arXiv:2602.02316v1 [math.ST] 2 Feb 2026

多元极值的 Kullback–Leibler 散度检验：理论与实践

Sebastian Engelke*日内瓦大学统计与信息科学研究所，瑞士；以及 Philippe Naveau †法国气候与环境科学实验室（法国国家科学研究中心）；以及 Chen Zhou 鹿特丹伊拉斯谟大学伊拉斯谟经济学院计量经济学研究所，荷兰

2026年2月3日

摘要

检验两个多元样本是否呈现相同的极值行为，是环境与气候科学等多个领域中的重要问题。尽管文献中已有一些临时采用的方法，但它们往往缺乏理论依据和统计保证。另一方面，极值理论为构建具有渐近理论依据的检验提供了理论基础。我们将这一理论与信息论和统计学中的基本概念——库尔贝克-莱布勒散度——相结合，提出一种检验在实际相关方向上的极值依赖结构是否相同的方法。在适当的假设下，我们推导了所提统计量在原假设和备择假设下的极限分布。值得注意的是，我们的检验计算速度快、易于实践者解读，因此在应用中颇具吸引力。模拟结果为检验功效提供了证据。在一个案例研究中，我们应用该方法表明，季节对法国强降雨不同聚合时段（日与小时）之间依赖强度有显著影响。

关键词： 两样本统计检验；多元正则变差；季节性降水极端事件；强度-历时-频率统计量

*S. Engelke 感谢瑞士国家科学基金会通过资助项目 186858 提供支持。

†P. Naveau 的部分研究工作得到了法国国家研究署的支持，资助项目包括：SHARE PEPR Maths-Vives 项目（法国2030 ANR-24-EXMA-0008）、EXSTA、PORC-EPIC、PEPR TRACCS 计划（PC4 EXTENDING, ANR-22-EXTR-0005）以及 PEPR IRIMONT（法国2030 ANR-22-EXIR-0003）。他还受益于 Geolearning研究讲席。

1 引言

在环境风险分析中，一个反复出现的问题是，判断两个多元极端事件在统计上是否无法区分。例如，这一问题是气候科学中极端事件归因领域的核心（如 Naveau 等人，2020年），该领域旨在比较存在与不存在人为强迫的世界中极端事件发生的概率。为识别复合气候极端事件的显著变化，气候学界采用了多种统计方法（如 Zscheischler 等人，2020），但这些方法缺乏统计保证。除了气候研究之外，金融领域的系统性风险分析（如 Embrechts 等人，1997年）和环境科学（如 Huser 和 Davison，2014；Buishand 等人，2008）中，也能找到其他例子，说明需要采用度量来比较多元极值行为。

例如，洪水风险管理者希望确定：除了不同雨量计之间的边际差异外，极值依赖是否会随季节、空间位置或时间聚合尺度而显著变化。为说明这一点，我们比较了法国波尔多市在2006–2023年期间6分钟尺度和小时尺度的日降水极大值；见图1左侧各面板，其中顶行和底行的散点图分别对应冬季和秋季。比较不同时间尺度（此处为6分钟与小时尺度）下季节对降雨聚合的影响，是水文学中的一个重要研究课题，至少半个世纪前便已出现；例如，美国气象局的强度-历时-频率曲线（Hershfield，1961）。关于不同聚合尺度下边际降雨分布的建模与解释，已有大量研究（例如，Ulrich 等人，2020；Haruna 等人，2023年）。自然会让人思考，联合极值行为是否也会随季节变化？也就是说，图1 中的灰色和黑色云团是否有所不同？多元极值理论为描述这一现象提供了数学基础

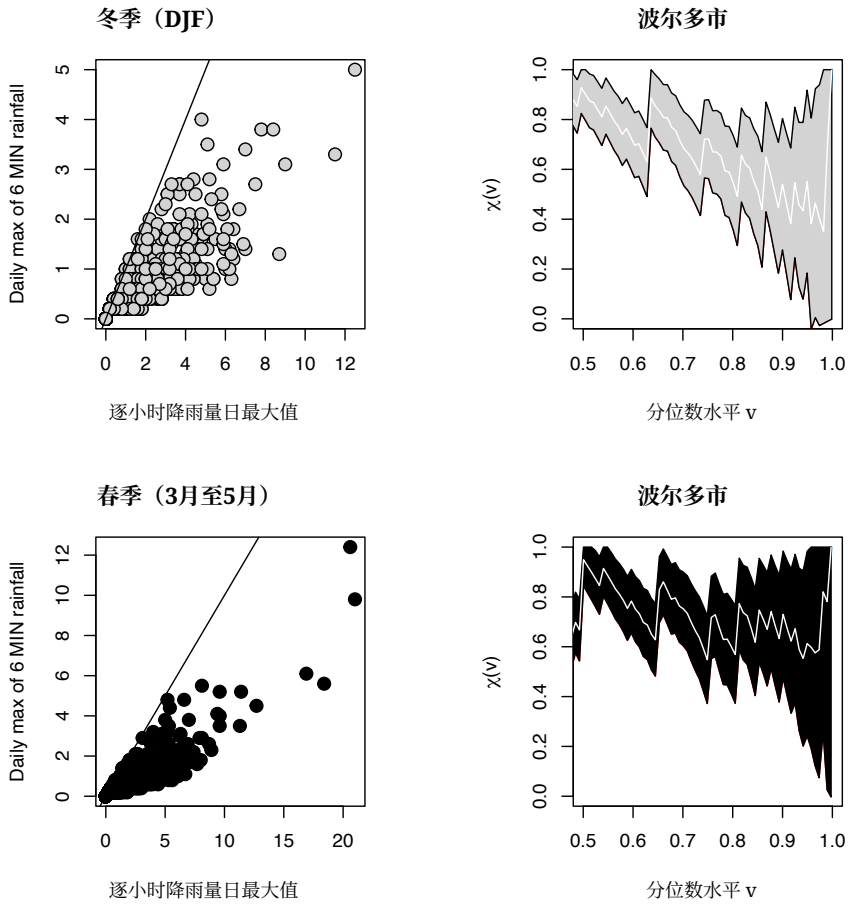


图1：左：2006–2023年冬季（上）和春季（下）期间在法国波尔多记录的每小时和6分钟降雨日最大值散点图。右：该数据的极值相关系数 $\chi(v)$，带有置信区间，用于衡量分位数水平 v以上的依赖强度；另见（3）。

极值依赖结构，并解答此类问题（例如，Beirlant 等人，2005年；de Haan 和 Ferreira，2006）。一种初步且略显简单的做法，是用极值依赖结构的汇总统计量（例如极值相关）来比较两个二元极值总体 χ （Coles 等人，1999年），其定义见下文中的（3）。对于波尔多降雨示例，图1 右侧各面板展示了极值相关的估计值（针对不同阈值）及其95%置信区间。然而，从视觉上和理论上都难以判断冬季与春季的极值是否

极值行为不同。从统计上来说，相关检验的检验功效会相当低。除了汇总统计量之外，还可以通过参数化或非参数化方式对完整的极值依赖进行建模（B¨ucher 等人，2017）。尽管这种方法可以提高检验功效，但它要求具备多元极值理论方面的专业知识，而且计算成本要高得多。在这项工作中，我们选择了一条折中路线：在保持与 χ 检验相近的计算简便性和可解释性的同时，增加灵活性以提高检验功效。

˜从数学上说，我们的目标是检验两个 d 维随机向量 $\boldsymbol{X} = (X_1, \ldots, X_d)$ 和 $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ 是否具有相同的极值依赖结构。记˜ ˜的第 j 个边缘分布为 F_j ， $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的第 j 个边缘分布分别为 F_j 和 G_j。为避免不同边缘分布（例如 6分钟降雨量分布与逐小时降雨分布）带来的影响，我们考虑分量为 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的向量

$$X_j = 1/(1 - F_j(\tilde{X}_j)), \qquad Y_j = 1/(1 - G_j(\tilde{Y}_j)), \quad j = 1, \ldots, d, \tag{1}$$

标准化为帕累托边缘。由于该变换不会改变 Copula，˜ ˜比较 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 之间的极值依赖结构可以通过比较 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 之间的极值依赖结构来实现。

对于多元数据，极端事件的定义并不明确。极值理论中的一种常见做法是用单变量风险泛函 $r\colon [0, \infty)^d \to [0, \infty)$ 对感兴趣的向量进行概括，该泛函刻画了在特定应用中尤为相关或影响重大的事件类型。比如，$\boldsymbol{X}$ 的一个极端观测定义为 r-超越集 $\{r(\boldsymbol{X}) > u\}$，其中高阈值 $u > 0$规定了极端事件的幅度（例如，Dombry 和 Ribatet，2015；de Fondeville 和 Davison，2022）。风险泛函的典型例子包括 $r(\boldsymbol{x}) = \max(x_1, \ldots, x_d)$、$r(\boldsymbol{x}) = \min(x_1, \ldots, x_d)$ 或 $r(\boldsymbol{x}) = \sqrt{x_1^2 + \cdots + x_d^2}$；见 $d = 2$ 中图2的三个面板。直观而言，我们的检验通过将超越集 $\{r(\boldsymbol{X}) > u\}$（该

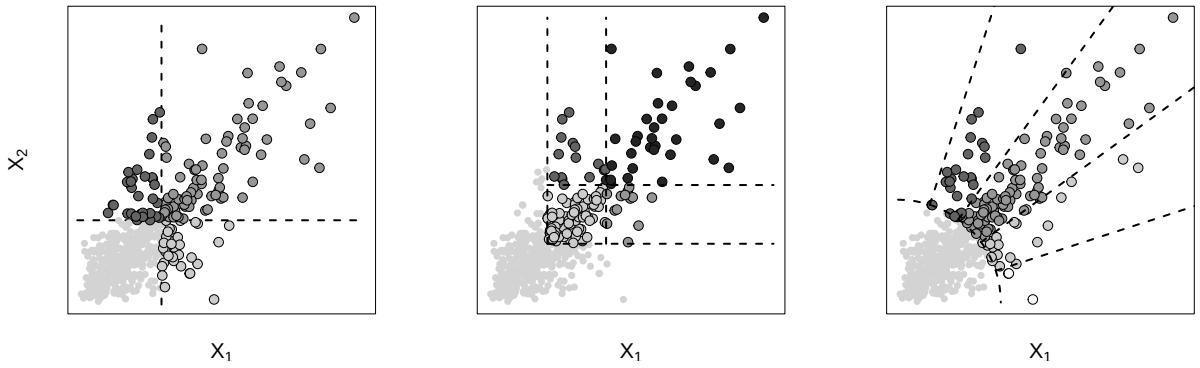


图2：三个风险泛函 $r(\boldsymbol{X})$，以及基于 $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ 的模拟双变量抽样得到的划分示例。浅灰色圆点表示非极端实现值，每个面板中圈出的点表示极端值，即 $\{\boldsymbol{X}_i : r(\boldsymbol{X}_i) > u\}$，其定义依据风险泛函 $r(\boldsymbol{x}) = \max(x_1, x_2)$（左侧面板）；$r(\boldsymbol{x}) = \min(x_1, x_2)$（中间面板）；$r(\boldsymbol{x}) = \sqrt{x_1^2 + x_2^2}$（右侧面板）。区域 $\{r(\boldsymbol{X}) > u\}$ 被划分为 K 个子集，以便使用右侧、中间和右侧面板分别对应的 $K = 3$、4 和 5 来估计 (2)。

图中颜色较深的点2）中 K 不同的区域，并检查这些集合中的极端事件数量在以下样本之间是否存在差异： $\boldsymbol{X}$ 和 $\boldsymbol{Y}$；有关此类划分的示例，请参见图2 中不同深浅的灰色点。更严格地说，如果 r 是一次齐次的，我们可以写成

$$\{r(\boldsymbol{X}) > u\} = u\Omega_r \text{ with } \Omega_r := \{\boldsymbol{x} \in [0, \infty)^d : r(\boldsymbol{x}) > 1\}.$$

对于 Ω_r 的博雷尔集的不相交划分 A_1，……， A_K，我们的主要假设是条件概率的极限存在

$$p_j := \lim_{u \to \infty} p_j(u) := \lim_{u \to \infty} P(\boldsymbol{X} \in uA_j \mid r(\boldsymbol{X}) > u) \in (0, 1), \qquad j = 1, \ldots, K. \tag{2}$$

这一假设适用于许多多元分布，尤其在常用的多元正则变差框架下成立（Resnick, 2008）。在

二元情形下，对 Ω_r with $r(\boldsymbol{x}) = \max(x_1, 0)$进行特定划分，可直接关联到极值相关

$$\chi(v) := p_1(u) = P(X_2 > u \mid X_1 > u) = P(\boldsymbol{X} \in uA_1 \mid r(\boldsymbol{X}) > u) \tag{3}$$

其中 $v = 1 - 1/u \in (0, 1)$ 和 $A_1 = [1, \infty) \times [1, \infty)$。这表明，我们的方法通过将超越集 $\{r(\boldsymbol{X}) > u\}$ 划分为更小的子区域，以捕捉该集合内更详细的信息，从而扩展了 χ 所提供的信息。

类似地，我们为随机向量 $\boldsymbol{Y}$定义概率 q_j 。本文的主要目标是检验原假设

$$H_0 : p_j = q_j, \quad \text{for all } j = 1, \ldots, K, \tag{4}$$

与备择假设相对，即这些概率中至少有一个不相等。如果 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的极值依赖结构一致，那么 H_0 成立；或者等价地，如果拒绝 H_0，那么极值依赖结构就存在差异。与 B¨ucher 等人（2017）所采用的完整极值依赖结构检验相比，较简单的原假设（4）在实践中往往更有意义，因为它易于解释。事实上，关注特定集合使我们能够定义一种计算快速、易于解释且具有良好渐近性质的检验统计量。

我们的检验依赖于多项Kullback–Leibler（KL）散度

$$D_K := \sum_{j=1}^{K} (p_j - q_j)(\log p_j - \log q_j), \tag{5}$$

两总体之间。散度 D_K 非负，且仅当 H_0ˆ成立时为零。我们基于$\boldsymbol{X}$ 和 $\boldsymbol{Y}$的样本构造理论 D_K 的经验估计量 D_K ，并用它来获得一个具有目标显著性水平 $\alpha \in (0, 1)$的统计检验。若 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 落入超越集的概率相近，则ˆ统计量 D_K 较小；若统计量较大，则拒绝 H_0。为使这一点形式化，我们推导了 D_K的渐近大样本理论。具体而言，我们证明，如果边缘分布函数

在（1）中已知，那么经过适当标准化的ˆ D_K渐近服从 χ^2（$K-1$）分布。另一方面，如果边缘分布未知，我们则使用经验分布函数进行标准化。此时的极限分布更为复杂，因此我们提出一种自助法程序来获得临界值。

我们基于 KL散度构建的极值依赖检验，其优势在于它在应用科学领域具有广泛的适用性。例如，Naveau 等人(2014) 利用 KL散度研究单变量气候极端事件。关于多变量框架，已有若干研究应用了我们的检验统计量 D_K （5），但关注的完全是应用，而非检验的统计性质。例如，Zscheischler 等人(2020) 评估了不同数据产品中的复合风雨极端事件是否具有相似的依赖结构。类似地，我们的检验也被用于模型评估，比较美国沿海地区海洋、河流和暴雨洪水驱动因子对应的观测数据与气候模型的极值依赖（Nasr 等人，2021）以及中国沿海地区的相关情况（Li 等人，2024）。Nasr 等人(2023年) 检验了复合沿海与内陆洪水之间依赖关系的时间变化。KL散度的另一种视角 D_K 是将其解释为衡量极值依赖的一般距离度量。沿着这一思路，Vignotto 等人(2021) 利用它根据大不列颠和爱尔兰的风-降水极端事件划分同质聚类。

第2 节介绍KL散度检验并推导其渐近理论，其中已知和未知边缘分布的情况分别在第2.3 节和2.4节中讨论。为评估检验的性能，第3节详细介绍了一项模拟研究。继图1中的案例研究之后，我们在第4 节考察季节周期是否会对6分钟和每小时极端降水之间的依赖结构产生显著影响；本研究的数据可在以下网址公开获取：
https://meteo.data.gouv.fr/
datasets/donnees-climatologiques-de-base-6-minutes/所有证明见附录。

2 双样本极值依赖KL散度检验

~ ~ ~ ~ ~ ~我们研究 d维随机向量 $\boldsymbol{X} = (X_1, \ldots, X_d)$和 $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$，并旨在检验它们是否具有相同的极值依赖结构。我们始终假设，式（1）中定义的标准化向量 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$均为多元正则变动的，因此概率 p_j 和 q_j 均有明确定义。

第2.1 节介绍多元正则变差的基本背景，并在第2.2节基于观测样本定义我们的检验统计量。实际上，如果~ ~的边缘分布以及 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的边缘分布已知，则可以获得这样的样本。对于这种情况，我们在第2.3节给出关于检验尺寸和检验功效的统计理论。第2.4节讨论边缘分布未知的情形。我们证明此时检验仍然有效，但临界值需要通过自助法程序获得。

2.1 背景与多元正则变差

我们考虑一个一般的随机向量 $\boldsymbol{X} = (X_1, \ldots\ldots, X_d$，其边缘分布已根据（1）标准化为帕累托边缘。为了刻画 $\boldsymbol{X}$的极值依赖，一种常见方法是假设其服从多元正则变差（MRV）（Resnick，2008）。这一假设等价于存在一个概率测度 ν，我们称之为超越分布，其定义为

$$\lim_{u\to\infty} P(\boldsymbol{X}/u \in B \,|\, \|\boldsymbol{X}\|_\infty > u) = \nu(B), \tag{6}$$

对于所有博雷尔集 $B \subset \{\boldsymbol{x} \geq 0 : \|\boldsymbol{x}\|_\infty > 1\}$，使得 $\nu(\partial B) = 0$。可以证明，对于任意 $t > 1$， ν 满足齐次性 $\nu(tB) = t^{-1}\nu(B)$。

测度 ν 描述随机向量 $\boldsymbol{X}$ 的尾部。事实上，对于固定的 u >0，式（6）左侧是归一化超越量的分布，也就是

至少有一个分量为极端值的观测值；见图 2 中颜色较深的点。由于每个分量 X_j， $j = 1$, ..., d 都服从相同的边际帕累托分布， ν刻画了超越值之间的依赖结构。将式（ $u6$）中的阈值趋于无穷大，可确保只考虑最极端的观测值，这对于后续为我们的检验程序建立恰当的理论至关重要。

备注 1。虽然每个 X_j 由变换(1) 得到的变量都具有标准帕累托边˜缘分布， F_j 但原始数据 X_j 的边缘分布可以是任意的。例如，有些分量可能具有重尾，而另一些分量则可能具有轻尾，甚至具有有限上端点。在气候科学中，这一点很重要，因为人们往往关注由尾部行为差异较大的变量构成的复合极端事件，例如温度与降水的联合极端事件（ *Zscheischler andSeneviratne*， *2017*）或风-降水极端事件（*Zscheischler et al.*， *2020*）。

在（6）中，超越值是指至少有一个分量为极端值的观测；对于多元数据，极端事件的定义通常取决于应用领域的具体情境。这一假设同样适用于其他类型事件的超越分布。为了建立一个通用框架，正如引言中所述，我们使用齐次风险泛函的概念 r： $[0, \infty)$ $^d \to [0, \infty)$ 满足 $r(t\boldsymbol{x}) = t\, r(\boldsymbol{x})$ 对任意 $t > 0$ 及 $x \in [0, +\infty)^d$。此外，我们假设 r是可测的，因此集合 Ω_r： $= \{\boldsymbol{x} \in [0, \infty)^d$： $r(\boldsymbol{x}) > 1\}$ 是一个有界且远离原点的博雷尔集。下面的例子给出了一些典型的风险泛函，以及相应集合 Ω 的划分r。

例1。 考虑风险泛函 $r(\boldsymbol{x}) = \max(x_1, ..., x_d)$。对于任意非空$I \subset \{1, \ldots, d\}$， 定义集合 $A_I = \{x \in \mathbb{R}^d$： $x_j > 1$， $j \in I$； $x_j \leq 1$， $j \in I^c\}$。则 $(A_I)_I$是 $K = 2^d - 1$ 集合的一个划分。图 *2* 左侧展示了 $d = 2$维空间中的这一划分。

例2。 考虑风险泛函 $r(\boldsymbol{x}) = \min(x_1, \ldots\ldots, x_d)$。对于任意 $I \subset \{1, \ldots, d\}$，定义集合 $A_I = \{x \in \mathbb{R}^d : \ x_j > 2, \ j \in I; \ 1 < x_j \leq 2, \ j \in I^c\}$。则 $(A_I)_I$ 是 $K = 2^d$ 集合的一个划分。图 2 中央展示了 $d = 2$维空间中的这一划分。

例 3。对于如上所述的一般风险泛函 r，令 $S_r = \left\{x \in \mathbb{R}^d : r(\boldsymbol{x}) = 1\right\}$ be˜ ˜相应的“单位球面”。对于 S_r 的任意划分 $A_1, \ldots, A_K$，集合 $A_j = \{x \in \mathbb{R}^d : x/r(\boldsymbol{x}) \in \tilde{A}_j, r(\boldsymbol{x}) > 1\}$, $j = 1, \ldots, K$ 构成 Ω_r 的一个划分。举例来说，图2右侧展示了维度为 $d = 2$ 、满足 $K = 5$ 时，风险泛函 $r(\boldsymbol{x}) = \sqrt{x_1^2 + x_2^2}$ 所对应的这一划分。

下一个结果表明，在满足多元正则变异假设时，对于齐次风险泛函 r，式（2）中的极限概率存在，并且可以用 ν的扩展定义简洁地表示：对于任意 $B \subset [0, \infty)^d \setminus \{0\}$ 远离原点的集合，也就是说， $\mu_B := \min\{\|\boldsymbol{x}\|_\infty : \ \boldsymbol{x} \in B\} > 0$，定义 $\nu(B) = \nu(B/\mu_B)/\mu_B$. 注意，这个扩展的 ν也称为指数测度（de Haan 和 Ferreira，2006，第6章），满足次数为 -1的齐次性，对所有 $t > 0$均成立。

命题1。 假设 (6) 中的多元正则变异假设对 $\boldsymbol{X}$ 成立，其指数测度为 ν。那么，对于任意博雷尔集 $B \subset \Omega_r$，若 $\nu(\partial B) = 0$，则

$$\lim_{u \to \infty} P(\boldsymbol{X} \in uB | r(\boldsymbol{X}) > u) = \frac{\nu(B)}{\nu(\Omega_r)}.$$

极限概率 p_j和 q_j， $j = 1, \ldots\ldots, K$，在（2）中可分别根据其超越分布 ν_X 和 ν_Y以类似方式表示。

2.2 Kullback–Leibler检验统计量

令 $\boldsymbol{X}_1, \ldots\ldots, \boldsymbol{X}_n$和 $\boldsymbol{Y}_1, \ldots\ldots, \boldsymbol{Y}_n$分别为随机向量 $\boldsymbol{X}$和 $\boldsymbol{Y}$的独立样本。假设这两个随机向量均满足多元正则变异假设（6）。设风险泛函 r和划分集合 $A_1, \ldots\ldots, A_K$ 均为固定值。下文中，

我们详细介绍一种算法，用于计算原假设（4）下的 Kullback–Leibler 散度检验统计量估计值，以评估两个样本在极值依赖方面的差异。为估计（2）中的极限概率，我们选择数据依赖阈值 u_n ，并对条件概率 $p_j(u_n)$进行经验估计。为此，记 $R^X_{1,\,n} \le \ldots \le R^X_{n,\,n}$ 为集合 $\{r(\boldsymbol{X}_i)\}_{i=1}^n$的次序统计量。对于某个$k_n \le n$ ，我们设定 $u_n^X = R^X_{n-k_n,\,n}$，因此恰有 k_n 个超越值，因为

$$k_n = \sum_{i=1}^{n} \mathbf{1}\{r(\boldsymbol{X}_i) > R^X_{n-k_n,n}\}.$$

p_j、 $j = 1$、……、 K的经验估计量形式如下

$$\hat{p}_j = \hat{p}_j(u_n^X) = \frac{1}{k_n} \sum_{i=1}^{n} \mathbf{1}\{\boldsymbol{X}_i \in R^X_{n-k_n,n} A_j\}. \tag{7}$$

经验估计量 ˆq_j<style id='4'}的 q_j， $j = 1$，……， K，可根据 $\boldsymbol{Y}$的样本以类似方式定义。将这些概率估计代入（ D_K （5）中，可得库尔贝克-莱布勒散度的经验估计量如下

$$\hat{D}_K = \hat{D}_K(n) := \sum_{j=1}^{K} (\hat{p}_j - \hat{q}_j)(\log \hat{p}_j - \log \hat{q}_j). \tag{8}$$

ˆ 在原假设下，后者可与 D_K的极限卡方分布的分位数进行比较，其中 $n \to \infty$，从而得到检验统计量的临界值；渐近分布的推导见第2.3 节。

˜ ˜在实际应用中，我们拥有从原始随机向量 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$、
˜ ˜ ˜ ˜记为 $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ 以及 $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$中抽取的独立同分布样本。因此，要获得从˜ ˜$\boldsymbol{X}$ 和 $\boldsymbol{Y}$中抽取的独立同分布样本，必须已知随机向量 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$的边缘分布。如果边缘分布未知，则可以根据数据通过参数化或非参数化方法进行估计。由于我们关注的是检验极值依赖，˜并且为避免参数假设，我们采用经验估计。记 $\boldsymbol{X}_i =$
˜ ˜ ˆ $(X_{i1}, \ldots, X_{id})$，并将基于这些样本的经验分布函数记为 F_j

$\{\tilde{X}_{ij}\}_{i=1}^{n}$。我们定义新样本 ˆ $\boldsymbol{X}_i$、 $i = 1, .., n$，其各项为

$$\hat{X}_{ij} = 1/\{1 - \hat{F}_j(X_{ij})\} = \frac{n+1}{n+1-\text{rank}(\tilde{X}_{ij})}, \quad j = 1, \ldots, d, \tag{9}$$

˜ ˜ ˜其中，rank(X_{ij})表示 X_{ij} 在样本 $\{X_{kj}\}_{k=1}^{n}$中的秩。类似地，我们为第二个样本定义ˆ$\boldsymbol{Y}_i$、 $i = 1$、.....、 n。这些样本被称为伪观测，因为ˆ ˆ$\boldsymbol{X}_1$、.....、 $\boldsymbol{X}_n$不再相互独立。这是因为边缘分布是根据同一数据集估计得到的。此外，由于秩的取值是确定的，共同边缘分布并不完全服从标准帕累托分布。

随后，我们通过使用这些伪观测计算（7）中的经验概率，并将其代入（8）中的散度定义，得到Kullback–Leibler检验统计量 D_K 。由于还需对经验边缘分布进行标准化，D_K 的极限分布更为复杂，通常只能通过自助法程序进行评估；详情请参见第3.1节。

算法 1 总结了我们的库尔贝克-莱布勒散度极值相依检验的各个步骤。

2.3 Theo已知边缘分布下的retical保证

在本节中，我们推导ˆ Kullback–Leibler 散度检验统计量 D_K 的渐近分布的理论保证，该统计量定义于第 2.2 节，随样本量$n \to \infty$。我们首先在已知边缘分布的假设下进行，从而可以获得抽自 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$的分布的独立同分布样本，这些样本用于ˆ构造检验统计量 D_K。我们假设 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 满足多元正则变化条件（6），尾指数分别为 α_X 与 α_Y，指数测度分别为 ν_X 与 ν_Y。

ˆ 为了展示 D_K 的渐近性质，我们需要针对风险泛函、可能的划分以及 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$的分布陈述一些技术条件。风险泛函 r 应与第 2.1 节中的一致，并附加限制 $\nu_X(\partial\Omega_r) =$

算法 1 Kullback–Leibler 散度检验

输入： 两个数据集 $\tilde{\boldsymbol{X}}_1,\ldots,\tilde{\boldsymbol{X}}_n$ 和 $\tilde{\boldsymbol{Y}}_1,\ldots,\tilde{\boldsymbol{Y}}_n$；一个风险泛函 r；划分集合 A_1，……，A_K；要使用的超越值数量 k_n；显著性水平 $\alpha \in$（0，1）。

输出： 检验统计量 $\hat{D}_K$、p 值，以及原假设是否被拒绝。

1: **过程** KULLBACK–LEIBLER DIVERGENCE TEST（k_n，α）

2: 标准化边缘分布：

a）若边缘分布已知，则对两个数据集均使用（1）；b）若边缘分布未知，则通过（9）利用经验分布函数为两个数据集获取伪观测，并使用这些变换后的样本。

3: 对每个数据集，选取 k_n 个超越值，并计算概率估计 $\hat{p}_j$ 和 $\hat{q}_j$，依据（7）。

4: 计算观测值 D_K，即（5）中的Kullback–Leibler散度，通过代入$\hat{p}_j$ 和$\hat{q}_j$。

5: 利用原假设下的分布的近似计算 p值：a）若边缘分布已知

$$p = \mathbb{P}\left(\chi^2(K-1) > k_n \hat{D}_K/2\right);$$

b）若边缘分布未知，第3.1节所述的自助法程序得到 B 个自助法版本$\hat{D}_K^{(1)}$，……，$\hat{D}_K^{(B)}$，即 H_0下检验统计量的相应版本，并且

$$p = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\{\hat{D}_K^{(b)} > \hat{D}_K\}.$$

6: **返回值：** D_K 的值、p 值；若 $p < \alpha$，则“拒绝 H_0”，否则“不拒绝 H_0”。

$\nu_Y(\partial\Omega_r) = 0$，其中 $\Omega_r = \{\boldsymbol{x} \in [0,\infty)^d : \ r(\boldsymbol{x}) > 1\}$。若超越分布具有密度，则此假设成立；大多数参数模型都符合这一情况（例如，Colesand Tawn，1991；Engelke 等人，2015）。

为了控制概率估计 $\hat{p}_j$ 相对于其理论极限的误差，我们可以将其分解

$$|\hat{p}_j - \nu_X(A_j)/\nu_X(\Omega_r)| \le \left|\hat{p}_j - P(\boldsymbol{X} \in u_n^X A_j \mid r(\boldsymbol{X}) > u_n^X)\right| + \left|P(\boldsymbol{X} \in u_n^X A_j \mid r(\boldsymbol{X}) > u_n^X) - \nu_X(A_j)/\nu_X(\Omega_r)\right|.$$

第二项误差是确定性近似偏差，源于阈值 u_n 为有限这一事实。该误差通常出现在极值统计中，并在如下标准二阶条件下随阈值增大而消失。我们假设存在二阶尺度函数 a_X，使得当 $u \to \infty$ 时，$a_X(u) \to 0$，并且对任意满足 $\nu_X(\partial B) = 0$ 的博雷尔集 $B \subset [0,+\infty)^d \setminus [0,\epsilon)^d$，

$$\mathbb{P}(\boldsymbol{X}/u \in B \,|\, \|\boldsymbol{X}\|_\infty > u) - \nu_X(B) = O(a_X(u)). \tag{10}$$

为了定义一大类可容许划分 $A_1, \ldots, A_K$（即集合 Ω_r 的划分），我们作如下假设。令 $\mathcal{B}(\Omega_r)$ 表示 Ω_r 的所有博雷尔子集，并令

$$\mathcal{A}_r = \{B \subset \mathcal{B}(\Omega_r) : tB \subset B \text{ for all } t > 1\} \tag{11}$$

为 Ω_r 的所有博雷尔子集，且其膨胀后仍属于同一集合。

假设1. 设 r 为齐次的可测风险泛函。设 $A_1, \ldots, A_K$ 为 $\Omega_r = \{\boldsymbol{x} \in [0,\infty)^d : \ r(\boldsymbol{x}) > 1\}$ 的一个划分，使得对所有 $j = 1, \ldots, K$，集合 A_j 可由 $\mathcal{A}_r$ 中的集合经有限次并集、交集与补运算表示。

我们可以很容易地验证，第2.1 节中的例子都满足这一假设。事实上，例3 中划分的集合已经包含在 $\mathcal{A}_r$中。在例1中，每一个

集合 A_I 可以表示为形如 $B_I = \{x \in \mathbb{R}^d$ 的集合的有限交集：$x_j > 1$，$j \in I\} \in \mathcal{A}_r$。类似地，在例2中，集合 A_I 是形如 $B_I = \{x \in \mathbb{R}^d$ 的集合的有限交集：$x_j > 2$，$j \in I\} \in \mathcal{A}_r$。更一般地，假设1 非常温和，并且对大多数实际相关的例子都成立。

ˆ 以下定理展示了检验统计量 D_K 在原假设和备择假设下的渐近行为。

定理 *2.* 设随机向量 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 满足条件 *(10)*，其指数测度分别为 ν_X 和 ν_Y，二阶尺度函数分别为 a_X 和 a_Y。假设风险泛函 r 和集合 Ω_r 的划分 $A_1, \ldots, A_K$ 满足假设*1*，其中 $\nu_X(\partial A_j) = \nu_Y(\partial A_j) = 0$，适用于 $j = 1, \ldots, K$。相应地，定义 $U_X(t) = \inf\{u: \mathbb{P}(\|\boldsymbol{X}\|_\infty > u) \leq 1/t\}$ 和 U_Y。设 k_n 为一个序列，使得当 $n \to \infty$ 时，$k_n \to \infty$，$k_n/n \to 0$，$\sqrt{k_n} a_X\{U_X(n/k_n)\} \to 0$ 且 $\sqrt{k_n} a_Y\{U_Y(n/k_n)\} \to 0$。

— —

在原假设 H_0 下，(4)中的库尔贝克-莱布勒散度依分布收敛于具有 $K-1$ 个自由度的卡方分布，即

$$\frac{k_n}{2} \hat{D}_K \xrightarrow{d} \chi^2(K-1), \quad n \to \infty. \tag{12}$$

在备择假设 $p_j \neq q_j$ 对某些 $j = 1, \ldots, K$ 成立时，我们有

$$\sqrt{k_n}\left(\hat{D}_K - D_K\right) \xrightarrow{d} N(0, \sigma^2), \quad n \to \infty,$$

其中 D_K 定义见 (5)，且

$$\sigma^2 = \sum_{j=1}^{K} D_{1j}^2 p_1(1-p_1) + D_{2j}^2 q_1(1-q_1) - 2 \sum_{1 \leq j_1 < j_2 \leq K} (D_{1j_1} D_{1j_2} p_{j_1} p_{j_2} + D_{2j_1} D_{2j_2} q_{j_1} q_{j_2})$$

with

$$D_{1j} = \frac{\partial D_K}{\partial p_j} = \log \frac{p_j}{q_j} + 1 - \frac{q_j}{p_j}, \text{ and } D_{2j} = \frac{\partial D_K}{\partial q_j} = \log \frac{q_j}{p_j} + 1 - \frac{p_j}{q_j},$$

对于 $1 \leq j \leq K$。

该定理给出了 p值在算法1中于原假设和已知边缘分布下所使用的渐近依据；见算法中的5a)。在备择假设下，由于$\hat{}D_K \xrightarrow{P} D_K > 0$，我们得到 $\frac{k_n}{2}\hat{}D_K \xrightarrow{P} \infty$。因此， $p-$值由算法1的5a)得到，并在 $n \to \infty$时收敛到1。这表明KL散度检验的渐近功效。

—

2.4 未知边缘分布下的理论保证

~ ~本节考虑 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的边缘分布未知的情形。如（9）中所述，首先使用各边缘的经验分布，将数据归一化到共同的近似帕累托边缘尺度。然后，对变换后的伪观测应用我们的 KL散度检验。

使用伪观测会影响概率估计量 $\hat{}p_j$ 和 $\hat{}q_j$的渐近行为，进而影响 KL散度统计量的渐近行为。一般地假设估计量 $\hat{}p_j$ 和 $\hat{}q_j$ 渐近服从 $\hat{}$ 正态分布，下面的定理给出了检验统计量 D_K 在原假设和备择假设下的渐近行为。虽然其极限分布不像上一节那样具有显式形式，但渐近理论为采用子样本自助法获取检验的临界值提供了理论依据；详情请参见第3节。

定理 3。 对于适当的 k_n，假设当 $n \to \infty$ 时，

$$\sqrt{k_n}(\hat{p}_j - p_j) = N_j^X + o_P(1) \quad and \quad \sqrt{k_n}(\hat{q}_j - q_j) = N_j^Y + o_P(1),$$

对所有 $1 \le j \le K$，其中 $(N_1^X, \ldots, N_K^X)$ 和 $(N_1^Y, \ldots, N_K^Y)$ 是两个相互独立的零均值正态随机向量，其协方差矩阵分别为 $\boldsymbol{\Sigma}^X$ 和 $\boldsymbol{\Sigma}^Y$。

在原假设 H_0 于(4)下，库尔贝克-莱布勒散度依分-

分布

$$\frac{k_n}{2}\hat{D}_K \xrightarrow{d} \sum_{j=1}^{K} \frac{1}{2p_j}(N_j^X - N_j^Y)^2, \quad n \to \infty. \tag{13}$$

在备择假设 $p_j \neq q_j$ 对某个 $1 \leq j \leq K$成立时，记 D_K 为(5) 中取 p_j时的总体版本，q_j。于是有

$$\sqrt{k_n}\left(\hat{D}_K - D_K\right) \xrightarrow{d} \sum_{j=1}^{K} \left(D_{1j}N_j^X + D_{2j}N_j^Y\right), \quad n \to \infty,$$

其中 D_{ij} 的定义与定理2 中 $i = 1$的定义相同，2 且 $1 \leq j \leq K$。

定理3 中的条件具有一般性，原假设和备择假设下的极限分布都仍未明确，且取决于未知的极值依赖参数。特别是，即使在原假设下的分布中，其渐近极限也不再像边缘分布已知时那样服从简单的卡方分布。因此，在第3.1节中，我们提出一种自助法程序，以获得原假设下检验的临界值。

下面我们验证，对于前几节中的三个例子，在一定的正则条件下，基于（9）中的伪观测构造的经验估计量 $\hat{p}_j$ 和 $\hat{q}_j$ 满足上述定理的所有假设。

例 4. 考虑风险泛函 $r(\boldsymbol{x}) = \max(x_1, \ldots, x_d)$，以及由集合 $A_I = \{x \in \mathbb{R}^d : x_j > 1, j \in I; x_j \leq 1, j \in I^c\}$ 构成的划分，其中任意 $I \subset \{1, \ldots, d\}$ 均非空；参见例1。定理 3 中关于经验估计量 $\hat{p}_j$ 和 $\hat{q}_j$ 渐近正态性的假设是否成立，与尾部依赖函数的估计有关；例如参见 *de Haan* 和 *Pereira*（2006）的第 7 章。我们还在附录中为 $d = 2$ 提供了一个具体示例。

例 5. 考虑风险泛函 $r(\boldsymbol{x}) = \min(x_1, \ldots, x_d)$ 以及由集合 $A_I = \{x \in \mathbb{R}^d : x_j > 2, j \in I; 1 < x_j \leq 2, j \in I^c\}$ 构成的划分，适用于任意非空

$I \subset \{1, \ldots, d\}$；见例 2。那么，定理 3 中关于经验估计量渐近正态性的假设是否成立，也与尾部依赖函数的估计有关。经验估计量 $\hat{}p_j$ 和 $\hat{}q_j$ 。

例 6. 考虑例 3 中 $d = 2$ 维的一个具体情形，其中 r 是 L_2范数 $r(\boldsymbol{x}) = \sqrt{x_1^2 + x_2^2}$，并配有角分量 $\theta(\boldsymbol{x}) = \arctan(x_2/x_1)$。令 $0 = \theta_0 < \theta_1 < \ldots < \theta_K = \pi/2$ 为一列角度。我们根据这些角度将单位球面S_r 划分为 K 个区域，由此得到集合$\Omega_r = \{\boldsymbol{x} \geq 0$ 的一个划分： $x_1^2 + x_2^2 > 1\}$

$$A_j = \{\boldsymbol{x} = (x_1, x_2) : r(\boldsymbol{x}) > 1, \theta(\boldsymbol{x}) \in (\theta_{j-1}, \theta_j]\}, \quad j = 1, \ldots, K,$$

它在假设1的意义下是可接受的。我们可以验证，在一些额外的温和条件下，经验估计量 $\hat{}p_j$ 和 $\hat{}q_j$ 满足定理 3 的假设；参见 *Einmahl 等人 (2001 年)*。

$\hat{}$ 一般而言， D_K 在原假设和备择假设下的渐近分布都没有简单的闭式形式。尽管如此，对于 $d = 2$ 在例 4中的特殊情形，我们推导出了这样的显式形式；详情见附录B 。特别值得注意的是，原假设下的极限服从缩放卡方分布，其缩放常数取决于未知的超越值分布 ν_X 和 ν_Y。

3 模拟

我们开展了若干模拟研究，以评估用于检验两个样本的极值依赖结构是否相同的库尔贝克-莱布勒散度检验的性能；实现细节见算法 1 。我们考察了多个方面，例如对风险函数和划分选择的敏感性、经验边际标准化的影响，以及备择假设的类型。

我们考虑两种不同的联结函数族。第一种是参数为 $\theta \in (0,1]$ (Gumbel, 1960年) 的逻辑斯蒂分布族，它属于极值联结函数类（Gudendorf 和 Segers，2010年）。其在均匀边缘分布上的分布函数为

$$F_\theta^{\log}(x_1,x_2)=\exp\left\{-\left[(-\log x_1)^{1/\theta}+(-\log x_2)^{1/\theta}\right]^\theta\right\}. \tag{14}$$

第二类是外幂 Clayton copula，它也有一个参数 $\theta \in (0,1]$, 和分布函数

$$F_\theta^{\text{clay}}(x_1,x_2)=\left\{\left[(x_1^{-1}-1)^{1/\theta}+(x_2^{-1}-1)^{1/\theta}\right]^\theta+1\right\}^{-1}.$$

可以验证，变换为标准帕累托边缘分布后，logistic Copula 和外幂 Clayton Copula 都是多元正则变差的，并且满足（6），且具有相同的扩展超越分布

$$\nu_\theta([\mathbf{0},\boldsymbol{\infty}]\setminus[\mathbf{0},\boldsymbol{x}])\propto\left(x_1^{1/\theta}+x_2^{1/\theta}\right)^\theta,\qquad \boldsymbol{x}\in[0,\infty)^2\setminus\{0\}$$

重要的是，即使两个模型的依赖参数 θ 相等，这两个 Copula$F_\theta^{\log}$和 $F_{\text{clay}\,\theta}$仍然不同；它们仅在极端尾部才会重合。

接下来，我们使用 R 软件包 `copula` （Hofert 等人，2023年）从 logistic Copula 和外幂 Clayton Copula 中生成独立样本。

3.1 自助法零分布

如第 2.4 节所述，当边缘分布未知时，检验统计量 D_K 的原假设分布没有简单的闭式表达式。因此，我们使用自助法程序来获取临界值。为此，我们无放回抽样 $n/2$ （假设 n 为偶数）个观测值 $\boldsymbol{X}_1$，……，$\boldsymbol{X}_n$。随后，计算该样本与剩余 $n/2$个样本之间的散度。我们重复这一过程 B 次，其中 B 为所需的自助样本数量。根据构造，这

得到 B 个样本 $\hat{D}^{(1)}(n/2)$，……，$\hat{D}^{(B)}(n/2)$，它们是在原假设下检验统计量 $D_K(n/2)$ 的样本。由于样本量仅为所需样本量的一半 n，我们针对这种情况进行校正

$$\hat{D}^{(i)}(n) = \hat{D}^{(i)}(n/2)/2, \quad i = 1, \ldots, B,$$

正如定理3中的收敛速率所示。

我们通过一项模拟研究来说明自助法方法的有效性，其中$\boldsymbol{X}_1$、……、 $\boldsymbol{X}_n$ 由外幂 Clayton Copula 生成，且 $n = 2000$。针对风险泛函 $r(\boldsymbol{x}) = \sqrt{x_1^2 + x_2^2}$ 计算散度，所用划分与例3 相同，由 $K = 4$ 个集合组成，超越值的数量设为 $k_n = 200$。为获得自助法零分布，我们按照上文所述对数据进行重抽样，其中$B = 1000$。为从真实零分布中获取样本，我们从相同的外幂 Clayton 分布中模拟 1000 个新数据集，并计算散度。图 3 左侧比较了自助法分布与真实零分布的直方图，并根据（12）进行了适当标准化；此处假设边缘分布已知。可以看到，两者吻合良好，也与理论卡方极限的密度（黑线）相当一致。当边缘分布未知且必须通过经验方法进行标准化时，根据（13）我们仍然知道收敛速率，但极限零分布不再与卡方分布一致。即使在这种情况下，自助样本也能很好地近似原假设下的分布。因此，可以从自助法分布中计算临界值，而不必使用卡方分位数。

3.2 检验功效

接下来，我们将考察检验在不同情形下的检验功效。我们生成 $n = 2000$个独立样本 $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ 以及 $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$，它们分别来自参数为 θ_X 和 θ_Y 的外幂 Clayton Copula ^。我们针对两个情形计算式（5）中的检验统计量 D_K

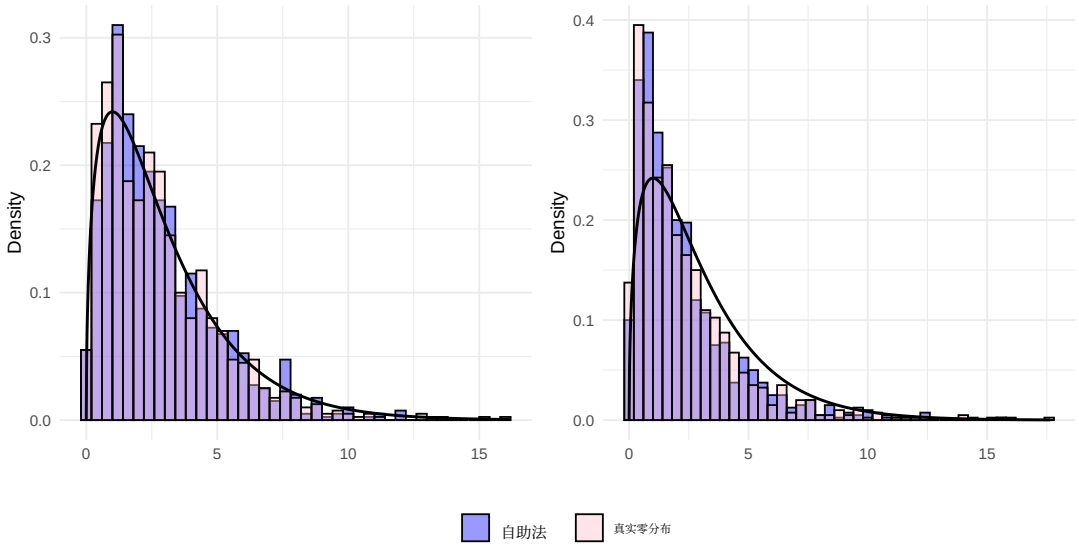


图 3：自助法分布（蓝色）和真实零分布（粉色）的直方图，以及具有 $K-1=3$ 个自由度的卡方分布的密度；左图对应边缘分布已知的情形，右图对应边缘分布未知、必须通过经验方式进行标准化的情形。

风险泛函 $r(\boldsymbol{x})$ =最大值(x_1, x_2) 和 $r(\boldsymbol{x}) = \sqrt{x_1^2 + x_2^2}$，划分如示例 1和 3 所示，分别由 $K=3$ 和 $K=5$ 集合构成；另请参见图 2。对于阈值u_n，我们选择一系列不同数量的较大次序统计量 k_n，将其用作风险泛函的超越值 r。

我们首先考虑已知边缘分布的情形，此时可以使用标准化（1）。定理 2 进而指出，检验统计量的渐近分布为 $\chi^2(K-1)$，其中K 是划分中的集合数量。图 4 左上面板展示了在 $\theta_X = \theta_Y = 0.45$ 情形下进行 500 次重复的模拟研究汇总，也就是说，原假设（4）成立。蓝色实线和虚线分别表示 500 个 KL ˆ 检验统计量值的均值 D_K 以及每个 k_n对应的逐点经验 5% 和 95% 分位数。黑色虚线表示 $\chi^2(K-1)$ 分布在 95% 水平下的临界值随超越值数量 k_n 变化的情况。正如预期的那样

根据理论，约 95% 的检验统计量实现值低于临界水平。橙色线也证实了这一点，它表示拒绝 H_0 的次数所占的百分比。在 k_n 的所有取值下，拒绝率约为 5%，表明该检验达到了预期的水平。

图 4 的上方中间面板显示了与之前相同的图，但此处数据是在备择假设 $0.45 = \theta_X \neq \theta_Y = 0.55$下模拟得到的。我们可以看到，检验区分依赖结构的能力很强，因为大多数 D_K 值都高于临界值。尽管 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 的依赖结构差异不大，但橙色曲线表明拒绝概率很高，因此检验功效也很高。

还有一种更有意思的情况：我们像之前一样，从外幂 Clayton Copula 中模拟样本 $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$，但第二总体的样本 $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ 则由逻辑斯蒂分布生成（14）。具体而言，我们将两个模型的依赖参数设为相同的值 $\theta_X = \theta_Y = 0.45$。由于极值依赖结构一致，原假设（4）此时成立。在图 4 的右上面板中，我们观察到，当 k_n 较大（阈值较低 u_n）时，检验会被拒绝；而当 k_n足够小时，则不再拒绝。这表明该检验关注的是极值依赖；这也与理论相符，理论要求 $k_n/n \to 0$。

当边缘分布未知、必须按照（9）进行经验归一化时，我们重复了所有模拟。我们使用第 2.4 节所述自助法程序得到的临界值。图 4 的底行显示，结果与已知边缘分布的情况相似，证实自助法程序是有效的。

我们对最大风险 $r(\boldsymbol{x}) = \max(x_1, x_2)$ 重复所有模拟，其具有 $K = 3$个集合。附录 9 D 中的图9 显示了定性上相似的结果，但检验功效通常低于基于具有 $K = 5$ 个集合的欧几里得风险泛函所得的结果。

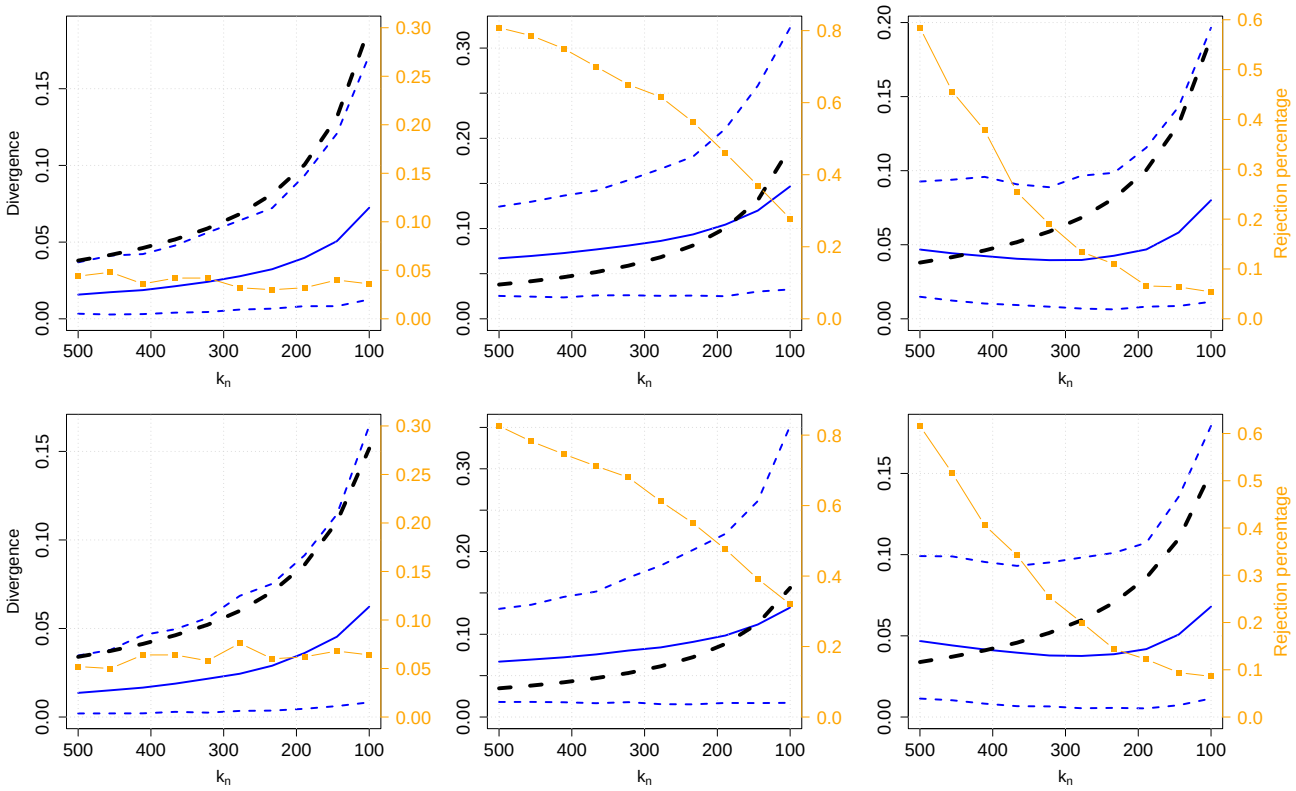


图4：基于具有 $K = 5$ 个集合的欧几里得风险泛函，在一系列超越值 k_n 下得到的 500 个样本中 KL检验统计量 D_K 的均值（蓝色实线）以及经验 5% 和 95% 分位数（蓝色虚线），并列出 95% 水平下的临界值（黑色虚线）和拒绝百分比（橙色线）。上排和底排分别显示已知边缘分布和未知边缘分布的结果。两个样本分别来自相同的分布（左）、具有不同极值依赖结构的分布（中）以及具有相同极值依赖结构的不同分布（右）。

3.3 K的作用

现在，我们以欧几里得风险泛函 $r(\boldsymbol{x}) = \sqrt{x_1^2 + x_2^2}$ 为例，研究划分中的集合数量 K 所产生的影响，其划分方式与例 3 中相同。作为进一步比较，我们还使用和风险泛函 $r(\boldsymbol{x}) = |x_1| + |x_2|$ ，沿角度进行类似划分，并采用不同的集合数量。作为基准，我们考虑例1中的最大风险泛函 $r(\boldsymbol{x}) = \max(x_1, x_2)$，其集合数量 $K = 3$为常数。首先，我们从两个具有已知边缘分布且依赖参数 $0.45 = \theta_X \neq \theta_Y = 0.55$不同的外幂 Clayton copula 中模拟 $n = 5000$ 个样本。对于 $k_n = 200$ 个超越值，图5左侧展示了不同风险泛函的拒绝概率随 K 变化的情况，结果基于 500 次重复。我们首先观察到，对于 $K = 2$ 个集合，两种角度划分都无法区分这两个分布。原因在于 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 都是对称的，因此落入上半部和下半部的点数大致相等。当 K 增加时，检验功效迅速提高，并在集合数量达到 4 或 5 个时趋于稳定。重要的是，其检验功效显著高于基于最大风险的简单检验。这表明，尽管我们的方法很简单，但仍具有足够的灵活性，能够优于更基础的统计检验。选择 K 意味着需要权衡偏差与方差。如果 K 过小，在比较 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$中的两种极值依赖结构时可能产生偏差，因为差异在少数几个集合中可能并不明显。如果 K 过大，那么每个集合中只包含少量观测值，估计散度 D_K时的变异性就会很高。

其次，使用 R 软件包 `grahpicalExtremes`(Engelke 等人，2024)，并采用 Dombry 等人 (2016) 中的精确模拟方法，我们从 $\boldsymbol{X}$ 中生成样本，该分布由 Coles and Tawn (1991) 提出；随后将其与来自外幂 Clayton Copula 的 $\boldsymbol{Y}$ 样本进行比较。我们为每个模型选择参数，使它们具有相同的极值相关 $\chi_X = \chi_Y$ (Coles 等人，1999年；Schlather

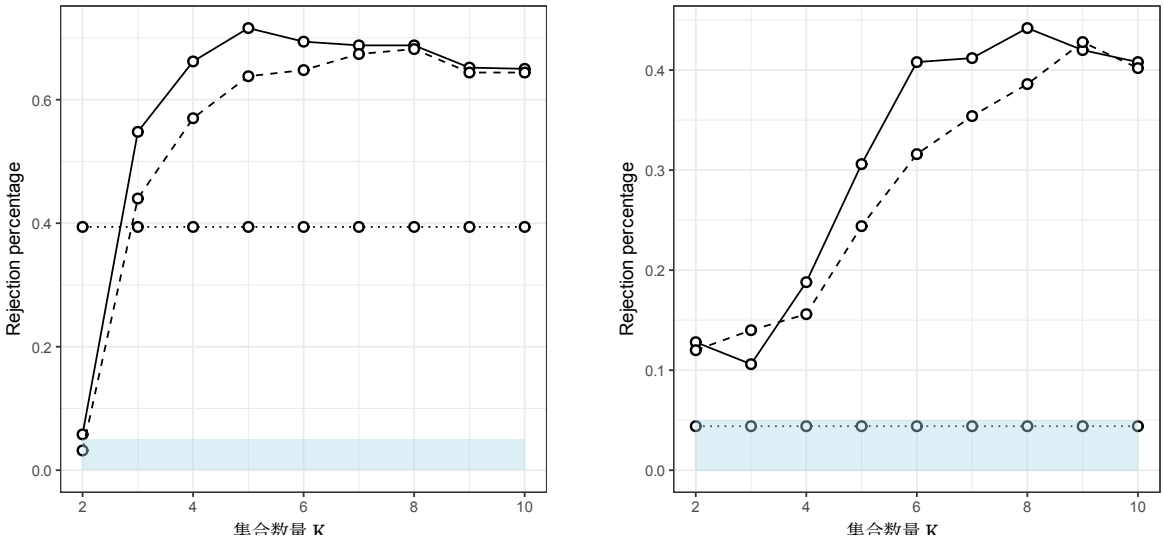


图5：对 Clayton 与 Clayton（左）以及狄利克雷与 Clayton（右）进行 500 次模拟时，H_0 的拒绝百分比，不同风险泛函的结果分别以最大风险（点线）、欧几里得范数（实线）和求和风险（虚线）表示。蓝色阴影区域表示显著性水平。

和 Tawn，2003)，这是极值依赖的一种常用汇总统计量。在这种情况下，由于极值依赖结构在对称性特征方面存在差异，H_0并不成立。对于最大风险泛函，我们可以将检验统计量 D_K 改写为极值相关的简单函数；见附录C。由于两个样本具有相同的 χ 系数，图5 右侧表明，最大风险无法区分这两种极值依赖结构（尽管它们并不相同）。两种角风险都能在集合数量较少时便检测出这种对称性差异。我们在此也观察到性能持续提升，甚至达到 $K = 8$ 个集合。

4 应用

在过去二十年里，一些国家气象服务机构依托可靠的雨量计网络，持续在全国范围内测量亚小时降水。这类

时间序列对于理解和预测山洪、城市排水问题，以及评估气候变化对极端降雨事件的影响至关重要。一个关键方面是理解降水强度、历时和频率之间的关系。从图形上看，可以针对单个地点绘制强度-历时-频率曲线（IDF曲线），这条曲线是设计水文建筑物时常用的工具（Ulrich等，2020）。统计上，强度-历时-频率曲线可以视为不同时间尺度（分钟、小时、周等）下聚合降雨边缘分布行为的概括。这些边缘分布可采用参数模型拟合，例如广义极值分布（Ulrich等，2020）或扩展帕累托分布（Haruna等，2023年）。季节周期的作用也是一个重要方面：例如，欧洲的风暴强度会随季节变化，这是由大气条件的变化以及季节间温差所致。具体而言，夏季和秋季较高的气温可能导致风暴更强，尤其表现为降雨量增加。相反，冬季风暴很可能由冷气团与来自大西洋的暖空气相互作用而形成。对于德国，Ulrich等人 （2021）研究了极端降雨的季节变化如何随不同聚合时间尺度而改变。同样，这类研究关注的是边缘分布的变化；然而，从水文学角度来看，评估季节周期是否也会影响不同聚合尺度之间的极值依赖具有重要意义。本节将说明，我们提出的KL散度提供了一种简便工具来回答这一问题。

我们研究法国气象部门 M´et´eo-France 的降雨时间序列。该部门提供一个公开数据库[1] ，其中包含 2006–2023 年期间的 6分钟降雨记录。为避免样本长度过短，我们在此研究四个季节的6分钟和逐小时降雨日最大值：冬季（DJF）、春季（MAM）、夏季（JJA）和秋季（SON）；图1 展示了波尔多市冬季（蓝色）和春季（橙色）的双变量样本。为了比较这两个时间尺度下强降雨的极值依赖，图6

[1]https://meteo.data.gouv.fr/datasets/donnees-climatologiques-de-base-6-minutes/

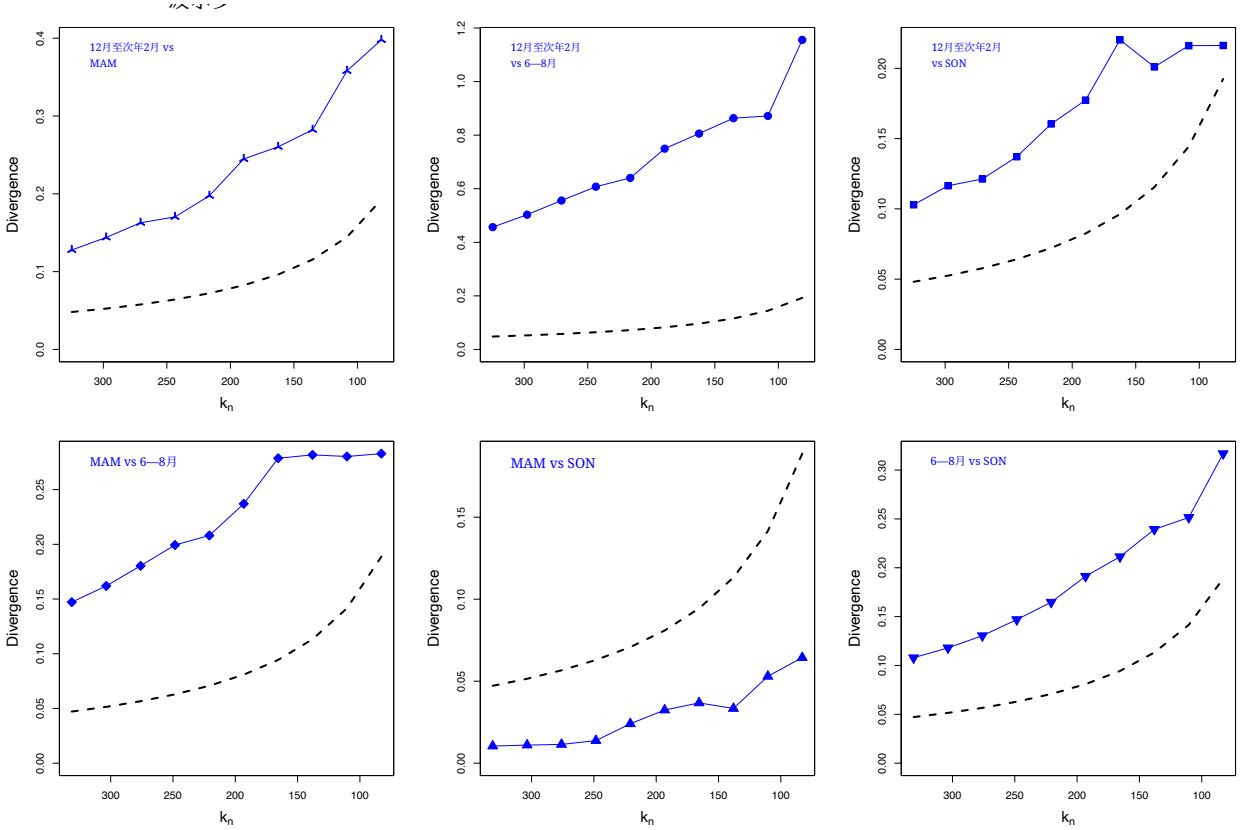


图6：对于法国波尔多市，每个面板使用基于欧几里得风险函数、具有 $K = 4$个集合的 KL 散度 D_K （蓝线），比较两个不同季节中 6分钟和每小时降雨量日最大值的极值依赖，横轴为超越值数量 k_n。95% 水平的临界值（黑色虚线）通过自助法程序获得。

ˆ 展示了我们的检验统计量 D_K 在（8）中随超越值数量 k_n 的变化情况，涵盖所有季节组合；此处我们固定 $K = 4$，但采用其他取值时结果也类似。

图1 中的极值相关性似乎表明不存在季节差异，但除春季与秋季外，我们的检验拒绝了所有季节对极值依赖相等的原假设。不同超越值数量下的结果均保持稳定，因为 D_K 始终高于 95% 水平的临界值（虚线）。

一个自然的问题是，这种季节对比是波尔多所特有的，还是在空间上具有普遍性。图 7 用橙色标记表明我们的检验是否拒绝原假设

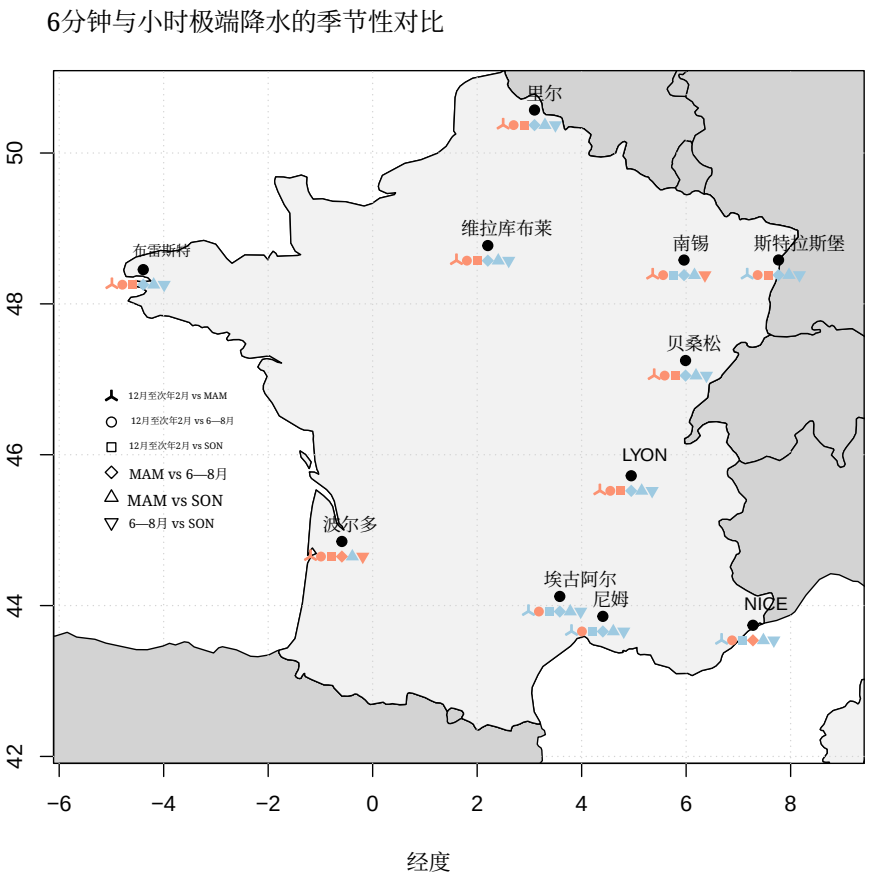


图7：对于法国的11个雨量计测站位置，橙色标记表示，基于 D_K 、包含 $K = 4$ 和 $k_n = 220$ 个超越值的KL散度检验在95%的临界水平下拒绝原假设。

（6分钟与小时降水极值之间无季节变化），采用 $K = 4$ 个集合，$k_n = 220$ 个超越值以及95%临界水平，在法国不同地点进行。总体而言，春季、夏季和秋季的表现相对相似，而冬季则明显不同。这一现象对西部、中部和北部站点尤为明显，如布雷斯特、里尔、维拉库布莱、贝桑松和波尔多。对于其他站点，冬季主要与春季形成对比。如果数量 K 划分中的集合过少，检验可能会损失检验功效。对于 $K = 2$ ，我们的检验本质上等价于基于极值相关的检验 χ。图8 表明此时差异无法被

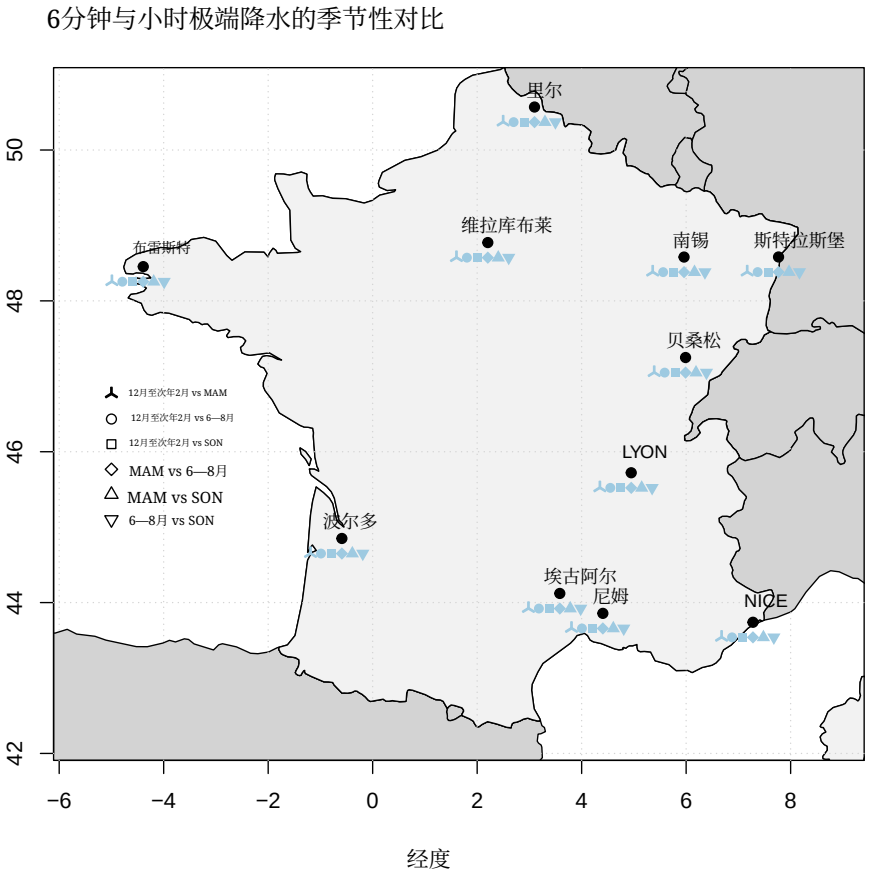


图8：与图 7 相同，但使用 $K = 2$ 代替 $K = 4$。

已通过该检验加以识别；另请参见图5右侧，以了解一种可能的解释。

从水文学视角来看，本研究强调，除了众所周知的不同聚合尺度下边缘分布的季节性效应外，不同时间尺度之间的极值依赖也可能受到季节性效应的影响。这一结果为IDF曲线建模领域开辟了新的研究方向，尤其是如何从物理机制上解释并纳入这一新认识。

参考文献

Beirlant, J.、Y. Goegebeur、J. Teugels 和 J. Segers（2005年8月）。《极值统计：理论与应用》。

B¨ucher, A. 和 H. Dette（2013）。尾部 copula 的乘子自助法及其应用。《Bernoulli》19(5A)，1655–1687。

B¨ucher, A., P. Kinsvater 和 I. Kojadinovic（2017）。检测多元极值分布中的依赖关系变化。《Extremes》20，53–89。Buishand, T. A., L. de Haan 和 C. Zhou（2008）。空间极值：在降雨问题中的应用。《Ann. Appl. Stat.》2，624–642。Coles, S., J. Heffernan 和 J. Tawn（1999）。极值分析中的相依性度量。《Extremes》2，339–365。Coles, S. G. 和 J. A. Tawn（1991）。极端多元事件建模。《J. R. Stat. Soc. Ser. B Stat. Methodol.》53，377–392。de Fondeville, R. 和 A. C. Davison（2022年9月）。函数型超阈值分析。《Journal of the Royal Statistical Society Series B》84(4)，1392–1422。de Haan, L. 和 A. Ferreira（2006）。《极值理论》。纽约：Springer。de Haan, L. 和 T. T. Pereira（2006）。空间极值：平稳情形的模型。《Ann. Statist.》34，146–168。Dombry, C., S. Engelke 和 M. Oesting（2016）。最大值稳定过程的精确模拟。《生物计量学》103，303–317。Dombry, C. 和 M. Ribatet（2015）。函数正则变差、帕累托过程与超阈值峰值。《Statistics and its Interface》8，9–17。Dudley, R.（1978）。经验测度的中心极限定理。《Annals of Probability》6(6)，899–929。Einmahl, H., L. de Haan 和 A. Sinha（1997）。极值分布谱测度的估计。《Stochastic Processes and their Applications》70(2)，143–171。Einmahl, J. H., V. I. Piterbarg 和 L. de Haan（2001）。极值分布谱测度的非参数估计。《Annals of Statistics》29(5)，1401–1423。Embrechts, P., C. Kl¨uppelberg 和 T. Mikosch（1997）。《极端事件建模：保险与金融应用》。伦敦：Springer。Engelke, S., S. A. Hitz, N. Gnecco 和 M. Hentschel（2024）。graphicalExtremes：图极值模型的统计方法。可从 https://CRAN.R-project.org/package=graphicalExtremes 获取，R 软件包版本 0.3.2。Engelke, S., A. Malinowski, Z. Kabluchko 和 M. Schlather（2015）。H¨usler–Reiss 分布与 Brown–Resnick过程的估计。《J. R. Stat. Soc. Ser. B Stat. Methodol.》77，239–265。Gudendorf, G. 和 J. Segers（2010）。极值copula。载于《Copula Theory and Its Applications》，第127–145页。Springer。Gumbel, E. J.（1960）。多维极值分布。《Publ. Inst. Statist. Paris》9，171–173。

Haruna, A., J. Blanchet, and A.-C. Favre (2023年). 对整个非零降水范围的强度-历时-频率曲线进行建模：模型比较。《水资源研究》59(6), e2022WR033362. Hershfield, D. M. (1961). 美国降雨频率图集：历时从30分钟到24小时、重现期从1年至100年。技术报告，气象局、商务部（华盛顿特区）。Hofert, M., I. Kojadinovic, M. Maechler, and J. Yan (2023年). copula：基于 Copula 的多变量依赖。R 软件包版本 1.1-3. Huser, R. and A. C. Davison (2014). 极端事件的时空建模。《英国皇家学会会刊 B 辑：统计方法论》76, 439–461. Li, Y., J. Gao, J. Yin, and S. Wu (2024). 评估中国海岸线复合极端风暴潮与降水的潜在影响。《天气与气候极端事件》45, 100702. Nasr, A. A., T. Wahl, M. M. Rashid, P. Camus, and I. D. Haigh (2021). 评估美国海岸线海洋、河流和降雨洪水驱动因素之间的依赖结构。《水文与地球系统科学》25(12), 6203–6222. Nasr, A. A., T. Wahl, M. M. Rashid, R. A. Jane, P. Camus, and I. D. Haigh (2023年). 美国本土相连部分海岸线周边复合沿海与内陆洪水驱动因素依赖关系的时间变化。《天气与气候极端事件》41, 100594. Naveau, P., A. Guillou, and T. Rietsch (2014). 一种用于检测气候极端事件变化的非参数熵方法。《英国皇家学会会刊 B 辑：统计方法论》76, 861–884. Naveau, P., A. Hannart, and A. Ribes (2020). 气候科学中极端事件归因的统计方法。《统计学及其应用年度评论》。出版商：Annual Reviews. Resnick, S. I. (2008). 《极值、正则变差与点过程》。纽约：Springer. Schlather, M. and J. A. Tawn (2003). 多变量与空间极值的依赖度量：性质与推断。《生物计量学》90, 139–156. Ulrich, J., F. S. Fauer, and H. W. Rust (2021). 对德国不同时间尺度极端降雨的季节变化进行建模。《水文与地球系统科学》25(12), 6133–6149. Ulrich, J., O. E. Jurado, M. Peter, M. Scheibel, and H. W. Rust (2020). 利用空间协变量对不同历时的强度-历时-频率曲线进行一致性估计。《水》12(11:3119). Vignotto, E., S. Engelke, and J. Zscheischler (2021). 对大不列颠和爱尔兰复合降水与风极端事件的双变量依赖关系进行聚类。《天气与气候极端事件》32, 100318. Zscheischler, J., O. Martius, S. Westra, E. Bevacqua, C. Raymond, R. Horton, B. Hurk, A. AghaKouchak, A. J´ez´equel, M. Mahecha, D. Maraun, A. Ramos, N. Ridder, W. Thiery, and E. Vignotto (2020, 06). 复合天气与气候事件的类型学。《自然综述：地球与环境》1, 1–15.

Zscheischler, J., P. Naveau, O. Martius, S. Engelke, and C. C. Raible (2020). 评估复合降水与极端风速的依赖结构。《地球系统动力学讨论》2020, 1–23.

Zscheischler, J. 和 S. I. Seneviratne（2017）。驱动因子的依赖性会影响与复合事件相关的风险。《科学进展》3。

补充材料

《多元极值的 Kullback–Leibler散度检验：理论与实践》

A 证明

命题证明*1*。取 $0 < c <$ 最小值 $\{\|\boldsymbol{x}\|_\infty : \ r(\boldsymbol{x}) > 1\}$。那么 $\Omega_r \subset \{\boldsymbol{x} : \ \|\boldsymbol{x}\|_\infty > c\}$。注意，$uB \subset u\Omega_r \subset \{\boldsymbol{x} : \ \|\boldsymbol{x}\|_\infty > uc\}$。我们有

$$\begin{aligned} P(\boldsymbol{X} \in uB | r(\boldsymbol{X}) > u) &= \frac{P(\boldsymbol{x} \in uB)}{P(\boldsymbol{x} \in u\Omega_r)} \\ &= \frac{P(\boldsymbol{x} \in uB, \|\boldsymbol{x}\|_\infty > uc)}{P(\boldsymbol{x} \in u\Omega_r, \|\boldsymbol{x}\|_\infty > uc)} \\ &= \frac{P(\boldsymbol{x} \in uB | \ \|\boldsymbol{x}\|_\infty > uc)}{P(\boldsymbol{x} \in u\Omega_r | \ \|\boldsymbol{x}\|_\infty > uc)} \end{aligned}$$

将 uc 视为 u 在（6）中的情形，并取 $u \to \infty$，可得

$$\lim_{u\to\infty} P(\boldsymbol{X} \in uB | r(\boldsymbol{X}) > u) = \frac{\nu(B/c)}{\nu(\Omega_r/c)} = \frac{\nu(B)}{\nu(\Omega_r)},$$

最后一行由 ν的齐次性得出。 □

为证明定理 2，我们首先在以下命题中建立估计量 $\hat{}p_j$ 的联合渐近行为。

命题4。 假设满足定理 *2*中的相同条件。那么，对于所有$1 \le j \le K$，联合地，当 $n \to \infty$时，

$$\sqrt{k_n}\,(\hat{p}_j - p_j) \xrightarrow{d} N_j - p_j \sum_{j=1}^{K} N_j,$$

其中， N_1、 N_2、……、 N_K 是均值为零、方差分别为 p_1、 p_2、……、 p_K 的独立正态分布随机变量。

考虑集合 $S(t,\ j)\ = tA_j$，其中 $t \in [a,b]$ 取任意固定的 $0 < a < b < +\infty$, $1 \leq j \leq K$。记 $\tilde{u}_n = U_X^{-1}\left(k_n/n\right)$。为了证明命题4，我们打算将Einmahl等人(1997年中的引理3.1应用于序列 $\{\boldsymbol{Z}_i^{(n)}\}_{i=1}^n = \{\boldsymbol{X}_i/\tilde{u}_n\}_{i=1}^n$ 和集合类 $\{S(t,\ j)\colon\ t \in [a,b],\ 1 \leq j \leq K\}$。为了检验其中的条件，我们先证明下面这个预备引理。

引理5。 假设假设条件 *1* 成立。那么，集合类 $\{S(t,j)\colon t \in [a,b], 1 \leq j \leq K\}$ 构成一个有限维*VC*类。此外，它们满足*Gaenssler*（*1983*年，第*108*页）中的条件*(SE)*

引理 *5*的证明。 根据假设条件 1，必定存在有限个博雷尔集，属于 $\mathcal{A}_r$，记为 $\{B_1, \ldots, B_L\}$ 使得对于所有 $j = 1,\ 2,\ \ldots,\ K,\ A_j$ 都可以通过对 $\{tB_l$ 中的集合进行有限次并集、交集和补运算来表示：$t \in [a,b],\ 1 \leq l \leq L\}$。不失一般性，我们假设 $B_l \subset \Omega_r$ 对于所有$l = 1,\ 2,\ \ldots,\ L$。

首先，我们证明，对于每个给定的 l，集合类 $\{tB_l\colon\ t \in [a,b]\}$ 是一个维度为3的VC类。对于 $a\Omega_r$中的任意3个点，如果至少有两个点落在aB_l之外，那么它们构成的子集就无法被该类打散。如果落在aB_l之外的点至多有一个，即至少有两个点 x_1 和 x_2 位于集合 aB_l内。由于 $tB_l \subset sB_l$ 对于任意$t > s$都成立，对于每个点 x_i，存在唯一的值 t_i，使得对于所有 $t < t_i$，$x_i \in tB_l$，并且对于所有 $t > t_i$，$x_i \in (tB_l)^c$。考虑 $t_1 < t_2$、$t_1 = t_2$ 和 $t_1 > t_2$这三种情况，可以直接得出，在每种情况下，都无法利用 tB_l（其中 $t \in [a,b]$）打散由两个点 x_1 和 x_2 构成的所有子集。

其次，由于将有限个有限维VC类相结合会得到一个有限维VC类，因此我们得到集合类 $\{tB_l\colon\ t \in [a,b],\ 1 \leq l \leq L\}$构成一个有限维VC类

最后，由于 $S(t,\ j)$ 中任意一个满足 $t \in [a,b],\ 1 \leq j \leq K$ 的集合都可以表示为一个

对$\{tB_l:\ t\in[a,b],\ 1\le l\le L\}$中的集合进行有限次并集、交集和补运算，类 $\{S(t,\ j):\ t\in[a,b],\ 1\le j\le K\}$ 也构成一个有限维 VC 类；参见Dudley (1978年)中的命题7.12。

关于条件(SE)的引理后半部分，只需考虑所有有理数 $t\in[a,b]$即可显而易见。

□

命题的证明 4。 回忆 $\tilde{u}_n=U_X^{-1}(k_n/n)$，这意味着 $P(\|\boldsymbol{X}\|_\infty>\tilde{u}_n)=\frac{k_n}{n}$且 $\tilde{u}_n\to\infty$ 当 $n\to\infty$时。进一步回忆 $\{\boldsymbol{Z}_i^{(n)}\}_{i=1}^n=\{\boldsymbol{X}_i/\tilde{u}_n\}_{i=1}^n$ 且 $S(t,\ j)\ =tA_j$。将二阶条件(10)应用于 A_j，可得当 $n\to\infty$时，

$$\frac{n}{k_n}\mathbb{P}(\boldsymbol{Z}_i^{(n)}\in S(t,j))=\frac{\mathbb{P}(\boldsymbol{X}_i\in\tilde{u}_ntA_j)}{P(\|\boldsymbol{X}\|_\infty>\tilde{u}_n)}\to\nu_X(tA_j)=t^{-1}\nu_X(A_j),\tag{15}$$

对所有 $t\in[a,b]$和 $j=1,\ 2,\ \ldots,\ K$均成立。这验证了Einmahl等人（1997年）中的引理3.1的条件(3.1)，其极限测度为 $\mu=\nu_X$。事实上，我们有一个更精确的极限关系：当 $n\to\infty$时，

$$\frac{n}{k_n}\mathbb{P}(\boldsymbol{Z}_i^{(n)}\in S(t,j))-t^{-1}\nu_X(A_j)=O(a_X(\tilde{u}_n)),\tag{16}$$

对所有 $t\in[a,b]$和 $j=1,\ 2,\ \ldots\ldots,\ K$，均一致成立。

在核查了上述引理 5 中的其他条件后，我们可以将Einmahl等人 (1997年中的引理3.1应用于序列 $\{\boldsymbol{Z}_i^{(n)}\}_{i=1}^n=\{\boldsymbol{X}_i/\tilde{u}_n\}_{i=1}^n$ 和集合类 $\mathcal{B}:=\{S(t,\ j):\ t\in[a,b],\ 1\le j\le K\}$ 并得到如下结果。在适当的 Skorokhod 构造下，存在一列定义在 W_n 上的有界、一致连续、零均值高斯过程 $S\in\mathcal{B}$，使得对任意 $S_1,\ S_2\in\mathcal{B}$

$$\mathrm{Cov}(W_n(S_1),W_n(S_2))=\mu(S_1\bigcap S_2),$$

且作为 $n\to\infty$,

$$\left|\sqrt{k_n}\left(\frac{1}{k_n}\sum_{i=1}^n 1_{\{\boldsymbol{X}_i\in t\tilde{u}_nA_j\}}-\frac{n}{k_n}\mathbb{P}(\boldsymbol{X}_i\in t\tilde{u}_nA_j)\right)-W_n(tA_j)\right|\xrightarrow{a.s.}0,$$

对于 $t \in [a,b]$ 和 $j = 1,\ 2,\ \ldots\ldots,\ l$，均一致。

由（16）以及条件 $\sqrt{k_n}a_X\left(\tilde{u}_n\right) \to 0$ 当 $n \to \infty$时，我们可以将确定性项 $\frac{n}{k_n}\mathbb{P}\left(\boldsymbol{X}_i \in t\tilde{u}_n A_j\right)$ 替换为其极限，从而得到当 $n \to \infty$时，

$$\left|\sqrt{k_n}\left(\frac{1}{k_n}\sum_{i=1}^{n}1_{\{\boldsymbol{X}_i \in t\tilde{u}_n A_j\}} - t^{-1}\nu_X(A_j)\right) - W_n(tA_j)\right| \xrightarrow{a.s.} 0,$$

对 $t \in [a,b]$ 和 $j = 1,\ 2,\ \ldots\ldots,\ l$一致成立。将两边除以 $\nu_X(\Omega_r)$，并将$\tilde{}$定义为高斯过程的级数 $W_n = W_n/\nu_X(\Omega_r)$，可得

$$\left|\sqrt{k_n}\left(\frac{1}{k_n\nu_X(\Omega_r)}\sum_{i=1}^{n}1_{\{\boldsymbol{X}_i \in t\tilde{u}_n A_j\}} - t^{-1}p_j\right) - \tilde{W}_n(tA_j)\right| \xrightarrow{a.s.} 0, \tag{17}$$

我们打算将（ t 式（17）中的）替换为 $t_n = u_n/\tilde{u}_n = R^{(X)}_{n-k_n,\ n}/\tilde{u}_n$。为此，我们首先需要推导 t_n的渐近行为。将极限关系（17）对所有$1 \le j \le K$ 相加，并注意到 $\{A_j\}_{j=1}^{K}$ 构成 Ω 的一个划分$_r$ ，可得当 $n \to \infty$时，对于$t \in [a,b]$

$$\sqrt{k_n}\left(\frac{1}{k_n\nu_X(\Omega_r)}\sum_{i=1}^{n}1_{\{r(\boldsymbol{X}_i)>t\tilde{u}_n\}} - t^{-1}\right) - \tilde{W}_n(t\Omega_r) \to 0 \quad \text{a.s.}$$

将 Vervaat 引理应用于这一极限关系，可得当 $n \to \infty$时，对于$v \in [b^{-1}, a^{-1}]$,

$$\sqrt{k_n}\left(\frac{R^{(X)}_{n-[k_n\nu_X(\Omega_r)v],n}}{\tilde{u}_n} - v^{-1}\right) - v^{-2}\tilde{W}_n\left(v^{-1}\Omega_r\right) \to 0 \quad \text{a.s.}$$

适当选择 a和 b，使得 $1/\nu_X(\Omega_r) \in [b^{-1}, a^{-1}]$，我们可以取 $v = 1/\nu_X(\Omega_r)$，并得到当 $n \to \infty$时，

$$\sqrt{k_n}\left(\frac{R^{(X)}_{n-k_n,n}}{\tilde{u}_n} - \nu_X(\Omega_r)\right) - (\nu_X(\Omega_r))^2\tilde{W}_n(\nu_X(\Omega_r)\cdot\Omega_r) \to 0 \quad \text{a.s.}$$

由于 $\nu_X(\Omega_r) \in [a,b]$，我们可以进行计划中的替换。将 t 替换为$t_n = u_n/\tilde{u}_n$ 在（17）中，可得当 $n \to \infty$时，

$$\sqrt{k_n}\left(\frac{1}{\nu_X(\Omega_r)}\hat{p}_j - \left(\frac{u_n}{\tilde{u}_n}\right)^{-1}p_j\right) - \tilde{W}_n(\nu_X(\Omega_r)\cdot A_j) \to 0 \quad \text{a.s.}$$

再结合对 $u_n/\tilde{u}_n$ 的渐近行为应用德尔塔方法，可得当 $n \to \infty$ 时，

$$\sqrt{k_n}\left(\frac{\hat{p}_j}{\nu_X(\Omega_r)} - \frac{p_j}{\nu_X(\Omega_r)}\right) - \left(\tilde{W}_n(\nu_X(\Omega_r)\cdot A_j) - p_j\tilde{W}_n(\nu_X(\Omega_r)\cdot\Omega_r)\right) \to 0 \quad \text{a.s.}$$

˜对于所有 $j = 1$、2、.....、K。回顾记号 $W_n = W_n/\nu_X(\Omega_r)$。因此我们得出，当 $n \to \infty$

$$\sqrt{k_n}(\hat{p}_j - p_j) \xrightarrow{a.s.} W_n(\nu_X(\Omega_r)\cdot A_j) - p_j W_n(\nu_X(\Omega_r)\cdot\Omega_r), \tag{18}$$

对于所有 $j = 1$、2、.....、K。记 $N_j := W_n(\nu_X(\Omega_r)\cdot A_j)$。显然，$N_j$ 是相互独立的正态分布随机变量，均值为零，且 $W_n(\nu_X(\Omega_r)\cdot\Omega_r) = \sum_{j=1}^{K} N_j$。

证明的剩余部分是计算每个 N_j 的方差，其中 j =1、2、.....、K：

$$\mathbb{V}(N_j) = \nu_X(\nu_X(\Omega_r)\cdot A_j) = \frac{\nu_X(A_j)}{\nu_X(\Omega_r)} = p_j.$$

命题得证。 □

现在我们可以着手证明主要定理。

定理 2 （及定理 3）的证明。 命题 4 表明，当 $n \to \infty$,

$$\sqrt{k_n}\,(\hat{p}_j - p_j) \xrightarrow{d} N_j^X := N_j - p_j\sum_{j=1}^{K} N_j,$$

其中 $N_j^X \sim N(0, p_j(1-p_j))$，且

$$\mathrm{Cov}(N_{j_1}^X, N_{j_2}^X) = -p_{j_1}p_{j_2},$$

对于 $1 \le j_1 < j_2 \le K$。

类似地，基于观测值 $\boldsymbol{Y}_1$，$\boldsymbol{Y}_2$，.....，$\boldsymbol{Y}_n$，我们构造条件概率估计量 ˆq_j。它们具有以下渐近性质：当 $n \to \infty$时，对所有$1 \le j \le K$ 联合成立

$$\sqrt{k_n}\,(\hat{q}_j - q_j) \xrightarrow{d} N_j^Y,$$

其中 $N_j^Y \sim N(0,\ q_j(1-q_j))$，且

$$\mathrm{Cov}(N_{j_1}^Y, N_{j_2}^Y) = -q_{j_1} q_{j_2},$$

对于 $1 \le j_1 < j_2 \le K$。由于从 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 中抽取的观测值相互独立，因此 $(N_1^X, N_2^X, \ldots, N_K^X)$ 和 $(N_1^Y, N_2^Y, \ldots, N_K^Y)$ 相互独立。

回顾式 (5中定义的多项Kullback–Leibler散度。我们推导其总体版本 D_K关于各个维度的偏导数如下：

$$\begin{aligned}
D_{1,j} &= \frac{\partial D_K}{\partial p_j} = \log\frac{p_j}{q_j} + 1 - \frac{q_j}{p_j},\\
D_{2,j} &= \frac{\partial D_K}{\partial q_j} = \log\frac{q_j}{p_j} + 1 - \frac{p_j}{q_j},\\
D_{11,j} &= \frac{\partial^2 D_K}{\partial p_j^2} = \frac{1}{p_j} + \frac{q_j}{p_j^2},\\
D_{22,j} &= \frac{\partial^2 D_K}{\partial q_j^2} = \frac{1}{q_j} + \frac{p_j}{q_j^2},\\
D_{12,j} &= \frac{\partial^2 D_K}{\partial p_j \partial q_j} = -\frac{1}{p_j} - \frac{1}{q_j}.
\end{aligned}$$

请注意，除上述列出的二阶偏导数外，所有其他二阶偏导数均为零。

通过对 D_K 进行泰勒展开，我们可以得到，当 $n \to \infty$ 时，

$$\begin{aligned}
\hat{D}_K = D_K &+ \sum_{j=1}^{K} \left(D_{1,j}(\hat{p}_j - p_j) + D_{2,j}(\hat{q}_j - q_j)\right)\\
&+ \frac{1}{2}\sum_{j=1}^{K} \left(D_{11,j}(\hat{p}_j - p_j)^2 + D_{22,j}(\hat{q}_j - q_j)^2 + 2D_{12,j}(\hat{p}_j - p_j)(\hat{q}_j - q_j)\right)\\
&+ o_P\left(\sum_{j=1}^{K}\left((\hat{p}_j - p_j)^2 + (\hat{q}_j - q_j)^2\right)\right).
\end{aligned}$$

若对所有 $1 \le j \le K$ 都有 $p_j = q_j$，则对所有 $1 \le j \le K$ 都有 $D_K = 0$，$D_{1,j} = D_{2,j} = 0$。因此，只剩下二阶项。此外$D_{11,j} = D_{22,j} = \frac{2}{p_j} = \frac{2}{q_j}$ 和

$D_{12,\ j} = -\frac{2}{p_j} = -\frac{2}{q_j}$。因此，由于 $n \to \infty$,

$$\begin{aligned}\frac{k_n}{2}\hat{D}_K &= \frac{k_n}{2}\sum_{j=1}^{K}\frac{1}{p_j}\left((\hat{p}_j - p_j)^2 + (\hat{q}_j - q_j)^2 - 2(\hat{p}_j - p_j)(\hat{q}_j - q_j)\right) + o_P(1)\\ &= \frac{k_n}{2}\sum_{j=1}^{K}\frac{1}{p_j}(\hat{p}_j - \hat{q}_j)^2 + o_P(1)\\ &\xrightarrow{d}\sum_{j=1}^{K}\frac{1}{2p_j}(N_j^X - N_j^Y)^2\end{aligned}$$

注意，　这正是定理 3 在原假设下的通用结果　.

在定理 2的设定下，我们可以进一步推导出如下极限分布。记 $\mathbf{L} = (L_1,\ L_2,\ \ldots\ldots,\ L_{K-1})^T$，其中 $L_j = N_j^X - N_j^Y$对 $j = 1$、2、……、K成立。于是有$\mathbf{L} \sim N(\mathbf{0},\ \Sigma)$，其中 $\Sigma = (\sigma_{j_1,\ j_2})_{1\le j_1,\ j_2\le K-1}$，且 $\sigma_{j,\ j} = 2p_j(1 - p_j)$以及 $\sigma_{j_1,\ j_2} = -2p_{j_1}p_{j_2}$对 $j_1 \neq j_2$成立。注意到 $\sum_{j=1}^{K} L_j = 0$，我们可以改写为

$$\sum_{j=1}^{K}\frac{1}{2p_j}(N_j^X - N_j^Y)^2 = \sum_{j=1}^{K-1}\frac{1}{2p_j}L_j^2 + \frac{1}{2p_K}\left(\sum_{j=1}^{K-1}L_j\right)^2 = \mathbf{L}^T\Sigma^{-1}\mathbf{L} \sim \chi^2(K-1).$$

在最后一步中，我们利用了以下事实：$\Sigma^{-1} = (\tilde{\sigma}_{j_1,\ j_2})_{1\le j_1,\ j_2\le K-1}$，其中 $\tilde{\sigma}_{j,\ j} = \frac{1}{2}\left(\frac{1}{p_j} + \frac{1}{p_K}\right)$且 $\tilde{\sigma}_{j_1,\ j_2} = \frac{1}{2p_K}$ 对 $j_1 \neq j_2$成立。可以验证，$\Sigma\Sigma^{-1} = \mathbf{I}_{K-1}$。

若对于某个 $1 \le j \le K$，$p_j \neq q_j$，则 $D_K > 0$。此外，$D_{1,\ j} \neq 0$ 且 $D_{2,\ j} \neq 0$ 当且仅当 $p_j \neq q_j$。因此，我们可以忽略二阶项，并得到当 $n \to \infty$ 时，

$$\sqrt{k_n}\left(\hat{D}_K - D_K\right) \xrightarrow{d}\sum_{j=1}^{K}\left(D_{1,j}N_j^X + D_{2,j}N_j^Y\right) \sim N(0, \sigma^2).$$

一般结果与定理3相同。在定理 2的设定下，还可以进一步计算 σ^2，最终得到其中所示的结果。

□

B 特例：最大风险泛函

最后，我们展示一个具体例子，其中 $d = 2$如例 4 所示，定理 3 中的条件成立；特别地，我们可以推导出原假设和备择假设下的极限分布

。记作 $A(w,\ v) := \left\{\boldsymbol{x}:\ x^{(1)} > w,\ x^{(2)} > v\right\}$ 对所有 $w,\ v \in [0, +\infty)$。该例实际上是区域 $\Omega_r := \{x:\ r(\boldsymbol{x}) > 1\}$ 的一种具体划分，

$$A_1 := A(1,1) = \{\boldsymbol{x} : \min(x_1, x_2) > 1\},$$

$$A_2 := A(0,1) \setminus A(1,1) = \{\boldsymbol{x} : 0 < x_1 \leq 1, x_2 > 1\},$$

$$A_3 := A(1,0) \setminus A(1,1) = \{\boldsymbol{x} : 0 < x_2 \leq 1, x_1 > 1\}$$

我们证明定理3 中的一般条件成立。我们进一步证明，在原假设下，极限分布为缩放卡方分布，并推导备择假设下极限正态分布的方差。

类似于（2），我们使用变换后的观测值，为这些集合及其 $\hat{}\ p_j,\ q_j$ 定义 $j = 1$, 2，3 以及 $\hat{}p_j$和 $\hat{}q_j$。回顾一下， $\boldsymbol{X}$ 的边缘分布服从标准帕累托分布。对于任意 6，我们将条件（ $A(w,\ v)$）应用于集合 $w,\ v \geq 1$ 并得到

$$\begin{aligned}\lim_{u\to\infty} u\mathbb{P}(\boldsymbol{X}/u \in A(w,v)) &= \lim_{u\to\infty} \mathbb{P}(\boldsymbol{X}/u \in A(w,v) |\, \|\boldsymbol{X}\|_\infty > u) \cdot uP(\|\boldsymbol{X}\|_\infty > u) \\ &= \frac{\nu_X(A(w,v))}{\nu_X(A(0,1))} =: \Lambda\left(\frac{1}{w}, \frac{1}{v}\right). \end{aligned} \tag{19}$$

我们注意到，对于所有 $\Lambda(1/w,\ 1/v) > 0$，$w,\ v \geq 1$。我们可以将 Λ 的定义扩展到集合 $\bar{}\mathbf{R}_+^2 = [0, +\infty]^2 \setminus \{(+\infty,\ +\infty)\}$。关于 Λ的偏导数，我们作出如下假设，见 B¨ucher 和 Dette (2013)。

假设（D）：假设 Λ 的偏导数存在，分别记为 Λ_1 和Λ_2。此外， Λ_l 在集合 $\{\boldsymbol{x}:\ 0 < x_l < +\infty\}$ 上连续，其中 $l = 1$,2。

最后但同样重要的是，我们假设10）所述的二阶条件对 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$成立。证明 p_j 和q_j 的估计量的渐近行为需要这一假设。以下推论给出了检验统计量的渐近理论。

推论 6。 假设条件 (10) 对随机向量成立 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$，其对应的 ν_X和 ν_Y二阶尺度函数分别为 a_X 和 a_Y。此外，假设假设（D）成立。

令 k_n 为一个序列，使得当 $n \to \infty$, $k_n \to \infty$, $k_n/n \to 0$, $\sqrt{k_n} a_X(n/k_n) \to 0$ 且 $\sqrt{k_n} a_Y(n/k_n) \to 0$时。

在原假设 $p_j = q_j$ 对所有 $j = 1,2,3$ 均成立的条件下，库尔贝克-莱布勒散度收敛如下：当 $n \to \infty$ 时，

$$\frac{k_n}{2} \hat{D}_K \xrightarrow{d} \left(\sum_{j=1}^{3} \frac{1}{p_j} - 1 \right) \frac{(\Lambda - 1)(\Lambda + 2\Lambda_1 \Lambda_2)}{\Lambda} \chi^2(1),$$

其中，$\Lambda = \Lambda(\mathbf{1}) = \nu_X(\Omega_r)$，$\Lambda_j = \Lambda_j(\mathbf{1})$ 对于 $j = 1, 2$ 是函数 $\Lambda(\cdot)$在点 $\mathbf{1}$处的偏导数。

在备择假设 $p_j \neq q_j$ 对某个 $j = 1,2,3$成立时，记 D 为 (5) 中带有 p_j, q_j的总体版本。那么，当 $n \to \infty$时，

$$\sqrt{k_n} \left(\hat{D}_K - D_K \right) \xrightarrow{d} N(0, \sigma^2),$$

其中

$$\begin{aligned} \sigma^2 = & [D_{1,1}(1+p_1) + D_{1,2}(-1+p_2) + D_{1,3}(-1+p_3)]^2 \frac{(\Lambda^X - 1)(\Lambda^X + 2\Lambda_1^X \Lambda_2^X)}{\Lambda^X} \\ & + [D_{1,1}(1+p_1) + D_{1,2}(-1+p_2) + D_{1,3}(-1+p_3)]^2 \frac{(\Lambda^Y - 1)(\Lambda^Y + 2\Lambda_1^Y \Lambda_2^Y)}{\Lambda^Y}, \end{aligned}$$

其中 $D_{i,j}$ 根据定理 *2*，针对 $i = 1, 2$ 和 $j = 1, 2, 3$ 定义。这里，基于 $\tilde{X}$ 和 $\tilde{Y}$的两个 Λ 函数分别记为 $\Lambda^X(\cdot)$ 和 $\Lambda^Y(\cdot)$。$\Lambda^X = \Lambda^X(1) = \nu_X(\Omega_r), \Lambda_j^X = \Lambda_j^X(1)$ 对于 $j = 1, 2$ 是函数 $\Lambda^X(\cdot)$ 在点 *1* 处的偏导数。Λ^Y, $\Lambda_j^Y (j = 1, 2)$ 根据函数 $\Lambda^Y(\cdot)$ 类似地定义。

根据该推论，我们发现，即使在原假设下，边缘变换也可能影响库尔贝克-莱布勒散度估计量的渐近行为。因此，由定理 χ^2 中的分布2导出的临界值可能并不有效。实际的渐近分布可能取决于尾部依赖结构。尽管如此，由于我们仍能在原假设和备择假设下以相同的收敛速度推导出渐近分布，因此可以在检验中使用自助法来导出临界值。

为证明该推论，我们首先指出，式（ Λ 19）中定义的量与 Λ_L 在B¨ucher 和 Dette (2013中的量一致，只要我们考虑随机向量 $\boldsymbol{U_X}$ ：$=(1/X^{(1)},\ 1/X^{(2)})$ 。注意，$\boldsymbol{U_X}$ 服从一个 Copula C_X ，并且

$$\lim_{t\to\infty} tC\left(\frac{1}{wt},\frac{1}{vt}\right)=\lim_{t\to\infty} t\mathbb{P}\left(1/X^{(1)}\le\frac{1}{wt},1/X^{(2)}\le\frac{1}{vt}\right)$$
$$=\lim_{t\to\infty} t\mathbb{P}\left(\boldsymbol{X}\in tA(w,v)\right)=\Lambda\left(\frac{1}{w},\frac{1}{v}\right).$$

接下来，我们验证 B¨ucher 和 Dette (2013) 中定理 2.2 所要求的所有条件均成立。首先，回顾一下，对于 $\boldsymbol{X}$ ，二阶条件 (10) 在二阶尺度函数 $a_X(u)$ 下成立。这意味着 (19) 中的极限关系具有由函数 $a_X(u)$ 决定的相同收敛速度。这等价于 B¨ucher 和 Dette (2013) 中的 (2.7)，其中 $A(t)=a_X(t)$。其次，中间序列 k_n 满足，当 $n\to\infty$ 时，$\sqrt{k_n}a_X($ $n/k_n)\to 0$。最后，我们的假设（D）与该文中的 (2.14) 相同。

—

应用 B¨ucher 和 Dette (2013) 中的定理 2.2 后，我们立即得到以下引理，它是证明推论 6 的主要工具。

引理 7. 回顾一下，$A(w,v)=\{\boldsymbol{x}:\ x^{(1)}>w,x^{(2)}>v\}$。定义

$$\hat{\Lambda}(x^{(1)},x^{(2)})=\frac{1}{k_n}\sum_{i=1}^{n}\mathbf{1}\left\{\hat{\boldsymbol{X}}_i\in\frac{n}{k_n}A\left(\frac{1}{x^{(1)}},\frac{1}{x^{(2)}}\right)\right\},$$

对于任意 $\boldsymbol{x}=(x^{(1)},x^{(2)})\in\bar{\mathbf{R}}_+^2=[0,+\infty]^2\setminus\{(+\infty,+\infty)\}$。于是，当 $n\to\infty$ 时，

$$\sqrt{k_n}(\hat{\Lambda}(\boldsymbol{x})-\Lambda(\boldsymbol{x}))\rightsquigarrow\mathbb{G}(\boldsymbol{x})-\Lambda_1(\boldsymbol{x})1\mathbb{G}(x^{(1)},+\infty)-\Lambda_2(\boldsymbol{x})\mathbb{G}(+\infty,x^{(2)}),$$

其中 Λ_1 和 Λ_2 是 Λ的两个偏导数，$\mathbb{G}$是均值函数为零、协方差结构为 $\mathbb{E}\mathbb{G}(\boldsymbol{x})\mathbb{G}(\boldsymbol{y})$ $=\Lambda(\boldsymbol{x}\wedge\boldsymbol{y})$的高斯随机场。这里的收敛发生在空间 $\mathcal{B}_\infty(\bar{\mathbf{R}}_+^2)$ 中，且该空间配备了良好的范数。关于该空间和范数的详细信息，请参见 *B¨ucher 和 Dette(2013)*。

ˆ推论*6*的证明。定义 $I_1(t)=\Lambda(1/t,1/t)$，对 $t>0$.....。我们指出，其理论对应量为 $\Lambda(1/t,1/t)=\nu_X(A(t,t))=\frac{1}{t}\nu_X(A_1)$ 。应用引理 7，其中将 $\boldsymbol{x}$替换

由 $(1/t, 1/t)$，并得到如下结果：当 $n \to \infty$ 时，

$$\begin{aligned}
&\sqrt{k_n}\left(I_1(t) - \frac{\nu_X(A_1)}{t}\right) \\
&\rightsquigarrow \mathbb{G}(1/t, 1/t) - \Lambda_1(1/t, 1/t)1\mathbb{G}(1/t, +\infty) - \Lambda_2(1/t, 1/t)\mathbb{G}(+\infty, 1/t) \\
&=: \mathbb{W}(t), \qquad (20)
\end{aligned}$$

其中，$\{\mathbb{W}(t)\}_{t\in(0,+\infty)}$ 是具有特定协方差结构的中心化高斯过程。我们只关注紧集 $t \in [a, b]$ 上的收敛，其中 $0 < a < b < +\infty$。因此，收敛在 $[a, b]$ 上是一致的。

^^ 接下来考虑 $I_2(t) = \Lambda(1/t, +\infty)$ 和 $I_3(u) = \Lambda(+\infty, 1/t)$。事实上，这两个量是非随机的，因为 ^$\boldsymbol{X}$ 的边缘分布是确定值 $n/1, n/2, \cdots, n/n$。容易验证，当 $n \to \infty$ 时，

$$\sqrt{k_n}(I_j(t) - 1/t) \to 0,$$

对所有 $t \in [a, b]$ 和 $j = 2, 3$. 该结果也可通过应用引理 7。

定义

$$I(t) = I_2(t) + I_3(t) - I_1(t) = \frac{1}{k_n}\sum_{i=1}^{n} \mathbf{1}\left\{r(\hat{\boldsymbol{X}}_i) > \frac{n}{k_n}t\right\},$$

其中 $r(\boldsymbol{x})$ =max($\boldsymbol{x}$)。因此，当 $n \to \infty$ 时，我们得到，

$$\sqrt{k_n}\left(I(t) - \frac{1}{t}\nu_X(\Omega_r))\right) \rightsquigarrow -\mathbb{W}(t),$$

对于 $t \in [a, b]$。

在适当的 Skorokhod 构造下，上述收敛可以是几乎必然收敛。特别地，取 $a = \nu_X(\Omega_r)/2$ 和 $b = 2\nu_X(\Omega_r)$。然后将 Vervaat 引理应用于这一极限关系，可得当 $n \to \infty$ 时，对 $v \in [1/2, 2]$,

$$\sqrt{k_n}\left(\frac{k_n}{n}R^{\hat{X}}_{n-[k_n v],n} - v^{-1}\nu_X(\Omega_r)\right) \rightsquigarrow -\frac{\nu_X(\Omega_r)}{v^2}\mathbb{W}\left(v^{-1}\nu_X(\Omega_r)\right).$$

令 $v = 1$，可得当 $n \to \infty$ 时，

$$\sqrt{k_n}\left(\frac{k_n}{n}R^{\hat{X}}_{n-k_n,n} - \nu_X(\Omega_r)\right) \rightsquigarrow -\nu_X(\Omega_r)\mathbb{W}\left(\nu_X(\Omega_r)\right). \qquad (21)$$

因此，我们可以在（$t = \frac{k_n}{n} R^{\hat{X}}_{n-k_n,\, n}$ 式（20）中替换。请注意，这样替换后得到的恰好是估计量 $\hat{p}_1$ 因为

$$I_1\left(\frac{k_n}{n} R^{\hat{X}}_{n-[k_n],n}\right) = \frac{1}{k_n}\sum_{i=1}^{n} \mathbf{1}\left\{\hat{\boldsymbol{X}}_i \in R^{\hat{X}}_{n-[k_n],n} A_1\right\} = \hat{p}_1$$

因此，我们得到当 $n \to \infty$

$$\sqrt{k_n}\left(\hat{p}_1 - \left(\frac{k_n}{n} R^{\hat{X}}_{n-[k_n],n}\right)^{-1} \nu(A_1)\right) \rightsquigarrow \mathbb{W}(\nu_X(\Omega_r)).$$

结合式（21）以及以下事实： $p_1 = \nu（A_1）/\nu（\Omega_r）$，我们得到当 $n \to \infty$,

$$\sqrt{k_n}\,(\hat{p}_1 - p_1) \rightsquigarrow (1 + p_1)\mathbb{W}(\nu_X(\Omega_r)).$$

类似地，考虑 $I_2(t) - I_1(t)$ 并将 $t = \frac{k_n}{n} R^{\hat{X}}_{n-k_n,\, n}$ 替换后,—可得当$n \to \infty$

$$\sqrt{k_n}\left(\hat{p}_2 - \left(\frac{k_n}{n} R^{\hat{X}}_{n-[k_n],n}\right)^{-1} \nu(A_2)\right) \rightsquigarrow -\mathbb{W}(\nu_X(\Omega_r)).$$

结合式（21）以及以下事实： $p_2 = \nu（A_2）/\nu（\Omega_r）$，我们得到当 $n \to \infty$,

$$\sqrt{k_n}\,(\hat{p}_2 - p_2) \rightsquigarrow (-1 + p_2)\mathbb{W}(\nu_X(\Omega_r)).$$

类似地，我们有当 $n \to \infty$,

$$\sqrt{k_n}\,(\hat{p}_3 - p_3) \rightsquigarrow (-1 + p_3)\mathbb{W}(\nu_X(\Omega_r)).$$

记 $N_X = \mathbb{W}(\nu_X(\Omega_r))$。我们得到估计量的联合收敛：当 $n \to \infty$,

$$\sqrt{k_n}\begin{pmatrix} \hat{p}_1 - p_1 \\ \hat{p}_2 - p_2 \\ \hat{p}_3 - p_3 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} 1 + p_1 \\ -1 + p_2 \\ -1 + p_3 \end{pmatrix} N_X.$$

在法ct 经过简单计算可得 $N_X = \mathbb{W}(\nu_X(\Omega_r))$ 的方差为

$$\mathbb{V}(N_X) = \frac{(\Lambda - 1)(\Lambda + 2\Lambda_1\Lambda_2)}{\Lambda},$$

其中 $\Lambda = \Lambda(\mathbf{1}) = \nu_X(\Omega_r)$， $\Lambda_j = \Lambda_j(\mathbf{1})$，对于 $j = 1$， 2。

我们可以用类似的方法推导 $\hat{}q_j$ $j = 1,2,3$ 的渐近行为，其中极限随机变量 N_Y 独立于 N_X。

在原假设 $p_j = q_j$ 下，对于 $j = 1$， 2， 3，类似于定理2的证明，我们得到当 $n \to \infty$时，

$$\begin{aligned}\frac{k_n}{2}\hat{D}_K &\xrightarrow{d} \frac{1}{2p_1}(1+p_1)^2(N_X-N_Y)^2 + \frac{1}{2p_2}(-1+p_2)^2(N_X-N_Y)^2 + \frac{1}{2p_3}(-1+p_3)^2(N_X-N_Y)^2 \\ &\stackrel{d}{=} \left(\sum_{j=1}^{3}\frac{1}{p_j} - 1\right)\frac{(\Lambda-1)(\Lambda+2\Lambda_1\Lambda_2)}{\Lambda}\chi^2(1)\end{aligned}$$

在备择假设下，类似于定理 2的证明，我们得到当$n \to \infty$时，

$$\begin{aligned}&\sqrt{k_n}\left(\hat{D}_K - D_K\right) \\ \xrightarrow{d}&[D_{1,1}(1+p_1) + D_{1,2}(-1+p_2) + D_{1,3}(-1+p_3)]N_X \\ &+ [D_{2,1}(1+q_1) + D_{2,2}(-1+q_2) + D_{2,3}(-1+2_3)]N_Y,\end{aligned}$$

由此得到渐近正态性。进一步计算可得出极限分布的方差，如推论中所示。

□

C D散度作为极值相关的函数

考虑附录 B 中相同的三个集合 A_1, A_2, A_3 所对应的最大风险泛函，其对应概率为 p_j, q_j， $j = 1, 2, 3$。注意，根据标准化，概率 $p_2 = p_3$ 和 $q_2 = q_3$。此外，注意到 $1 = p_1 + p_2 + p_3 = p_1 + 2p_2$，因此 $p_2 = (1 - p_1)/2$。与指数测度 ν_X 对应的极值相关系数为 $\chi_X = p_1/(p_1 + p_2) = 2p_1/(1 + p_1)$。求解可得 $p_1 = \chi_X/(2 - \chi_X)$，类似地， $q_1 = \chi_Y/(2 - \chi_Y)$。我们还得到 $p_2 = p_3 = (1 - \chi_X)/(2 - \chi_X)$， $q_2 = q_3$ 也同理。将其代入 KL

散度（5）可得

$$D_3 = \left(\frac{\chi_X}{2-\chi_X} - \frac{\chi_Y}{2-\chi_Y}\right)\left(\frac{\chi_X}{2-\chi_X} - \log\frac{\chi_Y}{2-\chi_Y}\right) + 2\left(\frac{1-\chi_X}{2-\chi_X} - \frac{1-\chi_Y}{2-\chi_Y}\right)\left(\frac{1-\chi_X}{2-\chi_X} - \log\frac{1-\chi_Y}{2-\chi_Y}\right).$$

这表明，最大风险泛函的 KL散度是极值相关系数的简单函数。

D 检验效能的更多模拟

与图 4类似，图 9 展示了风险泛函为最大风险泛函 $r(\mathbf{x}) = \max(x_1, x_2)$。

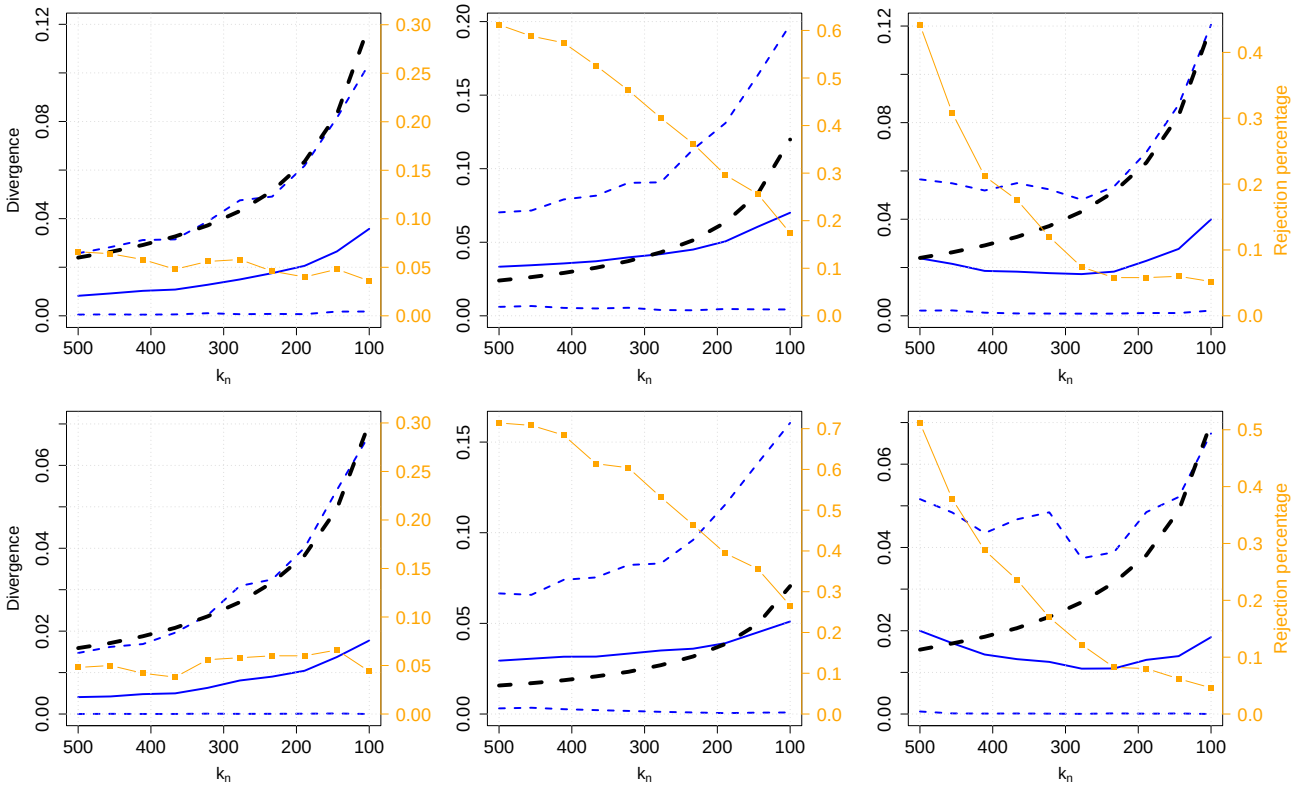


图 9：基于最大风险泛函、采用 $K = 3$ 集合的 ˆ 500 个样本的 KL检验统计量 D_K 的均值（蓝色实线）以及经验 5% 和 95% 分位数（蓝色虚线），对应一系列超越值 k_n；同时给出 95% 水平下的临界值（黑色虚线）和拒绝百分比（橙色线）。顶行和底行分别展示已知边缘分布和未知边缘分布的结果。两个样本分别来自相同分布（左）、极值依赖结构不同的分布（中）以及极值依赖结构相同但分布不同的分布（右）。