arXiv:2303.07986v1 [math.ST] 14 Mar 2023

最强（更强）检验统计量的刻画及其简单非参数应用

Albert Vexler，纽约州立大学布法罗分校生物统计学系，纽约州布法罗；Alan D. Hutson，罗斯威尔帕克综合癌症中心生物统计学与生物信息学系，纽约州布法罗

2023年3月15日

摘要

数据驱动的最强检验是一种统计假设决策工具：在所有相应的、给定检验水平的基于数据的检验中，它们针对固定原假设具有最大的检验功效。当底层数据分布已知时，可应用似然比原理实施最强检验。反过来，我们考虑以下问题：(a) 假设给定一个检验统计量，例如 T，如何变换 T 以提高检验功效？(b) 能否利用 T 生成最强检验？(c) 如何根据所需最强决策程序的某一属性来比较检验统计量？为探讨这些问题，我们通过匹配刻画，将“最强”这一术语与给定检验统计量的分布性质建立一一映射。这种刻画形式具有实际应用价值，并与充分性的一般原理相吻合。研究结果表明，要改进给定的检验，可以采用相关辅助统计量；这些统计量相对于所检验的假设，其分布不会发生变化。作为示例，我们在非参数情形下通过修改常用的 t 检验来说明该方法。基于生成数据和一个真实数据集开展的数值研究证实，所提出的方法在实践中具有应用价值。

关键词：辅助统计量；似然比；最强检验；非参数检验；充分性；t 检验；中位数检验。

1 引言

用于开发和考察基于数据的决策机制的方法，已在理论和实验统计框架中得到广泛确立。评估现代基于数据的检验算法时，通常采用近一个世纪前制定的标准和基础。在此背景下，关于全面的综述及相关示例，请参阅 Lehmann 和 Romano（2005）。比较统计检验时通常采用的标准如下：1）将所考察检验的第一类错误率（TIE）设定为相同水平，例如α；2）比较这些检验的检验功效水平。这一经典原则主要由 J. Neyman 和 E. S. Pearson 在 $1928–1938$ 期间发表的一系列重要论文中提出并倡导（例如，Lehmann, 1993）。在这一框架中，似然方法与似然比概念相关联，使得决策任务中能够开发出强大的统计推断工具。

鉴于此，似然比原理可广泛应用于各种决策问题，尽管在许多实际应用中，似然比检验并未得到完整明确的定义。有些情况下，根据所采用的估计方案不同，估计得到的参数似然比统计量可能有不同的形式。另一些情况下，相关似然函数可能相当复杂，例如，观测值属于相关纵向数据，且受到某种缺失数据机制的影响。还有一些非参数情形，使我们难以写出相应的似然比统计量。要提出指导如何构造、修改和／或选择检验统计量的方针，就有必要系统研究似然比的内在性质。

不失一般性，为了简化对本文主要目标的说明，我们给出以下形式化记号。假设我们观测到数据 D，目的是检验简单原假设 H_0，以对抗其简单备择假设H_1。为此，我们令 $T = T(D)$ 表示基于D 的实值一维统计量；若 $T(D) > C$，则该统计量支持拒绝 H_0，其中 C 是固定的检验阈值。我们分别用 Pr_k 和 E_k 表示在 H_k、 $k \in \{0, 1\}$下的概率和期望。在本文的大部分内容中，我们假设概率分布 Pr_0, Pr_1 的

D 相对于定义在 Υ上的给定 σ有限测度 τ 绝对连续，其中 Υ 是空间（例如 $\mathbb{X}$）中的一个集的可加类， D 服从该空间上的分布。于是，相对于 τ，存在非负概率密度函数 f_0 和 f_1，并满足

$$\Pr_k\left(D \in \upsilon\right) \;=\; \int_{\upsilon} f_k(x) d\tau(x), \quad \upsilon \in \Upsilon, \quad k \in \{0,1\}.$$

注意， f_0 和 f_1 不必属于同一个参数分布族。定义 T的概率密度函数 f_0^T和 f_1^T，使得

$$\Pr_k\left\{T(D) \le t\right\} \;=\; \int_{\mathbb{X}} I\left(T(x) \le t\right) f_k(x) d\tau(x) \;=\; \int_{\mathbb{R}^1} I\left(u \le t\right) f_k^T(u) du,$$

其中， I(.)表示示性函数，假设 H_k 成立，且 $k \in \{0,1\}$。在此框架下，如果对于所有 $x \in \mathbb{X}$， f_0(x) > 0 且 f_1(x) > 0，则似然比 $\Lambda = \Lambda$(D) $= f_1$(D)/f_0(D) 是最强检验（MP）检验统计量。当随机观测数据 D 为多维时， f_0 和 f_1 是相对于 τ 的广义联合概率密度（例如，Lehmann，1950）。

为简单起见，假设在研究者计划采用检验统计量 S(D) $= L$(Λ(D))的情况下，其中 L(u)是单调递增函数，其反函数为 W(L(u)) $= u$，我们假设使用检验统计量 T(D) $= W$(S(D))，且 T(D)是最强检验。

我们的研究以 Λ 的以下性质为出发点，该性质见 Vexler and Hutson（2018年），此处列出是为了确保完备性。

命题 1.1。 似然比统计量 Λ 满足 $f_1^{\Lambda}(u) = u\, f_0^{\Lambda}(u)$，对所有 $u > 0$ 均成立。

证明推迟至在线补充材料中。

一个有趣的观察是，基于似然比 Λ 的似然比 $f_1^{\Lambda}/f_0^{\Lambda}$ 本身构成了似然比 Λ，即 f_1^{Λ}(Λ) /f_0^{Λ}(Λ) $= \Lambda$。考虑观察到某个统计量的取值或单个数据点（记为 X）的情形。为了针对 H_0 与 H_1 作出决策，对 X 进行的最佳变换是比值 f_1^X(X)/f_0^X(X)。在情形 $X = \Lambda$ 下，观测到的统计量无法改进，并且变换 f_1^X(X)/f_0^X(X) 在 $X = \Lambda$ 处可逆，因为似然比 Λ 是方程 f_1^X(X)/f_0^X(X) $= X$ 的一个根。

暂且假设我们可以通过在命题 1.1 的表述中加入“当且仅当”这一说法来改进该命题，从而断言：统计量 T 具有似然比形式，当且仅当 $f_1^T(u) = u\, f_0^T(u)$。那么，由于 Λ 是最强检验统计量，在给定检验统计量 T 的情况下，我们将尝试调整 T 的结构，使 $f_1^T(u)$ 与 $u\, f_0^T(u)$ 之间的距离最小，至少对某些 u 的取值如此。在这一框架中，比较两个检验统计量，例如 T 和 B，如果满足例如 $\max_u \left| f_1^T(u) - u\, f_0^T(u) \right| \leq \max_u \left| f_1^B(u) - u\, f_0^B(u) \right|$，我们便可以选择 T。请注意，在非参数情形下，我们可以在许多情景中近似和/或估计检验统计量的分布函数。遗憾的是，命题 1.1 的简单表述无法用于刻画最强检验统计量。例如，当 $D = (X_1, X_2)$ 时，其中 X_1 服从正态分布，且 $\mathrm{E}_0(X_1) = 0$、$\mathrm{E}_1(X_1) = \mu$，以及 $\mathrm{E}_0(X_1^2) = \mathrm{E}_1(X_1 - \mu)^2,$，统计量 $T = f_1(X_1) / f_0(X_1)$ 满足 $f_1^T(u) = u\, f_0^T(u)$，而最强检验统计量为 $f_1(X_1, X_2) / f_0(X_1, X_2)$。也就是说，要刻画最强检验，必须修改命题 1.1。

注释 1.1. 本文避免使用“一致最强”这一术语，以便研究参数在检验程序中并不起关键作用的情形，同时也考虑一致最强检验不存在的情况，例如非参数检验命题。例如，设 H_0 推断观测值来自标准正态分布，而非标准逻辑斯蒂分布，条件为 H_1。在这种情况下，我们有似然比最强检验，但使用“一致最强”一词可能会造成误导。在第 3 节中，最强概念具有一种假设性的背景，我们旨在通过构建检验程序来逼近这一背景。（在这方面，也请参见注释 2.2 中的第 (iii) 项以及第 6 节中的注 (d)。）

本文的目标有两方面：(1) 通过扩展命题 1.1，我们描述一种刻画最强检验的方法；(2) 我们通过修正检验统计量，展示最强检验的理论刻画在实践中的实用性，而这种修正很容易应用。基于这一框架，第 2 节探讨检验统计量的分布性质与其成为最强检验能力之间的一一映射。我们证明的这一刻画，与 Bahadur（1955）在若干决策问题中广泛评估的充分性原理相一致。第 3 节中，为了展示所提出的最强检验刻画的潜在用途，我们运用第 2 节介绍的理论概念

来说明一种高效的原则：通过运用相关的辅助统计量，改进常用的检验程序。辅助统计量的分布不依赖于竞争假设。然而，我们表明，辅助统计量能够对有关目标假设的推断作出重要贡献。例如，尽管似乎很难胜过众所周知的单样本均值 t 检验，但我们认为，简单修正 t 检验统计量即可提高检验功效。具体而言，可以考虑总体偏度对样本均值分布的影响。第 3 节展示了在不存在最强检验决策机制时，如何在非参数假设下实施检验程序修正。随后，第 4 节通过实验评估证实，所提出的方案在各种情形下效率很高。此外，正如第 5 节所述，在分析一项与心肌梗死相关的生物标志物研究数据时，与已知方法相比，第 3 节提出的用于单样本中位数检验的方法，对检测数据分布中的不对称性更为敏感。最后，本文在第 6 节作总结讨论。

2 刻画与充分性

为了解本节的目的，下面给出一个说明性示例。

2.1 单样本均值检验

设 X_1, X_2 是来自均值为 μ、方差为 $\sigma^2 = 1$ 的正态总体的随机样本。我们考虑检验 H_0：$\mu = 0$ 对比 H_1：$\mu = \delta$，其中 δ 是一个固定值。我们给出这个简单的示例，旨在说明第 2.2 节中的结果，而非提供一种可与常规 t 检验相提并论的方法。在此例中，统计量 $X = (X_1 + X_2)/2$ 可用于最强检验。不过，对于评估竞争假设 H_0 和 H_1，有人可能会认为，例如统计量 $A = wX_1 + (1-w)X_2$ 是合理的，其中 $w \in [0, 1]$、$w \neq 0.5$ 满足某些条件。向量 $[\bar{X}, A]^\top$ 服从二元正态分布，其密度函数为二元正态密度函数，并且 $\mathrm{E}\left(\bar{X}\right) = \mathrm{E}(A) = \mu$、$\mathrm{var}\left(\bar{X}\right) = \mathrm{cov}\left(\bar{X}, A\right) = 1/2$，以及 $\mathrm{var}(A) = w^2 + (1-w)^2$。接下来，通过定义一个联合

基于 (X_1, X_2) 的统计量 A_1, A_2 的联合密度函数，形式为 $f^{A_1,A_2}_{\mu}(u,v)$，很容易看出，比值 $f^{\bar{X},A}_{\mu=\delta}(u,v)/f^{\bar{X},A}_{\mu=0}(u,v)$ 不依赖于 v，其中 v 是与 A 的分量相关的联合密度 $f^{\bar{X},A}_{\mu}$ 的自变量。特别是，这意味着，在考察两个数据点 $\bar{X}$ 和 A 后，我们可以通过构造用于检验 H_0 对比 H_1 的似然比形式的最强统计量 $f^{\bar{X},A}_{\mu=\delta}(\bar{X},A)/f^{\bar{X},A}_{\mu=0}(\bar{X},A)$，改进基于 $\left(\bar{X},A\right)$ 的决策机制；这只需要计算 $\bar{X}$。因此，我们不妨提出这样一个问题：上述观察能否推广，从而扩展命题 1.1？在这种情况下，若要给出最强概念的一项本质性质，似乎应当考虑与其他基于 D 的统计量之间的关系。

我们方法的第二个方面是刻画这样一种情景：例如，在构造检验时，统计量 A_1比统计量 A_2更可取。在此背景下，后文将会看到，如果比值 $f^{A_1,A_2}_{\mu=\delta}(u,v)/f^{A_1,A_2}_{\mu=0}(u,v)$ 不依赖于v，那么 A_1优于 A_2。为了说明这条规则的益处，我们不妨假设，在本小节中我们并不知道 X 是最佳统计量，从而可以考虑以下问题：在统计量 $T = aX_1 + (1-a)X_2$ 中，$a \in [0,1]$ 应取何值，才能使T对所有 $w \in [0,1]$均优于 $A = wX_1 + (1-w)X_2$？由于密度 $f^{T,A}_{\mu}(u,v)$ 是二元正态分布，简单的代数运算表明，若 $a^2 + (1-a)^2 = aw + (1-a)(1-w)$，则 $f^{A_1,A_2}_{\mu=\delta}(u,v)/f^{A_1,A_2}_{\mu=0}(u,v)$ 不依赖于 v。因此，解为 $a = 0.5$。事实上，我们知道，在这个框架中，T在 $a = 0.5$的条件下是最佳统计量。这个例子说明了我们的观点。

2.2 理论结果

本节将在下文的命题 2.1−2.5 中给出主要结果，从而确立最强检验的刻画。为确保内容完整，命题 2.1−2.5 的证明收录于在线补充材料中，并附有评论，以补充说明所得结果。命题 2.5 从充分性的角度重新审视了最强检验的刻画。

为推广命题 1.1，我们定义统计量 $A = A(D)$ 和 $B = B(D)$ 的联合密度函数，其形式为 $f^{A,B}_{\cdot,k}(u,v)$，前提是 H_k为真，$k \in \{0,1\}$。然后，似然比检验统计量 Λ 具有以下性质。

命题 2.1。 对于任意统计量 $A = A(D)$，我们有 $f^{\Lambda,A}_1(u,v) = u\, f^{\Lambda,A}_0(u,v)$，对所有 $u \geq 0$ 和 $v \in \mathbb{R}^1$均成立。

命题 2.1 是命题 1.1 的推广，因为 $f^{\Lambda,A}_1(u,v) = u\, f^{\Lambda,A}_0(u,v)$ 得出

$$f_1^{\Lambda}(u) = \int f_1^{\Lambda,A}(u,v)dv = u \int f_0^{\Lambda,A}(u,v)dv = u\, f_0^{\Lambda}(u).$$

下述论断表明，命题 2.1 可以进一步扩展，从而给出检验统计量分布能够产生最强决策方法的充要条件。定义 C 为检验阈值。

命题 2.2. 假设用于检验的统计量 $T = T(D)$ 满足 $\Pr_k(T \geq 0) = 1, k \in \{0,1\}$，当 $T > C$ 时拒绝 H_0。当且仅当（*iff*）对于任意统计量 $A = A(D)$ 及所有 $u \geq 0,\ v \in \mathbb{R}^1$，$T$ 是最强检验，$f^{T,A}_1(u,v) = u\, f^{T,A}_0(u,v)$。

请注意，Fisher 和 Robbins（2019年）采用了“T 严格非负”这一条件；在相应的第一类错误率渐近受控于 α 时，对检验统计量 T 进行单调对数变换，可能会提高基于 T 的检验功效。不过，要求 $T \geq 0$ 并非关键，因为如果我们评估的检验统计量（例如 $G = G(D)$）可能为负，那么单调变换 $T = g(G) \geq 0$ 可以在这种情况下发挥作用。

值得指出的是，在检验统计量的密度函数不存在的情景下，可以运用命题 2.2 证明中采用的论证，得到下面的结论。

命题 2.3. 检验统计量 $T(D) > 0$ 是最强检验，当且仅当对于每个关于 D 的函数 $g \in [0,1]$，均有 $\mathrm{E}_1\{g(D)\} = \mathrm{E}_0\{g(D)T(D)\}$。

注 2.1. 由于 $\mathrm{E}_1\{g(D)\} = \mathrm{E}_0\{g(D)\Lambda(D)\}$，条件 $\mathrm{E}_1\{g(D)\} = \mathrm{E}_0\{g(D)T(D)\}$意味着，对于每个 $g \in [0,1]$，都有 $\mathrm{E}_0[\{T(D) - \Lambda(D)\}\, g(D)] = 0$；这表示在 f_0 下，概率为一时有 $T = \Lambda$。还请注意，在命题 2.3 中，我们可以使用满足以下条件的 $g(D)$：对于所有 $m > 0$，均有 $\mathrm{E}\{g(D)\}^m = \mathrm{E}\{g(D)\}$。

命题 2.2 的步骤（2）证明中采用的方案得出以下结果。

命题 2.4。若对于所有 $u \geq 0, v \in \mathbb{R}^1$，比值 $f^{T_1,1}{}^{T_2}(u,v)/f^{T_1,0}{}^{T_2}(u,v) = u$，则统计量 $T_1 \geq 0$ 比统计量 T_2 更具检验功效。

值得注意的是，根据命题 1.1 和命题 2.2，对于任意基于 D 的统计量 $A = A(D)$，我们有 $f^{T,1}{}^{A}(u,v) = f^{T,0}{}^{A}(u,v)\ u$，于是 $f^{T,1}{}^{A}(u,v) = f^{T,0}{}^{A}(u,v) f_1^T(u)/f_0^T(u)$，若 T 是最强检验。也就是说，$f_1^{A|T}(u,v) = f_0^{A|T}(u,v)$，其中记号 $f_k^{A|T}$ 表示在 H_k，$k \in \{0,1\}$ 下给定 T 时 A 的条件密度函数。在这种情况下，当 A 与 T 独立时，我们得到 $f_1^A(v) = f_0^A(v)$，于是 $A = A(D)$ 无法区分这些假设。我们可以说 $A = A(D)$ 是辅助统计量，这意味着 $f_1^A = f_0^A$。这促使我们将上述结果与充分性原理联系起来。

根据 Bahadur（1955），在所考虑的框架中，如果对于每个 $A = A(D)$ 以及所有 $u \geq 0, v \in \mathbb{R}^1$，$f_1^{A|T}(u,\ v) = f_0^{A|T}(u,\ v)$，那么我们可以称 $T = T(D)$ 为充分检验统计量。在此背景下，上述陈述得出了以下结果。

命题 2.5. 以下说法等价：

(i) $T = T(D)$ 是充分的，且 $f_1^T(u)/f_0^T(u) = u$;

(ii) T 是用于检验竞争假设 H_0 和 H_1的最强统计量。

命题 2.5 论证了仅依据相应的充分统计量进行统计推断的合理性。

注释 2.2. 我们可以注意到以下事实：

(i) Kagan 和 Shepp（2005）举例说明了充分性悖论：一个不充分统计量仍能保留 Fisher 信息。(ii) 为了扩展命题 2.5，在某些情况下，可以采用与各种 Basu 定理类型结果相关的论述（例如，Ghosh，2002年）。据我们所知，Basu 定理尚未被直接应用于本文所探讨的问题。(iii) 在贝叶斯检验方法中（例如，Johnson，2013），命题 1.1 可以扩展至处理贝叶斯因子；在此背景下，可参见 Vexler（2021）的命题 5。然后，

命题 2.1 和命题 2.2 可以轻松修改，以便结合纳入的先验信息建立集成 MP 检验（Vexler 等人，2010年）。

3 应用

第 2 节阐述了 MP 刻画概念背后较为一般的理论框架。本节通过以易于实施的方式改进广受认可的统计检验，概述所提出的 MP 刻画原则的三个应用。希望所提出的 MP 刻画能够为统计实践中的决策算法开发、改进和比较带来不同的益处。

统计推断中一个常见的问题是，需要对给定的检验统计量进行调整，以提高相应决策方案控制 TIE率和检验功效的表现。例如，采用 Bartlett 校正类机制和／或位置调整技术，可以提高检验统计量零假设分布的渐近近似精度。在此背景下，我们请读者参阅以下示例：Hall 和 La Scala（1990）研究了如何调整非参数经验似然比；Chen（1995）介绍了用于评估非对称分布均值的 t 检验统计量的不同变换。最近，Fisher 和 Robbins（2019）提出使用对数变换，以期提高基于变换后统计量的检验的检验功效。

本节将按照下述简单思路，展示所考虑的 MP 原则的用法。假设我们有一个合理的检验统计量 T_o，并希望将其改进为T_o的形式，例如 T_N，使其近似满足以下断言：$f^{T_N}_{1}{}^{A}(u,v) = u\, f^{T_N}_{0}{}^{A}(u,v)$，对任意统计量 $A = A(D)$ 和所有 $u \geq 0,\ v \in \mathbb{R}^1$ 均成立。鉴于一般的非参数设定下不存在最强检验，若能至少针对某个统计量（尤其是某个辅助统计量）达到最强性质$f^{T_N}_{1}{}^{A}(u,v) = u\, f^{T_N}_{0}{}^{A}(u,v)$，将会很有吸引力。通俗地说，若有 A，且$f_1^A = f_0^A$，我们便能消除 A对 T_o的影响，从而构造出 T_N，使比值 $f^{T_N}_{1}{}^{A}(u,v)/f^{T_N}_{0}{}^{A}(u,v)$ 仅是 u 的函数。在这种情况下，命题 2.4 可以保证 T_N 优于 A。这可以通过 T_N 与 A 相互独立来实现，具体示例详见第 3.3 、 3.4 和 3.5 节。

通过 以下示例，我们旨在以直观的方式展示我们的方法。

3.1 辅助统计量的使用示例

We 先展示下面的玩具示例，以说明我们的核心思路。

观测到独立数据点 X_1 和 X_2；假设 $X_i \sim N(\mu,\ \sigma_i^2)$，且已知 $i \in [1,2]$ 和 $\sigma_1^2 \neq \sigma_2^2$。可以使用简单统计量 $T = 0.5(X_1 + X_2)$ 检验 H_0：$\mu = 0$，对比 H_1：$\mu > 0$。容易验证，$X_1 - X_2$ 是一个辅助统计量。现在我们考虑一种变换 T 并构造一个与 $X_1 - X_2$独立的修正检验统计量的方法。定义 $T_N = T + \gamma(X_1 - X_2)$，它是 T的变换版本，其中 γ是方程 $\mathrm{cov}\,(T_N,\ X_1 - X_2) = 0$ 的一个根。于是，我们得到 $\gamma = 0.5\,(\sigma_2^2 - \sigma_1^2)\ /(\sigma_2^2 + \sigma_1^2)$。因此，所得统计量

$$T_N \ = \ T + \frac{\sigma_2^2 - \sigma_1^2}{2\,(\sigma_2^2 + \sigma_1^2)}\,(X_1 - X_2) = \frac{\sigma_2^2 X_1 + \sigma_1^2 X_2}{\sigma_2^2 + \sigma_1^2} \ = \ \frac{X_1/\sigma_1^2 + X_2/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2}$$

这无疑是对初始统计量 T 的成功变换，它给出了最强检验统计量。例如，我们将检验功效记为

$P(a) = \mathrm{Pr}_{\mu=5}\{T + a(X_1 - X_2) > C(a)\}$, $C(a) : \mathrm{Pr}_{\mu=0}\{T + a(X_1 - X_2) > C(a)\} = 0.05$, 当 $\sigma_1 = 1$ 且 $\sigma_2 = 4$时。图 1(a) 展示了函数 $P(a) - P(0)$，即基于（$T + a(X_1 - X_2)$）的检验与基于 T 的检验在 $\alpha = 0.05$ 处的检验功效水平之差，该差值以函数 $Cov(a) = \mathrm{cov}\,(T + a(X_1 - X_2)\ X_1 - X_2)$ 为横轴绘制，其中 $a \in [-0.01, 0.9]$。正如预期，函数 $P(a) - P(0)$ 在$Cov(a) = 0$时达到最大值。此外，事实证明，使用修正检验统计量 $T + a(X_1 - X_2)$ 时，我们无需非常精确地近似方程 $Cov(a) = 0$，就能胜过基于 T 的检验。由此，我们可以直观地推断，变换后的检验统计量可以包含估计得到的元素，同时仍能使其决策算法具有良好的检验功效特性。

在各种情形下，我们不排除针对给定检验陈述存在多个辅助统计量的可能性。举例如下：假设我们分别观测到来自正态分布 $N(\mu, \sigma_1^2)$、$N(\mu, \sigma_2^2)$ 和 $N(\mu, \sigma_3^2)$ 的 X_1、X_2 和 X_3，其中 σ_i^2、$i \in [1,2,3]$ 已知。假设我们关注检验 H_0：$\mu = 0$ 对 H_1：$\mu > 0$。待修改的统计量为 $T =$

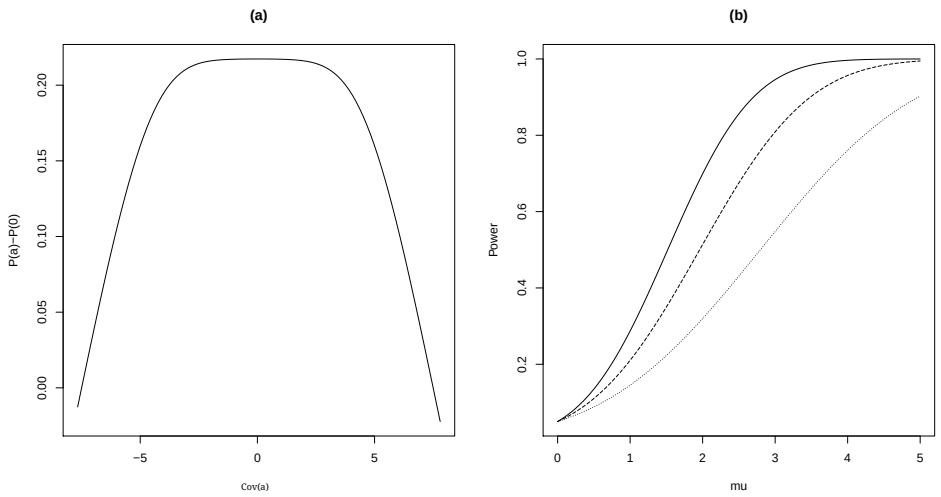


图 1：与第3.1节所示示例相关的图形评估。面板 (a) 绘出 $P(a) - P(0)$，即在 $\alpha = 0.05$ 水平下，基于 $(T + a(X_1 - X_2))$ 的检验的检验功效减去基于 T 的检验的检验功效，并以协方差 $Cov(a)$ =cov $(T + a(X_1 - X_2)\ X_1 - X_2)$ 为横轴，适用于 $a \in [-0.01, 0.9]$，其中 $T = 0.5(X_1 + X_2)$，$X_1 \sim N(\mu, 1)$，$X_2 \sim N(\mu, 42)$，$\mathrm{E}_0(X_i) = 0$，$\mathrm{E}_1(X_i) = 5$，$i \in [1, 2]$。面板 (b) 绘出在 $\alpha = 0.05$ 水平下的检验功效 $P_{T_O}(\mu) = \mathrm{Pr}_\mu\left\{T_O > C_\alpha^{T_O}\right\}$（实线）、$P_{T_N}(\mu) = \mathrm{Pr}_\mu\left\{T_N > C_\alpha^{T_N}\right\}$（长虚线）以及$P_T(\mu) = \mathrm{Pr}_\mu\left\{T > C_\alpha^T\right\}$（点线），其中 $T = (X_1 + X_2 + X_3)/3$，$T_N = T + \gamma(X_1 - X_2)$，$T_O = (X_1/\sigma_1^2 + X_2/\sigma_2^2 + X_3/\sigma_3^2)\ /(1/\sigma_1^2 + 1/\sigma_2^2 + 1/\sigma_3^2)$，$\gamma = (\sigma_2^2 - \sigma_1^2)\ (\sigma_2^2 + \sigma_1^2)^{-1}/3$，$X_1 \sim N(\mu, 1)$，$X_2 \sim N(\mu, 42)$，且 $X_3 \sim N(\mu, 32)$，适用于 $\mu \in [0, 5]$。

$(X_1 + X_2 + X_3)/3$。观测值 $X_1 - X_2$ 相对于 μ 是一个辅助统计量。定义 $T_N = T + \gamma(X_1 - X_2)$，其中 $\gamma = (\sigma_2^2 - \sigma_1^2)\ (\sigma_2^2 + \sigma_1^2)^{-1}/3$，由此得到 cov $(T + \gamma(X_1 - X_2)\ X_1 - X_2) = 0$。那么，显然 T_N 略优于 T，但就本例而言，$T_O = \sum_{i=1}^3 (X_i/\sigma_i^2)\ / \sum_{i=1}^2 (1/\sigma_i^2)$ 优于 T_N。定义检验功效 $P_T(\mu) = \mathrm{Pr}_\mu\left\{T > C_{0.05}^T\right\}$、$P_{T_N}(\mu) = \mathrm{Pr}_\mu\left\{T_N > C_{0.05}^{T_N}\right\}$ 和 $P_{T_O}(\mu) = \mathrm{Pr}_\mu\left\{T_O > C_{0.05}^{T_O}\right\}$，其中检验阈值 $C_{0.05}^T$、$C_{0.05}^{T_N}$ 和 $C_{0.05}^{T_O}$ 满足 $\mathrm{Pr}_{\mu=0}\left\{T > C_{0.05}^T\right\} = \mathrm{Pr}_{\mu=0}\left\{T_N > C_{0.05}^{T_N}\right\} = \mathrm{Pr}_{\mu=0}\left\{T_O > C_{0.05}^{T_O}\right\} = 0.05.$。图 1(b) 展示了当 $\sigma_1 = 1$、$\sigma_2 = 4$ 和 $\sigma_3 = 3$ 时，函数 $P_T(\mu)$、$P_{T_N}(\mu)$ 和 $P_{T_O}(\mu)$ 的表现。

因此，为了改进给定的检验统计量（例如 T），我们可以建议关注

相关的辅助统计量（例如 A），并对 T 进行修改，使其独立（或近似独立）于 A。

请注意，尽管辅助性这一概念指出，辅助统计量不会提供有关感兴趣参数的信息，但文献中已广泛探讨辅助统计量在参数估计中的不同作用。关于这一主题的全面综述，读者可参阅 Ghosh 等（2010）。例如，假设我们观测到来自二元正态分布的向量 $[X_i, Y_i]^\top$, $i \in [1, \ldots, n]$，其 E(X_1) = E(Y_1) = 0、var(X_1) = var(Y_1) = 1，且 corr(X_1, Y_1) = ρ，其中 $\rho \in (-1, 1)$ 未知。统计量 $U_1 = \sum_{i=1}^n X_i^2$和 $U_2 = \sum_{i=1}^n Y_i^2$是辅助统计量。根据 Ghosh 等（2010），为了定义 ρ的无偏估计量，可以建议使用统计量 $\sum_{i=1}^n X_i Y_i / U_j$, $j \in [1, 1]$。再举一例，在应用辅助统计量时，我们概述一种所谓的蒙特卡洛欺骗：这类基于模拟的方法只需生成少量样本，就能达到通常需要生成大量样本才能获得的统计精度。Boos 和 Hughes-Oliver（1998）讨论了以下程序。为了估计正态分布样本 $X_1, \ldots, X_n$的样本中位数 M的方差，蒙特卡洛欺骗方法使用 N个来自 $X_1, \ldots, X_n$ 的蒙特卡洛样本，估计 var $\left(M - \bar{X}\right)$（而非 var (M)），然后加上 var $\left(\bar{X}\right)$ = var (X_1) $/n$，从而得到 var (M)的有效估计值。为了证明这一框架合理，我们利用统计量 $V = (X_1 - X, \ldots, X_n - X)$ 是辅助统计量这一点。现在，由于 X 是完全充分的，根据 Basu 定理， X 和 V 相互独立。于是，X 与 V的样本中位数相互独立。因此，

$$\operatorname{var}(M) = \operatorname{var}\left(M - \bar{X} + \bar{X}\right) = \operatorname{var}\left(M - \bar{X}\right) + \operatorname{var}\left(\bar{X}\right).$$

显然，当 X_1 服从正态分布时，var (X_1) $/n$对 var (M) 的贡献远大于 var $\left(M - \bar{X}\right)$ 的贡献，而 var $\left(M - \bar{X}\right)$ 这一分量被建议通过模拟进行估计。这使得模拟估计的误差仅占 var (M) 的很小一部分（详情参见 Boos 和 Hughes-Oliver，1998 年）。

3.2 理论支持

上述观点可由以下结果支持。假设我们有一个检验统计量 $Y = Y(D)$，且比值 $L^Y(u) = f_1^Y(u)/f_0^Y(u)$ 是一个单调递增函数

并且存在反函数，记为 $W(u)$。在这种情景下， Y 可转化为 $Y_N = L^Y(Y)$ 的形式，从而意味着

$$f_k^{Y_N}(u) = \frac{d}{du}\Pr_k\left\{L^Y(Y) \le u\right\} = \frac{d}{du}\Pr_k\left\{Y \le W(u)\right\} = f_k^Y\left(W(u)\right)\frac{d}{du}W(u),\ k \in \{0,1\}.$$

这意味着似然比

$$f_1^{Y_N}\left(Y_N\right)/f_0^{Y_N}\left(Y_N\right) = f_1^Y\left(W(Y_N)\right)/f_0^Y\left(W(Y_N)\right) = f_1^Y\left(Y\right)/f_0^Y\left(Y\right) = Y_N.$$

（在此语境下，参见命题 1.1。）接下来，我们陈述一个命题。设统计量$A = A(D)$满足 $f_1^A = f_0^A$。假设我们有一个基于统计量 $T = T(D)$的决策程序，并且可以将 T 修改为 T_N，从而在 H_0和 H_1下，当 $T = \psi(T_N,\ A)$时，使 T_N 和 A 成为独立项，其中 ψ为二元函数。由此得出：

命题 3.1。 上文讨论的基于 T_N的检验优于基于 T的检验，若比值 $L^{T_N}(u) = f_1^{T_N}(u)/f_0^{T_N}(u)$ 是单调函数。

证明见在线补充材料。

请注意，命题 3.1 揭示了统计检验的功效与辅助性之间的一些联系。这两个概念看似无关，因为辅助统计量本身无法区分竞争假设 H_0 和 H_1。

命题 3.1 阐明了修改以下著名检验统计量的理由。

3.3 均值的单样本 t 检验

假设我们观测到独立同分布（i.i.d.）的数据点 $X_1,\ X_2,\ \ldots,\ X_n$，它们提供了 $D = \{X_1, \ldots, X_n\}$。对于检验假设 $H_0:\ \mu = 0$ 与 $H_1:\ \mu > 0$，其中 $\mu = \mathrm{E}X_1$，广为接受的统计量为 $T_o = n^{0.5}\bar{X}/\sigma$，其中 $\bar{X} = \sum_{i=1}^n X_i/n$，且$\sigma^2 =var(X_1)$。

在这一检验陈述中，统计量 $S^2 = \sum_{i=1}^n \left(X_i - \bar{X}\right)^2/(n-1)$（样本方差）似乎是关于 μ 的近似辅助统计量。另请注意，当 $X_1 \sim N(\mu, \sigma^2)$ 时， T_o 和 S^2 相互独立（在此情形下，T_o 是最强检验）。于是，我们将统计量 $T_N(\gamma)$ 记为 $= T_o + \gamma S^2$，并推导出一个 γ 的取值，比如 γ_0，以确保 cov $(T_N(\gamma_0)\ S^2) = 0$。

统计量 $T_N(\gamma_0)$ 是我们将要提出的修正检验统计量的基本组成部分。为此，我们定义 $\mu_k = \mathrm{E}\,(X_1-\mu)^k$, $k=3,4$，并运用 O'Neill (2014) 的结果来获得

$$\gamma_0 \;=\; \sigma^{-1}\mu_3 n^{0.5}/\left(\sigma^4\frac{n-3}{n-1}-\mu_4\right) \;\approx\; -\sigma^{-1}\mu_3 n^{0.5}/\mathrm{var}\left\{(X_1-\mu)^2\right\}.$$

在针对 H_0 与 H_1 进行检验时，使用 $T_N(\gamma_0)$ 的额外理由，可以用以下简单方式解释。显然，所述检验问题可以置于对 μ 进行置信区间估计的背景下处理。因此，检验 H_0: $\mu=0$ 的质量与相应决策方案中涉及的 μ 估计量的方差之间存在联系。（例如，t 检验 T_o 使用 X 来估计 μ。）为简单起见，考虑满足 $\mathrm{E}\,\tilde{T}_N(\gamma)=\mu$ 且 $\mathrm{var}\left\{\tilde{T}_N(0)\right\}=\mathrm{var}\left(\bar{X}\right)$ 的 $\tilde{T}_N(\gamma)=\bar{X}+\gamma(S^2-\sigma^2)$。为找到使 $\mathrm{var}\left\{\tilde{T}_N(\gamma)\right\}$ 最小的 γ 值，我们可以求解方程 $d\left[\mathrm{var}\left\{\tilde{T}_N(\gamma)\right\}\right]/d\gamma=0$，其中假设我们可以写成

$$\begin{aligned}\frac{d}{d\gamma}\mathrm{var}\left\{\tilde{T}_N(\gamma)\right\} &= \frac{d}{d\gamma}\mathrm{E}\left\{\bar{X}+\gamma\left(S^2-\sigma^2\right)-\mu\right\}^2 = \mathrm{E}\frac{d}{d\gamma}\left\{\bar{X}+\gamma\left(S^2-\sigma^2\right)-\mu\right\}^2\\ &= 2\mathrm{E}\left\{\bar{X}+\gamma\left(S^2-\sigma^2\right)-\mu\right\}\left(S^2-\sigma^2\right).\end{aligned}$$

那么，根

$$\gamma=-\mathrm{cov}\left(\bar{X},S^2\right)/\mathrm{var}\left(S^2\right)=\mu_3/\left\{\sigma^4\left(n-3\right)/\left(n-1\right)-\mu_4\right\}\approx-\mu_3/\mathrm{var}\left\{(X_1-\mu)^2\right\}$$

使 $\mathrm{var}\left\{\tilde{T}_N(\gamma)\right\}$ 最小，且为 $\gamma_0\sigma/n^{0.5}$，其中 γ_0 意味着 $\mathrm{cov}\,(T_N(\gamma_0)\;S^2)=0$。也就是说，$\mathrm{var}\left\{\tilde{T}_N(\gamma_0\sigma/n^{0.5})\right\}\le\mathrm{var}\,\left(\bar{X}\right)$。统计量 $T_N(\gamma_0)$ 包含乘以 $n^{0.5}/\sigma$ 的 $\bar{X}$。这证实了，对于各种 X_1 的分布情景，在检验 H_0 与 H_1 时，统计量 $T_N(\gamma_0)$ 的检验功效可能略高于 T_o。

最后，我们将检验统计量 $T_N(\gamma_0)$ 标准化，以便控制其 TIE率，并将其记为 α 水平的决策规则：如果满足以下条件，则拒绝原假设 H_0

$$T_N=\hat{\Delta}^{-0.5}\left[T_o-\frac{\hat{\mu}_3 n^{0.5}}{\sigma\,\hat{\mathrm{var}}\left\{(X_1-\mu)^2\right\}}\left(S^2-\sigma^2\right)\right]>z_\alpha,$$

其中，$\hat{}\mu_3=n^{-1}\sum_{i=1}^n\left(X_i-\bar{X}\right)^3$ 和$\hat{}\mathrm{var}\left\{(X_1-\mu)^2\right\}=n^{-1}\sum_{i=1}^n\left\{\left(X_i-\bar{X}\right)^2-\sigma^4\right\}^2$ 分别估计 μ_3 和$\mathrm{var}\left\{(X_1-\mu)^2\right\}$;

$$\hat{\Delta}=1-S^{-2}\hat{\mu}_3^2\left[n^{-1}\sum_{i=1}^n\left\{\left(X_i-\bar{X}\right)^2-S^4\right\}^2\right]^{-1}$$

是 $\Delta =$ var ($T_N(\gamma_0)$)的样本估计量；阈值 z_α 满足 Pr($Z > z_\alpha$) $= \alpha$，其中 $Z \sim N(0,\ 1)$。将 T_N改写为 $T_N = n^{0.5}\bar{Y}/\sigma$的形式可能很有意思，其中$Y_i = \hat{}\Delta^{-0.5}\left[X_i - \hat{}\right.$ $\left.\mu_3\left\{\left(X_i - \bar{X}\right)^2 n/(n-1) - \sigma^2\right\} / \hat{}\text{var}\ \left\{(X_1 - \mu)^2\right\}\right]$。在 H_0下，统计量 T_N渐近服从 N (0，1)分布。当然，在 $X_1 \sim N(\mu,\ \sigma^2)$的情形下，也就是说 T_o 是最强检验，我们有 $\mu_3 = 0$。采用 T_N的形式，实际上对底层数据的偏度进行了调整。

请注意，在上述检验统计量的变换中，我们实现了 T_N 与 S^2不相关，从而简化了非参数程序的构建。显然，等式 cov (T_N，S^2) $= 0$ 对于 T_N 与 S^2 的渐近独立性至关重要（例如，Ghosh 等人，2021年，第 181–206 页）。

第 4 节通过广泛的蒙特卡罗评估，展示了统计量 T_N 在检验 H_0 以对抗各种备择假设时的效率。

基于 T_o 和 T_N 的检验程序要求已知 σ。例如，可以采用自助法类型的策略、贝叶斯技术和/或 Bayarri 和 Berger（2000年）提出的基于 p 值的方法，来克服这一限制。本文仅指出，在一些实际应用中，假设 σ 已知是合理的，例如 Maity 和 Sherman（2006年）、Boos 和 Hughes-Oliver（1998年，第 3.2 节）以及 Johnson（2013年，第 1729 页）。在许多生物统计学研究中，生物标志物的取值经过缩放，使其方差 $\sigma^2 = 1$。我们还指出，改进 t 检验 T_o的简单检验统计量的发展，可能具有理论意义。

3.4 单样本中位数检验

关于均值效应与分位数效应比较的一个子问题可以表述如下：设 $X_1, X_2 \ldots, X_n$ 为连续独立同分布观测值，且 E$X_1 = 0$。我们感兴趣的是检验假设 H_0：$\nu = 0$，对比 H_1：$\nu > 0$，其中 ν 表示 X_1 的中位数。这个问题表述见于各种实际应用，例如检验线性回归残差是否对称、对安慰剂使用前后生物标志物测量值进行配对比较；又如，研究人员调查与饮用水放射性检测相关的数据，且总体均值已知；参见 Semkow 等人（2019年）第 4 节。

要检验 H_0，合理的做法是使用以下形式的统计量

$$T_o = 2n^{0.5} X_{(n/2)} f\left(X_{(n/2)}\right),$$

其中，$X_{(n/2)}$ 是基于次序统计量 $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ 的 ν 样本估计量，$1/f\left(X_{(n/2)}\right)$ 是尺度度量，f（u）是 X_1 的密度函数。统计量 $A = n^{0.5}\bar{X}/\sigma$ 可选为相对于 ν 近似辅助的统计量，因为 $\mathrm{E}X_1 = 0$。我们使用的已知事实是：(a) 统计量 T_o 和 A 渐近服从二元正态分布，其参数见 Ferguson (1998)；(b) 若 $X_1 \sim N$（0, 1），则 Basu 定理断言 X 和 $X_{(n/2)} - X$ 相互独立，因为 X 是完全充分统计量，而 $X_{(n/2)} - X$ 是辅助统计量。也就是说，采用与第 3.3 节所示推导类似的方法，我们可以通过聚焦于统计量来改进 T_o

$$T = \left\{\sigma T_o / \mathrm{E}\left|X_1 - \nu\right| - A\right\}\left\{\sigma^2\left(\mathrm{E}\left|X_1 - \nu\right|\right)^{-2} - 1\right\}^{-0.5}$$

满足 var (T) $\to 1$ 且 cov (T, A) $\to 0$，如 $n \to \infty$。因此，我们提出的检验如下：若满足以下条件，则拒绝 H_0:

$$T_N = \left\{2n^{0.5} X_{(n/2)} \hat{f}\left(X_{(n/2)}\right) S/\hat{w} - n^{0.5}\bar{X}/S\right\}\left\{S^2\hat{w}^{-2} - 1\right\}^{-0.5} > z_\alpha,$$

其中，$S^2 = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2/(n-1)$、$\hat{w} = \sum_{i=1}^{n}\left|X_i - X_{(n/2)}\right|/n$ 和 $\hat{f}$ 表示密度函数 f 的核估计量。为了估计 $f\left(X_{(n/2)}\right)$，我们建议使用 R 命令（R 开发核心团队，2012 年）：

$$\texttt{density}(\texttt{X}, \texttt{from} = \texttt{median}(\texttt{X}), \texttt{to} = \texttt{median}(\texttt{X}))[[2]][[1]]$$

显然，对于双侧检验 H_0：$\nu = 0$ 对 H_1：$\nu \neq 0$，我们可以采用拒绝规则：$T_N^2 > \chi_1^2(\alpha)$，其中阈值 $\chi_1^2(\alpha)$ 满足 $\Pr\left\{Z > \chi_1^2(\alpha)\right\} = \alpha$，随机变量 Z 服从自由度为 1 的卡方分布。

值得指出的是，本节所示的检验算法可以轻松扩展，用于对底层数据分布的分位数作出判断；详情请参见第 6 节。

3.5 对称中心检验

在许多实际应用中，例如检验治疗前后效应的配对检验，我们可能会关注检验配对观测值的对称中心是否为零。

为此，我们假设独立同分布的观测值 $X_1, \ldots, X_n$ 来自未知的对称分布 F，且 $\sigma^2 = \text{var}(X_1) < \infty$。根据 Bickel 和 Lehmann（2012 年），其自然位置参数（记为 ν）是 F 的对称中心。我们关注检验假设 $H_0: \nu = 0$ 与 $H_1: \nu > 0$。

统计量 $T_o = n^{0.5}\bar{X}/S$ 和 $T_1 = 2n^{0.5}X_{(n/2)}\hat{f}\left(X_{(n/2)}\right)$ 适合用于检验 H_0，其中 $S^2 = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2/(n-1)$ 和 $\hat{f}(u)$ 分别估计 $f(u) = dF(u)/du$。T_o 和 T_1 的各个组成部分将在第 3.3 节和第 3.4 节中详细说明。显然，若已知 F 是正态分布函数，则 T_o 优于 T_1；而若已知 F 是某个分布，例如随机变量 $\xi_1 - \xi_2$ 的分布，其中 ξ_1，ξ_2 是独立同分布的 $Exp(1)$，则 T_1 优于 T_o。

例如，考虑将统计量 T_o 作为待修改的检验统计量，而统计量 $A = \left(\bar{X} - X_{(n/2)}\right)D^{-0.5}$ 则发挥近似辅助统计量的作用，其中 $D = \sigma^2 - \mathrm{E}|X_1 - \nu|/f(\nu) + 1/(4f^2(\nu))$。然后，按照第 3.3 节和第 3.4 节中定义的概念和记号，我们提出在以下情况下拒绝 H_0:

$$
\begin{aligned}
T_N &= \left\{T_o + \delta n^{0.5}\left(\bar{X} - X_{(n/2)}\right)\hat{D}^{-0.5}\right\}V^{-0.5} > z_\alpha, \\
\hat{D} &= S^2 - \hat{w}/\hat{f}\left(X_{(n/2)}\right) + \left(4\hat{f}^2\left(X_{(n/2)}\right)\right)^{-1},\ \delta = \left\{\hat{w}/\left(2S\hat{f}\left(X_{(n/2)}\right)\right) - S\right\}\hat{D}^{-0.5}, \\
V &= 1 + \delta^2 + \frac{2\delta S}{\hat{D}} - \frac{\delta\hat{w}}{\hat{D}^{0.5}S\hat{f}\left(X_{(n/2)}\right)} = 1 + \frac{2\hat{w}}{\hat{D}\hat{f}\left(X_{(n/2)}\right)} - \frac{1}{\hat{D}}\left(\frac{\hat{w}}{2S\hat{f}\left(X_{(n/2)}\right)} + S\right)^2.
\end{aligned}
$$

我们将通过实验说明，基于 T_N 的检验能够结合基于 T_o 和 T_1 的检验所具有的优良功效特性。

注 3.1. 请注意，本节中的检验统计量 T_N 旨在改进统计量 T_o。在本节的框架下，不存在最强检验决策机制。因此，一般而言，可以假设我们能够找到在某些情形下优于基于 T_N 的检验的决策程序。

第 4 节对本节推导出的决策方案的性质进行了数值考察。

4 数值模拟

我们进行了蒙特卡罗研究，以探究第 3 节所述的、针对均值、中位数和对称中心的检验的拟议变换的表现。就非参数决策程序的评估而言，值得注意的是，在第 3.3、3.4 和 3.5 节的框架下，并不存在最强检验。因此，我们在各种 H_0/H_1-底层数据分布设计下，比较基于给定统计量 T_o 的检验与基于相应统计量 T_N（ T_o 的修正版）的检验。数值研究旨在确认，所提方法能够提升统计功效。在第 4.2 和 4.3 节中，为进行进一步比较，我们展示了常用于应用中的单样本 Wilcoxon-Mann-Whitney 检验的蒙特卡罗检验功效；研究者常借此检验考察假设 H_0 : $\nu = 0$ ，其中 ν 是观测值的中位数。请注意，在实践中，很难找到均值的单样本 t 检验的非参数替代方法。因此，在第4.1节评估 t 检验及其第 3.3 节定义的变换时，我们纳入了原始 t 统计量 T_o 的自助法（非参数重抽样）版本，以便与相应统计量 T_N 进行比较；我们预计，在若干非参数情形下，自助法 t 检验可能优于原始 t 检验（Efron，1992）。

为评估这些检验，我们针对不同分布生成了 55,000 个大小为 $n \in \{n_1, \ldots, n_J\}$ 的独立样本，分别对应设计 D_{km}、 $k \in \{0, 1\}$、 $m \in \{1, \ldots, M\}$。在此方案中，设计 D_{km}、 $m \in \{1, \ldots, M\}$ 分别符合假设 H_k、 $k \in \{0, 1\}$。所呈现的每项自助法模拟结果均基于 55,000 次重复试验，每次使用 1,000 个自助样本。

令记号 $T(D_{km})$ 表示针对设计 D_{km}, $k \in \{0, 1\}$, $m \in \{1, \ldots, M\}$ 进行的检验统计量 T 。为了评估所提出检验的实验特征，我们通过蒙特卡洛方法计算了以下量的估计值，记为 PowA 和 Pow：$\Pr_1\{T(D_{km}) > C_\alpha\}$ 和 $\Pr_1\{T(D_{km}) > q(D_{0m})\}$，其中 C_α 是与 T 的渐近 H_0 分布相关的 α 水平临界值，而 $q(D_{0m})$ 分别表示 $T(D_{0m})$ 的分布的 $100(1-\alpha)\%$ 分位数值。在 D_{0m}, $m \in \{1, \ldots, M\}$ 下计算的 PowA 准则考察了我们当前控制 TIE

率的能力，该检验基于 T，并采用 T 的近似 H_0 分布。在此框架下，在 D_{1m}，$m \in \{1,\ldots,M\}$ 条件下计算的 PowA 表示 T 的期望检验功效。假设我们能够准确控制相应的、基于 T 的检验的 TIE 率，则 Pow 的值可用于评估 T 的实际检验功效水平。理论上可以假设，我们能够校正 T，得到一个统计量（记为 T'），以便通过采用例如 Bartlett 型校正、位置调整和 / 或自助法技术，最小化距离 $|\Pr_0\{T'(D_{0m}) > C_\alpha\} - \alpha|$。在此框架下，可能需要对 $T(D_{0m})$ 的 H_0 分布作出精确的高阶近似。在若干情形下，Pow 可能表明，所研究检验的实际应用仍有改进空间。

4.1 均值的单样本t检验

为了考察在第 3.3 节中通过修正 t 检验生成的基于 T_N 的检验 T_o，采用了以下底层数据分布设计：D_{k1}：$X_1 \sim N(0.1k,1)$;D_{k2}：$X_i = 1 - \eta_i + 0.1k$，其中 $\eta_i \sim Exp(1)$;D_{k3}：$X_i = \eta_i - 1 + 0.1k$;D_{k4}：$X_i = (\xi_i - 2)/2 + 0.2k$，其中 $\xi_i \sim$ Weibull$(1,2)$; 其中 $k \in \{0,1\}$ 和 $i \in \{1,\ldots,n\}$。表1中的实验结果展示了基于 T_o 的 t 检验、基于 T_N 的修正版以及自助法检验 T_B（即 t 检验的自助法版本）的检验功效比较；假设这些检验的显著性水平 α 固定为 5%。

设计 D_{k1}、$k \in \{0,1\}$ 展示了 T_o 是最强检验的情景。在这些情况下，Pow 的值表明 T_o 略优于 T_N。设计 D_{k2}、$k \in \{0,1\}$ 对应负偏态分布。在这些情景中，T_N 明显略优于 T_o，与 T_o 相比，检验功效约提升 27%−30%。设计 D_{k3}、$k \in \{0,1\}$ 代表正偏态分布。所提出的检验 T_N（D_{13}）的检验功效约为 T_o（D_{13}）的两倍。不过，我们应注意，与 T_N（D_{03}）相关的渐近TIE率控制会受到 N-分布偏度的影响。根据在 D_{13} 下计算的 Pow 值，如果能改善与 T_N 相关的 TIE 率控制，程序 T_N（D_{k3}）、$k \in \{0,1\}$ 将明显优于策略 T_o（D_{k3}）、$k \in \{0,1\}$。为此，例如，可以考虑采用 Chen（1995）型方法。本文并不旨在改进检验算法

表1：以下统计量在 $\alpha = 0.05$下的蒙特卡洛拒绝率：t 检验统计量 T_o 及其在第 3.3 节中定义的修正版 T_N；t 检验统计量的自助法版本 T_B。

		$n = 150$		$n = 200$		$n = 250$		$n = 300$		$n = 350$	
设计检验		PowA	Pow	PowA	Pow	PowA	Pow	PowA	Pow	PowA	Pow
D_{01}	T_o	0.049	0.050	0.053	0.050	0.049	0.050	0.050	0.050	0.052	0.050
	T_N	0.050	0.050	0.054	0.050	0.050	0.050	0.051	0.050	0.053	0.050
	T_B	0.054	0.054	0.052	0.052	0.050	0.050	0.051	0.051	0.052	0.052
D_{11}	T_o	0.337	0.334	0.409	0.401	0.475	0.478	0.537	0.537	0.588	0.581
	T_N	0.337	0.333	0.412	0.399	0.477	0.476	0.539	0.533	0.590	0.579
	T_B	0.332	0.332	0.416	0.416	0.463	0.463	0.539	0.539	0.586	0.586
D_{02}	T_o	0.044	0.050	0.046	0.050	0.046	0.050	0.046	0.050	0.044	0.050
	T_N	0.044	0.050	0.043	0.050	0.044	0.050	0.043	0.050	0.043	0.050
	T_B	0.083	0.083	0.074	0.074	0.068	0.068	0.068	0.068	0.069	0.069
D_{12}	T_o	0.345	0.363	0.417	0.429	0.487	0.499	0.541	0.554	0.597	0.614
	T_N	0.479	0.503	0.591	0.615	0.679	0.704	0.748	0.774	0.805	0.826
	T_B	0.386	0.386	0.447	0.447	0.507	0.507	0.559	0.559	0.608	0.608
D_{03}	T_o	0.055	0.050	0.055	0.050	0.054	0.050	0.053	0.050	0.052	0.050
	T_N	0.068	0.050	0.068	0.050	0.067	0.050	0.067	0.050	0.065	0.050
	T_B	0.032	0.032	0.036	0.036	0.037	0.037	0.039	0.039	0.038	0.038
D_{13}	T_o	0.331	0.316	0.398	0.379	0.464	0.449	0.528	0.517	0.584	0.579
	T_N	0.644	0.580	0.732	0.663	0.801	0.749	0.854	0.810	0.890	0.856
	T_B	0.307	0.307	0.383	0.383	0.454	0.454	0.521	0.521	0.586	0.586
D_{04}	T_o	0.044	0.050	0.046	0.050	0.045	0.050	0.044	0.050	0.046	0.050
	T_N	0.042	0.050	0.043	0.050	0.044	0.050	0.044	0.050	0.044	0.050
	T_B	0.077	0.077	0.079	0.079	0.071	0.071	0.069	0.069	0.068	0.068
D_{14}	T_o	0.790	0.804	0.879	0.885	0.929	0.934	0.961	0.965	0.979	0.980
	T_N	0.940	0.950	0.981	0.984	0.994	0.995	0.998	0.999	0.999	1.000
	T_B	0.781	0.781	0.863	0.863	0.920	0.920	0.954	0.954	0.973	0.973

用于控制 T_N 的 TIE率。表1中所示的 Pow 准则计算值证实，基于 T_N 的策略是合理的。与 D_{04} 和 D_{14} 相关的结果支持上述结论。尽管在 D_{04} 下，相应的 PowA 值表明， T_N 的蒙特卡洛渐近 TIE 率低于与 T_o 相关的 TIE 率，但在 D_{14} 下，无论是 PowA 还是 Pow 情境，所提出的检验都优于 T_o 。

表1还报告了与基于 T_o 统计量自助法版本的检验之蒙特卡洛实现相关的实验结果，该版本记为 T_B，其中 X_1，…，X_n 均采用有放回重采样。在这些情形下，未对相应的 TIE率进行渐近近似。因此，我们将判据记为 PowA=Pow。所采用的自助法策略需要付出相当大的计算成本。然而，在 D_{01}， D_{11}， D_{02}， …， D_{14} 下，我们无法确认基于 T_B 的检验显著优于基于 T_o 的 t 检验。此外，在设计 D_{02} 和 D_{04} 下，不建议使用自助法 t 检验。

4.2 单样本中位数检验

为了深入了解第3.4节中定义的检验统计量 T_N 的运行特性，我们考察了与假设 H_0 ： $\nu = 0$ 和 H_1 ： $\nu > 0$ 相对应的底层数据分布的各种设计，其中 ν 表示 X_1 分布的中位数。为举例说明所开展的蒙特卡洛研究结果，我们采用以下方案： D_{01} ： $X_i = \eta_i - \xi_i$； D_{11}： $X_i = 1 - \eta_i$； D_{02}： $X_i \sim N(0, 4)$； D_{12}： $X_i = \exp(0.5) - \zeta_i$；其中 $\eta_i \sim Exp(1)$、 $\xi_i \sim Exp(1)$、 $\zeta_i \sim LN(0, 1)$ 和 $i \in \{1, \ldots, n\}$。在本研究中，根据第 3.4 节所述，我们实施了单样本单侧 Wilcoxon-Mann-Whitney 检验（记为 W）、基于 T_o 的检验及其修正版，以及基于 T_N 的检验。请注意，对于 W 检验，PowA=Pow 判据成立，因为 Wilcoxon-Mann-Whitney 检验统计量的 H_0 分布不依赖于底层数据分布。表2列出了在对 W、 T_o 和 T_N 进行广泛功效评估时观察到的典型结果；假设所考虑检验的显著性水平 α 固定为 5%。例如，在情景 $\{D_{11}, n = 50\}$ 中， T_N 提升了 T_0 的检验功效约 25%。

关于 第3.4节中推导的基于 T_N^2 的双侧检验，以下结果

表2：以下统计量在 $\alpha = 0.05$ 下的蒙特卡洛拒绝率：单样本 Wilcoxon-Mann-Whitney 检验统计量 (W)， T_o 及其修正版 T_N，定义见第3.4节。

		$n = 25$		$n = 50$		$n = 75$	
设计检验		PowA	Pow	PowA	Pow	PowA	Pow
D_{01}	W	0.048	0.048	0.049	0.049	0.049	0.049
	T_o	0.029	0.050	0.025	0.050	0.031	0.050
	T_N	0.030	0.050	0.027	0.050	0.028	0.050
D_{11}	W	0.213	0.213	0.319	0.319	0.417	0.417
	T_o	0.484	0.549	0.674	0.751	0.806	0.862
	T_N	0.660	0.725	0.899	0.936	0.973	0.985
D_{02}	W	0.048	0.048	0.050	0.050	0.050	0.050
	T_o	0.049	0.050	0.044	0.050	0.045	0.050
	T_N	0.048	0.050	0.042	0.050	0.045	0.050
D_{12}	W	0.394	0.394	0.609	0.609	0.755	0.755
	T_o	0.734	0.735	0.919	0.927	0.977	0.980
	T_N	0.843	0.849	0.982	0.986	0.998	0.999

分别举例说明双侧 W 检验、基于 T_o^2 的检验和基于 T_N^2 的检验所提供的相应蒙特卡洛功效评估：PowA $=0.223$、0.585 和 0.838，其中当$n = 50$ 且生成的数据满足 D_{11}时。请注意，在上述情景中，我们可以采用 Fisher 和 Robbins（2019）提出的方法。根据 Fisher 和 Robbins（2019），由于 $T_o^2 > 0$ 和 $T_N^2 > 0$ 是 $O_p\left(n^k\right)$，其中 $k = 0$ 和 $k = 1$ 在 H_0 和 H_1 下分别成立，因此值得考察检验统计量 $T_{o1}^2 = -n\log\left(1 - T_o^2/n\right)$ 和 $T_{N1}^2 = -n\log\left(1 - T_N^2/n\right)$。这些单调变换使基于 T_o^2 和 T_N^2 的策略的 PowA 分别略微提高了约 1.4% 和 1.1%。

4.3 对称中心检验

本节中，我们考察第3.5节提出的基于 T_o 检验的 T_N-修正版的实现。我们将基于 T_o 和 T_1 的检验，以及单样本单侧 Wilcoxon-Mann-Whitney 检验（W），与基于 T_N 的检验进行比较，比较所依据的设定如第3.5节所示。为举例说明数值研究的结果，采用了以下数据 $D = \{X_1, \ldots, X_n\}$ 生成设计：对于 $k \in \{0, 1\}$ 和 $i \in \{1, \ldots, n\}$，D_{k1}：$X_i \sim N\left(0.1k,\ 1\right)$，此时可预期 T_o 优于 T_1、T_N 和 W；D_{k2}：$X_i = \eta_i - \xi_i + 0.1k$，其中 η_i 和 ξ_i 是独立的Exp(1)分布随机变量，因此可预期 T_1 优于 T_o、T_N 和 W；D_{k3}：$X_i = \zeta_i + 0.1k$，其中 $\zeta_i \sim Unif\left(-1,\ 1\right)$；$D_{k4}$：$X_i = \epsilon_i - 0.5 + 0.1k$，其中$\epsilon_i \sim Beta$(0.5，0.5)。

表3总结了情景 D_{kj}, $k \in \{0, 1\}$,$j \in \{1, \ldots, 4\}$ 下计算得到的蒙特卡洛结果，其中 $n = 50,\ 150$，且假定检验的显著性水平 α 固定为 5%。观察结果如下：在 D_{01} 和 D_{11} 下，T_o 和 T_N 的表现非常相似；在 D_{02} 且 $n = 50$ 的情况下，T_N 在 TIE 率控制方面确实改善了 T_o；在 D_{12} 下，与 T_N 和 T_1 相关的测量指标 Pow 数值彼此接近，且高于 T_o 的数值；在 D_{kj} 下，$k \in \{0, 1\}$、$j \in \{3, 4\}$、T_N 展现出的蒙特卡洛检验功效特征优于 T_o 和 W。例如，在 D_{14} 且 $n = 50$ 的情况下，与 W、T_o 和 T_1 相比，T_N 的检验功效分别约提升 22%、23% 和 65%。

基于 所进行的蒙特卡洛研究，我们得出结论：所提出的检验

表3：单样本 Wilcoxon-Mann-Whitney 检验（W）以及第3.5节中定义的基于 T_o、 T_1 和 T_N 的检验在 $\alpha = 0.05$ 处的蒙特卡洛检验功效水平。

		$n = 50$		$n = 150$				$n = 50$		$n = 150$	
设计检验	PowA		Pow	PowA	Pow	设计检验	PowA		Pow	PowA	Pow
D_{01}	W	0.049	0.049	0.052	0.052	D_{03}	W	0.050	0.050	0.049	0.049
	T_o	0.052	0.050	0.053	0.050		T_o	0.054	0.050	0.050	0.050
	T_1	0.044	0.050	0.047	0.050		T_1	0.062	0.050	0.057	0.050
	T_N	0.054	0.050	0.054	0.050		T_N	0.052	0.050	0.048	0.050
D_{11}	W	0.169	0.169	0.321	0.321	D_{13}	W	0.301	0.301	0.645	0.645
	T_o	0.183	0.178	0.338	0.329		T_o	0.328	0.317	0.678	0.676
	T_1	0.128	0.140	0.231	0.238		T_1	0.191	0.160	0.345	0.319
	T_N	0.189	0.173	0.341	0.328		T_N	0.390	0.381	0.774	0.782
D_{02}	W	0.051	0.051	0.049	0.049	D_{04}	W	0.050	0.050	0.050	0.050
	T_o	0.058	0.050	0.049	0.050		T_o	0.054	0.050	0.053	0.050
	T_1	0.026	0.050	0.024	0.050		T_1	0.089	0.050	0.067	0.050
	T_N	0.044	0.050	0.035	0.050		T_N	0.046	0.050	0.048	0.050
D_{12}	W	0.144	0.144	0.280	0.280	D_{14}	W	0.643	0.643	0.965	0.965
	T_o	0.132	0.119	0.225	0.225		T_o	0.637	0.617	0.965	0.962
	T_1	0.089	0.149	0.206	0.315		T_1	0.292	0.179	0.493	0.431
	T_N	0.134	0.148	0.243	0.304		T_N	0.826	0.836	0.998	0.998

策略表现出　　在各种备择假设设计下，检验功效高且稳定。

5 真实数据示例

心脏血流受阻时，血栓通常会导致心肌梗死（MI），进而造成心肌损伤。根据美国疾病控制与预防中心的数据，心脏病是主要死因之一，影响约 20% 或更多的人口，且不受族裔差异影响，例如 Schisterman 等（2001）。

通过采用一项评估与心肌梗死相关生物标志物的研究样本，展示了所提出方法的应用。该研究以伊利县和尼亚加拉县年满 $35-79$ 岁的居民为对象。研究使用纽约州机动车辆管理局的驾驶执照名册作为 35 至 65 岁成年人的抽样框，而老年人样本（年龄 $65-79$）则从医疗保健融资管理局数据库中随机抽取。所谓“硫代巴比妥酸反应物”（TBARS）和“高密度脂蛋白”（HDL）胆固醇的生物标志物，常被用作区分患有（MI=1）和未患有（MI=0）心肌梗死疾病个体的判别因素，例如 Schisterman 等（2001）。

利用这 2,910 个生物标志物值构成的样本，估计与 MI$=1\}$病例相关的线性回归模型 $Y_i = a + bZ_i + \epsilon_i$中的参数 a 和 b，其中 $Y_1, \ldots, Y_{2910}$为经对数转换的高密度脂蛋白胆固醇测量值，$Z_1, \ldots, Z_{2910}$ 表示经对数转换的 TBARS 测量值，$\epsilon_i,\, i \geq 1,$ 表示回归残差，且 E$\epsilon_i = 0$。研究得出结论：$\epsilon_i \simeq Y_i - 4.034 - 0.045 Z_i,\, i \geq 1$（详见在线补充材料中的表 S1）。假设我们旨在根据 $n = 100$个生物标志物值，考察 ϵ_i 的分布，其中 MI $= 1$。在这种情况下，观测到样本均值和方差分别为 $\bar{\epsilon} = \sum_{i=1}^{n} \epsilon_i / n \simeq -0.002$ 和 $\sum_{i=1}^{n} (\epsilon_i - \bar{\epsilon})^2 / (n-1) \simeq 0.073$。图 2 展示了基于 $\epsilon_1, \ldots, \epsilon_{100}$对应值绘制的直方图。

为了检验 $H_0: \nu = 0$与 $H_1: \nu \neq 0$，其中 ν是 ϵ 的分布中位数，我们实施了基于 T_o^2 的双侧检验及其修正版，即第3.4节所述的基于 T_N^2 的检验，以及双侧 Wilcoxon-Mann-Whitney 检验（W）。尽管图 2 所示的直方图显示该分布相对于零点具有一定的不对称性，但基于 T_o^2 的检验和 W 检验的 p值分别为$= 0.071$ 和 p值$= 0.326$,

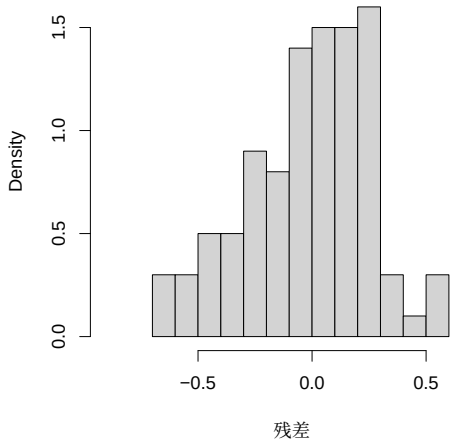


图 2：与回归残差相关的基于数据的直方图 ϵ_1、.....、 ϵ_n。

所提出的基于 T_N^2 的检验得到的 p值为$= 0.047$。随后，我们开展了自助法/刀切法类型的研究，以考察各检验统计量的检验功效表现。具体策略是从数据 $\{\epsilon_1, \ldots, \epsilon_n\}$ 中有放回地随机抽取一个样本量为 $n_b < 100$ 的样本，并在 5% 的显著性水平下检验 H_0。该策略重复 $10,000$ 次，以计算事件 $\{T_o^2$拒绝 $H_0\}$、 {W 拒绝 $H_0\}$，以及 $\{T_N^2$ 拒绝 $H_0\}$ 的频率。所得 T_o^2、W 和 T_N^2 的实验功效分别为：当 $n_b = 90$ 时为 0.238、0.106、0.535；当 $n_b = 80$ 时为 0.198、0.104、0.463；当 $n_b = 70$ 时为 0.176、0.100、0.415。随着样本量 n_b 增加，各检验的实验检验功效水平也随之提高。这项研究从实验角度表明，在评估关联关系 $Y_i = a + bZ_i + \epsilon_i$, $i \geq 1$ 的残差是否相对于零点呈不对称分布时，基于 T_N^2 的检验在功效特性方面优于经典方法。也就是说，在本研究中，与已知方法相比，所提出的检验预计对拒绝原假设 H_0： $\nu = 0$ 与 H_1 :$\nu \neq 0$ 更为敏感。

6 总结

本文提出了一个用于评估和构建高检验功效的基于数据的检验的理论框架。本文的贡献涉及刻画最强统计决策机制的原理。命题 2.2 通过分析相应似然比的行为，提供了一种方法，可将“最强”这一术语与检验统计量分布函数的性质一一对应。我们证明，所得的最强检验刻画可以与充分性原理联系起来。第 2 节介绍的概念已用于改进检验程序，方法是考虑相关的辅助统计量。在非参数情形下，本文提出的理论框架应用于展示单样本 t 检验、中位数检验和对称中心检验的有效修正版。通过模拟和一个真实数据示例，我们在基于相对偏斜分布样本的各种情景中，验证了所提出的非参数决策程序在保持较高检验功效方面的有效性。我们还提出以下几点说明。(a) 命题 2.4 和 3.1 可应用于不同的决策问题。(b) 对经典 t 检验进行有效修正，具有理论和应用价值。根据第 3.3 节对 t 检验所作的修正版涉及使用三阶中心矩的估计。当底层数据不对称时，该矩在调整零假设分布的某些 t 检验结构修正中发挥作用，例如 Chen（1995）。总体而言，本文提出的修正版与用于改善 t 检验类程序 TIE率控制的方法有所不同。因此，通常可以将上述方法的基本要素结合起来。(c) 第3.4节提出的方案可以轻松修改，用于构建¯分位数检验；这是基于如下观察： X 和样本的 p分位数（记作 $X_{(pn)}$）渐近服从二元正态分布，且

$$\begin{aligned}
&\operatorname{cov}\left(\bar{X}, X_{(pn)}\right) \;=\; f\left(\nu_p\right)^{-1} \mathrm{E}\left(X_1-\nu_p\right)\left\{pI\left(X_1>\nu_p\right)-(1-p)I\left(X_1 \leq \nu_p\right)\right\}, \\
&\operatorname{var}\left(X_{(pn)}\right) \;=\; f\left(\nu_p\right)^{-2} p(1-p), \;\; \Pr\left(X_1<\nu_p\right) \;=\; p.
\end{aligned}$$

(d) 第 3–5 节举例说明了经过处理的 MP 原则在非参数情形下的应用。在许多参数问题中，相应的似然比

没有显式形式，或形式非常复杂，例如，检验所依据的数据涉及纵向数据、相关观测、多变量结局、受不同类型误差影响的数据和/或缺失值机制。在此类情况下，可运用所考虑的 MP 原则来比较或构建检验。

决策算法种类繁多，涉及统计实践的大多数字段。因此，在一篇论文中探讨所有相关理论和示例并不现实。本文仅研究一种刻画一类最强检验机制的方法。也就是说，未来有许多颇具前景的研究方向，例如：(i) 考察一般最强检验刻画与 Basu 定理类结果（如 Ghosh，2002年）、充分性、完备性和辅助性等概念之间的关系，并探讨这些概念在不同决策策略表述下的情形。在这方面，例如可以提出这样一个研究问题：对于任意辅助统计量 A，何时可以声称统计量 T 是最强检验，当且仅当 T 和 A 相互独立？(ii) 可以在不同情境下，针对最强检验刻画开展各种参数应用和非参数应用。(iii) 命题 2.4 可用于比较并拓展对实际中不同统计程序的比较。(iv) 鉴于本文提出的最强检验原则，可以提出对检验统计量最优组合的相关评估（如 Berk 和 Jones，1978年）。(v) 可以分析依据最强检验刻画进行修正后的检验统计量的大样本性质。(vi) 或许可以将命题 3.1 纳入各种检验方法的发展之中，利用相应假设下对底层数据分布的刻画来定义相关辅助统计量。本文暂且将这些课题留待未来研究，并希望本文能使读者认识到，研究与基于 MP 数据的决策技术相关的不同方面及其刻画具有诸多益处。

补充材料

在线补充材料包含：文中理论结果的证明；以及表 S1，该表展示了第 5 节中以对数转换后的 TBARS 测量值为因子、对观测到的对数转换后的高密度脂蛋白胆固醇测量值进行线性回归拟合所涉及的方差分析。

参考文献

Bahadur, R. R.（1955 年）。充分性的刻画。《数学统计年刊》*26*，286–293。Bayarri, M. J. 和 J. O. Berger（2000年）。复合零模型的 P值。《美国统计学会会刊》*95*，1127–1142。Berk, R. H. 和 D. H. Jones（1978年）。检验统计量的相对最优组合。《斯堪的纳维亚统计学杂志》*5*，158–162。Bickel, P. J. 和 E. L. Lehmann（2012 年）。非参数模型的描述性统计量 I：引言。载于《*EL Lehmann* 选集》，第 465–471 页。施普林格。Boos, D. D. 和 J. M. Hughes-Oliver（1998年）。Basu 定理的应用。《美国统计学家》*52*(3)，218–221。Chen, L.（1995 年）。偏态分布均值的检验。《美国统计学会会刊》*90*，767–772。Efron, B.（1992）。自助法：重新审视刀切法。施普林格，纽约。Ferguson, T. S.（1998年）。样本均值与样本分位数的渐近联合分布。未发表。可访问：*http://www.math.ucla.edu/ tom/papers/unpublished/meanmed.pdf*。Fisher, T. J. 和 M. W. Robbins（2019年）。提高保守假设检验检验功效的一个简便技巧。《美国统计学家》*73*，232–242。Ghosh, I.、N. Balakrishnan 和 H. K. T. Ng（2021年）。《统计学进展：理论与应用：致敬 *Barry C. Arnold* 对统计科学的贡献》。施普林格，纽约。Ghosh, M.（2002年）。Basu 定理及其应用：个人主义视角的综述。*Sankhyā*：《印度统计学杂志》*64*，509–531。Ghosh, M.、N. Reid 和 D. A. S. Fraser（2010年）。辅助统计量：综述。《中华统计学志》*20*，1309–1332。

Hall, P. 和 B. La Scala（1990 年）。经验似然的方法与算法。《国际统计评论》/《国际统计评论》，109–127。Johnson, V. E.（2013）。一致最强贝叶斯检验。《统计年刊》*41*，1716–1741。Kagan, A. 和 L. A. Shepp（2005）。充分性悖论：一个保留 Fisher 信息的不充分统计量。《美国统计学家》*59*，54–56。Lehmann, E. L.（1950）。假设检验理论的若干原理。《数学统计年刊》*21*，1–26。Lehmann, E. L.（1993）。Fisher 与 Neyman-Pearson 的假设检验理论：一种理论还是两种? 《美国统计学会会刊》*88*，1242–1249。Lehmann, E. L. 和 J. P. Romano（2005）。《统计假设检验》。施普林格，纽约。Maity, A. 和 M. Sherman（2006）。一个方差未知时的双样本 t 检验。《美国统计学家》*60*，163–166。O'Neill, B.（2014）。抽样问题中的一些有用矩结果。《美国统计学家》*68*，282–296。R 开发核心团队（2012 年）。*R*：用于统计计算的语言与环境。R 统计计算基金会，维也纳，奥地利。ISBN 3-900051-07-0 http://www.R-project.org。Schisterman, E. F., D. Faraggi, R. Browne, J. Freudenheim, J. Dorn, P. Muti, A. D., B. Reiser, 和 T. M.（2001）。基于人群的样本中硫代巴比妥酸反应物与心血管疾病。《心血管风险杂志》*8*，1–7。Semkow, T. M., N. Freeman, U. F. Syed, D. K. Haines, A. Bari, J. Khan, A., K. Nishikawa, A. Khan, A. J. Burn, X. Li, 和 L. T. Chu（2019年）。卡方分布：新的推导与环境应用。《应用数学与物理杂志》*7*，1786–1799。

Vexler, A.（2021年）。有效的 p值及其期望值再探。统计数学学会年刊 *73*，227–248。

Vexler, A. 和 A. Hutson（2018年）。健康科学统计学：理论、应用与计算。CRC出版社，纽约。

Vexler, A.、C. Wu 和 K. Yu（2010年）。最优假设检验：从半贝叶斯因子到完全贝叶斯因子。*Metrika 71*，25–138。

补充材料”具有简单非参数应用的最（更）强检验统计量的刻画”

附录：证明

本附录包含确立本文所述命题所需的证明。

命题1.1的证明

考虑

$$\Pr_1\left(u-s\leq\Lambda\leq u\right)=\mathrm{E}_1 I\left(u-s\leq\Lambda\leq u\right)\;=\;\int I\left(u-s\leq\Lambda\leq u\right)f_1$$
$$=\;\int I\left(u-s\leq\Lambda\leq u\right)\frac{f_1}{f_0}f_0\;\;=\;\int I\left(u-s\leq\Lambda\leq u\right)\;\Lambda\;f_0,$$

其中 $I(.)$ 表示示性函数，而 u,s 不是随机变量。这意味着以下不等式成立

$$\begin{aligned}\Pr_1\left(u-s\leq\Lambda\leq u\right)\;&\leq\;\int I\left(u-s\leq\Lambda\leq u\right)uf_0\;\;=\;u\Pr_0\left(u-s\leq\Lambda\leq u\right),\\ \Pr_1\left(u-s\leq\Lambda\leq u\right)\;&\geq\;\int I\left(u-s\leq\Lambda\leq u\right)(u-s)f_0\\ &\qquad=\;(u-s)\Pr_0\left(u-s\leq\Lambda\leq u\right).\end{aligned}$$

将这些不等式除以 s 并利用 $s\to 0$，即可完成证明。

命题 2.1 的证明

我们可以写成

$$f_1^{\Lambda,A}(u,v) \;=\; \frac{d}{dv}\lim_{s\to 0}\frac{1}{s}\Pr_1\left(u-s<\Lambda<u, A<v\right),$$

其中

$$\begin{aligned}
&\Pr_1\left\{u-s<\Lambda(D)<u, A(D)<v\right\} = \mathrm{E}_1 I\left\{u-s<\Lambda(D)<u, A(D)<v\right\}\\
&\quad=\int I\left\{u-s<\Lambda(D)<u, A(D)<v\right\}\, f_1(D)\\
&\quad=\int I\left\{u-s<\Lambda(D)<u, A(D)<v\right\}\, \Lambda(D)\, f_0(D).
\end{aligned}$$

注意到

$$\begin{aligned}
\int I\left(u-s<\Lambda<u, A<v\right)\, \Lambda\, f_0(D) \;&\le\; u\int I\left(u-s<\Lambda<u, A<v\right)\, f_0(D),\\
\int I\left(u-s<\Lambda<u, A<v\right)\, \Lambda\, f_0(D) \;&\ge\; (u-s)\int I\left(u-s<\Lambda<u, A<v\right)\, f_0(D),
\end{aligned}$$

我们有

$$\begin{aligned}
\Pr_1\left(u-s<\Lambda<u, A<v\right) \;&\le\; u\Pr_0\left(u-s<\Lambda<u, A<v\right),\\
\Pr_1\left(u-s<\Lambda<u, A<v\right) \;&\ge\; (u-s)\Pr_0\left(u-s<\Lambda<u, A<v\right).
\end{aligned}$$

仿照命题 1.1 的证明，即可完成证明。

命题 2.2 的证明

命题 2.2 的证明包含以下两个步骤：(1) 当 $T=\Lambda$，即 T 是最强检验统计量时，命题 2.1 表明 $f^{T,\,1}{}^{A}(u,\ v) = u\, f^{T,\,0}{}^{A}(u,\ v)$。(2) 当满足 $f^{T,\,1}{}^{A}(u,\ v) =_u f^{T,\,0}{}^{A}(u,\ v)$ 时，我们考虑以下初等不等式：对于所有 B 和 C，

$$(B-C)\left\{I(B\ge C)-\delta\right\}\ge 0,$$

其中，$I(\cdot)$ 是示性函数，且 $\delta\in[0,1]$。令 $\delta = I(\Lambda\ge U)$ 表示具有固定阈值 U 的似然比检验的 H_0-拒绝规则，使得：若 $\delta=1$，则拒绝H_0。令 $B=T$，可得

$$\mathrm{E}_0\left\{(T-C)\, I\left(T\ge C\right)\right\}\ge \mathrm{E}_0\left\{(T-C)\, I\left(\Lambda\ge U\right)\right\}.$$

因此，固定阈值 C， U，使 TIE率 $\mathrm{Pr}_0(T \geq C) = \mathrm{Pr}_0(\Lambda \geq U)$，可得

$$\mathrm{E}_0 I(T \geq C)\, T \geq \mathrm{E}_0 I(\Lambda \geq U)\, T,$$

其中

$$\mathrm{E}_0 I(\Lambda \geq U)\, T = \mathrm{E}_0 I(\Lambda \geq U)\, f_1^{T,\Lambda}(T,\Lambda)/f_0^{T,\Lambda}(T,\Lambda),$$

因为 $u = f^{T,A}_{1}(u,v)/f^{T,A}_{0}(u,v)$，对于任意统计量 A和所有 $u \geq 0$均成立。也就是说，

$$\begin{aligned}
\mathrm{E}_0 I(\Lambda \geq U)\, T &= \iint I(v \geq U)\, u\, f_0^{T,\Lambda}(u,v)dudv \\
&= \iint I(v \geq U)\, \frac{f_1^{T,\Lambda}(u,v)}{f_0^{T,\Lambda}(u,v)} f_0^{T,\Lambda}(u,v)dudv = \iint I(v \geq U)\, f_1^{T,\Lambda}(u,v)dudv \\
&= \int I(v \geq U) \int f_1^{T,\Lambda}(u,v)dudv = \int I(v \geq U)\, f_1^{\Lambda}(v)dv \;=\; \mathrm{Pr}_1(\Lambda \geq U).
\end{aligned}$$

因此，

$$\mathrm{E}_0 I(T \geq C)\, T \geq \mathrm{Pr}_1(\Lambda \geq U),$$

其中

$$\begin{aligned}
\mathrm{E}_0 I(T \geq C)\, T &= \int I(u \geq C)\, u f_0^T(u)du = \int I(u \geq C)\, \frac{f_1^T(u)}{f_0^T(u)} f_0^T(u)du \\
&= \mathrm{Pr}_1(T \geq C),
\end{aligned}$$

因为 $f^{T,A}_{1}(u,v) = u\, f^{T,A}_{0}(u,v)$ 蕴含 $\int f^{T,A}_{1}(u,v)dv = u \int f^{T,A}_{0}(u,v)dv$，即 $f_1^T(u)/f_0^T(u) = u$。这就给出了

$$\mathrm{Pr}_1(T \geq C) \geq \mathrm{Pr}_1(\Lambda \geq U),$$

这意味着，若 $f^{T,A}_{1}(u,v) = u f^{T,A}_{0}(u,v)$，则 T是最强检验，因为 Λ是最强检验。证明完毕。

命题 2.3 的证明

显然， $T = \Lambda = f_1(D)/f_0(D)$，即 T是最强检验，蕴含 $\mathrm{E}_1\{g(D)\} = \mathrm{E}_0\{g(D)T(D)\}$。

现在，我们假设 T满足 $\mathrm{E}_1\{g(D)\} = \mathrm{E}_0\{g(D)T(D)\}$，且事件 $T \geq C$表示应拒绝 H_0。考虑以下初等不等式：对于所有 B 和 C，

$$(B - C)\{I(B \geq C) - \delta\} \geq 0, \;\; \delta \in [0,1].$$

设 $\delta = I(\Lambda \geq U)$ 表示似然比检验在固定阈值 U下的 H_0-拒绝规则，使得：若 $\delta = 1$且 $\mathrm{Pr}_0(T \geq C) = \mathrm{Pr}_0(\Lambda \geq U) = \alpha$，则拒绝 H_0。在上述初等不等式中，我们定义 $B = T$，从而得到

$$0 \leq \mathrm{E}_0(T - C)\{I(T \geq C) - I(\Lambda \geq U)\} = \mathrm{E}_0\{I(T \geq C)T\} - \mathrm{E}_0\{I(\Lambda \geq U)T\}.$$

由此可得

$$0 \leq \mathrm{E}_1 I(T \geq C) - \mathrm{E}_1 I(\Lambda \geq U),$$

由于对于每个 $g \in [0,1]$，$\mathrm{E}_0\{g(D)T(D)\} = \mathrm{E}_1\{g(D)\}$。但是，$\mathrm{Pr}_1(\Lambda \geq U) \geq \mathrm{Pr}_1(T \geq C)$，因为 Λ 是最强检验。因此，$\mathrm{Pr}_1(\Lambda \geq U) = \mathrm{Pr}_1(T \geq C)$，这意味着 T 是最强检验。证明完毕。

命题 2.4 的证明

为证明命题 2.4，我们注意到，如上所述（例如，参见命题 2.2 的证明），$f^{T_1,\ T_2}_{1,\ 1}(u,\ v) = uf^{T_1,\ T_2}_{1,\ 0}(u,\ v)$ 蕴含 $f_1^{T_1}(u) = uf_0^{T_1}(u)$，进而得到对于固定阈值 C，$\mathrm{E}_0 I(T_1 \geq C)\ T_1 = \mathrm{Pr}_1(T_1 \geq C)$。由于 $f^{T_1,\ T_2}_{1,\ 1}(u,\ v) = uf^{T_1,\ T_2}_{1,\ 0}(u,\ v)$，我们还可得

$$\mathrm{E}_0 I(T_2 \geq U)\, T_1 = \iint I(v \geq U)\, u\, f_0^{T_1,T_2}(u,v)dudv = \mathrm{Pr}_1(T_2 \geq U),$$

其中，固定的 U 满足 $\mathrm{Pr}_0(T_1 \geq C) = \mathrm{Pr}_0(T_2 \geq U) = \alpha$，$\alpha$ 表示检验的 TIE率。然后，按照与命题 2.2 的证明类似的方式，我们应用不等式 $\mathrm{E}_0 I(T_1 \geq C)\ T_1 \geq \mathrm{E}_0 I(T_2 \geq U)\ T_1$ 完成证明。

命题 2.5 的证明

根据文章中命题 2.5 之前给出的陈述，证明很直接。

命题 3.1 的证明

考虑二元特征函数

$$\begin{aligned}
\mathrm{E}_1 e^{\mathrm{i}t_1 T_N + \mathrm{i}t_2 T} &= \mathrm{E}_1 e^{\mathrm{i}t_1 T_N + \mathrm{i}t_2 \psi(T_N, A)} = \int\int e^{\mathrm{i}t_1 u + \mathrm{i}t_2 \psi(u,x)} f_1^{T_N, A}(u,x) du dx \\
&= \int\int e^{\mathrm{i}t_1 u + \mathrm{i}t_2 \psi(u,x)} f_1^{T_N}(u) f_1^{A}(x) du dx = \int\int e^{\mathrm{i}t_1 u + \mathrm{i}t_2 \psi(u,x)} f_1^{T_N}(u) f_0^{A}(x) du dx \\
&= \int\int L^{T_N}(u)\, e^{\mathrm{i}t_1 u + \mathrm{i}t_2 \psi(u,x)} f_0^{T_N}(u) f_0^{A}(x) du dx = \mathrm{E}_0 \left\{ L^{T_N}(T_N)\, e^{\mathrm{i}t_1 T_N + \mathrm{i}t_2 \psi(T_N, A)} \right\} \\
&= \mathrm{E}_0 \left\{ L^{T_N}(T_N)\, e^{\mathrm{i}t_1 T_N + \mathrm{i}t_2 T} \right\},
\end{aligned}$$

其中 $\mathrm{i}^2 = -1$ and $t_j \in \mathbb{R}^1$, $j \in \{1, 2\}$. 在该方程中，$f^{T_N,1\ T}(u, v)$ 和 $L^{T_N}(u)$ 的傅里叶变换 $f^{T_N,0\ T}(u, v)$ 等价，即 $f^{T_N,1\ T}(u, v) = L^{T_N}(u)\ f^{T_N,0\ T}(u, v)$。如文中命题 3.1 上文所述，由于比值 $L^{T_N}(u)$ 是单调递增函数，我们不失一般性地假设 $L^{T_N}(T_N) = T_N$。于是，我们得到

$$\mathrm{E}_1 e^{\mathrm{i}t_1 T_N + \mathrm{i}t_2 T} = \mathrm{E}_0 \left\{ T_N\, e^{\mathrm{i}t_1 T_N + \mathrm{i}t_2 T} \right\},$$

从而得到 $f^{T_N,1\ T}(u, v) = u\, f^{T_N,0\ T}(u, v)$。现在，借助命题 2.4，当 $L^Y(u)$ 是单调递增函数时，命题 3.1 的证明即告完成。类似地，我们可以考虑 $L^Y(u)$ 是单调递减函数的情形。证明完毕。

表 S1

表4：第 5 节中，使用经对数转换的 TBARS 测量值 $Z_i, i = 1, \ldots, 2910$ 作为因子，对观测到的经对数转换的高密度脂蛋白胆固醇测量值 $Y_i, i = 1, \ldots, 2910$ 拟合线性回归所得的方差分析。

残差：

Min	1Q	中位数	3Q	Max
-0.80510	-0.15987	0.03547	0.20073	0.60918

系数：

	估计值	标准误	t值	Pr(> |t|)
(截距)	4.0342	0.03131	207.371	< 2e-16
Z	-0.0448	0.00568	-7.786	9.55e-15

残差标准误：0.2693，自由度为2908

复相关决定系数：0.02042，调整后的决定系数：0.02008

F统计量：60.62，自由度为 1 和 2908，p值：9.552e-15