对于 i，$j=1,\ldots,N$；$i\neq j$，交互y_{ij}和y_{ji}尽管方向不同，却被假定服从相同的分布。移除算法 1 的第 3 步且不输入 $\boldsymbol{\gamma}$，并将第 4 步替换为其中的y_{ij} ~Pois $\left[\exp\left(\beta-||\boldsymbol{u}_i-\boldsymbol{u}_j||^2\right)\right]$，即可得到从 PoisLPCM 进行模拟的过程。本文对 PoisLPCM 的推断也基于变分贝叶斯，其中 PoisLPCM 的证据下界与式（6）共享某些项，如下文命题 7 所示。

命题 7 给定遵循期望 (5) 的变分分布，不考虑 $\{\tilde{\boldsymbol{v}}_{j\leftarrow i}\}$ 和 $\{\gamma_i\}$，并令 $\tilde{\boldsymbol{\Phi}}$：$=\{\tilde{\boldsymbol{U}},\tilde{\boldsymbol{\sigma}}^{\boldsymbol{2}},\tilde{\boldsymbol{\Pi}},\tilde{\eta},\tilde{\rho}^2,\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\omega}}^{\boldsymbol{2}},\tilde{\boldsymbol{\xi}},\tilde{\psi},\tilde{\boldsymbol{\delta}}\}$，则该模型的证据下界为

PoisLPCM 写作：

$$
\begin{aligned}
\mathcal{F}(\tilde{\mathbf{\Phi}}) = & \sum_{\substack{i,j=1,\\ i\neq j}}^{N} \left\{ y_{ij}\left[\tilde{\eta} - ||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j||^2 - d\left(\tilde{\sigma}_i^2 + \tilde{\sigma}_j^2\right)\right] - \frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2)^{d/2}} \right\} \\
& + \sum_{i=1}^{N}\sum_{k=1}^{K} \tilde{\pi}_{ik} \left\{ \frac{d}{2}\left[\mathbf{\Psi}(\tilde{\xi}_k) - \log\left(\tilde{\psi}_k\right)\right] - \frac{1}{2}\frac{\tilde{\xi}_k}{\tilde{\psi}_k}\left[||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{\mu}}_k||^2 + d(\tilde{\sigma}_i^2 + \tilde{\omega}_k^2)\right] \right\} \\
& + \sum_{i=1}^{N}\sum_{k=1}^{K} \tilde{\pi}_{ik} \left[\mathbf{\Psi}\left(\tilde{\delta}_k\right) - \mathbf{\Psi}\left(\sum_{g=1}^{K}\tilde{\delta}_g\right) - \log(\tilde{\pi}_{ik}) \right] \\
& + \sum_{k=1}^{K} \left\{ (\delta_k - \tilde{\delta}_k)\left[\mathbf{\Psi}\left(\tilde{\delta}_k\right) - \mathbf{\Psi}\left(\sum_{g=1}^{K}\tilde{\delta}_g\right)\right] + \log\Gamma(\tilde{\delta}_k) - \frac{1}{2\omega^2}\left(\tilde{\boldsymbol{\mu}}_k^T\tilde{\boldsymbol{\mu}}_k + d\tilde{\omega}_k^2\right) + \frac{d}{2}\log\left(\tilde{\omega}_k^2\right) \right\} \\
& + \sum_{k=1}^{K} \left\{ (\xi - \tilde{\xi}_k)\mathbf{\Psi}\left(\tilde{\xi}_k\right) - \xi\log\tilde{\psi}_k - \psi\frac{\tilde{\xi}_k}{\tilde{\psi}_k} + \log\Gamma(\tilde{\xi}_k) + \tilde{\xi}_k \right\} \\
& + \left[\sum_{i=1}^{N} \frac{d}{2}\log\left(\tilde{\sigma}_i^2\right) \right] - \frac{1}{2\rho^2}\left[(\tilde{\eta} - \eta)^2 + \tilde{\rho}^2\right] + \frac{1}{2}\log\left(\tilde{\rho}^2\right) - \log\Gamma\left(\sum_{k=1}^{K}\tilde{\delta}_k\right) + \texttt{const}.
\end{aligned}
\tag{17}
$$

命题 7 的证明见附录 A.4。与期望 (6) 相比，期望 (17) 中的大多数项与期望 (6) 中包含的对应项相同，而主要差异源于期望 (17) 中的第 1 个求和项，该项包含多个变分参数，包括 $\{\tilde{\boldsymbol{U}}, \tilde{\boldsymbol{\sigma}}^{\boldsymbol{2}}, \tilde{\eta}, \tilde{\rho}^2\}$。这些差异使得与此类项相关的 PoisLPCM 变分贝叶斯推断过程不同于 MLPCM 中的对应项，而其他不涉及混合潜在位置的过程保持不变。PoisLPCM 的变分贝叶斯推断算法的详细信息，以及为 PoisLPCM 定义的 PICL 准则，见附录 A.6。

A.6 PoisLPCM 的变分贝叶斯推断

在本附录中，我们详细介绍 PoisLPCM 的变分贝叶斯推断算法。回想一下，期望 (17) 中的大多数项与期望 (6) 中对应的项相同，而主要差异源自期望 (17) 中的第一个求和项，该项包含多个变分参数，包括 $\{\tilde{\boldsymbol{U}}, \tilde{\boldsymbol{\sigma}}^{\boldsymbol{2}}, \tilde{\eta}, \tilde{\rho}^2\}$。因此，这些差异导出了 PoisLPCM 变分贝叶斯算法所需的以下命题。

命题 8 对于 $\mathbf{\Phi}$ 的任意取值，当 $\epsilon > 0$ 足够小时，

1. 对于 $i = 1$、2、.....、N，$(\tilde{\boldsymbol{u}}_i, \tilde{\sigma}_i^2)$ 的自然梯度上升更新不会降低 *PoisLPCM* 的证据下界，其定义如下：

$$
\begin{cases}
\tilde{\boldsymbol{u}}_i^{\star} = \tilde{\boldsymbol{u}}_i^{\dagger} + \epsilon\tilde{\sigma}_i^{2\dagger}\frac{\partial\mathcal{F}}{\partial\tilde{\boldsymbol{u}}_i}\left(\tilde{\boldsymbol{u}}_i^{\dagger}, \tilde{\sigma}_i^{2\dagger}, \dots\right); \\
\tilde{\sigma}_i^{2\star} = \tilde{\sigma}_i^{2\dagger}\exp\left[\epsilon\frac{2}{d}\tilde{\sigma}_i^{2\dagger}\frac{\partial\mathcal{F}}{\partial\tilde{\sigma}_i^2}\left(\tilde{\boldsymbol{u}}_i^{\dagger}, \tilde{\sigma}_i^{2\dagger}, \dots\right)\right],
\end{cases}
\tag{18}
$$

其中，$(\cdot)^\star$ 和 $(\cdot)^\dagger$ 分别表示变分参数的新状态和当前状态，并且，

$$\begin{aligned}
\frac{\partial \mathcal{F}}{\partial \tilde{\boldsymbol{u}}_i} &= \sum_{\substack{j=1,\\ j\neq i}}^{N} \left\{ -2 \left[y_{ij} + y_{ji} - \frac{2\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2)^{d/2+1}} \right] (\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j) \right\} - \sum_{k=1}^{K} \left[\tilde{\pi}_{ik} \frac{\tilde{\xi}_k}{\tilde{\psi}_k} (\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{\mu}}_k) \right], \\
\frac{\partial \mathcal{F}}{\partial \tilde{\sigma}_i^2} &= \frac{d}{2} \cdot \left[-2 \left(\sum_{\substack{j=1,\\ j\neq i}}^{N} (y_{ij} + y_{ji}) \right) - \left(\sum_{k=1}^{K} \tilde{\pi}_{ik} \frac{\tilde{\xi}_k}{\tilde{\psi}_k} \right) + \frac{1}{\tilde{\sigma}_i^2} \right] - \sum_{\substack{j=1,\\ j\neq i}}^{N} \frac{2\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2)^{d/2+1}} \left[2\frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2} - d \right].
\end{aligned} \tag{19}$$

2. PoisLPCM 的$(\tilde{\eta}, \tilde{\rho}^2)$ 自然梯度上升更新不会降低证据下界，其定义如下：

$$\begin{cases}
\tilde{\eta}^\star = \tilde{\eta}^\dagger + \epsilon \tilde{\rho}^{2^\dagger} \frac{\partial \mathcal{F}}{\partial \tilde{\eta}} \left(\tilde{\eta}^\dagger, \tilde{\rho}^{2^\dagger}, \dots \right); \\
\tilde{\rho}^{2^\star} = \tilde{\rho}^{2^\dagger} \exp\left[\epsilon 2 \tilde{\rho}^{2^\dagger} \frac{\partial \mathcal{F}}{\partial \tilde{\rho}^2} \left(\tilde{\eta}^\dagger, \tilde{\rho}^{2^\dagger}, \dots \right) \right],
\end{cases} \tag{20}$$

其中，$(\cdot)^\star$ 和$(\cdot)^\dagger$ 分别表示变分参数的新状态和当前状态，并且，

$$\begin{aligned}
\frac{\partial \mathcal{F}}{\partial \tilde{\eta}} &= \left(\sum_{\substack{i,j=1,\\ i\neq j}}^{N} y_{ij} \right) - \frac{\tilde{\eta} - \eta}{\rho^2} - \sum_{\substack{i,j=1,\\ i\neq j}}^{N} \frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2)^{d/2}}, \\
\frac{\partial \mathcal{F}}{\partial \tilde{\rho}^2} &= -\frac{1}{2\rho^2} + \frac{1}{2\tilde{\rho}^2} - \frac{1}{2} \sum_{\substack{i,j=1,\\ i\neq j}}^{N} \frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2)^{d/2}}.
\end{aligned}$$

命题 8 中的自然梯度上升方案通常与命题 4 中所示的相应方案相同，关键区别在于一阶偏导数，其证明见附录 A.7.1。请注意，Pois-LPCM 的公式（19）中的一阶偏导数取决于变分参数 $\tilde{\boldsymbol{U}}$ 和 $\tilde{\boldsymbol{\sigma}}^{\boldsymbol{2}}$ 的整个集合，这与 MLPCM 的公式（12）中所示的一阶偏导数不同，后者仅取决于变分参数对 $(\tilde{\boldsymbol{u}}_i, \tilde{\sigma}_i^2)$。因此，如算法 3 所示，公式（18）和（19）中的自然梯度上升更新还需要以 $\tilde{\boldsymbol{U}}$ 和 $\tilde{\boldsymbol{\sigma}}^{\boldsymbol{2}}$ 中每个元素的最新状态为条件。如下文命题 9 所述，我们沿用命题 6 的思路，利用 PICL 模型选择准则来选取最佳簇数。命题 9 的证明与附录 A.3 中命题 6 的证明类似，只需先从积分对数似然中移除对数 $p(\hat{\boldsymbol{V}}|\hat{\boldsymbol{U}})$ 项，然后将附录 A.3.1 中所有的 $\hat{\boldsymbol{v}}_{j\leftarrow i}$ 替换为 $\hat{\boldsymbol{u}}_j$，即可轻松得到。

命题 9 给定遵循期望 (3)的先验分布（不考虑$\hat{\boldsymbol{\gamma}}$先验），以及 *PoisLPCM* 的模型参数点估计 $\hat{\boldsymbol{U}}$、$\hat{\boldsymbol{z}}$，

算法 3 PoisLPCM 的变分贝叶斯推断算法。

输入： $\boldsymbol{Y}, d, K, \eta, \rho^2, \omega^2, \xi, \psi, \boldsymbol{\delta}$，容差水平 tol，梯度上升步长$\{\epsilon_{\boldsymbol{u}_i}\}_{i=1}^N, \epsilon_\beta, \epsilon_{\boldsymbol{\Pi}}$，以及初始状态：

$$\tilde{\boldsymbol{\Phi}}^{(0)} = \{\tilde{\boldsymbol{U}}^{(0)}, \tilde{\boldsymbol{\sigma}}^{\boldsymbol{2}^{(0)}}, \tilde{\boldsymbol{\Pi}}^{(0)}, \tilde{\eta}^{(0)}, \tilde{\rho}^{2^{(0)}}, \tilde{\boldsymbol{\mu}}^{(0)}, \tilde{\boldsymbol{\omega}}^{\boldsymbol{2}^{(0)}}, \tilde{\boldsymbol{\xi}}^{(0)}, \tilde{\boldsymbol{\psi}}^{(0)}, \tilde{\boldsymbol{\delta}}^{(0)}\}.$$

设定 stop=FALSE 和 $t = 0$。计算初始状态下的证据下界：$\mathcal{F}(\tilde{\boldsymbol{\Phi}}^{(0)})$。

当 stop $\neq$ TRUE**时，执行**

初始化 $\tilde{\boldsymbol{\Phi}}^{(t+1)} = \tilde{\boldsymbol{\Phi}}^{(t)}$。

对于 $i = 1, \ldots, N$ **do**

1. 给定 $\{\tilde{\boldsymbol{u}}_i^{(t)}, \tilde{\boldsymbol{U}}_{(t+1)}/\tilde{\boldsymbol{u}}_i^{(t+1)}\}$和 $\{\tilde{\sigma}_i^{2^{(t)}}, \tilde{\boldsymbol{\sigma}}^{\boldsymbol{2}^{(t+1)}}/\tilde{\sigma}_i^{2^{(t+1)}}\}$，更新 $(\tilde{\boldsymbol{u}}_i^{(t+1)}, \tilde{\sigma}_i^{2^{(t+1)}})$，根据期望 (18) 并结合 $\epsilon = \epsilon_{\boldsymbol{u}_i}$。

如果 $\mathcal{F}(\tilde{\boldsymbol{u}}_i^{(t+1)}, \tilde{\sigma}_i^{2^{(t+1)}}, \ldots) < \mathcal{F}(\tilde{\boldsymbol{u}}_i^{(t)}, \tilde{\sigma}_i^{2^{(t)}}, \ldots)$ **则** $\epsilon_{\boldsymbol{u}_i} = \epsilon_{\boldsymbol{u}_i}/2$；重新执行步骤 1。

结束 if

结束 for

对于 $i = 1, \ldots, N$ **do**

对于 $k = 1, \ldots, K$ **do**

2. 更新 $\tilde{\pi}_{ik}^{(t+1)}$，按照期望（7）进行。

结束循环

结束循环

3. 按照期望（20），用 $\epsilon = \epsilon_\beta$ 更新（$\tilde{\eta}^{(t+1)}$，$\tilde{\rho}^{2^{(t+1)}}$）。

如果 $\mathcal{F}(\tilde{\eta}^{(t+1)}, \tilde{\rho}^{2^{(t+1)}}, \ldots) < \mathcal{F}(\tilde{\eta}^{(t)}, \tilde{\rho}^{2^{(t)}}, \ldots)$ **则** $\epsilon_\beta = \epsilon_\beta/2$; 重新运行步骤 3。

结束如果

对于 $k = 1, \ldots, K$ **do**

4. 更新 $\tilde{\boldsymbol{\mu}}_k^{(t+1)}$ 和 $\tilde{\omega}_k^{2^{(t+1)}}$ 按照期望 (8)。

5. 更新 $(\tilde{\xi}_k^{(t+1)}, \tilde{\psi}_k^{(t+1)})$，按照期望 (9)。

6. 使用 $\tilde{\delta}_k^{(t+1)}$ 根据期望（15）更新 $\epsilon = \epsilon_{\boldsymbol{\Pi}}$。

结束循环

如果 $\mathcal{F}(\tilde{\boldsymbol{\delta}}^{(t+1)}, \ldots) < \mathcal{F}(\tilde{\boldsymbol{\delta}}^{(t)}, \ldots)$ **则** $\epsilon_{\boldsymbol{\Pi}} = \epsilon_{\boldsymbol{\Pi}}/2$；对 7 重新执行步骤 $k = 1, \ldots, K$。

结束条件判断

如果 $\mathcal{F}(\tilde{\boldsymbol{\Phi}}^{(t+1)}) - \mathcal{F}(\tilde{\boldsymbol{\Phi}}^{(t)}) <$ tol **那么** stop=TRUE

结束条件判断

设置 $t = t + 1$。

结束循环

输出： $\{\tilde{\boldsymbol{\Phi}}^{(s)}, \mathcal{F}(\tilde{\boldsymbol{\Phi}}^{(s)}) : s = 1, 2, \ldots, t\}$.

PoisLPCM 的 *PICL* 准则写作：

$$PICL = \texttt{const} - \left(\sum_{\substack{i,j=1,\\ i\neq j}}^{N} y_{ij}\right)\log\left[\sum_{\substack{i,j=1,\\ i\neq j}}^{N}\exp\left(-||\hat{\boldsymbol{u}}_i-\hat{\boldsymbol{u}}_j||^2\right)\right] - \sum_{\substack{i,j=1,\\ i\neq j}}^{N}\left(y_{ij}||\hat{\boldsymbol{u}}_i-\hat{\boldsymbol{u}}_j||^2\right)$$
$$+\max_{\boldsymbol{\tau}}\sum_{g=1}^{K}\left[\frac{d}{2}\hat{n}_g\log(\tau_g)-\frac{d}{2}\log\left(\tau_g\hat{n}_g\omega^2+1\right)+\frac{1}{2}\frac{\tau_g^2\omega^2}{\tau_g\hat{n}_g\omega^2+1}\left\|\sum_{\hat{z}_i=g}\hat{\boldsymbol{u}}_i\right\|^2-\frac{\tau_g}{2}\left(\sum_{\hat{z}_i=g}||\hat{\boldsymbol{u}}_i||^2\right)\right]$$
$$-\frac{1}{2}K\log(N)+\left[\sum_{k=1}^{K}\log\Gamma(\hat{n}_k+\delta)\right]-\log\Gamma\left(N+K\delta\right)+\log\Gamma\left(K\delta\right)-K\log\Gamma(\delta),$$

其中 $\hat{}n_g := \sum_{i=1}^{N}\mathbb{1}(\hat{}z_i = g)$ 并且，

$$\texttt{const} = \left(\sum_{\substack{i,j=1,\\ i\neq j}}^{N} y_{ij}\right)\left[\log\left(\sum_{\substack{i,j=1,\\ i\neq j}}^{N} y_{ij}\right)-1\right]-\left[\sum_{\substack{i,j=1,\\ i\neq j}}^{N}\log(y_{ij}!)\right]-\frac{1}{2}\log[N(N-1)]-\frac{d}{2}N\log(2\pi),$$

而 $\max_{\boldsymbol{\tau}}(\cdot)$ 项通过命题 *6* 中所示的相同标准梯度上升更新得到。

A.7 PoisLPCM 的证据下界最大化

由于 PoisLPCM 的证据下界的闭式优化步骤、标准梯度步骤以及自然梯度上升方案，与 MLPCM 的对应步骤相同，且已在附录 A.2 中证明，因此只需证明命题 8 中包含的一阶偏导数。

A.7.1 命题 8 的证明

PoisLPCM 的 $(\tilde{\boldsymbol{u}}_i, \tilde{\sigma}_i^2)$ 的自然梯度上升方案（其中 $i = 1, 2, \ldots, N$）与附录 A.2.2 中为 MLPCM 证明的对应方案一致。主要区别在于证据下界的一阶偏导数，如下所示。

证据下界关于 $\tilde{\boldsymbol{u}}_i$（对于 $i = 1$，2，.....，N）的一阶偏导数为

写作：

$$\frac{\partial \mathcal{F}}{\partial \tilde{\boldsymbol{u}}_i} = \frac{\partial}{\partial \tilde{\boldsymbol{u}}_i} \sum_{\substack{j=1,\\ j\neq i}}^{N} \left\{ -y_{ij}||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j||^2 - \frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2)^{d/2}} \right\} + \frac{\partial}{\partial \tilde{\boldsymbol{u}}_i} \sum_{k=1}^{K} \left[-\frac{\tilde{\pi}_{ik}}{2} \frac{\tilde{\xi}_k}{\tilde{\psi}_k} ||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{\mu}}_k||^2 \right] +$$

$$+ \frac{\partial}{\partial \tilde{\boldsymbol{u}}_i} \sum_{\substack{j=1,\\ j\neq i}}^{N} \left\{ -y_{ji}||\tilde{\boldsymbol{u}}_j - \tilde{\boldsymbol{u}}_i||^2 - \frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_j - \tilde{\boldsymbol{u}}_i||^2}{1+2\tilde{\sigma}_j^2+2\tilde{\sigma}_i^2}\right)}{(1+2\tilde{\sigma}_j^2+2\tilde{\sigma}_i^2)^{d/2}} \right\}$$

$$= \sum_{\substack{j=1,\\ j\neq i}}^{N} \left[-2(y_{ij}+y_{ji})(\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j) - \frac{2\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2)^{d/2}} \frac{-2(\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j)}{1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2} \right] - \sum_{k=1}^{K} \left[\tilde{\pi}_{ik} \frac{\tilde{\xi}_k}{\tilde{\psi}_k} (\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{\mu}}_k) \right]$$

$$= \sum_{\substack{j=1,\\ j\neq i}}^{N} \left\{ -2 \left[y_{ij} + y_{ji} - \frac{2\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2)^{d/2+1}} \right] (\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j) \right\} - \sum_{k=1}^{K} \left[\tilde{\pi}_{ik} \frac{\tilde{\xi}_k}{\tilde{\psi}_k} (\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{\mu}}_k) \right].$$

证据下界关于 $\tilde{\sigma}_i^2$（对于 $i = 1,\ 2,\ .....,\ N$）的一阶偏导数写作：

$$\frac{\partial \mathcal{F}}{\partial \tilde{\sigma}_i^2} = \frac{\partial}{\partial \tilde{\sigma}_i^2} \sum_{\substack{j=1,\\ j\neq i}}^{N} \left[-y_{ij} \cdot d\tilde{\sigma}_i^2 - \frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2)^{d/2}} \right] +$$

$$+ \frac{\partial}{\partial \tilde{\sigma}_i^2} \sum_{\substack{j=1,\\ j\neq i}}^{N} \left[-y_{ji} \cdot d\tilde{\sigma}_i^2 - \frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_j - \tilde{\boldsymbol{u}}_i||^2}{1+2\tilde{\sigma}_j^2+2\tilde{\sigma}_i^2}\right)}{(1+2\tilde{\sigma}_j^2+2\tilde{\sigma}_i^2)^{d/2}} \right] - \frac{\partial}{\partial \tilde{\sigma}_i^2} \sum_{k=1}^{K} \left(\frac{\tilde{\pi}_{ik}}{2} \frac{\tilde{\xi}_k}{\tilde{\psi}_k} d\tilde{\sigma}_i^2 \right) + \frac{\partial}{\partial \tilde{\sigma}_i^2} \left[\frac{d}{2} \log\left(\tilde{\sigma}_i^2\right) \right]$$

$$= \sum_{\substack{j=1,\\ j\neq i}}^{N} [-(y_{ij}+y_{ji})d] - \sum_{k=1}^{K} \left(\frac{\tilde{\pi}_{ik}}{2} \frac{\tilde{\xi}_k}{\tilde{\psi}_k} d \right) + \frac{d}{2\tilde{\sigma}_i^2}$$

$$- \sum_{\substack{j=1,\\ j\neq i}}^{N} 2 \left[\frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2}\right) \cdot 2||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j||^2 (1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2)^{-2}}{(1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2)^{d/2}} - \frac{d}{2} \frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2}\right) \cdot 2}{(1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2)^{d/2+1}} \right]$$

$$= \frac{d}{2} \cdot \left[-2 \left(\sum_{\substack{j=1,\\ j\neq i}}^{N} (y_{ij}+y_{ji}) \right) - \left(\sum_{k=1}^{K} \tilde{\pi}_{ik} \frac{\tilde{\xi}_k}{\tilde{\psi}_k} \right) + \frac{1}{\tilde{\sigma}_i^2} \right] - \sum_{\substack{j=1,\\ j\neq i}}^{N} \frac{2\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2)^{d/2+1}} \left[2\frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2} - d \right].$$

□

PoisLPCM 的 $(\tilde{\eta}, \tilde{\rho}^2)$ 自然梯度上升方案与附录 A.2.2 中为 MLPCM 证明的相应方案一致。主要区别在于如下计算的证据下界的一阶偏导数。

证据下界关于 $\tilde{\eta}$ 的一阶偏导数写作：

$$\begin{aligned}\frac{\partial \mathcal{F}}{\partial \tilde{\eta}} &= \frac{\partial}{\partial \tilde{\eta}} \sum_{\substack{i,j=1,\\ i\neq j}}^{N} \left[y_{ij}\tilde{\eta} - \frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2)^{d/2}} \right] - \frac{\partial}{\partial \tilde{\eta}} \left[\frac{1}{2\rho^2}(\tilde{\eta}-\eta)^2 \right] \\ &= \left(\sum_{\substack{i,j=1,\\ i\neq j}}^{N} y_{ij} \right) - \frac{\tilde{\eta}-\eta}{\rho^2} - \sum_{\substack{i,j=1,\\ i\neq j}}^{N} \frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2)^{d/2}}.\end{aligned}$$

证据下界关于 $\tilde{\rho}^2$ 的一阶偏导数写作：

$$\begin{aligned}\frac{\partial \mathcal{F}}{\partial \tilde{\rho}^2} &= \frac{\partial}{\partial \tilde{\rho}^2} \sum_{\substack{i,j=1,\\ i\neq j}}^{N} \left[-\frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2)^{d/2}} \right] + \frac{\partial}{\partial \tilde{\rho}^2} \left[-\frac{\tilde{\rho}^2}{2\rho^2} + \frac{1}{2}\log\tilde{\rho}^2 \right] \\ &= -\frac{1}{2\rho^2} + \frac{1}{2\tilde{\rho}^2} - \frac{1}{2} \sum_{\substack{i,j=1,\\ i\neq j}}^{N} \frac{\exp\left(\tilde{\eta} + \frac{\tilde{\rho}^2}{2} - \frac{||\tilde{\boldsymbol{u}}_i - \tilde{\boldsymbol{u}}_j||^2}{1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2}\right)}{(1+2\tilde{\sigma}_i^2+2\tilde{\sigma}_j^2)^{d/2}}.\end{aligned}$$

□

A.8 2024年自我中心武器转让 K=5 潜在表示 s

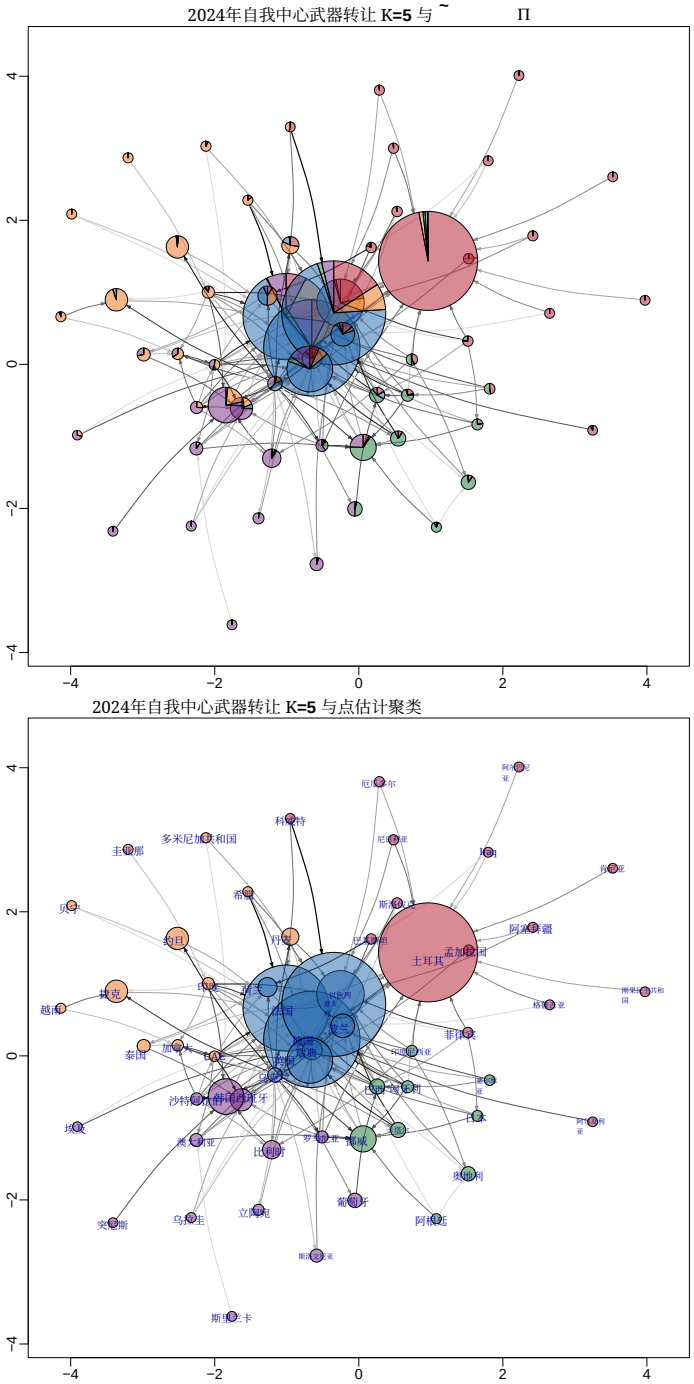

图 9：2024年自我中心武器转让真实数据应用：推断得到的 $K = 5$的潜在空间。对于两幅图，节点位置对应$\hat{\boldsymbol{U}}$，节点大小与推断得到的 $\{1/\hat{\gamma}_j = \tilde{b}_j/\tilde{a}_j\}$成正比，而边宽度和颜色与边权重成正比。上图：每个节点的饼图展示相应的群组分配变分概率。下图：节点颜色对应最大后验聚类；每个节点均标注相应的国家名称。