利用样本统计量进行协作均值估计（ColME）中的多重置信区间

Nikola Stankovi´c

摘要—个人数字设备和物联网（IoT）的迅猛发展，加剧了对协作式和去中心化学习框架的需求。由于这些设备持续生成敏感的高维数据，集中式传输往往不切实际，而纯粹的本地学习则会因数据逐渐累积而收敛缓慢。协作方法允许智能体利用彼此的信息来改进估计，从而缓解这些问题；然而，挑战在于数据分布的异质性。每个智能体通常都面临个性化学习问题，只有数据由相同分布生成、属于同一相似类的智能体之间开展协作才有益。本文研究此类异质环境下的个性化在线均值估计问题，其中每个智能体观测到的数据均来自其自身的 σ-次高斯分布。为应对异质性挑战，协作算法使智能体能够实时识别相似性类别，并利用同一类别中其他智能体的信息，以加快收敛并提高准确度。本研究基于现有方法展开：协作均值估计（colME）通过置信区间和智能体间的交互来改进估计；其基于图的扩展（C-colME 和 B-colME）则提升了分布式场景中的可扩展性和稳健性。由于方差估计在上述方法中起着关键作用

本文提出了一种能够准确、局部且实时估计方差的方法。文中介绍了一种与均值无关的在线方差估计量。此外，我们还将样本峰度估计纳入其中，无需明确依赖各智能体的均值，从而能够设计出更精细的置信区间，同时考虑数据在均值、方差以及高阶特征（分布）方面的差异。我们推导了样本标准差和样本峰度的置信区间估计量。结合样本均值估计方法，这些结果可用于设计一个统一流程，依据样本均值、样本方差和样本峰度联合构造多重置信区间。该框架支持在具有挑战性的情形下进行协同估计，例如不同类别的均值相似但方差不同，或均值和方差均相近而底层分布在高阶特征上存在差异；迄今为止，后者尚未纳入考量。我们以预言机情形作为基准，评估所提方法的准确度和收敛速度。最后，通过数值示例阐释理论结果。

arXiv:2603.01216v1 [eess.SP] 1 Mar 2026

关键词—**协同估计、均值估计、随机图、置信区间交集、样本方差、样本ku度、物联网、去中心化学习**。

一、引言

个人数字设备的发展及其广泛应用，加上计算技术和物联网（IoT）的进步，进一步提升了对去中心化协同计算范式的需求。这些设备本质上以数据为中心，所生成的信息往往敏感或数量庞大，因此集中式传输既不切实际，也不受欢迎。尽管可以选择独立进行本地处理，但这种孤立的学习方式可能导致收敛缓慢，尤其是在数据逐步累积的在线场景中。

过去几年，联邦学习（FL）领域针对涉及多个拥有本地数据集的智能体的协同估计与学习问题开展了广泛研究 [1], [2]。传统的联邦学习方法侧重于学习一个适用于所有参与智能体的全局模型，而客户端数据固有的统计异质性则推动了人们对个性化联邦学习技术的浓厚兴趣。这些技术旨在开发适配各个客户端特定数据分布的模型，以提高估计精度和收敛行为 [3]–[5], [5]–[8]。

个性化联邦学习中一种常见的策略是将客户端划分为多个聚类，随后为每个聚类训练一个专用模型 [3], [5], [8]。理想的聚类应将拥有相似局部最优模型的客户端归为一组，从而实现有效协作并加快收敛。然而，由于这些最优模型通常事先未知，模型学习与聚类识别的过程便密不可分。已有多项研究提出使用经验相似度度量，例如局部模型之间的欧氏距离或其更新的余弦相似度，作为一种实用的解决方案——度量[3], [5]。另一些研究则假设已掌握客户端数据分布之间距离的某种先验知识 [8], [9]。尽管如此，准确估计这些距离仍是一项重大挑战，尤其是在客户端局部数据可能有限的在线联邦学习场景中，正如 [9]所指出的那样。因此，相似度估计错误可能会显著减缓收敛速度，原因在于它可能阻碍有益的协作，或促使各方交换误导性信息。

在在线学习场景中，协作学习主要是在多臂赌博机（MAB）背景下得到研究的。然而，现有的大多数方法都考虑由多个智能体协作解决单个 MAB 实例，其中协作旨在加快学习进程，以实现共同目标。协作通常通过广播消息[10], [11],经由中心服务器 [12],实现，或通过网络图 [13]–[17]上的本地消息交换实现。在所有这些情形中，智能体共享一个共同的学习目标，协作主要用于降低不确定性，而非应对异质性。

近来，研究开始关注各智能体的臂均值存在差异的协作式 MAB 场景。 [18]的扩展研究考察了碰撞模型下存在异构奖励的情形，另一些研究则考虑联邦式或部分个性化的目标 [19]–[21]。与我们的工作一个关键区别在于，这些方法无需在线识别局部分布之间的关系来解决相应问题，因为相似性结构要么是预先假定的，要么并非收敛的关键因素。

联邦设置下的均值估计问题已被视为一项具有实际意义的基础构件 [22]–[24] ，应用领域包括智慧农业、电网管理和医疗健康 [25]。在此背景下， [26]提出了协作均值估计（ColME），这是一种在线、去中心化框架，智能体在其中与同伴交换信息，以改进其均值估计。然而，ColME 存在可扩展性局限，其收敛行为

N. Stankovi´c 就职于黑山波德戈里察的黑山大学（电子邮件：nikola.stankovic1@edu.ucg.ac.me）。作者谨向意大利都灵理工大学的 Emilio Leonardi 教授和 Franco Galante 博士致谢，感谢他们将作者引入这一研究主题，并在作者撰写硕士论文期间提出宝贵意见。

对智能体之间的异质性以及协作结构高度敏感。 [27] 的研究在协作均值估计的基础上，针对多维数据和更广泛的分布类别开发了可扩展算法，其中智能体根据感知到的相似性进行自组织。图 [28]–[33] 用作交互域，并采用消息传递和基于共识的变体 [27]。收敛速度深受智能体识别有用协作者的速度，以及信息随时间聚合的有效程度影响。

因此，协同均值估计的一项关键挑战在于，如何在线识别相似性结构，从而在加速收敛的同时实现有效协作。足够相似的智能体之间进行协作，能够显著降低估计误差；然而，不恰当或过于保守的协作策略可能会延缓学习。这一挑战在在线场景中尤为突出，因为数据是按顺序到达的，且必须在不确定性下作出协作决策。置信区间 [34]–[39]，用作确定相似性类别的关键工具 [26]。

本文通过研究方差及其他数据统计量的在线、局部和协同估计机制，探讨置信区间确定所面临的挑战；这些机制有助于高效应用置信区间和自适应图协作。

本文提出了一种在均值估计的协作场景中估计样本方差的方法，且不将智能体的均值作为最终目标。方差在局部进行估计。通过与邻近智能体协作，可以提高估计的准确度。本文考虑了整个系统中方差相同，以及方差在不同类别间存在差异的情况。此外，本文还在不使用智能体均值的情况下实现了样本峰度估计，并推导出估计的样本标准差和峰度的标准差。结合样本估计方法，本文利用这一结果提出了一种新的综合方法，在使用置信区间时基于样本均值、样本标准差和样本峰度进行估计。通过这种方法，我们得以在极其复杂的情况下考虑并采用协同估计；迄今为止，文献中尚未考虑过这些情况，例如

例如，类别的均值相近但标准差不同，或者类别的均值和标准差都相近，但数据分布不同，导致峰度不同。文中给出了实验示例，表明要成功实现相似类分离，必须使用全部三种样本估计量。为保持置信区间（从而避免错误剪枝）并同时实现快速收敛，对于接近这些区间边界的数据，引入了数据点权重。为此，采用加权图，并使用高斯核作为图权重。通过数值验证了该理论。

本文的组织结构如下。第II节介绍问题表述和协作均值估计框架。第III节介绍所提方法，用于估计样本方差和峰度。第IV节回顾用于均值估计的协作方法。第V节介绍多重置信区间。第VI节和第VII节给出多分类示例。第VIII节总结全文。

II. 问题定义

考虑 $A = N$ 个智能体，它们接收数据并能够进行一定处理，还能与其他智能体通信。本文关注的核心任务是识别数据分布相似或相同的智能体集合。检测到该集合后，通过交换数据进行协作处理，可以显著提高估计收敛速度和准确度。

我们假设每个智能体 $a \in \mathcal{A} = \{1, 2, , \ldots, A\}$接收一个随机样本 $x_a(t) \in \mathbb{R}^1$，该样本抽取自分布 D_a，具有期望均值 $\mu_a = \mathbb{E}\{x_a(t)\}$。信号的向量形式为：

$$\mathbf{x}(t) = \left[x_1(t), x_2(t), \cdots, x_A(t)\right]^T. \tag{1}$$

相似类： 从相同分布 D_a 中接收随机样本 $x_a(t) \in \mathbb{R}^1$，且具有相同期望均值 μ_a 的智能体构成一个相似类。智能体a所属的相似类记为 $\mathcal{C}_a$。

我们还假设，所有智能体中共有 C个相似类。若条件 $\Delta_{aa'} = |\mu_a - \mu_{a'}| = 0$ 成立，即两个智能体具有相同的均值，则智能体 a 和 a' 属于同一个相似类。

在协作均值估计（colME）中，每个智能体 $a \in \mathcal{A}$ 的目标是尽可能快地估计其均值，同时使用自己的数据以及属于同一相似类的某些协作智能体的数据 $\mathcal{C}_a$。

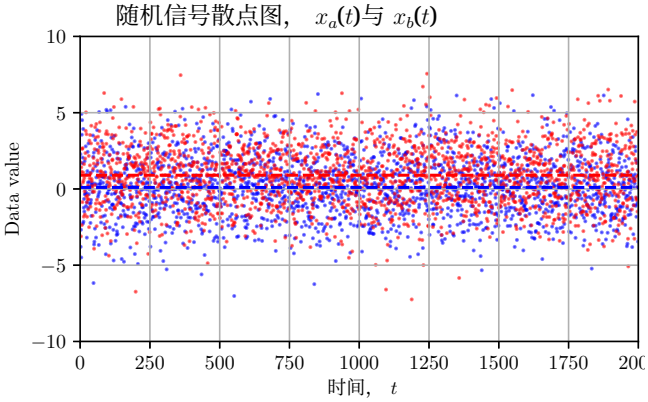


图1：两个来自不同相似类的智能体 a和 b的高斯数据示意图，$x_a(t)$、 $x_b(t)$（以红点和蓝点表示）；对于 $t \in [0, 2000]$，其中 $\mu_a = 0.1$、 $\mu_b = 0.9$、 $\Delta_{ab} = 0.8$ 和 $\sigma = 2$。

A. 局部均值

一个智能体在采集 t 个样本后（采样步长等于 1，即 $t = 1$、2、3、.....、T）得到的均值称为局部均值。其定义如下：

$$\bar{x}_{a,a}(t) = \frac{1}{t}\sum_{\tau=1}^{t} x_a(\tau) = \frac{1}{t}x_a(t) + \frac{t-1}{t}\bar{x}_{a,a}(t-1). \tag{2}$$

这个经验局部均值会缓慢趋近于真实均值 μ_a。如果我们能够确定智能体 a 所属的相似类 $\mathcal{C}_a$，或者至少确定该类中的一部分智能体，就能显著改善收敛。这样，我们便可以利用更多数据，通过不仅在时间维度上、还在具有相同均值的数据集合（协作集合）上进行平均，来提高均值估计的收敛速度。

所有智能体的均值 $a = 1$、2、.....、$A = N$ 可以写成向量形式：

$$\mathbf{X}(t) = \begin{bmatrix} \bar{x}_{1,1}(t) \\ \bar{x}_{2,2}(t) \\ \vdots \\ \bar{x}_{A,A}(t) \end{bmatrix} \text{ with } \mathbf{X}(t) = \frac{1}{t}\mathbf{x}(t) + \frac{t-1}{t}\mathbf{X}(t-1). \tag{3}$$

图2展示了数据中两个相似性类别 $\mu_a = 0.1$ 和 $\mu_b = 0.9$ 的局部均值的多次实现示例。红线表示具有均值 $\mu_a = 0.1$ 的 $N/2 = 100$ 个智能体的局部均值，而与相同数量的智能体对应的 $\mu_a = 0.9$ 的局部均值则以红色表示。我们可以看到，对于较小的 t，局部均值的实现可能会重叠；随着时间 t推移，它们随后会非常缓慢地收敛到各自的真实均值。

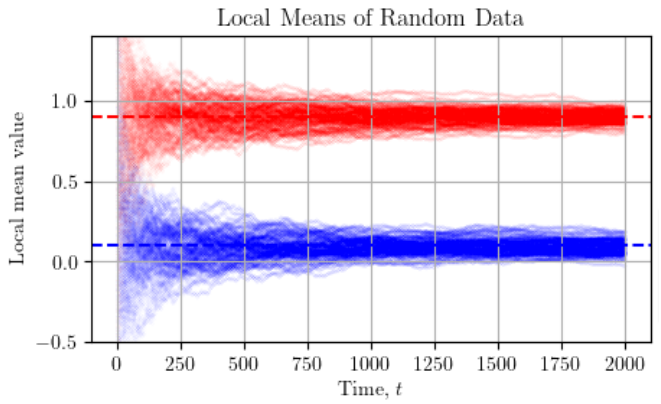


图2：数据 $x_a(t)$ 的局部均值 $\bar{x}_{aa}(t)$ 示意图，其中包含 2 个相似类和 $N = 200$ 个智能体（以红色和蓝色表示），$\mu_a = 0.1$、$\mu_b = 0.9$、$\Delta_{ab} = |\mu_a - \mu_b| = 0.8$，对应于 $t \in [0, 2000]$。数据 $x_a(t)$ 服从标准差为 $\sigma = 2$ 的高斯分布。

在某些方法中，也可以使用与局部均值 $\bar{x}_{a,a}(t)$ 密切相关的概念——局部和。其定义如下：

$$m_{a,a}(t) = \sum_{\tau=1}^{t} x_a(\tau) = x_a(t) + m_{a,a}(t-1) = t\bar{x}_{a,a}(t). \quad (4)$$

如果预先知道所有相似类，就可以得到估计均值的 **oracle 解**。然后，在每个时刻，通过对同一相似类 $a \in \mathcal{C}_a$ 中所有智能体的局部均值 $\bar{x}_{a,\,a}(t)$ 求平均，即可得到估计的预言机均值。数值示例将使用这个 oracle（可能达到的最佳）值来检验所提出解法的准确度。

B. 局部均值的置信区间

为了估计与所考虑的智能体 a 属于同一相似类的一组智能体，将使用置信区间。

在每个时刻，t，智能体接收新样本 $x_a(t)$ 后，即形成一个新的向量 $\mathbf{x}(t)$，它们会更新各自的局部均值。随后，每个智能体将其局部均值与其邻居的局部均值进行比较（邻域的定义将由算法指定），以尝试判断所考察的邻居是否属于同一相似类。

一般而言，随机变量（在此情况下为局部样本均值，$\bar{x}_{a,\,a}(t) = \frac{1}{t}\sum_{\tau=1}^{t} x_a(\tau)$）的置信区间定义为

$$\mathbb{I}_a(t) = [\bar{x}_{a,a}(t) - \beta_\delta(t), \bar{x}_{a,a}(t) + \beta_\delta(t)], \quad (5)$$

其中，$\beta_\delta(t)$规定了接下来将要讨论的置信区间宽度。

考虑随机变量 $\bar{x}_{a,\,a}(t)$，其相对于均值 μ_a 的概率密度函数是不对称的。置信区间根据随机变量的某个值 $\bar{x}_{a,\,a}(t)$ 与其均值 μ_a 之差小于给定较小概率 δ的概率来定义。

基于

$$\mathbb{P}\Big(|\bar{x}_{a,a}(t) - \mu_a| > \sigma B_\delta(t)\}\Big) < \delta, \quad (6)$$

其中，σ 是 $x_a(t)$的标准差，我们可以得出随机变量 $\bar{x}_{a,\,a}(t)$满足不等式

$$\mu_a - \sigma B_\delta(t) \le \bar{x}_{a,a}(t) \le \mu_a + \sigma B_\delta(t), \quad (7)$$

其成立的概率大于 $1 - \delta$。

对于 $\lim_{t\to\infty} B_\delta(t) = 0$ 的情况，我们可以指出，基于样本均值 $\frac{1}{t}\sum_{\tau=1}^{t} x_a(\tau)$ 对均值 μ_a 进行估计，是一个关于正确（PAC）问题：我们可以将估计误差缩小到任意小的值 ε，且其发生概率高于 $1 - \delta$，对于任意小的 δ 均如此；也就是说，对于任意小的 (ε, δ)，都存在 τ_a，使得对于任意 $t > \tau_a$，我们有

$$\mathbb{P}\Big(|\bar{x}_{a,a}(t) - \mu_a| < \varepsilon\Big) > 1 - \delta. \quad (8)$$

这一事实稍后将用于定义置信区间的两个关键性质，这些置信区间将用于协作均值估计。

特殊情形下的置信区间如下：

1) 高斯分布：如果局部均值 $\bar{x}_{a,\,a}(t)$可以视为一个高斯分布随机变量，那么它服从$\bar{x}_{a,\,a}(t) \sim \mathcal{N}(\mu_a, \sigma\frac{1}{\sqrt{t}})$。对于该分布，我们可以轻松计算给定$\delta$ 的置信区间，计算如下：

$$\sigma B_\delta(t) = \beta_\delta(t) = z_{1-\delta/2}\frac{\sigma}{\sqrt{(t)}}, \quad (9)$$

其中，各种 $1 - \delta$ 下若干常用值（$1 - \delta$，$z_{1-\delta/2}$）的值$z_{1-\delta/2}$ 如下：（0.95，1，96），(0.975，2.24)，(0.99，2.58)，(0.999，3.29)。

2) 学生 t 分布：如果样本数量较少，且标准差也由 $\hat{}\sigma$ 估计，则应使用学生 t 分布，其置信界定义为：$\beta_\delta(t) = \sigma B_\delta(t) = t_{1-\delta/2,\ \nu}\hat{}\frac{\sigma}{\sqrt{t}}$，其中，值$t_{1-\delta/2,\ \nu}$ 取决于 δ 和自由度ν。

3) σ-次高斯分布：σ-次高斯分布是一类概率分布，其特征是尾部不比方差为 σ的高斯分布更重。如果随机变量 $x_a(t)$服从 σ-次高斯分布，则：

$$\mathbb{E}[e^{\lambda(\mathrm{x}-\mu_a)}] \le e^{\frac{\sigma^2\lambda^2}{2}}, \quad \text{for all } \lambda \in \mathbb{R}, \quad (10)$$

其中随机变量用 x 表示。对于此分布，有：

$$\mathbb{P}(|\mathrm{x} - \mu_a| \ge t) \le 2e^{-\frac{t^2}{2\sigma^2}}, \quad \text{for all } t \ge 0. \quad (11)$$

这些界表明，偏离均值较大的概率会迅速衰减。对于方差为 σ 的次高斯分布随机变量，我们有 $\mathrm{Var}(\mathrm{x}) \le \sigma^2$。

σ 次高斯分布的示例包括：

- 均值为 $\mathrm{x} \sim \mathcal{N}(\mu_a, \sigma_1)$ 的高斯分布，其中 $\sigma_1 \le \sigma$；
- 拉德马赫分布：随机变量 x 以相等概率取值 +1 和 -1，是 1-次高斯分布；
- 在 $[-a, a]$ 范围内的均匀分布是 $\frac{a}{\sqrt{3}}$-次高斯分布；
- 取值为 $+a$ 和 $-a$ 且概率相等的对称伯努利分布是 a-次高斯分布。
- 独立次高斯分布的线性组合是 $\sqrt{a_i\sigma_i^2}$ 次高斯分布。
- 元素相互独立且服从 σ-次高斯分布的向量和矩阵服从 σ-次高斯分布。

置信区间（通过拉普拉斯方法，由 Maillard 引理推得）使用 [26], [27],[40]定义

$$\mathbb{P}\Big(\bar{x}_{a,a}(t) - \mu_a\Big) > \sigma\sqrt{\frac{2}{t}\Big(1 + \frac{1}{t}\Big)\ln\Big(\frac{\sqrt{t+1}}{\delta/2}\Big)}\Big) \quad (12)$$

作为拉普拉斯界

$$\beta_\delta(t) = \sigma B_\delta(t) = \sigma\sqrt{\frac{2}{t}\Big(1 + \frac{1}{t}\Big)\ln\Big(\frac{\sqrt{t+1}}{\delta/2}\Big)}. \quad (13)$$

4) 有界四阶矩分布：考虑一个具有有界四阶矩的分布，

$$\mathbb{E}\{(x_a(t)-\mu_a)^4\} \le \mu_4. \tag{14}$$

真实均值 μ_a 以大于 $1-\delta$ 的概率属于以 $\bar{x}_{a,a}(t)$ 为中心、以 [27] 为界的置信区间

$$\beta_\delta(t) = \sigma\sqrt[4]{\frac{2(\kappa+3)}{\delta/2}\frac{(1+\ln^2(t))}{t}}. \tag{15}$$

当所有变量均服从相同分布时，κ 对应于峰度

$$\kappa = \frac{\mathbb{E}\{(\mathrm{x}-\mu_a)^4\}}{\sigma^4}. \tag{16}$$

该结果产生的置信区间比使用拉普拉斯界得到的置信区间更紧。对于高斯分布$\bar{x}_{a,a}(t)$，我们有 $\kappa=3$。对于高斯分布和 $\delta=$ 0.01 we can write $\beta_\delta(t) = 8.32\sigma\sqrt[4]{(1+\ln^2(t))/t}$.

由于置信区间将在所介绍方法的收敛中发挥关键作用，我们将更详细地讨论其性质和行为。

置信区间的性质： 置信区间以高于 $1-\delta$ 的概率包含数据 $\bar{x}_{a,a}(t)$ 的期望真实值 μ_a 。当 $t\to\infty$ 时，它们还趋于 0。因此，对于足够大的 t:

- 同一相似类中智能体 a 和 a' 的两个置信区间将以较高概率相交，因为它们至少包含一个公共点，即真实期望值 μ_a。
- 不同相似性类别的智能体 a 和 a' 的两个置信区间以很高的概率不会相交，因为它们分别以不同的真实期望值 $\mu_a \neq \mu_{a'}$ 为中心，且置信区间的宽度趋于 0。

因此，置信区间交集可作为判断所考虑的智能体 a' 是否应归属于 a 所在相似性类别的指标。此标准用于确定相似性类别。

最初，假设所有邻居都属于 a所在的相似性类别。每经过一个时刻 t，所有智能体都会先检查自己与每个邻居的置信区间。

- 如果置信区间相交，则将该邻居保留在其邻域中。
- 如果置信区间不相交，则断开连接 a 和 a' 的边，该智能体不再被视为邻居（将邻接矩阵中该连接对应的值 1 设为 0，从而得到一个新的邻接矩阵）。因此，定义智能体邻域的邻接矩阵是时间相关的 $\mathbf{A}(t)$。

C. 期望分离时间

考虑来自两个不同相似性类别、均值分别为 μ_a 和 μ_b的两个智能体 a 和 b 。不失一般性，假设 $\mu_a>\mu_b$，即 $\Delta=\mu_a-\mu_b>0$。它们在时间 t 的局部均值分别记为 $\bar{x}_{aa}(t)$ 和 $\bar{x}_{bb}(t)$。置信区间的关键情形是当

$$\beta_\delta(t)+\beta_\delta(t-1) = \bar{x}_{aa}(t)-\bar{x}_{bb}(t). \tag{17}$$

假设 $\beta_\delta(t)\approx\beta_\delta(t-1)$，并对前一方程取期望值，可得:

$$2\beta_\delta(t) = \mu_a-\mu_b. \tag{18}$$

通过对该方程求解 t，在给定 $\mu_a-\mu_b$的情况下，我们得到期望分离时间 T_{sep}。一般而言，如果类别超过两个，则期望分离时间通过求解 $2\beta_\delta(t) = \min|\mu_a-\mu_b| = \min\{\Delta_{ab}\}$ 对所有 μ_a 和μ_b得到。

III. 样本方差估计

为了计算置信区间，我们必须知道数据 $x_a(t)$的标准差 σ。在关于colME的文献中，[26], [27], 一直假设随机分布的方差（标准差）是已知的。实际上并非如此，方差必须进行估计。此外，该估计必须是在线的，并且可以包含或不包含协作。方差估计依赖于均值，而由于均值是本问题的最终目标，因此希望得到方差的均值不变估计量。方差的在线、局部与协作的均值不变估计是本文的首要目标，将在下文介绍。

为了估计方差，我们将定义一个新的随机变量

$$d_a(t) = x_a(t)-x_a(t-1). \tag{19}$$

注意，新的随机变量 $d_a(t)$对所有 $a\in\mathcal{A}$, $\mathbb{E}\{d_a(t)\}=0$均具有零均值，因为我们假设给定智能体的精确期望值不随时间变化，$\mathbb{E}\{x_a(t)\}=\mu_a$。这意味着对于 $d_a(t)$所有智能体都属于同一个相似类，且具有已知的零均值。

我们可以仅利用最初的几个数据样本，高效地估计方差（若方差相同，所有智能体都可以协作进行方差估计，因为该估计与均值无关）。

方差估计基于以下事实:

$$\mathbb{E}\{d_a^2(t)\} = \mathbb{E}\{\Big(x_a(t)-x_a(t-1)\Big)^2\} \tag{20}$$

$$= \mathbb{E}\{\Big(x_a(t)-\mu_a-(x_a(t-1)-\mu_a)\Big)^2\} \tag{21}$$

$$= \mathbb{E}\{(x_a(t)-\mu_a)^2\}+\mathbb{E}\{(x_a(t-1)-\mu_a)^2\} = 2\sigma^2, \tag{22}$$

since $\mathbb{E}\{x_a(t)-\mu_a\}\mathbb{E}\{x_a(t-1)-\mu_a\}=0$.

在实际实现中，我们将考虑两种方法：（i）一种方法基于使用局部数据，不与邻域协作。此方法稍后将与局部均值结合使用。于是，智能体 a 在时刻 t 的估计方差（以及作为其平方根的标准差）为

$$\hat{\sigma}_a^2(t) = \frac{1}{2t}\sum_{\tau=1}^{t}\big(x_a(\tau)-x_a(\tau-1)\big)^2. \tag{23}$$

由于此方法不使用协作，因此可以（并且将会）用于各相似类方差不同情况下的估计，以及用于定义样本标准差的置信区间。

（ii）另一种方法仅假设在定义的邻域内进行智能体间的常规通信。这种方法具有实际意义，因为它不需要任何额外的通信链路。

智能体只能通过已建立的初始 r 条链路，在其定义的邻域（记为 $\mathcal{N}_a$）内进行通信。智能体会根据所采用的协作方法交换局部均值或局部和。在这两种情况下，我们都可以恢复最近的数据值 $x_a(t)$。假设交换的是局部均值，则在 $t=0$ 时，数据值可表示为

$$m_{a,a}(0)=x_a(0),\ \text{and}\ m_{a,a'}(0)=x_{a'}(0),\ \text{for}\ a'\in\mathcal{N}_a. \tag{24}$$

At $t=1$ we have

$$m_{a,a}(1)=x_a(0)+x_a(1),\ m_{a,a'}(1)=x_{a'}(0)+x_{a'}(1),\ a'\in\mathcal{N}_a. \tag{25}$$

然后，我们可以使用 $m_{a,\,a'}(1)-2m_{a,\,a'}(0)$ 计算 r 差值（适用于所有 r 相邻节点），如下所示

$$x_a(1)-x_a(0),\ \text{and}\ x_{a'}(1)-x_{a'}(0),\ \text{for}\ a'\in\mathcal{N}_a. \tag{26}$$

我们可以使用这些 r 值来估计方差。不过，差值的数量可能不足以进行准确的

估计，因为 r 通常较小。因此，应针对若干个初始时刻 t重复此过程，以积累足够的值，从而获得满意的估计。例如，对于每个$t = 2,\ 3 \ldots, 10 = T_s$，我们构造 r 个差值:

$$\begin{aligned} &x_a(2) - x_a(1), \text{ and } x_{a'}(2) - x_{a'}(1), \text{ for } a' \in \mathcal{N}_a, \\ &x_a(3) - x_a(2), \text{ and } x_{a'}(3) - x_{a'}(2), \text{ for } a' \in \mathcal{N}_a, \\ &\ldots \\ &x_a(10) - x_a(9), \text{ and } x_{a'}(10) - x_{a'}(9), \text{ for } a' \in \mathcal{N}_a, \end{aligned}$$

并估计

$$\hat{\sigma}_a^2(T_s) = \frac{1}{2T_s(|\mathcal{N}_a(T_s)|+1)} \sum_{\substack{a' \in \mathcal{N}_a(0)\cup\{a\} \\ t=1,\ldots,T_s}} \big(x_{a'}(t) - x_{a'}(t-1)\big)^2. \tag{27}$$

再次注意，在方差估计过程中，如果系统中所有智能体的方差相同，且类别（数据分布） D_a 仅在均值上有所不同，那么系统中的所有 A智能体都可以协作进行方差估计。因此，将这些值全部平方后求其均值。例如，若使用 10 个时刻和 12 个邻居 a'，则会得到 120 +12 个差值（包括中心节点 a）。

在最初的时刻，直到我们估计出方差之前，我们将其设得非常大，以确保所有置信区间都相交（因为它们通常在初始阶段会相交）， $\sigma_a = 10$。

局部求和情形：如果智能体交换的是和，那么

$$x_{a'}(t) = m_{a,a'}(t) - m_{a,a'}(t-1) = \sum_{\tau=1}^{t} x_{a'}(\tau) - \sum_{\tau=1}^{t-1} x_{a'}(\tau). \tag{28}$$

局部均值情形：智能体可以交换协作得到的局部均值 $\bar{x}_{aa}(t)$。然后，对于 $t \le T_s$，我们将数据值计算为

$$x_{a'}(t) = t\bar{x}_{a,a'}(t) - (t-1)\bar{x}_{a,a'}(t-1) \tag{29}$$

其中 $\bar{x}_{a,\ a'}(t) = \mu_a(t)$ 对于 $t \le T_s$。

如果均值估计算法的定义方式使得只交换协作均值，那么为避免邻近智能体在方差估计期间对交换的均值产生影响，并确保只交换局部均值，在基于一致性的方法中，我们可以使用 $\alpha(t) = 0$ 作为 $t \le T_s$，并在之后使用给定值 $\alpha(t)$。这也可以避免在初始阶段计算均值时进行智能体协作，因为此时邻居选择极不可靠。

该过程总结于算法 1 中。

算法 1 使用智能体邻域进行标准差的局部协同估计， σ

对于 每个 $t = 0$ 到 T **执行**
  对于 每个智能体 $a \in \mathcal{A}$ **执行**
    $s \leftarrow 0, \quad \hat{\sigma}_a \leftarrow 10$ （一个足够大的值，使得置信区间相交）
    如果 $t \le T_s$ **则**（使得 $T_s|\mathcal{N}_a| \sim 100$）
      更新 s:

$$s \leftarrow s + \sum_{a' \in \mathcal{N}_a \cup \{a\}} |x_{a'}(t) - x_{a'}(t-1)|^2$$

    否则，如果 $t = T_s$ **则**
      计算 $\hat{\sigma}_a$:

$$\hat{\sigma}_a \leftarrow \sqrt{\frac{s}{2T_s(|\mathcal{N}_a(0)|+1)}}$$

    结束 if
  结束循环
  计算置信区间和协作均值
结束循环

A. 中心四阶矩和峰度估计

对于基于四阶矩的置信区间，我们必须估计中心四阶矩

$$\mu_4 = \mathbb{E}\{(\mathrm{x} - \mu_a)^4\} \tag{30}$$

同样，这将利用以下性质来完成:

$$\begin{aligned} &\mathbb{E}\{(x_a(t) - (x_a(t-1))^4\} = \mathbb{E}\{(x_a(t) - \mu_a - (x_a(t-1) - \mu_a))^4\} \\ &= \mathbb{E}\{(x_a(t) - \mu_a)^4\} + 6\mathbb{E}\{(x_a(t) - \mu_a)^2\}\mathbb{E}\{(x_a(t-1) - \mu_a)^4\} \\ &\quad + \mathbb{E}\{(x_a(t-1) - \mu_a)^4\} = \mu_4 + 6\sigma^2\sigma^2 + \mu_4 \end{aligned}$$

对于均值随时间保持不变的对称数据分布。因此，可以使用以下方法进行估计:

$$\mu_4 = \frac{1}{2}\mathbb{E}\{(x_a(t) - (x_a(t-1))^4\} - 3\sigma^4$$

$$\hat{\mu}_{4a} = \frac{1}{2T_s(|\mathcal{N}(t)|+1} \sum_{\substack{a' \in \mathcal{N}_a(0)\cup\{a\} \\ \tau=1,\ldots,t}} \big(x_{a'}(\tau) - x_{a'}(\tau-1)\big)^4 - 3\hat{\sigma}_a^4$$

$$\hat{\kappa}_a = \frac{\hat{\mu}_{4a}}{\hat{\sigma}_a^4}$$

$$= \frac{1}{2T_s(|\mathcal{N}(t)|+1)\hat{\sigma}_a^4} \sum_{\substack{a' \in \mathcal{N}_a(0)\cup\{a\} \\ \tau=1,\ldots,t}} \big(x_{a'}(\tau) - x_{a'}(\tau-1)\big)^4 - 3.$$

对于任何高斯分布，峰度都是常数， $\kappa = 3$。在统计学中，峰度也用于衡量分布偏离高斯分布的程度，采用超额峰度 $\kappa - 3$；如果分布是高斯分布，其超额峰度始终为零。峰度较高的分布具有重尾。分布的峰度小于 3，意味着与正态分布相比，其尾部更轻。

IV. 具有估计方差的协作均值估计方法综述

基本方法（colMe）在整个智能体集合内进行通信，以实现协作处理 [25]。然而，该方法的计算和内存复杂度为 A量级，因此不适用于大型系统。为了降低系统复杂度，可以将给定的智能体集合映射为一个图，并将智能体视为节点。然后，我们从一个随机图开始，为每个节点设定起始度数 r （即为每个智能体建立 r 条与邻居通信的连接），如 B-colME 和 C-colME 中所示。接下来将简要回顾所有这些方法。

A. ColME

文献 [26] 研究了涉及多个智能体的场景中的均值估计问题。群体中的每个智能体（$\mathcal{A} = \{1, 2, \ldots, A\}$）都需要确定其分布的均值 μ_a。我们假设每个分布都是 σ 次高斯分布。该框架兼具在线性和协作性:

- 每个智能体依次从其分布中接收数据。
- 每个智能体都可以与一个智能体交互以交换数据。

这些智能体的运作方式类似于网络中的各种用户设备，它们同时运行，使每个设备都能在每个时间点收集一个新样本并与另一个智能体通信。

更正式地说，我们假设智能体之间的时间同步：在每个离散时间点 $t = 0$， 1， 2，....., T，一个智能体 a:

- 从其均值为μ_a的分布中重新获取样本 $x_a(t)$，并更新其局部均值估计，表示为 $\bar{x}_{aa}(t)$。
- 随后，该智能体选择另一个智能体 a' 进行查询。作为对其查询的回应，智能体 a从智能体 a'接收本地均值 $\bar{x}^t_{a'a'}$，该均值代表其独立样本的聚合信息a' ，并将其记录在自己的

数据集中，记为 $\bar{x}_{a,a'}(t)$，并附上样本数量$n_{a,a'}(t) = t$。智能体 a' 还可以发送其本地样本的聚合总和 $m^t_{a'a'}$，该总和存储为 $m^t_{aa'}$，并同时发送 $n_{a,a'}(t) = t$、$\bar{x}_{a,a'}(t) = m_{a,a'}(t)/n_{a,a'}(t)$ 或$m_{a,a'}(t) = n_{a,a'}(t)\bar{x}_{a,a'}(t)$。

- 每个智能体都会记录从其他智能体获得的所有近期本地平均值，包括其自身的本地平均值，

$$[\bar{x}_{a,1}(t), n_{a,1}(t), \ldots, \bar{x}_{a,A}(t), n_{a,A}(t)] \tag{31}$$

这些信息用于更新时间 t时每个智能体的均值估计，

$$\mu_a(t) = \frac{\sum_{a'} m_{a,a'}(t)}{\sum_{a'} n_{a,a'}(t)} = \frac{n_{a,a'}(t)\sum_{a'} \bar{x}_{a,a'}(t)}{\sum_{a'} n_{a,a'}(t)} \tag{32}$$

发生查询时，与多臂老虎机问题中仅获取一个样本不同，查询智能体能够全面了解截至当前时间的累积观测。这种方法比传统的老虎机设置能提供更丰富的数据。

在实践中，每个智能体 a都假设其初始邻域$\mathcal{N}a(0)$包含所有其他智能体

$$\begin{aligned}\mathcal{N}_0(0) &= [1, 2, 3, \ldots, A],\\ \mathcal{N}_1(0) &= [2, 3, 4, \ldots, A, 0],\\ &\ldots,\\ \mathcal{N}_A(0) &= [0, 1, 2, \ldots, A-1].\end{aligned}$$

在时刻 $t = 0$，智能体 a 访问智能体 $a + 1$（列表中的第一个元素）、交换数据，并判断被访问的智能体是否属于同一相似类（通过检查置信区间）。如果属于，则将邻域列表向左循环轮转，保留列表中的所有智能体，并更新列表

$$[\bar{x}_{a,1}(t), n_{a,1}(t), \ldots, (\bar{x}_{a,A}(t), n_{a,A}(t)])] \tag{33}$$

中的均值和项数。如果不属于，则删除邻域列表中的第一个智能体，剩余列表保持不变（受限轮询方法），并从数据列表中删除（或置零）相应的数据。

算法2 使用 colME 进行均值估计 1：输入：δ、ϵ、分布 D_a，2：为所有节点 a初始化有序邻域 $\mathcal{N}_a$，3：为所有 a 和 a'初始化已交换的局部均值与时间列表（$\bar{x}_{aa'}(0)$ $n_{aa'}(0)$），4：对于 $t = 1$ 到 T，执行 5：并行处理所有节点 $a \in \mathcal{A}$，6：从数据分布中抽取 $x_a(t)$，7：计算局部均值 $\bar{x}_{aa}(t)$，8：使用 $\bar{x}_{aa}(t)$和 $\bar{x}_{a'a'}(t-1)$计算 a与第一个智能体$a' = \mathcal{N}_a[0]$ 的置信区间，9：如果 a 与 a' 的置信区间不相交，则 10： 11：否则 12：向左循环轮转 $\mathcal{N}_a$，13：通过$\bar{x}_{aa'}(t)$和 $n_{aa'}(t)$更新 a 中的局部均值与时间列表 $= t$。14：结束 if，15：输出：智能体 a的协作均值，计算方法为对所有 a' 的 $n_{aa'}(t)\bar{x}_{aa'}(t)$求和，再除以$n_{aa'}(t)$之和。16：结束 for，17：

$$\mathcal{N}_a[0] = [\,],\ \bar{x}_{aa'}(t) = 0, \text{ and } n_{aa'}(t) = 0$$

结束 for

B. 基于图的 *C-colME* 和 *B-colME*

在大规模系统中，让智能体查询所有其他智能体，或维护随智能体数量线性增长的记忆A，可能并不可行。在实践中，每个智能体可以转而只关注规模可控的有限智能体子集，接下来将以图作为域来实现这一点。

随机图（具有假定连接的智能体）记作 $\mathcal{G}(\mathcal{A}, \mathcal{E})$，其中 $\mathcal{A}$ 是节点/智能体的集合，$\mathcal{E}$ 是边，表示节点之间的通信链路。具体而言，初始图将在 $t = 0$ 中为随机图 $\mathcal{G}(A, r)$ 的一个随机样本，其中每个节点恰有 r 条边。r 的选择非常重要，并已在 [27],图 3 中进行了详细讨论。

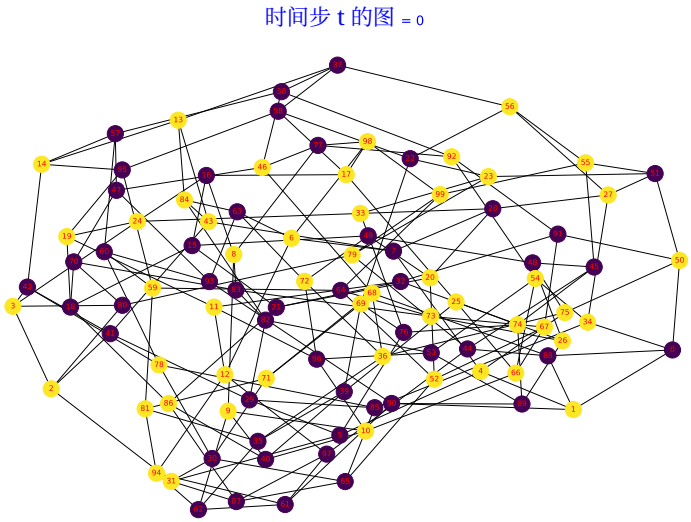


图 3：在 $t = 0$ 时，具有 $A = 100$ 个智能体/节点和 $C = 2$ 个类别的随机正则图，任意连通，包含 $r = 4$ 条通信链路，使用随机正则图 $\mathcal{G}(100, 4)$。

由于我们目前对类别一无所知，因此许多连接都是错误连接，这意味着不同类别的智能体建立了通信链路。任务如下。

- 随着时间推移消除错误连接，只保留同一类别智能体之间的连接。
- 定义一个合适的公式，利用均值和来自邻居的其他数据来改进估计。

基于共识的方法 (C-colME) 同时使用局部均值 $\bar{x}_{a,\ a}(t) = \frac{1}{t}\sum_{\tau=1}^{t} x_a(\tau)$，或以向量形式表示所有智能体的 $\mathbf{X}(t)$，以及利用估计得到的相似类别，在随机邻域 $\boldsymbol{\mu}(t)$内计算出的协作均值。请注意，如果在代码中计算局部和，那么局部均值向量为 $\mathbf{X}(t) = \mathbf{M}(t)/t$。

我们从 $\boldsymbol{\mu}(0) = \mathbf{X}(0)$ 开始，并通过组合局部均值 $\mathbf{X}(t)$（权重为 $(1 - \alpha(t))$）和先前的共识均值 $\boldsymbol{\mu}(t-1)$（权重为 $\alpha(t)$），计算所有后续的共识值，即

$$\boldsymbol{\mu}(t+1) = (1 - \alpha(t))\mathbf{X}(t) + \alpha(t)\mathbf{W}(t)\boldsymbol{\mu}(t) \tag{34}$$

其中$\mathbf{W}(t)$ 是接下来要讨论的权重矩阵。

- 为了得到无偏结果，显然必须使矩阵 $\mathbf{W}_t$ 各行之和为 $\mathbf{1}$。假设所有均值都等于 $\mathbf{1}$，则必须满足 $\mathbf{1} = (1-\alpha(t))\mathbf{1} + \alpha(t)\mathbf{W}(t)\mathbf{1}$，这意味着 $\mathbf{W}$的所有行之和都应为 $\mathbf{1}$。由于该矩阵是对称矩阵，其各列之和也为 $\mathbf{1}$，这种矩阵称为双随机矩阵。
- 同时，矩阵 $\mathbf{W}(t)$ 应发挥邻接矩阵 $\mathbf{A}(t)$（加权邻接矩阵）的作用，因为它应仅组合给定节点邻域内的 $\boldsymbol{\mu}(t-1)$ 的值，并期望该邻域在经过若干时间步后成为相似类的一部分。

- 我们假设矩阵 $\mathbf{W}(t)$ 的值对每个节点而言都是常数，并将其值取为

$$W_{i,j}(t) = W_{j,i}(t) = \frac{1}{\max\{D_i, D_j\} + 1} \text{ for } A_{i,j}(t) = 1 \tag{35}$$

以及 $W_{i,\ j}(t) = 0$ 对于 $A_{i,\ j} = 0$。值 $D(i)$ 是矩阵 $\mathbf{A}(t)$ 的度，即 $D(i) = \sum_j A_{i,\ j}(t)$

- 为了实现这一点 $\sum_j W_{i,\ j}(t) = 1$ 我们有

$$W_{i,i}(t) = 1 - \sum_j W_{i,j}(t) \tag{36}$$

算法 3 使用 C-colME 进行均值估计和方差估计

输入： 图 $G(\mathcal{A},\ \mathcal{E})$，其矩阵为 $\mathbf{A}(0)$，δ, ϵ，分布 D_a,
为所有节点 $a \in \mathcal{A}, s = 0,$ 初始化邻域 $\mathcal{N}_a(0)$
初始 $\hat{}\sigma(a) = 10$
对于 $t = 1$ 至 T **执行**
  对所有节点 $a \in \mathcal{A}$ **并行执行**
    从数据分布中抽取 $x_a(t)$
    计算局部均值 $\bar{x}_{aa}(t)$
    计算所有 a'的置信区间 $\in \mathcal{N}_a(t)$ 使用 $\bar{x}_{aa}(t)$ 和 $\bar{x}_{a'a'}(t-1)$
    如果 $t \le T_s$ **then**
      为 $a' \in$ 更新 $s \leftarrow s + (x_{a'}(t) - x_{a'}(t-1))^2$ $\mathcal{N}_a(t) \cup \{a\}$
    否则，如果 $t == T$ **then**
      Standard deviation $\hat{\sigma}(a) = \sqrt{\frac{s}{T_s(\mathcal{N}_a+1)}}$
    end if
    如果 a 和 a' 的置信区间不相交 **then**
      $\mathcal{N}_a(t) = \mathcal{N}_a(t) \setminus \{a'\}$, that is,
      Update the graph 邻接矩阵，$A_{aa'}(t) = A_{a'a}(t) = 0$
      计算双随机矩阵，$\mathbf{W}(t)$。
    结束 if
    交换局部和/协作均值/消息与邻居
    输出： 根据
    $\boldsymbol{\mu}(t+1) = (1 - \alpha(t))\mathbf{X}(t) + \alpha(t)\mathbf{W}(t)\boldsymbol{\mu}(t)$
  结束于
结束于

容易证明，C-colME 收敛到 oracle 解，[41]。在某个 $t = t_0$ 处通过置信区间检查完成剪枝，且 $\alpha(t)$ 达到 1 后，我们有：

$$\boldsymbol{\mu}(t+1) = (1 - \alpha(t))\mathbf{X}(t) + \alpha(t)\mathbf{W}(t)\boldsymbol{\mu}(t) \to \mathbf{W}^{t-t_0}(t_0)\boldsymbol{\mu}(t_0).$$

对于双随机矩阵，随着 t 增大，其幂 $\mathbf{W}^k$ 在每个图分量上趋近于全连接邻接矩阵。

消息传递算法（B-colME）方法中，智能体形成一个维度为 $d \times 2$ 的交换消息，其中d 是图中的距离（一个参数）。每个智能体首先计算截至实际时间瞬间t自身样本的总和

$$m_{1,1}^{a}(t) = \sum_{\tau=1}^{t} x_a(\tau). \tag{37}$$

所有智能体的这些总和 $a \in \mathcal{A}$ 可以视为列向量的元素，该列向量记为$\mathbf{M}(t)$。这些总和会作为消息的一部分进行交换（具体而言，它们是消息表的第一行，而总和中的样本数量 t则是消息表第一行的第二个元素）。在继续讨论接下来的行之前，我们先说明，这些消息会在相邻顶点之间交换。来自 a邻域的顶点 a'，记为 $\mathcal{N}_a$，会发送此信息

$$m_{1,1}^{a' \to a}(t) = \sum_{\tau=1}^{t} x_{a'}(\tau). \tag{38}$$

请注意，第一行的第二个元素 $m_{1\ 2}^{a'\to a},\ (t) = t$ 是样本数量。

节点 a 将来自 $\mathcal{N}_a$ 的消息中的所有值累积为 $\sum_{a' \in \mathcal{N}_a} m_{1\ 1}^{a'\to a},\ (t)$。

对于智能体，将所有邻居的总和相加，以矩阵形式可表示为整个智能体系统的形式

$$\mathbf{M_1}(t) = \mathbf{A}(t)\mathbf{M}(t) \tag{39}$$

由于图 $\mathcal{G}(\mathcal{A},\ \mathcal{E})$ 的对称邻接矩阵在某一时刻 t 记为 $\mathbf{A}(t)$，它在与每个节点的邻接智能体相对应的行中包含 1，因此将矩阵 $\mathbf{A}(t)$ 与向量 $\mathbf{M}(t)$ 相乘，恰好可以将每个节点所有邻居的总和相加；也就是说，对于每个 a，我们将其交换矩阵中所有邻居的第一元素 $m_{1\ 1}^{a'\to a},\ (t)$ 相加。

每个智能体的求和项 $\mathbf{M_1}(t)$ 的数量等于其邻居数量乘以 t。每个智能体的邻居数量等于矩阵$\mathbf{A}(t)$ 按行或按列求和的结果（矩阵为对称矩阵），即 $\mathbf{D1}(t) = [\sum_j A_{i,j}(t)\ i = 1,\ 2, \ldots,\ n]$

接下来，我们考虑消息表的第二行。这一行包含智能体 a邻居的一阶邻居的和。交换消息的第二行中的第一个元素是

$$m_{2,1}^{a' \to a}(t) = \sum_{a'' \in \mathcal{N}_{a'}, a'' \neq a} m_{1,1}^{a'' \to a'}(t-1). \tag{40}$$

这个交换消息元素给出了所考虑智能体的图中距离为 2 的智能体的和，a。对于所有 $a' \in \mathcal{N}_a$，这些和将被累加，并与项的总数（包含在消息表第二行的第二个元素中）一同用于估计协作均值。

V. 标准差和均值的双重置信区间

我们不再展示使用估计方差得到的 colME、B-colME 和 C-colME 结果（可见 [41]），而是转向一个更广泛、更复杂的情形：使用估计样本方差构造双重和三重置信区间（文献中尚未研究过）。

现有关于 colME 的研究仅考虑基于局部均值的置信区间。在**真实均值非常接近而标准差差异较为显著**的情况下，基于均值的置信区间实现分离收敛的速度会极其缓慢。不过，我们可以将标准差估计值及其置信区间纳入考虑，从而构造双重置信区间。

第III节已定义每个智能体在时刻 t 的估计标准差。仅使用所考虑的智能体 a 得到的局部值为

$$\hat{\sigma}_a(t) = \sqrt{\frac{1}{2t}\sum_{\tau=1}^{t}\big(x_a(\tau) - x_a(\tau-1)\big)^2} \tag{41}$$

$$\hat{\sigma}_a^2(t) = \frac{1}{2t}\sum_{\tau=1}^{t}\big(x_a(\tau) - x_a(\tau-1)\big)^2 = \frac{1}{2t}\sum_{\tau=1}^{t} d_a^2(\tau). \tag{42}$$

次高斯数据的标准差统计学性质相当复杂。为了估计样本方差估计量的方差，回顾一下，对于均值为零的 σ-次高斯随机

变量 $\mathbf{x}$，有 $\mathbb{E}\{\mathbf{x}^4\} \leq C\sigma^4$，其中某个常数 C 取决于具体的分布。对于高斯分布，C =3。对于存在峰度的分布，我们可以写成$\mathbb{E}\{\mathbf{x}^4\} = \kappa_x\sigma^4$。

首先，考虑 $d_a^2(t) = (x_a(t) - x_a(t-1))^2$ 的方差。利用 Var(x) = $\mathbb{E}(\mathbf{x}^2) - (\mathbb{E}(\mathbf{x}))^2$, ，对于 $d_a^2(t)$，我们可以写成

$$\mathrm{Var}(d_a^2(t)) = \mathbb{E}\{d_a^4(t)\} - (\mathbb{E}\{d_a^2(t)\})^2 = \kappa_d\sigma_d^4 - \sigma_d^4 \quad (43)$$

$$= (\kappa_d - 1)\sigma_d^4 = (\kappa_d - 1)(2\sigma^2)^2. \quad (44)$$

现在，我们将 $\hat{\sigma}_a^2(t)$ 的方差估计为

$$\mathrm{Var}(\hat{\sigma}_a^2(t)) = \mathrm{Var}(\frac{1}{2t}\sum_{\tau=1}^{t}(x_a(\tau) - x_a(\tau-1))^2) \quad (45)$$

$$= \frac{1}{4t^2}\mathrm{Var}(\sum_{\tau=1}^{t}(x_a(\tau) - x_a(\tau-1))^2) \quad (46)$$

$$= \frac{1}{4t^2}\sum_{\tau=1}^{t}(\kappa_d - 1)(2\sigma^2)^2 = \frac{(\kappa_d - 1)\sigma^4}{t}. \quad (47)$$

为确保 $d_a(t)$ $= x_a(t) - x_a(t-1)$ 的两个值在统计上相互独立，我们可以使用 $t = 1$，3，.....。然而，使用每个t 都产生了相同的统计结果。

估计样本标准差的方差涉及 Delta方法，对于 $g(x)$ = $\sqrt{x}$，其中 $g'(x)$ = $\frac{1}{2\sqrt{x}}$。使用此方法可得

$$\mathrm{Var}(g(\hat{\sigma}^2)) \approx [g'(\sigma^2)]^2\mathrm{Var}(\hat{\sigma}^2) = (\frac{1}{2\sqrt{\sigma^2}})^2\mathrm{Var}(\hat{\sigma}^2) \quad (48)$$

$$= \frac{\mathrm{Var}(\hat{\sigma}^2)}{4\sigma^2} = \frac{1}{4\sigma^2}\frac{(\kappa_d - 1)\sigma^4}{t} = \frac{(\kappa_d - 1)\sigma^2}{4t}. \quad (49)$$

因此，标准误差（估计标准差的标准差）为

$$SE_a(t) = \sqrt{\mathrm{Var}(g(\hat{\sigma}^2))} = \sqrt{\mathrm{Var}(\hat{\sigma})} = \frac{\sqrt{\kappa_d - 1}}{2}\frac{\sigma}{\sqrt{t}} \propto \frac{\sigma}{\sqrt{t}}. \quad (50)$$

接下来给出几个示例中将用到的分布的 κ_d 值。对于 $x_a(t)$ 的缩放拉德马赫（伯努利）分布，我们有 $\kappa = 1$；而 $\kappa_d = 2$，因为 $d(t)$ 是两个分布的线性和。对于由两个均匀分布组成的 $x_a(t)$，$\kappa = 2.4$ 和 $\kappa_d = 2.7$。对于高斯分布，$\kappa = \kappa_d = 3$。

由于我们在 Delta方法中使用了一阶近似，而这往往会低估方差，并且 $1/2 \leq \sqrt{\kappa_d - 1/2} \leq 1/\sqrt{2}$，对于所考虑的分布，我们决定进一步简化后续计算，对标准误差采用与局部均值相同的界限。我们通过将直方图与理论假定的标准差分布进行比较，对此进行了统计检验，见图 4。

基于前述假设和分析，我们可以对局部均值和估计标准差执行置信区间相交检验，定义如下：

$$\mathbb{I}_a(t) = [\bar{x}_{aa}(t) - \beta_\delta(t),\ \bar{x}_{aa}(t) + \beta_\delta(t)] \quad (51)$$

$$\mathbb{I}_a^\sigma(t) = [\hat{\sigma}_a(t) - \beta_\delta(t),\ \hat{\sigma}_a(t) + \beta_\delta(t)]. \quad (52)$$

形式上，我们可以将此情形视为一个二维数据问题，并使用无穷范数 [41]，这意味着如果任意一对置信区间 ($\mathbb{I}_a(t)$ $\mathbb{I}_{a'}(t)$) 或 ($\mathbb{I}_a^\sigma(t)$ $\mathbb{I}_{a'}^\sigma(t)$) 不相交，则图中智能体之间的边断开。所有其他计算（B-colME、C-colME）与文中其他部分相同。

示例：考虑具有两个相似性类别、均值非常接近（μ_a，$\mu_{a'}$） =（0.9，1.1）且标准差不同（σ_a，$\sigma_{a'}$） =（1.2，1.8）的数据。

根据第 II-C 节的结果，基于局部均值的置信区间的期望分离时间会很晚才出现，即

$$T_{sep} = 4,265. \quad (53)$$

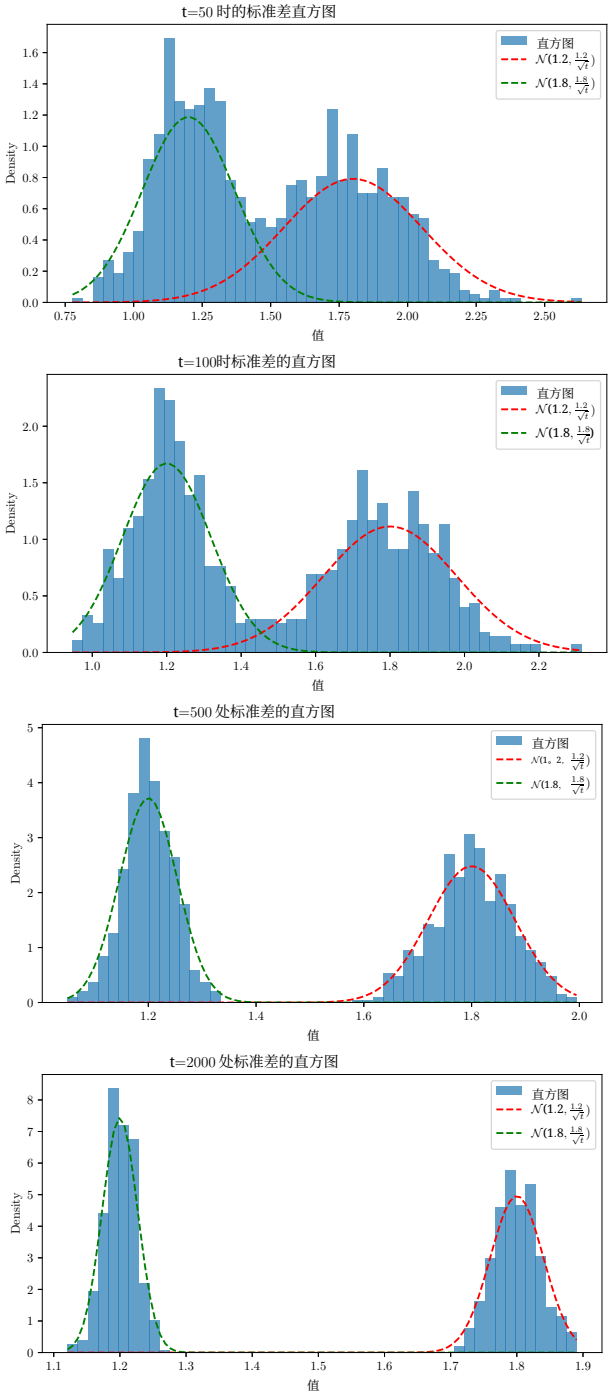


图4：不同时刻 t处估计样本标准差的直方图，随着数据数量随 t增加而收敛到所推导的理论分布（绿色与红色虚线）。对于$t = 500$ 和 $t = 2000$可见较高的准确度与一致性。

基于标准差的置信区间的期望分离时间由 $2\beta_\delta(t)$ = $\sigma_a - \sigma_{a'} = 1.8 - 1.2$ 得出：

$$T_{sep} = 416. \quad (54)$$

因此，基于标准差的分离将占主导地位并驱动该系统。

对于协作均值估计，我们使用了 B-colME，包含 $N = 1000$ 个智能体。接下来，图 4、图 5 和图 6 分别展示了标准差估计、均方误差和错误连接随时间的变化。可以看到，大约经过 $t = 1000$ 次迭代后，均方误差达到 oracle 解所在的区域。虽然我们同时检查了基于均值和基于标准差的置信区间交集，但所有错误连接的断开都是基于标准差区间完成的，因为它们的分离时间要短得多。

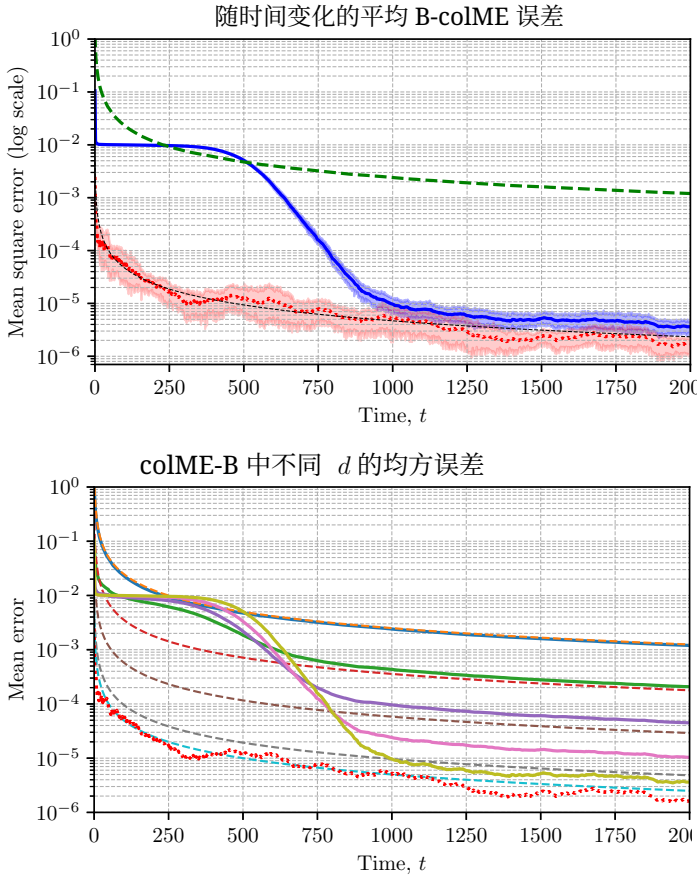


图 5：对具有相同均值的数据使用 B-colME，$N = 1000$，$\delta = 0.01$。10 次实现中各时刻 t 所有智能体的估计均方误差平均值。上图：局部估计（绿色虚线）；所提方法（蓝线）及其对应的自助法置信区间；oracle解（红色点线）及其对应的自助法置信区间。下图展示不同深度 $d = 0$，1、2、3、4 对应的均方误差值，其中 $d = 0$ 与局部估计重合，$d = 4$ 接近 oracle 解，其余值按相应顺序介于两者之间。理论线为虚线。

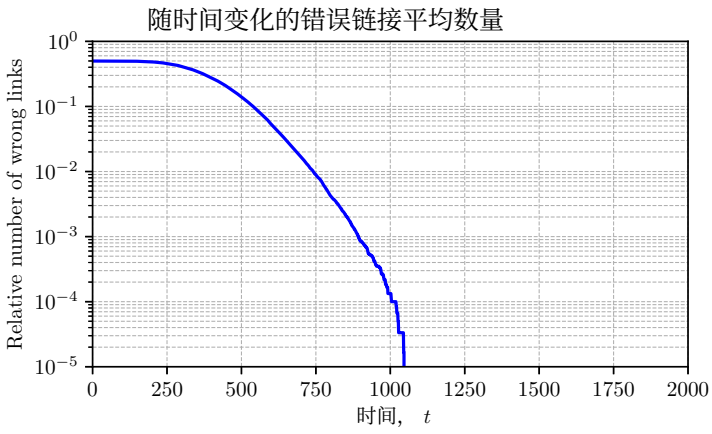


图 6：对具有相同均值的数据使用 B-colME：错误连接总数的相对值。

示例：我们还考虑了均值非常接近（0.9,1.1）、标准差不同（1.2, 1.8）的情况，使用 C-colME 而非 B-colME，并设置 $N = 1000$ 个智能体。结果如图 7 和图 8 所示。均方误差大约在 $t = 1000$时接近 oracle 值，此时错误连接的数量非常少。所有断开都是基于标准差置信区间完成的。

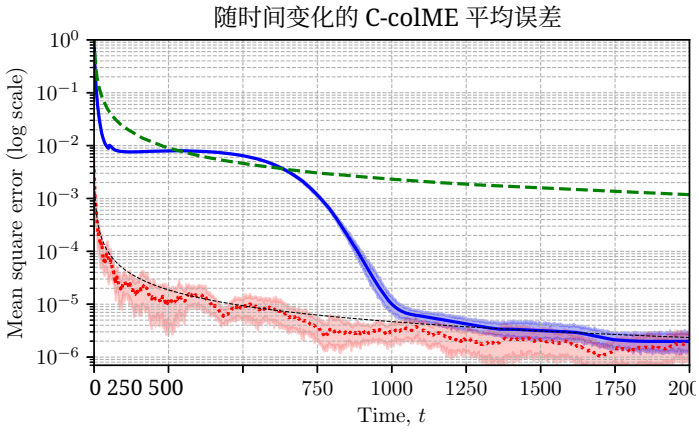


图 7：对具有相近均值的数据使用 C-colME，$N = 1000$，$\delta = 0.01$。10 次实现中所有智能体的估计均方误差总和的平均值。局部估计（绿色虚线）；所提方法（蓝线）及其对应的自助法置信区间；oracle解（红色点线）。

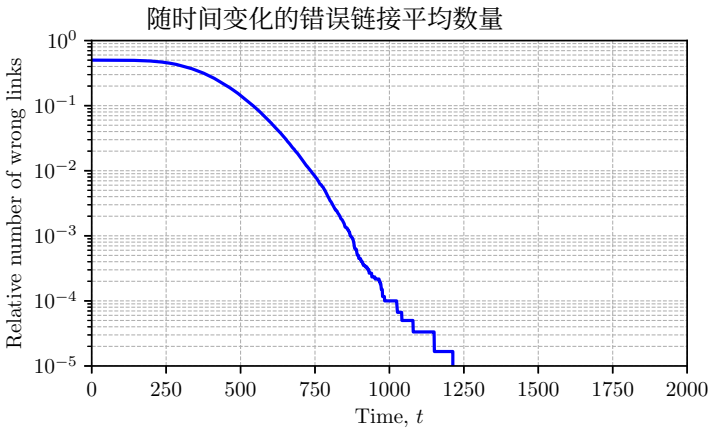


图 8：对均值相近的数据进行 C-colME：错误链接总数相对值。

VI. 峰度、标准差和均值的三重置信区间

现在考虑更复杂的情况：**真实均值和真实标准差都非常接近，但相似类的分布类型不同**。例如，一个服从高斯分布（峰度为 3），另一个服从均匀分布（峰度为 1.8）。在此，我们可以使用基于估计峰度的置信区间，

$$\hat{\kappa}_a(t) = \frac{1}{2t\hat{\sigma}_a^4(t)} \sum_{\tau=1}^{t} \big(x_a(\tau) - x_a(\tau-1)\big)^4 - 3, \qquad (55)$$

此外，还可以使用基于均值和标准差的置信区间。

次高斯数据的峰度统计学相当复杂。不过，当时刻数足够大时（这正是我们感兴趣的范围内的情况，其中 t包含数百个样本），可以证明，对于峰度的方差，我们可以使用费舍尔近似 [42]

$$\sigma^2_{a,kurtosis}(t) = \frac{24t(t-1)^2}{(t-3)(t-2)(t+3)(t+5)} \approx \frac{24}{t}. \qquad (56)$$

然后，我们可以对三个参数进行同时置信区间检验：局部均值、估计标准差和估计峰度，使用

$$\mathbb{I}_a(t) = [\bar{x}_{aa}(t) - \beta_\delta(t), \quad \bar{x}_{aa}(t) + \beta_\delta(t)] \tag{57}$$

$$\mathbb{I}_a^\sigma(t) = [\hat{\sigma}_a(t) - \beta_\delta(t), \quad \hat{\sigma}_a(t) + \beta_\delta(t)] \tag{58}$$

$$\mathbb{I}_a^\kappa(t) = [\hat{\kappa}_a(t) - z_\delta \frac{\sqrt{24}}{\sqrt{t}}, \quad \hat{\kappa}_a(t) + z_\delta \frac{\sqrt{24}}{\sqrt{t}}]. \tag{59}$$

对于峰度，我们使用了在高斯假定下定义的置信区间，但采用很高的概率 δ，因此区间更宽，因为该假定较为粗略。峰度的直方图始终与其假定分布一起绘制。

作为示例，考虑基于峰度的置信区间的期望分离时间：伯努利分布（期望峰度等于 1）与 4 个均匀随机数据之和（期望峰度为 2.7），置信概率分别为 $\delta = 0.001$ 和 $z_\delta = 3.89$。由形式为

$$2.7 - 1.0 = 2 \times 3.89\sqrt{24/t} \tag{60}$$

我们得到 $T_{sep} = 502$。

使用局部均值、局部标准差和局部峰度的问题可视为三维数据问题，其中每个智能体与

$$\mathbf{x}_a(t) = \Big(\bar{x}_{aa}(t), \quad \hat{\sigma}_a(t), \quad \hat{\kappa}_a(t)\Big). \tag{61}$$

在此框架下，我们将使用无穷范数，这意味着若任意一对置信区间：

$$(\mathbb{I}_a(t), \ \mathbb{I}_{a'}(t)) \text{ or } (\mathbb{I}_a^\sigma(t), \ \mathbb{I}_{a'}^\sigma(t)) \text{ or } (\mathbb{I}_a^\kappa(t), \ \mathbb{I}_{a'}^\kappa(t)) \tag{62}$$

不相交。

示例： 考虑均值接近（0.9、1.1）、标准差接近（1.9、2.1），但分布类型不同的情况。其中一类服从伯努利分布，峰度为 1；另一类是 4 个均匀分布随机变量之和，峰度为 2.7（峰度为 3 时接近高斯分布）。

在这种情况下，均值和标准差的置信区间会一直从 $t = 0$ 到 $t = 2000$进行检查。它们的期望分离时间均由 $2\beta_\delta(t) = \sigma_a - \sigma_{a'} = \mu_a - \mu_a$得出，即 $T_{sep} = 7,825$。由于峰度是通过数据的四次幂计算得出的，我们等待到 $t = 500$之后才进行峰度置信区间检查，此时重新连接选项处于启用状态 [41]。我们检查后得出结论：此情况下所有边断开都是基于峰度区间，因为均值和标准差太接近，其分离时间要晚得多。接下来给出 t 取不同值时的峰度直方图。

与 B-colME 的协作结果见图 10 以及图 11。

图 9 展示了不同时刻各智能体估计峰度的直方图，以及相应的理论分布。起初，这些分布相互重叠；但随着时间推移，直方图反映出底层数据分布的差异，凸显了基于峰度的类别分离。

图 10 展示了 B-colME 应用于均值和方差相近但分布不同的数据时，随时间变化的总均方误差（MSE）。上图显示固定邻域下的 MSE，下图则显示不同邻域大小下的 MSE。MSE 逐渐降低，表明随着智能体依据分布差异进行分组，估计效果有所改善。

图 11 展示了错误连接总数随时间的变化。尽管分布存在差异，类别间错误连接的数量变化趋势仍然一致，这反映了基于置信区间的边剪枝对网络结构的影响。

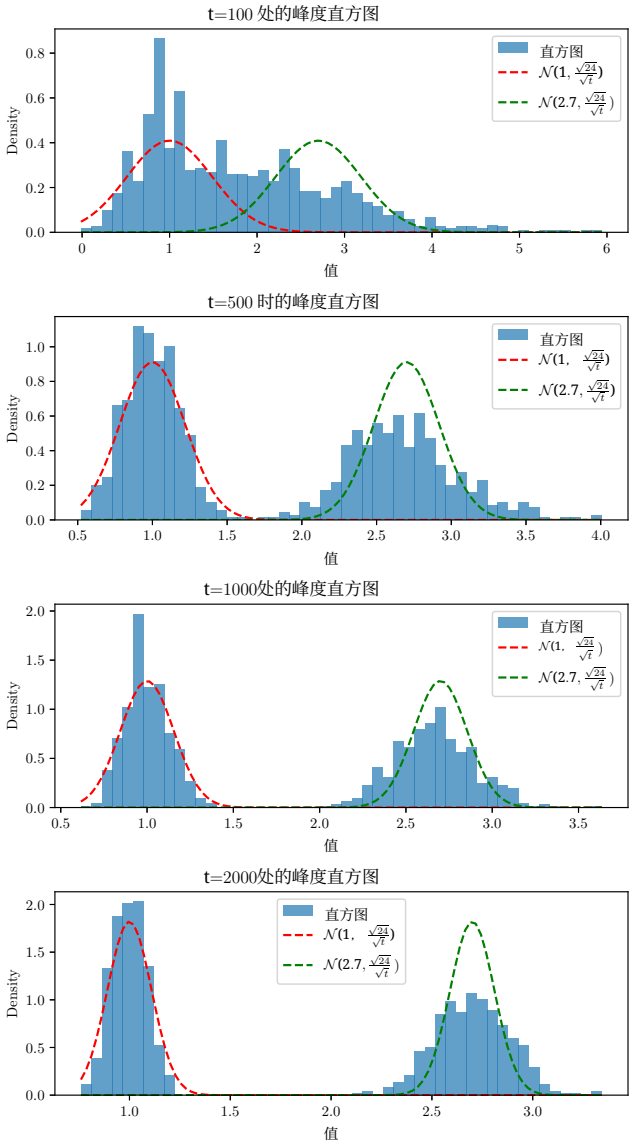


图9：不同时间点 t估计样本峰度的直方图；随着数据量增加并达到t，其逐渐收敛到推导出的理论分布（绿色和红色虚线）。

VII. 峰度、标准差和均值置信区间的多分类示例

现在考虑这样一种情况：例如，存在四个智能体类别，且无法采用单一置信区间方法。对于某些类别对，其均值十分接近，无法根据局部均值将其区分开来，但方差可能不同；而对于其他类别对，均值相距较远，但方差接近；还有一些类别对之间，只有分布类型存在显著差异。

具体而言，考虑以下类别及其对应的均值、标准差和分布类型：

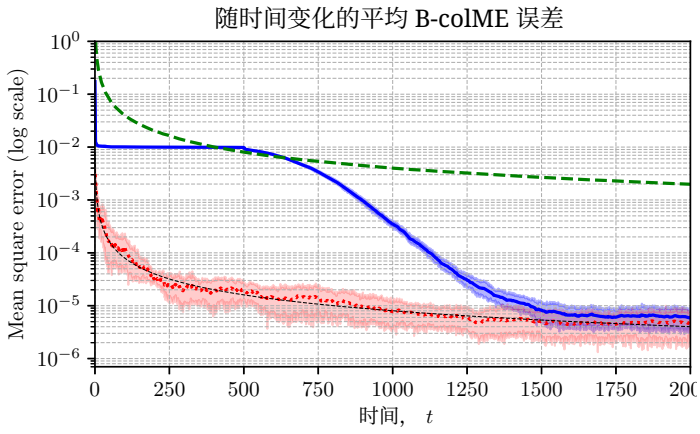
Class 1: $(\mu_1, \sigma_1) = (0.9, \ 1.8)$ sum of two uniform distributions

类别2: $(\mu_2, \ \sigma_2) = (0。1, \ 2。0)$ 两个均匀分布之和

类别3: $(\mu_3, \ \sigma_3) = (1。1, \ 1。2)$ 两个均匀分布之和

Class 4: $(\mu_4, \ \sigma_4) = (1.0, \ 1.9)$ Bernoulli distribution (63)

基于局部均值的置信区间、标准差置信区间和峰度的期望分离时间



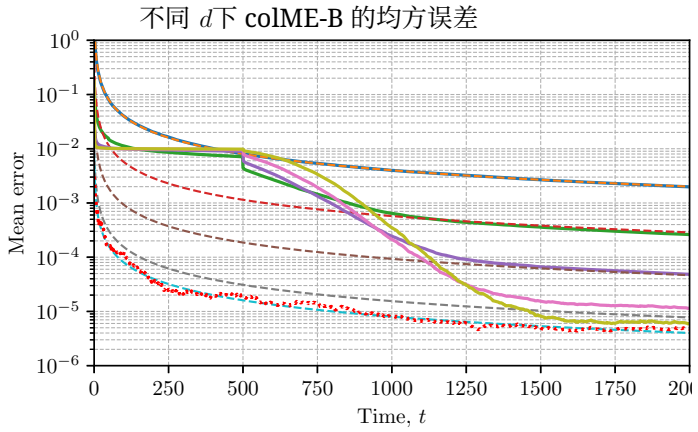


图 10：均值相同的数据上的 B-colME， $N = 1000$、$\delta = 0.01$。在时刻 t，所有智能体的估计总均方误差在 10 次实现上取平均值。上图中：局部估计（绿色虚线）；所提方法（蓝线）及其对应的自助法置信区间；oracle解（红色点线）及其对应的自助法置信区间。下图显示了不同深度 $d = 0$，1、2、3、4 下的均方误差值，其中 $d = 0$ 与局部估计重合，$d = 4$ 接近 oracle，而其他值则按相应顺序处于两者之间。理论解曲线为虚线。

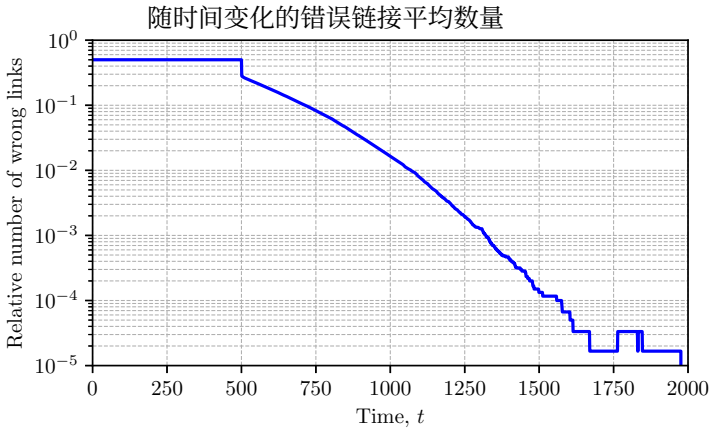


图 11：均值和方差相似但数据分布不同的数据上的 B-colME：错误连接总数相对值。由于峰度是使用数据的四次幂计算的，我们等待并在 $t = 500$ 之后应用峰度置信区间检查，且重新连接选项处于启用状态。

表 I：用彩色符号标示智能体分离时将使用的置信区间：区间 $(\mathbb{I}_a(t)\ \mathbb{I}_{a'}(t))$、$(\mathbb{I}^\sigma_a(t)\ \mathbb{I}^\sigma_{a'}(t))$ 和 $(\mathbb{I}^\kappa_a(t)\ \mathbb{I}^\kappa_{a'}(t))$ 按此顺序给出。红色圆圈表示不使用这种形式进行分离（期望分离时间非常晚），绿色表示可以使用此区间（分离时间合理）。当然，在所有情况下都会检查所有置信区间。

智能体类别（均值、标准差、峰度）区间

	1	2	3	4
1		(373, 7023, ∞)	(4264, 416, ∞)	(28552, 28552, 741)
2			(161, 258, ∞)	(306, 31890, 741)
3				(19685, 321, 741)
4				

基于的置信区间通过求解 $t = T_{sep}$以下等式来计算：

$$2\beta_\delta(t) = \mu_a - \mu_{a'} \tag{64}$$
$$2\beta_\delta(t) = \sigma_a - \sigma_{a'} \tag{65}$$
$$2z_\delta\sqrt{\frac{24}{t}} = \kappa_a - \kappa_{a'}. \tag{66}$$

它们在表 I 中以三元组形式给出，分别对应均值、标准差和峰度。

近似类别分离分析：

1）基于峰度的置信区间 的期望分离时间（对于峰度等于 2.4 的两个均匀分布之和以及峰度等于 1 的伯努利分布）由下式得到

$$2.4 - 1.0 = 2 \times 3.89\sqrt{24/t} \tag{67}$$

为 $T_{sep} = 741$。由于在局部峰度充分累积之前（例如直到 $t = 500$），我们不会允许根据该准则断开，因此该统计值可能会略高。

2) 基于标准差置信区间的期望分离时间，第一类别与第三类别之间约为 $T_{sep} = 416$，第三类别与第四类别之间为 $T_{sep} = 321$，平均值为369。

3) 基于局部均值置信区间的期望分离时间，第一与第二类别之间为$T_{sep} = 373$，第二与第三类别为 $T_{sep} = 161$，第二与第四类别之间为 $T_{sep} = 306$，平均值为280。

4) 统计平均分离时间与理论期望分离时间略有差异。在我们的示例中，基于局部均值的分离最先开始，且没有其他并行准则更早地分离边。随后，基于方差的分离会承担第二和第三类别的部分断开，从而使较高时间段内基于均值的断开数量减少，导致统计平均值略低。由于基于峰度的断开最后开始，其部分边会更早地被均值或标准差断开，从而使其平均分离时间略长。

程序记录了分离时间，其直方图如图 12 所示。对于每种置信区间，还计算并用垂直虚线标出了记录到的分离时间的均值。这些结果与上述近似理论分析一致。从表 I 可以看出，类别2最先分离（其最高期望分离时间为 373），接着类别3分离（最高期望分离时间为 416），最后类别1和类别4分离，其最高期望分离时间为 741（见图 15）。

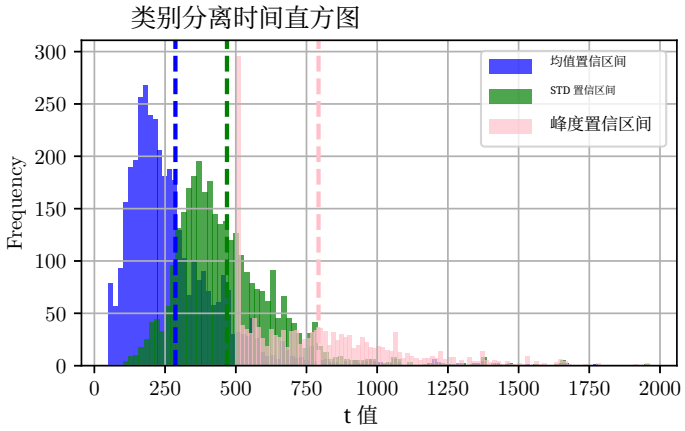


图 12：四类智能体的分离时间直方图，且无法使用单一置信区间方法。平均分离时间以竖直虚线表示。它们接近给定数据类别的理论期望分离时间。

这里我们使用了一个包含 $N = 1000$ 个智能体的 B-colME，并展示了均方误差和错误连接数随时间的变化，以及几个有代表性时刻的图。图 15、15 和 17。

我们还实现了包含智能体 $N = 1000$和均方误差的 C-colME。图 18。

加权图：最后，我们使用加权图来提高 B-colME 和 C-colME 的收敛速度，如 [41] 中所述。权重采用三个权重中最小的一个，对应于置信区间交集中最小的一个。我们计算了

$$aW_{a,a'}(t) = e^{-\left(2\frac{\bar{x}_{a,a}(t)-\bar{x}_{a',a'}(t)}{2\beta_\delta(t)}\right)^4} \text{when } A_{a,a'}(t) = 1. \quad (68)$$

然后我们计算了 $e^{-\left(2\frac{\hat{\sigma}_a(t)-\hat{\sigma}_{a'}(t)}{2\beta_\delta(t)}\right)^4}$ 并在其小于现有的 $AW_{a,a'}(t)$时，将该值用于$\overline{AW_{a,a'}(t)}$,

$$aW_{a,a'}(t) = \min\{AW_{a,a'}(t), e^{-\left(2\frac{\hat{\sigma}_a(t)-\hat{\sigma}_{a'}(t)}{2\beta_\delta(t)}\right)^4}\} \quad (69)$$

最后，对于 $A_{a,}$ a' v $=$ $(t) =$ $e^{-\left(2\frac{\hat{\kappa}_a(t)-\hat{\kappa}_{a'}(t)}{2z_\delta\sqrt{24/t}}\right)^4}$ $AW_{a\,a'}$ t 1，我们计算了 (t)，将其与现有的 $,( )$ 进行比较，并使用

$$aW_{a,a'}(t) = \min\{AW_{a,a'}(t), e^{-\left(2\frac{\hat{\kappa}_a(t)-\hat{\kappa}_{a'}(t)}{2z_\delta\sqrt{24/t}}\right)^4}\}. \quad (70)$$

结果如图19和图20所示，与图16和图18相比，收敛性明显改善。

图13展示了四个类别中智能体在不同时刻的峰度直方图。各分布最初相互重叠，但逐渐体现出四个类别之间的差异，表明可以根据分布进行类别分离。

图14展示了估计标准差随时间变化的直方图。随着智能体不断演化，各类别之间的变异性变得更加明显，凸显了方差在类别分离中的作用。最初，第一个类别与其他类别分离，而其余三个类别似乎仍合并在一起。一段时间后，$t = 2000$，可以观察到三个明显的峰值，但它们仍有显著重叠，与理论分布相符，并对应于 $t = 2500$ 时的四个类别。

图15展示了网络图的演化过程。最初，智能体之间的连接是任意的。随着时间推移，边根据均值、方差和峰度置信区间进行调整，逐步分离出四个类别，并最终将四个不同类别完全聚类，时间为 $t = 2000$。

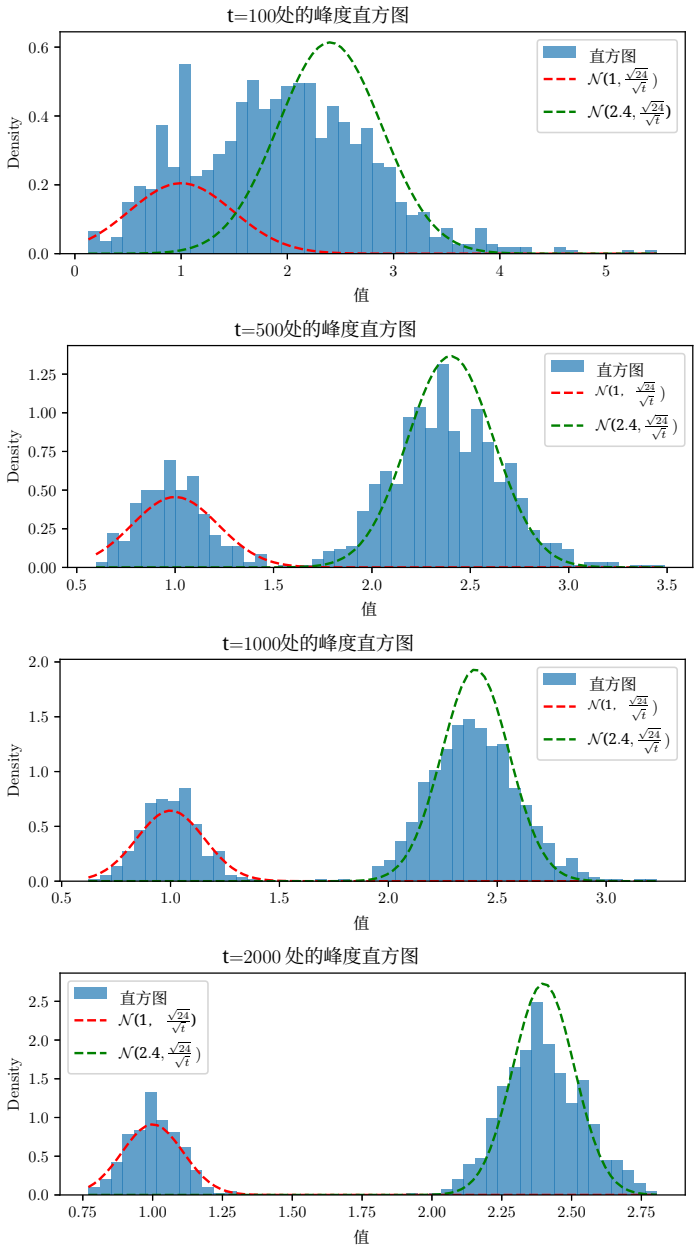


图13：不同时间时刻 t估计的样本峰度直方图。随着数据量增加，直方图趋近于推导出的理论分布（绿色和红色虚线），其中$t_。$

图 16 展示了 B-colME 在四个类别下的总均方误差随时间的变化。随着智能体正确地分离，均方误差逐渐降低，不同的邻域大小会影响收敛速度。所提算法的均方误差以蓝线表示，并附有相应的置信区间。为确认其收敛至 oracle 解，可以看到蓝线在略晚于 $t = 1800$ 时与红色虚线重合。对于 B-colME，所有方法（在不同深度下）都趋近于各自的理论解，其中最快的是在 $d = 1$ 时，约为 $t = 1000$，最慢的则是在 $d = 4$ 时。具有 $d = 4$ 的 B-colME 收敛至接近 oracle 均方误差的值，时间略早于 $t = 2000$

图17显示了随时间变化的错误连接总数，即类间连接数。随着分离过程推进，错误连接的数量逐渐减少，体现出类别区分效果的改善。

图 18 展示了 C-colME 在四个类别下的总均方误差。均方误差的下降表明，即使类别分布复杂，协作仍然成功，且估计准确。系统在 $t = 2000$ 时收敛至 oracle 解所在区域。这意味着，即使在这种复杂情况下，利用所引入置信区间的交集，四个类别也会在经过一定时间后彼此分离。

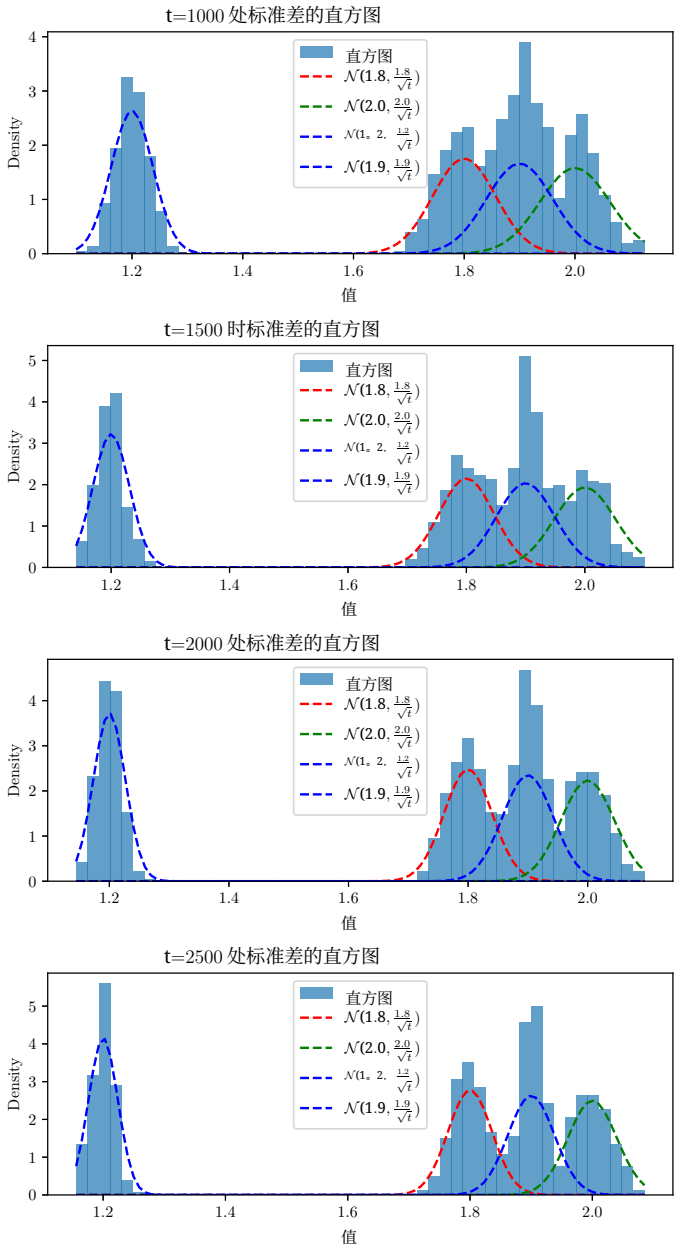


图 14：不同时间点 t 处估计的样本标准差直方图；随着数据量随 t 增加，直方图逐渐收敛至推导出的理论分布（虚线、蓝色、红色、绿色、蓝色和绿色曲线）。

图 19 和图 20 分别展示了 B-colME 和 C-colME 在加权图上的均方误差。加权通过突出最可信的连接并降低最不可信连接的权重，改善了收敛效果，从而降低了估计误差。与无权方法相比，此处收敛更快；对于 B-colME 和 C-col ME 两种情况，系统都在约 $t = 1000$ 时到达预言区域。

更多示例和实现细节请参见[41]。

VIII. 结论

。本文研究了去中心化、异质环境下的协作均值估计问题，尤其关注如何改善收敛行为。基于 ColME框架，我们考察了置信区间设计和基于图的协作如何影响智能体识别合适协作者并高效汇总信息的能力。本研究提出了一种去中心化框架，用于在线、局部地协同估计样本方差和峰度。该方法的一项关键贡献是能够

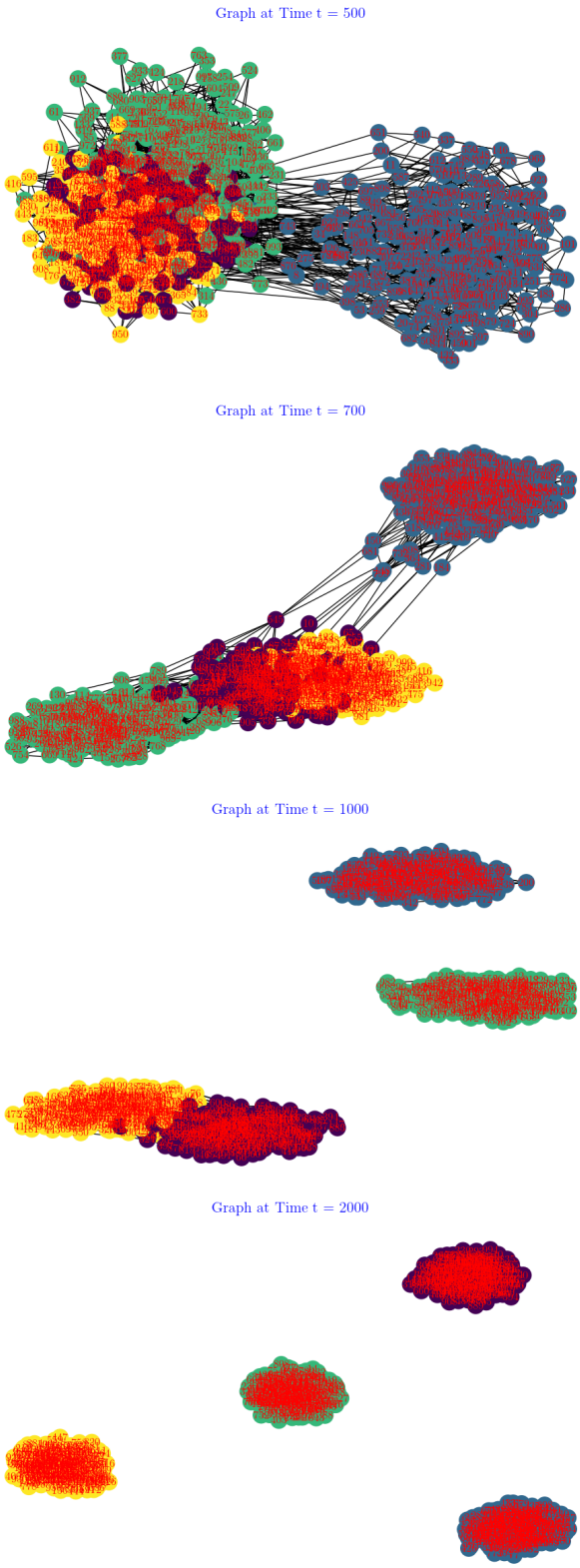
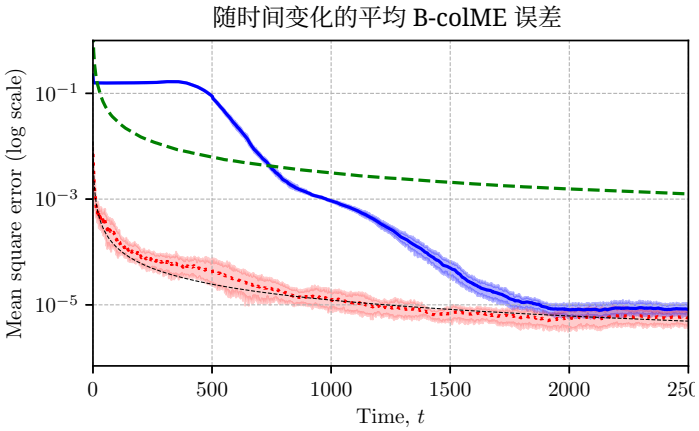


图 15：具有 $N = 1,\ 000$ 个数据的 B-colME，其中包含四类智能体，且无法使用单一置信区间方法。图中从上到下分别展示了 $t = 500$、$t = 700$、$t = 1000$ 和 $t = 2000$ 时的连接图。



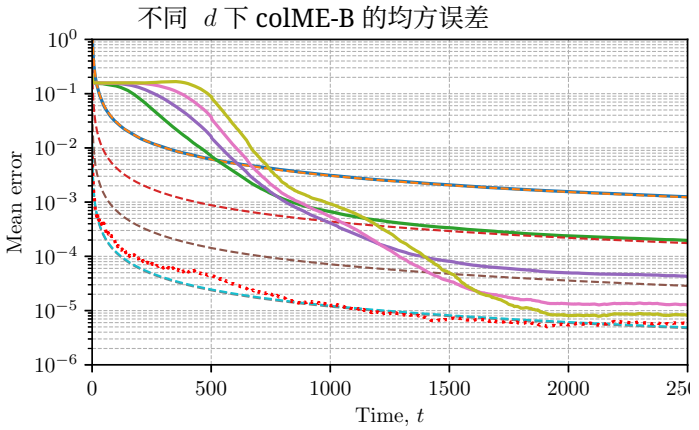


图 16：对均值相同的数据进行 B-colME，$N = 1000$，$\delta = 0.01$。10 次实现中，在时刻 t 对所有智能体的估计均方误差取平均值。上图：局部估计（绿色虚线）；所提方法（蓝线）及其对应的自助法置信区间；oracle 解（红色点线）及其对应的自助法置信区间。下图显示不同深度 $d = 0$，1，2，3，4 对应的均方误差值，其中 $d = 0$ 与局部估计重合，$d = 4$ 接近 oracle，其余值按相应顺序介于两者之间。理论解曲线为虚线。

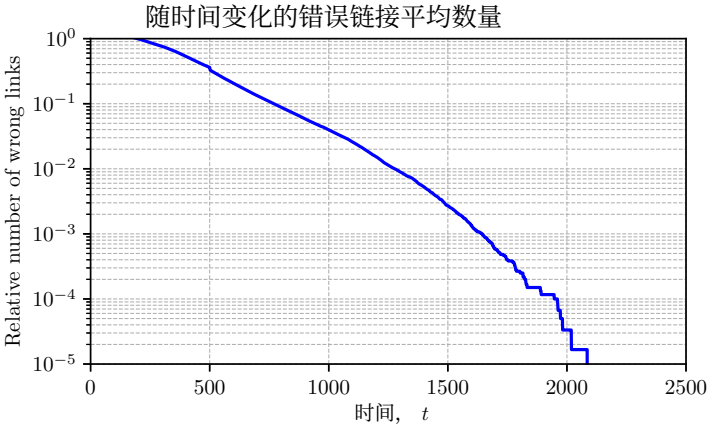


图 17：对包含 $N = 1,000$ 个数据的 B-colME，数据有四类智能体，且无法使用单一置信区间方法：错误链接总数的相对值。

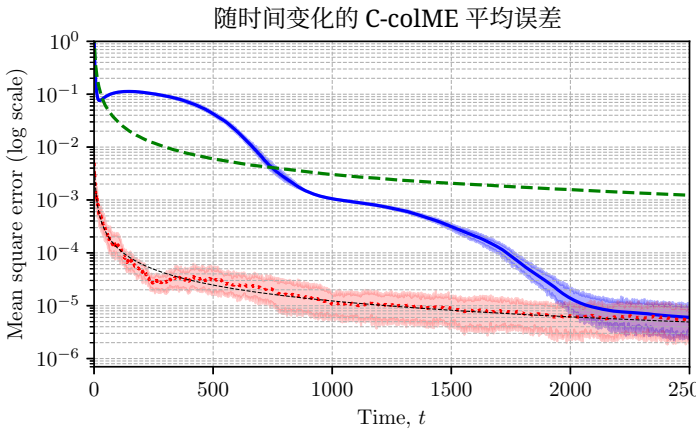
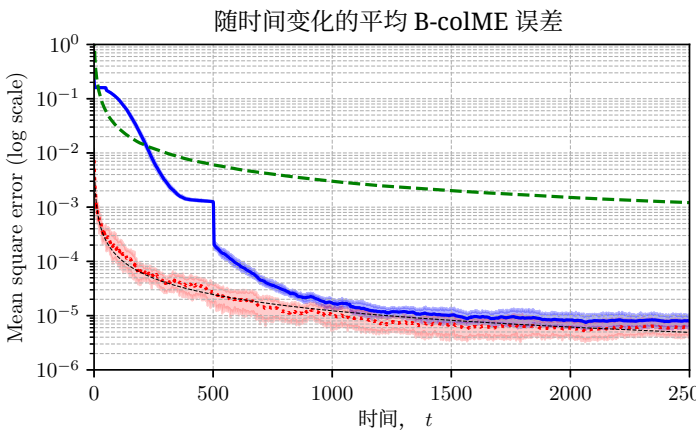


图 18：有四类智能体且无法采用单一置信区间方法时，数据量为 $N = 1,000$ 的数据的 C-colME。10 次实现中，各时刻 t 所有智能体的估计均方误差总和的平均值。



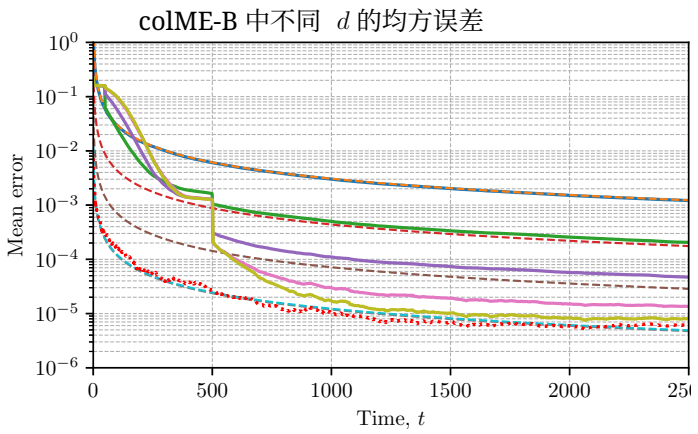


图 19：对均值相同的数据 $N = 1000, \delta = 0.01$ 进行 B-colME；有四类智能体，无法采用单一置信区间方法。10 次实现中，各时刻 t 所有智能体的估计均方误差总和的平均值。上图：局部估计（绿色虚线）；所提方法（蓝线）及其对应的自助法置信区间；oracle 解（红色点线）及其对应的自助法置信区间。下图显示不同深度的均方误差值 $d = 0, 1, 2, 3, 4$，其中 $d = 0$ 与局部估计重合，$d = 4$ 接近 oracle 解，其余值按相应顺序介于两者之间。理论线为虚线。

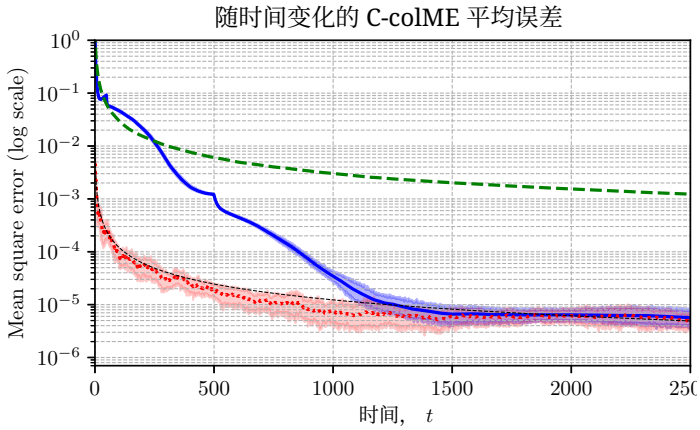


图20：数据量为 $N = 1{,}000$、采用加权图的C-colME，包含四个智能体类别，且无法使用单一置信区间方法。在10次实现中，于时刻 t对所有智能体估计的总均方误差取平均。由于峰度使用数据的四次幂计算，我们等待并在$t = 500$之后应用峰度置信区间检验，同时启用重新连接选项。

在无需访问智能体局部均值的情况下估计这些统计量，从而保持一定程度的计算独立性。我们同时处理了齐次系统以及复杂的多类别场景，在这些场景中，智能体可能具有相似的均值，但方差或分布形状各不相同。通过推导这些估计量的标准差，我们提出了一种基于样本均值、标准差和峰度构建自适应置信区间的综合方法。该方法能够在文献中此前尚未解决的困难情形下实现有效的基于图的协作——具体而言，即数据分布仅在高阶统计量上存在差异的情形。为优化收敛速度与防止错误剪枝之间的权衡，我们采用带高斯核的加权图，以降低靠近区间边界的数据点的影响。理论框架通过数值实验得到了严格验证，证实了同时使用这三种估计量对于稳健协作学习的有效性。

参考文献

[1] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings 等，“联邦学习：进展与开放问题”，机器学习基础与趋势®，第14卷，第1–2期，第1–210页，2021年。[2] A. Z. Tan, H. Yu, L. Cui 和 Q. Yang，“迈向个性化联邦学习”，*IEEE* 神经网络与学习系统汇刊，第34卷，第12期，第9587–9603页，2022年。[3] A. Ghosh, J. Chung, D. Yin 和 K. Ramchandran，“一种高效的聚类联邦学习框架”，载于神经信息处理系统进展，2020年。[4] A. Fallah, A. Mokhtari 和 A. Ozdaglar，“具有理论保证的个性化联邦学习：一种模型无关元学习方法”，载于神经信息处理系统进展，2020年。[5] F. Sattler, K.-R. M¨uller 和 W. Samek，“聚类联邦学习：隐私约束下的模型无关分布式多任务优化”，*IEEE* 神经网络与学习系统汇刊，第32卷，第8期，第3710–3722页，2021年。[6] T. Li, S. Hu, A. Beirami 和 V. Smith，“Ditto：通过个性化实现公平且稳健的联邦学习”，载于第*38*届国际机器学习会议论文集，2021年。[7] O. Marfoq, G. Neglia, A. Bellet, L. Kameni 和 R. Vidal，“分布混合下的联邦多任务学习”，载于*NeurIPS*，2021年。[8] S. Ding 和 W. Wang，“通过检测协作伙伴实现协作学习”，载于神经信息处理系统进展，第35卷，2022年，第15 629–15 641页。[9] M. Even, L. Massouli´e 和 K. Scaman，“个性化协作学习与联邦学习中的样本最优性”，载于神经信息处理系统进展，2022年。[10] E. Hillel, Z. S. Karnin, T. Koren, R. Lempel 和 O. Somekh，“多臂老虎机中的分布式探索”，载于*NIPS*，2013年。[11] C. Tao, Q. Zhang 和 Y. Zhou，“有限交互下的协作学习：多臂老虎机中分布式探索的紧界”，载于*FOCS*，2019年。[12] P.-A. Wang、A. Proutiere、K. Ariu、Y. Jedra 和 A. Russo，“多人多臂老虎机的最优算法”，发表于国际会议：关于人工智能与统计学。PMLR，2020，页码 4120–4129，2020a。[13] A. Sankararaman, A. Ganesh 和 S. Shakkottai，“多智能体多臂老虎机中的社会学习”，《ACM计算系统测量与分析会刊》，第3卷，第3期，第1–35页，2019年。[14] D. Mart´ınez-Rubio、V. Kanade和P. Rebeschini，“去中心化协作随机老虎机”，载于NIPS论文集，2019年。[15] Y. Wang、J. Hu、X. Chen和L. Wang，“分布式老虎机学习：高效通信下的近最优遗憾”，载于ICLR，2020年，2020b。[16] P. Landgren、V. Srivastava和N. E. Leonard，“多智能体多臂老虎机中的分布式协作决策”，Automatica，第125卷，文章109445，2021年。[17] U. Madhushani、A. Dubey、N. E. Leonard和A. Pentland，“向现实迈进一步：不完美通信下的协作老虎机”，载于NeurIPS，2021年。[18] E. Boursier和V. Perchet，“SIC-MMAB：多人多臂老虎机中的同步需要通信”，载于NeurIPS，2019年。[19] C. Shi、C. Shen和J. Yang，“具有个性化的联邦多臂老虎机”，载于AISTATS，2021年。[20] N. Karpov和Q. Zhang，“非独立同分布数据上有限通信下的协作最佳臂识别”，arXiv预印本arXiv:2207.08015，2022年。[21] C. R´eda、S. Vakili和E. Kaufmann，“老虎机中的近最优协作学习”，载于NeurIPS，2022年。[22] F. E. Dorner、N. Konstantinov、G. Pashaliev和M. Vechev，“在协作学习与优化中激励竞争者保持诚实”，载于神经信息处理系统进展，第36卷，2024年，第7659–7696页。[23] N. Tsoy、A. Mihalkova、T. N. Todorova和N. Konstantinov，“隐私敏感领域中联邦学习可带来可证明的互惠收益”，载于人工智能与统计国际会议，2024年。[24] F. Grimberg、M.-A. Hartley、S. P. Karimireddy和M. Jaggi，“最优模型平均：迈向个性化协作学习”，arXiv预印本arXiv:2110.12946，2021年。[25] E. Adi、A. Anwar、Z. Baig和S. Zeadally，“物联网的机器学习与数据分析”，Neural Computing and Applications，第32卷，第16205–16233页，2020年。[26] M. Asadi、A. Bellet、O.-A. Maillard和M. Tommasi，“在线个性化均值估计的协作算法”，arXiv预印本arXiv:2208.11530，2022年。[27] F. Galante、G. Neglia和E. Leonardi，“在线个性化均值估计的可扩展去中心化算法”，载于AAAI人工智能会议论文集，第39卷，第16期，2025年，第16699–16707页。[28] A. Ortega、P. Frossard、J. Kova˘cevi´c、J. M. Moura和P. Vandergheynst，“图信号处理：概述、挑战与应用”，《IEEE会刊》，第106卷，第5期，第808–828页，2018年。[29] G. Leus、A. G. Marques、J. M. Moura、A. Ortega和D. I. Shuman，“图信号处理：历史、发展、影响与展望”，IEEE信号处理杂志，第40卷，第4期，第49–60页，2023年。[30] D. K. Hammond、P. Vandergheynst和R. Gribonval，“谱图小波变换：基础理论与快速计算”，载于《图信号的顶点-频率分析》。施普林格，2018年，第141–175页。[31] I. Jestrovi´c、J. L. Coyle和E. Sejdi´c，“一种用于图上信号顶点-频率表示的快速算法”，信号处理，第131卷，第483–491页，2017年。[32] L. Stankovic、D. Mandic、M. Dakovi´c、M. Brajovic、B. Scalzo、S. Li和A. Constantinides，“图上的数据分析（第一部分）：图及图上的谱”，Foundations and Trends in Machine Learning，第13卷，第1期，第1–151页，2020年。[33] L. Stankovi´c、D. Mandic、M. Dakovi´c、B. Scalzo、M. Brajovi´c、E. Sejdi´c和A. G. Constantinides，“顶点-频率图信号处理：全面综述”，Digital Signal Processing，第107卷，文章102802，2020年。[34] J. Lerga和V. Sucic，“基于采用改进滑动成对ICI规则自适应调整的伪维格纳-维尔分布的非线性瞬时频率估计”，IEEE信号处理快报，第16卷，第11期，第953–956页，2009年。[35] J. Lerga、V. Kirin˘ci´c、D. Frankovi´c和I. ˘Stajduhar，“基于置信区间交集预处理的自适应状态估计器”，International Journal of Electrical Power & Energy Systems，第102卷，第413–420页，2018年。[36] V. Katkovnik和L. Stankovic，“使用具有变化窗长和数据驱动窗长的维格纳分布进行瞬时频率估计”，IEEE信号处理汇刊，第46卷，第9期，第2315–2325页，1998年。[37] N. A. Khan、M. Mohammadi、M. Ghafoor和S. A. Tariq，“基于卷积神经网络的时频图像增强及其在脑电信号分析中的应用”，多维系统与信号处理，第33卷，第3期，第863–877页，2022年。[38] M. Sadeghi、F. Behnia和R. Amiri，“用于从含噪测量中恢复信号的Savitzky–Golay滤波器窗选择”，IEEE仪器与测量汇刊，第69卷，第8期，第5418–5427页，2020年。[39] L. Das、J. Das和S. Nanda，“使用模糊自适应Gabor小波变换检测真核生物DNA中的外显子位置”，Genomics，第112卷，第6期，第4406–4416页，2020年。[40] O.-A. Maillard，“统计序贯决策的数学”，博士论文，里尔大学，科学与技术学院，2019年。[41] N. Stankovic，《方差感知ColME：增强去中心化在线均值估计》。硕士论文，都灵理工大学，2025年10月。[42] R. A. Fisher，“正态样本的偏离正态性的度量之分布矩”，《伦敦皇家学会会刊A辑：数学与物理学论文集》，第130卷，第812期，第16–28页，1930年。

Nikola Stankovic 出生于2001年12月28日，地点为黑山波德戈里察。他于2024年和2025年分别获得都灵理工大学（意大利都灵）的计算机工程理学学士学位和数据科学与工程理学硕士学位。目前，他是黑山大学的博士生。他的研究兴趣包括协同估计、联邦学习、人工神经网络以及工程领域中的人工智能方法。