渐近极小极大鲁棒似然比检验

G¨okhan G ¨ul

预防心脏病学与预防医学，心脏病学系，美因茨约翰内斯·古腾堡大学医学中心；临床流行病学与系统医学，血栓与止血中心，美因茨约翰内斯·古腾堡大学医学中心；德国心血管研究中心（DZHK）莱茵-美因伙伴站点，美因茨约翰内斯·古腾堡大学医学中心；德国美因茨朗根贝克街1号，邮编55131。电子邮箱：goekhan.guel@unimedizin-mainz.de

arXiv:2602.08174v1 [math.ST] 9 Feb 2026

摘要

本文针对分布不确定性下的渐近极小极大稳健假设检验，构建了一个统一框架，适用于贝叶斯和奈曼–皮尔逊两种表述（第一类和第二类）。本文考虑基于KL散度、 α-散度及其对称化变体的不确定类。利用Sion极小极大定理和Karush-Kuhn-Tucker条件，本文证明了所得稳健检验的存在性和唯一性，并以封闭的参数形式推导出最不利分布及相应的鲁棒似然比函数，从而可通过非线性方程组进行计算。本文证明，Dabak方法无法得到渐近极小极大鲁棒检验。本文提出的理论通过更系统、全面地推导稳健检验，推广了早期研究。数值模拟验证了理论结果，并展示了所推导稳健检验的表现。

索引词

假设检验、事件检测、鲁棒性、最不利分布、极小极大优化。

I. 引言

最优假设检验的设计依赖于一个假设，即观测数据的统计模型已被完全知晓 [1]。然而，底层分布往往无法完全获知。训练数据有限、非平稳环境或不完善的物理模型都可能导致模型失配，进而使实际性能显著偏离标称性能。为应对这些挑战，稳健假设检验应运而生，旨在即使真实分布不同于其标称描述，也能确保决策可靠 [2]。Huber 和 Strassen 的早期研究建立了一个原则性框架，通过一组概率测度来描述分布不确定性，从而涵盖相对于标称模型的可接受偏差 [3], [4]。这一表述奠定了极小极大鲁棒方法的基础：其重点从针对单一模型优化性能，转向确保性能能够抵御最不利的可接受替代情形。

在极小极大鲁棒框架中，假设每个假设下的真实分布都属于不确定类，这些不确定类通常由统计邻域或基于散度的约束定义 [3], [5]。目标是将可达到的最大风险降至最低，从而得到在最坏情况下仍保持最优的决策规则。由此得到的检验是基于最不利分布（LFDs）构造的似然比检验[1]，这些分布代表不确定类中的极端元素。这类形式化方法能够保证在有界模型扰动下的性能，也是鲁棒统计推断中的核心概念。

尽管保证性能至关重要，但极小极大鲁棒检验并非必然存在，其存在与否在很大程度上取决于不确定类的选择。如果在确定性决策规则类中不存在极小极大检验，那么在随机化决策规则中仍可能存在 [6]。然而，一个显著的局限是，这类检验通常仅对单个样本具有极小极大鲁棒性，无法在保持鲁棒性的同时扩展到多个样本 [7]。在不存在固定样本极小极大鲁棒检验时，一个自然的替代方案是考虑渐近极小极大鲁棒检验；随着样本数量增加，这类检验可使错误概率的指数衰减率最小。

[1]为简洁起见，除非可能引起混淆，否则不区分最不利分布及其对应的密度。

A. 相关工作

极小极大稳健假设检验奠基于 Huber 和 Strassen 的开创性研究。Huber 针对 ϵ-污染和全变差邻域建立了极小极大表述，确定了最不利分布，并证明相应的截尾似然比检验对于任意有限样本量 [3] 都具有极小极大鲁棒性。随后，研究将理论推广至 2-交替容量 [4] ，使极小极大鲁棒性的理论适用范围超越了污染型模型。在这些进展的基础上，该框架进一步扩展到其他不确定性表述，例如以 KL散度等信息论散度测度为基础的表述 [7]。

Dabak 和 Johnson 观察到，从 ϵ 污染模型中排除非光滑分布，可以得到更适合实际场景的不确定性模型，尤其是在建模误差为主要关注点的情况下。基于这一想法，他们提出使用 KL散度 来定义不确定性类，并针对渐近情形（即测量次数趋于无穷时）推导了相应的稳健检验 [8]。在若干假设下，Levy 证明，只要允许随机化决策规则，就可以为同一不确定性模型设计单样本极小极大稳健检验。随后，在后来的工作中，Levy 的全部假设均被去除 [7]。

KL散度表述的一个关键局限在于，散度度量与假设的先验概率都被设计为固定且不可调整 [7]。通过用 α 散度替换 KL散度，这些约束得以解除 [6]。值得注意的是，对于整个 α 散度邻域以及假设的任意先验概率，相应的极小极大鲁棒检验都是具有明确定义的随机化函数的删失似然比检验 [6]。

从更广泛的视角来看，近年来稳健假设检验的研究一直与分布鲁棒优化领域同步发展 [9]。人们提出了各种表述方式来刻画分布不确定性，包括基于Wasserstein的模型 [10],、Sinkhorn型扩展 [11], 以及基于核的不确定性集 [12], [13]。这些方法提供了具有强有力分析保证的凸或近似可处理表述。矩约束不确定性类 [14]和基于经验分布的方法 [15]进一步将稳健检验推广到数据驱动场景，并支持实际的有限样本实现。此外，还有研究探讨了去中心化 [16],、序贯 [17],以及对抗鲁棒的 [18]

场景。人们还研究了其他散度度量，例如基于海林格距离的度量，以在保留有限样本保证的同时，提高对分布设定错误的鲁棒性 [19], [20]。

B. 研究动机

1） [8], [21], 中的推导（随后在 [22], 中总结）并未得出渐近极小极大鲁棒的奈曼-皮尔逊（NP）检验。因此，要实现极小极大鲁棒性，就必须进行明确的推导和分析。2）渐近情形下的理论设计默认采用 NP 形式，这可能是因为由此得到的解更简单 [8], [23]。然而，正如 Chernoff 所示，NP 检验的错误概率指数衰减速度最慢 [24]。因此，为了使错误概率以最快速度衰减，有必要求得渐近极小极大鲁棒检验。

C. 论文概要及其贡献

本文建立了一个形式化框架，用于针对各种不确定类设计渐近极小极大鲁棒二元假设检验。本文在一般情形下证明了极小极大鲁棒检验的存在性和唯一性。根据 Karush-Kuhn-Tucker（KKT）条件，推导出最不利分布和鲁棒似然比函数（LRF）的参数形式，并可通过求解一组非线性方程将其明确表示出来。下文将总结本文的贡献及其与先前研究的关系。

1) 文中证明，任何渐近极小极大鲁棒检验都可以通过求解以下问题来设计

$$\min_{u\in(0,1)} \max_{(G_0,G_1)\in\mathscr{G}_0\times\mathscr{G}_1} D_u(G_0,G_1),$$

其中

$$D_u(G_0,G_1)=\int_\Omega {g_1}^u {g_0}^{1-u} d\mu$$

记作 u-亲和度（参见第三节-A）。

2) 对于基于 KL散度、 α-散度和对称化α-散度的不确定类，渐近极小极大稳健贝叶斯检验的最不利分布可用参数形式表示（见定理 IV.1、定理 IV.3 和定理 IV.4）。

3) 对于 KL 散度邻域，第一类和第二类渐近极小极大稳健奈曼–皮尔逊检验 的最不利分布可用参数形式表示。最不利分布的参数可通过求解四个非线性耦合方程得到（见 (26)），相应的检验不同于 [8], [21], [22]中推导的检验（见定理 IV.2 和注 IV.2）。这些结果为在一般不确定类下，针对贝叶斯和奈曼–皮尔逊设置设计稳健假设检验提供了系统框架。

D. 论文结构

本文其余部分安排如下。第二节介绍渐近极小极大鲁棒性。第三节分别在贝叶斯和 Neyman-Pearson 意义下推导渐近极小极大鲁棒性的方程，刻画鞍点值条件，并提出问题陈述。第四节针对基于 KL散度、 α-散度和对称化 α-散度的不确定类，求得最不利分布和渐近极小极大鲁棒检验。第五节通过仿真评估并展示理论推导结果。最后，第六节对全文进行总结。

E. 符号说明

本文通篇采用以下符号。大写字母表示概率分布和随机变量，相应的小写字母分别表示密度函数和观测值。黑体符号表示随机变量序列、观测值序列或联合函数。假设 $\mathcal{H}_0$ 和 $\mathcal{H}_1$ 分别与名义概率测度 F_0 和 F_1相关联，而相应的实际分布记为 G_0 和 G_1。概率分布集合记为 $\mathscr{G}_0$和 $\mathscr{G}_1$。每个概率测度（例如 $G[\cdot]$）都与其分布函数 $G(\cdot)$及密度函数 $g(\cdot)$相关联。记号$(\cdot)$表示最不利分布 $G_j \in \mathscr{G}_j$、相应的密度g_j或鲁棒似然比检验 δ。随机变量 $Y \sim G_j$ 的期望值记为 $\mathbb{E}_{G_j}[Y]$。后续运算的自变量（定义域上的取值）用 arg 表示。

II. 渐近极小极大鲁棒性

设 $\boldsymbol{Y} = (Y_1、\ldots\ldots、Y_n)$ 是一列 n 独立同分布的随机变量（r.v.s），每个变量的取值均属于 Ω。在 Y_k 假设下，其分布记为 G_j

假设 $\mathcal{H}_j$ 并非确切已知，而是属于不确定类 $\mathscr{G}_j$。给定 $\boldsymbol{Y}$，目标是判定以下哪个假设为真

$$\mathcal{H}_0 : Y_k \sim G_0, \quad G_0 \in \mathscr{G}_0,$$
$$\mathcal{H}_1 : Y_k \sim G_1, \quad G_1 \in \mathscr{G}_1, \tag{1}$$

其中 $k \in \{1, \ldots, n\}$。令 $P_F = G_0[\delta(\boldsymbol{Y}) = 1]$ 为虚警概率，$P_M = G_1[\delta(\boldsymbol{Y}) = 0]$ 为漏检概率，$\pi_0 = P(\mathcal{H}_0)$ 为假设 $\mathcal{H}_0$ 的先验概率。此外，令 $\delta : \boldsymbol{Y} \mapsto \{0, 1\}$ 为统计检验，并

$$P_E(\delta, G_0, G_1) = \pi_0 P_F(\delta, G_0) + (1 - \pi_0) P_M(\delta, G_1) \tag{2}$$

ˆ定义总体错误概率。于是，我们希望找到一个检验 δ 以及最不利分布 G_0 和 G_1，使其求解极小极大优化问题

$$\min_{\delta} \max_{(G_0, G_1) \in \mathscr{G}_0 \times \mathscr{G}_1} P_E(\delta, G_0, G_1). \tag{3}$$

ˆˆ令 $l = \hat{g}_1/\hat{g}_0$，定义 $X_k = \log l(Y_k)$，并用 $S_n(\boldsymbol{X}) = \frac{1}{n} \sum_{k=1}^{n} X_k$ 表示经验均值。由于极小极大问题 (3) 可由对应于最不利分布的（对数）似然比检验求解，因此鲁棒决策规则的形式为

$$\hat{\delta}(\boldsymbol{X}) = \begin{cases} 0, & S_n(\boldsymbol{X}) < t, \\ 1, & S_n(\boldsymbol{X}) \geq t, \end{cases} \tag{4}$$

其中，$t \in \mathbb{R}$ 是用于使最坏情况错误概率 P_E最小化的决策阈值。本文关注 n 足够大的渐近情形，尤其考察 $n \to \infty$ 时的行为。下文将陈述克拉默定理的扩展版本，以描述虚警和漏检概率 [25]的渐近衰减率。这些结果随后将作为推导极小极大最优性条件的基础，进而用于推导参数形式的最不利分布。

定理 II.1。令 $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ 为一列独立同分布随机变量，其中每个 Y_k 在 $\mathcal{H}_j$下均服从 $G_j \in \mathscr{G}_j$ 。此外，假设每个 X_k 的矩母函数有限

$$M_{X_k}^{j}(u) = \mathbb{E}_{G_j}\left[\exp\left(uX_k\right)\right] < \infty, \quad j \in \{0, 1\}. \tag{5}$$

对于由(4)给出的检验，若

$$\mathbb{E}_{G_0}[X_k] < t < \mathbb{E}_{G_1}[X_k], \tag{6}$$

那么，当 $n \to \infty$，虚警概率和漏检概率呈指数下降

$$\lim_{n\to\infty} \frac{1}{n} \log G_0[S_n(\boldsymbol{X}) > t] = -I_0(t) \tag{7}$$

$$\lim_{n\to\infty} \frac{1}{n} \log G_1[S_n(\boldsymbol{X}) \leq t] = -I_1(t) \tag{8}$$

其中速率函数为

$$I_j(t) = \sup_{u\in\mathbb{R}} \left(tu - \log M_{X_k}^j(u)\right), \quad j \in \{0, 1\}. \tag{9}$$

证明：分别对随机变量序列 $(X_k)_{k\geq 1}$ 在 $\mathcal{H}_0$ 下（其中每个随机变量都满足 $\mathbb{E}_{G_0}[X_k] < t$），以及随机变量序列 $(-X_k)_{k\geq 1}$ 在 $\mathcal{H}_1$ 下（其中每个随机变量都满足 $t < \mathbb{E}_{G_1}[X_k]$），各应用两次 Cramér 定理 [25]，即可得到结果（分别为 (7) 和 (8)）。■

根据定理 II.1，可以如下定义渐近极小极大鲁棒性。

定义 II.1。</style}形式为 (4) 的检验称为对阈值 t 渐近极小极大稳健，如果

$$\lim_{n\to\infty} \frac{1}{n} \log G_0 \left[S_n(\boldsymbol{X}) > t\right] \leq \lim_{n\to\infty} \frac{1}{n} \log \hat{G}_0 \left[S_n(\boldsymbol{X}) > t\right]$$

and

$$\lim_{n\to\infty} \frac{1}{n} \log G_1 \left[S_n(\boldsymbol{X}) \leq t\right] \leq \lim_{n\to\infty} \frac{1}{n} \log \hat{G}_1 \left[S_n(\boldsymbol{X}) \leq t\right]$$

对所有 $(G_0,\ G_1) \in \mathscr{G}_0 \times \mathscr{G}_1$ 均成立。

注 II.1. 定义 II.1 隐含地依赖正则性条件 (5) 和 (6)。这些假设较为宽松，并不苛刻。例如，只要不确定性类存在共同支配测度且具有共同支撑集，条件 (5) 就成立；而条件 (6) 则可通过选择合适的阈值（例如 $t = 0$）来确保。

1) 最优阈值：使错误概率渐近下降速率达到最大的 t 最优值尤为值得关注。Chernoff [24] 首先针对一对固定分布解决了这一问题。下文将给出一种解法，但起初并不假设数据样本来自 $G_0 \in \mathscr{G}_0$ 和 $G_1 \in \mathscr{G}_1$，参见定理 II.1。若实际的数据生成分布与极小极大检验的最不利分布不一致，这一视角对于刻画错误指数的渐近行为至关重要。

定理II.2。　　渐近意义下使错误概率最小的阈值为

$$\arg\min_{t} \lim_{n\to\infty} P_E(\hat{\delta}, G_0, G_1) = 0, \tag{10}$$

ˆ　　ˆ若 $G_0 := G_0$ 且 $G_1 :$
$= G_1$。

定理II.2的证明见附录 A。推导表明，当真实的数据生成分布与用于构造检验的分布一致，即 $G_j := G_j$ 时， $t = 0$ 保证是最优的。然而，在存在任何模型失配时，最优阈值可能非零。

III. 极小极大方程的推导与问题表述

下文将推导贝叶斯检验（$t = 0$）和奈曼–皮尔逊型渐近极小极大鲁棒检验。这一更广泛的问题表述既便于与早期结果进行比较，也提供了一个用于刻画渐近鲁棒性的统一框架。特别是，它凸显了与 Dabak 的推导 [8]之间的关键差异。

除非另有说明，渐近极小极大鲁棒检验指阈值为 $t = 0$ 的贝叶斯检验，这与最小化错误概率的目标一致。尽管定义 II.1 采用一般阈值以涵盖奈曼–皮尔逊型检验，但真正的极小极大鲁棒解出现在 $t = 0$。因此，以下鞍点值条件和问题表述均聚焦于此情形。

A. 贝叶斯检验

根据上一节，矩母函数 $M^0_{X_k}(u)$ 等价于 u-亲和度，其定义如下

$$D_u(G_0, G_1) = \int_\Omega {g_1}^u {g_0}^{1-u} d\mu,$$

其中， g_j 表示 G_j、 $j \in \{0, 1\}$ 的密度。事实上， D_u 是 Rényi 散度 [26] 的非缩放版本。以下列出后续推导中用到的 u-亲和度的一些重要性质：

1) D_u 是非负的，即 $D_u(G_0, G_1) \geq 0$2) D_u 关于 u连续且凸3) D_u 关于(g_0, g_1)连续且联合凹
4) $D_{u=0}(G_0, G_1) = D_{u=1}(G_0, G_1) = 1$

5) $D_u(G_0, G_1) \in [0,1]$，这可由性质1和H¨older不等式 [27] 得出 .

6) 若 $\hat{u} = \arg\min_u D_u(G_0, G_1)$，则 $\hat{u} \in (0, 1)$，由性质2和4可得。

7) D_u 与 α-散度的关系为 $D_\alpha(G_0, G_1) = (1 - D_u(G_0, G_1))/u(1-u)|_{u:=\alpha}$

性质1和 $4-7$ 可直接由定义和基本不等式得出。性质2和3继承自 α-散度已确立的性质，并使用性质 7 [28]。

将要推导的极小极大鲁棒检验是一个对数似然比检验(4)，其中检验 $\hat{}$ $\hat{}$ 和数据样本均使用最不利分布 G_0 和 G_1 构造。如定理II.2的证明所示，这为阈值 $t = 0$ 的选择提供了依据。因此，利用恒等式 $-\inf(x) = \sup(-x)$（对于 $x < 0$）以及 $\inf(-x) = -\sup(x)$，再结合 $t = 0$，可由(5)和(9)推导出极小极大方程：

$$\hat{g}_0 = \arg\sup_{G_0 \in \mathscr{G}_0} \left(\inf_{u_0 \in \mathbb{R}} \log \int_\Omega \left(\frac{\hat{g}_1}{\hat{g}_0} \right)^{u_0} g_0 d\mu \right),$$
$$\hat{g}_1 = \arg\sup_{G_1 \in \mathscr{G}_1} \left(\inf_{u_1 \in \mathbb{R}} \log \int_\Omega \left(\frac{\hat{g}_1}{\hat{g}_0} \right)^{u_1} g_1 d\mu \right). \tag{11}$$

这些耦合方程通常在数学上难以处理，尤其是当不确定性集合 $\mathscr{G}_0$ 和 $\mathscr{G}_1$ 为无限维时。然而，如果极小极大问题中上确界部分的解——即最不利密度 $\hat{g}_0$ 和 $\hat{g}_1$——已知，那么针对决策规则的相应优化就可以等价地用u-亲和代价函数表示为

$$C_0(\hat{G}_0, \hat{G}_1; \hat{u}_0) = \inf_{u_0 \in \mathbb{R}} \log D_{u_0}(\hat{G}_0, \hat{G}_1),$$
$$C_1(\hat{G}_0, \hat{G}_1; \hat{u}_1) = \inf_{u_1 \in \mathbb{R}} \log D_{1+u_1}(\hat{G}_0, \hat{G}_1). \tag{12}$$

命题 III.1. 当(12) 中的两个优化问题的解满足 $\hat{u}_0 = -\hat{u}_1$，$\hat{}$ $\hat{}$ $\hat{}$ $\hat{}$ 且 $\hat{u}_0 \in [0,1]$ 时，两者的解相同，即 $C_0(G_0, G_1; \hat{u}_0) = C_1(G_0, G_1; -\hat{u}_1)$。

证明： 由 D_u 的性质6可知，(12)中的下确界在唯一的极小值点 $\hat{u}_0 \in [0,1]$ 处取得。根据类似的推理，(12)中第二个方程的极小值点 $-\hat{u}_1$ 也位于区间 $[0,1]$ 内；另见(11)中耦合方程的结构。此外，由于 $-C_0$ 对应于 Chernoff距离，它关于自变量 [22,p. 82],对称，因此我们有

$$C_0(\hat{G}_0, \hat{G}_1; \hat{u}_0) = C_0(\hat{G}_1, \hat{G}_0; \hat{u}_0) = C_1(\hat{G}_0, \hat{G}_1; -\hat{u}_1).$$

至此，证明完成。 ■

命题 III.1 表明，这两个优化问题等价，并且对于 $\hat{u}_0 = -\hat{u}_1$，它们具有相同的解。因此，只需考虑其中一个即可。考虑第一个形式并去掉对数项后，待求解的问题可简化为

$$(\hat{g}_0, \hat{g}_1) = \arg \sup_{(G_0,G_1)\in\mathscr{G}_0\times\mathscr{G}_1} \inf_{u\in(0,1)} D_u(G_0, G_1). \tag{13}$$

这是可行的，因为对于 sup 和 inf，我们都有

$$\frac{\partial \log D_u}{\partial u} = 0 \Longrightarrow \frac{\partial D_u}{\partial u} = 0,$$

并且 log 是从 $[0,1]$ 到 $\mathbb{R}_{\leq 0}$的递增映射，参见 D_u的性质 5。

B. Neyman-Pearson 检验

渐近 Neyman–Pearson 检验的表述方式是：使其中一个错误指数达到其可能的最大衰减速率，而另一个错误指数则不可避免地以最小速率衰减。考虑对数似然比检验，将 $X_k := \log l(Y_k)$ 与阈值 t 进行比较，其中 $l = g_1/g_0$。对于 I类 NP 检验，阈值选择为

$$t_0 = \lim_{\epsilon\to 0} \mathbb{E}_{G_0}[\log l(Y_1)] + \epsilon$$

使得 P_F 渐近地保证低于任意 $\epsilon > 0$，而 P_M 具有最高的衰减速率。类似地，对于 II类 NP 检验，阈值选择为

$$t_1 = \lim_{\epsilon\to 0} \mathbb{E}_{G_1}[\log l(Y_1)] - \epsilon$$

使得 P_M 渐近地保证低于任意 $\epsilon > 0$，而 P_F 具有最高的衰减速率。对 (9) 中上确界内的表达式关于 u 求导，并令其等于零，可得相应速率函数 $I_j(t)$ 的驻定条件。特别地，在 $\mathcal{H}_j$ 下，选择 $t = t_j$ 意味着 $u \to 0$。类似地，在 $\mathcal{H}_0$ 下，$t = t_1$ 意味着 $u \to 1$；而在 $\mathcal{H}_1$ 下，$t = t_0$ 意味着 $u \to -1$。因此，对于 I类 NP 检验，可得 $I_0(t_0) = 0$ 和 $I_1(t_0) = -t_0 = D_{\mathrm{KL}}(G_0, G_1)$；而对于 II类 NP 检验，可得 $I_1(t_1) = 0$ 和 $I_0(t_1) = t_1 = D_{\mathrm{KL}}(G_1, G_0)$。因此，通过求解以下问题，可以得到渐近极小极大稳健 NP 检验

$$\begin{aligned}
&\text{Type-I NP-test:} && \min_{(G_0,G_1)\in\mathscr{G}_0\times\mathscr{G}_1} D_{\mathrm{KL}}(G_0, G_1) \quad \text{s.t. } g_0 > 0,\ g_1 > 0,\ \Upsilon(G_0) = 1,\ \Upsilon(G_1) = 1,\\
&\text{Type-II NP-test:} && \min_{(G_0,G_1)\in\mathscr{G}_0\times\mathscr{G}_1} D_{\mathrm{KL}}(G_1, G_0) \quad \text{s.t. } g_0 > 0,\ g_1 > 0,\ \Upsilon(G_0) = 1,\ \Upsilon(G_1) = 1.
\end{aligned} \tag{14}$$

值得注意的是，阈值 t_0 和 t_1 对应于式 (6) 所允许的两个极端值，而贝叶斯检验则使用中点 $t = 0$。

C. 鞍点值条件

本节将讨论鞍点值的存在性，从而讨论式 (13) 中极小极大优化问题的解。该解的唯一性取决于不确定类的选择，将在下一节中讨论。一般而言，鞍点值的存在性由以下方程的解来描述:

$$\min_{u\in(0,1)} \sup_{(G_0,G_1)\in\mathscr{G}_0\times\mathscr{G}_1} D_u(G_0,G_1) = \sup_{(G_0,G_1)\in\mathscr{G}_0\times\mathscr{G}_1} \min_{u\in(0,1)} D_u(G_0,G_1). \tag{15}$$

(15)中的等式通常只以不等式（$\geq$）成立，因此不能先验地假定它成立；要证明这一点，需要对目标函数和不确定性集进行严格分析。例如，可以借助Sion极小极大定理 [29]来证明。如附录B所示，应用该定理可保证(15)存在鞍点值，从而得到

$$D_{\hat{u}}(G_0,G_1) \leq D_{\hat{u}}(\hat{G}_0,\hat{G}_1) \leq D_u(\hat{G}_0,\hat{G}_1), \tag{16}$$

ˆ ˆ 其中，ˆu 是使 u最小化的 u，而 G_0 和 G_1 是最不利分布。

D. 问题陈述

ˆ ˆ 根据(16)，给定(G_0, G_1)，需要对 u 最小化目标函数 D_u；给定最小化的ˆu后，需要在(G_0, G_1)上最大化 $D_{\hat{u}}$。这可以简洁地写为

$$\begin{aligned}
\text{Maximization:}\quad & \hat{g}_0 = \arg\sup_{G_0\in\mathscr{G}_0} D_u(G_0,G_1) \quad \text{s.t. } g_0>0,\ \Upsilon(G_0)=\int_\Omega g_0\,d\mu = 1 \\
& \hat{g}_1 = \arg\sup_{G_1\in\mathscr{G}_1} D_u(G_0,G_1) \quad \text{s.t. } g_1>0,\ \Upsilon(G_1)=\int_\Omega g_1\,d\mu = 1 \\
\text{Minimization:}\quad & \hat{u} = \arg\min_{u\in(0,1)} D_u(\hat{G}_0,\hat{G}_1).
\end{aligned} \tag{17}$$

最大化阶段涉及两个相互耦合的独立约束优化问题。一旦将ˆg_0 和ˆg_1 表示为 u的函数，就可以求解以下最小化问题。

IV. 渐近极小极大鲁棒检验的推导

本节针对由(17)给出的极小极大优化问题，为各种不确定性类推导了最不利分布和渐近极小极大鲁棒检验。此外，还推导了渐近奈曼–皮尔逊检验。对库尔贝克–莱布勒（KL 散度）散度邻域给出了完整推导；当对 α散度以及对称化 α散度重复同一过程时，为清晰起见略去类似步骤。

A. KL 散度邻域

考虑不确定性类

$$\mathscr{G}_j = \{G_j : D_{\mathrm{KL}}(G_j, F_j) \le \epsilon_j\}, \quad j \in \{0, 1\}, \tag{18}$$

由 KL 散度诱导

$$D_{\mathrm{KL}}(G_j, F_j) = \int_\Omega \log\left(g_j/f_j\right) g_j d\mu,$$

其中， F_j 是 $\mathcal{H}_j$下的标称分布。KL 散度被视为经典信息散度 [30] ，并在早期研究中用于构造不确定性类 [8], [5]。它是一种平滑的距离，因此适用于处理建模误差 [7]。

1) 贝叶斯检验：针对 KL 散度邻域的渐近极小极大鲁棒检验可由以下定理表述。

定理 IV.1. 对于由 (18) 给出的不确定性类，最不利密度

$$\hat{g}_0 = \exp\left[\frac{-\lambda_0 - \mu_0}{\lambda_0}\right] \exp\left[\frac{(1-u)(\hat{g}_1/\hat{g}_0)^u}{\lambda_0}\right] f_0,$$
$$\hat{g}_1 = \exp\left[\frac{-\lambda_1 - \mu_1}{\lambda_1}\right] \exp\left[\frac{u(\hat{g}_1/\hat{g}_0)^{-1+u}}{\lambda_1}\right] f_1,$$

以及鲁棒似然比函数

$$\frac{\hat{g}_1}{\hat{g}_0} = \exp\left[\frac{-\mu_1}{\lambda_1} + \frac{\mu_0}{\lambda_0}\right] \exp\left[\frac{u(\hat{g}_1/\hat{g}_0)^{-1+u}}{\lambda_1} + \frac{(-1+u)(\hat{g}_1/\hat{g}_0)^u}{\lambda_0}\right] l \tag{19}$$

为(17)提供唯一解。此外，给定 u，通过求解可以得到拉格朗日参数 λ_0 和 λ_1，进而得到 μ_0 和 μ_1

$$\int_\Omega r_0 \log\left(r_0/s_0\right) f_0 d\mu/s_0 = \epsilon_0,$$
$$\int_\Omega r_1 \log\left(r_1/s_1\right) f_1 d\mu/s_1 = \epsilon_1,$$
$$\hat{g}_1/\hat{g}_0 = (r_1 s_0)/(r_0 s_1), \tag{20}$$

其中

$$s_0(\lambda_0) = \int_\Omega r_0(\lambda_0, \hat{g}_1/\hat{g}_0) f_0 d\mu = \int_\Omega \exp\left[\frac{(1-u)(\hat{g}_1/\hat{g}_0)^u}{\lambda_0}\right] f_0 d\mu = \exp\left[\frac{\lambda_0 + \mu_0}{\lambda_0}\right],$$
$$s_1(\lambda_1) = \int_\Omega r_1(\lambda_1, \hat{g}_1/\hat{g}_0) f_1 d\mu = \int_\Omega \exp\left[\frac{u(\hat{g}_1/\hat{g}_0)^{-1+u}}{\lambda_1}\right] f_1 d\mu = \exp\left[\frac{\lambda_1 + \mu_1}{\lambda_1}\right].$$

证明：考虑拉格朗日函数

$$L_0(g_0, g_1, \lambda_0, \mu_0) = D_u(G_0, G_1) + \lambda_0(\epsilon_0 - D_{\mathrm{KL}}(G_0, F_0)) + \mu_0(1 - \Upsilon(G_0)), \tag{21}$$

其中 λ_0 和 μ_0 是KKT 乘子。当所有KKT条件均满足时，(21)的解可以唯一确定 [31, 第 5]章，因为 L_0 是g_0的严格凹泛函，这是由于对每个 $\lambda_0 > 0$都有 $\partial^2 L_0/\partial g_0^2 < 0$ 。将(21)显式写出，可得

$$L_0(g_0, g_1, \lambda_0, \mu_0) = \int_\Omega \left[\left(\frac{g_1}{g_0} \right)^u - \lambda_0 \log \left(\frac{g_0}{f_0} \right) - \mu_0 \right] g_0 d\mu + \lambda_0 \epsilon_0 + \mu_0. \tag{22}$$

施加 KKT 乘子的驻定条件，并据此沿着 ψ_0 的方向对方程 (22) 求 Gâteaux 导数，得到

$$\int_\Omega \left[(1-u) \left(\frac{g_1}{g_0} \right)^u - \lambda_0 \log \left(\frac{g_0}{f_0} \right) - \lambda_0 - \mu_0 \right] \psi_0 d\mu,$$

这意味着

$$(1-u) {g_0}^{-u} {g_1}^u - \lambda_0 \log g_0 = \lambda_0 + \mu_0 - \lambda_0 \log f_0, \tag{23}$$

因为 ψ_0 是任意函数。类似地，对

$$L_1(g_0, g_1, \lambda_1, \mu_1) = D_u(G_0, G_1) + \lambda_1(\epsilon_1 - D_{\mathrm{KL}}(G_1, F_1)) + \mu_1(1 - \Upsilon(G_1)), \tag{24}$$

关于 g_1 沿着 ψ_1 的方向求 Gâteaux 导数，得到

$$u {g_1}^{-1+u} {g_0}^{1-u} - \lambda_1 \log g_1 = \lambda_1 + \mu_1 - \lambda_1 \log f_1. \tag{25}$$

根据 (23) 和 (25)，最不利密度可表示为鲁棒似然比函数 $\hat{}g_1/\hat{}g_0$ 和标称密度 f_0 和 f_1 的泛函，如定理 IV.1 所述。由于第二个拉格朗日函数 L_1 对每个 $\lambda_1 > 0$ 都是严格凹的，因此最不利分布的参数形式是唯一确定的。类似地，对于任意一对 $(G_0, G_1) \in \mathscr{G}_0 \times \mathscr{G}_1$，$D_u$ 关于 $u \in (0, 1)$ 都是严格凸的。这确保了对应于 $\hat{}g_0$ 和 $\hat{}g_1$ 的极小化 u 也是唯一的。为确定参数，对于任意给定的 u值，必须求解以下四个非线性方程组成的方程组：

$$\begin{aligned}
\Upsilon(\hat{G}_0(\lambda_0, \mu_0, \lambda_1, \mu_1)) &= 1 \\
\Upsilon(\hat{G}_1(\lambda_0, \mu_0, \lambda_1, \mu_1)) &= 1 \\
D_{\mathrm{KL}}(\hat{G}_0(\lambda_0, \mu_0, \lambda_1, \mu_1), F_0) &= \epsilon_0 \\
D_{\mathrm{KL}}(\hat{G}_1(\lambda_0, \mu_0, \lambda_1, \mu_1), F_1) &= \epsilon_1
\end{aligned} \tag{26}$$

以及附加条件 (19)。这个五元方程组可以在不失一般性的前提下化简为三元方程组。具体而言，根据前两个方程，参数 s_0 和 s_1可分别表示为 r_0 和 r_1的泛函，如定理 IV.1 所示。将 (r_0，s_0) 和 (r_1，s_1) 代入 (26) 的最后两个方程以及 (19)，即可得到 (20) 中给出的三元方程组。

■

备注 IV.1。 根据定理 IV.1，所采用的策略是先执行最大化步骤，以参数形式求得最不利密度，然后执行最小化步骤，确定最优的 $u \in (0, 1)$。尽管对于任意一对已知密度， D_u 关于 u 都是凸函数，但定理 IV.1 给出的最不利密度的参数形式不一定也具有这种凸性。具体而言，通过验证非负性 来确立凸性并非易事，而且反例很容易找到（见第 V 节）。这表明，确定 $u \in (0, 1)$ 的最小化取值，一种切实可行的方法是求解方程组 (20)，并附加一个显式地使 u 最小化的方程。

2) Neyman-Pearson 检验： I 类 NP 检验的解可由以下定理给出。

定理 IV.2。 渐近极小极大鲁棒 *I* 类 *NP* 检验的最不利密度为

$$\hat{g}_0 = \left(\lambda_1 + \frac{\lambda_1}{\lambda_0}\right)^{-\frac{1}{\lambda_0}} \exp\left[-1 - \frac{1+\mu_0}{\lambda_0}\right] W\left[\frac{\lambda_0 \exp\left[\frac{-\lambda_1-\lambda_1\mu_0+\lambda_0\mu_1}{(1+\lambda_0)\lambda_1}\right] l^{-\frac{\lambda_0}{1+\lambda_0}}}{(1+\lambda_0)\lambda_1}\right]^{-\frac{1}{\lambda_0}} f_0$$

$$\hat{g}_1 = \left(\lambda_1 + \frac{\lambda_1}{\lambda_0}\right)^{-\frac{1+\lambda_0}{\lambda_0}} \exp\left[-1 - \frac{1+\mu_0}{\lambda_0}\right] W\left[\frac{\lambda_0 \exp\left[\frac{-\lambda_1-\lambda_1\mu_0+\lambda_0\mu_1}{(1+\lambda_0)\lambda_1}\right] l^{-\frac{\lambda_0}{1+\lambda_0}}}{(1+\lambda_0)\lambda_1}\right]^{-\frac{1+\lambda_0}{\lambda_0}} f_0 \quad (27)$$

其中 W 是 *Lambert-W* 函数。

证明：因此 再次可以通过 KKT 乘子获得解。考虑拉格朗日函数 s

$$L_0(g_0, g_1, \lambda_0, \mu_0) = D_{\mathrm{KL}}(G_0, G_1) + \lambda_0(D_{\mathrm{KL}}(G_0, F_0) - \epsilon_0) + \mu_0(\Upsilon(G_0) - 1),$$
$$L_1(g_0, g_1, \lambda_1, \mu_1) = D_{\mathrm{KL}}(G_0, G_1) + \lambda_1(D_{\mathrm{KL}}(G_1, F_1) - \epsilon_1) + \mu_1(\Upsilon(G_1) - 1), \quad (28)$$

并按照之前的相同步骤，分别可以得到

$$g_1 = \exp\left[1 + \lambda_0 + \mu_0\right] g_0^{1+\lambda_0} f_0^{-\lambda_0}, \tag{29}$$

$$g_0 = g_1 \left(\mu_1 + \lambda_1 \left(1 + \log\left(g_1/f_1\right)\right)\right). \tag{30}$$

分别求解 (29) 和 (30) 中的 g_0 和 g_1，即可得到定理 IV.2 所述的第一类渐近极小极大鲁棒 NP 检验的最不利密度。 ■

注 IV.2。 第二类极小极大鲁棒 NP 检验也可以通过对目标函数 $D_{\mathrm{KL}}(G_1, G_0)$ 采用相同的步骤，或考虑 (27) 中给出的相同方程来得到。若采用后一种方法，则在优化之前，必须分别将 ϵ_0 和 f_0 与 ϵ_1 和 f_1 互换；得到最不利密度之后，再将 $\hat{}g_0$ 与 $\hat{}g_1$ 互换，参见 (14)。随后，直接求解 (26) 中针对 (27) 所给最不利密度的方程，即可得到相应参数。

这两种 NP 检验 实际上都是贝叶斯渐近极小极大鲁棒检验的极限情形

$$\max_{(G_0,G_1)\in\mathscr{G}_0\times\mathscr{G}_1} D_u(G_0, G_1) \equiv \min_{(G_0,G_1)\in\mathscr{G}_0\times\mathscr{G}_1} D_\alpha(G_0, G_1), \quad \forall u = \alpha \in (0, 1),$$

根据 D_u 的第七条性质（另见第 IV-B 节），当 $\alpha \to 1$ 和 $\alpha \to 0$ 时， $D_\alpha(G_0, G_1)$ 分别收敛于 $D_{\mathrm{KL}}(G_0, G_1)$ 和 $D_{\mathrm{KL}}(G_1, G_0)$。这一恒等式凸显了 D_u 的核心作用：它提供了一个统一框架，通过适当选择 u，即可推导出贝叶斯鲁棒检验和奈曼–皮尔逊鲁棒检验。

$\hat{}$ 有趣的是， $l = \hat{}g_1/\hat{}g_0$ 是通过 W关于 l 的一个非线性泛函。此外，与贝叶斯渐近极小极大鲁棒检验不同，其奈曼–皮尔逊（NP）对应检验的最不利密度仅表示为名义$\hat{}$分布的泛函，而不与 l耦合。尽管这种简化避免了对鲁棒似然比的依赖，却使所得最不利密度的解析结构更加复杂。

请注意，式 (14) 中的问题表述与 Dabak方法 [8], [21]不同；另见[22, 第 250–255]页。具体而言，Dabak检验是通过对所有 $G_1 \in \mathscr{G}_1$ 联合极小化 $I_0(t_1)$，并对所有 $G_0 \in \mathscr{G}_0$ 联合极小化 $I_1(t_0)$ 得到的。这使得最不利密度的闭式表达式更简单，但由此得到的检验并非本文所考虑意义下的渐近极小极大稳健检验（见第 V 节）。尽管如此，事实证明，对于序贯概率比检验（序贯概率比检验）[32], [7]的期望样本数而言，Dabak检验具有渐近极小极大稳健性。

B. α–散度邻域

基于 Kullback–Leibler 散度的不确定性类的一种自然推广，是通过采用 α-散度得到的，

$$D_\alpha(G_j, F_j) := \frac{1}{\alpha(1-\alpha)} \left(\int_\Omega \left((1-\alpha) f_j + \alpha g_j - g_j^\alpha f_j^{1-\alpha} \right) d\mu \right), \quad \alpha \in \Omega \setminus \{0, 1\},$$

它属于更广义的 f-散度类，并将各种散度度量作为特殊情形加以涵盖 [33],[28, p。 1537]; 例如， D_α 在 $\alpha \to 1$时退化为 Kullback–Leibler散度D_{KL}。与 α-散度邻域相关的最不利分布由以下定理刻画。

定理 IV.3。 基于 α–散度的不确定性类的最不利密度为

$$\begin{aligned} \hat{g}_0 &= \left(\frac{1-\alpha}{\lambda_0} \left(\mu_0 - (1-u) \left(\frac{\hat{g}_1}{\hat{g}_0} \right)^u \right) + 1 \right)^{\frac{1}{\alpha-1}} f_0, \\ \hat{g}_1 &= \left(\frac{1-\alpha}{\lambda_1} \left(\mu_1 - u \left(\frac{\hat{g}_1}{\hat{g}_0} \right)^{-1+u} \right) + 1 \right)^{\frac{1}{\alpha-1}} f_1, \end{aligned} \tag{31}$$

其中

$$\hat{l} = \left(\frac{\frac{1-\alpha}{\lambda_1} \left(\mu_1 - u \left(\frac{\hat{g}_1}{\hat{g}_0} \right)^{-1+u} \right) + 1}{\frac{1-\alpha}{\lambda_0} \left(\mu_0 - (1-u) \left(\frac{g_1}{g_0} \right)^u \right) + 1} \right)^{\frac{1}{\alpha-1}} l. \tag{32}$$

证明： 证明采用与前面相同的拉格朗日方法，即在（21）和（24）中用 D_α 替换 D_{KL} ，然后进行推导。 ■

这些参数通过求解与（32）联立的四个非线性方程，以类似方式获得。

1) 特殊情形：式(31)中的最不利分布可以对某些特殊的参数选择显式写出。例如，若 $\alpha = 1/2$ 且 $u = 1/2$，鲁棒似然比函数可简化为

$$\hat{l} = \sum_{k=0}^{2} c_k l^{\frac{k}{2}}$$

其中

$$\begin{aligned} c_0 &= \frac{0.25{\lambda_0}^2}{4{\lambda_0}^2{\lambda_1}^2 + {\lambda_0}^2{\mu_0}^2 + 4{\lambda_0}^2\lambda_1\mu_0}, \\ c_1 &= \frac{\lambda_0(2\lambda_0\lambda_1 + \mu_0\lambda_1)}{4{\lambda_0}^2{\lambda_1}^2 + {\lambda_0}^2{\mu_0}^2 + 4{\lambda_0}^2\lambda_1\mu_0}, \\ c_2 &= \frac{2\lambda_0\lambda_1 + \mu_0\lambda_1}{4{\lambda_0}^2{\lambda_1}^2 + {\lambda_0}^2{\mu_0}^2 + 4{\lambda_0}^2\lambda_1\mu_0}. \end{aligned}$$

C. 对称 $\alpha-$散度邻域

α-散度通常不是一种对称的差异度量，但在 $\alpha = 1/2$ 处是一个显著的例外。 $\alpha-$散度的对称化版本，

$$D_\alpha^s(G_j, F_j) = \frac{1}{\alpha(1-\alpha)} \left(\int_\Omega ((f_j^\alpha - g_j^\alpha)(f_j^{1-\alpha} - g_j^{1-\alpha}))d\mu \right), \quad \alpha \in \Omega \backslash \{0, 1\}$$

可通过

$$D_\alpha^s(G_j, F_j) = D_\alpha(G_j, F_j) + D_\alpha(F_j, G_j).$$

对称 α-散度属于 f-散度类 [32], [34], ，该类涵盖若干广为人知的对称度量，例如对称 Kullback–Leibler（KL）散度和对称 χ^2 散度 [28]。对称 α-散度邻域下的最不利分布由以下定理刻画。

定理 IV.4. 对称 α-散度邻域下的最不利密度由以下耦合非线性方程组确定，

$$\frac{\lambda_0}{1-\alpha} \left(\frac{\hat{g}_0}{f_0} \right)^{2\alpha-1} + \left((1-u) \left(\frac{\hat{g}_1}{\hat{g}_0} \right)^u - \frac{\lambda_0}{\alpha(1-\alpha)} - \mu_0 \right) \left(\frac{\hat{g}_0}{f_0} \right)^\alpha + \frac{\lambda_0}{\alpha} = 0, \tag{33}$$

$$\frac{\lambda_1}{1-\alpha} \left(\frac{\hat{g}_1}{f_1} \right)^{2\alpha-1} + \left(u \left(\frac{\hat{g}_1}{\hat{g}_0} \right)^{u-1} - \frac{\lambda_1}{\alpha(1-\alpha)} - \mu_1 \right) \left(\frac{\hat{g}_1}{f_1} \right)^\alpha + \frac{\lambda_1}{\alpha} = 0. \tag{34}$$

证明： 证明方法是类似于先前的推导，应用拉格朗日优化框架。

■

一般而言，求解 (33) 和 (34) 需要同时处理由相应拉格朗日约束导出的另外四个非线性方程，从而形成一个包含六个方程的方程组。然而，如果固定 α ，方程数量可减少至五个。其思路是将比值 $\hat{g}_0/f_0$ 和 $\hat{g}_1/f_1$ 表示为似然比 $\hat{g}_1/\hat{g}_0$ 的函数，即，

$$\frac{\hat{g}_0}{f_0} = h_0 \left(\frac{\hat{g}_1}{\hat{g}_0} \right), \quad \text{and} \quad \frac{\hat{g}_1}{f_1} = h_1 \left(\frac{\hat{g}_1}{\hat{g}_0} \right), \tag{35}$$

其中， h_0 和 h_1 是某些函数。这得到耦合关系

$$\hat{l} = \frac{h_1(\hat{l})}{h_0(\hat{l})} l, \tag{36}$$

这使得系统可以进一步简化。

表 I 仿真中使用的一对标称分布

缩写	在 $\mathcal{H}_0$ 下	在 $\mathcal{H}_1$ 下
d_1	$\mathcal{N}(-1,\ 1)$	$\mathcal{N}(1,\ 1)$
d_2	$\mathcal{N}(-1,\ 1)$	$\mathcal{N}(1,\ 4)$
d_3	$\mathcal{L}(0,\ 1)$	$f_{\mathcal{L}}(y)(\sin(2\pi y)\ +1)$

五、仿真

本节将对理论结果进行评估并加以举例说明。记号 $|_a^n$ 表示使用(a)-检验进行检验，而数据样本则取自与(n)-检验相对应的分布。在所有理论示例中，均考虑表 I 中列出的标称分布。记号 $\mathcal{N}(\mu,\ \sigma^2)$表示均值为 μ、方差为 σ^2的高斯分布，而 $\mathcal{L}(0,\ 1)$表示具有相应参数的标准拉普拉斯分布。相应的密度函数分别记为 $f_{\mathcal{N}}$ 和 $f_{\mathcal{L}}$。求解所有方程组时均使用阻尼牛顿法 [35]。下文将说明最不利分布、鲁棒似然比函数、方程参数，以及 D_u 的（非）凸性。

A. 最不利分布与稳健似然比函数

利用表 I 中以 d_1 表示的对称分布对及鲁棒性参数$\epsilon_0 = \epsilon_1 = 0.1$，研究距离度量的选择与其对鲁棒性的影响。对于这一设置，图 1 展示了 KL散度（$\alpha \to 1$）以及各种 α散度下，最不利分布与标称分布的密度。为清晰起见，未列出对称化的 α散度。从这个例子中可以观察到两点：

1) 最不利分布并非高斯分布（图中未显示，但已通过曲线拟合验证）。2) 最不利分布的方差与 α成反比。

在图 2中，展示了相应的似然比函数，其中包括带有 $\alpha = 2$的 α-散度的对称化版本，记为 $\alpha_s = 2$。这些示例表明，鲁棒性是通过对名义似然比函数进行顺时针旋转，随后进行截断操作来实现的。值得注意的是，旋转和截断的程度随 α 的减小而增大。对于较小的

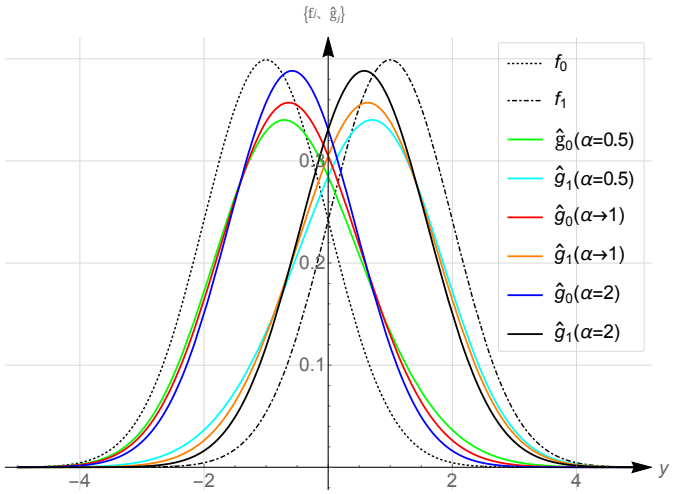

图 1. 表 I 中以 d_1 表示的标称分布，以及相对于其密度的对应最不利分布，针对$\epsilon_0 = \epsilon_1 = 0.1$，其中 $u = 0.5$ 对所有 α 均成立。

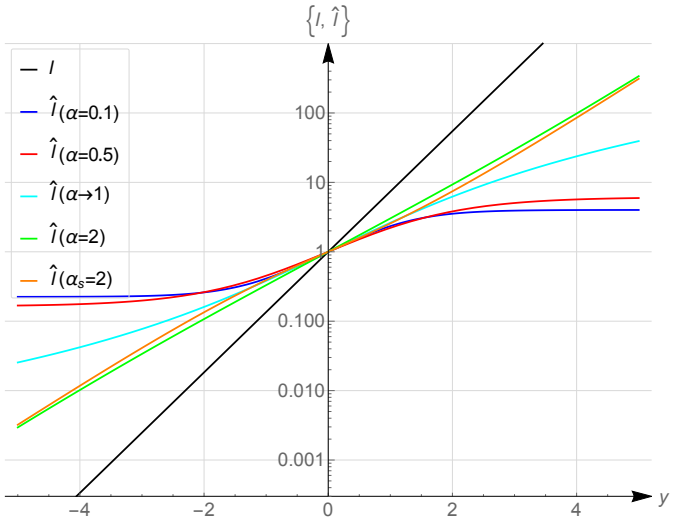

图 2. 与表 I 中以 d_1 缩写的标称分布相对应的鲁棒和名义似然比函数，其中 $\epsilon_0 = \epsilon_1 = 0.1$。对称情形 α_s 也包括在内。

α取值，鲁棒似然比函数类似于Huber截断似然比检验的一种平滑变体，其中名义似然比函数先被顺时针旋转，随后被截断 [36]。这种特定行为——即通过此类变换获得极小极大鲁棒性——据我们所知，在本工作中首次被观察到并予以报告。

下一个例子将考虑表 I 中缩写为 d_2 的非对称标称分布对。图 3 展示了最不利分布及标称分布的密度，图 4 则展示了对应于 $\epsilon_0 = \epsilon_1 = 0.1$ 的稳健似然比函数。可以

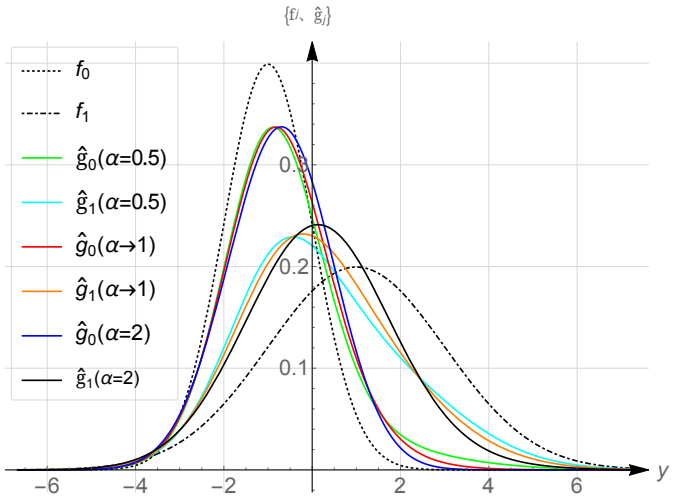

图3. 表 I 中缩写为 d_2 的标称分布以及 $\epsilon_0 = \epsilon_1 = 0.1$对应的最不利分布。

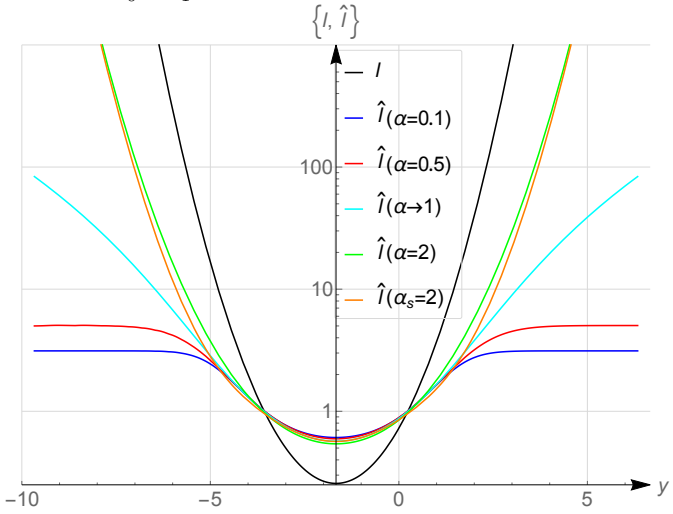

图4. 针对表 I 中缩写为 d_2 的标称分布以及 $\epsilon_0 = \epsilon_1 = 0.1$（包括对称情形 α_s）求得的鲁棒和名义似然比函数。 u 的最优值从 $\alpha = 0.1$ 到 $\alpha_s = 2$分别为0.95,0.67,0.56, 0.59, 0.61。

得出与前一个例子类似的结论。

在图2和图4中，名义似然比函数要么单调递增，要么先递减后递增。我们可以构造一个例子，使名义似然比函数反复递增和递减。这个情形既证实了相关非线性方程的可解性，也为下一节的凸性分析提供了一个例子。将表 I 中的标称分布简记为 d_3 。此外，令 $\epsilon_0 = \epsilon_1 = 0.05$，因为此时标称分布彼此更加接近。对于这一设置，图5和图6分别展示了最不利分布与标称分布的密度，以及相应的稳健似然比函数；这些结果对应于 KL

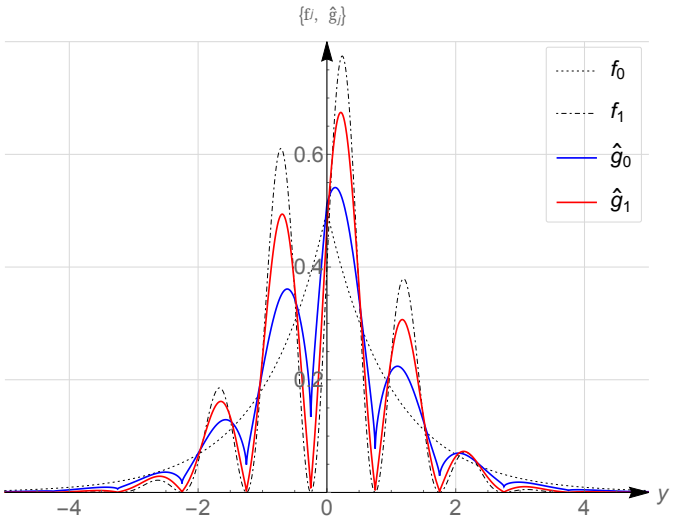

图 5。表 I 中以 d_3 缩写表示的标称分布，以及相对于其密度的对应最不利分布，其中 $\epsilon_0 = \epsilon_1 = 0.05$，且 $u = 0.46$。

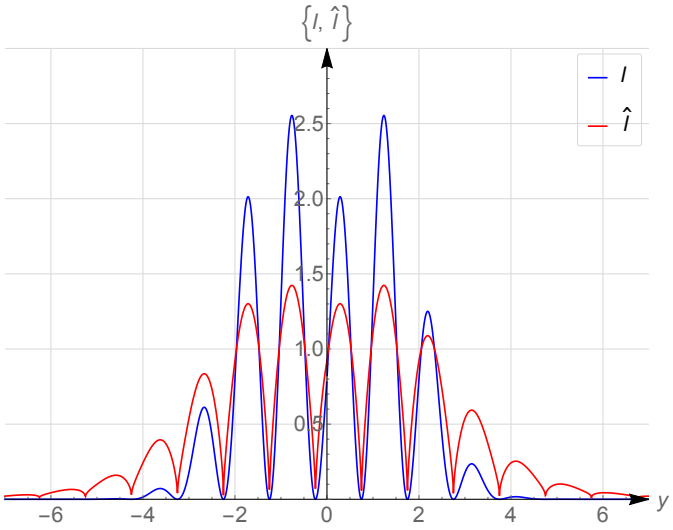

图 6。针对表 I 中以 d_3 缩写表示的标称分布以及 $\epsilon_0 = \epsilon_1 = 0.05$ 所得到的鲁棒和名义似然比函数。

散度邻域。与前面的例子类似，小于 1 的名义似然比函数会被放大，大于 1 的则会被减小。这意味着，与标称分布相比，最不利分布彼此更加接近（就 KL散度而言）。

B. D_u 的凸性与拉格朗日参数

如注 IV.1 所述，对于任意一对固定的已知分布，函数 D_u 关于 u 是凸的。然而，当分布本身

依赖于 u 时，例如在最不利分布的情形下，这种凸性未必普遍成立。本节将进一步通过实证研究这一现象，并评估其对本文所采用的极小极大优化策略的影响。

此处的主要目标有两个：首先，考察能否在优化过程中利用 D_u 的凸性（或非凸性）进行解析或数值分析；其次，为确保结果可复现且透明，展示并验证与先前计算得到的最不利分布相对应的最优拉格朗日参数。

为此，我们按如下步骤进行：首先，在 KL 散度邻域下，针对表 I 中以 d_1、d_2 和 d_3 表示的三组不同标称分布，求解第 III 节定义的非线性方程组；此外，在 $\alpha = 0.1$ 条件下，针对 d_2，在 α 散度下求解该方程组，以获得相应的最不利分布。随后，对每个 u 值，使用所得最不利分布计算 D_u，并将其绘制在图 7 中。这些结果表明，当最不利分布随 u 变化时，D_u 不一定关于 u 凸，从而支持注 IV.1 中的观察。

此外，对于每个 u，在 KL散度 情形下，当 $\epsilon_0 = \epsilon_1 = 0.1$、标称分布$d_1$ 和 d_2 时，KKT 乘子的最优值——λ_0、λ_1、μ_0 和 μ_1——如图 8 所示。这些图确认了可以根据拉格朗日参数一致地重构最不利分布及其对应的似然比函数，突出了数值过程的可复现性。

除了验证之外，该分析还为所采用的极小极大策略的顺序结构提供了实际依据——即先执行最大化步骤（以获得每个 u对应的最不利分布），随后对 u 进行最小化。当最不利分布依赖于 u 时，D_u 所表现出的非凸性使得在确定最不利密度之前先进行最小化、从而反转这一顺序在分析上并不可行。

C. 渐近极小极大鲁棒 *NP* 检验

本节给出了两种检验的数值结果：贝叶斯（a）检验被证明是渐近极小极大稳健的，而 Dabak 的（a^*）检验则被证明在 Neyman–Pearson (NP) 框架下不具备这种鲁棒性。此外，还纳入了从标称检验的分布中抽取的数据样本——为保持记号一致，称为（n）检验——以展示分布不匹配时的性能。这些结果有助于支持并说明前文各节开展的理论鲁棒性分析。

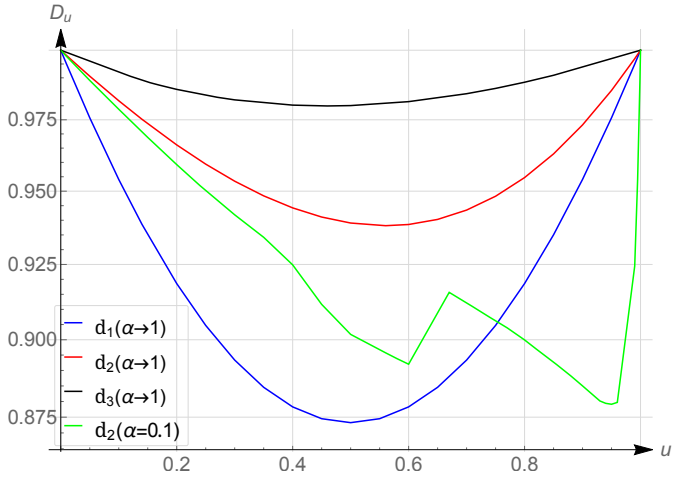

图 7. u-亲和度与 对于各种分布对以及不确定类所得最不利分布而言，作为 u 的函数 s.

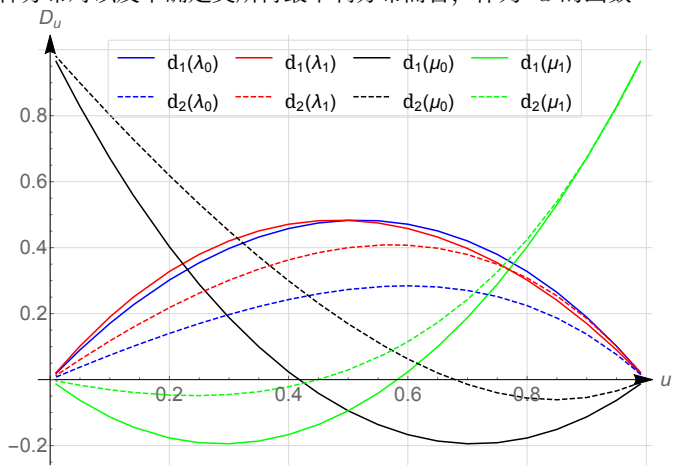

图 8. 两对标称分布及 KL 散度邻域（ $\epsilon_0 = \epsilon_1 = 0.1$）对应的 KKT参数。

为作说明，这里采用表 I 第 1 行所指定的标称分布，不确定性水平为 $\epsilon_0 = \epsilon_1 = 0.01$。图9和图10分别展示了速率函数 I_0 和 I_1随决策阈值 t变化的情况。对于（a^*）检验，只有阈值 $t = t_0 \approx -0.61$ 和$t = t_1 \approx 1.69$ 与之相关，因为 NP 第一类和第二类检验均使用这些阈值构造（参见第四节-A2）。

图 9 显示，在 $t = t_1$处， $I_0|_{a^*}^{a^*}$ 并非速率函数中的最小值。类似地，图 10 显示，在 $t = t_0$ 处， $I_1|_{a^*}^{a^*}$ 也并非最小值。数据由与（a）检验对应的最不利分布生成时，得到的衰减率最低。这些结果证实，Dabak检验在 NP表述下并非渐近极小极大稳健。

由于所选 该设置允许评估其他阈值，

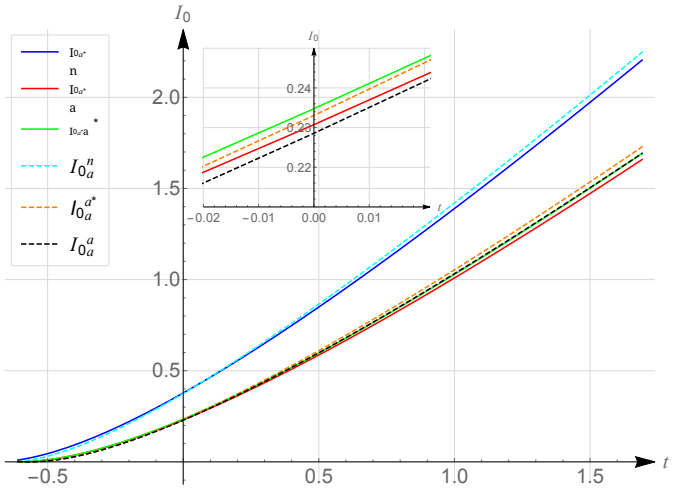

图9. 渐近极小极大鲁棒检验（(a)-检验）和Dabak检验（(a^*)-检验）的虚警概率渐近衰减率 I_0 。

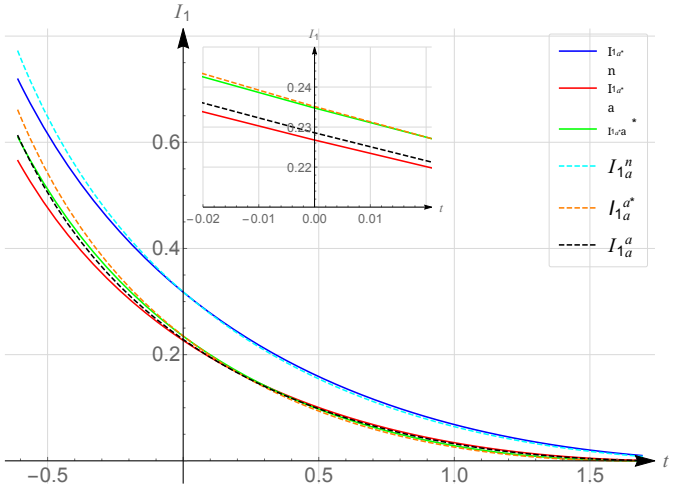

图10. 渐近极小极大鲁棒检验（(a)-检验）和Dabak检验（(a^*)-检验）的漏检概率渐近衰减率 I_1 。

同时还考察了在最不利分布不匹配情况下的(a)检验在 $t = 0$处的表现。在图9和图10的放大区域中，黑色虚线对应检验和数据均由(a)检验的最不利分布生成的情形。这些曲线始终呈现最小的错误指数。相比之下，其余虚线对应各种不匹配情形，产生了更大的衰减率。这表明，(a)检验的性能不会因不匹配而下降，这与其极小极大稳健性相符。

VI. 结论

本文提出了一个统一框架，用于在分布不确定性下构造渐近极小极大鲁棒检验。核心优化问题被表述为一个涉及 u-亲和度的鞍点问题：在预先定义的不确定性类中的分布对（G_0，G_1）上最大化 u-亲和度，并在 $u \in (0, 1)$ 上最小化。该表述使得能够在单一理论框架下推导出适用于贝叶斯和 Neyman–Pearson 表述的鲁棒检验。

本文考察了若干基于 f-散度的不确定性类，包括由 KL散度、 α-散度及其对称形式定义的邻域。针对这些不确定性类，利用 Karush-Kuhn-Tucker条件以闭式参数形式推导出最不利密度，从而得到便于处理的表达式，其中有些涉及 Lambert-W 函数等非线性函数。结果表明，鲁棒似然比函数是其名义对应函数的非线性变换——具体而言，是旋转和截断后的版本，尤其是在 α-散度邻域下。

数值模拟支持了理论发现，验证了所提出的贝叶斯（a）检验的鲁棒性，并证实 Dabak 的（a^*）检验在奈曼–皮尔逊设定下并非渐近极小极大稳健。除了验证稳健性保证外，仿真还为稳健似然比函数和拉格朗日参数的行为提供了实用见解。所有结果均与理论预测一致，并为该检验的极小极大最优性提供了可复现的证据。

附录 A

定理 II.2 的证明

考虑极小极大检验 (4) 的错误概率 (2)，为简便起见，仅用 n 和 t 表示为

$$P_E(n,t) = \pi_0 P_F(n,t) + (1-\pi_0) P_M(n,t).$$

由 (7) 和 (8) 可得

$$P_E(n,t) \approx C_F(n)\pi_0 \exp\left(-nI_0(t)\right) + C_M(n)(1-\pi_0)\exp\left(-nI_1(t)\right),$$

其中 C_F 和 C_M 满足

$$\lim_{n\to\infty} \frac{1}{n} \log C_F(n) = \lim_{n\to\infty} \frac{1}{n} \log C_M(n) = 0. \tag{37}$$

因此，错误概率的指数衰减率由 I_0 和 I_1决定。对于任意 t，若 $I_1(t) > I_0(t)$，则有

$$P_E(n,t) = \pi_0 C_F(n) \exp(-nI_0(t)) \text{ as } n \to \infty, \tag{38}$$

因为

$$\begin{aligned}\frac{P_E(n,t)}{\pi_0 C_F(n)\exp(-nI_0(t))} &= \frac{\pi_0 C_F(n)\exp(-nI_0(t)) + (1-\pi_0)C_M(n)\exp(-nI_1(t))}{\pi_0 C_F(n)\exp(-nI_0(t))} \\ &= 1 + C(n)\exp(-n(I_1(t) - I_0(t))) \to 1,\end{aligned} \tag{39}$$

其中

$$C(n) = \frac{(1-\pi_0)C_M(n)}{\pi_0 C_F(n)}. \tag{40}$$

这个论证成立，因为 $I_1(t) - I_0(t) > 0$ 和 $C(n)$ 是次指数的，这是因为由式 (37) 可得

$$\frac{1}{n}(\log((1-\pi_0)C_M(n)) - \log(\pi_0 C_F(n))) = 0 \text{ as } n \to \infty. \tag{41}$$

类似地，对于情形 $I_1(t) < I_0(t)$，我们有

$$P_E(n,t) = (1-\pi_0)C_M(n)\exp(-nI_1(t)) \text{ as } n \to \infty. \tag{42}$$

因此，由于 $n \to \infty$，我们有

$$P_E(n,t) = \begin{cases} \pi_0 C_F(n)\exp(-nI_0(t)), & I_1(t) > I_0(t) \\ (1-\pi_0)C_M(n)\exp(-nI_1(t)), & I_0(t) > I_1(t) \end{cases}, \tag{43}$$

这可以改写为

$$P_E(n,t) = \pi_0^a (1-\pi_0)^{1-a} C_F(n)^a C_M(n)^{1-a} \exp(-n\min\{I_0(t), I_1(t)\}) \tag{44}$$

其中 $a = \mathbf{1}_{\{I_1 > I_0\}}$。因此，当 $n \to \infty$

$$\begin{aligned}\min_t P_E(n,t) &\equiv \min_t \exp(-n\min\{I_0(t), I_1(t)\}) \equiv \min_t -n\min\{I_0(t), I_1(t)\} \\ &\equiv \max_t \min\{I_0(t), I_1(t)\} \equiv \min_t \max\{I_0(t), I_1(t)\},\end{aligned} \tag{45}$$

由于 $\pi_0^a(1-\pi_0)^{1-a}C_F(n)^a C_M(n)^{1-a}$为正且与 t 无关。根据 [32, 备注 5.2.2.]，I_0 和 I_1 分别是 u 的增函数和减函数。令 $h_j: u \mapsto t$ 为 (9) 中最大化 u 与 t 之间的映射。容易看出 h_j 是递增的，因为它

是凸函数 $\log\ M_{X_1}^j(u)$ [22, p. 77] 的导数。因此，$I_0(t) = I_0(h_0(u))$ 和 $I_1(t) = I_1(h_1(u))$ 分别也是增函数和减函数，因为

$$\frac{dI_0(h_0(u))}{du} = I_0^{'}(h_0(u))h_0^{'}(u) \geq 0,$$
$$\frac{dI_1(h_1(u))}{du} = I_1^{'}(h_1(u))h_1^{'}(u) \leq 0.$$

由于 I_0 和 I_1 同样分别是 t的增函数和减函数，并且由于对 G_j:=ˆG_j有 $M_{X_1}^1(u) = M_{X_1}^0(u+1)$，连同(9)可推出 $I_1(t) = I_0(t) - t$，因此 $I_0(0) = I_1(0)$成立，再结合 $\{t : I_1 > I_0\} \equiv \{t:\ t < 0\}$ 以及 $\{t :\ I_1 < I_0\} \equiv \{t:\ t > 0\}$，可以写出

$$I_m(t) = \min\{I_0(t), I_1(t)\} = \begin{cases} I_0(t), & t < 0 \\ I_1(t), & t > 0 \\ I_0(0) = I_1(0), & t = 0 \end{cases}. \tag{46}$$

因此，我们有

$$\arg\sup_t I_m(t) = 0. \tag{47}$$

注意，我们 ˆ需要定理 II.1中的 $G_j := G_j$。否则，(46)和(47)不一定成立。

APPENDIX B
EXISTENCE OF A SADDLE VALUE FOR (15)

定理 B.1（Sion极小极大定理的应用 [29]）。 当下列条件成立时，(15) 存在解：

- 目标函数 D_u 对所有 $u \in [0,1]$在$\mathscr{G}_0 \times \mathscr{G}_1$ 上为实值、上半连续且拟凹。*5*
- 目标函数 D_u 在 $[0,1]$ 上是下半连续且拟凸的，适用于所有（G_0, G_1） $\in \mathscr{G}_0 \times \mathscr{G}_1$。
- $[0,1]$ 是线性拓扑空间中的紧凸子集。
- $\mathscr{G}_0 \times \mathscr{G}_1$ 是线性拓扑空间中的凸子集。

证明：目标函数是实值函数，关于 u 和 (g_0, g_1) 连续，在所有 $u \in [0,1]$ 下关于 $\mathscr{G}_0 \times \mathscr{G}_1$ 联合凹，并在所有 (G_0, G_1) $\in \mathscr{G}_0 \times \mathscr{G}_1$ 下关于 $[0,1]$ 凸，参见 D_u 的第 2 和第 3 个性质。集合 $[0,1]$ 显然是凸集，且是闭、有界的，因此根据海涅-博雷尔定理 [37, 定理 2.41]，它在标准拓扑下是紧集。最后，$\mathscr{G}_0$ 和 $\mathscr{G}_1$ 是凸集，因为 D_f 是凸距离。因此，$\mathscr{G}_0 \times \mathscr{G}_1$ 也是凸集。

■

参考文献

[1] S. M. Kay，《统计信号处理基础》第 2 卷：检测理论。Prentice Hall PTR，1998 年 1 月。[2] B. C. Levy，《信号检测与参数估计原理》，第 1 版。施普林格，2008 年。[3] P. J. Huber，“概率比检验的稳健版本”，《数理统计年鉴》，第 36 卷，第 6 期，第 1753–1758 页，1965 年。[4] P. J. Huber 和 V. Strassen，“容量的极小极大检验与奈曼–皮尔逊引理”，《统计年鉴》，第 1 卷，第 2 期，第 251–263 页，1973 年。[5] B. C. Levy，“具有相对熵容差的稳健假设检验”，IEEE信息论汇刊，第 55 卷，第 1 期，第 413–421 页，2009 年。[6] G. G¨ul 和 A. M. Zoubir，“基于 α -散度的稳健假设检验”，《IEEE信号处理汇刊》，第 64 卷，第 18 期，第 4737–4750 页，2016 年 9 月。[7] G. G¨ul 和 A. M. Zoubir，“ 极小极大稳健假设检验 ”，IEEE 信息论汇刊，第 63 卷，第 9 期，第 5572–5587 页，2017 年。[8] A. G. Dabak 和 D. H. Johnson，“基于几何的稳健检测”，载于信息科学与系统会议论文集，约翰斯·霍普金斯大学，美国马里兰州巴尔的摩，1994 年 5 月，第 73–77 页。[9] H. Rahimian 和 S. Mehrotra，“分布鲁棒优化：综述”，Optimization Online，2019 年。[10] R. Gao、Y. Xie、L. Xie 和 H. Xu，“使用 Wasserstein不确定集进行稳健假设检验”，载于神经信息处理系统进展（NeurIPS），2018 年。[11] J. Wang、R. Gao 和 Y. Xie，“使用 Sinkhorn不确定集进行非凸稳健假设检验”，载于 2024 年 IEEE国际信息论研讨会（ISIT），2024 年，第 843–848 页。[12] Z. Sun 和 S. Zou，“基于核不确定集的稳健假设检验”，载于 2022 年 IEEE国际信息论研讨会（ISIT），2022 年，第 3309–3314 页。[13] A. Schrab 和 I. Kim，“数据损坏下的稳健核假设检验”，载于第 28 届国际人工智能与统计会议论文集，机器学习研究论文集，Y. Li、S. Mandt、S. Agrawal 和 E. Khan 编，第 258 卷，2025 年 5 月 3–5 日，第 2629–2637 页。[14] A. Magesh、Z. Sun、V. V. Veeravalli 和 S. Zou，“基于矩约束不确定集的稳健多重假设检验”，IEEE国际信息论研讨会（ISIT），2024 年，arXiv:2404.08589。[15] J. Wang 和 Y. Xie，“使用 Sinkhorn不确定集进行稳健假设检验的数据驱动方法”，载于 2022 年 IEEE国际信息论研讨会（ISIT），2022 年，第 3315–3320 页。[16] G. G¨ul，“并行传感器网络的极小极大稳健分布式假设检验”，IEEE信息论汇刊，第 67 卷，第 1 期，第 538–548 页，2021 年。[17] X. Cao、Y. Liu、Y. Li 和 J. Zhang，“对抗稳健序贯假设检验”，《统计计算与模拟杂志》，第 92 卷，第 14 期，第 2924–2942 页，2022 年。[18] M. Puranik、U. Madhow 和 R. Pedarsani，“用于对抗稳健假设检验的广义似然比检验”，arXiv:2105.00182，2021 年。[19] G. G¨ul 和 A. M. Zoubir，“基于平方海林格距离的稳健假设检验”，载于 2014 年第 22 届欧洲信号处理会议（EUSIPCO），2014 年，第 1083–1087 页。[20] A. T. Suresh 等，“海林格距离下的稳健假设检验与分布估计”，载于机器学习研究论文集，2021 年。[21] A. G. Dabak，“检测理论几何”，博士学位论文，莱斯大学，1993 年 5 月。

[22] B.C. Levy，《信号检测与参数估计原理》，第1版。Springer出版公司，2008年。[23] C. Pandit 和 S. Meyn，“最坏情形大偏差渐近及其在排队论和信息论中的应用”，《随机过程及其应用》，第116卷，第5期，第724–756页，2006年。[24] H. Chernoff，“基于观测值之和的假设检验的渐近效率度量”，《数理统计年鉴》，第23卷，第409–507页，1952年。[25] H. Cram´er，“关于概率论中的一个新极限定理”，Actualités Scientifiques et Industrielles，第736号，1938年。[26] T. van Erven 和 P. Harremos，“Rényi散度与Kullback-Leibler散度”，《IEEE信息论汇刊》，第60卷，第7期，第3797–3820页，2014年。[27] H. O，“关于一个平均值定理”，《哥廷根皇家科学学会及乔治-奥古斯特大学学报》，1889年，第338–47页，1889年。[28] A. Cichocki 和 S. ichi Amari，“α-、β-和γ-散度族：灵活且稳健的相似性度量”，《Entropy》，第12卷，第6期，第1532–1568页，2010年。[29] M. Sion，“关于一般极大极小定理”，《太平洋数学杂志》，第8卷，第1期，第171–176页，1958年。[30] F. Liese 和 I. Vajda，“统计学与信息论中的散度和信息”，《IEEE信息论汇刊》，第52卷，第10期，第4394–4412页，2006年。[31] D. P. Bertsekas、A. Nedi´c 和 A. Ozdaglar，《凸分析与优化》，Athena Scientific优化与计算丛书。Athena Scientific，2003年。[32] G. G¨ul，《鲁棒与分布式假设检验》，电气工程讲义。施普林格，2017年，第414卷。[33] F. Liese 和 I. Vajda，《凸统计距离》。Teubner，莱比锡，1987年，第130卷。[34] A. Guntuboyina、S. Saha 和 G. Schiebinger，“关于 f -散度的尖锐不等式”，《IEEE信息论汇刊》，第60卷，第1期，第104–121页，2014年1月。[35] D. Ralph，“通过路径搜索实现非光滑方程阻尼牛顿法的全局收敛”，《数学运筹研究》，第19卷，第2期，第352–389页，1994年。[36] P. J. Huber，“概率比检验的一个鲁棒版本”，《数理统计年鉴》，第36卷，第1753–1758页，1965年。[37] W. Rudin，《数学分析原理》，纯粹与应用数学国际丛书。巴黎：McGraw-Hill，1976年。