arXiv:2305.04330v1 [stat.ME] 7 May 2023

将仿射等变 Tyler M 估计量应用于椭圆分布的尾参数学习

Esa Ollila，高级会员，IEEE，Daniel P. Palomar，IEEE 会士，以及 Fr´ed´eric Pascal 高级会员，IEEE

摘要——我们提出通过求解散布矩阵的 Tyler M 估计量得到权重，并利用这些权重估计未指定椭圆对称分布的散布矩阵的尺度参数（特征值的均值）。随后，我们将所提出的基于 Tyler 权重的尺度估计（TWE）与归一化的 Tyler 权重结合，构造一个仿射等变 Tyler M 估计量，即加权样本协方差矩阵。接着，我们建立一个统一框架，用于估计椭圆分布的未知尾参数（例如，多元 t （MVT）分布的自由度（d.o.f.）ν ）。利用所提出的尺度 TWE，我们提出了一种新的 MVT 分布自由度参数稳健估计量，在重尾场景下表现出色，优于其他竞争方法。我们提供了实现所提方法的 R 软件包。

索引词—Tyler M 估计量、散布矩阵、协方差矩阵、椭圆分布

I. 引言

WE 模型中观测到的 p维观测值 $\mathbf{x}_1, \ldots, \mathbf{x}_n$ 服从某个未知的中心化（即关于原点对称的）椭圆对称（ES）分布 [1], [2]，且相互独立、同分布（ i.i.d. ），即为随机样本。连续随机向量 $\mathbf{x} \in \mathbb{R}^p$ 若具有如下形式的概率密度函数（pdf），则称其服从中心化的椭圆对称分布：

$$f(\mathbf{x}) = C_{p,g}|\mathbf{\Sigma}|^{-1/2} g(\mathbf{x}^\top \mathbf{\Sigma}^{-1}\mathbf{x}),$$

其中，$g: \mathbb{R}_{\geq 0} \to \mathbb{R}_{>0}$ 称为密度生成函数，$\mathbf{\Sigma} \succ 0$是正定对称矩阵参数，称为散布矩阵；$C_{p,\ g}$ 是归一化常数，确保f(x)的积分为 1。我们用 $\mathbf{x} \sim \mathcal{E}_p(0, \mathbf{\Sigma}, g)$ 表示这种情况。例如，当 $g(t) = \exp(-t/2)$时，可得到中心化多元正态（MVN）分布 $\mathcal{N}_p(0, \mathbf{\Sigma})$；而具有 $\nu > 0$ 自由度的多元 t （MVT）分布则在以下情况下得到

$$g(t) = (1 + t/\nu)^{-(p+\nu)/2}. \tag{1}$$

参数 $\nu > 0$ 是密度的尾参数。对于 $\nu \to \infty$，MVT分布退化为MVN分布，而$\nu = 1$ 对应于多元柯西分布。此外，椭圆对称分布的许多其他子类也通过一个由额外尾参数索引的密度生成函数进行参数化；该尾参数在实践中未知，需要估计。学习这一未知参数也是本文的目标之一。

Esa Ollila 任职于阿尔托大学信息与通信工程系，FI-00076 Aalto，芬兰（电子邮箱：esa.ollila@aalto.fi）。Daniel P. Palomar 任职于香港科技大学，香港（电子邮箱：palomar@ust.hk）。Fr´ed´eric Pascal 任职于法国国家科学研究中心—巴黎萨克雷大学、CentraleSup´elec，91190 Gif-sur-Yvette，法国（电子邮箱：frederic.pascal@centralesupelec.fr）。

我们关注于估计散布矩阵的尺度参数，该参数定义为其特征值的均值，

$$\eta = \frac{\mathrm{tr}(\mathbf{\Sigma})}{p} = \frac{1}{p}\sum_{i=1}^{p} \lambda_i, \tag{2}$$

其中，$\lambda_i > 0$ 表示 $\mathbf{\Sigma}$的第 i个特征值。形式上，如果 $\eta \equiv \eta(\mathbf{\Sigma})$满足 η (I) $= 1$ 且对所有$a > 0$ [3]都有 $\eta(a\mathbf{\Sigma}) = a\eta(\mathbf{\Sigma})$，则它是一个尺度参数。

Tyler M 估计量 [4] 是一种广受欢迎的散布矩阵稳健 M 估计量，在信号处理和统计学文献中均得到了广泛研究（例如，[5]–[13]）。Tyler M 估计量定义为以下不动点方程的解：

$$\hat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} \frac{p}{\mathbf{x}_i^\top \hat{\mathbf{\Sigma}}^{-1}\mathbf{x}_i}\mathbf{x}_i\mathbf{x}_i^\top \triangleq \mathcal{H}(\hat{\mathbf{\Sigma}}; \{\mathbf{x}_i\}). \tag{3}$$

请注意，我们也可以将映射 $\mathcal{H}(\cdot;\cdot)$写成以下形式：

$$\mathcal{H}(\hat{\mathbf{\Sigma}}; \{\mathbf{x}_i\}) = \frac{1}{n}\sum_{i=1}^{n} \hat{w}_i \mathbf{x}_i\mathbf{x}_i^\top \quad \text{with} \quad \hat{w}_i = \frac{p}{\mathbf{x}_i^\top \hat{\mathbf{\Sigma}}^{-1}\mathbf{x}_i},$$

其中，$\hat{}w_i,\ i = 1, \ldots, n$称为 Tyler 权重。Tyler M 估计量仅在缩放因子意义下唯一，因此通常约定考虑满足$\hat{}$tr($\mathbf{\Sigma}$) $= p$的解。因此，Tyler M 估计量实际上是形状矩阵（归一化散布矩阵） $\mathbf{\Lambda}$的估计量，其定义为 $\mathbf{\Lambda} = \mathbf{\Sigma}/\eta = p\mathbf{\Sigma}/\mathrm{tr}(\mathbf{\Sigma})$，并满足 $\mathrm{tr}(\mathbf{\Lambda}) = p$。

本文提出了一种基于 Tyler 权重 $\hat{}w_i$ 的尺度 η 估计量。随后，将所提出的尺度估计量与 $\hat{}$ Tyler M 估计量 $\mathbf{\Sigma}$ 联合用于构造散布矩阵 $\mathbf{\Sigma}$ 和协方差矩阵 R = cov(x) 的仿射等变鲁棒估计量（或其收缩版本）。这些研究成果将在第二节中介绍。然后，在第三节中，我们提出了一个统一框架，利用所提出的尺度统计量 $\hat{}\eta$ 估计椭圆分布的尾参数。在多元 t 分布的情况下，这将得到一种基于 Tyler 权重的新自由度参数估计量。最后，第四节通过模拟数据展示了所提出方法的有效性，并在第五节给出总结性评论。在 R软件包fitHeavyTail[14]，中，函数 fit_Tyler 实现了该方法。

II. 基于 Tyler 权重的尺度、散布和协方差矩阵估计

假设 $\mathbf{x} \sim \mathcal{E}_p(0, \mathbf{\Sigma}, g)$ 具有有限的二阶矩，则其协方差矩阵 $\mathbf{R} = \mathbb{E}[\mathbf{x}\mathbf{x}^\top]$ 满足

$$\mathbf{R} = \theta \cdot \mathbf{\Sigma} \quad \text{for} \quad \theta = \frac{\mathbb{E}[r^2]}{p}, \tag{4}$$

其中，$r^2 = \|\mathbf{\Sigma}^{-1/2}\mathbf{x}\|^2$ 是 x 相对于 $\mathbf{\Sigma}$ 的马氏距离平方，其概率密度函数为

$$f_{r^2}(t) = Ct^{p/2-1}g(t), \tag{5}$$

其中 $C = \int_0^\infty t^{p/2-1}g(t)\mathrm{d}t$。因此，$r^2$ 的概率密度函数与密度生成函数 g 一一对应。由 (4) 可知，散布矩阵 $\mathbf{\Sigma}$ 与协方差矩阵 R 成比例（假设 R 存在）。在 MVN 情形下，$\theta = 1$；而对于密度生成函数如 (1) 所示的多元 t 分布，对所有 $\nu > 2$ 均有 $\theta = \nu/(\nu-2)$。

A. 尺度估计量

$\hat{}$ 如前所述，Tyler M 估计量 $\mathbf{\Sigma}$ 会丢失尺度 η 的信息。然而，仍可以利用 Tyler 权重构造 η 的估计量。我们的基于 Tyler 权重的尺度估计量（TWE）定义为权重倒数的调和平均数，$1/\hat{}w_i$，即

$$\hat{\eta}_{\mathrm{TWE}} = \left(\frac{1}{n}\sum_{i=1}^n \hat{w}_i\right)^{-1} = \left(\frac{p}{n}\sum_{i=1}^n [\mathbf{x}_i^\top \hat{\mathbf{\Sigma}}^{-1}\mathbf{x}_i]^{-1}\right)^{-1}. \tag{6}$$

很容易验证，这是一个适当的尺度估计量：如果在缩放后的观测值 $\mathbf{x}_i^* = c\cdot\mathbf{x}_i, i = 1,\ldots,n$ 上计算 $\hat{}\eta^*_{\mathrm{TWE}}$，那么 $\hat{}\eta^*_{\mathrm{TWE}} = c^2\cdot\hat{}\eta_{\mathrm{TWE}}$。这是因为带迹约束的 Tyler M 估计量 $\mathbf{\Sigma}$（见 (3)）对数据缩放具有不变性，因此 $\hat{}\mathbf{\Sigma}^* = \hat{}\mathbf{\Sigma}$。

所提出的估计量 (6) 也可以由高维随机矩阵理论（RMT）范围内推导出的以下结果来解释，其中 $p, n \to \infty$ 满足 $n > p$，且其比值趋于常数：$p/n \to c \in (0,1)$。具体而言，设 $\hat{}\mathbf{\Sigma}$ 是 (3) 中满足 $\mathrm{tr}(\hat{}\mathbf{\Sigma}) = p$ 的 Tyler M 估计量。文献 [11], [13] 已证明，$\max_\ell |\eta\,\hat{}w_\ell - 1| \to 0$ 几乎必然成立。文献 [11] 的作者针对数据服从独立同分布的高斯分布 $\mathcal{N}_p(0, \mathbf{I})$ 的情形推导了这一结果，而 [13] 则将这些结果推广到了更一般的分布。因此，由于 $1/\hat{}w_i$ 集中于 η，(6) 是一种自然的稳健尺度估计量。基于 $1/\hat{}w_i$，还可以构造许多其他稳健尺度统计量，例如中位数、截尾均值等。在下一小节中，我们将说明为何所提出的调和平均数 (6) 是最自然的选择。

B. 散布矩阵的仿射等变估计量

$\hat{}$ 使用尺度估计量 $\hat{}\eta_{\mathrm{TWE}}$ 和 Tyler M 估计量 $\mathbf{\Sigma}$（满足 $\mathrm{tr}(\hat{}\mathbf{\Sigma}) = p$），我们可以构造散布矩阵的估计量

$$\hat{\mathbf{\Sigma}}_{\mathrm{TWE}} = \hat{\eta}_{\mathrm{TWE}}\cdot\hat{\mathbf{\Sigma}} \tag{7}$$

称为散布矩阵的 TWE。因此，$\hat{}\eta_{\mathrm{TWE}}$ 是由 $\hat{}\mathbf{\Sigma}_{\mathrm{TWE}}$ 得出的尺度统计量，因为 $\hat{}\eta_{\mathrm{TWE}} = \mathrm{tr}(\hat{}\mathbf{\Sigma}_{\mathrm{TWE}})/p$。$\hat{}$ 等价地，$\mathbf{\Sigma}_{\mathrm{TWE}}$ 的迹可以方便地计算为 Tyler 二次型的调和平均数：

$$\mathrm{tr}(\hat{\mathbf{\Sigma}}_{\mathrm{TWE}}) = \left(\frac{1}{n}\sum_{i=1}^n [\mathbf{x}_i^\top \hat{\mathbf{\Sigma}}^{-1}\mathbf{x}_i]^{-1}\right)^{-1}.$$

回顾式 (6)，我们可以将式 (7) 写成以下更直观的形式：

$$\hat{\mathbf{\Sigma}}_{\mathrm{TWE}} = \frac{1}{n}\sum_{i=1}^n \hat{v}_i\mathbf{x}_i\mathbf{x}_i^\top, \tag{8a}$$

$$\hat{v}_i = \frac{\hat{w}_i}{\frac{1}{n}\sum_{\ell=1}^n \hat{w}_\ell} = \frac{[\mathbf{x}_i^\top \hat{\mathbf{\Sigma}}^{-1}\mathbf{x}_i]^{-1}}{\frac{1}{n}\sum_{\ell=1}^n [\mathbf{x}_\ell^\top \hat{\mathbf{\Sigma}}^{-1}\mathbf{x}_\ell]^{-1}} \tag{8b}$$

其中，$\hat{}v_1, \ldots, \hat{}v_n$ 是归一化的 Tyler 权重，满足 $\frac{1}{n}\sum_{i=1}^n \hat{}v_i = 1$。式 (8a) 和 (8b) 表明，$\hat{}\mathbf{\Sigma}_{\mathrm{TWE}}$ 是加权样本协方差矩阵（SCM），其权重为 $\hat{}v_i$。

最后，我们将 Tyler的M估计方程 与我们的估计量 (8) 进行类比。首先需要注意的是，Tyler的M估计方程 (3) 所满足的不变性更广，不仅仅是关于数据矩阵缩放的不变性。也就是说，将单位范数归一化数据记为 $\tilde{}\mathbf{x}_i = \mathbf{x}_i/\|\mathbf{x}_i\|, i = 1,\ldots,n$，容易验证，(3) 中的不动点方程可改写为 $\hat{}\hat{}\ \mathbf{\Sigma} = \mathcal{H}(\mathbf{\Sigma}; \{\tilde{}\mathbf{x}_i\})$，因此是基于归一化数据的。此外，由于 $\mathrm{tr}(\hat{}\mathbf{\Sigma}) = p$，可得：

$$\mathrm{tr}(\hat{\mathbf{\Sigma}}) = \frac{1}{n}\sum_{i=1}^n \frac{p}{\tilde{\mathbf{x}}_i^\top \hat{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{x}}_i}\,\mathrm{tr}\left(\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^\top\right) = p,$$

或者等价地，$\frac{1}{n}\sum_{i=1}^n [\tilde{}\mathbf{x}_i^\top \hat{}\mathbf{\Sigma}^{-1}\tilde{}\mathbf{x}_i]^{-1} = 1$，因为 $\mathrm{tr}\left(\tilde{}\mathbf{x}_i\tilde{}\mathbf{x}_i^\top\right) = \hat{}\hat{}\,1$。由此可知，Tyler M 估计量 $\mathbf{\Sigma}$ 在 $\mathrm{tr}(\mathbf{\Sigma}) = p$ 的约束下，是以下不动点方程的解：

$$\hat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{i=1}^n \tilde{w}_i\tilde{\mathbf{x}}_i\tilde{\mathbf{x}}_i^\top \quad \text{with} \quad \tilde{w}_i = \frac{p\,[\tilde{\mathbf{x}}_i^\top \hat{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{x}}_i]^{-1}}{\frac{1}{n}\sum_{\ell=1}^n [\tilde{\mathbf{x}}_\ell^\top \hat{\mathbf{\Sigma}}^{-1}\tilde{\mathbf{x}}_\ell]^{-1}}, \tag{9}$$

其中，$\tilde{w}_\ell$ 是在归一化（单位范数）观测值 $\tilde{}\mathbf{x}_i$ 上计算得到的 Tyler 归一化权重。因此，尽管 Tyler 的 $\hat{}$M-estimator $\mathbf{\Sigma}$ 可以解释为基于归一化数据 $\{\tilde{}\mathbf{x}_i\}$ 的加权 SCM（如式（9）所示），但式（8）中的散布矩阵 TWE $\hat{}\mathbf{\Sigma}_{\mathrm{TWE}}$ 可以视为基于实际（未经归一化）数据 $\{\mathbf{x}_i\}$ 的加权 SCM。

值得指出的是，大多数稳健散布估计量都是仿射等变的，也就是说，对数据 $\mathbf{x}_i \mapsto \mathbf{A}\mathbf{x}_i, i = 1,\ldots,n$ 进行仿射变换，会使估计值发生如下变换：

$$\hat{\mathbf{\Sigma}}(\{\mathbf{A}\mathbf{x}_i\}) = \mathbf{A}\hat{\mathbf{\Sigma}}(\{\mathbf{x}_i\})\mathbf{A}^\top,\ \forall \mathbf{A} \in \mathbb{R}^{p\times p} \text{ invertible.} \tag{10}$$

例如，稳健的 Maronna [15] M估计量、S估计量 [16]，或 MM估计量 [17] 都是仿射等变的。然而，Tyler散布矩阵 $\mathbf{\Sigma}$ 不是仿射等变的，因为如式（11）所示，式（10）仅在相差一个乘法标量因子的意义下成立。仿射等变性是理想的，因为若 $\mathbf{x} \sim \mathcal{E}_p(0, \mathbf{\Sigma}, g)$，则 $\mathbf{A}\mathbf{x} \sim \mathcal{E}_p(0, \mathbf{A}\mathbf{\Sigma}\mathbf{A}^\top, g)$。因此，变换后的数据仍服从相同的椭圆分布，但散布矩阵参数会映射为 $\mathbf{A}\mathbf{\Sigma}\mathbf{A}^\top$。由此，对任何散布矩阵估计量，一个自然的要求是：在变换 $\mathbf{x}_i \mapsto \mathbf{A}\mathbf{x}_i$ 下，它应满足同样的等变原理。下面将对此进行说明。

引理 1. 散布矩阵的 TWE $\mathbf{\Sigma}_{\mathrm{TWE}}$ 是仿射等变的：$\mathbf{\Sigma}_{\mathrm{TWE}}(\{\mathbf{A}\mathbf{x}_i\}) = \mathbf{A}\,\mathbf{\Sigma}(\{\mathbf{x}_i\})\mathbf{A}^\top, \forall$ 可逆矩阵 $\mathbf{A} \in \mathbb{R}^{p\times p}$。

证明。很容易验证，Tyler M 估计量（满足 $\mathrm{tr}(\hat{}\mathbf{\Sigma}) = p$）具有如下等变性：若 $\hat{}\mathbf{\Sigma}^* =$

$\hat{\mathbf{\Sigma}}(\{\mathbf{x}_i^*\})$ 表示在数据 $\mathbf{x}_i^* = \mathbf{A}\mathbf{x}_i$, $i = 1, \ldots, n$ 上计算得到的 Tyler M 估计量（满足 $\mathrm{tr}(\hat{\mathbf{\Sigma}}^*) = p$），则

$$\hat{\mathbf{\Sigma}}^* = c\mathbf{A}\hat{\mathbf{\Sigma}}\mathbf{A}^\top, \; c = \frac{p}{\mathrm{tr}(\mathbf{A}\hat{\mathbf{\Sigma}}\mathbf{A}^\top)} \tag{11}$$

对于所有可逆矩阵 $\mathbf{A} \in \mathbb{R}^{p\times p}$。现在令 $\hat{w}_i^*$, $i = 1, \ldots, n$ 表示相应的 Tyler 权重。那么，将尺度估计量 (6) 计算为权重 $1/\hat{w}_i^* = [\mathbf{x}_i^*]^\top(\hat{\mathbf{\Sigma}}^*)^{-1}\mathbf{x}_i^*$ 的倒数的调和平均数，即为

$$\begin{aligned}\hat{\eta}_{\mathrm{TWE}}^* &= \left(\frac{1}{n}\sum_{i=1}^n \hat{w}_i^*\right)^{-1} = \left(\frac{p}{n}\sum_{i=1}^n [\mathbf{x}_i^\top\mathbf{A}^\top(\hat{\mathbf{\Sigma}}^*)^{-1}\mathbf{A}\mathbf{x}_i]^{-1}\right)^{-1} \\ &= c^{-1}\left(\frac{p}{n}\sum_{i=1}^n [\mathbf{x}_i^\top\hat{\mathbf{\Sigma}}^{-1}\mathbf{x}_i]^{-1}\right)^{-1} = c^{-1}\hat{\eta}_{\mathrm{TWE}}\end{aligned} \tag{12}$$

其中，在倒数第二个等式中，我们仅使用了（11）。因此可得

$$\begin{aligned}\hat{\mathbf{\Sigma}}_{\mathrm{TWE}}^* &= \hat{\eta}_{\mathrm{TWE}}^* \cdot \hat{\mathbf{\Sigma}}^* = c^{-1}\hat{\eta}_{\mathrm{TWE}} \cdot c\mathbf{A}\hat{\mathbf{\Sigma}}\mathbf{A}^\top \\ &= \mathbf{A}(\hat{\eta}_{\mathrm{TWE}}\hat{\mathbf{\Sigma}})\mathbf{A}^\top = \mathbf{A}\hat{\mathbf{\Sigma}}_{\mathrm{TWE}}\mathbf{A}^\top.\end{aligned}$$

□

对于高维数据的情况，也可以考虑散布矩阵的收缩 Tyler M 估计量，作为

$$\hat{\mathbf{\Sigma}}_{\mathrm{TWE},\beta} = \beta\hat{\mathbf{\Sigma}}_{\mathrm{TWE}} + (1-\beta)\hat{\eta}_{\mathrm{TWE}}\mathbf{I}, \tag{13}$$

其中，数据自适应收缩参数 $\beta \in [0,1]$ 的计算方法如[18, 第 IV.C 节]所述。然而，与 [18], 中的估计量不同，式 (13) 中的收缩 TWE 提供的是散布估计量，而非形状矩阵估计量。

C. 协方差矩阵的估计量

若密度生成函数 g （因而底层椭圆对称分布）已知，则可确定式（4）中的 θ ，并可利用关系式（4）得到协方差矩阵估计量 $\mathbf{R}_{\mathrm{TWE}} = \theta \cdot \mathbf{\Sigma}_{\mathrm{TWE}}$. 。例如，若数据服从 MVN 分布，则 $\theta = 1$；而在自由度为 ν 的多元 t 分布情况下，$\theta = \nu/(\nu-2)$ 。然而，底层参数族通常是已知的，但密度生成函数所对应的底层尾参数（记为 ν）未知。如第三节所示，我们可以利用 Tyler 权重估计 ν，记为$\hat{\nu}_{\mathrm{TWE}}$。由于 $\theta = h(\nu)$（参见下文式（15）），可计算协方差矩阵的 TWE: $\mathbf{R}_{\mathrm{TWE}} = \theta_{\mathrm{TWE}}\cdot\mathbf{\Sigma}_{\mathrm{TWE}}$，其中 $\theta_{\mathrm{TWE}} = h(\hat{\nu}_{\mathrm{TWE}})$ 。

III. 椭圆对称分布的尾参数估计

由（4）可推导出协方差矩阵的尺度参数 $\eta_{\mathrm{cov}} = p^{-1}\mathrm{tr}(\mathbf{R})$与散布矩阵的尺度 $\eta = p^{-1}\mathrm{tr}(\mathbf{\Sigma})$之间的如下关系：

$$\eta_{\mathrm{cov}} = \theta\eta \Leftrightarrow \theta = \eta_{\mathrm{cov}}/\eta. \tag{14}$$

注意到， η_{cov} 的一个自然估计量是 $p^{-1}\mathrm{tr}(\mathbf{S})$，其中 $\mathbf{S} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top$表示样本协方差矩阵（SCM）。另一方面，如果密度生成函数 g 仅确定到未知尾参数 ν，因此记为 $g_\nu(\cdot)$，那么式 (4) 中的 θ 是尾参数 ν的如下函数：

$$\theta = \int_0^\infty t f_{r^2}(t;\nu)\mathrm{d}t \triangleq h(\nu) \tag{15}$$

其中，定义于（5）中的概率密度函数$f_{r^2}(\cdot)$与 $g_\nu(\cdot)$之间是一一对应关系。在大多数实际情况下，我们不需要进行数值积分，因为通常可以推导出 $h(\nu)$的闭式表达式。然后，在求解逆映射 $\nu = h^{-1}(\theta) = h^{-1}(\eta_{\mathrm{cov}}/\eta)$后，算法1提供了一种估计椭圆对称分布尾参数的统一方法：

算法1：分布尾参数学习

输入：数据 $\{\mathbf{x}_i\}_{i=1}^n$
输出：估计的尾参数 ν
1. 计算 Tyler M 估计量 $\hat{\mathbf{\Sigma}}$ 以及式 (3) 中的权重 $\hat{w}_i$;
2. 计算（6）中的 $\hat{\eta}_{\mathrm{TWE}}$，并设置 $\hat{\theta}_{\mathrm{TWE}} = p^{-1}\mathrm{tr}(\mathbf{S})/\hat{\eta}_{\mathrm{TWE}}$
3. 利用 (15)，将 ν 估计为$\hat{\nu}_{\mathrm{TWE}} = h^{-1}(\hat{\theta}_{\mathrm{TWE}})$.

例如，如果 x 服从自由度为 ν >2 的多元 t 分布，则有 $\theta = h(\nu) = \frac{\nu}{\nu-2}$，由此可得如下关系：

$$\nu = h^{-1}(\theta) = \frac{2\theta}{\theta-1} \text{ for } \theta > 1. \tag{16}$$

注意，只有满足 $\nu > 2$ ，协方差矩阵 R 才存在。所得估计量 $\hat{\nu}_{\mathrm{TWE}}$ 与 [18, 算法 1], 中的估计量密切相关，简称 OPP 估计量。OPP 是一种迭代方法：根据自由度参数 $\nu^{(k)}$的当前估计值，迭代地（重新）计算自由度为 ν 的多元 t 分布的最大似然估计量（MLE） $\mathbf{\Sigma}$ 。随后计算 $\hat{\theta} = \mathrm{tr}(\mathbf{S})/\mathrm{tr}(\hat{\mathbf{\Sigma}})$，并通过式（16）更新 $\nu^{(k+1)} = h^{-1}(\hat{\theta})$。算法从初始值 $\nu^{(0)} = \hat{\nu}_{\mathrm{kurt}}$开始，对 $k = 0$、1、2.....进行迭代，直至收敛。其中，$\hat{\nu}_{\mathrm{kurt}}$ 是基于椭圆峰度的 ν估计值，该估计量由 [18], 提出，并称为峰度估计量。我们还在 [19], 中提出了 OPP 估计量的改进版本，但对于较大的 n 和 p而言，该版本并不实用。文献 [20]考虑了通过期望最大化（EM）方法对 ν 进行极大似然估计。然而，这种方法相当不稳定 [21]。

IV. 仿真研究

我们首先考虑散布矩阵 $\mathbf{\Sigma}$ 具有自回归模型（一阶自回归）结构的情形，即 $(\mathbf{\Sigma})_{ij} = \eta\varrho^{|i-j|}$，其中 $\eta = \mathrm{tr}(\mathbf{\Sigma})/p$ 是尺度参数， ϱ 是相关参数， $\varrho \in (-1,1)$。由于 Tyler M 估计量 $\hat{\mathbf{\Sigma}}$ 对数据缩放具有不变性，因此我们可以设定 $\eta = 1$ ，而不会偏向任何一个估计量。蒙特卡洛运行次数为 5,000，并从具有不同自由度参数 ν 的多元 t 分布中生成样本。

首先，我们研究 TWE 对自由度参数ν 的估计效果，并将其与 OPP 和峰度估计量进行比较。图1展示了 $p = 100$ 时以及当 $\nu = 5$ 或 $\nu = 3$时不同样本长度下的箱线图。可以看出，所提出的 TWE 达到了最高的准确度，同时具有最小的变异性。此外，对于 $\nu = 5$，其中位数值恰到好处。峰度估计量显然表现

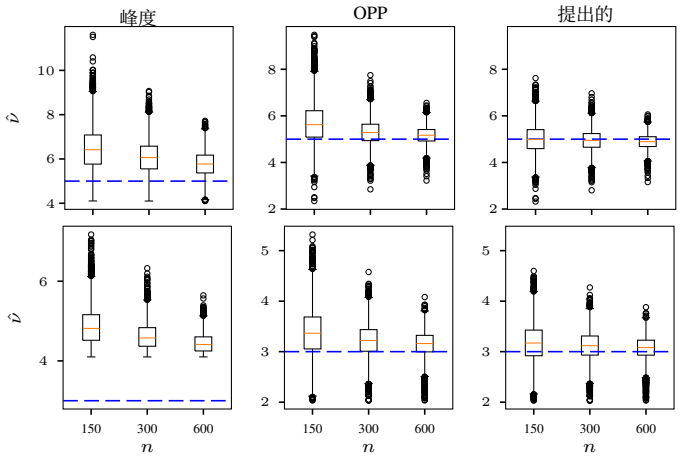

图 1：自由度估计值$\hat{\nu}$ 关于 n 的箱线图，其中$\nu = 5$ （上）和 $\nu = 3$ （下）； $p = 100$、 $\rho = 0.6$。

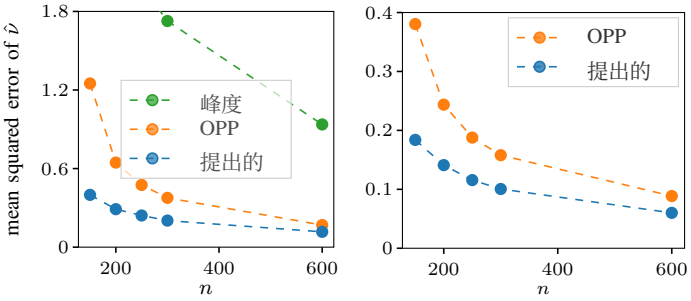

图 2：自由度估计值$\hat{\nu}$ 的均方误差随 n 的变化，其中 $\nu = 5$ （左图）和 $\nu = 3$ （右图）； $p = 100$， $\rho = 0.6$。

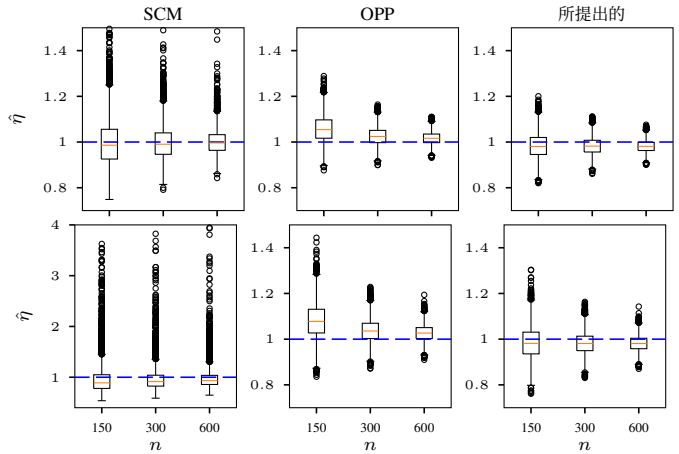

图3：当 $\nu = 5$ （上）和 $\nu = 3$ （下）时，尺度估计量$\hat{}\eta$ 随 n 变化的估计误差； $p = 100$， $\rho = 0.6$。

不佳，因为在 $\nu = 3$ 时四阶矩不存在。图2展示了平均均方误差（MSE），$(\hat{}\nu - \nu)^2$,，进一步说明了所提出的TWE 相较于其他竞争方法的优势和高准确度。

图3展示了不同尺度估计量η的箱线图。这里我们将 $\hat{}\eta_{\text{TWE}}$ 与 OPP 尺度估计量进行比较，其定义为 $\hat{}\eta_{\text{OPP}} = \text{tr}(\hat{}\mathbf{\Sigma})/p$，其中 $\hat{}\mathbf{\Sigma}$ 是基于$\nu = \hat{}\nu_{\text{OPP}}$的散布最大似然估计。我们还与 SCM 提供的尺度估计量进行比较，其定义为 $\hat{}\eta_{\text{cov}} = \text{tr}(\mathbf{S})/p$，但乘以

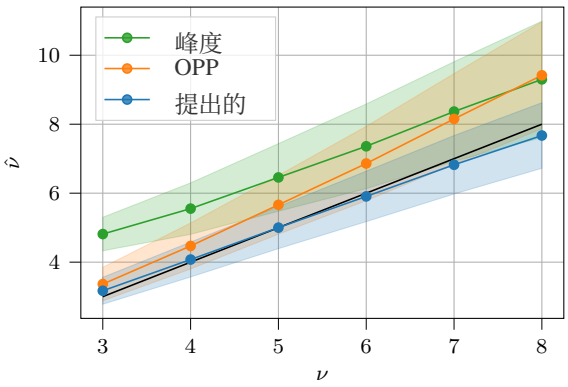

图4： ν、 $p = 100$、 $\rho = 0.6$、 $n = 150$ 的估计值。黑线表示真实值。

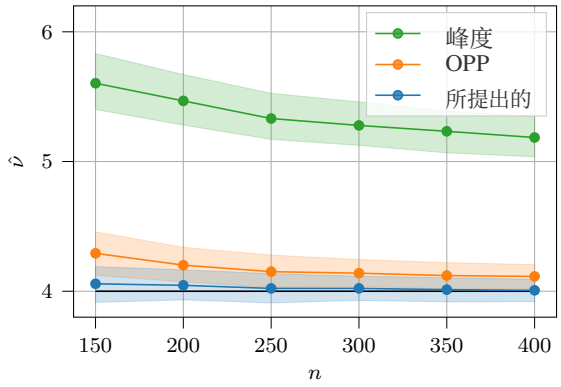

图 5：合成股票市场数据的 ν 估计值；$p = 100$、 $\nu = 4$。

$\theta^{-1} = (\nu - 2)/\nu$ 以获得 η的估计值；回顾（14）。从图3中可以看出，TWE略微低估了真实尺度，而 OPP 则高估了真实尺度。我们还注意到，对于 $\nu = 5$，SCM估计量明显无偏，但其变异性很大。图4展示了当 $n = 150$ 时（且 $p = 100$ 和$\varrho = 0.6$ 与前文相同），在不同 ν 取值范围内$\hat{}\nu$ 的中位数。对于所有自由度 $\nu \in [3, 8]$，所提出的TWE估计量都显著优于其他估计量。

现在我们考虑一个基于股票市场数据的示例。我们生成具有重尾的合成数据（$p = 100$ 资产），其服从自由度为 $\nu = 4$ 的多元 t 分布，协方差矩阵则根据标准普尔500指数成分股测得。图5比较了以下方法估计的 ν 值随观测数量的变化：峰度估计量、OPP 估计量 [18], 以及所提出的估计量。后者明显优于其他方法，展现出其在真实世界金融数据上的良好性能。

V. 总结性意见

我们提出了一种基于 Tyler M 估计量权重的椭圆分布尺度参数新型稳健估计量，并进一步利用该估计量构建了仿射等变 Tyler M 估计量。随后，我们提出了一种用于估计椭圆分布尾参数的统一框架。最后，值得注意的是，该方法可以直接推广至复值数据。

参考文献

[1] K.-T. Fang、S. Kotz 和 K.-W. Ng，《对称多元及相关分布》。伦敦：Chapman and Hall，1990。[2] E. Ollila、D. E. Tyler、V. Koivunen 和 H. V. Poor，“复椭圆对称分布：综述、新结果与应用”，IEEE 信号处理汇刊，第 60 卷，第 11 期，第 5597–5625 页，2012。[3] D. Paindaveine，“形状的规范定义”，《统计与概率快报》，第 78 卷，第 14 期，第 2240–2247 页，2008。[4] D. E. Tyler，“多元散布的无分布 M 估计量”，《统计年鉴》，第 15 卷，第 1 期，第 234–251 页，1987。[5] Y. Sun、P. Babu 和 D. P. Palomar，“正则化 Tyler 散布估计量：存在性、唯一性与算法”，IEEE 信号处理汇刊，第 62 卷，第 19 期，第 5143–5156 页，2014。[6] F. Pascal、Y. Chitour 和 Y. Quek，“广义稳健收缩估计量及其在空时自适应处理检测问题中的应用”，IEEE 信号处理汇刊，第 62 卷，第 21 期，第 5640–5651 页，2014。[7] E. Ollila 和 D. E. Tyler，“散布矩阵的正则化 M 估计量”，IEEE 信号处理汇刊，第 62 卷，6070，2014 22，6059 第。– 页。[8] Y. Chen、A. Wiesel 和 A. O. Hero，“高维协方差矩阵的稳健收缩估计”，IEEE 信号处理汇刊，第 59 卷，第 9 期，第 4097–4107 页，2011。[9] R. Couillet 和 M. McKay，“稳健收缩协方差矩阵估计量的大维度分析与优化”，《多元分析杂志》，第 131 卷，第 99–120 页，2014。[10] I. Soloveychik 和 A. Wiesel，“Tyler 协方差估计量的性能分析”，IEEE 信号处理汇刊，第 63 卷，第 2 期，第 418–426 页，2015。[11] T. Zhang、X. Cheng 和 A. Singer，“Tyler M 估计量的 Marčenko–Pastur 定律”，《多元分析杂志》，第 149 卷，123，2016 114页。–。[12] J. Goes、G. Lerman 和 B. Nadler，“通过阈值化 Tyler M 估计量进行稳健稀疏协方差估计”，《统计年鉴》，第 48 卷，第 1 期，第 86–110 页，2020。[13] E. Romanov、G. Kur 和 B. Nadler，“Tyler 和 Maronna 的 M 估计量：非渐近集中结果”，《多元分析杂志》，第 196 卷，文章编号 105184，2023。[14] D. P. Palomar、R. Zhou、X. Wang、F. Pascal 和 E. Ollila，fitHeavyTail：重尾条件下的均值与协方差矩阵估计，2023，R 软件包版本 0.2.0。 [在线]。网址：https://CRAN.R-project.org/package=fitHeavyTail [15] R. A. Maronna，“多元位置与散布的稳健 M 估计量”，《统计年鉴》，第 5 卷，第 1 期，第 51–67 页，1976。[16] P. L. Davies，“多元位置参数和离散矩阵的 S 估计量的渐近行为”，《统计年鉴》，第 1269–1292 页，1987。[17] K. S. Tatsuoka 和 D. E. Tyler，“非椭圆分布下 S 泛函和 M 泛函的唯一性”，《统计年鉴》，第 1219–1243 页，2000。[18] E. Ollila、D. P. Palomar 和 F. Pascal，“收缩协方差矩阵 M 估计量的特征值”，IEEE 信号处理汇刊，第 69 卷，第 256–269 页，2021。[19] F. Pascal、E. Ollila 和 D. P. Palomar，“改进多元 t 分布自由度参数的估计”，载于 2021 年第 29 届欧洲信号处理会议（EUSIPCO）。IEEE，2021，第 860–864 页。[20] C. Liu 和 D. B. Rubin，“使用 EM 及其扩展 ECM 和 ECME 对 t 分布进行极大似然估计”，《中华统计学志》，第 5 卷，第 1 期，第 19–39 页，1995。[21] C. Fernandez 和 M. F. J. Steel，“多元 Student-t 回归模型：陷阱与推断”，《生物统计学》，第 86 卷，167，1999 1，153第。– 页。