指数族中基于差分隐私充分统计量的噪声校准推断

Amir Asiaee[1] **Samhita Pal**[1]

[1]范德堡大学医学中心生物统计学系，美国田纳西州纳什维尔 {amir.asiaeetaheri, samhita.pal}@vumc.org

arXiv:2603.02010v1 [cs.LG] 2 Mar 2026

摘要

许多差分隐私（DP）数据发布系统要么输出差分隐私合成数据，让分析人员照常进行推断，这可能导致严重的误校准；要么输出差分隐私点估计，却没有原则性的方法来量化不确定性。本文针对指数族提出了一种简洁且易于处理的折中方案：仅发布差分隐私充分统计量，然后进行经过噪声校准的基于似然的推断，并可选择在后处理中生成参数化合成数据。我们的贡献包括：（1）在高斯机制下发布裁剪后的充分统计量以实现近似差分隐私的一般方案；（2）证明插件式差分隐私最大似然估计具有渐近正态性，给出明确的方差膨胀表达式，并构造有效的Wald型置信区间；（3）提出一种噪声感知的似然校正方法，该方法与插件式方法在一阶上等价，同时支持基于自助法的区间；（4）给出匹配的极小极大下界，证明隐私导致的失真率无法避免。由此得到的理论为发布具有原则性不确定性量化的差分隐私合成数据提供了具体的设计规则和实用处理流程，并在三个指数族和真实人口普查数据上进行了验证。

1 引言

合成数据常被宣传为一种兼顾隐私的原始数据集共享替代方案。然而，科学研究的目标很少只是获得一个“看起来逼真”的数据集；更重要的是，支持对总体参数（标准误、 p 值、覆盖率）进行有效推断，并且越来越多地支持得出有效的因果结论。近期基准测试表明，许多差分隐私合成数据方法保留了一些低维边缘分布，但如果分析人员将合成数据视同真实数据，在推断有效性和假设检验方面仍可能表现很差（Tao 等人，2022；Räisä 等人，2023；Perez 等人， 2024]）。这些失败并不令人意外：差分隐私会引入随机性，而在不确定性量化中必须将其纳入考量。

本文聚焦于数学上最易处理且应用最广泛的一类情形，即存在充分统计量的正则指数族模型。在指数族中，基于似然的推断仅依赖于数据集的经验充分统计量 $\bar{S} = n^{-1}\sum_i s(X_i)$。这使隐私机制与推断之间形成了自然且清晰的分离：（1）发布一个差分隐私加噪充分统计量 S；然后（2）仅根据 S 进行噪声校准推断。由于任何下游计算——包括参数估计、置信区间和参数化合成数据生成——都是已发布统计量的确定性函数，因此经后处理后，它们继承相同的差分隐私保证。

尽管差分隐私推断和差分隐私合成数据各自都受到关注 [Ferrando 等人，2022，Bernstein 和 Sheldon，2018，McKenna 等人， 2021, 2022],但很少有研究提供统一的处理框架，能够：（a）为基于差分隐私充分统计量的推断提供明确的渐近理论；（b）量化隐私导致的精确方差膨胀；以及（c）将参数化合成数据生成与同一推断框架联系起来。我们针对指数族填补了这一空白。

1.1 主要贡献与论文路线图

我们将一个处理流程形式化如下：

$$D = \{X_i\}_{i=1}^n \quad \longrightarrow \quad \widetilde{\bar{S}} \quad \longrightarrow \quad \widetilde{\theta} \quad (\longrightarrow\ D^{\mathrm{syn}}).$$

- **差分隐私机制：** 发布带噪的经验充分统计量 $\widetilde{S}$，使用高斯机制 [Dwork 等，2006；Dwork 和 Roth， 2014]。
- **推断：** 计算插件式差分隐私极大似然估计 $\theta_{\text{插件式}}$ 或噪声感知估计量 θ_{NA}。
- **可选的合成数据：** 样本 $D_{\text{合成}} \sim p(\cdot \mid \theta)$；通过后处理，合成数据继承差分隐私。

我们给出：(i) 明确的渐近方差膨胀，(ii)

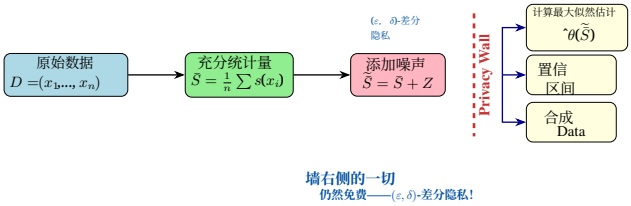


图 1： 处理流程概览。加噪充分统计量 $\widetilde{S}$ 是唯一受差分隐私保护的发布；通过后处理，所有下游任务都继承相同的 (ε, δ)-差分隐私保证。

有效的置信区间，以及 (iii) 在一个典型子类中的下界，其速率与上界相匹配。

2 相关工作

差分隐私估计与极小极大速率。差分隐私估计的基础理论表明，隐私会以统计准确性为代价。Dwork 等人 [2006] 提出了拉普拉斯机制和高斯机制，并根据查询的敏感度校准噪声。将差分隐私与经典统计推断联系起来的早期工作包括 Wasserman和Zhou [2010]。对于有界数据的均值估计，Smith [2011] 证明，当 n 相对于 $1/\varepsilon$ 较大时，(ε, δ)-差分隐私估计量能够达到接近非隐私估计量的速率；而 Duchi 等人 [2018] 给出的下界表明，在局部差分隐私模型中，$\Omega(1/(n\varepsilon))$ 的均方根误差失真不可避免，并且当 ε 足够小时，这种失真也会出现在中心化模型中。在中心化模型中，Cai 等人 [2021]建立了 (ε, δ)-差分隐私下参数估计的紧致极小极大速率，表明对于 d 维有界问题，隐私带来的 ℓ_2 风险代价为 $\Theta(d/(n^2\varepsilon^2))$；我们的定理 3 得出了 $d = 1$ 情形。对于高度依赖高斯噪声的场景，Dong 等人 [提出的高斯差分隐私 2022]等替代性隐私核算框架提供了另一种视角。

差分隐私推断与置信区间。当估计量受到扰动以保护隐私时，如何构造有效的置信区间是一项核心挑战。通用方法包括用于差分隐私置信区间的参数自助法 [Ferrando 等人，以及用于线性回归的差分隐私假设检验 [Alabi 和 Vadhan 2022]。Karwa 和 Vadhan [2018] 研究了高斯均值的有限样本差分隐私置信区间。Avella-Medina等人 [2023]通过加噪优化，开发了针对差分隐私M估计量的渐近推断方法。Awan和Wang [2025]研究了基于模拟的隐私化数据有限样本推断程序。Bernstein 和 Sheldon[2018], 开发了针对指数族的噪声感知贝叶斯推断方法，该方法将充分统计量中的差分隐私噪声纳入考量，并展示了如何将已知噪声分布纳入共轭后验更新。最近，Wang等人 [2025] 开发了通用差分隐私自助法框架，并进行了隐私感知方差估计；Covington等人 [2025]则为一大类统计函数提供了具有有效置信区间的无偏差分隐私估计量——量。Awan和Slavkovi´c [2018] 研究了二项比例及相关单参数族的差分隐私推断，采用能够实现精确有限样本推断的最优噪声机制。在图指数族模型中，Karwa和Slavkovi´c [2016] 研究了基于差分隐私加噪充分统计量（加噪度数）的推断以及合成图生成。

差分隐私合成数据生成。大规模实证基准测试表明，基于边际分布／工作负载的方法在表格数据合成方面可能优于基于生成对抗网络的方法，但推断有效性并非自动成立 [Tao 等人， 2022]。最先进的合成框架包括 Private-PGM 及其 MST/NIST-MST 实现 [McKenna 等人，2021], 工作负载自适应 AIM 机制 [McKenna 等人，2022],以及近期融入几何感知损失和最优传输的方法 [Donhauser 等人， 2024]。另一条互补的研究路线探讨如何分析差分隐私合成数据并正确量化不确定性，其中包括结合噪声感知合成数据生成与多重插补的方法[Räisä 等人， 2023] 以及警示对合成数据进行朴素分析会导致第一类错误膨胀的实证研究[Perez 等人， 2024]。

研究定位。我们的工作与 Bernstein 和 Sheldon [2018] 的研究思路最为接近，但有两点不同。首先，我们提供完整的频率学派渐近理论——包括明确的方差膨胀公式、Wald 置信区间以及相匹配的下界——以补充他们的贝叶斯处理。其次，我们表明，同一已发布的差分隐私充分统计量可以发挥双重作用：用于频率学派推断和参数化合成数据生成，且两者都继承相同的隐私保证。这在指数族设定下统一了“差分隐私推断”和“差分隐私合成数据”两方面的研究。我们指出，尽管各个理论组成部分——差分隐私估计量的 Delta 方法中心极限定理以及 Le Cam 下界——在差分隐私统计学文献中早已为人所知，但我们的贡献在于将它们整合到一个实用的处理流程中，并通过全面的实证验证，清晰展示朴素的合成数据分析会失效。

3 设置

3.1 指数族模型

设 $X \in \mathcal{X}$，并考虑一个自然参数为 $\theta \in \Theta \subset \mathbb{R}^d$ 的（最小）正则指数族：

$$p_\theta(x) = h(x) \exp\left(\langle \theta, s(x) \rangle - A(\theta)\right), \qquad (1)$$

其中 $s(x) \in \mathbb{R}^d$ 是充分统计量， $A(\theta)$ 是对数配分函数。给定独立同分布数据 $D = \{X_i\}_{i=1}^n$，经验充分统计量为

$$\bar{S}(D) = \frac{1}{n} \sum_{i=1}^{n} s(X_i) \in \mathbb{R}^d.$$

$\hat{}$（非隐私）最大似然估计量（MLE） $\theta\hat{}^{-}$满足 $\nabla A(\theta) = S(D)$，前提是 MLE 存在。

算法 1 差分隐私充分统计量发布与推断（指数族）

要求： 数据集 $D = \{X_i\}_{i=1}^n$，界限 B，隐私（ε，δ）

计算 $S = \frac{1}{n}\sum_{i=1}^n s(X_i)$ ▷ 充分统计量

集合 $\Delta_2 \leftarrow 2B/n$ ▷ ℓ_2 的敏感度 S

集合 $\sigma \leftarrow \mathrm{AGM}(\Delta_2, \varepsilon, \delta)$ ▷ 解析高斯校准 [Balle 和 Wang，2018]

样本 $Z \sim \mathcal{N}(0, \sigma^2 I_d)$ 并发布 $\widetilde{S} = \bar{S} + Z$

代入式： $\widetilde{\theta}_{\mathrm{plug}} \leftarrow (\nabla A)^{-1}(\widetilde{S})$

噪声感知： $\widetilde{\theta}_{\mathrm{NA}} \leftarrow \arg\max_{\theta\in\Theta} \log p(\widetilde{S} \mid \theta)$（第4.3节）

（可选）采样合成数据 $D^{\mathrm{syn}} \sim p(\cdot \mid \widetilde{\theta}_{\mathrm{plug}})$（或 $\widetilde{\theta}_{\mathrm{NA}}$）

3.2 差分隐私与高斯机制

两个数据集 D，D' 若仅相差一条记录，则称为邻居。如果对于所有相邻的 D，D' 以及可测集合 E，随机化机制 $\mathcal{M}$ 满足以下条件，则称其为 (ε, δ)-差分隐私：

$$\Pr(\mathcal{M}(D) \in E) \le e^{\varepsilon} \Pr(\mathcal{M}(D') \in E) + \delta.$$

我们使用高斯机制 [Dwork 等人，2006，Dwork 和 Roth，2014]。

假设 1（有界充分统计量）。存在 $B > 0$ 使得 $\|s(x)\|_2 \le B$ 对所有 $x \in \mathcal{X}$ 均成立。

在假设 1 下，$S(D)$ 的 ℓ_2 敏感度至多为 $\Delta_2 = 2B/n$。

4 机制与估计量

4.1 差分隐私充分统计量发布

这里，$\mathrm{AGM}(\Delta_2, \varepsilon, \delta)$ 表示解析高斯机制校准，它返回最小的 σ，使得向敏感度为 Δ_2 的查询添加 $Z \sim \mathcal{N}(0, \sigma^2 I_d)$ 后，该查询满足 (ε, δ)-差分隐私 [（Balle 和 Wang，2018]）。在所有实验中，我们均设定 $\delta = 1/n^2$；随后，解析高斯机制校准 [（Balle 和 Wang，2018]）确定最小的 σ。

定理 1（充分统计量发布的隐私性）。在假设 *1* 下，发布 $S = S(D) + Z$，其中 $Z \sim \mathcal{N}(0, \sigma^2 I_d)$，且 σ 如算法 *1* 所示，满足 (ε, δ)-差分隐私。此外，对 S 进行任何（随机化）后处理（包括 θ 和 D^{syn}）也满足 (ε, δ)-差分隐私。

证明概要。ℓ_2 的敏感度至多为 $2B/n$。所述 σ 是通过解析高斯机制校准选定的，该校准为高斯机制提供了一个 (ε, δ)-差分隐私保证，适用于任意 $\varepsilon > 0$ [Balle 和 Wang，2018]；根据高斯机制保证和后处理不变性 [Dwork 等人，2006，Dwork 和 Roth，2014]，即可得到差分隐私保证。 □

4.2 插件式差分隐私极大似然估计与渐近正态性

代入估计量通过求解 $\nabla A(\theta_{\mathrm{plug}}) = S$ 得到。在正则指数族中，$\nabla^2 A(\theta) = I(\theta)$ 等于单个样本的费希尔信息。

假设 2（正则性）。该模型是最小正则指数族，其中 θ_0 位于 Θ 的内部。费希尔信息 $I(\theta_0)$ 是正定的。

定理 2（渐近分布与方差膨胀）。在假设 *1–2* 下，插件式差分隐私极大似然估计满足

$$\sqrt{n}\left(\widetilde{\theta}_{\mathrm{plug}} - \theta_0\right) \xrightarrow{d} \mathcal{N}\big(0,\ I(\theta_0)^{-1} + n\sigma^2 I(\theta_0)^{-2}\big),$$

其中，σ^2 是算法 *1* 中高斯噪声的方差，且 $I(\theta_0)^{-2} := I(\theta_0)^{-1} I(\theta_0)^{-1}$。等价地，

$$\widetilde{\theta}_{\mathrm{plug}} - \theta_0 = \underbrace{O_p(n^{-1/2})}_{\textit{sampling}} + \underbrace{O_p(\sigma)}_{\textit{privacy}}.$$

当针对 (ε, δ)-差分隐私校准 σ 时，若 ℓ_2 敏感度 $\Delta_2 = 2B/n$，噪声尺度与 Δ_2 成正比，并随 ε 减小（至多相差 $\log(1/\delta)$ 因子），从而得到有界问题中熟知的 $O_p((n\varepsilon)^{-1})$ 隐私失真率。

证明概要。证明将误差分解为相互独立的抽样项和隐私项。*(i)* 多元中心极限定理（中心极限定理）给出 $\sqrt{n}(S - \mu(\theta_0)) \xrightarrow{d} \mathcal{N}(0, I(\theta_0))$，再通过均值参数逆映射 $(\nabla A)^{-1}$ 应用德尔塔方法，得到经典的 $I(\theta_0)^{-1}/n$ 方差。*(ii)* 在 $\bar{S}$ 附近对 $(\nabla A)^{-1}$ 作一阶泰勒展开，得到 $\widetilde{\theta}_{\mathrm{plug}} - \hat{\theta} = I(\hat{\theta})^{-1} Z + O_p(\|Z\|^2)$，由此产生 $\sigma^2 I(\theta_0)^{-2}$ 隐私方差。*(iii)* 由于高斯噪声 Z 与数据相互独立，两个方差分量相加。完整证明见附录 A。 □

推论 1（隐私何时能保持经典效率？）。在定理 *2* 的条件下，若 $n\varepsilon^2 \to \infty$（在 δ 中至多相差对数因子时等价于 $\varepsilon \gg n^{-1/2}$），则隐私引入的方差趋于零，且 θ_{plug} 渐近而言与非隐私最大似然估计同样有效率。若 $n\varepsilon^2$ 有界，则差分隐私噪声会导致不可忽略的方差膨胀。

4.3 噪声感知似然校正

代入估计量在光滑性条件下具有一阶正确性，但在实际应用中，我们通常会对 $s(x)$ 进行裁剪或其他净化处理，以限制敏感度，这会引入偏差。噪声感知推断会明确对差分隐私噪声以及（可选的）裁剪进行建模。

一种简单的方法是定义噪声统计量似然：

$$\ell_{\mathrm{NA}}(\theta; \widetilde{S}) := \log p(\widetilde{S} \mid \theta), \tag{2}$$

其中，$p(S \mid \theta)$ 由数据模型和差分隐私机制诱导。通常，这是一种难以处理的卷积，但在以下两种区制中可以处理：

•**中心极限定理区制：** 近似为 $S \mid \theta \approx \mathcal{N}(\nabla A(\theta),\ I(\theta)/n)$。于是 $\bar{S} \mid \theta \approx \mathcal{N}(\nabla A(\theta),\ I(\theta)/n + \sigma^2 I)$。

• **共轭／精确区制：** 对于某些分布族和先验，已有精确的噪声感知贝叶斯方法 [（Bernstein 和 Sheldon， 2018]）。

在中心极限定理区制下，噪声感知估计量求解一个广义最小二乘问题：

$$\widetilde{\theta}_{\mathrm{NA}} = \arg\min_{\theta\in\Theta} \big(\widetilde{\bar{S}} - \nabla A(\theta)\big)^{\top} \big(\tfrac{1}{n} I(\theta) + \sigma^2 I\big)^{-1} \big(\widetilde{\bar{S}} - \nabla A(\theta)\big).$$

数值稳定化。在有限样本中（尤其是在隐私保护强度较高时），估计得到的费希尔信息 $I(\theta)$ 可能条件数很差，噪声感知目标函数也可能变得平坦。在我们的实现中，我们采用以下数值保护措施：(i) 蒂赫诺夫正则化 $I(\theta) \leftarrow I(\theta) + \lambda I$，其中 $\lambda = \max(10^{-6},\ 0.01\,\sigma^2)$；(ii) 为 L-BFGS-B 优化器设置盒约束 $\theta_j \in [-10, 10]$；(iii) 在噪声感知目标函数中加入二次锚定惩罚项 $0.1\,\sigma^2 \sum_j (\theta_j - \theta_{\mathrm{plug},\,j})^2$；以及 (iv) 将方差对角元素裁剪至 $10^6/n$。这些选择纯粹是为了提高稳定性，不会改变一阶渐近理论。

命题 1（一阶等价性）。 在定理 *2* 的条件成立且 $I(\theta)$ 具有光滑性的情况下，噪声感知估计量 θ_{NA} 与代入估计量一阶等价：

$$\widetilde{\theta}_{\mathrm{NA}} - \widetilde{\theta}_{\mathrm{plug}} = o_p(n^{-1/2}) + o_p(\sigma),$$

因此，其渐近分布与定理 *2* 中的相同。

证明概要。在代入解处，$S - \nabla A(\theta_{\mathrm{plug}}) = 0$，这也使噪声感知的一阶条件为零。因此，两个估计量在最高阶上求解相同的方程；差异仅通过涉及 $\Sigma(\widetilde{\theta}_{\mathrm{plug}})^{-1} - \Sigma(\theta_0)^{-1}$ 的高阶项产生，而该项为 $O_p(n^{-1/2} + \sigma)$。详见附录 B。 □

4.4 置信区间与假设检验

定理 2 为每个坐标 θ_j 给出了一个 Wald 区间：

$$\mathrm{CI}_j = \left[\widetilde{\theta}_j \pm z_{1-\alpha/2}\sqrt{\widehat{\mathrm{Var}}(\widetilde{\theta})_{jj}}\right],$$

方差估计为

$$\widehat{\mathrm{Var}}(\widetilde{\theta}) = \frac{1}{n}\widehat{I}^{-1} + \sigma^2 \widehat{I}^{-2}, \quad \widehat{I} := I(\widetilde{\theta}).$$

其中，$\widehat{\mathrm{Var}}(\widetilde{\theta})_{jj}$ 表示 $\widehat{\mathrm{Var}}(\widetilde{\theta})$ 的第 j 个对角线元素。另一种方法是使用差分隐私参数自助法，以更好地捕捉有限样本效应和裁剪偏差 [Ferrando 等人， 2022]。自助抽样值从噪声感知分布 $p(S \mid \theta)$ 中生成，因此噪声感知似然模型可直接用于自助法；实现细节见附录 D。

5 下界

只有接近最优的上界才有用。我们给出了一个简单的下界，其速率与隐私项相匹配。

定理 3 (不可避免的 $\Omega(1/(n^2\varepsilon^2))$ 均方误差下界。考虑定义在 $\mathcal{X} = \{-1, +1\}$ 上的一维指数族，其 $s(x) = x$，自然参数为 $\theta \in \Theta \subset \mathbb{R}$。对于任意基于 n 个独立同分布样本的 (ε, δ)-差分隐私估计量 θ，其中 $\delta < 1/4$ 且 $\varepsilon \in (0, 1]$,

$$\inf_{\widehat{\theta}\ (\varepsilon,\delta)\text{-DP}} \sup_{\theta\in\Theta} \mathbb{E}_\theta\Big[(\widehat{\theta} - \theta)^2\Big] \ \geq\ \frac{c}{n^2\varepsilon^2},$$

其中普适常数为 $c > 0$。特别地，极小极大均方根误差至少为 $\sqrt{c}/(n\varepsilon)$。结合定理 *2*，这表明在该典型子类中，$1/(n\varepsilon)$ 隐私失真率（忽略 δ 中的对数因子）是极小极大最优的。

证明概要。将均值参数写作 $\mu(\theta) = \tanh(\theta)$。对于任意差分隐私估计量 θ，后处理估计量 $\widehat{\mu} = \widehat{\tanh}(\theta)$ 也满足差分隐私，并且满足 $|\widehat{\mu} - \mu(\theta)| \leq |\theta - \theta|$，因为 tanh 是 1-Lipschitz。因此，任何差分隐私均值估计的极小极大下界都可直接推广至 θ。所述速率可通过将 Cai 等人关于 (ε, δ)-差分隐私下有界均值估计的中心模型极小极大下界专门化得到 [2021]。完整论证见附录 C。 □

6 实验

这些实验旨在验证三项主张：(i) 定理 2 中的渐近方差公式能够准确预测有限样本行为；(ii) Wald 置信区间能够达到名义覆盖率；(iii) 比较裁剪条件下的代入式估计量与噪声感知估计量。

6.1 实验设计

所有实验均采用 (ε, δ)-差分隐私，并设置 $\delta = 1/n^2$；这是标准选择，可使 δ-失效概率相对于样本量可忽略。

模型与充分统计量。我们考虑三个经典指数族。

1. **高斯均值估计。** 已知 σ_0^2 时的数据 $X_i \sim \mathcal{N}(\mu, \sigma_0^2)$。充分统计量：$s(x) = x$。自然参数：$\theta = \mu/\sigma_0^2$。裁剪：将观测值截断至 $[-B, B]$。2. **逻辑回归。** 二元结果（X_i，Y_i），其中 $\Pr(Y = 1 \mid X) = \mathrm{Sigmoid}(\theta^\top X)$。充分统计量：$s(x,\ y) = yx \in \mathbb{R}^d$。裁剪：将特征投影至 $\|X\|_2 \leq B$。3. **泊松回归。** 计数结果 $Y_i \mid X_i \sim$ 泊松$(\exp(\theta^\top X_i))$。充分统计量：$s(x,\ y) = yx$。裁剪：将 Y 截断至 $[0, B_Y]$，并将特征投影至 $\|X\|_2 \leq B_X$。

对比的方法。

- **非隐私最大似然估计**（预言机）：标准最大似然估计，具有经典渐近置信区间。
- **差分隐私插件式方法（Wald）**：采用算法 1，并使用定理 2 中的方差公式计算 Wald 置信区间。
- **噪声感知差分隐私（Wald）**：采用第 4.3 节中的估计量，并使用相同的置信区间公式。
- **差分隐私参数自助法**：采用 Ferrando 等人的自助法程序 [2022]，并将其调整为适用于充分统计量发布；自助样本从含噪统计量的分布中抽取。~
- **朴素差分隐私合成数据**：生成 $D_{\text{合成}} \sim p(\cdot \mid \theta_{\text{代入}})$，然后对$D_{\text{合成数据}}$应用标准（非噪声感知差分隐私）最大似然估计和置信区间。

指标。

- **覆盖率**：在 1000 次独立运行中，名义水平为 95% 的置信区间的经验覆盖率（每次运行均使用新数据和差分隐私噪声）。
- **区间长度**：平均置信区间宽度，用于衡量推断精度。
- **参数均方误差**：$\mathbb{E}[\left\|\widetilde{\theta} - \theta_0\right\|^2]$，在重复试验中估计。
- **第一类错误**：在真实参数处对 H_0 的拒绝率：$\theta_j = \theta_{0,\ j}$
- **检验效能**：对于一系列备择假设 $\theta_j = \theta_{0,\ j} + \Delta$ 在 $\Delta \in \{0.1, 0.2, 0.5, 1.0\}$ 下的拒绝率。

6.2 模拟实验

实验 1：方差膨胀验证。对于带有 $d = 1$ 的高斯模型，计算 θ_{plug} 在 5000 次重复试验中的经验方差~，并与理论预测 $I(\theta_0)^{-1}/n + \sigma^2 I(\theta_0)^{-2}$ 进行比较。改变 $n \in \{100, 500, 1000, 5000\}$ 和 $\varepsilon \in \{0.1, 0.5, 1.0, 5.0, 10.0\}$。将经验方差与理论方差绘制为散点图；各点应落在对角线上。

实验 2：隐私范围内的覆盖率。对于每个模型以及每个 (ε, n) 组合，同时使用代入式估计量和噪声感知估计量计算 95% Wald 置信区间。在固定 n 的情况下，将覆盖率报告为 ε 的函数（反之亦然）。关键预测是：当 $n\varepsilon^2 \gg 1$ 时，所有方法的覆盖率应接近名义水平，但朴素的合成数据分析在 ε 较小时，其区间覆盖率会严重不足。

实验 3：裁剪条件下代入式与噪声感知方法的比较。对于带有 $d = 5$ 的逻辑回归，将裁剪半径B 从激进（$B = 1$）调整到宽松（$B = 10$），并测量偏差和覆盖率。我们测试当 B 相对于自然数据尺度较小时，噪声感知估计量是否表现出更小的偏差和更好的覆盖率。

实验 4：缩放定律验证。对于高斯模型，在双对数坐标下绘制不同 ε 取值时的均方误差与 n 的关系。理论预测为均方误差 $\approx c_1/n + c_2/(n^2\varepsilon^2)$，在双对数图上应呈现两种明显不同的斜率。找出隐私项开始占主导地位的交叉点。

6.3 真实数据实验

数据集。我们采用美国社区调查 (ACS) 公共使用微观数据样本，并遵循 Folktables 框架 [Ding 等人， 2021] 该框架提供标准化的预测任务。具体而言，我们针对加利福尼亚州 2018 年的 ACSIncome 任务（预测收入是否超过 50,000 美元）进行研究。删除含缺失值的行后，我们将所有特征标准化为均值为零、方差为一，再选取边际方差最高的 $d = 10$ 个协变量，从而得到一个逻辑回归问题。特征经过截断，范围为 $\|X\|_2 \le B$，其中 $B = 3.0$（标准化后），由此得到ℓ_2 敏感度 $\Delta_2 = 2B/n = 6/n$。较大的全样本（$n \approx 1.6$M）使我们能够将全数据最大似然估计视为真实值 θ_0。

实验方案。抽取大小为 $n \in \{500, 1000, 5000, 10000\}$ 的数据集作为子样本，并在 $\varepsilon \in \{0.5, 1.0, 2.0, 5.0\}$应用我们的处理流程。每种配置运行 500 次重复试验。报告每个回归系数的 Wald 置信区间覆盖率和区间长度，并对各坐标取平均。

6.4 合成数据评估

~生成参数化合成数据 D^{syn}，基于 θ 并评估：

- 在 $D_{\text{合成数据}}$ 上计算的点估计和置信区间（经噪声感知校正后）是否与差分隐私统计量发布所得结果一致。由于两者都是对相同的~样本 S进行后处理，估计量应当一致，但置信区间宽度可能因$n_{\text{合成数据}}$而异。
- 朴素分析（将 $D_{\text{合成数据}}$ 视为真实数据）如何导致不确定性校准失准，从而重现并扩展 Räisä 等人 [2023], 和 Perez 等人 [2024]的研究发现。我们预计覆盖不足会很严重，尤其是在 ε较小时。
- 合成样本量 $n_{\text{合成数据}}$ 对蒙特卡洛误差的影响，并验证$n_{\text{合成数据}} \gg n$ 是使合成抽样误差可以忽略不计所必需的。

6.5 结果

实验 1：方差膨胀验证。图2比较了差分隐私代入式估计量的经验方差与定理 2 中的理论预测，即$I(\theta_0)^{-1}/n + \sigma^2 I(\theta_0)^{-2}$。在高斯模型（$d = 1$）中，该结果简化为 $1/n + \sigma^2$，因此每个点对应一个（n， ε）对；如果有限样本行为与渐近公式一致，这些点应落在恒等线上。

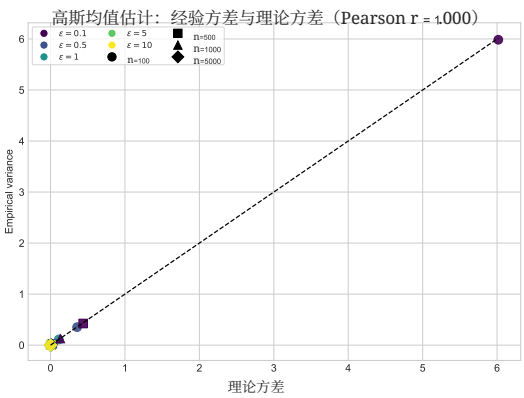


图2：高斯均值估计中，差分隐私插件估计量的经验方差与理论方差。每个点对应一种（n，ε）配置，颜色表示 ε，标记形状表示 n。与恒等线高度吻合，验证了定理 2 在有限样本下的相关性。

在所有隐私级别和样本量下，各点都紧密集中在对角线周围，呈现出近乎完美的线性一致性。理论方差与经验方差之间的皮尔逊相关系数为 $r = 0.9999978$，在 20 个（n，ε）设置中的最大相对误差为 3。67%。这证实了抽样噪声与隐私噪声的加性分解在有限样本下已经足够准确，支持在无需额外校准的情况下实际使用我们的 Wald方差公式。

实验 2：不同隐私范围下的覆盖率。图 3 展示了高斯、逻辑斯蒂和泊松模型在两种样本量下，95% 覆盖率随 ε 的变化情况。在所有模型和（n，ε）组合下，非隐私覆盖率保持在 $[0.941, 0.961]$，差分隐私插件式沃尔德方法保持在 $[0.768, 1.000]$，差分隐私噪声感知沃尔德方法保持在 $[0.768, 1.000]$，差分隐私自助法保持在$[0.765, 1.000]$。这三种方法中校准效果最差的隐私设置是泊松模型下的（n，ε）=（500，0。1），此时覆盖率约为 0。77。

朴素合成数据基线的覆盖率严重不足，范围为 0.014 至 0.844；最低覆盖率出现在泊松模型 (n, ε) = (500,0.1)。在 $\varepsilon = 0.1$ 下，其在所有模型和 n 上的平均覆盖率仅为 0.079。因此，图3和表1均表明，考虑隐私噪声对于有效的不确定性量化至关重要。

泊松模型在三个族中表现出最差的校准（覆盖率 ≈ 0.77 位于 n=500, ε=0.1），这反映出有限样本Wald失效：对于对数链接模型，$I(\theta) = X^\top \mathrm{diag}(\exp(X\theta))\, X/n$ 通过~凸函数 exp(·)依赖于 θ，因此根据詹森不等式，当隐私噪声较大时，$I(\theta)$会被系统性抬高，从而得到过窄的置信区间。随着 ε 增大，该效应消失。泊松覆盖率还进一步受到$s(x, y) = yx$高敏感度的制约：在 $B_X = 3, B_Y = 20$下，有效界限为 $B = 60$，从而在强隐私条件下产生较大噪声。

表1： 摘要：高斯模型 ($n = 1000$) 中不同方法和隐私级别下的 95%置信区间覆盖率。

方法	ε=0.1	ε=0.5	ε=1.0	ε=5.0	ε=10
非隐私最大似然估计	0.947	0.958	0.954	0.948	0.948
差分隐私插件式沃尔德方法	0.950	0.947	0.951	0.950	0.955
差分隐私噪声感知沃尔德方法	0.950	0.947	0.951	0.950	0.955
差分隐私参数自助法	0.944	0.945	0.945	0.945	0.948
朴素差分隐私合成数据	0.146	0.499	0.679	0.848	0.831

实验 2（精度权衡）：置信区间长度。图 4通过报告平均置信区间长度对覆盖率图加以补充。在较小的 ε下，有效的差分隐私方法会拓宽区间以吸收隐私不确定性，而朴素合成数据分析则保持人为的狭窄。这一模式解释了朴素基线同时出现的覆盖不足与表面上的精度。

随着 ε 增大，差分隐私感知方法的区间长度逐渐缩短，趋近于非隐私基准，这与推论 1 预测的渐近效率转变相符。总之，图 3 和图 4 表明，恰恰在隐私噪声最强的情况下，诚实校准需要更宽的区间。

实验 3：裁剪下的代入式方法与噪声感知方法。图5通过改变裁剪半径 B 并固定（n，ε）=(1000，1.0)，单独考察裁剪对逻辑斯蒂回归的影响。

结果显示出一条反映敏感度—准确度权衡的 U形偏差曲线。在强裁剪（$B \leq$1.0）下，裁剪偏差占主导：在 $B = 0.5$时，代入式方法的绝对偏差为 2.323，覆盖率为 0.116；噪声感知方法的偏差为 2.767，覆盖率为 0.088。在中等程度的裁剪（B = 2.0–3.0）下，两种方法都将偏差降至 0.16–0.29，并将覆盖率提高至 0.81–0.98。在宽松裁剪（B = 5.0–10.0）下，裁剪偏差消失，但 $\sigma \propto 2B/n$随 B线性增大，导致误差增加（偏差为 0.24–1.06）；覆盖率仍然良好（0.96–0.97），因为 Wald 区间正确地变宽。

在任何设置下，噪声感知估计量的表现都不优于代入式估计量，这与命题 1 一致：当中心极限定理对裁剪数据的近似较差时，二阶校正不仅没有帮助，反而可能有害。这些结果表明，应将 B 选择在特征的 2–3 个标准差处，在裁剪偏差与敏感度驱动的噪声之间取得平衡。

实验 4：缩放律验证。图 6 展示了多个 ε 值相对于样本量n的对数-对数均方误差曲线。经验曲线遵循理论形式 $c_1/n + c_2/(n^2\varepsilon^2)$，在较小的 n下处于隐私主导区间，在较大的 n下处于抽样主导区间。隐私方差不再大于抽样方差的第一个网格点，对于 $\varepsilon = 5$为 $n = 100$，对于 $\varepsilon = 2$为 $n = 500$，对于 $\varepsilon = 1$为 n =5，000，对于 $\varepsilon = 0.5$为 $n = 10$，000。在较大的n下，所有差分隐私曲线都趋近于非隐私 $1/n$ 基线，证实隐私开销会渐近消失。交叉顺序与理论一致：较大的 ε 会将转变推移至更小的

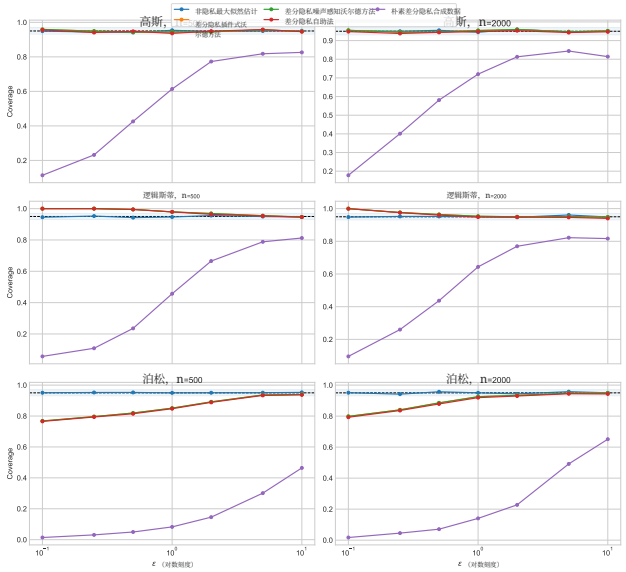


图3： 高斯、逻辑斯蒂和泊松模型在两个样本量下，不同隐私级别对应的 95% 区间覆盖率。阴影区域表示名义覆盖率附近可接受的校准范围。噪声校准的差分隐私方法的覆盖率接近名义值，而在低-ε 范围内，朴素合成数据分析的覆盖率不足。

过渡到更小的 n，这与定理 2 和定理 3 均一致。

实验 5：第一类错误和检验效能。图 7 展示了沃尔德检验在不同备择假设 Δ 和隐私级别下的拒绝率。当 $\Delta = 0$和 $\varepsilon = 1.0$时，差分隐私插件式方法/噪声感知差分隐私方法的第一类错误率为 0.0385，非隐私方法为 0.0515，朴素合成数据分析为 0.292。在不同隐私级别下，朴素合成数据分析在零假设下仍然是反保守的（第一类错误率从 $\varepsilon = 0.5$ 时的 0.488 降至 $\varepsilon = 5.0$时的 0.184）。

随着 Δ 增加，所有方法的检验效能均有所提高；例如，在 $\varepsilon = 1.0$ 和 $\Delta = 0.1$时，差分隐私方法的检验效能为 0.481，而非隐私方法为 0.873。到 $\Delta = 0.2$时，差分隐私方法的检验效能达到 0.967，非隐私方法则达到 1.000。在较低的 ε下，可以看到预期的隐私—检验效能权衡：校准方法牺牲检验效能以维持有效检验尺寸。因此，适当的噪声校准能够确保检验有效，同时保持对真实效应具有竞争力的敏感度。

实验 6：真实数据（ACS/Folktables）。图 8 总结了以全数据最大似然估计为真实值时，ACSIncome 逻辑回归在 (n, ε) 网格上的结果。在所有 (n, ε) 设置下，差分隐私插件式方法、噪声感知差分隐私和差分隐私自助法的平均覆盖率分别为 0.889、0.888 和 0.880，而朴素合成方法为 0.510。在 (n, ε) = (500,0.5) 设置下，朴素方法的表现最差，覆盖率为 0.134；此时，经过校准的差分隐私方法的覆盖率均接近 1.0。

经过第 4.3 节的数值稳定化后，所有设置下的置信区间长度都保持有限：差分隐私插件式方法的范围为 0.127 至 90.835，噪声感知差分隐私的范围为 0.127 至 93.912，DPbootstrap 的范围为 0.125 至 137.185。朴素合成数据区间要短得多（0.118 至 2.735），但

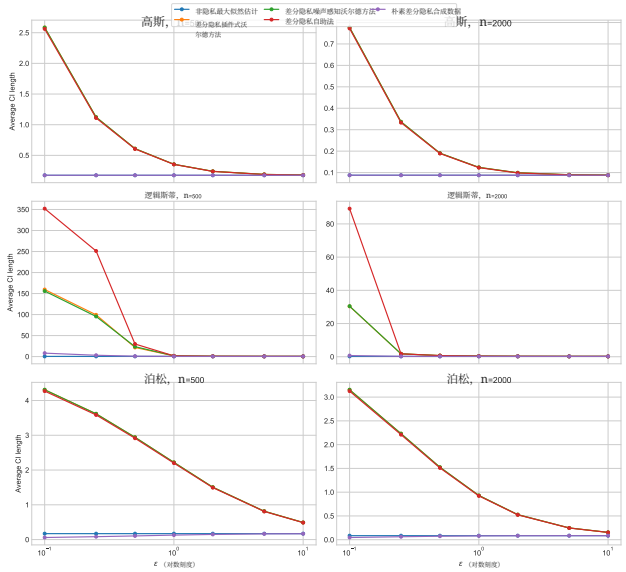


图4： 在与图3相同的设置下，平均置信区间长度随 ε 的变化。隐私保护较强（ε较小）时，噪声感知方法给出的区间较宽；随着 ε 增大，区间逐渐收窄，体现了预期的隐私—准确率权衡。朴素合成方法给出的区间仍然较窄，但校准不当。

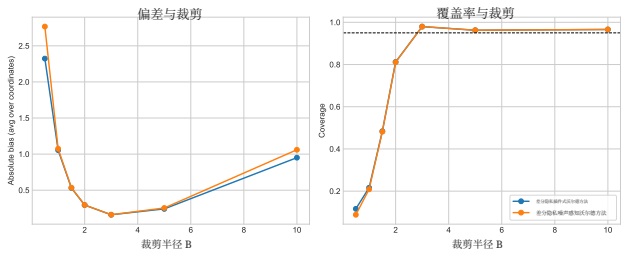


图 5： 逻辑回归裁剪研究：比较差分隐私插件式方法和差分隐私噪声感知估计量。左图：平均绝对偏差随裁剪半径 B的变化。右图：经验 95% 覆盖率随 B的变化。两种方法均呈现 U形偏差曲线：B 过小会导致裁剪偏差，而 B 过大会因敏感度升高而增加噪声。与命题 1一致，噪声感知估计量并不比代入式方法有优势。

覆盖率严重不足。这些结果表明，充分统计量发布流程在理想化的合成数据设置之外也具有实用性。

实验 7：合成数据推断评估。图 9 比较了三种分析模式随合成样本量 $\tilde{}$增至 n_{syn} 时的表现。基于 θ 的直接推断在所有比例下的覆盖率均为 0.952，而噪声感知合成数据分析在 $n_{\text{syn}}/n = 1, 5, 10, 50$ 下分别为 0.942, 0.943,0.954, 0.957。朴素合成数据分析在所有比例下的覆盖率都严重不足：0.685, 0.391,0.282, 0.136。

增加 n_{syn} 只会减少合成数据蒙特卡罗噪声，$\tilde{}$而不会减少从 θ继承的隐私失真。因此，仅增加合成样本量无法修复因忽略隐私噪声而导致的推断校准失准。

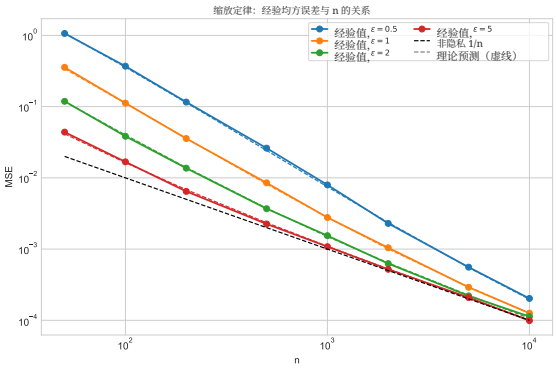


图6： 高斯均值估计的双对数均方误差缩放。实线表示每个 ε 的经验均方误差；颜色匹配的虚线表示理论预测；细黑虚线表示非隐私情形 $1/n$。观察到的双区间行为与预测的 $1/n + 1/(n^2\varepsilon^2)$ 结构一致。

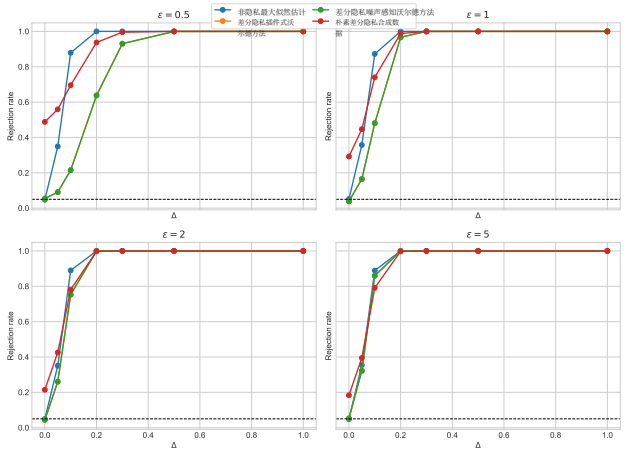


图7： 四个隐私级别下高斯沃尔德检验的检验效能曲线和第一类错误诊断。在 $\Delta = 0$ 时，经过校准的差分隐私方法保持在名义水平 0.05 附近，而朴素合成数据检验的错误率则偏高。检验效能随效应量增大而提升，并随着 ε 增大而趋近于非隐私情形。

7 讨论与扩展

实验要点。在所有实验中，定理 2 给出的方差膨胀公式都具有很高的准确性，经过校准的差分隐私推断显著优于朴素的合成数据分析；除极端有限样本情形外，代入式估计量与噪声感知估计量的表现相近，这与命题 1 一致。

超越指数族。充分统计量释放这一思路可扩展至M估计和广义矩估计法。在这些方法中，“充分统计量”变为经验矩条件 $n^{-1}\sum_i \psi(X_i;\theta)$，可对其进行隐私保护并用于噪声感知推断，自然地与基于工作负载的合成和因果推断流程相结合。

噪声整形与矩阵机制。对于高维 d，各向同性高斯噪声可能并非最优。矩阵机制 [Li 等人， 2015] 可以降低方差

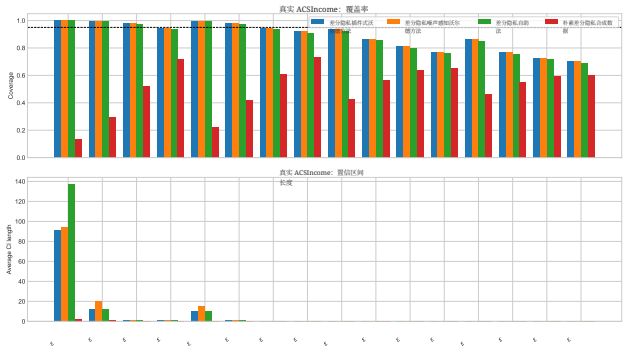


图 8： 在 $(n,\ \varepsilon)$ 种配置下，对真实数据 ACSIncome 进行评估。上方面板报告平均系数覆盖率，下方面板报告平均置信区间长度。差分隐私插件式方法和噪声感知方法的校准效果显著优于朴素合成数据分析，在低-n、低-ε 设置下提升最大。

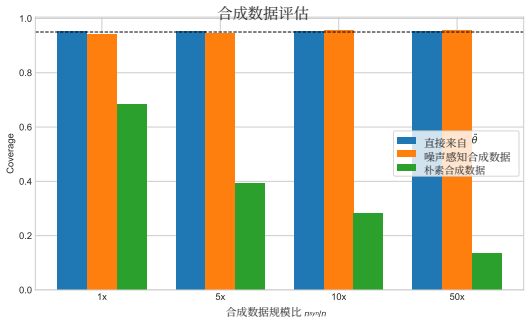


图 9： 比较直接差分隐私推断、噪声感知合成数据分析和朴素合成数据分析在不同合成样本量比例下的覆盖率。直接方法和噪声感知方法的覆盖率均接近名义覆盖率，而朴素分析则持续出现覆盖不足。增大合成数据集并不能消除隐私感知不确定性校正的必要性。

¯在目标方向上的方差，方法是发布线性变换 MS¯，并根据 MS 的敏感度校准噪声；将我们的理论扩展到这种情形很简单，只需对变换后的查询 MS 进行¯修改后的敏感度分析。

与差分隐私随机梯度下降的组合。当 B 先验未知且采用数据依赖裁剪时（如差分隐私随机梯度下降，DP-SGD），敏感度分析会变得更加复杂；研究这种相互作用是一个重要的实践方向。

多变量置信区域。我们的 Wald 区间是按坐标分别构造的；基于同一方差公式构造同时椭球区域的工作留待未来开展。

局限性。我们的处理流程要求使用具有已知充分统计量的正则指数族；扩展到半参数或非参数模型则需要采用不同的方法。对于凸链接函数的模型（例如泊松模型），当 ε 较低时，Wald 置信区间可能无法达到名义覆盖率，这是因为詹森不等式会使 $I(\theta)$ 膨胀；自助法置信区间或许有所帮助。裁剪半径 B 需要在裁剪偏差与噪声幅度之间权衡，这与偏差–准确性–隐私三难困境密切相关 [Kamath 等人, 2025]。

致谢

参考文献

Daniel Alabi 和 Salil Vadhan。差分隐私线性回归的假设检验。发表于神经信息处理系统进展，2022。

Marco Avella-Medina、James Bradshaw 和 Po-Ling Loh。通过加噪优化进行差分隐私推断。统计年刊，51(5)，2023。doi：10.1214/ 23-AOS2321。

Jordan Awan 和 Aleksandra Slavkovi´c。针对二项数据的差分隐私一致最强检验。载于神经信息处理系统进展，第 31 卷，2018 年。

Jordan Awan 和 Yuguo Wang。基于模拟的隐私化数据有限样本推断。美国统计学会会刊，120(551):1669–1682，2025 年。doi: 10.1080/01621459.2024.2427436。Borja Balle 和 Yu-Xiang Wang。改进差分隐私的高斯机制：解析校准与最优去噪。载于第 *35* 届国际机器学习会议（*ICML*）论文集，第 80 卷机器学习研究论文集，第 403–412 页。PMLR，2018 年。Garrett Bernstein 和 Daniel Sheldon。指数族的差分隐私贝叶斯推断。载于神经信息处理系统进展，第 31 卷，第 2924–2934 页，2018 年。T. Tony Cai、Yichen Wang 和 Linjun Zhang。隐私的代价：差分隐私下参数估计的最优收敛速率。统计学年鉴，49(5):2825–2850，2021。Christian Covington、Xi He、James Honaker 和 Gautam Kamath。差分隐私下的无偏统计估计与有效置信区间。华人统计学报，35:651–670，2025 年。Frances Ding、Moritz Hardt、John Miller 和 Ludwig Schmidt。Retiring Adult：用于公平机器学习的新数据集。载于神经信息处理系统进展，第 34 卷，第 6478–6490 页，2021。Jinshuo Dong、Aaron Roth 和 Weijie J. Su。高斯差分隐私。皇家统计学会会刊*B*辑：统计方法学，84(1):3–37，2022。doi: 10.1111/rssb.12454。Konstantin Donhauser、Javier Abad、Neha Hulkund 和 Fanny Yang。利用最优传输与粒子梯度下降进行隐私保护数据发布。载于第 *41* 届国际机器学习会议（*ICML*）论文集，2024。

John C. Duchi、Michael I. Jordan 和 Martin J. Wainwright。局部隐私估计的极小极大最优程序。美国统计学会会刊，113（521）：182–201，2018 年。

Cynthia Dwork 和 Aaron Roth。差分隐私的算法基础，第 9 卷，收录于理论计算机科学基础与趋势。Now Publishers，2014 年。

Cynthia Dwork、Frank McSherry、Kobbi Nissim 和 Adam Smith。为隐私数据分析中的敏感度校准噪声。载于理论密码学会议（*TCC*），第 265–284 页。施普林格，2006 年。

Cecilia Ferrando、Shufan Wang 和 Daniel Sheldon。面向差分隐私置信区间的参数自助法。载于国际人工智能与统计会议（*AISTATS*）第 *25* 届会议论文集，机器学习研究论文集第 151 卷，第 1598–1618 页。PMLR，2022 年。Gautam Kamath、Argyris Mouzakis、Matthew Regehr 和 Vatsal Singhal。统计估计中的偏差-准确性-隐私三难困境。美国统计学会会刊，120(552)：2338–2349，2025 年。doi：10.1080/01621459.2024.2443275。Vishesh Karwa 和 Aleksandra Slavkovi´c。使用加噪度数进行推断：差分隐私$beta$模型与合成图。统计年刊，44(1)，2016 年。doi：10.1214/15-AOS1358。Vishesh Karwa 和 Salil P. Vadhan。有限样本差分隐私置信区间。载于第 *9* 届理论计算机科学创新大会（*ITCS*），*LIPIcs*第 94 卷，第 44:1–44:9 页。Schloss Dagstuhl - Leibniz-Zentrum für Informatik，2018 年。doi：10.4230/LIPIcs.ITCS.2018.44。Chao Li、Gerome Miklau、Michael Hay、Andrew McGregor 和 Vibhor Rastogi。矩阵机制：在差分隐私下优化线性计数查询。*VLDB*期刊，24(6)：757–781，2015 年。Ryan McKenna、Gerome Miklau 和 Daniel Sheldon。赢得美国国家标准与技术研究院竞赛：一种可扩展的通用差分隐私合成数据方法。隐私与保密期刊，11(3)，2021 年。Ryan McKenna、Brett Mullins、Daniel Sheldon 和 Gerome Miklau。AIM：一种用于差分隐私合成数据的自适应迭代机制。*VLDB*基金会论文集，15(11)：2599–2612，2022 年。doi：10.14778/3551793.3551817。Ileana Montoya Perez、Parisa Movahedi、Valtteri Nieminen、Antti Airola 和 Tapio Pahikkala。差分隐私合成数据是否能带来合成发现

？医学信息方法，2024 年。doi：10.1055/a-2385-1355。在线发表于 2024 年 9 月 9 日。

Ossi Räisä、Joonas Jälkö、Samuel Kaski 和 Antti Honkela。使用差分隐私合成数据进行噪声感知统计推断。载于第 *26* 届人工智能与统计国际会议论文集（*AISTATS*），机器学习研究论文集第 206 卷，第 3620–3643 页。PMLR，2023。

Adam Smith。具有最优收敛速率的隐私保护统计估计。载于第 *43* 届 *ACM* 计算理论研讨会年会论文集（*STOC*），第 813–821 页。ACM，2011。

Yuchao Tao、Ryan McKenna、Michael Hay、Ashwin Machanavajjhala 和 Gerome Miklau。差分隐私合成数据生成算法的基准测试。*arXiv* 预印本 *arXiv:2112.09238*，2022。

Zhanyu Wang、Guang Cheng 和 Jordan Awan。差分隐私自助法：新的隐私分析与推断策略。机器学习研究杂志，26:1–57，2025。

Larry Wasserman 和 Shuheng Zhou。《差分隐私的统计框架》。美国统计学会会刊，105(489):375–389，2010。doi: 10.1198/jasa.2009.tm08651。

指数族中基于差分隐私充分统计量的噪声校准推断（补充材料）

Amir Asiaee[1] **Samhita Pal**[1]

[1]范德堡大学医学中心生物统计学系，美国田纳西州纳什维尔 {amir.asiaeetaheri, samhita.pal}@vumc.org

定理 2（渐近分布与方差膨胀）的证明

证明。 回忆一下， $\mu(\theta) := \nabla A(\theta) = \mathbb{E}_\theta[s(X)]$ 是均值参数映射，而 $(\nabla A)^{-1}$ 是它的逆映射。我们将其分解为：

$$\widetilde{\theta}_{\mathrm{plug}} - \theta_0 = (\nabla A)^{-1}(\widetilde{\bar{S}}) - \theta_0 = (\nabla A)^{-1}(\bar{S} + Z) - \theta_0.$$

步骤 *1*（抽样项）。 根据大数定律， $\bar{S} \xrightarrow{p} \mu(\theta_0)$。对 $s(X_1), \ldots, s(X_n)$ 应用多元中心极限定理，

$$\sqrt{n}\big(\bar{S} - \mu(\theta_0)\big) \xrightarrow{d} \mathcal{N}\big(0, \mathrm{Cov}_{\theta_0}[s(X)]\big) = \mathcal{N}(0, I(\theta_0)),$$

其中，最后一个等式利用了极小指数族的恒等式 $\mathrm{Cov}_{\theta_0}[s(X)] = \nabla^2 A(\theta_0) = I(\theta_0)$。由于 $(\nabla A)^{-1}$在 $\mu(\theta_0)$ 处可微，且其雅可比矩阵为 $I(\theta_0)^{-1}$，根据德尔塔方法可得 $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, I(\theta_0)^{-1})$。

步骤 *2*（隐私项）。对 $(\nabla A)^{-1}$ 在 $\bar{S}$ 处进行一阶泰勒展开，

$$\widetilde{\theta}_{\mathrm{plug}} - \hat{\theta} = (\nabla A)^{-1}(\bar{S} + Z) - (\nabla A)^{-1}(\bar{S}) = I(\hat{\theta})^{-1} Z + O_p(\|Z\|^2).$$

由于 $Z \sim \mathcal{N}(0, \sigma^2 I_d)$，我们有 $\|Z\| = O_p(\sigma)$，因此余项为 $O_p(\sigma^2) = o_p(\sigma)$。因为 $I(\hat{\theta}) \xrightarrow{p} I(\theta_0)$，所以我们有 $\widetilde{\theta}_{\mathrm{plug}} - \hat{\theta} = I(\theta_0)^{-1} Z + o_p(\sigma)$。

步骤 *3*（合并）。抽样项 $\theta - \theta_0$ 仅取决于 $\{X_i\}$，而 Z 与数据无关。因此这两项的协方差为零，并且

$$\sqrt{n}(\widetilde{\theta}_{\mathrm{plug}} - \theta_0) \xrightarrow{d} \mathcal{N}\Big(0,\ I(\theta_0)^{-1} + n\sigma^2 I(\theta_0)^{-2}\Big).$$

该速率分解源于这样一个事实：在高斯机制校准下， σ 与敏感度 $\Delta_2 = 2B/n$ 成正比（相差 $\log(1/\delta)$ 因子），从而得到有界问题的 $O_p((n\varepsilon)^{-1})$ 隐私失真率。 □

B 命题 1 的证明（一阶等价性）

证明。代入估计量满足 $\nabla A(\theta) = S$，而噪声感知估计量最小化加权二次型 $Q(\theta) = (\bar{S} - \nabla A(\theta))^\top \Sigma(\theta)^{-1} (\bar{S} - \nabla A(\theta))$，其中 $\Sigma(\theta) = I(\theta)/n + \sigma^2 I$。噪声感知的一阶条件为 $\nabla A(\theta)^\top [\nabla^2 A(\theta)] \Sigma(\theta)^{-1} (\widetilde{\bar{S}} - \nabla A(\theta)) = 0$。在代入解处， $\widetilde{\bar{S}} - \nabla A(\widetilde{\theta}_{\mathrm{plug}}) = 0$，这也使该表达式为零。因此，两个估计量在 leading order 下满足相同的方程。

更具体地，将 $Q(\theta)$ 在 θ_{plug} 处展开。差值 $\theta_{\mathrm{NA}} - \theta_{\mathrm{plug}}$ 满足牛顿型更新，其大小受 $\left\|\Sigma(\widetilde{\theta}_{\mathrm{plug}})^{-1} - \Sigma(\theta_0)^{-1}\right\| = O_p(n^{-1/2} + \sigma)$ 乘以低阶项所控制。根据标准的 M 估计论证，可得 $\widetilde{\theta}_{\mathrm{NA}} - \widetilde{\theta}_{\mathrm{plug}} = o_p(n^{-1/2}) + o_p(\sigma)$。 □

C 定理 3 的证明（下界）

^证明。令 $X \in \{-1, +1\}$，并将均值参数写为 $\mu(\theta) = \mathbb{E}_\theta[X] = \tanh(\theta)$。令 θ 为任意一个基于 n 个独立同分布样本的 (ε, δ)-差分隐私估计量，用其定义经后处理得到的均值估计量 $\widehat{\mu} := \tanh(\widehat{\theta})$。根据后处理不变性，$\widehat{\mu}$ 也满足 (ε, δ)-差分隐私。由于 tanh 是 1-Lipschitz 函数，因此逐点不等式 $(\widehat{\mu} - \mu(\theta))^2 \le (\widehat{\theta} - \theta)^2$ 成立，从而

$$\mathbb{E}_\theta\left[(\widehat{\theta} - \theta)^2\right] \;\ge\; \mathbb{E}_\theta\left[(\widehat{\mu} - \mu(\theta))^2\right].$$

因此，只需给出中心化（ε，δ）差分隐私下极小极大均值估计风险的下界。Cai 等人 [2021] 关于有界均值估计的极小极大下界表明，对于 $\varepsilon \in (0, 1]$ 和 $\delta < 1/4$，

$$\inf_{\widehat{\mu}\ (\varepsilon,\delta)\text{-DP}} \sup_{\mu \in [-1,1]} \mathbb{E}\left[(\widehat{\mu} - \mu)^2\right] \;\ge\; \frac{c}{n^2\varepsilon^2}$$

其中 $c > 0$ 是一个普适常数。将此下界应用于我们的伯努利族，并结合上述不等式，即可得到所述的 θ 下界。

□

D 补充实验细节

重复试验次数。所有模拟实验均采用足够的重复试验次数，以获得稳定的蒙特卡洛估计：实验 1 采用 5,000 次重复试验，实验 2–3 采用 1,000 次，实验 4–5 采用 2,000 次，实验 6（真实数据）采用 500 次，实验 7 采用 1,000 次。差分隐私参数自助法在模拟实验中使用 $B_{\text{boot}} = 500$ 次抽样，在真实数据实验中使用 $B_{\text{boot}} = 200$。

运行时间。在一台配备 8 个 CPU 核心的机器上，完整实验套件（实验 1–7，包含所有模型和配置）约需 4–6 小时。单项实验耗时从 2 分钟（仅使用高斯分布的实验 1）到 90 分钟不等（实验 6，真实数据，16 种 (n, ε) 配置各进行 500 次重复试验）。与代入式估计量相比，噪声感知估计量只增加了可忽略不计的开销，因为两者在每次重复试验中都只需执行一次优化步骤。

软件。所有实验均使用 Python、NumPy 和 SciPy 实现。实验 6 中的 ACS 数据通过 Folktables 软件包获取 [Ding et al., 2021]。