arXiv:2603.01237v1 [stat.ME] 1 Mar 2026

带异常检测规则的圆形数据稳健离散度量

Houyem Demni* Mia Hubert† Giovanni C. Porzio‡ Peter J. Rousseeuw§

2026年2月28日

摘要

圆形变量表示方向或周期观测，广泛存在于生物学、环境科学等领域。处理圆形数据时，一个重要问题是如何稳健地估计其离散度，避免异常点造成过度影响。本文将三种稳健离散度量从直线推广到圆上。我们通过影响函数和相对偏差曲线研究其稳健性。基于这些离散度量，可以推导出圆形分布参数的稳健估计量，从而得到冯·米塞斯分布集中参数和包裹正态分布离散参数的稳健估计量。本文求得了这些估计量的崩溃值和统计效率，并通过模拟研究对其进行比较。在表现最佳的估计量基础上，本文提出了一种稳健的圆形异常检测方法，并利用圆形小提琴图直观展示异常值。本文还分析了三个真实数据集。

关键词： 崩溃值、Circular 中位绝对偏差、Circular 最小中位数离散度、圆形最小截尾标准差、方向数据、离群值检测。

1 引言

在若干应用领域中，人们会遇到圆形数据，它由角度、方向和周期观测组成，并表示为圆上的点。圆形数据在生物科学中有广泛而有趣的应用，涉及多个领域，例如分布模型（Mardia 等人，2007年）、参数与非参数估计（Tishkovskaya 和 Blackwell，2021年）、回归（Hassanzadeh，2021年）和聚类（Ranalli 和 Maruotti，2020年）。另请参阅Ley 和 Verdebout （2017年，2019年），了解各种现代方法和数据示例。由于圆形数据位于有界非线性流形上，因此人们为其开发了专门的统计方法

*经济与法律系，卡西诺与南拉齐奥大学，意大利，houyem.demni@unicas.it
†统计与数据科学部，鲁汶大学，鲁汶，比利时，mia.hubert@kuleuven.be
‡经济与法律系，卡西诺与南拉齐奥大学，意大利，porzio@unicas.it§统计与数据科学部，鲁汶大学，鲁汶，比利时，peter@rousseeuw.net

分析，例如参见 Buttarazzi 等人（2018年）、Fern´andez-Dur´an （2007年）以及 Di Marzio 等人（2022年）。

位置和离散度的标准度量包括圆周均值和圆周标准差 CSD。其精确定义见第2节。然而，这些度量对离群值非常敏感。为说明这一点，我们考察了22只撒丁岛海星离开自然栖息地后移动的方向（Fisher，1995）。数据见图1，其中有两个异常观测值，我们用红色标出。完整数据集的 CSD 为0.58；剔除这两个特殊观测值后，CSD 大幅降至0.36。

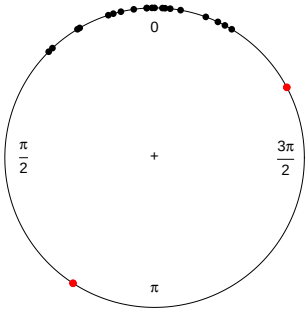


图1：撒丁岛海星移动方向的圆形图。两个异常观测值以红色标出。

为应对这种对异常值的敏感性，我们可以改用稳健度量。圆形数据的位置稳健度量之一是圆中位数。已有研究表明，它在单峰对称分布族中具有确立完善的稳健性性质（Ko和Guttorp，1988年）。无论使用完整的海星数据集，还是剔除两个可疑观测值，所得圆中位数都相同。

然而，与线性情形不同，迄今为止，圆离散度的稳健度量受到的关注较少。直线上的一种著名稳健离散度估计量是中位数绝对偏差（MAD），由Hampel等人（1986）和Ley等人（2013年）加以阐述。名为CMAD的圆形版本已被用于比较若干圆位置估计量的性能（Otieno和Anderson-Cook，2006）、降低凌星行星检测中的虚警率（Seader等人，2013年），以及在非参数估计脉冲星光变曲线的脉冲间歇区间时确定平滑参数（Schutte和Swanepoel，2016）。

尽管如此，尚未有人运用影响函数和崩溃值等稳健统计工具研究 CMAD。本文将对此展开研究。我们还将研究线性数据中另外两种稳健尺度度量——最小中位数平方（LMS）和最小截尾平方（LTS）方法——的圆形对应方法。这两种方法最初由 Rousseeuw (1984) 在回归背景下提出。我们将定义由此得到的 CLMS 和 CLTS 方法，并将其与

CMAD。

我们还将使用这些离散度量，为标准圆形分布的离散参数推导稳健估计量。我们将详细讨论冯·米塞斯分布的集中参数 κ 的估计，以及包裹正态分布的离散参数 σ 的估计。

最后，我们基于研究结果构造一种圆上的离群值检测规则，并介绍一种配套的图形工具，它是广为人知的小提琴图的一种变体。

本文的组织结构如下。第2节回顾圆形数据的一些背景概念。第3节定义 CMAD、CLMS 和 CLTS 离散度量。第4节推导它们的影响函数，并在第5节计算它们在污染下的相对偏差。第6节构造圆形分布离散参数的相应估计量，给出其崩溃值，并研究其效率。第7节介绍稳健异常检测规则和圆形小提琴图。第8节分析三个真实数据示例，第9节总结全文。

2 背景材料

圆形数据由单位圆上的点组成。选定原点和方向后，同一个点既可以用极坐标中的角度 $\theta \in [-\pi, \pi)$ 表示，单位为弧度；也可以用笛卡尔坐标表示为向量 $x = (x_1, x_2)^T = (\cos\theta, \sin\theta)^T$ 因此 $x \in \mathcal{S} := \{x : ||x|| = 1\}$。通过 $\theta = arctan2(x_2, x_1)$ 可以进行反变换。

注意，角度 θ 和 $\theta + 2\pi$ 在圆上表示同一个点。因此，圆形数据的运算是模 2π 运算。为简便起见，我们将用 $\theta_1 + \theta_2$ 代替 $\theta_1 + \theta_2 \bmod 2\pi$，其中 θ_1 和 θ_2 是圆上的两个点。

基于角度的记法比笛卡尔坐标记法更简洁，因此我们将主要使用角度。圆周随机变量可记为 Θ，其分布函数可记为 $F(\theta) = P(\Theta \leqslant \theta)$，其中 $\theta \in [-\pi, \pi)$。

在本文中，我们将反复使用单位圆上连接两点的最短弧的长度，并将其记为最短弧距离 $d(\theta_1, \theta_2) := \pi - |\pi - |\theta_1 - \theta_2||$。

2.1 Circular 随机变量的位置与离散度

圆周随机变量的位置通常用其圆周均值来衡量。设 Θ 为一个分布为 F 的圆周随机变量。其圆周均值定义为角度 $\mu(F) = \arctan2(\beta_1, \alpha_1)$，其中

$$\alpha_1 = \mathbb{E}(\cos\Theta) = \int_{-\pi}^{\pi} \cos\theta \, dF(\theta) \qquad \text{and} \qquad \beta_1 = \mathbb{E}(\sin\Theta) = \int_{-\pi}^{\pi} \sin\theta \, dF(\theta).$$

Θ 的圆周中位数集 则定义为使平均角距离最小的点集。

定义 1（圆周中位数集）。 随机变量 Θ 的圆周中位数集 M_Θ 定义为:

$$M_\Theta := \{\theta : \arg\min_\theta \mathbb{E}(d(\Theta, \theta))\}, \quad \theta \in [-\pi, \pi) \tag{1}$$

其中 $\mathbb{E}(T)$ 是（线性）随机变量 T 的期望值。

任一元素，在 M_Θ 中称为 圆中位数 或 Fr´echet 中位数。圆周中位数集可以是单点集、整个圆（例如在均匀分布情形下），甚至是不连通集。具有唯一中位数的圆形随机变量确实存在。例如，若某分布具有连续密度且为单峰，则其圆中位数是唯一的（Mardia，1972）。若 Θ具有唯一的圆中位数，我们将其记为 $\tilde{\mu}$。

定义 2（Ci rcular 方差）。 随机变量 Θ 的圆方差由下式给出

$$\mathrm{CV}(F) = 2(1-\rho) = \int_{-\pi}^{\pi} 2(1-\cos(\theta-\mu))dF(\theta) \tag{2}$$

其中 μ 是 Θ 的圆周均值，且

$$\rho = \left| \int_{-\pi}^{\pi} e^{i\theta}\, dF(\theta) \right| = \int_{-\pi}^{\pi} \cos(\theta-\mu)\, dF(\theta) \tag{3}$$

是平均合向量长度。

圆方差和平均合向量长度均有界，其中 $0 \leqslant \mathrm{CV}(F) \leqslant 2$ 和$0 \leqslant \rho \leqslant 1$。有些作者在（2）中省略因子 2，但我们更倾向于采用这个版本，因为随着集中参数增大，当分布集中在圆上一个很小的近似线性区域时，（2）对应于线性方差。这是因为，当 θ 趋近于 μ 时，$1-\cos(\theta-\mu)$ 的泰勒展开中，$2(1-\cos(\theta-\mu)$ 的首项为 $\theta-\mu)^2$, ；另见 Jammalamadaka和Sengupta（2001年）第 21 页。本文主要使用（2）的平方根，即圆标准差，记为

$$\mathrm{CSD}(F) = \sqrt{\mathrm{CV}(F)}\,.$$

如果圆上的测度在所有旋转下均保持不变，则称其为旋转不变。若旋转圆时测度也以相同方式旋转，则称该测度旋转等变。若圆形分布 F 关于方向 μ 反射对称，则称其为反射对称分布；其概率函数满足 $P(\Theta \in (\mu-\theta,\ \mu)) = P(\Theta \in (\mu,\ \mu+\theta))$，对所有 $\theta \in [-\pi, \pi)$ 均成立。

如果圆形分布 F 具有密度 f，且密度每隔 $2\pi/m$ 弧度重复一次，即对所有 θ，$f(\theta) = f(\theta + 2\pi/m)$，则称其为 m 阶旋转对称分布，其中 $m \in \mathbb{N}$。如果对所有 $\theta \in [-\pi, \pi$，$f(\theta) = f(\theta+\pi)$，则称其为对跖对称分布；也就是说，密度在旋转 π 弧度后保持不变。此外，如果该分布关于 μ 单调反射对称，即它是反射对称的，且对于 $\theta \in [0, \pi]$，$f(\mu+\theta)$ 非递增，那么密度会在 μ 处达到最大值，并向两个方向对称递减。

2.2 von Mises 分布与包裹正态分布

冯·米塞斯分布 $vM(\mu,\kappa)$ 是最常用于建模圆形数据的分布，有时也称为圆形正态分布。其密度为

$$f_{vM}(\theta;\mu,\kappa) := \frac{1}{2\pi I_0(\kappa)}\exp(\kappa\cos(\theta-\mu)),$$

其中， I_0 为第一类 0 阶修正贝塞尔函数。这里， $\mu\in[-\pi,\pi)$为位置参数， $\kappa\geqslant 0$ 为集中参数。位置参数 μ 同时是圆周均值、众数和圆中位数。其密度形状与正态分布相似，并且随着集中参数 κ 增大，逐渐趋近于正态分布。当 $\kappa=0$时，该分布退化为圆上的连续均匀分布。当 $\kappa>0$时，该分布为单峰分布，并且关于 μ反射对称。集中参数κ 与圆方差之间的关系由平均合向量长度 $\rho=I_1(\kappa)/I_0(\kappa)$表示，其中$I_1$ 为第一类 1 阶修正贝塞尔函数。因此， κ 由 $I_1(\kappa)/I_0(\kappa) = 1-\mathrm{CSD}^2/2$的解隐式定义；该方程没有解析表达式，必须通过数值方法求解。

另一种广泛用于单峰圆形数据的反射对称分布是包裹正态分布 $WN(\mu,\ \sigma^2)$，其密度为

$$f_{WN}(\theta;\mu,\sigma^2) := \frac{1}{\sigma\sqrt{2\pi}}\sum_{m=-\infty}^{\infty}\exp\left(\frac{-(\theta-\mu+2\pi m)^2}{2\sigma^2}\right).$$

其位置参数为 $\mu\in[-\pi,\pi)$，而 $0<\sigma^2<+\infty$ 是包裹到圆上的线性正态分布的方差。包裹正态分布的平均合向量长度为 $\rho=\exp(-\sigma^2/2)$，并且随着 σ^2 增大而减小。CSD 与 σ^2 直接相关，$\mathrm{CSD}^2=2(1-\exp(-\sigma^2/2))$。

3 三种圆形离散度度量

本节定义圆中位数绝对离差（CMAD）、圆形最小中位数离差（CLMS）和圆形最小截尾标准差（CLTS）。

3.1 圆中位数绝对离差（CMAD）

对于具有唯一圆中位数的分布 F，CMAD 定义为从该（圆）中位数出发的最短弧距离分布的（线性）中位数。

定义 3 (CMAD)。 设 $\mathcal{M}_1$ 为圆形随机变量集合，其唯一的圆中位数为 $\tilde{\mu}$。 $\Theta\in\mathcal{M}_1$ 的 CMAD 定义如下:

$$\mathrm{CMAD}(F) := \mathrm{Median}(d(\Theta,\tilde{\mu})) \tag{4}$$

其中，Median(T)表示（线性）随机变量 T的中位数。也就是说，Median(T)是满足以下条件的值集合 x： $P(T \leqslant x) \geqslant 0.5$且 $P(T \geqslant x) \geqslant 0.5$。如果该集合是一个区间，则取其中点作为中位值。

对于具有密度的圆形随机变量，CMAD(F)是以 $\tilde{\mu}$ 为中心的对称最短弧的长度的一半，该弧包含概率0.5：

$$\begin{aligned} \mathrm{CMAD}(F) &= \frac{1}{2}\, d(\theta_1, \theta_2), \\ \{\theta_1, \theta_2\} &= \underset{\theta_1,\theta_2 \in [-\pi,\pi)}{\arg\min} \{d(\theta_1, \theta_2) \,|\, P(\theta_1 \leqslant \Theta \leqslant \theta_2) = 0.5,\, d(\theta_1, \tilde{\mu}) = d(\theta_2, \tilde{\mu})\} \end{aligned} \tag{5}$$

圆中位数绝对离差（CMAD）非负。由于圆中位数具有旋转等变性，且最短弧距离具有旋转不变性，因此 CMAD 具有旋转不变性。它也具有反射不变性，因为圆中位数在反射下具有等变性，而最短弧距离具有不变性。

CMAD 是一种有界的圆离散度度量，满足 $0 \leqslant$ CMAD(F) $< \pi$。现在，我们将考察 CMAD 可取值的范围，以及该范围如何取决于分布。下一个引理的证明，以及其他所有理论结果的证明，均放在补充材料中。

引理1 (CMAD 的取值）**.** 设 $\Theta \in \mathcal{M}_1$ 为一个圆周随机变量，且 $\theta \in [-\pi, \pi)$.

$$\begin{aligned} &\textit{If } \exists \theta : P(\Theta = \theta) > 0.5, \quad \textit{then } \mathrm{CMAD}(F) = 0\,; \\ &\textit{If } \exists \theta : P(\Theta = \theta) = 0.5, \quad \textit{then } 0 \leqslant \mathrm{CMAD}(F) < \pi/2\,; \\ &\textit{If } \forall \theta : P(\Theta = \theta) < 0.5, \quad \textit{then } 0 < \mathrm{CMAD}(F) < \pi\,. \end{aligned}$$

根据引理1， π 是 CMAD 的一个上界。以下结果表明，该上界是紧的。

定理1. 对于遍历所有圆周随机变量的累积分布函数集合的 F，上确界$_F$CMAD(F) $=\pi$ 。

对于反射对称分布，CMAD 的值是该分布本身的一个分位数。以下结果可直接由公式（5）得出。

定理 2 （反射对称性下的 CMAD 值）。 设 $\Theta \in \mathcal{M}_1$ 关于某个 μ^*具有反射对称性。那么， Θ 的圆中位数是 μ^* 或 $\mu^* + \pi$。如果 μ^*，CMAD(F)等于连接 μ^* 与从 μ^* 自身沿顺时针方向或逆时针方向累积 *25%* 概率的点 ξ 的最短弧的长度。也就是说，

$$\xi : \int_{\mu^*}^{\xi} dF(\theta) = 0.25, \tag{6}$$

我们有：

$$\mathrm{CMAD}(F) = d(\xi, \mu^*). \tag{7}$$

If t Θ 的圆中位数为 $\mu^* + \pi$，方程(6) 和(7) 适用于 $\mu^* + \pi$.

当圆中位数不唯一时，CMAD 未定义。因此，我们定义了另外两种不依赖于圆中位数的离散度量。

3.2 圆形最小中位散布（CLMS）

圆形最小中位散布（CLMS）将直线上的最小中位数平方离散度量（Rousseeuw，1984；Rousseeuw 和 Leroy，1987）推广到了圆上。CLMS 是该随机变量到圆上一点的中位距离的最小值。

定义 4 （CLMS）。设 Θ 为分布为 F的圆周随机变量，且 θ 遍历圆上的所有点。则 CLMS 定义为：

$$\mathrm{CLMS}(F) := \min_{\theta} \mathrm{Median}(d(\Theta, \theta)).$$

等价地，CLMS 是包含至少一半概率分布的最短弧长度的一半：

$$\mathrm{CLMS}(F) = \frac{1}{2} \min\{d(\theta_1, \theta_2) | P(\theta_1 \leqslant \Theta \leqslant \theta_2) \geqslant 0.5\}, \tag{8}$$

其中 θ_1 和 θ_2 是圆上的两个点。CLMS是非负的，并且在旋转和反射下保持不变。我们现在研究CLMS达到其可达到的最小值和最大值的条件。

引理2 （CLMS的条件 $= 0$）。 设 Θ 为服从分布 F的圆周随机变量， Δ_θ 表示 θ处的圆形点质量分布。我们有：

$$\mathrm{CLMS}(F) = 0 \quad \textit{iff} \quad F = (1-\varepsilon)\Delta_\theta + \varepsilon G, \quad 0 \leqslant \varepsilon \leqslant 0.5,$$

对任意值 θ 以及任意分布G。

引理3 （CLMS的最大值）。 设 $\Theta \in [-\pi, \pi)$ 为服从分布F的圆周随机变量。则，

$$\max_{F} \ \mathrm{CLMS}(F) = \pi/2.$$

该结果成立，因为 CLMS(F) 不可能大于 $\pi/2$，因为 π 是圆上两点之间最短路径的最大长度。另一方面，CLMS $= \pi/2$ 刻画了一个分布类，如下述定理所示。

定理 3. 设 Θ 是一个分布为 F的圆周随机变量。那么

$$\mathrm{CLMS}(F) = \pi/2 \quad \Leftrightarrow \quad F \in H_{0.5}$$

其中， $H_{0.5}$ 是满足 $\int_{HC} dH_{0.5}(\theta) = 0.5$ 的分布类，条件对任意半圆 HC均成立。

分布类 $H_{0.5}$ 是中心对称（或对跖对称）分布类的子集H，因为它们满足 $\int_{HC} dH(\theta) \geqslant 0.5$ 对任意半圆 HC均成立，参见例如 Nagy 等人 (2024年）中的讨论。

请注意，CSD 与定理 3 中的 $\Leftarrow$ 具有类似的含义。由于$H_{0.5}$ 中的分布具有对跖对称性，其合向量长度为 0，因此其 CSD 取最大值$\sqrt{2}$。但是， $\Rightarrow$ 并不成立。例如，阶数为$m = 3$ 的旋转对称分布 F 满足 $CSD(F) = \sqrt{2}$，但它不属于 $H_{0.5}$。

第二点说明是将 CLMS 与 CMAD 作对比。根据定义，在$H_{0.5}$ 下，后者没有定义，因为这些分布不存在唯一的圆中位数。

另一方面，在单调旋转对称性下，CLMS 与 CMAD 取值相同，如下所述。

定理 4。 设 Θ 是一个连续的圆周随机变量，其分布 F 具有密度 f ，且关于某个 μ^* 反射对称并严格单峰，众数为 μ^*。设 ξ 是从 μ 本身开始累积 *25%* 概率的点，如 （6） 所示。那么

$$\mathrm{CLMS}(F) = d(\xi, \mu^*).$$

该定理成立，是因为最短弧关于众数方向 μ^* 对称。此外，根据连续性以及公式（8）中的概率约束，其长度的一半所对应的概率质量为 0.25。请注意，单调反射对称性是反射对称性的一种特殊情况，在此条件下，定理 2 对 CMAD 得到了相同的取值。

3.3 圆形最小截尾标准差（CLTS）

我们讨论的第三种稳健离散度量是圆形最小截尾标准差 CLTS，它将最小截尾平方（LTS）离散度（Rousseeuw，1984）从直线推广到圆。CLTS 是指至少包含 50% 概率质量的弧上 CSD 的最小值。

对于任意的 θ_1 和 θ_2，以及任意分布为 F的圆周随机变量 Θ ， 记 $F_{[\theta_1,\theta_2]}$为给定 $\theta_1 \leqslant \Theta \leqslant \theta_2$ 时 Θ 的条件分布，并记 $\mu_{[\theta_1,\theta_2]}$ 为其圆周均值。

定义 5 (CLTS)。 设 Θ 是分布为 F 的圆周随机变量。则圆形最小截尾标准差（CLTS）定义为

$$\mathrm{CLTS}(F) := \min_{\theta_1,\theta_2} \{\mathrm{CSD}(F_{[\theta_1,\theta_2]}) \mid P(\theta_1 \leqslant \Theta \leqslant \theta_2) \geqslant 0.5\}. \tag{9}$$

由于 CSD 具有非负性、旋转不变性和反射不变性，因此 CLTS 也具有这些性质。由式 (2) 可得

$$\mathrm{CLTS}(F) := \min_{\theta_1,\theta_2} \left\{ \sqrt{\frac{\int_{\theta_1}^{\theta_2} 2(1-\cos(\theta-\mu_{[\theta_1,\theta_2]}))dF(\theta)}{P(\theta_1 \leqslant \Theta \leqslant \theta_2)}} \mid P(\theta_1 \leqslant \Theta \leqslant \theta_2) \geqslant 0.5 \right\}. \tag{10}$$

引理4 (CLTS 的条件 $= 0$)。 设 Θ 为分布为 F的圆周随机变量， Δ_θ 为某个 θ中的点质量。那么

$$\mathrm{CLTS}(F) = 0 \quad \Leftrightarrow \quad F = (1-\varepsilon)\Delta_\theta + \varepsilon G, \quad 0 \leqslant \varepsilon \leqslant 0.5,$$

对于任意值 θ 以及任意分布 G。

对于 CLTS 的最大值，我们可以依据以下结果。

引理5。 设 F 为连续圆周均匀分布。也就是说，F 的密度为 $1/(2\pi)$。那么

$$\mathrm{CLTS}(F) = \sqrt{2(1-2/\pi)}.$$

该结果可由式（10）推得，因为每个半圆都恰好包含 0.5 的概率质量。我们猜想，不存在其他（离散或连续的）圆形分布，其 CLTS 能达到这一数值。

最后，在单调反射对称的情况下，可以推导出 CLTS 的一般方程。

定理5。 设 Θ 是一个连续的圆周随机变量，其分布 F 具有密度 f ，该密度关于某个 μ^* 反射对称且严格单峰，众数为 μ^*。设 ξ 是从 μ^* 起累积 *25%* 概率的点，如 (6) 所示。那么

$$\mathrm{CLTS}(F) = \sqrt{4\int_{\mu^*}^{\mu^*+\xi} 2(1-\cos(\theta-\mu^*))dF(\theta)}\,. \tag{11}$$

根据 (10)，且由于具有最小 CSD 的条件分布关于 μ^* 反射对称，因此此结论成立。

3.4 von Mises 分布和包裹正态分布的值、下确界与上确界

推论1. 令 $F = vM(\mu, \kappa > 0)$ 或 $F = WN(\mu, \sigma^2)$，并记 $q_1 = F^{-1}(1/4)$ 和 $q_3 = F^{-1}(3/4)$。则

$$\mathrm{CMAD}(F) = \mathrm{CLMS}(F) = q_3 - \mu \tag{12}$$

$$\mathrm{CLTS}(F) = \mathrm{CSD}(F_{[q_1,q_3]}) \tag{13}$$

这由定理2、定理4和定理5可得。下面给出冯·米塞斯分布（ $\kappa > 0$ and）以及包裹正态分布下 CMAD、CLMS、CLTS 的下确界和上确界。

推论2. 设 $\Theta \sim vM(\mu, \kappa > 0)$ 或 $\Theta \sim WN(\mu, \sigma^2)$。则

$$\begin{aligned} \inf\ \mathrm{CMAD}(F) = \inf\ \mathrm{CLMS}(F) = \inf\ \mathrm{CLTS}(F) = 0 \\ \sup\ \mathrm{CMAD}(F) = \sup\ \mathrm{CLMS}(F) = \pi/2 \\ \sup\ \mathrm{CLTS}(F) = \sqrt{(2(1-2/\pi)}\,. \end{aligned}$$

在冯·米塞斯分布处，下确界成立是因为引理1、引理2和引理4，并且因为 $F \xrightarrow{d} \Delta_\mu$ 对于 $\kappa \to +\infty$。上确界由(12)和引理5得出。成立的是$\lim_{\kappa\to 0}\mathrm{CMAD}(F) = \lim_{\kappa\to 0}\mathrm{CLMS}(F) = \frac{\pi}{2}$以及$\lim_{\kappa\to 0}\mathrm{CLTS}(F) = \sqrt{(2(1-2/\pi)}$。

在包裹正态分布处，我们有$\lim_{\sigma\to 0}\mathrm{CMAD}(F) = \lim_{\sigma\to 0}\mathrm{CLMS}(F) = \lim_{\sigma\to 0}\mathrm{CLTS}(F) = 0$，因为对于 $\sigma \to 0$ 该分布越来越集中于 μ^*附近。我们还有$\lim_{\sigma\to+\infty}\mathrm{CMAD}(F) = \lim_{\sigma\to+\infty}\mathrm{CLMS}(F) = \frac{\pi}{2}$，以及 $\lim_{\sigma\to+\infty}\mathrm{CLTS}(F) = \sqrt{(2(1-2/\pi)}$，因为对于 $\sigma \to +\infty$ 该分布趋近于均匀分布。

4 影响函数

泛函的稳健性可以通过不同方式衡量。首先，我们考察其影响函数，它用于评估少量污染所产生的影响。形式上，泛函 Hampel等人，1986年 在模型分布 S 下，点 θ 处的影响函数 F 定义为

$$\mathrm{IF}(\theta; S, F) := \frac{\partial}{\partial\varepsilon} S\big((1-\varepsilon)F + \varepsilon\Delta_\theta\big)\Big|_{\varepsilon=0+} = \lim_{\varepsilon\downarrow 0} \frac{S\big((1-\varepsilon)F + \varepsilon\Delta_\theta\big) - S(F)}{\varepsilon}\,. \tag{14}$$

CSD 的影响函数由 Ko和Guttorp （1988年）给出，形式如下:

$$\mathrm{IF}(\theta; \mathrm{CSD}, F) = \frac{1}{\mathrm{CSD}(F)}\Big(1 - \frac{\mathrm{CSD}^2(F)}{2} - \cos(\theta)\Big)$$

对于 $-\pi \leqslant \theta \leqslant \pi$，以及众数为 0 的反射对称圆分布 F。现在我们将计算 CMAD、CLMS 和 CLTS 泛函的影响函数。为了处理诸如 $(1-\varepsilon)F + \varepsilon\Delta_\theta$ 这样的受污染分布，我们将考虑以下类型的分布

$$P = a_0 P_0 + \sum_{j=1}^{k} a_j \Delta_{\theta_j} \tag{15}$$

其中， $a_0 > 0$, P_0 是圆上的一个连续密度为 f 的分布， $k < \infty$，所有 $a_j > 0$，所有 θ_j均位于区间 $]-\pi,\ \pi]$内，而 Δ_θ 是将全部概率质量集中在点 θ上的概率分布。由于 P是一个概率，因此必有 $a_0 + a_1 + \ldots + a_k = 1$。接下来我们定义累积

分布函数 G，定义为 $G(\theta) = \int_{-\pi}^{\theta} g(u)du$，其中 $-\pi \leqslant \theta \leqslant \pi$。因此，我们有 $G(-\pi) = 0$ 以及$G(\pi) = 1$。接下来，我们将 G 扩展到区间 $[-2\pi, 2\pi]$ 上，定义为 $G(\theta) = G(\theta + 2\pi) - 1$，其中 $-2\pi \leqslant \theta < -\pi$；以及 $G(\theta) = G(\theta - 2\pi) + 1$，其中 $\pi < \theta \leqslant 2\pi$。 G 是右连续的，因此我们可以按通常方式定义 G^{-1}。

对于任意分布 函数（形式如式 (15)），现在可以将泛函 CLMS(G) 写作

$$\mathrm{CLMS}(G) = \frac{1}{2}\inf\left\{G^{-1}\left(t + \frac{1}{2}\right) - G^{-1}\left(t\right)\ ;\ G(0) - 1 < t \leqslant G(0) + \frac{1}{2}\right\}. \tag{16}$$

定理 6 (CMAD 和 CLMS 的影响函数)。 设 F 为一个具有连续密度$f(\theta) > 0$ 的圆形分布，该分布关于 *0* 反射对称且严格单峰，众数为 *0*（即 f 在 $[0, \pi]$上严格递减）。那么，CMAD 和 CLMS 的影响函数分别为

$$\mathrm{IF}(\theta; S, F) = \frac{sign(|\theta| - F^{-1}(3/4))}{4f(F^{-1}(3/4))} \tag{17}$$

对于 $-\pi \leqslant \theta \leqslant \pi$，其中 $\mathrm{CMAD}(F) = \mathrm{CLMS}(F) = F^{-1}(3/4)$（根据定理 *2* 和定理 *4*）。

证明见附录 B。请注意，大多数经典圆形统计学模型都满足关于 F 的条件，包括任意冯·米塞斯分布（其中 $\mu = 0$ 和 $\kappa > 0$）以及任意包裹正态分布（其中 $\mu = 0$）。有趣的是，在这类分布 F 下，CMAD 和 CLMS 具有相同的影响函数，尽管它们是相当不同的估计量。

图 2 展示了当 $\mu = 0$ 和 $\kappa = 1/2$ 时，CMAD 和 CLMS 在冯·米塞斯分布下的影响函数。在区间 $[-\pi, \pi]$ 之外，它以 2π 为周期延续。

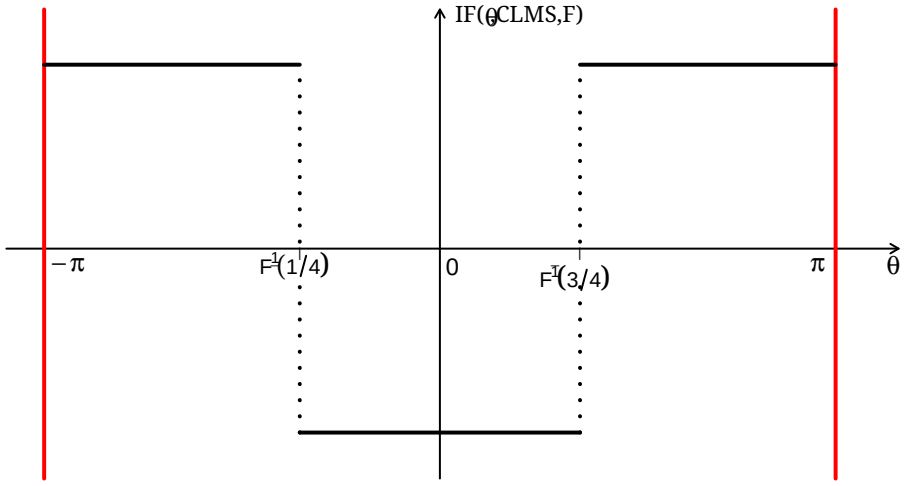


图 2：当 $\mu = 0$ 和 $\kappa = 1/2$ 时，CMAD 和 CLMS 在冯·米塞斯分布下的影响函数。

我们看到，影响函数是一个分段常数函数：对于由未受污染分布 F 确定尺度的区间之外的点，其取值为 $1/(4f(F^{-1}(3/4)))$，

而对于位于该区间内的点，则取相反的值。这意味着单个点的标准化影响相当小。当我们在区间外添加一个点时，CMAD 和 CLMS 只会略微增大；在区间内添加一个点时，它们只会略微减小。影响的大小由所谓的粗差敏感度 来衡量，其定义为

$$\gamma^*(S,F) = \sup_\theta |\mathrm{IF}(\theta; S, F)|$$

因此，对于 γ^*(CMAD F) $= 1/(4f(F^{-1}(3/4)))$ $= \gamma^*$(CLMS F)。CLTS 的影响函数看起来有所不同。

定理 7 (CLTS 的影响函数)。 设 F 为一个具有连续密度 $f(\theta) > 0$ 的圆形分布，该分布关于 *0* 反射对称且以 *0* 为众数严格单峰，也就是说， f 在 $[0,\pi]$上严格递减。那么 *CLTS* 估计量 S 在 $[-\pi,\pi]$ 上的影响函数为

$$\mathrm{IF}(\theta; S, F) = \frac{1}{S(F)}\Big[1 - \frac{S^2(F)}{2} - \cos(F^{-1}(3/4)) + 2I\big(|\theta| < F^{-1}(3/4)\big)\big(\cos(F^{-1}(3/4)) - \cos(\theta)\big)\Big]. \tag{18}$$

图 3 展示了冯·米塞斯分布下 CLTS 的影响函数，其中 $\mu = 0$ 和 $\kappa = 1/2$。请注意，影响函数的中心部分呈现 $-\cos(\theta)$ 的形状，而线性情形下最小截尾平方尺度估计量的影响函数在中心部分呈二次形式（Croux 和 Rousseeuw，1992年）。然而，当冯·米塞斯分布变得更加集中时（$\kappa \to \infty$），CLTS 的影响函数将变得无法与高斯分布下线性最小截尾平方的影响函数区分开来，因为 $-\cos(\theta)$ 在零点附近的泰勒展开为 $-1 + \theta^2/2 + o(\theta^3)$ 。此外，CMAD 和 CLMS 的影响函数也将变得无法与线性情形下尺度估计量 MAD 和 LMS 的影响函数区分开来。

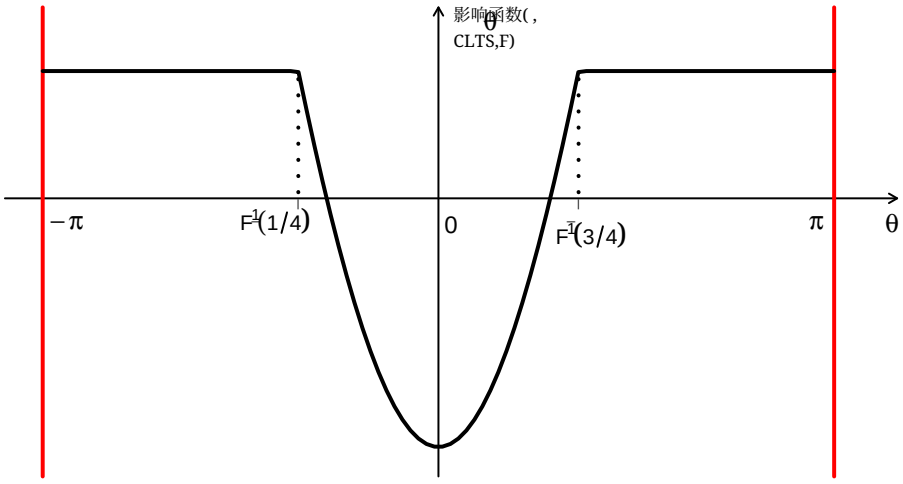


图 3：在冯·米塞斯分布下 CLTS 的影响函数，其中 $\mu = 0$ 和 $\kappa = 1/2$。

请注意，定理 6 和定理 7 给出的是 CMAD、CLMS 和 CLTS 的“原始”版本的影响函数。在下文中，我们将考虑变换 ζ，使它们成为参数模型中某个参数的相合估计量，因此，实际估计量将具有 $\zeta(S)$ 的形式，其中 S 是原始估计量。在这种情况下，我们可以通过链式法则计算最终估计量的影响函数：

$$\mathrm{IF}(\theta;\zeta(S),F)=\frac{\partial}{\partial\varepsilon}\zeta\Big(S((1-\varepsilon)F+\varepsilon\Delta_\theta)\Big)\Big|_{\varepsilon=0+}=\zeta'\big(S(F)\big)\mathrm{IF}(\theta;S,F)\,, \tag{19}$$

因此，它与原始估计量的影响函数成正比。

5 偏差曲线

影响函数衡量的是少量点污染对泛函的影响，而我们也可以研究其他类型和程度的污染所产生的影响。

此处我们考虑冯·米塞斯分布 F和污染分布$G_{\theta,\ \varepsilon}=(1-\varepsilon)F+\varepsilon\Delta_\theta$ 对于 $\theta\in[-\pi,\pi)$ 和 $0\leqslant\varepsilon<0.5$。我们采用点污染，因为它通常对尺度泛函的影响最大（Martin 和 Zamar，1989年；Croux 和 Haesbroeck，2001年）。我们将相对偏差在 θ 中定义为

$$rb(\theta;S,F,\varepsilon):=\frac{S(G_{\theta,\varepsilon})-S(F)}{S(F)} \tag{20}$$

其中 S 是 CSD、CMAD、CLMS 或 CLTS 泛函。公式（20）中的分母使不同泛函的取值更具可比性。绘制整个 θ取值范围内的相对偏差，即可得到相对偏差曲线。

为近似计算这些偏差曲线，我们在大小为 5000 的样本上进行计算。所有离散度量的样本版本均使用R中实现的函数计算。样本 CMAD 通过考虑公式（4）的经验对应量得到，并使用R 软件包 `circular` (Lund 和 Agostinelli，2013年)。根据（8），样本 CLMS 等于覆盖 $h=\lceil n/2\rceil+1$ 个样本点的最短弧长度的一半，其中 $\lceil\cdot\rceil$ 为向上取整函数。为得到样本 CLTS，我们计算大小为 $h=\lceil n/2\rceil+1$的所有连续子样本的 CSD，并取其中最小值。

图 4 展示了位置为 $\mu=0$ 、集中参数为 $\kappa\in\{2,5\}$ 的冯·米塞斯分布下的相对偏差曲线 F。在每种情形中，污染水平分别设为 $\varepsilon\in\{5\%,\ 10\%,\ 20\%\}$。

我们发现，离散度量的偏差曲线形状与其影响函数相似。当污染位于分布的中心区域时，偏差为负；在该区域之外，偏差为正。随着污染程度增加，影响也随之增大。污染对圆标准差 CSD 的影响远大于对 CMAD、CLMS 和 CLTS 的影响。在这三个稳健估计量中，CMAD 在分布外围区域的正偏差最大，而 CLTS 在中心区域的负偏差最大，但差异很小。

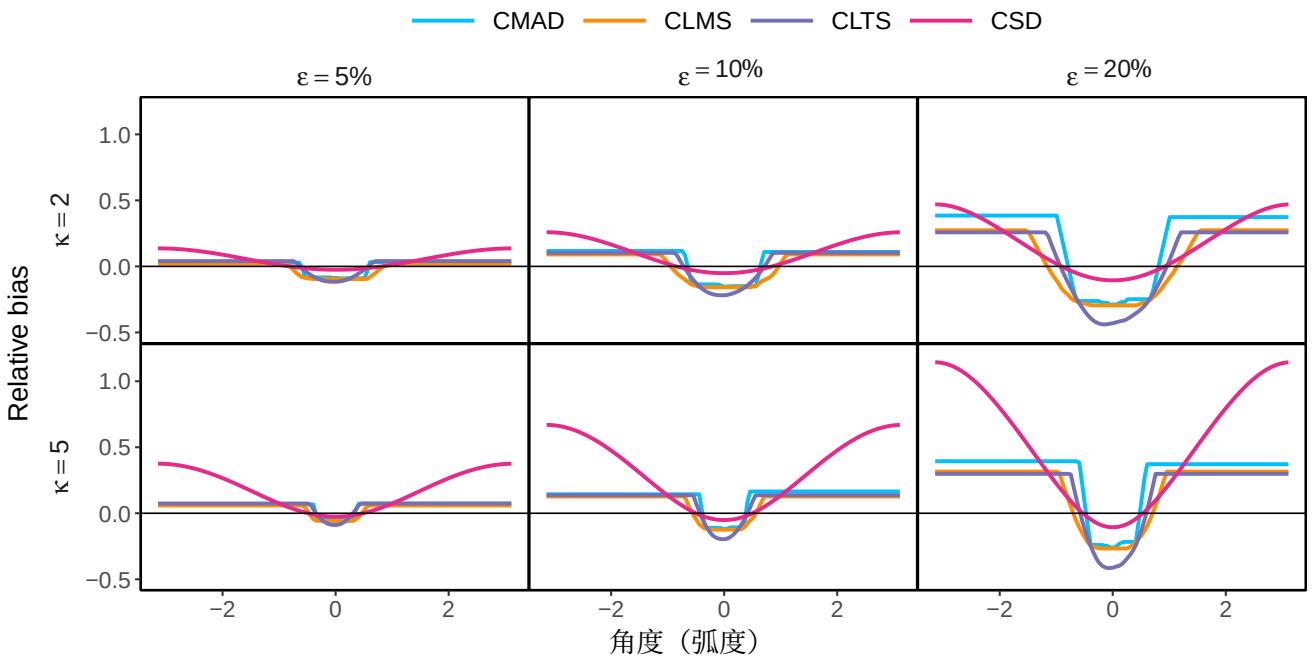


图 4：在满足 $\mu = 0$ 的冯·米塞斯分布下，CMAD、CLMS、CLTS 和 CSD 的相对偏差曲线。各行对应集中参数 $\kappa = 2$，5。各列对应污染率$\varepsilon = 5\%$，10%，20%。

κ 的取值对这三个稳健度量相对偏差的大小影响不大。另一方面，CSD 在较高的 κ 值下表现出大得多的相对偏差。原因如下。当 κ较高时，未污染数据紧密聚集在 μ周围，离散度量 CSD 的精度很高，变异性较低。因此，即使少量污染也会导致显著的相对变化。换言之，当 κ 较高时，CSD 的表现类似于直线上的经典标准差。

这些结果与Ko和Guttorp（1988年）的结果一致。他们考虑了影响函数的标准化版本，即将影响函数除以 $S(F)$。在（20）中，我们以相同方式对偏差进行标准化。他们在定理 4 中证明，经典离散度估计量的标准化影响函数会随着 κ 增大而无界，因此得出结论：该估计量不稳健。

从相对偏差的角度来看，我们的结果表明，所提出的三种稳健度量的表现都远优于 CSD。总体而言，CLMS 的偏差最小。

我们在均值偏移离群值模型下进行了类似分析，其中未污染分布和污染分布的 κ 值相同，仅位置不同。结果（未显示）与图 4 中的结果在定性上相似。

6 圆形分布参数的稳健估计

所提出的离散度量可用于获得圆形分布参数的稳健估计量。我们将以集中参数为 κ的冯·米塞斯分布，以及离散参数为 σ的包裹正态分布为例进行说明。

我们首先考察图5中 CMAD、CLMS 和 CLTS 随总体圆形标准差变化的数值结果。虚线表示它们的下确界和上确界。

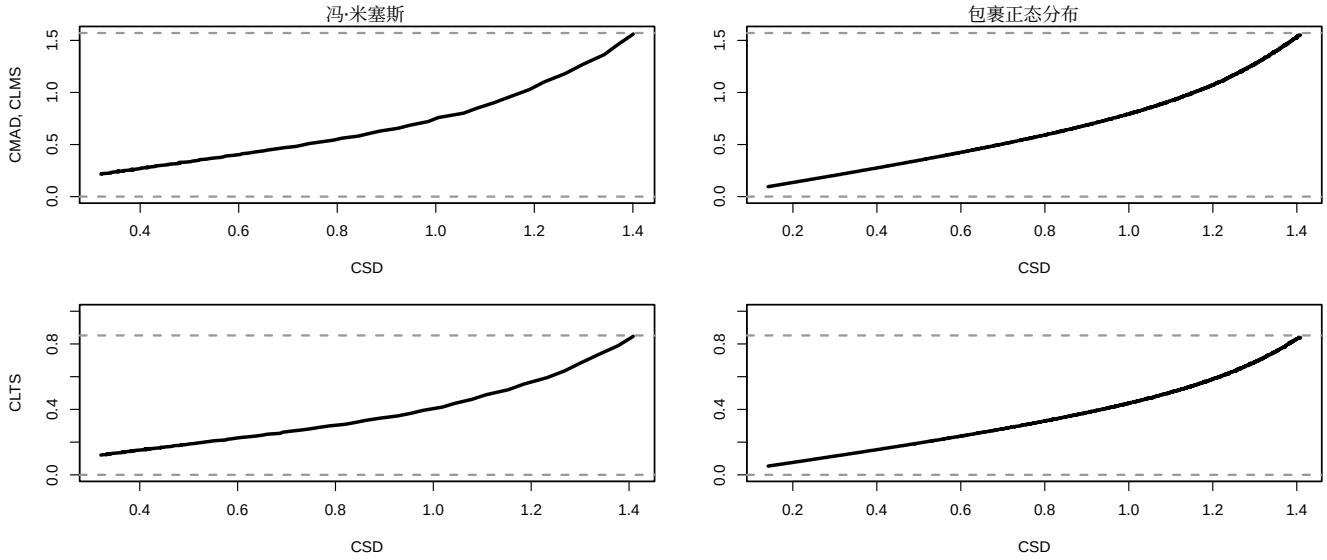


图5：在冯·米塞斯分布（左，$0 < \kappa \leqslant 10$）和包裹正态分布（$0 < \sigma \leqslant 3$）下，CMAD、CLMS 和 CLTS 随总体圆形标准差变化的取值。虚线表示这些圆形度量的下确界和上确界。

此外，总体圆形标准差本身是一个非线性函数，取决于冯·米塞斯分布的集中参数 κ以及包裹正态分布的 σ 。如图6所示，图中给出了 CSD 关于 κ 在 $0 < \kappa \leqslant 10$时，以及关于 σ 在 $0 < \sigma \leqslant 3.4$时的函数关系。

6.1 κ 和 σ的稳健一致估计量

我们现在利用这些关系来构造模型参数的稳健相合估计量：首先获得总体圆形标准差的稳健估计，然后通过适当的变换将其映射到相应参数 κ或 σ 。

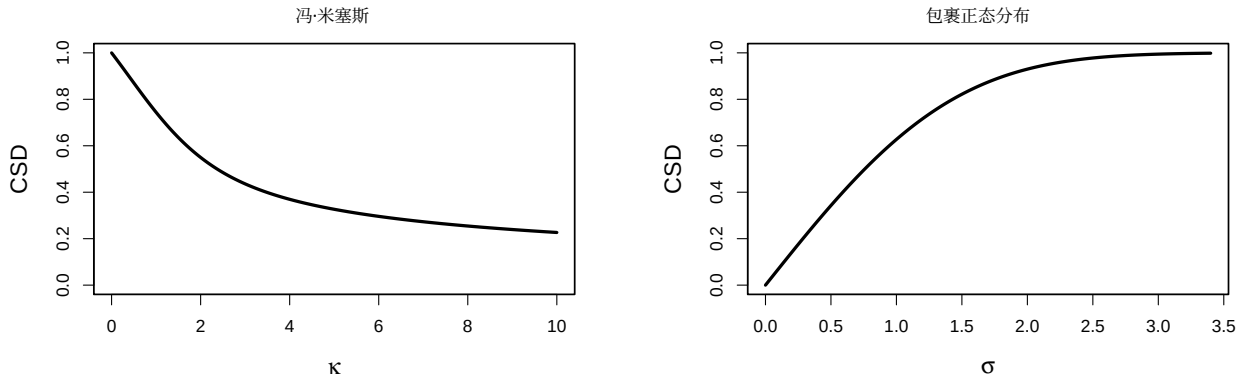


图6：CSD 随参数 κ 在冯·米塞斯分布和 σ 在包裹正态分布下的取值变化。

设 S_n 表示由样本 $\{\theta_1, ..., \theta_n\}$计算得到的稳健圆周离散度量。对于每个圆周模型，相应的总体离散度量 S 可以表示为底层CSD的函数，即

$$S = \eta(\mathrm{CSD}).$$

通过反演这一关系，我们得到总体圆形标准差的稳健估计量：

$$\widehat{\mathrm{CSD}}_n = \eta^{-1}(S_n).$$

由此总体圆形标准差的稳健估计量，我们进一步推导出相应参数的相合估计量。对于冯·米塞斯分布，我们通过数值反演函数 $A(\kappa) = I_1(\kappa)/I_0(\kappa)$，得到 κ 的估计量，从而得到一致性函数：

$$\widehat{\kappa} = A^{-1}(\widehat{\rho}) = A^{-1}\Big(1 - \frac{\widehat{\mathrm{CSD}}_n^{\,2}}{2}\Big).$$

对于包裹正态分布，我们同样构造尺度参数 σ 的一致性函数：

$$\widehat{\sigma} = \sqrt{-2\log(\widehat{\rho})} = \sqrt{-2\log\left(1 - \frac{\widehat{\mathrm{CSD}}_n^{\,2}}{2}\right)}\,.$$

在线性情形中，相应的一致性函数是常数，因此它成为相合因子。

6.2 CMAD、CLMS 和 CLTS 的崩溃值

泛函的一种常见稳健性度量是其崩溃值，定义为该泛函能够应对的离群值的最高比例。对于离散度/集中度估计量，分别定义内爆崩溃值和爆炸崩溃值。前者是使估计量降至零所需的最小污染比例

，后者则是使估计量达到其上确界所需的最小污染比例。

若始终有 $S(F_\psi) = \psi$，则泛函 S 是分布 F_ψ 的参数 ψ 的费希尔一致估计。

定义 6（泛函的崩溃值）。泛函 S 在圆形分布 F 处的内爆和爆炸崩溃值分别为

$$\varepsilon^*_{\text{imp}}(S, F) := \inf\{\varepsilon > 0 : \inf_H S((1-\varepsilon)F + \varepsilon H) = 0\}$$
$$\varepsilon^*_{\text{exp}}(S, F) := \inf\{\varepsilon > 0 : \sup_H S((1-\varepsilon)F + \varepsilon H) = \sup S\}$$

where H 遍历所有圆形分布。总体崩溃值为

$$\varepsilon^*(S, F) := \min\{\varepsilon^*_{\text{imp}}(S, F), \varepsilon^*_{\text{exp}}(S, F)\}.$$

设 ψ 表示一个圆周离散参数，ψ_{S} 是一个估计量，它由稳健离散度量 S 的严格单调一致性函数定义，其中 S 为 CMAD、CLMS 或 CLTS。各个 ψ_{S} 的崩溃值由相应稳健度量 S 的崩溃值决定，其上下界见引理 2 至 5。以下给出这些崩溃值。证明见补充材料 C 节。

引理 6（ψ_{CMAD} 的内爆崩溃值上界）。设 F 是一个单峰且对称的圆形分布，其离散参数为 ψ。我们暂时将定义 *6* 中的 H 限定为满足以下条件的分布：$G := (1-\varepsilon)F + \varepsilon H$ 具有唯一的圆中位数。则

$$\varepsilon^*_{\text{imp}}(\widehat{\psi}_{\text{CMAD}}, F) < 1 - 0.5/[P_F(\mu - \pi/2 < \Theta < \mu + \pi/2)]\,. \tag{21}$$

因此，对于冯·米塞斯分布和包裹正态分布，ψ_{CMAD} 的内爆崩溃值小于 *0.5*。

定理 8 (ψ_{CMAD}的总体崩溃值**)**。设 F 是一个单峰且对称的圆形分布，其离散度或集中参数为 $\psi > 0$。我们暂时将定义 *6* 中的 H 限制为满足以下条件的分布：$G := (1-\varepsilon)F + \varepsilon H$ 具有唯一的圆中位数。则

$$\varepsilon^*_{\text{exp}}(\widehat{\psi}_{\text{CMAD}}, F) = 0.5\,.$$

因此，ψ_{CMAD} 的总体崩溃值满足以下上界

$$\varepsilon^*(\widehat{\psi}_{\text{CMAD}}, F) < 1 - 0.5/[P_F(\mu - \pi/2 < \Theta < \mu + \pi/2)] < 0.5\,. \tag{22}$$

根据定理 8，ψ_{CMAD} 的崩溃值达不到可实现的最大值 0.5。这与线性数据的中位数绝对偏差形成对比，后者的崩溃值确实达到 0.5。

另一方面，尽管 CMAD 的崩溃值达不到 0.5，但其上界仍可能相当高。图7展示了在冯·米塞斯模型下，CMAD 的内爆崩溃值上界随集中参数 κ变化的情况。当 κ 低至 2 时，该上界仍可达到 0.459。当半圆 $P_F(\mu-\pi/2<\Theta<\mu+\pi/2)$的概率趋近于 1 时，上界也趋近于 0.5。请注意，$\widehat{\kappa}_{\mathrm{CMAD}}$是 CMAD 的单调递减函数，因此 CMAD 内爆时它会爆炸，但其崩溃值仍与定理 8 中相同。

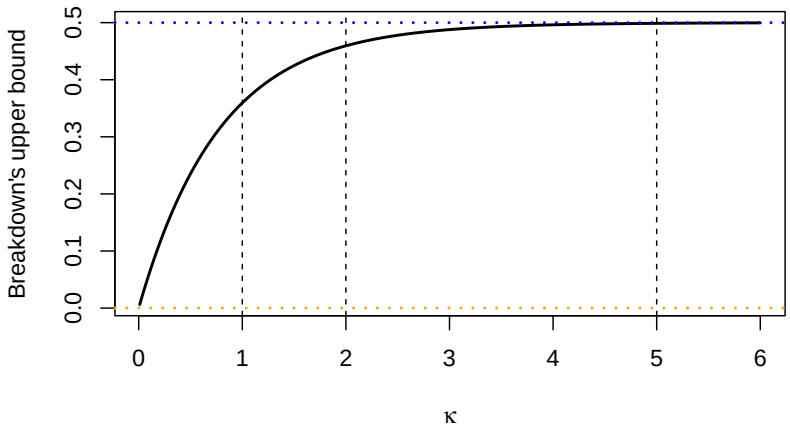


图7：根据定理 8，在冯·米塞斯分布下，$\widehat{\kappa}_{\mathrm{CMAD}}$ 的崩溃值上界是集中参数 κ 取值的函数。

基于 CLMS 的 ψ 估计量确实达到了可能的最高崩溃值 .

定理 9 ($\widehat{\psi}_{\mathrm{CLMS}}$**的崩溃值）。**设 F 是一个具有离散程度或集中参数 $\psi>0$ 的单峰对称分布，估计量为 ψ_{CLMS}。则

$$\varepsilon^*(\widehat{\psi}_{\mathrm{CLMS}},F)=0.5.$$

这可直接由引理 2 和 3 得出。

定理 10 ($\widehat{\psi}_{\mathrm{CLTS}}$**的崩溃值）。**设 F 是一个具有离散程度或集中参数 $\psi>0$ 的单峰对称分布，估计量为 ψ_{CLTS}。则

$$\varepsilon^*(\widehat{\psi}_{\mathrm{CLTS}},F)=0.5.$$

这一结果与定理 9 类似。特别是，内爆崩溃值由引理4得出，而爆炸崩溃值由引理5得出。

$\widehat{\psi}_{\mathrm{CMAD}}$、$\widehat{\psi}_{\mathrm{CLMS}}$ 和 $\widehat{\psi}_{\mathrm{CLTS}}$ 的崩溃性质直接取决于稳健圆形离散泛函 CMAD、CLMS 和 CLTS 的内缩与爆炸崩溃值，因此与所假设的参数族无关。

6.3 $\widehat{\psi}_{\mathrm{CMAD}}$、$\widehat{\psi}_{\mathrm{CLMS}}$和 $\widehat{\psi}_{\mathrm{CLTS}}$的效率

通过计算所提出的稳健估计量相对于假设的模型下最大似然估计量（MLE）的渐近相对效率（ARE），评估其效率。对于感兴趣参数 ψ，估计量 ψ 的 ARE 定义为

$$\mathrm{ARE}(\widehat{\psi}) = \mathrm{AVar}(\widehat{\psi}_{\mathrm{MLE}})/\mathrm{AVar}(\widehat{\psi})$$

其中，AVar 表示渐近方差。这些结果在补充材料的 D 节中推导得出，也可以通过数值方法计算。

对于冯·米塞斯分布，图8左侧面板展示了估计量$\widehat{\kappa}_{\mathrm{CMAD}}$、$\widehat{\kappa}_{\mathrm{CLMS}}$和 $\widehat{\kappa}_{\mathrm{CLTS}}$ 在参数 κ 从 1 到 10 的网格上的渐近相对效率。$\widehat{\kappa}_{\mathrm{CMAD}}$ 和$\widehat{\kappa}_{\mathrm{CLMS}}$ 的渐近相对效率相同，以橙色曲线表示；$\widehat{\kappa}_{\mathrm{CLTS}}$ 的渐近相对效率则以紫色表示。

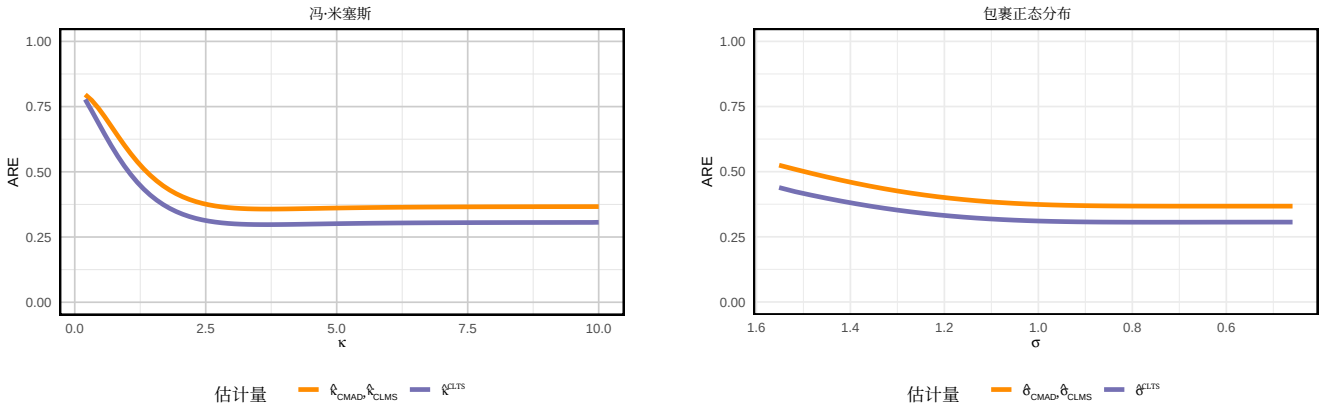


图8：基于 CMAD、CLMS 和 CLTS 的估计量的渐近相对效率曲线；该效率随冯·米塞斯分布的集中参数 κ以及包裹正态分布的离散参数 σ 变化。

我们看到，渐近相对效率随 κ 增大而单调下降。当 κ较低时，基于CMAD和CLMS的估计量相对于最大似然估计仍保持超过70%的效率。随着 κ增大，渐近相对效率下降至约37%。这一水平渐近线符合预期，因为当 $\kappa \to \infty$时，冯·米塞斯分布在局部上变得类似于直线上的正态分布，而线性中位数绝对偏差和最小中位数平方的渐近效率确实为36.74%。

基于CLTS的 κ 估计量的行为在较低的 κ时与之相似。当 κ 增大时，渐近相对效率略低于 $\widehat{\kappa}_{\mathrm{CMAD}}$ 和 $\widehat{\kappa}_{\mathrm{CLMS}}$的相应值。在 $\kappa \to \infty$ 的极限情形下，它变为30.67%，即直线上正态分布处最小截尾平方尺度估计量的效率，正如Croux 和 Rousseeuw （1992年）中所推导的那样。

图8右侧面板展示了估计量 $\widehat{\sigma}_{\mathrm{CMAD}}$、$\widehat{\sigma}_{\mathrm{CLMS}}$ 和 $\widehat{\sigma}_{\mathrm{CLTS}}$ 在包裹正态分布下的渐近相对效率。参数 σ 的取值范围为 0.45 至 1.55。请注意，为便于与左侧面板中冯·米塞斯分布的图进行比较，横轴上的取值是从右向左绘制的

，因为较低的尺度 σ 与较高的集中参数κ有类似的效果。橙色曲线表示 $\hat{\sigma}_{\text{CMAD}}$ 和 $\hat{\sigma}_{\text{CLMS}}$的渐近相对效率，紫色曲线表示 $\hat{\sigma}_{\text{CLTS}}$的渐近相对效率。同样，基于 CLTS 的估计量效率略低于基于 CLMS 的估计量；对于 $\hat{\sigma} \to 0$，我们得到与之前相同的极限： $\hat{\sigma}_{\text{CLMS}}$ 为 36.74%， $\hat{\sigma}_{\text{CLTS}}$为 30.67%。

综合两种分布模型，我们可以得出结论：基于 CLMS 的估计量的效率略优于基于 CLTS 的估计量。

6.4 $\hat{\psi}_{\text{CMAD}}$、$\hat{\psi}_{\text{CLMS}}$、$\hat{\psi}_{\text{CLTS}}$的稳健性：一项实证研究

到目前为止，我们研究了影响函数、相对偏差和崩溃值。这些量都基于泛函，也就是总体层面的量。当然，我们也关心有限样本下的表现。为此，我们进行了一项小规模模拟研究，考察了 10% 和 20% 污染的影响。

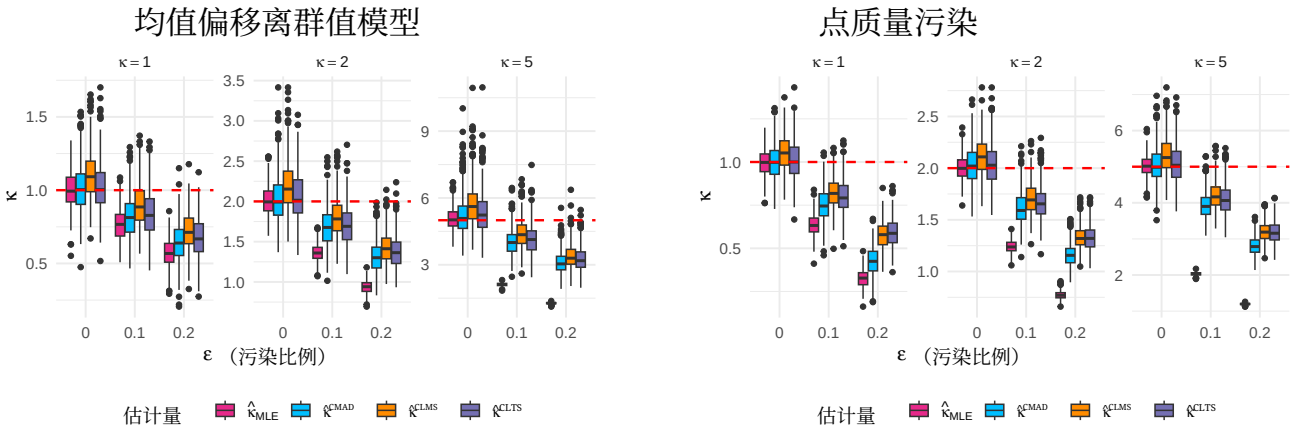


图 9：冯·米塞斯分布（ $\kappa \in \{1, 2, 5\}$）在含有比例为 ε的对跖污染时的模拟结果。图中展示了估计值 $\hat{\kappa}_{\text{MLE}}$、 $\hat{\kappa}_{\text{CMAD}}$、 $\hat{\kappa}_{\text{CLMS}}$和 $\hat{\kappa}_{\text{CLTS}}$ 的箱线图，以及表示真实 κ的红色虚线。

我们从冯·米塞斯分布开始，比较污染对最大似然估计量 $\hat{\kappa}_{\text{MLE}}$ 以及对 $\hat{\kappa}_{\text{CMAD}}$、 $\hat{\kappa}_{\text{CLMS}}$和 $\hat{\kappa}_{\text{CLTS}}$的影响。我们从冯·米塞斯分布 F 生成了200个点，其中 $\mu(F) = 0$，真实集中参数 κ 等于1、2和5。除了这一干净数据集之外，我们还将其20个点（10%）替换为来自冯·米塞斯分布 H 的点，该分布具有相同的 κ，但中心为对跖中心，即 $\mu(H) = \pi$。这种均值偏移离群值模型是对位置而言最严重的污染，参见例如Demni等人（2021）。同样的情形在20%污染下重复进行。实验重复了500次，结果展示在图9的左侧面板中。其中包含 $\hat{\kappa}$在 κ、 ε和估计量的每种组合下的箱线图。对于 $\varepsilon = 0$ （干净数据），四个估计量的中位数相近，并且

接近真实 κ处的红色虚线水平线，但 $\hat{\kappa}_{\mathrm{CLMS}}$ 除外，后者存在较小的向上偏差。每个箱体的高度反映估计量的准确度（效率）。正如预期的那样，对于干净数据，最大似然估计最为准确，但其他估计量的表现仍然相当不错。对于 $\varepsilon = 0.1$，情况发生了变化：最大似然估计受影响最大，其次是CMAD，然后是CLTS和CLMS。对于 $\varepsilon = 0.2$，这种影响被放大，最大似然估计的箱体远离且往往较小。

图 9 的右侧面板是使用相同干净数据进行模拟的结果，但现在 10% 和 20% 的污染数据全部放在一个单点上，即对跖点 π。这是点质量污染。我们看到，其效应在性质上与左侧面板相似。

对包裹正态分布重复了该模拟，此次离散参数 σ 取值为 0.5、1 和 1.2。污染同样属于均值偏移离群值类型以及点污染。图10 中的结果与冯·米塞斯分布的结果相当相似，同时考虑到污染下较高的估计离散度 $\hat{\sigma}$ 所起的作用与此前较低的集中参数 $\hat{\kappa}$相同。

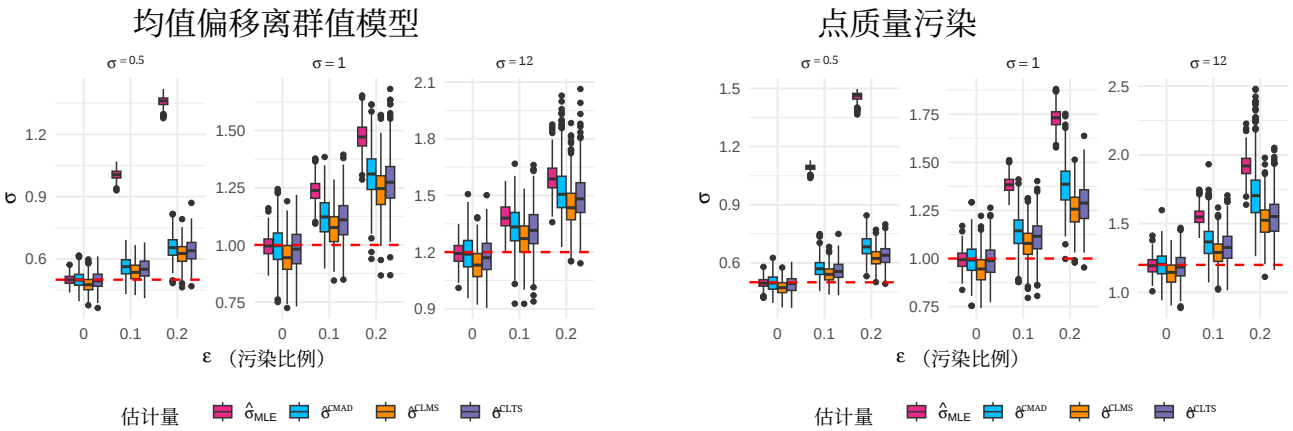


图10：对包裹正态分布进行模拟，其中 $\sigma \in \{0.5, 1, 1.2\}$，并含有比例为$\varepsilon$ 的对跖污染。图中展示了 $\hat{\sigma}_{\mathrm{MLE}}$、 $\hat{\sigma}_{\mathrm{CMAD}}$、 $\hat{\sigma}_{\mathrm{CLMS}}$和 $\hat{\sigma}_{\mathrm{CLTS}}$ 的箱线图，以及表示真实 σ的红色虚线。

总体而言，模拟结果表明，无论是在冯·米塞斯模型还是包裹正态模型中，最大似然估计量受到污染的影响都最大。三种稳健估计量受到的影响要小得多。其中，CLMS 似乎最稳健。结合相对偏差和渐近相对效率的结果来看，CLMS 似乎是三种稳健方法中表现最佳的。

7 稳健异常检测

本节将说明如何利用稳健估计在圆上进行离群值检测。

7.1 稳健圆形异常检测规则

人们日益认识到异常观测对数据分析的影响，以及圆形数据离群值检测方法的必要性。一些作者基于删除法提出了圆形离群值检测技术，参见例如 Collett (1980), Jammalamadaka和 Sengupta (2001年), 以及 Mardia(1975). 然而，这类方法往往会受到掩蔽效应的影响，即当另一个异常值存在时，某个异常点可能无法被检测出来。其中一些技术仅适用于样本量较小的情况，或者假定集中参数事先已知。

也有人提出了其他方法。然而，其中一些方法只能检测单个异常值（Abuzaid 等人，2012），另一些方法则在特定数据模型下通过模拟获得临界值（Mahmood 等人，2017），或者要求异常值与其余数据之间有明显的间隔（Mohamed 等人，2016）。

此外，Abuzaid Abuzaid (2020)、Greco 等人 (2021）以及 Sau 和 Rodriguez (2018）也提出了用于环面数据和球面数据的离群值检测方法。

基于稳健度量的一些替代方法也已得到发展。加权似然和最小差异方法均已适用于圆形情形（Agostinelli，2007）。事实证明，这些方法的性能很大程度上取决于某些参数的选择。在相关领域中，已有研究将一种稳健离群值检测工具用于圆形回归（Rana 等人，2016）。

稳健异常检测技术假设大多数数据遵循某种参数模型。随后对参数进行稳健估计，再根据数据与拟合模型之间的距离较大，将异常值标记出来（Rousseeuw 和 Hubert，2018年）。

Demni 和 Porzio (2023）通过假设数据服从冯·米塞斯分布，将这一方法应用于圆形数据。然而，所使用的稳健估计量 κ相当简单，并且在数据集中程度适中时会表现出较大的变异性。

按照这一思路，并利用上述结果，我们建议改用基于 CLMS 的估计量 $\widehat{\kappa}_{\mathrm{CLMS}}$ 来估计冯·米塞斯分布，并使用估计量 $\widehat{\sigma}_{\mathrm{CLMS}}$来估计包裹正态分布。我们还希望离群值检测规则具有旋转不变性。

该方法首先通过 Fr´echet 中位数 $\tilde{\mu}$稳健地估计位置，然后计算每个圆形观测 θ_i 到 $\tilde{\mu}$的最短弧距离 $d(\theta_i, \tilde{\mu})$。

随后将这些距离与截断值 $c_\alpha(\psi)$进行比较，其中 $\alpha > 0$ 是数据干净时期望被标记点的比例，而 ψ 是离散度或集中参数

底层分布的。我们使用基于 CLMS 的稳健估计量（例如 $\widehat{\kappa}_{\mathrm{CLMS}}$ 或 $\widehat{\sigma}_{\mathrm{CLMS}}$）来估计该参数。

当底层圆形分布关于其位置参数对称时，截断值 $c_\alpha(\psi)$ 定义为中心位于 $\mu = 0$ 的底层分布的 $1 - \alpha/2$ 分位数。因此，它对应于下列方程的数值解

$$\int_0^{c_\alpha(\psi)} f(\theta, \psi) d\theta = 1 - \alpha/2 \,, \tag{23}$$

其中，$f(\theta, \psi)$ 是具有 $\mu = 0$ 的分布的密度。接下来，每当满足以下条件时，我们就将圆形数据点 θ_i 标记为异常点

$$d(\theta_i, \tilde{\mu}) > c_\alpha(\widehat{\psi}_{\mathrm{CLMS}}) \,. \tag{24}$$

7.2 异常点可视化：圆形小提琴图

除了离群值检测规则外，我们还提供了一种配套的图形工具，称为圆形小提琴图。它旨在帮助进一步发掘数据分析洞见。

直线上的小提琴图由 Hintze 和 Nelson(1998年) 提出，旨在克服传统箱线图的一些局限。Tukey 提出箱线图时，计算能力有限，因此箱线图的简洁性是一项优势，但这也限制了它们可视化的信息量。小提琴图包含密度估计，能够揭示多峰性。有关线性小提琴图及其用途的概述，请参见 Aberasturi (2014年)。小提琴图已广泛应用于多个学科，包括生物统计学（Mekkes 等人，2024年）和心理学（Tanious 和 Manolov，2022年）。已有研究提出了小提琴图及相关图形工具的多种扩展形式（Montalcini 和 Rousseeuw，2025年）。

圆形数据的小提琴图必须遵循圆的几何结构。估计得到的密度被安置在单位圆的圆周上。随后，该密度曲线沿径向镜像到圆的内侧。两条密度曲线之间的区域形成一个平滑的环，其厚度反映局部数据集中程度。密度使用冯·米塞斯核进行非参数估计。密度在离群值检测规则的截断阈值处被截断，位于小提琴图之外的数据点被标记为离群值。

小提琴图内部融合了圆形箱线图（Buttarazzi 等人，2018年）。箱体表示四分位距，中位数两侧各包含 25% 的数据，从而突出显示中间 50% 的数据。中位数以一条短白线表示。中位数和箱体展现了位置和变异性的稳健汇总统计量。下一节将通过真实示例展示圆形小提琴图。

8 实际数据分析

本文考察了三个真实数据示例。所示结果基于使用冯·米塞斯分布拟合数据得出，但使用包裹正态分布也会得到基本相同的结果。

8.1 北方蟋蟀蛙数据

Ferguson 等人 (1967)收集了北方蟋蟀蛙归巢能力的数据。在黑暗的环境舱中经过30小时后，14只蛙被释放，并记录了它们的方向。结果以弧度报告，零弧度对应于东，角度按逆时针测量。数据绘制于图11的左侧面板。其圆形样本中位数 $\widehat{\mu} = 2.39$弧度由黑色箭头标示。集中参数的基于 CLMS 的稳健估计为 $\widehat{\kappa}_{\text{CLMS}} = 3.88$。由公式(23)得到截断值 $c_{0.01}(\widehat{\kappa}_{\text{CLMS}}) = 1.34$弧度。

右侧面板中的圆形小提琴图通过使用稳健估计 $\widehat{\kappa}_{\text{CLMS}}$来估计圆形密度而构建。密度经重新缩放后沿单位圆径向显示，形成对称的小提琴形状。一个观测被标记为异常值。这与文献一致：Collett (1980)通过删除法发现同一点（5.515弧度）为圆形异常值。

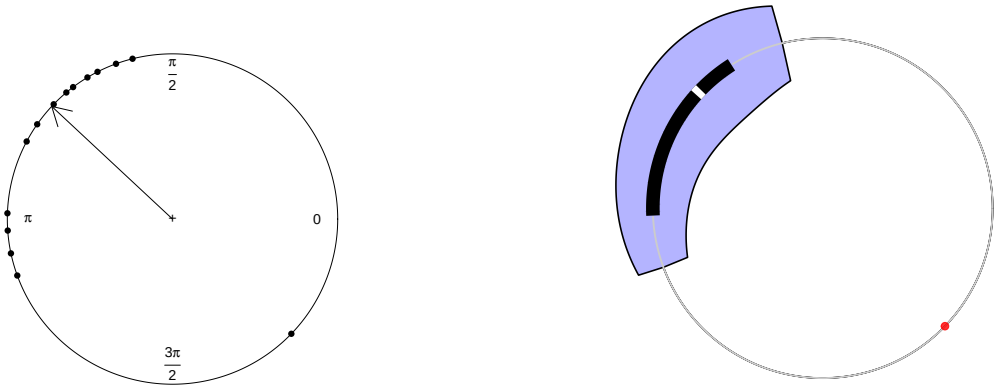


图 11：（左）北方蟋蟀蛙数据及其样本圆周中位方向（黑色箭头）的图，（右）其圆形小提琴图。

乍看之下，人们可能会觉得小提琴图位于圆内和圆外的部分并非完全镜像，但实际上它们是镜像对称的。这种观感源于圆形几何本身，包括其曲率以及圆形定义域的闭合性。在线性小提琴图中，密度以直轴为轴反射，因此轴两侧相同的距离对应相同的面积。而在圆上，镜像是沿径向方向进行的，这并不会对应相同的面积，也不会形成视觉上对称的形状。因此，人们可能会将镜像密度看作

一条厚度不一的连续带，而不是两个对称的半边。这种效果是圆形几何固有的，与具体的密度估计量或缩放选择无关。

8.2 撒丁岛海星数据

我们在图 1 中看到了撒丁岛海星数据。数据记录的是 22 只海星离开自然栖息地 11 天后的合成方向（Fisher，1995）。这些方向显示在图 12 的左侧面板中。标记为 0 弧度的方向（位于顶部）是海星朝海岸移动的“正确”方向。

在这个数据集中，有两个点远离数据中的大多数点：一只海星（位于逆时针方向 5.201 弧度处）沿着较长的路线朝海岸移动，而另一只（位于 2.565 弧度处）则朝远离海岸的方向移动。问题是，这两个点能否视为离群值。

如果采用 Collett （1980）中的单值删除技术，由于掩蔽效应，这两个点都不会被标记为异常值。

Kato 和 Eguchi （2016）也采用了单值删除技术。他们在逐一省略一个点的子样本上，将冯·米塞斯分布的位置和集中参数的稳健估计与相应的最大似然估计进行比较。他们得出的结论是，位于 2.565 处的点似乎是异常值，而位于 5.201 处的点是否异常则难以判断。这可以理解：如果数据集中保留了距离最远的点，那么省略第二远的点对最大似然估计不会产生很大影响。

序贯删除方法（即删除最偏离的观测值并重复该过程）在这些数据中也存在局限性。根据 Fisher (1995)中的分析，冯·米塞斯Q-Q图在视觉上显示，2.565 弧度处的点与其他点不一致。但删除该点后，剩余数据的冯·米塞斯Q-Q图看起来正常。针对是否偏离冯·米塞斯分布的不一致性检验未能拒绝无不一致性的原假设，因此，5.201 弧度处的点未被标记。

我们提出的离群值检测规则得出了不同的结论。根据基于CLMS的估计$\widehat{\kappa}_{\mathrm{CLMS}} = 7.92$，我们得到阈值 $c_{0.01}(\widehat{\kappa}_{\mathrm{CLMS}}) = 0.86$ 弧度。由此得到的圆形小提琴图见图 12 右侧面板。它清楚地将两个远端点都标记为离群值。这是因为$\widehat{\kappa}_{\mathrm{CLMS}}$ 仅使用 50% 的数据点来估计 κ，因此这两个远端点都未参与该计算。另一方面，如果在公式 (24) 中使用最大似然估计位置 $\widehat{\mu}_{\mathrm{MLE}}$ 和集中参数 $\widehat{\kappa}_{\mathrm{MLE}}$，则相应的截断值将为 $c_{0.01}(\widehat{\kappa}_{\mathrm{MLE}}) = 1.52$，此时只有距离最远的点会被标记。

8.3 果蝇幼虫数据

Pewsey 和 Jones (2012)记录了果蝇幼虫运动状态的变化，每秒测量一次

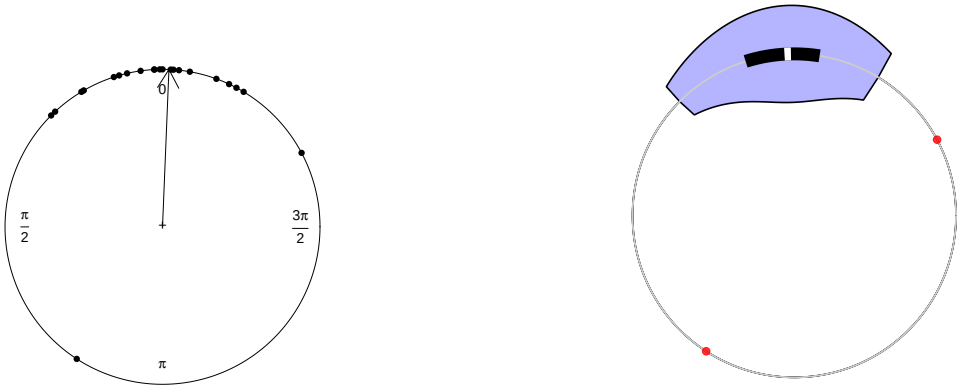


图12：左图为撒丁岛海星数据图，标出圆周中位方向（黑色箭头）；右图为圆形小提琴图。

持续三分钟。图13左侧面板显示了全部180个点，顶部标出了0弧度的参考方向。样本圆周中位数为 $\tilde{\mu} = 6.26$弧度（接近顶部），基于CLMS的集中参数估计为$\hat{\kappa}_{\text{CLMS}} = 15.58$。假设数据来自冯·米塞斯分布，相应的截断值为 $c_{0.01}(\hat{\kappa}_{\text{CLMS}}) = 0.60$。

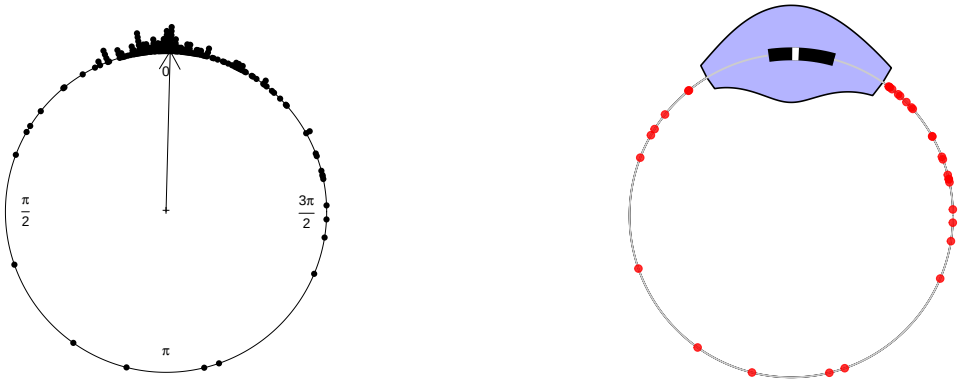


图13：左图为果蝇幼虫数据图，标出其圆中位数（黑色箭头）；右图为其圆形小提琴图。

在右侧面板的圆形小提琴图中，许多点被标记为异常值。此外，小提琴图与附近的“离群值”之间没有明显的间隙。这表明，这些数据并非由冯·米塞斯分布生成，而是来自尾部更重的分布。Pewsey 和 Jones (2012) 的研究结果确实表明，幼虫数据用逆巴切莱特分布建模效果更好。我们认为，圆形小提琴图有助于质疑观测数据所依据的分布模型。

9 结论

分析圆形数据时，能够估计其离散度十分重要。圆标准差等经典离散度量可能会受到异常值的不当影响。本文研究了三种稳健离散度估计量：CMAD、CLMS 和 CLTS，它们是线性数据中常用离散度量的扩展。其中只有 CMAD 以前被使用过，但尚未通过稳健统计工具进行分析。本文通过推导这三种方法的影响函数、相对偏差、崩溃值和渐近效率，对它们进行了分析，并针对干净数据和污染数据开展了模拟研究。综合比较所有结果后，我们得出结论：CLMS 表现最佳。

我们还构建了一种基于 CLMS 的离群值检测规则，其相较于早期方法不易受到掩蔽效应的影响。与之配套的是一种名为圆形小提琴图的新型图形展示方法。这些工具使我们能够分析三个潜在异常值数量各不相同的真实数据集。

尽管当前研究聚焦于圆形数据，但将其扩展到球面或环面等其他流形将具有重要意义。

软件 包含 R 代码以及用于复现实例的示例脚本的 zip 文件可在 https://wis.kuleuven.be/statdatascience/code/circular_r_code.zip 获取。

资助 H. Demni 的研究得到了佛兰德斯研究基金会（FWO）科学人员赴佛兰德斯访问资助项目 V500325N 的支持，同时也是项目 ECS 0000024“创新生态系统—罗马科技城”的一部分。该项目由欧盟下一代计划通过 MUR 法令第 1051 号（2022年6月23日）资助，隶属于 PNRR 第4项任务第2部分投资1.5—CUP H33C2200042000。G. Porzio 的工作由项目 ECS 0000024“创新生态系统—罗马科技城”资助。该项目由欧盟下一代计划通过 MUR 法令第 1051 号（2022年6月23日）资助，隶属于 PNRR 第4项任务第2部分投资1.5—CUP H33C2200042000。

参考文献

D. T. Aberasturi. 小提琴图。*Wiley StatsRef*：在线统计学参考资料，1-7，2014年。

A.H.Abuzaid. 在医学多元圆形数据中识别基于密度的局部离群值。《医学统计学》，39(21):2793–2798，2020年。

A. H. Abuzaid、A. G. Hussin、A. Rambli 和 I. Mohamed。《圆形数据不一致性新检验的统计方法》

。《统计学通讯：模拟与计算》，41(10):1882–1890，2012年。C. Agostinelli。《圆形数据的稳健估计》《计算统计与数据分析》，51(12):5867–5875，2007年。D. Buttarazzi、G. Pandolfo 和 G. C. Porzio。《圆形数据箱线图》《生物统计学》，74(4):1492–1501，2018年。D. Collett。《圆形数据中的离群值》《皇家统计学会杂志：*C*辑（应用统计）》，29(1):50–57，1980年。C. Croux 和 G. Haesbroeck。《基于子区间的稳健尺度估计量的最大偏差曲线》《计量学》，53:101–122，2001年。C. Croux 和 P. J. Rousseeuw。《一类基于子区间的高崩溃点尺度估计量》《统计学通讯：理论与方法》，21:1935–1951，1992年。H. Demni、A. Messaoud 和 G. C. Porzio。《基于距离的方向深度分类器：一项稳健性研究》《统计学通讯：模拟与计算》，52(11):5695–5713，2021年。H. Demni 和 G. C. Porzio。《圆形数据异常检测》。载于*SEAS 2023*短论文集，第63–68页，2023年。M. Di Marzio、S. Fensore、A. Panzera 和 C. C. Taylor。《含误差观测的圆形数据密度估计》《生物统计学》，78(1):248–260，2022年。D. E. Ferguson、H. F. Landreth 和 J. P. Mckeown。《北方蟋蟀蛙（Acris crepitans）的太阳罗盘定向》《动物行为》，15(1):45–53，1967年。J. J. Fern´andez-Dur´an。《基于非负三角和的圆形分布构建圆形–线性和圆形–圆形数据模型》《生物统计学》，63(2):579–585，2007年。N. I. Fisher。《圆形数据统计分析》《圆形数据统计分析》。剑桥大学出版社，1995年。L. Greco、G. Saraceno 和 C. Agostinelli。《多元圆形数据的包裹正态模型稳健拟合与异常检测》《统计学》，4(2):454–471，2021年。F. R. Hampel、E. M. Ronchetti、P. Rousseeuw 和 W. A. Stahel。《稳健统计：基于影响函数的方法》《稳健统计：基于影响函数的方法》。Wiley Interscience，纽约，1986年。F. Hassanzadeh。《多峰和偏斜圆形响应的平滑样条模型：气象学和海洋学中的应用》《环境计量学》，32(2):e2655，2021年。

J. L. Hintze 和 R. D. Nelson。小提琴图：箱线图与密度曲线的协同作用。《美国统计学家》，52(2):181–184，1998年。S. R. Jammalamadaka 和 A. Sengupta。圆统计学专题。世界科学出版社，2001。M. C. Jones 和 A. Pewsey。适用于圆形数据的逆 Batschelet 分布。《生物统计学》，68(1):183–193，2012。S. Kato 和 S. Eguchi。冯·米塞斯–Fisher 分布的位置和集中参数的稳健估计。《统计学论文》，57:205–234，2016。D. Ko 和 P. Guttorp。方向数据估计量的稳健性。《统计学年鉴》，16:609–618，1988年。C. Ley 和 T. Verdebout。现代方向统计学。 Chapman and Hall/CRC，2017。C. Ley 和 T. Verdebout。应用方向统计学。现代方法与案例研究。 CRC出版社，Taylor and Francis，2019。C. Leys、C. Ley、O. Klein、P. Bernard 和 L. Licata。检测离群值：不要使用均值周围的标准差，而应使用中位数周围的绝对偏差。《实验社会心理学杂志》，49:764–766，2013年。U. Lund 和 C. Agostinelli。Circular。R 软件包版本 0.5-0。https://cran.r-project.org/web/packages/circular/circular.pdf，2013年。E. A. Mahmood、S. Rana、H. Midi 和 A. G. Hussin。使用稳健圆距离检测单变量圆形数据中的离群值。《现代应用统计方法杂志》，16:418–438，2017。K. V. Mardia。方向数据统计学。学术出版社，伦敦，1972。K. V. Mardia。方向数据统计学。《英国皇家统计学会杂志 *B*辑：统计方法学》，37(3):349–371，1975。K. V. Mardia、C. C. Taylor 和 G. K. Subramaniam。角度数据的蛋白质生物信息学与双变量冯·米塞斯分布混合。《生物统计学》，63(2):505–512，2007。R. D. Martin 和 R. H. Zamar。正随机变量尺度的渐近最小最大偏差稳健 M 估计量。《美国统计学会杂志》，84(406):494–501，1989年。

N. J. Mekkes、M. Groot、E. Hoekstra、A. de Boer、E. Dagkesamanskaia、S. Bouwman、...、I. R. Holtman。利用自然语言处理识别神经退行性疾病的临床疾病轨迹。《自然医学》，30(4):1143–1153，2024年。I. B. Mohamed、A. Rambli、N. Khaliddin 和 A. I. N. Ibrahim。基于间距理论的圆形数据新型差异性检验。《统计学通讯：模拟与计算》，45(8):2904–2916，2016。C. M. Montalcini 和 P. J. Rousseeuw。bixplot：一种适用于双峰数据的箱线图变体。*arXiv*，2510.09276，2025年。S. Nagy、H. Demni、D. Buttarazzi 和 G. C. Porzio。用于方向数据分类的角深度理论。《数据分析与分类进展》，18(3):627–662，2024年。B. S. Otieno 和 C. M. Anderson-Cook。环境与生态圆形数据的优选方向度量。《环境与生态统计》，13:311–324，2006。S. Rana、E. A. Mahmood、H. Midi 和 A. G. Hussin。简单圆形回归模型中响应变量和解释变量离群值的稳健检测。《马来西亚数学科学杂志》，10(3):399–414，2016。M. Ranalli 和 A. Maruotti。含噪纵向圆形数据的基于模型的聚类及其在动物运动中的应用。《环境计量学》，31(2):e2572，2020。P. J. Rousseeuw。最小中位数平方回归。《美国统计学会杂志》，79:871–880，1984。P. J. Rousseeuw 和 A. M. Leroy。《稳健回归与离群值检测》。约翰·威立父子公司，1987。P. J. Rousseeuw 和 M. Hubert。基于稳健统计的异常检测。《威立跨学科评论：数据挖掘与知识发现》，8(2):e1236，2018年。W. D. Schutte 和 J. W. H. Swanepoel。SOPIE：一种用于非参数估计脉冲星光变曲线脉冲间歇区间的 R 软件包。《皇家天文学会月刊》，461(1):627–640，2016。S. Seader、P. Tenenbaum、J. M. Jenkins 和 C. J. Burke。用于凌星行星探测的 χ^2 判别器：基于开普勒数据。《天体物理学杂志增刊系列》，206(2):25，2013年。M. F. Sau 和 D. Rodriguez。用于方向数据和离群值检测的最小距离法。《数据分析与分类进展》，12:587–603，2018年。

R. Tanious 和 R. Manolov。小提琴图作为单案例实验设计荟萃分析中的可视化工具。*Methodology*，18(3):221–238，2022年。

S. V. Tishkovskaya与P. G. Blackwell。基于动物移动数据对异质环境的贝叶斯估计。《环境计量学》，32(6):e2679，2021。

补充材料

A：稳健离散度量

引理1的证明

为证明引理1，我们首先陈述一个中间引理。

引理7。 设 Θ 为一个圆周随机变量，且 $\theta^* \in [-\pi, \pi)$ 满足 $P(\Theta = \theta^*) \geqslant 0.5$，则圆周中位数集 $M_\theta = \{\theta^*\}$。也就是说， θ^* 是 Θ的圆中位数。

证明。 考虑一个值 $\delta \in [-\pi, \pi)$，使得 $\delta > \theta^*$。我们证明

$$\mathbb{E}(d(\Theta, \delta)) \geqslant \mathbb{E}(d(\Theta, \theta^*)) \quad \forall\, \delta > \phi^*. \tag{25}$$

由于情形 $\delta < \theta^*$ 可由对称性推出，因此根据圆中位数的定义，有 $M_\theta = \theta^*$。

令 $\gamma \in [-\pi, \pi)$。由弧距离的性质可知，

$$d(\gamma, \delta) - d(\gamma, \theta^*) = \begin{cases} d(\delta, \theta^*) & \text{if } \gamma \leqslant \theta^* \\ > -d(\delta, \theta^*) & \text{if } \gamma > \theta^* \end{cases} \qquad \forall\, \delta > \theta^*\,.$$

因此，考虑差值 $\mathbb{E}(d(\Theta, \delta)) - \mathbb{E}(d(\Theta, \theta^*))$ 并取条件期望，可得 $\mathbb{E}(d(\Theta, \delta) - d(\Theta, \phi^*)) \geqslant d(\delta, \theta^*)P(\Theta \leqslant \theta^*) - d(\delta, \theta^*)P(\Theta > \theta^*) = d(\delta, \theta^*)(P(\Theta \leqslant \theta^*) - P(\Theta > \theta^*)) \geqslant 0$，因为根据假设，$P(\Theta = \theta^*) \geqslant 0.5$。 □

注意，利用相同的论证，若 $P(\Theta = \theta^*) > 0.5$，则 $\mathbb{E}(d(\Theta, \delta)) > \mathbb{E}(d(\Theta, \theta^*))\ \forall\, \delta > \theta^*$，且 θ^* 是唯一极小值点。因此， $\Theta \in \mathcal{M}_1$。

引理1的证明如下，逐个推导等式。

证明。 如果 $\exists \theta$：$P(\Theta = \theta) > 0.5$，那么根据上述引理， $M_\theta = \theta$和 $\Theta \in \mathcal{M}_1$。因此，我们有 $P(d(\Theta, \theta) = 0) > 0.5$，且 $P(d(\Theta, \theta) \geqslant 0) > 0.5$。于是，根据线性中位数的定义，$\mathrm{CMAD}(F) = 0$。

若 $\exists \theta$：$P(\Theta = \theta) = 0.5$，则根据上述引理， θ属于集合 M_θ。如果该集合包含多个元素，则 CMAD 未定义。若该集合是单元素集，则距离分布 $d(\Theta, \theta)$ 的中位数可能等于零，也可能以零为（线性）中位数集的端点。第一种情况（例如当 $F = 0.5\Delta_\theta + 0.5G$，其中 Δ_θ 是位于 θ 的点质量分布，且 $G \sim$ 为均匀分布 $(-\pi, \pi)$ ）下，$\mathrm{CMAD}(F) = 0$。第二种情况（例如，若 $F = 0.2\Delta_{\theta-\pi/8} + 0.5\Delta_\theta + 0.3\Delta_{\theta+\pi/6}$），$\mathrm{CMAD}(F) > 0$，其中 CMAD 是一个区间的中点

该区间长度为正，且以零为端点（这是因为距离 θ 的分布的累积分布函数从零开始到某个点为止，函数值均为 0.5）。最后，由于距离分布有界于 π，我们也有 CMAD(F) $< \pi/2$。等号无法成立，因为这必然要求 $F = 0.5\Delta_\theta + 0.5\Delta_{\theta\pm\pi}$，而这确实属于 $\Theta \notin \mathcal{M}_1$ 的一种情形。

如果 $\forall\,\theta\, P(\Theta = \theta) < 0.5$，那么来自 θ 的距离分布不可能具有等于零的中位数，因此 CMAD(F) > 0。此外，该中位数不可能达到 π。由于圆是有界的，最短弧距离分布的最大可达值为 π，因此 CMAD(F) $\leqslant \pi$。然而，要使 CMAD $= \pi$，距离分布的（线性）中位数应等于其最大可达值。只有当至少一半的概率位于距离分布的上尾时，这种情况才可能发生；这又意味着至少一半的概率点位于圆上的某处。然而，根据上述引理，CMAD 要么等于 0，要么等于一个包含 0 的集合，而不会等于 π。因此，CMAD(F) $< \pi,\ \forall\,\Theta \in \mathcal{M}_1$。

□

定理1的证明 为证明该定理，我们将证明存在一个分布族，使 CMAD 的上确界为 π。首先，考虑分布

$$F = \lambda\Delta_0 + 0.5(1-\lambda)\Delta_{c\pi} + 0.5(1-\lambda)\Delta_{-c\pi},\quad 0 \leqslant \lambda < 0.5,\quad 0 < c < 1.$$

当 $c > 0$ 时。注意，我们令 $\lambda < 0.5$，因为如果 $\lambda \geqslant 0.5$，则根据引理1，我们会得到 $0 \leqslant$ CMAD(F) $< \pi/2$。情形 $c = 0$ 并不重要（我们会在 0 处得到一个点质量，且 CMAD(F) 再次等于零）。对于 $c = 1$，F 变为 $\lambda\Delta_0 + (1-\lambda)\Delta_\pi$，并且根据引理1，当 $\lambda < 0.5$ 时，CMAD(F) $= 0$。

对于这些分布，我们推导期望值 $\mathbb{E}(d(\theta,\phi))$，将其表示为一般角度$\theta$的函数。为简化起见，且根据分布关于 0 的对称性，不失一般性，我们只考察 $0 \leqslant \theta \leqslant \pi$。我们有：

$$\begin{aligned}\mathbb{E}(d(\Theta,\theta)) &= (1-\lambda)d(\theta,0) + 0.5\lambda d(\theta,c\pi) + 0.5\lambda d(\theta,-c\pi)\\ &= (1-\lambda)\theta + 0.5\lambda(d(\theta,c\pi) + d(\theta,-c\pi)).\end{aligned}$$

我们将区分三种情况。首先，假设 $0 \leqslant \theta \leqslant -c\pi + \pi$。若如此，则有$d(\theta,c\pi) + d(\theta,-c\pi) = 2c\pi$。若 $-c\pi + \pi \leqslant \theta \leqslant c\pi$，则 $d(\theta,c\pi) + d(\theta,-c\pi) = 2((c\pi - \theta) + (\pi - c\pi)) = 2(\pi - \theta)$。最后，若 $c\pi \leqslant \theta \leqslant \pi$，则 $d(\theta,c\pi) + d(\theta,-c\pi) = 2(1-c)\pi$。简而言之，对于$0 \leqslant \theta \leqslant \pi$，我们可以写成：

$$\mathbb{E}(d(\Theta,\theta)) = \lambda\theta + (1-\lambda) * \begin{cases} c\pi & 0 \leqslant \theta \leqslant (1-c)\pi \\ \pi - \theta & (1-c)\pi < \theta \leqslant c\pi \\ (1-c)\pi & c\pi < \theta \leqslant \pi\,. \end{cases} \tag{26}$$

这是一个分段线性函数，从 0 增加到 $(1-c)\pi$，从 $(1-c)\pi$ 下降至$c\pi$，然后又从 $c\pi$ 增加到 π。因此，将结果推广到整个圆后，中位数集 M_θ 只能是 $\{0\}$ 或 $\{c\pi, -c\pi\}$。在后一种情况下，CMAD 未定义。

我们关注第一种情况。根据圆中位数的定义，当 $\mathbb{E}(d(\Theta, 0)) < \mathbb{E}(d(\Theta, c\pi))$ 时，该情况成立。也就是说，根据公式（26），若

$$(1-\lambda)c\pi < \lambda c\pi + (1-\lambda)(1-c)\pi. \tag{27}$$

只要满足 $c < (1-\lambda)/(2-3\lambda)$，这个条件就成立。因此，圆周随机变量 Θ'满足

$$F' := \lambda\Delta_0 + 0.5(1-\lambda)\Delta_{c\pi} + 0.5(1-\lambda)\Delta_{-c\pi}, \quad 0 \leqslant \lambda < 0.5, \quad 0 < c < (1-\lambda)/(2-3\lambda), \tag{28}$$

按构造有 $M_{\theta'} = \{0\}$。此外，$\mathrm{CMAD}(F') = c\pi$。这是因为从 $M_{\theta'}$ 出发的距离分布是两个点质量分布的线性组合：一个位于 0 处（概率为 $\lambda < 0.5$），另一个位于 $c\pi$ 处。因此，鉴于 $\lim_{\lambda\to 0.5}(1-\lambda)/(2-3\lambda) = 1$,

$$\sup_{\lambda,\, c} \mathrm{CMAD}(F') = \sup_{\lambda,\, c} c\pi = \pi.$$

定理 3 的证明

($\Rightarrow$) 若 $\Theta \sim H_{0.5}$，则每个概率质量至少为 0.5 的弧，其长度恰好等于 π。由公式（8）可得结论。

（$\Leftarrow$）假设 Θ 服从分布 $H \neq H_{0.5}$。根据定义，存在至少一个半圆 HC^*，使得 $\int_{HC^*} dH(\theta) \neq 0.5$。若如此，则要么 $\int_{HC^*} dH(\theta) > 0.5$ 且 $\int_{HC^*} dH(\theta) < 0.5$，要么 $\int_{HC^*} dH(\theta) < 0.5$ 且 $\int_{HC^*} dH(\theta) > 0.5$（此处，$HC^*$ 是 HC^*的补集 $\overline{\mathcal{S}}$）。在这两种情况下，都至少存在一段长度为 $\ell(A) = a < \pi$ 的弧 A，使得 $\int_A dH(\theta) = 0.5$。因此，由公式（8）可得 $\mathrm{CLMS}(H) = a/2 < \pi/2$，这与假设矛盾。

B：影响函数

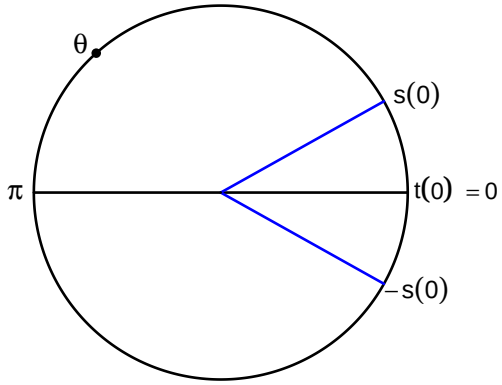
定理 6 的证明

我们先讨论 CMAD。未受污染分布 F 的总体圆形中位数是唯一的，因为 F 严格单峰，且该中位数等于对称中心 0。因此，从总体中位数出发的中位数弧长等于

$$\mathrm{CMAD}(F) = F^{-1}\Big(\frac{3}{4}\Big) - F^{-1}\Big(\frac{1}{2}\Big) = F^{-1}\Big(\frac{3}{4}\Big)\,.$$

为简洁起见，我们将 $q := F^{-1}(3/4)$ 记作.....

现在让我们污染 F，得到 $F_\varepsilon := (1-\varepsilon)F + \varepsilon\Delta_\theta$。累积分布函数$F_\varepsilon$ 在 $[-\pi, \theta[$ 上连续，其中 F_ε 的表达式为 $F_\varepsilon(u) = (1-\varepsilon)F(u)$。我们将受污染分布 F_ε 的圆中位数记作 $t(\varepsilon) := T(F_\varepsilon)$。未受污染分布具有 $\varepsilon = 0$，因此 $t(0) = 0$，如图 S1 所示。我们还将 F_ε 的 CMAD 记作 $s(\varepsilon)$，其中 $s(0)$ 是未受污染分布 F 的 CMAD。对于 $\varepsilon > 0$，图形会发生变化，因为 $t(\varepsilon)$ 和 $s(\varepsilon)$ 都可能移动。此时 CMAD 区间的端点变为$t(\varepsilon) - s(\varepsilon)$ 和 $t(\varepsilon) + s(\varepsilon)$。



Fi图 S1：CMAD 证明第一部分的示意图，其中 $q < \theta < \pi$.

我们首先考虑污染点 $-\pi < \theta < \pi$ 位于原区间之外的情形，即 $|\theta| > s(0)$。由于 F 的对称性，不失一般性，我们可以假设 $\theta > s(0)$，如图 S1 所示。注意，对于足够小的 ε，我们仍有 $\theta > t(\varepsilon) + s(\varepsilon)$，因为$\theta > q = \lim_{\varepsilon\downarrow 0} s(\varepsilon) = \lim_{\varepsilon\downarrow 0}(t(\varepsilon) + s(\varepsilon))$。

根据 CMAD 的定义，两个端点之间的质量必须为二分之一，因此

$$F_\varepsilon\big(t(\varepsilon)+s(\varepsilon)\big) - F_\varepsilon\big(t(\varepsilon)-s(\varepsilon)\big) = \frac{1}{2}$$
$$(1-\varepsilon)F\big(t(\varepsilon)+s(\varepsilon)\big) - (1-\varepsilon)F\big(t(\varepsilon)-s(\varepsilon)\big) = \frac{1}{2}\,.$$

对该等式两边关于 ε求导，得到

$$-1\Big(F(s(0)) - F(-s(0))\Big) + f(s(0))\Big(\frac{\partial t(\varepsilon)}{\partial\varepsilon} + \frac{\partial s(\varepsilon)}{\partial\varepsilon}\Big)$$
$$-f(-s(0))\Big(\frac{\partial t(\varepsilon)}{\partial\varepsilon} - \frac{\partial s(\varepsilon)}{\partial\varepsilon}\Big) = 0$$
$$-\frac{1}{2} + f(s(0))\Big(2\frac{\partial s(\varepsilon)}{\partial\varepsilon}\Big) = 0$$
$$\mathrm{IF}(\theta; \mathrm{CMAD}, F) = \frac{\partial s(\varepsilon)}{\partial\varepsilon} = \frac{1}{4f(F^{-1}(\frac{3}{4})}$$

其中 $\partial t_{(\varepsilon)}/\partial\varepsilon$ 已抵消。

现在考虑 $|\theta| < q$ 的情形。图 S2 展示了 $\theta > 0$ 的情况，但情形$\theta \leqslant 0$ 类似。对于足够小的 ε，有 $\theta < t(\varepsilon) + s(\varepsilon)$，因为 $y < s(0) = \lim_{\varepsilon \downarrow 0}(t(\varepsilon) + s(\varepsilon))$。

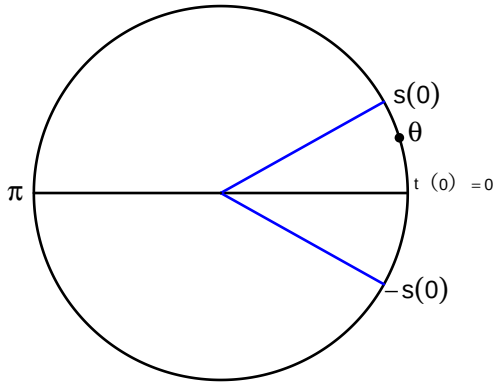


图 S2：CMAD 证明第二部分的示意图，其中 $-q < \theta < q$。

根据 CMAD 的定义，端点之间的 F_ε 的质量仍为二分之一，但由于$t(\varepsilon) + s(\varepsilon) > \theta$，$F_\varepsilon$ 的表达式多出一项 ε，因此得到

$$F_\varepsilon\big(t(\varepsilon)+s(\varepsilon)\big) - F_\varepsilon\big(t(\varepsilon)-s(\varepsilon)\big) = \frac{1}{2}$$
$$(1-\varepsilon)F\big(t(\varepsilon)+s(\varepsilon)\big) + \varepsilon - (1-\varepsilon)F\big(t(\varepsilon)-s(\varepsilon)\big) = \frac{1}{2}\,.$$

对该等式两边关于 ε 求导，得到

$$-1\Big(F(s(0)) - F(-s(0))\Big) + 1 + f(s(0))\Big(\frac{\partial t(\varepsilon)}{\partial\varepsilon} + \frac{\partial s(\varepsilon)}{\partial\varepsilon}\Big)$$
$$-f(-s(0))\Big(\frac{\partial t(\varepsilon)}{\partial\varepsilon} - \frac{\partial s(\varepsilon)}{\partial\varepsilon}\Big) = 0$$
$$-\frac{1}{2} + 1 + f(s(0))\Big(2\frac{\partial s(\varepsilon)}{\partial\varepsilon}\Big) = 0$$
$$\mathrm{IF}(\theta;\mathrm{CMAD},F) = \frac{\partial s(\varepsilon)}{\partial\varepsilon} = \frac{-1}{4f(F^{-1}(\frac{3}{4}))}\,.$$

结合 $|\theta| > q$ 和 $|\theta| < q$ 的表达式，即得定理中的公式。

现在我们转向　　圆形最小中位散布。由于对 F 的假设，可以通过将　　i-

目标函数的一阶导数设为零来求解式 (16) 的最小值:

$$\begin{aligned}
\frac{\partial}{\partial t}\frac{1}{2}\Big(F^{-1}(t+\frac{1}{2})-F^{-1}(t)\Big) &= 0\\
\frac{1}{f\big(F^{-1}(t+\frac{1}{2})\big)}-\frac{1}{f\big(F^{-1}(t)\big)} &= 0\\
F^{-1}(t+\frac{1}{2}) &= F^{-1}(1-t)\\
t+\frac{1}{2} &= 1-t\\
t &= \frac{1}{4}
\end{aligned}$$

因此

$$\mathrm{CLMS}(F)=\frac{1}{2}\Big(F^{-1}\Big(\frac{3}{4}\Big)-F^{-1}\Big(\frac{1}{4}\Big)\Big)=F^{-1}\Big(\frac{3}{4}\Big)\,.$$

我们再次记 $q := F^{-1}(3/4)$。

现在我们转而考虑受污染分布 $F_\varepsilon := (1-\varepsilon)F + \varepsilon\Delta_\theta$。我们首先考虑 $q < |\theta| \leqslant \pi$ 的情形。不失一般性，我们假设 $\theta > q$；由于对称性， $\theta < -q$ 的情形类似。图 S3 示意了这种情形。分布函数 F_ε 在 $[-\pi, \theta[$ 上连续，其中 F_ε 的表达式为 $F_\varepsilon(u) = (1-\varepsilon)F(u)$。因此，严格位于 θ 左侧的总质量为 $P_\varepsilon[-\pi < u < \theta] = (1-\varepsilon)F(\theta)$。对于任意 $\varepsilon < (F(\theta) - F(q))/F(\theta)$，均有 $(1-\varepsilon)F(\theta) > F(\theta) - F(\theta) + F(q) = F(q) = \frac{3}{4}$。因此，严格位于 θ 左侧的 F_ε 的总质量大于 3/4。这就是我们寻找最短半区间的位置。

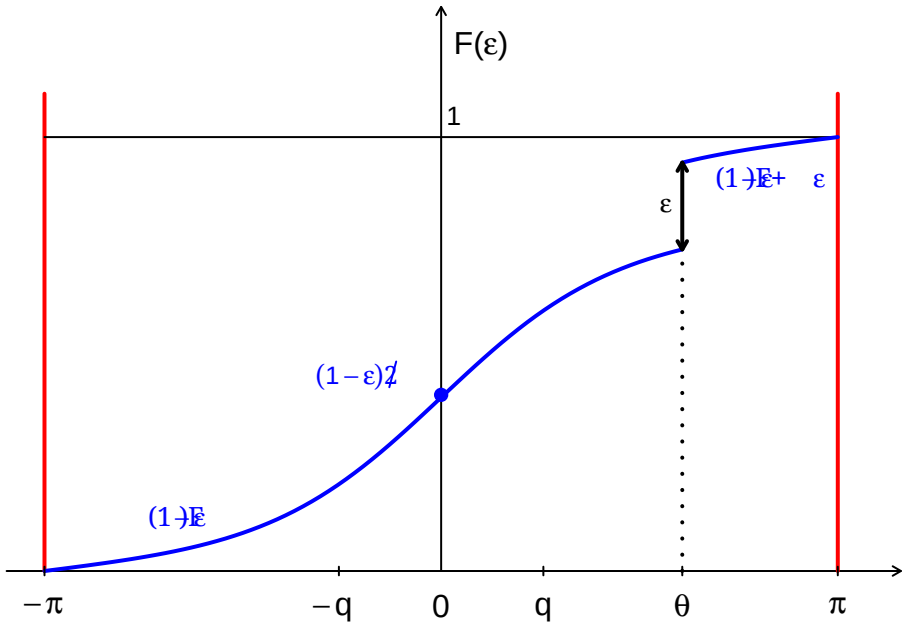


图 S3：CLMS 证明第一部分的示意图，其中 $q < \theta \leq \pi$。

在 F_ε 光滑的这一区域内对 (16) 求最小值，可得

$$\begin{aligned}
\frac{\partial}{\partial t}\frac{1}{2}\Big(F_\varepsilon^{-1}(t+\frac{1}{2})-F_\varepsilon^{-1}(t)\Big)&=0\\
\frac{\partial}{\partial t}\Big(F^{-1}\Big((t+\frac{1}{2})/(1-\varepsilon)\Big)-F^{-1}\Big(t/(1-\varepsilon)\Big)&=0\\
f\Big(F^{-1}\Big((t+\frac{1}{2})/(1-\varepsilon)\Big)\Big)&=f\Big(F^{-1}\Big(t/(1-\varepsilon)\Big)\Big)\\
F^{-1}\Big((t+\frac{1}{2})/(1-\varepsilon)\Big)&=F^{-1}\Big(1-t/(1-\varepsilon)\Big)\\
(t+\frac{1}{2})/(1-\varepsilon)&=(1-\varepsilon-t)/(1-\varepsilon)\\
t&=(1-2\varepsilon)/4
\end{aligned}$$

因此，CLMS(F_ε) 变为

$$\begin{aligned}
\mathrm{CLMS}(F_\varepsilon)&=\frac{1}{2}\Big(F_\varepsilon^{-1}\Big((1-2\varepsilon)/4+\frac{1}{2}\Big)-F_\varepsilon^{-1}\Big((1-2\varepsilon)/4\Big)\Big)\\
&=\frac{1}{2}\Big(F^{-1}\Big((3-2\varepsilon)/(4(1-\varepsilon))\Big)-F^{-1}\Big((1-2\varepsilon)/(4(1-\varepsilon))\Big)\Big)\,.
\end{aligned}$$

求导可得此 θ中的影响函数:

$$\begin{aligned}
\mathrm{IF}(\theta,\mathrm{CLMS},F)&=\frac{\partial}{\partial\varepsilon}\,\mathrm{CLMS}(F_\varepsilon)\Big|_{\varepsilon=0+}\\
&=\frac{1}{2f(F^{-1}(3/4))}\,\frac{\partial}{\partial\varepsilon}\,\frac{3-2\varepsilon}{4(1-\varepsilon)}\Big|_{\varepsilon=0+}\\
&\quad-\frac{1}{2f(F^{-1}(3/4))}\,\frac{\partial}{\partial\varepsilon}\,\frac{1-2\varepsilon}{4(1-\varepsilon)}\Big|_{\varepsilon=0+}\\
&=\frac{1}{2f(F^{-1}(3/4))}\,\frac{\partial}{\partial\varepsilon}\,\frac{1}{2(1-\varepsilon)}\Big|_{\varepsilon=0+}\\
&=\frac{1}{4\,f(F^{-1}(3/4))}\,.
\end{aligned}$$

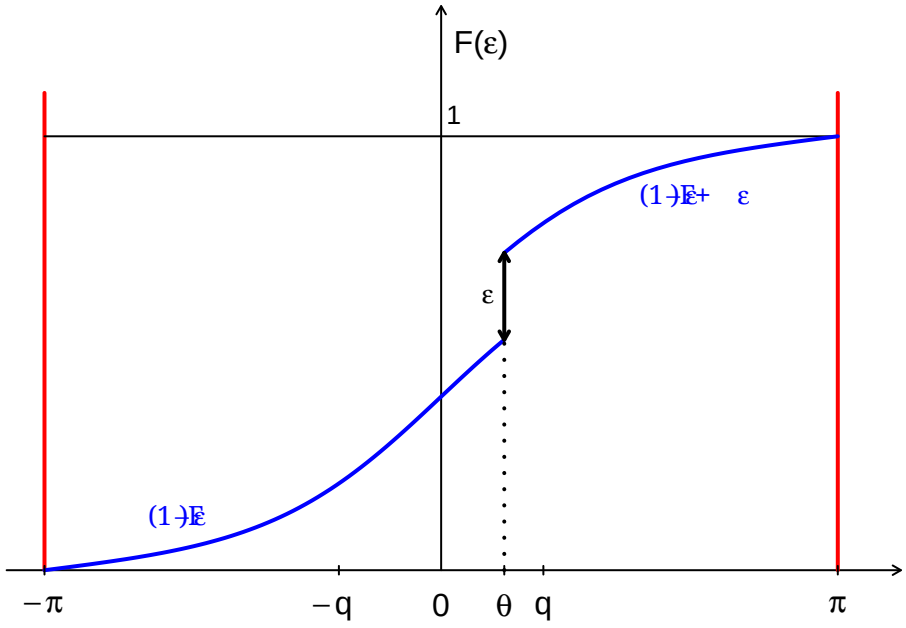


图 S4：CLMS 证明第二部分的示意图，其中 $-q < \theta < q$。

我们还会遇到如图 S4 所示的情况，其中 $-q < \theta < q$。区间的上端点现在位于满足 $F_\varepsilon(u) = (1-\varepsilon)F(u) + \varepsilon$ 的区域，因此最小化 (16) 可得

$$\begin{aligned}
\frac{\partial}{\partial t}\frac{1}{2}\Big(F_\varepsilon^{-1}(t+\frac{1}{2}) - F_\varepsilon^{-1}(t)\Big) &= 0 \\
\frac{\partial}{\partial t}\Big(F^{-1}\Big((t+\frac{1}{2}-\varepsilon)/(1-\varepsilon)\Big) - F^{-1}\Big(t/(1-\varepsilon)\Big) &= 0 \\
f\Big(F^{-1}\big((t+\frac{1}{2}-\varepsilon)/(1-\varepsilon)\big)\Big) &= f\Big(F^{-1}\Big(t/(1-\varepsilon)\Big)\Big) \\
F^{-1}\Big((t+\frac{1}{2}-\varepsilon)/(1-\varepsilon)\Big) &= F^{-1}\Big(1-t/(1-\varepsilon)\Big) \\
(t+\frac{1}{2}-\varepsilon)/(1-\varepsilon) &= (1-\varepsilon-t)/(1-\varepsilon) \\
t &= 1/4
\end{aligned}$$

因此 CLMS(F_ε) 变为

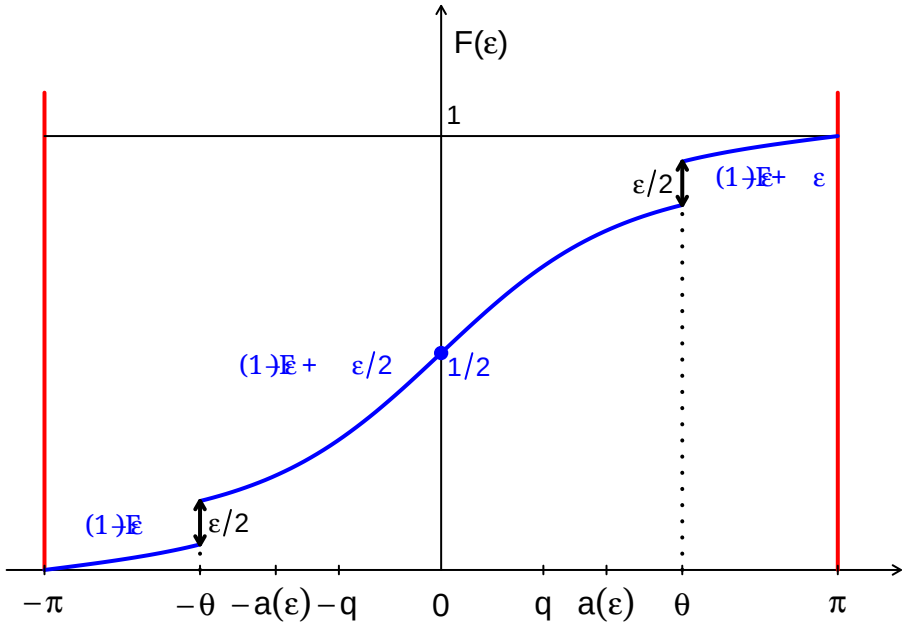
$$\begin{aligned}
\mathrm{CLMS}(F_\varepsilon) &= \frac{1}{2}\Big(F_\varepsilon^{-1}\Big(3/4\Big) - F_\varepsilon^{-1}\Big(1/4\Big)\Big) \\
&= \frac{1}{2}\Big(F^{-1}\Big((3/4-\varepsilon)/(1-\varepsilon)\Big) - F^{-1}\Big((1/4)/(1-\varepsilon)\Big)\Big) \\
&= \frac{1}{2}\Big(F^{-1}\Big((3-4\varepsilon)/(4(1-\varepsilon))\Big) - F^{-1}\Big(1/(4(1-\varepsilon))\Big)\Big)\,.
\end{aligned}$$

求导得到该 θ 中的影响函数:

$$\begin{aligned}
\mathrm{IF}(\theta, \mathrm{CLMS}, F) &= \frac{\partial}{\partial \varepsilon} \mathrm{CLMS}(F_\varepsilon)\Big|_{\varepsilon=0+} \\
&= \frac{1}{2f(F^{-1}(3/4))} \frac{\partial}{\partial \varepsilon} \frac{3-4\varepsilon}{4(1-\varepsilon)}\Big|_{\varepsilon=0+} \\
&\quad - \frac{1}{2f(F^{-1}(3/4))} \frac{\partial}{\partial \varepsilon} \frac{1}{4(1-\varepsilon)}\Big|_{\varepsilon=0+} \\
&= \frac{1}{2f(F^{-1}(3/4))} \frac{\partial}{\partial \varepsilon} \frac{1-2\varepsilon}{2(1-\varepsilon)}\Big|_{\varepsilon=0+} \\
&= \frac{-1}{4\, f(F^{-1}(3/4))} .
\end{aligned}$$

证明至此结束。

定理 7 的证明 我们将利用 IF 具有对称性的性质，因为 F 具有对称性，即 $\mathrm{IF}(-\theta, S, F) = \mathrm{IF}(\theta, S, F)$。因此，我们可以通过以对称方式污染 F 来计算 IF，令 $F_\varepsilon := (1-\varepsilon)F + \frac{\varepsilon}{2}\Delta_\theta + \frac{\varepsilon}{2}\Delta_{-\theta}$，如图 S5 所示。现在，累积分布函数 F_ε 在 $[-\pi, -\theta[$ 上连续，其中 $F_\varepsilon(u) = (1-\varepsilon)F(u)$，随后在 $-\theta$ 处向上跳跃 $\frac{\varepsilon}{2}$，接着为 $(1-\varepsilon)F(u) + \frac{\varepsilon}{2}$，然后在 θ 处再次向上跳跃，最后变为 $(1-\varepsilon)F(u) + \varepsilon$。与之前一样，$F_\varepsilon(-\pi) = 0$，且 $F_\varepsilon(\pi) = 1$。



Fi图 S5：CLTS 证明第一部分的示意图，其中 $q < \theta \leq \pi$。

我们先从 $|\theta| > q$ 且 $q = F^{-1}(3/4)$ 的情形开始。不失一般性地，我们可以假设 θ 为正，如图 S5 所示。我们将质量为 $1/2$ 且圆方差最小的区间记为$[-a(\varepsilon),\ a(\varepsilon)]$。对于足够小的 ε，有 $\theta > a(\varepsilon)$，因为我们知道 $\lim_{\varepsilon\downarrow 0} a(\varepsilon) = q < \theta$。这意味着 $[-a(\varepsilon),\ a(\varepsilon)]$ 完全位于 $F_\varepsilon(u) = (1-\varepsilon)F(u) + \frac{\varepsilon}{2}$ 的区域内，从而得到

$$\begin{aligned}\frac{3}{4} = F_\varepsilon(a(\varepsilon)) &= (1-\varepsilon)F(a(\varepsilon)) + \frac{\varepsilon}{2}\\ (1-\varepsilon)F(a(\varepsilon)) &= \frac{3-2\varepsilon}{4}\\ a(\varepsilon) &= F^{-1}\Big(\frac{3-2\varepsilon}{4-4\varepsilon}\Big).\end{aligned}$$

现在，$S(F_\varepsilon)$ 由公式 (10) 给出，其中 $\phi_1 = -a(\varepsilon)$ 和 $\phi_2 = a(\varepsilon)$。由 F_ε 在这两点的连续性可知，$P(-a(\varepsilon) \leqslant u \leqslant a(\varepsilon)) = \frac{1}{2}$（而不严格大于 $1/2$）。因此，(10) 变为

$$\begin{aligned}S(F_\varepsilon) &= 2\sqrt{\int_{-a(\varepsilon)}^{a(\varepsilon)} (1-\cos(u))dF_\varepsilon(u)}\\ &= 2\sqrt{0.5 - c(\varepsilon)}\end{aligned}$$

其中 $c(\varepsilon)$ 是积分

$$c(\varepsilon) := \int_{-a(\varepsilon)}^{a(\varepsilon)} \cos(u)dF_\varepsilon(u) = (1-\varepsilon)\int_{-a(\varepsilon)}^{a(\varepsilon)} \cos(u)f(u)du\,.$$

为了求得影响函数，我们必须对 $S(F_\varepsilon)$ 关于 $\varepsilon = 0+$ 求导，得到

$$\begin{aligned}\frac{\partial S(F_\varepsilon)}{\partial\varepsilon} &= 2\frac{\partial}{\partial\varepsilon}\sqrt{0.5-c(\varepsilon)} = \frac{2}{2\sqrt{0.5-c(0)}}\,\frac{\partial}{\partial\varepsilon}(-c(\varepsilon))\\ &= \frac{2}{S(F)}\Big[\int_{-a(0)}^{a(0)} \cos(u)f(u)du - \cos(a(0))f(a(0))\frac{\partial}{\partial\varepsilon}a(\varepsilon)\\ &\quad + \cos(-a(0))f(-a(0))\frac{\partial}{\partial\varepsilon}\big(-a(\varepsilon)\big)\Big]\\ &= \frac{2}{S(F)}\Big[\int_{-q}^{q} \cos(u)f(u)du - 2\cos(q)f(q)\frac{\partial}{\partial\varepsilon}F^{-1}\Big(\frac{3-2\varepsilon}{4-4\varepsilon}\Big)\Big]\\ &= \frac{2}{S(F)}\Big[c(0) - 2\cos(q)f(q)\frac{1}{f(F^{-1}(3/4))}\frac{\partial}{\partial\varepsilon}\Big(\frac{3-2\varepsilon}{4-4\varepsilon}\Big)\Big]\\ &= \frac{2}{S(F)}\Big[\frac{1}{2} - \frac{S^2(F)}{4} - \frac{1}{2}\cos(F^{-1}(3/4))\Big]\\ &= \frac{1}{S(F)}\Big[1 - \frac{S^2(F)}{2} - \cos(F^{-1}(3/4))\Big]\,.\end{aligned}$$

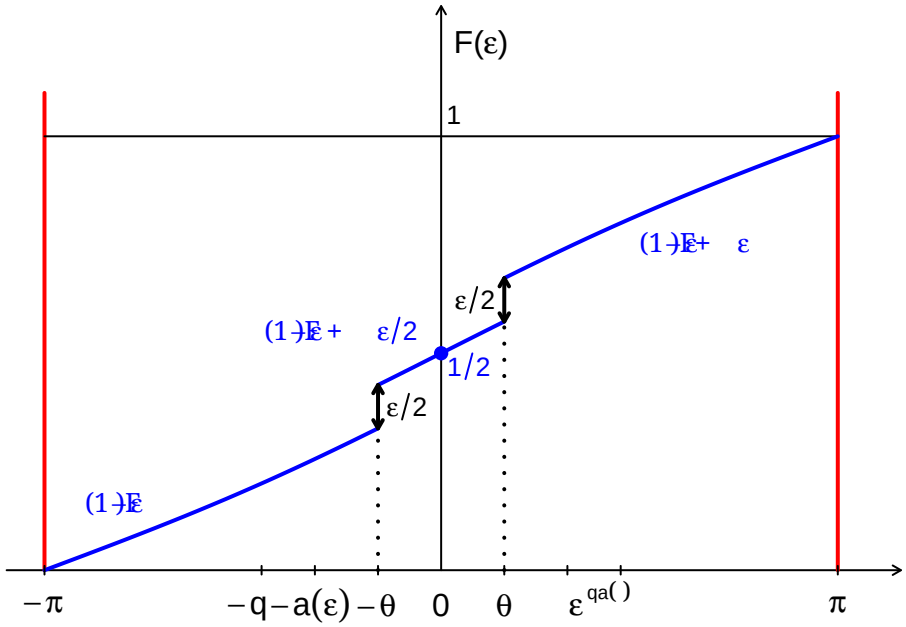


图 S6：CLTS 证明第二部分的示意图，其中 $-q < \theta < q$。

现在我们转而讨论 $|\theta| < q$ 所示的情形，如图 S6 所示。对于足够小的 ε，我们有 $\theta < a(\varepsilon)$，因为 $y < q = \lim_{\varepsilon \downarrow 0} a(\varepsilon)$。因此，$F_\varepsilon$ 的两个向上跳跃都发生在 $[-a(\varepsilon)\ a(\varepsilon)]$，内部，所以这一次

$$\begin{aligned}
\frac{3}{4} = F_\varepsilon(a(\varepsilon)) &= (1-\varepsilon)F(a(\varepsilon)) + \varepsilon \\
(1-\varepsilon)F(a(\varepsilon)) &= \frac{3-4\varepsilon}{4} \\
a(\varepsilon) &= F^{-1}\Big(\frac{3-4\varepsilon}{4-4\varepsilon}\Big)\,.
\end{aligned}$$

这里的 $S(F_\varepsilon)$ 也是 $[-a(\varepsilon)$ 上圆方差平方根的两倍 $a(\varepsilon)],$，但表达式会稍微变长，因为 $-\theta$ 和 θ 中的点质量位于区间内，因此现在

$$\begin{aligned}
c(\varepsilon) &= \int_{-a(\varepsilon)}^{a(\varepsilon)} \cos(u) dF_\varepsilon(u) \\
&= (1-\varepsilon)\int_{-a(\varepsilon)}^{a(\varepsilon)} \cos(u) dF(u) + \frac{\varepsilon}{2}\cos(\theta) + \frac{\varepsilon}{2}\cos(-\theta) \\
&= (1-\varepsilon)\int_{-a(\varepsilon)}^{a(\varepsilon)} \cos(u) dF(u) + \varepsilon\cos(\theta)\,.
\end{aligned}$$

这次我们得到

$$\begin{aligned}
\frac{\partial S(F_\varepsilon)}{\partial \varepsilon} &= \frac{\partial}{\partial \varepsilon} 2\sqrt{0.5 - c(\varepsilon)} = \frac{2}{2\sqrt{0.5 - c(0)}} \frac{\partial}{\partial \varepsilon}(-c(\varepsilon)) \\
&= \frac{2}{S(F)} \Big[\int_{-a(0)}^{a(0)} \cos(u) f(u) du - \cos(a(0)) f(a(0)) \frac{\partial}{\partial \varepsilon} a(\varepsilon) \\
&\quad + \cos(-a(0)) f(-a(0)) \frac{\partial}{\partial \varepsilon} \big(- a(\varepsilon) \big) - \frac{1}{2} \cos(\theta) - \frac{1}{2} \cos(-\theta) \Big] \\
&= \frac{2}{S(F)} \Big[\int_{-q}^{q} \cos(u) f(u) du - 2 \cos(q) f(q) \frac{\partial}{\partial \varepsilon} F^{-1}\Big(\frac{3 - 4\varepsilon}{4 - 4\varepsilon}\Big) - \cos(\theta) \Big] \\
&= \frac{2}{S(F)} \Big[c(0) - 2 \cos(q) f(q) \frac{1}{f(F^{-1}(3/4))} \frac{\partial}{\partial \varepsilon} \Big(\frac{3 - 4\varepsilon}{4 - 4\varepsilon}\Big) - \cos(\theta) \Big] \\
&= \frac{2}{S(F)} \Big[\frac{1}{2} - \frac{S^2(F)}{4} + \frac{1}{2} \cos(F^{-1}(3/4)) - \cos(\theta) \Big] \\
&= \frac{1}{S(F)} \Big[1 - \frac{S^2(F)}{2} + \cos(F^{-1}(3/4)) - 2 \cos(\theta) \Big] .
\end{aligned}$$

综合 $|\theta| > q$ 与 $|\theta| < q$ 的结果，即可得到该定理的陈述。

C：崩溃值

引理 6 的证明。 我们首先证明存在一个分布 H^*，使得崩溃值恰好等于公式（21）右端的表达式。然后，我们证明存在第二个分布 H^{**}，其崩溃值更低，尽管无法以闭式导出。

考虑分布 $G^* := (1 - \varepsilon)F + \varepsilon H^*$，其中 $H^* := 0.5\Delta_{\mu - \pi/2} + 0.5\Delta_{\mu + \pi/2}$。注意，由 H^* 关于 μ 的对称性，并且因为 F 在 μ 处具有唯一众数，且密度从 μ 到 $\mu \pm \pi$单调递减，我们可得 G^* 的中位数是唯一的，并且等于 F的中位数： $\tilde{\mu}_{G^*} = \tilde{\mu}_F = \mu$。

考虑使 CMAD(G^*) $= \pi/2$ 的 ε 的取值。根据 CMAD 的定义（公式 5），CMAD(G^*) $= \pi/2$，如果

$$P_{G^*}(\mu - \pi/2 < \theta < \mu + \pi/2) < 0.5 \tag{29}$$

并且如果

$$P_{G^*}(\mu - \pi/2 \leqslant \theta \leqslant \mu + \pi/2) \geqslant 0.5, \tag{30}$$

其中第一个条件确保在 G^* 下不存在短于 $\pi/2$ 、概率含量至少为 0.5 的弧。

第二个条件总是满足的。鉴于 $P_F(\mu - \pi/2 \leqslant \theta \leqslant \mu + \pi/2) \geqslant 0.5$和 $P_{H^*}(\mu - \pi/2 \leqslant \Theta \leqslant \mu + \pi/2) \geqslant 0.5$，它们的任意线性组合也都是 $\geqslant 0.5$。因此，对任意 $\varepsilon \leqslant 0.5$，(30) 都成立。

要满足第一个条件，必须施加一个约束。首先请注意，根据 H^* 的定义，$P_{H^*}(\mu-\pi/2<\Theta<\mu+\pi/2)=0$，因此

$$P_{G^*}(\mu-\pi/2<\Theta<\mu+\pi/2)=(1-\varepsilon)P_F(\mu-\pi/2<\Theta<\mu+\pi/2)\quad\forall\varepsilon.$$

因此，条件 (29) 变为

$$(1-\varepsilon)P_F(\mu-\pi/2<\Theta<\mu+\pi/2)<0.5,$$

若 $\varepsilon>1-0.5[P_F(\mu-\pi/2<\Theta<\mu+\pi/2)]^{-1}$，则该条件成立，因此

$$\inf\{\varepsilon:\mathrm{CMAD}(G^*)=\pi/2\}=1-0.5[P_F(\mu-\pi/2<\Theta<\mu+\pi/2)]^{-1}.$$

为完成证明，我们需要证明至少存在一个分布 H^{**}，使得 $\inf\{\varepsilon:\mathrm{CMAD}(G^{**})=\pi/2\}<\inf\{\varepsilon:\mathrm{CMAD}(G^*)=\pi/2\}$，其中 $G^{**}:=(1-\varepsilon)F+\varepsilon H^{**}$（否则，(21) 描述的就不是上限，而是崩溃值本身）。

设 $H^{**}:=(1-\alpha)\Delta_{\tilde{\mu}_{G^{**}}-\pi/2}+\alpha\Delta_{\tilde{\mu}_{G^{**}}+\pi/2}$，$0\leqslant\alpha\leqslant1$，$\alpha\neq0.5$。若如此，则 G^{**} 是非对称分布，并且 $\tilde{\mu}_{G^{**}}\neq\tilde{\mu}_F$，其中 $\tilde{\mu}_{G^{**}}$ 在圆上的位置取决于 α 的值，最重要的是，也取决于 ε 本身的值。因此，无法推导出闭式来求得 $\tilde{\mu}_{G^{**}}$ 的值。

另一方面，$\tilde{\mu}_{G^{**}}\in[\mu-\pi/2,\mu+\pi/2]$。首先，注意到对于任意的 ε 和 ψ，$\tilde{\mu}_{G^{**}}$ 都不可能与 μ对跖，因为 $\alpha\neq0.5$。此外，若 $\alpha<0.5$，则 $\tilde{\mu}_{G^{**}}\in[\mu-\pi/2,\mu]\ \forall\varepsilon,\ \psi$，反之亦然。

最终，为了使 $\mathrm{CMAD}(G^{**})=\pi/2$，从而使 ψ_{CMAD}崩溃，类比条件 (29) 和 (30)，我们必须满足

$$P_{G^{**}}(\tilde{\mu}_{G^{**}}-\pi/2<\Theta<\tilde{\mu}_{G^{**}}+\pi/2)<0.5 \tag{31}$$

and

$$P_{G^{**}}(\tilde{\mu}_{G^{**}}-\pi/2\leqslant\Theta\leqslant\tilde{\mu}_{G^{**}}+\pi/2)\geqslant0.5. \tag{32}$$

第二个条件始终成立，因为 $P_F(\tilde{\mu}_{G^{**}}-\pi/2\leqslant\theta\leqslant\tilde{\mu}_{G^{**}}+\pi/2)\geqslant0.5$，且 $P_{H^{**}}(\tilde{\mu}_{G^{**}}-\pi/2\leqslant\Theta\leqslant\tilde{\mu}_{G^{**}}+\pi/2)\geqslant0.5$；第一个条件成立是因为 $\tilde{\mu}_{G^{**}}\in[\mu-\pi/2,\mu+\pi/2]$，并且 F 对于每个包含 μ的半圆都有 $\geqslant0.5$ 概率。

要满足第一个条件，根据 H^{**} 的定义，$P_{G^{**}}(\tilde{\mu}_{G^{**}}-\pi/2<\Theta<\tilde{\mu}_{G^{**}}+\pi/2)=(1-\varepsilon)P_F(\tilde{\mu}_{G^{**}}-\pi/2<\Theta<\tilde{\mu}_{G^{**}}+\pi/2)$，我们需要 $(1-\varepsilon)P_F(\tilde{\mu}_{G^{**}}-\pi/2<\Theta<\tilde{\mu}_{G^{**}}+\pi/2)<0.5$，而这又在 $\varepsilon>1-0.5[P_F(\tilde{\mu}_{G^{**}}-\pi/2<\Theta<\tilde{\mu}_{G^{**}}+\pi/2)]^{-1}$ 时成立，因此

$$\inf\{\varepsilon:\mathrm{CMAD}(G^{**})=\pi/2\}=1-0.5[P_F(\tilde{\mu}_{G^{**}}-\pi/2<\Theta<\tilde{\mu}_{G^{**}}+\pi/2)]^{-1}.$$

证明完成，因为 $\inf\{\varepsilon:\mathrm{CMAD}(G^{**})=\pi/2\}<\inf\{\varepsilon:\mathrm{CMAD}(G^*)=\pi/2\}$given that $P_F(\mu-\pi/2<\Theta<\mu+\pi/2)>P_F(\tilde{\mu}_{G^{**}}-\pi/2<\Theta<\tilde{\mu}_{G^{**}}+\pi/2)$，且 $\tilde{\mu}_{G^{**}}\neq\mu$。

对于冯·米塞斯分布和包裹正态分布， $P_F(\mu-\pi/2<\Theta<\mu+\pi/2) > 0.5 \forall \psi > 0$。因此，式 (21) 给出的上界始终小于 0.5。

定理 8 的证明。根据爆炸崩溃值的定义， $\varepsilon^*_{\exp}(\psi_{\mathrm{CMAD}}) = \inf\{\varepsilon\colon \mathrm{CMAD}((1-\varepsilon)F+\varepsilon H) = 0\}$。由 Θ的连续性以及引理1可知，$\mathrm{CMAD}((1-\varepsilon)F+\varepsilon H) = 0$ 当且仅当 $H=\Delta_\theta$ 对某个 θ成立，且 $\varepsilon \geqslant 0.5$。结论得证。

D：渐近方差与相对效率

本节研究稳健离散度泛函 CMAD、CLMS 和 CLTS 在参数圆模型下的渐近行为。在给定模型中，对泛函进行变换，使其对于特定参数费希尔一致；这里该参数是冯·米塞斯模型的 κ 和包裹正态分布的 σ。我们将根据原泛函的渐近方差，推导变换后泛函的渐近方差。

在以下定理中，我们计算泛函 CMAD、CLMS、CLTS 及其导出估计量的渐近方差。

定理 11 （CMAD 和 CLMS 的渐近方差。）。 设 F 为密度函数为连续密度 $f(\theta) > 0$的圆形分布。令 $S(F)$ 表示 CMAD 或 CLMS 泛函，并令 $S_n = S(F_n)$ 表示基于经验分布 F_n的样本版本。S 的影响函数由定理 *6* 给出，估计量 S_n 的渐近方差为

$$\mathrm{AVar}(S_n) = \frac{1}{16\, f(F^{-1}(3/4))^2}\,.$$

证明 估计量 S_n 的渐近方差为

$$\mathrm{AVar}(S_n) = \mathbb{E}_F\left[\mathrm{IF}(\theta; S, F)^2\right],$$

其中 $\mathrm{IF}(\theta;\, S, F)$ 是 S在 F处的影响函数。对于 CMAD 或 CLMS，其影响函数为

$$\mathrm{IF}(\theta; S, F) = \frac{\operatorname{sign}\left(|\theta| - F^{-1}(3/4)\right)}{4\, f(F^{-1}(3/4))}\,.$$

平方后取期望可得

$$\mathbb{E}_F[\mathrm{IF}^2] = \int_{-\pi}^{\pi} \left(\frac{\operatorname{sign}(|\theta| - F^{-1}(3/4))}{4\, f(F^{-1}(3/4))}\right)^2 f(\theta)\, d\theta = \int_{-\pi}^{\pi} \frac{1}{16\, f(F^{-1}(3/4))^2} f(\theta)\, d\theta\,.$$

由于 $\int_{-\pi}^{\pi} f(\theta)\ d\theta = 1$，我们得到

$$\mathrm{AVar}(S_n) = \frac{1}{16\, f(F^{-1}(3/4))^2}\,.$$

定理12 (CLTS 的渐近方差。)。 设 F 为一个具有连续密度 $f(\theta) > 0$ 的圆形分布，该分布关于众数 0对称且严格单峰。设 $S(F)$ 表示 CLTS 泛函， $S_n = S(F_n)$ 表示基于经验分布 F_n的样本版本。 S 的影响函数由定理 7 给出，而 S_n 的渐近方差等于

$$\begin{aligned}\mathrm{AVar}(S_n) = \frac{1}{S^2(F)}\Bigg[&\left(1 - \frac{S^2(F)}{2} - \cos\big(F^{-1}(3/4)\big)\right)^2 \\ &+ 4\left(1 - \frac{S^2(F)}{2} - \cos\big(F^{-1}(3/4)\big)\right)\int_{F^{-1}(1/4)}^{F^{-1}(3/4)} \big(\cos(F^{-1}(3/4)) - \cos\theta\big)\, dF(\theta) \\ &+ 4\int_{F^{-1}(1/4)}^{F^{-1}(3/4)} \big(\cos(F^{-1}(3/4)) - \cos\theta\big)^2\, dF(\theta)\Bigg].\end{aligned}$$

证明 S 的影响函数为

$$\mathrm{IF}(\theta; S, F) = \frac{1}{S(F)}\Big[1 - \frac{S^2(F)}{2} - \cos(F^{-1}(3/4)) + 2\, I\big(|\theta| < F^{-1}(3/4)\big)\big(\cos(F^{-1}(3/4)) - \cos\theta\big)\Big],$$

因此，渐近方差等于

$$\begin{aligned}\mathrm{AVar}(S_n) &= \mathbb{E}_F\big[\mathrm{IF}(\theta; S, F)^2\big] \\ &= \frac{1}{S^2(F)}\,\mathbb{E}_F\Big[\Big(1 - \frac{S^2(F)}{2} - \cos F^{-1}(3/4) \\ &\qquad + 2\, I(|\theta| < F^{-1}(3/4))(\cos F^{-1}(3/4) - \cos\theta)\Big)^2\Big].\end{aligned}$$

展开平方 并利用 $F^{-1}(1/4) = -F^{-1}(3/4)$，可将积分写成 s

$$\begin{aligned}\mathrm{AVar}(S_n) = \frac{1}{S^2(F)}\Bigg[&\left(1 - \frac{S^2(F)}{2} - \cos(F^{-1}(3/4))\right)^2 \\ &+ 4\left(1 - \frac{S^2(F)}{2} - \cos F^{-1}(3/4)\right)\int_{F^{-1}(1/4)}^{F^{-1}(3/4)} (\cos F^{-1}(3/4) - \cos\theta)\, dF(\theta) \\ &+ 4\int_{F^{-1}(1/4)}^{F^{-1}(3/4)} (\cos F^{-1}(3/4) - \cos\theta)^2\, dF(\theta)\Bigg].\end{aligned}$$

定理13 （变换估计量的渐近方差。）。 令 S 表示CMAD、CLMS 或CLTS 泛函，并令 $\psi_S = \zeta(S)$ 为其对应的变换。令 $S_n = S(F_n)\widehat{\ }$和 $\psi_S = \zeta(S_n)$ 为基于经验分布 F_n的对应估计量。若 ζ 在 $S(F)$处连续可微，则 ψ 的渐近方差等于

$$\mathrm{AVar}(\widehat{\psi}_S) = \Big(\zeta'(S(F))\Big)^2\, \mathrm{AVar}(S_n).$$

证明。 ψ_S 的影响函数为

$$\mathrm{IF}(x;\psi_S,F)=\zeta'(S(F))\ \mathrm{IF}(x;S,F).$$

代入渐近方差公式

$$\mathrm{AVar}(\widehat{\psi}_S)=\int \mathrm{IF}(x;\psi_S,F)^2\,dF(x)$$

可得所述结果。 □

定理14 （$\widehat{\kappa}$ 在冯·米塞斯分布下的渐近方差。） 设 F 为集中参数为 κ的冯·米塞斯分布，且 CSD 为其总体圆标准差，该标准差与 κ 的关系为

$$\kappa=g(\mathrm{CSD})=A^{-1}\left(1-\frac{\mathrm{CSD}^2}{2}\right).$$

设 S_n 为CSD 的估计量，并通过 $\widehat{\kappa}=g\left(S_n\right)$ 定义 κ 的诱导估计量。那么， $\widehat{\kappa}$的渐近方差等于

$$\mathrm{AVar}(\widehat{\kappa})=\frac{\mathrm{CSD}^2}{A'(\kappa)^2}\,\mathrm{AVar}(S). \tag{33}$$

证明 对于冯·米塞斯分布

$$\rho=A(\kappa)=\frac{I_1(\kappa)}{I_0(\kappa)}\quad\text{and}\quad\mathrm{CSD}=\sqrt{2(1-\rho)}\,,$$

因此 $\rho=1-\mathrm{CSD}^2/2$。将 κ 表示为CSD的函数，得到

$$\kappa=A^{-1}\left(1-\frac{\mathrm{CSD}^2}{2}\right).$$

我们现在通过 $g(S):=A^{-1}(1-S^2/2)$定义函数 g，从而 $\kappa=g(S)$且 $\widehat{\kappa}=g(S_n)$。由于$A(\kappa)$连续可微且严格递增，函数 g 在CSD处可微，其导数为

$$g'(S)=\frac{d}{dR}A^{-1}(R)\Big|_{R=A(\kappa)}\frac{d}{dS}\left(1-\frac{S^2}{2}\right).$$

由于

$$\frac{d}{dR}A^{-1}(R)\Big|_{R=A(\kappa)}=\frac{1}{A'(\kappa)}\quad\text{and}\quad\frac{d}{dS}\left(1-\frac{S^2}{2}\right)=-S,$$

我们得到

$$g'(S(F))=-\frac{S(F)}{A'(\kappa)}=-\frac{\mathrm{CSD}}{A'(\kappa)}\,.$$

我们知道

$$\mathrm{AVar}(\widehat{\kappa})=\left(g'(S(F))\right)^2\ \mathrm{AVar}(S_n).$$

因此， $\widehat{\kappa}$ 的渐近方差确实为

$$\mathrm{AVar}(\widehat{\kappa})=\frac{\mathrm{CSD}^2}{\left(A'(\kappa)\right)^2}\,\mathrm{AVar}(S_n)\,.$$

定理 15 （包裹正态分布下 $\widehat{\sigma}$ 的渐近方差。）。 设 F 服从尺度参数为 σ 的包裹正态分布，并令 CSD表示圆标准差，它与 σ 的关系为

$$\sigma = g(\mathrm{CSD}) = \sqrt{-2\log\left(1 - \frac{\mathrm{CSD}^2}{2}\right)}.$$

设 S_n 为 *CSD* 的一个估计量，并定义 σ 的诱导估计量为 $\widehat{\sigma} = g(S_n)$。那么，渐近方差 $\mathrm{AVar}(\widehat{\sigma})$ 等于

$$\mathrm{AVar}(\widehat{\sigma}) = \frac{\mathrm{CSD}^2}{\sigma^2\left(1 - \frac{\mathrm{CSD}^2}{2}\right)^2}\,\mathrm{AVar}(S_n)\,. \tag{34}$$

证明 对于包裹正态分布，其平均合向量长度 is

$$\rho = e^{-\sigma^2/2}.$$

由于 ρ 与 CSD 满足关系 $\mathrm{CSD} = \sqrt{2(1-\rho)}$，因此有 $\rho = 1 - \mathrm{CSD}^2/2$。由此得到

$$e^{-\sigma^2/2} = 1 - \frac{\mathrm{CSD}^2}{2}.$$

解得 σ

$$\sigma = \sqrt{-2\log\left(1 - \frac{\mathrm{CSD}^2}{2}\right)}.$$

现在定义函数 g 如下

$$g(S) := \sqrt{-2\log\left(1 - \frac{S^2}{2}\right)},$$

于是 $\sigma = g(S)$ 且 $\widehat{\sigma} = g(S_n)$。函数 g 在 CSD 处连续可微，其导数为

$$g'(S) = \frac{d}{d\rho}\left(\sqrt{-2\log\rho}\right)\Big|_{\rho=1-\frac{S^2}{2}} \frac{d}{dS}\left(1 - \frac{S^2}{2}\right).$$

由于

$$\frac{d}{d\rho}\left(\sqrt{-2\log\rho}\right) = -\frac{1}{\rho\sqrt{-2\log\rho}} \quad \text{and} \quad \frac{d}{dS}\left(1 - \frac{S^2}{2}\right) = -S,$$

利用 $\rho = 1 - \frac{S^2}{2}$ 和 $\sigma = \sqrt{-2\log\rho}$，我们得到，

$$g'(S) = \frac{S}{\sigma\left(1 - \frac{S^2}{2}\right)}.$$

因为我们知道

$$\mathrm{AVar}(\widehat{\sigma}) = ((\eta^{-1})'(S(F)))^2\,\mathrm{AVar}(S_n)\,,$$

$\widehat{\sigma}$ 的渐近方差确实等于

$$\mathrm{AVar}(\widehat{\sigma}) = \frac{\mathrm{CSD}^2}{\sigma^2\left(1-\frac{\mathrm{CSD}^2}{2}\right)^2}\,\mathrm{AVar}(S_n)\,.$$

$\widehat{\psi}_{\mathrm{CMAD}}$、$\widehat{\psi}_{\mathrm{CLMS}}$和 $\widehat{\psi}_{\mathrm{CLTS}}$的效率

通过考察所提出的稳健估计量在假设模型下相对于最大似然估计量（MLE）的渐近相对效率（ARE），评估其效率。对于感兴趣参数 ψ （例如，冯·米塞斯分布的 κ ，或包裹正态分布的 σ ），估计量 ψ 的渐近相对效率定义为

$$\mathrm{ARE}(\widehat{\psi}) = \frac{\mathrm{AVar}(\widehat{\psi}_{\mathrm{MLE}})}{\mathrm{AVar}(\widehat{\psi})}.$$

对于冯·米塞斯分布，集中参数 κ 的最大似然估计量 $\widehat{\kappa}_{\mathrm{MLE}}$ 渐近服从正态分布，其方差为

$$\mathrm{AVar}(\widehat{\kappa}_{\mathrm{MLE}}) = \frac{1}{I(\kappa)},$$

其中，费希尔信息为

$$I(\kappa) = \frac{1}{2}\left(1+\frac{I_2(\kappa)}{I_0(\kappa)}\right) - \left(\frac{I_1(\kappa)}{I_0(\kappa)}\right)^2,$$

其中， $I_0(\cdot)$ 、 $I_1(\cdot)$ 和 $I_2(\cdot)$ 分别表示阶数为0、1和2的第一类修正贝塞尔函数。

在已知 μ的包裹正态模型下，尺度参数 σ 的最大似然估计量也渐近服从正态分布。具体而言，

$$\sqrt{n}\,(\widehat{\sigma}_{\mathrm{MLE}} - \sigma) \xrightarrow{d} \mathcal{N}\big(0,\, I(\sigma)^{-1}\big)\,,$$

其中， $I(\sigma)$ 表示可通过数值方法计算的费希尔信息。