arXiv:2301.07078v1 [stat.ML] 17 Jan 2023

自适应私有均值估计的快速算法

John Duchi[1,2] Saminul Haque[3] Rohith Kuditipudi[3]

斯坦福大学 [1]统计学、 [2]电气工程和 [3]计算机科学系

2023年1月

摘要

我们设计了一种 (ε, δ)-差分隐私算法，用于估计一个 d 维分布的均值，该分布的协方差未知 Σ，并且算法能根据 Σ 自适应调整。在多对数因子范围内，该估计量相对于诱导的马氏范数 $\|\cdot\|_\Sigma$ 达到最优收敛速率，计算时间为 $\widetilde{O}(nd^2)$，对次高斯分布具有近线性样本复杂度，允许 Σ 退化或秩较低，并能自适应地推广到次高斯性以外的情形。在此项工作之前，其他方法要么需要指数级计算时间，要么需要超线性增长 $n = \Omega(d^{3/2})$，才能在范数 $\|\cdot\|_\Sigma$ 下达到非平凡误差。

1 引言

只有当我们至少拥有一种在样本和计算方面都高效的均值估计量时，才能认为差分隐私理论已臻完善。在维度 d 和样本量 n 的对数因子范围内，我们实现了这两点。

更精确地说，设 P 是定义在 $\mathbb{R}^d$ 上的一个分布，其均值$\mu = \mathbb{E}_P[X]$和协方差 $\Sigma = \mathbb{E}_P[(X-\mu)(X-\mu)^T]$未知；设 $X_i \overset{\text{独立同分布}}{\sim} P, i \le n$。对于估计量 $\widehat{\mu}$，考虑协方差归一化误差 $\mathsf{err}_\Sigma(\widehat{\mu}, \mu) := (\widehat{\mu}-\mu)^T\Sigma^{-1}(\widehat{\mu}-\mu)$。我们给出一个 (ε, δ)-差分隐私估计量 $\widetilde{\mu}$，用于估计 μ；假设向量 $\Sigma^{-1/2}X_i$ 为次高斯，且 $n = \widetilde{\Omega}(d/\varepsilon^2)$，

$$\mathrm{err}_\Sigma(\widetilde{\mu}, \mu) = (\widetilde{\mu}-\mu)^T\Sigma^{-1}(\widetilde{\mu}-\mu) \le \widetilde{O}(1)\left[\frac{d + \log\frac{1}{\delta}}{n} + \frac{d^2\log^2\frac{1}{\delta}}{n^2\varepsilon^2}\right] \tag{1}$$

$\widetilde{}$以至少 $1-\delta$ 的概率成立，其中 $O(1)$ 项隐藏了对 $\Sigma^{-1/2}X$ 的次高斯参数以及 n 中对数因子的依赖。除去 $\log\frac{1}{\delta}$ 因子以及 n 中隐藏的对数因子外，该结果是最优的；此外，该方法可自然扩展到尾部更重的分布，并为其提供类似的近乎最优保证。

通过根据数据本身的协方差 Σ 来衡量误差，我们采用经典理论统计学中熟悉的效率目标：估计量应能适应协变量的结构，并具有（近乎）最优的协方差。当然，均值估计是统计学中最基本的问题之一；七十多年来，我们早已知道样本均值 $\overline{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$ 是有效的 [9, 30], ，达到最优误差$\mathbb{E}[(\overline{X}_n-\mu)^T\Sigma^{-1}(\overline{X}_n-\mu)] = \frac{d}{n}$，并在适当的矩假设下提供高概率保证 [39]。或许显而易见的是，样本均值能够适应分布的协方差：无论 Σ如何，样本均值都是有效的。

然而，当我们要求估计量满足隐私保护时，情况就没那么明朗了。自差分隐私[14, 13] 问世以来的约十六年间，它已成为保护敏感数据的事实标准，并取得了重大的理论进展和成功应用 [17,

3, 1, 28, 18, 12]——据我们所知，目前尚无计算高效的过程，能够针对总体 P 通过其协方差诱导出的自然马氏范数 $\|v\|_\Sigma = \sqrt{v^T\Sigma^{-1}v}$，达到阶最优的样本复杂度。Brown 等人[7] 指出了这一点，并提出了样本高效的过程，即使在Σ未知时，也能在马氏度量下实现较小误差。当协方差 Σ 已知时，只需根据Σ 截断数据，再为所得截尾均值添加协方差与 Σ 成比例的高斯噪声，便足以在近似差分隐私下以基本最优速率（1）私密估计 μ，因此只需 $n = \Omega(d)$个观测值即可估计 μ （例如参见 [6, 7]）。但在更符合现实的 Σ 未知情形下，据我们所知，所有先前工作要么要求样本量为 $n = \Omega(d^{3/2})$；要么计算难以处理，计算时间关于 n 或 d 呈指数级；要么假设 P 为各向同性。此外，其中许多工作还假设 P 服从高斯分布，而这是一项严格的假设，在实践中从未成立。更多讨论见第 1.1 节。

我们的贡献是一种多项式时间的隐私保护估计量（算法 4，PRIVMEAN），其误差与已知协方差时（等价地，数据为各向同性时）可达到的误差相差至多多对数因子。本质上，我们的估计量通过添加协方差与经验协方差的稳定估计成正比的高斯噪声，对经验均值的稳定估计进行隐私保护；其计算时间为 $\tilde{O}(nd^2)$，对于次高斯分布，其样本复杂度近乎线性；它允许 Σ为退化或低秩，并且自然地适用于次高斯性之外的情形。

1.1 相关工作

差分隐私与稳健统计之间有许多联系 [11], ：稳健统计主要致力于开发对离群点和受污染数据不敏感的估计量 [36, 24, 25, 19], ，而差分隐私则使得即使底层数据中的个体发生变化，估计器的输出（分布）也保持相似[14, 13, 11]。Tukey 和 Huber 开创稳健统计已有六十多年 [36, 24], ，但对受污染数据进行估计与推断的统计极限研究，在计算可处理性方面长期进展有限：直到过去十年，研究人员才开发出能够高效计算的方法，甚至可以稳健地估计样本均值 [10]。同样，学界也只是近来才阐明稳健估计中统计考量与计算考量之间的权衡[10]。

很自然会想知道，隐私问题中是否也存在此类权衡。例如，私有查询评估中的经典过程在自然的问题参数下需要指数级运行时间 [20, 15]。同样，在估计中，遵循 Dwork 和 Lei 的“提出、测试、发布”框架 [11], 的一些样本高效隐私估计量 [32, 7, 22] 需要检验给定统计量在移除若干组数据点后是否仍具稳健性，而这在高维情况下可能在计算上难以处理。在这些场景中，直到最近一年左右，才出现了能实现相当样本效率的计算高效估计量 [，例如 22, 27, 4, 2]。我们的均值估计问题是一个显著的例子：它看似简单，却不存在已知的次指数时间且样本高效的算法。具体而言，据我们所知，所有先前工作要么（i）具有指数级运行时间 [7, 31]；要么（ii）样本效率低下 [26, 31], ，要求样本量至少为 $n = \Omega(d^{3/2})$；要么（iii）实质上假设总体协方差 Σ 为各向同性 [26, 6, 23, 31]（严格来说，论文 [23] 允许任意协方差，但其估计量的平方误差至少随总体协方差的条件数 Σ 线性增长，这实际上等同于假设协方差为各向同性 [6]）。这里我们重点介绍了最相关的（近期）例子；更早期工作的综述见论文 [7] 。

这项工作 与我们最密切相关的是 Brown 等人 [7], 的工作，他们也考虑

协方差自适应均值估计，并且同样实现了（近乎）线性的样本复杂度。他们为自适应私有均值估计提供了一条路线，绕开了私有协方差估计；后者的样本复杂度必然为 $\Omega(d^{3/2})$（参见 Dwork 等人给出的下界 [16]），而他们首次实现了样本复杂度 $o(d^{3/2})$，更不用说线性复杂度了。然而，他们的估计器计算起来需要指数时间；此外，尽管其准确性分析与 Σ的条件数无关，但要求 Σ 是满秩的。最后，他们只考虑高斯分布和次高斯分布。

在同期开展的研究中，Hopkins 等人 [22] 提出了一种从隐私估计到鲁棒估计的通用归约，并利用该归约获得了样本复杂度（近乎）最优的隐私估计量。尽管他们的归约具有通用性，但由此得到的估计器仅在某些特殊情况下是高效的，例如，对于代数矩关系允许高效建模的高斯分布；而且，他们关于均值估计的结果假设协方差有界。他们拓展了这一研究方向 [27, 21]，借助平方和（SoS）松弛，对低效的隐私机制进行高效近似。尽管平方和估计器在技术上是高效的，但通常需要较长的多项式运行时间，因此难以扩展到高维场景或大规模数据。不过，与我们的估计量不同，这些估计器能够抵御一定比例数据受到污染的情况。

1.2 组织结构

下面简要介绍本文的结构。第 2 节介绍符号约定，并阐述本文所需的初步隐私定义。我们的主要估计量PRIVMEAN由两个主要部分组成：先以合理的准确性稳定地估计数据的协方差，再估计截断均值并为其添加噪声。我们在第 3节介绍算法，其中第 3.1 节给出协方差估计器，第 3.2 节介绍均值估计器，第 3.3 节给出完整过程；我们在第 PRIVMEAN 节分析其隐私性，并将部分必要的证明推迟到第 4节；部分证明则见第 6 节和第 7节。我们在第 5节分析准确性，并在该节介绍 ADAMEAN（算法 5），它使 PRIVMEAN能够根据观测数据的尺度进行自适应。

2 预备定义、隐私属性与机制

为了使后续内容更加流畅易懂，我们在此介绍符号，并重述全文所采用的隐私定义。我们还将回顾几种标准隐私机制，并给出其行为保证；对于并非完全标准的结果，我们会在附录中提供证明，以确保内容完整。

2.1 符号

半正定矩阵与范数 对于半正定（PSD）矩阵 $A \in \mathbb{R}^{d\times d}$，我们用 $\mathsf{Col}(A)$ 表示其列空间，用 $A^\dagger$ 表示其伪逆，而伪逆的平方根为 $A^{\dagger/2}$。我们用 $\Pi_A := A^\dagger A = A^{\dagger/2}A^{1/2} \in \mathbb{R}^{d\times d}$ 表示投影到 $\mathsf{Col}(A)$ 上的正交投影算子。利用核范数 $\|A\|_* = \sum_{i=1}^n \sigma_i(A)$（即 A 的奇异值之和），我们将半正定矩阵 A, B 之间类似距离的量定义为

$$d_{\mathrm{psd}}(A,B) := \begin{cases} \max\left\{\left\|A^{\dagger/2}(B-A)A^{\dagger/2}\right\|_*, \left\|B^{\dagger/2}(A-B)B^{\dagger/2}\right\|_*\right\} & \text{if } \mathsf{Col}(A)=\mathsf{Col}(B) \\ \infty & \text{otherwise,} \end{cases}$$

若 A 或 B 不是半正定矩阵，则令 $d_{\mathrm{psd}}(A,B) = \infty$。当 A 和 B 可逆时，$d_{\mathrm{psd}}(A,B) = \max\{\|A^{-1/2}BA^{-1/2} - I\|_*, \|B^{-1/2}AB^{-1/2} - I\|_*\}$，不过顺便指出，这并非一种距离。与 $A \succeq 0$ 对应的扩展值马氏范数 $\|\cdot\|_A$ 定义为

$$\|v\|_A^2 := \lim_{t\downarrow 0} v^T(A+tI)^{-1}v = \begin{cases} v^T A^\dagger v & v \in \mathsf{Col}(A) \\ +\infty & \text{otherwise.} \end{cases}$$

当 A 非奇异时，这是标准的 $\|v\|_A = \sqrt{v^T A^{-1} v}$，且该范数具有单调性：若 $A \preceq B$，则对所有 $v \in \mathbb{R}^d$ 都有 $\|v\|_A \geq \|v\|_B$。

集合与划分 对于集合 S、S'，定义距离 $d_{\mathrm{sym}}(S,S') := \max\{|S \setminus S'|, |S' \setminus S|\}$。给定整数 n 和 b，为简化起见，假设 b 整除 n，令 $\mathcal{P}_{n,\,b}$ 表示 $[n]$ 的所有划分组成的集合，其中构成划分的每个子集均包含 b 个元素。我们将 $\mathcal{P}_{n,\,b}$ 中的一个给定划分表示为子集元组 $\mathcal{S} = (S_1, \ldots, S_{n/b})$，其中每个 $S_j \subset [n]$ 都包含 b 个元素，且两两不相交。

分布 我们用 $W \sim \mathsf{Lap}(\sigma)$ 表示 W 服从尺度参数为 σ 的拉普拉斯分布，其概率密度为 $p(w) = \frac{1}{2\sigma}\exp(-|w|/\sigma)$。$X \sim \mathsf{N}(\mu, \Sigma)$ 表示 X 服从均值为 μ、协方差为 $\Sigma \succeq 0$ 的正态分布；若 Σ 不满秩，则表示 X 的支撑集为 $\mu +$ 列空间 Σ。

2.2 隐私定义与基本性质

为了便于推导，我们将使用分布之间的接近性（参见 [12, 第 3 章。5]），因此我们将差分隐私表述为一种分布接近性。

定义 1（(ε,δ)-接近）。若概率分布 P 和 Q 对于所有可测集 $A \subset \mathcal{X}$ 均满足以下条件，则称它们在分布上是 (ε,δ)-接近的，记作 $P \overset{d}{=}_{\varepsilon,\delta} Q$，

$$P(A) \leq e^\varepsilon Q(A) + \delta \quad \text{and} \quad Q(A) \leq e^\varepsilon P(A) + \delta.$$

类似地，若随机变量 X 和 Y 所诱导的分布满足：$\mathbb{P}(X \in \cdot) \overset{d}{=}_{\varepsilon,\delta} \mathbb{P}(Y \in \cdot)$，则称它们是 (ε,δ)-接近的，记作 $X \overset{d}{=}_{\varepsilon,\delta} Y$。

那么，差分隐私 [14, 13] 等价于这一接近性概念：从输入空间 $\mathcal{X}^n$ 到 $\mathcal{Y}$ 的随机化函数（或机制）M，当且仅当对于任意仅有一个元素不同的向量 $x, x' \in \mathcal{X}^n$，满足以下条件时，才是 (ε,δ)-差分隐私的，

$$M(x) \overset{d}{=}_{\varepsilon,\delta} M(x').$$

以下关于接近度的结果是标准的 [12, 第 3] 章。

引理 2.1（基本组合）。设 X, X', Y, Y' 为满足 $X \overset{d}{=}_{\varepsilon_X,\delta_X} X'$ 且 $Y \overset{d}{=}_{\varepsilon_Y,\delta_Y} Y'$ 的随机变量。那么 $(X,Y) \overset{d}{=}_{\varepsilon_X+\varepsilon_Y,\delta_X+\delta_Y} (X',Y')$。

引理 2.2（群组合）。设 $X_1, \ldots, X_k$ 为随机变量，且对每个 i 都有 $X_i \overset{d}{=}_{\varepsilon_i,\delta_i} X_{i+1}$。设 $\varepsilon_{>i} := \sum_{j=i+1}^{k-1} \varepsilon_j$，$\varepsilon = \sum_{i=1}^{k-1} \varepsilon_i$，且 $\delta = \sum_{i=1}^{k} e^{\varepsilon_{>i}} \delta_i$。那么 $X_1 \overset{d}{=}_{\varepsilon,\delta} X_k$。

引理 2.3（后处理）。设 X、Y、W 为随机变量。那么，对于任意函数 f，如果 $X \overset{d}{=}_{\varepsilon,\delta} Y$，则 $f(X,W) \overset{d}{=}_{\varepsilon,\delta} f(Y,W)$。

2.3 机制

我们使用若干已知机制，而我们的过程依赖于它们的分布接近性性质。首先是 TOPk 机制，它（近似地）返回样本中最大的 k个元素。在分析中，为了更容易追踪分布接近性，方便起见，我们将开发的、以噪声为参数的过程称为.....

算法1：Top-k DP（TOPk）

输入：数据 $x \in \mathbb{R}^p$，阈值 k
噪声：$\xi_1, \xi_2 \in \mathbb{R}^p$
输出：$R \subseteq [n]$ 使得 $|R| = k$, $\tilde{x} \in (\mathbb{R} \cup \{\perp\})^p$

$y_1 \leftarrow x + \xi_1$
$y_2 \leftarrow x + \xi_2$
$R \leftarrow$ 包含最大 k 个 $y_{1,j}$ 的索引集合
对 $j \in [p]$ 执行
  如果 $j \in R$，则
    $\tilde{x}_j \leftarrow y_{2,\ j}$
  else
    $\tilde{x}_j \leftarrow \perp$
返回 $\tilde{x}$

引理 2.4 (TOPk 机制，[34], 定理2.1)。令 γ，$\varepsilon \in \mathbb{R}_+$。令 x，$x' \in \mathbb{R}^p$ 满足 $\|x - x'\|_\infty \leq \gamma$。则对于 ξ_1，$\xi_2 \sim$ 拉普拉斯分布$(\frac{2k\gamma}{\varepsilon})^p$,

$$\texttt{TOPk}(x, k; \xi_1, \xi_2) \overset{d}{=}_{\varepsilon,0} \texttt{TOPk}(x', k; \xi_1, \xi_2).$$

由于我们的过程依赖于添加高斯噪声，因此我们需要两个关于正态分布的分布接近性结果。证明见附录A.1 和 A.2 。为完整起见，我们将证明一并列出，因为这两个结果都是对现有结果的调整 [12, 33]。

引理 2.5（均值不同的高斯分布）。设 $\mu_1, \mu_2 \in \mathbb{R}^d$，且 $\Sigma \in \mathbb{R}^{d\times d}$ 为半正定。假设 $\|\mu_1 - \mu_2\|_\Sigma \leq \rho$ 并定义

$$\tau = \begin{cases} \frac{\rho}{\varepsilon}\sqrt{2\log\frac{5}{4\delta}} & \text{if } 0 < \varepsilon \leq 1 \\ \rho / \left(\sqrt{2\log\frac{1}{\delta} + 2\varepsilon} - \sqrt{2\log\frac{1}{\delta}}\right) & \text{otherwise.} \end{cases}$$

那么 $\mathrm{N}(\mu_1, \tau^2\Sigma) \overset{d}{=}_{\varepsilon,\ \delta} \mathrm{N}(\mu_2, \tau^2\Sigma)$。

Brown 等人 [7, 引理 4.15] 基本上给出了下一个结果，但我们允许使用低秩协方差矩阵。

引理 2.6（协方差不同的高斯分布）。设 $\mu \in \mathbb{R}^d$，且 $\Sigma_1, \Sigma_2 \in \mathbb{R}^{d\times d}$ 为半正定，并满足 $d_{\mathrm{psd}}(\Sigma_1, \Sigma_2) \leq \gamma < \infty$。那么对于 $\varepsilon \geq 6\gamma\log(2/\delta)$，$\mathrm{N}(\mu, \Sigma_1) \overset{d}{=}_{\varepsilon,\ \delta} \mathrm{N}(\mu, \Sigma_2)$。

最后，我们给出拉普拉斯随机向量的标准保证 [例如，12，定理 3。6]。

引理 2.7（拉普拉斯机制）。设 α、$\beta > 0$ 和 $Z \overset{\text{独立同分布}}{\sim}$ 拉普拉斯分布(β/α)。那么，对于任意满足 $\|\eta\|_1 \leq \beta$ 的 $A \subseteq \mathbb{R}^m$ 和 $\eta \in \mathbb{R}^m$，

$$\mathbb{P}(Z \in A) \leq \exp(\alpha)\mathbb{P}(Z \in A + \eta).$$

3 个算法

由于我们的估计量及其完整分析相当复杂，我们在此概述一下我们的过程。我们分两个阶段计算该估计量，完整内容见第 4(PRIVMEAN) 节 3.3：先计算稳定协方差估计，再计算稳定均值估计。每个阶段都会谨慎地从数据中剔除离群点，并使用其余观测值计算插件量，以替代常用的插件均值和协方差。

在第一阶段（算法 2，COVSAFE），我们得到一个稳健但非隐私的估计 $\widehat{\Sigma}$协方差。为方便起见，假设 n 是偶数，我们将观测值两两配对，并先令

$$\widehat{\Sigma} := \frac{1}{n}\sum_{i=1}^{n/2}(x_i - x_{n/2+i})(x_i - x_{n/2+i})^T.$$

由于 $x_i - x_{n/2+i}$ 是对称的，我们可以修剪满足 $\|x_i - x_{n/2+i}\|_{\widehat{\Sigma}}$ 为$\widehat{}$大的观测值对（无论总体均值 μ 如何），在剩余观测值上重新计算 Σ，并重复此过程直至收敛。关键在于，这种修剪虽然不提供正式的稳健性$\widehat{}$保证，但对单个样本 x_i 的变化是稳定的，从而确保 Σ 本身稳定。

在第二阶段（算法 3，MEANSAFE）中，我们首先通过依据 $\|\cdot\|_{\widehat{\Sigma}}$ 剔除离群点，得到经验均值的稳健估计 $\widehat{\mu}$ 。使用 $\|x_i\|_{\widehat{\Sigma}}$ 来判断x_i 是否会对均值估计产生影响并不可靠，因为即使 $\|\mu\|_{\widehat{\Sigma}}$ 本身很大，该量对于非离群点也可能任意大；遗憾的是，成对观测值（如稳定协方差估计阶段所用）同样无济于事，因为如果x_i和x_j 以相同方式“偏离”， $\|x_i - x_j\|_{\widehat{\Sigma}}$ 可能会很小。相反，我们将 n 个观测值随机划分为若干组 S ，每组大小为 $O(\log \frac{n}{\delta})$；如果组 S 中任意两个观测值相对于 $\|\cdot\|_{\widehat{\Sigma}}$ 相距甚远，就修剪该组中的所有观测值，这样该组中至少有一对离群观测值。假设两个阶段修剪的观测值总数不太多——我们的许多分析都说明了如何使不同样本 x、x' 中被修剪的观测值保持稳定——我们令 $\widehat{\mu}$ 为未修剪观测值的经验均值，然后发布 $\widetilde{\mu} \sim \mathrm{N}(\widehat{\mu}, \sigma^2(\varepsilon,\delta)\widehat{\Sigma})$，其中隐私预算决定 $\sigma^2(\varepsilon,\delta)$。

3.1 稳定协方差估计

私有均值估计算法的第一个组成部分是协方差估计程序 COVSAFE，见算法 2。该程序从样本 $x \in (\mathbb{R}^d)^n$中移除适当异常的数据点对，然后使用剩余的数据点对来实际构造协方差。该程序在每次迭代 Σ_t 时维护剩余数据的经验协方差 t，使得 $\{\Sigma_t\}$ 构成一个在半正定序下非递增的矩阵序列，并将移除的索引存入一个逐步扩大的集合 R_t，适用于 $t = 1$、2、.....；因此，该程序必定会在至多 $n/2$ 轮索引移除后终止。为便于分析，COVSAFE 还会返回一个执行记录 Γ，其中包含移除的索引和迭代构造的协方差；如果数据极不稳定，以致移除了过多索引，则返回 $\perp$ 。

关键在于，协方差估计具有适当的稳定性（参见第 4 节稍后给出的条件 (C.i) 和(C.ii)），并且对于任意给定的输入，以高概率 x，该算法保证：移除索引 i 后，其输出变化很小；或者，如果数据的方差相对于其自身过大$\widehat{}$，该过程就会直接返回 $\Sigma = \perp$。为便于更清晰地描述第 4 节主要隐私结果所需的精确结果，对于一个假定的界 B、$\|x_i - x_j\|_{\Sigma}^2$可接受的离群点数量 m以及隐私随机性

算法 2：稳健协方差估计（COVSAFE）

输入：数据 $x_{1:n}$参数：阈值 B，阈值 m噪声：$z \in \mathbb{R}^{n/2+1}$、 $w \in \mathbb{R}$1 $x \leftarrow x_{1:n/2} - x_{n/2+1:n}$2 $R_0 \leftarrow \emptyset$，$\Sigma_0 \leftarrow \frac{1}{n}\sum_{i=1}^{n/2} \tilde{x}_i \tilde{x}_i^T$3 收敛 $\leftarrow$ 假， $t \leftarrow 0$4 当未收敛时执行5 $t \leftarrow t+1$， $R_t \leftarrow R_{t-1}$，截断 $\leftarrow 0$6 对于 $i \in [n/2] \setminus R_{t-1}$ 执行 7 如果 2 对数 $\|\tilde{x}_i\|_{\Sigma_{t-1}} + z_i + z_{n/2+1} >$对数（$B$）则8 $R_t \leftarrow R_t \cup \{i\}$9 截断 $\leftarrow$ 截断 $+ 1$10 如果截断 $= 0$ 则 11 收敛 $\leftarrow$ 真， $T \leftarrow t$12 $\Sigma_t \leftarrow \frac{1}{n}\sum_{i \in [n/2] \setminus R_t} \tilde{x}_i \tilde{x}_i^T$13 $\Gamma \leftarrow$ （$[R_i]_{t=0}^{T}$，$[\Sigma_t]_{t=0}^{T}$，T）14 如果 $|R_T| > m + w$ 则 15 返回 $\perp$， Γ16 返回 Σ_T， Γ

变量 Z 和 W 待指定，令

$$(\widehat{\Sigma}, \Gamma) := \texttt{COVSAFE}_{B,m}(x; Z, W), \tag{2a}$$

其中 $\Gamma = ([\Sigma_t]_{t \le T}, [R_t]_{t \le T}, T)$是中间协方差和移除的索引的执行记录，对于 $\tilde{x} = x_{1:n/2} - x_{n/2+1:n}$（如第 1 行所示），定义留一协方差

$$\widehat{\Sigma}_{-i} := \begin{cases} \widehat{\Sigma} - \frac{1}{n} 1\{i \in R_T\} \tilde{x}_i \tilde{x}_i^T & \text{if } \widehat{\Sigma} \neq \perp \\ \perp & \text{otherwise,} \end{cases} \tag{2b}$$

其为 $\widehat{\Sigma}$ 当 COVSAFE 未移除索引对（i， $n/2+i$）时 $\in [n]^2$。

3.2 稳定均值估计

私有均值估计算法的第二个组成部分是样本均值估计器，它会添加与估计的（半正定）噪声协方差相当的噪声，我们抽象地将其称为 $A \in \mathbb{R}^{d \times d}$。过程 MEANSAFE会移除数据 x_i 中的元素 x，这些元素根据 $\|x_i - x_{i'}\|_A$的度量，与数据主体“相距过远”；它通过随机化确保移除的索引具有适当的隐私保护。该算法使用 TOPk选出包含过多离群数据点的索引组，然后移除与这些组相关的所有数据。通过评估（随机）数据组，该过程实现隐私保护：如果大多数数据点按照协方差的度量都与某个中心点足够接近，那么直径较大的组就很少；添加或移除单个数据点 x_i 最多只会导致移除一个组，因此该方法可以私密地返回带噪经验均值。当许多数据点都是离群点时，无论任何单个数据点如何变化，该方法都很可能返回 $\perp$。

算法 3：稳健均值估计（MEANSAFE）

输入：数据 $x_{1:n}$，半正定矩阵 $A \in \mathbb{R}^{d\times d}$参数：阈值 B，批大小 b，阈值 k噪声： $\mathcal{S} = (S_1, \ldots, S_{n/b}) \in \mathcal{P}_{n,b}$, $z, z' \in \mathbb{R}^{n/b}$, $w \in \mathbb{R}$, $z^{\mathsf{N}} \in \mathbb{R}^d$输出：均值估计 $\widetilde{\mu}$1 对于 $j \in [n/b]$，执行2 $D_j \leftarrow \log(\mathrm{diam}_A(x_{S_j}))$~3 $D \leftarrow \mathrm{TOPk}(D, k; z, z')$4 $R \leftarrow \emptyset$, $t \leftarrow 0$ /* 将移除的索引初始化为空 */ 5 对于 $j \in [n/b]$，执行 6 如果 $\widetilde{D}_j \neq \perp$ 且 $\widetilde{D}_j > \log(\sqrt{B}/4)$，则7 $R \leftarrow R \cup S_j$8 $t \leftarrow t + 1$9 $\widehat{\mu} \leftarrow \frac{1}{n-|R|}\sum_{i\notin R} x_i$10 如果 $t > \frac{2k}{3} + w$，则11 $\widetilde{\mu} \leftarrow \perp$12 否则13 $\widetilde{\mu} \leftarrow \widehat{\mu} + A^{1/2} z^{\mathsf{N}}$14 $\Gamma \leftarrow (D, \widetilde{D}, R, t, \widehat{\mu})$15 返回 $\widetilde{\mu}, \Gamma$

为在第 4 节中使用，与 COVSAFE 一样，我们为 MEANSAFE 的输出指定符号。令 $x \in \mathbb{R}^{n\times d}$ 为任意样本， A 为任意半正定矩阵。对于定义下述界的参数 B（针对 $\|x_i - x_j\|_\Sigma^2$）、组大小 b、可接受的离群值数量 k以及隐私随机变量（$\mathcal{S}$、Z、Z'、W、Z^{N}），这些参数稍后均会给出，定义如下

$$(\widetilde{\mu}(x, A), \Gamma(x, A)) := \texttt{MEANSAFE}_{B,b,k}(x, A; \mathcal{S}, Z, Z', W, Z^{\mathsf{N}}). \tag{3}$$

3.3 私有均值估计算法

给定 COVSAFE 和 MEANSAFE，算法 4 (PRIVMEAN) 将二者结合（并采用适当的参数设置）来执行私有均值估计。首先，PRIVMEAN通过 ^计算稳定的协方差估计COVSAFE；假设返回的协方差估计为 $\Sigma \neq \perp$，则^使用MEANSAFE计算截尾均值，并为其添加协方差与 Σ 成比例的高斯噪声。第 2 节 4 中的定理表明，这些参数选择能够保证隐私。

我们简要说明 PRIVMEAN 的运行时间。COVSAFE 的 while 循环（从第 4 行开始）每次迭代都要进行一次 $d \times d$ 矩阵求逆，随后最多进行 $n \geq d$次矩阵-向量乘积，所需时间为 $O(nd^2)$。我们可以在不改变 COVSAFE 行为的前提下修改它，使其在 $m + W_{\mathrm{cov}}$ 次迭代后终止，因为拒绝超过 $m + W_{\mathrm{cov}}$ 个索引即可保证 COVSAFE（因而也包括 PRIVMEAN）返回 $\perp$（见第 14 行）。以高概率，我们有 $m + W_{\mathrm{cov}} = O(\frac{1}{\varepsilon}\log\frac{1}{\delta})$，从而运行时间为 $O(nd^2 \min\{n, \frac{1}{\varepsilon}\log\frac{1}{\delta}\})$。COVSAFE 的运行时间主导 MEANSAFE 的运行时间，因此总运行时间（以高概率）为 $O(nd^2 \min\{n, \frac{1}{\varepsilon}\log\frac{1}{\delta}\})$。另外，我们可以通过将 W_{cov} 截断至 $\frac{1}{\varepsilon}\log\frac{1}{\delta}$ 的量级，将这一期望运行时间转化为同阶的最坏情况运行时间，同时不影响 PRIVMEAN 的隐私性。

算法 4: 协方差自适应私有均值估计（PRIVMEAN）

输入 ： 数据 $x_{1:n}$
参数: 阈值 B, 隐私预算 (ε, δ)
输出: 均值估计 $\widetilde{\mu}$

稳健协方差估计

$m \leftarrow \frac{16}{\varepsilon}$ 对数 $\frac{1}{\delta}$, $m_{\max} \leftarrow m + \frac{16}{\varepsilon}$ 对数 $\frac{1+e^{\varepsilon/4}}{\delta}$

$\sigma_Z \leftarrow \frac{32\sqrt{e}B(m_{\max}+1)}{n\varepsilon}$, $\sigma_{W_{\text{cov}}} \leftarrow \frac{16}{\varepsilon}$

$Z_{\text{cov}} \sim$拉普拉斯分布$(\sigma_Z)^{n/2+1}$, $W_{\text{cov}} \sim$拉普拉斯分布$(\sigma_{W_{\text{cov}}})$

$\widehat{\Sigma}, \Gamma_{\text{cov}} \leftarrow$ COVSAFE$_{B,\ m}(x;\ Z_{\text{cov}},\ W_{\text{cov}})$

if $\widehat{\Sigma} = \perp$ 然后

  返回 $\perp$

私有均值估计

$b \leftarrow 1 +$ 对数$_2 \frac{6n^2}{\delta}$, $k \leftarrow \frac{24}{\varepsilon}$ 对数 $\frac{3}{\delta} - 3$

$\sigma_{\text{top}} \leftarrow \frac{8k}{n\varepsilon}\frac{B\sqrt{e}}{1-B\sqrt{e}/n}$, $\sigma_{\mathsf{N}} \leftarrow \frac{20b\sqrt{B}}{n\varepsilon}\exp(3\sigma_{\text{top}}$对数 $\frac{12n}{b\delta})$, $\sigma_{W_{\text{mean}}} \leftarrow \frac{8}{\varepsilon}$

$\mathcal{S} \sim$ 均匀分布$(\mathcal{P}_{n,b})$, $Z_{\text{top}}, Z'_{\text{t op}} \overset{\text{独立同分布}}{\sim}$ 拉普拉斯分布$(\sigma_{\text{top}})^{n/b}$, $W \sim$ 拉普拉斯分布$(\sigma_{W_{\text{mean}}})$, $Z^{\mathsf{N}} \sim \mathsf{N}(0, \sigma_{\mathsf{N}}^2 I_{d\times d})$

$\widetilde{\mu}, \Gamma_{\text{mean}} \leftarrow$ MEANSAFE$_{B,\ b,\ k}(x,\ \widehat{\Sigma}; \mathcal{S}, Z_{\text{top}}, Z'_{\text{t op}}, W, Z^{\mathsf{N}})$

返回 $\widetilde{\mu}$

4 主要隐私结果

对 PRIVMEAN的分析相当复杂，但其中有四个关键构件。前两个条件涉及我们所称的协方差估计的内部和外部留一稳定性，即 2a）和（2b）中 COVSAFE返回的结果。这些条件要求协方差估计（2）具有适当的稳定性，既要考虑移除一个对协方差估计有贡献的元素 $\widehat{}$（该估计基于输入 Σ x ），也要考虑两个输入之间的稳定性 x, x' 它们在 COVSAFE的第 1 行进行变换后，也就是 $\tilde{x} = x_{1:n/2} - x_{n/2+1:n}$ 和$\tilde{x}' = x'_{1:n/2} - x'_{n/2+1:n}$, 仅在一个元素上有所不同。令 $0 \le a < \infty$ 为稍后确定的常数，并令 $(0,\ 1)$ 为一个概率，考虑以下条件:

$\widehat{}$ $\widehat{}$(C.i) 内部留一稳定性。 设 Σ 和 Σ_{-i} 为 2) 式中 COVSAFE在任意大小为 x 的输入 n上的输出。则对于任意索引 $i \in [n/2]$, 以至少 $1-\gamma$的概率,

$$d_{\text{psd}}(\widehat{\Sigma}, \widehat{\Sigma}_{-i}) \le \frac{a}{n} \quad \text{or} \quad \widehat{\Sigma} = \perp .$$

$\widehat{}$ $\widehat{}$(C.ii) 外部留一稳定性。 设 Σ 和 Σ' 为 COVSAFE 在输入 x x' 、上的输出， n 其大小为 $\tilde{x}$ ， 且 $\tilde{x}'$ 与 $i \in [n/2]$仅在索引 Σ_{-i} 处不同， 其中$\widehat{\Sigma'_{-i}}$ 和 2b) 式中的定义相同。则

$$\widehat{\Sigma}_{-i} \overset{d}{=}_{\varepsilon,\delta} \widehat{\Sigma}'_{-i}.$$

接下来的两个条件涉及带噪截断均值估计（3）MEANSAFE的输出。第一个条件(C.iii) 基本上指出 MEANSAFE在输入x 和 x' 仅有一个元素不同时保持稳定，而第二个条件指出，当MEANSAFE应用于

相同的输入样本 x， x' 但协方差估计不同 A， A' 时保持稳定，只要 A， A' 按照条件(C.i)中的相同意义足够接近。

（C.iii）均值样本稳定性。设 $\widetilde{\mu}$（A， x）为均值 MEANSAFE 在输入协方差（3）上的输出 A 和数据 x，且设 x， x' 仅在一个元素上不同。那么

$$\widetilde{\mu}(x, A) \overset{d}{=}_{\varepsilon,\delta} \widetilde{\mu}(x', A).$$

（C.iv）均值协方差稳定性。如果 $d_{\text{psd}}(A, A') \leq \frac{a}{n}$，那么 $\widetilde{\mu}(x, A) \overset{d}{=}_{\varepsilon,\ \delta} \widetilde{\mu}(x, A')$ 。

条件 （C.i）– （C.iv）构成了证明 PRIVMEAN 满足差分隐私的基本隐私构件，而以下命题——为后续完整定理2 预热——说明如何相对轻松地综合这些条件以实现隐私。

命题 1。设样本 x, x' 仅在一个元素上不同，且 Σ 和 $\widetilde{\mu}(x, \Sigma)$ 以及 Σ' 和 $\widetilde{\mu}(x', \Sigma')$ 分别是输入 x 和 x' 对应的协方差估计值和均值估计值（2）和（3）。设条件（C.i）–（C.iv）成立。那么

$$\widetilde{\mu}(x, \widehat{\Sigma}) \overset{d}{=}_{4\varepsilon,(e^{3\varepsilon}+e^{\varepsilon})\delta+(e^{2\varepsilon}+1)(\delta+\gamma)} \widetilde{\mu}(x', \widehat{\Sigma}').$$

证明。 由于 x 和 x' 相邻，因此存在 $i \in [n/2]$ 使得 $\tilde{x}_{-i} = \tilde{x}'_{-i}$。我们得到一串近似分布相等；结合群体隐私（引理 2.2）所蕴含的分布接近性的传递性，该命题便显而易见。首先，我们证明条件 (C.i) 和 (C.iv) 意味着

$$\widetilde{\mu}(x, \widehat{\Sigma}) \overset{d}{=}_{\varepsilon,\delta+\gamma} \widetilde{\mu}(x, \widehat{\Sigma}_{-i}) \quad \text{and} \quad \widetilde{\mu}(x', \widehat{\Sigma}') \overset{d}{=}_{\varepsilon,\delta+\gamma} \widetilde{\mu}(x', \widehat{\Sigma}'_{-i}).$$

我们证明第一个等式，第二个等式的论证完全相同。将 x 视为固定值，令 $\mathcal{E}$ 表示事件： $d_{\text{psd}}(\Sigma, \Sigma_{-i}) \leq \frac{a}{n}$ 或 $\Sigma = \perp$。那么，对于任意可测集 O，有

$$\begin{aligned}
\mathbb{P}(\widetilde{\mu}(x, \widehat{\Sigma}) \in O) &= \mathbb{E}\left[\mathbb{P}(\widetilde{\mu}(x, \widehat{\Sigma}) \in O \mid \widehat{\Sigma}) 1\{\mathcal{E}\}\right] + \mathbb{E}\left[\mathbb{P}(\widetilde{\mu}(x, \widehat{\Sigma}) \in O \mid \widehat{\Sigma}) 1\{\mathcal{E}^c\}\right] \\
&\overset{(i)}{\leq} \mathbb{E}\left[\left(e^{\varepsilon}\mathbb{P}(\widetilde{\mu}(x, \widehat{\Sigma}_{-i}) \in O \mid \widehat{\Sigma}_{-i}) + \delta\right) 1\{\mathcal{E}\}\right] + \mathbb{P}(\mathcal{E}^c) \\
&\leq e^{\varepsilon}\mathbb{P}(\widetilde{\mu}(x, \widehat{\Sigma}_{-i}) \in O) + \delta + \gamma,
\end{aligned}$$

其中不等式（i）来自条件 (C.iv)，最后一个不等式则由条件 γ 中的概率界 (C.i)得出。其次，我们有如下分布近似

$$\widetilde{\mu}(x, \widehat{\Sigma}_{-i}) \overset{d}{=}_{\varepsilon,\delta} \widetilde{\mu}(x, \widehat{\Sigma}'_{-i})$$

根据条件 (C.ii)，这是因为后处理保持分布接近性（引理 2.3）。最后，我们从均值样本稳定性条件 (C.iii) 可知

$$\widetilde{\mu}(x, \widehat{\Sigma}') \overset{d}{=}_{\varepsilon,\delta} \widetilde{\mu}(x', \widehat{\Sigma}').$$

结合各个分布等式，我们得到

$$\widetilde{\mu}(x, \widehat{\Sigma}) \overset{d}{=}_{\varepsilon,\delta+\gamma} \widetilde{\mu}(x, \widehat{\Sigma}_{-i}) \overset{d}{=}_{\varepsilon,\delta} \widetilde{\mu}(x, \widehat{\Sigma}'_{-i}) \overset{d}{=}_{\varepsilon,\delta+\gamma} \widetilde{\mu}(x, \widehat{\Sigma}') \overset{d}{=}_{\varepsilon,\delta} \widetilde{\mu}(x', \widehat{\Sigma}').$$

应用引理 2.2。 □

最后，我们来到了主要的隐私定理。该定理验证了构成 PRIVMEAN的过程确实满足条件 (C.i)–(C.iv) ，且常数取值适当。我们在此陈述该定理，并给出一个证明；证明由若干引理组成，这些引理明确了条件中出现的常数，而它们的证明将在后文给出。

定理 2. 设 $B < \infty$， $\delta \in (0,\ 1)$，并设 x， $x' \in (\mathbb{R}^d)^n$ 是相邻样本，且设 $\varepsilon \le 8$。定义 $\delta' = (e^{3\varepsilon/4} + e^{\varepsilon/4})\delta + 2(e^{\varepsilon/2}+1)\delta$，并设 $m \in \mathbb{N}$ 与 PRIVMEAN的第 1 行中的定义相同。假设 $\delta \le \frac{1}{n}$，且 n 足够大，使得

$$n \ge \frac{128\sqrt{e}B\log\frac{n(1+e^{\varepsilon/4})}{\delta}}{\varepsilon}\left(m+1+\frac{16}{\varepsilon}\log\frac{1+e^{\varepsilon/4}}{\delta}\right) = O(1)\frac{B\log^2\frac{1}{\delta}}{\varepsilon^2}.$$

那么，$\texttt{PRIVMEAN}_{B,\ (\varepsilon,\delta)}(x)$ 是 (ε, δ')-差分隐私的。

简要说明一下，条件 $\varepsilon \le 8$ 仅为方便起见；稍作修改证明即可允许任意的 ε，代价是定理陈述更为复杂，但其中的n 仍保持同一量级。

证明。 根据命题 1，只需验证条件 (C.i)–(C.iv)，我们将证明它们在参数 $(\varepsilon/4,\ \delta)$ 下均成立。在整个证明过程中，值 $m \in \mathbb{N}$ （PRIVMEAN第 1 行）和参数 $B < \infty$ 均不再赘述，因为无论如何，隐私保证都成立。

首先，我们考虑关于协方差估计的条件 (C.i) 和 (C.ii) 。我们将在第 6 节证明接下来的两个引理。该节将先介绍证明所需的预备知识，然后再给出正式证明。第一个引理给出了验证条件 (C.i)（内部稳定性）的充分条件。设 $z \in \mathbb{R}^{n/2+1}$ 和 $w \in \mathbb{R}$ 是变量——为了便于稍后讨论隐私，它们将是随机变量，但在定义中我们先使用它们——并令

$$\widehat{\Sigma}(x,z,w), \left([R_t]_{t=0}^T, [\Sigma_t]_{t=0}^T, T\right) := \texttt{COVSAFE}_{B,m}(x; z, w), \tag{4a}$$

其中我们省略了执行记录 $([R_t], [\Sigma_t], T)$ 对 (x, z, w) 的依赖关系，并按照定义 (2) 重新定义 $\widehat{\Sigma}_{-i}$：

$$\widehat{\Sigma}_{-i}(x,z,w) := \widehat{\Sigma}(x,z,w) - \frac{1}{n}1\{i \in R_T\}\,\tilde{x}_i\tilde{x}_i^T \tag{4b}$$

当 $\widehat{\Sigma}(x,z,w) \ne \perp$ 时成立，否则为 $\perp$。于是我们有

引理 4.1（内部稳定性）。设 Z_j 独立同分布$\sim$ 拉普拉斯分布(σ)，且 $i \in [n/2]$。那么，以至少 $1 - \exp(-\frac{1}{4\sigma})$ 的概率，要么 $\widehat{\Sigma}(x, Z, w) = \perp$，要么

$$d_{\mathrm{psd}}(\widehat{\Sigma}(x,Z,w), \widehat{\Sigma}_{-i}(x,Z,w)) \le \frac{1}{1-B\sqrt{e}/n}\frac{B\sqrt{e}}{n}.$$

证明引理4.1的细节见第 6.2 节。接下来讨论关于外部稳定性的条件 (C.ii) $\widehat{\Sigma}$ $\widehat{\Sigma}$：COVSAFE 的留一协方差 $\Sigma_{-i}(x,z,w)$ 和 $\Sigma_{-i}(x',z,w)$，分别对应输入样本 x 和 x'，且随机化参数相同 z, w。回顾$\tilde{x} = x_{1:n/2} - x_{n/2+1:n}$ 和 $\tilde{x}' = x'_{1:n/2} - x'_{n/2+1:n}$，可得以下保证：

引理 4.2（外部稳定性）。设 $\gamma \in (0, 1)$， Z_j独立同分布$\sim$拉普拉斯分布(σ_Z)， $W \sim$拉普拉斯分布$(\sigma_{W_{\mathrm{cov}}})$ 且 $k \in \mathbb{N}$。定义 $m_{\max} = m + \sigma_{W_{\mathrm{cov}}}$对数$\frac{1}{\gamma}$和 $\alpha = \frac{1}{\sigma_{W_{\mathrm{cov}}}} + \frac{2\sqrt{e}B}{(m_{\max}+1)\,n\sigma_Z}$ 以及 $\beta = \frac{\gamma}{2} + \frac{n}{2}\exp(-\frac{1}{4\sigma_Z})$。对于所有 $i \in [n/2]$，若 $\tilde{x}_{-i} = \tilde{x}'_{-i}$，则

$$\widehat{\Sigma}_{-i}(x,Z,W) \overset{d}{=}_{2\alpha,(1+e^{\alpha})\beta} \widehat{\Sigma}_{-i}(x',Z,W).$$

该引理实际上表明，移除的索引集合 R 和 R' 是稳定的，并且由于它们$\widehat{\quad}$ $\widehat{\quad}$决定了 Σ_{-i} 和 Σ'_{-i}，因此二者彼此接近。引理4.2的证明见第4.2节。

接下来我们讨论 MEANSAFE 的保证，以验证条件 (C.iii) 和 (C.iv)。回顾定义 (3)：$\widetilde{\mu}(x, A)$ 是 MEANSAFE 在输入 $x \in (\mathbb{R}^d)^n$ （其 $A \in \mathbb{R}^{d\times d}$ 为半正定）时的输出；其参数包括上界 B、批大小 b、阈值 $k \in \mathbb{N}$，以及作为噪声的 $\mathcal{S}$、Z、 Z' 和 W 。我们令 Z、$Z' \in \mathbb{R}^{n/b}$、$W \in \mathbb{R}$ 为拉普拉斯随机变量，$Z^{\mathsf{N}} \in \mathbb{R}^d$ 为高斯分布， $\mathcal{S}$ 为将 $[n]$ 均匀随机划分为大小为 n/b 的块所得的划分；在给出分布近似保证时，我们会跟踪它们的尺度。

为更清晰地陈述一般的样本稳定性保证（我们可以用它来验证条件 (C.iii)），我们定义若干可确定取值的附加参数。令批大小 $b \in \mathbb{N}$和阈值 $k > 0$满足 $b \geq 4$和 $2b(k+1) \leq n$。令 β_1， $\gamma \in (0,\ 1)$，令 $\alpha \geq 0$，并令 $\sigma_{\text{top}} > 0$ 和 $\sigma_{W_{\text{mean}}} > 0$。定义常数

$$\Delta := \frac{5b\sqrt{B}}{2n} \exp\left(3\sigma_{\text{top}} \log \frac{2n}{b\gamma}\right),\ \beta_2 := \frac{1}{2}e^{-(k/3-1)/\sigma_{W_{\text{mean}}}} + \gamma + n^2 2^{1-b}$$

and

$$\sigma_{\mathsf{N}} = \begin{cases} \frac{\Delta}{\alpha}\sqrt{2\log\frac{5}{4\beta_1}} & \text{if } 0 \leq \alpha \leq 1 \\ \frac{\Delta}{\sqrt{2\log\frac{1}{\beta_1}+2\alpha}-\sqrt{2\log\frac{1}{\beta_1}}} & \text{otherwise.} \end{cases}$$

由此，我们得到一个均值样本稳定性结果，由此可推出条件 (C.iii)：

引理 4.3。设上述条件成立，并设 Z_j, Z'_j 独立同分布$\sim$ 拉普拉斯分布(σ_{top}) $W \sim$ 拉普拉斯分布$(\sigma_{W_{\text{mean}}})$ Z_j^{N}独立同分布$\sim\mathsf{N}(0, \sigma_{\mathsf{N}}^2)$ 满足 (3)。若 x 和 x' 相邻，则

$$\widetilde{\mu}(x, A) \overset{d}{=}_{\alpha+1/\sigma_{W_{\text{mean}}},\beta_1+\beta_2} \widetilde{\mu}(x', A).$$

证明见第 7.1 节。

论证的最后一个组成部分是证明条件 (C.iv)，即当 $\widetilde{\mu}(x, A)$ 和 $\widetilde{\mu}(x, A')$ 接近时，估计值 A, A' 也接近。为此，我们给出以下引理，其中噪声参数是一般的。

引理 4.4。设 $b, k \in \mathbb{N}$, $\beta \in (0, 1)$，且 $a, \sigma_{\mathsf{N}}, \alpha_2 > 0$。定义 $\alpha_1 = \frac{6a}{n}$对数 $\frac{2}{\beta}$，并定义噪声尺度 $\sigma_{\text{top}} = \frac{ka}{n\alpha_2}$。那么对于 Z_j, Z'_j 独立同分布 $\sim$ 拉普拉斯分布 (σ_{top}) Z_j^{N} 独立同分布 $\sim \mathsf{N}(0, \sigma_{\mathsf{N}}^2)$，若

$$d_{\text{psd}}(A, A') \leq \frac{a}{n} \quad \textit{then} \quad \widetilde{\mu}(x, A) \overset{d}{=}_{\alpha_1+\alpha_2,\beta} \widetilde{\mu}(x, A').$$

证明见第 7.2 节。

最后一步，我们将所有部分组合起来证明该定理。我们给出各个引理，使相应的条件（(C.i)–(C.iv)）在参数（$\varepsilon/4$， δ）下成立，之后便可直接应用命题 1 。由于不同引理之间依赖于噪声尺度，我们按一个略显反常的顺序进行，首先从

条件（C.ii）。对于 Z_j独立同分布$\sim$拉普拉斯分布(σ_Z) 和 $W \sim$拉普拉斯分布$(\sigma_{W_{\text{cov}}})$，我们使用引理 4.2 来保证条件（C.ii），即 $\widehat{\Sigma}_{-i} \overset{d}{=}_{\varepsilon/4,\delta} \widehat{\Sigma}'_{-i}$。根据引理陈述，我们有 $\widehat{\Sigma}_{-i} \overset{d}{=}_{2\alpha,(1+e^{\alpha})\beta} \widehat{\Sigma}'_{-i}$，其中 $\alpha = \frac{1}{\sigma_{W_{\text{cov}}}} + {}^{2\sqrt{e}B}(m_{\max}+1)\, n\sigma_Z$， $\beta = \frac{\gamma}{2} + \frac{n}{2}\exp(-\frac{1}{4\sigma_Z})$，且 $m_{\max} = m + \sigma_{W_{\text{cov}}}$对数 $\frac{1}{\gamma}$用于 PRIVMEAN 第 1 行中的 m （尽管隐私不取决于其值）。我们首先实现

$2\alpha \le \frac{\varepsilon}{4}$。设定 $\sigma_{W_{\text{cov}}} = \frac{16}{\varepsilon}$，只需令 σ_Z足够大，使得 $\frac{2\sqrt{e}B}{(m_{\max}+1)n\sigma_Z} \le \frac{\varepsilon}{16}$，即

$$\sigma_Z \ge \frac{32\sqrt{e}B(m_{\max}+1)}{n\varepsilon} = \frac{32\sqrt{e}B}{n\varepsilon}\left(m+1+\frac{16\log\frac{1}{\gamma}}{\varepsilon}\right).$$

对于 δ隐私部分，我们希望有 $(1+e^{\alpha})\beta \le \delta$。由于我们已保证 $\alpha \le \frac{\varepsilon}{4}$，取 $\gamma = \frac{\delta}{1+e^{\varepsilon/4}}$并确保 σ_Z足够小，使得 $n\exp(-\frac{1}{4\sigma_Z}) \le \gamma = \frac{\delta}{1+e^{\varepsilon/4}}$ 即可。显然，只需满足 $\frac{1}{\sigma_Z} \ge 4\log\frac{n}{(1+e^{\varepsilon/4})\,\delta}$，也就是说（代入 σ_Z）

$$n \ge \frac{128\sqrt{e}B\log\frac{n(1+e^{\varepsilon/4})}{\delta}}{\varepsilon}\left(m+1+\frac{16}{\varepsilon}\log\frac{1+e^{\varepsilon/4}}{\delta}\right)$$

保证 $\widehat{\Sigma}_{-i} \overset{d}{=}_{\varepsilon/4,\ \delta} \widehat{\Sigma}'_{-i}$。

条件 **(C.i)**。引理 4.1中，如果噪声的尺度 σ_Z 在 Z_j 独立同分布$\sim$ 拉普拉斯分布(σ_Z) 满足 $\exp(-\frac{1}{4\sigma_Z}) \le \gamma$，则条件 (C.i) 成立。选择 σ_Z 以满足上述条件 (C.ii) 以及对 n 的下界，显然都是充分的。

条件（C.iii）。引理 4.3 保证，若 Z_j^{N}独立同分布$\sim \mathsf{N}(0,\sigma_{\mathsf{N}}^2)$、 $W \sim$拉普拉斯分布$(\sigma_{W_{\text{mean}}})$，且$\mathcal{S} \sim$均匀分布$(\mathcal{P}_{n,b})$，则 PRIVMEAN 第 10 行对 MEANSAFE 的调用会给出 $\widetilde{\mu}(x,A) \overset{d}{=}_{\alpha+1/\sigma_{W_{\text{mean}}},\beta_1+\beta_2}\widetilde{\mu}(x',A)$，其中 Δ、 β_2和 σ_{N}如引理中所定义。要实现 $\alpha + \frac{1}{\sigma_{W_{\text{mean}}}} = \frac{\varepsilon}{4}$，取$\sigma_{W_{\text{mean}}} = \frac{8}{\varepsilon}$并选择 $\alpha = \frac{\varepsilon}{8}$。要实现 $\beta_1+\beta_2 \le \delta$，选择 $\beta_1 = \frac{\delta}{2}$，然后注意到，只要 $\gamma \le \frac{\delta}{6}$, $n^2 2^{1-b} \le \frac{\delta}{6}$（或 $b \ge$对数$_2 \frac{6n^2}{\delta}+1$）且 1 2 $\exp(-\frac{k/3+1}{\sigma_{W_{\text{mean}}}}) \le \frac{\delta}{6}$（或$k \ge \frac{24}{\varepsilon}$对数 $\frac{3}{\delta}-3$），就有 $\beta_2 \le \frac{\delta}{2}$。因此，我们得到

$$\sigma_{\mathsf{N}} = \frac{8\Delta}{\varepsilon}\sqrt{\log\frac{5}{2\delta}} = \frac{20\sqrt{B}b}{n\varepsilon}\exp\left(3\sigma_{\text{top}}\log\frac{12n}{b\delta}\right)$$

（对任意） $b \ge 2$对数 $\frac{6n}{\delta}+1$只要 $\frac{\varepsilon}{8} \le 1$。（否则，我们可以使用 σ_{N}前述引理 4.3 的另一取值，该引理的（ε， δ）-差分隐私保证由引理 2.5 提供了正当依据。）

条件（C.iv）。需要验证的最后一个条件是，对于足够接近的 A、 A'， $\widetilde{\mu}(x,A) \overset{d}{=}_{\varepsilon/4,\ \delta} \widetilde{\mu}(x,A')$。为此，我们使用引理 4.4，该引理保证当 $\alpha_1 = \frac{6a}{n}$对数 $\frac{2}{\delta}$ 时， $\widetilde{\mu}(x,A) \overset{d}{=}_{\alpha_1+\alpha_2,\ \delta} \widetilde{\mu}(x,A')$，其中我们通过引理 4.1 取 $a = \frac{B\sqrt{e}}{1-B\sqrt{e}/n}$，并令 $\alpha_2 > 0$为任意值。令$\alpha_2 = \frac{\varepsilon}{8}$，即可得到 $\sigma_{\text{top}} = \frac{8ka}{n\varepsilon}$。当 $n \ge \frac{48a}{\varepsilon}$对数 $\frac{2}{\delta}$ 时，我们有 $\alpha_1 \le \frac{\varepsilon}{8}$，因此所需的隐私条件成立。

进行适当的代入后，可得条件 (C.i)–(C.iv) 均以参数（$\varepsilon/4$， δ）成立。命题 1 即给出该定理。□

5 准确性分析

分析 PRIVMEAN 的第二个重要部分是其准确性。我们给出两项准确性结果：第一项（定理 3）讨论数据服从次高斯分布的情况，

其中，我们假设该方法对抽样分布的次高斯参数有所了解。当然，假设给定分布服从次高斯分布，或假设我们知道其次高斯范数，是不合理的。因此，我们在第 5.2 节通过一种能够适应数据实际尺度的过程扩展 PRIVMEAN。

在本节中，我们设 P 为定义在 $\mathbb{R}^d$ 上的分布，其均值为 μ，协方差为 Σ，并假设 $X_i \overset{\text{独立同分布}}{\sim} P, i = 1, \ldots, n$。经典的（非隐私）样本均值和协方差分别为 $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ 和 $\overline{\Sigma}_n = \frac{1}{n}\sum_{i=1}^{n/2}(X_i - X_{n/2+i})(X_i - X_{n/2+i})^T$。我们始终假设 P 具有某些集中性质，但要强调的是，我们的方法能够适应此处指定的参数。

假设 A1（样本集中性）。令 $c_1 \geq 64e$ 和 $\beta \in (0, 1)$。对于满足$\mathbb{E}[X] = \mu$ 且 $\mathrm{Cov}(X) = \Sigma$ 的 $X_i \overset{\text{独立同分布}}{\sim} P$，存在 $M < \infty$，使得事件

$$\mathcal{E}_{\text{samp}} := \left\{ \max_{i\in[n]} \|X_i - \mu\|_\Sigma^2 \leq M^2/c_1 \text{ and } \tfrac{1}{2}\Sigma \preceq \overline{\Sigma}_n \preceq \tfrac{3}{2}\Sigma \right\}$$

发生的概率至少为 $1 - \beta$。

了解在各种分布假设下我们预期 M 取何值会很有帮助。由于 $\mathbb{E}[\|X_i - \mu\|_\Sigma^2] = d$，常数 M^2 通常至少按 d的量级增长。下面给出更详细的样本。在每个例子中，我们令 $Z_i = \Sigma^{-1/2}(X_i - \mu)$ 为白化数据，并据此定义样本协方差 $\overline{\Sigma}_Z = \frac{1}{n}\sum_{i=1}^{n/2}(Z_i - Z_{n/2+i})(Z_i - Z_{n/2+i})^T$。由于$\|X_i - \mu\|_\Sigma = \|Z_i\|_2$ 和 $\overline{\Sigma}_n = \Sigma^{1/2}\overline{\Sigma}_Z\Sigma^{1/2}$，我们有如下等价关系

$$\mathcal{E}_{\text{samp}} = \left\{ \max_{i\in[n]} \|Z_i\|_2^2 \leq M^2/c_1 \text{ and } \left\|\overline{\Sigma}_Z - I\right\|_{\text{op}} \leq \frac{1}{2} \right\}.$$

示例 1（次高斯随机向量）：如果对于所有满足 $\|v\|_2 \leq 1$ 的 v s，标量随机变量 $\langle Z, v\rangle$ 是 τ^2-次高斯的，

$$M^2 \leq O(1)\tau^2 \left[d + \log\frac{n}{\beta} \right].$$

事实上，标准的覆盖论证（参见例如 [39, 第 5] 章或 [38, 第 4]章）表明，对于所有$t \geq 0$，$\mathbb{P}(\|Z\|_2 \geq t) \leq 4^d \exp(-ct^2/\tau^2)$，其中 $c > 0$ 是一个数值常数。将 t^2 替换为$O(1)(d\tau^2 + \tau^2 t^2)$，可得 $\mathbb{P}(\|Z\|_2 \geq C\tau\sqrt{d + t^2}) \leq \exp(-t^2)$；对于任意 $\gamma > 0$，令$t^2 = \log\frac{n}{\gamma}$，则以至少 $1 - \gamma$ 的概率有：

$$\max_{i\leq n} \|Z_i\|_2^2 \leq O(1)\tau^2 \left[d + \tau^2 \log\frac{n}{\gamma} \right].$$

为了控制协方差，我们使用 Vershynin [37, 定理 5.39]，，该定理表明，以至少 $1 - 2e^{-ct^2}$ 的概率，$\left\|\overline{\Sigma}_Z - I\right\|_{\text{op}} \leq O(1)\tau^2 \max\{\sqrt{d/n} + t/\sqrt{n}, d/n + t^2/n\}$，因此（忽略次高斯常数），对于 $n \gtrsim d$，令 $t^2 = O(1)\log\frac{1}{\gamma}$，可得以至少 $1 - \gamma$ 的概率有 $\left\|\overline{\Sigma}_Z - I\right\|_{\text{op}} \leq \frac{1}{2}$。令 $\gamma = \beta/2$。 ◊

样本 **2** $p \geq 4$ $\mathbb{E}[\|X_i - \mu\|_\Sigma^p] =$ （一般矩界）：$\mathbb{E}[\|Z_i\|_2^p] \leq \tau^p d^{p/2}$ $\tau \geq 1$ 假设对于某个值，我们有 ，其中 必然满足 。那么我们可以得到两个结果：第一个是渐近地 $M = o(n^{1/p})$，第二个则更具定量性。对于第一个结果，我们声称 $\max_{i\leq n}\|Z_i\|_2 / n^{1/p} \xrightarrow{a.s.} 0$。为此，注意对于任意 $\varepsilon > 0$，

$$\infty > \frac{1}{\varepsilon^p}\mathbb{E}[\|Z_1\|_2^p] = \int_0^\infty \mathbb{P}(\|Z_1\|_2^p \geq \varepsilon^p t)\, dt \geq \sum_{i=1}^\infty \mathbb{P}(\|Z_i\|_2^p \geq \varepsilon^p i).$$

根据博雷尔-坎特利引理，事件 $\|Z_n\|_2 \geq \varepsilon n^{1/p}$ 仅发生有限次，因此结论成立。同时，强大数定律保证 $\overline{\Sigma}_Z \overset{a.s.}{\to} I$。

对于更具体的参数，我们首先利用马尔可夫不等式得到

$$\mathbb{P}(\max_{i\in[n]} \|Z_i\|_2 > t) \leq \frac{n\mathbb{E}[\|Z_1\|_2^p]}{t^p} \leq \frac{n\tau^p d^{p/2}}{t^p},$$

因此，令$M \asymp \tau\sqrt{d}n^{1/p}/\beta^{1/p}$，我们有 $\max_i \|Z_i\|_2 \leq M/c_1$，其成立概率至少为 $1-\beta$。为证明协方差矩阵的集中性，我们应用 Chen 等人的 [8, 定理 A.1 第2], 部分 p，将其视为常数，得到

$$\begin{aligned} n\mathbb{E}\big[\,\big\|\overline{\Sigma}_Z - I\big\|_{\mathrm{op}}^{p/2}\,\big]^{2/p} &\lesssim \sqrt{n\log d}\sqrt{\mathbb{E}[\|Z\|_2^4]} + (n^{2/p}\log d)\mathbb{E}[\|Z_1\|_2^p]^{2/p} \\ &\lesssim \max\{\sqrt{n\log d}, n^{2/p}\log d\}\mathbb{E}[\|Z_1\|_2^p]^{2/p}, \end{aligned}$$

因此，由马尔可夫不等式可得

$$\mathbb{P}(\big\|\overline{\Sigma}_Z - I\big\|_{\mathrm{op}} > \tfrac{1}{2}) \lesssim \max\{n^{-p/4}\log^{p/4} d, n^{1-p/2}\log^{p/2} d\}\mathbb{E}[\|Z_1\|_2^p] \lesssim \frac{\tau^p(d\log d)^{p/2}}{n^{p/2-1}},$$

其界为 β当 $n \gtrsim$(τ²d对数 d)$^{p/(p-2)}\beta^{-2/(p-2)}$ 时。 ◊

5.1 PRIVMEAN 的准确性

只要假设A1 成立，我们就能给出所承诺的准确性保证。尽管并非严格必要，为了得到更简洁的结果，我们在陈述定理时假设 δ 不太小。全文中， c 表示一个数值常数，其取值可能随行变化。

定理 3. 令 $\varepsilon > 0$和 $e^{-d} \leq \delta \leq \frac{1}{n}$为隐私参数，并假设假设 *A1*成立。令 $B \geq M^2$，且假设 $n \geq \frac{c}{\varepsilon^2}B\log^2\frac{1}{\delta}$。令 $\widetilde{\mu} =$PRIVMEAN$_{B,\ \varepsilon,\ \delta}(X_{1:n})$。那么，以至少 $1-(\beta+5\delta)$ 的概率， $\widetilde{\mu} \neq \perp$ 和

$$\big\|\widetilde{\mu} - \overline{X}_n\big\|_\Sigma \leq \frac{c\sqrt{Bd}\log(\frac{1}{\delta})}{n\varepsilon}.$$

证明。我们首先证明，在事件 $\mathcal{E}_{\mathrm{samp}}$ 条件下，对 PRIVMEAN 中的随机性取概率，COVSAFE 和 MEANSAFE均以至少 $1-4\delta$ 的概率不剔除任何观测值，即两个过程中的移除索引集合 $R = \emptyset$ 均为空（因此，COVSAFE 中的第 7 行和 MEANSAFE 中的第 6 行都不会失败），从而 $\widetilde{\mu} = \overline{X}_n + \overline{\Sigma}_n^{1/2}Z^{\mathsf{N}}$。由于 $\overline{\Sigma}_n \preceq \frac{3}{2}\Sigma$ 在 $\mathcal{E}_{\mathrm{samp}}$ 上，我们有$\|\overline{\Sigma}_n^{1/2}Z^{\mathsf{N}}\|_\Sigma^2 \leq \frac{3}{2}\|Z^{\mathsf{N}}\|_2^2$。只要证明 $\|Z^{\mathsf{N}}\|_2 \leq {}^{c\sqrt{Bd}}{}_{\log(\frac{1}{\delta})n\varepsilon}$ 以至少 $1-\delta$ 的概率成立，并对这些事件及 $\mathcal{E}_{\mathrm{samp}}$ 使用并集界，即可得到结论。

重新整理第7 行COVSAFE中的条件后，元素 $X_i - X_{n/2+i}$ 仅当满足以下条件时才会在第一次迭代中被剪枝:

$$\begin{aligned} (Z_{\mathrm{cov}})_i + (Z_{\mathrm{cov}})_{n/2+1} &> \log(B) - \log(\big\|X_i - X_{n/2+i}\big\|_{\overline{\Sigma}_n}^2) \\ &\overset{(\star)}{\geq} \log(c_1 B/8M^2) \geq \log(c_1/8), \end{aligned}$$

其中，(⋆) 对所有 $i \in [n/2]$ 都成立，这是因为事件 $\mathcal{E}_{\mathrm{samp}}$ 上

$$\big\|X_i - X_{n/2+i}\big\|_{\overline{\Sigma}_n}^2 \leq 2\big\|X_i - X_{n/2+i}\big\|_\Sigma^2 \leq 4\|X_i - \mu\|_\Sigma^2 + 4\big\|X_{n/2+i} - \mu\big\|_\Sigma^2 \leq 8M^2/c_1. \quad (5)$$

由于 $c_1 \geq 64e$ 根据假设A1，如果 $\|Z_{\text{cov}}\|_\infty \leq 1/2$ ，那么第 7 行中的 COVSAFE 不会剪除任何条目，而只会在 $W_{\text{cov}} + m > 0$ 时将 $\overline{\Sigma}_n$ 传递给 MEANSAFE （参见第 14 行）。回想一下，$(Z_{\text{cov}})_j \overset{\text{独立同分布}}{\sim}$ 拉普拉斯分布(σ_Z)，其中 $j = 1, \ldots, n/2 + 1$，且 $\sigma_Z = \frac{32\sqrt{e}B}{n\varepsilon}(m_{\max}+1) \leq \frac{cB\,\log\frac{1}{\delta}}{n\varepsilon^2}$，因此对各条目应用并集界可得，以至少 $1-\delta$ 的概率

$$\|Z_{\text{cov}}\|_\infty \leq \frac{cB\log\frac{1}{\delta}}{n\varepsilon^2}\log\left(\frac{n/2+1}{\delta}\right) \leq \frac{1}{2},$$

其中，最后一个不等式由假设 $n \geq \frac{cB\,\log\frac{1}{\delta}}{\varepsilon^2}$ 和 $\delta \leq \frac{1}{n}$ 得出。另请回忆，$W_{\text{cov}} \sim$ 服从拉普拉斯分布 $\text{Lap}(\frac{16}{\varepsilon})$，且 $m = \frac{16}{\varepsilon}\log\frac{1}{\delta}$，因此 $W_{\text{cov}} + m > 0$ 以至少 $1 - \frac{\delta}{2}$ 的概率成立。

继续进入 PRIVMEAN 的下一阶段，MEANSAFE以 $A = \overline{\Sigma}_n$ 为输入，仅当满足以下条件时才会剪除索引 S_j

$$\widetilde{D}_j = D_j + (Z'_{\text{top}})_j > \log(\sqrt{B}/4).$$

根据我们用于推导不等式（5）的相同论证，在 $\mathcal{E}_{\text{samp}}$ 上，对所有 $j \in [n/b]$，我们有

$$D_j = \log(\text{diam}_{\overline{\Sigma}_n}(X_{S_j})) \leq \log(\sqrt{8M^2/c_1}),$$

因此，如果 $\|Z'_{\text{top}}\|_\infty \leq 1/2$ ，则

$$\widetilde{D}_j \leq \log(\sqrt{8M^2/c_1}) + \frac{1}{2} \leq \log(\sqrt{B}/4),$$

其中，最后一个不等式由 $c_1 \geq 64e$ 和 $B \geq M^2$ 得出。因此，MEANSAFE 不会剪除任何条目，并且只要 $W_{\text{mean}} + \frac{2k}{3} > 0$ ，$\widetilde{\mu} = \overline{X}_n + \overline{\Sigma}_n^{1/2} Z^{\mathsf{N}}$ （见第 10 行）。回忆一下，对于 $j = 1, \ldots, n/b$，$(Z'_{\text{top}})_j$ 均独立同分布$\overset{\text{且服从}}{\sim}\text{Lap}(\sigma_{\text{top}})$，并且 $\sigma_{\text{top}} = \frac{8k}{n\varepsilon}\frac{B\sqrt{e}}{1-B\sqrt{e}/n} \leq \frac{cB}{n\varepsilon^2}\log(\frac{1}{\delta})$。再由并集界可知，以至少 $1-\delta$ 的概率，

$$\|Z_{\text{cov}}\|_\infty \leq \frac{cB\log(\frac{1}{\delta})}{n\varepsilon^2}\log\frac{n}{b\delta} \leq \frac{1}{2},$$

其中，最后一个不等式由假设 $n \geq \frac{cB}{\varepsilon^2}\log(\frac{1}{\delta})$ 和 $\delta \leq \frac{1}{n}$ 得到。此外，$W_{\text{mean}} \sim$拉普拉斯分布$(\frac{8}{\varepsilon})$和 $\frac{2k}{3} = \frac{16}{\varepsilon}\log\frac{3}{\delta} - 2$，因此以至少 $1-\delta$的概率有 $W_{\text{mean}} - k > \frac{8}{\varepsilon}\log\frac{3}{\delta} - 2 \geq 0$。

因此，PRIVMEAN 在事件$\mathcal{E}_{\text{samp}}$上以至少 $1-4\delta$ 的概率返回 $\overline{X}_n + \overline{\Sigma}_n^{1/2} Z^{\mathsf{N}}$。回顾 $Z^{\mathsf{N}} \sim \mathsf{N}(0, \sigma_{\mathsf{N}}^2 I)$，其中 $\sigma_{\mathsf{N}} = \frac{20b\sqrt{B}}{n\varepsilon}\exp(3\sigma_{\text{top}}\log\frac{12n}{b\delta})$，并且由于 $\sigma_{\text{top}} \leq cB\log(\frac{1}{\delta})n\varepsilon^2$、$\delta \leq \frac{1}{n}$以及 $n \geq \frac{cB}{\varepsilon^2}\log(\frac{1}{\delta})$，我们有 $\sigma_{\mathsf{N}} \leq \frac{c\sqrt{B}}{n\varepsilon}\log(\frac{1}{\delta})$。对 χ^2分布的经典尾界根据[29, 引理 1]给出，以至少 $1-\delta$的概率有

$$\|Z^{\mathsf{N}}\|_2^2 \leq \sigma_{\mathsf{N}}^2\left[d + 2\sqrt{d\log\tfrac{1}{\delta}} + 2\log\tfrac{1}{\delta}\right] \leq \frac{cBd\log^2\frac{1}{\delta}}{n^2\varepsilon^2},$$

其中，最后一个不等式由对 σ_{N} 的界以及 $e^{-d} \leq \delta$ 的假设得出。 □

5.2 适应重尾数据

在实际应用中，我们可能并不具备先验的数据集中性质方面的知识。鉴于对于仅具有 p 阶矩 [5], 的随机变量，其均值的隐私估计量必然具有较慢的收敛速度，因此必须根据问题的实际尺度（以及矩的数量）进行自适应调整。为此，我们提出ADAMEAN，它会自动调整

阈值参数 B，方法是反复调用PRIVMEAN并将 B 加倍，直到 $\widetilde{\mu} \neq \perp$。关键在于，ADAMEAN终止时，有效的 B 至多为随机变量实际尺度的两倍$O(1)$ 最大值$_{i\le n}\|Z_i\|^2$。为确保无论调用PRIVMEAN多少次都能满足隐私要求，ADAMEAN会逐次减少分配给PRIVMEAN的隐私预算；具体而言，当 $\sum_{t=1}^{\infty} 1/t^2 = \pi^2/6$时，通过基本组合，我们可以将ADAMEAN的总隐私损失界定为 $\frac{\pi^2}{6}$ 倍的PRIVMEAN隐私损失。除在 B/d方面多出一个多对数因子外，ADAMEAN的准确性与PRIVMEAN相当，下面将对此作出说明。

算法 5：完全自适应私有均值估计（ADAMEAN）

输入 ：数据 $x_{1:n}$
参数：隐私预算 (ε, δ)
输出：均值估计 $\widetilde{\mu}$

对于 $t = 1, 2, \ldots$ **do**
  $\widetilde{\mu}_t \leftarrow$ PRIVMEAN$_{d2^{t-1},(\varepsilon/t^2,\delta/t^2)}(x)$
  若 $\widetilde{\mu}_t \neq \perp$，则
    返回 $\widetilde{\mu}_t$

定理 4（ADAMEAN 的准确性）。设 $\varepsilon > 0$ 和 $e^{-d} \le \delta \le \frac{1}{n}$ 为隐私参数，并且事件 $\mathcal{E}_{\text{samp}}$ 成立。设 $s = \max\{1, \log_2 \frac{4M^2}{d}\}$，并假设 $n \ge \frac{c \max\{d, M^2\} s^2}{\varepsilon^2} \log^2 \frac{s^2}{\delta}$。设 $\widetilde{\mu} =$ ADAMEAN$_{\varepsilon,\delta}(X_{1:n})$。那么，以至少 $1 - (5 + \pi^2/3)\delta$ 的概率，

$$\left\|\widetilde{\mu} - \overline{X}_n\right\|_\Sigma \le \frac{cs^2 \log(\frac{s^2}{\delta}) \max\{M\sqrt{d}, M \log(\frac{s^2}{\delta}), d\}}{n\varepsilon}. \tag{6}$$

证明。设 $t^\star$ 为满足以下条件的最小正整数 $M^2 \le d2^{t^\star - 1}$，并设 t_{stop} 为 ADAMEAN终止时的迭代次数（可能为无穷大）。注意 $t^\star \le s$。证明分为两部分：在事件 $\mathcal{E}_{\text{samp}}$发生时，我们首先证明 $t_{\text{stop}} > t^\star$ 或 $\widetilde{\mu}$ 以至少 6 的概率满足结论（$1 - \pi^2\delta/6$；其次，我们证明 ADAMEAN 以至少 $t_{\text{stop}} \le t^\star$ 的概率终止 $1 - 5\delta$。由并集界即可得到结论。

我们借助以下引理完成第一部分。

引理 5.1。设 $\varepsilon > 0$ 和 $e^{-d} \le \delta \le \frac{1}{n}$ 为隐私参数，并设 $B \ge 0$。假设事件 $\mathcal{E}_{\text{samp}}$成立，且令 $\widetilde{\mu} =$ PRIVMEAN$_{B,\ (\varepsilon,\delta)}(X_{1:n})$。那么，以至少 $1 - 2\delta$的概率（关于 PRIVMEAN的随机性），$\widetilde{\mu} = \perp$ 或

$$\left\|\widetilde{\mu} - \overline{X}_n\right\|_\Sigma \le \frac{c \log \frac{1}{\delta} \max\{M \log \frac{1}{\delta}, \sqrt{Bd}\}}{n\varepsilon}.$$

证明。假设 $\widetilde{\mu} \neq \perp$ 否则该结论显然成立。令 Σ 表示 COVSAFE 的协方差估计（即 Σ_T 在 COVSAFE 的最后一次迭代中），并令 $\widehat{\mu}$ 表示未被 MEANSAFE 剪除的观测值的经验均值，因此 $\widetilde{\mu} = \widehat{\mu} + \widehat{\Sigma}^{1/2} Z^{\mathsf{N}}$。根据 $\perp$ MEANSAFE 第 11 行中返回 $\perp$ 的条件，MEANSAFE至多剪除 $b(\frac{2k}{3} + W_{\text{mean}})$ 个点，因此

$$\begin{aligned}\left\|\widehat{\mu} - \overline{X}_n\right\|_\Sigma &\le \frac{b(\frac{2k}{3} + W_{\text{mean}}) \max_i \|X_i - \widehat{\mu}\|_\Sigma}{n} \\ &\le \frac{b(\frac{2k}{3} + W_{\text{mean}}) (\max_i \|X_i - \mu\|_\Sigma + \|\mu - \widehat{\mu}\|_\Sigma)}{n}\end{aligned}$$

$$\leq \frac{2b(\frac{2k}{3}+W_{\text{mean}})\max_i\|X_i-\mu\|_\Sigma}{n} \overset{(\star)}{\leq} \frac{2b(\frac{2k}{3}+W_{\text{mean}})M}{\sqrt{c_1}n},$$

其中 $(\star)$ 直接由 $\mathcal{E}_{\text{samp}}$ 得出。回顾 $k=\frac{24}{\varepsilon}\log\frac{3}{\delta}-3$ 和 $W\sim\text{Lap}(\frac{8}{\varepsilon})$，可知以至少 $1-\frac{\delta}{6}$ 的概率，$\frac{2k}{3}+W<\frac{24}{\varepsilon}\log\frac{3}{\delta}$。再回顾 $b=1+\log_2\frac{6n^2}{\delta}$，可知在 $\mathcal{E}_{\text{samp}}$ 上，

$$\left\|\widehat{\mu}-\overline{X}_n\right\|_\Sigma \leq \frac{2(1+\log_2\frac{6n^2}{\delta})(\frac{24}{\varepsilon}\log\frac{3}{\delta})M}{\sqrt{c_1}n} \leq \frac{cM\log^2\frac{1}{\delta}}{n\varepsilon}$$

以至少 $1-\frac{\delta}{6}$ 的概率。

同时，注意到 $\mathcal{E}_{\text{samp}}$ 意味着 $\Sigma\preceq\overline{\Sigma}_n\preceq 2\Sigma$ 因为修剪条目（第 7 行）不会 COVSAFE，只会使其协方差估计缩小。因此，$\|\widehat{\Sigma}^{1/2}Z^{\mathsf{N}}\|_\Sigma\leq\sqrt{2}\|\widehat{\Sigma}^{1/2}Z^{\mathsf{N}}\|_{\widehat{\Sigma}}=\sqrt{2}\|Z^{\mathsf{N}}\|_2$。根据与定理 3 中相同的论证，

$$\left\|Z^{\mathsf{N}}\right\|_2 \leq \frac{c\sqrt{Bd}\log(\frac{1}{\delta})}{n\varepsilon}$$

以至少 $1-\delta$ 的概率。

将前面的两个显示式结合起来，并取并集界，可得引理 5.1 。 □

应用引理 5.1 ，取映射 $B=d2^{t-1}$、$\varepsilon\mapsto\varepsilon/t^2$ 和 $\delta\mapsto\delta/t^2$，则对于任意 $1\leq t\leq t^\star$，在事件 $\mathcal{E}_{\text{samp}}$发生的条件下，以至少 $1-2\delta/t^2$ 的概率，要么 $\widetilde{\mu}_t=\perp$，要么

$$\left\|\widetilde{\mu}_t-\overline{X}_n\right\|_\Sigma \leq \frac{ct^2\log(t^2/\delta)\max\{M\log(t^2/\delta),d2^{(t-1)/2}\}}{n\varepsilon},$$

其中，后一种情况 $\widetilde{\mu}_t$ 满足公式（6），因为 $t\leq t^\star\leq s$。随后通过并集界，该事件对所有 $1\leq t\leq t^\star$ 同时成立的概率至少为 $1-\pi^2\delta/3$，因此要么 $t_{\text{stop}}>t^\star$，要么 ADAMEAN 终止，并且 $\widetilde{\mu}$ 满足结论（6）。

接下来证明的第二部分，回顾 $d2^{t^\star-1}\geq M^2$ 因此，应用定理 3 并令 $B=d2^{t^\star-1}$，可知在 $\mathcal{E}_{\text{samp}}$ 的条件下，$\widetilde{\mu}_{t^\star}\neq\perp$，因此 ADAMEAN在 $t_{\text{stop}}\leq t^\star$ 次迭代后终止，且其概率至少为 $1-5\delta/(t^\star)^2\geq 1-5\delta$。结论（6）得证。 □

示例 3（示例 1续）：在这种情况下，$\Sigma^{-1/2}X_i$是 τ^2-次高斯，因此假设 A1 中有 $M^2\lesssim\tau^2(d+\underline{\text{对数}}\ \frac{n}{\beta})$。于是，样本均值以至少 $1-\delta$ 的概率集中于 $\left\|\overline{X}_n-\mu\right\|_\Sigma\lesssim\tau\sqrt{(d+\text{对数}(1/\delta))/n}$，并且假设 $\delta\geq e^{-d}$，定理 4 随即推出：对 $\widetilde{\mu}=\text{ADAMEAN}_{\varepsilon,\,\delta}(X_{1:n})$，以至少 $1-O(\delta)$ 的概率有（忽略 n中的多对数因子）

$$\|\widetilde{\mu}-\mu\|_\Sigma=\widetilde{O}\left(\tau\sqrt{\frac{d}{n}}+\frac{\tau d\log\frac{1}{\delta}}{n\varepsilon}\right).$$

对于次高斯情形，该速率在相差一个 $\log\frac{1}{\delta}$ 因子和关于 n 的多对数因子的意义下是极小极大的（见 [35] 或 [26, 引理 6.7] ，其中给出了已知协方差矩阵时高斯均值估计的下界）。 ♢

例4（例2，续）：回顾一下，此处 $\mathbb{E}[\|X_i-\mu\|_\Sigma^p]\leq\tau^p d^{p/2}$ 对于 $p\geq 4$ 和$\tau\geq 1$。根据定理4，以至少 $1-5\delta$ 的概率，对 $\widetilde{\mu}=\text{ADAMEAN}_{\varepsilon,\,\delta}(X_{1:n})$，我们有

$$\|\widetilde{\mu}-\mu\|_\Sigma\leq\left\|\overline{X}_n-\mu\right\|_\Sigma+\widetilde{O}\left(\frac{\max_i\|X_i-\mu\|_\Sigma\sqrt{d}\log\frac{1}{\delta}}{n\varepsilon}\right)$$

只要经验协方差满足 $\frac{1}{2}\Sigma \preceq \overline{\Sigma}_n \preceq \frac{3}{2}\Sigma$。由于这一条件以常数概率成立，且 $\|\overline{X}_n - \mu\|_\Sigma \lesssim \sqrt{d/n}$ 也以常数概率成立，我们代入例2 中的 $\max_i \|X_i - \mu\|_\Sigma$ 界，得到该结论以（任意）常数概率成立：

$$\|\widetilde{\mu} - \mu\|_\Sigma = \widetilde{O}\left(\sqrt{\frac{d}{n}} + \frac{\tau d \log\frac{1}{\delta}}{n^{1-1/p}\varepsilon}\right).$$

结合 Barber 和 Duchi [5, 命题 4] 以及 Steinke 和 Ullman [35], 的极小极大下界，目前已知的最佳极小极大下界是：以常数概率,

$$\|\widetilde{\mu} - \mu\|_\Sigma \gtrsim \sqrt{\frac{d}{n}} + \tau\frac{d^{\frac{2p-1}{2p}}\log^{\frac{p-1}{2p}}\frac{1}{\delta}}{(n\varepsilon)^{1-1/p}}.$$

因此，自适应方法在 n方面达到了最优尺度，但在 ε 方面可能较松，且维度依赖性相差 $d^{1/2p}$ 倍。 ◊

6 稳定协方差估计的证明

在本节中，我们给出引理的证明4.1 和 4.2，不过我们先介绍一系列预备结果，以便实际证明这两个主要引理。在证明中，我们将从 4 行开始的 COVSAFE的 while 循环的每次执行称为一次 迭代，并将 COVSAFE的执行记录 Γ 用作追踪COVSAFE所有迭代中完整执行过程的便捷方式。

6.1 COVSAFE的性质

我们首先形式化描述 COVSAFE 执行过程的确定性性质，并给出使 COVSAFE 的输出相当稳定的条件。接下来，我们将利用这些性质给出一些集合，使噪声变量 Z 和 W 以高概率属于这些集合，从而保证稳定性。回顾记号 (4)： $\widehat{\Sigma}(x, z, w)$ 表示 COVSAFE 在输入样本 x 和噪声 $z \in \mathbb{R}^{n/2+1}, w \in \mathbb{R}$下的输出，其执行记录 $\Gamma = ([R_t]_{t\le T}, [\Sigma_t]_{t\le T}, T)$ 隐式依赖于 (x, z, w)；而 $\widehat{\Sigma}_{-i}(x, z, w)$ 表示相应的留一协方差。简记 $\widehat{\Sigma} = \widehat{\Sigma}(x, z, w)$ 和 $\widehat{\Sigma}_{-i} = \widehat{\Sigma}_{-i}(x, z, w)$，并取 $\tilde{x} = x_{1:n/2} - x_{n/2+1:n}$为 COVSAFE 第 1 行中的值。

引理 6.1 给出了在 $\tilde{x}_i$ 次迭代中修剪 $t+1$ 的COVSAFE的充要条件，即 $i \in R_{t+1}$；而引理 6.2 给出了曾经修剪 $\tilde{x}_i$ 的类似条件（也就是说，是否 $i \in R_T$）。

引理 6.1. 索引 $i \in R_{t+1}$ 当且仅当 $\log\left(\|\tilde{x}_i\|^2_{\Sigma_t}\right) + z_i + z_{n/2+1} > \log(B)$.

证明。 “若”方向可直接由向 R_{t+1} 中添加元素的条件得出（参见 COVSAFE的第 COVSAFE 行）。对于另一方向，若 $i \in R_{t+1}$，则（同样根据该条件）必有某个 $s \le t$ 满足

$$\log\left(\|\tilde{x}_i\|^2_{\Sigma_s}\right) + z_i + z_{n/2+1} > \log(B).$$

由于 $s \le t$，我们有 $R_s \subset R_t$，因此 $\Sigma_s \succeq \Sigma_t$，这又意味着对数 $\left(\|\tilde{x}_i\|^2_{\Sigma_t}\right) + z_i + z_{n/2+1} >$ 对数(B)。 □

引理 6.2. 索引 $i \notin R_T$ 当且仅当对数$(\|\tilde{x}_i\|^2_{\Sigma_T}) + z_i + z_{n/2+1} \le$ 对数(B)。

证明。 注意到 $\Sigma_{T-1} = \Sigma_T$，因为 COVSAFE 的内部 while 循环仅在算法于上一次迭代中未移除任何观测值时才终止（参见 COVSAFE 的第 10 行）。然后，通过将引理 6.1 应用于 $t = T-1$，即可得到该结论。 □

最后，根据移除的索引 $R_{T,-i}$ 和阈值 $m+w$，我们可以完全刻画 Σ_{-i}，如下方引理所述。

引理 6.3。 $\widehat{\Sigma}_{-i} = \frac{1}{n}\sum_{j \notin R_{T,-i} \cup \{i\}} \tilde{x}_j \tilde{x}_j^T$ 当且仅当 $|R_T| \le m+w$。

证明。 该结论可直接由 COVSAFE 的第 14 行中的返回条件得到，因为$\widehat{\Sigma} \neq \perp$意味着 $\widehat{\Sigma}_{-i} = \widehat{\Sigma} - \frac{1}{n} 1\{i \notin R_T\} \tilde{x}_i \tilde{x}_i^T$，其中 $\widehat{\Sigma} = \Sigma_T = \frac{1}{n}\sum_{j \notin R_T} \tilde{x}_j \tilde{x}_j^T$按定义成立。 □

6.2 引理4.1的证明

在整个证明中，我们简记 $\Sigma = \Sigma(x, Z, w)$ 和 $\Sigma_{-i} = \Sigma_{-i}(x, Z, w)$。假设$\widehat{\Sigma} \neq \perp$，否则结论显然成立，并回顾 $Z_j \overset{\text{独立同分布}}{\sim}$ 拉普拉斯分布(σ)。注意，如果 $i \in R_T$，则有 $\Sigma = \Sigma_{-i}$，根据定义可得；因此，我们只需考虑 $i \notin R_T$。接下来，由引理 6.2 可得

$$\log\left(\|\tilde{x}_i\|^2_{\widehat{\Sigma}}\right) + Z_i + Z_{n/2+1} \le \log(B)$$

对所有 $i \notin R_T$ 均成立，由此可知，只要 $Z_i + Z_{n/2+1} \ge -1/2$，就有 $\|\tilde{x}_i\|^2_{\widehat{\Sigma}} \le B\sqrt{e}$。现在我们利用 $\|\tilde{x}_i\|^2_{\widehat{\Sigma}} \le B\sqrt{e}$ 蕴含 $d_{\text{psd}}(\widehat{\Sigma}, \widehat{\Sigma}_{-i})$ 很小这一事实，这可由以下线性代数引理推出。

引理 6.4。设 $A \in \mathbb{R}^{d\times d}$ 为半正定，且 $a \in \mathbb{R}^d$ 满足$\|a\|^2_A < 1$。则

$$d_{\text{psd}}(A, A - aa^T) \le \frac{1}{1 - \|a\|^2_A} \|a\|^2_A .$$

证明。为简洁起见，定义 $C = A - aa^T$。我们首先证明 Col(C) = Col(A)。由于$\|a\|^2_A$ 是有限的，因此 $a \in$ Col(A)，于是 Col$(C) \subset$ Col(A)。另一方面，展开 C 可得

$$C = A^{1/2}(I - A^{\dagger/2} a a^T A^{\dagger/2}) A^{1/2} \succeq (1 - \|a\|^2_A) A, \tag{7}$$

由此可得 Col(C) =Col(A)。

由（7）还可得 $C^\dagger \preceq \frac{1}{1-\|a\|^2_A} A^\dagger$，因此

$$\left\|C^{\dagger/2}(A - C)C^{\dagger/2}\right\|_* = \left\|C^{\dagger/2} a a^T C^{\dagger/2}\right\|_* = \|a\|^2_C \le \frac{\|a\|^2_A}{1 - \|a\|^2_A}.$$

类似的计算可得 $\|A^{\dagger/2}(C - A)A^{\dagger/2}\|_* = \|a\|^2_A$，从而证明了该结论。 □

引理 6.4 直接表明 $\widehat{\Sigma}_{-i} = \widehat{\Sigma} - \frac{1}{n}\tilde{x}_i \tilde{x}_i^T 1\{i \in R_T\}$ 满足

$$d_{\text{psd}}(\widehat{\Sigma}, \widehat{\Sigma}_{-i}) \le \frac{1}{1 - B\sqrt{e}/n} \frac{B\sqrt{e}}{n}$$

只要 $Z_i + Z_{n/2+1} \ge -\frac{1}{2}$。为了证明这种情况以高概率发生，我们使用以下结果。该结果可由如下观察推出：若 $c \ge 0$，则对任意独立变量 X, Y，由并集界可得 $\mathbb{P}(X + Y > c) \le \mathbb{P}(X > c/2) + \mathbb{P}(Y > c/2)$：

观察 6.1. 设 $X, Y \overset{\text{独立同分布}}{\sim}$ 拉普拉斯分布(σ) 且 $c \ge 0$。则 $\mathbb{P}(X + Y > c) \le \exp(-\frac{c}{2\sigma})$。

我们看到，$\mathbb{P}(Z_i + Z_{n/2+1} < -\frac{1}{2}) \le \exp(-\frac{1}{4\sigma})$，正如所声称的那样。

6.3 引理的证明 4.2

引理 4.2 的证明分为四个步骤。证明的关键在于一个耦合论证：通过当前假设 $\tilde{x}_{-i} = \tilde{x}'_{-i}$，我们以谨慎的方式扰动 Z，使 COVSAFE 在 x 上的执行与其在 x' 上的执行相等，同时仅稍微改变 Z 的分布。我们在引理 6.5 中给出的第一步，是一个确定性引理，它通过噪声输入值 z，关联 COVSAFE 在相邻输入 x 和 x' 上移除的索引集合 R 和 R'。第二步和第三步分别由引理 6.6 和 6.7 构成，我们构造映射 $\pi : \mathbb{R}^{n/2+1} \to \mathbb{R}^{n/2+1}$，使得 Z 和 $\pi(Z)$具有相似的分布 $\widehat{\quad}$ $\widehat{\quad}$；此外，对于所有 z，$\Sigma_{-i}(x, z, W)$ 和 $\Sigma_{-i}(x', \pi(z)\ W)$（回顾定义 (4b)）也具有相似的分布，其中后一个分布近似使用了 W 中的随机性。引理 6.6 关联移除索引 R_T、$_{-i}$的分布，而引理 6.7 关联 COVSAFE 中止并返回 $\perp$的概率。在第 6.3.1 节中，我们最后综合这些中间引理，给出引理 4.2 的证明。

我们的第一步是建立一个确定性引理，用以关联各组移除的索引。

引理 6.5. 设 z、$z' \in \mathbb{R}^{n/2+1}$ 和 $w \in \mathbb{R}$，并设

$$\widehat{\Sigma}, \left([R_t]_{t=0}^{T}, [\Sigma_t]_{t=0}^{T}, T\right) := \texttt{COVSAFE}_{B,m}(x; z, w),$$
$$\widehat{\Sigma}', \left([R'_t]_{t=0}^{T'}, [\Sigma'_t]_{t=0}^{T'}, T'\right) := \texttt{COVSAFE}_{B,m}(x'; z', w).$$

假设 $\tilde{x}'_i = 0$。则有以下结论。

(a) 若对所有 $j \in R_{T,-i}$，均有 $z'_j + z'_{n/2+1} \geq z_j + z_{n/2+1}$，则 $R_{T,-i} \subset R'_{T',-i}$。
(b) 若对所有 $j \notin R'_{T',-i}$，均有 $z'_j + z'_{n/2+1} \geq z_j + z_{n/2+1}$，则 $R_{T,-i} \subset R'_{T',-i}$。

附加 另外假设 $n \geq 2B\sqrt{e}$ 和 $z_i + z_{n/2+1} \geq -1/2$。那么以下命题成立 *ld.*

(c) If $z'_j + z'_{n/2+1} \leq z_j + z_{n/2+1} - 2B\sqrt{e}/n$ *for all* $j \in R'_{T',-i}$, *then* $R'_{T',-i} \subset R_{T,-i}$.

(d) If $z'_j + z'_{n/2+1} \leq z_j + z_{n/2+1} - 2B\sqrt{e}/n$ *for all* $j \notin R_{T,-i}$, *then* $R'_{T',-i} \subset R_{T,-i}$.

证明。我们通过对 $t \in \mathbb{N}$进行归纳法证明每个断言。

先看第一个断言 (a)，显然 $R_{0,\ -i} = \emptyset \subset R'_{T',\ -i}$。现在，为进行归纳，假设 $R_{t,\ -i} \subset R'_{T',\ -i}$。若 $t = T$，则无需证明，因此我们取 $t < T$。假设 $\tilde{x}_{-i} = \tilde{x}'_{-i}$和 $\tilde{x}'_i = 0$意味着 $\Sigma_t \succeq \Sigma'_{T'}$。因此，对于所有$j \in R_{t,\ -i} \subset R_{T,\ -i}$，我们有

$$\log(B) \overset{(i)}{<} \log\left(\|\tilde{x}_j\|^2_{\Sigma_t}\right) + z_j + z_{n/2+1} \overset{(ii)}{\leq} \log\left(\|\tilde{x}_j\|^2_{\Sigma'_{T'}}\right) + z_j + z_{n/2+1}$$
$$\overset{(iii)}{\leq} \log\left(\|\tilde{x}_j\|^2_{\Sigma'_{T'}}\right) + z'_j + z'_{n/2+1},$$

其中，不等式（i）由引理 6.1 （取噪声 z）得到，不等式（ii）是因为$\Sigma_t \succeq \Sigma'_{T'}$，而不等式（$iii$）则是情形 (a)中的假设。引理 6.2（取数据 x'和噪声 z'）进而给出：$j \in R'_{T',\ -i}$当且仅当 $\log B < \log(\|\tilde{x}_j\|^2_{\Sigma'_{T'}}) + z'_j + z'_{n/2+1}$。由于$j \in R_{t,\ -i}$是任意的，因此我们有 $R_{t+1,\ -i} \subset R'_{T',\ -i}$。归纳步骤至此完成，因此 $R_{t,\ -i} \subset R'_{T'}$

$_{,-i}$ 对所有 $t \leq T$ 均成立，因此第一个结论得证。

命题(b) 的证明依赖于与第一个命题(a)类似的归纳论证。包含关系 $R_{T,-i} \subset R'_{T',-i}$ 等价于：若 $j \notin R'_{T',-i}$，则 $j \notin R_{T,-i}$。考虑

$j \notin R'_{T',-i}$，并从归纳假设 $R_{t,-i} \subset R'_{T',-i}$开始；只需证明 $j \notin R_{t+1,-i}$。由于 $\Sigma_t \succeq \Sigma'_{T'}$是根据 Σ_t的构造得到的，我们有

$$\begin{aligned}\log\left(\|\tilde{x}_j\|^2_{\Sigma_t}\right) + z_j + z_{n/2+1} &\le \log\left(\|\tilde{x}_j\|^2_{\Sigma'_{T'}}\right) + z_j + z_{n/2+1} \\ &\overset{(i)}{\le} \log\left(\|\tilde{x}_j\|^2_{\Sigma'_{T'}}\right) + z'_j + z'_{n/2+1} \le \log(B),\end{aligned}$$

其中，步骤（i）依据假设 $z_j + z_{n/2+1} \le z'_j + z'_{n/2+1}$，见第 (b) 部分，最后一个不等式则由引理 6.2得出。应用引理 6.1 以及不等式 $\log\left(\|\tilde{x}_j\|^2_{\Sigma_t}\right) + z_j + z_{n/2+1} \le \log B$，即可保证 $j \notin R_{t+1,\,-i}$ 如所需，从而完成了对断言 (b)的证明。

为证明断言 (c)，我们对 $R'_{t\text{进行归纳},\,-i}$ 当 $t \le T'$ 时，并且必须考虑 $\Sigma_T \not\succeq \Sigma'_t$的可能性，即使 $R'_{t,\,-i} \subset R_{T,\,-i}$，因为 Σ_T可能包含项 $\tilde{x}_i\tilde{x}_i^T$（即 $i \notin R_T$）。对于 $t = 0$，基例显然成立，因此假设 $R'_{t,\,-i} \subset R_{T,\,-i}$。若 $i \notin R_T$，则引理 6.2和假设 $z_i + z_{n/2+1} \ge -\frac{1}{2}$可保证

$$\log\left(\|\tilde{x}_j\|^2_{\Sigma_T}\right) \le \log B - z_i - z_{n/2+1} \le \log B + \frac{1}{2} = \log(B\sqrt{e}),$$

即 $\|\tilde{x}_j\|^2_{\Sigma_T} \le B\sqrt{e}$。我们需要以下关于正定矩阵的技术性观察，其证明暂且推迟。

观察 6.2。设 $\Sigma \in \mathbb{R}^{d\times d}$ 为半正定，$\alpha \ge 0$，且 $u \in \mathbb{R}^d$。定义$\Sigma' := \Sigma - \alpha uu^T$。若 $\|u\|^2_\Sigma \le \frac{1}{2\alpha}$，则 $\Sigma' \succeq \frac{1}{2}\Sigma$，并且对于任意 $v \in$列空间(Σ)，

$$\left|\log\left(\|v\|^2_\Sigma\right) - \log\left(\|v\|^2_{\Sigma'}\right)\right| \le 2\alpha\|u\|^2_\Sigma.$$

由于 $\|\tilde{x}_i\|^2_{\Sigma_T} \le B\sqrt{e}$，观察 6.2 适用于 $u = \tilde{x}_i$ 和 $\alpha = \frac{1}{n}$ 当 $n \ge 2B\sqrt{e}$时，因此

$$\log\left(\|v\|^2_{\Sigma_T - \frac{1}{n}1\{i\notin R_T\}\tilde{x}_i\tilde{x}_i^T}\right) \le \log\left(\|v\|^2_{\Sigma_T}\right) + \frac{2B\sqrt{e}}{n} \tag{8}$$

对于所有 v都成立，特别是，对于每个 $j \in [n/2]$，$v = \tilde{x}_j$ 也成立。另一方面，无论 $i \in R_T$ 是否成立，归纳假设 $R'_{t,\,-i} \subset R_{T,\,-i}$ 都保证

$$\Sigma'_t \succeq \Sigma_T - \frac{1}{n}1\{i \notin R_T\}\,\tilde{x}_i\tilde{x}_i^T. \tag{9}$$

考虑$j \in R'_{t+1,\,-i}$，那么，引理6.1 推出

$$\begin{aligned}\log B &< \log(\|\tilde{x}_j\|^2_{\Sigma'_t}) + z'_j + z'_{n/2+1} \\ &\overset{(i)}{\le} \log\left(\|\tilde{x}_j\|^2_{\Sigma_T - \frac{1}{n}1\{i\notin R_T\}\tilde{x}_i\tilde{x}_i^T}\right) + z'_j + z'_{n/2+1} \\ &\overset{(ii)}{\le} \log\left(\|\tilde{x}_j\|^2_{\Sigma_T}\right) + \frac{2B\sqrt{e}}{n} + z'_j + z'_{n/2+1} \\ &\overset{(iii)}{\le} \log\left(\|\tilde{x}_j\|^2_{\Sigma_T}\right) + z_j + z_{n/2+1}.\end{aligned}$$

这里，不等式（i）由序关系（9）得出；不等式（ii）成立，因为如果$i \in R_T$，那么 $\Sigma_T = \Sigma_T - \frac{1}{n}1\{i \notin R_T\}\,\tilde{x}_i\tilde{x}_i^T$，而如果 $i \notin R_T$，则不等式（8）成立；最后一个不等式（iii）根据假设和断言(c)得出。这就得到了归纳结论：$R'_{t+1,\,-i} \subset R_{T,\,-i}$，正如引理6.1 所示，$j \in R_{T,\,-i}$。

断言 (d)可通过与断言 在作必要修改后的证明基本相同的归纳论证得到，断言 (c)亦然。

观察 6.2的证明。 最后，我们回过头来证明所述观察结果。注意到$\Sigma' \succeq \frac{1}{2}\Sigma$ 可知 $u \in$ 列空间（Σ），因此

$$\Sigma - \alpha uu^T = \Sigma^{1/2} \underbrace{(I - \alpha \Sigma^{\dagger/2} uu^T \Sigma^{\dagger/2})}_{\succeq (1/2)I} \Sigma^{1/2} \succeq \frac{1}{2}\Sigma.$$

这也意味着列空间(Σ') = 列空间(Σ)。

要证明引理的其余部分，只需证明对于 $v \in$ 列空间(Σ)，有对数$(\|v\|_{\Sigma'}^2)$ $\leq$ 对数$(\|v\|_{\Sigma}^2)$ $+ 2\alpha \|u\|_{\Sigma}^2$，因为由 $\Sigma \succeq \Sigma'$ 可立即得到另一方向。注意到

$$(\Sigma - \alpha uu^T)^\dagger = \Sigma^{\dagger/2}(I - \alpha \Sigma^{\dagger/2} uu^T \Sigma^{\dagger/2})^{-1} \Sigma^{\dagger/2}.$$

由不等式 $I - \alpha \Sigma^{\dagger/2} uu^T \Sigma^{\dagger/2} \succeq (1 - \alpha \|u\|_{\Sigma}^2) I$，可得

$$(I - \alpha \Sigma^{\dagger/2} uu^T \Sigma^{\dagger/2})^{-1} \preceq (1 - \alpha \|u\|_{\Sigma}^2)^{-1} I \preceq (1 + 2\alpha \|u\|_{\Sigma}^2) I,$$

其中，最后一个不等式由假设 $\|u\|_{\Sigma}^2 \leq \frac{1}{2\alpha}$ 得出。将其与前面的公式结合可知

$$\Sigma'^{\dagger} \preceq (1 + 2\alpha \|u\|_{\Sigma}^2) \Sigma^{\dagger}$$

因此

$$\log\left(\|v\|_{\Sigma'}^2\right) \leq \log\left((1 + 2\alpha \|u\|_{\Sigma}^2) \|v\|_{\Sigma}^2\right) \leq \log\left(\|v\|_{\Sigma}^2\right) + 2\alpha \|u\|_{\Sigma}^2$$

如所需。 □

接下来我们讨论本节开头概述的第二步，它将 COVSAFE 在相邻输入 x 和 x' 上运行时移除索引 R_T 的分布联系起来。关键思路是构造一个确定性映射 π，使得根据引理 6.5, COVSAFE 在输入x 上使用噪声 z 的运行与在 x' 上使用噪声 π（z）的运行相似，并证明 π（Z）和 Z 的分布相似。引理 6.3 表明，离群值索引的集合 $\widehat{R}_T$，$_{-i}$ 完全决定 Σ_{-i}，但在 $\Sigma_{-i} = \perp$ 的情况下除外；当 $|\widehat{R}_T$，$_{-i}|$ 较大时，这种情况以高概率发生，因此下一个引理将控制移除索引集合的分布。为陈述该引理，我们需要几个其概率可加以控制的事件。回顾 Z_j 独立同分布 $\sim$ 拉普拉斯分布(σ_Z)，定义

$$\mathcal{E}_{\text{prune}} := \{z \in \mathbb{R}^{n/2+1} \mid z_j + z_{n/2+1} \geq -1/2 \text{ for all } j \in [n/2]\}. \tag{10}$$

为给证明余下部分的符号设定约定，我们将定义 (4a) 简记为

$$\begin{aligned} \widehat{\Sigma}, ([R_t]_{t=0}^T, [\Sigma_t]_{t=0}^T, T) &:= \texttt{COVSAFE}_{B,m}(x; Z, W) \\ \widehat{\Sigma}', ([R'_t]_{t=0}^{T'}, [\Sigma'_t]_{t=0}^{T'}, T') &:= \texttt{COVSAFE}_{B,m}(x'; Z, W), \end{aligned} \tag{11}$$

以及定义（4b）

$$\widehat{\Sigma}_{-i} = \widehat{\Sigma} - \frac{1}{n} 1\{i \notin R_T\} \tilde{x}_i \tilde{x}_i^T \quad \text{and} \quad \widehat{\Sigma}'_{-i} = \widehat{\Sigma}' - \frac{1}{n} 1\{i \notin R'_T\} \tilde{x}'_i \tilde{x}_i'^T,$$

其中，对任意向量 v，$\perp + v = \perp$。

我们有 对移除的索引给出以下分布保证，不论 W .

引理 6.6。设 $S \subset [n/2] \setminus \{i\}$ 并定义 $\alpha = \frac{2\sqrt{e}B}{(|S|+1)\, n\sigma_Z}$。若 $\tilde{x}'_i = 0$，则

(a) $\mathbb{P}(R_{T,-i} = S, Z \in \mathcal{E}_{\text{prune}}) \leq \exp(\alpha)\ \mathbb{P}(R'_{T',-i} = S)$.
(b) $\mathbb{P}(R'_{T',-i} = S, Z \in \mathcal{E}_{\text{prune}}) \leq \exp(\alpha)\ \mathbb{P}(R_{T,-i} = S)$.

证明。输入噪声 Z 完全决定了 R_T 和 $R'_{T'}$ 在 COVSAFE (参见 而 循环构造 R_t，行 4–12)。因此，我们可以定义产生给定移除索引集合的输入噪声集合 Z，令

$$\mathcal{Z}(S) := \{z \in \mathbb{R}^{n/2+1} \mid R_{T,-i} = S \text{ for } Z = z\}$$
$$\mathcal{Z}'(S) := \{z \in \mathbb{R}^{n/2+1} \mid R'_{T',-i} = S \text{ for } Z = z\},$$

因此 $Z \in \mathcal{Z}(S)$ 等价于 $R_{T,\ -i} = S$。只需证明

$$\begin{aligned} \mathbb{P}\left(Z \in \mathcal{Z}(S) \cap \mathcal{E}_{\text{prune}}\right) &\leq e^{\alpha}\mathbb{P}\left(Z \in \mathcal{Z}'(S)\right) \quad \text{and} \\ \mathbb{P}\left(Z \in \mathcal{Z}'(S) \cap \mathcal{E}_{\text{prune}}\right) &\leq e^{\alpha}\mathbb{P}\left(Z \in \mathcal{Z}(S)\right), \end{aligned} \tag{12}$$

如所述，(a) 由第一个界得出，而断言 (b) 由第二个界得出。

继续讨论第一个界，定义 $\eta \in \mathbb{R}^{n/2+1}$ 和 $\pi\colon\ \mathbb{R}^{n/2+1} \to \mathbb{R}^{n/2+1}$ 为

$$\eta_j := \begin{cases} 2B\sqrt{e}/n & j \in S \\ -2B\sqrt{e}/n & j = n/2+1 \\ 0 & \text{otherwise,} \end{cases} \qquad \pi(z) := z + \eta.$$

确定性移除引理 6.5 表明，对于 $Z \in \mathcal{E}_{\text{prune}}$，若令 $z = Z$ 和 $z' = \pi(z)$，使得 $z_j + z_{n/2+1} \geq z'_j + z'_{n/2+1}$ 对于 $j \in S$，则引理 (a) 和 (d) 的 6.5 给出

$$\pi(\mathcal{Z}(S) \cap \mathcal{E}_{\text{prune}}) \subset \mathcal{Z}'(S).$$

式（12）中的第一个界由引理 2.7 中的标准拉普拉斯比值界直接得到。事实上，我们有 Z_j 独立同分布$\sim$ 拉普拉斯分布(σ_Z) $=$ 拉普拉斯分布$(\|\eta\|_1, \frac{\sigma_Z}{\|\eta\|_1})$，并且 $\|\eta\|_1 = \frac{2\sqrt{e}B}{(|S|+1)\, n}$。接着令$\beta = \|\eta\|_1$，可得 $\beta/\sigma_Z \leq \alpha$，因此可应用引理 2.7，通过以下推导得到所需的界（12）：

$$\mathbb{P}(Z \in \mathcal{Z}(S) \cap \mathcal{E}_{\text{prune}}) \leq e^{\alpha}\mathbb{P}(Z \in \pi(\mathcal{Z}(S) \cap \mathcal{E}_{\text{prune}})) \leq e^{\alpha}\mathbb{P}(Z \in \mathcal{Z}'(S)).$$

第二个界（12）的证明基本相同，只是这次令

$$\eta_j := \begin{cases} 2B\sqrt{e}/n & j \in S \\ 0 & \text{otherwise,} \end{cases} \qquad \pi(z) = z + \eta.$$

于是，事件 $Z \in \mathcal{E}_{\text{prune}}$ 蕴含 $\pi(Z) \in \mathcal{E}_{\text{prune}}$，因为对于所有$j \in [n/2]$，$\pi(Z)_j + \pi(Z)_{n/2+1} \geq Z_j + Z_{n/2+1}$。因此，我们可以在设置 $z = \pi(Z), z' = Z$ 和 $R'_{T',\ -n} = S$ 的情况下应用引理 6.5 的情形(b) 和(c)，并沿用上述相同的论证。 □

我们在引理 4.2 证明开头概述的第三步，是建立以下关系：$\widehat{\ }$ $\widehat{\ }$COVSAFE 在相邻输入 x 和 x'上中止的概率。回想一下，Σ 和 Σ' 分别是 COVSAFE 在输入 COVSAFE 上的输出协方差 x 和 x'，定义（11）中对此有定义。

引理 6.7。令 $\sigma_Z, \sigma_W > 0$，Z_j独立同分布$\sim$拉普拉斯分布(σ_Z)，以及 $W \sim$拉普拉斯分布(σ_W) 在定义 (11) 中。如果 $\tilde{x}_{-i} = \tilde{x}'_{-i}$ 且 $\tilde{x}'_i = 0$，那么

(a) $\mathbb{P}(\widehat{\Sigma}=\perp) \leq \exp(\frac{1}{\sigma_W})\mathbb{P}(\widehat{\Sigma}'=\perp)$.

(b) $\mathbb{P}(\widehat{\Sigma}'=\perp, Z\in\mathcal{E}_{\text{prune}}) \leq \exp(\frac{2\sqrt{e}B}{n\sigma_Z})\mathbb{P}(\widehat{\Sigma}=\perp)$.

证明。令 f表示 W的概率密度，因此 $f(w)=\frac{1}{2\sigma_{W_{\text{cov}}}}\exp(-|w|/\sigma_{W_{\text{cov}}})$，于是| 对数 $\frac{f_{(w)}f(w-1)}{}\Big| \leq \frac{1}{\sigma_{W_{\text{cov}}}}$对所有 w都成立，并回顾 $m\in\mathbb{N}$ COVSAFE 中第 14 行的 COVSAFE阈值。接下来证明引理的第一个结论，我们有以下一系列不等式：

$$\begin{aligned}
\mathbb{P}\left(\widehat{\Sigma}=\perp\right) &\overset{(i)}{\leq} \int \mathbb{P}\left(|R_T|>m+w\right) f(w)dw \leq \int \mathbb{P}\left(|R_{T,-i}|>m+w-1\right) f(w)dw \\
&\overset{(ii)}{\leq} \int \mathbb{P}\left(\left|R'_{T',-i}\right|>m+w-1\right)\ f(w)dw \\
&\leq \exp(1/\sigma_{W_{\text{cov}}}) \int \mathbb{P}\left(\left|R'_{T',-i}\right|>m+w-1\right) f(w-1)dw \\
&\overset{(iii)}{=} \exp(1/\sigma_{W_{\text{cov}}})\mathbb{P}(\widehat{\Sigma}'=\perp).
\end{aligned}$$

这里，步骤（i）由引理 6.3 得出，即 $\Sigma'=\perp$ 当且仅当 $|R'_T|>m+w$。步骤（ii）依据引理 6.5(a) 中的耦合论证：由于在 COVSAFE(x; Z, W) 和 COVSAFE(x'; Z, W) 的两次执行中，噪声 Z 相同，因此我们有 $R_{T,-i}\subset R'_{T',-i}$。步骤（$iii$）再次应用引理 6.3，因为假设 $\widetilde{x}'_i=0$ 保证了 $R'_{T',-i}=R'_{T'}$（回想一下第 7 行中的拒绝阈值）。命题 (a) 得证。

对于命题 (b)，再次应用引理 6.3 可得

$$\mathbb{P}\left(\widehat{\Sigma}'=\perp, Z\in\mathcal{E}_{\text{prune}}\right) = \int \mathbb{P}\left(\left|R'_{T',-i}\right|>m+w, Z\in\mathcal{E}_{\text{prune}}\right) f(w)dw$$

and

$$\int \mathbb{P}\left(|R_{T,-i}|>m+w\right) f(w)dw \leq \mathbb{P}\left(\widehat{\Sigma}=\perp\right).$$

结合这些式子，只需证明对所有 w，都有

$$\mathbb{P}\left(\left|R'_{T',-i}\right|>m+w, Z\in\mathcal{E}_{\text{prune}}\right) \leq \exp\left(\frac{2\sqrt{e}B}{n\sigma_Z}\right)\mathbb{P}\left(|R_{T,-i}|>m+w\right). \tag{13}$$

为此，我们采用与引理 6.6 的证明中类似的方法，定义

$$\begin{aligned}
\mathcal{Z}(w) &:= \{z\in\mathbb{R}^{n/2+1} \mid |R_{T,-i}|>m+w \text{ for } Z=z\} \\
\mathcal{Z}'(w) &:= \{z\in\mathbb{R}^{n/2+1} \mid \left|R'_{T',-i}\right|>m+w \text{ for } Z=z\}
\end{aligned}$$

以及单坐标扰动 $\pi(z):=z+\eta$ 对于除 $\eta_{n/2+1}=\frac{2\sqrt{e}B}{n}$ 外全为零的向量。与引理 6.6 的证明类似，映射 π 保证 $z'=\pi(z)$ 对所有 j都满足 $z'_j+z'_{n/2+1}\leq z_j+z_{n/2+1}-\frac{2B\sqrt{e}}{n}$，这正是引理 (c)6.5中情形的条件，因此无论 $R'_{T',-i}$ 和 $R_{T,-i}$ 如何，$R'_{T',-i}\subset R_{T,-i}$。因此，对所有 $w\in\mathbb{R}$，以下结论同时成立：

$$\pi\left(\mathcal{Z}'(w)\cap\mathcal{E}_{\text{prune}}\right)\subset\mathcal{Z}(w).$$

注意到 $\|\eta\|_1=\frac{2\sqrt{e}B}{n}$，引理 2.7 关于拉普拉斯随机变量似然比的结果保证了

$$\mathbb{P}(Z\in\mathcal{Z}'(w)\cap\mathcal{E}_{\text{prune}}) \leq \exp\left(\frac{2\sqrt{e}B}{n\sigma_Z}\right)\mathbb{P}(Z\in\pi(\mathcal{Z}'(w)\cap\mathcal{E}_{\text{prune}})) \leq \exp\left(\frac{2\sqrt{e}B}{n\sigma_Z}\right)\mathbb{P}(Z\in\mathcal{Z}(w)),$$

which is equivalent to inequality (13). □

6.3.1 引理 4.2证明的最后步骤

结合引理 6.6 和 6.7，我们可以证明 COVSAFE的稳定性。回顾集合$\mathcal{E}_{\text{prune}}$，见式（10），以及 $W \sim$ 拉普拉斯分布$(\sigma_{W_{\text{cov}}})$，并另外定义

$$\mathcal{E}_{\text{thr}} := \left(-\infty, \sigma_{W_{\text{cov}}} \log \frac{1}{\gamma}\right].$$

我们论证的关键在于证明：当 x 和 x' 相邻，但 $\tilde{x}'_i = 0$时，如果噪声变量 Z, W 满足 $Z \in \mathcal{E}_{\text{prune}}$ 且 $W \in \mathcal{E}_{\text{thr}}$时，那么对于 Σ 和 Σ'，如调用（11）中所示，留一协方差 Σ_{-i} 和 Σ'_{-i} 相似。随后，我们界定各个事件发生的概率，并利用群组合论证，得到适用于任意 $\tilde{x}'_i$的引理。

有鉴于此，令 $A \in \mathbb{R}^{d\times d}$，并注意，对于任意固定的样本 x，$\widehat{\Sigma}$ 和 $\widehat{\Sigma}_{-i}$ 只能取有限多个值。对于

$$\alpha = \frac{1}{\sigma_{W_{\text{cov}}}} + \frac{2\sqrt{e}B(m_{\max}+1)}{n\sigma_Z},$$

我们证明，对于任意 $A \in \mathbb{R}^{d\times d} \cup \{\perp\}$，都有

$$\mathbb{P}\left(\widehat{\Sigma}_{-i} = A, Z \in \mathcal{E}_{\text{prune}}, W \in \mathcal{E}_{\text{thr}}\right) \le \exp(\alpha)\mathbb{P}\left(\widehat{\Sigma}'_{-i} = A\right) \quad \text{and} \tag{14}$$

$$\mathbb{P}\left(\widehat{\Sigma}'_{-i} = A, Z \in \mathcal{E}_{\text{prune}}, W \in \mathcal{E}_{\text{thr}}\right) \le \exp(\alpha)\mathbb{P}\left(\widehat{\Sigma}_{-i} = A\right). \tag{15}$$

引理6.7 已经推出不等式（14）和（15）对 $A = \perp$均成立，因此剩下的只需证明它们对 $A \in \mathbb{R}^{d\times d}$也成立。

首先处理不等式（14），令 $f(w) = \frac{1}{2\sigma_{W_{\text{cov}}}}\exp(-|w|/\sigma_{W_{\text{cov}}})$ 为W和 $S(A)$ 的概率密度：$= \{S \subset [n/2] \setminus \{i\} \mid A = \frac{1}{n}\sum_{j\notin S\cup\{i\}} \tilde{x}_j\tilde{x}_j^T\}$对 W进行边缘化可得

$$\begin{aligned}
&\mathbb{P}\left(\widehat{\Sigma}_{-i} = A, Z \in \mathcal{E}_{\text{prune}}, W \in \mathcal{E}_{\text{thr}}\right) \\
&\overset{(i)}{=} \int_{\mathcal{E}_{\text{thr}}} \mathbb{P}\left(R_{T,-i} \in S(A), |R_T| \le m + w, Z \in \mathcal{E}_{\text{prune}}\right) f(w)dw \\
&\le \int \mathbb{P}\left(R_{T,-i} \in S(A), |R_{T,-i}| \le \min\{m+w, m_{\max}\}, Z \in \mathcal{E}_{\text{prune}}\right) f(w)dw,
\end{aligned}$$

其中，步骤（i）由条件 $|R_T| \le m + w$ 当且仅当 $\Sigma_{-i} \neq \perp$ 成立可得，这来自引理 6.3；最后一个不等式则是因为 $\mathcal{E}_{\text{thr}} = \{w \mid w \le m_{\max}\}$。继续注意，对每个 $S \in S(A)$，我们有 $S = R_{T,\ -i}$ 其中 $|R_{T,\ -i}| \le \min\{m+w, m_{\max}\}$ 仅当$|S| \le \min\{m+w, m_{\max}\} \le m_{\max}$时成立，因此根据引理 (a) 的情形 6.6

$$\begin{aligned}
&\mathbb{P}(R_{T,-i} \in S(A), |R_{T,-i}| \le \min\{m+w, m_{\max}\}, Z \in \mathcal{E}_{\text{prune}}) \\
&\le \exp\left(\frac{2\sqrt{e}B(m_{\max}+1)}{n\sigma_Z}\right)\mathbb{P}\left(R'_{T',-i} \in S(A), \left|R'_{T',-i}\right| \le m+w\right).
\end{aligned}$$

回到上面的积分，通过积分并应用引理（14），我们得到不等式 6.3：

$$\int \mathbb{P}\left(R'_{T',-i} \in S(A), \left|R'_{T',-i}\right| \le m+w\right) f(w)dw = \mathbb{P}(\widehat{\Sigma}'_{-i} = A)$$

因为 $R'_{T',\ -i} = R'_{T'}$ 根据假设， $\tilde{x}'_i = 0$。

不等式（15）的证明基本相同，只是现在我们必须额外注意考虑 $i \in R_T$ 的可能性。与前面的积分不等式一样，引理6.3 给出

$$\begin{aligned}&\mathbb{P}\left(\widehat{\Sigma}'_{-i} = A, Z \in \mathcal{E}_{\text{prune}}, W \in \mathcal{E}_{\text{thr}}\right)\\&\leq \int_{\mathcal{E}_{\text{thr}}} \mathbb{P}\left(R'_{T',-i} \in S(A), \left|R'_{T',-i}\right| \leq \min\{m+w, m_{\max}\}, Z \in \mathcal{E}_{\text{prune}}\right) f(w)dw.\end{aligned}$$

在这种情况下，按照与上述相同的推理，我们应用引理（b）的情形 6.6 得到

$$\begin{aligned}&\mathbb{P}\left(R'_{T',-i} \in S(A), \left|R'_{T',-i}\right| \leq \min\{m+w, m_{\max}\}, Z \in \mathcal{E}_{\text{prune}}\right)\\&\leq \exp\left(\frac{2\sqrt{e}B(m_{\max}+1)}{n\sigma_Z}\right)\mathbb{P}\left(R_{T,-i} \in S(A), |R_{T,-i}| \leq m+w\right),\end{aligned}$$

so

$$\begin{aligned}&\mathbb{P}\left(\widehat{\Sigma}'_{-i} = A, Z \in \mathcal{E}_{\text{prune}}, W \in \mathcal{E}_{\text{thr}}\right)\\&\leq \exp\left(\frac{2\sqrt{e}B(m_{\max}+1)}{n\sigma_Z}\right)\int \mathbb{P}\left(R_{T,-i} \in S(A), |R_{T,-i}| \leq m+w\right) f(w)dw.\end{aligned}$$

注意到以下事实，我们便可对最后一个积分作上界估计：

$$\begin{aligned}\mathbb{P}\left(\widehat{\Sigma}_{-i} = A\right) &\overset{(\star)}{=} \int \mathbb{P}\left(R_{T,-i} \in S(A), |R_T| \leq m+w\right) f(w)dw\\&\geq \int \mathbb{P}\left(R_{T,-i} \in S(A), |R_{T,-i}| \leq m+w-1\right) f(w)dw,\end{aligned}$$

其中，$(\star)$由引理 6.3 得出，然后利用 $|$对数 $\frac{f(w)}{f(w+1)}| \leq \frac{1}{\sigma_{W_{\text{cov}}}}$对所有 w 成立，可知

$$\int \mathbb{P}\left(R_{T,-i} \in S(A), |R_{T,-i}| \leq m+w\right) f(w)dw \leq \exp\left(\frac{1}{\sigma_{W_{\text{cov}}}}\right)\mathbb{P}\left(\widehat{\Sigma}_{-i} = A\right),$$

代入 $\alpha = \frac{1}{\sigma_{W_{\text{cov}}}} + \frac{2\sqrt{e}B}{(m_{\max}+1)\ n\sigma_Z}$ 后，即得到不等式（15）。

结合不等式（14）和（15），得到引理 4.2。对于任意集合 $C \subset \mathbb{R}^{d\times d} \cup \{\perp\}$,

$$\begin{aligned}\mathbb{P}(\widehat{\Sigma}_{-i} \in C) &\leq \mathbb{P}(\widehat{\Sigma}_{-i} \in C, Z \in \mathcal{E}_{\text{prune}}, W \in \mathcal{E}_{\text{thr}}) + \mathbb{P}(Z \notin \mathcal{E}_{\text{prune}}) + \mathbb{P}(W \notin \mathcal{E}_{\text{thr}})\\&\leq e^{\alpha}\mathbb{P}(\widehat{\Sigma}'_{-i} \in C) + \mathbb{P}(Z \notin \mathcal{E}_{\text{prune}}) + \mathbb{P}(W \notin \mathcal{E}_{\text{thr}})\end{aligned}$$

根据不等式 (14)，立即可得以下界 $\mathbb{P}(W \notin \mathcal{E}_{\text{thr}}) = \mathbb{P}(W > \sigma_{W_{\text{cov}}}$ 对数$\frac{1}{\gamma}) = 1\,2\exp(-$ 对数 $\frac{1}{\gamma}) = \frac{\gamma}{2}$。类似地， $\mathbb{P}(Z \notin \mathcal{E}_{\text{prune}}) \leq \frac{n}{2}\exp(-\frac{1}{4\sigma_Z})$，根据观察结论 6.1。关于 $\widehat{}$的上界 $\mathbb{P}(\Sigma'_{-i} \in C)$ 类似，但使用不等式 (15)。

到目前为止，我们已经证明，如果 x 和 x'' 是相邻样本，且唯一差异在于差值 $\tilde{x}_i = x_i - x_{n/2+i}$ 可能非零，而 $\tilde{x}''_i = x''_i - x''_{n/2+i} = 0$，那么返回$\widehat{\ }$ $\widehat{\ }$ $\widehat{\ }$ $\widehat{\ }$回到记号（4），并令 $\Sigma_{-i} = \Sigma_{-i}$（$x$, Z, W）和 $\Sigma'_{-i} = \Sigma_{-i}$（$x''$, Z, W），

$$\widehat{\Sigma}_{-i}(x, Z, W) \overset{d}{=}_{\alpha,\beta} \widehat{\Sigma}_{-i}(x'', Z, W)$$

对于 $\alpha = \frac{1}{\sigma_{W\text{cov}}} + \frac{2\sqrt{e}B}{(m_{\max}+1)}\, n\sigma_Z$ 和 $\beta = \frac{\gamma}{2} + \frac{n}{2}\exp(-\frac{1}{4\sigma_Z})$。因此我们得出结论：如果 x' 是满足 $x'_{-i} = x_{-i}$的任意样本，

$$\widehat{\Sigma}_{-i}(x, Z, W) \overset{d}{=}_{\alpha,\beta} \widehat{\Sigma}_{-i}(x'', Z, W) \overset{d}{=}_{\alpha,\beta} \widehat{\Sigma}_{-i}(x', Z, W).$$

利用群组合（引理 2.2），我们得到

$$\widehat{\Sigma}_{-i}(x, Z, W) \overset{d}{=}_{2\alpha,\beta+e^{\alpha}\beta} \widehat{\Sigma}_{-i}(x', Z, W),$$

这就是所需的引理4.2。

7 均值估计的证明

本节给出引理 4.3 和 4.4 的证明。全文中，我们用撇号区分 MEANSAFE 在输入 x 与 x'（或 A 与 A'）上的输出，因此（例如）$\widehat{\mu}$ 对应于 MEANSAFE 在输入样本 x 时第 9 行的均值，或者 D'_j 对应于 MEANSAFE 在输入样本 x' 时第 2 行的对数直径。我们将使用式 (3) 中的函数 $\Gamma(x, A)$ 对此进行精确定义；该函数表示 MEANSAFE 在输入 x, A 时输出的执行记录。

7.1 引理 4.3 的证明

我们将 $\widetilde{\mu}(x, A)$ 和 $\widetilde{\mu}(x', A)$ 分别简写为 $\widetilde{\mu}$ 和 $\widetilde{\mu}'$，并展开相应的执行记录：

$$(D, \widetilde{D}, R, t, \widehat{\mu}) := \Gamma(x, A) \quad \text{and} \quad (D', \widetilde{D}', R', t', \widehat{\mu}') := \Gamma(x', A).$$

在我们的论证中，$i \in [n]$ 表示样本 x和x' 不同的索引，也就是说，$x_{-i} = x'_{-i}$，尽管我们可能有$x_i \neq x'_i$。

引理 4.3 证明中的核心思路是首先给出均值敏感度的上界，证明（以高概率）$\|\widehat{\mu} - \widehat{\mu}'\|_A$ 很小，除非离群条目x_j过多。我们在引理7.3 中，通过证明对于适当的子组大小 b（回顾一下，在 $\mathcal{S}$ 中将 $[n]$ 随机划分为大小为 n/b 的块，这是MEANSAFE算法中的做法），MEANSAFE 算法能够正确识别所有离群点，而不会修剪掉太多非离群数据点。第二步中，我们通过将敏感度上界与高斯分布较为标准的分布稳定性保证相结合，完成证明（第7.1.1节）；这些保证列于预备知识部分的第2节。

我们首先形式化两个性质，它们将有助于证明引理 7.3中的敏感度上界。回顾一下记号 t（分别为 t'），它表示 MEANSAFE 在输入 x 和 x'上执行第 5–8 行时剪枝的组数；而 R 和 R' 表示所有被剪枝索引构成的集合。接下来的两个引理中，引理 7.1 界定 R 与R' 之间以及 t 与 t'之间的差异；引理 7.2 则是一个一般引理，用于界定具有嵌套指标集的经验均值之差。这两个引理将在引理 7.3 中结合起来，以界定估计均值 $\widehat{\mu} = \frac{1}{n-|R|}\sum_{j\notin R} x_j$ 与 $\widehat{\mu}'$之间的差异。

在陈述引理 7.1之前，回顾对于集合 S，S'，有 $d_{\text{sym}}(S, S') = \max\{|S \setminus S'|, |S' \setminus S|\}$。

引理 7.1（被拒绝索引的稳定性）。设 t, t'和 R, R'如上所述。则 $|t - t'| \leq 1$和 $d_{\text{sym}}(R, R') \leq b$。

证明。令集合 $J := \{j \mid \widetilde{D}_j \neq \perp, \widetilde{D}_j \geq \log(\sqrt{B}/4)\}$ 索引由 MEANSAFE 在样本 x' 上执行时剪除的子组，并类似地定义相对于样本 x' 的 J' D'。则 $t = |J|$和 $R = \cup_{j\in J} S_j$，且 $t' = |J'|$和 $R' = \cup_{j\in J'} S'_j$。我们证明 $d_{\mathrm{sym}}(J, J') \leq 1$，由此立即得出结论 $|t - t'| \leq 1$；而结论 $d_{\mathrm{sym}}(R, R') \leq b$ 则由对所有 $j \in [n/b]$ 都有 $|S_j| = b$ 这一事实得出。

回忆一下，x 和 x' 仅在索引 i处不同，假设对于 $\ell \in [n/b]$，$i \in S_\ell$ 成立。那么对于所有 $j \neq \ell$，$x_{S_j} = x'_{S_j}$成立；特别地，$\mathrm{diam}_A(x_{S_j}) = \mathrm{diam}_A(x'_{S_j})$，因此对于 $j \neq \ell$，$D_j = D'_j$成立。于是，$D+Z$ 和 $D'+Z$中最大的 k 个元素的索引，也就是 TOPk 识别为$\widetilde{\ }$ $\widetilde{\ }$具有最大直径的那些子群（分别记为 $K = \{j \mid D_j \neq \perp\}$ 和 $K' = \{j \mid D'_j \neq \perp\}$），至多相差一个索引：$d_{\mathrm{sym}}(K, K') \leq 1$，且仅当$\ell$ 恰好属于 K 或 K'之一时等号成立。若 ℓ 既不属于 K 也不属于 K'，则 $J = J'$，结论$d_{\mathrm{sym}}(J, J') \leq 1$随即成立。否则，假设 $\ell \in K$，由界 $d_{\mathrm{sym}}(K, K') \leq 1$ 可得$\sim$ $\sim K \setminus \{\ell\} \subset K'$，因此 $D_{K\setminus\{\ell\}} = D'_{K\setminus\{\ell\}}$；若 $\ell \in K'$，则反之亦然。于是，根据 $J \subset K$ 和 $J' \subset K'$，$d_{\mathrm{sym}}(J, J') \leq 1$成立。 □

引理 7.2. 设 $\{y_1, \ldots, y_n\}$ 是一组任意的向量，且 $S \subset S' \subset [n]$。定义 $\mu_S := \frac{1}{|S|}\sum_{i\in S} y_i$ 和 $\mu_{S'} := \frac{1}{|S'|}\sum_{i\in S'} y_i$。则

$$\|\mu_S - \mu_{S'}\| \leq \frac{|S' \setminus S| \,\mathrm{diam}_{\|\cdot\|}(y_{S'})}{|S'|}.$$

证明。注意

$$\mu_S - \mu_{S'} = \mu_S - \left(\frac{|S|}{|S'|}\mu_S + \frac{1}{|S'|}\sum_{i\in S'\setminus S} y_i\right) = \frac{1}{|S'|}\sum_{i\in S'\setminus S}(\mu_S - y_i),$$

根据假设 $S \subset S'$，我们有

$$\max_{i\in S'\setminus S} \|y_i - \mu_S\| \leq \max_{j\in S, i\in S'} \|y_i - y_j\| \leq \mathrm{diam}_{\|\cdot\|}(S').$$

由此可得结论，因为$\|\mu_S - \mu_{S'}\| \leq \frac{1}{|S'|}\sum_{i\in S'\setminus S}$直径$_{\|\cdot\|}(y_{S'}) = \frac{|S'\setminus S|}{\text{直径}_{\|\cdot\|}(x_{S'})\ |S'|}$。 □

现在我们转向所概述的第一步，给出 $\|\widehat{\mu} - \widehat{\mu}'\|_A$ 的显式上界，除非发生以下事件：$\max\{t, t'\} = k$。回顾引理 4.3 的陈述中 $\exp(3\sigma_{\mathrm{top}} \log \frac{2n}{b\gamma})$ 的定义。

引理 7.3. 以至少 $1 - \gamma - n^2 2^{1-b}$ 的概率，$\max\{t, t'\} = k$ or $\|\widehat{\mu} - \widehat{\mu}'\|_A \leq \Delta$。

证明。我们首先证明，对于随机划分$\mathcal{S} \sim \mathrm{Uni}(\mathcal{P}_{n,\ b})$，以至少 $1 - n^2 2^{-b}$ 的概率，$\mathcal{S} = (S_1, \ldots, S_{n/b})$，

$$\mathrm{diam}_A(x_{R^c}) \leq \frac{1}{2}\exp(2\|Z\|_\infty + \|Z'\|_\infty)\sqrt{B}, \tag{16}$$

根据对称性，x' 也满足相同的界。为此，注意索引集合

$$J := \{j \in [n/b] \mid \widetilde{D}_j \neq \perp, \widetilde{D}_j \geq \log(\sqrt{B}/4)\},$$

MEANSAFE 通过并集 R 在第 5–8 行 构造移除的索引 $R = \cup_{j\in J} S_j$。界（16）的第一步是用集合 x_{R^c} 中各个组成集合的直径来界定其直径 R，以下一般引理允许这样做（证明见第 7.1.2 节）。

断言 7.4。设 $\{y_1, \ldots, y_n\}$ 为任意一组向量，且 $\mathcal{S} \sim$ 均匀分布($\mathcal{P}_{n,\ b}$)。以至少 $1 - n^2 2^{-b}$的概率，对所有指标集 $J \subset [n/b]$，集合 $S_J := \cup_{j \in J} S_j$ 满足 $\mathrm{diam}_{\|\cdot\|}(y_{S_J}) \leq 2 \max_{j \in J} \mathrm{diam}_{\|\cdot\|}(y_{S_j})$。

根据断言 7.4，不等式（16）可通过证明以下结论得出

$$\mathrm{diam}_A(x_{S_j}) \leq \exp(2\|Z\|_\infty + \|Z'\|_\infty)\sqrt{B}/4 \tag{17}$$

~对事件 $t < k$ 上的所有 $j \notin J$ 均成立。当 $t < k$ 时，存在一个索引 $\ell \in [n/b]$，使得 $D_\ell \neq \perp$ 且 $\widetilde{D}_\ell \leq \log(\sqrt{B}/4)$，也就是说，$\ell$ 是 $D + Z$ 中最大的 k 个元素之一，但 $\ell \notin J$。因此，对于满足 $D_j = \perp$ 的 $j \notin J$，若 $\log(\mathrm{diam}_A(x_{S_j})) + Z_j$ 不属于 $D + Z$ 中最大的 k 个元素（根据 TOPk 的构造），则有

$$\log(\mathrm{diam}_A(x_{S_j})) \leq \log(\mathrm{diam}_A(x_{S_\ell})) + 2\|Z\|_\infty .$$

~同时，对于所有满足 $D_j \neq \perp$ 的 $j \notin J$，包括 $j = \ell$，根据 J 的定义，我们立即得到

$$\log(\mathrm{diam}_A(x_{S_j})) + Z'_j \leq \log(\sqrt{B}/4).$$

断言（17），进而断言（16），由前面的两个公式即可得出。此外，通过对 MEANSAFE 的两次执行应用并集界，断言 7.4 给出

$$\max\{\mathrm{diam}_A(x_{R^c}), \mathrm{diam}_A(x'_{R'^c})\} \leq \exp(2\|Z\|_\infty + \|Z'\|_\infty)\frac{\sqrt{B}}{2} \quad \text{or} \quad \max\{t, t'\} = k \tag{18}$$

以至少 $1 - n^2 2^{1-b}$ 的概率。

现在我们可以通过以下断言来界定 $\|\widehat{\mu} - \widehat{\mu}'\|_A$ 对于 $\widehat{\mu} = \frac{1}{n-|R|}\sum_{j \notin R} x_j$ 以及 $\widehat{\mu} = \frac{1}{n-|R'|}\sum_{j \notin R'} x'_j$（本质上是多次应用三角不等式），其证明也留待后文（见第 7.1.3节）。

断言 7.5. $\|\widehat{\mu} - \widehat{\mu}'\|_A \leq {}^{4}\!/_{(b+1)n} \max\{\mathrm{diam}_A(x_{R^c}), \mathrm{diam}_A(x'_{R'^c})\}$。

利用断言 7.5，主要引理7.3 便可较快得出。结合显示式（18）以及以下事实（通过初等计算可得）：

$$\mathbb{P}(\max\{\|Z\|_\infty, \|Z'\|_\infty\} > \sigma_{\mathrm{top}} \log(2n/b\gamma)) \leq \gamma,$$

我们得到，以至少 $1 - \gamma - n^2 2^{1-b}$ 的概率，$\max\{t, t'\} = k$ 或

$$\|\widehat{\mu}' - \widehat{\mu}'\|_A \leq \frac{2(b+1)\sqrt{B}}{n} \exp\left(2\|Z\|_\infty + \|Z'\|_\infty\right) \leq \frac{2(b+1)\sqrt{B}}{n} \exp\left(3\sigma_{\mathrm{top}} \log \frac{2n}{b\gamma}\right).$$

回想一下g the a 假设批大小为 $b \geq 4$（因此 $2(b+1) \leq \frac{5}{2}b$），我们得到该引理。 □

7.1.1 完成引理的证明4.3

我们证明，对于任意（可测）事件 $O \subset \mathbb{R}^d \cup \{\perp\}$ 均有

$$\mathbb{P}(\widetilde{\mu} \in O) \leq e^{\alpha+1/\sigma_{W_{\mathrm{mean}}}} \mathbb{P}(\widetilde{\mu}' \in O) + \beta_1 + \beta_2, \tag{19}$$

其中， $\alpha > 0$ 和 $\beta_1 \in (0, 1)$ 通过以下方式确定 $Z^{\mathsf{N}} \sim \mathsf{N}(0, \sigma_{\mathsf{N}}^2 I)$ 的高斯噪声尺度

$$\sigma_{\mathsf{N}} = \begin{cases} \frac{\Delta}{\alpha}\sqrt{1.25 \log \frac{1}{\beta_1}} & \text{if } \alpha \le 1 \\ \frac{\Delta}{\sqrt{2 \log \frac{1}{\beta_1} + 2\alpha} - \sqrt{2 \log \frac{1}{\beta_1}}} & \text{otherwise,} \end{cases} \quad \text{and} \quad \beta_2 = \frac{1}{2} e^{-(k/3-1)/\sigma_{W_{\text{mean}}}} + \gamma + n^2 2^{1-b}.$$

另一方向由对称性可得。我们分别处理 $O \subset \mathbb{R}^d$ 和 $O = \perp$，并在最后合并这两种情况以证明结论（19）。首先假设 $O \subset \mathbb{R}^d$，下面的观察给出了 $\widetilde{\mu} \in O$的充要条件。

观察 7.1。设 $O \subset \mathbb{R}^d$。那么 $\widetilde{\mu} \in O$ 当且仅当 $t \le 2k/3 + W$ 且 $\widehat{\mu} + A^{1/2} Z^{\mathsf{N}} \in O$。

证明。 根据 MEANSAFE第 11 行中返回 $\perp$ 的条件，我们立即得到$\widetilde{\mu} = \perp \notin \mathbb{R}^d$ 当且仅当 $t > 2k/3 + W$；因此，条件 $t \le 2k/3 + W$ 是 $\widetilde{\mu} \in \mathbb{R}^d$的充要条件。由于根据定义， $\widetilde{\mu} = \perp$ 或 $\widetilde{\mu} = \widehat{\mu} + A^{1/2} Z^{\mathsf{N}}$，因此显然可得， $t \le 2k/3 + W$ 和 $\widehat{\mu} + A^{1/2} Z^{\mathsf{N}} \in O$ 共同足以得到 $\widetilde{\mu} \in O$。 □

对被拒绝的索引集合的数量 t和 $\widehat{\mu}$ 进行边缘化后，我们得到以下一系列不等式：

$$\begin{aligned}
&\mathbb{P}(\widetilde{\mu} \in O) \\
&= \mathbb{E}\left[\mathbb{P}(\widehat{\mu} + A^{1/2} Z^{\mathsf{N}} \in O \mid \widehat{\mu}) \mathbb{P}(t \le 2k/3 + W \mid t)\right] \\
&\overset{(i)}{\le} \mathbb{E}\left[\mathbb{P}(\widehat{\mu} + A^{1/2} Z^{\mathsf{N}} \in O \mid \widehat{\mu}) \mathbb{P}(t \le 2k/3 + W \mid t) 1\{\|\widehat{\mu}' - \widehat{\mu}'\|_A \le \Delta\}\right] \\
&\quad + \mathbb{E}\left[\mathbb{P}(t \le 2k/3 + W \mid t) 1\{\max\{t, t'\} = k\}\right] + \gamma + n^2 2^{1-b} \\
&\overset{(ii)}{\le} \mathbb{E}\left[\mathbb{P}(\widehat{\mu} + A^{1/2} Z^{\mathsf{N}} \in O \mid \widehat{\mu}) \mathbb{P}(t \le 2k/3 + W \mid t) 1\{\|\widehat{\mu}' - \widehat{\mu}'\|_A \le \Delta\}\right] \\
&\quad + \mathbb{P}(W \ge k/3 - 1) + \gamma + n^2 2^{1-b} \\
&= \mathbb{E}\left[\mathbb{P}(\widehat{\mu} + A^{1/2} Z^{\mathsf{N}} \in O \mid \widehat{\mu}) \mathbb{P}(t \le 2k/3 + W \mid t) 1\{\|\widehat{\mu}' - \widehat{\mu}'\|_A \le \Delta\}\right] + \beta_2 \qquad (20)
\end{aligned}$$

这里，步骤 (i) 成立是因为 $\|\widehat{\mu}' - \widehat{\mu}'\|_A \le \Delta$ 或最大值$\{t, t'\} = k$ 以至少 $1 - \gamma - n^2 2^{1-b}$ 的概率出现，这由引理 7.3得出；步骤 (ii) 成立是因为 $|t - t'| \le 1$ 由引理 7.1 可知，因此最大值$\{t, t'\} = k$意味着 $t \ge k-1$；最后一个等式由恒等式 $\mathbb{P}(W \ge k/3 - 1) = \frac{1}{2} e^{-(k/3-1)/\sigma_{W_{\text{mean}}}}$ 以及 β_2的定义得出。

继续推导，我们可以将前面表达式中的最后一个期望上界为

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{P}(\widehat{\mu} + A^{1/2} Z^{\mathsf{N}} \in O \mid \widehat{\mu}) \mathbb{P}(t \le 2k/3 + W \mid t) 1\{\|\widehat{\mu}' - \widehat{\mu}'\|_A \le \Delta\}\right] \\
&\overset{(i)}{\le} \exp(\alpha) \mathbb{E}\left[\mathbb{P}(\widehat{\mu}' + A^{1/2} Z^{\mathsf{N}} \in O \mid \widehat{\mu}') \mathbb{P}(t \le 2k/3 + W \mid t)\right] + \beta_1 \\
&\overset{(ii)}{\le} \exp(\alpha + 1/\sigma_{W_{\text{mean}}}) \mathbb{E}\left[\mathbb{P}(\widehat{\mu}' + A^{1/2} Z^{\mathsf{N}} \in O \mid \widehat{\mu}') \mathbb{P}(t' \le 2k/3 + W \mid t')\right] + \beta_1 \\
&= \exp(\alpha + 1/\sigma_{W_{\text{mean}}}) \mathbb{P}(\widetilde{\mu}' \in O) + \beta_1, \qquad (21)
\end{aligned}$$

其中步骤（i）源于带噪声 σ_N 和敏感度上界 $\|\widehat{\mu} - \widehat{\mu}'\| \le \Delta$ 的高斯机制的隐私性（引理 2.5）；步骤（ii）源于 $|t - t'| \le 1$，依据引理 7.1 以及 $W \sim$ 拉普拉斯分布($\sigma_{W_{\text{mean}}}$)；最后一个等式直接由观察 7.1应用

于此处对数据 MEANSAFE的执行 x'。结合不等式（20）和（21），当 $O \subset \mathbb{R}^d$时即可得到结论（19）。

对于 $O = \{\perp\}$ 的情况，我们有

$$\begin{aligned}\mathbb{P}(\widetilde{\mu} = \perp) &= \mathbb{E}\left[\mathbb{P}(t > 2k/3 + W \mid t)\right] \\ &\leq e^{\frac{1}{\sigma_{W_{\text{mean}}}}} \mathbb{E}\left[\mathbb{P}(t' > 2k/3 + W \mid t')\right] = \exp^{\frac{1}{\sigma_{W_{\text{mean}}}}} \mathbb{P}(\widetilde{\mu} = \perp).\end{aligned}$$

这里的两个等式源自 $\perp$ MEANSAFE第 11 行中关于返回MEANSAFE的条件，而不等式则是因为 $|t - t'| \leq 1$ 根据引理7.1，且 $W \sim$ 拉普拉斯分布$(\sigma_{W_{\text{mean}}})$。对于任意 O，结论 (19) 显然成立。

7.1.2 结论7.4的证明

考虑事件 $\mathcal{E}$：对于所有索引 $i_1, i_2 \in [n]$，若 $i_1 \in S_{j_1}$ 且 $i_2 \in S_{j_2}$，则有$\|y_{i_1} - y_{i_2}\| \leq 2\max\{\text{diam}(y_{S_{j_1}}), \text{diam}(y_{S_{j_2}})\}$。该结论在 $\mathcal{E}$ 上成立：对于任意 $J \subset [n/b]$ 和$S_J := \cup_{j \in J} S_j$，存在 $j_1, j_2 \in J$，满足 $i_1 \in S_{j_1}$ 且 $i_2 \in S_{j_2}$，使 $\text{diam}(y_{S_J}) = \|y_{i_1} - y_{i_2}\|$达到最大值，因此

$$\|y_{i_1} - y_{i_2}\| \leq 2\max\{\text{diam}(y_{S_{j_1}}), \text{diam}(y_{S_{j_2}})\} \leq 2\max_{j \in J} \text{diam}(y_{S_j}).$$

剩下要证明的是，$\mathcal{E}$ 发生的概率至少为 $1 - n^2 2^{-b}$。由于不同索引 i_1、$i_2 \in [n]$共有 $\binom{n}{2} \leq \frac{1}{2}n^2$个无序对，只要证明 $\|y_{i_1} - y_{i_2}\| > 2\max\{\text{diam}(y_{S_{j_1}}), \text{diam}(y_{S_{j_2}})\}$ 发生的概率至多为 2^{1-b}，即可由并集界得到结论。

接下来，令 $i_1, i_2 \in [n]$ 和 $i_1 \in S_{j_1}, i_2 \in S_{j_2}$，并令 $c = \frac{1}{2}\|y_{i_1} - y_{i_2}\|$。若 $i_1 = i_2$ 或 $j_1 = j_2$，则无需证明，因此假设 $i_1 \neq i_2$ 且 $j_1 \neq j_2$。令 $C_1 = \{i \in [n] \setminus \{i_1, i_2\} \|y_{i_1} - y_i\| < c\}$ 和 $C_2 = \{i \in [n] \setminus \{i_1, i_2\} \mid \|y_{i_2} - y_i\| < c\}$ 为满足以下条件的索引 i：y_i 分别接近 y_{i_1} 或 y_{i_2}。根据三角不等式，C_1 与 C_2 不相交，因此不失一般性，我们假设 $|C_1| \leq (n-2)/2$。

我们希望证明 $\text{diam}(y_{S_{j_1}}) \geq c$；为此，只需证明存在一个索引属于 $S_{j_1} \setminus \{i_1\}$ 而不属于 C_1。因此，只要证明 $S_{j_1} \setminus \{i_1\} \subset C_1$ 发生的概率至多为 2^{1-b}，就完成了证明。由于 $\mathcal{S} \sim$ 服从均匀分布 Uni($\mathcal{P}_{n,b}$)，集合 $S_{j_1} \setminus \{i_1\}$ 是$[n] \setminus \{i_1, i_2\}$ 的一个大小为 $b-1$的均匀分布子集。因此，它有 $\binom{n-2}{b-1}$ 个不同的可能取值，且有 $\binom{|C_1|}{b-1}$个值满足 $S_{j_1} \setminus \{i_1\} \subset C_1$。所以，$S_{j_1} \setminus \{i_1\} \subset C_1$ 的概率为

$$\mathbb{P}(S_{j_1} \setminus \{i_1\} \subset C_1) = \frac{\binom{|C_1|}{b-1}}{\binom{n-2}{b-1}} = \prod_{i=0}^{b-2} \frac{(|C_1| - i)_+}{n - 2 - i} \leq \left(\frac{|C_1|}{n-2}\right)^{b-1} \leq 2^{1-b},$$

其中，最后一个不等式是因为 $|C_1| \leq (n-2)/2$。

7.1.3 命题7.5 的证明

回顾一下，

$$\widehat{\mu} = \frac{1}{n - |R|} \sum_{j \notin R} x_j \quad \text{and} \quad \widehat{\mu}' = \frac{1}{n - |R'|} \sum_{j \notin R'} x'_j,$$

并定义

$$R_{\text{all}} := R \cup R', \quad \widehat{\mu}_{\text{all}} := \frac{1}{n - |R_{\text{all}}|} \sum_{j \notin R_{\text{all}}} x_j, \quad \widehat{\mu}'_{\text{all}} := \frac{1}{n - |R_{\text{all}}|} \sum_{j \notin R_{\text{all}}} x'_j.$$

引理 7.1 给出 $|R^c \setminus R^c_{\text{all}}| = |R_{\text{all}} \setminus R| \le b$，根据对批大小 b 和拒绝阈值 k 的假设，我们还得到 $|R_{\text{all}}| \le b + |R| \le b + kb \le \frac{n}{2}$。

应用引理 7.2 并取 $S = R^c_{\text{all}}$和 $S' = R^c$，可得

$$\|\widehat{\mu} - \widehat{\mu}_{\text{all}}\|_A \le \frac{|R^c \setminus R^c_{\text{all}}| \operatorname{diam}_A(x_{R^c})}{|R^c|} \le \frac{2b\operatorname{diam}_A(x_{R^c})}{n}$$

由于 $|R| \le |R_{\text{all}}| \le \frac{n}{2}$。再次应用引理 7.2，这次取数据集 x'，$S = R'^c_{\text{all}}$ 和$S' = R'^c$，可得 $\|\widehat{\mu}' - \widehat{\mu}'_{\text{all}}\|_A \le \frac{2b}{n}$ 直径$_A(x'_{R'^c})$。

现在我们给出 $\widehat{\mu}_{\text{all}} - \widehat{\mu}'_{\text{all}}$ 的界。回想一下，索引 i 是 x、x' 中唯一可能不同的索引（也就是说，$x_{-i} = x'_{-i}$），因此可以写作

$$\widehat{\mu}_{\text{all}} - \widehat{\mu}'_{\text{all}} = \frac{1}{n - |R_{\text{all}}|} \sum_{j \notin R_{\text{all}}} (x_j - x'_j) = \frac{\mathbf{1}\{i \notin R_{\text{all}}\}}{n - |R_{\text{all}}|}(x_i - x'_i).$$

如果 $i \in R_{\text{all}}$，则该差为 0。否则，$i \notin R$ 和 $i \notin R'$。由于$|R_{\text{all}}| \le \frac{n}{2}$，我们可以选取某个 $j' \notin R_{\text{all}} \cup \{i\}$。因为 $x_{j'} = x'_{j'}$，我们有$\|x_i - x_{j'}\|_A \le$直径$_A(x_{R^c})$，并且有 $\|x'_i - x_{j'}\|_A \le$直径$_A(x'_{R'^c})$。由三角不等式可得 $\|x_i - x'_i\|_A \le$2 max{直径$_A(x_{R^c})$，直径$_A(x'_{R'^c})$}，因此 $\|\widehat{\mu}_{\text{all}} - \widehat{\mu}'_{\text{all}}\|_A \le \frac{4}{n}$max{直径$_A(x_{R^c})$，直径$_A(x'_{R'^c})$}。结合以上结果，该结论立即由下式推出

$$\begin{aligned}\|\widehat{\mu} - \widehat{\mu}'\|_A &\le \|\widehat{\mu} - \widehat{\mu}_{\text{all}}\|_A + \|\widehat{\mu}_{\text{all}} - \widehat{\mu}'_{\text{all}}\|_A + \|\widehat{\mu}'_{\text{all}} - \widehat{\mu}'\|_A \\ &\le \frac{2b\operatorname{diam}_A(x_{R^c})}{n} + \frac{4\max\{\operatorname{diam}_A(x_{R^c}), \operatorname{diam}_A(x'_{R'^c})\}}{n} + \frac{2b\operatorname{diam}_A(x'_{R'^c})}{n}.\end{aligned}$$

7.2 引理 4.4 的证明

展开执行记录 $\Gamma(x, A)$ 和 $\Gamma(x, A')$（由 3）可得

$$(D, \widetilde{D}, R, t, \widehat{\mu}) := \Gamma(x, A) \quad \text{and} \quad (D', \widetilde{D}', R', t', \widehat{\mu}') := \Gamma(x, A'),$$

注意，给定这对 $(\widetilde{D}, A^{1/2}Z^{\mathsf{N}})$ 后，$\widetilde{\mu}(x, A)$ 与 A 独立（参见 MEANSAFE 的执行过程）；类似地，给定 $(\widetilde{D}', A'^{1/2}Z^{\mathsf{N}})$ 后，$\widetilde{\mu}(x, A')$ 与 A' 独立。因此，只要证明 $A^{1/2}Z^{\mathsf{N}} \overset{d}{=}_{\alpha_1,\beta} A'^{1/2}Z^{\mathsf{N}}$ 和 $\widetilde{D} \overset{d}{=}_{\alpha_2,0} \widetilde{D}'$，基本组合（引理 2.1）和后处理性质（引理2.3）就能推出所需结果 $\widetilde{\mu} \overset{d}{=}_{\alpha_1+\alpha_2,\beta} \widetilde{\mu}'$。

回顾 $Z^{\mathsf{N}} \sim \mathsf{N}(0, \sigma^2_{\mathsf{N}} I)$，我们有 $A^{1/2}Z^{\mathsf{N}} \sim \mathsf{N}(0, \sigma^2_{\mathsf{N}} A)$ 和 $A'^{1/2}Z^{\mathsf{N}} \sim \mathsf{N}(0, \sigma^2_{\mathsf{N}} A')$，因此 $A^{1/2}Z^{\mathsf{N}} \overset{d}{=}_{\alpha_1,\beta} A'^{1/2}Z^{\mathsf{N}}$ 可由假设 $d_{\text{psd}}(A, A') \le \frac{a}{n}$ 以及协方差不同的高斯分布之间的接近性（引理 2.6）立即得到。

为证明 $\widetilde{D} \overset{d}{=}_{\alpha_2,0} \widetilde{D}'$，我们作出以下观察，以界定 A 和 A'的对数马氏范数的敏感度。

观察 7.2. 假设 $A, A' \in \mathbb{R}^{d\times d}$和 $d_{\text{psd}}(A, A') \le \gamma < \infty$。那么，对于任意 $v \in$列空间(A),对数 $\|v\|_A -$对数 $\|v\|_{A'}| \le \gamma/2$。对于任意 $v \notin$列空间(A)，对数 $\|v\|_A =$对数 $\|v\|_{A'} = \infty$。

证明。注意到 $d_{\text{psd}}(A, A') < \infty$ 显然蕴含 A 和 A' 均为半正定，且它们的列空间相同，由此可立即得出第二个结论。对于 $v \in$ 列空间(A)，我们只需证明 $\log \|v\|_{A'} \le \frac{1}{2}\gamma + \log \|v\|_A$，因为反向不等式可由对称性得出。根据假设，

$$\left\|A'^{\dagger/2}(A - A')A'^{\dagger/2}\right\|_{\text{op}} \le d_{\text{psd}}(A, A') \le \gamma$$

因此 $A'^{\dagger/2}(A-A')A'^{\dagger/2} \preceq \gamma I$。用 $A'^{1/2}$ 进行共轭并整理各项，得到$\Pi_{A'}A\Pi_{A'} \preceq (1+\gamma)A'$。由于 $\Pi_{A'}=\Pi_A$，我们有 $\Pi_{A'}A\Pi_{A'}=A$，从而得到 $A \preceq (1+\gamma)A'$，或等价地得到 $A'^{\dagger} \preceq (1+\gamma)A^{\dagger}$。因此$\|v\|_{A'}^2 \le (1+\gamma)\|v\|_A^2$。对两边取平方根和对数即可证明该结论，因为 $\log(\sqrt{1+\gamma}) \le \frac{\gamma}{2}$。 □

这一观察，再结合我们的构造—— $\widetilde{\mu}(x,A)$ 和 $\widetilde{\mu}(x,A')$ 使用相同的（随机）划分 $\mathcal{S}=(S_1,\dots,S_{n/b})$—— 可知 $|D_j - D'_j| \le \frac{a}{2n}$ 对所有 $j \in [n/b]$ 都成立；因此$\|D-D'\|_\infty \le \frac{a}{2n}$ （取值为无穷大的元素对应的索引相同）。TOPk（引理 2.4）的接近性以及我们的选择 $\sigma_{\text{top}}=\frac{ka}{n\alpha_2}$ 随即给出 $\widetilde{D} \overset{d}{=}_{\alpha_2,0} \widetilde{D}'$。

8 讨论

经典统计学中的均值估计看似简单，但要在差分隐私约束下自适应且准确地估计均值，则需要精巧的方法。尽管我们已经提出了（近乎）极小极大最优的均值估计过程，仍有许多问题尚待解决，我们希望自己或其他研究者能够继续探索。从实践角度来看，虽然我们的过程可以实现，但为保证隐私而保留的数值常数因子，再加上 n 和 $\log\frac{1}{\delta}$ 中的对数因子，可能会使该过程难以有效应用。从理论角度来看，尝试去除我们界中出现的对数因子仍然很有意义。此外，虽然我们可以适应比次高斯矩界更弱的条件（通过 ADAMEAN 方法），但或许能够提出更精细的过程或更紧的分析，从而实现对维度 d 和隐私级别 ε 的最优依赖。与 p 阶矩已知的情形相比，我们的结果似乎大约松了一个 $(\sqrt{d}/\varepsilon)^{1/p}$ 因子（回顾样本 2 和 4）。在这些更一般的情形下，差分隐私估计究竟还有多大的改进空间，将是一个值得探究的问题。

A 标准隐私结果的证明

A.1 引理 2.5

第一种情况由 Dwork 和 Roth [12, 定理 3.22] 得出。对于第二种情况，我们使用 Mironov 的 R´enyi 差分隐私 [33]。分布 P 与 Q 之间的 R´enyi α-散度为 $D_\alpha(P\|Q) = \frac{1}{\alpha-1}$对数 $\int(\frac{dP}{dQ})^\alpha dQ$，并且 [33, 命题 3] 若 $D_\alpha(P\|Q) \le c$，则对于所有可测的 A 和 $\delta>0$，都有 $P(A) \le \exp(c+\log(1/\delta)\,\alpha-1)Q(A)+\delta$。~~高斯~~分布的 R´enyi 散度具有显式形式

$$D_\alpha\left(\mathsf{N}(\mu_1,\tau^2\Sigma)\|\mathsf{N}(\mu_2,\tau^2\Sigma)\right) = \frac{\alpha}{2\tau^2}\|\mu_1-\mu_2\|_\Sigma^2.$$

当 $\rho \ge \|\mu_1-\mu_2\|_\Sigma$ 时，我们令 $\beta=\alpha-1$，并可知对于满足 ε 的

$$\varepsilon = \frac{\rho^2}{2\tau^2}+\frac{\beta\rho^2}{2\tau^2}+\frac{\log(1/\delta)}{\beta}$$

我们得到 $\mathsf{N}(\mu_1,\tau^2\Sigma) \overset{d}{=}_{\varepsilon,\,\delta} \mathsf{N}(\mu_2,\tau^2\Sigma)$。选择 β 以最小化前述 ε，我们得到$\varepsilon = \frac{\rho^2}{2\tau^2}+\frac{\rho}{\tau}\sqrt{2}$对数$\frac{1}{\delta}$，并在 $\frac{\rho^2}{2}\eta^2+\sqrt{2}$对数$\frac{1}{\delta}\rho\eta-\varepsilon$ 中求解 $\eta=\frac{1}{\tau}$，得到

$$\tau = \frac{1}{\eta} = \frac{\rho}{\sqrt{2\log\frac{1}{\delta}+2\varepsilon}-\sqrt{2\log\frac{1}{\delta}}}$$

始终足以保证 $\mathsf{N}(\mu_1, \tau^2\Sigma) \overset{d}{=}_{\varepsilon,\ \delta} \mathsf{N}(\mu_2, \tau^2\Sigma)$。

A.2 引理证明 2.6

不失一般性，我们可以假设 $\mu = 0$。我们首先将问题归结为 Σ_1 和Σ_2 为满秩矩阵的情形。由于 $d_{\text{psd}}(\Sigma_1, \Sigma_2) < \infty$，我们立即得到存在一个向量空间 $V \subset \mathbb{R}^d$，使得 $V =$ 列空间$(\Sigma_1) =$ 列空间(Σ_2)。令 $k =$ 维数(列空间(Σ_1))，取 $U \in \mathbb{R}^{d\times k}$为满足 $V =$列空间(U) 的正交矩阵。随机变量 $X \sim \mathsf{N}(0, \Sigma_1)$ 和 $Y \sim \mathsf{N}(0, \Sigma_2)$ 的支撑集分别为 V，且乘以 U^T是 V与 $\mathbb{R}^k$之间的同构，因此 $X \overset{d}{=}_{\varepsilon,\delta} Y$当且仅当 $U^T X \overset{d}{=}_{\varepsilon,\delta} U^T Y$。当然，$U^T X \sim \mathsf{N}\left(0, U^T\Sigma_1 U\right)$且$U^T Y \sim \mathsf{N}\left(0, U^T\Sigma_2 U\right)$，并且 $U^T\Sigma_1 U$和 $U^T\Sigma_2 U$均为满秩矩阵。U的正交性给出 $d_{\text{psd}}(U^T\Sigma_1 U, U^T\Sigma_2 U) = d_{\text{psd}}(\Sigma_1, \Sigma_2) \le \gamma$。因此，只要证明该引理对满秩矩阵 $U^T\Sigma_1 U$和 $U^T\Sigma_2 U$成立，就能证明关于 Σ_1和 Σ_2的结论。

我们接下来讨论满秩情形，论证与 Brown 等人 [7, 引理 4.15] 类似。定义 $D_1 = \Sigma_1^{1/2}\Sigma_2^{-1}\Sigma_1^{1/2} - I$和 $D_2 = \Sigma_2^{1/2}\Sigma_1^{-1}\Sigma_2^{1/2} - I$。由于 D_1与 $\Sigma_2^{-1/2}\Sigma_1\Sigma_2^{-1/2} - I$具有相同的谱，根据假设可得 $\|D_1\|_* \le \gamma$和 $\|D_2\|_* \le \gamma$。

令 f_1为 $P_1 = \mathsf{N}(0, \Sigma_1)$的概率密度，$f_2$为 $P_2 = \mathsf{N}(0, \Sigma_2)$的概率密度。接下来，要证明 (ε, δ)-接近性，只需证明 $\ell(W) := \left|\text{对数 } \frac{f_1(W)}{f_2(W)}\right| \le \varepsilon$ 的概率至少为 $1 - \delta$，其中 W从 P_1 或 P_2中抽取。根据对称性，只需证明 $W \sim P_1$ 时的这一界。展开 ℓ，得到

$$\ell(w) = \left|\frac{1}{2}\log\frac{\det(\Sigma_2)}{\det(\Sigma_1)} + \frac{1}{2}w^T(\Sigma_2^{-1} - \Sigma_1^{-1})w\right| \le \frac{1}{2}\left|w^T(\Sigma_2^{-1} - \Sigma_1^{-1})w\right| + \frac{1}{2}\left|\log\frac{\det(\Sigma_2)}{\det(\Sigma_1)}\right|. \quad (22)$$

最后一项与 w 无关，且满足以下界

$$\begin{aligned}\left|\log\frac{\det(\Sigma_2)}{\det(\Sigma_1)}\right| &= \max\left\{\log\det(\Sigma_2^{1/2}\Sigma_1^{-1}\Sigma_2^{1/2}), \log\det(\Sigma_1^{1/2}\Sigma_2^{-1}\Sigma_1^{1/2})\right\} \\ &\le \max\{\mathrm{tr}(D_2), \mathrm{tr}(D_1)\} \le \max\{\|D_2\|_*, \|D_1\|_*\} \le \gamma,\end{aligned}$$

其中 对于任意正定的 A，第一个不等式成立，因为 $\log\det(A) \le \mathrm{tr}(A - I)$.

现在我们以高概率对右侧不等式（22）中的第一项进行界定。由于 $W \sim P_1$，白化后的随机变量 $Z = \Sigma_1^{-1/2}W \sim \mathsf{N}(0, I)$。因此可得

$$\left|W^T(\Sigma_2^{-1} - \Sigma_1^{-1})W\right| = \left|Z^T D_1 Z\right|,$$

因此，根据 Hanson-Wright 不等式 [例如 38，定理 6.2。1], 我们以至少 $1 - \delta$ 的概率有

$$\left|Z^T D_1 Z\right| \le |\mathrm{tr}(D_1)| + 2\|D_1\|_{\text{Fr}}\sqrt{\log\frac{2}{\delta}} + 2\|D_1\|_{\text{op}}\log\frac{2}{\delta} \le 5\gamma\log\frac{2}{\delta},$$

其中，该不等式成立是因为 $\|D_1\|_{\text{op}} \le \|D_1\|_{\text{Fr}} \le \|D_1\|_* \le \gamma$ 且 $\log\frac{2}{\delta} \ge 1$。因此，以至少 $1 - \delta$ 的概率，$\ell(W) \le 6\gamma\log\frac{2}{\delta} \le \varepsilon$。

参考文献

[1] M. Abadi、A. Chu、I. Goodfellow、B. McMahan、I. Mironov、K. Talwar 和 L. Zhang。差分隐私深度学习。载于第 23 届 ACM 计算机与通信安全会议（ACM CCS），第 308–318 页，2016 年。

[2] D. Alabi、P. K. Kothari、P. Tankala、P. Venkat 和 F. Zhang。高斯分布的隐私估计：高效、稳健且最优。arXiv:2212。*08018 [cs.DS]*, 2022。[3] 苹果差分隐私团队。大规模隐私学习，2017 年。可访问 https://machinelearning.apple.com/2017/12/06/learning-with-privacy-at-scale.html.

[4] H. Ashtiani 和 C. Liaw. 用于高斯分布学习及更广泛问题的私有多项式时间算法。载于第三十五届计算学习理论年会论文集，2022年。[5] R. F. Barber 和 J. C. Duchi. 隐私与统计风险：形式化方法与极小极大界。arXiv:1412.*4451 [*math.ST*]*, 2014.[6] S. Biswas、Y. Dong、G. Kamath 和 J. Ullman. CoinPress：实用的私有均值与协方差估计。载于神经信息处理系统进展第33卷，2020年。[7]G. Brown、M. Gaboardi、A. Smith、J. Ullman 和 L. Zakynthinou. 无需私有协方差估计的协方差感知私有均值估计。arXiv:2106.13329*[*cs.LG*]*, 2021.[8] R. Chen、A. Gittens 和 J. A. Tropp. 掩码样本协方差估计量：基于矩阵集中不等式的分析。信息与推断，即将发表，2012年。[9] H. Cram´er. 统计学数学方法。普林斯顿大学出版社，1946年。[10]I. Diakonikolas 和 D. M. Kane. 算法高维稳健统计。剑桥大学出版社，2022年。[11] C. Dwork 和 J. Lei. 差分隐私与稳健统计。载于第四十一届 ACM 计算理论年会论文集，第371–380页，2009年。[12] C. Dwork 和 A. Roth. 差分隐私的算法基础。理论计算机科学基础与趋势，9(3 & 4):211–407，2014年。[13] C. Dwork、K. Kenthapadi、F. McSherry、I. Mironov 和 M. Naor. 我们的数据，我们自己：通过分布式噪声生成实现隐私。载于密码学进展（EUROCRYPT 2006），2006年。[14] C. Dwork、F. McSherry、K. Nissim 和 A. Smith. 在隐私数据分析中按敏感度校准噪声。载于第三届密码学理论会议论文集，第265–284页，2006年。[15] C. Dwork、G. N. Rothblum 和 S. P. Vadhan. 提升与差分隐私。载于第51届计算机科学基础年会，第51–60页，2010年。[16] C. Dwork、K. Talwar、A. Thakurta 和 L. Zhang. Analyze Gauss：隐私保护主成分分析的最优界。载于第四十六届 ACM 计算理论年会论文集。ACM，2014年。[17] U. Erlingsson、V. Pihur 和 A. Korolova. RAPPOR：随机化可聚合隐私保护序数响应。载于第21届 ACM 计算机与通信安全会议（CCS）论文集，2014年。

[18] S. Garfinkel、J. Abowd 和 S. Powazek. 部署差分隐私时遇到的问题。载于2018年电子社会隐私研讨会（WPES）论文集，第133–137页，2018年。[19] F. Hampel、E. Ronchetti、P. Rousseeuw 和 W. Stahel. 稳健统计：基于影响函数的方法。约翰·威立父子出版公司，1986年。[20]M. Hardt 和 G. N. Rothblum. 用于隐私保护数据分析的乘法权重机制。载于第51届计算机科学基础年会，第61–70页，2010年。[21] S. B. Hopkins、G. Kamath 和 M. Majid. 通过平方和指数机制实现纯差分隐私下的高效均值估计。载于第五十四届 ACM 计算理论年会论文集，第1406–1417页。ACM，2022年。

[22] S. B. Hopkins、G. Kamath、M. Majid 和 S. Narayanan。稳健性意味着统计估计中的隐私。arXiv:2212.*05015* [cs.DS], 2022。

[23] Z. Huang、Y. Liang 和 K. Yi。差分隐私下的实例最优均值估计。载于《神经信息处理系统进展》第 34 卷，第 34 卷，页码 25993–26004，2021 年。

[24]P. J. Huber。位置参数的稳健估计。《数理统计年鉴》，35(1):73–101，1964 年。

[25] P. J. Huber 和 E. M. Ronchetti。《稳健统计》。John Wiley and Sons，第二版，2009 年。

[26] G. Kamath、J. Li、V. Singhal 和 J. R. Ullman。私密学习高维分布。arXiv:1805.*00216* [cs.DS], 2018。

[27]P. K. Kothari、P. Manurangsi 和 A. Velingker。通过稳定化凸松弛进行隐私保护稳健估计。载于《第三十五届计算学习理论年会论文集》，第 1–55 页，2022 年。

[28]A. N. D. A. D. Lauger、P. E. Singer、D. Kifer、J. P. Reiter、A. Machanavajjhala、S. L. Garfinkel、S. A. Dahl、M. Graham、V. Karwa、H. Kim、P. Leclerc、I. M. Schmutte、W. N. Sexton、L. Vilhuber 和 J. M. Abowd。美国人口普查局统计披露限制的现代化。在线访问地址：https://www2.census.gov/cac/sac/meetings/2017-09/statistical-disclosure-li模仿.pdf, 2017。[29] B. Laurent 和 P. Massart。通过模型选择自适应估计二次泛函。《统计年刊》，28(5):1302 – 1338，2000。[30] L. Le Cam。《统计决策理论中的渐近方法》。施普林格出版社，1986。[31]X. Liu、W. Kong、S. Kakade 和 S. Oh。稳健且满足差分隐私的均值估计。arXiv:2102.*09159* [cs.LG], 2021。[32] X. Liu、W. Kong 和 S. Oh。高维中的差分隐私与稳健统计。载于《第三十五届计算学习理论年会》论文集，第1167–1246页，2022年。

[33] I. Mironov。Rényi差分隐私。载于第三十届IEEE计算机安全基础研讨会（CSF），第263–275页，2017年。[34]G. Qiao、W. Su 和 L. Zhang。一次性差分隐私 top-k 选择。载于第三十八届国际机器学习会议论文集，2021年。[35] T. Steinke 和 J. Ullman。介于纯差分隐私与近似差分隐私之间。《隐私与保密期刊》，7(2):3–22，2017年。[36] J. W. Tukey。关于从污染分布中抽样的综述。普林斯顿大学，1960年。[37] R. Vershynin。随机矩阵非渐近分析导论。载于《压缩感知：理论与应用》，第5章，第210–268页。剑桥大学出版社，2012年。[38] R. Vershynin。《高维概率：数据科学中的导论与应用》。剑桥大学出版社，2019年。[39]M. J. Wainwright。《高维统计：非渐近视角》。剑桥大学出版社，2019年。