用于本地差分隐私框架的改进的 Christofides 机制

She Sun 人工智能研究院之江实验室 中国浙江杭州
sunshe@zhejianglab.com

Li Zhou 人工智能研究院之江实验室 中国浙江杭州
zhou.li@zhejianglab.com

Xiaoran Yan*人工智能研究院 之江实验室 中国浙江杭州
yanxr@zhejianglab.com

arXiv:2303.00857v2 [cs.CR] 3 Mar 2023

摘要

互联网技术的发展使得能够针对整个总体而非一定数量的样本进行分析，并导致对隐私保护的需求日益增加。本地差分隐私（LDP）是一种有效的隐私度量标准；然而，其均值估计的较大方差给应用带来了挑战。为解决这一问题，本文提出了一种新的本地差分隐私方法，即改进的 Christofides 机制。

本文比较了四种用于敏感主题调查的统计调查方法：改进的 Warner、Simmons、Christofides 和改进的 Christofides 机制。具体而言，我们对 Warner、Simmons 和 Christofides 机制进行了改进，使其能够从总体中不放回地抽取样本，以降低方差。此外，我们基于改进的 Christofides 机制，通过不放回抽取卡片，引入了一种称为改进的 Christofides 机制的新机制。研究发现，在将本地差分隐私作为隐私泄露的衡量指标时，在某些假设条件下，该机制的方差最小。现实世界中通常确实满足这一假设。实际上，在基于 HCOVANY 数据集（一个来自 IPUMS USA 的现实世界数据集）进行的实验中，我们将方差降低至改进的 Christofides 机制方差的 28.7%。这意味着，以本地差分隐私作为隐私泄露的衡量指标时，我们的方法能够获得更准确的估计。据我们所知，这是首次基于四种机制的对比分析，并将本地差分隐私作为统一的隐私泄露衡量指标，为本地差分隐私框架提出改进的 Christofides 机制。

PVLDB 引用格式：
She Sun、Li Zhou 和 Xiaoran Yan。面向本地差分隐私框架的改进的 Christofides 机制。PVLDB，16(1)：XXX-XXX，2023。
doi:XX.XX/XXX.XX

PVLDB 工件可用性：
源代码、数据和/或其他资源已发布于 URL_TO_YOUR_ARTIFACTS。

*通讯作者：yanxr@zhejianglab.com。本作品采用知识共享署名-非商业性使用-禁止演绎 4.0 国际许可协议授权。请访问 https://creativecommons.org/licenses/by-nc-nd/4.0/ 查看该许可协议的副本。如需进行超出该许可协议范围的任何使用，请发送电子邮件至 info@vldb.org 获取许可。版权所有权归所有者/作者所有。出版权授予 VLDB基金会。VLDB基金会会刊，第 16 卷，第 1 期，ISSN 2150-8097。doi:XX.XX/XXX.XX

1 引言

本地差分隐私（LDP）在 [24] 中作为第一个等价定义提出，并由 Duchi 等人 [11], 进一步发展。它是一种算法，通过让用户在本地随机扰动数据，并将扰动后的数据发送给（可能不可信的）数据收集方，从而量化隐私。由于数据收集方无法获取原始个人数据，这是一种强隐私模型 [3–5, 7, 9, 29, 30,33, 34]。许多行业组织都采用了本地差分隐私，包括 Apple [26], 谷歌 [13], 和微软 [10]。例如，Apple [26] 在 iOS 上部署本地差分隐私，以了解受欢迎的健康数据类型，从而改进未来的健康应用。谷歌 [13] 已在 Chrome 网页浏览器中采用 RAPPOR，向约 1400 万名选择向谷歌发送使用统计信息的受访者收集 Chrome 客户端数据。微软 [10] 在 Windows 10 中集成本地差分隐私，以收集用户使用特定应用的秒数。

在社会科学和流行病学研究中，对欺骗、赌博、滥用毒品或酒精以及反社会行为等敏感问题进行调查，可能会导致受访者拒答或给出不真实的回答。为解决这一问题，Warner 等人 [31] 于 1965年提出了一种称为随机应答的程序，这是最经典的本地差分隐私机制。此后，人们提出了许多用于调查敏感问题的机制 [8, 14, 16–18, 21, 22] ，其中两个具有代表性的机制是 [17]和 [8]。Simmons 等人 [17] 于 1967年提出，如果使用两个互不相关的问题（或陈述），合作程度将会提高。Christofides 等人 [8] 于 2003年提出了一种广义随机应答技术，用户只需回答数字，而不必回答是或否。

Waseda 等人 [32] 使用差分隐私评估了 Warner、Kuk [18], 负向调查机制 [14] 及其变体 [2] 。他们指出，这些机制在隐私与效用之间存在权衡，但既未比较不同机制的性能，也未提出改进方法。Giordano 等人 [15] 比较了在 Lanke [19], Leysi- effer 和 Warner [20], 提出的保密性度量水平相同的情况下，三种针对无关二分问题的机制（ [17, 23, 25]）的估计量方差，但同样没有提出降低方差的改进方法。

随着大数据和人工智能的发展，人们可以不再通过抽样，而是通过调查整个总体来收集数据，例如2020年美国人口普查和2020年美国总统选举。基于此，我们修改了三种机制（Warner、Simmons 和 Christofides 机制），通过从总体中不放回地抽取样本来开展敏感主题调查。由于从总体中不放回地抽取样本可以减少我们案例中估计均值的随机性，

我们改进了修改的 Christofides 机制，方法是从装置中不放回地抽取卡片。改进效果显著：如果使用本地差分隐私作为衡量隐私泄露的统一标准，那么当具有敏感属性的人群比例较小时，改进的 Christofides 机制的方差最小。事实上，根据我们基于真实世界数据集的实验，其方差降至修改的 Christofides 机制方差的 28.7%。因此，我们提出了一种新的本地差分隐私均值估计机制，它具有更小的方差，并能更好地促进数据提供者的合作。

贡献。我们的贡献主要有三点：

（1）我们通过从总体中不放回抽样，对改进的 Warner、Simmons 和 Christofides 机制进行分析。

（2）当具有敏感属性的人群比例较小时，我们对改进的 Christofides 机制作出了一项虽小但有效的调整，并提出了一种改进版 Christofides 机制。

（3）我们以本地差分隐私作为相同的隐私泄露度量，通过理论分析和数值模拟比较了改进的 Warner、Simmons、Christofides 以及改进版 Christofides 机制的方差。我们发现，当具有敏感属性的人群比例较小时，改进版 Christofides 机制的方差最小，并且可作为更优的本地差分隐私机制替代 Warner 机制。

文章结构。第2节介绍背景，包括本地差分隐私的定义，以及 Warner、Simmons 和 Christofides 机制的流程。第3节通过理论分析介绍改进的 Warner、Simmons 和 Christofides 机制的流程及方差。第4节介绍改进版 Christofides 机制并分析其方差。第5节以本地差分隐私作为隐私泄露的度量，分析改进的 Warner、Simmons、Christofides 以及改进版 Christofides 机制的方差。第6节通过理论分析和数值模拟比较四种机制的方差和最小样本量。第7节给出结论。

2 背景

2.1 本地差分隐私

在本地差分隐私设置中，用户使用算法 A 对私有值 x 进行扰动，并将 A(x) 发送给不可信收集方。收集方获得有关用户的统计信息。

定义 *2.1*（ε-本地差分隐私 *[11]*）。当且仅当对于任意一对输入值 x、x' 以及所有随机化输出 O，随机化机制 A 满足 ε-本地差分隐私时，有

$$\mathbf{P}[\mathcal{A}(x) = O] \leq e^{\varepsilon} \times \mathbf{P}[\mathcal{A}(x') = O].$$

符号 **P**表示概率。

2.2 三种随机应答机制

本节简要介绍 Warner 机制、Simmons 机制和 Christofides机制。

2.2.1 Warner 机制 *[31].* 假设总体中的每位受访者都属于组 A 或 $\bar{A}$，且需要通过调查估计组 A 的比例。每位受访者都需要对以下两项陈述之一作出回答 *yes* 或 *no*：

(a) 我是群体 A的一员。
(b) 我是群体 $\bar{A}$的一员。

受访者使用随机装置（例如通过掷一枚有偏硬币）在没有调查者在场的情况下，以概率 p（本文中为$p < 1/2$）回答陈述 (a)，并以概率 $1 - p$ 回答陈述 (b)。调查者（例如政府或学校）可以根据回答，对具有敏感属性的人群真实比例进行无偏估计。

2.2.2 Simmons 机制 *[17]*。值得注意的是，Warner 机制中的两个陈述都涉及敏感内容。在 Simmons 机制中，我们将 Warner 机制的第二个陈述改为一个无关陈述，因此与 Warner 机制相比，受访者的配合度更高。

Simmons 机制使用两个陈述：
(a) 我是 Group A 的成员。
(b) 我是 Group B 的成员。
，其中 Group B 中无关的陈述并不敏感。例如，提出的两种陈述可能是：
(a) 我在考试中作弊了。
(b) 我出生于上半年。

调查者（例如政府或学校）可以根据这些回答，获得对具有敏感属性的人群真实比例的无偏估计量。

*2.2.3 Christofides*机制 *[8]*。在 Christofides机制中，每位受访者都会得到一张卡片，该卡片按比例 p_1、p_2、...、p_L 分别生成整数 1、2、...、L（这些比例并不完全相等）。如果受访者具有敏感属性，则报告该整数与 $L + 1$ 相差多远；如果不具有敏感属性，则报告该整数与 0 相差多远。例如，假设 L 为 3，且受访者报告数字 3。这意味着受访者可能具有敏感属性且抽取的数字为 1，也可能不具有敏感属性且抽取的数字为 3。注意，当 $L = 2$时，Christofides机制与 Warner 机制基本等价。

调查者（例如政府或学校）可以根据这些回答，获得具有敏感属性人群真实比例的无偏估计。

3 分析三种修改后的机制

本节分析改进的 Warner、Simmons 和 Christofides 机制。我们假设整个流程由受访者完成，且调查者无法观察到这一过程，并且所有受访者都如实作答。

在 [31] [17] 和 [8], 中，作者假设从总体中进行有放回随机抽样。然而，随着互联网的发展，人们可以让样本数与总体数相等，例如美国人口普查 2020 [6] 和美国总统选举，这意味着进行无放回抽样。

令 x_i 表示受访者 i的敏感属性对应的整数：如果受访者属于敏感群体（即访谈时没有医疗保险覆盖的人群），则为 1；如果受访者不属于敏感群体，

（即在我们的实验中，访谈时有医疗保险覆盖的人群）。我们有

$$x_i = \begin{cases} 1, & \text{respondent } i \in \text{ Group } A \\ 0, & \text{respondent } i \in \text{ Group } \bar{A}. \end{cases}$$

3.1 分析改进的 Warner 机制

算法 1 展示了改进的 Warner 机制。我们从大小为 N 的总体中，不放回地抽取一个大小为 N 的随机样本。令 π_A 为总体中属于群体 A 的真实比例。

算法 1 改进的 Warner 机制

Input: $x_i (i = 1, 2, ..., N), p$
输出： X_i

抽取一个伯努利变量 u，使得 $P(u = 1) = p$
如果 $u < p$ **则**
    $X_i = x_i$
否则
    $X_i = 1 - x_i$
结束 if
返回 X_i
与下一位受访者进行实验 $i + 1$。

引入随机变量

$$X_i = \begin{cases} 1, & \text{the } i \text{ th respondent answers } yes \\ 0, & \text{the } i \text{ th respondent answers } no. \end{cases}$$

随机变量 $X_1, X_2, , X_N$ 彼此相关。从 N 位受访者中获得的是回答数为 $N_1 = \acute{\text{I}}_{i=1}^{N} X_i$。

根据全概率公式，回答问题 (a) 或 (b) 的受访者的理论比例约为 N_1/N，如下所示， yes

$$p\pi_A + (1-p)(1-\pi_A) \approx \frac{N_1}{N}.$$

参数 $\hat{}\pi_A$ 根据所有受访者的间接回答，通过估计量进行估计

$$\hat{\pi_A} = \frac{\frac{N_1}{N} - (1-p)}{2p-1}.$$

根据大数定律 [12]，估计量 $\hat{}\pi_A$ 是无偏估计。

$\hat{}\pi_A$ [28] 的方差为

$$\begin{aligned} \mathbf{Var}\hat{\pi_A} &= \frac{\mathrm{Var}(\sum_{i=1}^{N} X_i)}{N^2(2p-1)^2} \\ &= \frac{\pi_A p(1-p) + (1-\pi_A)(1-p)p}{N(2p-1)^2} \\ &= \frac{p(1-p)}{N(2p-1)^2}. \end{aligned} \tag{1}$$

与 [31] 有放回抽样相比，方差降低了 $[\pi_A(1-\pi_A)]/N$，即

$$\mathbf{Var}\hat{\pi_A} = \frac{\pi_A(1-\pi_A)}{N} + \frac{p(1-p)}{N(2p-1)^2}.$$

3.2 修改的 Simmons 机制分析

算法2展示了修改的 Simmons 机制。我们从大小为 N 的总体中抽取一个大小为 N 的不放回随机样本。设 π_A 为总体中属于A组的真实比例。

算法2 修改的 Simmons 机制

输入： $x_i (i = 1, 2, ..., N)$ p, π_B
输出： X_i

抽取一个伯努利变量 u，使得 $P(u = 1) = p$
如果 $u < p$ **则**
    $X_i = x_i$
否则
    抽取一个伯努利变量 v，使得 $P(v = 1) = \pi_B$
    如果 $v < \pi_B$ **那么**
        $X_i = 1$
    else
        $X_i = 0$
    结束如果
结束如果
返回 X_i
对下一位受访者 $i + 1$进行实验。

引入随机变量

$$X_i = \begin{cases} 1, & \text{the } i \text{ th respondent answers } yes \\ 0, & \text{the } i \text{ th respondent answers } no. \end{cases}$$

随机变量 X_1、 X_2、 X_N.相关。来自 N 位受访者的 是 回答数量为 $N_1 = \acute{\text{I}}_{i=1}^{N} X_i$。

根据全概率公式，受访者对问题 (a) 或 (b) 回答 yes 的理论比例近似等于 N_1/N，如下所示:

$$p\pi_A + (1-p)\pi_B \approx \frac{N_1}{N}.$$

参数 π_A 是根据所有受访者的间接回答，通过估计量进行估计的

$$\hat{\pi_A} = \frac{\frac{N_1}{N} - (1-p)\pi_B}{p}.$$

根据大数定律 [12]，估计量 $\hat{}\pi_A$ 是无偏估计量。

$\hat{}\pi_A$ 的方差为

$$\begin{aligned} \mathbf{Var}\hat{\pi_A} =& \frac{\mathbf{Var}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right)}{p^2} \\ =& \frac{\pi_A\left[p + (1-p)\pi_B\right]\left[1 - p - (1-p)\pi_B\right]}{Np^2} + \\ & \frac{(1-\pi_A)\left[(1-p)\pi_B\right]\left[1 - (1-p)\pi_B\right]}{Np^2} \\ =& \frac{\pi_B(1-p) - \pi_B^2(1-p)^2}{Np^2} + \\ & \pi_A\frac{1 - p - 2\pi_B(1-p)}{Np}. \end{aligned} \tag{2}$$

与采用有放回抽样的 [17] 相比，方差降低了 $[\pi_A(1-\pi_A)]/N$，

$$\mathbf{Var}\hat{\pi_A} = \frac{\pi_A(1-\pi_A)}{N} + \frac{\pi_B(1-p)-\pi_B^2(1-p)^2}{Np^2} + \pi_A\frac{1-p-2\pi_B(1-p)}{Np}.$$

3.3 分析修改的 Christofides 机制

算法 3 阐述了修改的 Christofides 机制。我们从总体大小为 N 的总体中进行有放回抽样，抽取大小为 N 的样本。为了以数学方式表述该算法，准备 M 张卡片，使其分别以比例 p_1、p_2、……、p_L 产生整数 1、2、……、L（这些比例并不全都相等）。令 π_A 表示总体中属于群组 A 的真实比例。

算法 3 修改的 Christofides 机制

输入： x_i（i = 1、2、……、N）L、p_1、p_2、……、p_L

输出： X_i

1: 准备一个由 M 张卡片组成的装置，每张卡片上标有一个整数 1，2，……，L，其比例为 p_1，p_2，……，p_L，分别;
2: 随机选择一张卡片，假设其编号为 k;
3: **如果** $x_i = 1$，**则**
4: $X_i = k$;
5: **否则**
6: $X_i = L + 1 - k$
7: **结束 if**
8: 返回 X_i;
9: 将卡片放回装置。
10: 对下一位受访者进行实验 $i + 1$。

为方便起见，我们引入以下记号：Y_i 是他/她抽取的数字，X_i 是他/她回答的数字。我们有

$$X_i = \begin{cases} 1, & \text{the } i \text{ th respondent answers } yes \\ 0, & \text{the } i \text{ th respondent answers } no \end{cases}$$

，其中

$$X_i = (L + 1 - Y_i)\, x_i + Y_i\,(1 - x_i)\,.$$

X 的期望为

$$\mathbf{E}X = \mathbf{E}Y + \pi(L + 1 - 2\mathbf{E}Y).$$

因此，π_A 的估计为

$$\hat{\pi}_A = \frac{\frac{1}{N}\sum_{i=1}^{N} X_i - \mathbf{E}(Y)}{L + 1 - 2\mathbf{E}(Y)}.$$

根据大数定律 [12]，估计量 $\hat{}\pi_A$ 是无偏估计。

定理 3.1（修改的 Christofides 机制方差）。$\hat{}$ 的方差是

$$\mathbf{Var}\hat{\pi_A} = \frac{\mathbf{Var}\left(\sum_{i=1}^{N} X_i\right)}{N^2(L + 1 - 2\mathbf{E}(Y))^2} = \frac{\mathbf{Var}Y}{N(L + 1 - 2\mathbf{E}(Y))^2}.$$

证明。 证明见附录A。 □

与采用有放回抽样的 [8] 相比，方差减少了 $[\pi_A(1-\pi_A)]/N$。

$$\mathbf{Var}\hat{\pi_A} = \frac{\pi_A(1-\pi_A)}{N} + \frac{\mathrm{Var}\,Y}{N(L + 1 - 2\mathbf{E}(Y))^2}.$$

通过对 Warner、Simmons 和 Christofides 进行无放回抽样改造，与采用有放回抽样的 [31] [17] 和 [8], 相比，方差减少了 $[\pi_A(1-\pi_A)]/N$。这是我们的首要贡献。

4 改进的CHRISTOFIDES机制

由于在我们的情形下，从总体中无放回地抽取随机样本可以降低估计均值的随机性。如果受访者无放回地抽取卡片，会发生什么？为澄清这一问题，我们提出了改进的 Christofides 机制，如算法 4 所示。

算法 4 改进的 Christofides 机制 输入：$x_i (i = 1, 2, ..., N)$ L, $p_1, p_2, ..., p_L$输出：X_i1：准备一个装置，其中包含 N张卡片，每张卡片上标有整数 1、2、...、L，其比例分别为 $p_1, p_2, ..., p_L$；2：随机抽取一张卡片，假设其数字为 k；3：如果 $x_i = 1$, 则 4：$X_i = k$；5：否则 6：$X_i = L + 1 - k$7：结束如果 8：返回 X_i；9：受访者保留这张卡片。（不要将卡片放回装置中。）10：对下一位受访者 $i + 1$进行实验。

我们从大小为 N 的总体中不放回地抽取大小为 N 的样本。为了用数学形式表述该算法，准备 N 张卡片，它们分别以 p_1、p_2、……、p_L 的比例产生整数 1、2、……、L（这些比例并非全部相等）。

为方便起见，我们引入以下记号：Y_i 表示他/她抽取的数量，X_i 表示他 / 她回答的数量。

我们有

$$X_i = \begin{cases} 1, & \text{the } i \text{ th respondent answers } yes \\ 0, & \text{the } i \text{ th respondent answers } no, \end{cases}$$

其中

$$X_i = (L + 1 - Y_i)\, x_i + Y_i\,(1 - x_i)\,.$$

X的期望为

$$\mathbf{E}X = \mathbf{E}Y + \pi(L + 1 - 2\mathbf{E}Y).$$

因此， π_A 的估计为

$$\hat{\pi}_A = \frac{\frac{1}{N}\sum_{i=1}^{N} X_i - \mathbf{E}(Y)}{L + 1 - 2\mathbf{E}(Y)}.$$

根据大数定律 [12]，估计量 $\hat{}\pi_A$ 是无偏估计。

定理 4.1（改进的 Christofides机制方差）。$\hat{}$的方差为

$$\begin{aligned}\mathbf{Var}\hat{\pi_A} &= \frac{\mathbf{Var}\left(\sum_{i=1}^{N} X_i\right)}{N^2(L + 1 - 2\mathrm{E}(Y))^2} \\ &= \frac{4\pi_A\,(1 - \pi_A)\,\mathbf{Var}(Y)}{(N-1)(L + 1 - 2\mathbf{E}(Y))^2}.\end{aligned} \tag{3}$$

证明。证明见附录 B。 □

由于改进的 Christofides 机制和改进的 Christofides机制对 π_A 的估计均为无偏估计，因此我们通过计算改进的 Christofides 机制（$\mathbf{Var}_{IC}$）与改进的 Christofides机制（$\mathbf{Var}_{MC}$）的方差之差，比较这两种机制的效用。

定理 4.2（通过理论分析比较修改的 Christofides 与改进的 Christofides 机制）。比较结果可总结如下：

(i.) When proportion π_A is in interval $\left(0, 1/2 - 1/2\sqrt{N}\right) \cup (1/2 + 1/2\sqrt{N},\ 1)$时，我们得到 **VarIC** 小于 **VarMC** 。

(ii.) 当比例 π_A 处于区间 $(1/2 - 1/2\sqrt{N},\ 1/2 + 1/2\sqrt{N})$时，我们得到 VarIC 大于 VarMC。

证明。

$$\begin{aligned}&\mathbf{Var_{IC}}\hat{\pi_A} - \mathbf{Var_{MC}}\hat{\pi_A} \\ &= \left[\frac{4\pi_A\,(1 - \pi_A)}{N-1} - \frac{1}{N}\right]\frac{\mathbf{Var}Y}{(L + 1 - 2\mathbf{E}Y)^2} \\ &= \left[\frac{-4N\pi_A^2 + 4N\pi_A - N + 1}{N(N-1)}\right]\frac{\mathbf{Var}Y}{(L + 1 - 2\mathbf{E}Y)^2} \\ &= -\frac{4(1/2 + 1/2\sqrt{N} - \pi_A)(1/2 - 1/2\sqrt{N} - \pi_A)\mathbf{Var}Y}{(N-1)(L + 1 - 2\mathbf{E}Y)^2}.\end{aligned}$$

只要 $\pi_A < 1/2 - 1/2\sqrt{N}$或 $\pi_A > 1/2 + 1/2\sqrt{N}$，我们就有

$$\mathbf{Var}_{IC}\hat{\pi_A} < \mathbf{Var}_{MC}\hat{\pi_A}. \quad \square$$

由于 N是一个较大的数，区间 $[1/2 - 1/2\sqrt{N},\ 1/2 + 1/2\sqrt{N}]$很小。例如，当 $N = 10^4$时，该区间为 $[0.495, 0.505]$. 特别是，由于 $1/2\sqrt{N} \le 1/4$（对于 $N \ge 3$），只要 $\pi_A < 0.25$ 或 $\pi_A > 0.75$，对于任意 $N(N > 3)$，我们就有 $\mathbf{Var}_{IC} < \mathbf{Var}_{MC}$。

此外，我们将在第 6 节证明，在以本地差分隐私作为隐私泄露的衡量标准时，在某些假设下，改进的 Christofides 机制的方差小于改进的 Warner 和 Simmons 机制的方差。该假设在现实世界中通常是成立的。我们通过不放回抽样卡片改进了 Christofides 机制。这是我们的第二项贡献。

5 使用本地差分隐私分析四种机制的方差

本节采用理论分析，研究四种机制的方差，并使用本地差分隐私作为衡量隐私泄露的统一标准。

5.1 使用本地差分隐私的改进 Warner 机制

定理5.1（Warner 机制）。改进的 *Warner* 机制满足 ε–差分隐私，其中 $\varepsilon = \ln[(1-p)/p]$ （不失一般性，假设 $p < 1/2$。）

证明。我们考虑四种情况：

情况1：受访者具有敏感属性，在此条件下回答是的概率为 p。

情况 2：受访者具有敏感属性，在此条件下回答否的概率为 $1-p$。

情况 3：受访者不具有敏感属性，在此条件下回答是的概率为 $1-p$。

情况 4：受访者不具有敏感属性，在此条件下回答否的概率为 p。

对于任意两个答案相同的受访者，比例的最大比值为 $(1-p)/p$。 □

将 $\varepsilon = \ln[(1-p)/p]$ 和 $p < 1/2$ 代入方程 (1)。方差 VarMW 可写为

$$\mathbf{Var}\hat{\pi_A} = \frac{e^{\varepsilon}}{N\,(e^{\varepsilon} - 1)^2}.$$

5.2 使用本地差分隐私的修改的 Simmons 机制

定理5.2（Simmons 机制）。修改后的 *Simmons* 机制满足 ε–差分隐私，其中

$$\varepsilon = \begin{cases} \ln\frac{p + (1-p)\pi_B}{(1-p)\pi_B} & ,\pi_B \le \frac{1}{2} \\ \ln\frac{p + (1-p)(1-\pi_B)}{(1-p)(1-\pi_B)} & ,\pi_B > \frac{1}{2}. \end{cases} \tag{4}$$

证明。我们考虑四种情况：

情况1：受访者具有敏感属性，在此条件下回答是的概率为 $p + (1-p)\pi_B$。

情况2：受访者具有敏感属性，在此条件下回答否的概率为 $(1-p)(1-\pi_B)$。

情况3：受访者不具有敏感属性，在此条件下回答是的概率为 $(1-p)\pi_B$。

情况4：受访者不具有敏感属性，在此条件下回答否的概率为 $p + (1-\pi_B)$。

对于答案相同的任意两位受访者，比例的最大比值为

$$\varepsilon = \begin{cases} \ln\frac{p + (1-p)\pi_B}{(1-p)\pi_B} & ,\pi_B \le \frac{1}{2} \\ \ln\frac{p + (1-p)(1-\pi_B)}{(1-p)(1-\pi_B)} & ,\pi_B > \frac{1}{2}. \end{cases} \quad \square$$

定理 5.3（修改的 Simmons 机制的方差与 epsilon 的关系）。当 $\pi_B = 1/2$ 时，方差 $\mathbf{Var}_{\mathrm{MS}}$ 取得如下最小值

$$\mathbf{Var}\hat{\pi_A} = \frac{e^{\varepsilon}}{N\left(e^{\varepsilon}-1\right)^2}.$$

这与改进的 Warner 机制相同。

Proof. *The proof is in Appendix C.* □

5.3 使用本地差分隐私的修改的 Christofides 机制

定理 5.4（修改的 Christofides 机制）。修改的 *Christofides* 机制满足 ε-差分隐私，其中

$$\varepsilon = \max\left(\ln\frac{p_{L+1-k}}{p_k}, k = 1, 2, \cdots, L\right).$$

证明。我们考虑四种情况：

情况 1：受访者具有敏感属性并抽取 k。在此条件下，回答 $L+1-k$ 的概率为 p_k。

情况 2：$L+1-k$ 受访者具有敏感属性并抽取 k。在此条件下，回答 k 的概率为 p_{L+1-k}。

情况 3：受访者不具有敏感属性，并抽取 k。在此条件下，回答 k 的概率为 p_k。

情况 4：受访者不具有敏感属性，并抽取 $L+1-k$。在此条件下，回答 $L+1-k$ 的概率为 p_{L+1-k}。

对于任意两个答案相同的受访者，概率的最大比值为

$$\varepsilon = \max\left(\ln\frac{p_{L+1-k}}{p_k}, k = 1, 2, \cdots, L\right).$$

□

我们固定数值 $L = 3$，因此隐私预算 ε 为

$$\varepsilon = \max\left\{\ln\frac{p_1}{p_3}, \ln\frac{p_3}{p_1}\right\}.$$

定理 5.5（修改的 Christofides 机制的方差与 epsilon 的关系）。对于情况 $L = 3$，修改的 *Christofides* 机制**方差**$_{\mathrm{MC}}$ 达到最小值

$$\mathbf{Var}\hat{\pi_A} = \frac{1}{4N}\left[\frac{(e^{\epsilon}+1)^2}{(e^{\epsilon}-1)^2(1-p_2)} - 1\right].$$

在比例满足以下条件时

$$p_1 = (1-p_2)/\left(e^{\varepsilon}+1\right), p_3 = \left[e^{\varepsilon}(1-p_2)\right]/\left(e^{\varepsilon}+1\right)$$

or

$$p_1 = \left[e^{\varepsilon}(1-p_2)\right]/\left(e^{\varepsilon}+1\right), p_3 = (1-p_2)/\left(e^{\varepsilon}+1\right).$$

证明。 证明见附录 D。 □

注意，当比例 $p_2 = 0$时，改进的 Christofides 机制与改进的 Warner 机制实质上等价。

5.4 使用本地差分隐私的改进的 Christofides 机制

定理5.6（改进的 Christofides 机制）。假设任何受访者都不知道其他人抽到的卡片以及其他人回答的数字。那么，改进版 *Christofides* 机制满足 ε-差分隐私，其中

$$\varepsilon = \max\left(\ln\frac{p_{L+1-k}}{p_k}, k = 1, 2, \cdots, L\right).$$

证明。我们考虑四种情况：

情况 1：受访者具有敏感属性并抽取 k。在此条件下回答 $L+1-k$ 的概率为 p_k。

情况 2：$L+1-k$ 受访者具有敏感属性并抽取 k。在此条件下回答的概率为 p_{L+1-k}。

情况 3：受访者不具有敏感属性并抽取 k。在此条件下回答 k 的概率为 p_k。

情况 4：受访者不具有敏感属性，并抽取 $L+1-k$。在此条件下回答 $L+1-k$ 的概率为 p_{L+1-k}。

对于答案相同的任意两名受访者，概率的最大比值为

$$\varepsilon = \max\left(\ln\frac{p_{L+1-k}}{p_k}, k = 1, 2, \cdots, L\right).$$

□

我们固定数值 $L = 3$，因此隐私预算 ε 为

$$\varepsilon = \max\left\{\ln\frac{p_1}{p_3}, \ln\frac{p_3}{p_1}\right\}.$$

定理5.7（改进的 Christofides 机制方差与 epsilon 的关系）。对于情况 $L = 3$，改进的 *Christofides* 机制的方差$\mathbf{Var}_{\mathrm{IC}}$ 达到最小值

$$\mathbf{Var}\hat{\pi_A} = \frac{\pi_A(1-\pi_A)}{(N-1)}\left[\frac{(e^{\epsilon}+1)^2}{(e^{\epsilon}-1)^2(1-p_2)} - 1\right]$$

在比例满足以下条件时

$$p_1 = (1-p_2)/\left(e^{\varepsilon}+1\right), p_3 = \left[e^{\varepsilon}(1-p_2)\right]/\left(e^{\varepsilon}+1\right)$$

or

$$p_1 = \left[e^{\varepsilon}(1-p_2)\right]/\left(e^{\varepsilon}+1\right), p_3 = (1-p_2)/\left(e^{\varepsilon}+1\right).$$

证明。证明与附录 D 中的证明类似。 □

6 使用本地差分隐私比较四种机制

由于改进的 Warner/Simmons 机制、Christofides 机制和改进的 Christofides 机制对 π_A的估计均为无偏估计，因此我们通过理论分析和数值模拟比较这四种机制的方差。

6.1 通过理论分析比较四种本地差分隐私机制

本节通过计算相同隐私预算 ε下的方差，比较这四种机制的效用。

定理 6.1（通过理论分析比较四种本地差分隐私机制）。比较结果可总结如下：

(i.) 当比例 π_A 处于区间 $(0,\pi_{A_1})\cup(\pi_{A_2},1)$内时，$\mathbf{Var_{IC}} < \mathbf{Var_{MW/MS}} < \mathbf{Var_{MC}}$。 π_{A_1} 和 π_{A_2} 见方程 *(5)*。

(ii.) 当比例 π_A 处于区间 $(\pi_{A_1}, 1/2-1/2\sqrt{N})\cup(1/2+1/2\sqrt{N}, \pi_{A_2})$内时，$\mathbf{Var_{MW/MS}} < \mathbf{Var_{IC}} < \mathbf{Var_{MC}}$。

(iii.) 当比例 π_A 处于区间 $(1/2-1/2\sqrt{N},1/2+1/2\sqrt{N})$内时，$\mathbf{Var_{MW/MS}} < \mathbf{Var_{MC}} < \mathbf{Var_{IC}}$。

证明。首先，根据定理 4.2，当比例 π_A处于区间 $(1/2-1/2\sqrt{N}, 1/2+1/2\sqrt{N})$ 时，我们有 $\mathbf{Var_{IC}} > \mathbf{Var_{MC}}$；而当比例 π_A 处于区间 $(0,\ 1/2-1/2\sqrt{N})\ \cup(1/2+1/2\sqrt{N},\ 1)$ 时，$\mathbf{Var_{IC}} < \mathbf{Var_{MC}}$。

其次，我们比较改进的 Christofides机制的方差 $\mathbf{Var_{MC}}$ 与改进的 Warner/Simmons 机制的方差（$\mathbf{Var_{MW/MS}}$）。

由于

$$\begin{aligned}&\mathbf{Var_{MC}}\hat{\pi_A} - \mathbf{Var_{MW/MS}}\hat{\pi_A}\\ &= \frac{(e^\varepsilon+1)^2 p_2}{4N(e^\varepsilon-1)^2(1-p_2)},\\ &\geq 0.\end{aligned}$$

改进的 Christofides机制的方差 $\mathbf{Var_{MC}}$ 大于改进的 Warner/Simmons 机制的方差 $\mathbf{Var_{MW/MS}}$。

接下来，我们比较改进的 Christofides 机制的方差 $\mathbf{Var_{IC}}$ 与改进的 Warner/Simmons 机制的方差 $\mathbf{Var_{MW/MS}}$。改进的 Christofides 机制与改进的 Warner/Simmons 机制的方差之差为

$$\begin{aligned}&\mathbf{Var_{IC}}\hat{\pi_A} - \mathbf{Var_{MW/MS}}\hat{\pi_A}\\ &= \frac{\pi_A(1-\pi_A)}{N-1}\left[\frac{(e^\varepsilon+1)^2}{(e^\varepsilon-1)^2(1-p_2)}-1\right] - \frac{e^\varepsilon}{N(e^\varepsilon-1)^2}\\ &= \frac{\left[4e^\varepsilon+(e^\varepsilon-1)^2 p_2\right]\pi_A(1-\pi_A) - (N-1)(1-p_2)e^\varepsilon}{N(N-1)(e^\varepsilon-1)^2(1-p_2)}\end{aligned}$$

Let

$$a = 4e^\varepsilon + \left(e^\varepsilon-1\right)^2 p_2, c = (N-1)(1-p_2)e^\varepsilon$$

我们有

$$\begin{aligned}&\mathbf{Var_{IC}}\hat{\pi_A} - \mathbf{Var_{MW/MS}}\hat{\pi_A}\\ &= \frac{-a\pi_A^2 + a\pi_A - c}{N(N-1)(e^\varepsilon-1)^2(1-p_2)}\\ &= \frac{-a\left(\pi_A-\pi_{A_1}\right)\left(\pi_A-\pi_{A_2}\right)}{N(N-1)(e^\varepsilon-1)^2(1-p_2)}\end{aligned}$$

$$\begin{aligned}\pi_{A_1} &= \frac{a-\sqrt{\Delta}}{2a}\\ \pi_{A_2} &= \frac{a+\sqrt{\Delta}}{2a}\\ \Delta &= a^2-4ac.\end{aligned}$$

$$\begin{aligned}\pi_{A_1} &= \frac{1}{2}-\frac{1}{2}\sqrt{1-\frac{4c}{a}}\\ &= \frac{1}{2}-\frac{1}{2}\sqrt{1-\frac{4e^\varepsilon(N-1)(1-p_2)}{N\left[4e^\varepsilon+p_2\left(e^{2\varepsilon}-2e^\varepsilon+1\right)\right]}}\\ &= \frac{1}{2}-\frac{1}{2}\sqrt{1-\frac{(N-1)(1-p_2)}{N\left[1+\frac{p_2}{4}\left(e^\varepsilon+\frac{1}{e^\varepsilon}-2\right)\right]}}.\end{aligned}$$

$$\begin{aligned}\pi_{A_2} &= \frac{1}{2}+\frac{1}{2}\sqrt{1-\frac{4c}{a}}\\ &= \frac{1}{2}+\frac{1}{2}\sqrt{1-\frac{4e^\varepsilon(N-1)(1-p_2)}{N\left[4e^\varepsilon+p_2\left(e^{2\varepsilon}-2e^\varepsilon+1\right)\right]}}\\ &= \frac{1}{2}+\frac{1}{2}\sqrt{1-\frac{(N-1)(1-p_2)}{N\left[1+\frac{p_2}{4}\left(e^\varepsilon+\frac{1}{e^\varepsilon}-2\right)\right]}}.\end{aligned}$$

Let

$$d = e^\varepsilon + \frac{1}{e^\varepsilon} - 2.$$

So

$$\begin{aligned}\pi_{A_1} &= \frac{1}{2}-\frac{1}{2}\sqrt{1-\frac{N-1}{N}\cdot\frac{1-p_2}{1+\frac{p_2 d}{4}}}\\ \pi_{A_2} &= \frac{1}{2}+\frac{1}{2}\sqrt{1-\frac{N-1}{N}\cdot\frac{1-p_2}{1+\frac{p_2 d}{4}}}.\end{aligned} \tag{5}$$

只要 $\pi_A < \pi_{A_1}$ 或 $\pi_A > \pi_{A_2}$，我们就有

$$\mathbf{Var_{IC}}\hat{\pi_A} - \mathbf{Var_{MW/MS}}\hat{\pi_A} < 0. \quad \square$$

改进的 Christofides 机制在区间 $[\pi_{A_1},\pi_{A_2}]$上的表现最差。区间 $[\pi_{A_1},\pi_{A_2}]$取决于 N、ε和 p_2。由于该区间关于 0.5 对称，在固定 $N=10^4$后，我们计算了不同的比例 p_2和隐私预算 ε 下的区间长度。结果见表1。区间 $[\pi_{A_1},\pi_{A_2}]$通常较小。

表1：不同比例 p_2 和隐私预算 ε下区间$[\pi_{A_1},\pi_{A_2}]$的长度。

p_2 \ ε	0.01	0.05	0.25	0.5
0.01	0.100	0.101	0.101	0.104
0.05	0.224	0.224	0.225	0.230

6.2 使用本地差分隐私通过数值模拟比较四种机制

为了通过数值模拟绘制四种机制的方差随隐私预算 ε 变化的曲线，我们固定四种机制的样本量 $N = 10^2$ 和比例 $\pi_A = 0.1$，固定修改的 Simmons 机制的比例 $\pi_B = 0.5$，以及改进的 Christofides 和改进的 Christofides 机制的比例 $p_2 = 0.5$。然后按照算法1至4进行实验。结果为运行 10000 次后的平均值。四种机制的理论分析和数值模拟结果如图1所示。数值模拟结果验证了理论分析值。结果表明，对于这四种机制，隐私预算 ε 越大，方差越小。

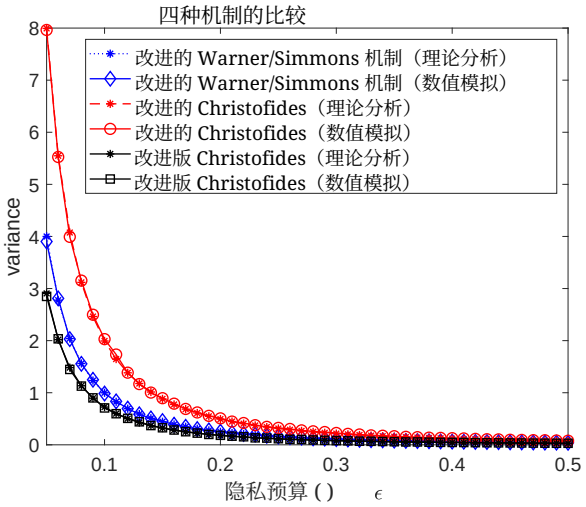


图1：在固定样本量 $N = 10^2$ 和比例 $\pi_A = 0.1$时，四种机制的方差随隐私预算 ε的变化（通过理论分析和数值模拟）。

随后，我们报告了在定理 6.1的不同条件下，通过数值模拟得到的四种机制的方差随隐私预算 ε 变化的曲线。我们固定四种机制的样本量为 $N = 9$，固定修改的 Simmons 机制的比例为 $\pi_B = 0.5$，固定修改的 Christofides 机制和改进的 Christofides 机制的比例为 $p_2 = 0.36$，并将比例 $\pi_A \in \{1/9, 2/9, 4/9\}$设定为这些值。结果见图2至图4。

图2显示，当比例 π_A 处于区间 $(0, \pi_{A_1}) \cup (\pi_{A_2}, 1)$时，改进的 Christofides 机制表现最佳，修改的 Warner/Simmons 机制次之，而修改的 Christofides 机制表现最差。

图3显示，当比例 π_A 位于区间 $(\pi_{A_1}, 1/2-1/2\sqrt{N}) \cup \left(1/2+1/2\sqrt{N}, \pi_{A_2}\right)$时，改进的 Warner/Simmons 机制表现最佳，改进的 Christofides 机制次之，而修改的 Christofides 机制最差。

图4显示，当比例 π_A 位于区间 $(1/2-1/2\sqrt{N}、1/2+1/2\sqrt{N})$ 时，改进的 Christofides 机制效果最好，改进的 Warner/Simmons 机制效果次之，而改进的 Christofides 机制效果最差。

在现实世界中，具有敏感属性的人所占比例通常较小。这意味着改进的 Christofides 机制在大多数情况下通常是最佳方法。

实际上，在第6.3节中，我们基于一个真实世界的数据集进行了实验。

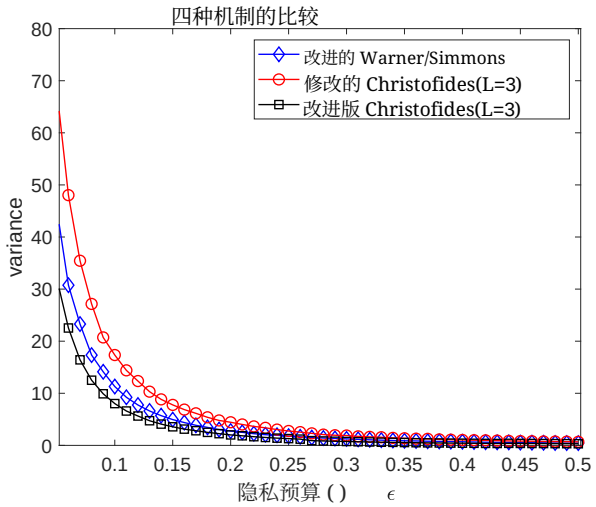


图2：固定样本量 $N = 9$ 和比例 $\pi_A = 1/9$时，四种机制的方差与隐私预算 ε 之间的关系（通过数值模拟）。

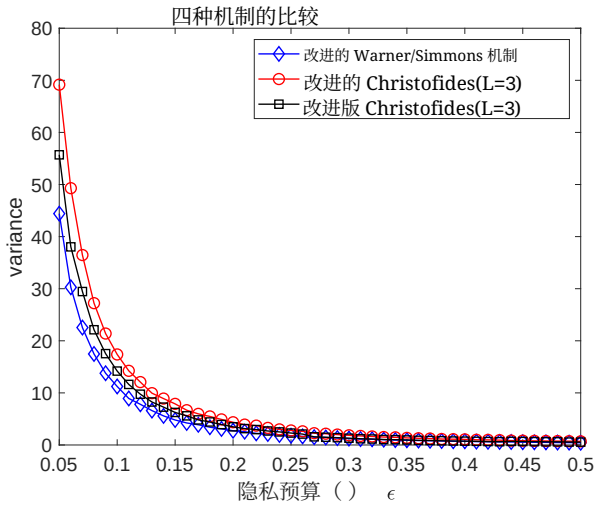


图 3：四种机制的方差随隐私预算 ε 的变化情况，固定样本量 $N = 9$ 和比例 $\pi_A = 2/9$（通过数值模拟）。

6.3 使用本地差分隐私在真实世界数据集上比较四种机制

数据集。*HCOVANY* 数据集 *[27]*。我们在一个开放的现实世界数据集上开展实验。HCOVANY 表示受访者在接受访谈期间是否拥有

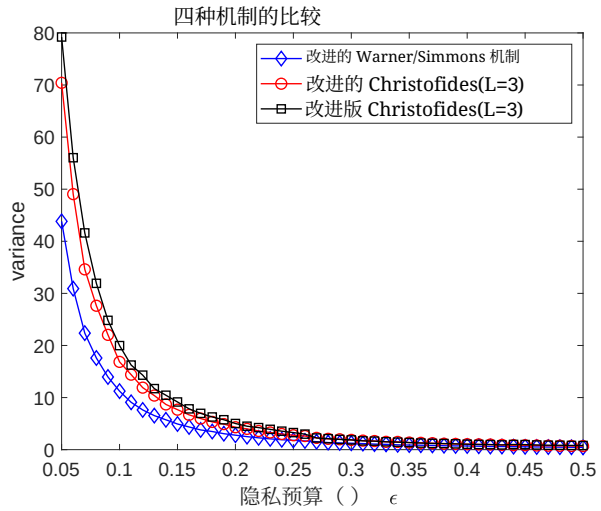


图4：在固定样本量 $N = 9$ 和比例 $\pi_A = 4/9$时，四种机制的方差随隐私预算 ε 变化的关系（通过数值模拟）。

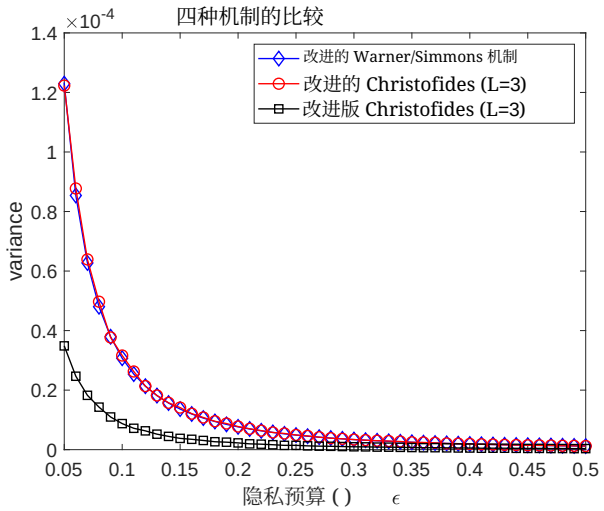


图 5：基于 HCOVANY 数据集的四种机制的方差与隐私预算ε 比较。

医疗保险覆盖，具体通过雇主提供的保险（HINSEMP）、私人购买的保险（HINSPUR）、Medicare（HINSCARE）、Medicaid 或其他政府保险（HINSCAID）、TRICARE 或其他军事医疗保险（HINSTRI），或退伍军人事务管理局提供的保险（HINSVA）来衡量。美国人口普查局不将仅通过印第安人卫生服务（HINSIHS）获得医疗保险覆盖的受访者视为拥有医疗保险覆盖，因为 IHS 保单并不总是全面的。编码 2 表示某人受所列保险类型覆盖（无论是本人直接投保，还是通过其他家庭成员的保单投保）；编码 1 表示某人未受覆盖。我们选取了*HCOVANY* 数据集 中的一个子集用于实验（该子集涵盖 2021 年美国总体中 1 % 人口的医疗保险覆盖情况，抽样规模为 $N = 3252599$）。

隐私预算 ε。根据哈佛关于数据集不同敏感性级别及各级别合理隐私损失参数的列表 [1]，我们将隐私预算 ε 设定为 0.05 至 0.5，因为医疗保险覆盖信息一旦泄露，可能会给个人带来实质性损害风险。

我们为修改的 Christofides 机制和改进的 Christofides 机制固定比例 $p_2 = 0.01$。随后按照算法1至4进行实验。我们报告了 10_4 次运行的平均结果。这四种机制的方差随隐私预算 ε 变化的比较如图 5 所示。

改进的 Warner 机制的方差与修改的 Simmons 机制相同。修改的 Christofides 机制的方差略大于改进的 Warner 机制和 Simmons 机制。改进的 Christofides 机制的方差在各机制中最小。事实上，基于 HCOVANY 数据集，我们将方差降至修改的 Christofides 机制方差的 28.7%。

$$\frac{\mathbf{Var}_{IC}\hat{\pi_A}}{\mathbf{Var}_{MC}\hat{\pi_A}} = \frac{4N\pi_A\,(1-\pi_A)}{N-1}$$
$$= \frac{4 \times 3252599 \times 0.0778 \times (1-0.0778)}{3252599-1}$$
$$= 28.7\%.$$

6.4 四种机制的最小样本量比较

由于统计学家的一个重要任务是在给定约束条件下确定最小样本数 N，因此我们比较了四种机制（改进的 Warner、Simmons、Christofides（$p_2 = 0.01$）和改进的 Christofides 机制（$p_2 = 0.01$）在给定比例 $\pi_A = 0.1$、方差 **方差** $\hat{}\pi_A = 0.1$ 和隐私预算 ε 条件下的最小样本量 N。结果如表 2所示。结果表明，对于这四种机制，隐私预算 ε 越大，所需样本量 N 越小。改进的 Warner 和 Simmons 机制所需的最小样本数 N 相同。改进的 Christofides 机制所需的最小样本数 N 略大于改进的 Warner/Simmons 机制。改进的 Christofides 机制所需的最小样本数 N 最少。

表 2：给定比例 N、方差 $\mathbf{Var}\,\hat{}\pi_A = 0.1$ 和隐私预算 ε时四种机制的最小样本量。

机制 \ 隐私预算	0.01	0.05	0.25	0.5
改进的 Warner/Simmons 机制	100000	4000	160	40
改进的 Christofides	101011	4040	161	40
改进版 Christofides	36365	1456	59	16

尤其是，使用改进的 Christofides 机制时，如果我们不知道该比例是否属于敏感群体，例如，在方差 **方差** ≤ 0.1、隐私预算 $\varepsilon = 0.01$ 和所选比例 $p_2 = 0.01$的约束条件下，由于方差在 $\pi_A = 0.5$时达到最大值，样本量 N应至少为 101011（ $N > 101010.3$）。

由于修改的 Simmons 机制的第二项陈述无关且不敏感，因此具有敏感属性的受访者比改进的 Warner 机制中的受访者更愿意配合。改进的 Christofides 机制和改进的 Christofides 机制只要求受访者报告数字。因此，具有敏感属性的受访者比改进的 Warner 机制和 Simmons 机制中的受访者更愿意配合。在改进的 Christofides 机制中，没有受访者知道其他人抽取的数字以及其他人回答的数字，这符合现实世界的情况。为在相同的隐私预算 ε 下获得更小的方差，受访者牺牲了知晓其他人抽取的数字和回答的数字的自由。

我们通过理论分析、数值模拟以及基于现实世界数据集的实验，将本地差分隐私作为衡量隐私泄露的通用指标，比较了改进的 Warner 机制、Simmons 机制、Christofides 机制和改进的 Christofides 机制的方差和最小样本量。结果，我们提出了适用于本地差分隐私框架的改进的 Christofides 机制。这是我们的第三项贡献。

7 结论

总之，我们通过不放回抽取样本，对 Warner、Simmons 和 Christofides 机制进行了改进，从而降低了各机制的方差。随后，我们通过不放回抽样卡片，进一步改进了改进的 Christofides 机制。结合理论分析、数值模拟以及基于现实世界数据集的实验，我们以本地差分隐私（LDP）作为隐私泄露的衡量标准，对改进的 Warner、Simmons、Christofides 和改进的 Christofides 机制进行了比较。我们发现，当具有敏感属性的人所占比例为 $\pi_A < \pi_{A_1}$或$\pi_A > \pi_{A_2}$时（这在现实世界中通常较为常见），改进的 Christofides 机制的方差最小；此外，在敏感问题调查和本地差分隐私场景中，改进的 Christofides 机制可替代 Warner 机制，因为它在隐私预算、方差和合作程度之间实现了更好的权衡。

附录

A 定理 3.1 的证明

定理 3.1（修改的 Christofides 机制方差）。$\hat{}$的方差为

$$\begin{aligned}\mathbf{Var}\hat{\pi_A} &= \frac{\mathbf{Var}\left(\sum_{i=1}^{N} X_i\right)}{N^2(L+1-2\mathbf{E}(Y))^2} \\ &= \frac{\mathbf{Var}Y}{N(L+1-2\mathbf{E}(Y))^2}.\end{aligned}$$

证明。首先，我们有

$$\begin{aligned}\mathbf{Var}\left(\sum_{i=1}^{N} X_i\right) =& N\pi_A\left[\sum_{k=1}^{L} k^2 p_{L+1-k} - \left(\sum_{k=1}^{L} k p_{L+1-k}\right)^2\right] + \\ & N(1-\pi_A)\left[\sum_{k=1}^{L} k^2 p_k - \left(\sum_{k=1}^{L} k p_k\right)^2\right] \\ =& N\pi_A\left[\sum_{k=1}^{L}(L+1-k)^2 p_k - \sum_{k=1}^{L} k^2 p_k\right] + \\ & N\pi_A\left[-(\sum_{k=1}^{L}(L+1-k)p_k)^2 + (\sum_{k=1}^{L} k p_k)^2\right] + \\ & N\left[\sum_{k=1}^{L} k^2 p_k - \left(\sum_{k=1}^{L} k p_k\right)^2\right] \\ =& N\pi_A\left[\sum_{k=1}^{L}((L+1)^2 - 2(L+1)k)p_k\right] - \\ & N\pi_A\left[-(\sum_{k=1}^{L}(L+1-k)p_k)^2 + (\sum_{k=1}^{L} k p_k)^2\right] + \\ & N\pi_A\left[(\sum_{k=1}^{L}((L+1)p_k)(\sum_{k=1}^{L}((L+1-2k)p_k)\right] \\ =& N\left[\sum_{k=1}^{L} k^2 p_k - \left(\sum_{k=1}^{L} k p_k\right)^2\right].\end{aligned}$$

Thus

$$\begin{aligned}\mathbf{Var}\hat{\pi_A} &= \frac{\mathbf{Var}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right)}{(L+1-2E(Y))^2} \\ &= \frac{\mathbf{Var}Y}{N(L+1-2E(Y))^2}. \quad \square\end{aligned}$$

B 定理 4.1 的证明

定理 4.1（改进的 Christofides 机制方差）。$\hat{}$ 的方差为

$$\begin{aligned}\mathbf{Var}\hat{\pi_A} &= \frac{\mathbf{Var}\left(\sum_{i=1}^{N} X_i\right)}{N^2(L+1-2\mathrm{E}(Y))^2} \\ &= \frac{4\pi_A\,(1-\pi_A)\,\mathbf{Var}(Y)}{(N-1)(L+1-2\mathbf{E}(Y))^2}.\end{aligned}$$

证明。由于 π_A的估计量

$$\hat{\pi_A} = \frac{\frac{1}{N}\sum_{i=1}^{N} X_i - \mathbf{E}Y}{(L+1-2\mathbf{E}Y)}.$$

其方差为

$$\begin{aligned}\mathbf{Var}\hat{\pi_A} &= \mathbf{Var}\left(\frac{\frac{1}{N}\sum_{i=1}^{N} X_i - \mathbf{E}Y}{L+1-2\mathbf{E}Y}\right) \\ &= \frac{1}{N^2(L+1-2\mathbf{E}Y)^2}\mathbf{Var}\sum_{i=1}^{N} X_i\end{aligned}$$

，其中

$$\begin{aligned}\mathbf{Var}\sum_{i=1}^{N}X_i &= \mathbf{E}(\sum_{i=1}^{N}X_i)^2-(\mathbf{E}\sum_{i=1}^{N}X_i)^2\\ &=\sum_{i=1}^{N}\mathbf{E}(X_i^2)+\sum_{i\neq j}\mathbf{E}\left(X_iX_j\right)-(\mathbf{E}\sum_{i=1}^{N}X_i)^2.\end{aligned}$$

在改进的 Christofides 机制中， X_i和 X_j不再是独立同分布的情况，这给证明带来了困难。此处，参与者和卡片均为不放回抽样，因此 Y_i和 Y_j、 x_i和 x_j相关，但 Y_i和 x_i相互独立。

首先，

$$\begin{aligned}&\sum_{i=1}^{N}\mathbf{E}\left(X_i^2\right)=\sum_{i=1}^{N}\mathbf{E}\left[(L+1-Y_i)^2x_i^2+Y_i^2x_i^2\right]\\ &=N\left[\mathbf{E}(L+1-Y)^2\mathbf{E}x_i^2+\mathbf{E}Y^2\mathbf{E}\left(1-x_i\right)^2\right]\\ &=N\left[(L+1)^2+\mathbf{E}Y^2-2(L+1)\mathbf{E}Y\right]\mathbf{E}x_i^2+\\ &N\mathbf{E}Y^2\mathbf{E}\left(1-2x_i+x_i^2\right)\\ &=N\mathbf{E}Y^2-2N(L+1)\pi_A\mathbf{E}Y+N(L+1)^2\pi_A.\end{aligned}\tag{6}$$

其次，
对于 $i\neq j$

$$\begin{aligned}\mathbf{E}X_iX_j&=\mathbf{E}\left((L+1-Y_i)x_i+Y_i(1-x_i)\right)\cdot\\ &\quad\left((L+1-Y_j)x_j+Y_j(1-x_j)\right),\\ &=\mathbf{E}(L+1-Y_i)(L+1-Y_j)\mathbf{E}x_ix_j\\ &+\mathbf{E}Y_iY_j\mathbf{E}(1-x_i)(1-x_j)\\ &+2\mathbf{E}Y_i(L+1-Y_j)\mathbf{E}(1-x_i)x_j.\end{aligned}$$

因为我们有

$$\begin{aligned}\mathbf{E}Y_iY_j&=\sum_{m=1}^{N}\sum_{n=1}^{N}mn\mathbf{P}(Y_i=m,\,Y_j=n)\\ &=\sum_{m=1}^{N}m\left(\sum_{k\neq m}^{N}np_m\frac{p_nN}{N-1}+mp_m\frac{p_mN-1}{N-1}\right)\\ &=\frac{N}{N-1}\mathbf{E}Y\sum_{m=1}^{N}mp_m-\frac{1}{N-1}\sum_{m=1}^{N}m^2p_l\\ &=(\mathbf{E}Y)^2-\frac{1}{N-1}\mathbf{Var}Y.\end{aligned}$$

and

$$\mathbf{E}x_i=\mathbf{P}(x_i=1)=\pi_A$$

$$\mathbf{E}x_ix_j=\mathbf{P}(x_i=1,\,x_j=1)=\frac{C_{N\pi_A}^2}{C_N^2}$$

$$\mathbf{E}(1-x_i)(1-x_j)(i\neq j)=\mathbf{P}(x_i=0,\,x_j=0)=\frac{C_{N-N\pi_A}^2}{C_N^2}$$

$$\mathbf{E}(1-x_i)x_j(i\neq j)=\mathbf{P}(x_i=0,\,x_j=1)=\frac{C_{N\pi_A}^1C_{N-N\pi_A}^1}{A_N^2}.$$

我们可以得到

$$\begin{aligned}\mathbf{E}X_iX_j&=\mathbf{E}[(L+1)^2-Y_i(L+1)+Y_iY_j]\mathbf{E}x_ix_j+\mathbf{E}Y_iY_j\\ &\quad\mathbf{E}(1-x_i)(1-x_j)+\mathbf{E}(L+1-Y_i)Y_j\mathbf{E}x_i(1-x_j)+\\ &\quad\mathbf{E}Y_i(L+1-Y_j)\mathbf{E}(1-x_i)x_j\\ &\quad=\mathbf{E}Y_iY_j[\mathbf{E}x_ix_j+\mathbf{E}(1-x_i)(1-x_j)-\mathbf{E}x_i(1-x_j)-\\ &\quad\mathbf{E}(1-x_i)x_j]+\mathbf{E}Y[-2(L+1)\mathbf{E}x_ix_j+(L+1)\mathbf{E}x_i\\ &\quad(1-x_j)+(L+1)\mathbf{E}(1-x_i)x_j]+(L+1)^2\mathbf{E}x_ix_j\\ &\quad=[N\pi_A(N\pi_A-1)+(N-N\pi_A)(N-N\pi_A-1)-\\ &\quad2N+\pi_A(N-N\pi_A)]\mathbf{E}Y_iY_j+[-2(L+1)N\pi_A(N\pi_A-\\ &\quad1)+2(L+1)N\pi_A(N-N\pi_A)]\mathbf{E}Y+(L+1)^2N\pi_A\\ &\quad(N\pi_A-1)\\ &\quad=[4N^2\pi_A^2-4N^2\pi_A+N^2-N]\mathbf{E}Y_iY_j+2(L+1)\\ &\quad N\pi_A[1+N-2N\pi_A]\mathbf{E}Y+(L+1)^2N\pi_A(N\pi_A-1)\\ &\quad=[4N^2\pi_A^2-4N^2\pi_A+N^2-N][(\mathbf{E}Y)^2-\frac{1}{N-1}\\ &\quad\mathbf{Var}Y]+2(L+1)N\pi_A[1+N-2N\pi_A]\mathbf{E}Y+(L+1)^2\\ &\quad N\pi_A(N\pi_A-1).\end{aligned}\tag{7}$$

第三，

$$\begin{aligned}\mathbf{E}(\sum_{i=1}^{N}X_i)^2&=N^2[\mathbf{E}Y+\pi_A(L+1-2\mathbf{E}Y)]^2\\ &=N^2[(L+1)\pi_A+(1-2\pi_A)\mathbf{E}Y]^2\\ &=N^2[(L+1)^2\pi_A^2+(1-2\pi_A)^2(\mathbf{E}Y)^2+2(L+1)\pi_A\\ &\quad(1-2\pi_A)\mathbf{E}Y].\end{aligned}\tag{8}$$

最后，结合方程(6)、(7)和(8)，估计量$\hat{}\pi_A$的方差为

$$\begin{aligned}\mathbf{Var}\hat{\pi_A}&=\frac{1}{N^2(L+1-2\mathbf{E}Y)^2}\mathbf{Var}\sum_{i=1}^{N}X_i\\ &=\frac{4\pi_A(1-\pi_A)\mathbf{Var}Y}{(N-1)(L+1-2\mathbf{E}Y)^2}.\quad\square\end{aligned}$$

C 定理 5.3 的证明

定理 5.3（修改的 Simmons 机制的方差与 epsilon 的关系）。当 $\pi_B=1/2$时，方差 $\mathbf{Var}\pi_A$ 取得如下最小值

$$\mathbf{Var}\hat{\pi_A}=\frac{e^\varepsilon}{N\left(e^\varepsilon-1\right)^2}.$$

证明。将 (4) 代入 (2)，方差可写为

(1) 当 $\pi_B\leq\frac{1}{2}$,

$$\begin{aligned}\mathbf{Var}\hat{\pi_A}=&\frac{\left(\frac{1}{e^\varepsilon-1}+\pi_A\right)\left(1+\frac{1}{(e^\varepsilon-1)\pi_B}\right)}{N}-\frac{1}{N\left(e^\varepsilon-1\right)^2}\\ &-\frac{\left(e^\varepsilon+1\right)\pi_A}{N\left(e^\varepsilon-1\right)}\end{aligned}$$

$$\min\mathbf{Var}\hat{\pi_A}=\mathbf{Var}\hat{\pi_A}|_{\pi_B=\frac{1}{2}}=\frac{e^\varepsilon}{N\left(e^\varepsilon-1\right)^2}.$$

(2) *When* $\pi_B > \frac{1}{2}$,

$$\begin{aligned}\mathbf{Var}\hat{\pi}_A =& \frac{\left(\frac{1}{e^{\varepsilon}-1}+\pi_A\right)\left(1+\frac{1}{(e^{\varepsilon}-1)(1-\pi_B)}\right)}{N} - \frac{1}{N\left(e^{\varepsilon}-1\right)^2} \\ & - \frac{\left(e^{\varepsilon}+1\right)\pi_A}{N\left(e^{\varepsilon}-1\right)}\end{aligned}$$

$$\min \mathbf{Var}\hat{\pi_A} = \mathbf{Var}\hat{\pi}_{A_{\pi_B=\frac{1}{2}}} = \frac{e^{\varepsilon}}{N\left(e^{\varepsilon}-1\right)^2}.$$

因此，估计量 $\hat{\pi}_A$ 的方差为

$$\mathbf{Var}\hat{\pi_A} = \frac{e^{\varepsilon}}{N\left(e^{\varepsilon}-1\right)^2}. \quad \square$$

D 定理 5.5 的证明

定理 5.5（修改的 Christofides 机制的方差与 epsilon 的关系）。对于情形 $L = 3$，修改的 *Christofides* 机制$\mathbf{Var}_{\mathrm{MC}}$ 的方差在以下比例条件下达到最小值

$$\mathbf{Var}\hat{\pi_A} = \frac{1}{4N}\left[\frac{(e^{\epsilon}+1)^2}{(e^{\epsilon}-1)^2(1-p_2)} - 1\right]$$

比例满足以下条件

$$p_1 = (1-p_2)/\left(e^{\varepsilon}+1\right), p_3 = \left[e^{\varepsilon}(1-p_2)\right]/\left(e^{\varepsilon}+1\right)$$

or

$$p_1 = \left[e^{\varepsilon}(1-p_2)\right]/\left(e^{\varepsilon}+1\right), p_3 = (1-p_2)/\left(e^{\varepsilon}+1\right).$$

证明。 Y的期望为

$$\begin{aligned}\mathbf{E}Y &= p_1 + p_2 + 3(1-p_1-p_2) \\ &= 3 - 2p_1 - p_2.\end{aligned}$$

Y的方差为

$$\begin{aligned}\mathbf{Var}Y &= p_1 + 4p_2 + 9(1-p_1-p_2) - (3-2p_1-p_2)^2 \\ &= 4p_1 + p_2 - 4p_1^2 - p_2^2 - 4p_1p_2.\end{aligned}$$

$$\begin{aligned}\mathbf{Var}\hat{\pi}_A &= \frac{\mathbf{Var}Y}{N(3+1-2\mathbf{E}Y)^2} \\ &= \frac{4p_1+p_2-4p_1^2-p_2^2-4p_1p_2}{N\left[3+1-2(3-2p_1-p_2)\right]^2} \\ &= \frac{4p_1+p_2-4p_1^2-p_2^2-4p_1p_2}{N(4p_1+2p_2-2)^2} \\ &= f(p_1,p_2).\end{aligned}$$

现在，让我们将函数 $f(p_1,p_2)$ 最小化。

由于

$$\begin{aligned}\frac{\partial f(p_1,p_2)}{\partial p_2} &= \frac{(1-4p_1-2p_2)(4p_1+2p_2-2)^2}{N(4p_1+2p_2-2)^4} - \\ &\quad \frac{4\left(4p_1+p_2-4p_1^2-p_2^2-4p_1p_2\right)(4p_1+2p_2-2)}{N(4p_1+2p_2-2)^4} \\ &= \frac{\left(p_1+\frac{1-p_2}{2}\right)\left(p_1-\frac{1-p_2}{2}\right)}{N(2p_1+p_2-1)^4} \\ &\neq 0.\end{aligned}$$

函数 $f(p_1,\ p_2)$ 没有极值。最大值点或最小值点必须是边界点。

由于

$$\frac{1}{e^{\varepsilon}}p_1 \le p_3 = 1-p_1-p_2 \le e^{\varepsilon}p_1$$

thus

$$\frac{1-p_2}{e^{\varepsilon}+1} \le p_1 \le \frac{e^{\varepsilon}}{e^{\varepsilon}+1}(1-p_2).$$

由于两个边界点的函数

$$\begin{aligned}f(p_1,p_2)|_{p_1=\frac{1-p_2}{e^{\varepsilon}+1}} &= \frac{4\frac{1-p_2}{e^{\varepsilon}+1}+p_2-4\left(\frac{1-p_2}{e^{\varepsilon}+1}\right)^2-p_2^2-4\frac{1-p_2}{e^{\varepsilon}+1}p_2}{N\left[4\frac{1-p_2}{e^{\varepsilon}+1}+2p_2-2\right]^2} \\ &= \frac{1}{4N}\left[\frac{(e^{\varepsilon}+1)^2}{(e^{\varepsilon}-1)^2(1-p_2)}-1\right]\end{aligned}$$

$$\begin{aligned}f(p_1,p_2)|_{p_1=\frac{e^{\varepsilon}(1-p_2)}{e^{\varepsilon}+1}} &= \frac{4\frac{e^{\varepsilon}(1-p_2)}{e^{\varepsilon}+1}+p_2-4(\frac{e^{\varepsilon}(1-p_2)}{e^{\varepsilon}+1})^2}{N[\frac{e^{\varepsilon}}{e^{\varepsilon}+1}(1-p_2)+2p_2-2]^2} - \\ &\quad \frac{p_2^2+4\frac{e^{\varepsilon}}{e^{\varepsilon}+1}(1-p_2)p_2}{N[\frac{e^{\varepsilon}}{e^{\varepsilon}+1}(1-p_2)+2p_2-2]^2}. \\ &= \frac{1}{4N}[\frac{(e^{\varepsilon}+1)^2}{(e^{\varepsilon}-1)^2(1-p_2)}-1]\end{aligned}$$

and

$$f(p_1,p_2)|_{p_1=\frac{1-p_2}{2}} \to \infty.$$

因此

$$\min f(p_1,p_2) = f\left(\frac{1-p_2}{e^{\varepsilon}+1},p_2\right) = f\left(\frac{e^{\varepsilon}(1-p_2)}{e^{\varepsilon}+1},p_2\right).$$

So

$$\mathbf{Var}\hat{\pi_A} = \frac{1}{4N}\left[\frac{(e^{\varepsilon}+1)^2}{(e^{\varepsilon}-1)^2(1-p_2)}-1\right]. \quad \square$$

致谢

本研究得到国家重点研发计划（2022YFB4501500、2022YFB4501504）和之江实验室（2022NF0AC01）的支持。本材料中表达的任何观点、研究结果、结论或建议均代表作者个人意见，不反映资助机构的观点。

参考文献

[1] Rafal Ablamowicz 和 Bertfried Fauser。2023年。哈佛安全数据分类。哈佛大学隐私工具项目。检索日期：2023年2月15日，网址：http://psiprivacy.org/ [2] Shunsuke Aoki、Masayuki Iwai 和 Kaoru Sezaki。2012年。有限负向调查：隐私保护参与式感知。载于*2012*年*IEEE*第一届云网络国际会议（*CLOUDNET*）。IEEE，第158–160页。[3] Hilal Asi、Vitaly Feldman 和 Kunal Talwar。2022年。本地差分隐私下均值估计的最优算法。载于国际机器学习会议。PMLR，第1046–1056页。[4] Ergute Bao、Yin Yang、Xiaokui Xiao 和 Bolin Ding。2021年。CGM：一种采用本地差分隐私的增强型流式数据收集机制。*VLDB*基金会会刊14，11（2021），2258–2270。[5] Raef Bassily、Kobbi Nissim、Uri Stemmer 和 Abhradeep Guha Thakurta。2017年。实用的本地隐私重频项。神经信息处理系统进展30（2017）。[6] 美国人口普查局。2020年。美国人口普查局。美国政府。检索日期：2023年2月23日，网址：https://onthemap.ces.census.gov/ [7] Yang Cao、Masatoshi Yoshikawa、Yonghui Xiao 和 Li Xiong。2018年。时间相关性下连续数据发布中的差分隐私量化。*IEEE*知识与数据工程汇刊31，7（2018），1281–1295。[8] Tasos C Christofides。2003年。一种广义随机应答技术。*Metrika*57（2003），195–200。[9] Graham Cormode、Samuel Maddock 和 Carsten Maple。2021年。本地差分隐私下的频率估计。*VLDB*基金会会刊14，11（2021），2046–2058。[10] Bolin Ding、Janardhan Kulkarni 和 Sergey Yekhanin。2017年。私密地收集遥测数据。神经信息处理系统进展30（2017）。[11] John C Duchi、Michael I Jordan 和 Martin J Wainwright。2013年。本地隐私与统计极小极大速率。载于*2013*年*IEEE*第*54*届计算机科学基础年会。IEEE，第429–438页。[12] Rick Durrett。2019年。概率论：理论与实例。第49卷。剑桥大学出版社。[13] Úlfar Erlingsson、Vasyl Pihur 和 Aleksandra Korolova。2014年。RAPPOR：随机可聚合隐私保护序数应答。载于*2014*年*ACM SIGSAC*计算机与通信安全会议。第1054–1067页。[14] Fernando Esponda。2006年。负向调查。*arXiv*预印本 *math/0608176*（2006）。[15] Sabrina Giordano 和 Pier Francesco Perri。2012年。基于相同隐私保护程度的无关问题模型效率比较。*Statistical Papers* 53，4（2012），987。[16] Christopher R Gjestvang 和 Sarjinder Singh。2006年。一种新的随机应答模型。英国皇家统计学会会刊：*B*辑（统计方法学）68，3（2006），523–530。[17] Bernard G Greenberg、Abdel-Latif A Abul-Ela、Walt R Simmons 和 Daniel G Horvitz。1969年。无关问题随机应答模型：理论框架。美国统计学会会刊 64, 326 (1969), 520–539. [18] Anthony YC Kuk. 1990. 间接询问敏感问题。*Biometrika* 77, 2 (1990), 436–438. [19] Jan Lanke. 1976. 关于随机化访谈中的保护程度。国际统计评论/国际统计评论 (1976), 197–203. [20] Frederick W Leysieffer 和 Stanley L Warner. 1976. 随机应答模型中的受访者风险与最优设计。美国统计学会会刊 71, 355 (1976), 649–656. [21] NS Mangat 和 Ravindra Singh. 1990. 一种替代性的随机应答程序。*Biometrika* 77, 2 (1990), 439–442. [22] Naurang S Mangat. 1994. 一种改进的随机应答策略。英国皇家统计学会会刊：*B*辑（方法论）56, 1 (1994), 93–95. [23] Pier Francesco Perri. 2008. Simmons模型的改进随机化装置。模型辅助统计与应用 3, 3 (2008), 233–239.[24] Sofya Raskhodnikova、Adam Smith、Homin K Lee、Kobbi Nissim 和 Shiva Prasad Kasiviswanathan. 2008. 我们能从隐私保护中学到什么。在第*54*届计算机科学基础年度研讨会论文集。531–540. [25] Sarjinder Singh、Stephen Horn、Ravindra Singh 和 NS Mangat. 2003. 关于使用改进的随机化装置估计敏感属性流行率。转型中的统计学 6, 4 (2003), 515–522. [26] Apple差分隐私团队. 2017. 在大规模场景中实现隐私保护学习。Apple。检索于2023年3月1日，网址：https://machinelearning.apple.com/research/learning-with-privacy-at-scale [27] 综合公共使用微观数据系列（IPUMS USA）. 2021. *HCO-VANY2021*数据集说明。综合公共使用微观数据系列（IPUMS USA）。检索于2023年2月15日，网址：https://usa.ipums.org/usa/acs_healthins.shtml [28] Tianhao Wang 和 JeremiahBlocki. 2017. 用于频率估计的局部差分隐私协议。在第*26*届*USENIX*安全研讨会论文集。 [29] Tianhao Wang、Ninghui Li 和 Somesh Jha. 2018. 局部差分隐私频繁项集挖掘。在*2018*年*IEEE*安全与隐私研讨会（*SP*）。IEEE，127–143. [30] Tianhao Wang、Milan Lopuhaa-Zwakenberg、Zitao Li、Boris Skoric 和 Ninghui Li. 2020. 具有一致性的局部差分隐私频率估计。在*NDSS'20*：*NDSS*研讨会论文集。[31] Stanley L Warner. 1965. 随机应答：一种消除回避性回答偏差的调查技术。美国统计学会会刊 60, 309 (1965), 63–69. [32] Atsushi Waseda 和 Ryo Nojima. 2016. 差分隐私下的随机应答机制分析。在信息安全：第*19*届国际会议，*ISC 2016*，美国夏威夷州檀香山，*2016*年*9*月*3*日至*6*日，论文集*19*。Springer，271–282.[33] Zhuolun Xiang、Bolin Ding、Xi He 和 Jingren Zhou. 2020. 基于度量的局部差分隐私下的线性计数与范围计数。在*2020*年*IEEE*国际信息论研讨会（*ISIT*）。IEEE，908–913. [34] Jianyu Yang、Tianhao Wang、Ninghui Li、Xiang Cheng 和 Sen Su. 2020. 局部差分隐私下的多维范围查询应答。*VLDB*基金会会刊 14, 3 (2020), 378–390.