arXiv:2303.02892v1 [stat.ME] 6 Mar 2023

参数极值的差分隐私置信区间

Xiaowen Fu[a*]、Yang Xiang[a、b†]、Xinzhou Guo[a‡]

[a] *Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong*

[b] *HKUST Shenzhen-Hong Kong Collaborative Innovation Research Institute,Futian, Shenzhen, China*

摘要

本文旨在为隐私保护下的参数极值构造有效且高效的置信区间。针对参数极值的常规统计推断通常会受到选择偏差问题的影响；而在许多极值参数的应用场景中，我们往往需要保护数据的隐私，因此这一问题变得更加突出。本文聚焦于指数族分布，提出了一种隐私化参数自助法，以解决差分隐私机制下参数极值问题中的选择偏差。尽管通常的隐私化参数自助法无法恰当地处理选择偏差，但我们证明，通过加入隐私化偏差校正项，所提出的参数自助法能够为参数极值带来有效且高效的置信区间。我们通过高斯情形和回归情形阐释所提出的方法，并通过数值实验展示该方法的优势。

1 引言

置信区间（CI）是总体未知参数估计值的一个合理取值范围。与点估计相比，置信区间不仅衡量参数的大小，还量化了估计不确定性，并构成统计推断的一项基础。置信区间广泛应用于不同的科学领域。例如，Sandercock (2015)指出，自20世纪80年代末以来，在医学期刊中报告置信区间已成为一项标准。

极值参数指总体中某些参数的最大值或最小值，在许多现实问题中具有实际意义。人们普遍

*电子邮箱地址：***xfuak@connect.ust.hk**
†电子邮箱地址：***maxiang@ust.hk**
‡电子邮箱地址：***xinzhoug@ust.hk**

认识到，由于选择偏差，为极值参数构造置信区间颇具挑战；简单使用极值参数的样本对应量，无法得出有效的统计推断。如何解决选择偏差问题，并为极值参数构造有效且高效的置信区间，是一个兼具方法论和实践意义的重要问题。在许多涉及极值参数的应用场景中，数据往往受到隐私保护，这使得该问题更加突出。

极值参数的一个重要应用场景是亚组分析，其旨在发现并确认总体内的治疗效应异质性。在临床试验中，某种治疗对部分患者比对其他患者更有效的情况十分常见；例如参见 Mologen (2018)。在这种情况下，极值参数可用于表示最佳亚组的治疗效应，而针对极值参数的有效且高效的置信区间能够帮助研究人员更好地理解该治疗，并了解该治疗在何处以及在多大程度上最为有效。尽管存在潜在益处，但通过直接访问临床试验数据来分析最佳亚组可能面临多种阻碍，尤其是鉴于健康数据的敏感性，隐私泄露日益受到关注，正如 Xiang 和 Cai (2021) 所讨论的那样。

为了保护隐私，人们提出了多种方案。例如，Sweeney (2002) 研究了前向 K-匿名性，Machanavajjhala 等人 (2007) 提出了 l-多样性，而 Li 等人 (2006) 研究了 t-接近性。本文聚焦于差分隐私（DP），这是一种由 Dwork 等人 (2006) 提出的广泛使用的方案。DP 旨在通过确保数据库中任意一次替换所产生的影响足够小，使对手无法推测真实数据集，从而保护隐私。正如 Dwork 等人 (2006) 所提出的，为实现 DP，我们通常会向估计值或数据中添加一定量的噪声。这显然会在统计推断中引入额外的随机性；而假设数据公开时所采用的常规置信区间或偏差校正方法，在私有数据场景下将无法得到令人满意的结果，正如 Dwork 等人 (2014) 所讨论的那样。

本文提出了一种在差分隐私机制下针对极值参数的有效且高效的置信区间。我们聚焦于指数族分布，并提出一种隐私化参数自助法，以解决差分隐私机制下极值参数的选择偏差。所提出的方法易于实现、效率高，并且可适用于不同的实际场景。

1.1 相关文献

极值参数去偏置信区间 众所周知，如果不进行适当调整，极值参数的置信区间将受到选择偏差的影响；参见 Thomas 和 Bornkamp（2017）以及 Magnusson 和 Turnbull（2013）。为构造极值参数的去偏置信区间，人们已经进行了一些尝试。Fuentes 等人（2018）以及 Hall 和 Miller（2010）提出了基于同时控制的有效置信区间，因此所得置信区间往往较为保守，效率也不高。Rosenkranz（2016）以及 Stallard 等人（2008）考虑了一些调整选择偏差的临时方法，但这些方法缺乏理论依据。Bornkamp 等（2017）以及 Woody 等（2022）考虑了贝叶斯方法，但该方法显然依赖于模型，并且通常缺乏频率学解释。近期，Guo 和 He（2021）、Guo 等人（2022）以及其他研究者提出了几种基于自助法的极值参数置信区间。尽管这些基于自助法的极值参数置信区间高效且理论依据充分，但它们是为公开数据设计的，不能直接应用

于差分隐私方案下。

差分隐私推断 在差分隐私下进行统计推断，已针对多种情形进行了研究。例如，Dimitrakakis 等人（2014）、Dimitrakakis 等人（2017）和 Zhang 等人（2016）研究了贝叶斯推断的差分隐私；Rogers 等人（2016）、Balle 等人（2020）和 Gaboardi 等人（2016）研究了差分隐私下的假设检验；Rinott 等人（2018）则研究了频数表中的差分隐私。也有一些研究探讨如何构造差分隐私置信区间。例如，Karwa 和 Vadhan （ 2017 ）提出了一种隐私算法，用于估计数据的取值范围，并据此推导出隐私置信区间。Du 等人（2020）提出了用于估计均值和方差的隐私模拟与分位数方法。Wang 等人（2018）提出了在差分隐私下构造经验风险估计置信区间的方法，该方法可用于逻辑回归和支持向量机（SVM）。

差分隐私自助法 为纠正偏差，通常会采用自助法程序。在差分隐私下，Covington 等人(2021) 提出了小批量自助法（BLB），用于私密地估计参数的抽样分布；Chadha 等人(2021) 提出了一种结合自助法的私有置信集；Brawner 和 Honaker (2018) 使用自助法结合差分隐私“免费”估计标准误；Dunsche 等人(2022) 提出了一种在纯差分隐私下结合自助法进行多元均值比较的检验；Ferrando 等人(2022) 提出了一种利用差分隐私参数自助法构造置信区间的方法。然而，现有的差分隐私自助法程序未能妥善处理选择偏差，也无法为极值参数提供有效的置信区间。

总结 据我们所知，在差分隐私下，通常的自助法无法直接用于处理选择偏差，而极值参数的去偏置信区间仍然缺失；本文旨在弥补这一空白。

1.2 贡献

我们工作的主要贡献可总结如下：(1) 提出一种隐私化参数自助法，以解决差分隐私下极值参数的选择偏差；(2) 在使用私有数据构造置信区间时，考虑噪声项所引入的随机性；(3) 将所提方法应用于高斯情形和回归情形，这两种情形在实践中有广泛应用；(4) 提出若干策略，以避免在干扰参数上浪费隐私预算。

2 预备知识

2.1 极值参数与选择偏差

令 $\beta = (\beta_1, \ldots, \beta_k) \in \mathbb{R}^k$ 表示总体中某些感兴趣的参数，并令 $\hat{}\ \beta_j$ 表示 β_j 对于 $j = 1, \cdots, k$ 的估计，这些估计来自数据集 $\mathcal{X}$。不失一般性，我们的目标是为最大值 $\beta_{\max} = \max_{j=1,\cdots,k} \beta_j$ 构造一个下置信限。

$\hat{}\ \hat{}$ 由于选择偏差， $\beta_{\max}$， $\beta_{\max} = \max_{j=1,\ \cdots,\ k} \beta_j$的样本对应量存在偏差 $\hat{}\ \hat{}$ 即使 β_j 与 β_j一致，单纯依赖 $\beta_{\max}$ 的置信区间也无效；理论推导见Nadarajah 和 Kotz(2008)。

2.2 差分隐私

差分隐私是一种隐私保护方案：如果更改数据集中的某个个体数据不会导致结果发生太大变化，对手可能就无法推测出真实的数据集。

定义 1（差分隐私，Dwork 等人（2006））如果一个机制 $\mathcal{A}$ 满足 ε -差分隐私（ε-DP），则对于所有仅在一个条目上不同的数据集对 x，$x^{'} \in \mathcal{X}$ 以及任意结果 $O \subseteq$range($\mathcal{A}$)，都有

$$|\ln(\frac{Pr(\mathcal{A}(x) \in O)}{Pr(\mathcal{A}(x^{'}) \in O)})| \leq \varepsilon$$

在差分隐私下，ε 是用于控制隐私泄露的参数，称为隐私预算。较小的 ε 表示更好的隐私保护，但可能会牺牲准确度。为了实现差分隐私，我们通常需要向数据中添加一定量的噪声，而噪声量通常由如下定义的敏感度决定：Dwork 等人（2006）。

定义 2（敏感度）函数 Γ的敏感度是满足对于所有 x，$x^{'} \in \mathcal{X}$ 仅在一个条目上不同的情况下，使得$S(\Gamma)$成立的最小数值S

$$||\Gamma(x) - \Gamma(x^{'})||_1 \leq S(\Gamma)$$

对于随机算法 Γ，为了实现 ε-差分隐私，我们通常考虑拉普拉斯机制。

定义 3（拉普拉斯机制）对于所有将数据集映射到$\mathbb{R}^d$的函数 Γ，$\Gamma(\mathbf{x}) + w$ 满足 ε-差分隐私，其中 $w = \{w_k\}_{k=1}^{d}$ 是添加的拉普拉斯噪声，其元素服从$w_k \sim$ Lap($S(\Gamma)/\varepsilon$)；Lap 表示尺度为 $S(\Gamma)/\varepsilon$的零均值拉普拉斯分布。

拉普拉斯机制会引入额外随机性以保护隐私，这通常会损害估计或推断的准确度和效率，因此我们需要对此进行适当处理。

差分隐私具有以下组合性质 Zhao（2017），稍后我们将利用这些性质为参数分配隐私预算，并在交叉验证中提供隐私保证。

顺序组合定理 设 M_i 各自提供 ε_i-差分隐私，则 $M_i(X)$的序列提供（$\sum_i \varepsilon_i$）-差分隐私。

并行组合定理 设 M_i 各自提供 ε-差分隐私。设 D_i 是输入域 D的任意互不相交子集。序列 $M_i(X \cap D_i)$提供 ε-差分隐私。

2.3 参数自助法

自助法是一种重抽样程序，通过有放回地抽取随机样本来模拟抽样过程，常用于统计推断和偏差校正。然而，有放回地抽取样本可能会导致隐私泄露。为避免在自助法中多次访问原始数据，我们可以考虑参数自助法，并假设已知数据集的分布族为 $X \sim \mathcal{M}(\alpha)$，其中 α 是该分布中的自然参数。设 θ 为关注参数，则针对 α 的参数自助法置信区间如算法1所示。

算法1： θ的参数自助法置信区间

数据： $\mathcal{X}$

置信区间： θ

基于 $\mathcal{X}$推导自然参数$\hat{}\alpha$ 的估计值;

当 $b = 1$： B **时**

  基于估计分布生成自助样本 $\mathcal{X}^* \sim \mathcal{M}(\hat{}\alpha)$;

  基于自助样本估计目标统计量 $\theta_b^* = \theta^*(\mathcal{X}^*)$

结束

基于 θ^* 推导置信区间$, \cdots, \theta_B^*$.

3 差分隐私极值参数置信区间

本节介绍极值参数的差分隐私置信区间框架。该框架基于指数族构建，可自然地扩展到其他参数分布族，但本文重点关注指数族分布。

3.1 指数族分布

对于来自数据集 $\mathcal{X} = \{x_i\}_{i=1}^n$ 的样本 x_i ，如果某个分布族的密度函数为

$$p(x_i; \alpha) = h(x_i) e^{\alpha^T T(x_i) - A(\alpha)}$$

其中， h 是基函数， α 是自然参数， T 是充分统计量函数， $A(\alpha)$ 是对数配分函数。指数族分布包含许多常见分布。

在实际应用中，我们感兴趣的参数 $\beta \in \mathbb{R}^{k_1}$ 可以视为 α 的函数；即 $\beta = f(\alpha)$。对于适当的干扰参数 $\gamma \in \mathbb{R}^{k_2}$，存在对 $\alpha = (\alpha_1, \alpha_2)$ 的一种划分，以及一个一一映射 $\mathbf{f}: (\alpha_1, \alpha_2) = \mathbf{f}(\beta, \gamma)$，其中 α_2 仅依赖于干扰参数 γ ，并且

$$\begin{cases} \alpha_1 = f_1(\beta, \gamma) \\ \alpha_2 = f_2(\gamma). \end{cases} \tag{1}$$

α_1 可以通过汇集所有包含 β 的项来生成，而 α_2 则很容易确定。经过这种重新参数化后，指数族分布可以改写为

$$p(x_i; \alpha_1, \alpha_2) = h(x_i) e^{\alpha_1^T T_1(x_i) + \alpha_2^T T_2(x_i) - A(\alpha_1, \alpha_2)} \tag{2}$$

其中 $(T_1(x_i)\ T_2(x_i)) = T(x_i)$。我们的目标是为 $\beta_{\max} = \max_{j=1,\cdots,k_1} \beta_j$ 构造隐私置信区间。

考虑数据集 $\mathcal{X} = \{x_i\}_{i=1}^n$ 并将式 1 代入式 2，取对数

似然为

$$\ln p(\mathcal{X};\beta,\gamma)=\sum \ln h(x_i)+f_1(\beta,\gamma)\sum T_1(x_i)$$
$$+f_2(\gamma)\sum T_2(x_i)-nA(\beta,\gamma)$$

其中，$A(\beta,\ \gamma)=A(f_1^{-1}(\beta,\ \gamma)\ f_2^{-1}(\gamma))$，并且 $\sum$ 在全文中表示 $\sum_{i=1}^{n}$ 的简化符号。然后，最大似然估计是方程3的一个解。

$$\begin{aligned}\frac{\partial \ln p(X;\beta,\gamma)}{\partial\beta}&=\frac{\partial f_1(\beta,\gamma)}{\partial\beta}\sum T_1(x_i)-n\frac{\partial A}{\partial\beta}=0\\ \frac{\partial \ln p(X;\beta,\gamma)}{\partial\gamma}&=\frac{\partial f_1(\beta,\gamma)}{\partial\gamma}\sum T_1(x_i)\\ &\quad+\frac{\partial f_2(\gamma)}{\partial\gamma}\sum T_2(x_i)-n\frac{\partial A}{\partial\gamma}=0\end{aligned}\tag{3}$$

我们写出该解，以便为第5节讨论的部分隐私化情形做准备。该解可写为

$$\begin{cases}\hat{\beta}=g_1(\sum T_1(x_i),\sum T_2(x_i))\\ \hat{\gamma}=g_2(\sum T_1(x_i),\sum T_2(x_i))\end{cases}\tag{4}$$

其中，g_1，g_2 是充分统计量的函数，用以表示 β，γ的最大似然估计。

3.2 隐私化参数自助法置信区间

为构建 $\beta_{\max}$的差分隐私置信区间，我们提出了一种隐私化参数自助法算法，以解决极值参数中的选择偏差。该算法包含三个关键要素：(1) 隐私化点估计；(2) 隐私化参数自助法；以及 (3) 隐私化偏差校正项。

隐私化点估计：为在点估计中实现差分隐私，根据方程（4），我们在第2步中向方程4中的充分统计量添加拉普拉斯噪声，具体如下

$$\begin{cases}\hat{\beta}^{priv}=g_1(\sum T_1(x_i)+w_1,\sum T_2(x_i)+w_2)\\ \hat{\gamma}^{priv}=g_2(\sum T_1(x_i)+w_1,\sum T_2(x_i)+w_2)\end{cases}\tag{5}$$

其中，$w_i\sim Lap(\triangle_i/\varepsilon_i)(i=1,\ 2)$为拉普拉斯噪声。 $\triangle_i$是充分统计量 $\sum T_i(x)$的敏感度，ε_i是分配给它的隐私预算。根据顺序组合理论，该框架满足$(\varepsilon_1+\varepsilon_2)$-差分隐私。

隐私化参数自助法：为避免重复访问数据，此处采用参数自助法。具体而言，遵循 Ferrando 等人（2022）的思路，我们基于估计模型中的$\hat{}\beta^{priv}$ 和$\hat{\gamma}^{priv}$生成自助法估计。由于我们至多只能从自助样本中推断出估计模型（即$\hat{}\beta^{priv}$ 和$\hat{\gamma}^{priv}$），根据并行组合理论，自助法之后该框架仍满足$(\varepsilon_1,\ \varepsilon_2)$-差分隐私。为考虑拉普拉斯机制所引入的随机性，我们添加拉普拉斯噪声，并在步骤 6 中计算自助法估计$\hat{}\ \beta^{*,\ priv}_{\ j}$。

隐私化偏差校正项：众所周知，常规自助法无法解决选择偏差；参见 Bornkamp 等人（2017）。遵循 Guo 和 He （2021）的思路，我们考虑使用修正自助法估计 $\hat{}\ \beta^{*,\ priv}_{\ j,\ modified}$ 来校正选择偏差，

$$\hat{\beta}^{*,priv}_{j,modified}=\hat{\beta}^{*,priv}_{j}+d_j,j=1,\cdots,k_1,$$

其中， d_j 是隐私化估计 $\hat{\beta}_j^{priv}$ 与基于原始数据集得到的其极值之间的距离

$$d_j = (1 - n^{r-0.5})(\hat{\beta}_{\max}^{priv} - \hat{\beta}_j^{priv}), j = 1, \cdots, k_1.$$

其中， n 是总体规模， $r \in$ (0,0 5)是调优参数，用于确定对估计值的调整幅度。 r越小，调整力度越大；以牺牲置信限的效率为代价，覆盖概率会更高。

校正后，该框架仍满足（$\varepsilon_1, \varepsilon_2$）-差分隐私。以此为例，所提出的算法总结于算法 2 中。

算法 2：极值问题中带偏差校正的隐私化参数自助法推断 数据： x_1, $\cdots, x_n$结果： $\beta_{\max}$的隐私化下置信限 1 对充分统计量添加拉普拉斯噪声，

并根据式 5 计算隐私化最大似然估计 $\hat{\beta}^{priv}, \hat{\gamma}^{priv}$ ； 2 估计隐私化极值 $\hat{\beta}_{\max}^{priv} = \max_{j=1,\cdots,k_1}\{\hat{\beta}_j^{priv}\}$； 3 计算偏差校正项$d_j = (1 - n^{r-0.5})(\hat{\beta}_{\max}^{priv} - \hat{\beta}_j^{priv})$ $j = 1, \cdots, k_1$； 4 当 $b = 1 : B$ 时，执行 5 基于指数族 (2) 和参数变换 (1) 生成自助样本$x_1^*, \cdots, x_n^* \sim p(f_1(\hat{\beta}^{priv}, \hat{\gamma})\ f_2(\hat{\gamma}^{priv}))$； 6 对充分统计量添加拉普拉斯噪声，并根据 (5) 计算隐私化最大似然估计 $\hat{\beta}^*$, ^{priv} ； 7 计算 $T^{*priv} = \sqrt{n}(\max\{\hat{\beta}^*, {}_{j}{}^{priv} + d_j - \hat{\beta}_{\max}^{priv}\})$ 8 结束 9 令 $c_\alpha = quantile(T^{*priv}, 1 - \alpha)$。置信水平为 $1 - \alpha$ 的下置信限为 $L^{priv} = \hat{\beta}_{\max}^{priv} - c_\alpha/\sqrt{n}$

3.3 交叉验证

为选择 r，我们建议采用数据自适应交叉验证方法。首先，我们考虑如下偏差缩减估计 $\hat{\beta}^{priv}{}_{\max\ reduced}$：

$$\hat{\beta}_{\max,reduced}^{priv} = \hat{\beta}_{\max}^{priv} - E^*[\beta_{\max,modified}^{priv,*} - \hat{\beta}_{\max}^{priv}], \tag{6}$$

其中， E^* 表示自助法分布下的期望 .

交叉验证的思路是选择 r，使 $\hat{\beta}^{priv}{}_{\max\ reduced}(r)$ 与 $\beta_{\max}^{priv}$ 之间的均方误差最小。令 $A = \{r_1, \cdots, r_m\}$ 表示取值范围为 (0，0.5) 的一组可能的调优参数，其中 $r_1 < \cdots < r_m$，且 m 为有限整数。根据我们的框架，可以使用以下算法在差分隐私下选择 $r \in A$。

对于交叉验证中的 j 折，我们记 $\hat{\beta}_j^{priv} = \{\beta_j^{priv}{}_{\cdot\ i}\}_{i=1}^{k_1}$其中 $\beta_{j\,i}^{priv}$,作为

$\hat{\beta}_j^{priv}$ 的第i项。根据并行组合理论，该框架将隐私预算预留为算法 2 中的预算。

算法3： 调优参数 r的交叉验证选择

数据： $x_1, \cdots, x_n$

结果： 调优参数 r的最优选择

将数据随机划分为 v 个（大小近似相等的）子样本；

对于 $l = 1, \cdots, m$ **do**

  对于 $j = 1, \cdots, v$，

    将第 j 个子样本用作参考数据，其余部分用作训练数据；

    使用训练数据，通过以下方式获得偏差缩减估计：通过 (6) $\hat{\beta}^{priv}$ 进行差分隐私 (r_l)，其中 r_l 为调优参数；

    使用参考数据估计 $\hat{\beta}_j^{priv}$；

    对于 $i = 1, \cdots, k_1$ **do**

      计算标准误 $\hat{\sigma}_{j\,i}^{priv}$，对于 $\hat{\beta}_{j\,i}^{priv}$；

      计算每种选择的准确度 $h_{j\,i}^{priv}(r_l) = (\hat{\beta}^{\text{最大值}}_{reduced,j} - \hat{\beta}_{j\,i}^{priv})^2 - (\hat{\sigma}_{j\,i}^{priv})^2$

    end

  end

结束

调优参数选为 $argmin_{r_l}\{\min_{i\in[k]} [\sum_{j=1}^{j=v} h_{j,i}^{priv}(r_l)]/v\}$

4 应用

本节将我们提出的框架应用于两个重要场景：（1）多元高斯；（2）线性回归。为简便起见，我们仅讨论隐私化点估计和隐私化参数自助法的实现，因为具体算法可自然地从算法1中的框架推导得出。

4.1 多元高斯情形

考虑一个 k 维多元高斯分布，其中 $\mathbf{x}_i \sim N(\mu, \Sigma)$（$i = 1, .., n$）

$\beta = \mu \in \mathbb{R}^k$ $\gamma = \Sigma \in \mathbb{R}^{k\times k}$ 为关注参数，

$\beta = \mu \in \mathbb{R}^k$ $\gamma = \Sigma \in \mathbb{R}^{k\times k}$ 为干扰参数。于是，

$\beta_{\max}$ 表示总体均值中的最大值，并且通常具有实际意义，例如临床研究中的最佳亚组效应。对于 $\mathbf{x}_i$，密度函数为

$$p(\mathbf{x}_i; \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{k}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{(\mathbf{x}_i-\mu)^T \Sigma^{-1} (\mathbf{x}_i-\mu)}{2}}$$

具有两个充分统计量： $T_1(\mathbf{x}_i) = \mathbf{x}_i, T_2(\mathbf{x}_i) = \mathbf{x}_i\mathbf{x}_i^T$。给定数据集 $\mathcal{X} = \{\mathbf{x}_i\}_{i=1}^n$，对数似然为

$$\ln p(\mathcal{X}; \mu, \Sigma) = -\frac{nk\ln(2\pi)}{2} - \frac{n\ln(|\Sigma|)}{2} - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \mu)^T\Sigma^{-1}(\mathbf{x}_i - \mu),$$

以及通过最大似然估计得到的隐私化点估计为

$$\begin{cases} \hat{\mu}^{priv} = \frac{1}{n}(\sum_{i=1}^{n}\mathbf{x}_i + w_1) \\ \hat{\Sigma}^{priv} = \frac{1}{n-1}(\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^T + w_2) \\ \qquad\quad - \frac{1}{n(n-1)}(\sum_{i=1}^{n}\mathbf{x}_i + w_1)(\sum_{i=1}^{n}\mathbf{x}_i + w_1)^T \end{cases} \tag{7}$$

其中， $w_i \sim Lap(\triangle_i/\varepsilon_i)(i = 1, 2)$ 是拉普拉斯噪声。 $\triangle_i$ 是充分统计量 $\sum T_i(x)$ 的敏感度，而 ε_i 是分配给它的隐私预算。

对于隐私估计算法，我们生成 $\mathbf{x}_1^*, \cdots, \mathbf{x}_n^* \sim N(\hat{\mu}^{priv},$

$$\hat{\mu}^{*,priv} = \frac{1}{n}(\sum_{i=1}^{n}\mathbf{x}_i^* + w_1^*) \tag{8}$$

其中， w_1^* 是由与 w_1 相同的分布生成的拉普拉斯噪声，用于考虑隐私保护中的额外随机性。利用 $\hat{\mu}^{priv}$ 和 $\hat{\mu}^{*,priv}$，我们可以按照算法 2 继续进行。

4.2 线性回归

我们考虑回归情形。尽管受到指数族的启发，我们仍采用了一些修改，以便更好地适应后文详述的参数自助法中的隐私保护。考虑线性模型 $y_i = \mathbf{x}_i^T\beta + e_i$，其中 $e_i \sim N(0, \sigma^2)$，β，$\mathbf{x}_i \in \mathbb{R}^k$，$y_i \in \mathbb{R}$ 对于 $i = 1, ...$，n。于是，$\beta_{\max}$ 表示最大的回归系数，通常具有实际意义，例如遗传关联研究中最强的信号。密度函数为

$$p(\mathbf{x}_i, y_i; \beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y_i - \mathbf{x}_i^T\beta)^2}{2\sigma^2}}$$

令 $X \in \mathbb{R}^{n\times k}$ 表示其第 i^{th} 行等于 $\mathbf{x}_i^T$ 的矩阵，$\mathbf{y}$，$\mathbf{e} \in \mathbb{R}^n$ 分别表示其第 i^{th} 个元素为 y_i 和 e_i 的向量。那么，线性回归问题变为 $\mathbf{y} = X\beta + \mathbf{e}$。其中，对数似然为

$$\ln P(X, \mathbf{y}; \beta, \sigma^2) = -\frac{\mathbf{y}^T\mathbf{y} - 2\beta^T X^T\mathbf{y} + \beta^T X^T X\beta}{2\sigma^2} - \frac{n}{2}\ln(\sigma^2),$$

以及通过最大似然估计得到的 β 的隐私化点估计为

$$\hat{\beta}^{priv} = (X^T X + w_1)^{-1}(X^T\mathbf{y} + w_2). \tag{9}$$

其中 $w_i \sim \text{Lap}(\triangle_i/\varepsilon_i)(i=1、2)$ 为拉普拉斯噪声， $\triangle_i$ 是充分统计量的敏感度， $\sum T_1 = X^T X$ 以及 $\sum T_2 = X^T \mathbf{y}$。继 Ferrando 等人 (2022) 之后，我们对 σ^2 采用带有额外拉普拉斯噪声的偏差校正估计。

$$\hat{\sigma}^{2,priv} = \frac{1}{n-k}\left[\sum_{i=1}^{n}(y_i - \mathbf{x}_i^T \hat{\beta}^{priv})^2\right] + w_3, \tag{10}$$

其中， $w_3 \sim \text{Lap}(\triangle_3/\epsilon_3)$ 是额外添加的拉普拉斯噪声，而 $\triangle_3$ 是方差的敏感度。该方案也适用于下一节将要讨论的部分隐私化设置。

如果我们使用估计模型$\mathbf{y} = X\hat{}\beta^{priv} + \mathbf{e}$ 生成自助样本，就需要多次访问原始数据 X ，并牺牲隐私预算来保护X。Ferrando 等人(2020) 建议改写式（9）中的隐私化最大似然估计$\hat{}\beta^{priv}$ ，并考虑以下 β自助法估计

$$\begin{aligned}\sqrt{n}\hat{\beta}^{*,priv} = &\sqrt{n}(S^{priv} + \frac{1}{n}w_1^*)^{-1} S^{priv}\hat{\beta}^{priv} \\ &+ (S^{priv} + \frac{1}{n}w_1^*)^{-1}(C^{*,priv} + \frac{1}{\sqrt{n}}w_2^*)\end{aligned} \tag{11}$$

其中 $S^{priv} = \frac{1}{n}X^T X + \frac{1}{n}w_1$，我们生成 $C^{*,priv} \sim N(0, \hat{\sigma}^{2,priv} S^{priv})$。 w_1， w_2是相应的拉普拉斯噪声，而 w_1^*， w_2^* 是从与 w_1， w_2 相同的分布中生成的拉普拉斯噪声，用于考虑隐私保护中额外引入的随机性。利用 $\hat{}\beta^{priv}$ 和 $\hat{}\ \beta^{*,priv}$，我们可以按照算法 2继续进行。

5 部分隐私方法

本节介绍一种节省隐私预算的策略：当关注参数 β 的计算和自助法仅依赖于部分充分统计量 T 时，可采用此策略。我们将讨论该策略在多元高斯和线性回归这两个应用中的实现。

5.1 一般隐私预算缩减

在一些实际应用中，如定理 1 所述，关注参数 β 的估计值的发布可能仅依赖于部分充分统计量。

定理 **1** *In the framework of exponential family shown in section 3.1 if* $\frac{\partial A}{\partial f_1} = ,\ (\beta)$ ___
l
l is some function, then the MLE has the form,

$$\begin{cases}\hat{\beta} = g_1(\sum T_1(x_i)) \\ \hat{\gamma} = g_2(\sum T_1(x_i), \sum T_2(x_i)).\end{cases} \tag{12}$$

从式 12 可知， $T_2(x)$ 与发布 $\hat{}\ \beta$ 毫无关系，这意味着在进行估计时，我们可以不向 $T_2(x)$ 添加拉普拉斯噪声，或者具体而言，节省隐私预算。令

$$\begin{cases}\hat{\beta}^{priv} = g_1(\sum T_1(x_i) + w_1) \\ \hat{\gamma} = g_2(\sum T_1(x_i) + w_1, \sum T_2(x_i)).\end{cases} \tag{13}$$

根据公式 13，我们获得的是 ε_1-差分隐私估计，而不是 $(\varepsilon_1+\varepsilon_2)$-差分隐私估计。此外，如果发布 β^* 与$\hat{\gamma}$ 无关，或仅取决于$\hat{\gamma}$ 的一部分，我们可能只

需在估计 γ 时对相关部分添加噪声，从而节省算法 2 中的隐私预算。我们阐述了高斯情形和回归情形下的部分私有化方案，更多细节见附录。

5.2 部分隐私化的多元高斯情形

在应用中，我们可能只关注部分总体均值。假设

$$\mathbf{x}_i = \begin{pmatrix} \mathbf{x}_i^1 \\ \mathbf{x}_i^2 \end{pmatrix} \sim \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right) \tag{14}$$

其中 $\mathbf{x}_i^1 \in \mathbb{R}^{n_1}$，$\mathbf{x}_i^2 \in \mathbb{R}^{n_2}$。那么 $\beta = \mu_1$ 是关注参数，而 $\gamma = (\mu_2, \Sigma)$ 是干扰参数，其中 $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$。我们可以验证，在此多元高斯情形下，$\frac{\partial A}{\partial f_1} = \frac{1}{\mu_1}$，满足定理 1 中的条件。请注意，在参数自助法中，我们只需估计 μ 和 Σ_{11}，而非全部参数，因此可以修改公式 7，并通过仅向充分统计量 $\sum_{i=1}^{n} \mathbf{x}_i^1, \sum_{i=1}^{n} \mathbf{x}_i^1 (\mathbf{x}_i^1)^T$ 添加拉普拉斯噪声 w_1, w_2，推导出部分隐私化极大似然估计。

我们估计

$$\hat{\mu}_1^{*,priv} = \frac{1}{n_1} \left(\sum_{i=1}^{n_1} \mathbf{x}_i^* + w_1^* \right)$$

其中，w_1^* 是从与 w_1相同的分布生成的拉普拉斯噪声。然后，我们按照算法 2 计算 $T^{*,priv}$ 以及隐私化置信下限。至于算法 2 第 3 步和第 7 步中的偏差选择校正部分，我们代入部分样本量 n_1。

通过这一修改，我们可以为 μ_2，Σ_{12}，Σ_{21} 和 Σ_{22}节省隐私预算。

5.3 含干扰参数的线性回归

在许多应用中，我们通常考虑线性回归模型 $y_i = \mathbf{z}_i^T \beta + \mathbf{x}_i^T \gamma + e_i$，$i = 1, \cdots, n$。以亚组分析为例，$\mathbf{z}_i \in \mathbb{R}^{k_1}$ 可以是亚组指标与处理指标之间的交互项，$\mathbf{x}_i \in \mathbb{R}^{k_2}$ 可以是治疗前协变量，而 y_i 是响应向量，$e_i \sim N(0, \sigma^2)$ $i.i.d.$ 是白噪声；参见Imai 等人 (2013)。那么，$\beta \in \mathbb{R}^{k_1}$ 是关注参数，$\beta_{\max}$ 表示最佳亚组效应，而 $\gamma \in \mathbb{R}^{k_2}$ 是干扰参数。

令 $Z \in \mathbb{R}^{n \times k_1}$ 表示一个矩阵，其 i^{th} 行等于 $\mathbf{z}_i^T$；$X \in \mathbb{R}^{n \times k_2}$ 表示一个矩阵，其 i^{th} 行等于 $\mathbf{x}_i^T$ 和 $\mathbf{y}$，$\mathbf{e} \in \mathbb{R}^n$ 表示向量，其 i^{th} 个元素分别为 y_i 和 e_i。

在随机试验等一些实际问题中，我们有 $Z^T X = 0$。我们可以验证，$A(\beta, \gamma, \sigma^2) = \frac{1}{2} \ln \sigma^2$。因此 $\frac{\partial A}{\partial \beta} = 0$。定理 1 的条件得到满足。接下来，我们按照式 9 的方案，仅通过向与 β 相关的两个充分统计量 $\sum T_1 = Z^T Z$ 和 $\sum T_3 = Z^T \mathbf{y}$ 添加拉普拉斯噪声 w_1，w_2，为 β 构造隐私化最大似然估计。

我们采用偏差校正估计 $\hat{\sigma}^{2,priv}$ 来估计 σ^2，并按照式 10 的形式，通过代入 $\hat{\beta}^{priv}$，$\hat{\gamma}$添加额外的拉普拉斯噪声。

类似地，我们构造自助法估计 $\hat{\beta}^{*,priv}$，其遵循式 11 的思路：

$$\begin{aligned}\sqrt{n}\hat{\beta}^{*,priv} = \sqrt{n}(S^{priv} + \frac{1}{n}w_1^*)^{-1}S^{priv}\hat{\beta}^{priv} \\ + (S^{priv} + \frac{1}{n}w_1^*)^{-1}(C^{*,priv} + \frac{1}{\sqrt{n}}w_2^*),\end{aligned} \tag{15}$$

其中 $S^{priv} = \frac{1}{n}Z^TZ + \frac{1}{n}w_1$，我们生成 $C^{*,priv} \sim N(0, \hat{\sigma}^{2,priv}S^{priv})$。$w_1$，$w_2$是相应的拉普拉斯噪声，而 w_1^*，w_2^* 是根据与 w_1，w_2 相同的分布生成的拉普拉斯噪声，用于考虑隐私保护中额外的随机性。有了 $\hat{\beta}^{priv}$ 和 $\hat{\beta}^{*,priv}$后，我们可以按照算法 2继续进行。

6 自助法理论

尽管通常的自助法估计 $\hat{\beta}^{*,priv}_{\max}$无法解决差分隐私下的选择偏差问题，但定理 2指出，在加入校正项且 $0 < r < 0.5$的情况下，我们在算法1中提出的方法能够为 $\beta_{\max}$ 提供有效的下置信限；并且该下置信限是有效的，因为当 n 趋于无穷时，它能够达到名义水平。

定理 2 *For any tuning parameter $0 < r < 0.5$, we have that $\hat{\beta}^{*,priv}_{\text{最大值}\ modified}$ is consistent:*

$$\begin{aligned}&\sup_{x\in R}|P^*(\sqrt{n}(\beta^*_{\max,modified} - \hat{\beta}^{priv}_{\max}) \le x) \\ &- P(\sqrt{n}(\hat{\beta}^{priv}_{\max} - \beta_{\max}) \le x)| \to 0\end{aligned}$$

如 $n \to \infty$，关于 P 依概率收敛。

对于部分隐私化情形，在证明中将 $\hat{\alpha}^{priv}$ 替换为 $\hat{\alpha} = (\hat{\beta}^{priv}, \gamma)$ 后，一致性仍然成立，详细内容见附录。交叉验证和偏差缩减估计的依据也可见附录。

7 实验

本节以多元高斯情形为例，通过蒙特卡洛模拟展示所提方法的优势。其他情形的结果见附录。

我们考虑一个简单的设置，其中数据为 $\mathbf{x_1}, \cdots, \mathbf{x_n} \sim N(\mu, \Sigma)$，其中 Σ 是单位矩阵。关注参数为 μ。我们生成大小为 $n = 800$ 的随机样本，并使用 1000 个蒙特卡洛样本评估经验覆盖率，以及与真实最大值和估计的 95% 置信下限之间的平均距离。我们考虑调优参数 $r = 1/30, 1/15, 1/10, 1/5$，以及通过交叉验证选出的调优参数。

为了进行比较，我们采用朴素隐私化方法，即利用正态近似构造置信区间，并考虑 Guo 和 He (2021) 中估计得到的隐私化极值及其标准误。我们还采用 Bonferroni 方法进行比较。非私有朴素法的结构相同，但不添加拉普拉斯噪声。我们还通过设置调优参数$r = 0.5$来比较一种半朴素自助法，这意味着我们在自助法中不添加偏差校正项。

为便于表述，我们使用以下缩写：(1) **PPB**：私有化参数自助法；(2)**NPB**：非私有参数自助法；(3) **ParPB**：部分私有化参数自助法；(4) **rPPB**：不添加相应拉普拉斯噪声以反映随机性的私有化参数自助法。我们考虑不同的情形，以展示所提方法的优势，具体如下。

偏差校正。我们以设置隐私预算 $\varepsilon = 1.5$ 为例，考察偏差校正的效果，并令 $\mu =$(0，0)或(0，1)。从表1可以看出，偏差校正在隐私设置和非隐私设置下实现名义覆盖率方面都发挥着重要作用。所提出的方法结合交叉验证表现良好。

真实参数		$\mu =$ (0，0)		$\mu =$ (0，1)	
标准		覆盖率	长度	覆盖率	长度
PPB	r =-∞	0.939	0.067	0.952	0.069
	r = 1/30	0.934	0.065	0.947	0.068
	r = 1/15	0.933	0.065	0.945	0.068
	r = 1/10	0.932	0.065	0.943	0.068
	r = 1/5	0.924	0.063	0.939	0.066
	r = 0.5	0.889	0.063	0.910	0.058
	cv	0.934	0.065	0.947	0.068
NPB	r =-∞	0.942	0.078	0.969	0.097
	r = 1/30	0.939	0.077	0.950	0.084
	r = 1/15	0.938	0.076	0.949	0.083
	r = 1/10	0.938	0.076	0.948	0.082
	r = 1/5	0.927	0.074	0.947	0.081
	r = 0.5	0.903	0.055	0.947	0.081
	cv	0.937	0.073	0.947	0.082
私有朴素法		0.875	0.050	0.894	0.082
非私有朴素法		0.905	0.052	0.953	0.083
私有邦费罗尼法		0.912	0.066	0.954	0.100
非私有邦费罗尼法		0.965	0.071	0.976	0.099

表1：高斯情形下的模拟结果（ $k = 2$）

自助法中的随机性。 按照**偏差校正**中的设定，我们通过跳过算法 2 的第 6 步，展示了考虑自助法中的随机性的重要性。在多元高斯情形下，这意味着我们不在式 8 中添加拉普拉斯噪声 w_1^*。结果如表2所示。可以观察到，覆盖率并不令人满意。因此，在使用私有数据构造置信区间时，考虑噪声项引入的随机性至关重要。

隐私预算。 在调优参数 $r = 1/10$ 固定的情况下，我们比较了差分隐私中的 5 个隐私预算 $\varepsilon = 0.1$，0.5，1， 1.5， 2，其余设定遵循**偏差校正**中的设定。覆盖率和长度的结果如图1和图7所示。

可以看出，长度与隐私保护之间存在权衡。 ε较小时，隐私保护更强，但代价是长度增加、效率降低。 ε较大时，由于拉普拉斯噪声引入的随机性较少，长度会缩短，但隐私保护会减弱。这也表明，只要相关，部分隐私方法就有助于提高效率。部分隐私化方法的实验结果见附录，结果表明这种修改效果良好，并能为极值参数提供效率更高的置信区间。

真实参数		$\mu = (0,\ 0)$		$\mu = (0,\ 1)$	
标准		覆盖率	长度	覆盖率	长度
rPPB	r =-∞	0.890	0.063	0.942	0.098
	r = 1/30	0.882	0.062	0.916	0.085
	r = 1/15	0.881	0.061	0.913	0.084
	r = 1/10	0.880	0.061	0.911	0.083
	r = 1/5	0.873	0.059	0.910	0.082
	r = 0.5	0.844	0.051	0.909	0.082
	cv	0.878	0.061	0.912	0.083

表2：高斯情形下无随机性的参数自助法（ $k = 2$）

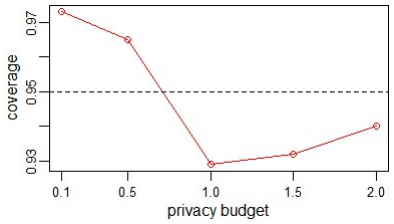

图1：调整隐私预算时的覆盖率

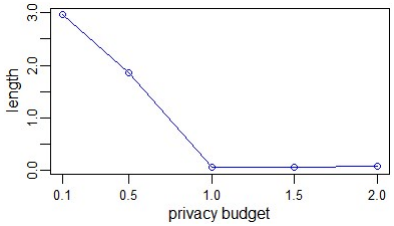

图2：调整隐私预算时的长度

参数维度。沿用**偏差校正**中的设置，但采用$k = 8$，表3报告了调优参数为 $r = 1/10$ 和 $r = 0.5$ 的参数自助法结果，以及朴素方法的结果。隐私预算设为 $\varepsilon = 3$。附录中还展示了 k 较大时的更多结果，这些结果表明，随着维度提高，我们的方法优势会愈发显著。

8 结论

我们提出了一种在隐私方案下为极值参数构造置信区间的方法，该方法高效且易于实现。我们通过理论分析和实验验证了该方法。借助精心设计的隐私化自助法流程，我们在差分隐私下恰当地校正了极值参数的选择偏差，并充分考虑了拉普拉斯噪声引入的随机性。我们还提出了一种部分隐私化策略，可帮助某些应用场景避免浪费隐私预算。

真实参数		$\mu = (0, \cdots, 0)$		$\mu = (0, \ \cdots, \ 0, \ 1)$	
标准		覆盖率	长度	覆盖率	长度
PPB	r = 1/10	0.931	0.058	0.954	0.069
	r = 0.5	0.717	0.054	0.941	0.087
私有朴素法		0.613	0.009	0.928	0.080

表3：含有 $k = 8$ 的高斯情形模拟结果

致谢

本研究得到香港科技大学 IEG19SC04 项目以及河套深港科技创新合作区项目 HZQB-KCZYB-2020083 的资助。

参考文献

Balle, B., Barthe, G., Gaboardi, M., Hsu, J., and Sato, T. (2020). 假设检验的解释与 Renyi 差分隐私。载于 *International Conference on Artificial Intelligence and Statistics* 2496 2506 ,

第 – 页。PMLR。Bornkamp, B., Ohlssen, D., Magnusson, B. P., and Schmidli, H.（2017）。亚组治疗效应估计的模型平均。 *Pharmaceutical statistics*, 16 2 133 142( ): – 。Brawner, T. and Honaker, J.（2018）。自助法推断与差分隐私：免费的标准误。 *Unpublished Manuscript*。Chadha, K., Duchi, J., and Kuditipudi, R.（ 2021 *NeurIPS* ）。 *2021 Workshop Privacy in Machine Learning* 私有置信集。载于 。Covington, C., He, X., Honaker, J., and Kamath, G.（2021）。差分隐私下的无偏统计估计与有效置信区间。 *arXiv preprint arXiv:2110.14465* 。

Dimitrakakis, C., Nelson, B., Mitrokotsa, A., and Rubinstein, B. I.（2014）。稳健且私有的贝叶斯推断。载于 *International Conference on Algorithmic Learning Theory* 291 305 ,

第 – 页。施普林格。Dimitrakakis, C., Nelson, B., Zhang, Z., Mitrokotsa, A., and Rubinstein, B. I.（2017）。通过后验抽样实现贝叶斯推断的差分隐私。 *Journal of machine learning research* 18 11 1 39 ,

( ): – 。Du, W., Foot, C., Moniot, M., Bray, A., and Groce, A.（2020）。差分隐私置信区间。 *arXiv preprint arXiv:2001.02285*。Dunsche, M., Kutta, T., and Dette, H.（2022）。差分隐私下的多元均值比较。载于 *International Conference on Privacy in Statistical Databases*, 第31–45页。施普林格。Dwork, C., McSherry, F., Nissim, K., and Smith, A.（2006）。私有数据分析中根据敏感度校准噪声。载于 *Theory of cryptography conference*, 第265–284页。施普林格。Dwork, C., Roth, A., 等（2014）。差分隐私的算法基础。*Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407。Ferrando, C., Wang, S., and Sheldon, D.（2020）。通用差分隐私置信区间。 *arXiv preprint arXiv:2006.07749*。Ferrando, C., Wang, S., and Sheldon, D.（2022）。差分隐私置信区间的参数自助法。载于 *International Conference on Artificial Intelligence and Statistics* 1598 1618 ,

第 – 页。PMLR。

Fuentes, C., Casella, G., and Wells, M. T.（2018）。选定总体均值的置信区间。 *Electronic Journal of Statistics*，12(1):58–79。Gaboardi, M., Lim, H., Rogers, R., and Vadhan, S.（2016）。差分隐私卡方假设检验：拟合优度与独立性检验。载于 *International conference on machine learning* 2111 2120 ，

第 – 页。PMLR。Guo, X. and He, X.（ 2021 *Journal* ）。 *of the American Statistical Association* 116 535 1498 1506
临床试验中选定亚组的推断。，( ): – 。Guo, X., Wei, W., Liu, M., Cai, T., Wu, C., and Wang, J.（2022）。评估他汀类药物使用相关的2型糖尿病异质性风险：来自电子健康记录数据的证据。 *arXiv preprint arXiv:2205.06960*。Hall, P. and Miller, H.（2010）。参数极值的自助法置信区间与假设检验。 *Biometrika*，97(4):881–892。Imai, K., Ratkovic, M.，等（2013）。随机化项目评估中的治疗效应异质性估计。 *The Annals of Applied Statistics*，7(1):443–470。Karwa, V. and Vadhan, S.（2017）。有限样本差分隐私置信区间。 *arXiv preprint arXiv:1711.03908*。Li, N., Li, T., and Venkatasubramanian, S.（2006）。t-接近度：超越 k-匿名与 l-多样性的隐私保护。载于 *2007 IEEE 23rd International conference on data engineering* 106 115 ，

第 – 页。IEEE。Machanavajjhala, A., Kifer, D., Gehrke, J., and Venkitasubramaniam, M.（2007）。l-多样性：超越 k-匿名的隐私保护。 *ACM Transactions on Knowledge Discovery from Data (TKDD)* 1 1 3 ，

( ): –es。Magnusson, B. P. and Turnbull, B. W.（2013）。纳入亚组选择的组序贯富集设计。 *Statistics in medicine*，32(16):2695–2714。Mologen（2018）。IMPULSE研究最终分析确认了积极的亚组结果及主要数据。 *MOLOGEN Press Releases*。Nadarajah, S. and Kotz, S.（2008）。两个高斯随机变量最大值/最小值的精确分布。 *IEEE Transactions on very large scale integration (VLSI) systems* 16 2 210 212 ，

( ): – 。Rinott, Y., O'Keefe, C. M., Shlomo, N., and Skinner, C.（2018）。频数表发布中的保密性与差分隐私。 *Statistical Science*，33 3 358 385( ): – 。Rogers, R., Roth, A., Smith, A., and Thakkar, O.（2016）。最大信息量、差分隐私与选择后假设检验。载于 *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)* 487 494 ，

第 – 页。IEEE。Rosenkranz, G. K.（2016）。通过模型选择进行临床试验探索性亚组分析。 *Biometrical Journal*，58(5):1217–1228。Sandercock, P. A.（2015）。置信区间简史：或者，别问"治疗有效吗？"，而要问"你有多确定它有效？"。 *Stroke*，46(8):e184–e187。

Stallard, N., Todd, S., and Whitehead, J. (2008). 两个正态均值中较大者经选择后的估计。 *Journal of Statistical Planning and Inference*, 138 6 1629 1638 ( ): – .
Sweeney, L. (
2002 *International*).
journal of uncertainty fuzziness and knowledge-based systems 10 05 557 570
k-匿名：一种保护隐私的模型。,, ( ): – . Thomas, M. and Bornkamp, B. (2017). 比较临床试验中亚组治疗效应估计的方法。 *Statistics in Biopharmaceutical Research*, 9 2
160 171( ): – . Wang, Y., Kifer, D., and Lee, J. (2018). 经验风险最小化的差分隐私置信区间。 *arXiv preprint arXiv:1804.03794*. Woody, S., Padilla, O. H. M., and Scott, J. G. (2022). 稀疏信号的最优选择后推断：一种非参数经验贝叶斯方法。 *Biometrika*, 109(1):1– 16. Xiang, D. and Cai, W. (2021). 健康数据的隐私保护与二次使用：策略和方法。 *BioMed Research International*, 2021. Zhang, Z., Rubinstein, B. I., and Dimitrakakis, C. (2016). 关于贝叶斯推断的差分隐私。载于
Thirtieth AAAI Conference on Artificial Intelligence。Zhao, J. (2017). 贝叶斯差分隐私的组合性质。载于*2017*
IEEE 28th Annual International Symposium on Personal Indoor and Mobile Radio ,,, Communications (PIMRC) 1 5
页码 – 。IEEE。