arXiv:2305.01839v1 [stat.ME] 3 May 2023

多元对称性：通过最优传输进行分布无关检验

Zhen Huang

哥伦比亚大学统计系，电子邮箱：
zh2395@columbia.edu和

Bodhisattva Sen*

哥伦比亚大学统计系，电子邮箱：
bodhi@stat.columbia.edu

2023年5月4日

摘要

符号检验（Arbuthnott，1710年）和 Wilcoxon 符号秩检验（Wilcoxon，1945）是最早的非参数检验范例之一。这些基于符号、（绝对）秩和带符号秩的检验方法，可用于一维对称性的分布无关检验。本文基于最优传输理论，提出一个用于多元对称性（包括中心对称性、符号对称性、球对称性等）分布无关检验的全新统一框架。我们的方法引出了分布无关的广义多元符号、秩和带符号秩的概念。因此，我们构建了符号检验和 Wilcoxon 符号秩检验的对应方法，它们保留了许多一维检验的优良性质。具体而言，所提检验在有限样本下均为精确的分布无关检验，并具有渐近正态极限，且能适用于多种多元对称性概念。我们研究了所提检验的一致性及其在局部备择下的表现，并证明所提广义Wilcoxon符号秩（GWSR）检验对位置平移备择尤其具有较高的检验功效。我们还证明，在一大类此类模型中，与 Hotelling 的 T^2 检验相比，我们的 GWSR 检验虽为非参数检验且具有精确的分布无关性，却不会损失（渐近）效率。对 GWSR 统计量进行适当的得分变换，可得到局部渐近最优检验。此外，我们的方法还可直接用于构造对称中心的分布无关置信集。

1 引言

对称性无处不在，从希腊雕塑中的双侧对称，到现代群论对自然界晶体结构的描述 [103, 93]。统计学中的对称性研究同样历史悠久、内容丰富。非参数检验最早的发表记录或许可追溯到 John Arbuthnott 的里程碑式论文 [4]：他实质上对单变量对称性进行了统计假设检验，并通过符号检验计算了一个 p值；例如，参见 [7, 41]。一些统计学家也将（经典）非参数统计学，尤其是基于秩的

*由 NSF 资助，项目编号为 DMS-2015376。

方法的起源，追溯至 Wilcoxon 1945 年论文的发表 [104], ，该论文提出了用于检验单变量对称性的 *Wilcoxon* 符号秩检验（WSR）。这两种经典检验都是分布无关的，并对非参数统计学的发展产生了巨大影响。

若单变量随机变量 X 满足对称（关于 0），则 X 与$-X$具有相同的分布。给定来自 X 分布的 n 个观测值 X_1、……、X_n，当 $|\sum_{i=1}^n \text{sign}(X_i)|$ 较大时，符号检验拒绝原假设 H_0 ：$X \overset{d}{=} -X$ ；这里，若 $X_i \geq 0$ ，则 $\text{sign}(X_i) = 1$ ，否则为-1。相比之下，Wilcoxon 符号秩检验首先对 $|X_1|$、……、$|X_n|$ 进行排序，得到其（绝对）秩 $\{\text{rank}(|X_i|)\}_{i=1}^n$，并在 $|\sum_{i=1}^n \text{sign}(X_i)\text{rank}(|X_i|)|$取较大值时拒绝 H_0。当X 服从连续分布时，经典的符号检验和带符号秩检验具有若干使其在实践中强大且实用的性质，包括：(i) 分布无关性，(ii) 符号与绝对秩的独立性（在 H_0 下），(iii) 简单的渐近正态极限分布，以及 (iv) 相较于 t 检验 [57, 22]具有较高的（相对）效率；关于这些结果的详细讨论，参见 [99, 第 15] 章。

自然会有人问，是否存在满足上述理想性质的这些检验的多维推广。在本文中，我们通过提出一个利用最优传输理论 [100, 101]来检验多元对称性的新框架，对这一问题给出了肯定的回答。

一维情形中的对称性是明确的（即，$X \overset{d}{=} -X$），而在多元情形下则存在多种对称性概念。其中最突出的是：(a) 中心对称性 [91,13, 80, 83, 32], (b) 符号对称性 [31, 84, 83], 以及 (c) 球对称性 [6, 39, 52, 2]；Serfling [93]对这一主题进行了全面综述。

在本文中，我们考虑以下用于多元对称性检验的一般设定。令$\mathcal{G}$ 为正交群 $\text{O}(p)$ 的一个紧子群（对于 $p \geq 1$）——即所有 $p \times p$ 实正交矩阵构成的群，其中群运算为通常的矩阵乘法。对于随机向量$\mathbf{X} \sim \mu \in \mathcal{P}_{\text{ac}}(\mathbb{R}^p)$，其中 $\mathcal{P}_{\text{ac}}(\mathbb{R}^p)$表示在 $\mathbb{R}^p$上具有勒贝格密度的所有博雷尔概率测度构成的类，我们感兴趣的是检验 $\mathcal{G}$-对称性这一假设：

$$\text{H}_0 : \mathbf{X} \overset{d}{=} Q\mathbf{X} \quad \text{for all} \quad Q \in \mathcal{G} \qquad \text{versus} \qquad \text{H}_1 : \text{not } \text{H}_0. \tag{1.1}$$

特别地，前文提到的对称性的典型例子对应于以下群 $\mathcal{G}$。

(a) 中心对称性：这里 $\mathcal{G} = \{I, -I\}$，其中 $I \equiv I_p$ 是 p阶的单位矩阵。中心对称性（或“反射”对称性、“对角”对称性、“简单”对称性或“对跖”对称性 [91, 13,80, 83, 32]）自然来源于配对数据（例如参见 [40, 第 5]节）：给定一个由 n对 $\{(\text{Y}_i, \text{Z}_i)\}_{i=1}^n$组成的样本，研究差值 $\text{X}_i := \text{Y}_i - \text{Z}_i$, $i = 1$, ……, n通常很有意义。在 Y_i 和 Z_i 可交换这一原假设下，X_i 将服从中心对称分布 [67, 40]。中心对称性的扩展包括角对称性 [86], ，它对应于 $\text{X}/\|\text{X}\| \overset{d}{=} -\text{X}/\|\text{X}\|$；此处 $\|\cdot\|$ 表示通常的欧几里得范数。 —— ——

(b) 符号对称性：这里 $\mathcal{G} = \{Q \in \mathbb{R}^{p\times p} : Q$ 是一个对角正交矩阵$\}$；因此 Q 的每个对角元素均为 ± 1。

符号对称性也称为正交象限对称性或边际对称性 [31, 84, 83]。这种符号变化不变性常用于位置-散布模型的假设中，该模型为稳健估计常规均值向量和协方差矩阵提供了工具 [83]；另见 Puri 和 Sen 的著作 [85] ，其中全面介绍了基于边际符号和秩的多元分析方法。

（c）球对称性：此处 $\mathcal{G} = \text{O}(p)$。
球对称分布是高斯分布的自然推广，并受到广泛关注 [33, 2]。另据了解，当偏离正态性假设的方向属于球对称族时，许多估计结果具有稳健性，在一般线性模型 [62, 10, 36], 充分降维 [69, 105],等领域均有应用。文献 [6, 39, 52, 2] 中研究了球对称性检验。

显然，O(p的其他子群，例如符号置换群[1] [35], ，也对应着一种对称性概念。H $\mathcal{G}$-对称模型由 H_0 在（1.1）中描述，其灵感来自 Beran 和 Millar 的工作 [9], ，并为以统一方式检验多种多元对称性概念提供了一个通用框架。

为构建针对 $\mathcal{G}$对称性一般假设的分布无关检验（1.1），我们首先指出（见第2节），一维符号、秩和带符号秩——它们是一元情形下分布无关推断的基石——可以作为一个优化问题的解，而该问题可推广至多维。由此得到的优化问题是一个最优传输（OT）问题 [100, 101] ，具有特定的代价函数，可以高效求解（见第2.3节）。

具体而言，给定来自 $\mu \in \mathcal{P}_{\mathrm{ac}}(\mathbb{R}^p)$ 的观测值 $\mathbf{X}1, \ldots, \mathbf{X}_n$ 以及（已知的）候选秩向量 $\mathbf{h}1, \ldots, \mathbf{h}_n \in \mathbb{R}^p$，广义符号 $S_n(\mathbf{X}_i)$ 被定义为 $\mathcal{G}$ 的一个元素，而广义秩的集合 $(R_n(\mathbf{X}_1) \ldots, R_n(\mathbf{X}_n))$ 是 $(\mathbf{h}1, \ldots, \mathbf{h}_n)$ 的一个排列；形式化定义见第2节第2.2小节。我们的方法导出了（$\mathcal{G}$-对称特定的）广义多元符号、广义秩和带符号秩的概念，这些概念同样具有分布无关性（在备择假设 H_0 下）。在一维情形下，如果 $(\mathbf{h}1, \ldots, \mathbf{h}_n)$ 选择为 $(\frac{1}{n}, \ldots, \frac{n}{n})$，那么我们便恢复了通常的符号和秩的概念。此外，与一维情形相对应，我们将广义符号检验统计量定义为 $\sum_{i=1}^n S_n(\mathbf{X}_i)$，而广义Wilcoxon符号秩（GWSR）检验的统计量则定义为 $\sum_{i=1}^n S_n(\mathbf{X}_i)R_n(\mathbf{X}_i)$；详情见第3节。

过去几年中，通过最优传输理论定义的多元秩 [43, 27], 已被有效地用于设计两样本问题中的分布无关假设检验程序 [15, 27], 用于独立性检验 [27, 95, 96, 94], 用于多元线性模型 [44, 45], 用于方向数据 [46],等。尽管我们的框架受到这些最新进展的启发，但它具有创新性和更强的通用性，并将通常意义下基于最优传输的多元秩 [43, 27] 作为一种特殊情形纳入其中。我们通过在最优传输问题中考虑一种更一般的代价函数来实现这一点，该代价函数自然源于假设检验问题（1.1），并且当 $\mathcal{G}$ 为平凡群时，退化为通常的平方欧几里得距离代价 $\{I\}$。此外，在第2.4节中，我们利用最优传输理论，对经验广义符号、秩和带符号秩的总体对应形式给出了几何刻画（参见定理2.1）。此外，我们在定理2.2中证明了这些样本估计量与其总体对应形式之间的一致性。我们的处理方法无需对数据分布作出任何矩条件假设 μ，因为我们采用了最优传输的几何刻画（由 Robert McCann 开创 [73]）。

经典 Wilcoxon 符号秩检验广受欢迎的重要原因之一，是 Hodges和Lehmann [57] 以及 Chernoff和Savage [22], 所取得的显著成果：作者证明，在位置平移备择下，Wilcoxon 检验相对于单样本 Student t-检验的渐近（Pitman）相对效率分别在采用恒等评分函数和高斯评分函数时，始终不低于 0.864 和 1，尽管前者对于所有样本量而言都是精确的分布无关检验。受这些结果启发，我们在第 3.4 节研究了按照 Pitman 的定义 [例如，参见 99，第 14 节。3], 所提出的 Wilcoxon 符号秩检验多元推广相对于 Hotelling T^2检验（通常 t-检验的多元推广）的（渐近相对）效率。特别地，我们证明，尽管所提出的 GWSR 检验完全不依赖于底层数据生成机制，但在多元分布的若干自然子族上，它仍满足 Hodges-Lehmann型和Chernoff-Savage型效率下界；参见定理 3.5-3.7。

现在简要概述本文的主要结果。

1. 分布无关性：在 H_0, 下

   (a) $S_n(\mathbf{X}_1) \ldots, S_n(\mathbf{X}_n)$ 独立同分布，且服从 $\mathcal{G}$ 上的均匀分布。(b) 秩 $(R_n(\mathbf{X}_1) \ldots, R_n(\mathbf{X}_n))$ 在所有 $n!$ 个 $(\mathbf{h}_1, \ldots, \mathbf{h}_n)$ 的排列中均匀分布；见命题 2.2。

2. 独立性：$S_n(\mathbf{X}_1) \ldots, S_n(\mathbf{X}_n))$ 在备择假设 H_0下相互独立（命题 2.2）。

[1]符号置换群由所有满足以下条件的矩阵构成：每行和每列恰有一个非零元素，且非零元素为 1 或 -1。

3. 渐近正态性：在备择假设 H_0 下，$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_n(X_i)$ 和 $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_n(X_i)R_n(X_i)$ 均依分布收敛于多元正态分布（定理 3.1）。4. 针对多元分布适当子类（例如乘积分布、椭圆对称分布等）的位置平移备择的相对效率：(a)（Hodges-Lehmann 现象 [57]）如果 $h_1, \ldots, h_n$ 和 $\mathcal{G}$ 确定一个均匀有效参考分布，那么 GWSR 检验相对于 Hotelling 的 T^2 检验的相对效率不会低于 0.864（精确表述见定理 3.5）。(b)（切尔诺夫-萨维奇现象 [22]）如果 $h_1, \ldots, h_n$ 和 $\mathcal{G}$ 确定一个高斯有效参考分布，那么 GWSR 检验相对于 Hotelling 的 T^2 检验的相对效率下界为 1（见定理 3.5–3.7）。5. 效率（针对位置平移备择）：得分变换 GWSR 检验族（其统计量为 $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_n(X_i)J(R_n(X_i))$，其中得分函数 $J: \mathbb{R}^p \to \mathbb{R}^p$）包含一个局部渐近最优检验（见定理 3.9），即在所有具有固定第一类错误率的检验中，该检验具有最大的（渐近）局部功效。这与一元情形中的类似结果相对应，参见 Van der Vaart [99, 推论 15.10]。

如果对于所有 $Q \in \mathcal{G}$，都有 $(X-\boldsymbol{\theta}) \overset{d}{=} Q(X-\boldsymbol{\theta})$，则称 X 关于 $\boldsymbol{\theta} \in \mathbb{R}^p$（通常称为 $\mathcal{G}$-对称性的中心）是 $\mathcal{G}$-对称的。我们的框架可以立即扩展，用于检验关于任意固定且已知的 $\boldsymbol{\theta}$ 的 $\mathcal{G}$-对称性假设。要检验关于一般中心 $\boldsymbol{\theta} \in \mathbb{R}^p$ 的对称性，只需将数据 $\{X_i\}_{i=1}^n$ 平移至 $\{X_i - \boldsymbol{\theta}\}_{i=1}^n$，即可将问题化归为 (1.1)。此外，如果 $\boldsymbol{\theta}$ 未知，我们提出的检验框架可用于构造关于 $\boldsymbol{\theta}$ 的分布无关置信集（见第 5 节）。另外，在第 4 节中，我们通过模拟研究展示了所提检验的优越性能（尤其是对于位置偏移模型），阐释了效率结果对样本量较小至中等的数据集的意义，并将我们提出的方法与现有的、针对不同对称性概念的检验方法进行了比较。

基于上述性质，我们认为，我们的广义多元符号与秩（通过最优传输定义）是经典概念的自然推广。据我们所知，我们的工作首次提出了多元分布无关符号与秩的概念（相对于任意群 $\mathcal{G}$），并在多维情形下为 (1.1) 构建了分布无关检验，其效率下界如上所述，与经典的 Wilcoxon 符号秩检验类似。

1.1 文献综述

多元对称性是非参数统计中的一个重要概念 [106]。关于多元对称性的假设检验已在文献中得到广泛研究，参见例如 Baringhaus [6], 、Diks 和 Tong [28]以及 Serfling [93]。

多维符号的扩展。 符号检验可能是最早且最流行的对称性检验之一 [4]。此后，文献中提出了若干多元符号的概念。作为一种流行的推广，分量式多元符号可追溯至 Bennett [8], 和 Chatterjee [20]。Bickel 推导了分量式符号检验针对 Hotelling's T^2 的渐近正态性和 Pitman效率 [12]。分量式秩和符号秩也被用于假设检验 [85, 59]。分量式符号的一个缺点是，其分布无关性要求 $\mathbf{X}$ 满足符号对称性（在中心对称性下，它并非分布无关）。我们将证明，当 2.4 中的广义符号所对应的群是 $\mathcal{G}$ 符号对称性对应的群时（见注释 2.4），它就化为分量式符号。

基于交叉方向的多元推广由 Randles 提出[86], ，推广了双侧单变量符号检验和 Blumen双变量符号检验 [14]。该检验统计量在非奇异线性变换下保持不变，并且对于具有椭圆方向的分布，它具有有限样本分布无关性 [86]。Chaudhuri 和 Sengupta [21]将 Randles 的统计量表示为数据依赖符号，并进一步统一了若干先前的仿射不变广义符号检验 [56, 102, 1]。随后，人们研究了这种变换—逆变换技术，以定义仿射等变多元中位数和分位数 [18, 19, 17, 92],以及仿射不变多元符号检验 [87] Hallin 和 Paindaveine 基于交叉方向推导出了局部渐近最大最小检验 [47] ，其前提是椭圆对称性。

M¨ott¨onen 和 Oja [75] 将多元符号定义为给定点方向上的单位向量，该向量称为空间符号。相应检验的 Pitman效率以及 Hodges-Lehmann估计量的一种推广，随后由 M¨ott¨onen 等人进行了分析 [77, 76]。然而，除球对称性外，上述各种符号概念都不是分布无关的。

Hallin 等人 [43] 提出了一种基于测度传输的新颖中心向外秩与符号定义，具有吸引人的性质，与我们基于 $\mathcal{G} = \{I\}$ 的框架类似。然而，他们的中心向外符号与经典符号及我们的方案在多个方面有所不同；例如，对于一维对称分布，中心向外符号在 $\{\pm 1\}$ 上并非独立同分布的均匀分布。

中心对称性检验。除了符号检验之外，还有一些中心对称性检验能够针对所有备择情形保持一致性。在一元情形下，McWilliams [74] 提出了一种基于游程统计量的简单检验。Dyckerhoff 等人 [29] 将 McWilliams 的方法扩展到利用数据深度检验双变量中心对称性。Einmahl 和 Gan [32], 提出了二元概率分布中心对称性的综合检验，这些检验在渐近意义下是分布无关的。另一类方法探究中心对称分布的特征函数性质，并利用经验特征函数构造检验统计量，既包括一元情形下的方法 [24, 34] ，也包括多维情形下的方法[39, 51, 53, 80, 54]。Henze 等人 [53]利用核密度估计量提出了中心对称性的仿射不变检验；另请参见 [54]。然而，这些检验通常并非渐近分布无关，因此推断时需要根据数据估计极限分布[80, 54]。

球对称性检验。 Marden 在 [71, 第 2 节4], 给出了 McWilliams 多维球对称性检验方法的直接推广，该方法在原假设下具有简单的渐近 χ^2_1 分布。Baringhaus [6], 提出了基于 von Mises 型检验统计量的旋转不变检验，由此得到的检验对所有备择情形均具有一致性。特征函数也被用于多维情形下的球对称性检验 [52, 39]。Kariya 和 Eaton [64] 考察了两类用于检验球对称性的备择情形，并给出了一致最强检验。Albisetti 等人 [2], 提出了一种新的假设检验方法，该方法基于如下事实：当且仅当对于任意两个垂直向量 $\mathbf{u}$ 和 $\mathbf{v}$ ，$\mathbf{X}$ 满足 $\mathbb{E}[\mathbf{u}^\top\mathbf{X}|\mathbf{v}^\top\mathbf{X}] = 0$时，$\mathbf{X}$ 才具有球对称性；其拒绝域通过自助法确定。

符号对称性检验。经典 Hotelling's T^2 检验的分析也已扩展至符号对称分布，相关研究包括 Efron [31], 、Pinelis [84], 等。在 Oja [83], 中，系统介绍了基于空间符号和秩的检验；针对观测具有符号对称性的情形，提出了空间符号检验和空间符号秩检验的符号变化版本。这些检验在多元 t-分布情形下，相对于 Hotelling's T^2 也具有良好的效率性质 [83, 第 8] 章。然而，这些方法尚无 Hodges-Lehmann [57] 和 Chernoff-Savage [22] 类型的下界。

1.2 论文结构

第2节定义多元符号、秩和带符号秩，并给出它们的总体对应量。第3节定义符号检验和 Wilcoxon 符号秩检验的多元对应检验。我们进一步研究这些所提出检验的统计性质，例如分布无关性、渐近正态性、一致性和效率。所提出检验的有限样本表现以及与竞争方法的比较见第4节。第 $\mathcal{G}$-对称分布的未知对称中心的一个分布无关置信集将在第5节中构造。我们结果的所有证明及进一步讨论均置于附录A-C中。

2 多元广义符号、秩和符号秩

2.1 预备知识

设 O(p) 表示 p 维空间中的正交群，即在矩阵乘法这一群运算下由所有 $\mathbb{R}^{p\times p}$ 正交矩阵构成的群。设 $\mathcal{G}$ 为正交群 O(p) 的一个紧子群。特别地， $\mathcal{G}$ 满足以下群公理：(i) $I \in \mathcal{G}$；其中 $I \equiv I_p$ 是

阶数为 p 的单位矩阵；(ii) 如果 $Q \in \mathcal{G}$，则 $Q^{-1} \in \mathcal{G}$；(iii) 如果 Q_1， $Q_2 \in \mathcal{G}$，则 $Q_1 Q_2 \in \mathcal{G}$；(iv) 如果 Q_1， Q_2， $Q_3 \in \mathcal{G}$，则 $(Q_1 Q_2) Q_3 = Q_1 (Q_2 Q_3)$。紧致性假设保证在 $\mathcal{G}$上存在唯一的均匀分布（也称为哈尔测度）；例如参见 Eaton [30]。

定义 2.1 (轨道)。 元素$\mathbf{x} \in \mathbb{R}^p$ 的轨道是指 $\mathbb{R}^p$中可由 $\mathcal{G}$的元素将$\mathbf{x}$变换到的元素集合，即 $\{Q\mathbf{x} : Q \in \mathcal{G}\}$。

请注意，如果$\mathbf{y}$ 属于$\mathbf{x}$的轨道，那么$\mathbf{x}$ 也属于$\mathbf{y}$的轨道。因此， $\mathbb{R}^p$ 可划分为若干轨道的不相交并。令 $B \subset \mathbb{R}^p$。我们定义集合

$$\mathcal{G}B := \{Q\mathbf{x} : Q \in \mathcal{G}, \mathbf{x} \in B\} \subset \mathbb{R}^p \tag{2.1}$$

作为 $\mathcal{G}$ 在 B 上的群作用的像。

定义 2.2 （群作用）。 $\mathcal{G}$ 在 $\mathcal{G}B$ 上的（左）群作用是一个函数 $\cdot : \mathcal{G} \times \mathcal{G}B \to \mathcal{G}B$，满足以下两条公理：*(i)* $I \cdot \mathbf{x} = \mathbf{x}$，以及 *(ii)* $Q_1 \cdot (Q_2 \cdot \mathbf{x}) = (Q_1 Q_2) \cdot \mathbf{x}$，对所有$\mathbf{x} \in \mathcal{G}B$以及 Q_1、$Q_2 \in \mathcal{G}$均成立。在我们的设定中，我们将群作用 $\cdot$ 取为通常的矩阵乘法运算。

定义2.3 (自由群作用)。 我们称 $\mathcal{G}$ 在 $\mathcal{G}B$ 上自由作用，如果对于$\mathbf{y} \in B$ 和 $Q \in \mathcal{G}$，$Q\mathbf{y} = \mathbf{y}$ 意味着 $Q = I$ （即，对于 $\mathcal{G}B$中的任意$\mathbf{x}$，只有 $\mathcal{G}$ 中的恒等映射才能使$\mathbf{x}$ 保持不变）。

直观而言，自由群作用有助于识别出将一个点映射到另一个点的唯一元素（即，对于 Q_1, $Q_2 \in \mathcal{G}$，若 $Q_1\mathbf{x} = Q_2\mathbf{x}$，则 $Q_1 = Q_2$）。令 $\mathcal{P}_{\mathrm{ac}}(\mathbb{R}^p)$ 表示所有具有勒贝格密度的 $\mathbb{R}^p$ 上的博雷尔概率测度所构成的类。若概率测度 μ 集中在集合 B 上，则称其集中于该集合，即 $\mu(B) = 1$。若 μ 是博雷尔测度，则 μ 的支撑是其所集中于的最小闭集。

定义 2.4 （推前分布）。 设 F： $\mathbb{R}^p \to \mathbb{R}^k$ （对于 $k \geq 1$）是一个函数，且 μ 是 $\mathbb{R}^p$ 上的一个分布。推前分布 $F\#\mu$ 定义为 $F(\mathbf{X})$ 的分布，其中$\mathbf{X} \sim \mu$。

定义 2.5 （循环单调性）。 设 c： $\mathbb{R}^p \times \mathbb{R}^p \to (-\infty, +\infty]$ 为一个函数。若子集$\Gamma \subset \mathbb{R}^p \times \mathbb{R}^p$满足：对于任意 $N \in \mathbb{N}$以及 Γ中的任意一组点$(\mathbf{x}_1, \mathbf{y}_1) \ldots, (\mathbf{x}_N, \mathbf{y}_N)$，以下不等式均成立，则称其为 c-循环单调的：

$$\sum_{i=1}^{N} c(\mathbf{x}_i, \mathbf{y}_i) \leq \sum_{i=1}^{N} c(\mathbf{x}_i, \mathbf{y}_{i+1}),$$

约定 $\mathbf{y}_{N+1} = \mathbf{y}_1$。

2.2 与最优传输的联系

对于随机向量 $\mathbf{X} \sim \mu \in \mathcal{P}_{\mathrm{ac}}(\mathbb{R}^p)$以及正交群 $\mathrm{O}(p)$ 的紧子群 $\mathcal{G}$，我们关注对 $\mathcal{G}$-对称性（1.1）假设的分布无关检验。在本小节中，我们将定义新的分布无关（有限样本）符号和秩的多元推广，它们将构成分布无关推断的基石。此外，这些符号和秩的推广——当 $p = 1$ 且 $\mathcal{G} = \{I, -I\}$时，它们会退化为通常的经典符号和秩概念——能够适应不同的多元对称性概念，并且具有许多与其经典对应检验相似的性质。我们借鉴最优传输（OT）理论中的思想来定义多元推广。尽管我们的方法受到通过 OT 定义多元秩的概念的启发 [43, 27]，我们的框架是新颖且更一般的。

令 ν 为定义在 $\mathbb{R}^p$ 上的某个已知且固定的博雷尔测度，称为参考分布；具体示例见注释3.3 。对于每个 n，我们考虑一个准蒙特卡洛[48, 66, 81] 或蒙特卡罗 [50, 88] 序列 $\{\mathbf{h}_{n,j} \in \mathbb{R}^p : j = 1, 2, \ldots, n\}$，并假设该序列已知且固定，使得由 $\{\mathbf{h}_{n,j}\}_{j=1}^{n}$ 得到的经验分布（记为 ν_n）在 $n \to \infty$ 时依分布收敛到 ν。因此， ν_n 提供了对 ν 的离散化。为简化记号，我们省略 n 并写作

$\mathbf{h}_j \equiv \mathbf{h}_{n,j}$。例如，$\{\mathbf{h}_j\}_{j=1}^n$ 可以取自 ν 的独立同分布观测 n，然后固定下来。如果$\nu = \mathrm{Uniform}(0,1)^p$，我们也可以令 $\{\mathbf{h}_j\}_{j=1}^n$ 为 Halton 序列 [48] 中的前 n 个点；还有其他选择（例如参见 Deb 和 Sen [27, 附录 D.3]）。正如我们稍后将看到的，要恢复经典一维符号和带符号秩，可以取 $\nu = \mathrm{Uniform}(0,1)$ 和 $\mathbf{h}_j = \frac{j}{n}$，其中 $j = 1、2、\ldots、n$。

令 $\mathcal{S}_n$ 为 $\{1,\ldots,n\}$ 的所有 $n!$ 个排列构成的集合。考虑以下优化问题。 on 问题：

$$\min\left\{\sum_{i=1}^n \|Q_i^\top \mathbf{X}_{\sigma(i)} - \mathbf{h}_i\|^2 : Q_i \in \mathcal{G}, \sigma \in \mathcal{S}_n\right\}. \tag{2.2}$$

令 $(\{Q_i\}_{i=1}^n, \hat{\sigma})$ 为 (2.2) 的极小化解。事实上，在温和的假设下，$\hat{\sigma}$ 实际上几乎必然唯一，且 $\{Q_i\}_{i=1}^n$ 在 $\mathcal{G}$ 诱导自由群作用时也几乎必然唯一（见命题 2.1）。定义代价函数

$$c(\mathbf{x},\mathbf{h}) := \min_{Q\in\mathcal{G}} \|Q^\top \mathbf{x} - \mathbf{h}\|^2, \qquad \text{for } \mathbf{x},\mathbf{h} \in \mathbb{R}^p, \tag{2.3}$$

使得 (2.2) 化简为 $\min_{\sigma\in\mathcal{S}_n} \sum_{i=1}^n c(\mathbf{X}_{\sigma(i)},\mathbf{h}_i)$——即在代价函数 $c(\cdot,\cdot)$ 下，从$\{\mathbf{X}_1,\ldots,\mathbf{X}_n\}$到 $\{\mathbf{h}_1,\ldots,\mathbf{h}_n\}$的最优传输（OT）问题。注意，代价函数满足$c(\mathbf{x},\mathbf{h}) = c(\mathbf{x}\ Q\mathbf{h}) = c(Q\mathbf{x},\mathbf{h})$，对于任意 $Q \in \mathcal{G}$ 均成立，即它在 $\mathcal{G}$ 中的变换下保持不变。

定义广义符号 S_n： $\{\mathbf{X}_1,\ldots,\mathbf{X}_n\} \to \mathcal{G}$ 以及广义秩 R_n： $\{\mathbf{X}_1,\ldots,\mathbf{X}_n\} \to \{\mathbf{h}_1,\ldots,\mathbf{h}_n\}$ 如下：

$$S_n(\mathbf{X}_i) = \widehat{Q}_{\hat{\sigma}^{-1}(i)}, \qquad \text{and} \qquad R_n(\mathbf{X}_i) = \mathbf{h}_{\hat{\sigma}^{-1}(i)}, \qquad \text{for } i = 1,\ldots,n. \tag{2.4}$$

第 i 个观测$\mathbf{X}_i$ 的广义符号秩定义为 $S_n(\mathbf{X}_i)R_n(\mathbf{X}_i)$。

注 *2.1*（当 $p = 1$ 时）。注意，当 $p = 1$ 且 $\mathcal{G} = \{1,-1\}$ 时，广义符号和秩（见 *(2.4)*）退化为经典的一维概念。实际上，在这种情况下，问题 *(2.2)* 直接给出通常的符号和秩。为说明这一点，注意对于 $x \in \mathbb{R}$ 和 $h \in [0,1]$，由 *(2.3)* 可得，$c(x,h) = \min\{(x-h)^2,(-x-h)^2\}$，由于实数轴上的线性排序，这可化简为 $c(x,h) = (|x|-h)^2$。现在很容易看出，对于所有 j，令 $h_j = \frac{j}{n}$ 后，*(2.2)* 可简化为

$$\min_{\sigma\in\mathcal{S}_n} \sum_{i=1}^n \left(|X_i| - \frac{\sigma(j)}{n}\right)^2 = \max_{\sigma\in\mathcal{S}_n} \sum_{i=1}^n |X_i| \frac{\sigma(j)}{n},$$

根据重排不等式（例如，参见 *[49, 定理 368]*），可得 $\hat{\sigma}(i)$等于 $|X_i|$的通常（绝对）秩，其中 $i = 1,\ldots,n$。因此，X_i 中绝对值第 j大的元素将映射为 j/n，且 X_i 的符号在 $X_i \geq 0$ 时为 *1*，在 -1 时为 $X_i < 0$。

当 $p > 1$ 时，由于 $\mathbb{R}^p$ 中缺乏典范排序，优化问题 (2.2) 没有显式解。然而，其方法是类似的；优化问题 (2.2) 以一一对应的方式将数据点 $\{\mathbf{X}_i\}_{i=1}^n$ 分配给典范多元秩向量 $\{\mathbf{h}_j\}_{j=1}^n$，使得 (2.2) 中的目标函数最小化。为了帮助理解广义符号和带符号秩（如 (2.4) 中所示），或许将代价函数表示为$c(\mathbf{x},\mathbf{h}) = \min_{Q\in\mathcal{G}} \|\mathbf{x} - Q\mathbf{h}\|^2$会更容易。因此，如果 $R_n(\mathbf{X}_i) = \mathbf{h}_j$，那么 $c(\mathbf{X}_i,\mathbf{h}_j) = \min_{Q\in\mathcal{G}} \|\mathbf{X}_i - Q\mathbf{h}_j\|^2$，因此，带符号秩 $S_n(\mathbf{X}_i)R_n(\mathbf{X}_i) \equiv S_n(\mathbf{X}_i)\mathbf{h}_j$ 是在 $\mathbf{h}_j$ 的轨道中与 $\mathbf{X}_i$ 最接近的元素（按照欧几里得范数衡量）。此外，$S_n(\mathbf{X}_i)$，即 $\mathbf{X}_i$ 的符号，是代价 $\|\mathbf{X}_i - Q\mathbf{h}_j\|^2$ 在 $\mathcal{G}$ 上的 argmin——它是 $\mathcal{G}$ 中将秩向量 $R_n(\mathbf{X}_i) \equiv \mathbf{h}_j$ 进行变换，使其与 $\mathbf{X}_i$ 最接近的正交矩阵。

图 1 展示了 $p = 1$ 和 $p = 2$ 在 $\mathcal{G} = \{-I, I\}$时的情形。图 1 左图给出了一个示例，其中包含来自 $N(02, 4I_2)$ 的 $n = 7$ 个数据点 $\mathbf{X}1,\ldots,\mathbf{X}_n$（标记为“•”），以及当 $\mathcal{G} = \{I,-I\}$ 时（对应于中心对称性）的广义秩。秩向量 $\mathbf{h}1,\ldots,\mathbf{h}_n$（标记为“+”）取自分布 $(|Z_1|, Z_2)$ 的 n 个独立同分布抽样，其中 Z_1, Z_2 是来自 $N(0,1)$ 的独立同分布抽样。图 1 中每条实线的长度平方就是运输每个数据点的成本，即 $c(\mathbf{X}_i, R_n(\mathbf{X}_i)) \equiv \|\mathbf{X}_i - S_n(\mathbf{X}_i)R_n(\mathbf{X}_i)\|^2$，其中 $i = 1,\ldots,n$。对于某些数据点 $\mathbf{X}_i$，若 $\min_{Q\in\{-I,I\}} \|Q^\top \mathbf{X}_i -$，则会先将 $\mathbf{X}_i$ 运输至 $-\mathbf{X}_i$，而不产生任何成本

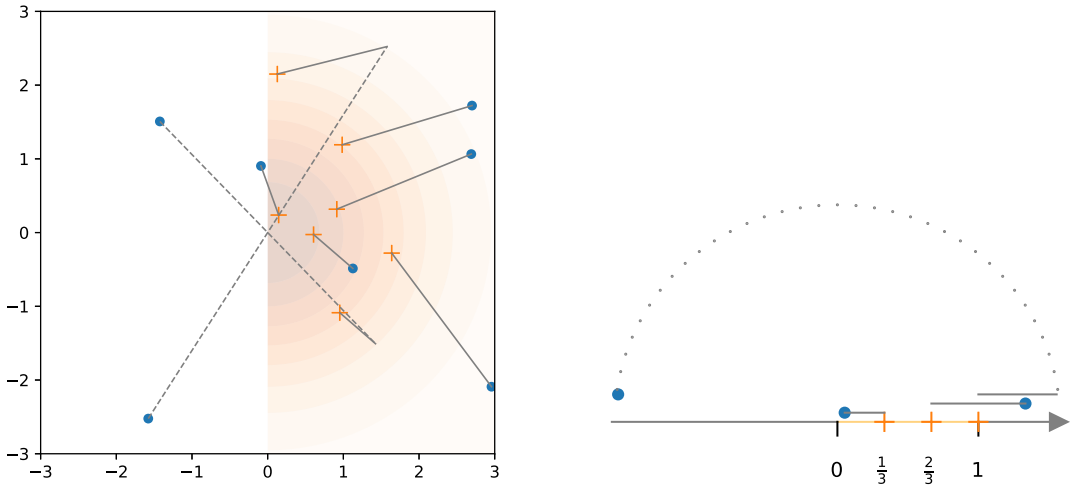


图 1：左：数据点 $\mathbf{X}1, \ldots, \mathbf{X}_n$（标记为“•”），其 $n = 7$ 取自 $N(02, 4I_2)$。秩向量 $\mathbf{h}1, \ldots, \mathbf{h}_n$（标记为“+”）是从 $(|Z_1|, Z_2)$ 的分布中进行 n 独立同分布抽样得到的，其中 Z_1, Z_2 独立同分布于 $N(0, 1)$。右（经典符号与秩）：$n = 3$ 数据点（标记为“•”）取自 $N(0, 4)$。秩（标记为“+”）为 $h_j = \frac{j}{n}$, $j = 1, \ldots, n$。每条实线的长度平方就是运输相应数据点的运输成本。虚线（点线）表示 $\mathbf{X}_i \mapsto -\mathbf{X}_i$（即 $S_n(\mathbf{X}_i) = -I$），之后再将其运输至 $\mathbf{h}_j$。

$R_n(\mathbf{X}_i)\|^2 = \| - \mathbf{X}_i - R_n(\mathbf{X}_i)\|^2$（如虚线所示）。在这种情况下，$S_n(\mathbf{X}_i)$（即 $\mathbf{X}_i$ 的符号）为 $-I$。如果直接将 $\mathbf{X}_i$ 映射到 $R_n(\mathbf{X}_i)$ 的平方距离，小于 $-\mathbf{X}_i$ 与 $R_n(\mathbf{X}_i)$ 之间的平方距离，那么 $S_n(\mathbf{X}_i) = I$。当 $p = 1$ 时（见图 1 右图），也可以观察到类似现象。区别仅在于，此时优化问题 (2.2) 可以显式求解（如注 2.1 所示）。

下面的命题（证明见附录 B.1）表明，在温和的假设下，广义秩和符号秩几乎处处唯一。尽管广义符号可能不唯一，但当 $\mathcal{G}$ 的群作用是自由的时，它们几乎处处唯一。

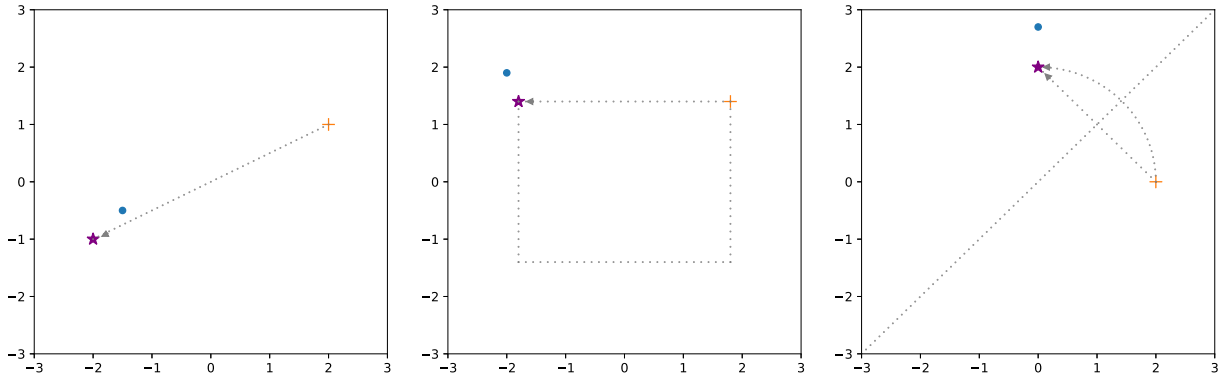
命题 2.1（样本秩、符号和符号秩的唯一性）。假设 $\mathbf{X} \sim \mu \in \mathcal{P}_{\text{ac}}(\mathbb{R}^p)$，并且假设任意两个$\mathbf{h}_j$都不在 $\mathcal{G}$ 的同一轨道上。

1. 则多元秩 $R_n(\cdot)$ 和带符号秩 $S_n(\cdot)R_n(\cdot)$ 几乎处处唯一。*2.* 若 $\mathbf{h}_1$、…、$\mathbf{h}_n \in B \subset \mathbb{R}^p$，且 $\mathcal{G}$ 在 $\mathcal{G}B$ 上自由作用（见(2.1)），则 $S_n(\cdot)$ 几乎处处唯一 *ue.*

通过例如取 $B = (0, \infty) \times \mathbb{R}^{p-1}$ 和 $B = (0, \infty)^p$，中心对称性和符号对称性都可以轻易获得自由群作用。因此，对于这些群，广义符号 $S_n(\cdot)$是唯一的；见图2a（中心对称性）和图2b（符号对称性）。

然而，对于球对称性，当 $p > 1$时，我们没有自由群作用：图 2c（球对称性）中，$\mathcal{G}$ =O(2)，且广义符号并不唯一——旋转 90 度和沿直线 $y = x$ 反射，都会将秩向量映射为带符号秩。然而，带符号秩是唯一的——它是 $\{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\| = 2\}$ 中距离数据点最近的点（即秩向量 $\mathbf{h}$，在图 + 中以“2c”表示，满足 $\|\mathbf{h}\| = 2$）。

注 *2.2*（当 *(2.2)* 的极小化解不唯一时）。如果 *(2.2)* 的极小化解不唯一，则按以下方式从所有可能的极小化解中随机抽取：如果 σ_1、…、$\sigma_k \in \mathcal{S}_n$ 均达到 $\sum_{i=1}^n c(X^{\sigma(i)}, h_i)$ 的最小代价，则令 $\hat{\sigma}$ 以概率 $\mathbb{P}(\hat{\sigma} = \sigma_j) = \frac{1}{k}$ 取值为 σ_j、$j = 1$、…、k。确定 $\hat{\sigma}$ 后，求极小化解 $\widehat{Q}_i := \operatorname{argmin}_{Q \in \mathcal{G}} \|Q_i^\top X^{\hat{\sigma}(i)} - h_i\|^2$。如果该极小化解不唯一，则引入另一个随机变量/向量（与其他所有变量/向量独立），从中选择 Q_i



(a) $\mathcal{G}$：中心对称性。　(b) $\mathcal{G}$：符号对称性。　(c) $\mathcal{G}$：球对称性。

图 2： 数据点（标记为“•”）、其秩向量（标记为“+”）以及其带符号秩（标记为“⋆”）的示意图。带符号秩是轨道中与数据点距离最近的秩向量的唯一点。对于 (c)，其中 $\mathcal{G}$ 对应球对称性，符号并不唯一，因为旋转 90 度和沿直线 $y = x$ 反射都会将秩向量映射为带符号秩。

在 $\mathcal{G}$ 条件为其使 $\|Q_i^\top \mathbf{X}^{\hat{\sigma}(i)} - \mathbf{h}_i\|^2$时的均匀分布；关于如何从该条件分布中抽样，参见第 3.3 节。

以下命题（证明见附录 B.2 ）表明，广义符号和带符号秩（若不唯一，可按备注2.2 所述通过随机化选取）自然地推广了经典对应检验的性质（当 $p = 1$时）。特别地，在 $\mathcal{G}$-对称性的原假设下，广义符号与广义秩相互独立，且任何基于它们的统计量在有限样本中都具有精确的分布无关性。

命题 2.2。 考虑式 (2.4)中定义的广义符号、秩和带符号秩。若它们不唯一（参见命题 2.1），则假设按照备注 2.2中所述通过随机化进行选择。则以下性质成立：

1. 广义秩（$R_n(X_1), \ldots, R_n(X_n)$）与$\{X1, \ldots, X_n\}$的任意次序统计量[2] 独立，并且在由$(h1, \ldots, h_n)$的全部 $n!$ 个排列构成的集合上服从均匀分布。*2.* 在 $H_0 : X \overset{d}{=} QX$，对所有 $Q \in \mathcal{G}$，$(R_n(X_1), \ldots, R_n(X_n))$ 与广义符号 $(S_n(X_1), \ldots, S_n(X_n))$ 相互独立。*3.* 在 H_0下，$S_n(X_1), \ldots, S_n(X_n)$ 独立同分布，且在 $\mathcal{G}$ 上服从均匀分布。

注 2.3 (无符号多元秩)。 若 $\mathcal{G} = \{I\}$，则 $S_n(\mathbf{X}_i) = I$ 和 $R_n(\mathbf{X}_i)$ 退化为 *Deb* 和 *Sen* 定义的（无符号）多元秩 *[27]*, ，并包含 *Hallin* 等人提出的中心向外秩 *[43]*。

注 2.4 (分量符号)。 对于符号对称性，当 ν_n 集中于 $(0, \infty)^p$ 时， $S_n(\mathbf{X}_i)$是一个对角矩阵，其对角元素恰好等于$\mathbf{X}_i$ *[8, 12,20, 85]*的分量符号。

2.3 计算复杂度

优化问题（2.2）可以高效求解。回顾一下， $c(\mathbf{x},\mathbf{h}) = \min_{Q\in\mathcal{G}} \|Q^\top \mathbf{x} - \mathbf{h}\|^2$是代价函数（见（2.3））， 当 O有限时，可在 $\mathcal{G}$ 时间内计算。随后，秩

[2]一个 *order statistic* 是数据 $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ [43]. 的无序版本

$R_n(\mathbf{X}_i)$ 可通过求解指派问题 [78, 11]，将 $\mathbf{X}1, \ldots, \mathbf{X}_n$ 指派给 $\mathbf{h}1, \ldots, \mathbf{h}_n$，并采用代价 $c(\cdot,\cdot)$；已有算法的最坏情况复杂度为 $O(n^3)$ [63]。得到秩后，可以获得 $S_n(\mathbf{X}_i) = \operatorname{argmin}_{Q\in\mathcal{G}} \|Q^\top \mathbf{X}_i - R_n(\mathbf{X}_i)\|^2$（关于如何为本文讨论的对称性的三个主要示例求得该极小化解，参见第 3.3 节和注 3.1 中的讨论）。

注释*2.5*（球对称性的快速计算）。当 $\mathcal{G} = O(p)$ 时，秩（以及带符号秩）的计算会更快（即时间复杂度为 $O(n \log n)$），因为 $c(x, h) = \min_{Q\in\mathcal{G}} \|Q^\top x - h\|^2 = \min_{Q\in\mathcal{G}} \{\|x\|^2 - 2x^\top Qh + \|h\|^2\} = (\|x\| - \|h\|)^2$，其中最后一步利用了这样一个事实：存在 $Q \in \mathcal{G}$ 将 h 映射到 x 的方向，例如 *Householder*反射 *[58]*。因此，若 X_i 在 $X1, \ldots, X_n$ 中欧几里得范数排名第 j，且 $\|h_1\| < \ldots < \|h_n\|$，那么 X_i 的秩将为 h_j。X_i 的带符号秩就是沿着 X_i 方向、长度为 $\|R_n(X_i)\|$ 的向量，即 $S_n(X_i) R_n(X_i) = \|R_n(X_i)\| X_i \| {}^i X_i \|$。 ——

2.4 总体对应形式

到目前为止，我们已通过优化问题（ S_n（·）和R_n（·）2.2）定义了多元广义符号和秩的样本版本。现在我们将证明，若 μ 和 ν满足有限二阶矩假设，则样本秩映射 R_n（·）收敛到总体秩映射 R（·），使得（$\mathbf{X}$ R（$\mathbf{X}$））解以下最优传输问题（亦称 Kantorovich 松弛 [101]）：

$$\inf_{(\mathbf{X},\mathbf{H}):(\mathbf{X},\mathbf{H})\in\Pi(\mu,\nu)} \mathbb{E}\left[c(\mathbf{X},\mathbf{H})\right], \quad \text{with } c(\mathbf{x},\mathbf{h}) = \min_{Q\in\mathcal{G}} \|Q^\top \mathbf{x} - \mathbf{h}\|^2. \tag{2.5}$$

这里，Π（μ，ν）由所有边缘分布为$\mathbf{X}$、$\mathbf{H}$的联合博雷尔分布（$\mathbf{X} \sim \mu$和$\mathbf{H} \sim \nu$）构成。可以看出，优化问题（2.5）是（2.2）的自然总体对应形式。

为证明 $S_n(\cdot)$ 和 $R_n(\cdot)$ 的收敛，我们需要参考分布 ν 和群 $\mathcal{G}$ 满足某些自然的相容性假设。具体而言，我们假设：

假设 1。存在一个博雷尔集 $B \subset \mathbb{R}^p$，使得 $\nu(B) = 1$，且对任意$\mathbf{x} \in \mathbb{R}^p$，$\mathbf{x}$ 的轨道：$\{Q\mathbf{x} : Q \in \mathcal{G}\}$ $\mathbf{x}$ 与 B 至多相交于一点。

该假设要求：对于 $\mathcal{G}$的任一轨道，至多只能取其中一个点作为秩向量（该向量属于 B）。这是因为，由于代价函数的形式$\mathbf{x}$ 传输到 $\mathcal{G}$轨道中的任意一点，代价都相同$(\cdot,\cdot)$（见（2.3）），因此无需纳入轨道中的两个代表点，否则可能破坏总体最优传输映射的唯一性。 [3]

为正式陈述我们的结果，我们先引入一些记号。

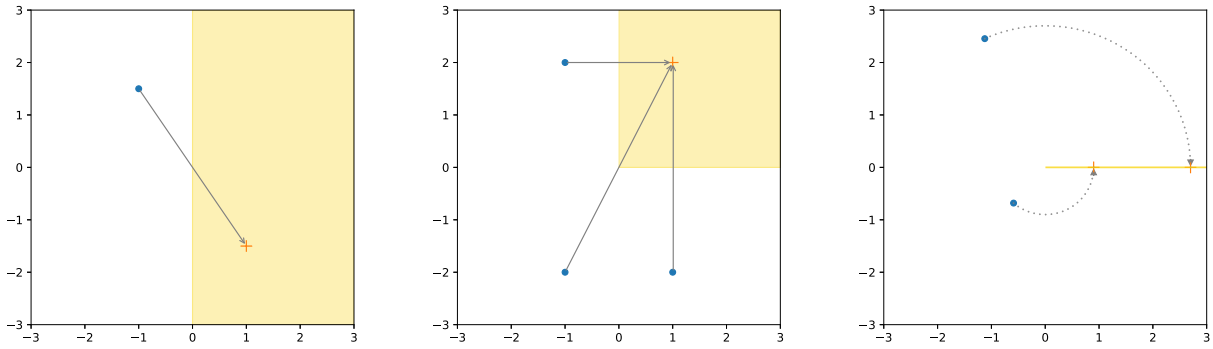
定义 2.6（商映射 q）。在假设 *1* 下，我们定义商映射 $q : \mathcal{G}B \to B$ 为$q(Q\mathbf{x}) = \mathbf{x}$，其中 $\mathbf{x} \in B$ 且 $Q \in \mathcal{G}$。注意，假设 *1* 保证了 q 定义良好。

请注意，上述定义与通常定义商映射的方式略有不同，例如，将一个点映射到一个等价类（即这里的轨道）的映射；参见 Munkres [79,第 2 章。22]相反，我们的商映射 q 将轨道 $\{Q\mathbf{x} : Q \in \mathcal{G}\}$ 中的所有点显式映射到$\mathbf{x}$ 中的一个代表点 B。

图 3 展示了商映射的示例。在图 3a 中，$B = (0,\infty) \times \mathbb{R}$ 和 $\mathcal{G} = \{-I, I\}$。这里，$q(\mathbf{x}) = -\mathbf{x}$ if $x_1 < 0$ and $q(\mathbf{x}) = \mathbf{x}$ if $x_1 > 0$，其中 $\mathbf{x} = (x_1, x_2)$。图 3b 展示了商映射 q 对三个点 $(-1, 2)$、$(-1, -2)$ 和 $(1, -2)$ 的作用，其中 $\mathcal{G}$ 是与符号对称性对应的群，且 $B = (0,\infty)^2$。这三个点都属于$\mathcal{G}$的同一轨道，并且被映射到同一个点 $(1, 2)$。在这种情况下，$q(\mathbf{x}) = (|x_1|, |x_2|)$。图 3c 展示了 $\mathcal{G} = O(2)$ 的商映射，对应于球对称性，且 $B = (0,\infty) \times \{0\}$。在这种情况下，商映射 q 可以写为 $q(\mathbf{x}) = (\|\mathbf{x}\|, 0)$。

以下结果（证明见附录B.4）指出，存在唯一的总体秩映射 R： $\mathbb{R}^p \to \mathbb{R}^p$ 使得（恒等映射 $R)\#\mu$ （即（$\mathbf{X}$ R的分布$\mathbf{X}$））其中

[3]设 $\mu = \nu = N(\mathbf{0}, I)$。那么 $R(\mathbf{x}) = \mathbf{x}$ 和 $R(\mathbf{x}) = -\mathbf{x}$ 都是损失为 0 的最优传输映射（见（2.5））， 前提是 $-I \in \mathcal{G}$。



(a) $\mathcal{G}$：中心对称性。 (b) $\mathcal{G}$：符号对称性。 (c) $\mathcal{G}$：球对称性。

图 3：不同群 $q(\cdot)$ 下商映射 $\mathcal{G}$（当 $p = 2$ 时）的示意图。此处，每个蓝色点（标记为“•”）都映射到其轨道中的一个代表点（标记为“+”），该代表点位于集合 B 中（黄色阴影区域）。

$\mathbf{X} \sim \mu$) 求解了（2.5），前提是最优成本有限。即使最优成本为无穷大，$R(\cdot)$ 仍可利用 c-循环单调性（参见定义 2.5）进行唯一刻画。

定理 2.1（总体秩和带符号秩映射）。假设 *1* 成立。记 $\mathbf{X}$ 的分布为 $\mu \in \mathcal{P}_{\mathrm{ac}}(\mathbb{R}^p)$。则存在一个 μ-几乎处处唯一的博雷尔可测映射$R: \mathbb{R}^p \to \mathbb{R}^p$，使得 $(\mathbf{X}\ R(\mathbf{X}))$ 在 $\Pi(\mu, \nu)$ 中具有唯一分布，且其支撑具有 c-循环单调性。此外，R 还具有以下性质：

(i) 令 q 为从 $\mathcal{G}B$ 到 B的商映射。则存在一个下半连续凸函数 $\psi: \mathbb{R}^p \to (-\infty, +\infty]$ 使得

$$R(\mathbf{x}) = q(\nabla\psi(\mathbf{x})) \quad (\mu\text{-}a.e.\ \mathbf{x}).$$

(ii) 令 $S \sim$ *Uniform(*$\mathcal{G}$*)* 与 $X \sim \mu$ 和 $H \sim \nu$ 独立。令 μ_S 和 ν_S 分别表示 SX 和 SH 的分布。那么 $\nabla\psi(\cdot)$ 是将 μ_S 推送至 ν_S的凸函数梯度，且几乎处处唯一。*(iii)* $\nabla\psi$ 在 $\mathcal{G}$的群作用下等变，即对于 $Q \in \mathcal{G}$，且 $x \in \mathbb{R}^p$为 ψ的可微点时，$\nabla\psi(Qx) = Q\nabla\psi(x)$。*(iv)* $\nabla\psi(X) \overset{a.s.}{=} S(X\ R(X))R(X)$，其中 $S(x, h) := \mathrm{argmin}_{Q\in\mathcal{G}} \|Q^\top x - h\|^2$。因此，$\nabla\psi(\cdot)$可视为总体符号秩映射。*(v)* 若 μ 和 ν 具有有限二阶矩，则 *(identity* $R)\#\mu$ 是总体最优传输问题 *(2.5)* 的唯一解。

我们来讨论上述结果中 (*i*)-(*v*) 的结论：(*i*) 表明，总体秩映射本质上是某个凸函数 $\psi(\cdot)$ 的梯度（参见 McCann 对最优传输映射的几何刻画 [73]，其中采用通常的平方损失），只需再与商映射 $q(\cdot)$ 复合即可将其映至 B。凸函数 $\psi(\cdot)$ 由 (*ii*) 刻画——其梯度将对称化数据分布 μ_S 推送至对称化参考测度 ν_S，且该函数唯一确定（这可由 McCann 的著名结果推出 [73]）。尽管 [73] 中的结果意味着$\nabla\psi(\cdot)$ 在 μ_S-几乎处处意义下唯一确定，但附录中的引理 B.3 表明，μ 相对于 μ_S绝对连续，因此 $\nabla\psi(\cdot)$ 在 μ-几乎处处意义下也唯一确定。在定理 2.1 的 (*iii*) 和 (*iv*) 中，我们展示了 $\nabla\psi(\cdot)$ 的一些“良好”性质：它保留了 $\mathcal{G}$ 的对称性，并且实际上是总体符号秩映射，这进一步促使我们研究广义（样本）符号秩 $S_n(\cdot)R_n(\cdot)$。请注意，上述结果中的 (*i*)-(*iv*) 不对 μ 和 ν作任何矩条件假设，并利用 c-循环单调性对秩映射作出几何刻画。

当 μ 和 ν 的二阶矩存在时，(v) 表明，总体秩映射确实是按最小化 (2.5) 意义下的最优传输映射。

注 2.6 （关于定理 2.1 的证明）。 上述结果的证明通过验证 c-次微分集合 *[101, 定理 5.30]*的唯一性， c从而证明 $\Pi(\mu,\nu)$ 中的 -循环单调测度是唯一的，且由一个 *Monge* 映射给出。该证明关键地利用了代价函数 $c(\cdot,\cdot)$ 的结构，例如， c-凸函数（定义见附录 *B.4*）在通常意义下仍是凸函数，其（几乎处处）梯度是 $\mathcal{G}$-等变的。

请注意，极小化解 $S(\mathrm{x},\mathrm{h}) = \operatorname{argmin}_{Q\in\mathcal{G}} \|Q^\top \mathrm{x} - \mathrm{h}\|^2$ 通常不一定唯一。例如，当 $\mathcal{G} = \mathrm{O}(p)$ 时，有许多正交矩阵可以将给定向量映射到给定方向（例如参见图 2c）。然而，在定理 2.1 的证明中，我们证明带符号秩$S(\mathrm{X}\ R(\mathrm{X}))R(\mathrm{X})$ 几乎必然唯一——它是 $R(\mathrm{X})$ 的轨道中的点，即 $\{QR(\mathrm{X}) : Q \in \mathcal{G}\}$，且该点与 X 的距离最近。

借助定理 $R(\cdot)$ 中定义的总体秩映射 $\nabla\psi(\cdot)$ 和符号秩映射，我们在下文陈述一个关于样本广义符号、秩和带符号秩的收敛定理。我们作出以下假设：

假设 2 （关于 ν_n 的弱收敛）。 经验测度 $\nu_n := \frac{1}{n}\sum_{i=1}^n \delta_{\mathbf{h}_i}$ 弱收敛于参考分布 ν。

下面的结果表明，这些样本量在平均 ρ-损失下收敛于其总体对应量。这里，令 $\rho(\cdot,\cdot)$ 为 $\mathbb{R}^p \times \mathbb{R}^p$ 上任意满足以下条件的连续非负函数：(i) $\rho(\mathrm{x},\mathrm{x}) = 0, \forall \mathrm{x} \in \mathbb{R}^p$；(ii) 存在 $C_1 > 0$，使得 $\rho(\mathrm{x},\mathrm{y}) \le C_1(\rho(\mathrm{x},0) + \rho(\mathrm{y},0)), \ \forall \mathrm{x},\mathrm{y} \in \mathbb{R}^p$；(iii) 存在 $C_2 > 0$，使得 $\rho(Q\mathrm{x},0) \le C_2(\rho(\mathrm{x},0)+1), \ \forall \mathrm{x} \in \mathbb{R}^p, Q \in \mathcal{G}$。 $\rho(\cdot,\cdot)$ 的例子包括通常的欧几里得距离 $\|\cdot - \cdot\|$、 $\mathbb{R}^p$ 上的任意范数，以及对于任意 $k>0$，满足 $= \|\mathrm{x}-\mathrm{y}\|^k$ 的 $\rho(\mathrm{x},\mathrm{y})$。

定理 2.2 (收敛到总体广义秩、符号和符号秩)。 假设 *1* 和*2* 成立。我们有：

1.（秩的收敛）若选择参考分布 ν，其集中于 B，使得商映射 $q : \mathcal{G}B \to B$ 定义为 $q(Qx) = x$（对于 $x \in B, Q \in \mathcal{G}$）关于通常的欧几里得拓扑是连续的，则 $R(\cdot) = q(\nabla\psi(\cdot))$ 是 μ-几乎处处连续的。此外，若$\mathbb{E}\rho(H_n,0) \to \mathbb{E}\rho(H,0) < \infty$ 当 $n\to\infty$ 时成立，其中 $H_n \sim \nu_n$ 且 $H \sim \nu$，则

$$\frac{1}{n}\sum_{i=1}^n \rho(R_n(\mathbf{X}_i), R(\mathbf{X}_i)) \xrightarrow{a.s.} 0, \quad \text{as } n\to\infty.$$

2.（符号的收敛）若 $\mathcal{G}$ 在 $\mathcal{G}B$上自由作用，则 $S(\mathbf{x}\ R(\mathbf{x})) := 使_{Q\in\mathcal{G}} \|Q^\top \mathbf{x} - R(\mathbf{x})\|^2$对 μ-几乎处处的 $\mathbf{x}$ 唯一确定，并且对于定义在 $\mathbb{R}^{p^2}\times\mathbb{R}^{p^2}$ 上的任意连续函数 $\widetilde{\rho}(\cdot,\cdot)$，只要$\widetilde{\rho}(Q,Q) = 0$，对所有 $Q \in \mathbb{R}^{p^2}$，我们都有

$$\frac{1}{n}\sum_{i=1}^n \widetilde{\rho}(S_n(\mathbf{X}_i), S(\mathbf{X}_i, R(\mathbf{X}_i))) \xrightarrow{a.s.} 0, \quad \text{as } n\to\infty.$$

3. (C 带符号秩的收敛）如果 $\mathbb{E}\rho(\mathbf{H}_n,\mathbf{0}) \to \mathbb{E}\rho(\mathbf{H},\mathbf{0}) < \infty$ 其中 $\mathbf{H}_n \sim \nu_n$, $\mathbf{H} \sim \nu$, *th* *en*

$$\frac{1}{n}\sum_{i=1}^n \rho(S_n(\mathbf{X}_i)R_n(\mathbf{X}_i), \nabla\psi(\mathbf{X}_i)) \xrightarrow{a.s.} 0, \quad \text{as } n\to\infty.$$

上述结果的证明见附录B.4。要证明秩的收敛，需要 q 具有连续性；而要证明 $S_n(\mathbf{X}_i)$的收敛，则需要将讨论范围限制在 $\mathcal{G}$ 具有自由群作用的情形（这包括中心对称性和符号对称性），以确保$S_n(\mathbf{X}_i)$以及 $S(\mathbf{X}_i, R(\mathbf{X}_i))$几乎必然唯一确定。符号秩的收敛所需的假设最弱，这再次说明基于符号秩的检验统计量可能

更可取。这一判断将在第3节得到证实。请注意，上述结果允许在比较样本与总体的符号、秩和符号秩时使用任何“合理的”损失函数。

3 分布无关的对称性检验

在本节中，我们将定义符号检验和Wilcoxon带符号秩检验的多元推广，它们与经典对应检验具有类似的性质（第3.1-3.2节）。随后，我们研究所提出检验的一致性（第3.3节），并陈述其相对于Hotelling的 T^2 检验的相对效率结果（第3.4节）。我们将在第3.4节介绍广义Wilcoxon符号秩（GWSR）检验的局部渐近最优性质。

3.1 广义符号检验

回顾（ $\mathcal{G}$-对称性的假设（见1.1），以及在H_0，下，广义符号$S_n(\mathbf{X}_1)\ldots,S_n(\mathbf{X}_n)$服从均匀分布Uniform($\mathcal{G}$)。因此，可以对 $\mathcal{G}$ 上的均匀性检验应用广义符号，从而定义符号检验的多元对应检验。

例 3.1 (卡方检验)。 如果 $\mathcal{G}=\{g_1,\ldots,g_k\}$ 是一个大小为 k的有限群，那么我们可以统计 $\mathcal{G}$中每个元素的观测频数，即 $Y_i:=\sum_{j=1}^n I(S_n(\mathbf{X}j)=g_i)$、 $i=1$、……、 k。在$\mathrm{H}_0,(Y_1,\ldots\ldots,Y_k)$ 下服从多项分布 $\left(n、\ \frac{1}{k}\mathbf{1}_k\right)$。此外，我们可以进行渐近 *Pearson*卡方检验：

$$\widetilde{T}_n:=\sum_{i=1}^k\frac{(Y_i-n/k)^2}{n/k}\xrightarrow{d}\chi^2_{k-1},\quad \text{as } n\to\infty.$$

~H_0 可以在 T_n取值较大时被拒绝。

然而，对于大小为 k 的群（例如，与符号对称性对应的群大小为 2^p，而~对于适中的 p也可能非常大），卡方检验统计量 T_n 收敛到其极限分布的速度可能较慢。因此，本文提出以下广义符号检验：

$$T_n:=\frac{1}{\sqrt{n}}\sum_{i=1}^n S_n(\mathbf{X}_i). \tag{3.1}$$

对于较大的 $\|T_n\|_F^2$ 值，我们拒绝原假设 H_0，其中 $\|\cdot\|_F$ 表示矩阵的 Frobenius范数。在 H_0 下，广义符号检验是分布无关的，因为符号 $S_n(\mathbf{X}_i)$ 独立同分布于 Uniform($\mathcal{G}$)。下面给出一个关于 H_0 下 $\|T_n\|_F^2$ 渐近行为的简单结果。请注意，该检验统计量也是经典符号检验的一种直接推广。

命题3.1 (广义符号检验). 我们得到以下推论：

1. 中心对称性： S_n（$\mathbf{X}_i$）在 $\mathcal{G}=\{I,-I\}$ 条件 H_0, 下独立同分布且服从均匀分布，并且 T_n （在 （3.1））因此在有限样本中等价于一个二项随机变量（经过线性变换）。 T_n 其渐近分布可由通常的一维中心极限定理得出：

$$\frac{1}{p}\|T_n\|_F^2\xrightarrow{d}\chi^2_1,\quad \text{as } n\to\infty.$$

2. 符号对称性：在原假设下， $S_n(\mathbf{X}_i)$ 是从对角正交矩阵上均匀分布独立同分布抽取的。等价地， $S_n(\mathbf{X}_i)$ 可以视为从 $\{1,-1\}^p$ 上的均匀分布独立同分布抽取。在这种情况下，

$$\|T_n\|_F^2\xrightarrow{d}\chi^2_p,\quad \text{as } n\to\infty.$$

3. 球对称性：在 H_0 下， $S_n(\mathbf{X}_i)$ 在正交群 $\mathrm{O}(p)$ 上独立同分布且服从均匀分布。在这种情况下，我们的广义符号检验表现如下：

$$p\|T_n\|_F^2\xrightarrow{d}\chi^2_{p^2},\quad \text{as } n\to\infty.$$

在 O(p) 上进行均匀性检验，最简单的方法可能是 Rayleigh 检验 [72, 第 13.2 节。2],当 $pn\mathrm{Tr}\left(S^{\top}S\right)$ 较大时，该检验会拒绝均匀性假设，其中 $S:=\frac{1}{n}\sum_{i=1}^{n}S_n(\mathbf{X}_i)$。注意，这恰好就是命题 3.1-3 中给出的广义符号检验。我们还注意到，当 $p=1$。时，上述三种情况都会化为经典双侧符号检验。

3.2 广义 Wilcoxon 带符号秩（GWSR）检验

GWSR 统计量可以自然地定义为 $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}S_n(\mathbf{X}_i)R_n(\mathbf{X}_i)$。我们还考虑一种更一般的得分变换版本：对于得分函数 $J:B\to\mathbb{R}^p$，定义

$$\mathbf{W}_n:=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}S_n(\mathbf{X}_i)J(R_n(\mathbf{X}_i)). \tag{3.2}$$

注意$\mathbf{W}_n$ 在 H 下是分布无关的$_0$ （见（1.1））根据命题 2.2。除了在有限样本中使用其精确分布外，还可以使用其渐近分布，当 $n\to\infty$时，这会得到一种计算简单的检验。

定义 3.1（有效参考分布）。我们将通过得分函数 $J(\cdot)$ 和参考分布 ν 得到的有效参考分布（*ERD*）定义为 $SJ(\mathbf{H})$的分布，其中 S ~均匀分布$(\mathcal{G})$ 与$\mathbf{H}\sim\nu$独立。

ERD 是 $\mathcal{G}$-对称的，即对于所有 $Q\in\mathcal{G}$，$QSJ(\mathbf{H})$ 与 $SJ(\mathbf{H})$ 具有相同的分布。令Σ_{ERD} 为 ERD 的协方差矩阵。对于较大的数值，GWSR 检验拒绝原假设 H_0。

$$\mathbf{W}_n^{\top}\Sigma_{\mathrm{ERD}}^{-1}\mathbf{W}_n. \tag{3.3}$$

在接下来的几个小节中，我们研究 GWSR 检验在原假设下以及（局部）相邻备择下的表现。为此，我们需要以下假设中的一部分，现列出如下，以方便读者查阅。

我们称 $J:\ \mathcal{G}B\to\mathbb{R}^p$ 在 $\mathcal{G}$ 的群作用下是等变的，如果对于所有 $Q\in\mathcal{G}$ 和 $\mathbf{x}\in\mathcal{G}B$，都有 $J(Q\mathbf{x})=QJ(\mathbf{x})$。如果 $J:\ B\to\mathbb{R}^p$ 可以扩展为定义在 $\mathcal{G}B$ 上的等变函数，那么 $\mathbf{W}_n$ 将是一个基于符号秩的统计量。

假设 3（关于得分函数 J）：$J:\ \mathcal{G}B\to\mathbb{R}^p$ 是 ν_S-几乎处处连续的，并且在 $\mathcal{G}$的群作用下等变。

Assu假设 4（ν_n 到 ν的矩收敛）。令 $\mathbf{H}_n\sim\nu_n=\frac{1}{n}\sum_{i=1}^{n}\delta_{\mathbf{h}_i}$ 和 $\mathbf{H}\sim\nu$。

（i）（一阶矩）$\mathbb{E}\|J(H_n)\|\to\mathbb{E}\|J(H)\|<\infty$，当 $n\to\infty$ 时。（ii）（二阶矩）$\mathbb{E}\|J(H_n)\|^2\to\mathbb{E}\|J(H)\|^2<\infty$，当 $n\to\infty$ 时。（iii）（二阶矩）$\mathbb{E}|J(H_n)_iJ(H_n)_j|\to\mathbb{E}|J(H)_iJ(H)_j|<\infty$，当 $n\to\infty$ 时，对于 i、$j\in\{1,\ldots,p\}$。

假设 5（关于群 $\mathcal{G}$）。（i）$\mathbb{E}$[均匀分布$(\mathcal{G})]=\mathbf{0}_{p\times p}$；（$ii$）$-I_p\in\mathcal{G}$。

请注意，在 ν_n 弱收敛于 ν 的条件下（见假设 2），假设 4-（iii）意味着假设 4-（ii），而假设 4-（ii）意味着假设 4-（i）。根据强大数定律，如果 $J\#\nu$ 具有有限二阶矩，且 ν_n 是从 ν 中独立同分布抽样得到的，那么ν_n 满足假设 4-（iii）。如果 $J(\cdot)$ 有界且连续，那么只要 ν_n 弱收敛于 ν，根据控制收敛定理，假设 4-（iii）也成立。类似地，假设 5-（ii）意味着假设 5-（i），而中心对称性、符号对称性和球对称性均满足假设 5-（ii）。

以下结果（证明见附录B.3）表明 $\mathbf{W}_n$ 具有渐近正态分布，与经典的 Wilcoxon 符号秩统计量相呼应。

定理 3.1 (W$_n$的渐近正态性**)**。假设条件 2、*4-(iii)* 和 *5-(i)* 成立，并且用于定义 $J(\cdot)$ 的得分函数 $\mathbf{W}_n$ *(*参见 (3.2)*)* 是 ν-几乎处处连续的。那么，在 H_0，

$$\mathbf{W}_n\xrightarrow{d}N\left(\mathbf{0}_p,\Sigma_{\mathrm{ERD}}\right).$$

3.3 检验的一致性

如果 X 具有 $\mathcal{G}$-对称性，那么根据 ψ 的等变性（见定理 2.1），对于所有 $Q \in \mathcal{G}$，$Q\nabla\psi(\mathbf{X})=\nabla\psi(Q\mathbf{X}) \overset{d}{=} \nabla\psi(\mathbf{X})$。因此，$\nabla\psi(\mathbf{X})$ 也具有 $\mathcal{G}$-对称性，并且如果假设 5-(i) 成立且 $J(\cdot)$ 具有等变性，则 $\mathbb{E}[J(\nabla\psi(\mathbf{X}))] = \mathbf{0}$。另一方面，如果 X 不具有 $\mathcal{G}$-对称性，那么 $\mathbb{E}[J(\nabla\psi(\mathbf{X}))]$ 可能不为 0。以下结果（定理 3.2；证明见附录 B.4）表明，在此假设下，我们的 GWSR 检验将是一致的。

定理 3.2（GWSR 检验的一致性）。假设 *1*、*2*、*3* 和 *4-(i)* 成立。则无分布 *GWSR* 检验（见 （3.3））对于满足以下条件的所有备择情形均具有一致性：$\mathbb{E}[J(\nabla\psi(\mathbf{X}))] \neq \mathbf{0}$。

自然会问，条件 $\mathbb{E}[J(\nabla\psi(\mathbf{X}))] \neq \mathbf{0}$ 是否对位置平移备择成立。下面的结果（证明见附录 B.6 ）表明确实如此，从而说明，对于 3.3）中的 $\mathcal{G}$-对称性（若 $-I \in \mathcal{G}$），GWSR 检验对中心对称分布的位置平移备择具有一致性。

推论 3.1（位置偏移模型中的一致性）。假设 *1*、*2*、*4-(i)* 和 *5-(ii)* 成立。设 $J(\mathbf{x}) = \mathbf{x}$，且 $\mathbf{X}$ 是由 $\boldsymbol{\Delta} \neq \mathbf{0}$平移得到的中心对称分布，即：

$$\mathbf{X} - \boldsymbol{\Delta} \overset{d}{=} \boldsymbol{\Delta} - \mathbf{X}.$$

如果 ψ （定理*2.1*中定义的）在正测度的开集上严格凸[5] ，那么 *GWSR* 检验是一致的。 μ

现在我们来讨论广义符号检验的一致性性质。回顾一下，$S_n(\mathbf{X}_i) \equiv S(\mathbf{X}_i, R_n(\mathbf{X}_i))$，其中 $S(\mathbf{x}, \mathbf{h}) = \operatorname{argmin}_{Q\in\mathcal{G}} \|Q^\top \mathbf{x} - \mathbf{h}\|^2$。当群作用是自由的时，$S(\mathbf{x}, \mathbf{h})$ 将在（恒等映射 R）#μ-几乎处处唯一（见定理 2.2）。例如：

1.（中心对称性）$S(\mathbf{x}, \mathbf{h}) = \operatorname{sign}(\mathbf{x}^\top\mathbf{h})I_p$ 唯一，当且仅当 $\mathbf{x}^\top\mathbf{h} \neq 0$。2.（符号对称性）$S(\mathbf{x}, \mathbf{h}) = \operatorname{diag}(\operatorname{sign}(\mathbf{x} \odot \mathbf{h}))$ 唯一，当 $\mathbf{x} \odot \mathbf{h} \neq \mathbf{0}$时，其中对于 $\mathbf{a}, \mathbf{b} \in \mathbb{R}^p$，$\mathbf{a} \odot \mathbf{b}$ 表示 $(a_1b_1, \ldots, a_pb_p)$，而 $\operatorname{sign}(\cdot)$ 是分量符号。

注 *3.1*（球对称性）。对于球对称性，极小化解 $S(x, h)$ 不唯一，因此需要从所有极小化解构成的集合中均匀随机地选取。我们可以通过引入一个独立的高斯向量 $\varepsilon \sim N(0, I_{p(p-1)})$，按如下方式选取极小化解：假设 $h, x \neq 0$。令 $v = h/\|h\|$，$w = x/\|x\|$，并令 $V and W$ 为 $p \times (p-1)$ 矩阵，使得$V^\top V = W^\top W = I_{p-1}$, $V^\top v = W^\top w = 0$。将 $S(x, h) \equiv S(x, h, \varepsilon)$ 生成为 x, h, ε 的函数：

(a) 将 $v^\top$ 设为第一行，并将其扩展为一个 $p \times p$ 正交矩阵 $[v \ V]^\top \in O(p)$。*(b)* 将 w 设为第一列，并添加 $p(p-1)$ 个独立同分布的标准正态随机变量 $\varepsilon \in \mathbb{R}^{p(p-1)}$，以构成一个 $p \times p$ 矩阵 $[w \ \varepsilon]$。按列的顺序对 $[w \ \varepsilon]$ 的各列应用格拉姆-施密特正交化，得到 $GS([w \ \varepsilon]) \in O(p)$。*(c)* 令 $S(x, h, \varepsilon) = GS([w \ \varepsilon])[v \ V]^\top$。

上述程序的有效性见附录 *A.2*。

下述结果（证明见附录 B.4）给出了本文所考虑的 $\mathcal{G}$ 三个主要示例（分别对应中心对称、符号对称和球对称）的广义符号检验具有一致性的充分条件。

定理 3.3（广义符号检验的一致性）。假设 *1* 和 *2* 成立。那么，当 H_0 的取值较大时拒绝该假设的广义符号检验 $\|T_n\|_F$具有一致性，当且仅当

- *(中心/符号对称性)* $\mathcal{G}$ 在 $\mathcal{G}B$ 上自由作用，且 $\mathbb{E}[S(\mathbf{X} \ R(\mathbf{X}))] \neq \mathbf{0}$。

[4]对于与$\mathbf{H} \sim \nu$独立的 $S \sim$均匀分布$(\mathcal{G})$，有 $\mathbb{E}[J(\nabla\psi(\mathbf{X}))] = \mathbb{E}[J(S\mathbf{H})] = \mathbb{E}[SJ(\mathbf{H})] = \mathbb{E}[S]\mathbb{E}[J(\mathbf{H})] = \mathbf{0}$。[5]$\psi$ 在 U 上称为严格凸，如果对任意$\mathbf{x}$、$\mathbf{y} \in U$及 $\alpha \in (0,1)$，都有 $\psi(\alpha\mathbf{x} + (1-\alpha)y) < \alpha\psi(\mathbf{x}) + (1-\alpha)\psi(\mathbf{y})$。

- *(球对称性)* $\mathcal{G} = \mathrm{O}(p)$，$\nu(\{\mathbf{0}\}) = 0$，以及 $\mathbb{E}[S(\mathbf{X}\ R(\mathbf{X})\ \varepsilon)] \neq \mathbf{0}$，其中 $S(\mathbf{x},\mathbf{h}\ \varepsilon)$ 在上述构造中给出（其要求 $\mathbf{h} \neq \mathbf{0}$）。

需要注意的是，即使在一元情形下，如果对称分布在原点的某个邻域内没有概率质量，经典符号检验也可能无法在位置平移备择下保持一致性。[6] 事实上，定理 $\mathbb{E}[S(\mathbf{X}\ R(\mathbf{X}))] \neq \mathbf{0}$ 中提到的一致性条件 3.3 直接推广了一维情形下的对应结论；见下文的备注。

注 *3.2*（当 $p = 1$时所提检验的一致性）。如果 $p = 1$ 和 ν 集中于 *(0, ∞)*，那么根据定理 *3.3*，广义符号检验对满足$\mathbb{E}[sign(X)] \neq 0$的备择情形具有一致性。这与经典双侧符号检验达到一致性所需的条件相同 *[*例如参见 *65*，第 *2* 节。*3]*。如果 $\nu =$ 服从 *Uniform(0, 1)*，那么 $R(x) = F_{|X|}(|x|)$，其中 $F_{|X|}(\cdot)$是 $|X|$的累积分布函数，并且 $\nabla\psi(x) = sign(x)F_{|X|}(|x|)$。定理 *3.2* 表明，采用恒等评分函数的 *GWSR* 检验对 $\mathbb{E}[sign(X)F_{|X|}(|X|)] \neq 0$具有一致性，这等价于经典 *Wilcoxon* 符号秩检验达到一致性的条件，即 $\mathbb{P}(X_1 + X_2 > 0) \neq \frac{1}{2}$，其中 X_1、X_2 是来自 μ的独立同分布随机变量；证明见附录 *B.5*。 _

3.4 GWSR 检验的效率

在本小节中，我们研究所提出的 GWSR 检验相对于 Hotelling T^2 检验的渐近相对效率（ARE）。首先，我们引入以下引理（证明见附录 B.7 ），该引理指出，在 H_0, 下，我们可以用总体对应量替代经验符号秩，同时保持 GWSR 统计量的极限分布不变。此类结果有时称为 H´ajek 表示（参见 H´ajek 等人 [42, 定理 1，第 6.1 节。7]），近年来在许多关于通过最优传输定义的多元秩的应用研究中均有出现[96, 44]。

引理 3.1 (H´ajek 表示)。 假设 *1*、*2*、*3*、*4-(ii)* 和 *5-(i)* 成立。那么，在 H_0,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_n(\mathbf{X}_i)J(R_n(\mathbf{X}_i)) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} S(\mathbf{X}_i, R(\mathbf{X}_i))J(R(\mathbf{X}_i)) \xrightarrow{L^2} 0.$$

上述引理结合 Le Cam 第三引理，得出以下关于相邻备择情形的中心极限定理；证明见附录 B.8 。

定理 3.4 （局部备择下的渐近分布）。 设$\mathbf{X}_1$、.....、$\mathbf{X}_n$ 独立同分布，具有定义在 $\mathbb{R}^p$ 上的勒贝格密度 $f(\cdot - \boldsymbol{\theta})$，其中 f 是一个 $\mathcal{G}$-对称分布的密度，且 $\boldsymbol{\theta} \in \mathbb{R}^p$。考虑以下假设检验:

$$\mathrm{H}_0 : \boldsymbol{\theta} = \mathbf{0}_p \qquad versus \qquad \mathrm{H}_1 : \boldsymbol{\theta} = \frac{\boldsymbol{\xi}_n}{\sqrt{n}}, \tag{3.4}$$

其中 $\boldsymbol{\xi}_n \to \boldsymbol{\xi} \in \mathbb{R}^p$ 当 $n \to \infty$时。假设条件 *1*、*2*、*3*、*4-(ii)* 和 *5-(i)* 均成立。若参数族 $\{f(\cdot - \boldsymbol{\theta})\}_{\boldsymbol{\theta}\in\mathbb{R}^p}$满足均方可微性（*A.1*附录中详述），则在 H_1 下：$\boldsymbol{\theta} = \frac{\boldsymbol{\xi}_n}{\sqrt{n}}$，

$$\mathbf{W}_n \xrightarrow{d} N\left(\boldsymbol{\gamma}, \Sigma_{\mathrm{ERD}}\right), \qquad where\ \boldsymbol{\gamma} := -\mathbb{E}_{\mathrm{H}_0}\left[J(\nabla\psi(\mathbf{X}))\frac{\boldsymbol{\xi}^\top\nabla f(\mathbf{X})}{f(\mathbf{X})}\right] \in \mathbb{R}^p.$$

上述定理意味着

$$\mathbf{W}_n^\top \Sigma_{\mathrm{ERD}}^{-1}\mathbf{W}_n \xrightarrow{d} \left\|\Sigma_{\mathrm{ERD}}^{-1/2}\boldsymbol{\gamma} + N(\mathbf{0}, I_p)\right\|^2,$$

这是一个非中心 χ^2 分布，具有 p 个自由度和非中心参数$\|\Sigma^{-1/2}{}_{\mathrm{ERD}}\boldsymbol{\gamma}\|^2$。注意，这个极限分布不直接依赖于 $\mathcal{G}$ ；它仅依赖于

[6]例如，如果 $X \sim$Un iform（$[-2,-1]\cup[1,2]$），那么符号检验对于位置偏移 $\delta < 1$ 将毫无检验功效。

关于 $J(\cdot)$， $\nabla\psi(\cdot)$（其还取决于 ν_S；见定理 2.1-(ii)），$\boldsymbol{\xi}\in\mathbb{R}^p$，以及$\mathbf{X}$ 在 H_0下的分布。我们将此检验与 Hotelling 的 T^2 统计量进行比较，其定义为

$$T^2 = n\overline{\mathbf{X}}_n^\top S_n^{-1}\overline{\mathbf{X}}_n, \tag{3.5}$$

其中 $S_n := \frac{1}{n-1}\sum_{i=1}^n(\mathbf{X}_i-\overline{\mathbf{X}}_n)(\mathbf{X}_i-\overline{\mathbf{X}}_n)^\top$ 是样本协方差矩阵。注意，当 $n\to\infty$时，$S_n \xrightarrow{p} \Sigma_{\mathbf{X}}:= \mathbb{E}[(\mathbf{X}-\mathbb{E}\mathbf{X})(\mathbf{X}-\mathbb{E}\mathbf{X})^\top]$，前提是后者有限。$\mathbf{X}\sim\mu$的绝对连续性意味着 $\Sigma_{\mathbf{X}}$ 是正定的。在相邻备择序列 $\boldsymbol{\theta}_n\equiv\boldsymbol{\xi}_n/\sqrt{n}$下，我们有

$$\sqrt{n}S_n^{-1/2}\overline{\mathbf{X}}_n - \Sigma_{\mathbf{X}}^{-1/2}\boldsymbol{\xi} = \sqrt{n}S_n^{-1/2}\left(\overline{\mathbf{X}}_n-\boldsymbol{\xi}_n/\sqrt{n}\right) + (S_n^{-1/2}\boldsymbol{\xi}_n - \Sigma_{\mathbf{X}}^{-1/2}\boldsymbol{\xi}) \xrightarrow{d} N(\mathbf{0}, I_p).$$

因此，在这一相邻备择下，Hotelling 的 T^2 统计量服从具有 p 个自由度且非中心参数为 $\|\Sigma^{-1/2}_{\mathbf{X}}\boldsymbol{\xi}\|^2$的非中心 χ^2分布。

定义 3.2（渐近相对效率）。对于 $\boldsymbol{\xi}\in\mathbb{R}^p\backslash\{\mathbf{0}\}$，考虑以下假设检验

$$\mathrm{H}_0:\boldsymbol{\theta}=\mathbf{0}_p \quad versus \quad \mathrm{H}_1:\boldsymbol{\theta}=\varepsilon_n\boldsymbol{\xi}_n,\quad \text{where } \varepsilon_n\to 0,\ \boldsymbol{\xi}_n\to\boldsymbol{\xi}$$

在位置偏移模型 $f(\cdot-\boldsymbol{\theta})$中，其中 f是定义在 $\mathbb{R}^p$上的 $\mathcal{G}$-对称分布的密度。对于$0<\alpha<\beta<1$，令 $N_1(\alpha,\beta,\boldsymbol{\xi})$为基于 $\mathbf{W}_n$ 的检验达到检验功效 β 所需的最小样本数（渐近水平为α），其中 $\boldsymbol{\theta}=\boldsymbol{\xi}$，即

$$N_1(\alpha,\beta,\boldsymbol{\xi}) := \min\left\{K\geq 1:\mathbb{P}_{\boldsymbol{\theta}=\boldsymbol{\xi}}\left(\mathbf{W}_K^\top\Sigma_{\mathrm{ERD}}^{-1}\mathbf{W}_K\geq\chi_p^2(\alpha)\right)\geq\beta\right\},$$

其中，$\chi_p^2(\alpha)\in\mathbb{R}$满足 $\mathbb{P}\left(\chi_p^2\geq\chi_p^2(\alpha)\right)=\alpha$。类似地，令 $N_2(\alpha,\beta,\boldsymbol{\xi})$表示基于 *Hotelling* 的 T^2统计量、达到检验功效 β 且渐近水平为 α的检验所需的最小样本数，其中 $\boldsymbol{\theta}=\boldsymbol{\xi}$。*GWSR* 检验相对于 *Hotelling* 的 T^2，沿方向 $\boldsymbol{\xi}$的渐近相对效率（*ARE*）定义为：

$$\mathrm{ARE}(\mathbf{W}_n,\overline{\mathbf{X}}_n;\boldsymbol{\xi}) = \lim_{n\to\infty}\frac{N_2(\alpha,\beta,\varepsilon_n\boldsymbol{\xi}_n)}{N_1(\alpha,\beta,\varepsilon_n\boldsymbol{\xi}_n)}.$$

一般而言，渐近相对效率（ARE）可能取决于 α、β、$\{\varepsilon_n\}$ 和 $\{\boldsymbol{\xi}_n\}$。然而，如果这些检验的渐近分布均为自由度相同的非中心卡方分布，如本例所示（因此定理 3.4 成立），那么渐近相对效率（ARE）不依赖于 α、β、$\{\varepsilon_n\}$，而且等于相应非中心性参数之比的平方（证明见附录 B.9）：

$$\mathrm{ARE}(\mathbf{W}_n,\overline{\mathbf{X}}_n;\boldsymbol{\xi}) = \frac{\|\Sigma_{\mathrm{ERD}}^{-1/2}\boldsymbol{\gamma}\|^2}{\|\Sigma_{\mathbf{X}}^{-1/2}\boldsymbol{\xi}\|^2}.$$

上述 对于多元分布的某些子族，渐近相对效率下界 s.

在一维情形下，当 $f(\cdot)$ 是 $N(0,\sigma^2)$的密度，其中未知的 $\sigma>0$为某个 t-检验时，Wilcoxon 符号秩检验相对于 $3/\pi\approx 0.95$ [57]的相对效率为 。以下命题（证明见附录B.10）推广了这一结果。文中还给出了一个 Chernoff-Savage 型结果，说明如果采用高斯参考分布（使用恒等评分函数），则就渐近相对效率而言，GWSR 检验不会劣于 Hotelling 的 T^2 检验。

命题 3.2（高斯情形）。假设 *1*、*2*、*4-(ii)* 和 *5-(i)* 成立。设 $f(\cdot)$ 是多元正态分布 $N(\mathbf{0}_p,\Sigma)$的密度，其中未知的正定协方差矩阵为 Σ。

1. 若 $ERD = N(0, I)$，且 $J(x) = x$，则对所有 $\boldsymbol{\xi}\neq\mathbf{0}\in\mathbb{R}^p$，$ARE(W_n, \overline{X_n}; \boldsymbol{\xi}) = 1$。*2.* 若 $ERD = Uniform(-1, 1)^p$，由 $J(x) = (2\Phi(x_1) - 1, \ldots, 2\Phi(x_p) - 1)^\top$ 得到，其中 Φ 是标准正态分布的累积分布函数（*c.d.f.*）[7]，则对所有$\boldsymbol{\xi}\neq\mathbf{0}$，$ARE(W_n, X_n; \boldsymbol{\xi}) = 3/\pi\approx 0.95$。

[7]请注意，此 ERD 不包括球面对称性检验，因为 $J(\cdot)$ 在 $\mathrm{O}(p)$ 的群作用下不是等变的。

3. 若ERD 是球面均匀分布[8], 且 $J(\mathbf{x}) = \mathbf{x}$，则对所有 $\boldsymbol{\xi} \neq \mathbf{0}$,

$$\mathrm{ARE}(\mathbf{W}_n, \overline{\mathbf{X}}_n; \boldsymbol{\xi}) = \kappa_p \geq \begin{cases} 0.95, & \text{for } p < 5, \\ 0.648, & \text{for all } p, \end{cases}$$

（精确值见附录*B.10*） κ_p。特别地， $\kappa_1 = 3/\pi$ 可化为 *Wilcoxon* 符号秩检验相对于 t检验 *[57]*的经典渐近相对效率。

请注意，只要 ERD 具有指定形式，上述结果对 $\mathcal{G}$ 的所有可能取值均成立。本节的所有结果都会呈现这一现象。一维情形下，经典 Wilcoxon 符号秩检验的一个显著结果是 Hodges-Lehmann 现象：对于位置平移备择，经典 Wilcoxon 符号秩检验相对于 t检验的相对效率绝不会低于 0。864 [57]。下述定理（证明见附录B.11）给出了这一结果的多元对应检验，但适用条件更为严格——假设多元向量的各分量相互独立。

定理 3.5 （独立分量情形）。在定理*3.4*的假设下，设 $\widetilde{\quad}$ $\widetilde{\quad}$ $\widetilde{\quad}$ 另外， $f(\mathbf{x}) = \prod_{i=1}^{p} f_i(x_i)$，对于$\mathbf{x} = (x_1, \ldots, x_p) \in \mathbb{R}^p$，其中 $f_1, \ldots, f_p$ 是方差有限且非零的一维勒贝格密度。

1. 若 *ERD* 服从均匀分布$(-1, 1)^p$，且 $J(\mathbf{x}) = \mathbf{x}$，则渐近相对效率$(\mathbf{W}_n, \overline{\mathbf{X}}_n; \boldsymbol{\xi}) \geq \frac{108}{125} \approx 0.864$。

进一 $\widetilde{\quad}$ 步，渐近相对效率下界取等，当且仅当对所有 $\xi_i \neq 0$， f_i 的密度为：

$$\widetilde{f}_i(x) = \frac{3}{20\sqrt{5}\sigma_i^3}(5\sigma_i^2 - x^2) I(|x| \leq \sqrt{5}\sigma_i), \quad \text{for some } \sigma_i > 0.$$

2. 若 *ERD* 服从 $N(\mathbf{0}, I)$，且 $J(\mathbf{x}) = \mathbf{x}$，则渐近相对效率$(\mathbf{W}_n, \overline{\mathbf{X}_n}; \boldsymbol{\xi}) \geq 1$。

进一 $\widetilde{\quad}$ 步，当且仅当对所有 $\xi_i \neq 0$， f_i 是正态分布的密度 n.

GWSR 检验相对于 Hotelling T^2 检验的渐近相对效率，也可在另一类丰富的多元概率分布上得到下界，即椭圆对称分布类；下述定理给出了这一结果（证明见附录 B.12 ）。

定理 3.6 （椭圆对称情形）。 在定理 *3.4*的假设下，进一步假设 $\mathbf{X}$ 在H_0 下具有有限二阶矩，且

$$f(\mathbf{x}) \propto (\det(\Sigma))^{-\frac{1}{2}} \underline{f}\left(\mathbf{x}^\top \Sigma^{-1} \mathbf{x}\right), \qquad \textit{for } \mathbf{x} \in \mathbb{R}^p,$$

对于某个未知的正定矩阵 Σ以及某个未知函数 $f: \mathbb{R}^+ \to \mathbb{R}^+$。在关于 f *[25, 43]* 的光滑性和可积性的标准正则性条件下（详见附录 *B.12*），有 $\underline{\quad}$

1. 若 *ERD* 为球面均匀分布，且 $J(\mathbf{x}) = \mathbf{x}$，则：

$$\mathrm{ARE}(\mathbf{W}_n, \overline{\mathbf{X}}_n; \boldsymbol{\xi}) \geq \frac{81}{500} \cdot \frac{(\sqrt{2p-1}+1)^5}{p^2(\sqrt{2p-1}+5)} \geq 0.648.$$

使上述渐近相对效率下界取到的分布见附录*B.12* 。

2. 若 *ERD* 为 $N(\mathbf{0}, I)$ 且 $J(\mathbf{x}) = \mathbf{x}$，则 渐近相对效率$(\mathbf{W}_n, \overline{\mathbf{X}_n}; \boldsymbol{\xi}) \geq 1$。

当且仅当 $f(\cdot)$ 为多元正态分布时，上述等式成立。

下面，我们考虑一类用于盲源分离 [16, 23,97, 25]和独立成分分析 [61, 90, 5]的生成模型，并推导出类似的渐近相对效率下界（证明见附录B.13 ）。

[8]也就是说，单位球面上的均匀分布与到原点距离在单位区间上的均匀分布的乘积分布 [43, 95, 45, 44]。

定理 *3.7*（盲源混合生成模型）。在定理 *3.4* 的假设下，进一步假设在 H_0 下，$X = AW$，其中 A 是某个未知的正交矩阵，W 具有 $\sim$ $\sim$ 独立分量，其勒贝格密度为 $f(w) = \prod_{i=1}^{p} f_i(w_i)$，其中 $w = (w_1, \ldots, w_p)$，且 $\sim$ $\sim f_1, \ldots, f_p$ 是单变量密度。假设 *ERD* 为 $N(0, I)$，其中 $J(x) = x$，并且在 H_0 下，X 具有有限二阶矩。那么：

$$\mathrm{ARE}(\mathbf{W}_n, \overline{\mathbf{X}}_n; \boldsymbol{\xi}) \geq 1.$$

当且仅当 $\widetilde{f}_i$ 是正态分布的密度时，等式成立，其中 $(A^\top \boldsymbol{\xi})_i \neq 0, i = 1, 2, \ldots, p$。

请注意，在上述所有渐近相对效率结果中，只要 ERD 具有所需形式，$\mathcal{G}$ 就不受限制。定理 3.7 再次说明，采用高斯 ERD 和恒等评分函数的优越性。基于这一事实，我们针对不同的 ν，建议在下文中为其选择一些参考测度 $\mathcal{G}$。

注3.3（参考测度 ν的示例）。这里基于得分函数 $J(\mathbf{x}) = \mathbf{x}$的规范选择，推荐几种参考测度 ν：

1. 我们首先推荐采用高斯 *ERD* 的参考分布 ν—— $N(0, I_p)$。例如，对于中心对称性，可以令 ν 为 $(|Z_1|, Z_2, \ldots, Z_p)$ 的分布（集中于 $B = (0, \infty) \times \mathbb{R}^{p-1}$），其中 Z_i均为独立同分布的标准正态分布；对于符号对称性，可以令 ν为 $(|Z_1|, |Z_2|, \ldots, |Z_p|)$ 的分布（此处 $B = (0, \infty)^p$）。对于球对称性，可以令 ν为 $\left(\sqrt{Z_1^2 + \ldots + Z_p^2}, 0, \ldots, 0\right)$ 的分布，且 $B = (0, \infty) \times \{0\}^{p-1}$。使用高斯 *ERD* 尤为有效，因为对于一大类模型，*GWSR* 检验相对于 *Hotelling* 的 T^2 检验的渐近相对效率始终不低于 *1*（见定理 *3.5–3.7*）。

2. 我们的第二项建议是采用具有均匀 *ERD* 的参考分布 ν ——均匀分布$(-1,1)^p$。例如，对于中心对称性，我们可以令 ν 为$(|U_1|, U_2, \ldots, U_p)$ 的分布（集中于 $B = (0, 1) \times (-1, 1)^{p-1}$），其中 U_i 相互独立同分布，服从均匀分布 $(-1, 1)$；对于符号对称性，我们可以令 ν 为 $(|U_1|, |U_2|, \ldots, |U_p|)$ 的分布（因此 $B = (0, 1)^p$）。注意，均匀分布 $(-1,1)^p$ 不具有球对称性，因此这种 *ERD* 不适用于球对称性假设检验。

3. 我们的第三项建议是采用具有球面均匀 *ERD* 的参考分布 ν，它可简化为 均匀分布$(-1,1)$ 当 $p = 1$，并且在利用最优传输进行分布无关推断的文献中得到了广泛倡导 *[43, 95, 45, 44]*。中心对称性、符号对称性和球对称性的参考测度也可以参照上述高斯 *ERD* 的描述，以类似方式构造。

到目前为止，我们已经说明了 GWSR 检验相对于 Hotelling 的 T^2 检验的渐近相对效率下界。一个自然的问题是，广义符号检验是否也具有类似性质？对于符号对称性，答案是肯定的：此时广义符号与各分量的符号一致；Bickel [12]对此进行了研究。下述定理直接推自 Bickel[12, 公式 (5.5)]。

定理 3.8（符号对称性下广义符号检验的渐近相对效率 [12]）。假设参考分布 ν_n 集中于$(0, +\infty)^p$，对所有 n均如此。设勒贝格密度 $f(\cdot)$ 具有符号对称性且协方差矩阵非退化，并且每个边缘密度 f_i 的众数均为 f 处的0，且在0处连续。考虑模型 $f(\cdot - \boldsymbol{\theta})$ 和 H_0: $\boldsymbol{\theta} = \mathbf{0}_p$，并针对相邻备择假设进行检验。那么，采用符号对称性的广义符号检验相对于 *Hotelling* 的 T^2 检验的相对效率不会低于 $1/3$。

在表1中，给出了广义符号检验（针对符号对称性）相对于 Hotelling 的 T^2 检验在若干分布下的相对效率（采用公式 [12, 公式（5.2）]），这与一维情形完全相同 [99, 表 14。1]。此外，通过在 0 处设计一个较大的密度，相对效率可以任意增大，这将使符号检验相较于 Hotelling 的 T^2 检验具有明显优势。不过，对于其他类型的对称性，渐近相对效率通常没有下界。渐近相对效率可能取决于 α、β、参考分布、$\boldsymbol{\xi}$的方向，以及

表 1：广义符号检验（用于检验符号对称性）相对于 Hotelling's T^2检验在若干符号对称分布下的相对效率。相对效率的最小可能值为1/3。

分布	渐近相对效率（符号检验/Hotelling's T^2）
拉普拉斯分布：$\prod_{i=1}^{p}$ 拉普拉斯分布$(0, b_i)$	2
正态分布：$\prod_{i=1}^{p} N(0, \sigma_i^2)$	$2/\pi$
逻辑回归：$\prod_{i=1}^{p}$ Logistic$(0, s_i)$	$\pi^2/12$
均匀分布：$\prod_{i=1}^{p}$ 均匀分布$(-c_i, c_i)$	1/3（下界）

维数 p。从第 4 节的模拟结果中，我们还发现，采用符号对称性的广义符号检验往往比采用中心对称性或球对称性的表现更好。因此，我们建议在进行广义符号检验时，优先选择符号对称性。

到目前为止，我们已经证明，对于位置平移备择的自然（且实用）子族选择以及适当的得分函数（主要是 $J(\mathbf{x}) = \mathbf{x}$），GWSR 检验的表现至少不逊于 Hotelling 的 T^2 检验。在以下结果中（证明见附录 B.14），我们通过两个独立的结果表明，存在一种经过得分变换的 Wilcoxon 带符号秩检验（采用适当选择的得分函数），该检验具有局部渐近最优性，即在所有具有固定第一类错误率的检验中，它具有最大的局部功效（针对相邻备择）。

定理 3.9（局部渐近最优性）。设 *\(X_1,\ldots,X_{\v1\}\)* 是来自 *\(\v2\)* 的一个样本，其勒贝格密度为 *\(\v3\}(\v4\})\)*，其中 *\(\v5\)* 关于 *\(\v7\}:\v8\)* 具有 *\(\v6\)*-对称性。在定理 *3.4* 的假设下，设 *\(\v9\)* 为在 *\(\v10\)* 处的非奇异费 *Fisher* 信息矩阵。假设 *\(\v11\)* 具有勒贝格密度。令 *\(\v12\}(y):\v13\}\sup_{\v14\}y\v15\}x \v16\}(x)\v17\})* 为 *\(\v18\)* 的 *Legendre-Fenchel* 变换，并假设

$$J(\mathbf{x}) := -\frac{\nabla f(\nabla\psi^*(\mathbf{x}))}{f(\nabla\psi^*(\mathbf{x}))}, \qquad for\ \mathbf{x} \in \mathcal{G}B, \tag{3.6}$$

用于定义$\mathbf{W}_n$（通过（3.2））。

1.（局部极大极小最优检验）设 $\boldsymbol{\xi} \neq \mathbf{0} \in \mathbb{R}^p$。对于任意一列水平为 $\alpha \in$（0、1）的检验$\{\phi_n\}_{n\geq 1}$，用于检验

$$\mathrm{H}_0 : \boldsymbol{\theta} = \mathbf{0}_p \qquad \text{versus} \qquad \mathrm{H}_1 : \boldsymbol{\theta} = \frac{\boldsymbol{\xi}}{\sqrt{n}} \tag{3.7}$$

其功效函数为 $\pi_n(\boldsymbol{\theta}) := \mathbb{E}_{\boldsymbol{\theta}}[\phi_n]$，且 $\pi_n\left(\frac{\boldsymbol{\xi}}{\sqrt{n}}\right)$ 在 H_1 下对任意 $\boldsymbol{\xi} \in \mathbb{R}^p$ 收敛，我们有

$$\inf_{\boldsymbol{\xi}:\|I_{\boldsymbol{\theta}_0}^{1/2}\boldsymbol{\xi}\|=c} \lim_{n\to\infty} \pi_n\left(\frac{\boldsymbol{\xi}}{\sqrt{n}}\right) \leq \inf_{\boldsymbol{\xi}:\|I_{\boldsymbol{\theta}_0}^{1/2}\boldsymbol{\xi}\|=c} \lim_{n\to\infty} \mathbb{P}_{\boldsymbol{\theta}_0+\frac{\boldsymbol{\xi}}{\sqrt{n}}}\left(\mathbf{W}_n^\top \Sigma_{\mathrm{ERD}}^{-1} \mathbf{W}_n \geq \chi_p^2(\alpha)\right),$$

对所有 $c > 0$ 均成立。

2.（局部最强检验）设 $\zeta:\ \mathbb{R}^p \to \mathbb{R}$ 在 $\boldsymbol{\theta}_0 = \mathbf{0}_p$ 处可微，且 $\zeta(\boldsymbol{\theta}_0) = 0$。设$\boldsymbol{\xi} \in \mathbb{R}^p$ 满足 $\boldsymbol{\xi}^\top \nabla\zeta(\boldsymbol{\theta}_0) > 0$。对于任意一列水平为 $\alpha \in$（0、1）的检验 $\{\phi_n\}_{n\geq 1}$，用于

$$\mathrm{H}_0 : \zeta(\boldsymbol{\theta}) \leq 0 \qquad \text{versus} \qquad \mathrm{H}_1 : \zeta(\boldsymbol{\theta}) > 0 \tag{3.8}$$

其功效函数为 $\pi_n(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}[\phi_n]$，我们有

$$\limsup_{n\to\infty} \pi_n\left(\frac{\boldsymbol{\xi}}{\sqrt{n}}\right) \leq \lim_{n\to\infty} \mathbb{P}_{\boldsymbol{\theta}_0+\frac{\boldsymbol{\xi}}{\sqrt{n}}}\left(\frac{\nabla\zeta(\boldsymbol{\theta}_0)^\top I_{\boldsymbol{\theta}_0}^{-1}\mathbf{W}_n}{\sqrt{\nabla\zeta(\boldsymbol{\theta}_0)^\top I_{\boldsymbol{\theta}_0}^{-1}\nabla\zeta(\boldsymbol{\theta}_0)}} \geq z_\alpha\right),$$

其中 z_α 满足 $\Phi(z_\alpha) = 1-\alpha$，且 $\frac{\nabla\zeta(\boldsymbol{\theta}_0)^\top I_{\boldsymbol{\theta}_0}^{-1}\mathbf{W}_n}{\sqrt{\nabla\zeta\ \boldsymbol{\theta}_0\ ^\top I_{\boldsymbol{\theta}_0}^{-1}\nabla\zeta\ ()(\boldsymbol{\theta}_0)}} \geq z_\alpha$ 是适当的（渐近）水平 α 基于 *GWSR* 统计量的检验 $\mathbf{W}_n$ 对于 (3.8)。

上述第一个结果表明，GWSR 检验在轮廓 $\{\boldsymbol{\xi} \in \mathbb{R}^p$ 上使最小（局部）检验功效达到最大：$\|I_{\boldsymbol{\theta}_0}^{1/2}\boldsymbol{\xi}\| = c\}$ 对于任意 $c > 0$；而第二个结果告诉我们，如果假设是单侧的，那么基于 GWSR 统计量 $\mathbf{W}_n$ 的检验对于任意固定方向 $\boldsymbol{\xi} \in \mathbb{R}^p$都能达到最大（局部）检验功效。尽管基于 $\mathbf{W}_n$ 的检验可以在局部渐近意义下达到最优，但最优得分函数 $J(\cdot)$（见（3.6））取决于底层分布 f，且通常未知。

例 *3.2*（高斯情形）。假设已知 $f(\cdot)$（在定理 *3.9* 中）服从高斯分布 $N(\mathbf{0}_p, \Sigma)$，其未知协方差矩阵 Σ 非奇异。于是，当 $SH \sim N(0, I_p)$ 时，定理 *3.9* 表明，最优得分函数（见（*3.6*））为 $J(x) = \Sigma^{-1/2}x$。尽管 Σ 未知，但可以通过样本协方差矩阵 $\widehat{\Sigma}$ 对其进行估计。定义 $\widehat{J}(x) = \widehat{\Sigma}^{-1/2}x$。于是，

$$\mathbf{W}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_n(\mathbf{X}_i)J(R_n(\mathbf{X}_i)) = \Sigma^{-1/2}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_n(\mathbf{X}_i)R_n(\mathbf{X}_i)\right)$$

其渐近分布与

$$\widehat{\Sigma}^{-1/2}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_n(\mathbf{X}_i)R_n(\mathbf{X}_i)\right). \tag{3.9}$$

这意味着，使用数据驱动得分函数 $J(\cdot)$ 与使用 $J(\cdot)$具有相同的渐近分布，因此基于 (3.9) 的 *GWSR* 检验也具有相同的局部渐近最优性质。

将上述例子推广，如果我们能够充分准确地估计最优得分函数 $J(\cdot) = -\frac{\nabla f(\nabla\psi^*(\cdot))}{f(\nabla\psi^*(\cdot))}$，使得在定理 3.9 中有

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_n(\mathbf{X}_i)J(R_n(\mathbf{X}_i)) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} S_n(\mathbf{X}_i)\widehat{J}(R_n(\mathbf{X}_i)) + o_p(1),$$

那么 GWSR 采用 $\widehat{J}$ 将如定理 3.9 所述，在局部渐近意义下达到最优。

然而，一般来说，估计最优得分函数 $J(\cdot) = -\frac{\nabla f(\nabla\psi^*(\cdot))}{f(\nabla\psi^*(\cdot))}$ 面临两个主要障碍。若无额外信息，这两个问题都很难解决。首先，$\nabla f(\cdot)$ 与 $f(\cdot)$ 未知，可能需要通过经验数据进行估计。原则上，估计对数密度函数的梯度是一个困难的非参数问题 [60]。Hyv¨arinen [60]，给出了一个积极的结果：作者提出了得分匹配法，通过最小化 $\mathbb{E}_{\mathbf{X}\sim f}\left[\|s_\tau(\mathbf{X}) - \nabla \log f(\mathbf{X})\|_2^2\right]$ 在 τ 上的经验版本，其中 s_τ：$\mathbb{R}^p \to \mathbb{R}^p$ 由 τ 参数化。在 Song 和 Ermon [98]，中，s_τ 采用神经网络；s_τ 也可以属于一个参数族 [60]。另一个障碍是估计 $\nabla\psi^*(\cdot)$（即从 ν_S 到 $P_{\boldsymbol{\theta}_0}$ 的传输映射）。尽管有了来自 $P_{\boldsymbol{\theta}_0}$ 的足够多数据，我们可以将 $\nabla\psi^*(\cdot)$ 估计到任意精度 [26, 70]，，但其收敛速率仍可能受到维数灾难的影响 [26, 70]。

4 有限样本性能

在本节中，我们考察分布无关方法在有限样本下的表现。我们首先展示第 3 节中的渐近相对效率结果，然后将我们提出的方法与文献中用于检验不同类型对称性的现有方法进行比较。本节将使用恒等评分函数$J(\mathbf{x}) = \mathbf{x}$。

4.1 渐近相对效率结果的实证验证

考虑对以下分布进行 $\mathcal{G}$-对称性检验：

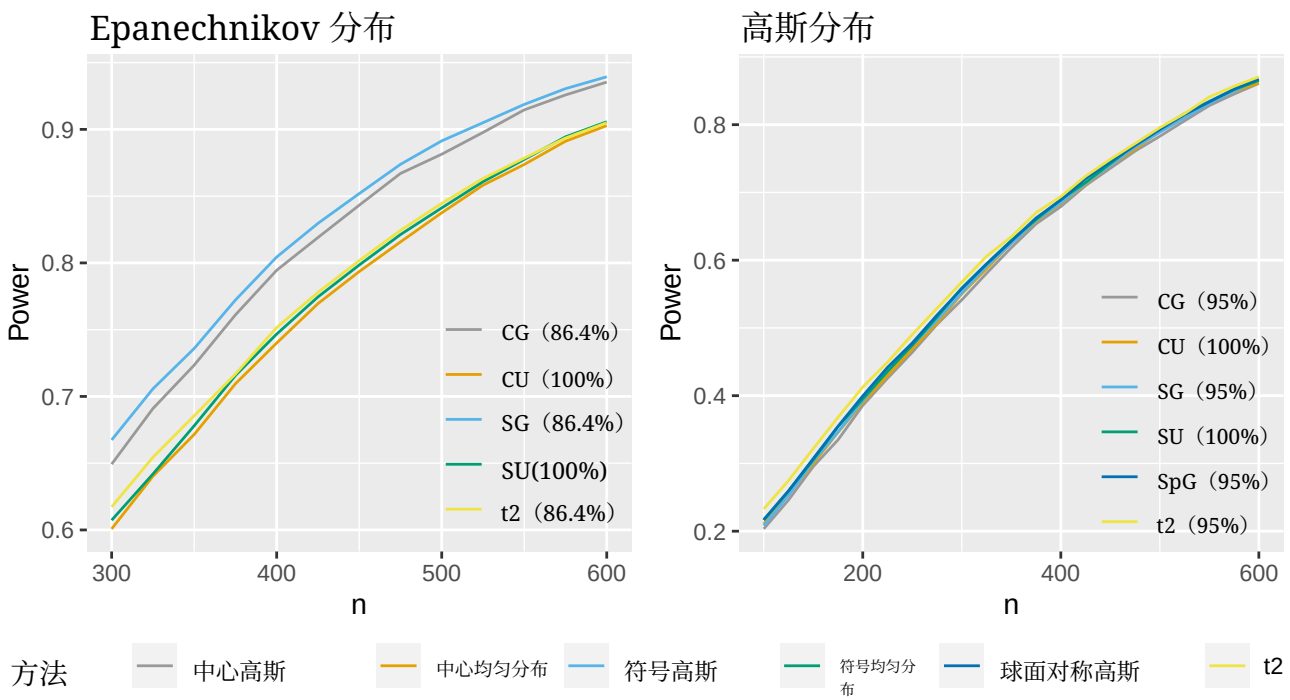


图 4：左：基于平移的埃潘尼奇科夫分布中样本量为 n 或 $0.864n$ （括号内标示百分比）的不同方法的功效曲线。右：基于平移的高斯分布中 n （或 $0.95n$）个样本的不同方法的功效曲线。这里，CG、CU、SG、SU 和 SpG 分别表示采用不同 ERD 的 GWSR 检验，用于检验不同的对称性；第一个字母表示所检验的对称性（C 表示中心对称，S 表示符号对称，Sp 表示球对称），第二个字母表示所采用的 ERD（G 表示高斯（$\mathbf{0}, I_p$），U 表示均匀分布$[-1,1]^p$）。“t2”表示 Hotelling's T^2 检验。

(a) 二元埃潘尼奇科夫分布（见定理 3.5），其 $\sigma_i = \frac{1}{\sqrt{5}}$ 且各分量相互独立，平移了 $0.05 \cdot \mathbf{1}_2$。此处检验中心对称性和符号对称性。(b) 平移了 $0.1 \cdot \mathbf{1}_2$的二元标准高斯分布。此处检验中心对称性、符号对称性和球对称性。

我们将考虑两类有效参考分布：(i) 高斯 ERD： $N(\mathbf{0},\ I_2)$，以及 (ii) 均匀 ERD：Uniform$(-1,\ 1)^2$。针对不同类型的对称性，我们采用备注 ν 3.3 中推荐的参考分布：

- 采用均匀 ERD 的中心对称性（CU）： $\nu \sim$ Uniform$(0,\ 1) \times$ Uniform$(-1,\ 1)$。
- 采用均匀 ERD 的符号对称性（SU）： $\nu \sim$ Uniform$(0,\ 1)^2$。
- 采用高斯 ERD 的中心对称性（CG）： $\nu \sim |N(0,\ 1)| \times N(0,\ 1)$。
- 高斯 ERD 下的符号对称性 (SG)： $\nu \sim |N(0,1)| \times |N(0,1)|$。
- 高斯 ERD 下的球对称性 (SpG)：$\nu \sim \sqrt{\chi^2_2} \times \{0\}$。

在每次重复中，通过从 ν 抽取 n 个独立同分布观测值并将其固定，得到经验参考分布 ν_n。我们提出的 GWSR 检验在水平 $\alpha = 0.05$、拒绝域为下式时的经验检验功效（基于 10^4 次重复）

$$\mathbf{W}_n^\top \Sigma_{\mathrm{ERD}}^{-1} \mathbf{W}_n \geq \chi^2_p(\alpha)$$

显示于图4，对应不同的样本量 n。同时给出了 Hotelling 的 T^2 检验功效，其拒绝域为：$T^2 \geq \chi^2_p(\alpha)$；见 (3.5)。

根据定理3.5 可知，(a) 中的 Epanechnikov 分布达到了均匀 ERD 的 0.864 下界。因此，对于均匀 ERD，使用 n 个样本时，我们采用中心对称性（CU）和符号对称性（SU）的方法，其检验功效几乎与使用T^2 个样本的 Hotelling 检验 $0.864n$ 功效相同；见图4左图中的黄色、橙色和绿色曲线。

如果 ERD 是高斯分布，那么根据定理 3.5，我们的方法相对于 Hotelling's T^2 的渐近相对效率严格大于 1。这也可以从图 4 的左图中看出——蓝色曲线和灰色曲线（分别对应 CG 和 SG）均严格位于黄色曲线（对应 Hotelling's T^2）之上；所有曲线均基于相同数量的样本（即 86.4% n）得到。此外，可以观察到蓝色曲线（SG）略高于灰色曲线（CG）；这是因为符号对称性施加的约束多于中心对称性（注意，Epanechnikov 分布同时满足这两种对称性假设）。

对于（b）中的平移高斯示例，我们知道，当使用高斯 ERD 时，渐近相对效率 ARE($\overline{\mathbf{W}}_n$,$\mathbf{X}_n$) ≥ 1 的下界可以达到（见定理 3.5 和 3.6）。这也可以从图 4 的右图中看出，其中黄色、深蓝色、灰色和蓝色四条曲线彼此重合。

我们的方法采用中心对称性（CG）、符号对称性（SG）或球对称性（SpG）中的任一种时，其检验功效几乎与使用相同样本数量的 Hotelling T^2 检验相同。如果我们的方法采用均匀 ERD，那么 ARE($\mathbf{W}_n$,$\mathbf{X}_n$) $= 3/\pi \approx 0.95$。因此，同样从图 4的右图可以看出，我们的方法采用均匀 ERD，并使用中心对称性（CU）或符号对称性（SU）以及 n 个样本时，其检验功效几乎与使用 T^2 个样本的 Hotelling $0.95n$检验相同。

4.2 功效比较

在本小节中，我们将方法与文献中用于检验不同类型对称性的现有程序进行比较。我们将使用推荐的具有恒等评分函数的高斯 ERD（见注3.3）。除了像前文一样通过从 ν 中随机抽取 $\mathbf{h}_i$'s 获得的 ν_n 外，我们还将使用准蒙特卡罗序列（ Halton 变换序列）来构造 ν_n，具体如下所述。Halton 序列 [48] $\{\mathrm{U}_i^p\}_{i\geq 1}$（分别为 $\{U_i^1\}_{i\geq 1}$）提供了 Uniform(0, 1)^p（分别为 Uniform(0, 1)）的离散化。令 $F_Z(\cdot)$ 为 $Z \sim N(0,1)$ 的分布函数， $F_{|Z|}(\cdot)$ 为 $|Z| \sim |N(0,1)|$ 的分布函数，并令 $F_{\|\mathbf{Z}\|}$为 $\|\mathrm{Z}\| \sim \|N(0, I_p)\|$ 的分布函数。对于中心对称性、符号对称性和球对称性，Halton 变换后的 ν_n's 分别由以下离散均匀分布给出：对于 $i = 1,\ldots,n$，在 $\left(F_{|Z|}^{-1}(\mathrm{U}_i^p,1)\ F_Z^{-1}(\mathrm{U}_i^p,2)\ldots,F_Z^{-1}(\mathrm{U}_i^p,d)\right)$ 上的离散均匀分布；对于 $i = 1,\ldots,n$，在$\left(F_{|Z|}^{-1}(\mathrm{U}_i^p,1)\ F_{|Z|}^{-1}(\mathrm{U}_i^p,2)\ldots,F_{|Z|}^{-1}(\mathrm{U}_i^p,p)\right)$ 上的离散均匀分布；以及对于 $i = 1,\ldots,n$，在$\left(F_{\|}^{-1}\|\mathrm{Z}\ (U_i^1)\ 0,\ldots,0\right)$ 上的离散均匀分布。

对于低维情形（即 $p = 2$），我们将对所用方法和 Hotelling 的 T^2 使用渐近 χ^2 临界值。在高维情形（即 $p = 50$），我们将使用 T^2 的精确分布（假设具有高斯性）以及我们的统计量（其分布无关）。具体而言，对于 GWSR 检验，给定参考点 $\mathbf{h}_1,\ldots,\mathbf{h}_n$，我们从 $\mathbf{W}_n$ 的零分布中抽取 1000 个观测，即 $\frac{1}{\sqrt{n}}\sum_{i=1}^n S_i\mathbf{h}_i$，其中 S_i 独立同分布于 Uniform($\mathcal{G}$)；若观测到的 $\|\mathbf{W}_n\|^2$ 值大于零分布中 95% 的观测值，则拒绝原假设。

我们考虑以下数据分布，以检验中心对称性（C1-C10）、符号对称性（S1-S10）和球对称性（Sp1-Sp10）：

C1. 多元高斯分布：$\mathbf{X} \sim N\left(\mathbf{0},\ \begin{pmatrix}2 & 1\\ 1 & 3\end{pmatrix}\right) + \lambda\cdot\mathbf{1}_2$； $n = 200$。C2. 多元 t分布，自由度为 1，位置参数为 $\lambda\cdot\mathbf{1}_2$，尺度参数为 $\begin{pmatrix}2 & 1\\ 1 & 3\end{pmatrix}$； $n = 200$。C3. 在 $[-1,0]^2\cup[0,1]^2$上的均匀分布，另加 $\lambda\cdot\mathbf{1}_2$； $n = 200$。C4. 指数分布： $n = 100$、 $p = 2$、 X_1， X_2 独立同分布地服从 Exp(1) -1。$\mathbf{X}$ 的均值为 $\mathbf{0}_p$。C5. 指数分布（较大样本）：除 $n = 200$ 外，与上述相同。

C6. 卡方分布： $n = 100$、 $p = 2$、 X_1、 X_2 独立同分布地服从 $\chi^2(1) - 1$。X 的均值为 $\mathbf{0}_p$。C7. 帕累托分布： $n = 200$、 $p = 2$、 X_1、 X_2 独立同分布地服从 Pareto(1) − 2。X 的中位数为 $\mathbf{0}_p$。C8. H_0下的 AR(1)： $n = 200$、 $p = 50$、 $X_1 = Z_1$、 $X_i = 0.5X_{i-1} + Z_i$、 $i = 2$、……、 p，其中 Z_i 独立同分布地服从 $N(0, 1)$。C9. H_1下的 AR(1)：除 X 的分布平移了 $0.15 \cdot \mathbf{1}_d$ 外，与上述相同。C10. AR(1)（H_1, 重尾）： $n = 200$、 $p = 50$、 Z_i 独立同分布地来自自由度为 1 的标准 t分布。$Y_1 = Z_1$、 $Y_i = 0.5Y_{i-1} + Z_i$、 $i = 2$、……、 p、 $X_i = Y_i + 0.9$、……、 p。S1. 多元高斯分布：X $\sim N\left(\mathbf{0}, \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}\right) + \lambda \cdot \mathbf{1}_2$； $n = 200$。S2. 多元 t分布，自由度为 1，位置参数为 $\lambda \cdot \mathbf{1}_2$，尺度参数为 $\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$； $n = 200$。S3. 在 $[-1, 1]^2$上的均匀分布，另加 $\lambda \cdot \mathbf{1}_2$； $n = 200$。S4. 相关性： $n = 200$、 $p = 2$， X $\sim N\left(\mathbf{0}, \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}\right)$。S5. 指数分布： $n = 100$、 $p = 2$、 X_1、 X_2 独立同分布地服从 Exp(1) − 1。S6. 卡方分布： $n = 100$、 $p = 2$、 X_1、 X_2 独立同分布地服从 $\chi^2(1) - 1$。S7. 拉普拉斯分布： $n = 200$、 $p = 2$、 X_1、 X_2 独立同分布地服从位置参数为 0.2、尺度参数为 1 的拉普拉斯分布。S8. 高维（H_0）： $n = 200$、 $p = 50$、 Z_1、……、 $Z_p \overset{i.i.d.}{\sim} N(0, 1)$， $X_i = \sin(i)Z_i$。S9. 高维（H_1）： $n = 200$、 $p = 50$、 Z_1、……、 $Z_p \overset{i.i.d.}{\sim} N(0, 1)$， $X_i = \sin(i)Z_i + 0.003$。S10. 高维（H_1, 重尾）： $n = 200$、 $p = 50$、 $X_i = \sin(i)Z_i + 0.003$；其中 Z 服从自由度为 1、位置参数为 $\mathbf{0}_p$ 、尺度参数为 I_p 的多元 t 分布。Sp1. 多元高斯分布：X $\sim N(0, I_2) + \lambda \cdot \mathbf{1}_2$； $n = 200$。Sp2. 多元 t分布，自由度为 1，位置参数为 $\lambda \cdot \mathbf{1}_2$，尺度参数为 I_2； $n = 200$。Sp3. 在单位圆盘 {x $\in \mathbb{R}^2$ 上的均匀分布： $\|\mathbf{x}\| \le 1\}$，另加 $\lambda \cdot \mathbf{1}_2$； $n = 200$。Sp4. 椭圆分布： $n = 200$、 $p = 2$， X $= (2Z_1, Z_2)^\top$，其中 Z_1、 Z_2 独立同分布地服从 $N(0, 1)$。Sp5. 相关性： $n = 200$、 $p = 2$， X $\sim N\left(\mathbf{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$，其中 $\rho = 0.6$。Sp6. 非球形均匀分布： $n = 1000$、 $p = 2$， X 服从$[-1, 1]^2$上的均匀分布。Sp7. 卡方分布： $n = 100$、 $p = 2$、 X_1、 X_2 独立同分布地服从 $\chi^2(1) - 1$。Sp8. 高维（H_0）： $n = 200$、 $p = 50$， X $\sim N(0, I_p)$。Sp9. 高维（H_1）： $n = 200$、 $p = 50$， X $\sim N(0, I_p) + 0.05 \cdot \mathbf{1}_p$。

Sp10. 高维（H_1, 重尾）： $n = 200$、 $p = 50$， $\mathbf{X}$ 服从具有 1 个自由度、位置参数为 $0.05 \cdot \mathbf{1}_p$ 、尺度参数为 I_p 的多元 t-分布。

对于中心对称性（C1-C10），我们将我们的 GWSR 检验（记为“OT-Wilcox”）、我们的广义符号检验（记为“OT-sign”）和 Hotelling 的 T^2 检验与 Dyckerhoff 等人提出的基于深度的检验 [29]（记为“DLP”），以及 Einmahl 和 Gan 提出的、比较相反区域经验测度的检验[32] （记为“EG”）进行比较。对于 EG，我们采用与 Einmahl 和 Gan 相同的参数设置 [32, 第 3]节。

对于符号对称性（S1-S10），我们将提出的方法——OT-Wilcox 和 OT-sign——与 Hotelling 的 T^2 检验、空间符号检验的符号变化版本（记为“SS”） [83, 第 6.1 节2]进行比较。

以及空间带符号秩检验（记为“SSR”） [83, 第 7 节。1]最后两种检验已在 R 软件包“MNM”中实现 [82]。

对于球对称性（Sp1-Sp10），我们将我们的方法与 Baringhaus [6] （记为“LB”）进行比较。LB 是一种针对多元球对称性的检验，对任何固定备择假设都具有一致性。对于 LB [6], , 需要指定一个函数 h 。当 $p = 2$ 时，我们采用 Baringhaus [6, 第 4], 节中的函数 $h(t) = \frac{t-1/4}{17/8-t}$ ，该函数可得到易于处理的渐近零分布。当 $p = 50$ 时，我们采用 Baringhaus [6, 第 5], 节中的函数 $h(t) = (1 - 2tw + w^2)^{-\lambda} - 1$ [6, 方程 3.10] , 其中 $\lambda = 24$ 和 $w = 1/4$, 并采用 Baringhaus [6, 方程 4.4] 所倡导的相同一阶近似来估算上尾概率，以计算临界值。我们还将我们的方法与 Henze 等人 [52] （记为“HHM”）进行比较。HHM 是一种基于经验特征函数的球对称性检验。我们采用 Henze 等人 [52, 第 5] 节建议的设置。[9]

所有假设检验程序的名义检验大小均设定为 $\alpha = 0.05$。表 2、3和 4 分别展示了不同方法在中心对称性、符号对称性和球对称性假设检验中的经验检验功效。对于我们的方法——OT-Wilcox 和 OT-sign—— 以及 Hotelling 的 T^2，经验检验功效是基于 10^4 次重复试验取平均值计算的。对于其他竞争方法，经验检验功效则基于 1000 次重复试验计算，因为其中一些方法的计算成本较高。

前三个示例（C1-C3、S1-S3、Sp1-Sp3）是位置偏移模型，这类备择情形下 OT-Wilcox 的检验功效较高。当 $\lambda = 0$ 时，这些分布分别具有中心对称性、符号对称性或球对称性，从表 2-4 可以看出，所有方法的检验大小都在 0.05 左右，符合预期。当 $\lambda > 0$ 时，在高斯情形（C1、S1、Sph1）下，Hotelling 的 T^2 检验表现最佳，而 OT-Wilcox 仅略逊一筹，这是命题 3.2 （或定理 3.5、3.6）的推论。在 t 分布示例（C2、S2、Sp2）中，Hotelling 的 T^2 检验几乎没有检验功效，而 OT-Wilcox 仍保持较高的检验功效。在 C2 和 S2 设置下，OT-sign 检验的检验功效甚至高于 OT-Wilcox。在轻尾均匀分布情形（设置 C3、S3、Sp3）下，Hotelling 的T^2 检验恢复了一些检验功效，但仍逊于 OT-Wilcox；在所有比较方法中，OT-Wilcox 的检验功效最高。

上述情形 C1-C3、S1-S3 和 Sp1-Sp3 表明，我们的 OT-Wilcox 是 Hotelling T^2 检验的一种非参数竞争方法，在检测对称分布的位置偏移时具有较高的检验功效。然而，对于非位置偏移备择，我们的方法可能检验功效较低，例如均值或中位数为 0 的偏斜备择。我们通过情形 C4-C7、S4-S6 和 Sph4-Sph7 进行说明：与 OT-sign 和大多数其他竞争方法相比，OT-Wilcox 的检验功效较低，但总体而言仍高于 Hotelling T^2 检验。请注意，OT-Wilcox 在情形 S4 和 Sp4-Sp6 下检验功效为零，这或许可以由定理 3.2解释：在检验符号对称性和球对称性时，对于中心对称备择，OT-Wilcox 不具一致性，因为 $\mathbb{E}[\nabla\psi(\mathbf{X})] = \mathbf{0}$ (其中 $J(\mathbf{x}) = \mathbf{x}$)，这是 $\nabla\psi(-\mathbf{x}) = -\nabla\psi(\mathbf{x})$ 的结果。

C8、S8 和 Sp8 对应于 H_0下的高维设定。可以看出，我们的方法和 Hotelling 检验 T^2 的检验大小都接近 0.05，符合预期。在 H_1 下（即设定 C9、S9、Sp9），OT-Wilcox 具有非平凡的检验功效，但与 Hotelling 检验 T^2 相比并不具竞争力（见设定 C9 和 S9 ）。这里的一个问题是，当维度 p 较高时，ν_n可能很难很好地逼近 ν 。例如，对于中心对称性， ν 的支撑由 2^{p-1} 个正交象限构成，但其中至多只有 n 个象限包含来自 ν_n的点。因此，从表2中的设定 C8 可以看出，基于随机抽样得到的ν_n 使 OT-Wilcox 的检验功效低于 T^2。使用基于准蒙特卡罗（Halton 变换）点得到的 ν_n 或许能改善这一情况，并显著提升 OT-Wilcox 的检验功效（见表2中的 C8）。然而，对于符号对称性（S8）、球对称性（Sp8），或任何低维示例，Halton 变换 ν_n 并未显著提升检验功效，因为这些情形下 ν 更容易逼近。在重尾高维设置下（即 C10、S10、Sp10），OT-Wilcox 的检验功效（和稳健性）明显优于 Hotelling 检验 T^2。

现在让我们比较上述其他方法的表现。请注意，DLP[29]和 EG [32] 是为检验双变量中心对称性而设计的，因此无法用于高维情形。LB [6] 在高维情形下毫无检验功效。[10] HHM [52] 实现了相对

[9]具体而言，我们使用带有超参数 $\mathcal{R} = 2$ 的柯尔莫哥洛夫–斯米尔诺夫型统计量。对于 $p = 2$，我们采用 8 个环和 9 个网格点，与 Henze 等人 [52, 第 5] 节中的设置相同。对于 $p = 50$，我们采用 8 个环和 200 个网格点）。[10]此时，上尾概率的一阶近似已不再准确—— [6, 公式 4 中的无限乘积4], 仅在 p 较小时才便于实际计算，而当 $p = 50$ 时则会变得极其庞大。

表 2：名义水平为 0.05 时，各竞争方法对中心对称性进行假设检验的检验功效。对于基于最优传输的方法，采用 ν_n 的两种选择（随机抽样/[Halton 变换]）。

多元高斯分布					
C1	T^2	OT-Wilcox	OT-sign	DLP [29]	EG [32]
$\lambda = 0.0$	0.06	**0.05**/[0.04]	**0.06**/[0.06]	0.05	0.04
$\lambda = 0.1$	0.15	**0.15**/[0.13]	**0.11**/[0.11]	0.06	0.14
$\lambda = 0.2$	0.49	**0.46**/[0.42]	**0.29**/[0.29]	0.08	0.46
$\lambda = 0.3$	0.84	**0.82**/[0.80]	**0.55**/[0.54]	0.12	0.85
$\lambda = 0.4$	0.98	**0.98**/[0.97]	**0.78**/[0.77]	0.20	0.98
多变量 t分布					
C2	T^2	OT-Wilcox	OT-sign	DLP [29]	EG [32]
$\lambda = 0.0$	0.03	**0.06**/[0.05]	**0.05**/[0.05]	0.05	0.04
$\lambda = 0.2$	0.03	**0.15**/[0.13]	**0.19**/[0.18]	0.07	0.21
$\lambda = 0.4$	0.04	**0.48**/[0.48]	**0.54**/[0.54]	0.12	0.62
$\lambda = 0.6$	0.08	**0.81**/[0.79]	**0.86**/[0.87]	0.26	0.94
$\lambda = 0.8$	0.12	**0.96**/[0.96]	**0.98**/[0.97]	0.49	1.00
多元均匀分布					
C3	T^2	OT-Wilcox	OT-sign	DLP [29]	EG [32]
$\lambda = 0.00$	0.05	**0.05**/[0.04]	**0.05**/[0.05]	0.05	0.05
$\lambda = 0.03$	0.10	**0.22**/[0.18]	**0.09**/[0.09]	0.08	0.13
$\lambda = 0.06$	0.27	**0.64**/[0.57]	**0.19**/[0.18]	0.15	0.45
$\lambda = 0.09$	0.55	**0.91**/[0.88]	**0.35**/[0.34]	0.24	0.80
$\lambda = 0.12$	0.81	**0.99**/[0.98]	**0.54**/[0.52]	0.37	0.97
其他备择情形					
C4-C7	T^2	OT-Wilcox	OT-sign	DLP [29]	EG [32]
指数分布	0.07	**0.10**/[0.09]	**0.63**/[0.66]	0.84	0.88
指数分布（L）	0.06	**0.15**/[0.15]	**0.91**/[0.93]	1.00	1.00
卡方分布	0.09	**0.21**/[0.19]	**0.83**/[0.88]	0.99	1.00
帕累托分布	0.84	**0.04**/[0.02]	**0.90**/[0.96]	1.00	1.00
高维设定					
C8-C10	T^2	OT-Wilcox	OT-sign	DLP [29]	EG [32]
AR(1) (H_0)	0.05	**0.05**/[0.05]	**0.04**/[0.04]	—	—
AR(1)（H_1）	0.97	**0.41**/[0.99]	**0.07**/[0.16]	—	—
AR(1)（H_1,重尾）	0.35	**0.24**/[0.89]	**0.06**/[0.23]	—	—

表3：名义水平为 0.05 时，用于检验符号对称性的各竞争方法的检验功效。对于基于最优传输的方法，采用了 ν_n 的两种选择（随机抽样/[Halton 变换]）。

多元高斯分布					
S1	T^2	OT-Wilcox	OT-sign	SSR [83]	SS [83]
$\lambda = 0.0$	0.06	**0.05**/[0.04]	**0.05**/[0.05]	0.05	0.05
$\lambda = 0.1$	0.21	**0.20**/[0.18]	**0.14**/[0.14]	0.19	0.16
$\lambda = 0.2$	0.64	**0.62**/[0.60]	**0.43**/[0.43]	0.61	0.52
$\lambda = 0.3$	0.94	**0.93**/[0.93]	**0.79**/[0.79]	0.94	0.87
$\lambda = 0.4$	1.00	**1.00**/[1.00]	**0.97**/[0.97]	1.00	0.99
多变量 t分布					
S2	T^2	OT-Wilcox	OT-sign	SSR [83]	SS [83]
$\lambda = 0.0$	0.03	**0.06**/[0.05]	**0.04**/[0.04]	0.05	0.05
$\lambda = 0.1$	0.04	**0.10**/[0.09]	**0.12**/[0.12]	0.09	0.12
$\lambda = 0.2$	0.05	**0.22**/[0.20]	**0.29**/[0.29]	0.24	0.34
$\lambda = 0.3$	0.06	**0.44**/[0.43]	**0.58**/[0.58]	0.48	0.66
$\lambda = 0.4$	0.06	**0.66**/[0.65]	**0.83**/[0.83]	0.72	0.89
多元均匀分布					
S3	T^2	OT-Wilcox	OT-sign	SSR [83]	SS [83]
$\lambda = .000$	0.05	**0.05**/[0.04]	**0.05**/[0.05]	0.06	0.05
$\lambda = .025$	0.12	**0.16**/[0.14]	**0.07**/[0.07]	0.10	0.08
$\lambda = .050$	0.33	**0.51**/[0.46]	**0.13**/[0.13]	0.29	0.18
$\lambda = .075$	0.65	**0.83**/[0.80]	**0.25**/[0.25]	0.58	0.37
$\lambda = .100$	0.89	**0.97**/[0.96]	**0.42**/[0.42]	0.83	0.60
其他备择情形					
S4–S7	T^2	OT-Wilcox	OT-sign	SSR [83]	SS [83]
相关性	0.06	**0.06**/[0.05]	**0.05**/[0.05]	0.05	0.05
指数分布	0.07	**0.14**/[0.12]	**0.93**/[0.93]	0.34	0.79
卡方分布	0.09	**0.30**/[0.28]	**1.00**/[1.00]	0.59	0.98
拉普拉斯分布	0.73	**0.81**/[0.79]	**0.92**/[0.92]	0.83	0.88
高维设定					
S8–S10	T^2	OT-Wilcox	OT-sign	SSR [83]	SS [83]
高维 (H_0)	0.04	**0.05**/[0.05]	**0.05**/[0.05]	0.05	0.05
高维 (H_1)	0.65	**0.30**/[0.28]	**0.49**/[0.49]	0.63	0.62
高维（H_1, 重尾）	0.05	**0.17**/[0.13]	**0.28**/[0.28]	0.27	0.35

表 4：在名义水平为 0.05 时，用于球对称性假设检验的各竞争方法的检验功效。对于基于最优传输的方法，采用了 ν_n 的两种选择（随机抽样/[Halton 变换]）。

多元高斯分布					
Sp1	T^2	OT-Wilcox	OT-sign	LB [6]	HHM [52]
$\lambda = 0.00$	0.06	0.05/[0.05]	0.05/[0.06]	0.05	0.05
$\lambda = 0.05$	0.15	0.14/[0.14]	0.07/[0.06]	0.08	0.10
$\lambda = 0.10$	0.42	0.42/[0.42]	0.12/[0.12]	0.18	0.26
$\lambda = 0.15$	0.77	0.76/[0.76]	0.21/[0.22]	0.38	0.54
$\lambda = 0.20$	0.95	0.95/[0.95]	0.36/[0.37]	0.62	0.84
多变量 t分布					
Sp2	T^2	OT-Wilcox	OT-sign	LB [6]	HHM [52]
$\lambda = 0.0$	0.03	0.07/[0.06]	0.05/[0.06]	0.04	0.05
$\lambda = 0.1$	0.05	0.16/[0.15]	0.09/[0.09]	0.21	0.19
$\lambda = 0.2$	0.06	0.46/[0.45]	0.20/[0.20]	0.75	0.60
$\lambda = 0.3$	0.07	0.79/[0.78]	0.39/[0.39]	0.97	0.94
$\lambda = 0.4$	0.10	0.96/[0.96]	0.62/[0.62]	1.00	0.99
多元均匀分布					
Sp3	T^2	OT-Wilcox	OT-sign	LB [6]	HHM [52]
$\lambda = .000$	0.06	0.05/[0.04]	0.05/[0.05]	0.04	0.05
$\lambda = .025$	0.14	0.21/[0.19]	0.06/[0.06]	0.04	0.09
$\lambda = .050$	0.40	0.61/[0.59]	0.09/[0.09]	0.05	0.25
$\lambda = .075$	0.75	0.90/[0.90]	0.14/[0.15]	0.07	0.55
$\lambda = .100$	0.96	0.99/[0.99]	0.23/[0.23]	0.11	0.83
其他备择情形					
Sp4-Sp7	T^2	OT-Wilcox	OT-sign	LB [6]	HHM [52]
椭圆分布	0.06	0.06/[0.06]	0.05/[0.05]	0.50	1.00
相关性	0.06	0.06/[0.06]	0.05/[0.05]	0.50	1.00
非球形均匀分布	0.05	0.05/[0.05]	0.05/[0.05]	0.07	0.05
卡方分布	0.09	0.27/[0.25]	0.45/[0.45]	1.00	1.00
高维设定					
Sp8-Sp10	T^2	OT-Wilcox	OT-sign	LB [6]	HHM [52]
高维 (H_0)	0.04	0.05/[0.05]	0.05/[0.05]	0.00	0.04
高维 (H_1)	0.54	0.68/[0.68]	0.05/[0.05]	0.00	0.15
高维 (H_1, 重尾)	0.05	0.35/[0.35]	0.05/[0.05]	0.00	0.11

在 Sp4、Sp5、Sp7 情形下表现相当不错，但在高维情形 Sp8-Sp10 下则不太理想。SSR [83] 和 SS [83] 在检验符号对称性方面表现相当出色，但在均匀分布发生位置偏移时（即 S3 情形），其检验功效甚至低于 Hotelling'sT^2。

在示例 S7 中，我们的广义符号检验相对于 Hotelling 的 T^2 渐近相对效率为 2（见表 1）；表 3 （设置 S7 ）显示，OT-sign 的检验功效确实高于 Hotelling 的 T^2。

5 分布无关置信集

我们称 $\mathbf{X} \sim \mu \in \mathcal{P}_{\text{ac}}(\mathbb{R}^p)$ 关于 $\boldsymbol{\theta}^* \in \mathbb{R}^p$ 具有 $\mathcal{G}$-对称性，如果对于所有$Q \in \mathcal{G}$都有 $\mathbf{X} - \boldsymbol{\theta}^* \overset{d}{=} Q(\mathbf{X} - \boldsymbol{\theta}^*)$。这里的 $\boldsymbol{\theta}^*$ 称为分布 μ 的对称中心。作为我们结果的直接推论，可以使用 GWSR 检验（或广义符号检验）来获得 $\boldsymbol{\theta}^*$ 的分布无关置信集。其思想只是对定义如下的假设集合 $\text{H}_{0,\boldsymbol{\theta}}$ 取逆：

$$\text{H}_{0,\boldsymbol{\theta}} : (\mathbf{X} - \boldsymbol{\theta}) \overset{d}{=} Q(\mathbf{X} - \boldsymbol{\theta}), \quad \text{for all } Q \in \mathcal{G},$$

对于 $\boldsymbol{\theta} \in \mathbb{R}^p$，使用基于数据的分布无关 GWSR 检验（见命题 2.2） $\{\mathbf{X}_i - \boldsymbol{\theta}\}_{i=1}^n$，检验水平为 $\alpha \in (0,1)$。由此可得一个关于 $1-\alpha$ 的精确（ $\boldsymbol{\theta}^*$）分布无关置信集：

$$\mathcal{C} := \{\boldsymbol{\theta} \in \mathbb{R}^p : \text{H}_{0,\boldsymbol{\theta}} \text{ is accepted using data } \{\mathbf{X}_i - \boldsymbol{\theta}\}_{i=1}^n\},$$

不对$\mathbf{X} \sim \mu \in \mathcal{P}_{\text{ac}}(\mathbb{R}^p)$作任何分布假设，唯一的假设是 $\mathcal{G}$-对称性。为便于实施，我们建议使用 $\boldsymbol{\theta}$ 的细网格来检验 $\text{H}_{0,\boldsymbol{\theta}}$ 并近似计算 $\mathcal{C}$。对于 $p \geq 2$，一种替代“网格化”的方法（网格化在计算上可能不可行）是检验 $\mathcal{G}$-对称性假设 $\text{H}_{0,\boldsymbol{\theta}}$ 对每个 $i = 1, \ldots, n$，在 $\boldsymbol{\theta} = \mathbf{X}_i$ 处进行检验，然后取被接受的 $\boldsymbol{\theta}$ 的凸包（该凸包可能更大）。

图 5 展示了分布无关的 95%置信集（以黄色标出；通过对 $\boldsymbol{\theta} \in \mathbb{R}^p$进行细网格划分得到），这些置信集是使用 GWSR 检验获得的，其中 $p = 2$ 采用高斯 ERD 和恒等评分函数，且：(i) $\mathcal{G} = \{-I, I\}$ （对应于中心对称性），以及 (ii) $\mathcal{G} = O(p)$（对应于球对称性）；关于我们对 3.3 的选择，见备注 ν （此处我们使用来自 ν 的随机样本来获得 ν_n，如第4.2节所述）。前两行展示了从高斯分布 $N(0.5 \cdot \mathbf{1}_2, I_2)$中抽取的数据点；第三行展示了从自由度为 1、位置参数为 $0.5 \cdot \mathbf{1}_2$、尺度参数为 I_2的多元t-分布中得到的数据点。每个示例中，真实的对称中心 $\boldsymbol{\theta}^*$ 用“+”标出。

可以看出，在所展示的情形中，我们基于 GWSR 检验构造的置信集都成功覆盖了真实的未知中心$\boldsymbol{\theta}^*$。此外，对于高斯数据的球对称性假设检验（见图5第一行），我们提出的基于 GWSR 检验的置信集与通过 Hotelling 的 T^2 检验得到的置信集几乎无法区分，即使样本量为 $n = 20$也是如此。对于高斯数据的中心对称性假设检验（见图5第二行），我们也观察到类似现象；不过，当 $n = 20$ 基于 Wilcoxon 符号秩检验构造的置信集边界更为参差不齐。

当高斯假设不成立且观测具有重尾特征时（见图5第三行），通过 Hotelling 的 T^2 检验得到的置信集会严重膨胀。不过，我们的分布无关置信集仍然相当小且可靠。

致谢

作者谨向 Nabarun Deb 致谢，感谢他提供的诸多有益讨论。

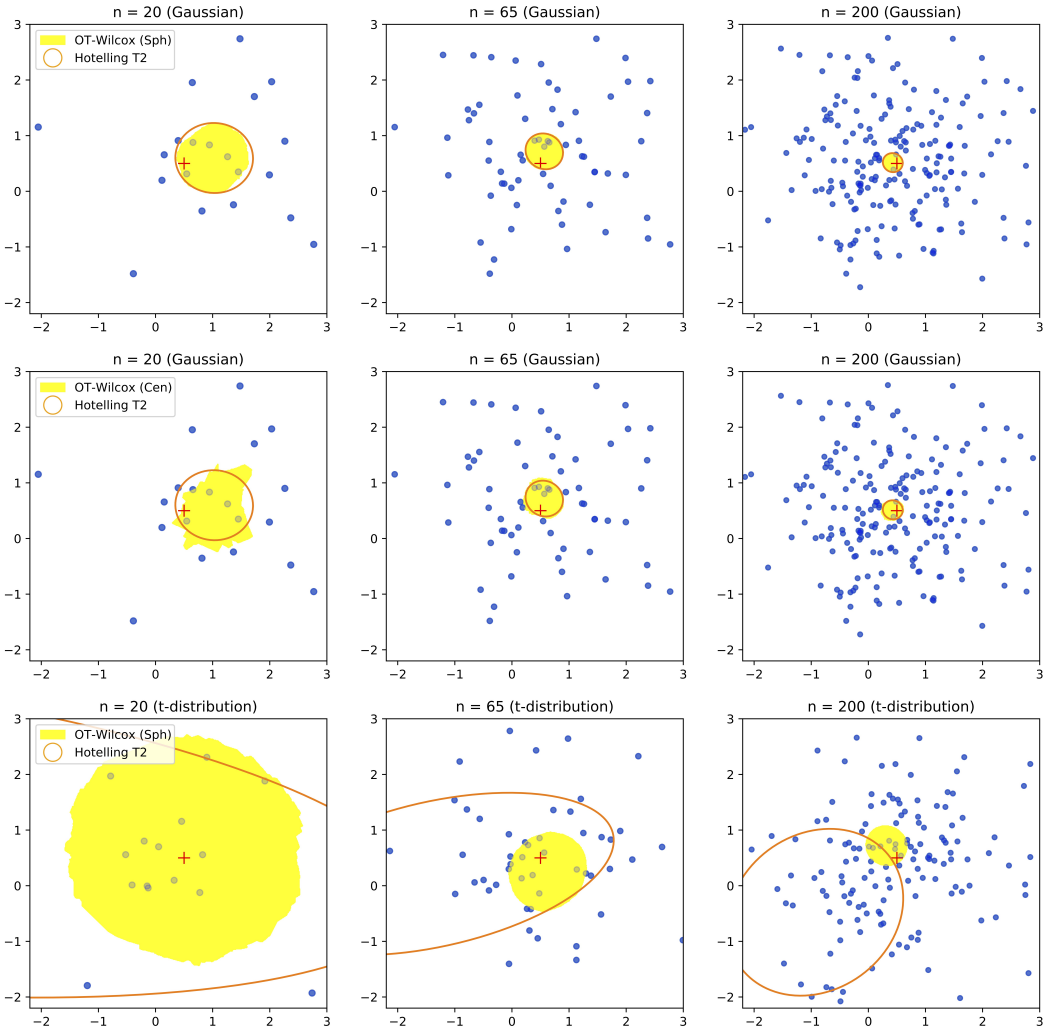


图5：通过反演 OT-Wilcox（用于中心对称性和球对称性）以及 Hotelling's T^2 检验得到的置信集。底层高斯分布的真实中心以“+”标记。数据分布为高斯分布（前两行）或自由度为 1 的 t-分布。

A 一些讨论

A.1 正则性条件

本节将介绍定理中所假设的正则性条件3.4，以研究针对局部连续备择的检验功效表现。我们采用保证局部渐近正态性（LAN）的标准光滑参数模型假设（例如，参见 Van der Vaart [99, 第 7]章）。考虑由对称分布 f的所有位置平移构成的参数族，即$\{\mathcal{P}_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta}\in\mathbb{R}^p} = \{f(\cdot-\boldsymbol{\theta}):\ \boldsymbol{\theta}\in\mathbb{R}^p\}$。假设：

1. 参数族 $\{\mathcal{P}_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta}\in\mathbb{R}^p}$ 在 $\boldsymbol{\theta}_0 = \mathbf{0}$ 处均方可微（QMD）（相关定义见 [68, 定义 12.2.1] ）。2. $\mathbb{E}_{\mathbf{X}\sim\mathcal{P}_{\boldsymbol{\theta}_0}}\left(\|\eta(\mathbf{X}\ \boldsymbol{\theta}_0)\|^2\right) < \infty$，其中 $\eta(\cdot,\ \boldsymbol{\theta}) :=\ ^{\nabla f}{}_{(\cdot-\boldsymbol{\theta})}\ {}_{f}{}_{(\cdot-\boldsymbol{\theta})}$ 为得分函数。另假设 $\boldsymbol{\theta}_0$处的 Fisher 信息（即 $I_{\boldsymbol{\theta}_0} := \mathbb{E}_{\mathbf{X}\sim\mathcal{P}_{\boldsymbol{\theta}_0}}\left[\eta(\mathbf{X}\ \boldsymbol{\theta})\ \eta(\mathbf{X}\ \boldsymbol{\theta})^{\top}\right]$）存在且可逆。

A.2 为球对称性生成随机符号

为证明3.1 节3.3 中的注释所述的过程会生成来自 Uniform(O(p)) 条件分布的样本（已知它将$\mathbf{v}$ 映射到$\mathbf{w}$），我们注意到，O(p) 上的均匀分布可以按以下方式生成 [30, 例 4。1]:

1. 首先从球面 $\{\mathbf{x}\in\mathbb{R}^p$ 上的均匀分布中抽取 $\omega:\ \|\mathbf{x}\| = 1\}$。2. 然后抽取 $p(p-1)$ 个独立同分布的标准正态随机变量 $\varepsilon\in\mathbb{R}^{p(p-1)}$，且它们与 ω 独立。3. 按列的顺序对 $[\omega\ \varepsilon]$ 的各列进行格拉姆-施密特正交化。由此得到正交矩阵 GS($[\omega\ \varepsilon]$)，它服从 O(p) 上的均匀分布 [30,例 4.1]。4. 定义 $f(\omega,\ \varepsilon):= \text{GS}([\omega\ \varepsilon])[\mathbf{v}\ V]^{\top}$。根据哈尔测度的定义（在群作用下不变），该矩阵也服从 O(p) 上的均匀分布（因为$[\mathbf{v}\ V]\in$ O(p)）。

注意，集合 $\operatorname{argmin}_{Q\in\mathcal{G}}\|Q^{\top}\mathbf{x} - \mathbf{h}\|^2$ 由 $\{Q\in\text{O}(p):\mathbf{w} = Q\mathbf{v}\}$描述。如果 $f(\omega,\varepsilon)\mathbf{v} = \mathbf{w}$，我们知道该条件等价于 GS($[\omega\ \varepsilon]$)$\mathbf{e}_1 = \mathbf{w}$，其中 $\mathbf{e}_1 = (1,0,\ldots,0)^{\top}$，且进一步等价于 $\omega = \mathbf{w}$。因此，根据给定 $\omega = \mathbf{w}$ 时 ω 与 ε 的独立性，$f(\omega,\varepsilon)$ 的条件分布等于 $f(\mathbf{w}\ \varepsilon)$ 的无条件分布。由此可知，该过程是有效的。 □

生成条件分布的第二种方法。来自

$$\{Q\in\text{O}(p):\ \mathbf{w} = Q\mathbf{v}\} = \{\mathbf{w}\mathbf{v}^{\top} + WUV^{\top}:\ U\in\text{O}(p-1)\},$$

其中 V 和 W 是 $p\times(p-1)$ 矩阵，满足 $V^{\top}V = W^{\top}W = I_{p-1}$, $V^{\top}\mathbf{v} = W^{\top}\mathbf{w} = \mathbf{0}$。我们可以先从 O($p-1$) 生成 U，然后 $\mathbf{w}\mathbf{v}^{\top} + WUV^{\top}$ 具有所需的条件分布。该过程的有效性可如下证明:

1. 抽取 p^2 个独立同分布的标准正态变量 $\varepsilon\in\mathbb{R}^{p\times p}$，组成 $p\times p$ 矩阵 $[\varepsilon_{ij}]$。2. 按列的顺序对 $[\varepsilon_{ij}]$ 的列应用格拉姆-施密特正交化。这会得到一个在 O(p) 上服从均匀分布的正交矩阵 GS($[\varepsilon_{ij}]$) [30, 示例 4.1]。3. 定义 $f(\varepsilon) := [\mathbf{w}\ W]\text{GS}([\varepsilon_{ij}])[\mathbf{v}\ V]^{\top}$，根据哈尔测度的不变性，该矩阵同样在 O($p$) 上服从均匀分布。

如果 $f(\varepsilon)\mathbf{v} = \mathbf{w}$，我们知道，该条件等价于 GS($[\varepsilon_{ij}]$) 的第一列为 $\mathbf{e}_1$。根据格拉姆-施密特过程，这又等价于事件$\{\varepsilon_{j1} = 0,$，其中 $j = 2,3,\ldots,p,$ 以及 $\varepsilon_{11} > 0\}$，而这仅取决于 $[\varepsilon_{ij}]$ 的第一列。根据独立性，在给定该条件时，从第二列到最后一列的其他 ε_{ij} 仍然是独立同分布的标准正态分布。因此，这意味着 GS($[\varepsilon_{ij}]$) 的右下角 $(p-1)\times(p-1)$ 子矩阵（记为

由 U) 条件性地服从 O($p-1$) 上的均匀分布，并且 GS($[\varepsilon_{ij}]$) $= \begin{pmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & U \end{pmatrix}$. 因此，给定 $\{f(\varepsilon)\mathbf{v} = \mathbf{w}\}$，均匀分布 $f(\varepsilon)$ 的条件分布具有 $[\mathbf{w}\ W] \begin{pmatrix} 1 & \mathbf{0}^\top \\ \mathbf{0} & U \end{pmatrix} [\mathbf{v}\ V]^\top = \mathbf{w}\mathbf{v}^\top + WUV^\top$ 的分布，其中 $U \sim$服从 Uniform(O($p-1$))。因此，第二个生成过程的有效性得以证明。

□

B 证明

B.1 命题 2.1 的证明

我们首先证明 $R_n(\cdot)$ 的唯一性。若唯一性不成立，则必然存在两个不同的（非随机）排列 σ_1 和 σ_2，使得 $\sum_{i=1}^n c(\mathbf{X}_i, \mathbf{h}_{\sigma_1(i)}) = \sum_{i=1}^n c(\mathbf{X}_i, \mathbf{h}_{\sigma_2(i)})$。不失一般性，假设 $\sigma_1(1) \neq \sigma_2(1)$。于是

$$c(\mathbf{X}_1, \mathbf{h}_{\sigma_1(1)}) - c(\mathbf{X}_1, \mathbf{h}_{\sigma_2(1)}) = \sum_{i=2}^{n} \left(c(\mathbf{X}_i, \mathbf{h}_{\sigma_2(i)}) - c(\mathbf{X}_i, \mathbf{h}_{\sigma_1(i)}) \right). \tag{B.1}$$

定义点 $\mathbf{x} \in \mathbb{R}^p$ 到集合 $A \subset \mathbb{R}^p$ 的距离为 $d(\mathbf{x}, A) := \inf_{\mathbf{y} \in A} \|\mathbf{x} - \mathbf{y}\|$。注意，$c(\mathbf{X}_1, \mathbf{h}_{\sigma_1(1)}) - c(\mathbf{X}_1, \mathbf{h}_{\sigma_2(1)}) = d(\mathbf{X}1, \mathcal{G}\mathbf{h}_{\sigma_1(1)})^2 - d(\mathbf{X}1, \mathcal{G}\mathbf{h}_{\sigma_2(1)})^2$。下面的引理表明，只要轨道 $\mathcal{G}\mathbf{h}_{\sigma_1(1)} \neq \mathcal{G}\mathbf{h}_{\sigma_2(1)}$ 不同，

$$\{\mathbf{x} \in \mathbb{R}^p : d(\mathbf{x}, \mathcal{G}\mathbf{h}_{\sigma_1(1)})^2 - d(\mathbf{x}, \mathcal{G}\mathbf{h}_{\sigma_2(1)})^2 = C\}$$

对于任意常数 C，其勒贝格测度为 0。考虑给定 $\mathbf{X}_2, \ldots, \mathbf{X}_n$ 时 $\mathbf{X}_1$ 的条件分布，可知 (B.1) 以概率 0 成立。因此，$\widehat{\sigma}$，从而 $R_n(\cdot)$ 是唯一的。

引理 B.1. 设 $A, B \subset \mathbb{R}^p$ 是两个不相交的勒贝格测度为 *0* 的紧集，且 C 为常数。那么

$$D := \{\mathbf{x} \in \mathbb{R}^p : d(\mathbf{x}, A)^2 - d(\mathbf{x}, B)^2 = C\}$$

具有勒贝格测度 *0*。

证明。 众所周知，距离函数 $f(\mathbf{x}) := d(\mathbf{x}, A)$ 是 1-利普希茨函数，因此几乎处处（Lebesgue 意义下）可微（根据拉德马赫定理）。如果 $\mathbf{x}_0 \notin A$ 是 f 的可微点，那么它在 A 上的投影必定唯一[11]，，否则就会存在两个方向，使得 f 沿这两个方向以最快速度（速率为 1）下降，这与 f 的可微性矛盾。此外，$\nabla f(\mathbf{x}_0)$ 是连接 $\mathbf{x}_0$ 与其在 A 上的投影的直线上的一个单位向量。

现在，取$\mathbf{x}_0 \in D$ 为一个不属于 A 或 B的点，且同时是$d(\mathbf{x}, A)$和 $d(\mathbf{x}, B)$的可微点。于是，对于$\mathbf{x}$ 在$\mathbf{x}_0$, 的某个邻域内，利用可微性的定义，$d(\mathbf{x}, A)^2 - d(\mathbf{x}, B)^2 = C$ 可写为:

$$2\left[d(\mathbf{x}_0, A)\nabla d(\mathbf{x}_0, A) - d(\mathbf{x}_0, B)\nabla d(\mathbf{x}_0, B)\right]^\top (\mathbf{x} - \mathbf{x}_0) + o(\|\mathbf{x} - \mathbf{x}_0\|) = 0. \tag{B.2}$$

设 $\mathbf{f} := 2\left[d(\mathbf{x}0, A)\nabla d(\mathbf{x}0, A) - d(\mathbf{x}0, B)\nabla d(\mathbf{x}0, B)\right] \in \mathbb{R}^p$。那么 $\mathbf{f}$ 非零；否则，$\mathbf{x}_0$ 在 A 和 B 上的投影相同，这与 A、B 互不相交相矛盾。因此，(B.2)

[11]为使这一唯一性论证更加严谨，假设$\mathbf{y}_0$（先验而言不唯一）实现了$\mathbf{x}_0$到 A的最小距离，即 $f(\mathbf{x}_0) = \|\mathbf{x}_0 - \mathbf{y}_0\|$。那么$\mathbf{y}_0$也实现了 A到线段[$\mathbf{x}_0$,$\mathbf{y}_0$]上任意一点的最小距离。这是因为，对于任意$\mathbf{y}' \in A$、$t \in [0,1]$以及$\mathbf{z} := \mathbf{x}_0 + t(\mathbf{y}_0 - \mathbf{x}_0)$，根据三角不等式，我们有:

$$\|\mathbf{z} - \mathbf{y}'\| \geq \|\mathbf{x}_0 - \mathbf{y}'\| - \|\mathbf{x}_0 - \mathbf{z}\| \geq f(\mathbf{x}_0) - \|t(\mathbf{y}_0 - \mathbf{x}_0)\| = (1-t)\|\mathbf{y}_0 - \mathbf{x}_0\| = \|\mathbf{z} - \mathbf{y}_0\|.$$

因此，$f(\mathbf{z}) = \|\mathbf{z} - \mathbf{y}_0\|$; 进一步，$f(\mathbf{z}) = f(\mathbf{x}_0) - (\mathbf{z} - \mathbf{x}_0)^\top \mathbf{y}\frac{0-0}{\|0-0\|}\mathbf{x}\,\mathbf{y}\,\mathbf{x}_0\|$。由于 f 在 $\mathbf{x}_0$, 处可微，且 $\mathbf{z}$ 足够接近 $\mathbf{x}_0$, ，我们有 $f(\mathbf{z}) = f(\mathbf{x}_0) + (\mathbf{z} - \mathbf{x}_0)^\top \nabla f(\mathbf{x}_0) + o(\|\mathbf{z} - \mathbf{x}_0\|)$，由此进一步得到 $\nabla f(\mathbf{x}_0) = -\mathbf{y}\frac{0-0}{\|0-0\|}\mathbf{x}\,\mathbf{y}\,\mathbf{x}_0\|$。因此，$\mathbf{y}_0$ 由 $\mathbf{x}_0$ 和梯度 $\nabla f(\mathbf{x}_0)$ 唯一确定，即 $\mathbf{y}_0 = \mathbf{x}_0 - f(\mathbf{x}_0)\nabla f(\mathbf{x}_0)$，故 $\mathbf{y}_0$ 是唯一的。

意味着集合 D 局部包含在一个厚度可任意小的超平面中；也就是说，对于任意$\delta > 0$，都存在 $\varepsilon_0 > 0$，使得对于所有 $\varepsilon < \varepsilon_0$,

$$D \cap B(\mathbf{x}_0, \varepsilon) \subset \{\mathbf{x} \in B(\mathbf{x}_0, \varepsilon) : |\mathbf{f}^\top(\mathbf{x} - \mathbf{x}_0)| \leq \delta \|\mathbf{x} - \mathbf{x}_0\|\}.$$

这里 $B(\mathbf{x}_0, \varepsilon) := \{\mathbf{x} \in \mathbb{R}^p : \|\mathbf{x} - \mathbf{x}_0\| < \varepsilon\}$ 是以 $\mathbf{x}_0$ 为中心、半径为 $\varepsilon > 0$ 的开球。因此，记 m 为勒贝格测度,

$$\limsup_{\varepsilon \to 0} \frac{m(D \cap B(\mathbf{x}_0, \varepsilon))}{m(B(\mathbf{x}_0, \varepsilon))} \leq \limsup_{\varepsilon \to 0} \frac{m(\{\mathbf{x} \in B(\mathbf{x}_0, \varepsilon) : |\mathbf{f}^\top(\mathbf{x} - \mathbf{x}_0)| \leq \delta \|\mathbf{x} - \mathbf{x}_0\|\})}{m(B(\mathbf{x}_0, \varepsilon))}.$$

通过进行平移和缩放，我们观察到

$$\frac{m(\{\mathbf{x} \in B(\mathbf{x}_0, \varepsilon) : |\mathbf{f}^\top(\mathbf{x} - \mathbf{x}_0)| \leq \delta \|\mathbf{x} - \mathbf{x}_0\|\})}{m(B(\mathbf{x}_0, \varepsilon))} = \frac{m(\{\mathbf{x} \in B(\mathbf{0}, 1) : |\mathbf{f}^\top \mathbf{x}| \leq \delta \|\mathbf{x}\|\})}{m(B(\mathbf{0}, 1))}$$

对所有 $\mathbf{x}_0 \in \mathbb{R}^p$ 和 $\varepsilon > 0$，均有。因此,

$$\limsup_{\varepsilon \to 0} \frac{m(D \cap B(\mathbf{x}_0, \varepsilon))}{m(B(\mathbf{x}_0, \varepsilon))} \leq \frac{m(\{\mathbf{x} \in B(\mathbf{0}, 1) : |\mathbf{f}^\top \mathbf{x}| \leq \delta \|\mathbf{x}\|\})}{m(B(\mathbf{0}, 1))}.$$

由于 δ 是任意的（并且当 δ 趋近于 0 时，上述显示右侧的集合收敛到超平面的一个子集），我们进一步得到

$$\lim_{\varepsilon \to 0} \frac{m(D \cap B(\mathbf{x}_0, \varepsilon))}{m(B(\mathbf{x}_0, \varepsilon))} = 0.$$

但根据勒贝格微分定理，对于 D中的几乎处处（Lebesgue 意义下）的 $\mathbf{x}_0$，

$$\lim_{\varepsilon \to 0} \frac{m(D \cap B(\mathbf{x}_0, \varepsilon))}{m(B(\mathbf{x}_0, \varepsilon))} = 1_D(\mathbf{x}_0) = 1.$$

比较最后两个显示可知， D 的勒贝格测度必为 0。 □

接下来，我们证明 $S_n(\cdot)$ 的唯一性。由于 $R_n(\mathbf{X}_1)$ 必须是 $\mathbf{h}_1, \ldots, \mathbf{h}_n$ 中的一个，如果有两个元素 $Q_1 \neq Q_2 \in \mathcal{G}$ 属于 $\operatorname{argmin}_{Q \in \mathcal{G}} \|Q^\top \mathbf{X}_1 - R_n(\mathbf{X}_1)\|^2$，那么必然存在 $\mathbf{h} \in \{\mathbf{h}_1, \ldots, \mathbf{h}_n\}$，使得

$$\|Q_1^\top \mathbf{X}_1 - \mathbf{h}\|^2 = \|Q_2^\top \mathbf{X}_1 - \mathbf{h}\|^2, \text{ or equivalently } \langle \mathbf{X}_1, Q_1\mathbf{h} - Q_2\mathbf{h} \rangle = 0.$$

自由群作用意味着 $Q_1\mathbf{h} - Q_2\mathbf{h} \neq \mathbf{0}$。因此，上述事件发生的概率为 0（因为 $\mathbf{X}_1$ 绝对连续）。

最后，我们将在下文证明带符号秩的唯一性。我们将使用以下引理。

引理 B.2. 设 $C \subset \mathbb{R}^p$ 是一个紧集。那么对于勒贝格几乎处处的点，它在 C 中都有唯一的最近点。

证明。 众所周知，到 C的距离，即 $f(\mathbf{x}) := d(\mathbf{x}, C)$，是 1-利普希茨函数，因此几乎处处勒贝格可微（根据拉德马赫定理）。对于任意一个属于 $\mathbf{x}_0$ 且不属于 f 的可微点 C，其在 C 上的投影必定唯一（参见引理 B.1的证明）。当 $\mathbf{x}_0 \in C$时，投影就是 $\mathbf{x}_0$ 本身，因此其唯一性显然成立。 □

由于 $S_n(\mathbf{X}_i) = \operatorname{argmin}_{Q \in \mathcal{G}} \|\mathbf{X}_i - QR_n(\mathbf{X}_i)\|^2$, $S_n(\mathbf{X}_i)R_n(\mathbf{X}_i)$ 是 $R_n(\mathbf{X}_i)$ 的轨道中距离 $\mathbf{X}_i$ 最近的点。但 $R_n(\mathbf{X}_i)$ 必定是 $\mathbf{h}1, \ldots, \mathbf{h}_n$ 中的一个。如果 $S_n(\mathbf{X}_i)R_n(\mathbf{X}_i)$ 不唯一，那么存在 $\mathbf{h} \in \{\mathbf{h}1, \ldots, \mathbf{h}_n\}$，使得 $\mathbf{X}_i$ 在 $\mathcal{G}\mathbf{h}$ 上的投影不唯一。然而，根据引理 B.2，对于几乎处处（Lebesgue 意义下）的 $\mathbf{x} \in \mathbb{R}^p$，$\mathbf{x}$ 在 $\mathcal{G}\mathbf{h}$ 中都有唯一的最近点。由于 $\mathbf{X}_i$ 绝对连续， $S_n(\mathbf{X}_i)R_n(\mathbf{X}_i)$ 几乎必然唯一。 □

B.2 命题 2.2 的证明

由于 $\mathbf{X}_1, \ldots, \mathbf{X}_n$ 是独立同分布的，使目标函数达到最小值的 $\widehat{\sigma}$

$$\sum_{i=1}^{n} c(\mathbf{X}_{\sigma(i)}, \mathbf{h}_i), \quad \text{where } c(\mathbf{x}, \mathbf{h}) = \min_{Q \in \mathcal{G}} \|Q^\top \mathbf{x} - \mathbf{h}\|^2,$$

取 $n!$ 个排列中的任意一个的可能性相同 [27]。因此，第一个结论成立。

为证明另外两个性质，假设 H_0 成立。我们考虑由 n 轨道 $\mathcal{G}\mathbf{X}_1$、.....、$\mathcal{G}\mathbf{X}_n$生成的 σ-域，以及为选择 $\widehat{\sigma}$而引入的可能随机性（如果 $\widehat{\sigma}$不唯一）。于是给定 $\mathcal{F}$后， $\widehat{\sigma}$ 是确定的，即 $\widehat{\sigma}$ 关于 $\mathcal{F}$可测。

给定 $\mathcal{F}$, $\widehat{Q}_i$ （见（2.2））定义为

$$\widehat{Q}_i = \operatorname*{argmin}_{S_i \in \mathcal{G}} \|S_i^\top \mathbf{X}_{\widehat{\sigma}(i)} - \mathbf{h}_i\|^2. \tag{B.3}$$

根据引理 B.3（见下文），当 X 具有 $\mathcal{G}$-对称分布时，X 的密度 f 满足$f(\mathrm{x}) = f(\mathrm{y})$，其中 x 和 y 位于 $\mathcal{G}$的同一轨道上。注意，给定 $\mathcal{F}$后，我们只知道 $\mathbf{X}_{\widehat{\sigma}(i)}$的轨道；由于 $\mathcal{G}$-对称性，$\mathbf{X}_{\widehat{\sigma}(i)}$ 在该轨道中的任意点出现的可能性相同，即 $\mathbf{X}_{\widehat{\sigma}(i)}$在条件 $\mathcal{F}$下服从 $\mathcal{G}$-轨道 $\mathcal{G}\mathbf{X}_{\widehat{\sigma}(i)}$上的均匀分布。对于任意 $Q \in \mathcal{G}$，定义$\widetilde{Q}_i$ 为

$$\widetilde{Q}_i := \operatorname*{argmin}_{S_i \in \mathcal{G}} \|S_i^\top Q^\top \mathbf{X}_{\widehat{\sigma}(i)} - \mathbf{h}_i\|^2. \tag{B.4}$$

如果上述极小化解不唯一，则从限制在 (B.4) 的极小化点集合上的均匀分布 Uniform($\mathcal{G}$) 中选取 $\widetilde{Q}_i$。由于 $\mathbf{X}_{\widehat{\sigma}(i)}$ 和 $Q^\top \mathbf{X}_{\widehat{\sigma}(i)}$ 在给定 $\mathcal{F}$时具有相同的条件分布，因此我们知道 $\widetilde{Q}_i$ 与 Q_i具有相同的条件分布。另一方面，如果将 (B.4) 中的 $S_i^\top Q^\top$ 视为一个整体，那么 (B.4) 就化简为 (B.3) 中的极小化问题，因此$Q\widetilde{Q}_i$ 与 $\widehat{Q}_i$具有相同的条件分布。因此，给定 $\mathcal{F}$时， $Q\widehat{Q}_i \overset{d}{=} \widehat{Q}_i$。这意味着，给定 $\mathcal{F}$，$\widehat{Q}_i$ 独立同分布，且服从 $\mathcal{G}$上的均匀分布。因此， $\{Q_i\}_{i=1}^n$ 与 $\mathcal{F}$独立，进而与 $\widehat{\sigma}$独立。因此，第二和第三个结论成立。

□

B.3 定理 3.1的证明

由于 $(S_n(\mathbf{X}_1) \ldots, S_n(\mathbf{X}_n))$ 与 $(R_n(\mathbf{X}_1) \ldots, R_n(\mathbf{X}_n))$ 相互独立，并且 $(R_n(\mathbf{X}_1) \ldots, R_n(\mathbf{X}_n))$ 是 $(\mathbf{h}_1, \ldots, \mathbf{h}_n)$ 的一个置换版本，因此 $\mathbf{W}_n$ 与其具有相同的分布

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_n(\mathbf{X}_i) J(\mathbf{h}_i),$$

根据假设 5-(i)，其均值为 0。我们首先证明协方差矩阵$\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[S_n(\mathbf{X}_i)J(\mathbf{h}_i)J(\mathbf{h}_i)^\top S_n(\mathbf{X}_i)^\top\right]$的收敛。注意， $f(\mathrm{x}) := \mathbb{E}_S\left[S\mathrm{x}\mathrm{x}^\top S^\top\right]$，其中 $S \sim$Uniform($\mathcal{G}$)，关于 x 连续（应用控制收敛定理即可得到）。于是，协方差矩阵可以写成 $\mathbb{E}f(J(\mathrm{H}_n))$，其中 $\mathrm{H}_n \sim \frac{1}{n}\sum_{j=1}^n \delta_{\mathbf{h}_j}$弱收敛于 $\mathrm{H} \sim \nu$。

为了证明 $\mathbb{E}f(J(\mathrm{H}_n)) \to \mathbb{E}f(J(\mathrm{H}))$（当 $n \to \infty$时），根据表示定理，我们可以假设 $\mathrm{H}_n \xrightarrow{a.s.} \mathrm{H}$（由假设 2 得出）。于是，由于 $J(\cdot)$ 几乎处处连续，根据连续映射定理， $J(\mathrm{H}_n) \xrightarrow{a.s.} J(\mathrm{H})$。结合 $\mathbb{E}|J(\mathrm{H}_n)_i J(\mathrm{H}_n)_j| \to \mathbb{E}|J(\mathrm{H})_i J(\mathrm{H})_j|$（由假设 4-($iii$) 得出），根据 Scheff´e 引理可得 $J(\mathrm{H}_n)J(\mathrm{H}_n)^\top \xrightarrow{L^1} J(\mathrm{H})J(\mathrm{H})^\top$。注意，$\|f(\mathrm{x})\|_F = \|\mathbb{E}_S\left[S\mathrm{x}\mathrm{x}^\top S^\top\right]\|_F \le \mathbb{E}_S[\|S\mathrm{x}\mathrm{x}^\top S^\top\|_F] = \mathbb{E}_S[\|\mathrm{x}\mathrm{x}^\top\|_F] = \|\mathrm{x}\mathrm{x}^\top\|_F$，其中用到了 Frobenius范数的凸性，以及对于正交的 S 和任意矩阵 A，有 $\|AS\|_F = \|SA\|_F = \|A\|_F$。因此，$\|f(J(\mathrm{H}_n))\|_F \le \|J(\mathrm{H}_n)J(\mathrm{H}_n)^\top\|_F$。结合 $J(\mathrm{H}_n)J(\mathrm{H}_n)^\top$ 在 L^1 中收敛这一事实，我们得到 $\|f(J(\mathrm{H}_n))\|_F$ 一致可积。由于 $f(J(\mathrm{H}_n)) \xrightarrow{a.s.} f(J(\mathrm{H}))$，进一步可得 $\mathbb{E}f(J(\mathrm{H}_n)) \to \mathbb{E}f(J(\mathrm{H}))$。

由于协方差收敛，我们只需检验林德伯格条件：

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\|S_n(\mathbf{X}_i)J(\mathbf{h}_i)\|^2 1_{\|S_n(\mathbf{X}_i)J(\mathbf{h}_i)\|\geq\varepsilon\sqrt{n}}\right]$$
$$=\frac{1}{n}\sum_{i=1}^{n}\|J(\mathbf{h}_i)\|^2 1_{\|J(\mathbf{h}_i)\|\geq\varepsilon\sqrt{n}}=\mathbb{E}\left[\|J(\mathbf{H}_n)\|^2 1_{\|J(\mathbf{H}_n)\|\geq\varepsilon\sqrt{n}}\right].$$

注意，由于 $J(\mathbf{H}_n)$依概率有界，$\mathbb{P}(\|J(\mathbf{H}_n)\|\geq\varepsilon\sqrt{n})\to 0$ 当 $n\to\infty$时，因此$\|J(\mathbf{H}_n)\|^2 1_{\|J(\mathbf{H}_n)\|\geq\varepsilon\sqrt{n}}$ 依概率收敛到 0。注意，$\|J(\mathbf{H}_n)\|^2$ 是一致可积的，因为 $J(\mathbf{H}_n)J(\mathbf{H}_n)^\top$ 在 L^1中收敛，因此 $\|J(\mathbf{H}_n)\|^2 1_{\|J(\mathbf{H}_n)\|\geq\varepsilon\sqrt{n}}$ 是一致可积的。于是，$\mathbb{E}\left[\|J(\mathbf{H}_n)\|^2 1_{\|J(\mathbf{H}_n)\|\geq\varepsilon\sqrt{n}}\right]\to 0$。因此，根据林德伯格-费勒中心极限定理，当 $n\to\infty$时，

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}S_n(\mathbf{X}_i)J(\mathbf{h}_i)\xrightarrow{d}N\left(\mathbf{0},\mathbb{E}f(J(\mathbf{H}))\right)\equiv N\left(\mathbf{0},\mathrm{Var}(SJ(\mathbf{H}))\right).$$

□

B.4 定理 2.1、2.2、3.2和3.3的证明

定义 B.1 (c-凸性 [101])。若函数 $\psi:\ \mathbb{R}^p\to\mathbb{R}\cup\{+\infty\}$ 满足，则称其为 c-凸函数；如果它不恒等于 $+\infty$，且存在 $\zeta:\ \mathbb{R}^p\to\mathbb{R}\cup\{\pm\infty\}$ 使得

$$\psi(\mathbf{x})=\sup_{\mathbf{y}\in\mathbb{R}^p}\left(\zeta(\mathbf{y})-c(\mathbf{x},\mathbf{y})\right),\quad\text{for all }\mathbf{x}\in\mathbb{R}^p.$$

其 c-变换是由下式定义的函数 ψ^c

$$\psi^c(\mathbf{y}):=\inf_{\mathbf{x}\in\mathbb{R}^p}\left(\psi(\mathbf{x})+c(\mathbf{x},\mathbf{y})\right),\quad\text{for all }\mathbf{y}\in\mathbb{R}^p.\tag{B.5}$$

其 c-次微分定义为

$$\partial_c\psi:=\left\{(\mathbf{x},\mathbf{y})\in\mathbb{R}^p\times\mathbb{R}^p:\psi^c(\mathbf{y})-\psi(\mathbf{x})=c(\mathbf{x},\mathbf{y})\right\}.$$

如果 $c(\cdot,\ \cdot)$ 是连续的，则 c-凸函数自动是下半连续（l.s.c.）且博雷尔可测 [101, 备注 5.5]。

我们考虑等价的代价函数 $c(\mathbf{x},\mathbf{y}):=-\max_{Q\in\mathcal{G}}\mathbf{x}^\top Q\mathbf{y}$。注意，按照我们对 c 的选择，$\psi(\mathbf{x})=\sup_{\mathbf{y},\ Q}\left(\zeta(\mathbf{y})+\mathbf{x}^\top Q\mathbf{y}\right)$ 是线性函数的上确界。这意味着，c-凸函数在通常意义下是凸的，因此在其定义域 [3] 上几乎处处勒贝格可微。还容易看出，任意 c-凸函数 ψ 都是 $\mathcal{G}$-不变的，即对于任意 $Q\in\mathcal{G}$，$\psi(Q\mathbf{x})=\psi(\mathbf{x})$；因此，$\nabla\psi(\cdot)$ 是 $\mathcal{G}$- 等变的，即当 $\mathbf{x}$ 和 $Q\mathbf{x}$ 是 ψ 的可微点时，$\nabla\psi(Q\mathbf{x})=Q\nabla\psi(\mathbf{x})$。

引理 B.3. 设$\mathbf{X}\sim\mu\in\mathcal{P}_{\mathrm{ac}}(\mathbb{R}^p)$，且 $\mathcal{G}$ 是正交群$\mathrm{O}(p)$的紧子群。设 S 服从 $\mathcal{G}$上的均匀分布，且与$\mathbf{X}$独立。则 $S\mathbf{X}\sim\mu_S$ 也是勒贝格绝对连续的，且 μ 相对于 μ_S绝对连续。

证明。假设$\mathbf{X}$ 具有勒贝格密度 $f(\cdot)$。下面证明 $S\mathbf{X}$ 具有如下勒贝格密度：对于$\mathbf{x}\in\mathbb{R}^p$，$g(\mathbf{x})=\mathbb{E}_S[f(S\mathbf{x})]$，，其中 $S\sim\mathrm{Uniform}(\mathcal{G})$。令 m 表示在 $\mathcal{G}$ 上的均匀分布（即 $\mathcal{G}$上的哈尔测度）。对于任意博雷尔集 $B\subset\mathbb{R}^p$，

$$\begin{aligned}
\mathbb{P}(S\mathbf{X}\in B)&=\int 1_{Q\mathbf{x}\in B}f(\mathbf{x})\mathrm{d}\mathbf{x}\,\mathrm{d}m(Q)\\
&=\int 1_{\mathbf{y}\in B}f(Q^{-1}\mathbf{y})\mathrm{d}\mathbf{y}\,\mathrm{d}m(Q)&&\text{(by a change of variable }\mathbf{y}=\mathrm{Q}\mathbf{x})\\
&=\int 1_{\mathbf{y}\in B}f(Q\mathbf{y})\mathrm{d}\mathbf{y}\,\mathrm{d}m(Q)&&\left(\text{as }S^{-1}\overset{d}{=}S\sim\mathrm{Uniform}(\mathcal{G})\right)\\
&=\int 1_{\mathbf{y}\in B}\mathbb{E}_S f(S\mathbf{y})\mathrm{d}\mathbf{y}.&&\text{(by Fubini's theorem)}
\end{aligned}$$

因此， $S\mathbf{X}$ 具有勒贝格密度 $g(\mathbf{x}) = \mathbb{E}_S f(S\mathbf{x})$，其中$\mathbf{x} \in \mathbb{R}^p$。

回顾一下，概率测度 μ 在 $\mathbb{R}^p$ 上的支撑是定义如下的闭集：

$$\operatorname{supp}(\mu) := \{\mathbf{x} \in \mathbb{R}^p : \forall \text{ open neighborhood } N_{\mathbf{x}} \text{ of } \mathbf{x}, \mu(N_{\mathbf{x}}) > 0\}.$$

若 μ 具有勒贝格密度 $f(\cdot)$，则在 $\operatorname{supp}(\mu)$ 上， $f(\mathbf{x})$ 几乎处处关于勒贝格测度成立；并且在 $\mathbb{R}^p \backslash \operatorname{supp}(\mu)$ 上，$f(\mathbf{x})$ 几乎处处关于勒贝格测度成立。

我们将证明，$\mathbf{X}$ 支撑中的每个点都属于 $S\mathbf{X}$ 的支撑，因此 $f_{(\cdot)}\ g_{(\cdot)}$ 定义良好（它是 ~~$\mathbf{X}$~~相对于 $S\mathbf{X}$的拉东-尼科迪姆导数）。由此即可确立 $\mathbf{X}$ 相对于 $S\mathbf{X}$的绝对连续性。

注意，每个元素 $Q \in \mathcal{G}$ 都属于 Uniform($\mathcal{G}$) 的支撑。[12] 特别地，单位矩阵 $I \equiv I_p$ 属于 Uniform($\mathcal{G}$) 的支撑。令 $\mathbf{x} \in \operatorname{supp}(\mu)$。对于 $\mathbf{x}$ 的任意邻域 $N_{\mathbf{x}}$，都存在一个开球 $B(\mathbf{x}\ \varepsilon)$ 包含于 $N_{\mathbf{x}}$。由群作用 $\mathcal{G} \times \mathbb{R}^p \to \mathbb{R}^p : (Q, \mathbf{x}) \mapsto Q\mathbf{x}$ 的连续性，存在 I 的一个邻域 V，使得对于所有 $Q \in V$ 以及 $\mathbf{y} \in B(\mathbf{x}\ \varepsilon/2)$，都有 $Q\mathbf{y} \in N_{\mathbf{x}}$。因此，由独立性可得， $\mathbb{P}(S\mathbf{X} \in N_{\mathbf{x}}) \geq \mathbb{P}(S \in V, \mathbf{X} \in B(\mathbf{x}\ \varepsilon/2)) = \mathbb{P}(S \in V)\mathbb{P}(\mathbf{X} \in B(\mathbf{x}\ \varepsilon/2)) > 0$，这里用到了 I、$\mathbf{x}$ 分别属于 Uniform($\mathcal{G}$) 和 μ 的支撑这一事实。因此，$\mathbf{X}$ 支撑中的每个点都属于 $S\mathbf{X} \sim \mu_S$ 的支撑。N

ote that $g(\cdot)$ is Leb斯格几乎处处正 iti于支撑上 (μ_S). Thus f对于任意 Borel set $B \subset \mathbb{R}^p$,

$$\mathbb{P}(\mathbf{X} \in B) = \int_{\operatorname{supp}(\mu_S) \cap B} f(\mathbf{x})\, \mathrm{d}\mathbf{x} = \int_{\operatorname{supp}(\mu_S) \cap B} \frac{f(\mathbf{x})}{g(\mathbf{x})} g(\mathbf{x})\, \mathrm{d}\mathbf{x} = \int_{\operatorname{supp}(\mu_S) \cap B} \frac{f(\mathbf{x})}{g(\mathbf{x})}\, \mathrm{d}\mu_S(\mathbf{x}).$$

因此，$\mathbf{X}$ 相对于 $S\mathbf{X}$ 绝对连续，其拉东-尼科迪姆导数为 $f_{(\cdot)}\ g_{(\cdot)}$。 —— □

定理 2.1 的证明。根据命题 C.1， $c(\cdot, \cdot)$ 的连续性意味着存在一个具有 c-循环单调支撑的 $\pi \in \Pi(\mu, \nu)$。根据命题 C.2，对于任意具有 c-循环单调支撑的 $\pi \in \Pi(\mu, \nu)$，存在一个 c-凸函数 ψ：$\mathbb{R}^p \to \mathbb{R} \cup \{+\infty\}$，使得 π 的支撑包含于 c-次微分集合中

$$\partial_c \psi := \{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^p \times \mathbb{R}^p : \psi^c(\mathbf{y}) - \psi(\mathbf{x}) = c(\mathbf{x}, \mathbf{y})\}.$$

由于 ψ 是 c-凸的，且 $c(\cdot, \cdot)$ 连续，因此 ψ 是通常意义下的下半连续凸函数。因此， ψ 在其定义域内几乎处处勒贝格可微（该定义域包含 μ 的支撑）。任取支撑于 π 中的一对 $(\mathbf{x}, \mathbf{y})$，并假设 ψ 在 $\mathbf{x}$ 处可微。令 $\nabla_{\mathbf{x}}(-c(\mathbf{x}, \mathbf{y}))$ 表示 $g(\mathbf{x}) := \max_{Q \in \mathcal{G}} \mathbf{x}^\top Q\mathbf{y}$ 的一个次梯度；该次梯度显然存在（因为 $g(\cdot)$ 是凸函数），且属于凸包 $\{Q\mathbf{y} : Q \text{ 是 } \max_{Q \in \mathcal{G}} \mathbf{x}^\top Q\mathbf{y}\}$ 的最大化点（即 $\mathbf{x}$ 处活跃线性函数的梯度；例如参见 [55, 定理 4.4.2]）。由于 $\psi(\cdot)$ 在 $\mathbf{x}$ 处可微，对于邻域 $\mathbf{0} \in \mathbb{R}^p$ 内的 $\mathbf{h} \in \mathbb{R}^p$，有

$$\psi(\mathbf{x} + \mathbf{h}) = \psi(\mathbf{x}) + \nabla\psi(\mathbf{x}) \cdot \mathbf{h} + o(\mathbf{h}). \tag{B.6}$$

现在，利用 c-变换的定义（见（B.5））：

$$\begin{aligned}
\psi(\mathbf{x} + \mathbf{h}) &\geq \psi^c(\mathbf{y}) - c(\mathbf{x} + \mathbf{h}, \mathbf{y}) \\
&\geq \psi^c(\mathbf{y}) - c(\mathbf{x}, \mathbf{y}) + \nabla_{\mathbf{x}}(-c(\mathbf{x}, \mathbf{y})) \cdot \mathbf{h} \quad \text{(by the definition of sub-gradient of } g(\cdot)) \\
&= \psi(\mathbf{x}) + \nabla_{\mathbf{x}}(-c(\mathbf{x}, \mathbf{y})) \cdot \mathbf{h}.
\end{aligned}$$

由于 $\mathbf{h}$ 可以取任意方向，将上述表达式与（B.6）进行比较，我们得到

$$\nabla\psi(\mathbf{x}) = \nabla_{\mathbf{x}}(-c(\mathbf{x}, \mathbf{y})). \tag{B.7}$$

这也意味着次梯度 $\nabla_{\mathbf{x}}(-c(\mathbf{x},\mathbf{y}))$ 是唯一的，即 $\{\widetilde{Q}\mathbf{y} : \widetilde{Q} \in \operatorname{argmax}_{Q \in \mathcal{G}} \mathbf{x}^\top Q\mathbf{y}\}$ 仅包含一个元素。令 $Q_0 := \operatorname{argmax}_{Q \in \mathcal{G}} \mathbf{x}^\top Q\mathbf{y}$ 为 $\mathbf{x}^\top Q\mathbf{y}$ 的唯一最大化元。接下来，

[12]否则，存在 $Q_0 \in \mathcal{G}$，以及 Q_0 的一个邻域 U，其概率为 0。考虑 $\mathcal{G}$ 的以下开覆盖： $\{QU : Q \in \mathcal{G}\}$。根据 $\mathcal{G}$ 的紧致性，我们可以找到一个有限子覆盖 $\cup_{i=1}^k Q_i U$，它覆盖 $\mathcal{G}$ 且概率为零，从而导出矛盾。

$\nabla\psi(\mathbf{x}) = Q_0\mathbf{y}$，使得 $\mathbf{y} = Q_0^{-1}\nabla\psi(\mathbf{x})$。由于轨道 $\{Q\nabla\psi(\mathbf{x}) : Q \in \mathcal{G}\}$ 与 B 至多相交于一点，因此我们有 $\mathbf{y} = q(\nabla\psi(\mathbf{x}))$（因为 $\mathbf{y} \in B$ 且 ν 集中在 B上）。

因此，对于 μ-几乎处处成立的 $\mathbf{x}$，只有一个 $\mathbf{y}$ 满足 $(\mathbf{x},\mathbf{y}) \in \partial_c\psi$。因此，$\pi$ 由确定性耦合给出，即 (恒等映射 $q\circ\nabla\psi)\#\mu$，从而证明了 R的存在性。尽管 $q\circ\nabla\psi$ 在 ψ的定义域上仅几乎处处（勒贝格意义下）有定义，但它存在一个 μ-几乎处处唯一的博雷尔可测版本（见 Villani [101, 定理 5。30]的证明），我们将其定义为 R（注意，$q\circ\nabla\psi$ 与 R 在一个测度为 μ-1 的博雷尔集上相同）。

令 $S \sim$ Uniform($\mathcal{G}$) 与其他所有对象相互独立。因此，$S \overset{d}{=} SQ$ 对任意 $Q \in \mathcal{G}$均成立。因此，对于 μ-几乎处处的 x，$S\nabla\psi(\mathbf{x}) \overset{d}{=} SR(\mathbf{x})$，因为 $\nabla\psi(\mathbf{x})$ 和 $R(\mathbf{x}) = q(\nabla\psi(\mathbf{x}))$ 属于同一轨道（即，对某个 $Q \in \mathcal{G}$，有 $R(\mathbf{x}) = Q\nabla\psi(\mathbf{x})$）。因此，$S\nabla\psi(\mathbf{X}) \overset{d}{=} SR(\mathbf{X}) \overset{d}{=} S\mathbf{H}$（回忆 $R(\mathbf{X}) \sim \mathbf{H}$）。由于 ψ 是一个 c-凸函数，$\nabla\psi(\cdot)$ 的 $\mathcal{G}$-等变性（见定义 B.1 后的讨论）由此推出 $\nabla\psi(S\mathbf{X}) \overset{d}{=} S\mathbf{H}$。由于 $\nabla\psi(\cdot)$——一个凸函数的梯度——将对称化版本 $S\mathbf{X}$ 推送至 $S\mathbf{H}$，根据 McCann 定理，[73]（无需任何矩条件）$\nabla\psi(\cdot)$ 由 μ_S 的绝对连续性决定，且在 μ_S-几乎处处意义下唯一。由于 μ 相对于 μ_S 绝对连续，$\nabla\psi(\cdot)$ 同样由此唯一确定，且在 μ-几乎处处意义下成立，从而证明了 R 的唯一性。

为证明 (iv)，注意根据定义

$$S(\mathbf{x},\mathbf{h}) = \operatorname*{argmin}_{Q\in\mathcal{G}} \|Q^\top\mathbf{x} - \mathbf{h}\|^2 = \operatorname*{argmin}_{Q\in\mathcal{G}} \|\mathbf{x} - Q\mathbf{h}\|^2 = \operatorname*{argmax}_{Q\in\mathcal{G}} \mathbf{x}^\top Q\mathbf{h},$$

因此，$S(\mathbf{x}\ R(\mathbf{x}))R(\mathbf{x})$ 是轨道 $\{QR(\mathbf{x}) : Q \in \mathcal{G}\}$ 中距离 x 最近的点。然而，如 (B.7) 所示，对于 μ-几乎处处的 x，$\nabla\psi(\mathbf{x})$ 是 $\{\widetilde{Q}R(\mathbf{x}) : \widetilde{Q} \in \operatorname{argmax}_{Q\in\mathcal{G}} \mathbf{x}^\top QR(\mathbf{x})\}$中的唯一元素。这意味着 $S(\mathbf{X}\ R(\mathbf{X}))R(\mathbf{X}) \overset{a.s.}{=} \nabla\psi(\mathbf{X})$。

最后，(v) 直接由 Villani [101, 定理 510] 得出。 □

至此，定理 2.1 已得证。接下来，我们将为定理2.2、3.2 和3.3进一步推导一些有用的结果。

引理 *B.4*（当 $\mathcal{G}$ 在 $\mathcal{G}B$ 上自由作用时，$S(\cdot,\cdot)$ 的连续性与唯一性）。在定理 *2.1* 的假设下，如果 $\mathcal{G}$ 在 $\mathcal{G}B$ 上自由作用，那么对于 *(identity R)#μ*-几乎处处的 *(x,y)*，$S(x,y) := \operatorname{argmin}_{Q\in\mathcal{G}} \|Q^\top x - y\|^2$ 是唯一且连续的。

证明。我们首先证明唯一性。假设存在两个极小化点 Q_1, Q_2，它们对应于 $\|Q^\top\mathbf{x} - \mathbf{y}\|^2$。那么 Q_1, Q_2 也都是 $\mathbf{x}^\top Q\mathbf{y}$ 的极大化点。由于对于 μ-几乎处处的 x 和 y，$= R(\mathbf{x})$，根据定理 2.1 的证明，$\{Q\mathbf{y} : Q$ 是 $\max_{S\in\mathcal{G}}\mathbf{x}^\top Q\mathbf{y}\}$ 的极大化点集合，且仅有一个元素 $\nabla\psi(\mathbf{x})$，因此必有 $Q_1\mathbf{y} = Q_2\mathbf{y} = \nabla\psi(\mathbf{x})$。这意味着 $Q_1^{-1}Q_2\mathbf{y} = \mathbf{y}$，进而得到 $Q_1 = Q_2$，因为 $\mathcal{G}$ 在 $\mathcal{G}B$ 上自由作用。因此，$\operatorname{argmin}_{Q\in\mathcal{G}} \|Q^\top\mathbf{x} - \mathbf{y}\|^2$ 几乎处处唯一（恒等映射 R）$\#\mu$。

接下来，我们证明其连续性。假设 $(\mathbf{x}_n, \mathbf{y}_n) \to (\mathbf{x}_0,\mathbf{y}_0)$，其中 $n \to \infty$，且 $S(\mathbf{x}, \mathbf{y})$ 在 $(\mathbf{x}_0, \mathbf{y}_0)$ 处唯一确定。考虑函数 $f_n\colon \mathbb{R}^{p\times p} \to \mathbb{R}$，定义为 $f_n(Q) = \|Q^\top\mathbf{x}_n - \mathbf{y}_n\|^2$，若 $Q \in \mathcal{G}$，否则 $f_n(Q) = +\infty$。令 $f\colon \mathbb{R}^{p\times p} \to \mathbb{R}$ 为由 $f(Q) = \|Q^\top\mathbf{x}_0 - \mathbf{y}_0\|^2$（若 $Q \in \mathcal{G}$）以及 $f(Q) = +\infty$（否则）定义的函数。对于任意 $Q \in \mathbb{R}^{p\times p}$，容易看出，对于每个满足 $\|Q_n - Q\|_F \to 0$ 的序列$\{Q_n\}_{n\geq 1} \subset \mathbb{R}^{p\times p}$，$\liminf_n f_n(Q_n) \geq f(Q)$（利用 f_n 的形式，该函数在 $\mathcal{G}$ 上连续，以及 $\mathcal{G}$ 的紧致性）。此外，对于某个满足 $\|Q_n - Q\|_F \to 0$ 的序列 $\{Q_n\}_{n\geq 1} \subset \mathbb{R}^{p\times p}$，我们也有 $\limsup_n f_n(Q_n) \leq f(Q)$。这意味着 f_n 上图收敛于 f [89,（命题 7.2]）。注意，f_n 和 f 是下半连续的，且具有有界水平集 $\{Q \in \mathbb{R}^{p\times p} : f_n(Q) \leq \alpha\}$，对所有 $\alpha \in \mathbb{R}$ 均成立。根据 Rockafellar 和 Wets 的定理 7.33],，f_n 的极小化解必然收敛于 f 的极小化解。因此，证明了 $S(\mathbf{x}, \mathbf{y})$ 的连续性。 □

以下　　证明定理 2.2、3.2 和 3. 时将用到以下弱收敛结果。　　3.

引理 *B.5*（经验秩与符号秩的联合弱收敛）。假设 *1* 和 *2* 成立。令 μ_n 表示 $\{X1, \ldots, X_n\}$ 和 $\mathcal{X}_n \sim \mu_n$上的经验分布。以概率 *1*，我们有 $(\mathcal{X}_n, R_n(\mathcal{X}_n)) \xrightarrow{w} (X\ R(X))$，且 $(\mathcal{X}_n, S_n(\mathcal{X}_n)R_n(\mathcal{X}_n)) \xrightarrow{w} (X\ \nabla\psi(X))$。

证明。 众所周知，μ_n 以概率 1 弱收敛到 μ。固定满足 $\mu_n(\omega) \xrightarrow{w} \mu$ 的 $\omega \in \Omega$（样本空间）。考虑随机测度序列（identity R_n）$\#\mu_n(\omega)$，其中 R_n

是经验秩映射，随机性来自根据 $\mu_n(\omega)$进行的随机抽取。由于 $R_n\#\mu_n = \nu_n$ 弱收敛（根据假设 2，收敛至 ν），且 $\mu_n(\omega)$弱收敛（收敛至 μ），因此(identity R_n)$\#\mu_n(\omega)$是紧的。于是，根据普罗霍罗夫定理，(identity R_n)$\#\mu_n(\omega)$的每个子序列都有一个进一步的子序列弱收敛。我们将证明，无论选择哪些子序列，该极限都是(identity R)$\#\mu$。因此，由于弱收敛是可度量化的，(identity R_n)$\#\mu_n(\omega)$弱收敛至同一个极限。

现在假设收敛子序列为 (identity R_{n_k})$\#\mu_{n_k}(\omega)$。根据 R_{n_k} 的最优性（例如，参见命题 C.1 的证明），我们知道 (identity R_{n_k})$\#\mu_{n_k}(\omega)$ 的支撑包含于一个 c-循环单调集合中。结合 $c(\cdot,\cdot)$ 的连续性可知，其弱极限的支撑也必定包含于一个 c-循环单调集合中（命题 C.1 的证明也给出了这一结论）。由定理 2.1 可知，$\Pi(\mu,\nu)$ 中唯一的 c-循环单调测度为 (identity R)$\#\mu$；由此可得 (identity R_n)$\#\mu_n(\omega) \xrightarrow{w}$ (identity R)$\#\mu$。或者等价地，令 $\mathcal{X}_n \sim \mu_n$，有 $(\mathcal{X}_n, R_n(\mathcal{X}_n)) \xrightarrow{w}$ (X R(X))。

接下来，我们利用上述弱收敛证明 $(\mathcal{X}_n, S_n(\mathcal{X}_n)R_n(\mathcal{X}_n)) \xrightarrow{w}$ (X $\nabla\psi$(X))。回顾一下，带符号秩是秩向量轨道中距离数据点最近的点。因此，定义 f(x, y) := Proj$_{\mathcal{G}\mathbf{y}}$(x) 为 x 在 y 的轨道上的投影，我们可以写成 $S_n(\mathcal{X}_n)R_n(\mathcal{X}_n) = f(\mathcal{X}_n, R_n(\mathcal{X}_n))$，且 $\nabla\psi$(X) = f(X R(X))。我们将证明 f(x, y) 在 (identity R)$\#\mu$ 几乎处处（关于 (x,y)）连续。我们在定理 2.1 的证明中已证明，带符号秩 f(x, y) = Proj$_{\mathcal{G}\mathbf{y}}$(x) 在 (identity R)$\#\mu$ 几乎处处（关于 (x,y)）唯一确定。若 x 在 $\mathcal{G}$y 中最近的点不唯一，则可将 f(x, y) 选为任意一个投影；我们可以证明连续性仍然成立。

设 ($\mathbf{x}_0$,$\mathbf{y}_0$) 是投影 Proj$_{\mathcal{G}\mathbf{y}_0}$($\mathbf{x}_0$) 唯一的点。我们用反证法。如果 f(x, y) 在 ($\mathbf{x}_0, \mathbf{y}_0$) 处不连续，则存在 $\delta > 0$，以及一个序列 ($\mathbf{x}_n$, $\mathbf{y}_n$) $\to$ ($\mathbf{x}_0, \mathbf{y}_0$)，使得对所有 n，$\|$Proj$_{\mathcal{G}\mathbf{y}_n}$($\mathbf{x}_n$) $-$Proj$_{\mathcal{G}\mathbf{y}_0}$($\mathbf{x}_0$)$\| \geq \delta$。定义点 x $\in \mathbb{R}^p$ 到集合 $A \subset \mathbb{R}^p$ 的距离为 d(x A) := inf$_{\mathbf{y}\in A}$(x, y)。令 $d_0 := d$(x0, $\mathcal{G}$y$_0$)，并设 Proj$_{\mathcal{G}\mathbf{y}_0}$($\mathbf{x}_0$) = g_0y$_0$ 是 $\mathcal{G}$y$_0$ 中达到该最小距离的点，即 $\|\mathbf{x}_0 - g_0\mathbf{y}_0\| = d_0$。注意到 d($\mathbf{x}_n$, $\mathcal{G}$y$_n$) $\leq d$(x$_n$, x$_0$) + d(x0, $\mathcal{G}$y$_n$) $\leq d$(x$_n$, x$_0$) + d(x0, g_0y$_n$) $\leq d$(x$_n$, x$_0$) + d(x0, g_0y$_0$) + d(g_0y0, g_0y$_n$) = $\|\mathbf{x}_n - \mathbf{x}_0\| + d_0 + \|\mathbf{y}_n - \mathbf{y}_0\|$。因此，$\|$Proj$_{\mathcal{G}\mathbf{y}_n}$($\mathbf{x}_n$) $-$ x$_0\| \leq \|$Proj$_{\mathcal{G}\mathbf{y}_n}$($\mathbf{x}_n$) $-\mathbf{x}_n\| + \|\mathbf{x}_n - \mathbf{x}_0\| =$ d(x$_n$, $\mathcal{G}$y$_n$) + $\|\mathbf{x}_n - \mathbf{x}_0\| \leq 2\|\mathbf{x}_n - \mathbf{x}_0\| + d_0 + \|\mathbf{y}_n - \mathbf{y}_0\| \to d_0$，因为 $n \to \infty$。因此，Proj$_{\mathcal{G}\mathbf{y}_n}$($\mathbf{x}_n$) 必有一个子序列收敛到 {x $\in \mathbb{R}^p$ 中的某个点：$\|\mathbf{x} - \mathbf{x}_0\| \leq d_0\} \cap \mathcal{G}\mathbf{y}_0$,，该集合仅包含一个点 Proj$_{\mathcal{G}\mathbf{y}_0}$($\mathbf{x}_0$)，这是由投影 Proj$_{\mathcal{G}\mathbf{y}_0}$($\mathbf{x}_0$) 的唯一性所致。这与对所有 n，$\|$Proj$_{\mathcal{G}\mathbf{y}_n}$(x$_n$) $-$Proj$_{\mathcal{G}\mathbf{y}_0}$($\mathbf{x}_0$)$\| \geq \delta > 0$ 矛盾。

现在，对 (x，y) $\mapsto$ (x f(x，y)) 和 ($\mathcal{X}_n$，$R_n(\mathcal{X}_n)$) $\xrightarrow{w}$ (X R(X)) 应用连续映射定理，可得 ($\mathcal{X}_n$, $S_n(\mathcal{X}_n)R_n(\mathcal{X}_n)$) $\xrightarrow{w}$ (X $\nabla\psi$(X))。 □

最后，我们给出一个简单的引理，它可作为证明定理2.2的便捷工具。

引理 B.6。假设条件 2 成立。对于连续函数 $f: B \to \mathbb{R}$，若有 $\frac{1}{n}\sum_{i=1}^n |f(\mathbf{h}_{n,i})| \to \int |f(\mathbf{x})| d\nu(\mathbf{x})$，则 $f\#\nu_n$ 一致可积。

证明。由于 $\mathbf{H}_n \sim \nu_n$ 弱收敛于 ν，通过选取一个几乎处处相等的代表元，我们可以假设 $\mathbf{H}_n \xrightarrow{a.s.} \mathbf{H}$；由连续映射定理，$f(\mathbf{H}_n) \xrightarrow{a.s.} f(\mathbf{H})$。由于矩也收敛，即 $\mathbb{E}|f(\mathbf{H}_n)| \to \mathbb{E}|f(\mathbf{H})|$，Scheffé 引理表明 $f(\mathbf{H}_n)$ 弱收敛于 $f(\mathbf{H})$，在 L^1 中成立。这意味着$\{f(\mathbf{H}_n)\}_{n\geq 1}$ 一致可积。 □

定理 2.2 的证明。我们首先陈述凸函数梯度的一条连续性性质。注意，凸函数 $\nabla\psi(\cdot)$ 的梯度在几乎处处（Lebesgue 意义下）唯一确定。在极少数梯度在某点 x 不存在的情形下，我们可以将 $\nabla\psi$(x) 定义为 x 处任意一个次梯度，即次微分集合 $\partial\psi$(x) 中的一个元素（例如参见 [73]）。凸函数定义域内部总是存在次梯度。$\nabla\psi(\cdot)$ 的这一概念在以下意义下是连续的：若 x $\in \mathbb{R}^p$ 是 $\psi(\cdot)$ 的一个可微点，且 $\mathbf{x}_n \to$ x 和 $\mathbf{y}_n \in \partial\psi(\mathbf{x}_n)$，则 $\mathbf{y}_n \to \nabla\psi$(x)。

现在证明定理 2.2 的第 1 部分。根据引理 B.5，以概率 1，($\mathcal{X}_n, R_n(\mathcal{X}_n)$) $\xrightarrow{w}$ (X R(X)) 且 ($\mathcal{X}_n, S_n(\mathcal{X}_n)R_n($ $\mathcal{X}_n$)) $\xrightarrow{w}$ (X $\nabla\psi$(X))，其中 $\mathcal{X}_n \sim \mu_n$。由于 As$q(\cdot)$ 是连续的，$R(\cdot) = q(\nabla\psi(\cdot))$ 是 μ-几乎处处连续的。因此，将连续映射定理应用于 f(x, y) = $\rho(R$(x), y)，($\mathcal{X}_n, R_n(\mathcal{X}_n)$) $\xrightarrow{w}$ (X R(X)) 蕴含 $\rho(R(\mathcal{X}_n)$ $R_n(\mathcal{X}_n))$ $\xrightarrow{w}$

$\rho(R(\mathbf{X}) \; R(\mathbf{X})) = 0$。注意，$R_n(\mathcal{X}_n) \sim \nu_{n\circ} \;\; \mathbb{E}\rho(\mathbf{H}_n, 0) \to \mathbb{E}\rho(\mathbf{H}, 0)$，结合引理 B.6，蕴含 $\rho(R_n(\mathcal{X}_n), 0)$ 一致可积。因此，$\rho(R(\mathbf{X}), 0)$ 作为 $\rho(R_n(\mathcal{X}_n), 0)$ 的弱极限，根据 Fatou 引理可积。随后，强大数定律表明 $\mathbb{E}\rho(R(\mathcal{X}_n), 0)$ 收敛到 $\mathbb{E}\rho(R(\mathbf{X}), 0)$。由于 $\rho(R(\mathcal{X}_n), 0)$ 依分布收敛到 $\rho(R(\mathbf{X}), 0)$，Scheff´e 引理再次表明 $\rho(R(\mathcal{X}_n), 0)$ 一致可积。因此，$\rho(R(\mathcal{X}_n) \; R_n(\mathcal{X}_n)) \lesssim \rho(R(\mathcal{X}_n), 0) + \rho(R_n(\mathcal{X}_n), 0)$ 一致可积，且其均值收敛到 0 的均值，即，

$$\frac{1}{n}\sum_{i=1}^{n} \rho(R(\mathbf{X}_i), R_n(\mathbf{X}_i)) \to 0.$$

接下来，我们证明符号的收敛。根据引理 B.4，并对 $f(\mathbf{x}, \mathbf{y}) = \widetilde{\rho}(S(\mathbf{x}, \mathbf{y}) \; S(\mathbf{x} \; R(\mathbf{x})), (\mathcal{X}_n, R_n(\mathcal{X}_n)) \xrightarrow{w} (\mathbf{X} \; R(\mathbf{X}))$ 应用连续映射定理，可得 $\widetilde{\rho}(S(\mathcal{X}_n, R_n(\mathcal{X}_n)) \; S(\mathcal{X}_n, R(\mathcal{X}_n))) \xrightarrow{w} 0$。由于 $\widetilde{\rho}$ 是连续的，且符号有界，因此我们可以应用有界收敛定理，从而其均值也收敛，即

$$\frac{1}{n}\sum_{i=1}^{n} \widetilde{\rho}(S(\mathbf{X}_i, R_n(\mathbf{X}_i)), S(\mathbf{X}_i, R(\mathbf{X}_i))) \to 0 \qquad \text{as } n \to \infty.$$

由于 $S(\mathbf{X}_i, R_n(\mathbf{X}_i)) = S_n(\mathbf{X}_i)$，因此收敛结果成立。

最后，我们证明经验符号秩的收敛。由于以概率 1，$(\mathcal{X}_n, S_n(\mathcal{X}_n)R_n(\mathcal{X}_n)) \xrightarrow{w} (\mathbf{X} \; \nabla\psi(\mathbf{X}))$（根据引理 B.5），对 $f(\mathbf{x}, \mathbf{y}) = \rho(\mathbf{y} \; \nabla\psi(\mathbf{x}))$ 应用连续映射定理，我们有 $\rho(S_n(\mathcal{X}_n)R_n(\mathcal{X}_n) \; \nabla\psi(\mathcal{X}_n)) \xrightarrow{w} 0$。我们已经证明 $\rho(R_n(\mathcal{X}_n), 0)$ 是一致可积的。由 $\rho(Q\mathbf{x}, 0) \lesssim \rho(\mathbf{x}, 0) + 1$ 可知，$\rho(S_n(\mathcal{X}_n)R_n(\mathcal{X}_n), 0)$ 也是一致可积的。同样地，由于 $\nabla\psi(\mathcal{X}_n)$ 和 $R(\mathcal{X}_n)$ 位于同一轨道上，$\rho(R(\mathcal{X}_n), 0)$ 的一致可积性意味着 $\rho(\nabla\psi(\mathcal{X}_n), 0)$ 是一致可积的。因此，$\rho(S_n(\mathcal{X}_n)R_n(\mathcal{X}_n) \; \nabla\psi(\mathcal{X}_n))$ 是一致可积的，且其均值收敛于 0 的均值，即

$$\frac{1}{n}\sum_{i=1}^{n} \rho(S_n(\mathbf{X}_i)R_n(\mathbf{X}_i), \nabla\psi(\mathbf{X}_i)) \to 0 \qquad \text{as } n \to \infty.$$

□

定理 *3.2* 和 *3.3* 的证明。只需证明以下结论：
1. $\frac{1}{n}\sum_{i=1}^{n} S_n(\mathbf{X}_i) \xrightarrow{a.s.} \mathbb{E}[S(\mathbf{X} \; R(\mathbf{X}))]$；2. $\frac{1}{n}\sum_{i=1}^{n} S_n(\mathbf{X}_i)J(R_n(\mathbf{X}_i)) \xrightarrow{a.s.} \mathbb{E}[J(\nabla\psi(\mathbf{X}))]$。

由引理 B.5 可知，对于几乎处处的 ω，(恒等映射 R_n)$\#\mu_n(\omega) \xrightarrow{w}$ (恒等映射 R)$\#\mu$。由引理 B.4，$S(\cdot, \cdot)$ 关于 (恒等映射 R)$\#\mu$ 几乎处处连续（对于球对称性，如注 3.1 所述，$S(\cdot, \cdot, \cdot)$ 关于 (恒等映射 R)$\#\mu \otimes P_\varepsilon$-几乎处处连续），通过找到一个几乎处处成立的代表元并应用控制收敛定理，积分收敛：

$$\frac{1}{n}\sum_{i=1}^{n} S_n(\mathbf{X}_i) = \frac{1}{n}\sum_{i=1}^{n} S(\mathbf{X}_i, R_n(\mathbf{X}_i), \varepsilon_i) \to \mathbb{E}[S(\mathbf{X}, R(\mathbf{X}), \varepsilon)].$$

为证明上述第 2 部分，令 $C \subset \mathbb{R}^p$ 为 $J(\cdot)$ 连续的点的集合。根据假设 3，$\nu_S(C) = 1$。由定理 2.1，由于 $\nabla\psi(\cdot)$ 将 μ_S 推送至 ν_S，我们有 $\mathbb{P}(\nabla\psi(S\mathbf{X}) \in C) = 1$。由于 μ 相对于 μ_S 绝对连续（根据引理 B.3），因此 $\mathbb{P}(\nabla\psi(\mathbf{X}) \in C) = 1$。这使我们可以应用连续映射定理。

根据引理 B.5，对于 $\mathcal{X}_n \sim \mu_n(\omega)$，我们有 $(\mathcal{X}_n, S_n(\mathcal{X}_n)R_n(\mathcal{X}_n)) \xrightarrow{w} (\mathbf{X} \; \nabla\psi(\mathbf{X}))$。特别地，根据 $J(\cdot)$ 的等变性和连续性，$S_n(\mathcal{X}_n)J(R_n(\mathcal{X}_n)) = J(S_n(\mathcal{X}_n)R_n(\mathcal{X}_n)) \xrightarrow{w} J(\nabla\psi(\mathbf{X}))$。注意，$\|S_n(\mathcal{X}_n)J(R_n(\mathcal{X}_n))\| = \|J(R_n(\mathcal{X}_n))\|$ 一致可积，因为：(i) $\mathbb{E}\|J(\mathbf{H}_n)\| \to \mathbb{E}\|J(\mathbf{H})\|$，(ii) $J(\mathbf{H}_n) \xrightarrow{w} J(\mathbf{H})$，以及 (iii) 应用 Scheff´e 引理。因此，我们得到一阶矩的收敛：

$$\frac{1}{n}\sum_{i=1}^{n} S_n(\mathbf{X}_i)J(R_n(\mathbf{X}_i)) \to \mathbb{E}[J(\nabla\psi(\mathbf{X}))].$$

□

B.5 注释3.2的证明

在备注 3.2 的假设下，$S(X, R(X)) = \text{sign}(X)$ 且 $R(X) = F_{|X|}(|X|)$，由于 $\text{sign}(\cdot)F_{|X|}(|\cdot|)$ 将 μ_S 传输到均匀分布$(-1, 1)$，并且表现为凸函数的梯度。因此根据定理 3.2，GWSR 检验对于 $\mathbb{E}\left[\text{sign}(X)F_{|X|}(|X|)\right] \neq 0$ 是一致的。设f 是 X的密度函数。那么：

$$
\begin{aligned}
&\mathbb{E}\left[\text{sign}(X)F_{|X|}(|X|)\right] > 0 \\
\Leftrightarrow &\int_0^\infty \left[\frac{f(x)}{f(x)+f(-x)} - \frac{f(-x)}{f(x)+f(-x)}\right] F_{|X|}(x)\left[f(x)+f(-x)\right]\mathrm{d}x > 0 \\
\Leftrightarrow &\int_0^\infty \left[f(x)-f(-x)\right] F_{|X|}(x)\mathrm{d}x > 0 \\
\Leftrightarrow &\int_0^\infty \left[f(x)-f(-x)\right] \int_0^x \left[f(y)+f(-y)\right]\mathrm{d}y\mathrm{d}x > 0 \\
\Leftrightarrow &\int_{0\le y\le x} \left[f(x)-f(-x)\right]\left[f(y)+f(-y)\right]\mathrm{d}x\mathrm{d}y > 0 \\
\Leftrightarrow &\int_{0\le y\le x} f(x)\left[f(y)+f(-y)\right]\mathrm{d}x\mathrm{d}y > \int_{0\le y\le x} f(-x)\left[f(y)+f(-y)\right]\mathrm{d}x\mathrm{d}y \\
\Leftrightarrow &\mathbb{P}(0\le X_2\le X_1) + \mathbb{P}(0\le -X_2\le X_1) > \mathbb{P}(0\le X_2\le -X_1) + \mathbb{P}(0\le -X_2\le -X_1) \\
\Leftrightarrow &\mathbb{P}(-X_1\le X_2\le X_1) > \mathbb{P}(X_1\le X_2\le -X_1) \\
\Leftrightarrow &\mathbb{P}(X_1+X_2\ge 0, X_2\le X_1) > \mathbb{P}(X_1+X_2\le 0, X_1\le X_2) \\
\Leftrightarrow &\frac{1}{2}\mathbb{P}(X_1+X_2\ge 0) > \frac{1}{2}\mathbb{P}(X_1+X_2\le 0) \\
\Leftrightarrow &\mathbb{P}(X_1+X_2>0) > \frac{1}{2}.
\end{aligned}
$$

将 X 替换为 $-X$，我们得到：

$$
\mathbb{E}\left[\text{sign}(X)F_{|X|}(|X|)\right] < 0 \quad \Leftrightarrow \quad \mathbb{P}(X_1+X_2>0) < \frac{1}{2},
$$

由此完成证明。 □

B.6 推论3.1的证明

根据定理 3.2，我们只需证明 $\mathbb{E}[\nabla\psi(\mathbf{X})] \neq \mathbf{0}$。

我们首先证明 $\nabla\psi(\mathbf{X})$ 和 $\nabla\psi(\mathbf{X}-2\boldsymbol{\Delta})$ 都可积。注意，$\mathbb{E}[\|\nabla\psi(\mathbf{X})\|] = \mathbb{E}[\|\mathbf{H}\|]$其中 $\mathbf{H} \sim \nu$；因此 $\nabla\psi(\mathbf{X}) \in L^1$。由于 $-I \in \mathcal{G}$ (根据假设 5-(ii))，

$$
2\boldsymbol{\Delta} - \mathbf{X} = \boldsymbol{\Delta} + (\boldsymbol{\Delta} - \mathbf{X}) \overset{d}{=} \boldsymbol{\Delta} + (\mathbf{X} - \boldsymbol{\Delta}) = \mathbf{X}. \tag{B.8}
$$

由于 $\nabla\psi$ 在 $\mathcal{G}$的群作用下是等变的，

$$
\nabla\psi(\mathbf{X}-2\boldsymbol{\Delta}) = -\nabla\psi(2\boldsymbol{\Delta}-\mathbf{X}) \overset{d}{=} -\nabla\psi(\mathbf{X}) \in L^1.
$$

为证明 $\mathbb{E}[\nabla\psi(\mathbf{X})] \neq \mathbf{0}$我们采用反证法。若 $\mathbb{E}[\nabla\psi(\mathbf{X})] = \mathbf{0}$，则（B.8）将推出$\mathbb{E}[\nabla\psi(\mathbf{X}-2\boldsymbol{\Delta})] = \mathbf{0}$因此：

$$
\mathbb{E}\left[\boldsymbol{\Delta}^\top\{\nabla\psi(\mathbf{X}) - \nabla\psi(\mathbf{X}-2\boldsymbol{\Delta})\}\right] = 0.
$$

根据凸函数的梯度的单调性：$2\boldsymbol{\Delta}^\top\{\nabla\psi(\mathbf{X}) - \nabla\psi(\mathbf{X}-2\boldsymbol{\Delta})\} \geq 0$。这意味着 $\boldsymbol{\Delta}^\top\{\nabla\psi(\mathbf{X}) - \nabla\psi(\mathbf{X}-2\boldsymbol{\Delta})\} = 0$ 几乎处处成立。

设 A 是关于 μ 的正测度开集，其中 ψ 是严格凸的。令 $\mathbf{x}_0 \in A$。考虑 $g(t) := \psi(\mathbf{x}_0 - t\boldsymbol{\Delta})$，$t \in \mathbb{R}$，它是一维凸函数，并且严格

凸于 0 的某个邻域内。这意味着 $g'(2) > g'(0)$，即 $\boldsymbol{\Delta}^\top \{\nabla\psi(\mathbf{x}_0) - \nabla\psi(\mathbf{x}_0 - 2\boldsymbol{\Delta})\} > 0$。由于对于 $\mathbf{x}_0 \notin A$，$\boldsymbol{\Delta}^\top \{\nabla\psi(\mathbf{x}_0) - \nabla\psi(\mathbf{x}_0 - 2\boldsymbol{\Delta})\} \geq 0$，我们得到

$$\begin{aligned}&\mathbb{E}\left[\boldsymbol{\Delta}^\top \{\nabla\psi(\mathbf{X}) - \nabla\psi(\mathbf{X} - 2\boldsymbol{\Delta})\}\right]\\ &= \mathbb{E}\left[\boldsymbol{\Delta}^\top \{\nabla\psi(\mathbf{X}) - \nabla\psi(\mathbf{X} - 2\boldsymbol{\Delta})\} \mathbf{1}_A(\mathbf{X})\right] + \mathbb{E}\left[\boldsymbol{\Delta}^\top \{\nabla\psi(\mathbf{X}) - \nabla\psi(\mathbf{X} - 2\boldsymbol{\Delta})\} \mathbf{1}_{A^c}(\mathbf{X})\right] > 0,\end{aligned}$$

这就导出了矛盾。

□

B.7 引理 3.1 的证明

注意到

$$\begin{aligned}&\mathbb{E}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n S_n(\mathbf{X}_i)J(R_n(\mathbf{X}_i)) - \frac{1}{\sqrt{n}}\sum_{i=1}^n S(\mathbf{X}_i, R(\mathbf{X}_i))J(R(\mathbf{X}_i))\right\|^2\\ &= \frac{1}{n}\mathbb{E}\left\|\sum_{i=1}^n S_n(\mathbf{X}_i)J(R_n(\mathbf{X}_i))\right\|^2 + \frac{1}{n}\mathbb{E}\left\|\sum_{i=1}^n S(\mathbf{X}_i, R(\mathbf{X}_i))J(R(\mathbf{X}_i))\right\|^2\\ &\quad - \frac{2}{n}\mathbb{E}\left(\sum_{i=1}^n S_n(\mathbf{X}_i)J(R_n(\mathbf{X}_i))\right)^\top \left(\sum_{i=1}^n S(\mathbf{X}_i, R(\mathbf{X}_i))J(R(\mathbf{X}_i))\right).\end{aligned} \tag{B.9}$$

由于 $(S_n(\mathbf{X}_1)\ldots, S_n(\mathbf{X}_n)) \perp\!\!\!\perp (R_n(\mathbf{X}_1)\ldots, R_n(\mathbf{X}_n))$ 且服从独立同分布的均匀分布$(\mathcal{G})$,

$$\frac{1}{n}\mathbb{E}\left\|\sum_{i=1}^n S_n(\mathbf{X}_i)J(R_n(\mathbf{X}_i))\right\|^2 = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left\|S_n(\mathbf{X}_i)J(R_n(\mathbf{X}_i))\right\|^2 = \mathbb{E}\left\|J(R_n(\mathbf{X}_1))\right\|^2,$$

因为所有交叉项的期望均为 0（根据假设5-(i)）。

根据 $\mathbf{X}_i$ 的 $\mathcal{G}$-对称性以及 $\nabla\psi$ 和 J 的等变性（此处 $S \sim$ 为均匀分布$(\mathcal{G})$），

$$\mathbb{E}[S(\mathbf{X}_i, R(\mathbf{X}_i))J(R(\mathbf{X}_i))] = \mathbb{E}[J(\nabla\psi(\mathbf{X}_i))] = \mathbb{E}[J(\nabla\psi(S\mathbf{X}_i))] = \mathbb{E}[S]\mathbb{E}[J(\nabla\psi(\mathbf{X}_i))] = \mathbf{0}.$$

这意味着

$$\frac{1}{n}\mathbb{E}\left\|\sum_{i=1}^n S(\mathbf{X}_i, R(\mathbf{X}_i))J(R(\mathbf{X}_i))\right\|^2 = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left\|S(\mathbf{X}_i, R(\mathbf{X}_i))J(R(\mathbf{X}_i))\right\|^2 = \mathbb{E}\left\|J(R(\mathbf{X}_1))\right\|^2.$$

为了简化式（B.9）中的第三项，我们首先证明，对于 $i \neq j$,

$$\mathbb{E}\left[\left(S_n(\mathbf{X}_i)J(R_n(\mathbf{X}_i))\right)^\top \left(S(\mathbf{X}_j, R(\mathbf{X}_j))J(R(\mathbf{X}_j))\right)\right] = 0.$$

我们首先对 \(X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_n\) 以及 \(X_j\) 的轨道进行条件化，即 \(\{v5\}X_j : \{v7\}\)。此时，\(\{v8\}(X_i)\{v10\}(\{v11\}(X_i))\) 已知，而由于 \(X_j\) 的 \(\{v14\}\) 对称性，\(X_j\) 仍然等概率地取轨道中的任意元素。因此，根据 \(\{v16\}\) 和 \(\{v17\}(\{v18\})\) 的等变性，\(\{v19\}(X_j,\{v21\}(X_j))\{v23\}(\{v24\}(X_j))\) \(\{v26\}(\{v27\}(X_j))\) 的条件期望为 0，因而其无条件期望也为 0。因此，

$$\begin{aligned}&\frac{1}{n}\mathbb{E}\left(\sum_{i=1}^n S_n(\mathbf{X}_i)J(R_n(\mathbf{X}_i))\right)^\top \left(\sum_{i=1}^n S(\mathbf{X}_i, R(\mathbf{X}_i))J(R(\mathbf{X}_i))\right)\\ &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\Big(S_n(\mathbf{X}_i)J(R_n(\mathbf{X}_i))\Big)^\top \Big(S(\mathbf{X}_i, R(\mathbf{X}_i))J(R(\mathbf{X}_i))\Big)\\ &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[J(S_n(\mathbf{X}_i)R_n(\mathbf{X}_i))^\top J(\nabla\psi(\mathbf{X}_i))\right].\end{aligned}$$

回想引理 B.5 表明，若 $\mathcal{X}_n \sim \mu_n(\omega)$，则有 $(\mathcal{X}_n, S_n(\mathcal{X}_n)R_n(\mathcal{X}_n)) \xrightarrow{w} (\mathbf{X}\ \nabla\psi(\mathbf{X}))$。因此，由连续映射定理及假设3,

$$\begin{aligned}&J(S_n(\mathcal{X}_n)R_n(\mathcal{X}_n))^\top J(\nabla\psi(\mathcal{X}_n)) + \|J(S_n(\mathcal{X}_n)R_n(\mathcal{X}_n))\|^2 + \|J(\nabla\psi(\mathcal{X}_n))\|^2\\ \xrightarrow{w}\ &J(\nabla\psi(\mathbf{X}))^\top J(\nabla\psi(\mathbf{X})) + \|J(\nabla\psi(\mathbf{X}))\|^2 + \|J(\nabla\psi(\mathbf{X}))\|^2.\end{aligned}$$

由于上述随机变量非负，根据 Fatou 引理，关于 $\mu_n(\omega)$ 的期望满足：

$$\begin{aligned}&\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\left[J(S_n(\mathbf{X}_i)R_n(\mathbf{X}_i))^\top J(\nabla\psi(\mathbf{X}_i)) + \|J(S_n(\mathbf{X}_i)R_n(\mathbf{X}_i))\|^2 + \|J(\nabla\psi(\mathbf{X}_i))\|^2\right]\\ &\geq \mathbb{E}\left[3\|J(\nabla\psi(\mathbf{X}))\|^2\right] = 3\mathbb{E}\|J(R(\mathbf{X}))\|^2.\end{aligned}$$

注意，上述结论对几乎处处的 ω 成立。因此，我们可以取通常的期望，并再次根据 Fatou 引理得到：

$$\begin{aligned}&\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\mathbb{E}\left[J(S_n(\mathbf{X}_i)R_n(\mathbf{X}_i))^\top J(\nabla\psi(\mathbf{X}_i)) + \|J(S_n(\mathbf{X}_i)R_n(\mathbf{X}_i))\|^2 + \|J(\nabla\psi(\mathbf{X}_i))\|^2\right]\\ &\geq \mathbb{E}\left[\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\left[J(S_n(\mathbf{X}_i)R_n(\mathbf{X}_i))^\top J(\nabla\psi(\mathbf{X}_i)) + \|J(S_n(\mathbf{X}_i)R_n(\mathbf{X}_i))\|^2 + \|J(\nabla\psi(\mathbf{X}_i))\|^2\right]\right]\\ &\geq \mathbb{E}\left[3\mathbb{E}\|J(R(\mathbf{X}))\|^2\right] = 3\mathbb{E}\|J(R(\mathbf{X}))\|^2.\end{aligned}$$

注意 $\mathbb{E}\|J(S_n(\mathbf{X}_i)R_n(\mathbf{X}_i))\|^2 = \mathbb{E}\|J(R_n(\mathbf{X}_1))\|^2$ 对所有 i 均成立，并且由 $\mathbb{E}\|J(R(\mathbf{X}))\|^2$的二阶矩收敛可知，它们收敛于 $J\#\nu_n$ (假设 4-(ii))。因此，

$$\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\mathbb{E}\left[J(S_n(\mathbf{X}_i)R_n(\mathbf{X}_i))^\top J(\nabla\psi(\mathbf{X}_i))\right] \geq \mathbb{E}\|J(R(\mathbf{X}))\|^2.$$

综合前面的讨论，我们得到

$$\begin{aligned}&\limsup_{n\to\infty}\mathbb{E}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n S_n(\mathbf{X}_i)J(R_n(\mathbf{X}_i)) - \frac{1}{\sqrt{n}}\sum_{i=1}^n S(\mathbf{X}_i, R(\mathbf{X}_i))J(R(\mathbf{X}_i))\right\|^2\\ &= \limsup_{n\to\infty}\left[\mathbb{E}\|J(R_n(\mathbf{X}_1))\|^2 + \mathbb{E}\|J(R(\mathbf{X}_1))\|^2 - \frac{2}{n}\sum_{i=1}^n\mathbb{E}\left[J(S_n(\mathbf{X}_i)R_n(\mathbf{X}_i))^\top J(\nabla\psi(\mathbf{X}_i))\right]\right]\\ &\leq \lim_{n\to\infty}\mathbb{E}\|J(R_n(\mathbf{X}_1))\|^2 + \mathbb{E}\|J(R(\mathbf{X}_1))\|^2 - 2\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\mathbb{E}\left[J(S_n(\mathbf{X}_i)R_n(\mathbf{X}_i))^\top J(\nabla\psi(\mathbf{X}_i))\right]\\ &\leq \mathbb{E}\|J(R(\mathbf{X}))\|^2 + \mathbb{E}\|J(R(\mathbf{X}))\|^2 - 2\mathbb{E}\|J(R(\mathbf{X}))\|^2 = 0,\end{aligned}$$

至此，证明完成。 □

B.8 定理 3.4 的证明

令 Q_n 为 H_1, 下 $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ 的分布，且 P_n 为 H_0下 $(\mathrm{x}_1, \ldots, \mathrm{X}_n)$ 的分布。在附录 A.1中的正则性条件下，若 $\boldsymbol{\xi}_n \to \boldsymbol{\xi} \in \mathbb{R}^p$，则根据 Van der Vaart [99, 定理 7。2],在 H_0下：

$$\log\frac{\mathrm{d}Q_n}{\mathrm{d}P_n} = \sum_{i=1}^n\log\frac{f(\mathbf{X}_i - \boldsymbol{\xi}_n/\sqrt{n})}{f(\mathbf{X}_i)} = -\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{\boldsymbol{\xi}^\top\nabla f(\mathbf{X}_i)}{f(\mathbf{X}_i)} - \frac{1}{2}\boldsymbol{\xi}^\top I_{\boldsymbol{\theta}_0}\boldsymbol{\xi} + o_p(1),$$

其中 $I_{\boldsymbol{\theta}_0}$ 是 $\boldsymbol{\theta}_0 = \mathbf{0}$处的费 Fisher 信息。

根据引理 3.1 以及多元中心极限定理：

$$\begin{pmatrix}\mathbf{W}_n\\ \log\frac{\mathrm{d}Q_n}{\mathrm{d}P_n}\end{pmatrix}=\begin{pmatrix}\frac{1}{\sqrt{n}}\sum_{i=1}^n J(\nabla\psi(\mathbf{X}_i))\\ -\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{\boldsymbol{\xi}^\top\nabla f(\mathbf{X}_i)}{f(\mathbf{X}_i)}-\frac{1}{2}\boldsymbol{\xi}^\top I_{\boldsymbol{\theta}_0}\boldsymbol{\xi}\end{pmatrix}+o_p(1)$$
$$\xrightarrow{d} N\left(\begin{pmatrix}\mathbf{0}\\ -\frac{\boldsymbol{\xi}^\top I_{\boldsymbol{\theta}_0}\boldsymbol{\xi}}{2}\end{pmatrix},\begin{pmatrix}\mathrm{Var}(J(\nabla\psi(\mathbf{X})))&\boldsymbol{\gamma}\\ \boldsymbol{\gamma}^\top&\boldsymbol{\xi}^\top I_{\boldsymbol{\theta}_0}\boldsymbol{\xi}\end{pmatrix}\right),$$

其中 $\boldsymbol{\gamma}=-\mathbb{E}_{\mathrm{H}_0}\left[J(\nabla\psi(\mathbf{X}))\frac{\boldsymbol{\xi}^\top\nabla f(\mathbf{X})}{f(\mathbf{X})}\right]$。现在，根据 Le Cam 第三引理 [99, 例 6。7]，在 H 下1，

$$\mathbf{W}_n\xrightarrow{d}N\left(\boldsymbol{\gamma},\mathrm{Var}\big(J(\nabla\psi(\mathbf{X}))\big)\right)=N\left(\boldsymbol{\gamma},\Sigma_{\mathrm{ERD}}\right).$$

□

B.9 定义的证明 3.2

我们首先考虑一个序列 $\boldsymbol{\theta}_n=\frac{\boldsymbol{\xi}_n}{\sqrt{n}}$，其中 $\boldsymbol{\xi}_n\to\boldsymbol{\xi}\neq\mathbf{0}$。根据附录 B.8，我们知道在$\boldsymbol{\theta}_n=\frac{\boldsymbol{\xi}_n}{\sqrt{n}}$下，

$$\Sigma_{\mathrm{ERD}}^{-1/2}\mathbf{W}_n\xrightarrow{d}\Sigma_{\mathrm{ERD}}^{-1/2}\boldsymbol{\gamma}+N(\mathbf{0},I_p),\quad\text{and}\quad\sqrt{n}S_n^{-1/2}\overline{\mathbf{X}}_n\xrightarrow{d}\Sigma_{\mathbf{X}}^{-1/2}\boldsymbol{\xi}+N(\mathbf{0},I_p),$$

其中 $\boldsymbol{\gamma}\equiv\boldsymbol{\gamma}(\boldsymbol{\xi})=-\mathbb{E}_{\mathrm{H}_0}\left[J(\nabla\psi(\mathbf{X}))\frac{\boldsymbol{\xi}^\top\nabla f(\mathbf{X})}{f(\mathbf{X})}\right]$.

因此，基于 n 个观测值的水平为 α 的 GWSR 检验和 Hotelling 的 T^2 检验，其检验功效收敛于：

$$\mathbb{P}\left(\left\|\Sigma_{\mathrm{ERD}}^{-1/2}\boldsymbol{\gamma}+N(\mathbf{0},I_p)\right\|^2\geq\chi_p^2(\alpha)\right),\quad\text{and}\quad\mathbb{P}\left(\left\|\Sigma_{\mathbf{X}}^{-1/2}\boldsymbol{\xi}+N(\mathbf{0},I_p)\right\|^2\geq\chi_p^2(\alpha)\right),$$

分别地，其中 $\chi_p^2(\alpha)\in\mathbb{R}$ 满足 $\mathbb{P}\left(\chi_p^2\geq\chi_p^2(\alpha)\right)=\alpha$。

如果我们使用 C_1n 个观测进行 GWSR 检验，并采用某个 $C_1>0$，那么在 $\boldsymbol{\theta}_n=\frac{\boldsymbol{\xi}_n}{\sqrt{n}}=\frac{\sqrt{C_1}\boldsymbol{\xi}_n}{\sqrt{C_1n}}$ 下，其中 $\sqrt{C_1}\boldsymbol{\xi}_n\to\sqrt{C_1}\boldsymbol{\xi}$，水平为 α 的 GWSR 检验功效收敛于

$$\mathbb{P}\left(\left\|\Sigma_{\mathrm{ERD}}^{-1/2}\sqrt{C_1}\boldsymbol{\gamma}+N(\mathbf{0},I_p)\right\|^2\geq\chi_p^2(\alpha)\right).$$

注意，对于给定的 $\beta\in(\alpha,1)$，存在唯一的 $C_1>0$，使得

$$\mathbb{P}\left(\left\|\Sigma_{\mathrm{ERD}}^{-1/2}\sqrt{C_1}\boldsymbol{\gamma}+N(\mathbf{0},I_p)\right\|^2\geq\chi_p^2(\alpha)\right)=\beta.$$

类似地，存在唯一的 $C_2>0$，使得在 $\boldsymbol{\theta}_n=\frac{\boldsymbol{\xi}_n}{\sqrt{n}}$ 下，基于 C_2n 个观测的水平为 α 的 Hotelling T^2 检验，其检验功效收敛于

$$\mathbb{P}\left(\left\|\Sigma_{\mathbf{X}}^{-1/2}\sqrt{C_2}\boldsymbol{\xi}+N(\mathbf{0},I_p)\right\|^2\geq\chi_p^2(\alpha)\right)=\beta.$$

这意味着，为了达到相同的检验功效 β，$\|\Sigma^{-1/2}{}_{\mathrm{ERD}}\sqrt{C_1}\boldsymbol{\gamma}\|=\|\Sigma^{-1/2}{}_{\mathbf{X}}\sqrt{C_2}\boldsymbol{\xi}\|$，并且

$$\lim_{n\to\infty}\frac{N_2\left(\alpha,\beta,\frac{\boldsymbol{\xi}_n}{\sqrt{n}}\right)}{N_1\left(\alpha,\beta,\frac{\boldsymbol{\xi}_n}{\sqrt{n}}\right)}=\lim_{n\to\infty}\frac{C_2n+o(n)}{C_1n+o(n)}=\frac{\|\Sigma_{\mathrm{ERD}}^{-1/2}\boldsymbol{\gamma}\|^2}{\|\Sigma_{\mathbf{X}}^{-1/2}\boldsymbol{\xi}\|^2}.$$

注意，对于任意满足 $\varepsilon_n\to0$ 和 $\boldsymbol{\xi}_n\to\boldsymbol{\xi}\neq0$ 的 $\varepsilon_n\boldsymbol{\xi}_n\to0$，$\{\varepsilon_n\boldsymbol{\xi}_n\}_{n\geq1}$ 的任意子序列都存在一个进一步的子序列，该子序列包含于形如 $\{\frac{\boldsymbol{\xi}_n}{\sqrt{n}}\}_{n\geq1}$ 且满足 $\widetilde{\boldsymbol{\xi}}_n\to\boldsymbol{\xi}$的序列中。这意味着 $\frac{1}{\sqrt{n}}$ 的速率并不重要，并且我们还有 $\lim_{n\to\infty}\frac{N_2(\alpha,\beta,\varepsilon_n\boldsymbol{\xi}_n)}{N_1(\alpha,\beta,\varepsilon_n\boldsymbol{\xi}_n)}=\frac{\|\Sigma^{-1/2}{}_{\mathrm{ERD}}\boldsymbol{\gamma}\|^2}{\|\Sigma^{-1/2}{}_{\mathbf{X}}\boldsymbol{\xi}\|^2}$。

B.10 命题证明 3.2

注意，本证明中使用的所有得分函数 $J(\cdot)$ 均满足假设 3，因此定理 3.4适用。

第 1 部分。（具有恒等评分函数的高斯 ERD）：由定理 2.1 可知，$\nabla\psi$ 是在平方损失下将 $\mu = \mu_S$ 传输到 $\nu_S = N(\mathbf{0},\ I)$ 的最优传输映射（在 H_0, 下），它也是将 μ 映射到 $N(\mathbf{0},\ I)$，且可表示为某个凸函数 [73]的梯度的唯一映射。

在这个高斯情形下，$\nabla\psi(\mathbf{x}) = \Sigma^{-1/2}\mathbf{x}$，它将 $N(\mathbf{0},\ \Sigma)$ 映射到 $N(\mathbf{0},\ I)$，并且是某个凸函数的梯度。令 $\mathbf{b} := \Sigma^{-1/2}\boldsymbol{\xi}$ 且 $\mathbf{G} \sim N(\mathbf{0},\ I)$。于是：

$$\begin{aligned}\|\Sigma_{\mathrm{ERD}}^{-1/2}\boldsymbol{\gamma}\|^2 = \|\boldsymbol{\gamma}\|^2 = \left\|\mathbb{E}_{\mathrm{H}_0}\left[\nabla\psi(\mathbf{X})\frac{\boldsymbol{\xi}^\top\nabla f(\mathbf{X})}{f(\mathbf{X})}\right]\right\|^2 &= \left\|\mathbb{E}\left[\Sigma^{-1/2}\mathbf{X}\cdot\boldsymbol{\xi}^\top\Sigma^{-1}\mathbf{X}\right]\right\|^2 \\ &= \left\|\mathbb{E}\left[\mathbf{G}\cdot\mathbf{b}^\top\mathbf{G}\right]\right\|^2 \\ &= \|\mathbf{b}\|^2 = \|\Sigma^{-1/2}\boldsymbol{\xi}\|^2.\end{aligned}$$

这意味着渐近相对效率$(\mathbf{W}_n, \overline{\ }\mathbf{X}_n;\ \boldsymbol{\xi})\ = 1$。

第 2 部分。（均匀 ERD）：在这种情形下，我们仍有 $\nu_S = N(\mathbf{0},\ I)$，因此 $\nabla\psi(\mathbf{x}) = \Sigma^{-1/2}\mathbf{x}$。对于 ERD =均匀分布$(-1, {}^-1)^p$, $\Sigma_{\mathrm{ERD}} = \frac{1}{3}I$。利用分部积分公式，

$$\begin{aligned}\|\Sigma_{\mathrm{ERD}}^{-1/2}\boldsymbol{\gamma}\|^2 = 3\|\boldsymbol{\gamma}\|^2 &= 3\left\|\mathbb{E}_{\mathrm{H}_0}\left[J(\nabla\psi(\mathbf{X}))\frac{\boldsymbol{\xi}^\top\nabla f(\mathbf{X})}{f(\mathbf{X})}\right]\right\|^2 \\ &= 3\sum_{j=1}^p\left(\int J(\nabla\psi(\mathbf{x}))_j\sum_{i=1}^p\xi_i\frac{\partial}{\partial x_i}f(\mathbf{x})\mathrm{d}\mathbf{x}\right)^2 \\ &= 3\sum_{j=1}^p\left(\sum_{i=1}^p\mathbb{E}\left[\xi_i\frac{\partial}{\partial X_i}J(\nabla\psi(\mathbf{X}))_j\right]\right)^2 \\ &= 3\sum_{j=1}^p\left(\sum_{i=1}^p\mathbb{E}\left[\xi_i\frac{\partial}{\partial X_i}\left(2\Phi(e_j^\top\Sigma^{-1/2}\mathbf{X})-1\right)\right]\right)^2 \\ &= 3\sum_{j=1}^p\left(\sum_{i=1}^p\mathbb{E}\left[2\xi_i\Sigma_{ji}^{-1/2}\phi(e_j^\top\Sigma^{-1/2}\mathbf{X})\right]\right)^2,\end{aligned}$$

其中，在第三行中我们使用了分部积分。这里，$\phi(\cdot)$ 表示标准正态分布的密度函数。根据 Deb 等人 [25, 附录]$\mathbb{E}\left[\phi$中的公式 (C.32)$(e_j^\top\Sigma^{-1/2}\mathbf{X})\right] = \frac{1}{2\sqrt{\pi}}$，上述公式可进一步简化为

$$3\sum_{j=1}^p\left(\sum_{i=1}^p\xi_i\Sigma_{ji}^{-1/2}\frac{1}{\sqrt{\pi}}\right)^2 = \frac{3}{\pi}\|\Sigma^{-1/2}\boldsymbol{\xi}\|^2.$$

从而得到所需结果。

第 3 部分。（球面 rical 均匀 ERD 与恒等评分函数）：写出 $\nabla\psi(\mathbf{x}) = (r_1(\mathbf{x})\ldots, r_p(\mathbf{x}))$。

对于球面均匀分布 $\Sigma_{\mathrm{ERD}}=\frac{1}{3p}I$。使用分部积分公式，

$$\begin{aligned}\|\Sigma_{\mathrm{ERD}}^{-1/2}\boldsymbol{\gamma}\|^2=3p\|\boldsymbol{\gamma}\|^2&=3p\left\|\mathbb{E}_{\mathrm{H}_0}\left[\nabla\psi(\mathbf{X})\frac{\boldsymbol{\xi}^\top\nabla f(\mathbf{X})}{f(\mathbf{X})}\right]\right\|^2\\&=3p\sum_{j=1}^{p}\left(\sum_{i=1}^{p}\mathbb{E}\left[\xi_i\frac{\partial}{\partial X_i}r_j(\mathbf{X})\right]\right)^2\\&=3\|\Sigma^{-1/2}\boldsymbol{\xi}\|^2\cdot\frac{1}{p}\left[\frac{1}{2^{p-1}}\cdot\frac{\Gamma(p-0.5)}{(\Gamma(p/2))^2}+\frac{\sqrt{2\pi}(p-1)}{2^{p/2}\Gamma(p/2)}\mathbb{E}[|G_1|^{p-2}]-\mathcal{C}_p\right]^2,\end{aligned}$$

其中最后一个等式由 Deb 等人 [25, 附录中的公式 (C.44)]及其中的记号可得。因此，渐近相对效率$(\mathbf{W}_n,\mathbf{X}_n;\boldsymbol{\xi})=\kappa_p$。□

B.11 定理3.5的证明

部分 1。（ERD $=$均匀分布$(-1,1)^p$，采用恒等评分函数）：在此情形下 $\Sigma_{\mathrm{ERD}}=\frac{1}{3}I$，且可以证明

$$\nabla\psi(\mathbf{x})=\left(2F_1(x_1)-1,\ldots,2F_p(x_p)-1\right)^\top,$$

其中 F_i 是 f_i的累积分布函数（cdf）。为说明这一点，注意上述映射将 μ映射为均匀分布$(-1,1)^p$，且它是以下凸函数的梯度：$2\sum_{i=1}^{p}\int_{-\infty}^{x_i}F_i(t_i)\mathrm{d}t_i-\mathbf{1}^\top\mathbf{x}$。因此，

$$\begin{aligned}\left\|\Sigma_{\mathrm{ERD}}^{-1/2}\mathbb{E}_{\mathrm{H}_0}\left[\nabla\psi(\mathbf{X})\frac{\boldsymbol{\xi}^\top\nabla f(\mathbf{X})}{f(\mathbf{X})}\right]\right\|^2&=3\left\|\mathbb{E}_{\mathrm{H}_0}\left[\nabla\psi(\mathbf{X})\frac{\boldsymbol{\xi}^\top\nabla f(\mathbf{X})}{f(\mathbf{X})}\right]\right\|^2\\&=3\sum_{i=1}^{p}\left(\mathbb{E}_{\mathrm{H}_0}\left[(2F_i(X_i)-1)\frac{\boldsymbol{\xi}^\top\nabla f(\mathbf{X})}{f(\mathbf{X})}\right]\right)^2\\&=12\sum_{i=1}^{p}\left(\mathbb{E}_{\mathrm{H}_0}\left[F_i(X_i)\frac{\boldsymbol{\xi}^\top\nabla f(\mathbf{X})}{f(\mathbf{X})}\right]\right)^2\\&=12\sum_{i=1}^{p}\left(\int F_i(x_i)\xi_i\widetilde{f}_i'(x_i)\mathrm{d}x_i\right)^2\\&=12\sum_{i=1}^{p}\left(\xi_i\int\widetilde{f}_i^2(x_i)\mathrm{d}x_i\right)^2.\end{aligned}$$

设 σ_i^2为 F_i的方差。则 $\|\Sigma^{-1/2}\mathbf{x}\,\boldsymbol{\xi}\|^2=\boldsymbol{\xi}^\top\Sigma^{-1}\mathbf{x}\;\boldsymbol{\xi}=\sum_{i=1}^{p}\frac{\xi_i^2}{\sigma_i^2}$。

从 Wilcoxon 检验相对于 t一维情形下的 -检验 [57], 的渐近相对效率推导可知:

$$\sigma_i^2\left(\int\widetilde{f}_i^2(x_i)\mathrm{d}x\right)^2\geq\frac{9}{125}.$$

这意味着

$$\mathrm{ARE}(\mathbf{W}_n,\overline{\mathbf{X}}_n;\boldsymbol{\xi})=\frac{\left\|\Sigma_{\mathrm{ERD}}^{-1/2}\mathbb{E}_{\mathrm{H}_0}\left[\nabla\psi(\mathbf{X})\frac{\boldsymbol{\xi}^\top\nabla f(\mathbf{X})}{f(\mathbf{X})}\right]\right\|^2}{\|\Sigma_{\mathbf{X}}^{-1/2}\boldsymbol{\xi}\|^2}\geq\frac{108}{125}.$$

第 2 部分。（具有恒等评分函数的高斯 ERD）：在这种情况下 $\Sigma_{\mathrm{ERD}}=I$，并且

$$\nabla\psi(\mathbf{x})=\left(\Phi^{-1}\circ F_1(x_1),\ldots,\Phi^{-1}\circ F_p(x_p)\right).$$

为说明这一点，请注意，上述映射将 μ 推送至 $N(\mathbf{0},\ I)$，且它是凸函数$\psi(\mathbf{x})$：
$=\sum_{i=1}^{p}\int_{-\infty}^{x_i}\Phi^{-1}\circ F_i(t_i)\mathrm{d}t_i$ 的梯度。因此，

$$\begin{aligned}\left\|\Sigma_{\mathrm{ERD}}^{-1/2}\mathbb{E}_{\mathrm{H}_0}\left[\nabla\psi(\mathbf{X})\frac{\boldsymbol{\xi}^\top\nabla f(\mathbf{X})}{f(\mathbf{X})}\right]\right\|^2&=\left\|\mathbb{E}_{\mathrm{H}_0}\left[\nabla\psi(\mathbf{X})\frac{\boldsymbol{\xi}^\top\nabla f(\mathbf{X})}{f(\mathbf{X})}\right]\right\|^2\\&=\sum_{i=1}^{p}\left(\mathbb{E}_{\mathrm{H}_0}\left[\Phi^{-1}\big(F_i(X_i)\big)\frac{\boldsymbol{\xi}^\top\nabla f(\mathbf{X})}{f(\mathbf{X})}\right]\right)^2\\&=\sum_{i=1}^{p}\left(\int\Phi^{-1}\big(F_i(x_i)\big)\xi_i\widetilde{f}_i'(x_i)\,\mathrm{d}x_i\right)^2\\&=\sum_{i=1}^{p}\left(\xi_i\int\frac{1}{\phi\left(\Phi^{-1}\big(F_i(x_i)\big)\right)}\widetilde{f}_i^2(x_i)\,\mathrm{d}x_i\right)^2,\end{aligned}\tag{B.10}$$

其中，最后一个等式由分部积分得到，且 ϕ 是 $N(0,\ 1)$ 的密度。由一维优化问题 [37, 定理 2。1] 可知

$$\sigma_i^2\left(\int\frac{1}{\phi\left(\Phi^{-1}\big(F_i(x_i)\big)\right)}\widetilde{f}_i^2(x_i)\,\mathrm{d}x_i\right)^2\geq 1,\tag{B.11}$$

对于任意 $\sigma>0$，极小化密度是 $N(\mathbf{0},\ \sigma^2)$ 的密度。这意味着：

$$\mathrm{ARE}(\mathbf{W}_n,\overline{\mathbf{X}}_n;\boldsymbol{\xi})=\frac{\sum_{i=1}^{p}\left(\xi_i\int\frac{1}{\phi\left(\Phi^{-1}\left(F_i(x_i)\right)\right)}\widetilde{f}_i^2(x_i)\,\mathrm{d}x_i\right)^2}{\sum_{i=1}^{p}\frac{\xi_i^2}{\sigma_i^2}}\geq 1.\tag{B.12}$$

□

B.12 定理3.6的证明

我们对 f 作出与 Deb 等人 [25, 附录 C.53]中相同的正则性条件：

- $\int_{\mathbb{R}^+}r^{d+1}f(r)\mathrm{d}r<\infty$。
- $\sqrt{f}(\cdot)$ 存在一个弱导数，记为 $\left(\sqrt{f}\right)'(\cdot)$。这意味着，对于所有 $\psi:\ \mathbb{R}^+\to\mathbb{R}^+$，其中 $\mathbb{R}^+\to\mathbb{R}^+$ 具有紧支撑且无穷次可微，我们有

$$\int_{\mathbb{R}^+}\sqrt{\underline{f}(r)}\psi'(r)\,\mathrm{d}r=-\int_{\mathbb{R}^+}\left(\sqrt{\underline{f}}\right)'(r)\psi(r)\mathrm{d}r.$$

- $\int_{\mathbb{R}^+}r^{d-1}\left[\left(\sqrt{\underline{f}}\right)'(r)\right]^2\mathrm{d}r<\infty.$

第 1 部分。（具有恒等评分函数的球面均匀 ERD）：设 $\mathbf{X}:=\Sigma^{-1/2}(\mathbf{X}-\boldsymbol{\theta})$，其在原假设下具有球面对称分布，且其协方差矩阵为 $\frac{1}{p}\mathbb{E}\|\mathbf{X}\|^2I_p$。因此，$\Sigma_{\mathbf{X}}=\frac{1}{p}\mathbb{E}\|\mathbf{X}\|^2\cdot\Sigma$ 和

$$\|\Sigma_{\mathbf{X}}^{-1/2}\boldsymbol{\xi}\|^2=\boldsymbol{\xi}^\top\Sigma_{\mathbf{X}}^{-1}\boldsymbol{\xi}=\frac{p}{\mathbb{E}\|\overline{\mathbf{X}}\|^2}\cdot\boldsymbol{\xi}^\top\Sigma^{-1}\boldsymbol{\xi}.$$

引自 Deb 等人 [25, 附录中的式 (C.55)]，

$$\left\|\Sigma_{\mathrm{ERD}}^{-1/2}\mathbb{E}_{\mathrm{H}_0}\left[\nabla\psi(\mathbf{X})\frac{\boldsymbol{\xi}^\top\nabla f(\mathbf{X})}{f(\mathbf{X})}\right]\right\|^2=\frac{3}{p}\cdot\boldsymbol{\xi}^\top\Sigma^{-1}\boldsymbol{\xi}\left(\mathbb{E}\overline{h}\left(\|\overline{\mathbf{X}}\|\right)+(p-1)\mathbb{E}\left[\frac{\overline{H}\left(\|\overline{\mathbf{X}}\|\right)}{\|\overline{\mathbf{X}}\|}\right]\right)^2,$$

其中，$\overline{H}$ 是 $\|\mathbf{X}\|$ 的分布函数，h 是相应的密度函数。因此，

$$\begin{aligned}\mathrm{ARE}(\mathbf{W}_n, \overline{\mathbf{X}}_n; \boldsymbol{\xi}) &= \frac{\left\|\Sigma_{\mathrm{ERD}}^{-1/2}\mathbb{E}_{\mathrm{H}_0}\left[\nabla\psi(\mathbf{X})\frac{\boldsymbol{\xi}^\top\nabla f(\mathbf{X})}{f(\mathbf{X})}\right]\right\|^2}{\|\Sigma_{\mathbf{X}}^{-1/2}\boldsymbol{\xi}\|^2} \\ &= \frac{3}{p^2}\cdot\mathbb{E}\|\overline{\mathbf{X}}\|^2\cdot\left(\mathbb{E}\overline{h}\left(\|\overline{\mathbf{X}}\|\right)+(p-1)\mathbb{E}\left[\frac{\overline{H}\left(\|\overline{\mathbf{X}}\|\right)}{\|\overline{\mathbf{X}}\|}\right]\right)^2 \\ &\geq \frac{81}{500}\cdot\frac{(\sqrt{2p-1}+1)^5}{p^2(\sqrt{2p-1}+5)},\end{aligned}$$

其中，不等式由 Deb 等人所求解的优化问题得出。[25, 附录中的式 (C.56)]。使其达到最小值的密度也由 Deb 等人给出。[25, 附录 C.53]。

Part 2.（具有恒等评分函数的高斯 ERD）：根据 Deb 等人 [25, 引理 C.2]，

$$\nabla\psi(\mathbf{X}) = \frac{\Sigma^{-1/2}(\mathbf{X}-\boldsymbol{\theta})}{\|\Sigma^{-1/2}(\mathbf{X}-\boldsymbol{\theta})\|}\cdot H_p^{-1}\left(\overline{H}\left(\|\Sigma^{-1/2}(\mathbf{X}-\boldsymbol{\theta})\|\right)\right),$$

其中，H_p 是一个 $\sqrt{\chi_p^2}$ 分布的分布函数。根据 Deb 等人，[25, 附录中的式 (C.60)]:

$$\begin{aligned}\mathrm{ARE}(\mathbf{W}_n, \overline{\mathbf{X}}_n; \boldsymbol{\xi}) &= \frac{\left\|\Sigma_{\mathrm{ERD}}^{-1/2}\mathbb{E}_{\mathrm{H}_0}\left[\nabla\psi(\mathbf{X})\frac{\boldsymbol{\xi}^\top\nabla f(\mathbf{X})}{f(\mathbf{X})}\right]\right\|^2}{\|\Sigma_{\mathbf{X}}^{-1/2}\boldsymbol{\xi}\|^2} \\ &= \frac{\mathbb{E}\|\overline{\mathbf{X}}\|^2}{p^3}\cdot\left\{\mathbb{E}\left[\frac{\overline{h}\left(\|\overline{\mathbf{X}}\|\right)}{h_p(H_p^{-1}(\overline{H}(\|\overline{\mathbf{X}}\|)))}\right]+(p-1)\mathbb{E}\left[\frac{H_d^{-1}\circ\overline{H}\left(\|\overline{\mathbf{X}}\|\right)}{\|\overline{\mathbf{X}}\|}\right]\right\}^2 \\ &\geq 1.\end{aligned}$$

当且仅当 Deb 等人的 [25, 附录中的式 (C.60)] 中的下确界达到时，等式成立；也就是说，当 $f(\cdot)$是一个多元正态分布时 [25, 定理 3.5]。 □

B.13 定理3.7 的证明

设 $f_i(\cdot)$ 的方差为 σ_i^2。则 $\mathbf{W}$ 的协方差矩阵为 $\mathrm{diag}(\sigma_1^2,\ldots,\sigma_p^2)$，且 $\Sigma_{\mathbf{X}} = \mathrm{Cov}(A\mathbf{W}) = A\,\mathrm{diag}(\sigma_1^2,\ldots,\sigma_p^2)A^\top$，$\|\Sigma^{-1/2}\mathbf{x}\boldsymbol{\xi}\|^2 = \boldsymbol{\xi}^\top A\,\mathrm{diag}\left(\frac{1}{\sigma_1^2},\ldots,\frac{1}{\sigma_p^2}\right)A^\top\boldsymbol{\xi}$。注意，在原假设下，对于 $A=(a_{i,j})$，当$\mathbf{x}=(x_1,\ldots,x_p)\in\mathbb{R}^p$，

$$f(\mathbf{x}) = \prod_{i=1}^p \widetilde{f}_i\left(\sum_{j=1}^p a_{j,i}x_j\right).$$

因此，使用 $w_i=\sum_{j=1}^p a_{j,\,i}x_j$ （或 $\mathbf{w}=A^\top\mathbf{x}$），

$$\frac{\boldsymbol{\xi}^\top\nabla f(\mathbf{X})}{f(\mathbf{X})} = \sum_{i,j=1}^p \xi_i a_{i,j}\frac{\widetilde{f}_j'}{\widetilde{f}_j}(w_j).$$

令 F_i 为与 f_i对应的分布函数。令$R_{\mathbf{W}}(\mathbf{w}) = \left(\Phi^{-1}\circ\widetilde{F}_1(w_1)\ldots,\Phi^{-1}\circ\widetilde{F}_p(w_p)\right)$。根据 Ghosal 和 Sen [38, 引理 A.8], $\nabla\psi(\mathbf{X}) = AR_{\mathbf{W}}(A^\top\mathbf{X})$。

因此，

$$\begin{aligned}\left\|\Sigma_{\mathrm{ERD}}^{-1/2}\mathbb{E}_{\mathrm{H}_0}\left[\nabla\psi(\mathbf{X})\frac{\boldsymbol{\xi}^\top\nabla f(\mathbf{X})}{f(\mathbf{X})}\right]\right\|^2 &= \left\|\mathbb{E}_{\mathrm{H}_0}\left[AR_{\mathbf{W}}(A^\top\mathbf{X})\sum_{i,j=1}^{p}\xi_i a_{i,j}\frac{\widetilde{f}_j'}{\widetilde{f}_j}(W_j)\right]\right\|^2 \\ &= \left\|\mathbb{E}_{\mathrm{H}_0}\left[R_{\mathbf{W}}(\mathbf{W})\sum_{i,j=1}^{p}\xi_i a_{i,j}\frac{\widetilde{f}_j'}{\widetilde{f}_j}(W_j)\right]\right\|^2 \\ &= \sum_{k=1}^{p}\left(\mathbb{E}_{\mathrm{H}_0}\left[\Phi^{-1}\circ\widetilde{F}_k(W_k)\sum_{i,j=1}^{p}\xi_i a_{i,j}\frac{\widetilde{f}_j'}{\widetilde{f}_j}(W_j)\right]\right)^2.\end{aligned}$$

在原假设下，当 $j\neq k$ 时，W_j 与 W_k 相互独立，因此 $\mathbb{E}_{\mathrm{H}_0}\left[\Phi^{-1}\circ\widetilde{F}_k(W_k)\frac{\widetilde{f}_j'}{\widetilde{f}_j}(W_j)\right]=0$。于是

$$\begin{aligned}\mathbb{E}_{\mathrm{H}_0}\left[\Phi^{-1}\circ\widetilde{F}_k(W_k)\sum_{i,j=1}^{p}\xi_i a_{i,j}\frac{\widetilde{f}_j'}{\widetilde{f}_j}(W_j)\right] &= \sum_{i=1}^{p}\xi_i a_{i,k}\cdot\mathbb{E}_{\mathrm{H}_0}\left[\Phi^{-1}\circ\widetilde{F}_k(W_k)\frac{\widetilde{f}_k'}{\widetilde{f}_k}(W_k)\right] \\ &\geq \sum_{i=1}^{p}\xi_i a_{i,k}\cdot\frac{1}{\sigma_k} = \frac{(A^\top\boldsymbol{\xi})_k}{\sigma_k},\end{aligned}$$

其中，不等式由 (B.10) 和 (B.11) 得出。因此，

$$\mathrm{ARE}(\mathbf{W}_n,\overline{\mathbf{X}}_n;\boldsymbol{\xi}) = \frac{\left\|\Sigma_{\mathrm{ERD}}^{-1/2}\mathbb{E}_{\mathrm{H}_0}\left[\nabla\psi(\mathbf{X})\frac{\boldsymbol{\xi}^\top\nabla f(\mathbf{X})}{f(\mathbf{X})}\right]\right\|^2}{\|\Sigma_{\mathbf{X}}^{-1/2}\boldsymbol{\xi}\|^2} \geq \frac{\sum_{k=1}^{p}\left(\frac{(A^\top\boldsymbol{\xi})_k}{\sigma_k}\right)^2}{\boldsymbol{\xi}^\top A\,\mathrm{diag}\left(\frac{1}{\sigma_1^2},\ldots,\frac{1}{\sigma_p^2}\right)A^\top\boldsymbol{\xi}} = 1.$$

□

B.14 定理的证明3.9

若模型 $(P_\theta:\boldsymbol{\theta}\in\Theta)$ 在均方意义下可微，则局部实验收敛到第 [99, 15 节3]中的高斯极限：

$$\left(P^n_{\boldsymbol{\theta}_0+\boldsymbol{\xi}/\sqrt{n}}:\boldsymbol{\xi}\in\mathbb{R}^p\right)\to\left(N(\boldsymbol{\xi},I_{\boldsymbol{\theta}_0}^{-1}):\boldsymbol{\xi}\in\mathbb{R}^p\right),$$

其中，$I_{\boldsymbol{\theta}_0}$ 是 $\boldsymbol{\theta}_0=\mathbf{0}_p$处的费 Fisher 信息。

1. 我们考虑对 (3.7) 进行假设检验。渐近表示定理 [99, 定理 15.1]表明，若存在一列水平为 α、功效函数为 $\pi_n(\cdot)$ 的检验，满足对于每个$\boldsymbol{\xi}$ (当 $n\to\infty$ 时) 及某个函数 π，均有 $\pi_n(\boldsymbol{\xi}/\sqrt{n})\to\pi(\boldsymbol{\xi})$，则 $\pi(\cdot)$ 是高斯极限实验中水平为 α 的检验的功效函数。现在考虑高斯极限实验：对于基于观测 $\mathrm{Y}\sim N(\boldsymbol{\xi},I_{\boldsymbol{\theta}_0}^{-1})$、检验 $\mathrm{H}_0:\boldsymbol{\xi}=\mathbf{0}_p$ 对 $\mathrm{H}_1:\boldsymbol{\xi}\neq\mathbf{0}_p$ 的问题，在 $\{\boldsymbol{\xi}:\|I_{\boldsymbol{\theta}_0}^{1/2}\boldsymbol{\xi}\|=c\}$ 上最大化最小检验功效的水平为 α 的检验，其临界域为 $\{\mathrm{y}\in\mathbb{R}^p:\|I_{\boldsymbol{\theta}_0}^{1/2}\,\mathrm{y}\|^2\geq\chi_p^2(\alpha)\}$，对所有 $c>0$ [99,问题 15.4] 均成立。这意味着

$$\inf_{\boldsymbol{\xi}:\|I_{\boldsymbol{\theta}_0}^{1/2}\boldsymbol{\xi}\|=c}\lim_{n\to\infty}\pi_n\left(\frac{\boldsymbol{\xi}}{\sqrt{n}}\right)\leq\mathbb{P}\left(\|N(\boldsymbol{\xi}_c,I_p)\|^2\geq\chi_p^2(\alpha)\right),\tag{B.13}$$

对于任意满足 $\|\boldsymbol{\xi}_c\|=c$ 的 $\boldsymbol{\xi}_c\in\mathbb{R}^p$。

接下来，我们将说明，通过选择适当的得分函数 $J(\cdot)$，我们可以达到上述最优检验功效（即（B.13）右侧的值）。根据定理 3.4，在 H_1，

$$\mathbf{W}_n\xrightarrow{d}N\left(\boldsymbol{\gamma},\Sigma_{\mathrm{ERD}}\right),$$

其中 $\boldsymbol{\gamma} = -\mathbb{E}_{\mathrm{H}_0}\left[J(\nabla\psi(\mathbf{X}))\boldsymbol{\xi}^\top \nabla f(\mathbf{X})\, f(\mathbf{X}) \right] \in \mathbb{R}^p$。在定理 2.1 中，我们证明 $\nabla\psi(\mathbf{x})$ 是平方损失下的最优传输映射，它将 $P_{\boldsymbol{\theta}_0}$ 推送至 ν_S。如果 ν_S 具有勒贝格密度，那么 McCann 定理 [73, 注释 16] 指出，$\nabla\psi(\cdot)$ 存在几乎处处逆映射 $\nabla\psi^*: \mathbb{R}^p \to \mathbb{R}^p$，使得 $\nabla\psi^*(\nabla\psi(\mathbf{x})) = \mathbf{x}$（$P_{\boldsymbol{\theta}_0}$ 几乎处处对 $\mathbf{x}$）且 $\nabla\psi \circ \nabla\psi^*(\mathbf{x}) = \mathbf{x}$（$\nu_S$ 几乎处处对 $\mathbf{x}$）。

对于 $J(\cdot) = -\frac{\nabla f}{(\nabla\psi^*(\cdot))\, f(\nabla\psi^*(\cdot))}$，$J(\nabla\psi(\mathbf{x})) = -\frac{\nabla f(\mathbf{x})}{f}(\mathbf{x})$ ($P_{\boldsymbol{\theta}_0}$a.e.-$\mathbf{x}$)。注意到，

$$J(Q\nabla\psi(\mathbf{x})) = J(\nabla\psi(Q\mathbf{x})) = -\frac{\nabla f(Q\mathbf{x})}{f(Q\mathbf{x})},$$

对任意 $Q \in \mathcal{G}$。由于 f 是 $\mathcal{G}$-对称的，对所有 $Q \in \mathcal{G}$ 和 $\mathbf{x} \in \mathbb{R}^p$，均有 $f(Q\mathbf{x}) = f(\mathbf{x})$，这进一步意味着 $\nabla f(Q\mathbf{x}) = Q\nabla f(\mathbf{x})$。因此，$J(Q\psi(\mathbf{x})) = QJ(\psi(\mathbf{x}))$，这表明 $J(\cdot)$ 是 $\mathcal{G}$-等变的（ν_S几乎处处）。取此 J，

$$\boldsymbol{\gamma} = \mathbb{E}_{\mathrm{H}_0}\left[\frac{\nabla f(\mathbf{X})}{f(\mathbf{X})}\frac{\nabla f(\mathbf{X})^\top}{f(\mathbf{X})}\right]\boldsymbol{\xi} = I_{\boldsymbol{\theta}_0}\boldsymbol{\xi},$$
$$\Sigma_{\mathrm{ERD}} = \mathrm{Var}_{\mathrm{H}_0}\left(J(\nabla\psi(\mathbf{X}))\right) = \mathrm{Var}_{\mathrm{H}_0}\left(\frac{\nabla f(\mathbf{X})}{f(\mathbf{X})}\right) = I_{\boldsymbol{\theta}_0}.$$

因此，在 H_1, $\mathbf{W}_n \xrightarrow{d} N(I_{\boldsymbol{\theta}_0}\boldsymbol{\xi}, I_{\boldsymbol{\theta}_0})$ 下，故

$$\mathbb{P}_{\boldsymbol{\theta}_0+\frac{\boldsymbol{\xi}}{\sqrt{n}}}\left(\mathbf{W}_n^\top \Sigma_{\mathrm{ERD}}^{-1}\mathbf{W}_n \geq \chi_p^2(\alpha)\right) \to \mathbb{P}\left(\|N(I_{\boldsymbol{\theta}_0}^{1/2}\boldsymbol{\xi}, I_p)\|^2 \geq \chi_p^2(\alpha)\right).$$

结合式（B.13），可得

$$\inf_{\boldsymbol{\xi}:\|I_{\boldsymbol{\theta}_0}^{1/2}\boldsymbol{\xi}\|=c} \lim_{n\to\infty} \pi_n\left(\frac{\boldsymbol{\xi}}{\sqrt{n}}\right) \leq \inf_{\boldsymbol{\xi}:\|I_{\boldsymbol{\theta}_0}^{1/2}\boldsymbol{\xi}\|=c} \lim_{n\to\infty} \mathbb{P}_{\boldsymbol{\theta}_0+\frac{\boldsymbol{\xi}}{\sqrt{n}}}\left(\mathbf{W}_n^\top \Sigma_{\mathrm{ERD}}^{-1}\mathbf{W}_n \geq \chi_p^2(\alpha)\right),$$

由此得到所需结果。

2. 回顾一下，$\pi_n(\cdot)$ 是一列水平为 α 的检验（针对式（3.8））的检验功效函数。根据 Van der Vaart [99,定理 15。4],

$$\limsup_{n\to\infty} \pi_n\left(\boldsymbol{\theta}_0 + \frac{\boldsymbol{\xi}}{\sqrt{n}}\right) \leq 1 - \Phi\left(z_\alpha - \frac{\boldsymbol{\xi}^\top\nabla\zeta(\boldsymbol{\theta}_0)}{\sqrt{\nabla\zeta(\boldsymbol{\theta}_0)^\top I_{\boldsymbol{\theta}_0}^{-1}\nabla\zeta(\boldsymbol{\theta}_0)}}\right), \tag{B.14}$$

其中，$\Phi(\cdot)$ 是标准正态分布的累积分布函数，且 z_α 满足 $\Phi(z_\alpha) = 1-\alpha$。

根据第 1 部分的证明，在 $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \frac{\boldsymbol{\xi}}{\sqrt{n}}$ 条件下，$\mathbf{W}_n \xrightarrow{d} N(I_{\boldsymbol{\theta}_0}\boldsymbol{\xi}, I_{\boldsymbol{\theta}_0})$，因此

$$\nabla\zeta(\boldsymbol{\theta}_0)^\top I_{\boldsymbol{\theta}_0}^{-1}\mathbf{W}_n \xrightarrow{d} N\left(\boldsymbol{\xi}^\top\nabla\zeta(\boldsymbol{\theta}_0), \nabla\zeta(\boldsymbol{\theta}_0)^\top I_{\boldsymbol{\theta}_0}^{-1}\nabla\zeta(\boldsymbol{\theta}_0)\right).$$

这意味着

$$\lim_{n\to\infty} \mathbb{P}_{\boldsymbol{\theta}_0+\frac{\boldsymbol{\xi}}{\sqrt{n}}}\left(\frac{\nabla\zeta(\boldsymbol{\theta}_0)^\top I_{\boldsymbol{\theta}_0}^{-1}\mathbf{W}_n}{\sqrt{\nabla\zeta(\boldsymbol{\theta}_0)^\top I_{\boldsymbol{\theta}_0}^{-1}\nabla\zeta(\boldsymbol{\theta}_0)}} \geq z_\alpha\right) = 1 - \Phi\left(z_\alpha - \frac{\boldsymbol{\xi}^\top\nabla\zeta(\boldsymbol{\theta}_0)}{\sqrt{\nabla\zeta(\boldsymbol{\theta}_0)^\top I_{\boldsymbol{\theta}_0}^{-1}\nabla\zeta(\boldsymbol{\theta}_0)}}\right),$$

其达到了（B.14）中的上界。

□

C 辅助结果

下面我们陈述并证明一些辅助结果。McCann [73, 引理 9] 和 Villani [101, 定理 510]中可以找到类似形式的结果。

命题 C.1。 如果 $c(\cdot,\ \cdot)$ 是连续的，那么存在 $\pi \in \Pi(\mu,\ \nu)$，其具有 c-循环单调支撑（见定义 2.5）。

证明。 令 $\mu_n := \frac{1}{n}\sum_{i=1}^n \delta_{\mathbf{x}_i}$ 和 $\nu_n := \frac{1}{n}\sum_{i=1}^n \delta_{\mathbf{y}_i}$ 满足 $\mu_n \xrightarrow{w} \mu$ 和 $\nu_n \xrightarrow{w} \nu$（例如，我们可以从 μ 中独立同分布地抽取$\mathbf{x}_i$，并从 ν 中独立同分布地抽取$\mathbf{y}_i$）。对于每个 n，求解

$$\min_{a_{ij}\geq 0} \sum_{i,j=1}^n a_{ij} c(\mathbf{x}_i, \mathbf{y}_j), \qquad \text{s.t. } \sum_{j=1}^n a_{ij} = 1, i = 1, \ldots, n, \text{ and } \sum_{i=1}^n a_{ij} = 1, j = 1, \ldots, n \tag{C.1}$$

以得到一个极小化解 (a_{ij}^n)。事实上，这是指派问题 [78, 11]，并且存在一个最优解，使得对于所有 i, j，$a_{ij}^n \in \{0, 1\}$。令 $\pi_n := \sum_{i,j=1}^n a_{ij}^n \delta_{(\mathbf{x}_i, \mathbf{y}_j)}$。那么 π_n 支撑于 S，即所有满足 $a_{ij}^n > 0$ 的有序对 $(\mathbf{x}_i, \mathbf{y}_j)$ 构成的集合。为证明 $\pi_n \in \Pi(\mu_n, \nu_n)$ 具有 c-循环单调支撑，我们采用反证法。假设存在 $N \in \mathbb{N}$ 和 $(\mathbf{x}_{i_1}, \mathbf{y}_{i_1}) \ldots, (\mathbf{x}_{i_N}, \mathbf{y}_{i_N})$ 属于 S，使得（其中 $\mathbf{y}_{i_{N+1}} = \mathbf{y}_{i_1}$）

$$c(\mathbf{x}_{i_1}, \mathbf{y}_{i_2}) + c(\mathbf{x}_{i_2}, \mathbf{y}_{i_3}) + \ldots + c(\mathbf{x}_{i_N}, \mathbf{y}_{i_{N+1}}) < c(\mathbf{x}_{i_1}, \mathbf{y}_{i_1}) + \ldots + c(\mathbf{x}_{i_N}, \mathbf{y}_{i_N}).$$

定义

$$\widetilde{\pi}_n := \pi_n + \varepsilon \sum_{l=1}^N \left(\delta_{(\mathbf{x}_{i_l}, \mathbf{y}_{i_{l+1}})} - \delta_{(\mathbf{x}_{i_l}, \mathbf{y}_{i_l})} \right).$$

假设 $\widetilde{\pi}_n = \sum_{i,\ j=1}^n \widetilde{a}_{ij}^n \delta_{(\mathbf{x}_i, \mathbf{y}_j)}$。那么，对于足够小的 $\varepsilon > 0$，$\widetilde{a}_{ij}^n \geq 0$，且 $(\widetilde{a}_{ij}^n)$ 在 (a_{ij}^n) 中的代价小于 (C.1) 的代价，由此产生矛盾。

因此，$\pi_n \in \Pi(\mu_n, \nu_n)$ 的支撑是 c-循环单调的。μ_n 和 ν_n的弱收敛意味着 $\{\mu_n\}$ 和 $\{\nu_n\}$ 是紧的，因此 $\{\pi_n\}$ 必定是紧的（Villani [101, 引理 4.4]）。根据普罗霍罗夫定理，存在一个子序列 $\{\pi_{n_k}\}_{k\geq 1}$，其弱收敛至某个概率测度 π。由于 $(\mathbf{x}, \mathbf{y}) \mapsto \mathbf{x}$ 和 $(\mathbf{x}, \mathbf{y}) \mapsto \mathbf{y}$ 是连续的，根据连续映射定理，π的边缘分布为 μ 和 ν。对于每个 n，π_n 的 c-循环单调性意味着，对于所有 $N \in \mathbb{N}$ 以及 $\pi_n^{\otimes N}$-几乎处处的 $(\mathbf{x}_1, \mathbf{y}_1) \ldots, (\mathbf{x}_N, \mathbf{y}_N)$，我们有（其中 $\mathbf{y}_{N+1} = \mathbf{y}_1$）

$$\sum_{i=1}^N c(\mathbf{x}_i, \mathbf{y}_i) \leq \sum_{i=1}^N c(\mathbf{x}_i, \mathbf{y}_{i+1}). \tag{C.2}$$

换言之，$\pi_n^{\otimes N}$集中在集合 $\mathcal{C}(N)$ 上，该集合由所有满足 (C.2) 的 $((\mathbf{x}_1, \mathbf{y}_1) \ldots, (\mathbf{x}_N, \mathbf{y}_N)) \in (\mathbb{R}^p \times \mathbb{R}^p)^N$组成。由于 $c(\cdot, \cdot)$ 是连续的，$\mathcal{C}(N)$ 是闭集合，因此根据 Portmanteau 定理，$\pi_{n_k}^{\otimes N}$ 的弱极限 $\pi^{\otimes N}$也集中在 $\mathcal{C}(N)$ 上。令 π 的支撑为 Γ。于是，$\Gamma^N = (\text{supp}(\pi))^N = \text{supp}(\pi^{\otimes N}) \subset \mathcal{C}(N)$。由于这对所有 N 都成立，Γ 是 c-循环单调的。

□

命题 C.2。 若 $c(\cdot,\ \cdot)$ 是连续的，且 $\pi \in \Gamma(\mu,\ \nu)$ 具有 c-循环单调支撑，则存在一个 c-凸函数 ψ，使得 $\partial_c \psi$（见定义B.1）包含 π的支撑。

证明。 令 Γ 表示 π 的支撑。固定 $(\mathbf{x}_0, \mathbf{y}_0) \in \Gamma$。定义

$$\begin{aligned}\psi(\mathbf{x}) := \sup_{m\in\mathbb{N}} \sup \big\{ &[c(\mathbf{x}_0, \mathbf{y}_0) - c(\mathbf{x}_1, \mathbf{y}_0)] + [c(\mathbf{x}_1, \mathbf{y}_1) - c(\mathbf{x}_2, \mathbf{y}_1)] \\ &+ \ldots + [c(\mathbf{x}_m, \mathbf{y}_m) - c(\mathbf{x}, \mathbf{y}_m)] : (\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_m, \mathbf{y}_m) \in \Gamma \big\}.\end{aligned}$$

根据定义，对于 $m = 1$ 以及 $(\mathbf{x}_1, \mathbf{y}_1) = (\mathbf{x}_0, \mathbf{y}_0)$，我们有 $\psi(\mathbf{x}_0) \geq 0$。另一方面，$c$-循环单调性 Γ 意味着 $\psi(\mathbf{x}_0) \leq 0$，因此 $\psi(\mathbf{x}_0) = 0$。将 $\mathbf{y}_m$ 重命名为 $\mathbf{y}$ 后，

$$\begin{aligned}\psi(\mathbf{x}) &= \sup_{\mathbf{y}\in\mathbb{R}^p} \sup_{m\in\mathbb{N}} \sup_{(\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_{m-1}, \mathbf{y}_{m-1}), \mathbf{x}_m} \big\{ [c(\mathbf{x}_0, \mathbf{y}_0) - c(\mathbf{x}_1, \mathbf{y}_0)] + [c(\mathbf{x}_1, \mathbf{y}_1) - c(\mathbf{x}_2, \mathbf{y}_1)] \\ &\quad + \ldots + [c(\mathbf{x}_m, \mathbf{y}) - c(\mathbf{x}, \mathbf{y})] : (\mathbf{x}_1, \mathbf{y}_1), \ldots, (\mathbf{x}_m, \mathbf{y}) \in \Gamma \big\} \\ &= \sup_{\mathbf{y}\in\mathbb{R}^p} [\zeta(\mathbf{y}) - c(\mathbf{x}, \mathbf{y})],\end{aligned}$$

其中 $\zeta(\mathbf{y}) := -\infty$ 如果不存在满足 $(\mathbf{x}_m, \mathbf{y}) \in \Gamma$ 的 $\mathbf{x}_m$，则为此情形；否则

$$\begin{aligned}\zeta(\mathbf{y}) := \sup_{m\in\mathbb{N}} \sup_{(\mathbf{x}_1,\mathbf{y}_1),\ldots,(\mathbf{x}_{m-1},\mathbf{y}_{m-1}),\mathbf{x}_m} \big\{ &[c(\mathbf{x}_0,\mathbf{y}_0) - c(\mathbf{x}_1,\mathbf{y}_0)] + [c(\mathbf{x}_1,\mathbf{y}_1) - c(\mathbf{x}_2,\mathbf{y}_1)] \\ &+ \ldots + [c(\mathbf{x}_m,\mathbf{y}) - c(\mathbf{x},\mathbf{y})] : (\mathbf{x}_1,\mathbf{y}_1),\ldots,(\mathbf{x}_m,\mathbf{y}) \in \Gamma \big\}.\end{aligned}$$

因此 $\psi(\cdot)$ is 是一个 c-凸函数。对于任意 $(\overline{\mathbf{x}}, \overline{\mathbf{y}}) \in \Gamma$，若 $(\mathbf{x}_m, \mathbf{y}_m) = (\mathbf{x}, \overline{\mathbf{y}})$ 满足 ψ 的定义，

$$\begin{aligned}\psi(\mathbf{x}) &\geq \sup_{m\geq 2} \sup_{(\mathbf{x}_1,\mathbf{y}_1),\ldots,(\mathbf{x}_{m-1},\mathbf{y}_{m-1})\in\Gamma} \big\{ [c(\mathbf{x}_0,\mathbf{y}_0) - c(\mathbf{x}_1,\mathbf{y}_0)] + [c(\mathbf{x}_1,\mathbf{y}_1) - c(\mathbf{x}_2,\mathbf{y}_1)] \\ &\quad + \ldots + [c(\mathbf{x}_{m-1},\mathbf{y}_{m-1}) - c(\overline{\mathbf{x}},\mathbf{y}_{m-1})] + [c(\overline{\mathbf{x}},\overline{\mathbf{y}}) - c(\mathbf{x},\overline{\mathbf{y}})] \big\} \\ &= \psi(\overline{\mathbf{x}}) + [c(\overline{\mathbf{x}},\overline{\mathbf{y}}) - c(\mathbf{x},\overline{\mathbf{y}})].\end{aligned}$$

因此，$\psi(\mathbf{x}) + c(\mathbf{x}, \mathbf{y}) \geq \psi(\overline{\mathbf{x}}) + c(\mathbf{x}, \overline{\mathbf{y}})$ 对 $\mathbf{x} \in \mathbb{R}^p$ 取下确界，可得 $\psi^c(\mathbf{y}) \geq \psi(\overline{\mathbf{x}}) + c(\mathbf{x}, \mathbf{y})$。根据定义 $\overline{\psi}^c$（定义 B.1），反向不等式始终成立，因此有 $\psi^c(\mathbf{y}) = \psi(\overline{\mathbf{x}}) + c(\mathbf{x}, \mathbf{y})$。这意味着 $(\mathbf{x}, \mathbf{y}) \in \partial_c\psi$，因此 $\Gamma \subset \partial_c\psi$。 □

参考文献

[1] Ajne, B.（1968）。圆周分布均匀性的简单检验。《生物计量学》，55：343–354。[2] Albisetti, I., Balabdaoui, F. 和 Holzmann, H.（2020）。球对称性和椭圆对称性检验。《多元分析杂志》，180：104667，19页。[3] Anderson, R. 和 Klee Jr, V.（1952）。凸函数与上半连续集族。《杜克数学杂志》，19(2)：349–357。[4] Arbuthnott, J.（1710年）。从男女出生人数中观察到的恒定规律论证神的眷佑。《哲学汇刊》，27.328：186–190。[5] Bach, F. R. 和 Jordan, M. I.（2003）。核独立成分分析。《机器学习研究杂志》，3(1)：1–48。[6] Baringhaus, L.（1991）。多元分布的球对称性检验。《统计年刊》，19(2)：899–917。[7] Bellhouse, D.（2001）。John Arbuthnot。载于《世纪统计学家》，第39–42页。施普林格。[8] Bennett, B. M.（1962）。关于多元符号检验。《英国皇家统计学会会刊*B*辑》，24：159–161。[9] Beran, R. J. 和 Millar, P. W.（1997）。多元对称模型。载于《*Lucien Le Cam*纪念文集》，第13–42页。施普林格。[10] Berk, R. 和 Hwang, J. T.（1989）。最小二乘估计量的最优性。《多元分析杂志》，30(2)：245–254。[11] Bertsekas, D. P.（1988）。拍卖算法：一种用于指派问题的分布式松弛方法。《运筹学年刊》，14(1-4)：105–123。[12] Bickel, P. J.（1965）。关于 Hotelling T^2 的一些渐近非参数竞争方法。《数学统计年刊》，36：160–173。[13] Blough, D. K.（1989）。通过投影寻踪实现多元对称性。《统计数学研究所年刊》，41(3)：461–475。[14] Blumen, I.（1958）。一种新的双变量符号检验。美国统计学会会刊，53(282)：448–456。[15] Boeckel, M., Spokoiny, V. 和 Suvorikova, A.（2018）。多元 Brenier 累积分布函数及其在非参数假设检验中的应用。*arXiv*预印本 *arXiv:1809.04090*。[16] Cardoso, J.-F.（1998）。盲信号分离：统计原理。《*IEEE*会刊》，86(10)：2009–2025。[17] Chakraborty, B.（2001）。关于仿射等变多元分位数。《统计数学研究所年刊》，53(2)：380–403。[18] Chakraborty, B. 和 Chaudhuri, P.（1996）。构造仿射等变多元中位数的一种变换与逆变换技术。《美国数学会会刊》，124(8)：2539–2547。[19] Chakraborty, B., Chaudhuri, P. 和 Oja, H.（1998）。空间中位数的变换—逆变换操作与角度检验。《统计学辛尼卡》，8(3)：767–784。

[20] Chatterjee, S. K. (1966). 位置的双变量符号检验。《数学统计年刊》，37:1771–1782.[21] Chaudhuri, P. and Sengupta, D. (1993). 多维符号检验：基于数据云几何的推断。《美国统计学会会刊》，88(424):1363–1370.[22] Chernoff, H. and Savage, I. R. (1958). 某些非参数检验统计量的渐近正态性与效率。《数学统计年刊》，29:972–994.[23] Comon, P. and Jutten, C. (2010). 《盲源分离手册：独立成分分析及其应用》。爱思唯尔科学出版社.[24] Cs¨org˝o, S. and Heathcote, C. R. (1987). 对称性检验。《生物计量学》，74(1):177–184.[25] Deb, N., Bhattacharya, B. B., and Sen, B. (2023). 基于最优传输的多元分布无关检验的 Pitman 效率下界。arXiv 预印本 arXiv:2104.01986.[26] Deb, N., Ghosal, P., and Sen, B. (2021). 通过重心投影使用代入估计量估计最优传输映射的速率。神经信息处理系统进展，34:29736–29753.[27] Deb, N. and Sen, B. (2023). 使用测度传输的基于多元秩的分布无关非参数检验。《美国统计学会会刊》，118(541):192–207.[28] Diks, C. and Tong, H. (1999). 多元概率分布对称性的检验。《生物计量学》，86(3):605–614.[29] Dyckerhoff, R., Ley, C., and Paindaveine, D. (2015). 双变量中心对称性的基于深度的游程检验。《统计数学研究所年刊》，67(5):917–941.[30] Eaton, M. L. (1989). 《群不变性在统计学中的应用》，NSF-CBMS 概率与统计区域会议系列第1卷。数理统计学会，Hayward, CA；美国统计学会，Alexandria, VA.[31] Efron, B. (1969). 对称条件下的 Student t-检验。《美国统计学会会刊》，64:1278– 1302.[32] Einmahl, J. H. J. and Gan, Z. (2016). 中心对称性检验。《统计规划与推断杂志》，169:27–33.[33] Fang, K. T., Kotz, S., and Ng, K. W. (1990). 《对称多元分布及相关分布》，统计学与应用概率专著第36卷。Chapman and Hall, Ltd., 伦敦.[34] Feuerverger, A. and Mureika, R. A. (1977). 经验特征函数及其应用。《统计年刊》，5(1):88–97.[35] Fire, M. (2006). 有符号排列群上的统计学。载于第18届形式幂级数与代数组合国际年会，FPSAC 2006.[36] Fourdrinier, D. and Wells, M. T. (1995). 一般线性模型中球对称分布损失函数的估计。《统计年刊》，23(2):571–592.[37] Gastwirth, J. L. and Wolff, S. S. (1968). 求得秩检验渐近功效下界的一种初等方法。《数学统计年刊》，39:2128–2130.[38] Ghosal, P. and Sen, B. (2019). 使用最优传输的多元秩与分位数：一致性、速率及非参数检验。arXiv 预印本 arXiv:1905.05340.[39] Ghosh, S. and Ruymgaart, F. H. (1992). 经验特征函数在若干多元问题中的应用。《加拿大统计学杂志》，20(4):429–440.[40] Gibbons, J. D. and Chakraborti, S. (2003). 《非参数统计推断》，统计学：教科书与专著第168卷。Marcel Dekker, Inc., 纽约，第四版.[41] Gibbons, J. D. and Chakraborti, S. (2011). 《非参数统计推断》。载于：Lovric, M. (编)《国际统计科学百科全书》，第977–979页。施普林格柏林海德堡，柏林，海德堡.[42] H´ajek, J., ˇSid´ak, Z., and Sen, P. K. (1999). 《秩检验理论》。概率与数理统计。学术出版社，圣迭戈，CA，第二版.[43] Hallin, M., del Barrio, E., Cuesta-Albertos, J., and Matr´an, C. (2021). 多维 d中的分布函数与分位数函数、秩与符号：测度传输方法。《统计年刊》，49(2):1139–1165.[44] Hallin, M., Hlubinka, D., and Hudecov´a, ˇS. (2020). 用于多输出回归和多元方差分析的完全分布无关中心向外秩检验。arXiv 预印本 arXiv:2007.15496.[45] Hallin, M., La Vecchia, D., and Liu, H. (2022a). 半参数 VARMA 模型的中心向外 R 估计。《美国统计学会会刊》，117(538):925–938.

[46] Hallin, M.、Liu, H. 和 Verdebout, T.（2022b）。基于非参数测度传输的方向数据方法。arXiv预印本 arXiv:2212.10345。[47] Hallin, M. 和 Paindaveine, D.（2002）。基于交错方向和伪马氏秩的多元位置最优检验。《统计年刊》，30(4)：1103–1133。[48] Halton, J. H.（1964）。算法247：底数取逆准随机点序列。《计算机协会通讯》，7(12)：701–702。[49] Hardy, G. H.、Littlewood, J. E. 和 P´olya, G.（1952）。《不等式》。剑桥：大学出版社，第二版。[50] Hastings, W. K.（1970）。使用马尔可夫链的蒙特卡罗抽样方法及其应用。《生物计量学》，57(1)：97–109。[51] Heathcote, C. R.、Rachev, S. T. 和 Cheng, B.（1995）。多元对称性检验。《多元分析杂志》，54(1)：91–112。[52] Henze, N.、Hl´avka, Z. 和 Meintanis, S. G.（2014）。通过经验特征函数检验球面对称性。《统计学》，48(6)：1282–1296。[53] Henze, N.、Klar, B. 和 Meintanis, S. G.（2003）。基于经验特征函数检验关于未指定点的对称性不变检验。《多元分析杂志》，87(2)：275–297。[54] Henze, N. 和 Mayer, C.（2020）。关于未知中心多元反射对称性的HKM检验又有新进展。《统计数学研究所年刊》，72(3)：741–770。[55] Hiriart-Urruty, J.-B. 和 Lemar´echal, C.（1993）。《凸分析与最小化算法》第一卷，数学科学基础丛书第305卷。施普林格出版社，柏林。[56] Hodges, Jr., J. L.（1955）。双变量符号检验。《数学统计年刊》，26：523–527。[57] Hodges, Jr., J. L. 和 Lehmann, E. L.（1956）。某些非参数竞争方法相对于 t-检验的效率。《数学统计年刊》，27：324–335。[58] Householder, A. S.（1958）。非对称矩阵的酉三角化。《计算机协会杂志》，5：339–342。[59] Hu˘skov´a, M.（1984）。对称性假设。载于《非参数方法》，《统计学手册》第4卷，第63–78页。爱思唯尔。[60] Hyv¨arinen, A.（2005）。通过分数匹配估计未归一化统计模型。《机器学习研究杂志》，6：695–709。[61] Hyv¨arinen, A. 和 Oja, E.（2000）。独立成分分析：算法与应用。《神经网络》，13(4)：411–430。[62] Jammalamadaka, S. R.、Tiwari, R. C. 和 Chib, S.（1987）。球面对称误差下线性模型的贝叶斯预测。《经济学快报》，24(1)：39–44。[63] Jonker, R. 和 Volgenant, A.（1987）。一种用于稠密和稀疏线性指派问题的最短增广路径算法。《计算》，38(4)：325–340。[64] Kariya, T. 和 Eaton, M. L.（1977）。球面对称性的稳健检验。《统计年刊》，5(1)：206–215。[65] Kolassa, J. E.（2020）。《非参数统计学导论》。CRC出版社。[66] Kuipers, L. 和 Niederreiter, H.（1974）。《序列的均匀分布》。《纯粹与应用数学》。Wiley-Interscience，纽约-伦敦-悉尼。[67] Lehmann, E. L.（1975）。《非参数统计：基于秩的统计方法》。Holden-Day概率与统计丛书。Holden-Day公司，旧金山；麦格劳-希尔国际图书公司，纽约-杜塞尔多夫。H. J. M. d'Abrera 特别协助。¨Lehmann, E. L. 和 Romano, J. P.（2005）。《统计假设检验》。Springer统计学教材。施普林格，纽约，第三版。[68] Li, B.、Zha, H. 和 Chiaromonte, F.（2005）。等高线回归：一种通用的降维方法。《统计年刊》，33(4)：1580–1616。[69] Manole, T.、Balakrishnan, S.、Niles-Weed, J. 和 Wasserman, L.（2021）。光滑最优传输映射的插件估计。arXiv预印本 arXiv:2107.12364。[70] Marden, J. I.（1999）。多元秩检验。载于《多元分析、试验设计与调查抽样》，第401–432页。Marcel Dekker公司，纽约。[71] Mardia, K. V. 和 Jupp, P. E.（2000）。《方向统计学》。Wiley概率与统计丛书。约翰·威立父子公司，奇切斯特。

[73] McCann, R. J.（1995）。单调保测映射的存在性与唯一性。《杜克数学杂志》，80(2):309–323。[74] McWilliams, T. P.（1990）。一种基于游程统计量的分布无关对称性检验。美国统计学会会刊，85(412):1130–1133。[75] M¨ott¨onen, J. 和 Oja, H.（1995）。多元空间符号与秩方法。《非参数统计》，5(2):201–213。[76] M¨ott¨onen, J.、Oja, H. 和 Serfling, R.（2005）。多元广义空间带符号秩方法。《统计研究杂志》，39(1):25–48。[77] M¨ott¨onen, J.、Oja, H. 和 Tienari, J.（1997）。论多元空间符号与秩检验的效率。《统计年刊》，25(2):542–552。[78] Munkres, J.（1957）。指派问题与运输问题的算法。《工业与应用数学学会杂志》，5:32–38。[79] Munkres, J. R.（2000）。拓扑学。Prentice Hall, Inc.，Upper Saddle River, NJ。[80] Ngatchou-Wandji, J.（2009）。多元分布的对称性检验。《统计方法论》，6(3):230–250。[81] Niederreiter, H.（1992）。随机数生成与准蒙特卡洛方法，应用数学CBMS-NSF区域会议系列第63卷。工业与应用数学学会（SIAM），Philadelphia, PA。[82] Nordhausen, K. 和 Oja, H.（2011）。多元 l_1 方法：MNM软件包。《统计软件杂志》，43(5):1–28。[83] Oja, H.（2010）。基于空间符号与秩的R多元非参数方法，统计学讲义第199卷。施普林格，纽约。[84] Pinelis, I.（1994）。对称性条件下的极值概率问题与Hotelling的 T^2 检验。《统计年刊》，22(1):357–368。[85] Puri, M. L. 和 Sen, P. K.（1971）。多元分析中的非参数方法。约翰·威立父子公司，纽约-伦敦-悉尼。[86] Randles, R. H.（1989）。一种基于交叉方向的分布无关多元符号检验。美国统计学会会刊，84(408):1045–1050。[87] Randles, R. H.（2000）。一种更简单、仿射不变的分布无关多元符号检验。美国统计学会会刊，95(452):1263–1268。[88] Robert, C. P. 和 Casella, G.（2004）。蒙特卡罗统计方法。Springer统计学教材。施普林格出版社，纽约，第二版。[89] Rockafellar, R. T. 和 Wets, R. J.-B.（1998）。变分分析，数学科学基础丛书第317卷。施普林格出版社，柏林。[90] Samarov, A. 和 Tsybakov, A.（2004）。非参数独立成分分析。《伯努利》，10(4):565–582。[91] Sen, P. K. 和 Puri, M. L.（1967）。论多元单样本位置问题的秩次检验理论。《数学统计年刊》，38:1216–1228。[92] Serfling, R.（2010）。多元分位数及相关函数的等变性与不变性，以及标准化的作用。《非参数统计》，22(7):915–936。[93] Serfling, R. J.（2014）。多元对称性与非对称性。《Wiley统计参考：在线统计参考》。[94] Shi, H.、Drton, M.、Hallin, M. 和 Han, F.（2023）。利用中心向外象限、Spearman和Kendall统计量进行多元独立性的半参数有效检验。ECARES工作论文。[95] Shi, H.、Drton, M. 和 Han, F.（2022a）。通过中心向外秩与符号进行分布无关的一致独立性检验。美国统计学会会刊，117(537):395–410。[96] Shi, H.、Hallin, M.、Drton, M. 和 Han, F.（2022b）。关于向量独立性的普遍一致且完全分布无关的秩检验。《统计年刊》，50(4):1933–1959。[97]Shlens, J.（2014）。独立成分分析教程。arXiv预印本 arXiv:1404.2986。[98] Song, Y. 和 Ermon, S.（2019）。通过估计数据分布的梯度进行生成建模。神经信息处理系统进展，32。[99] Van der Vaart, A. W.（1998）。渐近统计学，剑桥统计学丛书第3卷。

与概率数学。剑桥大学出版社，剑桥。[100] Villani, C. (2003). 最优运输专题，数学研究生丛书第58卷。美国数学学会，罗得岛州普罗维登斯。[101] Villani, C. (2009). 最优传输，数学基础丛书第338卷。施普林格出版社，柏林。[102] Watson, G. S. (1961). 圆上的拟合优度检验。《生物计量学》，48:109–114。[103] Weyl, H. (1952). 对称性。普林斯顿大学出版社，新泽西州普林斯顿。[104] Wilcoxon, F. (1945). 基于秩方法的个体比较。《生物计量学》，1(6):80–83。[105] Zeng, P. 和 Zhu, Y. (2010). 一种用于估计中心均值和中心子空间的积分变换法。《多元分析杂志》，101(1):271–290。[106] Zuo, Y. 和 Serfling, R. (2000). 关于若干稳健非参数位置测度相对于多元对称性一般概念的表现。《统计规划与推断杂志》，84(1- 2):55–79。