arXiv:2611.32955v1 [stat.CO] 24 Nov 2026

伪边际马尔可夫链的精确局部查询 s

作者：Zihan Xu

青岛大学数学与统计学院，中国青岛，*266071*xuzihan1@qdu.edu.cn

摘要

马尔可夫链的局部查询可能取决于被查询分量之外的历史更新。针对具有高斯图先验的伪边际链，我们利用有限的历史参数范围构造决策证书，从而在所有相邻记录均确定之前判定更新。在共享随机输入和精确算术的条件下，对于任意预先指定的有限查询序列，执行器都能还原参考参数、辅助变量和似然估计。我们对条件期望工作量进行界定，其中包括似然评估、历史范围构造和执行管理。对于指定的路径族和网格族，我们固定每个节点的历史长度以及工作量界中的所有其他参数。随着图规模增大，证书执行的预期增量工作量保持有界，而针对同一查询的指定缓存逐点评估器，其工作量则线性增长。在受控实验中，对于稀疏查询，证书执行比逐点执行和并行 Picard 执行更快；而对于完整查询，Picard 执行比证书执行更快。我们将该方法应用于正电子发射断层成像和 Landmine 数据的高斯过程分类。对于一个固定的数值断层成像后验分布，证书执行能在预算内完成所有区域平均分布容积查询，而使用相同初始缓存的完整执行一个也无法完成。在 Landmine 数据中，伪边际采样器与辅助变量增广采样器对两个位点之间对数长度尺度差值的后验汇总，在估计的蒙特卡洛不确定性范围内一致。在整个查询序列中，证书执行几乎判定了所有计划更新。

一些关键词：决策证书；图模型；局部查询；马尔可夫链蒙特卡洛；伪边际方法。

1. 引言

贝叶斯推断通常关注联合模型中一小部分参数的函数。示例包括正电子发射断层成像中的区域平均值，以及不同位点协方差长度尺度之间的对比量。在马尔可夫链蒙特卡洛（MCMC）中，这些输出可能取决于此前对查询子集之外参数的更新。当似然评估需要对潜变量进行积分时，例如在高斯
过程分类中（Filippone & Girolami, 2014），此类更新可能代价高昂。伪边际方法使用非负的无偏似然估计来构造一条链，使其不变分布具有所需的后验边际分布（Andrieu & Roberts, 2009）。我们研究如何从预先指定的参考链中恢复指定的联合输出，同时减少确定这些输出所需的历史计算量。

一类研究关注单次更新中的计算。在伪边际方法中，似然估计涉及计算成本与链混合之间的权衡（Doucet 等，2015；Sherlock 等，2015）。相关辅助变量可以减少似然比估计值中的噪声

c○ 2021 Biometrika 信托

（Deligiannidis 等，2018）。延迟接受使用廉价近似，在昂贵的评估之前筛选提议（Sherlock 等，2021）。子采样方法减少每次迭代中评估的似然项数量（Bardenet 等，2017）。一些方法使用控制变量和界限，以保持精确后验分布为不变分布（Cornish 等，2019；Prado 等，2026年）。

其他方法则重新组织参考链的模拟。预测性预取会并行评估可能的后续步骤，同时保留原有的轨迹顺序（Angelino 等人，2014）。Picard 方法将使用固定随机输入进行模拟表述为关于轨迹的不动点问题（Grazzi & Zanella，2026）。对于离散自旋系统，Feng 等人（2025）引入了向过去耦合的方法，通过追溯历史依赖关系并缓存已解析状态来恢复局部状态。Liu 等人（2026）提出的局部采样过程也利用了有关未解析状态的部分信息。

我们为具有局部似然和高斯图先验的伪边际链开发了一种精确的局部查询方法。完整记录包含一个参数、其辅助变量及其似然估计。对于加性随机游走和独立提议，在有限历史范围内，已解析状态和固定提议输入可为未解析的历史参数提供有限且可计算的范围。已解析的参数会为同一节点后续更新的范围提供锚点。在旧记录和候选记录固定且候选似然估计为正的情况下，对数接受比是邻近参数的仿射函数。这些范围可界定该比率，并能在所有相邻记录尚未解析之前，根据其原始均匀阈值确定是否更新。这种比较运用了回溯采样的原理（Papaspiliopoulos & Roberts，2008；Chanialidis 等人，2014；Li 等人，2016）。

递归执行器使用这些界限作为联合查询的决策证书，并缓存完整记录。我们证明，对于任意预先指定的有限查询集合和请求顺序，该执行器都会终止，并返回与完整参考执行相同的完整记录（定理1）。该结果假设初始记录、更新计划和随机输入共享，采用精确算术，且生成并评估每个候选项所需的时间有限。每个已调度候选项的似然估计至多计算一次。如果生成统计输出所用的额外随机输入也共享，那么两种执行还会返回相同的统计输出。

对于有界度图上的独立标准高斯提议，在满足收缩条件时，预期增量工作量有一个与查询数量成正比的常数上界，且该上界对图规模一致（推论 1）。采用这些提议时，定理 3 中的路径构造保证：对于每个固定的有限耦合强度，已解析块的数量具有几何尾界。两个结果都允许候选成本依赖于控制传播的提议输入，但要求存在一个公共可积成本界。对于指定的路径和网格调度，证书执行的预期增量工作量有界，而指定的缓存逐点评估器的预期增量工作量则与图规模呈线性关系（推论 2）。比较时使用相同的单记录查询、初始记录缓存和随机输入，并假设候选似然估计几乎必然为正。这些规模比较固定了模型参数、输入分布、每个节点的历史长度、初始参数界限、操作成本和公共成本界。我们还通过对提议输入和已揭示阈值进行条件化，界定已完成请求后的剩余工作量（定理 2）。工作量核算包括记录评估、输入与范围生成以及执行管理；准备和统计输出的成本则与增量工作量分开计算。

受控实验测量了块解析的工作量，并比较了针对相同查询的执行策略。对于所研究的双节点查询，与缓存逐点执行相比，证书执行所需的候选似然评估次数更少，且耗时少于逐点执行或并行 Picard 执行。在全图比较中，并行 Picard 执行更快。

在正电子发射断层成像中，我们推断左侧岛叶平均分布容积的后验分布。我们将证书构造扩展到由 Potts先验在体素间耦合的二元模型标签。固定求积和分箱定义了数值后验分布。从同一初始缓存出发，证书执行在时间预算内完成了所有预定的区域查询。完整执行在完成第一个查询之前就耗尽了时间预算。

对于 Landmine分类数据，我们推断两个站点的对数协方差长度尺度之差。伪边际采样器使用无偏似然估计，并联合更新每个位点的对数长度尺度和对数振幅。早期决策频繁发生，几乎每次计划更新都能在完整查询序列上得到解决。后续请求可能需要访问此前证书跳过的记录。后验摘要与以相同后验分布为目标的辅助增广采样器所得结果一致，差异在估计的蒙特卡洛不确定性范围内。

2. 参考链与查询

2.1。目标分布与参考链

我们针对具有标量节点参数的图模型开发该方法。设 $G=(V,\ E)$ 为一个有限无向图，其节点集为 $V=\{1,\ \ldots,\ n\}$，边集为 E。在节点 $j\in V$ 处，观测数据 y_j 的似然为 $l_j(\theta_j)$，其中 $\theta_j\in\mathbb{R}$ 为标量参数。对于 $\theta=(\theta_1,\ \ldots,\ \theta_n)\ \in\mathbb{R}^n$，考虑后验密度

$$\pi(\theta)=\frac{1}{Z}\left\{\prod_{j=1}^{n}\ell_j(\theta_j)\right\}\exp\left\{-\frac{1}{2}\sum_{j=1}^{n}a_j\theta_j^2-\frac{1}{2}\sum_{\{j,k\}\in E}w_{jk}(\theta_j-\theta_k)^2\right\},\tag{1}$$

其中，$a_j>0$ 是固定的节点系数，$w_{jk}>0$ 是固定的边权重，Z 是正且有限的归一化常数。似然项具有局部性，而先验分布则将邻近参数耦合起来。

在节点 j，令 $m_j(\mathrm{d}u_j\mid\theta_j)$ 表示辅助变量 $u_j\in\mathrm{U}_j$ 的模拟分布。假设$\widehat{\ }$有限的非负似然估计量 $l_j(\theta_j,u_j)$ 满足

$$\int_{\mathcal{U}_j}\widehat{\ell_j}(\theta_j,u_j)\,m_j(\mathrm{d}u_j\mid\theta_j)=\ell_j(\theta_j),\qquad\theta_j\in\mathbb{R}.$$

记作 $u=(u_1,\ldots,u_n)$，并令 $g(\theta)$ 表示式（1）中的指数因子。相应的伪边际目标（Andrieu & Roberts, 2009）为

$$\widetilde{\pi}(\mathrm{d}\theta,\mathrm{d}u)=\frac{g(\theta)}{Z}\left\{\prod_{j=1}^{n}\widehat{\ell_j}(\theta_j,u_j)\,m_j(\mathrm{d}u_j\mid\theta_j)\right\}\mathrm{d}\theta.$$

对 u 积分可恢复式（1）中的后验分布。节点 j 处的完整记录为 $R_j=\widehat{\ }(\theta_j,u_j,l_j(\theta_j,u_j))$。提议被拒绝时，我们保留 R_j。通过取 $l_j(\theta_j,u_j)\ =l_j(\theta_j)$，即可纳入确定性似然评估。

固定初始记录 $R_j^{(0)}$、$j\in V$，其似然估计为正，并设置一个包含 T 个节点更新的调度 j_1、……、$j_T\in V$。对于在节点 j 经过 t 次更新后的记录，记为 $R^{(t)}{}_j$。在节点 j 处，从 x 移动到 y 的参数提议具有固定、正且有限的密度 $q_j(x,\ y)$。在第 t 次更新时，令 $j=j_t$，并令 $(\theta,\ u)$ 表示经过 $t-1$ 次更新后的状态。抽取 $\vartheta\sim q_j(\theta_j,\ \cdot)$，然后抽取 $v\sim m_j(\cdot\mid\vartheta)$。

接受概率为

$$\alpha_t = \min\left\{1, \frac{g(\theta^{j\leftarrow\vartheta})\,\widehat{\ell_j}(\vartheta, v)\,q_j(\vartheta, \theta_j)}{g(\theta)\,\widehat{\ell_j}(\theta_j, u_j)\,q_j(\theta_j, \vartheta)}\right\}, \tag{2}$$

其中 $\theta^{j\leftarrow\vartheta}$ 是将 θ 的第 j 个坐标替换为 ϑ 后得到的。令 U_1、……、U_T 为 $(0,1]$ 上的独立均匀阈值，且与用于生成提议的随机输入相互独立。若 $U_t < \alpha_t$，则设 $R_j^{(t)} = (\vartheta$、v、$\widehat{\ell}_j(\vartheta$、$v))$；否则设为 $R_j^{(t)} = R_j^{(t-1)}$。所有其他节点记录保持不变。似然估计为零时，得到 $\alpha_t = 0$。每个节点核都保持 $\tilde{\pi}$。固定用于生成提议的底层随机输入和阈值 U_t，即可确定参考轨迹。这种固定输入表示也构成了马尔可夫链 Picard 模拟的基础（Grazzi & Zanella，2026年）。

2.2 查询与计算工作量

固定一个有限集合 $Q \subseteq V \times \{0$、……、$T\}$，其中包含请求的节点与时间对，并确定发出这些请求的顺序；该集合和顺序均独立于参考链的随机输入。请求的输出是带索引的集合

$$\mathcal{R}_Q = \left(R_j^{(t)}\right)_{(j,t)\in Q}. \tag{3}$$

此集合中的所有记录均来自参考链的同一次实现。令 $S \subseteq V$ 为 Q 中出现的节点集合，并记作 $r = |S|$。如果在相同的 K 个预定时刻 $0 \le t_1 < \cdots < t_K \le T$ 查询这些节点，则 $Q = S \times \{t_1$、……、$t_K\}$ 且 $|Q| = rK$。

对于上述共同查询时刻，固定一个可测实值输出函数 h，并记作 $Y_k = h\left((R^{(t_k)}{}_j)_{j\in S},\ \xi_k\right)$，$k = 1$、……、$K$。条件输出抽样使用具有预定分布的辅助输入 ξ_k，这些输入在不同的 k 之间相互独立，且独立于参考链。我们在不同执行方法之间共享这些输入；对于确定性输出，则省略这些输入。h 的选择包括参数对比、用于均值估计的区域条件均值，以及用于分位数估计的条件抽样的区域平均值。令 $\mathrm{I} \subseteq \{1$、……、$K\}$ 为非空且预先指定的保留索引集合。经验均值和正值比例为

$$\overline{Y} = \frac{1}{|\mathcal{I}|}\sum_{k\in\mathcal{I}} Y_k, \qquad \widehat{p}_+ = \frac{1}{|\mathcal{I}|}\sum_{k\in\mathcal{I}} \mathbf{1}\{Y_k > 0\},$$

其中 $\mathbf{1}\{\cdot\}$ 表示指示函数。经验分位数使用保留的输出和固定的插值规则。

固定一个加性成本模型。令 W_{setup} 表示准备工作，令 $W_{\mathrm{out}}(Q)$ 表示组装所请求的记录并计算其统计输出（包括任何条件抽样）所需的工作量。从准备好的状态开始，令 $W_{\mathrm{back}}(Q)$ 表示新的似然计算成本，令 $W_{\mathrm{manage}}(Q)$ 表示解析这些记录所需的其余工作量。我们将增量工作量和总工作量定义为

$$\begin{aligned} W_{\mathrm{inc}}(Q) &= W_{\mathrm{back}}(Q) + W_{\mathrm{manage}}(Q), \\ W_{\mathrm{total}}(Q) &= W_{\mathrm{setup}} + W_{\mathrm{inc}}(Q) + W_{\mathrm{out}}(Q). \end{aligned} \tag{4}$$

准备工作包括数据和图处理、初始记录和缓存，以及任何预先计算的随机输入或索引。管理工作包括候选生成、范围计算、缓存访问和调度；每项操作只计费一次。基于现有缓存进行的比较会标明该缓存，并单独核算其准备工作。我们将似然评估次数和测得的经过时间与这些工作量指标分开报告。

3. 使用决策证书执行3.1. 历史范围与决策证书

当固定参考轨迹中的所有组成部分都已确定时，记录即为已解析。在更新 t时，定义

$$p_k(t) = \max\left(\{0\} \cup \{s \in \{1, \dots, t-1\} : j_s = k\}\right), \qquad k \in V,$$

其中， $p_k(t)$ 是节点 k 在 t 之前最近一次计划更新的时间，零表示其初始记录。对于 $j = j_t$，旧记录为 $R^{(p_j(t))j}$。我们先解析这条记录，并将其与固定的提议输入结合，用以生成并评估候选项。式（2）中的接受比率使用了 $R^{(p_k(t))k}$ 的参数分量，涉及相邻节点 k。我们将这些参数置于区间内，以检验是否能在解析所有相邻记录之前判定此次更新。

界链在共享随机更新下传播可能状态的集合（Huber，2004）。Murdoch 与 Green（1998）利用 Metropolis–Hastings 接受概率的界限，在连续状态空间上耦合马尔可夫链。对于加性随机游走和独立提议，我们根据固定的提议输入和一条已解析记录构造明确的历史范围。从节点 k 在 $p_k(t)$ 时或之前最近一条已解析记录开始。记其参数为 x_0，将节点 k 后续的更新记为 $s_1 < \cdots < s_m \le p_k(t)$。对于随机游走， δ_i 是 s_i 时的固定提议增量；对于独立提议， η_i 是固定候选值。对于 $m \ge 1$， $R^{(p_k(t))k}$ 的参数分量位于

$$H_m = \begin{cases} \left[x_0 + \sum_{i=1}^{m} \min(0, \delta_i),\; x_0 + \sum_{i=1}^{m} \max(0, \delta_i) \right], & \text{random walk}, \\ \left[\min\{x_0, \eta_1, \dots, \eta_m\},\; \max\{x_0, \eta_1, \dots, \eta_m\} \right], & \text{independent proposal}. \end{cases} \tag{5}$$

对于 $m = 0$，取 $H_0 = \{x_0\}$。每次更新要么保留其先前的参数，要么采用其候选项，由此可通过归纳法得到这些界限。从节点 k 的一条新已解析记录中，重新计算相同后续时刻的范围，只会使这些范围更加紧致。

$\widehat{\ }$ $\widehat{\ }$将已解析的旧记录写为 $(x,\ u_j,\ l_j(x,\ u_j))$，将候选项写为 $(\vartheta,\ v,\ l_j(\vartheta,\ v))$。假设 $\widehat{\ }l_j(\vartheta,\ v) > 0$。令 $\mathrm{N}_j = \{k : \{j,\ k\} \in E\}$，并将 z_k写作 $R^{(p_k(t))k}$ 的参数分量。展开图先验可得对数Metropolis–Hastings比率

$$\Lambda_t = C_t + \sum_{k \in N_j} c_{tk} z_k, \qquad c_{tk} = w_{jk}(\vartheta - x),$$

$$C_t = \log \frac{\widehat{\ell}_j(\vartheta, v) q_j(\vartheta, x)}{\widehat{\ell}_j(x, u_j) q_j(x, \vartheta)} - \frac{1}{2}\left(a_j + \sum_{k \in N_j} w_{jk} \right)(\vartheta^2 - x^2).$$

此时， C_t 中的所有术语均已知。对于每个邻居，根据公式 (5) 获取一个区间 $z_k \in \underline{[z_k}, \bar{z}_k]$；对于已解析记录，则使用单元素集合。下界和上界为

$$\begin{aligned} \underline{\Lambda}_t &= C_t + \sum_{k \in N_j} \min\{c_{tk}\underline{z}_k, c_{tk}\bar{z}_k\}, \\ \overline{\Lambda}_t &= C_t + \sum_{k \in N_j} \max\{c_{tk}\underline{z}_k, c_{tk}\bar{z}_k\}. \end{aligned} \tag{6}$$

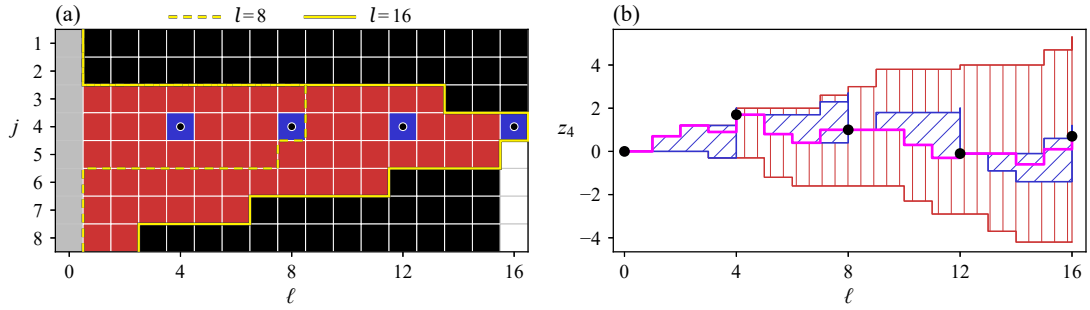


图 1. 具有固定加性随机游走输入的八节点路径上的历史解析和合法范围。每轮按顺序更新节点 1l 8，……，；节点 4 在第 4、8、12 和 16 轮更新后被查询。初始参数为零，节点精度为 1/2，边权重为 1/4，似然均为 1。（a）第 16 轮查询后的记录：蓝色点状单元格表示已查询，红色单元格表示其他已解析，黑色单元格表示未解析，灰色单元格表示初始状态，白色单元格表示当前查询后的状态。虚线和实线黄色前沿分别对应第 8 轮和第 16 轮。（b） z_4 的范围，即节点 4 在第l 轮更新后的参数。本面板按顺序显示四次查询。红色竖线阴影表示来自初始锚点的范围；蓝色斜线阴影表示来自最新已解析锚点的范围。每次查询时，蓝色范围都会收缩至返回的参数值。洋红色线条显示完整参考路径以供比较，初始状态和查询状态处以黑点标示。

替代文本(b)面板(a)已解析的各节点和更新轮次中参数范围。在每次查询时，这些范围都会收缩到参考路径。

因此 $\underline{\Lambda}_t \le \Lambda_t \le \overline{\Lambda_t}$，其中

$$\overline{\Lambda}_t - \underline{\Lambda}_t = |\vartheta - x| \sum_{k \in \mathcal{N}_j} w_{jk}(\overline{z}_k - \underline{z}_k).$$

已解析的邻居对该宽度的贡献为零。

接受检验对照一个固定的均匀阈值来细化界限，如同回顾抽样中那样（Chanialidis 等人，2014；Li 等人，2016）。零候选估计值会被立即拒绝。对于正的估计值，将 (6) 中的界限与原始阈值 U_t 进行比较：

$$\begin{cases} \text{accept}, & \log U_t < \min\{0, \underline{\Lambda}_t\}, \\ \text{reject}, & \log U_t \ge \min\{0, \overline{\Lambda}_t\}, \\ \text{resolve a neighbour}, & \text{otherwise}. \end{cases} \tag{7}$$

前两个条件中的每一个都是决策证书：它为当前区间内的所有邻居参数值固定参考决策。若两者均不成立，则解析一条未解析的记录$R^{(p_k(t))k}$，$k \in \mathrm{N}j$，并使用相同的旧记录、候选项和阈值，以更新后的界限重复该比较。每一次拒绝都保留完整的旧记录。例如，在只有一个邻居时，取 $C_t = -1/2$、 $c_{tk} = 1/4$ 和 $z_k \in [-1, 1]$。界限为 $[-3/4, -1/4]$,，因此 $\log U_t = -1$ 在该邻居被解析之前即可证明接受。图 1 说明了解析查询的两种效果：缓存历史不断增长，而新的锚点会收紧后续的参数范围。

3.2。查询执行与正确性

算法1按指定顺序将决策证书应用于 Q 中的请求。它采用递归求值和完整记录的共享缓存，其结构遵循局部吉布斯采样（Feng 等人，2025年；Liu 等人，2026年）。递归由显式栈实现。

算法1。使用决策证书执行查询。

输入：§2.1 中的参考规范、固定随机输入，以及按规定顺序排列的请求 Q 。输出：带索引的集合 Rϱ。 1 将共享记录缓存 C 初始化为 C $[j,0]=R_j^{(0)}$, $j\in V$，并将 Rϱ初始化为空； 2 过程 Resolve(j,τ):3 $t\leftarrow\max\big(\{0\}\cup\{s\in\{1,\dots,\tau\}:\ j_s=j\}\big)$; 4 如果 (j,t) 已缓存，则返回 C $[j,t]$;5 $R^{-}\leftarrow$Resolve($j,p_j(t)$);⌢6 从 R^{-} 和节点 t 处的固定输入中生成并保留候选记录 $R^{+}=(\vartheta,v,\mathrm{l}j(\vartheta,v))$，并对其求值一次；7 $R\leftarrow R^{-}$;⌢8 如果 $\mathrm{l}j(\vartheta,v)>0$ 则 9 重复执行 10 根据式 (5)，为 $R^{(p_k(t))_k}$, $k\in$ Nj 的参数分量构造区间，对缓存记录使用单元素集合；根据式 (6) 计算 $[\Lambda_t,\Lambda_t]$; 11 如果 $\log U_t<\min\{0,\Lambda_t\}$，则 $R\leftarrow R^{+}$；退出循环； 12 如果 $\log U_t\geq\min\{0,\Lambda_t\}$，则退出循环； 13 选择一个未解决的邻居 k，其参数区间宽度最大；若宽度相同，则选择节点索引最小者； 14 Resolve($k,p_k(t)$)；保留 R^{-},R^{+} 和 U_t，以供下一次比较； 15 C $[j,t]\leftarrow R$ ；将其参数分量用作节点 j 处新的范围锚点，并丢弃基于其先前锚点构造的范围； 16 返回 R； 17 对规定顺序中的每个 (j,τ) $\in Q$ 执行 18 将 Resolve(j,τ) 追加到 Rϱ，其索引为 (j,τ)；

我们针对§2.1中的有限参考构造，使用(5)中的两种提议构造之一，证明其正确性。所有计算均采用精确算术，并假设每个候选记录都能在有限时间内生成和评估。参考执行和查询执行使用相同的初始记录、更新计划和随机输入。

定理1。 在上述条件下，固定任意查询集 Q 和*§2.2*中规定的请求顺序。对于参考随机输入的每个实现，算法*1*都会终止，并返回$\mathcal{R}\varrho$ 中(3)所示的索引集合。每个缓存记录都与相应的完整参考记录一致，包括其参数、辅助变量和似然估计。在每次计划更新时，候选似然估计至多计算一次。

定理*1*的证明。 初始记录会被缓存。update t 发出的每次递归调用都会请求一条更早的记录。对于正的候选估计值，每次未决比较都会解析一条先前尚未解决的邻居记录。所有邻居都解析后，(6)中的区间就是一个单元素集合，而(7)会决定是否更新，包括恰好达到阈值的情况。在两种执行中，估计值为零时都会立即拒绝。对 t使用归纳法，再结合 Q的有限性，即可证明算法会终止。为证明正确性，对添加到缓存中的记录进行归纳，并从给定的初始记录开始。先解析自身的前驱记录，可确保每个节点缓存的更新构成其调度的一个前缀。

因此，未解决历史记录的范围从较早的已解析锚点开始。给定正确的缓存记录，旧记录和固定输入会生成与参考执行相同的完整候选项。对于正的估计值，(5)中的范围包含所需的参考参数，因此(6)涵盖参考对数比。由此，每个证书都会给出参考决策。接受时会存储完整候选项；拒绝时会存储完整旧记录。这保证了缓存正确性。由于节点的记录在其更新之间保持不变，请求时映射会为每个$(j,\ \tau)\in Q$返回$R_j^{(\tau)}$。候选项会在比较循环之前计算，并在递归调用期间保留；后续请求则使用缓存结果。因此，每个已调度的候选项最多计算一次。使用相同的ξ_k时，§2.2中的输出Y_k也一致，其规定的摘要也相同。□

正电子发射断层成像应用使用带有Potts先验的二元标签。一个未解析的历史标签的可行集合为$\{0,\ 1\}$，而一个已解析的标签给出一个单元素集合。补充材料§S4.1给出了由此产生的接受界限，并证明其与相应的完整参考执行一致。

Landmine应用在每个位点更新两个参数，其中只有对数长度尺度进入位点之间的交互。补充材料§S4.3推导了块接受界限，并证明其与完整参考执行一致。

4. 查询执行的工作量界

4.1。条件工作界

考虑算法1在前s个请求Q完成后立即执行的情形，其中$0\le s\le|Q|$。记C_s为该边界处的缓存，此时没有待处理的递归调用。令F收集用于在第T次更新中生成参数和辅助提议的原始输入。这些输入与接受阈值U_1、……、U_T独立。当某个阈值首次用于执行决策时，我们揭示该阈值的完整值。下一个要揭示的索引以及停止揭示阈值的决策，仅取决于F和已经揭示的值。令T_s按顺序记录已揭示的索引和值，并设$\mathsf{G}_s=\sigma(F,\ \mathsf{T}_s)$。这些信息确定了$\mathsf{C}_s$。在此边界处，每个具有已揭示阈值的更新都已缓存。候选估计值为零的更新无需揭示其阈值即可缓存。因此，未缓存更新的阈值仍未揭示；在给定G_s的条件下，它们在$(0,\,1]$上独立且服从均匀分布。补充材料§S1.2证明了这一自适应揭示顺序下的结论。根据定理1，对于未揭示阈值的每一种相容补全，所有缓存记录都与参考链一致。

对于每次未缓存更新t，为时刻$p_k(t)$的邻居参数构造规范范围。若所需记录位于C_s中，则使用单元素集合；否则，从不晚于$p_k(t)$的最新缓存锚点应用式（5）。利用参考旧记录和候选记录，当候选估计值为正时，这些范围给出规范的对数比界$[\Lambda_t^{\circ},\Lambda_t^{\circ}]$至（6）。选择一个有限、非负且关于$\mathsf{G}_s$可测的界$\omega_t$，使其对每个兼容的补全都适用于其宽度。相应的未决区间为$[\alpha_t,\alpha_t)$，其中$\alpha_t=\min\{1,\ \exp(\Lambda_t^{\circ})\}$，以及$\alpha_T=\min\{1,\ \exp(\Lambda_t^{\circ})\}$；对于估计值为零的情况，设$\overline{\alpha}_t=\alpha_t=0$。Grazzi & Zanella（2026年，补充材料，引理1）中的基本指数界给出

$$\overline{\alpha}_t-\underline{\alpha}_t\le p_t:=1-\exp(-\omega_t).\tag{8}$$

在参考调度中，将未缓存更新列为$\tau_1<\cdots<\tau_m$，其中m是它们的数量。对于$r=1,\ \ldots,\ m$，τ_r处的规范端点仅取决于G_s和$U_{\tau_1},\ \ldots,\ U_{\tau_{r-1}}$。将$V_{\tau_r}=(U_{\tau_r}-\alpha_{\tau_r})$设为模1。每次平移都在按此顺序使用$U_{\tau_r}$之前确定，因此连续的

条件化表明，给定 G_s 后，V_{τ_r} 是 $[0,1$ 上的独立均匀变量。因此，这些事件

$$B_t = \{V_t < p_t\}, \qquad \mathrm{pr}(B_t \mid \mathcal{G}_s) = p_t, \tag{9}$$

在未缓存更新 t 上具有条件独立性。从更新 t 到未解决的邻居进行递归调用需要 $U_t \in [\alpha_t, \alpha_t)$，因为执行范围细化了规范范围。根据（8），这意味着 B_t。每次调用后，执行都会保留 U_t，因此同一事件 B_t 涵盖了从更新 t 发出的所有邻居调用。

对于每个节点 j j T $V_s = j$ $j \neq$，令 E 表示截至时间 的未缓存更新集合，并设 $\{:\mathrm{E}\}$。称节点 $j \in V_s$ 在事件 $O_j = \bigcup_{t\in\mathcal{E}_j} B_t$ 上是开放的。根据（），这些事件具有条件独立性，其中

$$\pi_j := \mathrm{pr}(O_j \mid \mathcal{G}_s) = 1 - \exp\left(-\sum_{t\in\mathcal{E}_j} \omega_t\right). \tag{10}$$

令 $\mathsf{Q}_s \subseteq V_s$ 表示剩余请求中其记录不在 C_s 中的节点。记 I_j 为指示函数，表示在完成这些请求的过程中，节点 j 处至少有一条新记录得到解决。任何此类节点都可以从 Q_s 沿一条路径到达，且该路径在 j 之前的顶点均为开放的。令 $\Gamma_s(q, j)$ 包含从 q 到 j 的路径，且在由 V_s 诱导的图中不包含重复顶点。沿每条路径的条件独立性与并集界给出

$$\mathrm{pr}(I_j = 1 \mid \mathcal{G}_s) \le h_j := \min\left\{1, \sum_{q\in\mathsf{Q}_s}\sum_{\gamma\in\Gamma_s(q,j)}\prod_{v\in\gamma\setminus\{j\}} \pi_v\right\}. \tag{11}$$

查询节点处长度为零的路径权重为一；对空路径集合求和结果为零。乘积不包含终端节点，因为在进一步传播之前，解析其记录就会产生工作量。定理 2 将这些可达性界转换为工作量界，其中包括在未解决的邻居处生成范围。

定理 2. 在定理*1*的条件及前述构造下，令 W_s 表示从 $\mathcal{C}_s$ 开始，直至完成剩余请求及其统计输出所需的所有额外工作，采用*§2.2*中的加性成本模型。令 b_s、c_j、g_j 为有限、非负且 $\mathcal{G}_s$-可测的成本界，并且对所有剩余输入均一致有效。其中，b_s 涵盖额外的全局准备工作、查询开销和输出工作；c_j 涵盖在解析 j 处的新记录时计入的后端和管理工作；g_j 涵盖单独计费的提议输入和在 j 处的范围生成。每个节点成本界涵盖整个剩余块，包括重复操作，且每项操作只计费一次。令 J_j 表示是否发生单独计费的生成工作。令 $\mathcal{D}_s$ 为 V_s 上的一个 $\mathcal{G}_s$-可测有向关系，其中 (i, j) 记录从解析 i 到生成 j 的可能触发关系。假设 $J_j \le I_j + \sum_{i:\ (i,j)\in\mathcal{D}_s} I_i$，且工作量满足

$$W_s \le b_s + \sum_{j\in V_s} c_j I_j + \sum_{j\in V_s} g_j J_j. \tag{12}$$

令 $h_j^+ =$ 最小$\{1, h_j + \sum_{i:\ (i,\ j)\in\mathcal{D}_s} h_i\}$。那么，几乎必然，

$$\mathrm{E}(W_s \mid \mathcal{G}_s) \le b_s + \sum_{j\in V_s}\left(c_j h_j + g_j h_j^+\right). \tag{13}$$

首次进入未解析节点的条目构成一片以请求节点为根的森林。每条边都要求有一个开放父节点，从而得到路径界（11）。应用生成触发关系

条件并取条件期望即可证明该结果；参见补充材料第 S1.2 节。

4.2. 有界度图上的有限历史

现在，我们对独立提议输入下的条件工作量界取平均。固定整数$\Delta \geq 2$ 和 $L \geq 1$。考虑满足 $a_j = 1$、最大度不超过 Δ且边权重为$0 < w_{jk} \leq w < \infty$的式（1）。使用独立的标准高斯参数提议，并采用预先规定的调度，使每个节点至多更新 L 次。从初始记录缓存 C_0, 开始，其中 $|\theta_j^{(0)}| \leq B$，且 $B \geq 0$固定。图、初始记录、调度和有序查询集 Q 均在生成随机输入之前固定。令 F_j 包含节点 j处的原始参数及辅助提议输入，并补齐至 L 次更新。假设这些区块独立同分布，且独立于接受阈值。假设该图存在 D 种颜色，使得图距离至多为二的不同顶点颜色各异。始终可以取 $D = \Delta^2 + 1$。对于独立的标准高斯变量 $Y_1, \ldots, Y_L$，令 $A = 2(B + \text{最大值}_{1 \leq \ell \leq L} |Y_\ell|)$，并定义

$$q_D(w) = \left(\mathrm{E}\left[\left\{ 1 - \exp\left(-wLA_0 \sum_{i=1}^{\Delta} A_i \right) \right\}^D \right] \right)^{1/D},$$

其中 A_0、……、A_Δ 是 A 的独立副本。

推论 1。 在定理*1*的条件和前述构造下，设 $(\Delta - 1)q_D(w) < 1$。假设 $C(F_j)$ 界定了在节点 j 生成并评估至多 L 个候选记录所需的全部工作，但不包括已计入 W_{setup} 的工作。同一个非负函数 C适用于每个节点和图大小，并且 $c = \mathrm{E}\{C(F_j)\} < \infty$。使用单位成本的精确算术和索引访问，并对固定大小记录采用摊销单位成本的缓存操作。将图、数据、初始记录、随机输入和事件索引的准备成本计入W_{setup}。对于使用逐节点事件索引和缓存范围实现算法*1*的情况，从 C_0 解析 Q 的增量工作满足

$$\mathrm{E}\{W_{\mathrm{inc}}(Q)\} \leq \kappa |Q|,$$

其中 $\kappa < \infty$ 仅取决于 Δ、D、L、B、w、c 以及固定的操作成本。对于满足这些条件的图大小 n、图、调度和查询集，该界限是一致的。

该证明利用每个节点自身的输入块及其邻居的输入块，对其开启概率给出界限。距离二着色使得每种颜色内的这些界限相互独立。利用 H¨older 不等式和自避路径计数，可得压缩条件；去掉涉及端点的因子，则可控制其候选成本。论证详见补充材料第 S1.5 节。有限的矩形四邻域网格具有 $\Delta = 4$，并满足 $D = 5$：将颜色 $i + 2j$（模 5）赋予整数坐标为 (i, j) 的顶点。由于 $q_D(w) \to 0 \,\text{as}\, w \downarrow 0$，当耦合强度为充分小的正值时，该条件成立。一个可直接检验的充分条件是

$$4wL\Delta(\Delta - 1)\left\{ B + (Lm_D)^{1/D} \right\}^2 < 1,$$

其中 $m_D = \mathrm{E}(|Z|^D)$，$Z \sim N$取值范围为 $(0, 1)$。这由 $1 - \exp(-x) \leq x$（针对 $x \geq 0$）以及最大绝对提议值的 D阶矩上界推出。直接评估 $q_D(w)$ 会保留指数项，并可证明更宽的耦合范围。对于固定次数的系统扫描、固定的 B，w，以及满足该推论的候选生成和评估成本的共同上界，随着网格边长增大，同一 κ 仍然成立。

4.3。路径上的块执行

我们将算法1专用于路径 $1,\dots,n$，其中 $a_j = 1$，公共边权重为 $0 < w < \infty$，给定的初始记录满足 $\theta_j^{(0)} = 0$。参考调度依次访问 $n, n-1, \dots, 1$；在每个节点执行$L \ge 1$次连续更新；查询为 $Q = \{(1, nL)\}$。设 $Y_{j\ell}$, $j = 1,\dots, n$ 和 l =1,…, L 为相互独立的标准高斯参数候选值。可寻址源提供相互独立的原始提议输入和阈值，且这些输入和阈值在各节点具有相同的分布。地址指定节点、局部更新和输入类型；重复访问会返回相同的输入。输入块在首次需要时缓存，候选项则在其更新得到解决时进行评估。在块 j 期间，左邻居 $j-1$（若存在）在这些更新的参考时刻具有初始参数值零。右邻居 $j+1$（若存在）具有其块末值。在该块得到解决之前，其参数被包含在由 $0, Y_{j+1,1}, 1, \dots, Y_{j+1,L}$ 张成的区间内。未决比较会触发对块 $j+1$ 的求解，然后使用相同的旧记录、候选项和阈值继续。

完整的块处理过程见补充材料中的算法 S1。

令 F，F' 为节点原始提议输入块的独立副本。对于 F 的参数坐标 Y_1、.....、Y_L，设 $Y_0 = 0$ 并定义

$$r(F) = \max_{0 \le \ell \le L} Y_\ell - \min_{0 \le \ell \le L} Y_\ell, \qquad s(F) = \sum_{\ell=1}^{L} \max_{0 \le h < \ell} |Y_\ell - Y_h|.$$

这里，$r(F)$ 是该块的范围宽度，而 $s(F)$ 界定了其候选项与可能的旧参数之间距离之和。对于块 $j < n$ 的第 l 次更新，选择 (8) 中的 ω_t 为 $wr(F_{j+1}) \max_{0 \le h < \ell} |Y_{j\ell} - Y_{jh}|$，其中 $t = (n-j)\,L + l$ 且 $Y_{j0} = 0$。这些界限相加为 $ws(F_j)r(F_{j+1})$；在终端节点处，取 $\omega_t = 0$。在给定所有原始输入块的条件下，(10) 将每个块 $j < n$ 的规范开启概率界定为 $p(F_j,\ F_{j+1})$，其中

$$p(F, F') = 1 - \exp\{-ws(F)r(F')\}, \qquad \rho = \left[\mathrm{E}\{p(F, F')^2\}\right]^{1/2}.$$

对于每个固定的有限 w 和 L，这些范围几乎必然是有限的，因此 $0 \le \rho < 1$。相邻因子共享一个输入块；ρ 中的二阶矩控制了这种依赖关系。

定理 3。考虑满足上述路径规范和定理*1*条件的算法 *S1*。使用精确实数运算、对可寻址输入源的单位成本访问以及固定大小记录。补充记账模型为每次原始输入抽取、范围访问、块查找、阈值读取、候选值调度、证书比较、记录提交以及列出的控制操作各分配一个逻辑管理单元。令 $C(F_j)$ 统一界定在块 j 得到解析时构造并评估 L 候选项的其余全部工作量，其中对其可能的旧记录取一致上界。假设同一个非负函数 C 适用于每个节点，且$c = \mathrm{E}\{C(F)\} < \infty$，并定义

$$\delta = \mathrm{E}\{C(F')p(F, F')\}, \qquad \varepsilon = \mathrm{E}\{C(F')p(F, F')^2\}.$$

令 N 表示已解析块的数量。那么，对于每个整数 $h \ge 0$，

$$\mathrm{pr}(N \ge h+1) \le \rho^h, \qquad \mathrm{E}(N) \le \frac{1}{1-\rho}, \tag{14}$$

且增量工作量满足

$$\mathrm{E}\{W_{\mathrm{inc}}(Q)\} \le c + \delta + \frac{\rho(c\varepsilon)^{1/2}}{1-\rho} + 24L\left(1 + \frac{1}{1-\rho}\right). \tag{15}$$

对于固定的 w、L、输入分布、公共包络 C 和操作成本，此界对 n 一致。准备工作和输出工作在 (4) 中仍保留各自的项。

证明按奇数边和偶数边对传播因子进行分组，每组内的输入彼此独立。加权柯西–施瓦茨不等式保留了最终边与其候选成本之间的依赖关系。最多生成 $N+1$ 个输入块，其中包括任何未解析的边界区块；管理最多需要 $24L(N+1)$ 个逻辑单位。完整论证和运算次数见补充材料 §S1.3。

4.4。与逐点祖先评估的比较

我们将证书执行与逐点祖先评估器 B_{point} 进行比较。两种评估器都已知模型、图、调度和有序查询集 Q。它们从相同的准备好的初始记录缓存 C_0 开始，并使用相同的提议输入和接受阈值，同时遵循相同的输入访问和成本约定。对于每个请求的未缓存更新 t，B_{point} 首先解析其自身的前驱，然后生成并评估一次候选项。它会在各请求之间共享完整记录缓存，并在递归调用期间保留候选项和阈值。似然估计为零时立即拒绝。当提议参数等于旧参数时，决策使用两条局部记录和 U_t。对于节点 $j=j_t$ 上任何其他具有正似然估计的候选项，它首先解析所有邻居记录 $R^{(p_{k(t)})k}$、$k\in\mathrm{N}j$，重用缓存记录，并按固定顺序处理未缓存的邻居。随后，它计算式（2），并将接受概率与 U_t 比较。接受的候选项和保留的旧记录按照算法1进行缓存。以下推论比较同一索引集合 $\mathrm{R}Q$ 下两种执行规则的增量工作量。

推论 2. 假设候选似然估计在每次计划更新时几乎必然严格为正。考虑以下任一类情形，并采用所引结果中的输入和成本假设。对于定理 *3* 的路径，使用算法 *S1* 和 $Q=\{(1,nL)\}$。对于满足推论 *1* 的矩形四邻域网格，将 n 顶点从左上到右下逐行编号，按1、……、n的顺序执行 L 次扫描，并使用算法*1*及$Q=\{(n,nL)\}$。使用索引记录访问实现 B_{point}，并根据局部因子计算接受比率。令 N^{point} 调用表示其候选似然评估次数，并令 W^{point} 增量(Q) 和 W^{cert} 增量(Q) 分别表示逐点评估器和证书评估器的增量工作量。在每一类情形中，当 n 变化时，固定 L、输入分布、公共包络 C 以及相应工作量界中的所有其他参数。于是

$$\begin{aligned} N_{\mathrm{call}}^{\mathrm{point}} &= nL \quad \textit{almost surely},\\ \mathrm{E}\{W_{\mathrm{inc}}^{\mathrm{point}}(Q)\} &= \Theta(nL), \qquad \sup_{n}\mathrm{E}\{W_{\mathrm{inc}}^{\mathrm{cert}}(Q)\}<\infty. \end{aligned} \tag{16}$$

两个评估器都返回参考集合 $\mathcal{R}Q$。

对于路径，逐点规则会沿连续的右侧块传播。对于网格，最终查询会通过西侧和北侧邻居到达最后一次扫描的更新。解析自身的前驱后，便会纳入所有 nL 候选项。共同的成本包络和证书界限给出式 (16)；补充材料 §S1.4 提供了完整证明。

补充材料中的表 S1 总结了条件和工作量界。

5. 数值实验

5.1. 块解析与计算工作量

这些实验考察历史解析、执行策略和查询工作负载如何影响计算工作量。

我们从 §4.3 中的路径模型和块调度开始，其中 $a_j = 1$、公共边权重w，并采用独立的标准高斯参数提议。在每个节点，设

$$\ell_j(\theta) = \exp(-\theta^2/2), \qquad \widehat{\ell}_j(\theta, u) = 2\,\mathbf{1}\{u < 1/2\}\exp(-\theta^2/2),$$

其中 u 服从 (0,1) 上的均匀分布。这会得到一个无偏估计，其取零的概率为 1/2。所有初始记录均为 $R_j^{(0)} = (0, 1/4, 2)$。我们使用相同的参数候选值、辅助输入和接受阈值，将算法 S1 与完整执行进行比较，取值为 $Q = \{(1, nL)\}$。我们取 $n \in \{4, 16, 64, 256\}$、 $L \in \{1, 2, 4\}$ 和 $w \in \{0.01, 0.2, 1.5\}$。对于所有配置，我们使用相同的四个固定输入流。公共节点和更新地址处的输入完全相同。图 2(a) 比较了 L 的三个取值下， $w = 0.2$ 处的候选数量。补充材料 §S2.2 报告了三种边权重下的结果。

算法 S1 为 N 已解析块评估 LN 候选项；完整执行评估 nL。在 144 个固定配置中， N 的范围为 1 到 4。对于每个固定的 L、 w 和输入流，N 在四种路径长度下保持不变。返回的完整记录在每种配置下都与完整执行一致。

管理工作量范围为 12 到 176 个逻辑单位，处于定理 3 的界限 $24L(N + 1)$ 内。该计数包括对任何未解析的边界区块进行的输入和范围生成。似然评估、准备和统计输出成本在 (4) 中保留各自的项。

在第二个实验中，我们使用平路径，其中 $u \sim N(0, 1)$。我们固定 $L = 4$ 和 $w = 0.2$，并针对具有 $n \in \{32, 128, 512, 2048\}$ 的路径，对同一查询使用 100 个公共输入流。

在每个测试的路径长度下，证书执行平均解析 1.43 个区块，最多解析 6 个区块。其平均候选数量为 5.72，而指定的逐点评估器和完整执行对应的数量为 $4n$。在 400 次匹配比较中，三个执行器每次都返回相同的完整查询记录。图 2(b) 展示了随着 n 增大，候选评估所获得的改进。(c) 面板展示了总经过时间的改进，其中包括准备全部 n 条初始记录的时间。在 $n = 2048$ 下，逐点执行和完整执行的平均时间分别是平均证书时间的 160.7 倍和 155.7 倍。在此实现中，准备工作在 $n = 32$ 时占平均证书时间的 5.2%，在 $n = 2048$ 时占 73.3%。补充材料 §S2.2 给出了运算次数和时间明细。接下来，我们将使用粒子似然估计器考察这些成本。

5.2。相同查询的执行成本

我们比较两种模型的缓存逐点执行和证书执行：具有泊松观测的高斯潜在过程，以及具有切换活动和二项捕获的基因表达过程。两者均在含有128个节点的路径图上使用式（1）中的高斯图先验， $a_j = 1$和 $w_{jk} = 0$。8。粒子滤波器用于提供似然估计（Andrieu 等人，2010）。每个模型的两个模拟数据集产生六种数据、粒子估计器和参数提议的配置。对于基因表达模型，两种粒子估计器还使用不同的参数提议。完整的模型和计算设置见补充材料 §S3。每次运行均返回每个参考时间点节点1和节点2的参数历史。每种配置进行五次配对运行，每对运行使用相同的初始完整记录、更新计划、提议输入和接受阈值。§4.4中的逐点评估器和算法1都会缓存完整记录并重用已评估候选项。逐点执行和证书执行使用

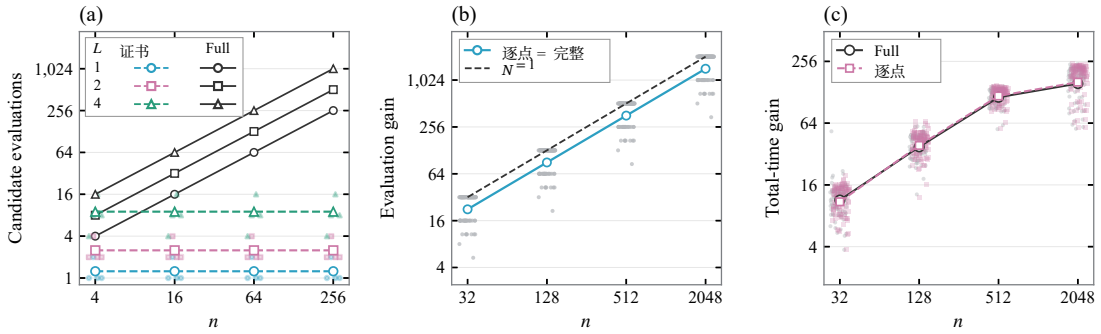


图 2. 路径上局部查询的工作量和时间收益。该查询请求在参考时间 nL 时，路径上节点 1 的完整记录；路径包含 n 个节点，每个节点有 L 次更新。面板 (a) 使用可能为零的估计值 $w = 0$。2, $L \in 1, 2, 4$ $w = 0\,2$ $L = 4\{\}$以及相同的四个固定输入流。面板 (b) 和 (c) 使用严格正估计值 。以及 100 个公共输入流。面板 (a) 显示候选数量；(b) 显示评估收益；(c) 显示总时间收益，包括准备时间。收益为基线成本除以证书成本。(b) 中逐点和完整候选数量相同。小标记表示各个输入流的数据点，并在水平方向错开，以区分重叠点。在 (a) 中，空心标记以线连接，表示平均计数。在 (b) 和 (c) 中，它们表示平均成本比。(b) 中的虚线表示收益为 n，此时仅解析一个块 ($N = 1$)。两个坐标轴均采用对数刻度。

替代文本：面板 (a) 显示，随着路径增长，证书评估次数保持稳定，而完整执行次数增加。面板 (b) 显示，相较于逐点执行和完整执行，评估收益不断增加。面板 (c) 显示，总时间收益增加得更为缓慢，并且在不同输入流之间存在差异。

四个工作进程进行初始化，并在控制进程中评估后续候选项。经过时间包括输入准备、初始化、随机输入和索引的构建、范围维护、查询评估、统计汇总，以及进程的启动和关闭。

返回的参数历史记录在全部 30 次配对运行中均一致。图 3 显示了相对于逐点执行的证书执行的候选项评估次数和总经过时间。在六种配置中，证书执行所需的平均候选项评估次数为逐点执行所需平均数量的 16.6–35.6%。每个执行器都会针对每次已解析的历史更新评估一个候选项。其平均经过时间为逐点时间的 19.7–47.7%，并且在每一对运行中都更快。在每种配置中，平均时间比均超过平均候选项数量比。绝对计时及其离散程度见补充材料 §S3.2。

我们使用 128 条保存的轨迹来研究缓存范围的作用（补充材料 §S3）。执行器在一些所需邻居记录仍未解析时，确定了 163,164 个决策。在这些比较中的每一次，我们都将旧记录、候选记录和接受阈值固定不变，并使用从初始状态传播而来的范围重新计算界限。替换范围使其中 101,784 个决策（62.4%）仍未确定。接下来，我们改变查询节点的集合，以研究其对计算工作量的影响。

5.3. 查询范围与执行成本

我们改变 P1 和 P2 中查询节点的数量，并比较证书、逐点和 Picard 执行。Picard 执行器将 exactOnlinePicard 算法（Grazzi&Zanella，2026年）扩展到完整的伪边际状态。它在滑动窗口内并行评估候选项，并推进已解析前缀。我们的实现使用四个工作进程、包含 16 次更新的窗口，以及已评估候选项的缓存。模型设置和计时约定与 §5.2 中相同。

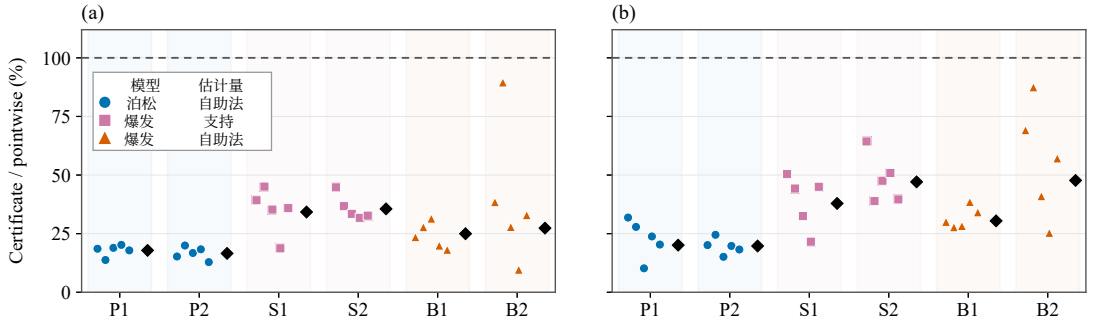


图 3。同一组查询的候选评估和经过时间。两个执行器都在路径上的每个参考时间返回节点 1 和 2 的参数历史记录，该路径包含 128 个节点。面板 (a) 和 (b) 分别显示候选似然评估和总经过时间。彩色点表示每种配置下五次配对运行中，证书成本占逐点成本的百分比。每组右侧的黑色菱形表示平均成本比，以百分比表示。虚线表示 100% 处的成本相等。P1 和 P2 使用泊松观测模型中的两个数据集。对于爆发式基因表达模型，S1 和 B1 共享一个数据集，S2 和 B2 共享另一个数据集。支持 筛选器根据观测值设定条件；S1 和 S2 使用独立于当前值的参数提议。所有自助法配置都使用随机游走提议。在每种配置中，五次配对运行从左到右的顺序在两个面板中相同。

替代文本：面板 (a) 显示，在六种配置的每组配对运行中，证书执行的候选评估次数都更少。面板 (b) 显示，每组配对运行的经过时间都更短。黑色菱形表示，平均时间的降幅百分比小于平均评估次数的降幅百分比。

每个任务在每个参考时间请求节点 1 至 m 的参数历史记录，查询范围为$m \in \{2、8、32、64、128\}$。我们为每个数据集使用五个随机输入流，并在全部五个查询范围中使用相同的输入流。在每个输入流中，三个执行器共享初始完整记录、更新计划和固定随机输入。每次执行都从全新缓存开始，并承担初始化开销。在数据集、输入流和查询范围的全部 50 种组合中，各执行器返回的参数历史记录均一致。

图 4 展示了历史工作量和相对于 Picard 的经过时间如何随查询范围变化。对于双节点查询，逐点执行在每个数据集中平均解析调度的约 67%，而证书执行仅解析 11–12%。当请求全部 128 个节点历史记录时，两个执行器都会解析完整调度。Picard 在每个查询范围下都会解析完整调度。

对于 2、8 和 32 个查询节点，证书执行在每次匹配运行中都快于 Picard。在 64 个节点时，证书执行在每个数据集中平均解析调度的约 61%。此时，其平均经过时间在两个数据集中都接近 Picard。对于完整查询，P1 和 P2 中证书执行的平均时间分别是相应 Picard 平均时间的 1.497 倍和 1.512 倍。

在 §5.2 的全部六种配置中，对于双节点查询，证书执行的平均耗时均少于 Picard（每种配置进行五次配对运行）。对于完整查询，在每种配置仅进行一次配对运行时，Picard 更快。两种执行器在全部 36 对运行中返回了完全相同的参数历史。

在 P1 和 P2 中，将 Picard 窗口从 16 次更新增大到 64 次更新后，在双节点查询和完整查询中，经过时间减少了 16.7–24.5%（四个工作线程，每种设置运行一次）。补充材料 §S3.2 给出了各项具体计时结果。

我们 xt 检查应用中的统计查询的历史工作量和经过时间 .
ne

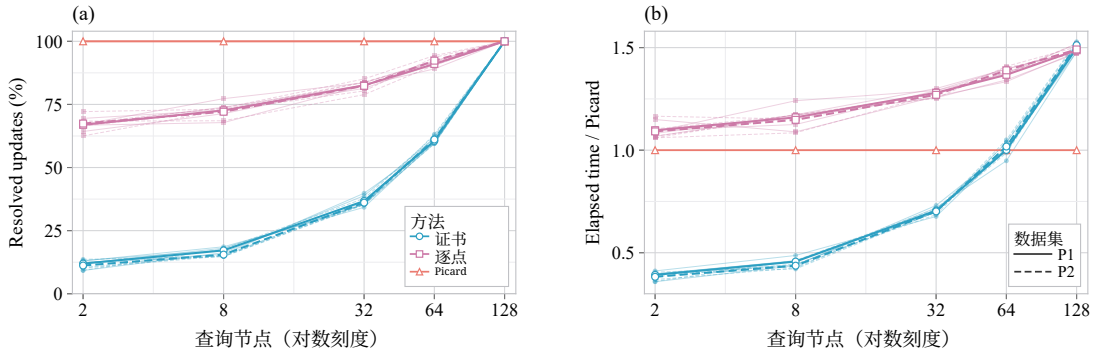


图4。查询范围、已解析历史和执行成本。每项任务请求节点 1 至 m 在 P1 或 P2 中每个参考时间的参数历史。(a) 面板显示已解析的计划更新百分比；(b) 面板显示相对于 Picard 的总经过时间。每个数据集的五个输入流由各方法和查询范围共享。细线和小点表示单个证书运行与逐点运行。较粗的线和空心标记在 (a) 中表示平均百分比，在 (b) 中表示平均时间比。方法和数据集图例适用于两个面板。Picard 曲线在 (a) 中均重合于 100%，在 (b) 中均重合于 1。横轴采用对数尺度；纵轴为线性尺度。

替代文本：面板 (a) 显示，随着查询范围扩大，证书执行和逐点执行解析的历史记录比例不断增加，最终覆盖完整历史记录。面板 (b) 显示，在两个数据集中，对于较小的查询，证书执行速度比 Picard 快；对于完整查询，速度则较慢。

6. 应用

6.1。正电子发射断层成像中的区域推断

我们使用来自 OpenNeuro 数据集 ds004869 的一次基线[^{11}C]MC1 扫描，研究正电子发射断层成像（PET）中的区域定量（Yan 等人，2025 年、2026年）。分析使用覆盖最初 90 分钟的 27 帧，以及所提供的血浆输入和全血曲线。在每个体素处，总分布容积 V_T 是在恒定血浆输入下，组织示踪剂浓度与血浆示踪剂浓度的平衡比率（Zhou 等人，2013 年）。我们的查询是预定义左侧岛叶区域内 1311 个体素的 V_T 的平均值。该区域位于左半球一个包含 41500 个体素的分析域内。我们考察关于组织区室数量的不确定性，以及模型标签之间的空间依赖性如何影响该区域平均值的推断。

我们使用单组织室和双组织室模型，并假设观测误差服从高斯分布，条件是给定的输入曲线和观测权重。动力学参数具有定义在有界区间上的独立对数均匀先验，噪声精度则具有伽马先验。二元模型标签的基础概率相等，并采用Potts先验，倾向于使面相邻体素的标签一致，耦合参数为0.25。推断使用模型标签上的马尔可夫链，并对动力学参数和噪声精度进行积分消去。对每个精度进行解析积分，并对动力学参数进行确定性正求积，从而得到模型证据。在每次更新时，参考链均匀地选择一个体素，并以1/2的概率提出任一标签。证书执行会对未解决邻居的Potts贡献进行界定，并且仅在接受检验需要时才解析其历史标签。

对于每个体素和模型，我们将 V_T 求积质量归入固定的对数分箱，并将每个分箱的质量置于其几何中点处。这些表格与空间标签后验分布共同定义了下文使用的数值后验分布。在每个规定的查询时间，我们解析区域标签向量，分别从相应的条件 V_T 表格中独立抽样，并对这些体素值求均值。所得的区域抽样值纳入了标签之间的空间依赖性以及 V_T 中的条件变异。我们通过对条件区域

均值求平均来估计后验均值，并根据区域抽样值的分位数估计区间端点。补充材料 §S4.1 给出了模型细节和数值规则。其中还证明，所请求的标签和区域抽样值与使用相同初始标签和随机输入进行完整执行的结果一致。

运行 4 条链，每条链进行 512 次迭代，并舍弃每条链的前 64 次迭代，得到空间后验均值 2.05372 mL cm^{-3}，以及 $[2.04246, 2.06468]$ 的 95% 等尾区间。将所有体素固定为双室模型，得到的均值为 2.05449；而将空间耦合设为零并对模型标签取平均，得到的均值为 2.05323。这些均值之间的最大差异为该空间后验区间宽度的 5.7%。在空间模型下，给定标签时 V_T 的变异占区域平均值估计后验方差的 99.7%。区域后验均值的蒙特卡罗标准误估计为 1.23×10^{-5} mL cm^{-3}。固定模型证据和抽样所得的标签后，使用未分箱的条件矩会使均值估计降低约 1.5% 的该标准误。补充材料 §S4.4 给出了基于秩的链诊断和区间端点误差估计（Vehtari 等人，2021）。§S4.1 中的图 S1 比较了针对 12 个预先指定的区域体素的条件 V_T 分布和模型证据所用的两种求积规则。

我们在第一条链上比较这两个执行器，使用相同的初始状态、随机输入和 2048 个指定的区域查询。两者都从相同的缓存启动，其中包含 2942 个体素–模型证据值和区域条件分布，并使用相同的积分程序。证书执行在 23.29 分钟内完成所有查询。完整执行在运行 45.03 分钟后达到预算上限并停止，期间完成了 725 次更新；首次查询安排在第 10,375 次更新之后。证书执行为区域查询构建了 177 个新的双区室证据值。完整执行则在首次区域查询之前，沿图调度推进并构建了 353 个证据值。图 5(a) 展示了区域后验分布，以及区域 V_T 如何随模型组成而变化。每次运行中，已完成的双区室构建耗时均占观测时间的 90% 以上（面板 b）。证书执行和完整执行的平均构建时间分别为 7.4 秒和 7.3 秒（面板 c）。经过时间从进程启动到退出计算，共享准备成本则在补充材料 §S4.2 中单独报告。

6.2。比较*Landmine*数据中的协方差尺度

我们比较Landmine雷达数据（Xue等，2007年）中两个站点的协方差长度尺度。每个观测值包含九个特征和一个指示地雷或杂波的二元标签。我们使用站点1–15中每个站点的200个观测值，这些站点代表植被茂密的区域。特征使用合并后的训练观测值进行标准化。在站点 j，我们采用Probit高斯过程模型，其平方指数协方差分量的长度尺度为 $\exp(z_j)$，信号标准差为 $\exp(\gamma_j)$。我们指定条件独立先验 $z_j \mid \mu \sim N(\mu,\ 0。75^2)$，其中 $\mu \sim N(\log 3,\ 0。75^2)$，$\mu$为共享均值。对数振幅$\gamma_j$具有独立标准正态先验。对 μ 进行积分消去后，可在完全图上为中心化对数长度尺度z_j − log 3给出高斯先验。我们查询联合状态（$z_1,\ z_2$），并通过 $d = z_1 - z_2$ 的均值、95%等尾区间以及其为正的概率来概括后验分布。d 为正表示标准化特征空间中站点1的协方差长度尺度更长。

在每个位点，期望传播都会为潜变量值提供一个高斯重要性分布。按照 Filippone & Girolami (2014) 的方法，我们对 32 次独立抽样的完整重要性权重取平均，从而得到一个非负且无偏的边际似然估计量。每次参数更新都会从一个与当前参数值无关的固定混合分布中联合提议 $(z_j,\ \gamma_j)$。只有对数长度尺度会出现在耦合各站点的先验项中。因此，证书执行会界定相邻对数长度尺度的范围，并计算振幅

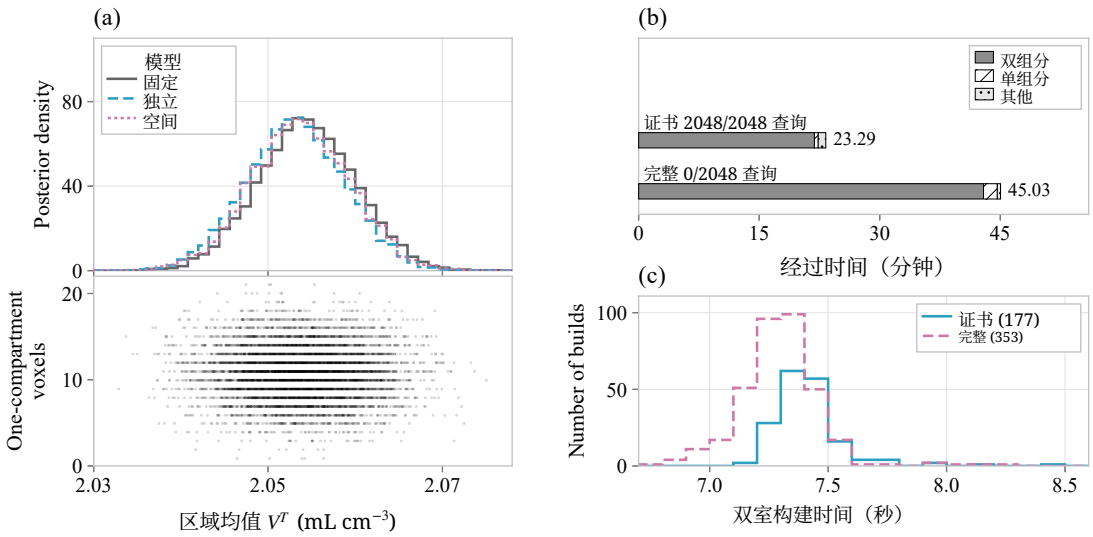


图 5. 区域 PET 推断及其计算成本。PET 表示正电子发射断层成像。(a) 上：固定双室建模、独立模型平均和空间模型平均下区域均值 V_T 的后验密度直方图，分别使用 8192、8192 和 7168 次抽样。下：7168 次区域均值 V_T 的空间联合抽样，以及分配给单室模型的体素数。(b) 观测到的经过时间，按已完成的双室证据构建、单室证据构建及其余经过时间（其他）划分。两个执行器均从同一缓存开始，并以第一条链的 2048 个区域查询为目标。证书执行完成 2048/2048 个查询；完整执行在达到时间预算前完成 0/2048 个查询。共享缓存和积分基底的构建不计在内。(c) 证书执行和完整执行分别完成的 177 次和 353 次双室证据构建的耗时直方图。构建时间包括数值积分和数据库写入。每项比较中的直方图使用相同的分箱。

替代文本：面板 (a) 展示了三种模型下相似的区域后验分布。下方的联合抽样将区域均值与分配至一个区室的体素数量联系起来。面板 (b) 展示了证书执行完成所有查询，而完整执行在未完成任何查询的情况下达到其时间预算。面板 (c) 展示了重叠的构建时间分布，其中证书执行下构建的证据更少。

先验，以及旧参数对和提议参数对处的完整混合密度。若拒绝提议，则保留完整的旧参数和辅助记录。补充材料 §S4.3 给出了中心化先验、接受界限及正确性论证。

作为比较，我们采用辅助变量增广（AA），在白化高斯变量的条件下进行协方差更新（Murray & Adams, 2010）。潜在过程及其高斯截距通过椭圆切片采样分别更新（Murray 等人，2010）。在每个位点，先进行十次此类成对更新，随后对 (z_j, γ_j) 进行一次联合 Metropolis–Hastings 更新。两个采样器均以协方差参数的同一边际后验为目标，我们分别运行了四条链。预处理、协方差细节和抽样日程详见补充材料 §S4.3；诊断遵循该补充材料 §S4.4。

表 1 比较了共享模型下 d 均值的伪边际（PM）和辅助变量增广（AA）估计值、其两个区间端点以及 $\mathrm{pr}(d > 0)$。每个差值的绝对值都小于 1.5 倍合并后的蒙特卡洛标准误估计值。图 6(a) 比较了 z_1、z_2 和 d 的伪边际和 AA 分位数，并展示了它们的带符号差值。

我们还使用这些站点相同的观测值、似然和各自的边际先验，分别对站点 1 和 2 进行拟合。两个模型都认为站点 1 的长度尺度比站点 2 更长的后验概率约为 60%。根据 d 伪边际抽样，d 的 95% 后验区间为 16.7%

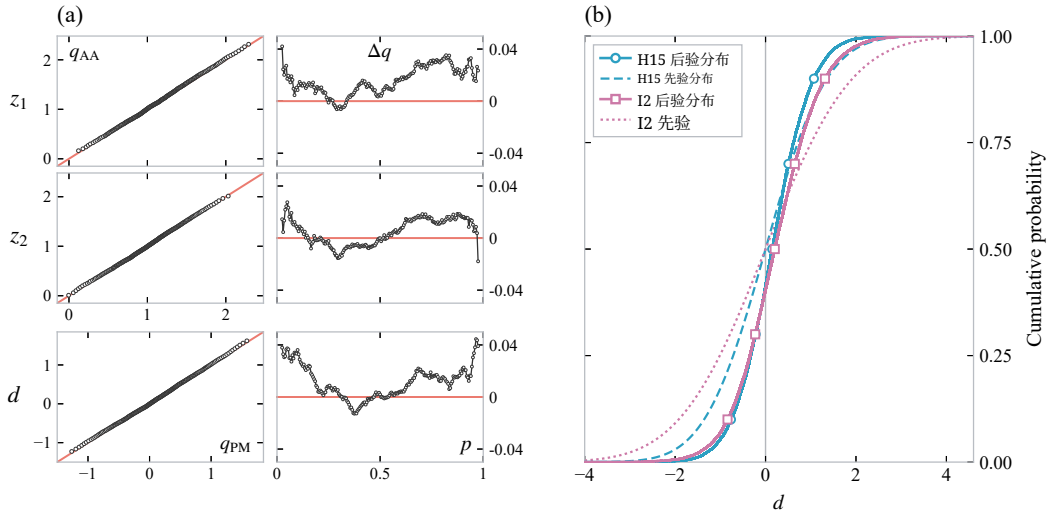


图 6. Landmine 数据的后验分布比较。(a) 共享 15 位点模型（H15）下伪边际（PM）样本与辅助变量增广（AA）样本的比较。从上到下各行分别对应 z_1、z_2 和 $d = z_1 - z_2$。左列绘制概率为 p 时的 AA 分位数 $q_{AA}(p)$ 与 PM 分位数 $q_{PM}(p)$ 的关系。右列绘制 $\Delta q = q_{AA}(p) - q_{PM}(p)$ 与 p 的关系。两列均使用 p 从 0.025 到 0.975 之间等间距取值的 191 个点。左列中的红线表示相等，右列中的红线表示差值为零。(b) H15（蓝色）和独立双位点模型（I2，粉色）的 d 先验与后验累积分布函数。实线表示两个模型基于 PM 样本得到的经验后验分布。虚线和点线分别表示 H15 和 I2 的先验分布。圆圈和方块标出概率水平为 0.1、0.3、0.5、0.7 和 0.9 时 H15 和 I2 的后验分位数。灰色竖线标出 $d = 0$。H15 使用 24,576 个 PM 抽样和 294,912 个 AA 抽样；I2 使用 24,576 个 PM 抽样。所有样本均为预热后四条链的合并结果。

替代文本：面板 (a) 显示，对于两个对数长度尺度及其差值，伪边际分位数与辅助变量增广分位数高度一致。相邻图表显示，在不同概率下存在细微偏差。面板 (b) 比较了先验和后验分布。与独立模型相比，共享模型对差值的后验分布更加集中。两个后验分布在零的两侧仍保留了相当大的概率。

表 1. *Landmine* 位点对比的后验汇总

模型	采样器	Mean	95% 区间	pr($d > 0$)
共享（H15）	PM	0.152 (0.009)	[−1.265, 1.581]	0.587 (0.005)
	AA	0.165 (0.015)	[−1.227, 1.620]	0.584 (0.008)
独立（I2）	PM	0.223 (0.010)	[−1.441, 1.973]	0.596 (0.006)

对比量 $d = z_1 - z_2$ 比较了位点 1 和 2 的对数长度尺度。H15 表示共享 15 点模型，I2 表示独立双位点模型。PM 表示伪边际；AA 表示辅助变量增广。括号中包含估计的蒙特卡罗标准误。区间为等尾后验区间。区间端点误差和链诊断见补充材料 §S4.4。替代文本：在共享模型中，两个采样器均给出位点对比为正的后验概率，接近 0.59。在两个模型中，95% 区间均跨越零。

在共享模型下更窄。共享模型使用来自 13 个额外位点的数据，并通过共同均值关联两个对数长度尺度。这种关联将 d 的先验方差从 2.25 减半至 1.125。图6(b)展示了两种模型下的先验累积分布函数和后验累积分布函数。

我们在每 15 次随机扫描更新后联合查询这两个位点，在每条链中共查询 8192 次。在四条链中，证书执行评估了预定 491520 个候选项中的 491306 个，留下 214 个（0.044%）未评估。当至少有一个所需的邻居记录尚未解析时，证书仍能确定 90.9% 的已评估更新。每次更新都需要其自身紧邻的前一个 PM记录，因此每个位点的已解析历史构成其更新计划的前缀。积分共享先验还将每个位点与其他 14 个位点关联起来。后续递归或查询可能需要较早证书跳过的记录，从而延长这些前缀（补充材料，图 S2）。计算工作量取决于它们在整个查询序列中累积的范围。

7. 讨论

决策证书允许未解析的历史参数通过范围参与接受计算。在完整的旧记录和候选记录可用时，这些范围可以针对固定阈值确定决策。定理1证明，这种构造保留了参考过程所查询的完整记录。在推论 2 的路径与网格族中，证书执行与指定的逐点祖先评估器具有不同的预期增量工作量级。这里，图会随着每个节点的历史长度增长，而工作量界中的所有其他参数均保持固定。对于相同的查询、参考调度和随机输入，这种差异依然成立，表明未解决的邻居历史的表示方式会改变恢复指定输出所需的工作量。

Anand和Jerrum（2022）从可能无限的自旋系统的吉布斯测度中，对有限窗口上的配置进行采样。在指数强空间混合和次指数图增长条件下，其预期运行时间与窗口大小呈线性关系。对于有限参考轨迹，我们的图与路径界限允许随机候选成本取决于控制传播的提议输入。这些界限假设所有节点和图大小均具有公共可积成本包络。

证书执行会在一系列查询之间复用缓存记录。后续请求也可能需要先前证书绕过的历史记录，例如 Landmine 中的情况。准备成本和并行评估也会影响局部执行带来的时间收益。对于较长的路径，在路径实验（§5.1）中，准备工作占证书时间的比例更大。更广泛的查询使泊松实验（§5.3）中的计时优势转向 Picard。相同请求输出的总经过时间综合体现了这些影响。

正确性结果假设使用精确算术。Landmine 的实现采用向外舍入的界限；如果完全解析后的比较仍无法判定，就会停止（补充材料，§S4.3）。

定理 2 中的条件界为分析节点历史记录增长时的累积工作量提供了起点。更精细的分析应将缓存范围的变化与后续请求所需的工作量联系起来。此类分析必须追踪各请求之间已揭示的阈值和保留的辅助记录，并且每项操作只计费一次。另一个互补方向是将算法 S1 的可寻址输入扩展到更广泛的预定调度类别。此扩展要求能够访问历史前驱和有效范围，而无需生成完整的随机输入序列。将这种访问能力与由显式请求驱动的输出例程结合起来，就能围绕统计任务所需的记录来组织输入生成、索引和输出工作。

资金支持

本研究未获得任何资金支持。

利益冲突

作者声明不存在利益冲突。

数据可用性

断层成像数据可从 OpenNeuro 获取，数据集为 ds004869，版本为版本1.4.1（Yan 等人，2026年）。Landmine 数据由 Xue 等人（2007 年）描述。本文使用的公开镜像详见补充材料 §S4.3。复现包的预留数字对象标识符为 https://doi.org/10.5281/zenodo.24506399。

补充材料

补充材料包含证明、分块执行流程和运算次数、模型与计算规格、个体计时记录以及应用诊断。

参考文献

Anand, K. & Jerrum, M. (2022). 通过强空间混合实现无限自旋系统的完美采样。*SIAM J. Comput.* **51**, 1280–1295. Andrieu, C., Doucet, A. & Holenstein, R. (2010). 粒子马尔可夫链蒙特卡罗方法。*J. R. Statist. Soc. B* **72**, 269–342. Andrieu, C. & Roberts, G. O. (2009). 高效蒙特卡罗计算的伪边际方法。*Ann. Statist.* **37**, 697–725. Angelino, E., Kohler, E., Waterland, A., Seltzer, M. & Adams, R. P. (2014). 通过并行预测预取加速马尔可夫链蒙特卡罗。载于第三十届人工智能不确定性会议论文集，N. L. Zhang & J. Tian 编。AUAI Press。第22–31页。Bardenet, R., Doucet, A. & Holmes, C. (2017). 关于高维数据的马尔可夫链蒙特卡罗方法。*J. Mach. Learn. Res.* **18**, 1–43. Chanialidis, C., Evers, L., Neocleous, T. & Nobile, A. (2014). MCMC 中的回溯采样及其在 COM-泊松回归中的应用。*Stat* **3**, 273–290. Cornish, R., Vanetti, P., Bouchard-Cˆot´e, A., Deligiannidis, G. & Doucet, A. (2019). 面向大型数据集精确贝叶斯推断的可扩展 Metropolis–Hastings 算法。载于第*36*届国际机器学习会议论文集，K. Chaudhuri & R. Salakhutdinov 编，第97卷机器学习研究论文集。PMLR。第1351–1360页。Deligiannidis, G., Doucet, A. & Pitt, M. K. (2018). 相关伪边际方法。*J. R. Statist. Soc. B* **80**, 839–870. Doucet, A., Pitt, M. K., Deligiannidis, G. & Kohn, R. (2015). 使用无偏似然估计量时马尔可夫链蒙特卡罗的高效实现。*Biometrika* **102**, 295–313. Fe ng, W., Guo, H., Wang, C., Wang, J. & Yin, Y. (2025). 朝着马尔可夫链蒙特卡罗去随机化迈进。*SIAM J. Comput.* **54**, 775–813. Filippone, M. & Girolami, M. (2014). 高斯过程的伪边际贝叶斯推断。*IEEE Trans. Pattern Anal. Mach. Intell.* **36**, 2214–2226. Grazzi, S. & Zanella, G. (2026). 基于 Picard 映射的 Metropolis 马尔可夫链并行计算。*Biometrika* **113**, asag022. Huber, M. (2004). 使用界链进行完美采样。*Ann. Appl. Probab.* **14**, 734–753. Li, C., Sra, S. & Jegelka, S. (2016). 矩阵逆形式的高斯求积及其应用。载于第*33*届国际机器学习会议论文集，M. F. Balcan & K. Q. Weinberger 编，第48卷机器学习研究论文集。PMLR。第1766–1775页。Liu, H., Wang, C. & Yin, Y. (2026). 超越局部均匀性的局部吉布斯采样。载于*2026*年 *ACM-SIAM* 离散算法年度研讨会论文集（*SODA*）。工业与应用数学学会。第996–1025页。Murdoch, D. J. & Green, P. J. (1998). 从连续状态空间进行精确采样。*Scand. J. Statist.* **25**, 483–502. Murray, I. & Adams, R. P. (2010). 潜在高斯模型协方差超参数的切片采样。载于神经信息处理系统进展，J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel & A. Culotta 编，第23卷。Curran Associates。第1732–1740页。

Murray, I., Adams, R. P. & MacKay, D. J. C. (2010). 椭圆切片采样。载于第十三届国际人工智能与统计会议, Y. W. Teh & M. Titterington 编，第9卷机器学习研究论文集。PMLR。第541–548页。Papaspiliopoulos, O. & Roberts, G. O. (2008). 狄利克雷过程层次模型的回顾式马尔可夫链蒙特卡罗方法。*Biometrika* **95**, 169–186. Prado, E., Nemeth, C. & Sherlock, C. (2026). 采用可扩展子采样的 Metropolis–Hastings 算法。*J. Am. Statist. Assoc.* 已在线发表。doi:10.1080/01621459.2026.2710382. Sherlock, C., Thiery, A. H. & Golightly, A. (2021). 延迟接受随机游走 Metropolis 算法的效率。*Ann. Statist.* **49**, 2972–2990. Sherlock, C., Thiery, A. H., Roberts, G. O. & Rosenthal, J. S. (2015). 关于伪边际随机游走 Metropolis 算法的效率。*Ann. Statist.* **43**, 238–275. Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. & B¨urkner, P.-C. (2021). 秩归一化、折叠与局部化：一种用于评估 MCMC 收敛性的改进 $\widehat{R}$（附讨论）。*Bayesian Anal.* **16**, 667–718. Xue, Y., Liao, X., Carin, L. & Krishnapuram, B. (2007). 使用狄利克雷过程先验进行分类的多任务学习。*J. Mach. Learn. Res.* **8**, 35–63. 地雷雷达数据：LandmineData.mat，公开的 GitHub 镜像位于 https: // github.com/daizhongxiang/Differentially-Private-Federated-Bayesian-Optimization/ blob/main/LandmineData.mat。Yan, X., Noergaard, M., Morse, C. L., Liow, J.-S., Hong, J., Greve, D., Telu, S., Kim, M.-J., Montero Santa- maria, J. A., Galassi, A., Feng, N., Williams Avram, S. K., Usdin, T. B., Wu, S., Zhang, A., Manly, L. S., Jenkins, M., Van Buskirk, M., Lee, A., Zoghbi, S. S., Pike, V. W., Zanotti-Fregonara, P. & Innis, R. B. (2025). 健康人群中环氧合酶-2（一种神经炎症潜在生物标志物）的 PET 定量。*J. Nucl. Med.* **66**, 398–404. Yan, X., Noergaard, M., Morse, C. L., Liow, J.-S., Hong, J., Telu, S., Kim, M.-J., Montero Santamaria, J. A., Galassi, A., Feng, N., Williams Avram, S. K., Usdin, T. B., Wu, S., Zhang, A., Manly, L., Jenk- ins, M., Van Buskirk, M., Lee, A., Zoghbi, S. S., Pike, V. W., Innis, R. B. & Zanotti-Fregonara, P. (2026). 健康人群中环氧合酶-2（COX- 2）的正电子发射断层成像（PET）定量：一种神经炎症潜在生物标志物。OpenNeuro 数据集 ds004869，版本1.4.1。[数据集]。doi:10.18112/openneuro.ds004869.v1.4.1. Zhou, Y., Aston, J. A. D. & Johansen, A. M. (2013). 区室模型的贝叶斯模型比较及其在正电子发射断层成像中的应用。*J. Appl. Statist.* **40**, 993–1016.

伪边际马尔可夫链的精确局部查询的补充材料

Zihan Xu

S1. 补充证明S1.1. 工作量界总结下文引用的符号、假设和编号结果均来自主论文。表 S1 汇总了工作量保证。

表 S1。 查询执行的设置和工作量界

结果	设置和执行器	关键条件	工作量界
定理 2	有限图；其余来自 C_s 的查询；算法1	规范范围；成本界关于 G_s 可测	关于……的条件界 (13) 中的 $E(W_s \mid G_s)$
推论 1	有界度图；在最多 L 每个节点的更新次数；算法1	独立高斯提议；$(\Delta-1)q_D(w) < 1$；常见可积 C	$E\{W_{\text{inc}}(Q)\} \le \kappa\lvert Q\rvert$，一致地 in n
定理 3	下降路径块；$Q = \{(1, nL)\}$；算法 S1	独立高斯提议；固定有限 w，L；通常可积 C	$E(N) \le (1-\rho)^{-1}$；增量工作量由 (15) 界定，关于 n 一致地
推论 2	路径和网格调度在 §4.4 中；证书执行以及 B_{point}	相应族的条件相应族；严格正估计	预期增量工作量：$\Theta(nL)$ 用于 B_{point}；$O(1)$ 用于证书执行

该表列出了关键条件；完整假设和操作成本见所引用的结果。这里，n 表示图顶点的数量，N 表示已解析路径块的数量。对 n 一致的界限将 L、输入分布以及适用的模型和成本参数固定。定理 2 涉及W_s，包括b_s 中的额外准备工作和输出工作；其他各行涉及 W_{inc}，其设置和输出成本按式 (4) 计入。

替代文本：四项结果将图和查询条件与工作量界联系起来。在指定的路径和网格上，证书执行的预期增量工作有界，而逐点评估器的预期增量工作则随图规模线性增长。

S1.2. 条件工作量界

*定理 2 的证明。*我们首先确立未揭示阈值的条件分布。给定F，这些阈值最初服从乘积均匀分布。每次揭示时，当前执行记录决定接下来选择哪个索引，或是否在请求边界停止揭示。如果选中了某个索引，则在以相应阈值为条件后，其余未揭示坐标仍相互独立且服从均匀分布。停止决策仅使用当前执行记录和F，因此不会对这些坐标施加任何额外条件。对最多 T 次揭示进行归纳，可得在 G_s 条件下仍服从相同的乘积均匀分布。缓存由 G_s 决定，并且在此边界上，所有具有已揭示阈值的更新都已返回并被缓存。因此，未缓存更新对应的阈值构成一个子集，且服从该乘积分布。

固定 G_s，并将这些更新列为参考调度中的 $\tau_1 < \cdots < \tau_m$。在此条件下，所有缓存记录均为固定值。因此，可以根据 F 和较早的未缓存阈值重构 τ_r 处的参考旧记录和候选项。τ_r 处的规范端点

因此相对于 $\mathrm{H}_{r-1}=\sigma(\mathrm{G}_s,\ U_{\tau_1},\ \ldots\ldots,\ U_{\tau_{r-1}})$ 可测，其中 $\mathrm{H}_0=\mathrm{G}_s$。给定 H_{r-1}，偏移量 α_{τ_r} 是固定的，而 U_{τ_r} 服从均匀分布。模 1 平移保持这一分布，因此

$$\mathrm{pr}(V_{\tau_r}\in A\mid \mathcal{H}_{r-1})=\mathrm{Leb}(A),$$

其中，A 是 $[0,1)$的任意博雷尔子集，Leb 表示勒贝格测度。每个更早的 $V_{\tau_i}, i<r,$ 都是 H_{r-1}-可测的。因此，连续条件化可得所有 V_{τ_r} 条件独立，进而事件 B_t也条件独立，因为每个 p_t 都是 G_s-可测的。

仅当阈值位于当前未决带内时，才会调用未解决的邻居。范围细化使该带包含在规范带内。因此，来自更新 t 的调用意味着$0\le V_t<\alpha_t-\alpha_{_t}\le p_t$，进而意味着 B_t。对于此更新之后的每次调用，这一蕴含同样成立，因为这些调用保留了原始阈值。集合 $\mathrm{E}j$ 互不相交且是 G_s-可测的，因此节点事件 O_j 条件独立，其概率为 (10)。

先考虑未缓存记录进入节点的情形。来自请求的进入会在 Q_s中产生一个根节点；否则，将发起调用的节点指定为父节点。在发起调用的节点处，对前一条记录的调用不会添加边。每个父节点都先于其子节点进入，因此这些连接构成一片森林。根据前述论证，每条边都要求父节点开放。每当 $I_j=1$时，沿父节点追溯便会得到 $\Gamma_s(q,\ j)$中的一条路径，其中某个 $q\in \mathrm{Q}_s$，且 j之前的每个顶点都开放。沿每条路径的条件独立性结合并集界，便可得出 (11)。

生成触发条件和另一个并集界给出 $\mathrm{pr}(J_j=1\mid \mathrm{G}_s)\ \le h_j^{+}$。对（12）取条件期望得到（13），因为所有成本界都是 G_s-可测的。节点界包括对自身前驱记录所进行的工作，而生成界包括记录仍未解析的节点处的范围。□

S1.3。路径块界

定理 *3* 的证明。 查询返回时，已解析块构成一个前缀 1、……、N，因为一个块只能请求其右邻居。记 $p_i=p(F_i,\ F_{i+1})$ 和 $C_j=C(F_j)$。对于 $0\le h<n$，令 M_h^{o} 和 M_h^{e} 分别表示 p_i 在 1、……、h 中奇数索引和偶数索引上的乘积，其中空乘积等于 1。以分配给节点地址的所有原始输入块为条件。到达块 $h+1$ 需要前 h 个块均处于规范开放状态，因此 §S1.2 中建立的条件独立性给出

$$\mathrm{pr}(N\ge h+1\mid F_1,\ldots,F_n)\le\prod_{i=1}^{h}p_i.$$

取平均并应用柯西–施瓦茨不等式可得

$$\begin{aligned}\mathrm{pr}(N\ge h+1)&\le \mathrm{E}(M_h^{\mathrm{o}}M_h^{\mathrm{e}})\\&\le\left[\mathrm{E}\{(M_h^{\mathrm{o}})^2\}\mathrm{E}\{(M_h^{\mathrm{e}})^2\}\right]^{1/2}=\rho^h.\end{aligned}$$

每个奇偶组使用互不相交的输入块对，因此其二阶矩是因子 $\mathrm{E}(p_i^2)\ =\rho^2$ 的乘积。两个组中共有 h 个因子，这就得到最终等式。对于 $h\ge n$，尾概率为零，对尾概率求和即可证明（14）。

块 $h+1$ 处的候选成本由 $C_{h+1}\mathbf{1}\{N\ge h+1\}$ 界定。由于 C_{h+1} 由输入块决定，根据上面的条件界可得

$$\mathrm{E}\{C_{h+1}\mathbf{1}\{N\ge h+1\}\}\le\mathrm{E}(C_{h+1}M_h^{\mathrm{o}}M_h^{\mathrm{e}}).$$

期望候选成本在 c 根节点处至多为 c，在存在块 2 时至多为 $\mathrm{E}(C_2p_1)\ =\delta$。对于 $2\le h<n$，变量 $C^{1/2}{}_{h+1}M_h^{\mathrm{o}}$ 和 $C^{1/2}{}_{h+1}M_h^{\mathrm{e}}$ 的二阶矩有限，因为

它们的平方由 C_{h+1}界定。因此，柯西–施瓦茨不等式给出

$$\begin{aligned}\mathrm{E}(C_{h+1}M_h^{\mathrm{o}}M_h^{\mathrm{e}}) &\le \left[\mathrm{E}\{C_{h+1}(M_h^{\mathrm{o}})^2\}\mathrm{E}\{C_{h+1}(M_h^{\mathrm{e}})^2\}\right]^{1/2}\\ &= (c\varepsilon)^{1/2}\,\rho^{h-1}.\end{aligned}$$

在包含 p_h 的组中，最后一个因子和端点成本贡献 $\mathrm{E}(C_{h+1}p_h^2) = \varepsilon$。该组中的所有其他因子都使用与 (F_h, F_{h+1}) 不相交的区块。另一个奇偶组不使用任何 F_{h+1}，因此成本贡献为 c。其余 $h-1$ 个因子在平方根之前各贡献 ρ^2。δ 和 ε 均至多为 c，因为 $0 \le p_i \le 1$。对区块求和可得

$$\mathrm{E}\left(\sum_{j=1}^{N} C_j\right) \le c + \delta + (c\varepsilon)^{1/2}\sum_{h=2}^{\infty}\rho^{h-1} = c + \delta + \frac{\rho(c\varepsilon)^{1/2}}{1-\rho}.$$

设 G 为生成的输入区块数量。只有已解析前缀的紧邻右邻居可以在未被解析的情况下生成，因此 $G \le \min\{n, N+1\}$。每个已解析区块至多请求其右邻居一次，并至多进行 $L+1$ 次证书比较。因此，§S2.2 中的核算将管理成本界定为 $24L(N+1)$ 个逻辑单位，其中包括输入生成以及未解析边界区块处的范围。将其期望值加到候选成本界限中，并使用 (14)，即可证明 (15)。□

S1.4。同查询分离

证明。更新时的旧参数是其初始值，或同一节点上更早的候选项。因此，独立连续提议几乎必然不同于旧参数。结合严格正性，这迫使 B_{point} 在每次决策前解析所有邻居记录。在降序路径调度中，解析块 $j < n$ 需要块 $j+1$ 的最后一条记录，因此查询会遍历每个块。在网格的最终扫描中，若某顶点的西侧和北侧邻居存在，它们会在调度中排在该顶点之前。从右下角顶点开始，这些依赖关系会遍历每个顶点在最终扫描中的更新。随后解析自身的前驱记录，会纳入这两类结构中每个节点的所有 L 更新。缓存使每个候选项只评估一次，因此调用次数为 N^{point} 次 $= nL$。每次评估至少产生一个工作单位。公共包络 $C(F_j)$ 将其余候选项工作量限制在 $\sum_{j=1}^{n} C(F_j)$ 以内，期望为 nc。有界度、索引访问和固定 L 带来 $O(nL)$ 的管理工作量，从而证明 B_{point} 的所述阶数。路径的证书界限由定理 3 得出，网格的证书界限则由推论 1 得出，因为 $|Q| = 1$。精确的逐点决策和定理1保证与 $\mathrm{R}\varrho$ 一致。□

S1.5。推论 *1* 的颜色类论证

推论 *1* 的证明。在初始记录缓存处，$\mathrm{G}_0 = \sigma(F)$。对于每个节点 j，令 A_j 为由 F_j 形成的 A 的值。历史范围允许在节点 j 的每次更新时选择 $\omega_t = wA_j\sum_{k\in N_j} A_k$。此类更新至多有 L 次，因此

$$\pi_j \le \bar{\pi}_j := 1 - \exp\left(-wLA_j\sum_{k\in N_j} A_k\right).$$

写作 $q = q_D(w)$。使用 A g 的独立副本将邻居和补齐为 Δ 项，可得 $\mathrm{E}(\bar{\pi}_j^D) \le q^D$。

令 $\mathrm{K}_1, \ldots, \mathrm{K}_D$ 为颜色类。同一颜色类中的不同顶点具有不相交的闭邻域 $\{j\} \cup \mathrm{N}j$，因此它们的 $\bar{\pi}_j$ 取决于不相交的输入块，并且相互独立。对于任意顶点集 $\mathrm{A} \subseteq V$，令 $\mathrm{A}_d = \mathrm{A} \cap \mathrm{K}_d$。跨颜色类应用 H¨older 不等式，并利用独立性

在每个颜色类内得到

$$\mathrm{E}\left(\prod_{v\in\mathcal{A}}\bar{\pi}_v\right)\le\prod_{d=1}^{D}\left\{\mathrm{E}\left(\prod_{v\in\mathcal{A}_d}\bar{\pi}_v^{D}\right)\right\}^{1/D}=\prod_{v\in\mathcal{A}}\{\mathrm{E}(\bar{\pi}_v^{D})\}^{1/D}\le q^{|\mathcal{A}|}.$$

空乘积等于一。

对于一条自避路径 $\gamma=(v_0,\ \ldots\ldots,\ v_\ell)$，令 $\mathrm{P}\gamma=\{v_i:\ 0\le i<l\}$ 为其端点之前的顶点。(11) 中的条件可达性界为这条路径赋予权重 $\prod_{v\in\mathcal{P}_\gamma}\pi_v$，根据前述不等式，其期望至多为 q^{ℓ}。从给定顶点出发、长度为 $l\ge 1$ 的路径至多有 $\Delta(\Delta-1)^{\ell-1}$ 条。对于 $a\in\{0,\ \Delta\}$，定义

$$\Xi_a=1+\sum_{\ell=1}^{\infty}\Delta(\Delta-1)^{\ell-1}q^{(\ell-a)_+},$$

其中 $x_+=\max\{x,\ 0\}$。两个级数都收敛，因为其尾项的比率为 $(\Delta-1)q<1$。令 I_j 表示在 j 处至少有一条新记录得到解析。对端点求和，并对至多 $|Q|$ 个查询节点求和可得

$$\mathrm{E}\left(\sum_{j}I_j\right)\le|Q|\Xi_0.$$

对于一条以 j 为终点的路径，从 $\mathrm{P}\gamma$ 中移除 j 的所有邻居，并记为 $\mathrm{P}'_\gamma=\mathrm{P}\gamma\setminus\mathrm{N}j$。这至多移除 Δ 个顶点。每个剩余因子 $\bar{\pi}_v$ 仅取决于 F_j 以外的区块，因此它们的联合乘积与 $C(F_j)$ 独立。由于 $0\le\bar{\pi}_v\le 1$，我们得到

$$\begin{aligned}\mathrm{E}\left\{C(F_j)\prod_{v\in\mathcal{P}_\gamma}\bar{\pi}_v\right\}&\le\mathrm{E}\left\{C(F_j)\prod_{v\in\mathcal{P}'_\gamma}\bar{\pi}_v\right\}\\&=c\,\mathrm{E}\left(\prod_{v\in\mathcal{P}'_\gamma}\bar{\pi}_v\right)\le cq^{|\mathcal{P}'_\gamma|}\le cq^{(\ell-\Delta)_+}.\end{aligned}$$

这里 $|\mathrm{P}'_\gamma|\ge(l-\Delta)_+$和 $0<q<1$。成本 $C(F_j)$ 是 G_0-可测的，因此我们可以先将条件可达性界限乘以它，再取期望。对所得的路径界限求和可得

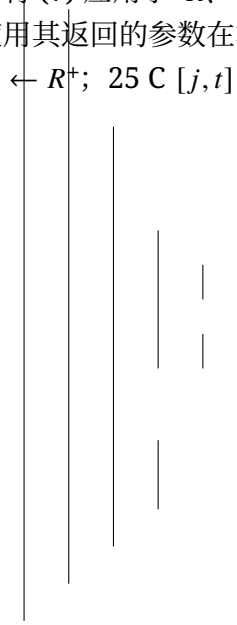
$$\mathrm{E}\left(\sum_{j}C(F_j)I_j\right)\le c|Q|\Xi_\Delta.$$

还需计算候选成本界之外的管理工作。每个进入的节点最多解决 L 次更新。每次更新最多进行 $\Delta+1$ 次比较，并且对每个未解决的邻居最多请求一次。使用逐节点事件索引时，每个历史查找或范围计算对每个邻居最多检查 L 个条目。将每次范围计算的成本计入请求该计算的更新，包括记录仍未解决的节点上的范围。根据推论的成本约定，原始输入和事件索引已预先准备好。候选生成和评估的成本已由 $C(F_j)$ 涵盖。因此，剩余的管理工作至多为 $K(|Q|+\sum_j I_j)$，其中 $K<\infty$ 仅取决于 Δ，L 以及固定的操作成本。合并这些界限可得

$$\mathrm{E}\{W_{\mathrm{inc}}(Q)\}\le\{c\Xi_\Delta+K(1+\Xi_0)\}|Q|,$$

算法 S1。路径上的区块查询执行。

输入：§4.3 中的路径规范和可寻址输入。输出：完整记录 $R_1^{(nL)}$。1 初始化 C $[j,0]=R_j^{(0)}$、 $j=1$、……、 n，并为输入块及其区间创建空缓存；2 过程 GetBlock(j)：3 如果 j 处的输入块未被缓存，则 4 生成并缓存其 L 参数和原始辅助输入；5 缓存由 0、Y_{j1}、……、Y_{jL} 跨越的区间；6 过程 ResolveBlock(j)：7 $\tau\leftarrow(n-j+1)L$；8 如果 (j、 τ) 已缓存，则返回 C $[j,\tau]$；9 GetBlock(j)；$R\leftarrow$C $[j,0]$；10 对 l 从 $\leftarrow 1$ 到 L 执行 11 $t\leftarrow(n-j)\,L+$l；⌢12 使用缓存输入在 (j、l) 处一次性构造并计算 $R^{+}=(Y_{j\ell}$、 v、l$_j(Y_{j\ell}$、 $v))$；⌢13 如果 l$_j(Y_{j\ell}$、 $v)>0$，则 14 从其地址检索并保留 U_t；15 当 $j>1$ 时，将单元素集合 {0} 用作左邻居；16 如果 $j<n$，则 17 如果 ($j+1$、$(n-j)\,L$) 已缓存，则 18 将其参数用作右侧单元素集合；19 否则 20 GetBlock($j+1$)；使用其右侧的缓存区间；21 计算 (6)，并将 (7) 应用于 R、 R^{+}、 U_t；22 如果比较结果未决，则 23 ResolveBlock($j+1$)；使用其返回的参数在右侧重复比较，同时保留 R、 R^{+}、 U_t；24 如果比较接受，则 $R\leftarrow R^{+}$；25 C $[j,t]\leftarrow R$；26 返回 R；27 返回 ResolveBlock(1)。



这就证明了推论 $\kappa=c\Xi_{\Delta}+K(1+\Xi_0)$。 □

S2. 区块过程与操作计数

S2.1. 区块执行

算法 S1 将执行器专用于主论文 §4.3 中的降序路径调度。在块 j 的参考时刻，左邻居仍处于其初始状态，右邻居则处于其区块结束状态。即使调用方已前进到不同状态，递归执行仍会使用这些历史状态。第一行中的初始缓存属于准备工作。输入块及其范围会在收到请求时生成，重复的地址会返回相同的随机输入。

表 S2. 路径实现的运算次数

操作	上界
调度	$LN + 2N$
证书比较	$LN + N$
栈推入、弹出和递归请求	$3N$
候选值调度、提交和拒绝处理	$3LN$
块查找	$N + LN$
阈值读取	LN
参数和原始辅助输入生成	$2LG$
范围元素访问（$L + 1$）	G

这些是指定实现的逻辑操作单位。固定大小记录、索引访问和摊销缓存操作的成本均为单位成本。后端评估、初始记录和数据准备仍分别计费。

替代文本：调度和证书比较的规模随已解析块及其长度而变化。输入生成和范围访问取决于生成了多少个块的输入。

S2.2。逻辑工作单位

令 G 为已生成输入区块的数量， N 为已解析块的数量。查询完成时，已解析块构成一个前缀，且只有其紧邻的右邻居可以在尚未解析时生成。因此， $N \le G \le \min\{n, N + 1\}$。每个已解析块最多请求其右邻居一次。该请求返回后，对于调用块后续的所有更新，邻居的块末记录都会被缓存。因此，一个块最多进行 L 次初始比较，并在递归返回后进行一次比较。候选估计值为零时，会在获取其阈值之前被拒绝。每个生成的块提供 L 个参数输入和 L 个原始辅助输入，并根据 $L + 1$ 个值构造一次其范围。重复访问会使用缓存输入和范围。表 S2 将候选值调度与 $C(F_j)$所涵盖的构造和评估工作分开计数。一个逻辑单位包括执行所述操作所需的固定标量算术和记录访问。对于范围构造，一个逻辑单位涵盖一个值的两个端点比较。前六行合计至多为 $7LN + 7N$ 个逻辑单位。最后两行额外增加至多$(3L + 1)G$ 个单位。由于 $L \ge 1$ 且 $G \le N + 1$，总计满足

$$\begin{aligned} 7LN + 7N + (3L + 1)G &\le 18LN + 4L \\ &\le 18L(N + 1) \le 24L(N + 1). \end{aligned}$$

在其他固定操作价格不变的情况下，将此上界乘以它们的最大价格。这些是所列执行操作的成本，准备和统计输出按照（4）另行计费。

对于估计值可能为零的实验，按该顺序考虑 $w = 0$。01、0。2、1。5。对于 $L = 1$，在三个权重下，证书候选项平均计数均为 1.25；对于 $L = 2$，分别为 2.5、2.5 和 3.5。对于 $L = 4$，相应的均值为 5、9 和 12。每个均值都使用相同的四个固定输入流，并且在 $n \in \{4、16、 64、 256\}$ 中保持不变。

对于正估计值实验，我们在每个节点设置 $R_j^{(0)} = (0，\ 0，\ e^{-1/2})$，并以对数刻度存储估计值。对于 $u \sim N(0，\ 1)$，恒等式 $\mathrm{E}\{\exp(u - 1/2)\} = 1$ 表明 $\mathit{li}(\theta) = \exp(-\theta^2/2)$具有无偏性。对于每个输入流，四种路径长度下，公共节点和更新地址处的参数候选值、辅助输入和接受阈值均相同。每个执行器都会准备自己的初始记录，并从全新缓存开始。

表 S3. 正估计路径实验的证书执行时间均值（毫秒）

n	准备	增量	总计
32	0.023	0.420	0.451
128	0.079	0.430	0.517
512	0.249	0.389	0.645
2048	1.371	0.490	1.870

替代文本：随着路径规模从 32 个节点增加到 2048 个节点，平均准备时间从 0.023 毫秒上升到 1.371 毫秒。平均增量时间保持在 0.389 至 0.490 毫秒之间。

对于每个输入流，四种路径长度下已解析块的数量保持不变。在每种路径长度下，N 的样本标准差为 0.807。在全部 400 次证书运行中，$G = N + 1$。管理使用 54–281 个逻辑单位，并且每次运行都满足界限 $24L(N + 1)$。

表 S3 报告了插桩实现中证书执行的平均耗时。每一行都使用相同的 100 个输入流。增量耗时包括输入生成、范围维护和候选评估。总耗时还包括准备、查询输出，以及各单独计时阶段之外的耗时。随附文件 data/positive path timings.csv 和 data/positive path_summary.csv 报告了各次运行及其汇总结果。

S3. 受控实验

S3.1. 模型与固定配置

图 3 中的六种配置使用了四个已保存的模拟数据集。图 4 使用 P1 和 P2。它们的路径上均有 128 个标量参数，节点精度为 1，边权重为 0.8。参数先验是式 (1) 中的高斯图先验。P1 和 P2 使用数据集 0 和 1。S1 和 B1 使用数据集 8；S2 和 B2 使用数据集 9。在每个基因活性数据集中，S 和 B 配置使用不同的粒子估计器和参数提议。每一对中的数据和统计模型相同。

对于 P1 和 P2，潜在过程和观测值为

$$X_{j1} \sim N\{0, 0.4^2/(1 - 0.7^2)\}, \qquad X_{jt} \mid X_{j,t-1} \sim N(0.7X_{j,t-1}, 0.4^2),$$
$$Y_{jt} \mid X_{jt}, \theta_j \sim \text{Poisson}\{8e^{\theta_j + X_{jt}}\},$$

其中，X_{jt} 是潜在状态，Y_{jt} 是节点 j 在观测时间 t的观测计数。每个节点有 16 个观测值。自举粒子滤波器使用 64 个粒子，并在每个后续观测时进行多项式重采样。参数提议为 $N(\theta_j, 0.35^2)$，调度包含 128×192 次更新。

对于 S1、S2、B1 和 B2，启动子 $G_j(t) \in \{0, 1\}$ 以速率 $0.7e^{\theta_j}$ 开启，以速率 0.9 关闭。分子以速率 $3G_j(t)$ 生成，且每个分子以速率 0.8 降解。启动子和分子计数均从零开始。在八个时点进行计数观测，时点之间相隔 0.5 个时间单位；对于每个存在的分子，其独立捕获概率为 0.95。给定一个启动子保持区段，存活分子数服从二项分布，新产生的存活分子数服从泊松分布。该实现对这些转移进行模拟，不截断分子计数。

B1 和 B2 使用包含 16 个粒子的自助滤波器，以及 $N(\theta_j,\ 0。35^2)$ 参数提议。S1 和 S2 使用 16 个粒子、观测条件计数端点，以及独立的 $N(0,\ 1)$ 参数提议。给定一个区间内的启动子路径，观测到的存活分子数是相互独立的二项分布变量与泊松变量之和。它们的有限卷积给出观测权重；随后根据观测条件抽取未观测到的存活分子。当启动子初始处于关闭状态且观测计数超过初始分子计数时，会对首次激活发生在该区间内这一条件进行处理。其概率计入重要性权重。所有四种配置均使用 128×48 更新。归一化参数提议密度计入接受比率。所有粒子估计值均使用增量权重均值的乘积（Andrieu 等人，2010）。

每次比较都共享初始完整记录、节点选择输入、参数输入、粒子随机种子和接受阈值。初始估计通过相同的初始化程序获得，并计入每次执行的开销。稀疏任务会在每个参考时间返回两个节点的历史记录。完整任务会返回全部 128 条历史记录。五次稀疏重复和一次完整查询重复构成固定计时设计，执行器顺序在各次重复中循环轮换。它们不是额外的独立数据集。

缓存比较使用来自八个泊松数据集和八个基因活性数据集的 128 条保存的轨迹，每个数据集运行八次。我们纳入每个数据集指定的全部八次运行。每次运行都使用上述 128 节点路径先验，并请求节点 1 和 2 的完整参数历史。六个泊松数据集的暴露量为 8，另外两个为 1。六个基因活性数据集的捕获概率为 0.95，另外两个为 0.65。其余模型参数如上所述。泊松运行使用 64 个自助法粒子、 $N(\theta_j, 0.35^2)$ 个参数提议和 128×192 次更新。基因活性运行使用 16 个粒子，并采用受观测条件约束的计数端点、独立的 $N(0, 1)$ 参数提议和 128×48 次更新。

对于每个至少有一个所需邻居记录尚未解析的决策，我们保留旧记录、候选记录和接受阈值。我们使用相同的提议输入从初始状态传播历史范围，并重新计算接受界限。如果这些替代界限使决策仍无法确定，我们就将该决策计入受缓存范围辅助的决策数中。无论决策发生在第一次测试时，还是发生在此前的邻居请求之后，均计入该数量。

S3.2。测得的成本和单次运行

实验使用 Windows 11、配备 24 个逻辑处理器的 Intel64 Family 6 Model 183 处理器、Python 3.12.14 和 NumPy 2.3.5。各次运行均在同一主机上按顺序部署。共有四个似然工作线程可用于公共初始化和 Picard 候选评估。证书和逐点生产更新在控制进程中进行评估。Picard 窗口为 16。数值库线程限制为 1。这些比较测量的是任务延迟；工作线程作业时间可以重叠。

经过时间是保存的算法时间加上输入加载开销。它包括初始化、工作进程启动与关闭、随机输入、索引、查询执行以及规定的汇总计算。模拟数据生成、结果审计和审计序列化不计入该生产计时器。随附的 data/execution timings.csv 包含全部 180 条记录的部署数据，包括执行器、查询范围、重复次数、候选数量和各计时组成部分。表 S4–S5 汇总了主论文中使用的比较结果。每次运行的配对记录见 data/paired timings.csv。

在这六种配置下，双节点任务中证书执行的平均解决比例介于 2.69% 和 11.95% 之间。对于完整任务，所有计划更新均已解决。在每种配置中，两种范围使用相同的初始记录和随机输入，在

表 S4。 稀疏查询：测得的经过时间（秒）

Case	逐点	证书	平均比率	配对比率范围
P1	12.995 (6.837)	2.613 (1.227)	0.201	0.102–0.319
P2	18.888 (1.409)	3.727 (0.893)	0.197	0.151–0.245
S1	7.996（1.370）	3.030（0.881）	0.379	0.216–0.504
S2	6.699（1.104）	3.155（0.302）	0.471	0.389–0.643
B1	5.022 (1.859)	1.529 (0.397)	0.304	0.277–0.385
B2	2.725 (1.198)	1.300 (0.287)	0.477	0.254–0.874

括号内为五次运行的标准差。平均比率为平均证书时间除以平均逐点时间。配对比率范围表示五次运行中证书时间与逐点时间比率的最小值和最大值。

替代文本：在每次配对运行中，证书执行都比逐点执行更快。在六个案例中，其平均经过时间为逐点执行均值的 20% 至 48%。

表 S5。 查询范围及与 *Picard* 执行的比较

	稀疏查询：五次运行的均值			所有查询：一次运行		
Case	证书	Picard	比率	证书	Picard	比率
P1	2.613	8.370	0.312	28.035	12.275	2.284
P2	3.727	10.945	0.341	25.697	10.373	2.477
S1	3.030	7.125	0.425	32.094	7.402	4.336
S2	3.155	7.287	0.433	34.309	7.587	4.522
B1	1.529	4.157	0.368	22.988	4.286	5.363
B2	1.300	4.187	0.310	19.014	4.600	4.134

所有时间均以秒为单位。稀疏证书运行与表 S4 中的运行相同。每个比率均为证书时间除以 Picard 时间。替代文本：在全部六个案例中，对于稀疏查询，证书执行的均值时间都短于 Picard。对于涵盖所有查询的每次单独运行，Picard 都更快。

首次配对运行。Picard 的候选项评估次数和并行轮次在这六次比较中均在两个范围内一致。

附加的范围实验使用 P1 和 P2，以及上文给出的模型、粒子滤波器和提议设置。我们请求节点 1 至 m的历史记录，其中 $m \in \{2、8、32、64、128\}$。对于每个数据集，我们在每个范围使用五个随机输入流。在每个流中，三个执行器在每个范围共享初始完整记录和随机输入。每次执行均需支付初始化成本，并从全新缓存开始。Picard 窗口固定为 16 次更新，使用四个似然工作线程。

对于每个数据集、输入流和查询范围，请求的历史记录在各执行器之间一致。在 64 个节点时，P1 中证书时间与 Picard 时间的各个比率范围为 0.948 至 1.037，P2 中则为 0.987 至 1.050。随附的数据/query scope timings.csv 包含全部 150 次部署。文件 data/query scope summary.csv 给出了平均成本及其标准差。

对于窗口检查，我们使用第一个输入流，并请求两个节点历史记录或全部128个。对于每个数据集和查询范围，我们分别在窗口大小为4、16和64次更新时运行一次 Picard。使用4个工作线程时，将窗口从16增大到64，会使 P1 中的候选调用增加4.8%，P2 中增加5.3%。P1 中的并行轮次从2,955降至842，P2 中则从2,951降至849。在全部四项比较中，Picard 时间均有所缩短。随附的数据/query window timings.csv 文件包含全部12次运行。

S4. 应用规格与诊断

S4.1. 断层成像模型与数值目标

输入数据为 OpenNeuro ds004869 中受试者01的基线会话，版本1.4.1（Yan 等人，2025、2026年）。我们使用前27帧，覆盖90分钟，并结合已发布的血浆曲线、全血曲线和观测权重。该数据集依据知识共享CC0公有领域奉献协议发布。原始人体研究经美国国立卫生研究院综合神经科学机构审查委员会批准，遵循方案19-M-0079和20-M-0082。

我们使用提供的空间变换和最近邻插值，将发布的解剖学标签映射到 PET 图像网格。标签 1035（左侧岛叶）经皮层脑掩膜限制后，定义了 1311 个查询体素。我们连接左侧皮层脑掩膜中面相邻的体素对。分析域由包含查询体素的连通分量构成，共有 41500 个体素。

活度浓度以 $\mathrm{kBq\,mL^{-1}}$ 表示，时间以分钟计。卷积网格包含 6000 个等间距点。从帧中点减去 -0.12912561648752363 分钟的延迟，以获得评估时间。血液分数固定为 $v_b = 0.05$。

对于单组织模型，组织曲线为 $K_1 (C_p * e^{-k_2 \cdot}) (t)$。对于双组织模型，组织曲线为 $C_p * \ c_f e^{-\lambda_f} \ + c_s e^{-\lambda_s} \quad t \qquad C_p \qquad * \qquad \lambda_f = \ k_2 + (\cdot\cdot) \, (\,)$，其中 表 $k_3 + k_4 + g \quad 2 \ \lambda_s = 2k_2 k_4 \quad k_2 + k_3 + k_4 + g \quad g = \quad k_2 - k_4 \ ^2 + k_3^2 + 2k_3 \ k_2 + k_4 \ ^{1/2} \ c_f =$ 示血浆曲线，表示卷积，$(\,)/$ 、$/(\,)$、$\{(\,) \, (\,)\}$ 、$K_1 (\lambda_f - k_3 - k_4)/g$和 $c_s = K_1 (k_3 + k_4 - \lambda_s)/g$。测量曲线由组织曲线乘以 $(1 - v_b)$，再加上全血曲线乘以 v_b 得到。代码在固定网格上使用矩形求和卷积，并通过线性插值将结果映射到帧中点。

给定动力学参数和模型标签，观测值具有精度为 $w_t \tau_j$ 的独立高斯误差，其中 w_t 是给定权重，$\tau_j \sim \mathrm{Gamma}(2,\ 1)$ 采用率参数化。动力学参数的对数服从相互独立的均匀先验。原始尺度上的界限为 $K_1,\ k_2,\ k_3 \in [0。0001,\ 0。5]$，以及 $k_4 \in [0。01,\ 0。2]$；单组织模型仅使用 $K_1,\ k_2$。这里，K_1 的单位为 $\mathrm{mL\ cm^{-3}\ min^{-1}}$，其他速率的单位为 $\mathrm{min^{-1}}$。单组织模型的分布容积为 K_1/k_2，双组织模型的分布容积为 $(K_1/k_2)(1 + k_3/k_4)$。

对噪声精度进行解析积分后，似然与 $(1 + \mathrm{RSS}/2)^{-(2+27/2)}$ 成正比，其中 RSS 为加权残差平方和。双组织模型在 $\log k_2$、$\log k_3$ 和 $\log k_4$ 上分别使用 96 个高斯–勒让德点，并在每个正宽度的 $\log K_1$ 区间上使用 32 个点。单组织模型在每个正宽度的 $\log K_1$ 区间上使用 64 个点。对于其外层 $\log k_2$ 积分，自适应高斯–克朗罗德积分确定先验区间的一个划分。在每个最终子区间上应用 32 点高斯–勒让德求积规则，得到用于计算证据和体积分布的正求积权重。该划分针对固定体素组计算，组成员关系在查询前指定。

对于两个模型，K_1 分段边界为先验端点，以及将条件最小二乘中心加上尺度乘以 -256、-64、-16、-4、-1、0、1、4、16、64、256 后截取到先验范围内所得的值。取对数得到对数 K_1 分段，宽度为零的分段不作贡献。中心、尺度和数值规则在 specifications/pet/integration.py和 specifications/pet/quadrature.py 中指定。所有划分和求积值都是固定数据、组及数值设置的确定性函数。每个正原子都被分配到一个固定的对数体积箱中，其对数宽度至多为 0.002，并以该箱的几何中点表示。

该图包含 $n = 41500$ 个体素和 85619 条面相邻边。为模型标签写出 $b = (b_1,\ \ldots\ldots,\ b_n) \in \{0,\ 1\}^n$，其中 0 表示一个组织区室，1 表示两个组织区室。该

数值标签目标为

$$\widetilde{\pi}_B(b) \propto \left\{\prod_{j=1}^{n} \widetilde{Z}_j(b_j)\right\} \exp\left\{\beta \sum_{\{j,k\}\in E} \mathbf{1}\{b_j = b_k\}\right\},$$

其中， $\widetilde{Z}_j(k) > 0$ 是体素 j 和标签 k对应的确定性求积证据， E 是边集， $\beta = 0.25$，而 $\mathbf{1}\{\cdot\}$ 表示指示函数。相同的标签先验概率会在此表达式中抵消。

每次参考更新都会均匀地选取一个体素，并以相同的概率提议标签 0 或 1，且该提议独立于当前标签。在体素 j进行更新 t时， 记 x和 y为旧标签和提议标签，并记 $U_t \in (0, 1]$ 为独立均匀的接受输入。若 $y = x$，则对数接受比为零，且两次执行都会返回 x ，无需计算模型证据或解析邻居标签。否则，根据提议的对称性，对数接受比为

$$\Lambda_t = \log \widetilde{Z}_j(y) - \log \widetilde{Z}_j(x) + \beta \sum_{k:\{j,k\}\in E} \left[\mathbf{1}\{y = b_k\} - \mathbf{1}\{x = b_k\}\right],$$

其中， b_k 是更新 t之前紧邻的邻居标签。

对于 $y \neq x$， 令 K_t 为在当前比较时，所需历史标签已缓存的邻居集合。记 r_t 为剩余邻居的数量。定义

$$a_t = \log \widetilde{Z}_j(y) - \log \widetilde{Z}_j(x) + \beta \sum_{k\in K_t} \left[\mathbf{1}\{y = b_k\} - \mathbf{1}\{x = b_k\}\right].$$

每个未解决的邻居贡献 $-\beta$或 β，因此 $\Lambda_t \in [a_t - \beta r_t, a_t + \beta r_t]$。若 $\log U_t < \min\{0, a_t - \beta r_t\}$，则接受此次更新；若 $\log U_t \geq \min\{0, a_t + \beta r_t\}$，则拒绝。未决比较会在所需的较早时刻解决一个未解决的邻居，并使用相同的提议和均匀变量重复测试。

标签记录按体素和更新时间缓存；每次调用都会先解决该体素自身的前驱，再执行测试。证据值是确定性的，并按体素和模型标签分别缓存。固定一个有限的更新计划、一个有限的有序查询列表、初始标签和所有随机输入。使用精确算术，并假设每个所需的求积表都能在有限时间内求得。每次递归调用都会请求一个更早的历史标签，而每次未决比较都会解决一个此前未解决的邻居。一旦所有邻居都已解决，该区间便是单元素集合，测试会据此判定是否更新，包括恰好等于阈值的情况。给定正确的缓存标签，该区间包含参考对数比，因此每份证书都会得出参考决策。对更新时间归纳可证，算法必然终止，且结果与完整参考标签一致。对称的 Metropolis规则保持 $\widetilde{\pi}_B$，对更新体素进行均匀选择也会保持它。

对于体素 j 和标签 k，定义条件体积分布

$$H_{jk}(\mathrm{d}v) = \sum_{r} p_{jkr}\, \delta_{v_{jkr}}(\mathrm{d}v),$$

其中， r索引固定分箱， p_{jkr}是分箱 r中的归一化求积质量， v_{jkr}是其几何中点，而 δ_v 是在 v处的单位质量。标签和区域体积的联合数值目标为

$$\widetilde{\Pi}(b, \mathrm{d}v_R) = \widetilde{\pi}_B(b) \prod_{j\in R} H_{j,b_j}(\mathrm{d}v_j),$$

其中， R是包含 1311 个体素的左岛叶区域， $v_R = (v_j)_{j \in R}$。在每个查询时间点，独立的输出均匀分布生成$v_j \sim H_{j,\ b_j}$（对于 $j \in R$），区域抽样值为 $|R|^{-1} \sum_{j \in R} v_j$。相同的请求标签和输出均匀分布会产生相同的区域抽样值。对条件区域均值取平均可估计后验均值；区域抽样值的经验分位数可估计区间端点。蒙特卡罗误差是相对于这一固定数值目标进行评估的。

我们运行四条链，每条链进行 512 次名义扫描。一次名义扫描由 41,500 次随机扫描更新组成。四个初始标签向量分别为全零、全一、按固定体素顺序交替排列的零和一，以及反向交替排列。我们每 10,375 次更新请求一次区域输出，因此每条链有 2048 次查询。丢弃前 256 次查询（64 次扫描）后，每条链剩下 1792 个标签向量，总计 7168 个。

我们使用保存的求积分布条件矩，评估以区间中点替代支撑点的影响。证据值和保存的四条标签链保持不变。使用未分箱的条件矩后，估计的区域均值从 2.0537176909 降至 2.0537175120 mLcm^{-3}。相应的后验方差总计估计值降低了 0.0033%。

我们还使用每个抽样区间的下端点和上端点，界定经验区域分位数的变化。对于保存的预热后抽样结果，经验 2.5% 分位数的相应界限为 $[2.04042, 2.04451]\ \mathrm{mLcm}^{-3}$。经验 97.5% 分位数的相应界限为 $[2.06261, 2.06675]\ \mathrm{mL\ cm}^{-3}$。这些界限针对保存的抽样结果中区间位置所造成的影响，不包含蒙特卡洛不确定性和求积误差。

图 S1 比较了 12 个预先指定的区域体素在两种有限求积规则下的条件累积分布函数（CDF）和对数证据。对于双组织模型，体素 9 的累积分布函数差异最大，且对数证据的绝对变化也最大。在体素 6，双组织累积分布函数差异为 0.01141。相应的对数证据增加了 3.73×10^{-5}。这些体素的单组织对数证据变化最多为 1.01×10^{-12} 。

S4.2。断层成像成本核算

两个被比较的执行器都从相同的 2942 个缓存的体素—模型证据值和区域条件分布开始。我们测量每个进程从启动到退出的用时，其中包括缓存加载、随机输入生成和输出写入。完整执行器使用相应的优化证据计算、缓存值和当前状态表示。其预算在 725 次更新后用尽，此时尚未到达事件 10375 处的第一个请求输出。它完成了 353 次新的双组织证据计算，而证书执行完成了 177 次。测得的任务用时分别为 45.03 分钟和 23.29 分钟；只有证书执行完成了规定的 2048 次查询。

可追溯至初始缓存的历史组件总计约为 7.01 小时：区域集成耗时 6.35 小时，区域外集成耗时 0.65 小时，基底准备耗时约 56 秒。这些是去重后的历史组件，并非最终实现的实测冷启动时间。这些组件中已包含的群构造费用仅计入一次。因此，记录的任务比较衡量的是从现有缓存开始的执行过程。完整任务的完成时间尚未观测到。随附的准备组件账本列出了用于该核算的源事件。

S4.3。*Landmine*模型与抽样日程

我们使用公开镜像中的 LandmineData.mat：https://github.com/daizhongxiang/Differentially-Private-Federated-Bayesian-Optimization/blob/main/LandmineData.mat。站点1–15属于原始的高度叶状群（Xue等，2007年）。我们在不使用标签的情况下选择训练索引。在每个位点内，我们随机排列

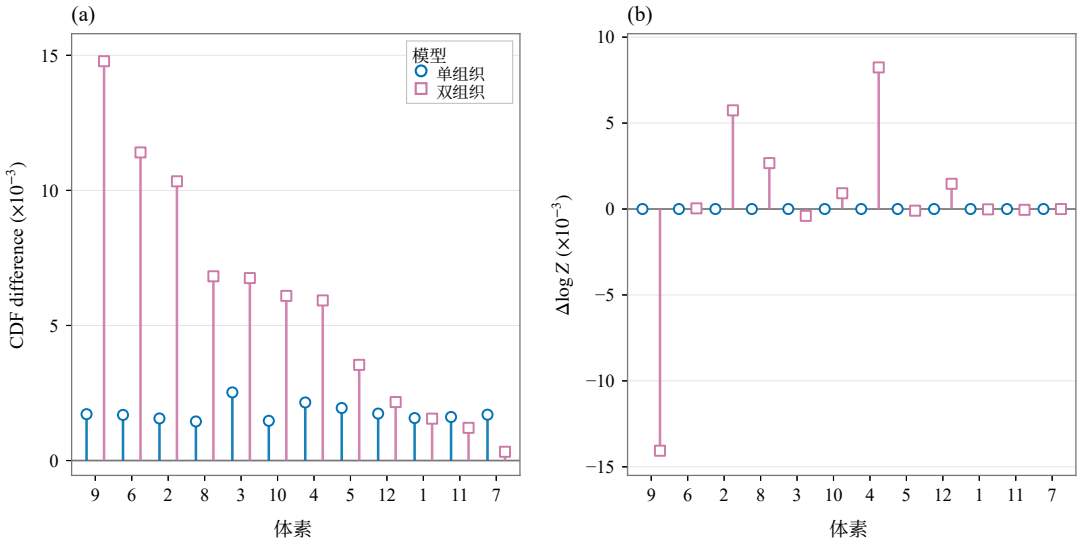


图 S1. 12 个预先指定的左岛叶体素对有限求积规则的敏感性。(a) 两种规则下分箱条件 V_T 分布的累积分布函数（CDF）之间的最大绝对差异。(b) 对数证据的变化， $\Delta\log Z = \log Z_{\text{high}} - \log Z_{\text{low}}$，其中 Z_{low} 和 Z_{high} 表示相应的求积证据值。对于单组织模型，内层高斯–勒让德阶数从 32 增至 64。对于双组织模型，每个外层阶数从 48 增至 96，内层阶数为 32。体素索引用保存顺序标识这 12 个体素。两个面板均按双组织累积分布函数差异递减的顺序排列这些索引。图例适用于两个面板。每条垂直线段连接其数值与零。

替代文本：对于双组织模型，面板 (a) 中体素 9 的累积分布差异最大以及图 (b) 中对数证据降幅最大。单组织对数证据在全部 12 个体素中变化不大。

相同特征行组成的组，并保留完整的组，直到选取的行数至少达到200行。保存的训练集在每个位点恰好有200行。特征使用合并后的训练均值和总体标准差进行中心化和尺度缩放。固定索引和缩放输入随规格说明一并提供。

每个二元标签编码为 $y \in \{-1,\ 1\}$，其条件似然为 $\Phi(y f)$，其中 Φ 是标准正态分布函数， f 是相应的潜在值。在位点 j，潜在高斯协方差的元素为

$$K_{j,rs} = e^{2\gamma_j} \exp\{-\|x_{jr} - x_{js}\|^2/(2e^{2z_j})\} + 4 + 10^{-6}\mathbf{1}_{\{r=s\}},$$

其中 x_{jr} 是训练行 r的标准化特征向量， z_j 是对数长度尺度，γ_j 是对数信号标准差。常数项是位点内公共截距的方差，最后一项仅加到对角元素上。给定 μ后，对数长度尺度在条件上相互独立，其中 $z_j \mid \mu \sim N(\mu,\ \sigma^2)$ 和 $\mu \sim N(\log 3,\ \sigma^2)$，其中 $\sigma^2 = 0。75^2$。对数振幅 γ_j 具有相互独立的 $N(0,\ 1)$ 先验，且独立于 μ和对数长度尺度。

对于 $n = 15$，记 $\zeta_j = z_j - \log 3$ 和 $\zeta = (\zeta_1, \ldots, \zeta_n)^{\mathsf{T}}$。对 μ进行积分后，得到协方差 $\Sigma_\zeta = \sigma^2(I_n + \mathbf{1}\mathbf{1}^{\mathsf{T}})$ 和精度

$$\Sigma_\zeta^{-1} = \sigma^{-2}\left(I_n - \frac{\mathbf{1}\mathbf{1}^{\mathsf{T}}}{n+1}\right),$$

其中 I_n 是单位矩阵，$\mathbf{1}$ 是全一向量。等价地，中心化先验密度为

$$p(\zeta) \propto \exp\left[-\frac{\sum_j \zeta_j^2 + \sum_{j<k}(\zeta_j - \zeta_k)^2}{2\sigma^2(n+1)}\right].$$

这是完全图上的 (1) 中的图先验，其中 $a_j = w_{jk} = \{\sigma^2(n+1)\}^{-1} = 1/9$。独立双站点分析 (I2) 对 z_1 和 z_2 使用相互独立的 $N(\log 3,\ 1。125)$ 先验。

记作 $\vartheta_j = (z_j,\ \gamma_j)$，并令 $g_j(\ f \mid \vartheta_j)$ 表示通过期望传播得到的归一化高斯重要性密度。对于 32 个独立抽样 $f_{(m)j} \sim g_j(\cdot \mid \vartheta_j)$，估计量为

$$\widehat{\ell}_j(\vartheta_j, u_j) = \frac{1}{32}\sum_{m=1}^{32} \frac{p(y_j \mid f_j^{(m)})\,\phi_{K_j}(f_j^{(m)})}{g_j(f_j^{(m)} \mid \vartheta_j)},$$

其中 $p(y_j \mid f) = \prod_r \Phi(y_{jr} f_r)$，$\phi_{K_j}$ 是协方差为 K_j 的中心高斯密度，而 u_j 包含用于生成这些抽样的独立标准正态输入（Filippone & Girolami, 2014）。高斯重要性密度具有完整支持，因此该估计量非负，且其期望为边际似然。由此得到的伪边际目标以协方差参数后验为其边际分布（Andrieu & Roberts, 2009）。

固定独立参数提议为

$$q_j(\vartheta) = 0.9\, t_5(\vartheta; \mu_j^q, 1.8 S_j) + 0.1\, \phi_2\{\vartheta; (\log 3, 0)^{\mathsf{T}}, \mathrm{diag}(1.125, 1)\},$$

其中 t_5 是具有五个自由度和所述尺度矩阵的归一化二元 Student 密度，ϕ_2 是二元高斯密度，而 μ_j^q，S_j 是保存的经验均值和协方差。这些常数在采样前固定，并在 specifications/landmine/proposal.json 中提供。完整的归一化混合密度会进入 Hastings 比率。

考虑位点 j 处从 $\vartheta_j = (\log 3 + x,\ \gamma)$ 到 $\vartheta_j' = (\log 3 + y,\ \gamma')$ 的更新，其中 $x,\ y$ 是旧的和提议的中心化对数长度尺度。对于正的旧估计值和提议估计值，定义

$$A_t = \log \widehat{\ell}_j' - \log \widehat{\ell}_j - \tfrac{1}{2}\{(\gamma')^2 - \gamma^2\} - \frac{n(y^2 - x^2)}{2\sigma^2(n+1)} + \log q_j(\vartheta_j) - \log q_j(\vartheta_j'),$$

$$c_t = \frac{y - x}{\sigma^2(n+1)},$$

其中 t 是更新索引，$\widehat{l}_j$，$\widehat{l}_j'$ 是完整的旧似然估计和提议似然估计。对数接受比为 $\Lambda_t = A_t + c_t \sum_{k \neq j} \zeta_k$，其中邻居值取自更新 t 之前的即时状态。

对于未解决的邻居，取其最新已解决前驱以及其中所有固定提议的中心化对数长度尺度的区间包络。每个接受的值都是这些提议之一，而每次拒绝都会保留更早的值，因此该区间包含所需的历史值。对于已解决的历史值，使用单元素集合，并对每个邻居区间写作 $[\underline{\zeta}_{kt}, \overline{\zeta}_{kt}]$。然后

$$\underline{\Lambda}_t = A_t + \sum_{k \neq j} \min\{c_t \underline{\zeta}_{kt}, c_t \overline{\zeta}_{kt}\}, \qquad \overline{\Lambda}_t = A_t + \sum_{k \neq j} \max\{c_t \underline{\zeta}_{kt}, c_t \overline{\zeta}_{kt}\}$$

界定参考对数接受比率。在整个比较过程中，应用式 (7) 中的接受/拒绝检验，始终保留相同的提议、辅助输入和均匀阈值。

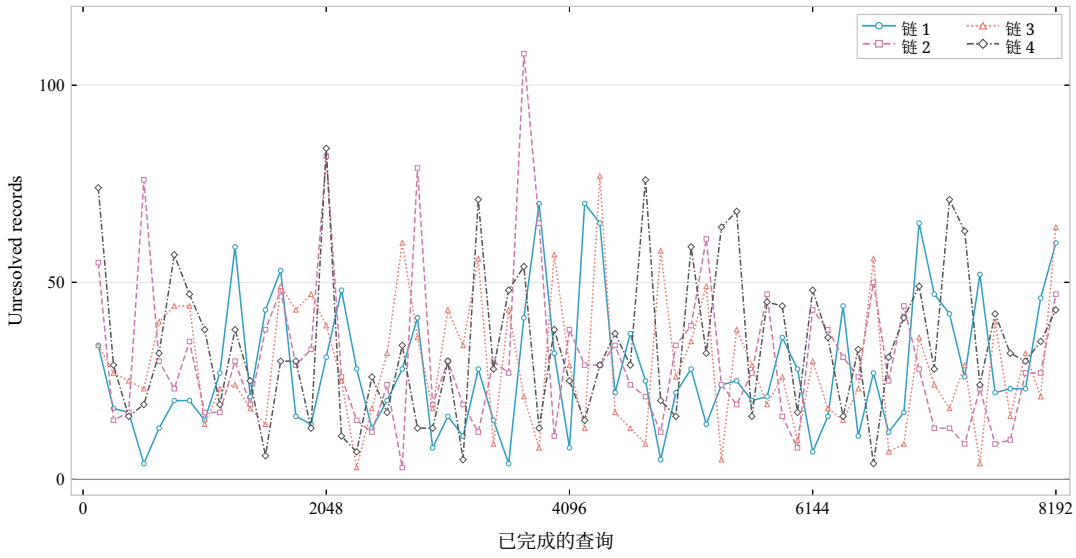


图 S2. 共享 15 点 Landmine 模型进行伪边际抽样期间的未解析历史。每进行 15 次随机扫描更新后，位点 1 和 2 会被联合查询一次。横轴表示已完成的联合查询次数 q。纵轴表示 $15q - C(q)$，其中 $C(q)$ 统计由查询 q 解析的计划更新次数。四条曲线分别跟踪一条链的 8192 次查询，其中包括前 2048 次预热查询。每隔 128 次查询记录一个点，并用线段连接。
替代文本：四条链中的未解析记录数均反复升降。在最终查询时，每条链的记录数范围为 43 至 64。

每条缓存记录都包含 (ϑ_j, u_j, l_j)，且每个查询都会先解析自己的前驱。接受时会存储完整的提议记录，而拒绝时则保留完整的旧记录。固定一个有限更新计划、一个有限有序查询列表、具有正似然估计的初始记录，以及共同的随机输入。使用精确算术，并假设每条候选记录都能在有限时间内生成和评估。每次递归调用都会请求一条更早的记录，且每次未决比较都会解析一个先前尚未解决的邻居。一旦所有所需邻居都得到解析，界限便会重合，检验也会判定是否更新，包括阈值相等的情况。提议的估计值为零时会立即拒绝。在缓存记录正确的前提下，历史区间包含参考参数，且每个证书都给出参考决策。对更新时间进行归纳，可证明算法会终止，且每条返回记录都与完整参考执行一致。实现采用向外舍入界，并在完全解析的区间仍无法判定时停止。

对于共享 15 点模型（H15），伪边际抽样使用四条链，每条链进行 8192 次名义扫描，每次扫描执行 15 次随机扫描更新。前 2048 次扫描被丢弃，因此每条链剩下 6144 个联合帧。图 S2 展示了四条伪边际链在连续查询后未解析记录的数量。在链 4 中，该数量在查询 2048 次到 2176 次之间从 84 降至 11。执行器在此区间内解析了此前的全部 84 条记录。查询 2176 次时剩余的 11 条未解析记录属于查询 2048 次之后的更新。

辅助增广采样器将潜向量写作 $f_j = L_j \nu_j + b_j \mathbf{1}$，其中 ν_j 具有标准正态先验，$b_j \sim N(0,\ 4)$，且 $L_j L_j^{\top}$ 是不含常数截距项的协方差。对 ν_j 和 b_j 进行十对椭圆切片更新后，再对 $(\zeta_j,\ \gamma_j)$ 进行联合对称 Metropolis–Hastings 更新（Murray & Adams，2010；Murray 等人，2010）。站点按系统方式访问

。我们在单独的试运行中调整参数提议尺度，每条链的每个位点使用一个尺度。试运行包括 256 次调节扫描，目标接受率为 0.234，随后进行 2048 次扫描并固定尺度。报告的四条链从试运行结束时的状态开始，并使用这些固定尺度。试运行抽样不纳入报告样本。每条链运行 49152 次扫描，并使用相同的转移核额外运行 32768 次。丢弃合并后的 81920 次扫描中的前 8192 次后，每条链剩余 73728 帧。I2 使用四条链，每条链进行 8192 次系统二位点扫描，其中丢弃 2048 次。它使用相同的固定参数提议，并在接受比率中采用 I2 先验。

S4.4。诊断与蒙特卡罗误差

我们采用 Vehtari 等（2021）的秩标准化折叠分链诊断。每条链都被分成两个等长的部分。相同的值取其平均秩。我们将每个合并秩 r 映射为正态得分 $\Phi^{-1}\{(r-3/8)/(S+1/4)\}$，其中 S 是跨分链的保留抽样值数量，Φ^{-1} 是标准正态分布函数。报告的 R 取秩抽样值与其相对于合并中位数的秩绝对偏差中的较大值。主体有效样本量使用秩标准化抽样值。尾部有效样本量取 5% 和 95% 经验分位数处指示函数对应的较小估计值。自相关估计使用初始正配对和单调配对和。该实现将有效样本量限制为保存的抽样值数量。

对于均值，断层成像分析报告原始尺度有效样本量对应的标准误估计值与四条链均值的标准误中的较大者。Landmine 还包括批长度为 128 和 256 的非重叠批均值估计，并取全部四个估计值中的最大值。该规则同样适用于方向指示变量。对于分位数概率p，我们根据保留抽样值中小于或等于经验 p分位数的指示函数，估计有效样本量 n_e。按照 Vehtari 等（2021）的方法，我们计算 $\mathrm{Beta}\{n_e p+1, n_e(1-p)+1\}$ 的 16% 和 84% 分位数。我们将这两个概率映射到保留抽样值的尺度上。断层成像使用界定次序统计量，而 Landmine 使用线性经验分位数插值。上下映射值之差的一半用于估计蒙特卡罗标准误。这些是条件于所选统计目标和数值目标的估计模拟误差。

表 S6 和表 S7 分别报告了断层成像和 Landmine 的链诊断。表 S8 报告了 Landmine 对比量 d 的蒙特卡罗标准误估计值。

对于 Landmine，我们对 $R \le 1.01$ 并使用以下阈值：$\widehat{d}$，主体有效样本量和尾部有效样本量均至少为 400。我们要求 d 的后验均值以及 $d>0$ 的后验概率的蒙特卡洛标准误估计值不超过 0。02。 d 的每个区间端点的蒙特卡罗标准误估计值必须不超过 0。03。任何 $\widehat{}$ 链中恒定的方向指示变量也会被标记。在断层成像中，我们检查 $R<1$。01，以及条件区域均值、区域抽样和双组织比例的主体有效样本量是否至少为 400。区域平均值后验均值的蒙特卡罗标准误估计值必须不超过估计均值的 0。5%。对于 95% 等尾区间的每个端点，我们要求指示变量有效样本量至少为 400，且蒙特卡洛标准误估计值不超过区间宽度的 5%。这些检查并不构成对偏离平稳性程度的界限。

S5. 配套记录

配套源代码包包含固定的模型配置、定义数值规则和诊断的代码摘录、记录的计时结果以及文件标识清单。这些文件标识了本补充材料中使用的量。它们记录了所报告的分析；

表 S6。 断层扫描链诊断

数量	$\widehat{R}$	总体有效样本量	尾部有效样本量
条件区域均值	1.007	535	961
区域抽样	1.000	7113	7014
双组织比例	1.009	497	991

ESS 表示有效样本量。双组织比例是被分配到双组织区室模型的区域体素所占的比例。四条链中的每一条均保留 1792 个查询。对于 V_T 的区域平均值，95%等尾后验区间在其下端点和上端点处估计的蒙特卡罗标准误分别为 $1。93 \times 10^{-4}$ 和 $1。57 \times 10^{-4}\ \mathrm{mL\ cm^{-3}}$。这些标准误分别为区间宽度的 0。87% 和 0。71%。

替代文本：条件区域均值的主体有效样本量为 535，区域抽样的主体有效样本量为 7113，双组织比例的主体有效样本量为 497。三项收敛统计量都接近 1。

表 S7。*Landmine* 链诊断

模型	采样器	数量	$\widehat{R}$	总体有效样本量	尾部有效样本量
H15	PM	z_1	1.000	7603	7534
H15	PM	z_2	1.000	7505	8072
H15	PM	d	1.000	8273	11546
H15	AA	z_1	1.001	2284	4064
H15	AA	z_2	1.002	1996	4534
H15	AA	d	1.001	2384	4745
I2	PM	z_1	1.000	12451	14376
I2	PM	z_2	1.000	14028	14718
I2	PM	d	1.000	13177	16449

H15 表示共享 15 位点模型，I2 表示独立双位点模型。PM 和 AA 分别表示伪边际抽样和辅助变量增广抽样。ESS 表示有效样本量。对比量 $d = z_1 - z_2$ 比较位点 1 和 2 的对数长度尺度。所有分析均使用四条链。每条 H15 PM 或 I2 PM 链保留 6144 帧；每条 H15 AA 链保留 73728 帧。

替代文本：在两个模型中，所有报告的收敛统计量均介于 1.000 和 1.002 之间。在共享模型中，伪边际抽样对所有

三个量。

该软件包不会启动新的后验模拟。清单中的代码库路径和源代码哈希值可用于在公开发布以供复现之前，将其与完整项目进行匹配。公共数据集标识符见参考文献。

表 S8. 地雷数据集蒙特卡洛标准误

模型	采样器	Mean	方向概率	下端点	上端点
H15	PM	0.00868	0.00513	0.0188	0.0212
H15	AA	0.0147	0.00836	0.0180	0.0254
I2	PM	0.00971	0.00648	0.0139	0.0241

H15 表示共享 15 位点模型，I2 表示独立双位点模型。PM 和 AA 分别表示伪边际抽样和辅助变量增广抽样。对比量 $d = z_1 - z_2$ 比较位点 1 和 2 的对数长度尺度。各列给出了其后验均值、取正值的概率以及 95%等尾后验区间两个端点的蒙特卡洛标准误估计值。对于每个 H15 汇总量，我们通过对 PM 和 AA 估计值各自的标准误平方求和后开平方，来估计两者之差的蒙特卡洛标准误。

AI文本：在三项分析中，位点对比量为正的概率的蒙特卡罗标准误范围为 0.00513 至 0.00836，区间端点的蒙特卡罗标准误范围为 0.0139 至 0.0254。