arXiv:2603.02356v1 [cs.LG] 2 Mar 2026

学习最优搜索策略

Stefan Ankirchner * Maximilian Philipp Thiel †

2026年3月4日

摘要

我们以停车问题为例，探讨如何学习最优搜索策略。在该问题中，停车机会按照一个未知的非齐次泊松过程到达。最优策略是一种阈值型停止规则，由一个无差异位置刻画。我们提出一种算法，通过估计积分跳跃强度而非强度函数来学习这一阈值。我们证明，该算法在一大类环境中均能实现对数遗憾增长。此外，我们还证明了对数极小极大遗憾下界，从而确立了所提方法的增长最优性。

*2020*年数学学科分类：主要分类：60G40，93E35；次要分类：62L05，68Q32。关键词：最优停止、停车问题、强化学习、遗憾。

引言

停车问题是搜索问题的一个经典例子。为描述其标准版本，假设一个智能体正沿街道行驶，她不能掉头，并且只能看到下一个车位是否空闲，而看不到其后哪些车位空闲。如果智能体到达一个空闲车位，她必须决定是否占用它。一旦放弃某个空闲车位，便永远放弃。决策规则可以建模为停时。停车问题在于找到使所占用车位到某个给定目标位置的期望距离最小的停时。

*Stefan Ankirchner，数学研究所，耶拿大学，Inselplatz 5，07743 耶拿，德国。电子邮箱：s.ankirchner@uni-jena.de.

†Maximilian Thiel，数学研究所，耶拿大学，Inselplatz 5，07743 耶拿，德国。电子邮箱：maximilian.thiel@uni-jena.de.

要确定最优停止时间，需要知道空闲停车位位置的分布。但如果智能体不知道该分布呢?

停车问题也是一种通常需要连续进行多轮求解，而非只解决一次的问题范例，例如每天早晨驾车前往工作地点时。如果停止代理不知道空闲车位的分布，就可以通过观察每一轮中截至实际选择车位之前的空闲车位位置，逐渐学习该分布。

在本文中，我们探讨了什么样的策略适合在一轮又一轮中选择停车位。我们在连续时间模型的框架下进行研究，其中空闲停车位按照非齐次泊松过程到达。在该模型中，最优停止规则属于阈值型：存在一个位置 b^* ，在该位置之后，选择第一个空闲车位是最优的。位置 b^* 可以被刻画为无差异水平：如果 b^* 处的车位空闲，那么智能体在选择该车位和选择下一个空闲车位之间没有差异。

我们假设停止代理不知道泊松过程的跳跃强度。我们提出了一种称为无差异水平更新（ILU）的特定算法，该算法根据截至目前所获得的观测结果，估计积分跳跃强度，并在每一轮确定停止规则。

为了评估我们算法的质量，我们计算遗憾的增长率，即选择最优停止规则时，目标位置的期望距离与最小期望距离之差的累积值。事实上，我们证明，对于一大类分布，ILU算法所产生的遗憾均以对数级增长。

ILU算法的一个关键特性是，它不估计跳跃强度函数，而是估计积分后的跳跃强度。对于后者，我们有均方误差（MSE）以 $1/n$ 的速率趋于零的估计量，其中 n 是独立过程观测的数量。我们假设跳跃强度连续可微一次，这意味着每轮遗憾可以用估计量的MSE来界定。因此，MSE为 $\mathcal{O}(1/n)$ 时，累积遗憾会以对数级增长。我们指出，对强度函数本身的任何估计量（例如核估计量），其收敛速率都慢于 $\mathcal{O}(1/n)$，因此会导致遗憾以快于对数级的速度增长。

为了证明 ILU算法表现良好，我们证明极小极大遗憾也按对数级增长。这意味着，不存在一种算法能在所有环境中都使遗憾率以比对数更慢的速度增长。从这个意义上说，ILU算法是选择停车位的好方法。

尽管我们是在停车问题的背景下展示研究结果，但

其底层方法并非仅适用于这一应用。它们适用于一类具有随机机会到达特征的时序与搜索问题。为简化阐述，我们以停车问题作为典型示例来展开方法；只需进行少量且自然的调整，就应当能够将其扩展到其他设定。

与文献的比较

将停车问题表述为停止问题的文献十分丰富，例如可参见 [12] ，其中给出了离散时间情形下的早期参考文献。连续时间停车问题的一个版本已在[14]中加以考虑。关于已求解的停车问题变体的最新综述，我们参阅 [8] 的第2.5章。

本文所述的ILU算法可视为基于模型的强化学习算法的一个例子。强化学习（RL）通常被理解为包含那些对模型和系统分布几乎不作假设的算法，例如 q-学习算法。因此，RL算法具有普遍适用性。然而，对于某些应用，这些算法可能相当低效。

为了获得用于学习随机控制问题解的高效算法，自然应尽可能多地利用关于模型的信息。本文提出的ILU算法正是这样一种算法：它利用了最优停止规则为阈值型这一事实，即存在某个位置，在该位置之后，选取第一个空闲车位是最优的。此外，该算法还利用了空闲车位按照非齐次泊松过程到达这一事实。该算法的一个关键组成部分是对泊松过程积分跳跃强度的估计。

关于模型驱动的强化学习算法，已有相当多的文献研究状态动力学由马尔可夫决策过程描述的模型。综述请参见 [11] 。针对连续时间随机控制问题的模型驱动强化学习算法，目前仍相当少见。 [7]提出了一种针对连续时间线性二次问题的初步方法。作者提出了一种加权最小二乘算法，用于学习漂移率，进而求得最优控制。文中未给出遗憾的增长率，而且似乎尚未确定该增长率。 [2, 10, 15]研究了分阶段、有限时间范围设定下的线性二次问题，并给出了非渐近遗憾界。

文章 [13] 研究传播器模型，并描述了实现次线性遗憾的算法。 [6, 3, 5, 4] 在非参数、遍历设定下考察奇异控制问题和脉冲控制问题，并获得了

幂类型的次线性遗憾率。文章 [1]介绍了一种引导算法，以便使过程尽可能接近某条给定路径；该算法利用了过程由布朗运动驱动且漂移率来自某个未知有界区间这一事实。

1 连续时间中的停车问题

在本节中，我们总结连续时间停车问题的一些结果。

假设停车位位于区间 $[S,\infty)$上，其中$S \in (-\infty,0)$。假设空闲停车位按照跳跃强度为 λ 的非齐次泊松过程到达： $[S,\infty) \to (0,\infty)$。更精确地说，设$(N_t^\lambda)_{t\geq S}$ 为定义在某个概率空间$(\Omega\ \mathcal{F}, P)$上的右连续过程，使得增量相互独立，并且对所有 $s \leq t$，$(N_t^\lambda - N_s^\lambda)$服从参数为 $\int_s^t \lambda(u)du$的泊松分布。我们将过程(N_t)的跳跃时刻解释为智能体沿街道从 S 向右行驶时可以使用的空闲停车位的位置。

在本节中，我们假设跳跃强度函数 λ为停止代理所知。在下一节中，我们将去掉这一假设，并假设停止代理不知道 λ，但可以通过观察跳跃过程(N_t)的独立样本来学习它。

我们假设智能体希望停车的位置尽可能接近目标位置 0，使期望距离最小。搜索车位的策略可以用决策规则来描述：对于每个车位 $t \in [S,\infty)$，若其空闲，决策规则会决定是否选择它。

很容易证明，只需考虑以下阈值型决策规则：存在一个阈值 $b \in [S,0]$，超过该阈值后，遇到的第一个空闲车位就会被接受；而在 b之前，任何空闲车位都会被放弃。事实上，对于任意决策规则，都可以构造一个阈值规则，使其期望距离不大于原规则。

我们将阈值为 $b \in [S,0]$ 的决策规则建模为停时$\tau_b := \inf\{t \geq b : N_t > N_b\}$。注意，$\tau_b$ 是 N 在b之后的首次跳跃时间；我们将 τ_b 解释为 b之后第一个空闲停车位的位置。

注意，尽可能将汽车停在靠近目标位置处这一目标，等价于寻找使 $E|\tau_b|$ 达到最小值的阈值 b 的问题。我们将满足 $E|\tau_{b^*}| = \min_{b\in[S,0]} E[|\tau_b|]$ 的 b^* 称为最优阈值。

下一个定理给出了 b^* 成为最优阈值的一个充分条件。

定理 1.1. 设 $b^* \in [S, 0)$ 满足

$$\int_{b^*}^{0} e^{\int_y^0 \lambda(u)du}\, dy = \int_0^{\infty} e^{-\int_0^y \lambda(u)du}\, dy. \tag{1.1}$$

则 b^* 是一个最优阈值。

证明。 首先注意到，对所有 $b \in [S, 0]$ ，我们有

$$E|\tau_b| = -\int_b^0 y\lambda(y)e^{-\int_b^y \lambda(u)du}\, dy + \int_0^{\infty} y\lambda(y)e^{-\int_b^y \lambda(u)du}\, dy.$$

接下来观察到，要使 b 最优，只需它满足一阶条件$\frac{\partial E|\tau_b|}{\partial b} = 0$。我们现在证明(1.1)等价于一阶条件。事实上，

$$\frac{\partial E|\tau_b|}{\partial b} = \lambda(b)\left[b - \int_b^0 y\lambda(y)e^{-\int_b^y \lambda(u)du}\, dy + \int_0^{\infty} y\lambda(y)e^{-\int_b^y \lambda(u)du}\, dy\right],$$

因此$\frac{\partial E|\tau_b|}{\partial b} = 0$ 等价于

$$b = \int_b^0 y\lambda(y)e^{-\int_b^y \lambda(u)du}\, dy - \int_0^{\infty} y\lambda(y)e^{-\int_b^y \lambda(u)du}\, dy.$$

利用 $\lambda(y)e^{-\int_b^y \lambda(u)du} = -\frac{\partial}{\partial y}e^{-\int_b^y \lambda(u)du}$并应用分部积分，可将最后一个方程改写为

$$\begin{aligned}
b &= -\int_b^0 y\frac{\partial}{\partial y}e^{-\int_b^y \lambda(u)du}\, dy + \int_0^{\infty} y\frac{\partial}{\partial y}e^{-\int_b^y \lambda(u)du}\, dy \\
&= -\left[ye^{-\int_b^y \lambda(u)du}\right]_b^0 + \int_b^0 e^{-\int_b^y \lambda(u)du}\, dy + \left[ye^{-\int_b^y \lambda(u)du}\right]_0^{\infty} - \int_0^{\infty} e^{-\int_b^y \lambda(u)du}\, dy \\
&= -(0 - b) + \int_b^0 e^{-(\int_b^0 \lambda(u)du - \int_y^0 \lambda(u)du)}\, dy + 0 - \int_0^{\infty} e^{-(\int_b^0 \lambda(u)du + \int_0^y \lambda(u)du)}\, dy \\
&= b + e^{-\int_b^0 \lambda(u)du}\left(\int_b^0 e^{\int_y^0 \lambda(u)du}\, dy - \int_0^{\infty} e^{-\int_0^y \lambda(u)du}\, dy\right).
\end{aligned}$$

这意味着，一阶条件等价于

$$\int_b^0 e^{\int_y^0 \lambda(u)du}\, dy = \int_0^{\infty} e^{-\int_0^y \lambda(u)du}\, dy.$$

□

下面我们用积分跳跃强度来重写（1.1）的左端

$$\Lambda(y) := \int_y^0 \lambda(u)\, du,\ y \in [S, 0].$$

因此，如果 b^* 满足该方程

$$\int_{b^*}^0 e^{\Lambda(y)}\, dy = \int_0^\infty e^{-\int_0^y \lambda(u)du}\, dy, \tag{1.2}$$

那么 b^* 就是一个最优阈值。

最后要指出，由于假设 λ 为正，因此至多存在一个满足方程 (1.1) 的实数 $b^* \in [S, 0]$ 。

最优性准则的直观解释

注意，对于所有 $t > 0$

$$P(\tau_0 > t) = P(N_t^\lambda - N_0^\lambda = 0) = e^{-\int_0^t \lambda(u)du},$$

因此，(1.1) 的右端与期望 $E[\tau_{b^*}|\tau_{b^*} > 0]$ =$E(\tau_0)$ $= \int_0^\infty P(\tau_0 > t)\ dt.$ 相同。换句话说，(1.1) 的右侧恰好对应于停车位 0 之后首次跳跃时间的期望值。因此，右侧衡量的是驾驶员经过停车位 0 后的预期成本。

方程 (1.1) 的左侧等于 $\overset{|b^*|-E[|\tau_{b^*}|1_{\{\tau_{b^*}<0\}}]}{P}{}_{(\tau_{b^*}\geq 0)}$ 。重新排列各项可得，(1.1) 等价于

$$|b^*| = E|\tau_{b^*}]. \tag{1.3}$$

方程(1.3)将 b^* 表征为这样一个位置：智能体在此处对于是否在车位空闲时占用该车位，还是继续前行直至遇到下一个空闲车位，持无差异态度。

最优性差距

我们用.....表示

$$\Delta(b) := E|\tau_b| - E|\tau_{b^*}| \tag{1.4}$$

选择阈值为 $b \in [S, 0]$ 的停止规则，而非最优阈值 b^*，所产生的最优性差距。下一个引理总结了最优性差距的一些性质。

引理 1.2。 Δ 可微。此外，若 λ 连续可微，则 Δ二阶连续可微，且二阶导数为

$$\begin{aligned}&\Delta''(b) \qquad (1.5)\\ =&\lambda'(b)\left[b-\int_b^0 y\lambda(y)e^{-\int_b^y\lambda(u)du}\,dy+\int_0^\infty y\lambda(y)e^{-\int_b^y\lambda(u)du}\,dy\right]\\ &+\lambda(b)\left[1+b\lambda(b)-\int_b^0 y\lambda(y)\lambda(b)e^{-\int_b^y\lambda(u)du}\,dy+\int_0^\infty y\lambda(y)\lambda(b)e^{-\int_b^y\lambda(u)du}\,dy\right].\end{aligned}$$

证明。 这些结论很容易证明。 □

我们以考察 λ为常数这一特殊情况来结束本节。

推论 1.3。 假设 λ 为大于$\ln(2)/|S|$的常数。那么$b^*=-\ln(2)/\lambda$ 是一个最优阈值，并且

$$\Delta_\lambda(b)=\frac{1}{\lambda}(2e^{\lambda b}-1)-b-\frac{\ln(2)}{\lambda}. \qquad (1.6)$$

2 学习最优规则

根据方程 (1.1)，已知强度函数 λ即可确定最优停止时间 τ_{b^*}。下文假设智能体不知道强度函数 λ。不过，我们假设智能体必须连续进行多轮停止问题，并且在每一轮中，智能体都会观察跳跃过程，直到选定的停止时间。借助这些观测，智能体可以估计真实的强度函数 λ。因此，在第 n 轮中，智能体可以利用前几轮的观测结果来近似最优停止规则：0，1，.....，$n-1$。

设 $(N_t^0)_{t\in[S,\infty)}$，$(N_t^1)_{t\in[S,\infty)}$，.....是可测空间 $(\Omega\ \mathcal{F})$ 上的一个随机过程序列。假设对于每个正的可测强度函数 λ：$[S,\infty)\to(0,\ \infty)$，都存在一个概率测度 P_λ，使得 $(N_t^0)_{t\in[S,\infty)}$，$(N_t^1)_{t\in[S,\infty)}$，.....是一个独立的泊松过程序列，其跳跃强度为 λ。

与上一节类似，我们用 τ_b^i 表示N^i 在时间 $b\in[S,0]$之后的首次跳跃时间。此外， P_λ 下的最优性差距记为 $\Delta_\lambda(b)$。

定义2.1 （策略）。 策略 $(\pi_n)_{n\geq0}$ 是一个随机变量序列，其取值属于 $[S,0]$ ，且对于， π_n 关于 $\mathcal{F}_n$ 可测，其中

$\mathcal{F}_0 = \{\emptyset, \Omega\}$，且 $\mathcal{F}_n := \bigvee_{k=0}^{n-1} \sigma(N^{\lambda,\ k}_{\cdot\ t} : t \in [S, \tau^k_{\pi_k}])$ 对每个 $n \geq 1$. 都成立。（回忆一下，若 $(\mathcal{A}_i)_{i\in I}$ 是由集合 I 索引的一族 σ-代数，则 $\bigvee_{i\in I} \mathcal{A}_i$ 表示包含所有 $\mathcal{A}_i$ 的最小 σ- 代数。）我们用 Π. 表示策略集合

我们将 π_n 解释为智能体在第 n轮为停止规则选择的阈值。

对于给定的正可测强度函数 λ，我们将任意策略（π_n）$_{n\geq 0}$ 在第 $T \in \mathbb{N}$ 轮的遗憾定义为

$$\mathcal{R}^{\pi}_{\lambda}(T) := \sum_{n=0}^{T} E_{\lambda}[\Delta_{\lambda}(\pi_n)],$$

其中， Δ_{λ} 是定义的最优性间隙函数。

接下来，我们介绍一种特定算法，该算法对应的策略所产生的遗憾会随着轮次数趋于无穷而对数级增长。随后我们将证明，不存在遗憾增长速度慢于对数级的策略。从这个意义上说，该算法是渐近最优的。

算法 (无差异水平更新（ILU）)。

1. 初始化 $\mathcal{I} = \{0\}$。
2. 在第 0 轮选择阈值为 0 的停时。
3. 在任意一轮 $n \geq 1$ 执行以下操作:

   (a) 计算

$$\hat{\Gamma}(y) := \frac{\sum_{i\in\mathcal{I}}(N^i_0 - N^i_y)}{|\mathcal{I}|}, \quad y \in [S, 0], \tag{2.1}$$

   and

$$\hat{\varphi} := \frac{1}{|\mathcal{I}|} \sum_{i\in\mathcal{I}} \tau^i_0 \tag{2.2}$$

   ˆ (b) 确定 b，使得

$$\int_{\hat{b}}^{0} e^{\hat{\Gamma}(y)}\, dy = \hat{\varphi}. \tag{2.3}$$

   ˆ ˆ 如果不存在解 $b \in [S, 0]$，则令 $b = S$。

ˆ (c) 选择阈值为 b 的停时。(d) 如果停时取大于零的值，则 $\mathcal{I} \leftarrow \mathcal{I} \cup \{n\}$。

备注：注意， $\mathcal{I}$ 是算法在 0 之后停止的轮次集合。我们将 $\mathcal{I}$ 称为完全信息轮次的集合。量$N_0^j - N_y^j$表示第 j轮中区间 $[y, 0]$ 上的跳跃次数，其中 $y \in [S, 0]$。函数 $\Gamma(y)$是积分跳跃强度函数$\Lambda(y)$的估计量， $y \in [S, 0]$。此外，φ 是 $E_\lambda(\tau_0^j)$的估计量，即 0 之后的期望首次跳跃时间。

由于 Γ 和φ 仅使用 $\mathcal{I}$中各轮次的观测值，因此 ILU算法所选阈值序列是定义2.1意义下的策略。

3 主要结果

我们首先定义一个由光滑强度函数 λ构成的环境类。

定义 3.1 （环境类 $\mathcal{M}(L)$， $L \in (1, \infty)$）。当且仅当

1. λ连续可微；
2. λ 下界为 $\ln(2)/|S| + 1/L$；
3. λ 和 λ' 在 $[S, \infty)$ 上由 L 控制，即对于所有 $u \in [S, \infty)$， $\lambda(u) \leq L$ 且 $|\lambda'(u)| \leq L$。

注意，性质 2 保证了以下结论。

引理 3.2. 对于任意 $\lambda \in \mathcal{M}(L)$，存在唯一的 $b^* \in (S, 0)$ 满足(1.1)；此外，该 b^*满足 $b^* - S \geq |S| - \left(\frac{\ln(2)}{|S|} + \frac{1}{L}\right)^{-1} \ln(2) > 0$.

证明。 b^* 的最小可能值记为 $b^*_{\min}$，由恒等于最小值的 $\lambda \in \mathcal{M}(L)$ 取得。因此 $b^*_{\min}$ 满足

$$\int_{b^*_{min}}^{0} e^{-(\frac{\ln(2)}{|S|} + \frac{1}{L})u} du = \left(\frac{\ln(2)}{|S|} + \frac{1}{L}\right)^{-1}. \tag{3.1}$$

经过直接计算可得

$$b^*_{min} = -\left(\frac{\ln(2)}{|S|} + \frac{1}{L}\right)^{-1} \ln(2) > -\frac{|S|}{\ln(2)} \ln(2) = S.$$

□

下文中，我们用 $\mathcal{R}_\lambda^{ILU}(T)$ 表示在概率测度 P_λ下，ILU算法在第 $T \in \mathbb{N}$ 轮产生的遗憾。

定理 3.3 (ILU在类 $\mathcal{M}(L)$上的上界。 存在一个常数$C \in \mathbb{R}$，它仅取决于 S 和 L，使得对于所有 $T \in \mathbb{N}$，均有

$$\sup_{\lambda \in \mathcal{M}(L)} \mathcal{R}_\lambda^{ILU}(T) \leq C \ln(T+1).$$

证明见第 5 节。

定理 3.4 (类 $\mathcal{M}(L)$上的下界。 设 L 大于$\ln(2)/|S|+1/L$，因此 $\mathcal{M}(L)$ 非空。那么存在 $c \in (0, \infty)$，使得对于所有 $T \in \mathbb{N}$

$$\inf_{policy\ \pi} \sup_{\lambda \in \mathcal{M}(L)} \mathcal{R}_\lambda^{\pi}(T) \geq c \ln(T).$$

证明见第 6 节。
定理 3.3 和定理 3.4 表明，ILU算法具有渐近最优的增长率。

4 估计阈值的均方误差

我们采用第 2 节的设定。为简化记号，我们省略上标零，并记为 $\tau_b = \tau_b^0$。注意，在任意 P_λ 下，对于任意 $i \geq 1$， τ_b 与 τ_b^i 具有相同的分布。

以下令 $\lambda \in \mathcal{M}(L)$。为简化记号，我们在 P_λ、 E_λ、Var_λ 等记号中省略 λ。

引理 4.1。 我们有 $Var(\tau_0) < \infty$。

证明。 由于 $\lambda \in \mathcal{M}(L)$，我们有 $\lambda(u) \geq C$ 对所有 $u \in [S, \infty)$，其中 $C :=$

ln(2)/$|S| + \frac{1}{L}$。此外，

$$\begin{aligned} E(\tau_0^2) &= \int_0^\infty P(\tau_0^2 > t)dt = \int_0^\infty P(\tau_0 > \sqrt{t})dt \\ &= 2\int_0^\infty P(\tau_0 > u)udu \\ &= 2\int_0^\infty e^{-\int_0^u \lambda(y)dy} udu \\ &\le 2\int_0^\infty e^{-Cu} udu \\ &= 2\left([-\frac{1}{C}e^{-Cu}u]_0^\infty + \frac{1}{C}\int_0^\infty e^{-Cu} \right) \\ &= \frac{2}{C^2} < \infty. \end{aligned}$$

显然， $|E(\tau_0)| < \infty$。因此 $\mathrm{Var}(\tau_0) < \infty$。 □

我们首先在完全信息这一附加假设下分析 ILU算法。采用这一假设可以让我们更容易处理这些估计量。

定义 4.2 (完全信息策略)。 策略 $(\pi_n^{full})_{n\ge 0}$ 是一个随机变量序列，其取值属于 $[S, 0]$，且 π_n关于 $\mathcal{F}_n$可测，其中 $\mathcal{F}_0 = \{\emptyset, \Omega\}$ 且 $\mathcal{F}_n := \bigvee_{k=0}^{n-1} \sigma(N^{\lambda,\ t}{}^{k} : t \in [S, \tau_0^k])$ 对每个 $n \ge 1.$ 都成立。我们将策略集合记为 Π^{full}.

接下来，我们考虑估计量

$$\hat{\Lambda}_n(y) := \frac{\sum_{i=0}^{n-1}(N_0^i - N_y^i)}{n},$$

and

$$\hat{\tau}_{0,n} := \frac{1}{n}\sum_{i=0}^{n-1} \tau_0^i,$$

当 $n \ge 1$ 时。

^请注意，(Λ_n) 和 $(\hat{\tau}_{0,\ n})$ 正是 ILU算法中的估计量 $\hat{\Gamma}$ 和 $\hat{\varphi}$ 的对应项，前提是始终可以获得完全信息。

^ ^**定义 4.3.** 设 $b_0 := 0.$ 对于每个 $n \in \mathbb{N}$，令 b_n 为 $[S, 0]$ 中满足以下条件的实数

$$\int_{\hat{b}_n}^0 e^{\hat{\Lambda}_n(y)}\, dy = \hat{\tau}_{0,n}. \tag{4.1}$$

^ ^ 如果不存在满足方程 (4.1) 的 $b_n \in [S, 0]$ s，我们令 $b_n = S.$

注意，$(\hat{b}_n) \in \Pi^{full}$，但不一定是定义 2.1 意义下的策略。

引理 4.4. 对所有 $y \in [S, 0]$，我们有均方误差$(\hat{\Lambda}_n(y)) = \frac{1}{n}\Lambda(y)$，且$\hat{\Lambda}_n(y)$是一个无偏估计量。

证明. 显然。 □

引理 4.5. 可知均方误差$(\hat{\tau}_{0,n}) = \frac{1}{n}$ 方差(τ_0)，且$\hat{\tau}_{0,n}$ 是一个无偏估计量。

证明. 显然。 □

引理 4.6. 对所有 $n \in \mathbb{N}$，我们有

$$E(\sup_{x\in[S,0]}(\hat{\Lambda}_n(x) - \Lambda(x))^2) \le 4\Lambda(S)\frac{1}{n} \le 4L|S|\frac{1}{n} \tag{4.2}$$

证明. 首先，注意到 $M(y) := \Lambda_n(-y) - \Lambda(-y)$，$y \in [0, |S|]$，是关于自然滤过$(\mathcal{F}_y)_{y\in[0,|S|]}$的时间连续鞅，其中 $\mathcal{F}_y = \sigma(N_0^i - N_t^i: -y \le t \le 0, i = 0, \dots, n-1)$. 注意 $(M(y))$ 是具有左连续路径的平方可积鞅。因此，Doob 的 L^2–极大不等式适用，我们得到

$$\begin{aligned} E(\sup_{x\in[S,0]}(\hat{\Lambda}_n(x) - \Lambda(x))^2) &= E[\sup_{y\in[0,|S|]} M(y)^2] \\ &\le 4E[M(|S|)^2] = E((\hat{\Lambda}_n(S) - \Lambda(S))^2). \end{aligned}$$

由引理 4.4 可得

$$E(\sup_{x\in[S,0]}(\hat{\Lambda}_n(x) - \Lambda(x))^2) \le 4\Lambda(S)\frac{1}{n},$$

并利用简单估计 $\Lambda(S) \le L|S|$，我们得到 (4.2)。 □

命题 4.7。 设 $M := \min\{m \in \mathbb{N}_0 : S + m\frac{1}{2L} \ge 0\}$。那么，对于所有$n \in \mathbb{N}$, 我们有

$$\begin{aligned} E\left(\left(\hat{b}_n - b^*\right)^2\right) \le &\left(S^2\frac{4\Lambda(S) + \mathrm{Var}(\tau_0)}{\varepsilon^2}\right)\frac{1}{n} + \\ &+ \left(S^2 4\Lambda(S)(M+1) + 2\mathrm{Var}(\tau_0) + 8|S|\int_S^0 e^{2\Lambda(y)}\Lambda(y)\,dy\right)\frac{1}{n}, \end{aligned} \tag{4.3}$$

其中 $0 < \varepsilon <$最小值$\{E(\tau_0)\, 1, \frac{b^*-S}{3+2E(\tau_0)}\}$.

证明。 固定 $n \in \mathbb{N}$，先假设我们处于事件 $B_n := \{|S| > \hat{}|b_n|\}$ 上。

注意，$\hat{b}_n$ 的确定方式使得

$$\int_{\hat{b}_n}^{0} e^{\hat{\Lambda}_n(y)}\, dy = \hat{\tau}_{0,n}.$$

此外，最优的 b^* 与未知的真实强度函数 λ满足

$$\int_{b^*}^{0} e^{\Lambda(y)}\, dy = E(\tau_0).$$

因此

$$\int_{b^*}^{0} e^{\Lambda(y)}\, dy - \int_{\hat{b}_n}^{0} e^{\hat{\Lambda}_n(y)}\, dy = E(\tau_0) - \hat{\tau}_{0,n}.$$

这等价于

$$\int_{b^*}^{0} e^{\Lambda(y)}\, dy - \int_{\hat{b}_n}^{0} e^{\Lambda(y)}\, dy + \int_{\hat{b}_n}^{0} \left(e^{\Lambda(y)} - e^{\hat{\Lambda}_n(y)}\right) dy = E(\tau_0) - \hat{\tau}_{0,n}. \tag{4.4}$$

1st 情形：令^ $b_n \geq b^*$。由（4.4）可得：

$$\begin{aligned}\int_{b^*}^{\hat{b}_n} e^{\Lambda(y)}\, dy &= E(\tau_0) - \hat{\tau}_{0,n} - \int_{\hat{b}_n}^{0} \left(e^{\Lambda(y)} - e^{\hat{\Lambda}_n(y)}\right) dy \\ &= E(\tau_0) - \hat{\tau}_{0,n} + \int_{\hat{b}_n}^{0} e^{\Lambda(y)} \left(e^{\hat{\Lambda}_n(y)-\Lambda(y)} - 1\right) dy. \end{aligned} \tag{4.5}$$

接下来，我们从下方估计方程 (4.5) 的左侧，并从上方估计其右侧。为此，我们定义

$$\begin{aligned} A &:= \int_{b^*}^{\hat{b}_n} e^{\Lambda(y)}\, dy, \\ B &:= E(\tau_0) - \hat{\tau}_{0,n} + \int_{\hat{b}_n}^{0} e^{\Lambda(y)} \left(e^{\hat{\Lambda}_n(y)-\Lambda(y)} - 1\right) dy. \end{aligned}$$

注意，

$$A \geq (\hat{b}_n - b^*) \min_{y\in[S,0]} e^{\Lambda(y)} \geq \hat{b}_n - b^*;$$

$$B \leq |E(\tau_0) - \hat{\tau}_{0,n}| + \int_{S}^{0} e^{\Lambda(y)} \left|e^{\hat{\Lambda}_n(y)-\Lambda(y)} - 1\right| dy.$$

请注意，对于所有的 $x \in [-1,1]$，我们都有

$$|e^x - 1| \le 2|x|. \tag{4.6}$$

定义 $C_n := \left\{ \left| \hat{\Lambda}_n(y) - \Lambda(y) \right| \le 1, \quad \forall y \in [S,0] \right\}$.

那么在 C_n 上，我们有

$$\left| e^{\hat{\Lambda}_n(y)-\Lambda(y)} - 1 \right| \le 2 \left| \hat{\Lambda}_n(y) - \Lambda(y) \right|.$$

特别地，利用 A 和 B 的估计，我们得到

$$\hat{b}_n - b^* \le |E(\tau_0) - \hat{\tau}_{0,n}| + 2 \int_S^0 e^{\Lambda(y)} \left| \hat{\Lambda}_n(y) - \Lambda(y) \right| dy.$$

2nd 情形：设 $\hat{} \; b_n < b^*$。则 (4.4) 意味着

$$-\int_{\hat{b}_n}^{b^*} e^{\Lambda(y)} \, dy = E(\tau_0) - \hat{\tau}_{0,n} + \int_{\hat{b}_n}^0 e^{\Lambda(y)} \left(e^{\hat{\Lambda}_n(y)-\Lambda(y)} - 1 \right) dy,$$

这等价于

$$\int_{\hat{b}_n}^{b^*} e^{\Lambda(y)} \, dy = (\hat{\tau}_{0,n} - E(\tau_0)) + \int_{\hat{b}_n}^0 e^{\Lambda(y)} \left(- \left(e^{\hat{\Lambda}_n(y)-\Lambda(y)} - 1 \right) \right) dy. \tag{4.7}$$

下面我们从下方估计方程 (4.7) 的左侧，并从上方估计其右侧。为此，我们定义

$$A := \int_{\hat{b}_n}^{b^*} e^{\Lambda(y)} \, dy,$$

$$B := (\hat{\tau}_{0,n} - E(\tau_0)) + \int_{\hat{b}_n}^0 e^{\Lambda(y)} \left(- \left(e^{\hat{\Lambda}_n(y)-\Lambda(y)} - 1 \right) \right) dy.$$

Then

$$A \ge (b^* - \hat{b}_n) \min_{y \in [S,0]} e^{\Lambda(y)} \ge b^* - \hat{b}_n;$$

$$B \le |E(\tau_0) - \hat{\tau}_{0,n}| + \int_S^0 e^{\Lambda(y)} \left| e^{\hat{\Lambda}_n(y)-\Lambda(y)} - 1 \right| dy.$$

令 C_n 与第一种情况相同。那么在 C_n 上有

$$\left| e^{\hat{\Lambda}_n(y)-\Lambda(y)} - 1 \right| \le 2 \left| \hat{\Lambda}_n(y) - \Lambda(y) \right|.$$

因此，

$$b^* - \hat{b}_n \leq |E(\tau_0) - \hat{\tau}_{0,n}| + 2\int_S^0 e^{\Lambda(y)} \left|\hat{\Lambda}_n(y) - \Lambda(y)\right| \, dy.$$

总之，在任何情况下，如果出现 B_n 和 C_n，我们有

$$\left|\hat{b}_n - b^*\right| \leq |E(\tau_0) - \hat{\tau}_{0,n}| + 2\int_S^0 e^{\Lambda(y)} \left|\hat{\Lambda}_n(y) - \Lambda(y)\right| \, dy. \tag{4.8}$$

因此，利用估计式 $(e+f)^2 \leq 2(e^2+f^2)$ $e,\ \ f \in \mathbb{R}$

$$\begin{aligned} E\left(\left|\hat{b}_n - b^*\right|^2 \mathbf{1}_{B_n}\mathbf{1}_{C_n}\right) \leq & 2\Big\{E\left[(E(\tau_0) - \hat{\tau}_{0,n})^2 \mathbf{1}_{B_n}\mathbf{1}_{C_n}\right] + \\ & + 4E\left[\mathbf{1}_{B_n}\mathbf{1}_{C_n}\left(\int_S^0 e^{\Lambda(y)}\left|\hat{\Lambda}_n(y) - \Lambda(y)\right| \, dy\right)^2\right]\Big\}. \\ \leq & 2\Big\{E\left[(E(\tau_0) - \hat{\tau}_{0,n})^2\right] + \\ & + 4E\left[\left(\int_S^0 e^{\Lambda(y)}\left|\hat{\Lambda}_n(y) - \Lambda(y)\right| \, dy\right)^2\right]\Big\}. \end{aligned} \tag{4.9}$$

利用积分形式的詹森不等式，我们进一步得到

$$\begin{aligned} & E\left(\left|\hat{b}_n - b^*\right|^2 \mathbf{1}_{B_n}\mathbf{1}_{C_n}\right) \\ & \leq 2\Big\{E\left[(E(\tau_0) - \hat{\tau}_{0,n})^2\right] + 4E\left[|S|\left(\int_S^0 \left(e^{\Lambda(y)}\left|\hat{\Lambda}_n(y) - \Lambda(y)\right|\right)^2 dy\right)\right]\Big\} \\ & = 2\Big\{\mathrm{MSE}(\hat{\tau}_{0,n}) + 4|S|\int_S^0 e^{2\Lambda(y)}\mathrm{MSE}(\hat{\Lambda}_n(y))\, dy\Big\}. \end{aligned}$$

利用引理 4.4 和 4.5，我们进一步得到

$$E\left(\left|\hat{b}_n - b^*\right|^2 \mathbf{1}_{B_n}\mathbf{1}_{C_n}\right) \leq \frac{1}{n}\left(2\mathrm{Var}(\tau_0) + 8|S|\int_S^0 e^{2\Lambda(y)}\Lambda(y)\, dy)\right). \tag{4.10}$$

总之，我们有

$$\begin{aligned} & E((\hat{b}_n - b^*)^2) \\ & = E((\hat{b}_n - b^*)^2\mathbf{1}_{B_n^c}) + E((\hat{b}_n - b^*)^2\mathbf{1}_{B_n}\mathbf{1}_{C_n^c}) + E((\hat{b}_n - b^*)^2\mathbf{1}_{B_n}\mathbf{1}_{C_n}) \\ & \leq S^2P(B_n^c) + S^2P(C_n^c) + \frac{1}{n}\left(2\mathrm{Var}(\tau_0) + 8|S|\int_S^0 e^{2\Lambda(y)}\Lambda(y)\, dy)\right). \end{aligned} \tag{4.11}$$

根据引理 4.5、引理 4.6 和下文的引理 4.8，我们有

$$\begin{aligned} P(B_n^c) &\leq \frac{E(||\hat{\Lambda}_n - \Lambda||^2) + \mathrm{Var}(\hat{\tau}_{0,n})}{\varepsilon^2} \\ &\leq \frac{4\Lambda(S) + \mathrm{Var}(\tau_0)}{\varepsilon^2} \frac{1}{n} \end{aligned} \tag{4.12}$$

其中 $0 < \varepsilon < \min\{E(\tau_0)\, 1,\, \frac{b^* - S}{3 + 2E(\tau_0)}\}$.

此外，利用下文引理 4.10 中的估计 (4.14)，我们有

$$P(C_n^c) \leq (4\Lambda(S)(M+1))\, \frac{1}{n}. \tag{4.13}$$

结合估计 (4.12)、(4.13) 和 (4.11)，可得 (4.7)。 □

引理 4.8. 设 $B_n := \{|S| > |b_n|\}, n \in \mathbb{N}$. 则有

$$P(B_n^c) \leq \frac{E(||\hat{\Lambda}_n - \Lambda||_\infty^2) + Var(\hat{\tau}_{0,n})}{\varepsilon^2},$$

其中 $0 < \varepsilon < \min\{E(\tau_0)\, 1,\, \frac{b^* - S}{3 + 2E(\tau_0)}\}$.

证明. 我们希望确定 $\varepsilon := \varepsilon(\lambda) > 0$ 使得 $||\Lambda_n - \Lambda||_\infty < \varepsilon$ 且 $|E(\tau_0) - \tau_{0,n}| < \varepsilon$ 蕴含 $|b_n| < |S|$。设 $\delta := b^* - S$。

我们通过反证法证明，$\varepsilon < \min\{1, \frac{\delta}{3 + 2E(\tau_0)}\}$ 意味着 $|b_n| < |S|$。

假设 $|b_n| = |S|$。我们有

$$\begin{aligned} \int_{\hat{b}_n}^0 e^{\hat{\Lambda}_n(u)} du &\geq \int_{\hat{b}_n}^0 e^{\Lambda(u) - \varepsilon} du = \int_S^0 e^{\Lambda(u) - \varepsilon} du = e^{-\varepsilon} \int_{b^* - \delta}^0 e^{\Lambda(u)} du \\ &= e^{-\varepsilon} \left(\int_{b^*}^0 e^{\Lambda(u)} du + \int_{b^* - \delta}^{b^*} e^{\Lambda(u)} du \right) = e^{-\varepsilon} \left(E(\tau_0) + \int_{b^* - \delta}^{b^*} e^{\Lambda(u)} du \right) \\ &\geq e^{-\varepsilon} \left(E(\tau_0) + \delta \min_{u \in [S, 0]} e^{\Lambda(u)} \right) = e^{-\varepsilon} \left(E(\tau_0) + \delta \right). \end{aligned}$$

此外

$$E(\tau_0) + \varepsilon \geq \hat{\tau}_{0,n} \geq \int_{\hat{b}_n}^0 e^{\hat{\Lambda}_n(u)} du \geq e^{-\varepsilon} \left(E(\tau_0) + \delta \right).$$

因此，若 $e^{\varepsilon}(E(\tau_0)+\varepsilon) < (E(\tau_0)+\delta)$，就会产生矛盾。注意，对于 $0 \le \varepsilon \le 1$，有 $e^{\varepsilon} \le 1+\varepsilon+\varepsilon^2$。由此可得，对于 $\varepsilon \in [0,1]$

$$\begin{aligned} e^{\varepsilon}(E(\tau_0)+\varepsilon) &\le (1+\varepsilon+\varepsilon^2)(E(\tau_0)+\varepsilon) \le (1+2\varepsilon)(E(\tau_0)+\varepsilon) \\ &\le E(\tau_0)+\varepsilon+2\varepsilon E(\tau_0)+2\varepsilon^2 \\ &\le E(\tau_0)+(3+2E(\tau_0))\varepsilon. \end{aligned}$$

当 $\varepsilon < \min\{1, \frac{\delta}{3+2E(\tau_0)}\}$ 时，我们有 $e^{\varepsilon}(E(\tau_0)+\varepsilon) \le E(\tau_0)+(3+2E(\tau_0))\varepsilon < E(\tau_0)+\delta$，因此矛盾。

注意到 $E(\tau_0)-\varepsilon \ge 0$，因此最终我们有 $\varepsilon < \min\{E(\tau_0), 1, \frac{\delta}{3+2E(\tau_0)}\}$。最后，

$$P(B_n^c) \le P(||\hat{\Lambda}_n - \Lambda|| > \varepsilon) + P(|E(\tau_0) - \hat{\tau}_{0,n}| > \varepsilon) \le \frac{E(||\hat{\Lambda}_n - \Lambda||_\infty^2) + \mathrm{Var}(\hat{\tau}_{0,n})}{\varepsilon^2}$$

□

在下文中，令 $\tilde{C}_n(y,r) := \left\{ \left| \hat{\Lambda}_n(y) - \Lambda(y) \right| \le r \right\}, r \in \mathbb{R}, n \in \mathbb{N}$。特别地，

$$C_n = \bigcap_{y \in [S,0]} \tilde{C}_n(y,1),$$

其中事件 C_n 的定义与命题 4.7 的证明中相同。

引理 4.9。令 $y \in [S,0]$ 并定义 $k := \frac{1}{2L}$。那么对所有 $t \in [y, y+k \wedge 0]$ 和 $\omega \in \tilde{C}_n(y,1) \cap \tilde{C}_n(y+k,1)$：

$$\left| \hat{\Lambda}_n(t) - \Lambda(t) \right| \le 1.$$

证明。不失一般性，设 $y+k<0$。若 $y+k \ge 0$，则在以下讨论中将 $y+k$ 替换为 0，所得估计仍然成立。

设 $\omega \in \tilde{C}_n(y,1) \cap \tilde{C}_n(y+k,1)$。由此可得 $\left| \hat{\Lambda}_n(y,\omega) - \Lambda(y) \right| \le \frac{1}{2}$ 且 $\left| \hat{\Lambda}_n(y+k,\omega) - \Lambda(y+k) \right| \le \frac{1}{2}$。此外，由定义直接可得 $\Lambda_n(y_1,\omega) \ge \Lambda_n(y_2,\omega)$，若 $y_1 \le y_2$ 且 $y_1, y_2 \in [S,0]$。因此，对任意 $t \in [y, y+k]$，下列结论成立：

$$\begin{aligned} \hat{\Lambda}_n(t,\omega) - \Lambda(t) &\le \hat{\Lambda}_n(y,\omega) + \Lambda(y) - \Lambda(y) - \Lambda(t) \\ &\le \frac{1}{2} + L(t-y) \\ &\le \frac{1}{2} + Lk = 1. \end{aligned}$$

此外

$$\begin{aligned}\Lambda(t) - \hat{\Lambda}_n(t,\omega) &\le \Lambda(t) - \Lambda(y+k) + \Lambda(y+k) - \hat{\Lambda}_n(y+k,\omega)\\ &\le Lk + \frac{1}{2} = 1.\end{aligned}$$

□

引理 4.10。设 $M := \min\{m \in \mathbb{N}_0 : S + m\frac{1}{2L} \ge 0\}$。则

$$P(C_n^c) \le 4\Lambda(S)(M+1)\frac{1}{n}, \tag{4.14}$$

其中 $C_n = \left\{\left|\hat{\Lambda}_n(y) - \Lambda(y)\right| \le 1, \forall y \in [S,0]\right\}$.

证明。我们有 $M = \min\{m \in \mathbb{N}_0 : S + m\frac{1}{2L} \ge 0\}$。定义 $y_m := S + m\frac{1}{2L}$，满足 $m \in \{0,1,\dots,M-1\}$。此外，设 $y_M := 0$。根据引理 4.9，对所有 $t \in [S,0]$ 和 $\omega \in \tilde{C}_n(y_0, \frac{1}{2}) \cap \tilde{C}_n(y_1, \frac{1}{2}) \cap \cdots \cap \tilde{C}_n(y_M, \frac{1}{2})$，均有：

$$\left|\hat{\Lambda}_n(t) - \Lambda(t)\right| \le 1.$$

特别是，以下内容适用：

$$\begin{aligned}\tilde{C}_n\Big(y_0, \frac{1}{2}\Big) \cap \cdots \cap \tilde{C}_n\Big(y_M, \frac{1}{2}\Big) &\subseteq \left\{\left|\hat{\Lambda}_n(y) - \Lambda(y)\right| \le 1, \forall y \in [S,0]\right\}\\ &= C_n.\end{aligned}$$

因此，利用 $C_n^c \subseteq \left(\tilde{C}_n(y_0, \frac{1}{2}) \cap \tilde{C}_n(y_1, \frac{1}{2}) \cap \cdots \cap \tilde{C}_n(y_M, \frac{1}{2})\right)^c$ 以及概率测度的次可加性，我们得到

$$\begin{aligned}P(C_n^c) &\le P\left(\left(\tilde{C}_n\Big(y_0, \frac{1}{2}\Big) \cap \tilde{C}_n\Big(y_1, \frac{1}{2}\Big) \cap \cdots \cap \tilde{C}_n\Big(y_M, \frac{1}{2}\Big)\right)^c\right)\\ &\le \sum_{i=0}^{M} P\left(\tilde{C}_n\Big(y_i, \frac{1}{2}\Big)^c\right)\\ &= \sum_{i=0}^{M} P\left(\left\{\left|\hat{\Lambda}_n(y_i) - \Lambda(y_i)\right| > \frac{1}{2}\right\}\right)\\ &\le \sum_{i=0}^{M} \frac{1}{(1/2)^2}\mathrm{Var}(\hat{\Lambda}_n(y_i))\\ &= \sum_{i=0}^{M} 4\Lambda(y_i)\frac{1}{n}\\ &\le 4\Lambda(S)(M+1)\frac{1}{n}.\end{aligned}$$

□

5 定理 3.3 的证明

设 (π_n) 是 ILU算法选取的阈值序列。回顾一下，(π_n) 是定义2.1意义下的策略。我们需要证明，存在一个常数 C，使得对所有 $\lambda \in \mathcal{M}(L)$ 和 $T \in \mathbb{N}$，都有$\mathcal{R}_\lambda^\pi(T) \leq C \ln(T+1)$。

我们利用序列 (b_n) 所蕴含的最优性差距，估计期望最优性间隙 $E_\lambda[\Delta_\lambda(\pi_n)]$。为此，我们对 $n \geq 0$ 定义

$$r_\lambda(n) := E_\lambda[\Delta_\lambda(\hat{b}_n)].$$

接下来，我们递归地定义序列 $\sigma(j)$、$j \geq 0$，如下所示。令 $\sigma(1) = 0$，且对所有 $j \geq 1$

$$\sigma(j+1) = \min\{n > \sigma(j) : \tau_{\pi_n}^n > 0\}.$$

注意，$\sigma(j)$ 表示第 j 次获得完全信息的轮次。此外，$E_\lambda(\sigma(j+1) - \sigma(j))$ 恰好是等待一个获得完全信息的轮次所需的期望等待时间。该期望值的上界为 $E_\lambda(\sigma(j+1) - \sigma(j)) \leq e^{L|S|}$。最后，令 $\mathcal{I}_n$ 表示搜索的前 n 轮中获得完全信息的轮次所构成的随机集合。我们有 $\mathcal{I}_0 = \emptyset$，且对 $n \geq 1$，$\mathcal{I}_n = \{\sigma(j) : \sigma(j) \leq n-1\}$。此外，注意到$|\mathcal{I}_n| = \max\{k : \sigma(k) \leq n-1\}$，且 $\sigma(|\mathcal{I}_n|) < n \leq \sigma(|\mathcal{I}_n|+1)$。

关键的观察是，现在对于所有 $n \geq 0$，我们都有

$$E_\lambda[\Delta_\lambda(\pi_n)] = E_\lambda[r_\lambda(|\mathcal{I}_n|)]. \tag{5.1}$$

注意，$n = 0$ 的方程是平凡的，因为 $E_\lambda[\Delta_\lambda(\pi_0)] = E_\lambda[\Delta_\lambda(0)] = \hat{} E_\lambda[\Delta_\lambda(b_0)] = r_\lambda(0)$。为了证明 $n \geq 1$的方程，我们可以假设 $\hat{}\hat{}$ (b_n) 与 (π_n) 和 $(\mathcal{I}_n)$ 相互独立（在概率空间的一个独立副本上定义 (b_n)）。那么，在事件 $\{|\mathcal{I}_n| = i\}$ 上，阈值 π_n 与 b_i 具有相同的分布。因此，$E_\lambda[\Delta_\lambda(\pi_n)1_{\{|\mathcal{I}_n|=i\}}] = E_\lambda[\Delta_\lambda(b_i)1_{\{|\mathcal{I}_n|=i\}}] = \hat{} E_\lambda[\Delta_\lambda(b_i)]P_\lambda(|\mathcal{I}_n| = i) = r_\lambda(i)P_\lambda(|\mathcal{I}_n| = i)$。由此可得，

$$E_\lambda[\Delta_\lambda(\pi_n)] = \sum_{i=0}^{n-1} r_\lambda(i)P_\lambda(|\mathcal{I}_n| = i) = E_\lambda[r_\lambda(|\mathcal{I}_n|)].$$

由 (5.1) 得

$$\mathcal{R}_\lambda^\pi(T) = \sum_{n=0}^{T} E_\lambda\left[r_\lambda(|\mathcal{I}_n|)\right].$$

现在我们利用 λ 属于环境类 $\mathcal{M}(L)$。这意味着λ 连续可微，因此 Δ_λ 二阶连续可微（见引理 1.2）。因此，利用 $\Delta_\lambda(b^*) = \Delta'_\lambda(b^*) = 0$，根据泰勒定理可得

$$\Delta_\lambda(b) = \frac{1}{2}\Delta''_\lambda(\theta)(b - b^*)^2 \tag{5.2}$$

对于某个 θ，介于 b 和 b^* 之间。

此外，我们定义 $c := \sup_{\lambda\in\mathcal{M}(L)} \sup_{b\in[S,0]} \Delta''_\lambda(b)$。由于 $\lambda \in \mathcal{M}(L)$，根据方程 (1.5) 可以推出，$c$ 仅取决于 S 和 L，且$c < \infty$。因此，对于所有 λ 和 $b \in [S, 0]$，我们有 $\Delta_\lambda(b) \leq \frac{1}{2}c(b - b^*)^2$。于是，

$$\sup_{\lambda\in\mathcal{M}(L)} r_\lambda(n) \leq \frac{1}{2}cE_\lambda[(\hat{b}_n - b^*)^2].$$

为方便记号，令 $L_{low} := \ln(\frac{2}{|S|} + \frac{1}{L})$。注意，$L_{low}$ 恰好是环境类 $\mathcal{M}(L)$ 的性质 2 所给出的下界，并且仅取决于环境类参数 S 和 L。以下估计对所有 $\lambda \in \mathcal{M}(L)$ 均成立

1. $\Lambda(S) \leq |S|L$;
2. $\min\{m \in \mathbb{N}_0 : S + m\frac{1}{2L} \geq 0\} \leq \lceil 2|S|L\rceil \leq 2|S|L + 1$;
3. $E(\tau_0) \geq \frac{1}{L}$;
4. $\mathrm{Var}(\tau_0) \leq 2L_{low}^2$.

最后一个估计由引理 4.1 的证明得出。此外，利用引理 3.2，我们有

$$\frac{b^* - S}{3 + 2E(\tau_0)} \geq \frac{|S| - L_{low}^{-1}\ln(2)}{3 + 2L_{low}^{-1}}.$$

因此，命题 4.7 蕴含 $\sup_{\lambda\in\mathcal{M}(L)} r_\lambda(n) \leq \frac{1}{2}cD\frac{1}{n}$，其中

$$D = \left[\left(S^2\frac{4L|S| + 2L_{low}^2}{(\min\{\frac{1}{L}, \frac{|S|-L_{low}^{-1}\ln(2)}{3+2L_{low}^{-1}}\})^2}\right) + \left(4L_{low}^2 + 8|S|^3Le^{2|S|L} + |S|^3 4L(2|S|L + 2)\right)\right]\frac{1}{n}.$$

因此，对于所有 $\lambda \in \mathcal{M}(L)$，我们有

$$\begin{aligned}
\mathcal{R}_\lambda^\pi(T) = E_\lambda\left[\sum_{n=0}^{T} r_\lambda(|\mathcal{I}_n|)\right] &\le E_\lambda(\tau_0^0) + E_\lambda\left[\sum_{n=1}^{T} r_\lambda(|\mathcal{I}_n|)\right] \\
&\le L_{low}^{-1} + E_\lambda\left[\sum_{j=1}^{|\mathcal{I}_T|}(\sigma(j+1)-\sigma(j))r_\lambda(j)\right] \\
&\le L_{low}^{-1} + \sum_{j=1}^{T} E_\lambda[\sigma(j+1)-\sigma(j)]r_\lambda(j) \\
&\le L_{low}^{-1} + e^{L|S|}\sum_{j=1}^{T} r_\lambda(j) \\
&\le L_{low}^{-1} + e^{L|S|}\frac{1}{2}c\,D\ln(T+1) \\
&\le L_{low}^{-1}\frac{\ln(T+1)}{\ln(2)} + e^{L|S|}\frac{1}{2}c\,D\ln(T+1) \\
&= \left(\frac{L_{low}^{-1}}{\ln(2)} + e^{L|S|}\frac{1}{2}c\,D\right)\ln(T+1).
\end{aligned}$$

请注意，前一个不等式的右端不依赖于 λ。因此，我们已经证明了该定理。

6 下界：定理 3.4 的证明

下面我们证明定理 3.4。我们需要证明极小极大遗憾

$$\inf_{\pi\in\Pi}\sup_{\lambda\in\mathcal{M}(L)}\mathcal{R}_\lambda^\pi(T)$$

至少按对数级增长。论证的关键在于，只需为 $\mathcal{H}(L) := \{\lambda \in \mathcal{M}(L) : \lambda$ 常数$\}$ 推导出一个对数下界；该子类由常强度函数构成。事实上，请注意

$$\inf_{\pi\in\Pi}\sup_{\lambda\in\mathcal{M}(L)}\mathcal{R}_\lambda^\pi(T) \ge \inf_{\pi\in\Pi}\sup_{\lambda\in\mathcal{H}(L)}\mathcal{R}_\lambda^\pi(T).$$

因此， $\mathcal{H}(L)$ 的对数下界意味着整个环境类 $\mathcal{M}(L)$ 也有一个下界。

请注意，如果未知强度函数属于类 $\mathcal{H}(L)$，那么停止代理在每一轮中观察到的都是齐次泊松过程。

因此，强度函数的估计实际上只是一个单参数估计，即对 $\lambda \in [a,b]$ 的估计，其中 $a := \ln(\frac{2}{|S|} + \frac{1}{L})$ 且 $b := L$。

此外，我们假设每一轮都拥有完全信息，即每一轮都可以观测 $[S,0]$ 上的齐次泊松过程。由于完全信息策略类大于部分信息策略类，因此我们有

$$\inf_{\pi\in\Pi}\sup_{\lambda\in\mathcal{H}(L)}\mathcal{R}_\lambda^\pi(T) \geq \inf_{\pi\in\Pi^{full}}\sup_{\lambda\in\mathcal{H}(L)}\mathcal{R}_\lambda^\pi(T). \tag{6.1}$$

此外，为了简化下面的分析，我们引入所谓的截断完全信息策略类

$$\Pi^{cut} := \left\{\pi^{cut} = (\pi_n^{cut})_{n\geq 1} | \exists \pi^{full} \in \Pi^{full} \text{ s.t. } \forall n \geq 0 : \pi_n^{cut} = \min\{\pi_n^{full}, -\frac{\ln(2)}{L}\}\right\}.$$

其思路是，对于 $\lambda \in \mathcal{H}(L)$，我们知道 $b^* \in (S, -\frac{\ln(2)}{L}]$，由推论 1.3 可知）。因此，将每个策略在 $-\frac{\ln(2)}{L}$ 处截断不会增大遗憾。令 $\pi \in \Pi^{full}$ 和 π^{cut} 分别为相应的截断策略。固定 $n \in \mathbb{N}_0$.

情况 1： 我们有 $\pi_n < -\frac{\ln(2)}{L}$。因此

$$\Delta(\pi_n) = E|\tau_{\pi_n}| - E|\tau_{b^*}| = E|\tau_{\pi_n^{cut}}| - E|\tau_{b^*}| = \Delta(\pi_n^{cut})$$

情况 2： 我们有 $\pi_n \in [L^*, 0]$，其中 $L^* := -\frac{\ln(2)}{L}$。根据定理 1.3，我们知道 $\lambda \in \mathcal{H}(L)$ 的成本函数为 $\Delta_\lambda(b) = \frac{1}{\lambda}(2e^{\lambda b} - 1) - b - \ln(2)\lambda$，对于 $b \in [S,0].$。我们想要证明，对于所有 $x \in [L^*,0].$，$\Delta_\lambda(x) \geq \Delta_\lambda(L^*)$。我们有 $\Delta'_\lambda(x) = 2e^{\lambda x} - 1.$。回想一下，$\lambda \in [\ln(2)/|S| + \frac{1}{L}, L]$，且 $x \in [L^*,0].$。由此直接得到 $\lambda x \geq -\ln(2)$，因此对于所有 $x \in [L^*,0].$，$\Delta'_\lambda(x) \geq 0$。由此可得

$$\Delta_\lambda(\pi_n) \geq \Delta_\lambda(\pi_n^{cut}).$$

总之，将策略替换为截断策略不会使遗憾率变差，因此我们得到

$$\inf_{\pi\in\Pi}\sup_{\lambda\in\mathcal{H}(L)}\mathcal{R}_\lambda^\pi(T) \geq \inf_{\pi\in\Pi^{full}}\sup_{\lambda\in\mathcal{H}(L)}\mathcal{R}_\lambda^\pi(T) = \inf_{\pi\in\Pi^{cut}}\sup_{\lambda\in\mathcal{H}(L)}\mathcal{R}_\lambda^\pi(T).$$

因此，针对截断策略的下界同样适用于定义 2.1 中的策略。

在估计常数跳跃强度时，观察 $(N_t^j)_{t\in[S,0]}$，就充分性而言，等价于观察 $N_0^j \sim Poi(\lambda|S|)$。因此，在下文中我们可以假设，参数 λ 的任何估计量仅使用值 N_0^0、N_0^1、N_0^2、.....。注意，后一序列是 $i.i.d.$

并且服从参数为 $\lambda|S|$ 的泊松分布。不失一般性，我们取 $|S| = 1$。~否则，只需在以下考虑中将 λ 替换为 $\lambda := |S|\lambda \in [|S|a, |S|b]$

根据引理 1.3，对于任意常数 λ ，我们有 $E_\lambda(|\tau_{b^*}|) = \ln(2)\ \lambda$ 。最优阈值b^* 与给定的常数强度 λ之间存在一一对应关系。事实上，我们有 $b^*(\lambda) = -\frac{\ln(2)}{\lambda}$ 以及$\lambda(b^*) = -\frac{\ln(2)}{b^*}$. 。我们利用这一对应关系，将寻找使均方误差最小的截断策略问题，转化为确定一个使均方误差达到最小的 λ估计量的问题。

命题 6.1。 存在常数 $c \in (0, \infty)$，使得对于所有 $x \in [S, 0]$ 和 $\lambda \in \mathcal{H}(L)$，我们都有

$$\Delta_\lambda(x) \geq c(x - b^*(\lambda))^2.$$

证明。 我们在最优阈值 $b^*(\lambda)$ 附近对 $\Delta_\lambda(x)$ 作泰勒展开。注意到所有必要的光滑性条件均已满足，因此可得

$$\Delta_\lambda(x) = \Delta_\lambda(b^*(\lambda)) + \Delta'_\lambda(b^*(\lambda))(x - b^*(\lambda)) + \Delta''_\lambda(\nu)(x - b^*(\lambda))^,$$

其中 $\nu \in [\min\{x, b^*(\lambda)\},\ \max\{x, b^*(\lambda)\}]$. 。根据定义和一阶最优性条件，我们知道 $\Delta_\lambda(b^*(\lambda)) = \Delta'_\lambda(b^*(\lambda)) = 0.$ 。此外，对于 $\lambda \in \mathcal{H}(L)$，我们知道 $\Delta_\lambda(x) = \frac{1}{\lambda}(2e^{\lambda x} - 1) - x - \ln(2)\ \lambda$ ˙。由此可得，对于所有 x， $\Delta''_\lambda(x) = 2\lambda e^{\lambda x} \geqslant 0$。因此，我们有

$$\Delta_\lambda(x) \geq c(x - b^*(\lambda))^2,$$

其中 $c := \inf^{\lambda \in [a,b]} \inf_{x \in [S,0]} \Delta''_\lambda(x) > 0.$ □

命题 6.1 意味着，对于所有 $(\pi_n) \in \Pi^{cut}$ ，我们有

$$E_\lambda(\Delta_\lambda(\pi_n)) \geq cE_\lambda[(\pi_n - b^*(\lambda))^2] = cE_\lambda\left[(b^*(-\frac{\ln(2)}{\pi_n}) - b^*(\lambda))^2\right]$$
$$\geq \tilde{c}E_\lambda\left[(-\frac{\ln(2)}{\pi_n} - \lambda)^2\right], \tag{6.2}$$

其中 $\tilde{c} := c \min^{\lambda \in [a,b]} (b^*)'(\lambda) = c\frac{\ln(2)}{L^2}$。

我们可以将 $-\frac{\ln(2)}{\pi_n}$ 解释为由策略 $(\pi_n) \in \Pi^{cut}$ 定义的参数 λ 的相应估计量。由 (6.2) 可直接得出

$$\mathcal{R}^\pi_\lambda(T) = \sum_{n=1}^{T} E_\lambda(\Delta_\lambda(\pi_n)) \geq \tilde{c}\sum_{n=1}^{T} E_\lambda\left[(\frac{-\ln(2)}{\pi_n} - \lambda)^2\right]$$

因此

$$\inf_{\text{policies }\pi}\sup_{\lambda\in\mathcal{H}(L)}\mathcal{R}_\lambda^\pi(T)\geq\inf_{\pi\in\Pi^{cut}}\sup_{\lambda\in\mathcal{H}(L)}\mathcal{R}_\lambda^\pi(T)\geq\inf_{\text{estimator }\hat{\lambda}}\sup_{\lambda\in[a,b]}\tilde{c}\sum_{n=1}^{T}E_\lambda[(\hat{\lambda}_n-\lambda)^2].$$

总而言之，我们已将求出环境类 $\mathcal{M}(L)$上极小极大遗憾下界的问题，转化为利用观测值 N_0^0， N_0^1， N_0^2，.....估计 $\lambda\in\mathcal{H}(L)$时求出极小极大风险下界的问题。

接下来注意到，极小极大风险可以用先验密度 q对应的贝叶斯风险给出一个下界。

$$\begin{aligned}\inf_{\text{estimator }\hat{\lambda}}\sup_{\lambda\in[a,b]}E_\lambda\left[\sum_{n=1}^{T}(\hat{\lambda}_n-\lambda)^2\right]&\geq\inf_{\text{estimator }\hat{\lambda}}\int_{[a,b]}E_\lambda\left[\sum_{n=1}^{T}(\hat{\lambda}_n-\lambda)^2\right]q(\lambda)d\lambda\\&=\inf_{\text{estimator }\hat{\lambda}}\sum_{n=1}^{T}\int_{[a,b]}E_\lambda\left[(\hat{\lambda}_n-\lambda)^2\right]q(\lambda)d\lambda.\end{aligned}\tag{6.3}$$

回忆一下， $Beta(3,3)$–分布在 $[0,1]$ 上的密度为$q_{[0,1]}(x)=30x^2(1-x)^2$. 不难证明，$q_{[a,b]}(x)=\frac{1}{b-a}q_{[0,1]}(\frac{x-a}{b-a})=\frac{30}{(b-a)^5}(x-a)^2(b-x)^2$是 $[a,b]$上的密度函数。下文中，我们选取密度为$q=q_{[a,b]}$的先验分布。

现在，我们对 (6.3) 的右端应用所谓的 van Trees不等式。例如， [9] 的第 2 章中介绍了这个不等式。对于密度为 $q=q_{[a,b]}$ 的特定先验分布，应用 van Trees不等式的所有条件均得到满足，尤其是条件 $q(a)=q(b)=0$，因此我们得到

$$\inf_{\text{estimator }\hat{\lambda}}\sum_{n=1}^{T}\int_{[a,b]}E_\lambda\left[(\hat{\lambda}_n-\lambda)^2\right]q(\lambda)d\lambda\geq\sum_{n=1}^{T}\frac{1}{\mathbb{I}_{q_{[a,b]}}+\int_{[a,b]}\mathbb{I}_{p^{(n)}}(\lambda)q(\lambda)d\lambda},$$

其中 $\mathbb{I}_{q_{[a,b]}}:=\int\left[\frac{\partial}{\partial\lambda}\ln(q(\lambda))\right]^2q(\lambda)d\lambda$ ，而 $\mathbb{I}_{p^{(n)}}(\lambda)$ 表示泊松分布的 n 重乘积分布的 Fisher 信息。回忆一下，$\mathbb{I}_{p^{(n)}}=n\,\mathbb{I}_{p^{(1)}}=\frac{n}{\lambda}$。

为推导出 $\mathbb{I}_{q_{[a,b]}}$ 的值，我们首先计算 $\mathbb{I}_{q_{[0,1]}}$。注意， $\mathbb{I}_{q_{[0,1]}}=\int\frac{q'(\lambda)^2}{q(\lambda)}d\lambda$，因此

$$\mathbb{I}_{q_{[0,1]}}=\int_0^1\frac{60^2x^2(1-x)^2(1-2x)^2}{30x^2(1-x)^2}dx=120\int_0^1(1-2x)^2dx=40.$$

接下来注意到 $q'_{[a,b]}(x) = \frac{1}{b-a} q'_{[0,1]}(\frac{x-a}{b-a}) \frac{1}{b-a}$，因此

$$\begin{aligned}
\mathbb{I}_{q_{[a,b]}} &= \int_a^b \frac{\frac{1}{(b-a)^4}(q'_{[0,1]}(\frac{x-a}{b-a}))^2}{\frac{1}{b-a} q_{[0,1]}(\frac{x-a}{b-a})} dx \\
&= \int_0^1 \frac{1}{(b-a)^3} \frac{q'_{[0,1]}(u)}{q_{[0,1]}(u)} (b-a) du \\
&= \frac{40}{(b-a)^2},
\end{aligned}$$

其中代入 $u = \frac{x-a}{b-a}$。综上，我们得到

$$\begin{aligned}
\sum_{n=1}^{T} \frac{1}{\mathbb{I}_{q_[a,b]} + \int_{[a,b]} \mathbb{I}_{p^{(n)},n}(\lambda) q(\lambda) d\lambda} &= \sum_{n=1}^{T} \frac{1}{\frac{40}{(b-a)^2} + n \int_{[a,b]} \frac{1}{\lambda} q(\lambda) d\lambda} \\
&\geq \sum_{n=1}^{T} \frac{1}{\frac{40}{(b-a)^2} + n \frac{1}{a} \int_{[a,b]} q(\lambda) d\lambda} \\
&\geq C' \ln(T),
\end{aligned}$$

其中 $C' = \frac{1}{\frac{40}{(b-a)^2} + \frac{1}{a}}$。因此

$$\tilde{c} \inf_{\text{estimator } \hat{\lambda}} \sup_{\lambda \in [a,b]} E_\lambda \left[\sum_{n=1}^{T} (\hat{\lambda}_n - \lambda)^2 \right] \geq \tilde{c} C' \ln(T) = C \ln(T),$$

其中 $C = \tilde{c} C'$。总之，

$$\inf_{\pi \in \Pi} \sup_{\lambda \in \mathcal{M}(L)} \mathcal{R}_\lambda^\pi(T) \geq \inf_{\text{estimator } \hat{\lambda}} \sup_{\lambda \in [a,b]} \tilde{c} E_\lambda \left[\sum_{n=1}^{T} (\hat{\lambda}_n - \lambda)^2 \right] \geq C \ln(T).$$

因此，定理 3.4 得证。

致谢

我们感谢Nabil Kazi-Tani和Michael Neumann进行了富有成效的讨论。
本研究由德国研究基金会（DFG）资助——项目编号547236699。

参考文献

[1] S. Ankirchner、S. Christensen、J. Kallsen、P. L. Borne和S. Perko。学习在布朗噪声下操控系统。*arXiv*预印本 *arXiv:2410.03221*，2024年。[2] M. Basei、X. Guo、A. Hu和Y. Zhang。有限时间范围内分阶段连续时间线性二次强化学习的对数遗憾。机器学习研究杂志，23(178)：1–34，2022年。[3] S. Christensen、N. Dexheimer和C. Strauch。数据驱动的最优停止：纯探索分析。*arXiv*预印本 *arXiv:2312.05880*，2023年。[4] S. Christensen和C. Strauch。脉冲控制问题的非参数学习——探索与利用。应用概率年刊，33(2)：1569–1587，2023年。[5] S. Christensen、C. Strauch和L. Trottner。学习反射：数据驱动随机控制策略的统一方法。*Bernoulli*，30(3)：2074–2101，2024年。[6] S. Christensen、A. H. Thomsen和L. Trottner。多维反射问题的数据驱动规则。即将发表于*SIAM / ASA*《不确定性量化杂志》，*arXiv*预印本 *arXiv:2311.06639*，2024年。[7] T. E. Duncan、L. Guo和B. Pasik-Duncan。自适应连续时间线性二次高斯控制。*IEEE*自动控制汇刊，44(9)：1653–1662，2002年。[8]T. S. Ferguson。最优停止及其应用，2006年。网址 *https://www. math. ucla. edu/˜ tom/Stopping/Contents. html*，2018年。[9] R. D. Gill和B. Y. Levit。van Trees不等式的应用：贝叶斯Cram´er-Rao界。1995年。[10] X. Guo、A. Hu和Y. Zhang。通过反馈控制的稳定性分析，研究含跳跃的线性凸模型的强化学习，2021年。[11] T. M. Moerland、J. Broekens、A. Plaat和C. M. Jonker。基于模型的强化学习：综述。机器学习基础与趋势，16(1)：1–118，2023年。

[12] J. MacQueen 和 R. G. Miller, Jr. 最优持久性策略。运筹学，8:362–380，1960年。[13] E. Neuman 和 Y. Zhang. 传播器模型的次线性遗憾统计学习，2023年。[14] M. Sakaguchi 和 M. Tamaki. 关于空间随机出现的最优停车问题。1982年。[15] L. Szpruch、T. Treetanthiploet 和 Y. Zhang. 线性-凸模型下连续时间情节式强化学习的探索-利用权衡。*arXiv* 预印本 *arXiv:2112.10264*，2021年。