从每个单词一个标签进行随机自动机的多项式时间学习

程怡霖
中国上海复旦大学；中国北京中关村
学院

arXiv:2611.32959v1 [cs.LG] 24 Nov 2026

摘要

我们研究一种二进制确定性有限自动机，其 $2n$ 个状态转移独立地从其 n 个状态中均匀选取。学习器观察长度为 m的独立均匀词，每个词都标有一个表示是否接受的比特。对于每个固定的 $\beta > 0$，我们给出一个多项式时间算法；除去概率为 $o(n^{-\beta})$ 的一组转移表，该算法均能奏效。对于其余每张转移表，任取一个起始状态和接受集合，预测器在新的、长度为 m 的均匀词上的误差至多为 ϵ；这一结论以至少 $1-\gamma$ 的概率成立，该概率是针对训练样本和学习器的随机选择而言的。其样本量和运行时间关于 n、 m、 $1/\epsilon$ 和 $\log(2/\gamma)$ 均为多项式，且次数仅取决于 β。尽管每个词只揭示一个输出，但比较两个连续历史长度下的标签矩，可以让我们估计从同一隐藏状态出发的两个延续的带符号输出的期望乘积。

我们还反驳了 Fish 和 Reyzin 关于相遇时间的猜想 4。对于在抽取转移表之前固定的两个互异状态，以及每个固定的 $c > 0$，将它们映射到不同状态的长度为 $\lfloor cn \rfloor$ 的词所占比例依概率收敛于 e^{-c}。在典型的转移表上，要将这一比例降至$1/n$，渐近地需要 $n\log n$ 个字母。部分证明步骤已在 Lean 中形式化。

关键词： 随机自动机、被动学习、同步、矩方法

1. 引言

考虑一个逐字母读取词的确定性有限自动机。被动学习器可以观察输入词以及该词是否被接受，但看不到沿途访问的状态。Fish 和 Reyzin [1] 提出了这样的问题：当转移表是随机的且输入词服从均匀分布时，这种学习器能够多高效地预测接受情况？对于一个固定的起始状态，穷举搜索需要遍历 $n^{2n}2^n$ 种转移表和接受集合的组合。对于长度为多项式的词$m = \Omega(n)$，他们的命题 5 给出了一种有条件的 $2^{O(n)}$时间算法，其误差为多项式倒数：如果某个线性长度的公共后缀能以足够高的概率使任意两条轨迹合并，那么支集在该后缀上的傅里叶系数就足以进行预测。

我们的多项式时间学习器采用了不同的方法。其重构步骤需要来自同一隐藏状态的两个延续的响应，然而每个训练词仅揭示了对一个延续的响应
一。我们的学习器通过比较两个连续历史长度下的标签矩，来估计两个此类带符号输出的期望乘积。第 2 节的相遇时间定律也解释了命题 5 的同步前提为何不成立：经过 $\lfloor cn \rfloor$个字母后，对于任意固定的 $c > 0$，仍有趋近于 e^{-c} 的比例的所有词会将两个指定起点送至不同状态。。

© 2026年 Y. Cheng。

模型。 $2n$ 是 f_0, f_1 : $[n] \to [n]$ 的条目彼此独立且均匀分布。对于一个词 $w = w_1 \cdots w_m$，我们用 $f_w(x)$ 表示从 x 出发、读入 w 后到达的状态。设 b : $[n] \to \{0,1\}$ 为接受集合的指示函数，$\xi = 2b - 1$ 为其符号编码， s 为指定初始状态。在训练长度 m 下，每个样本都是一个独立的数对

$$(W, b(f_W(s))), \qquad W \sim \mathrm{Unif}(\{0,1\}^m).$$

所有样本共享相同的状态转移和接受集合。对于固定的带标签目标 M，预测器的风险 $\widehat{\ }$ $\widehat{\ }$为 $R_{\ell,\ M}(f) = \Pr_{|V|=\ell}[f(V) \neq b(f_V(s))]$, 其中 V是长度为 ℓ的均匀词。对接受集合均匀的保证允许在绘制转移表之后再选择 b；在附录中，逐状态比特向量称为接受掩码。除非指明底数，否则对数均为自然对数。

定理 *1*（在观测长度下学习）对于每个固定的 $\beta > 0$，存在一个被动算法$\mathcal{A}_\beta$，使得对于每个 n，存在一个转移表集合 $\mathcal{G}_{n,\beta}$，其概率为 $1 - o(n^{-\beta})$，并具有以下性质。固定 $\mathcal{G}_{n,\beta}$中的任意转移表、任意接受集合和起始状态、任意 $m \geq 1$ 以及任意 $\epsilon, \gamma \in (0, 1/2)$。 $\mathcal{A}_\beta$ 的唯一输入是 n、m、 ϵ、 γ 以及如上所述的独立样本（$W, b(f_W(s))$）。以至少 $1 - \gamma$ 的概率（关于训练样本和学习器的随机选择），它输出一个预测器 $\frown$ $\frown f$，满足 $R_{m,M}(f) \leq \epsilon$。其样本量和比特运行时间关于 n、 m、 $1/\epsilon$和 $\log(2/\gamma)$ 均为多项式，其度仅取决于 β。

取 $\gamma = n^{-\beta}/2$，对于所有充分大的 n，在表、样本和学习器的随机选择上，联合失败概率至多为 $n^{-\beta}$。训练词和测试词的长度始终相同，均为 m。

推论 1′（公平接受比特）存在一个绝对常数 $c > 0$，使得对于每个固定的整数$D \geq 1$，都存在一个双批次学习器 $\mathcal{A}'_D$，并具有如下保证。对于每个指定起点，$1 \leq m \leq n^D$ 和 ϵ，$\gamma \in (0,\ 1/2)$，它仅接收 n、 m、 ϵ、 γ 以及相互独立的带终端标签的词。在独立公平接受比特的条件下，返回风险大于 ϵ 的预测器的概率至多为 $O_D(n^{-c}) + \gamma$，其中概率取自表、接受位、训练样本以及学习器的任意随机选择。对于固定的 D，其样本量和比特运行时间关于 n、 $1/\epsilon$、$\log(2/\gamma)$ 均为多项式。

附录 B.1.7 直接证明了这一基线方法。其样本界远不实用：即使在 $m = n$ 时，第一批样本的规模关于 573722 的度为 n，仅相差对数因子（第 6 节）。

学习器如何使用一个标签。固定一个短历史和两个延续测试。对于每个测试，我们估计以该历史为结尾、后接该测试的词的带符号标签均值。总体均值的乘积对应于两个独立的隐藏起点读取相同历史后，分别读取各自测试的情形。我们先在一个长度上，再在下一个长度上，对这些乘积按历史取平均。在两个长度上始终互异的状态对，其贡献比例几乎相同。因此，加权差值估计的是在一个单个状态上对这些测试的期望乘积（第3节）。这些乘积足以区分未来标签不同的循环状态（第4节）。当表中存在一个较小的闭合分量，可能使运行在词的早期就被吸收时，还会出现另一项困难；第5 节说明在这种情况下，如何保留输入前缀中的相关信息。附录C.7 在原始随机表分布下综合这两种情况，并明确给出窗口长度、精度和样本量。

2. 共同词下的相遇时间

令 τ 为公共随机词首次将两个预先指定的互异初始状态映射到同一状态的时刻。给定表后，令 $q_i(t) = \Pr_W(\tau > t \mid A, i)$表示长度为$t$的词中将两个起点映射到不同状态的词所占比例。在二元独立同分布模型中，对于固定的 $c \geq 0$和$\beta > 0$,

$$q_i(\lfloor cn \rfloor) \xrightarrow{\mathbb{P}_A} e^{-c}, \qquad q_i(\lfloor \beta n \log n \rfloor) = n^{-\beta + o_{\Pr}(1)}. \tag{1}$$

这两个极限都是针对转移表的概率而言的。在一张典型的转移表上，对数窗口的初始阶段过去后，一个仍然互异的指定对在下一个字母处相遇的条件概率为 $(1 + o(1))/n$。由于两个转移行相同而选出的对会在第一个字母处相遇。Quattropani 和 Sau [2, 预印本，定理 2.2 和推论 2。5] 证明了由独立 词驱动的两条游走的几何相遇定律，其中每一行的不同出发目标互异。他们的定律驳斥了独立游走需要线性长度的断言，并且还模拟了共同词的情形。Fish 和 Reyzin 猜想中的共同输入涉及公式（1），在该猜想中，两条游走重复使用相同的状态转移；附录 A.2 对此进行了证明。学习算法需要一个相关的量，即在对数窗口中存活下来的各对的分布，下一节将解释终端标签如何使我们能够获得该分布。

3. 从终端标签提取同状态乘积

记 π 为大型常返分量上的平稳分布， $T_a g(x) = g(f_a(x))$，以及 $m_v(g) = \pi T_v g$。对于 $g = T_u \xi$， $m_v(g)$ 是从 π 抽取的状态读取 vu 时的期望终止符号。在训练数据中，对后缀 vu 进行条件化后，会留下一个来自指定起点的均匀随机前缀。若该前缀足够长以完成混合，则条件均值以指定的逆多项式精度逼近 $m_v(g)$。对于较短的训练长度，学习器将完整词制成表。重构步骤需要 $m_v(gh)$，其中两个测试作用于同一隐藏状态。

在固定的表上定义

$$S_l = \mathbb{E}_{|w|=l} (\pi T_w)^{\top} (\pi T_w), \qquad J_l(v; g, h) = \mathbb{E}_{|w|=l} m_{wv}(g) m_{wv}(h).$$

学习器分别估计每个乘积中的两个均值。 S_l 的非对角部分行 u_l 在由同一字母驱动的有序互异对上的次随机核 Q 下演化，而 S_l 的两个边缘分布均等于 π。令 $x_l = (u_{l+1}\mathbf{1})/(u_l\mathbf{1})$ 以及 $q_l = 1 - x_l$。若在一步之内，非对角部分行的变化主要体现在其总质量上，而形状变化很小，则从 S_{l+1} 中减去 $x_l S_l$ 后基本上只剩下对角线。精确地说，

$$\left\| \frac{S_{l+1} - x_l S_l}{q_l} - \operatorname{diag}(\pi) \right\|_{\text{ent}} \leq \frac{2 \| u_{l+1} - x_l u_l \|_1}{q_l},$$

其中 $\| \cdot \|_{\text{ent}}$ 表示各项绝对值之和。因子 2 用于计入非对角部分的余项，以及为保持边缘分布而进行的对角线修正。对于足够长的对数级l 右侧小于任意指定的逆多项式，而 $q_l \asymp 1/n$。将此应用于$T_v g$ 和 $T_v h$ 即可得到图1中的乘法规则。

该规则的有限样本代价由 q_l 决定。如果 J 的两个估计值都在其总体值的 η 范围内，且截断比率 $\widehat{x} \in [0, 1]$ 在 x_l 的 $d \leq q_l/2$ 范围内，那么商的变化幅度至多为 $4\eta/q_l + 4d/q_l^2$。事实上，其分母至少为 $q_l/2$；替换所导致的变化

History before v	Product of two terminal means
$\lvert w\rvert = l$	$J_l = \mathbb{E}_w[m_{wv}(g)m_{wv}(h)]$
$\lvert w\rvert = l+1$	$J_{l+1} = \mathbb{E}_w[m_{wv}(g)m_{wv}(h)]$
$(J_{l+1} - x_l J_l)/(1-x_l)$	$\longrightarrow \quad \pi T_v(gh) = \sum_z (\pi T_v)_z \, g(z)h(z)$

图 1：同一张转移表上的两个窗口。在线上方，均值的每个乘积对应两个独立的平稳初始状态，它们读出相同的历史 w 和后缀 v。对两行加权并相减，会抵消仍然互异的状态对，余项除外，其上界见正文。在线下方，两个测试函数都作用于同一个状态 z。

x_l 单独为 $(\widehat{x} - x_l)(J_{l+1} - J_l)/[(1-\widehat{x})q_l]$，且 J 的两个值都位于 $[-1,1]$中。即使它们的词分箱存在重叠，只要原始均值同时足够准确即可。附录 B.1 在总体层面界定了该近似的界限。

该比值 x_l 本身可以从数据中估计。对于符号读出，$B_l = J_l(\varnothing;\ \xi,\ \xi)$ 满足$1 - B_l = u_l$ $(1 - \xi_i\xi_j)$ 因为 $\xi_i^2 = 1$，所以$(1 - B_{l+1})/(1 - B_l)$ 估计 x_l。对于标签并非常数的核心，即使接受集合是在抽取转移表后才选定的，分母中的不匹配质量也有一个逆多项式下界（附录 B.4.1）；常数核心则可通过其终端符号频率识别。递归应用乘法规则，可以估计任意固定度数的短符号测试乘积；每项估计都只使用训练长度下普通带标签词的词分箱。

为什么非对角余项很小。考虑同一已揭示的表上一个公共词对的两个副本，由相互独立的词驱动。一旦四条轨迹占据了四个新状态，有序匹配两个端点对在主导阶上需要与已命名状态发生两次碰撞。如果两个连接之后还剩下 i 个字母待读取，则副本词必须在最后 $\max(i,j)$ 个字母上保持一致。因此，$1/n^2$ 的系数为

$$\sum_{i,j\geq 0} 2^{-\max(i,j)} = \sum_{j\geq 0} (2j+1)2^{-j} = 6.$$

配对块的单行及其平均行出现相同的系数，因此，减去这两个二阶矩即可控制中心化块的弗罗贝尼乌斯范数平方。为此，两个矩界必须在同一个表事件上取得。对成功查询记录的整个已揭示森林进行条件化后，所有其他单元格仍相互独立；在之后的局部搜索中，两个多余的边需要另外两次命中，由此得到相应的下界（引理 10）。改变一行出边只会通过读取该行的轨迹扰动中心化块，因此，利用独立行上的集中性，可将该界提升到没有小型终端分量的表上任意给定的目标置信度（附录B.3）。

4. 恢复循环响应

如果对任意延续，两个状态产生的输出都相同，则称这两个状态预测等价。学习器首先识别这些循环等价类，以及它们的转移和接受符号。在此阶段，这些类只能确定到重新标记的程度，初始状态不起作用。

特征是由以不同首字母开头的测试函数相乘得到的。中心化格拉姆矩阵估计结合异常对的有界度图，为每个递归预测类提供一个线性泛函，使其能够在较小误差范围内隔离出对应的行；固定的张量幂则抑制剩余的交叉项。所得字典矩阵 V 每个类对应一行，具有满行秩，并且 $\sigma_{\min}(V) \geq n^{-C}$，其中指数取决于预先规定的置信度。下文使用的短词类特征也能区分所有类。

罕见的抵消需要谨慎处理。固定一个对数窗口，并考虑一个匹配，其中h 对的有序后继行不同，但存活概率极小。连续两次二叉分叉会迫使至少出现 $3h$ 次与已命名目标顶点的碰撞，而候选匹配至多有 n^{2h} 个；另有一项单独计数可界定与单个状态相连的此类对的数量。因此，除概率为 $o(n^{-\beta})$ 的目标异常事件外，只剩下有界数量的后继行不等的异常对。每一对要么沿一条有界路径到达具有正存活系数的区域，要么在合并之前一直留在这个有限集合中。这两种情况下，两个常返类的响应语言之间的任何差异，都可由一个对数长度的词见证（附录 B.6）。

从可观测矩阵到离散状态。列出长度至多为 $L+1$的词上的符号测试函数，其中 $L = C \log n$ 是区分深度。令s_j 表示在常返类上取值的测试函数 j，并令 ω 表示它们的平稳权重。乘法规则估计

$$D = V^{\top} \operatorname{diag}(\omega) V, \qquad D_j = V^{\top} \operatorname{diag}(\omega s_j) V.$$

对于给定的初始符号模式 a_1，……，a_d，构造 $K(a) = (d - \frac{1}{2}) D - \sum_{j \leq d} a_j D_j$。在相应的对角因子中，与该模式匹配的类的系数为 $-1/2$，而其他每个类的系数至少为 $3/2$。因此，缺失的模式会给出一个半正定矩阵；而对于一个已实现的模式，将伪逆应用于匹配类的坐标，会得到一个负的瑞利商，其幅度至少为 $\min_{\omega} \sigma_{\min}(V)^2/2$。因此，利用一个已知的小偏移量进行有理正定性测试，便可使用估计矩阵区分这两种情况。在符号模式树的每一层，至多有 n 个分支存活。空测试函数的符号给出某个类的输出，而在每个较短的测试词前置一个字母即可确定其后继。一旦解码出状态转移，便可沿着测试词依次进行这些离散状态转移，从而完成预测。无需进一步对估计矩阵进行乘法来放大其误差。

在指定起点处，一个较短的前缀决定运行进入该机器的位置。为了准确预测较长的词，在异常事件概率为 $O(1/n)$ 的表事件上，每个 g字母前缀都会到达核心，而前缀剩余的瞬态行为可以并入一台至多具有 n 个状态的机器（附录 B.5）。对于允许正误差的预测，只需舍弃少量前缀，并利用新的带标签样本比较候选的初始核心状态。

5. 小型闭合分量学习

以 $1/n$ 量级的概率，表中会包含一个两种字母都形成自环的状态，因此，承诺失败概率为 n^{-2} 的学习器必须能够处理这类表。终止样本从不揭示运行何时进入闭合分量，而在较大的闭合分量内，最终状态可能一直变化。尽管如此，仍可以学习平均标签：只需询问运行进入该分量后，标签中哪一部分可能取决于前缀。

设 T 是由 s 个状态组成的闭集，并暂且假设已知其两个内部映射，但不知道状态标识符或输出。对于长度为 r 的后缀，令 E_r 为常值函数以及 s^2 个指示函数所张成的空间，其中每个指示函数表示 T 中某个指定状态是否能到达另一个指定状态。如果运行已经处于 T 内，其整个后缀响应都属于 E_r。揭示下一个输入位会将该空间从 E_r 扩大到 $L^2(\{0,1\})\otimes E_{r-1}$。对于任意归一化后的新增正交方向 d，可观测响应 $Yd(W_{j+1:m})$ 在给定一个以 T 内状态结束的前缀时，其均值为零；而给定整个前缀时，其二阶矩至多为一。因此，原任务可归约为 $O((m+1)(s^2+1))$ 个回归问题，其目标在 T 上为零，且这些回归的误差在平方损失下相加。

在每次此类回归中，具有显著残差信号的一个外部状态都存在一个具有可检测均值的短后缀。在该后缀之前插入一个混合块，会使可观测的条件响应与避开 T 的单状态存活系数成正比；对该系数设置阈值，即可识别出残差会迅速变小的状态。两个在所有测试上都一致的候选隐藏起点，沿着共同延续会面临三种可能：它们合并，其中一个比另一个更早达到阈值，或者它们剩余的残差对预测贡献很小。在第二种情况下，一个可观测的短词可以将它们区分开来。由此得到一种算法，利用实际测试词的前缀追踪剩余的区别（附录 C.6）。

在单个干净基础表上，该论证可同时适用于所有此类修改： T可以在检查表之后选择，$|T| \le n^{1-a}$ 对于固定的 $a>0$，而 T的出边行可以替换为任意映射到 T的函数。学习器需要抽象内部作用，否则可以在时间 $s^{O(s)}$内枚举所有动作。在独立同分布模型中，一个给定的 s-集以概率$(s/n)^{2s}$成为闭集。通过饱和小型闭集，并仅对其出边行重新抽样，我们证明：除了概率为 $o(n^{-\beta})$ 的表异常之外，存在一个具有 s 的干净补全，其界仅取决于 β，因此枚举的成本只依赖于 β这一常数。完整构造及其量词见附录C.2–C.7。

算法 1. 在观测长度下学习。固定 β，并使用附录C.7中的窗口、容差和列表上限。输入由 n、 m、 ϵ、 γ以及带终端标签的词组成。

1. 从后缀分箱均值构建循环核心的候选项：用终端符号校准双窗口差值，构造 D 和 D_j，并用第4节的移位符号模式测试解码循环响应，包括响应为常数的情况。在新词上拟合前缀预测。

2. 此外，尝试至多作用于 $\lceil \beta+2 \rceil$ 个状态的每一种二元作用。对于每次试验，根据观测到的词和标签构造粗粒度后缀响应及正交细节任务，分别拟合并验证它们的读出，然后将它们相加、截断并进行阈值化。如果指定窗口超过词长，则改用直接词分箱。

3. 在新的完整词上比较所有由此得到的分类器，并返回经验分类误差最小的一个。

置信度的作用。 令 S_k 表示每个终端分量都包含多于k 个状态这一事件。对于固定的指数 A，干净表估计的形式为 $\Pr(S_k \cap G_A^c) = o(n^{-A})$，其中 G_A 仅涉及状态转移。对于典型表，中心块估计可通过一种仅含一条多余弧的初等搜索得到。为得到更小的失败概率，附录 B.3允许存在数量有界的多余弧，然后利用独立行编辑下的集中性。

附录 C.7 将此估计用于 $A = 2\beta + 10$，应用于一个保留所有外部行的补全；引理 24 则在实际删除集合对应的同一补全上应用。表一旦固定，新标注词便足以选出一个候选项，使训练失败概率达到 γ。

6. 窗口长度和样本成本

对于 $m = n$ 带有独立公平符号的基线构造（附录 B.1），观测一个指定的长度为 L 的后缀会使样本量增加一个 2^L 的因子。当 $d = \lceil 10 \log_2 n \rceil$、$R = 7d + 3$ 和 $t = 4096R$时，我们有 $L = 2t + 4d + 6 = 57348d + 24582$，因此第一批的大小为 $O(n^{573722}(\log n + \log(8/\gamma)))$，其中 γ 是训练失败概率（在附录 B.1 中记为 δ）。岭矩阵有 $p = 2^{t+1} = O(n^{286720})$ 行，直接进行稠密求逆需要 $O(p^3) = O(n^{860160})$ 次算术运算。这些数值针对该特定调度中的单个批次和单次操作。对于定理 1，指数还取决于小分量截断值、块压缩以及解码器所需的精度；附录 C.7 给出了每个阶段的成本。

在不同滞后下测量近似值。对于每个规模 n，我们抽取了 24 个独立表，并计算了总体误差 $\| (S_{t+1} - x_t S_t)/(1 - x_t) - $ 对角线 $\pi \|_{\mathrm{ent}}$。从 $\pi \otimes \pi$ 开始，我们在两个字母映射下传播成对分布，并对共同词进行精确平均。表 1给出了具有唯一闭类和正存活质量的表的中位数，以及$q_t > 10^{-12}$。在四个滞后下，我们使用了相同的抽样结果。

表 1：总体对角线提取误差（绝对值项之和）。最后一列给出了在每个所示滞后下误差有定义的表的数量。

n	$t = 3$	$t = 10$	$t = 20$	$t = 40$	24 个表中有定义的数量
6	0.261	0.00413	1.31×10^{-5}	1.14×10^{-10}	所有滞后下均为 20
10	1.15	0.0290	3.71×10^{-4}	7.75×10^{-8}	所有滞后下均为 20
16	1.72	0.121	2.57×10^{-3}	1.48×10^{-6}	22, 22, 22, 21
32	4.30	0.309	8.14×10^{-3}	6.33×10^{-6}	所有滞后下均为 20
64	8.77	0.732	0.0202	1.60×10^{-5}	所有滞后

解码器恢复的内容。在 72 个独立表中（$n = 6$、10、16 各 24 个），有 61 个具有带非恒定标签的唯一常返类。我们在每个滞后下测试了全部 61 个表。为将解码与第 4 节中的条件化论证分离开来，我们使用了长度不超过 $\lceil \log_2 n \rceil$ 的单符号检验，并额外使用一个字母来识别签名和后继。对于预言机移位 $(\min_x \omega_x)\sigma_{\min}(V)^2/4$，我们从隐藏表中提取质量和特征，将具有相同签名的质量相加，并为每个签名保留一行归一化行。成对和三元递归对公共字母取平均。在滞后为 3、10、20、40、60、80、120 时，解码器分别恢复了 61 个目标中的 4、30、58、58、58、58、58 个；使用精确的隐藏矩时，在相同移位下恢复了全部 61 个目标。

从滞后 20 开始，同样的三个核心未能满足标签比率保护条件 $1 - x > 10^{-5}$，而从每个核心的二元图中的对角线开始进行反向搜索，结果表明恰好这三个核心不同步。在抽样 $(n,\ \mathrm{rep}) = (6,\ 18)$ 中，两个字母分别执行单位算子和交换。在抽样 $(6,\ 14)$ 中，合并两行相同的转移行后，三个状态上仍有置换作用（合并的两个状态带有不同标签）。在抽样 $(16,\ 7)$ 中，每个字母都会交换块 $\{7, 15\}$ 和

$\{8, 10\}$，因此，起始于相反块中的状态永远不会相遇。精确矩比较直接给出了隐藏矩。

随机二进制表无法同步的概率 $\Theta(1/n)$ [3], 足以对每个 $\beta \geq 1$产生影响。在不存在小型终端分量的良性事件上，(36) 给出 $\rho(Q) < 1$。随后，每一对状态都能在某个词的作用下合并，连续合并便会将像集缩减为一个状态。附录 C.7 界定了该层上的例外情况。对于保证范围内剩余的表，学习器会枚举有界的闭合动作，并利用观测到的前缀。该实验仅测试递归核心解码器。

Lean 源码。 该 Lean代码库包含 22 个模块，用于形式化部分证明步骤，包括四源系数，以及删除状态后使用的确定性界。

7. 相关工作

自动机的平均情形学习。 Freund等人 [4] 在随机游走的每一步都观测一个标签。Angluin和Chen [5] 观测每个词到达的状态，因此他们的统计查询可以估计状态概率向量。对于独立同分布的随机转移，他们对常返分量的恢复界假设观测矩阵满秩，并取决于该矩阵的伪逆。在这里，每个词只揭示一位信息。随机转移表决定了由此产生的矩是否能区分循环响应，以及需要以多高的精度对其进行估计。

Maranzatto [6, 第 5.2–5.4] 节研究了偏置输入游走。学习器在每次预测后观察一个标签，并且可以通过付出一次错误的代价进行重置。该工作第 5.4.2 节还使用两个独立均匀的字母映射来研究单个字母的循环。我们的学习器只观察每个独立词的最终标签，并在相同长度下进行预测。

矩方法与从生成字符串中学习。 可观测算子模型与谱算法 [7; 8; 9] 使用相关的矩分解。在两条轨迹中使用共同输入，使我们能够根据终端标签估计一个隐藏的乘积。Angluin等人 [10] 证明了针对依赖目标的输入分布的统计查询下界。我们的输入分布是在固定长度下的均匀分布。Mossel和Roch [11] 表明，学习一般的隐马尔可夫模型至少与学习带噪声奇偶校验一样困难，并在非奇异条件下给出了一种算法。Clark和Thollard [12] 学习由概率确定性自动机生成的字符串分布，其界限取决于状态可区分性以及各状态上的最大期望字符串长度。在我们的问题中，输入分布是已知的，而未知函数为每个词分配一个接受位。

8. 讨论与开放问题

第 6 节的第一批样本量很大，因为长后缀很少见，而且重构会放大其均值估计中的误差。表1 显示，在所测试的表上，窗口短得多时总体误差仍然很小。能否在这些长度下证明关于 $\|u_{l+1} - x_l u_l\|_1/(1 - x_l)$ 的更紧界？若要降低样本指数，这样的界必须满足校准和解码的精度要求。

第二个问题是，多项式次数能否独立于目标置信度指数 β 选取。我们的枚举规模随着学习器必须处理的最大隐藏闭合分量而增长

。此外，双窗口方法能否在字母数超过两个的字母表上恢复循环响应，目前也尚无定论。

参考文献

[1] B. Fish 和 L. Reyzin。《开放问题：学习随机自动机的相遇时间》。COLT，PMLR 65，8–11，2017。[2] M. Quattropani 和 F. Sau。《随机确定有限自动机上随机游走的相遇》。Stochastic Process. Appl. 166，104225，2023。预印本。[3] M. V. Berlinkov。《关于可同步性的概率》。CALDAM，LNCS 9602，73–84，2016。[4] Y. Freund、M. Kearns、D. Ron、R. Rubinfeld、R. E. Schapire 和 L. Sellie。《通过随机游走高效学习典型有限自动机》。Inf. Comput. 138(1)，23–48，1997。[5] D. Angluin 和 D. Chen。《从均匀字符串和状态信息中学习随机DFA》。ALT，LNCS 9355，119–133，2015。[6] T. J. Maranzatto。《用于学习和信息论的组合方法》。博士学位论文，伊利诺伊大学芝加哥分校，2024，第5章（与 K. Dubin、M. Michelen 和 L. Reyzin 合作）。[7] H. Jaeger。《离散随机时间序列的可观测算子模型》。Neural Comput. 12(6)，1371–1398，2000。[8] D. Hsu、S. M. Kakade 和 T. Zhang。《一种学习隐马尔可夫模型的谱算法》。J. Comput. Syst. Sci. 78(5)，1460–1480，2012。[9] B. Balle 和 M. Mohri。《通过约束矩阵补全进行一般加权自动机的谱学习》。神经信息处理系统进展 25，2168–2176，2012。[10] D. Angluin、D. Eisenstat、L. Kontorovich 和 L. Reyzin。《利用统计查询学习随机结构的下界》。ALT，LNCS 6331，194–208，2010。[11] E. Mossel 和 S. Roch。《学习非奇异系统发育树和隐马尔可夫模型》。Ann. Appl. Probab. 16(2)，583–614，2006。[12] A. Clark 和 F. Thollard。《概率确定有限状态自动机的PAC可学习性》。J. Mach. Learn. Res. 5，473–497，2004。[13] V. Yu. Protasov。《表面维数、铺砖与同步自动机》。SIAM J. Math. Anal. 52(4)，3463–3486，2020。扩展预印本（2021），定理8。[14] L. Addario-Berry、B. Balle 和 G. Perarnau。《随机r出向图的直径和平稳分布》。Electron. J. Combin. 27(3)，P3.28，2020。[15] M. Rudelson 和 R. Vershynin。《Hanson–Wright 不等式与次高斯集中》。Electron. Commun. Probab. 18，第82号，1–9，2013。[16] E. J. McShane。《函数值域的扩张》。Bull. Amer. Math. Soc. 40(12)，837–842，1934。[17] C. McDiarmid。《关于有界差分法》。载于 J. Siemons（编），《组合学综述》，1989，London Math. Soc. Lecture Note Ser. 141，148–188。剑桥大学出版社，1989。[18] F. Bassino、J. David 和 C. Nicaud。《Moore状态最小化算法的平均复杂度》。STACS，LIPIcs 3，123–134，2009。[19] G. Chapuy 和 G. Perarnau。《随机自动机的短同步词》。ACM Trans. Algorithms 21(4)，第41篇，2025。

推论的证明思路 1′。

$$\underbrace{\text{first meetings}}_{\text{Appendix A}} \longrightarrow \underbrace{\text{moments and readout}}_{\text{Appendix B.1}} \xrightarrow{\text{length schedule (D.3.1)}} \underbrace{\text{target failure } O_D(n^{-c})}_{\text{Appendix B.1.7}}.$$

任意的保证β 使用附录中的独立补全论证 C.7。

附录 A. 在共同词下的首次相遇

两个指定状态。对于 $i = \{x, y\}$，当 $x \neq y$ 时，令 τ 为首次相遇时间。记 $q_i(t) = \Pr_W(\tau > t \mid A,\ i)$ 为词的存活比例。

定理 2（共享词相遇时间）对于每个指定不同对 $i = i_n$，以及任意固定的 $c \geq 0$,

$$q_i(\lfloor cn \rfloor) \xrightarrow{\mathbb{P}_A} e^{-c}, \qquad \frac{1}{n}\mathbb{E}_W[\tau \mid A, i] \xrightarrow{\mathbb{P}_A} 1.$$

在 $t = \lfloor \beta n \log n \rfloor$ 时，对每个固定的 $\beta > 0$，存活概率为 $n^{-\beta + o_{\Pr}(1)}$。

在猜想 4 中，$\lambda = 1$ 已经与固定的c极限矛盾。然而，一对具有相同行转移的顶点会在第一个字母上相遇。

A.1. 对数窗口内的首次相遇

对于互异的无序对，令 Q 为对角线被杀死后的次随机转移核。设$f_t(i) = \Pr_W(\tau = t \mid A,\ i)$。我们先估计单个滞后下的 f_r，再利用存活递推得到较晚时刻的界。设

$$\ell = \lfloor (\log_2 n)/5 \rfloor, \quad d = \lceil 10 \log_2 n \rceil, \quad r = 4d + 1, \quad k = \lceil (\log n)^3 \rceil, \quad M = 2kr + 2k + 10.$$

已命名顶点是指先前指定的起始状态、查询源或揭示的目标。在 k 个副本的 r 步过程中，至多有 $2kr$ 次单元查询；计入初始状态和任何预先指定的终端有序对后，在每次延续中，已命名顶点的总数都小于 M。不同副本的词相互独立，且使用同一张表。

引理 3 除去概率为 $O(n^{-1/5})$ 的事件，对于每个顶点以及大小至多为 M的每个集合 H，一个均匀输入游走接下来的 d 步全部留在 H 中的概率至多为 n^{-3}。

证明 设 G 表示以下事件：每次向外进行广度优先搜索、揭示距离为 $< \ell$ 的顶点发出的所有弧时，发现树之外的弧至多有一条。每个根节点至多探测 $N = 2^{\ell+1}$ 个表项。每次新的探测命中已命名顶点的概率至多为 $(N + 1)/n$。这样的两次探测代价至多为 $\binom{N}{2}((N + 1)/n)^2$。对 n 个根节点求和。在 G 上，从 x 到指定 y、长度为 ℓ 的词至多有 $\ell + 1$ 个：只要指定唯一额外弧的经过次数，就能确定其间的有向树路径段。此处将平行标记弧视为不同的弧。对于所有足够大的 n，经过 ℓ 步后到达 H 的概率至多为 $M(\ell + 1)2^{-\ell} \leq n^{-1/6}$。在连续的块末时刻应用马尔可夫性质，直至时间 d。 ■

仅在首次查询表项时才揭示该表项，并且只在需要时才揭示输入比特。若新探针的目标顶点已经命名，则称该探针为一次返回。读取先前已揭示的单元格不算作返回。在每一步中使用固定顺序，并立即标记每个目标顶点。给定已揭示历史，未查询单元的取值服从均匀分布。一次试验从一个安全的部分表开始；若从每个已命名顶点出发，仅沿当前已揭示的表项行进 d 个公平比特的概率至多为 n^{-3}，则该部分表是安全的。通过在遇到未知表项时终止行走，即可根据部分表计算此性质。在 G 上，此处考虑的每个部分历史都通过测试，这是引理 3的结论。安全性检查在部分表上进行，以确保未查询单元仍服从其条件乘积分布。

引理 4 若一对互异的顶点中，两个当前顶点都从未作为源被查询过，则称该顶点对为新配对。从一个安全的部分历史出发，在接下来的 $2d$ 步的任何延续中，指定顶点的数量始终不超过 M。在没有获得新配对的情况下，经过 $2d$ 步仍未合并的概率至多为

$$\eta = 4dM/n + 2n^{-3} + d^2 2^{-d}.$$

证明 排除这些步骤中的返回，代价至多为 $4dM/n$。令 H 表示初始命名集合，其中包含每条初始暴露弧的两个端点。在此事件下，一条路径若要在前 d 步始终留在 H 内，就只能使用初始已知的表项。根据安全性，任一坐标发生这种情况且不返回的概率至多为 $2n^{-3}$。

两个坐标现在都在时间 d 前离开 H。先离开的坐标会在 H 之外发现一条延伸的简单路径。如果两者同时离开，它们的新目的地就是互异的新源。如果第二个坐标通过一条新表项离开，那么两者再次都是新源。唯一的另一种离开方式，是沿着第一个行走者之前经过的离开弧：在第二个行走者离开之前，这是从 H 到其补集的唯一暴露弧。第二个行走者沿着第一个行走者的路径前进，但存在严格为正的时间滞后。在发生分歧之前，它始终落后，而先行者则不断发现新顶点。路径上某个顶点处首次出现不同的输入字母时，会走到一个未查询表项，此后两个顶点都是新源。只要存活至 窗口结束，这一论证就一直适用。因此，到 $2d$ 仍未成功就意味着，从 $\{1,\dots,d\}$ 中两个互异退出时间之后紧接着开始的长度为 d 的子串相同。对于固定的偏移量，即使子串重叠，其相同的概率也为 2^{-d}：每个相等关系都会将较后的比特确定为较早比特的值。对这些偏移量求和时，使用给定其初始历史后，该试验未来词的无条件分布。 ∎

引理 5 令 $a = (1 + 1/\text{对数 } n)/n$ 和 $L = (\text{对数 } n)^{12}$。以高概率，

$$\max_i f_r(i) \le a, \qquad \max_{i,j} Q^r(i,j) \le L/n^2.$$

此外，对于每个给定的 i，存在 $b_n = 1/n + O((\text{对数}\, n)^{10}/n^2)$ 以及 $\mathbb{E}[\mathbf{1}_G f_r(i)^2] \le b_n^2$。

证明 尝试在前 $2d$ 步内获取一个新配对；如仍有必要，则在接下来的 $2d$ 步内继续尝试。在实际进行的任何尝试开始时，若部分表不安全则中止。截断成功是指在第 r 步发生的一次普通首次相遇（对于条目界，则是在指定配对处的一次普通存活到达），且期间未发生此类中止。获取新配对失败的代价将在下文计入，但这本身不会中止试验。所考虑的成功要求存活至严格早于最后一步。通过迭代条件化，在仍符合条件的情况下两次尝试均失败的代价至多为

η^2。获取后，在最后一步之前至多还有 $2r$ 个探针。其中出现两次或更多次返回的条件概率至多为 $\binom{2r}{2}(M/n)^2$。

在相遇之前，一个起初为新鲜的坐标会一直停留在新源，直到在其自身的某个探针上返回。因此，若最终步骤之前严格至多发生一次返回，那么两个最终探针中至少有一个是新探针。即使以最终字母为条件，并以另一个目标为条件（当另一条目不同时），它的目标仍是独立均匀的。给定这样一段步骤前历史，其中当前状态互异，并给定最终字母，两个表格单元格互异；两者相等的概率恰为 $1/n$。对返回次数的限制在这些最终探针之前结束。截断单副本成功概率的上界为

$$b_n = 1/n + \eta^2 + \binom{2r}{2}(M/n)^2.$$

若要落在某个特定的互异无序对上，获取之后在最终步骤前没有返回，就会留下两个新探针，其概率至多为 $2/n^2$。一次最终步骤前的返回，其概率至多为 $2rM/n$，而目标概率至多为 $2/n$。如前所述，将两次返回或两次失败尝试计入后，可得上界 $b'_n = O((\log n)^{10}/n^2)$。

对于每个固定的表，普通副本成功在各个词之间相互独立。在 G上，普通成功会通过每项安全测试，因此 $\mathbb{E}[\mathbf{1}_G f_r(i)^k]$ 至多等于连续 k 次截断成功的概率。迭代均匀的历史条件界可得 b_n^k，且无需在 G 上对未使用的条目进行条件化。使用两个副本即可得到所述的二阶矩。同样的论证给出$\mathbb{E}[\mathbf{1}_G(Q^r(i,\ j))^k] \le (b'_n)^k$。应用马尔可夫不等式以及大小至多为 n^2 和 n^4 的并集界；$n^2(b_n/a)^k = o(1)$ 和 $n^4(n^2 b'_n/L)^k = o(1)$。 ■

A.2. 从首次相遇到存活概率

令 $S_j = q_i(jr)$ 以及 $u = ra$。在引理5所述事件上，对于 $t \ge r$，

$$\Pr_i(\tau = t) = (Q^{t-r} f_r)(i) \le a q_i(t-r).$$

因此，对于 $j \ge 1$，有 $S_{j+1} \ge S_j - uS_{j-1}$。对于一个指定不同对，通过揭示路径直到出现第一个重复顶点，计算出在第 r 步之前相遇的期望概率为 $O(r^2/n)$。记 $\gamma = n^{-1/4}$；以高概率 $S_1 \ge 1-\gamma$。同样的条件界也适用于从 s 出发、由两个不同的首字母到达的那一对顶点；将这些目标顶点相等视为立即失败：先揭示这两个条目，并在整个探索过程中将一个重复的已命名顶点计入其中。对于较大的 n，$u/(1-\gamma) \le \gamma$。每当 $S_j/S_{j-1} \ge 1-\gamma$ 时，下一个比值至少为 $1 - u/(1-\gamma) \ge 1-\gamma$。因此

$$S_j \ge (1-\gamma)\big(1 - u/(1-\gamma)\big)^{j-1}, \qquad j \ge 1. \tag{2}$$

引理6 以高概率，集合 $B = \{i :\ \ nf_r(i) < 1 - 1/\text{对数 } n\}$ 的大小至多为 $n^{2-1/10}$。

证明 对于每个固定的 i，要求直到时间 $r-1$ 为止的每个目标顶点都是新的，然后要求最后一个目标顶点中的第一个是新的，而第二个与它相同。对于固定的词，其概率恰为 $(n-2)_{2r-1}/n^{2r}$，其中分子是一个下降阶乘：两条路径通过 $2r$ 个不同的单元格，发现了 $2r-1$ 个互异的新顶点。这至少为 $(1 - O(r^2/n))/n$。每条此类查询记录中已揭示的单元格规定了一个有向图，其底层无向图是一棵由两条内部不相交的路径组成、并在共同的终端顶点处结束的树。其余条目是

独立均匀分布。对于每个搜索根，构建一个辅助无向多重图，其中包含整棵植入树和搜索根；如果搜索根位于树之外，则最初将其视为一个孤立顶点。该辅助图是森林。运行常规广度优先搜索，并将每条新查询到的未植入弧加入辅助图。遇到新目标顶点不会改变圈秩，而命中一个已命名顶点会使圈秩至多增加一。实际搜索图是辅助多重图的子图，其中包括任何经过的植入弧。因此，搜索中出现两条多余弧，就要求在新查询的未植入单元格上发生两次指定目标命中。根据查询记录，这类单元格的目标顶点仍然是均匀分布的。每个根节点至多进行 $N = 2^{\ell+1}$ 次探针，

$$\Pr(G^{c} \mid \text{clean transcript}) \le n\binom{N}{2}\left(\frac{N+2r+3}{n}\right)^{2} = O(n^{-1/5}).$$

这对均匀分布的植入词及其指定顶点均成立。因此 $\mathbb{E}[\mathbf{1}_{G} n f_r(i)] \ge 1 - O(n^{-1/5} + r^{2}/n)$。结合双副本界可得 $\mathbb{E}[\mathbf{1}_{G}(n f_r(i) - 1)^{2}] = O(n^{-1/5} + (\log n)^{10}/n)$。对所有顶点对求和，并应用马尔可夫不等式估计超过平方阈值的计数。 ∎

令 $b = (1 - 1/\log n)/n$ 以及 $v = L n^{-1/10}$。对于 $t \ge 2r$，引理 5 和 6 给出

$$(Q^{t-r}\mathbf{1}_{B})(i) \le v q_i(t-2r), \qquad (Q^{t-r} f_r)(i) \ge b\big(q_i(t-r) - v q_i(t-2r)\big).$$

由于 q_i 单调不增，$q_i(t) \le (1-b) q_i(t-1) + b v q_i(t-2r)$。对于 $\lambda = 1 - b(1-2v)$ 以及所有足够大的 n，$\lambda^{1-2r} \le 2$。利用 $q_i(t)$ 进行归纳，初始得到 ≤ 1

$$\max_{i} q_i(t) \le \lambda^{t-2r}, \qquad t \ge 2r. \tag{3}$$

对于给定的一对对象，也可以直接从这些界读取条件风险率。在事件 $S_1 \ge \theta = 1 - \gamma$ 上，上述块比值对每个 $j \ge 0$ 都给出 $S_{j+1}/S_j \ge \theta$。对于 $t \ge 2r$，因此块内的单调性给出 $q_i(t-r)/q_i(t-1) \le \theta^{-2}$ 以及 $q_i(t-2r)/q_i(t-1) \le \theta^{-3}$。将两个首次相遇估计分别除以 $q_i(t-1) > 0$，得到

$$b(1 - v\theta^{-3}) \;\le\; \Pr_{i}(\tau = t \mid \tau > t-1, A) \;\le\; a\theta^{-2}.$$

两个端点均为 $(1 + o(1))/n$，且在 t 中均匀分布。由于 $q_i(2r-1) \ge \theta^{2}$，将条件生存概率相乘可得 $\log q_i(t) = o(1) - (1 + o(1))t/n$，对于 $t \ge 2r$ 成立。在 $t = \lfloor cn \rfloor$ 处，这给出 e^{-c} 存活；在 $t = \lfloor \beta n \log n \rfloor$ 处，它给出 $n^{-\beta + o(1)}$。将 (2) 和 (3) 相加，并利用块内的单调性，可得 $(1 - o(1))\, n \le \mathbb{E}_{W}[\tau \mid A, i] \le 2r + (1-\lambda)^{-1} = (1 + o(1))n$。这就证明了定理 2。

推论 7 写下 $p(A) = \lim_{t\to\infty} \Pr_{W}(W_{1:t}$ 不是重置词 $\mid A)^{1/t}$ 对于 *[13]* 的同步参数。那么 $n(1 - p(A)) \to 1$ 依概率成立。

证明 对于固定的表，通过对起始对取并集，极限内的概率介于 $\|Q^{t}\|_{\infty} = \max_i q_i(t)$ 和 $\binom{n}{2}\|Q^{t}\|_{\infty}$ 之间。谱半径公式给出 $p(A) = \rho(Q)$。在固定 n 的联合良好事件上，对于一个指定的起始对取根，并令 $t \to \infty$ 代入式 (2)；对于所有起始对则代入式 (3)。这给出 $(1 - u/(1-\gamma))^{1/r} \le p(A) \le \lambda$。两个端点都是 $1 - (1 + o(1))/n$，因为 $n \to \infty$。 ∎

附录 B。从终端矩重构

B.1. 从终端标签进行多项式学习

我们首先给出 $m = n$处的固定调度构造。之后的置信度细化使用相同的可观测数组。下面的所有样本都是长度为 n 的独立均匀词，并附有各自的终止符号。起始状态 s在目标被抽样之前就已固定。带标签目标的良好事件将独立于准确度和置信度参数 ϵ， δ。

令 $T_a(x, f_a(x)) = 1$ 为字母矩阵，其作用于列测试函数上为 $(T_a g)(x) = g(f_a(x))$，并令 $P = (T_0 + T_1)/2$。对于一个词 $v = v_1 \cdots v_l$，记作 $T_v = T_{v_1} \cdots T_{v_l}$。在本附录中， Q 是定义在 $N = n(n-1)$ 个有序不同点对上的公共词核，在相等时取零。其有序条目由引理 5 中相应的无序条目界定。令

$$d = \lceil 10 \log_2 n \rceil, \qquad r = 4d + 1, \qquad w = 2d + 1, \qquad R = 3w + d.$$

随机有向图以高概率具有唯一的闭类 $\mathcal{K}$ 以及满足下式的平稳行分布 π

$$\max_x \pi_x = n^{-1+o_{\Pr}(1)}, \qquad \min_{x \in \mathcal{K}} \pi_x \geq n^{-2}. \tag{4}$$

这里的最小值取遍正的平稳质量。上述事实由 [14, 定理 2]推出；对于二元出度，控制最小正质量的指数严格小于二。

B.1.1. 四个源与中心化有序对块

使用第 A.1 节的首次查询揭示滤过。一个返回 是一个目标顶点已被指定的新单元；读取已揭示单元不产生额外开销。在每一步内按固定顺序处理互异单元，并立即指定其目标顶点。在给定任何具体历史的条件下，未查询单元是相互独立的均匀状态。

引理 8（四个新源）假设四个互异顶点没有已揭示的出向单元，并且有 m 个指定顶点，其中包括这四个顶点。路径 *1* 和 *2* 接收一个长度为 l 的均匀词 W，路径 *3* 和 *4* 接收一个独立的长度为 l 的均匀词 W'。成功意味着最终的有序端点满足 $1 = 3$ 和 $2 = 4$，且两个端点互异。在给定任何具体先前历史的条件下，至多再返回两次即成功的概率为

$$\sum_{k,h=0}^{l-1} 2^{-\max(k,h)} \frac{(n-m)_{4l-k-h-2}}{n^{4l-k-h}}. \tag{5}$$

对于 m、$l = O(\log n)$，成功的总概率至多为 $6/n^2 + O((m+l)^6/n^3)$。对于四个互异的起始顶点且没有先前查询，（5）中的贡献至少为

$$\frac{6 - (4l+6)2^{-l}}{n^2}\big(1 - O(l^2/n)\big).$$

证明 首先将四个源声明为前沿。在之后的每个时刻，前沿由上一步中新命名的已占用顶点组成。前沿中每个顶点的所有出边

均未查询。首先公开前沿提议；如果在一个前沿顶点请求了两个字母，则计为一次分裂。如果 F 是前沿大小，那么

$$F_{\text{next}} = F + \#\{\text{分裂}\} - \#\{\text{从前沿返回}\} + \#\{\text{来自其他源的新目的地}\}. \tag{6}$$

每个新命名的目标顶点在其步骤结束时都已被占用。一次成功匹配以$F_{\text{end}} \le 2$结束。从四个顶点出发，并将（6）中的返回次数限制为至多两次，因此必然有$F_{\text{end}} = 2$，恰好两次前沿返回、没有分裂，也没有来自其他源的新发现。

在这些限制下，离开前沿的路径无法重新进入前沿。一条暴露的弧指向一个旧状态，而来自前沿之外的源所产生的新单元将导致额外的返回，或导致来自此类源的新发现。由于四条路径最终都在前沿上结束，因此它们从未离开前沿。于是，每次返回都会命中同一步骤中较早生成的目标顶点。恰好存在两次同步连接，且此后都不再分裂。指定的最终匹配迫使这些连接为$1 = 3$ 和 $2 = 4$。

设 k、h 分别表示两个连接之后的字母数。、的最后 $\max(k,h)$ 个字母 W、W'必须一致，这是一个概率为 $2^{-\max(k,h)}$的事件。对于相容的固定词，共有 $4l - k - h$个互异的单元查询，其中两个满足指定等式，其余所有目标顶点都必须是不同的新状态。这给出（5），即使两个连接发生在同一步骤也成立。至少出现三次返回的概率至多为

$$\binom{4l}{3}\left(\frac{m+4l}{n}\right)^3$$

通过序贯条件化，并在必要时填充查询列表。最后，$\sum_{k,\ h<l} 2^{-\max(k,h)} = \sum_{j<l}(2j+1)2^{-j} = 6 - (4l+6)2^{-l}$。当 $m = 4$时，下降阶乘与其首项幂的相对差为$O(l^2/n)$，从而证明了下界。 ∎

令 G 表示引理 3中使用的事件：每次广度优先探索至深度 $\ell = \lfloor(\log_2 n)/5\rfloor$时，发现树之外至多有一条弧。于是 $\mathbb{P}_A(G^c) = O(n^{-1/5})$。若从每个已命名顶点出发，仅沿其已暴露单元行进 d 步的概率至多为n^{-3}，则称部分表是安全的。引理 3 表明，在 G 上，以下所有部分表都是安全的。安全性根据已揭示历史进行检查，同时保留未使用单元的乘积分布。

引理 9（获取四个源） 从一个安全的部分表开始，在每个延续上都有 O（对数 n）个指定顶点。在 $w = 2d + 1$ 步内，两个副本均存活但其四个最终顶点并非互异的未查询源的概率为 O（（对数 n）$^2/n$）。这些副本最初可能共享顶点。

证明 窗口内任意位置的一次返回的代价为 $O(w \log n/n)$。在其补集上，经过 d 步后仍位于初始命名集合中的坐标，在整个过程中都经过了已揭示单元；安全性界限表明，对于四个坐标中的每一个，该事件的概率至多为 n^{-3}。路径离开该集合后，新顶点都具有唯一的带标签入边，因此无法重新进入该集合，也无法再次访问外部顶点。如果一个终点顶点与另一条路径当前或此前访问过的外部顶点重合，则唯一入边链会将这两次访问都追溯到同一条退出弧。终止于该处的路径至少又经过了 d 步，因此两次访问具有相同的最后 d 个字母。对于不同副本，这些子串相互独立。对于同一副本内的路径，沿同一条弧同时退出会使该路径对失效；不同的退出时间则迫使同一词的偏移子串相等。在固定偏移量下，这一事件的概率同样为 2^{-d}，其中包括

重叠子串，根据引理 4中的子串双射。对六个坐标对以及$O(d^2)$偏移量求和。该并集界使用固定偏移量，而不根据实际退出时间对词比特进行条件化。每个此前查询过的源都已被一条路径占据，因此同样的论证也涵盖了窗口内此前查询过的终点顶点。总代价为 $O(w \log n/n) \ + 4n^{-3} + O(d^2 2^{-d})$。

■

Put

$$K = Q^R, \qquad a = N^{-1}\mathbf{1}^\top K, \qquad B = K - \mathbf{1}a.$$

除非另有说明，以下所有矩阵范数均为欧几里得算子范数或弗罗贝尼乌斯范数。

引理 10（中心有序块）以高概率，

$$\|B\|_{\mathrm{op}} \le \|B\|_{\mathrm{F}} \le n^{-1/20}, \qquad \|a\|_2 \le \frac{\log n}{n}.$$

事实上 $n\|a\|_2 = O_{\Pr}(1)$。

证明 对于一个有序对 i，其平方行范数 $\|K_i\|_2^2$ 是两个从 i开始、位于同一张表上的独立复制词之间发生有序终端匹配的条件概率。尝试引理 9 的至多三个窗口，在死亡或首次获取四个新源时停止，并中止不安全的尝试。如果 $\eta = O\left((\log n)^{\,2}/n\right)$，则根据迭代条件化，在两对均存活的情况下，三次不成功的安全尝试所付出的代价至多为 η^3。停止规则仅使用已揭示历史。我们首先在原始分布下界定这一截断实验，然后插入 $\mathbf{1}_G$，因为截断实验与实际实验在 G上一致。获取之后，剩余长度介于 d 和 R之间。引理8 适用于实际的已揭示历史。在 G 上不存在安全中止，因此，对 i，均有，

$$\mathbb{E}[\mathbf{1}_G \|K_i\|_2^2] \le \frac{6}{n^2} + O\!\left(\frac{(\log n)^6}{n^3}\right). \tag{7}$$

所有条件 此处的界均基于可观测历史，而非在 G 条件下对未使用单元进行条件化。

该量 $\|a\|_2^2$ 则从两个独立均匀有序不同点对出发生成副本。这四个源互异的概率为 $1 - O(1/n)$。考虑长度为5的式 (R) 中所计数的两次返回轨迹。每条轨迹都会在底层无向多重图中生成一片森林：四条路径按指定点对连接，并共享其延续。给定词及答案后，所有未查询单元仍相互独立且均匀分布。

对每个根节点，以这片完整森林初始化一次辅助探索；如有必要，将该根节点作为孤立顶点加入其中。若根节点的广度优先搜索中出现两条多余弧，则辅助多重图的圈秩至少为二。从森林出发，这需要两个命中指定顶点的新探测。至多有 $N_0 = 2^{\ell+1}$ 个新探测，且每个探测的条件命中概率至多为 $(N_0 + 4R + 5)/n$。对根节点应用并集界，可得一致的相对估计

$$\mathbb{P}_A(G^c \mid \text{two-return transcript}) \le n\binom{N_0}{2}\left(\frac{N_0 + 4R + 5}{n}\right)^2 = O(n^{-1/5}).$$

因此，引理8 给出

$$\mathbb{E}[\mathbf{1}_G \|a\|_2^2] \ge \frac{6 - O(n^{-1/5} + R^2/n + R2^{-R})}{n^2}. \tag{8}$$

对于每个固定的表，方差恒等式为

$$\|B\|_{\mathrm{F}}^2 = \sum_i \|K_i\|_2^2 - N\|a\|_2^2.$$

这两个矩估计使用同一个事件 G，且具有相同的首项 $6/n^2$。从 (8) 的和中减去 (7)，得到 $\mathbb{E}[\mathbf{1}_G\|B\|_{\mathrm{F}}^2] = O(n^{-1/5} + (\log n)^6/n)$。马尔可夫不等式证明了第一个界。由詹森不等式和 (7) 得 $n\|a\|_2 = O_{\Pr}(1)$，因此得到关于 a的所述高概率界。 ■

B.1.2. 混合、历史秩与可观测乘积

块 K 与交换一对端点的操作可交换，其平均行 a 是对称的。对于 $f\colon [n] \to [-1,1]$令 $d_f(x, y) = f(x) - f(y)$。在相等处令其为零，得到

$$(Q^l d_f)(x,y) = P^l f(x) - P^l f(y).$$

在反对称函数上，K 的作用为 B，且保持反对称性。由于 $\|d_f\|_2 \le 2n$，引理 10，再对 y 关于 π取平均，便可同时对每个 $k \ge 1$

$$\sup_x \|P^{kR}(x,\cdot) - \pi\|_{\mathrm{TV}} \le 2n^{1-k/20}. \tag{9}$$

该界 限包括从 $\mathcal{K}$外部开始的情形；通过全变差收缩，后续步骤仍保持这一性质 .

记 $\mu_v = \pi T_v$ 并定义

$$S_l = \mathbb{E}_{|v|=l}\mu_v^{\top}\mu_v.$$

设 u_l 为 S_l 和 $H_l = u_l\mathbf{1}$的有序非对角元素组成的行。S_l 的两个边缘分布均为 π，且 $u_l = u_0Q^l$。对于每个固定的 C，对 $l \le CR$一致地有

$$H_l = 1 - o_{\Pr}(1). \tag{10}$$

事实上，一个指定不同对只有在一次新的返回之后，才会在 CR 步内相遇；其退火平均概率为 $O((CR)^2/n)$。以高概率，有序对上的这些相遇概率之和至多为 $n(\log n)^4$。其来自表选初始分布 $\pi\otimes\pi$ 的贡献由$\|\pi\|_\infty^2 n(\log n)^4 = o_{\Pr}(1)$ 利用（4）可得界。初始对角线的质量至多为 $\|\pi\|_\infty$。这就证明了（10），而无需假设 π 与该表相互独立。

在（3）之前的不等式，由线性性可推广到互异对上的每个非负初始子分布。结合（10），可得

$$H_l - H_{l+1} = (1 + o_{\Pr}(1))/n \tag{11}$$

在每个固定倍数 $l = kR \ge 2r$。这里只需要这些首次相遇估计；其证明见第A.1–A.2节。

Fix

$$t = 4096R, \qquad h = t+1, \qquad x_t = H_{t+1}/H_t, \qquad q = 1 - x_t.$$

为了比较 u_t 和 u_{t+1}的形状，将各行归一化 $p_i = u_{iR}/H_{iR}$ 和 $p_i' = u_{1+iR}/H_{1+iR}$。二者均按 $\mathcal{L}(y) = yK/(yK\mathbf{1})$更新。这些序列中的所有分母都大于 $1/2$，根据 (10)。引理 10 给出

$$\|\mathcal{L}(y)\|_2 \le 2\log n/n + 2n^{-1/20}\|y\|_2.$$

经过 40 个块后，两个序列的范数均至多为 5 log n/n。对于概率行 x， y，令 $z = \mathcal{L}(y)$。精确差值为

$$\mathcal{L}(x) - \mathcal{L}(y) = \frac{(x-y)B(I - \mathbf{1}z)}{xK\mathbf{1}}.$$

由于 $\|\mathbf{1}z\|_{\mathrm{op}} = \sqrt{N}\|z\|_2$，对于足够大的 n，后续更新至多将这些差异缩小 $n^{-1/25}$。转换为绝对值元素范数至多带来 $\sqrt{N} \le n$ 的代价，因此

$$\|u_{t+1} - x_t u_t\|_1 \le 2n\, n^{-(4096-40)/25}.$$

结合（11）以及 S_l的相等边缘分布，可得

$$q \asymp n^{-1}, \qquad C = \frac{S_{t+1} - x_t S_t}{q}, \qquad \|C - \mathrm{diag}(\pi)\|_{\mathrm{ent}} \le n^{-150}. \tag{12}$$

这里 $\|M\|_{\mathrm{ent}} = \sum_{x,\ y} |M$（$x$， y）$|$。为证明最后一个断言，非对角绝对质量至多为 $\|u_{t+1} - x_t u_t\|_1/q$。由于 $C\mathbf{1} = \pi^\top$，对角误差至多也是同一数量。特别地， S_t 的半正定性与（4）蕴含，在核心坐标上，

$$S_h \succeq cn^{-3} I \tag{13}$$

对于某个绝对常数 $c > 0$。

对于有界测试函数 g_0， g_1 以及一个词 v，令

$$m_v(g) = \pi T_v g, \qquad J_l(v; g_0, g_1) = \mathbb{E}_{|w|=l} m_{wv}(g_0) m_{wv}(g_1).$$

公式（12）给出了带误差的一致恒等式

$$\left| m_v(g_0 g_1) - \frac{J_{t+1}(v; g_0, g_1) - x_t J_t(v; g_0, g_1)}{1 - x_t} \right| \le n^{-150}. \tag{14}$$

事实上， $J_l = (T_v g_0)^\top S_l (T_v g_1)$，并且 $T_v(g_0 g_1) = (T_v g_0)(T_v g_1)$逐元素成立。

利用单个随机接受掩码可以观测比值 x_t。令 ξ 为其符号向量，并设

$$B_l = \mathbb{E}_{|w|=l} m_w(\xi)^2 = \xi^\top S_l \xi.$$

给定该表，$\mathbb{E}_\xi B_t = \mathrm{tr}\, S_t = 1 - H_t$。由于 $0 \le B_t \le 1$， （10）和马尔可夫不等式推出 $1 - B_t \ge 1/2$ 在概率趋于一的带标签目标事件上成立。由于 $\xi_x^2 = 1$， （12）进而得到

$$\left| \frac{1 - B_{t+1}}{1 - B_t} - x_t \right| \le 2qn^{-150}. \tag{15}$$

B.1.3. 双分支特征格拉姆矩阵

设 $j = r$。用长度为-$j + 1$的词对 $\omega = (w_0,\ w_1)$ 为一个特征编索引，并设

$$Z_\omega(x) = \xi_{f_0 w_0}(x) \xi_{f_1 w_1}(x), \qquad k = 2^{2(j+1)}.$$

如果两个物理状态的有序后继行（$f_0(x)$ $f_1(x)$）一致，则它们属于同一个行类。在 $\mathcal{K}$ 上，每个行类选取一个代表元，并令 V 的条目为 $V(x,\ \omega) = Z_\omega(x)/\sqrt{k}$。特征在同一个行类上为常数。

引理 11（双分支格拉姆界）以表和符号为条件，其概率趋于 *1*，

$$VV^{\top} \succeq (1/2 - o_{\Pr}(1))I.$$

证明 对于均匀分布的长度为 l 的词 W，记

$$D_l(x,y) = \Pr_W(f_W(x) = f_W(y) \mid A), \qquad G_l(x,y) = \mathbb{E}_W \xi_{f_W(x)} \xi_{f_W(y)}.$$

给定 A，$G_l - D_l$ 是独立符号的中心二次型。对于 $x \neq y$，对称系数矩阵的弗罗贝尼乌斯范数平方至多为 $(\log n)^{12}/(2n^2)$，在 $l = j$处由引理 5得出。在 $j+1$处也有相同的界，这是在第一步之后应用 j 步界得到的。其算子范数至多为其弗罗贝尼乌斯范数。由 Hanson–Wright 不等式 [15]以及对根节点应用并集界，可得

$$\max_{x,y} |G_l(x,y) - D_l(x,y)| \le C \log^7 n/n =: \eta_n, \qquad l = j, j+1, \tag{16}$$

以高概率成立。根节点相同时，差为零。

接下来，我们记录所需的仅依赖表的暴露界。对于 $x \neq y$，令

$$m_{xy} = \prod_{a=0}^{1} D_{j+1}(f_a(x), f_a(y)).$$

对于互异的 x、y、z，在适用时，均有

$$\mathbb{E} m_{xy} = O(j^4/n^2), \tag{17}$$

$$\mathbb{E}[m_{xy} m_{yz}] = O(j^8/n^4), \tag{18}$$

$$\mathbb{E}\left[\mathbf{1}_{(f_0(x), f_1(x)) \neq (f_0(y), f_1(y))} \prod_{v \in \{0,1\}^2} D_j(f_v(x), f_v(y)) \right] = O(j^6/n^3). \tag{19}$$

所有这些估计都是在原始表上进行退火平均的，即在选取核心代表元之前。下面给出一个同时处理共享根节点的返回计数证明（18）。

为每个所需的相遇分别指定一个对比坐标：将 $+1$ 放在其第一条路径上，将-1 放在其第二条路径上，并使所有尾部词在各坐标之间相互独立。在一个顶点处，将这些带符号系数收集成一个行向量。在一个同步轮次中，新前沿顶点没有任何已查询的出向单元。其出向提议向量之和等于其当前系数行。如果它们的张成空间具有秩 p_l，且当前新行具有秩 r_l，那么 $p_l \ge r_l$。非返回提议会指向一个新状态。通过从该状态的最终系数中减去随后返回到同一状态的提议，可以恢复其向量。此前已暴露的弧不可能指向本轮中首次创建的状态。因此，如果本轮发生了 c_l 次新单元格回访，则下一前沿的系数满足

$$r_{l+1} \ge p_l - c_l \ge r_l - c_l. \tag{20}$$

该论证依次处理互异单元格并立即标记目标，因此允许同一源对应多个坐标。当所有必需的相遇都发生时，系数行处处为零。

对于（17），初始固定字母提案张成两个对比坐标，每个字母对应一个。要实现消失，至少需要通过（20）返回两次。对于（18），在每个字母块内，x、y、z处的系数行分别为 $(1,\ 0)$、$(-1,\ 1)$、$(0,\ -1)$。它们张成秩为二的空间。两个字母块占据互不相交的坐标，因此初始提案秩为四，至少需要四次返回。

对于 (19)，先揭示两个固定的首字母，每个字母各用一次对比，并令 c 表示这一层的返回次数。若 $c \le 1$，剩余的新鲜秩至少为 $2-c$。针对固定的第二个字母复制每个坐标。第二层的提议分裂为两个互不相交的坐标块，使该秩加倍。消失总共至少需要 $c + 2(2-c) \ge 3$ 次返回。若 $c = 2$ 且之后没有返回，则每个新命名的第一层目的地处的系数都必须已经为零。此时，两个非返回提议各自只能由同一首字母坐标中的相反提议抵消。这就迫使 $f_a(x) = f_a(y)$ 对两个 a 都成立，而这正是被排除的相同行情形。情形$c \ge 3$ 已经具有所需的返回次数。每次计算使用 $O(j)$ 次查询和状态。序贯条件化将 q 次返回的概率限制在 $O(j^{2q}/n^q)$，从而证明 (17)–(19)。

考虑对角线为 1 且非对角线元素为 m_{xy} 的对称矩阵，并将其限制在互异的行类上。以高概率，边 $m_{xy} > n^{-1/3}$ 构成一个匹配。事实上，根据（18），任意两条相接的此类边同时出现的概率至多为 $n^3 n^{2/3} O(j^8/n^4) = o(1)$。其余元素的平方和至多为

$$n^{-1/3} \sum_{x \neq y} m_{xy} = o_{\Pr}(1)$$

根据（17）。

对于 $v \in \{0,1\}^2$，令 $d_v = D_j(f_v(x)\ f_v(y))$。将（19）对所有行不等的行对求和可知，它们的乘积之和 $\prod_v d_v$ 依概率趋于零。因此，对这些行对一致地有：

$$\min_v d_v \le \left(\sum_{\text{unequal rows } x,y} \prod_v d_v \right)^{1/4} = o_{\Pr}(1).$$

由于 $D_{j+1}(f_a(x)\ f_a(y)) = [D_j(f_{a0}(x)\ f_{a0}(y)) + D_j(f_{a1}(x)\ f_{a1}(y))]/2$，一条分支至多为 $1/2 + o_{\Pr}(1)$，另一条至多为 1。因此，每条重边的权重至多为 $1/2 + o_{\Pr}(1)$。支集为匹配的矩阵可分解为二乘二块，因此结构矩阵的最小特征值至少为 $1/2 - o_{\Pr}(1)$。

实际的 Gram 矩阵元素为 $\prod_a G_{j+1}(f_a(x)\ f_a(y))$。根据 (16)，其非对角线误差至多为$\eta_n(d_0 + d_1) + \eta_n^2$，此时 $d_a = D_{j+1}(f_a(x)\ f_a(y))$。从指定的不同行根出发要相遇，必须发生返回，因此

$$\sum_{x \neq y} D_{j+1}(f_a(x), f_a(y))^2 = O_{\Pr}(nj^2).$$

为得到此界，使用 $D^2 \le D$ 和单次返回概率 $O(j^2/n)$，其中包括初始的固定字母查询。因此，弗罗贝尼乌斯扰动平方为 $O_{\Pr}(\eta_n^2 nj^2 + n^2\eta_n^4) = o_{\Pr}(1)$。对角线 Gram 矩阵元素恰好为 1。由此得证该引理。 ∎

B.1.4. 估计可观测数组

令 c_* 为 $\mathcal{K}$ 上的行类数量，M 为它们的物理状态指示矩阵。M 的每一行都恰有一个 1；其各列互不相交且非空。令

$$p = 2^h, \qquad d_h = 2^{-\lceil h/2 \rceil}, \qquad F(u,\cdot) = d_h \pi T_u M \quad (|u| = h).$$

由于 $1/2 \le p d_h^2 \le 1$，13式给出

$$F^\top F = p d_h^2 M^\top S_h M \succeq c' n^{-3} I.$$

行类具有定义良好的确定性字母转移 U_a。元素为$b_a([x]) = \xi_{f_a(x)}$ 的向量 b_a 也定义良好。具有相同后继行的状态可能具有不同的当前符号；它们的状态转移和这些末字母检验相同。定义

$$H = FV, \qquad H_a = FU_aV, \qquad g_a = Fb_a. \tag{21}$$

引理11和（13）推出

$$s_* := \sigma_{\min,+}(H) \ge n^{-2} \tag{22}$$

当 n足够大时。（21）中的所有五个数组的范数至多为一：每个矩阵元素的绝对值不超过 $d_h/\sqrt{k}$，每个向量元素的绝对值不超过 d_h，且 $p d_h^2 \le 1$。

对于每个特征 $\omega = (w_0, w_1)$，将（14）应用于 $g_0(x) = \xi_{f_{0w_0}(x)}$和 $g_1(x) = \xi_{f_{1w_1}(x)}$。使用 $v = u$和 $v = ua$ 分别恢复 H 和 H_a的元素。两种情况的归一化均为 $d_h/\sqrt{k}$，包括长度为$h+1$的词 ua。普通均值 $m_{ua}(\xi)$乘以 d_h后，可恢复 g_a。（14）右侧的每个因子都是在给定后缀条件下计算的普通终端均值。

下面给出一种有限样本实现。令

$$L = 2t + j + 5, \qquad N_1 = \left\lceil C 2^L n^{242} \big(L + \log(8/\delta)\big) \right\rceil.$$

对于长度不超过 L 的每个后缀，通过其所在分箱中的经验符号平均值来估计其终端均值；若分箱为空，则取零。此类分箱少于 2^{L+1} 个，且每个分箱的概率至少为 2^{-L}。给定分箱计数后，其中的符号是独立同分布的，其均值为相应的条件均值。由 Chernoff 不等式、Hoeffding不等式和并集界可知，以至少 $1 - \delta/2$ 的概率，同时误差不超过 n^{-120}。不同长度分箱之间无需相互独立。

若 $n \ge L + 2500R$，则每个后缀之前未观测到的填充至少有 $2500R$ 个字母。公式 (9) 将其条件终端均值与平稳均值之差限制在 $4n^{-124}$ 以内。若 $n < L + 2500R$，则改用完整的长度为 n 真值表：在此情形下，随着 $L + 2500R = O(\log n)$，其 2^n 个条目为多项式数量。通过优惠券收集过程，以置信度$1 - \delta$ 学得这些条目。

$\widehat{\ }$ $\widehat{\ }$ $\widehat{\ }$根据经验均值计算 B_t, B_{t+1}。如果 $1 - B_t < 1/4$，则返回一个常数。否则令 $\widehat{x}_t = (1 - \widehat{B}_{t+1})/(1 - \widehat{B}_t)$，若 $1 - \widehat{x}_t < n^{-2}$，则返回一个常数。在良好的目标事件和同时发生的采样事件下，当 n 足够大时，这两个守卫条件都不会触发。由等式 (15) 和采样界可得 $|\widehat{x}_t - x_t| = O(n^{-120})$。每个经验 J_l 的乘积误差为 $O(n^{-120})$。在 (14) 中，除以 q 会带来 $O(n)$ 的代价，而扰动比值带来的代价至多为其误差的 $O(n^2)$ 倍。将所得未归一化乘积估计裁剪至 $[-1, 1]$。其误差为 $O(n^{-118})$。

乘以 $d_h/\sqrt{k}$ 后，整个矩阵的弗罗贝尼乌斯误差至多为其最大的未归一化误差，因为 $p d_h^2 \le 1$。对向量应用同样的观察，并乘以因子 d_h，即可控制这些向量。因此，估计数组同时满足

$$\|\widehat{H} - H\|_{\mathrm{op}}, \quad \|\widehat{H}_a - H_a\|_{\mathrm{op}}, \quad \|\widehat{g}_a - g_a\|_2 \le \tau := n^{-100}. \tag{23}$$

这种归一化使固定历史长度 $h = t + 1$ 足以满足条件化和数值精度的要求。

B.1.5. 有理数运算中的稳定线性表示

F的满列秩和 V的满行秩意味着 $VH^{\dagger}F = I_{c_*}$。令 $K_a = H_a H^{\dagger}$。对于每个非空词 v,

$$K_v = FU_v VH^{\dagger}.$$

由于 $\|F\| \le 1$、 $\|U_v\| \le \sqrt{n}$、$\|V\| \le \sqrt{n}$和 $\|H^{\dagger}\| = 1/s_*$，每个真实字母乘积（包括空乘积）的范数至多为

$$B_* = n/s_* \le n^3.$$

对于核心行类 i，其读出$e_i^{\top} VH^{\dagger}$ 的范数至多为 $1/s_*$，因为相应的字典行的范数为一。对于长度为正的后缀 $z = z_1 \cdots z_l$，该读出作用于

$$K_{z_1} \cdots K_{z_{l-1}} g_{z_l} \tag{24}$$

返回其在类 i中的精确终止符号。

使用有理岭逆

$$\lambda = n^{-26}, \qquad R_\lambda(X) = X^{\top}(XX^{\top} + \lambda I)^{-1}, \qquad \widehat{K}_a = \widehat{H}_a R_\lambda(\widehat{H}).$$

奇异值分解给出 $\|R_\lambda（H）- H^{\dagger}\| \le \lambda / s_*^3$。预解式恒等式和（23）给出

$$\|R_\lambda(\widehat{H}) - R_\lambda(H)\| \le \tau/\lambda + 3\tau/\lambda^2, \qquad \|R_\lambda(\widehat{H})\| \le (2\sqrt{\lambda})^{-1}.$$

因此 $\|\widehat{K}_a - K_a\| \le a_* = O(n^{-20})$。在错误位置展开乘积，并对每个中间真实乘积使用 B_* 的界，可得长度为 l 时的误差上界为

$$B_*\big[(1 + B_* a_*)^l - 1\big].$$

估计的末端向量至多增加 $\tau B_*（1 + B_* a_*）^{l}$。与任意真实读出相乘的代价至多为 $1/s_*$。因此，总误差为 $o(1)$ 对 $l \le n$一致成立。特别地，当 n 足够大时，（24）的估计版本在真实读出下，对于每个后缀和每个核心行类，其带符号间隔至少为 $1/2$ 。

所有归一化都是有理数： d_h 是二的幂，且$\sqrt{k} = 2^{j+1}$。经验均值、矩阵乘积和岭逆都可以通过对多项式维度的数组执行多项式数量的有理运算来计算。有理数求逆具有多项式比特复杂度。对于词求值，取两个字母矩阵的公分母。至多包含 n 个矩阵的乘积，其比特长度至多为单个字母比特预算的 n 倍，再加上 $O(n\log\ p)$。保护条件、截断和正的岭正则化确保该过程在每个数据集上都有定义。这些观察结果给出了多项式比特复杂度以及多项式算术复杂度。

B.1.6. 拟合起始读出并证明风险界

选择前缀长度

$$b = R\left\lceil 20\left(1 + \log_n \frac{64}{\epsilon}\right)\right\rceil.$$

向下舍入 ϵ 到二进制有理数容差后，就可以通过将其与 n的幂进行整数比较来计算这个上界。在相同的表事件中，(9)给出

$$\Pr_{|u|=b}\left(f_u(s) \notin \mathcal{K}\right) \le \epsilon/8, \qquad 2^b \le (n/\epsilon)^{O(1)}. \tag{25}$$

如果 $b \geq n$，则通过优惠券收集学习完整真值表。其样本量为 $O(2^n(n + \log(1/\delta)))$，运行时间关于 n、$1/\epsilon$、$\log(1/\delta)$ 是多项式的，因为在此情形下 $2^n \leq 2^b$。

假设 $b < n$。对于非空后缀 z，定义有理特征向量

$$\widehat{\phi}(z) = \widehat{K}_{z_1} \cdots \widehat{K}_{z_{|z|-1}} \widehat{g}_{z_{|z|}}.$$

取第二个独立的样本批次，并按前 b 个字母进行划分。在每个前缀分箱中，求解有理线性不等式，要求线性读出 $\widehat{\ }(\phi)$ 的带符号训练间隔至少为 1。如果分箱为空或无可行解，则使用常数预测器。对于每个以 $\mathcal{K}$ 结尾的前缀，根据上述间隔界，其真实读出的两倍对所有样本都可行。因此，返回的预测器在这些良好分箱的固定并集上训练误差为零。

以第一批样本为条件，所有可能的读出构成一个半空间类，其中为 2^b 个前缀分别设有独立的块。加入一个仿射坐标即可涵盖常数回退情形，并使VC维至多为 $d_* = (p+1)2^b$。将误差指示函数乘以良好分箱的固定指示函数，不会改变这一VC界。这个受限误差类包含一个恒等于零的元素，而返回的预测器与之相容。因此，可实现情形的VC界表明

$$N_2 = O\left(\frac{d_* \log(2/\epsilon) + \log(8/\delta)}{\epsilon}\right)$$

样本足以使良好分箱上的风险至多为 $\epsilon/2$，其置信度至少为 $1 - \delta/2$。根据（25），其他分箱的总质量至多为 $\epsilon/8$。学习器不需要知道它们的单位算子或任何隐藏状态。

两批样本以及所有计算的规模均为关于 n、$1/\epsilon$和$\log(1/\delta)$的多项式。它们的两个失败概率之和至多为 δ。在固定的有限初始范围上采用真值表过程可保持这些界。证明中使用的所有目标事件仅取决于该表和标签，其概率为 $1 - o(1)$；因此，对于每个 ϵ，在 δ 具有公平接受符号时，相同的目标事件均成立。附录C.7给出了推广到每个接受集合以及每个固定目标置信度指数的结果。特别地，令 $\epsilon = n^{-\alpha}$，即可对每个固定的 $\alpha > 0$获得多项式复杂度，且目标事件与样本的联合失败概率至多为 $\delta + o(1)$。

B.1.7. 原始调度的多项式异常目标界

推论12 存在一个绝对常数 $c > 0$，使得对于每个固定的整数 $D \geq 1$，附录 *D.3.1* 中的原始调度会在目标抽取之前确定一个两批次算法 $\mathcal{A}'_D$。该算法允许使用良好目标集合$\mathcal{G}_{n,\,D}^{\mathrm{qt}}$，满足

$$\Pr_{A,b}\left((\mathcal{G}_{n,D}^{\mathrm{qt}})^c\right) = O_D(n^{-c}).$$

这里的接受位是相互独立的公平比特。对于每个指定起点以及 $m \leq n^D$、$\mathcal{A}'_D$，仅给定参数和带终端标签的词时，其条件风险至多为 ϵ，训练概率至少为 $1-\delta$；所需样本数和位运算次数关于 n、$1/\epsilon$、$\log(2/\delta)$ 均为多项式（对固定的D而言）。同一目标事件涵盖所有这些长度和精度，因此目标和样本上的联合失败概率至多为 $O_D(n^{-c}) + \delta$。

证明 我们保留原始调度的所有窗口长度、精度参数和拟合规则。证明使用较弱的平稳质量界以及不同的平均有序对行界，同时保留该算法所需的数值阈值 $\sigma_{\min,\,+}(H) \geq n^{-2}$。以下所有估计均适用于充分大的 n；固定的初始范围则由真值表约定覆盖。

中心化块与混合。 设 G为引理中使用的广度优先搜索事件10。其展示的截断矩估计给出

$$\Pr(G^c) = O(n^{-1/5}), \qquad \mathbb{E}[\mathbf{1}_G \|B\|_F^2] = O(n^{-1/5} + n^{-1}\log^6 n), \qquad \mathbb{E}[\mathbf{1}_G \|a\|_2^2] = O(n^{-2}).$$

最后一个断言由式（7）根据詹森不等式得出。因此，事件

$$G, \qquad \|B\| \le n^{-1/16}, \qquad \|a\|_2 \le n^{-1+1/100} \tag{26}$$

发生的概率至多为

$$O(n^{-1/5} + n^{-3/40} + n^{-7/8}\log^6 n + n^{-1/50}).$$

Q^R 在反对称检验函数上的作用等同于 B的作用。因此，对于 P的任意平稳分布 π，

$$\sup_x \|P^{kR}(x,\cdot) - \pi\|_{\mathrm{TV}} \le 2n^{1-k/16}. \tag{27}$$

这也证明了 π的唯一性、从每个状态出发的收敛性，以及其闭不可约支集 $\mathcal{K}$的唯一性。无需单独估计巨型常返类的概率。

具有多项式尾部的平稳最小值。 用 $\eta_2 < 1$表示二进制直径定理中的常数，该定理由 [14], 提出，并令 $a_0 = (1-\eta_2)/2 > 0$。其定理 1 的证明（第4.1节，第 10–12 页）给出：对于某个 $\gamma_0 > 0$,

$$\Pr(\mathrm{diam}(A) > (1+\eta_2+a_0)\log_2 n) = O(n^{-\gamma_0}).$$

取 $\varepsilon_0 = \min\{a_0/2, 1/20\}$。 [14] 中的引理 4 在这一固定松弛量下给出的逐顶点失败概率为 $O(n^{-1-\delta_{\mathrm{ABP}}})$。接着，命题 5 以及第 12 页上的并集界允许我们取 $\gamma_0 = \min\{1, \delta_{\mathrm{ABP}}\} > 0$。取$x$ 使得 $\pi_x \ge 1/n$。对于每个 $y \in \mathcal{K}$，都存在一条从 x到 y 的路径，其长度至多为该直径。沿路径的平稳性（其概率至少为 $2^{-\mathrm{diam}(A)}$）因此给出，当 $\zeta = (1-\eta_2)/2$ 时，

$$\pi_y \ge \tfrac{1}{2} n^{-3+\zeta}. \tag{28}$$

此论证仅使用直径尾部和（27）；它不需要平稳极限定理的尾部速率。

通过独立预热实现平稳历史的存活。 令 $t = 4096DR$ 和 $t_{\max} = t+1$。对于每个固定的表，取两个独立均匀状态作为起点，分别用长度为$128R$的独立词驱动它们，然后再用一个长度为 $t_{\max}$ 的共同词驱动它们。令 C_A 为在此实验中两者端点相同的概率。在随机表上揭示这两条路径会用到 $O_D(\log n)$ 个单元格和状态。初始状态相同的概率为 $1/n$；否则，最终状态相同要求一个新鲜单元格的目标命中某个先前出现过的状态。因此，序贯条件化给出

$$\mathbb{E}_A C_A = O_D(\log^2 n / n). \tag{29}$$

在（26）中，两个状态分别经过独立预热后的分布，根据（ $2n^{-7}$ 与 π，由（27）可知，与 $4n^{-7}$ 相差不超过 $\pi \otimes \pi$。应用公共词核并观察相等性会保持这一全变差界。因此

$$1 - H_{t_{\max}} \le C_A + 4n^{-7}.$$

将马尔可夫不等式应用于（29）即可证明

$$H_l \geq 1 - n^{-1/4} \quad (0 \leq l \leq t_{\max}) \tag{30}$$

此外，还有一个概率为 $O_D(n^{-3/4}\log^2 n)$ 的事件。 H_l 的单调性使得这在 l 中成为同时发生的一个事件。特别地，该估计考虑了 π 对转移表的依赖性。

首次相遇损失与归一化收缩。 引理5的高阶矩证明中， $k = \lceil (\log n)^3 \rceil$将其另外两项失败界定为 $n^2 (b_n/a)^k$和 $n^4 (n^2 b'_n/L)^k$。二者都以快于 n的每个逆幂的速度递减；已计入的 G 失败是剩余项。引理 6 证明中展示的平方偏差估计将其基数断言的失败界定为

$$O(n^{-1/10}\log^2 n + n^{-9/10}\log^{12} n).$$

其热核不等式和首次相遇不等式根据线性性适用于任何非负初始子分布，包括依赖于转移表的平稳非对角分布。结合（30），在时刻 $t+1$ 它们给出

$$q = 1 - H_{t+1}/H_t \asymp n^{-1}.$$

两条归一化历史序列中的所有分母均大于 $1/2$。由（26）得到的范数递推为

$$\|\mathcal{L}(y)\|_2 \leq 2n^{-1+1/100} + 2n^{-1/16}\|y\|_2.$$

经过 40 个块后，范数至多为 $5n^{-1+1/100}$。因此，原证明中的精确差值恒等式给出了收缩因子 $Cn^{-1/16+1/100} \leq n^{-1/25}$ ，适用于所有足够大的 n。所以，原来的$4096D$ 个块仍然意味着

$$\left\| \frac{S_{t+1} - x_t S_t}{q} - \operatorname{diag}(\pi) \right\|_{\text{ent}} \leq n^{-150D}. \tag{31}$$

In事实上，在除以 q 之前的下界为 $2n\, n^{-(4096D-40)/25}$，从而保留了所需的余量。

对于公平接受符号， $\mathbb{E}_\xi B_t = \operatorname{tr} S_t = 1 - H_t \leq n^{-1/4}$。因此， $B_t \leq 1/2$ ，除非条件概率至多为 $2n^{-1/4}$。这在带标签目标事件上验证了相同的校准保护条件和精度界，并且失败概率为多项式级别。

特征条件化。 双分支 Gram 证明中给出的退火矩估计表明，在任意固定且足够小的数值阈值下，都具有如下速率。连接到同一顶点的两条重边具有概率 $O(j^8 n^{-1/3})$。轻元素的弗罗贝尼乌斯范数平方至多为 $n^{-1/3}\sum_{x\neq y} m_{xy}$，其期望为 $O(j^4 n^{-1/3})$。不等行上的孙节点碰撞乘积之和的期望为 $O(j^6/n)$。马尔可夫不等式使轻元素范数变小，并使每条重边至多为 $1/2$ 加上一个小常数，且这些结论具有多项式级别的失败概率。因此，匹配块论证为结构格拉姆矩阵给出了一个固定的正下界。

成对热核事件具有刚才记录的速率。在此事件上，条件 Hanson–Wright 不等式和并集界结合一个足够大的绝对常数，可得 $\max |G_l - D_l| \leq C\log^7 n/n$ ，失败概率为 $O(n^{-2})$。最后，一次返回估计给出

$$\mathbb{E}\sum_{x\neq y} D_{j+1}(f_a(x), f_a(y))^2 = O(nj^2).$$

要求此和至多为 $n^{3/2}$ 的代价是 $O(j^2 n^{-1/2})$。在该事件上，平方弗罗贝尼乌斯符号扰动至多为

$$O(n^{-1/2}\log^{14} n + n^{-2}\log^{28} n) = o(1).$$

选择足够小的固定结构阈值，可使 $VV^\top \succeq \frac{1}{4}I$ 位于此前各事件的多项式有界并集之外。在选择任何行代表元之前，所有界都对物理配对成立。

原有的数值调度仍然适用。 的半正定性 S_t，(31)，以及 $q \asymp 1/n$，和 (28)给出在核心坐标上的

$$S_{t+1} = x_t S_t + q\,\mathrm{diag}(\pi) + E_t \succeq c_1 n^{-4+\zeta} I, \qquad \|E_t\| \le q n^{-150D}.$$

行类的聚合和有理归一化意味着 $F^\top F \succeq c_2 n^{-4+\zeta} I$。因此

$$\sigma_{\min,+}(FV) \ge c_3 n^{-2+\zeta/2} \ge n^{-2}$$

当 n 足够大时成立。这正是原有岭正则化和词乘积误差估计所使用的数值下界。其填充界和前缀界也仍然有效，因为 (27) 强于它们所使用的混合界。因此，原有的样本和计算保证均保持不变。

证明中的每个目标事件都与 ϵ、δ 无关。取它们的并集，例如可以选择

$$c = \min\{\gamma_0/2, 1/100\} > 0.$$

这些常数可以依赖于固定的长度指数 D （见29）。这证明了所述目标界以及条件失败到联合失败的陈述。该论证针对原始调度；要将其平稳质量替换为（28），则需要对采用更强平稳最小值的较小调度进行单独的数值核算。 ■

B.2. 用于准确学习的单一转移表条件

在 $N = n(n-1)$ 个有序不同点对上，令 Q 为被杀死的公共字母转移矩阵。以下充分条件以更高的置信度隔离了目标概率问题。

定理 13 固定 A、D、c_R、$\kappa > 0$，令 $R = \lceil c_R \log n \rceil$（更一般地，使用明确指定的可比对数整数）、 $K = Q^R$ 和 $a = N^{-1}\mathbf{1}^\top K$。存在一个多项式时间终端标签学习器，以及概率为 $1 - o(n^{-A})$ 的带标签目标事件，使得在这些事件上，它能够在满足以下条件的每个表上成功

$$\|K - \mathbf{1}a\|_{2\to 2} \le n^{-\kappa}.$$

在任意指定长度 $m \le n^D$ 下，其误差至多为 ϵ，样本概率为 $1-\gamma$。所需资源关于 n、$1/\epsilon$、$\log(1/\gamma)$ 均为多项式，且指数固定。学习器仅接收 n 和带标签的词。将 κ 固定缩减后，转移条件也可以替换为 $\|K - (K\mathbf{1})a/(a\mathbf{1})\| \le n^{-\kappa}$。此处，在下文所用的事件上，$a\mathbf{1} > 0$。

证明 下文的掩码事件仅要求对满足块条件和所述几何估计的表成立。其他表自动满足这一蕴含。额外事件包括后缀字典定理、引理26中的同时碰撞界、第 61 页上指数为 $A+2$ 的有限距离界、一致短词纤维界，以及接受掩码校准。这些事件的选取均独立于样本、准确度和训练长度。

一致纤维。对于每个固定的 $T = O(\log n)$ 和每个固定的 B_0，除概率$O(n^{-B_0})$之外，每个长度至多为 T 的词的每个纤维，其大小均为 $M = (\log n)^{O(1)}$。一种证明方法是使用吸收器学习器中的幽灵根界。采样 $k = \lceil (\log n)^2 \rceil$ 个互异的均匀分布根节点。对于任意固定的词和目标，先标记所有根节点和目标，然后依次揭示 k 条完整路径。若所有端点都等于目标，则已揭示路径构成的无向图必须将 k 个根节点与目标连通。若目标与这些根节点互异，则在 $O(kT)$ 个探针/状态中，需要 k 次新鲜返回，以连接到已命名的连通分量。若某个根节点就是目标，那么初始重合的代价为 k/n，并且还需要 $k-1$ 次返回。这个连通分量论证也将先前查询过的弧的重复使用计算在内；一次新鲜返回至多能连接两个连通分量。因此，纤维 Z 的阶乘矩为 $\mathbb{E}(Z)_k/(n)_k \le [C(k(T+1))^2/n]^k$，其中已增大绝对常数，以涵盖根节点重合的情况。对 $\log n$ 的一个足够大的固定幂次应用马尔可夫不等式，可将尾部界定为 $e^{-\Omega((\log n)^2)}$。对状态以及至多 2^{T+1} 个词取并集，即可证明该估计。对于调度所需的 $\log n$ 的任意固定倍数，该结论都成立。

因此，a的每个元素都至多为 M^2/N，且$\|a\|_2 \le M/\sqrt{N}$。为比较这两个表条件，令 $g = \mathbf{1} - Q\mathbf{1}$。同字母入度至多为 M，因此通过对具有相同目标的状态对数量求和，$\|g\|_2^2 \le nM$。每个确定性消亡词对映射的行和至多为一，并且在时间 R之前列和至多为 M^2。由舒尔界，$\|Q^l\|_{2\to 2} \le M$ 对于 $l \le R$。因此，$\|\mathbf{1} - K\mathbf{1}\|_{2\to 2} \le \sum_{l<R} \|Q^l g\|_2 \le R\sqrt{n}M^3$。由此可知 $\rho = a\mathbf{1} = 1 - o(1)$，并且两个中心化矩阵之间的秩一差为$(K\mathbf{1} - \rho\mathbf{1})a/\rho$，其范数为 $O(RM^{5/2}/\sqrt{n})$。对于任一条件，取固定的 $0 < \kappa' < \min(\kappa, 1/2)$作为新的普通中心化指数。

历史校准。对反对称差分 $f(x) - f(y)$ 的作用给出唯一的普通平稳分布，且最差行混合误差为 $2n^{1-k\kappa'}$ 经过 kR 步后。经过对数烧入期后，全变差误差为 n^{-3}，均匀初始分布的原子概率根据纤维界至多为 M/n。因此 $\pi_{\max} \le M/n + n^{-3}$。对于每个短历史 v，平稳行 $\mu_v = \pi T_v$ 的最大原子概率至多为 $M\pi_{\max}$。对每个所需的对数 t，历史碰撞质量 $\mathrm{tr}\ S_t$ 至多为 $M\pi_{\max} = o(1)$，因此当 n 足够大时，所有相关存活分母都至少为 $3/4$。对于 $B_t = \xi^\top S_t \xi$，公平掩码期望为 $\mathrm{tr}\, S_t$。翻转 y 处的比特，至多会使该统计量改变 $4\mathbb{E}_v \mu_v(y) = 4\pi_y$。因此，根据有界差分，除条件掩码概率超多项式地小的情况外，有 $B_t \le 1/2$。于是，同步碰撞引理给出

$q = 1 - H_{t+1}/H_t \ge c/[n(1+t+G)^4] \ge n^{-2}$，对任意固定的对数历史系数均成立，且指数不依赖于该系数。它仅使用普通的单游走混合和新鲜的最终槽位。直径估计给出 $\pi_{\min,\ +} \ge n^{-S}$，其中常数 S 取决于 A。

任意精度的对角线提取。附录中的归一化历史更新D.2.3具有与 $\|a\|_2 \le M/\sqrt{N}$相同的差分公式。经过固定数量的块后，其行的范数为 $O(M/n)$；之后的差分至多收缩到原来的 $n^{-\kappa'/2}$。在 $t = kR$处，将非对角差异乘以 n 以改变范数，再除以 $q \ge n^{-2}$，代价至多为 n^3。对角线的匹配使用两个平稳边缘分布。因此，对于任意固定的 K_0，都可以选择 $k = O_{K_0,\ \kappa'}(1)$，使 $\operatorname{diag}\pi$ 在绝对值元素范数下逼近至 n^{-K_0}。此外，一旦 $K_0 > S+4$，由 S_t 的半正定性，在循环状态上也有 $S_{t+1} \succeq cn^{-(S+2)}I$。此下界中的指数与所选的 k无关。

使用定理 25的字典，并令 M_0 为循环支集内预测类的隶属矩阵。由于 $M_0^\top M_0 \succeq I$，数组 $H = FV$（其中 F与原估计量中的相同）的最小正奇异值至少为 n^{-C}，对于固定的 $C = C(A,\ S)$，且与$K_0,\ k$无关。读出和真实的词乘积范数因而都有多项式界。选择固定的$P_* > 20(C+D+1)$。附录 D.2.3 中的岭正则化和乘积展开只需要归一化数组精度 n^{-4P_*}，即可在长度达到 n^D时获得统一的带符号间隔 $1/2$。要在固定张量字典中达到该精度，使用附录 D.2的平衡乘积递归。其固定深度至多为 $d_* = \lceil \log_2 k_A \rceil$。在每一层，截断使因子保持有界，除法使用 $q^{-1} \le n^2$，而扰动比值产生的代价为其误差的 $O(n^4)$ 倍。取原始均值精度和对角线容差至多为 n^{-K_0}，对于 $K_0 > 4P_* + 8(d_*+2) + S$，根据归纳法即已足够。经验保护条件可以使用 $\widehat{q} \ge n^{-3}$；在该精度下，真实下界能够满足该条件。使用足够长的对数填充预热，以及长度为 $O_{A,\ D,\ c_R,\ \kappa}(\log n)$ 的后缀分箱。对于被分块的较短词，学习真值表。归一化字典行可以乘以一个与单位范数相差不超过两倍的公共二进制因子，同时保持所有数组运算为有理数。

预测等价的商具有确定性转移和终止输出。因此，原有的第二个独立样本面板和逐前缀线性可行性问题仍然适用，并采用相同的间隔证明和 VC 分析。对数级常规混合界也能提供所需的前缀长度，使非递归质量至多为 $\epsilon/8$。条件成功和资源界随之成立。本证明中的所有目标谓词在选定$A,\ D,\ c_R,\ \kappa$后，均使用固定的对数级系数。没有任何谓词要求每个起始对都具有正确的首次命中密度 $1/n$。 ∎

B.3. 行编辑下中心化成对核的集中性

更改转移表的一行可能影响杀死配对核中的许多项。我们对轨迹经过编辑行的起始对求和，从而界定其中心化块的弗罗贝尼乌斯范数平方的变化。这样便可应用独立表行的集中性。

引理14 设 Q_B 为二元表 B, $N = n(n-1)$, $J = I - \mathbf{1}\mathbf{1}^\top/N$, *and* $F(B) = \|JQ_B^R\|_F^2$的杀死有序对核。考虑满足以下条件的任意表类

$$\max_{i,j} Q_B^R(i,j) \le L/n^2, \qquad \max_{|v|<R,z} |f_v^{-1}(z)| \le M.$$

在这个类中，对于行汉明距离为 h 的任意两张表 B、C，

$$|F(B) - F(C)| \le 8RMLh/n.$$

证明 对两张表上的每一对起始状态，都使用同一字符串运行，并将墓地结果也包括在内。在活动坐标即将读取某个 h 个编辑行之前，它们的结果始终一致。对于每个固定的时刻 $j < R$ 和输入前缀，至多有 hM 个第一坐标起点会落在 B 中的某个编辑行上；第二起点至多有 $n-1$ 种选择。交换坐标后，同样的界也成立。去掉存活条件只会使计数增大。记 $K_B = Q_B^R$，由普通耦合不等式并对所有起始状态求和，得到

$$\sum_{i,t} |K_B(i,t) - K_C(i,t)| \le 4R(n-1)hM.$$

这两个矩阵 JK_B，JK_C的所有元素都属于 $[-L/n^2, L/n^2]$。由于 J是自伴投影，

$$F(B) - F(C) = \langle JK_B + JK_C, K_B - K_C \rangle_F.$$

这些界直接给出了结论，适用于两个端点，无论中间是否有表属于该类。

■

定理 15 对于每个固定的 A，$D > 0$ 存在常数 K_A，c_R 以及 $\kappa > 0$，并存在一个多项式终端标签学习器；除概率为 $o(n^{-A})$ 的带标签目标事件外，在原始独立同分布模型下，所有终端强连通分量大小均大于 K_A 的每张表，都能以规定误差 ϵ 在任意规定的 $m \le n^D$下学得，且样本失败概率至多为 γ。多项式关于 n，$1/\epsilon$，$\log(2/\gamma)$。特别地，对于这些表，定理 *13*的分块谓词在 $o(n^{-A})$之外成立。

证明 取 K_A 如定理 25所述，并增加一个单位以留出额外余量。在其向外搜索事件发生时，若不存在大小至多为 K_A的终端分量，则 $j = \lfloor (\log_2 n)/8 \rfloor$ 处的最大普通热项为 $O_A(n^{-b_A})$，且 $b_A > 0$成立。因此，从任一起始状态出发，在 d步内停留在任意给定的多对数规模命名集合中的概率，对于某个足够大的固定倍数 n^{-3}，至多为 d的 j；这是由连续各块末端处的马尔可夫性质得出的。此结论同样适用于仅沿该命名集合上任意部分表的条目进行跟随的情形。搜索失败的原始概率为 $o(n^{-A})$。

下面明确给出共享表矩步骤。取 $d = qj$，令 q 为一个大于 $\max(100, 8/b_A)$ 的固定整数，$r = 4d+1$、$k = \lceil (\log n)^3 \rceil$ 以及 $H = 2kr + 2k + 10$。在搜索事件和无小终端分层上，至多包含 H 个命名状态的每段部分历史都是安全的：仅沿着其已知条目行走 d 步，在所有 q 个块结束时都留在该集合中，其概率至多为 $(HC_A n^{-b_A})^q \le n^{-3}$。将此交集记为 E。对于固定的有序对 i、t，令 $Z = Q^r(i,t)$。从 i开始运行 k 个独立词，保留各词中的所有已暴露单元，并在第一次试验前命名 t。一次试验仅在每次尝试获取开始时因引理 5 的安全测试而被截断。给定完整的过往查询记录，未来字母和未使用单元服从其原始分布律。在存活条件下，两次获取失败的代价至多为 η^2，其中 $\eta = 4dH/n + 2n^{-3} + d^2 2^{-d}$。获取后，按照在最终字母之前严格发生零次、一次或至少两次返回进行分裂。因此，截断到达的条件概率至多为

$$b_n = \eta^2 + 2/n^2 + (2rH/n)(2/n) + \tbinom{2r}{2}(H/n)^2 = O_A((\log n)^{10}/n^2).$$

这些界在第一次测试中止时同样适用，此时成功概率为零。给定完整表，普通到达具有联合概率 Z^k。在 E 下没有试验中止；依据原始分布律对部分记录反复应用条件界，可得

$$\mathbb{E}[\mathbf{1}_E Z^k] \le \Pr(\text{all } k \text{ truncated trials arrive}) \le b_n^k.$$

因此 $\Pr(E, \max_{i,t} Q^r(i,t) > 2b_n) \le n^4 2^{-k}$。取 $R = \max(r, R_0)$，其中 R_0 是引理 10 中的四步随机游走滞后。左乘次随机的 Q^{R-r} 可保持条目界。对于此 R，定理 13 的一致纤维界以超多项式小概率失效。因此，除原始事件概率为 $o(n^{-A})$ 的情形外，引理14 中的类 $\mathcal{G}$（其中 L、M 是 $\log n$ 的固定幂次）包含所有所需的表。它的概率也为 $1 - o(1)$：对 $1 \le s \le K_A$ 取并集，可将任意此类闭集出现的概率界定为 $\sum_s \binom{n}{s}(s/n)^{2s} = O_A(1/n)$。

对于典型表，先前的四步随机游走估计（引理 10）给出 $\|JQ^{R_0}\|_F^2 \le n^{-u}$，其概率为 $1 - o(1)$，其中平方范数指数为 $u = 1/10$。这仅是一个粗略的中位数估计。一致纤维将 Q^{R-R_0} 的最大列和界定为 M^2，其最大行和则不超过一（将长度 R 纳入纤维事件）。因此，在交集上，由算子范数的次乘性可得 $F(B) \le M^2 n^{-u}$。令 d_H 表示行汉明距离，并令 $c_n = 8RML/n$。对 F 的限制，使用 McShane 的利普希茨延拓 [16]，其定义为 F

$$\widetilde{F}(B) = \min_{C \in \mathcal{G}} [F(C) + c_n d_H(B, C)].$$

它在 $\mathcal{G}$ 上等于 F，且在任意一次行编辑下的变化至多为 c_n。在原始分布下，所有 n 行相互独立~。McDiarmid不等式 [17] 推出，F 的均值与中位数相差不超过 $O(c_n\sqrt{n})$，并且 ¯

$$\Pr\{\widetilde{F} > n^{-2\kappa}\} \le \exp[-n^{1-4\kappa}/(\log n)^{O(1)}]$$

对于任意固定的 $0 < 2\kappa < \min(u,\ 1/2)$。事实上，根据前述典型界，其中位数至多为 $M^2 n^{-u}$。此论证不需要在 $\mathcal{G}$ 条件下的独立性，且扩展仅用于证明。特别是，平方范数 F 就是有界差分观测量。如果 S 是无小型终端层，那么 $\Pr(S,\ \|JQ^R\|_{2\to 2} > n^{-\kappa}) \le \Pr(S \setminus \mathcal{G}) + \Pr(\widetilde{F} > n^{-2\kappa})$ 因为$\|JQ^R\|_{2\to 2}^2 \le F$ 在 $\mathcal{G}$ 上。应用定理 13。 ■

B.4. 公共后缀决定输入历史的几何结构

记 λ 为 i.i.d. r 字母表中状态上的均匀行，$r \ge 2$ 固定。对于词 $u,\ v$，令$s(u,\ v)$ 表示它们的最长公共后缀的长度，并令 $u \preceq v$ 表示 u 是 v的后缀（包括相等的情况）。对于任意固定的 $C > 0$ 和 $c < 1/2$，以关于该表的超多项式置信度，同时对所有 $|u|,\ |v| \le C \log n$，

$$n\langle \lambda T_u, \lambda T_v \rangle = 1 + s(u,v) + O(n^{-c}). \tag{32}$$

将每一行都条件化为目标顶点互异，不会改变这一结论，因为该事件的概率有一个正的下界。在中心块定理的无小终端分层上，对于任意固定的置信度指数 A，除去 $o(n^{-A})$，

$$n\langle \pi T_u, \pi T_v \rangle = 1 + s(u,v) + \frac{\mathbf{1}\{u \preceq v \text{ or } v \preceq u\}}{r-1} + O(n^{-c}). \tag{33}$$

这里的小终点截断可以依赖于 A；第一个公式不需要遍历性假设。**证明** 对于固定的 u， v， a， b，定义 D 为在独立均匀起点 i， j 下，状态 $f_u(i)$ $f_v(j)$ 互异，且它们分别在字母 a， b 作用下的像相等的概率。取$T = 1 + \max(|u|, |v|)$，在表先验下

$$\mathbb{E}[nD] = 1 + O(T^4/n).$$

这里也给出了长度不等的历史的详细分析。首先以起始状态互异为条件，并将两者都标记出来。按照每条路径固定的时间顺序，依次展开两条完整历史，但不包括最后一个字母。除概率 $O(T^4/n^2)$ 外，对已命名顶点的新回答至多只有一个。若没有这样的回答，则两条路径的终点都互异且是新的。若只有一次返回且终点互异，则至少有一个终点仍未被查询为源。如果这次返回是在某个已遍历连通分量中的自相交，那么另一条轨迹就是一条新的简单路径。如果它连接了两个初始连通分量，那么底层无向多重图就是一片森林。此时每条路径都是简单路径。只有当两个互异的终端顶点各自都曾严格出现在另一条路径内部时，它们才可能先前已作为源出现；但这将形成一个有向环。这些情况也涵盖了第二个根节点的轨迹尚未开始前与其接触的情形。在此类历史中，最后的单元格至多有一个已被展开；两个互异端点处的单元格彼此不同。相等的条件概率恰为 $1/n$。仅无返回的情况就给出下界 $(1 - O(T^2/n))/n$。起始状态相等的概率为 $1/n$；以起始状态相等为条件，若没有新鲜返回，所有已揭示的路径都位于一棵向外树中。只有当两个互异端点各自位于另一根节点路径的内部时，它们才可能都已作为源出现，这是不可能的。因此，它们额外的条件成功概率至多为 $1/n + O(T^2/n)$，其中包括一次返回的例外情况。这就证明了在原始独立单元格先验分布下的期望估计。

在 $M = (\log\, n)^{O(1)}$ 时间内的均匀纤维事件上 T，根据定理 13的增长矩证明，其补事件的概率为超多项式小。对于两个在 h 行上存在差异的表，对 D 差异有贡献的事件，必须在至少一条轨迹中于最终探针之前访问某个编辑行，并且在其中一个表中具有相同的最终端点。统计正向变化时，整个计数过程都使用那个表；统计负向变化时也同样处理。对于每个词，至多有 ThM 个会使用某个编辑行的初始状态。固定这样一个起始状态后，其最终端点在对侧起点中的纤维大小至多为 M。因此 $|nD - nD'| \le 4ThM^2/n$。这在良好纤维集合上全局成立。在该集合上，将统计量通过 McShane 的行汉明距离 Lipschitz 延拓进行扩展 [16] ，并将其截断至 $[0, n]$。其期望与未扩展统计量的期望相差至多 n 乘以例外事件的概率。由 McDiarmid不等式 [17] ，并对多项式数量的词、字母和长度取并集，可得 $nD = 1 + O(n^{-c'})$ 对任意固定的$c < c' < 1/2$同时以超多项式置信度成立。

如果两个末尾字母相同，那么这两个扩展历史的内积等于其前驱的内积，再加上恰好 D。如果末尾字母不同，则内积为 D 加上前端点等于某个 z 且满足 $f_a(z) = f_b(z)$的概率。对于每对不同的字母，这类状态的数量服从 $\mathrm{Bin}(n, 1/n)$ ，并且除超多项式小概率事件外，不超过 $\log^2 n$ 。对于使用指定历史的两个独立起点，每个状态的质量至多为 M^2/n^2 。最后，任何与空历史的内积都恰好等于 $1/n$ 。递归至多进行 $C \log n$ 步，即可证明（32）。

使用长度为 $G = O_{A,\, c}(\log n)$ 的普通随机字母进行燃烧阶段，其长度足以使全变差 n^{-4}从 λ 和 $r^{-G} \le n^{-2}$。在 u 之前插入两个相互独立的此类预热词， v 并在扩大的固定时域取平均（ 32 ）。对于后缀不可比较历史，公共后缀保持不变。在可比较情形中，额外长度的均值为 $\sum_{i=1}^{G} r^{-i} = (1 - r^{-G})/(r - 1)$：条件是

较长的扩展词中，相关的较短预热阶段是独立均匀的。两个经过燃烧阶段的端点行可替换为 π，代价为 O（n^{-4}）的内积。这证明了平稳公式。
■

例如，在平稳层上的每个对数长度二叉历史 t、 $n\,\mathrm{tr}\,S_t = t + 2 + O(n^{-c})$ 处。更明确地说，固定一个满足 $\sum_a d_a = 0$ 的系数单位向量 $d \in \mathbb{R}^r$。定义$H_w(d) = \sqrt{n}(r-1)/r \sum_a d_a \pi T_{aw}$。对于该系数空间的一组固定正交归一基，所有由长度至多为 $C \log n$ 的词 w 和基向量索引的向量 $H_w(d)$，其 Gram 矩阵元素为$\mathbf{1}_{\{w=w'\}}\langle d, d'\rangle + O_r(n^{-c})$。对于 $w = w'$ 对角线元素和字母不相等的元素，缩放前相差 $1 + 1/(r-1)$。对于不可比较的节点，所有子节点之间的交叉元素都相等；对于不同的可比较节点，交叉元素只取决于较短节点所添加的字母，而另一个系数和为零。因此，任意至多包含 $n^{1/2-\eta}$ 个此类向量的子族，其 Gram矩阵 $I + o$ 在算子范数下均为 (1)，并且对每个固定的 $\eta > 0$ 同时成立。这是关于普通平稳历史分布的陈述。终端标签学习器还需要隐状态乘积构造。

B.4.1. 每个接受方向都存在常数相对间隙

在附录 B.4 的大终端层上，固定一个对数长度时间范围 $t \le C \log n$。对每个后缀 w 令 $\mu_w = \pi T_w$。在二元模型中存在一个绝对 $c > 0$，使得除原始概率为 $o(n^{-A})$的表事件外（对每个固定的 A分别选择截断值），在此时间范围内，对每一对 u， v 以下结论同时成立：

$$\sum_{x:\pi_x>0} \frac{\mu_u(x)\mu_v(x)}{\pi_x} \ge c(1 + s(u,v)).$$

因此，对常返支集上的每个实值测试函数 h，即使该测试函数是根据表选取的，以下结论也均匀成立：

$$\mathbb{E}_{|w|=t} \operatorname{Var}_{\mu_w} h \ge c \operatorname{Var}_\pi h, \tag{34}$$

$$\frac{1}{2}\sum_{i\ne j} \nu(i,j)(h_i - h_j)^2 \ge c' \operatorname{Var}_\pi h. \tag{35}$$

在第二个不等式中， ν 是被杀死的公共字母对核的拟平稳行，而 c'也是绝对常数。这些结论为稀有接受标签以及同一张表中的稠密线性组合提供相对界。它们并未为任意单侧符号字典的最小奇异值提供下界。

三个指定历史的碰撞界。对于等长词 a_1， a_2， a_3，其长度为$b = O(\log\, n)$，令 $p_j = \lambda T_{a_j}$，其中 λ 在状态上均匀分布。用 s_{ij} 表示公共后缀长度，并用 s_-， s_+ 表示其最小值和最大值。在超多项式置信度下，对这些词及所有此类长度均匀成立：

$$n^2 \sum_x p_1(x)p_2(x)p_3(x) \le C_0(1 + s_+)(1 + s_-) + o(1).$$

从三个配备互异人工源的独立均匀状态开始，因此每次起始抽取本身就是一次新鲜均匀探针。通过同步轮次展开这三条轨迹，

以这三个探针开始。使用 $O(b+1)$ 个探针和状态时，三个或更多新鲜返回的代价为$O((b+1)^6/n^3)$。成功的重合会连接三个初始连通分量，因此至少要使用两次返回。恰好使用两次时，合并后的多重图是一棵无向树：两次返回都连接互异的连通分量，而其他每条展开的边都会发现一个新状态。每条轨迹都是简单路径，并在同一个顶点结束。因此，交集必须同步发生并永久保持，因为在这棵树中，通往公共端点的剩余有向简单路径是唯一的。连接后共享的边使用同一个字母；对于重复的物理头部也是如此：端点相同的两个互异带标签槽位会形成一个重边环，在此处被排除。

选择要合并的第一对，以及它们在合并处和最终合并处的两个剩余步数。其中包括一次同步三重合并。这些数字分别至多为 s_+ 和 s_-。在一次成功的双返回历史上，每一轮中的每个当前源都是首次被使用：若在先前深度出现，将与其到终点顶点的唯一剩余距离相矛盾。第零轮的同时诞生使用人工新探测。为了对两个指定合并的概率给出上界，施加历史可测条件：所需源槽位是新的，且这些轮次中该对连通分量仍然互异。在合并处，按固定顺序读取互异的新探测；两个指定等式各自具有条件代价 $1/n$ ，包括在一轮中来自三个槽位的情形。联合上界为 $1/n^2$。并合界在缩放后以误差 $O((b+1)^6/n)$在期望意义下证明该结论。

关于同时的最大纤维界 $M=(\log n)^{O(1)}$，左侧统计量的行编辑常数为 $6(b+1)M^3/n$。事实上，只有当选定路径读取了某一编辑行时，三重重合才可能发生变化。对每个编辑行，每个词至多有 bM 个受影响的起始状态；对于每个这样的起始状态，至多有 M^2 种其他起始状态的选择能给出所需的最终端点。对相应端点表上的正向和负向变化分别计数。将归一化统计量延拓并截断到 $[0,n^2]$，并在取期望之前选取一个超多项式级别优良的纤维事件。有界差分以及对词的多项式级并集界现在给出一致的 $o(1)$ 偏差。这就证明了所述界。

从三元组到加权重叠。给 u,v 添加前缀：在前面接上两个独立均匀的燃烧阶段词，长度为 $G=O_{A,C}(\log n)$，从而得到长度相等的 a_1,a_2，长度为 $b=t+G$。独立均匀地取一个长度为 b 的 a_3。这三行的平均分布在全变差距离下分别与 μ_u,μ_v,π相差不超过 n^{-5}。给定前两个词后，重叠长度 s_{13},s_{23} 具有几何上尾，且二阶矩有界。利用 $s_+\le s_{12}+s_{13}+s_{23}$ 和 $s_-\le s_{13}$ 可得 $\mathbb{E}[(1+s_+)(1+s_-)\mid a_1,a_2]\le C(1+s_{12})$。剩余均值至多为 $C'(1+s(u,v))$；其中也计入燃烧阶段增加几何有界长度时 u,v相等的情形。归一化后替换三行平均分布的代价为 $O(n^{-3})$。因此

$$n^2\sum_x \mu_u(x)\mu_v(x)\pi_x \le C''(1+s(u,v)).$$

结合双历史公式，利用柯西–施瓦茨不等式，通过 $(\sum_x \mu_u\mu_v)^2 \le (\sum_x \mu_u\mu_v/\pi)\,(\sum_x \mu_u\mu_v\pi)$可得所述下界。所有行均由正常返质量支撑。

均匀选择一个后缀，并按分布 μ_w抽取其端点。条件期望算子 $B:\ L^2(\pi)\to L^2(W)$ 将 h 映射到 $\mu_w h$。随机自伴算子 BB^* 相对于均匀后缀的密度为 $\sum_x \mu_u(x)\mu_v(x)/\pi_x$，因此在每一对点上都不小于 c 倍的平稳分布。减去这个常数核，再除以 $1-c$，即可得到另一个平稳马尔可夫压缩算子。对于中心化测试函数，$\|B\|^2\le 1-c$，由此证明（34）。

对于（35），在同一个后缀上运行两个独立的 π 起始状态，直至对数 PF 燃烧阶段。存活行 u_t 与（$u_t\mathbf{1}$）ν 之间的全变差距离，对任意固定指定的E至多为 n^{-E}，只需增大燃烧阶段的系数。此外，$\frac{1}{2}\sum u_t\,(i,\ j)\ (h_i^- - h_j)^{\,2} = \mathbb{E}_w \mathsf{Var}_{\mu_w} h$。将非常数 h 归一化，使其振荡幅度为 $\mathbf{1}$。由直径和混合估计可得 $\mathsf{Var}_\pi h \geq \pi_{\min,\ +}/2 \geq n^{-C_A}/2$。选取 $E > C_A + 2$，即可使加性误差对 h均匀有界。由于 $u_t\mathbf{1} \leq 1$，这就证明了第二个结论，例如取 $c' = c/2$即可。极限碰撞系数中的常数不依赖于固定的对数窗口乘数。

B.5. 识别递归机器并预测每个长输入

瞬态状态可能永远不会作为指定长度训练词的端点出现。即便如此，对于典型目标，我们仍能恢复一个离散自动机，正确预测每个足够长的词，包括概率呈指数级小的词。

引理 16（恢复递归穆尔机）在具有固定置信度指数的干净中心化区块事件上，通过预测等价最小化得到的递归自动机，包括其两个命名的字母转移及其接受符号，可由终端标签算法在状态置换意义下确定。该算法使用指定长度为 $m \geq C\log n$ 的独立均匀词，其中 C 固定且充分，并在 n,$\log(2/\gamma)$ 中使用多项式数量的样本，以获得条件成功 $1-\gamma$。该机器中的具体起始状态不属于此次识别的内容。

证明 记 $\mathcal{C}$ 为真实递归机器的预测类，V为这些类上的提升字典，并 ω 为它们的正平稳权重。根据附录 D.2和B.2，V 具有单位行，且 $\sigma_{\min}(V) \geq n^{-C_0}$，而 $\min\omega \geq n^{-C_1}$。互异的类在长度至多为 $L = O(\log n)$的词上具有不同的符号签名。固定所有长度至多为 $L+1$的词的完整列表，包括空单词。令 s_j 为经过第 j 个此类词后的符号，并将其视为定义在类上的向量，定义

$$D = V^\top \operatorname{diag}(\omega) V, \qquad D_j = V^\top \operatorname{diag}(\omega s_j) V.$$

这里的乘积 ωs_j 是逐元素的。每个元素都是至多 $2k_A+1$ 个简短符号测试函数乘积的平稳期望的已知标量倍数。因此，只需选择足够长的对数滞后，递归终端专用乘积估计器便可将每个所需元素估计到 n 的任意固定逆幂精度。字典维度和列表长度 d 都是 n的固定多项式。通过采用足够精细的元素精度，可以同时得到算子范数下误差为 $o(\eta/d)$ 的对称有理数估计，其中 η 是介于 $n^{-(C_1+2C_0+3)}/2$和 $n^{-(C_1+2C_0+3)}$之间的正二进有理数。已知的实归一化因子可以舍入到这一精度，或者可以对其列应用统一的正缩放。

对于由 j个符号 $a_1,\ldots,a_j$组成的候选字符串，考虑

$$K(a) = \sum_{\ell \leq j} (D - a_\ell D_\ell)/2 - D/2.$$

若符号串不存在，则该矩阵为半正定，因为对应的整数不匹配计数在每个类上都至少为一。若它出现在类 x 中，选取 $u = V^\dagger e_x$，使得 $Vu = e_x$ 且 $\|u\| \leq n^{C_0}$。其关于 $K(a)$ 的瑞利商至多为 $-\omega_x/(2\|u\|^2)$ $\leq -n^{-(C_1+2C_0)}/2$。因此，$\widehat{\ }$经验矩阵 $K(a)+\eta I$ 当且仅当模式不存在时为正定。正定性可以

通过有理消元或顺序主子式进行测试。将已实现的模式用两种符号扩展，并保留未通过此测试的子模式。在任意深度，通过测试的模式至多有 n 个。对于任意经验输入，将搜索深度限制在 n 。因此，最终列表中每个常返类恰好对应一个扩展后的符号签名。其空词位上的比特就是该类的接受符号。对于每个字母 a，测试函数 aw、 $|w| \le L$ 上的取值确定了该状态转移到的唯一符号签名，因为其后继是常返状态。所有步骤的比特复杂度均为多项式。平稳量通过从指定起点进行未观测随机填充来获得，与乘积估计量中的做法相同，即使该起点位于核心之外也是如此。 ■

定理 17（对每个长词均正确的正确自动机）在原始独立同分布二元模型中，接受位彼此独立且公平；预先固定起始状态，并令 $g = \lceil 3\log_2 n \rceil$。根据长度为 n 且带有唯一最终标签的独立均匀训练词，多项式时间算法返回一个至多具有 n 个状态的 *DFA*。除去概率为 O（$1/n$）的目标事件，在条件样本概率至少为 $1-\gamma$的情况下，其预测值与目标在长度大于 g的每个词上均一致。样本量关于 n、$\log(2/\gamma)$ 是多项式的，并且与期望的预测误差无关。

证明 固定行编辑、后缀字典、谱估计和入流质量估计的干净事件，其指数为 $A = 3$。根据闭集界，有界终端分量的原始概率为 $O(1/n)$。存在一个大小大于 $3n/4$ 、外部概率为 $O(1/n)$的闭强连通核心 $\mathcal{K}$ 。指定某个这样的集合的事件彼此不相交，并且仅取决于其内部行。在给定该集合的条件下，外部槽位是独立均匀状态。若固定起始状态位于外部，则它能够到达一个完全位于外部且不进入核心的环的概率为 $O(1/n)$。事实上，枚举一条长度为 $a \ge 0$ 的简单初始路径和一条长度为 $b \ge 1$ 的简单环，两者仅在进入环的位置相遇。至多有 2^{a+b} 种字母选择，以及至多 $(n/4)^{a+b-1}$ 种新状态选择，最后一个顶点已预先确定。不同源槽位的代价为$n^{-(a+b)}$。对 a，b 求和即可得到该界。若不存在这样的环，则长度为g 的纯外部路径必须是简单路径，其概率至多为 $2^g(1/4)^g$。因此，在剩余事件上，长度为 g 的每个词都会将起始状态带入核心。

令 $x \sim y$ 表示两个有序物理出边行的均值相等。固定一个足够大的 $t = jR = O(\log n)$。我们声称，除去一个 $O(1/n)$ 表事件外，

$$\inf_{x \not\sim y} \delta_{xy} Q^t \mathbf{1} \ge 1/32.$$

对于每个指定的配对，沿着由四个字母构成的完整二叉树揭示两条起始路径。在其 16 个叶节点处，附加长度为 $t-4$ 的独立均匀延续，并在每对叶节点内共享。条件性的全部相遇概率等于 16 个叶节点相遇概率的乘积。分叉前缀中只有常数个单元格提案，因此在那里发生三次新鲜返回的概率为 $O(n^{-3})$。否则，记 c_j 为第 1 轮至第 4 轮的返回次数，其总和至多为二。令r_j 表示限制在新发现的终态上的独立前缀对比量的秩。(20) 的证明给出 $r_1 \ge 2 - c_1$, $r_{j+1} \ge 2r_j - c_{j+1}$：在下一轮之前，将每个对比量分叉为位于不相交字母块中的两个坐标。若 $c_1 = 2$ 且 $r_1 = 0$，则两个新发现的顶点必须在同一坐标上分别收到相反的首字母提案。这要求 $x \sim y$，而此处已排除这种情况。因此，对于 $c_1 = 2$ ，仍有 $r_1 \ge 1$。这些不等式给出$r_4 \ge 4$。对于 $c_1 = 0$ 或 1，当所有剩余返回都发生在第

二轮时，下界最小。对于 $c_1 = 2$，之后的每次分叉都是新鲜的。任意延续要消除这一秩，至少需要再发生四次新鲜返回。根据对 $O(t)$ 个单元格和 $O(t)$ 个命名目标进行顺序揭示，它们的联合概率为 $O(t^8/n^4)$。因此，不等行的指示函数与该乘积的乘积的期望，上界为 $O(n^{-3} + t^8/n^4)$。若全部 16 个叶节点的相遇概率都超过 $1/2$，则它们的乘积超过2^{-16}。由马尔可夫不等式以及对 $n(n-1)$ 种可能起始状态取并集，即可证明该结论：否则，某一条指定的四字母分支的存活概率至少为 $1/2$。

令 ν 为杀死对链的单位质量左 Perron 行， ρ为其特征值。附录D.1 中的秩一燃烧阶段给出：对于足够大的固定值 j，对起始对均有 $\|\delta_{xy}Q^t - (\delta_{xy}Q^t\mathbf{1})\nu\|_1 \le n^{-4}$ 。此外， $1-\rho = (1+o(1))/n$。在 R 处的对链热核上界给出 $\max_{i,\,k}\nu(i,\ k) \le \rho^{-R}(\log n)^{O(1)}/n^2$。因此， $\nu\{(i,\ k) : \xi_i \neq \xi_k\} = 1/2 + o(1)$ 的条件概率以超多项式概率成立：非对角部分上的二次型均值为零，弗罗贝尼乌斯范数至多为 $(\log n)^{O(1)}/n$，因此可以应用 Hanson–Wright 不等式。在 $\ell = n - g$ 处，与 $x,\ y$ 具有不同最终标签的条件概率，因而对 $x \not\sim y$一致地有一个绝对常数下界；这是将 $Q^{\ell-t}$ 应用于燃烧后的行并使用 $\rho^{\ell-t} \to e^{-1}$得到的。对于 $x \sim y$，任意正长度下该概率均为零。

首先根据引理 16恢复最小化的递归穆尔机。将其中每个由两个后继状态组成的向量称为 行类：当两个恢复状态的有序后继状态对一致时，它们共享一个行类。这些抽象类在核心上与物理 $\sim$一致。事实上，相同的物理行具有相同的后继特征；而不同的物理行在长度为 ℓ 的输入上有正比例的差异，因此即使最小化后也不可能具有相同的后继状态。取一个独立样本组。对于每个长度为 p 的前缀 g，将每个核心候选在样本剩余后缀上的预测与观测到的符号进行比较。真实端点位于核心中。所有错误的行类都具有常数级的不匹配质量，而正确类中的成员在剩余长度为正时始终一致。因此， $O(2^g(g + \log n + \log(2/\gamma)))$ 个额外的独立词就能通过二项集中和并集界，为所有前缀识别出正确的行类。

还需要将带前缀的核心缩减为至多 n 个状态。对于物理行类，令 $U([x]) = ([f_0(x)], [f_1(x)])$。在给定事件上，这个映射在整个物理表上都是良定义且单射的：两个此类元组相等意味着它们在每个双字母词之后的轨迹都一致，而不同的行在时间 t时具有正的存活率。对于所有递归行类，使用这些有序元组建立查找表。保留完整的恢复穆尔机，并为来自前缀树的条目选择每个递归行类的一个代表状态。每个叶节点 p在深度 g 处都带有其已识别的递归行类令牌。按深度递减的顺序对每对子标记进行内部化：如果该对子在查找表中，则复用对应的行类令牌；否则创建一个令牌，并将其附加到该对子上。每个新令牌对应一个新的 DFA 状态，其符号任意，并连接到其两个子标记的代表状态。根据构造过程进行归纳可知，令牌相等当且仅当所表示的物理行类相等。事实上，子标记表示实际的子类，而 U 是单射，即使这些令牌是在不同深度进行内部化的。因此，新增令牌至多有 $n - |\mathcal{K}|$ 个，而完整的最小化核心至多有 $|\mathcal{K}|$ 个状态。使用根节点代表作为初始状态，并在其恢复的穆尔自动机中保留所有真实的核心边。限制所有计数，并在保护条件失败时使用单状态默认值。

对于任意长度为g 的输入前缀，模拟的行类令牌都是真实令牌：每次状态转移，包括完整核心内的状态转移，都具有适当的子标记。于是，其模拟状态是循环的，而真实状态位于同一个物理行类中。每个非空延续都会给出相同的输出，因为等行状态具有相同的物理后继。这就同时证明了任意未来词长下的断言。 ∎

命题18（目标先验障碍对精确置信度的影响）对于任何基于 N 个独立均匀分布的长度为m的词及其终止位的学习规则，在原始先验下，至少对一个长度为m 的词产生误差的联合概率至少为

$$\frac{1}{n}(1-1/n)^{2n-1}(1-2^{-m})^{N}.$$

即使学习器知道完整的转移表，这一点仍然成立。特别是，对于 $m = n$ 和多项式 N，下界为 $(e^{-2}+o(1))/n$。

证明 以概率 $2n^{-1}(1-1/n)^{2n-1}$，起始状态的两个槽位中恰有一个是自环，另一个离开该状态，并且没有其他状态的槽位指向起始状态。重复的自环字母是唯一一个长度为 m 、以起始状态结束的输入。如果没有样本使用它，那么其中公平的接受位就与所有观测及学习器的随机位相互独立。对该词进行的测试以条件概率 $1/2$失败。 ■

B.5.1. 接受集上的一致精确置信界

在一组概率为 $1-O(1/n)$ 的独立同分布二元转移表上，定理 n 17 在训练长度 下的正确精确结论对每个接受集合都成立，包括在检查转移表后才选定的接受集合。所有样本量和运行时间的多项式都可以独立于接受集合预先固定；样本置信度以特定目标为条件。割估计中的一个多对数倒数级常数足以保证这一一致界。

对于表事件，取定理 17、热估计与公共后缀估计以及（35）中的大分量中心块、对数纤维、前缀和四秩分叉。巨分量、规定起始状态的外部前缀事件以及四秩分叉合计耗费 $O(1/n)$；所有所需的小终端失败概率都可以指定为更高的逆幂。记 ψ 为非负的右佩龙列，并将其归一化，使秩一投影为 $\psi\nu$，其中 ν 的质量为 1。对于所有固定的精度指数 E 以及足够长的固定对数燃烧阶段 l，

$$\|Q^{l}-\rho^{l}\psi\nu\|_{2\to 2}\le n^{-E}.$$

块时间之间的幂次由纤维界和交换性保证；在下文中，l 始终可以是一个完整的块。归一化以及不等行的 $1/32$ 存活性给出$\psi_{xy}\ge c$，当 $x\not\sim y$时一致成立。对于 $z_i=2b_i-1$，记 $\alpha=\nu(1-z_iz_j)$。如果核心包含两种颜色，则（35）蕴含 $\alpha\ge c\pi_{\min,\,+}\ge n^{-C_0}$，因为非常数符号的方差至少为 $2\pi_{\min,\,+}$。否则，核心是一个输出类；在长度为 n 时，每个输入都会在其最终字母之前到达该类。根据单游走混合，非常数情形下，两种终端符号的概率在训练长度 n时都至少为 $\pi_{\min,\,+}/2$。多项式规模样本组可以区分这两种情形；当只观测到一种符号时，则采用常数级后备方案。

中心提升的标签无关结构。 选择 t_0 一个足够大的固定对数块倍数。令 $D_{t_0}(x,\ y)$ 表示相遇概率，令 $G_j(x,\ y)$ 表示经过一个共同 j-词后的标签乘积的均值。扩展 $\psi_{xx}=0$。对有界函数 1 和 $1-z_iz_j$ կիրառ用成对 Perron 估计，可在掩码上均匀得到

$$D_{t_0}(x,y)=1-\rho^{t_0}\psi_{xy}+O(n^{-E}),\qquad G_j(x,y)=1-\alpha\rho^{j}\psi_{xy}+O(n^{-E}),$$

对于 j, t_0 足够大的所选对数倍数。在对有序对上的有界向量应用算子界时，额外保留一个逆幂。取 $j = t_0$，其几何精度指数为 $E > C_0 + 3$；它在确定特征维数和之后的矩提取精度之前固定。记 $P_* = I - \mathbf{1}\pi$。令 $l_w(x) = [z_{f_w(x)} - \pi T_w z]/2$，并使用基特征$l_{w_0}(f_0 x) l_{w_1}(f_1 x)$，其中采用相互独立的长度为 j 的词，并按词数进行通常的单位归一化。除以 $\alpha/4$ 后的单分支 Gram 矩阵逐元素等于 $P_* D_j P_*^\top + O(n^{-E}/\alpha)$，因为两个中心常数矩阵均为零。其对角线为 $1 + O(\sqrt{j/n})$，每个非对角元素与$D_j(x,\ y)$ 相差 $O(\sqrt{j/n})$。事实上，根据历史 Gram，有 $(D_j \pi^\top)_x \le \sqrt{\mathbb{E}\|\pi T_w\|_2^2} = O(\sqrt{j/n})$。取$E$ 足够大，以便将近似误差吸收到该界中。

根据附录 D.2中的仅依赖表的星形界，在此固定的对数 j 处应用该界，并令折叠集合为空，则每个状态只有固定数量的 B 个重邻居，其阈值为 n^{-a}在 $\prod_i D_j(f_i x, f_i y)$ 处，且 $a > 0$是固定的小量。其证明只使用固定星形度下的 $O(j)$ 个探针，并允许增加一个新的 $\log n$倍数。因此，两次中心化后的基础 Gram 矩阵除以 $(\alpha/4)^2$,后，对角线为 $1 + o(1)$，而在这些重边之外，绝对值不超过 $O(n^{-a})$。对于结构表事件上的每个非恒定掩码，这些估计同时成立。

对于不同的物理行，在长度为 j 后，其两个符号不匹配的概率为$\frac{1}{2}\alpha\rho^j\psi_{xy} + O(n^{-E})$ > 0。这为每一对这样的物理行提供了一个短区分词。同一物理行中的两个成员要么当前符号不同，要么预测等价。因此，长度不超过 j 的短测试能够区分所有预测类，且对掩码均匀成立，无需使用并集界。应用附录 D.2 中的对偶向量构造：对中心化双分支特征取固定次张量幂，并为重邻居加入 B 个短测试因子。在分析中将基础特征除以 $(\alpha/4)$ 后，对角线为 $1 + o(1)$，而轻边重叠的大小为 $O(n^{-a})$。取固定次张量幂后， $O(\sqrt{j/n})$对角线误差仍为 $o(1)$，因此对偶评估矩阵的对角线有正的下界，非对角线元素的绝对值行和与列和均为 $o(1)$。对偶行的范数为多项式有界，因为其基础范数有界，每个选择器读取一个长度为 $O(\log n)$的测试，且每行对应的选择器数至多为 B。返回未缩放特征只需付出 α^{-1}的固定次幂代价。因此，该学习器的字典维数为多项式且有界，并且在不同的循环预测类上，其最小奇异值至少为 n^{-C_2}，对这些掩码均匀成立。

可观测乘积与精确路由。只有在设定好这些固定的精度指数之后，才能选择乘积提取滞后 l。对于平稳历史矩 S_l 和 $B_l = z^\top S_l z$，恒等式 $1 - B_l = u_l(1 - z_i z_j)$ 与烧灼行近似在每个非恒定掩码上给出 $1 - B_l \ge c\alpha$。事实上，根据平稳历史格拉姆矩阵，行质量为 $1 - O(l/n)$。结合 $x_l = (u_{l+1}\mathbf{1})/(u_l \mathbf{1})$，根据成对佩龙速率以及任意精确的对数烧入期，我们有 $1 - x_l \asymp n^{-1}$。随后，对校准比率 $(1 - B_{l+1})/(1 - B_l)$ 的估计以及对角乘法规则仅损失多项式精度，如附录 B.2 所示。中心化特征还使用标量 $\pi T_w z$，它本身是以相同精度估计的终端均值。展开其固定度数的乘积或应用平衡递归，只需付出固定幂次的代价。任何采样步骤都不要求从同一训练轨迹中观测两个响应。在移位状态下使用中心化特征时，$f_a(x)$ 保持与 l_w 中相同的标量 $\pi T_w z$。每个这样的标量以及每个原始终端均值都以 1 为界。在固定度数的有界乘积中，将这些标量改变 ζ 会通过对各因子进行望远镜求和，使其值改变 $O_A(\zeta)$。考虑到矩阵坐标的多项式数量，便得到引理 16 中的算子精度。该引理的因子分解及其见证 $Vu = e_x$ 使用奇异值下界；此处允许有界但不相等的特征行范数。半正定

模式检验因此恢复了最小化循环机。根据同一个 Perron公式，不同行之间长度为 $n-g$ 的差异至少为 $c\alpha$ ；在第二样本面板中，一个因子 $O(n^{C_0})$ 已足够。外部状态和行内化证明与掩码无关，从而证明了该结论。

B.6. 有限例外集

在独立同分布二元表中，给定的有序对未必具有均匀的正存活系数。两个状态上的两个字母可能产生相同的后继，或者只有一个被反复选取的字母能够使一对状态保持互异。我们在同一张表上统计后一类障碍的互不相交副本。

引理 19 固定 A、$C>0$ 和一个整数 $3\le t\le C\log n$。除去一个概率为 $o(n^{-A})$的原始表事件，至多有 B_A 个具有不同有序出边行 的无序对，在经过 t个公共字母后仍保持互异的概率小于 $1/8$ 。这里的 B_A 可以独立于对数乘数 C 固定（对于每个固定的 C，所需的初始 n 可能有所不同）。

证明 固定一个由 k 个有序对组成的匹配（所有 $2k$ 个起始状态各不相同）。从每个有序对开始运行四个双字母前缀，并为所得的每个有序对分别附加一个长度为 $t-2$的均匀分布延续，在该有序对内部共享其字母。令 E 表示原始的 k 次行比较均不相等，且所有 $4k$ 个叶对均相遇。我们声称，实现 E 至少需要 $3k$ 次新单元格回访。在揭示表之前，标记所有 $2k$个起始状态。一次回访是指首次揭示一个单元格，而该单元格的头在此次揭示之前已被命名；重新读取单元格无需付出代价。

首先从所有起始状态暴露两个字母，并带有 c_1 个返回。存在 $4k-c_1$ 个新状态。存在 $2k$ 个对比，每一对与每一个字母各对应一个。将它们的端点差向量限制到新状态上。它们的秩为 r_1。至多有 k 个原始对比具有相同端点，因为每一对中的两个有序行互不相同。每一个这样的对比至多只能解释一个新状态。任一原始不等对比上的每一个新状态都位于相应受限对比的支集中。从实际列中选取一组受限对比的基；其 r_1 列各自的支集大小至多为二，并且它们共同覆盖每一个受限对比的支集。由此可得

$$4k-c_1\le k+2r_1.$$

将每个对比再用两个固定字母分叉一次，分成两个互不相交的坐标块。新前沿上的出向提议向量现在所张成的维数至少为 $2r_1$。在随后每一轮中，每个叶对比使用一个坐标，并对其两个端点取相反的符号。在一个新状态上，两个出向提议向量之和等于当前系数向量，即使互异的叶选择不同的字母。设 r 为该轮之前新状态上的秩，r' 为该轮结束时新发现状态上的秩，而 b 为该轮中新鲜返回的数目。一个旧单元不能指向新发现的状态。每一个发现某状态的提议等于其最终系数向量减去此后对该状态的返回贡献。因此，所有新鲜源提议都位于最终新状态向量与 b 个返回向量的张成空间中。这给出 $r\le r'+b$；在分叉轮中，同一论证从提议秩至少为 $2r_1$开始。在最终等式处所有系数均为零，因此第一轮之后至少出现$2r_1$ 个返回。这也证明了该断言在不同的叶延续随后于同一状态选择不同字母时仍然成立。

共有 $O_k(t)$ 个提议的单元格和状态。对原始独立同分布单元格进行序贯条件化，并对返回时间取并集，可得 $\Pr(E)=O_k(t^{6k}n^{-3k})$。在每个根节点

在该匹配中，其存活概率均小于 $1/8$ 的表上，其四个等可能的双字母分支的条件存活概率均小于 $1/2$。如果所有行都不相同，则由独立延拓可知，在该表上 E 的概率至少为2^{-4k}（该表已固定）。候选匹配至多有 n^{2k} 个。任意大小为 k 的坏匹配的概率为 $O_k(t^{6k}n^{-k})$。

此外，这个坏边图以压倒性固定幂次置信度具有有界最大度。如果根节点以低于 $1/8$ 的概率存活，则在每个指定首字母之后的相遇概率都超过 $3/4$。这两个相遇概率的乘积大于 $9/16$。对于一个在 h 个互异邻点上、度为 h 的星图，沿每个首字母进行一次分叉，并对所有 $2h$ 个对比使用独立延拓。邻点的 $2h$ 单例提议坐标具有满初始秩；最终消失至少需要 $2h$ 次新鲜返回。因此，根据马尔可夫性并对中心和邻点取并集，任何此类坏星图的概率都可由 $O_h(t^{4h}n^{1-h})$ 上界。取 $h > A+2$ 和 $k > A+1$。极大匹配的端点覆盖所有坏边。度小于 h的端点少于 $2k$ 个，因此覆盖的此类边少于 $2kh$ 条。

■

推论20 在置信度指数为 A 的干净中心化区块事件上，除去 $o(n^{-A})$，记ψ_{xy} 为右存活系数，并用单位质量左 *Perron* 行 ν 对其归一化：$Q^{\ell} = \rho^{\ell}\psi\nu + O_{2\to 2}(n^{C_A}e^{-\theta_A \ell})$。对于所有不同的 x、y，均匀地有：要么 $\psi_{xy} \geq c_A > 0$，要么 $\psi_{xy} = 0$。在后一种情况下，它们的公共词活对永远不会达到对数级存活至少为 $1/8$ 的有序对，并且对于所有 $\ell \geq 0$，其存活至多为 $C_A e^{-c_A \ell}$。在这种后一种情况下，任何具有不等行的对，在合并或最终形成等行对之前，始终处于至多包含 B_A 个无序对的集合中。关系$\psi_{xy} = 0$ （其中 $\psi_{xx} = 0$）是一个右同余。

证明 选择一个足够长的固定对数块倍数 t。由 $(Q^t\mathbf{1})_{xy} \geq 1/8$ 定义的好对，对于大 n，根据燃烧近似和 $\rho^t \leq 1$，具有 $\psi_{xy} \geq 1/9$。如果这样的有序对可以到达，则到好集合的最短活路径至多经过 B_A 个坏的非等行对；一对等行会在下一步进入墓地状态。最短路径不会重复同一个无序对。特征向量恒等式与这条路径的质量给出 $\psi_{xy} \geq 2^{-B_A}/9$。如果无法到达好集合，则可达活子矩阵的 t 步行和低于 $1/8$。由于 $\rho^t \to 1$，根据其最大范数和特征向量恒等式，将非负特征向量限制在这些状态上所得的向量为零。每一个这样的有序对都有一条通往墓地状态的路径：否则，它在所有时刻的存活率都将为一。最短路径的长度至多为 $B_A + 2$ （允许末尾出现一对等行），长度为此的马尔可夫块的吸收概率至少为 $2^{-(B_A+2)}$，从而得到所显示的指数衰减。最后，根据端点相等的逐路径传递性，在共同词上的存活满足 $q_{xz}(\ell) \leq q_{xy}(\ell) + q_{yz}(\ell)$。除以 ρ^{ℓ}，并令 ℓ 在给定的有限表上趋于无穷，从而得到 ψ 的三角不等式。右特征向量恒等式使其零关系在每个字母作用下保持闭合。

■

恰好区分所有循环预测类。对于任意在干净常返类上 z 非常数的接受符号，记 $g_z(i,j) = 1 - z_i z_j$ 和 $p = \pi\{z = 1\}$。那么 (35) 给出 $\alpha = \nu g_z \geq 4c'p(1-p) \geq 2c'\pi_{\min,\,+} \geq n^{-C_A}$，因为两种颜色都出现了。从状态对 i 出发，在时刻 t' 的不一致概率为 $(Q^{t'}g_z)_i/2$。在 Perron近似中，$Q^{t'}g_z$ 的~~绝对分~~量误差至多为 $2\sqrt{n}(n-1)~~n^{-E}~~$，因为 $\|g_z\|_2 \leq 2\sqrt{n}(n-1)$。先选择 $E > C_A + 2$ 及其对数烧入期 t'，再选择 z。主项 $\rho^{t'}\psi_i\alpha$ 对每个好对都大于此误差，因为 $\rho^{t'} = 1 - o(1)$ 且 $\psi_i \geq 1/9$。因此，同一个表事件也能为检查表后才选定的接受集合提供见证

。对于任何能够到达好集合的状态对，都先加上通往此测试的短路径前缀。如果无法到达，那么每条活路径都会停留在有界坏集合中，且可能在某一步到达一个等行对，并于下一步合并。因此，通往任意当前符号不相等状态的最短路径至多为 $B_A + 1$。由此，任意两个预测上不同的循环状态，都能在长度至多为 $t' + B_A + 1$ 的词上区分。核心常数情形只有一个预测类。

附录 B.5.1 中基于中心化双分支 Gram 矩阵和有界度选择器的证明 A现在使用这些短见证，且置信度指数仍可任意指定为 A。PSD 特征测试在单个训练长度 $m \geq C_A \log n$ 下重构出 n，m的$\log(2/\gamma)$最小化循环穆尔机，所需时间和样本数均为多项式级。对于这一推断，目标事件是在干净大终端层上，所有掩码和所有指定起点的均匀分布。当核心为常数时，另取一个样本即可观察到其占主导地位的符号；如果起点是外部状态，则可通过一个频率阈值去除其非典型符号，该阈值低于非恒定核心质量下限，并高于外部燃烧阶段的误差。逆多项式几何阈值在采样精度确定之前固定。

附录 C. 被动回归与隐藏闭合分量

C.1. 每个状态上的未知随机输出

固定一个表，并为每个状态指定一个数 $h_i \in [-1, 1]$。给定一个以 i 结尾的词，观测到的响应 $Y \in \{-1, 1\}$ 的均值为 h_i。在给定各自词的条件下，不同样本中的输出相互独立。两个相邻状态的均值可以任意接近，也可以相同。任务是估计每个词上的条件分布。

定理 21 预先固定一个指定起点，并令 $g = \lceil 3\log_2 n \rceil$。除去 $O(1/n)$ 原始表 概率，对于均值的每一种赋值，使用长度为 n 的均匀词及其单个带噪终止输出的学习器，都会返回一个至多有 n 个状态的确定性转移自动机，并为每个状态估计一个均值。以至少 $1 - \gamma$ 的样本与内部成功概率，其条件均值在每个长度大于 g 的词上都与目标相差不超过 η。其运行时间和样本量关于 n、$1/\eta$、$\log(2/\gamma)$ 均为多项式。表事件取决于指定起点，并且对所有均值赋值均相同。该断言是关于 n 的渐近结论。

证明 使用附录 B.5.1 中仅转移的适当事件。特别地，从起点出发的每个 g-词都以常返类中的状态结束，每个常返质量都大于n^{-C_0}，中心化块具有多项式收缩，并且不同的物理有序行具有至少为绝对常数的右 Perron 系数。在整个证明中，行令牌是物理后继的有序对。共享同一行的两个状态，其实际均值可能不同。它们的响应在每个非空延续之后都相同。

寻找排放的尺度。 对于循环平稳行向量，写作 $D = \max_{\mathcal{K}} h_i - \min_{\mathcal{K}} h_i, \pi$；对于在所有长度为 l 的历史上对 $\mu_w = \pi T_w$ 的二阶矩，写作 S_l。令

$$B_l = h^\top S_l h, \quad M_2 = \pi h^2, \quad \beta = \tfrac{1}{2} \sum_{i \neq j} \nu(i, j)(h_i - h_j)^2.$$

这里，标量值状态检验的幂是逐点定义的。式（35）给出 $\beta \geq c$方差$_\pi h \geq cn^{-C_0} D^2/2$。方差的下界使用最大和最小原子，它们各自的

质量至少为 $\pi_{\min,\ +}$。当 u_l 非对角历史行以及 $w_0 = (\pi \otimes \pi)_{\mathrm{off}}\psi$ 有上界且远离零时，精确方差恒等式为

$$M_2 - B_l = \tfrac{1}{2}\sum_{i\neq j} u_l(i,j)(h_i - h_j)^2 = w_0\beta\rho^l + O(D^2 n^{C_1} e^{-\theta l}).$$

它只使用循环对。在任意足够长的固定对数烧入期内，对于余项中任意预先指定的固定相对精度，增量 $\Delta_l = B_{l+1} - B_l$ 是 $w_0\beta\rho^l(1-\rho) + O(D^2 n^{-E})$。特别地，对于固定的 $j_0 = C_2 \log n$，$\Delta_{j_0} \geq D^2 n^{-C_3}$。将 h 以 πh 为中心并不会改变这一增量，因为每个历史行的均值都是 π。

令 c_n 为从指定起点开始的实际长度为 n 的均值。对于足够大的 n，每个长度为 $\leq j_0 + 1$ 的历史至少会在其后留下 g 个不受约束的输入字母。其填充状态属于 $\mathcal{K}$，并且对于每个所需的固定 E'，其分布与 π 之间的全变差为 $n^{-E'}$。因此，每个真实的条件终端均值与 $\mu_w h$ 的差至多为 $n^{-E'}D$，且 $|c_n - \pi h| \leq n^{-E'}D$。长度为 $j_0 + 1$ 时的平稳中心化平方均值平均至少为 Δ_{j_0}。对于固定的 C_4，我们得出：在这些历史中，最大的绝对条件偏差 R 相对于 c_n 满足 $n^{-C_4}D \leq R \leq D$。对于 $D = 0$ 且 $R = 0$时，这一点同样成立。

使用一个独立样本组来估计所有这些条件均值以及 c_n 到 $e = \eta n^{-C_4-5}$。使用估计中心 $\widehat{c}$ 计算 $\widehat{R}$。如果 $\widehat{R} \leq \eta n^{-C_4}/4$，则当 n足够大时，真实范围小于 $\eta/2$；在所有所示词上同时返回估计常数 $\widehat{c}$ 即可。否则取 $S = 4n^{C_4}(\widehat{R} + 3e)$，其中幂向上取整为整数指数。上述不等式给出 $S \geq D + e$、$S \geq c\eta$和 $S/D \leq Cn^{C_4}$。因此，归一化核心均值 $z_i = (h_i - \widehat{c})/S$ 位于 $[-1, 1]$ 中，且范围至少为 cn^{-C_4}。对于这些归一化均值，后续所有几何和重构精度都可以固定为 n的逆幂，与 η无关。归一化样本响应的幅度可能为 $O(1 + 1/\eta)$；这会使样本量增加一个多项式因子，而它们在相关后缀分箱上的条件均值仍保持有界。经验中心和尺度在新面板上固定。

物理行上的一个字典。对于 z，相同的增量公式在适当选择的乘积滞后 $l = O(\log n)$ 下给出 $\Delta_{l+1}/\Delta_l = \rho + O(n^{-E''})$。由于归一化范围和平稳正下界都是多项式的逆量，分母现在至少为 $n^{-C}(1-\rho)$ 阶。应在该下界之下施加经验保护条件，同时还要满足 $1 - \widehat{\rho} \geq 1/(4n)$。独立估计的终端均值之积给出 B_l；观测到的 Y^2 并不是对z_i^2的测量。恢复得到的 ρ 和 $(S_{l+1} - \rho S_l)/(1-\rho)$ 提供了与附录 D.2.2 中相同的对角乘积递推。对于下方给定的固定几何参数，只有在确定所需乘积的最终多项式精度后，才能选择 l 和原始分箱精度。随后，所有窗口都固定为 $\log n$ 的倍数，并且在具有足够未使用填充的情况下，均可容纳于训练长度 n 之内。

为了解所需的秩，在固定的大对数 j 处，将 $D_j(x,\ y)$记为相遇概率，将$G_j(x,\ y)$记为两个归一化均值在同一个 j-词上的期望乘积。在核心上均匀地，

$$G_j(x,y) = \tfrac{1}{2}[(P^j z^2)_x + (P^j z^2)_y] - \rho^j \psi_{xy}\beta_z + O(n^{-E}), \qquad D_j(x,y) = 1 - \rho^j\psi_{xy} + O(n^{-E}).$$

这里，β_z 是针对 z 给出的狄利克雷量，并且有一个固定逆幂下界。对 $I - \mathbf{1}\pi$ 进行双重中心化会消去两个单坐标项。对于有界特征 $(z_{f_w(x)} - \pi T_w z)/2$，因此，将中心化的单分支 Gram 矩阵除以 $\beta_z/4$ 后，所得结果与中心化的D_j 相符，且误差在所需的几何误差范围内。双分支乘积 Gram 矩阵的对角线为 $1 + o(1)$，经过

相应的归一化后，其绝对值条目 $O(n^{-a})$ 在有界度图之外。这就是附录B.5.1中的纯表格星图证明。选择恰好足以压过该β_z 下限和此阈值的几何精度。

对于任意两个互异的循环行令牌 x，y，相同的成对计算给出一个长度为 j 的单词 w，满足 $|z_{f_w(x)} - z_{f_w(y)}| \geq n^{-C_5}$。事实上，均方差为 $2\rho^j\psi_{xy}\beta_z + O(n^{-E})$。对于每个重邻居 y，选择一个针对 x 的此类测试，并将线性选择因子相乘

$$\frac{z_{f_w(v)} - z_{f_w(y)}}{z_{f_w(x)} - z_{f_w(y)}}.$$

它们在 x 处取值为 1，在各自的邻居处取值为 0；其乘积在整个核心上的绝对值均受 $n^{O(1)}$ 限制。双分支特征的一个充分大的固定张量幂，会使其所有轻量评估值即使乘以此因子后也小于 n^{-2}。每个选择器都是常数和长度为 j 的测试的线性泛函，且系数范数为多项式级别。相同的对偶向量计算在互异的循环行令牌上给出 $\sigma_{\min}(V) \geq n^{-C_6}$。这里，$V$ 列出有界且未经缩放的字典；各行的范数不必相同。只有长度为正的测试会进入 V，因此无论该物理行当前的发射值如何，它在每个物理行上都是常数。这里没有使用单个发射值之间的最小间隔。

用一个随机标量为行命名。 对于每个 $d = 2^j$测试 $z_w(x) = z_{f_w(x)}$，独立地从 $\{1, \ldots, N\}$、$N = 16n^2$中均匀抽取 u_w。令 $g_*(x) = \sum_w u_w z_w(x) / (Nd)$。在给定除某个区分系数之外的所有条件下，两行的标量差值每当该系数改变一个单位时，至少改变$n^{-C_5}/(Nd)$。因此，以至少 $3/4$ 的概率，所有行令牌在 g_*中彼此至少相隔 $\delta_* = n^{-C_5}/(4Nd)$。该标量的绝对值不超过一。

我们描述一种能够识别充分分离性并恢复读取状态转移所需数值的测试函数。对于我们正测试的任意有界线性组合 f，终端递推以任意指定的固定逆幂精度估计这些矩阵

$$D_V = V^\top \operatorname{diag}(\omega) V, \quad D_f = V^\top \operatorname{diag}(\omega f) V, \quad D_{f^2} = V^\top \operatorname{diag}(\omega f^2) V.$$

这里 ω 是物理行质量。对一个固定多项式规模的线性组合求平方，会产生多项固定度数的乘积。D_V的正谱由前述字典下限和平稳质量下限给出的 $\lambda_0 = n^{-C_7}$从下方界定。计算其对称估计中大于 $\lambda_0/2$ 的特征值数量，即可得到行令牌的确切数量 r；有理惯性或具有间隙的多项式精度特征值求解器足以完成此任务。

对于网格位置 a 和半径 s_0，使用正则化正定性检验进行扫描，作用于

$$D_{f^2} - 2aD_f + (a^2 - s_0^2) D_V.$$

如果所有实际的 f 值与其距离至少为 s_0，则该矩阵为半正定。如果某个值与 a的距离小于 $s_0/4$，则见证 $V^\dagger e_x$ 给出的瑞利商至多为 $-s_0^2\lambda_0/2$（如有必要，将λ_0 减小至共同乘积下限 $\min \omega\, \sigma_{\min}(V)^2$）。因此，施加一个较小的正对角平移，并使估计误差更小，就能确保每个被接受的网格点都在真实值的 s_0 范围内，且每个真实值都有一个被接受的点与其相距不超过网格间距 $s_0/8$的 $s_0/4$。扫描一个略大于已知有界范围的区间，以包含端点。

取 $s_0 \ll \delta_*$，并贪心地保留彼此至少相隔 $\delta_*/2$ 的 g_* 中的点。如果存在 r 个这样的点，则每个点代表不同的物理行，所有行都得到表示，且它们的真实中心

彼此相隔 $\delta_*/2 - 2s_0$。如果满足更强的采样间隔条件，则对于任何这样的贪心扫描，都将有 r 个点。抽取 $O(\log(2/\gamma))$ 个相互独立的系数组，直到该可验证的测试函数成功为止。所有底层的原始矩估计都可以在抽取之前于同一事件上同时完成。对于 $a \in \{0,1\}$，计算相同的扫描以

$$f = g_* + t_*(g_* \circ f_a) \quad \text{and} \quad f = g_* + t_*(z \circ f_a), \qquad t_* = \delta_*/16.$$

每个新中心都位于对应旧中心的 t_* 范围内，且不同中心之间仍保持分离。因此，所有接受点都可以与其唯一的旧中心匹配。减去位置后再除以 t_*，产生的误差至多为 $2s_0/t_*$。对于第一次扰动，取最近的恢复行令牌；当 $2s_0/t_* \ll \delta_*$时，这就是准确的后继令牌。第二次扰动给出该字母边上的临时均值。选择 s_0 足够小，使其归一化均值误差也成为远小于 $\sqrt{\beta_z}$ 的固定逆幂。所有这些尺度都仅关于 n呈逆多项式关系。特别是，在扫描中只返回近似的实数位置并不会导致指数级分支。

路由并估计最终输出。在独立样本组上，针对每个长度为g 的前缀，使用终端比特上的平方损失分别拟合入口行令牌。对于任意固定后缀，期望的噪声平方损失等于其条件方差加上相对于真实均值的平方误差。正确令牌具有均匀较小的归一化均值误差。两个不同令牌的均值平方差异至少为 $c\beta_z$，在长度为$(n-g)$ 的延续上由成对公式得出。因此，在将临时边均值反归一化并将其截断至 $c'S^2\beta_z$ 之后，总体损失间隔至少为$[-1,1]$（相对于对应令牌真实均值的均匀误差不会增加）。该间隔关于 n、$1/\eta$呈逆多项式关系，因为 $S \geq c\eta$。对所有前缀分箱应用 Hoeffding不等式即可选出准确的行令牌。按照附录B.5中的方法，将后继元组向后内部关联至根节点。这就给出了从时间 g 起每个输入上的真实行令牌轨迹。新的暂态令牌在下一个循环字母之后不会对读出产生影响。

最后，通过新面板模拟恢复出的路由，直至倒数第二个状态，并以其行令牌和最后一个字母为条件。根据混合性质，每个循环令牌在时间 $n-1$ 的概率至少为其平稳质量的一半；最后一个字母服从公平分布。通过对这些多项式质量分箱中的终端响应进行普通平均，将每条循环边的均值估计至 $\eta/16$。这一步使用实际词分布，并且无论要求的精度多高都能实现，因为它不存在平稳近似偏差。每条循环边的真实输出都属于该边的物理头部。

将行表示转换为状态上带有均值的普通机器。对于每个循环令牌，将所有进入它的循环边归为一组。在这些估计均值中，采用贪心方式保留一个两两距离大于 $\eta/2$ 的极大子集。由于同一状态的估计值相差至多为 $\eta/8$，每个保留的估计值都必须来自该行中不同的状态标识符。为每个保留值创建一个状态。将每条循环边指向其真实后继令牌对应的、距离最近的保留状态，并为令牌的每个状态配置该令牌的两条出边。此类边上的均值误差至多为 $\eta/2+\eta/16$。根据闭包性和不可约性，所有循环令牌都至少有一条进入的循环边。当此前缀程序首次进入某个循环令牌时，可以选择其任意一个保留代表元，因为其下一个字母已经具有正确的边输出。每个新的暂态令牌至多添加一个状态。所得状态数至多为 n，且每个长度为$> g$ 的词都会在其最后一个字母上经过一条正确的循环边。每个分支都会在其多项式时间界内停止；若达到上限，则返回默认预测器。上述失败界使用独立批次。 ■

C.2. 学习真实的终端均值：干净表

本节将干净表构造扩展到任意函数的平方损失 h： $[n] \to [-1, 1]$。观测值相互独立，其条件均值为 $h(f_W(s))$，且条件二阶矩至多为一。第C.5 节中的变换后的响应满足这些矩条件。所有保证均针对训练和测试中相同长度的 m。

固定目标异常事件指数 A。使用该指数下的干净事件来处理单游走混合和平稳质量界、中心化有序对块及其佩龙投影、相对间隙（35）、B.5.1 节的重邻居图 以及推论 20的二分法。右 Perron 系数为零 ψ_{xy} 意味着 x， y 会在均匀分布的 j-词上合并，除非发生概率为 $C_A e^{-c_A j}$的事件。正系数至少为 c'_A。包含对角线的零关系是一个等价关系：使用经过第三个状态产生分歧的并集界，以及另一种情况下的正存活系数。根据非负特征向量方程，它是一个右同余。一个商类可能包含两个不同的当前均值。我们恢复其状态转移，并在观测长度处拟合一个单独的读出。

选择一个可测尺度。在常返类上，令 $D = \max h - \min h$。C.1 节中的推导给出 $\beta_h = \nu((h_x - h_y)^2)/2 \geq n^{-C_1} D^2$。对于足够大的固定 $j = C_2$，当 $\log n$ 时，平稳增量 $B_{j+1} - B_j$ 至少为 $D^2 n^{-C_3}$。这里， B_j 使用了每个历史均值的两个独立估计的乘积。因此，长度不超过$j+1$ 的最大绝对中心化均值 R 满足 $n^{-C_4} D \leq R \leq D$。采用长度为 $C \log(n/\epsilon)$ 的常规填充，并以绝对误差 $e = \epsilon n^{-C_4 - 10}$ 估计所有均值和中心，其中包括混合偏差。若 $\widehat{R} \leq \epsilon n^{-C_4}/16$，则当规模较大时，常返范围至多为 $\epsilon/8$ n $S =$，常数中 $8n^{C_4}$ $\widehat{R} + 8e$ $S \geq$ $D + 2e$ $c\epsilon$ $S/D \leq Cn^{C_4}$ $z =$ $h - \widehat{c}$ $/S$ 心便足够。否则，( ) 满足 max(,) 和 。对于 ( )，常返测试函数的绝对值至多为一，并且 $\beta_z \geq n^{-B_0}$，其中 B_0 为固定值。连续可观测增量之比可用于估计成对速率：利用 $1 - \rho \geq n^{-C}$ 和正的平稳投影系数，其分母可由一个固定的逆幂下界所约束。现在，双历史乘法规则可以估计任意所需的固定阶矩。确定所需精度后，再选择其滞后。归一化响应的条件二阶矩为 $O(\epsilon^{-2})$。对独立样本组采用中位数均值法，可达到所需的加性精度和置信度。额外的不可见填充 $O(\log(n/\epsilon))$ 可控制该尺度上的瞬态贡献。

以商代表元为条件。令 $t_w(x) = z(f_w(x))$，并令 $l_w = (t_w - \pi t_w)/2$ 适用于$|w| = j$，其中 $j = C_j \log n$ 中的常数将在最后确定。独立抽取 $d_1 = n^{K_1} + O(1)$ 个测试函数，并构造向量 $t(x)$，其分量为这些测试函数的取值及该常数，再除以 $\sqrt{d_1 + 1}$。独立地抽取 $d_0 = n^{K_0} + O(1)$ 个词对，构造向量 $v(x)$，其分量为 $l_{w_0}(f_0 x) l_{w_1}(f_1 x)/\sqrt{d_0}$ 的乘积。将 $d_1 + 1$、 d_0 舍入为相同多项式阶的四次幂。对于互异的类，根据成对投影，当 j 足够大时， $\mathbb{E}_w(t_w(x) - t_w(y))^2 \geq c_A \beta_z$。有界性表明，对于每一对代表元，以至少 .99 的概率可以得到采样间隔 n^{-B_2}；对于固定的 K_1、 B_2，该概率仅取决于 B_0、 A。

对于中心化格拉姆矩阵，使用

$$\mathbb{E}_w t_w(x) t_w(y) = \tfrac{1}{2}(P^j z^2(x) + P^j z^2(y)) - \beta_z \rho^j \psi_{xy} + O(n^C e^{-cj}).$$

双重中心化会消去单坐标项。因而，第B.5.1 节中剩余的中心化双分支估计，在将乘积格拉姆矩阵除以 $(\beta_z/4)^2$之后仍然适用。其对角线为 $1 + o(1)$，而在一个固定度数的图 $O(n^{-a})$ 之外，其各项为 B。中心化校正使用对数长度下的 $\max_w \|\pi T_w\|_\infty \leq n^{-1+o(1)}$。逐元素采样精度

可通过选取一个固定的 K_0 并对 β_z取下整得到。两个样本数的指数均与 C_j无关。可以在初始的 $j_* = C_* \log n$ 处固定例外图，此时投影误差可忽略不计。对于之后任意固定倍数 j，$\|\psi\|_\infty \le 2$ 和 $1-\rho \le n^{-1+o(1)}$ 使单分支生存核的变化至多为 $2|j-j_*|(1-\rho) + O(n^{-E}) = n^{-1+o(1)}$。每个先前较轻的双分支项在 $a < 1$下仍为 $O(n^{-a})$。因此，在确定容差后选取 C_j，即可使用同一个有界度图。

使用 $V(x) = v(x)^{\otimes p} \otimes t(x)^{\otimes B}$，并为每个类取一个代表元，从而得到行矩阵 $\bar V$。对于 x 的重邻居 y，使用其采样见证，将 x 的对偶测试函数乘以 $(t_w - t_w(y))/(t_w(x) - t_w(y))$。这些因子具有多项式系数范数。一个足够大的固定 p 可使归一化对角线后所有轻交叉评估至多为 n^{-3}。所得评估矩阵在算子范数意义下与单位算子的距离不超过 $o(1)$。它给出 $\sigma_{\min}(\bar V) \ge n^{-C_V}$，其中 C_V 在 C_j 之前固定。另一方面，在同一个类内，归一化 V 和 t 向量的期望平方距离至多为 $C_{p,\,B} e^{-c_A j}$，这是由耦合、右同余以及有界测试函数的望远镜乘积得到的。对于之后所需的任意 $\delta_0 = n^{-K}$，马尔可夫不等式和并集界表明，通过令 C_j 足够大，同一类的所有循环对的距离以至少 .99 的概率至多为 δ_0。这种选择顺序避免了标量分离量依赖于 2^j。

命名类及其后继。 令 ω 为类质量，并令 $\bar D = \bar V^\top \operatorname{diag}(\omega) \bar V$。根据平稳下界和前述奇异值界，其正特征值超过已知下界 $\lambda_0 = n^{-C_\lambda}$。真实矩 $D_V = \mathbb{E}_\pi V^\top V$ 与 $\bar D$的距离不超过 $2\delta_0$。构造有界标量 $g = \sum_i u_i t_{w_i}/(N d_1)$，其中 u_i 在 $[N]$、$N = 64 n^2$ 上独立均匀分布。它在 t 中的系数范数至多为二。随着其系数变化，一个采样见证以至少$n^{-B_2}/(N d_1)$的步长区分一对代表元。对其他系数进行条件化并取并集界，可得同时隙距 $\Delta = n^{-C_\Delta}$，其概率至少为。98。在一个类中，g 在 $2\delta_0$范围内为常数。

对于 $f = g$ 和 $f = g + u(g \circ f_a)$，$u = \Delta/16$，计算矩 $D_f = \mathbb{E}_\pi f V^\top V$ 和 D_{f^2}。它们的估计值与理想类矩阵的距离在 $O(\delta_0 + \eta)$ 以内，第二种情况下的编码值为 $g_C + u g_{f_a C}$。计算高于 $\lambda_0/2$ 的特征值数量，以确定类的数量。在有理数网格上考虑

$$D_{f^2} - 2b D_f + (b^2 - s_0^2) D_V .$$

如果每个类值都至少距离 $b\,s_0$，其理想矩阵就是半正定的。距离 $s_0/4$ 以内的类，通过其对偶坐标可产生至多为 $-15\lambda_0 s_0^2/16$ 的负瑞利商。加上平移量 $\lambda_0 s_0^2/10$，并保留具有负方向的网格点。网格间距为 $s_0/8$ 时，每个真实值附近都有一个距离不超过 $s_0/4$ 的保留点，而每个保留点到某个真实值的距离都不超过 s_0。理想编码值之间仍相隔 $\Delta - 2u$。将距离为 $2s_0$ 的保留点聚类，即可为每个类恢复一个中心。将其编码中心与原始中心匹配，减去并除以 u，然后读取最近的后继中心。选择 $s_0 \ll \Delta^2$，然后选择 δ_0，$\eta \ll \lambda_0 s_0^2$。δ_0 容差决定 C_j；η 决定乘法滞后。这些选择可实现精确路由，且每次试验的成功概率至少为二分之一。在相同的固定维度下，使用全新的独立试验 $O(\log(2/\gamma))$ 次，将所有扫描限制在其多项式上限内，并验证所得预测器。

读出。 燃烧阶段 $g_0 = C_A \log(2nm/\epsilon)$ 个前导字母。以任意小的固定乘数计，结束于常返类之外的箱的质量为 $O(\epsilon)$。在每个箱中，尝试每台拟合机器的每个初始类。跟踪该类，直到包含 $L = C'_A \log(2nm/\epsilon)$ 个字母的最后窗口。在窗口边界处及最后 L 个字母上，对其类拟合一个有界回归表。对于正确的路由试验和初始类，固定物理代表元的边界类的比较器 $h(f_w(x_C))$

平方风险至多为 $4C_A e^{-c_A L}$，并且在该处的物理混合上均匀成立。每个单元格表至多有 $n2^L$ 个条目。可能取值的逆多项式网格的对数覆盖数为 $O(n2^L \log(2/\epsilon))$。拟合可以使用逐单元格截断经验均值；将原始响应截断于 $O(1/\epsilon)$，这样根据二阶矩假设，均匀条件偏差为 $O(\epsilon)$。在每个初始分箱的留出样本上进行均值的中位数比较，以选出候选者：对于每个有界预测器，下面的中心化得分具有有界条件二阶矩。期望损失 $\widehat{h}^2 - 2Y\widehat{h}$ 与均方风险之差仅为一个公共项。调整燃烧阶段、拟合和验证误差中的常数，使风险达到 ϵ。如果 m 太短，无法容纳任何所需窗口及其填充，则 2^m 关于 n，m，$1/\epsilon$ 为多项式级，并且直接进行全词回归即可。历史和前缀分箱的数量、位精度以及所有试验规模均满足所述多项式界。

C.3. 均值为常数的闭合分量上的回归

随机终端观测是一个有界随机变量 $Y \in [-1, 1]$，其 $\mathbb{E}[Y \mid W] = h(f_W(s))$，其中 $h : [n] \to [-1, 1]$。给定独立的训练样本，我们在平方损失下估计条件均值，并允许方差取决于整个词。只有一个物理状态的陷阱会自动具有常数均值；同样的论证也适用于均值相同的任意有界闭集。

谱输入与记号。 在有限吸引域约化中，对闭合分量使用干净修复。将闭集折叠为一个特殊状态 x。按照该约化中的做法修复其出弧。通过在 x 处终止修复过程，可以恢复原来的外部链。对于$b = 1$，2，令 P_b 为互异 b元组上的干净转移核，并在坐标间重合时终止。记 $P_1 = P$和 $P_2 = Q$。在 $D = C \log n$ 时，干净热核与佩龙估计给出 $P_b^D(i, j) \le M/n^b$，以及一个秩一近似 $P_b^D = \rho_b^D h_b \nu_b + E_b$，对于任意预先指定的固定 B，只需增大 C即可使其满足 $\|E_b\|_2 \le n^{-B}$；此外还有 $\nu_b \mathbf{1} = \nu_b h_b = 1$、$\|h_b\|_\infty = O(1)$、$\|\nu_b\|_\infty \le M/n^b$。此处及下文中，$M$可以是固定的多重对数函数，也可以是 $n^{o(1)}$。短词纤维上的一致上界适用于长度为 $O(D)$ 的词的每个子串。从均匀不同状态出发，在下文所用事件上，第 $D+1$ 步的干净成对游走首次相遇概率至少为 $n^{-1-o(1)}$。所有这些结论都针对干净修复，包括其选择；标记与该修复相互独立并非必要。

从干净游走中移除一个状态。将 U_b 写作同样的核：当任一坐标访问 x 时被置零，并对包含 x 的元组以及 $A_b = U_b^D$ 的元组补零。于是

$$\|U_b^{3D} - A_b P_b^D A_b\|_2 \le n^{-1+o(1)}.$$

事实上，在两个外部块中，每个固定端点的概率至多为 M/n^b。第一次燃烧阶段之后，在任意后续的中间块时刻，杀死轨迹在禁片上的质量至多为 $O(M/n)$（利用由 P_b^j 对 $j \ge D$ 的支配及其行次随机性）。命中该禁片后，最终端点的概率至多为 M/n^b。对 $D+1$ 次取并集，可将差的每个条目界定为 $n^{-b-1+o(1)}$。算子范数至多为行数乘以条目界。由于 ν_b 是 P_b 的左特征测度，并且在 ν_b 作用下经过 $2D$ 步触及标记而损失的质量至多为$n^{-1+o(1)}$，因此首要的秩一算子具有非零特征值 $\rho_b^D \nu_b U_b^{2D} h_b = 1 - n^{-1+o(1)}$。在这一估计中，干净对杀死率为 $n^{-1+o(1)}$。根据其外部因子的热核界，秩一算子的范数为$n^{o(1)}$。其误差（包括中间的干净近似）为$n^{-1+o(1)}$。

我们直接使用预解式。对于特征值为 λ 且接近 1 的秩一 J，$(z-J)^{-1}=z^{-1}I+J/[z(z-\lambda)]$。在围道 $|z|=n^{-1/4}$ 和 $|z-\lambda|=n^{-1/4}$上应用诺伊曼界，再结合投影算子秩的连续性，可得到一个简单的顶部佩龙根，以及一个补空间，其 k次幂的范数至多为 $n^{O(1)}n^{-k/4}$。将此应用于 U_b^{3D}；由行次随机性，中间幂的范数至多为 $n^{b/2}$。特别地，对于某个固定的 $a>0$ 和 C_0，

$$U_b^j=\zeta_b^jH_b\mu_b+F_{b,j},\qquad \|F_{b,j}\|_{\infty\to\infty}\le n^{C_0}e^{-aj}. \tag{S1}$$

对 $\mu_b\mathbf{1}=\mu_bH_b=1$进行归一化。根据同一个秩一公式，围绕 λ 的围道半径也可以取为 $n^{-1+o(1)}$，但使用稍大的 $o(1)$ 指数。因此 $1-\zeta_b\le n^{-1+o(1)}$。对于足够大的固定 $D'=C'\log n$，(S1) 和行次随机性意味着 $\|H_b\|_\infty\le 2$；除去均匀分布在 $n^{o(1)}$ 个例外单状态之外，$H_1(y)\ge 1/3$。对于最后一个结论，对初始状态求和后，在 D' 步内接触 x 的概率至多为 $(D'+1)M$，这是因为对每个固定初始片段都有纤维界，再对词取平均即可。

两个根的定量分离。令 $\sigma=\zeta_1,\tau=\zeta_2$。记 C 为一步内从安全不同对到安全共同目的地的核。从均匀不同状态出发，在 $D'+1$ 处发生的首次相遇在删除操作后仍然存在，并且以至少 $1/3$ 的概率在其端点处具有 $H_1\,n^{-1-o(1)}$。为说明这一点，固定一个完整词并将其揭示出来。如果某个坐标在时刻 j访问该标记，则其初始状态属于一个大小为 M的纤维。对于每个这样的状态，另一个坐标必须映射到其最终端点，这同样是一个大小至多为 M的纤维。因此，对 j取并集至多会移除 $O(D'M^2/n^2)$。在 $n^{o(1)}$ 个例外状态中的某个状态处进行最终连接，代价至多为 $n^{-2+o(1)}$。在原始的不同起点条件化下，这些界均一致成立。因此，$b=2$ 的 (S1) 加上足够长的燃烧阶段意味着$\mu_2CH_1\ge n^{-1-o(1)}$。令 $\Phi(i,j)=H_1(i)+H_1(j)$。逐坐标地，对两个单独特征函数求和所得的单位算子恒等式给出 $\sigma\Phi\ge U_2\Phi+2CH_1$，其中多出的部分来自只有一个坐标访问标记的步。左乘 μ_2 和 $\mu_2\Phi\le 4$ 得到 $\sigma-\tau\ge n^{-1-o(1)}$。

单侧可观测量能量。减去共同陷阱均值 θ，在标记处将 $z=h-\theta$ 延拓为零，并用 U_w 表示沿特定词的单次杀死转移。令 $B_l=\mathbb{E}_{|w|=l}(\mu_1U_wz)^2$。对于由共同词驱动的两个相互独立的 μ_1 状态，互异安全状态对上的质量和安全单例上的质量分别按 $u_{l+1}=u_lU_2$、$v_{l+1}=v_lU_1+u_lC$ 演化，其初值分别为 $\mu_1\otimes\mu_1$ 的非对角部分和对角部分。因此

$$B_l=A\sigma^l\mu_1(z^2)+L(z)\tau^l+O(n^{C_1}e^{-a'l}),\qquad A=v_0H_1+\sum_{j\ge0}\sigma^{-j-1}u_0U_2^jCH_1. \tag{S2}$$

一种推导方法利用（S1）展开三角递推。两个慢项的卷积为（$\sigma^l-\tau^l$）/（$\sigma-\tau$）；单个快项可以延拓为其无限加权和，并带有指数级小尾项。快项–快项卷积为O（$n^{2C_0}le^{-al}$）。这些估计对于有界的 z和 σ，$\tau\ge1-o(1)$均一致成立。A中的所有项均非负，并且根据柯西–施瓦茨不等式，$A\ge\sum_i\mu_1(i)^2H_1(i)\ge1/(2n)$。特别地

$$B_{l+1}\ge B_{l+1}-\tau B_l\ge n^{-3}\sigma^l\mu_1(z^2)-O(n^{C_1}e^{-a'l}). \tag{S3}$$

学习器从不以避开标记为条件。

仅使用观测到的词和响应。 在外部状态起点 y处，对于剩余任务长度r 以及长度为 $l+1$的可见后缀 w，实际的条件残差均值为 $e_y U_1^{r-l-1} U_w z$。未作条件化的目标均方残差等于 $e_y U_1^r z^2$。对于检测阈值 $d \in (0,1)$，令其为$l = \lceil C_2 \log(2nm/d) \rceil$。若 $r < 2l+2$，则枚举整个剩余词。否则，(S1) 和 (S3) 表明，只要均方残差超过 d，就有某个后缀分箱的绝对均值残差至少为 $(2nm/d)^{-C_3}$。确实，令 $s_y = \sigma^{r-l-1} H_1(y)$，我们有 $s_y \sigma^{l+1} \mu_1 z^2 \geq d/2$ 且 $s_y \geq d/8$。后缀上的可观测残差平方均值至少为 $n^{-3} s_y^2 \sigma^l \mu_1 z^2 - O(n^{C_4} e^{-a'' l})$。主项为 $\Omega(n^{-3} d^2)$；通过选择 C_2，可以使误差更小。每个后缀分箱的概率都为逆多项式量级。在陷阱内，每个残差都恰好为零。

使用原有限吸引域构造中的安全常返类 S、安全直径 h_0，以及精确长度的安全路由：对于任意长度至少为 G 的路径，都可从任意 $y \in S$ 到达任意 $z \in S$。所有安全字母都留在 S 中。如果根节点相对于常数 θ 的风险较小，就纳入常数候选项。否则，固定一个安全前缀 p，使其满足 $|p| \leq W$，并令 $L = W + G$。如果 $m' \leq L$，则使用全词字典。如果 $m' > L$，则对根节点任务的前 L 个字母取平均：存在一个这样的前缀 v，其条件均方残差至少为 $\epsilon/16$。它始终安全，因为一旦进入，剩余部分的残差就恒为零。从 p 的端点出发，经过恰好 $L - |p|$ 步安全地到达 v 的端点。因此，给定 p，剩余字母上的残差风险至少为 $d = 2^{-L}\epsilon/16$。不需要变换时钟。对于陷阱内部的前缀，真实延续是常数。因此，刚刚证明的后缀分箱测试可以提供所有所需前缀上的成员判定。使用独立样本估计每个所需的条件均值。前缀和后缀长度均为对数级，并集界带来的开销是多项式级的。在网格上猜测 θ，网格间距小于检测裕量的十分之一，并将真实分支纳入候选项。最终验证通过比较平方损失来进行；即使条件方差任意，损失差仍是均方风险差的无偏估计。修正后的干净核心分支使用单独的有界回归学习器。此证明将高置信度的有限吸引域几何结构，以及干净的热估计和首次命中估计作为输入。具有非恒定均值的有界闭分量还包含 (S2) 未涵盖的额外内部模态。

C.4. 在例外分量上学习任意终端均值

定理 22 固定 $A > 0$。在概率为 $1 - o(n^{-A})$ 的独立同分布二元转移表集合上，对于每个起始状态、每个 $m \geq 1$ 以及每个向量 $h \in [-1,1]^n$，以下结论成立。样本是独立对 (W, Y)，其中 W 在长度为 m 的词上服从均匀分布，且 $|Y| \leq 1$、$\mathbb{E}[Y \mid W] = h(f_W(s))$。给定这些样本，学习器会返回一个预测器，使其对同样长度的独立词的终端均值的均方误差至多为 ϵ。条件样本失败概率至多为 γ。其样本数量和算术成本关于 n、m、$1/\epsilon$ 以及 $\log(2/\gamma)$ 是多项式的，次数取决于 A。对于伯努利数据，位运算也具有同样的结论。对于实值数据，在通常的位模型中，若能够访问有界响应的连续二进制位，并将这些二进制位的成本计入其中，则结论同样成立。

前一节中的几何结构和零均值闭分量学习器，是唯一依赖于具体表的输入。下面的归约适用于异常状态上的任意确定性有限作用，也适用于依赖输入的响应方差。特别地，它适用于每个词仅有一次伯努利观测、且成功概率依赖于状态的情形。平方误差是相对于条件均值而言的；新观测到的平方损失则包含不可约响应方差。

对后缀的动作进行条件化。设 T 为表约化中的有界饱和闭集。若它为空，则使用干净的有界回归结果。否则，猜测其抽象

转移动作。将其 $s = |T|$ 个状态编号后，至多有 s^{2s} 种可能；对于每个固定的 A，这是一个常数。不会提供或猜测任何内部状态的均值。对于正确的动作，记$G(v)$： $T \to T$ 为由 v 诱导的变换，并记 G_a 为一个单个字母。考虑以下递增的信息集序列：

$$\mathcal{F}_j = \sigma\big(W_{1:j}, G(W_{j+1:m})\big), \qquad 0 \le j \le m.$$

确实如此， $G(av) = G(v) \circ G_a$，因此揭示下一个字母和剩余的变换，会保留先前的信息。最初只能看到 $G(W)$；最后则能看到整个词。

在还剩 $r = m - j \ge 1$ 个字母时进行分裂，记 $B = (a, G(v))$ 和 $C = G(av)$ 为 $av = W_{j+1:m}$。在新的均匀后缀下，定义 $p_b = \Pr(B = b)$、 $p_c = \Pr(C = c)$。对于每个概率为正的 b，都存在唯一的 $c = c(b)$，且 $0 < p_b \le p_c$。这些概率可由猜测的动作得知：将概率向量（其对应的变换至多有 s^s 个）与长度为 $0, \dots, m$ 时的双字母转移规则相乘。它们的二进制分子和分母有$O(m+1)$ 位。令

$$\psi_b(av) = \mathbf{1}_{B=b} - (p_b/p_c)\mathbf{1}_{C=c}, \qquad g_{j,b}(x) = \mathbb{E}_{a,v}\left[\psi_b(av)h(f_{av}(x))\right].$$

差值的绝对值至多为一，并且在每个粗粒度单元上均值为零。如果 $x \in T$，那么$h(f_{av}(x)) = h(C(x))$ 仅取决于 C。由此可知，$g_{j,b}$ 在整个饱和闭集上都为零，其中包括其瞬态状态。

从每个训练词中保留其前 j 个字母，并使用 $Y\psi_b(W_{j+1:m})$ 作为其响应。其条件均值恰好是 $g_{j,\ b}(f_{W_{1:j}}(s))$。给定可见前缀时，后缀是随机的，而关于 Y 的定义性条件均值假设就是所使用的全部内容。对于 $j > 0$，适用此前的、在 T 上具有常数均值的学习器，其已知内部均值为零。对于 $j = 0$，通过样本均值估计这个单一数值。训练过程中不会测试物理状态、判断后缀是否进入 T，也不会要求对某个词给出第二个响应。

重构预测。令 $M_j = \mathbb{E}[Y \mid \mathcal{F}_j]$。对于终止于物理状态 x 的实际前缀，在单元格 $B = b$ 上，条件增量恰好为

$$M_{j+1} - M_j = g_{j,b}(x)/p_b. \tag{F1}$$

这是通过相减 $h(f_{av}(x))$ 的两个条件单元均值得到的。起始项 M_0 是至多 s^s 个可观测变换单元格的函数。使用普通单元均值对其进行拟合。最后$M_m = h(f_W(s))$。因此，未识别的发射向量只出现在 T 上均值为零的辅助任务中。

为了界定增量总和的误差，令 $K = 2s^s, L = m+1, \eta = \epsilon/(64KL^2)$，且$e = \epsilon\eta/(64KL^2)$。跳过概率小于 η 的细化单元格，并在该处预测增量为零。在某一时刻对所有被跳过单元格求和，其贡献至多为 $4K\eta$，因为 $|M_{j+1} - M_j| \le 2$。在其他单元格上，将每个 $g_{j,b}$ 学习到均匀前缀上的平方误差为 e，再除以 p_b 并截断至 $[-2, 2]$。该前缀与 B相互独立。因此，在给定时刻，预测增量的总均方误差至多为

$$4K\eta + \sum_{b:p_b \ge \eta} e/p_b \le 4K\eta + Ke/\eta.$$

对于 M_0，在有限单元上进行平方误差 $\epsilon/(8L^2)$的有界拟合。可以舍弃质量为$\epsilon/(64KL^2)$的单元格，并对其余部分应用条件霍夫丁不等式。将这些 L项相加，柯西–施瓦茨不等式

将总平方误差界定为 L 乘以其平方误差之和，低于 $\epsilon/2$。将最终预测截断为 $[-1,1]$。通过为每次拟合分配$O(\gamma/(LK))$的失败概率，所有条件训练保证可以同时成立。独立样本批次就足够了。当每个单独批次都具有独立同分布的行时，在各次拟合之间共享样本也能保留这一并集界。

对每个允许的有限作用和规模运行该构造，包括干净表候选项。使用新样本依据有界经验平方损失进行选择，使风险的加性误差为 $\epsilon/8$。期望损失之差等于相对于真实条件均值的误差之差。所有良好表的量词都先于辅助均值和长度的选择。因此，对于固定的 A，资源多项式在 h，m，ϵ，γ中仍保持均匀。对于实值输入，首先将每个 Y替换为 $Y-U$ 的符号，并在 $[-1,1]$中使用独立均匀的 U。新的响应具有条件均值 Y，并且只取两个值。给定总样本数 N 的上界，使用每个响应 $O(\log(2N/\gamma))$比特进行比较；如果区间重叠，则中止。对 U的抽样进行并集界，并以其响应为条件，可使任何中止的概率至多为 $\gamma/4$。在所有其他执行中，该比较与精确舍入完全相同。将剩余的失败额度用于拟合和验证。此后所有观测权重均为有理数，且比特长度为多项式级别。

C.5. 对给定有限作用进行二次数量的后缀测试

前述归约有一个版本，其代价关于给定例外作用中的状态数是多项式的，即使其变换幺半群包含指数多个元素。我们将任意终端均值的估计归约为对在例外集上为零的均值进行估计。该归约产生具有有界条件二阶矩的独立观测，而后者正是学习器所要求的。

对于大小为 s 的给定作用，令 E_r 为在均匀分布 $L^2(\{0,1\}^r)$ 下由指示函数 $e_{xz}(v)=\mathbf{1}\{f_v(x)=z\}$（对于 x，$z\in T$）以及常数所张成的空间。其维数至多为 s^2+1。整个输入上的空间 $V_j=L^2(\{0,1\}^j)\otimes E_{m-j}$ 是嵌套的：$e_{xz}(av)$ 的每个单字母限制都是指示函数 $e_{f_a(x)\,z}(v)$。特别地，V_m 是整个函数空间。令 $r=m-j$，定义

$$D_r=\big(L^2(\{0,1\})\otimes E_{r-1}\big)\ominus E_r.$$

于是 $V_{j+1}\ominus V_j=L^2(\{0,1\}^j)\otimes D_r$。为 E_m 和每个D_r选择相互正交的有理基，并将其元素 d 归一化，使平方范数为 $1/4\le v_d=\mathbb{E}d^2\le 1$。

对每个 $d\in D_r$，定义 $g_d(y)=\mathbb{E}_{|v|=r}d(v)h(f_v(y))$。对于 $y\in T$，这恰好为零，因为$h(f_v(y))\in E_r$。对于任意 y，由柯西–施瓦茨不等式可得 $|g_d(y)|\le 1$。使用实际的训练前缀和响应 $Yd(W_{j+1:m})$，在长度 j处学习 g_d。如果 $\mathbb{E}[Y^2\mid W]\le 1$，那么给定整个前缀，变换后的响应的条件二阶矩至多为一，因为其独立后缀服从均匀分布。投影到 $V_{j+1}\ominus V_j$ 上的相应系数为 $g_d(f_{W_{1:j}}(s))/v_d$。E_m 上的起始系数是普通的标量响应相关性除以相同的归一化因子。最终的正交和给出终端均值。对于任何根据训练数据计算出的拟合系数函数，在最终截断之前，新词上的平方误差会沿时间和细节方向恰好相加，这是由所展示的空间正交性所致。因此，估计的 g_d中每个误差 e 至多贡献 $4e$。使用 $e=\epsilon/(32(m+1)(s^2+1))$，且拟合次数至多为 $2(m+1)(s^2+1)$。该界与变换幺半群的大小以及其最罕见后缀变换的概率无关。

所有计算仅使用给定的抽象状态转移。利用由公共字母驱动的 T^2 上的双坐标链，计算指示函数生成元的内积。对于首字母

分裂，在匹配的字母上将相应的内积乘以 $1/2$，否则乘以零。先处理粗生成元，再处理细生成元，并省略零主元。使用原始生成元上的精确有理数 Gram 系统，将每个新生成元投影去除先前选定的线性无关生成元所张成的空间。粗生成元之后得到的新正交向量张成 D_r。Gram 矩阵元素的公共分母整除 2^r。维度为 $O(s^2)$ 时的行列式界表明，投影系数的分子和分母各需要 $O(s^2(r + \log(2s)))$ 比特，另差一个固定常数因子。使用 2 的整数次幂即可将平方范数表示为 $[1/4, 1]$。因此，即使特征值非零且指数级微小，也只需多项式数量级的比特。方差界或样本数中不存在逆特征值因子。每个特征都可以通过沿观测到的后缀传播 s 个状态，并结合有理系数来计算。

有界条件方差下的响应估计。 零点-T 证明在条件均值估计、拟合有限终端读出以及新验证中读取随机响应。矩构造将分别估计的均值相乘；它不会要求对同一个观测响应取高次幂。通过独立组均值的中位数估计每个所需均值。在指定词分箱内，条件二阶矩仍至多为一。对于具有 k 个单元格的读出，估计这些单元格的概率及其未归一化响应均值；当预测值和真实均值均有界于一时，总质量为 $O(\eta/k)$ 的单元格，其总成本为 $O(\eta)$。对于其他单元格，只需达到逆多项式绝对精度，并对输出进行截断。对于有界候选函数 f 的验证，可使用 $f(W)^2 - 2Y f(W)$，或使用这些分数的成对差值。对于单个分数和 $|f| \le 1$，在新面板上，给定拟合预测器的条件 L^2 范数至多为 $1 + 2\|Y\|_2 \le 3$。其期望等于均方风险，至多相差一个对所有候选函数都相同的项。无需响应的四阶矩。相同的修改也适用于干净分量：其中所有响应都通过基本词分箱均值、终端读出和验证进入。有限维随机字典及其集中性使用潜在状态上的有界均值函数。若需要有界的统计查询实现，对于一个二阶矩至多为 V 的线性查询 X，将其截断至 $T = 2V/\eta$ 会带来至多 $\eta/2$ 的均值偏差，所得函数除以 T 后有界。仅将此方法用于单个基本估计；若只截断一次响应，可能会改变例外集合上的逐词均值。对于比特输入数据，即使长度为 r 的原子上最大的归一化后缀特征也至多为 $2^{r/2}$，上述有理数界也涵盖其算术成本。

C.6. 对隐藏闭合修改的鲁棒性

闭合异常区域的大小可能随 n增长。设 A 是一个独立同分布转移表，且满足常规干净事件；该事件涵盖单坐标和双坐标热估计与投影估计、短词纤维界、常规平稳下界、单步均匀起始首次相遇估计，以及干净有界回归学习器。给定这些输入，以下韧性断言是确定性的。对于其二元应用，附录 C.7 列出了在干净补全上使用的转移谓词，并针对每个固定置信度指数，给出了无小终端分层上这些谓词联合失败的上界。固定 $a > 0$。观察到 A后，选择大小为 $T \subset [n]$ 的 $s \le n^{1-a}$，并将从 T出发的每个状态转移替换为指向 T的任意状态转移。在 T之外，保持原有字母不变。接受均值可以同时依赖这两个表。后续机器（包括其固定初始状态）是隐藏的。

定理 23 对于干净表的所有此类修改，以下结论同时成立：若向被动学习器提供除 n之外的 T 上的抽象内部动作，则它能在训练长度为 m时学得所得终端均值。对于具有指定状态依赖条件均值的有界响应，若风险为 ϵ 且条件训练失败概率为 γ，则样本数和比特复杂度关于n、 m、$1/\epsilon$及 $\log(2/\gamma)$ 呈多项式增长，其次数取决于 a和干净事件常数。 T中的状态标识符、原表或外部表，以及所有终端均值均保持隐藏。若不提供抽象动作，则只要其大小的上界为 s ，计算开销便至多增加乘法因子 $s^{O(s)}$。更准确地说，每个动作对应的固定几何子学习器数量为 $O((m+1)(s^2+1))$。

根据 C.5 节，只需使用条件二阶矩均匀有界的观测值，学习一个在 T的所有位置都消失的有界均值 h。我们可以假设 $s \geq 1$。不进行修改的情形就是干净学习器。下面的所有常数都可以依赖于 a。对于大小至多为 $n^{1-a/2}$的每个集合 B，令 U_1^B表示在访问 B时终止的干净单步游走，令 U_2^B表示在发生合并或访问 B中的任一位置时终止的干净成对游走。它们在禁用起点处均以零填充。C.3 节中的中间块论证对所有这些集合同时成立。事实上，对于 $b=1$，禁片包含 $O(|B|n^{b-1})$ 个元组，因此内部访问造成的算子误差至多为 $|B|n^{-1+o(1)} \leq n^{-a/2+o(1)}$。为将干净误差与删除代价分离，取附录 C.3 中的干净燃烧时间 D，其中$P_b^D = \rho_b^D h_b \nu_b + E_b$。令 $V = (U_b^B)^D$，并令 $L_B = \rho_b^D V h_b \nu_b V$。于是

$$\|(U_b^B)^{3D} - L_B\|_2 \leq \|V E_b V\|_2 + |B| n^{-1+o(1)}.$$

事实上，内部访问估计将第一个算子与 $V P_b^D V$ 进行比较。根据行和与列和界，短词纤维给出 $\|V\|_2 \leq (\log\ n)^{O(1)}$。增加干净燃烧时间会使干净余项按多项式阶小到任意固定阶。秩一特征值为 $\lambda_B = \rho_b^D \nu_b V^2 h_b$。在 ν_b $2D+1$ $|B|n^{-1+o(1)}$ $\nu_b P_b^j = \rho_b^j \nu_b$ $\|\nu_b\|_\infty \leq$ 下，对访 $n^{-b+o(1)}$ 1 B 问时间应用并集界的代价为，因为，且 $\|h_b\|_\infty = O(\ )$。因此 $0 \leq \rho_b^{3D} - \lambda_B \leq |B| n^{-1+o(1)}$。当为空时，两个干净项仍然存在：中间块余项，以及对于成对情形因相遇造成的损失。在半径按多项式阶足够小的围道上应用秩一预解式，并按其中所示取块幂，可均匀地得到，

$$(U_b^B)^k = (\zeta_b^B)^k H_b^B \mu_b^B + F_{b,k}^B, \qquad \|F_{b,k}^B\|_{\infty\to\infty} \leq n^{C_a} e^{-c_a k}, \quad 1 - \zeta_b^B \leq n^{-c_a}. \tag{R1}$$

归一化非负因子满足 $\mu_b^B \mathbf{1} = \mu_b^B H_b^B = 1, \|H_b^B\|_\infty \leq 2$。这里，只需一个固定的$O_a(\log n)$ 燃烧阶段，就已能得到最后这个界。对于实际的 T，写作 $H = H_1^T, \sigma = \zeta_1^T$。至多有$n^{o(1)}s$ 个状态满足 $H < 1/3$：从每个状态出发，在足够长的固定对数烧入期内，计算访问 T的概率并对所有状态求和，同时在每个时刻使用一致纤维界。然后应用 (R1)。这个估计将使经数据测试扩大的 T处于同一个谱区间内。

引理 24（删除状态后的谱分离） 固定 $0 < a \leq 1$。在 ***(R1)*** 的干净表事件以及附录 *B.4* 的一致历史估计成立时，以下结论对所有大小至多为 $B \subseteq [n]$ 的 n^{1-a} 同时成立。记 $\sigma_B = \zeta_1^B$和 $\tau_B = \zeta_2^B$。令 C_B：若下一个字母在 B 之外将一个存活有序对合并，则将其映射到该对的公共端点；否则质量为零。采用 ***(R1)*** 中的归一化，

$$\mu_2^B C_B H_1^B = \frac{1+o(1)}{n}, \qquad \sigma_B - \tau_B \geq \frac{1-o(1)}{n}.$$

这两种误差对 B 都是一致的。在 $B = \varnothing, n(\sigma_B - \tau_B) \to 1$ 在此事件上成立，因此一致下界中的首项常数是尖锐的。

证明 固定事件中的一张表，并记 $s = |B|$, $H = H_1^B$，在 B 上将 H 延拓为零。取$D = \lceil c\log\, n\rceil$，使得两个核在 D 处的行误差（R1）均不超过 $e = n^{-5}$。根据（R1）之前的禁切片估计，同一个 D 对每个 B 都适用。令 $M = (\log n)^{O(1)}$ 为此表上长度不超过 $D+1$ 的所有词纤维的上界。由（R1），$\sigma_B^D = 1 - o(1)$ 均匀成立，且$\|(U_1^B)^D\mathbf{1} - \sigma_B^D H\|_\infty \le e$。因此，$\|H\|_\infty \le (1+e)/\sigma_B^D = 1 + o(1)$。对所有单一起始状态，将截至时间 D 访问 B 的概率相加，得到

$$\sum_y (1 - H(y))_+ \le (D+1)sM + ne.$$

事实上，在每个时刻，对于每个固定的词，最多有 sM 个起始状态处于 B 中，其中包括时刻零的访问。

设 u 在 $N = n$ 个按序排列的互异起始状态上服从均匀分布。在未删除的表上，两个独立均匀起点以 $D+1$ 首次相遇的概率为 $(1 + o(1))/n$：将附录 B.4 的互异到相等估计应用于以前 D 个字母构成的历史，并将第 $(D+1)$ 个字母作为下一个字母，然后对词取平均。以两个均匀起点互异为条件，会使这一质量乘以 $n/(n-1)$。对于固定的词、坐标和访问时间，该坐标至多有 sM 个起始状态可以访问 B。在终止步骤，每个起始状态至多能与 M 个对侧起点相遇。因此，删除任何访问过 B 的相遇，代价至多为 $2(D+2)sM^2/N = o(1/n)$。在指定的终止状态处，至多有 M^2 个有序起始状态可以相遇。在下界中，对每个剩余的相遇赋予 H 的权重，代价至多为$M^2[(D+1)sM + ne]/N = o(1/n)$。上界使用 $\|H\|_\infty \le 1 + o(1)$。我们得到

$$u(U_2^B)^D C_B H = (1 + o(1))/n.$$

均匀历史格拉姆估计给出了在删除前于 D 前相遇的概率 $O(D/n)$。任一坐标在 D 前访问 B 的概率至多为 $2(D+1)sM/n$。因此 $u(U_2^B)^D\mathbf{1} = 1 - o(1)$，而 (R1) 给出 $\tau_B^D u H_2^B = 1 + o(1)$。再次应用 (R1)，这次将其用于有界测试函数 $C_B H$，即可证明通量公式。以上所有损失仅使用了 B 的大小以及适用于整张表的相同纤维界和核界。因此，对于检查表后选出的集合，它们同样适用。

在存活点对上，置 $\Phi(i,\ j) = H(i) + H(j)$。对两个坐标使用相同的下一个字母，应用单步行走特征函数方程：

$$\sigma_B \Phi = U_2^B \Phi + 2C_B H + V_B.$$

这里，$V_B \ge 0$ 是仅有一个坐标进入 B 时，剩余坐标的期望 H-值。乘以 μ_2^B 得到 $(\sigma_B - \tau_B)\mu_2^B \Phi \ge 2\mu_2^B C_B H > 0$。由于 $0 < \mu_2^B \Phi \le 2 + o(1)$，下界随即成立。对于 $B = \varnothing$，普通游走具有 $\sigma_B = 1$ 和 $H = \mathbf{1}$，而 $V_B = 0$。于是，单位算子给出 $1 - \tau_B = \mu_2^B C_B \mathbf{1} = (1 + o(1))/n$。 ∎

随附的 Lean 模块 DeletionRootGap 验证了在给定通量和投影输入的情况下所述的有限计数不等式、单字母平衡、标量间隙及锐度推论。

应用引理 24 于实际的修改后的集合 T。式 (S2) 中的卷积分母为 $\sigma_T - \tau_T$；该引理将其逆界定为 $(1 + o(1))n$，并给出下文所用的能量估计 (S3)。

使低存活集合可观测。设 $u = \epsilon/[10^4(m+1)^2]$。如果指定起点处的信号均方至多为 $\epsilon/8$，那么零预测器已经达到这一风险。否则，使用

根节点处的对数长度后缀证明 (S1)–(S3)，从而得到某个长度为 $l = O_a(\log(2nm/\epsilon))$ 的词 w，使得

$$|K| = |\sigma^{m-l}\mu_1^T U_w^T h| \geq (2nm/\epsilon)^{-C_a}. \tag{R2}$$

事实上，从根节点出发的后缀条件均值为 $KH(s_0)$，误差可小到任意程度的多项式误差，其中 s_0 是指定起点，而 $H(s_0) \leq 2$。当所需窗口重叠时，较短的 m可由全词字典处理。对于长度至多为所需访问时域和测试时域之和的每个前缀 p，以 p 为前缀、以 w 为后缀的普通带标签词给出条件均值 b_p。根据（R1），

$$\sigma^{|p|} b_p = KH(f_p(s_0)) + O(n^{C_a} e^{-c_a(m-|p|-l)}). \tag{R3}$$

如果 p 进入闭集，则其真实端点位于 T 中，且两侧的均值都为零。由于下面的可见前缀和后缀长度都是对数级的，词箱概率的乘积仍为逆多项式。如果 m 小于它们长度之和加上必要的对数混合裕度，则全词字典的规模为多项式级。

在数值 $H(y)$ 中，选择一个阈值 $\alpha \in [u, 2u]$，使其与所有这些数的距离至少为 $u/[3(n+1)]$（在这至多 n 个点处分割区间）。令 $D = \{y : H(y) < \alpha\}$。它包含 T，大小为 $sn^{o(1)} \leq n^{1-a/2}$，并且在所有测试前缀处同时具有这一分离性。枚举 w（一个符号）、关于 $\sigma \in [0,1]$的精细有理数网格，以及 $|K|\alpha$ 在 $[0,2]$中的阈值。根据（R2）、（R3）以及单位区间上的 $|z^k - z'^k| \leq k|z - z'|$，网格间距取为 $\epsilon/(2nm)$的多项式级大小即可。再结合小于 $|K|u/(n+1)$固定分数的经验误差，其中一个分支便能准确识别每个查询端点是否属于 D。数值 K、H 和 D 本身从不提供。一旦轨迹到达 D，在至少还剩 t个字母时忽略它，代价很小：对于 $y \in D$ 和 $r \geq t$，$\mathbb{E}h(f_V(y))^2 \leq 2u + n^{C_a} e^{-c_a t}$，这由（R1）对 T和 $h|_T = 0$得出。

存活对的单侧进入。首先，假设 T 与干净常返类相交。其平稳分布 π 根据平稳下界具有 $\pi(D) \geq n^{-C}$。存在一个固定的对数 d_0，以及定义在互异安全状态对上的函数 b 、 $0 \leq b \leq 1$，它给出在合并之前，它们首次访问 D 的时间在第 $2d_0$ 步或之前发生差异的概率。我们断言 $\mu_2^D b \geq n^{-C'}$。从相同的长度为 $3d_0$ 的随机词出发，选取两个独立均匀状态。根据常规混合，第一个状态以至少 $\pi(D)/2$ 的概率最终到达 D。它在前 d_0 步中已经访问过 D 后才到达该状态的概率为 $o(\pi(D))$：根据定长纤维界，后一次访问的概率为 $n^{-a/2+o(1)}$；而从该次访问起至少还剩下 $2d_0$ 个常规步骤，因此条件最终概率至多为 $2\pi(D)$。在给定整个词和第一个状态的条件下，第二个独立均匀状态在这 $3d_0$ 步中的任一步或在起始时访问 D 的概率至多为 $(3d_0+1)M|D|/n = o(1)$，且对该词均匀成立。如果它从未访问过 D，而第一个状态最终到达 D，那么二者不可能在第一次访问之前合并。这就从均匀不同状态出发，以概率 $\Omega(\pi(D))$ 存活 d_0 步，随后在 $2d_0$ 内发生单侧进入。取 d_0 相对于平稳下界足够大，并应用 (R1)。如果两个亮状态在长度至多为 t 的每个词上的首次进入行为都相同，那么它们经过 $t - 2d_0$ 步后的杀死状态对在该测试中对 b 的质量为零。行投影和 $\mu_2^D b \geq n^{-C'}$ 表明，经过 t 步后仍然不同且二者均为亮态的概率至多为 $n^{C''} e^{-c_a t}$。这里使用的是 D 本身对应的杀死状态对；在同时进入两个不同暗状态之后，路径的倾向不起任何作用。

对于任意剩余长度 $r \geq 2t$，用具有相同入口特征的另一个亮态替换一个亮态，会使共享未来上的最终响应产生期望平方差$O(u + n^{C''} e^{-c_a t})$。如果二者都在 t 内进入，则以该停止前缀为条件。每个均值的期望平方至多为 $2u + o(u)$，因为至少还剩 t 个字母。如果二者都保持亮态并在物理上合并

则差异为零。仍然不同的概率代价至多为前述界限的四倍。

在没有安全直径的情况下覆盖相关状态。 使用长度为 $g = C_a \log(2nm/\epsilon)$ 的块，使其足以达到 (R1) 所需的任意多项式精度。选择同阶的 t，使先前的耦合误差至多为 u。l、g和t 中的常数都可以从固定表指数中选取，而 (R3) 所需的额外分离精度则由剩余的自由中间字母承担。如有必要，在确定这三个长度后，将 m 的字典截断扩大一次。记作 $\eta = u/n$。当 $\mu_1^D(z) \geq \eta$ 时，称 z 为重状态。若根风险超过 $\epsilon/8$，则当 $m - g \geq t$ 时，其经过 g 步仍保持亮态的概率至少为 $c\epsilon$：在此前任何一次进入时，剩余的均方响应至多为 $2u + o(u)$。因此，(R1) 给出 $(\zeta_1^D)^g H_1^D(s_0) \geq c'\epsilon$。对于每个重状态，通过将其绝对误差选为小于 $c'\epsilon\eta/2$，$(U_1^D)^g(s_0, z)$ 中相应的项均为正。因而，从实际根节点出发，每个重状态都有一个长度恰为 g 的可达的全亮前缀。从任意亮态出发，在一个新的 g-块中始终保持亮态并最终到达非重状态的概率，依据 (R1) 至多为 $2n\eta + o(u)$。使用表选出的集合都满足这两项断言。无需为每个外部状态提供访问词或平稳质量为正的条件。

计算根节点和所有全亮的 g 字母访问前缀的首次进入签名，长度最多为 t。每个签名存储一个代表元。给定一个代表元和一个新的 g 字母块，通过检查其各个前缀上的测试函数，可以判断是否命中过暗态。如果安全，块之后的测试函数会给出其端点签名。如果有已存储的代表元，则用它替换；否则以输出 0 终止。所有必需的前缀长度都不超过 $2g + t$，所有枚举和数据箱的规模都是多项式级的。重端点总有一个已存储的代表元。在留下 $2t$ 个最终字母的同时处理此类字母块。终止路径的输出为 0。否则，最终代表元和剩余的短后缀构成一个键。对于比较预测器，为每个键指定从实际物理代表元读取该后缀得到的真实响应。即使算法不知道该值，这个键仍有一个固定且定义良好的比较值。

比较连续的混合体：它们分别执行前 j 次和前 $j - 1$ 次字母块操作，然后对其余部分使用真实的表。在新字母块之前，给定一个活跃代表元，暗态命中造成的平方差至多为 $2u + o(u)$；每一种可能的命中之后都至少还剩 $2t$ 个未来字母。缺失签名带来的概率代价为 $2n\eta + o(u)$，此概率是从该代表元本身计算的，与其被选中的方式无关。具有代表元的安全端点与其替代代表元具有相同签名，在新剩余词上造成的平方差为 $O(u + n^{C''} e^{-c_a t})$。已经终止的历史彼此一致。对至多 m 次替换应用 L^2 中的三角不等式，得到最终均方比较误差 $O(m^2 u)$。将 u 中固定的数值分母取足够大，以覆盖这些通用的有界输出代价。用新观测拟合有界终端查找表，并验证所有分支。即使对于错误分支，代表元也至多有 $2^g + 1$ 个，因此该运行时间界对每个样本都成立。

如果 T 与干净常返类不相交，则改为在足够长的对数前缀处分裂。除了总词质量为 $O(\epsilon)$ 的部分外，路径要么已经到达 T（此时终端均值为零），要么在到达干净常返类后、触及 T 之前便已离开，且之后永远不会再触及它。这可通过在首次进入之前耦合干净过程，并结合常规的对数混合及其在核心上的平稳质量得出。在每个分箱内，同时拟合一个常数和干净全起点回归学习器。验证会在不知道哪种几何结构成立的情况下，从候选项中进行选择。对于在 T 上为零的任务，该构造无需使用状态标识符或内部动作，也不依赖 s 的多项式

次数。二次规模细节约简，以及对可选动作猜测进行的新验证，现在即可证明该定理。

C.7. 在原始独立同分布先验下补全外部状态

现在，我们将行编辑定理应用于原始随机表。对于每个固定的置信度指数A，我们允许任意依赖状态的终止均值，并对学习器隐藏例外闭集上的动作。

使用指数为 $A' = 2A + 10$ 的清洁估计。存在一个常数 k，使得不含大小至多为 k 的闭合分量的独立同分布表，在原始例外概率 $o(n^{-A'})$ 之外满足这些估计。清洁输入包括清洁回归证明中列出的所有均匀有界均值估计，以及均匀起始首次相遇下界。令 T 为原始表中所有大小至多为 k 的闭集的并集。对于任意固定的 $d > A + 1$，若 $|T| > d$，则依次加入这些集合会产生一个大小介于 d 和 $d + k$ 之间的闭集。出现这些大小中的任意一个 j 的概率至多为 $\binom{n}{j}(j/n)^{2j} = O_j(n^{-j})$。因此，在 $o(n^{-A})$ 之外，该并集的大小有界，记为 $s \le d$。其中不涉及任何预先指定的输出值。根据二项尾界和并集界，除超多项式小概率事件外，原图的最大入度至多为 $\log^2 n$ 。

对于一个指定的非空 T（大小为 $s \le d$），在$[n]$中独立均匀地重新采样其 $2s$ 个字母单元格，同时保留外部行。在仅以 T 的指定闭包为条件时，这个补全后的表恰好服从通常的独立同分布律。我们还需要给出一个下界：在这个 T 确实是上述并集的事件上，这样的补全不含小闭类的概率。 T之外的每个状态，其原始前向可达集大小都大于 k。除至多 $s(1 + \log^2 n)^{k+1} = o(n)$个状态外，其前 k 个前向层都避开 T，这是由原始逆邻域界得到的。对于任何剩余状态，这些层已经包含超过 k 个互异顶点：如果它们稳定在至多 k的大小，则整个前向轨道将是一个小闭集。以 $1 - o(1)$的概率，所有重新采样的首项都会落在这样的状态上。一个与 T 不相交的小闭集在原始表中也会是闭集，因而会被包含在 T中。一个与 T 相交的闭集现在包含这些首项中的一个，以及其超过 k 个互异的外部后代。因此，对于每个足够大的 n ，在每个保留的原始外部状态及实际并集事件上，补全不含小闭类的概率至少为 $1/2$。

如果这样的原始表不存在良好补全，那么至少有一半的补全抽样既没有小闭类，又不满足清洁估计。在分子中，可以舍弃实际并集和逆度限制，只对指定的闭合条件进行条件化。对各种选择和大小求和后，可将这部分例外的上界写为

$$2\sum_{s=1}^{d}\binom{n}{s}(s/n)^{2s}o(n^{-A'}) = o(n^{-A}).$$

当 T 为空时，原有的干净保证仍然适用。否则，取一个干净补全作为见证，并保持所有外部行不变。实际的表由该补全得到，方法是将 T 中的各行替换为其真实的闭合动作。应用定理 23，枚举大小不超过 d的每个动作。由此可得完整的同长度终端均值学习结果，其联合失败概率为 $o(n^{-A}) + \gamma$ ，且对于每个固定的 A，所需资源为多项式级别。确定性末位是无噪声响应这一特殊情形；将均值预测器在 $1/2$处设为阈值，使用平方误差 $\epsilon/4$ ，即可得到分类风险 ϵ。对于伯努利响应，当平方误差 $2\sqrt{\epsilon}$时，超额贝叶斯分类错误也至多为 ϵ

用于逼近成功概率时。若存在依赖于状态的正间隔，则相对于最优终端符号的误差可由平方误差除以间隔的平方来界定。

使用终端标签选择分类器。 对于定理 1 中的无噪声分类表述，给定用于构造候选者的所有样本和随机硬币。按照上述方法对其预测均值进行阈值处理和截断。在平方误差容差 $\epsilon/8$ 且条件失败概率至多为 $\gamma/2$的情况下，对正确分支运行前述拟合。于是，其分类风险至多为 $\epsilon/2$。如果有 J 个候选者，则抽取一批新的、包含 $\lceil 8\epsilon^{-2}\log(4J/\gamma)\rceil$ 个带标签的词。Hoeffding不等式和并集界表明，以至少 $\epsilon/4$ 的概率，每个经验分类风险与其条件总体风险之差均不超过 $1-\gamma/2$。选择经验最小化器，其风险至多为 ϵ。定义正确表补全的事件仅取决于转移。条件训练失败和该表异常共同给出了定理 1 中所述的两个概率。

定理1 的样本与精度安排。 这里，观测值是一个完整的均匀分布的m字词以及一个比特 Y。令 $A' = 2\beta + 10$，取相应的干净阈值 k，并令$d = \lceil \beta + 2 \rceil$。枚举至多具有 d 个状态的每个二元作用，包括空作用在内，总计至多 $1 + \sum_{s=1}^{d} s^{2s}$ 个选择。干净补全仅用于证明，所枚举的候选者中有一个具有较低风险。算法始终从原始带标签机器中进行抽样。

对于二元数据，在行编辑估计中取 $a = 1/2$，最终取 $d \le n^{1/2}$。所需的干净事件仅涉及状态转移。在指数为 A' 且额外保留一个单位的松弛量时，取中心化对块与附录 B.3 中的均匀对热度界、附录 B.2 中的短词纤维界以及下文的有限距离界的交集。中心化块的反对称部分给出常规混合，而来自均匀根的纤维界则给出最大的平稳原子界。从最重的平稳状态出发的一条路径提供平稳下界。附录 B.4 中针对给定共同历史的异点到相等估计，对历史取平均后，给出均匀根一步相遇下界。附录 D.3 中的幂估计给出对数燃烧行投影。特别地，如果 H_i 是来自两个独立平稳起始状态的存活质量，则在 $i = l$、$l+1$ 处的两个平稳历史 Gram 公式给出 $H_l - H_{l+1} = (1 + o(1))/n$ 和 $H_l = 1 - O(l/n)$。在足够长的固定对数燃烧阶段进行行投影，可得$|H_{l+1} - \rho H_l| = O(n^{-3})$。因此，在同一层上，$n(1-\rho) = 1 + o(1)$，从而提供分母保护。最后，取历史重叠（附录 B.4.1）、有界重邻居度（附录 D.2）和异常对匹配（附录 B.6）的表事件的交集。仅使用这些论证中的结构性界限。于是，对于符号掩码，它们的推论在常返类上同时成立。在 β 固定后，固定的对数乘数只有有限多个。无小分量层上的表异常为 $o(n^{-A'})$；因此，上述重新抽样计算适用于该交集。

令 S_k 表示独立同分布表不含大小至多为 k 的终端分量这一事件，并令G 表示刚才描述的交集。以下预算针对原始独立行。每个固定的对数窗口都在绘制表之前选定。

事件或估计值	其置信度的使用位置
四源矩	引理10 给出了原始分布下的典型中位数。其一个盈余搜索只需以 $1-o(1)$ 的概率成功。
洁净对热度与中心化块	附录 B.3 允许依赖于 A'的截断，控制 S_k上的热度，并对以下的行汉明扩展应用有界差分
历史与输出测试	在独立同分布律下。 S_k 的交集与失败代价为 $o(n^{-A'})$。一致纤维和中心化块带来常规混合；有限距离估计给出平稳下界。 附录 B.4、B.4.1 和 B.6 给出相遇、方差和成对分离估计
补全与数据	在该层上以相同预算并留有余量。$\Pr(S_k \cap G^c) = o(n^{-A'})$ 进入指定闭包的重新抽样总和上文。由此产生的原始表异常为 $o(n^{-\beta})$。在条件于由此产生的目标后，候选拟合和留出选择合计耗费至多为 γ。

见证补全有两种可能。令 $\mathcal{K}$ 表示其常返类。如果 $T \cap \mathcal{K}$为空，则燃烧长度为 $g_0 = C_\beta \log(2nm/\epsilon)$ 的前缀。在首次访问 T之前，将原始游走与干净游走耦合起来。该前缀最终位于 $T \cup \mathcal{K}$ 之外的概率至多为干净混合误差，并将其选为低于 $\epsilon/16$。以 $\mathcal{K}$ 结尾的前缀会在其剩余字母上产生一个普通的干净二元实验。以 T结尾的前缀会产生一个至多具有 d个状态的 DFA。在每个前缀分箱中拟合干净二元学习器，并枚举所有带有其掩码和起始状态的此类小型机器。在每个分箱内使用新标注词进行选择。给定前缀后，剩余字母是独立均匀的比特。对于正确候选者条件风险至多为 $\epsilon/8$ 、且每个良好分箱中的选择开销至多为 $\epsilon/8$ 时，组合预测器的风险小于 $\epsilon/2$。每个前缀分箱的概率为 2^{-g_0}，因此在分箱内收集样本会使样本量增加 2^{g_0} 倍。将失败容许度分配给各个分箱。对于 $T = \varnothing$ ，同样的论证只使用干净学习器。如果前缀或干净窗口无法容纳，则使用下面的短词规则。

对于 $T \cap \mathcal{K} \neq \varnothing$，在这个相同的见证补全上应用引理 24，其中 $B = T$和 $a = 1/2$。由于 $|T| \le d$，两个被杀死的根节点至少相隔 $(1 - o(1))/n$。入射前核仅使用外部行，而这些外部行与原始表一致。这为（S2）中的零均值细节拟合提供分母。使用附录 C.5 的正交细节检验。将其总数（包括粗粒度检验）限定为 $J = 3(m+1)(d^2+1)$，并令 $r = \epsilon/(32J)$，$U = 2n(m+1)/r$。对于前缀长度为 j 的细节检验，将一个新的训练样本 (W, Y) 变换为 $(W_{1:j}, Y d_*(W_{j+1:m}))$。学习器根据候选动作计算 d_* 。对于真实动作，变换后的响应在 T 上的均值为零。在给定完整 j 字母前缀的条件下，其二阶矩至多为一。在给定该输入的任意分箱后，同样的界仍然成立。不会根据隐藏状态进行拒绝。粗粒度均值（包括适用时的 $j = 0$ ）只需进行标量估计。表 2 按依赖顺序列出了其余选择。一旦指定了 A' 处的干净指数，对数窗口中的所有常数便随之固定。

要读取残差样本行中的指数，写作 $P + l \le c_w \log_2 U + O_\beta(1)$ 和$g + t \le c_h \log_2 U + O_\beta(1)$。显示的阈值为 $z^{-1} = O((n+1)U^2)$，因此一个词箱面板使用 $O_\beta((n+1)^2 U^{c_w+4} \log(2B/\delta))$ 条记录。终端键的数量满足 $H = O_\beta(U^{2c_h})$，从而读出面板包含 $O_\beta(U^{4c_h} r^{-2} \log(2HL/\delta))$ 条记录。窗口系数c_w， c_h 取决于在目标指数处选择的结构常数；这些是在候选枚举和验证之前、按面板计算的界。

干净行 v， k_s 关于 n 是多项式，且 h_* 由 β固定。模式矩阵是至多 $2k_s + 1$ 个矩阵的和，系数按绝对值计。其算子误差至多为

表 2：干净解码器和残差拟合的参数选择。

数量	选择与用途
干净二进制	令 v 为字典维度， k_s 为符号测试函数的数量， h_* 为固定乘积
解码器	递推深度，且 $\lambda_* = w_* v_*^2$，其中 w_*， v_* 是确定性干净下界
	分别针对 $\pi_{\min,\ +}$， $\sigma_{\min}(V)$。根据 Ap- 选取签名深度 $L_s = C_\beta \log\ n$
	附录 B.6。在所有解码器矩阵中使用元素容差 $e_* = \lambda_*/[32v(k_s+1)]$ ，并
	在半正定测试中平移 $\lambda_*/4$。
干净分箱精度	令 b_* 为非恒定核心的 $1 - B_l$ 干净逆多项式下界。
	采用原始精度，并使总体对角误差至多为 $b_* e_*/[C_*(16n)^{2h_*+6}]$。
	这里， C_* 界定了归一化特征的固定次数展开中的系数和
	（多项式界即可）。 选择整块乘积滞后和
	前缀混合预热 在 达到此精度后，两者均为 $O_\beta(\log n)$。使用词箱采样行
	如下所示，采用干净箱概率和容差。
残差预算	$u = r/[10^4(m+1)^2]$, 重质量 u/n。务必包含零预测器。
后缀与信号	取可见长度 $l = 1 + \lceil C_l \log\ U \rceil$ ，并在 $l-1$ 处应用 (S3)。如果信号均方
	在指定起点处超过 $r/8$，则 (R2) 中存在某个后缀满足 $\lvert K\rvert \geq k_0 = cr/n^2$
	对于绝对值足够小的 $c > 0$ 和足够大的 n。
进入阈值	令 $z = k_0 u/[48(n+1)]$。对于缩放阈值，网格间距为 z，而对于
	σ，其中 P 是下方最长测试前缀。
替换	固定单侧进入测试所用的 $d_0 = C_\beta \lceil \log n \rceil$ 。取 $t \geq 2d_0$ ，使存活
窗口	余项除以逆多项式见证质量后至多为 u。
	取 g ，使 (R1) 的误差至多为 $c' ru/n$ ，其中 c' 足够小 > 0. 两者都是
	$O_\beta(\log\ U)$。令 $P = 2g + t$。
未使用中间部分	要求 $n^{C_a} \exp[-c_a(j - P - l)] \leq z$ 满足 (R3)， $j \geq 2l$ 用于信号估计，并且
	$j \geq g + 2t$。如果某个条件不满足， $j = O_\beta(\log\ U)$，所有 2^j 值都可以直接拟合。
词分箱	使用长度至多为 P 的前缀和长度为 l 的后缀。每个概率 $p \geq$
	$2^{-(P+l)}$。对于 B 个查询词箱， $O(2^{P+l} z^{-2} \log(2B/\delta))$ 条记录足以保证同时达到——
	绝对误差 z。
终端读出	最多有 $H = 1 + (1+2^g)2^{2t+g}$ 个键。估计每个键的质量和未归一化响应
	均值，在新记录上达到 $r/(256H)$。在 $[-1, 1]$ 中拟合值。
选择	通过对 $f^2 - 2Y'f$ 的独立估计，比较 L 个有界残差预测器
	绝对误差为 $r/8$， 其中 Y' 是变换后的响应。 这需要
	每个共享面板需 $O(r^{-2} \log(2L/\delta))$ 条记录。

$(2k_s + 1)ve_* < \lambda_*/8$，而当前模式存在一个大小至少为 $\lambda_*/2$的负方向。在校准历史比率时，两个 B 估计中的误差 e会带来 $O(e/b_*)$的代价；每次后续对角乘法在其两个参数上的代价至多为 $O(n)$，在比率上的代价至多为 $O(n^2)$。显示的余量涵盖固定的递推（数值常数增大 C_* ），其中包括可观测中心化标量中的误差。在估计词分箱之前，总体偏差由烧毁行界决定；签名字典的维数不受之后这一燃烧阶段影响。干净行仅使用二进制核心；需包含两个常数分类器。对于不相交分支，所有这些估计都在相应的前缀分箱内进行，并使用其自身的填充。

在此处及下文中，每个批次都独立于定义其分箱和候选项的量，或者同时覆盖其中一个固定列表。具体而言，对于概率为 p 的可观测分箱，使用 $Y' \mathbf{1}_{\text{bin}}/p$。其方差至多为 $1/p$，因此对独立组均值取中位数即可得到所示的分箱成本。放大之前，各组只需达到常数成功概率。对于终端读出，在给定学习得到的键映射的条件下，质量 p_i 和未归一化响应 b_i 对其风险的贡献为 $\sum_i (p_i f_i^2 - 2b_i f_i)$ 加上一个常数。这两个量的绝对误差 a 至多会带来

$6Ha$ 的超额风险（针对估计得分的极小化器）。该拟合使用 $O((H/r)^2\log(2HL/\delta))$ 条记录，并允许键的质量任意小。验证得分的二阶矩至多为十。直接进行全词回归的界可通过对键进行相同的计算并取 $H=2^j$ 得到。

例如，指定起点 s 和另一个状态 x 可以满足 $f_0(s)=f_0(x)=s$，以及 $f_1(s)=f_1(x)=x$。x 的不同选择给出互不相交的事件，其总概率为 $(n-1)/n^4$。当接受位相反时，每个非空词的标签由其最后一个字母决定。在此事件上，任何常数分类器的误差均为 $1/2$。该动作属于所枚举的动作之一；其全部响应已经位于粗粒度后缀空间中。

对于信号行，令 $s_*=\sigma^{j-l}H_1^T(s_0)\le 2$ 表示 j-字母细节任务。若对于均匀分布的 j-字母词 W，有 $\mathbb{E}_W[h(f_W(s_0))^2]>r/8$，则由（S1）得到 $s_*\sigma^l\mu_1^T(h^2)\ge r/16$，从而有 $s_*\ge r/16$。在 $l-1$ 处应用（S3），并乘以 s_*^2。由于 $0<\sigma\le 1$，后缀分箱均值的平方的均值的主导下界至少为 $n^{-3}(r/16)^2$。根据 C_l 和 $j\ge 2l$ 的选择，（S3）以及（S1）中的前缀投影所产生的误差均小于该下界的一半。因此，某个长度为 l 的后缀具有绝对总体均值至少为 $c_1 n^{-3/2}r$。在根节点处，（R3）用 $KH_1^T(s_0)$ 近似该均值，其中 $H_1^T(s_0)\le 2$；这给出了网格所使用的较弱下限 $k_0=cr/n^2$。在 $[u,2u]$ 中选择一个真实边界，使其与每个 $H_1^T(y)$ 的距离为 $u/[3(n+1)]$。混合偏差、分箱估计、阈值舍入以及 $\sigma^{|p|}$ 的舍入，在（R3）中各自至多造成 z 的误差（将分箱均值截断到 $[-1,1]$）。因此，某个枚举分支能够识别所有经过测试的入口特征。附录 C.6 的区块替换耦合在平方损失中对每个块至多收取 $32u$：提前命中至多收取 $3u$，缺失重代表元收取 $3u$，同时进入之后收取 $12u$，互异显著幸存者收取 $4u$，并为谱余项留有余量。随后，对至多 m 个块应用 L^2 三角不等式，得到比较器风险至多为 $32m^2u<r/16$。对于低信号任务，零比较器的风险至多为 $r/8$。读出拟合和选择在两种情况下都能将风险控制在 r 以内。分别拟合并验证每个细节任务，然后再将它们汇总为试验动作；无需对候选列表取笛卡尔积。阈值、速率、后缀、键及其位长度的数量均关于 U 为多项式级。即使零比较器已经足够，也要枚举正信号分支，并在每张表上尝试两种几何分支。将每个拟合列表和数值扫描都限制在其规定的多项式预算内，包括错误动作猜测的情况。没有任何采样量使用未知阈值或物理代表元。

细节向量彼此正交，且平方范数至少为 $1/4$。在同时成功的情况下，总和的平方误差至多为 $4Jr=\epsilon/8$，而截断只会改善误差。在 $1/2$处进行阈值化，可得到一个分类风险至多为 $\epsilon/2$的二元候选。通过对多项式规模的拟合列表和动作列表使用并集界，将至多 $\gamma/2$ 分配给构造过程；对于重复使用拟合选择的情况，使用独立批次。置信度放大只涉及这些列表大小的对数。附录 C.7 中的新鲜最终分类面板会使失败概率进一步增加 $\gamma/2$。由于表异常为$o(n^{-\beta})$，取 $\gamma=n^{-\beta}/2$即可使总失败概率至多为 $n^{-\beta}$，对所有充分大的 n均成立。

固定置信度指数下的有限距离界。 对于每个固定的 $A>0$，存在一个常数C_A，使得独立同分布二元表中的最大有限有向距离至多为 $C_A\log n$，除去概率为 $O_A(n^{-A})$的情况。

证明 在目标处进行反向搜索时，令 m 表示已处理目标的数量，令 z 表示当前前沿。每个未访问源点都有两个独立单元格，其在已处理目标之外均匀分布。因此，其下一层前沿服从二项分布，均值为 $z(1-u)(2-u)$，$u=z/(n-m)$。在某个前沿达到 $K_0=\lceil k\log n\rceil$ 之前，以及在 $J_0=\lceil J\log n\rceil$ 轮过程中，对于较大的 n，其均值至少为 $1.8z$。对于$\theta=1/5$ 和 $a_0=1.8(1-e^{-\theta})-\theta>0$，在零点或 K_0 处消失的概率至多为 $e^{\theta K_0-a_0J_0}$

剩余正数且较小。将 J 选得足够大，即可得到任意所需的多项式指数。消失会遍历整个入分量。

到达 K_0 后，要求至少增长.4，直到前沿大小达到 $n/100$。在前一次成功后，已处理大小至多为 $J_0K_0 + 2.5z$，因此其均值仍高于 $1.8z$。Chernoff界将增长失败的概率限定为 $e^{-z/25}$；选择 $k \geq 25(A+6)$ 后，它们的并集以及随后对目标取并集的概率都足够小。

在将大型前沿 F 查询为逆向目标之前停止。以该查询记录为条件化条件，每个未访问源点的单元格在已处理目标之外仍保持独立均匀分布。从指定的外部起始状态出发，在命中之前的每次新探测命中 F 的概率至少为 $1/100$ 。未命中 F 而耗尽探测意味着该目标不可达。否则， $O_A(\log n)$ 次未命中的概率为 $O(n^{-A-5})$。在这么多次新探测内命中，则可得到一条长度为该次数加逆向深度的路径。对起始状态和目标取并集即可证明结论。所有条件化都由明确的逐单元格排除构成；在乘积测度论断中不存在对全局图事件的条件化。∎

在唯一常返类中，从质量至少为 $1/n$ 的平稳原子出发的这一路径长度给出 $\pi_{\min,+} \geq n^{-1}2^{-C_A\,\text{对数}\,n}$。

附录 D. 辅助估计

D.1. 存活对的速率

对于每个固定的 $A > 0$ ，存在一个整数 $b(A,\ r)$，使得除去概率为 $o(n^{-A})$ 的原始事件外，每个至多具有 $b(A,\ r)$ 个状态且不含终端分量的表，都对每个固定的 $c < 1/2$ 满足，

$$n(1-\rho(Q)) = 1 + O_{A,r,c}(n^{-c}). \tag{36}$$

这里， Q 作用于在其首个公共端点处被杀死的有序不同点对，而 $\rho(Q)$ 是其谱半径。行单射表通过条件化继承这一估计。

为清晰起见，固定 r 时，中心化证明在有界盈余搜索中使用 $\ell = \lfloor(\text{对数}_r\, n)/8\rfloor$ 。共有 $O_r(r^{\ell})$ 个槽位，并具有相同的盈余尾部。在到达指定点的局部部分中，除该点外的所有点至少都有一个已表示槽位，并且至多有 K 个顶点的全部 r 个槽位已被表示，这里使用了附录 D.2 的超额计数。每 $K+1$ 个字母的退出概率至少为 $r^{-(K+1)}$。在对数烧入期，这给出相同的安全历史热协议。在四源锚点计算中，将 $2^{-\max(k,h)}$ 替换为 $r^{-\max(k,h)}$，因为两个独立词必须共享其最后的 $\max(k,h)$ 个字母。对于两个矩，其首项常数均为 $\sum_{i\geq 0}(2i+1)r^{-i} = r(r+1)/(r-1)^2$。每个前沿顶点至多请求两个互异字母，每个副本请求一个；所有新单元概率和森林扣除均不变。单盈余典型锚点使用 $\ell_0 = \lfloor(\text{对数}_r\, n)/5\rfloor$、$O_r(r^{\ell_0})$ 个槽位，以及充分长的 $d = O_r(\text{对数}\, n)$；重复偏移词的代价为 r^{-d}。因此，同样的多项式小量级典型弗罗贝尼乌斯范数给出了中位数。附录 B.3 的汉明扩展、纤维界和集中性均适用。只需要表几何，无需符号或掩码条件。

因此， $Q^{jR} = \lambda^j \Pi + E^j$ 与 $\lambda = 1 - o(1)$， $\|E\| \leq n^{-\kappa/2}$以及一个具有归一化公共左行向量的非负秩一投影 ν，根据附录D.3中的预解式论证。这个 Q^R 主特征值是单的且孤立的，因此 $\nu Q = \rho\nu$ 满足 $\rho^R = \lambda$：交换性保持特征子空间，正性给出正根，并且 Q 的所有特征值的模至多为$\lambda^{1/R}$。令 $g = \mathbf{1} - Q\mathbf{1}$， $\nu g = 1 - \rho$。给两个独立均匀起点赋予一个公共随机词

长度为 $t = jR$，其中 j足够大且固定。它们未归一化的存活行 u_t 随后满足$\|u_t - (u_t\mathbf{1})\nu\|_1 \le n^{-4}$，这是由附录D.3中的秩一幂估计得到的。附录B.4 中的均匀历史估计给出 $u_t\mathbf{1} = 1 - O(t/n)$ 以及 $nu_tg = 1 + O(n^{-c})$，其使用了共同历史的一步异态到同态界并取平均。相除得到（36）。选择一个略大于 c 的指数，即可吸收所有对数因子。

D.2. 对所有预测类进行短后缀测试

对于一个固定的带标签自动机，$x \equiv y$ 如果它们在每个有限词（包括空单词）之后的接受响应都一致，那么它们满足右同余。本节中，字典在转移表或接受掩码绘制之前就已指定。

定理 25 对于每个固定的 $A > 0$，都存在一个显式多项式维字典，其中包含至多 k_A 个符号的乘积，这些符号位于长度不超过 $J_A \log n$ 的词末尾。在整个原始二元模型上，除概率为 $o(n^{-A})\, n$ 的情况外，互异预测类上的单位行字典的最小奇异值至少为 n^{-C_A}。该字典在绘制转移表和共享的公平掩码之前就已指定。此处的界适用于固定的 A。

证明 取 $j = \lfloor (\log_2 n)/8 \rfloor$，并取一个固定整数 K，使得 $3(K+1)/4 > A + 2$。除概率为 $O_K(n^{1-3(K+1)/4})$ 的情况外，每次深度为 j 的向外搜索至多有 K 条多余弧。事实上，一次搜索使用 $O(2^j)$ 个探针和状态。至少 $K+1$ 次未发现需要数量相当的新命名目标命中；对所有探针和所有根取并集即可得到该界。

对于两条分支上的 j 个字母延续，令 v_x 具有坐标 $2^{-j}\xi_{f_{0w_0}(x)}\,\xi_{f_{1w_1}(x)}$，因此$\|v_x\| = 1$。其 Gram 元素为 $K_{xy} = \prod_{a=0}^{1} G_j(f_a(x),\ f_a(y))$，其中 $G_j(u,\ v)$ 是两个符号的公共单词期望乘积。

令 T为完整表中大小至多为 K 的所有终端强连通分量的并集。对于任意给定的起始状态、端点 $z \notin T$及其深度为 j 的搜索，只保留在这个有限搜索图中存在一条路径通向 z 的顶点，以及它们之间的所有搜索槽位。如果从未发现 z，则下面的断言显然成立。否则，剩余子图是连通的：其中所有顶点都能到达 z。根据连通搜索的子图环秩界，它具有 $E - V \le K - 1$。除可能的 z 之外，其所有顶点至少都有一个留在内部的槽位。因此，至多有 K 个顶点的两个槽位都在内部，因为拥有两个此类槽位的顶点数等于 $E - V$加上没有此类槽位的顶点数。如果无法从某个顶点取得未表示的槽位，则其整个前向轨道将是闭合的，并由完全表示的行组成。它将包含一个顶点数至多为 K 的终端分量。这些顶点中的每一个都能在子图内部到达 z，因此该闭合分量包含 z，矛盾。通向未表示槽位的最短路径长度至多为 $K+1$（包括该槽位），因为此前的所有顶点都是互异的、完全表示的行。因而，一个随机的（$K+1$）块以至少 $2^{-(K+1)}$的概率使所表示的槽位保持不变。从起始状态出发并以 z 结束的每条实际长度为 j 的路径，在整个过程中都停留在所表示的槽位中：其所有源都出现在深度 j 之前，并且它们的槽位都已被探索。因此，对所有起始状态和所有$z \notin T$均匀地，

$$P^j(x,z) \le (1 - 2^{-(K+1)})^{\lfloor j/(K+1) \rfloor} \le C_K n^{-b_K}, \qquad b_K = \frac{-\log_2(1 - 2^{-(K+1)})}{8(K+1)} > 0.$$

选取一个整数 $M > A + 2$。若 $|T| > M$，其连通分量的某个并集的大小属于 $[M+1, M+K]$。每个指定的大小为 t 的集合以概率 $(t/n)^{2t}$ 成为闭集。对这些固定大小求并，得到 $\Pr(|T| > M) = O_{M,\ K}(n^{-M-1})$。

对于闭集 U，若 $|U| \le M$，则将其中的状态折叠为一个吸收态（如果 U 为空，则保留原始表）。在普通源处，给定闭集这一条件后，各独立槽位命中每个指定的普通状态 $^{-}$ 的概率为 $1/n$，命中吸收态的概率为 $|U|/n$。令 D_j 表示 $^{-}\quad ^{-}\qquad ^{-}$ 两个折叠端点相等的概率，并令 $\bar{m}_{xy} = \prod_a D_j(f_a(x)\ f_a(y))$，其中在吸收态处对内部状态求值。对于每个中心以及 d 个互异的普通邻居（均不同于该中心），这些 d 权重乘积的条件期望为 $O_{M,d}(j^{4d}n^{-2d})$。为说明这一点，执行星形计算中使用的 $2d$ 分支对比量。将吸收态和所有起始状态标记为时间零时刻。每个普通邻居的两个初始出边槽位分别在其两个对比坐标中提供单元素系数向量，因此完整的初始提议秩为 $2d$，即使中心已经处于吸收态也是如此。跟踪新发现的普通前沿状态上的秩。先前已揭示的单元格和吸收态都无法导向新发现的状态。与 (20) 中一样，每次真实的新鲜返回至多使该秩降低一。最终折叠后相等时，所有系数均为零，因此至少需要 $2d$ 次新鲜命中。共有 $O_d(j)$ 个槽位和状态时，每次命中的条件概率为 $O_{d,M}(j/n)$，其中也包括可能不均匀的吸收态。除闭集条件外，没有对外部表施加任何条件。

选择 $0 < a < \min(1/8,\ b_K/8)$ 和 d，使得 $d(1-a) > A+2$。对中心和 d 普通邻居应用并集界，可得所有边$\bar{m}_{xy} > n^{-a}$的星图的条件概率为 $O_{M,\ d}(j^{4d}n^{1-d(1-a)})$。为证明可以使用由表选出的 $U = T$，将此界乘以 $(t/n)^{2t}$，并对所有满足 $t \le M$ 的候选集合求和，其中也包括空集。它们的期望总数为 $O_M(1)$。在由此得到的事件上，相对于实际折叠后的 T，每个物理状态至多有 $M + d - 1$ 个重邻居。

对于固定的表和两个起始状态，将带符号相关性 G_j 拆分为折叠端点相同的结果，以及所有其余结果。第一项的绝对值至多为D_j，即使内部符号不同也是如此。其余部分的系数矩阵仅在互异的物理端点上有支集，且其中至少一个端点位于 T 之外。根据上述边缘估计，每个这样的有序系数均不超过 $C_K n^{-b_K}$。它们的总和至多为 1。因此，对称化后其弗罗贝尼乌斯范数平方至多为 $C_K n^{-b_K}$，且对角线为零。对于原始的单个公平掩码应用条件 Hanson–Wright 不等式，再对起始对取并集，可知余项在所有位置均不超过 n^{-2a}，且失败概率超多项式小。因此，对于足够大的 n，每个轻基 Gram 元素的绝对值至多为 $2n^{-a}$。

设 $B = M + d - 1$。在这个重边图中，类代表的度至多为 B。它们的最短区分词有一个单独的估计，该估计在给定任意状态转移表的条件下成立。

短见证。对于任意固定的、带有公平独立掩码的表，除概率为$n^2 2^{-L}$的情况外，每个不等价对都能由长度至多为 $L = \lceil (A+4)\log_2 n \rceil$的词区分。对于一个固定的词对，令 $e_{f_w(x)} - e_{f_w(y)}$ 的张成空间随 $|w| \le l$ 增长。如果它停止增长，那么它在两个字母的作用下均保持不变，因此包含所有更长的关联差向量。否则，它在 L 处的秩至少为 L（对于一个互异对）。所有这些关联差向量均与公平掩码正交的概率恰为$2^{-\mathrm{rank}}$，因为符号必须在其端点图的每个连通分量上保持一致。这是 Bassino、David 和 Nicaud 对 Moore 最小化进行平均情形分析时所依据的随机终态论证 [18, 引理 3.2–3.3]。

所有预测方向的对偶向量。令 t_x 为由一个值为 1 的常数坐标和坐标 $2^{-|w|}\xi_{f_w(x)}$（$|w| \le L$）组成的归一化向量。其归一化前的平方范数为 $S_L = 1 + \sum_{l=0}^{L} 2^{-l} = 3 - 2^{-L}$。这一选择使较大的权重集中在少数最短的

测试上。取展开后的单位行

$$V_x = v_x^{\otimes p} \otimes t_x^{\otimes B}, \qquad p > 2/a.$$

在一个预测类上，所有行都相同。考虑任意一组代表元和一个代表元x。对于代表元中的每个重邻居 y，选择一个能够将其与 x 区分开的短符号。将相应因子的乘积 $(1 + \xi_{f_w(x)} \xi_{f_w(z)})/2$在 z 处求值，可得到一个选择器$g_x(z) \in \{0,1\}$。它在 x 处的值为一，在重邻居上为零。将其实现为 $t_z^{\otimes B}$ 的一个线性泛函：每个因子使用常数和一个符号坐标，未使用的槽位读取常数。系数范数至多为 $(\sqrt{3}\,2^L)^B$，因为读取长度为 l 的归一化符号的代价至多为$\sqrt{3}\,2^l$。与 $v_x^{\otimes p}$ 张量积后，可得到一个具有相同范数、求值为 $K_{xz}^p g_x(z)$ 的对偶向量。所得方阵具有单位对角线，其非对角线元素的绝对值至多为 $(2n^{-a})^p$，因此它与单位算子的算子距离至多为 $n(2n^{-a})^p = o(1)$。对偶向量矩阵的范数至多为 $\sqrt{n}(\sqrt{3}\,2^L)^B$。因此，将这些求值作用于特征矩阵的转置，可得最小奇异值至少为$[2\sqrt{n}(\sqrt{3}\,2^L)^B]^{-1}$。选择器和物理恒等式仅用于证明这一不等式。学习器无需知道任何依赖于目标的选择，即可列出字典。

■

仅终端观测的接口。 假设某个相邻历史统计量以逐项范数将对角线逼近到 n^{-K}，其损失 q 至少为 n^{-C_1}，并且其可观测校准与原始平稳终端均值可在对数级历史上估计到 n^{-K}。此处 K可取为足够大的固定常数，以满足所需的最终精度。新字典的一个坐标是至多 $k = 2p + B$ 个普通符号测试的乘积，其中计入用于填充的常数因子。将这些因子分成两个均衡组，并递归使用（14）。在前缀为 v 的节点上，对每个取 $|w| = t$ 或 $t+1$的 wv 调用两个子节点估计值。在叶节点上，一个终端均值或一个已知常数即可。当 k 固定时，栈深度为 $\lceil \log_2 k \rceil$，且所有所需后缀长度均为 $O_{A,\,K}(\log n)$。将每个估计值截断到 $[-1,1]$。每一层对子节点误差的放大至多为 $O(q^{-1})$，对校准误差的放大至多为 $O(q^{-2})$。因此，根节点处固定的多项式倒数精度只需叶节点处固定的多项式倒数精度。所有坐标以及所需的全部词分箱数量均为多项式级；单位行归一化可用介于二分之一倍与一倍单位归一化之间的二进制因子代替。真实预测商具有定义良好的确定性字母转移和终端读出。在给定所述对角近似的条件下，由此构造出这些测试。仍需对获得该近似的概率给出界，以控制随机目标上的失败。

D.2.1. 短指定历史之后相遇质量的同时下界

为了归一化被动估计量，我们需要对可观测对角分数给出一个逆多项式下界。我们通过计数两条指定历史之后的相遇来得到该下界，这两条历史可以不同。

引理 26 固定 $C > 0$ 与 $T = \lceil C \log n \rceil$。以失败概率 $\exp[-\Omega_C(n/\log^4 n)]$, , 对长度至多为 T 的任意词 u、v 以及任意两个末字母 a、b 同时成立：至少存在 $c_0 n/(T+1)^4$ 个互异有序对 (i, j) 满足

$$f_u(i) \neq f_v(j), \qquad f_a(f_u(i)) = f_b(f_v(j)).$$

这里 $c_0 > 0$是绝对的。每组起始状态还可以从构成见证的试验中无重复地选取。

证明 设定 $k = \lfloor n/[100(T+1)^2] \rfloor$。先用第一个词进行 k 次试验，然后用第二个词进行 k 次试验。在每次试验中，按照固定规则选择一个未命名起点，并沿着指定词前进。立即为每个新回答命名，并在首次遇到已命名顶点时中止，包括本次试验中访问过的顶点。每个仍处于活动状态的源都是新命名的，并且拥有尚未查询的出向单元。在任何时刻，已命名顶点数至多为 $2k(T+1)$。因此，在给定任意过去信息的条件下，一次试验的失败概率至多为 $2kT(T+1)/n \leq 1/50$。根据条件伯努利支配，除概率 $2e^{-\Omega(k)}$外，每组至少有 $k/2$次成功。成功路径彼此顶点不相交，且所有此类路径的终止状态都没有暴露任何出向单元。即使之后失败的试验访问了较早的端点，也会在查询该端点之前停止。该论证也包括空单词。

揭示第一个组的成功端点上的字母 a。列表中任一头与先前头重复的概率至多为 k/n。除以概率 $e^{-\Omega(k)}$ 外，至少有 $k/4$个互异的头。在第二组的所有端点处，所要求的 b-单元格仍保持独立均匀分布，即使 $a = b$也是如此。每个单元格以至少 $k/(4n)$的概率命中一个第一组的头。由 Chernoff 界可知，至少有 $k^2/(16n)$次成功命中，除以概率 $e^{-\Omega(k^2/n)}$外。这些命中产生互异的有序见证对，因为它们的第二起点彼此不同；第一个起点可以重复使用。对少于 4^{T+2} 种词和字母选择应用并集界，即可完成证明。 ∎

设 π 为一个平稳分布，并假设均匀行 λ 在对数时间 G满足 $\|\lambda P^G - \pi\|_{\mathrm{TV}} \leq n^{-3}$。对于任意给定的长度为 $t \leq T - G$的共同历史 v，两个独立同分布的 π起点在经过 v 后到达互异端点，并在下一个给定字母处相遇的概率至少为 $c_1/[n(T+1)^4]$。具体而言，首先让两个独立均匀起点分别进行独立的 G-字母燃烧阶段。对于每一对燃烧词 u_1、u_2，将引理 26 应用于 $u_1 v$、$u_2 v$，然后取平均。用 $\pi \otimes \pi$ 替换燃烧后的初始乘积分布，代价至多为 $2n^{-3}$。因此，随后以平稳存活概率进行的任何归一化只会增大该下界；特别地，在对历史和下一个字母取平均后，有 $1 - H_{t+1}/H_t \geq c_1/[n(T+1)^4]$。这里既不假设由表选取的初始对，也不假设不同相遇事件之间相互独立。此处 n 的指数与对数历史长度中的固定常数无关。

在相同的对数词尺度下，一致短词纤维界还可为无条件单步相遇质量提供一个乘以 $1/n$ 的多对数上界。若每个相关纤维的大小至多为 M，则均匀独立起点对每个不同端点对赋予的质量至多为 M^2/n^2。在两个固定字母下，共享同一目标的端点对数量由单字母入度界可知至多为 nM。如上所述进行燃烧阶段、取平均并转移。若存活分母以下界为一个常数，则条件损失具有相同的上界。

作为参考，单组抽样步骤还可同时为所有 T字词给出秩 $\Omega(n/(T+1)^2)$，其中 $T \leq cn^{1/3}$ 且 $c > 0$ 足够小。对 2^T 个词取并集，并针对 $e^{-\Omega(n/T^2)}$。对于每个固定的 $\delta > 0$，Chapuy 和 Perarnau 在 [19, 定理 1 中4] 证明，以高概率，不存在长度小于 $n^{1/3-\delta}$的重置词。为进行校准，我们对所有对数级历史使用引理 26的同时相遇界。

D.2.2. 在干净表层学习任意接受集

固定一个置信度指数 A。在大终端中心块层上，排除原始表异常 $o(n^{-A})$ 后，一个终端标签学习器适用于每个二元输出赋值、每个训练长度 $m \geq 1$以及每个误差 $\epsilon > 0$。对于条件样本置信度 $1-\gamma$，其样本预算和比特运行时间关于 n、m、$1/\epsilon$ 以及 $\log(2/\gamma)$ 均为多项式。所有这些选择均使用同一个表事件。

设 $U = 2nm/\epsilon$。对于任意两个状态，令 $d_l(x,\ y)$ 表示它们在均匀分布的 l 字母延续上的不一致概率。干净谱块给出一个右 Perron 系数 $\psi_{xy} \geq 0$ 和一个共同的失配系数 $\alpha/2$，其中对于接受符号 z，$\alpha = \nu(1 - z_i z_j)$。对于所有掩码、状态对和 l，以下结论一致成立；存在绝对多项式前因子及依赖于 A 的常数 $\theta > 0$，

$$|d_l(x,y) - \rho^l \psi_{xy}\alpha/2| \leq n^{O_A(1)} e^{-\theta l}.$$

对于 $x = y$，所有项均为零。这可由块分解推出，余项至多为$n^{-\kappa/2}$ 每块 $R = O_A(\log n)$；较短的剩余块由均匀纤维在算子范数上界定。如果核心同时包含两种符号，则 (35) 给出 $\alpha \geq c\pi_{\min,\ +} \geq n^{-C_A}$。选择 $L = C'_A \log U$ 足够长，使得 L 处的余项至多为 $\eta/2$，$\eta = \epsilon/[32(m+1)]$。如果两个状态在长度不超过 L 的所有词上具有相同的符号签名，那么

$$d_l(x,y) \leq \eta \qquad (l \geq 0).$$

对于 $l \leq L$，它们的响应完全相同。对于更大的 l，使用 $d_L = 0$ 来控制那里的非负首项；其大小在之后的时刻只会减小。这样就不必要求每个物理上互异的有序行都具有正的长词信号。

识别短签名。 在固定的 $j = C''_A \log n$ 下，使用中心化双分支特征和附录 B.5.1 中的重星估计。这里只涉及仅基于表的 Gram 估计和星计数。在阈值 n^{-a} 处，根据附录 D.2，度数为 d 的星的失败概率为 $O_d(j^{4d} n^{1-d(1-a)})$；选取 $d(1-a) > A+2$。下文的签名论证使用每一对各自的短测试，无需分叉存活事件。现在，对偶选择器遍历长度至 $L+1$的测试；必要时增大 L，使其超过 $j+2$。特征在截至 $L+1$的签名上为常数。根据定义，互异的此类签名必然会在这些测试之一上有所不同，即使它们的长延续差异极小。轻中心化 Gram 条目的阈值、重度数上界以及张量因子的数量，只需要固定的 j 个输入。所选选择器的范数为 $O(2^{L+1})^B$，因此互异扩展签名上的字典具有最小奇异值 $U^{-O_{A(1)}}$。其坐标由有界的中心化短测试计算得出，且归一化已知；所有未知的中心化常数都是普通的平稳终端均值。类别权重至少为 $\pi_{\min,\ +}$。

引理 16 中的半正定特征测试现在会精确枚举常返平稳分布支集内、截至$L+1$ 的扩展签名；它只测试这些扩展类上取值为常数的带符号坐标。它使用固定数量的有界终端测试函数的乘积，其长度现在可达到 $O_A(\log\ U)$。为完整起见，相邻的矩校准在这些较长窗口上仍然准确。对于两个独立的平稳初始状态，在 l 处的非对角历史行是$u_l = u_0 Q^l = \rho^l w\nu + e_l$，其中 $w = u_0\psi$ 在该层上由绝对常数上下界控制：在固定的对数烧入期后，其质量为 $1 - O((\log n)/n)$，见附录 B.4。e_l 的绝对值逐项范数至多为 $n^{O_A(1)}e^{-\theta l}$。在确定所需的多项式倒数精度后，选取一个长度为 $C\log U$ 的完整块。然后，相对误差 $\|e_l\|_1/\rho^l$ 可以小于

任意预先指定的 U^{-1}固定幂次，因为 $\rho \geq 1-2/n$。特别地，当 n足够大时，在固定 C 之后，ρ^{-l} 的代价至多为 U。当 $B_l = z^\top S_l z$时，

$$1 - B_l = u_l(1 - z_i z_j), \qquad 1 - B_l \geq c\rho^l \alpha.$$

可观测比值（$1 - B_{l+1}$）/（$1 - B_l$）以所需的基本精度逼近 ρ。由于$1 - \rho \asymp n^{-1}$，（$S_{l+1} - \rho S_l$）/（$1 - \rho$）以所需精度在绝对值元素范数下逼近平稳对角线，这源于抵消首个非对角行以及两个边缘分布相等。具体而言，令 $D_l = 1 - B_l$。在此事件上，$D_l \geq c\rho^l\alpha \geq U^{-C_A}$，并且 $|D_{l+1}/D_l - \rho| \leq C\|e_l\|_1/$（$\rho^l\alpha$），其中常数已为下一行适当增大。原始绝对误差 $\zeta < D_l/4$ 会使其比值发生 O（ζ/D_l）的扰动。将所需比值精度的指数选为至少比最终对角线精度高三个 U 的幂，以便允许除以 $1-\rho$ 并扰动其分母。在$D_l \geq U^{-C'_A}$ 和 $1 - D_{l+1}/D_l \geq 1/$（$4n$）上使用带截断的经验保护条件，并在指数和常数上留出余量。先固定所需精度，再固定滞后系数 C。平衡递归（包括对标量中心化均值的估计）在 U中只损失固定幂次。利用后缀分箱和未见过的随机填充，在一个训练长度处计算原始均值。在确定精度后，每个所需后缀和填充的长度均为 O_A(log U)；如果字母数量过少，则使用整词字典，此时该字典包含 $2^m = U^{O_{A(1)}}$ 个条目。有限的初始范围也可以改用有界状态枚举。

一个近似转移系统。对于截至 L 的每个互异签名，任选一个实际出现的扩展签名。对于字母 a，使用其在 aw 处针对 $|w| \leq L$ 的条目，转移到由此得到的短签名。使用其空词符号作为输出。该机器的每一步都可以视为：沿当前物理代表元的真实后继前进，然后将其替换为具有相同短签名的代表元。对于任意长度的均匀未消耗延续，这种替换使最终符号发生变化的概率至多为 η。因此，至多m 次替换的望远镜式混合带来的风险至多为 $m\eta$；再考虑初始代表元选择所需的一次额外替换，总风险至多为 $(m+1)\eta$。此比较是在给定先前字母和第一个独立拟合样本组的条件下，使用独立均匀的剩余字母进行的。

对于指定起点，使用 $g = O_A$(log U) 个随机前缀比特；根据常规混合，其最终落在核心区域之外的概率至多为 $\epsilon/16$。真实核心分箱由至多 n 个重构状态中的一个近似。在每个分箱中，选择独立样本组上经验损失最小的状态；每个分箱使用足够多的样本，可使超额风险至多为 $\epsilon/8$，且联合失败概率至多为 $\gamma/2$，这由 Hoeffding不等式和并集界得出。如果 $m \leq g$，则使用字典。在新验证中将两个常数符号都作为候选项；即使训练长度下存在标签不同的罕见外部词，它们也能处理核心区域常值掩码。每个数值算法在数据失败时都会设定上限，而目标谓词仅使用独立于 ϵ，m 的固定表格窗口。这就证明了该结论。

D.2.3. 低于任意固定幂次的联合失败的终端标签学习$n^{-\beta}$，$\beta < 1$

固定 $\beta < b < 1$，其中 b 为有理数，并设 $c = \kappa = (1-b)/32$。我们沿用附录 B.1 中的可观测数组、有理岭反演和独立前缀拟合。只有它们的窗口和精度取决于 β。令 $\ell = \lfloor c\log_2 n \rfloor$、$d = \lceil 12\log_2 n \rceil$、$r = 4d+1$、$w = 2d+1$、$R = 3w + d$、$j = r$、$N = n(n-1)$ 和 $\Lambda = \log^{12} n$。取 $K = Q^R$，设 $a = N^{-1}\mathbf{1}^\top K$ 和 $B = K - \mathbf{1}a$。记 G 为如下事件：每棵深度为 ℓ 的向外广度优先树至多有一条额外弧。对根节点取并集界

并对 $O(2^{\ell})$ 个指定顶点中的两次命中取并集，得到 $\Pr(G^{c}) = O(n^{-1+4c})$。在 G 条件下，一个 ℓ 字母长的游走以至多 $(\ell+1)2^{-\ell}$ 的概率终止于某个指定顶点：指定额外弧的使用次数，即可确定其间的树路径。对于任意 $(\log^{4} n)$ 个已暴露状态，在 d 步内始终留在其中的概率，对于足够大的 n，小于 n^{-3}；这是通过在不相交的块中应用端点界得到的。

两个中心化估计。将 $f_r(i)$ 记为从有序对 i 出发在第 r 步首次相遇的概率。用于成对核的暴露估计给出

$$\mathbb{E}\left[\mathbf{1}_{G}\left(\|B\|_{F}^{2}+N^{-1}\sum_{i}(nf_r(i)-1)^{2}\right)\right]=O(n^{-1+4c}\log^{20} n).$$

下面说明 c 较小时所需的变化。在存活对中经过 $2d$ 步尝试获取新源后，在可观测安全历史上，获取失败的概率为 $O(d\log^{4} n/n)+2n^{-3}+d^{2}2^{-d}$。进行两次尝试并完成最终暴露后，对于指定的不同端点对，首次相遇概率至多为 $1/n+O(\log^{10} n/n^{2})$，或为 $O(\log^{10} n/n^{2})$。在同一张表上复制，并按照成对核证明中的方法检查部分暴露单元格的安全性。两个副本将 $f_r(i)$ 的截断二阶矩控制在第一个上界的平方以内。对于下界一阶矩，将两条路径植入，使其仅在最后一步相连；其概率为 $(n-2)_{2r-1}/n^{2r}$。在给定整个植入树的条件下，带有两条额外弧的向外搜索仍需两次命中新鲜的指定顶点，此时包括 $O(r)$ 个植入状态。其所有根节点均失败的概率仍为 $O(n^{-1+4c})$。在同一事件 G 上作差，即可得到所显示的首次命中估计。对于 B，四源枚举给出上行二阶矩 $6/n^{2}+O(\log^{6} n/n^{3})$。其下行均值矩将两个连接都植入为一片森林，并应用完全相同的条件界，此时包含 $O(R)$ 个额外状态；该矩至少为 $[6-O(n^{-1+4c}+R^{2}/n+R2^{-R})]/n^{2}$。减去 N 倍的这个下界，即得到 B 的所显示估计。

使用 $\lceil \log^{3} n \rceil$ 个副本来建立端点界，并对起始状态与终止状态对取并集，可得 $\max_{i,\,j} Q^{r}(i,\;j)\le \Lambda/n^{2}$ 在 G 上成立，但需排除一个超多项式小的例外集合。通过在前面添加状态转移，该界可推广到之后的每个时刻。现在由马尔可夫不等式可知，在$O(n^{-1+4c+2\kappa}\log^{20} n)$之外，

$$\|B\|\le n^{-\kappa},\quad N^{-1}\sum_{i}(nf_r(i)-1)^{2}\le n^{-2\kappa},\quad \|a\|_{2}\le \sqrt{\Lambda}/n.$$

最后一个不等式使用了条目界和 $a\mathbf{1}\le 1$。在反对称测试函数上，Q^{kR} 的作用等同于 B^{k}，因此普通游走有唯一的平稳分布 π，且在时刻kR，其与该分布之间的最大全变差距离至多为 $2n^{1-k\kappa}$。

平稳质量和对角校准。对于每个固定的 $A>0$，在概率 $O_{A}(n^{-A})$之外，所有有限有向距离至多为 $C_{A}\log n$。在处理了 m 个目标且当前前沿为 z的反向搜索中，下一个前沿在 $n-m-z$ 个未见行上服从参数为 $2u-u^{2}$,$u=z/(n-m)$的二项分布。在至多 $J\log n$ 轮且 $0<z<k\log n$的情况下，其均值至少为 $1.8z$。对于 $\theta=1/5$，在任一阈值处终止的指数马尔可夫迭代表明，保持较小且为正的概率至多为 $\exp(\theta\lceil k\log n\rceil-[1.8(1-e^{-\theta})-\theta]\lfloor J\log n\rfloor)$。在 k之后取足够大的 J。从规模$k\log n$开始，要求增长至 $n/100$，即增长因子达到$1\text{。}4$。在先前成功的情况下 $m\le O(\log^{2} n)+2.5z$，条件失败为 $\le e^{-z/25}$；取 $k\ge 25(A+6)$。如果反向搜索耗尽，则所有前驱都已找到。否则，在探测其大型前沿之前，每个未访问源处的每个向外新插槽都有至少 的概率命中该前沿。

经过 $O_A(\log n)$ 次未命中后，剩余失败概率为 $O(n^{-A-5})$，而未命中即耗尽意味着不可达。对起始状态和目标取并集。取 $A=2$ 并沿着从平稳质量最大的平稳状态出发的最短路径，可得对于固定整数 S，$\min_{x\in\operatorname{supp}\pi}\pi_x\ge n^{-S}$。

令 $S_t=\mathbb{E}_v[(\pi T_v)^\top(\pi T_v)]$ 对 $|v|=t$, $H_t=1-\operatorname{tr} S_t$，并且 $B_t=\xi^\top S_t\xi$ 表示独立公平符号。对于每个固定的 $T=O_{\beta,D}(\log n)$，在前述混合事件上，$1-H_T$ 的截断期望为 $O_{\beta,D}(\log^2 n/n)$。事实上，独立地让两个均匀起点经过 kR 个字母的燃烧阶段，直到它们的联合分布与 $\pi\otimes\pi$ 的距离不超过 $2n^{-10}$，然后使用一个共同词，长度为 T 个字母。起初互异的端点，只有在经过 $O(\log n)$ 个探针后命中一个新的指定目标，才会发生碰撞。单调性和马尔可夫不等式给出 $H_l\ge 3/4$，其上限为 T，失败概率仍具有相同的阶。由于 $\mathbb{E}_\xi B_t=1-H_t$ 且 $B_t\ge 0$，在选定的 t 处进行校准，代价也仅为 $O(\log^2 n/n)$。对于质量至多为一的子分布 μ，且 $l\ge 2r$，热核界给出 $\|\mu Q^{l-r}\|_2\le\sqrt{\Lambda}/n$。它与 $(nf_r(i)-1)_i$ 的配对绝对值为 $\le\sqrt{\Lambda}n^{-\kappa}$。应用于平稳非对角子分布后，在我们选定的 t 处得到下一项条件损失 $q=1-H_{t+1}/H_t\in[1/(2n),2/n]$：$t+1-r$ 处的质量介于 $H_t\ge 3/4$ 和一之间。

双分支分离。使用引理 11 的权重 m_{xy}。在物理状态上，按权重降序贪心地保留一个匹配。每条被省略的边都可将其自身权重与一条相邻且权重不低于它的保留边的乘积作为独特的计费对象。因此，其权重平方可由所有相邻边乘积之和覆盖。公式（18）将预期计费界定为 $O(j^8/n)$。除去该失败阶 $1/20$ 外，被省略的对称矩阵的弗罗贝尼乌斯范数至多为 $1/20$。根据（19），除去概率 $O(j^6/n)$ 外，对于每一对不同行，四个孙节点中至少有一个的碰撞概率至多为 $1/10$。于是，在行代表元上，保留边的权重至多为 0.55。限制在代表元上的单位对角结构矩阵的下界至少为 $0.45-1/20$。对于带符号扰动，通过增大其常数，公式（16）以失败 $O(n^{-2})$ 成立。根据单次返回界，对于互异的根和字母，分支碰撞概率平方之和的期望为 $O(nj^2)$。在阈值 $n^{2-\kappa}$ 处，其失败概率为 $O(j^2n^{-1+\kappa})$。取 $\eta=O(\log^7 n/n)$ 后，弗罗贝尼乌斯扰动平方现在为 $O(\eta^2n^{2-\kappa}+n^2\eta^4)=o(1)$，从而得到 $VV^\top\succeq I/4$。

数值参数和样本参数。设 $C=S+3$、$P_*=20(C+D+1)$、$k_0=\lceil 4/\kappa\rceil+1$、$K_*=k_0+\lceil 2(8P_*+6)/\kappa\rceil$ 以及 $t=K_*R$、$h=t+1$、$L=2t+j+5$。归一化存活历史行 y 更新为 $z=yK/(yK\mathbf{1})$。两个平稳历史的分母至少为 $1/2$。经过 k_0 个块后，根据递推 $\|z\|_2\le 2\sqrt{\Lambda}/n+2n^{-\kappa}\|y\|_2$，它们的范数至多为 $5\sqrt{\Lambda}/n$。对于第二个归一化行 x，更新差异为 $(x-y)B(I-\mathbf{1}z)/(xK\mathbf{1})$，从而后续收缩至多为 $n^{-\kappa/2}$。比较相差一个步骤的历史，转换为逐元素范数的代价至多为 n，除以 q 的代价至多为 $2n$。由于边缘分布一致，对角线误差可由非对角线误差涵盖。因此

$$\left\|q^{-1}(S_{t+1}-(1-q)S_t)-\operatorname{diag}\pi\right\|_{\mathrm{ent}}\le n^{-8P_*}.$$

由此可知，$S_h\succeq c'n^{-S-1}I$ 在循环状态上且可观测矩阵 $H=FV$ 具有 $\sigma_{\min,+}(H)\ge n^{-C}$。使用 $1-B_t\ge 1/2$ 进行校准的确定性比率偏差至多为 $2qn^{-8P_*}$。

使用填充 $R\lceil(6P_*+4)/\kappa\rceil$ 以及原始终端均值精度 n^{-6P_*}。较短长度使用其完整真值表。对于较长长度，第一批包含 $O(2^Ln^{12P_*}(L+\log(8/\delta)))$ 个词。同时进行箱集中性、填充并除以 q，可得到归一化数组误差 $\tau\le n^{-4P_*}$。采用岭惩罚 n^{-P_*} 后，字母算子误差为 $O(n^{-P_*+3C})$。精确词乘积的范数至多为

至多为 n^{C+1}，而真实读出范数至多为 n^{C}；按照原证明在错误位置之间展开，可将长度不超过 n^{D} 时的读出误差界定为 $O(n^{D+6C+2-P_*} + n^{2C+1-4P_*}) = o(1)$。因此，一致间隔超过 $1/2$。在指定精度 ϵ下，取长度为 $b_\epsilon = R\lceil(2+\log_n(64/\epsilon))/\kappa\rceil$的前缀。其非递归质量至多为 $\epsilon/8$，且 $2^{b_\epsilon} \le (n/\epsilon)^{O_\beta(1)}$。独立前缀拟合在固定递归分箱上使用维度为 $(2^h+1)2^{b_\epsilon}$ 的可实现情形的VC界。其误差在那里至多为 $\epsilon/2$。如果 $b_\epsilon \ge m$，则使用完整真值表。所有保护条件、数组归一化和有理数舍入均采用与原算法相同的位复杂度界。

表异常事件和掩码异常事件的总和为 $O_{\beta,\ D}((n^{-1+4c+2\kappa} + n^{-1+\kappa})\log^{20} n + n^{-2}) = o(n^{-\beta})$。这些异常事件仅取决于目标、 D 和 β。这证明了在 $\beta < 1$下公平标签、多项式长度情形的所声称速率。定理 1 中的扩展使用附录 C.7。

D.3. 近似入口签名和二阶置信度

当入口语言能够区分商的不同有序行时，入口检验可以准确识别该商。在必须以更高置信度覆盖的表上，这一条件可能失效。相反，我们只要求：在独立延续上，用具有相同短签名的另一个隐藏状态替换某个隐藏状态时，影响很小。这样便可以对最后几个字母进行普通的监督拟合。

定理27 固定 $D \ge 1$。对于每个 $A > 2$，存在一个固定整数 K和一个多项式时间被动学习器，使得在原始带标签目标概率为 $o(n^{-A})$的事件之外，它以样本概率 ϵ 达到误差 $1-\gamma$，且对每个预先指定的 $m \le n^{D}$ 均成立，前提是该表具有唯一的吸收行，并且不存在大小至多为 K的其他非空闭集。此处“其他”不包括其单元素集合。其样本量和运行时间关于 n、$1/\epsilon$、$\log(2/\gamma)$均为多项式，其中常数仅取决于 D、A。结合定理 *15* 可得，无条件目标异常事件至多为 $(8+o(1))/n^2$。

证明 取 K 大于 $A+4$，并将所有局部搜索截断值设为指数 $A+3$，以进行干净修复。除去概率为 $o(n^{-A})$的原始事件外，给定表中的所有有限有向距离至多为 $h = C_A \log n$，以下固定时域对应的所有对数长度词纤维均为多对数级，并且没有闭集的大小落在 $[K+1, n/4]$ 中。对于最后一个断言，求和 $\binom{n}{s}(s/n)^{2s} \le (es/n)^{s}$。对于固定的$K < s \le 2K+10$，该和为 $O(n^{-K-1})$；对于更大的 $s \le \sqrt{n}$，使用 $(e/\sqrt{n})^{s}$；对于剩余范围，使用 $(e/4)^{s}$。因此，在目标层上，每个外部状态都可以到达多于 $n/4$ 个状态，因为其可达集是闭集。

修复见证。令 x 为吸收态。仅替换其所在的行为独立均匀单元，从而得到表 B；为两张表赋予相同的接受符号。在独立同分布模型中，给定指定的吸收行后，该修复恰好服从原始的干净分布。对于层上具有纤维界的固定目标，至多有 $O_K((\log n)^{O(1)})$个状态能在不超过 K步内到达 x：将这类词对应的有界集合上的纤维求和即可。该集合之外的新头部可在 B中不超过 K步内到达多于 K个状态。它的前 K 个搜索层级与原图中的层级一致，而在原图中，可达集包含多于 K 个状态。因此，不可能生成一个至多包含 K 个状态且避开 x 的闭集。于是，以条件修复概率 $1-o(1)$，干净表没有小终端分量。

如果修复具有此性质，并且满足指数为 $A+3$ 的高置信度中心化块和干净学习器条件、普通直径和固定时域纤维事件，以及下述公平掩码平衡，则称其为良好修复。每个附加失败事件与无小终端分量条件相交后的原始概率为 $o(n^{-A-2})$。由于干净最大平稳质量至多为 $1/3$，掩码在每个

符号上的平稳质量至少为 $n^{-1}(\log n)^{O(1)}$，因此具有超多项式条件置信度。在指定层上不存在良好见证的概率，至多是发生失败且无小终端分量时联合修复概率的两倍。对于指定的吸收行乘以 n^{-2}，并对 x 求和。这就界定了不存在良好修复的概率。学习器仅使用原始表中的观测结果。在下文中，P, Q, π 指一个良好的干净修复，而 $\mathcal{K}$ 指其常返支撑。其界对每个指定初始状态都成立。所有未标记转移都与观测目标一致。

外部标签间隙。令 $L = P|_{[n]\setminus\{x\}}$ 为杀死的普通游走，并令 $S_j = L^j \mathbf{1}$。选择 $G = O(\log\, n)$，使得任意时刻 $\geq G$ 的每一条干净行与 π 的全变差距离都不超过 $d = n^{-8}$。在本段中，将 S_j 在 x 处延拓为零，包括 $j = 0$。令 $U = \{z \neq x :\ S_G(z) \geq 1/2\}$。其补集的规模为多对数规模：根据短词纤维界 M，对初始状态求和后，干净链在时刻 G 前访问 x 的期望次数至多为 $(G+1)M$。因此 $\pi(U) = 1 - o(1)$。对于任意 $z \in U$，条件于前 G 步存活。在接下来的 G 步中，命中 x 的概率至多为 $e = 2G(\pi_x + d) = o(1)$，因为在这些时刻中的每一个时刻，无条件干净端点的质量都是 $\leq \pi_x + d$。相应地，行 $L^{2G}(z,\cdot)$ 的质量至少为 $1/3$，归一化后与 π 的距离不超过 $e + d$。对于任何初始分布，只要其 G 步存活概率至少为 $1/2$，同样的结论也成立。根据并集界，距离 π 不超过 $e + d$ 的分布具有这一存活性质$e + d + (G+1)\pi_x$。经过迭代，再加上少于 $2G$ 步的终端块，可知从 $z \in U$ 出发，在每个 $j \geq 2G$，归一化存活端点分布都与 π 相差 $o(1)$，并且关于 j 均匀。对于终端块，可以以代价 $e + d$ 将输入分布与平稳分布耦合，而平稳分布的命中代价至多为 $(2G+1)\pi_x$。

我们记录相应的存活比较。对于 $M_0 = L^{2G}$, $f \geq 0$, $H = \max\, f$，以及$v = \pi f$，由 P^{2G}的支配可得 $\max\, M_0 f \leq v + dH$。只要后者为正，安全行就给出 $\min_U M_0 f \geq (v - (e+d)H)/3$。因此，对于所有足够大的 n，$\pi f \geq H/8$意味着同时有 $\pi M_0 f \geq (\max\, M_0 f)/8$和 $\min_U M_0 f \geq (\max\, M_0 f)/8$。初始列 S_j，$0 \leq j < 2G$具有$\pi S_j \geq 1 - (2G+1)\pi_x$。迭代给出：对于每个 $j \geq 2G$，均有 $\min_U S_j \geq \max S_j/8$。

从任意外部状态 y 出发，U 中至少有一个状态可在至多 h 步内到达。通往该状态的路径避开吸收行。沿着这条选定的路径，然后利用存活比较和平衡掩码，可知对于每个 $r \geq h + 2G$ 以及任一符号 b，有

$$\Pr_y\{\text{survive to } r,\ \xi_{f_{W_r}(y)} = b\} \ \geq\ 2^{-h}(\max_z S_r(z))/32.$$

如果从指定起点 s 出发，在长度为 m 时，常数预测 ξ_x 的误差大于 $\epsilon/8$，那么对于每个这样的 $r \leq m$，这至少为 $2^{-h}\epsilon/256$。

一个后缀可以区分一对物理持久对。首先，假设 $x \in \mathcal{K}$。令 $N = n(n-1)$。在良好修复 $K_0 = Q^R = \mathbf{1}a + E$ 上，$\|E\| \leq n^{-\kappa}$、$\sqrt{N}\|a\|_2 \leq (\log\, n)^{O(1)}$，且 $a\mathbf{1} = 1 - o(1)$。存在一个质量为一的非负行向量 ν 以及指数中的一个绝对常数，使得对于每个 $j \geq 1$，

$$\max_i \|Q^{jR}(i,\cdot) - (Q^{jR}\mathbf{1})_i \nu\|_1 \leq 2n(n^{-\kappa/2})^j. \tag{*}$$

以下是初等秩一论证的详细说明。设 $u = \mathbf{1}/\sqrt{N}$、$v = \sqrt{N}a$、$\alpha = v\overline{u}$。标量映射 $z \mapsto \alpha + vE(zI - E)^{-1}u$ 将 $[\alpha - 1/8, \alpha + 1/8]$ 映射到自身，并且根据 Neumann级数和 $\|v\| = n^{o(1)}$，它是一个压缩映射。其不动点 $\lambda = 1 - o(1)$ 具有右特征向量和左特征向量（$I - E/\lambda)^{-1}u$ 以及 $v(I - E/\lambda)^{-1}$。将二者的乘积除以它们的标量内积，即得到秩一

投影 Π。同样的级数给出 $\|\lambda\Pi - uv\| \le n^{-\kappa+o(1)}$。然后 $K_0^j = \lambda^j\Pi + (K_0 - \lambda\Pi)^j$，其中 $\|K_0 - \lambda\Pi\| \le n^{-\kappa/2}$。令 $\lambda^{-j}K_0^j$ 取极限即可证明 $\Pi \ge 0$。其所有非零行具有相同的归一化 ν。转到行绝对质量时，代价至多为 $\sqrt{N}$；将其标量系数替换为实际的存活质量，误差至多会加倍。这就证明了 (*)。

写作 $g_x(i,j) = \mathbf{1}\{$恰好是 $i,j = x\}$ 中的一种。使用两个独立的 π 起始状态，它们具有共同的长度为 t 的历史；其中 j_0 是充分大的固定值。它们在 g_x 处的非对角质量为 $2(\pi_x - S_t(x,x))$，这里 $S_t(x,x)$ 表示它们的联合对角项。对于每个确定性的此类历史，根据干净纤维界，其行 πT_v 的最大原子概率为 $o(1)$。对历史取均值后，结果为 π。因此，$S_t(x,x) \le o(1)\pi_x$。由于根据直径有 $\pi_x \ge n^{-C}$，且 (*) 适用于非对角子概率，我们便得到：对所有充分大的 n，$\nu g_x \ge \pi_x$。

令 $H_y(v)$ 表示在观测表上到 v 结束时（包括时间零）进入吸收态。若 y,z 在所有长度不超过 $t = jR$ 的词上具有相同的 H 签名，那么它们的干净轨迹不可能在时间 t 恰好只有一个端点位于 x：较早的首次访问会将二者区分开，而同时访问则会使干净路径从此永远相同。因此，(*) 将它们在时间 t 的干净存活概率限制为 $2n^{1+C}(n^{-\kappa/2})^j$。它们在时间 t 的观测终点只有在干净成对游走存活时才可能不同，因为沿该词发生的任何首次访问都是同时的。选取 $j = O_{D,A}(1 + \log(2m/\epsilon)/\log n)$，使得该界为 $\le \epsilon/(16m)$。因此 $t = O_{D,A}(\log n + \log(2/\epsilon))$。证明只需要初始 j_0R 和块 R 处的固定时域纤维；(*) 对所有 j 均有效。

使用近似签名进行学习。 如果 $m < 2h + t + 2G + 2$，则整词字典具有多项式域大小。否则，通过在长度 $h + 1 + t$ 以内收集每个前缀分箱中的两个符号，或每个前缀分箱中的一个符号，可以推断非恒定分支中的所有进入指示变量。事实上，已吸收箱是纯的，而其他每个箱以至少 $2^{-h}\epsilon/256$ 的概率包含两个符号，如上所示。每个箱的输入概率至少为 $2^{-(h+1+t)}$。因此，根据并集界，只需多项式数量的样本，其中包括 $\log(2/\gamma)$。计算长度 $h+1$ 以内所有前缀的 t-签名。每个可达状态都有一个长度至多为 h 的访问词。对于每个签名，使用长度至多为 h 的最短访问代表，并沿每个字母转移到扩展访问词所对应签名的一个代表元。将空访问词用作其签名的初始代表。对不一致数据施加确定性上限和默认值。

让这台机器运行前 $m - t$ 个字母。在每次替换时，当前代表元的真实后继与其新的代表元具有相同的短签名。根据（*），在剩余的独立延续下，它们给出不同最终符号的概率至多为 $\epsilon/(16m)$，因为其至少有 t 个字母。为进行形式化的望远镜论证，对每个 i进行比较：先模拟 i 个代表元步骤，然后在剩余的所有步骤中沿用物理表。相邻比较仅以截至第 i步的前缀为条件。将不同意的概率相加可知，在 t字母后缀上，从最终代表元得到的物理响应的误差至多为 $\epsilon/16$。在独立样本组上，通过经验风险最小化，在（代表元，末尾 t 个字母）上拟合一个二元查找表。其定义域大小至多为 $n2^t$，因此根据Hoeffding不等式和并集界，需要 $O((n2^t + \log(2/\gamma))/\epsilon^2)$个样本。这允许出现签名碰撞，并不要求有限签名构成右同余。

一个外部吸收态。不妨设 $x \notin \mathcal{K}$。将前 $g = O_{A,\ D}(\log n + \log(2/\epsilon))$ 个字母分入前缀分箱。取 g 足够大，使得利用中心化块在连续幂次处的常规混合界，偏离 $\mathcal{K}$ 的干净质量至多为 $\epsilon/16$。对 $m \le g$使用整词字典。对于每个分箱，在后缀上训练带有未知指定起始状态的常规干净学习器，同时使用两个常数预测器和独立验证。对于任何已经命中

常数的前缀，预测都是正确的。若尚未命中 x，且其干净端点位于 $\mathcal{K}$中，则后缀标签恰好是从该端点出发的干净目标的标签。其余分箱的输入概率总和至多为 $\epsilon/16$。由于燃烧阶段、字典和压缩映射都是均匀的，干净条件学习器的假设在其目标事件上对所有初始状态同时成立。在 2^g 个分箱之间分摊样本置信度，并在每个分箱中收集足够多的独立均匀后缀，会带来一个多项式因子。一个公共的独立验证面板在这些学习器、字典和常数预测器之间进行选择，同时将内部精度降低固定倍数。所有选择都使用带有单个最终符号的完整词。

最后，在大连通分量和分层结果中选择 $A > 2$，并通过验证汇总它们的候选项。如果两个层都不适用，则存在大小为 $s \in [2, K]$ 的某个闭集（两个吸收态也会形成这样的集合）。其概率至多为 $\binom{n}{2}(2/n)^4 + \sum_{s=3}^{K} \binom{n}{s}(s/n)^{2s} \le (8 + o(1))n^{-2}$。这就得到了所声称的统一例外。 ∎

D.3.1. 多项式词长上的一致性

以下是定理1的固定调度版本中多项式长度对应的参数调整。绘制表格前，先固定整数 $D \ge 1$。在整个构造过程中使用 $t = 4096DR$，$h = t + 1$ 和 $L = 2t + j + 5$。对于固定的 D，首次相遇界通过 $t + 1$ 给出 $H_l = 1 - o_{\mathrm{Pr}}(1)$，根据单调性，该结果对 $l \le t + 1$ 均成立。归一化历史在相同的前 40 个块之后收缩。转换为逐元素范数并除以 $q \asymp 1/n$后，残差为 $O(n^{2-(4096D-40)/25}) \le n^{-150D}$，当 n较大时成立。因此，历史和特征的奇异值界仍然给出 $s_* \ge n^{-2}$。针对这一个 t的目标事件适用于每个 $m \le n^D$。

若 $m < L + 2500DR$，则用优惠券收集界收集长度为 m 的真值表：$2^m = n^{O(D)}$。否则，填充偏差至多为 $4n^{1-125D}$。使用 $C2^L n^{242D}(L + \log(8/\delta))$ 个样本，分箱误差为 $O(n^{-120D})$。经过校准和归一化后，数组误差的上界为 $\tau = n^{-100D}$。令 $\lambda = n^{-26D}$。岭偏差至多为 n^{6-26D}，预解式扰动为 $O(n^{-48D})$，因此一个估计字母的误差为 $a_* = O(n^{-20D})$。对于长度为 $l \le n^D$ 的词，带有 $B_* \le n^3$ 的乘积界为 $lB_*a_* = O(n^{3-19D})$。应用读出后，词和终止向量的误差分别为 $O(n^{8-19D} + n^{5-100D})$ 和 $= o(1)$。因此，在整个范围内，间隔保持不变。使用与上文相同的 b 来拟合前缀；若 $b \ge m$，则切换到真值表，因为 $2^m \le 2^b$。m 个有理矩阵乘积的比特长度至多为每个字母的多项式界的 m 倍。因此，在所述范围内，样本量和比特运行时间均为多项式级。置信界适用于训练时使用的长度。