arXiv:2601.00760v1 [stat.AP] 2 Jan 2026

# 生成模型中广义贝叶斯推断的无梯度集合变换方法

Diksha Bhandari[1]*, Sebastian Reich[1]

Abstract. 复杂生成模型中的贝叶斯推断常受制于难以处理的似然缺失，以及计算高维模拟器梯度不可行。现有用于广义贝叶斯推断的无似然方法通常依赖基于梯度的优化或重参数化，这些方法计算成本高昂，且往往不适用于黑箱模拟器。为克服这些局限，我们提出一种基于最大均值差异、用于广义贝叶斯推断的无梯度集成变换朗之万动力学方法。该方法依赖基于集合的协方差结构，而非模拟器导数，因此无需访问正向模型的梯度即可实现稳健的后验近似，使其适用于更广泛的无似然模型。该方法具有仿射不变性，计算效率高，并能稳健应对模型设定错误。通过对设定正确的混沌动力系统，以及包含污染数据的模型设定错误生成模型进行数值实验，我们证明，与现有基于梯度的方法相比，所提出的方法可达到相当或更高的精度，同时大幅降低计算成本。

## 1. Introduction

在现代机器学习和统计推断中，许多我们关注的生成模型仅通过模拟器隐式定义，而非通过显式的概率密度定义。对于这类模型，通常可以针对任意给定的参数值 $\theta \in \Theta$，轻松生成来自分布 $P_\theta$ 的独立同分布（i.i.d.）样本，但相应的似然函数无法以闭式形式获得。由于缺乏易于处理的似然函数，最大似然估计或通过后验采样进行贝叶斯推断等经典推断技术，难以甚至无法直接应用。

为应对这一挑战，越来越多的研究聚焦于无似然或基于模拟器的推断方法。这些方法以模拟样本与观测样本之间的比较取代似然评估。在这一类方法中，基于最大均值差异（MMD）的方法尤为

[1]University of Potsdam, Institute of Mathematics, Karl-Liebknecht Str. 24/25, D-14476 Potsdam, Germany*E-mail address*: diksha.bhandari@uni-potsdam.de, sebastian.reich@uni-potsdam.de.2020 *Mathematics Subject Classification.* 62F15, 65C05, 62-08.*Key words and phrases.*广义贝叶斯推断、无梯度、仿射不变性、朗之万动力学、相互作用粒子系统。

*通讯作者：DikshaBhandari。本研究由德国研究联合会（DFG）资助，项目编号：318763901-SFB1294。

强大。最大均值差异是一种基于核的统计距离，用于量化概率分布之间的差异，并且仅凭样本即可高效估计。在频率学派框架中，最大均值差异构成了最小距离估计量的基础，这类估计量具有一致性、渐近正态性，并且对模型设定错误具有鲁棒性 [4]。近期，基于 MMD 的方法也通过 MMD-Bayes 后验被引入贝叶斯框架。这种基于评分规则的形式无需进行似然评估即可实现贝叶斯推断 [6, 30]。现有研究主要依赖基于梯度的技术（如随机梯度朗之万动力学（SGLD））来探索该后验分布 [30]。然而，这些方法需要获取模拟器的梯度，或依赖重参数化技术；对于复杂的现实世界模型，这些条件往往无法满足或不切实际。

本文建立在两条互补的研究脉络之上：（i）基于损失函数的广义贝叶斯推断；（ii）用于采样的集合卡尔曼滤波技术。具体而言，仿射不变方法（如仿射不变相互作用朗之万动力学（ALDI） [11] 和集合卡尔曼采样器（EKS）[12, 10]）已作为近似后验分布的相互作用粒子系统得到广泛研究。然而，现有方法的大多数表述都局限于标准的 $\ell_2$损失函数。我们通过引入一种新颖的无梯度集成变换朗之万动力学方法，利用最大均值差异损失进行广义贝叶斯推断，从而拓展了这条研究路线。所提出的框架以 MMD-Bayes 后验为目标，同时避免使用显式似然梯度，因此能够对各种难以处理或设定错误的模型进行稳健推断。我们的方法依托由朗之万型流驱动的相互作用粒子系统，通过结合基于样本的更新与集成间的相互作用来近似目标后验。此外，尽管许多基于损失函数的推断方法对模型参数 [25], 的仿射变换较为敏感，我们提出的方法具有仿射不变性，因此在 $\theta$的重参数化下具有更高的稳定性和可解释性。

**1.1. 基于损失的广义贝叶斯推断。** 在经典贝叶斯推断中，似然函数表示观测数据所提供的有关未知参数 $\theta$ 的全部信息。它形式化了假定的数据生成过程，并作为核心组成部分，通过贝叶斯公式更新先验信念以形成后验分布，

$$\pi_{\text{post}}(\theta) \propto \pi_{\text{prior}}(\theta)\, \pi(y \mid \theta), \tag{1}$$

其中 $\pi_{\text{prior}}(\theta)$表示有关参数的先验知识，而 $\pi(y \mid \theta)$表示似然，用于描述模型输出 $G(\theta)$ 与观测数据 $y$ 之间的概率关系。考虑一族概率分布

$$\mathcal{P}_\Theta(\mathcal{X}) = \left\{ P_\theta \text{ defined on } \mathcal{X} \subset \mathbb{R}^N \,:\, \theta \in \Theta \subset \mathbb{R}^D \right\}.$$

在模型设定正确的情况下，真实数据生成分布 $P_{\text{data}}$ 属于该参数族，这意味着存在一个真实参数值 $\theta^* \in \Theta$，使得$P_{\theta^*} = P_{\text{data}}$。如果模型具有可处理密度 $\pi(\cdot, \theta)$，则可将最大似然估计量（MLE）计算为

$$\theta^*_{\text{MLE}} = \arg\max_{\theta \in \Theta} \frac{1}{N} \sum_{n=1}^{N} \log \pi(y^n \mid \theta),$$

对于独立同分布样本 $\{y^n\}_{n=1}^N \sim P_{\text{data}}$。在这种情形下，随着观测值数量增加，贝叶斯后验几乎必然收敛到以真实参数 $\theta^*$为中心的多元正态分布，其协方差与 MLE 的渐近采样协方差一致，正如伯恩斯坦-冯·米塞斯定理 [13]所述。然而在实践中，许多现实模型，尤其是基于模拟器的模型或无似然模型，都不具有闭式或可微的似然。此外，$P_{\text{data}}$ 属于参数族 $\mathcal{P}_\Theta(\mathcal{X})$这一假设往往并不成立。近似前向模型、受污染或异质数据，以及结构性模型误差都可能导致模型设定错误。在这种情况下，不存在真实参数值，贝叶斯后验渐近地集中于使其与真实数据分布之间的 Kullback–Leibler（KL）散度最小的参数，

$$\theta^*_{\text{KL}} = \arg\min_\theta \ \text{KL}(P_\theta, P_{\text{data}}).$$

虽然这提供了一个有理论依据的渐近极限，但由此得到的后验分布可能表现出较差的稳健性，尤其是在存在离群值或重尾噪声时，因为 KL 散度会对尾部差异施以重罚。

为应对这些挑战，基于损失的广义贝叶斯推断提供了一个更宽泛、更灵活的框架。它不要求使用似然函数，而是通过另一种损失函数或评分规则来定义后验分布，以量化观测数据与模型预测之间的差异。形式上，广义贝叶斯后验 [3]定义为

$$\pi_{\text{post}}(\theta \mid y) \propto \pi_{\text{prior}}(\theta)\ \exp\big(-w\, l(y,\theta)\big), \tag{2}$$

其中，$l(y,\theta)$ 是选定的损失函数，$w$ 是用于平衡先验信息与观测数据影响的缩放参数。在对 $l$ 施加适当的正则条件后，广义后验满足伯恩斯坦–冯·米塞斯结论：随着样本量增加，后验渐近正态，并以损失函数的极小值点为中心 [26, 30]。

这种表述的灵活性在于损失函数 $l$ 的选择可根据具体问题进行定制。常见示例包括能量评分、核评分 [14], 或最大均值差异等稳健评分规则，这些规则都能增强模型错设和离群值情况下的稳健性。当 $l(y,\theta) = -\log\ \pi(y \mid \theta)$ 且 $w = 1$ 时，广义后验 (2) 就退化为 (1) 中的标准贝叶斯后验，对应于 KL 散度。若选择 $w \neq 1$ 并采用相同的对数损失，则得到分数后验 [18, 2]。总之，基于损失的广义贝叶斯推断将贝叶斯范式扩展到似然不可用或难以计算的复杂情形，为现代无似然推断提供了一种有理论依据且稳健的替代方案。

**1.2. 相关工作。** 基于核的方法凭借其非参数灵活性和坚实的理论保证，在统计推断中日益受到欢迎。其中，最大均值差异已成为一种特别有力的工具，可通过再生核希尔伯特空间（RKHS）嵌入来比较概率分布 [15]。过去十年间，最大均值差异已广泛应用于无似然推断，并推动了频率学派估计量和贝叶斯形式的发展。从频率学派的角度来看，Briol 等人 [4]对最小最大均值差异估计量的统计性质进行了严格分析，确立了其一致性和渐近正态性方面的结果。他们还提出了一种

最小距离估计框架，通过最小化模拟模型输出与观测数据之间的最大均值差异来估计参数。他们的工作还提出了一种自然梯度下降型算法，以高效计算这些估计量，使最大均值差异成为一种实用且具有坚实理论依据的无需显式似然的推断方法。

基于这些思路，Cherief-Abdellatif 和 Alquier [6] 提出了一种名为 MMD-Bayes 的贝叶斯对应方法，其中贝叶斯公式中的经典似然项被基于核的伪似然所取代。这一形式使得人们能够在各种无法获得似然的生成模型中进行贝叶斯推断。MMD-Bayes 面临的主要挑战之一是，所得后验分布在解析上难以处理。为解决这一问题，Cherief-Abdellatif 和 Alquier [6] 开发了一种变分推断方法，并提出随机梯度算法来近似 MMD-Bayes 后验。近期，Pacchiardi、Khoo 和 Dutta[30] 将 MMD-Bayes 框架拓展至基于评分规则的广义贝叶斯推断这一更广泛的范畴。他们的形式采用自适应随机梯度朗之万动力学和伪马尔可夫链蒙特卡罗方案[27, 20, 1] ，从而对所得后验分布进行采样。尽管这些方法有效，但它们依赖于根据有限样本估计的核梯度，并且每次迭代都需要针对提议样本进行大量模拟，因而计算开销可能很大。 [30] 所采用的广义贝叶斯形式建立在 Bissiri、Holmes 和 Walker [3] 以及 Jewson、Smith 和 Holmes[19], 的奠基性工作之上；在这些工作中，似然被任意损失函数或评分规则所取代，用以衡量模型预测与观测数据之间的差异。在这一框架下，MMD-Bayes 方法自然对应于选择核评分作为评分规则。贝叶斯合成似然（BSL） [32] 等相关方法也可以在同一广义贝叶斯形式下加以解释。另一个互补的研究方向则利用其他核差异度，探索基于不同损失函数的贝叶斯形式。例如，Matsubara 等人 [25]研究了基于核斯坦差异的广义贝叶斯推断[7, 22], ，并表明由此得到的后验分布兼具一致性与稳健性。与我们聚焦于完全无似然模型不同，他们的方法适用于双重不可解模型，此类模型的似然只能确定到一个归一化常数。此外，Dellaporta 等人 [9] 提出了一种用于稳健贝叶斯推断的非参数方法，即引入 MMD 后验自助法，将后验自助法与基于MMD的估计量相结合，以实现对模型错设的稳健性。

人们还开发了其他一些方法，以便在无需显式似然评估的情况下进行贝叶斯推断。近似贝叶斯计算 (ABC) [8,21, 31] 直接比较观测到的汇总统计量与模拟得到的汇总统计量，但其准确性很大程度上取决于汇总统计量和距离度量的选择。较新的进展包括神经网络方法，例如神经密度比估计 [34] ，以及基于合成似然的变分方法；这两类方法都旨在将无似然推断扩展到复杂的高维模型。

1.3. **本文的概要与贡献**。 本文提出一种新颖的无梯度、仿射不变集成变换方法，用于广义贝叶斯推断。该方法在无似然设置中采用最大均值差异（MMD）作为损失函数。所提出的方法基于

基于集成的朗之万动力学原理，无需进行似然评估或计算模拟器梯度，即可高效、稳健地近似后验分布。通过在一系列生成模型上开展广泛的数值实验，我们表明，该方法不仅在模型设定正确的情形下能够实现稳健且准确的推断，在模型错设时也能如此，例如观测值受到污染或涉及混沌动力系统的情况。此外，与标准贝叶斯推断和基于梯度的朗之万动力学进行系统比较后发现，本方法的估计精度与它们相当，且在许多情况下更优，同时对模型设定错误具有更强的稳健性。

本文其余部分安排如下。第 2 节介绍最大均值差异及其经验估计量的定义，并给出最小最大均值差异估计量的形式。第 3 节阐述如何在广义贝叶斯推断框架中使用最大均值差异，并提出用于无似然推断的最大均值差异-贝叶斯形式。第 4 节详细介绍所提出的无梯度无梯度集成变换朗之万动力学方法，包括其推导、仿射不变性以及实际实现。最后，第 5 节针对各类基准问题进行经验评估，并将所提方法的性能与其他推断方法进行比较。

## 2. 框架与定义

**2.1. 符号说明。** 我们用 $\mathcal{X} \subset \mathbb{R}^N$ 表示数据空间，用 $\Theta \subset \mathbb{R}^D$ 表示参数空间。令 $\mathcal{P}(\mathcal{X})$ 表示 $\mathcal{X}$ 上的博雷尔概率测度集合。一般的概率测度记作 $Q \in \mathcal{P}(\mathcal{X})$；若其相对于勒贝格测度的密度存在，则记作 $\pi_Q(\cdot)$。

我们考虑一个由参数 $\theta \in \Theta$ 索引的生成模型族，该模型族通过一族概率测度 $\mathcal{P}_\Theta(\mathcal{X}) = \{P_\theta \in \mathcal{P}(\mathcal{X}) : \theta \in \Theta\}$. 定义。当 $P_\theta$ 关于勒贝格测度具有密度时，我们将其记作 $\pi(\cdot \mid \theta)$。

真实数据生成分布记作 $P_{\text{data}}$，其密度（若可得）记作 $\pi_{\text{data}}(\cdot)$。即使无法获得似然，我们仍可以抽取独立同分布样本 $\{y^n\}_{n=1}^N \sim P_{\text{data}}$。

**2.2. 生成模型设定。** 设 $(\mathcal{U}, \mathcal{F}, \mathbb{U})$ 为一个用作潜在随机源的概率空间。生成模型 $\mathcal{P}_\Theta(\mathcal{X})$ 定义为由 $\Theta$ 参数化的一族概率测度，使得对于任意 $\theta \in \Theta$，我们都可以从 $P_\theta$ 中抽取独立同分布样本 $\left\{x^j\right\}_{j=1}^J \subset \mathcal{X}$。这些实现值通过函数 $G_\theta : \mathcal{U} \to \mathcal{X}$, 获得，该函数称为生成器 或模拟器，它将潜在样本映射到数据空间。因此， $G_\theta$ 在 $\mathcal{X}$ 上诱导的概率分布就是推前测度 $P_\theta = G_{\theta\#}\mathbb{U}$。要从该模型生成 $n$ 个独立样本，我们首先抽取独立同分布样本 $\{u^j\}_{j=1}^J \sim \mathbb{U}$，然后将生成器应用于每个样本：

$$x^j = G_\theta(u^j), j = 1, \ldots, J.$$

这种表述使得即使似然函数 $\pi(x \mid \theta)$ 难以处理或不存在，也能够从生成模型中进行模拟。我们的目标是找出一个参数$\theta^*$ ，使相应的模型 $P_{\theta^*}$ 在某种由损失函数或统计准则定义的“接近”意义上最接近 $P_{\text{data}}$。

此外，为了构建无似然推断框架的基础，我们将讨论支撑本方法的关键数学概念。具体而言，

我们将定义最大均值差异——一种广泛用于比较概率分布的基于核的散度度量，并介绍其经验形式。我们还将介绍最小 MMD 估计量，它用于确定能够使模型生成的数据分布与观测数据分布之间的差异最小化的参数。这些工具构成了后续章节所研究的广义贝叶斯框架的基础。

**2.3. 最大均值差异。** 最大均值差异首次由 [15], 提出，是一种广受欢迎的基于核的统计度量，用于比较概率分布。在无似然推断中，MMD 是一种强有力的工具，可用于量化模型生成分布与观测数据之间的差异。

设 $\mathcal{H}_k$ 是一个再生核希尔伯特空间（RKHS），其关联的核 $k: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ 为正定对称核。设 $\mathcal{F}$ 是一个函数类$f: \mathcal{X} \to \mathbb{R}$，则任意两个概率分布之间的最大均值差异定义为它们相对于 $\mathcal{F}$的 $P, Q \in \mathcal{P}(\mathcal{X})$

$$\mathrm{MMD}_{\mathcal{F}}(P, Q) := \sup_{f \in \mathcal{F}} \left[\mathbb{E}_{x \sim P}[f(x)] - \mathbb{E}_{y \sim Q}[f(y)]\right]. \tag{3}$$

一般而言，MMD 是概率分布上的一种伪度量（即具有对称性、满足三角不等式，且 $\mathrm{MMD}_{\mathcal{F}}(P, P) = 0$ [4]）。当 $\mathcal{F}$是 $\mathcal{H}_k$中的单位球 $\mathcal{F} := \{f \in \mathcal{H}_k : \|f\|_{\mathcal{H}_k} \leq 1\}$ 时，无需显式计算上确界，因而便于计算。利用 RKHS 的再生性质（$f(x) = \langle f, k(x, \cdot)\rangle_{\mathcal{H}_k}$），MMD 可表示为两个分布的核均值嵌入之间的距离:

$$\mathrm{MMD}(P, Q) = \|\mu_P - \mu_Q\|_{\mathcal{H}_k}, \tag{4}$$

其中，$\mu_P := \mathbb{E}_{x \sim P}[k(x, \cdot)]$ 和 $\mu_Q := \mathbb{E}_{y \sim Q}[k(y, \cdot)]$ 分别是 $P$ 和 $Q$在 $\mathcal{H}_k$中的核均值嵌入。已有研究表明，在 $\mathbb{R}^D$上，高斯核和拉普拉斯核具有特征性，因此这两种核对应的最大均值差异是 $\mathbb{R}^D$ [16]上分布的度量。基于这些性质，本文将重点研究用于最大均值差异的高斯核

$$k(x, y) = \exp\left(-\frac{\|x - y\|_2^2}{2\gamma^2}\right),$$

其中，$\gamma$ 是标量带宽。平方 MMD 在应用中尤其有用，因为它具有由 [16]给出的简洁表达式

$$\begin{aligned}\mathrm{MMD}^2(P, Q) &= \|\mu_P - \mu_Q\|_{\mathcal{H}_k}^2 \\ &= \mathbb{E}_{x, x' \sim P}\left[k(x, x')\right] + \mathbb{E}_{y, y' \sim Q}\left[k(y, y')\right] - 2\,\mathbb{E}_{x \sim P, y \sim Q}\left[k(x, y)\right]. \end{aligned} \tag{5}$$

在实践中，我们无法直接访问分布 $P$ 或 $Q$，只能获取它们的样本。给定从模型模拟得到的独立同分布样本 $\{x^j\}_{j=1}^J \sim P_\theta$，以及从真实数据分布中抽取的 $\{y^n\}_{n=1}^N \sim P_{\mathrm{data}}$，我们使用以下

MMD 的无偏经验估计$^2(P_\theta,\ P_{\text{data}})$:

$$\begin{aligned}\widehat{\text{MMD}}^2(P_\theta, P_{\text{data}}) &= \mathbb{E}_{x,x'\sim P_\theta}\left[k(x,x')\right] + \mathbb{E}_{y,y'\sim P_{\text{data}}}\left[k(y,y')\right] - 2\,\mathbb{E}_{x\sim P_\theta, y\sim P_{\text{data}}}\left[k(x,y)\right] \\ &= \frac{1}{J(J-1)}\sum_{\substack{j,j'=1\\ j\neq j'}}^{J} k(x^j, x^{j'}) + \frac{1}{N(N-1)}\sum_{\substack{n,n'=1\\ n\neq n'}}^{N} k(y^n, y^{n'}) \\ &\quad - \frac{2}{JN}\sum_{j=1}^{J}\sum_{n=1}^{N} k(x^j, y^n). \end{aligned} \tag{6}$$

**2.4. 最小 MMD 估计量。** 给定参数集 $\theta \in \Theta$，最小 MMD 估计量（定义见 [4], ）是使参数化模型族 $P_\theta$ 与未知数据生成分布 $P_{\text{data}}$之间的平方 MMD 最小化的参数值:

$$\hat{\theta}_{\text{MMD}} := \arg\min_{\theta\in\Theta} \text{MMD}^2(P_\theta, P_{\text{data}}). \tag{7}$$

对于 $\{y^n\}_{n=1}^N \sim P_{\text{data}}$ 独立同分布。在下一节中，我们将重点介绍一种基于 MMD 估计量的无似然推断贝叶斯框架。

## 3. Likelihood-free inference using maximum mean discrepancy

传统的标准贝叶斯推断需要能够访问似然函数$\pi(\cdot \mid \theta)$，以更新关于参数 $\theta$ 的先验信念。然而，在许多复杂模型中，尤其是由模拟器定义的模型，似然要么不可用，要么计算上难以处理。为应对这些挑战，广义贝叶斯推断提供了一种灵活的替代方案，允许使用其他数据拟合目标（例如一般损失函数或适当评分规则 [3, 30]）来更新先验。在假设的无似然设定下，给定损失函数 $l(y,\ \theta)$，广义贝叶斯后验定义为

$$\pi_{\text{post}}(\theta \mid y) \propto \pi_{\text{prior}}(\theta)\exp\big(-w\ l(y,\theta)\big), \tag{8}$$

其中 $w > 0$ 是学习率。更广泛地说，即使没有似然函数，也可以通过使用其他差异度量进行贝叶斯更新。

我们通常可以使用生成器 $G_\theta(\cdot)$，它定义了参数化模型分布 $P_\theta$，并且可以轻松地从中进行模拟。因此，合适的$l(y,\ \theta)$选择应当是能够利用生成的样本获得估计量的选择。分布差异度量构成了一类特别有用的损失函数。其中一种度量是第 2.3 节定义的最大均值差异，它利用再生核希尔伯特空间嵌入来量化两个分布之间的差异。如 [6], 所示，MMD-Bayes 后验具有理想的性质：它是一致的，并且对模型设定错误具有稳健性，即使在标准（基于似然的）后验无法集中于真实参数的情况下也是如此。因此，即使模型存在中度设定错误，最小化 MMD 也能找出与数据生成过程最接近的模型分布，从而实现稳健推断。给定未知的真实数据生成分布 $P_{\text{data}}$ 和参数化模型分布 $P_\theta$，则对于正定核 $k$，$P_\theta$ 与 $P_{\text{data}}$ 之间的平方 MMD 如式（6）所示。将 $\text{MMD}^2$ 代入广义贝叶斯公式（8），可得以下 MMD-Bayes 后验分布

$$\pi_{\text{MMD}}(\theta \mid Y) \propto \pi_{\text{prior}}(\theta)\exp\left(-\beta\ \text{MMD}^2(P_\theta, P_{\text{data}})\right), \tag{9}$$

其中 $Y = \{y^n\}_{n=1}^N \sim P_{\text{data}}$，而 $\beta > 0$ 是温度参数，用于调整数据拟合项相对于先验的影响。该后验分布会集中于使生成分布 $P_\theta$ 在 MMD 意义下最接近观测分布 $P_{\text{data}}$ 的参数值附近。这一框架在无似然推断中特别有吸引力，因为这类推断中的模型是通过模拟器而非显式似然隐式定义的。由于 MMD-Bayes 只要求能够从 $P_\theta$ 进行模拟并计算核函数 $k$，因此在实践中适用范围广泛。请注意，MMD-Bayes 后验（9）中的 $\beta$ 表示观测值相对于先验的影响程度。对于标准贝叶斯更新，在设定正确的情形下，$\beta$ 固定为 1 [37]。对于设定错误的模型，多项研究 [17,18]探讨了使用 $\beta \neq 1$，并提出了多种选择标准[3, 24, 25]。本文采用 [3], 的方法，该方法也在 [30],中实现，计算上易于处理，且无需从精确后验中抽取样本。本文不探讨最优 $\beta$ 的选择，并指出这仍是一个活跃的研究领域，最好针对具体问题进行研究。

注 *3.1*。 [30]提出采用一种基于样本的评分规则，记作 $S(P_\theta, y)$，并将其作为广义贝叶斯框架（8）中的损失函数 $l(y, \theta)$。为进行推断，他们采用以所得评分规则后验分布为目标的马尔可夫链蒙特卡罗技术。由于无法从评分规则后验分布中精确采样，他们采用伪边际蒙特卡罗马尔可夫链（PM-MCMC）方法：对每个提出的 $\theta^\dagger$，利用来自 $P_{\theta^\dagger}$ 的模拟进行采样，以得到一个近似分布。然而，降低这一近似的误差需要在每个 MCMC 步骤中增加模拟次数，从而造成显著的计算成本。这限制了该方法在模拟成本高昂的复杂模型中的实用性。

如前所述，基于损失的广义贝叶斯推断框架并不局限于选择最大均值差异作为差异度量，而是可以扩展到采用评分规则的更一般框架。将 $l(y, \theta)$设为适当评分规则 $S(P_\theta, y)$，即可得到 [30]中介绍的评分规则后验分布 $\pi_S(\theta \mid y)$。具体而言，在（8）中代入核评分，即可恢复 MMD-Bayes 后验。关于核评分后验（因而也包括 MMD-Bayes 后验）的渐近正态性、有限样本泛化界以及对离群值稳健性的定量界的详细证明和讨论，请参阅 [29]。

在下文中，我们提出一种基于集成变换朗之万动力学的新型无梯度方法，用于从 MMD-Bayes 后验分布 (9) 中采样。该方法利用集合结构，无需依赖显式梯度信息，也无需承担通常与大量模拟预算相关的高昂计算成本，即可高效地探索后验分布。

## 4. Ensemble transform methods for likelihood-free inference

近期，无似然推断领域的研究探索了基于梯度的蒙特卡洛和变分技术，用于在似然难以计算的情形下从 MMD-Bayes 后验 $\pi_{\text{MMD}}$中采样 [30, 6]。本文基于这些进展，提出一种以过阻尼朗之万动力学为基础的集成方法。在时间 $s > 0$，其核心思想是演化一个粒子集 $\{\theta_s^m\}_{m=1}^M$，将后验近似为 $s \to \infty$；这一过程仅使用从模拟输出中获得的信息，无需梯度

即可访问正向模型 $G_{\theta_s}(\cdot)$。在时间 $s=0$，令 $\{\theta_0^m\}_{m=1}^M$ 表示从先验分布 $\pi_{\text{prior}}$中抽取的 $M$ 个参数组成的集成。我们对随机噪声输入 $\{u^j\}_{j=1}^J \sim \mathbb{U}$进行采样，并将相应的模型输出通过生成器 $G_{\theta_s^m}$ 变换为

$$x_s^{mj} = G_{\theta_s^m}(u^j), \qquad j = 1, \ldots, J,$$

对于每个 $\theta_s^m$, $m = 1$、.....、$M$。每个 $\theta_s^m \in \mathbb{R}^D$ 表示一个位于$D$维空间中的参数向量，而 $\{x_s^{mj}\}_{j=1}^J$ 则是参数化模型对应的实现值。在本文中，我们约定上标表示集合索引，或表示 $N$维向量中的第 $i$ 个元素；下标则表示时间索引。我们首先定义经验均值和协方差如下：

$$\bar{\theta}_s = \frac{1}{M}\sum_{m=1}^M \theta_s^m, \quad \bar{x}_s^j = \frac{1}{M}\sum_{m=1}^M x_s^{mj}, \tag{10}$$

$$C_s = \frac{1}{M}\sum_{m=1}^M (\theta_s^m - \bar{\theta}_s)(\theta_s^m - \bar{\theta}_s)^{\mathrm{T}}, \tag{11}$$

$$C_s^{\theta x^j} = \frac{1}{M}\sum_{m=1}^M (\theta_s^m - \bar{\theta}_s)(x_s^{mj} - \bar{x}_s^j)^{\mathrm{T}}. \tag{12}$$

在下一节中，我们将介绍一个无梯度相互作用粒子系统，该系统在无限时间范围内演化粒子 $\{\theta_s^m\}_{m=1}^M$，以获得式（9）中定义的近似后验分布 $\pi_{\text{MMD}}$ 的样本。

**4.1. 无梯度集成变换朗之万动力学**。本节介绍一种基于过阻尼朗之万动力学的相互作用粒子系统，以及将其用于无似然模型广义贝叶斯推断的无梯度近似方法。我们的方法受到仿射不变相互作用朗之万动力学（ALDI）框架 [11], 的启发；在该框架中，经验协方差结构充当朗之万动力学的预条件器。这种设计通过适应后验分布的局部几何结构来提高效率。文献 [11] 中的工作侧重于使用 $l_2$-损失函数从目标后验分布中采样，而我们将这种方法扩展到一般损失函数，特别是最大均值差异（MMD），以从 MMD-Bayes 后验中采样。对于时间 $s > 0$，定义集成为 $\{\theta_s^m\}_{m=1}^M$，相应的模型输出为 $\{x_s^{mj}\}_{j=1}^J$，其中 $m = 1$、.....、$M$。我们引入一个集合偏差矩阵

$$\theta' = (\theta^1 - \bar{\theta}_s, \theta^2 - \bar{\theta}_s, \ldots, \theta^J - \bar{\theta}_s) \tag{13}$$

因此，将经验协方差 $C_s$ 改写为紧凑形式

$$C_s = \frac{1}{J}\theta'\theta'^{\,\mathrm{T}}. \tag{14}$$

受 [11], 中工作的启发，我们选择 $C_s$ 作为预条件矩阵。因此，对于粒子 $\theta_s^m$以及势函数 $\mathcal{V}(\theta_s^m) \in \mathbb{R}^{D\times M}$，我们考虑以下类型的有限规模 IPS

$$\mathrm{d}\theta_s^m = -C_s \nabla_{\theta^m}\, \mathcal{V}(\theta_s^m)\, \mathrm{d}s + \frac{D+1}{M}(\theta_s^m - \bar{\theta}_s)\, \mathrm{d}s + \sqrt{2} C_s^{1/2}\, \mathrm{d}W_s^m, \tag{15}$$

其中，$W_s^m$表示一个 $D$维布朗运动。

注释4.1。按照 [11], ，在大集合极限（$M > D$）下，可以利用集合偏差定义协方差矩阵 $C_s$ 的广义非对称平方根，具体如下:

$$C_s^{1/2} = \frac{1}{\sqrt{J}}\theta'$$

即， $C_s = C_s^{1/2}(C_s^{1/2})^{\mathrm{T}}$. 这种广义平方根在高维场景中特别有用，因为计算 $C_s$ 的 Cholesky 分解在计算上可能代价过高。

注 4.2。 对于固定的 $D$，式 (15) 中的修正项 $\frac{D+1}{M}(\theta_s^m - \bar{\theta}_s)$ 是从所需目标测度中采样所必需的，尤其是在粒子数 $M$ 较少时（即，$M < D$）。此外，当 $M \to \infty$ 时，该修正项趋于零。关于该相互作用粒子系统及其平均场极限的详细推导，请参阅 [11, 28]。

对于基于 MMD 的无似然推断，我们选择势函数 $\mathcal{V}(\theta_s^m)$ 作为负对数后验分布 (15)

$$\mathcal{V}(\theta_s^m) = -\log \pi_{\mathrm{prior}}(\theta^m) + \beta\, \mathrm{MMD}^2(P_{\theta_s^m}, P_{\mathrm{data}}), \tag{16}$$

对于 $m = 1,$ ....., $M$。我们使用简写 $\mathrm{MMD}^2(P_{\theta_s^m},\ P_{\mathrm{data}}) := \mathrm{MMD}^2(\theta_s^m)$。因此，得到的相互作用粒子系统随机微分方程为

$$\begin{aligned} \mathrm{d}\theta_s^m =& -C_s\left(\nabla_{\theta^m} \log \pi_{\mathrm{prior}}(\theta^m) - \beta \nabla_{\theta^m} \mathrm{MMD}^2(\theta_s^m)\right) \mathrm{d}s \\ &+ \frac{D+1}{M}(\theta_s^m - \bar{\theta}_s)\,\mathrm{d}s + \sqrt{2}C_s^{1/2}\,\mathrm{d}W_s^m \end{aligned} \tag{17}$$

对于 $m = 1, \ldots, M,$ 其中 $W_s^m$表示一个 $D$ 维标准布朗运动。此演化过程需要计算 $\nabla_{\theta^m}$最大均值差异$^2(\theta_s^m)$。该表达式需要正向模型 $G_{\theta_s^m}(u^j)$ 的雅可比矩阵。对于复杂模拟器，计算该雅可比矩阵通常不切实际，甚至在计算上不可行，即使只有单个粒子轨迹 $\theta_s$ 也是如此。因此，为了克服这一挑战，我们借鉴 EKS [10], 和 ALDI 框架 [11], 提出一种基于最大均值差异、利用集合协方差统计量进行贝叶斯推断的无梯度近似。这些方法通过基于集成的协方差估计来近似参数到输出的关系，而非直接计算模型导数。如 [35,33], 所示，对于一般的非线性正向算子，我们可以通过以下方式表示参数与输出之间的互协方差

$$C_s^{\theta x^j} \approx C_s\, \nabla_{\theta^m} G_{\theta_s^m}(u^j), \tag{18}$$

对于 $m = 1, \ldots, M,$ 其中 $C_s^{\theta x^j} \in \mathbb{R}^{D\times N}$ 是参数集合与模型输出之间的样本互协方差 (12), $C_s \in \mathbb{R}^{D\times D}$ 是如 (11) 中定义的参数协方差矩阵。

备注 4.3。 近似式 (18) 称为统计线性化 ，可用于在任何贝叶斯推断采样（或优化）方法中替代导数。因此，即使是用于逆问题的单粒子采样方法，只要其依赖于正向模型的导数，也可以转化为用于采样的无导数相互作用粒子系统。此外，如 [5, 11], 所示，统计线性化假设保留了 IPS (17) 的仿射不变性，并且在线性情形下是精确的，即当 $G$是线性映射时，$C_s^{\theta x^j} = C_s\, \nabla_{\theta^m} G_{\theta_s^m}(u^j)$。

$\widehat{\text{MMD}}^2$我们引入一种对 $\nabla_{\theta^m}$最大均值差异（$\theta_s^m$）的近似。为了计算关于 a $\theta^m$ 的梯度，我们只对经验估计$\widehat{\text{MMD}}^2$最大均值差异（定义见 (6)）中依赖于 $\theta^m$的项求导。该梯度可写为

$$\nabla_{\theta^m}\widehat{\text{MMD}}^2(\theta_s^m) = \frac{2}{J(J-1)}\sum_{\substack{l,j=1\\ l\neq j}}^{J}\nabla_{\theta^m}G_{\theta_s^m}(u^j)\nabla_{x^{mj}}k(x_s^{mj},x_s^{ml}) - \frac{2}{JN}\sum_{j=1}^{J}\sum_{n=1}^{N}\nabla_{\theta^m}G_{\theta_s^m}(u^j)\nabla_{x^{mj}}k(x_s^{mj},y^n). \tag{19}$$

请注意，文献 [4]，研究了针对单个粒子$\theta$ 的梯度（19）的U统计量近似，其中梯度是对所有无序样本对取平均。将近似式（18）代入（19），得到

$$\begin{aligned}\nabla_{\theta^m}\widehat{\text{MMD}}^2(\theta_s^m) &= C_s^{-1}\left(\frac{2}{J(J-1)}\sum_{\substack{l,j=1\\ l\neq j}}^{J}C_s^{\theta x^j}\nabla_{x^{mj}}k(x_s^{mj},x_s^{ml}) - \frac{2}{JN}\sum_{j=1}^{J}\sum_{n=1}^{N}C_s^{\theta x^j}\nabla_{x^{mj}}k(x_s^{mj},y^n)\right)\\ &= C_s^{-1}\ g_s^{mj},\end{aligned} \tag{20}$$

其中

$$g_s^{mj} := \frac{2}{J(J-1)}\sum_{\substack{l,j=1\\ l\neq j}}^{J}C_s^{\theta x^j}\nabla_{x^{mj}}k(x_s^{mj},x_s^{ml}) - \frac{2}{JN}\sum_{j=1}^{J}\sum_{n=1}^{N}C_s^{\theta x^j}\nabla_{x^{mj}}k(x_s^{mj},y^n). \tag{21}$$

将得到的梯度近似（20）与朗之万动力学（17）结合，便得到无梯度集成变换朗之万动力学（GF-ETLD）的公式，其相互作用粒子系统为

$$\begin{aligned}\mathrm{d}\theta_s^m &= -\Big(C_s\nabla_{\theta^m}\log\pi_{\text{prior}}(\theta^m) - \beta\ g_s^{mj}\Big)\mathrm{d}s\\ &\quad + \frac{D+1}{M}(\theta_s^m-\bar{\theta}_s)\,\mathrm{d}s + \sqrt{2}\ C_s^{1/2}\,\mathrm{d}W_s^m,\end{aligned} \tag{22}$$

对于 $m=1$、.....、$M$，其中 $W_s^m$ 表示 $D$ 维标准布朗运动， $g_s^{mj}$ 定义见（21）。给定固定步长 $\Delta s > 0$，我们采用 Euler–Maruyama 方法对公式（22）进行离散化。这定义了一个无梯度朗之万型相互作用粒子系统，当 $s\to\infty$ 时，该系统以（9）中的后验分布为目标。按照文献 [11]，中的讨论，我们现在证明，无梯度集成变换朗之万动力学（22）在模型参数的仿射变换下保持不变。在此语境中，仿射不变性意味着，在对参数空间进行任何可逆仿射重新参数化后，粒子系统仍保持其形式和演化规律。

**引理4.4** (GF-ETLD 的仿射不变性)。 相互作用粒子系统(17)及其无梯度表述(22)具有仿射不变性。也就是说，在可逆仿射变量变换 $\theta = A\tilde{\theta}+b$ 下， $\{\tilde{\theta}_s^m\}_{m=1}^M$ 的无梯度 *IPS*

满足

$$\begin{aligned}\mathrm{d}\tilde{\theta}_s^m = &-\tilde{C}_s \nabla_{\tilde{\theta}^m} \log \tilde{\pi}_{\mathrm{prior}}(\tilde{\theta}^m)\,\mathrm{d}s \;-\; \beta\, \tilde{g}_s^{mj}\,\mathrm{d}s \\ &+ \frac{D+1}{M}\big(\tilde{\theta}_s^m - \bar{\tilde{\theta}}_s\big)\,\mathrm{d}s + \sqrt{2}\,\tilde{C}_s^{1/2}\,\mathrm{d}W_s^m, \end{aligned} \tag{23}$$

证明。令 $\theta \in \mathbb{R}^D$ 为模型参数向量，并考虑如下形式的仿射变换

$$\theta = A\tilde{\theta} + b, \tag{24}$$

对任意可逆的 $A \in \mathbb{R}^{D\times D}$ 以及任意平移 $b \in \mathbb{R}^D$。在 $\tilde{\theta}$ 坐标系下诱导的概率密度函数定义为

$$\tilde{\pi}_s(\tilde{\theta}) := |\det A|\; \pi_s(A\tilde{\theta} + b). \tag{25}$$

对于粒子 $\{\theta_s^m\}_{m=1}^M$，我们将变换后的集成 $\{\tilde{\theta}_s^m\}_{m=1}^M$ 定义为 $\tilde{\theta}_s^m := A^{-1}(\theta_s^m - b)$。因此，我们有

$$C_s = A\tilde{C}_s A^{\mathrm{T}} \quad \text{and} \quad C_s^{1/2} = A\tilde{C}_s^{1/2}.$$

此外，对于函数 $h(\theta_s^m) = h(\theta_s^m) = h(A\theta_s^m + b)$，可得

$$\nabla_{\tilde{\theta}^m}\, \tilde{h}(\tilde{\theta}_s^m) = \nabla_{\tilde{\theta}^m} h(A\tilde{\theta}_s^m + b) = A^{\mathrm{T}} \nabla_{\theta^m} h(\theta_s^m) \tag{26}$$

此外，

$$C_s^{\theta x^j} = A\, C_s^{\tilde{\theta} x^j}, \tag{27}$$

对于 $j = 1, \ldots, J$。将这些结果代入式 (22)，可得式 (23)，从而证明了所提出的 GF-ETLD 方法具有仿射不变性。该结果对于基于梯度的 IPS (17) 也成立。□

在第 5 节中，我们通过一系列数值实验展示该方法在模型设定正确时的有效性、在模型错设情形下的稳健性，以及相较于现有基于梯度的技术的计算优势。

## 5. Numerical experiments

本节将介绍一组数值实验，旨在考察所提出的无梯度集成变换朗之万动力学方法在模型设定正确和模型错设情形下进行广义贝叶斯推断的表现。我们考虑三种不同类型的生成模型，以评估该方法的性能、稳健性和计算效率。首先，我们在显式似然难以求解的随机动力系统上评估该方法。随后，我们考察涉及合成位置模型的错设情形，包括高斯分布和均匀分布模型，以评估该方法抵御数据污染的稳健性。在所有情形中，我们都将无梯度集合变换方法与现有基线方法进行比较。

**5.1. 设定正确情形。** 本节聚焦于设定正确的情形，使我们能够在理想化条件下考察该方法，并在不受模型设定错误这一额外挑战影响的情况下，单独评估 GF-ETLD 算法的性能。为此，我们选择随机版本的 Lorenz96 模型，因为它具有混沌动力学、高维度以及在大气建模中的实际相关性。本实验作为基准，用于评估该方法在复杂动力系统中准确恢复潜在参数的能力。

5.1.1. 随机 *Lorenz96* 模型。 在本实验中，我们考察 GF-ETLD 方法应用于参数化随机 Lorenz96 模型的广义贝叶斯推断。Lorenz96 系统 [23] 是数据同化和动力系统领域的经典测试平台，常用于模拟大气系统中的非线性混沌动力学，包括慢变量和快变量。此处，我们使用一种改进的随机版本，该版本由 [36], 提出，其中快变量对慢变量的影响通过一个依赖于一组参数的随机有效项来近似。此处，真实似然不可得，因此无法从精确后验中采样。

我们考虑以下耦合微分方程组所描述的 Lorenz96 模型随机变体

$$\frac{dy_k}{dt} = -y_{k-1}(y_{k-2} - y_{k+1}) - y_k + F - g_k, \quad \text{for } k = 1, \ldots, K, \tag{28}$$

其中， $y_k(t)$ 表示时刻 $t$ 第 $k$ 个变量的状态，且各索引按循环方式排列，即 $y_{K+1} = y_1$、 $y_{K+2} = y_2$ 等。随机强迫项 $g(y_k, t; \theta)$ 依赖于参数向量 $\theta = (b_0, b_1, \sigma_e, \phi)$，用于模拟未解析的快尺度过程对每个慢变量 $y_k(t)$ 的影响；将微分方程按步长 $\Delta t$ 离散化后，该项定义为

$$g(y_k, t; \theta) = b_0 + b_1 y_k + \phi\, r(t - \Delta t) + \sigma_e \sqrt{1 - \phi^2}\, \eta(t), \tag{29}$$

其中， $\theta = (b_0, b_1, \phi, \sigma_e)$ 是未知参数。残差过程 $r(t)$ 遵循以下一阶离散时间自回归过程（AR(1)）：

$$r(t) = \phi r(t - \Delta t) + \sigma_e \sqrt{1 - \phi^2}\, \eta(t), \tag{30}$$

其中， $\eta(t) \sim \mathcal{N}(0, 1)$ 是独立同分布的标准高斯噪声， $\phi$ 是自回归系数， $\sigma_e > 0$ 控制随机项的幅度。随机强迫以相同但相互独立的方式，通过共享的 AR(1) 过程进入每个慢变量方程，从而使噪声与状态动力学耦合。

我们的目标是推断参数 $\theta = (b_0, b_1, \phi, \sigma_e) \in \mathbb{R}^4$。我们固定 $K = 8$ 个变量，选择 $F = 10$，并使用四阶龙格-库塔方法在时间区间 $[0, T]$ 上对模型进行数值积分， $T = 2.5$，采用固定时间步长 $\Delta t = 1/40$，得到 $K = T/\Delta t = 33$ 个时间步，数据维度为 $33 \times 8 = 264$。我们使用真实参数 $\theta^* = (2.0, 0.8, 0.9, 1.7)$ 对随机 Lorenz96 模型进行模拟，以获得该系统的观测数据。我们的目标是使用 GF-ETLD 方法，在给定观测轨迹的情况下从 $\theta$ 上的 MMD-Bayes 后验中进行采样。我们从先验 $\pi_{\text{prior}} = \mathcal{N}([1.0, 0.0, 0.0, 1.0], (2\,\mathrm{I}, \mathrm{I}, 2\,\mathrm{II}))$ 中初始化粒子集 $\{\theta^m\}_{m=1}^M$ ，并使用 $M = 200$ 和步长 $\Delta s = 1e{-}3$ 运行 GF-ETLD 采样器。

为定量评估参数推断的准确性，我们计算四个参数 $\theta = (b_0,\ b_1,\ \phi,\ \sigma_e)$ 的后验均值与真实参数值之间的均方根误差（RMSE）。我们将所提出的 GF-ETLD 方法得到的 RMSE 与基于核评分的自适应随机梯度朗之万动力学（adSGLD）方法进行比较，该方法由 [30]提出。此方法使用核评分规则的随机梯度进行后验采样。表1汇总了两种方法在各参数上的均方根误差。结果表明，在所有参数上，GF-ETLD 的均方根误差均低于或接近 adSGLD，凸显了其在随机 Lorenz96 模型的无似然设定下的表现。然而，由于

| 方法 | $b_0$ | $b_1$ | $\phi$ | $\sigma_e$ |
|---|---|---|---|---|
| GF-ETLD | 0.0081 | 0.0011 | 0.0021 | 0.0199 |
| adSGLD | 0.0023 | 0.0016 | 0.0035 | 0.0074 |

TABLE 1. 随机 Lorenz96 模型中各参数后验均值的均方根误差。

adSGLD 方法依赖于对生成器进行梯度评估，因此计算开销更大；我们将在第 5.3 节进一步考察这一点。

这些结果凸显了 GF-ETLD 在随机动力系统上进行高效无似然推断的有效性，尤其是在无法获取模拟器梯度或计算代价过高的情形下。该实验表明，对于由含有噪声过程的微分方程刻画且似然难以处理的动力系统，GF-ETLD 作为推断工具具有良好的实际适用性。

5.2. **模型错设情形**。 现在，我们考察更具挑战性的模型错设情形，即观测数据与假定的数据生成过程并不完全吻合。这在现实场景中十分常见，因为数据污染和分布偏移屡见不鲜。我们的目标是评估 GF-ETLD 在此类条件下的鲁棒性。我们采用两个规模较小但颇具启发性的生成模型：(i) 高斯位置模型和 (ii) 均匀位置模型，以展示 GF-ETLD 在不同程度数据污染下的表现。我们将考察即使似然与真实数据分布不一致，该方法能否仍保持准确推断。

5.2.1. 作为生成模型的高斯模型。 我们评估所提出的无梯度集成变换朗之万动力学方法从 MMD-Bayes 后验分布中采样的性能。本实验聚焦于一个单变量正态位置模型，其中均值参数 $\theta$ 未知，而标准差 $\sigma = 1$ 固定。

为生成数据，我们定义正向模型 $G_\theta(u) = \theta + u$，其中 $u \sim \mathcal{N}(0,\ 1)$，因此 $x^j = G_\theta(u^j) \sim \mathcal{N}(\theta,\ 1)$，对 $j = 1,\ \ldots,\ J$ 均成立。我们通过考虑数据污染来考察模型设定错误下的稳健性。具体而言，生成 $N = 150$ 个样本，其中比例为 $1-\epsilon$ 的样本取自 $\mathcal{N}(\theta^*,\ 1)$，其真实 $\theta^* = 0$；比例为 $\epsilon$ 的样本取自 $\mathcal{N}(z,\ 1)$，其 $z = 10$。污染程度 $\epsilon \in \{0, 0.1, 0.2\}$ 控制数据中离群值的占比。

我们在这个问题上比较三种推断方法。首先，使用精确似然和高斯先验 $\theta \sim \mathcal{N}$ 计算标准贝叶斯后验（2，1）。这为理想条件下的结果提供了基准，但预计随着污染增加，效果会变差。其次，我们采用基于核评分 [30]的 adSGLD 方法。最后，我们应用 GF-ETLD 采样器，其中一组参数 $\{\theta_0^m\}_{m=1}^M$ 从先验分布中初始化，并使用第 4.1 节推导的相互作用朗之万动力学进行演化。我们使用 $M = 10$ 个粒子，步长为$\Delta s = 10^{-3}$，并将系统演化 12 个时间步。

图 1 展示了不同离群值污染水平下得到的后验密度。我们观察到，随着污染水平不断升高，GF-ETLD 方法依然保持稳健。当 $\epsilon$从 0.0 增至 0.2 时，GF-ETLD

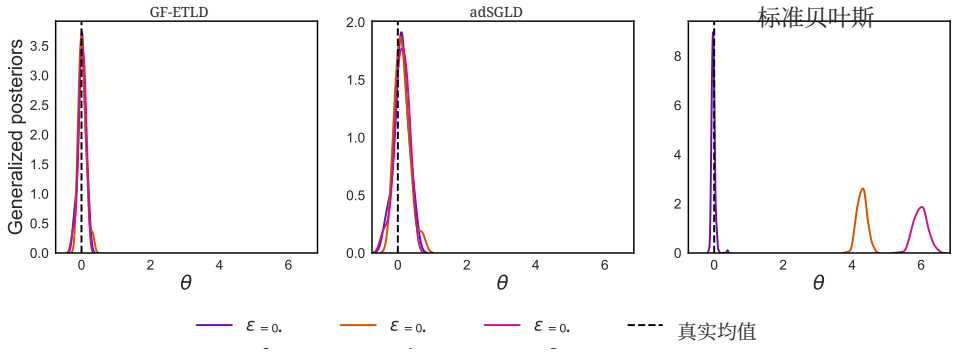


FIGURE 1. 污染水平逐渐增加时的后验密度 $\epsilon = 0.0$，0。1，0。2。垂直虚线表示真实参数值 $\theta^* = 0$。左：GF-ETLD（MMD-Bayes）；中：adSGLD（核得分）；右：标准贝叶斯后验分布。

后验分布仍持续集中于真实均值 $\theta^* = 0$附近。这表明该方法能够降低数据生成过程中离群值的影响。adSGLD 的表现与 GF-ETLD 相当，但其后验分布略为分散。此外，adSGLD 的计算成本更高，因为它每次迭代所需的前向模拟次数不断增加——这一点将在下一节讨论。相比之下，标准贝叶斯后验对污染非常敏感。即使在轻度至中度污染水平 $\epsilon = 0.1$下，0。2，后验分布也会显著偏离真实值，无法捕捉真实参数，反映出其对模型设定错误的敏感性。这些结果凸显了在稳健的无似然推断中采用基于集成的方法的价值。

图 2 进一步展示了 GF-ETLD 方法的稳健性，该图显示了后验均值与真实位置参数 $\theta^* = 0$ 之间的均方根误差（RMSE）如何随离群值比例 $\epsilon$变化。我们观察到，当污染水平不超过 $\epsilon = 0.5$时，RMSE 始终较低且几乎保持不变，这表明即使存在相当比例的离群值，GF-ETLD 后验分布仍能准确估计真实参数。只有当离群值比例超过 50% 时，估计误差才会显著增大，并在 $\epsilon$接近 1 时急剧上升。这一发现证实，GF-ETLD 方法在中等程度的污染下能够进行稳健推断。

总之，这项实验凸显了 GF-ETLD 对不断增加的离群值污染的稳健性。即使在严重污染的情况下，该方法仍能保持后验分布集中于真实参数附近，其表现优于标准贝叶斯方法，并以更低的计算成本达到与 adSGLD 相当的效果。因此，GF-ETLD 尤其适用于真实场景。

5.2.2. 均匀分布位置参数估计。 我们进一步考察 GF-ETLD 在均匀分布位置模型中的表现。与贝叶斯推断中通常采用的高斯假设相比，这一设定的偏离更大。在此实验中，我们估计均匀分布的位置参数 $\theta$

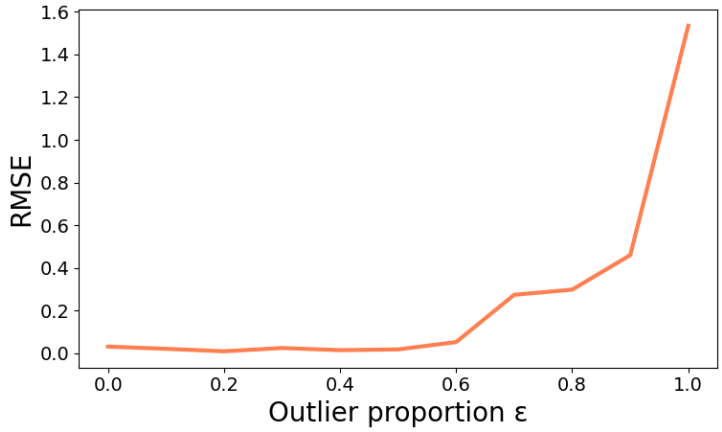


FIGURE 2. GF-ETLD 在错设高斯位置模型下获得的后验均值的均方根误差，随离群值比例 $\epsilon$ 的变化。

$\mathcal{U}([\theta - 1.0, \theta + 1.0])$，其真实值为 $\theta^* = 1$。我们生成 $N = 100$ 个独立同分布的观测值，并按照前一项实验中的错设情形，将比例为 $\epsilon \in [0, 0.4]$ 的样本替换为从高斯分布 $\mathcal{N}(10, 1)$ 中抽取的离群值。我们通过计算后验均值与真实参数 $\theta^*$ 之间的均方根误差，比较无梯度集成变换朗之万动力学方法与标准贝叶斯后验分布的表现。由于 adSGLD 方法的结果与 GF-ETLD 相当，因此此处不再单独报告。图3展示了随着 $\epsilon$ 增大，均方根误差的变化。我们发现，标准贝叶斯后验分布对少量离群值也高度敏感，均方根误差急剧上升。相比之下，GF-ETLD 方法始终保持显著较低的误差，展现出对污染的稳健性，并凸显了其在错设情形下的有效性。

该实验进一步印证了 GF-ETLD 在稳健性方面的优势。GF-ETLD 在高斯和均匀分布错设情形下均表现稳定，证明了其对一大类似然自由推断问题的普适性。

5.3. **计算时间**。 最后，我们评估 GF-ETLD 方法与 adSGLD 和标准贝叶斯推断相比的计算效率。本节将表明，GF-ETLD 不仅在推断准确度和稳健性方面具有竞争力或更优，而且计算成本也更低。在高维模型或基于模拟的模型中，梯度评估或较大的集成规模会带来沉重的计算负担，因此这一特性至关重要。

我们提供了多次独立运行中记录的各项实验计算时间。高斯位置估计和均匀分布位置估计实验重复了 $L = 10$ 次，随机 Lorenz96 实验重复了 5 次。每次运行都会生成一个新数据集，并为 $\theta$ 获得 $L$ 个后验样本。表 2 记录了各方法生成 $L$ 个样本的平均耗时。

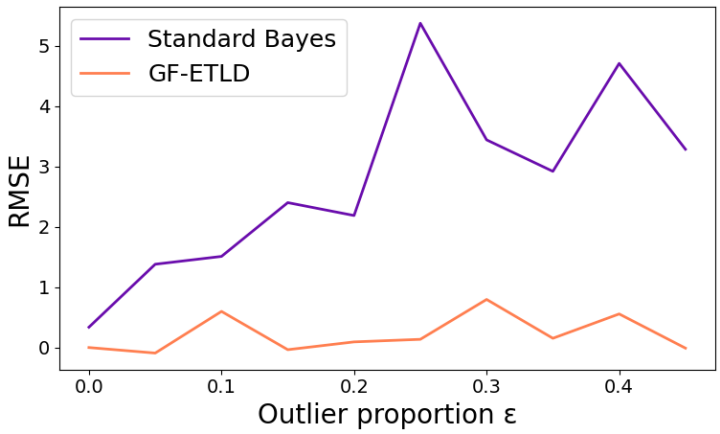


FIGURE 3. 随着离群值污染比例的增加，GF-ETLD 和标准贝叶斯方法对均匀位置参数的后验估计的均方根误差 $\epsilon$

| **方法** | **高斯** | **均匀随机Lorenz96** | |
|---|---|---|---|
| GF-ETLD | 1.244 | 0.016 | 1.515 |
| adSGLD | 8.586 | 6.169 | 55.416 |

TABLE 2. 平均时钟时间（秒）。

表 2 的结果凸显了所提出的 GF-ETLD 方法的计算效率。在全部三种实验设置中，GF-ETLD 生成后验样本所需的时间始终显著少于 adSGLD，速度提升达一至两个数量级。因此，GF-ETLD 为无似然问题提供了一种实用且可扩展的推断工具，尤其适用于无法获得模拟器梯度、计算模拟器梯度代价高昂，或必须尽量减少前向模拟次数的情形。

## 6. CONCLUSIONS

本文提出了一种基于最大均值差异的无梯度集成变换朗之万动力学（GF-ETLD）算法，用于广义贝叶斯推断。该方法利用基于集合的协方差结构，无需模拟器梯度或显式似然评估即可进行无似然推断。针对随机Lorenz96系统以及受污染的高斯模型和均匀位置模型开展的实证研究表明，GF-ETLD 能够恢复准确的后验分布，同时对离群值和模型设定错误保持稳健。在这些基准测试中，它以显著更低的计算成本，达到了与基于梯度的采样器相当或更高的准确度。这些

研究结果表明，GF-ETLD 是一种实用且可扩展的工具，可用于高维生成模型的无似然推断，并兼顾统计稳健性与计算效率。未来研究可进一步建立形式化的理论保证，探索通过子采样、局部化或丢弃策略提升可扩展性，并将该框架扩展至更多类型的真实世界生成模型，包括目前仍鲜有研究的多模态目标分布。

## References


[1] Christophe Andrieu 和 Gareth O. Roberts. “高效蒙特卡洛计算的伪边际方法”。载于：《统计年鉴》37.2（2009年4月），第697–725页。 DOI: 10.1214/07-AOS574。 URL: https://mathsci net.ams.org/mathscinet/relay-station?mr=2502648.[2] Anirban Bhattacharya、Debdeep Pati 和 Yun Yang. “贝叶斯分数后验”。载于：《统计年鉴》47.1（2019年2月）。出版方：数理统计学会，第39–66页。 ISSN: 0090-5364, 2168-8966。 DOI: 10.1 214/18-AOS1712。 URL: https://projecteuclid.org/journals/annals-of-statistics/volume-47/issue-1/Bayesian-fractional-posteriors/10.1214/18-AOS1712.full。[3] P. G. Bissiri、C. C. Holmes 和 S. G. Walker. “更新信念分布的一般框架”。载于：《皇家统计学会杂志》*B*辑：统计方法学78.5（2016年11月），第1103–1130页。 ISSN: 1369- 7412。 DOI: 10.1111/rssb.12158。 URL: https://www.ncbi.nlm.nih.gov/pm c/articles/PMC5082587/.[4] Francois-Xavier Briol 等人。基于最大均值差异的生成模型统计推断。arXiv:1906.05944 [cs, math, stat]。2019年6月。URL: http://arxiv.org/abs/1906.05944（访问于2024年6月20日）。[5] Edoardo Calvello、Sebastian Reich 和 Andrew M. Stuart. “集合卡尔曼方法：平均场视角”。英文。载于：《数值数学报》34（2025年7月），第123–291页。 ISSN: 0962-4929, 1474-0508。 DOI: 10.1017/S0962492924000060。URL: https://www.cambridge.org/core/product/identifier/S0962492924000060/type/journal_article。[6] Badr-Eddine Cherief-Abdellatif 和 Pierre Alquier. “MMD-Bayes：基于最大均值差异的稳健贝叶斯估计”。英文。载于：近似贝叶斯推断进展研讨会论文集。ISSN：2640-3498。PMLR，2020年2月，第1–21页。 URL: https://proceedings.mlr.p ress/v118/cherief-abdellatif20a.html。[7] Kacper Chwialkowski、Heiko Strathmann 和 Arthur Gretton. “核拟合优度检验”。英文。载于：第*33*届机器学习国际会议论文集。ISSN：1938-7228。PMLR，2016年6月，第2606–2615页。 URL: https://proceedings.mlr.press/v48/chwialkowski16.html。[8] Pierre Del Moral、Arnaud Doucet 和 Ajay Jasra. “用于近似贝叶斯计算的自适应序贯蒙特卡洛方法”。英文。载于：统计与计算22.5（2012年9月），第1009–1020页。 ISSN: 1573-1375。 DOI: 10.1007/s11222-011-9271-y。 URL: https://doi.org/10.1007/s11222-011-9271-y（访问于2025年7月13日）。

[9] Charita Dellaporta 等人。“基于模拟器模型的稳健贝叶斯推断：MMD后验自助法”。英文。载于：第*25*届人工智能与统计国际会议论文集。ISSN：2640-3498。PMLR，2022年5月，第943–970页。 URL: https://proceedings.mlr.press/v151/dellaporta22a.html（访问于2025年7月13日）。[10] Zhiyan Ding 和 Qin Li. “集合卡尔曼采样器：平均场极限与收敛性分析”。载于：*SIAM*数学分析杂志53.2（2021年1月）。出版方：工业与应用数学学会，第1546–1578. ISSN: 0036-1410. DOI: 10.1137/20M1339507. URL: https://epubs.siam.org/doi/10.1137/20M1339507（访问于 06/26/2025）。[11] Alfredo Garbuno-Inigo、Nikolas N¨usken 和 Sebastian Reich。《用于贝叶斯推断的仿射不变相互作用朗之万动力学》。英文。载于：SIAM应用动力系统杂志 19.3（2020年1月），第 1633–1658 页。 ISSN: 1536-0040. DOI: 10.1137/19M1304891. URL: https://epubs.siam.org/doi /10.1137/19M1304891.[12] Alfredo Garbuno-Inigo 等。《相互作用朗之万扩散：梯度结构与集合卡尔曼采样器》。英文。载于：SIAM应用动力系统杂志 19.1（2020年1月），第 412–441 页。 ISSN: 1536-0040. DOI: 10.1137/19M1251655. URL: https://epubs.siam.org/doi/10.1137/19M1251655.[13] J. K. Ghosh、Mohan Delampady 和 Tapas Samanta。《贝叶斯分析导论：理论与方法》。英文。Springer统计学丛书。纽约：Springer，2006。 ISBN: 978-0-387-40084-6.[14] Tilmann Gneiting 和 Adrian E Raftery。《严格适当评分规则、预测与估计》。载于：《美国统计学会杂志》102.477（2007年3月），第 359–378 页。 DOI: 10.1198/016214506000001437. URL: https://doi.org/10.1198/016214506000001437.[15] Arthur Gretton 等。《双样本问题的核方法》。载于：神经信息处理系统进展。第 19 卷。麻省理工学院出版社，2006。 URL: https:// papers.nips.cc/paper_files /paper/2006/hash / e9fb2eda3d9c55a0d89c98d6c54b5b3e-Abstract.html.[16] Arthur Gretton 等。《核双样本检验》。载于：《机器学习研究杂志》13.25（2012），第 723–773 页。 ISSN: 1533-7928. URL: http: //jmlr.org/papers/v13/gretton12a.html.[17] Peter Gr¨unwald 和 Thijs van Ommen。《错设线性模型中贝叶斯推断的不一致性及其修正方案》。载于：《贝叶斯分析》12.4（2017年12月），第 1069–1103 页。 ISSN: 1936-0975, 1931-6690. DOI: 10.1214/17-BA1085. URL: https://projecteuclid.org/journals/bayesian-analysis/volume-12/issue-4/Inconsistency-of-Bayesian-Inference-for-Misspecified-Linear-Models-and-a/10.1214/17-BA1085.full.[18] C. C. Holmes 和 S. G. Walker。《在一般贝叶斯模型中为幂似然赋值》。载于：Biometrika 104.2（2017年6月），第 497–503 页。ISSN: 0006-3444. DOI: 10.1093/biomet/asx010. URL: https://doi.org/10 .1093/biomet/asx010.[19] Jack Jewson、Jim Q. Smith 和 Chris Holmes。《使用一般散度准则的贝叶斯推断原理》。英文。载于：熵 20.6（2018年6月）。期号：6；出版商：多学科数字出版研究所，第 442 页。ISSN: 1099-4300. DOI: 10.3390/e20060442. URL: https://www.mdpi.com/10 99-4300/20/6/442.

[20] Andrew Jones 和 Ben Leimkuhler。《用于驱动分子系统采样的自适应随机方法》。载于：《化学物理杂志》135.8（2011年8月），第 084125 页。 ISSN: 0021-9606. DOI: 10 . 1063 /1. 3626941. URL: https://doi.org/10.1063/1.3626941.[21] Jarno Lintusaari 等。《近似贝叶斯计算的基础与最新进展》。载于：《系统生物学》66.1（2017年1月），第 e66–e82 页。 DOI: 10.1093/sysbio/syw077. URL: https://doi.org/10.1093/sysbi o/syw077.[22] Qiang Liu、Jason Lee 和 Michael Jordan。《用于拟合优度检验的核化斯坦差异》。英文。载于：第33届国际机器学习会议论文集。ISSN：1938-7228。PMLR，2016年6月，第 276–284 页。 URL: https://proceedings.mlr.press/v48/liub16.html.[23] Edward N. Lorenz。《可预测性——一个部分解决的问题》。载于：《天气与气候的可预测性》。Renate Hagedorn 和 Tim Palmer 编。剑桥：剑桥大学出版社，2006，第 40–58 页。 ISBN: 978-0-521-84882-4.DOI: 10.1017/CBO9780511617652.004. URL: https://www.cambridge.org / core/books/predictability-of-weather-and-climate/predictabilit y-a-problem-partly-solved/3221BDE379DEB669BA52C66263AF3206.[24] S P Lyddon、C C Holmes 和 S G Walker。《一般贝叶斯更新与损失似然自助法》。载于：Biometrika 106.2（2019年6月），第 465–478 页。ISSN: 0006-3444. DOI: 10.1093/biomet/asz006. URL: https://doi.org/10 .1093/biomet/asz006.[25] Takuo Matsubara 等。《难以处理的似然下的稳健广义贝叶斯推断》。载于：《皇家统计学会杂志B辑：统计方法学》84.3（2022年4月），第 997–1022 页。 ISSN: 1369-7412. DOI: 10.1111 /rssb.12500. URL: https://doi.org/10.1111/rssb.12500.[26] Jeffrey W. Miller。《广义后验分布的渐近正态性、集中性与覆盖率》。载于：《机器学习研究杂志》22.168（2021），第 1–53 页。 ISSN: 1533-7928. URL: http://jmlr.org/papers/v22/20- 469.html.[27] Christopher Nemeth 和 Paul Fearnhead。《随机梯度马尔可夫链蒙特卡罗》。载于：《美国统计学会杂志》116.533（2021年1月），第 433–450 页。 ISSN: 0162-1459. DOI: 10.1080/01621459.2020.1 847120. URL: https://doi.org/10.1080/01621459.2020.1847120（访问于 07/13/2025）。[28] Nikolas N¨usken 和 Sebastian Reich。《关于 Garbuno-Inigo、Hoffmann、Li 和 Stuart 的〈相互作用朗之万扩散：梯度结构与集合卡尔曼采样器〉的注记》。arXiv:1908.10890 [math] 版本：1。2019年8月。 DOI: 10.48550/arXiv.1908.10890. URL: http://arxiv.org/abs/1908.10890.[29] L. Pacchiardi。《使用评分规则对生成模型进行统计推断》。英文。博士论文。牛津大学，2022。 URL: https://ora.ox.ac . uk/objects/uuid:67db26b1-3dbf-4087-bbd1-7297769ff37f.[30] Lorenzo Pacchiardi、Sherman Khoo 和 Ritabrata Dutta。《广义贝叶斯

n

无似然推断”。载于：《电子统计学杂志》18.2（2024年1月）。DOI: 10.1214/24-EJS2283。 URL: https://projecteuclid.org/journals/e lectronic-journal-of-statistics/volume-18/issue-2/Generalized-Bayesian-likelihood-free-inference/10.1214/24-EJS2283.full。

[31] Mijung Park、Wittawat Jitkrittum 和 Dino Sejdinovic。“K2-ABC：利用核嵌入进行近似贝叶斯计算”。英文。载于：第*19*届人工智能与统计国际会议论文集。ISSN：1938-7228。PMLR，2016年5月，第398–407页。 URL: https://proceedings.mlr.press/v51/park16.html。[32] L. F. Price 等。“贝叶斯合成似然”。载于：计算与图形统计学杂志27.1（2018年1月），第1–11页。 ISSN: 1061-8600。DOI: 10.1080/10618600.2017.1302882。 URL: https://doi.org/10.1080/10618600.2017.1302882。[33] Claudia Schillings 和 Andrew M. Stuart。用于逆问题的集合卡尔曼滤波分析。arXiv:1602.02020 [math]。2016年9月。 DOI: 10.48550/arXiv.1602.02020。 URL: http://arxiv.org/abs/1602.02020。[34] Owen Thomas 等。“通过比率估计进行无似然推断”。英文。载于：贝叶斯分析17.1（2022年3月）。出版方：数理统计学会。 ISSN: 1936-0975。 DOI: 10.1214/20-ba1238。 URL: https://projec teuclid.org/journals/bayesian-analysis/volume-17/issue-1/Likelihood-Free-Inference-by-Ratio-Estimation/10.1214/20-BA1238.full。[35] Simon Weissmann。“非线性正向模型的集合卡尔曼反演的梯度流结构与收敛性分析”。载于：逆问题38.10（2022年10月），第105011页。 ISSN: 0266-5611, 1361-6420。 DOI: 10.1088 /1361-6420/ac8bed。 URL: https://iopscience.iop.org/article/10.1088/1361-6420/ac8bed。[36] Daniel S. Wilks。“随机参数化对 Lorenz ’96系统的影响”。英文。载于：英国皇家气象学会季刊131.606（2005年），第389–407页。 ISSN: 1477-870X。 DOI: 10.1256/qj.04.03。 URL: https: //onlinelibrary.wiley.com/doi/abs/10.1256/qj.04.03。[37] Arnold Zellner。“最优信息处理与贝叶斯定理”。载于：《美国统计学家》42.4（1988年11月），第278–280页。 ISSN: 0003-1305。DOI: 10.1080/00031305.1988.10475585。