arXiv:2302.03592v2 [math.ST] 8 Feb 2023

# 针对双样本问题的二部排序方法

Stephan Cl´emen¸con[1], Myrto Limnios *[2], Nicolas Vayatis[3]

[1] stephan.clemencon@telecom-paris.fr，巴黎电信、LTCI、巴黎综合理工学院，19 place Marguerite Perey，帕莱索，91120，法国。[2] myli@math.ku.dk，哥本哈根大学数学科学系，Universitetsparken 5，2100 哥本哈根，丹麦。[2]nicolas.vayatis@ens-paris-saclay.fr，巴黎´萨克雷大学，巴黎´西岱大学´、巴黎萨克雷高等师范学院、法国国家科学研究中心、SSA、法国国家健康与医学研究院、博雷利中心，4 avenue des Sciences，吉夫叙尔伊韦特，91190，法国。

**摘要**

双样本问题是指检验独立样本是否来自同一（未知）分布，在许多领域都有应用。高维情形下对该问题的研究受到广泛关注，尤其是因为大数据时代的信息获取过程往往涉及多个来源，而这些来源缺乏有效控制，从而产生可能表现出严重抽样偏差的数据集。经典方法依赖于计算经验分布之间的差异度量，但会面临维数灾难；为此，我们提出一种基于统计学习并扩展秩检验的备择方法。当经过适当设计时，该方法能够在单变量情形下检测出相对于原假设的微小偏离。由于 $\mathbb{R}^d$ 上缺乏自然序，且当 $d \geq 2$ 时，该方法通过两个步骤实现。首先，为两个样本分别赋予标签（正类与负类），并将其分成两部分；随后，利用应用于第一部分的二部排序算法，借助学习得到的实值评分函数在 $\mathbb{R}^d$ 上定义一个预序；然后对剩余观测的评分应用秩检验，以检测分布中可能存在的差异。该方法学习如何将数据投影到实数轴上，其效果近似于对原始多元分布之间的似然比进行（任意单调变换），因此几乎不受维数影响；在忽略排序模型偏差问题的同时，还保留了单变量秩检验的优势。我们基于近期关于两样本线性秩过程的集中性结果，证明了非渐近误差界；实验研究表明，本文提出的方法优于作为自然竞争者的其他备择方法。

## 1 引言

（非参数）统计假设检验问题，即双样本问题，在统计学和机器学习中具有核心重要性。基于定义在同一概率空间（$\Omega$ $\mathcal{F}$， $\mathbb{P}$）上的两个相互独立的独立同分布样本 $\{\mathbf{X}_1, \ldots\ldots, \mathbf{X}_n\}$ 和 $\{\mathbf{Y}_1, \ldots\ldots, \mathbf{Y}_m\}$，这些随机变量取值于同一个空间 $\mathcal{Z}$（通常为 $\mathbb{R}^d$，满足 $d \geq 1$），并分别来自（未知的）概率分布 $G$ 和 $H$，且有 $n$， $m \geq 1$，它

*通讯作者。作者按字母顺序排列。

旨在检验分布 $G$ 和 $H$ 是否相同，即检验原假设（复合） $\mathcal{H}_0$ : $G = H$，以对抗备择假设 $\mathcal{H}_1$ : $G \neq H$。

这个一般性问题很容易表述，却无处不在。它在许多领域都有应用，尤其是临床试验：人们希望确定，在接受不同治疗的两个统计总体中进行的一组测量所呈现的波动，究竟只是抽样现象所致，还是治疗效果所致。它也可用于多模态数据融合，以判断两个数据集能否合并。事实上，选择偏差/抽样偏差如今也是机器学习领域的一大关注点。正如近期的理论和实证研究所指出的（参见例如Bertail 等人 [2021] 或 Wang 等人 [2019]），对训练数据（即使数据量庞大）的采集过程控制不当，也会严重损害所学习预测规则的泛化能力。

统计文献中已经提出了多种专门方法，参见 Lehmann 和 Romano [2005] 第 6.9 节，或 van der Vaart 和 Wellner [1996]第 3.7 章。$\widehat{\quad}$ $\widehat{\quad}$其中许多方法首先计算底层分布的非参数估计量 $G_n$ 和 $H_m$ （通常是经验分布的平滑版本），$\widehat{\quad}$ $\widehat{\quad}$随后计算二者之间的距离或信息论伪度量 $\mathcal{D}(G_n, H_m)$，以衡量其差异（例如两样本 Kolmogorov-Smirnov $\widehat{\quad}$ $\widehat{\quad}$统计量），并在统计量 $\mathcal{D}(G_n, H_m)$取值“较大”时拒绝原假设 $\mathcal{H}_0$；例如参见 Biau 和 Gyorfi [2005] 或 Gretton 等人 [2012] 。除了计算方面的困难，以及为使统计量渐近枢轴化（即其极限分布不依赖于参数）而必须确定适当的标准化之外（实际中这通常需要使用重抽样/自助法技术），实施此类插件程序时面临的主要问题与维数灾难有关。事实上，这类程序需要在维数可能极高的特征空间上对分布进行一致估计 $d \in \mathbb{N}^*$。

本文提出并分析的方法截然不同，其灵感来自单变量情形下的秩检验，参见 Lehmann 和 Romano [2005] 第 6.8 节。将 $G$ 和 $\{\mathrm{X}1, \ldots, \mathrm{X}_n\}$ 分别称为“正”分布和“正”样本，将 $H$ 和 $\{\mathrm{Y}1, \ldots, \mathrm{Y}_m\}$ 分别称为“负”分布和“负”样本，这类方法包括计算“正秩”的某个函数，即正例在大小为 $N = n + m$ 的合并样本中的秩。将这类统计量称为双样本秩统计量；在单变量双样本问题中使用它们的依据，自然在于它们在零假设下是枢轴的（即忽略并列值时，正秩在 $\mathcal{H}_0$ 下服从 $\{1, \ldots, N\}$ 上的均匀分布），因此所得检验没有偏差。双样本线性秩统计量这一概念，即通过得分生成函数 $u \in (0, 1) \mapsto \phi(u)$ 对正秩除以 $(N + 1)$ 后的值进行映射，并对映射结果求和（可加权），在统计学中已得到广泛研究。对于适当选择的 $\phi$，相关秩检验已知具有渐近功效方面的最优性质，能够刻画其检测接近零假设的备择假设的能力，参见 H´ajek 和 Sid´ak [1967] 或 van der Vaart [1998] 第 15 章。例如，广泛用于检验一个分布是否随机序较另一个分布大的 Mann-Whitney-Wilcoxon“秩和”统计量，对应于 $\phi(u) = u$ 的情形，并且在渐近检测小位移方面具有最优性。我们在此提出的方法则与之形成对比

的是基于预先给定的深度函数，将这类技术直接扩展到多元框架（参见 Liu 和 Singh [1993]）。我们的出发点是观察到，在单变量情形下，双样本秩统计量是相对于概率测度对 $(H, G)$ 的 ROC 曲线统计版本的概括。在 $\mathcal{H}_0$ 下，ROC 曲线与单位正方形 $[0,1]^2$ 的对角线重合。特别地，经过仿射变换后，秩和统计量无非就是经验 ROC 曲线下面积（简称曲线下面积）。该方法采用二分技巧，分两步实施。首先，利用一半样本学习如何借助评分函数 $s$：$\mathcal{Z} \to \mathbb{R}$ 对 $\mathcal{Z}$ 中的多元观测进行排序；该函数将 $\mathbb{R}$ 上的自然顺序映射到特征空间 $\mathcal{Z}$，例如似然比 $dG/dH(z)$（或其任一递增变换）即可做到这一点（即，来自 $H$ 或 $G$ 的观测，其评分 $s(Z)$ 越高，理想情况下就越可能是从 $G$ 抽取的 $Z$）。接下来，对另一半样本的（单变量）评分应用秩检验。因此，该方法的第一步归结为求解一个二部排序问题，通常可将其表述为 ROC/曲线下面积优化问题，参见 Cl´emen¸con 和 Vayatis [2009b]或 Cl´emen¸con 和 Vayatis [2010]。第二步是检验所得经验 ROC 曲线是否显著偏离对角线；秩检验中使用的得分生成函数 $\phi(u)$ 决定了偏离的性质。借助 Cl´emen¸con 等人 [2021]，近期建立的双样本线性秩过程（即双样本线性秩统计量的集合）的集中性结果，我们从非渐近视角分析了这一（无偏）两阶段检验程序。对于非参数备择假设类，我们证明了基于评分函数 $s(z)$的检验的第一类和第二类错误界；该评分函数使二部排序性能准则（概括 ROC 曲线）的统计对应量最大化，而该对应量的形式为双样本线性秩统计量。我们还从实证角度深入研究了所提出的两阶段方法检测“微小”偏离的能力；在一定程度上，即使在极高维度下，这种能力也得以保留。本文进行了广泛的实验研究，涵盖多种双样本问题，并将基于排序的检验与文献中记载的其他非参数方法进行比较。值得强调的是，尽管 ROC 曲线的概念已被广泛用于评估任意单变量检验统计量的优劣（并在两类误差之间寻找适当的权衡），本文提出的方法是首次尝试在一般多元框架中运用这一概念来设计检验程序。最后请注意，两阶段检验方法的一个非常初步版本曾在会议论文 Cl´emen¸con 等人 [2009] 中提出，其仅涉及曲线下面积优化和“秩和”检验统计量。本文在更一般的框架下提出基于排序的检验并分析其性能：基于 Cl´emen¸con 等人 [2021] 近期建立的结果，本文考虑了更广泛的一类基于排序的检验，对其检验功效进行了有限样本分析，并给出了更全面的实验结果。

本文结构安排如下。在第2节中，我们列出主要符号，详细回顾双样本问题的统计框架，并简要介绍传统方法，尤其是单变量情形下的秩检验。此外，我们还探讨二部排序与双样本问题之间的联系，并解释我们所倡导方法背后的依据。我们提出的新型检验程序及其理论保证将在第3节中阐述。

实验结果将在第4节中呈现并讨论，若干总结性意见则汇集于第5 节。技术证明和详细内容见附录部分。

## 2 背景与预备知识

本节介绍主要符号，并在非参数框架下对双样本问题进行形式化表述。我们简要回顾解决该问题的现有方法，特别关注当 $\mathcal{Z} \subset \mathbb{R}^d$ 时，维数为 1 至 $d$ 的情形下的秩检验，以及通过 ROC分析对其进行的解释。我们还回顾与二部排序任务相关的概念和结果；该任务可视为优化 ROC 曲线（的概括形式）的问题，因为后续章节提出并分析的方法正是基于后者。此处及全文中，任意事件 $\mathcal{E}$ 的示性函数记为 $\mathbb{I}\{\mathcal{E}\}$，任意点 $x$ 处的狄拉克质量记为 $\delta_x$，任意累积分布函数 $W(t)$ 在 $\mathbb{R} \cup \{+\infty\}$ 上的广义逆记为 $W^{-1}(u) = \inf\{t \in (-\infty, +\infty] : W(t) \geq u\}$, $u \in [0,1]$。我们分别用 $u \in \mathbb{R} \mapsto \lfloor u \rfloor$ 和 $u \in \mathbb{R} \mapsto \lceil u \rceil$ 表示向下取整函数和向上取整函数。对于任意有界函数 $\psi : (0, 1) \to \mathbb{R}$，我们还定义 $||\psi||_\infty = \sup_{u \in (0,1)} |\psi(u)|$。全文中，粗体符号表示多元变量，例如，对于一元随机变量，我们记为 $X$ 和 $Y$；而当其取值空间为多元可测空间 $\mathcal{Z}$（通常是 $\mathbb{R}^d$ 的子集，且 $d \geq 2$ 可能成立）时，则使用 $\mathbf{X}$ 和 $\mathbf{Y}$ 。

### 2.1 双样本问题——非参数表述

设 $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ 和 $\{\mathbf{Y}_1, \ldots, \mathbf{Y}_m\}$ 是从可测空间 $\mathcal{Z} \subset \mathbb{R}^d$ 上的概率分布 $H$ 和 $G$ 中抽取的相互独立的 独立同分布样本。在最一般的双样本问题中，不对分布 $H$ 和 $G$ 作任何假设，目标是检验复合假设：

$$\mathcal{H}_0 : G = H \text{ against the alternative } \mathcal{H}_1 : G \neq H \ , \tag{2.1}$$

基于这两个样本。经典方法是考虑概率（伪）度量$\mathcal{D}$ ，该度量定义在 $\mathcal{Z}$上所有概率分布构成的空间上。根据一个简单的观察，即在原假设下$\mathcal{D}(G, H) = 0$ ，一种自然的检验程序是计算分布 $G$ 和 $H$ 的估计值，通常采用经验分布：

$$\widehat{G}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\mathbf{X}_i} \text{ and } \widehat{H}_m = \frac{1}{n} \sum_{j=1}^{m} \delta_{\mathbf{Y}_j} \ ,$$

$\widehat{\ }$ $\widehat{\ }$并在 $\mathcal{H}_0$ 统计量取“较大”值时拒绝 $\mathcal{D}(G_n, H_m)$；例如，参见Biau和Gyorfi [2005] 。可以采用各种度量或伪度量来衡量两个概率分布之间的差异。关于概率测度空间中的度量及其应用的精彩论述，请参见Rachev [1991] 。典型的例子包括卡方距离、Kullback-Leibler散度、海林格距离、Kolmogorov-Smirnov距离及其以下类型的推广：

$$\mathrm{MMD}(\mathrm{G}, \mathrm{H}) = \sup_{\mathrm{f} \in \mathcal{F}} \left| \int_{\mathrm{x} \in \mathcal{Z}} \mathrm{f}(\mathrm{x}) \mathrm{G}(\mathrm{dx}) - \int_{\mathrm{x} \in \mathcal{Z}} \mathrm{f}(\mathrm{x}) \mathrm{H}(\mathrm{dx}) \right| \tag{2.2}$$

其中，$\mathcal{F}$表示一个足够丰富的函数类$f:\ \mathcal{Z}\to\mathbb{R}$，使得 MMD(G, H) =0 当且仅当 $G=H$，参见 Gretton 等人 [2012] 的定理 5。在单变量情形下，当$\mathcal{F}$ 为所有半直线构成的集合时，若 $\mathcal{F}$ 是一个特定的再生核希尔伯特空间（RKHS），则称量 (2.2) 为最大均值差异，它自然退化为通常的柯尔莫哥洛夫–斯米尔诺夫统计量。除了计算上的困难，以及为使检验统计量 (2.2) 渐近地成为枢轴量（即其极限分布与参数无关）而必须确定适当的标准化方式、并确定合适的临界阈值（实际中通常需要使用自助法技术）之外，尝试实施此类插件程序时面临的主要问题与维数灾难有关。事实上，这类程序需要在维数可能非常高的特征空间 $d\in\mathbb{N}^*$ 上对分布进行一致估计。然而，在选择再生核希尔伯特空间 $\mathcal{H}$ 的单位球作为 $\mathcal{F}$，以便高效计算上确界 (2.2) 时，在一定程度上可以规避这一困难，参见 Gretton 等人 [2007]和 Bach 等人 [2008]。最后，计算机科学文献中一个相关问题将双样本问题称为性质测试，例如参见 Rubinfeld 和 Sudan [1996],、 Goldreich 等人 [1998]。本文所采用的检验方法 (2.1) 在本质上截然不同，其灵感源自基于秩统计量概念的传统技术；为便于说明，下面先回顾其在特定单维情形下的应用。

## 2.2 单变量情形——秩检验与ROC 分析

在一维设置下，解决双样本问题的一种经典方法是根据 $\mathbb{R}$ 上的自然顺序对观测数据进行排序，并根据正例在合并样本中的秩作出决策:

$$\forall i\in\{1,\ \ldots,\ n\},\ \ \mathrm{Rank}(X_i)=N\widehat{F}_N(X_i)\ ,$$

其中 $\widehat{F}_N(t)=(1/N)(\sum_{i=1}^n\mathbb{I}\{X_i\le t\}+\sum_{j=1}^m\mathbb{I}\{Y_j\le t\})$，且 $N=n+m$。为简单起见，假设分布 $G$ 和 $H$ 均为连续分布（并列值出现的概率为零）。秩检验的基本思想很简单：在原假设 $\mathcal{H}_0$下，正例的秩与分布无关，且在$\{1,\ \ldots,\ N\}$上均匀分布。一种常见的做法是计算“正秩”之和，由此得到著名的秩和曼-惠特尼-威尔科克森统计量，参见 Wilcoxon [1945],

$$\widehat{W}_{n,m}=\sum_{i=1}^n\mathrm{Rank}(X_i)\ .\tag{2.3}$$

它被广泛用于检验 $\mathcal{H}_0$ ，以对立备择假设：两个分布之一随机序较大于另一个。若 $G(\mathrm{d}t)$ 随机序较大[1] 于 $H(\mathrm{d}t)$，即，当 $H(t)\ge G(t)$对所有 $t\in\mathbb{R}$均成立时，检验统计量$\widehat{W}_{n,\ m}$在 $\mathcal{H}_0$ 下的分布表（即便存在观测值并列的情况，也可通过将并列值赋予平均秩来处理，参见Cheung 和 Klotz [1997]），因此无需渐近近似结果，也能构造出具有适当显著性水平的检验。若两个累积分布函数满足

[1]给定概率空间上的两个分布函数 $H\ \ t\ \ G\ \ t\ \ \mathbb{R}\cup\{+\infty\}\ \ G\ \ t\ \ stochastically$ (d ) 和 (d )，当且仅当对于 $larger\ \ H\ \ t\ \ t\in\mathbb{R}\ \ G\ t\ \le H\ t\ \ H\le_{sto}G$ 任意，都有 ( ) ( ) 时，称 (d ) 随机序较大于 (d )。记作：。经典地，$G$ 随机序较大于 $H$ 的充要条件是存在 $(X,\ Y)$ 与 $(G,H)$ 的耦合，即 $i.e.$ 一对定义在同一概率空间上的随机变量，其第一和第二边缘分布分别等于 $H$ 和 $G$，并且 $X\le Y$ 以概率 1 成立。

由关系式 $G(t) = H(t-\theta)$连接，其中 $\theta \geq 0$ (即当处理效应被建模为加性时，检验 $\mathcal{H}_0$ : $\theta = 0$)时，检验统计量（2.3）在极限实验 $\theta \searrow 0$中渐近地一致最强，参见例如van der Vaart 第 15.5 节中的推论 15.14 [1998]。'正秩'的其他泛函也可用作双样本问题的检验统计量。尤其是，下文定义的双样本线性秩统计量类构成了一个丰富的泛函集合。

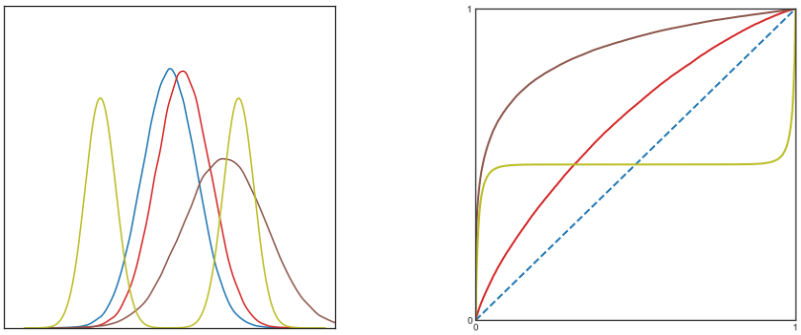
**定义 1.** (Two-sample linear rank sta-tistics) 设 $\phi\colon [0,1] \to [0,1]$ 为非递减函数。基于随机样本 $\{X_1, \ldots, X_n\}$ 和 $\{Y_1, \ldots, Y_m\}$ 的得分生成函数为'$\phi(u)$ 的双样本线性秩统计量定义为

$$\widehat{W}^{\phi}_{n,m} = \sum_{i=1}^{n} \phi\left(\frac{\mathrm{Rank}(X_i)}{N+1}\right) . \tag{2.4}$$

对于 $\phi(u) = u$，统计量 (2.4) 与 $W_{n,m}/(N+1)$ 相同。在 $\mathcal{H}_0$ 下，统计量 (2.4) 都是分布无关的，因此特别适合用于检测分布 $G$ 与 $H$之间的差异。通过列表给出它们在零假设下的分布，可以据此设计检验水平为 $\alpha \in (0,1)$ 的无偏检验。得分生成函数 $\phi$ 的选择可由可能预期的两种分布之间的差异类型（例如尺度或位置差异）来指导，并由此得到能够检测相对于 $\mathcal{H}_0$的某些类型"小幅"偏离的局部最有力检验程序。关于秩统计量渐近理论的论述，参见 Serfling 的第 9 章 [1980]或 van der Vaart 的第 13 章[1998] 。例如，对于 $u_0 \in (0,1)$，选择$\phi(u) = u \times \mathbb{I}\{u \geq u_0\}$，或选择 $\phi(u) = u^q$ 且满足 $q > 1$，会增强最高秩的作用，参见 Cl´emen.con 和 Vayatis [2007]或 Rudin [2006]。我们还要强调，Cl´emen.con 等人 [2021],近期研究了双样本线性秩过程（即双样本线性秩统计量的集合）的集中性质，研究动机是将 (2.4) 解释为与样本对 $(H,G)$ 相对应的 ROC 曲线的标量统计摘要。基于这些结果，附录部分证明了 (2.4) 的指数尾概率界（见其中的定理 10），并将其用于第 3 节的理论分析。

ROC分析。ROC 曲线是描述两个单变量概率分布 $G$ 和 $H$之间差异的金标准工具。这一具有函数性质的准则 $\mathrm{ROC}_{H,G}$可定义为概率-概率图：$t \in \mathbb{R} \mapsto (1-H(t)\ \ 1-G(t))$，其中可能出现的跳跃通过线段连接，以确保所得曲线连续。按照这一约定，可以将与分布函数对 $(H,G)$ 相关的 ROC 曲线看作一条取值于 $[0,1]$, 的 c`ad-l`ag（即右连续且左极限存在）非递减映射的图像，该映射在满足 $G \circ H^{-1}(1-\alpha) = 1-\alpha$的点 $\alpha$ 处由 $\alpha \in (0,1) \mapsto 1 - G \circ H^{-1}(1-\alpha)$ 定义。记 $\mathcal{Z}_H$ 和 $\mathcal{Z}_G$ 分别为 $H$ 和 $G$ 的支撑集。注意，该曲线在单位正方形 $[0,1]^2$ 内连接点 $(0, 1-G(\mathcal{Z}_H))$ 和 $(H(\mathcal{Z}_G), 1)$；并且在不存在平台区的情况下（为简便起见，我们在此作此假设，而不将特征空间限制为 $G$的支撑集），曲线 $\alpha \in (0,1) \mapsto \mathrm{ROC}_{G,H}(\alpha)$ 是 $\alpha \in (0,1) \mapsto \mathrm{ROC}_{H,G}(\alpha)$ 关于欧几里得平面的主对角线（即方程为"$\beta = \alpha$"的直线）作轴对称所得的像。请注意，当且仅当 $H = G$，即 $\mathcal{H}_0$ 成立时，曲线 $\mathrm{ROC}_{H,G}$ 与主对角线重合。因此，ROC 曲线这一概念提供了一种以枢轴方式考察两个分布差异的可视化工具，见图 1。例如，当且仅当曲线 $\mathrm{ROC}_{H,G}$ 始终位于主对角线上方，且 $\mathrm{ROC}_{H,G}$ 与单位正方形的左上角重合时，单变量分布 $G(dt)$ 随机序较大于 $H(dt)$

当且仅当当且仅当。分布 $H$的本质上确界小于分布 $G$的本质下确界时。



a. 概率分布　　b. ROC 曲线

图1：分布对及其对应 ROC 曲线示例。“负”分布 $H$ 以蓝色表示，三个典型的“正”分布及其对应的 ROC 曲线分别以红色（位置模型）、棕色（位置-尺度模型）和绿色（混合模型）表示。

基于样本 $\{X_1, \ldots, X_n\}$ 和 $\{Y_1, \ldots, Y_m\}$，通过计算经验累积分布函数 $H_m(t) = (1/m)\sum_{j=1}^{m} \mathbb{I}\{Y_j \le t\}$ 和 $G_n(t) = (1/n)\sum_{i=1}^{n} \mathbb{I}\{X_i \le t\}$（针对 $t \in \mathbb{R}$），并绘制经验 ROC 曲线，可得到 ROC 的统计版本$\widehat{\mathrm{H}, \mathrm{G}}$：

$$\widehat{\mathrm{ROC}}_{\mathrm{H,G}} = \mathrm{ROC}_{\widehat{\mathrm{H}}_{\mathrm{m}}, \widehat{\mathrm{G}}_{\mathrm{n}}} \ . \tag{2.5}$$

请注意，ROC 曲线 (2.5) 完全由合并样本中正例所占据的秩集合决定 $\{$秩$(X_i)$: $i = 1, \ldots\ldots, n\}$。中断

p.....的点　p分段线性曲线　(2.5) 必然　y 属于 g 属于集合　gridp点:

$$\{(j/m,\ i/n):\ j \in \{1, \ldots, m-1\},\ i \in \{1, \ldots, n-1\}\} \ .$$

将与样本 $\{X_1, \ldots, X_n\}$ 相关的次序统计量记为 $X_{(i)}$，即，Rank$(X_{(n)}) > \cdots >$秩$(X_{(1)})$，ROC 曲线 (2.5) 是连接阶梯曲线各跳跃点的连续折线:

$$\alpha \in [0,1] \mapsto \sum_{j=1}^{m} \widehat{\gamma}_j \times \mathbb{I}\{\alpha \in [(j-1)/m,\ j/m[\} \ , \tag{2.6}$$

其中，对于所有 $j \in \{1, \ldots, m\}$，我们定义

$$\widehat{\gamma}_j = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{j \ge N - \mathrm{Rank}(X_{(n-i+1)}) - i + 2\} \ .$$

经验 ROC 曲线（2.5）仅表示为秩（$X_i$）的函数。因此，它的任何汇总指标都是双样本秩统计量，也就是说，是“正秩”的可测函数。特别是，选择 ROC 曲线下面积（简称 AUC）的经验对应量作为 ROC 曲线的汇总指标十分常见，其定义如下

$$\mathrm{AUC}_{\mathrm{H,G}} := \int_0^1 \mathrm{ROC}_{\mathrm{H,G}}(\alpha)\mathrm{d}\alpha = \mathbb{P}\{\mathrm{Y} < \mathrm{X}\} + \frac{1}{2}\mathbb{P}\{\mathrm{X} = \mathrm{Y}\} \ . \tag{2.7}$$

事实上，经验 ROC 曲线（2.5）的 ROC曲线下面积，也称为一致对率，可以很容易地证明与秩和曼-惠特尼-威尔科克森统计量（2.3）仅相差一个仿射变换

$$\widehat{W}_{n,m} = nm\mathrm{AUC}_{\widehat{\mathrm{H}}_{\mathrm{m}},\widehat{\mathrm{G}}_{\mathrm{n}}} + \frac{\mathrm{n}(\mathrm{n}+1)}{2} .$$

更一般地，与不同于 $\phi(u) = u$ 的得分生成函数相关的双样本线性秩统计量（2.4），可以作为经验 ROC 曲线的替代性概括指标，并衡量单位正方形内偏离主对角线的不同方式；在零假设下，该主对角线与 $\mathrm{ROC}_{\mathrm{H},\ \mathrm{H}}$ 重合。顺便指出，在 $\mathcal{H}_0$下，我们有

$$\mathrm{AUC}_{\mathrm{H,G}} - 1/2 = \int_{\alpha=0}^{1} \{\mathrm{ROC}_{\mathrm{H,G}}(\alpha) - \alpha\}\,\mathrm{d}\alpha = 0 .$$

与（2.3）类似，统计量（2.4）是枢轴的。为了量化它们在完整样本量 $n$，$m$ 增加时的波动，必须控制合并数据集中“正”/“负”观测值的比例。令 $p \in (0,\ 1)$ 为正例的“理论”比例。对于 $N \geq 1/p$，我们假设 $n = \lfloor pN \rfloor$ 和 $m = \lceil (1-p)N \rceil = N - n$。定义混合概率分布 $F = pG + (1-p)H$。当 $N$趋于无穷时，$\widehat{W}^{\phi}_{n,\ m}/n$ 的渐近均值为：

$$\begin{aligned} W_\phi(H,G) = \mathbb{E}[\phi \circ F(X)] = \frac{1}{p}\int_0^1 \phi(u)\mathrm{d}u \\ - \frac{1-p}{p}\int_0^1 \phi\left(p(1-\mathrm{ROC}_{\mathrm{H,G}}(\alpha)) + (1-\mathrm{p})(1-\alpha)\right)\mathrm{d}\alpha , \quad (2.8) \end{aligned}$$

参见Cl´emen,con 等人 [2021]第 3 节。在所考虑的单变量分布对 $(H, G)$ 没有任何歧义时，为简洁起见，我们记作 $W_\phi$。注意，在 $\mathcal{H}_0$ 下，我们有

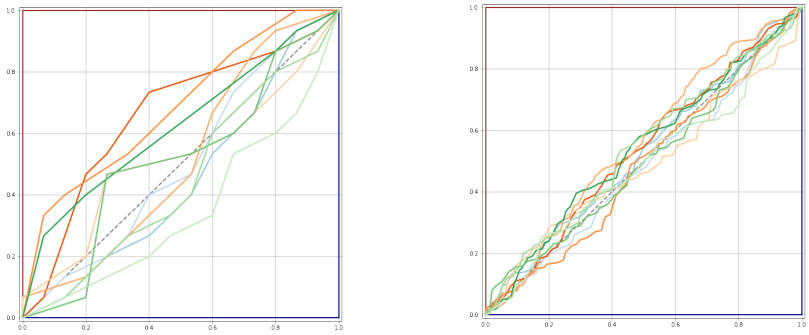
$$W_\phi = \int_0^1 \phi(u)\mathrm{d}u .$$

在 ROC 空间中构建双样本检验。在 $\mathcal{H}_0$ 下，理论 ROC 曲线与 $[0,1]^2$ 的主对角线重合：对于所有 $\alpha \in (0,1)$ 以及 $\mathbb{R}$ 上的任意分布 $H$，$\mathrm{ROC}_{\mathrm{H,H}}(\alpha) = \alpha$。此外，由于经验 ROC 曲线本身是秩的函数，因此在零假设下，它也是一个（函数型）枢轴统计量。当概率分布 $H$ 连续时，所有可能的经验 ROC 曲线在 $\mathcal{H}_0$ 下等概率。因此，在 $\binom{N}{n}$ 不太大的情况下，由于基于正样本和负样本（样本量分别为 $n$ 和 $m$）构成的所有可能经验 ROC 曲线的集合 $\mathcal{C}_{n,m}$ 的基数为 $\binom{N}{n}$，其中所有折线均可枚举（见图 2）；对于任意 $\alpha \in \left\{i/\binom{N}{n} : i = 1,\ \ldots\ldots,\ \binom{N}{n}\right\}$，都可以构造一个水平为 $\alpha$ 的容忍/预测区域 $\mathcal{R}_\alpha \subset \mathcal{C}_{n,m}$，即一个基数为 $\alpha\binom{N}{n}$ 的子集 $\mathcal{R}_\alpha \subset \mathcal{C}_{n,m}$。于是，可以考虑这样一种检验：当经验 ROC 曲线 $\widehat{\mathrm{ROC}}_{H,G}$ 落在 $\mathcal{R}_\alpha$ 之外时，拒绝 $\mathcal{H}_0$。在空间 $\mathbb{D}([0,1])$ 中构建临界域的一种自然方法是：考虑 c`ad-l`ag 映射 $[0,1] \to [0,1]$，并在水平为 $\alpha$ 下定义假设检验 $\mathcal{H}_0$；固定 $\mathbb{D}([0,1])$ 上的一个伪距离 $D$，按与第一条对角线的距离递增顺序排列 $\mathcal{C}_{n,m}$ 中的 $\binom{N}{n}$ 条曲线，并保留由最接近对角线的 $\lceil \binom{N}{n}(1-\alpha) \rceil$ 条曲线构成的子集 $\mathcal{R}_\alpha$，这里的距离按所选距离 $D$ 衡量。选择由 $L_1$-范数定义的距离时，自然会得到曼-惠特尼-威尔科克森检验。不过，为此还可以考虑许多函数距离。

**注释 1。** （The 曲线下面积 as a statistical distance）容易证明，

$$\mathrm{AUC_{H,G}} = \frac{1}{2} + \int_{-\infty}^{\infty} \{\mathrm{H(t)} - \mathrm{G(t)}\} \mathrm{dH(t)} \ , \tag{2.9}$$

更多细节请参见附录部分。因此，当 $H$ 在随机意义上小于$G$时，量曲线下面积$_{\mathrm{H,\ G}} - 1/2$ 等于累积分布函数 $L_1$（$H$）与 $H$（$t$）$and G$（$t$）之间的 $L_1$（$H$）-距离。



a. 具有 $n = m = 15$

b. 具有 $n, m = 200, 150$

图2：原假设下经验 ROC 曲线模拟示例 s.

将双样本秩检验推广到多元框架。由于只要 $d \geq 2$， $\mathbb{R}^d$ 上就不存在任何“自然顺序”，现有方法探索了多元秩统计量的不同概念。除了基于分量秩和 Copula 的方法（例如参见 Puri and Sen [1971] 或 Lung-Yut-Fong 等人 [2015]）——这些方法只有在对概率模型作出强假设时才能保证有效——统计深度旨在为多元分布 $P$（$dx$）的支撑集中的点定义一种由中心向外的排序，以模拟秩，参见 Mosler [2013]。具体而言，相对于 $P$的深度函数是一种有界、非负的博雷尔可测映射 $D_P$： $\mathbb{R}^d \to \mathbb{R}_+$，它为 $\mathbb{R}^d$ 中的多元点定义了一个预序，并有望确定任意点 $x \in \mathbb{R}^d$ 相对于概率测度 $P$的中心性。因此，靠近质量“中心”的点 $x \in \mathbb{R}^d$ 深度最大，即 $D_P$（$x$）是深度函数取值中的较高值之一。半空间深度由 Tukey 在其开创性研究中首次提出 [1975], ： $\mathbb{R}^d$ 中点 $x$ 相对于 $P$的半空间深度，是对所有满足 $x \in S$的闭半空间 $S \subset \mathbb{R}^d$ 取质量 $P$（$S$）的最小值。此后发展出许多备择方法，例如参见 Liu [1990], 、Koshevoy and Mosler [1997], 、Chaudhuri [1996], 、Oja [1983], 、Vardi and Zhang [2000], 、Chernozhukov 等人 [2017], 、Deb and Sen [2021] 或 Beirlant 等人 [2020]。Zuo and Serfling [2000], 提出了统计深度的公理化分类体系，为系统比较其优缺点提供了方法。按照 Liu 和 Singh 的建议 [1993], ，假设已选定某种深度概念，将双样本秩检验推广到多元框架的一种自然方法是：取两个可用样本中较大的一个，例如 X 样本，再用其中一部分计算相对于 $S$的深度的样本版本 $D_{\widehat{S}}$。然后，对由未参与深度估计步骤的 $\mathrm{X}_i$各点的深度值构成的一元样本，以及由 $\mathrm{Y}j$各点的深度值构成的一元样本，应用一元双样本秩检验。

基于深度的秩检验的主要局限自然在于所采用的深度概念会产生影响：不同的选择会凸显分布的不同特征，因此可能导致不同的判断。关于基于由中心向外的分布的齐性检验，请参见 Hallin 等人 [2021], 第 2.1 小节和附录A.2；若要全面了解相关内容，并回顾关于空间秩（和符号）这一概念的大量文献，可参见例如 M¨ott¨onen 和 Oja [1995], 、M¨ott¨onen 等人 [1997], 、M¨ott¨onen 等人 [2005]。这类模型通常用于检验齐性假设 $\mathcal{H}_0$ 下经典的参数备择（例如位置、尺度）。关于这些方法的全面综述，请参见 Oja [2010]；关于高维情形，请参见 Chakraborty 和 Chaudhuri [2015, 2017] 。最后，为求完整，还要指出一些研究半参数或参数秩的方法，这些方法基于马氏统计距离，但仅适用于椭圆分布类，例如参见 Um 和 Randles [1998], 、Hallin 和 Paindaveine[2002a、b， 2008]。

正如我们将看到的，下一小节概述并在第 3 节中详细分析的方法，与依赖统计深度的方法有一些相似之处；但不同的是，为了尽可能检测两个样本之间分布的偏离，用于将多元观测“投影”到实数轴上的映射是专门根据数据学习得到的。下一小节将准确说明，如何利用任何评分函数$s$： $\mathcal{Z} \to \mathbb{R}$，即与数对（$H$， $G$）相关的二部排序问题的解，来将 ROC 分析和双样本秩统计量扩展到多元框架下的双样本问题。

### 2.3 二部排序——我们方法背后的原理

二部排序的目标是基于“正例”和“负例”样本{X1, . . . , $X_n$} 和 {Y1, . . . , $Y_m$}，学习如何为任意新的观测值 Z1, . . . , $Z_k$评分。这些观测值各自为“正例”或“负例”，也就是说，在没有先验知识的情况下，它们分别来自 $G$ 或 $H$，从而以较高概率使正例大多位于所得列表的顶部。定义 $\mathcal{Z}$ 上的全预序[2] 的一种自然方法，是借助评分规则将其映射到 $\mathbb{R} \cup \{+\infty\}$ 上的自然序，即可测映射 $s : \mathcal{Z} \to (-\infty, \infty]$。于是， $\mathcal{Z}$ 上的预序 $\preccurlyeq_s$ 定义如下：对于所有 $(x, x') \in \mathcal{Z}$,$x \preccurlyeq_s x'$ ，当且仅当 $s(x) \leq s(x')$。我们用 $\mathcal{S}$ 表示所有评分函数构成的集合。候选评分函数 $s(z)$在 $\mathcal{S}$ 中区分正、负统计总体的能力，通常通过 ROC 曲线 $\mathrm{ROC}_{H_s,G_s}(\alpha) = \mathrm{ROC}(s\ \alpha)$来评估，其中 $H_s$ 和$G_s$ 分别是 $H$ 和 $G$ 经映射 $s(z)$得到的推前概率分布。它为评估排序性能提供了一种直观工具：ROC(s .)曲线越接近单位正方形的左上角，评分规则 $s$就越好。因此，ROC 曲线在所有评分函数的集合上诱导出一个偏预序：对于任意一对评分函数 $s_1$和 $s_2$，若对于所有$\alpha \in [0, 1]$，$\mathrm{ROC}(s1, \alpha) \leq \mathrm{ROC}(s2, \alpha)$，则称 $s_2$ 比 $s_1$ 更准确。根据标准的奈曼–皮尔逊论证，最准确的评分规则是似然比 $\Psi(z) = dG/dH(z)$的递增变换。具体而言，Cl´emen¸con 和 Vayatis [2009b], 在其中证明的命题 4 表明，最优元素集合为

$$\mathcal{S}^* = \left\{ s \in \mathcal{S},\ \forall (z,\ z') \in \mathcal{Z}^2,\ \Psi(z) < \Psi(z') \Rightarrow s^*(z) < s^*(z') \right\} . \tag{2.10}$$

[2]集合 $\mathcal{Z}$ 上的预序 $\preccurlyeq$ 是 $\mathcal{Z}$上的自反且传递的二元关系。若对所有 $(z,\ z') \in \mathcal{Z}^2$， $total$成立当且仅当 $z \preccurlyeq z'$ 或 $z' \preccurlyeq z$ 成立。

并且，对于所有 $(s, \alpha) \in \mathcal{S} \times (0, 1)$:

$$\mathrm{ROC}(s, \alpha) \leq \mathrm{ROC}^*(\alpha) ,$$

其中 $\mathrm{ROC}^*(\cdot) = \mathrm{ROC}(\Psi, \cdot) = \mathrm{ROC}(s^*, \cdot)$ 对任意 $s^* \in \mathcal{S}$均成立。请注意，这条最优曲线是单调不减且凹的，因此始终位于单位正方形的主对角线上方（更多性质见附录A）。现在，可以用更具定量性的方式重新表述二部排序任务：目标是基于训练随机样本 $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ 和 $\{\mathbf{Y}_1, \ldots, \mathbf{Y}_m\}$，构建评分函数 $s(z)$，使其 ROC 曲线尽可能接近 $\mathrm{ROC}^*$。衡量两条曲线偏差的一种典型方法是考虑它们在上确界范数下的距离：

$$d_\infty(s, s^*) = \sup_{\alpha \in (0,1)} |\mathrm{ROC}(s, \alpha) - \mathrm{ROC}^*(\alpha)| . \tag{2.11}$$

这个量度量的是 ROC 曲线之间的距离（或者相关评分函数等价类之间的距离；任一评分函数的 ROC 曲线在严格递增变换下保持不变），而非评分函数本身之间的距离。由于实际中未知 ROC 曲线*，主要困难在于，无法直接获得与（函数型）损失（2.11）相对应的统计量。 Cl´emen¸con 和 Vayatis [2009b]（另见Cl´emen¸con 和 Vayatis [2010]）证明，二部排序可以视为一系列代价敏感分类问题的叠加，并以自适应方式进行“离散化”，从而应用经验风险最小化，并获得 意义下的统计保证。该过程的代价是，近似步骤会带来额外的偏差项。另一种方法是，利用 ROC 空间中的 范数来衡量候选评分规则 $s$ 的性能。注意，在这种情况下，损失可以分解如下：

$$d_1(s, s^*) = \int_0^1 |\mathrm{ROC}(s, \alpha) - \mathrm{ROC}^*(\alpha)| \mathrm{d}\alpha = \int_0^1 \mathrm{ROC}^*(\alpha) \mathrm{d}\alpha - \int_0^1 \mathrm{ROC}(s, \alpha) \mathrm{d}\alpha , \tag{2.12}$$

将与最优 ROC 曲线的 $L_1$ 距离降至最低，归结为最大化曲线下面积 $\mathrm{ROC}(s, \cdot)$:

$$\mathrm{AUC}(s) := \mathrm{AUC}_{H_s, G_s} = \mathbb{P}\{s(\mathbf{Y}) < s(\mathbf{X})\} + \frac{1}{2}\mathbb{P}\{s(\mathbf{Y}) = s(\mathbf{X})\} ,$$

其中，X 和 Y 是定义在同一概率空间上的随机变量，相互独立，其分布分别为 $G$ 和 $H$；记 $G_s$ 和 $H_s$ 分别为 $s(X)$ 和 $s(Y)$ 的分布。标量性能指标 AUC(s) 在 $\mathcal{S}$ 上定义了一个全预序，其最大值记为 $\mathrm{AUC}^* = \mathrm{AUC}(s^*)$，其中 $s^* \in \mathcal{S}^*$。通过最大化 AUC 准则的经验版本来进行二部排序，已在多篇文章中得到研究，包括 Agarwal 等人 [2005] 以及 Cl´emen¸con 等人 [2008]。考虑一个类 $\mathcal{S}_0 \subset \mathcal{S}$，通过求解以下问题得到的评分规则的AUC差距置信界为：

$$\max_{s \in \mathcal{S}_0} \mathrm{AUC}_{\widehat{H}_{s,m}, \widehat{G}_{s,n}} , \tag{2.13}$$

其中 $\widehat{H}_{s,\,m} = (1/m) \sum_{j=1}^m \delta_{s(\mathbf{Y}_j)}$ 和 $\widehat{G}_{s,\,n} = (1/n) \sum_{i=1}^n \delta_{s(\mathbf{X}_i)}$ 已得到确立。请注意，在评分规则候选项类 $\mathcal{S}_0$ 上最大化 AUC(s) 的经验版本，归结为最大化秩和准则：

$$\widehat{W}_{n,m}(s) := \sum_{i=1}^n \mathrm{Rank}(s(\mathbf{X}_i)) ,$$

$\widehat{\ }$ $\widehat{\ }$其中，对于 $i \in \{1, \ldots, n\}$和 $F_s$，Rank($s(\mathrm{X}_i)$) $= nF_{s,N}(s(\mathrm{X}_i))$，$_N(t) = (1/N) \sum_{i=1}^{n} \mathbb{I}\{s(\mathrm{X}_i) \leq t\} + (1/N) \sum_{j=1}^{m} \mathbb{I}\{s(\mathrm{Y}_j) \leq t\}$ 适用于 $t \in \mathbb{R}$。Cl´emen¸con 等人（另见 Cl´emen¸con 和 Vayatis [2009a]）研究了评分规则 $s(z)$ 的性能；这些规则最大化了其他标量准则，其形式为基于单变量样本 $\{s(\mathrm{X}_1) \ldots, s(\mathrm{X}_n)\}$ 和 $\{s(\mathrm{Y}_1) \ldots, s(\mathrm{Y}_m)\}$ 的双样本线性秩统计量。具体而言，考虑一个非递减的得分生成函数 $\phi(u)$，$W_\phi$-排序性能准则定义如下：

$$W_\phi(s) = \mathbb{E}\left[(\phi \circ F_s)(s(\mathbf{X}))\right] = \frac{1}{p}\int_0^1 \phi(u)\mathrm{d}u - \frac{1-p}{p}\int_0^1 \phi\left(p(1-\mathrm{ROC(s},\ \alpha)) + (1-\mathrm{p})(1-\alpha)\right)\mathrm{d}\alpha\ , \quad (2.14)$$

其中，对于任意 $s \in \mathcal{S}$，$F_s = pG_s + (1-p)H_s$。采用这一记号，我们有$W_\phi(s) \leq W_\phi^*(H, G) := W_\phi(s^*)$，对于任意 $s^* \in \mathcal{S}^*$；当 $\phi$严格递增时，准则 $W_\phi$ 的最大化解集合与 $\mathcal{S}^*$一致，参见 Cl´emen¸con 等人 [2021]的命题 6。为简便起见，当所考虑的概率分布对 $(H, G)$ 没有歧义时，我们将 $W_\phi^*(H, G) = W_\phi^*$。当 $\phi(u) = u$时，最大化式 (2.14) 归结为进行 AUC 最大化；然而，根据所选的得分生成函数 $\phi$，评分函数 $s(z)$ 所诱导的预序中的某些具体模式可能会得到不同程度的突出。例如，对于 $\phi(u) = u^q$ 且满足 $q > 1$的情况，或任何其他在 0 附近快速趋近于 0、而在 1 附近取值高得多的得分生成函数 $\phi$，式 (2.14) 的值基本上由 ROC 曲线 ROC(s $\cdot$) 在 0 附近的行为决定（即 $s(\mathrm{X})$ 取较高值的概率），正如 Cl´emen¸con 等人 [2021]在第 2.3 小节中所讨论的那样。

**基于排序的双样本秩检验**。双样本检验程序依据这样的观察：对于适当的得分生成函数，ROC 曲线* 偏离 $[0, 1]^2$, 的主对角线，以及 $W_\phi^*$偏离 $\int_0^1 \phi\ (u)\ \mathrm{d}u$，在理论上提供了一种自然的方法来衡量 $G$ 与 $H$ 之间的差异。如下述命题所示，只要零假设成立，这些偏离就等于零。

**命题 2**。以下断言等价。

*(i)* 假设‘$\mathcal{H}_0 :\ H = G$’成立。

*(ii)* 相对于由配对$(H, G)$ 定义的二部排序问题的最优ROC 曲线与 $[0, 1]^2$的对角线重合

$$\forall \alpha \in (0, 1),\ \mathrm{ROC}^*(\alpha) = \alpha\ .$$

*(iii)* 对任意得分生成函数 $\phi(u)$，我们有

$$W_\phi^* = \int_0^1 \phi(u)\mathrm{d}u\ .$$

（*iv*）存在严格递增的得分生成函数 $\phi\ (u)$，使得：

$$W_\phi^* = \int_0^1 \phi(u)\mathrm{d}u\ .$$

（*iv*）我们有 曲线下面积$^* = 1/2$。

此外，我们还有：

$$\mathrm{AUC}^* - 1/2 = \mathbb{E}[|\Psi(\mathbf{Y}) - 1|] \ . \tag{2.15}$$

我们还回顾一下，与分布对 $(H, G)$ 相关的最优 ROC 曲线，与单变量分布对 $(H_{s^*}, G_{s^*})$ 相关的最优 ROC 曲线相同；并且对于任意 $s^* \in \mathcal{S}^*$，$\mathrm{d}G/\mathrm{d}H(z) = \mathrm{d}G_{s^*}/\mathrm{d}H_{s^*}(s^*(z))$，参见 Clémençon 和 Vayatis[2009b] 的推论 5。因此，最优 ROC 曲线$^*$ 是衡量两个多元分布之间差异的一种非常自然且全面的方式，它将 ROC 分析扩展到定义在 $\mathbb{R}$ 上的分布；该基础分析已在第 2.2 小节中回顾，下面的例子对此作了说明。

**例1.** （Multivariate Gaussian populations）考虑定义在 $\mathbb{R}^d$ 上的两个高斯分布 $H$ 和 $G$，它们具有相同的正定协方差矩阵 $\Gamma$，均值$\theta_-$ 和 $\theta_+$ 分别位于 $\mathbb{R}^d$ 中，且假设二者不同。作为似然比对数的一个递增变换，评分函数

$$z \in \mathbb{R}^d \mapsto s(z) = \langle z,\ \Gamma^{-1}(\theta_+ - \theta_-)\rangle$$

是最优的，其中 $\langle .\ ,\ . \rangle$ 表示 $\mathbb{R}^d$上的通常欧几里得内积。由于它是线性的，推前测度 $H_s$ 和 $G_s$ 都是单变量高斯分布。记$\Delta(t) = (1/\sqrt{2\pi}) \int_{-\infty}^{t} \exp(-u^2/2)\mathrm{d}u$, $t \in \mathbb{R}$为中心化标准单变量高斯分布的累积分布函数，可以立即验证，对于所有 $\alpha \in (0,1)$，最优ROC 曲线为：

$$\mathrm{ROC}^*(\alpha) = 1 - \Delta\left(\Delta^{-1}(1-\alpha) - \sqrt{\mathrm{s}(\theta_+ - \theta_-)}\right).$$

此外，我们还有

$$\mathrm{AUC}^* - 1/2 = 1 - \exp\{\mathrm{s}(\theta_+ - \theta_-)\} \ .$$

因此，命题 2允许以多种方式重新表述非参数检验问题（2.1），例如：

$$\mathcal{H}_0 : \ \mathrm{AUC}^* = 1/2 \quad vs. \quad \mathcal{H}_1 : \mathrm{AUC}^* > 1/2 \ , \tag{2.16}$$

或者等价地，写作：

$$\mathcal{H}_0 : \ W_\phi^* = \int_0^1 \phi(u)\mathrm{d}u \quad vs. \quad \mathcal{H}_1 : \ W_\phi^* > \int_0^1 \phi(u)\mathrm{d}u \ , \tag{2.17}$$

对于任意给定的严格递增评分生成函数 $\phi(u)$。值得注意的是，上述表述都是单向的，这与单变量情形下的经典秩和检验形成对比：事实上，$G^*$ 总是随机序大于 $H^*$，而对于任意两个单变量概率分布，其中一个不一定随机序大于另一个。

由于实际中未知最优 ROC 曲线及其汇总指标（如曲线下面积 AUC$^*$ 或 $W_\phi^*$），我们提出一种分两步实施的双样本问题求解方法。将样本 $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ 和 $\{\mathbf{Y}_1, \ldots, \mathbf{Y}_m\}$ 分成两半：1）首先，基于“正”样本和“负”样本的前半部分求解二部排序问题，得到评分函数 $\widehat{s}(z)$，具体如前一小节所述；2）然后，使用所得评分函数 $\widehat{s}(z)$映射剩余样本数据，并对其进行单变量秩检验。后续各节将从理论和实证两方面证明，除了近乎无偏之外，这类检验程序还能检测出相对于零假设极其微小的偏离。

# 3 基于排序的秩检验：双样本问题

本节将详细介绍前一节观察结果所引出的双样本方法论，并讨论其可能的实际应用。随后，我们将在特定假设下，从非渐近的角度分析其理论性质（检验水平和检验功效）。

## 3.1 方法与实现

我们说明如何根据两个独立的2.3 中概述的一般思路，基于对两个独立独立同分布 样本 $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ 和 $\{\mathbf{Y}_1, \ldots, \mathbf{Y}_m\}$ 的观测，来有效地应用这一思路，其中 $n,\ m \geq 1$。令 $\alpha \in (0,1)$ 为目标显著性水平，即例如所期望的第一类错误率。如前所述，检验程序主要涉及两个要素：

1. 一个二部排序算法 $\mathcal{A}: \mathcal{Z}^{n+m} \to \mathcal{S}_0$ 作用于一个类 $\mathcal{S}_0 \subset \mathcal{S}$，并为任意一组训练观测值 $\mathscr{D}_{n,\ m} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \cup \{\mathbf{y}_1, \ldots, \mathbf{y}_m\}$ 指定一个评分函数 $\mathcal{A}(\mathscr{D}_{n,\ m})$，其取值属于 $\mathcal{S}_0$；2. 一个显著性水平为 $\alpha \in (0,1)$ 的双样本秩检验 $\Phi_\alpha^\phi$：$(-\infty,\ +\infty]^{n+m} \to \{0,1\}$ 其检验结果取决于任意合并单变量数据集 $\mathscr{D}_{n,\ m} = \{x_1,\ \ldots,\ x_n\} \cup \{y_1,\ \ldots,\ y_m\}$ 的秩 $\mathrm{Rank}(x_1) \ldots, \mathrm{Rank}(x_n)\}$ 以及得分生成函数 $\phi$。

有了这两个组成部分，依赖于基于排序的双样本秩检验计算的方法分为两个主要步骤实施，如图 3所示。

在详细讨论上述一般原则可能采取的各种形式之前，有必要先作几点说明。

说明 *2.* (Bipartite ranking algorithms)如第*2.3*节所述，统计学习文献中记载的绝大多数二部排序算法，都是针对特定评分函数类 $\mathcal{S}_0$ 求解 $M$-估计问题。所要最大化的准则是曲线下面积，或其某种（平滑化／凹化／惩罚化）变体，例如式 *(2.14)*；其最优元素集合与 $\mathcal{S}^*$ 的某个子集相同。参见 *Freund* 等人 *[2003]*、*Rakotomamonjy [2004]*、*Rudin* 等人 *[2005]*、*Rudin [2006]* 或 *Burges* 等人*[2007]*。在各种复杂度假设下，已有研究为 $\mathcal{S}_0$ 的经验最大化者的缺陷建立了置信上界形式的泛化结果，参见 *Cl´emen¸con* 等人 *[2008]*、*Agarwal* 等人 *[2005]*、*Cl´emen¸con* 和 *Vayatis [2007]*、*Cl´emen¸con* 等人 *[2011]*、*Menon* 和 *Williamson [2016]* 以及 *Cl´emen¸con* 等人 *[2021]*。对于其他方法，也已建立了更强的理论保证（即上确界范数偏差 *(2.11)* 的界）；这类方法将 *ROC* 优化视为一系列代价敏感二元分类问题，并结合 $M$-估计与非线性逼近方法，参见 *Cl´emen¸con* 和 *Vayatis [2009b]*、*Cl´emen¸con* 和 *Vayatis [2010]*、*Cl´emen¸con* 和 *Vayatis [2009c]*，或 *Cl´emen¸con* 等人 *[2013a]*。

**注 3。** (2-split trick) 如上所述，近乎最优的评分函数通常通过 $M$-估计技术来学习。因此，它们对训练观测值的依赖可能很复杂，通常很难明确表示。为此，采用 2-折分技巧，以便于分析量 (3.2)的波动。于是，在给定二部排序步骤中所用子样本的条件下，泛函 (3.2) 是一个双样本秩统计量。我们还要强调，

**基于排序的双样本秩检验**

**输入。** 两个独立且独立同分布的 样本 $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ 和 $\{\mathbf{Y}_1, \ldots, \mathbf{Y}_m\}$，样本量为 $n$，$m \geq 2$，取值于 $\mathcal{Z}$；子样本量为 $n' < n$ 和 $m' < m$；在评分函数类 $\mathcal{S}_0$ 上运行的二部排序算法 $\mathcal{A}$，该类定义在 $\mathcal{Z}$上；检验水平为 $\alpha \in$ 的一元双样本秩检验 $\Phi_\alpha^\phi$ (0,1)。

**二分技巧。** 将每个原始样本分成两个子样本

$$\{\mathbf{X}_1, \ldots, \mathbf{X}_{n'}\} \cup \{\mathbf{X}_{1+n'}, \ldots, \mathbf{X}_n\} \text{ and } \{\mathbf{Y}_1, \ldots, \mathbf{Y}_{m'}\} \cup \{\mathbf{Y}_{1+m'}, \ldots, \mathbf{Y}_m\}$$

1. **二部排序。** 基于训练数据 $\mathscr{D}_{n'}$运行二部排序算法 $\mathcal{A}$，$_{m'} = \{\mathbf{X}_1$、.....、$\mathbf{X}_{n'}\} \cup \{\mathbf{Y}_1$、.....、$\mathbf{Y}_{m'}\}$，生成评分函数

$$\widehat{s} = \mathcal{A}(\mathscr{D}_{n',m'}) \ . \tag{3.1}$$

2. **单变量秩检验。** 构造单变量样本

$$\{\widehat{s}(\mathbf{X}_{1+n'}), \ldots, \widehat{s}(\mathbf{X}_n)\} \text{ and } \{\widehat{s}(\mathbf{Y}_{1+m'}), \ldots, \widehat{s}(\mathbf{Y}_m)\} \ ,$$

最终通过计算二元量来确定检验结果 y

$$\Phi_\alpha \left(\text{Rank}(\widehat{s}(\mathbf{X}_{1+n'})), \ldots, \text{Rank}(\widehat{s}(\mathbf{X}_n))\right) \ , \tag{3.2}$$

其中 Rank($t$) $= \sum_{i=1+n'}^{n} \mathbb{I}\{\widehat{s}(\mathbf{X}_i) \leq t\} + \sum_{j=1+m'}^{m} \mathbb{I}\{\widehat{s}(\mathbf{Y}_j) \leq t\}$。

图 3：基于排序的双样本秩检验程序。

从实际角度来看，如何恰当地选择用于二部排序的子样本和用于秩检验的子样本，既至关重要又十分困难，因为很难预先知道排序任务的复杂度。

现在，我们提出几种实现图中概述的方法3的方式。接下来，我们将在特定情形下对这些方式进行理论研究，并在第4节中对其性能进行深入的实证考察。

**基于排序的双样本线性秩检验。** 最简单的实现方式是考虑基于一元双样本线性秩统计量（2.4）的检验，该统计量由给定的得分生成函数 $\phi$刻画，见定义 1。如第 2.2 节所述，在单变量情形下，这种统计量在齐性假设 $\mathcal{H}_0$ 下是枢轴的。其零假设概率分布很容易制成表格，即使（$n$, $m$）取值非常大也不例外，因为如今可用的计算能力已十分强大。对于所有 $n$, $m \geq 1$ 以及任意 $\alpha \in$ （0,1），由此可确定分位数:

$$q_{n,\,m}^{\phi}(\alpha) = \inf_{t \geq 0} \left\{ \mathbb{P}_{\mathcal{H}_0} \left\{ \frac{1}{n} \widehat{W}_{n,m}^{\phi} - \int_0^1 \phi(u) \mathrm{d}u \leq t \right\} \geq 1 - \alpha \right\} \ , \tag{3.3}$$

以及临界域：

$$\left\{\frac{1}{n}\widehat{W}_{n,m}^{\phi} > \int_0^1 \phi(u)\mathrm{d}u + q_{n,\,m}^{\phi}(\alpha)\right\} \tag{3.4}$$

在单变量情形下，发生概率小于 $\alpha$ under $\mathcal{H}_0$，并将检验的显著性水平定义为 $\alpha$：

$$\Phi_{\alpha}^{\phi}(\mathscr{D}_{n,m}) = \mathbb{I}\left\{\frac{1}{n}\widehat{W}_{n,m}^{\phi} > \int_0^1 \phi(u)\mathrm{d}u + q_{n,\,m}^{\phi}(\alpha)\right\}, \tag{3.5}$$

基于单变量样本 $\mathscr{D}_{n,\,m} = \{X_1, \ldots, X_n\} \cup \{Y_1, \ldots, Y_m\}$。如第 2.3 节所述，对于双样本检验问题的重述形式（2.16）或（2.17），在多变量情形下只有单侧检验 $\Phi_{\alpha}^{\phi}$ 才有意义。这与单变量情形不同，后者无需进行二部排序。Cl´emen.con 等人 [2021] 研究了一种自然的二部排序方法，即基于（多变量）训练数据，$_{m'}$ 在类 $\mathcal{S}_0$上最大化性能准则（2.14），该准则的统计版本基于训练数据；即求解以下优化问题：

$$\max_{s\in\mathcal{S}_0} \widehat{W}_{n',m'}^{\phi}(s), \tag{3.6}$$

其中，对于任意评分函数 $s(z)$，我们定义：

$$\widehat{W}_{n',m'}^{\phi}(s) = \sum_{i=1}^{n'} \phi\left(\frac{\mathrm{Rank}(s(\mathbf{X}_i))}{N'+1}\right), \tag{3.7}$$

其中 $N' = n'+m'$，量 $\mathrm{Rank}(s(\mathbf{X}_i))/(N'+1)$ 是 $i = 1, \ldots, n'$ 时 $F_s(s(\mathbf{X}_i))$ 的自然经验对应量。Cl´emen.con 等人 [2021] 研究了问题 (3.6) 解的泛化能力，以及用于近似求解 (3.6) 的（基于梯度上升的）优化策略。因此，考虑在步骤 1 中获得的 (3.6) 的一个解 $\widehat{s}$，基于评分后的两个样本在步骤 2 中构建的检验：

$$\mathscr{D}_{n'',m''}(\widehat{s}) = \{\widehat{s}(\mathbf{X}_{1+n'}), \ldots, \widehat{s}(\mathbf{X}_n)\} \cup \{\widehat{s}(\mathbf{Y}_{1+m'}), \ldots, \widehat{s}(\mathbf{Y}_m)\}\,, \tag{3.8}$$

其中 $n'' = n - n'$ 和 $m'' = m - m'$，写作

$$\Phi_{\alpha}^{\phi}(\mathscr{D}_{n'',m''}\,(\widehat{s})) = \mathbb{I}\left\{\frac{1}{n''}\widehat{W}_{n'',m''}^{\phi}(\widehat{s}) > \int_0^1 \phi(u)\mathrm{d}u + q_{n'',m''}^{\phi}(\alpha)\right\}. \tag{3.9}$$

在特定假设下，尤其是与类 $\mathcal{S}_0$ 和得分生成函数 $\phi$相关的假设下，文献第 3.9 节研究了检验（3.2）的非渐近性质。

注 *4.* （Combining multiple ranking-based two-sample linear rank tests）正如 *Cl´emen.con* 等人所强调的，*[2021]*, 所选的得分生成函数 $\phi$决定了该量 $W_{\phi}^{*}$ 以某种方式概括 *ROC** 。由于我们很难事先知道哪种 $\phi$ 最能捕捉 *ROC** 在实践中偏离对角线的方式（通过标量 $W_{\phi}^{*} - \int_0^1 \phi(u)du \geq 0$），一种自然的策略是同时实施若干基于排序的双样本线性秩检验，并采用多重假设检验和集成学习中的常用原则。例如，可以考虑 $K \geq 1$ 得分生成函数 $\phi_1$、……、 $\phi_K$，并配合显著性水平 $\alpha_1$、……、 $\alpha_K$，构成检验集成$\{\Phi_{\alpha_k}^{\phi_k}: k = 1$、……、 $K\}$，以及组合

$$\sup_{1\leq k\leq K} \Phi_{\alpha_k}^{\phi_k}\left(\mathscr{D}_{n'',m''}(\widehat{s})\right). \tag{3.10}$$

尽管可以保证 (3.10) 的第一类错误率小于 $\alpha$ （只需选择 $\alpha_k$<style id='15}'，使得 $\sum_{k=1}^{K}\alpha_k = \alpha$），但由于所结合的检验以及不同 $\phi_k$之间存在依赖关系，研究其性质会面临重大困难。因此，这类方法的研究有待进一步开展。

**基于排序的检验在ROC 空间中。** 如第 2.2 节所强调的，在 $\mathcal{H}_0$ 下，经验 ROC 曲线本身就是一个双样本秩统计量（因此也是枢轴的）。因此，步骤 *2* 中涉及的检验可以基于置信区域 $\mathcal{R}_{n'',m''}(\alpha)\subset \mathcal{C}_{n'',m''}$ 其水平为$1-\alpha$，针对在$\mathcal{H}_0$下、基于样本量分别为 $n''$ 和 $m''$ 的单变量样本得到的经验 ROC 曲线。使用步骤 *1*中得到的评分函数 $\widehat{s}$，绘制与单变量分布相关的经验 ROC 曲线:

$$\widehat{H}_{\widehat{s},n''} = \frac{1}{n''}\sum_{i=1+n'}^{n}\delta_{\widehat{s}(\mathbf{X}_i)} \text{ and } \widehat{G}_{\widehat{s},m''} = \frac{1}{m''}\sum_{j=1+m'}^{n}\delta_{\widehat{s}(\mathbf{Y}_j)} \tag{3.11}$$

于是临界域写作

$$\left\{\mathrm{ROC}_{\widehat{\mathrm{H}}_{\widehat{\mathrm{s}},\mathrm{m}''},\,\widehat{\mathrm{G}}_{\widehat{\mathrm{s}},\mathrm{n}''}} \notin \mathcal{R}_{\mathrm{n}'',\mathrm{m}''}(\alpha)\right\}. \tag{3.12}$$

如第2.2节所指出的，给定 ROC 空间中的一个（伪）度量（例如 上确界范数），置信区域 $\mathcal{R}_{n'',m''}(\alpha)$自然可以对应于 $\mathcal{C}_{n'',m''}$ 中与对角线的距离小于特定阈值 $t_\alpha$ 的分段线性曲线集合（也可以要求曲线位于对角线上方，与 ROC* 相同）。对于 $L_1$-距离，当选择 $\phi(u) = u$ 时，这种实现与前一种实现一致。

## 3.2 理论保证——非渐近误差界

现在从非渐近视角分析第 3.1 节所述的基于排序的双样本线性秩检验的性质。令 $\alpha \in (0, 1)$。根据定理 3.9，检验$\alpha$ 按构造满足显著性水平3 ，证明见下文。

**定理 3。** (Type-I error bound) 设 $\phi(u)$ 为得分生成函数，且 $n,\ m \geq 2$。固定 $\alpha \in (0,\ 1)$。在原假设 $\mathcal{H}_0$下，检验 (3.9) 的第一类错误小于 $\alpha$

$$\mathbb{P}_{\mathcal{H}_0}\left\{\Phi_\alpha^{\phi}\left(\mathscr{D}_{n'',m''}(\widehat{s})\right) = 1\right\} \leq \alpha\ , \tag{3.13}$$

对所有 $1 \leq n'' < n$ 和 $1 \leq m'' < m$均成立。

证明直接由附录部分 B 中详述的条件化论证得出。（2.4）的分布很容易制成表格，而对于任意 $n,\ m \geq 1$，参见第 3.1 节，均可通过数值计算得到分位数（3.3）。考虑以下与得分生成函数 $\phi$ 的光滑性相关的假设。

假设 **1.** 得分生成函数 $\phi:\ [0,1] \mapsto \mathbb{R}$ 单调不减且二阶连续可微。

在假设 1下，以下结果给出了（3.3）的一个上界，当 $n,\ m$ 同时趋于无穷大时，该上界趋于 0（因此 $n/(n+m) \to p$ 以 $1/(n+m)$ 的速率趋于 $p \in (0,\ 1)$）。

**命题 4**。 假设 *1* 成立。设 $p \in (0,\ 1)$ 以及 $N \geq 1/p$。令 $n = \lfloor pN \rfloor$ 且 $m = \lceil (1-p)N \rceil = N - n$。那么，对于任意 $\alpha \in (0,\ 1)$，有：

$$q^{\phi}_{n,m}(\alpha) \leq \sqrt{\frac{\log(18/\alpha)}{CN}}\ , \tag{3.14}$$

其中 $C = 8^{-1}$ 最小值 $\left(p/\|\phi\|^2_{\infty}, (p\|\phi'\|^2_{\infty})^{-1}, ((1-p)\|\phi'\|^2_{\infty})^{-1}\right)$。

证明见附录 B.6。该结论可直接由附录 B 中的定理 10所建立的双样本线性秩统计量尾概率界得出；鉴于（2.4）满足中心极限定理，该界具有预期的阶，参见例如 van der Vaart 的定理 13.25 [1998]。

此外，还可以利用 Cl´emen¸con 等人 [2021]证明的非渐近结果来控制第二类错误。他们针对复杂度受控的评分函数类 $\mathcal{S}_0$，给出了经验准则（3.7）与理论准则（2.14）之间最大偏差的概率不等式，以及问题（3.6）解的 $W_{\phi}$-排序性能缺失的上置信界。这些结果依赖于对统计量（3.7）应用的线性化技术，以及双样本$U$过程的集中界。如下文所示，并且由 $\widehat{s} \in \mathcal{S}_0$可知，这些结果能够为以下量建立尾界

$$\frac{1}{n''}\widehat{W}^{\phi}_{n'',m''}(\widehat{s}) - \int_0^1 \phi(u)\mathrm{d}u = \left\{\frac{1}{n''}\widehat{W}^{\phi}_{n'',m''}(\widehat{s}) - W_{\phi}(\widehat{s})\right\} + \left\{W_{\phi}(\widehat{s}) - W^*_{\phi}\right\} + \left\{W^*_{\phi} - \int_0^1 \phi(u)\mathrm{d}u\right\}. \tag{3.15}$$

公式（3.15）右侧第三项（非负）衡量偏离齐性假设的程度 $\mathcal{H}_0$，而第二项对应二部排序步骤本身固有的误差（图 步骤 *1* 中的3）。应用这些结果还需要以下附加假设。

假设 **2.** 令 $M > 0$。对于所有 $s \in \mathcal{S}_0$，随机变量 $s(\mathbf{X})$ 和 $s(\mathbf{Y})$ 均为连续变量，其密度函数二阶可微，且索*оболев* $\mathcal{W}^{2,\infty}$- 范数[3] 有界于 $M < +\infty$。

假设 **3.** 评分函数类 $\mathcal{S}_0$ 是一个具有有限 VC维数$\mathcal{V} < +\infty$的 VC类。

关于 Vap- nik–Chervonenkis (VC) 函数类的定义，请参阅 van der Vaart 和 Wellner [1996] 的第 2.6.2 节。Cl´emen¸con 等人 [2021] 精确证明，在上述假设下，且当 $n'/(n'+m') \sim n''/(n''+m'') \sim p$ 时，经验最大化者 $\widehat{s}$ 的解的 $W_{\phi}$-排序性能亏损，即 (3.15) 右侧的第二项，为 $O_{\mathbb{P}}(1/\sqrt{n' \wedge m'})$ 阶（忽略~~可能的~~偏差模型；该模型源于 $\mathcal{S}_0 \cap \mathcal{S}^*$ 可能为空，见其中的推论 7 ），

[3]回顾一下，索оболев空间 $\mathcal{W}^{2,\infty}$ 是由所有满足以下条件的博雷尔函数 $h\colon \mathbb{R} \to \mathbb{R}$ 构成的空间： $h$ 及其一阶和二阶弱导数 $h'$ 和 $h''$ 几乎处处有界。记 $||.||_{\infty}$为由博雷尔且本质有界函数构成的勒贝格空间 $L_{\infty}$ 的范数，则 $\mathcal{W}^{2,\infty}$ 在赋予范数$||h||_{2,\infty} = \max\{||h||_{\infty}, ||h'||_{\infty}, ||h''||_{\infty}\}$后是一个巴拿赫空间。

而第一项为 $O_{\mathbb{P}}(1/\sqrt{n'' \wedge m''})$ 阶（参见其中定理 5 的论证）。

考虑量 $W_\phi^* - \int_0^1 \phi(u)\mathrm{d}u$ 来描述偏离零假设的程度$\mathcal{H}_0$（见命题 2），以及二部排序步骤中固有的偏差模型 $W_\phi^* - \sup_{s\in\mathcal{S}_0} W_{\phi(s)}$（将其表述为经验 $W_\phi$-排序性能最大化时），我们引入定义在 $\mathcal{Z}$ 上的两类（非参数）概率分布对。

**定义 5.** 设 $\varepsilon > 0$ 和 $\phi(u)$ 为得分生成函数。我们用 $\mathcal{H}_1(\varepsilon)$ 表示满足以下条件的所有概率分布对 $(H, G)$ 所对应的备择假设集合，这些分布定义在 $\mathcal{Z}$ 上：

$$W_\phi^* - \int_0^1 \phi(u)\mathrm{d}u \geq \varepsilon \ ,$$

其中，回顾一下，对于任意 $s^* \in \mathcal{S}^*$，有 $W_\phi^* = W_\phi(s^*) = W_\phi^*(H, G)$。

**定义6.** 设 $\delta > 0$, $\phi(u)$ 为得分生成函数， $\mathcal{S}_0$ 为评分函数类。我们用 $\mathcal{B}(\delta)$ 表示满足以下条件的所有定义在 $\mathcal{Z}$上的概率分布对 $(H, G)$:

$$W_\phi^* - \sup_{s\in\mathcal{S}_0} W_\phi(s) \leq \delta \ .$$

下面的定理给出了基于排序的秩检验（3.9）第二类错误的速率界，其取决于该程序中涉及的两个合并样本的大小 $\alpha$和 $N'$，分别为 $N'' = N - N'$，即，它们分别 用于二部排序，以及基于所学习的评分函数执行秩检验。

定理 7. (Type-II error bound) 设 $\phi(u)$ 为一个得分生成函数，且 $\varepsilon > \delta > 0$。固定 $\alpha \in (0, 1)$。假设假设1-3均成立。设 $p \in (0, 1)$，使得$N' \wedge N'' \geq 1/p$。令 $n' = \lfloor pN' \rfloor$ 和 $m' = \lceil (1-p)N' \rceil = N' - n'$，以及 $n'' = \lfloor pN'' \rfloor$和 $m'' = \lceil (1-p)N'' \rceil = N'' - n''$。j 那么，存在常数 $C_1$和 $C_2 \geq 24$，它们取决于 $(\phi, \mathcal{V})$，使得检验 (3.9) 的第二类错误一致地满足以下界:

$$\sup_{(H,G)\in\mathcal{H}_1(\varepsilon)\cap\mathcal{B}(\delta)} \mathbb{P}_{H,G}\left\{\Phi_\alpha^\phi(\mathscr{D}_{n'',m''}(\widehat{s})) = 0\right\} \leq 18\exp\left\{-\frac{CN''}{16}(\varepsilon-\delta))^2\right\}$$
$$+ C_2\exp\left\{-\frac{N'}{8C_2}p(p\wedge(1-p))(\varepsilon-\delta)\log\left(1+\frac{\varepsilon-\delta}{32C_1(p\wedge(1-p))}\right)\right\} \ , \quad (3.16)$$

一旦 $N'' \geq 4\log(18/\alpha)/(C(\varepsilon-\delta)^2)$ 且 $N' \geq 16C_1^2/(p(\varepsilon-\delta)^2)$，常数 $C$为命题 4中涉及的常数，$C_j$则为定理13中涉及的常数。

详细证明见附录 B.7。根据定义 5 和 6，通过要求 $\varepsilon > \delta$，我们在此考虑满足以下条件的样本对 $(H, G)$：类 $\mathcal{S}_0$ 在二部排序问题中引起的模型偏差足够小，即小于偏离原假设的最小程度 $\varepsilon$。上述界表明，当 $\varepsilon - \delta > 0$ 固定时，随着 $N'$ 和 $N''$ 同时趋于无穷，第二类错误的上界一致地以指数速度趋于零：界（3.16）右侧的第二项是学习阶段固有的，而第一项对应于单变量秩检验的第二类错误上界。因此，当 $N'$ 足够大时，基于排序的秩检验的第二类错误与基于单变量样本 $\{\Psi(\mathbf{X}_{1+n'}) \ldots, \Psi(\mathbf{X}_n)\}$ 和 $\{\Psi(\mathbf{Y}_{1+m'}) \ldots, \Psi(\mathbf{Y}_m)\}\}$ 的秩检验的第二类错误具有相同的量级。当偏离原假设的程度 $\varepsilon$ 趋于 0 或 $+\infty$ 时，这两个界呈现出不同的变化：

当合并样本量 $N'$和 $N''$固定时，随着 $\varepsilon$趋于无穷，单变量秩检验的第二类错误相关上界相较于学习界可忽略不计；而当 $\varepsilon$ 趋于 0 时，该上界的恶化速度则更快。另请注意，在上述结果中，二部排序步骤（步骤 *1*）所涉及的合并样本中正例的比例 $p$ 与秩检验（步骤 *2*）所用合并样本中的比例相同；本分析的两个步骤均使用相同的秩统计量 $\widehat{W}^{\phi}$。然而，如果可以根据所用检验统计量的离散枢轴条件分布所对应的检验水平 $\alpha$来指导 $N''$ 的选择，那么定理 7 所确立的界仍无法在实践中调整每个 $N$ 步骤 1 和 步骤 2所需使用的训练观测值数量。正如第 4节将要讨论的，只需将样本分成两部分，并将合并数据集中的较大部分留给学习阶段（例如 $N' \sim 4N/5$），或许就能获得令人满意的结果。

注释 *5*。（Relation to minimax testing）与非参数极小极大检验文献相比，所提出的、$\mathcal{H}_1$ 从 $W_\phi$-准则意义下定义备择假设 $\mathcal{H}_1$ 的方法是新颖的。后者通过使用关于某一特定度量的距离来衡量底层分布之间的差异，从而对其进行表述；例如，*Lam-Weil* 等人 *[2022]* 研究了由 $L_1$-范数定义的离散分布的局部极小极大分离率，*Carpentier* 等人 *[2018]* 则研究了稀疏线性回归中的 $L_2$-范数。事实上，在 *ROC* 空间中考虑 $W_\phi$-准则时，定义 *5* 提供了类似的解释；当根据公式 *(2.15)* 以及 *Cl´emen¸con* 等人 *[2021]*的注释 *5* 选择 $\phi(u) = u$ 时，这一点尤为明显。因此，可以对当前统计量的第二类错误的极小极大性质进行表述，详见 *Limnios [2022],* 其中的第 *6.4.2* 章讨论。

最后我们指出，对于 ROC 空间中基于上确界范数的临界域（3.12）所对应的秩检验，也可以完全以相同的方式建立与定理7 类似的结果，只是必须使用 Cl´emen¸conandVayatis [2010]或Cl´emen¸conand Vayatis [2009b] 中建立的（上确界范数下的）界，而不是 Cl´emen¸con 等人 [2021]中建立的界。

## 4 说明性数值实验

本节通过展示基于合成数据集开展的各种数值实验结果，阐释前文所述并针对 $W_\phi$排序性能优化情形进行分析的方法。除与多元框架中的某些常用双样本检验相比，实证展示其成功检测齐性假设的（较小）偏离的能力外，这些实验还旨在说明所涉及的各种要素的影响，尤其是得分生成函数 $\phi$ 的选择以及所使用的二部排序算法。我们首先详细介绍实验中考察的双样本问题的基于排序的方法的各种变体。接着说明所考虑的双样本问题，最后总结并讨论实证结果。其他实验结果将推迟至附录部分C。用于开展这些实验的所有 Python 代码均可在线获取，网址为`https://github.com/MyrtoLimnios/twosampleranktest` 以便复现。

### 4.1 实际实现 - 超参数

第3.1小节介绍的两阶段检验程序可以通过多种方式实现，具体取决于所使用的二部排序算法 $\mathcal{A}$ 以及所选择的得分生成函数。

二部排序算法（步骤 1）。如第 2.3 小节所述（另见注释 3），机器学习文献中记载的大多数技术都将二部排序表述为成对分类问题，参见 Cl´emen¸con 等人 [2008] 以及 Menon 和 Williamson [2016]。因此，可以将常用的分类算法（例如支持向量机、神经网络、提升）应用于样本对 $(Z, Z')$：当 $Z$ 和 $Z'$来自同一个多元分布（$H$ 或 $G$）时，赋予其标签 $-1$；否则赋予标签 $+1$。具体而言，在本研究的实验中，我们实现了带$L_2$ 损失的线性版本 RankSVM（rSVM2，参见 Joachims [2002]）、RankNN（rNN，参见 Burges 等人 [2005]）和 RankBoost（rBoost，参见 Freund 等人 [2003]）。除成对方法外，研究者还提出了基于有向二叉树的递归划分方法，用于优化经自适应离散化的 ROC 曲线版本（在上确界范数下具有显著的泛化保证），参见 Cl´emen¸con 和 Vayatis [2009b]以及 Cl´emen¸con 等人 [2011]（另见 Cl´emen¸con 和 Vayatis [2010] 以及 Cl´emen¸con 和 Vayatis [2009c]）。本文实现了稳定性更高的集成学习版本，称为 Ranking Forest（rForest，参见 Cl´emen¸con 等人 [2013a]）；其实际性能尤其在 Cl´emen¸con 等人 [2013b]中得到了评估。此外，正如 Cl´emen¸con 等人 [2021] 所述（参见其中第 4 节），通过核平滑程序可以获得经验 $W_\phi$-排序性能准则的正则化版本，并可通过（随机）梯度上升算法对其进行最大化；该算法与第 3.2 小节中从理论上分析的算法相近。

**单变量双样本秩检验（步骤 *2*）**。设 $\hat{s}$ 为第一步的结果。为执行步骤 *2*，我们考虑检验统计量 $\widehat{W}_{n,\,m}^{\phi}(\hat{s})$，其得分生成函数如下：$\phi_{MWW}(u) = u$，它对应于经典的 Mann-Whitney-Wilcoxon 统计量（MWW，Wilcoxon [1945]），以及 $\phi_{RTB}(u) = u\mathbb{I}\{u \geq u_0\}$，针对 $u_0 \in (0,\ 1)$（RTB，Cl´emen¸con 和 Vayatis [2007]），以聚焦最高秩次以及 ROC 曲线在原点附近的表现。
对于给定的水平 $\alpha$，通过将分位数$q_{n'',\ m''}^{\phi}(\alpha)$ 作为临界阈值来控制第一类错误，而对于下文所述的工具分布对，则通过蒙特卡罗方法估计第二类错误。如第2.2小节所述，秩统计量是枢轴的，分位数 $q_{n'',\ m''}^{\phi}(\alpha)$ 已针对经典的 $\phi$ 选项制成表格（例如，可通过 Python 中开放访问的 SciPy 库轻松获取），也可以借助现代计算框架轻松计算，即使 $n'', m''$ 的值非常大也不例外。
我们首先考察固定设计下，不相似性/差异 参数 $\varepsilon > 0$变化时的两种概率，然后固定不相似性参数，让维数$d$ 增加。

**双重划分技巧**。由于最具挑战性的步骤无疑是第一步，即学习如何对观测值进行排序，使其顺序尽可能接近似然比 $\Psi(z)$ 所诱导的顺序，因此会将大得多比例的数据分配给二部排序。具体而言，我们取 $N'/N = 4/5$ (使得 $N'' = N/5$)。

**基准测试**。将所提出的基于排序的方法与四种最先进的多变量非参数双样本检验进行比较：即采用高斯

核（参见 Gretton 等人 [2007, 2012]）的无偏（平方）最大均值差异（最大均值差异）检验、基于图的 Wald-Wolfowitz 游程检验（FR）——由Friedman 和 Rafsky [1979]，提出并推广至多变量情形——基于度量的能量检验（能量）（参见 Sz´ekely 和 Rizzo [2013]），以及Tukey在 Liu 和 Singh [1993] 中记载的基于统计深度的程序，用 Tukey 统计深度（Tukey [1975]）推广一元双样本秩检验；更多细节请参见第2.2小节。请注意，最大均值差异 和能量 在 $\mathcal{H}_0$下并非严格意义上的分布无关检验，因此采用置换方法进行校准。我们还指出，Tukey 的实现使用了与基于排序的方法相同的秩统计量查表结果。

**评估标准**。 对于每项实验，在所有水平 $\alpha \in (0, 1)$ 下，各检验程序的第一类错误频率（$\varepsilon = 0$）和第二类错误频率（$\varepsilon > 0$）均基于 $B \geq 1$ 次蒙特卡罗重复计算，并报告所得的 $p$-值分布。我们还研究了在固定 $\varepsilon$ 和样本量 $N$的情况下，维数 $d$ 增加对检验功效的影响。

**实验参数**。 在所有实验中，合并样本的大小为 $N$ =2000，且两组样本数量相等，即 $n = m = N/2$ （对应于 $p = 1/2$）。蒙特卡罗重复次数为 $B = 100$。RTB 得分生成函数的参数 $u_0$ 取自集合 $\{0.7, 0.8, 0.9\}$。对于基准测试，零分布通过 $B_{perm} = 1000$ 次置换进行估计，相关超参数则在范围 $\{1e-3, 1e-2, 1e-1, 1, 5, 10, 15, 20, 25, 30, 1e2, 1e3\}$ 内优化。图中还展示了置信水平为 95% 的逐点置信区间。

### 4.2 合成数据集

我们通过以下双样本问题，展示所提出的基于秩的检验族的性能。我们考虑各种位置高斯模型和尺度高斯模型，以便将其与（最优）似然比统计量的性能进行比较，参见示例1。我们还检验从混合模型和较不经典（重尾）统计分布中抽取的样本的齐性。有关所选分布参数以及真实 ROC 曲线图（包括 ROC）的更多信息，请参见附录部分C 。对于 *，记 $d \geq 1$为 $0_d \in \mathbb{R}^d$ 零向量，记 $\mathbf{1}_d \in \mathbb{R}^d$ 为单位向量（即 所有坐标均为 1），记 $\mathbb{I}_d \in \mathbb{R}^{d\times d}$ 为单位矩阵，记 $S_d^+(\mathbb{R})$ 为实对称半正定 $d \times d$ 矩阵锥。

**位置高斯模型**。 两个样本 $\mathbf{X} \sim \mathcal{N}_d(\mu_X, \Sigma)$ 和 $\mathbf{Y} \sim \mathcal{N}_d(\mu_Y, \Sigma)$ 独立抽取，其中 $\Sigma \in S_d^+(\mathbb{R})$、 $\varepsilon \in \{0.0, 0.02, 0.05, 0.08, 0.1\}$ 如下。

(L1) 我们设定 $\mu_Y = 0_d$ 和 $\mu_X = (\varepsilon/\sqrt{d}) \times \mathbf{1}_d$。考虑两种协方差矩阵模型：第一个坐标与所有其他坐标负相关，其余 $d-1$ 个坐标则：相互独立 (L1−)；正相关 (L1+)。

对于此类模型，类 $\mathcal{S}^*$ 可以明确给出，详见示例1。

**尺度高斯模型**。 两个样本 $\mathbf{X} \sim \mathcal{N}_d(0_d, \Sigma_X)$ 和 $\mathbf{Y} \sim \mathcal{N}_d(0_d, \Sigma_Y)$ 相互独立地抽取，且 $\Sigma_X$, $\Sigma_Y$ 取自 $S_d^+(\mathbb{R})$，定义如下。

(S1) 递减相关性。 令 $\Sigma_{X,i,j} = \alpha^{|i-j|}$ 和$\Sigma_{Y,i,j} = \beta^{|i-j|}$ ，其中 $i, j \leq d$，且$d = 20$、$\beta = 0.2$和 $\alpha = \beta + \varepsilon$、 $\varepsilon \in \{0, 0.1, 0.2, 0.3\}$。

(S2) 等相关样本。 令 $\Sigma_X = (1-\alpha)\mathbb{I}_d + \alpha \mathbf{1}_d \mathbf{1}_d^T$和 $\Sigma_Y = (1-\beta)\mathbb{I}_d + \beta \mathbf{1}_d \mathbf{1}_d^T$，其中 $d \in \{20,\ 30,\ 60,\ 100\}$、 $\beta = 0.3$和 $\alpha = \beta + \varepsilon$、 $\varepsilon \in \{0, 0.05, 0.1, 0.15\}$。

对于高斯位置模型，这种情况下很容易明确写出最优评分函数，即二次函数 $s_{\theta^*}(\cdot)$ $= \langle \cdot, \theta^* \cdot \rangle$，其中 $\theta^* = \Sigma_X^{-1} - \Sigma_Y^{-1}$ 例如。
**非高斯模型**。 我们还考虑具有更重尾部的生成模型，以构建其他双样本问题。

(T1) 柯西分布。我们独立生成 X $= (X_1, X_2, X_3)$ 和 Y $= (Y_1, Y_2, Y_3)$，使得 $X_1, X_2 \overset{i.i.d}{\sim}$ 服从 Cauchy($\varepsilon$, 1)，且 $X_3$, $Y_1$, $Y_2$, $Y_3 \overset{i.i.d}{\sim}$ 服从 Cauchy(0, 1)，其中$\varepsilon \in \{0, 0.05, 0.15, 0.20\}$, $d = 3$。(T2) 对数正态位置。我们生成 X 和 Y，使得 log(X) 和 log(Y) 取自具有 $\varepsilon \in \{0, 0.1, 0.2, 0.3\}$, $d = 4$ 的 (L1) 模型。(T3) 对数正态尺度。我们生成 X 和 Y，使得 l og X 和 log Y 取自具有 $\varepsilon \in \{0, 0.1, 0.2, 0.3\}$, $d = 20$ 的 (S1) 模型。

模型 (L1) 用于 Cl´emen¸con 等人 [2021]在二部排序情境中，模型 (S1)、(S2) 和 (T1) 见 Deband Sen [2021], 而模型 (T2) 和 (T3) 则广泛用于单变量情形。

### 4.3 结果与讨论

现在，我们讨论通过所提出的方法针对此前描述的双样本问题所获得的数值结果。我们根据步骤 1 所选的二部排序算法以及函数 $\phi$，将该方法的表现与一些最先进的非参数检验进行比较，比较依据是 I/II 类错误计数。我们自然关注各方法在偏离原假设较小时拒绝 $\mathcal{H}_0$ 的能力（不同实验中所考虑的偏离类型也有所不同），即 当 $\varepsilon \to 0$ 时，同时控制第一类错误。具体而言，我们讨论所有方法的：1）在水平范围 $\alpha \in$ (0,1) 内控制第一类错误的能力（见图）；2）其 $p$ 值随 $\varepsilon$ 变化的分布（见箱线图）；以及 3）随着特征空间维数 $d$ 增大，在水平范围 $\alpha \in$ (0,1) 内拒绝原假设的能力（见图）。对各种检验进行调参，使其在控制第一类错误的同时，最大化固定水平 $\alpha = 0.05$ 下的检验功效。我们在所考虑的最困难双样本问题上对它们进行比较，即 取 $\varepsilon$ 的最小值（见表格）。这里展示的是得分生成函数 $\phi(u) = u$ 对应的结果，该函数对应于 MWW 检验统计量。在附录部分还给出了采用其他得分生成函数的一些结果，以展示 $\phi$ 的选择可能如何影响双样本检验（步骤 2）的表现，尤其是 RTB 版本的表现。除图表外，我们还依次回顾了针对所考虑的各个双样本问题所得的结果。

首先，我们分析高斯位置模型的结果。如表 1 所示，对于这类双样本问题，当 $\varepsilon$ 值较小时，基于排序的检验显然比最先进方法检验更有能力拒绝原假设，同时对第一类错误的控制也更好，具体见图 4 (a) 和 5 (a)，尤其是 `rForest`。与最先进方法检验相比，基于排序的方法的 $p$ 值分布方差更低、检验功效更高，尤其是在 $\varepsilon = 0.02$ （L1-）值非常小时，见图 4 (b-d)，以及图 5 (b-d)

针对 (L1+)，汇总结果见表1。

对于高斯尺度模型，我们研究了这些方法在高维情形下的表现。结果清楚地表明，所有基于排序的检验对第一类错误的控制都优于最先进的方法，见图 6 (a)。在较低维数下，最先进方法检验的检验功效具有竞争力，见表 2，但 Tukey 除外，它的 $p$ 值方差较大，检验功效较低。具体而言，rForest 和 rNN 在较小的$\varepsilon$ （见列 (S2) $\varepsilon = 0.1$ 和 (S3) $\varepsilon = 0.05$）情况下的拒绝率最高，而最先进的方法的拒绝率则非常低。此外，随着特征空间 $d$ 维数的增加，无论检验水平 rNN 的经验检验功效始终高于其他任何方法，且这一结论对任意检验水平均成立 $\alpha \in (0,\ 1)$，见图 11 中的 (S2) 和 $d > 30$。

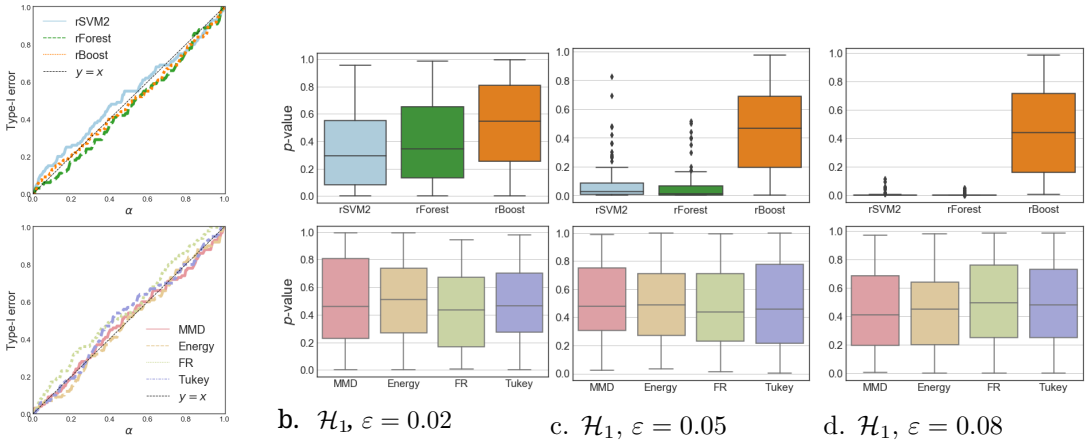
对于第三类模型，我们分析了图 $p$中8 （对应 (T1)）、图 9 （对应 (T2) 和 (L1+)）以及图 10 （对应 (T3) 和 (S1)）中的值分布。 rBoost 和 最大均值差异 在 (T1)（即多元柯西分布）下表现相似。 rForest 和 最大均值差异 在 (T2) 和 (T3) 下的结果均具有可比性。总体而言，在三种模型的备择假设下，rForest 在最小的 $\varepsilon$下表现出较高的经验拒绝率。最后，基于秩的算法对原假设下拒绝率的控制更好，见图 9(a) 和图 10 (a)。

最后，我们测试了增加 $d$ 对所有方法在备择假设下的拒绝率相对于水平范围 $\alpha \in (0,1)$ 对（S2）双样本问题的影响。图11 清楚地表明，对于最小的 $\varepsilon = 0.05$，基于排序的方法 rNN 的经验检验功效始终高于最先进的方法，在所考虑的最高维数 $d = 100$处差异尤为显著。当 $\varepsilon$ 增加时，rNN 和最大均值差异的表现相近。表3 汇总了 $\alpha = 0.05$ 的估计检验功效，并展示了基于排序的方法的检验功效随维数增加而不降低，而最先进的方法总体上具有恒定或下降的检验功效相对于维数，见能量 $\varepsilon = 0.1$。

总之，我们通过经验结果说明了基于排序的秩检验的竞争力，尤其是在位置和病态统计问题中，因为在较大的水平范围 $\alpha$内，这些方法总体上清楚地控制了经验第一类错误，同时在非常小的偏离 $\varepsilon \to 0$下，备择假设中的拒绝率相近或更高。在这些实验中，步骤 1中表现最好的基于排序的算法是 rForest，而表现较差的是 rBoost。相比之下，最先进方法检验总体上能够在大多数模型中控制第一类错误，但代价是算法复杂度较高，这是由于置换次数很多。此外，在非常小的 $\varepsilon$ 下，它们的经验检验功效低于基于排序的方法。基于深度的检验（Tukey）的表现尤其不佳。最后， $p$-值的分布在 $\varepsilon$ 减小时显示出较大的方差，并且对于最先进的方法而言方差尤其更大，而与底层概率模型无关。

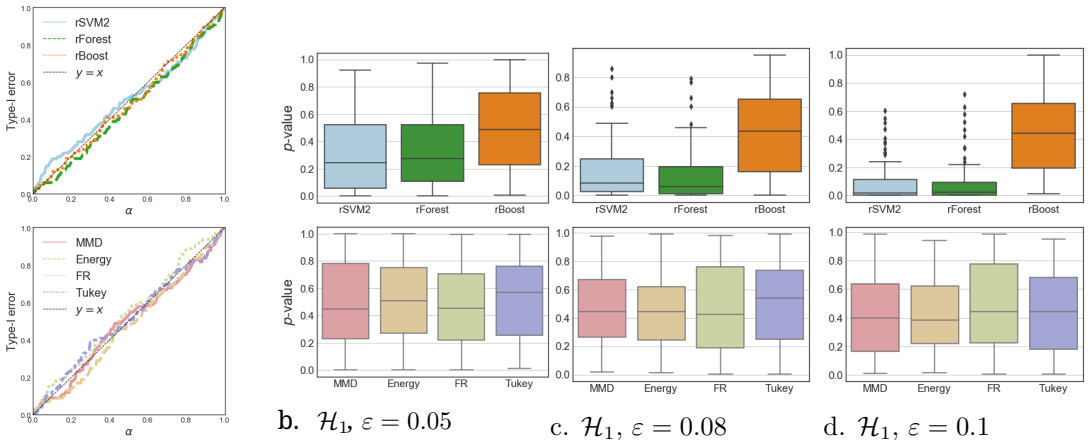
## 5 结论

本文提出了一种全新且灵活的双样本问题解决方法，该问题在统计学和机器学习应用中无处不在。
该方法基于以下观察：一旦与所观测两个样本的概率分布 $H$ 和 $G$ 相关的二部排序问题的最优 ROC 曲线不再与对角线重合，零假设/同质性假设 $\mathcal{H}_0$ 就不再成立




a. $\mathcal{H}_0$, $\varepsilon = 0.0$ b. $\mathcal{H}_1$, $\varepsilon = 0.02$ c. $\mathcal{H}_1$, $\varepsilon = 0.05$ d. $\mathcal{H}_1$, $\varepsilon = 0.08$

图4：图(a)中在显著性水平 $\alpha \in$ (0,1) 下，针对 $\mathcal{H}_0$ 的齐性假设的拒绝率。图(b-d)中，在 $\mathcal{H}_1$ 下，针对 (L1-) 且具有 $\phi(u) = u$, $n = m = 1000$, $d = 6$, $B = 100$ 的 $p$-值箱线图。排序算法为 rSVM2、rForest、rBoost，并与最大均值差异、能量、FR 和 Tukey 进行比较。

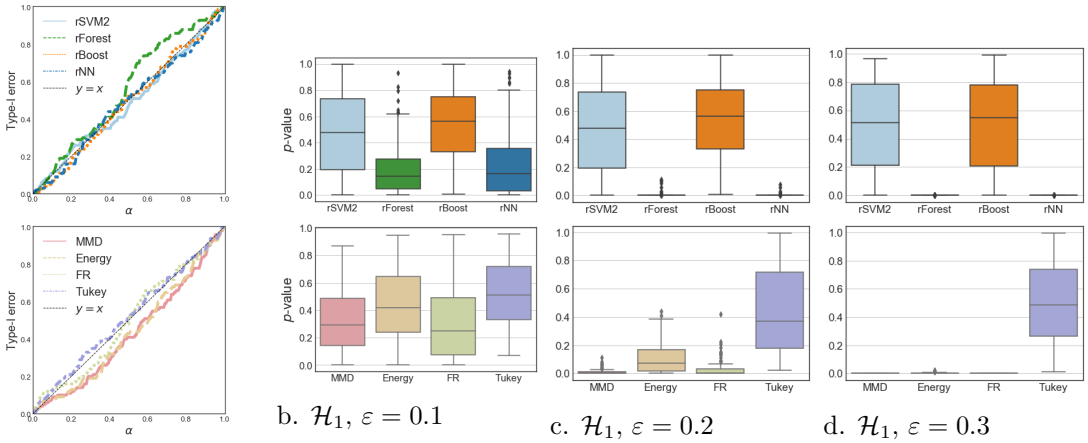



a. $\mathcal{H}_0$, $\varepsilon = 0.0$ b. $\mathcal{H}_1$, $\varepsilon = 0.05$ c. $\mathcal{H}_1$, $\varepsilon = 0.08$ d. $\mathcal{H}_1$, $\varepsilon = 0.1$

图5：图(a)中在显著性水平 $\alpha \in$ (0,1) 下，针对 $\mathcal{H}_0$ 的齐性假设的拒绝率。图(b-d)中，在 $\mathcal{H}_1$ 下，针对 (L1+) 且具有 $\phi(u) = u$, $n = m = 1000$, $d = 6$, $B = 100$ 的 $p$-值箱线图。排序算法为 rSVM2、rForest、rBoost，并与最大均值差异、能量、FR 和 Tukey 进行比较。

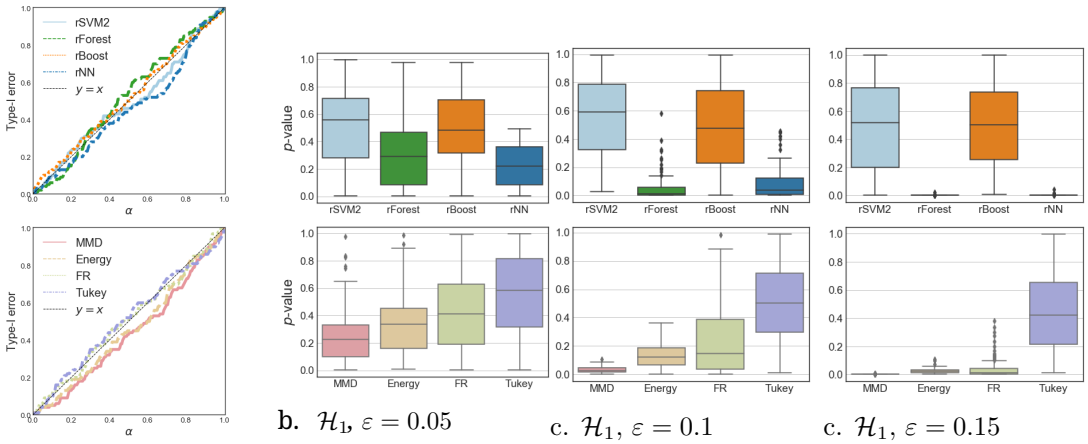
（此时，它必然会在 ROC 空间的某些点处位于对角线之上）。由于 ROC 曲线$^*_{H,\,G}$无法直接观测，所提出的方法包括一个统计学习步骤，即基于样本的第一部分来解决二部排序问题。接下来，使用（ROC$^*_{H,\,G}$的）经验版本来评估其偏离 $\mathcal{H}_0$ 的程度。在实践中，第二步归结为执行经典的双样本秩检验：利用第一步构建的排序规则，对未用于学习阶段的那部分样本进行排序，并基于所得秩进行检验。

我们进行了严谨的理论分析，证明该方法产生的两类误差均可通过非渐近方式加以控制。除此之外，




a. $\mathcal{H}_0$, $\varepsilon = 0.0$ b. $\mathcal{H}_1$, $\varepsilon = 0.1$ c. $\mathcal{H}_1$, $\varepsilon = 0.2$ d. $\mathcal{H}_1$, $\varepsilon = 0.3$

图 6：在图 (a) 中，针对 $\mathcal{H}_0$ 的齐性假设拒绝率与显著性水平 $\alpha \in$ （0，1）的关系。图 (b-d) 展示了在 $\mathcal{H}_1$ 下，(S1) 中 $p$ 值的箱线图，其中$\phi(u) = u$, $n = m = 1000$, $d = 20$, $B = 100$。排序算法包括 rSVM2、rForest、rBoost、rNN，并与最大均值差异、能量、FR 和 Tukey 进行比较。




a. $\mathcal{H}_0$, $\varepsilon = 0.0$ b. $\mathcal{H}_1$, $\varepsilon = 0.05$ c. $\mathcal{H}_1$, $\varepsilon = 0.1$ c. $\mathcal{H}_1$, $\varepsilon = 0.15$

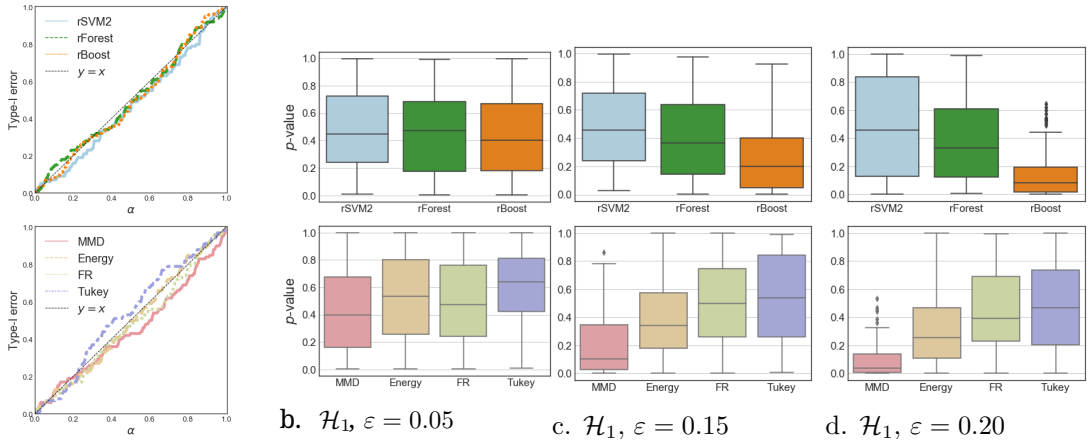
图 7：在图 (a) 中，针对 $\mathcal{H}_0$ 的齐性假设拒绝率与显著性水平 $\alpha \in$ （0，1）的关系。图 (b-d) 展示了在 $\mathcal{H}_1$ 下，(S2) 中 $p$ 值的箱线图，其中$\phi(u) = u$, $n = m = 1000$, $d = 30$, $B = 100$。排序算法包括 rSVM2、rForest、rBoost、rNN，并与最大均值差异、能量、FR 和 Tukey 进行比较。

我们还开展了实验研究，提供了有力的实证证据，表明与一些流行的备择假设相比，基于秩的方法具有诸多优点：总体而言，该方法在高维情形下表现更为稳健，在越来越接近原假设时也更能抵御性能下降，并能适应多种偏离齐性的方式。然而，要使此处提出的一般方法在实践中能够系统地表现良好，仍有若干问题有待研究。

尤其可以提出两个开放问题：如何分配现有数据，以用于统计学习和进行秩检验；以及如何开发一种方法，高效地结合多个基于排序的秩检验，从而有望提

| | 模型（L1-），$d=6$ | | | | 模型（L1+），$d=6$ | | | |
|---|---|---|---|---|---|---|---|---|
| **原假设拒绝率** | 在 $\mathcal{H}_0$ 条件下 | 在 $\mathcal{H}_1$ 条件下 | | | 在 $\mathcal{H}_0$ 条件下 | 在 $\mathcal{H}_1$ 下 | | |
| **方法** | $\varepsilon=0.0$ | $\varepsilon=0.02$ | $\varepsilon=0.05$ | $\varepsilon=0.08$ | $\varepsilon=0.0$ | $\varepsilon=0.05$ | $\varepsilon=0.08$ | $\varepsilon=0.1$ |
| rSVM2 | 0.11 ± 0.31 | **0.22 ± 0.49** | 0.60 ± 0.49 | 0.96 ± 0.20 | 0.10 ± 0.30 | **0.23 ± 0.42** | 0.37 ± 0.48 | **0.66 ± 0.48** |
| **rForest** | **0.05 ± 0.22** | 0.15 ± 0.36 | **0.71 ± 0.46** | **0.98 ± 0.14** | **0.05 ± 0.22** | 0.19 ± 0.39 | **0.48 ± 0.50** | 0.62 ± 0.49 |
| rBoost | 0.09 ± 0.29 | 0.07 ± 0.26 | 0.08± 0.27 | 0.11 ± 0.31 | 0.06 ± 0.24 | 0.10± 0.30 | 0.07 ± 0.26 | 0.06 ± 0.24 |
| MMD | **0.05 ± 0.22** | 0.04 ± 0.20 | 0.03 ± 0.17 | 0.04 ± 0.20 | 0.08 ± 0.27 | 0.06 ± 0.24 | 0.06 ± 0.24 | 0.06 ± 0.24 |
| 能量 | 0.06 ± 0.24 | 0.05 ± 0.22 | 0.03 ± 0.17 | 0.06 ± 0.24 | 0.06 ± 0.24 | 0.05 ± 0.22 | 0.04 ± 0.20 | 0.02 ± 0.14 |
| FR | 0.09 ± 0.29 | 0.09 ± 0.29 | 0.09 ± 0.29 | 0.06 ± 0.24 | 0.08 ± 0.27 | 0.10± 0.30 | 0.09 ± 0.29 | 0.07 ± 0.26 |
| Tukey | 0.06 ± 0.24 | 0.05 ± 0.22 | 0.08 ± 0.27 | 0.07 ± 0.26 | **0.05 ± 0.22** | 0.08 ± 0.27 | 0.07 ± 0.26 | 0.09 ± 0.29 |
| | **模型 (S1)，** $d=20$ | | | | **模型 (S2)，** $d=30$ | | | |
| **原假设拒绝率** | 在 $\mathcal{H}_0$ 下 | 在 $\mathcal{H}_1$ 下 | | | 在 $\mathcal{H}_0$ 下 | 在 $\mathcal{H}_1$ 下 | | |
| **方法** | $\varepsilon=0.0$ | $\varepsilon=0.1$ | $\varepsilon=0.2$ | $\varepsilon=0.3$ | $\varepsilon=0.0$ | $\varepsilon=0.05$ | $\varepsilon=0.1$ | $\varepsilon=0.15$ |
| rSVM2 | 0.06 ± 0.24 | 0.03± 0.17 | 0.06 ± 0.24 | 0.09 ± 0.29 | 0.04 ± 0.20 | 0.04 ± 0.20 | 0.04 ± 0.20 | 0.04 ± 0.20 |
| **rForest** | 0.05 ± 0.22 | **0.30 ± 0.46** | 0.95 ± 0.22 | **1.00** | **0.03 ± 0.17** | 0.19 ± 0.39 | 0.72 ± 0.45 | **1.00** |
| rBoost | 0.04 ± 0.20 | 0.02 ± 0.14 | 0.08 ± 0.27 | 0.10 ± 0.30 | 0.09 ± 0.29 | 0.04 ± 0.20 | 0.05 ± 0.22 | 0.03 ± 0.17 |
| rNN | **0.03 ± 0.17** | 0.27 ± 0.45 | **0.98 ± 0.14** | **1.00** | 0.04 ± 0.20 | **0.20 ± 0.40** | 0.54 ± 0.50 | **1.00** |
| MMD | 0.04 ± 0.20 | 0.08 ± 0.27 | 0.93 ± 0.26 | **1.00** | 0.04 ± 0.20 | 0.13 ± 0.34 | **0.79 ± 0.41** | **1.00** |
| 能量 | 0.06 ± 0.24 | 0.07 ± 0.26 | 0.39 ± 0.49 | **1.00** | 0.04 ± 0.20 | 0.11 ± 0.31 | 0.18± 0.39 | 0.93 ± 0.26 |
| FR | 0.10 ± 0.30 | 0.19 ± 0.39 | 0.82 ± 0.39 | **1.00** | 0.06 ± 0.24 | 0.06 ± 0.24 | 0.28± 0.45 | 0.78 ± 0.42 |
| Tukey | 0.08 ± 0.27 | 0.00 | 0.02 ± 0.14 | 0.04 ±0.20 | 0.04 ± 0.20 | 0.04 ± 0.20 | 0.04± 0.20 | 0.05 ± 0.22 |

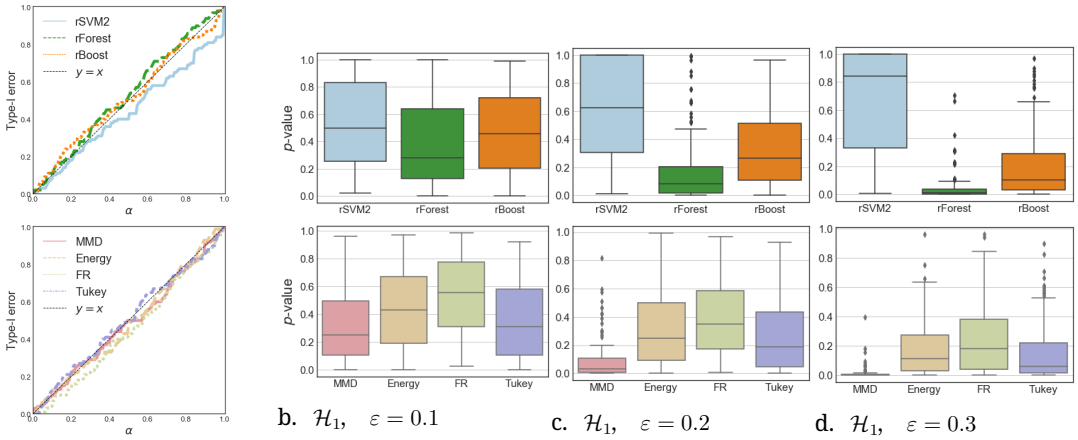
表1：在显著性水平 $\alpha=0.05$ 下，对于位置模型 (L1-) 和 (L1+) 以及尺度模型 (S1) 和 (S2)，当 $\mathcal{H}_0$ 和 $\mathcal{H}_1$ 时齐性假设的拒绝率； $n=m=1000$， $\pm$ 为其标准差，置信水平为 95%。对于排序方法，仅展示与 MWW 检验相关的结果。粗体标出的方法在所有算法中，在 $\mathcal{H}_0$ 下使拒绝率最小，同时在 $\mathcal{H}_1$下使拒绝率最大。




a. $\mathcal{H}_0$, $\varepsilon=0.0$　b. $\mathcal{H}_1$, $\varepsilon=0.05$　c. $\mathcal{H}_1$, $\varepsilon=0.15$　d. $\mathcal{H}_1$, $\varepsilon=0.20$

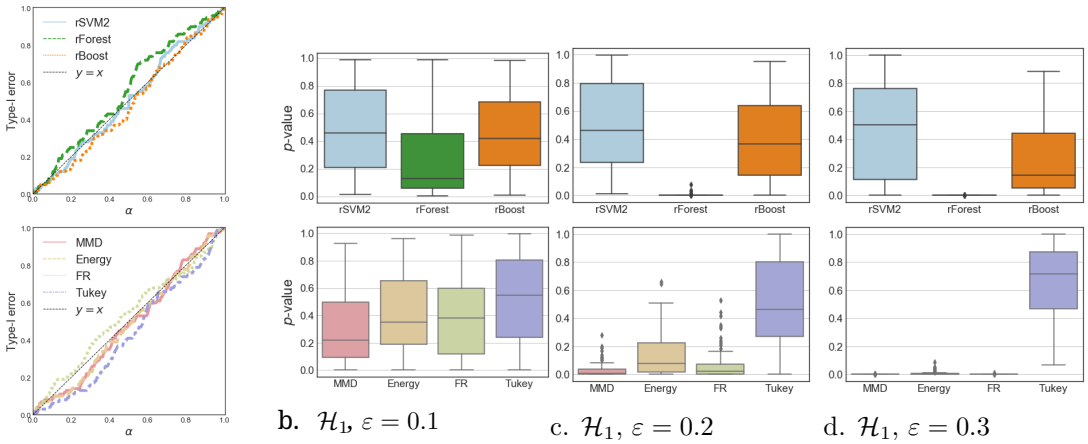
图8：图(a)中，在显著性水平 $\alpha\in$ (0,1) 下，针对 $\mathcal{H}_0$ 的齐性假设的拒绝率。图(b-d)中，在 $\mathcal{H}_1$ 下，(T1) 中带有$\phi(u)=u$, $n=m=1000$, $d=3$, $B=100$ 时 $p$-值的箱线图。排序算法 rSVM2、rForest、rBoost 与最大均值差异、能量、FR 和 Tukey 进行比较。

升性能（因为事先并不知道哪种得分生成函数 $\phi$ 最能捕捉所呈现的偏离齐性类型），这些问题都需要进一步研究。




a. $\mathcal{H}_0$, $\varepsilon = 0.0$ b. $\mathcal{H}_1$, $\varepsilon = 0.1$ c. $\mathcal{H}_1$, $\varepsilon = 0.2$ d. $\mathcal{H}_1$, $\varepsilon = 0.3$

图9：图（a）中，在 $\mathcal{H}_0$ 下相对于显著性水平 $\alpha \in$ （0，1）的齐性假设拒绝率。图（b-d）中，在 $\mathcal{H}_1$ 下的 $p$值箱线图，针对（T2）及（L1-），其中 $\phi$（$u$） $= u$、 $n = m = 1000$、 $d = 4$、 $B = 100$。排序算法包括rSVM2、rForest、rBoost，并与最大均值差异、能量、FR和Tukey进行比较。




a. $\mathcal{H}_0$, $\varepsilon = 0.0$ b. $\mathcal{H}_1, \varepsilon = 0.1$ c. $\mathcal{H}_1, \varepsilon = 0.2$ d. $\mathcal{H}_1, \varepsilon = 0.3$

图10：图(a)中，针对齐性假设的拒绝率与显著性水平 $\alpha \in$ (0,1) 的关系，其中 $\mathcal{H}_0$ 。图(b-d)中，针对 (T3) 与 (S1)， $\mathcal{H}_1$ 下 $p$ 值的箱线图，其中 $\phi(u) = u$, $n = m = 1000$, $d = 20$, $B = 100$。排序算法 rSVM2、rForest、rBoost 与最大均值差异、能量、FR 和 Tukey 进行比较。

# 参考文献

S. Agarwal、T. Graepel、R. Herbrich、S. Har-Peled 和 D. Roth。ROC 曲线下面积的泛化界。机器学习研究期刊，6:393–425，2005。

F. Bach、Z. Harchaoui 和 E. Moulines。使用核 Fischer 判别分析进行齐性检验。载于神经信息处理系统进展 *2008*年。麻省理工学院出版社，马萨诸塞州剑桥，2008年。

| | 模型 (T1), $d = 3$ | | | |
|---|---|---|---|---|
| **原假设拒绝率** | **在 $\mathcal{H}_0$ 下** | | **在 $\mathcal{H}_1$ 下** | |
| **方法** | $\varepsilon = 0.0$ | $\varepsilon = 0.05$ | $\varepsilon = 0.15$ | $\varepsilon = 0.20$ |
| rSVM2 | **0.05 ± 0.22** | 0.06 ± 0.24 | 0.06 ± 0.24 | 0.14 ± 0.35 |
| rForest | 0.06 ± 0.24 | **0.08 ± 0.27** | 0.14 ± 0.35 | 0.17 ± 0.38 |
| rBoost | **0.05 ± 0.22** | 0.06 ± 0.24 | 0.26 ± 0.44 | 0.41 ± 0.5 |
| MMD | 0.06 ± 0.24 | 0.04 ± 0.20 | **0.35 ± 0.48** | **0.57 ± 0.50** |
| 能量 | **0.05 ± 0.22** | 0.05 ± 0.22 | 0.09 ± 0.29 | 0.15 ± 0.36 |
| FR | **0.05 ± 0.22** | 0.05 ± 0.22 | 0.08 ± 0.27 | 0.06 ± 0.24 |
| Tukey | 0.06 ± 0.24 | 0.07 ± 0.26 | 0.04 ± 0.20 | 0.07 ± 0.26 |
| | 模型 (T2), $d = 4$ | | | |
| **原假设拒绝率** | **在 $\mathcal{H}_0$ 下** | | **在 $\mathcal{H}_1$ 下** | |
| **方法** | $\varepsilon = 0.0$ | $\varepsilon = 0.10$ | $\varepsilon = 0.20$ | $\varepsilon = 0.30$ |
| rSVM2 | 0.03 ± 0.17 | 0.06 ± 0.24 | 0.04 ± 0.20 | 0.03 ± 0.17 |
| rForest | 0.05 ± 0.22 | 0.12 ± 0.33 | 0.43± 0.50 | 0.80 ± 0.40 |
| rBoost | 0.06 ± 0.24 | 0.09 ± 0.29 | 0.18 ± 0.39 | 0.34 ± 0.48 |
| MMD | 0.05 ± 0.22 | **0.18 ± 0.39** | **0.56 ± 0.50** | **0.92 ± 0.27** |
| 能量 | 0.06 ± 0.24 | 0.08 ± 0.27 | 0.16 ± 0.37 | 0.32± 0.47 |
| FR | **0.01 ± 0.10** | 0.02 ± 0.14 | 0.08 ± 0.26 | 0.28 ± 0.45 |
| Tukey | 0.04 ± 0.20 | 0.12 ± 0.33 | 0.27 ± 0.45 | 0.46 ± 0.50 |
| | 模型 (T3), $d = 20$ | | | |
| **原假设拒绝率** | **在 $\mathcal{H}_0$ 之下** | | **在 $\mathcal{H}_1$ 之下** | |
| **方法** | $\varepsilon = 0.0$ | $\varepsilon = 0.1$ | $\varepsilon = 0.2$ | $\varepsilon = 0.3$ |
| rSVM2 | 0.05 ± 0.22 | 0.06 ± 0.24 | 0.05 ± 0.22 | 0.13 ± 0.34 |
| **rForest** | 0.04 ± 0.20 | **0.23 ± 0.42** | **0.99 ± 0.10** | **1.00** |
| rBoost | 0.05 ± 0.22 | 0.07 ± 0.26 | 0.11 ± 0.31 | 0.24 ± 0.43 |
| MMD | 0.07 ± 0.26 | 0.15 ± 0.36 | 0.83 ± 0.38 | 1.00 |
| 能量 | 0.06 ± 0.24 | 0.09 ± 0.29 | 0.40 ± 0.49 | 0.98 ± 0.14 |
| FR | 0.07 ± 0.26 | 0.09 ± 0.29 | 0.63 ± 0.49 | 1.00 |
| Tukey | **0.03 ± 0.17** | 0.05 ± 0.22 | 0.04 ± 0.20 | 0.00 |

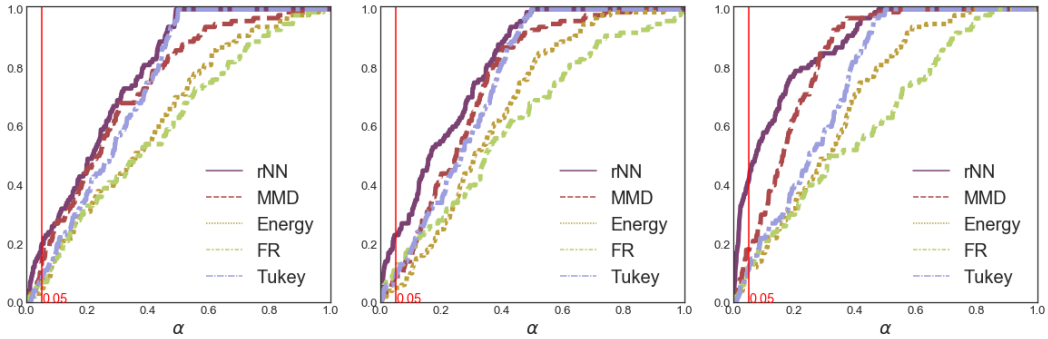
表2：在显著性水平 $\alpha = 0.05$ 下，对于模型 (T1-3)，在 $\mathcal{H}_0$ 和 $\mathcal{H}_1$ 条件下，齐性假设的拒绝率； $n = m = 1000$、 ± 为其 95% 标准差。对于排序方法，仅列出与 MWW 检验相关的结果。粗体显示的方法在所有算法中，使 $\mathcal{H}_0$ 条件下的拒绝率最小，同时使 $\mathcal{H}_1$ 条件下的拒绝率最大。

| **原假设拒绝率** | **在 $\mathcal{H}_1$ 下，** ($\varepsilon = 0.05$) | | | **在 $\mathcal{H}_1$ 下，** ($\varepsilon = 0.1$) | | |
|---|---|---|---|---|---|---|
| **方法** | $d = 30$ | $d = 60$ | $d = 100$ | $d = 30$ | $d = 60$ | $d = 100$ |
| **rNN** | **0.20 ± 0.40** | **0.23 ± 0.42** | **0.42 ± 0.50** | 0.54 ± 0.50 | **0.84± 0.37** | 0.93 ± 0.26 |
| MMD | 0.13 ± 0.34 | 0.09 ± 0.29 | 0.18 ± 0.39 | **0.89 ± 0.31** | 0.76 ± 0.43 | **0.99 ± 0.10** |
| 能量 | 0.04 ± 0.20 | 0.05 ± 0.22 | 0.11 ± 0.31 | 0.27 ± 0.45 | 0.22 ± 0.42 | 0.23 ± 0.42 |
| FR | 0.06 ± 0.24 | 0.11 ±0.31 | 0.13 ± 0.34 | 0.39 ± 0.49 | 0.43 ± 0.50 | 0.38 ± 0.49 |
| Tukey | 0.08 ± 0.27 | 0.09 ± 0.29 | 0.11 ± 0.31 | 0.11 ± 0.31 | 0.07 ± 0.26 | 0.11 ± 0.31 |

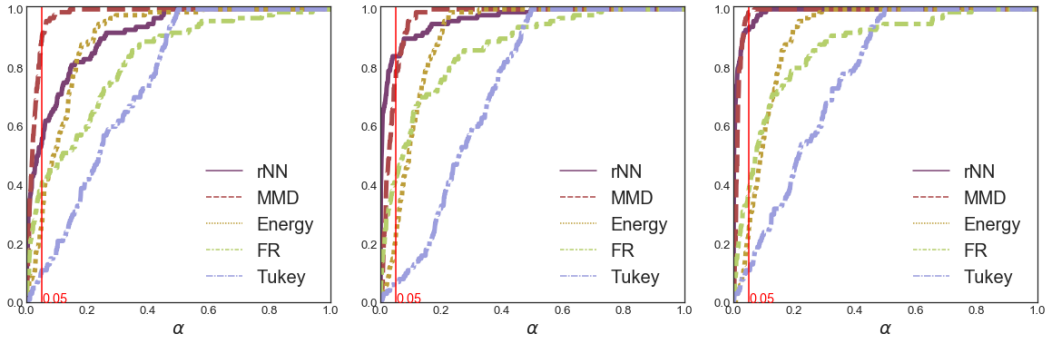
表3：尺度模型（S2）中在显著性水平$\alpha = 0.05$ 下对齐性假设的拒绝率， $\mathcal{H}_1$ ， ± 的标准差为95%。对于 rNN，仅展示与 MWW 检验相关的结果。粗体方法在所有算法中，在 $\mathcal{H}_1$下的拒绝率最高。

J. Beirlant、S. Buitendag、E. del Barrio、M. Hallin 和 F. Kamper。中心向外分位数与多元风险的度量。《保险：数学与经济学》，95：79–100，2020年。

P. Bertail、S. Cl´emen¸con、Y. Guyonvarch 和 N. Noiry。从有偏数据中学习：一种半参数方法。载于第*38*届国际机器学习会议论文集，收录于第139卷《机器学习研究论文集》，第803–812页。PMLR，2021年7月18–24日。




a. $d = 30, \varepsilon = 0.05$ b. $d = 60, \varepsilon = 0.05$ c. $d = 100, \varepsilon = 0.05$




e. $d = 30, \varepsilon = 0.1$ f. $d = 60, \varepsilon = 0.1$ g. $d = 100, \varepsilon = 0.1$

图11：对于特征空间维数不断增加的（S2），在 $\mathcal{H}_1$ 下，针对显著性水平 $\alpha \in$（0，1）的齐性假设拒绝率，其中$\phi(u) = u$、 $n = m = 1000$、 $B = 100$ 和 $\varepsilon = 0.05$ （a-c）、$\varepsilon = 0.1$ （e-g）。排序算法为 rNN，并与最大均值差异、能量、FR 和 Tukey 进行比较。

G. Biau 和 L. Gyorfi。关于非参数 $l_1$齐性检验统计量的渐近性质。*IEEE* 信息论汇刊，51(11)：3965–3973，2005年。C. Burges、T. Shaked、E. Renshaw、A. Lazier、M. Deeds、N. Hamilton 和 G. Hullen-der。使用梯度下降学习排序。载于第*22*届国际机器学习会议，第89–96页，2005年。C. Burges、R. Ragno 和 Q. Le。使用非光滑代价函数学习排序。载于神经信息处理系统进展，第19卷。麻省理工学院出版社，2007年。A. Carpentier、O. Collier、L. Comminges、A. Tsybakov 和 Y. Wang。稀疏线性回归中的检验极小极大速率。自动化与遥控，80，2018年。A. Chakraborty 和 P. Chaudhuri。针对无限维数据的 Wilcoxon-Mann-Whitney 型检验。*Biometrika*，102(1)：239–246，2015年。A. Chakraborty 和 P. Chaudhuri。基于均值、空间符号和空间秩的高维数据检验。《统计年刊》，45(2)：771–799，2017年。P. Chaudhuri。关于多元数据中分位数的几何概念。美国统计学会会刊，91(434)：862–872，1996年。

V. Chernozhukov、A. Galichon、M. Hallin 和 M. Henry。Monge–Kantorovich 深度、分位数、秩与符号。《统计年刊》，45(1):223–256，2017年。Y. K. Cheung 和 J. H. Klotz。使用链表计算 Mann–Whitney–Wilcoxon 分布。《中华统计学志》，7:805–813，1997年。S. Cl´emen¸con 和 N. Vayatis。最佳实例排序。机器学习研究期刊，8:2671–2699，2007年。S. Cl´emen¸con 和 N. Vayatis。基于线性秩统计量的经验性能最大化。载于神经信息处理系统进展，第3559卷计算机科学讲义，第1–15页。施普林格，2009a。S. Cl´emen¸con 和 N. Vayatis。基于树的排序方法。*IEEE* 信息论汇刊，55(9):4316–4336，2009b。S. Cl´emen¸con 和 N. Vayatis。最优 ROC 曲线的自适应估计与一种二部排序算法。载于*ALT'09* 会议论文集，2009c。S. Cl´emen¸con 和 N. Vayatis。叠加分类器：一种实现最优评分的实用方法。构造性逼近，32(3):619–648，2010。S. Cl´emen¸con、G. Lugosi 和 N. Vayatis。U 统计量的排序与经验风险最小化。《统计年刊》，36(2):844–874，2008年。S. Cl´emen¸con、M. Depecker 和 N. Vayatis。AUC 最大化与双样本问题。载于神经信息处理系统进展，第3559卷计算机科学讲义，第1–15页。施普林格，2009年。S. Cl´emen¸con、M. Depecker 和 N. Vayatis。二部排序的自适应划分方案。机器学习，43(1):31–69，2011年。S. Cl´emen¸con、M. Depecker 和 N. Vayatis。排序森林。机器学习研究期刊，14:39–73，2013a。S. Cl´emen¸con、M. Depecker 和 N. Vayatis。非参数评分学习算法的经验比较——treerank 算法及其他方法。模式分析及其应用，16(4):475–496，2013b。S. Cl´emen¸con、M. Limnios 和 N. Vayatis。双样本秩过程的集中不等式及其在二部排序中的应用。电子统计学杂志，15(2): 4659 – 4717，2021年。N. Deb 和 B. Sen。使用测度传输进行基于多元秩的分布无关非参数检验。美国统计学会会刊，第1–16页，2021年。Y. Freund、R. D. Iyer、R. E. Schapire 和 Y. Singer。一种用于组合偏好的高效提升算法。机器学习研究期刊，4:933–969，2003年。J. H. Friedman 和 L. C. Rafsky。Wald-Wolfowitz 检验和 Smirnov 双样本检验的多元推广。《统计年刊》，7(4):697 – 717，1979年。

O. Goldreich、S. Goldwasser 和 D. Ron。性质测试及其与学习和近似的联系。*ACM*学报，45(4):653–750，1998年。A. Gretton、K. M. Borgwardt、M. J. Rasch、B. Scholkopf 和 A. Smola。一种用于双样本问题的核方法。载于神经信息处理系统进展 *19*。麻省理工学院出版社，马萨诸塞州剑桥，2007年。A. Gretton、K. M. Borgwardt、M. J. Rasch、B. Scholkopf 和 A. Smola。一种核双样本问题。机器学习研究期刊，13:723–773，2012年。J. H´ajek 和 Z. Sid´ak。秩检验理论。学术出版社，1967年。M. Hallin 和 D. Paindaveine。基于交方向和伪马哈拉诺比斯秩的多元位置最优检验。《统计年刊》，30(4):1103 – 1133，2002a。M. Hallin 和 D. Paindaveine。基于交方向和伪马哈拉诺比斯秩的最优程序：检验多元椭圆白噪声是否存在 ARMA 依赖。伯努利，8(6):787 – 815，2002b。M. Hallin 和 D. Paindaveine。散布矩阵齐性的最优秩检验。《统计年刊》，36(3):1261 – 1298，2008年。M. Hallin、E. del Barrio、J. Cuesta-Albertos 和 C. Matr´an。d 维中的分布函数与分位数函数、秩和符号：一种测度传输方法。《统计年刊》，49(2):1139 – 1165，2021年。W. Hoeffding。有界随机变量之和的概率不等式。美国统计学会会刊，58(301):13–30，1963。T. Joachims。使用点击流数据优化搜索引擎。载于第八届 *ACM SIGKDD* 知识发现与数据挖掘国际会议论文集，KDD ’02，第133–142页。计算机协会，2002年。G. Koshevoy 和 K. Mosler。多元分布的类锥体修剪。《统计年刊》，25(5):1998–2017，1997年。J. Lam-Weil、A. Carpentier 和 B. K. Sriperumbudur。离散分布接近性检验的局部极小极大速率。伯努利，28(2):1179 – 1197，2022年。E. L. Lehmann 和 J. P. Romano。统计假设检验。施普林格，2005年。M. Limnios。高维中的秩过程与统计应用。学位论文，巴黎萨克雷大学，2022年。Liu。基于随机单纯形的数据深度概念。《统计年刊》，18 (1):405–414，1990年。R. Y. Liu 和 K. Singh。基于数据深度和多元秩检验的质量指数。美国统计学会会刊，88(421):252–260，1993年。A. Lung-Yut-Fong、C. L´evy-Leduc 和 O. Capp´e。使用秩统计量对多元数据进行齐性检验和变点检测。法国社会 ´与 ´统计学›会刊，156(4):133–162，2015年。

A. K. Menon 和 R. C. Williamson. 二部排序：风险理论视角。机器学习研究期刊，7:1–102，2016年。K. Mosler. 深度统计。载于稳健性与复杂数据结构：向 *Ursula Gather* 致敬的纪念文集，第17–34页。施普林格，2013年。J. M¨ott¨onen 和 H. Oja. 多元空间符号与秩方法。《非参数统计杂志》，5(2):201–213，1995年。J. M¨ott¨onen、H. Oja 和 J. Tienari. 关于多元空间符号与秩检验的效率。《统计年刊》，25(2):542–552，1997年。J. M¨ott¨onen、H. Oja 和 R. Serfling. 多元广义空间符号秩方法。《统计研究杂志》，39(1):19–42，2005年。D. Nolan 和 D. Pollard. $U$-过程：收敛速率。《统计年刊》，15 (2):780 – 799，1987。H. Oja. 多元分布的描述性统计。《统计与概率快报》，1(6):327–332，1983年。H. Oja. 使用$R$的多元非参数方法：基于空间符号与秩的方法，第199卷，统计学讲义。施普林格纽约出版社，2010年。M. L. Puri 和 P. K. Sen. 多元分析中的非参数方法。约翰·威立父子出版公司，1971年。A. Rachev. 概率度量与随机模型的稳定性。威立，1991年。A. Rakotomamonjy. 使用支持向量机优化ROC曲线下面积。载于第一届人工智能中的*ROC*分析研讨会论文集，2004年。R. Rubinfeld 和 M. Sudan. 多项式的稳健刻画及其在程序测试中的应用。《*SIAM*计算杂志》，25(2):252–271，1996年。C. Rudin. 使用P范数推动进行排序。载于*COLT 2006*论文集，第4005卷计算机科学讲义，第589–604页，2006年。C. Rudin、C. Cortes、M. Mohri 和 R. E. Schapire. 基于间隔的排序与提升殊途同归。载于*COLT 2005*论文集，第3559卷计算机科学讲义，第63–78页。施普林格，2005年。R. J. Serfling. 数理统计逼近定理。约翰·威立父子出版公司，1980年。G. J. Sz´ekely 和 M. L. Rizzo. 能量统计：一类基于距离的统计量。《统计规划与推断杂志》，143(8):1249–1272，2013年。J. W. Tukey. 数学与数据图示。载于国际数学家大会论文集，第2卷，第523–531页，1975年。Y. Um 和 R. H. Randles. 多元多样本位置问题的非参数检验。《中华统计学志》，8(3):801–812，1998年。

A. van der Vaart. 渐近统计学。剑桥大学出版社，1998年。

A. van der Vaart 和 J. Wellner。弱收敛与经验过程。施普林格出版社，纽约，1996年。Y. Vardi 和 C.-H. Zhang。多变量 $L_1$中位数及相关数据深度。美国国家科学院院刊，97(4):1423–1426，2000年。M. Wang、W. Deng、J. Hu、X. Tao 和 Y. Huang。野外种族人脸：通过信息最大化自适应网络减少种族偏差。载于*2019* 年 *IEEE/CVF* 国际计算机视觉大会，*ICCV 2019*，第 692–702 页。IEEE，2019年。F. Wilcoxon。基于排序方法的个体比较。《生物统计学》，1:80–83，1945年。Y. Zuo 和 R. Serfling。统计深度函数的一般概念。《统计年刊》，28(2):461–482，2000年。

## ROC 曲线的性质

为完整起见，回顾 ROC 曲线的基本性质。参见 例如 第 17 个命题，见 Cl´emen¸con 和 Vayatis [2009b]。

**命题 8**。 (Properties of the ROC curve) 对于任意概率分布 $H$和 $G$ 以及任意评分函数 $s$： $\mathcal{Z} \to \mathbb{R}$，以下性质成立：

*1.* **极限值**。我们有：

$$\mathrm{ROC}(\mathrm{s}, 0) = 0 \quad \textit{and} \quad \mathrm{ROC}(\mathrm{s}, 1) = 1 \ .$$

*2.* **不变性**。 对于任意严格递增函数 $T$： $\mathbb{R} \to \mathbb{R}$，对所有 $\alpha \in$(0,1)，均有：$\mathrm{ROC}(\mathrm{T} \circ \mathrm{s}\ \alpha) = \mathrm{ROC}(\mathrm{s}\ \alpha)$。

*3.* **凹性**。 如果似然比 $\mathrm{d}G_s/\mathrm{d}H_s$ 是单调函数，则ROC曲线是凹的。特别地，最优曲线 $\mathrm{ROC}^*$ 是凹的。

*4.***线性部分**。 如果似然比 $\mathrm{d}G_s/\mathrm{d}H_s$ 在评分函数 $s$的值域中的某个区间上为常数，那么相应定义域上的 ROC 曲线将呈现一段线性部分。此外，$\mathrm{ROC}^*$ 在 $[\alpha_1, \alpha_2]$ 上线性，当且仅当 $\Psi(z) = (\mathrm{d}G/\mathrm{d}H)(z)$ 在该子集上为常数

$$\{z \in \mathcal{Z} / \ Q^*(\alpha_2) \leq \Psi(z) \leq Q^*(\alpha_1)\} \ ,$$

其中 $Q^*(\alpha)$ 表示随机变量 $\Psi(\mathbf{Y})$在水平 $1 - \alpha \in (0,1)$ 处的分位数。

*5.* **可微性**。 假设分布 $H$ 和 $G$是连续的。那么，评分函数 $s$ 的 ROC 曲线可微，当且仅当分布 $H_s$ 和$G_s$ 是连续的。

# B 技术证明

## B.1 公式 (2.9) 的证明

给定两个取实数值、相互独立且连续的随机变量 $X \sim G$ 和 $Y \sim H$，注意到以下事实即可直接得出证明:

$$\int_{-\infty}^{\infty} \{H(t) - G(t)\} \mathrm{d}H(t) = \frac{1}{2} - \int_{-\infty}^{\infty} G(t) \mathrm{d}H(t) \ .$$

因此，我们有:

$$\begin{aligned} \mathrm{AUC_{H,G}} = \mathbb{P}\{\mathrm{Y} \leq \mathrm{X}\} = \mathbb{E}[\mathbb{E}[\mathbb{I}\{\mathrm{Y} \leq \mathrm{X}\}] \mid \mathrm{Y}] = \mathbb{E}_{\mathrm{Y} \sim \mathrm{H}}[1 - \mathrm{G}(\mathrm{Y})] \\ = \frac{1}{2} + \int_{-\infty}^{\infty} \{H(t) - G(t)\} \mathrm{d}H(t) \ . \end{aligned}$$

## B.2 命题 2 的证明

断言 $(i)$ 与 $(ii)$ 之间的等价性可直接由以下结果推出，该结果证明见 Cl´emen˙con 和 Vayatis [2009b] （参见推论 5 及其中命题 6 的证明），为清晰起见，现重述如下。

**引理 9**。 *(Cl´emen˙con 和 Vayatis [2009b])* 以下结论以概率 *1* 成立:

$$\Psi(\mathbf{Z}) = \frac{\mathrm{d}G_{\Psi}}{\mathrm{d}H_{\Psi}}(\Psi(\mathbf{Z})) \ ,$$

**Z** 表示**X** 或 **Y**。

The 断言（$ii$）与其他断言等价，利用公式（2.14即可立即得出 ).

## B.3 双样本线性秩统计量的指数不等式

下面陈述的结果给出了双样本线性秩统计量（见定义 1）的指数型尾界，其中该统计量以其渐近均值（2.8）为中心化。

**定理 10**。 设 $p \in (0,1)$ 且 $N \geq 1/p$。令 $n = \lfloor pN \rfloor$ 且 $m = \lceil (1-p)N \rceil = N - n$。设 $\{X_1, \ldots, X_n\}$ 和 $\{Y_1, \ldots, Y_m\}$ 为两个相互独立的独立同分布随机样本，分别取自单变量概率分布 $G$ 和 $H$ 。假设假设*1*成立。那么，对于所有 $t > 0$，有:

$$\mathbb{P}\left\{\frac{1}{n}\widehat{W}_{n,m}^{\phi} - W_{\phi} > t\right\} \leq 18 \exp\{-CNt^2\} \ , \tag{B.1}$$

其中 $C = 8^{-1} \min\left(p/\|\phi\|_{\infty}^2, \ (p\|\phi'\|_{\infty}^2)^{-1}, \ ((1-p)\|\phi'\|_{\infty}^2)^{-1}\right)$。

证明。证明基于 Cl´emen˙con 等人用于建立双样本线性秩过程集中不等式的线性化技术 [2021], , 参见其中的定理 5。考虑满足假设 1 的 $\phi$ 。将 $\phi$ 在 $NF_N(X_i)/(N+1)$ 处围绕 $F(X_i)$ 对 $1 \leq i \leq n$ 的二阶泰勒展开 $\phi$相加，可得统计量 $\widehat{W}_{n,m}^{\phi}$ 的几乎必然分解，参见 Cl´emen˙con 等人的式 (B.3,4) [2021]。应用

Hoeffding 分解于其一阶项后，该项具有双样本$U$-统计量的形式，我们将分解写为如下形式：

$$\frac{1}{n}\widehat{W}^{\phi}_{n,m} - W_\phi = \widehat{W}_\phi - W_\phi + \frac{1}{n}\left(\widehat{V}^X_n - \mathbb{E}\left[\widehat{V}^X_n\right]\right) + \frac{1}{n}\left(\widehat{V}^Y_m - \mathbb{E}\left[\widehat{V}^Y_m\right]\right) + \frac{1}{n}\mathcal{R}_{n,m} \ , \quad \text{(B.2)}$$

其中

$$\begin{aligned}
\widehat{W}_\phi &= \frac{1}{n}\sum_{i=1}^{n}(\phi\circ F)(X_i) \ , \\
\widehat{V}^X_n &= \frac{n-1}{N+1}\sum_{i=1}^{n}\int_{X_i}^{+\infty}(\phi'\circ F)(u)\mathrm{d}G(u) \ , \\
\widehat{V}^Y_m &= \frac{n}{N+1}\sum_{j=1}^{m}\int_{Y_j}^{+\infty}(\phi'\circ F)(u)\mathrm{d}G(u) \ .
\end{aligned}$$

量 $\mathcal{R}_{n,\ m}$ 对应于泰勒–拉格朗日余项，以及从 Hoeffding 分解中继承的阶为 $\mathcal{O}_{\mathbb{P}}(N^{-1})$ 的其他统计量。下面的引理给出了后者的尾界。

**引理 11**。 假设定理 *10* 中的假设均成立。那么，对于所有 $t>0$和 $N\geq 2$，若 $Nt \geq 128\|\phi'\|^2_\infty/(p\|\phi''\|_\infty)$，则有：

$$\mathbb{P}\left\{|\mathcal{R}_{n,m}|>t\right\} \leq 12\exp\left\{-\frac{Nt}{12\kappa_p\|\phi''\|_\infty}\right\} \ , \quad \text{(B.3)}$$

否则

$$\mathbb{P}\left\{|\mathcal{R}_{n,m}|>t\right\} \leq 12\exp\left\{-\frac{\alpha_p N^2t^2}{512\|\phi'\|^2_\infty}\right\} \ , \quad \text{(B.4)}$$

其中 $\alpha_p = \min(p/(1-p),\ 1)$，$\kappa_p = \max(p,\ 1-p)$。

对于任意 $t>0$，将经典指数尾不等式直接应用于（Hoeffding [1963], 中证明的）式（B.2）右侧前三项，可得对任意 $t>0$

$$\begin{aligned}
\mathbb{P}\left\{|\widehat{W}_\phi - W_\phi|>t\right\} &\leq 2\exp\left\{-\frac{2pNt^2}{\|\phi\|^2_\infty}\right\}, \\
\mathbb{P}\left\{\frac{1}{n}\left|\widehat{V}^X_n - \mathbb{E}\left[\widehat{V}^X_n\right]\right|>t\right\} &\leq 2\exp\left\{-\frac{2Nt^2}{p\|\phi'\|^2_\infty}\right\}, \\
\mathbb{P}\left\{\frac{1}{n}\left|\widehat{V}^Y_m - \mathbb{E}\left[\widehat{V}^Y_m\right]\right|>t\right\} &\leq 2\exp\left\{-\frac{2Nt^2}{(1-p)\|\phi'\|^2_\infty}\right\} \ .
\end{aligned}$$

然后，将上述四个尾界应用于阈值 $t/4$，结合展开式（B.2）和并集界，即可得出证明。

□

## B.4 引理 11 的证明

考虑 $\mathbb{R}^2$上的非对称有界核：

$$\begin{aligned}
k(x,x') &= \mathbb{I}\{x'\leq x\}(\phi'\circ F)(x) \ , \\
\ell(x,y) &= \mathbb{I}\{y\leq x\}(\phi'\circ F)(x) \ .
\end{aligned}$$

余项过程分解如下

$$\mathcal{R}_{n,m} = \widehat{R}_{n,m} + \frac{n(n-1)}{N+1} U_n(k) + \frac{nm}{N+1} U_{n,m}(\ell) + \widehat{T}_{n,m} \ ,$$

其中

$$\widehat{R}_{n,m} = \left( \frac{n}{N+1} - p \right) \sum_{i=1}^{n} \left( \mathbb{E}[k(X_i, X) \mid X_i] - \mathbb{E}[k(X, X')] \right) + \left( \frac{m}{N+1} - 1 + p \right) \sum_{i=1}^{n} \left( \mathbb{E}[k(X_i, Y) \mid X_i] - \mathbb{E}[k(X, Y)] \right)$$

$(X,\ X')$ 和 $Y$ 表示两个相互独立的随机变量，其分布分别为 $G \otimes G$ 和 $H$，且独立于 $X_i$; $U_n(k)$ 是基于 $\{X_1, \ldots, X_n\}$、核为 $k$, $U_{n,\ m}$的二阶单样本退化 $U$统计量；$\ell(\ell)$ 是基于两个样本 $\{X_1, \ldots, X_n\}$ 和 $\{Y_1, \ldots, Y_m\}$、核为 $\ell$, $T_{n,\ m}$ 的度数为$\widehat{}(1,\ 1)$ 的二样本退化 $U$统计量；$_m$ 是 $\phi$ 在二阶展开中的泰勒-拉格朗日积分余项，即

$$\widehat{T}_{n,m} = \sum_{i=1}^{n} \int_{F(X_i)}^{N\widehat{F}_N(X_i)/(N+1)} \left( \frac{N\widehat{F}_N(X_i)}{N+1} - u \right) \phi''(u) \mathrm{d}u \ . \tag{B.5}$$

显然有

$$|\mathcal{R}_{n,m}| \leq |\widehat{R}_{n,m}| + p^2 N |U_n| + p(1-p) N |U_{n,m}| + |\widehat{T}_{n,m}| \ .$$

现在，我们为上述界右侧求和中涉及的每一项给出一个尾界。首先，注意到映射 $x \mapsto \mathbb{E}[k(x,\ X)] = G(x)\phi' \circ F(x)$ 和 $x \mapsto \mathbb{E}[k(x,\ Y)] = H(x)\phi' \circ F(x)$ 均有界于 $\|\phi'\|_\infty$，且 $|n/(N+1) - p| \leq 1/N$以及 $|m/(N+1) - (1-p)| \leq 1/N$，将 Hoeffding不等式应用两次并结合并集界可直接得到，对任意 $t > 0$,

$$\mathbb{P}\left\{ \frac{1}{n} |\widehat{R}_{n,m}| > t/4 \right\} \leq 4 \exp\left\{ -\frac{pN^3 t^2}{32 \|\phi'\|_\infty^2} \right\} . \tag{B.6}$$

两个退化 $U$过程的尾界由 Cl´emen¸con 等人 [2021],的引理27给出，现为清晰起见重述如下。

**引理12**。 设 $P$ 和 $Q$ 分别是可测空间 $\mathcal{X}$ 和 $\mathcal{Y}$上的两个概率分布。考虑基于从 $P$和 $Q$ 分别抽取的独立同分布随机样本 $\{X_1, \ldots, X_n\}$ 和$\{Y_1, \ldots, Y_m\}$构造的、具有有界核 $\theta$: $\mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ 的退化二样本 $U$- 统计量，其阶数为 $(1, 1)$。对于所有 $t > 0$，有:

$$\mathbb{P}\left\{ U_{n,m}(\theta) \geq t \right\} \leq \exp\left\{ -\frac{nmt^2}{32 c_\theta^2} \right\} \ ,$$

其中 $c_\theta =$上确界$_{(x,y) \in \mathcal{X} \times \mathcal{Y}} |\theta(x, y)| < +\infty$。

引理 12 直接应用于 $U_{n,\ m}$ 可得，对所有 $t > 0$,

$$\mathbb{P}\left\{ \frac{p(1-p)N}{n} |U_{n,m}| > \frac{t}{4} \right\} \leq 2 \exp\left\{ -\frac{pN^2 t^2}{512(1-p)\|\phi'\|_\infty^2} \right\} \ , \tag{B.7}$$

类似地，根据Nolan 和 Pollard 中的引理 3 [1987]，我们有

$$\mathbb{P}\left\{\frac{p^2N}{n}|U_n| > \frac{t}{4}\right\} \le 2\exp\left\{-\frac{N^2t^2}{512\|\phi'\|_\infty^2}\right\} . \tag{B.8}$$

最后，由公式 (B.5) 可得

$$\begin{aligned}
&\frac{1}{n}|\widehat{T}_{n,m}| \le \|\phi''\|_\infty \left(\sup_{t\in\mathbb{R}}\left(\widehat{F}_N(t) - F(t)\right)^2 + \frac{1}{(N+1)^2}\right) \\
&\le 3p^2\|\phi''\|_\infty \sup_{t\in\mathbb{R}}\left(\widehat{G}_n(t) - G(t)\right)^2 + 3(1-p)^2\|\phi''\|_\infty \sup_{t\in\mathbb{R}}\left(\widehat{H}_m(t) - H(t)\right)^2 + \frac{13\|\phi''\|_\infty}{N^2} .
\end{aligned}$$

考虑上述界右侧涉及的各项，应用两次经典的 Dvoretzky–Kiefer–Wolfowitz 不等式，并注意到第三项相较于前两项可以忽略，我们得到：

$$\mathbb{P}\left\{\frac{1}{n}|\widehat{T}_{n,m}| > \frac{t}{4}\right\} \le 4\exp\left\{-\frac{Nt}{12\kappa_p\|\phi''\|_\infty}\right\}, \tag{B.9}$$

其中 $\kappa_p = \max(p, 1-p)$。结合并集界以及公式 (B.6)、(B.7)、(B.8) 和 (B.9)，即可完成证明。

### B.5 定理 3 的证明

只需注意到几乎必然有

$$\begin{aligned}
&\mathbb{P}_{\mathcal{H}_0}\left\{\frac{1}{n''}\widehat{W}^{\phi}_{n'',m''}(\widehat{s}) > \int_0^1 \phi(u)\mathrm{d}u + q^{\phi}_{n'',m''}(\alpha) \mid \mathbf{X}_1, \ldots, \mathbf{X}_{n'}, \mathbf{Y}_1, \ldots, \mathbf{Y}_{m'}\right\} \\
&\qquad = \mathbb{P}_{\mathcal{H}_0}\left\{\frac{1}{n''}\widehat{W}_{n'',m''} > \int_0^1 \phi(u)\mathrm{d}u + q^{\phi}_{n'',m''}(\alpha)\right\} \le \alpha
\end{aligned}$$

并对上式关于 $\{\mathbf{X}_1、\ldots\ldots、\mathbf{X}_{n'}、\mathbf{Y}_1、\ldots\ldots、\mathbf{Y}_{m'}\}$ 取期望。

### B.6 命题4的证明

注意，根据定理10，我们有

$$\mathbb{P}_{\mathcal{H}_0}\left\{\frac{1}{n}\widehat{W}^{\phi}_{n,m} - \int_0^1 \phi(u)\mathrm{d}u > \sqrt{\frac{\log(18/\alpha)}{CN}}\right\} \le \alpha ,$$

由此得到所需的界。

### B.7 定理7的证明

令 $\alpha \in (0, 1)$ 和 $\widehat{s} \in \mathcal{S}_0$ 为基于两个样本 $\mathscr{D}_{n',\,m'} = \{\mathbf{X}_1, \ldots\ldots, \mathbf{X}_{n'}\} \cup \{\mathbf{Y}_1, \ldots\ldots, \mathbf{Y}_{m'}\}$ 的步骤 *1* 解。首先注意，对于所有 $(H, G) \in \mathcal{H}_1(\varepsilon)$，都有：

$$\mathbb{P}_{H,G}\left\{\Phi_\alpha^\phi(\mathscr{D}_{n'',m''}(\widehat{s}))=0\right\}=\mathbb{P}_{H,G}\left\{\frac{1}{n''}\widehat{W}^\phi_{n'',m''}(\widehat{s})-\int_0^1\phi(u)\mathrm{d}u\le q^\phi_{n'',m''}(\alpha)\right\}$$
$$\le\mathbb{P}_{H,G}\left\{2\sup_{s\in\mathcal{S}_0}\left|\frac{1}{n'}\widehat{W}_{n',m'}(s)-W_\phi(s)\right|+\left|\frac{1}{n''}\widehat{W}^\phi_{n'',m''}(\widehat{s})-W_\phi(\widehat{s})\right|\ge\varepsilon-\delta-\sqrt{\frac{\log(18/\alpha)}{CN''}}\right\}$$
$$\le\mathbb{P}_{H,G}\left\{2\sup_{s\in\mathcal{S}_0}\left|\frac{1}{n'}\widehat{W}_{n',m'}(s)-W_\phi(s)\right|\ge\frac{\varepsilon-\delta}{4}\right\}+\mathbb{P}_{H,G}\left\{\left|\frac{1}{n''}\widehat{W}^\phi_{n'',m''}(\widehat{s})-W_\phi(\widehat{s})\right|\ge\frac{\varepsilon-\delta}{4}\right\},\tag{B.10}$$

其中 $C=8^{-1}\min\left(p/\|\phi\|_\infty^2,\ (p\|\phi'\|_\infty^2)^{-1},\ ((1-p)\|\phi'\|_\infty^2)^{-1}\right)$，使用上述分解（3.15）和以下界

$$W^*_\phi-W_\phi(\widehat{s})\le 2\sup_{s\in\mathcal{S}_0}\left|\frac{1}{n'}\widehat{W}_{n',m'}(s)-W_\phi(s)\right|+\delta\ ,$$

以及命题 4、定义 5、并集界和 $N''\ge 4\log(18/\alpha)/(C(\varepsilon-\delta)^2)$。将定理 10应用于单变量样本：

$$\{\widehat{s}(\mathbf{X}_{1+n'}),\ \ldots,\ \widehat{s}(\mathbf{X}_n)\}\text{ and }\{\widehat{s}(\mathbf{Y}_{1+m'}),\ \ldots,\ \widehat{s}(\mathbf{Y}_m)\}\ ,$$

在给定条件 $\mathscr{D}_{n',m'}=\{\mathbf{X}_1,\ldots,\mathbf{X}_{n'}\}\cup\{\mathbf{Y}_1,\ldots,\mathbf{Y}_{m'}\}$ 下（因此也给定了在步骤 *1*中得到的评分函数 $\widehat{s}$），我们几乎必然有

$$\mathbb{P}_{H,G}\left\{\left|\frac{1}{n''}\widehat{W}^\phi_{n'',m''}(\widehat{s})-W_\phi(\widehat{s})\right|\ge\frac{1}{2}\left(\varepsilon-\delta-\sqrt{\frac{\log(18/\alpha)}{CN''}}\right)\mid\mathscr{D}_{n',m'}\right\}$$
$$\le 18\exp\left(-\frac{CN''(\varepsilon-\delta)^2}{16}\right)\ .$$

取期望后，上述界给出

$$\mathbb{P}_{H,G}\left\{\left|\frac{1}{n''}\widehat{W}^\phi_{n'',m''}(\widehat{s})-W_\phi(\widehat{s})\right|\ge\frac{1}{2}(\varepsilon-\delta)\right\}\le 18\exp\left\{-\frac{CN''(\varepsilon-\delta)^2}{16}\right\}\ .\tag{B.11}$$

我们还回顾以下结果，该结果为 $W_\phi$-排序性能与其经验版本之间的最大偏离提供了尾界，详细常数见相应证明。

**定理 13.** *(Clémençon et al. [2021],* 定理 *5)* 假设定理 7 中的假设均成立。那么，存在常数 $C_1,\ C_2\ge 24$，它们取决于 $(\phi,\ \mathcal{V})$，使得对所有 $C_4\ge C_1$（取决于 $\phi$）以及所有 $t>0$:

$$\mathbb{P}\left\{\sup_{s\in\mathcal{S}_0}\left|\frac{1}{n'}\widehat{W}^\phi_{n',m'}(s)-W_\phi(s)\right|>t\right\}\le C_2\exp\{-C_3pN't^2\}\ ,$$

假设 $C_1/\sqrt{pN'}\le t\le C_4(p\wedge(1-p))$，其中 $C_3=\log(1+C_4/(4C_1))/(C_2C_4)>0$取决于 $\phi$、$\mathcal{V}$。

将上述定理应用于 $t=(\varepsilon-\delta)/8=C_4(p\wedge(1-p))$，可得：

$$\mathbb{P}_{H,G}\left\{2\sup_{s\in\mathcal{S}_0}\left|\frac{1}{n'}\widehat{W}_{n',m'}(s)-W_\phi(s)\right|\geq\frac{\varepsilon-\delta}{4}\right\}$$
$$\leq C_2\exp\left\{-\frac{N'p(p\wedge(1-p))}{8C_2}(\varepsilon-\delta)\left(1+\frac{\varepsilon-\delta}{32C_1(p\wedge(1-p))}\right)\right\}\ ,\quad \text{(B.12)}$$

一旦 $N'\geq 16C_1^2/(p(\varepsilon-\delta)^2)$. 将 (B.12) 与 (B.11) 和 (B.10) 结合起来，我们就得到所需的界

$$\mathbb{P}_{H,G}\left\{\Phi_\alpha^\phi(\mathscr{D}_{n'',m''}(\widehat{s}))=0\right\}$$
$$\leq 18\exp\left\{\frac{-CN''(\varepsilon-\delta)^2}{16}\right\}+C_2\exp\left\{-\frac{N'p(p\wedge(1-p))}{8C_2}(\varepsilon-\delta)\log\left(1+\frac{\varepsilon-\delta}{32C_1(p\wedge(1-p))}\right)\right\}.$$

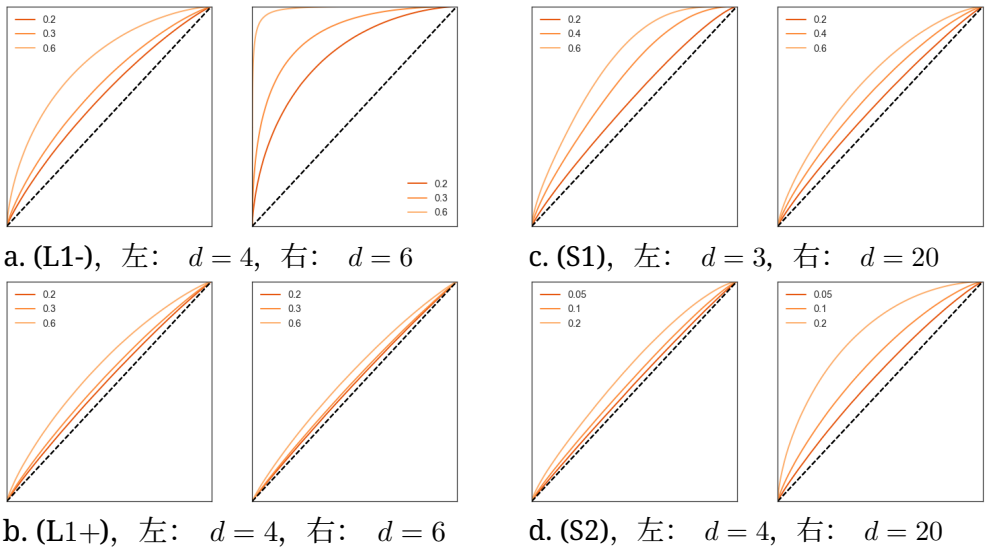
# C 附加数值实验

**合成数据集的精确参数。**

(L1-) 对于 $d=4$, $\Sigma$的对角线为 (2，6，1，5), (−1，0，0), (−1，0), (−1); $d=6$. 对于$d=6$, 矩阵对角线等于 (2，6，1，5，4，3), (−1，0，0，0，0), (−1，0，0，), (−1，0，0), (−1), (−1)。

(L1+) 对于 $d=4$, $\Sigma$的对角线为 (6，4，5，3), (−2，4，2), (−3，0), (−2); $d=6$. 对于$d=6$, 矩阵对角线等于 (6，5，5，3，2，3), (−2，4，2，1，0), (−3，0，0，1), (−2，1，1), (−3，2), (−2)。

ROC* **位置模型和尺度模型的曲线。**

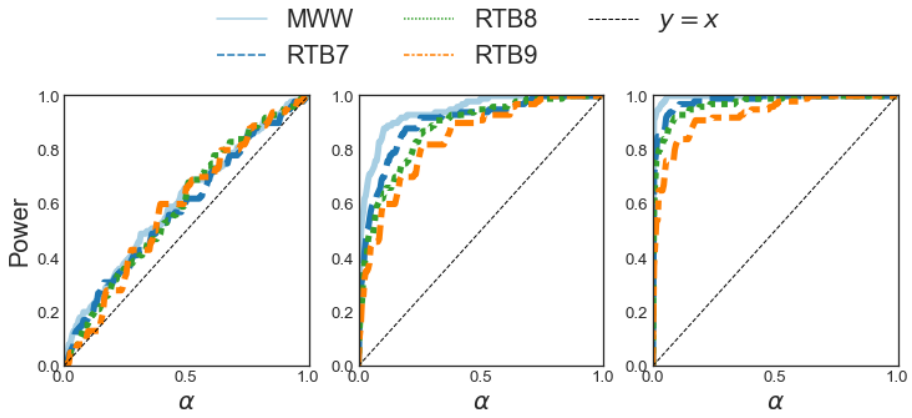



a. (L1-)，左：$d=4$，右：$d=6$　　c. (S1)，左：$d=3$，右：$d=20$

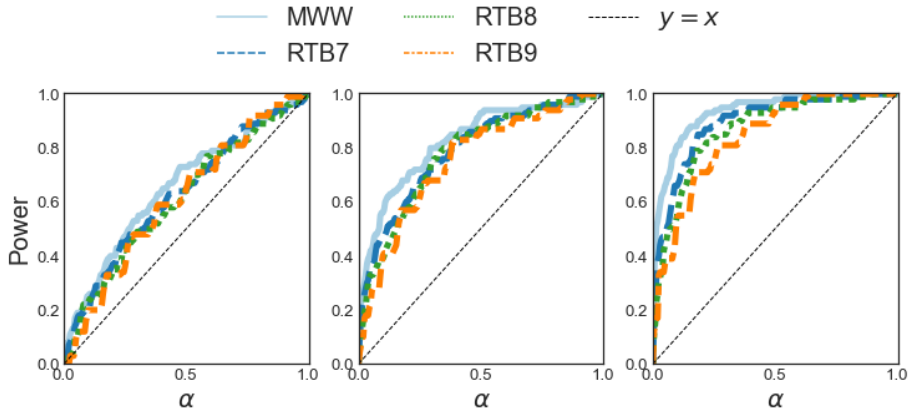
b. (L1+)，左：$d=4$，右：$d=6$　　d. (S2)，左：$d=4$，右：$d=20$

图 12：位置模型和尺度模型的真实 ROC 曲线（ROC*）随差异参数 $\varepsilon\in\{0.2,0.3,0.6\}$ 变化的情况：(L1+) 和 (L1-) 对应 $\varepsilon\in\{0.2,0.4,0.6\}$，(S1) 对应 $\varepsilon\in\{0.05,0.1,0.2\}$，(S2) 对应相应项。

**其他数值实验。** 以下图表展示了第 4 节中介绍的一些实验的数值结果，但其中为 步骤 2 实现的得分生成函数是 RTB： $\phi: \ u \mapsto u\mathbb{I}\{u \geq u_0\}$，其中 $u_0 \in \{0.7, 0.8, 0.9\}$。我们使用`rForest` 作为步骤 *1* 的二部排序算法。我们展示了在备择假设下，齐性假设的经验拒绝率相对于 检验水平 $\alpha \in$ （0，1）。




a. (L1-) , $\varepsilon \in \{0.02, 0.05, 0.08\}$ 从左到右




b. (L1+), $\varepsilon \in \{0.05, 0.08, 0.1\}$ 从左到右

图13：在备择假设下，针对模型 (L1-) (a) 和 (L1+) (b)，在 $\mathcal{H}_1$ 条件下，齐性假设的拒绝率与显著性水平$w.r.t.$ $\alpha \in$ (0,1) 的关系。使用 rForest 的步骤 *1*，使用得分生成函数 RTB $\phi$ 的步骤 *2*： $u \mapsto u\mathbb{I}\{u \geq u_0\}$ 与$u_0 \in \{0.7, 0.8, 0.9\}$ $resp.$ 分别对应 RTB7、RTB8、RTB9。与 MWW 的曲线进行比较。