经经验熵归一化的二进制词调整柯尔莫哥洛夫复杂度

Brani Vidakovic，德克萨斯农工大学
统计学系

arXiv:2512.21193v1 [stat.CO] 24 Dec 2025

有限二进制词的柯尔莫哥洛夫复杂度既反映算法结构，也反映词中符号的经验分布。符号频率与二分之一相差较大的词，组合丰富度较低，因此按标准定义衡量时显得不那么复杂。本文提出一种熵归一化复杂度度量，即将词的柯尔莫哥洛夫复杂度除以其观测到的零和一分布的经验熵。这种调整将内在描述复杂度与符号不平衡带来的纯组合效应区分开来。对于可构造交换测度下的 Martin-L¨of 随机序列，调整后的复杂度线性增长并收敛于一。一种病态构造表明，底层测度的正则性至关重要。所提出的框架以自然的方式联系起柯尔莫哥洛夫复杂度、经验熵与随机性，并为随机性检验和结构化二进制数据分析提供了应用方向。

关键词：柯尔莫哥洛夫复杂度、熵归一化、算法随机性、通用先验。

AMS学科分类：68Q30, 60G09。

1 引言

柯尔莫哥洛夫复杂度由 Kolmogorov (1965)、Zvonkin 和 Levin (1970) 以及 Levin (1974) 的开创性研究引入，为有限对象和无限序列提供了一种算法复杂度概念。如果一个二进制词在通用前缀机上输出它的最短程序长度基本等于词本身的长度，就认为它是复杂的；而高度结构化的词则可以用短得多的描述来表示。

一个经典且广为人知的观察是，柯尔莫哥洛夫复杂度会受到符号经验分布的显著影响。如果一个长度为 n 的词包含 n_1 个 1，那么其组合类的对数大小约为 $nH(p)$，其中 $p = n_1/n$ 和 H 表示二元熵。由于 $H(p)$ 会随符号频率而变化，原始柯尔莫哥洛夫复杂度无法充分区分内在描述结构与符号不平衡所造成的组合效应。

为解决这一问题，我们分析一种熵归一化复杂度尺度

$$KA(x) = \frac{K(x)}{H(\nu_{\ell(x)})}, \tag{1}$$

其中， $K(x)$ 表示有限二进制词 x 的柯尔莫哥洛夫复杂度， $H(\nu_{\ell(x)})$ 则表示其观测到的零和一分布的经验熵。对于常值字（经验熵为零），这种归一化没有定义；但对于其他所有有限词，它都有意义。Zvonkin 和 Levin (1970) 首次指出了这种归一化，但未作深入分析。

本文的主定理表明，对于可构造可交换测度下的 Martin L¨of 随机序列，比值 $KA(\omega^m)/m$ 收敛于 1。从这个意义上说，经验熵为随机序列的描述复杂度提供了恰当的归一化尺度。

这种熵归一化表述建立在早期研究的基础上；在那些研究中，算法复杂性、Martin L¨of 检验、通用先验和奥卡姆型原则被纳入统一框架加以发展，并特别关注 Ω 上的测度以及 Schnorr 式复杂性概念（Schnorr，1971；Vidakovic，1988年；Zvonkin 与 Levin，1970）。算法信息领域一再强调组合熵与描述复杂性之间的概念区别；例如，参见 Kramosil（1982）和 Calude（2002），其中信息量被视为一种内在的结构属性，不能仅由符号频率还原。

本归一化在有限词层面凸显了经验熵的作用，此时，不平衡带来的组合效应在取任何渐近极限之前就已明确且可量化。这种有限样本视角在本文后续部分尤为重要；届时，我们将在第8节中开发有效替代量并构建可实施的随机性检验。

这项研究也与早期试图将二进制序列中的算法结构与组合效应区分开来的工作密切相关。Vidakovic（1983）提出了一种基于过程的条件复杂度（KP 复杂度），并引入了一种通过条件复杂度之差定义的信息测度。该框架强调了条件对象和生成过程在量化信息与随机性方面的作用，并确立了复杂度增长率与 Schnorr 型和 Martin-L¨of 型随机性之间的渐近联系。

从互补的视角来看，Stojanovic 和 Vidakovic（1987年）研究了二进制词的有效组合复杂度测度。研究表明，复杂度深受词重影响，明显偏离平衡状态必然会降低可达到的复杂度。这项工作还确立了一种香农型集中现象：对于固定长度，大多数词的复杂度都接近其组合类所允许的最大值。

本文引入的熵归一化柯尔莫哥洛夫复杂度，可视为这些思想在算法层面的对应形式。通过明确剥离符号不平衡的组合贡献，这种归一化提取出内在的描述复杂度，并直接在柯尔莫哥洛夫复杂度与算法随机性的框架内运作。此外，这种方法可自然推广至条件情形和互信息情形（见第7节），从而为有效组合测度、基于过程的条件复杂度以及信息论中的熵率搭建起统一的桥梁。

2 结构安排与贡献

本文针对有限二进制词提出一种经验熵归一化的柯尔莫哥洛夫复杂度，并将由此得到的归一化量与 $\Omega = \{0,1\}^{\mathbb{N}}$ 上常规（可构造）概率测度下的 Martin–L¨of 随机性联系起来。

全文围绕一个核心比率展开，

$$R(x) = \frac{K(x)}{\ell(x)\, H(\nu_{\ell(x)})}, \tag{2}$$

该比率将一个词的柯尔莫哥洛夫复杂度与由其符号频率决定的经验组合基线进行比较。式（1）中的量 $KA(x)$ 只是相应的尺度，$KA(x) = \ell(x)R(x)$。从概念上讲，$R(x)$ 是主要统计量：分子和分母都以比特为单位衡量，该比率直接量化偏离壳层典型行为的程度。经验熵基线在第4节中作了正式定义，并在图1中展示。

全文采用的记号约定

我们严格区分有限词与无限序列前缀的记号，并对经验频率采用一致的记法。

若 x 表示一个有限二进制词，则 $\ell(x)$ 表示其长度，$w(x)$ 表示其重量，也就是 x 中 1 出现的次数。1 的经验比例为

$$p(x) = \frac{w(x)}{\ell(x)},$$

其对应的经验分布为

$$\nu_{\ell(x)} = (p(x), 1 - p(x)).$$

如果 $\omega \in \Omega$ 表示一个无限二进制序列，那么 ω^m 表示其长度为 m 的前缀。该前缀的重量为 $w(\omega^m)$，其中 1 的经验比例为

$$p_m(\omega) = \frac{w(\omega^m)}{m},$$

相应的经验分布记为 $\nu_m(\omega)$。在全文中，m 专指前缀长度，而 $\ell(x)$ 表示有限词的长度。

调整后的复杂度与归一化原理

经典计数论证表明，对于固定整数（m，k），壳层

$$S(m,k) = \{x \in \{0,1\}^m : w(x) = k\}$$

的基数为 $\binom{m}{k}$。因此，描述一个壳层典型词的自然组合基线是

$$\begin{aligned} \log_2 |S(m,k)| &= \log_2 \binom{m}{k} \\ &= mH(k/m) + O(\log m). \end{aligned} \tag{3}$$

这一观察促使我们采用基于熵的归一化，从内在算法结构中剥离符号不平衡的影响。

对于具有 $H(\nu_{\ell(x)}) > 0$ 的有限词 x，我们通过式（2）定义归一化复杂度比，并将相应的调整后复杂度尺度定义为

$$KA(x) = \ell(x)\, R(x) = \frac{K(x)}{H(\nu_{\ell(x)})}. \tag{4}$$

$R(x)$ 的值接近 1，表明 x 的描述成本与其经验分布所隐含的壳基准相符；而显著低于 1 的值则表明，存在仅凭符号频率无法解释的算法正则性。第4节将这一直觉进一步明确为精确的计数结论，并将其作为第4节中 Martin–L¨of 检验构造的关键步骤。

贡献

本文有四项主要贡献。

（i）有限词的经验熵归一化。我们对二进制词的经验熵基线进行了形式化，并推导出精确的壳界，表明对于固定经验分布，只有指数级小比例的词会表现出给定程度的复杂度亏损。

（ii）Martin-L¨随机序列的极限定理。对于相对于正则可构造测度 μ 而 Martin L¨ 随机的无限序列 ω，我们建立了如下渐近恒等式

$$\begin{aligned} \frac{K(\omega^m)}{m} &\longrightarrow h(\mu), \\ R(\omega^m) &= \frac{K(\omega^m)}{m\, H(\nu_m(\omega))} \longrightarrow 1, \end{aligned} \tag{5}$$

其中 $h(\mu)$ 是该测度的熵率。证明见第4节，随后紧接着在第4.1节讨论正则性。

（iii）正则性假设的必要性。 我们构造了一个病态反例，表明如果不对 μ施加适当的正则性条件，经验熵可能无法提供正确的归一化尺度，从而促使我们采用高阶经验摘要。

（iv）有效替代量与可实施检验。 通过用可计算上界替代Kolmogorov复杂性，我们获得了有效统计量

$$R_{\mathrm{eff}}(x) = \frac{K_{\mathrm{eff}}(x)}{\ell(x)\, H(\nu_{\ell(x)})},$$

从而引出第8节提出并在图2中展示的操作性随机性检验。

本文结构如下。第3节介绍符号和预备知识。第4节阐述 $R(x)$和 $KA(x)$的基本性质。第4节证明主要极限定理，第4.1节解释正则性为何至关重要。第5节和第6节探讨算法解释和应用解释。第7节讨论条件扩展和互扩展，第8节整合有效检验逻辑。第9节为结论。

3 预备知识

令 $\Omega = \{0,1\}^{\mathbb{N}}$ 表示所有无限二进制序列构成的空间。对于 $\omega \in \Omega$，其长度为 m 的前缀记作 $\omega^m = \omega_1\omega_2\cdots\omega_m$。对于有限二进制词$x = x_1x_2\cdots x_{\ell(x)}$，与之关联的柱集为

$$\Gamma_x = \{\omega \in \Omega : \omega^{\ell(x)} = x\}. \tag{6}$$

柱集生成 Ω 上的 Borel σ–代数。

我们通过标准双射将有限词与非负整数对应起来

$$x \;\mapsto\; 2^{\ell(x)} - 1 + \sum_{i=1}^{\ell(x)} x_i\, 2^{\ell(x)-i}. \tag{7}$$

若不会引起混淆，同一符号既表示一个字，也表示其对应的整数。这一约定与早先在（Vidakovic，1983）中的用法一致。

全文中，我们使用忽略加性常数的标准支配关系。对于定义在同一域上的函数 $a(\cdot)$ 和 $b(\cdot)$，我们写作

$$a \preceq b \tag{8}$$

若存在常数 C，使得对所有相关的 u，均有 $a(u) \le b(u) + C$。反向关系记作 $a \succeq b$，而 $a \asymp b$ 表示相互支配。下文几乎一律以 $\preceq$，$\succeq$，$\asymp$的形式陈述比较关系，并避免将其改写为显式不等式。这些关系在算法信息论中频繁出现，并在（Downey与Hirschfeldt，2010；Li & Vit´anyi，2008；Nies，2009；Vidakovic，1983）中得到明确使用。

令 $\mathcal{P}$ 表示作用于有限二进制字符串的所有部分递归函数类。对于 $F \in \mathcal{P}$，有限词 x 相对于 F的Kolmogorov复杂性为

$$K_F(x) = \begin{cases} \min\{\ell(p) : F(p) = x\}, & \text{if such } p \text{ exists}, \\ \infty, & \text{otherwise.} \end{cases} \tag{9}$$

存在一个通用的 $F_0 \in \mathcal{P}$，使得对于所有 $G \in \mathcal{P}$，

$$K(x) \equiv K_{F_0}(x) \preceq K_G(x). \tag{10}$$

我们将 $K(x)$ 解释为以比特为单位的描述长度。

（Downey与Hirschfeldt，2010；Li与Vit´anyi，2008；Nies，2009）的专著全面介绍了柯尔莫哥洛夫复杂度及其核心性质。

设 μ 为 Ω 上的概率测度。定义在有限词上的函数 V 称为相对于构造性测度 μ 的 Martin L¨of 检验，如果存在递归函数$g(m) \downarrow 0$，使得对每个整数 $m \geq 1$，

$$\mu\Big(\big\{\omega \in \Omega : \sup_{r\geq 1} V(\omega^r) \geq m\big\}\Big) \leq g(m). \tag{11}$$

存在一个通用的 Martin L¨of 检验 U，满足对所有检验 V，$U(x) \preceq V(x)$（Martin-L¨of, 1966）。

一个序列 $\omega \in \Omega$ 是 相对于 μ如果它是 ***Martin L¨of*** 随机的

$$\forall \text{ Martin Löf tests } V, \qquad \sup_{r\geq 1} V(\omega^r) < \infty. \tag{12}$$

等价地，如果 U 是通用的，那么

$$\omega \text{ is Martin Löf random w.r.t. } \mu \quad \Longleftrightarrow \quad \sup_{r\geq 1} U(\omega^r) < \infty. \tag{13}$$

若 x 是有限二进制词，我们用 $w(x)$表示其重量，即 x中 1 的数量。定义

$$p(x) = \frac{w(x)}{\ell(x)}, \qquad q(x) = 1 - p(x), \qquad \nu_{\ell(x)} = (p(x), q(x)). \tag{14}$$

若 $\omega \in \Omega$，则 $w(\omega^m)$表示前缀 ω^m 中 1 的数量，并且

$$p_m(\omega) = \frac{w(\omega^m)}{m}, \qquad \nu_m(\omega) = (p_m(\omega), 1 - p_m(\omega)). \tag{15}$$

我们使用以每个符号的比特数为单位的二元熵，

$$H(\nu_{\ell(x)}) = -p(x)\log_2 p(x) - q(x)\log_2 q(x), \tag{16}$$

对于前缀，也类似地定义 $H(\nu_m(\omega))$。量 $\ell(x)H(\nu_{\ell(x)})$表示与 x具有相同经验分布的词所构成的类型类（壳层）的对数大小，忽略低阶项。这定量地说明了为何 $\ell(x)H(\nu_{\ell(x)})$ 是与 $K(x)$ 比较时的正确尺度；这也正是比值 $R(x)$ 在 (2) 中采用的尺度。

由于 $K(x)$ 和 $\ell(x)H(\nu_{\ell(x)})$ 的计量单位都是比特，因此诸如 $K(x)/(\ell(x)H(\nu_{\ell(x)}))$ 这样的比值在量纲上是一致的。改变对数的底会以相同的常数缩放这两个量，因此不会影响以 $\preceq$，$\succeq$，$\asymp$ 表述的比较。

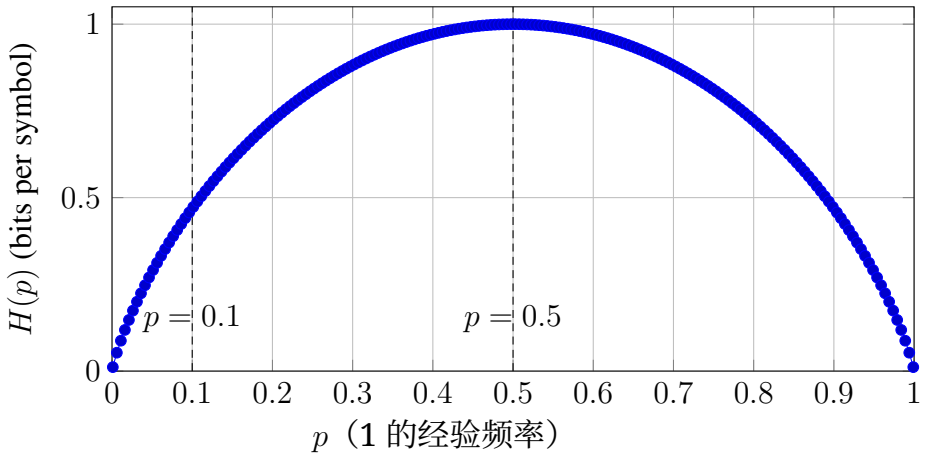


图1：二元经验熵 $H(p)$ 随符号不平衡程度的变化。经验频率为 p的有限词，其组合基线为 $mH(p)$，而非 m。

3.1 可计算测度、半测度与通用先验

主定理要求 Ω 上的底层测度满足温和的正则性条件。我们采用 Vidakovic（1981、1984、1988年）提出的可计算测度和半可计算测度框架。

代数由柱集生成的概率半测度 μ 称为可计算的，如果存在一个递归过程，给定一个有限词 x，即可将 $\mu(\Gamma_x)$近似到任意精度。半可计算测度具有有效的有理数上下界，且其收敛速率可计算。具有可计算参数的伯努利测度和一般独立同分布测度都属于此类。

作用于 Ω 且将可计算测度映射为可计算测度的变换，会在半可计算测度上诱导出一种自然的偏序，并引出通用先验的概念。粗略地说，在 Ω 上所有半可计算测度中，通用先验在相差一个乘法常数的意义下支配其他所有测度，因此，它是在与有效近似性相容的前提下信息量最低的选择。本文中，经验熵归一化在有限词层面发挥着类似作用：它剔除符号不平衡的组合贡献，留下由柯尔莫哥洛夫复杂度刻画的剩余描述成分。第5节将再次运用这一视角，以通用编码和最小描述长度（MDL）的语言来解释 $R(x)$。

最后，我们指出，基于过程的复杂性和信息版本为理解随机性与亏损提供了另一种视角。具体而言，Vidakovic（1983）引入了过程条件复杂度，并通过条件复杂度之差定义了信息量。这种基于复杂度差的视角，与第6节和第8节随后引入的熵亏损量自然契合。

4 调整后的复杂度 $KA(x)$

本节将形式化定义调整后的尺度 $KA(x)$ 和比率 $R(x)$，并明确阐述促使我们进行归一化的基于壳层的直觉。关键的技术手段是一个壳层计数陈述（步骤 3），它在步骤 4 中被转化为 Martin L¨of 检验；这正是我们稍后将再次采用的逻辑模式：用可计算描述长度替代不可计算的 $K(x)$，并将由此得到的亏损转化为可实施的统计量（第6节和第8节）。主要关注渐近定理的读者可以略读步骤 0–3，直接阅读第4节中的定理陈述。

定义与归一化原理

将熵归一化柯尔莫哥洛夫复杂度定义为

$$KA(x) = \frac{K(x)}{H(\nu_{\ell(x)})}. \tag{17}$$

$KA(x)$ 在 $H(\nu_{\ell(x)}) = 0$ 时未定义，也就是说，对于常值字而言未定义。对于非常值字，它是有限的，并提供一种无尺度的描述复杂度测度。实际上，常值字可以单独处理：它们的不平衡程度最大，而且可以轻易压缩。有意义的情形是 $0 < p(x) < 1$，此时经验熵提供了一个有意义的组合基线。

两种标准基线为这种归一化提供了依据。首先，对于固定长度 $\ell(x) = n$ 和固定重量 $w(x) = k$，壳层

$$S(n,k) = \{x \in \{0,1\}^n : w(x) = k\}$$

其基数为 $\binom{n}{k}$，因此壳层典型索引成本满足

$$\log_2 \binom{n}{k} = nH(k/n) + O(\log n),$$

其中，$O(\log n)$ 项对应不可避免的自定界开销。其次，对于任意长度为 n 的词 x，都存在一种一致编码方案：给定 n 后，先用 $O(\log n)$ 比特描述其重量 $w(x)$，然后在壳层 $S(n, w(x))$ 中为 x 编制索引。因此，

$$K(x \mid n) \preceq \log_2 \binom{n}{w(x)} + \log n. \tag{18}$$

因此，量 $H(\nu_{\ell(x)})$ 提供了合适的逐符号组合基准，而调整后的复杂度 $KA(x) = K(x)/H(\nu_{\ell(x)})$ 则根据经验熵壳层的有效大小对柯尔莫哥洛夫复杂度进行重新缩放。

定义无量纲归一化比率很有用

$$R(x) = \frac{K(x)}{\ell(x)\, H(\nu_{\ell(x)})}, \tag{19}$$

每当 $H(\nu_{\ell(x)}) > 0$。对于正则测度下的随机对象， $R(x)$应接近 1。显著低于 1 的值表明，除符号不平衡之外还存在其他结构。稍后，在第6节和第8节中，我们将再次讨论 $R(x)$，并说明如何利用 $K(x)$的可计算上界得到有效比率，再将其校准为壳内检验。

可交换测度与德菲内蒂约化

设 ω 是关于 Ω 上某个构造性可交换测度 μ 的 Martin L¨随机。可交换性意味着 μ是独立同分布伯努利测度的混合（de Finetti，1937年；Diaconis 和 Freedman，1980年），对于 μ典型序列，1 的经验频率收敛到某个极限 $p \in [0,1]$。在构造性情形下，我们假设（作为主定理中所用正则性假设的一部分），关于 μ 的 Martin L¨随机性可有效约化到一个伯努利分量；也就是说，存在参数 p ，使得 ω 关于伯努利测度 μ_p是 Martin L¨随机。这正是可交换性在论证中发挥作用之处，也说明了为何正确的归一化涉及单符号经验熵 $H(\nu_m(\omega))$。

定理

对于具有极限频率 p的序列 ω ，

$$\lim_{m\to\infty} \frac{KA(\omega^m)}{m} = 1, \qquad \text{equivalently} \qquad \lim_{m\to\infty} \frac{K(\omega^m)}{m} = H(p). \tag{20}$$

证明的写法突出了壳缺陷量，因为当 $K(\cdot)$被可计算编码长度取代后，同一亏损随后会再次作为有效检验统计量出现（第 6 节和第 8 节）。

证明

记 $k_m = w(\omega^m)$和 $p_m = k_m/m$。根据定义， $\nu_m(\omega) = (p_m,\ 1-p_m)$。由于 ω 在正则可交换 μ下是 Martin L¨ 随机的，经验频率收敛，

$$p_m \to p, \tag{21}$$

因此，由二元熵的连续性可得

$$H(\nu_m(\omega)) \to H(p). \tag{22}$$

因此,

$$\frac{KA(\omega^m)}{m} = \frac{K(\omega^m)}{m\,H(\nu_m(\omega))} \tag{23}$$

当且仅当其收敛到 1

$$\frac{K(\omega^m)}{m} \to H(p). \tag{24}$$

我们通过比较 $K(\omega^m)$ 与壳基准 $\log_2 \binom{m}{k_m}$，并证明逐符号差值趋于零，来证明 (24)。论证分为两部分：通过壳层编码对 $K(\omega^m)$ 给出一个构造性的上界（步骤 2），以及通过计数引理和 Martin L¨of 检验构造获得亏损控制（步骤 3–4）。

S步骤 0：壳层与基准。对于整数 $(m,\ k)$，定义壳层（类型类）

$$S(m,k) = \{x \in \{0,1\}^m : w(x) = k\}, \tag{25}$$

因此 $|S(m,\ k)| = \binom{m}{k}$。量 $\log_2 \binom{m}{k}$ 是对$S(m,\ k)$ 中的词进行最优索引所需的成本，另有加性 $O(\log\ m)$ 开销。

步骤 1：斯特林近似给出正确的壳层速率。令 $k = k_m$ 且 $p_m = k_m/m$。标准斯特林界表明

$$\log_2 \binom{m}{k} = mH(p_m) + O(\log m), \tag{26}$$

其中，余项在 k的中心范围内是显式且一致的。由于$p_m \to p$，我们得到

$$\frac{1}{m} \log_2 \binom{m}{k_m} \to H(p). \tag{27}$$

这是 $K(\omega^m)/m$的目标速率。

步骤 2：通过壳层编码得到 $K(\omega^m)$ 的显式上界。固定 m 和 $k = k_m$。 $S(m,\ k)$与 $\{0, 1, \ldots, \binom{m}{k} - 1\}$ 之间存在一个由字典序索引给出的可计算双射。因此，存在一种前缀无歧义码，可通过以下方式描述任意 $x \in S(m,\ k)$：(i) 对$(m,\ k)$的自定界描述，以及 (ii) x 在 $S(m,\ k)$中的索引。这给出了一个一致的界

$$K(x) \preceq \log_2 \binom{m}{k} + O(\log m), \tag{28}$$

对于所有长度为 m、重量为 k 的词 x。将此应用于 $x = \omega^m$，可得

$$K(\omega^m) \preceq \log_2 \binom{m}{k_m} + O(\log m). \tag{29}$$

除以 m 并使用 (27)，得到

$$\limsup_{m\to\infty} \frac{K(\omega^m)}{m} \le H(p). \tag{30}$$

步骤 3：计数引理：壳层内亏损较大的情况很少。定义一个字词的壳层亏 x 长度为 m 且重量为 k 时的损

$$d(m,k;x) = \log_2 \binom{m}{k} - K(x). \tag{31}$$

固定 (m,k) 和一个整数 $t \ge 1$，并考虑集合

$$A_{m,k}(t) = \{x \in S(m,k) : d(m,k;x) \ge t\}. \tag{32}$$

Then

$$|A_{m,k}(t)| \le 2^{-t}\binom{m}{k}. \tag{33}$$

事实上，条件 $d(m,k;x) \ge t$ 等价于 $K(x) \le \log_2 \binom{m}{k} - t$。长度至多为 $\log_2 \binom{m}{k} - t$ 的程序不超过 $2^{\log_2 \binom{m}{k} - t}$ 个，因此不同的输出也至多有这么多个。因此，壳层中亏损至少为 t 的字词所占比例至多为 2^{-t}。

步骤 4：将计数引理转化为马丁-洛夫¨检验。对于每个 和每个，定义前缀长度为 且亏损至少为 的序列集合：

$$B_m(t) = \{\omega \in \Omega : d(m, w(\omega^m); \omega^m) \ge t\}. \tag{34}$$

现在我们来界定伯努利测度 μ_p 下 $B_m(t)$ 的概率。对于任意长度为 m 且重量为 k 的字词 x，有

$$\mu_p(\Gamma_x) = p^k(1-p)^{m-k}. \tag{35}$$

H度，对 k 求和，然后再对每个壳层内的非典型集合 $A_{m,\,k}(t)$ 求和，

$$\begin{aligned}
\mu_p(B_m(t)) &= \sum_{k=0}^{m} \sum_{x\in A_{m,k}(t)} \mu_p(\Gamma_x) \\
&= \sum_{k=0}^{m} |A_{m,k}(t)|\, p^k(1-p)^{m-k} \\
&\le \sum_{k=0}^{m} 2^{-t}\binom{m}{k} p^k(1-p)^{m-k} \\
&= 2^{-t}(p+(1-p))^m \\
&= 2^{-t}.
\end{aligned} \tag{36}$$

现在，通过以可计算的方式对 m 加权来定义一个全局检验，例如

$$\mathcal{U}_t = \bigcup_{m\geq 1} B_m(t + 2\log_2(m+1)). \tag{37}$$

将之前的界中的 t 替换为 $t + 2\log_2(m+1)$，得到

$$\begin{aligned}\mu_p(\mathcal{U}_t) &\leq \sum_{m\geq 1} \mu_p\Big(B_m(t + 2\log_2(m+1))\Big) \\ &\leq \sum_{m\geq 1} 2^{-(t+2\log_2(m+1))} \\ &= 2^{-t}\sum_{m\geq 1}(m+1)^{-2} \;<\; c\,2^{-t},\end{aligned} \tag{38}$$

其中 c 为绝对常数。族 $\{\mathcal{U}_t\}_{t\geq 1}$ 是一致递归可枚举的（因为 $K(x)$ 是上半可计算的，因此谓词 $d(m,\ k; x) \geq r$ 是可有效枚举的），因而构成关于 μ_p 的 Martin L¨of 检验。

由于 ω 关于 μ_p 是 Martin L¨of 随机的，因此它避开了这个检验。因此存在一个常数 t_0，使得对于所有充分大的 m，

$$d(m, k_m; \omega^m) < t_0 + 2\log_2(m+1). \tag{39}$$

等价地，

$$\log_2 \binom{m}{k_m} - K(\omega^m) = O(\log m). \tag{40}$$

两边除以 m 得

$$\frac{1}{m}\Big(\log_2 \binom{m}{k_m} - K(\omega^m)\Big) \to 0. \tag{41}$$

步骤 5：结论。 合并 (27) 和 (41)：

$$\begin{aligned}\frac{K(\omega^m)}{m} &= \frac{1}{m}\log_2 \binom{m}{k_m} - \frac{1}{m}\Big(\log_2 \binom{m}{k_m} - K(\omega^m)\Big) \\ &\to H(p) - 0 = H(p).\end{aligned} \tag{42}$$

因此，根据（22），

$$\frac{KA(\omega^m)}{m} = \frac{K(\omega^m)}{m\,H(\nu_m(\omega))} \to 1, \tag{43}$$

如所述。

注。 论证的核心是亏损控制：在每个壳层内，具有较大亏损 $\log_2 \binom{m}{k} - K(x)$ 的词构成一个指数级微小的子集，而 Martin L¨of 随机性排除了出现无穷多个大亏损的可能性。这种复杂度之差的逻辑与（Vidakovic，1983）中发展出的基于过程的信息测度相似；在那里，信息通过条件过程复杂度之差来定义，并与序列的随机性性质相关联。在后续各节中，我们实际上通过用可计算描述长度替换 $K(x)$ 来重用这一逻辑，从而将亏损转化为检验统计量；具体请参见第6节中的有效比率以及第8节中汇总的检验规则。

4.1 非遍历测度与病态测度

对底层测度 μ 施加的正则性假设，对于极限定理的有效性和可解释性至关重要。关键在于，定理中使用的基线是单符号经验熵 $H(\nu_m(\omega))$，它捕捉边缘符号频率，却忽略了高阶约束。如果生成机制在保持边缘分布表面上均衡的同时施加强依赖，那么即使可容许字符串的真实组合增长率较小，$H(\nu_m(\omega))$ 仍可能接近 1。这正是需要采用基于经验块熵或条件熵的高阶归一化的情形。稍后在第7节引入条件经验基准时，我们也会再次提到这一点。

为具体说明这一点，考虑一个以相同概率输出 00、01 和 11 这几个块，且从不输出块 10 的过程。更精确地说，在 Ω 上定义一个测度μ，其支撑集为满足以下条件的序列 ω：对每个 $k \geq 1$，长度为 2 的块 $(\omega_{2k-1},\ \omega_{2k})$ 都属于 $\{00, 01, 11\}$。该测度受到很强的约束，既不可交换，也不是独立同分布的。

对于 μ–典型序列，1 的边际频率仍为 $1/2$，因为每个块中 1 的期望数量等于 1，因此经验分布满足

$$\nu_m(\omega) \to (1/2, 1/2), \qquad H(\nu_m(\omega)) \to 1. \tag{44}$$

然而，长度为 m 的可容许词的数量为 $3^{m/2}$ 量级（相差一个与奇偶性有关的因子），因为每个不相交对都有 3 种选择。因此，支撑集的真实组合增长率为

$$\frac{1}{m} \log_2(3^{m/2}) = \frac{1}{2} \log_2 3 < 1. \tag{45}$$

对于 μ–随机序列 ω（相对于这一受约束块测度而言是随机的），ω^m 的柯尔莫哥洛夫复杂度与可容许前缀的数量的对数相对应，因此可以预期

$$\frac{K(\omega^m)}{m} \to \frac{1}{2} \log_2 3, \tag{46}$$

同时， H（ν_m（ω））$\to 1$。因此，

$$\frac{K(\omega^m)}{m\,H(\nu_m(\omega))} \to \frac{(1/2)\log_2 3}{1} = \frac{1}{2}\log_2 3, \tag{47}$$

严格小于 1。

这个例子表明，当底层测度具有无法从单符号经验分布中看出的依赖结构时，熵归一化的极限定理可能会失效。它也阐明了正则性假设的作用：在可交换独立同分布混合情形中，单比特经验熵是恰当的基线；而在受约束或非遍历情形中，为了获得稳定且易于解释的归一化，可能有必要用块熵、经验条件熵或其他高阶经验摘要来替代 $H(\nu_m)$。因此，第7节后文提出的条件归一化并非仅仅是形式上的扩展，而是针对这种隐性依赖的实际解决方案。

5 算法解释与复杂性理论变体

熵调整复杂度 $KA(x)$ 通过剔除由符号不平衡导致的纯组合贡献，揭示内在的算法结构。对于由正则可构造测度生成的 Martin L¨ 随机序列，归一化比值

$$\frac{KA(\omega^m)}{m} = \frac{K(\omega^m)}{m\,H(\nu_m(\omega))} \tag{48}$$

如第4节所证，其收敛于1。这种一致性表明，Kolmogorov复杂度达到了由观测到的符号频率所决定的最优经验熵率。本节从几个互补的角度阐释这一事实，并说明熵归一化如何自然地融入现有的算法与复杂性理论框架。

通用先验与壳内归一化

在Ω上的半可计算测度类中，通用半测度在相差乘法常数的意义下支配所有其他测度，因此会给Kolmogorov复杂度较低的对象赋予相对较大的概率质量（Vidakovic，1988年）。从这一角度看，熵归一化可视为通用先验原则的局部类比，仅适用于由$\nu_{\ell(x)}$确定的固定经验壳层。

以壳层为条件，可以消除与符号不平衡相关的组合贡献。经验熵 $H(\nu_{\ell(x)})$ 决定了该壳层的对数大小，而 $K(x)$ 则衡量了相对于通用参考机的描述成本。因此

$KA(x)$ 量化了在与其经验分布一致的最不具信息性的先验下，x 相对于典型壳元素的算法偏差。这一视角明确说明了为何形如

$$\ell(x)H(\nu_{\ell(x)}) - K(x) \tag{49}$$

在主定理的证明（第4节）以及后文提出的有效检验程序（第8节）中都发挥着核心作用。

测度敏感的随机性概念

熵归一化与更精细的算法随机性概念相兼容，这些概念明确纳入了测度结构。基于过程的复杂度以及 Schnorr 风格的表述，会根据可计算概率测度或生成机制对描述进行条件化（Vidakovic，1988年）。对于固定的经验熵，调整后的复杂度 $KA(x)$ 表现为 $K(x)$ 的重新缩放版本，并且在加性常数范围内保留所有顺序比较；本文通篇使用第3节引入的支配关系 $\preceq$ 来表达这一点。

在无限序列的层面上，$KA(\omega^m)$ 的渐近行为与经典复杂度率定理相呼应；在这些定理中，底层测度的熵率决定了算法复杂度的增长率。与此同时，第4.1节中的例子表明，对于依赖过程或受约束过程，单符号经验熵可能无法作为合适的基线。这一观察促使我们在第7节提出条件归一化，其中经验联合结构取代边缘符号频率，成为恰当的组合参照。

资源受限变体

从计算角度来看，自然会考虑调整后的复杂度的资源受限变体。令 $K^t(x)$ 表示时间有界柯尔莫哥洛夫复杂度，其中程序必须在时间 $t(\ell(x))$ 内停机。于是可以定义

$$KA^t(x) = \frac{K^t(x)}{H(\nu_{\ell(x)})}, \qquad R^t(x) = \frac{K^t(x)}{\ell(x)\,H(\nu_{\ell(x)})}. \tag{50}$$

在固定时间界下，这些量是全递归的，因此可计算。它们在理想算法描述与计算约束下可行的编码之间，搭起了一座有原则依据的桥梁。第8节将明确阐述这座桥梁：其中，$K(x)$ 被替换为具体的可计算替代量 $K_{\text{eff}}(x)$，所得的归一化比率则用作操作性检验统计量。

与最小描述长度和通用编码的联系

熵归一化也与统计模型选择和通用编码的思想自然契合。在最小描述长度框架中，模型复杂性与数据拟合度通过两部分编码取得平衡（Barron & Clarke，1990年；Barron, Rissanen, & Yu，1998年；Rissanen，1978年）。在本文的设定中，量 $\ell(x)H(\nu_{\ell(x)})$ 表示与经验模型相关的随机码长，而 $K(x)$ 则体现结构描述长度。二者的比率为，

$$R(x) = \frac{K(x)}{\ell(x)\, H(\nu_{\ell(x)})}, \tag{51}$$

衡量符号频率本身所能揭示的信息之外的算法结构。

对于由平稳信源生成的典型序列，算术编码等通用编码方案可实现接近$\ell(x)H(\nu_{\ell(x)})$的码长。第4节的主定理表明，对于正则测度下的马丁-洛夫随机序列，只要考虑有限前缀，柯尔莫哥洛夫复杂度本身就与这一经验基线相符，误差至多为$O(\log\ m)$。因此，

$$R(\omega^m) \to 1, \tag{52}$$

这为将熵归一化复杂度作为评估有限二进制数据中内在随机性与结构的适当尺度提供了算法层面的依据。

6 调整后复杂度的统计、应用与有效用途

一旦将第4节的抽象壳层论证重新应用于有限数据，熵归一化复杂度框架便兼具概念性和实践性解释。一个核心汇总统计量是归一化随机性得分

$$R(x) = \frac{K(x)}{nH(\nu_n)}, \qquad n = \ell(x), \tag{53}$$

该得分将一个词的描述成本与其经验分布所隐含的组合基线进行比较。 $R(x)$的值接近1，表明 x 相对于其经验壳层而言在算法上具有典型性；而显著低于1的值则表明，除了符号频率本身所能解释的因素之外，还存在算法结构。当不同样本中1的比例差异很大时，这种归一化在应用中至关重要，因为它可以防止将稀疏性或不平衡误解为正则性。

抽象计数引理（33）可以先被解读为一种原型统计结论：在固定的经验壳层内，大缺陷出现的概率呈指数级稀少。本节将展示，当这一原理与柯尔莫哥洛夫复杂度的有效近似相结合时，它如何成为一种操作性工具。

潜在应用包括随机性检验、基于复杂度的假设检验、符号序列结构评估、生物数据与基因组数据分析，以及对阈值化过程（如小波系数选择）所产生的二进制模式进行评估。在所有这些场景中，熵归一化确保比较是相对于恰当的组合基线进行，而不是依据原始字长。

壳层与典型性再探

设 $x \in \{0,1\}^n$ 是一个有限二进制词，其 $n = \ell(x)$，重量为 $w(x) = k$。相应的壳层（类型类）

$$S(n,k) = \{x \in \{0,1\}^n : w(x) = k\} \tag{54}$$

恰好包含 $\binom{n}{k}$ 个词，它们具有相同的经验分布 $\nu_n = (p, 1-p)$，且满足 $p = k/n$。因此，

$$\log_2 |S(n,k)| = \log_2 \binom{n}{k} = nH(p) + O(\log n), \tag{55}$$

这是有限词版本的斯特林近似（26），该近似用于主定理的证明中。

若一个词 $x \in S(n,k)$ 的相对于 (n,k) 的条件柯尔莫哥洛夫复杂度达到壳层所允许的最大阶，则称其为**壳层典型**。

$$K(x \mid n,k) \asymp \log_2 \binom{n}{k}. \tag{56}$$

标准的计数论证表明，复杂度显著较低的壳层非典型字只占 $S(n,k)$ 中极小的一部分；这正是（33）中抽象量化的现象。

归一化为何重要：平衡壳层与不平衡壳层

通过比较具有不同经验分布的壳层典型词，可以说明熵归一化的作用。考虑两个长度为 $n = 1000$ 的词。设 x 恰好包含 500 个 1，并以壳层典型的方式排列，使得 $p_x = 0.5$ 和 $H(\nu_x) = 1$。设 y 恰好包含 100 个 1，也以壳层典型的方式排列，使得 $p_y = 0.1$ 和

$$H(\nu_y) = -0.1 \log_2(0.1) - 0.9 \log_2(0.9) \approx 0.469. \tag{57}$$

壳层典型性意味着

$$K(x) \approx nH(\nu_x) = 1000, \qquad K(y) \approx nH(\nu_y) \approx 469. \tag{58}$$

尽管绝对复杂度存在显著差异，但调整后的尺度一致：

$$KA(x) \approx 1000, \qquad KA(y) \approx 1000. \tag{59}$$

该归一化消除了不平衡的平凡组合效应，并分离出内在的描述复杂度。

与结构化词的对比同样具有启发性。设 $z = 010101\ldots01$ 是长度为 $n = 1000$ 的交替词。那么， $p_z = 0.5$ 和 $H(\nu_z) = 1$，但由于一个简短程序即可生成该模式，$K(z)$ 的量级为 $\log_2 n$ 。因此， $KA(z)$ 的量级为 $\log_2 n$，远低于 n。这正是预期的行为： z 与随机平衡词具有相同的经验分布，却表现出低得多的算法复杂度。

有效替代量与操作统计量

柯尔莫哥洛夫复杂度不可计算，因此 $KA(x)$ 和 $R(x)$ 都无法被精确求得。有效替代量是通过将 $K(x)$ 替换为可计算上界而得到的。由于 $K(x)$ 是上半可计算的，存在一个可计算序列 $K_t(x)$，当 $t \to \infty$ 时，该序列递减并趋近于 $K(x)$。对于固定的 t，定义

$$KA_t(x) = \frac{K_t(x)}{H(\nu_n)}, \qquad R_t(x) = \frac{K_t(x)}{nH(\nu_n)}. \tag{60}$$

通过固定一个可计算的无损编码 C 并设定 $K_C(x) = L_C(x)$，即可得到一种完全有效的替代方案，即 x 在 C下的码长。该操作统计量

$$R_C(x) = \frac{L_C(x)}{nH(\nu_n)} \tag{61}$$

是可计算的，并且具有直接的解释。编码定理保证

$$K(x) \preceq L_C(x), \tag{62}$$

因此， $R_C(x)$ 为 $R(x)$ 提供了一个有原则的上界替代量。

两个明确的有效界尤其具有启发性。字面编码给出 s

$$K_{\text{len}}(x) = n, \tag{63}$$

而组合壳层编码对 $(n,\ w(x))$ 进行编码，并在 $S(n,\ w(x))$ 内对 x 建立索引，得到

$$K_{\text{comb}}(x) = \log_2 \binom{n}{w(x)} + O(\log n). \tag{64}$$

给定任意这样的可计算上界 $K_{\text{eff}}(x)$，定义有效归一化复杂度

$$R_{\text{eff}}(x) = \frac{K_{\text{eff}}(x)}{nH(\nu_n)}. \tag{65}$$

基于亏损的检验

使用熵亏损而非比值通常更方便。定义

$$d_A(x) = nH(\nu_n) - K(x), \tag{66}$$

及其有效类比

$$d_A^C(x) = nH(\nu_n) - L_C(x). \tag{67}$$

固定 $m \geq 1$，并在 $d_A^C(x) \geq m$ 时拒绝其随机性，等价地，当

$$R_C(x) \leq 1 - \frac{m}{nH(\nu_n)}. \tag{68}$$

这一规则是主定理证明中所用亏损逻辑的实际体现：将观测到的描述长度与壳基准进行比较，并标记差距异常大的情形。这一构造与算法统计学中的随机性亏损密切相关；在算法统计学中，典型性由描述长度与基于熵的基准之间的差距来量化。当基准取为经验熵 $nH(\nu_n)$，即对应经验熵壳层的大小时，$d_A(x)$ 与 Vereshchagin-Vit´anyi-Shen 随机性亏损框架在对数项范围内一致（Vereshcha- gin & Shen, 2015; Vereshchagin & Vit´anyi, 2004）。

有限示例说明

考虑二进制词

$$x = 01010001001000001010000100000100001, \tag{69}$$

长度为 $n = 35$，包含 $w(x)$ 个 $= 91$。经验熵为

$$H(\nu_n) \approx 0.821, \qquad nH(\nu_n) \approx 28.7 \text{ bits}. \tag{70}$$

字面编码产生 $R_{\text{len}}(x)$ 个 ≈ 1.22，而组合壳层编码产生 $R_{\text{comb}}(x)$ 个 $\approx 1。09$。选择 $m = 5$，得到阈值 $c(5) \approx 0.83$，因此在这种粗略的有效编码方案下，不拒绝随机性。

熵归一化确保该检验以恰当的组合基线为参照来比较词串，并检测真正的算法结构，而非平凡的符号失衡。下一节将这一思路扩展为以条件经验基线和联合经验基线取代单符号经验熵，从而能够系统地纳入辅助信息和依赖结构。

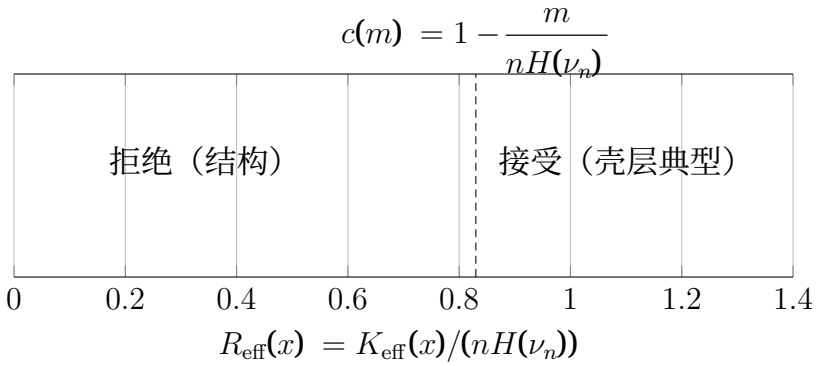


图2：以归一化比值 $R_{\text{eff}}(x)$ 表示的玩具有效随机性检验。对于固定的亏损阈值 m，只要 $R_{\text{eff}}(x) \le c(m) = 1 - m/(nH(\nu_n))$，就拒绝随机性。

7 条件与互调整复杂度

熵归一化框架自然地扩展到条件情形和互信息情形，提供了一种有原则的方法，能够在存在辅助信息时剔除经验联合结构的影响。与无条件情形一样，归一化旨在将纯粹的组合效应与真正的算法依赖区分开来。从概念上说，本节可视为对第4.1节中病态示例所引出问题的回答：如果单符号边缘分布未能反映真实约束，就必须扩展基线所采用的经验摘要。

设 $x, y \in \{0,1\}^n$ 为长度均为 n 的有限二进制词。定义经验联合频率如下：

$$p_{ab} = \frac{1}{n} \#\{i : x_i = a,\ y_i = b\}, \qquad a, b \in \{0,1\}. \tag{71}$$

令 $p_{\cdot b} = \sum_a p_{ab}$ 且 $p_{a|b} = p_{ab}/p_{\cdot b}$，其中 $p_{\cdot b} > 0$。则经验条件熵为

$$H_{\text{emp}}(X \mid Y) = \sum_b p_{\cdot b} \Big(- \sum_a p_{a|b} \log_2 p_{a|b} \Big). \tag{72}$$

我们定义条件调整复杂度及其归一化比率如下：

$$KA(x \mid y) = \frac{K(x \mid y)}{H_{\text{emp}}(X \mid Y)}, \qquad R(x \mid y) = \frac{K(x \mid y)}{nH_{\text{emp}}(X \mid Y)}, \tag{73}$$

当 $H_{\text{emp}}(X \mid Y) > 0$ 时。相应的条件熵亏损

$$d_A(x \mid y) = nH_{\text{emp}}(X \mid Y) - K(x \mid y) \tag{74}$$

是 Levin 随机性亏损的一个特例（Levin，1974年）。

在此处及下文中，符号 X 和 Y 并不表示由底层随机过程生成的抽象随机变量，而是指由固定有限词 $x = (x_1, \ldots, x_n)$ 和 $y = (y_1, \ldots, y_n)$ 诱导的经验随机变量：从 $\{1, \ldots, n\}$ 中均匀抽取一个索引 i，并令 $X(i) = x_i$、$Y(i) = y_i$。因此，所有熵均根据由 (x, y) 确定的经验联合分布计算。这一约定在类型理论和经验分布中十分常见（Cover 与 Thomas，2006），并在将复杂性与熵率联系起来时被明确采用（Li 与 Vit´anyi，2008）。

有限对象之间的依赖性也可以通过算法互信息来研究。两个有限词 x 和 y 之间的算法互信息定义为（相差一个加性常数）

$$I(x : y) \asymp K(x) + K(y) - K(x, y), \tag{75}$$

并衡量 x 和 y 共享的算法信息量（G´acs、Tromp 与 Vit´anyi，2001；Kolmogorov，1965；Levin，1974年）。从经验角度看，依赖性反映在联合符号频率中。令 X 和 Y 表示由 (x_i, y_i) 诱导的经验随机变量，则经验互熵为

$$I_{\mathrm{emp}}(X; Y) = H_{\mathrm{emp}}(X) + H_{\mathrm{emp}}(Y) - H_{\mathrm{emp}}(X, Y). \tag{76}$$

我们将互调整复杂度定义为

$$KA(x : y) = \frac{I(x : y)}{I_{\mathrm{emp}}(X; Y)}, \tag{77}$$

只要 $I_{\mathrm{emp}}(X; Y) > 0$。这种归一化消除了依赖性中纯粹的统计成分，分离出经验相关性之外的算法依赖。当 x 和 y 是来自联合独立同分布源的典型样本时，$I(x : y)$ 集中在 $nI_{\mathrm{emp}}(X; Y)$ 附近，且互调整复杂度仍接近于一。相反，即使经验互熵很小，确定性或结构性关系也可能产生较大的 $I(x : y)$，从而导致 $KA(x : y)$ 取较大值。因此，$KA(x : y)$ 提供了一种与尺度无关的算法依赖度量，与调整后的复杂度在无条件情形和条件情形中的作用相对应。

现在，我们通过一个具体的有限示例来说明条件归一化，这与前面在第6节中讨论的无条件情形相对应。令

$$x = 01010001001000001010000100000100001 \tag{78}$$

为长度为 $n = 35$ 的同一个二进制词。和之前一样，$w(x) = 9$ 以及

$$H(\nu_n) \approx 0.82, \qquad nH(\nu_n) \approx 28.8 \text{ bits}, \tag{79}$$

因此，相对于其无条件壳层，x 除了符号不平衡外，不表现出显著的正则性。

通过令.....来引入辅助信息

$$y_0 = 00100011 \tag{80}$$

并将其周期性延拓，截断至长度 $n = 35$ 以获得一个词 y。构造坐标逐点配对（x_i， y_i），得到表1所示的经验联合计数。

$x\backslash y$	0	1	总计
0	19	7	26
1	6	3	9
总计	25	10	35

表1：（x， y）的经验联合计数。

由表1可得，经验条件概率为

$$P(X = 1 \mid Y = 0) = \frac{6}{25}, \qquad P(X = 1 \mid Y = 1) = \frac{3}{10}, \tag{81}$$

相应的条件熵为

$$H(X \mid Y = 0) \approx 0.79, \qquad H(X \mid Y = 1) \approx 0.88. \tag{82}$$

按 $P(Y = 0) = 25/35$ 和 $P(Y = 1) = 10/35$ 加权，得到

$$H_{\text{emp}}(X \mid Y) = \frac{25}{35} H(X \mid Y = 0) + \frac{10}{35} H(X \mid Y = 1) \approx 0.47. \tag{83}$$

因此，条件熵基线满足

$$nH_{\text{emp}}(X \mid Y) \approx 16.5 \text{ bits}, \qquad nH(\nu_n) \approx 28.8 \text{ bits}. \tag{84}$$

这一化简反映了这样一个事实：以 y 为条件会限制组合壳。坐标被划分为具有 $Y = 0$ 和 $Y = 1$ 的坐标；在每个子集中， x 的可行模式由相应的条件类型决定。因此，量 $nH_{\text{emp}}(X \mid Y)$ 表示与观测到的联合结构相容的条件壳中词的对数规模。

一种自然的有效条件描述长度，是在这个条件壳中对 x 进行编码，例如，分别对$Y = 0$ 和 $Y = 1$ 坐标中取值为 1 的位置进行索引。由此得到有效上界 $K_{\text{eff}}(x \mid y)$，以及相应的比率

$$R_{\text{eff}}(x \mid y) = \frac{K_{\text{eff}}(x \mid y)}{nH_{\text{emp}}(X \mid Y)} \approx 1.09. \tag{85}$$

定性结论与无条件情形相似。相对于由辅助信息 y所诱导的条件组合基线，字词 x 并未表现出显著的额外规律性。若 x 几乎可以由 y确定，那么 $H_{\text{emp}}(X \mid Y)$ 就会

接近于零，而条件壳代码将能够实现短得多的描述长度，使 $R_{\mathrm{eff}}(x \mid y)$ 大幅低于 1。因此，条件熵归一化提供了一种更精细的壳内比较，能够将真正的算法结构与辅助信息所施加的统计约束区分开来。下一节的条件亏损正是采用这一条件基线。

8 有效版本与检验

柯尔莫哥洛夫复杂度不可计算，因此既无法精确求得调整后的复杂度 $KA(x)$，也无法精确求得其条件变体和互信息变体。不过，将 $K(\cdot)$替换为可计算上界后，前几节建立的熵归一化框架自然得到具有实际意义的有效对应版本。这些有效版本保留了经验熵基线的作用，并直接导出可实现的程序，用于检验随机性和检测有限数据中的结构。

从逻辑角度看，本节是第4节所用证明策略的操作层面对应。在那里，利用每个组合壳层内的亏损构造了一个 Martin L¨of 检验，并证明了渐近极限定理。这里则以明确且可计算的描述长度重用同一亏损思想，将抽象论证转化为具体的统计工具。

设 $K_{\mathrm{eff}}(x)$ 是任意一个全递归函数，它在相差一个加法常数的意义下支配柯尔莫哥洛夫复杂度，即

$$K(x) \preceq K_{\mathrm{eff}}(x). \tag{86}$$

最简单的例子是字面编码，

$$K_{\mathrm{len}}(x) = \ell(x) = n, \tag{87}$$

它逐字输出 x 的比特，无需任何建模假设。虽然粗略，但它提供了一个通用基线。

一种信息量更大的有效上界利用了前文引入的壳结构。首先对 $n = \ell(x)$ 和重量 $w(x)$ 进行编码，然后在相应壳层的 $\binom{n}{w\,x()}$ 个字中对 x 编制索引。由此得到

$$K_{\mathrm{comb}}(x) = \log_2 \binom{n}{w(x)} + O(\log n), \tag{88}$$

这是 Stojanovic 和 Vidakovic（1987年）在有效组合复杂度的背景下明确研究过的一种形式。更一般地，任何固定的无损压缩方案 C 都会诱导出一个可计算描述长度 $L_C(x)$，满足

$$K(x) \preceq L_C(x), \tag{89}$$

其中支配常数仅取决于所选的编码。这类编码长度在最小描述长度理论和通用编码中发挥着核心作用（Gr¨unwald，2007年）。

给定所选的有效上界 $K_{\mathrm{eff}}(x)$，我们将有效归一化调整复杂度定义为

$$R_{\mathrm{eff}}(x) = \frac{K_{\mathrm{eff}}(x)}{nH(\nu_n)}, \qquad n = \ell(x), \tag{90}$$

只要 $H(\nu_n) > 0$。在条件和互信息情形中，同样的构造适用于第7节引入的经验基线 $nH_{\mathrm{emp}}(X \mid Y)$ 和 $nI_{\mathrm{emp}}(X;Y)$。在所有情况下，分母表示相关经验壳层的对数大小，而分子则是可计算描述长度。因此， $R_{\mathrm{eff}}(x)$ 将 x 的观测描述成本与其经验熵所预测的成本进行比较。对于壳层典型词以及效率合理的编码界， $K_{\mathrm{eff}}(x)$ 接近 $nH(\nu_n)$，且 $R_{\mathrm{eff}}(x)$ 接近一；而显著低于一的值则表明，除了符号不平衡之外还存在算法结构。

使用相应的熵亏损通常更为方便，

$$d_A^{\mathrm{eff}}(x) = nH(\nu_n) - K_{\mathrm{eff}}(x), \tag{91}$$

它是针对经验伯努利壳层的 Levin 型随机性亏损的有效类比。固定整数 $m \geq 1$，并考虑拒绝区域

$$\mathcal{R}_m = \{x : d_A^{\mathrm{eff}}(x) \geq m\}. \tag{92}$$

根据主定理证明中使用的同一计数论证，在所有长度为 n 且具有固定经验分布的字串中，至多有 $2^{nH(\nu_n)-m}$ 个属于 $\mathcal{R}_m$。因此，在与该经验分布一致的独立同分布伯努利测度下，拒绝概率至多为 2^{-m} 的某个倍数（常数取决于所选的编码界）。因此，参数 m 起到显著性水平的作用：当 $d_A^{\mathrm{eff}}(x) \geq m$ 时拒绝，可在不受符号比例失衡影响的情况下控制假阳性。这在有效层面上与第4节中 Martin–L¨of 检验构造所得的概率界（38）相呼应。

同一机制可直接扩展到条件调整复杂度和互相调整复杂度。一个可计算界 $K_{\mathrm{eff}}(x \mid y)$ 给出条件亏损

$$d_A^{\mathrm{eff}}(x \mid y) = nH_{\mathrm{emp}}(X \mid Y) - K_{\mathrm{eff}}(x \mid y), \tag{93}$$

从而可以检验 x 中的算法结构是否超出了侧信息 y所能解释的范围。同样地，用有效上界替代算法互信息 $I(x : y)$，便可形成实用的判据，用于检测超越经验相关性的算法依赖。

从最小描述长度的角度来看，量 $nH(\nu_n)$ 表示与经验模型相关的随机码长，而 $K_{\text{eff}}(x)$ 表示显式的结构描述长度。因此，熵归一化确保了经验拟合与描述性结构之间的清晰分离，这与两部分最小描述长度编码的思想一致（Barron 等人，1998年；Gr¨unwald，2007年）。相较于早期的有效组合测度（Stojanovic 和 Vidakovic，1987年），当前表述将经验熵置于归一化的核心，并将无条件、条件和互信息概念整合到一个统一且可操作的框架中。

9 讨论与结论

本文提出的熵归一化复杂度框架遵循一个简单的组织原则：应相对于恰当的经验组合基准来评估描述结构。柯尔莫哥洛夫复杂度衡量内在的算法描述长度，而经验熵则量化与观测到的符号频率相容的组合类规模。用经验熵对复杂度进行归一化，可以将这两种效应区分开来，并阐明在剔除平凡的不平衡后还剩下多少结构。

这一视角为香农熵、通用先验与算法复杂度之间的经典联系（Chaitin，1975；Solomonoff，1964a、1964b）提供了有限且局部的对应。具体而言，它明确指出，熵本身无法捕捉算法结构（Kramosil，1982）：经验熵相同的两个词，其描述复杂度可能截然不同；反过来，原始复杂度值也无法直接在不同经验分布之间进行比较。调整后的量 $KA(x)$ 和$R(x)$ 通过将描述长度置于恰当的经验尺度上，解决了这一不匹配问题。

自然地可以想到若干扩展和改进。本文目前的研究聚焦于二进制词和单符号经验熵，但这一归一化原理可以直接推广到更大的字母表。更重要的是，可以用块熵、经验条件熵或其他能够捕捉高阶依赖的概括量来替代$H(\nu_n)$。当符号级边缘分布无法反映生成机制的真实组合约束时，这类改进至关重要。第4.1节中的病态示例生动地说明了这一点：当强依赖性在边缘分布层面不可见时，就需要采用高阶经验基准。第7节提出的条件归一化可以视为朝这一方向迈出的第一步。

该框架也自然地契合算法统计学的思想。在这里，经验分布扮演粗略模型的角色，而与之相关的亏损衡量的是相对于该模型的残余算法不规则性。进一步的研究可以将经验模型、分块模型或条件模型纳入一个明确的模型类，并将经熵归一化的亏损解释为随机性指标，从而更直接地关联到最小充分统计量和随机性亏损的经典概念。条件构造与互构造已经展现出迈向这种综合的可能性。

计算方面的考量还引出了更多问题。第5节中介绍的调整后复杂度的资源受限变体，对于有效检验以及理解可行的编码约束如何影响结构检测尤为重要。第8节展示了可计算描述长度如何导出假阳性率受控的可实施检验，但在简单零模型下进行系统的有限样本校准和功效分析仍是尚待解决的问题。若能建立此类校准，将有助于进一步加强算法随机性理论与实际统计方法学之间的联系。

从应用角度看，在不平衡或稀疏性本身就是常态而非例外的情形下，熵归一化尤其自然。在多分辨率表示中，例如小波骨架、阈值化系数模式或稀疏符号编码，不平衡是默认状态。原始复杂度度量往往会混淆稀疏性与结构，而熵归一化复杂度则为跨尺度比较描述性内容提供了一种有原则的方法。因此，将当前框架扩展到多尺度或层次化情形，或可为分析结构化信号和符号数据提供实用工具。

最后，自然会想到量子类比。在 Schumacher 的量子数据压缩理论中，冯·诺依曼熵决定了量子系综的可压缩性（Schumacher，1995）。量子调整复杂度需要用态熵取代经典经验熵，并以密度算符而非符号频率为参照来衡量算法结构。这样的概念或许能够区分混合态或纠缠态中的结构性量子关联与随机噪声，从而有望进一步加强算法信息论、量子编码和量子统计推断之间的概念联系。

总之，熵归一化柯尔莫哥洛夫复杂度提供了一个灵活且统一的框架，将算法信息、经验熵和统计建模联系起来。通过明确剔除经验分布的组合贡献，它得到的量既具有理论可解释性，又具有实际操作意义。这一视角阐明了复杂性在随机性、依赖性和结构检测中的作用，并为进一步的理论发展和应用开辟了多种途径。

致谢。 作者感谢 H. O. Hartley讲席基金会以及德克萨斯农工大学美国国家科学基金会项目编号2515246提供的部分支持。

参考文献

Barron, A. R., & Clarke, B. M. (1990年). 贝叶斯方法的信息论渐近性质. 《*IEEE*信息论汇刊》, *36*(3), 453–471. doi: 10.1109/ 18.54892 Barron, A. R., Rissanen, J., & Yu, B. (1998年). 编码与建模中的最小描述长度原理. 《*IEEE*信息论汇刊》, *44*(6), 2743–2760.

Calude, C. S. (2002). 信息与随机性：算法视角（第2版）. 施普林格. Chaitin, G. J. (1975年). 一种与信息论形式上完全相同的程序大小理论. 《ACM期刊》, 22(3), 329–340. doi: 10.1145/321892.321894 Cover, T. M., & Thomas, J. A. (2006). 信息论要素（第2版）. 霍博肯, 新泽西州：威利. de Finetti, B. (1937年). La pr´evision：逻辑规律与主观来源. Annales de l’Institut Henri Poincar´e, 7(1), 1–68. 检索自 http://eudml.org/doc/ 79004（提出了可交换性，以及将无限可交换序列表为独立同分布律混合的表示。）Diaconis, P., & Freedman, D. (1980年). 有限可交换序列. 《概率年刊》, 8(4), 745–764.（de Finetti定理的有限样本版本，为有限可交换序列给出了界和近似。）Downey, R. G., & Hirschfeldt, D. R. (2010). 算法随机性与复杂性. 施普林格. G´acs, P., Tromp, J., & Vit´anyi, P. M. B. (2001). 算法统计学. 《IEEE信息论汇刊》, 47(6), 2443–2463. Gr¨unwald, P. D. (2007年). 最小描述长度原理. 剑桥, 马萨诸塞州：麻省理工学院出版社. Kolmogorov, A. N. (1965). 信息定量定义的三种方法. 《信息传输问题》, 1(1), 1–7. Kramosil, I. (1982). 关于信息与复杂性. 《控制论》, 18(6), 472–485. Levin, L. A. (1974年). 信息守恒（非增长）定律及概率论基础的若干方面. 《信息传输问题》, 10(3), 206–210. Li, M., & Vit´anyi, P. M. B. (2008). 柯尔莫哥洛夫复杂度及其应用导论（第3版）. 施普林格. Martin-L¨of, P. (1966). 随机序列的定义. 《信息与控制》, 9(6), 602–619. Nies, A. (2009). 可计算性与随机性. 牛津大学出版社. Rissanen, J. (1978). 以最短数据描述进行建模. 《自动化》, 14(5), 465–471. doi: 10.1016/0005-1098(78)90005-5 Schnorr, C.-P. (1971). 随机序列定义的统一方法. Mathematical Systems Theory, 5(3), 246–258. Schumacher, B. (1995). 量子编码. 《物理评论A辑》, 51(4), 2738–2747. doi: 10.1103/PhysRevA.51.2738 Solomonoff, R. J. (1964a). 归纳推理形式理论. 第一部分. 《信息与控制》, 7(1), 1–22. doi: 10.1016/S0019-9958(64)90223-2 Solomonoff, R. J. (1964b). 归纳推理形式理论. 第二部分. 《信息与控制》, 7(2), 224–254. doi: 10.1016/S0019-9958(64)90224-4 Stojanovic, S., & Vidakovic, B. (1987年). 二进制词组合复杂度度量的一些性质. Publications de l’Institut Math´ematique, 42(56),

143–147. Vereshchagin, N., & Shen, A. (2015). 重新审视算法统计学。arXiv:1504.04950[cs.IT].（预印本见 https://arxiv.org/abs/1504.04950）Vereshchagin, N., & Vit´anyi, P. (2004). Kolmogorov 的结构函数及其在模型选择基础理论中的应用。《*IEEE*信息论汇刊》，*50*(12), 3265–3290. Vidakovic, B. (1981). 复杂性与随机性（未发表的硕士论文）。贝尔格莱德大学。（硕士论文，塞尔维亚语，183页）Vidakovic, B. (1983). 信息量过程测度的一些特征。数学研究所出版物´学，*33*(47), 235–238. Vidakovic, B. (1984). 关于 Martin-L¨of 随机性测度对有限二进制词的一些性质。载于会议论文集：代数与逻辑（第 171–176 页）。克罗地亚萨格勒布。Vidakovic, B. (1988年). 关于算法复杂度、通用先验与奥卡姆剃刀。*Resenhas IME USF*，*3*(4), 359–390. Zvonkin, A. K., & Levin, L. A. (1970). 有限对象的复杂性以及信息与随机性概念的发展。《信息传输问题》，*6*(3), 1–12.