arXiv:2603.04479v1 [stat.ML] 4 Mar 2026

科拉茨停止时间的贝叶斯建模：概率机器学习视角

Nicolò Bonacorsi* Matteo Bordoni†

2026年3月6日

摘要

我们从概率机器学习的视角研究 $n \le 10^7$ 范围内的考拉兹总停止时间 $\tau(n)$。从经验上看，$\tau(n)$是一种偏斜且过度离散程度很高的计数，其算术异质性十分显著。我们开发了两种互补模型。首先，贝叶斯分层负二项回归（NB2-GLM）利用简单的协变量（对数 n 和模 8 的剩余类 n ）预测 $\tau(n)$，并通过后验分布和后验预测分布量化不确定性。其次，我们提出一种基于奇数块分解的机制性生成近似：对于奇数 m，写作 $3m+1 = 2^{K(m)}m'$，其中 m' 为奇数，且 $K(m) = v_2(3m+1) \ge 1$；对这些块长度进行随机化后，即可得到一种通过狄利克雷-多项式更新校准的随机逼近。在留出数据上，NB2-GLM 的预测似然显著高于奇数块生成器。将块长度分布按 m 模 8 进行条件化，可显著改善生成器的分布拟合，这表明低阶模结构是 $\tau(n)$异质性的关键驱动因素。

1 引言

令 $T:\ \mathbb{N} \to \mathbb{N}$ 为科拉茨映射

$$T(n) = \begin{cases} n/2, & n \text{ even}, \\ 3n+1, & n \text{ odd}, \end{cases}$$

并定义总停止时间

$$\tau(n) = \min\{t \ge 0 :\ T^t(n) = 1\}.$$

科拉茨猜想断言 $\tau(n) < \infty$ 对所有 n 都成立，但至今仍未解决 [1–4]。本文并不试图证明该猜想，而是将 n 视为随机变量，并研究其在大范围内诱导出的经验分布律 $\tau(n)$ 。

*美国纽约哥伦比亚大学应用数学系。电子邮箱：nb3328@columbia.edu
†美国纽约哥伦比亚大学数学系。电子邮箱：mb5333@columbia.edu

数据集与问题。 我们固定 $N = 10^7$并计算 $\tau(n)$对所有 $1 \le n \le N$，从而得到$D_N = \{(n, \tau(n))\}_{n=1}^{N}$。我们的目标是进行预测和解释： 哪些简单的概率模型能够预测并解释 $\tau(n)$ 所呈现的分布形态及其算术异质性？

两个互补模型。 我们构建了：**(i)** 贝叶斯分层负二项回归，将 $\tau(n)$ 视为过度离散计数，其条件均值取决于 $\log n$和 $n \bmod 8$ (第 3节)；以及 **(ii)** 一种机制生成式奇数块模型，在加速奇数到奇数动力学中将块长度 $K = v_2(3m + 1)$ 随机化（第 4节）。第 5 节比较了模型在留出数据上的预测性能。关于回归变量和奇数块启发式方法的进一步依据见附录 C。

关于随机性的说明（工作似然）。 映射 T 和停止时间$\tau(n)$ 是确定性的。全文中，“概率模型”一词均按标准统计学意义使用，指一种工作似然：我们将 n 视为随机变量（例如，在评估中从$\{1, \dots, N\}$中均匀抽样），并通过参数化随机族对 $\tau(n)$ 在 n之间产生的变异进行建模。这在不假设考拉兹动力学中存在物理噪声的前提下，提供了不确定性量化和有原则的预测比较。

2 数据与探索性分析

大规模计算 $\tau(n)$

对每个 $n \le 10^7$ 进行直接模拟的代价很高，因为轨迹很长。我们采用动态规划：在计算 $\tau(i)$时，如果轨道到达某个 $m < i$，就停止计算，并设置 $\tau(i) = k + \tau(m)$，复用缓存值。Numba 即时编译可加速核心循环 [5]。详情见附录 A。

分布形态与过度离散

图 1 总结了 $\tau(n)$ 的边缘分布（整数对齐分箱的直方图叠加核密度估计曲线）。重右尾现象表明有必要采用过度离散计数似然。表 1报告了 $N = 10^7$的关键统计摘要，包括离散比 $R = \widehat{\mathrm{Var}}(\tau)/\widehat{\mathbb{E}}[\tau] \approx 24.56 \gg 1$，这排除了将泊松分布作为现实的噪声模型 [6, 7]。

尺度效应与算术异质性

图2 绘制了 $\tau(n)$ 与 n 的关系（对数x 轴）：均值随 n 缓慢增长，离散程度也随之增大（异方差性）。可见的条带反映了算术结构；一个稳健的代理变量是模小的 2 的幂所得的剩余类，因此我们将 n 模 8 作为回归模型中的低维分类特征。

3 方法 1：贝叶斯负二项回归

经验值 $R \gg 1$ 表明应采用负二项似然。我们使用 NB2 均值-离散度参数化：在给定 (μ_n, α) 的条件下，

$$Y_n = \tau\ (n)\ \ |\mu_n, \alpha \sim \mathrm{NB}(\mu_n, \alpha), \qquad \mathbb{E}[Y_n \mid \mu_n, \alpha] = \mu_n, \quad \mathrm{Var}(Y_n \mid \mu_n, \alpha) = \mu_n + \alpha\,\mu_n^2,$$

with 参数映射及 PyMC 实现说明，见附录 D .

分层 NB2-GLM

我们通过对数链接将 μ_n 与协变量关联起来，并按模 8 的剩余类设置随机截距：

$$\log \mu_n = \eta_n, \tag{1}$$

$$\eta_n = \beta_0 + \beta_{\log}\, \log n + u_{\mathrm{mod8}(n)}, \tag{2}$$

$$u_r \mid \sigma_u \sim \mathcal{N}(0, \sigma_u^2), \qquad r = 0, 1, \ldots, 7. \tag{3}$$

这种分层结构在各同余类之间进行部分汇聚，从而得到稳定的类别特异性偏差，同时控制过拟合。

推断与预测评估

我们采用弱信息先验，并在附录 B.2 中对其进行完整说明。我们在 PyMC [8] 中使用 NUTS [9]，在大小为 $N_{\mathrm{fit}} = 50{,}000$ 的训练子样本上进行拟合，并在大小为 $N_{\mathrm{test}} = 50{,}000$的独立测试子样本上进行评估。两个子样本均从 $\{1, \ldots, N\}$ 中采用固定的随机数生成器种子进行均匀无放回抽样（确切流程见附录 B）。对于马尔可夫链蒙特卡洛，我们运行 2 条链，每条链包含 1000 个调优步数和 1000 次抽样，并将目标接受率设为 0.9。

图 3 展示了留出数据的后验预测检验（PPC）：模型能够很好地匹配主体部分的分布，并略微夸大最右侧尾部，这与二次NB2方差一致。

4 方法 2：随机奇数块生成模型

回归具有预测能力，但属于唯象的。因此，我们基于奇数块分解构建一种机制性近似。当 m 为奇数时，$3m + 1$ 为偶数，可以写成

$$3m + 1 = 2^{K(m)} m', \qquad m' \text{ odd}, \qquad K(m) = v_2(3m + 1) \geq 1.$$

因此，一次奇数更新后会经过 $K(m)$ 个减半步骤，直到返回奇数状态。

加速奇数到奇数动力学

令 m_0 为从 n 出发遇到的第一个奇数迭代（经过 $v_2(n)$次初始减半步骤后）。定义加速映射

$$m_{j+1} = \frac{3m_j + 1}{2^{K(m_j)}}.$$

总停止时间可分解为

$$\tau(n) = v_2(n) + \sum_{j=0}^{J-1} \Big(1 + K(m_j)\Big),$$

其中， J 是满足 $m_J = 1$的第一个索引。对于确定性动力学，这个恒等式是精确的。

随机块长度生成器

生成式近似　　将 $K(m_j)$替换为随机序列（K_j），其概率质量函数（p_k）为$k\geq 1$　　：

$$M_{j+1} \approx \left\lfloor \frac{3M_j + 1}{2^{K_j}} \right\rceil_{\text{odd}}, \qquad \tau_{\text{gen}}(n) = v_2(n) + \sum_{j=0}^{J-1} (1 + K_j). \tag{4}$$

取整 $\lfloor \cdot \rceil_{\text{odd}}$ 反映了这样一个实现细节：由于整数运算，随机更新可能产生一个偶数中间值；为保持奇数到奇数结构，我们将其投影回奇数状态空间。

经典启发式表明 K 接近几何分布，其参数为 $\mathbb{P}(K = k) \approx 2^{-k}$， $k \geq 1$。在这一基准下， $\mathbb{E}[3/2^K] = 1$ 但对数漂移为负（附录 F），这与对数尺度上的收缩一致。

奇数投影的定义 $\lfloor \cdot \rceil_{\text{odd}}$。在生成器中，我们首先将结果舍入到最近整数，然后投影到奇数状态空间，以实现奇数到奇数的更新：令 $\lfloor x \rceil$ 表示舍入到最近整数（遇到中点时取偶数），并定义

$$\lfloor x \rceil_{\text{odd}} = \begin{cases} 1, & \lfloor x \rceil \leq 1, \\ \lfloor x \rceil, & \lfloor x \rceil \text{ odd}, \\ \lfloor x \rceil + 1, & \lfloor x \rceil \text{ even and } \lfloor x \rceil > 1. \end{cases}$$

这与用于生成第5节结果的实现一致　　.

校准 p_k 和后验预测检验

我们根据观测到的块长度 p_k 估计 $K_i = v_2(3m_i + 1)$，使用狄利克雷先验，从而得到共轭狄利克雷后验 [10]（附录E）。

在实现中，我们将 K 离散化到$\{1, \ldots, K_{\max}\}$，其中通过将 K_{cap} =限制为$\min(K, K_{\max})$来设置 $K_{\max} = 30$，因此最后一个类别聚合了尾部事件$\{K \geq K_{\max}\}$。对于 $k = 1, \ldots, K_{\max}$，我们使用浓度为 $\alpha_k = 1$ 的对称狄利克雷先验。

图 4 在对数尺度上比较经验频率与几何分布参照；这直接检验了关键的建模假设。图 5 展示了 (p_k) 的后验不确定性。图 7 通过分布型后验预测检验，将生成的停止时间分布与经验停止时间进行比较。

5 模型比较：预测准确度与机制忠实度

后验预测检验是定性的。为了在公平的基础上比较回归方法和生成方法，我们在留出数据上使用一项 适当评分规则 [11]：对数预测得分。在测试集 $\{(n_i,\ y_i)\}_{i=1}^{N_{\text{test}}}$且 $y_i = \tau(n_i)$ 的情况下，定义

$$\text{Score} = \sum_{i=1}^{N_{\text{test}}} \log p(y_i \mid n_i, \mathcal{D}_{\text{fit}}).$$

对于 NB2-GLM，$p(y_i \mid n_i,\ \mathcal{D}_{\text{fit}})$ 是通过对后验抽样中的 NB 概率质量函数取平均得到的。对于奇数块生成器，我们通过蒙特卡罗模拟估计预测概率质量函数：$\hat{}p_i(y) = \frac{1}{S_{\text{MC}}} \sum_{s=1}^{S_{\text{MC}}} \mathbf{1}\{\tau_{\text{gen}}^{(s)}(n_i) = y\}$ 并使用 $\varepsilon = 10^{-12}$ 进行数值稳定处理，对 $\sum_i \log(\hat{}p_i(y_i) + \varepsilon)$ 评分。每个测试点使用 $S_{\text{MC}} = 40$ 次蒙特卡罗重复。需要强调的是，生成器的对数评分是理想对数预测得分的蒙特卡罗近似，因为 $\hat{}p_i(y_i)$ 是根据有限次重复估计的。

此外，我们报告经验值与预测累积分布函数之间的分布距离（一阶瓦瑟斯坦距离 [12]），以概括整体形状失配情况。表 2 给出了报告格式。

解读。表2 显示，分层 NB2 广义线性模型在留出集上的预测似然远胜其他模型：其对数评分为 -2.73×10^5 （平均每个观测值 -5.46 ），而全局奇数块生成器G2为 -1.17×10^6，条件生成器G3为 -1.08×10^6。换言之，该回归模型对实际观测到的停止时间赋予了显著更高的概率。就分布拟合而言，NB2-GLM的 W_1 距离也最小（$W_1 \approx 3.20$）；G3较G2有明显改善（$W_1 \approx 5.43$ 对比17.59），但仍逊于回归模型。

因此，从纯粹的预测角度和适当评分规则来看，NB2-GLM（M3）是所考察模型中的最佳模型。不过，从机制角度看，奇数块生成器仍颇具吸引力：它通过随机化的块长度 $K = v_2(3m+1)$ 及其对 $m \bmod 8$的算术依赖性，解释了重尾和异质性。条件变体G3缩小了预测性能上的差距，同时保留了这种结构性解释；当优先考虑机制忠实性而非原始对数评分时，它便是一个自然的候选模型。

6 讨论

由此得出两个互补的结论。首先，一个紧凑的分层 NB2 广义线性模型，以对数 n 和 n 模 8 刻画了 $\tau(n)$ 的大部分预测结构，并通过后验与后验预测检验进行不确定性量化（图 3）。其次，奇数块分解给出一个机制生成器，其成功取决于对 $K = v_2(3m+1)$ 的分布及其算术依赖性的准确建模（图 4–7）。条件变体在两种观点之间架起桥梁：在回归中表现为随机效应的模结构，在生成器中成为显式的条件信息。

未来工作包括将条件结构扩展到二的更高次幂，在（K_j）中纳入显式状态依赖性，并利用生成器的基于似然的校准，使评分规则表现与机制可解释性相一致。

参考文献

[1] Jeffrey C Lagarias。考拉兹问题及其推广。*The American Mathe-matical Monthly*，92(1):3–23，1985。[2] 陶哲轩。几乎所有考拉兹映射轨道都能达到几乎有界的数值。*Forumof Mathematics, Pi*，10:e12，2022。doi: 10.1017/fmp.2022.8。[3] Riho Terras。正整数上的停止时间问题。*Acta Arithmetica*，30(3): 241–252，1976。[4] 西奈。统计（3x + 1）问题。*Communications on Pure and AppliedMathematics*，56(7):1016–1028，2003。doi: 10.1002/cpa.10084。[5] Siu Kwan Lam、Antoine Pitrou 和 Stanley Seibert。Numba：基于 LLVM 的 Python JIT编译器。在 *Proceedings of the Second Workshop on the LLVM Compiler Infrastructurein HPC*，第 1–6 页，2015 年。[6] A Colin Cameron 和 Pravin K Trivedi。计数数据的回归分析。剑桥大学出版社，第 2 版，2013 年。[7] Andrew Gelman、John B Carlin、Hal S Stern、David B Dunson、Aki Vehtari 和 Donald B Rubin。贝叶斯数据分析。CRC 出版社，第 3 版，2013 年。[8] Matthew D Hoffman 和 Andrew Gelman。无U型转弯采样器：在哈密顿蒙特卡罗中自适应设置路径长度。*Journal of Machine Learning Research*，15(1): 1593–1623，2014。[9] Oriol Abril-Pla 等。PyMC：Python 中现代且全面的概率编程框架。*PeerJ Computer Science*，9:e1516，2023。[10] Kevin P Murphy。机器学习：概率视角。MIT 出版社，2012 年。[11] Tilmann Gneiting 和 Adrian E Raftery。严格恰当评分规则、预测与估计。*Journal of the American statistical Association*，102(477):359–378，2007。[12] Cédric Villani。最优传输：古今。Springer Science & Business Media，2009 年。

附录

构建数据集的方法

当 N 达到 10^7, 时，由于某些初始值可能需要经历很长的轨迹，朴素的停止时间计算实现会变得过于缓慢。为了获得可扩展的计算引擎，我们使用 Python，结合 NumPy 数组和 Numba 的即时（JIT）编译功能 [5]。Numba 将核心循环编译为高效的机器代码，大幅减少 Python 开销，从而能够在标准的类似 Colab 的运行环境中计算 τ（n）内的所有 $n \le N$ 。

我们还利用了动态规划捷径。计算给定 i 时的 $\tau(i)$，轨迹 (n_k) 最终可能会经过数值 $m < i$。此时，到达 1 所需的剩余步数恰好是此前已计算过的 $\tau(m)$。因此，我们可以终止迭代并设置

$$\tau(i) = k + \tau(m),$$

复用缓存值。对于较大的 N，这种复用尤为有效，因为许多轨迹会汇入已探索过的路径段。

这一阶段的输出是数据集

$$\mathcal{D}_N = \{(n, \tau(n)) : \ 1 \le n \le N\},$$

以包含 N 行的表格形式存储。在后续分析中，我们会为每一行添加 n 的简单派生特征（例如 $\log n$ 和 $n \bmod 8$），作为回归模型中的协变量。

B 可复现性细节（划分、先验和评估）

B.1 训练/测试方案

我们使用随机数生成器种子 123，从{1, . . . ,{v8} 中均匀且不重复地抽取 $N_{\text{fit}} = 50{,}000$ 个索引，并将剩余部分作为测试候选集。随后，我们再次使用同一随机数流，从剩余部分中均匀且不重复地抽取 $N_{\text{test}} = 50{,}000$ 个测试索引。

B.2 M3 的先验和马尔可夫链蒙特卡洛设置

对于分层 NB2 广义线性模型（M3），我们采用：

$$\beta_0 \sim \mathcal{N}(0, 10^2), \qquad \beta_{\log} \sim \mathcal{N}(0, 5^2), \qquad \sigma_u \sim \text{HalfNormal}(1), \qquad \alpha \sim \text{HalfNormal}(5),$$

其中 $u_r = \sigma_u z_r$, $z_r \sim \mathcal{N}(0, 1)$ 适用于 $r = 0, \ldots, 7$。我们使用 NUTS（PyMC）进行抽样，采用 2 条链、每条链 1000 个调优步数和 1000 个后验抽样，目标接受率为 0.9，并使用固定的随机种子。

B.3 生成器评估设置

对于奇数块生成器，我们将模拟停止时间循环的上限设为 `MAX_STEPS` $= 200{,}000$ 步，以避免在随机逼近下出现病态的非吸收。对数评分对每个测试点使用 $S_{\mathrm{MC}} = 40$ 次蒙特卡洛重复，$\varepsilon = 10^{-12}$，并使用固定的随机数生成器种子，以确保完全可复现。

C 建模选择与启发式方法（依据）

本附录阐述了本文的主要建模选择：(i) NB2-GLM 中使用的回归变量；(ii) 负二项似然；以及 (iii) 奇数块生成器及其变体的设计。本文并非旨在提出新理论，而是要使这些建模决策清晰透明且有理有据。

C.1 为什么将 $\tau(n)$ 建模为过度离散计数（采用 NB2 而非泊松分布）

响应变量 $\tau(n)$ 是非负整数值，因此采用计数似然是合理的。泊松模型要求 $\mathrm{Var}(Y \mid x) \approx \mathbb{E}[Y \mid x]$，而我们的数据显著违背了这一要求：$\widehat{\quad}$ $\widehat{\quad}$经验离散比 $\mathrm{Var}(\tau)/\mathbb{E}[\tau]$ 远高于 1（表 1）。因此，我们采用允许方差增长速度超过均值的负二项似然。

我们采用 NB2 均值–离散度参数化

$$\mathrm{Var}(Y \mid \mu, \alpha) = \mu + \alpha\mu^2,$$

因为它同时捕捉了两个经验上可观察到的特征：(i) 显著的过度离散，以及 (ii) 条件均值越大，离散程度也越大（异方差性）。在实践中，当数据呈现重尾和方差膨胀时，NB2 也是应用计数回归中的标准默认选择。

C.2 为什么将 $\log n$ 作为主要尺度回归变量

一个基本的经验事实是，$\tau(n)$ 随 n 缓慢增长，但并非线性增长（图2）。直接使用n 作为回归变量，会迫使模型引入过多曲率，以拟合一个从图形上看更接近对数增长的趋势。变换$\log n$ 提供了一个简单的一维尺度概括，它：(i) 压缩动态范围 $1 \le n \le 10^7$，(ii) 符合观察到的典型停止时间缓慢增长的规律，且 (iii) 在对数链接下能获得稳定的数值推断。

简言之，$\log n$ 是能够捕捉宏观趋势的最小尺度特征，且不会引入更难解释或论证的灵活非线性基底，从而使我们能够保持模型简约性和可解释性。

C.3 为什么纳入剩余类 n 模 8 （以及为什么将其作为随机效应）

除了尺度效应之外，考拉兹动力学还呈现出明显的算术分层：将 $\tau(n)$对照 n 绘制时，2（图中）会形成带状分布，这表明粗略的模特征已经能够解释异质性中相当一部分。我们采用 n 模 8 作为低维分类协变量，原因如下：

- 它是一个精简的“二的幂”模特征（规模足够小，不会形成高维分类设计），
- 它与考拉兹步骤中 2-adic 结构的重要性相吻合（反复减半），并且
- 它在经验上具有稳健性：即使在这一分辨率下，分带现象也已清晰可见。

我们将八个类别的偏移量建模为层级式（部分汇聚）随机效应 $u_r \sim \mathcal{N}(0, \sigma_u^2)$，而不是八个互不相关的固定效应。这一选择能稳定推断，并通过在数据缺乏有力证据支持较大差异时，将各类别的偏差收缩至 0，来避免过拟合。这也符合我们的建模目标：我们希望量化“模 8 很重要”这一点，同时保持模型精简。

C.4 为什么我们没有添加更多回归变量

还可以加入其他预测变量（例如 $v_2(n)$、更多样的剩余类，或 $\log n$ 的非线性函数）。我们有意保持回归模型简洁，以保留其可解释性：重点并非不惜一切代价最大化预测准确度，而是展示一组最精简且有理论依据的协变量，已足以带来强劲的预测表现和实用的不确定性量化。

C.5 奇数块生成器：为何它是一种自然的机制近似

奇数块分解将唯一的“扩张”步骤（奇数 $\mapsto 3m+1$）与其后的一系列确定性减半步骤区分开来。对于奇数 m，将 $3m+1 = 2^{K(m)}m'$写成 m' （其中 m' 为奇数），即可将动力学转化为奇数到奇数的跳跃，以及一个块长度 $K(m)$。由此可得到一种机制近似：用随机替代分布取代$K(m)$ 对 m 的复杂确定性依赖关系。

基准启发式 $\Pr(K = k) \approx 2^{-k}$ 体现了一种直觉： $3m+1$ 在其 2-adic 赋值方面大致表现得像一个“随机偶整数”，因此 2 的因子个数近似服从几何分布。校准后的变体（ G2/G3）保留了相同的机制，但依据数据估计块长度分布，使这种近似可以接受检验，而不再只是纯粹的启发式。

C.6 为什么对 m 模 8 进行条件化是首个有意义的细化

全局独立同分布块长度模型（G2）对强烈的算术异质性取平均。最小限度的改进是对 p_k 按与 2-adic 结构直接相关的低阶剩余类进行条件化。对 m 模 8 进行条件化是仍具信息量的最简单选择：它维度较低（8 个类），却已能捕捉估计块长度分布中的系统性差异（图 6）。G3 的后验预测检验有所改善，且 W_1 距离有所降低，这表明生成器从这些模信息中获益颇多。

C.7 为什么采用这些评估指标

我们报告两项互补的诊断指标：

- **对数预测得分**（适当评分规则）：它评估整个预测分布，并对为观测结果赋予更高概率的模型给予更高评分。这是用于比较预测表现的主要规范性指标。

- W_1（**一阶瓦瑟斯坦距离**）：它概括经验边际分布与预测边际分布之间的整体分布差异；在存在重尾，且可视化后验预测检验显示形状差异时，这一指标尤其有参考价值。

对于奇数块生成器，预测分布没有闭式形式，因此通过在每个测试点进行蒙特卡洛“命中”频率估计对数评分，重复次数为 S_{MC}。这给出了适当评分的近似值；增加 S_{MC} 可提高估计精度，但计算成本也会更高。

D 负二项回归：参数化形式与性质

本附录概述正文中使用的负二项分布（NB）和负二项广义线性模型（GLM）约定。

D.1 经典(r, p) 参数化

经典负二项分布可定义为在成功概率为$p \in$（0,1）的独立同分布伯努利试验中，第 r次成功之前观察到的失败次数 Y 所服从的分布：

$$Y \sim \mathrm{NB}(r, p), \qquad r > 0,\ p \in (0, 1).$$

其概率质量函数为

$$\Pr(Y = y \mid r, p) = \frac{\Gamma(y + r)}{\Gamma(r)\,\Gamma(y + 1)} (1 - p)^y p^r, \qquad y = 0, 1, 2, \ldots,$$

and

$$\mathbb{E}[Y] = r\,\frac{1 - p}{p}, \qquad \mathrm{Var}(Y) = r\,\frac{1 - p}{p^2}.$$

D.2 均值–离散度 (μ, α) 参数化（NB2）

对于回归，我们采用 NB2 约定

$$\mathbb{E}[Y \mid \mu, \alpha] = \mu, \qquad \mathrm{Var}(Y \mid \mu, \alpha) = \mu + \alpha\,\mu^2,$$

其中， $\alpha > 0$ 控制过度离散， $\alpha \to 0$ 产生泊松极限。映射到（r, p）的关系为

$$r = \frac{1}{\alpha}, \qquad p = \frac{1}{1 + \alpha\mu}.$$

D.3 实现说明：映射到 PyMC

在 PyMC 中， `NegativeBinomial(mu=mu, alpha=alpha_PyMC)`使用

$$\mathrm{Var}(Y \mid \mu, \alpha_{\mathrm{PyMC}}) = \mu + \frac{\mu^2}{\alpha_{\mathrm{PyMC}}}.$$

因此，为匹配上述 NB2 约定，我们设定

$$\alpha_{\mathrm{PyMC}} = \frac{1}{\alpha}.$$

D.4 泊松–伽马混合视图

若 $\Lambda \sim$服从伽马分布(r， c)，且 $Y \mid \Lambda \sim$服从泊松分布(Λ)，则 $Y \sim$服从NB(r， p)，其中 $p = c/($ $c + 1$)。在NB2形式下，取 $r = \alpha^{-1}$ 与 $c = \alpha^{-1}/\mu$ 可得到方差(Y) $= \mu + \alpha\mu^2$，从而将过度离散解释为潜在速率异质性。

E 狄利克雷–多项式更新

若 K_1，……， K_m 为概率为 p_1，……， $p_{K_{\max}}$ 的独立同分布的分类变量，且先验为 $p \sim$狄利克雷分布(α)，则后验为狄利克雷分布($\alpha + c$)，其中 $c_k = \sum_{i=1}^{m} \mathbf{1}\{K_i = k\}$。后验均值和方差为

$$\mathbb{E}[p_k \mid K_{1:m}] = \frac{\alpha_k + c_k}{\sum_j (\alpha_j + c_j)}, \qquad \mathrm{Var}(p_k \mid K_{1:m}) = \frac{(\alpha_k + c_k)\,(\alpha_0 + m - (\alpha_k + c_k))}{(\alpha_0 + m)^2(\alpha_0 + m + 1)},$$

其中 $\alpha_0 = \sum_j \alpha_j$。

F 对数尺度随机游走近似

对于较大的状态，忽略加性 $+1$ 项，奇数块更新满足 $M_{j+1} \approx$($3/2^{K_j}$)M_j，因此在对数尺度上

$$X_{j+1} \approx X_j + \log_2 3 - K_j, \qquad X_j = \log_2 M_j.$$

即使在 $\mathbb{E}[3/2^K] = 1$ 的情况下，詹森不等式也意味着 $\mathbb{E}[\log_2($$3/2^K$$)] < \log_2 \mathbb{E}[3/2^K] = 0$, 从而在 X_j 中产生负漂移，并表明在较小数值处吸收是常见现象。

图目录

1 $1 \le n \le N$ 的 $\tau(n)$ 经验分布（整数对齐分箱，宽度为 2），叠加了基于大型子样本计算的核密度估计，以降低噪声。这为过度离散计数似然提供了依据.13 2 $\tau(n)$ 与 n （对数-x）的散点图。均值缓慢增加，并且大致是 $\log n$ 的线性函数；随着 n 增大，离散程度也随之增加。条带状分布表明存在模结构，因此采用 $\log n$ 和 $n \bmod 8$ 作为协变量.13 3 分层 NB2-GLM（模型 M3）的后验预测检验。PPC 对主体部分的拟合良好，并轻微高估了极端右尾质量。 14 4 $K = v_2(3m+1)$（奇数 $m \le N$）的经验块长度分布 $\hat{}p_k$ ，与对数-y尺度上的几何分布参照 2^{-k} 进行比较。此图用于评估“几何 K”启发式。 14 5 (p_k) 的狄利克雷后验（对数尺度），带有不确定性误差线，并与几何分布参照 2^{-k}进行比较.15 6 在奇数剩余类 $m \bmod 8$ 条件下 p_k 的后验均值（热力图）。这揭示了独立同分布几何分布之外系统性的算术依赖性。 15 7 PPC叠加图：经验值 τ 与 G1（几何 2^{-k}）、G2（全局校准 p_k）和 G3（在 $m \bmod 8$ 条件下的 p_k ）进行比较。条件化降低了均值偏差，并改善了主体部分的分布一致性.16 8 $n = 27$ 的对数尺度示例轨迹：确定性考拉兹与随机奇数块近似.16

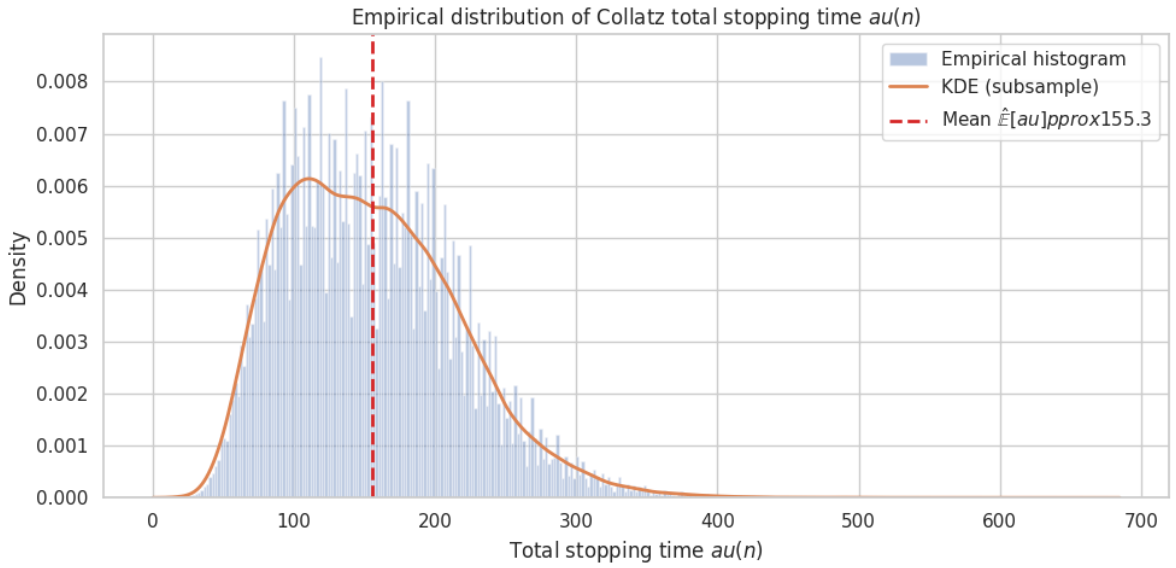


图 1： $1 \le n \le N$ 中 $\tau(n)$ 的经验分布（整数对齐分箱，宽度为 2），叠加了基于大型子样本计算的核密度估计，以减少噪声。这为采用过度离散计数似然提供了依据。

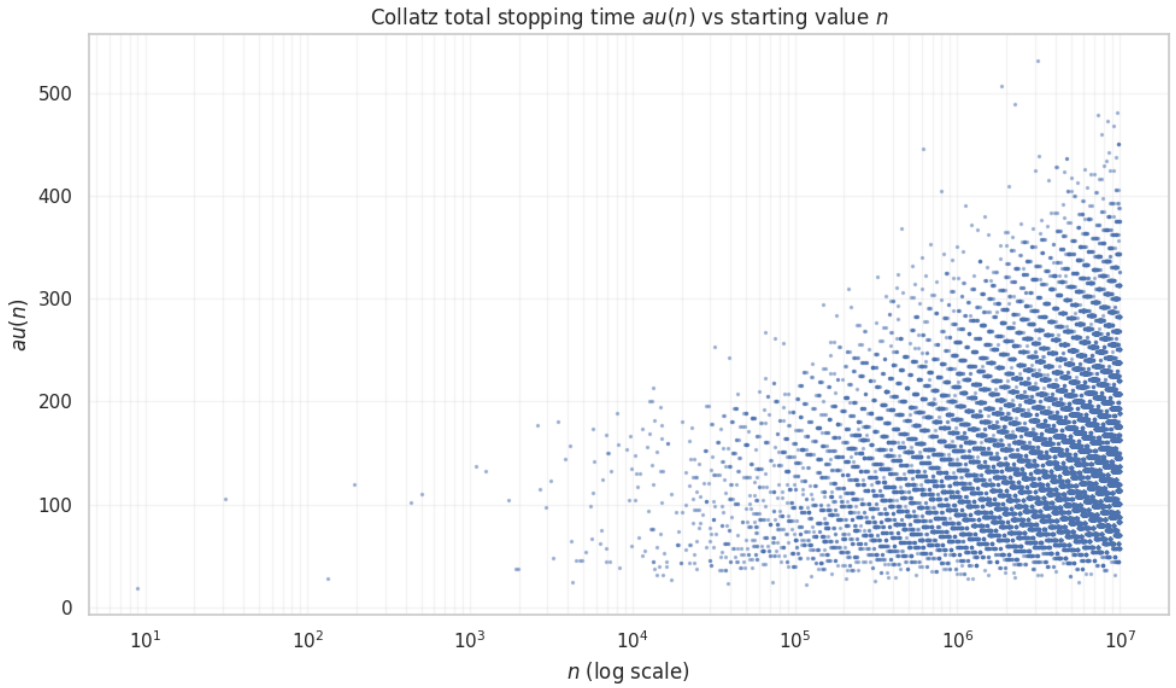


图 2： $\tau(n)$ 与 n （对数-x）的散点图。均值缓慢增加，且关于 $\log n$ 近似呈线性，而离散程度随 n 增大；条带状分布表明存在模结构，因此将 $\log n$ 和 $n \bmod 8$ 作为协变量。

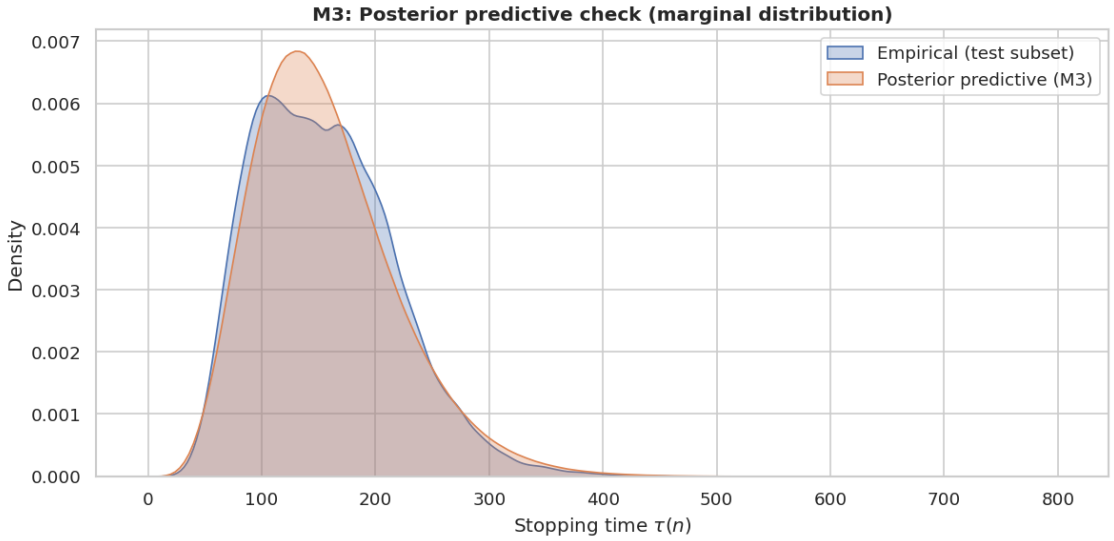


图3：分层 NB2 广义线性模型（模型 M3）的后验预测检验。后验预测检验与主体部分吻合良好，并略微高估了极端右尾质量。

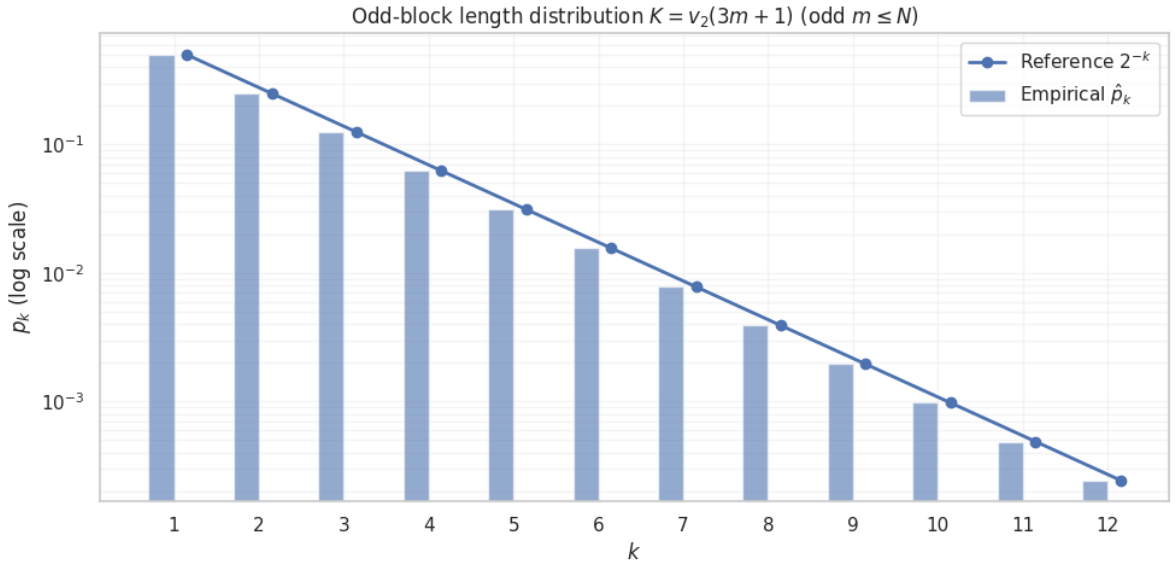


图4： $K = v_2(3m+1)$（奇数 $m \le N$）的经验块长度分布 $\hat{p}_k$ 与对数-y 上的几何分布参照 2^{-k} 的对比。该图用于评估“几何 K”启发式方法。

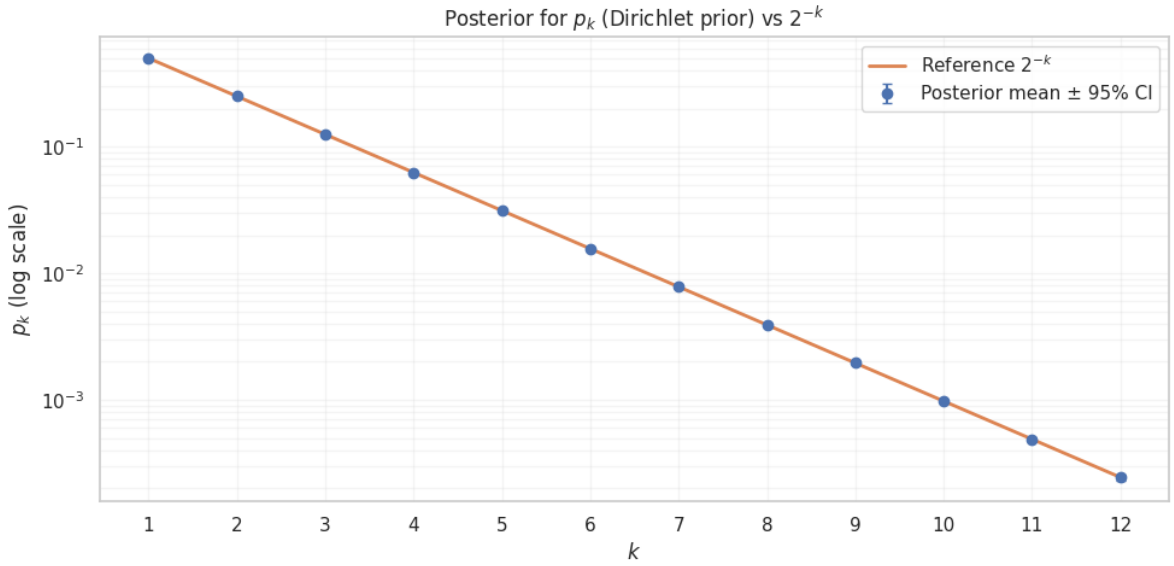


图5：(p_k) 的狄利克雷后验（对数尺度），带有不确定性误差线，并与几何分布参照 2^{-k} 进行比较。

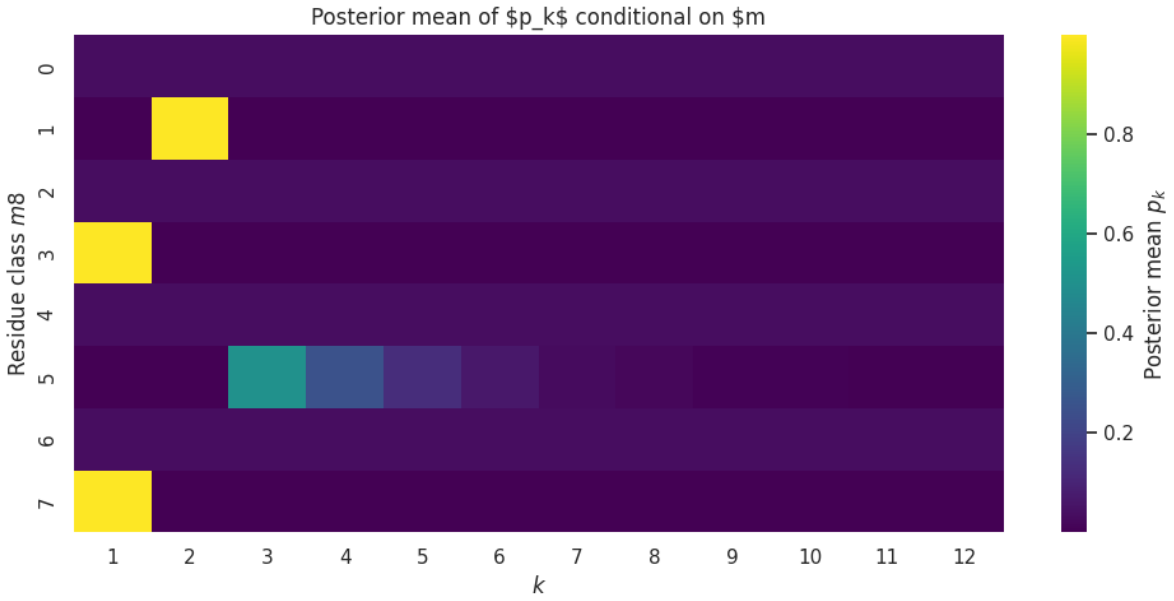


图6：在奇数剩余类 m 模 8 条件下 p_k 的后验均值（热力图）。这揭示了超出独立同分布几何分布范围的系统性算术依赖性。

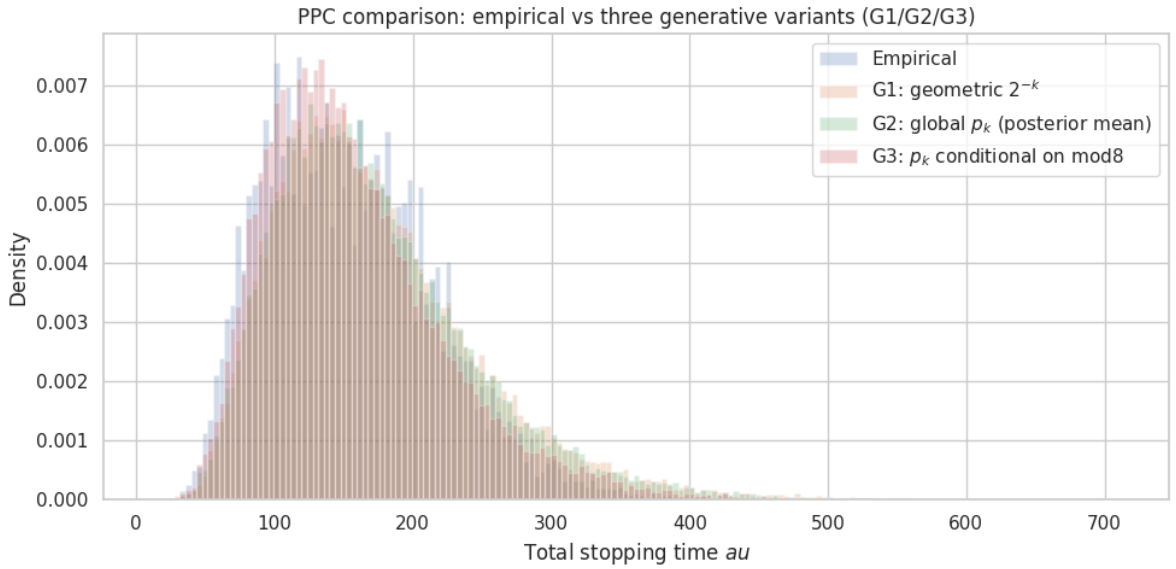


图7：PPC叠加图：经验值 τ 与 G1（几何 2^{-k}）、G2（全局校准 p_k）和 G3（p_k 条件于 m 模 8）。条件化减小了均值偏差，并提升了主体部分的分布一致性。

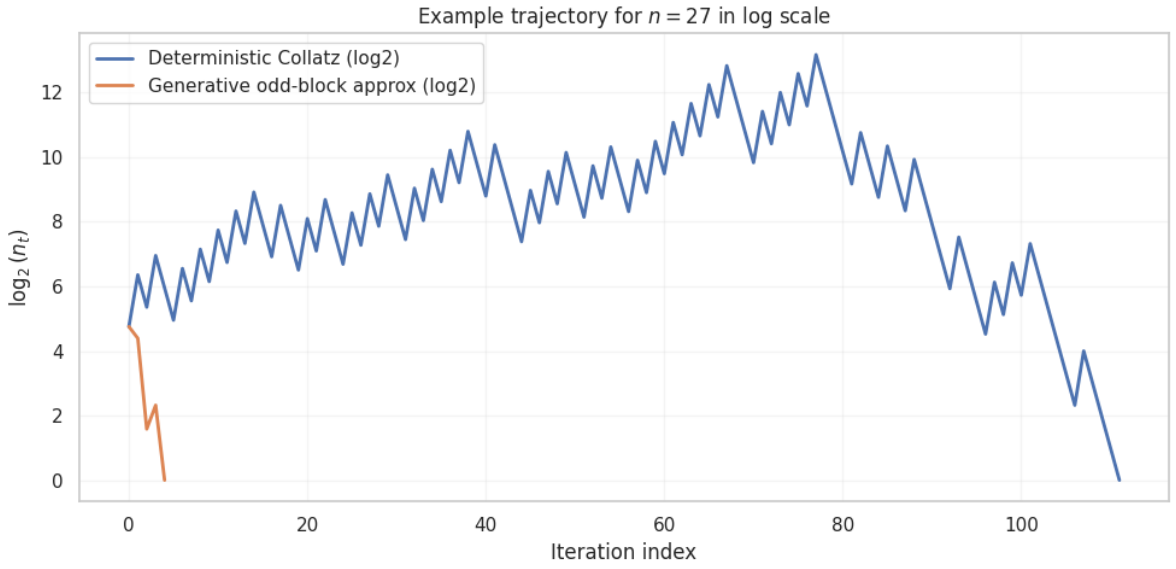


图8： $n = 27$ 的对数尺度示例轨迹：确定性考拉兹与随机奇数块近似。

表目录

的描述性统计 $\tau(n)$，共 $1 \le n \le N$。.182 使用恰当对数预测得分和分布距离对留出数据进行定量比较。所有得分均基于 $N_{\text{test}} = 50{,}000$ 个留出对计算，其中生成对数得分通过每个测试点进行 $S_{\text{MC}} = 40$ 次蒙特卡洛重复估计。.18

统计量	数值
N	10,000,000
$\tau_{\min}$	0
$\tau_{\max}$	685
$\widehat{\mathbb{E}}[\tau]$	155.272
$\widehat{\text{方差}}(\tau)$	3814.045
$\widehat{\text{方差}}(\tau)/\widehat{\mathbb{E}}[\tau]$	24.564

表1： $1 \le n \le N$ 上 $\tau(n)$ 的描述性统计。

模型	对数评分（越高越好）	W_1 距离（越低越好）
NB2-GLM（M3）	$-272\,911.95$	3.199
奇数块 G2（全局 p_k）	$-1\,165\,983.43$	17.594
奇数块 G3（$p_k \mid m$ 模 8）	$-1\,079\,086.65$	5.434

表2：使用恰当对数预测得分和分布距离对留出数据进行定量比较。所有得分均在 $N_{\text{test}} = 50,000$个留出对上计算，生成对数得分通过每个测试点 $S_{\text{MC}} = 40$ 次蒙特卡洛重复估计。