arXiv:2301.09024v1 [math.ST] 21 Jan 2023

各向异性高斯分布均值与协方差估计的统计最优稳健方法

ArshakMinasyan* 和 Nikita Zhivotovskiy†

2023年1月24日

摘要

假设 $X_1, \dots, X_N$ 是由 N 个相互独立的高斯向量构成的样本，其中有 ε 污染，向量取值于 R^d，均值为 μ，协方差为 Σ。在强 ε 污染模型中，我们假设对手将原始高斯样本中比例为 ε 的向量替换成任意其他向量。我们证明，存在一个均值估计量 $\widehat{\mu}$，其满足以下形式的界，且该结论以至少 $1-\delta$ 的概率成立：

$$\|\widehat{\mu}-\mu\|_2 \leqslant c\left(\sqrt{\frac{\operatorname{Tr}(\Sigma)}{N}}+\sqrt{\frac{\|\Sigma\|\log(1/\delta)}{N}}+\varepsilon\sqrt{\|\Sigma\|}\right),$$

其中，$c>0$ 是一个绝对常数，$\|\Sigma\|$ 表示 Σ的算子范数。在相同的污染高斯分布设定下，我们构造了协方差矩阵 Σ 的一个估计量 $\widehat{\Sigma}$，该估计量以至少 $1-\delta$的概率满足：

$$\left\|\widehat{\Sigma}-\Sigma\right\| \leqslant c\left(\sqrt{\frac{\|\Sigma\|\operatorname{Tr}(\Sigma)}{N}}+\|\Sigma\|\sqrt{\frac{\log(1/\delta)}{N}}+\varepsilon\|\Sigma\|\right).$$

这两个结果在乘法常数因子意义下都是最优的。尽管近年来人们对稳健统计的兴趣显著增加，但在典型的高斯分布情形下，同时获得两个与维数无关的界仍是一个未解决的问题。事实上，此前已知的若干结果要么依赖于维度，且要求 Σ 接近单位矩阵；要么对污染水平 ε的依赖并非最优。

作为分析的一部分，我们推导了高斯分布、折叠正态分布和卡方分布的中心次序统计量的尖锐集中不等式。

1 稳健多元均值估计

稳健统计学中最早得到严格研究的问题，大概是污染高斯分布的均值（或位置参数）估计 [Hub64]。这一问题的自然延伸，是研究数据受到恶意对手污染时的多元高斯均值估计问题。对于未受污染的数据，著名的 Borell、Tsirelson-Ibragimov-Sudakov [Bor75、TIS76] 高斯集中不等式给出了样本均值性能的尖锐非渐近界：若 $X_1 \dots,$ X_N是取值于$\mathbb{R}^d$ 的相互独立高斯随机向量，其均值为 μ、协方差为 Σ，那么，以至少 $1-\delta$的概率，

$$\left\|\frac{1}{N}\sum_{i=1}^N X_i-\mu\right\|_2 \leqslant \sqrt{\frac{\operatorname{Tr}(\Sigma)}{N}}+\sqrt{\frac{2\|\Sigma\|\log(1/\delta)}{N}}. \tag{1}$$

*CREST、ENSAE、巴黎综合理工学院，arshak.minasyan@ensae.fr
†加州大学伯克利分校统计学系，zhivotovskiy@berkeley.edu

我们要解决的问题是：当恶意对手污染了所有观测值中的 ε 比例，且对手同时知晓“干净”样本和我们的估计量时，如何估计高斯随机向量的均值 μ 。我们聚焦于多元情形，不对受污染的观测值作任何假设。这种模型通常称为强污染模型 [DK19, DKP20]（关于我们模型的精确定义，见 [DM22, Definition 1]），其中包含许多已知的污染模型，例如 Huber 的 ε-污染模型 [Hub64]。这样的高斯向量污染样本称为 ε-污染样本。显然，即使只有一个离群值，样本均值也可能遭到破坏，因此我们将为另一种估计量给出式（1）的类似结果。尽管近年来稳健统计学取得了重大进展（详见近期综述，其中对统计学 [LM19a] 和算法 [DK19] 方面均有详细讨论），但在允许 ε-强对抗性污染时，仍没有高斯集中不等式（1）的尖锐类似结果。尽管高斯情形在历史上是稳健统计理论的起点，但即便从信息论视角出发、不考虑计算方面的问题，我们要研究的问题仍未解决。在陈述第一个界之前，我们还需要给出一个定义。给定协方差矩阵 Σ，其有效秩定义为

$$\mathbf{r}(\Sigma) = \frac{\mathrm{Tr}(\Sigma)}{\|\Sigma\|},$$

其中，$\mathrm{Tr}(\Sigma)$ 是矩阵 Σ 的迹。显然，对于 $d \times d$ 协方差矩阵 Σ，$1 \leqslant \mathbf{r}(\Sigma) \leqslant d$；但若数据分布具有各向异性，且由若干主方向决定，则该值可能小得多。现在我们可以给出第一个界。

定理 1（高斯情形下的稳健均值估计）。存在绝对常数 c_1，$c_2 > 0$，使得以下结论成立。假设 $X_1, \ldots, X_N$ 是一个由均值为 μ 、协方差矩阵为 Σ 的高斯随机向量构成的 ε- 污染样本，取自 $\mathbb{R}^d$。令 $\varepsilon < c_1$，则存在一个估计量 $\widehat{\mu}$，使其以至少 $1-\delta$ 的概率满足：

$$\|\widehat{\mu} - \mu\|_2 \leqslant c_2 \sqrt{\|\Sigma\|} \left(\sqrt{\frac{\mathbf{r}(\Sigma)}{N}} + \sqrt{\frac{\log(1/\delta)}{N}} + \varepsilon \right).$$

相差一个乘法常数因子以内，没有任何高斯均值估计量能取得更好的表现。

尽管我们的结果具有一般性是其主要优势，但我们暂且聚焦于各向同性情形，即 $\Sigma = I_d$。在这种情形下，如果对手至多污染样本中的 $O(\sqrt{dN})$ 个元素，那么即使在数据未受污染的设定下，我们仍能获得样本均值的最优性能（1）。这种依赖关系与更常见的 $\sqrt{\varepsilon}$- 污染水平依赖关系形成对比；已知在二阶矩假设下可以达到后者 [LM21, DKP20]。特别是，后者的结果只有在离群值数量不超过 $O(d)$时，才能保持最优性能。

我们分析的另一个有趣之处在于，我们对样本量不作任何假设。相比之下，现有那些在各向同性情形下达到最优界的估计量 [CGR18、DK19] 要求 $N \geqslant c(d + \log(1/\delta))$，或者 $N \geqslant cd\varepsilon^{-2}$ 如 [DL22], 中那样$c > 0$ 其中 6是某个绝对常数。我们将在第 6节遇到一些较弱的假设，但这只发生在为我们的估计量调整一个实值参数时。

在各向异性次高斯分布的均值估计问题中，目前已知的最优界由 Lugosi 和 Mendelson 给出 [LM21]。这些作者提出了截尾均值估计量的一个多变量版本，可达到以下误差率

$$\|\widehat{\mu}_{\mathrm{LM}} - \mu\|_2 \leqslant c \sqrt{\|\Sigma\|} \left(\sqrt{\frac{\mathbf{r}(\Sigma)}{N}} + \sqrt{\frac{\log(1/\delta)}{N}} + \varepsilon \sqrt{\log\left(\frac{1}{\varepsilon}\right)} \right).$$

Dalalyan 和本文第一作者 [DM22] 也针对另一种估计量给出了相同的误差率，但需要额外假设 Σ 已知，不过他们使用的是计算高效算法。人们可能会认为，额外出现的 $\sqrt{\log(1/\varepsilon)}$ 项是 [LM21] 和 [DM22]分析中的一种伪影。事实上并非如此；已知对于次高斯分布，这一项是必不可少的 [CDG19, LM21]。像 [LM21] 那样对观测值进行截尾，可能会丢失高斯分布的一些特有性质。因此，我们的结果需要采用不同的估计量和不同的分析方法。额外出现的 $\sqrt{\log(1/\varepsilon)}$ 项从计算角度来看也很有意思。具体而言，[DKS17] 的作者认为，对于任何多项式时间统计查询算法，误差界中的因子 $\sqrt{\log(1/\varepsilon)}$ 都是不可避免的（另见 [HL19]）。我们注意到，在丰富的稳健统计文献中，已有若干多项式时间估计量被提出，它们对污染水平 [LRV16, DKK+17, DK19, DM22, BMD22]的依赖为 $\varepsilon\sqrt{\log(1/\varepsilon)}$。

值得一提的是，有若干结果表明，在各向同性情形下，ε的线性依赖可以实现，此时 Σ 为单位矩阵（或接近单位矩阵）。关于 Tukey中位数、方向相关中位数以及 Stahel-Donoho均值中位数的分析，分别可参见 [CGR18、DK19、DL22]。除了各向同性假设在处理真实数据时相当严格之外，VC 类型论证或球面覆盖论证也是现有依赖维数的估计量分析中不可或缺的部分。遗憾的是，这些论证无法帮助我们证明定理1中的与维数无关的界。众所周知，椭球的高斯复杂度界（对应于我们结果中的$\sqrt{\mathrm{Tr}(\Sigma)}$ 项）既无法通过椭球覆盖得到，也无法通过 Dudley积分或VC 类型论证得到。在第2节讨论的高斯协方差估计问题中，这一特点尤为突出。关于这些问题的深入分析，请参阅 Talagrand 专著中的第2.5章[Tal14] 。

我们的分析从一维情形下高斯分布的样本中位数这一广为人知的性质出发。下文中，我们用 Med(·) 表示样本中位数。若 $X_1, \dots, X_N$ 是由独立标准高斯变量构成的一个 ε-污染样本，其均值为 μ、方差为 σ^2，那么，以至少 $1-\delta$ 的概率，

$$|\mathrm{Med}(X_1, \dots, X_N) - \mu| \leqslant c\sigma \left(\sqrt{\frac{\log(1/\delta)}{N}} + \varepsilon \right),$$

当 $N \geqslant c\log(1/\delta)$ 且 ε 小于某个绝对常数时。推广到高维情形时，我们不采用 Tukey 中位数（其精确分析仅在各向同性情形下已知） [CGR18], , 也不采用 [DL22], 中的 Stahel-Donoho 型估计量，而是基于我们所称的平滑中位数估计量。

设 $x_1, \dots, x_N$ 是 $\mathbf{R}^d$ 中的一组向量，且 $\xi = (\xi_1, \dots, \xi_N)$ 是 $\mathbf{R}^N$ 中的一个零均值随机向量，其协方差矩阵 H 由 $H_{i,j} = \beta^{-1}\langle x_i, x_j\rangle$ 给出，其中 $i, j = 1, \dots, N$。也就是说，H 与原始数据的格拉姆矩阵成正比。这里，$\beta > 0$ 是满足 $\mathbf{r}(\Sigma)/10 \leqslant \beta \leqslant 10\mathbf{r}(\Sigma)$ 的任意整数（由统计学家选取）。对于任意方向 $v \in S^{d-1}$，我们关注下述量，并称之为平滑中位数：

$$\mathrm{SmoothMed}_v(x_1, \dots, x_N) = \mathbf{E}_\xi \,\mathrm{Med}(\langle x_1, v\rangle + \xi_1, \dots, \langle x_N, v\rangle + \xi_N).$$

注意到 $\mathrm{SmoothMed}_v(x_1, \dots, x_N)$ 是 $x_1, \dots, x_N$和 v的函数。定理 1 中的估计量形式简单。给定一个 ε污染样本 $X_1, \dots, X_N$，我们设

$$\widehat{\mu} = \operatorname*{arg\,min}_{\nu \in \mathbb{R}^d} \sup_{v \in S^{d-1}} \left| \mathrm{SmoothMed}_v(X_1, \dots, X_N) - \langle \nu, v\rangle \right|. \tag{2}$$

注 1. 在第 6 节中，我们讨论如何仅根据污染样本选择满足以下条件的整数 β $\mathbf{r}(\Sigma)/10 \leqslant \beta \leqslant 10\mathbf{r}(\Sigma)$ 。重要的是，在调整该参数时，我们避免使用样本拆分方法。

从实际角度来看，我们的估计量的计算复杂度随维度呈指数增长。这是所有现有估计量的典型局限，这些估计量在 ε 高斯情形[1]下与污染水平呈线性依赖。不过，当维度 d 足够小时，可以用适当的 ε-网替代对球面 S^{d-1} 的计算，并利用蒙特卡洛抽样技术在该网的所有元素上对平滑积分进行一致近似。

平滑中位数的出现源于能够保证我们的界具有无维数依赖性的证明技术。我们的方法采用所谓的*PAC*-贝叶斯引理，O. Catoni及其合作者率先将其应用于 [AC11、Cat16、CG17] 重尾设定下的均值/协方差估计/线性回归。我们的应用进一步发展了这些技术，但将其用于对抗性污染情形。第4节还进行了补充讨论。

符号说明。全文中， c、 c_1、 c_2等表示绝对常数，这些常数可能随行而变化。对于两个半正定矩阵 A和 B，若 $B-A$为半正定矩阵，我们记作 $A \preceq B$。符号 $\|\cdot\|$根据上下文表示矩阵的算子范数或向量的欧几里得范数。令S_+^d 表示 $d \times d$ 半正定矩阵的集合。符号 I_d 表示单位矩阵 $d \times d$ 。我们用$\mathbf{1}$（A）表示事件 A 的示性函数。对于任意整数 N， $[N]$是集合 $\{1, \ldots, N\}$的简写。对于随机变量 Y和 $\alpha \in [1,2]$，其奥利奇范数 ψ_α 定义如下

$$\|Y\|_{\psi_\alpha} = \inf\{c > 0 : \mathbf{E} \exp(|Y|^\alpha / c^\alpha) \leqslant 2\}.$$

按照标准约定，我们称 $\|\cdot\|_{\psi_2}$ 为次高斯范数，称 $\|\cdot\|_{\psi_1}$ 为次指数范数。令 $\mathcal{KL}(\rho, \gamma) = \int \log\left(\frac{d\rho}{d\gamma}\right) d\rho$表示测度 ρ 与 γ 之间的库尔贝克-莱布勒散度，其中 $\rho \ll \gamma$。记号 $\rho \ll \gamma$ 表示测度 ρ 关于测度 γ 绝对连续。我们定义次序统计量。给定一组实数 $x_1, \ldots, x_n$，令 $x_{(1)}, \ldots, x_{(n)}$ 表示它们的非递减重排。也就是说， $x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(n)}$。对于 $\alpha \in [0,1]$，假设 αn 是一个整数，我们令

$$\mathrm{Quant}_\alpha(x_1, \ldots, x_n) = x_{(\alpha n)}.$$

特别地，为简单起见，假设 n 是奇数，则样本中位数为

$$\mathrm{Med}(x_1, \ldots, x_n) = x_{((n+1)/2)}.$$

相关文献。稳健统计是一个发展成熟的领域，近年来已有多部阐释性著作出版。在我们的语境中，最相关的是 Lugosi 和 Mendelson 的综述[LM19a], 以及 Diakonikolas 和 Kane 的综述 [DK19]。稳健统计的一些经典参考文献包括教材[HRRS80、Hub81、RL05]。讨论协方差估计时，我们参考综述 [KMR+19] ，其重点是重尾分布。我们还提及了几篇近期关于协方差估计的论文[CGR18、MW22、AZ22、OR22], ，其重点是对抗性污染。

与定理 1中采用欧几里得范数不同，一些作者关注的是马氏范数。也就是说，目标是构造一个估计量 $\widehat{\mu}$，使得 $(\widehat{\mu}-\mu)^\mathsf{T}\Sigma^{-1}(\widehat{\mu}-\mu)$ 是

[1]请注意，现有的估计器 此类估计器会导致依赖于维度的 界限。

以高概率足够小。看来，关于该范数的界必然依赖于维度，而简单的 VC 型/球面覆盖论证就足以得到最优收敛速率 [DL22]。类似的观察对协方差估计问题同样成立。我们关注算子范数，在这种情况下，分析可以得到与维度无关的界。对于文献中通常分析的（加权）弗罗贝尼乌斯范数，情况则并非如此。

2 协方差估计

现在我们转向一个更具挑战性的协方差估计问题。为简单起见，我们假设未污染分布的均值为零。首先，我们需要给出不允许任何污染时不等式（1）的一个精确类比。直到最近，Koltchinskii 和 Lounici [KL17]才证明了这样的结果。他们的分析基于二次过程的泛型链。这种非平凡的方法源于一个难点：难以用有效秩 d替代 $\mathbf{r}(\Sigma)$。下面表述他们的结果。假设 $Y_1,\dots,Y_N$ 是 $\mathbb{R}^d$ 中相互独立的零均值高斯向量，其协方差为 Σ。存在绝对常数 $c_1, c_2 > 0$使得，以至少 $1-\delta$ 的概率，

$$\left\| \frac{1}{N}\sum_{i=1}^{N} Y_i Y_i^{\mathsf{T}} - \Sigma \right\| \leqslant c_1 \|\Sigma\| \left(\sqrt{\frac{\mathbf{r}(\Sigma)}{N}} + \sqrt{\frac{\log(1/\delta)}{N}} \right), \tag{3}$$

前提是 $N \geqslant c_2(\mathbf{r}(\Sigma)+\log(1/\delta))$。在允许对抗性污染时，目前已知最精确的无维度依赖结果可由 Abdalla 和本文第二作者的界推出 [AZ22]。他们的工作提出了一种基于截尾均值的估计量，其收敛速率为

$$\left\| \widehat{\Sigma}_{\mathrm{AZh}} - \Sigma \right\| \leqslant c_1 \|\Sigma\| \left(\sqrt{\frac{\mathbf{r}(\Sigma)}{N}} + \sqrt{\frac{\log(1/\delta)}{N}} + \varepsilon \log\left(\frac{1}{\varepsilon}\right) \right), \tag{4}$$

只要 $N \geqslant c_2(\mathbf{r}(\Sigma) + \log(1/\delta))$。上述界对任意次高斯分布均成立，且一般而言无法改进。不过，与均值估计的情形类似，我们预期在高斯情形下，关于污染参数 ε 的依赖关系会更优。我们还注意到，基于中位数均值估计量 4的界（ [MZ20] ）的较简单版本，对污染水平的依赖仅为 $\sqrt{\varepsilon}$。

备注 2。 顺便一提，在无污染设定下，若希望式（3）中的首项常数达到尖锐值，则结合 Han [Han22] 的最新结果与 [KZ20]中推导的二阶集中不等式，可以获得近乎最优的性能。在 [Zhi21]中可以找到次高斯情形下具有显式常数的类似界。

在我们的分析中，我们首先作出一些额外假设。我们的估计量依赖于一些额外参数，这些参数可以仅根据观测到的 ε污染样本进行预估计。关于预估计程序的仔细分析留待第 6 节讨论。本节其余部分中，我们假设可以获得以下量。

1. 存在一个整数 β 和一个实数 ω，分别满足

$$\frac{1}{10}\mathbf{r}(\Sigma) \leqslant \beta \leqslant 10\mathbf{r}(\Sigma), \quad \text{and} \quad \frac{1}{10}\|\Sigma\| \leqslant \omega \leqslant 10\|\Sigma\|.$$

2. 设 H 是一个已知的半正定矩阵。假设我们知道一个满足以下条件的实数 $\alpha = \alpha(H)$

$$|\alpha - \operatorname{Tr}(\Sigma H)| \leqslant c \operatorname{Tr}(\Sigma H) \left(\sqrt{\frac{\mathbf{r}(\Sigma) + \log(1/\delta)}{N}} + \varepsilon \right), \tag{5}$$

其中 $c > 0$ 是某个绝对常数。

3. 我们可以访问一个满足以下条件的半正定矩阵 G

$$\frac{1}{10}\Sigma \preceq G, \quad \text{and} \quad \operatorname{Tr}(G) \leqslant 10 \operatorname{Tr}(\Sigma). \tag{6}$$

除矩阵 G 外，我们只需调整实值参数。这通常可以在最小假设 $N \geqslant c(\mathbf{r}(\Sigma) + \log(1/\delta)$ 下完成，其中 $c > 0$ 为某个绝对常数。注意，假设 $\Sigma \preceq 10G$ 并不意味着 Σ 在算子范数下接近 G。同时，该假设要求对 Σ 的最小奇异值进行一定控制。特别地，我们证明，只要 $N \geqslant c(d + \log(1/\delta)$，就总能基于污染数据高效地构造这样的矩阵 G，同时保持我们上界的无维数依赖性。此外，我们的保证似乎对矩阵 G 的选择是统一的。对于满足 (6) 的任何可接受的 G，我们都可以在同一数据上多次重新运行估计量，而不会影响其性能。我们将在第6节对此进行正式讨论。

定理 2（高斯情形下的稳健协方差估计）。存在绝对常数c_1，$c_2 > 0$ 使得以下结论成立。假设 $X_1, \ldots, X_N$ 是由零均值高斯向量组成的 ε-污染样本，取值于 $\mathbb{R}^d$，协方差为 Σ。令 $\varepsilon < c_1$，则存在一个估计量 $\widehat{\Sigma} = \widehat{\Sigma}_{\alpha,\ \beta,\ \omega,\ G,\ \varepsilon}(X_1, \ldots, X_N)$，以至少 $1 - \delta$ 的概率满足，

$$\left\|\widehat{\Sigma} - \Sigma\right\| \leqslant c_2 \|\Sigma\| \left(\sqrt{\frac{\mathbf{r}(\Sigma)}{N}} + \sqrt{\frac{\log(1/\delta)}{N}} + \varepsilon\right).$$

在相差乘法常数因子的意义下，没有任何协方差矩阵估计量能做得更好.

现在我们已经可以定义我们的估计量了。

我们首先构造如下分布。对于任意 $v \in S^{d-1}$，令 ρ_v 为 $\mathbf{R}^d$ 上的一个分布，其密度 f_v 为

$$f_v(x) = \frac{1}{p(2\pi\beta^{-1})^{d/2}} \exp\left(-\frac{\beta\|x - v\|^2}{2}\right) \mathbf{1}\left(\|G^{1/2}(x - v)\| \leqslant 100\sqrt{\omega}\right).$$

这里 $p > 0$ 是归一化因子。假设 θ 是服从 ρ_v 的随机向量。令 $H = \mathbf{E}_{\rho_v}(\theta - v)(\theta - v)^{\mathsf{T}}$ 为 ρ_v 的协方差矩阵（由 ρ_v 关于 v 的对称性可知，矩阵 H 不依赖于 v）。对于特意选取的某个绝对常数 $c > 0$ 和 $\alpha = \alpha(H)$，定义集合

$$\mathcal{H} = \left\{\Gamma \in \mathbb{S}_+^d : |\operatorname{Tr}(\Gamma H) - \alpha| \leqslant c \operatorname{Tr}(\Gamma H)\left(\sqrt{\frac{\mathbf{r}(G) + \log(1/\delta)}{N}} + \varepsilon\right); \Gamma \preceq 10G; \|\Gamma\| \leqslant 10\omega\right\}. \tag{7}$$

对于 ε-污染样本 $X_1, \ldots, X_N$，我们的估计量定义如下：

$$\widehat{\Sigma} = \operatorname*{arg\,min}_{\Gamma \in \mathcal{H}} \sup_{v \in S^{d-1}} \mathbf{E}_{\rho_v} \left|\operatorname{Med}\left(|\langle X_1, \theta\rangle|, \ldots, |\langle X_N, \theta\rangle|\right) - \Phi^{-1}(3/4)\sqrt{\theta^{\mathsf{T}}\Gamma\theta}\right|.$$

该估计量是我们平滑中位数估计量的一个更复杂的版本。首先，我们不再使用高斯平滑测度，而是将此分布限制在一个椭圆集$\{x \in \mathbb{R}^d : \ \|G^{1/2}(x-v)\| \leqslant 100\sqrt{\omega}\}$内。其次，我们需要限制输出矩阵的特征值，并引入集合 $\mathcal{H}$。我们的估计量与最小化所谓的 中位数绝对偏差 有关（相关定义见 [DG92] ）。定理 2的证明利用了这样一个事实： $|\langle X, \theta\rangle|$ 的分位数与相应的方差紧密相关。这一点体现在我们的估计量定义中出现的项 $\Phi^{-1}(3/4)$ 上。

3 样本分位数的集中不等式

本节针对从若干正则分布中抽取的独立同分布观测值，推导其分位数的次高斯和次指数集中不等式。样本分位数的分析是统计学中的一个经典问题。Kolmogorov 的早期工作 [Kol31] 聚焦于证明某些对称分布的样本中位数的中心极限定理。随后，研究重点转向该极限分布的显式展开 [Bur97]。另一条研究路线则着重考察次序统计量分布的显式公式。关于次序统计量，可参阅 David 和 Nagaraja 的专著 [DN04] ；关于极值理论——这一研究样本中最小值和最大值性质的领域——可参阅 De Haan 和 Ferreira 的专著[DHF07] 。最后，许多作者着重分析样本分位数的 Bahadur 表示（例如，参见 [Bah66、Kie67], [Sha03，定理 5。11]）。遗憾的是，无论是样本分位数分布的精确表达式，还是各种渐近展开，都无法导出我们所关注的精确集中不等式。

关于样本分位数的集中不等式，目前所知不多。Shao 的专著中给出了一些显式的非渐近界 [Sha03，第 5 节。3], 其中通过将问题归约为伯努利随机变量的集中性来处理。Boucheron 和 Thomas 的研究中也给出了一些相关的集中不等式 [BT12], ，不过其界对我们的目的而言不够精确。具体而言，这些作者为高斯分布的样本中位数给出了一个次指数集中不等式，而我们的结果将导出次高斯集中不等式。关于样本分位数的较新结果受到稳健统计学问题的启发。事实上， [CGR18、ABM19] 中的分析为样本分位数提供了一些精确界，不过，现有的界仅能在偏离均值足够小时推出次高斯集中性。另一类结果由 Bobkov 和 Ledoux 提出[BL19]。他们的结果为对数凹分布提供了精确的集中不等式（请回想一下，高斯分布是对数凹的），但由于其一般性，只能得到次指数尾部[2]。

不过，我们的方法相当简单，据我们所知，文献中尚未明确采用过。证明样本分位数的集中不等式时，我们考虑两种情形。对于较小的偏差，我们利用密度函数的正则性，并按照 [Sha03、CGR18、ABM19、DK19、Xia19], 中的做法，将问题化归为伯努利随机变量的集中性；对于较大的偏差，则利用分布的次高斯／次指数尾部。由此可得到所需的尖锐集中不等式。第 7 节讨论了对我们分析的一些直接推广。

在给出第一个集中不等式之前，先回顾一下，标准高斯分布的累积分布函数记为 $\Phi(\cdot)$。其逆函数记为 $\Phi^{-1}(\cdot)$。

引理 1 （高斯分位数的集中性）。 存在绝对常数 c_1， $c_2 > 0$ 使得以下结论成立。设 $\varepsilon \in [0, 1/4]$。不失一般性，假设 $(1/2 \pm \varepsilon) N$ 是

[2]对于实数轴上的均匀分布这一特殊情形，Bobkov 和 Ledoux [BL19] 给出了一个次高斯集中不等式，该不等式适用于均匀分布的所有次序统计量，见 [0, 1]。

整数。令 Y_1，……，Y_N 为相互独立的标准高斯随机变量样本。于是，对于任意 $t \geqslant 0$,

$$\Pr(|Y_{((1/2\pm\varepsilon)N)} - \Phi^{-1}(1/2 \pm \varepsilon)| \geqslant t) \leqslant 2\exp(-c_1 N t^2).$$

等价地,

$$\left\|Y_{((1/2\pm\varepsilon)N)} - \Phi^{-1}(1/2 \pm \varepsilon)\right\|_{\psi_2} \leqslant \frac{c_2}{\sqrt{N}}.$$

证明。我们只分析分位数 $Y_{((1/2+\varepsilon)N)}$，因为对 $Y_{((1/2-\varepsilon)N)}$ 的分析相同。我们将尾部的两部分分别进行分析。首先，我们证明

$$\Pr(Y_{((1/2+\varepsilon)N)} - \Phi^{-1}(1/2 + \varepsilon) \geqslant t) \leqslant \exp(-c_1 N t^2).$$

此分析也分为两种情形。对于某个绝对常数 $C > 0$，我们先证明上述不等式在 $0 \leqslant t \leqslant C$ 时成立，然后再处理 $t \geqslant C$的情形。在第一种情形中，我们采用将问题标准地化归为二项分布尾部的方法（类似计算见 [Sha03，定理 5。9] 以及 [Xia19]）。设 Z_1，……，Z_N 为期望为 $p = 1 - \Phi(s)$ 的相互独立伯努利随机变量，其中某个固定的 $s \in \mathbb{R}$。我们有

$$\Pr(Y_{((1/2+\varepsilon)N)} \geqslant s) = \Pr\left(\sum\nolimits_{i=1}^{N} Z_i > (1/2 - \varepsilon)N\right).$$

令 $s = \Phi^{-1}(1/2 + \varepsilon) + t$，并利用 $\mathrm{E}Z_i = 1 - \Phi(\Phi^{-1}(1/2 + \varepsilon) + t)$ 得到,

$$\Pr(Y_{((1/2+\varepsilon)N)} \geqslant \Phi^{-1}(1/2 + \varepsilon) + t) = \Pr\left(\frac{1}{N}\sum_{i=1}^{N} Z_i - \mathbf{E}Z_i > \Phi(\Phi^{-1}(1/2 + \varepsilon) + t) - \frac{1}{2} - \varepsilon\right).$$

记 $\varphi_+(t) = \Phi(\Phi^{-1}(1/2 + \varepsilon) + t) - 1/2 - \varepsilon$，由 Hoeffding 不等式应用于相互独立的伯努利随机变量可得，只要 $\varphi_+(t) \geqslant 0$,

$$\Pr(Y_{((1/2+\varepsilon)N)} - \Phi^{-1}(1/2 + \varepsilon) \geqslant t) \leqslant \exp\left(-2N(\varphi_+(t))^2\right). \tag{8}$$

我们来求函数 φ_+ 的下界。根据密度公式，可得

$$\begin{aligned}
\varphi_+(t) &= \Phi(\Phi^{-1}(1/2 + \varepsilon) + t) - \Phi(\Phi^{-1}(1/2 + \varepsilon)) \\
&\geqslant \frac{t}{\sqrt{2\pi}} \exp\left(-(\Phi^{-1}(1/2 + \varepsilon) + t)^2/2\right) \\
&\geqslant \frac{t}{\sqrt{2\pi}} \exp\left(-(\Phi^{-1}(3/4) + t)^2/2\right).
\end{aligned}$$

我们将这些计算与 t 取大值时的尾部行为结合起来。由于 $\varepsilon \leqslant 1/4$，我们需要所有观测值中至少有 $N/4$ 个（不失一般性，假设其为整数）大于 $\Phi^{-1}(1/2 + \varepsilon) + t$。这可以如下控制

$$\begin{aligned}
\Pr(Y_{((1/2+\varepsilon)N)} - \Phi^{-1}(1/2 + \varepsilon) \geqslant t) &\leqslant \binom{N}{N/4} \left(\Pr(Y_1 \geqslant \Phi^{-1}(1/2 + \varepsilon) + t)\right)^{N/4} \\
&\leqslant 2^N \left(\Pr(Y_1 \geqslant t)\right)^{N/4} \\
&\leqslant 2^N \exp(-Nt^2/8) \\
&= \exp(N\log(2) - Nt^2/8) \\
&\leqslant \exp(-Nt^2/16),
\end{aligned}$$

当 $t \geqslant 4\sqrt{\log(2)}$ 时。不等式（8）以及 $\varphi_+(t)$ 的下界 给出

$$\Pr(Y_{((1/2+\varepsilon)N)} - \Phi^{-1}(1/2+\varepsilon) \geqslant t) \leqslant \exp\left(\frac{-2Nt^2}{2\pi\exp((\Phi^{-1}(3/4)+4\sqrt{\log(2)})^2)}\right),$$

当 $0 \leqslant t \leqslant 4\sqrt{\log(2)}$ 时。结合这两种情形并调整绝对常数，我们证明了一个上尾界。下面证明下尾界。证明过程类似，但计算略有不同。我们要证明，对任意 $t \geqslant 0$,

$$\Pr(Y_{((1/2+\varepsilon)N)} - \Phi^{-1}(1/2+\varepsilon) \leqslant -t) \leqslant \exp(-c_1 N t^2).$$

对于任意 $s \in \mathbb{R}$ 和 Z_1，我们有……，Z_N 如上

$$\Pr(Y_{((1/2+\varepsilon)N)} \leqslant s) = \Pr\left(\sum_{i=1}^{N} Z_i \leqslant (1/2-\varepsilon)N\right).$$

定义

$$\varphi_-(t) = \Phi(\Phi^{-1}(1/2+\varepsilon) - t) - 1/2 - \varepsilon.$$

类似地，我们有

$$\Pr(Y_{((1/2+\varepsilon)N)} - \Phi^{-1}(1/2+\varepsilon) \leqslant -t) \leqslant \exp\left(-2N(\varphi_-(t))^2\right).$$

现在，我们按如下方式给数量 $|\varphi_-(t)|$ 一个下界

$$\begin{aligned}
|\varphi_-(t)| &= \Phi(\Phi^{-1}(1/2+\varepsilon)) - \Phi(\Phi^{-1}(1/2+\varepsilon) - t) \\
&\geqslant \frac{t}{\sqrt{2\pi}} \exp\left(-\max\{(\Phi^{-1}(1/2+\varepsilon))^2, (\Phi^{-1}(1/2+\varepsilon)-t)^2\}/2\right) \\
&\geqslant \frac{t}{\sqrt{2\pi}} \exp\left(-\max\{(\Phi^{-1}(3/4))^2, t^2\}/2\right).
\end{aligned}$$

如上所述，对于较大的 t，我们需要得到其尾界。我们需要至少有 $N/4$ 的观测值小于 $\Phi^{-1}(1/2+\varepsilon) - t$。在下文中，我们假设 $t \geqslant 2\Phi^{-1}(3/4)$。这可以按如下方式控制

$$\begin{aligned}
\Pr(Y_{((1/2+\varepsilon)N)} - \Phi^{-1}(1/2+\varepsilon) \leqslant -t) &= \Pr(-Y_{((1/2+\varepsilon)N)} \geqslant t - \Phi^{-1}(1/2+\varepsilon)) \\
&\leqslant \binom{N}{N/4} \left(\Pr(-Y_1 \geqslant t - \Phi^{-1}(1/2+\varepsilon))\right)^{N/4} \\
&\leqslant 2^N \left(\Pr(Y_1 \geqslant t - \Phi^{-1}(3/4))\right)^{N/4} \\
&\leqslant 2^N \left(\Pr(Y_1 \geqslant t/2)\right)^{N/4} \\
&\leqslant 2^N \exp(-Nt^2/32) \\
&\leqslant \exp(-Nt^2/64),
\end{aligned}$$

只要 $t \geqslant$ 最大值$\{2\Phi^{-1}(3/4)8\sqrt{\log(2)}\}$。下尾的证明如下。并集界完成证明。最后，我们关于 ψ_2- 范数的界由 [Ver16，命题 2.5。2]。

□

我们的第二个结果给出了一个类似的集中界，适用于从 χ_1^2 分布中抽取的独立同分布观测值的经验分位数。该分布与标准高斯随机变量的平方所服从的分布一致。记其累积分布函数为 $F_{\chi_1^2}(\cdot)$，其反函数为 $F_{\chi_1^2}^{-1}(\cdot)$。主要区别在于，此处我们只能证明次指数尾界。我们指出，考虑具有 χ_k^2 分布且自由度为 $k \geqslant 2$ 的情形时，所需的集中不等式可由对数凹性和 [BL19，引理 6 得出。5]。

引理 2（χ_1^2 分布的分位数）。存在一个绝对常数 $c_1 > 0$，使得下述结论成立。不失一般性，假设 $(1/2 \pm \varepsilon)N$ 为整数。令 $Y_1,\dots,Y_N$ 为一个由独立同分布 χ_1^2 随机变量组成的样本，且 $\varepsilon \in [0, 1/4]$。那么，

$$\left\| Y_{((1/2\pm\varepsilon)N)} - F_{\chi_1^2}^{-1}(1/2 \pm \varepsilon) \right\|_{\psi_1} \leqslant \frac{c_1}{\sqrt{N}}.$$

证明。首先，χ_1^2 的密度 g 为

$$g(x) = \frac{\exp(-x/2)}{\sqrt{2\pi x}}, \quad x > 0. \tag{9}$$

也容易证明 $F_{\chi_1^2}^{-1}(1/2) = (\Phi^{-1}(3/4))^2$。再次使用相同的记号，我们记 $\varphi_+(t) = F_{\chi_1^2}(F_{\chi_1^2}^{-1}(1/2+\varepsilon)+t) - 1/2 - \varepsilon$。重复引理 1 的证明过程，当 $\varphi_+(t) \geqslant 0$ 时，我们有

$$\Pr(Y_{((1/2+\varepsilon)N)} - F_{\chi_1^2}^{-1}(1/2+\varepsilon) \geqslant t) \leqslant \exp\left(-2N(\varphi_+(t))^2\right).$$

利用（9），可得

$$\begin{aligned}
\varphi_+(t) &= F_{\chi_1^2}(F_{\chi_1^2}^{-1}(1/2+\varepsilon)+t) - F_{\chi_1^2}(F_{\chi_1^2}^{-1}(1/2+\varepsilon)) \\
&\geqslant \frac{t\exp(-F_{\chi_1^2}(F_{\chi_1^2}^{-1}(1/2+\varepsilon)+t))}{\sqrt{2\pi F_{\chi_1^2}(F_{\chi_1^2}^{-1}(1/2+\varepsilon)+t)}}.
\end{aligned}$$

只要 $\varepsilon \leqslant 1/4$ 且 $t \geqslant 0$ 被某个绝对常数所界定，就可得到次高斯尾部。根据 χ^2 分布 [LM00 的集中性，引理 1] 对于 $t \geqslant 0$，有

$\Pr(Y_1 \geqslant 1 + \sqrt{2t} + t) \leqslant \exp(-t)$，因此，当 $t \geqslant 4 + 2\sqrt{3}$ 时，$\Pr(Y_1 \geqslant t) \leqslant \exp(-t/2)$.

由于 $\varepsilon \leqslant 1/4$，我们需要所有观测值中至少有 $N/4$ 个高于 $F_{\chi_1^2}^{-1}(1/2+\varepsilon)+t$。因此，只要 $t \geqslant 4 + 2\sqrt{3}$，就有

$$\begin{aligned}
\Pr(Y_{((1/2+\varepsilon)N)} - F_{\chi_1^2}^{-1}(1/2+\varepsilon) \geqslant t) &\leqslant \binom{N}{N/4} \left(\Pr(Y_1 \geqslant \Phi^{-1}(1/2+\varepsilon)+t)\right)^{N/4} \\
&\leqslant \binom{N}{N/4} \left(\Pr(Y_1 \geqslant t)\right)^{N/4} \\
&\leqslant 2^N \exp(-Nt/8) \\
&= \exp(N\log(2) - Nt/8) \\
&\leqslant \exp(-Nt/16),
\end{aligned}$$

其中，最后一个不等式还需要 $t \geqslant 16\log(2)$。结合上述界并调整绝对常数 $c_1 \geqslant 0$，我们得到

$$\Pr(Y_{((1/2+\varepsilon)N)} - F_{\chi_1^2}^{-1}(1/2+\varepsilon) \geqslant t) \leqslant \exp(-c_1 N \min\{t, t^2\}).$$

我们继续推导下尾界。我们希望证明，对于任意 $t \geqslant 0$,

$$\Pr(Y_{((1/2+\varepsilon)N)} - F_{\chi_1^2}^{-1}(1/2+\varepsilon) \leqslant -t) \leqslant \exp(-c_2 N t^2),$$

其中 c_2 是一个绝对常数。对于 $0 \leqslant t < F_{\chi_1^2}^{-1}(1/2+\varepsilon)$，我们定义

$$\varphi_-(t) = F_{\chi_1^2}^{-1}(F_{\chi_1^2}^{-1}(1/2+\varepsilon) - t) - 1/2 - \varepsilon.$$

使用相同的论证，我们可以证明

$$\Pr\left(Y_{((1/2+\varepsilon)N)} - F_{\chi_1^2}^{-1}(1/2+\varepsilon) \leqslant -t\right) \leqslant \exp\left(-2N(\varphi_-(t))^2\right).$$

我们有

$$|\varphi_-(t)| = F_{\chi_1^2}(F_{\chi_1^2}^{-1}(1/2+\varepsilon)) - F_{\chi_1^2}(F_{\chi_1^2}^{-1}(1/2+\varepsilon) - t) \geqslant \frac{t\exp\left(-F_{\chi_1^2}^{-1}(1/2+\varepsilon)\right)}{\sqrt{2\pi F_{\chi_1^2}^{-1}(1/2+\varepsilon)}}.$$

由于 $\varepsilon \leqslant 1/4$ 我们便在以下情形下完成证明： $t \leqslant F_{\chi_1^2}^{-1}(1/2+\varepsilon)$。注意，由于 $Y_{((1/2+\varepsilon)N)}$ 非负，我们可以将此界推广到所有 $t > F_{\chi_1^2}^{-1}(1/2+\varepsilon)$。第6 节中的引理 4 完成证明。对第 ($1/2-\varepsilon$)个分位数的分析重复同样的步骤。

□

我们的最终结果为（标准）半正态分布证明了一个类似的界。具体而言，我们要证明标准高斯随机变量绝对值的分位数满足集中不等式。将该分布的累积分布函数记为$\Phi_{\mathrm{H}}(\cdot)$，其反函数记为 $\Phi_{\mathrm{H}}^{-1}(\cdot)$。

引理 3 （半正态分布的分位数）。 存在一个绝对常数 c_1 ，使得以下结论成立。不失一般性，假设 $(1/2 \pm \varepsilon)N$ 为整数。令$Y_1, \ldots, Y_N$为独立半正态随机变量构成的样本，且 $\varepsilon \in [0, 1/4]$。那么，

$$\left\|Y_{((1/2\pm\varepsilon)N)} - \Phi_{\mathrm{H}}^{-1}(1/2 \pm \varepsilon)\right\|_{\psi_2} \leqslant \frac{c_1}{\sqrt{N}}.$$

该结果的证明重复了引理1 和引理2证明中使用的相同计算。细节从略。

4 辅助结果

下一节将介绍本文中用到的若干技术结果。首先给出一个通常称为*PAC*-贝叶斯引理的界，它直接源自相对熵的 Donsker–Varadhan 变分公式 [DV75]。

引理 4。 假设 X 是在某个可测空间 $\mathcal{X}$上定义的随机变量。另假设 Θ （称为参数空间）是$\mathbb{R}^d$的子集。设 γ 是 Θ 上的一个分布（称为先验分布），并令 ρ 为 Θ 上的任意分布（称为后验分布），满足 $\rho \ll \gamma$。设 $f\colon \mathcal{X} \times \Theta \to \mathbb{R}$，使得$\mathbf{E}_X \exp(f(X,\theta))$ 几乎处处关于 γ有限。那么，我们有

$$\Pr_X\left(\textit{for all } \rho \ll \gamma \ :\ \mathbf{E}_\rho f(X,\theta) \leqslant \mathbf{E}_\rho \log(\mathbf{E}_X \exp(f(X,\theta))) + \mathcal{KL}(\rho,\gamma) + t\right) \geqslant 1 - e^{-t}.$$

近期多篇关于重尾分布均值与协方差估计的论文中，采用的一项关键论证 [CG17、Giu18、AZ22、OR22], 是巧妙地运用这一引理，从而绕过球面覆盖和 VC 类型论证。引理4 将在我们的分析中发挥同样的关键作用。不过，以往对这一引理的应用，是用（截断）随机变量的和代替 $f(X,\ \theta)$（在这种情况下， X 本质上是一个由相互独立的随机变量组成的向量 X_1, ..., X_N），而我们则利用引理4与特定单变量分布的样本分位数之间的相互作用。引理4 的证明及其部分应用见[CG17、Zhi21]。

4.1 后验分布的分析

我们分析中的另一个技术层面，是在稳健估计的背景下引入截断后验分布。对于给定的半正定矩阵 G 和 $r \geqslant 0$，我们将均值为 $v \in S^{d-1}$ 、协方差为 $\beta^{-1} I_d$ 的多元高斯分布截断如下。定义密度函数

$$f_v(x) = \frac{1}{p(2\pi\beta^{-1})^{d/2}} \exp\left(-\frac{\beta\|x-v\|^2}{2}\right) \mathbf{1}\left(\|G^{1/2}(x-v)\| \leqslant r\right), \tag{10}$$

其中 $p > 0$ 是一个归一化因子。下面给出以下结果。

引理 5 （截断后验分布的性质）。 设 r， $\beta > 0$ 并设 G 表示定义 (10) 中的一个半正定矩阵。设 Σ是零均值随机向量X 在$\mathbb{R}^d$ 中的一个协方差矩阵，满足

$$\frac{1}{10}\Sigma \preceq G, \quad \textit{and} \quad \operatorname{Tr}(G) \leqslant 10 \operatorname{Tr}(\Sigma).$$

Let γ 是 $\mathbb{R}^d$ 中的一个高斯测度，其均值为零，协方差为 $\beta^{-1} I_d$。此外，如果 ly

$$r \geqslant \sqrt{20\beta^{-1} \operatorname{Tr}(\Sigma)},$$

那么我们有

$$\mathcal{KL}(\rho_v, \gamma) \leqslant \log(2) + \beta/2.$$

此外，设 θ 服从 ρ_v。那么，$\mathbf{E}_{\rho_v}\theta = v$，并且对于 θ 的实现几乎必然有

$$\theta^{\mathsf{T}}\Sigma\theta \leqslant 2\|\Sigma\| + 20r^2.$$

证明。 我们利用如下事实：对于服从 ρ_v 的 θ ，有 $\mathbf{E}_{\rho_v}\theta = v$。这是由密度关于 v 的对称性得出的。令 g 表示均值为零、协方差为 $\beta^{-1} I_d$ 的高斯随机向量的密度。为了控制 $\mathcal{KL}(\rho_v, \mu)$，我们写作

$$\begin{aligned}
\int \log\left(\frac{f_v(x)}{g(x)}\right) f_v(x)dx &= \mathbf{E}_{\rho_v} \log\left(\frac{1}{p}\exp\left(\frac{-\beta\|\theta-v\|^2 + \beta\|\theta\|^2}{2}\right)\right) \\
&= \log\left(\frac{1}{p}\right) + \mathbf{E}_{\rho_v}\left(\frac{-\beta\|v\|^2 + 2\beta\langle\theta, v\rangle}{2}\right) \\
&= \log\left(\frac{1}{p}\right) + \frac{\beta}{2}.
\end{aligned}$$

为了证明所需的不等式，我们注意到

$$p = \Pr(\|G^{1/2}W\| \leqslant r),$$

其中 W 是协方差为 $\beta^{-1} I_d$ 的零均值高斯随机向量。由于 $\operatorname{Tr}(G) \leqslant 10\operatorname{Tr}(\Sigma)$，简单计算可得

$$\Pr(\|G^{1/2}W\| \geqslant r) \leqslant \mathbf{E}W^{\mathsf{T}}GW/r^2 = \beta^{-1}\operatorname{Tr}(G)/r^2 \leqslant 10\beta^{-1}\operatorname{Tr}(\Sigma)/r^2 \leqslant 1/2,$$

只要 $r \geqslant \sqrt{20\beta^{-1}\operatorname{Tr}(\Sigma)}$。因此，在此假设下 $p \geqslant 1/2$，从而 $\log(1/p) \leqslant \log 2$。这就证明了第一个不等式。利用矩阵 G 的第二个性质，我们有

$$\theta^{\mathsf{T}}\Sigma\theta \leqslant 2v^{\mathsf{T}}\Sigma v + 2(\theta-v)^{\mathsf{T}}\Sigma(\theta-v) \leqslant 2v^{\mathsf{T}}\Sigma v + 20(\theta-v)^{\mathsf{T}}G(\theta-v) \leqslant 2\|\Sigma\| + 20r^2.$$

该结论随即得证。 □

我们的下一个结果将混合次高斯/次指数尾界转化为对 $\|\cdot\|_{\psi_1}$-范数的界。为完整起见，我们给出这一标准推导。

引理 6. 假设随机变量 X 对所有 $t \geqslant 0$都满足，

$$\Pr(|X| \geqslant t) \leqslant 2\exp(-K\min\{t^2, t\}),$$

其中 $K > 1$ 是某个常数。那么，存在一个绝对常数 $c > 0$，使得 t

$$\|X\|_{\psi_1} \leqslant \frac{c}{\sqrt{K}}.$$

证明。 我们只需计算 X的矩。对于固定的 $p \geqslant 1$，有

$$\begin{aligned}
\mathbf{E}|X|^p &= \int_0^\infty \Pr(|X|^p \geqslant t)\, dt = \int_0^\infty \Pr(|X| \geqslant t) p t^{p-1}\, dt \leqslant 2\int_0^\infty \exp(-K\min\{t^2, t\}) p t^{p-1} dt \\
&\leqslant 2\int_0^\infty \exp(-Kt^2) p t^{p-1} dt + 2\int_0^\infty \exp(-Kt) p t^{p-1} dt = \frac{1}{K^{p/2}} p\Gamma(p/2) + \frac{1}{K^p} 2p\Gamma(p) \\
&\leqslant \frac{3p(p/2)^{p/2}}{K^{p/2}} + \frac{2p^p}{K^p} \leqslant \frac{3p^p}{K^{p/2}} + \frac{2p^p}{K^p} \leqslant \frac{5p^p}{K^{p/2}},
\end{aligned}$$

其中 $\Gamma(\cdot)$表示伽马函数，并且我们对所有 $\Gamma(x) \leqslant 3x^x$ 使用了 $x \geqslant 1/2$ 以及$p\Gamma(p) = \Gamma(p+1) \leqslant p^p$。最后，[Ver16、命题 2.7.1(b)] 推出所需的界。 □

5 主要结果的证明

我们首先证明第一个主要结果，该结果表明，在（2）中定义的估计量在稳健均值估计问题中达到了最优误差界。我们将在本节末讨论结果的最优性。

定理*1*的证明。首先，根据我们估计量的定义，有

$$\widehat{\mu} = \operatorname*{arg\,min}_{\nu \in \mathbb{R}^d} \sup_{v \in S^{d-1}} |\mathbf{E}_{\rho_v} \operatorname{Med}(\langle X_1, \theta\rangle, \ldots, \langle X_N, \theta\rangle) - \langle \nu, v\rangle|,$$

其中， ρ_v 是定义在 R^d 上、均值为 v 、协方差为 $\beta^{-1} I_d$ 的多元高斯分布。根据三角不等式和我们估计量的定义，有

$$\begin{aligned}
\|\widehat{\mu} - \mu\| &= \sup_{v \in S^{d-1}} \langle \widehat{\mu} - \mu, v\rangle \\
&\leqslant \sup_{v \in S^{d-1}} |\mathbf{E}_{\rho_v} \operatorname{Med}(\langle X_1, \theta\rangle, \ldots, \langle X_N, \theta\rangle) - \langle \widehat{\mu}, v\rangle| \\
&\quad + \sup_{v \in S^{d-1}} |\mathbf{E}_{\rho_v} \operatorname{Med}(\langle X_1, \theta\rangle, \ldots, \langle X_N, \theta\rangle) - \langle \mu, v\rangle| \\
&\leqslant 2 \sup_{v \in S^{d-1}} |\mathbf{E}_{\rho_v} \operatorname{Med}(\langle X_1, \theta\rangle, \ldots, \langle X_N, \theta\rangle) - \langle \mu, v\rangle| \\
&= 2 \sup_{v \in S^{d-1}} |\mathbf{E}_{\rho_v} \operatorname{Med}(\langle X_1 - \mu, \theta\rangle, \ldots, \langle X_N - \mu, \theta\rangle)|.
\end{aligned}$$

我们需要对最后一个量进行界定。从现在起，不失一般性地假设$\mu = 0$。假设 Y_1,... , Y_N是一个均值为零、协方差为 Σ的独立高斯随机变量组成的未受污染样本。也就是说，在 $X_1, \ldots, X_N$中，至多有 εN个元素不同于它们的

$Y_1, \ldots, Y_N$个对应元素。注意，污染样本在任意方向上的投影的样本中位数，不可能与未受污染样本对应投影的 $1/2 \pm \varepsilon$ 分位数相差太远。形式化地说，假设样本中位数和 $1/2 \pm \varepsilon$ 样本分位数均唯一，则几乎必然有 ρ_v

$$\begin{aligned}\operatorname{Quant}_{\frac{1}{2}-\varepsilon}(\langle Y_1,\theta\rangle,\ldots,\langle Y_N,\theta\rangle) &\leqslant \operatorname{Med}(\langle X_1,\theta\rangle,\ldots,\langle X_N,\theta\rangle)\\ &\leqslant \operatorname{Quant}_{\frac{1}{2}+\varepsilon}(\langle Y_1,\theta\rangle,\ldots,\langle Y_N,\theta\rangle),\end{aligned}$$

因此，对 ρ_v 取期望，可得

$$\begin{aligned}\mathbf{E}_{\rho_v}\operatorname{Quant}_{\frac{1}{2}-\varepsilon}(\langle Y_1,\theta\rangle,\ldots,\langle Y_N,\theta\rangle) &\leqslant \mathbf{E}_{\rho_v}\operatorname{Med}(\langle X_1,\theta\rangle,\ldots,\langle X_N,\theta\rangle)\\ &\leqslant \mathbf{E}_{\rho_v}\operatorname{Quant}_{\frac{1}{2}+\varepsilon}(\langle Y_1,\theta\rangle,\ldots,\langle Y_N,\theta\rangle).\end{aligned}$$

因此，对于任意 $v \in S^{d-1}$，我们有

$$\begin{aligned}|\mathbf{E}_{\rho_v}\operatorname{Med}(\langle X_1,\theta\rangle,\ldots,\langle X_N,\theta\rangle)| &\leqslant \left|\mathbf{E}_{\rho_v}\operatorname{Quant}_{\frac{1}{2}+\varepsilon}(\langle Y_1,\theta\rangle,\ldots,\langle Y_N,\theta\rangle)\right|\\ &\quad+\left|\mathbf{E}_{\rho_v}\operatorname{Quant}_{\frac{1}{2}-\varepsilon}(\langle Y_1,\theta\rangle,\ldots,\langle Y_N,\theta\rangle)\right|.\end{aligned}$$

这两项的分析方法类似。我们只分析第一项。注意，由于球对称性，
$S_N = \{\langle Y_1,\theta\rangle/\sqrt{\theta^\mathsf{T}\Sigma\theta},\ldots,\langle Y_N,\theta\rangle/\sqrt{\theta^\mathsf{T}\Sigma\theta}\}$ 由相互独立的标准高斯随机变量组成（在我们的情形下，$\theta \neq 0$ 几乎必然成立）。我们有

$$\begin{aligned}\left|\mathbf{E}_{\rho_v}\operatorname{Quant}_{\frac{1}{2}+\varepsilon}(\langle Y_1,\theta\rangle,\ldots,\langle Y_N,\theta\rangle)\right| &\leqslant \left|\mathbf{E}_{\rho_v}\sqrt{\theta^\mathsf{T}\Sigma\theta}\left(\operatorname{Quant}_{\frac{1}{2}+\varepsilon}(S_N)-\mathbf{E}\operatorname{Quant}_{\frac{1}{2}+\varepsilon}(S_N)\right)\right|\\ &\quad+\left|\mathbf{E}_{\rho_v}\sqrt{\theta^\mathsf{T}\Sigma\theta}\cdot\left(\mathbf{E}\operatorname{Quant}_{\frac{1}{2}+\varepsilon}(S_N)-\Phi^{-1}(1/2+\varepsilon)\right)\right|\\ &\quad+\mathbf{E}_{\rho_v}\sqrt{\theta^\mathsf{T}\Sigma\theta}\cdot\Phi^{-1}(1/2+\varepsilon)\\ &= (I)+(II)+(III).\end{aligned}$$

为了给 (I) 给出上界，我们希望应用引理 4。固定 $\lambda > 0$，并令 γ 为 $\mathbb{R}^d$ 上均值为零、协方差为 $\beta^{-1}I_d$ 的多元高斯分布。标准公式表明$\mathcal{KL}(\rho_v,\gamma) = \beta/2$。因此，由引理 4 可知，对所有 $v \in S^{d-1}$，以下结论同时以至少 $1-\delta$ 的概率成立，

$$\begin{aligned}&\lambda\mathbf{E}_{\rho_v}\sqrt{\theta^\mathsf{T}\Sigma\theta}\left(\operatorname{Quant}_{\frac{1}{2}+\varepsilon}(S_N)-\mathbf{E}\operatorname{Quant}_{\frac{1}{2}+\varepsilon}(S_N)\right)\\ &\quad\leqslant \mathbf{E}_{\rho_v}\log\mathbf{E}\exp\left(\lambda\sqrt{\theta^\mathsf{T}\Sigma\theta}\left(\operatorname{Quant}_{\frac{1}{2}+\varepsilon}(S_N)-\mathbf{E}\operatorname{Quant}_{\frac{1}{2}+\varepsilon}(S_N)\right)\right)+\beta/2+\log(1/\delta).\end{aligned}$$

由于中心化使 ψ_2-范数最多乘以一个绝对常数因子（例如参见 [Ver16，引理 2.6。8]），由引理 1，存在某个绝对常数 $c_1 > 0$，使得

$$\left\|\operatorname{Quant}_{\frac{1}{2}+\varepsilon}(S_N)-\mathbf{E}\operatorname{Quant}_{\frac{1}{2}+\varepsilon}(S_N)\right\|_{\psi_2} \leqslant \frac{c_1}{\sqrt{N}}.$$

因此，根据 [Ver16、命题 2.5.2 的 (v)] (在给定 θ的条件下，我们在该结果中取 $\lambda\sqrt{\theta^\mathsf{T}\Sigma\theta}$ 而非 λ)，对于某个绝对常数 $c_2 > 0$,

$$\begin{aligned}\mathbf{E}_{\rho_v}\log\mathbf{E}\exp\left(\lambda\sqrt{\theta^\mathsf{T}\Sigma\theta}\left(\operatorname{Quant}_{\frac{1}{2}+\varepsilon}(S_N)-\mathbf{E}\operatorname{Quant}_{\frac{1}{2}+\varepsilon}(S_N)\right)\right) &\leqslant \frac{\mathbf{E}_{\rho_v}c_2\lambda^2\theta^\mathsf{T}\Sigma\theta}{N}\\ &= \frac{c_2\lambda^2(v^\mathsf{T}\Sigma v+\beta^{-1}\operatorname{Tr}(\Sigma))}{N}\\ &\leqslant \frac{11c_2\lambda^2\|\Sigma\|}{N},\end{aligned}$$

其中，最后几行基于直接计算以及我们对 β 的选择（我们有 $\beta^{-1} \leqslant 10\|\Sigma\|/$迹$(\Sigma)$）。对 λ $\mathbf{E}_{\rho_v}\sqrt{\theta^\mathsf{T}\Sigma\theta}\left(\mathrm{Quant}_{\frac{1}{2}+\varepsilon}(S_N) - \mathbf{E}\,\mathrm{Quant}_{\frac{1}{2}+\varepsilon}(S_N)\right)$的界关于 $\lambda > 0$ 进行优化，并且由于 $\beta \leqslant 10$ 迹$(\Sigma)/\|\Sigma\|$，我们得到该结论对所有 S^{d-1} 均一致成立，

$$\mathbf{E}_{\rho_v}\sqrt{\theta^\mathsf{T}\Sigma\theta}\left(\mathrm{Quant}_{\frac{1}{2}+\varepsilon}(S_N) - \mathbf{E}\,\mathrm{Quant}_{\frac{1}{2}+\varepsilon}(S_N)\right) \leqslant c_3\sqrt{\frac{\mathrm{Tr}(\Sigma) + \|\Sigma\|\log(1/\delta)}{N}},$$

其中， $c_3 > 0$ 是一个绝对常数。对 $\lambda < 0$ 重复上述证明并使用并集界，我们将此界推广为上界（I）。

现在我们着重于给 (II) 定上界。注意，对于任意标量 C，有 $\|C\|_{\psi_2} = |C|/\sqrt{\log 2}$。利用这一观察以及詹森不等式和引理 1 ，对于某个 $c_4 > 0$，有

$$\begin{aligned}\left|\mathbf{E}_{\rho_v}\sqrt{\theta^\mathsf{T}\Sigma\theta}\cdot\left(\mathbf{E}\,\mathrm{Quant}_{\frac{1}{2}+\varepsilon}(S_N) - \Phi^{-1}(1/2+\varepsilon)\right)\right| &\leqslant \frac{c_4\sqrt{\log 2}\cdot\mathbf{E}_{\rho_v}\sqrt{\theta^\mathsf{T}\Sigma\theta}}{\sqrt{N}} \\ &\leqslant \frac{c_4\sqrt{\log 2}\cdot\sqrt{11\|\Sigma\|}}{\sqrt{N}}.\end{aligned}$$

最后，我们给（III）定上界。首先注意，函数 $\Phi^{-1}(\cdot)$在闭区间 $[1/2-\varepsilon, 1/2+\varepsilon]$ 上对于 $\varepsilon \in [0, 1/4]$是局部利普希茨连续的。我们如下计算并界定 Φ^{-1} 的局部利普希茨常数

$$\frac{d}{dx}\Phi^{-1}(1/2+x) = \sqrt{2\pi}\exp\left((\Phi^{-1}(1/2+x))^2/2\right) \leqslant \sqrt{2\pi}\exp\left((\Phi^{-1}(3/4))^2/2\right) \leqslant 4,$$

这里我们用到了函数 $\exp(\Phi^{-1}(\cdot)^2)$ 是递增的这一事实。由于标准高斯分布是对称的，因此有 $\Phi^{-1}(1/2) = 0$。因此， $\Phi^{-1}(1/2+\varepsilon)$ 的界为

$$\Phi^{-1}(1/2+\varepsilon) = \Phi^{-1}(1/2+\varepsilon) - \Phi^{-1}(1/2) \leqslant 4\varepsilon.$$

因此，我们有

$$\mathbf{E}_{\rho_v}\sqrt{\theta^\mathsf{T}\Sigma\theta}\cdot\Phi^{-1}(1/2+\varepsilon) \leqslant 4\varepsilon\sqrt{11\|\Sigma\|}.$$

结合上述界即可完成证明。 □

现在我们可以证明第二个主要结果了。

定理 2的证明。 回忆一下， H 是 ρ_v 的协方差矩阵，且不依赖于方向$v \in S^{d-1}$。注意，$\mathbf{E}_{\rho_v}\theta^\mathsf{T}\Sigma\theta = v^\mathsf{T}\Sigma v + \mathrm{Tr}(\Sigma H)$。此外，由于我们选择的参数意味着 $p \geqslant 1/2$ 在（10）中，因此有

$$H \preceq 2\beta^{-1}I_d, \quad \text{and} \quad \|H\| \leqslant 2\beta^{-1}.$$

我们还注意到 $\mathbf{r}(G) =$迹$(G)/\|G\| \leqslant 100\mathbf{r}(\Sigma)$。利用三角不等式，以及将估计量的定义与集合 $\mathcal{H}$ 的定义（见7）相结合，可得

$$\begin{aligned}\|\widehat{\Sigma} - \Sigma\| &= \sup_{v\in S^{d-1}}\left|v^\mathsf{T}\Sigma v - v^\mathsf{T}\widehat{\Sigma}v\right| \\ &\leqslant \sup_{v\in S^{d-1}}\left|\mathbf{E}_{\rho_v}\theta^\mathsf{T}(\Sigma-\widehat{\Sigma})\theta\right| + |(\mathrm{Tr}(\widehat{\Sigma}H) - \alpha(H)) - (\mathrm{Tr}(\Sigma H) - \alpha(H))| \\ &\leqslant \sup_{v\in S^{d-1}}\left|\mathbf{E}_{\rho_v}\theta^\mathsf{T}(\Sigma-\widehat{\Sigma})\theta\right| + c\,\mathrm{Tr}((\Sigma+\widehat{\Sigma})H)\left(\sqrt{\frac{\max\{\mathbf{r}(G),\mathbf{r}(\Sigma)\}+\log(1/\delta)}{N}} + \varepsilon\right) \\ &\leqslant \sup_{v\in S^{d-1}}\left|\mathbf{E}_{\rho_v}\theta^\mathsf{T}(\Sigma-\widehat{\Sigma})\theta\right| + 2\beta^{-1}c\,\mathrm{Tr}(\Sigma+10G)\left(\sqrt{\frac{100\mathbf{r}(\Sigma)+\log(1/\delta)}{N}} + \varepsilon\right) \\ &\leqslant \sup_{v\in S^{d-1}}\left|\mathbf{E}_{\rho_v}\theta^\mathsf{T}(\Sigma-\widehat{\Sigma})\theta\right| + 202\beta^{-1}c\,\mathrm{Tr}(\Sigma)\left(\sqrt{\frac{100\mathbf{r}(\Sigma)+\log(1/\delta)}{N}} + \varepsilon\right).\end{aligned}$$

由于 $\beta^{-1} \leqslant 10\mathbf{r}(\Sigma)$，最后一个不等式中的最后一项不超过定理 2陈述中的收敛速率。现在，我们只需对前一个公式中不等式最后一行的第一项进行界定。首先需要进行一些辅助计算。根据集合 $\mathcal{H}$的定义，我们有 ρ_v几乎必然

$$\begin{aligned}\sqrt{\theta^{\mathsf{T}}\Sigma\theta} + \sqrt{\theta^{\mathsf{T}}\widehat{\Sigma}\theta} &\leqslant \sqrt{2v^{\mathsf{T}}\Sigma v + 2(\theta-v)^{\mathsf{T}}\Sigma(\theta-v)} + \sqrt{2v^{\mathsf{T}}\widehat{\Sigma}v + 2(\theta-v)^{\mathsf{T}}\widehat{\Sigma}(\theta-v)} \\ &\leqslant \sqrt{2\|\Sigma\| + 20(\theta-v)^{\mathsf{T}}G(\theta-v)} + \sqrt{20\omega + 20(\theta-v)^{\mathsf{T}}G(\theta-v)} \\ &\leqslant \sqrt{2\|\Sigma\| + 20r^2} + \sqrt{20\omega + 20r^2} \\ &\leqslant c_1\sqrt{\|\Sigma\|},\end{aligned}$$

其中 $c_1 > 0$ 是某个绝对常数。这推出以下几行

$$\begin{aligned}&\sup_{v\in S^{d-1}} \left|\mathbf{E}_{\rho_v}\theta^{\mathsf{T}}(\Sigma - \widehat{\Sigma})\theta\right| \\ &\leqslant \sup_{v\in S^{d-1}} \mathbf{E}_{\rho_v}\left|\sqrt{\theta^{\mathsf{T}}\Sigma\theta} - \sqrt{\theta^{\mathsf{T}}\widehat{\Sigma}\theta}\right|\left|\sqrt{\theta^{\mathsf{T}}\Sigma\theta} + \sqrt{\theta^{\mathsf{T}}\widehat{\Sigma}\theta}\right| \\ &\leqslant c_1\sqrt{\|\Sigma\|}\sup_{v\in S^{d-1}} \mathbf{E}_{\rho_v}\left|\sqrt{\theta^{\mathsf{T}}\Sigma\theta} - \sqrt{\theta^{\mathsf{T}}\widehat{\Sigma}\theta}\right| \\ &\leqslant c_1\sqrt{\|\Sigma\|}\sup_{v\in S^{d-1}} \mathbf{E}_{\rho_v}\left|\mathrm{Med}\left(|\langle X_1,\theta\rangle|,\ldots,|\langle X_N,\theta\rangle|\right)/(\Phi^{-1}(3/4)) - \sqrt{\theta^{\mathsf{T}}\widehat{\Sigma}\theta}\right| \\ &\quad + c_1\sqrt{\|\Sigma\|}\sup_{v\in S^{d-1}} \mathbf{E}_{\rho_v}\left|\mathrm{Med}\left(|\langle X_1,\theta\rangle|,\ldots,|\langle X_N,\theta\rangle|\right)/(\Phi^{-1}(3/4)) - \sqrt{\theta^{\mathsf{T}}\Sigma\theta}\right| \\ &\leqslant 2c_1\sqrt{\|\Sigma\|}\sup_{v\in S^{d-1}} \mathbf{E}_{\rho_v}\left|\mathrm{Med}\left(|\langle X_1,\theta\rangle|,\ldots,|\langle X_N,\theta\rangle|\right)/(\Phi^{-1}(3/4)) - \sqrt{\theta^{\mathsf{T}}\Sigma\theta}\right|,\end{aligned}$$

^其中在最后一行，我们使用了 Σ 的定义以及 $\Sigma \in \mathcal{H}$。我们着重对最后一个表达式进行上界估计。令 $Y_1,\ldots,Y_N$为未受污染的版本，即我们的 ε-污染样本。采用定理 1证明中的相同论证，我们有

$$\begin{aligned}&\sup_{v\in S^{d-1}} \mathbf{E}_{\rho_v}\left|\mathrm{Med}\left(|\langle X_1,\theta\rangle|,\ldots,|\langle X_N,\theta\rangle|\right) - \Phi^{-1}(3/4)\sqrt{\theta^{\mathsf{T}}\Sigma\theta}\right| \\ &\leqslant \sup_{v\in S^{d-1}} \mathbf{E}_{\rho_v}\left|\mathrm{Quant}_{\frac{1}{2}+\varepsilon}\left(|\langle Y_1,\theta\rangle|,\ldots,|\langle Y_N,\theta\rangle|\right) - \Phi^{-1}(3/4)\sqrt{\theta^{\mathsf{T}}\Sigma\theta}\right| \\ &\quad + \sup_{v\in S^{d-1}} \mathbf{E}_{\rho_v}\left|\mathrm{Quant}_{\frac{1}{2}-\varepsilon}\left(|\langle Y_1,\theta\rangle|,\ldots,|\langle Y_N,\theta\rangle|\right) - \Phi^{-1}(3/4)\sqrt{\theta^{\mathsf{T}}\Sigma\theta}\right|.\end{aligned}$$

我们只分析第一项。注意，由于球对称性，我们有$S_N = \{|\langle Y_1,\theta\rangle|/\sqrt{\theta^{\mathsf{T}}\Sigma\theta}\ ,\ldots,\ |\langle Y_N,\theta\rangle|/\sqrt{\theta^{\mathsf{T}}\Sigma\theta}\}$ 由相互独立的半正态随机变量组成（在我们的情形下， $\theta \neq 0$ 几乎必然成立）。根据三角不等式，我们有

$$\begin{aligned}&\sup_{v\in S^{d-1}} \mathbf{E}_{\rho_v}\sqrt{\theta^{\mathsf{T}}\Sigma\theta}\left|\mathrm{Quant}_{\frac{1}{2}+\varepsilon}(S_N) - \Phi^{-1}(3/4)\right| \\ &\leqslant \sup_{v\in S^{d-1}} \mathbf{E}_{\rho_v}\sqrt{\theta^{\mathsf{T}}\Sigma\theta}\left(\left|\mathrm{Quant}_{\frac{1}{2}+\varepsilon}(S_N) - \Phi_{\mathrm{H}}^{-1}(1/2+\varepsilon)\right| - \mathbf{E}\left|\mathrm{Quant}_{\frac{1}{2}+\varepsilon}(S_N) - \Phi_{\mathrm{H}}^{-1}(1/2+\varepsilon)\right|\right) \\ &\quad + \sup_{v\in S^{d-1}} \mathbf{E}_{\rho_v}\mathbf{E}\sqrt{\theta^{\mathsf{T}}\Sigma\theta}\left|\mathrm{Quant}_{\frac{1}{2}+\varepsilon}(S_N) - \Phi_{\mathrm{H}}^{-1}(1/2+\varepsilon)\right| \\ &\quad + \sup_{v\in S^{d-1}} \mathbf{E}_{\rho_v}\sqrt{\theta^{\mathsf{T}}\Sigma\theta}\left|\Phi^{-1}(3/4) - \Phi_{\mathrm{H}}^{-1}(1/2+\varepsilon)\right| \\ &= (I) + (II) + (III).\end{aligned}$$

我们希望应用引理 4 来控制 (I)。固定 $\lambda > 0$，并令 γ 为 $\mathbb{R}^d$ 上的一个零均值且协方差为 $\beta^{-1} I_d$的多元高斯分布。引理 5意味着，对于我们选择的参数 $\mathcal{KL}(\rho_v, \gamma) \leqslant \log(2) + \beta/2$。记

$$Q(S_N) = \left|\mathrm{Quant}_{\frac{1}{2}+\varepsilon}(S_N) - \Phi_{\mathrm{H}}^{-1}(1/2+\varepsilon)\right| - \mathbf{E}\left|\mathrm{Quant}_{\frac{1}{2}+\varepsilon}(S_N) - \Phi_{\mathrm{H}}^{-1}(1/2+\varepsilon)\right|.$$

注意到，在给定 θ 的条件下，随机变量 $Q(S_N)$ 是随机变量 $|\mathrm{Quant}_{\frac{1}{2}+\varepsilon}(S_N) - \Phi^{-1}{}_{\mathrm{H}}(1/2+\varepsilon)|$ 的中心化版本，其 $\|\cdot\|_{\psi_2}$ 由引理 3 控制。由于中心化使 ψ_2-范数至多乘上一个绝对常数因子，因此我们有（在给定 θ 的条件下）$\|Q(S_N)\|_{\psi_2} \leqslant \frac{c_2}{\sqrt{N}}$，其中某个绝对常数为 $c_2 > 0$。由引理 4 可知，对所有 $v \in S^{d-1}$，以下结论同时以至少 $1-\delta$ 的概率成立：

$$\lambda \mathbf{E}_{\rho_v} \sqrt{\theta^{\mathsf{T}}\Sigma\theta}\, Q(S_N) \leqslant \mathbf{E}_{\rho_v} \log \mathbf{E} \exp\left(\lambda\sqrt{\theta^{\mathsf{T}}\Sigma\theta}\, Q(S_N)\right) + \beta/2 + \log(2/\delta).$$

因此，根据 [Ver16、命题 2.5.2 的 (v)]（在给定 θ 的条件下，在此结果中取 $\lambda\sqrt{\theta^{\mathsf{T}}\Sigma\theta}$ 而非 λ），重复定理 1 证明中的步骤，可得某些绝对常数c_3，$c_4 > 0$,

$$\mathbf{E}_{\rho_v} \log \mathbf{E} \exp\left(\lambda\sqrt{\theta^{\mathsf{T}}\Sigma\theta}\, Q(S_N)\right) \leqslant \mathbf{E}_{\rho_v} \frac{c_3\lambda^2\theta^{\mathsf{T}}\Sigma\theta}{N} \leqslant \frac{c_4\lambda^2\|\Sigma\|}{N}.$$

结合这些界并对 λ 进行优化，我们以至少 $1-\delta$ 的概率，同时对所有 $v \in S^{d-1}$ 有，

$$(I) \leqslant c_5\|\Sigma\|^{1/2}\left(\sqrt{\frac{\mathbf{r}(\Sigma)}{N}} + \sqrt{\frac{\log(1/\delta)}{N}}\right),$$

其中 $c_5 > 0$ 是某个绝对常数。现在我们来界定项 (II)。与定理 1的证明类似，我们使用引理 3 得到：对于某个绝对常数 $c_6 > 0$，有以下界

$$\mathbf{E}_{\rho_v}\mathbf{E}\sqrt{\theta^{\mathsf{T}}\Sigma\theta}\left|\mathrm{Quant}_{\frac{1}{2}+\varepsilon}(S_N) - \Phi_{\mathrm{H}}^{-1}(1/2+\varepsilon)\right| \leqslant c_6\sqrt{\frac{\|\Sigma\|}{N}}.$$

为了界定 (III)，我们首先注意到 $\Phi^{-1}{}_{\mathrm{H}}(1/2) = \Phi^{-1}(3/4)$。现在我们证明，差值 $\Phi^{-1}{}_{\mathrm{H}}(1/2+\varepsilon) - \Phi^{-1}{}_{\mathrm{H}}(1/2)$ 对于 $\varepsilon \in [0, 1/4]$ 有界于 ε（相差一个乘法常数）。与定理 1 的证明中用于标准高斯分布的分位数函数的论证类似，我们计算并界定 $\Phi^{-1}{}_{\mathrm{H}}(1/2+x)$ 在 $x \in [0, 1/4]$ 时的导数如下

$$\frac{d}{dx}\Phi_{\mathrm{H}}^{-1}(1/2+x) = \sqrt{\frac{\pi}{2}}\exp\left((\Phi_{\mathrm{H}}^{-1}(1/2+x))^2/2\right) \leqslant \sqrt{\frac{\pi}{2}}\exp\left((\Phi_{\mathrm{H}}^{-1}(3/4))^2/2\right) \leqslant 3.$$

因此，对于某个 $c_7 > 0$,

$$\mathbf{E}_{\rho_v}\sqrt{\theta^{\mathsf{T}}\Sigma\theta} \cdot \left|\Phi_{\mathrm{H}}^{-1}(1/2+\varepsilon) - \Phi_{\mathrm{H}}^{-1}(1/2)\right| \leqslant c_7\varepsilon\sqrt{\|\Sigma\|}.$$

结合得到的界，我们完成了证明。 □

我们估计量的统计最优性。 下面简要讨论我们所声称的界的最优性。最优性结论可直接由已有下界推出。 [CGR18中的界、定理 2.2 和定理 32] 表明，定理 1 和定理 2 对于污染水平的依赖均为最优，并且具有正确的与维数无关的参数速率。对于协方差估计，其余项的最优性已在 [AZ22第 5]节中详细描述。均值估计问题的匹配下界见[LM19b]。

6 调节未知参数

我们现在重点讨论估计量中若干参数的调节。为简单起见，我们假设 ε 已知，或者已知其一个上界 ε_0 ，使得$\varepsilon \leqslant \varepsilon_0 < 1/2$。这是文献中的标准假设 [LM21]。请注意，至少在均值估计中，ε 的取值仅用于调节参数 β。估计其他参数最常见的方法是样本拆分：将样本划分为若干个大小相同的独立样本块。对于每个数据块，我们都可以界定其中污染点的数量。这样，我们就能针对任意 $\varepsilon \in [0, c]$陈述结果，其中 c 是一个足够小的绝对常数。我们的分析有一个有趣之处：可以在同一样本上调节不同的参数。下面将对此进行更详细的讨论。

处理依赖关系。显然，在强污染设定下，对手可以使上述数据块彼此依赖。也就是说，任何子样本中的离群值都可能依赖于整个样本。一些作者隐含地假设，对样本进行拆分会得到相互独立的子样本。例如，[LM21 中对截尾均值估计量的分析，定理 1] 利用了这种独立性。在 Huber 污染模型中，这种独立性成立，但在一般的强污染模型中并不成立。在强污染模型下处理样本拆分步骤，需要进行一些额外的稳定性分析。关于这种方法的一个例子，参见 [DKLP22，第 6] 节。

现在我们证明，我们的方法可以在同一样本上调节参数。因此，我们的结果在强污染模型下无需额外假设即可成立。为清晰起见，我们仅关注均值估计问题。假设给定一个大小为 N 的 ε 污染样本，记为 S_N。给定 S_N，我们首先找到一个整数 $\beta = \beta(S_N)$，使得其以至少 $1 - \delta/2$ 的概率满足，

$$\mathbf{r}(\Sigma)/10 \leqslant \beta(S_N) \leqslant 10\mathbf{r}(\Sigma). \tag{11}$$

随后，我们在同一个样本 S_N 上用 $\beta = \beta(S_N)$ 计算式（2）定义的估计量。将式（11）成立的事件记为 E。我们将说明，由于引理 4 的性质，这种依赖不会带来额外的技术问题。首先注意到，由于 β 是整数，我们可以对至多 $10\mathbf{r}(\Sigma)$ 个先验高斯分布 γ 使用并集界，以处理 β 对 S_N 的潜在依赖。可以验证，应用并集界不会改变定理 1 的界。重要的是，引理 4 的结果对于后验分布 ρ_v 是一致的，并且允许 β 依赖于样本，只要 $\mathcal{KL}(\rho_v, \gamma) = \beta(S_N)/2 \leqslant 5\mathbf{r}(\Sigma)$，而这在事件 E 上成立。最后，可以很容易验证，在同一事件 E 上，对项的所需上界

$$\mathbf{E}_{\rho_v} \log \mathbf{E} \exp\left(\lambda\sqrt{\theta^\top \Sigma \theta}\left(\mathrm{Quant}_{\frac{1}{2}\pm\varepsilon}(S_N) - \mathbf{E}\,\mathrm{Quant}_{\frac{1}{2}\pm\varepsilon}(S_N)\right)\right),$$

定理 1 的证明中出现的 $\beta = \beta(S_N)$ 不会受到影响。这一论证使我们能够在估计量中使用 $\beta(S_N)$。

类似的思路也可以应用于协方差估计问题。为避免不必要的技术细节，我们假设在协方差估计中，确实可以将样本划分为若干数据块，且对手不得在这些数据块之间建立依赖关系。这涵盖了许多标准污染模型，包括 Huber 的 ε-污染模型。

估计 β 和 ω。该步骤可由已有结果推出。特别地，在高斯情形下，[AZ22] 中的命题 6 给出了一个估计量 ω，满足 $\|\Sigma\|/4 \leqslant \omega \leqslant 4\|\Sigma\|$ 只要$N \geqslant c(\mathbf{r}(\Sigma) + \log(1/\delta))$，其中 $c > 0$ 是一个绝对常数。我们还需要估计$\mathrm{Tr}(\Sigma)$。

该问题可归结为均值估计。由于我们只需知道 ε 的线性依赖不会发挥任何作用，因为我们只需要知道$\mathrm{Tr}(\Sigma)$，精度达到一个乘法常数因子即可。特别地，可以使用任意一个次高斯均值估计器，在$\mathbb{R}$ 中（参见 [LM19a] 以获取精确定义），该估计器能够容忍强污染，并给出关于污染水平的 $\sqrt{\varepsilon}$依赖关系，以找到τ，满足$\mathrm{Tr}(\Sigma)/2 \leqslant \tau \leqslant 2\mathrm{Tr}(\Sigma)$，只要 ε 足够小且 $N \geqslant c\log(1/\delta)$。由此我们可以找到一个整数 β，满足（11）。

构造矩阵 G。我们讨论如何构造一个满足以下条件的半正定矩阵G

$$\Sigma \preceq 10G, \quad \text{and} \quad \mathrm{Tr}(G) \leqslant 10\,\mathrm{Tr}(\Sigma). \tag{12}$$

以下结果表明，只要 $N \geqslant c(d+\log(1/\delta))$，我们就能高效地构造这样的矩阵，其中 $c>0$ 是某个绝对常数。

命题 1。存在绝对常数 c，$c_1>0$，使得以下结论成立。设X 是一个取值于 $\mathbb{R}^d$、协方差为 Σ 的高斯零均值向量。令 X_1，……，X_N 为 ε-污染的 X独立副本集合。固定 $\delta \in (0,1)$。假设 $\varepsilon \leqslant c$ 且 $N \geqslant c_1(d+\log(1/\delta))$。那么，以至少 $1-\delta$ 的概率，对于所有满足 $|I'| = cN$ 的 $I' \subseteq [N]$，以下结论同时成立

$$\Sigma \preceq \frac{10}{N}\sum_{i\in[N]\setminus I'} X_i X_i^{\mathsf{T}}.$$

此外，在同一事件上，存在 $I \subseteq [N]$ 使得 $|I| = cN$，并且

$$\frac{1}{N}\sum_{i\in[N]\setminus I} \|X_i\|^2 \leqslant 10\,\mathrm{Tr}(\Sigma).$$

该结果立即表明，矩阵 $G = \frac{1}{N}\sum_{i\in[N]\setminus I} X_i X_i^{\mathsf{T}}$满足所需性质（12）。要找到这个集合，只需找到一个大小为 I 的集合 εN，使得$\sum_{i\in[N]\setminus I}\|X_i\|^2 \leqslant 10N$ 迹(Σ)。只需移除范数最大的 εN 个向量即可做到。

证明。不失一般性，假设 cN是一个整数。任取一个大小为 $I \subset [N]$ 的 $2cN$。令 $Y_1,\ldots,Y_N$ 表示一个未受污染样本。此类子集的总数至多为 $\binom{N}{2cn} \leqslant \left(\frac{2e}{c}\right)^{2cN}$。根据 Oliveira [Oli16 中的定理 4.1，取 $\mathrm{h}=3$]，并对所有大小为 I 的集合应用并集界 $2cN$，可得

$$\Sigma\left(1 - 27\sqrt{\frac{d + 4cN\log(\frac{2e}{c}) + 2\log(2/\delta)}{N-2cN}}\right) \preceq \frac{1}{N-2cN}\sum_{i\in[N]\setminus I} Y_i Y_i^{\mathsf{T}}.$$

当 c 足够小时，且对于足够大的 $c_1>0$，$N \geqslant c_1(d+\log(1/\delta))$ 成立时，在同一事件上，我们有

$$\Sigma \preceq \frac{10}{N}\sum_{i\in[N]\setminus I} Y_i Y_i^{\mathsf{T}}.$$

注意，由于每一项 $Y_i Y_i^{\mathsf{T}}$都是半正定矩阵，且 $\varepsilon \leqslant c$，因此对于任意大小为 cN的 I'，都存在一个大小为 $2cN$ 的集合 I，使得

$$\sum_{i\in[N]\setminus I} Y_i Y_i^{\mathsf{T}} \preceq \sum_{i\in[N]\setminus I'} X_i X_i^{\mathsf{T}}.$$

事实上，要构造这样的集合 I，我们考虑将污染点集合与集合 I' 取并集（如果该并集的基数小于 $2cN$，我们可以添加任意其他元素）。这意味着，在我们的假设下，对于所有大小为 cN的 $I' \subset [N]$，其概率至少为 $1-\delta$，

$$\Sigma \preceq \frac{10}{N} \sum_{i \in [N] \setminus I'} X_i X_i^{\mathsf{T}}.$$

现在我们考虑该陈述的第二部分。结合高斯集中不等式 [BLM13，例 5。7] 和 [Ver16，命题 2.5。2], 我们得到，存在一个绝对常数 $c_2 > 0$ 使得

$$\big\| \|X\| - \mathbf{E}\|X\| \big\|_{\psi_2} \leqslant c_2 \sqrt{\|\Sigma\|}.$$

现在验证 $\big\| \|X\|^2 - \mathbf{E}\|X\|^2 \big\|_{\psi_1} \leqslant c_3 \|\Sigma\|$ 是标准的，其中 $c_3 > 0$ 是一个绝对常数。根据伯恩斯坦不等式 [Ver16，定理 2.8。1] 以及并集界，对于所有 $I \subset [N]$，$|I| = cN$以至少 $1-\delta$的概率，同时存在某个绝对常数$c_4 > 0$ 使得

$$\begin{aligned} \sum_{i \in [N] \setminus I} \|Y_i\|^2 &\leqslant N \operatorname{Tr}(\Sigma) + c_4 \|\Sigma\| \left(\sqrt{N(\log(1/\delta) + cN \log(e/c))} + \log(1/\delta) + cN \log(e/c) \right) \\ &\leqslant 10N \operatorname{Tr}(\Sigma). \end{aligned}$$

最后一个不等式成立的前提是 c足够小且 c_1 足够大。我们选择 I为与污染点集合相对应的集合。对于该集合 I，在同一事件上，我们有

$$\frac{1}{N} \sum_{i \in [N] \setminus I} \|X_i\|^2 = \frac{1}{N} \sum_{i \in [N] \setminus I} \|Y_i\|^2 \leqslant 10 \operatorname{Tr}(\Sigma).$$

该结论由并集界得出。 □

估计 $\alpha(H)$。我们通过分析实值参数 $\alpha = \alpha(H)$来得出结论，该参数定义于 (5) 中。下文中，H 是一个已知的半正定矩阵。当允许对 ε存在略微次优的依赖关系时，我们可以使用$\mathbb{R}$中的截尾均值估计量分析（参见[LM21，定理 1]）。遗憾的是，当我们关注对污染水平的线性依赖时，分析会变得更加复杂。回顾一下，我们希望找到$\alpha = \alpha(H)$，使得其以至少 $1-\delta$ 的概率满足，

$$|\alpha - \operatorname{Tr}(\Sigma H)| \leqslant c \operatorname{Tr}(\Sigma H) \left(\sqrt{\frac{\mathbf{r}(\Sigma) + \log(1/\delta)}{N}} + \varepsilon \right).$$

我们给出一种估计量，它在几乎所有有意义的情形下都能达到这一误差率。更准确地说，我们将额外假设 $\delta \geqslant \exp(-\sqrt{\mathbf{r}(\Sigma)})$，或者假设 $\log d \leqslant \mathbf{r}(\Sigma)$。下文中，$e_1, \dots, e_d$ 表示 $\mathbb{R}^d$ 中的标准基。

命题 2. 存在绝对常数 $c, c_1, c_2 > 0$ 使得以下结论成立。假设 X 是 $\mathbb{R}^d$ 中的一个高斯零均值向量，其协方差为 Σ。令 $X_1, \dots, X_N$ 为 ε-污染的 X 的独立副本集合。取定 $\delta \in (0,1)$。假设 $\varepsilon \leqslant c$ 且 $N \geqslant c_1 \log(1/\delta)$。那么，以至少 $1-\delta$ 的概率，有

$$\begin{aligned} &\left| (\Phi^{-1}(3/4))^{-2} \sum_{i=1}^{d} \operatorname{Med}\left(\langle e_i, H^{1/2} X_1 \rangle^2, \dots, \langle e_i, H^{1/2} X_N \rangle^2 \right) - \operatorname{Tr}(\Sigma H) \right| \\ &\qquad \leqslant c_1 \operatorname{Tr}(\Sigma H) \left(\frac{\log(1/\delta)}{\sqrt{N}} + \varepsilon \right). \end{aligned} \tag{13}$$

此外，若 $N \geqslant c_2(\log d + \log(1/\delta))$，则在同一事件下，有

$$\left|(\Phi^{-1}(3/4))^{-2}\sum_{i=1}^{d}\operatorname{Med}\left(\langle e_i, H^{1/2}X_1\rangle^2,\ldots,\langle e_i, H^{1/2}X_N\rangle^2\right) - \operatorname{Tr}(\Sigma H)\right| \leqslant c_1 \operatorname{Tr}(\Sigma H)\left(\sqrt{\frac{\log d + \log(1/\delta)}{N}} + \varepsilon\right).$$

证明。令 $Y_1,\ldots,Y_N$ 表示未受污染样本，并令 Y 为 $\mathbb{R}^d$ 中的零均值高斯分布，其协方差为 Σ。由于 $\operatorname{Tr}(\Sigma H) = \operatorname{Tr}(H^{1/2}\Sigma H^{1/2})$，根据三角不等式以及定理 1证明中的论证，我们有

$$\begin{aligned}
&\left|\sum_{i=1}^{d}\operatorname{Med}\left(\langle e_i, H^{1/2}X_1\rangle^2,\ldots,\langle e_i, H^{1/2}X_N\rangle^2\right) - (\Phi^{-1}(3/4))^2\operatorname{Tr}(\Sigma H)\right| \\
&\leqslant \sum_{i=1}^{d}\left|\operatorname{Med}\left(\langle e_i, H^{1/2}X_1\rangle^2,\ldots,\langle e_i, H^{1/2}X_N\rangle^2\right) - (\Phi^{-1}(3/4))^2\|\Sigma^{1/2}H^{1/2}e_i\|^2\right| \\
&\leqslant \sum_{i=1}^{d}\left|\operatorname{Quant}_{1/2+\varepsilon}\left(\langle e_i, H^{1/2}Y_1\rangle^2,\ldots,\langle e_i, H^{1/2}Y_N\rangle^2\right) - (\Phi^{-1}(3/4))^2\|\Sigma^{1/2}H^{1/2}e_i\|^2\right| \\
&\quad + \sum_{i=1}^{d}\left|\operatorname{Quant}_{1/2-\varepsilon}\left(\langle e_i, H^{1/2}Y_1\rangle^2,\ldots,\langle e_i, H^{1/2}Y_N\rangle^2\right) - (\Phi^{-1}(3/4))^2\|\Sigma^{1/2}H^{1/2}e_i\|^2\right|.
\end{aligned}$$

我们只考虑第一个求和项，因为第二个求和项的分析类似。注意，由于球对称性，随机变量 $\langle e_i, H^{1/2}Y\rangle^2/\|\Sigma^{1/2}H^{1/2}e_i\|^2$ 服从 χ_1^2 分布。记

$$S_{N,i} = \{\langle e_i, H^{1/2}Y_1\rangle^2/\|\Sigma^{1/2}H^{1/2}e_i\|^2,\ldots,\langle e_i, H^{1/2}Y_N\rangle^2/\|\Sigma^{1/2}H^{1/2}e_i\|^2\}.$$

利用前一个公式中的记号、三角不等式以及 $F_{\chi_1^2}^{-1}(1/2) = (\Phi^{-1}(3/4))^2$,这一事实，我们得到

$$\begin{aligned}
&\left|\operatorname{Quant}_{1/2+\varepsilon}\left(\langle e_i, H^{1/2}Y_1\rangle^2,\ldots,\langle e_i, H^{1/2}Y_N\rangle^2\right) - (\Phi^{-1}(3/4))^2\|\Sigma^{1/2}H^{1/2}e_i\|^2\right| \\
&\leqslant \|\Sigma^{1/2}H^{1/2}e_i\|^2\left|\operatorname{Quant}_{1/2+\varepsilon}(S_{N,i}) - \mathbf{E}\operatorname{Quant}_{1/2+\varepsilon}(S_{N,i})\right| \\
&\quad + \|\Sigma^{1/2}H^{1/2}e_i\|^2\left|\mathbf{E}\operatorname{Quant}_{1/2+\varepsilon}(S_{N,i}) - F_{\chi_1^2}^{-1}(1/2+\varepsilon)\right| \\
&\quad + \|\Sigma^{1/2}H^{1/2}e_i\|^2\left|F_{\chi_1^2}^{-1}(1/2+\varepsilon) - F_{\chi_1^2}^{-1}(1/2)\right| \\
&= (I)_i + (II)_i + (III)_i.
\end{aligned}$$

根据引理 2 ，存在某个 $c_1 > 0$,

$$\|(I)_i + (II)_i\|_{\psi_1} \leqslant \frac{c_1\|\Sigma^{1/2}H^{1/2}e_i\|^2}{\sqrt{N}}, \quad \text{and therefore,} \quad \left\|\sum_{i=1}^{d}\big((I)_i + (II)_i\big)\right\|_{\psi_1} \leqslant \frac{c_1\operatorname{Tr}(\Sigma H)}{\sqrt{N}},$$

其中，最后一个表达式由三角不等式得出。利用 χ_1^2 分布的累积分布函数反函数的精确形式，以及用于界定半正态分布分位数之差的相同技巧，可以验证，对于任意 $\varepsilon \leqslant 1/4$ ，都有 $\left|F_{\chi_1^2}^{-1}(1/2 \pm \varepsilon) - F_{\chi_1^2}^{-1}(1/2)\right| \leqslant c_2\varepsilon$，其中 $c_2 > 0$ 是一个绝对常数。由此可直接得到

$$(III)_i = \|\Sigma^{1/2}H^{1/2}e_i\|^2\left|F_{\chi_1^2}^{-1}(1/2+\varepsilon) - F_{\chi_1^2}^{-1}(1/2)\right| \leqslant c_2\|\Sigma^{1/2}H^{1/2}e_i\|^2\varepsilon.$$

因此，对于 $\varepsilon \leqslant 1/4$，我们有 $\sum_{i=1}^{d}(III)_i \leqslant c_2\varepsilon$迹$(\Sigma H)$。结合以上计算，并利用 [Ver16 的尾界、命题 2.7，1], 我们证明不等式 (13)。

为了证明该界的第二部分，我们针对项 $(I)_i$ 提出一种略有不同的分析。记

$$S'_{N,i} = \{|\langle e_i, H^{1/2}Y_1\rangle|/\|\Sigma^{1/2}H^{1/2}e_i\|, \dots, |\langle e_i, H^{1/2}Y_N\rangle|/\|\Sigma^{1/2}H^{1/2}e_i\|\},$$

并注意到 $S'_{N,\ i}$ 由相互独立的半正态随机变量组成。我们有

$$(I)_i = \|\Sigma^{1/2}H^{1/2}e_i\|^2 \left|\mathrm{Quant}_{1/2+\varepsilon}\left(S'_{N,i}\right) - \sqrt{\mathbf{E}\,\mathrm{Quant}_{1/2+\varepsilon}\left(S_{N,i}\right)}\right| \times \left|\mathrm{Quant}_{1/2+\varepsilon}\left(S'_{N,i}\right) + \sqrt{\mathbf{E}\,\mathrm{Quant}_{1/2+\varepsilon}\left(S_{N,i}\right)}\right|.$$

首先，我们对上一个公式中的第二个乘数进行界定。根据引理 3，以至少 $1-\delta$ 的概率，我们有

$$\mathrm{Quant}_{1/2+\varepsilon}(S'_{N,i}) + \sqrt{\mathbf{E}\,\mathrm{Quant}_{1/2+\varepsilon}\left(S_{N,i}\right)} \leqslant 2\Phi_{\mathrm{H}}^{-1}(1/2+\varepsilon) + c_3\sqrt{\frac{\log(1/\delta)}{N}}.$$

现在注意到，在 $\varepsilon \in [0, 1/4]$ 且 $N \geqslant c_4\log(1/\delta)$ 的条件下，上一行中的最后一个表达式有一个绝对常数上界。再次使用引理 3 并结合并集界，以至少 $(I)_i$ 的概率，我们对项 $1-\delta$作如下界定

$$(I)_i \leqslant c_4\|\Sigma^{1/2}H^{1/2}e_i\|^2\sqrt{\frac{\log(1/\delta)}{N}}.$$

根据并集界，对所有 $i \in [d]$，以至少 $1-\delta$ 的概率，我们有：

$$(I)_i \leqslant c_3\|\Sigma^{1/2}H^{1/2}e_i\|^2\sqrt{\frac{\log d + \log(1/\delta)}{N}},$$

whe永不 $N \geqslant c_2(\log d + \log(1/\delta))$。对所有 $i \in [d]$ 求和，即完成证明。 □

7 总结性讨论

由此引出了几个自然的问题。首先，是否存在能够达到我们所给界限的计算高效估计量。众所周知，在强污染模型中，获得一个关于 ε呈线性依赖、且达到定理1界限的多项式时间算法是一项颇具挑战的问题，即使协方差矩阵为单位矩阵也是如此。从计算角度看，协方差估计是一个更加困难的问题。据我们所知，目前尚不清楚是否存在多项式时间算法，即使对污染水平的依赖弱得多，达到定理2的统计性能。 $\sqrt{\varepsilon}$-依赖。

一个较简单的问题是，我们的界限能否推广到高斯情形之外。答案是肯定的；我们之所以只进行显式的高斯计算，只是为了让证明更便于读者理解。具体而言，定理1 的证明仅用到了该分布的以下性质：

1. $X-\mu$ 的分布关于原点对称。

2. 该分布具有球对称性。也就是说，对于任意 $v \in S^{d-1}$，$\langle X-\mu、\ v\rangle/\sqrt{v^{\mathsf{T}}\Sigma v}$的分布都不依赖于 v。将该分布的密度函数记为 f。

3. 与密度 f相对应的累积分布函数 F 的逆函数满足 $|F^{-1}(1/2 \pm \varepsilon) - F^{-1}(1/2)| \leqslant c\varepsilon$，其中某个 $c > 0$和足够小的 ε满足条件。4. 对所有 $x \in [F^{-1}(1/2 - \varepsilon)\ F^{-1}(1/2 + \varepsilon)]$，密度函数 f 与零之间至少相差某个绝对常数。5. 与密度函数 f 相对应的分布是次高斯分布。也就是说，对于服从该分布的 Y，我们有 $\|Y\|_{\psi_2} \leqslant c$，其中某个 $c > 0$满足条件。

几乎逐字沿用我们的证明思路，可以分析这些更一般的分布。值得探究的是，能否免除次高斯尾部假设（性质 5）。事实上，假设满足性质 1-4，并且 $\Sigma = I_d$，结合我们的方法与 [DK19 命题 1 中的分析。3], 可以构造一个估计量 $\widehat{\mu}$，使其以至少 $1 - \delta$ 的概率满足

$$\|\widehat{\mu} - \mu\| \leqslant c\left(\sqrt{\frac{d + \log(1/\delta)}{N}} + \varepsilon\right),$$

只要 $N \geqslant c_1(d + \log(1/\delta))$。这里 c，$c_1 > 0$ 是一些绝对常数。Chen、Gao 和 Ren [CGR18，第4]节，也给出了一个不需要次高斯假设的类似界。在我们的情形中，应用引理 4 时，需要次高斯假设（性质 5）来控制矩母函数；而 [DK19，命题 13] 中的证明基于对 ε-网使用并集界；对于该方法，在“大偏差”区域无需次高斯尾部。

最后，若不对样本量和置信水平作出额外假设，定理 2 中的某些参数相当难以估计。或许可以采用其他方法，例如 Lepskii方法 [Lep90]，以此作为调整这些参数的另一种途径。

致谢。 作者感谢 Ankit Pensia 就不同污染模型之间的差异进行讨论，感谢 Vladimir Ulyanov 就样本分位数的渐近分布进行讨论，也感谢 Arnak Dalalyan 进行了许多富有见地的讨论并提出了有益的意见。AM 的研究工作得到了 Investissements d’Avenir (ANR-11-IDEX-0003/Labex Ecodec/ANR-11-LABX-0047) 项目和 FAST Advance资助的支持。

参考文献

[ABM19] Jason Altschuler、Victor-Emmanuel Brunel 和 Alan Malek。污染多臂老虎机中的最优臂识别。机器学习研究杂志，20(91):1–39，2019。[AC11] Jean-Yves Audibert 和 Olivier Catoni。稳健线性最小二乘回归。*The Annals of Statistics*，39(5):2766–2794，2011。[AZ22] Pedro Abdalla 和 Nikita Zhivotovskiy。协方差估计：针对对抗性污染和重尾的最优无维数依赖保证。*arXiv* 预印本 *arXiv:2205.08494*，2022。[Bah66] Raj Bahadur。关于大样本分位数的一则注记。《数理统计年鉴》，37(3):577–580，1966。[BL19] Sergey Bobkov 和 Michel Ledoux。一维经验测度、次序统计量与 *Kantorovich* 输运距离，美国数学会回忆录第 1259 卷。美国数学会回忆录。美国数学会，2019。[BLM13] Stéphane Boucheron、Gábor Lugosi 和 Pascal Massart。集中不等式：独立性的非渐近理论。牛津大学出版社，2013。[BMD22] Amir-Hossein Bateni、Arshak Minasyan 和 Arnak S. Dalalyan。通过迭代谱降维得到的近似极小极大稳健均值向量估计量。*arXiv* 预印本 *arXiv:2204.02323*，2022。

[Bor75] Christer Borell。高斯空间中的布伦-闵可夫斯基不等式。《数学新进展》，30(2):207–216，1975。[BT12] Stéphane Boucheron 和 Maud Thomas。次序统计量的集中不等式。《概率电子通讯》，17:1–12，2012。[Bur97] Marat Burnashev。参数中位数估计的渐近展开。《概率论及其应用》，41(4):632–645，1997。[Cat16] Olivier Catoni。针对格拉姆矩阵和随机设计最小二乘回归的 PAC-贝叶斯界。arXiv 预印本 arXiv:1603.05229，2016。[CDG19] Yu Cheng、Ilias Diakonikolas 和 Rong Ge。近线性时间的高维稳健均值估计。载于第三十届 ACM-SIAM 离散算法年会论文集，第 2755–2771 页。SIAM，2019。[CG17] Olivier Catoni 和 Ilaria Giulini。针对矩阵、向量和线性最小二乘回归的无维数依赖 PAC-贝叶斯界。arXiv 预印本 arXiv:1712.02747，2017。[CGR18] Mengjie Chen、Chao Gao 和 Zhao Ren。Huber 污染模型下的稳健协方差与散布矩阵估计。The Annals of Statistics，46(5):1932–1960，2018。[DG92] David L Donoho 和 Miriam Gasko。基于半空间深度和投影离群度的位置估计的崩溃性质。The Annals of Statistics，第 1803–1827 页，1992。[DHF07] Laurens De Haan 和 Ana Ferreira。《极值理论：导论》。Springer 科学与商业媒体，2007。[DK19] Ilias Diakonikolas 和 Daniel M Kane。算法高维稳健统计的最新进展。arXiv 预印本 arXiv:1911.05911，2019。[DKK+17] Ilias Diakonikolas、Gautam Kamath、Daniel M. Kane、Jerry Li、Ankur Moitra 和 Alistair Stewart。稳健（在高维情形下）也可以很实用。载于第 34 届国际机器学习会议论文集，ICML 2017，第 70 卷，第 999–1008 页，2017。[DKLP22] Ilias Diakonikolas、Daniel M Kane、Jasper CH Lee 和 Ankit Pensia。重尾分布的抗离群值稀疏均值估计。arXiv 预印本 arXiv:2211.16333，2022。[DKP20] Ilias Diakonikolas、Daniel M Kane 和 Ankit Pensia。通过稳定性实现具有次高斯速率的抗离群值均值估计。神经信息处理系统进展，33:1830–1840，2020。[DKS17] Ilias Diakonikolas、Daniel M Kane 和 Alistair Stewart。高维高斯分布和高斯混合分布稳健估计的统计查询下界。载于 2017 年 IEEE 第 58 届计算机科学基础年会（FOCS），第 73–84 页，2017。[DL22] Jules Depersin 和 Guillaume Lecué。Stahel–Donoho 均值中位数对对抗性污染和重尾数据的稳健性。《信息与推断：IMA 期刊》，12，2022。[DM22] Arnak S Dalalyan 和 Arshak Minasyan。高斯均值的一体化稳健估计量。The Annals of Statistics，50(2):1193–1219，2022。[DN04] Herbert A David 和 Haikady N Nagaraja。《次序统计量》。约翰·威立父子公司，2004。[DV75] Monroe D. Donsker 和 Srinivasa Varadhan。长时间下某些马尔可夫过程期望的渐近求值，I。《纯粹与应用数学通讯》，28(1):1–47，1975。[Giu18] Ilaria Giulini。无维数依赖的稳健格拉姆算子估计。《伯努利》，24(4B):3864–3923，2018。[Han22] Qiyang Han。样本协方差的精确谱范数误差。arXiv 预印本 arXiv:2207.13594，2022。[HL19] Samuel B Hopkins 和 Jerry Li。稳健均值估计有多难？载于学习理论会议论文集，第 1649–1682 页。PMLR，2019。[HRRS80] Frank R Hampel、Peter J Rousseeuw、Elvezio M Ronchetti 和 Werner A Stahel。稳健统计：基于影响函数的方法。威立，1980。[Hub64] Peter J Huber。位置参数的稳健估计。《数理统计年鉴》，35(1):73–101，1964。[Hub81] Peter J Huber。稳健统计。威立概率与数理统计丛书，1981。[Kie67] Jack Kiefer。关于样本分位数的 Bahadur 表示。《数理统计年鉴》，38(5):1323–1342，1967。[KL17] Vladimir Koltchinskii 和 Karim Lounici。样本协方差算子的集中不等式与矩界。《伯努利》，23(1):110–133，2017。

[KMR+19] Yuan Ke、Stanislav Minsker、Zhao Ren、Qiang Sun 和 Wen-Xin Zhou。重尾分布的用户友好型协方差估计。Statistical Science，34(3)：454–471，2019。[Kol31] 安德雷·柯尔莫哥洛夫。误差理论中的中位数方法。Rec. Math. Moscou，38(3-4)：47–50，1931。[KZ20] Yegor Klochkov 和 Nikita Zhivotovskiy。通过熵方法针对无界条目的统一 Hanson-Wright 型集中不等式。Electronic Journal of Probability，25，2020。[Lep90] Oleg V Lepskii。高斯白噪声中的自适应估计问题。Teoriya Veroyatnostei i ee Primeneniya，35(3)：459–470，1990。[LM00] Beatrice Laurent 和 Pascal Massart。通过模型选择进行二次泛函的自适应估计。Annals of Statistics，第1302–1338页，2000。[LM19a] Gábor Lugosi 和 Shahar Mendelson。重尾分布下的均值估计与回归：综述。Foundations of Computational Mathematics，19(5)：1145–1190，2019。[LM19b] Gábor Lugosi 和 Shahar Mendelson。关于一般范数的近最优均值估计量。Probability theory and related fields，175(3)：957–973，2019。[LM21] Gábor Lugosi 和 Shahar Mendelson。稳健多元均值估计：截尾均值的最优性。The Annals of Statistics，49(1)：393–410，2021。[LRV16] Kevin A. Lai、Anup B. Rao 和 Santosh S. Vempala。均值与协方差的无先验估计。载于 IEEE 57 674，2016 2016，第665届计算机科学基础年会，FOCS，第–页。[MW22] Stanislav Minsker 和 Lang Wang。协方差矩阵的稳健估计：对抗性污染及其他问题。arXiv 预印本 arXiv:2203.02880，2022。[MZ20] Shahar Mendelson 和 Nikita Zhivotovskiy。在 $L_4 - L_2$ 范数等价性下的稳健协方差估计。The Annals of Statistics，48(3)：1648–1664，2020。[Oli16] Roberto I Oliveira。随机二次型的下尾及其在普通最小二乘法中的应用。Probability Theory and Related Fields，166(3-4)：1175–1194，2016。[OR22] Roberto I Oliveira 和 Zoraida F Rico。改进的协方差估计：重尾下的最优稳健性与次高斯保证。arXiv 预印本 arXiv:2209.13485，2022。[RL05] Peter J Rousseeuw 和 Annick M Leroy。稳健回归与异常值检测，第589卷。约翰·威立父子公司，2005。[Sha03] Jun Shao。数理统计。Springer 科学与商业媒体，2003。[Tal14] Michel Talagrand。随机过程的上界与下界：现代方法与经典问题，第60卷。Springer 科学与商业媒体，2014。[TIS76] Boris S Tsirelson、Ildar A Ibragimov 和 Vladimir N Sudakov。高斯样本函数的范数。载于《第三届日苏概率论研讨会论文集》，第20–41页。Springer，1976。[Ver16] Roman Vershynin。高维概率：导论及其应用，第47卷《剑桥统计与概率数学丛书》。剑桥大学出版社，2016。[Xia19] Dong Xia。独立非同分布随机变量百分位数的非渐近界。Statistics & Probability Letters，152：111–120，2019。[Zhi21] Nikita Zhivotovskiy。通过变分原理得到的独立矩阵之和与简单张量的无维数依赖界。arXiv 预印本 arXiv:2108.08198，2021。