arXiv:2603.07887v1 [cs.LG] 9 Mar 2026

拒绝、重采样、重复：理解语言模型推理中的并行推理

NoahGolowich Fan Chen DhruvRohatgi RaghavSinghal微软研究院麻省理工学院麻省理工学院纽约大学nzg@cs.utexas.edu fanchen@mit.edu drohatgi@mit.edu singhal.raghav@gmail.comCarlesDomingo-Enrich Dylan J. Foster Akshay Krishnamurthy微软研究院微软研究院微软研究院carlesd@microsoft.com dylanfoster@microsoft.com akshaykr@microsoft.com

2026年3月10日

摘要

聚合并剪枝多个样本的推理时方法已成为引导大语言模型的一种强大范式，但我们尚未对其准确性与成本之间的权衡形成任何有原则的理解。本文借助粒子滤波算法（如序贯蒙特卡洛（SMC））的视角，提出了一条严格研究此类方法的路径。给定一个基础语言模型和一个过程奖励模型，该模型用于估计预期终端奖励，我们提出以下问题：在给定一定数量的过程奖励评估的情况下，我们能够多准确地从目标分布中进行采样？从理论上看，我们确定了：（1）能够为序贯蒙特卡洛提供非渐近保证的简单判据；（2）对序贯蒙特卡洛的算法改进；以及（3）所有粒子滤波方法所面临的一项根本性限制。在实证方面，我们证明，尽管理论判据并不一定决定其最终准确性，但确实有效地制约了序贯蒙特卡洛的采样误差，这表明可能需要超越采样的理论视角。

1 引言

近期关于大语言模型（LLM）的研究表明，推理时干预能够引导模型生成质量更高的输出，因而具有重要价值。例如，现有研究尝试结合使用来自大语言模型的并行生成（Madaan 等人，2025年；Brown 等人，2024年）、对此类生成结果进行聚合的技术（Wang 等人，2023年；Fu 等人，2025年；Zhao 等人，2025年），以及将其剪枝或压缩，以便纳入上下文窗口供后续模型调用的方法（Yang 等人，2024年；Wu 等人，2025年）。即使不进行额外训练，这类干预也能显著提升模型在数学推理和问答等具有挑战性的任务上的性能。然而，这些方法大多是临时拼凑的，因为我们缺乏统一的理论框架，既无法解释不同推理时干预的益处，也无法指导算法设计，使其朝着最有效的方法发展。

在本文中，我们提出以下问题：

我们能否以有原则的方式分析聚合并行生成和对其进行剪枝等推理时干预？

为此，我们研究借助过程奖励模型进行引导式大语言模型生成（Wang 等人，2025a；Puri 等人，2025b）这一范式，其中过程奖励模型会对部分生成内容进行数值评分。在这一范式中，聚合和剪枝可通过粒子滤波算法自然实现，包括广受欢迎的序贯蒙特卡洛（*SMC*）方法。随着词元的生成，这些算法利用过程奖励模型自适应地剪枝和复制多个粒子（即基础模型生成的部分内容）。与 Best-of-N等较简单的干预方法相比，这些策略在实证中具有稳健优势；例如，参见图 1或Puri 等人（2025b）。

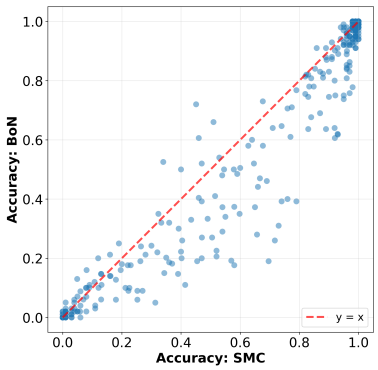


图 1：序贯蒙特卡洛（SMC）（使用 N 个粒子）与 Best-of-N 在 Math500 问题上的性能对比；这里取 $N = 32$。每个点代表一个不同的问题；因此，在大多数问题上，使用 N 个粒子的 SMC 的性能优于 Best-of-N。详见第4.5节。

尽管粒子滤波方法已广泛应用于从后验推断（Pitt 和 Shephard，1997年）、机器人定位（Montemerlo 等人，2002年），到引导大语言模型（Lipkin 等人，2025年；Lew 等，2023；Loula 等人，2025年；Grand 等人，2025年；Puri 等人，2025a；Zhao 等人，2024年；Feng 等人，2024年）和扩散模型（Skreta 等人，2025年；Singhal 等人，2025年）等各种任务，但针对这类方法建立的非渐近保证仍然很少。本文致力于应对这一挑战，对序贯蒙特卡洛进行通俗易懂的分析，并指出更复杂的粒子滤波方法的优势与局限。

迈向推理时干预理论。越来越多的推理时干预不仅利用模型进行生成和评分，还执行拼接（Fu 等，2025年）或摘要生成（Yu 等，2025年）等更复杂的操作。尽管我们的设定未直接涵盖这些操作，但我们认为，本研究是构建理解（并改进）此类干预的框架所必需的第一步。

1.1 设定：具有不完美过程奖励的引导式生成

对于固定的提示，语言模型在某个集合 H 动作上定义了序列的分布 $\pi_{\mathsf{ref}} \in \Delta(\mathcal{A}^H)$。[1] 可以将 $\mathcal{A}$视为词元的空间，或更一般地视为词元的分块。给定序列 $a_{1:h}$，我们可以从 $\pi_{\mathsf{ref}}(a_{h+1} = \cdot \mid a_{1:h})$ 中高效采样，从而能够从 π_{ref}进行自回归采样。

我们将引导语言模型获得某种期望奖励的任务建模为一个采样问题，这与近期关于大语言模型推理和后训练的研究一致（Rohatgi 等人，2025；Geuter 等人，2025；Foster 等人，2025；Xiong 等人，2024；Zhu 和 Lu，2026）。给定对奖励函数 $r^\star : \mathcal{A}^H \to \mathbb{R}_{\geq 0}$ 的查询权限，我们旨在从倾斜分布 $\pi_H^\star(a_{1:H}) \propto \pi_{\mathsf{ref}}(a_{1:H}) r^\star(a_{1:H})$ 中采样。例如，在数学推理的场景中，$r^\star(a_{1:H}) \in \{0, 1\}$ 用于衡量 $a_{1:H}$ 是否正确解答了提示。

不完美的过程奖励。遵循Rohatgi 等人(2025年)，我们从理论上将过程奖励模型（PRM）形式化为一个近似价值函数。给定一个 $a_{1:h}$ 任意长度的序列，PRM 会提供$\widehat{\ }$对 $V(a_{1:h}) \in \mathbb{R}_{\geq 0}$ 的期望奖励的 $a_{1:h}$ 在 π_{ref}条件下的估计，即 $V^\star(a_{1:h}) := \mathbb{E}_{a'_{1:H} \sim \pi_{\mathsf{ref}}}[r^\star(a'_{1:H}) \mid a'_{1:h} = a_{1:h}]$。

在实践中，人们已投入大量努力，专门训练Transformer模型使其成为过程奖励模型（Lightman 等人，2023年；Wang 等人，2024年，2025b；Zhang 等人，2025年），以用于某些问题，例如数学推理。如果无法获得这种适配于所需奖励函数 $r^\star$ 的微调后的过程奖励模型，则还有其他几种常用的$\widehat{\ }$选择可用于 V——例如，LLM 作为评判者（Stephan 等人，2024年）或 LLM 对数概率（Karan 与 Du，2025年；Lew 等人，2023年）；另见第 4 节。

[1]为简化起见，我们假 设语言模型生成的文本 长度固定。

在这一设定中，关键的算法挑战在于过程奖励模型并不完美。为应对这一挑战，已提出多种并行与串行方法，包括序贯蒙特卡洛（Puri 等人，2025b）、块拒绝采样（Mudgal 等人，2024年）以及回溯（Botta 等人，2025年）。迄今为止，可证明保证仅已知适用于回溯（Rohatgi 等人，2025年）——一种本质上为串行的方法。我们问：在能够访问 $\sim\pi_{\mathsf{ref}}$ 和 V（如上所述）时，哪些结构性质能够使针对 $\pi_H^\star$ 的近似采样借助粒子滤波等高效并行方法得以实现？

1.2 理论贡献：对大语言模型推理中粒子滤波方法的原则性分析

贡献一：SMC 成功的简单判据。我们的首个结果指出了使 SMC 获得成功的两个关键性质。对于 $h \in [H]$，令 $\pi_h^\star, \widehat{\pi}_h \in \Delta(\mathcal{A}^h)$ 表示真实和近似的中间目标分布，分别定义为 $\pi_h^\star(a_{1:h}) \propto \pi_{\mathsf{ref}}(a_{1:h})V^\star(a_{1:h})$ 和 $\widehat{\pi}_h(a_{1:h}) \propto \pi_{\mathsf{ref}}(a_{1:h})V(a_{1:h})$。接下来我们证明：

定理 1.1 (定理 3.2的推论)。假设以下条件成立。
(1) 有界的动作级覆盖：对所有 h，$\frac{\pi_h^\star(a_{1:h})}{\pi_{\mathsf{ref}}(a_{1:h})} \le C_{\mathsf{act}}$。*(2)* $\pi_{h+1}^\star$ 有界的 χ^2-散度（根据 $\widehat{V}$ 与 $V^\star$ 之间的误差控制，参见事实 *A.2*）：$D_{\chi^2}(\pi_h^\star \,\|\, \widehat{\pi}_h) \le C_{\chi^2}$ 对所有 h。

然后是序贯蒙特卡洛 (SMC) 使用 N 来自分布的粒子样本 $\widehat{\mu}$ 满足 $D_{\mathsf{TV}}(\widehat{\mu},\ \pi^\star) \le \sqrt{\frac{H^2 C_{\mathsf{act}}(C_{\chi^2}+1)}{N}}$.
这一界限强化了 Rohatgi 等人 (2025年提出的基于回溯的算法 VGB 的保证。此外，序贯蒙特卡洛的并行运行时间为 $O(H)$ 而 VGB 的运行时间为 $\Omega(H^2)$（且本质上是顺序执行的）。更多讨论见 附录 A 。

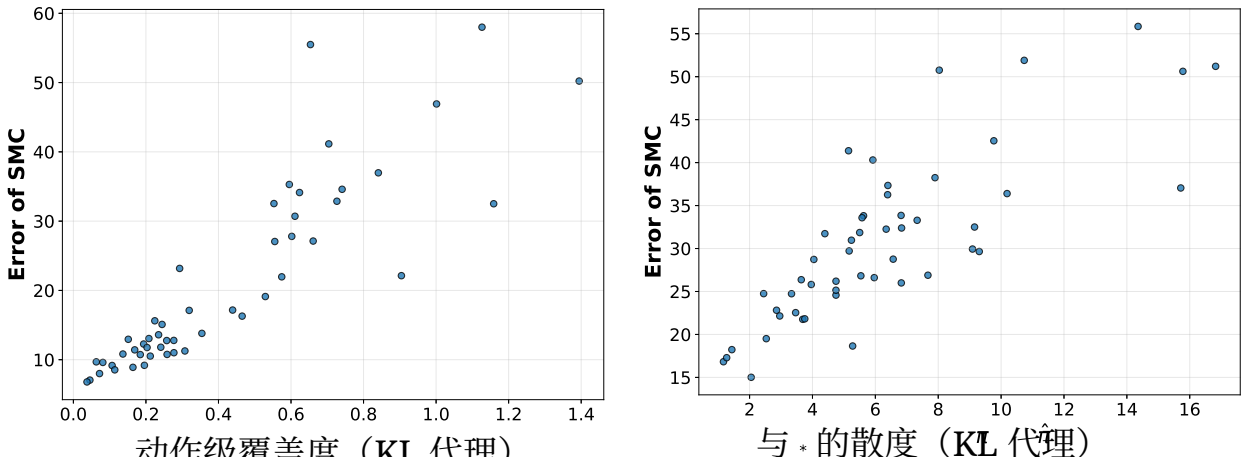
备注 *1.2*（与序贯蒙特卡洛改进方法相关文献的联系）。有趣的是，上述两项关键性质与以往致力于改进序贯蒙特卡洛算法的各种研究成果相当吻合。两个条件中后者（即 V 近似于 $V^\star$）与序贯蒙特卡洛扭曲目标分布方面的研究有关。具体而言，上文引入的分布 $\widehat{\pi}_h$ 通常称为（中间）目标分布，以往的研究探讨了如何选择（或学习）V，从而使 $\widehat{\pi}_h$ 尽可能接近 $\pi_h^\star$，对所有$h \in [H]$ 均如此（*Guarniero* 等人，*2017*；*Heng* 等人，*2017*；*Zhao* 等人，*2024*年；*Lawson* 等人，*2022*），鉴于定理 *1.1*，这一目标十分自然。

此外，第一个条件（关于动作级覆盖）与大量关于改变提议分布的文献相关 $\pi_{\mathsf{ref},\ h}$ 以更好地逼近局部最优提议分布 （*Naesseth* 等人，*2019*年；*Doucet* 等人，*2000*年），其被定义为分布 $\widehat{\pi}_{h+1}(a_{h+1} = \cdot \mid a_{1:h})$。若事件$\widehat{\pi}_{h+1} \approx \pi_{h+1}^\star$发生，则局部最优提议分布的动作级覆盖接近 1，即最小的可能值。因此，只要 V 足够准确，修改$\pi_{\mathsf{ref},\ h}$ 以逼近局部最优提议，便是降低动作级覆盖的一种方法。[2] 摘要生成，我们的结果有助于统一现有关于序贯蒙特卡洛变体的文献，表明它们实际上是在最小化关键量 定理 ***1.1***。

贡献二：超越序贯蒙特卡洛。 序贯蒙特卡洛之外的粒子滤波算法能否获得更强的保证？首先，我们证明，如果 $\widehat{\pi}_h$ 接近 $\pi_h^\star$，且在更强的 L_∞ 意义下，则有可能实现 指数衰减 于采样误差（而非多项式衰减，如 定理 1.1中那样），方法是将序贯蒙特卡洛包裹在外层拒绝采样循环中（定理 3.6）。

其次，我们观察到，即使过程奖励模型是完美的，因而 $C_{\chi^2} = 0$，序贯蒙特卡洛仍需要$\Omega(\sqrt{H})$个粒子才能达到非平凡的采样准确性（命题 C.8），尽管针对这一特殊情形存在一种简单的单粒子算法（Rohatgi 等人，2025年）。我们证明，一种改进算法——带拒绝采样的序贯蒙特卡罗（*SMC-RS*，算法 *2*）——可以避免这一病态现象，并恢复定理 1.1 以及单粒子保证（定理 3.8）。

[2]事实上，将 $\widehat{\pi}_{h+1}$ 用作提议分布，并在序贯蒙特卡洛中适当调整权重，可得到一种收敛性质更优的变体（见 第 3.3 节 和 附录 D.3）。



（a）动作级覆盖度。 （b） $D_{\mathsf{KL}}(\pi_h^\star \,\|\, \widehat{\pi}_h)$ 。

图 2：我们关于提示切换任务的理论的实证验证（参见第 4 节）。（a）我们在许多提示上改变$\widehat{\ }$动作级覆盖度（通过 KL 散度代理进行衡量），同时保持 $V = V^\star$固定，并观察到动作级覆盖度可以预测序贯蒙特卡洛的采样误差。（b）我们固定 $\pi_{\mathsf{ref}} = \pi^\star$ （从而使动作级覆盖度保持不变），并观察到 $D_{\mathsf{KL}}(\pi_h^\star \,\|\, \widehat{\pi}_h)$ 可以预测序贯蒙特卡洛的采样误差。

贡献三：粒子滤波的局限。序贯蒙特卡洛（以及我们的改进）的一个关键局限在于，$\widehat{\ }$即使当 π_{ref} 具有有界的动作级覆盖（即 $C_{\mathsf{act}} = O(1)$），且近似价值函数 V 与真实值相差不超过常数因子（因而 $C_{\chi^2} = O(1)$）时，我们的算法仍需要维持$\mathrm{poly}(H)$个粒子，因而需要超线性的工作量。在定理 3.9 中，我们证明，要完全避开这一下界，就需要前瞻：任何短视方法至少需要$\Omega(\log\ H/\log\log\ H)$ 个粒子（定理 3.9）。弥合这一差距仍是开放问题。

补充说明：回溯是一种粒子滤波形式。在附录 E中，我们展示了粒子滤波算法与基于回溯的方法 VGB（Rohatgi 等人，2025年）之间一个有趣的联系——此前，VGB 是唯一已知能够在过程奖励模型错误下保持稳健的大语言模型推理时引导算法。VGB 是一种（单粒子）马尔可夫链方法，它会生成词元，但有时也会删除最近生成的词元。我们证明，VGB 的执行过程可以与某种粒子滤波方法的执行过程耦合，这或许能为理解这两种方法的优势提供新的视角。

1.3 实证贡献：理论能否预测大语言模型中序贯蒙特卡洛的性能?

我们通过实证研究，探讨定理 1.1 所确定的量是否确实能够预测序贯蒙特卡洛在大语言模型采样问题上的性能。首先，我们考察一项简单的“提示切换”任务：基础模型 π_{ref} 对应于给定某个提示 $\mathrm{p}_{\mathsf{ref}}$时从大语言模型中采样，而 $\pi^\star$ 对应于给定另一个提示 $\mathrm{p}^\star$ 时从该大语言模型中采样$\widehat{\ }$（详情见第 4 节）。通过改变 $\mathrm{p}_{\mathsf{ref}}$、$\mathrm{p}^\star$和过程奖励模型 V，我们$\widehat{\ }$可以控制定理 1.1 中的各个量。首先，在图 2a 中，我们考察一个数据集，其中该数据集内的所有提示均满足 $V = V^\star$ （因此 $C_{\chi^2} = 0$），但动作级覆盖各不相同；我们观察到，动作级覆盖（通过 KL 散度代理指标衡量）与序贯蒙特卡洛的采样误差之间存在很强的相关性。其次，在图 2b 中，我们考察一个数据集，其中所有提示的动作级覆盖都相同，但 $\widehat{\ }$相对于 $V^\star$ 的准确性程度各不相同；我们再次观察到，$D_{\mathsf{KL}}(\pi_h^\star \,\|\, \widehat{\pi}_h)$ 与序贯蒙特卡洛的采样误差之间存在很强的相关性。（我们报告 KL 散度，而不是 χ^2 散度，因为估计后者所需的样本复杂度高得难以承受。）

在数学推理任务上评估序贯蒙特卡洛。最后，我们考察序贯蒙特卡洛在包括 AIME 和 Math500 在内的具有挑战性的数学推理基准上的性能。在图 1中，我们针对 Math500 中的每个问题，绘制了使用 N 个粒子的序贯蒙特卡洛的准确性，并与 Best-of-N （使用 $N = 32$）的准确性进行对比；这些量均通过对每个问题运行所需算法 100 次来估计。由于大多数点位于该直线下方

$y = x$，我们看到序贯蒙特卡洛相较于 Best-of-N 在所有问题上均以一致的方式提升了性能，从而扩展了Puri 等人（2025b）；Chatziveroglou（2025年）中的先前观察；他们观察到，序贯蒙特卡洛相较于 Best-of-N 在平均意义上于此类基准上有所提升。尽管已有强有力的证据表明序贯蒙特卡洛的优越性，正如我们在第4.5节中进一步讨论的那样，过程奖励模型 V（以卡方散度的估计来衡量，该散度介于 $\widehat{\pi}_h$ 与 $\pi_h^\star$之间）似乎与性能并不呈正相关，因为通常更大的散度会导致更高的准确性。我们将构建一个能够刻画此类基准上性能的更精细框架，留作一个引人深思的开放问题。

路线图。 第 2 节 介绍了我们理论结果的形式化设定。 第3节 包含我们的主要理论结果，而 第 4 节 包含我们的实验结果。

2 预备知识

我们的理论适用于一种推广了 第1.1节的设定。它应当能够涵盖扩散模型引导等设定，尽管我们并未具体研究这类设定。

令 $\mathcal{X}_0$、$\mathcal{X}_1$、……、$\mathcal{X}_H$ 为任意集合，并令 $\mathcal{X} := \mathcal{X}_0 \sqcup \mathcal{X}_1 \sqcup \cdots \sqcup \mathcal{X}_H$ 为它们的不交并。令 $\pi_{\mathsf{ref}} = (\pi_{\mathsf{ref},h})_{h=0}^{H-1}$ 为一组转移核 $\pi_{\mathsf{ref},h} : \mathcal{X}_h \to \Delta(\mathcal{X}_{h+1})$。不失一般性，假设 $\mathcal{X}_0 = \{\perp\}$ 是单元素。转移核诱导出一个马尔可夫链 $x_0 \to x_1 \to x_2 \cdots \to x_H$，其中 $x_{h+1} \sim \pi_{\mathsf{ref},h}(\cdot \mid x_h)$ 对于 $0 \le h \le H-1$成立；在上下文明确时，我们省略下标 h。我们用$x_{1:H} \sim \pi_{\mathsf{ref}}$ 表示从该马尔可夫链中抽取的样本。令 $\pi_h \in \Delta(\mathcal{X}_h)$为 x_h所诱导的边缘分布。

备注 2.1（特殊情况：自回归生成）。 在第*1*节讨论的大语言模型特例中，我们有 $\mathcal{X}_h = \mathcal{A}^h$，而 $\pi_{\mathsf{ref},\,h}(\cdot \mid a_{1:h})$ 是转移核，它会采样 a_{h+1} （给定 $a_{1:h}$）并将其附加到 $a_{1:h}$。

与 第1节中的设定一样，给定终端奖励函数 $r^\star : \mathcal{X}_H \to \mathbb{R}_{\ge 0}$，价值函数 $V^\star : \mathcal{X} \to \mathbb{R}_{\ge 0}$对于 $x \in \mathcal{X}_h$，定义为 $V^\star(x) = \mathbb{E}_{x_{1:H} \sim \pi_{\mathsf{ref}}}[r^\star(x_H) \mid x_h = x]$。注意，对于 $V^\star(x_H) = r^\star(x_H)$， $x_H \in \mathcal{X}_H$。
对于 $h \in [H]$，定义分布 $\pi_h^\star \in \Delta(\mathcal{X}_h)$ 为 $\pi_h^\star(x) \propto \pi_h(x) \cdot V^\star(x)$。可以验证，分布 $\pi_h^\star$ 由具有如下核的马尔可夫链诱导：

$$\pi^\star(x_{h+1} \mid x_h) := \frac{\pi_{\mathsf{ref},h}(x_{h+1} \mid x_h) V^\star(x_{h+1})}{V_h^\star(x_h)}, \tag{1}$$

如 $\sum_{x_h} \pi_h^\star(x_h)\pi^\star(x_{h+1} \mid x_h) = \pi_{h+1}^\star(x_{h+1})$。[3]

部分奖励模型。 我们以更一般设定下自然的方式扩展 PRM 的定义：将 PRM 记为 $V : \mathcal{X} \to \mathbb{R}_{\ge 0}$，并且对 $x_H \in \mathcal{X}_H$ 满足 $V(x_H) = r^\star(x_H)$。我们定义 $\widehat{V}(x) = \mathbb{E}_{x \sim \pi_h}[V(x)]$，其中 $h \in [H]$，并令 $\widehat{\pi}_h \in \Delta(\mathcal{X}_h)$ 由 $\widehat{\pi}_h(x) = \pi_h(x) \cdot V(x)/Z_h$ 定义。

3 粒子滤波方法的理论分析

3.1 序贯蒙特卡洛成功的简单判据

第一项标准是有界的动作级覆盖：

假设 3.1. 存在参数 $C_{\mathsf{act}} \ge 1$，使得对于所有 $h \ge 0$, $V_{h+1}^\star(x_{h+1}) \le C_{\mathsf{act}} V_h^\star(x_h)$ 以及任意 $x_{h+1} \in \mathrm{supp}(\pi_{\mathsf{ref}}(\cdot \mid x_h))$。

我们将此条件称为动作级覆盖，因为$\frac{V^\star(x_{h+1})}{V^\star(x_h)} = \frac{\pi^\star(x_{h+1}|x_h)}{\pi_{\mathsf{ref},h}(x_{h+1}|x_h)}$ 恰好是条件分布 $\pi^\star$ 与$\pi_{\mathsf{ref},h}$ 之间的密度比。[4]

[3]我们指出，以下记法略有滥用：我们用不带下标的 $\pi^\star$ 表示公式 (1) 中的转移核，并用带下标的 $\pi_h^\star$ 表示相应的中间分布。 [4]我们还指出，假设 3.1 可以放宽为平均意义（参见定理 C.6）。

算法 1 序贯蒙特卡洛

输入： 转移核 π_{ref}，价值函数 V。**参数：** 粒子数 $N \geq 1$。初始化 $\widehat{\nu}_0 = \delta_\perp$。**对于** $h = 1, \dots, H$ **执行**定义 $W_h := \prod_{i=1}^{h} \frac{W_i}{N}$。**输出选项1：** 输出 $x_H \sim \widehat{\nu}_H$。

Sample N particles $\tilde{x}_{h-1}^1, \dots, \tilde{x}_{h-1}^N \overset{\text{i.i.d.}}{\sim} \widehat{\nu}_{h-1}(\cdot)$.

Generate $x_h^i \sim \pi_{\mathsf{ref},h-1}(\cdot \mid \tilde{x}_{h-1}^i)$ independently for $i \in [N]$.

Define weights $w_h^i = \frac{\widehat{V}(x_h^i)}{\widehat{V}(\tilde{x}_{h-1}^i)}$.

Define weighted empirical measure $\widehat{\nu}_h := \frac{1}{W_h} \sum_{i=1}^N w_h^i \delta_{x_h^i}$, where $W_h = \sum_{i=1}^N w_h^i$.

输出选项 2： 以 $\min\left\{ \frac{\widehat{W}_H}{2 \, \widehat{V}_\infty \mathsf{C}(\perp)}, 1 \right\}$ 的概率输出 $x \sim \widehat{\nu}_H$，否则重新启动算法。

第二项判据要求 χ^2-散度在 $\pi_h^\star$ 与 $\widehat{\pi}_h$ 之间对所有 h均有界。定量而言，这两项判据为序贯蒙特卡洛提供了如下保证，并可直接推出 定理 1.1：

定理 3.2。 在 假设 *3.1*下，具有 N 个粒子的序贯蒙特卡洛达到

$$D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H], \pi_H^\star) \leq \sqrt{\frac{C_{\mathsf{act}}}{N}} \left(H + \sum_{h=1}^{H-1} \sqrt{D_{\chi^2}(\pi_h^\star \,\|\, \widehat{\pi}_h)} \right).$$

我们指出，要对 $D_{\chi^2}(\pi_h^\star \,\|\, \widehat{\pi}_h)$ 进行界定，只需 平均情形下 的 V 与 $V^\star$ (事实 A.2) 足够接近。序贯蒙特卡洛的现有理论（Schweizer，2012年；Zhu 和 Lu，2026）可以得到类似的界，但要求最坏情形下 足够接近：

假设 3.3。 存在参数 $\mathsf{C}_\infty \geq 1$，使得 $\mathsf{C}_\infty^{-1} \leq \frac{V^\star(x)}{\widehat{V}(x)} \leq \mathsf{C}_\infty$ 对任意 $x \in \mathcal{X}$均成立。

尽管 Schweizer(2012年) 的证明技巧在一定程度上可以放宽这一假设，但无法恢复定理 3.2 (参见附录 C.3)。此外，正如下文所示，我们可以进一步强化定理 3.2，使其能够容忍重尾误差在 V中。

3.1.1 处理尾部行为

尽管 χ^2-散度在定理 3.2 中给出了易于理解的上界，$D_{\chi^2}(\pi_h^\star \,\|\, \widehat{\pi}_h)$ 却对的尾部行为十分敏感，具体而言，是相对于 $\widehat{\pi}_h$ 而言的 $\pi_h^\star$ (等价地，V 相对于 $V^\star$的尾部行为)；在重尾误差的情况下，它可能会发散，从而使定理 3.2 变得毫无意义。

然而，序贯蒙特卡洛的目标是从一个近似正确的分布中生成样本，因此尾部行为只应体现为错误概率。这引出了以下分布接近度的概念（Chen 等人，2025年），它可以用 χ^2-散度来界定，反之则不成立——特别是，即使 $D_{\chi^2}(\pi_h^\star \,\|\, \widehat{\pi}_h) = \infty$也可以任意小。

定义 3.4。 对于参数 $M \geq 1$, $h \in [H]$，我们定义

$$\mathrm{Cov}_M(\pi_h^\star \,\|\, \widehat{\pi}_h) = \mathbb{P}_{x \sim \pi_h^\star} \left(\frac{\pi_h^\star(x)}{\widehat{\pi}_h(x)} \geq M \right).$$

定理 3.5。 在 假设 *3.1*下，具有 N 个粒子的序贯蒙特卡洛可对任意 $M \geq 1$实现以下结果：

$$D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H], \pi_H^\star) \leq H \sqrt{\frac{M C_{\mathsf{act}}}{N}} + \sum_{h=1}^{H-1} \mathrm{Cov}_M(\pi_h^\star \,\|\, \widehat{\pi}_h).$$

算法 2 带拒绝采样的序贯蒙特卡罗

输入： 转移核 π_{ref}，价值函数 V。**参数：** 初始样本量 $N \geq 1$，参数 $\eta \geq 1$。将 $\mathcal{S}_0$ 初始化为 N 份副本。**对于** $h = 1, \cdots, H$ **执行**初始化多重集合 $\mathcal{S}_h = \{\}$。**当** $|\mathcal{S}_h| < N$ **执行**采样 $x_{h-1} \sim \mathrm{Unif}(\mathcal{S}_{h-1})$。采样 $x_h \sim \pi_{\mathsf{ref},h-1}(\cdot \mid x_{h-1})$。以概率 $V_{(x_h)}\eta\widehat{V}(x_{h-1})$，更新 $\mathcal{S}_h \leftarrow \mathcal{S}_h \cup \{x_h\}$。

输出： $x \sim \widehat{\nu}_H = \mathrm{Unif}(\mathcal{S}_H)$。

让我们将定理 3.5 的保证与定理 3.2 的保证进行比较：定理 3.5 中的覆盖项 $\mathsf{Cov}_M(\pi_h^\star \,\|\, \widehat{\pi}_h)$ 扮演着定理 3.2 中卡方散度 $D_{\chi^2}(\pi_h^\star \,\|\, \widehat{\pi}_h)$ 的角色，刻画了 $\widehat{\pi}_h$ 偏离 $\pi_h^\star$ 的程度。不过，一个关键区别在于，对于 M 的任意固定取值，定理 3.5 中 $\mathbb{E}[\widehat{\nu}_H]$ 与 $\pi_H^\star$ 之间的全变差误差并不会随着 $N \to \infty$ 趋近于 0，也就是说，我们要以加性方式为质量 $\mathsf{Cov}_M(\pi_h^\star \,\|\, \widehat{\pi}_h)$ 付出代价。

3.2 超越序贯蒙特卡洛：在 L_∞-条件下实现更快收敛

下文将展示，对序贯蒙特卡洛（算法 1的输出选项 2）进行简单修改后，在最坏情况下，当 V 与 $V^\star$ 足够接近时，可实现指数级（而非多项式级）收敛（假设 3.3）。

定理 3.6。 在假设 *3.1* 和 *3.3*成立时，算法 *1* 采用输出选项 ***2*** 从满足 μ 和 $D_{\mathsf{TV}}(\mu, \pi_H^\star) \leq \delta$的分布中采样，只要 $N \geq \Omega(HC_\infty^4 C_{\mathsf{act}}^2 \log(C_\infty H/\delta))$，所需外层循环步数的上界为 $O(C_\infty^2 \log(1/\delta))$。

基本思路是在序贯蒙特卡洛外层套上拒绝采样循环。从定性上看，这与 Rohatgi 等人(2025年提出的基于回溯的方法 VGB 的收敛速率相匹配；该方法需要相同的假设。

3.3 超越序贯蒙特卡洛：近乎完美的过程奖励模型

即使当 $\widehat{V} = V^\star$（即过程奖励模型是完美的）时，定理 3.2 中的上界仍然按 $\frac{H}{\sqrt{N}}$ 缩放。事实上，这是序贯蒙特卡洛的根本局限，而非分析的人为产物：即使当 $V = V^\star$ 且 $C_{\mathsf{act}} = 2$时，序贯蒙特卡洛也无法实现 $o(1)$ 采样误差，除非至少有 $N \geq \Omega(\sqrt{H})$ 个粒子（命题 C.8）。[5] 请注意，$V = V^\star$，仅采样 $x_h \sim \widehat{\pi}(x_h \mid x_{h-1})$ 的单粒子算法便能精确地从 $\pi^\star$ 中采样，因此确实有望改进 $H/\sqrt{N}$ 的上界。

为缓解这一问题，我们引入带拒绝采样的序贯蒙特卡罗（SMC-RS，算法 2）。直观而言，SMC-RS 解决了如下问题：为了采样新的粒子，序贯蒙特卡洛会归一化加权经验测度，从而在粒子之间引入干扰。SMC-RS 则改用拒绝采样，因此无论其他粒子如何，条件分布 $x_h \mid x_{h-1}$ 都是

$$\widehat{\pi}(x_h \mid x_{h-1}) \propto \pi_{\mathsf{ref}}(x_h \mid x_{h-1})\widehat{V}(x_h).$$

这可以看作是在掌握 V 的信息后，对基础提议分布 $\pi_{\mathsf{ref}}(\cdot \mid x_{h-1})$ 的细化。特别是，当 $N = 1$时，SMC-RS 会退化为动作级拒绝采样（Yang 和 Klein，2021）；根据公式 (1)，当 $V = V^\star$ 时，它是一个精确采样器（定义见附录 A）。我们在以下动作级覆盖条件下分析 SMC-RS；从概念上看，我们认为该条件等价于假设 3.1（从形式上看，在满足假设 3.3 的 L_∞ 条件下，除以 C_∞^2 的因子外二者等价）。如果我们作出更弱的假设，即控制 $\pi^\star$ 与 $\widehat{\pi}$ 之间的卡方散度（如下方定理 3.8 所述），则需要假设 3.7，以确保算法 2 中参数 η 的有界取值足够。

[5] 我们将确定最优 H依赖性的问题留作开放问题。

假设 3.7。 存在参数 $C_{\mathsf{act}} \geq 1$，使得对于所有 $h \geq 0$、$V_{h+1}(x_{h+1}) \leq C_{\mathsf{act}} V_h(x_h)$ 以及任意 $x_{h+1} \in \mathrm{supp}(\pi_{\mathsf{ref}}(\cdot \mid x_h))$。

定理 3.8。 在 假设 *3.7*下，假设 *SMC-RS*（算法 *2*）采用 $\eta \geq C_{\mathsf{act}}$。那么，使用 N 个粒子的 *SMC-RS* 可达到

$$D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H], \pi^\star_H) \leq \frac{1}{\sqrt{N}} \sum_{h=1}^{H} \sqrt{D_{\chi^2}(\pi^\star_h \,\|\, \widehat{\pi}_h)},$$

并且其期望时间复杂度为 $O(NH\eta)$。

特别地，假设 $\widehat{V}$ 是近似准确的，即对于所有 $h \in [H]$，$D_{\chi^2}(\pi^\star_h \,\|\, \widehat{\pi}_h) \leq \epsilon^2$。那么 SMC-RS 可达到

$$D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H], \pi^\star_H) \leq \frac{\epsilon H}{\sqrt{N}}.$$

若 $\epsilon \leq \frac{1}{H}$，则 SMC-RS 仅使用 $N = O$个粒子，即可达到采样误差 $o(1)$。这与序贯蒙特卡洛形成鲜明对比：根据 命题 C.8，$N \geq \Omega(\sqrt{H})$ 个粒子是必要的，即使 $\widehat{V} = V^\star$ 且 $C_{\mathsf{act}} = O(1)$（注意，在此设定下，$C_{\mathsf{act}} = C_{\mathsf{act}} = O(1)$ 同样成立，因此 SMC-RS 确实能够成功）。

最后，SMC-RS 也有与定理 3.5 （重尾误差）和定理 3.6 （快速收敛率）类似的结果；分别见定理 D.3和定理 D.5。

3.4 粒子滤波方法的局限性

SMC（及其变体）的一项主要缺点是，当 PRM 不完美时，要避免误差放大，粒子数量 N 必须至少随时域 H线性增长。本节将说明，对于所有短视粒子滤波方法，时域依赖都是必要的。粗略地说，短视粒子滤波方法是指这样一种算法：在每一步 $\mathcal{S}_h$ 维护一组粒子 h，且不使用后续步骤 $k > h$ 的 PRM 数据来确定集合 $\mathcal{S}_h$。SMC 和 SMC-RS 是短视方法的两个例子；形式化定义见附录 F 。

我们证明，即使 PRM 存在轻微（常数因子）的误差，任何短视粒子滤波算法若要对 $\pi^\star_H$ 获得非平凡的覆盖度，也需要近乎 Ω(log H) 个粒子：

定理 *3.9*。设 $N(H) := \log(H)/(4 \log\log(H))$。不存在满足以下保证的短视粒子滤波算法 *Alg*：对于所有充分大的 H，*Alg* 使用至多 $N(H)$ 个粒子；并且在假设 *3.1* 和 *3.3* 以及 $C_{\mathsf{act}} = 2$ *and* $C_\infty = e^3$ 条件下，*Alg* 的输出分布 ν 满足：对于所有事件 $\mathcal{E} \subset \mathcal{X}_H$，$\nu(\mathcal{E}) \geq H^{-1/5}$且 $\pi^\star_H(\mathcal{E}) \geq 1/2$。

直接推论是，这类算法都无法达到 $D_{\mathsf{TV}}(\nu, \pi^\star_H) \leq 1/3$。相比之下，若 PRM 完美（即 $\mathsf{C}_\infty = 1$），单个粒子即可实现零误差（Rohatgi等人，2025年）。此外，定理 3.9 尤其引人注目，因为在假设 3.3 下，每个中间分布 $\widehat{\pi}_h$ 的覆盖度均为常数$\pi^\star_h$（即，对于所有 $x \in \mathcal{X}_h$，$\widehat{\pi}_h(x)/\pi^\star_h(x) \geq \mathsf{C}_\infty^{-2}$）。定理 3.9 表明，短视方法不可能在不使用超常数量粒子的情况下，归纳地“维持”这种覆盖度；这也引出了一个问题：某种形式的前瞻能否提高计算效率。

4 语言模型实验

接下来，我们将考察 第 3.1 节中的量在多大程度上能够预测序贯蒙特卡洛在涉及大语言模型的采样问题上的性能。

4.1 受控设定：提示切换任务

评估 SMC 并理解其性能如何随上述各量变化的一项关键挑战，是对这些量进行测量以及测量 SMC 本身的性能都很困难。尤其是，我们需要测量涉及目标分布的各量 $\pi^\star_h$，而目标分布通常并不容易获得。虽然先前的研究（例如 Zhao 等人 (2024年）已研究如何评估

SMC 的性能，并使用该算法对配分函数的估计值 $Z := \mathbb{E}_{\pi_{\mathtt{ref}}}[V^\star(x_H)]$, 但只有在 SMC 足够准确时，这类技术才能保证给出紧确界。此外，这些方法并未直接解决估计动作级覆盖等其他量所面临的挑战。因此，在本节中，我们考虑一种替代方法，并采用一种能够方便地满足以下条件的设定： $\pi_{\mathtt{ref}}$， $\pi^\star$ 均可高效采样。

特别地，我们考虑提示切换任务，其中 $\pi_{\mathtt{ref}}$ 和 $\pi^\star$ 分别对应语言模型在不同提示下的输出分布。形式上，我们固定一个语言模型 M，并考虑第1节中的设定，其中 $\mathcal{A}$ 是 M 的词元空间，且 $\mathcal{X}_h = \mathcal{A}^h$。提示切换任务的每个实例由 3 个提示 ($\mathtt{p}_{\mathtt{ref}}$, $\mathtt{p}^\star$, p) 确定，具体如下。给定这些提示的选择，我们令 $\pi_{\mathtt{ref}}$ 为 M 对提示 $\mathtt{p}_{\mathtt{ref}}$的输出分布，并将其记为 $\pi_{\mathtt{ref}}(\cdot) = \mathsf{M}(\cdot \mid \mathtt{p}_{\mathtt{ref}})$。接着，我们选择终端奖励函数为 $r^\star(a_{1:H}) := \mathsf{M}(a_{1:H} \mid \mathtt{p}^\star) / \mathsf{M}(a_{1:H} \mid \mathtt{p}_{\mathtt{ref}})$，由此可得，对于任意中间步骤 h，$V^\star(a_{1:h}) = \mathsf{M}(a_{1:h} \mid \mathtt{p}^\star) / \mathsf{M}(a_{1:h} \mid \mathtt{p}_{\mathtt{ref}})$。因此，$\pi^\star_h(a_{1:h}) = \mathsf{M}(a_{1:h} \mid \mathtt{p}^\star)$。由于从 M 中采样并计算序列级概率都是可行的，因此从 $\pi_{\mathtt{ref}}, \pi^\star$ 中采样以及计算 $V^\star$ 都是可行的。最后，提示 p 决定 V，具体方式将在下文讨论。

评估。为了评估序贯蒙特卡洛在提示切换任务上的表现，我们采用以下指标。该指标受到更复杂的“大语言模型评审”技术以及利用对数概率刻画大语言模型输出分布的方法启发（Golowich 等人，2025年）。给定两个关于长度为 H 的序列的分布 ν、ν'，以及一组提示 $\mathcal{P}$，我们将 $\mathcal{P}$-对数概率差异 $\mathrm{LPD}_{\mathcal{P}}(\nu, \nu')$ 定义为：在提示 $\mathcal{P}$ 下，从 ν 和 ν' 采样得到的序列所获平均词元级对数概率之差，其中对所有位置 h 求和。形式化地，对于 $\nu, \nu' \in \Delta(\mathcal{A}^H)$，有：

$$\mathrm{LPD}_{\mathcal{P}}(\nu, \nu') := \sum_{\mathtt{p} \in \mathcal{P}} \sum_{h=1}^{H} \left| \mathbb{E}_{a_{1:H} \sim \nu} \left[\log \mathsf{M}(a_h \mid \mathtt{p}, a_{1:h-1}) \right] - \mathbb{E}_{a'_{1:H} \sim \nu'} \left[\log \mathsf{M}(a'_h \mid \mathtt{p}, a'_{1:h-1}) \right] \right|.$$

换言之，$\mathcal{P}$-对数概率差异将 $\mathcal{P}$ 中的提示用作测试函数，以衡量两个分布 ν、ν' 之间的差异。可以将其视为一种非常简单（但更稳定）的“大语言模型评审”指标：我们不再从 M 生成内容来评估 ν、ν'，而是直接使用模型 M 给出的对数概率。在我们的评估中，我们选择 $\mathcal{P} = \{\mathtt{p}_{\mathtt{ref}}, \mathtt{p}^\star\}$；鉴于提示切换任务中最显著的两个提示是 $\mathtt{p}_{\mathtt{ref}}$ 和 $\mathtt{p}^\star$，这种选择十分自然。

4.2 评估对 PRM 准确度的依赖

为评估 PRM 准确度对 SMC 性能的影响，我们固定 $\asymp \pi_{\mathtt{ref}}$ 与 $\pi^\star$ 之间的动作级覆盖（具体而言，通过固定 $\mathtt{p}_{\mathtt{ref}} = \mathtt{p}^\star$），并改变 PRM V。具体来说，我们考虑上述提示切换任务中的若干 $k = 50$ 个实例：每个实例 $i \in [k]$ 由三个提示 $((\mathtt{p}_{\mathtt{ref}})^{(i)}, (\mathtt{p}^\star)^{(i)}, \mathtt{p}^{(i)})$ 以及一个标量 α 指定，该标量将在下文公式 (2) 中使用。在本小节中，所有实例共用参考提示 $(\mathtt{p}_{\mathtt{ref}})^{(i)}$ 和目标提示 $(\mathtt{p}^\star)^{(i)}$，因此我们分别记作 $\mathtt{p}_{\mathtt{ref}}$、$\mathtt{p}^\star$。此外，实际上 $\mathtt{p}_{\mathtt{ref}} = \mathtt{p}^\star$ ——该提示要求模型写一篇短篇故事（例如：“写一个场景，讲述一条龙与最后一个人类王国谈判缔结和平。”）。[6] 每个实例 $i \in [k]$ 的 $\widehat{V}$ 取值各不相同，记为 $\widehat{V}(i)$，其定义如下：我们固定 k 个提示 $\mathtt{p}^{(1)}, \ldots, \mathtt{p}^{(k)}$，它们都与 $\mathtt{p}_{\mathtt{ref}}$ 相同，只是分别要求以 k 种不同风格讲述故事（例如：“. . . 将其写成一篇新闻文章。”）。对于每个这样的提示 $\mathtt{p}^{(i)}$，我们将分布 $\pi^{(i)}(a_{1:H}) = \mathsf{M}(a_{1:H} \mid \mathtt{p}^{(i)})$ 记为 $\pi^{(i)}$，并将其视为 $\pi^\star$ 的“扰动”版本。然后我们定义

$$\widehat{V}^{(i)}(a_{1:h}) := \frac{\pi^\star(a_{1:h})}{\pi_{\mathtt{ref}}(a_{1:h})} \cdot \left(\frac{\pi^{(i)}(a_{1:h})}{\pi^\star(a_{1:h})} \right)^{(1-h/H)\cdot\alpha}, \tag{2}$$

对于某个参数 $\alpha > 0$，它控制着 $\widehat{V}^{(i)}$ 的准确性。为了直观地理解这个定义，简化起见，假设 $\alpha = 1$：当 h 较小时，$\widehat{V}^{(i)}(a_{1:h}) \approx \pi^{(i)}(a_{1:h}) / \pi_{\mathtt{ref},\, h}(a_{1:h})$；而在最后一步，$h = H$，$\widehat{V}_{(i)}(a_{1:H}) = V^\star(a_{1:H})$。特别是在早期步骤，$\widehat{V}_{(i)h}$ 看起来像是“扰动后”目标分布 $\pi^{(i)}$ 的价值函数，这意味着当 h 较小时，$\widehat{V}_h^{(i)}$ 会产生“误导”，因为它“指向”的是 $\pi^{(i)}$，而不是

[6] 虽然在此设定中从 $\mathtt{p}^\star$ 采样是轻而易举的，但我们强调，此处的主要目标是研究序贯蒙特卡洛在 PRM 准确度方面的性能，而不是解决一个具有挑战性的采样问题。

$\pi^\star$。当 h 较大时，我们会减弱噪声，以确保从 $V_H^\star$ 进行采样在信息论上是可行的（即，$\widehat{V}_{(i)\,H} = V_H^\star$）。

提示 $\mathrm{p}^{(i)}$（由 GPT-5 生成）的选择存在自然变化，导致得到的函数 $\widehat{V}_{(i)\,h}$ 相对于 $V_h^\star$的准确性水平各不相同。将 $\widehat{\pi}_h^{(i)} \in \Delta(\mathcal{A}^h)$ 定义为 $\widehat{\pi}_h^{(i)}(a_{1:h}) \propto \pi_{\mathsf{ref}}(a_{1:h}) \cdot \widehat{V}^{(i)}(a_{1:h})$。原则上可以估计 $D_{\chi^2}\left(\pi_h^\star \,\|\, \widehat{\pi}_h^{(i)}\right)$，但我们发现，估计 χ^2 散度所需的比率$V^\star(a_{1:h})/\widehat{V}_{(i)}(a_{1:h})$的方差过高，难以接受。因此，我们改为估计 $D_{\mathsf{KL}}\left(\pi_h^\star \,\|\, \widehat{\pi}_h^{(i)}\right)$，以衡量 $\pi_h^\star$ 与 $\widehat{\pi}_h^{(i)}$ 之间的散度。由于 KL散度可以写成对数概率的期望，这种方法要稳定得多（更多详情请参见附录 G.2）。

在我们的实验中，我们使用 Qwen3-0.6B 作为语言模型 M，并将时域设为 $H = 64$ 个词元。对于每个 $i \in [k]$ 的取值，我们使用给定的 PRM $\widehat{V}^{(i)}$运行粒子数为 $N = 32$ 的序贯蒙特卡洛，并从其输出分布 $\widehat{\nu}_H^{(i)}$ 中采样。在图 2b 中，我们绘制了 $D_{\mathsf{KL}}\left(\pi_h^\star \,\|\, \widehat{\pi}_h^{(i)}\right)$ 的取值与序贯蒙特卡洛采样误差之间的关系，采样误差通过 $\mathcal{P}$-logprob 差异 $\mathrm{LPD}_{\mathcal{P}}(\mathbb{E}[\widehat{\nu}_H^{(i)}], \pi_H^\star)$ 衡量（通过运行 200 次序贯蒙特卡洛试验进行估计）。我们观察到这两个量之间存在明显的相关性，从而验证了我们的理论发现：PRM 准确度（以 $\widehat{\pi}_h$ 和 $\pi_h^\star$之间的散度衡量）能够预测序贯蒙特卡洛的采样误差。在附录 G.2 中，我们针对 $\mathrm{p}_{\mathsf{ref}}$、$\mathrm{p}^\star$、$\mathrm{p}^{(i)}$、$\alpha$ 的其他取值重复了该实验，并观察到类似结果。

4.3 评估对动作级覆盖的依赖性

接下来，我们评估动作级覆盖对序贯蒙特卡洛性能的影响。我们考虑与上一小节相同的设定，唯一不同的是，对于每个实例 $i \in [k]$，我们令目标分布（记作 $(\pi^\star)^{(i)}$）与 $\pi^{(i)}$ 相同。具体而言，每个实例由元组 $(\mathrm{p}_{\mathsf{ref}}, \mathrm{p}^{(i)}, \mathrm{p}^{(i)})$ 确定。此外，我们令 $\alpha = 0$，使得 $\widehat{V}^{(i)} = (V^\star)^{(i)}$。在此设定下，动作级覆盖取决于 i。我们再次使用 Qwen3-0.6B，将时域设为 $H = 32$ 个词元，并使用 $N = 32$ 个粒子运行序贯蒙特卡洛。

虽然 $(\pi^\star)^{(i)}$ 与 π_{ref} 之间（按 L_∞ 的含义，依据 假设 3.1）的动作级覆盖度极大，实际上几乎不可能计算，但我们注意到 $D_{\mathsf{KL}}\left((\pi^\star)^{(i)} \,\|\, \pi_{\mathsf{ref}}\right)$ 可视为动作级覆盖的“平均情形代理量”，这是由链式法则可知的：

$$D_{\mathsf{KL}}\left((\pi^\star)^{(i)} \,\|\, \pi_{\mathsf{ref}}\right) = \sum_{h=1}^{H} \mathbb{E}_{a_{1:h} \sim (\pi^\star)_h^{(i)}} \left[\log \frac{(\pi^\star)_h^{(i)}(a_h \mid a_{1:h-1})}{\pi_{\mathsf{ref},h}(a_h \mid a_{1:h-1})}\right].$$

图 2a 绘制 $\frac{1}{H} D_{\mathsf{KL}}\left((\pi^\star)^{(i)} \,\|\, \pi_{\mathsf{ref}}\right)$ 与 $\mathcal{P}$-对数概率差异 $\mathrm{LPD}_{\mathcal{P}}(\mathbb{E}[\widehat{\nu}_H^{(i)}], \pi_H^\star)$；我们再次观察到这两个量之间存在明显的相关性，这与我们的理论一致。

4.4 对粒子数量的依赖

最后，在图 3 中，我们展示了提示切换任务中 SMC 的性能如何随粒子数量 N 变化，实验设置与第 4.2 节类似。具体而言，我们考虑 50 个不同的提示元组 $(\mathrm{p}^{(i)}\ \mathrm{ref}\ (\mathrm{p}^\star)^{(i)}, \mathrm{p}^{(i)})$，并针对每个元组使用一个 PRM $\widehat{V}_{(i)}$，该模型根据公式 (2) 从 $\mathrm{p}^{(i)}$ 定义，且$\alpha = 1$。除 SMC 外，我们还绘制了 (a) 使用 $\widehat{V}^{(i)}$ 的序贯重要性采样（附录 A），以及 (b)“best-of-N”算法的性能；该算法从 π_{ref} 中采样 N 个序列，并输出根据终端奖励 $(V^\star)^{(i)}(a_{1:H})$ 得分最高的序列。我们观察到，SMC 的表现始终优于这两种基线方法，而且其性能会随着 N 增加而提升；后一个结论与我们的理论结果一致。

4.5 序贯蒙特卡洛在数学问题求解任务中的应用

最后，我们研究了如何使用序贯蒙特卡洛来提升大语言模型在数学问题求解任务（包括 AIME 和 Math500）上的性能。在这一设定中，基础分布 π_{ref} 对应于通过提示要求解决数学问题的语言模型的输出分布，而终端奖励函数$r^\star(a_{1:H}) \in \{0, 1\}$则是答案 $a_{1:H}$是否正确的指示函数。因此，$\pi^\star$就是基础模型

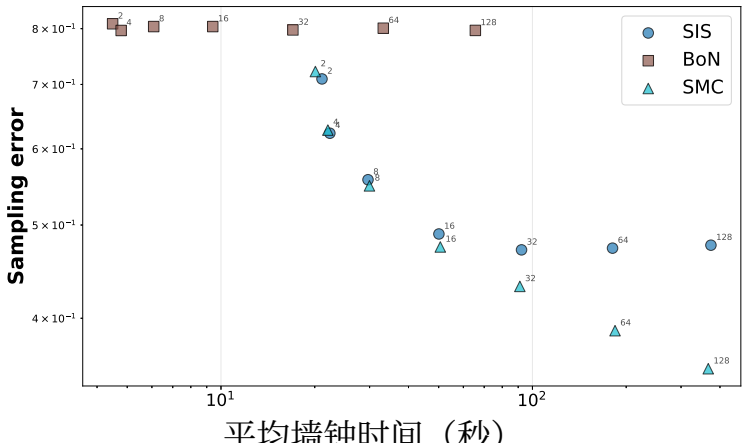


图 3：SMC 在提示切换任务（第 4.4 节）上的性能。每个点表示使用一定数量粒子的 SMC $N \in \{2, 4, 8, \ldots, 128\}$。我们对每个数据点的所有试验以及所有数据点的采样误差和墙钟时间取平均值。SMC 的性能始终优于序贯重要性采样（SIS）和 Best-of-N （BoN）基线方法。

分布 在生成正确答案的条件下，而 $V^\star(a_{1:h})$ 是在给定前缀 $a_{1:h}$ 的条件下， $\pi_{\texttt{ref}}$ 生成正确答案的概率。

有若干针对数学问题进行微调的开源过程奖励模型，据称能够对 $\widehat{}$给出良好的估计。在我们的实验中，我们将 PRM V 设为 Qwen2.5-Math-PRM-7B，并将基础模型 $\pi_{\texttt{ref}}$ 设为 Qwen2.5-1B-Instruct。在将 SMC 应用于数学推理任务时，由于目标只是找到某个 正确答案，通常会将 SMC 与 Best-of-N 采样结合使用，具体 $\widehat{}$方式如下：在 SMC 结束时，我们选择取值最高的单个粒子 V（$a_{1:H}$）。请注意，这不同于标准的 Best-of-N 采样过程；在该过程中，会从 $\sim\pi_{\texttt{ref}}$ 中独立地 抽取 N 个样本，并选择 V值最高的样本。

序贯蒙特卡洛能提升数学问题上的性能。 许多论文（Puri 等人，2025b；Chatziveroglou，2025年）报告称，序贯蒙特卡洛通常比 Best-of-N （以及其他基线方法）性能更好。不过，这类结果通常是对数据集中所有问题取平均后报告的，未能考虑各个问题之间的差异。例如，目前尚不清楚序贯蒙特卡洛在N 所有问题上的性能是否都优于 Best-of-N，还是说在数据集中相当一部分问题上，序贯蒙特卡洛的性能实际上可能更差（与 Best-of-相比）。[7] 显然，这一区别对于任何解释序贯蒙特卡洛性能的理论都至关重要。

在图 1中，我们估算了 Math500 中每个问题上序贯蒙特卡洛和 Best-of-N 的性能（每个问题进行 100 次试验），并绘制了序贯蒙特卡洛与 Best-of-N的准确性对比图。尽管两种算法的相对性能有所波动，但我们发现，对于绝大多数问题，序贯蒙特卡洛的准确性至少与 Best-of-N相当。在图 5中，我们在 AIME24 和 AIME25 上也观察到了类似的结论。

序贯蒙特卡洛与过程奖励模型准确度。既然我们已经确认，序贯蒙特卡洛能在 Math500 和 AIME 的大多数问题上持续提升准确性，接下来便要探究哪些因素会影响其在$\widehat{}$特定问题上的性能。鉴于人们已投入大量精力来设计更好的过程奖励模型 V (Lightman 等人，2023；Wang 等人，2024年，2025b；Zhang 等人，2025年），我们的理论自然引出了一个后续问题： $\pi_h^\star$ 与 $\widehat{\pi}_h$ 之间的散度（例如，用 χ^2-散度 $D_{\chi^2}(\pi_h^\star \,\|\, \widehat{\pi}_h)$ 衡量）能否预测序贯蒙特卡洛的性能。

为了解决这一问题，我们考察一系列过程奖励模型 $\widehat{V}^{(\lambda)}$ ，方法是改变λ 过程奖励模型用于生成最终概率的逆温度参数。在图 4中，我们选取 Math500 中的 10 道问题，并针对 λ的每个取值，绘制 $D_{\chi^2}(\pi_h^\star \,\|\, \widehat{\pi}_h)$ 的经验估计值与给定 $\widehat{V}^{(\lambda)}$时序贯蒙特卡洛的准确性之间的关系。

[7]事实上，许多先前的研究都指出，序贯蒙特卡洛可能会出现众数坍缩等病态现象，从而降低 Best-of-N 可供选择的回答多样性。

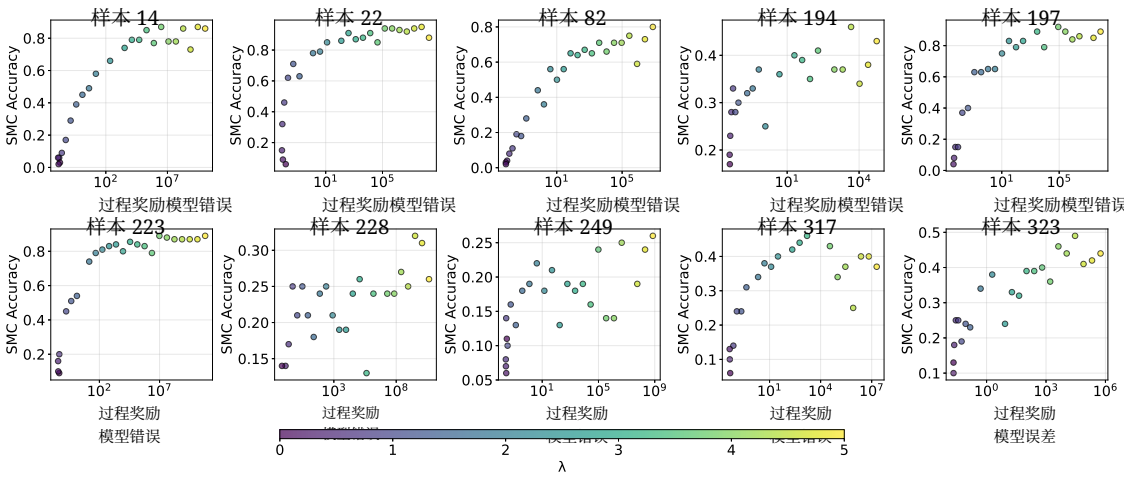


图 4：过程奖励模型误差（以 $D_{\chi^2}(\pi_h^\star \,\|\, \widehat{\pi}_h)$ 衡量）对 10 道 Math500 问题的序贯蒙特卡洛准确性的影响。不同颜色对应逆温度参数 λ 的不同取值，该参数用于参数化 $\widehat{V}^{(\lambda)}$。图中展示的是序贯蒙特卡洛结束时选取的随机粒子（而非最佳粒子）的序贯蒙特卡洛准确性。

有趣的是，我们观察到，较大的 χ^2-散度往往会带来更高的准确性，这与我们预测相反结果的理论发现形成对比。我们推测，较大的 λ（较低的温度）会带来更好的性能，因为它们能更有效地剔除错误的部分解；然而，它们也可能导致 $\widehat{\pi}_h$ 漏掉 $\pi_h^\star$的一些众数，从而使 χ^2-散度急剧增大。因此，为了准确评估序贯蒙特卡洛在这类问题上的性能，我们需要对框架进行修改：目标不是从分布意义上近似 $\pi_H^\star$，而只是覆盖其部分概率质量。

5 讨论

本文在多种假设下，为粒子滤波算法建立了理论保证，这些假设涉及参考分布 π_{ref}、目标分布 $\pi^\star$ 与过程奖励模型 V（通过诱导分布 $\widehat{\pi}$）之间的关系。下面我们重点介绍一些值得探究的未来研究方向：

- 我们的上界与下界之间存在多项式间隙，例如它们在时域上的依赖关系 H，如第 1.2 节所述；此外，全变差距离随 N变化的衰减速度也存在差异：尽管定理 3.2 和3.8的上界以 $1/\sqrt{N}$的速度衰减，但命题 C.8的下界仅以 $1/N$的速度衰减。

- 我们能否为 SMC（以及其他推理时采样算法）的性能开发一些指标？这些指标所要求的条件弱于全变差距离，却仍能有意义地刻画输出的“有用性”。例如，对于数学问题，通常只需要答案正确，解答的具体措辞可能并不那么重要。

- 我们能否为粒子滤波算法或其他类型的推理时采样过程提供可证明保证？这些过程所依赖的 $\pi_h^\star$ 与 $\widehat{\pi}_h$ 之间的差异度量较弱。实验中（例如$D_{\chi^2}(\pi_h^\star \,\|\, \widehat{\pi}_h)$）观察到的数值第 4.2节至第 4.4节中的数值极大（尤其是远大于我们实验中观察到的 N 数值），然而粒子滤波算法仍然能够有效利用 V。

参考文献

Edoardo Botta、Yuchen Li、Aashay Mehta、Jordan T. Ash、Cyril Zhang 和 Andrej Risteski。《验证器辅助语言生成的查询复杂度》。载于第四十二届国际机器学习会议，2025年。网址：https://openreview.net/forum?id=9oIjvaDhoN。

Bradley Brown，Jor dan Juravsky、Ryan Ehrlich、Ronald Clark、Quoc V Le、Christopher Ré 和 Azali a

Mirhoseini。《大型语言猴子：通过重复采样扩展推理计算》。*arXiv* 预印本 *arXiv:2407.21787*，2024年。F Cérou、P Del Moral 和 A Guyader。《未归一化费曼–卡茨粒子模型的一个非渐近定理》。载于*Annales de l'IHP Probabilités et statistiques*，第47卷，第629–649页。Gauthier-Villars，2011年。Giannis Chatziveroglou。《A* 解码：令牌高效推理扩展》，2025年。网址：https://arxiv.org/abs/2505.13672。Fan Chen、Audrey Huang、Noah Golowich、Sadhika Malladi、Adam Block、Jordan T Ash、Akshay Krish- namurthy 和 Dylan J Foster。《覆盖原则：预训练如何助力后训练》。*arXiv* 预印本 *arXiv:2510.15020*，2025年。Pierre Del Moral。《费曼–卡茨公式》。载于费曼–卡茨公式：谱系与相互作用粒子系统及其应用，第47–93页。施普林格，2004年。Arnaud Doucet、Simon J. Godsill 和 Christophe Andrieu。《贝叶斯滤波的序贯蒙特卡洛采样方法》。*Statistics and Computing*，10(3):197–208，2000年。Shengyu Feng、Xiang Kong、Shuang Ma、Aonan Zhang、Dong Yin、Chong Wang、Ruoming Pang 和 Yiming Yang。《通过扭曲序贯蒙特卡洛对数学问题进行逐步推理》。*arXiv* 预印本 *arXiv:2410.01920*，2024年。Dylan J. Foster、Zakaria Mhammedi 和 Dhruv Rohatgi。《高效强化学习是否需要良好的基础？基础模型在探索中的计算作用》。*arXiv* 预印本 *arXiv:2503.07453*，2025年。Yichao Fu、Xuewei Wang、Yuandong Tian 和 Jiawei Zhao。《自信地深入思考》。载于 *NeurIPS 2025* 高效推理研讨会，2025年。Jonathan Geuter、Youssef Mroueh 和 David Alvarez-Melis。《用于大语言模型高效测试时对齐的引导式推测推理》。*arXiv* 预印本 *arXiv:2506.04118*，2025年。Noah Golowich、Allen Liu 和 Abhishek Shetty。《logit 序列揭示语言模型的低秩结构》，2025年。网址：https://arxiv.org/abs/2510.24966。Gabriel Grand、Joshua B. Tenenbaum、Vikash K. Mansinghka、Alexander K. Lew 和 Jacob Andreas。《自我引导语言模型》。载于语言建模会议（*COLM*），2025年。arXiv:2504.07081。Peter Guarniero、Adam M. Johansen 和 Anthony Lee。《迭代辅助粒子滤波器》。美国统计学会期刊，112(520):1636–1647，2017年。Jeremy Heng、Adrian Bishop、George Deligiannidis 和 Arnaud Doucet。《受控序贯蒙特卡洛》。*arXiv* 预印本 *arXiv:1708.08396*，2017年。Aayush Karan 和 Yilun Du。《基于采样的推理：你的基础模型比你想象的更聪明》，2025年。网址：https://arxiv.org/abs/2510.14901。J. Lawson、L. Raventos、A. Warrington 和 S. Linderman。《Sixo：基于扭曲目标的平滑推理》。*CoRR*，abs/2206.05952，2022年。Holden Lee 和 Matheau Santana-Gijzen。《利用软分解推导多模态分布上序贯蒙特卡洛的收敛界》。*arXiv* 预印本 *arXiv:2405.19553*，2024年。Alexander K. Lew、Zhi-Xuan Tan、Gabriel Grand 和 Vikash K. Mansinghka。《使用概率程序对大语言模型进行序贯蒙特卡洛引导》。*arXiv* 预印本 *arXiv:2306.03081*，2023年。Hunter Lightman、Vineet Kosaraju、Yura Burda、Harri Edwards、Bowen Baker、Teddy Lee、Jan Leike、John Schulman、Ilya Sutskever 和 Karl Cobbe。《让我们逐步验证》。*arXiv* 预印本 *arXiv:2305.20050*，2023年。Benjamin Lipkin、Benjamin LeBrun、Jacob Hoover Vigly、João Loula、David R. MacIver、Li Du、Jason Eisner、Ryan Cotterell、Vikash K. Mansinghka、Timothy J. O'Donnell、Alexander K. Lew 和 Tim Vieira。《使用自适应加权拒绝采样从语言模型快速生成受控文本》。*arXiv* 预印本 *arXiv:2504.05410*，2025年。

João Loula、Benjamin LeBrun、Li Du、Ben Lipkin、Clemente Pasti、Gabriel Grand、Tianyu Liu、Yahya Emara、Marjorie Freedman、Jason Eisner、Ryan Cotterell、Vikash K. Mansinghka、Alexander K. Lew、Tim Vieira 和 Timothy J. O'Donnell。通过序贯蒙特卡洛对大语言模型进行句法与语义控制。发表于国际学习表征会议，2025年。arXiv:2504.13139。Lovish Madaan、Aniket Didolkar、Suchin Gururangan、John Quan、Ruan Silva、Ruslan Salakhutdinov、Manzil Zaheer、Sanjeev Arora 和 Anirudh Goyal。重新思考思考令牌：将大语言模型视为改进算子。*arXiv*预印本 *arXiv:2510.01123*，2025年。Joseph Marion 和 Scott C Schmidler。序贯蒙特卡洛和自适应路径选择的有限样本 l_2 界。*arXiv*预印本 *arXiv:1807.01346*，2018年。Michael Montemerlo、Sebastian Thrun、Daphne Koller 和 Ben Wegbreit。FastSLAM：同时定位与建图问题的一种因子化解法。发表于*AAAI*人工智能大会论文集，2002年。Sidharth Mudgal、Jong Lee、Harish Ganapathy、YaGuang Li、Tao Wang、Yanping Huang、Zhifeng Chen、Heng-Tze Cheng、Michael Collins、Trevor Strohman、Jilin Chen、Alex Beutel 和 Ahmad Beirami。基于语言模型的受控解码。发表于第四十一届国际机器学习大会，2024年。网址：https://openreview.net/forum?id=bVIcZb7Qa0。C. A. Naesseth、F. Lindsten 和 T. B. Schön。使用嵌套序贯蒙特卡洛进行高维滤波。《*IEEE*信号处理汇刊》，67(16):4177–4188，2019年。doi：10.1109/TSP.2019.2926035。Michael K. Pitt 和 Neil Shephard。通过模拟进行滤波：辅助粒子滤波器。技术报告 W13，牛津大学纳菲尔德学院，1997年。Isha Puri、Shivchander Sudalairaj、Guangxuan Xu、Kai Xu 和 Akash Srivastava。使用基于粒子的蒙特卡洛方法，以概率推断方法实现大语言模型的推理时扩展。*arXiv*预印本 *arXiv:2502.01618*，2025a。Isha Puri、Shivchander Sudalairaj、Guangxuan Xu、Kai Xu 和 Akash Srivastava。使用基于粒子的蒙特卡洛方法，以概率推断方法实现大语言模型的推理时扩展。*arXiv*预印本 *arXiv:2502.01618*，2025b。Dhruv Rohatgi、Abhishek Shetty、Donya Saless、Yuchen Li、Ankur Moitra、Andrej Risteski 和 Dylan J Foster。驯服不完美的过程验证器：回溯的采样视角。*arXiv*预印本 *arXiv:2510.03149*，2025年。Nikolaus Schweizer。多模态场景下序贯马尔可夫链蒙特卡洛方法的非渐近误差界。*arXiv*预印本 *arXiv:1205.6733*，2012年。Raghav Singhal、Zachary Horvitz、Ryan Teehan、Mengye Ren、Zhou Yu、Kathleen McKeown 和 Rajesh Ranganath。扩散模型推理时扩展与引导的通用框架。*arXiv*预印本 *arXiv:2501.06848*，2025年。Marta Skreta、Tara Akhound-Sadegh、Viktor Ohanesian、Roberto Bondesan、Alán Aspuru-Guzik、Arnaud Doucet、Rob Brekelmans、Alexander Tong 和 Kirill Neklyudov。扩散模型中的费曼-卡茨校正器：退火、引导与专家乘积。*arXiv*预印本 *arXiv:2503.02819*，2025年。Andreas Stephan、Dawei Zhu、Matthias Assenmacher、Xiaoyu Shen 和 Benjamin Roth。从计算到裁决：考察大语言模型评判器在数学推理任务中的表现。*arXiv*预印本 *arXiv:2409.04168*，2024年。Kaiwen Wang、Jin Peng Zhou、Jonathan Chang、Zhaolin Gao、Nathan Kallus、Kianté Brantley 和 Wen Sun。用于高效思维链推理的价值引导搜索。*arXiv*预印本 *arXiv:2505.17373*，2025a。Kaiwen Wang、Jin Peng Zhou、Jonathan Chang、Zhaolin Gao、Nathan Kallus、Kiante Brantley 和 Wen Sun。用于高效思维链推理的价值引导搜索。发表于神经信息处理系统进展，2025b。arXiv:2505.17373。Peiyi Wang、Lei Li、Zhihong Shao、R. X. Xu、Damai Dai、Yifei Li、Deli Chen、Y. Wu 和 Zhifang Sui。

Math-Shepherd：无需人工标注，逐步验证并强化大语言模型。*arXiv*预印本 *arXiv:2312.08935*，2024年。Xuezhi Wang、Jason Wei、Dale Schuurmans、Quoc V Le、Ed H. Chi、Sharan Narang、Aakanksha Chowdhery 和 Denny Zhou。自洽性可提升语言模型中的思维链推理效果。发表于第十一届国际学习表征会议，2023年。网址：https://openreview.net/forum?id=1PL1NIMMrw。Nick Whiteley、Nikolas Kantas 和 Ajay Jasra。时间齐次 Feynman–Kac 公式的粒子近似线性方差界。《随机过程及其应用》，122(4):1840–1865，2012年。Xixi Wu、Kuan Li、Yida Zhao、Liwen Zhang、Litu Ou、Huifeng Yin、Zhongwang Zhang、Xinmiao Yu、Dingchu Zhang、Yong Jiang 等。Resum：通过上下文摘要释放长时域搜索智能。*arXiv*预印本 *arXiv:2509.13313*，2025年。Wei Xiong、Hanze Dong、Chenlu Ye、Ziqi Wang、Han Zhong、Heng Ji、Nan Jiang 和 Tong Zhang。基于人类反馈的迭代式偏好学习：弥合 KL 约束下基于人类反馈的强化学习的理论与实践。*arXiv*预印本 *arXiv:2312.11456*，2024年。Kevin Yang 和 Dan Klein。Fudge：利用未来判别器实现可控文本生成。发表于*2021*年北美计算语言学协会年会论文集：人类语言技术，第3511–3535页，2021年。Ling Yang、Zhaochen Yu、Tianjun Zhang、Shiyi Cao、Minkai Xu、Wentao Zhang、Joseph E Gonzalez 和 Bin Cui。思想缓冲区：利用大语言模型进行思想增强推理。神经信息处理系统进展，37:113519–113544，2024年。Yue Yu、Qiwei Di、Quanquan Gu 和 Dongruo Zhou。论测试时计算的局限：通过序列奖励过滤提升推理效果。*arXiv*预印本 *arXiv:2512.04558*，2025年。Zhenru Zhang、Chujie Zheng、Yangzhen Wu、Beichen Zhang、Runji Lin、Bowen Yu、Dayiheng Liu、Jingren Zhou 和 Junyang Lin。数学推理中开发过程奖励模型的经验教训。*arXiv*预印本 *arXiv:2501.07301*，2025年。Stephen Zhao、Rob Brekelmans、Alireza Makhzani 和 Roger Grosse。通过扭曲序贯蒙特卡洛实现语言模型中的概率推断。*arXiv*预印本 *arXiv:2404.17546*，2024年。Wenting Zhao、Pranjal Aggarwal、Swarnadeep Saha、Asli Celikyilmaz、Jason Weston 和 Ilia Kulikov。多数意见并非总是正确：用于解聚合的强化学习训练。*arXiv*预印本 *arXiv:2509.06870*，2025年。Youheng Zhu 和 Yiping Lu。论（近似）奖励模型在推理时扩展中的作用。*arXiv*预印本 *arXiv:2602.01381*，2026年。

附录目录

A 与先前工作的比较 17 B 技术工具 17

C 序贯蒙特卡洛的证明 **18**
C.1 序贯蒙特卡洛的有用引理.18 C.2 定理 3.2 的证明.19 C.3 现有分析的不足.20 C.4 定理 3.5 的证明.21 C.5 定理 3.6 的证明.22 C.6 具有精确价值函数的序贯蒙特卡洛下界.24

D SMC-RS 的证明 **25**
D.1 定理 3.8 的证明.26 D.2 重尾误差下 SMC-RS 的分析.26 D.3 SMC-RS 的更快收敛.27

E 回溯与粒子滤波的统一视角 **28**
E.1 VGB 背景.29 E.2 SMC-IND 的定义.29 E.3 将 VGB 与 SMC-IND 耦合.30

F 定理 3.9 的证明 31
F.1 短视粒子滤波算法.31 F.2 专门设定与定理的正式表述.31 F.3 定理 F.5 的证明.32

G 补充实验细节 **37**
G.1 其他评估方法.38 G.2 第 4.2 节的详细内容.38 G.3 第 4.3 节的详细内容.39 G.4 AIME 上的序贯蒙特卡洛与 Best-of-N 对比.39

与先前工作的比较

本节讨论我们的理论保证与先前研究的比较。我们将讨论范围限定在自回归生成设置，因为过去尚未从理论上研究更一般的设置。

动作级拒绝采样 / 序贯重要性采样（Yang 和 Klein，2021）。该算法通过条件分布 $a_{1:H}$ 进行自回归采样

$$\widehat{\pi}(a_{h+1} \mid a_{1:h}) \propto \pi_{\mathsf{ref}}(a_{1:h+1} \mid a_{1:h})\widehat{V}(a_{1:h+1}).$$

其预期运行时间为 $O(HC_{\mathsf{act}})$ 可通过拒绝采样 / 重要性采样实现，Rohatgi 等人（2025年），且根据公式 (1)，若 $\pi^\star$，则可从 $V = V^\star$中精确采样。然而，即使满足假设 3.3且$\mathsf{C}_\infty = 1+\epsilon$，仍可能产生$\Omega(\epsilon\sqrt{H})$采样误差，Rohatgi 等人（2025年），例如当 $\epsilon = \Theta(1)$时，该误差界便毫无意义。

VGB（Rohatgi 等人，2025年）。该算法定义为在自回归生成树上进行随机游走，其状态空间为 $\mathcal{A}^0 \sqcup \cdots \sqcup \mathcal{A}^H$，其中 $a_{1:h}$的父节点为 $a_{1:h-1}$。Rohatgi 等人在（2025年）中的第一个主要结果表明，该随机游走可从 $\pi^\star$中采样，且全变差误差不超过 δ，所需时间约为 $O(H^2 C_{\mathsf{act}}\mathsf{C}_\infty^4 \log(1/\delta))$。这一保证与定理 3.6对序贯蒙特卡洛的保证基本相当，不过序贯蒙特卡洛中的语言模型评估可并行化为 H 轮，而 VGB 中的语言模型评估则本质上是顺序进行的。[8]

Rohatgi 等人（2025年）的第二个主要结果表明，在以下 平均情形接近性条件下， V 与 $V^\star$之间满足该条件，且动作级覆盖有界时，VGB 会从一个近似覆盖 $\pi^\star$的分布中进行采样。

假设 A.1 （Rohatgi 等人（2025年）的平均情形假设）。

$$\max\left\{\mathbb{E}^{\pi^\star}\left[\frac{\widehat{V}(a_{1:h})}{V^\star(a_{1:h})}\right], \mathbb{E}^{\pi^\star}\left[\frac{V^\star(a_{1:h})}{\widehat{V}(a_{1:h})}\right]\right\} \le C.$$

我们作出如下观察：

事实 A.2。 对于每个 $0 \le h \le H$，均有

$$D_{\chi^2}(\pi_h^\star \,\|\, \widehat{\pi}_h) = \mathbb{E}^{\pi^\star}\left[\frac{\widehat{V}(a_{1:h})}{V^\star(a_{1:h})}\right] \cdot \mathbb{E}^{\pi^\star}\left[\frac{V^\star(a_{1:h})}{\widehat{V}(a_{1:h})}\right] - 1.$$

因此，在假设 A.1 以及动作级覆盖有界的条件下，定理 3.2 表明，使用$N \le (C+1)^2 H^2/\epsilon^2$ 个粒子的序贯蒙特卡洛会从一个与 ϵ在全变差距离上相差 $\pi^\star$的分布中采样。请注意，在全变差距离上的接近性比近似覆盖提供了更强的保证。

B 技术工具

命题 B.1。 对于任意 $R \ge 0$、独立随机变量 X_1、$\cdots$、X_n，有

$$\mathbb{E}\left|\sum_{i=1}^n (X_i - \mathbb{E}[X_i])\right| \le \sqrt{nR \max_i \mathbb{E}|X_i|} + 2n \max_i \mathbb{E}[|X_i| \mathbb{I}\{|X_i| \ge R\}].$$

证明。 我们定义 $Y_i = \min\{\max\{X_i, -R\}, R\}$ 以及 $Z_i = X_i - Y_i$。于是，我们可以界定

$$\mathbb{E}\left|\sum_{i=1}^n (X_i - \mathbb{E}[X_i])\right| \le \mathbb{E}\left|\sum_{i=1}^n (Y_i - \mathbb{E}[Y_i])\right| + \sum_{i=1}^n \mathbb{E}|Z_i - \mathbb{E}[Z_i]|.$$

[8]附录 E 的结果表明，可以通过粒子滤波来并行化 VGB，但尚不清楚这种模拟在多大范围内能够高效运行。

请注意 $\mathbb{E}|Z_i - \mathbb{E}[Z_i]| \le 2\,\mathbb{E}|Z_i| \le 2\,\mathbb{E}[|X_i|\mathbb{I}\{|X_i| \ge R\}]$。此外，

$$\left(\mathbb{E}\left|\sum_{i=1}^{n}(Y_i - \mathbb{E}[Y_i])\right|\right)^2 \le \mathbb{E}\left(\sum_{i=1}^{n}(Y_i - \mathbb{E}[Y_i])\right)^2$$
$$= \sum_{i=1}^{n}\mathbb{E}(Y_i - \mathbb{E}[Y_i])^2 \le \sum_{i=1}^{n}\mathbb{E}\,Y_i^2 \le R\sum_{i=1}^{n}\mathbb{E}|X_i|,$$

其中我们使用 e $|Y_i| \le \min\{|X_i|, R\}$。结合上述不等式即可完成证明 ve. □

C 序贯蒙特卡洛的证明

记号。 我们定义 $\mathcal{H}_h = \{(\tilde{x}_\ell^i, x_\ell^i)\}_{\ell\in[h],i\in[N]}$ 为序贯蒙特卡洛执行截至第 算法 1 次迭代结束时的完整历史 h。我们用 $\mathbb{E}_h[\cdot]$ 表示关于 $\mathcal{H}_h$的条件分布。我们记作$Z := \mathbb{E}_{x_{1:H}\sim\pi_{\text{ref}}}[r^\star(x_H)]$ 并注意到，对于所有 $Z = \mathbb{E}_{x\sim\pi_h}[V^\star(x)]$ 均成立 h。

C.1 序贯蒙特卡洛的有用引理

我们首先证明以下引理。

引理 C.1。 对于任意步骤 h 和任意函数 $g : \mathcal{X}_h \to \mathbb{R}$，有

$$\mathbb{E}_{h-1}\left[\widehat{W}_h \cdot \mathbb{E}_{\widehat{\nu}_h}[g]\right] = \widehat{W}_{h-1} \cdot \mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}}\left[\frac{\mathbb{E}_{x_h\sim\pi_{\text{ref}}(\cdot|x_{h-1})}\left[\widehat{V}(x_h)\cdot g(x_h)\right]}{\widehat{V}(x_{h-1})}\right]. \tag{3}$$

下一个引理可通过对 引理 C.1 进行归纳应用得到。

引理 C.2。 对于任意 $g_h : \mathcal{X}_h \to \mathbb{R}$，有

$$\mathbb{E}\left[\widehat{W}_h \cdot \mathbb{E}_{\widehat{\nu}_h}[g_h]\right] = \mathbb{E}_{x_h\sim\pi_h}\left[\widehat{V}(x_h)g_h(x_h)\right]. \tag{4}$$

特别地，$\mathbb{E}[W_h] = Z_h = \mathbb{E}[V(x_h)]$ 对于 $h \ge 1$ 和

$$\mathbb{E}\left[\widehat{W}_H\widehat{\nu}_H(x)\right] = Z\cdot\pi_H^\star(x), \qquad \forall x \in \mathcal{X}_H. \tag{5}$$

最后，我们将 算法 1 的全变差误差上界化为 W_H 的偏差。

引理 C.3。 对于 *算法 1*，有

$$D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H], \pi_H^\star) \le \frac{1}{2Z}\,\mathbb{E}|\widehat{W}_H - Z|.$$

证明引理 C.1 和引理 C.2。为证明公式（3），注意到

$$\mathbb{E}_{h-1}\left[\widehat{W}_h \cdot \mathbb{E}_{\widehat{\nu}_h}[g]\right]$$
$$= \widehat{W}_{h-1} \cdot \mathbb{E}_{h-1}\left[\frac{1}{N}\sum_{i=1}^{N} w_h^i g(x_h^i)\right]$$
$$= \widehat{W}_{h-1} \cdot \mathbb{E}_{h-1}\left[\frac{1}{N}\sum_{i=1}^{N}\frac{\widehat{V}(x_h^i)}{\widehat{V}(\tilde{x}_{h-1}^i)}g(x_h^i)\right].$$

由于给定 $\mathcal{H}_{h-1}$时， $\{(\tilde{x}_{h-1}^i,\ x_h^i)\}$ 独立同分布，对于每个 $i \in [N]$，我们有

$$\mathbb{E}_{h-1}\left[\frac{\widehat{V}(x_h^i)}{\widehat{V}(\tilde{x}_{h-1}^i)}g(x_h^i)\right] = \mathbb{E}_{\tilde{x}_{h-1}\sim\hat{\nu}_{h-1}}\mathbb{E}_{x_h\sim\pi_{\text{ref}}(\cdot|\tilde{x}_{h-1})}\left[\frac{\widehat{V}(x_h)}{\widehat{V}(\tilde{x}_{h-1})}g(x_h)\right],$$

这就证明了公式（3）。接下来，我们利用公式（3），论证公式（4）可通过归纳法成立（注意基例显然成立$\widehat{\quad}\quad\widehat{\quad}$，因为 $W_0 = V(\perp)$ 且 $\widehat{\nu}_0 = \delta_\perp$）：

$$\begin{aligned}\mathbb{E}\left[\widehat{W}_h \cdot \mathbb{E}_{\widehat{\nu}_H}[g_h]\right] =& \mathbb{E}\left[\widehat{W}_{h-1} \cdot \mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}}\left[\frac{\mathbb{E}_{x_h\sim\pi_{\mathsf{ref}}(\cdot|x_{h-1})}\left[\widehat{V}(x_h)\cdot g_h(x_h)\right]}{\widehat{V}(x_{h-1})}\right]\right] \\ =& \mathbb{E}\left[\widehat{V}(x_{h-1})\cdot\frac{\mathbb{E}_{x_h\sim\pi_{\mathsf{ref}}(\cdot|x_{h-1})}[\widehat{V}(x_h)g_h(x_h)]}{\widehat{V}(x_{h-1})}\right] \\ =& \mathbb{E}[\widehat{V}(x_h)g_h(x_h)],\end{aligned}$$

$\widehat{\quad}\quad\widehat{\quad}$如所期望。我们得到 $\mathbb{E}[W_h] = \mathbb{E}[V(x_h)]$，这是通过使用 $g_h \equiv 1$得到的。最后，公式（5）可通过设定 $g_H(x) = \mathbb{I}\{x_H = x\}$于公式（4）中，并利用 $V(x_H) = V^\star(x_H)$对于 $x_H \in \mathcal{X}_H$成立。 □

证明引理 C.3。根据公式（5），我们可以写出

$$\begin{aligned}2Z D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H], \pi^\star_H) =& \sum_{x\in\mathcal{X}_H}\left|Z\,\mathbb{E}[\widehat{\nu}_H(x)] - Z\pi^\star_H(x)\right| = \sum_{x\in\mathcal{X}_H}\left|Z\,\mathbb{E}[\widehat{\nu}_H(x)] - \mathbb{E}[\widehat{W}_H\widehat{\nu}_H(x)]\right| \\ \le& \sum_{x\in\mathcal{X}_H}\mathbb{E}\left[\left|\widehat{W}_H - Z\right|\widehat{\nu}_H(x)\right] = \mathbb{E}\left[\left|\widehat{W}_H - Z\right|\cdot\sum_{x\in\mathcal{X}_H}\widehat{\nu}_H(x)\right] = \mathbb{E}\left|\widehat{W}_H - Z\right|.\end{aligned}$$

□

C.2 证明定理 3.2

我们证明一个稍强的上界。

命题 C.4. 序贯蒙特卡洛（算法 *1*）使用 N 个粒子可达到

$$D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H], \pi^\star_H) \le \frac{1}{\sqrt{N}}\sum_{h=1}^{H}\sqrt{\mathbb{E}_{\pi_{\mathsf{ref}}}[\widehat{V}(x_{h-1})]\,\mathbb{E}_{\pi_{\mathsf{ref}}}\left[\frac{V^\star(x_h)^2}{\widehat{V}(x_{h-1})}\right]}.$$

定理 3.2 随后由以下事实得出：$V^\star(x_h) \le C_{\mathsf{act}}V^\star(x_{h-1})$ 在 $x_h \sim \pi_{\mathsf{ref}}(\cdot \mid x_{h-1})$下几乎必然成立，并且

$$\mathbb{E}_{\pi_{\mathsf{ref}}}[\widehat{V}(x_{h-1})]\,\mathbb{E}_{\pi_{\mathsf{ref}}}\left[\frac{V^\star(x_{h-1})^2}{\widehat{V}(x_{h-1})}\right] = 1 + D_{\chi^2}\left(\pi^\star_{h-1} \,\|\, \widehat{\pi}_{h-1}\right).$$

□

以下是命题 C.4的证明。我们定义 $\rho(x) = {}^{V^\star}{}_{(x)}\,\widehat{V}(x)$ 以及

$$M_h = \widehat{W}_h \cdot \mathbb{E}_{x_h\sim\widehat{\nu}_h}[\rho(x_h)].$$

$\widehat{\quad}$于是公式（3）结合 $g(x) = \rho(x)$ 表明(M_h)是一个鞅，且具有 $M_0 = Z$和 $M_H = W_H$。然后我们可以分解

$$\mathbb{E}|\widehat{W}_H - Z| \le \sum_{h=1}^{H}\mathbb{E}|M_h - M_{h-1}|.$$

接下来，我们注意到

$$\begin{aligned}M_h - M_{h-1} =& \widehat{W}_{h-1}\left(\frac{1}{N}\sum_{i=1}^{N}w_h^i\rho(x_h^i) - \mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}}[\rho(x_{h-1})]\right) \\ =& \widehat{W}_{h-1}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{V^\star(x_h^i)}{\widehat{V}(\tilde{x}_{h-1}^i)} - \mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}}[\rho(x_{h-1})]\right).\end{aligned}$$

由于给定 $\mathcal{H}_{h-1}$ 时，元组 $(\tilde{x}_{h-1}^i, x_h^i)$ 是独立同分布的，因此我们可以给出上界

$$\frac{1}{\widehat{W}_{h-1}^2}\mathbb{E}_{h-1}(M_h - M_{h-1})^2 \le \frac{1}{N}\mathrm{Var}_{x_{h-1}\sim\widehat{\nu}_{h-1}, x_h\sim\pi_{\mathsf{ref}}(\cdot|x_{h-1})}\left[\frac{V^\star(x_h)}{\widehat{V}(x_{h-1})}\right]$$
$$\le \frac{1}{N}\mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}}\left[\frac{\mathbb{E}_{x_h\sim\pi_{\mathsf{ref}}(\cdot|x_{h-1})}[V^\star(x_h)^2]}{\widehat{V}(x_{h-1})^2}\right] =: \Delta_{h-1}.$$

因此，根据引理 C.2，可得

$$\mathbb{E}[\widehat{W}_{h-1}]\,\mathbb{E}[\widehat{W}_{h-1}\Delta_{h-1}] = \frac{1}{N}\mathbb{E}_{\pi_{\mathsf{ref}}}[\widehat{V}(x_{h-1})]\,\mathbb{E}_{\pi_{\mathsf{ref}}}\left[\frac{V^\star(x_h)^2}{\widehat{V}(x_{h-1})}\right].$$

因此，我们可以界定

$$\mathbb{E}|\widehat{W}_H - Z| \le \sum_{h=1}^H \mathbb{E}[\mathbb{E}_{h-1}|M_h - M_{h-1}|]$$
$$\le \sum_{h=1}^H \mathbb{E}\left[\widehat{W}_{h-1}\sqrt{\Delta_{h-1}}\right] \le \sum_{h=1}^H \sqrt{\mathbb{E}[\widehat{W}_{h-1}]\cdot\mathbb{E}\left[\widehat{W}_{h-1}\Delta_{h-1}\right]}$$
$$\le \frac{1}{\sqrt{N}}\sum_{h=1}^H \sqrt{\mathbb{E}_{\pi_{\mathsf{ref}}}[\widehat{V}(x_{h-1})]\,\mathbb{E}_{\pi_{\mathsf{ref}}}\left[\frac{V^\star(x_h)^2}{\widehat{V}(x_{h-1})}\right]}.$$

结合上述不等式，即完成证明。 □

C.3 现有分析的不足

大多数关于序贯蒙特卡洛的非渐近分析，所作假设至少与 L_∞ 假设 3.3 中的接近性 (Del Moral，2004年；Cérou 等人，2011年；Whiteley 等人，2012年）同样强。另一类研究（Marion 和 Schmidler，2018年；Lee 和 Santana-Gijzen，2024年）则对转移核作出假设，因此这些假设与本文中的假设不可比较。

我们注意到，如果只假设 χ^2-接近性（定理 3.2），这些方法可能无法捕捉序贯蒙特卡洛的收敛性。这是因为许多方法会显式或隐式地控制方差 $(\mathrm{Var}(W_h))_{h\in[H]}$（Cérou 等人，2011年；Whiteley 等人，2012年；Schweizer，2012年等）。相比之下，我们对定理 3.2 的证明是通过绝对偏差 $\mathbb{E}|M_H - Z|$ 得出的。看似更自然的做法是研究方差 $\mathbb{E}(M_H - Z)^2$，但必须面对这样一个事实：$M_H = \widehat{W}_H = \prod_{h=1}^H\left(\frac{1}{N}\sum_{i=1}^N w_h^i\right)$ 是一个序贯乘积。如下方引理所示，当 χ^2 散度 $(D_{\chi^2}(\pi_h^\star \,\|\, \widehat{\pi}_h))_{h\in[H]}$ 有界时，$\mathrm{Var}(\widehat{W}_h)$ 未必能得到控制。

引理 C.5。 假设 $V(\perp) = 1$。对于任意 $h \in [H]$，均有

$$\mathrm{Var}(\widehat{W}_h) = \mathbb{E}(\widehat{W}_h - \widehat{Z}_h)^2 \ge \frac{\mathrm{Var}_{x_1\sim\pi_1}(\mathbb{E}[\widehat{V}(x_h)\mid x_1])}{N}.$$

特别地，对于 $H \ge 2$，存在一个问题实例 $(V^\star,\ V)$ 使得 $D_{\chi^2}(\pi_h^\star \,\|\, \widehat{\pi}_h) \le 2$，并且假设 *3.1* 和 假设 *3.7* 均在 $C_{\mathsf{act}} = 2$下成立，但使用 N 个粒子的序贯蒙特卡洛会导致

$$\max_{h\in[H]}\mathrm{Var}(\widehat{W}_h) \ge \frac{2^{\lfloor H/2\rfloor} - 2}{N}.$$

因此，如果通过对$\mathrm{Var}(W_h)$的上界来证明序贯蒙特卡洛的收敛性，那么除非我们有 $N \ge \exp(\Omega(H))$ 个粒子，否则收敛速率可能毫无意义。例如，Schweizer (2012年) 通过以下关系研究序贯蒙特卡洛（参见其中的引理 2.1），对于 $f: \mathcal{X}_H \to [0,1]$:

$$\mathbb{E}\left(\mathbb{E}_{\widehat{\nu}_H}[f] - \mathbb{E}_{\pi_H^\star}[f]\right)^2 \le 2\mathrm{Var}(\widehat{W}_H\cdot\mathbb{E}_{\widehat{\nu}_H}[f]) + 2\mathrm{Var}(\widehat{W}_H).$$

因此，Schweizer(2012年) 给出了如下形式的上界

$$N \cdot \mathrm{Var}(\widehat{W}_H \cdot \mathbb{E}_{\widehat{\nu}_H}[f]) \le C_0(f) + C_1(f) \cdot \epsilon_N,$$

其中， C_0， C_1 是两个依赖于问题的泛函，且 ϵ_N 定义为使得

$$\epsilon_N \ge \max_{h \in [H]} \mathrm{Var}(\widehat{W}_h).$$

这意味着，Schweizer(2012年)中的分析无法得到定理 3.2 （以及定理 3.5）的上界。

引理 C.5 的证明对于任意 $h \le \ell$，我们定义 $\widetilde{V}_\ell(x_h) = \mathbb{E}[\widehat{V}(x_\ell) \mid x_h]$ 和 $\rho_\ell(x_h) = \frac{V_\ell(x_h)\,\widehat{V}(x_h)}{}$，并定义 $\widetilde{M}_{h,\ell} = W_h\,\mathbb{E}_{\widehat{\nu}_h}[\rho_\ell]$。不难验证，$(M_{h,\ell})_{h\in[\ell]}$ 是一个鞅，且 $M_{0,\ell} = Z_\ell$ 以及 $\widetilde{M}_{\ell,\ell} = W_\ell$。

因此，我们可以给出下界

$$\begin{aligned}
\mathbb{E}(\widehat{W}_\ell - \widehat{Z}_\ell)^2 &\ge \mathbb{E}(M_{1,\ell} - \widehat{Z}_\ell)^2 \\
&= \mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^N \frac{\widetilde{V}_\ell(x_1^i)}{\widehat{V}(\tilde{x}_0^i)} - \mathbb{E}_{\widehat{\nu}_0}[\rho_\ell]\right)^2\right] \\
&= \frac{1}{N}\Big(\mathbb{E}_{x_1\sim\pi_1}[\widetilde{V}_\ell(x_1)^2] - \widehat{Z}_\ell^2\Big),
\end{aligned}$$

其中注意到，对于 $i \in [N]$， $\tilde{x}_0^i = \perp$。

现在，考虑 $n = \lfloor H/2 \rfloor$。考虑以下简单示例：

- 对于每个 $h \in [H]$， $\mathcal{X}_h = \{(h,0),(h,1)\}$，且 $P((h+1,i) \mid (h,i)) = 1$ 对于 $i \in \{0,1\}$。
- 初始分布由 $\pi_1((1,0)) = \frac{1}{2^n} = \pi_1((1,1))$ 给出。
- 奖励函数为 $r \equiv 1$，因此对于所有 $x \in \mathcal{X}$， $V^\star(x) = 1$。
- 对于每个 $h \in [H]$，我们设定 $\widehat{V}((h,1)) = 1$，并且对于 $h \le n$， $\widehat{V}((h,0)) = 2^h$，而对于 $h \ge n$， $\widehat{V}((h,0)) = 2^{(2n-h)^+}$。

那么，很明显，假设 3.1 和 假设 3.7 均在 $C_{\mathsf{act}} = 2$ 下成立，并且 $\widehat{\pi}_h$ 由 $\widehat{\pi}((h,0)) = \frac{p\widehat{V}((h,0))}{p\widehat{V}((h,0)) + 1 - p}$ 给出，因此 $\widehat{\pi}((h,0)) \in [p, \frac{1}{2}]$。这立即意味着对于任意 $h \in [H]$， $D_{\chi^2}(\pi_h^\star \,\|\, \widehat{\pi}_h) \le 2$。

然而，随后我们可以计算 $\widetilde{V}_n((1,0)) = 2^n$，从而得到 $\mathbb{E}[\widetilde{V}_n(x_1)^2] \ge p \cdot 2^{2n} = 2^n$。这就得到了所需的结论。 □

C.4 定理 3.5的证明

我们证明以下 定理 3.5 的一般形式。

定理 C.6。 序贯蒙特卡洛与 N 粒子对于任意 $M \ge 1$、 $\eta \ge 1$ 均可实现以下结果 ，

$$D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H], \pi_H^\star) \le H\sqrt{\frac{M\eta}{N}} + \sum_{h=1}^H \big[\mathrm{Cov}_M(\pi_h^\star \,\|\, \widehat{\pi}_h) + \mathrm{Cov}_{\eta,h}^{\mathsf{act}}(\pi^\star \,\|\, \pi_{\mathsf{ref}})\big],$$

其中我们定义

$$\begin{aligned}
\mathrm{Cov}_{\eta,h}^{\mathsf{act}}(\pi^\star \,\|\, \pi_{\mathsf{ref}}) &:= \mathbb{P}_{x_{h-1}\sim\pi_{h-1}^\star, x_h\sim\pi^\star(\cdot|x_{h-1})}\left(\frac{\pi^\star(x_h \mid x_{h-1})}{\pi_{\mathsf{ref}}(x_h \mid x_{h-1})} \ge \eta\right) \\
&= \mathbb{P}_{x_{h-1}\sim\pi_{h-1}^\star, x_h\sim\pi^\star(\cdot|x_{h-1})}\left(\frac{V^\star(x_h)}{V^\star(x_{h-1})} \ge \eta\right).
\end{aligned}$$

量 $\mathrm{Cov}^{\mathsf{act}}_{\eta,h}(\pi^\star \,\|\, \pi_{\mathsf{ref}})$ 是 定义 3.4 的动作级类比。在 假设 3.1 下，当 $\eta \geq C_{\mathsf{act}}$ 时，这个附加项为 定理 3.5 的结果。

证明。我们沿用定理 3.2 （附录 C.2）的证明中的记号。回忆

$$M_h - M_{h-1} = \widehat{W}_{h-1}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{V^\star(x_h^i)}{\widehat{V}(\tilde{x}_{h-1}^i)} - \mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}}[\rho(x_{h-1})]\right),$$

其中，元组（$\tilde{x}_{h-1}^i$，x_h^i）在给定 $\mathcal{H}_{h-1}$的条件下是独立同分布的。由命题 B.1，对于任意固定参数 $R>0$，我们可以给出如下界

$$\begin{aligned}
&\frac{1}{\widehat{W}_{h-1}}\,\mathbb{E}_{h-1}|M_h - M_{h-1}| \\
&\leq \sqrt{\frac{R}{N}\,\mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}}\left[\frac{V^\star(x_{h-1})}{\widehat{V}(x_{h-1})}\right]} + 2\,\mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1},x_h\sim\pi_{\mathsf{ref}}(\cdot|x_{h-1})}\left[\frac{V^\star(x_h)}{\widehat{V}(x_{h-1})}\mathbb{I}\left\{\frac{V^\star(x_h)}{\widehat{V}(x_{h-1})}\geq R\right\}\right]
\end{aligned}$$

由引理 C.2，可得

$$\mathbb{E}\left[\widehat{W}_{h-1}\sqrt{\mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}}[\rho(x_{h-1})]}\right] \leq \sqrt{\mathbb{E}[\widehat{W}_{h-1}]\,\mathbb{E}[\widehat{W}_{h-1}\,\mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}}[\rho(x_{h-1})]]} \leq \sqrt{\widehat{Z}_{h-1}\cdot Z},$$

类似地，对于任意 $\eta>0$,

$$\begin{aligned}
&\mathbb{E}\left[\widehat{W}_{h-1}\,\mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1},x_h\sim\pi_{\mathsf{ref}}(\cdot|x_{h-1})}\left[\frac{V^\star(x_h)}{\widehat{V}(x_{h-1})}\mathbb{I}\left\{\frac{V^\star(x_h)}{\widehat{V}(x_{h-1})}\geq R\right\}\right]\right] \\
&= \mathbb{E}_{x_{h-1}\sim\pi_{h-1},x_h\sim\pi_{\mathsf{ref}}(\cdot|x_{h-1})}\left[V^\star(x_h)\mathbb{I}\left\{\frac{V^\star(x_h)}{\widehat{V}(x_{h-1})}\geq R\right\}\right] \\
&\leq \mathbb{E}_{x_{h-1}\sim\pi_{h-1},x_h\sim\pi_{\mathsf{ref}}(\cdot|x_{h-1})}\left[V^\star(x_h)\mathbb{I}\left\{\frac{V^\star(x_{h-1})}{\widehat{V}(x_{h-1})}\geq \eta^{-1}R\right\} + V^\star(x_h)\mathbb{I}\left\{\frac{V^\star(x_h)}{V^\star(x_{h-1})}\geq\eta\right\}\right] \\
&= Z\cdot\left[\mathbb{P}_{x_{h-1}\sim\pi^\star_{h-1}}\left(\frac{V^\star(x_{h-1})}{\widehat{V}(x_{h-1})}\geq\eta^{-1}R\right) + \mathbb{P}_{x_{h-1}\sim\pi^\star_{h-1},x_h\sim\pi^\star(\cdot|x_{h-1})}\left(\frac{V^\star(x_h)}{V^\star(x_{h-1})}\geq\eta\right)\right],
\end{aligned}$$

其中我们对 $x_{h-1}\in\mathcal{X}_{h-1}$ 使用 $\pi_{h-1}(x_{h-1})\cdot V^\star(x_{h-1}) = \pi^\star_{h-1}(x_{h-1})\cdot Z$，并对 $=\pi^\star(x_h\,|x_{h-1})\cdot Z$ 使用 $\pi_{\mathsf{ref}}(x_h\mid x_{h-1})\cdot V^\star(x_h)$。因此，我们可以选择 $R=M\eta\frac{Z}{\widehat{Z}_{h-1}}$，从而推出

$$\begin{aligned}
\frac{1}{Z}\,\mathbb{E}|M_h - M_{h-1}| \leq \sqrt{\frac{M\eta}{N}} &+ 2\mathbb{P}_{x_{h-1}\sim\pi^\star_{h-1}}\left(\frac{V^\star(x_{h-1})}{\widehat{V}(x_{h-1})}\geq M\cdot\frac{Z}{\widehat{Z}_{h-1}}\right) \\
&+ 2\mathbb{P}_{x_{h-1}\sim\pi^\star_{h-1},x_h\sim\pi^\star(\cdot|x_{h-1})}\left(\frac{V^\star(x_h)}{V^\star(x_{h-1})}\geq\eta\right).
\end{aligned}$$

求和即可完成证明。 □

C.5 定理 3.6 的证明

我们先证明以下引理。

引理 C.7. 在 假设 *3.3* 和 假设 *3.1*下，对于任意 $\delta\in(0,1)$，如果我们使用序贯蒙特卡洛（算法 *1*），并采用 $N=\Omega(HC_\infty^4 C_{\mathsf{act}}^2\log(H/\delta))$ 个粒子，那么以至少 $1-\delta$的概率，有 $|\widehat{W}_H - Z|\leq Z/2$。

证明。 我们沿用定理 3.2(附录 C.2的证明中的记号。回想一下，

$$M_h - M_{h-1} = \widehat{W}_{h-1}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{V^\star(x_h^i)}{\widehat{V}(\tilde{x}_{h-1}^i)} - \mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}}[\rho(x_{h-1})]\right),$$

其中，元组 $(\tilde{x}^i_{h-1},\ x^i_h)$ 在给定 $\mathcal{H}_{h-1}$的条件下独立同分布。注意，根据 假设 3.1，几乎必然有

$$0 \le \frac{V^\star(x^i_h)}{\widehat{V}(\tilde{x}^i_{h-1})} = \frac{V^\star(x^i_h)}{V^\star(\tilde{x}^i_{h-1})} \cdot \frac{V^\star(\tilde{x}^i_{h-1})}{\widehat{V}(\tilde{x}^i_{h-1})} \le C_{\mathsf{act}}\mathsf{C}_\infty.$$

利用随机变量 $Y \in [0, A]$ 的性质，我们可以界定 $\mathbb{E}[e^{\lambda(Y-\mathbb{E}[Y])}] \le e^{\lambda^2 A^2/8}$，因此对于任意 $\lambda \in \mathbb{R}$,

$$\begin{aligned}
&\mathbb{E}_{h-1}[\exp(\lambda(M_h - M_{h-1}))] \\
=&\mathbb{E}_{h-1}\left[\exp\left(\lambda\widehat{W}_{h-1}\cdot\left(\frac{1}{N}\sum_{i=1}^N \frac{V^\star(x^i_h)}{\widehat{V}(\tilde{x}^i_{h-1})} - \mathbb{E}_{\widehat{\nu}_{h-1}}[\rho(x_{h-1})]\right)\right)\right] \\
\le&\exp\left(\frac{1}{N}\lambda^2\mathsf{C}_\infty^2 C_{\mathsf{act}}^2\widehat{W}_{h-1}^2\right) \le \exp\left(\frac{1}{N}\lambda^2\mathsf{C}_\infty^4 C_{\mathsf{act}}^2 M_{h-1}^2\right).
\end{aligned}$$

根据 Freedman不等式，对于任意 $h \in [H]$，以概率 $1-\delta$，我们有

$$|M_h - Z| \le \sum_{i=1}^{h-1}\frac{\lambda\mathsf{C}_\infty^4 C_{\mathsf{act}}^2 M_i^2}{N} + \frac{\log(2/\delta)}{\lambda}.$$

取 $\lambda = 4\log(2 \quad H/\delta)/Z$并利用并集界，可得以概率 $1-\delta$，对于所有 $h \in [H \quad]$,

$$|M_h - Z| \le \frac{4\log(2/\delta)\mathsf{C}_\infty^4 C_{\mathsf{act}}^2}{N}\sum_{i=0}^{h-1}\frac{M_i^2}{Z} + \frac{Z}{4}. \tag{6}$$

由于我们选择了 $N = \Omega(H\log(H/\delta)\mathsf{C}_\infty^4 C_{\mathsf{act}}^2)$，由对 h 进行归纳法可知，在 式 (6) 成立这一事件下，我们有 $|M_h - Z| \le Z/2$ 对于所有 $h \in [H]$。特别地，这意味着 $|W_H - Z| \le Z/2$ 其概率至少为 $1-\delta$。 □

定理 定理 3.6 的证明。 在任意给定的一轮中，令 $\widehat{p}$ 为 算法 1 使用选项 2 时的输出分布（当 算法 1重新启动时，我们将其视为输出符号“重新启动”）。那么，算法输出任意 $x \in \mathcal{X}_H$ 的概率为

$$\widehat{p}(x) := \mathbb{E}\left[\min\left\{1, \frac{\widehat{W}_H}{2\mathsf{C}_\infty\widehat{Z}_1}\right\}\cdot\widehat{\nu}_H(x)\right] \le \frac{Z\cdot\pi^\star_H(x)}{2\mathsf{C}_\infty\widehat{Z}_1},$$

其中我们使用了引理 C.2。此外，利用$2\mathsf{C}_\infty\widehat{Z}_1 \ge 2Z \ge \widehat{W}_H$ 以概率$1 - {}_2{}^{\delta}\,\mathsf{C}_\infty^2$ 由引理 C.7，对于任意子集 $X \subset \mathcal{X}_H$,

$$\widehat{p}(X) = \mathbb{E}\left[\min\left\{1, \frac{\widehat{W}_H}{2\mathsf{C}_\infty\widehat{Z}_1}\right\}\cdot\widehat{\nu}_H(X)\right] \ge \frac{Z\cdot\pi^\star_H(x)}{2\mathsf{C}_\infty\widehat{Z}_1} - \frac{\delta}{2\mathsf{C}_\infty^2}.$$

特别地，成立$2 \quad {}^{Z}\,\mathsf{C}_\infty\widehat{z}_1 \quad - \quad {}_4{}^{\delta}\,\mathsf{C}_\infty^2 \le \widehat{p}(x \neq \text{重启}) \le {}_2 \quad {}^{Z}\,\mathsf{C}_\infty\widehat{z}_1$，因此利用 $\widehat{Z}_1/Z \le \mathsf{C}_\infty$，我们知道对于任意 $X \subset \mathcal{X}_H$

$$\pi^\star_H(X) - \delta \le \widehat{p}(X \mid \text{not restart}) \le (1+\delta)\pi^\star_H(X),$$

因此 $D_{\mathsf{TV}}(\pi^\star_H, \widehat{p}(\cdot \mid \text{不重新启动})) \le \delta$。最后，利用拒绝采样的保证，并注意到 $\widehat{Z}_1/Z \le \mathsf{C}_\infty$，我们知道外层循环步数上界为 $O(\mathsf{C}_\infty^2\log(1/\delta))$。 □

C.6 具有精确值函数的 SMC 下界

在下面的命题中，我们将证明，当 $V = V^\star$ 时，序贯蒙特卡洛对时域的依赖并非分析过程造成的假象。

命题 C.8. 存在一个问题实例，使得 $V = V^\star$ 且 假设 *3.1* 在 $C_{\mathsf{act}} = 2$下成立；然而，具有 N 个粒子的序贯蒙特卡洛必须承担

$$D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H], \pi_H^\star) \geq c \min\left\{1, \frac{\sqrt{H}}{N}\right\}.$$

证明。 我们按如下方式构造一个问题实例。固定 $\lambda > 0$。我们考虑 $\mathcal{X}_h = \{0,1\}^h$，并定义转移核 P和价值函数 $V = V^\star$ 如下：

- 对于任意 x，令 $\pi_{\mathsf{ref}}(\cdot \mid x) = \mathsf{unif}(\{(x,0), (x, 1)\})$。
- 令 $V = V^\star : \mathcal{X} \to \mathbb{R}$ 递归地定义为

$$V^\star(\perp) = 0, \qquad V^\star(x,0) = V^\star(x), \quad V^\star(x,1) = (1+\lambda)V^\star(x).$$

- 那么，有 $C_{\mathsf{act}} \leq 1 + \lambda$，且 $\pi_H^\star = \mathsf{Ber}(p)^{\otimes H}$，其中 $p = \frac{1+\lambda}{2+\lambda}$。 ——

在这一构造下，我们可以将序贯蒙特卡洛改写如下（其中 $\widehat{\nu}_0 = \delta_\perp$）。

- 对于 $h \geq 1$，采样 N个粒子 $\tilde{x}_{h-1}^1$、……、$\tilde{x}_h^N$独立同分布$\sim \widehat{\nu}_{h-1}(\cdot)$，并独立生成 $a_h^i \sim$均匀分布$(\{0,1\})$。
- 为粒子 $x_h^i = (\tilde{x}_{h-1}^i, a_h^i)$赋予权重 $w_h^i = 1 + \lambda a_h^i$。
- 定义加权经验测度 $\widehat{\nu}_h := \frac{1}{W_h} \sum_{i=1}^N w_h^i \delta_{x_h^i}$，其中 $W_h = N + \lambda \sum_{i=1}^N a_h^i$。

我们提出以下断言，并将在下文证明。

断言 C.9。 有 $\mathbb{E}[\widehat{\nu}_H] =$伯努利$(p_N)^{\otimes H}$，其中 p_N定义见 公式 *(7)*。

断言 C.10. </style}它成立，即，其中 $(1+\lambda)\ (2+\lambda)^3$。 ———

现在，我们可以给出下界

$$D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H], \pi_H^\star) = D_{\mathsf{TV}}(\mathrm{Bin}(H,p), \mathrm{Bin}(H,p_N)) \geq c \min\{1, \sqrt{H}|p - p_N|\},$$

其中 $c > 0$ 是一个绝对常数。取 $\lambda = 1$ 即可完成 命题 C.8 的证明。 □

我们证明以下辅助断言：

断言 C.9的证明。 根据定义，对于任意 $x \in \mathcal{X}_{h-1}$ 和 $a \in \{0,1\}$，

$$\widehat{\nu}_h(x,1) = \frac{1}{W_h} \sum_{i=1}^N w_h^i \mathbb{I}\{\tilde{x}_{h-1}^i = x, a_h^i = a\}.$$

注意到 $\{a_h^i\}$ 和 $\{\tilde{x}_{h-1}^i\}$ 相互独立，我们可以计算

$$\begin{aligned}\mathbb{E}[\widehat{\nu}_h(x,1)] &= \mathbb{E}\left[\frac{1}{W_h} \sum_{i=1}^N w_h^i \mathbb{I}\{a_h^i = 1\} \cdot \mathbb{I}\{\tilde{x}_{h-1}^i = x\}\right] \\ &= \mathbb{E}\left[\widehat{\nu}_{h-1}(x) \cdot \frac{(1+\lambda)\sum_{i=1}^N a_h^i}{N + \lambda \sum_{i=1}^N a_h^i}\right].\end{aligned}$$

注意， $A_h = \sum_{i=1}^N a_h^i \sim$ 二项分布$(N,\ 1\ 2)$ 与 $\widehat{\nu}_{h-1}$ 独立。于是，我们知道

$$\mathbb{E}[\widehat{\nu}_h(x,1)] = p_N\, \mathbb{E}[\widehat{\nu}_{h-1}(x)], \qquad p_N = (1+\lambda)\, \mathbb{E}_{A \sim \mathrm{Bin}(N, \frac{1}{2})}\left[\frac{A}{N + \lambda A}\right]. \tag{7}$$

递归地应用此方程即可得到所需结果。 □

断言的证明 断言 C.10。 令 $Y = \frac{2A}{N} - 1$使得 Y_h 在 $[-1,1]$上对称分布。于是我们可以写出

$$\frac{N}{N+\lambda A} = \frac{2}{2+\lambda(Y+1)} = \frac{2}{2+\lambda} \cdot \frac{1}{1+\frac{\lambda}{2+\lambda}Y} = \frac{2}{2+\lambda} \sum_{k\geq 0} (-1)^k \left(\frac{\lambda Y}{2+\lambda}\right)^k.$$

注意，对于奇整数 k 和 $\mathbb{E}\, Y^2 = \frac{1}{N}$，我们可以得出 $\mathbb{E}\, Y^k = 0$

$$\mathbb{E}\left[\frac{N}{N+\lambda A}\right] \geq \frac{2}{2+\lambda}\left(1 + \left(\frac{\lambda}{2+\lambda}\right)^2 \frac{1}{N}\right).$$

重新整理可得

$$p_N = (1+\lambda)\,\mathbb{E}\left[\frac{A}{N+\lambda A}\right] = \frac{1+\lambda}{\lambda}\left(1 - \mathbb{E}\left[\frac{N}{N+\lambda A}\right]\right) \leq \frac{1+\lambda}{2+\lambda} - \frac{2\lambda(1+\lambda)}{(2+\lambda)^3 N}.$$

□

D SMC-RS 的证明

与 附录 C 类似，我们为 算法 2 引入以下量：

$$W_h := \mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}, x_h\sim\pi_{\mathsf{ref}}(\cdot|x_{h-1})}\left[\frac{\widehat{V}(x_h)}{\widehat{V}(x_{h-1})}\right], \qquad \widehat{W}_h := \prod_{\ell=1}^{h} W_\ell.$$

我们定义 $\mathcal{H}_h = \{\mathcal{S}_1, \cdots, \mathcal{S}_h\}$ 为 SMC-RS（算法 2）执行至第 h 次迭代结束时的完整历史。我们记 $\mathbb{E}_h[\cdot]$ 为关于 $\mathcal{H}_h$的条件分布。

引理 D.1。 对于 $h \in [H]$ 步的 *SMC-RS*（算法 *2*），给定 $\mathcal{H}_{h-1}$，集合 $\mathcal{S}_h$ 由来自分布 $\widehat{\mu}_h$ 的独立同分布样本组成，该分布为

$$\widehat{\mu}_h(x_h) = \frac{1}{W_h}\,\mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}}\left[\pi_{\mathsf{ref}}(x_h \mid x_{h-1})\frac{\widehat{V}(x_h)}{\widehat{V}(x_{h-1})}\right], \qquad \forall x_h \in \mathcal{X}_h.$$

因此，对于任意 $g_h : \mathcal{X}_h \to \mathbb{R}$，有

$$\mathbb{E}\left[\widehat{W}_h \cdot \mathbb{E}_{\widehat{\nu}_h}[g_h]\right] = \mathbb{E}\left[\widehat{W}_h \cdot \mathbb{E}_{\widehat{\mu}_h}[g_h]\right] = \frac{1}{\widehat{Z}_1}\,\mathbb{E}_{x_h\sim\pi_h}\left[\widehat{V}(x_h) g_h(x_h)\right]. \tag{8}$$

引理 D.1的证明。 第一个结论直接由拒绝采样的分析得出：对于任意 $x_h \in \mathcal{X}_h$，根据贝叶斯法则，有

$$\mathbb{P}(x_h \mid \text{accept}) \propto \mathbb{P}(\text{accept} \mid x_h) \cdot \mathbb{P}(x_h) \propto \widehat{\mu}_h(x_h).$$

为证明第二个论断，我们证明以下结果，由此式（8）可通过归纳法得到：

$$\mathbb{E}_{h-1}[W_h \cdot \mathbb{E}_{\widehat{\nu}_h}[g_h]] = \mathbb{E}_{h-1}[W_h \cdot \mathbb{E}_{\widehat{\mu}_h}[g_h]] = \mathbb{E}_{x_{h-1}\sim\widehat{\nu}_{h-1}, x_h\sim\pi_{\mathsf{ref}}(\cdot|x_{h-1})}\left[\frac{\widehat{V}(x_h) g_h(x_h)}{\widehat{V}(x_{h-1})}\right].$$

□

特别地，对于 $h \geq 1$， $\mathbb{E}[W_h] = Z_h/Z_1$ 且

$$\mathbb{E}\left[\widehat{W}_H \widehat{\nu}_H(x)\right] = Z \cdot \pi_H^\star(x)/\widehat{Z}_1, \qquad \forall x \in \mathcal{X}_H.$$

进一步，我们定义 $\rho(x) = V^\star(x)\widehat{V}(x)$ 以及

$$M_h = \widehat{W}_h \cdot \mathbb{E}_{x_h \sim \widehat{\nu}_h}[\rho(x_h)].$$

那么 引理 D.1 与 $g(x) = \rho(x)$ 蕴含 (M_h) 是一个鞅，且 $M_0 = Z$ 并且 $M_H = W_H$。因此，类似于 引理 C.3，我们得到以下上界。

引理 D.2. 对于 算法 *2*，有

$$D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H], \pi_H^\star) \leq \frac{1}{2Z}\,\mathbb{E}|\widehat{W}_H - Z| \leq \frac{1}{2Z}\sum_{h=1}^{H} \mathbb{E}|M_h - M_{h-1}|.$$

D.1 定理 3.8 的证明

注意到 $M_h - M_{h-1} = W_h(\mathbb{E}_{\widehat{\nu}_h}[\rho] - \mathbb{E}_{\widehat{\mu}_h}[\rho])$，且 $\widehat{\nu}_h = \mathsf{unif}(\mathcal{S}_h)$，其中 $\mathcal{S}_h$ 由来自 $\widehat{\mu}_h$ 的独立同分布样本组成。然后，我们可以界定

$$\mathbb{E}_{h-1}|\mathbb{E}_{\widehat{\nu}_h}[\rho_h] - \mathbb{E}_{\widehat{\mu}_h}[\rho_h]| \leq \sqrt{\mathbb{E}_{h-1}\left|\frac{1}{N}\sum_{x_h \in \mathcal{S}_h} \rho(x_h) - \mathbb{E}_{\widehat{\mu}_h}[\rho]\right|^2} = \sqrt{\frac{\mathbb{E}_{\widehat{\mu}_h}[\rho(x_h)^2] - (\mathbb{E}_{\widehat{\mu}_h}[\rho(x_h)])^2}{N}},$$

因此

$$\begin{aligned}
\sqrt{N}\,\mathbb{E}|M_h - M_{h-1}| &\leq \mathbb{E}\left[\widehat{W}_h \sqrt{\mathbb{E}_{\widehat{\mu}_h}[\rho(x_h)^2] - (\mathbb{E}_{\widehat{\mu}_h}[\rho(x_h)])^2}\right] \\
&\leq \sqrt{\mathbb{E}[\widehat{W}_h]\left(\mathbb{E}\left[\widehat{W}_h\,\mathbb{E}_{\widehat{\mu}_h}[\rho(x_h)^2]\right] - \mathbb{E}\left[\widehat{W}_h(\mathbb{E}_{\widehat{\mu}_h}[\rho(x_h)])^2\right]\right)} \\
&\leq \sqrt{\widehat{Z}_h\,\mathbb{E}_{x_h \sim \pi_h}\left[\frac{V^\star(x_h)^2}{\widehat{V}(x_h)}\right] - Z^2} \\
&= Z\sqrt{\mathbb{E}_{x_h \sim \pi_h^\star}\left[\frac{\pi^\star(x_h)}{\widehat{\pi}(x_h)}\right] - 1} = Z\sqrt{D_{\chi^2}(\pi_h^\star \,\|\, \widehat{\pi}_h)},
\end{aligned}$$

其中第三行使用了 $\mathbb{E}[\widehat{W}_h]\,\mathbb{E}\left[\widehat{W}_h(\mathbb{E}_{\widehat{\mu}_h}[\rho(x_h)])^2\right] \geq \left(\mathbb{E}[\widehat{W}_h\,\mathbb{E}_{\widehat{\mu}_h}[\rho(x_h)]]\right)^2 = Z^2$ (柯西不等式和引理 D.1)，最后一行则根据 $\pi^\star$、 $\widehat{\pi}$ 和 χ^2-散度的定义得出。求和后即完成证明。

□

D.2 重尾误差下 SMC-RS 的分析

与 SMC 类似，我们证明 SMC-RS 也具有仅取决于覆盖度的上界（定义 3.4）。

定理 D.3。 在 假设 *3.7* 下，假设 *SMC-RS*（算法 *2*）采用 $\eta \geq C_{\mathsf{act}}$ 进行实例化。则对于任意 $M \geq 1$，算法 *2* 达到

$$D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H], \pi_H^\star) \leq H\sqrt{\frac{M}{N}} + \sum_{h=1}^{H} \mathrm{Cov}_M(\pi_h^\star \,\|\, \widehat{\pi}_h).$$

定理 D.3 的收敛速度比序贯蒙特卡洛（定理 3.5）提高了 C_{act}倍，不过时间复杂度中也会再次出现相同的倍数。期望时间复杂度上界与定理 3.8一致。

证明。 我们沿用定理 3.8 (附录 D.1）的证明。根据命题 B.1，有

$$\mathbb{E}_{h-1}|\mathbb{E}_{\widehat{\nu}_h}[\rho_h] - \mathbb{E}_{\widehat{\mu}_h}[\rho_h]| = \mathbb{E}_{h-1}\left|\frac{1}{N}\sum_{x_h\in\mathcal{S}_h}\rho(x_h) - \mathbb{E}_{\widehat{\mu}_h}[\rho]\right|$$
$$\leq \sqrt{\frac{R}{N}\,\mathbb{E}_{x_h\sim\widehat{\mu}_h}[\rho(x_h)]} + 2R\,\mathbb{E}_{x_h\sim\widehat{\mu}_h}[\rho(x_h)\mathbb{I}\{\rho(x_h)\geq R\}],$$

其中，我们使用 $\mathcal{S}_h$，它由来自 $\widehat{\mu}_h$ 的 N 个独立同分布样本组成。因此，我们可以界定

$$\mathbb{E}|M_h - M_{h-1}| \leq 2\sqrt{\frac{R}{N}}\,\mathbb{E}\left[\widehat{W}_h\sqrt{\mathbb{E}_{x_h\sim\widehat{\mu}_h}[\rho(x_h)]}\right] + 2\,\mathbb{E}\left[\widehat{W}_h\,\mathbb{E}_{x_h\sim\widehat{\mu}_h}[\rho(x_h)\mathbb{I}\{\rho(x_h)\geq R\}]\right]$$

此外，根据 引理 D.1，我们有

$$\mathbb{E}\left[\widehat{W}_h\sqrt{\mathbb{E}_{x_h\sim\widehat{\mu}_h}[\rho(x_h)]}\right] \leq \sqrt{\mathbb{E}[\widehat{W}_h]\,\mathbb{E}[\widehat{W}_h\,\mathbb{E}_{x_h\sim\widehat{\mu}_h}[\rho(x_h)]]} \leq \sqrt{\widehat{Z}_h\cdot Z},$$

and

$$\mathbb{E}\left[\widehat{W}_h\cdot\mathbb{E}_{x_h\sim\widehat{\mu}_h}[\rho(x_h)\mathbb{I}\{\rho(x_h)\geq R\}]\right] = Z\cdot\mathbb{P}_{x\sim\pi_h^\star}(\rho(x_h)\geq R).$$

选择 $R = M\frac{Z}{\widehat{Z}_h}$ 可得

$$\mathbb{E}|M_h - M_{h-1}| \leq 2Z\sqrt{\frac{M}{N}} + 2Z\mathrm{Cov}_M(\pi_h^\star\,\|\,\widehat{\pi}_h).$$

对其求和即可完成证明。 □

D.3 SMC-RS 的更快收敛

在本节中，我们提出并分析 SMC-RS 的一种修改版本，称为带重启的 SMC-RS（算法 3）。在以下假设下，我们证明算法 3 实现了指数收敛（类似于 定理 3.6 中关于序贯蒙特卡洛的结论）。

假设 D.4. 对于每个 x_h，至多有 η 个 x_{h+1} 值，使得 $\pi_{\mathsf{ref}}(x_{h+1}\mid x_h) > 0$，且这些值可以高效枚举。

算法 3 扩散蒙特卡洛

输入： 转移核 π_{ref}，价值函数 V。参数：初始样本量 $N\geq 1$。初始化 $\mathcal{S}_0$，使其为 N 个 $\perp$ 的副本。对于 $h=1、\cdots、H$，执行
对于每个 $x_{h-1}\in\mathcal{S}_{h-1}$，计算 $V(x_{h-1}) = \mathbb{E}_{\pi_{\mathsf{ref}}}[V(x_h)\mid x_{h-1}]$。定义 $W_{h-1} = \frac{1}{N}\sum_{x_{h-1}\in\mathcal{S}_{h-1}}\frac{V}{\widehat{V}}(x_{h-1})(x_{h-1})$。定义加权经验测度 $\widetilde{\mu}_{h-1} = \frac{1}{NW_{h-1}}\sum_{x_{h-1}\in\mathcal{S}_{h-1}}\frac{V}{\widehat{V}}(x_{h-1})(x_{h-1})\delta_{x_{h-1}}$。为 $i\in[N]$ 独立生成 $\tilde{x}_{h-1}^i\sim\widetilde{\mu}_{h-1}$、$x_h^i\sim\widehat{\pi}(\cdot\mid\tilde{x}_{h-1}^i)$。定义 $\mathcal{S}_h = \{x_h^i\}_{i\in[N]}$ 和 $\widehat{\nu}_h =$Unif$(\mathcal{S}_h)$。

定义 $\widehat{W}_H = \prod_{h=1}^H W_h$ 以概率 $\min\left\{\frac{\widehat{W}_H}{\widehat{Z}_1},\ 1\right\}$ 输出 $x\sim\widehat{\nu}_H$，否则重新启动该算法

输出 m.

我们证明以下保证：

定理 D.5. 假设 假设 *3.3* 和 假设 *D.4* 均成立，并记 $\epsilon := \mathsf{C}_\infty - 1$。此外，~假设我们有 $Z \le Z_1 \le JZ$。那么，算法 *3* 采用输出选项 ***2*** 从满足以下条件的分布中进行采样μ:

$$D_{\mathsf{TV}}(\mu, \pi_H^\star) \le \delta,$$

只要 $N \ge \Omega(H\mathsf{C}_\infty^2\epsilon^2\log(JH/\delta))$，使用的外层循环步数上界为 $O(J^2\log(1/\delta))$。

在 假设 D.4 下，我们总能选择 $\widetilde{Z}_1 = \mathsf{C}_\infty\widehat{Z}_1$，使得 $Z \le \widetilde{Z}_1 \le \mathsf{C}_\infty^2 Z$。此外，我们可以单独运行 ~ ^ ~算法 3 1 次，仅用于估计 Z，所选的 $Z_1 = 2W_H$ 满足 $Z \le Z \le 4Z$ 以高概率成立。

类似于定理 3.6的证明，我们首先陈述并证明以下引理（作为引理 C.7的对应版本）。定理 D.5随后可由附录 C.5中的论证推出。

引理 D.6。在定理 *D.5*的条件下，对于任意 $\delta \in (0,1)$，若使用 算法 *3*运行 *SMC-RS*，且使用 $N = \Omega(H\mathsf{C}_\infty^2\epsilon^2\log(H/\delta))$ 个粒子，则以至少 $1-\delta$的概率，有 $|\widehat{W}_H - Z| \le Z/2$。

证明 引理 D.6。 我们沿用 定理 3.8 （附录 D.1）的证明。回顾

$$\frac{M_h - M_{h-1}}{\widehat{W}_h} = \frac{1}{N}\left(\sum_{i=1}^N \rho(x_h^i) - \mathbb{E}_{\widehat{\mu}_h}[\rho]\right),$$

其中 $x_h^1, \cdots, x_h^N$是来自 $\widehat{\mu}_h$（给定 $\mathcal{H}_{h-1}$）的 N独立同分布样本。注意，由{style id='11'>假设 3.3可知，对于任意 $x_h \in \mathcal{X}_h$，都有$\rho(x_h) - 1 \in [-\epsilon, \epsilon]$。这意味着 $\mathbb{E}_{x_h\sim\widehat{\mu}_h}[e^{\lambda(\rho(x_h) - \mathbb{E}_{\widehat{\mu}_h}[\rho])}] \le e^{\lambda^2\epsilon^2/8}$对于 $\lambda \in \mathbb{R}$成立，因此

$$\begin{aligned}
&\mathbb{E}_{h-1}[\exp(\lambda(M_h - M_{h-1}))] \\
=&\mathbb{E}_{h-1}\left[\exp\left(\lambda\widehat{W}_h \cdot \left(\frac{1}{N}\left(\sum_{i=1}^N \rho(x_h^i) - \mathbb{E}_{\widehat{\mu}_h}[\rho]\right)\right)\right)\right] \\
\le&\exp\left(\frac{1}{N}\lambda^2\epsilon^2\widehat{W}_h^2\right) \le \exp\left(\frac{1}{N}\lambda^2\mathsf{C}_\infty^2\epsilon^2 M_{h-1}^2\right),
\end{aligned}$$

其中我们使用了 $M_{h-1} = \mathbb{E}_{h-1}[M_h] = \widehat{W}_h\,\mathbb{E}_{\widehat{\mu}_h}[\rho] \ge \mathsf{C}_\infty^{-1}\widehat{W}_h$。

作者: Fr ，可知对于任意 $h \in [H]$，以概率 $1-\delta$，有 th at

$$|M_h - Z| \le \sum_{i=1}^{h-1}\frac{\lambda\mathsf{C}_\infty^2\epsilon^2 M_i^2}{N} + \frac{\log(2/\delta)}{\lambda}.$$

选取 $\lambda = 4\log(2H/\delta)/Z$ 并应用并集界，可得以概率 $1-\delta$，对所有 $h \in [H]$,

$$|M_h - Z| \le \frac{4\log(2/\delta)\mathsf{C}_\infty^2\epsilon^2}{N}\sum_{i=0}^{h-1}\frac{M_i^2}{Z} + \frac{Z}{4}.$$

由于我们选取了 $N = \Omega(H\log(H/\delta)\mathsf{C}_\infty^2\epsilon^2)$，由 h 上的归纳法可知，在 成立的事件中，我们有 $|M_h - Z| \le Z/2$ 对所有 $h \in [H]$。特别地，这意味着 $|W_H - Z| \le Z/2$ 以至少 $1-\delta$ 的概率成立。

□

E 关于回溯与粒子滤波的统一视角

我们的主要理论结果表明，在关于^的假设相同的情况下，粒子滤波方法（如序贯蒙特卡洛）能够获得与 V 相当的强Rohatgi 等人(2025年提出的基于回溯的方法 VGB 相同的保证，同时将并行运行时间从 $\widetilde{O}(H^2)$ 缩短至 $\widetilde{O}(H)$。本节将说明二者的联系更为深刻：对于任意问题实例，VGB 的执行过程都可以与 SMC-IND 的执行过程耦合起来。SMC-IND 是序贯蒙特卡洛的一种变体，其中每个粒子的子粒子都是独立采样的。

我们将在陈述结果（定理 E.1）之后再给出形式化定义；这里先作几点说明以提供背景。下文将讨论范围限定在自回归设定（备注 2.1）中，VGB 即是在这一设定下由Rohatgi 等人(2025年定义的。回顾一下，VGB 实现的是一种单粒子随机游走，其允许的转移不仅可以向当前粒子添加一个词元（即 $a_{1:h} \mapsto a_{1:h+1}$），还可以进行回溯（即 $a_{1:h} \mapsto a_{1:h-1}$）。出于技术上的考虑，我们在随机游走的状态空间中增添了一个状态 $\mathfrak{s}$，它仅与空字符串相邻 $\perp$。关于 SMC-IND 和 VGB 随机游走的形式化定义，请参见附录 E.1和E.2 。

定理 E.1。 设 $\mathcal{S}_{\mathsf{DMC}}$ 为 *SMC-IND* 在具有 N 初始粒子的情况下访问过的（随机）粒子多重集。设 $\mathcal{S}_{\mathsf{VGB}}$ 为 *VGB*（从 $\mathfrak{s}$ 开始）在第 $N+1$ 次访问 $\mathfrak{s}$ 之前访问过的（随机）粒子多重集。那么存在一种耦合，使得 $\mathcal{S}_{\mathsf{DMC}} = \mathcal{S}_{\mathsf{VGB}} \setminus \{\mathfrak{s}\}$（计入重数）几乎必然成立。

直观地说，SMC-IND 访问的每个粒子 $a_{1:h}$ 都对应一个时间区间：VGB 随机游走从 $a_{1:h}$出发，在 $a_{1:h}$结束，并且从未访问过父节点 $a_{1:h-1}$（其中，我们将 $\perp$ 的父节点视为$\mathfrak{s}$）。这表明，回溯以抑制误差放大的范式 Rohatgi 等人 (2025年) 从某种意义上说，等价于一种粒子滤波。

附录概要。 在附录 E.1中，我们介绍 VGB 随机游走（出于技术原因略作修改）。在 附录 E.2 中，我们介绍 SMC-IND 随机过程。在 附录 E.3 中，我们证明定理 E.1。

E.1 VGB 背景

（非懒惰）VGB 随机游走 Rohatgi 等人 （2025年）定义在状态空间 $\mathcal{X} = \mathcal{X}_0 \sqcup \cdots \sqcup \mathcal{X}_H$ 上，其中 $\mathcal{A}$是词元空间，并且 $\mathcal{X}_h := \mathcal{A}^h$。出于技术上的便利（具体而言，是为了在将 VGB 与 SMC-IND 进行耦合时避免第 0 步的分情况讨论），我们引入一个辅助状态 $\mathfrak{s}$，它并不在原始定义中。我们将转移核 $\mathbb{P}_{\mathsf{VGB}}$ 定义如下。首先，$\mathfrak{s}$ 总是转移到 $\perp$:

$$\mathbb{P}_{\mathsf{VGB}}(\perp \mid \mathfrak{s}) = 1.$$

其次，任一 $a_{1:h} \in \mathcal{A}^h$ （满足 $0 \le h < H$）都可以转移到“子节点” $a_{1:h+1}$ 或“父节点” $a_{1:h-1}$:

$$\mathbb{P}_{\mathsf{VGB}}(a_{1:h+1} \mid a_{1:h}) = \frac{\pi_{\mathsf{ref}}(a_{h+1} \mid a_{1:h})\widehat{V}(a_{1:h+1})}{\widehat{V}(a_{1:h}) + \sum_{a'_{h+1}} \pi_{\mathsf{ref}}(a'_{h+1} \mid a_{1:h})\widehat{V}(a_{1:h}, a'_{h+1})},$$

$$\mathbb{P}_{\mathsf{VGB}}(a_{1:h-1} \mid a_{1:h}) = \frac{\widehat{V}(a_{1:h})}{\widehat{V}(a_{1:h}) + \sum_{a'_{h+1}} \pi_{\mathsf{ref}}(a'_{h+1} \mid a_{1:h})\widehat{V}(a_{1:h}, a'_{h+1})},$$

约定 $a_{1:0} = \perp$ （即空字符串）以及 $a_{1:-1} = \mathfrak{s}$。第三，任一 $a_{1:H} \in \mathcal{A}^H$ 都必须转移到其父节点:

$$\mathbb{P}_{\mathsf{VGB}}(a_{1:H-1} \mid a_{1:H}) = 1.$$

原始 VGB 随机游走的转移核Rohatgi 等人 （2025年）在从不访问 $\mathfrak{s}$ 的条件下是相同的。此外，由于 $\mathfrak{s}$ 只能从 $\perp$ 到达（并且只转移到 $\perp$），我们的修改对 VGB 执行的影响仅体现为：随机地将序列 （$\mathfrak{s}$ $\perp$）的副本插入 VGB 的随机轨迹中。

为便于记号，宜将转移核 $\mathbb{P}_{\mathsf{down}}$ 定义为以“向下”移动为条件的 VGB 转移核，即

$$\mathbb{P}_{\mathsf{down}}(a_{1:h+1} \mid a_{1:h}) = \frac{\pi_{\mathsf{ref}}(a_{h+1} \mid a_{1:h})\widehat{V}(a_{1:h+1})}{\sum_{a'_{h+1}} \pi_{\mathsf{ref}}(a'_{h+1} \mid a_{1:h})\widehat{V}(a_{1:h}, a'_{h+1})}.$$

E.2 SMC-IND 的定义

在算法 4中，我们描述 SMC-IND 如何随机构造给定问题实例的粒子多重集 $\mathcal{S}_0, \ldots, \mathcal{S}_H$。请注意，这里没有指定输出机制，因为 SMC-IND 与 VGB 之间的耦合将应用于整个粒子多重集。

算法 4 采用独立采样的序贯蒙特卡洛（SMC-IND）

输入：自回归语言模型 π_{ref}，价值函数 V。参数：初始样本量 $N \geq 1$。将 $\mathcal{S}_0$ 初始化为 N 份 $\perp$的副本。对于 $h = 1$、$\cdots$、H，执行：初始化多重集合 $\mathcal{S}_h = \{\}$。对于 $a_{1:h-1} \in \mathcal{S}_{h-1}$，执行：从 $\mathsf{Geom}(\mathbb{P}_{\mathsf{VGB}}(a_{1:h-2} \mid a_{1:h-1}))$ 中采样 $D \sim$（约定为 $a_{1:-1} = \mathsf{s}$；见附录 E.1）。对于 $1 \leq i \leq D$，执行：采样 $a_{1:h} \sim \mathbb{P}_{\mathsf{down}}(\cdot \mid a_{1:h-1})$。更新 $\mathcal{S}_h \leftarrow \mathcal{S}_h \cup \{a_{1:h}\}$。

E.3 VGB与SMC-IND的耦合

我们将 SMC-IND 与 VGB 的执行过程解释为随机构造粒子森林。随后我们证明，当随机性被适当地耦合时，这些森林实际上是相同的（不计 s 的出现次数）。

无限粒子 SMC-IND 过程。 使用 $N = \infty$ 执行 SMC-IND，自然会定义一个有根森林 $\mathcal{F}_{\mathsf{DMC}}$，其中每个节点都由 $\mathcal{S}_0 \sqcup \cdots \sqcup \mathcal{S}_H$ 中的一个粒子标记。根节点是 $\mathcal{S}_0$ 中的粒子。对于任意 $h > 0$，$\mathcal{S}_h$ 中某个粒子的父节点，是 $\mathcal{S}_{h-1}$ 中孕育它的粒子。请注意，任意粒子 $a_{1:h}$ 在 $\mathcal{S}_h$ 中可能出现多次（甚至无限多次），因此 $\mathcal{F}_{\mathsf{DMC}}$ 中可能包含多个标签相同的节点。标记为 $a_{1:h}$ 的节点，其父节点的标签始终为 $a_{1:h-1}$。

无限时间 VGB 过程。 接下来，我们将说明如何把自回归树 $\mathcal{T}$ 中 VGB 的执行解释为一个有根树 $\mathcal{F}_{\mathsf{VGB}}$，与上文类似。固定一条含有 $x_0 =$ 个 s 的 VGB 无限轨迹 $(x_0, x_1, x_2, x_3, \ldots)$。假设 s 在该轨迹中出现无限多次。$\mathcal{F}_{\mathsf{VGB}}$ 的根节点以 $\mathbb{N}$ 为索引，并标记为 s。其他每个节点都以一个区间 $[t, t'] \subset \mathbb{N}$ 为索引，该区间满足：在 $\mathcal{T}$ 中，$x_{t-1} = x_{t'+1}$ 是 $x_t = x_{t'}$ 的父节点，且该父节点未出现在 $\{x_t, x_{t+1}, \ldots, x_{t'}\}$ 中。此节点标记为 x_t，并且它在 $\mathcal{F}_{\mathsf{VGB}}$ 中的父节点是唯一一个标签为 x_{t-1} 且区间包含 $[t, t']$ 的节点。

引理 E.2。 假设 *VGB* 几乎必然会无限多次访问$\mathfrak{s}$。存在一种对 $\mathcal{F}_{\mathsf{DMC}}$ 和 $\mathcal{F}_{\mathsf{VGB}}$ 的随机性的耦合，使得 $\mathcal{F}_{\mathsf{DMC}}$ 几乎必然等于森林 $\mathcal{F}_{\mathsf{VGB}} \setminus \{\mathfrak{s}\}$。

证明。 我们逐层构造耦合。根据引理的假设，显然，$\mathcal{F}_{\mathsf{DMC}}$ 的第一层几乎必然与 $\mathcal{F}_{\mathsf{VGB}}$ 的第一层相同。现在固定 $h \geq 1$，并以 $\mathcal{F}_{\mathsf{VGB}}$ 的前 h 层的森林结构和标签 x_t（但不包括索引 $[t, t']$ 本身）为条件。任取一个标签为 $a_{1:h-1}$ 的节点 ι。

考虑 VGB 随机游走在两次访问 $a_{1:h-1}$ 之间的分布律。游走被限制在以 $a_{1:h-1}$ 为根的子树内；但由于该游走具有马尔可夫性，其分布律不受条件化影响（并且它也独立于第 h 层其他节点 $\iota' \neq \iota$ 的子节点）。令 $p = \mathbb{P}(a_{1:h-2} \mid a_{1:h-1})$（约定当 $h = 1$ 时为 $a_{1:h-2} = \mathsf{s}$）。节点 ι 的子节点数 D'，等于游走在返回 $a_{1:h-2}$ 之前，于 $\mathcal{T}$ 中前往 $a_{1:h-1}$ 的某个子节点的次数。因此，对于任意 $k \in \{0, 1, 2, \ldots\}$，有 $\mathbb{P}[D' = k] = (1-p)^k p$。也就是说，$D'$ 是成功概率为 p 的几何随机变量。此外，ι 的每个子节点的标签相互独立，且服从 $\mathbb{P}_{\mathsf{down}}(\cdot \mid a_{1:h-1})$。因此，可以将耦合扩展到 ι 的子节点。

□

定理定理 E.1的证明。 前 N 棵 $\mathcal{F}_{\mathsf{DMC}}$ 树可以与由 SMC-IND 诱导的森林耦合，其中包含 N 个初始粒子。同样，前 N 棵 $\mathcal{F}_{\mathsf{VGB}}$ 树恰好对应于 VGB 轨迹中访问 $\mathfrak{s}$ 之前的前缀，该前缀中 $N+1$ 次访问了该状态。由此即可得出结论引理 E.2。

□

定理 3.9 的证明 F

在本附录中，我们将正式给出证明定理 3.9所需的定义，并证明该定理。在附录 F.1中，我们将形式化定义短视粒子滤波算法。在附录 F.2 中，我们陈述定理 F.5，它可直接推出定理 3.9。在附录 F.3中，我们给出直觉说明，并正式证明定理 F.5。

F.1 短视粒子滤波算法

定义 F.1。 具有 N 粒子的 短视粒子滤波算法定义如下。该算法计算一个多重集随机序列 $\mathcal{S}_0$，……， $\mathcal{S}_H$，其中对于所有 $0 \le h \le H$， $\mathcal{S}_h \subset \mathcal{A}^h$ 均成立，并满足以下性质：

1. 几乎必然地，对于所有 $1 \le h \le H$，对每个 $y_{1:h} \in \mathcal{S}_h$，都有 $y_{1:h-1} \in \mathcal{S}_h$。
2. 对所有 h， $|\mathcal{S}_h| \le N$ 几乎必然成立。
3. 对所有 h，$\mathcal{S}_0$，……， $\mathcal{S}_h$的分布律由$(\pi_{\mathsf{ref}}(a_{1:h}) \colon a_{1:h} \in \mathcal{A}^h)$ 和$(\widehat{V}(a_{1:k}) \colon a_{1:k} \in \mathcal{A}^k,\ k \le h)$ 确定。

随后，该算法（随机地）输出 $\mathcal{S}_H$ 中的任意一个元素。

第一个条件是对何为粒子滤波方法的自然形式化，即 $\mathcal{S}_h$ 中的粒子不能“凭空出现”；相反，它必须在 $\mathcal{S}_{h-1}$ 中有一个父节点。第二个条件断言，粒子数量始终至多为 N。第三个条件是“ 短视 ”条件，它要求在分布层面上，算法不使用后续步骤 $k > h$ 的信息来确定多重集 $\mathcal{S}_h$。

注记 F.2 （序贯蒙特卡洛及其变体）。 从 算法 *1* 可以直接看出，序贯蒙特卡洛确实是一种短视粒子滤波算法，带拒绝采样的序贯蒙特卡洛（算法 *2*）亦然。

注记 F.3 （计算效率）。 我们指出，定义 *F.1* 并未对计算效率或查询复杂度施加任何上界；特别地，短视粒子滤波算法可以先检查 $\widehat{V}$（$a_{1:h}$），针对所有 $a_{1:h} \in \mathcal{A}^h$，然后再决定粒子集合 $\mathcal{S}_h$；它并不限于例如仅查询 $\mathcal{S}_{h-1}$ 中粒子的子节点。这只会让我们的不可能性结果更强。尽管如此，证明一个类似于 定理 *3.9* 、适用于 所有 进行 $o(H^2)$ 次查询、其对象为 π_{ref} 和 $\widehat{V}$ 的算法的类似结果，是一个有趣的开放问题。这样的结果在形式上不可比于 定理 *3.9*。

F.2 特殊设定与定理的正式表述

由于我们要证明一个下界，因此只需考虑第 2 节一般设定的一个特殊情形。因此，我们将范围限定在自回归生成设置（备注 2.1）中，并进一步令 $\pi_{\mathsf{ref}} :=$ 均匀分布（$\mathcal{A}^H$）其中 $\mathcal{A} = \{0,1\}$。

我们通过指定某个分布 $\mu \in \Delta(\mathcal{A}^H)$ 和函数 $(\mu_h)_h$ （其中 $\mu_h : \mathcal{A}^h \to \mathbb{R}_{\ge 0}$）来定义一个问题实例。这些共同诱导出 $V^\star$， $\widehat{V}$，其定义如下：

$$V^\star(a_{1:h}) := \frac{\mu(a_{1:h})}{\pi_{\mathsf{ref}}(a_{1:h})}$$

and

$$\widehat{V}(a_{1:h}) := \frac{\widehat{\mu}_h(a_{1:h})}{\pi_{\mathsf{ref}}(a_{1:h})}.$$

注意，目标分布 $\pi^\star$ 恰好就是 μ，而定义 F.1 的第三个条件恰好确保 $(\mathcal{S}_0, \ldots, \mathcal{S}_h)$ 的分布律由 $(\widehat{\mu}_0, \ldots, \widehat{\mu}_h)$ 确定。要使假设 3.1 对参数 C_{act} 成立，只需对所有 $a_{1:h+1} \in \mathcal{A}^{h+1}$ 都有 $\mu(a_{h+1} \mid a_{1:h}) \le C_{\mathsf{act}}/2$ 。此外，当且仅当 $\widehat{\mu}$ 在以下意义下对 μ 是 C_∞-准确的，假设 3.3 才对参数 C_∞ 成立：

定义 F.4。 我们称 $\widehat{\mu}$ 对 $\mu \in \Delta(\mathcal{A}^H)$ 是 κ-准确的，当且仅当满足

$$\frac{\widehat{\mu}(a_{1:h})}{\mu(a_{1:h})} \in [1/\kappa, \kappa]$$

对所有 $a_{1:h} \in \mathcal{A}^h$，并且对所有 $a_{1:H} \in \mathcal{A}^H$，还有 $\widehat{\mu}(a_{1:H}) = \mu(a_{1:H})$。

我们可以n ow正式陈述本附录的主要结果，该结果直接蕴含定理 3。 9.

定理 F.5. 定义 $N(H) := \log(H)/(4\log\log(H))$ 对所有 $H \in \mathbb{N}$。不存在具有如下保证的短视粒子滤波算法 Alg：固定充分大的 H。Alg 至多使用 $N(H)$ 个粒子。对任意 $\mu \in \Delta(\mathcal{A}^H)$，对任意 e^3-准确的 $\widehat{\mu}$：$\mathcal{A}^{\leq H} \to \mathbb{R}_{\geq 0}$，Alg 在 $\widehat{\mu}$ 上的执行具有满足下述条件的输出 $o_{1:H}$

$$\mathbb{P}_{o_{1:H} \sim \mathtt{Alg}}[o_{1:H} \in \mathcal{E}] \geq H^{-1/5}$$

对所有满足 $\mu(\mathcal{E}) \geq 1/2$ 的事件 $\mathcal{E} \subset \mathcal{A}^H$。

证明定理 3.9。 由定理 F.5以及前文关于假设 3.1和3.3的讨论直接得出。

□

F.3 证明定理 F.5

F.3.1 直觉与形式构造

为证明 定理 F.5，我们构造一族问题实例（每个问题实例由以下各项的选择确定：μ 和$\widehat{\mu}_0$，.....，$\widehat{\mu}_H$），使得任何至多使用 N个粒子的短视粒子滤波算法，都必须在该族中至少一个问题实例上失败。我们通过递归地构造来实现这一点：首先为 $N = 1$构造一个问题实例，然后构造 $N = 2$，依此类推。分析也将采用类似的归纳方式。

设 $\mathrm{H}\colon \mathbb{N} \to \mathbb{N}$ 是一个待确定的函数。假设对于所有 $N > 1$，$\mathrm{H}(N)$ 都是 $\mathrm{H}(N-1)+2$ 的倍数。对于每个 $N \in \mathbb{N}$ 和 $y^\star \in \mathcal{A}^{\mathfrak{H}(N)}$，我们递归地（关于 N）定义 $\mu^N(\cdot; y^\star) \in \Delta(\mathcal{A}^{\mathfrak{H}(N)})$ 和 $\widehat{\mu}^N(\cdot; y^\star)\colon \mathcal{A}^{\leq \mathfrak{H}(N)} \to \mathbb{R}_{\geq 0}$ 如下。

构造的直觉。首先，我们为 $N = 1$ 构造提供直觉，该构造基本遵循 Rohatgi 等人（2025）中的例 3.2。我们令 $\mu^1(\cdot; y^\star)$ 为乘积分布，其中边缘分布 $\mu^1(y_h)$ 偏向于 $y^\star_h$。接着，我们令 $\widehat{\mu}^1(\cdot; y^\star)$ 相较于 $\mu^1(\cdot; y^\star)$ 有所“延迟”，也就是说，在步骤 h 的边缘分布偏向于 $y^\star_{h-1}$，而非 $y^\star_h$。因此，$\widehat{\mu}^1_1(\cdot; y^\star) \ldots, \widehat{\mu}^1_h(\cdot; y^\star)$ 的数据与 $y^\star_h$ 独立；所以，正如引理 F.9 所形式化的，任何 1 粒子的短视粒子滤波方法都必然产生较大的采样误差。

接下来，我们为 $N > 1$ 构造提供直觉。时域 $H = \mathrm{H}(N)$ 被划分为大小为 $K = \mathrm{H}(N-1)+2$ 的分块。同样，$\mu^N(\cdot; y^\star)$ 是乘积分布，其在步骤 h 的边缘分布偏向于 $y^\star_h$。在分块 $[iK+1, (i+1)K]$ 中，我们通过将对 $y^\star_{iK+1}$ 的依赖“延迟”到步骤 $(i+1)K$，并在中间步骤嵌入一个 $(N-1)$ 粒子构造的副本，来定义 $\widehat{\mu}^N(\cdot; y^\star)$。直观地说，任何 N 粒子的短视粒子滤波算法在步骤 $iK+1$ 都有两种选择：要么不可逆地猜测 $y^\star_{iK+1}$（即所有粒子 $y \in \mathcal{S}_{iK+1}$ 都取相同的 y_{iK+1} 值），要么将粒子分开，使其中一些粒子的 y_{iK+1} 值各不相同。引理 F.11 对这一基本思路进行了形式化：两种选择都会导致采样误差；前者是因为在步骤 $iK+1$ 无法获得关于 $y^\star_{iK+1}$ 的数据，后者则是由 $(N-1)$ 粒子构造的归纳假设所致。需要多个分块，是为了放大由第一种选择引起的误差。

要使这一论证严谨成立，需要进行仔细的概率分析，以确保两种情况下的误差不会“相互抵消”。在继续分析之前，我们先将这一构造形式化。

设 $\gamma_1, \gamma_2, \cdots \in (0, 1/2)$ 是一组待确定的参数序列。

形式构造：基例（$N = 1$）。 写出 $H = \mathfrak{H}(1)$。我们定义

$$\mu^1(y_h \mid y_{1:h-1}; y^\star) := (1/2 - \gamma_1)^{\mathbb{I}[y_h \neq y^\star_h]} (1/2 + \gamma_1)^{\mathbb{I}[y_h = y^\star_h]}$$

and

$$\widehat{\mu}^1(y_h \mid y_{1:h-1}; y^\star) := \begin{cases} 1/2 & \text{if } h = 1 \\ \mu_1(y_{h-1} \mid y_{1:h-2}; y^\star) & \text{if } 1 < h < H \\ 2\mu_1(y_{H-1:H} \mid y_{1:H-2}; y^\star) & \text{if } h = \mathfrak{H}(1) \end{cases}.$$

形式构造：递归情形（$N>1$）。定义 $K:=\mathrm{H}(N-1)+2$。对于任意非负整数i 以及满足 $iK<h\le(i+1)K$ 的 $h\in[\mathrm{H}(N)]$，我们定义

$$\mu^N(y_h\mid y_{1:h-1};y^\star):=\begin{cases}(1/2-\gamma_N)^{\mathbb{I}[y_h\neq y_h^\star]}(1/2+\gamma_N)^{\mathbb{I}[y_h=y_h^\star]} & \text{if } h=iK+1 \text{ or } h=(i+1)K\\ \mu^{N-1}(y_h\mid y_{iK+2:h-1};y^\star_{iK+2:(i+1)K-1}) & \text{otherwise}\end{cases}$$

and

$$\widehat{\mu}^N(y_h\mid y_{1:h-1};y^\star):=\begin{cases}1/2 & \text{if } h=iK+1\\ \widehat{\mu}^{N-1}(y_h\mid y_{iK+2:h-1};y^\star_{iK+2:(i+1)K-1}) & \text{if } iK+1<h<(i+1)K\\ 2\mu^N(y_{iK+1}\mid y_{1:iK};y^\star)\mu^N(y_{(i+1)K}\mid y_{1:(i+1)K-1};y^\star) & \text{if } h=(i+1)K\end{cases}$$

在以下引理中，我们检验所需的准确性条件是否对所有 N 和 $y^\star$都成立。

引理 F.6. 对于任意 $N\ge1$和 $y^\star\in\mathcal{A}^{\mathfrak{h}(N)}$，有 $\widehat{\mu}^N(\cdot;\,y^\star)$ 是 $\exp(3\sum_{i=1}^N\gamma_i)$-准确的，对于 $\mu^N(\cdot;\,y^\star)$。**证明。**我们对 N 使用归纳法。固定 $N=1$。令 $H=\mathrm{H}(1)$。注意到 $\widehat{\mu}^1(y_{1:H};\,y^\star)=\mu^1(y_{1:H};\,y^\star)$，并且对于任意 $h<H$，

$$\begin{aligned}\frac{\widehat{\mu}^1(y_{1:h};y^\star)}{\mu_1(y_{1:h};y^\star)}&=\frac{1}{2\mu^1(y_h\mid y_{1:h-1};y^\star)}\\&\in[1/(1+2\gamma_1),1/(1-2\gamma_1)]\\&\in[e^{-3\gamma_1},e^{3\gamma_1}].\end{aligned}$$

固定 $N>1$，并假设该结论对 $\widehat{\mu}^{N-1}$成立。令 $H=\mathrm{H}(N)$，且 $K=\mathrm{H}(N-1)+2$。对于任意$h\in[H]$，若存在某个 i使得 $h=(i+1)K$，则

$$\begin{aligned}&\widehat{\mu}^N(y_{iK+1:(i+1)K}\mid y_{1:iK};y^\star)\\&=\frac12\cdot\left(\prod_{h'=iK+2}^{(i+1)K-1}\widehat{\mu}^{N-1}(y_{h'}\mid y_{iK+2:h'-1};y^\star_{iK+2:(i+1)K-1})\right)\\&\quad\cdot2\mu^N(y_{iK+1}\mid y_{1:iK};y^\star)\mu^N(y_{(i+1)K}\mid y_{1:(i+1)K-1};y^\star)\\&=\mu^{N-1}(y_{iK+2:(i+1)K-1};y^\star_{iK+2:(i+1)K-1})\cdot\mu^N(y_{iK+1}\mid y_{1:iK};y^\star)\mu^N(y_{(i+1)K}\mid y_{1:(i+1)K-1};y^\star)\\&=\mu^N(y_{iK+2:(i+1)K-1}\mid y_{1:iK+1};y^\star)\cdot\mu^N(y_{iK+1}\mid y_{1:iK};y^\star)\mu^N(y_{(i+1)K}\mid y_{1:(i+1)K-1};y^\star)\\&=\mu^N(y_{iK+1:(i+1)K}\mid y_{1:iK};y^\star).\end{aligned}$$

因此，有 $\widehat{\mu}^N(y_{1:(i+1)K};\,y^\star)=\mu^N(y_{1:(i+1)K};\,y^\star)$，特别地，$\widehat{\mu}^N(y_{1:H};\,y^\star)=\mu^N(y_{1:H};\,y^\star)$。对任意 $iK+1\le h<(i+1)K$，我们有

$$\begin{aligned}\widehat{\mu}^N(y_{1:h};y^\star)&=\mu^N(y_{1:iK};y^\star)\widehat{\mu}^N(y_{iK+1:h}\mid y_{1:iK};y^\star)\\&=\mu^N(y_{1:iK};y^\star)\cdot\frac12\cdot\widehat{\mu}^{N-1}(y_{iK+2:h};y^\star_{iK+2:(i+1)K-1})\\&\in[e^{-3\sum_{n=1}^{N-1}\gamma_n},e^{3\sum_{n=1}^{N-1}\gamma_n}]\cdot\mu^N(y_{1:iK};y^\star)\cdot\frac12\cdot\mu^{N-1}(y_{iK+2:h};y^\star_{iK+2:(i+1)K-1})\\&\in[e^{-3\sum_{n=1}^{N}\gamma_n},e^{3\sum_{n=1}^{N}\gamma_n}]\cdot\mu^N(y_{1:iK};y^\star)\cdot\mu^N(y_{iK+1}\mid y_{1:iK};y^\star)\cdot\mu^{N-1}(y_{iK+2:h};y^\star_{iK+2:(i+1)K-1})\\&\in[e^{-3\sum_{n=1}^{N}\gamma_n},e^{3\sum_{n=1}^{N}\gamma_n}]\cdot\mu^N(y_{1:h};y^\star)\end{aligned}$$

证毕。 □

为了证明，任何 N 粒子的短视粒子滤波方法都无法覆盖 $\mu^N(\cdot;\,y^\star)$(对于某个 $y^\star$ 的取值），我们将证明存在一些事件 $\mathcal{F}_N(y^\star)$，它们在 $\mu^N(\cdot;\,y^\star)$(对于任意 $y^\star$）下发生的概率极高，但在任何 N 粒子的短视粒子滤波方法的执行过程中（对于均匀随机选取的 $y^\star$）发生的概率极低。定义 F.7 对这些事件作了形式化定义。定义 F.8则将我们将通过归纳法证明的关于这些事件的论断形式化。

定义 F.7（关键事件）。固定 $N = 1$ 和 $H = \mathsf{H}(1)$。我们定义

$$\mathcal{F}_1(y^\star) := \{o_{1:H} : \#\{1 < h < H : o_h = y_h^\star\} \geq (1/2 + \gamma_1/2)(H-2)\}.$$

固定 $N > 1$，并令 $H := \mathfrak{h}(N)$ 和 $K := \mathfrak{h}(N-1) + 2$。我们定义集合 $\mathcal{F}_N(y^\star)$ 如下：

$$\mathcal{F}_N^\circ(y^\star) := \bigcap_i \left\{ o_{1:H} : o_{iK+2:(i+1)K-1} \in \mathcal{F}_{N-1}(y^\star_{iK+2:(i+1)K-1}) \right\}$$

$$\mathcal{F}_N^\square(y^\star) := \{o_{1:H} : \#\{i : o_{iK+1} = y^\star_{iK+1}\} \geq 1/2 + \gamma_N/2\}$$

$$\mathcal{F}_N(y^\star) := \mathcal{F}_N^\circ(y^\star) \cap \mathcal{F}_N^\square(y^\star).$$

定义 F.8 (分析的归纳假设**)**。固定 $N \geq 1$。对于参数 $\delta_{N,1}, \delta_{N,2} \in (0,1)$，我们用 $\mathsf{IH}_N(\delta_{N,1}, \delta_{N,2})$ 表示以下假设。存在满足以下性质的集合 $\mathcal{F}_N(y^\star) \subseteq \mathcal{A}^{\mathfrak{h}(N)}$：

- 对于所有 $y^\star \in \mathcal{A}_{\mathfrak{h}(N)}$，均有
$$\mu^N(\mathcal{F}_N(y^\star); y^\star) \geq 1 - \delta_{N,1}$$
- 对于任意 N-粒子粒子滤波算法 $\mathtt{Alg}$，均有
$$\mathop{\mathbb{E}}_{y^\star \sim \mathsf{Unif}(\mathcal{A}^{\mathfrak{h}(N)})} \mathbb{P}_{o \sim \mathtt{Alg}}[o \in \mathcal{F}_N(y^\star)] \leq \delta_{N,2}$$
其中，概率是针对 $\mathtt{Alg}$ 在实例 $\widehat{\mu}^N(\cdot; y^\star)$ 上运行所得的输出 o 而言的。

F.3.2 分析的基例

引理 F.9。对于任意 $y^\star \in \mathcal{A}^{\mathfrak{h}(1)}$，有

$$\mu^1(\mathcal{F}_1(y^\star); y^\star) \geq 1 - \exp\left(-\gamma_1^2 \frac{\mathfrak{h}(1) - 2}{16}\right).$$

此外，对于任意 1 粒子粒子滤波算法 $\mathtt{Alg}$，

$$\mathop{\mathbb{E}}_{y^\star \sim \mathsf{Unif}(\mathcal{A}^{\mathfrak{h}(1)})} \mathbb{P}_{o \sim \mathtt{Alg}}[o \in \mathcal{F}_1(y^\star)] \leq \exp\left(-\gamma_1^2 \frac{\mathfrak{h}(1) - 2}{12}\right)$$

其中 $\mathtt{Alg}$ 在实例 $\widehat{\mu}^1(\cdot;\ y^\star)$ 上运行。

证明。令 $X \sim \mu^1(\cdot; y^\star)$。那么，对于每个 $h \in [H]$，$X_1, \ldots, X_H$ 相互独立，且 $\mathbb{P}[X_h = y_h^\star] = 1/2 + \gamma_1$。由切尔诺夫界可得

$$\begin{aligned}\mu^1(\mathcal{F}_1(y^\star); y^\star) &= \mathbb{P}[\#\{1 < h < H : X_h = y_h^\star\} > (1/2 + \gamma_1/2)H] \\ &\geq 1 - \exp(-\gamma_1^2(\mathfrak{h}(1) - 2)/16)\end{aligned}$$

这证明了引理的第一个结论。接下来，固定任意一个 1 粒子粒子滤波算法 Alg。采样 $y^\star \sim \mathsf{Unif}(\mathcal{A}^{\mathfrak{h}(1)})$，并令 $\mathcal{S}_0, \ldots, \mathcal{S}_H$ 为在 $\widehat{\mu}^1(\cdot; y^\star)$ 上运行 Alg 所得到的多重集随机序列。令 o 为 Alg 的输出。那么，对于所有 $0 \leq h \leq H$，$\mathcal{S}_h = \{o_{1:h}\}$。由于 $\mathcal{S}_{0:h}$ 的分布律由 $(\widehat{\mu}_0^1(\cdot; y^\star) \ldots, \widehat{\mu}_h^1(\cdot; y^\star))$ 确定，根据 $\widehat{\mu}^1(\cdot; y^\star)$ 的定义，$(\mathcal{S}_{0:h}, y^\star_{1:h-1})$ 对每个 $1 < h < H$ 都与 $y_h^\star$ 独立。因此，在给定 $(\mathcal{S}_{0:h-1}, y^\star_{1:h-1})$ 的条件下，$y_h^\star$ 与 $\mathcal{S}_h$ 独立，于是有

$$\mathbb{P}[o_h = y_h^\star \mid \mathcal{S}_{0:h-1}, y^\star_{1:h-1}] = \frac{1}{2}$$

f或所有 $1 < h < H$。引理的第二个断言随后由 Azuma 不等式得出 ty. □

F.3.3 分析的归纳步骤

引理 F.10. 固定 $N > 1$。假设 $\mathsf{IH}_{N-1}(\delta_{N-1,1},\delta_{N-1,2})$ 成立。设 $\mathfrak{H}(N-1) \geq 2$。对于任意 $y^\star \in \mathcal{A}^{\mathfrak{H}(N)}$，有

$$\mu^N(\mathcal{F}_N(y^\star); y^\star) \geq 1 - \frac{\mathfrak{H}(N)}{\mathfrak{H}(N-1)}\delta_{N-1,1} - \exp\left(-\gamma_N^2 \frac{\mathfrak{H}(N)}{16 \cdot \mathfrak{H}(N-1)}\right)$$

证明。 令 $H := \mathsf{H}(N)$，且令 $K := \mathsf{H}(N-1) + 2$。令 $o \sim \mu^N(\cdot; y^\star)$。那么，对每个 i，$o_{iK+2:(i+1)K-1}$ 的分布律恰好是 $\mu^{N-1}(\cdot; y^\star_{iK+2:(i+1)K-1})$。因此，

$$\begin{aligned}
&\mu^N(\mathcal{F}_N^\circ(y^\star); y^\star) \\
&\geq 1 - \sum_i \mu^N(o_{iK+2:(i+1)K-1} \notin \mathcal{F}_{N-1}(y^\star_{iK+2:(i+1)K-1})) \\
&= 1 - \frac{H}{K}\left(1 - \mu^{N-1}(\mathcal{F}_{N-1}(y^\star_{iK+2:(i+1)K-1}); y^\star_{iK+2:(i+1)K-1})\right) \\
&\geq 1 - \frac{H}{K}\delta_{N-1,1} \\
&\geq 1 - \frac{\mathfrak{H}(N)}{\mathfrak{H}(N-1)}\delta_{N-1,1}.
\end{aligned}$$

接下来，注意随机变量 $(o_{iK+1})_i$ 相互独立，并且对每个 i 都有 $\mathbb{P}[o_{iK+1} = y^\star_{iK+1}] = 1/2 + \gamma_N$。因此，

$$\mu^N(\mathcal{F}_N^\square(y^\star); y^\star) \geq 1 - \exp(-\gamma_N^2(H/K)/8) \geq 1 - \exp(-\gamma_N^2(\mathfrak{H}(N)/\mathfrak{H}(N-1))/16$$

由 Hoeffding不等式可知，只要 $\mathsf{H}(N-1) \geq 2$ 即成立。该引理由并集界得证。 □

引理 F.11。 固定 $N > 1$ 和 $\delta_{N-1,1}, \delta_{N-1,2}, \delta_{N,2} \in (0, 1/2)$。假设$\mathsf{IH}_{N-1}(\delta_{N-1,1}, \delta_{N-1,2})$ 成立。令$H := \mathfrak{H}(N)$，且令 $K := \mathfrak{H}(N-1) + 2$。如果

$$\gamma_N > 4\delta_{N-1,2} + \sqrt{\frac{8K\log(2/\delta_{N,2})}{H}} + \frac{8K}{H}\log(4/\delta_{N,2}),$$

那么，对于任意的 N 粒子粒子滤波算法 $\mathtt{Alg}$，有

$$\mathop{\mathbb{E}}_{y^\star \sim \mathsf{Unif}(\mathcal{A}^{\mathfrak{H}(N)})} \mathbb{P}_{o \sim \mathtt{Alg}}[o \in \mathcal{F}_N(y^\star)] \leq \delta_{N,2}$$

其中，$\mathtt{Alg}$ 在实例 $\widehat{\mu}^N(\cdot; y^\star)$上运行。

证明。考虑通过采样 $y^\star \sim \mathsf{Unif}(\mathcal{A}^{\mathfrak{H}(N)})$ 并在 $\widehat{\mu}^N(\cdot; y^\star)$ 上运行 Alg 所诱导的随机过程。固定 i 并以 $\mathcal{S}_{0:iK}$ 为条件。令 $\mathcal{E}^i$ 为如下事件：对于某个 y，$y' \in \mathcal{S}_{(i+1)K-1}$，有 $y_{iK+1} \neq y'_{iK+1}$。对于每个$b \in \{0,1\}$，定义过程 $\mathcal{T}_{(0,\ b)}$ . . . ，$\mathcal{T}_{(K-2,\ b)}$，其中 $\mathcal{T}_{(0,\ b)} = \{\varnothing\}$，并且对于每个 $h' > 0$，$\mathcal{T}_{(h',\ b)}$ 由 ..... 的前 $\leq N-1$ 个元素构成

$$\{y_{iK+2:h'+iK+1} : y_{1:h'+iK+1} \in \mathcal{S}_{h'+iK+1} \wedge y_{iK+1} = b \wedge y_{iK+2:h'+iK} \in \mathcal{T}^{(h'-1,b)}\}$$

按照 $\mathcal{A}^{h'}$ 的某种任意排序。注意，$\mathcal{T}_{(0,b)}, \ldots, \mathcal{T}_{(K-2,b)}$ 描述了在 $\widehat{\mu}^{N-1}(\cdot; y^\star_{iK+2:(i+1)K-1})$ 上运行一个 $(N-1)$ 粒子的短视粒子滤波算法。为说明这一点，我们检查定义 F.1 的条件。首先，几乎必然有 $|\mathcal{T}^{(h',b)}| \leq N-1$ 对所有 h' 均成立。其次，根据构造，若 $y_{1:h'} \in \mathcal{T}_{(h',b)}$，则 $y_{1:h'-1} \in \mathcal{T}_{(h'-1,b)}$。第三，对于任意 $0 \leq h \leq K-2$，$\mathcal{T}_{(0,b)}, \ldots, \mathcal{T}_{(h',b)}$ 的分布律由 $\mathcal{S}_{0:h'+iK+1}$ 的分布律决定；根据 Alg 的定义，后者由 $(\widehat{\mu}_h^N(\cdot; y^\star) : h \leq h'+iK+1)\}$ 决定；而根据 $\widehat{\mu}^N$ 的定义，后者由 $(\widehat{\mu}_k^{N-1}(\cdot; y^\star_{iK+2:(i+1)K-1}) : 0 \leq k \leq h')$ 决定（以及与 $y^\star_{iK+2:(i+1)K-1}$ 独立的随机性）。

由归纳假设可知，

$$\mathbb{P}[\exists y \in \mathcal{T}^{(K-2,b)} \cap \mathcal{F}_{N-1}(y^\star_{iK+2:(i+1)K-1})] \leq \delta_{N-1,2}.$$

在事件 $\mathcal{E}^i$ 中，几乎必然有

$$\{y_{iK+2:(i+1)K-1} : y_{1:(i+1)K-1} \in \mathcal{S}_{(i+1)K-1}\} = \mathcal{T}^{(K-2,0)} \cup \mathcal{T}^{(K-2,1)}.$$

此外，Alg 的输出 $o_{1:H}$ 几乎必然满足 $o_{1:(i+1)K-1} \in \mathcal{S}_{(i+1)K-1}$，因此

$$\begin{aligned}
&\mathbb{P}[o_{iK+2:(i+1)K-1} \in \mathcal{F}_{N-1}(y^\star_{iK+2:(i+1)K-1}) \mid \mathcal{S}_{0:iK}] \\
&\le \mathbb{P}[\exists y_{1:(i+1)K-1} \in \mathcal{S}_{(i+1)K-1} : y_{iK+2:(i+1)K-1} \in \mathcal{F}_{N-1}(y^\star_{iK+2:(i+1)K-1}) \mid \mathcal{S}_{0:iK}] \\
&\le \mathbb{P}[\overline{\mathcal{E}}^i \mid \mathcal{S}_{0:iK}] + \mathbb{P}[\exists y \in \mathcal{T}^{(K-2,0)} \cup \mathcal{T}^{(K-2,1)} : y_{iK+2:(i+1)K-1} \in \mathcal{F}_{N-1}(y^\star_{iK+2:(i+1)K-1}) \mid \mathcal{S}_{0:iK}] \\
&\le \mathbb{P}[\overline{\mathcal{E}}^i \mid \mathcal{S}_{0:iK}] + 2\delta_{N-1,2}.
\end{aligned}$$

接下来，注意到在事件 $\mathcal{E}^i$中，成立 $o_{iK+1} =$ 众数$\{y_{iK+1} : \ y \in \mathcal{S}_{(i+1)K-1}\}$。注意到$\mathcal{S}_{(i+1)K-1}$ 与 $y^\star_{iK+1}$ 独立，因此

$$\mathbb{P}[\mathrm{mode}\{y_{iK+1} : y \in \mathcal{S}_{(i+1)K-1}\} = y^\star_{iK+1} \mid \mathcal{S}_{0:iK}] = \frac{1}{2}.$$

因此，我们有

$$\mathbb{P}[o_{iK+1} = y^\star_{iK+1} \mid \mathcal{S}_{0:iK}] \le \frac{1}{2} + \mathbb{P}[\mathcal{E}^i \mid \mathcal{S}_{0:iK}].$$

由此可得

$$\begin{aligned}
&\mathbb{P}[o_{iK+1} = y^\star_{iK+1} \mid \mathcal{S}_{0:iK}] \\
&\le \frac{1}{2} + \mathbb{P}[o_{iK+2:(i+1)K-1} \notin \mathcal{F}_{N-1}(y^\star_{iK+2:(i+1)K-1}) \mid \mathcal{S}_{0:iK}] + 2\delta_{N-1,2}.
\end{aligned}$$

定义随机变量

$$W_i := \mathbb{I}[o_{iK+1} = y^\star_{iK+1}] - \mathbb{P}[o_{iK+1} = y^\star_{iK+1} \mid \mathcal{S}_{0:iK}].$$

则 W_0，……，$W_{H/K}$是由 $(\mathcal{S}_{0:iK})_{i=0}^{H/K-1}$ 诱导的滤过下的鞅差序列，因此对任意$\delta \in (0,\ 1)$,

$$\mathbb{P}\left[\sum_{i=0}^{H/K-1} W_i > \sqrt{8(H/K)\log(1/\delta)}\right] \le \delta.$$

因此，以至少 $1-\delta$ 的概率，有

$$\begin{aligned}
&\frac{K}{H}\sum_{i=0}^{H/K-1} \mathbb{I}[o_{iK+1} = y^\star_{iK+1}] \\
&\le \frac{1}{2} + 2\delta_{N-1,2} + \sqrt{\frac{8K\log(1/\delta)}{H}} \\
&\qquad + \frac{K}{H}\sum_{i=0}^{H/K-1} \mathbb{P}[o_{iK+2:(i+1)K-1} \notin \mathcal{F}_{N-1}(y^\star_{iK+2:(i+1)K-1}) \mid \mathcal{S}_{0:iK}].
\end{aligned}$$

根据Freedman不等式，以至少为 $1-\delta$ 的概率，有

$$\begin{aligned}
&\sum_{i=0}^{H/K-1} \mathbb{P}[o_{iK+2:(i+1)K-1} \notin \mathcal{F}_{N-1}(y^\star_{iK+2:(i+1)K-1}) \mid \mathcal{S}_{0:iK}] \\
&\le 2\sum_{i=0}^{H/K-1} \mathbb{I}[o_{iK+2:(i+1)K-1} \notin \mathcal{F}_{N-1}(y^\star_{iK+2:(i+1)K-1})] + 8\log(2/\delta).
\end{aligned}$$

我们得出，以至少为 $1-2\delta$的概率，

$$\begin{aligned}
&\frac{K}{H}\sum_{i=0}^{H/K-1}\mathbb{I}[o_{iK+1}=y^\star_{iK+1}]\\
&\le\frac{1}{2}+2\delta_{N-1,2}+\sqrt{\frac{8K\log(1/\delta)}{H}}+\frac{8K}{H}\log(2/\delta)\\
&\quad+\frac{2K}{H}\sum_{i=0}^{H/K-1}\mathbb{I}[o_{iK+2:(i+1)K-1}\notin\mathcal{F}_{N-1}(y^\star_{iK+2:(i+1)K-1})].
\end{aligned}$$

以该事件为条件。若进一步有 $o_{1:H}\in\mathcal{F}_N(y^\star)$，则我们同时有

$$\frac{K}{H}\sum_{i=0}^{H/K-1}\mathbb{I}[o_{iK+1}=y^\star_{iK+1}]\ge\frac{1}{2}+\frac{\gamma_N}{2}$$

and

$$\frac{K}{H}\sum_{i=0}^{H/K-1}\mathbb{I}[o_{iK+2:(i+1)K-1}\notin\mathcal{F}_{N-1}(y^\star_{iK+2:(i+1)K-1})]=0.$$

但这样一来

$$\frac{\gamma_N}{2}\le 2\delta_{N-1,2}+\sqrt{\frac{8K\log(1/\delta)}{H}}+\frac{8K}{H}\log(2/\delta).$$

令 $\delta:=\delta_{N,2}/2$。那么这个界与引理的假设矛盾，因此事实上 $o_{1:H}\notin\mathcal{F}_N(y^\star)$ 几乎必然成立（在前述条件下）。由此可得 $\mathbb{P}[o_{1:H}\in\mathcal{F}_N(y^\star)]\le 2\delta=\delta_{N,2}$，如所述。 □

F.3.4 完成归纳

固定某个 N。对所有 n，定义 $\mathrm{H}(n):=N^{4n}$。对所有 n，定义 $\gamma_n:=1/N$。

引理 F.12。有

$$\mu^N(\mathcal{F}_N(y^\star);y^\star)\ge 1-\exp(-\Omega(N^2))$$

对于所有 $y^\star$，以及

$$\mathop{\mathbb{E}}_{y^\star\sim\mathsf{Unif}(\mathcal{A}^{\mathfrak{H}(N)})}\mathbb{P}_{o\sim\mathtt{Alg}}[o\in\mathcal{F}_N(y^\star)]\le 4\exp(-N).$$

证明。 由引理 F.9 和F.10 可知

$$\mu^N(\mathcal{F}_N(y^\star);y^\star)\ge 1-\sum_{n=1}^{N}\frac{\mathfrak{H}(N)}{\mathfrak{H}(n-1)}\exp(-\gamma_n^2\cdot\Omega(\mathfrak{H}(n)/\mathfrak{H}(n-1)))\ge 1-N^{4N+1}\exp(-\Omega(N^2)).$$

接下来注意，如果我们设定 $\delta_{n,\ 2}:=4\exp(-N)$ 那么，对于每个 引理 F.11，其条件均得到满足，只要 n超过一个充分大的常数 N。因此，将引理 F.11 与关于 N的归纳法结合起来（其中引理 F.9 为 $N=1$提供保证）可证明

$$\mathop{\mathbb{E}}_{y^\star\sim\mathsf{Unif}(\mathcal{A}^{\mathfrak{H}(n)})}\mathbb{P}_{o\sim\mathtt{Alg}}[o\in\mathcal{F}_n(y^\star)]\le 4\exp(-N)$$

对于每个 $n\le N$，按需进行。 □

定理 F.5的证明。 这直接由引理 F.6 和F.12得出，再结合以下事实： N-粒子构造的时域为 $H=\mathfrak{H}(N)=N^{4N}$，因此 $N\ge\log(H)/(4\log\log(H))$。 □

G 其他实验细节

在本节中，我们将进一步详细介绍实验设置。

G.1 其他评估方法

首先，我们会进一步说明背景，并阐述选择 第 4.1 节中提出的 logprob-差异评估指标的动机。

一种评估序贯蒙特卡洛的方法已在以往研究中使用：利用以下事实，即 W_H 在 算法 1 中计算出的值 $\mathbb{E}[W_H] = Z$ 满足，并且差值 $Z - \mathbb{E}[\log W_H]$ 是 KL 散度 $D_{\mathsf{KL}}(\mathbb{E}[\widehat{\nu}_H] \,\|\, \pi_H^\star)$ 的上界，其中该散度衡量序贯蒙特卡洛的输出分布 $\mathbb{E}[\widehat{\nu}_H]$ 与目标分布 $\pi_H^\star$ 之间的差异（参见 Zhao 等人 (2024年) 的命题 5）。[9] 然而，这种方法存在若干问题：首先，这只是一个 上界，无法保证其足够紧；此外，它只适用于序贯蒙特卡洛，因此无法将其性能与其他基线方法进行比较。最后，即使在我们所采用的相对简单的设定中，粒子数量也 N 过少，无法获得实际的 分布 接近程度，来衡量 $\mathbb{E}[\widehat{\nu}_H]$ 与目标分布 $\pi_H^\star$ 之间的差异（例如，动作级覆盖度过高）。

因此，我们旨在采用一种较弱的分布接近度概念，使我们能够 (a) 高效地进行测量，并且 (b) 有望使其与序贯蒙特卡洛所生成样本的质量相关。$\mathcal{P}$-logprob 差异显然满足条件 (a)。此外，由于其误差具有单边性（即 $\mathrm{LPD}_{\mathcal{P}}(\nu,\ \nu) = 0$），且在我们的所有实验中都取正值，我们认为它是提示切换任务中的一个良好候选指标。

第 4.2 节的详细信息

对于第 4.2 节中的提示切换任务，提示 $\mathsf{p}_{\mathsf{ref}}$、$\mathsf{p}^{(i)}$ 均由 GPT-5.2 生成（具体而言，使用的是 gpt-5.2_2025-12-11 版本）。

为估计 $D_{\mathsf{KL}}\left(\pi_h^\star \,\|\, \widehat{\pi}_h^{(i)}\right)$，我们写作

$$
\begin{aligned}
D_{\mathsf{KL}}\left(\pi_h^\star \,\|\, \widehat{\pi}_h^{(i)}\right) &= \mathbb{E}_{a_{1:h}\sim\pi_h^\star}\left[\log \frac{\pi_h^\star(a_{1:h})}{\widehat{\pi}_h^{(i)}(a_{1:h})}\right] \\
&= \mathbb{E}_{a_{1:h}\sim\pi_h^\star}\left[(1-h/H)\cdot\alpha\cdot\log\frac{\pi^\star(a_{1:h})}{\pi^{(i)}(a_{1:h})}\right] + \log \widehat{Z}_h^{(i)}, \qquad (9)
\end{aligned}
$$

其中

$$
\widehat{Z}_h^{(i)} := \mathbb{E}_{\pi_{\mathsf{ref}}}[\widehat{V}^{(i)}(a_{1:h})] = \mathbb{E}_{\pi^\star}\left[\left(\frac{\pi_h^{(i)}(a_{1:h})}{\pi_h^\star(a_{1:h})}\right)^{(1-h/H)\cdot\alpha}\right]. \qquad (10)
$$

我们通过从 $\pi_h^\star$进行蒙特卡洛轨迹采样，来估计 式 $\widehat{Z}_h^{(i)}$ 中的第一项；并根据 $\pi^\star$进行蒙特卡洛轨迹采样，来估计 $\log\ \widehat{Z}_h^{(i)}$，依据式 (10)（每项均使用 1000 个样本）。

请注意，我们对 $\widehat{Z}_h^{(i)}$ 的估计可能非常嘈杂，因为期望值内部没有对数。当我们在 Eq. (9) 中完全不添加（我们对）项 $\log\ \widehat{Z}_h^{(i)}$ 时重复实验，我们观察到了类似的结果。因此，我们对 $\widehat{Z}_h^{(i)}$ 的估计中的噪声似乎不太可能影响我们的结论。

图表详情。 在 图 2b 中，我们使用 公式 (2) 中定义的 PRM，其中 $\alpha = 2$。此外，KL 散度是在第 $h = 32$ 步测量的（回想一下，$H = 64$）；对于 h 的其他取值，我们也观察到了类似的结果。实验在 32GB NVIDIA V100 GPU 上运行。绘制在 图 2b 中各点的皮尔逊相关系数为 0.81。

在 图 7 中，我们绘制了与 图 2b 中相同实验的其他运行结果，使用了数据集的另外 3 种不同选择（即，对 $\mathsf{p}_{\mathsf{ref}}$ 和 $\mathsf{p}^{(i)}$ 的不同选择）。最后，在 图 8 中，我们对全部 4 个数据集重复了相同的实验，但使用 $\alpha = 1$。

[9] 此外，在引理 C.3 中，我们证明差值 $|\widehat{W}_H - Z|$ 是全变差距离 $D_{\mathsf{TV}}(\mathbb{E}[\widehat{\nu}_H]$的上界 $\pi_H^\star)$。

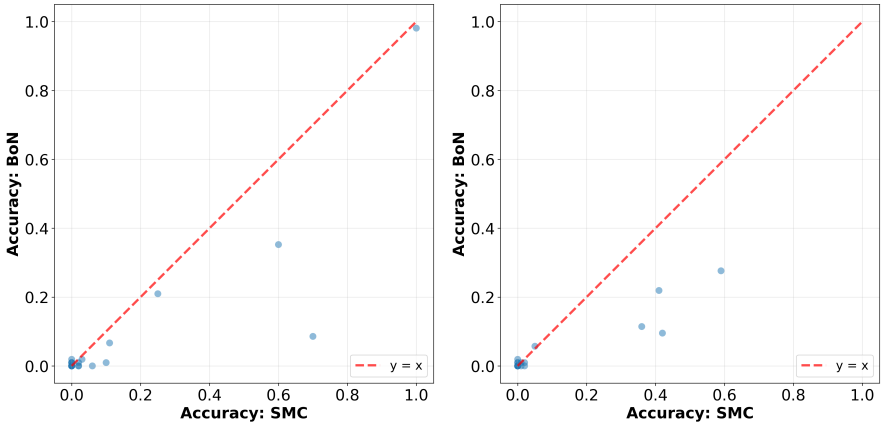
G.3 第4.3节的详细信息

在图 6中，我们针对图 2a 重复了实验，并为数据集额外选择了 4 种设置（即，对$\mathsf{p}_{\mathsf{ref}}$ 和$\mathsf{p}^{(i)}$的选择 $i \in [k]$）。皮尔逊相关系数 r 显示在每个子图的子标题中。实验在配备 32GB 显存的 NVIDIA V100 GPU 上运行。

G.4 序贯蒙特卡洛与 Best-of-N 在 AIME 上的比较

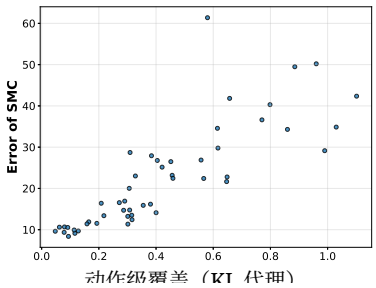
在第4.5节关于数学问题的实验中，我们使用 Qwen2.5-1B-Instruct 作为基础模型 π_{ref}，并使用^Qwen2.5-Math-PRM-7B 作为过程奖励模型 V。对于序贯蒙特卡洛和 Best-of-N，我们使用了 $N = 32$ 个粒子/样本。我们以温度0.8从基础模型采样，并在序贯蒙特卡洛的实现中使用系统重采样。AIME 最多生成 3072 个词元，Math500 最多生成 2048 个词元。

与将词元空间作为 $\mathcal{A}$ （如我们在提示切换实验中的做法）来实现序贯蒙特卡洛不同，我们将 $\mathcal{A}$ 设为由若干个 B 词元组成的分块空间；这样做显著提高了序贯蒙特卡洛的效率，因为调用过程奖励模型是主要的计算瓶颈。我们为^Math500 设置 $B = 64$，为 AIME 设置 $B = 128$。此外，我们观察到，奖励模型 V（$a_{1:h}$）在接收到作为输入的部分补全 $a_{1:h}$ （其在句子或数学表达式中间被截断）时，往往会给出不可靠的答案。因此，实际上，我们会在分隔符（例如换行符、句点）最后一次出现的位置截断所有分块。

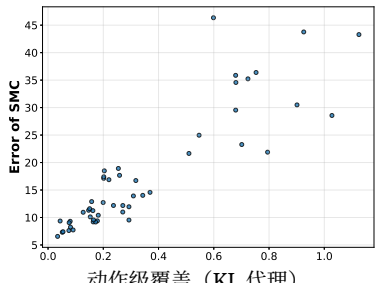


（a）AIME 2024年　　（b）AIME 2025年

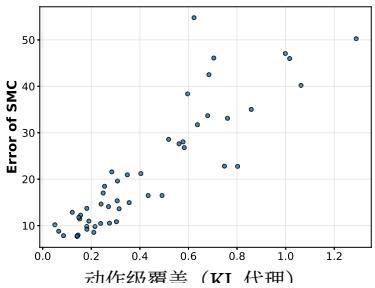
图 5：序贯蒙特卡洛与 Best-of-N 在 AIME 问题上的性能（每个点代表一个不同的问题）。与 图 1类似，大多数点位于直线 $y = x$下方，表明序贯蒙特卡洛在大多数问题上的性能优于 Best-of-N。



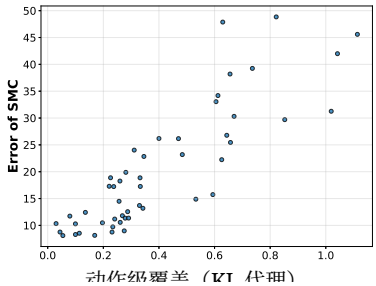
(a) $r = 0.83$



(b) $r = 0.89$

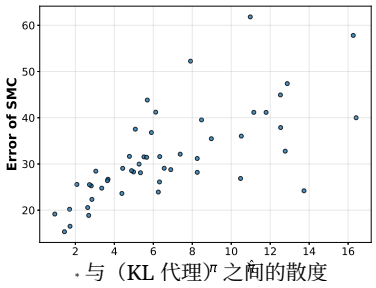


(c) $r = 0.88$

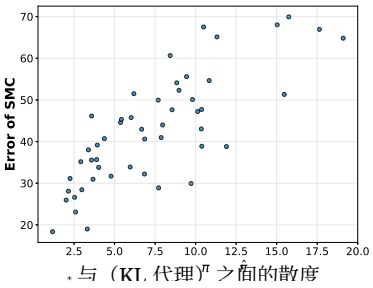


(d) $r = 0.86$

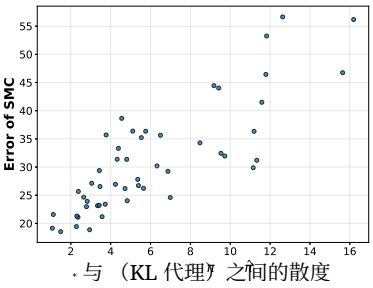
图 6：图中所示实验的其他实例 图 2a，采用不同的数据集（由 GPT-5.2 生成）。标题显示皮尔逊相关系数。



(a) $r = 0.70$

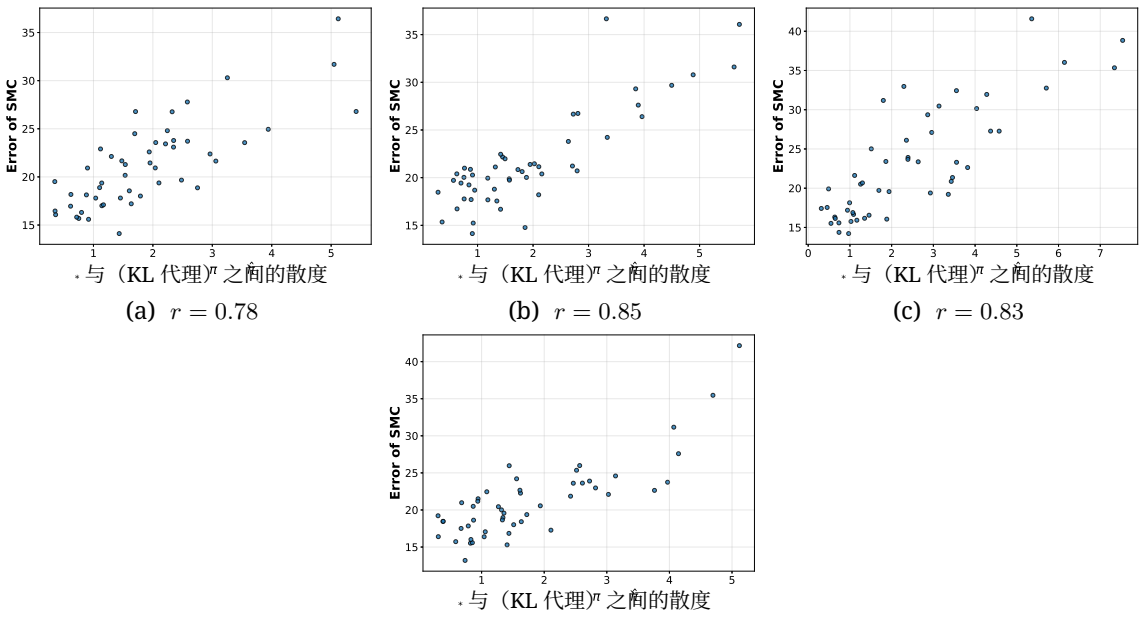


(b) $r = 0.79$



(c) $r = 0.83$

图 7：对图 2b中所示实验的更多实例，使用不同的数据集（由 GPT-5.2 生成）。图注显示皮尔逊相关系数。



(a) $r = 0.78$ (b) $r = 0.85$ (c) $r = 0.83$

(d) $r = 0.82$

图 8：使用 4 个数据集所示实验的额外实例化，见图 2b 和 7，区别在于使用 $\alpha = 1$。图注显示皮尔逊相关系数。