arXiv:2301.01653v2 [stat.ME] 6 Aug 2023

同时方向性推断

Ruth Heller*特拉维夫大学统计与运筹学系 ruheller@gmail.com Aldo Solari 米兰比可卡大学经济、管理与统计系 solari.aldo@gmail.com

2023年8月

摘要

我们研究对 $n > 1$ 参数符号进行推断的问题。我们的目标是为正参数和负参数（或非正参数）的数量提供 $1 - \alpha$ 事后置信界。该保证对所有参数子集同时成立。我们的建议如下：首先利用数据为每个参数选择假设检验的方向；然后针对选择调整单侧假设的 p值，并使用调整后的 p值，对选出的 n 单侧假设进行同时推断。假设单侧假设的 p值具有单调似然比密度且相互独立，则调整过程十分直接。我们证明，所提供的界限比现有方法更紧（往往紧得多），而且至多可在多项式时间内求得。我们展示了同时事后界在评估不同研究或亚组中的治疗效应时的实用价值。具体而言，我们给出有益研究数量以及有害（或非有益）研究数量的紧下界；此外，我们还对各项研究的效应方向作出结论，同时保证至少出现一次错误推断的概率不超过 0.05。

*作者衷心感谢意大利与以色列科学合作丽塔·莱维-蒙塔尔奇尼奖。

关键词：条件推断；方向性决策；荟萃分析；多重检验；划分原理；同时置信界。

1 引言

设 $\theta = (\theta_1, ..., \theta_n)$ 为由 n 个未知实值参数组成的向量。一种传统的分析方法是进行双侧多重检验，例如，使用一种能够在预先指定的显著性水平 α 下控制族错误率（FWER）的程序，检验点零假设族$H_1: \theta_1 = 0, ..., H_n: \theta_n = 0$。如果第 i 个零假设被拒绝，则结论是 $\theta_i \neq 0$。Tukey（1991）认为，这样的结论并不令人满意，分析应当转而对参数符号作出结论，即 $\theta_i > 0$，或 $\theta_i < 0$。拒绝后，人们很容易根据数据对参数符号作出结论。例如，对于被拒绝的点零假设，如果其点估计为正，则得出其为正的结论；否则得出其为负的结论。然而，如果在拒绝点零假设 $\theta_i = 0$ 后，我们判定 $\theta_i > 0$，而实际情况是 $\theta_i < 0$（或者判定 $\theta_i < 0$，而实际情况是 $\theta_i > 0$），就可能发生方向性错误，通常称为第三类错误。因此，至少发生一个第一类错误或第三类错误的概率，以下简称方向性族错误率（dFWER），可能大于 α，即使 FWER $\leq \alpha$ 也是如此。关于在检验统计量相互独立但服从柯西分布时，使用 Holm 程序（Holm, 1979）检验点零假设族却无法实现控制的示例，参见 Shaffer（1980）。

对于满足单调似然比（MLR）的独立检验统计量 θ_i（Karlin 和 Rubin 1956年），已有研究表明，现有的双侧多重检验程序能够控制方向性家族错误率：Holm程序（Shaffer，1980年）、Hochberg程序（Finner，1994；Liu，1997年），以及更一般地说，任何闭合检验程序（Finner，1999年）。例如，对于连续指数族（其中 p值是充分统计量的单调变换）， p值密度族满足单调似然比性质。对于检验单个均值时观测值服从正态分布且方差未知的情况，或比较两个均值时假设方差相等但未知的情况，单调似然比性质同样成立（Lehmann 和 Romano，2005年）。

与其采用现有的双侧检验程序来判断每个参数的符号，不如考虑检验由以下 n 对给出的 $2n$ 单侧假设这一问题：

$$H_i^- : \theta_i \leq 0, \quad H_i^+ : \theta_i \geq 0, \quad i = 1, \ldots, n. \tag{1}$$

对于由（1）中的 $2n$ 假设构成的假设族，多重检验程序的拒绝中内置了方向性决策，因此常规错误率（FWER）与方向性错误率（dFWER）一致。令 p_i 和 q_i 分别为 H_i^- 和 H_i^+ 的 p值，其中 $i = 1$，……，n。

Shaffer（1974年）针对 $n = 1$ 考虑了（1），并将其称为方向性假设对。Holm（1979）；Shaffer（1980，2002）；Bauer 等人（1986）；Finner（1999）；Guo 和 Romano（2015）等人考虑了多个方向性假设对。Bauer 等人（1986）证明，在没有额外分布假设的情况下，对于（1）中假设族的 $2n$ p 值，要在 α 家族错误率控制显著性水平下显著优于 Holm程序并不可行。遗憾的是，拒绝次数可能远少于基于双侧 p 值的 Holm程序（后者要保证有效性需要更多分布假设）；详见 § 1.1。

在这项工作中，我们不仅提供用于作出正向和负向发现的高效程序，还提供用于界定正参数和负参数总数，以及任何指标集$\mathcal{I} \subseteq \{1, \ldots, n\}$中正参数和负参数数量的程序。令 $n^+(\mathcal{I})$ 和 $n^-(\mathcal{I})$ 分别表示 $\mathcal{I}$中取正值和负值的参数数量：

$$n^+(\mathcal{I}) = |\{i \in \mathcal{I} : \theta_i > 0\}|, \quad n^-(\mathcal{I}) = |\{i \in \mathcal{I} : \theta_i < 0\}|.$$

为简化记号，我们使用 n^+ 和 n^-，分别代替 $n^+([n])$ 和 $n^-([n])$，其中 $[n]$ 是指标集 $\{1, \ldots, n\}$。我们的第一个目标是针对任意 $\mathcal{I}$，为 $n^+(\mathcal{I})$ 和 $n^-(\mathcal{I})$ 提供紧下界。更正式地说，我们旨在提供泛函 $\ell_\alpha^+(\cdot)$、$\ell_\alpha^-(\cdot)$，使得以下保证成立：

$$\mathbb{P}_\theta\Big(n^+(\mathcal{I}) \geq \ell_\alpha^+(\mathcal{I}),\ n^-(\mathcal{I}) \geq \ell_\alpha^-(\mathcal{I}),\ \text{for all } \mathcal{I}\Big) \geq 1 - \alpha. \tag{2}$$

我们将在本文中证明，相较于基于 $\mathcal{I}$ 中各个方向的发现数量所得的界限，可以得到紧得多的下界（即，对于 $n^+(\mathcal{I})$ 的下界，$|i \in \mathcal{I} : H_i^-$ 被拒绝 $|$；对于 $n^-(\mathcal{I})$ 的下界，$|i \in \mathcal{I} : H_i^+$ 被拒绝 $|$）。

在各种应用中，紧致界限都很重要。例如，考虑评估治疗在多个亚组（或队列）中的治疗效应。设 θ_i 为亚组 i 的平均治疗效应；对于满足 $\theta_i > 0$ (n^+) 的亚组数量，以及满足 $\theta_i < 0$ (n^-) 的亚组数量，若其下界均为正，则说明治疗效应存在异质性。这可能具有重大的临床重要性，因为它让研究者知道，至少对某些亚组而言，治疗

是有害而非有效的（即存在定性交互作用，Gail 和 Simon，1985年；Zhao 等人，2019年）。

在某些情况下，只需对正向和非正向结果进行推断即可，即我们只关心 $n^+(\mathcal{I})$ 和 $n^-(\mathcal{I}) + n^0(\mathcal{I})$ 的下界，其中 $n^0(\mathcal{I}) = |\mathcal{I}| - n^+(\mathcal{I}) - n^-(\mathcal{I}) = |i \in \mathcal{I}: \ \theta_i = 0|$。例如，在临床试验的系统综述中，评估具有正向治疗效应的研究数量界限（即 $\theta_i > 0$）十分重要；而关注 $n^- + n^0$ 的正下界也很重要，因为这表明治疗可能无效，甚至可能有害（IntHout 等人，2016）。更一般地，在可重复性分析（Bogomolov 和 Heller，2022）中，如果关注的是确定多个研究中是否存在正向关联（即 $\theta_i > 0$），那么以 $1-\alpha$ 置信度给出 $n^+(\mathcal{I})$ 以及 $n^-(\mathcal{I}) + n^0(\mathcal{I})$ 的下界就具有重要意义：$n^+(\mathcal{I})$ 的下界给出了可以有把握地断言、沿所需正向方向达到的最低可重复程度；$n^-(\mathcal{I}) + n^0(\mathcal{I})$ 的下界则提供了可重复性上限的证据——正下界意味着至少有一些研究未能重复。

因此，我们希望提供 $\tilde{\ell}^+_\alpha(\cdot)$, $\tilde{\ell}^-_\alpha(\cdot)$ 使得满足以下保证：

$$\mathbb{P}_\theta\Big(n^+(\mathcal{I}) \geq \tilde{\ell}^+_\alpha(\mathcal{I}),\ n^-(\mathcal{I}) + n^0(\mathcal{I}) \geq \tilde{\ell}^-_\alpha(\mathcal{I}),\ \text{for all } \mathcal{I}\Big) \geq 1-\alpha. \tag{3}$$

由于 $n^-_\alpha(\mathcal{I}) \leq n^-_\alpha(\mathcal{I}) + n^0(\mathcal{I})$，满足 (2) 的泛函 $\ell^+_\alpha(\mathcal{I})$ 和 $\ell^-_\alpha(\mathcal{I})$ 也满足 (3)。因此，若针对要求较宽松的 (3)，就有可能改进下界。为此，我们修改 (1) 中的 H^+_i，去掉等号。因此，对于 $i = 1, \ldots, n$，所考虑的假设为：

$$H^-_i : \theta_i \leq 0, \quad K_i : \theta_i > 0, \quad i = 1, \ldots, n. \tag{4}$$

假设对中恰有一对为真，也就是说，在式（4）定义的 $2n$ 原假设族中，恰有 n 个单侧原假设为真。因此，可以设计出比双侧检验中通常采用的程序更有检验效能的程序，以识别正向发现和非正向发现（Bauer 等人（1986）；Guo 和 Romano，2015）。Bauer 等人（1986）指出，使用 Bonferroni 检验式（4）中定义的 $2n$ 假设时，发现是依据 Bonferroni 截断值 α/n作出的，该值是对假设族（1）使用的 Bonferroni 截断值的两倍。Shaffer（1986）的研究可以解释这一点：他指出，执行 Bonferroni 程序时，家族错误率控制所需的适当校正因子并非受检验假设的数量，而是可能同时为真的原假设的最大数量（Goeman 等人，2010）。在 § 1.1 中，我们将回顾用于式（4）的其他家族错误率和错误发现率控制程序。

与式（1）中的假设对相比，我们推导的界限在考虑式（4）中的假设对时始终更紧。因此，除非需要同时推断正参数和负参数，否则应始终优先采用基于式（4）中假设对的方法。Tukey（1991）也支持对式（4）进行检验。针对单对假设，Tukey（1991）这样论述：“统计学家历来提出了错误的问题，而且愿意用谎言作答——而且往往是彻头彻尾的谎言。他们问‘A 和 B 的效应是否不同？’，并愿意回答‘不’。我们对世界的了解告诉我们，对于任意 A 和 B，它们的效应总是不同的——哪怕只差在某个小数位上。因此，问‘效应是否不同？’是愚蠢的。我们首先应该回答的问题是：‘我们能否判断 A 的效应相对于 B 的效应的方向？’换句话说，我们能否确定从 A 到 B 的方向？是‘向上’、‘向下’，还是‘不确定’？”因此，如果点零假设从不为真，那么显然应优先选择式（4）中的假设，而不是式（1）中的假设。据 Tukey（1991）以及 Jones 和 Tukey（2000）所述，这种情况很常见。

1.1 检验 (1) 和 (4) 的家族错误率与错误发现率控制程序

令 $x_i = \min(p_i, q_i)$ $i = 1, \ldots, n$ 为 $2n$ 单侧假设中 n 最小的 p 值。令 $x_{(1)} \leq \ldots \leq x_{(n)}$ 为这些值的排序值。给定临界值 $\alpha_1 \leq \ldots \leq \alpha_n$，方向性推断程序如下：

1. 令

$$R = \begin{cases} \max\{r : x_{(r)} \leq \alpha_i,\ \forall i \leq r\} & \text{for a step-down procedure ,} \\ \max\{r : x_{(r)} \leq \alpha_r\} & \text{for a step-up procedure .} \end{cases} \tag{5}$$

2. 对于 $r \leq R$，若 $x_{(r)}$ 是 H_i^- 的 p 值，则声明 $\theta_i > 0$；否则，若检验 (1)，则声明 $\theta_i < 0$；若检验 (4)，则声明 $\theta_i \leq 0$。

对于检验 (1) 的家族错误率控制，无论检验统计量之间具有何种依赖结构，均可采用有效程序：对 $2n$ p 值使用 Holm程序，即临界值为 $\alpha_i = \alpha/(2n - i + 1)$ $i = 1, \ldots, n$ 的逐步向下程序。

以下程序已证明，在额外假设分布具有一致有效性的条件下，能够控制家族错误率（Guo 和 Romano，2015；Zhao 等人，2019；

表1：现有和拟议程序的概览，包括错误准则和所考虑的推断类型。临界值对应于公式 (5) 中概述的逐步程序。前四行中的程序要求 n 检验统计量具有条件有效性且相互独立，即满足 § 3 中的条件 (A1)-(A2)。最后两行中的程序仅要求 n检验统计量相互独立。

准则	程序	推断类型	临界值 α_i
FWER	Holm n 双侧（Shaffer，1980年）	正/负	$\alpha/\{2(n-i+1)\}$
	Guo-Romano（Guo 和 Romano，2015）	正/非正	$\alpha/(n-i+1+\alpha)$
FDP	方向性闭合检验（§ 3）	正/负	-
	自适应检验划分法（§ 4）	正/非正	-
FDR	Benjamini-Hochberg（Benjamini 和 Yekutieli，2005年）	正/负	$i\alpha/2n$
	Guo-Romano（Guo 和 Romano，2015）	正向/非正向	$i\alpha/n$

Ellis 等人，2020）。若对于所有 $\tau \in (0,1)$，给定 $x_i \leq \tau$ 的有效 p-值 x_i x_i/τ 均有效，则称其具有一致有效性；也就是说，若零假设成立，则

$$\mathbb{P}(x_i \leq t \mid x_i \leq \tau) \leq \frac{t}{\tau} \quad \forall \quad 0 \leq t \leq \tau \leq 1.$$

一致有效性的一个充分条件是，针对 θ_i 的检验统计量具有单调似然比性质。实际上，只需要求对 $\tau = 1/2$ 满足上述不等式即可（这是较弱的条件有效性，定义见第 3 节中的 $(A1)$）。这些程序详见表1的前两行。

请注意，第 1 行的程序对 $2n$ p 值采用的临界值高于 Holm程序（其临界值为 $\alpha_i = \alpha/(2n-i+1)$ $i = 1, \ldots, n$），因此可能带来多得多的发现。此外，第 2 行程序的临界值几乎是第 1 行程序的两倍，所以如果发现正向和非正向参数（而非负向参数）就足够了，应优先采用该程序。

表1最后两行的程序分别对正向和负向发现（第 5 行）以及正向和非正向发现（第 6 行）提供错误发现率控制。就发现数量而言，第 6 行的程序始终优于第 5 行的程序，因为其临界值是后者的两倍。

我们以简要概述本文的贡献来结束本节，其中包括表1第3行和第4行中的两个程序。

1.2 我们贡献的简要概述

闭合检验程序不仅可用于发现，也可用于对 $\mathcal{I}$ 中真实发现的数量给出同时置信下界，或等价地，对 $\mathcal{I}$ 中的错误发现比例（FDP）给出同时置信上界（Goeman and Solari 2011；Blanchard et al. 2020，见第 2 节）。这些界限通常称为事后置信界。使用闭合检验程序检验点零假设时，真实发现的数量就是非零参数的数量。非零参数数量的下界并未揭示该集合中正参数和负参数的数量，因此可能并不令人满意。更具体地说，若有一个函数 $\ell_\alpha(\cdot)$ 满足

$$\mathbb{P}_\theta\Big(n^+(\mathcal{I}) + n^-(\mathcal{I}) \geq \ell_\alpha(\mathcal{I}),\ \text{for all}\ \mathcal{I}\Big) \geq 1-\alpha. \tag{6}$$

如何提取满足 (2) 的下界尚不明确。

为了求得满足 (2) 的 $\ell_\alpha^+(\cdot)$ 和 $\ell_\alpha^-(\cdot)$，我们提出以下两步方法，以下称为方向性闭合检验（DCT）：首先，根据数据从每一对中选出要检验的假设；其次，对选出的 n单侧假设应用 α 显著性水平闭合检验程序（见第 2 节）。当然，第二步必须对基于同一数据进行的第一步选择作出调整。采用这一方法，对于任意 $\mathcal{I}$，我们都可以为 $n^+(\mathcal{I})$ 和 $n^-(\mathcal{I})$ 提供同时置信界；此方法建立在 Goeman 和 Solari（2011）的工作之上，他们展示了如何为闭合检验程序获得同时界限。具体而言，我们可以为 n^+ 和 n^-提供紧下界，也可以给出正向和负向发现，并保证任何错误推断的概率至多为 α。

有趣的是，针对双侧 p 值的闭合检验程序（由 Finner（1999）提出，用于控制方向性家族错误率）隐含地从 DCT 程序的第一步开始。然而，将针对双侧 p 值的闭合检验视为 DCT 程序有若干优势。首先，认识到第一步是选择待检验的单侧假设后，后续推断便明确针对所选的单侧假设，因此能够控制通常的错误率（家族错误率），无需专门定义方向性错误率（dFWER）；关于错误率概念过多可能造成的危害，参见 Benjamini（2010）的讨论。其次，除了对单个参数的符号作出结论外，DCT 还会如式（11）所述，同时给出 $n^+(\mathcal{I})$ 和 $n^-(\mathcal{I})$ 的下界。第三，第 3 节中的充分假设（$A1$）比 Shaffer（1980）、Finner（1999）、Sarkar 等人（2004）以及 Guo 和 Romano（2015）为控制方向性家族错误率所考虑的分布假设更为一般。

DCT 的另一个优势是，其程序的第二步可以先使用 Zhao 等人（2019）提出的先进方法检验定性交互作用（即检验所有参数均为非正或所有参数均为非负这一零假设）。因此，检测定性交互作用的检验效能与 Zhao 等人（2019）的方法相同；此外，DCT 无需额外承担多重性代价，即可为感兴趣参数的任意子集给出正参数和负参数数量的发现及置信下界。详情参见第 5.2 节。

值得注意的是，一旦我们获得正参数和负参数数量的置信下界，就可以自动推导出相应的上界；也就是说，如果 $n^+(\mathcal{I}) \geq \ell_\alpha^+(\mathcal{I})$ $n^-(\mathcal{I}) \geq \ell_\alpha^-(\mathcal{I})$ 成立，那么 $n^+(\mathcal{I}) \leq |\mathcal{I}| - \ell_\alpha^-(\mathcal{I})$ $n_\alpha^-(\mathcal{I}) \leq |\mathcal{I}| - \ell_\alpha^+(\mathcal{I})$ 也成立。

为了找到满足 (3) 的 $\ell_\alpha^+(\cdot)$ 和 $\ell_\alpha^-(\cdot)$，我们建议对由 $\{(-\infty, 0], (0, \infty)\}^n$ 定义的 2^n 个正交象限应用划分原理（参见第 2 节）。每个正交象限都定义了一个正交象限假设，即 θ 位于该正交象限中。共有 2^n 个正交象限假设，但其中恰有一个为真。因此，只需对每个假设进行显著性水平为 α 的检验。这与 DCT 不同，DCT 需要进行多重检验校正（由于对于零向量 $\theta = 0$，式 (1) 中的 $2n$ 个假设都成立，因此必须进行校正）。

我们展示了提取界限的一般方法，以及一种具体的多项式时间算法；该算法在均匀意义上优于 DCT 界限 $\ell_\alpha^+(\cdot)$ $\ell_\alpha^-(\cdot)$，因此 $\tilde{\ell}_\alpha^+(\cdot) \geq \ell_\alpha^+(\cdot)$，$\tilde{\ell}_\alpha^-(\cdot) \geq \ell_\alpha^-(\cdot)$。

在第 2 节中，我们回顾本研究将要用到的必要统计学背景，以及部分符号和定义。接着，我们在第 3 节详细介绍 DCT，并在第 4 节介绍分割法。在第 5 节中，我们展示所提出方法的应用：用于亚组分析、在定性交互作用检验后进行推断，以及改进荟萃分析。在第 6 节中，我们通过模拟比较和对照这些方法，并提出一些实用建议。此外，我们在表1中量化了相较于旨在控制家族错误率或错误发现率的方法所取得的改进（其下界仅通过计算各个方向上的拒绝次数得出）。最后的总结见第 7 节。所有证明均见补充材料（SM）第 A 节。可在 `https://aldosolari.github.io/directionalinference/` 获取用于复现示例的 R 代码。

2 同步推断背景

令 $H_1, \ldots, H_n$ 为 n 基准零假设，或基假设，它们构成推断的基本单元。对应的 p 值为 $x_1, \ldots, x_n$。每个 x_i 都是有效的 p 值

对于各自的零假设 H_i，即当 H_i 为真时，其分布为均匀分布或随机占优于均匀分布。对于 $\mathcal{I} \subseteq \{1, \ldots, n\}$，令 $H_{\mathcal{I}}$: $\bigcap_{i \in \mathcal{I}} H_i$ 为 $\mathcal{I}$ 中 $|\mathcal{I}|$ 个基假设的交集假设。若 $\forall i \in \mathcal{I}$，$H_i$ 为真，则该假设为真。对 $H_{\mathcal{I}}$ 的显著性水平为 α 的检验记为 $\phi_{\mathcal{I}} \in \{0, 1\}$，称为局部检验。它满足：当 $\theta \in H_{\mathcal{I}}$ 时，$\mathbb{P}_\theta(\phi_{\mathcal{I}} = 1) \leq \alpha$。

闭合检验程序 闭合检验程序通过仅当 $H_{\mathcal{I}}$ 被所有包含 $\mathcal{I}$ 的交集假设所拒绝时才拒绝 $H_{\mathcal{I}}$，来校正多重检验中的局部检验；也就是说，若 $\phi_{\mathcal{I}} = 1$，其中 $\bar{\phi}_{\mathcal{I}} = \min\{\phi_{\mathcal{J}} : \mathcal{J} \supseteq \mathcal{I}\}$。Marcus 等人（1976年）证明，调整后的检验 $\bar{\phi}_{\mathcal{I}}$ 具有家族错误率控制：$\mathbb{P}_\theta\left(\bar{\phi}_{\mathcal{I}} = 0 \text{ 对所有 } \mathcal{I} \subseteq \mathcal{T}\right) \geq 1 - \alpha$ 均成立，其中 $\mathcal{T}$ 是真实基本假设的（未知）指标集。发现是指被拒绝的基本假设，$\mathcal{D}_\alpha = \{i \in [n] : \phi_i = 1\}$。

同时事后界 Goeman 和 Solari（2011年）证明，$\mathcal{I}$, $|\mathcal{I} \setminus \mathcal{T}|$ 中真实发现数量的下界为

$$\ell_\alpha(\mathcal{I}) = \min(|\mathcal{I} \setminus \mathcal{J}| : \mathcal{J} \subseteq \mathcal{I}, \bar{\phi}_{\mathcal{J}} = 0), \tag{7}$$

并保证对所有 $\mathcal{I}\Big) \geq 1 - \alpha.$，$\mathbb{P}_\theta\Big(|\mathcal{I} \setminus \mathcal{T}| \geq \ell_\alpha(\mathcal{I})$。对于双侧假设，这就是（6）中的保证。注意，$\ell_\alpha(\cdot)$ 可用于给错误发现比例提供上界，即 $|\mathcal{I} \cap \mathcal{T}|/|\mathcal{I}|$：以至少 $(1 - \alpha)$ 的置信度，$|\mathcal{I} \cap \mathcal{T}|/|\mathcal{I}|$ 至多为 $(|\mathcal{I}| - \ell_\alpha(\mathcal{I}))/|\mathcal{I}|$。

用于检验交集假设的合并函数。合并函数 f：$[0, 1]^{|\mathcal{I}|} \longrightarrow [0, 1]$ 将 H_i、$i \in \mathcal{I}$ 的 $|\mathcal{I}|$ p 值映射为有效的 p 值，用于检验 $H_{\mathcal{I}}$。局部检验为

$$\phi_{\mathcal{I}} = \mathbb{1}\{f(\{x_i, i \in \mathcal{I}\}) \leq \alpha\}.$$

对于相互独立的 p值，一种常用的合并函数是 Fisher 方法，它对各种信号都具有出色的检验效能性质（Benjamini 和 Heller，2008；Hoang 和 Dickhaus，2022），其形式为 $f_{\text{Fisher}}(x_1, \ldots, x_d) = \mathbb{P}(\chi^2_{2d} \geq -2\sum_{i=1}^{d} \log x_i)$，其中 χ^2_{2d} 是一个具有 $2d$ 自由度的卡方随机变量。令 $x_{(1)} \leq \ldots \leq x_{(d)}$ 为排序值。另一种常用的合并函数是 Simes[1]，$f_{\text{Simes}}(x_1, \ldots, x_d) = \min_{1 \leq i \leq d}\left\{\frac{d}{i} x_{(i)}\right\}$。它

[1]即使被合并的 p 值呈正相关依赖，这种合并方法也能为交集假设生成有效的 p 值，Goeman 等人（2019）。

本质上基于单个分位数，因此当正（或负）参数的数量较大时，无法充分利用合并证据。Bogomolov（2023）提出了以下组合函数 $\min_{1\le i\le d}\left\{\frac{2(\sum_{l=1}^{d}\mathbb{1}_{(x_l>0.5}+1}{i}x_{(i)}\right\})$，它结合使用了原假设数量的代入估计量。除交集仅包含两个假设的情况外，它可能远小于 Simes。因此，在这种情况下，改用 Simes组合方法是合理的。因此，我们将考虑的修正Simes（mSimes）组合函数为：

$$f_{\mathrm{mSimes}}(x_1,\ldots,x_d)=\begin{cases}\min_{1\le i\le d}\left\{\frac{2(\sum_{l=1}^{d}\mathbb{1}(x_l>0.5)+1)}{i}x_{(i)}\right\} & \text{if } d>2,\\ \min_{1\le i\le d}\left\{\frac{d}{i}x_{(i)}\right\} & \text{otherwise.}\end{cases} \tag{8}$$

我们对 $f(\cdot)$ 的假设适用于上述所有合并函数，具体如下：

$(A0)$ 合并函数满足单调性：对于所有 $i=1,\ldots,d$，$f(x_1,\ldots,x_d)\le f((x_1',\ldots,x_d'))$；并且满足对称性：对于 $(1,\ldots,d)$ 的任意置换 $(j_1,\ldots,j_d)$，$f(x_1,\ldots,x_d)=f(x_{j_1},\ldots,x_{j_d})$。

用于计算事后界的计算复杂度 计算 $\ell_\alpha(\mathcal{I})$ 需要评估数量呈指数级增长的检验，这妨碍了其在实践中的应用。然而，对于特定的合并函数，可以利用捷径在多项式时间内得出闭合检验结果。对于满足 $(A0)$ 的合并函数，已有用于家族错误率控制（Dobriban，2020）和同时错误发现比例对照（Goeman 和 Solari，2011；Goeman 等人，2021）的二次时间捷径。Tian等人（2021）指出，如果合并函数还满足可分性条件（定义见Tian等人（2021）的附录B），那么在对 p 值进行初步排序后，错误发现比例和家族错误率的捷径都可以缩短至线性时间。Fisher 合并函数以及基于 p 值广义均值的合并函数（Vovk 和 Wang，2020）均满足可分性条件。对于 Simes 合并函数，Goeman 等人（2019）所述的捷径可在线性时间内计算 $\ell_\alpha(\mathcal{I})$，但需先执行一个耗时为线性对数时间的预处理步骤。

划分原理 闭合检验是控制家族错误率的一项基本原理，但还有另一项同样基本的原理，称为划分原理（Finner 和 Strassburger，2002）。其核心思想是将参数空间划分为互不相交的区域，并同时检验真实参数值是否位于每个区域内。由于真实参数值恰好位于其中一个区域内，因此

只有一个真实的零假设，其余假设均为假。因此，如果针对真实零假设的检验能够控制 I 类错误概率，那么家族错误率也能得到控制。针对基假设 $H_1, \ldots, H_n$ 的分割程序，为每个 $\mathcal{I} \subseteq [n]$ 定义分割假设

$$J_{\mathcal{I}} = \Big\{ \bigcap_{i \in \mathcal{I}} H_i \Big\} \cap \Big\{ \bigcap_{j \in [n] \setminus \mathcal{I}} H_j^c \Big\},$$

其中，H_j^c 是第 j 个备择假设，即 H_j 的补集。当且仅当所有 H_i、$i \in \mathcal{I}$ 均为真，且所有 H_i、$i \in [n] \setminus \mathcal{I}$ 均为假时，分割假设 $J_{\mathcal{I}}$ 才成立。发现是所有满足以下条件的索引 $i \in [n]$：所有满足 $\mathcal{I} \ni i$ 的 $J_{\mathcal{I}}$ 均在显著性水平 α 下被拒绝。

闭合检验与分割法是实现家族错误率控制的等价原理：对于每个闭合检验程序，都存在一个拒绝完全相同假设的分割程序。同样，对于每个分割程序，都存在一个拒绝完全相同假设的闭合检验程序（Goeman 等人，2021）。Stefansson 等人（1988）和 Finner 等人（2021）指出，与闭合检验相比，划分原理在构造置信区间和置信集时更为方便。由于（4）中的假设族有 n 个为真的原假设，它们恰好分别属于 2^n 个可能正交象限中的一个，因此我们建议在 § 4 中使用划分原理，以获得 $n^+(\mathcal{I})$ 和 $n^0(\mathcal{I}) + n^-(\mathcal{I})$ 的置信集。

3 方向性闭合检验

令 $p = (p_1, \ldots, p_n)$ 表示 $(H_1^-, \ldots, H_n^-)$ 的 p-值向量，令 $q = (q_1, \ldots, q_n)$ 表示 $(H_1^+, \ldots, H_n^+)$ 的 p-值向量。为简便起见，我们假设检验统计量是连续的，且 $q_i = 1 - p_i, i \in [n]$。关于 p-值，我们的假设如下：

($A1$**)** 每个 p-值在条件下均有效：

$$\sup_{\theta \in H_i^-} \mathbb{P}_\theta(2p_i \le x \mid p_i \le 1/2) \le x \quad \forall x \in [0,1],$$
$$\sup_{\theta \in H_i^+} \mathbb{P}_\theta(2q_i \le x \mid p_i > 1/2) \le x \quad \forall x \in [0,1].$$

($A2$**)** $p_1, \ldots, p_n$ 相互独立。

我们关注对所选的一族 n 方向性假设 $\{H_i^-\colon\ p_i \le 1/2\} \cup \{H_i^+\colon\ p_i > 1/2\}$. 进行同时推断。我们提出以下程序。

程序 3.1（方向性闭合检验）。

步骤 1 选择要检验的 n 单侧假设：$\{H_i^-\colon p_i \le 1/2\} \cup \{H_i^+\colon p_i > 1/2\}$。令 $\mathcal{S}^- = \{i\colon p_i \le 1/2\}$ 和 $\mathcal{S}^+ = \{i\colon p_i > 1/2\}$ 分别表示我们检验 H_i^- 和 H_i^+ 时所用的索引。

步骤 2：使用条件 p-值 $\{2p_i, i \in \mathcal{S}^-\} \cup \{2q_j, j \in \mathcal{S}^+\}.$，对 n 个选定的单侧假设应用显著性水平为 α 的闭合检验程序。交集假设

$$H_{\mathcal{I}} : \Big\{ \bigcap_{i \in \mathcal{I} \cap \mathcal{S}^-} H_i^- \Big\} \cap \Big\{ \bigcap_{i \in \mathcal{I} \cap \mathcal{S}^+} H_i^+ \Big\}$$

当且仅当所有满足 $i \in \mathcal{I}$ 的选定假设均为真时，该假设为真。对于所有 $\mathcal{I} \subseteq [n]$，在显著性水平为 α 时，对 $H_{\mathcal{I}}$ 进行检验，所用的局部检验为

$$\phi_{\mathcal{I}} = \mathbb{1}\{f(\{2p_i, i \in \mathcal{I} \cap \mathcal{S}^-\} \cup \{2q_j, j \in \mathcal{I} \cap \mathcal{S}^+\}) \le \alpha\} \tag{9}$$

使用合并函数 $f(\cdot)$。

注意，$f(\{2p_i, i \in \mathcal{I} \cap \mathcal{S}^-\} \cup \{2q_j, j \in \mathcal{I} \cap \mathcal{S}^+\} = f(2 \min(p_i, q_i)\ i \in \mathcal{I})$，因此，对 $H_{\mathcal{I}}$ 的检验与使用双侧 p-值对 $\mathcal{I}$ 中的双侧原假设交集进行的检验相同。我们用 $\mathcal{S}$ 表示在 DCT 程序 3.1 的步骤 1 中作为条件的符号向量：

$$\mathcal{S} = (sign(p_1 - 1/2), \ldots, sign(p_n - 1/2)), \quad sign(p_i - 1/2) = \begin{cases} -1 & \text{if } p_i \le 1/2, \\ 1 & \text{if } p_i > 1/2. \end{cases}$$

在给定符号向量 $\mathcal{S} \in \{-1, 1\}^n$ 的条件下，拒绝所选假设中真零假设的交集假设的概率至多为 α，因为在式 (9) 的局部检验中，我们组合了（条件）p 值，而给定 $\mathcal{S}$ 时，这些值各自的分布都不小于均匀分布。我们将在下一个命题中对此进行形式化。

命题 3.1. 设 $p_1, \ldots, p_n$ 满足条件 $(A1)$–$(A2)$，并设 θ 为参数的真实未知向量。那么，在程序 3.1 第 1 步中选取用于检验的 n 中，$\mathcal{T} = \{\{i\colon \theta_i \in H_i^-\} \cap \mathcal{S}^-\} \cup \{\{i\colon \theta_i \in H_i^+\} \cap \mathcal{S}^+\}$ 是原假设为真的单侧假设的索引。则对 $\mathcal{T}$ 中假设的交集进行检验，满足条件第一类错误控制：

$$\sup_{\theta' \in H_{\mathcal{T}}} \mathbb{P}_{\theta'}\Big(\phi_{\mathcal{T}} = 1 \mid \mathcal{S}\Big) \le \alpha. \tag{10}$$

命题 3.1 表明，以至少 $1-\alpha$ 的概率，程序 3.1 中对交集假设的所有拒绝都是正确的；在给定 $\mathcal{S}$ 的条件下如此，因此无条件下也如此：$\mathbb{P}_\theta\left(\bar{\phi}_{\mathcal{I}}=0\ \forall\, \mathcal{I}\subseteq\mathcal{T}\mid\mathcal{S}\right)\geq 1-\alpha$，$\mathbb{P}_\theta\left(\bar{\phi}_{\mathcal{I}}=0\ \forall\, \mathcal{I}\subseteq\mathcal{T}\right)\geq 1-\alpha$，其中 $\bar{\phi}_{\mathcal{I}}=\min\{\phi_{\mathcal{J}}:\ \mathcal{J}\supseteq\mathcal{I}\}$，对于 (9) 中的 $\phi_{\mathcal{I}}$。

现在，我们明确展示如何根据程序 3.1，获得关于 $n^+(\mathcal{I})$ 和 $n^-(\mathcal{I})$ 的发现和置信界。令 $\mathcal{D}_\alpha^+=\{i\in\mathcal{S}^-:\ \phi_i=1\}$ 和 $\mathcal{D}_\alpha^-=\{i\in\mathcal{S}^+:\ \phi_i=1\}$ 分别为正向发现和负向发现的指标集，其中 $\phi_i=\min\{\phi_{\mathcal{J}}:\ \mathcal{J}\ni i\}$。该程序对所有 $i\in\mathcal{D}_\alpha^+$ 得出结论 $\theta_i>0$，并对所有 $i\in\mathcal{D}_\alpha^-$ 得出结论 $\theta_i<0$，同时保证在给定符号向量 $\mathcal{S}$ 的情况下，关于参数符号至少得出一个错误结论的概率至多为 α。

更一般地，对于任意 $\mathcal{I}\subseteq[n]$，我们可以如下使用 (7)。对于 $\mathcal{I}\cap\mathcal{S}^-$，$\ell_\alpha(\mathcal{I}\cap\mathcal{S}^-)$ 是正参数数量的下界：$n^+(\mathcal{I}\cap\mathcal{S}^-)\geq\ell_\alpha(\mathcal{I}\cap\mathcal{S}^-)$。类似地，对于 $\mathcal{I}\cap\mathcal{S}^+$：$n^-(\mathcal{I}\cap\mathcal{S}^+)\geq\ell_\alpha(\mathcal{I}\cap\mathcal{S}^+)$。由 (7) 和 (10) 可得以下界限

$$\ell_\alpha^+(\mathcal{I})=\ell_\alpha(\mathcal{I}\cap\mathcal{S}^-),\quad \ell_\alpha^-(\mathcal{I})=\ell_\alpha(\mathcal{I}\cap\mathcal{S}^+) \tag{11}$$

满足以下条件

$$\mathbb{P}_\theta\Big(n^+(\mathcal{I})\geq\ell_\alpha^+(\mathcal{I}),\ n^-(\mathcal{I})\geq\ell_\alpha^-(\mathcal{I})\text{ for all }\mathcal{I}\mid\mathcal{S}\Big)\geq 1-\alpha. \tag{12}$$

特别是对于 $\mathcal{I}=[n]$，$1-\alpha$ confidence bounds for n^+ and n^- 分别为 $\ell_\alpha^+=\ell_\alpha(\mathcal{S}^-)$ 和 $\ell_\alpha^-=\ell_\alpha(\mathcal{S}^+)$。注意，正向和负向发现由$\mathcal{D}_\alpha^+=\{i$给出：$\ell_\alpha^+(i)=1\}$，$\mathcal{D}_\alpha^-=\{i:\ \ell_\alpha^-(i)=1\}$，且 $\ell_\alpha^+(\mathcal{I})\geq|\mathcal{D}_\alpha^+|$，$\ell_\alpha^-(\mathcal{I})\geq|\mathcal{D}_\alpha^-|$始终成立。在 § 5 - § 6 中，我们将展示，对于 $\mathcal{I}=[n]$，下界与发现数量之间的差距可能很大；对于 Fisher 合并函数，这一差距比对于 mSimes 和 Simes 更大。最后请注意，$\ell_\alpha(\mathcal{I})$ 是非零参数数量 $n^+(\mathcal{I})+n^-(\mathcal{I})$ 的下界，并且 $\ell_\alpha(\mathcal{I})\geq\ell_\alpha^+(\mathcal{I})+\ell_\alpha^-(\mathcal{I})$ 始终成立。

计算 (11) 至多需要评估数量呈指数级增长的局部检验 (9)，但对于特定的合并函数，可以通过捷径在多项式时间内得出闭合检验结果，详情见第 2 节。因此，即使 n 适中，计算时间也在合理范围内。

3.1 相关的 p-值

若局部检验满足 (10)，则 DCT 有效。充分条件为 (A0)-(A2)。然而，对于 p-值或检验统计量之间的特定依赖关系，即使 (A2) 不满足，(10) 也可能仍然成立。

一个具有相关 p-值且局部检验满足 (10) 的特定情形，是检验 n 总体正态均值，其中总体方差 σ^2未知，但在所有总体中相同。这就是常见的单因素方差分析情形。我们将使用记号 $N(\mu,\tau)$、 χ^2_ν、 t_ν 和 $F_{k,\ \nu}$分别表示正态分布、卡方分布、中心t分布和中心F分布（其参数或自由度分别对应各分布）；而 $t_{\nu,\ x}$、 $F_{k,\ \nu,\ x}$ 分别表示 x分位数，其中t_ν 和 $F_{k,\ \nu}$。

均值的检验统计量 θ_i为 $y_i/(k_i s)$，其中 $y_i \sim N(\theta_i, k_i\sigma)$，而 $\nu s^2/\sigma^2 \sim \chi^2_\nu$则对应于适当的 k_i和 ν，它们取决于样本量 $i=1$、.....、 n。因此，对于零向量 $\theta=0$，对所有 $\mathcal{I}\subseteq[n]$， $y_i/(k_i s) \sim t_\nu$且 $\frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}}\frac{y_i^2}{k_i^2 s^2} \sim F_{|\mathcal{I}|,\ \nu}$。尽管 y_1、.....、 y_n相互独立，但检验统计量 $y_i/(k_i s)$ $i=1$、.....、 n 并不独立，因为 σ是利用所有观测值估计得到的。 p值为 $p_i=\mathbb{P}(t_\nu \geq y_i/(k_i s))$ 和 $q_i=\mathbb{P}(t_\nu \leq y_i/(k_i s))$。

一个自然的合并函数是用于 F检验的合并函数

$$f(\{2p_i, i\in\mathcal{I}\cap\mathcal{S}^-\}\cup\{2q_j, j\in\mathcal{I}\cap\mathcal{S}^+\} = f(2\min(p_i,q_i), i\in\mathcal{I}) = \mathbb{P}\left(F_{|\mathcal{I}|,\nu} \geq \frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}}\frac{y_i^2}{k_i^2 s^2}\right).$$

为说明这是一个有效的合并函数，注意对于任意 $\theta\in H_{\mathcal{I}}$ （在 DCT程序 3.1 的步骤 2 中定义）：

$$\begin{aligned}
&\mathbb{P}_\theta\left(\frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}}\frac{y_i^2}{k_i^2 s^2} \geq F_{|\mathcal{I}|,\nu,1-x} \geq \mid \mathcal{S}\right) = \mathbb{E}\left(\mathbb{P}_\theta\left(\frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}}\frac{y_i^2}{k_i^2 s^2} \geq F_{|\mathcal{I}|,\nu,1-x} \mid \mathcal{S}, s\right)\right)\\
&\leq \mathbb{E}\left(\mathbb{P}_0\left(\frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}}\frac{y_i^2}{k_i^2 s^2} \geq F_{|\mathcal{I}|,\nu,1-x} \mid \mathcal{S}, s\right)\right) = \mathbb{P}_0\left(\frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}}\frac{y_i^2}{k_i^2 s^2} \geq F_{|\mathcal{I}|,\nu,1-x} \geq \mid \mathcal{S}\right) \leq x,
\end{aligned}$$

其中，期望是对 s 的分布取值（该分布与 $\mathcal{S}$ 相互独立）；第二行中的第一个不等式由 Al Mohamad 等（2020）的引理 1 得出，最后一个不等式则源于均值向量为 $\theta=0$ 时 F 检验的有效性。对于双侧 p 值的 dFWER 控制，DCT 程序与 Finner （1999）提出的闭合F检验程序一致。它补充了 Finner（1999）提出的 dFWER 控制分析，还为所有 $\mathcal{I}\subseteq[n]$. 提供了针对 n^+ （$\mathcal{I}$）和 n^- （$\mathcal{I}$）的同时事后界

4 正参数和非正参数的同时推断

4.1 同时置信下界的一般方法

~ ~我们通过划分原理（见第 2 节），推导出针对 $n^+(\mathcal{I})$ 和$n^-(\mathcal{I}) + n^0(\mathcal{I})$ 的同时 $(1-\alpha)$ 置信下界 $\ell_\alpha^+(\mathcal{I})$ 和 $\ell_\alpha(\mathcal{I})$。核心思路是将参数空间 $\mathbb{R}^n$ 划分为互不相交的子空间：考虑这些正交象限 2^n

$$\Theta_{\mathcal{K}} = \{\theta \in \mathbb{R}^n : \theta_i > 0 \text{ for all } i \in \mathcal{K}, \theta_j \leq 0 \text{ for all } j \notin \mathcal{K}\}$$

对所有 $\mathcal{K}$均成立，因此恰有一个 $\Theta_{\mathcal{K}}$ 包含真实参数 θ。

每个正交象限 $\Theta_{\mathcal{K}}$ 都对应一个零假设 $J_{\mathcal{K}}$： $\theta \in \Theta_{\mathcal{K}}$ 该正交象限包含真实参数。且当且仅当所有 K_i、 $i \in \mathcal{K}$ 均为真，并且所有 H_j^-、 $j \notin \mathcal{K}$ 均为真时，该假设为真：

$$J_{\mathcal{K}} : \{\bigcap_{i\in\mathcal{K}} K_i\} \cap \{\bigcap_{i\notin\mathcal{K}} H_i^-\}.$$

令 $\psi_{\mathcal{K}} \in \{0,1\}$ 为假设 $J_{\mathcal{K}}$的显著性水平为 α 的局部检验，其中对所有 $\mathcal{K}$均成立。假设 $J_{\mathcal{K}}$两两互斥，且其中只有一个为真。对于真实的 $J_{\mathcal{K}}$，$\sup_{\theta \in \Theta_{\mathcal{K}}} \mathbb{P}_\theta(\psi_{\mathcal{K}} = 1) \leq \alpha$。

考虑假设 $J_v(\mathcal{I})$： $n^+(\mathcal{I}) = v$，它可以等价地表示为$J_v(\mathcal{I})$: $\bigcup_{\mathcal{K}:|\mathcal{K}\cap\mathcal{I}|=v} J_{\mathcal{K}}$。那么，当且仅当所有满足 $|\mathcal{K} \cap \mathcal{I}| = v$ 的 $J_{\mathcal{K}}$ 均被拒绝时， $J_v(\mathcal{I})$ 才被拒绝，并记

$$\psi_v(\mathcal{I}) = \min\{\psi_{\mathcal{K}} : \mathcal{K} \subseteq [n] \text{ and } |\mathcal{K} \cap \mathcal{I}| = v\} \tag{13}$$

用于 $J_v(\mathcal{I})$ 的统计检验。未能在显著性水平 α 下拒绝 $J_v(\mathcal{I})$ 的所有 v 值构成了 $n^+(\mathcal{I})$ 的一个 $(1-\alpha)$ 置信集。

命题 4.1。设 $\mathcal{N}^+(\mathcal{I})_\alpha = \{v \in \{0,\ldots,|\mathcal{I}|\}: \psi_v(\mathcal{I}) = 0\}$.。那么，对于任意 $\theta \in \mathbb{R}^n$， $\mathbb{P}_\theta(n^+(\mathcal{I}) \in \mathcal{N}^+(\mathcal{I})_\alpha$ 对所有 $\mathcal{I}) \geq 1-\alpha$。此外，由~ $\ell_\alpha^+(\mathcal{I}) = \min(\mathcal{N}^+(\mathcal{I})_\alpha)$ $\ell_\alpha^-(\mathcal{I}) = |\mathcal{I}| - \max(\mathcal{N}^+(\mathcal{I})_\alpha)$ 给出的下界满足 (3)。

上述构造具有很强的通用性，适用于任何有效的局部检验。具体而言，这意味着，如果 p 值之间存在依赖关系，只要

所用的局部检验对于每个正交象限假设中汇总的 n p 值之间的依赖关系都有效，就可以计算置信集。有关 p 值存在依赖关系的示例，请参见 § C。

接下来，我们将证明，把 DCT 中使用的局部检验映射为用于检验正交象限假设的局部检验，可以在整体上优于 DCT。

4.2 对 $n^+(\mathcal{I})$ 和 $n^-(\mathcal{I}) + n^0(\mathcal{I})$ 的界限，相较于 DCT 的均匀改进

我们可以使用 DCT 局部检验（下文称为自适应局部检验）来检验正交象限假设 $J_{\mathcal{K}}$，方法如下：不检验这个交集假设，而是检验一个更大的交集假设，该假设仅合并来自 $\mathcal{S}^-$ 和 $\mathcal{S}^+$ 的 p 值；也就是说，只检验 DCT 程序第一步中选出的单侧假设的交集。这个更大的交集假设的索引属于 $\{\mathcal{K}^c \cap \mathcal{S}^-\} \cup \{\mathcal{K} \cap \mathcal{S}^+\}$。自适应局部检验为

$$\begin{aligned} \psi_{\mathcal{K}} &= \phi_{\{\mathcal{K}^c \cap \mathcal{S}^-\} \cup \{\mathcal{K} \cap \mathcal{S}^+\}} \\ &= \mathbb{1}\{f(\{2p_i, i \in \mathcal{K}^c \cap \mathcal{S}^-\} \cup \{2q_j, j \in \mathcal{K} \cap \mathcal{S}^+\}) \leq \alpha\}. \end{aligned} \tag{14}$$

这是 $J_{\mathcal{K}}$ 的有效局部检验，因为

$$\Big\{\bigcap_{i \in \mathcal{K}^c} H_i^-\Big\} \cap \Big\{\bigcap_{j \in \mathcal{K}} K_j\Big\} \subseteq \Big\{\bigcap_{i \in \mathcal{K}^c \cap \mathcal{S}^-} H_i^-\Big\} \cap \Big\{\bigcap_{j \in \mathcal{K} \cap \mathcal{S}^+} H_j^+\Big\}.$$

这是检验正交象限假设的良好局部检验吗？Zhao 等人（2019）指出，对于一个仅合并低于某个阈值的条件 p 值的交集假设检验，其检验效能性质非常出色。由于在检验每个正交象限假设时，我们仅合并数据支持单侧备择假设的条件 p 值，因此预计自适应局部检验会具有 Zhao 等人（2019）所观察到的检验效能优势。具体而言，当参数向量具有混合符号时，我们预计其检验效能优于合并交集中所有 p 值的非自适应局部检验；而当参数向量不具有混合符号时，其检验效能也不会有太大损失。

然而，计算问题依然存在：直接应用（13）需要指数时间，因为恰有 v 个正参数的正交象限共有 $\binom{|\mathcal{I}|}{v}$ 个。对于满足条件（$A0$）的合并函数 $f(\cdot)$，可以通过先对以下各集合中的条件 p 值排序，将计算时间降至多项式级别： $\mathcal{S}^- \cap \mathcal{I}$、$\mathcal{S}^+ \cap \mathcal{I}$、 $\mathcal{S}^- \cap \mathcal{I}^c$、 $\mathcal{S}^+ \cap \mathcal{I}^c$。考虑检验 $n^+(\mathcal{I}) = v$ 且恰有 k 个正参数属于 $\mathcal{S}^+ \cap \mathcal{I}$ 的问题。由于 $f(\cdot)$ 是单调的，具有最大

p 值的正交象限是合并以下各项的正交象限：$\mathcal{I}\cap\mathcal{S}^-$ 中最大的 $|\mathcal{I}\cap\mathcal{S}^-|-(v-k)$ 个 $2p_i$；$\mathcal{I}\cap\mathcal{S}^+$ 中最大的 k 个 $2q_i$；$\mathcal{I}^c\cap\mathcal{S}^-$ 中介于 0 到最大的 $|\mathcal{I}^c\cap\mathcal{S}^-|$ 个 $2p_i$；$\mathcal{I}^c\cap\mathcal{S}^+$ 中介于 0 到最大的 $|\mathcal{I}^c\cap\mathcal{S}^+|$ 个 $2q_i$。因此，我们只需考虑 $|\mathcal{S}^-\cap\mathcal{I}^c|\times|\mathcal{S}^+\cap\mathcal{I}^c|$ 种可能的组合，就能找到最大的合并 p 值。考虑 k 和 v 的所有可能取值，可知复杂度为 $\mathcal{O}(|\mathcal{I}|^2|\mathcal{I}^c|^2)$。详细算法见附录 D。

命题 4.2。设 $p_1,\dots,p_n$ 满足条件 $(A1)-(A2)$，且组合函数 $f(\cdot)$ 满足条件 $(A0)$。那么，使用自适应局部检验 (14)，分割程序 (AP) 将分别为 $n^+(\mathcal{I})$ 和 $n^-(\mathcal{I})+n^0(\mathcal{I})$ 返回下界 $\tilde{\ell}^+_\alpha(\mathcal{I})$ 和 $\tilde{\ell}^-_\alpha(\mathcal{I})$，计算量至多为 $\mathcal{O}(|\mathcal{I}|^2\cdot\max(1,|\mathcal{I}^c|^2))$。这些界限分别满足以下条件覆盖保证（以符号向量 $\mathcal{S}$ 为条件）和无条件覆盖保证：

$$\mathbb{P}_\theta\Big(n^+(\mathcal{I})\ge\tilde{\ell}^+_\alpha(\mathcal{I}),\ n^-(\mathcal{I})+n^0(\mathcal{I})\ge\tilde{\ell}^-_\alpha(\mathcal{I}),\ \text{for all }\mathcal{I}\mid\mathcal{S}\Big)\ge 1-\alpha, \tag{15}$$

$$\mathbb{P}_\theta\Big(n^+(\mathcal{I})\ge\tilde{\ell}^+_\alpha(\mathcal{I}),\ n^-(\mathcal{I})+n^0(\mathcal{I})\ge\tilde{\ell}^-_\alpha(\mathcal{I}),\ \text{for all }\mathcal{I}\Big)\ge 1-(1-2^{-n})\alpha>1-\alpha. \tag{16}$$

条件 $(1-\alpha)$ 置信界，即满足 (15) 的界限，具有无条件 $1-\alpha(1-2^{-n})$ 置信保证，即满足 (16)。因此，如果不需要条件保证，则可以在显著性水平 $\tilde{\alpha}=\alpha/(1-2^{-n})$ 下实施带分割的自适应局部检验。使用 $\tilde{\alpha}$ 属于 Goeman 和 Solari（2022）的整体方法框架；该方法将选择和条件化视为实现有用推断并提供无条件错误保证的手段。

以（$1-\alpha$）置信度，$\tilde{\ell}^+_{\tilde{\alpha}}(\mathcal{I})\le n^+(\mathcal{I})\le|\mathcal{I}|-\tilde{\ell}^-_{\tilde{\alpha}}(\mathcal{I})$。这些界限比置信度为（$1-\tilde{\alpha}$）时更紧。具体而言，当 $f(\{2p_i,i\in\mathcal{S}^-\})\le\tilde{\alpha}$ 或 $f(\{2q_j,j\in\mathcal{S}^+\})\le\tilde{\alpha}$ 时，n^+ 或 n^-+n^0 的总体下界分别为正。相比之下，DCT 为校正多重性而进行了更多检验，而且局部检验的显著性水平为 $\alpha<\tilde{\alpha}$。因此，使用 DCT 时，检测 n^+ 和 n^-+n^0 的正下界的检验效能较低。使用分割法在显著性水平为 α 时进行家族错误率控制所发现的数量也可能多于 DCT，但这仅仅是因为其显著性水平为 $\tilde{\alpha}>\alpha$。多重性校正实质上相同，因为只有当所有包含 $i\in[n]$ 的自适应局部检验都被拒绝时，分割法才会将其判定为发现；而这些检验恰好就是针对所有 $\mathcal{I}\ni i$ 的局部检验。基于同样的理由，对于 $n^+(\mathcal{I})$，若 $\mathcal{I}\subseteq\mathcal{S}^-$，以及对于 $n^-(\mathcal{I})+n^0(\mathcal{I})$，若 $\mathcal{I}\subseteq\mathcal{S}^+$，在使用分割法和 DCT 且局部检验的显著性水平均为 α 时，下界相同。

表 2：对于每个正交象限假设（行），当实现的符号向量为 $\mathcal{S} = (sign(p_1 - 1/2)\ sign(p_2 - 1/2))$ 时，$n = 2$ 所对应的分割程序所使用的自适应局部检验（列）。

	$\mathcal{S} = (-1, -1)$	$\mathcal{S} = (-1, 1)$	$\mathcal{S} = (1, -1)$	$\mathcal{S} = (1, 1)$
$H_1^- \cap H_2^-$	$f(2p_1, 2p_2)$	$2p_1$	$2p_2$	1
$H_1^- \cap K_2$	$2p_1$	$f(2p_1, 2q_2)$	1	$2q_2$
$K_1 \cap H_2^-$	$2p_2$	1	$f(2q_1, 2p_2)$	$2q_1$
$K_1 \cap K_2$	1	$2q_2$	$2q_1$	$f(2q_1, 2q_2)$

命题 4.3。对于满足 $(A0)$ 的给定合并函数，令 $\ell_\alpha^+(\cdot)$ $\ell_\alpha^-(\cdot)$ 为满足 (11) 的 ˜ ˜DCT 下界，令 $\ell_\alpha^+(\cdot)$ $\ell_\alpha^-(\cdot)$ 为应用 (14) 中的 ˜ ˜显著性水平 α 自适应局部检验所提取的界限。则 $\ell_\alpha^+(\cdot) \geq \ell_\alpha^+(\cdot)$，且 $\ell_\alpha^-(\cdot) \geq \ell_\alpha^-(\cdot)$。此外，˜ ˜ ˜若 $\mathcal{I} \subseteq \mathcal{S}^-$，则 $\ell_\alpha^+(\mathcal{I}) = \ell_\alpha^+(\mathcal{I})$ 且 $\ell_\alpha^-(\mathcal{I}) = \ell_\alpha^-(\mathcal{I}) = 0$；若 $\mathcal{I} \subseteq \mathcal{S}^+$，则 $\ell_\alpha^-(\mathcal{I}) = \ell_\alpha^-(\mathcal{I})$ 且 $\ell_\alpha^+(\mathcal{I}) = \ell_\alpha^+(\mathcal{I}) = 0$。

4.3 n = 2 时的推断

为了加深对采用自适应局部检验的分割程序的理解，我们考察 $n = 2$ 的基本情形。表 2 展示了所用的自适应局部检验如何随选择事件而变化，而图 1 左上方的图则显示了使用 Simes 合并函数 f_{Simes} 对 n^+ 进行推断的区域。

通过考察图 1 的顶行，我们可以清楚地看到，与 DCT（右上图）相比，采用自适应局部检验的分割程序（左上图）具有明显的功效增益。一个值得注意之处是，分割区域在左上和右下象限显著扩大，从而增强了分割程序检测正效应和非正效应的能力。性能的提升还伴随着一个优势：该程序以更高的显著性水平运行，这由$\tilde{\alpha} = (1 - (1 - 2^{-n}))$ $\alpha = 4\alpha/3$确定。功效的提升也伴随着一项限制，即分割程序只能对非正效应作出推断。例如，如果 $p_1 > 1 - \frac{\alpha}{4}$，分割程序会推断 $\theta_1 \leq 0$。相比之下，DCT程序可以给出更明确的结论，推断 $\theta_1 < 0$。

图 1 左下方的图展示了基假设 H_1^-、 H_2^-、 K_1 和 K_2 的拒绝域。值得注意的是，如果我们观察到 (p_1, p_2)，且满足 $1/2 < p_1 < 1 - 2\alpha/3$

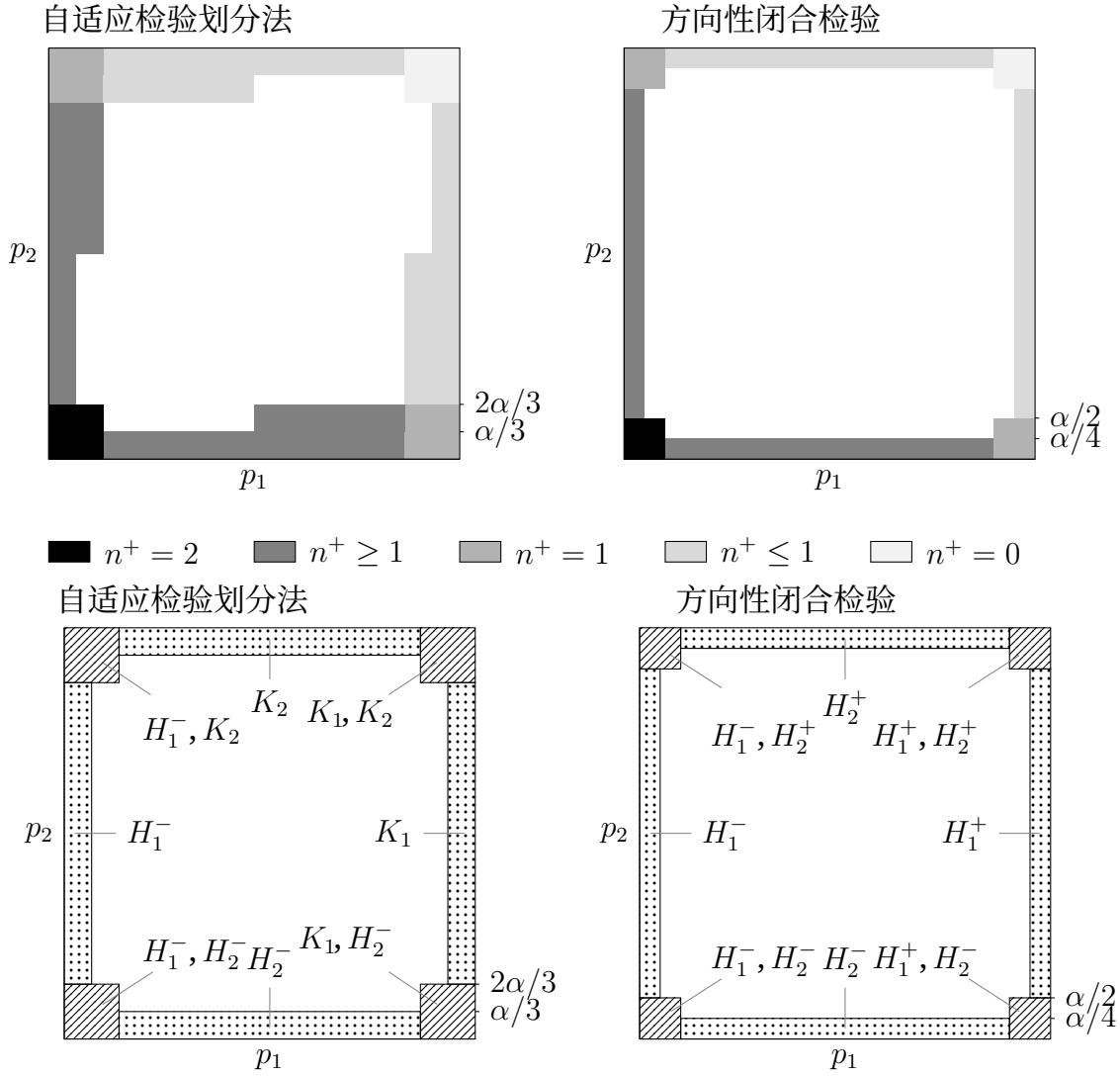


图 1：上排展示了自适应局部检验划分程序（左上图）和采用 Simes 合并函数的方向性闭合检验程序（右上图）中，对 n^+ 进行（$1-\alpha$）置信度推断所对应的区域。灰度表示推断的类型。下排展示了自适应局部检验划分法（左下图）和方向性闭合检验（右下图）中导致拒绝基假设的区域。图案表示在显著性水平 α下拒绝的假设数量：一个（点）或两个（斜线）。每幅图均基于$\alpha = 0.2$。

对于 $\alpha/3 < p_2 < 2\alpha/3$，即使我们能够推断出 $n^+ \geq 1$，也不会拒绝任何基假设。相比之下，在采用 Simes 合并函数的 DCT 中（右下图），每当我们能够对 n^+ 作出有信息量的推断时，都会同时拒绝基假设 H_1^-、H_2^-、K_1 和 K_2。对于 $n > 2$，情况并非如此；使用 Fisher 合并函数处理 $n = 2$ 时也是如此（有关此点以及与其他 FWER 控制程序的比较，请参见 § B）。图 1 左图中的区域形状是非单调的，也就是说，一对较大的 p 值可能会带来更大的 n^+ 下界。例如，若我们观察到 (p_1, p_2)，且 $1/2 < p_1 < 1-2\alpha/3$ 和 $\alpha/3 < p_2 < 2\alpha/3$，那么我们会通过拒绝 $H_1^- \cap H_2^-$ 得出 $n^+ \geq 1$，因为 $2p_2 \leq \tilde{\alpha}$（见表 2 第 3 列）。然而，如果 p_1 降至 $\alpha/3 < p_1 < 1/2$，我们就不会作出任何推断。这个结果相当违反直觉，因为 p_1 值越小，支持 $\theta_1 > 0$ 的证据就越强。这是采用自适应检验而非非自适应检验，也就是进行选择的结果。尽管如此，该程序仍颇具吸引力，因为在给定选择（即要合并哪些单侧假设的 p 值）的条件下，拒绝域是单调的。

图 2 展示了在一系列参数 (θ_1, θ_2) 和 $\alpha = 5\%$ 下，分割法相较于 DCT 的检验效能优势。p 值由服从高斯分布 $N(\theta_i, 1)$ 的独立检验统计量 θ_i 得出，$i = 1, 2$。非平凡界概率（左图）指 (p_1, p_2) 落在图 1 上方图中非白色区域的概率。预期发现数（右图）是图 1 下方图中所示拒绝次数的加权平均值，权重由相应的概率决定。增益显著，尤其是在参数符号相反时，非平凡界概率的提升更为明显。例如，当 $(\theta_1, \theta_2) = (2, -1.5)$ 时，DCT 的概率为 54.1%，而分割法为 71.6%，相差 17.5%。这一差值几乎是两种效应均为正时观测到的 $61.7\% - 55.1\% = 6.6\%$ 的三倍；此时 $(\theta_1, \theta_2) = (2, 1.5)$。

5 应用

5.1 亚组分析

在亚组分析中，通常会采用双侧检验，以判断各亚组中的干预效应。一个更简单但仍有信息价值的目标，是为 n^+ 以及 n^- 或 $n^- + n^0$ 提供下界。Fisher 等人（1983）的数据（由 Gail 和 Simon（1985年）重新分析）包括接受 PFT 治疗（L-苯丙氨酸

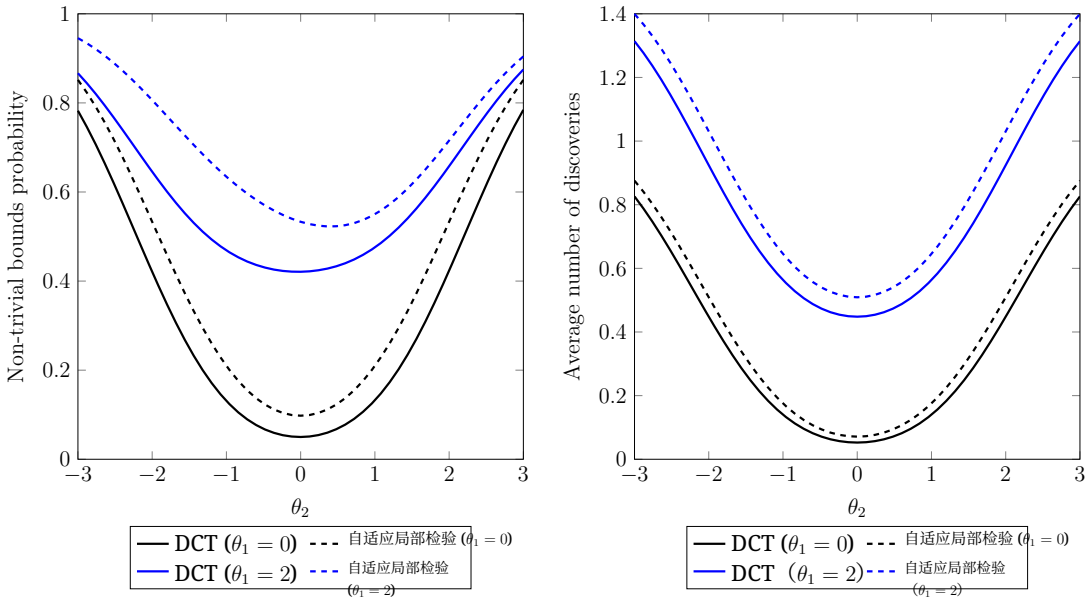


图 2：观测到非平凡界限的概率（左图）和预期发现数（右图）随 θ_2 的变化。置信水平为 95%，采用方向性闭合检验（DCT，实线）和带分割的自适应局部检验（自适应局部检验，虚线）。黑色曲线对应 $\theta_1 = 0$，蓝色曲线对应 $\theta_1 = 2$。

氮芥、5-氟尿嘧啶和他莫昔芬）与接受 PF 治疗（L-苯丙氨酸氮芥和 5-氟尿嘧啶）的患者进行比较。该研究共纳入 1260 名患者，分为 $n = 4$ 个亚组，并根据年龄和孕酮受体水平进行分类。设 m_i^j 为第 i个亚组中接受治疗 $j \in \{\text{PFT、PF}\}$ 的样本量。表3列出了 $\hat{}\pi_i^j$，即 i亚组中接受治疗 j 的患者在 3 年时的无病生存概率 π_i^j 的 Kaplan-Meier 估计值。相应的标准误 SE($\hat{}\pi_i^j$) 根据 Greenwood公式（Greenwood等人，1926年）计算。

然而，变量 $\sqrt{m_i^j}(\hat{}\pi_i^j - \pi_i^j)$ 的分布收敛于一个均值为 0、方差取决于（未知）生存概率 π_i^j 的正态变量；若不存在删失，其方差可简化为 $\pi_i^j(1 - \pi_i^j)$。对于删失数据，Cutler和Ederer（1958年）建议用有效样本量 $\tilde{m}_i^j = \hat{}\pi_i^j(1 - \hat{}\pi_i^j)/\text{SE}^2(\hat{}\pi_i^j)$ 替代 m_i^j。Anderson等人（1982年）证明，$\sqrt{\tilde{m}_i^j}(\hat{}\pi_i^j - \pi_i^j)$ 的分布收敛于 $N(0, \pi_i^j(1 - \pi_i^j))$。这为反正弦平方根变换 $g(\hat{}\pi_i^j) = \arcsin\sqrt{\hat{}\pi_i^j}$ 提供了依据：根据德尔塔法，$\sqrt{\tilde{m}_i^j}\{g(\hat{}\pi_i^j) - g(\pi_i^j)\}$ 的渐近方差为 $1/4$，因此该变换能有效稳定方差。随后，可利用标准化差异 $\hat{}\theta_i = \{g(\hat{}\pi_i^{\text{PFT}}) - g(\hat{}\pi_i^{\text{PF}})\}/\sqrt{1/(4\tilde{m}_i^{\text{PFT}}) + 1/(4\tilde{m}_i^{\text{PF}})}$ 的大样本正态近似，计算 H_i^- 的 $\hat{}$ p 值 $p_i = 1 - \Phi(\theta_i)$，其中 $\Phi(\cdot)$ 是 $N(0, 1)$ 的累积分布函数（ Klein 等人， 2007 年）。

根据数据支持的方向应用假设选择的第一步，DCT程序选择了 $H_1^-, H_2^+, H_3^+, H_4^+$。因此，第二步检验所用的 p 值分别为 $0.0202, 1 - 0.9490, 1 - 0.7774$ 和 $1 - 0.9966$，对应第 1、2、3 和 4 组。该程序如图3上方面板所示。

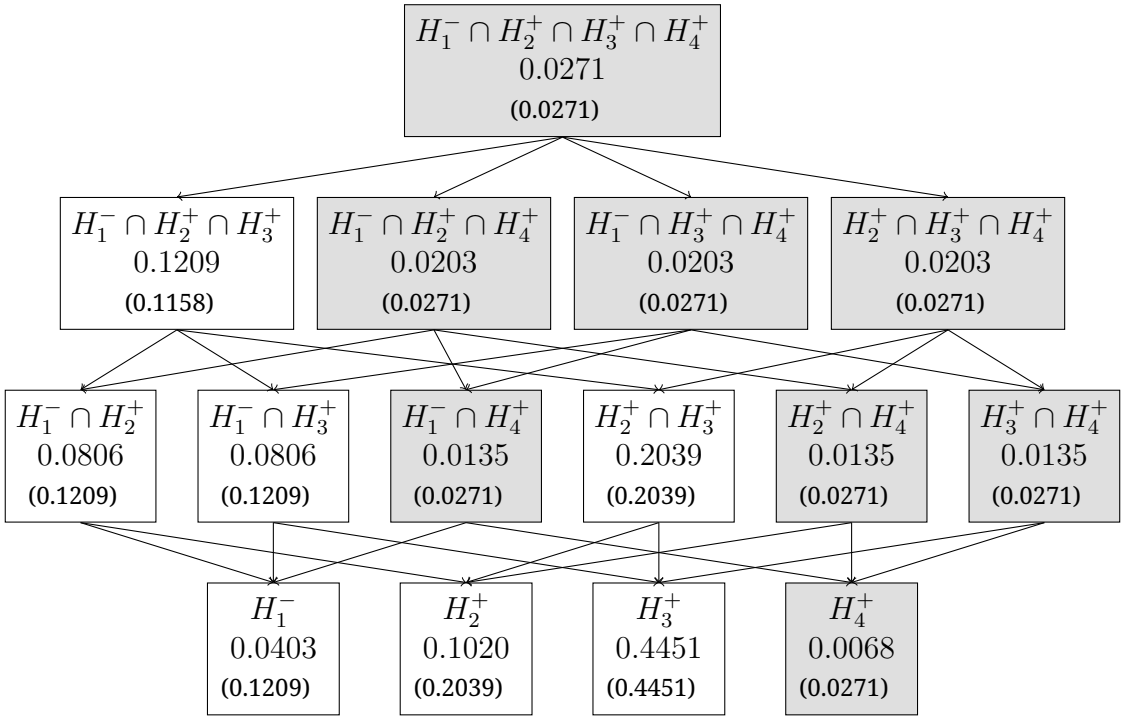
在显著性水平 $\alpha = 0.05$ 下应用 DCT，我们以 95% 置信度得到 $0 \le n^+$ 和 $1 \le n^-$。此外，我们得出结论 $\theta_4 < 0$，因为闭合检验拒绝了 H_4^+。由于闭合检验未拒绝 H_1^-，我们得到 n^+ 的平凡下界；由于闭合检验拒绝了 $H_2^+ \cap H_3^+ \cap H_4^+$，我们得到 n^- 的非平凡下界。

采用相同的带分割自适应局部检验的程序如图 4 所示。我们可以从这些未被拒绝的正交象限假设中提取 n^+ 的置信集，如下所示。由于每个未被拒绝的正交象限假设中正参数的数量至少为 1，至多为 3，因此界限为 $1 \le n^+ \le 3$，置信度为 95%。这优于使用 DCT 得到的 n^+ 界限。或者，我们也可以看到，下界为 1，因为 $H_1^- \cap H_2^- \cap H_3^- \cap H_4^-$ 被拒绝了（通过对 H_1^-进行局部检验，因为其余坐标 K_2、K_3、K_4 被选中进行检验）。此外，由于每个未被拒绝的分割零假设中的第四个单独假设均为非正，

表3：接受治疗 j治疗（j =PFT、PF）且属于由年龄和孕酮受体（PR）水平定义的亚组 i 的乳腺癌患者，在 3 年时的无病生存概率分析 $\hat{\pi}_i^j$ （$i = 1$：年龄 < 50、PR < 10；$i = 2$：年龄 ≥ 50、PR < 10；$i = 3$：年龄 < 50、PR ≥ 10；$i = 4$：年龄 ≥ 50、PR $\geq$10）。转载自 Gail 和 Simon（1985年）的表 2。最后两行分别为 PFT 与 PF 治疗组在 3 年时无病生存概率（经反正弦平方根变换）的（标准化）差值，以及使用大样本正态近似计算的 H_i^- 的 p 值。

亚组 i	1		2		3		4	
治疗 j	PFT	PF	PFT	PF	PFT	PF	PFT	PF
$\hat{\pi}_i^j$	.599	.436	.526	.639	.651	.698	.639	.790
SE($\hat{\pi}_i^j$)	.0542	.0572	.0510	.0463	.0431	.0438	.0386	.0387
$\hat{\theta}_i$	2.051		-1.635		-0.764		-2.708	
p_i	.0202		.9490		.7774		.9966	

程序 3.1（方向性闭合检验）



程序 5.1（定性交互的闭合检验）

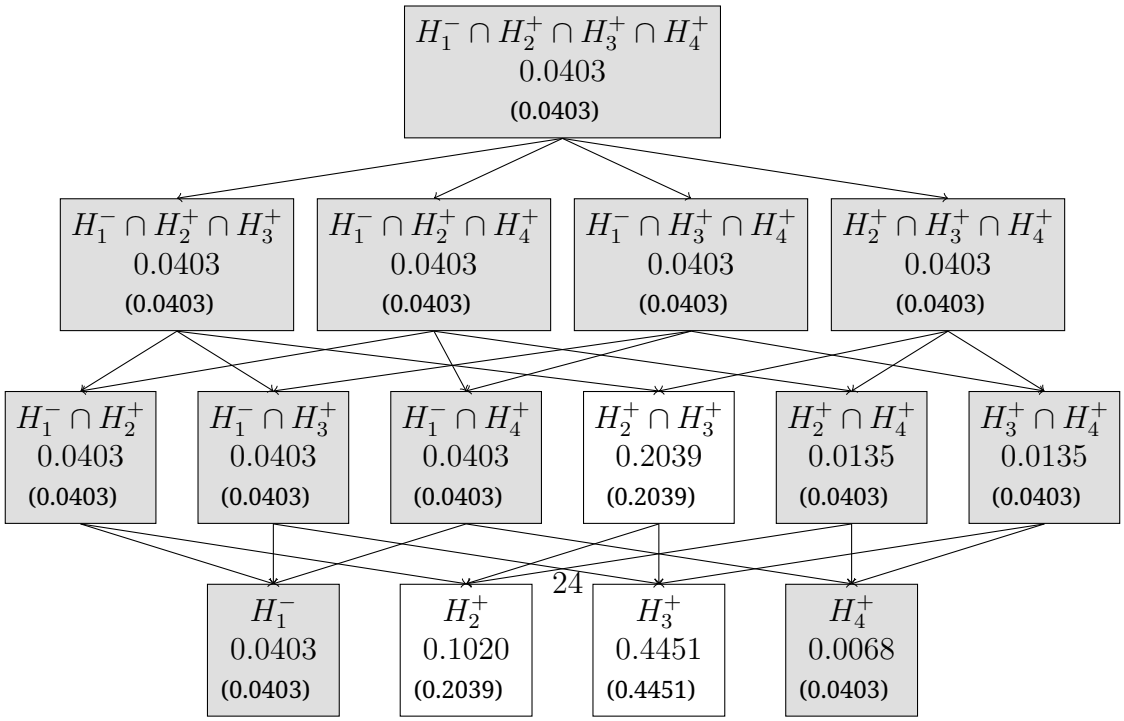


图3：表3中的示例。上图：采用 Simes 局部检验的程序 3.1（方向性闭合检验）。下图：针对定性相互作用的程序 5.1（闭合检验），以及针对 $\{H_1^-\} \cup \{H_2^+ \cap H_3^+ \cap H_4^+\}$ 的 Zhao 等人（2019）检验。每个节点代表一个闭合检验交集假设，箭头表示子集关系。每个交集假设下方标出了相应的 p-值和闭合检验调整后的p-值（括号内）。灰色节点表示被 5%水平闭合检验程序拒绝的假设。

我们进一步得出结论：$\theta_4 \leq 0$。与 DCT 得出的结论 $\theta_4 < 0$ 相比，这是一个较弱的结论。

5.2 检验 并进行后续推断，以检验定性交互作用 e

如果主要关注的是检验参数向量 θ 是否具有混合符号，则可以调整 § 5.3 中描述的方法，仅在拒绝无定性交互作用假设 H_0：$\{n^+ = 0\} \cup \{n^- = 0\}$ 后才应用 DCT。Gail 和 Simon（1985年）分析了表3中的数据，以判断能否在 $\alpha = 5\%$ 显著性水平下证明存在定性交互作用。然而，Gail 和 Simon（1985年）提出的似然比检验所得的 p 值为 0.0877。

Zhao 等人（2019年）的检验采用选择阈值 $\tau = 1/2$，当 $H_{\mathcal{S}^-}$ 被 $f(\{2p_i, i \in \mathcal{S}^-\})$ 拒绝且 $H_{\mathcal{S}^+}$ 被 $f(\{2q_i, i \in \mathcal{S}^+\})$ 拒绝时，拒绝 H_0：$\{n^+ = 0\} \cup \{n^- = 0\}$。这是一个有效的检验，因为 $\{n^+ = 0\} \subseteq H_{\mathcal{S}^-}$ 和 $\{n^- = 0\} \subseteq H_{\mathcal{S}^+}$。使用 Simes 组合检验，$p$ 值为 max(0.0403, 0.0203) $= 0.0403$。

我们提出以下针对定性交互作用的程序：首先对 H_0 进行 Zhao 等人（2019）的全局检验；如果拒绝该假设，则继续使用 DCT程序。

程序 5.1（定性交互的闭合检验）。

步骤 1 应用程序 2.1 的步骤 1

步骤 2 对假设族 $\{H_i^- : p_i \leq 1/2\} \cup \{H_i^+ : p_i > 1/2\}$，应用显著性水平为 α 的闭合检验程序。以显著性水平 α 检验 $H_{\mathcal{I}}$，使用局部检验

$$\phi_{\mathcal{I}} = \begin{cases} \mathbb{1}\{\max\{f(\{2p_i, i \in \mathcal{S}^-\}), f(\{2q_i, i \in \mathcal{S}^+\})\} \leq \alpha\} & \text{if } \mathcal{I} \supseteq \mathcal{S}^- \text{ or } \mathcal{I} \supseteq \mathcal{S}^+ \\ \mathbb{1}\{f(\{2p_i, i \in \mathcal{I} \cap \mathcal{S}^-\} \cup \{2q_j, j \in \mathcal{I} \cap \mathcal{S}^+\}) \leq \alpha\} & \text{otherwise} \end{cases}$$

如果拒绝了 H_0，那么 n^+ 和 n^- 的下界都具有信息性，即 $n^+ >$0，$n^- > 0$。从图3的底部图中可以看到，在5%的显著性水平下，程序 5.1拒绝了除 $H_2^+ \cap H_3^+$、H_2^+和 H_3^+之外的所有假设，得到 $1 \leq n^+ \leq 3$、$1 \leq n^- \leq 3$ 和 $\theta_1 > 0$、$\theta_4 < 0$。这一结果优于DCT程序 3.1得到的结果（$1 \leq n^- \leq 4$和 $\theta_4 < 0$）。然而，对于程序 5.1而言，拒绝 H_0 至关重要：如果不拒绝 H_0，闭合检验程序 5.1不会拒绝任何交集假设。例如，在 $\alpha = 3\%$下，程序 3.1给出的结论与 $\alpha = 5\%$下相同，而程序 5.1没有信息性，因为没有作出任何拒绝。

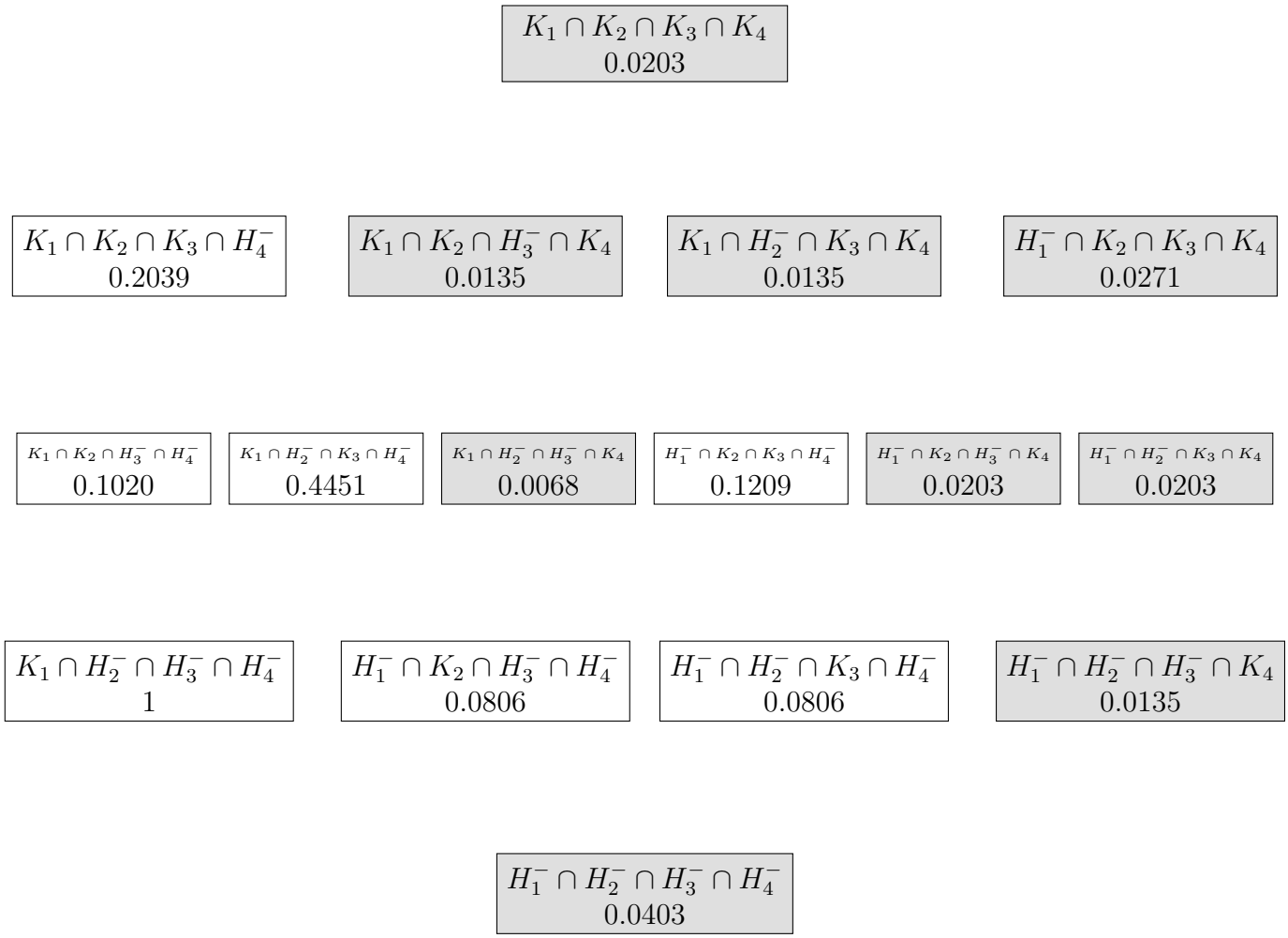


图4：表3中的示例。每个方框代表一个正交象限假设。在每个正交象限假设下方，是通过Simes局部检验得到的相应 p值。灰色方框表示在5%的显著性水平下被分割程序拒绝的正交象限假设。

请注意，采用自适应检验的分割程序可以使用程序 5.1 的局部检验，从而得到

$$\psi_{\mathcal{K}} = \begin{cases} \mathbb{1}\{\max\{f(\{2p_i, i \in \mathcal{S}^-\}), f(\{2q_i, i \in \mathcal{S}^+\})\} \leq \alpha\} & \text{if } \mathcal{K} \subseteq \mathcal{S}^- \text{ or } \mathcal{K} \subseteq \mathcal{S}^+ \\ \mathbb{1}\{f(\{2p_i, i \in \mathcal{K}^c \cap \mathcal{S}^-\} \cup \{2q_j, j \in \mathcal{K} \cap \mathcal{S}^+\}) \leq \alpha\} & \text{otherwise} \end{cases}$$

将该程序应用于表3中的数据，在显著性水平为 $\alpha = 5\%$ 时，得出的结论与在相同显著性水平下使用程序 5.1 得出的结论相同。

5.3 增强荟萃分析

荟萃分析的第一步通常是计算全局零假设的 p 值，以传达反对“所有研究中均不存在关联”这一零假设的证据强度。拒绝全局零假设并不意味着排除了这样的可能性：结果仅由单项研究中的非零关联所致（例如，由于该研究队列的具体情况或偏倚）。因此，人们会关注对具有正向效应的研究数量给出紧确的下界和上界。此外，对于非平凡的界限，自然也会尝试找出存在关联的研究。如果需要对正向和负向关联进行推断，则应使用 DCT。如果只需推断 n^+ 的下界和上界，并识别正向和非正向关联，则应使用带分割的自适应局部检验。由于典型的荟萃分析中各项研究相互独立，且每项研究中的检验统计量近似服从正态分布，因此所需条件 $(A1) - (A2)$ 得以满足。

举一个具体例子，我们来考察 Cooper 等人（2003）的数据，该数据由 Konstantopoulos（2011年）公开，其中研究了改良校历（假期更频繁但时间更短）对学生学业成绩的影响。该荟萃分析使用了11个学区中 $n = 56$ 所学校的研究。实验设计为两组比较（改良校历与传统校历），需要计算标准化均值差 $y_i \sim N(\theta_i, 1)$，其中 $\theta_i > 0$ 表示正效应，即采用改良校历的组学业成绩更高。Zhao 等人（2019）表明，在 5%的显著性水平下，有证据支持存在定性交互作用，即改良校历对学生学业成绩的影响在至少一项研究中为正，在至少另一项研究中为负。

表4展示了我们的分析结果。使用 DCT 时，n^+ 的上界是平凡界限 $n = 56$。带分割的自适应局部检验给出了信息量大得多的界限：使用 Fisher 合并法时，n^+ 的 95%置信界为 10 至 53，并且我们有四项个体发现；使用 mSimes 时，n^+ 的 95%置信界为

表4：改良校历对学生学业成绩的影响：使用不同合并方法和程序时，n^+和 n^- 的区间以及相应的发现数量 $|\mathcal{D}^+|$ 和 $|\mathcal{D}^-|$，置信度为 95%。结果包括程序 3.1（DCT）、带分割的自适应局部检验（AP），以及 Guo 和 Romano（2015）中用于家族错误率和错误发现率控制的程序（最后两行）。对于 DCT，$|\mathcal{D}^-|$ 是负参数的发现数量；对于其他方法，则是非正参数的发现数量。

		n^+			$n^- \setminus n^- + n^0$		
局部检验	程序	下界	上界	$\lvert\mathcal{D}^+\rvert$	下界	上界	$\lvert\mathcal{D}^-\rvert$
Fisher	DCT	9	56	4	0	47	0
	AP	10	53		3	46	
Simes	DCT	8	56	8	0	48	0
	AP	8	55		1	48	
修正 Simes	DCT	8	56	8	0	48	0
	AP	8	54		2	48	
自适应 LRT	DCT	9	56	5	0	47	0
	AP	11	53		3	45	
Guo-Romano	FWER	8	55	8	1	48	1
	FDR	12	52	12	4	44	4

8 至 54，并且我们有八项个体发现。这些界限比按各方向的发现数量计算出的界限更窄，因为即使基假设未被拒绝，用于确定界限的局部检验也可能拒绝交集假设。使用 Guo 和 Romano （ 2015 ）针对家族错误率控制设计的程序，n^+ 的 95% 置信界为 8 至 55，且有九项个体发现。使用 Guo 和 Romano（2015）针对错误发现率控制设计的程序，我们得到了 n^+ 最窄的界限，即 12 至 52。不过，这些并非 95%置信界，因为该程序不提供同时置信保证。

使用 Fisher 合并函数，可获得 $n^+(\mathcal{I}_k)$ 关于 $k = 1, \ldots, n$ 的事后界，其中 $\mathcal{I}_k$ 表示估计效应 $|y_i|$（按绝对值）最大的前 $k = |\mathcal{I}_k|$ 所学校的索引。

6 模拟

我们开展模拟研究，以评估所提出方法的性能。我们主要关注 DCT程序 3.1 和带分区的自适应局部检验（AP）程序，并考察几种可能有用的不同合并方法。命题 4.3 表明，AP 的界限至少与 DCT 的界限一样紧。我们旨在量化界限收紧所带来的潜在收益。为此，我们采用两种具有代表性的合并方法：Fisher（一种基于求和的方法）和修正 Simes（8）（一种基于分位数的方法）。基于求和的检验（如 Fisher）在界限方面表现更好，但在发现方面表现更差。在补充材料中，我们还考察了自适应似然比检验（Al Mohamad 等人，2020），这是一种针对正态检验统计量的求和型检验；此外还考察了 Simes—— 在我们的模拟中，无论是界限还是（方向性）决策，其表现都不如 mSimes。

我们将所提出的程序与 Guo 和 Romano（2015）中的家族错误率和错误发现率控制程序进行比较，这些程序已在 § 1.1 中介绍。这些程序旨在识别正参数和非正参数，因此能够控制方向性错误率。我们通过统计所声称为正向发现和非正向发现的发现数量，分别从这些程序中提取 $n^+(\mathcal{I})$ 和 $n^-(\mathcal{I}) + n^0(\mathcal{I})$ 的下界。家族错误率控制程序对于发现而言是一个很好的竞争者，但在下界方面表现不佳。错误发现率控制程序的相关性较低，因为与所考虑的其他所有程序不同，其正向发现或非正向发现的数量超过发现集合中真实正参数和非正参数数量的概率可能远大于 α，参见补充材料中的图 10。然而，我们仍将其纳入比较，因为有趣的是，尽管没有置信度保证，它在下界方面未必更好。

我们的数据生成机制如下。总共有 $n = 50$ 个参数，其中有 $n^+ \in \{0, \ldots, n\}$ 个正参数和 $n^- \in \{0, \ldots, n - n^+\}$ 个负参数。其余参数均为零。检验统计量独立地取自标准差为 1、均值以参数估计值为中心的高斯分布 $\theta_i \sim N(\theta_i, 1)$。$H_i^-$ 的单侧 p 值为 $p_i = 1 - \Phi(\theta_i)$ $i = 1$，....., n。我们的分析基于 $B = 2000$ 个数据生成机制，以及 $\alpha = 0.05$。

我们在图 5 和补充图 11 中展示了各程序所得正参数和非正参数数量下界之和的平均值（界限的平均长度为 n减去该平均值），以及发现的参数数量。我们的主要发现如下。首先，AP 给出的界限比 DCT 紧得多。其次，就界限而言，基于求和的检验优于基于分位数的检验，也优于 Guo 和 Romano（2015）的 FWER 控制程序。在信号相当弱的数据生成机制中，它们在界限方面也可能远优于

Guo 和 Romano（2015）的 FDR 控制程序。第三，就发现而言，mSimes 的表现与 Guo 和 Romano（2015）的 FWER 控制程序不相上下，而且二者都远优于基于求和的检验。根据模拟结果，我们可以给出以下一般性建议：如果除了界限之外，发现也很重要，或者预计远离零的参数占比很小，那么 mSimes 是不错的选择。否则，最好采用基于求和的组合方法，例如 Fisher 或 ALRT。

7 结论

在程序 3.1 中，选择步骤 1 的代价是将 p-值放大 2 倍。这是合理的代价，因为它不比双侧检验的代价更高。还可以更一般地考虑选择 $\{H_i^+ : \ p_i \leq \tau\} \cup \{H_i^- : \ p_i > \tau\}$，其中 $\tau \in$ 取值为 (0, 1)。这样的选择步骤会将 p_is 放大 $1/\tau$，并将 q_is 放大 $1/(1-\tau)$。当检测一个方向比检测另一个方向更重要时，这可能会有用。更一般地，每个参数都可以有各自预先设定的选择阈值 τ_i、$i = 1$、……、n。特别地，如果参数 θ_i 的推断方向事先已知，则可以设定 $\tau_i \in \{0, 1\}$。因此，选择具有更一般的形式；方向性闭合检验或分割程序按所述方式进行，使用条件 p-值 $\{p_i/\tau_i : \ p_i \leq \tau_i\} \cup \{q_i/(1-\tau_i) : \ p_i > \tau_i\}$。重要的是，这些条件 p-值也可以使用相同的计算捷径；只要 p_1、……、p_n 相互独立且均匀有效，推断就是有效的。

一个尚待解决的问题是如何处理相关的 p 值。对于 § 3.1 中讨论的非常具体的结构化依赖关系，可以将基于符号向量进行条件化的程序与适当的局部检验结合使用。若不进行条件化，只要局部检验对 p 值之间的特定依赖关系有效，就可以采用 § 4.1 中详述的分割程序。我们在 § C 中给出了一个小例子，并将对相关检验统计量的无条件方法进行更全面的考察留待未来研究。

本研究聚焦于事后界。为简明起见，我们给出的例子规模较小。但原则上，这些方法也可用于中等规模和大规模检验。如 § 6 所示，当信号较弱且稀疏时，这些界限可能比 FDR 控制程序所得的界限更紧。具体而言，我们的程序可能适用于研究独立个体的纵向数据。例如，在神经影像学中，对某一个体进行实验得到的 p 值

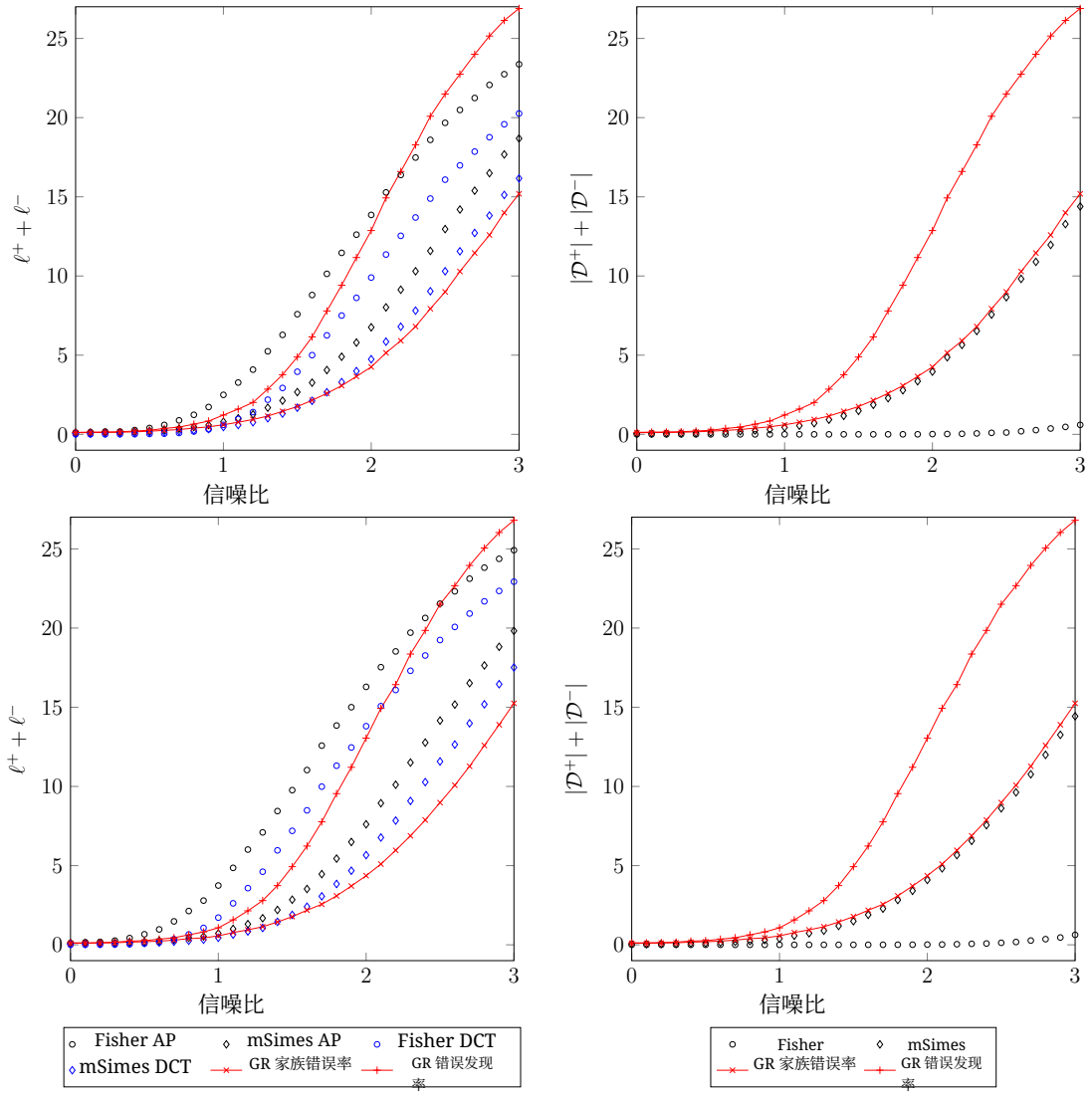


图 5：正参数和非正参数下界之和的平均值（左列，越高越好）以及被拒绝的基假设数量（右列，越高越好）随信噪比（即非零 θ_i 值的绝对值）变化的情况。比较的方法如下：采用 Fisher 合并函数的 AP 和 DCT（AP FISHER 和 DCT Fisher），以及采用 mSimes 合并函数的 AP 和 DCT（AP mSIMES 和 DCT mSimes）；用于家族错误率控制的 Guo-Romano 程序（GR FWER）和用于错误发现率控制的 Guo-Romano 程序（GR FDR）。 $n^+ = n^- = 15$ 在第一行； $n^+ = 30$ 和 $n^- = 0$在第二行。右列中没有DCT程序，是因为它们与 AP程序相同（命题 4.3）。请注意，对于 GR-FWER 和 GR-FDR，正参数和非正参数下界之和等于其发现数量，因此每行各面板中的曲线均相同，31。

脑部感兴趣区域中的结果可能会针对数百名个体进行报告。 $n^+(\mathcal{I})$和 $n^-(\mathcal{I})$或 $n^-(\mathcal{I}) + n^0(\mathcal{I})$的下界可能非常值得关注， $\mathcal{I} = [n]$ 如此，各种 $\mathcal{I} \subset [n]$也如此。

对于大规模检验，已有研究提出使用负对照或掩蔽方法来推断符号。例如，Leung和Tran（2023）（在Lei和Fithian（2018）的研究基础上）建议将单侧假设的 p 值围绕0.25或0.75进行反射。对于错误发现率控制，这些方法基于Barber和Candès（2015）的knockoff+ 方法，因此， $\mathcal{I}$ 中估计的错误发现比例，是 $\mathcal{I}$ 中的负对照估计数量加一，再除以 $|\mathcal{I}|$。分子中额外加一，使这些程序对于小规模检验或非零参数极少的情况不太适用。一个尚待解决的问题是，如何将其推断扩展到关于 $n^+(\mathcal{I})$和 $n^-(\mathcal{I})$的事后同时界。为此，可以考虑在Li等人（2022）的研究基础上继续推进。接下来，比较这两种方法将很有意思：一种是对符号进行条件化，并像本文一样使用调整后的 p值；另一种是对掩蔽后的 p值进行条件化，并采用knockoff方法。

在高维荟萃分析中，研究彼此独立，但每项研究都考察许多（例如 m个）特征。我们能否先筛选特征，再给出同时事后界？例如，对于神经影像学应用，我们该如何处理多个感兴趣区域，并为每个区域给出紧确界限，同时控制相关误差度量？如果从 m 个特征中选出 $|\mathcal{R}|$ 个，并在显著性水平 $|\mathcal{R}|\alpha/m$下构造事后界，那么就能控制一种相关误差。Benjamini 和 Bogomolov（2014）提出，这种误差与所选特征的平均家族错误率有关，即在所有选定特征中，至少出现一次错误拒绝的特征所占比例的期望。简言之，我们考虑的误差是：在选定特征中，至少有一个事后界未能覆盖真实值的比例的期望。未来，我们计划对这一程序及其针对特定和一般依赖关系的理论保证进行详细分析。

参考文献

Al Mohamad, D., Van Zwet, E. W., Cator, E., and Goeman, J. J. (2020). 约束似然比检验的自适应临界值。Biometrika, 107(3):677–688。

Anderson, J. R., Bernstein, L. 和 Pike, M. C.（1982年）。寿命表分析中生存概率和分位数的近似置信区间。《生物计量学》，第407–416页。

Barber, R. F. 和 Cand`es, E. J.（2015）。通过仿冒变量控制错误发现率。《统计年刊》，43(5):2055–2085。Bauer, P., Hackl, P., Hommel, G. 和 Sonnemann, E.（1986年）。对单侧假设对进行多重检验。《计量学》，33(1):121–127。Benjamini, Y.（2010年）。同时推断与选择性推断：当前成果与未来挑战。《生物统计学杂志》，52(6):708–721。Benjamini, Y. 和 Bogomolov, M.（2014）。对多个假设族进行选择性推断。《英国皇家统计学会杂志：B辑（统计方法学）》，76(1):297–318。Benjamini, Y. 和 Heller, R.（2008）。部分合取假设的筛选。《生物计量学》，64(4):1215–1222。Benjamini, Y., Heller, R. 和 Daniel, Y.（2009）。复杂研究中的选择性推断。《英国皇家统计学会杂志：A辑》，367(1906):4255–4271。Benjamini, Y. 和 Yekutieli, D.（2005年）。针对选定参数调整错误发现率的多重置信区间。《美国统计学会会刊》，100(469):71–81。Blanchard, G., Neuvial, P. 和 Roquain, E.（2020）。利用参考族对假阳性进行事后置信界估计。《统计年刊》，48(3):1281–1303。Bogomolov, M.（2023）。依赖情形下的部分合取假设检验及其在荟萃分析中的应用。《电子统计学杂志》，17(1):102–155。Bogomolov, M. 和 Heller, R.（2022年）。多项研究中的可重复性。arXiv:2210.00522。Bohrer, R. 和 Schervish, M. J.（1980年）。针对参数符号的最优多重决策规则。《美国国家科学院院刊》，77(1):52–56。Bretz, F., Hothorn, T. 和 Westfall, P.（2016）。使用 R 进行多重比较。CRC出版社。Cooper, H., Valentine, J. C., Charlton, K. 和 Melson, A.（2003）。调整后的校历对学生学业成绩及学校与社区态度的影响。《教育研究评论》，73(1):1–52。

Cutler, S. J. and Ederer, F.（1958年）。在生存分析中最大限度地利用生命表法。《慢性病杂志》，8(6):699–712。Dobriban, E.（2020）。针对可交换局部检验的快速闭合检验。Biometrika，107(3):761–768。Dunnett, C. W.（1955）。一种将多种治疗与对照进行比较的多重比较程序。《美国统计学会会刊》，50(272):1096–1121。Ellis, J. L., Pecanka, J., and Goeman, J. J.（2020）。通过条件化提高区间假设多重检验的检验效能。《生物统计学》，21(2):e65–e79。Finner, H.（1994）。多重假设检验：一般理论、具体问题及其与其他多重决策程序的关系。教授资格论文，第四数学系，特里尔大学。Finner, H.（1999）。逐步多重检验程序与方向性错误的控制。《统计年刊》，27(1):274–289。Finner, H. and Strassburger, K.（2002）。划分原理：多重决策理论中的有力工具。《统计年刊》，第1194–1213页。Finner, H., Tang, S.-Y., Cui, X., and Hsu, J. C.（2021）。置信集、置信方向和决策路径的分割法。载于《多重比较手册》，第57–90页。Chapman and Hall/CRC。Fisher, B., Redmond, C., Brown, A., Wickerham, D., Wolmark, N., Allegra, J., Escher, G., Lippman, M., Savlov, E., Wittliff, J., Fisher, E., Plotkin, D., Bowman, D., Wolter, J., Bornstein, R., Desser, R., and Frelick, R.（1983）。原发性乳腺癌中肿瘤雌激素受体和孕酮受体水平对他莫昔芬和化疗疗效的影响。《临床肿瘤学杂志》，1(4):227–241。Fisher, R. A.（1934 年）。《研究工作者的统计方法》。第五版。Oliver and Boyd，爱丁堡和伦敦。Gail, M. and Simon, R.（1985）。检验治疗效应与患者子集之间的定性交互作用。《生物计量学》，第361–372页。Goeman, J. and Solari, A.（2022年）。选择性推断的条件方法与无条件方法。预印本，参见 https://arxiv.org/abs/2207.13480。

Goeman, J., Solari, A., and Stijnen, T.（2010）。三侧假设检验：对优效性、等效性和非劣效性进行同时检验。医学统计学，29:2117–2125。Goeman, J. J., Hemerik, J., and Solari, A.（2021）。只有闭合检验程序才适用于控制错误发现比例。《统计年刊》，49(2):1218–1238。Goeman, J. J., Meijer, R. J., Krebs, T. J., and Solari, A.（2019）。大规模多重假设检验中同时控制所有错误发现比例。Biometrika，106(4):841–856。Goeman, J. J. and Solari, A.（2011）。《探索性研究的多重检验》。Statistical Science，26(4):584–597。Greenwood, M. 等（1926年）。关于癌症自然病程的报告。《关于癌症自然病程的报告》，(33)。Guo, W. and Romano, J. P.（2015）。在独立及某些依赖情形下逐步控制方向性错误。统计规划与推断杂志，163:21–33。Hoang, A.-T. and Dickhaus, T.（2022年）。可重复性分析中独立 p 值的合并：一项比较研究。统计计算与模拟杂志。Holm, S.（1979）。一种简单的序贯拒绝式多重检验程序。斯堪的纳维亚统计学杂志，第65–70页。IntHout, J., Ioannidis, J. P. A., Rovers, M. M., and Goeman, J. J.（2016）。呼吁在荟萃分析中常规报告预测区间。BMJ OPEN，6(7)。Jaljuli, I., Benjamini, Y., Shenhav, L., Panagiotou, O. A., and Heller, R.（2022年）。量化系统综述中的可重复性和一致性。生物制药研究统计学，0(0):1–14。Jones, L. and Tukey, J.（2000）。一种合理表述显著性检验的方法。心理学方法，5(4):411–414。Karlin, S. and Rubin, H.（1956年）。单调似然比分布的决策程序理论。《数学统计年刊》，第272–299页。

Klein, J. P., Logan, B., Harhoff, M. 和 Andersen, P. K.（2007年）。在固定时间点分析生存曲线。《医学统计学》，26(24)：4505–4519。Konstantopoulos, S.（2011年）。三级元分析中的固定效应和方差分量估计。《研究综合方法》，2(1)：61–76。Lehmann, E. L. 和 Romano, J. P.（2005年）。检验统计假设。Springer统计学丛书。Springer，第三版。Lei, L. 和 Fithian, W.（2018年）。AdaPT：一种利用辅助信息进行多重检验的交互式程序。《皇家统计学会期刊B辑：统计方法学》，80(4)：649–679。Leung, D. 和 Tran, N.（2023）。方向性错误发现率控制的自适应程序。预印本，见 https://arxiv.org/abs/2201.05828。Li, J., Maathuis, M. 和 Goeman, J.（2022年）。通过仿冒变量和闭合检验获得同时错误发现比例界限。预印本，见 https://arxiv.org/abs/2212.12822。Liu, W.（1997年）。利用逐步上升多重检验控制方向性错误。《统计与概率快报》，31(3)：239–242。Marcus, R., Peritz, E. 和 Gabriel, K. R.（1976年）。闭合检验程序，特别涉及有序方差分析。《Biometrika》，63(3)：655–660。Romano, J. P., Shaikh, A. 和 Wolf, M.（2011年）。多重检验中的一致性与闭合方法。《国际生物统计学杂志》，7(1)。Sarkar, S. K., Sen, P. K. 和 Finner, H.（2004年）。关于多重检验中的两个结果。《讲义专著系列》，第89–99页。Shaffer, J. P.（1974年）。双向无偏程序。《美国统计学会会刊》，69(346)：437–439。Shaffer, J. P.（1980年）。利用分阶段多重检验程序控制方向性错误。《统计年刊》，8(6)：1342–1347。Shaffer, J. P.（1986年）。修正的序贯拒绝式多重检验程序。《美国统计学会会刊》，81(395)：826–831。

Shaffer, J. P.（2002年）。多重性、方向性（III类）错误与零假设。《心理学方法》，7(3)：356。Spjotvoll, E.（1972年）。关于某些多重比较程序的最优性。《数学统计年刊》，43(2)：398–411。Stefansson, G., Kim, W. 和 Hsu, J.（1988年）。关于多重比较中的置信集。《统计决策理论及相关专题 IV》（SS Gupta 和 JB Berger 编），2，89–104。Tian, J., Chen, X., Katsevich, E., Goeman, J. 和 Ramdas, A.（2021年）。依赖性下的大规模同时推断。arXiv预印本，arXiv:2102.11253。Tukey, J.（1991年）。多重比较的哲学。《Statistical Science》，6(1)：100–116。Vovk, V., Wang, B. 和 Wang, R.（2022年）。任意依赖下合并 p 值的可容许方法。《统计年刊》，50(1)：351–375。Vovk, V. 和 Wang, R.（2020）。通过平均合并 p 值。《Biometrika》，107(4)：791–808。Zhao, Q., Small, D. S. 和 Su, W.（2019年）。当许多 p值一致保守时的多重检验：及其在教育干预中检验定性交互作用的应用。《美国统计学会会刊》，114(527)：1291–1304。

A 证明

A.1 命题 3.1 的证明

证明。根据程序的第 2 步可知， $\phi_{\mathcal{T}}$是 $\{2p_i, i \in \mathcal{S}^- \cap \mathcal{T}\} \cup \{2q_j, j \in \mathcal{S}^+ \cap \mathcal{T}\}$的函数。给定符号向量后，局部检验的有效性源于其结合了有效的 p值，因为根据假设（$A1$），每个条件 p值的有效性都得到保证，并且根据假设（$A2$），它们彼此独立。因此 $\forall\, \theta \in \mathbb{R}^n$,

$$\mathbb{P}_{\{\theta_i: i \in \mathcal{T}\}}(\phi_{\mathcal{T}} = 1 \mid \{sign(p_i - 1/2) : i \in \mathcal{T}\}) \leq \alpha.$$

此外，由于这些 p值彼此独立（假设（$A2$）），可知

$$\mathbb{P}_{\theta}(\phi_{\mathcal{T}} = 1 \mid \mathcal{S}) = \mathbb{P}_{\{\theta_i: i \in \mathcal{T}\}}(\phi_{\mathcal{T}} = 1 \mid \{sign(p_i - 1/2) : i \in \mathcal{T}\}).$$

至此，证明完成。

□

A.2 命题 4.1 的证明

证明。假设真实的 θ 对于某个 $\mathcal{K} \subseteq [n]$ 属于 $\Theta_{\mathcal{K}}$。对于每个 $\mathcal{I}$，我们有 $\psi_{|\mathcal{K}\cap\mathcal{I}|}(\mathcal{I}) \leq \psi_{\mathcal{K}}$。因此，若 $\psi_{\mathcal{K}} = 0$，则对于所有 $\mathcal{I}$，$|\mathcal{K} \cap \mathcal{I}| = n^+(\mathcal{I}) \in \mathcal{N}^+(\mathcal{I})_\alpha$，于是

$$\mathbb{P}_\theta(n^+(\mathcal{I}) \in \mathcal{N}^+(\mathcal{I})_\alpha \ \forall \mathcal{I}) \geq \mathbb{P}_\theta(\psi_{\mathcal{K}} = 0) \geq 1 - \alpha.$$

此外，事件 $\{n^+ \mathcal{I} \in \mathcal{N}^+ \mathcal{I}_\alpha\}$ $\{\mathcal{N}^+ \mathcal{I}_\alpha \leq n^+ \mathcal{I} \leq \mathcal{N}^+ \mathcal{I}_\alpha\}$ 意味着对所有 $\mathcal{I}$，$\min(\cdot)(\cdot) \max(\cdot)$，因此 $\ell_\alpha^+(\mathcal{I}) = \min(\mathcal{N}^+(\mathcal{I})_\alpha)$，$\ell_\alpha^-(\mathcal{I}) = |\mathcal{I}| - \max(\mathcal{N}^+(\mathcal{I})_\alpha)$ 满足 (3)。□

A.3 命题 4.2 的证明

证明。命题 D.1 表明，算法 1 返回由采用自适应局部检验 (14) 的分割程序导出的界限 $\ell_\alpha^+(\mathcal{I})$ 和 $\ell_\alpha^-(\mathcal{I})$，计算量至多为 $O(|\mathcal{I}|^2 \cdot \max(1, |\mathcal{I}^c|^2))$。

假设对某个 $\mathcal{K} \subseteq [n]$，真实的 $\theta \in \Theta_{\mathcal{K}}$，并令 $\mathcal{T} = \{\mathcal{K}^c \cap \mathcal{S}^-\} \cup \{\mathcal{K} \cap \mathcal{S}^+\}$。分割程序通过使用 $\psi_{\mathcal{K}} = \phi_{\tilde{\mathcal{T}}}$ 检验正交象限假设 $J_{\mathcal{K}}$，并

$$\sup_{\theta \in \Theta_{\mathcal{K}}} \mathbb{P}_\theta(\phi_{\mathcal{K}} = 1 \mid \mathcal{S}) \leq \sup_{\theta \in H_{\tilde{\mathcal{T}}}} \mathbb{P}_\theta(\phi_{\tilde{\mathcal{T}}} = 1 \mid \mathcal{S}) \leq \alpha$$

其中，第一个不等式由 $J_{\mathcal{K}} \subseteq H_{\tilde{\mathcal{T}}}$ 得出，第二个不等式由命题 3.1 得出。这证明，对于真实正交象限假设 $J_{\mathcal{K}}$ 的检验 $\phi_{\mathcal{K}}$，在给定符号向量 $\mathcal{S}$ 的条件下，I 类错误的显著性水平为 α。

此外，

$$\mathbb{P}_\theta(\tilde{\mathcal{T}} = \emptyset) = \mathbb{P}_\theta(\mathcal{S}^- = \mathcal{K}) = \prod_{i \in \mathcal{K}} \mathbb{P}_\theta(p_i \leq 0.5) \prod_{j \notin \mathcal{K}} \mathbb{P}_\theta(p_j > 0.5) \geq \mathbb{P}_0(\mathcal{S}^- = \mathcal{K}) \geq 2^{-n}. \tag{17}$$

Thus

$$\mathbb{P}_\theta(\psi_{\mathcal{K}} = 1) = \sum_{\mathcal{S}} \mathbb{P}_\theta(\phi_{\tilde{\mathcal{T}}} = 1 | \mathcal{S}) \mathbb{P}_\theta(\mathcal{S}) \leq \alpha(1 - \mathbb{P}_\theta(\tilde{\mathcal{T}} = \emptyset)) \leq \alpha(1 - 2^{-n}).$$

这确保了对真实正交象限假设 $J_{\mathcal{K}}$ 的检验 $\phi_{\mathcal{K}}$，其 I 类错误的无条件显著性水平为 $\tilde{\alpha} = \alpha(1 - 2^{-n})$。另请参阅 Al Mohamad 等（2020）补充材料第 1.2 节，其中讨论了自适应似然比检验的无条件第一类错误控制。

前述结果与命题 4.1 一起推出了条件覆盖率和无条件覆盖率（15）和（16）。

□

A.4 命题 4.3 的证明

˜证明。为了证明对于每个 $\mathcal{I}$，$\ell_\alpha^+(\mathcal{I}) \leq \ell_\alpha^+(\mathcal{I})$，我们采用反证法。˜假设存在 $\exists\mathcal{I}$，使得 $\ell_\alpha^+(\mathcal{I}) > \ell_\alpha^+(\mathcal{I})$。我们有 $\ell_\alpha^+(\mathcal{I}) = \ell_\alpha^+(\mathcal{I} \cap \mathcal{S}^-)$，并且根据 Goeman 等（2021）中的引理 1:

$$\ell_\alpha^+(\mathcal{I} \cap \mathcal{S}^-) = \min_{\mathcal{V} \subseteq [n]} (|(\mathcal{I} \cap \mathcal{S}^-) \setminus \mathcal{V}| : \phi_\mathcal{V} = 0). \tag{18}$$

˜由于在显著性水平 α 下，关于 $n^+(\mathcal{I})$ 等于 $\ell_\alpha^+(\mathcal{I})$ 的零假设未被拒绝，˜因此存在 $\exists\mathcal{K}$，使得 $|\mathcal{K} \cap \mathcal{I}| = \ell_\alpha^+(\mathcal{I})$，且 $J_\mathcal{K}$ 在显著性水平 α 下未被拒绝。等价地，存在 $\exists\mathcal{U} = (\mathcal{K}^c \cap \mathcal{S}^-) \cup (\mathcal{K} \cap \mathcal{S}^+)$，使得 $\phi_\mathcal{U} = 0$ 并且

$$|(\mathcal{I} \cap \mathcal{S}^-) \setminus \mathcal{U}| = |\mathcal{I} \cap \mathcal{S}^- \cap \mathcal{K}| \leq \tilde{\ell}_\alpha^+(\mathcal{I}) < \ell_\alpha^+(\mathcal{I}).$$

另一方面，由公式（18）可知，$\ell_\alpha^+(\mathcal{I}) \leq |\mathcal{I} \cap \mathcal{S}^-) \setminus \mathcal{U}|$，这与˜假设 $\ell_\alpha^+(\mathcal{I}) < \ell_\alpha^+(\mathcal{I})$ 相矛盾。

$\ell_\alpha^-(\mathcal{I}) \leq$˜对每个 $\mathcal{I}$，都可以用类似的方法证明 $\ell_\alpha^-(\mathcal{I})$。

为了证明 ˜对于每个 $\mathcal{I} \subseteq \mathcal{S}^-$，写作：$\ell_\alpha^+(\mathcal{I}) = \ell_\alpha^+(\mathcal{I})$

$$\begin{aligned}
\tilde{\ell}_\alpha^+(\mathcal{I}) &= \min\{v \in \{0, \ldots, n\} : \psi_v(\mathcal{I}) = 0\} \\
&= \min\{v \in \{0, \ldots, n\} : \min_{\mathcal{K} \subseteq [n] : |\mathcal{K} \cap \mathcal{I}| = v} \phi_{(\mathcal{K}^c \cap \mathcal{S}^-) \cup (\mathcal{K} \cap \mathcal{S}^+)} = 0\} \\
&= \min_{\mathcal{K} \subseteq [n]} \{|\mathcal{K} \cap \mathcal{I}| : \phi_{(\mathcal{K}^c \cap \mathcal{S}^-) \cup (\mathcal{K} \cap \mathcal{S}^+)} = 0\} \\
&= \min_{\mathcal{V} \subseteq [n]} \{|\mathcal{V}^c \cap \mathcal{I}| : \phi_{(\mathcal{V} \cap \mathcal{S}^-) \cup (\mathcal{V}^c \cap \mathcal{S}^+)} = 0\}
\end{aligned}$$

若 $\mathcal{I} \subseteq \mathcal{S}^-$，则对于任意 $\mathcal{V}$，$\mathcal{I} \setminus \{(\mathcal{V} \cap \mathcal{S}^-) \cup (\mathcal{V}^c \cap \mathcal{S}^+)\} = |\mathcal{I} \setminus \mathcal{V}|$，因此

$$\tilde{\ell}_\alpha^+(\mathcal{I}) = \min_{\mathcal{V} \subseteq [n]} \{|\mathcal{I} \setminus \mathcal{V}| : \phi_{(\mathcal{V} \cap \mathcal{S}^-) \cup (\mathcal{V}^c \cap \mathcal{S}^+)} = 0\} = \min_{\mathcal{V} \subseteq [n]} \{|\mathcal{I} \setminus \mathcal{V}| : \phi_\mathcal{V} = 0\} = \ell_\alpha^+(\mathcal{I}).$$

˜$\ell_\alpha^-(\mathcal{I}) = \ell_\alpha^-(\mathcal{I})$ 对每个 $\mathcal{I} \subseteq \mathcal{S}^+$ 都成立，这可用类似的方法证明。

□

B 情形 $n = 2$

图 6 展示了使用 Fisher 组合函数时，自适应局部检验划分（左上图）和方向性闭合检验（右上图）针对基假设的拒绝域。

如果交集假设 $H_{\mathcal{I}}$ 的拒绝意味着其至少一个分量假设 H_i、 $i \in \mathcal{I}$被拒绝（Romano 等人，2011年），我们就称闭合检验程序具有相容性。由图 6 的右上图和右下图可见，采用 Fisher 组合检验的方向性闭合检验程序也不具有相容性。例如，当 $\alpha = 0.2$ 如图所示时，如果（p_1、 p_2） =（0.101，0.101），则 Fisher 组合检验 f_{Fisher}（$2p_1$、 $2p_2$）会拒绝 $H_1^- \cap H_2^-$。然而，由于 $2p_i > \alpha$、 $i = 1$，2， H_1^-和 H_2^-均未被拒绝。

如图 1 所示，对于 $n > 2$，采用 Simes 局部检验的方向性闭合检验程序也不具有相容性，尽管对于 $n = 2$，它具有相容性。

从图 7 可以看出，当 $\alpha = 0.3$ and (p_1, p_2) =(0.19, 0.19) 时，我们能够拒绝 $H_1^- \cap K_1$ and $K_1 \cap H_2^-$，因为 $2p_i < 2\alpha/3$，其中 $i = 1$，2。然而，我们不能拒绝 $H_1^- \cap H_2^-$，因为 $f_{\text{Fisher}}(2p_1, 2p_2) = 0.423 > \alpha$。因此，我们得出结论，$n^+$ 不能等于 1，即 $n^+ \in \mathcal{N}_\alpha^+ = \{0, 2\}$。不过，在这种情况下，如果我们分别将 n^+ 的下界和上界计算为 min($\mathcal{N}_\alpha^+$) 和 max($\mathcal{N}_\alpha^+$)，得到的结果 0 和 2 并无信息价值。

最后，图 8 展示了针对假设族 (4) 的两种 FWER 控制程序的 $n = 2$ 拒绝域。第一种是 Guo 和 Romano（2015）提出的 Holm 型程序，要求满足 § 3 中的条件 (A1)-(A2)。第二种程序是 Sid´ak 型程序，也称为最初由 Spjotvoll（1972）提出、后经 Bohrer 和 Schervish （ 1980 ）修改的 “Sp 交叉法 ”，它只要求 n p- 值相互独立。其作为分割程序的推导见 § C。

C 基于无条件p-值的分割程序

基于无条件 p-值 p_1、……、 p_n 和 q_1、……、 q_n 的分割程序通过以下方式检验每个分割假设 $J_{\mathcal{K}}$

$$\psi_{\mathcal{K}} \quad = \quad \mathbb{1}\{f(\{p_i, i \in \mathcal{K}^c\} \cup \{q_j, j \in \mathcal{K}\}) \leq \alpha\}.$$

若假设 p_1、……、 p_n 相互独立，我们可以使用 Sid´ak 的合并函数 $f_{\text{Šidák}}$(x_1、……、 x_d) $= 1 - (1 - \min(x_1$、……、 $x_d))^d$，并得到以下家族错误率控制程序：若 $p_i \leq 1 - (1 - \alpha)^{(1/n)}$，则声明 $\theta_i > 0$；若 $q_i \leq 1 - (1 - \alpha)^{(1/n)}$，则声明 $\theta_i \leq 0$。图 8 右图展示了针对 $n = 2$ 的拒绝域。

有几种合并函数适用于 p-值之间的任意依赖 p_1、……、 p_n（Vovk等人，2022年）。其中最广为人知的或许是 Bonferroni 合并函—

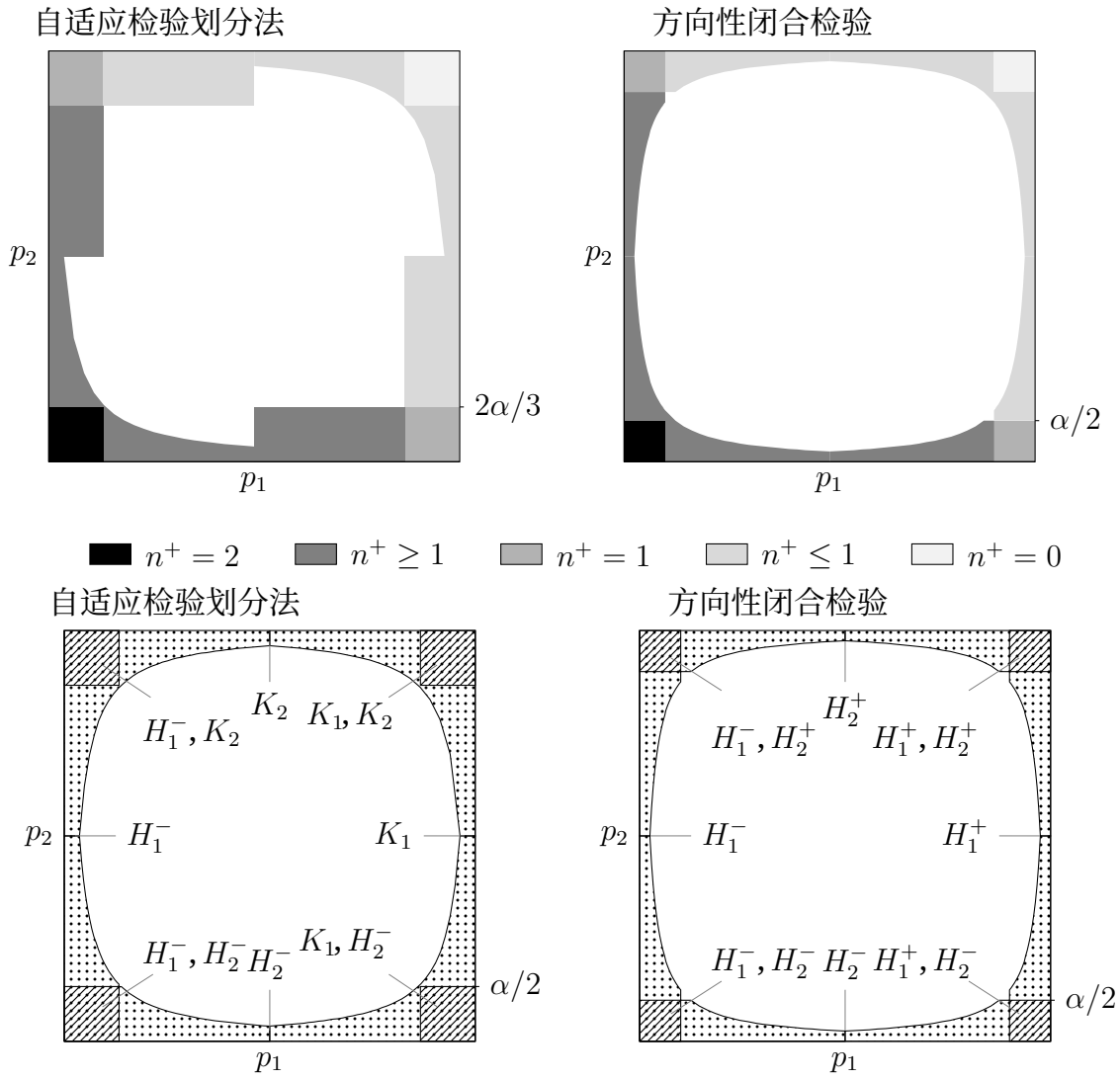


图 6：上排展示了使用 Fisher 合并函数时，自适应检验划分程序（左上图）和方向性闭合检验程序（右上图）中，对 n^+ 作出 $(1-\alpha)$ 置信度推断的区域。灰度表示推断类型。下排展示了自适应局部检验划分法（左下图）和方向性闭合检验（右下图）中导致拒绝基假设的区域。图案表示在显著性水平 α 下拒绝的假设数量：一个（点）或两个（斜线）。每幅图均基于 $\alpha = 0.2$。

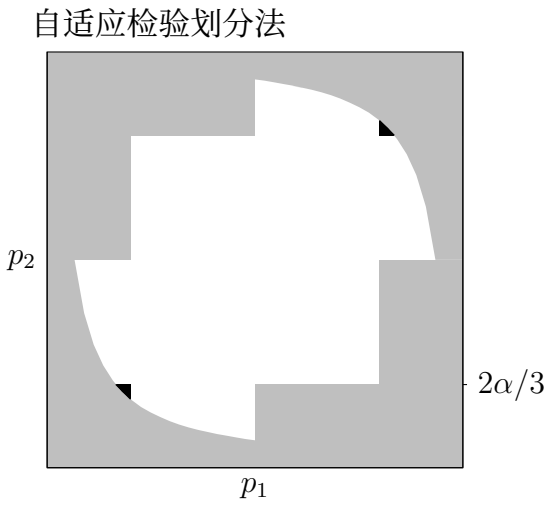


图 7：采用自适应检验和 Fisher 合并函数的分割程序，在 (p_1, p_2) 属于黑色区域时，以 $n^+ \in \{0, 2\}$ $(1-\alpha)$ 置信度得出结论。该图基于 $\alpha = 0.3$。

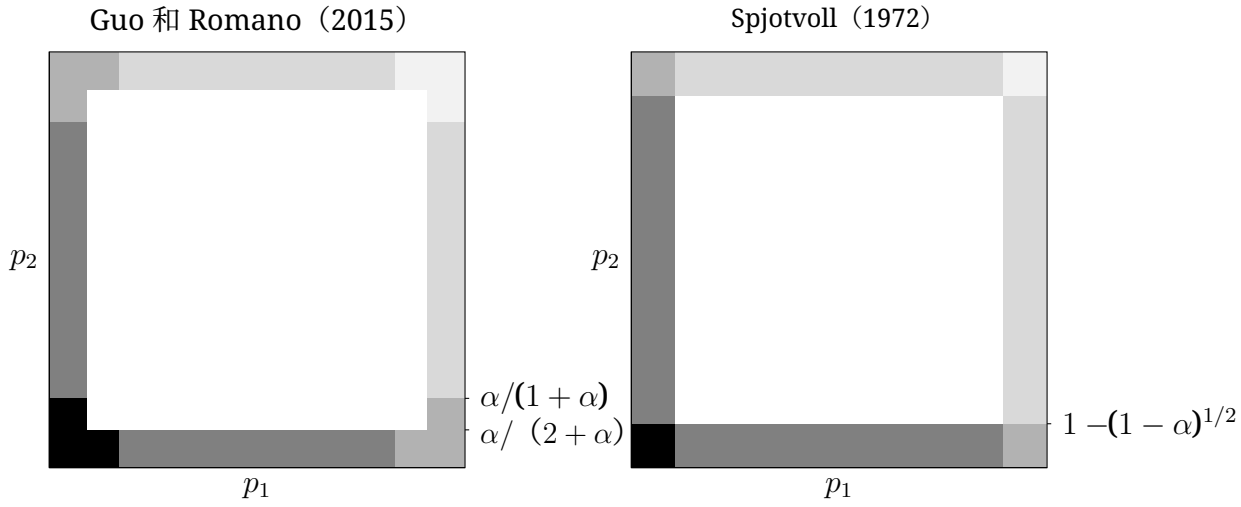


图 8：Guo 和 Romano（2015）的 Holm 型程序（左图）以及 Spjotvoll（1972）的 ˇSid´ak 型程序（右图）。每幅图均基于 $\alpha = 0.2$。

当 $f_{\text{Bonferroni}}(x_1,\ldots,x_d) = d\min(x_1,\ldots,x_d)\wedge 1$。在分割程序中使用 $f_{\text{Bonferroni}}$ 时，可得到 Bauer 等人（1986）的 FWER 控制程序：若 $p_i \le \alpha/n$，则判定 $\theta_i > 0$；若 $q_i \le \alpha/n$，则判定 $\theta_i \le 0$。

在单因素设计中，将 n 个组与共同对照组进行比较时，会出现正相关的 p-值 $p_1,\ldots,p_n$。为简单起见，假设每组有 m 个观测值。感兴趣参数为 $\theta_i = \mu_i - \mu_0$，相应的检验统计量为 $\hat{\theta}_i = (\hat{\mu}_i - \hat{\mu}_0)/\sqrt{2\hat{\sigma}^2/m}$，服从自由度为 $nm-1$ 的 Student t 分布。对于 $i=1,\ldots,n$，p-值为 $p_i = \mathbb{P}(t_{nm-1} \ge \theta_i)$。值得注意的是，$\hat{}$ p-值之间存在相关性，因为 $(\theta_1,\ldots,\theta_n)$ 服从自由度为 $nm-1$、相关矩阵非对角元素均等于 $1/2$ 的 n 维 t 分布。用于 $J_{\mathcal{K}}$ 的邓尼特检验（1955）为

$$\psi_{\mathcal{K}} \quad = \quad \mathbb{1}\{\max(\{\hat{\theta}_i, i\in\mathcal{K}^c\}\cup\{-\hat{\theta}_j, j\in\mathcal{K}\}) > c_\alpha(\mathcal{K})\} \tag{19}$$

其中，$c_\alpha(\mathcal{K})$ 表示具有相关矩阵的 n 维 t 分布中最大值的 $(1-\alpha)$ 分位数，该相关矩阵的非对角元素的符号取决于 $\mathcal{K}$。

为说明这一方法，我们考虑 Bretz 等人（2016）的 `recovery` 数据。一家公司开展了一项研究，以评估专用加热毯在帮助维持术后体温方面的效果。研究在手术患者中测试了四种毯子，分别标记为 b0、b1、b2 和 b3，以测量它们对恢复时间的影响。b0 毯子是各家医院已在使用的标准选项。研究的重点是评估患者的恢复时间，以分钟计；这些患者被随机分配到四种治疗之一。恢复时间越短，表明治疗效果越好。本研究要回答的关键问题是，与 b0 相比，b1、b2 或 b3 类型的毯子是否会改变恢复时间。为对数据进行正式分析，我们假设采用单因素设计

$$y_{ij} = \gamma + \mu_i + \epsilon_{ij}$$

其中，$\gamma+\mu_i$ 表示毯子 bi, $i=0$ 的预期恢复时间，1、2、3，且误差 $\epsilon_{ij} \sim N(0,\sigma^2)$ 独立同分布。

邓尼特程序（1955）是解决多对一比较问题的标准程序。在缺乏关于效应方向的具体信息时，我们通常将双侧原假设设定为 $H_i:\ \theta_i = 0$，其中$\theta_i = \mu_i - \mu_0$ 和 $i=1$，2、3。

表5列出了将单步和逐步向下 Dunnett 检验程序应用于点零假设 $H_i:\ \theta_i = 0$ 时经多重性调整的 p 值。尽管逐步向下 Dunnett 检验程序的检验效能更高，但单步 Dunnett 检验程序提供了

表5：针对 `recovery` 数据的 Dunnett 检验程序：双侧和单侧假设的单步及逐步向下调整p-值。

假设	检验统计量	p-值	Dunnett 调整后的 p-值	
			单步	逐步向下
$H_1 : \theta_1 = 0$	-1.330	0.191	0.456	0.192
$H_2 : \theta_2 = 0$	-4.656	< 0.001	< 0.001	< 0.001
$H_3 : \theta_3 = 0$	-1.884	0.067	0.884	0.127
$H_1^+ : \theta_1 \geq 0$	-1.330	0.096	0.241	0.096
$H_2^+ : \theta_2 \geq 0$	-4.656	< 0.001	< 0.001	< 0.001
$H_3^+ : \theta_3 \geq 0$	-1.884	0.034	0.092	0.064
$H_1^- : \theta_1 \leq 0$	-1.330	0.904	0.996	0.996
$H_2^- : \theta_2 \leq 0$	-4.656	1	1	1
$H_3^- : \theta_3 \leq 0$	-1.884	0.966	1	0.997

同时置信区间。这些区间使我们能够得出方向性结论，而不会有发生第三类错误的风险（Finner 等，2021）。在显著性水平 $\alpha = 0.1$ 下，我们得出结论：与标准毯子 b0 相比，毯子 b2 导致的恢复时间显著缩短。为便于比较，还提供了单侧 Dunnett 检验程序的结果，假设事先已掌握有关效应方向的具体信息。

表 6 显示了使用 Dunnett 检验 (19) 对正交象限假设进行检验所得的 p 值。p 值通过 `multcomp` R 包的 `glht` 函数计算。表 7 列出了每一对 $H_i^-: \theta_i \leq 0, K_i: \theta_i > 0$ 的多重性校正后 p 值。在显著性水平 $\alpha = 0.1$ 下，我们得出结论：毯子 b2 和 b3 的恢复时间不超过标准毯子 b0。

表 6：使用 Dunnett 检验的分割程序： `recovery` 数据中正交象限假设的 p值。

正交象限假设	检验统计量	临界值 $\alpha = 0.1$	p值
$K_1 \cap K_2 \cap K_3$	4.656	1.843	< 0.001
$K_1 \cap K_2 \cap H_3^-$	4.656	1.873	< 0.001
$H_1^- \cap K_2 \cap K_3$	4.656	1.867	< 0.001
$K_1 \cap H_2^- \cap K_3$	1.884	1.867	0.097
$K_1 \cap H_2^- \cap H_3^-$	1.330	1.867	0.259
$H_1^- \cap K_2 \cap H_3^-$	4.656	1.866	< 0.001
$H_1^- \cap H_2^- \cap K_3$	1.884	1.874	0.098
$H_1^- \cap H_2^- \cap H_3^-$	-1.330	1.843	0.996

表 7：使用 Dunnett 检验的分割程序： `recovery` 数据中基假设的校正后 p值。

假设	调整后的 p 值	假设	调整后的 p 值
H_1^-: $\theta_1 \leq 0$	0.996	$K_1 : \theta_1 > 0$	0.259
H_2^-: $\theta_2 \leq 0$	0.996	$K_2 : \theta_2 > 0$	< 0.001
H_3^-: $\theta_3 \leq 0$	0.996	$K_3 : \theta_3 > 0$	0.098

D 使用自适应检验的分割程序的精确捷径

我们首先讨论如何计算 n^+ 和 $n^- + n^0$ 对应的下界 $\tilde{\ell}_\alpha^+$ 和 $\tilde{\ell}_\alpha^-$。假设我们观测到大小为 $|\mathcal{S}^-| = s$ 的 $\mathcal{S}^-$，并希望检验式 (14) 中的检验是否在显著性水平 α 下拒绝所有满足 $|\mathcal{K}| = k$ 的假设 $J_\mathcal{K}$：

$$f_k = \max_{\mathcal{K}:|\mathcal{K}|=k} f(\{2p_i, i \in \mathcal{S}^- \cap \mathcal{K}^c\} \cup \{2q_i, i \in \mathcal{S}^+ \cap \mathcal{K}\}) \quad \leq \quad \alpha.$$

范德蒙德卷积

$$\binom{n}{k} \quad = \quad \sum_{v=0}^{k} \binom{s}{k-v} \binom{n-s}{v}$$

指出，对于每个 $v = 0$，.....，k，都存在 $\binom{s}{k-v}\binom{n-s}{v}$ 个大小为 k 的集合 $\mathcal{K}$，使得 $|\mathcal{S}^+ \cap \mathcal{K}| = v$ 且 $|\mathcal{S}^- \cap \mathcal{K}| = k - v$（或等价地，$|\mathcal{S}^- \cap \mathcal{K}^c| = s - k + v$）。于是，最大化问题变为

$$f_k = \max_{v \in \{\max(0,k-s),\ldots,\min(k,n-s)\}} \max_{\substack{\mathcal{K}:|\mathcal{S}^+\cap\mathcal{K}|=v,\\ |\mathcal{S}^-\cap\mathcal{K}|=k-v}} f(\{2p_i, i \in \mathcal{S}^- \cap \mathcal{K}^c\} \cup \{2q_i, i \in \mathcal{S}^+ \cap \mathcal{K}\})$$

对于任意递增函数 f，最大值的一个解是取最大的 v 个 p-值 q_i，其中 $i \in \mathcal{S}^+$，以及最大的 $s - k + v$ 个 p-值 p_i，其中 $i \in \mathcal{S}^-$，即

$$f_k = \max_{v \in \{\max(0,k-s),\ldots,\min(k,n-s)\}} f(\{2p_{(k-v+1)}, \ldots, 2p_{(s)}\} \cup \{2q_{(n-s-v+1)}, \ldots, 2q_{(n-s)}\}). \quad (20)$$

为了计算 $k = 0$、.....、n对应的 (20) 中的 f_k，我们可以使用嵌套循环。内层循环的迭代次数（即 (20) 中的索引 v，其取值范围从 $\max(0, k-s)$ 到 $\min(k, n-s)$）取决于外层循环索引的值（即 k，其取值范围从 0 到 n）以及 $\mathcal{S}^-$ 的大小 s。这两个循环的总复杂度为 $O(n^2)$。当 n 为偶数时，最大迭代次数出现在 $s \in \{n/2 - 1, n/2, n/2 + 1\}$；当 n 为奇数时，最大迭代次数出现在 $s \in \{(n-1)/2, (n+1)/2\}$；最小迭代次数出现在 $s \in \{0, n\}$。

我们刚才讨论的程序只是以下方法的一个特例，该方法用于推导一般子集 $\mathcal{I}$ 的下界 $\ell_\alpha^+(\mathcal{I})$ 和 $\ell_\alpha^-(\mathcal{I})$。

˜命题 D.1。对于任意 $\mathcal{I} \subseteq [n]$，算法 1 返回通过采用带有自适应局部检验 (14) 的分割程序所推导出的下界 $\ell_\alpha^+(\mathcal{I})$ 和 $\mathcal{l}_\alpha^-(\mathcal{I})$，计算量至多为 $O(|\mathcal{I}|^2 \cdot \max(1, |\mathcal{I}^c|^2))$。特别地，计算 H_i^- 和 K_i 的校正后 p-值需要 $O(n^2)$ 时间。

证明。假设我们观测到大小为 $|\mathcal{S}^-| = s$ 的 $\mathcal{S}^-$，对于任意 $\mathcal{I} \subseteq [n]$ 和任意 $v \in \{0, \ldots, |\mathcal{I}|\}$，我们想要检查式 (14) 中的检验是否会在显著性水平 α 下拒绝所有满足 $|\mathcal{K} \cap \mathcal{I}| = v$ 的假设 $J_\mathcal{K}$（或者等价地，检查 $J_v(\mathcal{I})$: $n^+(\mathcal{I}) = v$ 是否在显著性水平 α 下被拒绝）：

$$f_v(\mathcal{I}) = \max_{\mathcal{K}:|\mathcal{K}\cap\mathcal{I}|=v} f(\{2p_i, i \in \mathcal{S}^- \cap \mathcal{K}^c\} \cup \{2q_i, i \in \mathcal{S}^+ \cap \mathcal{K}\}) \leq \alpha.$$

算法 1 计算

$$f_{v,u}(\mathcal{I}) = \max_{\substack{\mathcal{K}:|\mathcal{K}\cap\mathcal{I}|=v,\\ |\mathcal{K}\cap\mathcal{I}^c|=u}} f(\{2p_i, i \in \mathcal{S}^- \cap \mathcal{K}^c\} \cup \{2q_i, i \in \mathcal{S}^+ \cap \mathcal{K}\})$$

使得 $f_v(\mathcal{I}) = \max\{f_{v,u}(\mathcal{I})\ u = 0, \ldots, |\mathcal{I}^c|\}$。式 (9) 中的函数 $f(\cdot)$ 将 $2p_i$ 与 $i \in \mathcal{S}^- \cap \mathcal{K}^c$ 以及 $2q_i$ 与 $i \in \mathcal{S}^+ \cap \mathcal{K}$ 结合起来。记 $\mathcal{K} = (\mathcal{K} \cap \mathcal{I}) \cup (\mathcal{K} \cap \mathcal{I}^c)$ 和 $\mathcal{K}^c = (\mathcal{K}^c \cap \mathcal{I}) \cup (\mathcal{K}^c \cap \mathcal{I}^c)$，则有

$$\begin{aligned} f_\mathcal{K}(\mathcal{I}) &= f(\{2p_i, i \in \mathcal{S}^- \cap \mathcal{K}^c\} \cup \{2q_i, i \in \mathcal{S}^+ \cap \mathcal{K}\}) \\ &= f(\{2p_i, i \in \mathcal{S}^- \cap \mathcal{K}^c \cap \mathcal{I}\} \cup \{2p_i, i \in \mathcal{S}^- \cap \mathcal{K}^c \cap \mathcal{I}^c\} \\ &\quad \cup \{2q_i, i \in \mathcal{S}^+ \cap \mathcal{K} \cap \mathcal{I}\} \cup \{2q_i, i \in \mathcal{S}^+ \cap \mathcal{K} \cap \mathcal{I}^c\}). \end{aligned}$$

考虑范德蒙德卷积:

$$\binom{|\mathcal{I}|}{v} = \sum_{k=0}^{v} \binom{|\mathcal{S}^- \cap \mathcal{I}|}{v-k} \binom{|\mathcal{S}^+ \cap \mathcal{I}|}{k}, \quad \binom{|\mathcal{I}^c|}{u} = \sum_{j=0}^{u} \binom{|\mathcal{S}^- \cap \mathcal{I}^c|}{u-j} \binom{|\mathcal{S}^+ \cap \mathcal{I}^c|}{j}.$$

第一个卷积说明，对于每个 $k \in \{0, \ldots, v\}$，存在 $\binom{|\mathcal{S}^-\cap\mathcal{I}|}{v-k}\binom{|\mathcal{S}^+\cap\mathcal{I}|}{k}$ 个集合 $\mathcal{K}$，使得 $|\mathcal{K} \cap \mathcal{I}| = v$，其中 $|\mathcal{S}^+ \cap \mathcal{K} \cap \mathcal{I}| = k$ 且 $|\mathcal{S}^- \cap \mathcal{K} \cap \mathcal{I}| = v - k$。同样，第二个卷积说明，对于每个 $j \in \{0, \ldots, u\}$，存在 $\binom{|\mathcal{S}^-\cap\mathcal{I}^\rfloor|}{u-j}\binom{|\mathcal{S}^+\cap\mathcal{I}^c|}{j}$ 个集合 $\mathcal{K}$，使得$|\mathcal{K} \cap \mathcal{I}^c| = u$，其中 $|\mathcal{S}^+ \cap \mathcal{K} \cap \mathcal{I}^c| = j$ 且 $|\mathcal{S}^- \cap \mathcal{K} \cap \mathcal{I}^c| = u - j$。于是，最大化问题变为

$$f_{v,u}(\mathcal{I}) = \max_{\substack{k\in\{k_1,\ldots,k_2\},\ \mathcal{K}:|\mathcal{S}^+\cap\mathcal{K}\cap\mathcal{I}|=k,|\mathcal{S}^-\cap\mathcal{K}\cap\mathcal{I}|=v-k,\\ j\in\{j_1,\ldots,j_2\}\quad |\mathcal{S}^+\cap\mathcal{K}\cap\mathcal{I}^c|=j,|\mathcal{S}^-\cap\mathcal{K}\cap\mathcal{I}^c|=u-j}} \max f_\mathcal{K}(\mathcal{I})$$

算法 1：计算下界的快捷方法 $\tilde{\ell}_\alpha^+(\mathcal{I})$ 和 $\tilde{\ell}_\alpha^-(\mathcal{I})$ 推导得出带分割的自适应局部检验推导得出。

输入 ：右尾 p-值 p_1、……、p_n；合并函数 $f(\cdot)$；子集 $\mathcal{I} \subset [n]$；显著性水平 α

输出：置信界 $l_\alpha^+(\mathcal{I})$ 和 $u_\alpha^+(\mathcal{I})$，p-值 $f_v(\mathcal{I})$，用于 $J_v(\mathcal{I})$：$n^+(\mathcal{I}) = v$，用于 $v = 0$、……、$|\mathcal{I}|$

初始化：
$\mathcal{S}^- = \{i \in [n] : p_i \le 1/2\}$、$|\mathcal{S}^-| = s$、$f_s(\mathcal{I}) = 0$；
一个具有 $a_1 \ge \ldots \ge a_{|\mathcal{S}\cap\mathcal{I}|}$ 个排序值的 $= (a_1$、……、$a_{|\mathcal{S}-\cap\mathcal{I}|})$，对应 $\{2p_i : i \in \mathcal{S}^- \cap \mathcal{I}\}$；
b $= (b_1, \ldots, b_{|\mathcal{S}^-\cap\mathcal{I}^c|})$，其中 $b_1 \ge \ldots \ge b_{|\mathcal{S}^-\cap\mathcal{I}^c|}$ 个排序值来自 $\{2p_i : i \in \mathcal{S}^- \cap \mathcal{I}^c\}$；
c $= (c_1, \ldots, c_{|\mathcal{S}^+\cap\mathcal{I}|})$，其中 $c_1 \ge \ldots \ge c_{|\mathcal{S}^+\cap\mathcal{I}|}$ 个排序值来自 $\{2q_i : i \in \mathcal{S}^+ \cap \mathcal{I}\}$；
d $= (d_1, \ldots, d_{|\mathcal{S}^+\cap\mathcal{I}^c|})$，其中 $d_1 \ge \ldots \ge d_{|\mathcal{S}^+\cap\mathcal{I}^c|}$ 个排序值来自 $\{2q_i : i \in \mathcal{S}^+ \cap \mathcal{I}^c\}$；

对于 $v \leftarrow 0$ 到 $|\mathcal{I}|$，执行
  对于 $u \leftarrow 0$ 到 $|\mathcal{I}^c|$，执行
    对于 $k \leftarrow \max(0, v - |\mathcal{S}^- \cap \mathcal{I}|)$ 到 $\min(v, |\mathcal{S}^+ \cap \mathcal{I}|)$，执行
      对于 $j \leftarrow \max(0, u - |\mathcal{S}^- \cap \mathcal{I}^c|)$ 到 $\min(u, |\mathcal{S}^+ \cap \mathcal{I}^c|)$，执行
        $f_{k,j} = f\ (\{a_1, \ldots, a_{k-v+|\mathcal{S}^-\cap\mathcal{I}|}\} \cup \{b_1, \ldots, b_{j-u+|\mathcal{S}^-\cap\mathcal{I}^c|}\} \cup \{c_1, \ldots, c_k\} \cup \{d_1, \ldots, d_j\})$；
      end
    end
    $f_{v,u}(\mathcal{I}) \leftarrow$最大$\{f_{k,j}\}$；
  end
  $f_v(\mathcal{I}) = \max\{f_{v,u}(\mathcal{I})\ u = 0, \ldots\ldots, |\mathcal{I}^c|\}$；
结束
$\tilde{\ell}_\alpha^+(\mathcal{I}) \leftarrow \min(v \in \{0, \ldots, s\} : f_v(\mathcal{I}) > \alpha)$；
$\tilde{\ell}_\alpha^-(\mathcal{I}) \leftarrow |\mathcal{I}| -$最大值$(v \in \{s, \ldots, |\mathcal{I}|\} : f_v(\mathcal{I}) > \alpha)$；
返回 $\tilde{\ell}_\alpha^+(\mathcal{I})$, $\tilde{\ell}_\alpha^-(\mathcal{I})$，$f_0(\mathcal{I})$……，$f_{|\mathcal{I}|}(\mathcal{I})$；

其中 $k_1 = \max(0, v - |\mathcal{S}^- \cap \mathcal{I}|)$、$k_2 = \min(v, |\mathcal{S}^+ \cap \mathcal{I}|)$、$j_1 = \max(0, u - |\mathcal{S}^- \cap \mathcal{I}^c|)$ 和 $j_2 = \min(u, |\mathcal{S}^+ \cap \mathcal{I}^c|)$。

对于任意递增函数 $f(\cdot)$，最大值在以下取值处取得：最大的 k p 个值 $q_i$with $i \in \mathcal{S}^+ \cap \mathcal{I}$，最大的 $k - v + |\mathcal{S}^- \cap \mathcal{I}|$ p 个值 p_i with $i \in \mathcal{S}^- \cap \mathcal{I}^c$，最大的 jp 个值 q_i with $i \in \mathcal{S}^+ \cap \mathcal{I}$，以及最大的 $j - u + |\mathcal{S}^- \cap \mathcal{I}^c|$ p 个值 p_i with $i \in \mathcal{S}^- \cap \mathcal{I}$:

$$f_{v,u}(\mathcal{I}) = \max_{\substack{k \in \{k_1, \ldots, k_2\} \\ j \in \{j_1, \ldots, j_2\}}} f(\{a_1, \ldots, a_{k-v+|\mathcal{S} \cap \mathcal{I}|}\} \cup \{b_1, \ldots, b_{j-u+|\mathcal{S} \cap \mathcal{I}^c|}\} \cup \{c_1, \ldots, c_k\} \cup \{d_1, \ldots, d_j\})$$

其中 $a_1 \geq \ldots \geq a_{|\mathcal{S}^- \cap \mathcal{I}|}$、$b_1 \geq \ldots \geq b_{|\mathcal{S}^- \cap \mathcal{I}^c|}$、$c_1 \geq \ldots \geq c_{|\mathcal{S}^+ \cap \mathcal{I}|}$ 和 $d_1 \geq \ldots \geq d_{|\mathcal{S}^+ \cap \mathcal{I}^c|}$ 分别表示 $\{2p_i : i \in \mathcal{S}^- \cap \mathcal{I}\}$、$\{2p_i : i \in \mathcal{S}^- \cap \mathcal{I}^c\}$、$\{2q_i : i \in \mathcal{S}^+ \cap \mathcal{I}\}$ 和$\{2q_i : i \in \mathcal{S}^+ \cap \mathcal{I}^c\}$ 的排序值。

算法 1通过嵌套循环计算 $f_{v,u}(\mathcal{I})$。外层循环执行 $\min(v, |\mathcal{S}^+ \cap \mathcal{I}|) - \max(0, v - |\mathcal{S}^- \cap \mathcal{I}|)$ 次。每次执行外层循环时，内层循环执行 $\min(v, |\mathcal{S}^+ \cap \mathcal{I}|) - \max(0, u - |\mathcal{S} \cap \mathcal{I}^c|)$ 次。因此，计算$f_{v,u}(\mathcal{I})$ 的复杂度为 $O(|\mathcal{I}||\mathcal{I}^c|)$。

计算 $f_v(\mathcal{I})$（其中 $v = 0, \ldots, |\mathcal{I}|$）的复杂度为 $O(|\mathcal{I}|^2|\mathcal{I}^c|^2)$，因为需要计算 $f_{v,u}(\mathcal{I})$（其中 $u = 0, \ldots, |\mathcal{I}^c|$ ）以及 $v = 0, \ldots, |\mathcal{I}|$。如果 $\mathcal{I} = \{i\}$，则需要 $O(n^2)$ 来计算调整后的 p 值 $\bar{p}_i = f_0(\{i\})$ 和 $\bar{q}_i = f_1(\{i\})$，分别对应 H_i^- 和 K_i。

□

E n^+ 和 $n^- + n^0$的自适应下界

我们推导出 l_α^+和 l_α^-的下界，适用于 n^+ 和 $n^- + n^0$，使得

$$\mathbb{P}_\theta\left(l_\alpha^+ \leq n^+,\ l_\alpha^- \leq n^- + n^0 \mid \mathcal{S}\right) \geq 1 - \alpha. \tag{21}$$

这些界限可能比采用 $\mathcal{I} = [n]$ 自适应划分更紧。值得注意的是，这些界限与 $\mathcal{I} = [n]$ 自适应划分程序中的界限之间的差距很小，因此，如果关注正向发现和非正向发现，建议采用自适应划分。这一发现基于对广泛数据生成情形的考察（因篇幅所限未列出）。通常，分析人员除了关注 n^+ 和 $n^- + n^0$的下界外，还关注正向发现和非正向发现。由于进行进一步推断的成本很低，我们建议使用自适应划分，以找出 n^+ 和 $n^- + n^0$ 的下界并提供发现结果。不过，如果只关注 n^+ 和 $n^- + n^0$的下界，则应使用下述程序 E.1，因为它

在计算上要简单得多，并且其界限至少与程序 1 中采用 $\mathcal{I} = [n]$ 时一样紧，命题 E.2对此进行了形式化说明。

我们简要回顾部分合取（PC）假设的检验，以便将新颖的程序 E.1 置于背景中。令 $H^{r/n}$：$n^+ \leq r-1$ 表示 PC 零假设，即在 H_1^-、.....、H_n^- 中至多有 $r-1$ 个假设为假；令 $K^{r/n}$：$n^- + n^0 \leq r-1$ 表示 PC 零假设，即在 K_1、.....、K_n 中至多有 $r-1$ 个假设为假。对于 $r=1$，$H^{r/n}$ 表示所有参数均不为正的全局零假设。对于 $r=2$，拒绝 $H^{r/n}$ 可确立正方向上的最小可重复性（Benjamini 等人，2009；Jaljuli 等人，2022年）。

由于 $H^{r/n}$ 为假当且仅当所有大小为 $n-r+1$ 的交集假设均为假（Benjamini 等人，2009），因此，对于 $H^{r/n}$，有效的 p 值 $p^{r/n}$ 是所有由 $n-r+1$ 个零假设构成的交集中，最大的交集假设 p 值:

$$p^{r/n} = \max_{\{\mathcal{I}:\mathcal{I}\subseteq[n],|\mathcal{I}|=n-r+1\}} p_{\mathcal{I}},$$

其中，$p_{\mathcal{I}}$ 是交集假设 $\cap_{i\in\mathcal{I}} H_i^-$ 的 p 值。类似地，对于 $K^{r/n}$，一个有效的 p 值 $q^{r/n}$ 为

$$q^{r/n} = \max_{\{\mathcal{I}:\mathcal{I}\subseteq[n],|\mathcal{I}|=n-r+1\}} q_{\mathcal{I}},$$

其中，$q_{\mathcal{I}}$ 是交集假设 $\cap_{i\in\mathcal{I}} K_i$ 的 p 值。

对于满足单调性与对称性条件 (A0) 的 f，$p^{r/n} = f(p_{(r)}, \ldots, p_{(n)})$ 是一个有效的 p 值，满足

$$\sup_{\theta\in H^{r/n}} \mathbb{P}_\theta(p^{r/n} \leq x) \leq x \quad \forall x \in [0,1].$$

对于最不利参数配置（LFC）$\theta_{LFC} \in H^{r/n}$，该不等式取等号，即 $p^{r/n}$ 服从均匀分布；在此配置下，与 $r-1$ 参数相对应的 p 值为零（几乎必然），而与其余 $n-r+1$ 个参数相对应的 p 值服从均匀分布。例如，在正态均值问题中，检验 $H^{r/n}$ 时的 LFC 配置为 $r-1$ 个参数趋于正无穷，且 $n-r+1$ 个参数为零；检验 $K^{r/n}$ 时的 LFC 配置为 $r-1$ 个参数趋于负无穷，且 $n-r+1$ 个参数为零。

例如　例如，使用 Fisher 合并法（Fisher，1934 年）得到的 PC p-值为　　:

$$p^{r/n} = \mathbb{P}\Big(\chi^2_{2(n-r+1)} \geq -2\sum_{k=r}^{n}\log(p_{(k)})\Big), \quad q^{r/n} = \mathbb{P}\Big(\chi^2_{2(n-r+1)} \geq -2\sum_{k=r}^{n}\log(q_{(k)})\Big)$$

其中，$p_{(1)} \leq \ldots \leq p_{(n)}$和 $q_{(1)} \leq \ldots \leq q_{(n)}$分别表示 p_1、……、p_n以及q_1、……、q_n的排序值。请注意，对于连续检验统计量，$1 - p_{(k)} = q_{(n-k+1)}$。

只考虑 $\mathcal{S}^-$ 中针对 $H^{r/n}$的假设，我们可以避免纳入这样的 p 值：当原假设为真时，它们随机占优于均匀分布且大得多。因此，与 § 3 中一样，我们将仅限于由数据指引的方向，因此将使用以下方法检验 $H^{r/n}$

$$p^{r/n} = \max_{\{\mathcal{I}:\mathcal{I}\subseteq[n],|\mathcal{I}|=n-r+1\}} f(\{2p_i : i \in \mathcal{I} \cap \mathcal{S}^-\}) = \begin{cases} f\left(2p_{(r)}, \ldots, 2p_{(|\mathcal{S}^-|)}\right) & \text{if } r \leq |\mathcal{S}^-|, \\ 1 & \text{otherwise.} \end{cases} \quad (22)$$

并使用以下方法检验 $K^{r/n}$

$$q^{r/n} = \max_{\{\mathcal{I}:\mathcal{I}\subseteq[n],|\mathcal{I}|=n-r+1\}} f(\{2q_i : i \in \mathcal{I} \cap \mathcal{S}^+\}) = \begin{cases} f\left(2q_{(r)}, \ldots, 2q_{(|\mathcal{S}^+|)}\right) & \text{if } r \leq |\mathcal{S}^+|, \\ 1 & \text{otherwise.} \end{cases} \quad (23)$$

程序 E.1（自适应 PC 检验）。

步骤 1 执行程序 3.1 的步骤 1。

步骤 2 按顺序检验 $\{H^{r/n}: \ r = 1, \ldots, |\mathcal{S}^-|\}$，使用式 (22) 中的 $p^{r/n}$，显著性水平为 α。遇到第一个不拒绝的情况时停止，$p^{r/n} > \alpha$。令 l_α^+为拒绝次数，其中 $l_\alpha^+ \in \{0, \ldots, |\mathcal{S}^-|\}$。若 $l_\alpha^+ = n$，则返回 $l_\alpha^+ = n, l_\alpha^- = 0$；否则进入下一步。

步骤 3 按顺序检验 $\{K^{r/n}: \ r = 1, \ldots, n - |\mathcal{S}^-|\}$，使用式 (23) 中的 $q^{r/n}$，显著性水平为 α。遇到第一个不拒绝的情况时停止，$q^{r/n} > \alpha$。令 l_α^- 为拒绝次数，其中 $l_\alpha^- \in \{0, \ldots, n - |\mathcal{S}^-|\}$。返回 l_α^+和 l_α^-（其中 $l_\alpha^+ + l_\alpha^- \leq n$）。

注意，如果按顺序进行检验，并在每一步持续至 n，程序的界限将保持不变。这是因为 $p^{(|\mathcal{S}^-|+1)/n} > \alpha$ 和 $q^{(n-|\mathcal{S}^-|+1)/n} > \alpha$。

保证覆盖率在以下命题中得到了形式化表述。

命题 E.1. 设 $\left\{p^{r/n}: \ r \in [n]\right\}$ 是关于 $\{H^{r/n}: \ r \in [n]\}$ 的有效条件 PC p 值，且 $\left\{q^{r/n}: r \in [n]\right\}$ 是关于 $\{K^{r/n}: \ r \in [n]\}$ 的有效条件 PC p 值。则l_α^+和 l_α^- 满足（21）。此外，无条件覆盖率为

$$\mathbb{P}_\theta\left(l_\alpha^+ \leq n^+, \ l_\alpha^- \leq n^- + n^0\right) \geq 1 - (1 - 2^{-n})\alpha. \quad (24)$$

证明。假设 θ（真实参数向量）具有 $n^+ = t$ 个正坐标。当且仅当 $p^{(t+1)/n} > \alpha$ 和 $q^{(n-t+1)/n} > \alpha$时，有 $l_\alpha^+ \le t \le n - l_\alpha^-$。该结果源于：在给定 $\mathcal{S}^-$的条件下，只可能在一个方向上犯错误。具体而言，如果 $t > |\mathcal{S}^-|$，则 $p^{(|\mathcal{S}^-|+1)/n} > \alpha$，因为不可能拒绝 $H^{(|\mathcal{S}^-|+1)/n}$，所以不可能在下界方面出错。因此，如果 $t > |\mathcal{S}^-|$，则

$$\mathbb{P}_\theta(t \notin [l_\alpha^+, n - l_\alpha^-] \mid \mathcal{S}) = \mathbb{P}_\theta(t > n - l_\alpha^- \mid \mathcal{S}) \le \mathbb{P}_\theta(q^{(n-t+1)/n} \le \alpha \mid \mathcal{S}) \le \alpha.$$

类似地，如果 $t < |\mathcal{S}^-|$，那么 $q^{(n-|\mathcal{S}^-|+1)/n} > \alpha$，因为不可能拒绝 $K^{(n-|\mathcal{S}^-|+1)/n}$，所以不可能在上界方面出错。因此，如果 $t < |\mathcal{S}^-|$，那么

$$\mathbb{P}_\theta(t \notin [l_\alpha^+, n - l_\alpha^-] \mid \mathcal{S}) = \mathbb{P}_\theta(t < l_\alpha^+ \mid \mathcal{S}) \le \mathbb{P}_\theta(p^{(t+1)/n} \le \alpha \mid \mathcal{S}) \le \alpha.$$

如果 $t = |\mathcal{S}^-|$，那么由于下界至多为 $|\mathcal{S}^-|$ ，且上界至少为$n - |\mathcal{S}^+|$，因此不可能在任一界限上出错。因此，未覆盖 n^+ 的无条件错误为

$$\begin{aligned}
&\mathbb{P}_\theta(t \notin [l_\alpha^+, n - l_\alpha^-]) = \mathbb{E}\left(\mathbb{P}_\theta(t \notin [l_\alpha^+, n - l_\alpha^-] \mid \mathcal{S})\right) \\
&= \mathbb{E}_\theta\left\{\mathbb{I}(t > |\mathcal{S}^-|)\mathbb{P}_\theta(t > n - l_\alpha^- \mid \mathcal{S}^-) + \mathbb{I}(t < |\mathcal{S}^-|)\mathbb{P}_\theta(t < l_\alpha^+ \mid \mathcal{S}^-)\right\} \\
&\le \alpha\mathbb{E}_\theta\left(\mathbb{I}(t > |\mathcal{S}^-|) + \mathbb{I}(t < |\mathcal{S}^-|)\right) = \alpha\mathbb{P}_\theta(t \ne |\mathcal{S}^-|) \le \alpha(1 - 2^{-n}),
\end{aligned}$$

其中，最后一个不等式由 (17) 得出。 □

˜ ˜命题 E.2. 算法 1 返回下界 ℓ_α^+ 和 ℓ_α^-。这些下界至多与程序 E.1 得到的结果一样好。

证明. 只需注意，如果 $k < s$，那么 (20) 中的索引 v 从 $\max(0, k-s) = 0$ 开始，方法是计算 PC 条件 p 值 $p^{k+1/n} = f(\{2p_{(k+1)}, \ldots, 2p_{(s)}\})$。那么 $p^{k+1/n} > \alpha$ 意味着

$$f_k = \max_{\mathcal{K}:|\mathcal{K}|=k} f(\{2p_i, i \in \mathcal{S}^- \cap \mathcal{K}^c\} \cup \{2q_i, i \in \mathcal{S}^+ \cap \mathcal{K}\}) \ge p^{k+1/n} > \alpha$$

对于任意 $k \in \{0, \ldots, s-1\}$，即，算法 1 中针对 n^+ 的下界 ℓ_α^+ 小于或等于程序 E.1 的下界。同样，如果 $k > s$ ，则式 (20) 中的索引 v 从 $\max(0, k-s)$ 开始 $= k - s$ ，通过计算 PC 条件 p-值 $q^{n-k+1/n} = f(\{2q_{(n-k+1)}, \ldots, 2q_{(n-s)}\})$，因此 $q^{n-k+1/n} > \alpha$ 意味着

$$f_k = \max_{\mathcal{K}:|\mathcal{K}|=k} f(\{2p_i, i \in \mathcal{S}^- \cap \mathcal{K}^c\} \cup \{2q_i, i \in \mathcal{S}^+ \cap \mathcal{K}\}) \ge q^{n-k+1/n} > \alpha$$

˜对于任意 $k \in \{s+1, \ldots, n\}$，即，算法 1 中 $n^- + n_0$ 的下界 ℓ_α^- 小于或等于程序 E.1 的上界。 □

注释 E.1. 如果 $n^0 > 0$，那么程序 E.1 给出的下界未能覆盖 n^+、n^- 中至少一个的概率可能超过 α。如果 $n^0 = n$ 且所有 p 值均服从均匀分布，那么$\mathbb{P}$（$p^{1/n} \leq \alpha \cup q^{1/n} \leq \alpha$）$\approx 2\alpha$；随着 n^0 减小，下界未能覆盖至少一个参数的概率会从 2α 降至 α（对于 $n_0 = 0$）。

注释 E.2. 对于组合函数 $f()$，当且仅当在上述程序中 $l_\alpha^+ \geq 1$ 且 $l_\alpha^- \geq 1$ 时，Zhao 等人（2019）关于定性交互作用的检验在显著性水平 α 下被拒绝。因此，如果假设 $n^0 = 0$ 是合理的，那么上述程序便能很好地补充“存在定性交互作用”这一结论：它以 $(1-\alpha)$ 置信度给出经检验参数 n^+ 的（区间）估计。更一般地，对于预先定义的 $1 \leq a < b \leq n-1$，关于“$n^+ < a$ 或 $n^+ > b$”这一广义定性交互作用原假设的显著性水平为 α 的检验，其拒绝规则如下：若 $l_\alpha^+ \geq a$ 且 $n - l_\alpha^- \leq b$，则拒绝。为说明这是一个显著性水平为 α 的检验，考虑满足 $n^+(\theta) \notin [a,b]$ 的零假设值 θ。不失一般性，假设 $n^+(\theta) > b$。错误拒绝广义定性交互作用真零假设的概率为

$$\mathbb{P}_\theta(l_\alpha^+ \geq a \ and \ n - l_\alpha^- \leq b) \leq \mathbb{P}_\theta(n - l_\alpha^- \leq b) \leq \mathbb{P}_\theta(n - l_\alpha^- < n^+) \leq \alpha.$$

E.1 关于 n^+ 的一般置信界限说明

如果 $p^{r/n}$ 是检验 $H^{r/n}$（针对 $r = 1$、.....、n）的有效 p 值，那么

$$l_\alpha^+(p) \ = \ \max\{l \in \{0,\ldots,n\} : p^{r/n} \leq \alpha \text{ for } r = 0,\ldots,l\} \tag{25}$$

则对所有 $\theta \in \Theta$，均满足 $\mathbb{P}_\theta\big(l_\alpha(p) \leq n^+\big) \geq 1-\alpha$，其中 $p^{0/n} \equiv 0$，因为 $H^{0/n}$：$n^+ < 0$ 始终为假。类似地，如果 $q^{r/n}$ 是检验 $K^{r/n}$（针对 $r = 1$、.....、n）的有效 p 值，那么

$$l_\alpha^-(p) \ = \ \max\{u \in \{0,\ldots,n\} : q^{r/n} \leq \alpha \text{ for } r = 0,\ldots,u\} \tag{26}$$

则对所有 $\theta \in \Theta$，均满足 $\mathbb{P}_\theta\big(n^+ \leq n - l_\alpha^-(p)\big) \geq 1-\alpha$，其中 $q^{0/n} \equiv 0$，因为 $K^{0/n}$：$n^+ > n$ 始终为假。为简化记号，我们常将 $l_\alpha^+(p)$ 和 $l_\alpha^-(p)$ 分别简写为 l_α^+ 和 l_α^-，但这些界限当然是 p 值向量 p 的函数。

通过直接应用 Bonferroni 不等式可以证明，如果每个界限采用显著性水平 $\alpha/2$，即在（25）中使用 $l_{\alpha/2}^+$，在（26）中使用 $l_{\alpha/2}^-$，那么 $\mathbb{P}_\theta\big(l_{\alpha/2}^+ \leq n^+ \leq n - l_{\alpha/2}^-\big) \geq 1-\alpha$。Jaljuli 等人（2022）使用了这些界限，以补充系统综述中的荟萃分析。

然而，在每个方向上使用 $\alpha/2$（而不是自适应PC检验程序中的 α）进行校正是保守的。直观地说，校正力度应当较弱，因为

在给定的配置下，因超出一个界限而出错的概率远大于因超出另一个界限而出错的概率。为说明这一点，请注意

$$\mathbb{P}_\theta(n^+ \notin [l^+_{\alpha/2}, n - l^-_{\alpha/2}] = \mathbb{P}_\theta(n^+ < l^+_{\alpha/2}) + \mathbb{P}_\theta(n^+ > n - l^-_{\alpha/2})$$

在检验 PC零假设时，对于以下最不利参数配置（最不利配置，LFCs），其值为 $\alpha/2$：正参数值为 $p_i = 0$ （几乎必然），非正参数值的 p 值服从均匀分布；或者正参数值的 p 值（实际上）服从均匀分布，非正参数值为 $q_i = 0$ （几乎必然）。为此，注意，对于正参数为 n^+ 的最不利配置 $p_i = 0$， $p^{(n^+ +1)/n} \sim U$ 服从 (0, 1) 均匀分布，而 $q^{(n-n^+ +1)/n} = 1$ 几乎必然成立。对于正参数为 $n - n^+$ 的最不利配置 $q_i = 0$， $q^{(n-n^+ +1)/n} \sim U$ 服从 (0, 1) 均匀分布，而 $p^{(n^+ +1)/n} = 1$ 几乎必然成立。若下界被违反，则可能发生未覆盖，即 $p^{(n^+ +1)/n} \leq \alpha/2$；若上界被违反，也可能发生未覆盖，即 $q^{(n-n^+ +1)/n} \leq \alpha/2$。因此

$$\mathbb{P}_{LFC}(n^+ \notin [l^+_{\alpha/2}, n - l^-_{\alpha/2}]) = \mathbb{P}(U \leq \alpha/2) = \alpha/2.$$

我们猜想，当下界l^+_α 且 l^-_α时，覆盖保证通常为（$1 - \alpha$）。尤其是，当检验统计量来自一维指数族且连续时。无损一般性地，假设前 t 个坐标为正，即 θ_1、.....、 $\theta_t > 0$ 为正，且 θ_{t+1}、.....、 $\theta_n \leq 0$和 $n^+ = t$。因此，我们猜想，对于一大类有效的 PC p-值，以下优化问题的解为 α：

$$\begin{aligned} \max_\theta \quad & \mathbb{P}_\theta(p^{(t+1)/n} \leq \alpha) + \mathbb{P}_\theta(q^{(n-t+1)/n} \leq \alpha) \\ \text{s.t.} \quad & \theta_i > 0, i = 1, \ldots, t, \\ & \theta_j \leq 0, j = t+1, \ldots, n. \end{aligned}$$

要说明这对无条件组合检验（即使用不对符号向量$\mathcal{S}$进行条件化的局部检验）一般而言并不成立，请看下面这个示例。设 $n = 2$ 和 $n^+ = 1$ 满足 $\theta_1 = 0$ ，且 θ_2 为正。假设来自 H_i、 x_i的一个 p-值的分布如下： $\mathbb{P}(x_i = \alpha) = \alpha, \mathbb{P}(x_i = 1) = 1 - \alpha$.类似地，来自 K_i、 $y_i = 1 - x_i$的一个 p-值的分布如下： $\mathbb{P}(y_i = \alpha) = \alpha, \mathbb{P}(y_i = 1) = 1 - \alpha$. 这些 p-值是有效的，因为

$$\mathbb{P}_{H_i}(p_i \leq a) \leq a, \ \mathbb{P}_{K_i}(q_i \leq a) \leq a, \ \forall a \in [0, 1].$$

无条件 PC p 值（即，它是由不以符号向量 $\mathcal{S}$ 为条件的局部检验得出的）对于 $H^{2/2}$ 和 $K^{2/2}$ 分别为 $p^{2/2} = \max(p_1, p_2)$，其分布为

$\max(1-x_1, x_2)$，以及 $q^{2/2} = \max(q_1, q_2)$，其分布为 $\max(x_1, 1-x_2)$。因此：

$$\begin{aligned}
&\mathbb{P}_\theta(1 \notin [l_\alpha^+, n - l_\alpha^-] = \mathbb{P}_\theta(1 < l_\alpha^+) + \mathbb{P}_\theta(1 < l_\alpha^-) \\
&= \mathbb{P}_\theta(\max(p^{1/2}, p^{2/2}) < \alpha) + \mathbb{P}_\theta(\max(q^{1/2}, q^{2/2}) < \alpha) \\
&= \mathbb{P}((1-x_1, x_2) = (0, \alpha)) + \mathbb{P}((x_1, 1-x_2) = (\alpha, 0)) = 2 \times \alpha \times (1-\alpha) > \alpha.
\end{aligned}$$

F 应用：增强荟萃分析

图 9 展示了使用 Fisher 合并函数得到的 $n^+(\mathcal{I}_k)$ 同时 95% 置信区间，其中 $\mathcal{I}_k$ 是估计效应 $|y_i|$（绝对值）最大的前 $k = |\mathcal{I}_k|$ 所学校的索引。例如，在排名前 38 的学校中，我们观察到至少 9 个正效应和至少 1 个非正效应；在排名前 51 的学校中，我们观察到至少 10 个正效应和至少 2 个非正效应，依此类推。

G 补充模拟结果

对于正文第 6 节所述的 $n = 50$ 参数模拟设置，我们提供以下补充结果。

首先，对于 Guo 和 Romano（2015）的 FDR 控制程序 GR-FDR，我们给出图 10 所考虑设置下的实际覆盖保证。具体而言，我们给出估计概率（基于每种数据生成方式下的 2000 次模拟运行）：发现集中的正向发现数和非正向发现数分别不超过正参数数和非正参数数。覆盖率可能远低于 0.95 的覆盖保证，因为该程序仅保证 FDP 的期望不超过 α。这与所有其他能够提供 0.95 覆盖保证的程序形成对比。

其次，对于 DCT 和分割法，我们比较了四种合并方法：Fisher、ALRT、Simes 和 mSimes。图 11 显示，mSimes 优于 Simes：它给出的界限更紧，发现也更多。就界限而言，Fisher 表现最佳，ALRT 紧随其后。就发现而言，情况则相反（ALRT 得到的发现略多于 Fisher），但两者都远逊于 Simes。

最后，我们指出，我们还进行了其他模拟（因篇幅所限未予展示），在每种数据生成设置中改变参数值，并考虑更多配置——

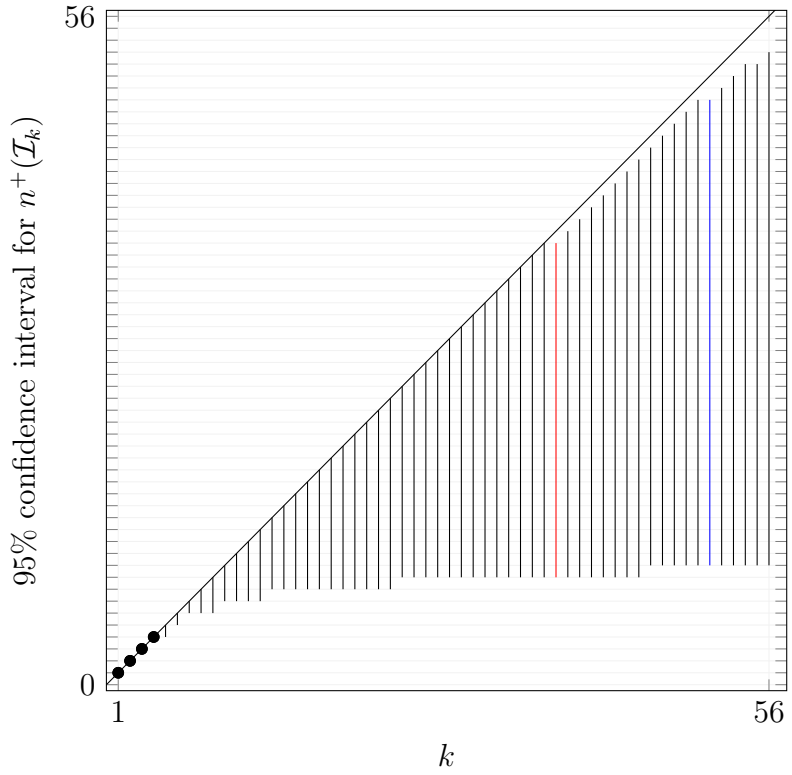


图 9：使用 Fisher 合并函数得到的 $n^+(\mathcal{I}_k)$ 同时 95% 置信区间，其中 $\mathcal{I}_k$ 是标准化效应（绝对值）最大的前 $k = |\mathcal{I}_k|$ 所学校的索引。例如，在前 38 所学校中，至少有 9 个正效应和至少 1 个非正效应（红色区间）；或者在前 51 所学校中，至少有 10 个正效应和至少 2 个非正效应（蓝色区间）。圆点表示退化区间 $[k, k]$。

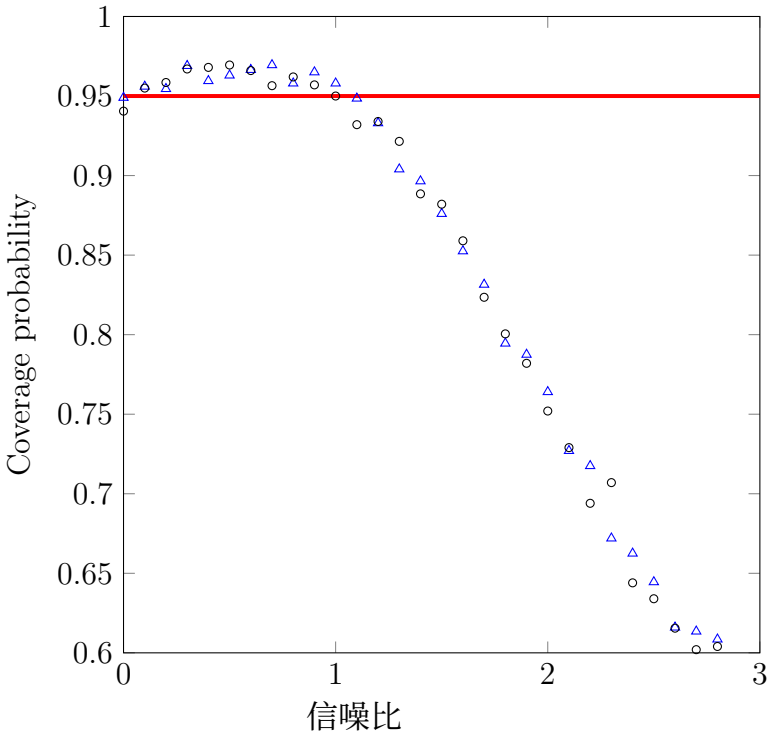


图 10：对于发现集为 $\mathcal{R}$ 的 GR-FDR程序，$n^+(\mathcal{R})$ 和 $n^-(\mathcal{R}) + n^0(\mathcal{R})$ 的覆盖概率随信噪比的变化，设置如下：$n^+ = n^- = 15$ （黑色圆点）；$n^+ = 30$ 和 $n^- = 0$ （蓝色三角形）。红色水平线表示目标覆盖概率 0.95。

n^+ 和 n^-的.....。上述关于不同合并方法相对性能的定性结论保持不变。

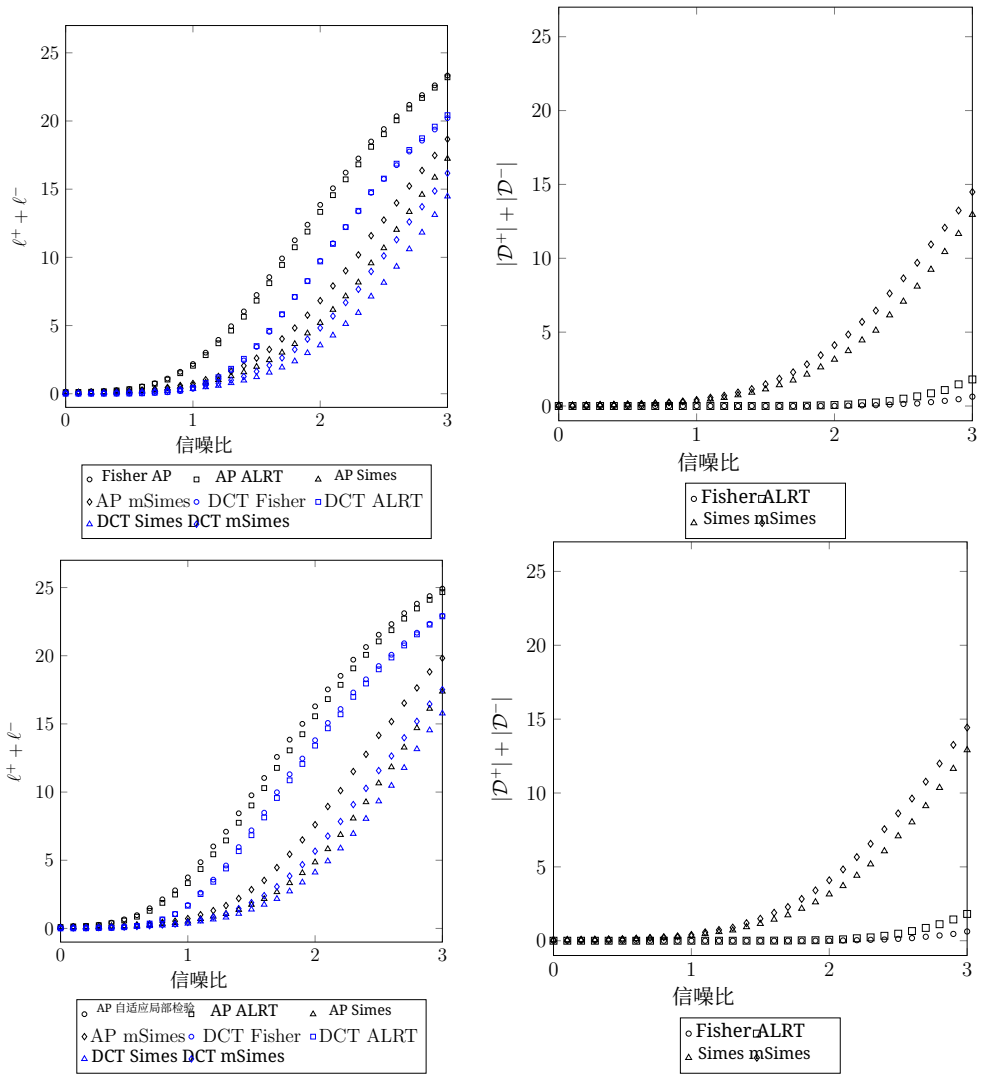


图 11：对于以下方法，正参数与非正参数下界之和的平均值（左列）以及被拒绝的个体假设数量（右列）随信噪比的变化：采用合并函数 Fisher（AP FISHER）、ALRT（AP ALRT）、Simes（AP SIMES）和修正 Simes（AP mSIMES）的 AP；采用合并函数 Fisher（DCT FISHER）、ALRT（DCT ALRT）、Simes（DCT SIMES）和修正 Simes（DCT mSIMES）的 DCT。 $n^+ = n^- = 15$ 在第一行； $n^+ = 30$ 和$n^- = 0$ 在第二行。右列中未显示 DCT程序，是因为它们与 AP程序相同。