arXiv:2611.32430v1 [stat.CO] 24 Nov 2026

用于非均匀超图的跨阶信息引导估计

Rongzhuo Nie[1, 2*]、Menghao Wang[1, 2*]、Sanguo Zhang[1,2],Mingyang Ren[3†]、Guojun Zhu[1, 2†]

[1]数学科学学院，中国科学院大学，北京，中国[2]大数据挖掘与知识管理重点实验室，中国科学院，北京，中国[3]数学科学学院，上海交通大学，上海，中国

摘要

许多现实世界中的交互涉及节点群组，而非节点对。非均匀超图将此类交互记录为不同阶数的超边，其中，阶表示超边中包含的节点数。完整估计需要评估每一条候选超边，其数量会随阶数呈指数增长，而经典方法要么牺牲高阶信息，要么加剧计算负担。不同阶数并非彼此孤立：低阶信息计算成本低且信息丰富，可以揭示高阶超边可能出现的位置。我们在一个所有阶数共享潜在位置的潜在空间模型下提出**CINE**（跨阶信息引导负采样估计）。CINE 从二阶开始，逐阶向上推进。在每个更高阶数下，CINE 保留所有存在的超边，并根据低阶得分引导采样一小部分负样本，同时校正已知的纳入概率，以保持完整估计的目标。由此，CINE 让用户能够选择计算时间与估计精度之间的权衡：保留阶数和采样预算决定所需时间，而随着采样预算增加，所保留的信息在理论上会逐渐逼近全数据信息。在模拟实验中，CINE 仅需完整估计一小部分计算时间，便可达到接近完整估计的精度；在 MADStat 中，对更大的团队进行分析揭示了研究关联。

关键词： 非均匀超图；潜空间模型；计算效率；负采样；跨阶信息。

*Rongzhuo Nie 和 Menghao Wang 对本研究贡献相同。
†通讯作者。邮箱：renmingyang17@mails.ucas.ac.cn；zhuguojun23@mails.ucas.ac.cn。

1 引言

过去几十年来，网络数据的统计分析主要聚焦于成对交互，即一条边是否连接两个节点（Hoff 等，2002；Fienberg 2012年）。然而，在现实世界中，许多交互会同时发生在一组节点之间：一篇论文的合作者、一次会议的参与者、一次化学反应中的分子，以及同一张购物小票上的商品（Battiston et al. 2020）。将这种群体交互拆分为成对的边，会丢失哪些节点共同交互的信息：由 $\{a, b, c, d\}$ 中的四位作者共同发表的一篇论文，与由六个作者节点对 $\{a, b\}$、$\{a, c\}$ ……、$\{c, d\}$ 分别发表的六篇论文，会产生完全相同的成对边（Turnbull et al. 2024）。超图保留了这一信息。在超图中，一条超边可以连接任意数量的节点，而超边中的节点数量就是其阶数，因此普通边是阶数为二的超边。由于实际中的群体规模各不相同，一个超图通常包含多个阶数的超边，我们将这类超图称为非均匀超图（Zhen and Wang 2023）。因此，非均匀超图的统计建模已成为网络分析的一个重要方向，并应用于科学合作和生物网络（Contisciani et al. 2022；Turnbull et al. 2024）。

为了估计非均匀超图的生成机制，一类常见的统计模型会在目标函数（例如负对数似然）中评估每条候选超边，并判断该候选超边是否存在（Dumitriu andWang 2026）。这里，阶数为 m 的候选超边是指由 m 个不同节点组成的任意集合。候选超边的数量会随着阶数增加而急剧增长，而存在的超边仅占候选超边中的极小一部分。例如，如图 1(a)所示，在 MADStat 数据（Ji等（2022））中，候选超边数量从五阶时的约 1.97×10^{21} 增至七阶时的 1.05×10^{29}，而存在的超边数量则从 1,308 降至 201。……的计算时间

因此，评估所有候选超边的完整估计，其计算量会随阶数呈指数增长；对于中等规模的超图，几乎不可能对所有阶数进行完整估计。因此，现有研究通常会在估计前压缩超图，最常见的方法是截断或投影。如图 1(b)–1(d) 所示，截断会丢弃所有高于指定阶数的超边（Zhen and Wang2023；Lucas etal. 2026；Zhu et al.2026），而投影则会将高阶超边拆分为低阶超边（Agarwal et al. 2006；Zhou et al. 2006；Chodrow et al. 2021）。针对均匀超图（其中所有超边的阶数相同）的方法也有很多（Ghoshdastidarand Dukkipati 2017；Lyu et al. 2023）。为了应用这些方法，许多研究会先采用填充，将每条低阶超边补齐至最高阶数，方法是添加一个空节点——一种没有真实对应物的人工节点。由于最高阶数往往仍然过高，难以处理，之后还会采用截断或投影（Zhenand Wang 2023；Zhu et al.2026）。截断会舍弃高阶超边中的信息，而投影会将高阶信息混入低阶，并对大型超边赋予过高权重：一篇由七位作者合著的论文会被转化为 35 次三位作者合作，从而使估计偏向这七位作者（Yoonet al. 2020；Contisciani et al. 2022）。填充不会丢失信息，但会将每一阶都提升至最高阶数并添加一个人工节点，因此计算量不减反增。由此，非均匀超图的估计面临两难：要么花费大量计算时间处理所有阶数，要么牺牲高阶信息。理想情况下，分析者可以设定可承担的计算时间，并在允许更多时间时获得逐渐接近完整估计的估计精度。要实现这一目标，需要一种无需评估每条高阶候选超边、却能保留每条高阶超边的方法。除了截断、投影和填充之外，也有一些研究在不压缩数据的情况下处理非均匀超图（Turnbullet al.2024；Yu andZhu 2025）。其中一类研究

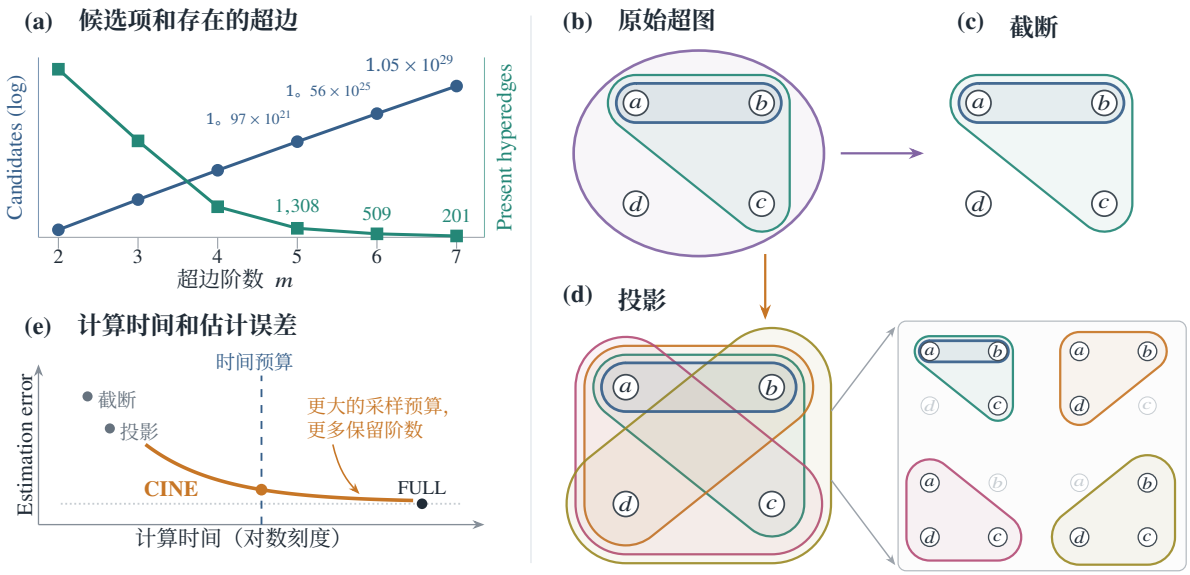


图 1：非均匀超图中候选超边数量的增长与压缩。(a) MADStat 中不同阶数下候选超边和存在的超边数量。(b) 一个包含 $\{a, b\}$、$\{a, b, c\}$ 和 $\{a, b, c, d\}$ 的原始超图。每个彩色区域包围一个超边的成员。(c) 在三阶处进行截断会丢弃 $\{a, b, c, d\}$。(d) 投影将 $\{a, b, c, d\}$ 替换为其三节点子集，同时保留现有的低阶超边。右侧图版展示了所得的超边。(e) 计算时间与估计误差之间的示意性权衡。截断与投影速度快，但会丢失高阶信息；FULL 精确但速度慢；CINE 则描绘出一条介于两者之间的曲线，用户可通过选择保留阶数和采样预算来确定其位置。

仅评估候选超边的随机子集，并通过病例-对照抽样（Raftery 等，2012；Lerner 和 Lomi，2023）或负采样（Shi 和 Evans 2023）针对采样对似然进行校正，这一思想与噪声对比估计（Gutmann 和 Hyvärinen，2012）密切相关。由于均匀抽取的子集中的绝大多数候选超边并不存在，对估计的贡献几乎为零，因此这些工作必须评估大量候选超边以保持估计精度，而计算时间再次增加。局部病例-对照抽样（Fithian and Hastie2014）通过试拟合抽取信息量更高的样本，但在每条目标阶候选超边上评估试拟合，恰恰就是会导致计算量爆炸的计算过程。另一类方法利用稀疏性，围绕存在的超边组织计算： Yuand Zhu（2025）对存在的超边求和似然贡献，并以解析方式计算归一化项；Turnbull 等 （2024）则利用重叠计数评估似然

存在的超边与几何诱导超边之间的重叠计数。归根结底，这些工作中的每一种都仅在单个阶的范围内加速该阶的估计。

然而，所有这些工作都忽略了一个最简单的事实：非均匀超图的各个阶并不是彼此孤立的岛屿。同一批节点会跨越不同阶出现。例如，在 MADStat（Ji等（2022））中，6,616 条四阶或更高阶超边中有 2,421 条包含一个节点对或三元组，而该节点对或三元组在更早年份已经构成了低阶超边。此外，不同阶之间的信息与计算并不匹配。低阶的候选超边较少，而存在的超边较多，因此计算成本低且信息丰富。高阶的候选超边很多，而存在的超边较少，因此计算成本高且信息稀疏。因此，低阶能够揭示哪些高阶候选超边值得评估；我们将低阶提供的有关高阶超边的信息称为跨阶信息。我们从二阶开始，此时计算成本最低，然后每次提升一阶。在每个新阶上，跨阶信息不必仅来自相邻阶，而是用于指导候选超边的评估选择。这样一来，需要评估的候选超边数量不再随阶数呈爆炸式增长，同时每条高阶超边都能保留在估计中，从而保持估计精度。

从统计角度看，每条候选超边都对应一个二元观测：若该候选超边存在，则观测值为 1；否则为 0。对于这类数据，潜在空间模型（Hoff 等，2002；Turnbull et al. 2024）用一个潜在位置表示每个节点，潜在位置彼此更接近的一组节点更有可能形成超边。潜在空间模型已扩展到各种超图数据： Yuan 和 Qu (2023) 对成对链接和高阶链接进行联合建模，Zhen 和 Wang(2023) 通过空节点填充嵌入非均匀超图，而 Wu 等 (2026) 则对含有重复超边的稀疏非均匀超图进行建模。在本文中，我们令所有阶数共享同一组潜在

位置，因此，在较低阶数下估计得到的潜在位置可以直接为每条高阶候选超边评分，并传递跨阶信息。基于此，我们提出 CINE（跨阶信息引导的负采样估计）。我们将不存在的候选超边称为负例。在每个更高阶数下，CINE 保留所有存在的超边，仅评估由 CINE 根据低阶得分选出的一小部分负样本，并利用已知的纳入概率（即负样本被采样的概率）校正估计，使 CINE 的估计目标与完整估计的目标一致（第2节）。因此，CINE 的计算时间随存在的超边数量和采样的负样本数量增长，而非随目标阶数下候选超边的数量增长，同时其估计精度仍接近完整估计。由此，计算时间与估计精度之间的权衡由用户自行选择，如图1(e) 所示：给定时间预算，可在预算范围内选择保留阶数（注2），而增加采样预算则能使估计精度更接近完整估计（第3节）。

我们的贡献主要有四点。(i) 我们揭示了非均匀超图估计中计算时间与估计精度之间的两难关系，并表明跨阶信息可将这一两难转化为由用户选择的权衡：CINE 在任意时间预算下都保留每一条现有超边，并且随着时间预算增加而逐渐接近完整估计。(ii) 我们提出了一种潜在空间模型，其中所有阶共享潜在位置；同时提出 CINE 这一估计方法，在跨阶信息的指导下逐阶进行估计。(iii) 我们证明 CINE 保留了全数据信息中的大部分，并推导出非渐近误差界，其中潜在位置达到信息量最大的保留阶数所对应的速率，且采样预算的影响得到了明确体现；这表明，在

显著降低计算成本的情况下，CINE 可达到与完整估计相当的估计精度。(iv) 在模拟中，CINE 的估计误差低于以下方法：Yuan 和 Qu (2023)、Zhen 和 Wang (2023) 以及 Wu 等(2026)；与其主要竞争方法相比，CINE 仅需极少的计算时间，并且其精度会随着采样预算的增加而持续提升。在 MADStat 数据（Ji等（2022））上，通过 CINE 纳入规模更大的团队，揭示了具有相近研究兴趣的统计学家之间的关联。

2 方法论

2.1 非均匀超图的共享潜在空间模型

超图记为 $G = (V, E)$，其中 $V = \{v_1, \ldots, v_n\}$ 是一个由 n 个节点组成的集合，每条超边 $e \in E$ 都是 V 的一个无序子集，其阶为该子集的基数。我们关注非均匀超图，其中超边的阶可能不同。超边存在并不要求其真子集也必须作为超边存在，这适用于记录共同参与事件的数据：特定群体参与的事件，与仅有该群体部分成员参与的事件，是不同的观测。

令 H 表示 G 中存在的超边阶数。纳入某个阶数会增加计算成本，且有些阶数中观测到的超边极少。因此，我们在保留的超边阶数集合 M ⊆H 上拟合模型，该集合可按照备注 2在时间预算内选取。我们将 $\Omega_m = \{e \subseteq [n] : |e| = m\}$ 定义为所有可能的阶数为m的超边集合，称为阶数为m 的候选超边（简称候选项），其中 $[n] = \{1, 2, \ldots, n\}$。二元指示变量 $A_e^{(m)} = \mathbf{1}\{e \in E\}$ 记录候选项 e 是否存在于 G中，其中 $\mathbf{1}\{\cdot\}$ 表示指示函数。 $A_e^{(m)} = 1$ 的候选项是存在的超边，而 $A_e^{(m)} = 0$ 的候选项称为负例。我们用 $E_m^+ = \{e \in \Omega_m : A_e^{(m)} = 1\}$ 表示阶数为m 的存在超边。现实世界中的超图

通常是稀疏的：只有一小部分候选超边存在，而我们关注的是这类稀疏超图。

为参数化每个候选超边的出现概率，我们采用潜在空间建模（Hoff 等，2002；Ma 等，2020），为每个节点分配一个潜在位置。具体而言，每个节点 v_i 都与一个潜在位置 $\boldsymbol{z}_i \in \mathbb{R}^r$相关联，其中 r 为潜在维度。将这些向量按行收集可得到潜在位置矩阵 $\boldsymbol{Z} \in \mathbb{R}^{n\times r}$，该矩阵对于 $\mathcal{M}$ 中的所有阶数都是共享的。

对于候选超边 $e \in \Omega_m$，我们假设其出现概率 $p_m(e;\ \boldsymbol{Z},\ \beta_m)$ 仅取决于共享潜在位置矩阵 $\boldsymbol{Z}$ 和特定于阶的截距 β_m，通过逻辑斯蒂模型表示为：

$$p_m(e;\boldsymbol{Z},\beta_m) := \text{logit}^{-1}(\eta_m(e;\boldsymbol{Z},\beta_m)) \quad \text{and} \quad \eta_m(e;\boldsymbol{Z},\beta_m) := \beta_m + s_m(e;\boldsymbol{Z}), \tag{1}$$

其中 $\text{logit}(x) = \log\{x/(1-x)\}$，而评分 $s_m(e;\boldsymbol{Z})$ 概括了 e 中节点之间的总体潜在亲和力，其定义为：

$$s_m(e;\boldsymbol{Z}) = \left\{ r\binom{m}{2} \right\}^{-1/2} \sum_{\{i,j\}\subseteq e} \boldsymbol{z}_i^\top \boldsymbol{z}_j. \tag{2}$$

通过 $\{r\binom{m}{2}\}^{1/2}$ 进行归一化，可根据潜在维度和节点对数量调整评分的尺度。截距 β_m 决定评分为零时的出现对数优势，并使基准超边流行程度能够随阶数变化。对于固定的 m 和 β_m，评分越高，出现概率越大。当超边表示事件中的共同参与时，评分具有直观的解释：正如群体成员两两共享的事件数量之和反映了成员之间关系的紧密程度，该评分将成对潜在亲和力汇总为整个群体共同参与某个事件的倾向。

给定 $\boldsymbol{Z}$ 和 β_m后，假设指示变量 $A_e^{(m)}$相互独立，且服从$A_e^{(m)} \sim$ 伯努利$\{p_m(e;\ \boldsymbol{Z}, \beta_m)\}$分布。潜在位置并不唯一，因为任意正交矩阵 $\boldsymbol{O} \in \mathbb{R}^{r\times r}$ 都会使 $s_m(e;\ \boldsymbol{ZO}) = s_m(e;\ \boldsymbol{Z})$，进而使 $p_m(e;\ \boldsymbol{ZO}, \beta_m) = p_m(e;\ \boldsymbol{Z}, \beta_m)$。我们始终施加中心化约束 $\sum_{i=1}^n \boldsymbol{z}_i = \boldsymbol{0}$ ，在此约束下，模型在公共正交变换意义下可识别。

命题 1（可识别性）。假设 $2 \in \boldsymbol{M}$ 和 $n \geq r + 2$。在上述中心化约束下，对于所有 $m \in \boldsymbol{M}$ 和 $e \in \Omega_m$，概率 $p_m(e;\ \boldsymbol{Z}, \beta_m)$ 唯一确定格拉姆矩阵 $\boldsymbol{ZZ}^\top$ 和阶数特定截距 $\{\beta_m\}_{m\in\mathcal{M}}$。因此，共享潜在位置矩阵 $\boldsymbol{Z}$ 可确定至一个共同的正交变换。因此，我们将相差一个正交变换的潜在位置矩阵视为等价，并使用格拉姆矩阵 $\boldsymbol{ZZ}^\top$。

2.2 跨阶信息引导估计

我们的目标是估计 $\boldsymbol{Z}$ 和 $\{\beta_m\}_{m\in\mathcal{M}}$。$G$ 的完整伯努利损失是在所有候选项上计算的伯努利负对数似然：

$$\begin{aligned}\ell(\boldsymbol{Z}, \{\beta_m\}_{m\in\mathcal{M}}) &= -\sum_{m\in\mathcal{M}}\sum_{e\in\Omega_m}\left[A_e^{(m)} \log p_m(e; \boldsymbol{Z}, \beta_m) + \{1 - A_e^{(m)}\} \log\{1 - p_m(e; \boldsymbol{Z}, \beta_m)\}\right] \\ &= \sum_{m\in\mathcal{M}}\sum_{e\in\Omega_m}\left[\log\{1 + \exp\{\eta_m(e; \boldsymbol{Z}, \beta_m)\}\} - A_e^{(m)}\eta_m(e; \boldsymbol{Z}, \beta_m)\right].\end{aligned} \tag{3}$$

在（3）中，每个阶数为 m 的负例都会产生严格为正的项 $\log\{1+\exp(\beta_m + s_m(e;\ \boldsymbol{Z}))\}$，因此即使对于稀疏超图，完整估计也必须评估所有 $\sum_{m\in\mathcal{M}} \binom{n}{m}$个候选项。由于大多数候选项都是负例，我们纳入所有存在的超边，并仅对一部分负样本进行抽样。

采样由跨阶信息引导。我们要求 $2 \in \mathcal{M}$ （命题 1），并从阶数为二、计算成本最低的阶开始。首先，如图版（1）所示，位于

图 2中，我们通过最小化二阶损失，将模型拟合到二阶候选项上，即完整伯努利损失的二阶分量：

$$(\widetilde{\boldsymbol{Z}}, \tilde{\beta}_2) \in \underset{\substack{\boldsymbol{Z}, \beta_2 \\ \sum_{i=1}^n \boldsymbol{z}_i = \mathbf{0}}}{\arg\min} \sum_{1 \le i < j \le n} \left[\log\left\{ 1 + \exp\left(\beta_2 + \frac{\boldsymbol{z}_i^\top \boldsymbol{z}_j}{\sqrt{r}} \right) \right\} - A^{(2)}_{\{i,j\}} \left(\beta_2 + \frac{\boldsymbol{z}_i^\top \boldsymbol{z}_j}{\sqrt{r}} \right) \right].$$

$\widetilde{\boldsymbol{Z}}$共享潜在位置矩阵的初步估计 $\boldsymbol{Z}$ 计算起来很简单，因为二阶候选集仅包含 $\binom{n}{2} = O(n^2)$ 个元素。

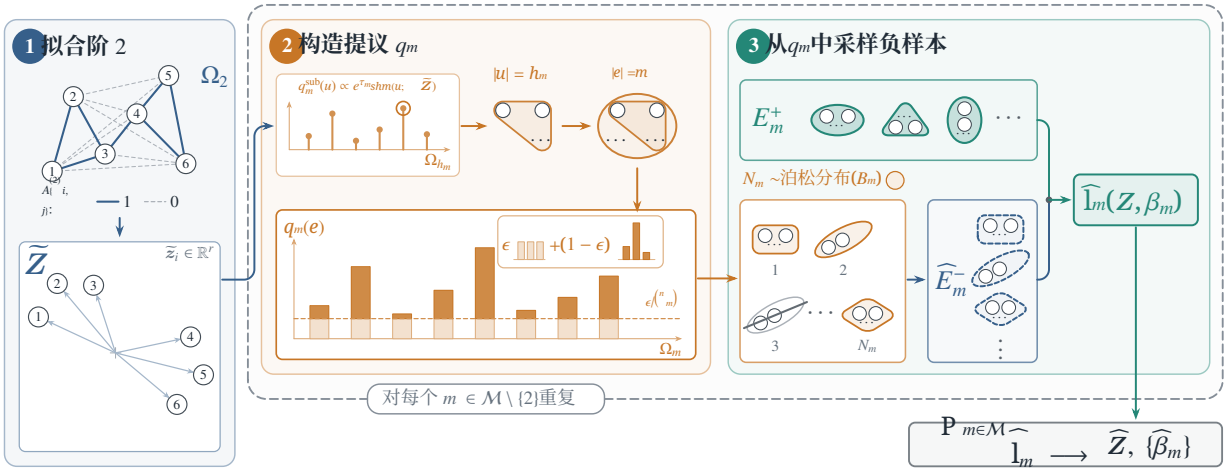


图2：CINE 工作流程。(1) CINE 在所有二阶候选项$\widetilde{\Omega}_2$ 上最小化二阶损失，并获得共享潜在位置的初步估计 $\boldsymbol{Z}$。(2) 对于每个保留阶数 $m > 2$，阶数为 h_m（小于 m 的最大保留阶数）的候选项的初步分数，定义了选择概率 $q_m^{\text{sub}}(u)$。将选中的 u 补充 $m - h_m$ 个均匀抽取的节点，并与权重为 ϵ 的均匀抽样混合，得到式（4）中的提议分布 q_m。(3) CINE 从 q_m E_m^+ 中抽取 $N_m \sim$个泊松分布(B_m)候选项，保留所有存在的超边以及抽样得到的负样本 $\widehat{E}_m^-$，并构造校正损失 $\ell_m(\boldsymbol{Z}, \beta_m)$，其偏移项为 $-\log \pi_m(e)$。步骤(2)和(3)对每个 $m \in \mathrm{M} \setminus \{2\}$ 重复执行，并在式（6）中最小化 CINE 损失 $\sum_{m\in \mathrm{M}} \ell_m$，得到 $\boldsymbol{Z}$ 和 $\{\beta_m\}$。

其次，如图版 2 的图版 (2) 所示，对于每个目标阶数 $m \in \mathcal{M}$ 且满足 $m > 2$ 的情况，我们根据 $\boldsymbol{Z}$构造跨阶提议分布。令 $h_m = \max\{k \in \mathcal{M} : k < m\}$ 表示低于 m的最大保留阶数；由于 $2 \in \mathcal{M}$，该阶数是定义良好的。对于每个 $e \in \Omega_{h_m}$（无论其是否存在），我们计算初步评分 $s_{h_m}(e; \boldsymbol{Z}) = \{r\binom{h_m}{2}\}^{-1/2} \sum_{\{i,j\} \subseteq e} \tilde{\boldsymbol{z}}_i^\top \tilde{\boldsymbol{z}}_j$，其中 $\tilde{\boldsymbol{z}}_i^\top$表示 $\widetilde{\boldsymbol{Z}}$的第 i 行。

对于每个目标阶数 m，一次抽样以如下概率选择 $e \in \Omega_m$：

$$q_m(e) = \frac{\epsilon}{\binom{n}{m}} + \frac{1-\epsilon}{\binom{n-h_m}{m-h_m}} \sum_{\substack{u \subseteq e \\ |u|=h_m}} q_m^{\mathrm{sub}}(u), \tag{4}$$

其中，固定混合权重 $\epsilon \in (0,1]$ 是从 Ω_m 中均匀采样的概率，而 $q_m^{\mathrm{sub}}(u) = \exp\{\tau_m s_{h_m}(u; \widetilde{\boldsymbol{Z}})\} / \sum_{v \in \Omega_{h_m}} \exp\{\tau_m s_{h_m}(v; \widetilde{\boldsymbol{Z}})\}$ 是选取 $u \in \Omega_{h_m}$ 的概率。缩放因子 $\tau_m = \{\binom{m}{2} / \binom{h_m}{2}\}^{1/2}$ 用于校正不同阶数间评分归一化的差异：如果目标候选超边 e 内部的初步成对内积 $\widetilde{\boldsymbol{Z}}\widetilde{\boldsymbol{Z}}^\top$ 全部相等，那么对于包含于 e 中的每个 $u \in \Omega_{h_m}$，均有 $\tau_m s_{h_m}(u; \boldsymbol{Z}) = s_m(e; \boldsymbol{Z})$。概率 $q_m^{\mathrm{sub}}(u)$ 会优先选择初步分数较高的 h_m 阶候选超边，且仅取决于 $\boldsymbol{Z}$ 和保留阶数。

从式(4)中进行一次抽取的过程如下。以概率 ϵ从Ω_m中均匀抽取 e。以概率 $1-\epsilon$，以概率 $q_m^{\mathrm{sub}}(u)$选择 $u \in \Omega_{h_m}$，并从 $[n] \backslash u$中无放回地均匀抽取 $m-h_m$个节点来补全 u。分母 $\binom{n-h_m}{m-h_m}$统计了 u的可能补全方式数量，而求和遍历包含于 e中的所有h_m 阶候选超边，因此对于每个 $e \in \Omega_m$，$\sum_{e \in \Omega_m} q_m(e) = 1$ 以及 $q_m(e) \geq \epsilon / \binom{n}{m} > 0$。与均匀抽样（$\epsilon = 1$）相比，该提议分布更偏向于包含初步分数较高的低阶候选超边的m 阶候选超边。在采样预算有限的情况下，将采样导向具有更强潜在关联的节点组，可以得到信息量更高的负样本。混合权重 ϵ控制均匀抽样的贡献：取值越大，提议分布越接近均匀分布；取值越小，则越强调初步分数。第三，如图2的图版(3)所示，对于每个目标阶数 m，我们生成数量为 $N_m \sim$ 的泊松分布(B_m)提议抽样，并从 q_m中独立地有放回抽取 N_m 个候选项。令 $C_e^{(m)}$表示 $e \in \Omega_m$ 被抽取的次数，则 $C_e^{(m)} \mid N_m = k \sim$

二项分布$\{k, q_m(e)\}$。根据全概率定律,

$$\begin{aligned}\Pr\{C_e^{(m)} = 0\} &= \sum_{k=0}^{\infty} \Pr\{N_m = k\} \Pr\{C_e^{(m)} = 0 \mid N_m = k\} \\ &= \sum_{k=0}^{\infty} \exp(-B_m) \frac{B_m^k}{k!} \{1 - q_m(e)\}^k = \exp\{-B_m q_m(e)\}.\end{aligned}$$

e的纳入概率，即 e 至少被抽取一次的概率，因此为 $\pi_m(e) = 1 - \exp\{-B_m q_m(e)\}$。从抽取结果中去除重复项和已存在的超边后，得到抽样负样本集合 $\widehat{E}_m^- = \{e \in \Omega_m : A_e^{(m)} = 0,\ C_e^{(m)} \geq 1\}$，其期望大小至多为 B_m。因此，我们将 B_m 称为阶数m的采样预算。我们通过预算比 B_m/μ_m来设置 B_m ；若 μ_m 未知，则通过 $B_m/|E_m^+|$ 设置。该比率控制保留的全数据的信息比例（第3.1节）。

为考虑负采样导致的纳入概率不等，我们为纳入的候选项构造条件伯努利似然。对于每个目标阶数 $m > 2\hat{}$以及候选项 $e \in \Omega_m$，令 $R_e^m = \mathbf{1}\{e \in E_m^+ \cup E_m^-\}$ 表示 e 是否纳入估计样本。在以下推导中，我们始终以初步估计为条件，并将提议概率 $q_m(e)$和采样预算 B_m 视为固定值。由于每条存在的超边都会被纳入，而负例以概率$\pi_m(e)$纳入，因此有 $\Pr\{R_e^m = 1 \mid A_e^{(m)} = 1\} = 1$ 以及 $\Pr\{R_e^m = 1 \mid A_e^{(m)} = 0\} = \pi_m(e)$。由贝叶斯定理和逻辑斯蒂模型可得:

$$\begin{aligned}\Pr\{A_e^{(m)} = 1 \mid R_e^m = 1\} &= \frac{\Pr\{A_e^{(m)} = 1, R_e^m = 1\}}{\Pr\{R_e^m = 1\}} \\ &= \frac{p_m(e; \boldsymbol{Z}, \beta_m)}{p_m(e; \boldsymbol{Z}, \beta_m) + \{1 - p_m(e; \boldsymbol{Z}, \beta_m)\}\pi_m(e)} \\ &= \frac{\exp\{\eta_m(e; \boldsymbol{Z}, \beta_m) - \log \pi_m(e)\}}{1 + \exp\{\eta_m(e; \boldsymbol{Z}, \beta_m) - \log \pi_m(e)\}}.\end{aligned}$$

因此，以纳入为条件会将已知偏移项 $-\log \pi_m(e)$ 引入原始对数优势中。在局部病例-对照抽样（Fithian and Hastie 2014）中也会出现类似的偏移项

以及噪声对比估计（Gutmann and Hyvärinen2012），而 CINE 的提议分布 q_m由低阶分数构建，无需预先评估阶为 m 的候选超边。偏移项 $-\log \pi_m(e)$ 同时适用于存在的超边和采样得到的负样本，并在参数估计过程中保持不变。以初步估计为条件时，泊松构造会产生相互独立的候选计数，并且这些计数也独立于目标阶超边指示变量。结合超边指示变量所假定的独立性，这些性质意味着：在给定纳入指示变量的条件下，被纳入候选项的结果是相互独立的伯努利变量，其概率如上所述。对它们的乘积似然取负对数，可得到阶 m 校正损失，即条件负对数似然：

$$\widehat{\ell}_m(\boldsymbol{Z}, \beta_m) = \sum_{e \in E_m^+ \cup \widehat{E}_m^-} \Big[\log\left\{1 + \exp\left(\eta_m(e; \boldsymbol{Z}, \beta_m) - \log \pi_m(e)\right)\right\} - A_e^{(m)} \left\{\eta_m(e; \boldsymbol{Z}, \beta_m) - \log \pi_m(e)\right\} \Big]. \tag{5}$$

对于 $m = 2$，每个二阶候选项都被纳入，因此我们设定 $\pi_2(e) = 1$，并且 ℓ_2 与二阶损失相同。据此，我们定义 CINE 估计量如下：

$$\left(\widehat{\boldsymbol{Z}}, \{\widehat{\beta}_m\}_{m \in \mathcal{M}}\right) \in \underset{\substack{\boldsymbol{Z}, \{\beta_m\}_{m \in \mathcal{M}} \\ \sum_{i=1}^n \boldsymbol{z}_i = \boldsymbol{0}}}{\arg\min} \sum_{m \in \mathcal{M}} \widehat{\ell}_m(\boldsymbol{Z}, \beta_m). \tag{6}$$

算法1 概述了 CINE 的步骤。

注 1 (序贯细化：CINE++)。 *CINE* 根据同一个初步估计，为所有更高阶构造提议分布$\widetilde{\boldsymbol{Z}}$。一种称为 *CINE++* 的序贯扩展按阶数递增的顺序处理保留阶数：对于每个 $m \in \mathcal{M}$ 满足 $m > 2$，*CINE++* 根据基于所有更低保留阶数得到的最新潜在位置估计，构造 $q_m(e)$ 在 (4)，如同 *CINE* 一样采样负例，并联合更新共享的

潜在位置和阶数特定截距，使用截至 m 的所有保留阶数。因此，跨阶数积累的信息会指导后续的负采样，第 4 节表明，*CINE++*还能进一步提高估计精度。

2.3 计算

本小节从单个候选超边出发，比较 CINE 与完整估计的计算成本。对于 $e \in \Omega_m$，恒等式 $\sum_{\{i,j\}\subset e} \boldsymbol{z}_i^\top \boldsymbol{z}_j = \{\|\sum_{i\in e} \boldsymbol{z}_i\|^2 - \sum_{i\in e}\|\boldsymbol{z}_i\|^2\}/2$ 使评分 $s_m(e;\boldsymbol{Z})$ 可在 $O(mr)$ 次运算内完成计算。对于 $i \in e$，同一向量和也可重复用于计算 $\partial s_m(e;\boldsymbol{Z})/\partial \boldsymbol{z}_i = (\sum_{j\in e}\boldsymbol{z}_j - \boldsymbol{z}_i)/\{r\binom{m}{2}\}^{1/2}$，且 e 之外的节点导数为零。这些 m 个梯度向量各需 $O(r)$ 次运算，因此，计算一个 m阶候选项的损失贡献及其梯度需要 $O(mr)$ 次运算。

现在我们分析 CINE 各阶段的计算复杂度。首先，CINE 通过最小化二阶损失函数获得初步估计 $\widetilde{\boldsymbol{Z}}$。此步骤涉及全部 $\binom{n}{2}$ 二阶候选项，因此每次评估二阶损失函数及其梯度的成本为 $O(n^2 r)$。其次，对于每个满足 $m > 2$ 的 $m \in \mathbf{M}$，CINE 都会针对每个 $u \in \Omega_{h_m}$ 评估 $s_{h_m}(u;\boldsymbol{Z})$。每个得分需要执行 $O(h_m r)$ 次操作，而对得分进行指数化和归一化以获得 $q_m^{\mathrm{sub}}(u)$ 还需要额外执行 $O\{\binom{n}{h_m}\}$ 次操作。因此，对于每个目标阶数，提议构造的成本为 $O\{r h_m \binom{n}{h_m}\}$。第三，生成每个 m 阶候选项并检查其是否存在，需要处理其 m 节点索引，预计需要 $O(m)$ 次操作。处理 N_m 候选项因此对于每个目标阶数的期望采样成本为 $O(mB_m)$，因为 $\mathbb{E}(N_m) = B_m$。最后，CINE 使用所包含的候选项 $E_m^+ \cup E_m^-$，联合估计 $\boldsymbol{Z}$ 和 $\{\beta_m\}_{m\in\mathcal{M}}$。由于每个候选项需要执行 $O(mr)$ 次操作，且 $\mathbb{E}|E_m^-| \le B_m$，因此每次评估 CINE 损失及其梯度的期望成本为 $O[r\sum_{m\in\mathbf{M}} m(|E_m^+| + B_m)]$。因此，计入每个损失函数的一次评估后，总体期望计算成本

为 $O[rn^2 + \sum_{m\in\mathcal{M}}\{rh_m\binom{n}{h_m} + mr(|E_m^+| + B_m)\}]$, ，其中考虑了初步估计、提议构造、采样和最终优化。

对于完整估计，在每个超边阶数 $m \in \mathbf{M}$ 下，每次损失函数和梯度评估都涉及全部 $\binom{n}{m}$ 候选项。对每个候选项应用相同的 $O(mr)$ 计算，每次评估的成本为 $O\{r\sum_{m\in\mathbf{M}} m\binom{n}{m}\}$。接下来，我们比较两种方法在 $n \to \infty$ 下的计算成本。对于固定的一组超边阶数，$h_m < m$意味着 $h_m\binom{n}{h_m}/\{m\binom{n}{m}\} = (h_m/m)\prod_{j=h_m+1}^{m} j/(n-j+1) = O(n^{h_m-m}) \to 0$。如果 M 至少包含一个大于二的阶数，那么初步估计的成本也满足 $n^2/\{\sum_{m\in\mathbf{M}} m\binom{n}{m}\} \to 0$。此外，如果对于每个 $m \in$M 都有 $|E_m^+| + B_m = o\{\binom{n}{m}\}$，那么 $\{\sum_{m\in\mathbf{M}} m(|E_m^+| + B_m)\}/\{\sum_{m\in\mathbf{M}} m\binom{n}{m}\} \to 0$。综合这些关系可得$rn^2 + \sum_{m\in\mathbf{M}}\{rh_m\binom{n}{h_m} + mr(|E_m^+| + B_m)\} = o\{r\sum_{m\in\mathbf{M}} m\binom{n}{m}\}$，由此可知，CINE 的渐近计算成本远低于完整估计。这里，条件 $|E_m^+| + B_m = o\{\binom{n}{m}\}$ 仅要求存在的超边数量和采样预算的增长速度低于候选超边的数量；在渐近稀疏情形下，这是一个较弱的要求。

因此，CINE从两方面节省计算：每个高阶损失函数仅涉及存在的超边和采样得到的负样本，且提议分布由低阶候选项构建，其数量的增长速度远慢于目标阶候选项。这一组合也使CINE有别于现有策略。截断与投影会改变数据，从而改变估计目标，而CINE保留每一条存在的超边以及完整估计的目标（命题2）。均匀负采样保留该目标，但其大多数抽样缺乏信息量，而局部病例-对照抽样只有通过对每个目标阶候选项评分才能找到有信息量的负样本。CINE则通过低阶分数找到它们，且第3 节表明，CINE保持了完整估计误差界的阶。

算法 1 跨阶信息引导负采样估计（CINE）输入：节点集 $[n]$；保留阶数M，其中 $2 \in \mathrm{M}$；存在的超边集合$\{E_m^+\}_{m\in\mathcal{M}}$；潜在维度 r；抽样预算$\{B_m : m \in \mathrm{M}\setminus\{2\}\}$；混合权重 ϵ。
$\widehat{\quad}$ $\widehat{\quad}$输出：估计的潜在位置 $\boldsymbol{Z}$ 以及截距$\{\beta_m\}_{m\in\mathcal{M}}$。 ~ ~1：最小化二阶损失以得到$(\boldsymbol{Z}, \beta_2)$。2：对每个 $m \in \mathrm{M}\setminus\{2\}$执行 3：根据在固定的 $\widetilde{\boldsymbol{Z}}$处对$\Omega_{h_m}$ 上的得分计算 q_m^{sub}，并由(4)得到 q_m 。4：抽取 N_m ~泊松分布(B_m)，并从q_m中有放回地独立采样 N_m 个候选项。$\widehat{\ }$5：去除重复候选项和存在的超边，得到 E_m^-。6：对每个 $e \in E_m^+ \cup \widehat{E}_m^-$计算 $q_m(e)$和 $\pi_m(e)$。7：结束循环 8：联合最小化二阶损失与(5)中的校正损失，得到CINE$\widehat{\ }$ $\widehat{\ }$估计量$\{\boldsymbol{Z}, \{\beta_m\}_{m\in\mathcal{M}}\}$。

在实践中，我们使用 L-BFGS-B 和解析梯度来实现算法1 ，并通过约化坐标参数化来施加中心化约束。补充材料的 B.1 节介绍了初始化和停止准则。

备注*2*（在时间预算下选择保留阶数）。上述成本分析$\widetilde{\ }$为选择 M 提供了一个简单指南。在计算 $\boldsymbol{Z}$后，我们记录 t_{ref} 从 $\binom{n}{2}$ 阶候选项构建三阶提议分布所需的时间。由于从h 阶候选超边构建提议分布的成本为 $O\{rh\binom{n}{h}\}$，忽略常数后，从 h阶候选项构建提议分布所需的时间$\widehat{t}(h) = t_{\mathrm{ref}}\, h\binom{n}{h}/\{2\binom{n}{2}\}$。由于提议构造是成本随阶数增长最快的步骤，保留阶数集合 M 需要大致$\sum_{m\in M}{}_{m>2}\,\widehat{t}(h_m)\widetilde{\ }$，用于在已花费于 $\boldsymbol{Z}$的时间之外进行提议构造。给定时间预算，只要预测时间保持在预算范围内，我们就按从低到高的顺序添加阶数。由于一个阶数的提议分布可以从任何更低的保留阶数构建，因此可以跳过某个中间阶数，以使更高阶数仍处于预算范围内。例如，在示例*1*中，$M = \{2, 3, 4, 5, 7\}$，且 $(n, m_{\max}) = (300, 7)$，从而避免从所有六阶候选项构建提议分布。确定 M 后，剩余时间决定预算比，因为采样和损失函数评估的成本随 B_m增长。在 $n = 100$ 和 $m_{\max} = 7$下，集合 $\{2, 3, 4, 5, 7\}$ 已经在大约

两秒（图3）。

3 理论

本节我们为 CINE 提供理论保证。在本节中，保留阶数集合 M 是固定的，并且不依赖于 n。我们首先介绍一些记号。令 $\boldsymbol{Z}^* = (\boldsymbol{z}_1^*, \ldots, \boldsymbol{z}_n^*)^\top$ 和 $\{\beta_m^*\}_{m\in\mathcal{M}}$ 分别表示真实潜在位置矩阵和阶数特定截距。给定 $\boldsymbol{Z}^*$，阶数为 m 的存在超边数量的条件期望为 $\mu_m = \mathbb{E}(|E_m^+| \mid \boldsymbol{Z}^*) = \sum_{e\in\Omega_m} p_m(e; \boldsymbol{Z}^*, \beta_m^*)$。我们进一步定义 $\varrho_{m,n} = \mu_m / \binom{n}{m} = \sum_{e\in\Omega_m} p_m(e; \boldsymbol{Z}^*, \beta_m^*)/\binom{n}{m}$，其表示这些候选超边中存在超边的期望比例。对于每个满足 $m > 2$ 的 $m \in$ M，令 F_m 表示全数据的信息，即式（3）中完整伯努利损失的阶数为 m 的分量的费舍尔信息；令 J_m 表示式（5）中校正损失所贡献的期望信息。这些矩阵表示为：

$$F_m = \sum_{e\in\Omega_m} p_m(e; \boldsymbol{Z}^*, \beta_m^*)\{1 - p_m(e; \boldsymbol{Z}^*, \beta_m^*)\}\nabla\eta_m(e; \boldsymbol{Z}^*, \beta_m^*)\nabla\eta_m(e; \boldsymbol{Z}^*, \beta_m^*)^\top,$$

$$J_m = \sum_{e\in\Omega_m} \frac{p_m(e; \boldsymbol{Z}^*, \beta_m^*)\{1 - p_m(e; \boldsymbol{Z}^*, \beta_m^*)\}\pi_m(e)}{p_m(e; \boldsymbol{Z}^*, \beta_m^*) + \{1 - p_m(e; \boldsymbol{Z}^*, \beta_m^*)\}\pi_m(e)}\nabla\eta_m(e; \boldsymbol{Z}^*, \beta_m^*)\nabla\eta_m(e; \boldsymbol{Z}^*, \beta_m^*)^\top,$$

其中，$\nabla\eta_m(e; \boldsymbol{Z}^*, \beta_m^*)$ 表示对数优势 $\eta_m(e; \boldsymbol{Z}, \beta_m)$ 关于潜在位置参数和截距参数的梯度，并在 $(\boldsymbol{Z}^*, \beta_m^*)$ 处进行评估。矩阵 F_m 量化了所有m阶候选项所提供的信息，而 J_m 则衡量了 CINE 采样后校正损失所保留的信息。

3.1 抽样性质

我们首先研究提议分布的理论性质。我们从对初步估计的以下假设开始。

假设 1（提议的独立性）。 对于每个 $m \in M$（满足 $m > 2$），用于构造 q_m 的潜在位置估计 $\boldsymbol{Z}$ 与阶为 m 的超边指示变量 $\{A_e^{(m)}: \quad e \in \Omega_m\}$ 相互独立，并满足 $\mathbb{P}\{\|\widetilde{\boldsymbol{Z}}\|_{\mathrm{F}} < \infty\} = 1$。

假设 1 包含两个部分。首先，独立性条件确保，在给定 $\boldsymbol{Z}$ 的条件下，阶 m 指示变量仍然相互独立，并保持其原有的出现概率 $p_m(e; \boldsymbol{Z}^*, \beta_m^*)$。CINE 和 CINE++ 均满足该条件，因为阶 m 的估计仅使用较低阶数，因此本节中的所有结果均适用于这两种方法，其中 $\boldsymbol{Z}$ 表示构造 q_m 的估计量。其次，有限性条件使每个 $\exp\{\tau_m s_{h_m}(u; \boldsymbol{Z})\}$ 以及它们在有限集合 Ω_{h_m} 上的和几乎处处有限且严格为正，因此概率 $q_m^{\mathrm{sub}}(u)$ 是有效的。基于这一假设，我们建立所提出抽样策略的以下性质。

命题 2（总体目标的保持）。 在假设 *1* 下，对于每个$m \in \mathcal{M}$ 满足 $m > 2$ 以及每个有限备选参数 $(\boldsymbol{Z}', \beta_m')$，以下不等式几乎处处成立：

$$\mathbb{E}\{\widehat{\ell}_m(\boldsymbol{Z}', \beta_m') - \widehat{\ell}_m(\boldsymbol{Z}^*, \beta_m^*) \mid \widetilde{\boldsymbol{Z}}\} \geq 0.$$

当且仅当对每个 $p_m(e; \boldsymbol{Z}', \beta_m') = p_m(e; \boldsymbol{Z}^*, \beta_m^*)$ $e \in \Omega_{m\cdot}$，等式成立。此外，我们有：

$$\left(\boldsymbol{Z}^*, \{\beta_m^*\}_{m\in\mathcal{M}}\right) = \underset{\substack{\boldsymbol{Z}', \{\beta_m'\}_{m\in\mathcal{M}} \\ \sum_{i=1}^n \boldsymbol{z}_i' = 0}}{\arg\min} \mathbb{E}\left\{\sum_{m\in\mathcal{M}} \widehat{\ell}_m(\boldsymbol{Z}', \beta_m')\right\} = \underset{\substack{\boldsymbol{Z}', \{\beta_m'\}_{m\in\mathcal{M}} \\ \sum_{i=1}^n \boldsymbol{z}_i' = 0}}{\arg\min} \mathbb{E}\{\ell(\boldsymbol{Z}', \{\beta_m'\}_{m\in\mathcal{M}})\}.$$

命题2 表明，期望CINE损失与期望完整伯努利损失具有相同的极小化解。因此，尽管我们的抽样策略提高了估计样本中存在的超边所占的比例，偏移项 $-\log \pi_m(e)$ 在（5）中保留了完整伯努利损失的优化目标。因此，命题2 为通过最小化 CINE 损失来估计 $\boldsymbol{Z}$ 和 $\{\beta_m\}_{m\in\mathcal{M}}$ 提供了理论依据

由（5）构造。接下来，我们量化 CINE 保留了多少全数据信息。**定理1**（信息保证）。在假设*1*下，对于每个 $m \in \mathcal{M}$，满足 $m > 2$，以及每个固定采样预算 $0 < B_m < +\infty$，如果存在常数 $\rho_m > 0$ 使得$\pi_m(e) \geq \rho_m p_m(e;\ \boldsymbol{Z}^*,\ \beta_m^*)$ 对于每个 $e \in \Omega_m$都成立，那么我们有：

$$\frac{\rho_m}{1+\rho_m} F_m \preceq J_m \preceq F_m.$$

定理 1将 CINE 保留的信息与所有m阶候选超边所能提供的信息联系起来。特定阶的量 ρ_m 控制着保留信息比例的下界。当 ρ_m 较大时，该界保证负采样造成的信息相对损失较小。由于二阶损失被完整计入，每个保留的高阶项都会为共享潜在位置贡献至少其全数据信息的 $\rho_m/(1+\rho_m)$。推论 1 通过采样预算表示 ρ_m。

推论1。在定理 *1* 的条件下，若存在常数 $a \in (0,1]$使得 $q_m(e) \geq a\, p_m(e;\, \boldsymbol{Z}^*, \beta_m^*)/\mu_m$ 对每个 $e \in \Omega_m$均成立，则几乎必然地，定理 *1* 中的结论对满足 $\rho_m \geq \frac{aB_m}{\mu_m + aB_m p_{m,\max}}$的某个 ρ_m 成立，其中 $p_{m,\max} := \max_{e\in\Omega_m} p_m(e;\, \boldsymbol{Z}^*, \beta_m^*)$。此外，我们有：

$$\frac{aB_m}{\mu_m + aB_m} F_m \preceq J_m \preceq F_m.$$

推论 1 中的附加条件 $q_m(e) \geq a\, p_m(e; \boldsymbol{Z}^*, \beta_m^*)/\mu_m$要求分配给每个候选超边的提议概率至少达到其归一化出现概率的 a 一定比例。该条件自动成立：补充材料中的命题 A.2~表明，对于每个 $\boldsymbol{Z}$，只要 a 远离零，且真实得分有界、 $\epsilon > 0$，该条件就成立；同时，式（4）中的第二项会随着 a 越来越~$\boldsymbol{Z}$ 准确，以及低阶分数与高阶得分越吻合而增大。在此

条件下，推论 1 给出了 ρ_m 关于 B_m、 μ_m和$p_{m的下界}$max。特别地，对于固定的 $a > 0$，当 $B_m/\mu_m \to \infty$时，所保留的全数据信息比例保证趋近于 1。例如，如果提议概率与归一化出现概率相匹配（$a = 1$），那么第 $B_m = 4\mu_m$ 4 节中使用的采样预算在每个目标阶数下至少保留 80% 的全数据信息。因此，预算比是在计算时间和保留的信息之间进行权衡；估计误差方面，定理 2 误差水平中的 n/B_m 项也表明了这一点。

3.2 非渐近误差界

⌢在本小节中，我们为 CINE 估计量 $\boldsymbol{Z}$ 和$\widehat{\{\beta_m\}_{m\in\mathcal{M}}}$建立非渐近误差界。在该模型下，真实参数 $\boldsymbol{Z}^*$ 和 $\{\beta_m^*\}_{m\in\mathcal{M}}$ 在中心化约束 $\mathbf{1}^\top \boldsymbol{Z} = 0$下最小化期望完整伯努利损失 $\mathbb{E}\{\ell(\boldsymbol{Z},\{\beta_m\}_{m\in\mathcal{M}})$ 。分析依赖于以下假设。

假设2. 对于固定常数 $L > 0$和 $c_p \in (0,1)$，$\max_{i\in[n]} \|\boldsymbol{z}_i^*\|_2 \le L$和$\max_{m\in\mathcal{M}} \max_{e\in\Omega_m} p_m(e;\boldsymbol{Z}^*,\beta_m^*) \le 1 - c_p$。

假设 2 中的第一个条件对真实潜在位置的 ℓ_2-范数作出界定，确保随着节点数量增加，得分 $s_m(e;\boldsymbol{Z}^*)$ 始终一致有界。潜在因子的相关界也见 Yuan 和 Qu (2023)。第二个条件要求超边出现概率一致地与 1 保持正距离。因此，$0 < \varrho_{m,n} \le 1 - c_p$，所以在所有 m 阶候选超边中，存在超边的期望比例也与 1 保持正距离。对于本文所考虑的稀疏超图，这一条件自然成立。Ma 等人 (2020) 在分析潜空间模型时也采用了相关的边概率上界。

假设 3. 存在一个正数常量 c_1，使得 $\lambda_{\min}\{(\boldsymbol{Z}^*)^\top \boldsymbol{Z}^*/n\} \ge c_1$。

假设 3 对以下各项的特征值施加了下界$(\boldsymbol{Z}^*)^\top \boldsymbol{Z}^*/n$。该条件

确保 $\boldsymbol{Z}^*$ 具有满列秩，从而能够根据内积，在正交变换意义下稳定地恢复潜在位置。在 Wu 等人 (2026) 中也出现了类似的特征值条件。下面的定理给出了 CINE估计量的估计误差上界。

定理 2（非渐近误差界）。在假设 2 和 3 下，若存在充分大的常数 $K > 0$，使得 $n\varrho_{2,n} \geq K \log n$、$n^{m-1}\varrho_{m,n} \geq K \log n$ 和 $B_m \geq Kn\log n$ 对 $m \in \mathcal{M} \setminus \{2\}$ 成立，则对于误差水平 $\omega_{2,n} = 1/(n\varrho_{2,n})$ 和 $\omega_{m,n} = 1/(n^{m-1}\varrho_{m,n}) + n/B_m$，在 $m > 2$ 下，以下界以至少 $1 - C/n$ 的概率成立，其中常数 $C > 0$ 与 n 无关：

$$\frac{1}{\sqrt{n}} \inf_{O^\top O = I_r} \|\widehat{\boldsymbol{Z}}O - \boldsymbol{Z}^*\|_{\mathrm{F}} \leq C\sqrt{\log^3 n \min_{m\in\mathcal{M}} \omega_{m,n}}, \tag{7a}$$

$$|\widehat{\beta}_m - \beta_m^*| \leq C\sqrt{\log n\, \omega_{m,n}}, \qquad m \in \mathcal{M}. \tag{7b}$$

定理 2 给出了 CINE估计量的估计误差上界，并且这些界对每个其 CINE 损失不超过真实参数对应损失的中心化参数都成立，而不仅仅对全局最小化点成立。每个保留阶数 m 都有其自身的误差水平 ω_{mn}，潜在位置则以信息量最大的保留阶数对应的水平进行估计，因此，保留更高阶数只会使 (7a) 中的界更紧。每个截距 β_m 均以其自身阶数对应的水平进行估计。对于 $m > 2$，ω_m的第一项 $_n$ 反映了阶数为 m 的存在的超边数量有限，而第二项 n/B_m则源于负采样；负采样将阶数为 m 的观测的有效数量从 μ_m 减少到 $\mu_m B_m/(\mu_m + B_m)$。当对于每个 $m \in \mathcal{M}$ 都有 $\log n\, \omega_{mn} = o(1)$，且 $\log^3 n \min_{m\in\mathcal{M}} \omega_{mn} = o(1)$ 时，定理 2 中的所有界均随 $n \to \infty$趋于零，从而证明 CINE 估计量具有一致性（潜在位置允许正交变换）。补充材料的 A.6 节表明，当像第 5 节中那样从有限多个阶数中根据数据选择保留阶数时，这些界仍可同时成立。

为评估负采样对参数估计的影响，我们将 CINE 估计量的误差界与完整估计的误差界进行比较。具体而言，我们将潜在位置矩阵及阶数特定截距的完全估计量 $\bar{}$ $\bar{}$定义为 $\{\boldsymbol{Z}, \{\beta_m\}_{m\in\mathcal{M}}\}$ 在 ($\ell(\boldsymbol{Z}, \{\beta_m\}_{m\in\mathcal{M}})$ 中完整伯努利损失 $\ell(\boldsymbol{Z}, \{\beta_m\}_{m\in\mathcal{M}}$的极小化器，且满足 $\sum_{i=1}^{n} \boldsymbol{z}_i = \mathbf{0}$。下面的命题给出了完全估计量的估计误差界。

命题 3（完全估计量的误差界）。 在假设 2 和3下，假设$n^{m-1}\varrho_{m,\ n} \geq K \log n$ 对每个 $m \in \boldsymbol{M}$ 都成立，包括 $m = 2$，且 $K > 0$ 足够大。对于某个与 n无关的常数 $C > 0$ ，在给定 $\boldsymbol{Z}^*$的条件下，关于超边指示变量，以至少 $1 - C/n$ 的概率，完全估计量满足定理2 中的误差界，其中每个$\omega_{m,\ n}$中的项 n/B_m 均被移除。

定理2 与命题3 仅在$\omega_{m,\ n}$中的项 n/B_m 上有所不同。如果 $\mu_m = O(B_m)$，那么 $n/B_m = O\{1/(n^{m-1}\varrho_{m,\ n})\}$，因此 CINE 与完整估计的所有误差界具有相同的阶。特别是，当每个 $m > 2$都满足 $B_m \asymp \mu_m$ 时，CINE 损失的计算平均至多使用 $\mu_m + B_m = O(\mu_m)$ 个候选超边，而完整伯努利损失则需要使用该阶的全部 $\binom{n}{m}$ 个候选项。因此，CINE 在大幅降低计算成本的同时，保留了完整估计误差界的阶。

4 仿真

我们通过三个示例评估了 CINE，分别考察估计准确性与计算时间、采样预算和信号强度的影响，以及二阶信息和高阶信息的作用。

示例1。 我们考察了估计精度和计算时间如何随 n和最大超边阶数 $m_{\max}$变化。我们设置 $n \in \{100, 300\}$和 $m_{\max} \in \{3,5,7\}$，其中 $c = 1.8$。此处， c 表示信号强度，用于控制 $\boldsymbol{Z}$的尺度。我们固定期望平均参与度 $m\mu_m/n$，以及包含一个阶-m 超边的期望数量 a

节点，在 $2\mu_2/n = 12$ 对于 $m = 2$ 以及 $m\mu_m/n = 20$ 对于 $m = 3$，……，$m_{\max}$。预算比固定为 $B_m/\mu_m = 4$ 对于 $m = 3$，……，7。

示例 2。 我们研究了采样预算和信号强度对估计精度的影响。我们固定了 $n = 100$ 和 $m_{\max} = 7$，并将 $c \in \{1。4、1。6、1。8、2。0\}$ 以及预算比 $B_m/\mu_m \in \{4、6、8\}$ 设为不同值，针对 $m = 3$、……、7 进行了实验。对于 $m = 2$，我们将期望平均参与度固定为$2\mu_2/n = 12$；对于 $m = 3$，则固定为 $m\mu_m/n = 20$、……、7。

示例 3。 我们研究了估计精度如何随二阶及高阶信息的数量变化而变化。我们固定了 $n = 100$、$m_{\max} = 7$ 和 $c = 1.8$。对于 $m = 3$、……、7，我们考虑了 $2\mu_2/n \in \{8、12、16\}$ 和 $m\mu_m/n \in \{16、20、24\}$的所有组合。对于 $m = 3$、……、7，预算比固定为 $B_m/\mu_m = 4$。

对于每种设置，我们均生成了 100 次重复，潜在维度为 $r = 2$。对于每个节点i，我们设定 $\boldsymbol{z}_i^{\mathrm{raw}} = c\exp(0.12\xi_i)(\cos\ \theta_i,\ \sin\ \theta_i)^\top + \boldsymbol{\varepsilon}_i$，其中 $\theta_i \sim \mathrm{Uniform}(0,\ 2\pi)$、$\xi_i \sim N(0,\ 1)$ 和 $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0},\ 0.12^2\boldsymbol{I}_2)$ 在各节点间相互独立，并将潜在位置中心化，使其满足 $\boldsymbol{z}_i = \boldsymbol{z}_i^{\mathrm{raw}} - n^{-1}\sum_{j=1}^n \boldsymbol{z}_j^{\mathrm{raw}}$。对于每个 $m = 2$，……，$m_{\max}$ 和 $e \in \Omega_m$，我们独立生成 $A_e^{(m)}\ |\boldsymbol{Z} \sim$ 伯努利$\{p_m(e)\}$，其中 β_m 的取值使得，对于各示例中指定的期望平均参与度 $m\mu_m/n$，$\sum_{e\in\Omega_m} p_m(e;\ \boldsymbol{Z},\ \beta_m) = \mu_m$。

我们将 CINE 实现中的 $\mathcal{M} = \{2,\ \ldots\ldots,\ m_{\max}\}$ 设为默认的保留阶数集合。对于$(n,\ m_{\max}) = (300,\ 7)$ 在示例1中，我们采用了 $\mathcal{M} = \{2,\ 3,\ 4,\ 5,\ 7\}$，其选取方式如备注2所述。我们将混合权重固定为 $\epsilon = 0.20$。

为进行比较，我们考虑了以下方法。(1) FULL：使用阶数为 2、……、$m_{\max}$ 的所有候选项进行完整估计。(2) Uniform：在 CINE 中固定 $\epsilon = 1$，同时保持保留阶数和采样预算不变。(3) 截断：仅保留二阶和三阶数据，并使用这两个阶数下的所有候选项拟合伯努利模型。(4) 投影：保留原始二阶数据，并将每个

阶数至少为三的现有超边投影到其所有三节点子集上。重复的三元组视为同一个现有超边，并使用所得的二阶和三阶数据拟合伯努利模型。(5) Yuan 和 Qu （2023）提出的联合链接嵌入（JLE）。(6) Zhen 和 Wang （2023）提出的超图嵌入模型（HEM）。(7) Wu 等人 （2026）提出的潜在嵌入方法，以下简称 HGE。对于 HGE，我们固定其估计得到的节点嵌入，并按照 CINE 的做法，根据（1）获得出现概率。

我们使用三项性能指标。(1) 潜在位置恢复：令 $\boldsymbol{Z}$ 表示中心化后的估计潜在位置矩阵，并令 $\boldsymbol{O}^* = \arg\min_{\boldsymbol{O}^\top \boldsymbol{O} = \boldsymbol{I}_r} \|\boldsymbol{Z}\boldsymbol{O} - \boldsymbol{Z}\|_F^2$。我们计算相对均方误差（相对 MSE）$\mathrm{RelMSE}_{\boldsymbol{Z}} = \|\widehat{\boldsymbol{Z}}\boldsymbol{O}^* - \boldsymbol{Z}\|_F^2 / \|\boldsymbol{Z}\|_F^2$. (2) 出现概率恢复：令 $\widehat{p}_m(e)$ 表示 m阶候选项 e 的估计出现概率。对于 HEM 和 JLE，我们使用其各自拟合模型所定义的概率。对于 JLE，我们为每个拟合阶数设置单独的截距，以适应不同阶数之间稀疏性的差异。对于其余方法，使用各方法的估计值计算 $\widehat{p}_m(e) = p_m(e; \boldsymbol{Z}, \beta_m)$。对于每个评估阶数，我们计算 $\mathrm{RelMSE}_{p,m} = \{\sum_{e \in \Omega_m} [\widehat{p}_m(e) - p_m(e)]^2\} / \{\sum_{e \in \Omega_m} p_m(e)^2\}$。我们还报告所有阶数的总体相对均方误差 $\mathrm{RelMSE}_{p,\mathrm{all}} = \{\sum_{m=2}^{m_{\max}} \sum_{e \in \Omega_m} [\widehat{p}_m(e) - p_m(e)]^2 / \{\sum_{m=2}^{m_{\max}} \sum_{e \in \Omega_m} p_m(e)^2\}$。(3) 计算时间：我们记录每种方法的实际运行时间，包括初始化。为降低大型候选集所带来的计算成本，我们采用了用于数据生成和概率评估的加速程序，具体说明见补充材料的 B.2 节。

对于示例1，表1 显示，CINE 在估计精度与计算时间之间取得了良好的平衡。在上方面板（$n = 100$ 和 $m_{\max} = 5$）中，FULL 使用了所有候选项，获得了最低的相对均方误差（$\mathrm{RelMSE}_{\boldsymbol{Z}} = 0.013$，$\mathrm{RelMSE}_{p,\ \mathrm{all}} = 0.032$），但平均需要 187.581 秒。相比之下，Uniform、截断和投影的速度快得多（0.459–0.545 秒），但相对均方误差较大

表1：最大超边阶数为 $m_{\max} = 5$ 的示例1中的相对均方误差和计算时间。条目为 100 次重复的均值（标准差）。在除 FULL 之外的方法中，每个图版的每列中均值相对均方误差最低的结果以粗体显示，第二低的结果加下划线。

面板 A： $n = 100$。							
方法	相对均方误差						时间（秒）
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	
FULL	*0.013（0.002）*	*0.032（0.003）*	*0.025（0.003）*	*0.065（0.012）*	*0.093（0.024）*	*0.115 (0.038)*	*187.581 (57.363)*
CINE++	**0.020** (0.003)	**0.050** (0.006)	**0.039**(0.004)	**0.111** (0.026)	**0.165** (0.051)	**0.200** (0.068)	0.416（0.033）
CINE	0.022 (0.004)	0.059 (0.010)	0.045 (0.006)	0.133 (0.038)	0。210（0.102）	0.264 (0.155)	0.780 (0.269)
Uniform	0.029（0.004）	0。129 (0.040)	0.065 (0.008)	0.380 (0.121)	1.129（0.706）	2。224 (1.947)	0.545（0.190）
截断 0.030 (0.004)		0.076 (0.008)	0。057 (0.006)	0.170（0.036）	0.273（0.074）	0.364（0.125）	0.472 (0.137)
投影	0.023 (0.004)	0。116（0.073）	0.047（0.007）	0.320（0.150）	1。236（0.978）	2.845 (3.034)	0.459（0.133）
JLE	0.442 (0.155)	1.348 (2.127)	0.569（0.434）	2.853(4.922)	18.901(49.314)	128.818(364.153)	1493.131(361.728)
HEM	0.621 (0.215)	1.000 (0.000)	1.000 (0.000)	0.999 (0.000)	0.991 (0.003)	0.990（0.014）	801。447（143.901）
HGE	0.058 (0.025)	0.298（0.123）	0.121（0.025）	0.937 (0.365)	3.582 (2.880)	8.700 (10.412)	6.658 (2.732)
面板 B： $n = 300$。							
方法	相对均方误差						时间（秒）
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	
FULL†	-	-	-	-	-	-	-
CINE++	**0.016**(0.001)	**0.054** (0.004)	0.051 (0.004)	**0.117** (0.014)	**0.170** (0.027)	**0.213** (0.041)	13.091 (5.255)
CINE	0.017（0.001）	0.058 (0.004)	0.054 (0.004)	0.127 (0.015)	0.184（0.030）	0.234（0.041）	11.292 (5.201)
Uniform	0.025 (0.002)	0.114 (0.013)	0.099 (0.009)	0.415 (0.117)	0.889 (0.357)	1.462 (0.718)	2.904（0.357）
截断 0.027 (0.002)		0.085（0.007）	0.080 (0.006)	0.191（0.023）	0.282 (0.045)	0.373（0.071）	8.835 (1.735)
投影	0.018(0.001)	**0.054** (0.006)	**0.050**(0.004)	0.127 (0.039)	0.205 (0.118)	0.289 (0.233)	6.332 (1.151)
JLE†	-	-	-	-	-	-	-
HEM†	-	-	-	-	-	-	-
HGE	0.042 (0.023)	0.195 (0.088)	0.158 (0.050)	0.881 (0.750)	2.768（4.051）	5.559 (9.986)	65。894（27.516）

† 图版 B 未列出 FULL、JLE 和 HEM，因为每种方法进行 100 次重复实验都需要超过 30 天 s.

（RelMSE$_{\boldsymbol{Z}}$：0.023–0.030；RelMSE$_{p,\ \text{all}}$：0.076–0.129）。JLE、HEM 和 HGE 的计算时间（6.658–1493.131 秒）长于这三种方法，相对均方误差仍高于 CINE。除 FULL 外（RelMSE$_{\boldsymbol{Z}} = 0.022$；RelMSE$_{p,\ \text{all}} = 0.059$），CINE 的相对均方误差低于所有竞争方法，而计算时间仅为 0.780 秒，比 FULL 快约 240 倍。CINE++ 进一步降低了相对均方误差。在下方面板（$n = 300$）中，两个面板所报告的所有方法，其潜在位置恢复效果均有所提升；CINE 和 CINE++ 再次实现了最低的潜在位置误差。补充材料中的表 B.1–B.3 显示了相同的趋势。

图 3 展示了图 1(e) 沿备注 2中的保留阶数变化所体现的权衡。我们使用 $n = 100$ 和 $m_{\max} = 7$ 重新考察示例1，并使用四个嵌套集合的保留

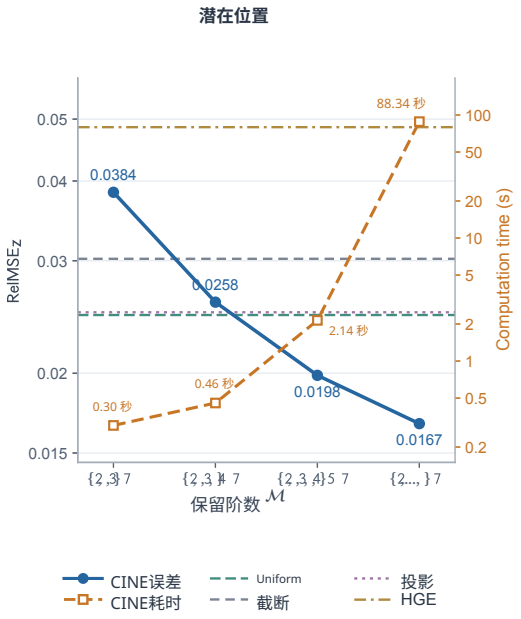


图3：示例1中，采用不同保留阶数时 CINE 的潜在位置恢复和计算时间，其中 $n = 100$ and$m_{\max} = 7$。蓝色实线和橙色虚线分别表示对数坐标轴上的 RelMSE$_Z$ and 计算时间。水平线标示 Uniform、截断、投影和 HGE 的 RelMSE$_Z$，其计算时间分别为 0.57、0.40、0.43 和 8.30 秒。

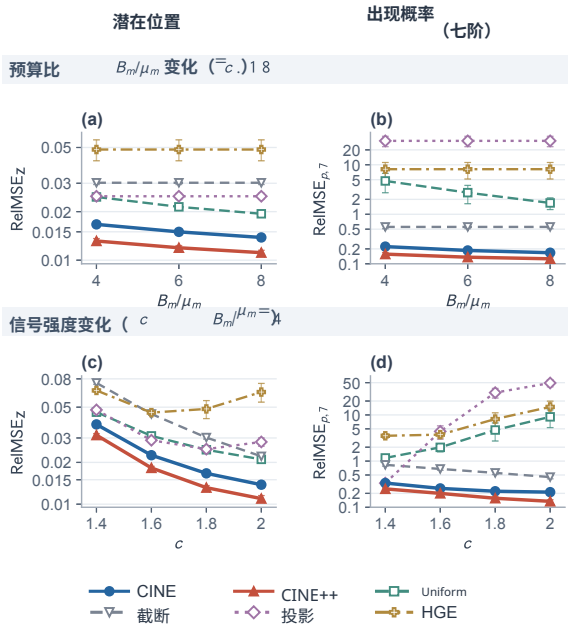


图 4：示例 2 中的潜在位置和七阶出现概率恢复。左列和右列分别报告 RelMSE$_Z$and RelMSE$_{p,\ 7}$。面板 (a)–(b) 展示预算比 B_m/μ_m 在 $m = 3, \ldots, 7$ 时的变化，$c = 1.8$; 面板 (c)–(d) 展示信号强度 c 在 $B_m/\mu_m = 4$ 时的变化。

阶数进行拟合，分别为 {2、3、7} 到 {2、.....、7}。每增加一个阶数，都会降低 RelMSE$_Z$，但会延长计算时间。当 $\mathcal{M} = \{2、3、4、5、7\}$ 时，CINE 在 2.14 秒内达到 RelMSE$_Z = 0.020$，优于 Uniform 、截断和投影方法。此时，CINE 的速度和准确性均优于 HGE；后者在 8.30 秒内达到 RelMSE$_Z = 0.049$ （补充材料中的表 B.3）。保留所有阶数可将 RelMSE$_Z$ 降至 0.017，但耗时 88.34 秒。

对于示例 2，图4 展示了潜在位置和七阶出现概率的相对均方误差如何随预算比和信号强度变化。补充材料中的表 B.4–B.7 报告了完整结果。当 $c = 1.8$，预算比从 4 提高到 8，使相对均方误差$_Z$ 从 CINE 的0。017 降至0。014 ，同时计算时间从88。3 增至140。5 秒（图4(a) 和表 B.6），这体现了预算比所控制的权衡。当预算比固定为 4 时，两种

随着信号增强，CINE 和 CINE++ 的相对均方误差均有所下降。相比之下，随着信号强度增加，HGE、投影和 Uniform 得出的出现概率相对均方误差更大。对于投影和 HGE，当 c从 1。8 增至 2。0 时，潜在位置的相对均方误差也有所增加。

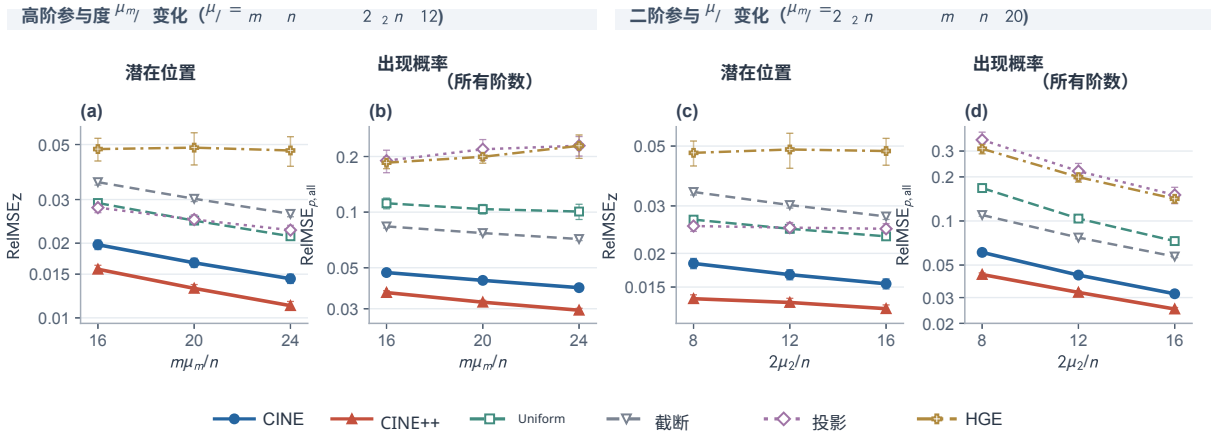


图5：示例 3 中潜在位置和总体出现概率恢复。左列和右列分别报告 RelMSE$_Z$ 和 RelMSE$_{p,\text{all}}$。面板 (a)–(b) 改变共同的高阶参与度 $m\mu_m/n$，同时将二阶参与固定为 $2\mu_2/n = 12$。面板 (c)–(d) 改变二阶参与，同时将高阶参与固定为 $m\mu_m/n = 20$，取值范围为$m = 3, \ldots\ldots, 7$。

对于示例 3，图 5 总结了潜在位置和总体出现概率恢复，其中面板 (a)–(b) 改变高阶参与度，面板 (c)–(d) 改变二阶参与度。补充材料中的表 B.8–B.10 报告了完整结果。在保持另一项不变的情况下，增加二阶参与度或高阶参与度，均会降低 CINE 和 CINE++ 的相对均方误差。例如，当高阶参与度固定为 $m\mu_m/n = 20$ 时，将二阶参与度从8 增加到 16，使 CINE 的潜在位置相对均方误差从 0。018 降至 0。015，其总体概率相对均方误差则从0。061降至 0。032。

《补充材料》第 B.4 节通过比较 CINE 与 JLE、HGE 和 HEM，评估模型设定错误下的稳健性；数据分别由示例1中的模型以及与各竞争方法对应的机制生成。

在竞争机制下，与 JLE、HGE 和 HEM 的比较中，CINE 的 RelMSE$_{p,\ \text{all}}$ 分别仅增长 3。9、1。2 和 2。1 倍。在示例1的模型下，JLE、HGE 和 HEM 的误差则分别增长 72。4、1。6 和 59。4 倍。

5 真实数据

我们将 CINE 应用于基于统计学家多属性数据集（MADStat）构建的合著超图。该数据集包含 1975—2015 年间 36 种期刊中 47,311 位作者发表的 83,331 篇论文的书目信息和引文记录（Ji等（2022））。我们聚焦于 2005—2015 年间发表在 《统计学年鉴》、《生物统计学》、*Journal of the American StatisticalAssociation* 和 《皇家统计学会杂志》*B*辑 上的论文。在这些论文中，我们利用合著图筛选作者：该图连接至少共同发表过两篇论文的节点对，而不考虑团队规模。我们将其最大连通分量中的 335 位作者称为核心成员。纳入所有至少涉及两位核心成员的论文的全部作者后，该集合扩展至 644 位作者。随后，我们保留完整作者团队均属于该集合的所有合著论文，共得到 811 篇论文。这些论文定义了 625 条超边，其中每个不同的完整作者团队由一条超边表示。在 625 条超边中，有 408 条涉及三位或以上作者，其中 126 条涉及至少四位作者。随着阶数升高，观测日益稀疏：超边数量从三阶的 282 条减少至四阶的 89 条和五阶的 20 条。五阶以上仅有 11 条六阶超边、3 条七阶超边、2 条八阶超边和 1 条十阶超边。因此，我们考虑了两组保留阶数，$\mathcal{M}_1 = \{2、3\}$ 和 $\mathcal{M}_2 = \{2、\ldots\ldots、5\}$，并通过估计的潜在位置及其在留出数据上的预测性能比较两种 CINE 拟合结果。在两种拟合中，我们将潜在维度设为 $r = 2$，并将每个保留阶数的采样预算设为 $B_m = 16|E_m^+|$，其预算比高于模拟实验中的设置，因为 MADStat 超图较为

小。图 6展示了这些潜在位置，它们概括了共同署名的模式，也可能反映相关的研究兴趣。对于每组保留阶数，我们将属于不同团队数量最多的十位作者称为枢纽，并以姓名标注。两种拟合结果的前十名中有六位作者重合。当保留四阶和五阶时，Donglin Zeng、Joseph Ibrahim 和 Hua Liang 才进入前十名。Zeng、Ibrahim 和 Liang 所参与的不同团队数量分别从 10 增加至 18、从 7 增加至 15，以及从 9 增加至 18。规模更大的团队还揭示了这些作者的其他共同署名联系。在图版 (b) 中，Zeng 的 29 位不同合作者中，有 17 位在我们的数据集中从未与他出现在任何两作者或三作者团队中。同样，Ibrahim 的 22 位合作者中有 13 位、Liang 的 27 位合作者中有 17 位从未与他们出现在任何两作者或三作者团队中。

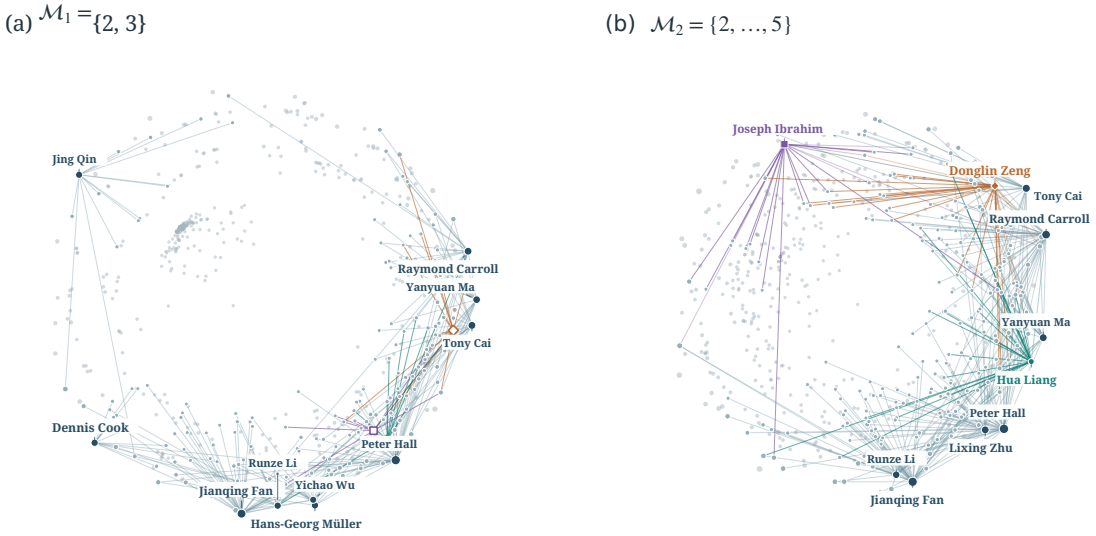
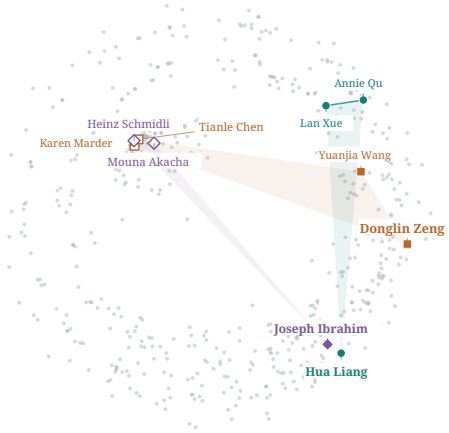


图 6：2005年至2015年 MADStat 合著超图中的核心作者。面板 (a) 和 (b) 展示了 CINE 在$\mathcal{M}_1 = \{2,3\}$ 和$\mathcal{M}_2 = \{2,\ldots,5\}$下估计的 644 位作者的潜在位置。

纳入规模更大的团队也会改变估计得到的潜在位置，并揭示不同的关联模式。例如，Yuanjia Wang 是 Zeng 的 17 位新增合作者之一。图 7 显示，在 $\mathcal{M}_2$ 下，Wang 比在 $\mathcal{M}_1$ 下更靠近 Zeng，而且也更靠近 Karen Marder 和 Tianle Chen。这四位作者曾共同参与一项由五位作者合著的

(a) CINE, $\mathcal{M}_1 = \{2, 3\}$



(b) CINE, $\mathcal{M}_2 = \{2, \ldots, 5\}$

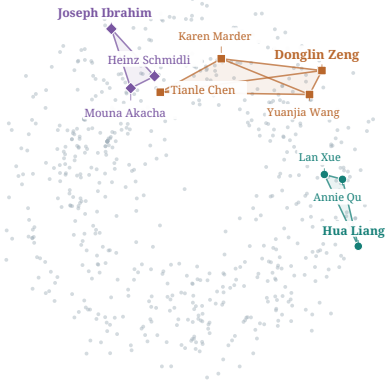


图7：基于2005年至2015年的合著超图估计的作者潜在位置。面板 (a) 和 (b) 展示了在 $\mathcal{M}_1 = \{2, 3\}$ 和 $\mathcal{M}_2 = \{2, \ldots, 5\}$ 下的 CINE 拟合结果。

亨廷顿病年龄依赖性风险预测研究。Wang 与 Marder 和 Chen 之间距离的缩短，与这三位作者此前在拟合样本之外开展的合作研究相符。在 2012年 的一项研究中，他们利用先证者及其亲属的数据，估计亨廷顿病发病年龄随 CAG重复长度 变化的分布。另一个例子涉及 Joseph Ibrahim。纳入规模更大的团队后，Mouna Akacha 和 Heinz Schmidli 在潜在空间中与他的距离更近。Ibrahim、Akacha 和 Schmidli 曾与 Donglin Zeng 和 Qingxia Chen 合作开展一项五位作者合著的研究，探讨多发性硬化症试验中复发事件的治疗效应。这一模式与他们此前关于相关事件时间、患者脱落和复发率的研究相符。

第三个例子涉及 Hua Liang、Annie Qu 和 Lan Xue。这三位作者都已出现在当前团队中$\mathcal{M}_1$, ，其中 Qu 和 Xue 彼此距离较近。在$\mathcal{M}_2$,下，这三位作者之间的节点对距离均有所缩短，其中 Qu 和 Xue 大幅向 Liang 靠近。他们早期的工作为这一模式提供了背景。Liang 和 Qu 此前曾从不同角度研究变量选择。Liang 和

合作者在加性部分线性模型中研究了这一问题，而 Qu 和 Cho 则为相关数据开发了惩罚二次推断函数。Xue 此前曾与这两位作者合作。她与 Liang 在一项拟合样本之外的研究中开发了用于加性系数模型的多项式样条估计量，而她与 Qu 则研究了相关数据中的模型选择。随后，这些研究方向在一篇由 Li Wang 参与的四位作者论文中汇合；该论文将多项式样条与惩罚二次推断函数相结合，以选择相关数据中的线性效应和非线性效应。综合来看，这些例子表明，较大规模的团队合作可能会揭示统计学家之间相关的研究兴趣，而仅从较小规模的团队合作中则不易看出这些兴趣。

为评估跨阶信息在预测中的作用，我们在留出数据上比较$\mathcal{M}_1$ 和$\mathcal{M}_2$ 下的 CINE。我们以 0.2 的概率将每个三阶候选项分配到测试集，这使 282 条三阶存在的超边中有 55 条进入测试集，其余 227 条保留在训练数据中。令$\mathcal{T}_3^{+}$和$\mathcal{T}_3^{-}$分别表示测试集中三阶的存在超边和负样本。我们用受试者工作特征曲线下面积（AUC）来衡量判别能力：

$$\mathrm{AUC} = \frac{1}{|\mathcal{T}_3^{+}|\,|\mathcal{T}_3^{-}|} \sum_{e \in \mathcal{T}_3^{+}} \sum_{f \in \mathcal{T}_3^{-}} \Big[\mathbf{1}\{\widehat{p}_3(e) > \widehat{p}_3(f)\} + \tfrac{1}{2}\mathbf{1}\{\widehat{p}_3(e) = \widehat{p}_3(f)\} \Big],$$

其中， $\widehat{p}_3(e)$ 是候选项 e的拟合出现概率。AUC 越大，判别能力越强。在$\mathcal{M}_1,$ 下，CINE 的 AUC 达到 0.737；保留四阶和五阶后，AUC 提升至 0.846。在$\mathcal{M}_2$ 下，采用相同的抽样预算时，Uniform 的 AUC 为 0.757。因此，由二阶信息引导负采样可将 AUC 从 0.757 提升至 0.846。由此可见，跨阶信息能够改善共同署名预测。

该论文的 C 节　补充材料比较了潜在位置和预测

性能，以及 CINE 与本数据集模拟研究中所考虑的其他方法的性能。总体而言，纳入更大的团队有助于 CINE 揭示具有相关研究兴趣的统计学家之间的关联，并且在$\mathcal{M}_2$条件下，CINE 在所考虑的方法中实现了最高的 AUC。

6 讨论

除了节省计算资源外，CINE 还使高阶信息得以应用。在 MADStat 分析中，保留四阶和五阶改变了估计得到的潜在位置。由于低阶信息能够指导高阶计算，同时不改变估计目标，用户可以根据时间预算选择保留阶数（注2），并将预算比设为所需的精度，而不必受完整估计不可行性的限制。

同样的分析也适用于记录不同规模群体交互的其他数据，例如通信、社会接触和生物医学共现数据，尤其是在小群体常见而大群体罕见的情况下。对于超大规模 n，超越成对亲和度的评分，以及逐个节点扩展候选项的采样器，都是自然的未来研究方向。**补充材料**。在线补充材料包含证明和额外的理论结果（第3节），以及额外的数值结果（第4节和第5节）。**披露声明**。作者声明不存在潜在利益冲突。

经费支持。Nie、Wang、Zhang 和 Zhu 获得了中国国家自然科学基金（编号：12571298）、中国教育部基础学科与交叉学科突破计划（JYB2025XDXM612）以及中央高校基本科研业务费的部分资助。

数据和代码可用性。MADStat数据（Ji等（2022））可在以下网址公开获取：https://github.com/ZhengTracyKe/MADStat。用于复现模拟

以及真实数据分析的代码以 zip 文件形式随补充材料一并提供。

参考文献

Agarwal, S., Branson, K. 和 Belongie, S.（2006），图上的高阶学习，载于‘第23届国际机器学习会议论文集’，ACM，第17–24页。

Battiston, F., Cencetti, G., Iacopini, I., Latora, V., Lucas, M., Patania, A., Young, J.-G. 和 Petri, G.（2020），‘超越成对交互的网络：结构与动力学’，《物理学报告》**874**，1–92。

Chodrow, P. S., Veldt, N. 和 Benson, A. R.（2021），‘生成式超图聚类：从块模型到模块度’，《科学进展》**7**(28)，eabh1303。

Contisciani, M., Battiston, F. 和 De Bacco, C.（2022），‘超图中超边与重叠社群的推断’，《自然·通讯》**13**，7229。

Dumitriu, I.和Wang, H.-X.（2026年），《一般非均匀超图随机块模型的最优与精确恢复》，《统计年鉴》 **54**(1)，48–73。

Fienberg, S. E.（2012年），《网络分析统计模型简史与开放挑战》，《计算与图形统计学杂志》**21**(4)，825–839。

Fithian, W.和Hastie, T.（2014年），《局部病例-对照抽样：不平衡数据集中的高效子抽样》，《统计年鉴》 **42**(5)，1693–1724。

Ghoshdastidar, D.和Dukkipati, A.（2017），《均匀超图划分：可证明的张量方法与抽样技术》，《机器学习研究杂志》 **18**(50)，1– 41。

Gutmann, M. U. 和 Hyvärinen, A.（2012年），《未归一化

统计模型的噪声对比估计及其在自然图像统计中的应用》，《机器学习研究杂志》**13**(11)，307–361。

Hoff, P. D., Raftery, A. E. 和 Handcock, M. S.（2002），《社会网络分析的潜在空间方法》，*Journal of the American Statistical Association* **97**（460），1090–1098。

Ji, P., Jin, J., Ke, Z. T. 和 Li, W.（2022），《统计学家的共被引网络与合著网络》，《商业与经济统计学杂志》**40**（2），469–485。

Lerner, J. 和 Lomi, A.（2023），《多元交互网络的关系型超事件模型》，《英国皇家统计学会会刊*A*辑：社会统计》**186**（3），577–600。

Lucas, M., Gallo, L., Ghavasieh, A., Battiston, F. 和 DeDomenico, M.（2026），《动力学中的高阶网络可约性》，《自然·通讯》**17**，1551。

Lyu, Z.、Xia, D. 和 Zhang, Y.（2023年），“高阶网络的潜在空间模型与广义张量分解”，《计算与图形统计学杂志》**32**(4)，1320–1336。

Ma, Z.、Ma, Z. 和 Yuan, H.（2020），“利用边协变量对大型网络进行通用潜在空间模型拟合”，《机器学习研究杂志》**21**(4)，1–67。

Raftery, A. E.、Niu, X.、Hoff, P. D. 和 Yeung, K. Y.（2012年），“使用病例对照近似似然对潜在空间网络模型进行快速推断”，《计算与图形统计学杂志》**21**(4)，901–919。

Shi, F.和 Evans, J.（2023），《研究内容与情境之间令人意外的组合与影响力相关，并且会随着来自遥远学科的科学界局外人而出现》，《自然·通讯》**14**，1641。

Turnbull, K.、Lunagómez, S.、Nemeth, C.和 Airoldi, E. M.（2024），《超图数据的潜在空间建模》，*Journal of the American Statistical Association* **119**（548），2634–2646。

Wu, S.、Xu, G. 和 Zhu, J.（2026），“一种用于建模具有多重性的非均匀高维稀疏超图的通用潜在嵌入方法”，*Journal of the AmericanStatistical Association*，doi: 10.1080/01621459.2026.2724030。

Yoon, S.-e.、Song, H.、Shin, K. 和 Yi, Y.（2020），“超图中何时以及需要多少高阶信息？超边预测案例研究”，发表于《2020年万维网大会论文集》，ACM，第2627–2633页。

Yu, X. 和 Zhu, J.（2025），“具有多样性和异质流行度的超图建模”，*Journal of the American Statistical Association* **120**（551），1491–1502。

Yuan, Y. 和 Qu, A.（2023），“用于多层级链接预测的高阶联合嵌入”，*Journal of the American Statistical Association***118**（543），1692–1706。

Zhen, Y. 和 Wang, J.（2023），《通过图嵌入检测一般超图中的社区》，*Journal of the American Statistical Association* **118**(543)，1620–1629。

Zhou, D.、Huang, J. 和 Schölkopf, B.（2006），《超图学习：聚类、分类与嵌入》，载于《神经信息处理系统进展》，第19卷，麻省理工学院出版社，第1601–1608页。

Zhu, G.、Wang, H. 和 Ren, M.（2026），《有向超图：张量表示、嵌入建模与社区检测》，arXiv:2611.29655。

RongzhuoNie、Menghao Wang、Sanguo Zhang、MingyangRen 和 Guojun Zhu 所著《用于非均匀超图的跨阶信息引导估计》的补充材料

本补充材料的组织结构如下。第A 节介绍理论推导，包括模型可辨识性、辅助引理以及主要结果的证明。第B 节给出实现细节、更多模拟结果以及模型设定错误比较。第C 节通过比较 CINE 与模拟研究中考察的其他方法估计得到的作者位置，以及它们在留出数据上的预测性能，分析这些方法在 MADStat 数据上的表现。

理论推导

证明按照主文稿的顺序展开。我们首先证明主文稿第 2.1 节中的可识别性。接着证明第 3.1 节的三项抽样结果，以及第 3.2 节中 CINE 和完全估计量的误差界。两个估计量所用的技术引理与定理 2 的证明一并给出。最后两个小节进一步讨论信息和提议覆盖。

位置不使用粗体，并记作 $M_m = \binom{n}{m}$。所有概率陈述均以真实位置 Z^* 为条件，并固定真实截距 $\{\beta_m^*\}$ 。我们对每个阶的候选集采用独立伯努利模型。总体恒等式针对一个固定的保留阶数集合 $\mathcal{M}$ ，其中包含二阶。若根据同一组指示变量选择集合并以此为条件，则伯努利概率可能发生变化。对于误差界，我们还给出一种同时论证，以便在有限个阶数中进行筛选。维数 r 和最大阶数均固定。在 CINE 损失函数中，混合权重 $\epsilon \in (0, 1]$ 固定，抽样预算为确定值，且在最终优化过程中提议概率和偏移量保持不变。公式 (1)–(5) 指主文稿中的公式。CINE 的误差界见该文稿中的 (7a)–(7b)，完全估计的误差界则见其中的命题 3。

A.1 命题1的证明

证明。 设（Z，$\{\beta_m\}$）和（Z'，$\{\beta'_m\}$）是对每个候选项具有相同概率的两个中心化参数值。记 $G = ZZ^\top$ 和 $G' = Z'Z'^\top$。由于 $2 \in \mathcal{M}$，二阶对数优势相等可得：

$$G'_{ij} - G_{ij} = c, \qquad i \neq j, \qquad c = \sqrt{r}\,(\beta_2 - \beta'_2).$$

中心化意味着 $G\mathbf{1} = G'\mathbf{1} = 0$。因此，$G' - G$ 的每个对角元素都是 $-(n-1)c$。结合 $P = I_n - \mathbf{1}\mathbf{1}^\top/n$，可得：

$$G' - G = -ncP.$$

若 $c > 0$，则 $G = G' + ncP$ 在 $\mathbf{1}^\perp$ 上正定，因为 G' 半正定，且 P 是该子空间上的恒等映射。由此可得 $\mathrm{rank}(G) \geq n - 1 > r$。若 $c < 0$，同样的论证给出 $\mathrm{rank}(G') \geq n - 1 > r$。两个 Gram 矩阵的秩至多为 r，因此 $c = 0$、$G' = G$ 和 $\beta_2' = \beta_2$。在每一高阶，(2) 中的得分现在相同。该阶任意候选项的对数优势相等即给出 $\beta_m' = \beta_m$。

最后，将 $\sum_i a_i z_i$ 映到 $\sum_i a_i z_i'$ 的映射保持平方范数，因为二者均由 $a^\top G a$ 给出。因此，它是这两个张成空间之间一个良定义的等距变换。将这些张成空间的标准正交基延拓为 $\mathbb{R}^r$ 的基，可得正交矩阵 O，使得 $Z' = ZO$。反之，每一个这样的变换都保持全部概率。该论证同样涵盖秩亏位置矩阵。

□

A.2 命题 2 的证明

证明。固定 $m > 2$ 并以 $\widetilde{Z}$ 为条件。根据假设 1 以及模型中指示变量的相互独立性，目标指示变量仍相互独立，且服从原有概率的伯努利分布。分析设置固定候选集和采样预算，并规定在给定初步估计的条件下，抽样随机化与目标指示变量相互独立。泊松构造给出了条件概率生成函数：

$$\mathbb{E}\left\{ \prod_{e \in \Omega_m} t_e^{C_e^{(m)}} \,\middle|\, \widetilde{Z} \right\} = \exp\left[B_m \left\{ \sum_e q_m(e) t_e - 1 \right\} \right] = \prod_e \exp\{ B_m q_m(e)(t_e - 1) \}.$$

因此，候选计数是均值为 $B_m q_m(e)$ 的相互独立的泊松变量，并且与目标指示变量独立。每条存在的超边都会被纳入，而负例以 $\pi_m(e) > 0$ 的概率被纳入。因此，

$$\begin{aligned} \mathbb{P}(A_e^{(m)} = 1, R_{me} = 1 \mid \widetilde{Z}) &= p_m(e), \\ \mathbb{P}(A_e^{(m)} = 0, R_{me} = 1 \mid \widetilde{Z}) &= \{1 - p_m(e)\}\pi_m(e). \end{aligned} \tag{A.1}$$

在此证明中，定义：

$$c_e = p_m(e) + \{1 - p_m(e)\}\pi_m(e), \qquad D_e = \frac{p_m(e)}{c_e},$$
$$D_e' = \frac{p_m(e; Z', \beta_m')}{p_m(e; Z', \beta_m') + \{1 - p_m(e; Z', \beta_m')\}\pi_m(e)}.$$

这些是真实参数与备择参数下的纳入概率和条件发生概率。它们均严格介于零与一之间。在两者处使用相同的提议分布和偏移项

参数值处，并在求导过程中保持固定。对校正损失取期望，可得：

$$\begin{aligned}&\mathbb{E}\{\widehat{\ell}_m(Z',\beta'_m)-\widehat{\ell}_m(Z^*,\beta^*_m)\mid\widetilde{Z}\}\\&\quad=\sum_{e\in\Omega_m}c_e\ \mathrm{KL}\{\mathrm{Bern}(D_e)\parallel\mathrm{Bern}(D'_e)\}\geq 0.\end{aligned}\tag{A.2}$$

这证明了命题 2 的条件超额损失不等式。由于所有权重均为正，当且仅当 $D'_e = D_e$ 对每个候选项成立时，等号成立。逻辑斯蒂链接函数是一一映射，并且两边使用相同的有限偏移项。因此，等号成立等价于 $\eta_m(e; Z', \beta'_m) = \eta_m(e; Z^*, \beta^*_m)$，或者等价地，所有候选项的出现概率均相等。

对于 CINE 损失，ℓ_2 是二阶损失，即式（3）的二阶分量，其中包含所有二阶候选项且偏移项为零。其无条件超额期望损失为：

$$\sum_{e\in\Omega_2}\mathrm{KL}\{\mathrm{Bern}(p_2(e))\parallel\mathrm{Bern}(p_2(e;Z',\beta'_2))\}.$$

对于每个高阶，取（A.2）的无条件期望。每一项均非负，且当且仅当每个所包含候选项的概率均一致时，它们的和为零。期望完整伯努利损失具有相同的等号成立条件，因为其超额期望损失是在所有阶数上对相应伯努利散度求和。因此，这两个联合总体准则具有相同的最小化解。

上述所用的期望均为有限值。对于固定的有限备选参数对数优势 η'，存在超边的期望贡献为 $p\log(1+\pi e^{-\eta'}) \leq p\log(1+e^{-\eta'})$。负例的期望贡献为 $(1-p)\pi\log(1+e^{\eta'}/\pi) \leq (1-p)\log(1+e^{\eta'})$。后一个不等式成立，是因为对于每个 $c > 0$，$x\log(1+c/x)$ 在 $x > 0$ 上递增。这些界对 $\pi \in (0, 1]$，均一致，且候选项只有有限个。当 $n \geq r + 2$ 时，命题 1 将公共极小化器集合明确地确定为：

$$\{(Z^*O, \{\beta^*_m\}_{m\in\mathcal{M}}) : O^\mathsf{T}O = I_r\}.$$

因此，真实参数属于命题 2 中的公共 argmin 集合。通过上述散度论证，即使没有这一秩阈值，这两个集合也相等。□

A.3 定理1的证明

证明。 给定一个有限的初步估计，并省略阶和候选项索引。CINE 与完整估计的信息系数之比为：

$$\frac{\pi}{p+(1-p)\pi}.$$

由于 $p+(1-p)\pi-\pi=p(1-\pi)\geq 0$，该比率至多为一。若 $\pi\geq\rho_m p$，则其至少为：

$$\frac{1}{p/\pi+1-p}\geq\frac{1}{\rho_m^{-1}+1}=\frac{\rho_m}{1+\rho_m}.$$

将这些不等式乘以半正定矩阵 $p_m(e)\{1-p_m(e)\}\nabla\eta_m(e)\nabla\eta_m(e)^\intercal$，并对候选项求和。由此，对于满足所述条件的每个初步估计，可得到定理1中的两个矩阵界。因此，J_m 在此比较中表示以初步估计为条件的期望信息。当纳入概率界成立时，这些矩阵界几乎处处成立。

□

A.4 推论1的证明

证明。固定一个保留阶数 $m>2$，并以一个满足推论1中提议界的有限初步估计为条件。根据假设 1，这种有限性几乎处处成立。我们用 J_m 表示 CINE 提议分布的信息矩阵。对于每个候选项和 $0<\mu_m<\infty$，有限伯努利模型给出 $0<p_m(e)<1$。根据所假设的提议界，

$$B_m q_m(e)\geq\frac{aB_m p_m(e)}{\mu_m}\qquad(e\in\Omega_m).$$

对于 $x\geq 0$，不等式 $e^x\geq 1+x$ 意味着 $1-e^{-x}\geq x/(1+x)$。因此，泊松纳入概率公式给出：

$$\begin{aligned}\frac{\pi_m(e)}{p_m(e)}&=\frac{1-\exp\{-B_m q_m(e)\}}{p_m(e)}\\&\geq\frac{aB_m}{\mu_m+aB_m p_m(e)}\geq\frac{aB_m}{\mu_m+aB_m p_{m,\max}}.\end{aligned}$$

因此，对于以下选择，所有候选项均满足纳入概率条件：

$$\rho_m=\frac{aB_m}{\mu_m+aB_m p_{m,\max}},$$

这就证明了第一个断言。

为建立更精确的信息界，我们直接使用$e^x-1\geq x$ 和提议界。对于每个候选项，

$$\begin{aligned}\frac{\pi_m(e)}{p_m(e)+\{1-p_m(e)\}\pi_m(e)}&=\left[1+\frac{p_m(e)}{\exp\{B_m q_m(e)\}-1}\right]^{-1}\\&\geq\left(1+\frac{\mu_m}{aB_m}\right)^{-1}=\frac{aB_m}{\mu_m+aB_m}.\end{aligned}$$

该系数至多为一，因为 $p_m(e)+\{1-p_m(e)\}\pi_m(e)-\pi_m(e)=p_m(e)\{1-\pi_m(e)\}\geq 0$。每个候选项贡献半正定矩阵 $p_m(e)\{1-p_m(e)\}\nabla\eta_m(e)\nabla\eta_m(e)^\intercal$，并将其加到 F_m 上。将这些矩阵乘以系数界，并对 Ω_m 求和，可得：

$$\frac{aB_m}{\mu_m+aB_m}F_m\preceq J_m\preceq F_m.$$

该论证适用于每个满足所假设提议界的有限初步估计，因此在该推论的条件下确立了所述的几乎必然结论。

□

A.5 定理 2 的证明

在本小节中，$\theta = (Z,\ \{\beta_m\})$ 表示一个有限备选参数，而 $\theta^* = (Z^*,\ \{\beta_m^*\})$ 表示真实参数。为证明之用，定义：

$$T_2 = \mu_2, \qquad T_m = \frac{\mu_m B_m}{\mu_m + B_m} \quad (m > 2).$$

$\widehat{\ }$ $\widehat{\ }$记 $L(\theta) = \sum_m \ell_m(Z,\ \beta_m)$，并定义：

$$\gamma_m(e;\theta) = \eta_m(e; Z, \beta_m) - \eta_m(e; Z^*, \beta_m^*), \qquad \Delta_m(\theta) = \frac{1}{M_m} \sum_{e \in \Omega_m} \min\{\gamma_m(e;\theta)^2, 1\}.$$

这些量仅用于证明。常数 $c,\ C > 0$ 可在各个公式之间发生变化，并且仅依赖于定理中列出的固定量。它们均不依赖于备选参数。我们首先在整个有限参数空间上获得似然控制，然后恢复大多数行，最后利用二阶负样本界定每一行。

对于概率计算，在初步估计存在时，以可测方式选取它；否则，赋予它一个相对于二阶数据可测的任意有限值。任何辅助的平局打破随机性都与所有指示变量独立。该约定定义了例外数据集上的提议分布；下文的证明将确立，在所声称的事件上，原始二阶极小化器确实存在。因此，在该事件上，这一约定不会改变主文稿中的任何一个优化问题。在分析更高阶时，以二阶数据和该辅助随机性为条件。此时，初步估计和偏移量均已固定。在每个更高阶内，指示变量仍然相互独立，泊松候选计数也彼此独立，并且与这些指示变量独立。对于 CINE++，逐阶应用相同的条件化方式：分析第 m 阶时，以较低保留阶数的数据和采样随机性为条件，并按照上述方式以可测方式选取中间估计。这样便固定了用于构造 q_m 的估计，同时使第 m 阶的指示变量和候选计数具有相同的条件分布，因此下文的每一步同样适用于 CINE++。

真实参数假设的推论。设 $S_m = L^2\sqrt{\binom{m}{2}/r}$。真实得分位于$[-S_m, S_m]$。根据恒等式：

$$p_m^*(e) = \{1 - p_m^*(e)\} \exp\{\beta_m^* + s_m(e; Z^*)\},$$

假设2给出：

$$c_p e^{\beta_m^* - S_m} \le p_m^*(e) \le e^{\beta_m^* + S_m}.$$

进行平均化并使用$\mu_m = M_m \varrho_{m,\ n}$ 可得:

$$|\beta_m^* - \log \varrho_{m,n}| \le C, \qquad c\varrho_{m,n} \le p_m^*(e) \le C\varrho_{m,n}, \qquad 1 - p_m^*(e) \ge c_p. \tag{A.3}$$

这些断言仅涉及真实参数。

引理 A.1 (无约束似然控制)。 对于每个节点 i，令 L_{-i} 表示删除与其相关的 $n-1$二阶项后的 *CINE* 损失；保留所有高阶项。对于每个 $t \ge 1$，以至少 $1 - 2e^{-t}$的概率，对每个有限的 θ 及每个 i同时成立:

$$\widehat{L}(\theta) - \widehat{L}(\theta^*) \ge c \sum_m T_m \Delta_m(\theta) - C(nr \log n + t), \tag{A.4}$$

$$\widehat{L}_{-i}(\theta) - \widehat{L}_{-i}(\theta^*) \ge -C(nr \log n + t). \tag{A.5}$$

T同一事件也涵盖单独的二阶损失，无论是否删除一个相关集合 .

证明。 对于一个候选项，抑制 m， e 并写为:

$$c_e = p^* + (1 - p^*)\pi, \quad D^* = \frac{p^*}{c_e}, \quad D(\theta) = \mathrm{logit}^{-1}\{\eta(\theta) - \log \pi\}.$$

对二阶候选项取 $\pi = c_e = 1$ 。定义:

$$h^2(a,b) = (\sqrt{a} - \sqrt{b})^2 + (\sqrt{1-a} - \sqrt{1-b})^2, \qquad V_{\mathcal{I}}(\theta) = \sum_{e \in \mathcal{I}} c_e h^2\{D^*, D(\theta)\}.$$

这里的 $\mathcal{I}$ 要么是二阶候选项的一个确定性集合，要么是某个保留的高阶上的候选项集合。在后一种情况下，期望是以二阶数据和辅助随机性为条件的。包含律对固定的替代项给出:

$$\mathbb{E} \exp\left\{ -\frac{\widehat{L}_{\mathcal{I}}(\theta) - \widehat{L}_{\mathcal{I}}(\theta^*)}{2} \right\} = \prod_{e \in \mathcal{I}} \left[1 - \frac{1}{2} c_e h^2\{D^*, D(\theta)\} \right] \le \exp\{-V_{\mathcal{I}}(\theta)/2\}. \tag{A.6}$$

特别是，我们没有使用 logit函数或 logit 差值的任何上界。

我们给出一致化论证，因为有界参数覆盖论证在此处并不足够。设 $N = \sum_m M_m \le C n^{\max \mathcal{M}}$和 $\delta = (N+1)^{-4}$。将 $[0,1]$ 划分为长度至多为δ的区间，并根据包含每个 $D_e(\theta)$的区间对备选方案进行分组。每个内部阈值由以下各项的符号决定:

$$\eta_m(e; Z, \beta_m) - \log \pi_m(e) - \mathrm{logit}(k\delta), \qquad 0 < k\delta < 1,$$

关于 $d = nr + |\mathcal{M}|$ 实参数的至多二次多项式。令 $H \le CN/\delta$ 表示所使用的阈值数量。如有需要，加入常数多项式以使 $H \ge d$，而不改变

可实现符号模式的数量。根据 Goldberg 和 Jerrum （1995）的推论 2.1，包含零在内的模式数量至多为 $(16e H/d)^d$。在端点采用固定约定时，符号决定量化区间。因此， J 非空单元的数量满足：

$$\log J \le Cd \log(N/\delta) \le Cnr \log n. \tag{A.7}$$

对环境参数空间进行计数也涵盖其中心化子集。我们将这一计数分别应用于二阶候选项的确定性集合，以及在给定初步估计的条件下应用于每个高阶。在后一种情况下，阈值在条件分布下是固定的。

对于一个非空单元，令其坐标区间为 $[l_e, u_e]$ ，选取一个参数代表元 θ_j，并令 $D_{e,\ j} = D_e(\theta_j)$。其似然函数的上界为：

$$\prod_{e \in \mathcal{I}} \{u_e^{A_e}(1-l_e)^{1-A_e}\}^{R_e}.$$

该包络函数不一定是归一化的似然函数。其平方根似然比的期望至多为：

$$\exp\{-V_{\mathcal{I}}(\theta_j)/2 + 2N\sqrt{\delta}\}.$$

事实上，利用 $|\sqrt{x} - \sqrt{y}| \le \sqrt{|x-y|}$，每个纳入候选项的亲和度与代表元的亲和度之差至多为 $2\sqrt{\delta}$；未纳入的候选项贡献为 ~~1。~~ 此外，

$$|V_{\mathcal{I}}(\theta) - V_{\mathcal{I}}(\theta_j)| \le 4N\sqrt{\delta} \le 4$$

在一个单元内。由马尔可夫不等式以及对 J 个包络函数应用并集界可得，失败概率至多为 e^{-u}，

$$\widehat{L}_{\mathcal{I}}(\theta) - \widehat{L}_{\mathcal{I}}(\theta^*) \ge V_{\mathcal{I}}(\theta) - C(nr\log n + u + 1) \quad \text{for every finite } \theta. \tag{A.8}$$

将此结果无条件地应用于所有二阶候选项，以及应用于排除一个关联集合的 n 个集合中的每一个，并使用 $u = t + \log(n+1)$。它们的总失败概率至多为 e^{-t}。对于每个保留的高阶项，在条件下应用该结果，并取 $u = t + \log \max(1,\ |\mathcal{M}| - 1)$。对这些阶数应用并集界并进行积分，可得它们合计的失败概率至多为 e^{-t} 。将这些不等式求和，得到一个失败概率至多为 $2e^{-t}$的公共事件；固定的阶数已被吸收到 C中。该过程既不需要二阶损失函数与高阶损失函数之间相互独立，也不需要不同阶数的抽样流之间相互独立。

接下来需要建立亲和度与 logit 差之间的联系。对于所有 $u,\ a \in \mathbb{R}$，满足 $\sigma = \text{logit函数}^{-1}$，

$$h^2\{\sigma(u), \sigma(u+a)\} \ge c\,\sigma(u)\{1-\sigma(u)\}\min(a^2, 1). \tag{A.9}$$

为验证这一点，令 $\psi(u) = \text{反正切}(e^{u/2})$。则 $\psi'(u) = \sqrt{\sigma(u)\{1-\sigma(u)\}}/2$，其对数导数的绝对值至多为 $1/2$，且左侧为 $4\sin^2\{(\psi(u+a) - \psi(u))/2\}$。对于 $|a| \le 1$，对 ψ'进行积分即可得到所述的二次下界。对于 $|a| > 1$，由单调性可将界简化为 $a = 1$ 或 $a = -1$。

根据（A.3）和均匀混合，

$$\pi_m(e) \ge 1 - \exp\{-\epsilon B_m / M_m\} \ge c \min(B_m/M_m, 1).$$

因此，

$$c_e D^*(1-D^*) = \frac{p_m^*(e)\{1-p_m^*(e)\}\pi_m(e)}{p_m^*(e) + \{1-p_m^*(e)\}\pi_m(e)} \ge c\,\frac{T_m}{M_m}. \tag{A.10}$$

对于 $m = 2$，这可直接由 $T_2 = \mu_2$得到。对于 $m > 2$，首先使用 $c\varrho_{m,\,n} b/(\varrho_{m,\,n} + b)$ 作为下界，其中 $b = \min(B_m/M_m,\ 1)$，然后将 $B_m \le M_m$ 和 $B_m > M_m$ 分开。结合（A.8）–（A.10）即可证明（A.4）。式（A.8）中的非负亲和度给出（A.5）及其仅涉及二阶的版本。

□

引理 A.2（从截断的 logit 差异中恢复）。设 Z^* 是中心化的，并满足真实行界和假设 *3*。对于充分大的 n，存在常数 c、$C > 0$，使得每个满足 $\Delta_2(\theta) \le c$ 的有限备选参数都存在一个满足以下条件的正交矩阵 O：

$$\frac{1}{n}\sum_i \min\{\|O^\top z_i - z_i^*\|_2^2, 1\} + |\beta_2 - \beta_2^*|^2 \le C\Delta_2(\theta). \tag{A.11}$$

此外，设 $E_O(\theta) = n^{-1}\sum_i \min\{\|O^\top z_i - z_i^*\|_2^2, 1\}$。对于任意正交矩阵 O 和任意 $m \in \mathcal{M}$，如果$E_O(\theta) \le c$ 且 $\Delta_m(\theta) \le c$，则：

$$|\beta_m - \beta_m^*|^2 \le C\{E_O(\theta) + \Delta_m(\theta)\}. \tag{A.12}$$

此引理不要求备择项经过中心化，也不要求对其幅度作出限制。

证明。我们首先选取三个小型参考节点集合，其真实位置张成潜在空间。令$k = r+1$, $x_i = ((z_i^*)^\top,\ 1)^\top$，以及 $X = [Z^*, \mathbf{1}]$。中心化条件与信号假设推出：

$$n^{-1}X^\top X \succeq \min(c_1, 1) I_k, \qquad \|x_i\|_2 \le \sqrt{L^2+1}.$$

当 n充分大时，至多删除 $3k$ 行仍能使归一化二阶矩保持一个固定的正下界。若 I 是从剩余节点中均匀选取的 k 个互不相同节点构成的有序集合，则由柯西–比内公式可得：

$$\mathbb{E}\{\det(X_I)^2\} = \frac{k!\det(X_{\text{rem}}^\top X_{\text{rem}})}{(N_{\text{rem}})_k} \ge c > 0,$$

其中，$(N_{\text{rem}})_k$ 是下降阶乘。Hadamard 不等式将该平方行列式的上界限定为一个固定常数。因此，其概率至少为某个固定的 $p_0 > 0$, $\sigma_{\min}(X_I) \ge \kappa > 0$。依次选取三个互不相交的

有序集合 I, J, K，即可使这三个集合均具有该性质，且概率至少为 p_0^3。

对于给定的备选方案，令 $e_{ij} = \min\{\gamma_2(\{i,j\};\theta)^2,\ 1\}$ 以及 $e_{ii} = 0$。如果 $S = I \cup J \cup K$，则 Uniform 选择给出：

$$\mathbb{E}\left\{ \sum_{\substack{a<b \\ a,b\in S}} e_{ab} + \frac{1}{n}\sum_i \sum_{a\in S} e_{ia} \right\} \le C\Delta_2.$$

马尔可夫不等式和前述的正概率表明，存在三个互不相交的集合，满足：

$$\min\{\sigma_{\min}(X_I), \sigma_{\min}(X_J), \sigma_{\min}(X_K)\} \ge \kappa, \qquad \sum_{\substack{a<b \\ a,b\in S}} e_{ab} + \frac{1}{n}\sum_i \sum_{a\in S} e_{ia} \le C\Delta_2. \tag{A.13}$$

当 $\Delta_2 = 0$ 时，在该选择下误差总和几乎处处为零，因此结论仍然成立。如果Δ_2 足够小，则参考节点之间的每个 logit函数误差都未被截断，且绝对值至多为 $C\sqrt{\Delta_2}$。对于每个备选方案，这是一个确定性的存在性论证；不要求所选节点与数据具有独立性。

接下来，我们将在不假设二阶截距有界的情况下恢复它。令 $A = Z_I^*$、 $B = Z_J^*$，并令A'， B'为相应的备选行矩阵。当 $\alpha = \sqrt{r}(\beta_2 - \beta_2^*)$ 时，交叉块逻辑值给出：

$$A'B'^{\mathsf{T}} = AB^{\mathsf{T}} - \alpha \mathbf{1}\mathbf{1}^{\mathsf{T}} + E, \qquad \|E\|_{\mathrm{F}} \le C\sqrt{\Delta_2}.$$

$A'B'^{\mathsf{T}}$ 和 AB^{T} 的秩均至多为 $r < k$。令 $D = AB^{\mathsf{T}}$，秩一扰动的行列式恒等式给出：

$$0 = \det(D + E) - \alpha \mathbf{1}^{\mathsf{T}} \operatorname{adj}(D + E)\mathbf{1}.$$

在 $E = 0$处， α的系数为 $\det[A, \mathbf{1}]\det[B, \mathbf{1}]$，其绝对值至少为 κ^{2k}。其扰动为 $O(\sqrt{\Delta_2})$，而 $\det(D + E) = O(\sqrt{\Delta_2})$，因为 $\det D = 0$。所有真实块条目均有界。因此，

$$|\beta_2 - \beta_2^*| \le C\sqrt{\Delta_2}. \tag{A.14}$$

在每个参考集中，选取 r 行组成一个条件良好的 $r \times r$ 真实矩阵，记为 A_0， B_0， C_0。这样的行确实存在，因为：

$$\sum_{\substack{I_0 \subset I \\ |I_0| = r}} \det(Z_{I_0}^*)^2 = \det\{(Z_I^*)^{\mathsf{T}} Z_I^*\} \ge \kappa^{2r},$$

且仅有 $r+1$ 项；行界控制其最大奇异值。类似的选择适用于 J, K。设 A_0', B_0', C_0' 是它们对应的替代矩阵。由 (A.13) 和 (A.14)，三个交叉乘积$A_0'B_0'^{\mathsf{T}}$， $C_0'B_0'^{\mathsf{T}}$，以及 $C_0'A_0'^{\mathsf{T}}$ 均与相应的真实乘积相差不超过 $C\sqrt{\Delta_2}$ 。因此，它们均可逆，且一致地接近固定的良态矩阵。考虑以下精确恒等式：

$$A_0'A_0'^{\mathsf{T}} = (A_0'B_0'^{\mathsf{T}})(C_0'B_0'^{\mathsf{T}})^{-1}(C_0'A_0'^{\mathsf{T}}), \tag{A.15}$$

这意味着 $\|A_0'A_0'^{\mathsf{T}} - A_0A_0^{\mathsf{T}}\|_{\mathrm{F}} \le C\sqrt{\Delta_2}$。极分解以及矩阵平方根在一致正定矩阵上的 Lipschitz连续性，现可得到一个正交矩阵 O，满足 $\|A_0'O - A_0\|_{\mathrm{F}} \le C\sqrt{\Delta_2}$。若 P, Q 是两个正平方根，则 $P^2 - Q^2 = P(P-Q) + (P-Q)Q$；与 $P-Q$ 取 Frobenius内积，可将其范数上界为某个常数乘以 $\|P^2 - Q^2\|_{\mathrm{F}}$。由交叉乘积可得，对于 $B_0'O - B_0$ 和 $C_0'O - C_0$也有相同的界。特别地， $A_0'O$ 及其逆矩阵一致有界。

对于 A_0 块外的一个节点，令 $\gamma_2(i,\ I_0)$ 表示该节点与该块之间的 logit 差的行向量。这些连接给出：

$$z_i^{\mathsf{T}}O = \left[(z_i^*)^{\mathsf{T}}A_0^{\mathsf{T}} + \sqrt{r}\{\gamma_2(i, I_0) - (\beta_2 - \beta_2^*)\mathbf{1}^{\mathsf{T}}\}\right](A_0'O)^{-\mathsf{T}}.$$

如果所有这些连接误差的绝对值均不超过 $1/2$，则平方行误差至多为 $C\{\Delta_2 + \sum_{a\in I_0} e_{ia}\}$。如果至少有一个超过 $1/2$，则截断平方行误差至多为 $1 \le 4\sum_{a\in I_0} e_{ia}$。 A_0块中的行已有误差 $O(\sqrt{\Delta_2})$。对其进行平均并应用（A.13），即可证明（A.11）中的位置部分。

—

最后，固定一个正交矩阵 O，使得 $E_O(\theta) \le c$，并令 $G = \{i:\ \|O^{\mathsf{T}}z_i - z_i^*\| \le 1/2\}$。于是 $|G^c| \le 4nE_O(\theta)$，且 G 上的平方行误差之和至多为 $nE_O(\theta)$。完全包含于 G 中的 m 阶候选项所占的比例至少为 $1 - 4mE_O(\theta)$。对于包含于 G中的一条边，对齐行范数均至多为 $L + 1/2$。展开（2）中的每个成对内积，并应用柯西–施瓦茨不等式，可将平方得分差限定为 $C\sum_{i\in e}\|O^{\mathsf{T}}z_i - z_i^*\|^2$，因为 m 是固定的。每个节点出现在 $\binom{n-1}{m-1} = mM_m/n$ 个候选项中。求和可得：

$$\frac{1}{M_m}\sum_{\substack{e\in\Omega_m \\ e\subset G}}\{s_m(e;Z) - s_m(e;Z^*)\}^2 \le CE_O(\theta). \tag{A.16}$$

它们各自的评分差也由一个固定常数界定。如果 $|\beta_m - \beta_m^*|$ 超过该常数加一，则它们的所有 logit 差都会超过一，这与 $\Delta_m \le c$矛盾。因此，截距差是有界的。这些候选项上的 logit 差现在也有界，所以它们的未截断平方平均值至多为 $C\Delta_m$。利用 $\beta_m - \beta_m^* = \gamma_m(e) - \{s_m(e;\ Z) - s_m(e;\ Z^*)\}$，对这些候选项取平均，并调用（A.16），即可证明（A.12）。

□

引理 A.3 (从高一阶恢复)。 设 Z^* 为中心化的，并满足真实行界和假设 *3*；固定 $m \in \mathcal{M}$，使得 $m > 2$。对于充分大的 n，存在常数 c、$C > 0$，使得每个满足 $\Delta_m(\theta) \le c$ 的有限备选参数都存在一个满足以下条件的正交矩阵 O：

$$\frac{1}{n}\sum_i \min\{\|O^{\mathsf{T}}z_i - z_i^*\|_2^2, 1\} \le C\Delta_m(\theta). \tag{A.17}$$

证明。 在此证明中，记 $\phi(x) = \min\{x^2, 1\}$, $\bar{\Delta} = \Delta_m(\theta)$, $\vartheta_m = \{r\binom{m}{2}\}^{-1/2} \le 1$，以及 $\chi_{ij} = z_i^{\mathsf{T}}z_j - (z_i^*)^{\mathsf{T}}z_j^*$为 $i \ne j$。根据主文稿中的 (2),

$$\gamma_m(e;\theta) = \beta_m - \beta_m^* + \vartheta_m\sum_{\{i,j\}\subset e}\chi_{ij}.$$

我们使用两个初等不等式：对于实数 $x_1, \ldots, x_p$ 和 $0 < t \le 1$,

$$\phi(x_1 + \cdots + x_p) \le p \sum_{q=1}^{p} \phi(x_q), \qquad \phi(x) \le t^{-2}\phi(tx). \tag{A.18}$$

截距以及两个候选项共有的节点对会在对数几率差中相互抵消。对于互不相同的 i、j、k、l 和 $\mathcal{R} \subset [n] \setminus \{i, j, k, l\}$，且满足 $|\mathcal{R}| = m - 2$,

$$\gamma_m(\mathcal{R} \cup \{i, j\}) - \gamma_m(\mathcal{R} \cup \{k, j\}) - \gamma_m(\mathcal{R} \cup \{i, l\}) + \gamma_m(\mathcal{R} \cup \{k, l\}) = \vartheta_m \Xi_{ij,kl}, \tag{A.19}$$

其中 $\Xi_{ij,\,kl} = \chi_{ij} - \chi_{kj} - \chi_{il} + \chi_{kl}$。对于互不相同的 i、i' 和 $\mathcal{U} \subset [n] \setminus \{i, i'\}$，且满足 $|\mathcal{U}| = m - 1$,

$$\gamma_m(\mathcal{U} \cup \{i\}) - \gamma_m(\mathcal{U} \cup \{i'\}) = \vartheta_m \sum_{s \in \mathcal{U}} (\chi_{is} - \chi_{i's}). \tag{A.20}$$

对于一个节点 a，令 $\bar{\Delta}_a$ 为包含 a 的候选项中 $\phi\{\gamma_m(e;\theta)\}$ 的平均值；对于一个节点对 $\{a, b\}$，令 $\bar{\Delta}_{ab}$ 为包含二者的候选项中的平均值。均匀选取的节点或节点对给出 $\mathbb{E}\bar{\Delta}_a = \mathbb{E}\bar{\Delta}_{ab} = \bar{\Delta}$。当 n 足够大时，在避开给定两个节点的候选项中，$\phi\{\gamma_m(e;\,\theta)\}$ 的平均值至多为 $2\bar{\Delta}$；而在包含 a 且避开给定一个节点的候选项中，其平均值至多为 $2\bar{\Delta}_a$。

我们首先选取两个参考节点。令 (k, l) 为从不同节点中均匀选取的有序对，设 $\mathcal{V} = [n] \setminus \{k, l\}$，并令 $Q(k, l)$ 为 $\phi(\Xi_{ij,kl})$ 在 $\mathcal{V}$ 中所有不同节点的有序对 (i, j) 上的平均值。由 (A.19) 和 (A.18)，对于 (A.19) 中的四个候选项 e_q 以及任意可容许的 $\mathcal{R}$，有 $\phi(\Xi_{ij,kl}) \le 4\vartheta_m^{-2} \sum_{q=1}^{4} \phi\{\gamma_m(e_q;\theta)\}$。当 $(i, j, k, l, \mathcal{R})$ 均匀分布时，每个 e_q 在 Ω_m 上均匀分布，因此 $\mathbb{E}Q(k, l) \le 16\vartheta_m^{-2}\bar{\Delta}$。将马尔可夫不等式应用于 $Q(k, l)$、$\bar{\Delta}_k$、$\bar{\Delta}_l$ 和 $\bar{\Delta}_{kl}$，可得一对 (k, l) 满足：

$$Q(k, l) \le C\bar{\Delta}, \qquad \max(\bar{\Delta}_k, \bar{\Delta}_l, \bar{\Delta}_{kl}) \le 5\bar{\Delta}. \tag{A.21}$$

考虑 $\mathcal{V}$ 内部的节点对。定义 $f_i = \chi_{il} - \chi_{kl}$、$g_j = \chi_{kj}$ 以及 $\zeta_{ij} = \chi_{ij} - f_i - g_j = \Xi_{ij,\,kl}$。根据 (A.21)，$\phi(\zeta_{ij})$ 在 $\mathcal{V}$ 中不同节点的有序对上的平均值至多为 $C\bar{\Delta}$。设 $(i, i', \mathcal{U})$ 为均匀的，其中 i、$i' \in \mathcal{V}$ 不同，且 $\mathcal{U} \subset \mathcal{V} \setminus \{i, i'\}$、$|\mathcal{U}| = m - 1$ 不同。(A.20) 中的两个候选项在 $\mathcal{V}$ 内的候选项中均匀分布，因此由 (A.18) 得到 $\mathbb{E}\phi\{\sum_{s \in \mathcal{U}}(\chi_{is} - \chi_{i's})\} \le 8\vartheta_m^{-2}\bar{\Delta}$。由于：

$$\sum_{s \in \mathcal{U}} (\chi_{is} - \chi_{i's}) = (m - 1)(f_i - f_{i'}) + \sum_{s \in \mathcal{U}} (\zeta_{is} - \zeta_{i's}),$$

(A.18) 得：

$$\phi(f_i - f_{i'}) \le (2m - 1)\Big[\phi\Big\{\sum_{s \in \mathcal{U}} (\chi_{is} - \chi_{i's})\Big\} + \sum_{s \in \mathcal{U}} \phi(\zeta_{is}) + \sum_{s \in \mathcal{U}} \phi(\zeta_{i's})\Big].$$

i, s 和 i', s 均是 $\mathcal{V}$ 中由不同节点组成的均匀有序对，因此 $\mathbb{E}\phi(f_i - f_{i'}) \le C\bar{\Delta}$，并且存在 $i_0 \in \mathcal{V}$ 使得 $|\mathcal{V}|^{-1}\sum_{i\in\mathcal{V}}\phi(f_i - f_{i_0}) \le C\bar{\Delta}$。根据对称性 $\chi_{is} = \chi_{si} = f_s + g_i + \zeta_{si}$，同一个和等于 $(m-1)(g_i - g_{i'}) + \sum_{s\in\mathcal{U}}(\zeta_{si} - \zeta_{si'})$，从而得到 $j_0 \in \mathcal{V}$ 使得 $|\mathcal{V}|^{-1}\sum_{j\in\mathcal{V}}\phi(g_j - g_{j_0}) \le C\bar{\Delta}$。令 $\bar{\chi} = f_{i_0} + g_{j_0}$。则 $\chi_{ij} - \bar{\chi} = \zeta_{ij} + (f_i - f_{i_0}) + (g_j - g_{j_0})$，因此 $\phi(\chi_{ij} - \bar{\chi})$ 在 $\mathcal{V}$ 中不同节点的有序对上的平均值满足：

$$\frac{1}{|\mathcal{V}|(|\mathcal{V}|-1)}\sum_{\substack{i,j\in\mathcal{V}\\ i\neq j}}\phi(\chi_{ij} - \bar{\chi}) \le C\bar{\Delta}. \tag{A.22}$$

接下来考虑与一个参考节点 $a \in \{k, l\}$ 相交的节点对。令 $(j, j', k', \mathcal{R})$ 在不同的 j、j'、$k' \in \mathcal{V}$ 以及 $\mathcal{R} \subset \mathcal{V}\setminus\{j, j', k'\}$、$|\mathcal{R}| = m-2$ 下满足 Uniform。在 (A.19) 中用 (a, j, k', j') 替换 (i, j, k, l) 后，两个候选项包含 a，且避开另一个参考节点；另两个则位于 $\mathcal{V}$ 内。因此 $\mathbb{E}\phi(\Xi_{aj,k'j'}) \le 4\vartheta_m^{-2}(4\bar{\Delta}_a + 4\bar{\Delta}) \le C\bar{\Delta}$。由于 $\chi_{aj} - \chi_{aj'} = \Xi_{aj,k'j'} + (\chi_{k'j} - \bar{\chi}) - (\chi_{k'j'} - \bar{\chi})$，由 (A.22) 得到 $\mathbb{E}\phi(\chi_{aj} - \chi_{aj'}) \le C\bar{\Delta}$，因此存在 $j_a \in \mathcal{V}$，使得在 $\bar{\chi}_a = \chi_{aj_a}$ 下，

$$\frac{1}{|\mathcal{V}|}\sum_{j\in\mathcal{V}}\phi(\chi_{aj} - \bar{\chi}_a) \le C\bar{\Delta}. \tag{A.23}$$

令 $(i', \mathcal{U})$ 为均匀分布，且 $i' \in \mathcal{V}$ 和 $\mathcal{U} \subset \mathcal{V}\setminus\{i'\}$, $|\mathcal{U}| = m-1$。根据 (A.20)，$\mathbb{E}\phi\{\sum_{s\in\mathcal{U}}(\chi_{as} - \chi_{i's})\} \le 2\vartheta_m^{-2}(2\bar{\Delta}_a + 2\bar{\Delta}) \le C\bar{\Delta}$，并且：

$$\sum_{s\in\mathcal{U}}(\chi_{as} - \chi_{i's}) = (m-1)(\bar{\chi}_a - \bar{\chi}) + \sum_{s\in\mathcal{U}}\{(\chi_{as} - \bar{\chi}_a) - (\chi_{i's} - \bar{\chi})\}.$$

对由 (A.18) 得到的不等式两边取期望，并利用 (A.22) 和 (A.23)，可得 $\phi(\bar{\chi}_a - \bar{\chi}) \le C\bar{\Delta}$。因此：

$$\frac{1}{|\mathcal{V}|}\sum_{j\in\mathcal{V}}\phi(\chi_{aj} - \bar{\chi}) \le C\bar{\Delta}, \qquad a \in \{k, l\}. \tag{A.24}$$

对于节点对 $\{k, l\}$，令 $(k', l', \mathcal{R})$ 均匀分布，其中 $k', l' \in \mathcal{V}$ 和 $\mathcal{R} \subset \mathcal{V}\setminus\{k', l'\}$, $|\mathcal{R}| = m-2$ 各不相同。在 (A.19) 中，将 (i, j, k, l) 替换为 (k, l, k', l') 后，四个候选项分别包含两个参考节点、仅包含 l、仅包含 k，以及两者都不包含，因此 $\mathbb{E}\phi(\Xi_{kl,k'l'}) \le 4\vartheta_m^{-2}(\bar{\Delta}_{kl} + 2\bar{\Delta}_l + 2\bar{\Delta}_k + 2\bar{\Delta}) \le C\bar{\Delta}$。由于 $\chi_{kl} - \bar{\chi} = \Xi_{kl,k'l'} + (\chi_{k'l} - \bar{\chi}) + (\chi_{kl'} - \bar{\chi}) - (\chi_{k'l'} - \bar{\chi})$，由 (A.22) 和 (A.24) 可得 $\phi(\chi_{kl} - \bar{\chi}) \le C\bar{\Delta}$。结合三类对，

$$\binom{n}{2}^{-1}\sum_{i<j}\phi(\chi_{ij} - \bar{\chi}) \le C\bar{\Delta}. \tag{A.25}$$

最后，令 θ' 为具有相同位置 Z 和二阶截距 β_2^* 的替代方案。于是 $\gamma_2(\{i, j\}; \theta') = (\bar{\chi}_{ij} - \bar{\chi})/\sqrt{r}$，而 (A.25) 给出 $\Delta_2(\theta') \le C\Delta_m(\theta)$。当 c 足够小时，引理 A.2 适用于 θ'，并得到满足 (A.17) 的正交矩阵 O，因为 θ' 和 θ 具有相同的位置。

□

引理 A.4 （各个方向上的二阶负例）。在假设 2–3 下，存在常数

a、 c、 $C>0$，使得以至少 $1-Cne^{-cn}$ 的概率，对每个节点 i 和每个单位向量 $v\in\mathbb{R}^r$，都有：

$$\left|\left\{j\neq i : A^{(2)}_{\{i,j\}}=0,\ v^{\mathsf{T}}z^*_j\geq a\right\}\right|\geq cn. \tag{A.26}$$

证明。 对于单位 v，令 $x_j=v^{\mathsf{T}}z^*_j$。这些数均值为零，绝对值至多为 L，且均方至少为 c_1。因此，

$$\frac{1}{n}\sum_j(x_j)_+=\frac{1}{2n}\sum_j|x_j|\geq\frac{c_1}{2L}.$$

结合 $b=c_1/(4L)$，可得：

$$n^{-1}|\{j : x_j\geq b\}|\geq c_1/(4L^2).$$

取一个确定性的 $b/(2L)$-网覆盖单位球面。由于 r、 L、 c_1 均为固定值，其大小可由某个常数界定。对于一个网向量和一个节点 i，至少有 cn 个其他真实位置的投影至少为 b，且它们的二阶指示变量相互独立，其负概率至少为 c_p。由切尔诺夫界可知，在该集合中至少有 $c'n$ 个负样本，除非发生概率至多为 $e^{-c''n}$ 的事件。对网和所有节点应用并集界，失败概率至多为 Cne^{-cn}。对于任意单位 v，最近的网向量使任意投影的变化至多为 $b/2$。这就以 A.26证明了式($a=b/2$。

□

定理 *2* 的证明。 在整个证明中，我们可以取 n 充分大。事实上，$\varrho_{2,\ n}\leq 1$ 且 $n\varrho_{2,\ n}\geq K\log n$ 推出 $n/\log n\geq K$。增大 K 即可排除任何固定的有限整数集 $n\geq 2$。在引理 $t=\log n$ 中取 A.1 ，并将其事件与引理 A.4的事件取交。联合概率至少为$1-2/n-Cne^{-cn}\geq 1-C'/n$。在本证明中，记 $a_n=(nr+1)\log n$，并定义：

$$b_n=\log(e/\varrho_{2,n})+a_n/n\geq 1.$$

阶数固定，因此对于充分大的 n， $M_m=\binom{n}{m}\asymp n^m$ 在 $m\in\mathcal{M}$ 上一致成立。另外，由于

$$\tfrac{1}{2}\min(\mu_m,B_m)\leq T_m\leq\min(\mu_m,B_m)\quad(m>2).$$

定理 2 的密度和预算条件给出：

$$\min_{m\in\mathcal{M}}T_m\geq cKn\log n,\qquad\frac{a_n}{\min_{m\in\mathcal{M}}T_m}\leq\frac{C}{K}. \tag{A.27}$$

常数 K 将在下文的论证中选取得足够大。考虑整个中心化次水平集：

$$\mathcal{S}_n=\{\theta : \mathbf{1}^{\mathsf{T}}Z=0,\ \widehat{L}(\theta)\leq\widehat{L}(\theta^*)\}.$$

该集合包含真实参数；我们暂不假设极小化器存在。对于每个 $\theta \in \mathcal{S}_n$，（A.4）意味着：

$$\sum_m T_m \Delta_m(\theta) \le C a_n, \qquad \Delta_m(\theta) \le C a_n / T_m. \tag{A.28}$$

令 $m_\star \in \operatorname{argmax}_{m \in \mathcal{M}} T_m$。选取 K 足够大，使得 (A.28) 和 (A.27) 满足引理 A.2 和 A.3的小误差条件。当 $m_\star = 2$时，使用引理 A.2 ；否则使用引理 A.3 ，可得到一个正交对齐，使得 $E_O(\theta) \le C a_n / T_{m_\star}$。随后，利用引理 A.2 中的截距部分，可得对每个 $m \in \mathcal{M}$:

$$\frac{1}{n} \sum_i \min\{\|O^\mathsf{T} z_i - z_i^*\|^2, 1\} \le \frac{C a_n}{T_{m_\star}}, \qquad |\beta_m - \beta_m^*|^2 \le C a_n \Big(\frac{1}{T_{m_\star}} + \frac{1}{T_m} \Big) \le \frac{2 C a_n}{T_m}. \tag{A.29}$$

现在我们控制每一行，包括那些未能通过 (A.29) 准确拟合的行。令 a 为引理 A.4中的投影常数，设 $b = \min(a/2, 1/2)$，并定义：

$$G = \{j : \|O^\mathsf{T} z_j - z_j^*\| \le b\}.$$

由方程 (A.29) 可得 $|G^c| \le C n a_n / T_{m_\star}$。如有必要，在 (A.27) 中增大 K ，删除这些节点后，对每个节点和方向而言，(A.26) 中被删除的负样本至多占一半。对于 G中剩余的节点，对齐后的拟合投影至少为 $a/2$。该引理中的事件对所有方向均成立，并且允许任意删除这么多个节点，因此适用于数据依赖的 G 及其对齐。

令 $\widehat{L}_i = \widehat{L} - \widehat{L}_{-i}$ 为与 i 相连的二阶损失之和。由次水平不等式和（A.5）可得：

$$\widehat{L}_i(\theta) \le \widehat{L}_i(\theta^*) + C a_n \le C n \log(e/\varrho_{2,n}) + C a_n. \tag{A.30}$$

最后一个不等式是确定性的：每个真实的二阶损失至多为 $\log 2 + |\beta_2^*| + L^2/\sqrt{r}$，且（A.3）给出了 $|\beta_2^*|$ 的界。若 $O^\mathsf{T} z_i \ne 0$，则在（A.26）中选取其单位方向。对于至少 cn 个连接 i 与 G 中节点的二阶负例，

$$\eta_2(\{i, j\}; \theta) \ge \beta_2 + \frac{a}{2\sqrt{r}} \|z_i\|.$$

由于二阶负例的损失为 $\log(1 + e^\eta) \ge (\eta)_+$，因此可得：

$$\widehat{L}_i(\theta) \ge c n (c' \|z_i\| - |\beta_2|)_+. \tag{A.31}$$

该不等式在 $z_i = 0$ 时也成立。由（A.29）和（A.3），$|\beta_2| \le C + \log(1/\varrho_{2,\,n})$。比较（A.30）和（A.31）可证明，该结论在整个 $\mathcal{S}_n$ 上一致成立，

$$\max_i \|z_i\| \le C b_n. \tag{A.32}$$

在此步骤之前，并未使用候选行有界这一条件。将（A.32）与相同的对齐方式结合

在（A.29）中可得：

$$\frac{1}{n}\|ZO - Z^*\|_{\mathrm{F}}^2 \le \frac{Cb_n^2}{n}\sum_i \min\{\|O^\top z_i - z_i^*\|^2, 1\} \le Cb_n^2 a_n / T_{m_\star}. \tag{A.33}$$

这与（A.29）中的截距界适用于 $\mathcal{S}_n$的每个元素。

接下来，我们证明其存在性，包括用于构造提议分布的初步估计。引理 A.1 中的无条件二阶事件，结合引理 A.4，使得上述论证可以仅针对二阶损失函数进行。只需要 $T_2 = \mu_2$，其规模条件由（A.27）得出。因此，其以真实参数为中心的真实损失次水平集在位置和截距上均有界。该集合非空且闭，因此由连续性可知，二阶损失函数存在有限的全局最小化点。这样，在调用条件高阶事件之前便已确立其存在性；在此二阶事件下，证明开头所述的有限扩展从未用到。

对于 CINE 损失，（A.29）给出了 $\mathcal{S}_n$ 中所有截距的界，而（A.32）则给出了所有位置的界。由于所有偏移量均固定且有限，对于每个 n，$\mathcal{S}_n$ 都是闭集，因而是紧集。它非空且包含真实参数。CINE 损失在该集合上取得最小值，而集合外的每个点的损失都大于真实参数处的损失。因此，该最小值也是整个中心化定义域上的全局最小值。对于任意 CINE 全局最小化点 $\widehat{\theta}$，根据定义的最优性以及 θ^* 的可行性，有：

$$\widehat{L}(\widehat{\theta}) = \inf_{\theta:\mathbf{1}^\top Z=0} \widehat{L}(\theta) \le \widehat{L}(\theta^*).$$

因此，所有此类极小化器都属于 $\mathcal{S}_n$。

现在我们用密度和采样预算来表示这些界。二阶密度条件给出 $\varrho_{2,\,n} \ge K\log\, n/n$，因此 $\log(e/\varrho_{2,\,n}) \le C\log\, n$。由于 r 固定，$a_n \le Cn\log\, n$ 且 $b_n \le C\log\, n$。令 $T_2 = \mu_2 = \binom{n}{2}\varrho_{2,\,n}$ 以及 $1/T_m = 1/\{\binom{n}{m}\varrho_{m,n}\} + 1/B_m$ 对 $m > 2$成立，这些界给出：

$$\frac{a_n}{T_m} \le C\log n\,\omega_{m,n} \quad (m \in \mathcal{M}), \qquad \frac{a_n}{T_{m_\star}} = \min_{m\in\mathcal{M}} \frac{a_n}{T_m} \le C\log n \min_{m\in\mathcal{M}} \omega_{m,n},$$

其中 $\omega_{m,\,n}$ 是定理 2 中的误差水平。将这些界代入（A.33）以及（A.29）中的截距界，并取平方根，即可在同一事件上证明（7a）和（7b）。

为保持一致性，假设对每个 $m \in \mathcal{M}$ 均有$\log n\,\omega_{m,n} = o(1)$，且$\log^3 n \min_{m\in\mathcal{M}} \omega_{m,n} = o(1)$。$\omega_{m,n}$ 的每一求和项均为非负，因此密度与采样预算阈值最终成立，并且(7a)与(7b)的右端趋于零。由于这些界以至少 $1 - C/n$的概率同时成立，对齐后的位置误差与最大绝对截距误差依概率收敛到零。在有限极小化器不存在的例外事件上，任意有限延拓给出相同的收敛结果。

□

A.6 命题 3 的证明

证明。采用第A.5节的记号，并将主文稿(3)中的完整伯努利损失记为 $\ell(\theta)$。每个候选项均被纳入，因此在每一阶上有 $R_{me} = \pi_m(e) = c_e = 1$ 。因而亲和度计算中的概率 D^* 与 $D(\theta)$即为通常的伯努利概率。在给定 Z^*的条件下，所有指示变量相互独立。由(A.3)，(A.9)中的权重满足 $p^*_m(e)\{1-p^*_m(e)\} \geq c\varrho_{m,n}$。因此，(A.6)–(A.8)中的乘积亲和度与一致界可直接应用于每一阶上带有$T_m = M_m\varrho_{m,n} = \mu_m$ 的完整伯努利损失。在移除与每个节点 i相关联的二阶候选项并保留所有高阶项之后，应用同一界。如引理A.1那样取 $t = \log n$ ，则以至少 $1 - 2/n$的概率，对所有有限的 θ 以及所有 i同时成立，

$$\ell(\theta) - \ell(\theta^*) \geq c \sum_{m \in \mathcal{M}} \mu_m \Delta_m(\theta) - Ca_n, \qquad \ell_{-i}(\theta) - \ell_{-i}(\theta^*) \geq -Ca_n,$$

其中 $a_n = (nr+1)\log n$。将此事件与引理 A.4中的事件取交集。联合概率至少为 $1 - C/n$。此处所有常数均与抽样设计无关。

考虑任意中心化的 θ ，满足 $\ell(\theta) \leq \ell(\theta^*)$。第一个不等式给出 $\Delta_m(\theta) \leq Ca_n/\mu_m$。由于阶数固定，密度条件给出 $\min_m \mu_m \geq cK n\log n$。取 K 充分大，可使所有截断误差对于引理 A.2 和 A.3而言足够小。令 $m_\star \in \operatorname{argmax}_{m\in\mathcal{M}} \mu_m$，这些引理将对齐后的截断位置均方误差界定为 $Ca_n/\mu_{m_\star}$，而引理 A.2 中关于截距的部分给出 $|\beta_m - \beta^*_m|^2 \leq 2Ca_n/\mu_m$ 对每个 $m \in \mathcal{M}$均成立。

为得到通常的位置均方误差，使用上面的第二个似然不等式，将与 i 相邻的二阶候选项的损失界定为其在真实参数处的取值加上 Ca_n。这些二阶项在完整伯努利损失和 CINE 损失中完全相同。因此，(A.30)–(A.33) 中的论证给出 $\max_i \|z_i\| \leq Cb_n$和 $n^{-1}\inf_{O^\top O = I_r} \|ZO - Z^*\|_{\mathrm{F}}^2 \leq Cb_n^2 a_n/\mu_{m_\star}$，其中 $b_n = \log(e/\varrho_{2,n}) + a_n/n$。这些行和截距界在整个中心化次水平集上成立。该集合包含真实参数，并且由连续性可知是闭集，因此是紧集。完整伯努利损失在该集合上达到其全局最小值，且每个全局最小化点都属于该集合。最后， $a_n \leq Cn\log n$、 $b_n \leq C\log n$以及 $\mu_m = \binom{n}{m}\varrho_{m,n}$。代入并开平方即可得到命题 3 中的界，这些界与定理 2 中的界相同，但取 $T_m = \mu_m$，即不含项 n/B_m。

对于命题 3 后续的比较，取完整估计和 CINE 事件的交集。根据并集界，它们的联合概率至少为 $1 - C'/n$。对于每个固定的 $m > 2$， $\mu_m = O(B_m)$ 意味着：

$$\frac{n\log n}{B_m} \leq C\frac{n\log n}{\mu_m} \leq C'\frac{\log n}{n^{m-1}\varrho_{m,n}}.$$

因此，每个 $\omega_{m,n}$ 至多是其不含该项 n/B_m对应量的常数倍，所以 CINE 和完整估计的所有界具有相同的阶。CINE 包含每条存在的超边。所包含的不同负样本数量至多等于泊松抽样次数，其均值为 B_m。

因此，阶为 m 的损失函数项的期望数量至多为 $\mu_m + B_m$。对于 $B_m \asymp \mu_m$，除以 $\binom{n}{m}$即可得到所述的 $O(\varrho_{m,\,n})$ 期望评估比例界。

□

阶数筛选后的误差界。 假设考虑的阶数属于一个包含 2 的固定有限集合$\mathcal{D}$，且筛选规则使用负采样前的观测。对于每个包含 2 且满足相关密度和采样预算条件的子集$S \subseteq \mathcal{D}$，使用其自身的h_m、固定采样预算以及第2节的采样规则。给定观测后，可以预先为每个此类设计生成抽样随机化。随后，选定设计的拟合结果与筛选后再运行该设计具有相同的分布。对于每个确定性的 S，定理 2 和命题 3 分别给出的误差界，其失效概率至多为 C/n。对至多 $2^{|\mathcal{D}|-1}$ 个设计应用并集界，可得一个概率至少为 $1 - C'/n$ 的事件，在该事件上所有符合条件的设计均满足相应误差界。因此，只要选定集合包含 2 且满足相应条件，这些误差界就在同一事件上对其成立。这是在原始观测律下的一个同时性结论。上述精确的总体恒等式，对每个固定设计均使用该观测律。

A.7 条件信息与提议分布比较

对于严格正的提议分布 q，将条件信息矩阵记为 $J_m(q)$，其中 $\pi_m(e) = 1 - \exp\{-B_m q(e)\}$。CINE 提议分布给出 $J_m = J_m(q_m)$。以下结果验证其解释，并对两种采样提议分布进行精确比较。求导时，固定中心化位置和各阶截距的坐标，并将提议分布视为固定。

命题 A.1 (条件信息恒等式)。 在假设 *1* 下，对于每个目标阶数 $m > 2$,

$$\mathrm{Cov}\{\nabla \widehat{\ell}_m(Z^*, \beta_m^*) \mid \widetilde{Z}\} = \mathbb{E}\{\nabla^2 \widehat{\ell}_m(Z^*, \beta_m^*) \mid \widetilde{Z}\} = J_m(q_m). \tag{A.34}$$

在真实参数处，条件总体得分为零。对于具有相同采样预算的两个严格正的提议分布 q_1、q_0，设 $\pi_j(e) = 1 - \exp\{-B_m q_j(e)\}$ 和 $c_j(e) = p_m(e) + \{1 - p_m(e)\}\pi_j(e)$。则，

$$\begin{aligned} J_m(q_1) - J_m(q_0) = \sum_{e \in \Omega_m} & \frac{p_m(e)^2\{1 - p_m(e)\}\{\pi_1(e) - \pi_0(e)\}}{c_1(e)c_0(e)} \\ & \times \nabla\eta_m(e)\nabla\eta_m(e)^{\mathsf{T}}. \end{aligned} \tag{A.35}$$

证明。 以初步估计为条件，并省略阶数和候选项下标。令 $p = p_m(e)$、$\pi = \pi_m(e)$、$A = A_e^{(m)}$、$R = R_{me}$ 和 $D = p/\{p + (1-p)\pi\}$。真实参数处的候选梯度和海森矩阵为:

$$R(D - A)\nabla\eta, \qquad RD(1 - D)\nabla\eta\nabla\eta^{\mathsf{T}} + R(D - A)\nabla^2\eta.$$

根据（A.1），

$$\mathbb{E}\{R(D - A) \mid \widetilde{Z}\} = p(D - 1) + (1 - p)\pi D = 0.$$

因此，条件得分的均值为零。由于潜在对数优势是非线性的，第二个海森项也存在，且其期望同样为零。此外，

$$\begin{aligned}\mathbb{E}\{R(D-A)^2 \mid \widetilde{Z}\} &= p(1-D)^2 + (1-p)\pi D^2 \\ &= [p+(1-p)\pi]D(1-D) = \frac{p(1-p)\pi}{p+(1-p)\pi}.\end{aligned}$$

同一系数是 $RD(1-D)$ 的条件期望。结合假设 1，命题 2 证明中的泊松分割论证表明，各候选项之间条件独立。对它们的协方差和期望海森矩阵求和，即可证明 (A.34)。

最后，

$$\frac{\pi_1(e)}{c_1(e)} - \frac{\pi_0(e)}{c_0(e)} = \frac{p_m(e)\{\pi_1(e)-\pi_0(e)\}}{c_1(e)c_0(e)}.$$

将每个候选项对全数据的信息的贡献乘入并求和，即可证明 (A.35)。重新分配固定采样预算可能会提高某些纳入概率，同时降低另一些纳入概率。因此，这一差值中的各项可能为正，也可能为负。仅仅保证全数据信息占比更高，并不能确立不同提议分布之间的矩阵序关系。

□

A.8 跨阶提议的覆盖范围

接下来的结果将基于 h_m 的补全规则与推论1中的提议条件联系起来。固定$m>2$，设定 $h=h_m$，并令 Δ_m^{prop}为 $e\in\Omega_m$ 上的取值范围：

$$d_m(e) = \log\left\{\binom{m}{h}^{-1}\sum_{\substack{u\subset e\\|u|=h}}\exp\{\tau_m s_h(u;\widetilde{Z})\}\right\} - s_m(e;Z^*). \tag{A.36}$$

因此， $\Delta_m^{\mathrm{prop}} = \max_e d_m(e) - \min_e d_m(e)$ 衡量补全后的权重偏离目标阶得分模式的程度。

命题 A.2 （提议覆盖）。 在假设 *1* 下，对于每个目标阶数，几乎必然有，

$$q_m(e) \geq a_m\frac{p_m(e)}{\mu_m} \quad (e\in\Omega_m), \qquad a_m = \epsilon\frac{\mu_m}{M_m p_{m,\max}} + (1-\epsilon)e^{-\Delta_m^{\mathrm{prop}}}(1-p_{m,\max}) \in (0,1]. \tag{A.37}$$

特别地，

$$\frac{a_m B_m}{\mu_m + a_m B_m}F_m \preceq J_m \preceq F_m. \tag{A.38}$$

如果真实行范数至多为 L，则令 $S_m = L^2\sqrt{\binom{m}{2}/r}$。于是，无论 $\widetilde{Z}$的准确性如何， $a_m \geq \epsilon e^{-2S_m}$。更具体地说，令：

$$\delta_m = \max_{i<j}|\widetilde{z}_i^{\mathsf{T}}\widetilde{z}_j - (z_i^*)^{\mathsf{T}}z_j^*|, \qquad D_m = \max_{\substack{e\in\Omega_m,\ u\subset e\\|u|=h_m}}\{\tau_m s_{h_m}(u;Z^*) - s_m(e;Z^*)\}.$$

于是：

$$0 \le D_m \le 2S_m, \qquad \Delta_m^{\mathrm{prop}} \le D_m + 2\sqrt{\binom{m}{2}/r}\,\delta_m. \tag{A.39}$$

第一个界给出了提案特定信息常数。当真实得分有界且混合权重远离零时，仅均匀成分就能提供一个常数下界。更准确的初步估计可以通过 Δ_m^{prop}改善覆盖界，尽管它未必能对两个估计量的实际风险排序。注意，即使初步估计完美，候选项内异质性项D_m 仍然存在。

证明。令 $w(u) = \exp\{\tau_m s_h(u;\, \widetilde{Z})\}$，且 $W(e) = \binom{m}{h}^{-1} \sum_{u\subset e,\ |u|=h} w(u)$。每个阶为-$h$的子集都有 $\binom{n-h}{m-h}$ 种补全方式，因此：

$$\sum_{e\in\Omega_m} W(e) = \frac{\binom{n-h}{m-h}}{\binom{m}{h}} \sum_{u\in\Omega_h} w(u).$$

因此，提议分布的引导成分等于 $W(e)/\sum_f W(f)$。由（A.36），

$$\frac{W(e)}{\sum_f W(f)} \ge e^{-\Delta_m^{\mathrm{prop}}} \frac{e^{s_m(e;Z^*)}}{\sum_f e^{s_m(f;Z^*)}}.$$

令 $v_e = \exp\{\beta_m^* + s_m(e;\, Z^*)\} = p_m(e)/\{1-p_m(e)\}$。由于$\sum_f v_f \le \mu_m/(1-p_{m,\max})$ 且 $v_e \ge p_m(e)$，最后一个比率至少为 $(1-p_{m,\max})p_m(e)/\mu_m$。均匀成分满足 $1/M_m \ge p_m(e)/(M_m p_{m,\max})$。结合这两个成分即可证明（A.37）。 a_m 的两个成分均非负，且它们的和为正，并且 $a_m \le 1$。推论1给出（A.38），该结论对初步估计逐点成立。

统计包含每个节点对的 h 子集可得：

$$\binom{m}{h}^{-1} \sum_{\substack{u\subset e\\ |u|=h}} \tau_m s_h(u; Z^*) = s_m(e; Z^*). \tag{A.40}$$

事实上，一个节点对属于 $\binom{m-2}{h-2}$ 个子集，并且 $\binom{m-2}{h-2}/\binom{m}{h} = \binom{h}{2}/\binom{m}{2}$。因此， $\tau_m s_h(u;\, Z^*) - s_m(e; Z^*)$ 的值在每个候选项内的平均值为零。它们的最大值 D_m 为非负数，并且由詹森不等式可得：

$$0 \le \log\left\{ \binom{m}{h}^{-1} \sum_{u\subset e,\ |u|=h} e^{\tau_m s_h(u;Z^*) - s_m(e;Z^*)} \right\} \le D_m.$$

将 Z^* 替换为 $\widetilde{Z}$到低阶分数中后，每个指数的变化幅度至多为 $\sqrt{\binom{m}{2}/r}\,\delta_m$。其总和的对数变化幅度也至多为相同数值。因此，其范围至多增加该数值的两倍，从而证明了（A.39）中的第二个界。

真实行界使 $s_m(e; Z^*)$ 和 $\tau_m s_h(u; Z^*)$ 的绝对值均不超过 S_m，因此$D_m \le 2S_m$。对于得分 $s_1 \le s_2$，逻辑斯蒂模型给出：

$$\frac{p(s_1)}{p(s_2)} = e^{s_1 - s_2} \frac{1 + e^{\beta_m^* + s_2}}{1 + e^{\beta_m^* + s_1}} \ge e^{-2S_m}.$$

由此可得 $\mu_m/(M_m p_{m,\max}) \geq e^{-2S_m}$，从而证明了关于 a_m 的所述下界。 □

B 实现细节与补充模拟结果

B.1 数值实现

我们使用带有解析梯度的 L-BFGS-B 实现主文稿中的算法 1，并通过约化坐标参数化来施加中心化约束。令 $\mathbf{1}_n$ 为由 n 个 1 组成的向量，$\boldsymbol{I}_{n-1}$为 $(n-1)\times(n-1)$ 单位矩阵，$\boldsymbol{e}_i$ 为 i中的第 $\mathbb{R}^n$个标准坐标向量。令 $\boldsymbol{Q} \in \mathbb{R}^{n\times(n-1)}$ 的列为一组标准正交基，张成与 $\mathbf{1}_n$正交的补空间。我们通过对 $[\boldsymbol{e}_1 - \boldsymbol{e}_n, \ldots, \boldsymbol{e}_{n-1} - \boldsymbol{e}_n]$ 进行 QR 分解得到 $\boldsymbol{Q}$，并针对约化坐标矩阵 $\boldsymbol{U} = (U_{ij}) \in \mathbb{R}^{(n-1)\times r}$进行优化，其中 $\boldsymbol{Z} = \boldsymbol{Q}\boldsymbol{U}$、$\boldsymbol{Q}^\top\boldsymbol{Q} = \boldsymbol{I}_{n-1}$和 $\boldsymbol{Q}^\top\mathbf{1}_n = \mathbf{0}$。该参数化在每次迭代时均满足中心化约束$\mathbf{1}_n^\top \boldsymbol{Z} = \mathbf{0}^\top$。$\boldsymbol{U}$的每个元素均属于 $[-6, 6]$，且每个截距均位于一个有界区间内。

对于初步估计，我们从三个矩阵开始。第一个矩阵由二阶存在超边的邻接矩阵构造而成，并从每个非对角元素中减去二阶存在超边的密度。我们取 r 个最大特征值对应的特征向量，并将每个特征向量乘以其特征值的平方根；若所选特征值小于 0。1，则将其替换为 0。1。其余两个初始矩阵的元素均为相互独立的标准正态分布元素。对每个初始矩阵按列进行中心化和标准化，并将二阶截距初始化为 logit函数{(|E_2^+|+0.5)/(|Ω_2|+1)}。我们从每个初始值出发最小化二阶损失，并将目标函数值最小的拟合结果作为 ($\widetilde{\boldsymbol{Z}}$, $\widetilde{\beta}_2$)。CINE 的联合优化从 $\widetilde{\boldsymbol{Z}}$ 开始，每个保留阶数 m 的截距初始化为 logit函数{(|E_m^+|+0.5)/(|Ω_m|+1)}。对于 CINE++，之后每次联合优化都以所有较低保留阶数下估计的潜在位置作为初始值，截距也按相同方式初始化。

为停止优化，令 f 表示 CINE 损失除以一个固定的归一化常数，即在模拟中纳入当前拟合的各阶之和上的 $\sum_m \mu_m$，其中 μ_m 由设计已知，而在真实数据分析中则为观测计数 $\sum_m |E_m^+|$。对于初步估计，二阶损失除以 $|E_2^+|$。令 $\boldsymbol{v}$ 收集 $\boldsymbol{U}$ 的各个条目以及所纳入各阶的截距，并记第 k 次迭代时 $f^{(k)} = f(\boldsymbol{v}^{(k)})$。令 Π 表示上述边界上的投影，则投影梯度为 $\boldsymbol{g}^{(k)}{}_{\text{proj}} = \boldsymbol{v}^{(k)} - \Pi\{\boldsymbol{v}^{(k)} - \nabla_{\boldsymbol{v}} f(\boldsymbol{v}^{(k)})\}$。当 $\|\boldsymbol{g}^{(k)}{}_{\text{proj}}\|_\infty \leq \text{tol}_\text{g}$、$\{f^{(k)} - f^{(k+1)}\}/\max\{1, |f^{(k)}|, |f^{(k+1)}|\} \leq \text{tol}_\text{f}$，或达到 $k_{\max}$ 次迭代后，我们停止优化。我们使用 $\text{tol}_\text{g} = 3 \times 10^{-6}$ 和 $\text{tol}_\text{f} = 10^{-10}$，其中在模拟中取 $k_{\max} = 260$，在真实数据分析中取 $k_{\max} = 10^4$。

B.2 数据生成与评估加速

对于较大的候选集，我们使用高斯辅助构造来加速数据生成。具体而言，我们通过高斯–埃尔米特求积和针对子集大小的动态规划，近似计算截距校准所需的指数得分和。随后，通过泊松抽样生成存在的超边

并进行拒绝校正；事件稀疏化在去重之前执行，以在数值精度范围内保留独立伯努利抽样机制。该过程避免了在生成过程中枚举所有候选超边。我们在示例1中将此实现用于四阶至七阶，参数为 $(n,\ m_{\max}) = (300,\ 7)$；在示例 2 和 3 中则用于五阶至七阶。

为加快相对均方误差的评估，我们采用并行分块求和，在完整候选集上累加平方误差和平方真实值和。在每个分块内同时评估真实参数和拟合得分，以减少候选项的重复枚举和内存分配。

对于示例1中的 $(n,\ m_{\max}) = (300,\ 7)$，二阶和三阶在其完整候选集上进行评估，而四阶至七阶则对每一阶使用 10^7 个独立重要性样本。评估分布将均匀成分与由真实潜在位置和可用的拟合潜在位置构造的高斯/子集混合分布相结合。使用基于归一化评估概率的重要性权重来估计相对均方误差。所有方法共享相同的评估样本和权重，以确保可比性。蒙特卡洛标准误与重复试验间变异性分别记录。

B.3 附加结果

表 B.1– B.3 报告示例1中和 $m_{\max} \in \{3, 5, 7$的相对均方误差 $\}$.

表 B.1：示例1中最大超边阶数为 $m_{\max} = 3$ 时的相对均方误差和计算时间。表中数值为 100 次重复实验的均值（标准差）。在 FULL 以外的方法中，每个图版各列的平均相对均方误差最低值以粗体显示，次低值以下划线标示。

面板 A：$n = 100$。

Method	相对均方误差				Time (秒)
	$\boldsymbol{Z}$	总体	二阶	三阶	
FULL	*0.030*	*0.073*	*0.058*	*0.170*	*0.227*
	(0.005)	*(0.009)*	*(0.007)*	*(0.035)*	*(0.042)*
Cine++	0.042	0.113	0.085	0.295	0.211
	(0.007)	(0.016)	(0.011)	(0.068)	(0.012)
Cine	0.042	0.113	0.085	0.295	0.300
	(0.007)	(0.016)	(0.011)	(0.068)	(0.026)
Uniform	0.049	0.194	0.105	0.755	0.395
	(0.007)	(0.047)	(0.013)	(0.265)	(0.067)
Truncating	**0.030**	**0.073**	**0.058**	**0.170**	0.259
	(0.005)	(0.009)	(0.007)	(0.035)	(0.023)
Projecting	**0.030**	**0.073**	**0.058**	**0.170**	0.286
	(0.005)	(0.009)	(0.007)	(0.035)	(0.028)
JLE	0.879	3.652	1.484	17.691	2.293
	(0.182)	(0.939)	(0.521)	(6.306)	(0.727)
HEM	1.452	0.944	0.957	0.863	1.185
	(0.456)	(0.009)	(0.005)	(0.040)	(0.042)
HGE	0.076	0.317	0.166	1.233	2.233
	(0.026)	(0.110)	(0.041)	(0.497)	(0.372)

面板 B：$n = 300$。

Method	相对均方误差				Time (s)
	$\boldsymbol{Z}$	总体	二阶	三阶	
FULL	*0.027*	*0.085*	*0.081*	*0.189*	*9.495*
	(0.002)	*(0.006)*	*(0.006)*	*(0.021)*	*(1.099)*
Cine++	0.035	0.119	0.112	0.291	0.658
	(0.003)	(0.008)	(0.008)	(0.038)	(0.028)
Cine	0.035	0.119	0.112	0.291	1.465
	(0.003)	(0.008)	(0.008)	(0.038)	(0.135)
Uniform	0.042	0.179	0.153	0.769	2.327
	(0.003)	(0.016)	(0.012)	(0.156)	(0.383)
Truncating	**0.027**	**0.085**	**0.081**	**0.189**	9.047
	(0.002)	(0.006)	(0.006)	(0.021)	(1.074)
Projecting	**0.027**	**0.085**	**0.081**	**0.189**	9.080
	(0.002)	(0.006)	(0.006)	(0.021)	(1.073)
JLE	0.793	1.502	1.188	8.531	35.878
	(0.089)	(0.973)	(0.773)	(5.767)	(12.584)
HEM	1.649	0.982	0.987	0.862	16.416
	(0.452)	(0.002)	(0.001)	(0.029)	(0.209)
HGE	0.076	0.279	0.232	1.306	12.317
	(0.070)	(0.124)	(0.086)	(1.085)	(0.505)

表 B.2：最大超边阶数为 $m_{\max} = 5$ 时示例1中的相对均方误差和计算时间。数值为 100 次重复的均值（标准差）。在除 FULL 之外的方法中，每个图版的每列中均值相对均方误差最低的以粗体显示，第二低的加下划线。

面板 A： $n = 100$。

Method	相对均方误差						Time （秒）
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	
FULL	*0.013* *(0.002)*	*0.032* *(0.003)*	*0.025* *(0.003)*	*0.065* *(0.012)*	*0.093* *(0.024)*	*0.115* *(0.038)*	*187.581* *(57.363)*
Cine++	**0.020** (0.003)	**0.050** (0.006)	**0.039** (0.004)	**0.111** (0.026)	**0.165** (0.051)	**0.200** (0.068)	0.416 (0.033)
Cine	0.022 (0.004)	0.059 (0.010)	0.045 (0.006)	0.133 (0.038)	0.210 (0.102)	0.264 (0.155)	0.780 (0.269)
Uniform	0.029 (0.004)	0.129 (0.040)	0.065 (0.008)	0.380 (0.121)	1.129 (0.706)	2.224 (1.947)	0.545 (0.190)
Truncating	0.030 (0.004)	0.076 (0.008)	0.057 (0.006)	0.170 (0.036)	0.273 (0.074)	0.364 (0.125)	0.472 (0.137)
Projecting	0.023 (0.004)	0.116 (0.073)	0.047 (0.007)	0.320 (0.150)	1.236 (0.978)	2.845 (3.034)	0.459 (0.133)
JLE	0.442 (0.155)	1.348 (2.127)	0.569 (0.434)	2.853 (4.922)	18.901 (49.314)	128.818 (364.153)	1493.131 (361.728)
HEM	0.621 (0.215)	1.000 (0.000)	1.000 (0.000)	0.999 (0.000)	0.991 (0.003)	0.990 (0.014)	801.447 (143.901)
HGE	0.058 (0.025)	0.298 (0.123)	0.121 (0.025)	0.937 (0.365)	3.582 (2.880)	8.700 (10.412)	6.658 (2.732)

面板 B： $n = 300$。

Method	相对均方误差						Time (s)
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	
FULL†	-	-	-	-	-	-	-
Cine++	**0.016** (0.001)	**0.054** (0.004)	0.051 (0.004)	**0.117** (0.014)	**0.170** (0.027)	**0.213** (0.041)	13.091 (5.255)
Cine	0.017 (0.001)	0.058 (0.004)	0.054 (0.004)	0.127 (0.015)	0.184 (0.030)	0.234 (0.041)	11.292 (5.201)
Uniform	0.025 (0.002)	0.114 (0.013)	0.099 (0.009)	0.415 (0.117)	0.889 (0.357)	1.462 (0.718)	2.904 (0.357)
Truncating	0.027 (0.002)	0.085 (0.007)	0.080 (0.006)	0.191 (0.023)	0.282 (0.045)	0.373 (0.071)	8.835 (1.735)
Projecting	0.018 (0.001)	**0.054** (0.006)	**0.050** (0.004)	0.127 (0.039)	0.205 (0.118)	0.289 (0.233)	6.332 (1.151)
JLE†	-	-	-	-	-	-	-
HEM†	-	-	-	-	-	-	-
HGE	0.042 (0.023)	0.195 (0.088)	0.158 (0.050)	0.881 (0.750)	2.768 (4.051)	5.559 (9.986)	65.894 (27.516)

† 面板 B 省略了 FULL、JLE 和 HEM，因为每一项进行 100 次重复都需要超过 30 天。

表 B.3：示例1中最大超边阶数为 $m_{\max} = 7$ 时的相对均方误差和计算时间。表中数值为 100 次重复实验的均值（标准差）。短横线表示结果不可用。除 FULL 外的各方法中，每个图版各列的相对均方误差均值最低值以粗体显示，次低值以下划线标出。

面板 A： $n = 100$。

Method	相对均方误差								Time (秒)
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	六阶	七阶	
FULL†	-	-	-	-	-	-	-	-	-
Cine++	**0.013** (0.002)	**0.033** (0.004)	**0.026** (0.003)	**0.072** (0.016)	**0.100** (0.028)	**0.125** (0.041)	**0.139** (0.046)	**0.157** (0.063)	75.724 (10.788)
Cine	0.017 (0.004)	0.043 (0.008)	0.033 (0.006)	0.096 (0.021)	0.139 (0.038)	0.177 (0.081)	0.204 (0.084)	0.222 (0.111)	88.339 (23.346)
Uniform	0.025 (0.003)	0.104 (0.031)	0.055 (0.007)	0.310 (0.087)	0.904 (0.634)	1.640 (1.598)	2.879 (4.190)	4.711 (10.141)	0.572 (0.123)
Truncating	0.030 (0.004)	0.077 (0.007)	0.059 (0.006)	0.177 (0.030)	0.292 (0.074)	0.388 (0.122)	0.475 (0.172)	0.555 (0.225)	0.402 (0.090)
Projecting	0.025 (0.006)	0.219 (0.143)	0.063 (0.014)	0.648 (0.296)	3.292 (2.483)	9.227 (8.851)	18.422 (20.215)	30.391 (36.674)	0.427 (0.096)
JLE†	-	-	-	-	-	-	-	-	-
HEM†	-	-	-	-	-	-	-	-	-
HGE	0.049 (0.037)	0.199 (0.078)	0.103 (0.041)	0.632 (0.237)	1.806 (1.402)	3.648 (4.780)	5.822 (10.134)	8.137 (15.215)	8.304 (3.109)

面板 B： $n = 300$。

Method	相对均方误差								Time (s)
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	六阶	七阶	
FULL	-	-	-	-	-	-	-	-	-
Cine++	**0.014** (0.001)	**0.044** (0.003)	**0.042** (0.003)	**0.096** (0.012)	**0.137** (0.021)	**0.171** (0.030)	**0.200** (0.035)	**0.244** (0.056)	458.161 (75.637)
Cine	0.015 (0.001)	0.049 (0.004)	0.047 (0.004)	0.107 (0.014)	0.154 (0.023)	0.192 (0.033)	0.224 (0.043)	0.279 (0.071)	350.313 (86.265)
Uniform	0.024 (0.001)	0.104 (0.009)	0.091 (0.007)	0.372 (0.081)	0.768 (0.264)	1.290 (0.589)	0.904 (0.357)	2.902 (1.805)	1.433 (0.373)
Truncating	0.027 (0.002)	0.084 (0.006)	0.079 (0.006)	0.188 (0.022)	0.278 (0.045)	0.367 (0.076)	0.460 (0.116)	0.558 (0.167)	7.772 (1.722)
Projecting	0.016 (0.002)	0.065 (0.025)	0.052 (0.012)	0.294 (0.224)	0.853 (0.969)	1.650 (2.091)	2.672 (3.542)	3.915 (5.339)	6.831 (1.463)
JLE	-	-	-	-	-	-	-	-	-
HEM	-	-	-	-	-	-	-	-	-
HGE	0.040 (0.026)	0.176 (0.075)	0.143 (0.050)	0.781 (0.567)	2.224 (2.659)	3.818 (5.222)	5.785 (8.529)	7.839 (12.092)	116.437 (36.695)

† 面板 A 未列出 FULL、JLE 和 HEM，因为每种方法完成 100 次重复都需要超过 30 天。

表格B.4–B.7 报告了示例 2 中关于 $n = 100$ 和 $m_{\max} = 7$的结果。

表 B.4：示例 2 中 $n = 100$、$m_{\max} = 7$ 和 $c = 1.4$ 的相对均方误差和计算时间。数值为 100 次重复的均值（标准差）。在每种设置下，最低的两个平均相对均方误差分别以粗体和下划线标出。

面板 A：$B_m/\mu_m = 4$。

Method	相对均方误差								Time (秒)
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	六阶	七阶	
FULL†	-	-	-	-	-	-	-	-	-
Cine++	**0.031** (0.004)	**0.040** (0.005)	**0.035** (0.004)	**0.102** (0.018)	**0.147** (0.029)	**0.195** (0.045)	**0.221** (0.049)	**0.250** (0.063)	69.638 (11.854)
Cine	0.038 (0.006)	0.048 (0.007)	0.042 (0.006)	0.126 (0.025)	0.189 (0.051)	0.254 (0.080)	0.287 (0.086)	0.335 (0.117)	79.771 (32.851)
Uniform	0.046 (0.005)	0.070 (0.011)	0.056 (0.007)	0.242 (0.070)	0.456 (0.209)	0.703 (0.409)	0.897 (0.600)	1.170 (0.990)	0.525 (0.140)
Truncating	0.075 (0.009)	0.096 (0.009)	0.083 (0.008)	0.259 (0.035)	0.410 (0.073)	0.547 (0.122)	0.686 (0.185)	0.830 (0.266)	0.398 (0.090)
Projecting	0.048 (0.006)	0.053 (0.006)	0.047 (0.005)	0.129 (0.027)	0.194 (0.074)	0.247 (0.122)	0.293 (0.162)	0.334 (0.194)	0.400 (0.080)
JLE	-	-	-	-	-	-	-	-	-
HEM	-	-	-	-	-	-	-	-	-
HGE	0.066 (0.009)	0.134 (0.022)	0.096 (0.014)	0.592 (0.146)	1.327 (0.530)	2.155 (1.043)	2.760 (1.330)	3.532 (1.965)	3.301 (0.116)

面板 B：$B_m/\mu_m = 6$.

Method	相对均方误差								Time (s)
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	六阶	七阶	
FULL†	-	-	-	-	-	-	-	-	-
Cine++	**0.029** (0.004)	**0.037** (0.004)	**0.032** (0.004)	**0.092** (0.014)	**0.132** (0.027)	**0.174** (0.041)	**0.198** (0.045)	**0.222** (0.054)	94.763 (17.330)
Cine	0.034 (0.005)	0.043 (0.006)	0.038 (0.005)	0.111 (0.020)	0.165 (0.042)	0.221 (0.067)	0.252 (0.070)	0.286 (0.085)	110.428 (41.317)
Uniform	0.041 (0.005)	0.060 (0.008)	0.049 (0.006)	0.191 (0.050)	0.346 (0.145)	0.523 (0.273)	0.647 (0.371)	0.838 (0.604)	0.697 (0.188)
Truncating	0.075 (0.009)	0.096 (0.009)	0.083 (0.008)	0.259 (0.035)	0.410 (0.073)	0.547 (0.122)	0.686 (0.185)	0.830 (0.266)	0.398 (0.090)
Projecting	0.048 (0.006)	0.053 (0.006)	0.047 (0.005)	0.129 (0.027)	0.194 (0.074)	0.247 (0.122)	0.293 (0.162)	0.334 (0.194)	0.400 (0.080)
JLE	-	-	-	-	-	-	-	-	-
HEM	-	-	-	-	-	-	-	-	-
HGE	0.066 (0.009)	0.134 (0.022)	0.096 (0.014)	0.592 (0.146)	1.327 (0.530)	2.155 (1.043)	2.760 (1.330)	3.532 (1.965)	3.301 (0.116)

面板 C：$B_m/\mu_m = 8$。

Method	相对均方误差								Time (s)
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	六阶	七阶	
FULL†	-	-	-	-	-	-	-	-	-
Cine++	**0.028** (0.004)	**0.035** (0.004)	**0.031** (0.004)	**0.088** (0.015)	**0.126** (0.027)	**0.167** (0.042)	**0.188** (0.042)	**0.214** (0.055)	119.906 (20.086)
Cine	0.032 (0.005)	0.040 (0.005)	0.035 (0.005)	0.103 (0.018)	0.152 (0.035)	0.203 (0.059)	0.232 (0.062)	0.262 (0.072)	137.141 (49.030)
Uniform	0.037 (0.004)	0.054 (0.007)	0.045 (0.005)	0.163 (0.038)	0.283 (0.103)	0.422 (0.193)	0.509 (0.239)	0.642 (0.343)	0.867 (0.236)
Truncating	0.075 (0.009)	0.096 (0.009)	0.083 (0.008)	0.259 (0.035)	0.410 (0.073)	0.547 (0.122)	0.686 (0.185)	0.830 (0.266)	0.398 (0.090)
Projecting	0.048 (0.006)	0.053 (0.006)	0.047 (0.005)	0.129 (0.027)	0.194 (0.074)	0.247 (0.122)	0.293 (0.162)	0.334 (0.194)	0.400 (0.080)
JLE	-	-	-	-	-	-	-	-	-
HEM	-	-	-	-	-	-	-	-	-
HGE	0.066 (0.009)	0.134 (0.022)	0.096 (0.014)	0.592 (0.146)	1.327 (0.530)	2.155 (1.043)	2.760 (1.330)	3.532 (1.965)	3.301 (0.116)

† 省略 FULL、JLE 和 HEM，因为对 100 次重复而言，每种方法都需要超过 30 天。

表 B.5：示例 2 中 $n = 100$、$m_{\max} = 7$ 和 $c = 1.6$ 的相对均方误差和计算时间。数值为 100 次重复中的均值（标准差）。在每种设置下，均值相对均方误差最低的两项分别以粗体和下划线显示。

面板 A：$B_m/\mu_m = 4_\circ$

Method	相对均方误差								Time (秒)
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	六阶	七阶	
FULL	-	-	-	-	-	-	-	-	-
Cine++	**0.018** (0.003)	**0.034** (0.004)	**0.028** (0.003)	**0.084** (0.016)	**0.124** (0.025)	**0.160** (0.041)	**0.186** (0.047)	**0.200** (0.056)	62.084 (10.432)
Cine	0.023 (0.003)	0.042 (0.006)	0.035 (0.005)	0.106 (0.020)	0.157 (0.035)	0.201 (0.057)	0.235 (0.071)	0.256 (0.083)	64.192 (20.539)
Uniform	0.031 (0.004)	0.079 (0.014)	0.054 (0.006)	0.277 (0.080)	0.640 (0.392)	1.037 (0.774)	1.551 (1.497)	1.986 (2.042)	0.496 (0.127)
Truncating	0.045 (0.005)	0.084 (0.008)	0.068 (0.007)	0.216 (0.029)	0.350 (0.066)	0.463 (0.111)	0.571 (0.162)	0.676 (0.224)	0.348 (0.068)
Projecting	0.029 (0.004)	0.076 (0.032)	0.047 (0.007)	0.270 (0.138)	0.845 (0.813)	1.736 (2.204)	2.925 (4.348)	4.396 (7.182)	0.365 (0.075)
JLE	-	-	-	-	-	-	-	-	-
HEM	-	-	-	-	-	-	-	-	-
HGE	0.046 (0.008)	0.145 (0.030)	0.091 (0.017)	0.570 (0.149)	1.388 (0.704)	2.300 (1.614)	3.090 (2.429)	3.744 (2.832)	4.438 (0.651)

面板 B：$B_m/\mu_m = 6_\circ$

Method	相对均方误差								Time (s)
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	六阶	七阶	
FULL	-	-	-	-	-	-	-	-	-
Cine++	**0.017** (0.002)	**0.031** (0.004)	**0.026** (0.003)	**0.075** (0.015)	**0.110** (0.022)	**0.142** (0.036)	**0.166** (0.049)	**0.178** (0.051)	84.045 (13.522)
Cine	0.020 (0.003)	0.038 (0.005)	0.031 (0.004)	0.093 (0.018)	0.137 (0.030)	0.175 (0.049)	0.204 (0.059)	0.222 (0.067)	93.636 (33.285)
Uniform	0.027 (0.003)	0.065 (0.010)	0.046 (0.005)	0.211 (0.048)	0.453 (0.226)	0.707 (0.433)	1.004 (0.686)	1.296 (1.132)	0.665 (0.176)
Truncating	0.045 (0.005)	0.084 (0.008)	0.068 (0.007)	0.216 (0.029)	0.350 (0.066)	0.463 (0.111)	0.571 (0.162)	0.676 (0.224)	0.348 (0.068)
Projecting	0.029 (0.004)	0.076 (0.032)	0.047 (0.007)	0.270 (0.138)	0.845 (0.813)	1.736 (2.204)	2.925 (4.348)	4.396 (7.182)	0.365 (0.075)
JLE	-	-	-	-	-	-	-	-	-
HEM	-	-	-	-	-	-	-	-	-
HGE	0.046 (0.008)	0.145 (0.030)	0.091 (0.017)	0.570 (0.149)	1.388 (0.704)	2.300 (1.614)	3.090 (2.429)	3.744 (2.832)	4.438 (0.651)

面板 C：$B_m/\mu_m = 8_\circ$

Method	相对均方误差								Time (s)
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	六阶	七阶	
FULL	-	-	-	-	-	-	-	-	-
Cine++	**0.016** (0.002)	**0.029** (0.003)	**0.024** (0.003)	**0.070** (0.013)	**0.102** (0.021)	**0.131** (0.033)	**0.151** (0.037)	**0.165** (0.048)	104.561 (16.777)
Cine	0.019 (0.003)	0.035 (0.005)	0.029 (0.004)	0.085 (0.017)	0.125 (0.027)	0.160 (0.042)	0.185 (0.049)	0.203 (0.059)	109.699 (30.810)
Uniform	0.025 (0.003)	0.057 (0.008)	0.042 (0.005)	0.175 (0.039)	0.351 (0.153)	0.538 (0.304)	0.739 (0.439)	0.917 (0.655)	0.811 (0.213)
Truncating	0.045 (0.005)	0.084 (0.008)	0.068 (0.007)	0.216 (0.029)	0.350 (0.066)	0.463 (0.111)	0.571 (0.162)	0.676 (0.224)	0.348 (0.068)
Projecting	0.029 (0.004)	0.076 (0.032)	0.047 (0.007)	0.270 (0.138)	0.845 (0.813)	1.736 (2.204)	2.925 (4.348)	4.396 (7.182)	0.365 (0.075)
JLE	-	-	-	-	-	-	-	-	-
HEM	-	-	-	-	-	-	-	-	-
HGE	0.046 (0.008)	0.145 (0.030)	0.091 (0.017)	0.570 (0.149)	1.388 (0.704)	2.300 (1.614)	3.090 (2.429)	3.744 (2.832)	4.438 (0.651)

表 B.6：示例 2 中 $n = 100$、$m_{\max} = 7$ 和 $c = 1.8$ 的相对均方误差和计算时间。数值为 100 次重复的均值（标准差）。在每种设置下，均值相对均方误差最低的两项分别以粗体和下划线标示。

面板 A：$B_m/\mu_m = 4$。

Method	相对均方误差								Time (秒)
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	六阶	七阶	
FULL	-	-	-	-	-	-	-	-	-
Cine++	**0.013** (0.002)	**0.033** (0.004)	**0.026** (0.004)	**0.072** (0.015)	**0.100** (0.029)	**0.125** (0.043)	**0.139** (0.051)	**0.157** (0.062)	75.724 (17.360)
Cine	0.017 (0.004)	0.043 (0.008)	0.033 (0.006)	0.096 (0.021)	0.139 (0.038)	0.177 (0.081)	0.204 (0.084)	0.222 (0.111)	88.339 (23.346)
Uniform	0.025 (0.003)	0.104 (0.031)	0.055 (0.007)	0.310 (0.087)	0.904 (0.634)	1.640 (1.598)	2.879 (4.190)	4.711 (10.141)	0.572 (0.123)
Truncating	0.030 (0.004)	0.077 (0.007)	0.059 (0.006)	0.177 (0.030)	0.292 (0.074)	0.388 (0.122)	0.475 (0.172)	0.555 (0.225)	0.402 (0.090)
Projecting	0.025 (0.006)	0.219 (0.143)	0.063 (0.014)	0.648 (0.296)	3.292 (2.483)	9.227 (8.851)	18.422 (20.215)	30.391 (36.674)	0.427 (0.096)
JLE	-	-	-	-	-	-	-	-	-
HEM	-	-	-	-	-	-	-	-	-
HGE	0.049 (0.037)	0.199 (0.078)	0.103 (0.041)	0.632 (0.237)	1.806 (1.402)	3.648 (4.780)	5.822 (10.134)	8.137 (15.215)	8.304 (3.109)

面板 B：$B_m/\mu_m = 6$。

Method	相对均方误差								Time (s)
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	六阶	七阶	
FULL	-	-	-	-	-	-	-	-	-
Cine++	**0.012** (0.002)	**0.029** (0.004)	**0.023** (0.003)	**0.063** (0.013)	**0.090** (0.024)	**0.110** (0.038)	**0.126** (0.043)	**0.136** (0.049)	80.270 (22.223)
Cine	0.015 (0.003)	0.038 (0.006)	0.030 (0.005)	0.082 (0.017)	0.117 (0.028)	0.149 (0.060)	0.171 (0.061)	0.186 (0.071)	110.807 (29.592)
Uniform	0.021 (0.003)	0.082 (0.023)	0.047 (0.006)	0.234 (0.061)	0.614 (0.380)	1.061 (0.805)	1.622 (1.647)	2.732 (5.601)	0.760 (0.184)
Truncating	0.030 (0.004)	0.077 (0.007)	0.059 (0.006)	0.177 (0.030)	0.292 (0.074)	0.388 (0.122)	0.475 (0.172)	0.555 (0.225)	0.402 (0.090)
Projecting	0.025 (0.006)	0.219 (0.143)	0.063 (0.014)	0.648 (0.296)	3.292 (2.483)	9.227 (8.851)	18.422 (20.215)	30.391 (36.674)	0.427 (0.096)
JLE	-	-	-	-	-	-	-	-	-
HEM	-	-	-	-	-	-	-	-	-
HGE	0.048 (0.026)	0.201 (0.070)	0.102 (0.026)	0.632 (0.212)	1.884 (1.441)	3.963 (5.687)	6.351 (11.195)	8.463 (16.383)	5.528 (1.523)

面板 C：$B_m/\mu_m = 8$。

Method	相对均方误差								Time (s)
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	六阶	七阶	
FULL	-	-	-	-	-	-	-	-	-
Cine++	**0.011** (0.002)	**0.027** (0.003)	**0.022** (0.003)	**0.058** (0.012)	**0.083** (0.022)	**0.101** (0.035)	**0.116** (0.041)	**0.126** (0.045)	101.437 (27.584)
Cine	0.014 (0.003)	0.035 (0.006)	0.027 (0.005)	0.076 (0.015)	0.108 (0.026)	0.135 (0.052)	0.154 (0.058)	0.167 (0.062)	140.544 (35.916)
Uniform	0.019 (0.002)	0.069 (0.016)	0.042 (0.005)	0.194 (0.051)	0.469 (0.269)	0.785 (0.554)	1.118 (0.854)	1.693 (2.289)	0.940 (0.233)
Truncating	0.030 (0.004)	0.077 (0.007)	0.059 (0.006)	0.177 (0.030)	0.292 (0.074)	0.388 (0.122)	0.475 (0.172)	0.555 (0.225)	0.402 (0.090)
Projecting	0.025 (0.006)	0.219 (0.143)	0.063 (0.014)	0.648 (0.296)	3.292 (2.483)	9.227 (8.851)	18.422 (20.215)	30.391 (36.674)	0.427 (0.096)
JLE	-	-	-	-	-	-	-	-	-
HEM	-	-	-	-	-	-	-	-	-
HGE	0.048 (0.026)	0.201 (0.070)	0.102 (0.026)	0.632 (0.212)	1.884 (1.441)	3.963 (5.687)	6.351 (11.195)	8.463 (16.383)	5.528 (1.523)

表 B.7：示例 2 中 $n = 100$、$m_{\max} = 7$ 和 $c = 2.0$ 的相对均方误差和计算时间。表中数值为 100 次重复实验的均值（标准差）。在每种设定下，均值最低的两个相对均方误差分别以粗体和下划线标示。

面板 A：$B_m/\mu_m = 4$。

Method	相对均方误差								Time (s)
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	六阶	七阶	
FULL	-	-	-	-	-	-	-	-	-
Cine++	**0.011** (0.003)	**0.033** (0.004)	**0.025** (0.005)	**0.062** (0.013)	**0.087** (0.028)	**0.105** (0.040)	**0.126** (0.052)	**0.135** (0.055)	71.734 (25.248)
Cine	0.014 (0.004)	0.045 (0.011)	0.033 (0.008)	0.087 (0.026)	0.128 (0.055)	0.162 (0.087)	0.191 (0.128)	0.213 (0.153)	66.557 (28.336)
Uniform	0.021 (0.003)	0.151 (0.093)	0.056 (0.007)	0.325 (0.105)	1.126 (0.860)	2.515 (2.668)	5.247 (7.560)	9.089 (19.242)	0.460 (0.153)
Truncating	0.022 (0.003)	0.071 (0.007)	0.051 (0.005)	0.142 (0.030)	0.234 (0.070)	0.315 (0.111)	0.385 (0.151)	0.450 (0.191)	0.360 (0.101)
Projecting	0.028 (0.010)	0.474 (0.274)	0.085 (0.022)	0.947 (0.335)	4.669 (2.088)	13.255 (7.824)	27.846 (20.651)	49.154 (43.606)	0.360 (0.099)
JLE	-	-	-	-	-	-	-	-	-
HEM	-	-	-	-	-	-	-	-	-
HGE	0.064 (0.051)	0.339 (0.186)	0.133 (0.056)	0.751 (0.320)	2.594 (2.226)	6.166 (7.590)	10.809 (15.923)	14.909 (25.264)	6.373 (1.398)

面板 B：$B_m/\mu_m = 6$。

Method	相对均方误差								Time (s)
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	六阶	七阶	
FULL	-	-	-	-	-	-	-	-	-
Cine++	**0.010** (0.003)	**0.029** (0.004)	**0.023** (0.005)	**0.054** (0.012)	**0.074** (0.023)	**0.091** (0.033)	**0.103** (0.039)	**0.112** (0.044)	88.050 (23.561)
Cine	0.012 (0.004)	0.039 (0.009)	0.029 (0.007)	0.074 (0.020)	0.107 (0.041)	0.132 (0.061)	0.154 (0.078)	0.173 (0.100)	89.452 (37.417)
Uniform	0.018 (0.003)	0.115 (0.072)	0.048 (0.006)	0.245 (0.077)	0.754 (0.563)	1.590 (1.550)	3.051 (4.326)	5.727 (12.901)	0.612 (0.203)
Truncating	0.022 (0.003)	0.071 (0.007)	0.051 (0.005)	0.142 (0.030)	0.234 (0.070)	0.315 (0.111)	0.385 (0.151)	0.450 (0.191)	0.360 (0.101)
Projecting	0.028 (0.010)	0.474 (0.274)	0.085 (0.022)	0.947 (0.335)	4.669 (2.088)	13.255 (7.824)	27.846 (20.651)	49.154 (43.606)	0.360 (0.099)
JLE	-	-	-	-	-	-	-	-	-
HEM	-	-	-	-	-	-	-	-	-
HGE	0.064 (0.051)	0.339 (0.186)	0.133 (0.056)	0.751 (0.320)	2.594 (2.226)	6.166 (7.590)	10.809 (15.923)	14.909 (25.264)	6.373 (1.398)

面板 C：$B_m/\mu_m = 8$。

Method	相对均方误差								Time (s)
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	六阶	七阶	
FULL	-	-	-	-	-	-	-	-	-
Cine++	**0.009** (0.003)	**0.027** (0.004)	**0.021** (0.004)	**0.049** (0.011)	**0.068** (0.022)	**0.080** (0.028)	**0.094** (0.037)	**0.101** (0.040)	121.951 (33.539)
Cine	0.011 (0.003)	0.035 (0.007)	0.027 (0.006)	0.066 (0.016)	0.095 (0.034)	0.116 (0.053)	0.135 (0.061)	0.152 (0.082)	110.336 (44.622)
Uniform	0.017 (0.003)	0.094 (0.047)	0.043 (0.006)	0.202 (0.068)	0.566 (0.401)	1.129 (0.959)	1.988 (2.423)	3.167 (5.459)	0.767 (0.254)
Truncating	0.022 (0.003)	0.071 (0.007)	0.051 (0.005)	0.142 (0.030)	0.234 (0.070)	0.315 (0.111)	0.385 (0.151)	0.450 (0.191)	0.360 (0.101)
Projecting	0.028 (0.010)	0.474 (0.274)	0.085 (0.022)	0.947 (0.335)	4.669 (2.088)	13.255 (7.824)	27.846 (20.651)	49.154 (43.606)	0.360 (0.099)
JLE	-	-	-	-	-	-	-	-	-
HEM	-	-	-	-	-	-	-	-	-
HGE	0.064 (0.051)	0.339 (0.186)	0.133 (0.056)	0.751 (0.320)	2.594 (2.226)	6.166 (7.590)	10.809 (15.923)	14.909 (25.264)	6.373 (1.398)

表格B.8–B.10 报告示例 3 中 $n = 100$, $m_{\max} = 7$, $c = 1.8$和 $B_m/\mu_m = 4$的结果。

表 B.8：示例 3 中二阶参与 $2\mu_2/n = 8$ 的相对均方误差和计算时间。条目为 100 次重复的均值（标准差）。在每种设置下，最低的两个平均相对均方误差分别以粗体和下划线标出。

面板 A： $m\mu_m/n = 16$ 对于 $m = 3, \ldots, 7$。

Method	相对均方误差								Time （秒）
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	六阶	七阶	
FULL†	-	-	-	-	-	-	-	-	-
Cine++	**0.017** (0.003)	**0.050** (0.006)	**0.040** (0.006)	**0.091** (0.017)	**0.129** (0.033)	**0.160** (0.046)	**0.179** (0.059)	**0.192** (0.064)	60.100 (21.088)
Cine	0.022 (0.005)	0.068 (0.019)	0.053 (0.013)	0.130 (0.050)	0.192 (0.073)	0.262 (0.293)	0.289 (0.311)	0.318 (0.262)	33.037 (4.044)
Uniform	0.032 (0.005)	0.183 (0.069)	0.093 (0.013)	0.454 (0.155)	1.410 (1.067)	2.665 (2.517)	4.796 (7.711)	7.857 (18.886)	0.271 (0.038)
Truncating	0.040 (0.006)	0.120 (0.014)	0.092 (0.011)	0.232 (0.042)	0.377 (0.098)	0.500 (0.155)	0.615 (0.218)	0.727 (0.292)	0.272 (0.037)
Projecting	0.028 (0.007)	0.316 (0.231)	0.097 (0.028)	0.739 (0.402)	3.513 (2.769)	9.421 (9.306)	18.576 (21.674)	30.929 (41.617)	0.274 (0.037)
JLE	-	-	-	-	-	-	-	-	-
HEM	-	-	-	-	-	-	-	-	-
HGE	0.048 (0.024)	0.291 (0.125)	0.145 (0.039)	0.732 (0.293)	2.266 (1.908)	5.018 (6.738)	8.014 (14.413)	11.588 (25.086)	2.748 (0.354)

面板 B： $m\mu_m/n = 20$ 对于 $m = 3, \ldots, 7$。

Method	相对均方误差								Time (s)
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	六阶	七阶	
FULL†	-	-	-	-	-	-	-	-	-
Cine++	**0.014** (0.002)	**0.043** (0.006)	**0.033** (0.005)	**0.072** (0.014)	**0.102** (0.026)	**0.126** (0.038)	**0.144** (0.047)	**0.155** (0.052)	50.756 (7.358)
Cine	0.018 (0.004)	0.061 (0.016)	0.045 (0.011)	0.105 (0.034)	0.152 (0.051)	0.198 (0.160)	0.229 (0.160)	0.242 (0.149)	40.184 (5.102)
Uniform	0.027 (0.004)	0.168 (0.056)	0.078 (0.012)	0.339 (0.100)	0.999 (0.644)	1.822 (1.512)	3.126 (3.661)	5.463 (12.572)	0.344 (0.058)
Truncating	0.034 (0.004)	0.110 (0.011)	0.078 (0.008)	0.190 (0.032)	0.311 (0.076)	0.414 (0.126)	0.508 (0.181)	0.598 (0.241)	0.287 (0.043)
Projecting	0.025 (0.006)	0.358 (0.234)	0.093 (0.024)	0.650 (0.292)	3.277 (2.433)	9.135 (8.655)	18.188 (19.742)	29.976 (35.770)	0.294 (0.045)
JLE	-	-	-	-	-	-	-	-	-
HEM	-	-	-	-	-	-	-	-	-
HGE	0.047 (0.026)	0.311 (0.122)	0.141 (0.033)	0.640 (0.217)	1.955 (1.550)	4.167 (6.015)	6.684 (11.836)	8.943 (17.925)	3.414 (0.447)

面板 C： $m\mu_m/n = 24$ 对于 $m = 3, \ldots, 7$。

Method	相对均方误差								Time (秒)
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	六阶	七阶	
FULL†	-	-	-	-	-	-	-	-	-
Cine++	**0.012** (0.002)	**0.039** (0.005)	**0.028** (0.004)	**0.061** (0.012)	**0.086** (0.022)	**0.110** (0.035)	**0.125** (0.042)	**0.142** (0.064)	58.274 (7.997)
Cine	0.016 (0.004)	0.055 (0.013)	0.039 (0.009)	0.086 (0.024)	0.126 (0.041)	0.164 (0.112)	0.188 (0.144)	0.205 (0.122)	44.037 (4.019)
Uniform	0.023 (0.003)	0.156 (0.072)	0.067 (0.010)	0.264 (0.085)	0.771 (0.695)	1.509 (2.077)	2.571 (4.439)	4.261 (11.116)	0.348 (0.047)
Truncating	0.029 (0.004)	0.101 (0.010)	0.068 (0.007)	0.159 (0.026)	0.262 (0.059)	0.349 (0.102)	0.428 (0.151)	0.504 (0.205)	0.261 (0.034)
Projecting	0.023 (0.005)	0.364 (0.222)	0.089 (0.023)	0.553 (0.226)	2.747 (1.764)	7.624 (6.516)	15.246 (15.532)	25.229 (28.765)	0.262 (0.036)
JLE	-	-	-	-	-	-	-	-	-
HEM	-	-	-	-	-	-	-	-	-
HGE	0.047 (0.032)	0.348 (0.296)	0.137 (0.047)	0.578 (0.245)	1.875 (2.071)	4.332 (8.405)	7.983 (23.072)	12.993 (54.124)	3.659 (0.274)

† 由于对 100 次重复而言，每项都需要超过 30 天，因此省略 FULL、JLE 和 HEM。

表 B.9：示例 3 中采用二阶参与时的相对均方误差和计算时间 $2\mu_2/n = 12$。结果为 100 次重复的均值（标准差）。每种设置下，均值最低的两个相对均方误差分别以粗体和下划线标出。

面板 A：$m\mu_m/n = 16$ 对于 $m = 3, \ldots, 7$。

Method	相对均方误差								Time (秒)
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	六阶	七阶	
FULL	-	-	-	-	-	-	-	-	-
Cine++	**0.016** (0.003)	**0.037** (0.004)	**0.031** (0.004)	**0.089** (0.019)	**0.128** (0.037)	**0.161** (0.051)	**0.176** (0.055)	**0.191** (0.065)	65.840 (18.478)
Cine	0.020 (0.004)	0.047 (0.008)	0.039 (0.007)	0.116 (0.024)	0.169 (0.047)	0.217 (0.073)	0.242 (0.089)	0.267 (0.125)	39.526 (6.115)
Uniform	0.029 (0.004)	0.112 (0.038)	0.065 (0.007)	0.412 (0.133)	1.286 (1.130)	2.436 (2.952)	4.595 (9.846)	6.544 (12.057)	0.316 (0.061)
Truncating	0.035 (0.005)	0.084 (0.009)	0.068 (0.007)	0.214 (0.038)	0.350 (0.092)	0.464 (0.149)	0.570 (0.209)	0.670 (0.273)	0.290 (0.059)
Projecting	0.028 (0.006)	0.190 (0.135)	0.066 (0.016)	0.737 (0.407)	3.538 (2.825)	9.534 (9.510)	18.808 (22.001)	31.249 (41.890)	0.305 (0.049)
JLE	-	-	-	-	-	-	-	-	-
HEM	-	-	-	-	-	-	-	-	-
HGE	0.048 (0.026)	0.185 (0.068)	0.104 (0.029)	0.724 (0.278)	2.204 (1.761)	4.747 (5.740)	7.487 (11.686)	10.713 (19.720)	3.290 (0.599)

面板 B：$m\mu_m/n = 20$ 对于 $m = 3, \ldots, 7$。

Method	相对均方误差								Time (s)
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	六阶	七阶	
FULL	-	-	-	-	-	-	-	-	-
Cine++	**0.013** (0.002)	**0.033** (0.004)	**0.026** (0.004)	**0.072** (0.015)	**0.100** (0.029)	**0.125** (0.043)	**0.139** (0.051)	**0.157** (0.062)	75.724 (17.360)
Cine	0.017 (0.004)	0.043 (0.008)	0.033 (0.006)	0.096 (0.021)	0.139 (0.038)	0.177 (0.081)	0.204 (0.084)	0.222 (0.111)	88.339 (23.346)
Uniform	0.025 (0.003)	0.104 (0.031)	0.055 (0.007)	0.310 (0.087)	0.904 (0.634)	1.640 (1.598)	2.879 (4.190)	4.711 (10.141)	0.572 (0.123)
Truncating	0.030 (0.004)	0.077 (0.007)	0.059 (0.006)	0.177 (0.030)	0.292 (0.074)	0.388 (0.122)	0.475 (0.172)	0.555 (0.225)	0.402 (0.090)
Projecting	0.025 (0.006)	0.219 (0.143)	0.063 (0.014)	0.648 (0.296)	3.292 (2.483)	9.227 (8.851)	18.422 (20.215)	30.391 (36.674)	0.427 (0.096)
JLE	-	-	-	-	-	-	-	-	-
HEM	-	-	-	-	-	-	-	-	-
HGE	0.049 (0.037)	0.199 (0.078)	0.103 (0.041)	0.632 (0.237)	1.806 (1.402)	3.648 (4.780)	5.822 (10.134)	8.137 (15.215)	8.304 (3.109)

面板 C：$m\mu_m/n = 24$ for $m = 3, \ldots, 7$.

Method	相对均方误差								Time (s)
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	六阶	七阶	
FULL	-	-	-	-	-	-	-	-	-
Cine++	**0.011** (0.002)	**0.029** (0.004)	**0.022** (0.003)	**0.059** (0.012)	**0.084** (0.021)	**0.104** (0.034)	**0.119** (0.040)	**0.133** (0.052)	86.721 (19.093)
Cine	0.014 (0.003)	0.039 (0.006)	0.029 (0.005)	0.080 (0.017)	0.117 (0.030)	0.148 (0.057)	0.170 (0.062)	0.186 (0.071)	42.754 (2.793)
Uniform	0.021 (0.003)	0.101 (0.049)	0.048 (0.006)	0.247 (0.081)	0.736 (0.802)	1.506 (2.893)	2.643 (6.281)	4.537 (14.726)	0.326 (0.013)
Truncating	0.026 (0.004)	0.072 (0.007)	0.051 (0.005)	0.150 (0.025)	0.247 (0.060)	0.329 (0.101)	0.403 (0.145)	0.471 (0.191)	0.231 (0.018)
Projecting	0.023 (0.005)	0.229 (0.141)	0.060 (0.014)	0.551 (0.227)	2.753 (1.799)	7.679 (6.671)	15.385 (15.901)	25.444 (29.423)	0.256 (0.019)
JLE	-	-	-	-	-	-	-	-	-
HEM	-	-	-	-	-	-	-	-	-
HGE	0.047 (0.033)	0.229 (0.169)	0.099 (0.037)	0.575 (0.232)	1.828 (1.888)	4.155 (7.582)	7.532 (20.124)	11.981 (45.577)	3.977 (0.211)

表 B.10：示例 3 中二阶参与的相对均方误差和计算时间 $2\mu_2/n = 16$。数值为 100 次重复的均值（标准差）。在每种设定下，最低的两个平均相对均方误差分别以粗体和下划线标示。

面板 A： $m\mu_m/n = 16$ 对于 $m = 3,..., 7$。

Method	相对均方误差								Time (秒)
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	六阶	七阶	
FULL	-	-	-	-	-	-	-	-	-
Cine++	**0.015** (0.003)	**0.028** (0.004)	**0.024** (0.004)	**0.085** (0.017)	**0.125** (0.039)	**0.155** (0.050)	**0.170** (0.058)	**0.189** (0.068)	64.438 (18.882)
Cine	0.018 (0.004)	0.035 (0.007)	0.030 (0.006)	0.110 (0.025)	0.164 (0.055)	0.214 (0.087)	0.229 (0.086)	0.265 (0.123)	41.513 (8.297)
Uniform	0.027 (0.004)	0.078 (0.027)	0.048 (0.006)	0.396 (0.129)	1.249 (1.090)	2.397 (2.979)	4.378 (8.687)	7.105 (19.327)	0.308 (0.066)
Truncating	0.032 (0.004)	0.062 (0.006)	0.052 (0.005)	0.203 (0.036)	0.336 (0.088)	0.448 (0.144)	0.552 (0.206)	0.652 (0.276)	0.266 (0.052)
Projecting	0.027 (0.006)	0.130 (0.090)	0.049 (0.010)	0.745 (0.415)	3.612 (2.909)	9.813 (9.895)	19.506 (23.285)	32.657 (45.361)	0.299 (0.061)
JLE	-	-	-	-	-	-	-	-	-
HEM	-	-	-	-	-	-	-	-	-
HGE	0.048 (0.027)	0.130 (0.045)	0.079 (0.025)	0.707 (0.257)	2.078 (1.597)	4.296 (4.781)	6.593 (9.310)	9.186 (14.830)	3.818 (0.829)

面板 B： $m\mu_m/n = 20$ 适用于 $m = 3,..., 7$。

Method	相对均方误差								Time (s)
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	六阶	七阶	
FULL	-	-	-	-	-	-	-	-	-
Cine++	**0.012** (0.002)	**0.025** (0.003)	**0.020** (0.003)	**0.069** (0.012)	**0.100** (0.029)	**0.122** (0.042)	**0.144** (0.052)	**0.151** (0.058)	57.215 (11.631)
Cine	0.015 (0.003)	0.032 (0.006)	0.026 (0.005)	0.089 (0.018)	0.133 (0.041)	0.168 (0.068)	0.190 (0.073)	0.213 (0.098)	50.978 (10.937)
Uniform	0.023 (0.003)	0.073 (0.023)	0.041 (0.005)	0.301 (0.086)	0.880 (0.624)	1.607 (1.509)	2.927 (4.719)	5.466 (16.693)	0.390 (0.109)
Truncating	0.027 (0.004)	0.057 (0.005)	0.045 (0.004)	0.169 (0.029)	0.280 (0.071)	0.374 (0.118)	0.459 (0.170)	0.540 (0.225)	0.271 (0.058)
Projecting	0.025 (0.006)	0.150 (0.098)	0.046 (0.010)	0.654 (0.299)	3.346 (2.527)	9.435 (9.074)	18.944 (20.890)	31.397 (38.109)	0.310 (0.056)
JLE	-	-	-	-	-	-	-	-	-
HEM	-	-	-	-	-	-	-	-	-
HGE	0.048 (0.028)	0.142 (0.046)	0.077 (0.024)	0.620 (0.202)	1.799 (1.297)	3.658 (4.910)	5.743 (9.451)	7.548 (13.205)	4.831 (1.261)

面板 C： $m\mu_m/n = 24$ for $m = 3,..., 7$.

Method	相对均方误差								Time (s)
	$\boldsymbol{Z}$	总体	二阶	三阶	四阶	五阶	六阶	七阶	
FULL	-	-	-	-	-	-	-	-	-
Cine++	**0.011** (0.002)	**0.023** (0.003)	**0.018** (0.003)	**0.057** (0.010)	**0.083** (0.021)	**0.104** (0.030)	**0.119** (0.039)	**0.135** (0.053)	58.489 (10.278)
Cine	0.014 (0.003)	0.030 (0.006)	0.022 (0.005)	0.076 (0.016)	0.115 (0.033)	0.147 (0.055)	0.164 (0.053)	0.188 (0.075)	67.078 (17.930)
Uniform	0.020 (0.003)	0.071 (0.035)	0.036 (0.004)	0.238 (0.075)	0.708 (0.748)	1.447 (2.689)	2.614 (6.227)	4.854 (18.179)	0.472 (0.167)
Truncating	0.024 (0.003)	0.054 (0.005)	0.039 (0.004)	0.143 (0.025)	0.239 (0.058)	0.320 (0.100)	0.393 (0.146)	0.461 (0.195)	0.285 (0.088)
Projecting	0.022 (0.005)	0.159 (0.099)	0.043 (0.009)	0.554 (0.231)	2.792 (1.838)	7.831 (6.845)	15.762 (16.449)	26.185 (30.702)	0.340 (0.094)
JLE	-	-	-	-	-	-	-	-	-
HEM	-	-	-	-	-	-	-	-	-
HGE	0.048 (0.033)	0.162 (0.118)	0.076 (0.031)	0.569 (0.232)	1.756 (1.826)	3.848 (7.108)	6.839 (18.961)	10.925 (43.246)	5.977 (2.468)

B.4 模型设定错误比较

为考察所提出的潜在空间模型之外的性能，我们对 CINE 与 JLE、HGE 和 HEM 分别进行了三组成对比较。每组比较均包括由示例1所使用模型生成的数据，以及由与竞争方法相关的机制生成的数据。我们固定 $n = 100$，并将潜在维度设为 $r = 2$，每种设置重复试验100次。在与 JLE 和 HEM 的比较中，我们评估了二阶和三阶超边。在与 HGE 的比较中，我们评估了二阶至七阶超边。这三组比较分别汇总。

对于 JLE型机制，各坐标独立生成，为 $z_{i\ell} = S_{i\ell}U_{i\ell}$，其中 $\Pr(S_{i\ell} = 1) = \Pr(S_{i\ell} = -1) = 1/2$，且 $U_{i\ell} \sim$ 服从 Uniform(0.6, 1)。令 $\boldsymbol{x}_i^{(m)} = \mathrm{diag}(\boldsymbol{w}_m)\boldsymbol{z}_i$，其中 $\boldsymbol{w}_2 =(1, 0.2)^\top$，且 $\boldsymbol{w}_3 =(0.2, 1)^\top$。出现概率为 $\mathrm{logit}\{p_2(\{i, j\})\} =(\boldsymbol{x}_i^{(2)})^\top \boldsymbol{x}_j^{(2)}$，以及 $\mathrm{logit}\{p_3(\{i, j, k\})\} = \sum_{\{u,v\}\subseteq\{i,j,k\}}(\boldsymbol{x}_u^{(3)})^\top \boldsymbol{x}_v^{(3)} + 4\sum_{\ell=1}^{2} \delta_{ijk,\ell} = 1$，否则为 -1。在给定潜在位置的条件下，二元响应独立生成。我们按照80%、10%和10%的比例，随机将二阶候选项划分为训练集、验证集和测试集。按照 Yuan 和 Qu（2023）中的分层抽样方案，若某个三阶候选项的三个构成二阶候选项均属于二阶训练集，则该三阶候选项符合条件，而不论其响应值如何。我们随机选取5%的符合条件的三元组用于训练，并将剩余的符合条件的三元组等量划分至验证集和测试集。两种方法拟合时均仅使用训练数据。评估使用完整的二阶和三阶候选集，包括候选集之外的三元组。

对于 HGE型机制，我们生成了 $L = 10000$ 个事件。n 个节点被分为两个规模相同的组，其潜在位置为 $\boldsymbol{z}_i = \boldsymbol{e}_{g_i} + \boldsymbol{\xi}_i$，其中 $\boldsymbol{e}_1$ 和 $\boldsymbol{e}_2$ 是标准基向量。$\sim$ 偏差 $\boldsymbol{\xi}_i$ 和初始事件位置 $\boldsymbol{f}_\ell$ 独立地从 $N_2(\boldsymbol{0}, \boldsymbol{I}_2)$ 中抽取，条件是它们落在 $[-1, 1]^2$ 内。我们将事件位置中心化为 $\boldsymbol{f}_\ell = \widetilde{\boldsymbol{f}}_\ell - L^{-1}\sum_{h=1}^{L} \widetilde{\boldsymbol{f}}_h$，并独立生成 $\alpha_i \sim$ Uniform(-0.1,0.1)。给定这些参数后，关联指示变量相互独立，且服从参数为 $q_{\ell i} =$ logit 函数$^{-1}(-4 + \alpha_i + \boldsymbol{f}_\ell^\top \boldsymbol{z}_i)$。的伯努利变量。对于一个完整团队$e$，它在 L 个事件中至少出现一次的概率为 $p(e) = 1 - \prod_{\ell=1}^{L}\{1 - r_\ell(e)\}$，其中$r_\ell(e) = \prod_{i\in e} q_{\ell i}\prod_{i\notin e}(1 - q_{\ell i})$。HGE 使用了完整的关联矩阵，包括空事件、单元素集合、重复团队以及阶数大于七的事件。对于 CINE，我们保留了二阶至七阶的超边，并将同一完整团队的重复出现合并为一条超边。

对于 HEM 型机制，我们独立生成了两个社区中心 $\boldsymbol{\mu}_1$，$\boldsymbol{\mu}_2 \sim N_2(\boldsymbol{0}$，$\boldsymbol{I}_2)$，并以概率 $1/2$将每个节点分配到其中一个社区。给定其社区标签 g_i，节点i 的潜在坐标为 $\boldsymbol{a}_i = \boldsymbol{\mu}_{g_i} + 0.5\boldsymbol{\varepsilon}_i$，其中 $\boldsymbol{\varepsilon}_i \sim N_2(\boldsymbol{0}$，$\boldsymbol{I}_2)$相互独立。将空节点坐标固定为 $(1/\sqrt{2}$，$1/\sqrt{2})^\top$后，对于 $e \in \Omega_m$，$m = 1$，2，3，出现概率为 $p_m(e) =0.3\ \mathrm{logit}^{-1}\left\{2^{-(3-m)/2}\sum_{\ell=1}^{2}\prod_{i\in e} a_{i\ell}\right\}$。给定潜在坐标后，超边出现指示变量按照概率为 $p_m(e)$ 的伯努利随机变量独立生成。

我们使用主文稿第4节定义的总体相对均方误差 RelMSE$_{p,\ \text{all}}$ 评估出现概率恢复情况；与 JLE 和 HEM 比较时，按二阶和三阶汇总，与 HGE 比较时，按二阶至七阶汇总。我们在每次重复实验中计算 RelMSE$_{p,\ \text{all}}$，并报告其在 100 次重复中的平均值。在每项比较中，Correct 列报告由该方法自身机制生成的数据的平均值；对于 CINE，该机制即示例1中的模型。Misspecification 列则报告由另一种方法的机制生成的数据的平均值。我们还报告比率 = 模型设定错误/正确和增量 = 模型设定错误 −正确；数值越小，表示模型设定错误下估计精度的下降越轻。

表 B。 11：CINE 与 JLE 之间的模型设定错误比较。正确设定和模型设定错误报告平均总体概率 相对均方误差，基于 100 次重复。

方法	正确	模型设定错误	比率	增量
CINE	0.1132	0.4374	**3.8633**	**0.3242**
JLE	0.0504	3.6520	72.4076	3.6016

表B.12：CINE 与 HGE 之间的模型设定错误比较。正确设定和模型设定错误分别报告了 100 次重复实验中的平均总体概率相对均方误差。

方法	正确	模型设定错误	比率	增量
CINE	0.0427	0.0520	**1.2172**	**0.0093**
HGE	0.1246	0.1993	1.6003	0.0748

表 B.13：CINE 与 HEM 之间的模型设定错误比较。正确设定和模型设定错误报告了 100 次重复中平均总体概率相对均方误差。

方法	正确	模型设定错误	比率	增量
CINE	0.1132	0.2370	**2.0934**	**0.1238**
HEM	0.0159	0.9443	59.4492	0.9284

表 B.11、B.12和B.13 显示，在全部三组成对比较中，与其他竞争方法相比，CINE 在概率估计误差的相对增量和绝对增量方面均较小。这些结果表明，CINE 对模型设定错误具有较强的稳健性，在不同生成机制下表现更加稳定。

C 其他真实数据结果

我们进一步使用相同的 644 位作者和潜在维度 $r = 2$，将 CINE 与 Uniform、截断、投影和 HGE 进行比较。对于每种方法，我们考虑两组保留阶数： $\mathcal{M}_1 = \{2, 3\}$ 和$\mathcal{M}_2 = \{2, \ldots, 5\}$。这些方法利用可用协作信息的方式各不相同。对于每个保留阶数 $m > 2$，Uniform 使用与 CINE 相同的抽样预算 B_m，但在每一阶内均匀采样负样本。截断仅保留原始的二作者团队和三作者团队，因此在 $\mathcal{M}_1$ 和 $\mathcal{M}_2$下得到相同的拟合结果。在 $\mathcal{M}_1$下，投影与截断得到相同的拟合结果。在 $\mathcal{M}_2$下，它

还会将每个规模更大的团队表示为其所有三作者子集，并且每个不同的三元组只计数一次。由于计算成本较高，我们省略了 $\mathcal{M}_2$ 下 FULL、HEM 和 JLE 的拟合结果。候选超边的数量从 $\mathcal{M}_1$ 下的 4.45×10^7 增加到 $\mathcal{M}_2$下的 9.16×10^{11}，使得穷举拟合在计算上颇为困难。

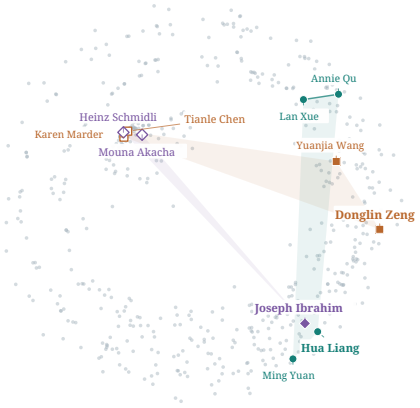
不同方法估计出的潜在位置存在显著差异，尤其是在总体配置上。如图 C.3 和 C.4所示，在 $\mathcal{M}_1$ 下，截断与投影拟合完全相同。在图 C.3、C.5和C.4 的左侧面板中，193 位未出现在任何已观测到的二人或三人团队中的作者几乎位于同一点。在截断与投影拟合中，其余大多数作者沿着一个环分布，中心区域基本空置。对于投影方法，这种环状结构在 $\mathcal{M}_2$ 下仍然存在，但各位作者的相对位置发生了显著变化。在这类配置中，作者之间的邻近程度主要取决于他们沿环分布的相对位置。当两个群体位于环的相对两侧时，同时与两个群体合作的作者可能看起来接近其中一组，却远离另一组，因此仅凭估计的位置很难辨明其与两个群体的联系。

除了总体配置上的这些差异外，各方法在纳入规模更大的团队会如何影响作者之间估计距离方面也有所不同。例如，如图 C.2 和 C.4所示，在 Uniform 和投影拟合中，Annie Qu 与 Donglin Zeng 的位置很近，在 $\mathcal{M}_1$ 下却相距甚远。 $\mathcal{M}_2$相较之下，在 CINE 拟合中，纳入规模更大的团队后，两人之间的距离有所缩短。鉴于两人都曾从事变量选择和纵向数据方面的相关研究，这些差异值得关注。在拟合样本中，Zeng 与合作者于 2008年研究了半参数回归模型中用于变量选择的惩罚估计函数。随后，Qu 与合作者在拟合样本之外的一项 2012年研究中，为高维纵向数据提出了惩罚广义估计方程。之后，Qu 和 Cho 于 2013年研究了相关数据的惩罚二次推断函数，并明确指出 Zeng 与合作者早期的研究是 2012年方法的基础。这些研究为两位作者之间的联系提供了具体的方法论依据。两人在纵向回归方面的研究也体现了另一层联系。在拟合样本中的一篇 2015年论文里，Zeng 与合作者研究了稀疏异步纵向数据。他们的论文明确讨论了 Qu 和 Li 的早期研究；该研究将惩罚样条与二次推断函数相结合，用于估计纵向数据的变系数模型。由于截断会舍弃由四位或更多作者组成的团队，因此在两组阶数下得到的位置完全相同。因此，截断无法体现 CINE 中观察到的组内距离缩短，包括 Hua Liang、Annie Qu 和 Lan Xue 之间更近的位置。Annie Qu 和 Lan Xue 之间还存在另一处差异。在 $\mathcal{M}_1$ 下，两人在投影拟合和 CINE 中的位置都很近。而在 $\mathcal{M}_2$下，两人之间的距离在投影拟合中增大，在 CINE 中则缩短。鉴于她们曾多次合作研究模型选择，这些差异值得关注。在一项 2010年的三作者研究中，她们与 Jianhui Zhou 合作，将多项式样条与惩罚二次推断函数结合，用于相关数据的模型选择；该研究已纳入 $\mathcal{M}_1$之下。保留在 $\mathcal{M}_2$

其中还包括她们与 Li Wang 和 Hua Liang 于 2014年开展的另一项合作研究，该研究提出了广义加性部分线性模型的估计与模型选择方法。她们的合作也延伸至拟合样本之外；在那里，她们于 2012年研究了高维变系数模型的变量选择。这些研究为她们在 CINE 中位置更近提供了进一步背景。不同方法之间的类似对比也体现在 Ming Yuan 与 Hua Liang、Annie Qu 和 Lan Xue 的关系上。如图 C.1所示，纳入规模更大的团队后，在 CINE 拟合中，Yuan 与另外三位作者的距离均缩短；而在 HGE 中，这三段距离则均增大。她们在变量选择方面的研究有助于理解这一模式。Yuan 和 Yi Lin 于 2006年提出了用于变量组选择的组 LASSO。随后，Qu 和 Xue 在 2010年的一项三作者研究中，与 Jianhui Zhou 合作，将她们对加性成分的分组处理与 Yuan 和 Lin 的研究联系起来。她们还在拟合样本之外的一项 2012年研究中，将组 LASSO 框架应用于高维变系数模型。Liang 与合作者在 2014年的一项四作者研究中，使用组 LASSO 及其自适应版本，对动态基因调控网络进行变量选择。Liang 和 Yuan 还共同与 Hui Zou 合作。Yuan 和 Zou 于 2008年研究了复合分位数回归与变量选择；而 Liang 和 Zou 于 2012年合作研究了高维部分线性模型中的变量选择。此外，一项由 Li Wang、Xue、Qu 和 Liang 合作的四作者研究已纳入 $\mathcal{M}_2$ ，该研究使用分组样条系数进行模型选择，并明确将其方法与 Yuan 和 Lin 的研究联系起来。

我们还在主分析所使用的留出测试集上评估了其他方法。该测试集包含55条现有超边和8,863,072个三阶负例。我们使用训练数据拟合每种方法，并在各方法中评估相同的测试候选项。对于HGE，我们通过对每个候选三元组的中心化作者位置的三个成对内积求和，为其评分。当纳入由四名和五名作者组成的团队时，CINE的AUC从 $\mathcal{M}_1$ 下的0.737增至 $\mathcal{M}_2$ 下的0.846。在 $\mathcal{M}_2$下，Uniform、截断、投影和HGE的AUC分别达到0.757、0.801、0.788和0.775。因此，在 $\mathcal{M}_2$下，CINE在这些方法中的AUC最高。

(a) CINE, $\mathcal{M}_1 = \{2, 3\}$



(b) CINE, $\mathcal{M}_2 = \{2, \ldots, 5\}$

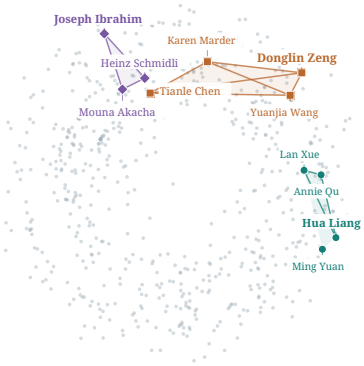
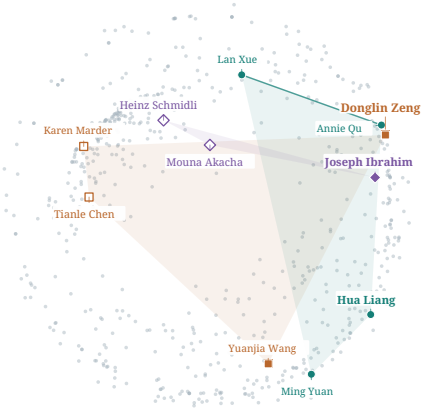


图 C.1：根据2005年至2015年的合著超图估计的作者潜在位置。面板 (a) 和 (b) 分别展示了在 $\mathcal{M}_1 = \{2, 3\}$ 和 $\mathcal{M}_2 = \{2, \ldots, 5\}$ 下的 CINE 拟合结果。

(a) Uniform, $\mathcal{M}_1 = \{2, 3\}$



(b) Uniform, $\mathcal{M}_2 = \{2, \ldots, 5\}$

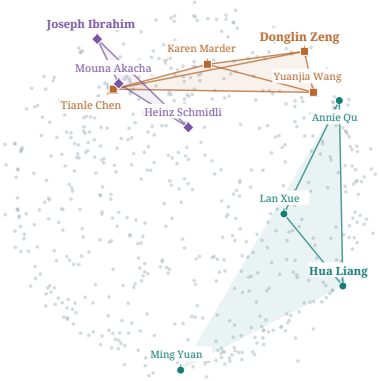
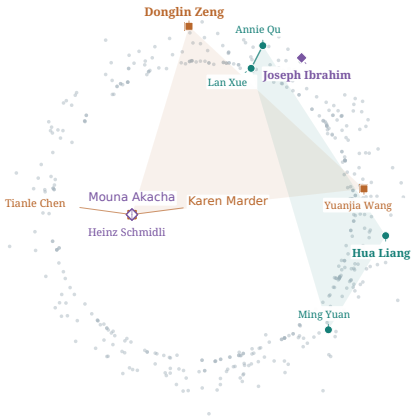


图 C.2：根据2005年至2015年的合著超图估计的作者潜在位置。面板 (a) 和 (b) 分别展示了在 $\mathcal{M}_1 = \{2, 3\}$ 和 $\mathcal{M}_2 = \{2, \ldots, 5\}$ 下的 Uniform 拟合结果。

（a）截断，$\mathcal{M}_1 = \{2, 3\}$



（b）截断，$\mathcal{M}_2 = \{2, ..., 5\}$

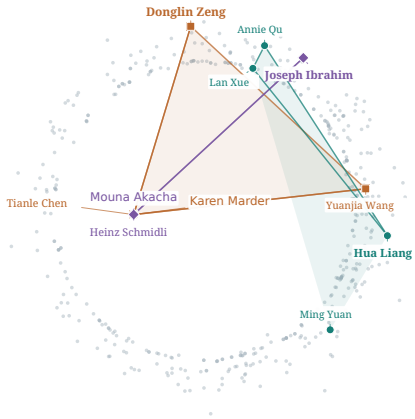
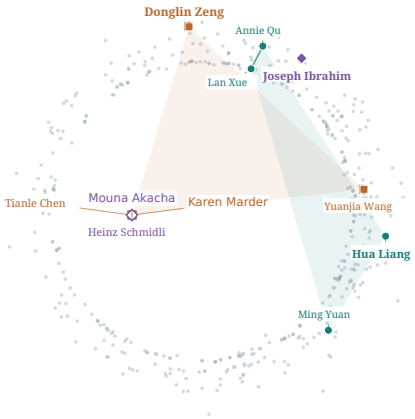


图 C.3：根据2005年至2015年的合著超图估计的作者潜在位置。面板 (a) 和 (b) 分别显示在 $\mathcal{M}_1 = \{2, 3\}$ 和 $\mathcal{M}_2 = \{2, \ldots, 5\}$ 下的截断拟合。

(a) 投影，$\mathcal{M}_1 = \{2, 3\}$



(b) 投影，$\mathcal{M}_2 = \{2, ..., 5\}$



图 C.4：根据2005年至2015年的合著超图估计的作者潜在位置。面板 (a) 和 (b) 分别显示在 $\mathcal{M}_1 = \{2, 3\}$ 和 $\mathcal{M}_2 = \{2, \ldots, 5\}$ 下的投影拟合。

(a) HGE, $\mathcal{M}_1 = \{2, 3\}$

(b) HGE, $\mathcal{M}_2 = \{2, ..., 5\}$



图 C .5：基于2005年至2015年共同署名超图估计的作者潜在位置 .
图版（ a）和（b）分别展示了 $\mathcal{M}_1 = \{2, 3\}$ 和 $\mathcal{M}_2 = \{2, \ldots, 5\}$下的 HGE 拟合结果。

参考文献

Goldberg, P. W. 和 Jerrum, M. R.（1995），“由实数参数化的概念类的 Vapnik-Chervonenkis 维数的界”，*Machine Learning***18**（2–3），131–148。

Yuan, Y. and Qu, A. (2023), ‘多层级链接预测的高阶联合嵌入’, *Journal of theAmerican Statistical Association***118**(543), 1692–1706.