arXiv:2603.02492v1 [cs.IT] 3 Mar 2026

分布类的 e变量与随机性检验

Georgii Potapov[1] 和 Yuri Kalnishkan[1]

[1]伦敦大学皇家霍洛威学院

2026年2月

摘要

E变量是一种相对较新的统计假设检验方法，近年来得到了长足发展。

本文介绍 e-变量可近似性的方法，并运用该方法发展出一种通用近似技术，使我们能够为重要且广泛应用的分布类构造 e-变量。

E-变量最初基于 Levin 在算法信息论中提出的（平均有界）随机性检验概念。我们证明，构造 e-变量的方法可用于针对某一分布类，显式构造一个随机性检验。

1 引言

1.1 历史

统计假设检验问题，即判定给定数据能否由某个特定假设描述的问题，在科学研究和实际应用中都自然地具有重要意义。

传统上，假设检验使用 p-值。直观而言， p-值是在零假设成立的条件下，给定数据上的某个选定统计量至少与其观测值同样极端的概率。这定义了一个导出统计量，我们称之为 p-变量。尽管 p-值被广泛使用，却缺少某些“自然”的性质（例如，对于数据中随机选取的一个初始片段， p-变量在该片段上的值通常本身并不是 p-变量），因此很容易被误用。

E-变量是统计检验，可视为用于统计假设检验的 p-变量的一种灵活且通用的替代方案。形式上，e-变量是一种非负统计检验，其期望值至多为 1（[13]）。

使用这类变量可以灵活地进行实验设计，针对给定的备择假设检验原假设，并且避免常用 p-值的许多缺点。 E-变量易于组合，这简化了荟萃分析；在相互独立的数据批次上对 e-变量连乘，所得过程是超鞅，因此在随机停止时刻仍是一个 e-变量（[15, 4]）。

在这项工作中，我们重新审视 e-变量的概念起源——随机性亏损。Levin（[8], [7]）提出的随机性亏损是算法信息论中的一个概念，可粗略地定义为一种最优随机性检验，其指数是一个 e-变量。通俗地说，随机性亏损是一个函数，用来统计其参数——一个随机对象——所具有的规律数量，并衡量一次实验结果有多么不可信（[10]）。

论文 [5] 和 [14]从算法随机性的角度，对随机的"伯努利比特序列"给出了不同但相关的定义。这些定义之间的差异对应于独立同分布比特序列的分布与可交换分布之间的差异，并可通过序列中 1 的个数的随机性亏损来量化。

论文 [14] 和 [12], 中的一项结果在我们的论文中被称为定理 2。该结果指出，就随机性亏损而言，整数 k 在 0 到 n 之间取值并由某种二项分布生成，与 k 从整数区间 $[s_j\colon s_{j+1}-1] \ni k$上的均匀分布中生成，二者大致同样可信；其中， $\{s_1 < s_2 \ldots < s_m\}$ 是一个仅依赖于 n的集合。

相对于某一分布族的随机性检验在 [1], 中得到了更为普遍的研究，其中通过针对康托空间和有效闭测度类的一般拓扑论证，证明了这类检验的存在性。

1.2 我们的贡献

我们的研究目标之一，是将 [14] 和 [1] 的结果加以推广，使其更适用于实际的机器学习。

关于算法随机性，我们的主要结果是定理5和定理6；这两个定理表明，对于若干重要的分布族，例如区间上的均匀分布、泊松分布、正态（高斯）分布和柯西分布，针对某一分布族的随机性检验具有下半可计算性。此外，我们的结果还明确地将相对于某一分布族的随机性表示为通用均匀随机性检验。

与 [1], 中的类似表述相比，我们在 $\mathbb{R}^n$ 中开展研究，而不是在康托集上开展研究。我们所证明的结果适用于这样一些分布类：它们在概率分布空间中并不构成拓扑闭集（例如， $[0$ 上的均匀分布序列： $n]$应当收敛于一个在所有数上均为零的测度，而该测度并不是有限整数区间上的均匀测度）。

定理5和定理6可由定理3和定理4推出，后两者可视为前两者的非算法的纯统计对应项。我们提供了一种统一的方法来

将针对简单假设 $\{\mathbb{P}_\theta\}$ 的任意一组 e 变量转化为针对复合假设 $\{\mathbb{P}_\theta\}_{\theta\in\Theta}$ 的一个 e 变量。我们的构造既不基于反向信息投影，也不基于对 p 变量的校准；它可以将针对简单假设的任意 e 变量作为黑箱，用来生成针对相应复杂假设的有效 e 变量。

我们的统计结果和算法结果都建立在降低复杂度的参数估计之上，这也可以视为最小描述长度原理的另一种理论依据。

2 预备知识

2.1 e变量

P值（以及相应的统计量，我们称之为 p变量）是统计假设检验中常用的工具。

在以下定义中， $\mathcal{H}$ 是一个概率分布集合（来自**假**设）。

定义 2.1 （p变量）。 若对每个分布 $\mathbb{P}\in\mathcal{H}$ ，都有一个统计量（可测函数） P，则称其为***p*变量**。

$$\mathbb{P}(P(X)\le p)\le p$$

对所有 $p\in[0,1]$。

定义 2.2 （e-变量）。 若统计量 $E\colon\ \mathcal{X}\to[0,+\infty]$ 在假设 $\mathcal{H}$ 下满足以下条件，则称其为 e-**变量**：

$$\mathbb{E}E_{\mathbb{P}}(X)\le 1.$$

对每个概率分布 $\mathbb{P}\in\mathcal{H}$ 都成立。

尽管 p-变量在科学领域中应用更广泛，但如果 P 是一个 p-变量，且 E 是一个 e-变量，那么对于所有 $\kappa\in(0,1)$， $1/E$ 是一个 p-变量，而 $\kappa P^{\kappa-1}$ 是一个 e-变量。 p-变量与 e-变量之间的转换在 [15] 中有详细讨论。

对于具有概率密度 p_0 和 p_1 的简单零假设和简单备择假设，比值 $p_1(x)/p_0(x)$ 在零假设下是一个 e-变量。利用反向信息投影（[6, 4]），这一结论可推广到复合假设的情形。

e-变量理论是在随机性的统一检验理论的基础上发展而来的；后者将在下一小节中定义，适用于可测函数的一般情形。最终，这一理论发展成为一个独立的研究领域（[13, 9]）。

2.2 可计算函数与半可计算函数

我们采用可计算函数以及下（上）半可计算函数的标准定义，该定义遵循 [3] 和 [1]。这一方法与

该主题的其他研究（如 [11]）相一致，并且通常与 [16]中所述的第二型可计算性理论相关联。

定义 2.3。函数 $f\colon \mathbf{D} \to [-\infty,\infty]$ 若其下图 $\{(x,t)\colon t<f(x)\}$ 可表示为并集 $\bigcup_{(m,n)\in W} B_m \times (-\infty, r_n)$，其中 $W \subseteq \mathbb{N}\times\mathbb{N}$ 为可枚举集，且 $(r_i)_i$ 和 $(B_i)_i$ 分别是 $\mathbb{Q}$ 以及$\mathbf{D}$的一个可数拓扑基的预先指定枚举，则称其为下半可计算（l.s.c.）。

我们也称 f 为上半可计算（u.s.c.），如果 $-f$ 是下半可计算的；当 f 既下半可计算又上半可计算时，称其为可计算。

可计算性的一个更一般定义（定理5和定理6需要用到）如下。

定义 2.4。一般而言，如果对 X和 Y的某些预先指定的可数基枚举$\mathbf{B}^X = (B_1^X, B_2^X, \ldots)$和$\mathbf{B}^Y = (B_1^Y, B_2^Y, \ldots)$（分别）成立：

$$\bigcup_k f^{-1}\left(B_k^Y\right) \times B_k^Y = \bigcup_{(m,n)\in W} B_m^X \times B_n^Y$$

对于某个可枚举集 W。

备注。请注意，下半可计算性意味着下半连续性，上半可计算性意味着上半连续性，而可计算性意味着连续性。

本文考虑定义在 $\mathbb{R}\times\mathfrak{P}(\mathbb{R})\times\{0,1\}^*$ 或 $\mathbb{N}\times\mathfrak{P}(\mathbb{N})\times\{0,1\}^*$ 上的函数的（半）可计算性。稍微不那么形式化地说，我们可以认为，如果存在一台神谕图灵机，能够从下方逼近函数 f，并通过神谕访问其输入 $u=(x,\ \mathbb{P},\ s)$ 中的“无限”分量，那么函数 f 在这样的定义域上是下半可计算的。

对于实数或分布的神谕，在有理输入 $\varepsilon>0$ 上会返回该对象的一个 ε 近似。

概率测度 $\mathbb{P}$ 的一个 ε 近似可以是一个测度 $\mathbb{P}_\varepsilon = \sum_{i=1}^n p_i\delta_{x_i}$，它集中在有限个有理点 x_i 上，并为这些点赋予有理概率 p_i，且 $\mathbb{P}_\varepsilon$ 与 $\mathbb{P}$ 在莱维-普罗霍罗夫度量下的距离为 ε，也就是说，对于所有可测集 A，我们有

$$\mathbb{P}(A) < \mathbb{P}_\varepsilon(A^\varepsilon) + \varepsilon, \quad \mathbb{P}(A^\varepsilon) > \mathbb{P}_\varepsilon(A) - \varepsilon,$$

其中，A^ε 是 A的 ε邻域。

2.3 随机性检验

定义 2.5（随机性检验）。对于可测空间 $\mathcal{X}$，考虑函数 $f(\cdot|\cdot,\ \cdot)\colon \mathcal{X}\times\mathfrak{P}(\mathcal{X})\times\mathcal{Y}\to[0,\infty]$，其参数包括对象 $x\in\mathcal{X}$、定义在 $\mathcal{X}$ 上的概率分布 $\mathbb{P}$ 以及来自 $y\in\mathcal{Y}$ 的辅助对象。若 f 满足以下条件，我们称其为一致随机性检验（或 Levin 意义下的随机性检验）：

1. 是下半可计算函数（假设在 $\mathcal{X}$，$\mathfrak{P}(\mathcal{X})$ 和 $\mathcal{Y}$ 上具有适当的可计算结构）；
2. 对所有概率测度 $\mathbb{P}$ 和所有字符串 y，函数 $f(\cdot \mid \mathbb{P},\ y)$ 是关于简单假设 $\{\mathbb{P}\}$ 的 e-变量。

从实践角度看，下半可计算性是一项非常宽松的限制，使我们能够构造通用随机性检验（如下所述）。从哲学角度看，下半可计算性体现了这样一种思想：如果某个 e-变量以选定的显著性水平拒绝原假设，那么最终必定能够根据以足够高但有限的精度测得的数据确立这一点。

下一个定理—定义是一个众所周知的结果。

定理1（通用均匀检验的存在与定义（[8, 3, 11]））。对于具有形式 $\{0,1\}^*$、$\{0,1\}^\omega$、$\mathbb{R}^n$ 和 $\mathcal{Y}$ 的 $\mathcal{X}$，其中 $\mathcal{Y}$ 是上述集合之一或这些集合上的有界测度集合，存在一个一致随机性检验 $\mathbf{t}$，使得对于任意一致随机性检验 f，都存在一个仅取决于$\mathbf{t}$ 和 f 的常数 $C_f > 0$，使不等式

$$\mathbf{t}(x \mid \mathbb{P}, y) \geq C_f f(x \mid \mathbb{P}, y)$$

对任意 x、$\mathbb{P}$、y 均成立。

任何此类检验都称为通用均匀随机性检验，或简称通用检验。

从现在起，我们将使用符号 $=^+$、$\leq^+$, $\geq^+$, $=^\times$, $\leq^\times$, $\geq^\times$ 表示等式或不等式在加法常数或乘性常数的意义下成立。

对于任意通用检验 $\mathbf{t}_1$ 和 $\mathbf{t}_2$，都有 $\mathbf{t}_1 =^\times \mathbf{t}_2$,，因此在本文余下部分中，我们固定一个任意的通用检验，并将其记为 $\mathbf{t}$。

本文主要研究以下对象。

定义 2.6. 对于概率分布族 $\mathcal{P}$，我们将相对于 $\mathcal{P}$ 的随机性检验定义为下确界投影

$$\mathbf{t}^{\mathcal{P}}(x \mid y) = \inf_{\mathbb{P} \in \mathcal{P}} \mathbf{t}(x \mid \mathbb{P}, y).$$

本文旨在为下确界投影寻找显式表达式，从而直接证明，针对某个分布族的随机性检验是一个下半连续函数。

我们将对许多重要的分布族 $\mathcal{P}$证明这一点，具体列举如下。

我们的方法推广并扩展了 [12, 14], 的方法，其中证明了以下定理。

定理 2（[12, 14]）。对于任意自然数 n，令$\mathrm{Bern}_n = \{\mathrm{Bern}(n, p)\}_{p \in [0,1]}$表示定义在 $\{0,1\}^n$ 上的、具有独立同分布比特的分布族，并令$\mathrm{Bin}_n = \{\mathrm{Bin}(n, p)\}_{p \in [0,1]}$

表示定义在 $[0:\ n] = \{0, 1, \ldots, n\}$上的二项分布族。那么对于任意字符串 $x \in \{0,1\}^n$ 和任意数值 $k \in [0:\ n]$

$$\begin{aligned}\mathbf{t}^{\mathrm{Bern}_n}(x) &=^{\times} \mathbf{t}(x \mid \mathrm{Bern}(n, \hat{p}(|x|_1/n))),\\ \mathbf{t}^{\mathrm{Bin}_n}(k) &=^{\times} \mathbf{t}(k \mid \mathrm{Bern}(n, \hat{p}(k/n)))\\ &=^{\times} \mathbf{t}\left(k \mid \ \mathcal{U}\left(\hat{p}^{-1}(\hat{p}(k/n))\right)\right),\end{aligned}$$

其中$\hat{}p$ 是一个可计算函数，它将形如 m/n,$m \in [0:\ 1]$, 的有理数映射到"网" $\left\{\sin^2\left(\frac{\pi t}{2\lfloor\sqrt{n}\rfloor}\right)\right\}_{t\in[1:\lfloor\sqrt{n}\rfloor]}$中最近的元素。

我们将证明以下分布族的类似结果：

- $[0$上的均匀分布：$n] = \{0, 1, \ldots, n\}$, $\{\mathcal{U}([0:\ n])\}_{n\in\mathbb{N}}$,
- 泊松分布，$\{\mathrm{Pois}(\lambda)\}_{\lambda>0}$,
- 实区间上的均匀分布，$\{\mathcal{U}([0,\theta])\}_{\theta>0}$,
- 固定方差的正态分布 $\{\mathcal{N}(\mu, 1)\}_{\mu\in\mathbb{R}}$,
- 固定均值的正态分布 $\{\mathcal{N}(0,\ \sigma^2)\}_{\sigma^2>0}$,
- 固定尺度的柯西分布 $\{$柯西分布$(x_0,\ 1)\}_{x_0\in\mathbb{R}}$。

我们的方法与 [12, 14] 中的方法有很大不同。我们将证明我们的结果（在第 4 节中进行了精确定式），方法是说明所讨论的所有分布族都是我们所谓的 e-变量可逼近的；这本身意味着，对于每个分布的 e-变量，它们都允许为该分布族构造某种 e-变量。

2.4 组织结构

在第 3 节中，我们将阐述一些关于不包含任何可计算性要求的 e-变量的技术性结果：第 3 小节给出 e-变量可逼近性的定义，并列出关于这一概念的主要结果；第 3.2 小节描述在第 3 小节中构造的 e-变量的连续插值版本，我们将在第 4 节中利用这些版本使我们的函数成为下半连续函数。

在第 4 节中，我们将利用关于 e-变量的这些结果，为所述分布族构造下半连续的下确界投影近似。大部分证明留待附录给出。

在第5节中，我们将讨论定义的选择、结果的适用范围以及进一步研究的展望。

所有非标准符号均列于附录中。

3 E-变量可逼近性

3.1 关于 e-变量可逼近性的主要结果

这里我们介绍本文中的主要概念之一，即分布族的 e-变量可逼近性概念。

定义 3.1. 我们称分布族 $\mathcal{H}$ 为 e-变量可逼近的（从网 $\mathcal{S}$）（因子为 C），如果存在一个至多可数集$\mathcal{S}$（称为网）、一个分布集 $\{\mathbb{P}_s\}_{s\in\mathcal{S}} \subset \mathcal{H}$、一个函数$\hat{s}$：$\mathbb{R} \to \mathcal{S}$（称为估计量）以及一个常数 $C > 0$，使得对于任意一族非负统计检验 $\{e_s\}_{s\in\mathcal{S}}$，若对所有 s，检验 e_s 在简单假设 $\{\mathbb{P}_s\}$ 下是一个 e-变量检验，则检验 $e(x) = \frac{1}{C} e_{\hat{s}(x)}(x)$ 在假设 $\mathcal{H}$下是一个 e-变量检验。

这个定义乍看可能有些人为，因此我们在第5节提供一些论证和示例。

以下引理给出了易于验证的条件，以确立e-变量可逼近性。

引理1。设 $\mathcal{H} = \{\mathbb{P}_\theta\}_{\theta\in\Theta}$ 是定义在 $\mathbb{R}^n$ 上、具有密度函数$\{p_\theta\}_{\theta\in\Theta}$ 的分布族，由 $\Theta \subseteq \mathbb{R}$ 中的元素作为参数，并设 $\mathcal{S} \subseteq \Theta$ 是一个至多可数集，使得对于所有 $\theta \in \Theta$，后继$_{\mathcal{S}}(\theta) = \min\{s \in \mathcal{S} \mid \theta < s\}$ 和 前驱$_{\mathcal{S}}(\theta) = \max\{s \in \mathcal{S} \mid s < \theta\}$ 都有定义，但可能不包括那些大于 $\mathcal{S}$ 最大元素或小于其最小元素的情形。设 $d(\cdot\|\cdot)$：$\Theta \times \Theta \to [0, +\infty]$ 是一个定义良好的函数。设 $\hat{s}$：$\mathbb{R}^n \to \mathcal{S}$ 是一个统计量，且 g：$\mathbb{R}^n \to \Theta$ 是一个函数。现在考虑以下性质：

(p i) 对于每个 $\theta \in \Theta$、$s \in \mathcal{S}$ 和 $x \in \hat{s}^{-1}(s) \cap \{y \mid p_\theta(y) > 0\}$，我们有

$$\log \frac{p_\theta(x)}{p_s(x)} = d(g(x)\|s) - d(g(x)\|\theta);$$

(p ii) 存在一个常数 $c' > 0$，使得对于任意 $s \in \mathcal{S}$ 以及任意满足 $x \in \hat{s}^{-1}(s)$ 的 x，都有

$$d(g(x)\|s) < c'.$$

(p iii) 对所有 $s \in \mathcal{S}$ 以及所有 $x \in \hat{s}^{-1}(s)$

$$\mathrm{pred}_{\mathcal{S}}(s) \le \mathrm{pred}_{\mathcal{S}}(g(x)) \le \mathrm{succ}_{\mathcal{S}}(g(x)) \le \mathrm{succ}_{\mathcal{S}}(s).$$

(p iv) 存在一个常数 $\alpha > 0$，使得对于所有 θ_1、$\theta_2 \in \Theta$、$\theta_1 < \theta_)$，都有

$$d(\theta_1\|\theta_2) \ge (1+\alpha)\log(k-1), \quad d(\theta_2\|\theta_1) \ge (1+\alpha)\log(k-1),$$

对于 $k = |\mathcal{S} \cap (\theta_1, \theta_2)|$[1]。

[1]对于此陈述，我们将非正数的对数视为 $-\infty$

若数据$(\mathcal{H}, \mathcal{S}, \Theta, \hat{s}, g, d)$满足所有性质 *(p i)–(p iv)*，则 $\mathcal{H}$ 可由网 $\mathcal{S}$ 和估计量$\hat{s}$进行 e-变量逼近，且常数因子 C 可取为 $C = \exp(c')(7 + 2/\alpha)$。

引理2. 在引理*1*中，条件 *(p iv)* 由以下两个条件推出：

(p i') 若 $\theta_1 \le \theta_2 \le \theta_3$ 或 $\theta_1 \ge \theta_2 \ge \theta_3$，则

$$d(\theta_1 \| \theta_3) \ge d(\theta_1 \| \theta_2) + d(\theta_2 \| \theta_3);$$

(p ii') 存在一个常数 $c > 0$，使得对于任意相邻的 $s, s' \in \mathcal{S}$，都有

$$d(s \| s') > c, \quad d(s' \| s) > c.$$

若引理*1*中的数据 $(\mathcal{H}, \mathcal{S}, \Theta, \hat{s}, g, d)$ 满足所有性质 *(p i)–(p iii)*，并且满足 *(p i')* 和 *(p ii')*，则 $\mathcal{H}$ 可由网 $\mathcal{S}$ 和估计量 $\hat{s}$进行 e-变量逼近，且常数因子 C 可取为$C = \exp(c') \left(\underline{5 + \quad}^{2}{}_{\mathrm{e}^{c}-1}\right)$。

这两个引理均通过仔细估计右端各项得以证明，

$$\mathbb{E}_\theta e_{\hat{s}(X)}(X) = \sum_{s \in \mathcal{S}} \int_{\hat{s}^{-1}(s)} e_s(x) \frac{p_\theta(x)}{p_s(x)} p_s(x) \mathrm{d}x$$

从而为总和提供一致上界。

我们将看到，在许多情况下，检查两个引理之一的条件都很简单。

特别是，对于许多标准指数族，这一点很容易验证。

引理 3. 设 $\{\mathbb{P}_\eta\}_{\eta \in H}$ 是一个由 H（$\mathbb{R}$ 中的一个开区间，可能是无界区间）参数化的指数族，其规范参数为 η，支撑集为 $D \subset \mathbb{R}^n$，密度函数为 $p_\eta(x) = h(x) \exp(\eta T(x) - A(\eta))$。

1. 如果存在函数 $\hat{}\eta : D \to H$，对所有 $x \in D$ 都满足似然方程 $T(x) = \mathbb{E}_{\hat{\eta}(x)} T(X)$，那么引理*1*的条件 *(p i)* 对$d = \mathrm{kl}$ 和 $g = \hat{}\eta$成立。

2. 引理 *2* 的条件 *(p i')* 对 $d = \mathrm{kl}$成立。

定理 3. 以下分布族是 e-变量可逼近的。当 $\hat{s} =$取整$_{\mathcal{S}}$ 时，省略函数 $\hat{s}$。

- $\{$二项分布$(n, p)\}_{p \in (0,1)}$, 来自网 $\mathcal{S} = \left\{\sin^2\left(\frac{\pi t}{2\lfloor\sqrt{n}\rfloor}\right)\right\}_{0 < t < \sqrt{n}}$;
- $\{\mathcal{U}(\{0, 1, \ldots, n\})\}_{n \in \mathbb{N}}$, 来自网 $\mathcal{S} = \{2^n\}_{n \in \mathbb{N}}$, $\hat{s}(n) = 2^{\lceil \log^2 n \rceil}$;
- $\{$泊松分布$(\lambda)\}_{\lambda \in \mathbb{R}_{>0}}$，来自网 $\{t^2\}_{t \in \mathbb{N}}$。

以下连续分布族也可以通过 e-变量近似：

- $\{\mathcal{U}([0,\theta])\}_{\theta>0}$，来自网 $\mathcal{S}=\{2^n\}_{n\in\mathbb{Z}}$，$\hat{s}(x)=2^{\lceil \text{对数}_2\, x\rceil}$
- $\{\mathcal{N}(\mu,1)^{\otimes n}\}_{\mu\in\mathbb{R}}$，来自网 $\mathcal{S}=\underline{\underline{\frac{\alpha}{\sqrt{n}}\mathbb{Z}}}$ 对于 $\alpha>0$，$\hat{s}(x_{1:n})=\text{取整}_{\mathcal{S}}\left(\frac{1}{n}\underline{\sum}_k x_k\right)$，因子可能取决于 α，但不取决于 n；
- $\{\mathcal{N}(0,\sigma^2)^{\otimes n}\}_{\sigma\in\mathbb{R}_{>0}}$，来自网 $\mathcal{S}=\underline{\underline{\left\{\left(1+\frac{1}{\sqrt{n}}\right)^k\right\}_{k\in\mathbb{Z}}}}$，$\hat{s}(x)=\text{取整}_{\mathcal{S}}\left(\frac{\|x\|^2}{n}\right)$，因子不取决于 n；
- $\{\text{柯西分布}(x_0,1)\}_{x_0\in\mathbb{R}}$，来自网 $\mathcal{S}=\mathbb{Z}$。

对于任意 $\varepsilon\le 1/5$，并令 r^ε 为定义在 $\mathbb{R}\setminus(\mathbb{Z}+0.5)^\varepsilon$ 上的任意函数，按 取整$_{\mathbb{Z}}$，且对于每个半整数 ε 的每个邻域，定义为 n 或 $n+1$（可能因 n 而异），则 $\{\mathcal{N}(\mu,1)\}_{\mu\in\mathbb{R}}$ 和 $\{\text{柯西分布}(x_0,1)\}_{x_0\in\mathbb{R}}$ 都是可由 $\mathcal{S}=\mathbb{Z}$ 和 $\hat{s}=r^\varepsilon$ 进行 e-变量逼近的。

推论。对于所有 n 和 $x_{1:n}\in\mathbb{R}^n$，令 $r_n(x_{1:n})=\text{取整}_{\underline{\underline{\frac{1}{\sqrt{n}}\mathbb{Z}}}}\left(\frac{1}{n}\underline{\sum}_k x_k\right)$。对于某个常数 $C>0$，以下结论成立。

对于任意 $\{e_\mu\}_{\mu\in\mathbb{R}}$，一族可测函数，其中每个 e_μ 都是关于 $\{\mathcal{N}(\mu,1)\}$ 的 e-变量，函数

$$e(x_{1:n})=\frac{1}{C}\prod_{k=1}^{n}e_{r_n(x_{1:n})}(x_k)$$

是一个 e-变量，适用于 $\{\mathcal{N}(\mu,1)^{\otimes n}\}_{\mu\in\mathbb{R}}$。

该推论可由定理 3 以及以下事实立即得出：e-变量的乘积是适用于乘积分布的 e- 变量。

3.2 插值逼近

取整$_{\mathcal{S}}$ 的不连续性使得该函数对于实数上的分布而言不可计算，这使我们无法将结果直接应用于随机性亏损。

对于取整$_{\mathbb{R}}$，我们希望“平滑” n 与 $n+1$之间的过渡；一般而言，对于取整$_{\mathcal{S}}$ 也是如此。

定义 3.2。我们称族 $\{\mathbb{P}_\theta\}_{\theta\in\Theta}$ 可从网 $\mathcal{S}\subset\Theta$ 进行连续 e-变量逼近，如果存在一个函数集 $\{\hat{s}(\cdot,s)\}_{s\in\mathcal{S}}$，构成一个单位分解，以及一个常数 $C>0$，使得定义如下的函数 e

$$e(x)=\frac{1}{C}\sum_{s\in\mathcal{S}}e_s(x)\hat{s}(x,s)$$

是一个 e变量，当且仅当所有 e_s 都是相应 $\{\mathbb{P}_s\}$的 e变量。

定理 4. 对于 $\mathcal{P} = \{\mathbb{P}_\theta\} = \{\mathcal{N}(\theta, 1)\}_{\theta\in\mathbb{R}}$ 或 $\{$柯西分布$(\theta, 1)\}_{\theta\in\mathbb{R}}$，对于任意满足 e_n 是 $\{\mathbb{P}_n\}$的 e变量的$\{e_n\}_{n\in\mathbb{Z}}$，存在一个常数 $C > 0$，使得对于任意 $\varepsilon \le 1/5$，函数

$$e^\varepsilon(x) = \frac{1}{C}\sum_{n=-\infty}^{\infty} e_n(x)\mathbb{I}^{2\varepsilon}_{\left[n-\frac{1}{2}+\varepsilon, n+\frac{1}{2}-\varepsilon\right]}(x)$$

是关于 $\mathcal{P}$ 的一个 e变量。

注意，对于任意 x，求和中至多只有两项非零。该证明展示了 e变量可逼近性质的灵活性 rty.

证明。注意，连续的 $[0,1]$ 值函数与一个 e 变量相乘后，仍然是一个 e变量。对于给定的族 $\{e_n\}_{n\in\mathbb{Z}}$，按如下方式定义两个新的族：$\{e_n^{\text{even}}\}_{n\in\mathbb{Z}}$ 和 $\{e_n^{\text{odd}}\}_{n\in\mathbb{Z}}$。令 e_n^{even} 在 n 为偶数时等于$e_n \cdot \mathbb{I}^{2\varepsilon}_{[n-0.5-\varepsilon, n+0.5+\varepsilon]}$，在 n 为奇数时等于 0；e_n^{odd} 的定义类似。

令 $\hat{s}_{\text{even}}$ 为舍入函数 $\text{round}_{\mathbb{Z}}$，并在 ε邻域内将半整数舍入到最近的偶数，从而重新定义该函数。以类似方式定义 $\hat{s}_{\text{odd}}$。

那么存在某个 $C_0 > 0$，使得 $e^{\text{even}}(x) = \frac{1}{C_0} e^{\text{even}}_{\hat{s}_{\text{even}}(x)}(x)$ 和$e^{\text{odd}}(x) = \frac{1}{C_0} e^{\text{odd}}_{\hat{s}_{\text{even}}(x)}(x)$ 都是相对于 $\mathcal{P}$ 的 e 变量。两个 e 变量的算术平均数显然是一个 e变量，因此对于 $C = 2C_0$，我们得到

$$e^\varepsilon(x) = \frac{e^{\text{even}}(x) + e^{\text{odd}}(x)}{2} = \frac{1}{C}\sum_{n=-\infty}^{\infty} e_n(x)\mathbb{I}^{2\varepsilon}_{\left[n-\frac{1}{2}+\varepsilon, n+\frac{1}{2}-\varepsilon\right]}(x)$$

也是相对于 $\mathcal{P}$ 的 e-变量。 □

4 随机性亏损下半可计算逼近的下确界-投影

如果函数 $(\theta \mapsto \mathbb{P}_\theta)$ 是可计算的，我们就称分布族 $\{\mathbb{P}_\theta\}_{\theta\in\Theta}$ 是可计算的。

引理 4。 对于离散集 $\mathcal{X}$ 上的可计算分布族 $\mathcal{P} = \{\mathbb{P}_\theta\}_{\theta\in\Theta}$，若其可通过带有可计算估计量$\hat{s}$和因子 C 的网 $\mathcal{S}$ 进行 e-变量逼近，则存在一个随机性检验 f：$\mathcal{X} \times \mathfrak{P}(\mathcal{X}) \times \mathcal{Y} \to [0, \infty]$，使得

$$f(x \mid \mathbb{P}_\theta, y) =^\times \mathbf{t}(x \mid \mathbb{P}_{\hat{s}(x)}, y)$$

对于所有 $\theta \in \Theta$。

证明思路。这种证明技巧在该领域很常见。由于$\mathbf{t}$是下半可计算的，且 $\mathbb{P}$可计算，因此我们可以枚举一列具有有限描述的有界连续分段线性函数 $0 = g_0 \le g_1 \le \dots$，满足 $g_n(x, y) \uparrow \mathbf{t}(x \mid \mathbb{P}_{\hat{s}(x)}, y)$。对于这些函数，我们可以估计

积分 $\int g_n(x,y)\mathbb{P}(\mathrm{d}x)$。不失一般性，我们设 $C \geq 1$ 为有理数。于是可以定义

$$f(x \mid \mathbb{P}, y) = \frac{1}{2} \sup_n \left\{ g_n(x,y) \mid \int g_n(x,y)\mathbb{P}(\mathrm{d}x) < 2C \right\}.$$

由 e-变量可逼近性可知，积分小于 $2C$ 的条件对 $\mathbb{P} \in \mathcal{P}$ 总是成立，因此在这种情况下有 $f =^{\times} \mathbf{t}$ 。 □

定理 5. 对于具有可计算估计量 $\hat{s}$ 且可由 e-变量逼近的可计算族 $\mathcal{P} = \{\mathbb{P}_\theta\}_{\theta \in \Theta}$ ，有

$$\mathbf{t}^{\mathcal{P}}(x \mid y) =^{\times} \mathbf{t}(x \mid \mathbb{P}_{\hat{s}(x)}, y).$$

证明。令 f 为引理 4 中的一致检验。

对于任意 θ

$$f(x \mid \mathbb{P}_\theta, y) =^{\times} \mathbf{t}(x \mid \mathbb{P}_{\hat{s}(x)}, y),$$

其中右侧项不依赖于 θ。

于是我们得到，对于任意 θ

$$f(x \mid \mathbb{P}_\theta, y) =^{\times} \mathbf{t}(x \mid \mathbb{P}_{\hat{s}(x)}, y) \geq \mathbf{t}^{\mathcal{P}}(x \mid y).$$

根据随机性亏损的最优性，我们还可以得到

$$\mathbf{t}(x \mid \mathbb{P}_\theta, y) \geq^{\times} f(x \mid \mathbb{P}_\theta, y) =^{\times} \mathbf{t}(x \mid \mathbb{P}_{\hat{s}(x)}, y).$$

由于 $\geq^{\times}, =^{\times}$ 中的常数不依赖于 x, θ, y，我们可以对所有 θ 取下确界，得到

$$\mathbf{t}^{\mathcal{P}}(x \mid y) \geq^{\times} \mathbf{t}(x \mid \mathbb{P}_{\hat{s}(x)}, y),$$

由此证明得证。 □

引理5. 对于一个可计算分布族 $\mathcal{P} = \{\mathbb{P}_\theta\}_{\theta \in \Theta}$ ，若其可从一个一致可计算网 $\mathcal{S} = \{s_i\}_{i \in \mathcal{I}}$ 连续地按 e 变量逼近，且$\mathcal{I} \subset \mathbb{N}$ 的$\hat{s}$ 关于两个自变量均为下半连续，并具有因子 C，那么存在一个随机性检验 $f\colon \mathcal{X} \times \mathfrak{P}(\mathcal{X}) \times \mathcal{Y} \to [0, \infty]$，使得

$$f(x \mid \mathbb{P}_\theta, y) =^{\times} \sum_{s \in \mathcal{S}} \mathbf{t}(x \mid \mathbb{P}_s, y)\hat{s}(x, s)$$

对所有 $\theta \in \Theta$ 均成立。

引理6。令 $\mathcal{P} = \{\mathbb{P}_\theta\}_{\theta \in \mathbb{R}}$ 为 {$\mathcal{N}$（θ，1）}$_{\theta \in \mathbb{R}}$ 或 {柯西分布（θ，1）}$_{\theta \in \mathbb{R}}$。令n 为一个整数，且 $x \in [n, n+1]$。那么

$$\mathbf{t}(x \mid \mathbb{P}_n, y) =^{\times} \mathbf{t}(x \mid \mathbb{P}_{n+1}, y).$$

注意到存在一个定义在分布上的可计算函数，它能从 $\mathcal{P}$中的元素恢复参数，由此可以使用与引理 4 相同的技巧证明该引理。此类函数的一个例子是：对于给定的 $\mathbb{P}$ ，计算 $\mathbb{E}_{\mathbb{P}}\arctan(X)$，并假设 $\mathbb{P}$ 来自 $\mathcal{P}$，从而恢复可能的参数。

定理 6. 设 $\mathcal{P} = \{\mathbb{P}_\theta\}_{\theta\in\mathbb{R}}$ 为 $\{\mathcal{N}(\theta,\ 1)\}_{\theta\in\mathbb{R}}$ 或 $\{$柯西分布$(\theta,\ 1)\}_{\theta\in\mathbb{R}}$。则

$$\mathbf{t}^{\mathcal{P}}(x \mid y) =^{\times} \mathbf{t}(x \mid \mathbb{P}_{\lfloor x\rfloor}, y).$$

证明。 设 f 为引理5中的一致检验，并根据定理 4 取$\hat{s}(x,n) = \mathbb{I}^{0.2}_{[n-0.4,n+0.4]}(x)$。那么，正如定理5的证明中所述，我们可以断言

$$\mathbf{t}^{\mathcal{P}}(x \mid y) =^{\times} \mathbf{t}(x \mid \mathbb{P}_{\lfloor x\rfloor}, y)\hat{s}(x,\lfloor x\rfloor) + \mathbf{t}(x \mid \mathbb{P}_{\lfloor x\rfloor+1}, y)\hat{s}(x,\lfloor x\rfloor+1).$$

现在，由引理6可得

$$\mathbf{t}(x \mid \mathbb{P}_{\lfloor x\rfloor}, y)\hat{s}(x,\lfloor x\rfloor) + \mathbf{t}(x \mid \mathbb{P}_{\lfloor x\rfloor+1}, y)\hat{s}(x,\lfloor x\rfloor+1) \geq^{\times} \mathbf{t}(x \mid \mathbb{P}_{\lfloor x\rfloor}, y),$$

因此

$$\mathbf{t}^{\mathcal{P}}(x \mid y) \geq^{\times} \mathbf{t}(x \mid \mathbb{P}_{\lfloor x\rfloor}, y) \geq \mathbf{t}(x \mid y).$$

□

5 讨论

5.1 对定义3.1的讨论

在第 3 节中，我们引入了 e-变量可逼近性的概念。这个概念可能不太直观，因此我们来看几个例子，说明一些看似更直观的 e-变量组合方式会如何失效。

首先，我们将说明，使用简单的下确界投影（$x \mapsto \inf_\theta e_\theta(x)$）可能会导致不可测函数。

例 1. 考虑正态分布族 $\{\mathcal{N}(\mu,1)\}_{\mu\in\mathbb{R}}$以及某个不可测集 $L \subset \mathbb{R}$。对每个 $\mu \in L$ ，令 $e_\mu(x) = \mathbb{I}\{x \neq \mu\}$；对所有 $\mu \notin L$ ，令 $e_\mu \equiv 1$。于是，所有函数 e_μ 都是可测的，且$\mathbb{E}_\mu e_\mu(X) = 1$，但函数 $\inf_\mu e_\mu(x) = \mathbb{I}\{x \notin L\}$ 不可测，因此不能用作统计检验。

备注。 对可数集取下确界可以保持可测性，但这种做法比我们提出的方法更弱（相差一个乘性常数）；对我们而言，同样重要的是，它无法保持下半可计算性。

接下来的两个例子表明，对于非平凡的族，我们不能对 $\hat{s}$ 使用最大似然估计量

例 2. 考虑正态分布族 $\{\mathcal{N}(\mu,\ 1)\}_{\mu\in\mathbb{R}}$ 以及任意常数 $C>$)。对于每个 μ 值，我们定义函数 $e_\mu(x) = C\cdot 10^{100}\mathbb{I}\{x = \mu\}$。显然，每个 e_μ 都是可测的，且 $\mathbb{E}_\mu e_\mu(X) = 0 \le 1$，因此所有 e_μ 都是 e-变量；但与此同时，对于所有 μ，都有 $\mathbb{E}_\mu \frac{1}{C} e_{\mathrm{MLE}(x)}(x) = \frac{1}{C}\mathbb{E}_\mu e_x(x) = 10^{100} > 1$。

例 3. 考虑泊松分布族 $\{$泊松分布$(\lambda)\}_{\lambda>0}$，则最大似然估计$(x) = x$。对于每个自然数 n，定义函数 $e_n(x) = 1_{\text{泊松分布}(n)(\{n\})}\mathbb{I}\{x = n\} = \exp(n)\frac{n}{n^n}\mathbb{I}\{x = n\}!$。那么，每个 e_n 都是可测的，且$\mathbb{E}_n e_n(X) = 1$，因此所有 e_n 都是 e-变量；但与此同时，对所有 $\lambda > 0$，都有 $\mathbb{E}_\lambda e_{\mathrm{MLE}(X)}(X) = \mathbb{E}_\lambda e_X(X) = \exp(X)\ {}^{X!}/_{X^X} > 1$。

此外，由于 $\exp(x)\frac{x}{x^x} \rightharpoonup \sqrt{2\pi x!}$，对于任意 $C > 0$，总存在足够大的 λ，使得 $\mathbb{E}_\lambda \frac{1}{C} e_X(X) > 1$。

5.2 进一步研究

我们认为，有几个方向值得进一步研究。

定理 3 表明，对于同一个分布族，可能存在许多不同的网；对于同一个网，甚至也可能存在许多不同的估计量。这自然引出了一个问题：如何刻画这些网，以及哪些网和估计量能够实现最优归一化因子。

我们工作的估计量基于这样一种观点：数据包含有关分布的信息，但这些信息会受到随机噪声的干扰，因此我们需要对数据进行“修剪”，以减轻噪声的影响，同时保留有用信息。这一观点与最小描述长度原理密切相关。进一步探究 MDL原理背后的理论如何用于改进我们的结果，或许会很有意思。

参考文献

[1] Laurent Bienvenu 等人。《相对于一类测度的算法检验与随机性》。载于：斯捷克洛夫数学研究所论文集274（2011），第34–89页。网址：https://api.semanticscholar.org/CorpusID:3178723。[2] Peter Gacs。《描述复杂性与随机性讲义》。2021年。arXiv：2105.04704 [cs.IT]。网址：https://arxiv.org/abs/2105.04704。[3] Peter Gács。《一般空间上的算法随机性统一检验》。载于：理论计算机科学341.1（2005），第91–137页。ISSN：0304-3975。DOI：https://doi.org/10.1016/j.tcs.2005.03.054。网址：https://www.sciencedirect.com/science/article/pii/S030439750500188X。

[4] Peter Grünwald、Rianne de Heide 和 Wouter Koolen。《安全检验》。载于：英国皇家统计学会会刊*B*辑：统计方法86.5（2024年3月），第1091–1128页。ISSN：1369-7412。DOI：10.1093/jrsssb/qkae011。电子版：https://academic.oup.com/jrsssb/article-pdf/86/5/1091/60648648/qkae011.pdf。网址：https://doi.org/10.1093/jrsssb/qkae011。[5] 安德烈·N·柯尔莫哥洛夫。《信息论与概率论的逻辑基础》。载于：*IEEE*信息论汇刊14（1968），第662–664页。网址：https://api.semanticscholar.org/CorpusID:11402549（访问日期：2024年5月28日）。[6] Tyron Lardy、Peter Grünwald 和 Peter Harremoës。《反向信息投影与最优E统计量》。载于：*IEEE*信息论汇刊70.11（2024），第7616–7631页。DOI：10.1109/TIT.2024.3444458。[7] 列昂尼德·A·列文。《随机性守恒不等式；数学理论中的信息与独立性》。载于：信息与控制61.1（1984），第15–37页。ISSN：0019-9958。DOI：https://doi.org/10.1016/S0019-9958(84)80060-1。网址：https://www.sciencedirect.com/science/article/pii/S0019995884800601。[8] 列昂尼德·阿纳托利耶维奇·列文。《随机性的统一检验》。载于：科学院学报。第227卷。第1期。俄罗斯科学院。1976年，第33–35页。[9] Aaditya Ramdas 和 Ruodu Wang。《使用E值进行假设检验》。载于：基础与前沿Ⓡ 统计学1.1-2（2025），第1–390页。ISSN：2978-4212。DOI：10.1561/3600000002。网址：http://dx.doi.org/10.1561/3600000002。[10] Alexander Shen。《随机性检验：理论与实践》。载于：逻辑与计算领域 *III*。A. Blass 等人编。第12180卷。计算机科学讲义。施普林格出版社，2020年5月，第258–290页。DOI：10.1007/978-3-030-48006-6_18。网址：https://hal-lirmm.ccsd.cnrs.fr/lirmm-03065320（访问日期：2024年5月28日）。[11] V. G. Vovk 和 V. V. V'Yugin。《贝叶斯方法的经验有效性》。载于：英国皇家统计学会会刊*B*辑（方法论）55.1（2018年12月），第253–266页。ISSN：0035-9246。DOI：10.1111/j.2517-6161.1993.tb01482.x。电子版：https://academic.oup.com/jrsssb/article-pdf/55/1/253/49173327/jrsssb_55_1_253.pdf。网址：https://doi.org/10.1111/j.2517-6161.1993.tb01482.x。[12] V.G Vovk。《伯努利模型参数的学习》。载于：计算机与系统科学杂志55.1（1997），第96–104页。ISSN：0022-0000。DOI：https://doi.org/10.1006/jcss.1997.1502。网址：https://www.sciencedirect.com/science/article/pii/S00220000979

15026.

[13] Vladimir Vovk. “非算法随机性理论”。载于：逻辑与计算领域 *III*：献给 *Yuri Gurevich 80* 岁寿辰的论文集。施普林格，2020年，第 323–340 页。[14] Vladimir Vovk. 关于伯努利性的概念。2016年。arXiv: 1612.08859[math.ST]。（访问于 2024年5月28日）。[15] Vladimir Vovk 和 Ruodu Wang。“E值：校准、组合与应用”。载于：《统计年鉴》49.3（2021年），第 1736–1754 页。doi: 10.1214/20-AOS2020。url: https://doi.org/10.1214/20-AOS2020。[16] K. Weihrauch. 可计算分析：导论。《理论计算机科学文献》。EATCS系列。施普林格柏林海德堡，2000年。isbn: 9783540668176。url: https://books.google.co.uk/books? id= OPolVWVFDJYC.

一种记法

- e 表示常数 e $= 2.71\ldots$，以区别于 e，后者将用于命名 e-变量

- $|A|$ 和 $\#A$ 可以互换使用，表示集合 A 的基数
- $|x|$ 表示一个数 x 的绝对值
- $\|x\|$ 表示一个点 $x \in \mathbb{R}^n$ 的欧几里得范数
- $x_{1:n}$ 表示 $\mathbb{R}^n$ 中的一个点，其坐标为 $x_1, x_2, \ldots, x_n$。
- $a \leq^+ b$ 表示 $a \leq b + C$，其中 C为某个数值常数；$\geq^+$也类似

- $a \leq^\times b$ 表示 $a \leq Cb$，其中 $C > 0$为某个数值常数；$\geq^\times$也类似

- $a =^+ b$ 表示 $a \leq^+ b$ 和 $b \leq^+ a$ 都成立；$a =^\times b$也类似
- $[l : r]$ 表示集合 $\{n \in \mathbb{Z} \mid l \leq n \leq r\}$
- A^ε 是集合 A的 ε-邻域
- $\mathbb{I}_A$ 表示集合 A的示性函数
- $\mathbb{I}_A^\varepsilon$ 表示 $\mathbb{I}_A$ 的有界连续逼近，该逼近在A上取值为 1，在 A^ε 的补集上取值为 0，并在 $A^\varepsilon \setminus A$上从 1 到 0 线性插值

- 取整$_\mathcal{S}$ 表示将实数映射到离散集 $\mathcal{S} \subset \mathbb{R}$中最近元素的函数；当存在两个这样的元素时，就本文而言，取整$_\mathcal{S}$可以是其中任意一个

- $\mathrm{kl}(\theta_1\|\theta_2)$表示Kullback–Leibler散度$\mathrm{KL}(\mathbb{P}_{\theta_1}\|\mathbb{P}_{\theta_2})$，此时族 $\{\mathbb{P}_\theta\}_{\theta\in\Theta}$ 由上下文唯一确定

  当族 $\{\mathbb{P}_\theta\}_{\theta\in\Theta}$ 由上下文唯一确定时，• $\mathbb{E}_\theta$ 表示 $\mathbb{E}_{\mathbb{P}_\theta}$

- $\mathfrak{P}(\mathcal{X})$ 表示 $\mathcal{X}$ 上所有概率测度的集合
- $\mathbf{t}$ 表示一个通用的均匀随机性检验，该检验为本文任意选定并固定

- 为避免与单元素集合混淆，指标集总是在指标集记号中明确指定，也就是说，记号 $\{\mathbb{P}_\theta\}$ 总是表示一个单元素集合，绝不是 $\{\mathbb{P}_\theta\}_{\theta\in\Theta}$ 的缩写

B 关于 e-变量可逼近性的证明

B.1 引理的证明

引理*1*的证明。 我们要证明 $\frac{1}{C}e_{\hat{s}(X)}(X)$ 是一个 e-变量。也就是说，我们需要证明，对于所有 $\theta\in\Theta$

$$\int_{\mathbb{R}^n} e_{\hat{s}(x)}(x)\mathbb{P}_\theta(\mathrm{d}x)\le C.$$

首先，对于每个 $s\in\mathcal{S}$，我们将 $\mathcal{X}_s=\hat{s}^{-1}(s)$ 记为 $\subseteq\mathbb{R}^n$。现在，

$$\begin{aligned}\int_{\mathbb{R}^n} e_{\hat{s}(x)}(x)\mathbb{P}_\theta(\mathrm{d}x)&=\sum_{s\in\mathcal{S}}\int_{\mathcal{X}_s} e_s(x)p_\theta(x)\mathrm{d}x\\&=\sum_{s\in\mathcal{S}}\int_{\mathcal{X}_s}\frac{p_\theta(x)}{p_s(x)}e_s(x)p_s(x)\mathrm{d}x\\&\overset{\text{by (p i)}}{=}\sum_{s\in\mathcal{S}}\int_{\mathcal{X}_s}\exp\Big[d(g(x)\|s)-d(g(x)\|\theta)\Big]e_s(x)p_s(x)\mathrm{d}x\end{aligned}$$

将 $C_s=C_{s,\ \theta}=\sup_{x\in\mathcal{X}_s} d(g(x)\|s)\ -d(g(x)\|\theta)$

$$\begin{aligned}&\le\sum_{s\in\mathcal{S}}\exp(C_s)\int_{\mathcal{X}_s} e_s(x)p_s(x)\mathrm{d}x\\&\le\sum_{s\in\mathcal{S}}\exp(C_s)\underbrace{\int_{\mathbb{R}^n} e_s(x)p_s(x)\mathrm{d}x}_{\le 1,\text{ by def. of } e\text{-var.}}\\&\le\sum_{s\in\mathcal{S}}\exp(C_s).\end{aligned}$$

现在我们必须估计总和 $\sum_{s\in\mathcal{S}}\exp(C_s)$。

根据（p ii），我们有

$$C_{s,\theta} \le c' - \inf_{x \in \mathcal{X}_s} d(g(x)\|\theta),$$

并且，由于 d 的非负性，可得 $C_s \le c'$.

对于某个 k，只要 $s \ge$ 后继$_{\mathcal{S}}^{(k+1)}(\theta)$，就有前驱$_{\mathcal{S}}(s) \ge$后继$_{\mathcal{S}}^{(k)}(\theta)$，并且根据 (p iii)，对所有 $x \in \mathcal{X}_s$:

$$g(x) > \mathrm{pred}_{\mathcal{S}}(g(x)) \ge \mathrm{pred}_{\mathcal{S}}(s) \ge \mathrm{succ}_{\mathcal{S}}^{(k)}(\theta) > \theta,$$

并且 $|\mathcal{S} \cap (g(x), \theta)| \ge k$，根据 (p iv)，我们有:

$$d(g(x)\|\theta) \ge (1+\alpha)\log(k-1), \quad \forall x \in \mathcal{X}_s,$$

因此，在这种情况下，$C_{s,\theta} \le c' - (1+\alpha)\log(k-1)$。

同理，对于 $s \le$ 前驱$_{\mathcal{S}}^{(k+1)}(\theta)$，我们也得到相同的界。

综上所述，我们得到

$$\begin{aligned}
\sum_{s\in\mathcal{S}} \exp(C_{s,\theta}) &\le \left(\sum_{\substack{s \le \mathrm{pred}_{\mathcal{S}}^{(3)}(\theta) \\ s \in \mathcal{S}}} + \sum_{\substack{\mathrm{succ}_{\mathcal{S}}^{(3)}(\theta) \le s \\ s \in \mathcal{S}}} + \sum_{\substack{s \in \mathcal{S} \\ \mathrm{pred}_{\mathcal{S}}^{(3)}(\theta) < s \\ s < \mathrm{succ}_{\mathcal{S}}^{(3)}(\theta)}} \right) \exp(C_{s,\theta}) \\
&\le \sum_{k=1}^{\infty} \mathrm{e}^{c' - (1+\alpha)\log k} + \sum_{k=1}^{\infty} e^{c' - (1+\alpha)\log k} + 5\mathrm{e}^{c'} \\
&= \mathrm{e}^{c'} \left(5 + 2\sum_{k=1}^{\infty} \frac{1}{k^{1+\alpha}} \right) \\
&\le e^{c'} \left(5 + 2\left(1 + \frac{1}{\alpha} \right) \right) \\
&= \exp(c') \left(7 + \frac{2}{\alpha} \right),
\end{aligned}$$

由此完成证明。 □

引理2的证明。 证明过程与引理1完全相同，直到得出 $C_{s,\,\theta} \le c' - \inf_{x\in\mathcal{X}_s} d(g(x)\|\theta)$，以及 $C_s \le c'$.

由（p i'）和 d 的非负性可知，对于任意 $\theta_1' \le \theta_1 \le \theta_2 \le \theta_2'$:

$$d(\theta_1\|\theta_2) \le d(\theta_1'\|\theta_2'), \quad d(\theta_2\|\theta_1) \le d(\theta_2'\|\theta_1').$$

对于后继$_{\mathcal{S}}(s) <$ 前驱$_{\mathcal{S}}(\theta) \le \theta$，由 (p iii) 及反复应用

（p i'）可得:

$$\begin{aligned}
d(g(x)\|\theta) &\geq d(\mathrm{succ}_{\mathcal{S}}(g(x))\|\,\mathrm{pred}_{\mathcal{S}}(\theta)) \\
&\geq \sum_{\substack{t\in\mathcal{S}\\ \mathrm{succ}_{\mathcal{S}}(g(x))\leq t\\ t<\mathrm{pred}_{\mathcal{S}}(\theta)}} d(t\|\,\mathrm{succ}_{\mathcal{S}}(t)) \\
&\overset{(pii')}{\geq} \sum_{\substack{t\in\mathcal{S}\\ \mathrm{succ}_{\mathcal{S}}(g(x))\leq t\\ t<\mathrm{pred}_{\mathcal{S}}(\theta)}} c \\
&\overset{(piii)}{\geq} \sum_{\substack{t\in\mathcal{S}\\ \mathrm{succ}_{\mathcal{S}}(s)\leq t\\ t<\mathrm{pred}_{\mathcal{S}}(\theta)}} c \\
&\geq (|\mathcal{S}\cap(s,\theta)|-1)\,c.
\end{aligned}$$

同理，对于 $\theta \leq$后继$_{\mathcal{S}}(\theta) <$前驱$_{\mathcal{S}}(s)$，我们有

$$d(g(x)\|\theta) \geq (|\mathcal{S}\cap(\theta,s)|-1)c.$$

综上，我们得到

$$\begin{aligned}
\sum_{s\in\mathcal{S}}\exp(C_{s,\theta}) &\leq \left(\sum_{\substack{s\leq \mathrm{pred}_{\mathcal{S}}^{(2)}(\theta)\\ s\in\mathcal{S}}} + \sum_{\substack{\mathrm{succ}_{\mathcal{S}}^{(2)}(\theta)\leq s\\ s\in\mathcal{S}}} + \sum_{\substack{s\in\mathcal{S}\\ \mathrm{pred}_{\mathcal{S}}^{(2)}(\theta)<s\\ s<\mathrm{succ}_{\mathcal{S}}^{(2)}(\theta)}}\right)\exp(C_{s,\theta}) \\
&\leq \sum_{k=0}^{\infty}\mathrm{e}^{c'-kc} + \sum_{k=0}^{\infty}\mathrm{e}^{c'-kc} + 3\mathrm{e}^{c'} \\
&\leq \frac{2\mathrm{e}^{c'+c}}{\mathrm{e}^{c}-1} + 3\mathrm{e}^{c'}, \\
&= \exp(c')\left(3+2\left(1+\frac{1}{\mathrm{e}^{c}-1}\right)\right) \\
&= \exp(c')\left(5+\frac{2}{\mathrm{e}^{c}-1}\right),
\end{aligned}$$

至此，证明完毕。 □

引理 *3* 的证明。我们将用到这样一个事实：对于典范指数族

$$\begin{aligned}
\mathrm{kl}(\eta\|\eta') =& (\eta-\eta')\mathbb{E}_{\eta}T(X) - (A(\eta)-A(\eta')) \\
=& (\eta-\eta')\frac{\mathrm{d}}{\mathrm{d}\eta}A(\eta) - (A(\eta)-A(\eta')).
\end{aligned}$$

1. 注意,

$$\begin{aligned}\log\frac{p_\eta(x)}{p_{\eta'}(x)}=&\log\frac{h(x)\exp\left(\eta T(x)-A(\eta)\right)}{h(x)\exp\left(\eta' T(x)-A(\eta')\right)}\\=&(\eta-\eta')T(x)-(A(\eta)-A(\eta'))\\=&(\eta-\eta')\mathbb{E}_{\hat\eta(x)}T(X)-(A(\eta)-A(\eta'))\\=&(\hat\eta(x)-\eta')\mathbb{E}_{\hat\eta(x)}T(X)-(A(\hat\eta(x))-A(\eta'))\\&-\left[(\hat\eta(x)-\eta)\mathbb{E}_{\hat\eta(x)}T(X)-(A(\hat\eta(x))-A(\eta))\right]\\=&\mathrm{kl}(\hat\eta(x)\|\eta')-\mathrm{kl}(\hat\eta(x)\|\eta)\end{aligned}$$

2. 我们使用布雷格曼散度 $\mathcal{B}_A(\eta'\|\eta)=\mathrm{kl}(\eta\|\eta')$ 的勾股等式:

$$\begin{aligned}\mathrm{kl}(\eta_3\|\eta_1)-\mathrm{kl}(\eta_2\|\eta_1)-\mathrm{kl}(\eta_3\|\eta_2)&=\mathcal{B}_A(\eta_1\|\eta_3)-\mathcal{B}_A(\eta_1\|\eta_2)-\mathcal{B}_A(\eta_2\|\eta_3)\\&=(\eta_1-\eta_2)\left(\left(\frac{\mathrm{d}}{\mathrm{d}\eta}A\right)(\eta_2)-\left(\frac{\mathrm{d}}{\mathrm{d}\eta}A\right)(\eta_3)\right)\end{aligned}$$

由于累积量函数 A 是凸函数，函数 $\frac{\mathrm{d}}{\mathrm{d}\eta}A$ 是非递减的，因此，只要 $\eta_1\le\eta_2\le\eta_3$ 或 $\eta_1\ge\eta_2\ge\eta_3$，差值 $(\eta_1-\eta_2)$ 和 $\left(\left(\frac{\mathrm{d}}{\mathrm{d}\eta}A\right)(\eta_2)-\left(\frac{\mathrm{d}}{\mathrm{d}\eta}A\right)(\eta_3)\right)$ 的符号就相同，这意味着

$$\mathrm{kl}(\eta_3\|\eta_1)\ge\mathrm{kl}(\eta_3\|\eta_2)+\mathrm{kl}(\eta_2\|\eta_1),$$

由此完成证明。 □

B.2 定理3的证明

均匀分布的证明。 对于离散和连续均匀分布族，证明可以直接进行。

对于 $2^n<x\le 2^{n+1}$，我们得到 $\hat s(x)=2^{n+1}$，并且

$$\begin{aligned}\mathbb{E}_N e_{\hat s(X)}(X)&=\sum_{x=0}^{N}\frac{1}{N+1}e_{\hat s(x)}(x)\\&=\sum_n\sum_{2^n<x\le 2^{n+1}}\frac{1}{N+1}e_{\hat s(x)}(x)\\&\le\sum_n\frac{2^{n+1}+1}{N+1}\underbrace{\sum_{x\le 2^{n+1}}\frac{1}{2^{n+1}+1}e_{2^{n+1}}(x)}_{\mathbb{E}_{2^{n+1}}e_{2^{n+1}}(X)\le 1}\\&\le\sum_n\frac{2^{n+1}+1}{N+1}\le 3,\end{aligned}$$

这意味着，对于所有 N，$\mathbb{E}_N\frac{1}{3}e_{\hat s(X)}(X)\le 1$。

连续情形的证明基本相同。 □

泊松分布的证明。概率密度函数 $p_\lambda(n)$ 的泊松分布(λ)为

$$p_\lambda(n) = \mathrm{e}^{-\lambda}\frac{\lambda^n}{n!} = \frac{1}{n!}\exp(n\log\lambda - \lambda).$$

为了应用引理 3，我们改用参数化 $\eta = \log\lambda$，即$\mathbb{E}_\eta f = \mathbb{E}_{\mathrm{Pois}(\mathrm{e}^\eta)} f$。

注意到 $n = \mathbb{E}_{\log n} X$，因此根据引理 3，对于$d = \mathrm{kl}$ 和 $g = \log$，我们有 (p i) 和 (p i')。

对分布进行参数化时，若使用 η，我们需要处理 $\mathcal{S}_H = \{2\log t\}_{t\in\mathbb{N}_{\geq 1}}$。此外，我们还有

$$\mathrm{kl}(\eta_1\|\eta_2) = \mathrm{e}_1^\eta(\eta_1 - \eta_2) - (\mathrm{e}^{\eta_1} - \mathrm{e}^{\eta_2}) = \lambda_1\log\frac{\lambda_1}{\lambda_2} - (\lambda_1 - \lambda_2) = \mathrm{kl}_\lambda(\lambda_1\|\lambda_2).$$

为了证明 (p ii)，任取 t^2 和 $n \in \hat{s}^{-1}(t^2)$。注意到 $|n - t^2| \leq t$。我们有：

$$\begin{aligned}
\mathrm{kl}(\log n\|\log t^2) &= \mathrm{kl}_\lambda(n\|t^2)\\
&= n\log\frac{n}{t^2} - (n - t^2)\\
&= n\log\left(1 + \frac{n - t^2}{t^2}\right) - (n - t^2)\\
&\leq \left(\frac{n}{t^2} - 1\right)(n - t^2)\\
&= \frac{(n - t^2)^2}{t^2}\\
&\leq 1.
\end{aligned}$$

性质（p iii）显然成立，因为$\hat{s}$ 会将一个数映射到最近的完整平方之一——即 $\mathcal{S}$的元素。

性质 (p ii') 的证明与 (p ii) 类似。

$$\begin{aligned}
\mathrm{kl}(\log t^2\|\log(t+1)^2) &= t^2\log\frac{t^2}{(t+1)^2} - (t^2 - (t+1)^2)\\
&= -2t^2\log\left(1 + \frac{1}{t}\right) + 2t + 1\\
&\geq -2t^2\frac{1}{t} + 2t + 1\\
&= 1.
\end{aligned}$$

$$\begin{aligned}\mathrm{kl}(\log(t+1)^2 \| \log t^2) &= (t+1)^2 \log \frac{(t+1)^2}{t^2} - ((t+1)^2 - t^2) \\ &= 2(t+1)^2 \log\left(1 + \frac{1}{t}\right) - 2t - 1 \\ &\geq 2(t+1)^2 \left(\frac{1}{t} - \frac{1}{2t^2}\right) - 2t - 1 \\ &= \frac{(t+1)^2(2t-1) - (2t+1)t^2}{t^2} \\ &= \frac{2t^2 - 1}{t^2} \geq 1.\end{aligned}$$

由于现在我们已具备所有 (p i)–(p iii)、(p i')、(p ii')，由引理 2 可得 e-值可逼近性$\{\mathrm{Pois}(\lambda)\}_{\lambda>0}$。 □

固定方差的正态分布的证明。密度p_μ为 $\mathcal{N}(\mu, 1)^{\otimes n}$ 是

$$p_\mu(x_{1:n}) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\|x_{1:n}^2\|/2} \exp\left(\mu \sum_k x_k - \frac{n\mu^2}{2}\right),$$

并且，由于 $\sum_k x_k = \mathbb{E}_{\sum x_k} \sum_k X_k$根据引理 3，我们得到性质（p i）和（p$_n$i'），适用于 $d(\theta_1 \| \theta_2) = \mathrm{kl}(\theta_1 \| \theta_2) = \frac{n}{2}(\theta_1 - \theta_2)^2$ 以及 $g(x_{1:n}) = \frac{\sum_k x_k}{n}$。

为证明性质（p ii），注意到 $x_{1:n} \in \hat{s}^{-1}(s)$ 意味着 $\left|\frac{\sum_k x_k}{n} - s\right| \leq \frac{\alpha}{2\sqrt{n}}$，因此

$$d(g(x_{1:n}) \| s) = \frac{n}{2}(g(x_{1:n}) - s)^2 \leq \frac{n}{2}\frac{\alpha^2}{4n} = \frac{\alpha^2}{8}.$$

由于$\hat{s}(x) = \text{取整}_{\mathcal{S}}(g(x))$，性质（p iii）显然成立。

网的连续元素 s, s' 之间的差恰好是$\frac{\alpha}{\sqrt{n}}$，因此

$$d(s \| s') = \frac{n}{2}\frac{\alpha^2}{n} = \frac{\alpha^2}{2},$$

这证明了性质（p ii'），并完成了证明。 □

零均值的正态分布。我们有

$$p_{\sigma^2}(x_{1:n}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\|x_{1:n}\|^2}{2\sigma^2} - \frac{n}{2}\log\sigma^2\right).$$

由于 $\|x_{1:n}\|^2 = \sum_k x_k^2$ 以及所有坐标都是独立同分布的，我们有 $T(x_{1:n}) = \mathbb{E}_{g(x_{1:n})} T(X_{1:n})$，对于 $T(x_{1:n}) = -\frac{1}{2}\sum_k x_k^2$和 $g(x_{1:n}) = \frac{\|x_{1:n}\|^2}{n}$。

由于典则参数化 $\eta = 1/\sigma^2$ 会颠倒顺序，当我们用方差定义 $d = \mathrm{kl}$时，仍然得到反向三角不等式（p i'），因此可以继续在标准参数化下进行讨论。

对于两个方差 σ_1^2, σ_2^2，我们有

$$\mathrm{kl}(\sigma_1^2 \| \sigma_2^2) = \frac{n}{2}\left(\frac{\sigma_1^2}{\sigma_2^2} - \log\frac{\sigma_1^2}{\sigma_2^2} - 1\right).$$

由于 $\hat{s}(x) = \text{取整}_{\mathcal{S}}(g(x))$，因此我们立即得到 (p iii)。我们还知道，对于任意 $s = \hat{s}(x)$，都有 $g(x)/s \in \left[\frac{\sqrt{n}}{1+\sqrt{n}}, \frac{1+\sqrt{n}}{\sqrt{n}}\right]$；对于相邻的 $s,\ s' \in \mathcal{S}$，有 $s/s' \in \left\{\frac{\sqrt{n}}{1+\sqrt{n}}, \frac{1+\sqrt{n}}{\sqrt{n}}\right\}$。

令 $h(t) = \frac{n}{2}(t - \log t - 1)$，则 $\mathrm{kl}(\sigma_1^2 \| \sigma_2^2) = h\left(\sigma_1^2/\sigma_2^2\right)$。

首先，我们将证明 (p ii')。

$$\begin{aligned}
h\left(\frac{1+\sqrt{n}}{\sqrt{n}}\right) &= \frac{n}{2}\left(1 + \frac{1}{\sqrt{n}} - \underbrace{\log\left(1 + \frac{1}{\sqrt{n}}\right)}_{\le \frac{1}{\sqrt{n}} - \frac{1}{4n}} - 1\right) \\
&\ge \frac{n}{2}\left(\frac{1}{\sqrt{n}} - \frac{1}{\sqrt{n}} + \frac{1}{4n}\right) \\
&\ge \frac{1}{8}, \\
h\left(\frac{\sqrt{n}}{1+\sqrt{n}}\right) &= \frac{n}{2}\left(1 - \frac{1}{1+\sqrt{n}} - \underbrace{\log\left(1 - \frac{1}{1+\sqrt{n}}\right)}_{\le -\frac{1}{1+\sqrt{n}} - \frac{1}{4(1+\sqrt{n})^2}} - 1\right) \\
&\ge \frac{n}{2}\left(-\frac{1}{1+\sqrt{n}} + \frac{1}{1+\sqrt{n}} + \frac{1}{4\left(1+\sqrt{n}\right)^2}\right) \\
&\ge \frac{1}{32},
\end{aligned}$$

因此，对于网中的任意两个相邻元素 $s,\ s'$，都有 $d(s\|s') \ge 1/32$。

现在，为证明 (p ii)，设 t 位于 $-\frac{1}{1+\sqrt{n}}$ 和 $\frac{1}{\sqrt{n}}$ 之间。

$$\begin{aligned}
h(1+t) &= \frac{n}{2}((1+t) - \underbrace{\log(1+t)}_{\ge t - t^2} - 1) \\
&\le \frac{n}{2}(t - t + t^2) \\
&\le \frac{|t|^2 n}{2} \\
&\le \frac{1}{n}\frac{n}{2} = \frac{1}{2}
\end{aligned}$$

□

柯西分布的证明。对于柯西分布，我们有 $p_\theta(x) = \pi\,1+\,^{1}\,((x-\theta)^2)$，因此

$$\log \frac{p_\theta(x)}{p_s(x)} = \log\left(1+(x-s)^2\right) - \log\left(1+(x-\theta)^2\right).$$

取 $g(x) = x$，$d(\theta_1\|\theta_2) = \log\left(1+(\theta_1-\theta_2)^2\right) \geq \log 1 = 0$。

对于 $\hat{s} = r^\varepsilon$，我们有 $x \in \hat{s}^{-1}(n) \implies x \in [n-0.5-\varepsilon, n+0.5+\varepsilon]$，因此

$$d(x\|n) \leq \log(1+(0.5+\varepsilon)^2) \leq \log 2.$$

对于 r^ε，性质（p iii）可直接由同一观察得出。

现在，假设 θ_1 和 θ_2 之间有 k 个整数。那么 $|\theta_1-\theta_2| \geq k-1$ 和

$$d(\theta_1\|\theta_2) = \log\left(1+(\theta_1-\theta_2)^2\right) \geq 2\log(k-1).$$

该结果现由引理1得出。 □

正态分布的证明 r^ε。该证明与取整$_{\mathcal{S}}$ 的证明基本相同，但界限略弱。

□

C 随机性检验的证明

第 4 节中的所有引理都由定理 4.1.1 推出。我们在 [2], 中重述该定理。

定理 7。设 ϕ_e, $e \in \mathbb{N}$, 是所有定义在$\mathcal{X} \times \mathfrak{M}(\mathcal{X}) \times \mathcal{Y} \to [0,\infty]$. 上的下半连续函数的一个枚举，其中 $\mathfrak{M}(\mathcal{X})$ 是 $\mathcal{X}$ 上所有有界测度的集合。设 $b\colon \mathcal{Y} \times \mathfrak{M}(\mathcal{X}) \to [0,\infty]$ 是一个下半连续函数。存在一个递归函数 $e \mapsto e'$，具有以下性质。

对于每个 e，函数 $\phi_{e'}$ 处处有定义，且其积分不超过 $b(\mathbb{P}, y)$。

对于每个 $e, y, \mathbb{P}$，若 $\mathbb{E}_{\mathbb{P}}\phi_e(X \mid \mathbb{P}, y) < b(\mathbb{P}, y)$，则 $\phi_{e'}(\cdot \mid \mathbb{P}, y) =^{\times} \phi_e(\cdot \mid \mathbb{P}, y)$。