第九章 差分隐私统计推断

本章导读

随着人工智能发展，大数据在各行各业的广泛应用，人们对于隐私和数据安全的关注度不断提高。数据中往往涉及大量个人隐私信息，在某种程度上，隐私不可怕，可怕的是用户的行为可通过大数据分析被预测出来进而导致隐私的泄露。伴随着层出不穷的隐私泄露问题和社会对数据安全的要求，数据作为智能时代的基石，不可能因噎废食，完全放弃。如何对隐私数据进行保护以防止敏感信息泄露已经成为当前面临得艰巨任务。针对富含敏感信息的数据集，直接发布参数的估计将会泄露隐私，因为参数的估计中包含了数据集的信息，攻击者可通过自己所拥有的背景信息和数据集变化导致的参数估计改变来推测出数据集中的敏感信息。针对隐私数据分析，差分隐私技术可通过添加少量的噪声来达到高级别的隐私保护。

9.1 差分隐私基本概念

差分隐私由 Dwork 于 2006 年首次提出，它给出了隐私的严格数学定义，并且做到了隐私保护与数据实用间的权衡。差分隐私要求一个随机算法作用于两个相近的数据集，它们输出结果的概率分布应当足够接近。首先我们定义两个只相差一条数据的数据集为相邻数据集，则差分隐私的具体定义如下:

定义 9.1.1.（差分隐私） 给定两个相邻数据集 D_1, D_2，随机化函数 $\mathcal{M}: D \to \mathbb{R}^d$，则 $\mathcal{M}$ 满足 $(\varepsilon, \delta) -$ 差分隐私当且仅当满足如下条件:

$$\Pr[\mathcal{M}(D_1) \in S] \leq \Pr[\mathcal{M}(D_2) \in S] \times e^{\varepsilon} + \delta \qquad (9.1.1)$$

其中 $S \subset \text{Range}(\mathcal{M}), \varepsilon$ 称为隐私预算，δ 为松弛项。δ 为 0 时，我们又称算法 $\mathcal{M}$ 满足 $\varepsilon -$ 差分隐私，$\delta > 0$ 时算法 $\mathcal{M}$ 以 δ 的概率不满足 $\varepsilon -$ 差分隐私。(ε, δ) 的大小体现了隐私保护的程度，ε 和 δ 越接近于零，隐私保护的程度越高。

9.1.1 噪声机制

实现差分隐私的一般手段为在算法中加人随机噪声扰动，常用的噪声机制有拉普拉斯机制、高斯机制、指数机制。添加的噪声量与被扰动目标的敏感度和隐私预算有关，其中敏感度定义如下：

定义 9.1.2.（敏感度）对于任意两个相邻数据集 D_1, D_2，算法 $f: D \to \mathbb{R}^d$ 的敏感度定义为:

$$\Delta = \max_{D_1, D_2 \text{ 相邻}} \|f(D_1) - f(D_2)\| \tag{9.1.2}$$

其中 $\|\cdot\|$ 取 l_1- 或 l_2- 范数分别可以得到 l_1- 敏感度和 l_2- 敏感度，对应不同的噪声机制。敏感度衡量了单个数据改变对算法输出的影响，为了保证差分隐私，用于扰动的随机噪声大小将正比于算法的敏感度。

定义 9.1.3.（拉普拉斯机制）对于算法 $f: D \to \mathbb{R}^d, f$ 的 l_1 - 敏感度为 Δ_1，则拉普拉斯机制:

$$\mathcal{M}(D) = f(D) + \eta \tag{9.1.3}$$

满足 $\varepsilon-$ 差分隐私，其中 $\eta \sim \mathrm{Lap}\left(\frac{\Delta_1}{\varepsilon}\right)^d$ 。

定义 9.1.3 的分析如下：

一维拉普拉斯分布的密度函数为:

$$f(x \mid b) = \frac{1}{2b}\exp\left(-\frac{|x|}{b}\right)$$

因此当 $\mathcal{M}(D_1), \mathcal{M}(D_2)$ 得到同一个输出 y 时，其概率之比为:

$$\frac{\Pr[\mathcal{M}(D_1)=y]}{\Pr[\mathcal{M}(D_2)=y]}=\frac{\Pr[\eta_1=y-f(D_1)]}{\Pr[\eta_2=y-f(D_2)]}=\prod_{i=1}^{d}\left(\frac{\exp\left(-\frac{\varepsilon|y_i-f(D_1)_i|}{\Delta}\right)}{\exp\left(-\frac{\varepsilon|y_i-f(D_2)_i|}{\Delta}\right)}\right)$$

$$=\prod_{i=1}^{d}\exp\left(\frac{\varepsilon(|y_i-f(D_1)_i|-|y_i-f(D_2)_i|)}{\Delta}\right)$$

$$\leq\prod_{i=1}^{d}\exp\left(\frac{\varepsilon|f(D_1)_i-f(D_2)_i|}{\Delta}\right)$$

$$=\exp\left(\frac{\varepsilon\|f(D_1)-f(D_2)\|_{l_1}}{\Delta}\right)$$

$$\leq\exp(\varepsilon),$$

根据定义 9.2.1 ，拉普拉斯机制满足 $\varepsilon-$ 差分隐私。

一维拉普拉斯分布的概率密度函数为 $f(x\mid b)=\frac{1}{2b}\exp\left(\frac{-|x|}{b}\right)$ 。 ε 越接近于 0 ，所加人的拉普拉斯噪声越大，隐私保护程度就越强，但同时输出结果所包含的真实信息也会被更加混淆。拉普拉斯机制使算法严格满足 $\varepsilon-$ 差分隐私，而高斯机制通过加人正态分布的噪声，使算法满足松他的 (ε,δ) - 差分隐私，在某些情况下能使结果包含更多的真实信息。

定义 9.1.4.（高斯机制）对于算法 $f:D\to\mathbb{R}^d,f$ 的 l_2 - 敏感度为 Δ_2，则高斯机制:

$$\mathcal{M}(D)=f(D)+\eta \tag{9.1.4}$$

满足 (ε,δ) - 差分隐私，其中 $\eta\sim\mathcal{N}\left(0,\frac{2\Delta_2^2\ln\left(\frac{1.25}{\delta}\right)}{\varepsilon^2}\right)^d$ 。

定义 9.1.4 的分析如下：

若算法输出维度 $d=1$， 则有 $\Delta f=\Delta_1 f=\Delta_2 f$ 因此相邻数据集其隐私损失的绝对值为

$$\left|\log\frac{e^{-1/2\sigma^2x^2}}{e^{-1/2\sigma^2(x+\Delta f)^2}}\right| = \left|\log e^{(-1/2\sigma^2)[x^2-(x+\Delta f)^2]}\right|$$
$$= \left|-\frac{1}{2\sigma^2}[x^2-(x^2+2x\Delta f+\Delta f^2)]\right|$$
$$= \left|\frac{1}{2\sigma^2}(2x\Delta f+\Delta f^2)\right|$$

为了保证隐私损失以至少 $1-\delta$ 的概率被 ε 约束，我们需要

$$\Pr[|x| \geq \sigma^2\varepsilon/\Delta f - \Delta f/2] < \delta$$

也就是 σ 要满足

$$\Pr[x \geq \sigma^2\varepsilon/\Delta f - \Delta f/2] < \delta/2$$

我们始终假设 $\varepsilon \leq 1 \leq \Delta f$ 。利用正态分布的尾部界限

$$\Pr[x > t] \leq \frac{\sigma}{\sqrt{2\pi}}e^{-t^2/2\sigma^2}$$

为了使尾部界限的概率小于 $\delta/2$，我们取 $t = \sigma^2\varepsilon/\Delta f - \Delta f/2$，我们有

$$\log\big((\sigma^2\varepsilon/\Delta f - \Delta f/2)/\sigma\big) + (\sigma^2\varepsilon/\Delta f - \Delta f/2)/2\sigma^2 > \log(2/\sqrt{2\pi}\delta)$$

若 $\sigma = c\Delta f/\varepsilon$，则我们需要找到 c 的约束。首先为了使上式对数内非负，需要 $c \geq 3/2$，对于第二项，有

$$\left(\frac{1}{2\sigma^2}\frac{\sigma^2\varepsilon}{\Delta f}-\frac{\Delta f}{2}\right)^2 = \frac{1}{2\sigma^2}\left[\Delta f\left(\frac{c^2}{\varepsilon}-\frac{1}{2}\right)\right]^2$$
$$= \left[\Delta f\left(\frac{c^2}{\varepsilon}-\frac{1}{2}\right)\right]^2\left[\frac{\varepsilon^2}{c^2(\Delta f)^2}\right]\frac{1}{2}$$
$$= \frac{1}{2}\left(\frac{c^2}{\varepsilon}-\frac{1}{2}\right)\frac{\varepsilon^2}{c^2}$$
$$= \frac{1}{2}(c^2-\varepsilon+\varepsilon^2/4c^2)$$

由于 $\varepsilon \leq 1, c^2-\varepsilon+\varepsilon^2/4c^2$ 在 $c \geq 3/2$ 时导数为正，因此 $c^2-\varepsilon+\varepsilon^2/4c^2 \geq$

$c^2 - 8/9$，需要满足

$$c^2 - 8/9 > 2\log\left(\sqrt{\frac{2}{\pi}}\frac{1}{\delta}\right)$$

也就是需要 $c^2 > 2\log(\sqrt{2/\pi}) + 2\log(1/\delta) + \log\left(e^{8/9}\right) = \log(2/\pi) + \log\left(e^{8/9}\right) + 2\log(1/\delta)$,由于 $(2/\pi)e^{8/9} < 1.55$，所以只需满足 $c^2 > 2\log(1.25/\delta)$ 。

将 $\mathbb{R}$ 分为 $\mathbb{R} = R_1 \cup R_2$， 其中 $R_1 = \{x \in \mathbb{R}: |x| \le c\Delta f/\varepsilon\}$ 和 $R_1 = \{x \in \mathbb{R}: |x| > c\Delta f/\varepsilon\}$ 。给定任意子集 $S \subseteq \mathbb{R}$ ，定义

$$S_1 = \{f(x) + x \mid x \in R_1\}$$
$$S_2 = \{f(x) + x \mid x \in R_2\}$$

我们有

$$\begin{aligned}\Pr_{x\in\mathcal{N}(0,\sigma^2)}[f(x) + x \in S] &= \Pr_{x\in\mathcal{N}(0,\sigma^2)}[f(x) + x \in S_1] + \Pr_{x\in\mathcal{N}(0,\sigma^2)}[f(x) + x \in S_2] \\ &\le \Pr_{x\in\mathcal{N}(0,\sigma^2)}[f(x) + x \in S_1] + \delta \\ &\le e^{\varepsilon}\left(\Pr_{x\in\mathcal{N}(0,\sigma^2)}[f(x) + x \in S_1]\right) + \delta\end{aligned}$$

因此对于一维输出高斯机制满足 (ε, δ) – 差分隐私。

对于高维的情况，定义 $\Delta f = \Delta_2 f$ 。令 v 为任意满足 $\| v \| \le \Delta f$ 的向量。对于相邻数据集 x, y 我们感兴趣的是 $v = f(x) - f(y)$，因为这是我们需要用噪声混淆的。与一维一样，我们的隐私损失为

$$\begin{aligned}\left|\log\frac{e^{(-1/2\sigma^2)\|x-\mu\|^2}}{e^{(-1/2\sigma^2)\|x+v-\mu\|^2}}\right| &= \left|\log e^{(-1/2\sigma^2)[\|x-\mu\|^2 - \|x+v-\mu\|^2]}\right| \\ &= \left|\frac{1}{2\sigma^2}(\| x \|^2 - \| x + v \|^2)\right|\end{aligned}$$

因为球对称法向量的分布依赖于绘制其组成法向量的正交基，因此我们可以在一个与 v 对齐的基上工作。给定偏差 $b_1, \cdots, b_m, \lambda_i \sim \mathcal{N}(0, \sigma^2)$， 然后定义 $x^{[i]} = \lambda_i b_i$，最后令 $x = \sum_{i=1}^{m} x^{[i]}$ 。假设 b_1 与 v 平行，我们关注 $|\| x \|^2 - \| x + v \|^2|$ 。

考虑底为 $v+x^{[1]}$，边 $\sum_{i=2}^{m}x^{[i]}$ 正交于 v 的直角三角形，其斜边为 $x+v$ 。

$$\| x+v \|^2 = \left\|v+x^{[1]}\right\|^2 + \sum_{i=2}^{m} \left\|x^{[i]}\right\|^2$$

$$\| x \|^2 = \sum_{i=1}^{m} \left\|x^{[i]}\right\|^2$$

因为 v 与 x_1 平行，我们有 $\left\|v+x^{[1]}\right\|^2 = (\| v \| +\lambda_1)^2$ 。因此 $\| x+v \|^2 - \| x \|^2 = \| v \|^2 + 2\lambda_1 \| v \|$ 。由于 $\| v \| \leq \Delta f$，且 $\lambda \sim \mathcal{N}(0,\sigma^2)$，所以我们又回到了一维的情况，用 λ_1 代替式 (A.1)：

$$\left|\frac{1}{2\sigma^2}(\| x \|^2 - \| x+v \|^2)\right| \leq \left|\frac{1}{2\sigma^2}(2\lambda_1\Delta f - \Delta f^2)\right|$$

剩余证明过程同上。

一维高斯分布的概率密度函数为 $f(x \mid \mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$ 。拉普拉斯机制输出结果的标准差为 $O\left(\frac{d\Delta_1}{\varepsilon}\right)$，高斯机制输出结果的标准差为 $O\left(\frac{\sqrt{\ln\frac{1}{\delta}}d\Delta_2}{\varepsilon}\right)$ 。随着输出结果的维度 d 提升，算法的 l_2- 敏感度将比 l_1- 敏感度更小，从而使得高斯机制在高维上输出结果的标准差更小。在同样隐私预算 ε 的情况下，高斯机制输出的结果更接近真实值。

拉普拉斯机制和高斯机制都是作用在输出为连续型数据的算法上的。当算法的输出为离散值时，指数机制通过随机选择输出来实现差分隐私。具体来说，给定候选集合 $\mathcal{R}$ ，指数机制根据效用函数 $u: D\times\mathcal{R}\to R$ 的大小及其敏感度随机选择一个元素 $r\in\mathcal{R}$ 作为输出。效用函数的敏感度定义为:

$$\Delta u \triangleq \max_{r\in\mathcal{R}} \max_{D_1,D_2\ \text{相邻}} |u(D_1,r)-u(D_2,r)|$$

定义 9.1.5. （指数机制） 随机算法 $\mathcal{M}_E(D,u,\mathcal{R})$ 随机输出一个元素 $r\in\mathcal{R}$,

输出概率正比于 $\exp\left(\frac{\varepsilon}{2\Delta u}u(D,r)\right)$，满足 $\varepsilon-$ 差分隐私。

定义 9.1.5 的分析如下：

对于相邻的数集 D_1, D_2，通过指数机制后的结果 $\mathcal{M}_E(D_1,u,\mathcal{R}), \mathcal{M}_E(D_2,u,\mathcal{R})$ 都返回元素 r 时，其概率之比为:

$$
\begin{aligned}
\frac{\Pr[\mathcal{M}_E(D_1,u,\mathcal{R})=r]}{\Pr[\mathcal{M}_E(D_2,u,\mathcal{R})=r]} &= \frac{\left(\frac{\exp\left(\frac{\varepsilon u(D_1,r)}{2\Delta u}\right)}{\sum_{r'\in\mathcal{R}}\exp\left(\frac{\varepsilon u(D_1,r')}{2\Delta u}\right)}\right)}{\left(\frac{\exp\left(\frac{\varepsilon u(D_2,r)}{2\Delta u}\right)}{\sum_{r'\in\mathcal{R}}\exp\left(\frac{\varepsilon u(D_2,r')}{2\Delta u}\right)}\right)} \\
&= \left(\frac{\exp\left(\frac{\varepsilon u(D_1,r)}{2\Delta u}\right)}{\exp\left(\frac{\varepsilon u(D_2,r)}{2\Delta u}\right)}\right)\cdot\left(\frac{\sum_{r'\in\mathcal{R}}\exp\left(\frac{\varepsilon u(D_2,r')}{2\Delta u}\right)}{\sum_{r'\in\mathcal{R}}\exp\left(\frac{\varepsilon u(D_1,r')}{2\Delta u}\right)}\right) \\
&\le \exp\left(\frac{\varepsilon\left(u(D_1,r)-u(D_2,r)\right)}{2\Delta u}\right)\cdot\left(\frac{\sum_{r'\in\mathcal{R}}\exp\left(\frac{\varepsilon(u(D_1,r')+\Delta u)}{2\Delta u}\right)}{\sum_{r'\in\mathcal{R}}\exp\left(\frac{\varepsilon u(D_1,r')}{2\Delta u}\right)}\right) \\
&\le \exp\left(\frac{\varepsilon}{2}\right)\cdot\exp\left(\frac{\varepsilon}{2}\right)\cdot\left(\frac{\sum_{r'\in\mathcal{R}}\exp\left(\frac{\varepsilon u(D_1,r')}{2\Delta u}\right)}{\sum_{r'\in\mathcal{R}}\exp\left(\frac{\varepsilon u(D_1,r')}{2\Delta u}\right)}\right) \\
&= \exp(\varepsilon)
\end{aligned}
$$

根据定义 9.2.1 ，指数机制满足 $\varepsilon-$ 差分隐私。

9.1.2 差分隐私的性质

定理 9.1.1.（后处理性） 令 $\mathcal{M}:D\to\mathcal{Y}$ 为满足 $(\varepsilon,\delta)-$ 差分隐私的随机算法，函数 $f:\mathcal{Y}\to\mathcal{Z}$, 则 $\mathcal{M}' \overset{\text{def}}{=} f(\mathcal{M}):D\to\mathcal{Z}$ 也满足 $(\varepsilon,\delta)-$ 差分隐私。

定理 9.1.1 的证明的如下：

证明. 因为任意随机映射都可以分解为确定性函数的凸组合，而差分隐私机制

的凸组合是差分隐私的。也就是对于相邻数据集 D_1 和 D_2，任意的 $S \subseteq \mathcal{Z}$，令 $T = \{y \in \mathcal{Y}: f(y) \in S\}$, 我们有

$$\begin{aligned}\Pr\left[f\left(M(D_1)\right) \in S\right] &= \Pr[M(D_1) \in T] \\ &\leq \exp(\varepsilon)\Pr[M(D_2) \in T] + \delta \\ &= \exp(\varepsilon)\Pr[M(D_2) \in S] + \delta\end{aligned}$$

即满足 (ε, δ) - 差分隐私。

所有不访问隐私数据本身，而仅对差分隐私算法输出做变换的操作叫做后处理，满足差分隐私的结果经过后处理后同样满足差分隐私。在设计差分隐私算法的过程中，如果直接扰动结果较为困难，可以选择对算法过程的中间量进行扰动，根据该定理，经过后续变换的输出同样能够满足差分隐私。

差分隐私的对数据集的保护作用会随着重复的输出而减小，比如不断输出一个数据集的最大值，即便它加人了噪声混淆，它们的均值也会收玫到真实的最大值。因此我们需要了解差分隐私机制的隐私保护程度在组合的过程中是如何减小的。

定理 9.1.2.（简单组成原理） 令 $\mathcal{M}_i: D \to \mathcal{R}_i$ 分别为满足 $(\varepsilon_i, \delta_i)$ - 差分隐私的随机算法，其中 $i \in [k]$，则它们的组合 $\mathcal{M}_{[k]}: D \to \prod_{i=1}^{k} \mathcal{R}_i, \mathcal{M}_{[k]}(D) = (\mathcal{M}_1(D), \ldots, \mathcal{M}_k(D))$ 满足 $\left(\sum_{i=1}^{k} \varepsilon_i, \sum_{i=1}^{k} \delta_i\right)-$ 差分隐私。

定理 9.1.2 的证明如下：

证明. 定义隐私损失

$$c_{\mathcal{M}}(o, D_1, D_2) \overset{\text{def}}{=} \log \frac{\Pr[\mathcal{M}(D_1) = o]}{\Pr[\mathcal{M}(D_2) = o]}$$

算法 $\mathcal{M}_i$ 满足 $(\varepsilon_i, \delta_i)$ - 差分隐私，则有

$$\Pr\left[c_{\mathcal{M}_i}(o, D_1, D_2) > \varepsilon_i\right] \leq \delta_i$$

进而有

$$\Pr\left[\sum_{i=1}^{k} c_{\mathcal{M}_i}(o, D_1, D_2) > \sum_{i=1}^{k} \varepsilon_i\right] \leq \sum_{i=1}^{k} \delta_i$$

即满足 $\left(\sum_{i=1}^{k}\varepsilon_i, \sum_{i=1}^{k}\delta_i\right)$ - 差分隐私。

该定理说明了对于同一个数据集重复地查询，其隐私保护程度将线性地下降。然而该定理所保证的隐私预算上界 $\left(\sum_{i=1}^{k}\varepsilon_i, \sum_{i=1}^{k}\delta_i\right)$ 是过于宽松的。高级组成原理证明了一个更紧的隐私预算上界。

定理 9.1.3.（高级组成原理） 令 $\mathcal{M}_i: D \to \mathcal{R}_i$ 均为满足 $(\varepsilon, \delta) -$ 差分隐私的随机算法，其中 $i \in [k]$，则它们的组合 $\mathcal{M}_{[k]}: D \to \prod_{i=1}^{k}\mathcal{R}_i, \mathcal{M}_{[k]}(D) = (\mathcal{M}_1(D), \ldots, \mathcal{M}_k(D))$ 满足 $(\varepsilon', k\delta + \delta')$ - 差分隐私，其中

$$\varepsilon' = \sqrt{2k\ln(1/\delta')}\varepsilon + k\varepsilon(e^{\varepsilon} - 1)$$

组成原理给出了任意差分隐私算法组合后的所能保证的隐私保护程度。结合差分隐私的后处理性和组成原理，我们得以考虑对复杂的算法加人差分隐私保护。

定理 9.1.3 的证明如下：

证明. 由隐私损失的定义，有如下引理

引理 A.1. 若 $(\mathcal{M}_1, \mathcal{M}_2, \cdots, \mathcal{M}_k)$ 均满足 $(\varepsilon, \delta) -$ 差分隐私，则有

$$\mathbb{E}\left[c_{\mathcal{M}_i}(o, D_1, D_2) \mid \mathcal{M}_{i:i-1}\right] \leq \varepsilon(e^{\varepsilon} - 1)$$

引理 A.2. 若随机变量 $(C_1, C_2, \cdots, C_k)$ 均满足 $|C_i| < \alpha$ 且 $\mathbb{E}[C_i \mid C_1 = c_1, \cdots, C_{i-1} = c_{i-1}] \leq \beta$，则

$$\Pr\left[\sum_{i=1}^{k} C_i > k\beta + z\alpha\sqrt{k}\right] \leq e^{-z^2/2}$$

在引理 A.2 中，C_i 对映隐私损失随机变量 $c_{\mathcal{M}_i}(o, D_1, D_2)$，令 $\alpha = \varepsilon, \beta =$

$\varepsilon(e^{\varepsilon}-1), z=\sqrt{2\log(1/\delta')}$，有

$$\Pr\left[\sum_{i=1}^{k} c_{\mathcal{M}_i}(o, D_1, D_2) > \sqrt{2k\log(1/\delta')}\varepsilon + k\varepsilon(e^{\varepsilon}-1)\right] \leq \delta',$$

因此组合后满足 $(\varepsilon', k\delta+\delta')-$ 的差分隐私。

9.2 差分隐私参数估计

令 $\mathcal{P}$ 表示支持于集合 $\mathcal{X}$ 上的分布族，令 $\boldsymbol{\theta}: \mathcal{P} \to \Theta \subseteq \mathbb{R}^d$ 为感兴趣的统计量，设从某概率分布 $P \in \mathcal{P}$ 中抽样出 n 个独立同分布的数据集 $\boldsymbol{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n) \in \mathcal{X}^n$ 。有了数据，我们可以通过估计量 $M(\boldsymbol{X}): \mathcal{X}^n \to \Theta$ 去估计参数 $\boldsymbol{\theta}(P)$，其中估计量 $M(\boldsymbol{X})$ 来自于所有满足 $(\varepsilon, \delta)-$ 差分隐私的估计量集合 $\mathcal{M}_{\varepsilon,\delta}$ 。估计量 $M(\boldsymbol{X})$ 的表现由它到真实值 $\boldsymbol{\theta}(P)$ 的距离衡量：令 $\rho: \Theta \times \Theta \to \mathbb{R}^+$为由在 Θ 上的范数 $\|\cdot\|$ 导出的度量，即 $\rho(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) = \|\theta_1 - \theta_2\|$，并令 $l: \mathbb{R}^+ \to \mathbb{R}^+$为一递增函数，在差分隐私约束下估计 $\boldsymbol{\theta}(P)$ 的极大极小风险定义为

$$\inf_{M \in \mathcal{M}} \sup_{\varepsilon,\delta} \mathbb{E}[l(\rho(M(\boldsymbol{X}), \boldsymbol{\theta}(P)))] \tag{9.2.1}$$

这个量刻画了最优的 (ε, δ) - 差分隐私的估计量在 $\mathcal{P}$ 上最差的表现。不同于(9.2.1)式，常用的无约束的极大极小风险

$$\inf_{M} \sup_{P \in \mathcal{P}} \mathbb{E}\left[l\left(\rho\left(M(\boldsymbol{X}), \boldsymbol{\theta}(P)\right)\right)\right] \tag{9.2.2}$$

为 “隐私的成本”。

我们主要关注于估计 d 维亚高斯分布的数据。若 d 维实数随机变量 $\boldsymbol{x}$ 服从参数为 σ 的亚高斯分布，则对于 $\boldsymbol{\mu} = \mathbb{E}\boldsymbol{x}$ 和任意向量 $\|v\|_2 = 1$ ，有

$$\mathbb{E}\exp(\lambda\langle \boldsymbol{x} - \boldsymbol{\mu}, \boldsymbol{v}\rangle) \leq \exp(\lambda^2\sigma^2), \forall \lambda \in \mathbb{R}$$

9.2.1 低维均值估计

本节我们考虑均值向量为 $\Theta=\{\boldsymbol{\mu}\in\mathbb{R}^d:\|\boldsymbol{\mu}\|_\infty<\delta\}$ 的 d 维亚高斯分布类 $\mathcal{P}(\sigma,d,\Theta)$ 。

定理 9.2.1. 令 $\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\}$ 为分布 $\mathcal{P}(d,\sigma,\Theta)$ 中独立同分布的样本，均值为 $\mathbb{E}\boldsymbol{x}_1=\boldsymbol{\mu}$ 的独立同分布的样本。若 $0<\varepsilon<1$，对于给定的 $\omega>0, n^{-1}\exp(-n\varepsilon)<\delta<n^{-(1+\omega)}$ 其中 $\log(\delta)/\log(n)$ 对于 n 不递增，$d/\log(1/\delta)\gtrsim 1$ 且 $n\gtrsim\sqrt{d/\log(1/\delta)}/\varepsilon$，则有

$$\inf_{M\in\mathcal{M}_{\varepsilon,\delta}}\sup_{\mathcal{P}(d,\sigma,\Theta)}\mathbb{E}\|M(\boldsymbol{X})-\boldsymbol{\mu}\|_2^2\gtrsim\sigma^2\left(\frac{d}{n}+\frac{d^2\log(1/\delta)}{n^2\varepsilon^2}\right)\qquad(9.2.3)$$

该极大极小下界在均值估计问题中隐私的成本：当 $d\log(1/\delta)/n\varepsilon^2\gtrsim 1$ 时隐私的成本主导了统计的风险。

定理 9.2.1 的证明如下：

证明. 引理 A.3. 令 $\boldsymbol{Y}=\{\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_n\}$ 为从给定的向量 $\boldsymbol{Z}=\{z_1,z_2,\cdots,z_m\}$ 中的放回抽样，其中 $n=km,k\geq 1$ 。存在一种对 $\boldsymbol{Z}$ 的选择，其中每个 $z_i\in\{-\sigma,\sigma\}^d, m=c_1\sqrt{d/\log(1/\delta)}\gtrsim 1$ 且 $k\asymp\log(1/\delta)/\varepsilon$，对于每个 $0<\varepsilon<1$ 的 $(\varepsilon,\delta)-$ 差分隐私的估计，有

$$\mathbb{E}\|M(\boldsymbol{Y})-\mathbb{E}\boldsymbol{y}_1\|_2\gtrsim\sigma\sqrt{d}$$

对于某个给定的常数 $\omega>0$，有 $n^{-1}\exp(-n\varepsilon)<\delta<n^{-(1+\omega)}$，且 $\log(\delta)/\log(n)$ 对于 n 不递增。

对于定理 9.3.1，只需证明极大极小下界的第二项，因为第一项为亚高斯均值估计的统计极大极小下界。

对于 $i\in[n]$，考虑有 $1-\alpha$ 的概率 $\boldsymbol{x}_i=\boldsymbol{0}\in\mathbb{R}^d$ 并有 α 的概率 $\boldsymbol{x}_i=\boldsymbol{y}_i$，其中$\boldsymbol{y}_i$ 服从引理 A.3 中的离散均匀分布。当 $n\gtrsim\sqrt{d\log(1/\delta)}/\varepsilon$，存在某个 $0<\alpha<1$ 满足 $\alpha n\asymp\sqrt{d\log(1/\delta)}/\varepsilon_\circ\boldsymbol{x}_i$ 的分布为均值 $\boldsymbol{\mu}\in\Theta$ 参数 σ 的亚高斯分布。

考虑随机指标集 $\mathcal{S}=\{i\in[n]:\boldsymbol{x}_i\neq\mathbf{0}\}$ 。对于所有 $M\in\mathcal{M}_{\varepsilon,\delta}$，我们有

$$\mathbb{E}\parallel M(\boldsymbol{X})-\boldsymbol{\mu}\parallel_2\geq\sum_{\mathcal{S}=S\subseteq[n],|S|\leq n\alpha}\mathbb{E}[\parallel M(\boldsymbol{X})-\boldsymbol{\mu}\parallel_2|\,\mathcal{S}=S]\Pr(\mathcal{S}=S).$$

对于每个固定的 S，定义 $\tilde{M}(\boldsymbol{X}_S)=\alpha^{-1}M\big(\{\boldsymbol{x}_i,i\in S\}\cup\{\mathbf{0}\}^{n-|S|}\big)$ 。我们注意到 $\tilde{M}(\boldsymbol{X}_S)$ 为关于 $\boldsymbol{X}_S=\{\boldsymbol{x}_i:i\in S\}$ 的 (ε,δ) - 差分隐私算法，因为修改 $\boldsymbol{X}_S$ 中的任何单个数据对 M 造成的隐私损失与 $\tilde{M}$ 的相同。同样有 $\boldsymbol{\mu}=\mathbb{E}\boldsymbol{x}_1=\alpha\mathbb{E}\boldsymbol{y}_1$ 。因此我们有

$$\begin{aligned}\mathbb{E}[\parallel M(\boldsymbol{X})-\boldsymbol{\mu}\parallel_2|\,\mathcal{S}=S]&\geq\mathbb{E}\big[\|\alpha\tilde{M}(\boldsymbol{X}_S)-\alpha\mathbb{E}\boldsymbol{y}_1\|_2\,|\,\mathcal{S}=S\big]\\&\geq\alpha\mathbb{E}\big[\|\tilde{M}(\boldsymbol{X}_S)-\mathbb{E}\boldsymbol{y}_1\|_2\big]\gtrsim\alpha\sigma\sqrt{d}\asymp\sigma\frac{d\sqrt{\log(1/\delta)}}{n\varepsilon}\end{aligned}$$

对于最后一个不等式，我们引用引理 A.3 中证明的下界，因为 $\boldsymbol{X}_S$ 的样本量至多为 $\alpha n\asymp\sqrt{d\log(1/\delta)}/\varepsilon$ 。证明完毕。

本节中我们展示了极大极小下界 (9.2.3) 可以由差分隐私估计量达到，从而暗示了低维均值估计中隐私成本的严格表征。

令 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$ 为从 $\mathbb{R}^d$ 上参数为 σ 的亚高斯分布中抽取的独立同分布的样本，令 $\mathbb{E}\boldsymbol{x}_1$ 为 $\boldsymbol{\mu}\in\mathbb{R}^d$ 并假设存在常数 $c=O(1)$，有 $\parallel\boldsymbol{\mu}\parallel_\infty<c$ 。我们考虑下面基于高斯机制的简单算法:

算法 9.2.1：差分隐私的均值估计

输入：数据集 $X=\{x_i\}_{i\in[n]}$，隐私参数 ε,δ，截断水平 R 。

1 计算 $\bar{\boldsymbol{X}}_R$ ：对于 $j\in[d],\bar{\boldsymbol{X}}_{R,j}=n^{-1}\sum_{i\in[n]}\Pi_R\big(x_{ij}\big)$;

2 计算 $\hat{\boldsymbol{\mu}}=\boldsymbol{X}_R+\boldsymbol{\omega}$，其中 $\boldsymbol{\omega}\sim N_d\left(\mathbf{0},\frac{4R^2d\log(1/\delta)}{n^2\varepsilon^2}\cdot\boldsymbol{I}\right)$;

输出：$\hat{\mu}$ 。

截断的步骤保证了对于相邻数据集 $\boldsymbol{X}$ 和 $\boldsymbol{X}', \|\overline{\boldsymbol{X}}_R - \overline{\boldsymbol{X}'}_R\|_2 < 2R\sqrt{d}/n$，因此使得高斯机制适用。当$R$选择得使大部分数据都被保护， $\hat{\boldsymbol{\mu}}$ 为均值 $\boldsymbol{\mu}$ 的精确估计。

定理 9.2.2. 若存在常数 $T < \infty$ 使得 $\| \boldsymbol{x} \|_\infty < T$ 以概率 1 成立，设置 $R = T$ 满足

$$\mathbb{E} \| \hat{\boldsymbol{\mu}} - \boldsymbol{\mu} \|_2^2 \lesssim \sigma^2 \left(\frac{d}{n} + \frac{d^2 \log(1/\delta)}{n^2 \varepsilon^2} \right)$$

否则，选择 $R = K\sigma\sqrt{\log n}$ 对于足够大的 K 满足

$$\mathbb{E} \| \hat{\boldsymbol{\mu}} - \boldsymbol{\mu} \|_2^2 \lesssim \sigma^2 \left(\frac{d}{n} + \frac{d^2 \log(1/\delta) \log n}{n^2 \varepsilon^2} \right)$$

其中第一中情况适用于有界支集的分布，如伯努利分布，其收玫率正对映 (9.2.3) 中的下界。第二种情况包括无界的亚高斯分布如高斯分布，其收玫率对映下界再加上 $O(\log n)$的差距。总的来说， 上界和下界表明 $(\varepsilon, \delta) -$ 的差分隐私算法对于低维均值估计的成本为 $\tilde{O}\left(\frac{d^2 \log(1/\delta)}{n^2 \varepsilon^2}\right)$ 。

定理 9.2.2 的证明如下：

证明. 通过选择 R 和 $\| \boldsymbol{\mu} \|_\infty \leq c = O(1)$，我们有

$$\| \hat{\boldsymbol{\mu}} - \boldsymbol{\mu} \|_2^2 \leq 2 \| \boldsymbol{\omega} \|_2^2 + 2 \| \overline{\boldsymbol{X}} - \boldsymbol{\mu} \|_2^2$$

只要我们取期望，就能根据 $\boldsymbol{\omega}$ 的分布和 $\boldsymbol{x}$ 的亚高斯性得到结论。

需要注意的是算法 9.2.1 缺少实用性：截断水平 R 为调优参数，需要被正确设置才能保持收玫速率；这里我们用稍简单的算法是为了理论分析隐私的成本。

9.2.2 低维稀疏均值估计

本节我们考虑估计参数为 σ 的亚高斯分布随机向量的均值，其中均值向量为 s^*- 稀疏的。具体来说，我们用均值向量的集合 $\Theta=\{\boldsymbol{\mu}\in\mathbb{R}^d:\|\boldsymbol{\mu}\|_0\leq s^*,\|\boldsymbol{\mu}\|_\infty<1\}$ 来索引这类分布，并将其记为 $\mathcal{P}(\sigma,d,s^*,\Theta)$ 。令 $\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\}$ 为从参数为 σ 的亚高斯分布中抽出的独立同分布样本，其均值向量为 $\boldsymbol{\mu}\in\Theta$ 。我们希望构建一个关于隐私参数 $(\varepsilon,\delta),d,n,s^*$ 和 σ 的函数作为 $\inf_{M\in\mathcal{M}_{\varepsilon,\delta}}\sup_{\mathcal{P}(\sigma,d,s^*,\Theta)}\mathbb{E}\|M(\boldsymbol{X})-\boldsymbol{\mu}\|_2^2$ 的下界。

定理 9.2.3. 若 $s^*=o(d^{1-\omega})$ 对于某个给定的 $\omega>0,0<\varepsilon<1$ 且 $\delta<n^{-(1+\omega)}$ 对于某个给定的 $\omega>0$，我们有

$$\inf_{M\in\mathcal{M}_{\varepsilon,\delta}}\sup_{\mathcal{P}(\sigma,d,s^*,\Theta)}\mathbb{E}\|M(\boldsymbol{X})-\boldsymbol{\mu}\|_2^2\gtrsim\sigma^2\left(\frac{s^*\log d}{n}+\frac{(s^*\log d)^2}{n^2\varepsilon^2}\right)\tag{9.2.4}$$

值得注意的是该下界中的隐私项于统计项类似，都只对数依赖于维度 d，这表明尽管有 (ε,δ) - 差分隐私的约束，高维的均值估计依然可行。

定理 9.2.3 的证明如下:

证明. 定义攻击追踪

$$\mathcal{A}_{\boldsymbol{\mu},s^*}(\boldsymbol{x},M(\boldsymbol{X}))=\langle(\boldsymbol{x}-\boldsymbol{\mu})_{\mathrm{supp}(\boldsymbol{\mu})},M(\boldsymbol{X})-\boldsymbol{\mu}\rangle$$

下面将简写为 $A_i=\mathcal{A}_{\boldsymbol{\mu},s^*}(\boldsymbol{x}_i,M(\boldsymbol{X})),A_i'=\mathcal{A}_{\boldsymbol{\mu},s^*}\left(\boldsymbol{x}_i,M(\boldsymbol{X}_i')\right)$ 。

引理 A.4. 令 $\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\}$ 为从 $\mathcal{N}_d(\boldsymbol{\mu},\sigma^2\boldsymbol{I})$ 中抽出的独立同分布样本，其中 $\boldsymbol{\mu}\in\Theta_0$ 若 $s^*=o(d^{1-\omega})$ 对于某个固定的 $\omega>0$，对于所有满足 $\forall\boldsymbol{\mu}\in\Theta,\mathbb{E}_{\boldsymbol{X}|\boldsymbol{\mu}}\|M(\boldsymbol{X})-\boldsymbol{\mu}\|_2^2=o(1)$的 $(\varepsilon,\delta)-$ 差分隐私估计量 M，以下为真:

(1)对于所有 $i\in[n]$，令 $\boldsymbol{X}_i'$ 为只对 $\boldsymbol{X}$ 替换 $\boldsymbol{x}_i$ 得到的数据集，则有

$$\mathbb{E}\mathcal{A}_{\boldsymbol{\mu},s^*}\left(\boldsymbol{x}_i,M(\boldsymbol{X}_i')\right)=0,\mathbb{E}\left|\mathcal{A}_{\boldsymbol{\mu},s^*}\left(\boldsymbol{x}_i,M(\boldsymbol{X}_i')\right)\right|\leq\sigma\sqrt{\mathbb{E}\|M(\boldsymbol{X})-\boldsymbol{\mu}\|_2^2}$$

(2)存在某个在 Θ 上的先验分布 $\pi=\pi(\mu)$ 满足

$$\sum_{i\in[n]}\mathbb{E}_{\boldsymbol{\pi}}\mathbb{E}_{\boldsymbol{X}|\boldsymbol{\mu}}\mathcal{A}_{\boldsymbol{\mu},s^*}\left(\boldsymbol{x}_i,M(\boldsymbol{X}_i')\right)\gtrsim\sigma^2 s^*\log d$$

引理 A.5. 若 M 为 $(\varepsilon,\delta)-$ 差分隐私算法，其中 $0<\varepsilon<1$ 且 $\delta>0$，则对所有的 $T>0$,有

$$\mathbb{E}A_i\leq\mathbb{E}A_i'+2\varepsilon\mathbb{E}|A_i'|+2\delta T+\int_T^\infty\Pr(|A_i|>t)$$

对于定理 9.3.3，只需证明其极大极小下界的第二项，因为第一项为稀疏均值估计的统计极大极小下界。根据引理 A.4 的第一部分和引理 A.5，对于所有 $\boldsymbol{\mu}\in\Theta$，我们有

$$\sum_{i\in[n]}\mathbb{E}_{\boldsymbol{X}|\boldsymbol{\mu}}A_i\leq 2n\varepsilon\sigma\sqrt{\mathbb{E}_{\boldsymbol{X}|\boldsymbol{\mu}}\parallel M(\boldsymbol{X})-\boldsymbol{\mu}\parallel_2^2}+2n\delta T+n\int_T^\infty\Pr(|A_i|>t)$$

对于尾部概率，因为每个 $\boldsymbol{\mu}\in\Theta$ 都被假设为满足 $\parallel\boldsymbol{\mu}\parallel_0\leq s^*$ 且 $\parallel\boldsymbol{\mu}\parallel_\infty<1$，

$$\Pr(|A_i|>t)\leq\Pr\left(\chi_{s^*}^2>t^2/4s^*\sigma^2\right)\leq\exp\left(-\frac{t^2}{c_1s^*\sigma^2}+s^*\right)$$

对某常数 c_1 成立。通过选择 $T=\sqrt{c_1}\sigma s^*\sqrt{\log(1/\delta)}$，我们有

$$\sum_{i\in[n]}\mathbb{E}_{\boldsymbol{X}|\boldsymbol{\mu}}A_i\leq 2n\varepsilon\sigma\sqrt{\mathbb{E}_{\boldsymbol{X}|\boldsymbol{\mu}}\parallel M(\boldsymbol{X})-\boldsymbol{\mu}\parallel_2^2}+c_2\sigma s^*n\delta\sqrt{\log(1/\delta)}$$

结合引理 A.4 的第二部分，有

$$\sigma^2 s^*\log d\leq\mathbb{E}_{\boldsymbol{\pi}}\sum_{i\in[n]}\mathbb{E}A_i\leq 2n\varepsilon\sqrt{\mathbb{E}_{\boldsymbol{\pi}}\mathbb{E}_{\boldsymbol{X}|\boldsymbol{\mu}}\parallel M(\boldsymbol{X})-\boldsymbol{\mu}\parallel_2^2}+c_2\sigma s^*n\delta\sqrt{\log(1/\delta)}$$

由于对于某个 $\omega>0,\delta<n^{-(1+\omega)}$，所以对于所有的 $(\varepsilon,\delta)-$ 差分隐私的 M，有

$$\mathbb{E}_{\boldsymbol{\pi}}\mathbb{E}_{\boldsymbol{X}|\boldsymbol{\mu}} \parallel M(\boldsymbol{X}) - \boldsymbol{\mu} \parallel_2^2 \gtrsim \sigma^2 \frac{(s^* \log d)^2}{n^2 \varepsilon^2}$$

由于贝叶斯风险总是最大风险的下界，因此证明完毕。

令 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ 为从 $\mathbb{R}^d$ 上参数为 σ 的亚高斯分布中抽取的独立同分布的样本，其均值为 $\mathbb{E}\boldsymbol{x}_1 = \boldsymbol{\mu} \in \mathbb{R}^d$ 并假设 $\parallel \boldsymbol{\mu} \parallel_0 \leq s^*$，存在常数 $c = O(1)$，有 $\parallel \boldsymbol{\mu} \parallel_\infty < c$ 。我们将提出差分隐私的稀疏均值 $\boldsymbol{\mu}$ 的估计算法。算法将以较高的水平选择（截断）样本均值较大的坐标,并将剩余的坐标设置为零。

算法 9.2.2：差分隐私的稀疏均值估计

输入： 数据集 $X = \{x_i\}_{i \in [n]}$，隐私参数 ε, δ，截断水平 R，稀疏性 s 。

1 计算 $\overline{\boldsymbol{X}}_R$ ：对于 $j \in [d], \overline{\boldsymbol{X}}_{R,j} = n^{-1} \sum_{i \in [n]} \Pi_R(x_{ij})$;

2 初始化 $S = \emptyset$;

3 for i in 1 to s do

4 生成 $\omega_i \in \mathbb{R}^d$，其中 $\omega_{i1}, \omega_{i2}, \cdots, \omega_{id} \overset{\text{i.i.d}}{\sim} \text{Lap}\left(\frac{2R}{n} \cdot \frac{2\sqrt{3s\log(1/\delta)}}{\varepsilon}\right)$;

5 向 S 中添加 $j^* = \arg\max_{j \in [d] \setminus S} \left|\overline{\boldsymbol{X}}_{R,j}\right| + \omega_{ij}$;

6 令 $\tilde{P}_s(\overline{\boldsymbol{X}}_R) = \overline{\boldsymbol{X}}_{RS}$;

7 生成 $\tilde{\omega}$，其中 $\tilde{\omega}_1, \cdots, \tilde{\omega}_d \overset{\text{i.i.d}}{\sim} \text{Lap}\left(\frac{2R}{n} \cdot \frac{2\sqrt{3s\log(1/\delta)}}{\varepsilon}\right)$;

输出： $\hat{\boldsymbol{\mu}} = \tilde{P}_s(\overline{\boldsymbol{X}}_R) + \tilde{\omega}_S$ 。

其中截断的步骤保证了对于相邻数据集 $\boldsymbol{X}$ 和 $\boldsymbol{X}', \|\overline{\boldsymbol{X}}_R - \overline{\boldsymbol{X}'}_R\|_\infty < 2R/n$ 。算法 9.2.2 可以得到稀疏均值 $\boldsymbol{\mu}$ 的精确估计量，如下定理所述:

定理 9.2.4. 若对于足够大的常数 K，有 $R = K\sigma\sqrt{\log n}, s \geq s^*$ 且 $s \asymp s^*$，则下式至少以概率 $1 - c_1\exp(-c_2\log n) - c_1\exp(-c_2\log d)$ 成立:

$$\mathbb{E}\,\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|_2^2 \lesssim \sigma^2\left(\frac{s^*\log d}{n} + \frac{(s^*\log d)^2\log(1/\delta)\log n}{n^2\varepsilon^2}\right)$$

对于一般的设置 $\delta = n^{-(1+\omega)}$，算法 9.3.2 的收玫率接近定理 9.3.3 的下界，差距为 $\log^2 n$ 。

定理 9.2.4 的证明如下：

证明. 引理 A.6. 令 S 和 $\{\boldsymbol{\omega}\}_{i\in[s]}$ 为算法 9.3.2 中的定义。对于所有满足 $|R_1| = |R_2|$ 的 $R_1 \subseteq S$ 和 $R_2 \subseteq S^c$ 和所有 $c > 0$ ，我们有

$$\|\boldsymbol{v}_{R_2}\|_2^2 \leq (1+c)\|\boldsymbol{v}_{R_1}\|_2^2 + 4(1+1/c)\sum_{i\in[s]}\|\boldsymbol{\omega}_i\|_\infty^2$$

对于定理 9.3.4，令 S, S^* 分别为 $\hat{\boldsymbol{\mu}}$ 和 $\boldsymbol{\mu}$ 的支集。通过选择 $R = K\sigma\sqrt{\log n}$ 和 $\|\boldsymbol{\mu}\|_\infty \leq c = O(1)$，我们有

$$\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|_2^2 \leq 2\|\tilde{\boldsymbol{\omega}}_S\|_2^2 + 2\|(\overline{\boldsymbol{X}} - \boldsymbol{\mu})_{S\cap S^*}\|_2^2 + \|\overline{\boldsymbol{X}}_{S\cap(S^*)^c} - \boldsymbol{\mu}_{S^*\cap S^c}\|_2^2$$

其最后一项有

$$\begin{aligned}\|\overline{\boldsymbol{X}}_{S\cap(S^*)^c} - \boldsymbol{\mu}_{S^*\cap S^c}\|_2^2 &= \|\overline{\boldsymbol{X}}_{S\cap(S^*)^c} - \overline{\boldsymbol{X}}_{S^*\cap S^c} + \overline{\boldsymbol{X}}_{S^*\cap S^c} - \boldsymbol{\mu}_{S^*\cap S^c}\|_2^2 \\ &\leq 4\|\overline{\boldsymbol{X}}_{S\cap(S^*)^c}\|_2 + 4\|\overline{\boldsymbol{X}}_{S^*\cap S^c}\| + 2\|(\overline{\boldsymbol{X}} - \boldsymbol{\mu})_{S^*\cap S^c}\|_2^2\end{aligned}$$

由于我们假设 $s^* = |S^*| \leq |S| = s$，结合引理 A.6，我们有

$$\|\overline{\boldsymbol{X}}_{S^*\cap S^c}\| \leq 2\|\overline{\boldsymbol{X}}_{S\cap(S^*)^c}\|_2 + 8\sum_{i\in[s]}\|\boldsymbol{\omega}_i\|_\infty^2$$

结合式（A.4）和（A.3）以及（A.2），我们有

$$\| \hat{\boldsymbol{\mu}} - \boldsymbol{\mu} \|_2^2 \leq 2\|(\overline{\boldsymbol{X}} - \boldsymbol{\mu})_{S^*}\|_2^2 + 12\|\overline{\boldsymbol{X}}_{S\cap(S^*)^c}\|_2 + 32\sum_{i\in[s]} \|\boldsymbol{\omega}_i\|_\infty^2 + 2\|\tilde{\boldsymbol{\omega}}_S\|_2^2.$$

对于前两项，由于 $|S| = s \asymp s^*$，我们有

$$2\|(\overline{\boldsymbol{X}} - \boldsymbol{\mu})_{S^*}\|_2^2 + 12\|\overline{\boldsymbol{X}}_{S\cap(S^*)^c}\|_2 \lesssim s^* \| \overline{\boldsymbol{X}} - \boldsymbol{\mu} \|_\infty^2$$

其中 $\overline{\boldsymbol{X}} - \boldsymbol{\mu}$ 为零均值参数为 $\sigma/\sqrt{n}$ 的亚高斯随机向量。根据亚高斯分布标准的尾部界限有 $\| \overline{\boldsymbol{X}} - \boldsymbol{\mu} \|_\infty^2 < C\sigma^2 \log d/n$ 以大于 $1 - c_1\exp(-c_2 \log n)$ 的概率成立。对于式（A.5）的最后两项，我们有如下引理。

引理 A.7. 考虑 $\boldsymbol{\omega} \in \mathbb{R}^k$，其中 $\omega_1, \omega_2, \cdots, \omega_k \overset{i.i.d}{\sim} \mathrm{Lap}(\lambda)$ 。对任意 $C > 1$，

$$\Pr(\| \boldsymbol{\omega} \|_2^2 > kC^2\lambda^2) \leq ke^{-C}$$
$$\Pr(\| \boldsymbol{\omega} \|_\infty^2 > C^2\lambda^2 \log^2 k) \leq e^{-(C-1)\log k}$$

在这里 $\lambda = 4R\sqrt{3s\log(1/\delta)}/n\varepsilon$ 且 $k = d$ 。则

$$32\sum_{i\in[s]} \|\boldsymbol{\omega}_i\|_\infty^2 + 2\|\tilde{\boldsymbol{\omega}}_S\|_2^2 \lesssim \sigma^2 \frac{(s^*\log d)^2 \log(1/\delta)\log n}{n^2\varepsilon^2}$$

以大于 $1 - c_1\exp(-c_2 \log d)$ 的概率成立。结合两个大概率的界即完成证明。

9.3 差分隐私假设检验

本节我们关注于高维随机分布的差分隐私假设检验。一个假设 $\mathcal{H}$ 表示为 $\mathcal{X}$ 上的一类分布，可以由单个分布或具有一定约束的分布族组成。在分布检验中，我们有两个不相交的假设 $\mathcal{H}_0$ 和 $\mathcal{H}_1$，即两个假设中不包含相同的分布。然后给定来自 $\mathcal{D}$ 的 n 个独立同分布的样本，其中 $\mathcal{D}$ 为 $\mathcal{H}_0 \cup \mathcal{H}_1$ 中的一个分布，我们的目标是确定 $\mathcal{D} \in \mathcal{H}_0$ 还是 $\mathcal{D} \in \mathcal{H}_1$ 。我们现在先给出假设检验的正式定义，然后再正式定义隐私假设检验。

定义 9.3.1.（假设检验）给定 $n \in \mathbb{N}$ 为样本量，令 $\mathcal{M}: (\mathbb{R}^d)^n \to \{0,1\}$ 为以 $X^{(1)}, \cdots, X^{(n)} \in \mathbb{R}^d$ 为输入的算法。给定不相交的假设 $\mathcal{H}_0$ 和 $\mathcal{H}_1$，我们称算法 $\mathcal{M}$ 为可以区分 $\mathcal{H}_0$ 和 $\mathcal{H}_1$ 的假设检验算法当满足下列条件：

(1)对于任意分布 $\mathcal{D} \in \mathcal{H}_0$，若每个 $X^{(i)}$ 为 $\mathcal{D}$ 中的独立同分布抽样，则对于 $\boldsymbol{X} = \left(X^{(1)}, \cdots, X^{(n)}\right)$，有 $\Pr[\mathcal{M}(\boldsymbol{X}) = 0] \geq \frac{2}{3}$，其中概率空间由抽样 $X^{(1)}, \cdots, X^{(n)} \leftarrow \mathcal{D}$和算法 $\mathcal{M}$ 的随机性构成。

(2)对于任意分布 $\mathcal{D}' \in \mathcal{H}_1$，若每个 $X^{(i)}$ 为 $\mathcal{D}'$ 中的独立同分布抽样，则 $\Pr[\mathcal{M}(\boldsymbol{X}) = 1] \geq \frac{2}{3}$。

定义 9.3.2.（差分隐私假设检验）给定 $n \in \mathbb{N}$ 为样本量，令 $\mathcal{M}: (\mathbb{R}^d)^n \to \{0,1\}$ 为以 $X^{(1)}, \cdots, X^{(n)} \in \mathbb{R}^d$ 为输入的算法。给定不相交的假设 $\mathcal{H}_0$ 和 $\mathcal{H}_1$，参数 $0 \leq \varepsilon, \delta \leq 1$ 我们称算法 $\mathcal{M}$ 可以 (ε, δ) - 隐私地区分 $\mathcal{H}_0$ 和 $\mathcal{H}_1$ 当满足下列条件：

(1)算法 $\mathcal{M}$ 可以区分 $\mathcal{H}_0$ 和 $\mathcal{H}_1$。

(2)算法 $\mathcal{M}$ 满足 (ε, δ) - 差分隐私（其中数据集为 $\mathbb{R}^d$，输出的集合为 $\{0,1\}$）。

我们的目的是设计一个算法 $\mathcal{M}$ 可以 $(\varepsilon, \delta)-$ 隐私地区分零假设 $\mathcal{H}_0$ 和备择假设 $\mathcal{H}_1$，其中样本量 n 会尽可能地小。我们同样也对构造一个有效的算法感兴趣，即运行时间是样本量 n 和维数 d 的多项式。注意到我们的隐私保护要在最坏的情况下有效，也就是对任意相邻数据集 $\boldsymbol{X}, \boldsymbol{X}' \in (\mathbb{R}^d)^n$ 成立，即便这些数据集不是从任意的分布中抽样产生的，然而我们设计的检验统计量只需要平均地适用从分布中抽样的典型数据集。因此我们希望找到一个替代的统计量 $\hat{T}$，在经典数据集上与 T 接近，而全局的敏感度更小。

定义 9.3.3.（李普希兹延拓）对于定义在域 $\mathcal{X}$ 上的函数 T，集合 $\mathcal{C} \subset \mathcal{X}, T$ 由 $\mathcal{C}$ 的一个李普希兹延拓 $\hat{T}$ 被定义为在所有数据集上满足：

(1)延拓 $\hat{T}$ 与 T 在 $\mathcal{C}$ 上相同，即对任意 $X \in \mathcal{C}$，有 $\hat{T}(X) = T(X)$。

(2)$\hat{T}$ 在所有 $\mathcal{X}$ 上的敏感度至多是 T 在 $\mathcal{C}$ 上的敏感度，即 $\Delta_{\hat{T}} \leq \max_{X,X' \in \mathcal{C}} T(X) - T(X')$ 。

本节研究的主要问题，可能也是最简单的问题，是已知协方差矩阵 Σ 的多元高斯分布的隐私同一性检验。在同一性检验中，我们的目标是区分样本只从单一假设分布 $\mathcal{N}(\mu^*, \Sigma)$中独立同分布地抽取的零假设，和样本从 $\mathcal{N}(\mu, \Sigma)$ 中独立同分布地抽取的备择假设，其中均值 μ “远” 不同于 μ^* 。对于已知协方差的高斯分布的同一性检验，我们可以先将其简化为伯努利生成分布的一致性检验。

定理 9.3.1. 存在一个映射 $F: \mathbb{R}^d \to \{\pm 1\}^d$ 和一个常数 $c > 0$ 使得下面结果成立。对于 $\mu \in \mathbb{R}^d$，用 P_μ 表示当 X 从 $\mathcal{N}(\mu, \mathbb{I}_{d\times d})$ 抽样时 $F(X)$ 的分布，则

(1)若 $\mu = \mathbf{0}$，则 $P_0 = \mathcal{U}_d$ 为 $\{\pm 1\}^d$ 上的均匀分布；

(2)P_μ 为 $\{\pm 1\}^d$ 上的生成分布，并满足 $\|P_\mu - \mathcal{U}_d\|_1 \geq c \cdot \|\mathcal{N}(\mu, \mathbb{I}_{d\times d}) - \mathcal{N}(\mathbf{0}, \mathbb{I}_{d\times d})\|$；

此外，F 在线性时间内可计算。

定理 9.3.1 的证明如下：

证明. 引理 A.8. 给定 $\tau \in (0,1]$ 令 P, Q 为 $\{\pm 1\}^d$ 上的生成分布，均值分别为 $\mu, \nu \in [-1,1]^d$，对所有 i 满足 $-1 + \tau \leq \nu_i \leq 1 - \tau$ 。则 P 和 Q 的 L_1 距离满足

$$c_\tau \| \mu - \nu \|_2 \leq \| P - Q \|_1 \leq C_\tau \| \mu - \nu \|_2$$

其中 $C_\tau, c_\tau > 0$ 为两个只依赖于 τ 的常数。并且我们可以取 $C_\tau = 1/\sqrt{\tau\left(1 - \frac{\tau}{2}\right)}$ 。

对于定理 9.3.1，映射 $F: \mathbb{R}^d \to \{\pm 1\}^d$ 是根据坐标定义的，通过设置 $F(x)_i \overset{\text{def}}{=} \operatorname{sgn}(x_i)$对于所有 $i \in [d]$；因此说明了定理 9.3.1 的第一点，计算时间效率的陈述还表明 P_μ 为一个生成分布。我们下面证明剩下的第二点。任给 $\mu \in$

$\mathbb{R}^d$，为了方便定义 $\alpha \overset{\text{def}}{=} \| \mathcal{N}(\mu, \mathbb{I}_{d\times d}) - \mathcal{N}(\mathbf{0}, \mathbb{I}_{d\times d}) \| \in [0,2]$；由于任给 $\mu, \nu \in \mathbb{R}^d$ 正态分布满足

$$\frac{1}{100} \| \mu - \nu \|_2 \leq \| \mathcal{N}(\mu, \mathbb{I}_{d\times d}) - \mathcal{N}(\nu, \mathbb{I}_{d\times d}) \|_1 \leq 9 \| \mu - \nu \|_2$$

因此我们有 $\| \mu \|_2 \geq \frac{\alpha}{9}$。对任意 $i \in [d]$，有

$$\begin{aligned} \mathbb{E}_{X\sim\mathcal{N}(\mu,\mathbb{I}_{d\times d})}[F(X)_i] &= 2\Pr[F(X)_i = 1] - 1 = 2\Pr[X_i > 0] - 1 = \mathrm{Erf}\left(-\mu_i/\sqrt{2}\right) - 1 \\ &= -\mathrm{Erf}\left(\mu_i/\sqrt{2}\right) \end{aligned}$$

其中 $\mathrm{Erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^{x} \exp(-t^2) dx$ 为误差函数。因此对于 P_μ 的均值向量 μ' 满足

$$\begin{aligned} \|\mu'\|_2^2 &= \sum_{i=1}^{d} \mathrm{Erf}\left(-\mu_i/\sqrt{2}\right)^2 \geq 0.84^2 \sum_{i=1}^{d} \min(\mu_i^2/2, 1) = 0.84^2 \min(\| \mu \|_2^2/2, d) \\ &\geq 0.84^2 \alpha^2/162 \end{aligned}$$

最后一个不等式根据 $\| \mu \|_2$ 的下界得到，这说明 $\|\mu'\|_2 > \alpha/12$。最后应用引理 A.8，存在常数 $c > 0$ 使得 $\left\|P_\mu - \mathcal{U}_d\right\|_1 \geq c\alpha$。

对于伯努利分布的一致性检验，我们采用统计量

$$T(X) = \sum_{i=1}^{d} (\bar{X}_i^2 - n)$$

我们首先基于李普希兹延拓方法给出了一种计算不有效的隐私一致性检验算法。

算法 9.3.1：李普希兹延拓检验（LIPSCHITZEXTENSIONTEST）

输入： 样本 $X = \left\{X^{(1)}, \cdots, X^{(n)}\right\}$，参数 $\varepsilon, \Delta > 0, \beta \in (0,1]$。

1 定义集合 $\mathcal{C}(\Delta) = \left\{X \in \{\pm 1\}^{n\times d} | \forall j \in [n]|, \left\langle X^{(j)}, \bar{X} \right\rangle | \leq \Delta\right\}$;

2 令 $\hat{T}(\cdot)$ 为 T 由 $\mathcal{C}(\Delta)$ 到所有 $\{\pm 1\}^{n\times d}$ 的 $4\Delta-$ 李普希兹延拓;

3 生成噪声 $r\sim \mathrm{Lap}(4\Delta/\varepsilon)$ 并令 $z\leftarrow \hat{T}+r$;

4 if $z>10n\sqrt{d}+4\Delta\log(1/\beta)/\varepsilon$ then

输出: 拒绝

输出: 接受

算法 9.3.2: 基于李普希兹延拓的隐私一致性检验

输入: 从P^n抽样的样本 $X=\{X^{(1)},\cdots,X^{(n)}\}\in\{\pm 1\}^{n\times d}$, 参数 $\varepsilon,\alpha,\beta>0$ 。

1 $M\leftarrow\lceil\log n\rceil,\varepsilon'\leftarrow\varepsilon/M,\beta\leftarrow 1/(10n)$;

2. $\Delta^{(1)}\leftarrow nd$ 且 $\Delta^{*}\leftarrow 1000\max\left(d,\sqrt{nd},\log(1/\beta)/\varepsilon'\right)\cdot\log(1/\beta)$;

3. for $m\leftarrow 1$ to $M-1$ do

4. if $\Delta^{(m)}\leq\Delta^{*}$ then

5. 令 $\Delta^{(M)}\leftarrow\Delta^{(m)}$ 并推出循环;

6. else

7. if $\mathrm{LIPSCHITZEXTENSIONTEST}\left(X,\varepsilon',\Delta^{(m)},\beta\right)$ 输出拒绝 then

输出: 拒绝

8. 令 $\Delta^{(m+1)}\leftarrow 11\left(d+\sqrt{nd}+\frac{\Delta^{(m)}}{n\varepsilon'}+\sqrt{\frac{\Delta^{(m)}}{\varepsilon'}}\right)\log\frac{1}{\beta}$;

9. 定义集合 $\mathcal{C}\left(\Delta^{(M)}\right)=\left\{X\in\{\pm 1\}^{n\times d}|\forall j\in[n]|,\left\langle X^{(j)},\bar{X}\right\rangle|\leq\Delta^{(M)}\right\}$;

10. 令 $\hat{T}(\cdot)$ 为由 $\mathcal{C}\left(\Delta^{(M)}\right)$ 到所有 $\{\pm 1\}^{n\times d}$ 的 $4\Delta^{(M)}$ - 李普希兹延拓;

11. 生成噪声 $r\sim\mathrm{Lap}\left(4\Delta^{(M)}/\varepsilon\right)$ 并令 $z\leftarrow\hat{T}+r$;

12. if $z > \frac{1}{4}n(n-1)\alpha^2$ then

输出： 拒绝

输出：接受

定理 9.3.2. （不有效的上界） 给定 $\mu^* \in \mathbb{R}^d, \Sigma \in \mathbb{R}^{d\times d}$ 为已知正定协方差矩阵。并给定参数 $0 < \alpha, \varepsilon \leq \frac{1}{2}$ 。算法 4.2 需要的样本量为

$$n = \tilde{O}\left(\frac{d^{1/2}}{\alpha^2} + \frac{d^{1/3}}{\alpha^{4/3}\cdot\varepsilon^{2/3}} + \frac{1}{\alpha\cdot\varepsilon}\right)$$

可以 $(\varepsilon, 0)-$ 隐私地区分仅由 $\mathcal{N}(\mu^*, \Sigma)$ 构成的零假设 $\mathcal{H}_0$ 和由所有 $\mathcal{N}(\mu, \Sigma)$ 构成的备择假设 $\mathcal{H}_1$ ，其中 $\sqrt{(\mu-\mu^*)^T\Sigma^{-1}(\mu-\mu^*)} \geq \alpha$ 。

定理 9.3.2 的证明如下：

证明. 首先算法 9.3.2 满足 $(\varepsilon, 0)-$ 差分隐私，因为根据李普希兹延拓的性质，$\hat{T}(X)$ 的敏感度至多为 $D \overset{\text{def}}{=} \frac{\varepsilon}{C}\alpha^2N^2$ 。

由于 X 的每列都是从 $\mathcal{N}(0, \mathbb{I})$ 中抽出的样本，则有至少 0.99 的概率有 $X \in \mathcal{C}$ 且 $\| \bar{X} \|^2 \leq Nd \pm CN\sqrt{d}$ 。因此有至少 0.99 的概率有 $\hat{T}(X) = \tilde{T}(X) \leq CN\sqrt{d}$，也就是说有至少 0.9 的概率有 $\hat{T}(X) + \mathrm{Lap}(D/\varepsilon) \leq CN\sqrt{d} + 10\frac{D}{\varepsilon}$，我们希望它最大为 $\frac{\alpha^2N^2}{2}$ 。实际上，如果 C 足够大，则 $10\frac{D}{\varepsilon} \leq \frac{1}{4}\alpha^2N^2$，因此我们只需要 $\Delta \leq D = \frac{\varepsilon}{C}\alpha^2N^2$ 和 $CN\sqrt{d} \leq \frac{1}{4}\alpha^2N^2$ 。

若 X 的每列都是从 $\mathcal{N}(\mu, \mathbb{I})$ 中抽出的样本，其中 $\alpha \leq \| \mu \| \leq 2\alpha$，则至少有 0.99 的概率有 $X \in \mathcal{C}$ 且 $\| \bar{X} \|^2 \geq Nd + N^2\alpha^2 - C(N\sqrt{d} + \alpha N\sqrt{N})$ 。因此至少有 0.99 的概率有 $\hat{T}(X) = \tilde{T}(X) \geq N^2\alpha^2 - C(N\sqrt{d} + \alpha N\sqrt{N})$，也就是说至少 0.9 的概率有 $\hat{T}(X) + \mathrm{Lap}(D/\varepsilon) \geq N^2\alpha^2 - C(N\sqrt{d} + \alpha N\sqrt{N}) - 10\frac{D}{\varepsilon}$ 我们希望它最小为

$\frac{\alpha^2N^2}{2}$ 。若 C 足够大，则 $10\frac{D}{\varepsilon} \leq \frac{1}{4}\alpha^2N^2$，因此我们只需要 $\Delta \leq D$ 和 $C(N\sqrt{d} + \alpha N\sqrt{N}) \leq \frac{1}{4}\alpha^2N^2$ 。

因此，我们的算法只要 N 足够大使得 $\Delta = 6L\left(\sqrt{Nd} + \alpha N + \frac{C}{\varepsilon}\right) \leq \frac{\varepsilon}{C}\alpha^2N^2$ 且 $C(N\sqrt{d} + \alpha N\sqrt{N}) \leq \frac{1}{4}\alpha^2N^2$ 就足够精确和隐私。由于 C 为固定的大的常数，且 L 为 N 的对数大小，所以满足

$$N \geq \tilde{O}\left(\frac{d^{1/2}}{\alpha^2} + \frac{d^{1/3}}{\alpha^{4/3} \cdot \varepsilon^{2/3}} + \frac{1}{\alpha \cdot \varepsilon}\right)$$

证明完毕。

我们给出了这种不有效的算法，其样本复杂度的上界。更简单地说，我们的目标是给定从多元高斯抽出的一些样本，隐私地区分均值固定为 μ^* 和均值远离 μ^* 的多维高斯分布。我们对远离的概念取决于量 $\sqrt{(\mu - \mu^*)^T\Sigma^{-1}(\mu - \mu^*)}$，也就是曼哈顿距离 $d_\Sigma(\mu^*, \mu)$ 。

接下来，我们将展示定理 3.2 的上界是紧的，即便算法是 $(0, \varepsilon)$ - 差分隐私的，而不是 $(\varepsilon, 0)$ - 差分隐私的。具体来说，我们证明了：

定理 9.3.3. （**下界**）任何可以 $(0, \varepsilon)$ - 隐私地区分 $\mathcal{H}_0$ 和 $\mathcal{H}_1$ 的算法，其样本复杂度至少为

$$n = \Omega\left(\frac{d^{1/2}}{\alpha^2} + \frac{d^{1/3}}{\alpha^{4/3} \cdot \varepsilon^{2/3}} + \frac{1}{\alpha \cdot \varepsilon}\right)$$

定理 9.3.3 的证明如下：

证明. 首先我们已知即便是没有隐私约束，$N \geq \Omega\left(\frac{\sqrt{d}}{\alpha^2}\right)$ 和 $N \geq \Omega\left(\frac{1}{\alpha\varepsilon}\right)$ 。因此足以证明 $N \geq \Omega\left(\frac{d^{1/3}}{\alpha^{4/3}\varepsilon^{2/3}}\right)$ 。

首先对于足够大的常数 $C, d \geq C$ 且 $\frac{d}{\alpha^2} \geq CN$ 。考虑零假设 $\mathcal{H}_0$，样本 $X =$

$\{X^{(1)},\cdots,X^{(N)}\}$ 独立同分布于 $\mathcal{N}(0,\mathbb{I})$，而在备择假设 $\mathcal{H}_1$ 中样本 $X=\{X^{(1)},\cdots,X^{(N)}\}$ 独立同分布于 $\mathcal{N}(\mu,\mathbb{I})$。因为 $\mathcal{D}$ 只支撑于 $\|\mu\|_2\geq\alpha$，所以备择假设有效。在这种条件下，我们有若算法可以 $(0,\varepsilon)-$ 地区分零假设 $\mathcal{H}_0$ 和备择假设 $\mathcal{H}_1$，则 $N\geq\Omega\left(\frac{d^{1/3}}{\alpha^{4/3}\varepsilon^{2/3}}\right)$。

最后我们考虑 $d\leq C$ 或 $\frac{d}{\alpha^2}\leq CN$ 的情况。若 $d\leq C$，则 $d=O(1)$，也就是说 $\frac{d^{1/3}}{\alpha^{4/3}\varepsilon^{2/3}}\leq O\left(\frac{1}{\alpha\varepsilon}+\frac{\sqrt{d}}{\alpha^2}\right)$。这是因为当 $d=O(1)$ 时，$\frac{1}{\alpha^{4/3}\varepsilon^{2/3}}$ 为 $\frac{1}{\alpha\varepsilon}$ 和 $\frac{1}{\alpha^2}$ 的加权平均。因此当 $d=O(1)$ 时 $\Omega\left(\frac{1}{\alpha\varepsilon}+\frac{\sqrt{d}}{\alpha^2}\right)$ 的下界也就是 $\Omega\left(\frac{d^{1/3}}{\alpha^{4/3}\varepsilon^{2/3}}\right)$ 的下界。若 $\frac{d}{\alpha^2}\leq CN$，则 $N\geq\Omega\left(\frac{d}{\alpha^2}\right)$。此外，我们知道 $N>\Omega\left(\frac{1}{\alpha\varepsilon}\right)$。可以简单的验证 $\frac{d^{1/3}}{\alpha^{4/3}\varepsilon^{2/3}}$ 是 $\frac{d}{\alpha^2}$ 和 $\frac{1}{\alpha\varepsilon}$ 的加权平均，因此我们有 $N\geq\Omega\left(\frac{d^{1/3}}{\alpha^{4/3}\varepsilon^{2/3}}\right)$，证明完毕。

由于我们为定理 9.3.2 设计的算法 9.3.2 运行时间非常慢，因此一个自然的问题是，如果算法必须在 n 和 d 的多项式时间内运行，多少样本是必要的。事实上，我们可以设计以下算法，只需要稍微多一点的样本数量，但运行效率很高。

算法 9.3.3：有效的隐私一致性检验

输入：从 P^n 抽样的样本 $X=\{X^{(1)},\cdots,X^{(n)}\}\in\{\pm1\}^{n\times d}$，参数 $\varepsilon,\delta,\alpha>0$。

令 $\bar{X}\leftarrow\sum_{j=1}^{n}X^{(j)}$;

第一步：预处理

令 $r_1\sim\mathrm{Lap}(2/\varepsilon)$ 且令 $z_1\leftarrow\max_{i\in[d]}|\bar{X}_i|+r_1$;

if $z_1>\sqrt{2n\log\frac{d}{\delta}}+\frac{2}{\varepsilon}\log\frac{1}{\delta}$ then

| **输出：拒绝**

4 令 $\tilde{X} \leftarrow \bar{X} + R$，其中 $R \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_{d\times d})$ 且 $\sigma = \frac{\sqrt{8d\log(5/4\delta)}}{\varepsilon}$;

5 令 $\Delta_\delta \leftarrow 16\left(d\log\frac{d}{\delta} + \frac{d}{n\varepsilon^2}\log^2\frac{1}{\delta} + \sqrt{nd}\sqrt{\log\frac{d}{\delta}\cdot\log\frac{n}{\delta}} + \frac{\sqrt{d}}{\varepsilon}\log\frac{1}{\delta}\sqrt{\log\frac{n}{\delta}}\right)$;

6 令 $r_2 \sim \mathrm{Lap}(1/\varepsilon)$ 且令 $z_2 \leftarrow \left|\left\{j \in [n]: \left|\langle X^{(j)}, \tilde{X}\rangle\right| > \Delta_\delta + \frac{4d}{\varepsilon}\sqrt{\log\frac{5}{4\delta}\cdot\log\frac{n}{\delta}}\right\}\right| + r_2$;

7 if $z_2 > \frac{\log(1/\delta)}{\varepsilon}$ then

|输出：拒绝

第二步：过滤

8 for $j = 1, \cdots, n$ do

9 if $\left|\langle X^{(j)}, \tilde{X}\rangle\right| > \Delta_\delta + \frac{4d}{\varepsilon}\sqrt{\log\frac{5}{4\delta}\cdot\log\frac{n}{\delta}}$ then

10 令 $\hat{X}^{(j)} \leftarrow U^{(j)}$，其中 $U^{(j)} \sim \mathcal{U}_d$;

else

令 $\hat{X}^{(j)} \leftarrow X^{(j)}$;

第三步：添加噪声和阈值

13 令 $r_3 \sim \mathrm{Lap}\left(\left(4\Delta_\delta + \frac{48d}{\varepsilon}\sqrt{\log\frac{5}{4\delta}\cdot\log\frac{n}{\delta}}\right)/\varepsilon\right)$ 且令 $z_3 \leftarrow \hat{T}(X) + r_3$;

14 if $z_3 > \frac{1}{4}n(n-1)\alpha^2$ then

| 输出：拒绝

输出：接受

不同于算法 9.3.2 那样迭代地降低敏感度，算法 9.4.3 采用预处理步骤来初步地

降低敏感度。

定理 9.3.4.（有效的上界） 算法 9.4.3 可以在 n 和 d 的多项式时间内 $(4\varepsilon, 14\delta)-$ 隐私地区分假设 $\mathcal{H}_0$ 和 $\mathcal{H}_1$，需要的样本量为

$$n = \tilde{O}\left(\frac{d^{1/2}}{\alpha^2} + \frac{d^{1/4}}{\alpha \cdot \varepsilon}\right)$$

定理 9.3.4 的证明如下：

证明. 假设对于足够大的常数 $C, \gamma N^2 \geq C\left(\frac{L\log N}{\varepsilon} + \frac{\log^2 N}{\varepsilon^2}\right)$ 有 $N^2\alpha^2 \geq C(N\sqrt{d} + \alpha N\sqrt{N})$。则对于任意相邻数据集 X, X' 对映的矩阵为 V, V' 有 $|\Pr[\mathcal{A}(V) = 0] - \Pr[\mathcal{A}(V') = 0]| \leq 4\varepsilon$, 即算法为 $(0, 4\varepsilon)$ - 差分隐私的。然后若 $X^{(1)}, \cdots, X^{(N)} \overset{i.i.d}{\sim} \mathcal{N}(0, \mathbb{I})$, 则 V 的样本量至多为 $\frac{C}{4} \cdot \frac{N\sqrt{d}+\alpha N\sqrt{N}}{R} \leq \frac{1}{4}\frac{N^2\alpha^2}{R} = \frac{1}{4}\gamma N^2$ 以至少 0.99 的概率成立，也就是说算法以至少 0.97 的概率输出 0。最后，若对于 $\alpha \leq \| \mu \|_2 \leq 2\alpha, X^{(1)}, \cdots, X^{(N)} \overset{i.i.d}{\sim} \mathcal{N}(\mu, \mathbb{I})$, 则 V 的样本量至少为 $\frac{N^2\alpha^2}{R} - \frac{C}{4}\frac{N\sqrt{d}+\alpha N\sqrt{N}}{R} \geq \frac{N^2\alpha^2}{R} - \frac{1}{4}\frac{N^2\alpha^2}{R} = \frac{3}{4}\gamma N^2$ 即算法以至少 0.97 的概率输出 1。

因此我们只需要 N 满足 $N^2\alpha^2 \geq C(N\sqrt{d} + \alpha N\sqrt{N})$ 且 $\gamma N^2 \geq C(N\sqrt{d} + \alpha N\sqrt{N})$。由于 $R = \tilde{O}(\sqrt{d})$ 并且 $L = \sqrt{N} + \alpha\frac{N}{\sqrt{d}}$, 所以

$$N = \tilde{O}\left(\frac{\sqrt{d}}{\alpha^2} + \frac{1}{\alpha^2} + \frac{d^{1/3}}{\alpha^{4/3}\varepsilon^{2/3}} + \frac{d^{1/4}}{\alpha\varepsilon} + \frac{1}{\alpha\varepsilon}\right) = \tilde{O}\left(\frac{\sqrt{d}}{\alpha^2} + \frac{d^{1/4}}{\alpha\varepsilon}\right)$$

其中 $\frac{1}{\alpha^2}$ 和 $\frac{1}{\alpha\varepsilon}$ 是较小的项并且 $\frac{d^{1/3}}{\alpha^{4/3}\varepsilon^{2/3}}$ 为 $\frac{\sqrt{d}}{\alpha^2}$ 和 $\frac{d^{1/4}}{\alpha\varepsilon}$ 加权几何平均因此可忽略。

9.4 习题

1. 简单阐述差分隐私的定义。

2. 试阐述函数机制的基本思想。

3. 试阐述差分隐私假设检验的基本思想和步骤。

4. 试阐述差分隐私算法的具有的基本性质。

5. 试考虑线性回归模型的差分隐私估计的步骤，并结合模拟数据和实际数据给出差分隐私参数估计的效果。

6. 试考虑如何做到差分隐私参数估计的隐私保护性和可用性之间的权衡。