arXiv:2302.01544v1 [cs.LG] 3 Feb 2023

帕累托赌博机中采用非信息先验的汤普森采样的最优性

Jongyeo ng Lee[1,2] Junya Honda[3,2] Chao-Kai Chiang[1] Masashi Sugiyama[2] ,[1]

[1] 东京大学 [2] 理化学研究所人工智能项目 [3] 京都大学

摘要

在随机多臂老虎机问题中，一种称为汤普森采样（TS）的随机概率匹配策略已在各种奖励模型中展现出优异表现。除经验表现之外，已有研究证明，TS 在若干模型中能够达到渐近问题相关下界。然而，其最优性主要是在轻尾模型或属于指数族的单参数模型下得到探讨。本文研究 TS 在具有重尾且由两个未知参数刻画的帕累托模型中的最优性。具体而言，我们讨论采用概率匹配先验的 TS 的最优性，其中包括杰弗里斯先验和参考先验。我们首先证明，采用某些概率匹配先验的 TS 能够达到最优遗憾界。随后，我们证明，采用其他先验（包括杰弗里斯先验和参考先验）的 TS 具有次优性。尽管如此，我们发现，若采用截断过程，使用杰弗里斯先验和参考先验的 TS 能够达到渐近下界。这些结果表明，应谨慎选择非信息先验以避免次优性，并展示了截断过程在基于 TS 的策略中的有效性。

1 引言

在多臂老虎机（MAB）问题中，智能体选择一只臂，并且只能观测所选臂的奖励，这就是部分反馈 [Robbins，1952，Thompson， 1933]。进一步假设，在随机多臂老虎机问题中，奖励由相应臂的分布生成 [Bubeck等人，2012]。由于只能获得部分观测，智能体必须估计未知分布，以判断哪只臂是最优臂，同时避免选择会造成资源损失的次优臂。因此，智能体必须应对探索与利用之间的两难。

在这个问题中，汤普森采样（TS）是一种随机化的贝叶斯策略，它根据某只臂为最优臂的后验概率来选择该臂。由于其出色的实证表现，该策略得到了广泛采用 [Chapelle and Li, 2011, Russo et al., 2018]。在实证成功的基础上，研究者针对多种奖励模型对 TS 进行了理论分析，例如伯努利模型 [Agrawal and Goyal, 2012, Kaufmann et al., 2012], 一维指数族 [Korda 等人， 2013], 高斯模型 [Honda and Takemura，2014], 以及有界支撑模型 [Baudry et al., 2021, Riou and Honda，2020] 并证明了 TS 的渐近最优性。如果某个算法能够达到 Lai 等人针对单参数模型 [1985] 以及 Burnetas and Katehakis 针对多参数或非参数模型 [1996] 推导的问题相关理论下界，则称该算法具有渐近最优性。需要注意的是，任何合理算法的表现都不可能优于这些下界。

除了依赖问题的遗憾分析外，还有一些研究探讨了汤普森采样的与问题无关或先验无关界 [Agrawal andGoyal，2017，Bubeck and Liu，2013，Russo and Van Roy，

2016]。本文研究非信息先验的选择如何影响汤普森采样在任意给定问题实例上的表现。换言之，我们关注汤普森采样的渐近最优性如何取决于非信息先验的选择。

汤普森采样的渐近最优性主要是在单参数模型中得到研究，而其在多参数模型下的最优性尚未得到充分研究。据我们所知，汤普森采样在非紧多参数模型中的渐近最优性，目前仅在高斯老虎机中得到证明 [Honda and Takemura， 2014] 其中均值和方差均未知。他们证明，采用均匀先验的汤普森采样是最优的，而采用杰弗里斯先验和参考先验的汤普森采样则无法达到下界。均匀先验之所以有效，是因为它会频繁选择看似次优的臂。其保守性源于它适度高估了当前次优臂可能成为最优臂的后验概率。

本文研究尾函数具有重尾特性的双参数帕累托模型。我们首先推导出帕累托模型理论下界中问题相关常数的闭式表达式；这并非易事，不同于指数族中的情形。基于这一结果，我们证明，采用某些概率匹配先验的汤普森采样能够达到最优下界。据我们所知，这是关于双参数帕累托老虎机模型的首个此类结果。

我们进一步证明，采用不同概率匹配先验（称为乐观先验）的汤普森采样，其期望遗憾为多项式级。因此，从期望角度来看，选择非信息先验时采取保守策略更有利于避免次优性。尽管如此，我们证明，若对形状参数加入截断过程，采用杰弗里斯先验或参考先验的汤普森采样便可达到最优遗憾界。我们的贡献总结如下：

- 我们证明了在不同先验选择下，汤普森采样的渐近最优性或次优性，从而表明，在双参数帕累托模型中，非信息先验的选择至关重要。
- 我们提供了另一种实现最优性的方法：对后验分布的参数空间进行截断，而非寻找最优先验。

本文的组织结构如下。在第2节中，我们建立帕累托分布下的随机多臂老虎机问题，并推导其遗憾下界。基于非信息先验的选择及其对应的后验分布，我们建立帕累托模型的汤普森采样，并在第3节提出另一种基于汤普森采样的算法，以解决杰弗里斯先验和参考先验的次优性问题。在第4节中，我们给出汤普森采样及采用截断过程的汤普森采样的最优性主要结果，并在第6节概述其证明。第5节给出支持理论分析的数值结果。

2 预备知识

本节将对随机多臂老虎机问题进行形式化描述，并推导出帕累托赌博机期望遗憾下界中出现的问题相关常数的精确形式。

2.1 符号说明

我们考虑随机的 K 臂老虎机问题，其中奖励由参数固定的 Pareto分布生成。智能体在每一轮 $t \in \mathbb{N}$ 选择一个臂 a in $[K] := \{1, \ldots, K\}$，并观测到一个来自 Pa(κ_a， α_a) 的独立同分布奖励，其中 Pa(κ， α) 表示 Pareto

分布，其参数为尺度参数 $\kappa > 0$ 和形状参数 $\alpha > 0$。其密度函数形式如下：

$$f^{\mathrm{Pa}}_{\kappa,\alpha}(x) = \frac{\alpha\kappa^{\alpha}}{x^{\alpha+1}}\mathbb{1}[x \geq \kappa], \tag{1}$$

其中，$\mathbb{1}[\cdot]$ 表示指示函数。我们考虑一个多臂老虎机模型，其中参数 $\theta_a = (\kappa_a, \alpha_a) \in \mathbb{R}_+ \times (1, \infty)$ 对智能体而言未知。我们将服从 $\mathrm{Pa}(\theta_a)$ 的随机变量的均值记为 $\mu_a = \mu(\theta_a) := \frac{\kappa_a\alpha_a}{\alpha_a - 1}$。注意，$\alpha > 1$ 是臂具有有限均值的必要条件，而有限均值是定义次优性差距 $\Delta_a := \max_{i\in[K]} \mu_i - \mu_a$ 所必需的。为简单起见，不失一般性地假设臂 1 具有最大均值，即 $\mu_1 = \max_{i\in[K]} \mu_i$。设 $j(t)$ 为第 $t \in \mathbb{N}$ 轮选择的臂，$N_a(t) = \sum_{s=1}^{t-1} \mathbb{1}[j(s) = a]$ 表示截至第 t 轮选择臂 a 的轮次数。那么，第 T 轮的遗憾为

$$\mathrm{Reg}(T) = \sum_{t=1}^{T} \Delta_{j(t)} = \sum_{a=2}^{K} \Delta_a N_a(T+1).$$

令 $r_{a,n}$ 表示由臂 a 生成的第 n 个奖励。对于帕累托分布，在给定 κ, α 个奖励及其分布的情况下，臂 a 的 n 最大似然估计量（MLEs）如下所示 [Malik,1970]:

$$\hat{\kappa}_a(n) = \min_{s\in[n]} r_{a,s} \sim \mathrm{Pa}(\kappa_a, n\alpha_a), \quad \hat{\alpha}_a(n) = \frac{n}{\sum_{s=1}^{n} \log(r_{a,s}) - n\log\hat{\kappa}_a(n)} \sim \mathrm{IG}(n-1, n\alpha_a), \tag{2}$$

其中 $\mathrm{IG}(n, \alpha)$ 表示形状参数为 $n > 0$、尺度参数为 $\alpha > 0$ 的逆伽马分布。注意，Malik[1970] 进一步证明了 $\hat{\alpha}(n)$ 与 $\hat{\kappa}(n)$ 之间的随机独立性。

2.2 渐近下界

Burnetas andKatehakis [1996] 给出了期望遗憾的问题相关下界。因此，任何一致快速收敛策略（即满足 $\mathrm{Reg}(T) = o(t^{\alpha})$ 对于所有 $\alpha \in (0, 1)$ 均成立的策略）都必须满足

$$\liminf_{T\to\infty} \frac{\mathbb{E}[\mathrm{Reg}(T)]}{\log T} \geq \sum_{a=2}^{K} \frac{\Delta_a}{\inf_{\theta:\mu(\theta)>\mu_1} \mathrm{KL}(\mathrm{Pa}(\kappa_a, \alpha_a), \mathrm{Pa}(\theta))}, \tag{3}$$

其中 $\mathrm{KL}(\cdot, \cdot)$ 表示 Kullback-Leibler（KL）散度。注意，在依赖问题的分析中，多臂老虎机模型 $(\theta_a)_{a\in[K]}$ 被视为一个固定常数。

Pareto分布之间的KL散度为

$$\mathrm{KL}(\mathrm{Pa}(\kappa_1, \alpha_1), \mathrm{Pa}(\kappa_2, \alpha_2)) = \begin{cases} \log\left(\frac{\alpha_1}{\alpha_2}\right) + \alpha_2 \log\left(\frac{\kappa_1}{\kappa_2}\right) + \frac{\alpha_2}{\alpha_1} - 1 & \text{if } \kappa_2 \leq \kappa_1, \\ \infty & \text{otherwise.} \end{cases}$$

这里的散度有时会变为无穷大，因为尺度参数 κ 决定了帕累托分布的支撑集。我们将式（3）中针对 $a \neq 1$ 的分子记作

$$\begin{aligned} \mathrm{KL}_{\inf}(a) &:= \inf_{\theta:\mu(\theta)>\mu_1} \mathrm{KL}(\mathrm{Pa}(\kappa_a, \alpha_a), \mathrm{Pa}(\theta)) \\ &= \inf_{\theta\in\Theta_a} \log\frac{\alpha_a}{\alpha} + \alpha\log\frac{\kappa_a}{\kappa} + \frac{\alpha}{\alpha_a} - 1, \end{aligned}$$

其中

$$\Theta_a = \{(\kappa, \alpha) \in (0, \kappa_a] \times (0, \infty) : \mu(\kappa, \alpha) > \mu_1\}. \tag{4}$$

请注意， Θ_a 允许期望奖励为无穷大的参数（$\alpha \in (0, 1]$），尽管我们考虑的是一个满足 $\alpha_a > 1$ 对所有 $a \in [K]$ 都成立的多臂老虎机模型，因此次优性差距 Δ_a 是有限的。这意味着$\mathrm{KL}_{\inf}(a)$ 并不取决于智能体是否考虑某个臂的期望奖励可能为无穷大。于是，我们可以将式（3）中的下界简单地改写为

$$\liminf_{T\to\infty} \frac{\mathbb{E}[\mathrm{Reg}(T)]}{\log T} \geq \sum_{a=2}^{K} \frac{\Delta_a}{\mathrm{KL}_{\inf}(a)}.$$

以下　　引理给出了该下确界的闭式表达，其证明见附录B　　.

引理1。对于任意臂 $a \neq 1$，有

$$\mathrm{KL}_{\inf}(a) = \log\left(\alpha_a \frac{\mu_1 - \kappa_a}{\mu_1}\right) + \frac{1}{\alpha_a} \frac{\mu_1}{\mu_1 - \kappa_a} - 1.$$

2.3 与有界矩模型的关系

在多臂老虎机领域，已有若干基于上置信界（UCB）的算法被提出，用于处理方差无穷的重尾模型，但需对矩作出额外假设 [Bubeck等人， 2013]。一个主要假设是，任意臂 a 的矩满足 $\mathbb{E}[|r_{a,n}|^\gamma] \leq v$，其中某个固定的 $\gamma \in [1, 2)$ 以及已知的 $v < \infty$ [Bubeck等人， 2013]。注意， γ以下 X 的概率密度函数的$\mathrm{Pa}(\kappa,\ \alpha)$的第

$$\mathbb{E}\left[X^\gamma\right] = \begin{cases} \infty & \alpha \leq \gamma, \\ \frac{\alpha\kappa^\gamma}{\alpha-\gamma} & \alpha > \gamma, \end{cases}$$

这意味着，帕累托模型与有界矩模型互不包含。

最近，Agrawal等人 [2021] 提出了一种基于 KL-UCB 的渐近最优算法，该算法要求在每一轮求解优化问题。由于有界矩模型通常只涵盖某些 Pareto分布，KL-UCB 的已知最优性结果并不一定意味着其在（3）意义下的最优性。

3 汤普森采样与概率匹配先验

汤普森采样是贝叶斯视角下的一种策略，其中先验分布的选择十分重要。尽管在选择先验分布时可以利用关于参数的先验知识，但实践中并不总能获得此类信息。为应对这类情形，我们考虑基于费 Fisher 信息（FI）矩阵的非信息先验，该先验不对未知参数作任何假设。

对于一个随机变量 X，其概率密度为 $f(\cdot|\boldsymbol{\theta})$，FI 定义为得分函数的方差，其中得分函数是 $\log f$ 关于 $\boldsymbol{\theta}$ 的偏导数，具体如下 [Cover和Thomas， 2006]:

$$I_{ij} = [I(\boldsymbol{\theta})]_{ij} := \mathbb{E}_X\left[\left(\frac{\partial}{\partial\theta_i}\log f(X|\boldsymbol{\theta})\right)\left(\frac{\partial}{\partial\theta_j}\log f(X|\boldsymbol{\theta})\right)\middle|\boldsymbol{\theta}\right]. \tag{5}$$

众所周知，当模型满足 FI 正则条件 [Schervish, 2012]时，式（5）中的费舍尔信息矩阵与$\log f(X|\boldsymbol{\theta})$ 的海森矩阵的负期望值相同。然而，$\mathrm{Pa}(\kappa, \alpha)$ 并不

满足这一条件，因为它是一个参数支持族。因此，对于 X，其密度函数如式（1）所示，可根据式（5）求得其费舍尔信息矩阵如下 [Li 等人, 2022]:

$$I(\kappa,\alpha)=\begin{bmatrix}\frac{\alpha^2}{\kappa^2} & 0\\ 0 & \frac{1}{\alpha^2}\end{bmatrix}=\begin{bmatrix}I_{11}(\kappa)I_{11}(\alpha) & 0\\ 0 & I_{22}(\alpha)\end{bmatrix}, \tag{6}$$

其中 $I_{11}(\kappa)=\frac{1}{\kappa^2}, I_{11}(\alpha)=\alpha^2$，以及 $I_{22}(\alpha)=\frac{1}{\alpha^2}$。注意，$I_{11}$ 与 $-\mathbb{E}\left[\frac{\partial^2}{\partial\kappa^2}\log f^{\mathrm{Pa}}_{\kappa,\alpha}(X;\,\theta)\,\middle|\,\theta\right]=\alpha\kappa^2$ 不同。

根据（6），杰弗里斯先验和参考先验分别为 $\pi_{\mathrm{J}}(\kappa,\alpha)\propto\sqrt{\det(I)}=\frac{1}{\kappa}$ 和 $\pi_{\mathrm{R}}(\kappa,\alpha)\propto\sqrt{I_{11}(\kappa)I_{22}(\alpha)}=\frac{1}{\kappa\alpha}$。这里，由参数正交性可知，反向参考先验与参考先验相同 [Datta,1996, Datta and Ghosh, 1995]。

根据参数正交性，当 κ 是关注参数且 α 是干扰参数时，概率匹配先验为

$$\pi_{\mathrm{P}}(\kappa,\alpha)\propto\sqrt{I_{11}}g_1(\alpha)=\frac{\alpha}{\kappa}g_1(\alpha)$$

对于任意 $g_1(\alpha)>0$ [Tibshirani, 1989]。本文考虑先验分布 $\pi(\kappa,\ \alpha)\propto\frac{\alpha^{-k}}{\kappa}$，适用于 $k\in\mathbb{Z}$，因为情形 $k=0,\ 1$ 分别对应杰弗里斯先验和（反向）参考先验。

注 1。本文讨论的帕累托分布有时也称为 Pareto 类型 1 分布 [Arnold, 2008]。另一方面，Kim 等人 [2009] 针对一种称为洛马克斯分布的特殊 Pareto 类型 2 分布推导出了若干非信息先验 [Lomax, 1954], 其中费舍尔信息矩阵可以用负海森矩阵表示。

在多参数情形下，众所周知，杰弗里斯先验存在许多问题 [Datta 和 Ghosh, 1996, Ghosh, 2011]。例如，众所周知，在 Neyman-Scott 问题中，杰弗里斯先验会导致误差方差的不一致估计量 [参见 Berger 和 Bernardo, 1992, 例 3]。这可能是汤普森采样采用杰弗里斯先验时，在多参数分布场景下会产生多项式阶期望遗憾的一个可能原因。有关概率匹配先验和杰弗里斯先验的更多细节，请参见附录 E。

3.1 抽样程序

令 $\mathcal{F}_t:=(j(s),r_{j(s),\,N_{j(s)}(s)})_{s=1}^{t-1}$ 表示截至第 t 轮的历史记录。在先验 $\frac{\alpha^{-k}}{\kappa}$ 且 $k\in\mathbb{Z}$ 的条件下，臂 a 的形状参数的边缘后验分布为

$$\alpha_a\mid\mathcal{F}_t\sim\mathrm{Erlang}\left(N_a(t)-k,\frac{N_a(t)}{\hat{\alpha}_a(N_a(t))}\right), \tag{7}$$

其中，厄朗分布$(s,\ \beta)$ 表示形状参数为 s、速率参数为 β 的埃尔朗分布。注意，我们要求 $\bar{n}\geq\max\{2,k+1\}$ 个初始轮次，以避免 α 的不适当后验分布和最大似然估计。当形状参数 α_a 给定为 β 时，κ_a 的条件后验分布的累积分布函数（CDF）为

$$\mathbb{P}\left[\kappa_a\leq x\mid\mathcal{F}_t,\alpha_a=\beta\right]=\left(\frac{x}{\hat{\kappa}_a(N_a(t))}\right)^{\beta N_a(t)}, \tag{8}$$

若 $0<x\leq\hat{\kappa}_a(N_a(t))$。由于可以按照与 Sun 等人 [2020],相同的步骤推导后验分布，详细推导见附录 E.3。在第 t轮次，我们将抽样得到的尺度参数和形状参数记为

算法1 STS/ STS-T

1：参数： $k \in \mathbb{Z}$, $\bar{n} = \max\{2, k+1\}$。**2：初始化：每个臂** $\bar{n}$ **次。3：循环：4：抽样** $\tilde{\alpha}_a(t) \sim$ **厄朗分布** $\left(N_a(t) - k,\ {}^{N_a}(t)\hat{\alpha}_a(N_a(t))\right)$。**5：** $\bar{\alpha}_a(N_a(t)) \leftarrow \min(N_a(t), \hat{\alpha}_a(N_a(t)))$。**6：抽样** $\tilde{\alpha}_a(t) \sim$ **厄朗分布** $\left(N_a(t) - k,\ {}^{N_a}(t)\bar{\alpha}_a(N_a(t))\right)$。**7：若** $\{a \in [K] :\ \tilde{\alpha}_a(t) \leq 1\} \neq \emptyset$ **则 8：选择** $j(t) = \arg\min_{a\in[K]} \tilde{\alpha}_a(t)$。**9：否则 10：对每个** $a \in [K]$**从** $u_a \sim U(0, 1)$**中抽样。11：** $\tilde{\kappa}_a(t) = \hat{\kappa}_a(N_a(t))u^{1/({}^{N_a}(t)\tilde{\alpha}_a(t))_a}$。**12：选择并执行** $j(t) = \arg\max_{a\in[K]} {}^{\tilde{\kappa}_a}(t)\tilde{\alpha}_a(t)\tilde{\alpha}_a(t)_{-1}$**13：** $=$ $\mathbf{argmax}_{a\in[K]}\tilde{\mu}_a(t)$。**14：结束若**

臂的参数 a 由 $\tilde{\kappa}_a(t)$ 和 $\tilde{\alpha}_a(t)$分别表示，相应的平均奖励表示为 $\tilde{\mu}_a(t) := \mu(\tilde{\kappa}_a(t)\ \tilde{\alpha}_a(t))$。我们首先从式（7）的边缘后验分布中抽样得到形状参数。然后，给定抽样得到的形状参数，我们利用逆变换抽样，根据式（8）中条件后验分布的累积分布函数抽样得到尺度参数。基于这一序贯过程的汤普森采样称为序贯汤普森采样（STS），可表述为算法 1。

在下一节给出的定理 3 中，根据式（3）的下界，采用 Jeffreys 先验和参考先验的 STS 被证明是次优的。其次优性主要源于当小 $N_1(t) = n$ 时，$\hat{\alpha}_1(n)$ 被高估后，式（7）中的后验分布所呈现的行为。为克服这些问题，我们提出了 STS-T 策略，即带截断的 STS 变体：在式（7）中，将 $\hat{\alpha}(n)$ 替换为 $\bar{\alpha}(n) := \min(n, \hat{\alpha}(n))$。请注意，这种截断过程专门用于式（7）和式（8）的后验抽样。我们证明，采用 Jeffreys 先验和参考先验的 STS-T 可达到定理4中的最优遗憾界。

3.2 先验参数 k的解释

厄朗分布是伽马分布的一种特殊情况，其形状参数为正整数。若随机变量 X 服从厄朗分布$(s,\ \beta)$，则其概率密度的形式为

$$f^{\mathrm{Er}}_{s,\beta}(x) = \frac{\beta^s}{\Gamma(s)} x^{s-1} e^{-\beta x} \mathbb{1}[x \in \mathbb{R}_+], \tag{9}$$

其中，$s \in \mathbb{N}$ 和 $\beta > 0$ 分别表示形状参数和率参数。于是，在 $x > 0$处的累积分布函数为

$$F^{\mathrm{Er}}_{s,\beta}(x) = \frac{\int_0^{\beta x} t^{s-1} e^{-t} \mathrm{d}t}{\Gamma(s)} = \frac{\gamma(s, \beta x)}{\Gamma(s)}, \tag{10}$$

其中，$\gamma(\cdot,\ \cdot)$ 表示下不完全伽马函数。由于 $\gamma(s+1,\ x) = s\gamma(s,\ x) - x^s e^{-x}$ 成立，可以观察到，对于任意 $x > 0$

$$F^{\mathrm{Er}}_{s,\beta}(x) \geq F^{\mathrm{Er}}_{s+1,\beta}(x). \tag{11}$$

根据 STS 和 STS-T 的抽样程序， $\tilde{\mu}$ 仅在 $\tilde{\alpha} > 1$ 成立时依赖于 $\tilde{\kappa}$ ，因为 $\tilde{\alpha} \leq 1$ 导致 $\tilde{\mu}(\cdot, \tilde{\alpha}) = \infty$。因此，对于（8）中的任意 $\beta > 1$ ，当 $N_a(t)$充分大时， $\tilde{\kappa}$ 将集中于$\hat{}\kappa = n$。因此， $\tilde{\mu}$ 将主要由 $\tilde{\alpha}$ 和$\hat{}\kappa$决定，其中 k 的选择会根据（7）影响 $\tilde{\alpha}$ 的抽样。由（11）可知，随着形状参数 $n-k$ 减小，抽取较小 $\tilde{\alpha}$ 的概率会增加。因此，对于相同的 n，随着 k 增加，次优臂的 $\tilde{\mu}$会增加。换言之，随着 k 增加，抽取较大 $\tilde{\mu}$ 的概率也会增加。因此，具有较大 k 的汤普森采样会成为一种保守策略，可能频繁选择当前的次优臂。相反，具有较小 k 的先验分布会产生一种乐观策略，专注于选择当前最优臂。

4 主要结果

本节给出采用不同 $k \in \mathbb{Z}$ 时 STS 和 STS-T 的遗憾界。首先，我们展示当先验分布为 $\pi(\kappa, \alpha) \propto \frac{\alpha^{-k}}{\kappa}$ with $k \in \mathbb{Z}_{\geq 2}$ 时，STS 的渐近最优性。

定理 2. 假设臂 1 是唯一的最优臂，且具有有限均值。对于每个 $a \in [K]$，令ε_a =最小值 $\left\{ \frac{\kappa_a(\kappa_a+1)}{\alpha_a} \right.$ μ_a $\frac{\kappa_a\delta_a}{\mu_a+\delta_a-\kappa_a()+\kappa_a\delta_a}$, $\frac{\kappa_a\delta_a}{(1+\mu_a+\delta_a)}$ $\left.\right\}$其中 $\delta_a = \frac{\Delta_a}{2}$对于 $a \neq 1$ 且 δ_1 =最小值$_{a\neq 1}\,\delta_a$。给定任意 $\epsilon \in (0,$ 最小值$_{a\in[K]}\,\varepsilon_a)$，采用 $k \in \mathbb{Z}_{\geq 2}$ 的STS 的期望遗憾满足如下界

$$\mathbb{E}[\mathrm{Reg}(T)] \leq \sum_{a=2}^{K} \frac{\Delta_a \log T}{D_{a,k}(\epsilon)} + \mathcal{O}\left(\epsilon^{-2}\right),$$

其中 $D_{a,\,k}(\epsilon)$ 是一个函数，满足对于任意固定的 $k \in \mathbb{Z}$，$\lim_{\epsilon\to 0} D_{a,\,k}(\epsilon) = \mathrm{KL}_{\inf}(a)$ 。

将定理 $\epsilon = o(1)$ 中的 2代入，可知 STS 在 $k \in \mathbb{Z}_{\geq 2}$ 下满足

$$\liminf_{T\to\infty} \frac{\mathbb{E}[\mathrm{Reg}(T)]}{\log T} \leq \sum_{a=2}^{K} \frac{\Delta_a}{\mathrm{KL}_{\inf}(a)},$$

这表明，STS 在下界（3）意义下具有渐近最优性。

接下来，我们将证明，在下文的定理中，STS 在 $k \in \mathbb{Z}_{\leq 1}$ 下无法达到渐近界。沿用高斯老虎机证明中的思路 [Honda and Takemura, 2014], 我们考虑双臂老虎机问题，并假设已知次优臂的完全信息，以简化分析。我们进一步假设两个臂具有相同的尺度参数 $\kappa_1 = \kappa_2$。

定理 3. 考虑一个双臂老虎机问题，其中 $\kappa_1 = \kappa_2$ 且 $1 < \alpha_1 < \alpha_2$。当 $\tilde{\alpha}_1(t)$ 和 $\tilde{\kappa}_1(t)$分别从（7）和（8）中的后验分布抽样，且 $\tilde{\mu}_2(t) = \mu_2$ 成立时，存在常数$C(\alpha_1, \alpha_2) > 0$ 满足

$$\liminf_{T\to\infty} \frac{\mathbb{E}[\mathrm{Reg}(T)]}{\log T} \geq C(\alpha_1, \alpha_2),$$

其中， $C(\alpha_1, \alpha_2) > \mathrm{KL}_{\inf}(2)$ 在某些实例中成立。特别是，对于 $k \in \mathbb{Z}_{\leq 0}$，存在满足以下条件的 $C'(\alpha_1, \alpha_2)$

$$\liminf_{T\to\infty} \frac{\mathbb{E}[\mathrm{Reg}(T)]}{\sqrt{T}} \geq C'(\alpha_1, \alpha_2).$$

根据定理 2 和 3，我们发现，当考虑最大化期望奖励。

尽管已有研究表明，采用杰弗里斯先验（$k = 0$）和参考先验（$k = 1$）的 STS 并非最优，但我们证明，改进后的算法 STS-T 可在 $k \in \mathbb{Z}_{\geq 0}$的情况下达到最优遗憾界。

定理4。 沿用定理 2中的相同记号，STS-T 在 $k \in \mathbb{Z}_{\geq 0}$ 下的期望遗憾有如下界

$$\mathbb{E}[\mathrm{Reg}(T)] \leq \sum_{a=2}^{K} \frac{\Delta_a \log T}{D_{a,k}(\epsilon)} + \mathcal{O}(\epsilon^{-m}),$$

其中 $m = \max(2,\ 3-k)$。

根据定理 2 和 4，要达到式（3）中的下界，有两种选择：使用带最大似然估计量的保守先验，或使用带截断样本的中度乐观先验。由于初始化步骤要求每个臂至少被选择 $\max(2, k+1)$ 次，如果臂的数量 K 较大，那么使用杰弗里斯先验或参考先验搭配截断估计量会是更好的选择。另一方面，如果模型可能包含 α较大的臂，而截断在较小的 n时可能会带来问题，那么最好使用带保守先验的 STS 。

5 实验

本节展示数值结果，以说明 STS 和 STS-T的性能，并为我们的理论分析提供支持。我们以4臂多臂老虎机模型 $\boldsymbol{\theta}_4$ 为例，参数见表 1；在该模型中，次优臂的 κ 分别小于、等于和大于最优臂。的参数为 $\boldsymbol{\mu} = (4.55, 3.2, 2.74, 3)$ ，且方差无穷大。更多实验结果见附录 H。

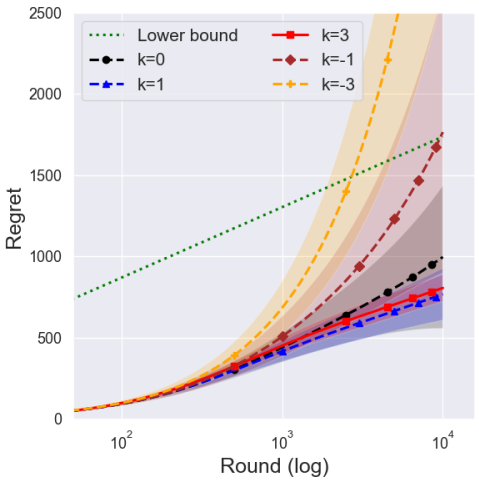
表 1：四臂老虎机模型 $\boldsymbol{\theta}_4$。

	臂 1	臂 2	臂3	臂 4
κ	1.3	1.2	1.3	1.5
α	1.4	1.6	1.9	2.0

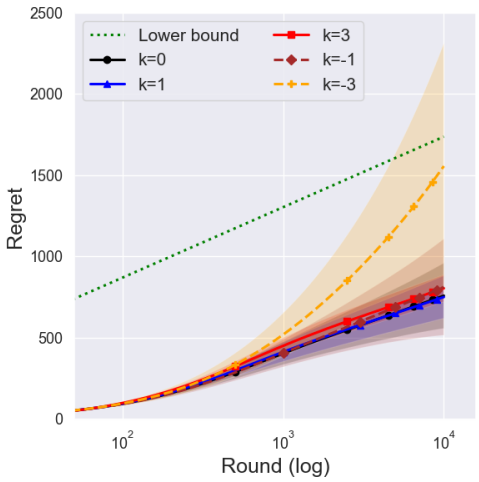
图 1 展示了先验分布参数取不同值时所提出策略的累积遗憾 k。实线表示对能够达到式（3）中最优下界的先验分布进行 100,000 次独立运行所得的平均累积遗憾，虚线则表示无法达到该下界的先验分布对应的结果。绿色点线表示问题相关下界，阴影区域表示四分之一标准差。

在图 2 和 3中，我们考察了STS 与 STS-T 之间在相同 k下的差异。实线表示 100,000 次独立运行所得的平均累积遗憾。阴影区域和虚线分别表示中央 99% 区间和遗憾的上0.05% 部分。

杰弗里斯先验（$k = 0$）在图1(a)中，与先验分布 $k = 1$、3相比，杰弗里斯先验似乎导致了更高阶的遗憾，而先验分布4 在这种情况下表现最佳。正如定理4 所述，它在STS-T下的表现有所改善，与 $k = 1$、3的表现相近。图2(a) 展示了这种改善可能的原因：在STS-T下，遗憾的中央99%区间明显缩小。由于采用杰弗里斯先验（$k = 0$）的STS次优性源于一种极端情况：它以较小的概率导致多项式遗憾，因此，这种区间缩小有助于降低采用杰弗里斯先验的STS-T 的期望遗憾。

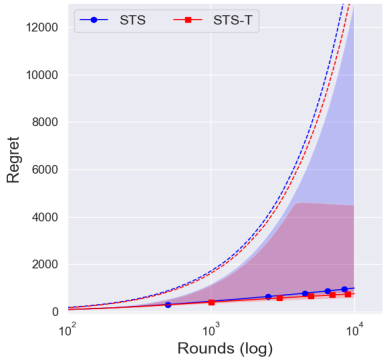


(a) 不同 k 下 STS 的累积遗憾

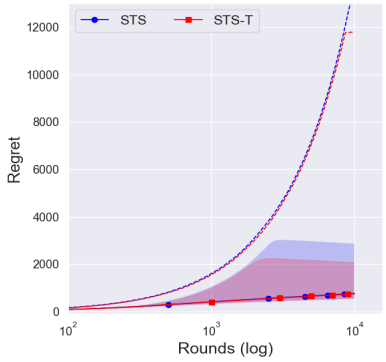


(b) 不同 k 下 STS-T 的累积遗憾

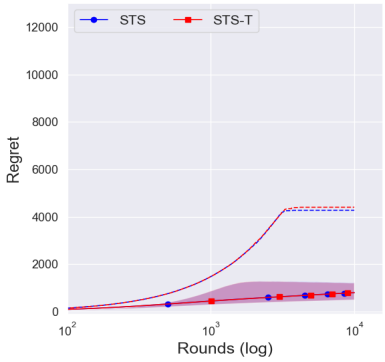
图1：实线表示在 100,000 次独立运行中，能够达到（3）中最优下界的先验分布对应的平均累积遗憾。虚线表示无法达到（3）中最优下界的先验分布对应的平均累积遗憾。阴影区域表示四分之一标准差。绿色点线表示基于引理 1 的问题相关下界。



(a) 杰弗里斯先验 $k = 0$



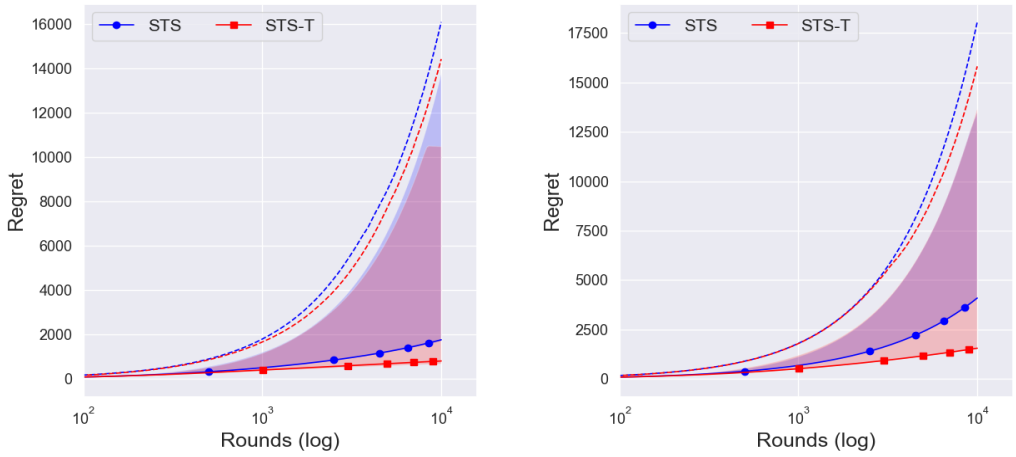
(b) 参考先验 $k = 1$



(c) 具有 $k = 3$ 的先验

图2：实线表示对独立运行100,000次所得遗憾的平均值。阴影区域和虚线分别表示中央99%区间和遗憾的上0.05%。

参考先验（$k = 1$）参考先验的表现与渐近最优先验 $k = 3$相近，尽管定理3表明，在某些实例中，STS下的参考先验并非最优。与杰弗里斯先验($k = 0$)类似，图2(b)所示，采用参考先验（$k = 1$）的 STS-T 下的遗憾中央99%区间比STS 下更窄，不过其缩减幅度相较于杰弗里斯先验要小得多。这意味着参考先验比杰弗里斯先验更加保守。



(a) 先验为 $k = -1$　　(b) 先验为 $k = -3$

图3：实线表示对独立运行 100,000 次所得遗憾的平均值。阴影区域和虚线分别表示中央99%区间和遗憾的最高 0.05%。

保守先验（$k = 3$）有趣的是，图2(c)显示，截断过程不会影响遗憾的中央99% 区间，甚至会降低最高0.05%部分的性能。注意，0.05% 遗憾最高部分的 $k = 3$ 远低于 $k = 0,1$，这体现了图2中保守先验的稳定性。

由于采用截断过程是为了防止对 $k \in \mathbb{Z}_{\leq 1}$ 而言成问题的极端情况，因此可以自然地看到，使用 $k = 3$ 时，STS 与 STS-T 之间没有差异。这意味着 $k = 3$ 足够保守，因此截断过程不会影响整体性能。

乐观先验（$k < 0$）在图1(a)中可以看到，$k = -1$ 和 $k = -3$的平均遗憾增长速度远快于 $k = 0$，1，3 在STS 策略下，这说明了采用先验分布 $k \in \mathbb{Z}_{<0}$的STS的次优性。

由于图1中的乐观先验（$k < 0$）在 STS-T 下表现更好，我们可以检验在采用乐观先验进行后验抽样时，截断过程的有效性。然而，如图3(a)和图3(b)所示，使用 $k = -1$、-3 时，STS 和 STS-T 的遗憾中央 99% 区间并无明显差异，这可能意味着带有 $k \in \mathbb{Z}_{<0}$ 的先验过于乐观。因此，我们可能需要采用更保守的截断过程，例如使用 $\bar{\alpha}_a(n) = \max(\sqrt{n}, \hat{\alpha}_a(n))$ 或 $\max(\log n, \hat{\alpha}_a(n))$ 的方法，这会导致有限时间范围内的遗憾更大。

6 最优结果的证明概要

本节给出定理2 和定理4的证明概要，详细证明见附录C。注意，定理3 的证明推迟至附录D。

我们首先考虑由下式定义的最大似然估计量上的良好事件：

$$\begin{aligned}\mathcal{K}_{a,n}(\epsilon) &:= \{\hat{\kappa}_a(n) \in [\kappa_a, \kappa_a + \epsilon]\} \\ \mathcal{A}_{a,n}(\epsilon) &:= \{\hat{\alpha}_a(n) \in [\alpha_a - \epsilon_{a,l}(\epsilon), \alpha_a + \epsilon_{a,u}(\epsilon)]\} \\ \mathcal{E}_{a,n}(\epsilon) &:= \mathcal{K}_{a,n}(\epsilon) \cap \mathcal{A}_{a,n}(\epsilon),\end{aligned}$$

其中 $n \in \mathbb{N}$ 和

$$\epsilon_{a,l}(\epsilon) = \frac{\epsilon\alpha_a^2}{1+\epsilon\alpha_a}, \quad \epsilon_{a,u}(\epsilon) = \frac{\epsilon\alpha_a^2(\kappa_a+1)}{\kappa_a - \epsilon\alpha_a(\kappa_a+1)}. \tag{12}$$

注意，对于任意 $n \geq \alpha_a + 1$，在$\mathcal{A}_{a,n}(\epsilon)$上，$\bar{\alpha}_a(n) = \hat{\alpha}_a(n)$成立。这里，我们设定 ε_a，使得对于任意 $\epsilon \leq \varepsilon_a$，在 $\mathcal{E}_a(\epsilon)$上$\hat{\mu}_a \in [\mu_a - \delta_a, \mu_a + \delta_a]$ 成立。定义最优臂样本上的一个事件

$$\mathcal{M}_\epsilon(t) := \{\tilde{\mu}_1(t) \geq \mu_1 - \epsilon\}.$$

那么，第 T 轮次的期望遗憾可分解如下：

$$\begin{aligned}\mathbb{E}[\mathrm{Reg}(T)] &= \mathbb{E}\left[\sum_{t=1}^{T} \Delta_{j(t)}\right] \\ &= \sum_{a=2}^{K} \Delta_a \left(\bar{n} + \sum_{t=\bar{n}K+1}^{T} \mathbb{E}[\mathbb{1}[j(t)=a]]\right) \\ &\leq \Delta_2 \sum_{t=\bar{n}K+1}^{T} \left(\mathbb{E}\left[\mathbb{1}[j(t)\neq 1, \mathcal{K}_{1,N_1(t)}(\epsilon), \mathcal{M}_\epsilon^c(t)]\right] + \mathbb{E}\left[\mathbb{1}[j(t)\neq 1, \mathcal{K}_{1,N_1(t)}^c(\epsilon), \mathcal{M}_\epsilon^c(t)]\right]\right) \\ &\quad + \sum_{a=2}^{K} \Delta_a \Bigg\{\bar{n} + \sum_{t=\bar{n}K+1}^{T} \Bigg(\mathbb{E}\left[\mathbb{1}[j(t)=a, \mathcal{M}_\epsilon(t), \mathcal{E}_{a,N_a(t)}(\epsilon)]\right] \\ &\qquad\qquad + \mathbb{E}\left[\mathbb{1}[j(t)=a, \mathcal{M}_\epsilon(t), \mathcal{E}_{a,N_a(t)}^c(\epsilon)]\right]\Bigg)\Bigg\},\end{aligned}$$

其中 $\mathcal{E}^c$ 表示 $\mathcal{E}$ 的补集。引理 5–8 完成了定理 2 和 4的证明，证明见附录C。

引理 5。 在 $k \in \mathbb{Z}_{\geq 2}$条件下，STS

$$\sum_{t=\bar{n}K+1}^{T} \mathbb{E}\left[\mathbb{1}[j(t)\neq 1, \mathcal{K}_{1,N_1(t)}^c(\epsilon), \mathcal{M}_\epsilon^c(t)]\right] \leq \mathcal{O}(\epsilon^{-2}).$$

在 $k \in \mathbb{Z}_{\geq 0}$条件下，STS-T 也成立，

$$\sum_{t=\bar{n}K+1}^{T} \mathbb{E}\left[\mathbb{1}[j(t)\neq 1, \mathcal{K}_{1,N_1(t)}^c(\epsilon), \mathcal{M}_\epsilon^c(t)]\right] \leq \mathcal{O}(\epsilon^{-m}),$$

其中 $m = \max(2, 3-k)$。

虽然引理 6 对遗憾的主要项有贡献，但引理 5 的证明是遗憾分析中的主要难点。我们发现，对于 STS，我们的分析无法得到有限的 $k \in \mathbb{Z}_{<2}$上界，因此设计了STS-T 来解决这类问题。

引理 6. 在 STS 和 STS-T 满足 $k \in \mathbb{Z}$ 的情况下，对于任意 $a \in [K]$，结论成立

$$\sum_{t=\bar{n}K+1}^{T} \mathbb{E}[\mathbb{1}[j(t)=a, \mathcal{M}_\epsilon(t), \mathcal{E}_{a,N_a(t)}(\epsilon)]] \leq \max(0,k)+1+\frac{1}{\alpha_a \epsilon}\mathbb{1}[k>0]+\frac{\log T}{D_{a,k}(\epsilon)}.$$

其中 $D_{a,\,k}(\epsilon) > 0$ 是满足$\lim_{\epsilon\to 0} D_{a,\,k}(\epsilon) = \mathrm{KL}_{\inf}(a)$ 的有限问题确定性常数。

由于较大的 k 会产生更保守的策略，并且需要每个臂进行更多初始轮次，因此较大的k 可能会在有限时间范围 T内导致更大的遗憾，这对应于引理6中讨论的遗憾分量。因此，该引理意味着策略需要在一定程度上保持保守，而过于保守则会在有限时间内导致更大的遗憾。

引理 7。在STS和STS-T 下，其中 $k \in \mathbb{Z}_{\geq 0}$,

$$\sum_{t=\bar{n}K+1}^{T} \mathbb{E}\left[\mathbb{1}[j(t)\neq 1, \mathcal{K}_{1,N_1(t)}(\epsilon), \mathcal{M}_\epsilon^c(t)]\right] \leq \mathcal{O}(\epsilon^{-1}).$$

引理 7 的关键是将关于 $\tilde{\mu}_1(t)$、 $\mathcal{M}_\epsilon(t)$ 的项转换为关于 $\tilde{\alpha}_1(t)$ 的项。由于对于 $\alpha \leq 1$，若 $\mu(\kappa, \alpha) = \infty$成立，则无论 $\tilde{\kappa}_1$ 取何值， $\tilde{\mu}_1 = \mu(\tilde{\kappa}_1, \tilde{\alpha}_1)$ 都会变为无穷大；若 $\tilde{\alpha}_1 \leq 1$ 成立，这意味着 $\mathbb{P}[\mathcal{M}_\epsilon^c(t)\ \tilde{\alpha}_1(t) \leq 1] = 0$。因此，要证明引理 7，只需考虑 $\tilde{\alpha}_1(t) > 1$ 成立的情况。尽管 STS 和 STS-T 下 $\tilde{\alpha}_1$ 的密度函数不同，但给定 $\alpha_1 = \tilde{\alpha}_1$ 时，$\tilde{\kappa}_1$ 的条件累积分布函数相同，如式 (8) 所示:

$$\mathbb{P}[\tilde{\kappa}_1 \leq x | \mathcal{F}_t, \tilde{\alpha}_1 = \alpha_1] = \left(\frac{x}{\hat{\kappa}_1(N_1(t))}\right)^{\tilde{\alpha}_1 N_1(t)}.$$

因此，当 $N_1(t)$ 和 $\tilde{\alpha}_1(t)$ 足够大时， $>$ 1($\tilde{\kappa}_1(t)$) 将以高 ˜ ˜ ˜概率集中于 ˆ$\kappa_1(N_1(t))$，后者在事件 $\{\mathcal{K}_{1,N_1(t)}(\epsilon)\}$ 下接近其真实值 κ_1。因此， $\tilde{\mu}_1 = \frac{\kappa_1\alpha_1}{\tilde{\alpha}_1 - 1} \geq \frac{\kappa_1\alpha_1}{\tilde{\alpha}_1 - 1} = \mu(\kappa_1, \tilde{\alpha}_1)$ 以高概率成立，这意味着 $\mathbb{P}[\mathcal{K}_{1,N_1(t)}(\epsilon)\ \mathcal{M}_\epsilon^c(t)|\mathcal{F}_t]$ 可以由 $\mathbb{P}[\mathcal{K}_{1,N_1(t)}(\epsilon)\ \tilde{\alpha}_1(t) \geq c|\mathcal{F}_t]$ 进行粗略界定，其中 $c > 1$ 为某些问题相关常数。由于给定 $\mathcal{F}_t$ 时 $\mathcal{K}_1$ 是确定的，我们有

$$\mathbb{P}[\mathcal{K}_{1,N_1(t)}(\epsilon), \tilde{\alpha}_1(t) \geq c|\mathcal{F}_t] = \mathbb{1}[\mathcal{K}_{1,N_1(t)}(\epsilon)]\mathbb{P}[\tilde{\alpha}_1(t) \geq c|\mathcal{F}_t],$$

这意味着，在事件 $\{\mathcal{K}_{1,N_1(t)}(\epsilon)\}$ 下，对于两种策略， $\tilde{\mu}_1(t)$ 主要由 $\tilde{\alpha}_1(t)$ 的取值决定。在这种情况下，STS 和 STS-T 的行为类似于尺度参数已知的帕累托分布中的汤普森采样，其中$\tilde{\mu}_1(t) := \mu(\kappa_1, \tilde{\alpha}_1(t))$，对于 $t \in \mathbb{N}$ 成立。这里，尺度参数已知的帕累托分布属于一维指数族，Korda 等人 [2013] 证明了采用杰弗里斯先验的汤普森采样的最优性。由于在单参数帕累托模型中，采用杰弗里斯先验时 α 的后验分布是形状为 $N_1(t)+1$ 的厄朗分布，因此我们可以利用 Korda 等人 [2013] 的结果，并结合厄朗分布的一些性质（例如性质 (11)），来证明引理 7。

引理 8。 在 STS 和 STS-T 下，对于任意 $a \neq 1$，在 $k \in \mathbb{Z}$ 的条件下均成立

$$\sum_{t=\bar{n}K+1}^{T} \mathbb{E}\left[\mathbb{1}[j(t)=a, \mathcal{E}^c_{a,N_a(t)}(\epsilon)]\right] \leq \mathcal{O}\left(\epsilon^{-2}\right).$$

引理8 描述了所选臂的估计量与其真实参数相差较远时所引起的遗憾。这与通常对汤普森采样的分析一样，并不难处理。事实上，该引理的证明很直接，因为可以根据式（ $\mathbb{P}[\mathcal{K}_a^c]$ 和 $\mathbb{P}[\mathcal{K}_a, \mathcal{A}_a^c]$ 中ˆκ_a 和ˆα_a 的分布，轻松推导出2）的上界。

7 结论

我们研究了帕累托分布下的多臂老虎机问题。帕累托分布具有重尾并服从幂律。以往关于汤普森采样的大多数研究都集中于一维分布或轻尾分布，而我们关注的是由未知尺度参数和形状参数刻画的帕累托分布。通过依次从边缘后验分布和条件后验分布中抽取参数，我们能够实现高效的抽样程序。我们首次证明，在这种设定下，采用恰当先验分布的汤普森采样可以达到依赖于问题的最优遗憾界。尽管研究表明，直接应用汤普森采样时，Jeffreys先验和参考先验并非最优，但我们证明，加入截断过程后，它们能够达到最优遗憾界。实验结果支持了我们的理论结论，表明保守先验的最优性，以及截断过程对Jeffreys先验和参考先验的有效性。

致谢

JL 获得了 JST SPRING 项目资助，资助编号为 JPMJSP2108。CC 和 MS 获得了东京大学人工智能与超越研究所的资助。

参考文献

Shipra Agrawal 和 Navin Goyal。《多臂老虎机问题的汤普森采样分析》。载于学习理论会议，第 39–1 页。PMLR，2012 年。

Shipra Agrawal 和 Navin Goyal。《汤普森采样的近最优遗憾界》。计算机协会会刊，64(5)：1–24，2017 年。

Shubhada Agrawal、Sandeep K Juneja 和 Wouter M Koolen。《重尾老虎机中的遗憾最小化》。载于学习理论会议，第 26–62 页。PMLR，2021 年。

Barry C Arnold. 帕累托分布与广义帕累托分布。载于收入分布与洛伦兹曲线建模，第119–145页。施普林格，2008。Dorian Baudry、Romain Gautron、Emilie Kaufmann和Odalric Maillard。适用于支撑集感知型条件风险价值老虎机的最优汤普森采样策略。载于国际机器学习会议，第716–726页。PMLR，2021。James O Berger和José M Bernardo。论参考先验方法的发展。贝叶斯统计，4(4):35–60，1992。Jose M Bernardo。贝叶斯推断的参考后验分布。英国皇家统计学会会刊：*B*辑（方法学），41(2):113–128，1979。Sébastien Bubeck和Che-Yu Liu。汤普森采样的无先验与依赖先验的遗憾界。载于神经信息处理系统进展，第26卷，第638–646页，2013。Sébastien Bubeck、Nicolo Cesa-Bianchi等。随机与非随机多臂老虎机问题的遗憾分析。机器学习基础与趋势，5(1):1–122，2012。

Sébastien Bubeck、Nicolo Cesa-Bianchi和Gábor Lugosi。重尾老虎机。*IEEE* 信息论汇刊，59(11):7711–7717，2013。Apostolos N Burnetas和Michael N Katehakis。序贯分配问题的最优自适应策略。应用数学进展，17(2):122–142，1996。Olivier Chapelle和Lihong Li。对汤普森采样的实证评估。载于神经信息处理系统进展，第24卷。Curran Associates, Inc.，2011。Thomas M Cover和Joy A Thomas。信息论基础。Wiley-Interscience，第2版，2006。Gauri Sankar Datta。为多个参数函数的贝叶斯推断提供频率学派有效性的先验。*Biometrika*，83(2):287–298，1996。Gauri Sankar Datta和Malay Ghosh。关于非信息先验的一些评论。《美国统计学会杂志》，90(432):1357–1363，1995。Gauri Sankar Datta和Malay Ghosh。论非信息先验的不变性。《统计年鉴》，24(1):141–159，1996。Gauri Sankar Datta和Trevor J Sweeting。概率匹配先验。《统计学手册》，25:91–114，2005。Thomas J DiCiccio、Todd A Kuffner和G Alastair Young。位置尺度模型中精确概率匹配先验的简单分析。《美国统计学家》，71(4):302–304，2017。Malay Ghosh。客观先验：面向频率学派的介绍。《统计科学》，26(2):187–202，2011。Junya Honda和Akimichi Takemura。高斯老虎机中汤普森采样的最优性取决于先验。载于人工智能与统计国际会议，第33卷，第375–383页。PMLR，2014。Harold Jeffreys。概率论。牛津大学出版社，1998。Emilie Kaufmann、Nathaniel Korda和Rémi Munos。汤普森采样：渐近最优有限时间分析。载于国际算法学习理论会议，第7568卷，第199–213页。施普林格，2012。Dal-Ho Kim、Sang-Gil Kang和Woo-Dong Lee。帕累托分布的非信息先验。《韩国数据与信息科学学会杂志》，20(6):1213–1223，2009。Nathaniel Korda、Emilie Kaufmann和Remi Munos。一维指数族老虎机的汤普森采样。载于神经信息处理系统国际会议，第26卷，第1448–1456页。Curran Associates, Inc.，2013。Tze Leung Lai、Herbert Robbins等。渐近有效的自适应分配规则。应用数学进展，6(1):4–22，1985。Mingming Li、Huafei Sun和Linyu Peng。帕累托分布的Fisher–Rao几何与杰弗里斯先验。《统计学通讯：理论与方法》，51(6):1895–1910，2022。

Kenneth S Lomax。《企业失败：失败数据分析的另一个示例》。《美国统计学会杂志》，49(268):847–852，1954。Henrick John Malik。《帕累托分布参数的估计》。*Metrika*，15(1):126–132，1970。Rahul Mukerjee 和 Malay Ghosh。《二阶概率匹配先验》。*Biometrika*，84(4):970–975，1997。Charles Riou 和 Junya Honda。《基于汤普森采样的有界奖励分布多臂老虎机算法》。载于国际算法学习理论会议，第777–826页。PMLR，2020。Herbert Robbins。《序贯实验设计的若干方面》。美国数学学会通报，58(5):527–535，1952。Christian P Robert 等。《贝叶斯选择：从决策理论基础到计算实现》。施普林格，第2版，2007。Daniel Russo 和 Benjamin Van Roy。《汤普森采样的信息论分析》。机器学习研究杂志，17(1):2442–2471，2016。Daniel J Russo、Benjamin Van Roy、Abbas Kazerouni、Ian Osband、Zheng Wen 等。《汤普森采样教程》。机器学习基础与趋势，11(1):1–96，2018。Mark J Schervish。《统计学理论》。施普林格科学与商业媒体，2012。Marvin K Simon 和 Dariush Divsalar。《高斯概率积分相关问题的一些新变化》。*IEEE*通信汇刊，46(2):200–210，1998。Charles Stein。《信任区间与置信区间之间存在巨大差异的一个示例》。数学统计年刊，30(4):877–880，1959。Fupeng Sun、Yueqi Cao、Shiqiang Zhang 和 Huafei Sun。《基于信息几何的帕累托模型贝叶斯推断》。*Entropy*，23(1):45，2020。William R Thompson。《根据两组样本的证据，一个未知概率大于另一个未知概率的可能性》。*Biometrika*，25(3/4):285–294，1933。Robert Tibshirani。《多个参数中某一个参数的非信息先验》。*Biometrika*，76(3):604–608，1989。Roman Vershynin。《高维概率：应用于数据科学的导论》，第47卷。剑桥大学出版社，2018。David L. Wallace。《Student t分布和卡方分布正态近似的界》。数学统计年刊，30(4):1121–1130，1959。BL Welch 和 HW Peers。《基于加权似然积分的置信点公式》。英国皇家统计学会会刊：*B*辑（方法学），25(2):318–329，1963。

A 符号说明

表2–5 总结了本文中使用的符号。

表 2：老虎机问题的符号说明。

符号	含义
K	臂的数量 .
T	时间范围 .
$j(t)$	第 t轮所选臂的索引。
$k \in \mathbb{Z}$	先验参数，详见第 3 节。
$\bar{n} =$ 最大值$(2, k+1)$	初始轮次，以避免不适当后验分布。
$N_a(t)$	截至第 t轮选择臂 a 的次数。
$r_{a,n}$	由臂 a生成的第n个奖励。
$\mu(\theta)$	服从 Pa(θ) 的随机变量的期望值。
$\mu_a = \mu(\theta_a)$	臂 a的期望奖励。
Δ_a	臂 a的次优性差距。
$\delta_a = \frac{\Delta_a}{2}$适用于 $a \neq 1$	臂 a的次优性差距的一半。
$\delta_1 :=$ 最小值$_{a \neq 1}\, \delta_a$	定义为次优性差距的最小值。
$\mathcal{F}_t = (j(s), r_{j(s),N_{j(s)}})_{s=1}^{t-1}$	截至第 t轮的历史记录。
$\mathbb{P}_t[\cdot] := \mathbb{P}[\cdot\|\mathcal{F}_t]$	给定 $\mathcal{F}_t$的条件概率。
$g_a(c, \alpha)$	对于 $c \geq 1$，Pa$\left(\frac{\kappa_a}{c}, \alpha\right)$ 与 Pa(κ_a, α_a)之间的 KL 散度。
$h_a(c, \boldsymbol{\mu}) = h_a(c)$	满足 $\mu\left(\frac{\kappa_a}{c}, \alpha\right) \geq \mu_1$ 的 α 的上界，对于 $c \geq 1$。

表 3：概率分布和估计量的符号说明

符号	含义
$\mathrm{Pa}(\kappa,\ \alpha)$	尺度参数为 $\kappa>0$ 、形状参数为 $\alpha>0$ 的帕累托分布。
$f_{\mathrm{Pa}_{\kappa,\ \alpha}}(x)$	$\mathrm{Pa}(\kappa,\alpha)$ 的概率密度函数。
厄朗分布(s,β)	形状参数为 $s>0$ 、速率参数为 $\beta>0$ 的埃尔朗分布。
$f_{\text{误差}s,\ \beta}(x)$	厄朗分布(s,β) 的概率密度函数。
$F_{\text{埃尔}s,\ \beta}(x)$	厄朗分布$(s,\ \beta)$在 $x>0$ 处的累积分布函数。
$\text{逆伽马分布}(s,\ \beta)$	形状参数为 $s>0$ 、尺度参数为 $\beta>0$ 的逆伽马分布。
$F_n(x)$	自由度为 n 的卡方分布的累积分布函数。
$\Gamma(s)$	伽马函数。
$\gamma(s,\ x)$	下不完全伽马函数。
$\Gamma(s,\ x)$	上不完全伽马函数。
$\hat{\kappa}_a(n),\ \hat{\ }\alpha_a(n)$	臂的尺度参数和形状参数的最大似然估计量 a 在 n 次观测后，定义见（2）。
$\tilde{\kappa}_a(t)\ \ \tilde{\alpha}_a(t)$	第 t 轮从后验分布中采样的参数，见（8）和（7）。
$\bar{\alpha}_a(n)\ =\text{最大值}(\hat{\ }\alpha_a(n)\ n)$	形状参数的截断估计量。
$\alpha_{a,n}$	一个临时记号，可分别替换为 $\hat{\ }\alpha_a(n)$（STS）和 $\bar{\alpha}_a(n)$（STS-T）。
$\hat{\ }\mu_a(n)\ =\mu(\hat{\kappa}_a(n),\ \hat{\ }\alpha_a(n))$	观测 n 后由最大似然估计量计算出的平均奖励。
$\tilde{\mu}_a(t)\ =\mu(\tilde{\kappa}_a(t)\ \tilde{\alpha}_a(t))$	第 t轮时由采样参数 $\tilde{\kappa}_a(t)$ 和 $\tilde{\alpha}_a(t)$ 计算出的平均奖励。
$\theta_a=(\kappa_a,\ \alpha_a)$	臂 $a\in[K]$的真实参数元组。
$\hat{\theta}_{a,n}=(\hat{\kappa}_a(n),\hat{\ }\alpha_a(n))$	臂 a 经过 n 次观测后的最大似然估计量元组。
$\bar{\theta}_{a,\ n}=(\hat{\kappa}_a(n)\ \bar{\alpha}_a(n))$	臂 a 经过 n 次观测后的、采用截断过程的估计量元组。
$\theta_{a,\ n}=(\hat{\kappa}_a(n)\alpha_{a,\ n})$	一个可由两者替代的临时记法 $\hat{\theta}_{a,\ n}$ (STS) 和 $\bar{\theta}_{a,\ n}$ (STS-T)。

表 4：遗憾分析的符号说明

符号	含义
$D_{a,\ k}(\epsilon)$	一个函数对遗憾分析的主项有贡献，该主项定义于（42）。
$\mathcal{K}_{a,\ n}(\epsilon)$	在经过 n 次观测后的第 t 轮，κ 的最大似然估计接近其真实值的事件。
$\mathcal{A}_{a,\ n}(\epsilon)$	在经过 n 次观测后的第 t 轮，α 的最大似然估计接近其真实值的事件。
$\mathcal{E}_{a,\ n}(\epsilon)$	$\mathcal{K}_{a,\ n}(\epsilon)$与$\mathcal{A}_{a,\ n}(\epsilon)$的交集。
$\mathcal{M}_\epsilon(t)$	第 t轮时，最优臂的样本均值接近其真实平均奖励的事件。
$p_n(x\|\theta_{1,\ n})$	给定$\theta_{1,\ n}$，观测臂 1 n 次后，$\{\tilde{\mu}_1(t)\ \le\mu_1-x\}$ 的概率。
$G_k(x;\ n)$	式（20）中埃尔朗分布累积分布函数的另一种表达式。

表5：确定性常数的符号说明

符号	含义
ε_a	满足对任意 $\epsilon\le\varepsilon_a$，$\hat{\ }\mu_a(n)\ \in[\mu_a-\delta_a,\mu_a+\delta_a]$ on $\mathcal{E}_{a,n}(\epsilon)$ 的依赖于问题的常数。
$\epsilon_{a,l}(\epsilon),\epsilon_{a,l}(\epsilon)$	用于控制事件$\mathcal{A}_{a,n}(\epsilon)$ 下 $\hat{\ }\alpha_a(\epsilon)$ 偏差的常数。
$\rho_{\theta_1}(\epsilon)\ \bar{\rho}=\rho_{\theta_1}(\epsilon/2)$	与其真实形状参数 α_1 的差异，以满足 $\mu(\kappa_a,\ \alpha+\rho_{\mu_1}(\epsilon))\ge\mu_1-\frac{\epsilon}{2}$。
$C_1(\mu_1,\epsilon,n)=C_{1,n}$	一个小于1 在（20）中。
$C_2(\mu_1,\ \epsilon,\ k)$	$\mathcal{K}^c$ 下 $p_n(\epsilon\|\cdot)$ 的一致界 $_{1,n}(\epsilon)\cap\mathcal{A}_{1,\ n}(\epsilon/2)$。
$c_{\mu_1}(\epsilon)$	与 $\mathcal{O}(\epsilon^{-2})$相关的小常数。

B 引理1 的证明

引理1。 对于任意臂 $a \neq 1$，有

$$\mathrm{KL}_{\inf}(a) = \log\left(\alpha_a \frac{\mu_1 - \kappa_a}{\mu_1}\right) + \frac{1}{\alpha_a}\frac{\mu_1}{\mu_1 - \kappa_a} - 1.$$

证明。 回顾定义

$$\mathrm{KL}_{\inf}(a) = \mathrm{KL}_{\inf}(\mathrm{Pa}(\theta_a), \mathrm{Pa}(\theta)) := \inf_{\theta \in \Theta_a} \log\frac{\alpha_a}{\alpha} + \alpha \log\frac{\kappa_a}{\kappa} + \frac{\alpha}{\alpha_a} - 1,$$

其中 $\theta = (\kappa,\ \alpha)$ 以及 Θ_a 定义于（4）。

这里，我们考虑 Θ_a的划分，

$$\begin{aligned}
\Theta_a^{(1)} &= \{(\kappa, \alpha) \in (0, \kappa_a] \times (0, 1] : \mu(\kappa, \alpha) > \mu_1\} = (0, \kappa_a] \times (0, 1] \\
\Theta_a^{(2)} &= \left\{(\kappa, \alpha) \in (0, \kappa_a] \times (1, \infty) : \mu(\kappa, \alpha) = \frac{\kappa\alpha}{\alpha - 1} > \mu_1\right\},
\end{aligned} \tag{13}$$

其中 $\Theta_a^{(1)} \cup \Theta_a^{(2)} = \Theta_a$。 因此，有

$$\mathrm{KL}_{\inf}(a) = \min\left(\inf_{\theta \in \Theta_a^{(1)}} \log\frac{\alpha_a}{\alpha} + \alpha\log\frac{\kappa_a}{\kappa} + \frac{\alpha}{\alpha_a} - 1,\ \inf_{\theta \in \Theta_a^{(2)}} \log\frac{\alpha_a}{\alpha} + \alpha\log\frac{\kappa_a}{\kappa} + \frac{\alpha}{\alpha_a} - 1\right).$$

对于 $(\kappa,\ \alpha) \in \Theta_a^{(1)}$, $\mu(\kappa,\ \alpha) = \infty$ 无论 κ 如何都成立。因此，我们得到

$$\begin{aligned}
\inf_{\theta \in \Theta_a^{(1)}} \log\frac{\alpha_a}{\alpha} + \alpha\log\frac{\kappa_a}{\kappa} + \frac{\alpha}{\alpha_a} - 1 &= \inf_{\alpha \in (0,1]} \log\frac{\alpha_a}{\alpha} + \frac{\alpha}{\alpha_a} - 1 \\
&= \log\alpha_a + \frac{1}{\alpha_a} - 1,
\end{aligned}$$

其中，最后一个等式成立是因为 $\log x + \frac{1}{x} - 1$ 对于 $x \geq 1$ 是递增函数 .

令 $\frac{\kappa_a}{\kappa} = c \geq 1$ ，使从 $\mathrm{Pa}(\theta_a)$ 到 $\mathrm{Pa}(\kappa,\ \alpha)$ 的 KL 散度定义良好。根据其在$\Theta_a^{(2)}$式（13）中的定义，任意 $\theta = (\kappa,\ \alpha) \in \Theta_a^{(2)}$都满足 $\frac{\kappa\alpha}{\alpha - 1} \geq \mu_1$，即:

$$\frac{\kappa_a\alpha}{c(\alpha - 1)} \geq \mu_1 \Leftrightarrow \alpha \leq \frac{c\mu_1}{c\mu_1 - \kappa_a} =: h_a(c, \boldsymbol{\mu}) = h_a(c).$$

注意，以下结论成立:

$$h_a(1) = \frac{\mu_1}{\mu_1 - \kappa_a} \leq \frac{\mu_a}{\mu_a - \kappa_a} = \alpha_a$$

因为 $\frac{x}{x-y}$ is d关于 $x \geq y$递减。然后，我们可以重写 KL 散度的下确界 a s

$$\mathrm{KL}_{\inf}(a) = \min\left(\log\alpha_a + \frac{1}{\alpha_a} - 1, \inf_{c \geq 1} \inf_{\alpha \leq h_a(c)} g_a(\alpha, c)\right),$$

其中 $g_a(\alpha,\ c) := \log\frac{\alpha_a}{\alpha} + \alpha\log\ c + \frac{\alpha}{\alpha_a} - 1$ 满足

$$\frac{\partial g_a(\alpha, c)}{\partial\alpha} = \frac{1}{\alpha_a} + \log c - \frac{1}{\alpha}. \tag{14}$$

那么，当 $\alpha = \frac{\alpha_a}{1+\alpha_a \log c}$ 成立且 $\frac{\alpha_a}{1+\alpha_a \log c} < h_a(c)$ 时，可以得到内层下确界，其中 $g_a\left(\frac{\alpha_a}{1+\alpha_a}\log c,\ c\right) = \log(1+\alpha_a \log c)$。

令 $c_a^* \geq 1$ 为满足以下条件的确定性常数

$$h_a(c_a^*) = \frac{c_a^*\mu_1}{c_a^*\mu_1 - \kappa_a} = \frac{\alpha_a}{1+\alpha_a \log c_a^*} \Leftrightarrow (\mu_1\alpha_a)c_a^* \log c_a^* + (\mu_1 - \alpha_a\mu_1)c_a^* = -\alpha_a\kappa_a \tag{15}$$

从而 $h_a(c) \geq \frac{\alpha_a}{1+\alpha_a \log c}$ 对任意 $c \geq c_a^*$ 均成立。由于 $ax\log(x) + bx = -c$ 的解为 $\exp\left(W\left(-\frac{ce^{b/a}}{a}\right) - \frac{b}{a}\right)$ Lambert W 函数的主分支 $W(\cdot)$，因此可通过求解等式（c_a^*（15）得到

$$c_a^* = \exp\left(W\left(-\frac{\kappa_a}{\mu_1}e^{\frac{1}{\alpha_a}-1}\right) + 1 - \frac{1}{\alpha_a}\right). \tag{16}$$

注意，$\frac{\kappa_a}{\mu_1}e^{\frac{1}{\alpha_a}-1} \leq \frac{\kappa_a}{\mu_a}e^{\frac{1}{\alpha_a}-1} \leq \left(1-\frac{1}{\alpha_a}\right)e^{\left(1-\frac{1}{\alpha_a}\right)} \leq e^{-1}$ 成立，因此 c_a^* 为实数值。这里，我们考虑主分支以确保 $c_a^* \geq 1$，因为其他分支上的解 $W_{-1}(\cdot)$ 小于 1，不在我们的关注范围内。

令 $A_a = 1 - \frac{1}{\alpha_a}$，由于 $\alpha_a > 1$ 和 $B_a = \frac{\kappa_a}{\mu_1} \leq \frac{\kappa_a}{\mu_a} = \frac{\alpha_a - 1}{\alpha_a} = A_a$，它为正。因此，我们可以将 c_a^* 改写为

$$c_a^* = e^{A_a}e^{W(-B_ae^{-A_a})} = e^{A_a}e^{-A_a}\frac{-B_a}{W(-B_ae^{-A_a})}. \qquad \because e^{W(x)} = \frac{x}{W(x)}$$

由于 Lambert W 函数的主分支 $W(x)$ 在 $x \geq -\frac{1}{e}$ 上单调递增，因此有

$$0 > W(-B_ae^{-A_a}) \geq W(-B_ae^{-B_a}) = -B_a,$$

这意味着 $c_a^* \geq 1$。因此，g_a 的下确界可写为

$$\begin{aligned}\inf_{c\geq 1}\inf_{\alpha \leq h_a(c)} g_a(\alpha, c) &= \min\left(\inf_{c\in[1,c_a^*]} g_a(h_a(c), c),\ \inf_{c\geq c_a^*}\log(1+\alpha_a\log c)\right) \\ &= \min\left(\inf_{c\in[1,c_a^*]} g_a(h_a(c), c), \log(1+\alpha_a\log c_a^*)\right),\end{aligned}$$

其中，我们遵循如下约定：空集上的下确界定义为无穷大。代入 c_a^* 到（16）中，可得

$$\log(1+\alpha_a\log c_a^*) = \log\left(\alpha_a + W\left(-\frac{\kappa_a}{\mu_1}e^{\frac{1}{\alpha_a}-1}\right)\right).$$

考虑以下不等式：

$$\begin{aligned}\log\left(\alpha_a + W\left(-\frac{\kappa_a}{\mu_1}e^{\frac{1}{\alpha_a}-1}\right)\right) &\geq \log\left(\alpha_a + W\left(-\frac{\kappa_a}{\mu_a}e^{\frac{1}{\alpha_a}-1}\right)\right) \\ &= \log\left(\alpha_a + W\left(\frac{1-\alpha_a}{\alpha_a}e^{\frac{1}{\alpha_a}-1}\right)\right) \\ &= \log\left(\alpha_a + \frac{1}{\alpha_a} - 1\right),\end{aligned} \tag{17}$$

其中，第一个不等式成立，是因为 Lambert W 函数的主分支 $W(x)$ 关于 $x \in [-1/e, 0)$ 单调递增且为负。

接下来要找出 $\inf_{c\in[1,c_a^*]} g_a(h_a(c), c)$ 的闭式解。根据 $h_a(c) = \frac{c\mu_1}{c\mu_1-\kappa_a}$ 的定义，我们有 ~~$h_a'(c) = -\frac{\mu_1\kappa_a}{(c\mu_1-\kappa_a)^2}$~~ 且

$$\begin{aligned}
\frac{\partial g_a(h_a(c), c)}{\partial c} &= \frac{\partial}{\partial c}\left(\log\frac{\alpha_a}{h_a(c)} + h_a(c)\log c + \frac{h_a(c)}{\alpha_a} - 1\right) \qquad \because g_a(x,c) = \log\frac{\alpha_a}{x} + x\log c + \frac{x}{\alpha_a} - 1 \\
&= -\frac{h_a'(c)}{h_a(c)} + h_a'(c)\log c + \frac{h_a(c)}{c} + \frac{1}{\alpha_a}h_a'(c) \\
&= \frac{\kappa_a}{c(c\mu_1-\kappa_a)} - \frac{\mu_1\kappa_a}{(c\mu_1-\kappa_a)^2}\log c + \frac{\mu_1}{c\mu_1-\kappa_a} - \frac{\kappa_a\mu_1}{\alpha_a(c\mu_1-\kappa_a)^2} \\
&= \frac{\kappa_a}{c(c\mu_1-\kappa_a)} - \frac{\mu_1\kappa_a}{(c\mu_1-\kappa_a)^2}\log c + \mu_1\frac{c\mu_1-\kappa_a-\frac{\kappa_a}{\alpha_a}}{(c\mu_1-\kappa_a)^2}. \qquad (18)
\end{aligned}$$

由于（18）中的第一项在 $c \geq 1$ 和 $\mu_1 \geq \mu_a > \kappa_a$时为正，因此我们来考察最后两项在$c \in [1, c_a^*]$时的情况，

$$\begin{aligned}
-\frac{\mu_1\kappa_a}{(c\mu_1-\kappa_a)^2}\log c + \mu_1\frac{c\mu_1-\kappa_a-\frac{\kappa_a}{\alpha_a}}{(c\mu_1-\kappa_a)^2} &= \frac{\mu_1}{(c\mu_1-\kappa_a)^2}\left(c\mu_1-\kappa_a-\frac{\kappa_a}{\alpha_a}-\kappa_a\log c\right) \\
&= \frac{\mu_1}{(c\mu_1-\kappa_a)^2}\left(\mu_1-\kappa_a-\frac{\kappa_a}{\alpha_a}+(c-1)\mu_1-\kappa_a\log c\right) \\
&= \frac{\mu_1}{(c\mu_1-\kappa_a)^2}\left(\mu_1-\kappa_a-\frac{\kappa_a}{\alpha_a}+\mu_1\left(c-\frac{\kappa_a}{\mu_1}\log c-1\right)\right).
\end{aligned}$$

这里，

$$\mu_1-\kappa_a-\frac{\kappa_a}{\alpha_a} \geq \mu_a-\kappa_a-\frac{\kappa_a}{\alpha_a} = \frac{\kappa_a\alpha_a}{\alpha_a-1}-\kappa_a-\frac{\kappa_a}{\alpha_a} = \frac{\kappa_a}{\alpha_a(\alpha_a-1)} > 0,$$

并且 $c - \frac{\kappa_a}{\mu_1}\log c - 1$ 关于 c 递增，因此对于 $c \geq 1$，有 $c - \frac{\kappa_a}{\mu_1}\log c - 1 \geq 0$。因此，$\frac{\partial}{\partial c} g_a(h_a(c), c)$ 在 $c \geq 1$ 时为正，即 $g_a(h_a(c), c)$ 关于 $c \geq 1$ 是递增函数。

因此，对于 $c \in [1, c_a^*]$，我们有，

$$\inf_{c\in[1,c_a^*]} g_a(h_a(c), c) = g_a\left(h_a(1), 1\right) = g_a\left(\frac{\mu_1}{\mu_1-\kappa_a}, 1\right) = \log\left(\alpha_a\frac{\mu_1-\kappa_a}{\mu_1}\right) + \frac{1}{\alpha_a}\frac{\mu_1}{\mu_1-\kappa_a} - 1.$$

记作 $X_a = \alpha_a\frac{\mu_1-\kappa_a}{\mu_1} \in [1, \alpha_a)$，其中仅当 $\mu_a = \mu_1$ 时才会发生 $X_a = 1$。于是，对于 $\alpha_a > 1$，有

$$\log(X_a) + \frac{1}{X_a} - 1 \leq \log\alpha_a + \frac{1}{\alpha_a} - 1 \leq \log\left(\alpha_a + \frac{1}{\alpha_a} - 1\right) \leq \log(1 + \alpha_a\log c_a^*),$$

其中最后一个不等式来自（17）中的结果。因此，我们有

$$\begin{aligned}
\mathrm{KL}_{\inf}(a) &= \min\left(\log\alpha_a + \frac{1}{\alpha_a} - 1, \inf_{c\in[1,c_a^*]} g_a(h_a(c), c), \log(1+\alpha_a\log c_a^*)\right) \\
&= \log\left(\alpha_a\frac{\mu_1-\kappa_a}{\mu_1}\right) + \frac{1}{\alpha_a}\frac{\mu_1}{\mu_1-\kappa_a} - 1,
\end{aligned}$$

至此，证明结束。 □

C 定理2和4的引理证明2 和 4

本节给出主要结果所需引理的证明。

为避免冗余，当同一结果同时适用于 $\hat{\alpha}_a(n)$ 和$\bar{\alpha}_a(n)$ 时，我们使用临时记号 $\alpha_{a,n}$ 。使用 $\alpha_{a,n}$ 记号时，可根据所考虑的策略，将其替换为 $\hat{\alpha}_a(n)$ 或 $\bar{\alpha}_a(n)$。例如，有如下结论：

$$\alpha_{a,n} \leq 1 \Leftrightarrow \begin{cases} \hat{\alpha}_a(n) \leq 1 & \text{under STS policy,} \\ \bar{\alpha}_a(n) \leq 1 & \text{under STS-T policy.} \end{cases}$$

$\hat{}$ 类似地，在可以将记号 $\theta_{a,n} := (\hat{\kappa}_a(n), \alpha_{a,n})$ 替换为 $\theta_{a,n} = (\hat{\kappa}_a(n), \hat{\alpha}_a(n))^-$ 和 $\theta_{a,n} = (\hat{\kappa}_a(n)\ \bar{\alpha}_a(n))$ 时，对于任意臂 $a \in [K]$ 和 $n \in \mathbb{N}$，我们都使用该记号。基于 $\theta_{a,n}$ 记号，我们给出样本均值小于给定值的后验概率不等式，并在附录C.5中提供证明。

引理 9。对于任意臂 $a \in [K]$ 和固定的 $t \in \mathbb{N}$，令 $N_a(t) = n$。对于任意正的 $\xi \leq \frac{y}{y-\kappa_a}$ 和 $k \in \mathbb{Z}$，有

$$\mathbb{1}[\hat{\kappa}_a(n) \leq y]\mathbb{P}[\tilde{\mu}_a(t) \leq y|\theta_{a,n}] \leq \int_{\xi}^{\infty} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\mathrm{d}x + \left(\frac{y}{\mu((\kappa_a,\xi))}\right)^n \int_1^{\xi} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\mathrm{d}x,$$

其中 $f^{\mathrm{Er}}_{s,\beta}(\cdot)$ 表示形状参数为 $s \in \mathbb{N}$ 、速率参数为$\beta > 0$的埃尔朗分布的概率密度函数。

根据 $\theta_{1,n}$ 记号，我们将后验分布中经过 n 次试验后抽取的样本小于 $\mu_1 - x$ 的概率记为

$$p_n(x|\theta_{1,n}) := \mathbb{P}[\tilde{\mu}_1 \leq \mu_1 - x|\hat{\kappa}_1(n), \alpha_{1,n}]. \tag{19}$$

设 $K(\epsilon) = (\kappa_1 + \epsilon, \mu_1 - \epsilon)$ 是 $\mathbb{R}$ 上的一个开区间。下面的引理 10 给出了在$\theta_{1,n}$条件下 p_n 的上界。

引理 10。给定 $\epsilon > 0$，定义一个正的问题相关常数 $\rho = \rho_{\theta_1}(\epsilon) := 2^{\kappa_1\epsilon/(\mu_1-\epsilon/2-\kappa_1)(\mu_1-\kappa_1)}$。若 $n \geq \bar{n} = \max(2, k+1)$，则对于 $k \in \mathbb{Z}_{\geq 0}$

$$p_n(\epsilon|\theta_{1,n}) \leq \begin{cases} e^{-n}, & \text{if } \hat{\kappa}_1(n) \geq \mu_1 - \epsilon, \\ h(\mu_1, \epsilon, n), & \text{if } \hat{\kappa}_1(n) \in K(\epsilon), \alpha_{1,n} \leq \alpha_1 + \rho, \\ C_1(\mu_1, \epsilon, n)G_k(1/\alpha_{1,n}; n) + 1 - G_k(1/\alpha_{1,n}; n) & \text{if } \hat{\kappa}_1(n) \in K(\epsilon), \alpha_{1,n} \geq \alpha_1 + \rho, \end{cases}$$

其中

$$h(\mu_1, \epsilon, n) := e^{-n\frac{3\epsilon}{4\mu_1}}\left(1 - \frac{1}{2}e^{-nc_{\mu_1}(\epsilon)}\right) + \frac{1}{2}e^{-nc_{\mu_1}(\epsilon)}$$

$$C_1(\mu_1, \epsilon, n) := \left(\frac{\mu_1 - \epsilon}{\mu_1 - \epsilon/2}\right)^n \leq e^{-n\frac{\epsilon}{2\mu_1-\epsilon}} < 1.$$

$$G_k(x; n) := F^{\mathrm{Er}}_{n-k,nx}(\alpha_1 + \rho) \tag{20}$$

适用于 F^{Er}，定义见（*10*）。这里，$c_{\mu_1}(\epsilon) = \zeta - \log(1+\zeta) = \mathcal{O}(\epsilon^{-2})$ 和 $\zeta = \frac{\epsilon}{4\mu_1-2\epsilon} \in (0, 1)$ 是 μ_1 和 ϵ 的确定性常数。

注意，$\mu((\kappa_1, \alpha_1+\rho)) = \mu_1 - \epsilon/2$ 成立，并且存在一个问题相关常数 $C_2(\mu_1, \epsilon, k) < 1$，使得对于任意 $n \geq \bar{n} = \max(2, k+1)$ 和 $\epsilon > 0$

$$h(\mu_1, \epsilon, n) \leq 1 - C_2(\mu_1, \epsilon, k). \tag{21}$$

C.1 引理 5 的证明

首先陈述一个在证明中会用到的众所周知的事实。

事实 11. 当 $X \sim \text{Erlang}(n, \beta)$ 的率参数为 β 时，$\frac{1}{X}$ 服从形状参数为 $n \in \mathbb{N}$、尺度参数为 $\beta \in \mathbb{R}_+$ 的逆伽马分布，即 $\frac{1}{X} \sim \text{IG}(n, \beta)$。 $\square$

引理 5. 在 STS下，$k \in \mathbb{Z}_{\geq 2}$，

$$\sum_{t=\bar{n}K+1}^{T} \mathbb{E}\left[\mathbb{1}[j(t) \neq 1, \mathcal{K}^c_{1,N_1(t)}(\epsilon), \mathcal{M}^c_\epsilon(t)]\right] \leq \mathcal{O}(\epsilon^{-2}).$$

在 STS-T 下，$k \in \mathbb{Z}_{\geq 0}$，

$$\sum_{t=\bar{n}K+1}^{T} \mathbb{E}\left[\mathbb{1}[j(t) \neq 1, \mathcal{K}^c_{1,N_1(t)}(\epsilon), \mathcal{M}^c_\epsilon(t)]\right] \leq \mathcal{O}(\epsilon^{-m}),$$

其中 $m = \max(2, \ 3-k)$。

证明。考虑以下在 STS 和 STS-T 下均成立的分解：

$$\begin{aligned}
&\sum_{t=K\bar{n}+1}^{T} \mathbb{1}[j(t) \neq 1, \mathcal{K}^c_{1,N_1(t)}(\epsilon), \mathcal{M}^c_\epsilon(t)] \\
&\qquad = \sum_{n=\bar{n}}^{T} \sum_{t=K\bar{n}+1}^{T} \mathbb{1}\left[j(t) \neq 1, \mathcal{K}^c_{1,N_1(t)}(\epsilon), \mathcal{M}^c_\epsilon(t), N_1(t) = n\right] \\
&\qquad = \sum_{n=\bar{n}}^{T} \sum_{m=1}^{T} \mathbb{1}\left[m \leq \sum_{t=K\bar{n}+1}^{T} \mathbb{1}\left[j(t) \neq 1, \mathcal{M}^c_\epsilon(t), \mathcal{K}^c_{1,N_1(t)}(\epsilon), N_1(t) = n\right]\right].
\end{aligned}$$

注意到

$$m \leq \sum_{t=K\bar{n}+1}^{T} \mathbb{1}[j(t) \neq 1, \mathcal{K}^c_{1,N_1(t)}(\epsilon), \mathcal{M}^c_\epsilon(t), N_1(t) = n]$$

意味着 $\tilde{\mu}_1(t) \leq \mu_1 - \epsilon$ 在 $\{t$ 上连续发生了 m 次：$\mathcal{K}^c_{1,\ N_1(t)}(\epsilon)\ \mathcal{M}^c_\epsilon(t)\ N_1(t) = n\}$。因此，我们有

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=K\bar{n}+1}^{T} \mathbb{1}[j(t) \neq 1, \mathcal{K}^c_{1,N_1(t)}(\epsilon), \mathcal{M}^c_\epsilon(t)]\right] &\leq \mathbb{E}\left[\sum_{n=\bar{n}}^{T} \sum_{m=1}^{T} \mathbb{1}[\mathcal{K}^c_{1,n}(\epsilon)] p_n(\epsilon|\theta_{1,n})^m\right] \\
&\leq \sum_{n=\bar{n}}^{T} \mathbb{E}\left[\mathbb{1}\left[\mathcal{K}^c_{1,n}(\epsilon)\right] \frac{p_n(\epsilon|\theta_{1,n})}{1 - p_n(\epsilon|\theta_{1,n})}\right] \qquad (22)
\end{aligned}$$

对于 $p_n(\cdot|\cdot)$，其定义见（19）。从现在起，我们固定 $n \geq \bar{n}$，并为简洁起见省略 n 中$\hat{}\kappa_1(n)$、$\hat{}\alpha_1(n)$ 和 $\bar{\alpha}_1(n)$的参数。

在STS **下，** $k \in \mathbb{Z}_{\geq 2}$**：通过应用引理**10并结合（21），在STS 下， $k \in \mathbb{Z}_{\geq 0}$我们可以将（22）中的期望分解为

$$\mathbb{E}\left[\mathbb{1}\left[\mathcal{K}_{1,n}^{c}(\epsilon)\right]\frac{p_n(\epsilon|\hat{\theta}_{1,n})}{1-p_n(\epsilon|\hat{\theta}_{1,n})}\right] \leq \mathbb{P}[\hat{\kappa}_1 \geq \mu_1 - \epsilon]\frac{e^{-n}}{1-e^{-n}} + \mathbb{P}[\hat{\kappa}_1 \in K(\epsilon), \hat{\alpha}_1 < \alpha_1 + \rho]\frac{h(\mu_1,\epsilon,n)}{C_2(\mu_1,\epsilon,k)} + \underbrace{\mathbb{E}_{\hat{\theta}_{1,n}}\left[\frac{\mathbb{1}[\hat{\kappa}_1 \in K(\epsilon), \hat{\alpha}_1 > \alpha_1+\rho]}{G_k(1/\hat{\alpha}_1;n)(1-C_{1,n})}\right]}_{(*)}, \tag{23}$$

其中我们记作 $C_{1,n} = C_1(\mu_1,\epsilon,n)$。为简便起见，定义 $z := {}^{1}\hat{\alpha}_1$，其中 $z \backsim$ Erlang$(n-1, n\alpha_1)$ 由事实（11）可知成立，因为式（2）中 $\hat{}\alpha_1 \sim$ IG$(n-1, n\alpha_1)$。根据 $\hat{}\kappa$ 与 $\hat{}\alpha$ 的独立性，以及式（9）和式（2）中 z 与 $\hat{}\alpha$ 的分布，我们有

$$\begin{aligned}(*) &= \int_0^{\frac{1}{\alpha_1+\rho}} z^{n-2}e^{-n\alpha_1 z}\frac{(n\alpha_1)^{n-1}}{\Gamma(n-1)}\int_{\hat{\kappa}_1 \in K(\epsilon)}\frac{f^{\mathrm{Pa}}_{\kappa_1,n\alpha_1}(\hat{\kappa}_1)}{G_k(z;n)(1-C_{1,n})}\mathrm{d}\hat{\kappa}_1\mathrm{d}z \\ &= \mathbb{P}[\hat{\kappa}_1 \in K(\epsilon)]\int_0^{\frac{1}{\alpha_1+\rho}}\frac{z^{n-2}e^{-n\alpha_1 z}}{G_k(z;n)(1-C_{1,n})}\frac{(n\alpha_1)^{n-1}}{\Gamma(n-1)}\mathrm{d}z.\end{aligned}$$

代入式（10）中的累积分布函数，可得

$$\begin{aligned}G_k(z;n) &= F^{\mathrm{Er}}_{n-k,nz}(\alpha_1+\rho) \\ &= \frac{1}{\Gamma(n-k)}\int_0^{n(\alpha_1+\rho)z} e^{-t}t^{n-k-1}\mathrm{d}t \\ &\geq \frac{e^{-n(\alpha_1+\rho)z}}{\Gamma(n-k)}\int_0^{n(\alpha_1+\rho)z} t^{n-k-1}\mathrm{d}t \\ &= \frac{e^{-n(\alpha_1+\rho)z}}{\Gamma(n-k+1)}(n(\alpha_1+\rho)z)^{n-k}.\end{aligned} \tag{24}$$

因此，

$$\begin{aligned}\frac{(*)}{\mathbb{P}[\hat{\kappa} \in K(\epsilon)]} &\leq \int_0^{\frac{1}{\alpha_1+\rho}}\frac{\Gamma(n-k+1)}{(n(\alpha_1+\rho)z)^{n-k}(1-C_{1,n})}e^{n(\alpha_1+\rho)z}\frac{(n\alpha_1)^{n-1}}{\Gamma(n-1)}z^{n-2}e^{-n\alpha_1 z}\mathrm{d}z \\ &= \frac{\Gamma(n-k+1)}{\Gamma(n-1)(1-C_{1,n})}(\alpha_1+\rho)^{k-1}\left(\frac{\alpha_1}{\alpha_1+\rho}\right)^{n-1}n^{k-1}\int_0^{\frac{1}{\alpha_1+\rho}} z^{k-2}e^{n\rho z}\mathrm{d}z \\ &\leq \frac{\Gamma(n-k+1)}{\Gamma(n-1)(1-C_{1,n})}(\alpha_1+\rho)^{k-1}e^{-\frac{\rho}{\alpha_1+\rho}(n-1)}\frac{n^{k-1}}{(n\rho)^{k-2}}\int_0^{\frac{1}{\alpha_1+\rho}}(n\rho z)^{k-2}e^{n\rho z}\mathrm{d}z.\end{aligned} \tag{25}$$

令 $n\rho z = y$，我们可以将式（25）中的积分界定为

$$\begin{aligned}\frac{n^{k-1}}{(n\rho)^{k-2}}\int_0^{\frac{1}{\alpha_1+\rho}}(n\rho z)^{k-2}e^{n\rho z}\mathrm{d}z &= \rho^{-(k-1)}\int_0^{\frac{n\rho}{\alpha_1+\rho}} y^{k-2}e^{y}\mathrm{d}y \\ &\leq \rho^{-(k-1)}e^{n\frac{\rho}{\alpha_1+\rho}}\int_0^{\frac{n\rho}{\alpha_1+\rho}} y^{k-2}\mathrm{d}y \qquad (26) \\ &= \frac{e^{n\frac{\rho}{\alpha_1+\rho}}}{k-1}\left(\frac{n}{\alpha_1+\rho}\right)^{k-1}, \quad \text{if } k \in \mathbb{Z}_{\geq 2} \qquad (27)\end{aligned}$$

其中，式（27）仅在 $k \in \mathbb{Z}_{\geq 2}$ 时成立，因为式（26）中的积分在以下情况下发散： $k \in \mathbb{Z}_{\leq 1}$。将式（27）应用于式（25），可得 $k \in \mathbb{Z}_{\geq 2}$

$$
\begin{aligned}
(\divideontimes) &\leq \mathbb{P}[\hat{\kappa} \in K(\epsilon)] \frac{e^{\frac{\rho}{\alpha_1+\rho}}}{1-C_{1,n}} \frac{n^{k-1}}{k-1} \frac{\Gamma(n-k+1)}{\Gamma(n-1)} \\
&\leq \frac{e^{1-\frac{\epsilon\alpha_1}{\kappa+\epsilon}n}}{1-C_{1,n}} \frac{\Gamma(n-k+1)}{\Gamma(n-1)} \frac{n^{k-1}}{k-1} = \mathcal{O}(ne^{-n\epsilon}),
\end{aligned} \tag{28}
$$

其中 (28) 可由引理15 和 $\frac{\rho}{\alpha_1+\rho} < 1$得到。结合 (28) 与 (23) 及 (22)，我们得到对于 $k \in \mathbb{Z}_{\geq 2}$

$$
\begin{aligned}
\mathbb{E}\left[\sum_{t=K\bar{n}+1}^{T} \mathbb{1}[j(t) \neq 1, \mathcal{K}^c_{1,N_1(t)}(\epsilon), \mathcal{M}^c_\epsilon(t)]\right] &\leq \sum_{n=\bar{n}}^{T} \left(\frac{e^{-n}}{1-e^{-n}} + \frac{e^{-n\frac{3\epsilon}{4\mu_1}} + \frac{1}{2}e^{-nc_{\mu_1}(\epsilon)}}{C_2(\mu_1,\epsilon,k)} + (\divideontimes) \right) \\
&\leq \sum_{n=\bar{n}}^{T} \mathcal{O}(e^{-n}) + \mathcal{O}(e^{-n\epsilon}) + \mathcal{O}(e^{-n\epsilon^2}) + \mathcal{O}(ne^{-n\epsilon}) \\
&= \mathcal{O}(1) + \mathcal{O}(\epsilon^{-1}) + \mathcal{O}(\epsilon^{-2}) + \mathcal{O}(\epsilon^{-2}),
\end{aligned}
$$

由此即得在 STS 下且满足 $k \in \mathbb{Z}_{\geq 2}$ 时的证明。

在 S 下 汤普森采样-T 下 $k \in \mathbb{Z}_{\geq 0}$： 在 STS-T下，我们有以下不等式，而非 (23)：

$$
\begin{aligned}
\mathbb{E}\left[\mathbb{1}\left[\mathcal{K}^c_n(\epsilon)\right] \frac{p_n(\epsilon|\bar{\theta}_{1,n})}{1-p_n(\epsilon|\bar{\theta}_{1,n})}\right] \leq \mathbb{P}[\hat{\kappa}_1 \geq \mu_1 - \epsilon] \frac{e^{-n}}{1-e^{-n}} + \mathbb{P}[\hat{\kappa}_1 \in K(\epsilon), \bar{\alpha}_1 < \alpha_1 + \rho] \frac{h(\mu_1,\epsilon,n)}{C_2(\mu_1,\epsilon,k)} \\
+ \underbrace{\mathbb{E}_{\bar{\theta}_{1,n}}\left[\frac{\mathbb{1}[\hat{\kappa}_1 \in K(\epsilon), \bar{\alpha}_1 \in (\alpha_1+\rho, n]]}{G_k(1/\bar{\alpha}_1; n)(1-C_{1,n})}\right]}_{(\star)}.
\end{aligned} \tag{29}
$$

由 $\mathbb{1}[\bar{\alpha}_1(n) < n] = \mathbb{1}[\bar{\alpha}_1(n) = \hat{\alpha}_1(n)]$,可得

$$
(\star) = \mathbb{E}_{\hat{\theta}_{1,n}}\left[\frac{\mathbb{1}[\hat{\kappa}_1 \in K(\epsilon), \hat{\alpha}_1 \in (\alpha_1+\rho, n)]}{G_k(1/\hat{\alpha}_1; n)(1-C_{1,n})}\right] + \mathbb{E}_{\bar{\theta}_{1,n}}\left[\frac{\mathbb{1}[\hat{\kappa}_1 \in K(\epsilon), \bar{\alpha}_1 = n]}{G_k(1/\bar{\alpha}_1; n)(1-C_{1,n})}\right].
$$

由于 $\mathbb{1}[\bar{\alpha}_1(n) = n] = \mathbb{1}[\hat{\alpha}_1(n) \geq n]$成立，且 $\hat{\kappa}$与 $\hat{\alpha}$相互独立，因此对于 $z = {}^1\!/_{\hat{\alpha}_1}$ ~厄朗分布$(n-1, n\alpha_1)$

$$
\frac{(\star)}{\mathbb{P}[\hat{\kappa}_1 \in K(\epsilon)]} \leq \underbrace{\int_{\frac{1}{n}}^{\frac{1}{\alpha_1+\rho}} \frac{z^{n-2}e^{-n\alpha_1 z}}{G_k(z;n)(1-C_{1,n})} \frac{(n\alpha_1)^{n-1}}{\Gamma(n-1)} \mathrm{d}z}_{(\dagger)} + \underbrace{\frac{1}{G_k(1/n;n)(1-C_{1,n})} \mathbb{P}\left[\frac{1}{\hat{\alpha}_1} \leq \frac{1}{n}\right]}_{(\diamond)}, \tag{30}
$$

其中，可以对$(\divideontimes)$ 应用相同的技巧，推导出$(\dagger)$的一个上界。按照（25）和（26）中的相同步骤，可得

$$
\int_{\rho}^{\frac{n\rho}{\alpha+\rho}} y^{k-2} \mathrm{d}y \leq \begin{cases} \left(\frac{n\rho}{\alpha+\rho}\right)^{k-1}, & \text{if } k \in \mathbb{Z}_{\geq 2}, \\ \log\left(\frac{n}{\alpha+\rho}\right), & \text{if } k = 1, \\ \rho^{k-1}/(1-k), & \text{if } k \in \mathbb{Z}_{k\leq 0}, \end{cases}
$$

作为（26）中积分的对应项。按照（27）和（28）中的相同步骤，可得

$$(\dagger) \leq \begin{cases} \frac{\Gamma(n-k+1)}{\Gamma(n-1)} \frac{en^{k-1}}{k-1}, & \text{if } k \in \mathbb{Z}_{\geq 2}, \\ (n-1)\log\left(\frac{n}{\alpha+\rho}\right), & \text{if } k = 1, \\ \frac{\Gamma(n-k+1)}{\Gamma(n-1)} \frac{e}{(1-k)(\alpha+\rho)^{1-k}}, & \text{if } k \in \mathbb{Z}_{k\leq 0}. \end{cases} \tag{31}$$

注意，$(\diamond)$ 中的概率项与速率参数为 $n\alpha_1$、在 $\frac{1}{n}$ 处取值的 $\text{Erlang}(n-1, n\alpha_1)$ 分布的累积分布函数相同，因为 $\widehat{\alpha}_1 \sim \text{IG}(n-1, n\alpha_1)$ 来自（2）。因此，我们有

$$\begin{aligned} (\diamond) &= \frac{1}{1-C_{1,n}} \frac{1}{G_k(1/n;n)} \frac{\gamma(n-1,\alpha_1)}{\Gamma(n-1)} \\ &\leq \frac{e^{\alpha_1+\rho}}{1-C_{1,n}} \frac{\Gamma(n-k+1)}{(\alpha_1+\rho)^{n-k}} \frac{\gamma(n-1,\alpha_1)}{\Gamma(n-1)} \quad \text{by (24)} \\ &\leq \frac{e^{\alpha_1+\rho}}{1-C_{1,n}} \frac{\Gamma(n-k+1)}{\Gamma(n-1)} \frac{\alpha_1^{n-1}}{(\alpha_1+\rho)^{n-k}} \qquad (32) \\ &\leq \frac{e^{\alpha_1+\rho}}{1-C_{1,n}} \frac{\Gamma(n-k+1)}{\Gamma(n-1)} \frac{1}{(\alpha_1+\rho)^{1-k}} \\ &= \mathcal{O}(n^{2-k}), \qquad (33) \end{aligned}$$

其中（32）由 $\gamma(s,x) \leq x^s$ 对任意 $s \geq 1$ 以及 $x > 0$成立。因此，结合（31）和（33）与（30），以及 $\mathbb{P}[\hat{\kappa} \in K(\epsilon)] = \mathcal{O}(e^{-n\epsilon})$，可得

$$(\star) \leq \begin{cases} \mathcal{O}(ne^{-n\epsilon}), & \text{if } k \in \mathbb{Z}_{\geq 2} \\ \mathcal{O}(n\log(n)e^{-n\epsilon}), & \text{if } k = 1, \\ \mathcal{O}(n^{2-k}e^{-n\epsilon}), & \text{if } k \in \mathbb{Z}_{\leq 0}. \end{cases} \tag{34}$$

结合（34）与（29）和（22），对于 $k \in \mathbb{Z}_{\geq 0}$，可得

$$\begin{aligned} \mathbb{E}\left[\sum_{t=K\bar{n}+1}^{T} \mathbb{1}[j(t) \neq 1, \mathcal{K}^c_{1,N_1(t)}(\epsilon), \mathcal{M}^c_\epsilon(t)]\right] &\leq \sum_{n=\bar{n}}^{T} \left(\frac{e^{-n}}{1-e^{-n}} + \frac{e^{-n\frac{3\epsilon}{4\mu_1}} + \frac{1}{2}e^{-nc_{\mu_1}(\epsilon)}}{C_2(\mu,\epsilon,k)} + (\star)\right) \\ &\leq \sum_{n=\bar{n}}^{T} \Big(\mathcal{O}(e^{-n}) + \mathcal{O}(e^{-n\epsilon}) + \mathcal{O}\left(e^{-n\epsilon^2}\right) \\ &\qquad + \mathcal{O}\left(\psi(n,k)e^{-n\epsilon}\right)\Big) \\ &= \mathcal{O}(1) + \mathcal{O}(\epsilon^{-1}) + \mathcal{O}(\epsilon^{-2}) + \mathcal{O}\left(\epsilon^{-\max(2,3-k)}\right), \end{aligned}$$

其中

$$\psi(n,k) = n\mathbb{1}[k \geq 2] + n\log(n)\mathbb{1}[k = 1] + n^{2-k}\mathbb{1}[k \leq 0].$$

请注意，针对术语 $(\star)$ 的分析同样适用于满足 $k \in \mathbb{Z}_{<0}$ 的 STS-T。不过，与 $k \in \{0,1\}$ 的情况不同，满足 $k \in \mathbb{Z}_{<0}$ 的先验分布在引理 10 中，在事件 $\{\hat{\kappa}_1 \in K(\epsilon)\ \bar{\alpha}_1(n) \leq \alpha_1 + \rho\}$ 下还存在其他问题，此时上界变为常数 1 2。 □

C.2 引理 6 的证明

首先，我们陈述一个证明中会用到的众所周知的事实。

事实12. 当 $X \sim \text{Erlang}(n, \beta)$ 的率参数为 β 时， $2\beta X$ 服从自由度为 $2n$ 的卡方分布，即 $2\beta X \sim \chi^2_{2n}$。

引理 6. 在 STS 和 STS-T 下，对于 $k \in \mathbb{Z}$，对任意 $a \in [K]$，均有

$$\sum_{t=\bar{n}K+1}^{T} \mathbb{E}[\mathbb{1}[j(t)=a, \mathcal{M}_\epsilon(t), \mathcal{E}_{a,N_a(t)}(\epsilon)]] \le \max(0,k)+1+\frac{1}{\alpha_a \epsilon}\mathbb{1}[k>0]+\frac{\log T}{D_{a,k}(\epsilon)}.$$

其中 $D_{a,k}(\epsilon) > 0$ 是满足 $\lim_{\epsilon\to 0} D_{a,k}(\epsilon) = \mathrm{KL}_{\inf}(a)$ 的有限问题确定性常数。

证明。 根据采样规则，有

$$\sum_{t=\bar{n}K+1}^{T} \mathbb{P}\left[j(t)=a, \tilde{\mu}_1(t) \ge \mu_1-\epsilon, \mathcal{E}_{a,N_a(t)}(\epsilon)\right] \le \sum_{t=\bar{n}K+1}^{T} \mathbb{P}\left[j(t)=a, \tilde{\mu}_a(t) \ge \mu_1-\epsilon, \mathcal{E}_{a,N_a(t)}(\epsilon)\right].$$

固定一个时间索引 t，并记 $\mathbb{P}_t[\cdot] = \mathbb{P}[\cdot \mid \mathcal{F}_t]$ 和 $N_a(t) = n$。为简化符号说明，我们省略 $\tilde{\kappa}_a(t)$ $\tilde{\alpha}_a(t)$ 和 $\tilde{\mu}_a(t)$ 中的参数 t，以及 $\hat{\kappa}_a(n), \hat{\alpha}_a(n)$ $\bar{\alpha}_a(n)$ 中的参数 n。

由于根据其后验分布，$\tilde{\kappa}_a \in (0, \hat{\kappa}_a]$ 成立，因此若 $\tilde{\alpha}_a \ge \frac{\mu_1-\epsilon}{\mu_1-\epsilon-\hat{\kappa}_a}$ 成立，则对任意 $\tilde{\kappa}_a$，$\tilde{\mu}_a = \frac{\kappa_a \alpha_a}{\tilde{\alpha}_a - 1} \le \mu_1 - \epsilon$ 均成立。回顾一下，$f^{\mathrm{Er}}_{n-k,\alpha_a,n}(\cdot)$ 表示率参数为 α_a、n 的厄朗分布 $\left(n-k, \frac{n}{\alpha_{a,n}}\right)$ 的密度函数，它是 STS 和 STS-T 下 $\tilde{\alpha}$ 的边缘后验分布。根据式 (8) 中 $\tilde{\kappa}$ 的累积分布函数，若 $\hat{\kappa}_a < \mu_1 - \epsilon$，则

$$\begin{aligned}
\mathbb{P}_t\left[\tilde{\mu}_a \ge \mu_1-\epsilon\right] &= \mathbb{P}_t\left[\tilde{\alpha}_a \le 1\right] + \mathbb{P}_t\left[\tilde{\kappa}_a \ge \frac{\tilde{\alpha}_a - 1}{\tilde{\alpha}_a}(\mu_1-\epsilon) \cap \tilde{\alpha}_a \in \left(1, \frac{\mu_1-\epsilon}{\mu_1-\epsilon-\hat{\kappa}_a}\right)\right] \\
&= \int_0^1 f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\mathrm{d}x + \int_1^{\frac{\mu_1-\epsilon}{\mu_1-\epsilon-\hat{\kappa}}} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\mathbb{P}_t\left[\tilde{\kappa}_a \ge \frac{x-1}{x}(\mu_1-\epsilon)\right]\mathrm{d}x \\
&= \int_0^1 f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\mathrm{d}x + \int_1^{\frac{\mu_1-\epsilon}{\mu_1-\epsilon-\hat{\kappa}}} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\left(1-\left(\frac{x-1}{\hat{\kappa}x}(\mu_1-\epsilon)\right)^{nx}\right)\mathrm{d}x \\
&= \int_0^{\frac{\mu_1-\epsilon}{\mu_1-\epsilon-\hat{\kappa}}} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\mathrm{d}x - \int_1^{\frac{\mu_1-\epsilon}{\mu_1-\epsilon-\hat{\kappa}}} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\left(\frac{x-1}{\hat{\kappa}x}(\mu_1-\epsilon)\right)^{nx}\mathrm{d}x.
\end{aligned}$$

由于 $\frac{x}{x-y}$ 关于 $y < x$ 单调递增，且 $\hat{\kappa} \le \kappa + \epsilon$ 在 $\mathcal{E}$ 上成立，因此对于 $\mathcal{E}$

$$\frac{\mu_1-\epsilon}{\mu_1-\epsilon-\hat{\kappa}} \le \frac{\mu_1-\epsilon}{\mu_1-\epsilon-(\kappa+\epsilon)}.$$

Let

$$\eta_a(\epsilon) = \frac{\kappa_a(\Delta_a-\epsilon)-\epsilon\mu_a}{(\mu_a-\kappa_a)(\mu_1-\kappa_a-2\epsilon)} > 0 \tag{35}$$

是一个仅取决于模型和 ϵ 的确定性常数，并且满足

$$\begin{aligned}\alpha_a - \eta_a(\epsilon) &= \frac{\mu_a}{\mu_a - \kappa_a} - \frac{\kappa_a(\Delta_a - \epsilon) - \epsilon\mu_a}{(\mu_a - \kappa_a)(\mu_1 - \kappa_a - 2\epsilon)} \\ &= \frac{\mu_a\mu_1 - \kappa_a\mu_a - 2\epsilon\mu_a - \kappa_a(\mu_1 - \mu_a - \epsilon) + \epsilon\mu_a}{(\mu_a - \kappa_a)(\mu_1 - \kappa_a - 2\epsilon)} \\ &= \frac{\mu_1(\mu_a - \kappa_a) - \epsilon(\mu_a - \kappa_a)}{(\mu_a - \kappa_a)(\mu_1 - \kappa_a - 2\epsilon)} \\ &= \frac{\mu_1 - \epsilon}{\mu_1 - \kappa_a - 2\epsilon}.\end{aligned}$$

由于 $\eta_a(\epsilon) > 0$，因此对任意 $\epsilon \in (0, \varepsilon_a)$

$$\alpha_a - \eta_a(\epsilon) = \frac{\mu_1 - \epsilon}{\mu_1 - \kappa_a - 2\epsilon} \leq \frac{\mu_a}{\mu_a - \kappa_a} = \alpha_a.$$

注意到 $\frac{\mu}{\mu - \kappa} = \alpha$ 成立，且在固定 κ 的情况下，μ 变为 μ'，即 $\frac{\mu'}{\mu' - \kappa}$，这意味着形状参数 α' 应取何值才能满足 $\mu((\kappa, \alpha')) = \mu'$。例如，$\theta = (\kappa_a + \varepsilon_a, \alpha_a)$ 满足 $\mu(\theta) \leq \mu_a + \frac{\delta_a}{2}$。因此，如果 $\mu((\kappa_a + \varepsilon_a, \alpha)) = \mu_1 - \epsilon > \mu_a + \frac{\delta_a}{2}$，那么 α 应小于 α_a。因此，我们有

$$\begin{aligned}&\mathbb{1}[\mathcal{E}_{a,n}(\epsilon)]\mathbb{P}_t\left[\tilde{\mu}_a \geq \mu_1 - \epsilon\right] \\ &\quad\leq \mathbb{1}[\mathcal{E}_{a,n}(\epsilon)]\Bigg(\int_0^{\frac{\mu_1-\epsilon}{\mu_1-\epsilon-\hat{\kappa}}} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\mathrm{d}x \\ &\qquad\qquad - \int_1^{\frac{\mu_1-\epsilon}{\mu_1-\epsilon-\hat{\kappa}}} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\left(\frac{x-1}{\hat{\kappa}x}(\mu_1 - \epsilon)\right)^{nx}\mathrm{d}x\Bigg) \\ &\quad\leq \mathbb{1}[\mathcal{E}_{a,n}(\epsilon)]\int_0^{\frac{\mu_1-\epsilon}{\mu_1-\epsilon-\hat{\kappa}}} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\mathrm{d}x \qquad (36) \\ &\quad\leq \mathbb{1}[\mathcal{E}_{a,n}(\epsilon)]\int_0^{\alpha_a-\eta_a(\epsilon)} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\mathrm{d}x = \mathbb{1}[\mathcal{E}_{a,n}(\epsilon)]\mathbb{P}_t[\tilde{\alpha}_a(t) \leq \alpha_a - \eta_a(\epsilon)]. \qquad (37)\end{aligned}$$

因此，通过取期望并利用事实 12，我们有

$$\begin{aligned}\mathbb{P}\left[\tilde{\mu}_a(t) \geq \mu_1 - \epsilon, \mathcal{E}_{a,n}(\epsilon)\right] &\leq \mathbb{P}[\tilde{\alpha}_a \leq \alpha_a - \eta_a(\epsilon), \mathcal{E}_{a,n}(\epsilon)], \\ &= \mathbb{P}\left[Z \leq \frac{2n}{\alpha_{a,n}}\left(\alpha - \eta_a(\epsilon)\right), \mathcal{E}_{a,n}(\epsilon)\right] \qquad (38)\end{aligned}$$

其中 Z 是一个服从自由度为 $2(n-k)$ 的卡方分布的随机变量，即 $Z \sim \chi^2_{2n-2k}$。

C.2.1 在STS下

这里，我们首先考虑 STS 的情形，其中我们将 $\alpha_{a,n}$ 替换为 $\hat{}\alpha_a(n)$。

由于在 $\mathcal{E}_{a,\ n}(\epsilon)$ 上成立 $\hat{}\alpha_a \in [\alpha_a - \epsilon_{a,l}, \alpha_a + \epsilon_{a,u}]$，因此有

$$\frac{1}{\alpha_a} - \epsilon\left(1 + \frac{1}{\kappa_a}\right) = \frac{1}{\alpha_a + \epsilon_{a,u}} \leq \frac{1}{\hat{\alpha}_a} \leq \frac{1}{\alpha_a - \epsilon_{a,l}} = \frac{1}{\alpha_a} + \epsilon \qquad (39)$$

根据ϵ_a的定义，$l(\epsilon)$ 以及 ϵ_a，$u(\epsilon)$ 见式（12）。

将α_a替换为n，其中$\hat{}\alpha_a(n)$，并在（38）中应用（39），可得

$$\begin{aligned}\mathbb{P}\left[\tilde{\mu}_a(t) \geq \mu_1 - \epsilon, \mathcal{E}_{a,n}(\epsilon)\right] &\leq \mathbb{P}\left[Z \leq \frac{2n}{\hat{\alpha}_a}\left(\alpha_a - \eta_a(\epsilon)\right), \mathcal{E}_{a,n}(\epsilon)\right] \\ &\leq \mathbb{P}\left[Z \leq 2n\left(\frac{1}{\alpha_a} + \epsilon\right)\left(\alpha_a - \eta_a(\epsilon)\right)\right] \\ &= \mathbb{P}\left[Z \leq 2(n-k)\frac{n}{n-k}\left(\frac{1}{\alpha_a} + \epsilon\right)\left(\alpha_a - \eta_a(\epsilon)\right)\right]. \end{aligned} \tag{40}$$

含有 $k \in \mathbb{Z}_{\leq 0}$的先验分布。首先考虑情形 $k \in \mathbb{Z}_{\leq 0}$，此时有 $\binom{n}{n-k} \leq 1$。可以得到

$$\begin{aligned}\mathbb{P}\left[\tilde{\mu}_a(t) \geq \mu_1 - \epsilon, \mathcal{E}_{a,n}(\epsilon)\right] &\leq \mathbb{P}\left[Z \leq 2(n-k)\frac{n}{n-k}\left(\frac{1}{\alpha_a} + \epsilon\right)\left(\alpha_a - \eta_a(\epsilon)\right)\right] \\ &\leq \mathbb{P}\left[Z \leq 2(n-k)\left(\frac{1}{\alpha_a} + \epsilon\right)\left(\alpha_a - \eta_a(\epsilon)\right)\right].\end{aligned}$$

注意，在定理 ε_a 2 中，2 的定义被设定为满足 $\left(\frac{1}{\alpha} + \epsilon\right)\left(\alpha - \eta_a(\epsilon)\right) < 1$ 对任意 $\epsilon \leq \varepsilon_a$均成立。因此，我们可以应用引理17，其表明

$$\mathbb{P}\left[Z \leq 2(n-k)\left(1 - \frac{\eta_a(\epsilon)}{\alpha_a} + \epsilon(\alpha_a - \eta_a(\epsilon))\right)\right] \leq e^{-(n-k)D_{a,k}(\epsilon)}, \tag{41}$$

其中

$$\begin{aligned}D_{a,k}(\epsilon) := -\log\left(1 - \frac{\eta_a(\epsilon)}{\alpha_a} + (\max(0,k)+1)\epsilon(\alpha_a - \eta_a(\epsilon))\right) \\ - \frac{\eta_a(\epsilon)}{\alpha_a} + (\max(0,k)+1)\epsilon(\alpha_a - \eta_a(\epsilon))\end{aligned} \tag{42}$$

i是一个仅取决于模型参数、 ϵ和先验参数 k 的有限常数 .

对于任意 $n_a > 0$，将（41）应用于（38）可得

$$\begin{aligned}&\sum_{t=\bar{n}K+1}^{T} \mathbb{E}[\mathbb{1}[j(t) = a, \tilde{\mu}_1(t) \geq \mu_1 - \epsilon, \mathcal{E}_{a,N_a(t)}(\epsilon)]] \\ &\qquad \leq \sum_{t=\bar{n}K+1}^{T} \mathbb{P}[j(t) = a, \tilde{\mu}_a(t) \geq \mu_1 - \epsilon, \mathcal{E}_{a,n}(\epsilon)] \\ &\qquad \leq n_a + \sum_{t=\bar{n}K+1}^{T} \mathbb{P}[\tilde{\mu}_a(t) \geq \mu_1 - \epsilon, \mathcal{E}_{a,N_a(t)}(\epsilon), N_a(t) \geq n_a] \\ &\qquad \leq n_a + \sum_{t=\bar{n}K+1}^{T} e^{-(n_a-k)D_{a,k}(\epsilon)} \\ &\qquad \leq n_a + \sum_{t=\bar{n}K+1}^{T} e^{-n_a D_{a,k}(\epsilon)} = n_a + Te^{-n_a D_{a,k}(\epsilon)}.\end{aligned}$$

令 $n_a = \log^{T}_{D_{a,k,\epsilon,()}}$ 由此完成了先验分布为 $k \in \mathbb{Z}_{\leq 0}$ 的情形。

先验分布为 $k \in \mathbb{Z}_{>0}$ 接下来，考虑情形 $k \in \mathbb{Z}_{>0}$。回顾一下，当$k > 0$时，我们首先会对每个臂进行 $k+1$ 次试验，这意味着 $n - k > 0$。对于 $n \geq \frac{1}{\alpha\epsilon} + k + 1$，有

$$\frac{n}{n-k}\left(\frac{1}{\alpha} + \epsilon\right) \leq \frac{1}{\alpha} + (k+1)\epsilon. \tag{43}$$

将（43）应用于（38），可得对于 $n \geq \frac{1}{\alpha\epsilon} + k + 1$,

$$\mathbb{P}[\tilde{\alpha}_a \leq \alpha_a - \eta_a(\epsilon), \mathcal{E}_{a,N_a(t)}(\epsilon)] \leq \mathbb{P}\left[Z \leq 2(n-k)\left(1 - \frac{\eta_a(\epsilon)}{\alpha_a} + (k+1)\epsilon(\alpha_a - \eta_a(\epsilon))\right)\right].$$

类似地，应用引理 17，可知对于 $n \geq \frac{1}{\alpha_a\epsilon} + k + 1$ —

$$\mathbb{P}[\tilde{\alpha}_a \leq \alpha_a - \eta_a(\epsilon), \mathcal{E}_{a,N_a(t)}(\epsilon)] \leq e^{-(n-k)D_{a,k}(\epsilon)}, \tag{44}$$

其中 $D_{a,\ k}$（ϵ）定义见（42）。

当 $k \in \mathbb{Z}_{>0}$ 时，令 $n_a \geq \frac{1}{\alpha_a\epsilon} + k + 1$ 为任意值。再次将（44）应用于（38），可得

$$\begin{aligned}
&\sum_{t=\bar{n}K+1}^{T} \mathbb{E}[\mathbb{1}[j(t) = a, \tilde{\mu}_1(t) \geq \mu_1 - \epsilon, \mathcal{E}_{a,N_a(t)}(\epsilon)]] \\
&\qquad \leq \sum_{t=\bar{n}K+1}^{T} \mathbb{P}[j(t) = a, \tilde{\mu}_a(t) \geq \mu_1 - \epsilon, \mathcal{E}_{a,n}(\epsilon)] \\
&\qquad \leq n_a + \sum_{t=\bar{n}K+1}^{T} \mathbb{P}[\tilde{\mu}_a(t) \geq \mu_1 - \epsilon, \mathcal{E}_{a,N_a(t)}(\epsilon), N_a(t) \geq n_a] \\
&\qquad \leq n_a + \sum_{t=\bar{n}K+1}^{T} e^{-(n_a-k)D_{a,k}(\epsilon)} = n_a + Te^{-(n_a-k)D_{a,k}(\epsilon)}.
\end{aligned}$$

令 $n_a = k + 1 + \frac{1}{\alpha_a\epsilon} + \text{对数}D_{a\ k}{}^{T}{}_{\epsilon,()}$ 即可完成对具有 $k > 0$ 的先验分布情形的证明。

C.2.2 在STS-T下

这里，我们考虑 STS-T 的情形，其中将 $\alpha_{a,n}$ 替换为 $\bar{\alpha}_a(n) = \min(\hat{\alpha}_a(n)\ n)$。根据 $\bar{\alpha}_a(n)$ 的定义，对于 $\epsilon \leq \varepsilon_a$，有

$$\forall n \geq \alpha_a + 1 : \mathbb{1}[\bar{\alpha}_a(n) = \hat{\alpha}_a(n), \mathcal{A}_{a,n}(\epsilon)] = 1.$$

因此，对于 $n \geq \alpha_a + 1$，可以直接将 STS 上的分析应用于 STS-T。

考虑在条件 $\bar{\alpha}_a(n) = n < \alpha_a + 1$ 成立的情形$\mathcal{A}_{a,\ n}(\epsilon)$。将$\alpha_{a,\ n}$ 替换为 n，在（38）中，并按照（38）和（41）中的相同步骤，可得对任意 $k \in \mathbb{Z}$,

$$\begin{aligned}
\mathbb{P}\left[\tilde{\mu}_a(t) \geq \mu_1 - \epsilon, \mathcal{E}_{a,n}(\epsilon)\right] &\leq \mathbb{P}\left[Z \leq \frac{2n}{n}\left(\alpha_a - \eta_a(\epsilon)\right), \mathcal{E}_{a,n}(\epsilon)\right] \\
&\leq \mathbb{P}\left[Z \leq 2(n-k)\frac{1}{n-k}\left(\frac{1}{\alpha_a} + \epsilon\right)\left(\alpha_a - \eta_a(\epsilon)\right), \mathcal{E}_{a,n}(\epsilon)\right] \\
&\leq \mathbb{P}\left[Z \leq 2(n-k)\left(\frac{1}{\alpha_a} + \epsilon\right)\left(\alpha_a - \eta_a(\epsilon)\right), \mathcal{E}_{a,n}(\epsilon)\right] \\
&\leq e^{-(n-k)D_{a,k}(\epsilon)},
\end{aligned}$$

其中 $D_{a,\ k}(\epsilon)$定义见（42）。因此，根据第 C.2.1 节中的分析，可得相同的结果。

C.2.3 $D_{a,\ k}(\epsilon)$的渐近行为

最后，我们证明 $\lim_{\epsilon\to 0} D_{a,\ k}(\epsilon) = \mathrm{KL}_{\inf}(a)$。根据它们对 $\eta_a(\epsilon)$ 在 (35) 中以及 $\Delta_a = \mu_1 - \mu_a$的定义，我们有

$$\begin{aligned}\lim_{\epsilon\to 0}\eta_a(\epsilon) &= \lim_{\epsilon\to 0}\frac{\kappa_a(\Delta_a-\epsilon)-\epsilon\mu_a}{(\mu_a-\kappa_a)(\mu_1-\kappa_a-2\epsilon)}\\ &= \frac{\kappa_a\Delta_a}{(\mu_a-\kappa_a)(\mu_1-\kappa_a)} = \frac{\kappa_a(\mu_1-\mu_a)}{(\mu_a-\kappa_a)(\mu_1-\kappa_a)} = (\alpha_a-1)\frac{\mu_1-\mu_a}{\mu_1-\kappa_a}.\end{aligned}$$

于是，有

$$\begin{aligned}\lim_{\epsilon\to 0} 1-\frac{\eta_a(\epsilon)}{\alpha_a} &= 1-\left(\frac{\alpha_a-1}{\alpha_a}\frac{\mu_1-\mu_a}{\mu_1-\kappa_a}\right)\\ &= \frac{\alpha_a(\mu_1-\kappa_a)-(\alpha_a-1)(\mu_1-\mu_a)}{\alpha_a(\mu_1-\kappa_a)}\\ &= \frac{\alpha_a(\mu_a-\kappa_a)+\mu_1-\mu_a}{\alpha_a(\mu_1-\kappa_a)}\\ &= \frac{\mu_a}{\mu_a-\kappa_a}\frac{(\mu_a-\kappa_a)}{\alpha_a(\mu_1-\kappa_a)}+\frac{\mu_1-\mu_a}{\alpha_a(\mu_1-\kappa_a)} \qquad \because \alpha_a = \frac{\mu_a}{\mu_a-\kappa_a}\\ &= \frac{\mu_1}{\alpha_a(\mu_1-\kappa_a)} =: \frac{1}{X_a},\end{aligned}$$

因此，根据 $D_{a,\ k}(\epsilon)$ 在 (42) 中的定义

$$\begin{aligned}\lim_{\epsilon\to 0} D_{a,k}(\epsilon) = \lim_{\epsilon\to 0}\Bigg[&-\log\left(1-\frac{\eta_a(\epsilon)}{\alpha_a}+(\max(0,k)+1)(\alpha_a-\eta_a(\epsilon))\epsilon\right)\\ &\qquad -\frac{\eta_a(\epsilon)}{\alpha_a}+(\max(0,k)+1)(\alpha_a-\eta_a(\epsilon))\epsilon\Bigg]\\ &= -\log\left(1-\lim_{\epsilon\to 0}\frac{\eta_a(\epsilon)}{\alpha_a}\right)-\lim_{\epsilon\to 0}\frac{\eta_a(\epsilon)}{\alpha_a}\\ &= -\log\frac{1}{X_a}+\frac{1}{X_a}-1 = \log X_a+\frac{1}{X_a}-1\\ &= \log\left(\alpha_a\frac{\mu_1-\kappa_a}{\mu_1}\right)+\frac{\mu_1}{\alpha_a(\mu_1-\kappa_a)}-1 = \mathrm{KL}_{\inf}(a),\end{aligned}$$

其中，最后一个等式由引理1得出。 □

C.3 引理7的证明

首先，我们陈述两个证明中用到的众所周知的事实。

事实 13. 当 $X\sim\mathrm{Pa}(\kappa,\ \alpha)$ 的尺度参数为 $\kappa\in\mathbb{R}$、速率参数为 $\alpha\in\mathbb{R}_+$ 时，$\log\left(\frac{X}{\kappa}\right)$服从速率参数为 α 的指数分布，即 $\log\left(\frac{X}{\kappa}\right)\sim\mathrm{Exp}(\alpha)$。 —

事实 14. 设 X_1，……， X_n 是独立同分布的随机变量，服从率参数为 α 的指数分布，即对任意 $i \in [n]$， X_i i.i.d.$\sim$ Exp(α)。那么，它们的和服从形状参数为 $n \in \mathbb{N}$、速率参数为 α 的厄朗分布，即 $\sum_{i=1}^{n} X_i \sim$ Erlang(n， α)。

引理 7。 在STS和STS-T 下，且 $k \in \mathbb{Z}_{\geq 0}$,

$$\sum_{t=\bar{n}K+1}^{T} \mathbb{E}\left[\mathbb{1}[j(t) \neq 1, \mathcal{K}_{1,N_1(t)}(\epsilon), \mathcal{M}_\epsilon^c(t)]\right] \leq \mathcal{O}(\epsilon^{-1}).$$

证明。 考虑尺度参数已知为 κ 、属于一维指数族的帕累托分布，观察到 $n = N_1(t)$ 个奖励后，形状参数 $\alpha^{\text{one}} > 0$ 的后验分布为 $k \in \mathbb{Z}$

$$\alpha^{\text{one}} \mid \mathcal{F}_t \sim \text{Erlang}\left(n - k + 1, X_n\right), \tag{45}$$

其中 $X_n = \sum_{s=1}^{n} \log(r_{1,s}) - n\log(\kappa_1)$。注意， $X_n \sim$ 厄朗分布(n, α_1) 由事实 13 和 14可知。令 $\tilde{\alpha}_1^{\text{one}}$ 为从 (45) 中的后验分布抽取的样本。那么，对于一维帕累托老虎机，由 (10) 可得

$$\mathbb{P}[\tilde{\mu}_1(t) \leq \mu_1 - \epsilon | \mathcal{F}_t] = \mathbb{P}\left[\tilde{\alpha}_1^{\text{one}} \geq \beta \,|\, \mathcal{F}_t\right] = \frac{\Gamma\left(n - k + 1, \beta X_n\right)}{\Gamma(n - k + 1)},$$

其中我们记 $\beta = \frac{\mu_1 - \epsilon}{\mu_1 - \epsilon - \kappa_1}$满足 $\mu(\kappa_1, \beta) = \mu_1 - \epsilon$。因此，引理 21 可写为

$$\begin{aligned}
\sum_{t=1}^{T} \mathbb{E}\left[\mathbb{1}\left[j(t) \neq 1, \mathcal{M}_\epsilon^c(t)\right]\right] &= \sum_{t=1}^{T}\sum_{n=1}^{T} \mathbb{E}\left[\mathbb{1}\left[j(t) \neq 1, \mathcal{M}_\epsilon^c(t), N_1(t) = n\right]\right] \\
&= \sum_{t=1}^{T}\sum_{n=1}^{T} \mathbb{E}\left[\mathbb{P}\left[j(t) \neq 1, \mathcal{M}_\epsilon^c(t), N_1(t) = n \,|\, \mathcal{F}_t\right]\right] \\
&= \sum_{t=1}^{T}\sum_{n=1}^{T} \int_0^\infty \frac{\Gamma(n+1, \beta x)}{\Gamma(n+1)} \frac{\alpha_1^n}{\Gamma(n)} x^{n-1} e^{-\alpha_1 x} \mathrm{d}x \leq \mathcal{O}(\epsilon^{-1}),
\end{aligned}$$

其中，我们在最后一个等式中代入了埃尔朗分布的密度函数。

另一方面，对于尺度参数未知的双参数帕累托老虎机，根据全期望定律，有

$$\begin{aligned}
\mathbb{E}[\mathbb{1}[j(t) \neq 1, \mathcal{K}_{1,N_1(t)}(\epsilon), \mathcal{M}_\epsilon^c(t)]] &= \mathbb{E}_{\hat{\kappa}_1, \hat{\alpha}_1}\left[\mathbb{P}[j(t) \neq 1, \mathcal{K}_{1,N_1(t)}(\epsilon), \mathcal{M}_\epsilon^c(t) | \mathcal{F}_t]\right] \\
&= \mathbb{E}_{\hat{\kappa}_1, \hat{\alpha}_1}\left[\mathbb{1}[\mathcal{K}_{1,N_1(t)}(\epsilon)] \mathbb{P}[j(t) \neq 1, \mathcal{M}_\epsilon^c(t) | \mathcal{F}_t]\right],
\end{aligned}$$

其中，最后一个等式成立是因为 $\mathcal{K}$ 由历史记录 $\mathcal{F}_t$确定。

由引理9 可知，对于任意 $y = \mu_1 - \epsilon$， $\xi \leq \frac{\mu_1 - \epsilon}{\mu_1 - \epsilon - \kappa_1} = \beta$ 均成立

$$\begin{aligned}
&\mathbb{1}[\mathcal{K}_{1,n}(\epsilon)] \mathbb{P}[\tilde{\mu}_1(t) \leq \mu_1 - \epsilon | \mathcal{F}_t] \\
&\quad \leq \mathbb{1}[\mathcal{K}_{1,n}(\epsilon)] \left(\left(\frac{\mu_1 - \epsilon}{\mu((\kappa_1, \xi))} \right)^n \int_1^\xi f^{\text{Er}}_{n-k, \frac{n}{\alpha_{1,n}}}(x) \mathrm{d}x + \int_\xi^\infty f^{\text{Er}}_{n-k, \frac{n}{\alpha_{1,n}}}(x) \mathrm{d}x \right) \\
&\quad \leq \mathbb{1}[\mathcal{K}_{1,n}(\epsilon)] \left(\left(\frac{\mu_1 - \epsilon}{\mu((\kappa_1, \xi))} \right)^n \left(1 - \frac{\Gamma\left(n - k, \frac{n}{\alpha_{1,n}} \xi\right)}{\Gamma(n-k)} \right) + \frac{\Gamma\left(n - k, \frac{n}{\alpha_{1,n}} \xi\right)}{\Gamma(n-k)} \right)
\end{aligned} \tag{46}$$

这是 1 和 $\left(\begin{smallmatrix}\mu_1-\epsilon\\ \mu\ \ \kappa_1\end{smallmatrix}\xi((\,,\,))\right)^n$ 的凸组合。因此，(46) 式右侧随 $\Gamma\left(n-k,\ \frac{n}{\alpha_1 n}\xi\right)$，( ) 的增大而增大。根据 $\Gamma(n,x)$ 的定义，对于任意正数 $y>0$，都有 $\Gamma(n,x)\geq\Gamma(n,x+y)$，且 $\Gamma(n+1,x)=n\Gamma(n,x)+x^n e^{-x}$。由于对任意 $n\in\mathbb{N}$，$\hat{\alpha}_1(n)\leq\bar{\alpha}_1(n)$ 均成立，因此对于 $k\in\mathbb{Z}_{\geq 0}$，有

$$\frac{\Gamma\left(n-k,\frac{n}{\bar{\alpha}_1(n)}\xi\right)}{\Gamma(n-k)}\leq\frac{\Gamma\left(n-k,\frac{n}{\hat{\alpha}_1(n)}\xi\right)}{\Gamma(n-k)}\leq\frac{\Gamma\left(n,\frac{n}{\hat{\alpha}_1(n)}\xi\right)}{\Gamma(n)}.$$

记 $Y_n:=\hat{\alpha}_1(n)\sum_{i=1}^n\log(r_{1,\,s})-n\log(\hat{\kappa}_1(n))$，其服从形状参数为 $n-1$、速率参数为 α_1 [Malik 的厄朗分布， 1970]。对 $\hat{\kappa}_1(n)$ 取期望可得，对任意 $\xi\leq\beta$，有

$$\begin{aligned}
&\mathbb{E}_{\hat{\kappa}_1}[\mathbb{1}[\mathcal{K}_{1,n}(\epsilon)]\mathbb{P}[\tilde{\mu}_1(t)\leq\mu_1-\epsilon|\mathcal{F}_t]]\\
&\qquad\leq\int_{\kappa_1}^{\kappa_1+\epsilon}\left(\left(\frac{\mu_1-\epsilon}{\mu((\kappa_1,\xi))}\right)^n\left(1-\frac{\Gamma(n,\xi Y_n)}{\Gamma(n)}\right)+\frac{\Gamma(n,\xi Y_n)}{\Gamma(n)}\right)\mathbb{P}[\hat{\kappa}_1(n)=x]\mathrm{d}x\\
&\qquad=\mathbb{P}[\mathcal{K}_{1,n}(\epsilon)]\left(\left(\frac{\mu_1-\epsilon}{\mu((\kappa_1,\xi))}\right)^n\left(1-\frac{\Gamma(n,\xi Y_n)}{\Gamma(n)}\right)+\frac{\Gamma(n,\xi Y_n)}{\Gamma(n)}\right)\\
&\qquad=\left(1-\left(\frac{\kappa_1}{\kappa_1+\epsilon}\right)^{n\alpha_1}\right)\left(\left(\frac{\mu_1-\epsilon}{\mu((\kappa_1,\xi))}\right)^n\left(1-\frac{\Gamma(n,\xi Y_n)}{\Gamma(n)}\right)+\frac{\Gamma(n,\xi Y_n)}{\Gamma(n)}\right),
\end{aligned}$$

其中，在最后一个等式中，我们使用了 $\hat{\kappa}_1(n)\sim\mathrm{Pa}(\kappa_1,\ n\alpha_1)$ 式（2）。

因此，在双参数帕累托分布下，对于任意 $\xi\leq\beta$，在满足 $k\in\mathbb{Z}_{\geq 0}$ 的情况下，STS 和 STS-T 均有如下结论：

$$\begin{aligned}
&\mathbb{E}_{\hat{\kappa}_1,\hat{\alpha}_1}[\mathbb{1}[\mathcal{K}_{1,n}(\epsilon)]\mathbb{P}[\tilde{\mu}_1(t)\leq\mu_1-\epsilon|\mathcal{F}_t]]\\
&\leq\left(1-\left(\frac{\kappa_1}{\kappa_1+\epsilon}\right)^{n\alpha_1}\right)\int_0^\infty\left(\left(\frac{\mu_1-\epsilon}{\mu((\kappa_1,\xi))}\right)^n\left(1-\frac{\Gamma(n,\xi y)}{\Gamma(n)}\right)+\frac{\Gamma(n,\xi y)}{\Gamma(n)}\right)\frac{\alpha_1^{n-1}}{\Gamma(n-1)}y^{n-2}e^{-\alpha_1 y}\mathrm{d}y.
\end{aligned}$$

因此，引理 21 通过仔细选择 $n\in\mathbb{N}$，得出对任意 $\xi\leq\beta$ 满足

$$\begin{aligned}
&\left(1-\left(\frac{\kappa_1}{\kappa_1+\epsilon}\right)^{n\alpha_1}\right)\int_0^\infty\left(\left(\frac{\mu_1-\epsilon}{\mu((\kappa_1,\xi))}\right)^n\left(1-\frac{\Gamma(n,\xi y)}{\Gamma(n)}\right)+\frac{\Gamma(n,\xi y)}{\Gamma(n)}\right)\frac{\alpha_1^{n-1}}{\Gamma(n-1)}y^{n-2}e^{-\alpha_1 y}\mathrm{d}y\\
&\qquad\leq\int_0^\infty\frac{\Gamma(n+1,\beta y)}{\Gamma(n+1)}\frac{\alpha_1^n}{\Gamma(n)}y^{n-1}e^{-\alpha_1 y}\mathrm{d}y.
\end{aligned}$$

注意，当我们考虑满足 $k=-1$ 的 STS 时，需要找到满足以下条件的 $\xi'\leq\beta$：

$$\begin{aligned}
&\left(1-\left(\frac{\kappa_1}{\kappa_1+\epsilon}\right)^{n\alpha_1}\right)\\
&\quad\times\int_0^\infty\left(\left(\frac{\mu_1-\epsilon}{\mu((\kappa_1,\xi'))}\right)^n\left(1-\frac{\Gamma(n+1,\xi' y)}{\Gamma(n+1)}\right)+\frac{\Gamma(n+1,\xi' y)}{\Gamma(n+1)}\right)\frac{\alpha_1^{n-1}}{\Gamma(n-1)}y^{n-2}e^{-\alpha_1 y}\mathrm{d}y\\
&\qquad\leq\int_0^\infty\frac{\Gamma(n+1,\beta y)}{\Gamma(n+1)}\frac{\alpha_1^n}{\Gamma(n)}y^{n-1}e^{-\alpha_1 y}\mathrm{d}y.
\end{aligned}$$

由任意正的 $\Gamma(n,x)\geq\Gamma(n,x+y)$ $x,y>0$ 和 $\xi'\leq\beta$ 可知，对于任意 $x>0$，有

$$\frac{\Gamma(n+1,\xi' x)}{\Gamma(n+1)}\geq\frac{\Gamma(n+1,\beta x)}{\Gamma(n+1)}.$$

因此，对于 $k\in\mathbb{Z}_{\leq -1}$，我们可能无法应用Korda 等人 [2013]的结果。 □

C.4 引理 8 的证明

我们首先陈述关于事件 $\mathcal{K}$ 和 $\mathcal{A}$的两个引理。

Le理15。 对于任意算法和 $a \in [K]$，对所有 $\epsilon > 0$、 $t > 0$ 和 $n \in \mathbb{N}$，均有

$$\mathbb{P}\left[\mathcal{K}^c_{a,N_a(t)}(\epsilon), N_a(t) = n\right] \leq \exp\left(-\frac{\alpha_a \epsilon}{\kappa_a + \epsilon} n\right).$$

引理 16。 对于任意算法以及任意 $a \in [K]$，对所有 $\epsilon \in \left(0, {}_{\alpha_a}{}^{\kappa_a(\kappa_a+1)}\right)$ 和 $t > 0$，以及$n \geq \bar{n}$，均有

$$\mathbb{P}\left[\mathcal{A}^c_{a,N_a(t)}(\epsilon), \mathcal{K}_{a,N_a(t)}(\epsilon), N_a(t) = n\right] \leq 2\exp\left(-\frac{\alpha_a^2 \epsilon^2}{4} n\right),$$

引理 8。 在 STS 和 STS-T 下，对于任意 $k \in \mathbb{Z}$，均有 $a \neq 1$

$$\sum_{t=\bar{n}K+1}^{T} \mathbb{E}\left[\mathbb{1}[j(t) = a, \mathcal{E}^c_{a,N_a(t)}(\epsilon)]\right] \leq \mathcal{O}\left(\epsilon^{-2}\right).$$

证明。 由引理 15 和 16可知，对于 $n \geq \bar{n}$,

$$\begin{aligned}
\mathbb{P}\left[\mathcal{E}^c_{a,N_a(t)}(\epsilon), N_a(t) = n\right] &= \mathbb{P}\left[\mathcal{K}^c_{a,N_a(t)}(\epsilon), N_a(t) = n\right] + \mathbb{P}\left[\mathcal{A}^c_{a,N_a(t)}(\epsilon), \mathcal{K}_{a,N_a(t)}(\epsilon), N_a(t) = n\right] \\
&\leq \exp\left(-\frac{\alpha_a \epsilon}{\kappa_a + \epsilon} n\right) + 2\exp\left(-\frac{\alpha_a^2 \epsilon^2}{4} n\right).
\end{aligned}$$

由于 $\{j(t) = a,\ \mathcal{E}^c_{a,\ n}(\epsilon)\ N_a(t) = n\}$ 对于任意 $n \in \mathbb{N}$ 只出现一次，因此

$$\begin{aligned}
\sum_{t=\bar{n}K+1}^{T} \mathbb{E}\left[\mathbb{1}\left[j(t) = a, \mathcal{E}^c_{a,N_a(t)}(\epsilon)\right]\right] &= \sum_{t=\bar{n}K+1}^{T} \sum_{n=\bar{n}}^{T} \mathbb{E}\left[\mathbb{1}\left[j(t) = a, \mathcal{E}^c_{a,N_a(t)}(\epsilon), N_a(t) = n\right]\right] \\
&\leq \sum_{n=\bar{n}}^{\infty} \mathbb{E}\left[\mathbb{1}\left[\mathcal{E}^c_{a,N_a(t)}(\epsilon), N_a(t) = n\right]\right] \\
&= \sum_{n=\bar{n}}^{\infty} \mathbb{P}\left[\mathcal{K}^c_{a,N_a(t)}(\epsilon), N_a(t) = n\right] \\
&\qquad + \mathbb{P}\left[\mathcal{A}^c_{a,N_a(t)}(\epsilon) \cap \mathcal{K}_{a,N_a(t)}(\epsilon), N_a(t) = n\right] \\
&\leq \sum_{n=\bar{n}}^{\infty} \exp\left(-\frac{\alpha_a \epsilon}{\kappa_a + \epsilon} n\right) + 2\exp\left(-\frac{\alpha_a^2 \epsilon^2}{4} n\right).
\end{aligned}$$

由于 $\exp(-an)$ 随 n 增大而递减，其中 $a > 0$，因此有

$$\sum_{n=2}^{\infty} \exp(-an) \leq \int_1^{\infty} \exp(-an)\mathrm{d}n = \frac{\exp(-a)}{a},$$

由此完成证明。 □

C.4.1 引理15的证明

Lemma 15. 对于任意算法和 $a \in [K]$，对于所有 $\epsilon > 0$、$t > 0$ 和 $n \in \mathbb{N}$，均有

$$\mathbb{P}\left[\mathcal{K}^c_{a,N_a(t)}(\epsilon), N_a(t) = n\right] \leq \exp\left(-\frac{\alpha_a \epsilon}{\kappa_a + \epsilon} n\right).$$

证明。由于 $\hat{}\kappa_a(n) \sim \mathrm{Pa}(\kappa_a, n\alpha_a)$ 对任意 $n \in \mathbb{N}$ 属于 (2) 时均成立，因此有

$$\begin{aligned}\mathbb{P}\left[\mathcal{K}^c_{a,N_a(t)}, N_a(t) = n\right] &= \mathbb{P}\left[\hat{\kappa}_a(N_a(t)) \geq \kappa_a + \epsilon, N_a(t) = n\right] \\ &= \left(\frac{\kappa_a}{\kappa_a + \epsilon}\right)^{n\alpha_a} \leq \exp\left(-\frac{\alpha_a \epsilon}{\kappa_a + \epsilon} n\right),\end{aligned}$$

由此证明得证。 □

C.4.2 引理 16的证明

引理 16。对于任意算法以及任意 $a \in [K]$，对所有 $\epsilon \in \left(0, \frac{\kappa_a}{\alpha_a(\kappa_a+1)}\right)$ 和 $t > 0$，以及$n \geq \bar{n}$，均有

$$\mathbb{P}\left[\mathcal{A}^c_{a,N_a(t)}(\epsilon), \mathcal{K}_{a,N_a(t)}(\epsilon), N_a(t) = n\right] \leq 2\exp\left(-\frac{\alpha_a^2 \epsilon^2}{4} n\right),$$

证明。固定一个时间索引 t，并记 $\mathbb{P}_t[\cdot] = \mathbb{P}[\cdot \mid \mathcal{F}_t]$ 和 $N_a(t) = n$。为简化符号说明，我们省略 $\hat{}\kappa_a(n)$ 和 $\hat{}\alpha_a(n)$ 中的参数 n。

令 $r'_{a,k}$ 为臂 a 的 $(r_{a,s})_{s=1}^n$ 中第 k 个顺序统计量，其中 $r'_{a,1} \leq r'_{a,2} \cdots \leq r'_{a,n}$。根据 α_a 的最大似然估计定义，

$$\begin{aligned}\mathbb{P}[\hat{\alpha}_a \leq \alpha_a - \epsilon_{a,l}(\epsilon), \mathcal{K}_{a,N_a(t)}(\epsilon), N_a(t) = n] &\leq \mathbb{P}\left[\frac{n}{\sum_{s=1}^n \log r'_{a,s} - n \log r'_{a,1}} \leq \alpha_a - \epsilon_{a,l}(\epsilon)\right] \\ &= \mathbb{P}\left[\frac{n}{\alpha_a - \epsilon_{a,l}(\epsilon)} \leq \sum_{s=1}^n \log \frac{r'_{a,s}}{r'_{a,1}}\right] \\ &= \mathbb{P}\left[\frac{n}{\alpha_a - \epsilon_{a,l}(\epsilon)} \leq n \log \frac{\kappa}{r'_{a,1}} + \sum_{s=1}^n \log \frac{r'_{a,s}}{\kappa}\right] \\ &\leq \mathbb{P}\left[\frac{n}{\alpha_a - \epsilon_{a,l}(\epsilon)} \leq \sum_{s=1}^n \log \frac{r_{a,s}}{\kappa_a}\right] \\ &\leq \mathbb{P}\left[\epsilon \leq \frac{1}{n}\sum_{s=1}^n \log \frac{r_{a,s}}{\kappa_a} - \frac{1}{\alpha_a}\right],\end{aligned}$$

其中，第一个等号由（2）中的最大似然估计量定义得到；第二个不等式成立，是因为从帕累托分布中生成的任何样本都不可能小于其尺度参数 κ；最后一个不等式则由$\epsilon_{a,\,l}(\epsilon)$（12）在（12）中的定义得到。

同理，可以推导出

$$\begin{aligned}\mathbb{P}[\hat{\alpha}_a \geq \alpha_a + \epsilon_{a,u}(\epsilon), \mathcal{K}_{a,N_a(t)}(\epsilon), N_a(t) = n] &\leq \mathbb{P}\left[\sum_{s=1}^{n} \log \frac{r_{a,s}}{\kappa_a} \leq \frac{n}{\alpha_a + \epsilon_{a,u}(\epsilon)} + n \log \frac{r'_1}{\kappa} \cap \mathcal{K}\right] \\ &\leq \mathbb{P}\left[\sum_{s=1}^{n} \log \frac{r_{a,s}}{\kappa} \leq \frac{n}{\alpha_a + \epsilon_{a,u}(\epsilon)} + n \log \frac{\kappa_a + \epsilon}{\kappa_a}\right] \\ &\leq \mathbb{P}\left[\sum_{s=1}^{n} \log \frac{r_{a,s}}{\kappa_a} \leq \frac{n}{\alpha_a + \epsilon_{a,u}(\epsilon)} + \frac{n\epsilon}{\kappa_a}\right] \\ &\leq \mathbb{P}\left[\frac{1}{n}\sum_{s=1}^{n} \log \frac{r_{a,s}}{\kappa_a} - \frac{1}{\alpha_a} \leq -\epsilon\right],\end{aligned}$$

其中第二个不等式成立是因为 $r'_{a,\ 1} = \hat{}\kappa_a \leq \kappa_a + \epsilon$ 在 $\mathcal{K}_{a,\ n}$上成立，第三个不等式由$\log(1+x) \leq x$ 对 $x > -1$成立得出，最后一个不等式则来自 $\epsilon_{a\text{的定义},\ u}(\epsilon)$。根据事实 13,$y_{a,\ s} := \log \binom{r_{a,s}}{\kappa_a} \sim \text{Exp}(\alpha_a)$，最后一个概率可以视为指数分布随机变量之和的偏差。

对于指数分布 Exp(α)，若满足以下条件，我们称其满足参数为 b 的伯恩斯坦条件:

$$\mathbb{E}\left[M_k\right] \leq \frac{1}{2} k! \frac{1}{\alpha^2} b^{k-2} \quad \text{for } k = 3, 4, \ldots,$$

其中 M_k 表示 k 阶中心矩。对于 Exp(α_a)，有

$$\mathbb{E}\left[M_k\right] = \frac{!k}{\alpha_a^k} \leq \frac{k!}{2} \frac{1}{\alpha_a^2} \left(\frac{1}{\alpha_a}\right)^{k-2},$$

其中 $!k$ 是 k 的次阶乘，且满足 $!k \leq {}^{k!}_{\ } \leq k! 2$ 当 $k \geq 3$时。因此，参数为 α_a 的指数分布满足参数为 $\frac{1}{\alpha_a}$ 的伯恩斯坦条件，故其为参数 $\left(\frac{2}{\alpha_a^2}, \frac{2}{\alpha_a}\right)$ 的次指数分布。因此，应用引理 18，可得

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{s=1}^{n} y_{a,s} - \frac{1}{\alpha_a}\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{n}{4}\min\{\alpha_a^2\epsilon^2, \alpha_a\epsilon\}\right).$$

注意，对于 $\epsilon < {}^{\kappa_a}_{\alpha_a\ (\kappa_a+1)}$，有

$$\begin{aligned}\mathbb{P}[\hat{\alpha}_a \leq \alpha_a - \epsilon_{a,l}(\epsilon) \cap \mathcal{K}_{a,n}] &\leq \mathbb{P}\left(\frac{1}{n}\sum_{s=1}^{n} y_{a,s} - \frac{1}{\alpha_a} \geq \epsilon\right) \\ \mathbb{P}[\hat{\alpha}_a \geq \alpha_a + \epsilon_{a,u}(\epsilon) \cap \mathcal{K}_{a,n}] &\leq \mathbb{P}\left(\frac{1}{n}\sum_{s=1}^{n} y_{a,s} - \frac{1}{\alpha_a} \leq -\epsilon\right),\end{aligned}$$

对于 $\epsilon_{a,\ l}(\epsilon) = \frac{\epsilon\alpha_a^2}{1 \mp \epsilon\alpha_a}$ 和 $\epsilon_{a,\ u}(\epsilon) = \frac{\epsilon\alpha_a^2}{(\kappa_a+1)\,\kappa_a - \epsilon\alpha_a(\kappa_a+1)}$，它们满足 $\lim_{\epsilon\to 0} \max\{\epsilon_{a,l}(\epsilon)\, \epsilon_{a,\ u}(\epsilon)\} = 0_+$。因此，通过恢复原始符号，我们得到

$$\begin{aligned}\mathbb{P}[\mathcal{A}^c_{a,N_a(t)}(\epsilon), \mathcal{K}_{a,N_a(t)}(\epsilon), N_a(t) = n] &= \mathbb{P}[\hat{\alpha}_a(n) \leq \alpha_a - \epsilon_{a,l}(\epsilon), \mathcal{K}_{a,N_a(t)}, N_a(t) = n] \\ &\quad + \mathbb{P}[\hat{\alpha}_a(n) \geq \alpha_a + \epsilon_{a,u}(\epsilon), \mathcal{K}_{a,N_a(t)}, N_a(t) = n] \\ &\leq 2\exp\left(-\frac{\alpha_a^2\epsilon^2}{4} n\right),\end{aligned}$$

对于 $\epsilon < \frac{1}{\alpha_a}$，其中有 $\alpha_a > 1$。 □

C.5 引理 9 的证明9

引理 9。对于任意臂 $a \in [K]$ 和固定的 $t \in \mathbb{N}$，令 $N_a(t) = n$。对于任意正数 $\xi \leq \frac{y}{y-\hat{\kappa}_a}$ 和 $k \in \mathbb{Z}$，有

$$\mathbb{1}[\hat{\kappa}_a(n) \leq y]\mathbb{P}[\tilde{\mu}_a(t) \leq y|\theta_{a,n}] \leq \int_{\xi}^{\infty} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\mathrm{d}x + \left(\frac{y}{\mu((\kappa_a,\xi))}\right)^n \int_1^{\xi} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\mathrm{d}x,$$

其中 $f^{\mathrm{Er}}_{s,\beta}(\cdot)$ 表示形状参数为 $s \in \mathbb{N}$、速率参数为$\beta > 0$的厄朗分布的概率密度函数。

证明。固定一个时间索引 t，使得 $N_a(t) = n$，并记为 $\mathbb{P}_t[\cdot] = \mathbb{P}[\cdot \mid \mathcal{F}_t]$。为简化符号说明，我们省略 $\hat{}$ $\kappa_a(n)$ 和 $\hat{}$ $\alpha_a(n)$ 中的参数 n，以及 $\tilde{\alpha}_k(t)$、$\tilde{\alpha}_a(t)$ 和 $\tilde{\mu}_a(t)$ 中的参数 t。

当 $\frac{\alpha_a}{\alpha_a-1} \leq y$ 成立时，因为如果 $\hat{\kappa}_a < y$，则无论其在 $(0, \hat{\kappa}_a)$ 取值如何，$\tilde{\mu}_a \leq y$ 都成立，如果 $\hat{\kappa}_a < y$，那么

$$\tilde{\alpha}(t) \geq \frac{y}{y-\hat{\kappa}_a} \implies \tilde{\mu}_a \leq y. \tag{47}$$

当 $1 < \tilde{\alpha}(t) < \frac{y}{y-\hat{\kappa}_a}$ 时，

$$\tilde{\mu}_a = \tilde{\kappa}_a \frac{\tilde{\alpha}_a}{\tilde{\alpha}_a - 1} \leq y \Leftrightarrow \tilde{\kappa}_a \leq y\frac{\tilde{\alpha}_a - 1}{\tilde{\alpha}_a}. \tag{48}$$

由于 $\tilde{\alpha}_a \leq 1$意味着 $\tilde{\mu}_a = \infty$，由（47）和（48），可得

$$\begin{aligned}
\mathbb{P}_t[\tilde{\mu}_a \leq y] &= \int_1^{\frac{y}{y-\hat{\kappa}_a}} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\mathbb{P}_t\left[\tilde{\kappa}_a \leq \frac{x-1}{x}y\right]\mathrm{d}x + \int_{\frac{y}{y-\hat{\kappa}_a}}^{\infty} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\mathrm{d}x \\
&= \int_1^{\frac{y}{y-\hat{\kappa}_a}} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\left(\frac{x-1}{\hat{\kappa}_a x}y\right)^{nx}\mathrm{d}x + \int_{\frac{y}{y-\hat{\kappa}_a}}^{\infty} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\mathrm{d}x,
\end{aligned} \tag{49}$$

其中我们记作 $\mathbb{P}_t[\cdot] = \mathbb{P}[\cdot|\mathcal{F}_t]$ 并在（8）中恢复了（49）中的累积分布函数。任取有限的 $y' > y$ 并令$\xi := \frac{y'}{y'-\kappa_a} < \frac{y}{y-\kappa_a}$ 使得 $\mu((\kappa_a, \xi))) = y'$。由于 $\frac{a}{a-b}$ 关于 $a > b > 0$单调递减，因此可知

$$\begin{aligned}
\mathbb{P}_t[\tilde{\mu}_a \leq y] &= \int_1^{\frac{y}{y-\hat{\kappa}_a}} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\left(\frac{x-1}{\hat{\kappa}_a x}y\right)^{nx}\mathrm{d}x + \int_{\frac{y}{y-\hat{\kappa}_a}}^{\infty} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\mathrm{d}x \\
&\leq \int_1^{\frac{y'}{y'-\hat{\kappa}_a}} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\left(\frac{x-1}{\hat{\kappa}_a x}y\right)^{nx}\mathrm{d}x + \int_{\frac{y'}{y'-\hat{\kappa}_a}}^{\infty} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\mathrm{d}x \\
&\leq \int_1^{\frac{y'}{y'-\kappa}} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\left(\frac{x-1}{\hat{\kappa}_a x}y\right)^{nx}\mathrm{d}x + \int_{\frac{y'}{y'-\kappa}}^{\infty} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\mathrm{d}x \\
&\leq \int_1^{\xi} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\left(\frac{x-1}{\kappa x}y\right)^{n}\mathrm{d}x + \int_{\xi}^{\infty} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\mathrm{d}x \qquad (50) \\
&\leq \left(\frac{\xi-1}{\kappa\xi}y\right)^n \int_1^{\xi} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\mathrm{d}x + \int_{\xi}^{\infty} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\mathrm{d}x \qquad (51) \\
&= \left(\frac{y}{\mu((\kappa,\xi))}\right)^n \int_1^{\xi} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\mathrm{d}x + \int_{\xi}^{\infty} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{a,n}}}(x)\mathrm{d}x,
\end{aligned}$$

其中（50）由 $\hat{\kappa}_a \geq \kappa$得出，并且我们在（51）中利用了 $\frac{x-1}{x}$ 的单调递增性质。 □

C.6 引理10的证明

引理10。给定 $\epsilon > 0$，定义一个正的问题相关常数 $\rho = \rho_{\theta_1}(\epsilon) := 2\frac{\kappa_1\epsilon}{(\mu_1 - \epsilon/2 - \kappa_1)(\mu_1 - \kappa_1)}$。若 $n \geq \bar{n} = \max(2, k+1)$，则对于 $k \in \mathbb{Z}_{\geq 0}$

$$p_n(\epsilon|\theta_{1,n}) \leq \begin{cases} e^{-n}, & \text{if } \hat{\kappa}_1(n) \geq \mu_1 - \epsilon, \\ h(\mu_1, \epsilon, n), & \text{if } \hat{\kappa}_1(n) \in K(\epsilon), \alpha_{1,n} \leq \alpha_1 + \rho, \\ C_1(\mu_1, \epsilon, n) G_k(1/\alpha_{1,n}; n) + 1 - G_k(1/\alpha_{1,n}; n) & \text{if } \hat{\kappa}_1(n) \in K(\epsilon), \alpha_{1,n} \geq \alpha_1 + \rho, \end{cases}$$

其中

$$\begin{aligned} h(\mu_1, \epsilon, n) &:= e^{-n\frac{3\epsilon}{4\mu_1}} \left(1 - \frac{1}{2} e^{-n c_{\mu_1}(\epsilon)}\right) + \frac{1}{2} e^{-n c_{\mu_1}(\epsilon)} \\ C_1(\mu_1, \epsilon, n) &:= \left(\frac{\mu_1 - \epsilon}{\mu_1 - \epsilon/2}\right)^n \leq e^{-n\frac{\epsilon}{2\mu_1 - \epsilon}} < 1. \\ G_k(x; n) &:= F^{\mathrm{Er}}_{n-k, nx}(\alpha_1 + \rho) \end{aligned} \tag{20}$$

对于 F^{Er}，其定义见 *(10)*。这里，$c_{\mu_1}(\epsilon) = \zeta - \log(1+\zeta) = \mathcal{O}(\epsilon^{-2})$ 以及 $\zeta = \frac{\epsilon}{4\mu_1 - 2\epsilon} \in (0,1)$ 分别是 μ_1 和 ϵ的确定性常数。

证明。与其他证明类似，固定 t 并令 $N_1(t) = n$。为简化符号说明，我们省略$\tilde{\kappa}_1(t)$ $\tilde{\alpha}_1(t)$ 和 $\tilde{\mu}_1(t)$ 中的参数 t，以及 $\hat{\kappa}_1(n)$，$\hat{\alpha}_1(n)$ $\bar{\alpha}_1(n)$ 中的参数 n。

情形 1。在 $\{\hat{\kappa}_1 \geq \mu_1 - \epsilon\}$ 在条件 $\{\hat{\kappa}_1 \geq \mu_1 - \epsilon\}$ 下，事件 $\{\tilde{\mu}_1 \leq \mu_1 - \epsilon\}$ 最终由 $\tilde{\kappa}$ 的取值确定，因为 $\{\tilde{\kappa}_1 \in (\mu_1 - \epsilon, \hat{\kappa}_1]\}$ 是满足 $\{\tilde{\mu}_1 > \mu_1 - \epsilon\}$ 的充分条件。因此，若 $\hat{\kappa}_1 \geq \mu_1 - \epsilon$，则

$$p_n(\epsilon|\theta_{1,n}) = \int_1^\infty f^{\mathrm{Er}}_{n-k, \frac{n}{\alpha_{1,n}}}(x) \left(\frac{\mu_1 - \epsilon}{\hat{\kappa}_1} \frac{x-1}{x}\right)^{nx} \mathrm{d}x.$$

于是，

$$\begin{aligned} \mathbb{1}[\hat{\kappa}_1 \geq \mu_1 - \epsilon] p_n(\epsilon|\theta_{1,n}) &= \mathbb{1}[\hat{\kappa}_1 \geq \mu_1 - \epsilon] \left(\int_1^\infty f^{\mathrm{Er}}_{n-k, \frac{n}{\alpha_{1,n}}}(x) \left(\frac{\mu_1 - \epsilon}{\hat{\kappa}_1} \frac{x-1}{x}\right)^{nx} \mathrm{d}x\right) \\ &\leq \mathbb{1}[\hat{\kappa}_1 \geq \mu_1 - \epsilon] \int_1^\infty f^{\mathrm{Er}}_{n-k, \frac{n}{\alpha_{1,n}}}(x) \left(1 - \frac{1}{x}\right)^{nx} \mathrm{d}x \\ &\leq \mathbb{1}[\hat{\kappa}_1(n) \geq \mu_1 - \epsilon] \int_1^\infty f^{\mathrm{Er}}_{n-k, \frac{n}{\alpha_{1,n}}}(x) e^{-n} \mathrm{d}x \\ &\leq \mathbb{1}[\hat{\kappa}_1(n) \geq \mu_1 - \epsilon] e^{-n}, \end{aligned}$$

其中第二个不等式由 $\mu_1 - \epsilon \leq \hat{\kappa}_1$ 得出。

情形 2。在 $\{\hat{}\kappa_1 \in K(\epsilon)\alpha_{1,n} \leq \alpha_1 + \rho\}$ 通过应用引理 9，其中 $y = \mu_1 - \epsilon$，可得对于任意 $\xi \leq \frac{\mu_1 - \epsilon}{\mu_1 - \epsilon - \kappa_1}$，有

$$\begin{aligned}\mathbb{1}[\hat{\kappa}_1 < \mu_1 - \epsilon, \alpha_{1,n} \leq \alpha + \rho] p_n(\epsilon|\theta_{1,n}) &\leq \left(\frac{\mu_1 - \epsilon}{\mu((\kappa_1, \xi))}\right)^n \int_1^{\xi} f^{\mathrm{Er}}_{n-k, \frac{n}{\alpha_{1,n}}}(x)\mathrm{d}x + \int_{\xi}^{\infty} f^{\mathrm{Er}}_{n-k, \frac{n}{\alpha_{1,n}}}(x)\mathrm{d}x \\ &\leq \left(\frac{\mu_1 - \epsilon}{\mu((\kappa_1, \xi))}\right)^n \int_0^{\xi} f^{\mathrm{Er}}_{n-k, \frac{n}{\alpha_{1,n}}}(x)\mathrm{d}x + \int_{\xi}^{\infty} f^{\mathrm{Er}}_{n-k, \frac{n}{\alpha_{1,n}}}(x)\mathrm{d}x.\end{aligned} \tag{52}$$

令 $\bar{\rho} := \rho_\theta(\epsilon/2)$。那么，它满足 $\mu((\kappa_1, \alpha_1 + \bar{\rho})) = \mu_1 - \frac{\epsilon}{4}$ 且

$$\alpha_1 + \bar{\rho} = \frac{\mu - \epsilon/4}{\mu - \epsilon/4 - \kappa_1} < \frac{\mu - \epsilon}{\mu - \epsilon - \kappa_1},$$

其中，不等式成立是因为函数 $\frac{x}{x-y}$ 关于 $x > y$具有递减性质。将 ξ 替换为 $\alpha_1 + \bar{\rho}$ 代入式（52），得到

$$\begin{aligned}&\mathbb{1}[\hat{\kappa}_1 < \mu_1 - \epsilon, \alpha_{1,n} \leq \alpha_1 + \rho] p_n(\epsilon|\bar{\theta}_{1,n}) \\ &\quad\leq \mathbb{1}[\hat{\kappa}_1 < \mu_1 - \epsilon, \alpha_{1,n} \leq \alpha_1 + \rho]\left(\left(\frac{\mu_1 - \epsilon}{\mu_1 - \epsilon/4}\right)^n \int_0^{\xi} f^{\mathrm{Er}}_{n-k, \frac{n}{\alpha_{1,n}}}(x)\mathrm{d}x + \int_{\alpha_1+\bar{\rho}}^{\infty} f^{\mathrm{Er}}_{n-k, \frac{n}{\alpha_{1,n^{(n)}}}}(x)\mathrm{d}x\right) \\ &\quad\leq \mathbb{1}[\hat{\kappa}_1 < \mu_1 - \epsilon, \alpha_{1,n} \leq \alpha_1 + \rho]\left(e^{-n\left(\frac{3\epsilon}{4\mu_1 - \epsilon}\right)}(1 - \mathbb{P}[\tilde{\alpha}_1 \geq \alpha_1 + \bar{\rho}]) + \mathbb{P}[\tilde{\alpha}_1 \geq \alpha_1 + \bar{\rho}]\right).\end{aligned} \tag{53}$$

设 Z_n 为服从自由度为 n 的卡方分布的随机变量，并用 $F_n(\cdot)$ 表示 Z_n的累积分布函数。那么，有

$$\begin{aligned}\mathbb{P}\left[\tilde{\alpha}_1 \geq \alpha_1 + \bar{\rho}, \alpha_{1,n} \leq \alpha_1 + \rho\right] &= \mathbb{P}\left[Z \geq \frac{2n}{\alpha_{1,n}}(\alpha_1 + \bar{\rho}), \alpha_{1,n} \leq \alpha_1 + \rho\right] \quad \text{by Fact 12} \\ &\leq \mathbb{P}\left[Z \geq 2n\frac{\alpha_1 + \bar{\rho}}{\alpha_1 + \rho}\right] \\ &\leq \mathbb{P}\left[Z \geq 2n\frac{\mu_1 - \epsilon/4}{\mu_1 - \epsilon/2}\right] \\ &= 1 - F_{2n-2k}(2n(1 + \zeta)),\end{aligned} \tag{54}$$

其中 $\zeta = \frac{\epsilon}{4\mu_1 - 2\epsilon} \in (0, 1)$。应用引理 19，若 $n\zeta > -k$，则有

$$\begin{aligned}&\mathbb{P}[\tilde{\alpha}_1 \geq \alpha_1 + \bar{\rho}, \alpha_{1,n} \leq \alpha_1 + \rho] \\ &\qquad\leq 1 - F_{2n-2k}\left(2n(1 + \zeta)\right) \\ &\qquad< \frac{1}{2}\frac{\sqrt{2\pi}(n-k)^{n-k-1/2}e^{-(n-k)}}{\Gamma(n-k)}\mathrm{erfc}\left(\sqrt{n(\zeta + k) - (n-k)\log\frac{n(1+\zeta)}{n-k}}\right),\end{aligned}$$

其中，$\Gamma(\cdot)$ 表示伽马函数。对于 $n \geq 1/2$，根据斯特林公式可得

$$\sqrt{2\pi}n^{n-1/2}e^{-n} \leq \Gamma(n) \leq \sqrt{2\pi}e^{1/6}n^{n-1/2}e^{-n},$$

由此可得

$$\mathbb{P}[\tilde{\alpha}_1 \geq \alpha_1 + \bar{\rho}, \alpha_{1,n} \leq \alpha_1 + \rho] < \frac{1}{2}\mathrm{erfc}\left(\sqrt{n(\zeta+k) - (n-k)\log\frac{n(1+\zeta)}{n-k}}\right). \tag{55}$$

注意到，$(n-k)\log\frac{n}{(1+\zeta)n-k} > 0$根据$n\zeta > -k$的假设总是成立，其中先验分布满足$k \in \mathbb{Z}_{\geq 0}$，且不论$n$为何。因此，若$\zeta + k \leq 0$，则式（55）中互补误差函数的自变量为负。这使得式（55）中的上界大于或等于1 2。因此，对于满足$k \in \mathbb{Z}_{<0}$的先验分布，由于$\zeta \in (0,1)$，式（55）中的右侧项有一个常数上界。

由于互补误差函数是递减函数，对于满足$k \in \mathbb{Z}_{\geq 0}$的先验分布，由式（55）可得

$$\mathbb{P}[\tilde{\alpha}_1 \geq \alpha_1 + \bar{\rho}, \alpha_{1,n} \leq \alpha_1 + \rho] \leq \frac{1}{2}\mathrm{erfc}\left(\sqrt{n(\zeta - \log(1+\zeta)}\right),$$

其中代入$k = 0$。通过变量代换，对于任意$x \geq 0$，互补误差函数有如下上界 [Simon and Divsalar, 1998]:

$$\mathrm{erfc}(x) \leq e^{-x^2},$$

这意味着

$$\mathbb{P}[\tilde{\alpha}_1 \geq \alpha_1 + \bar{\rho}, \alpha_{1,n} \leq \alpha_1 + \rho] \leq \frac{1}{2}e^{-nc_{\mu_1}(\epsilon)}, \tag{56}$$

其中$c_{\mu_1}(\epsilon) = \zeta - \log(1+\zeta) > 0$在$\mu_1$上是一个确定性常数，且$\epsilon$。结合式（56）和式（53），可得

$$\mathbb{1}[\hat{\kappa}_1 < \mu_1 - \epsilon, \alpha_{1,n} \leq \alpha_1 + \rho] p_n(\epsilon|\theta_{1,n}) \leq e^{-n\frac{3\epsilon}{4\mu_1}}\left(1 - \frac{1}{2}e^{-nc_{\mu_1}(\epsilon)}\right) + \frac{1}{2}e^{-nc_{\mu_1}(\epsilon)} =: h(\mu_1, \epsilon, n).$$

根据幂级数 展开式$\log(1+x)$，我们有$\log(1+x) \geq x - \frac{x^2}{2} + \frac{x^3}{3}$，其中$x \in (0, 1)$，并且

$$\begin{aligned} c_{\mu_1}(\epsilon) = \zeta - \log(1+\zeta) &\leq \frac{\zeta^2}{2} - \frac{\zeta^3}{3} = \frac{\zeta^2}{6}(3 - 2\zeta) \\ &\leq \frac{\zeta^2}{2} = \mathcal{O}(\epsilon^{-2}). \end{aligned}$$

情形 3. 在$\{\hat{\kappa}_1 \in K(\epsilon)\, \alpha_{1,\ n} \geq \alpha_1 + \rho\}$应用引理 9，取$y = \mu_1 - \epsilon$和$\xi = \alpha_1 + \rho$，我们得到

$$\begin{aligned} \mathbb{1}[\hat{\kappa}_1 < \mu_1 - \epsilon] p_n(\epsilon|\theta_{1,n}) &\leq \left(\frac{\mu_1 - \epsilon}{\mu_1 - \epsilon/2}\right)^n \int_1^{\alpha_1+\rho} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{1,n}}}(x)\mathrm{d}x + \int_{\alpha_1+\rho}^{\infty} f^{\mathrm{Er}}_{n-k,\frac{n}{\alpha_{1,n}}}(x)\mathrm{d}x \\ &= C_1(\mu_1, \epsilon, n)\mathbb{P}\left[\tilde{\alpha}_1 \in [1, \alpha_1 + \rho] \mid \alpha_{1,n}\right] + \mathbb{P}\left[\tilde{\alpha}_1 \geq \alpha_1 + \rho \mid \alpha_{1,n}\right] \\ &\leq C_1(\mu_1, \epsilon, n)\mathbb{P}\left[\tilde{\alpha}_1 \leq \alpha_1 + \rho \mid \alpha_{1,n}\right] + \mathbb{P}\left[\tilde{\alpha}_1 \geq \alpha_1 + \rho \mid \alpha_{1,n}\right] \\ &= C_1(\mu_1, \epsilon, n)G_k(1/\alpha_{1,n}; n) + (1 - G_k(1/\alpha_{1,n}; n)), \end{aligned} \tag{57}$$

其中$A_n = C_1(\mu_1, \epsilon, n) := \left(\frac{\mu_1-\epsilon}{\mu_1-\epsilon/2}\right)^n \leq e^{-n\frac{\epsilon}{2\mu_1-\epsilon}} < 1$。由于$\tilde{\alpha}_1$服从厄朗分布$\left(n-k, \frac{n}{\alpha_{1,n}}\right)$，因此有

$$\mathbb{P}\left[\tilde{\alpha}_1 \leq \alpha_1 + \rho \mid \alpha_{1,n}\right] = \frac{\gamma\left(n-k, \frac{n(\alpha_1+\rho)}{\alpha_{1,n}}\right)}{\Gamma(n-k)},$$

其中 $\gamma(\cdot,\cdot)$ 表示下不完全伽马函数。因此，令

$$G_k(x;n) := \frac{\gamma\left(n-k, n(\alpha_1+\rho)x\right)}{\Gamma(n-k)}$$

证明至此结束。 □

D 定理3的证明

如定理 2 的证明所示，式（26）中的积分项在 $k \in \mathbb{Z}_{\leq 1}$ 不受$\hat{}\alpha$限制时发散。本节给出定理 3 在 $k \in \mathbb{Z}_{\leq 0}$下的部分证明，说明此类条件对于实现渐近最优性是必要的。

定理 3 的证明。考虑一个具有 $\theta_1 = (\kappa, \alpha_1)$ 和 $\theta_2 = (\kappa, \alpha_2)$ 的双臂老虎机问题。令$\gamma = \max\{\lceil \alpha_2 \rceil, \lfloor \alpha_2 \rfloor + 1\}$ 和 $m = \frac{\gamma}{\gamma-1}$，使得 $\frac{\mu_2}{m} = \kappa \frac{\alpha_2}{\alpha_2-1} \frac{\gamma-1}{\gamma} > \kappa$。假设 $1 < \alpha_1 < \alpha_2$ 和 $\tilde{\mu}_2(s) = \mu_2 = \kappa \frac{\alpha_2}{\alpha_2-1}$ 对所有 $s \in \mathbb{N}$均成立。回顾一下，对于先验分布 $k \leq 1$，STS 从每个臂都玩两次开始，即 $N_a(s) \geq 2$ 对所有 $a \in \{1,2\}$均成立。对于 $T \geq 5$，我们有

$$\begin{aligned}
\mathbb{E}[\mathrm{Reg}(T)] &= \Delta_2 \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}[j(t)=2]\right] \\
&\geq \Delta_2 \mathbb{E}\left[\sum_{t=5}^{T} \mathbb{1}[j(t)=2, N_1(t)=2]\right].
\end{aligned}$$

根据定义 $N_1(\cdot)$、$\{j(s) \neq 2,\ N_1(s) = 2\}$ 意味着 $N_1(t) > 2$ 对于 $t > s$。因此，对于任意 $t \geq 5$ ，

$$\begin{aligned}
\{j(t)=2, N_1(t)=2\} &\Leftrightarrow \{\forall s \in [1, t-4] : j(s+4) = 2\} \\
&\Leftrightarrow \{\forall s \in [1, t-4] : \tilde{\mu}_1(s+4) < \mu_2\}.
\end{aligned}$$

令 $T' = T-4$，则有

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=5}^{T} \mathbb{1}[j(t)=2, N_1(t)=2]\right] &= \mathbb{E}\left[\sum_{t=5}^{T} \mathbb{1}\left[\forall s \in [1,t-4] : \tilde{\mu}_1(s+4) < \mu_2\right]\right] \\
&= \mathbb{E}\left[\sum_{s=1}^{T'} \left(\mathbb{P}[\tilde{\mu}_1 \leq \mu_2 | \hat{\kappa}_1(2), \hat{\alpha}_1(2)]\right)^s\right]. \qquad (58)
\end{aligned}$$

注意，对所有 $s \geq 2$，$\hat{}\kappa_1(N_1(s)) = \hat{}\kappa_1(2)$ 和 $\hat{}\alpha_1(N_1(s)) = \hat{}\alpha_1(2)$ 均成立，因为对所有$s \geq 2$，只有 $j(s) = 2$ 成立。这里，我们首先给出 $\mathbb{P}[\tilde{\mu}_1 \leq \mu_2 | \hat{}\kappa_1, \hat{}\alpha_1]$的下界。由于 $\frac{\mu_2}{m} = \frac{\kappa\alpha_2}{\alpha_2-1} \frac{\gamma-1}{\gamma} > \kappa$ 成立，我们可以考虑 $\hat{}\kappa_1(2) \leq \frac{\mu_2}{m}$发生的情况。

由（49）可知，对 $y \geq \hat{}\kappa_1(n)$ 有

$$\begin{aligned}
\mathbb{P}_t[\tilde{\mu}_a \leq y] &= \int_1^{\frac{y}{y-\hat{\kappa}(n)}} f^{\mathrm{Er}}_{n-k, \frac{n}{\alpha_n}}(x) \left(\frac{x-1}{\hat{\kappa}x} y\right)^{nx} \mathrm{d}x + \int_{\frac{y}{y-\hat{\kappa}(n)}}^{\infty} f^{\mathrm{Er}}_{n-k, \frac{n}{\alpha_n}}(x) \mathrm{d}x \qquad \text{by (49)} \\
&\geq \int_{\frac{y}{y-\hat{\kappa}(n)}}^{\infty} f^{\mathrm{Er}}_{n-k, \frac{n}{\alpha_n}}(x) \mathrm{d}x.
\end{aligned}$$

令 $n=2$和 $y=\mu_2$，则对 $k\in\mathbb{Z}_{\leq 1}$

$$\begin{aligned}\mathbb{P}[\tilde{\mu}_1\leq\mu_2|\hat{\kappa}_1(2),\hat{\alpha}_1(2)] &\geq \mathbb{1}\left[\hat{\kappa}\leq\frac{\mu_2}{m}\right]\int_{\frac{\mu_2}{\mu_2-\hat{\kappa}}}^{\infty} f^{\mathrm{Er}}_{2-k,\frac{2}{\hat{\alpha}}}(x)\mathrm{d}x \\ &\geq \mathbb{1}\left[\hat{\kappa}\leq\frac{\mu_2}{m}\right]\int_{\gamma}^{\infty} f^{\mathrm{Er}}_{2-k,\frac{2}{\hat{\alpha}}}(x)\mathrm{d}x \qquad (59)\\ &= \mathbb{1}\left[\hat{\kappa}\leq\frac{\mu_2}{m}\right]\frac{\Gamma(2-k,\frac{2\gamma}{\hat{\alpha}})}{\Gamma(2-k)}, \qquad (60)\end{aligned}$$

其中（59）由 $\mu_2-\hat{\kappa}_1(2)\leq\mu_2-\mu_2/m=m-1=\gamma=\lceil\alpha_2\rceil$ 得出，且 $\Gamma(\cdot,\cdot)$ 是上不完全伽马函数。为简化符号说明，在以下各节中，我们省略 $\tilde{\mu}_1(t)$、$\hat{\kappa}_1(n)$ 和 $\hat{\alpha}_1(n)$ 中 n 和 t 的自变量。

D.1 先验分布 $k\in\mathbb{Z}_{\leq 0}$

注意，$\Gamma(n,x)$ 关于 n 是递增函数（固定 x 时）。因此，（60）意味着，如果参考先验的遗憾下界大于该下界，那么所有满足 $k\in\mathbb{Z}_{\leq 0}$的先验分布都是次优的。因此，我们考虑 $k=1$的情况，此时可将（60）改写为

$$\mathbb{P}[\tilde{\mu}_1\leq\mu_2|\hat{\kappa}_1,\hat{\alpha}_1]\geq\mathbb{1}\left[\hat{\kappa}_1\leq\frac{\mu_2}{m}\right]\frac{\Gamma(1,\frac{2\gamma}{\hat{\alpha}_1})}{\Gamma(1)}=e^{-\frac{2\gamma}{\hat{\alpha}_1}}. \qquad (61)$$

由于 $\hat{\alpha}_1(2)\sim\mathrm{IG}(1,2\alpha_1)$ 在（2）中，$z:=2\gamma\hat{\alpha}$ 服从率参数为 α_1/γ 的指数分布，即$z\sim\mathrm{Exp}(\alpha_1/\gamma)$。结合（61）与（58），得到

$$\begin{aligned}\mathbb{E}\left[\sum_{t=5}^{T}\mathbb{1}[j(t)=2,N_1(t)=2]\right] &\geq \mathbb{E}_{\hat{\kappa},z}\left[\sum_{s=1}^{T'}\left(\mathbb{1}[\hat{\kappa}\leq\mu_2/m]e^{-z}\right)^s\right] \\ &= \mathbb{P}[\hat{\kappa}\leq\mu_2/m]\mathbb{E}_{z\sim\mathrm{Exp}(\alpha_1/\gamma)}\left[\sum_{s=1}^{T'}e^{-zs}\right], \qquad (62)\end{aligned}$$

其中我们利用了 $\hat{\alpha}$ 和 $\hat{\kappa}$ 的随机独立性。这里，

$$\begin{aligned}\mathbb{E}_{z\sim\mathrm{Exp}(\alpha_1/\gamma)}\left[\sum_{s=1}^{T'}e^{-zs}\right] &= \mathbb{E}_{z\sim\mathrm{Exp}(\alpha_1/\gamma)}\left[(1-e^{-zT'})\frac{e^{-z}}{1-e^{-z}}\right] \\ &= \int_0^{\infty}(1-e^{-xT'})\frac{e^{-x}}{1-e^{-x}}e^{-\frac{\alpha_1}{\gamma}x}\mathrm{d}x \\ &\geq \int_0^{\infty}(1-e^{-xT'})\frac{e^{-2x}}{1-e^{-x}}\mathrm{d}x \qquad \text{by } \frac{\alpha_1}{\gamma}<1 \\ &\geq \left(1-\frac{1}{e}\right)\int_{\frac{1}{T'}}^{\infty}\frac{e^{-2x}}{1-e^{-x}}\mathrm{d}x \\ &= \left(1-\frac{1}{e}\right)\left[\log(e^x-1)-z+e^{-z}\right]_{x=\frac{1}{T'}}^{\infty} \\ &\geq \left(1-\frac{1}{e}\right)\left(\log T'+1-\frac{3}{2T'}\right), \qquad (63)\end{aligned}$$

其中最后一个不等式由其幂级数展开得出

$$\log(e^x-1)-x+e^{-x}\geq \log(x)+1-\frac{3}{2}x$$

且$\lim_{x\to\infty}\log(e^x-1)-x+e^{-x}=0$。将（63）与（62）和（58）结合起来，并对$\hat{}\kappa_1(2)\sim \mathrm{Pa}(\kappa_1, 2\alpha_1)$进行初等计算，可得

$$\begin{aligned}\mathbb{E}[\mathrm{Reg}(T)] &\geq \Delta_2\left(1-\left(\frac{m\kappa}{\mu_2}\right)^{2\alpha_1}\right)\left(1-\frac{1}{e}\right)\left(\log T'+1-\frac{3}{2T'}\right)\\ &= \Delta_2\left(1-\left(\frac{m\kappa}{\mu_2}\right)^{2\alpha_1}\right)\left(1-\frac{1}{e}\right)\left(\log(T+4)+1-\frac{3}{2(T+4)}\right).\end{aligned}$$

因此，在参数为 $k\in\mathbb{Z}_{\leq 1}$ 的 STS 下，存在常数 $C(\alpha_1,\alpha_2)$，使得

$$\liminf_{T\to\infty}\frac{\mathbb{E}[\mathrm{Reg}(T)]}{\log T}\geq C(\alpha_1,\alpha_2).$$

D.2 先验 $k\in\mathbb{Z}_{\leq 0}$

与上一节类似，只需考虑 $k=0$，此时可将（60）重写为

$$\mathbb{P}[\tilde{\mu}_1\leq\mu_2|\hat{\kappa}_1,\hat{\alpha}_1]\geq \mathbb{1}\left[\hat{\kappa}_1\leq\frac{\mu_2}{m}\right]\frac{\Gamma(2,\frac{2\gamma}{\hat{\alpha}_1})}{\Gamma(2)}. \tag{64}$$

根据上不完全伽马函数的定义，可得

$$g(z):=\Gamma(2,z)=\int_z^\infty x^1e^{-x}\mathrm{d}x=e^{-z}(z+1),$$

作为 e^{-z} 在（62）中的对应形式，采用相同的符号 $z=\frac{2\gamma}{\hat{\alpha}_1}\sim\mathbf{Exp}\left(\frac{\alpha_1}{\gamma}\right)$。因此，将 e^{-zs} 代入（62）中的 $g(z)^s$，可得

$$\begin{aligned}\mathbb{E}_z\left[\sum_{s=1}^{T'}(g(z))^s\right] &\geq \mathbb{E}_z\left[\mathbb{1}[z\in(0,1]]\sum_{s=1}^{T'}(g(z))^s\right]\\ &\geq \mathbb{E}_z\left[\mathbb{1}[z\in(0,1]]\sum_{s=1}^{T'}(1-z^2)^s\right]\\ &= \mathbb{E}_z\left[\mathbb{1}[z\in(0,1]](1-(1-z^2)^{T'})\frac{1-z^2}{z^2}\right],\end{aligned}$$

其中在第二个不等式中，我们用到了这一事实 $z\in[0,1], g(z)\geq 1-z^2$。由于 $z\in\left(0,\frac{1}{\sqrt{T'}}\right]$，$(1-z^2)^{T'}\leq\frac{1}{1+T'z^2}$成立，因此我们有 $1-(1-z^2)^{T'}\geq\frac{T'z^2}{1+T'z^2}$。

应用这一事实，对于 $T' > 1$，我们有

$$\begin{aligned}\mathbb{E}_z\left[\sum_{s=1}^{T'}(g(z))^s\right] &\geq \mathbb{E}_z\left[\frac{T'(1-z^2)}{1+T'z^2}\mathbb{1}\left[z \in \left(0, \frac{1}{\sqrt{T'}}\right]\right]\right] \\ &\geq \mathbb{E}_{z\sim\mathrm{Exp}(\alpha_1/\gamma)}\left[\left(\frac{T'}{2}-\frac{1}{2}\right)\mathbb{1}\left[z \in \left(0, \frac{1}{\sqrt{T'}}\right]\right]\right] \\ &= \left(\frac{T'}{2}-\frac{1}{2}\right)\int_0^{\frac{1}{\sqrt{T'}}} \frac{\alpha_1}{\gamma} e^{-\frac{\alpha_1}{\gamma}z}\mathrm{d}z \\ &= \left(\frac{T'}{2}-\frac{1}{2}\right)\left(1-e^{-\frac{\alpha_1}{\gamma\sqrt{T'}}}\right).\end{aligned} \tag{65}$$

注意到 $e^{-x} \leq 1-\frac{x}{2}$ 对 $x < 1$成立，因此

$$\begin{aligned}\mathbb{E}_z\left[\sum_{s=1}^{T'}(g(z))^s\right] &\geq \left(\frac{T'}{2}-\frac{1}{2}\right)\left(1-e^{-\frac{\alpha_1}{\gamma\sqrt{T'}}}\right) \\ &\geq \left(\frac{T'}{2}-\frac{1}{2}\right)\frac{\alpha_1}{2\gamma\sqrt{T'}} = \frac{\alpha_1}{4\gamma}\left(\sqrt{T'}-\frac{1}{\sqrt{T'}}\right).\end{aligned} \tag{66}$$

将（66）应用于（58），我们可得，对 $k \in \mathbb{Z}_{\leq 0}$ 和 $T' = T-4 > 1$,

$$\begin{aligned}\mathbb{E}[\mathrm{Reg}(T)] &\geq \Delta_2\frac{\alpha_1}{4\gamma}\left(1-\left(\frac{m\kappa}{\mu_2}\right)^{2\alpha_1}\right)\left(\sqrt{T'}-\frac{1}{\sqrt{T'}}\right) \\ &= \mathcal{O}(\sqrt{T}).\end{aligned}$$

注意到由 $m = \frac{\gamma}{\gamma-1} = \frac{\lceil\alpha_2\rceil}{\lceil\alpha_2\rceil-1}$ 的定义可知，$m\frac{\kappa}{\mu_2} = m\left(1-\frac{1}{\alpha_2}\right) < 1$ 成立。因此，在采用先验分布 $k \in \mathbb{Z}_{\leq 0}$ 的 STS 下，存在一个常数 $C'(\alpha_1, \alpha_2) > 0$，使得

$$\liminf_{T\to\infty}\frac{\mathbb{E}[\mathrm{Reg}(T)]}{\sqrt{T}} \geq C'(\alpha_1, \alpha_2).$$

□

E 先验分布与后验分布

本节详细介绍多参数模型下的杰弗里斯先验和概率匹配先验问题。更多细节可参见本节所引文献。

E.1 存在干扰参数时杰弗里斯先验的问题 s

杰弗里斯先验被定义为与费舍尔信息矩阵行列式的平方根成正比，从而在对参数进行任意一一重参数化后仍保持不变 [Jeffreys, 1998]。然而，众所周知，当模型包含干扰参数时，杰弗里斯先验会面临诸多问题 [Datta 和 Ghosh, 1995, Ghosh, 2011]。因此，对于

多参数模型 [Berger 和 Bernardo， 1992]。例如，对于位置—尺度族，Jeffreys 建议使用其他先验，这些先验与精确匹配先验相同 [DiCiccio 等人， 2017]。

如正文所述，已知在 Neyman-Scott 问题中，杰弗里斯先验会导致方差的估计量不一致 [参见 Berger 和 Bernardo，1992，例 3。]。另一个例子是 Stein 的例子 [Stein， 1959], 其中考虑了具有共同方差的高斯分布模型。在该例中，杰弗里斯先验导致了不令人满意的后验分布，因为在二次损失下，基于杰弗里斯先验的广义贝叶斯估计量被其他估计量支配 [参见 Robert 等，2007，例 3.5.9。]。注意，Bernardo [1979] 证明了参考先验不存在此类问题，这也解释了为什么在多参数赌博机问题中，参考先验的表现优于杰弗里斯先验。

E.2 概率匹配先验

概率匹配先验是一类非信息先验，旨在实现贝叶斯区间估计的覆盖概率与频率学派区间估计的覆盖概率之间的统一 [Tibshirani，1989，Welch 和 Peers， 1963]。因此，在概率匹配先验下，某些区间的后验概率与频率学派的覆盖概率完全匹配或渐近匹配。如果某些区间的后验概率与置信区间完全匹配，则称该先验为精确匹配先验。在贝叶斯区间估计与频率学派的覆盖概率无法完全匹配、但差异较小时，称其为 k阶匹配先验。两种概率之间的差异由余项度量，通常记为 $\mathcal{O}(n^{-\frac{k}{2}})$，其中$n$ 为样本量， k 为匹配的阶数[1]。

例如，设 $\theta \in \mathbb{R}_+$ 为感兴趣参数。对于某些先验分布 $\pi(\theta)$，在观测到 n 个样本 X_n 后，令 $\psi(\theta|X_n)$ 表示后验分布。然后，对于任意 $\alpha \in (0, 1)$，定义 $\theta_\alpha > 0$ 使得

$$\int_0^{\theta_\alpha} \psi(\theta|X_n)\mathrm{d}\theta = \alpha.$$

当 $\pi(\theta)$ 是二阶概率匹配先验时，有

$$\mathbb{P}[\theta \leq \theta_\alpha|X_n] = \alpha + \mathcal{O}(n^{-1}).$$

当 $\pi(\theta)$ 是精确概率匹配先验时，有

$$\mathbb{P}[\theta \leq \theta_\alpha|X_n] = \alpha.$$

有关更多详细信息，请参阅 Datta 和 Sweeting [2005] 以及Ghosh [2011] 及其中的参考文献。

E.3 后验分布的推导细节

本节将详细推导后验分布。

设某个臂的观测为 $\boldsymbol{r} = (r_1, \ldots, r_n)$，且 $q(n) = \sum_{s=1}^n \log r_s$。那么，根据贝叶斯定理，后验概率密度为

$$p(\kappa, \alpha \mid \boldsymbol{r}) = \frac{p(\boldsymbol{r}|\kappa, \alpha)p(\kappa, \alpha)}{\int_0^\infty \int_0^\infty p(\boldsymbol{r} \mid \kappa, \alpha)p(\kappa, \alpha)\mathrm{d}\kappa\mathrm{d}\alpha},$$

[1] 一些论文将先验称为 k阶匹配先验，当余项为 $\mathcal{O}\left(n^{-\frac{k+1}{2}}\right)$ [DiCiccio 等人， 2017]。本文遵循Mukerjee 和 Ghosh [1997] 以及Datta 和 Sweeting [2005]所使用的符号说明。

其中

$$p(\boldsymbol{r} \mid \kappa, \alpha) = \alpha^n \kappa^{n\alpha} \left(\prod_{s=1}^{n} r_{a,s} \right)^{-\alpha-1} \mathbb{1}[\kappa \le \hat{\kappa}(n)]$$
$$= \alpha^n \kappa^{n\alpha} \exp(-q(n)(\alpha+1)) \mathbb{1}[\kappa \le \hat{\kappa}(n)].$$

通过对给定的先验 $k \in \mathbb{Z}$直接计算，我们得到

$$\begin{aligned}
\int_0^\infty \int_0^\infty p(\boldsymbol{r} \mid \kappa, \alpha) p(\kappa, \alpha) \mathrm{d}\kappa \mathrm{d}\alpha &= \int_0^\infty \int_0^\infty p(\boldsymbol{r} \mid \kappa, \alpha) \frac{\alpha^{-k}}{\kappa} \mathrm{d}\kappa \mathrm{d}\alpha \\
&= \int_0^\infty \alpha^{n-k} \exp(-q(n)(\alpha+1)) \int_0^{\hat{\kappa}} \kappa^{n\alpha-1} \mathrm{d}\kappa \mathrm{d}\alpha \\
&= \int_0^\infty \frac{\alpha^{n-k-1}}{n} e^{-q(n)} \exp(-\alpha(q(n) - n \log \hat{\kappa})) \mathrm{d}\alpha \\
&= \frac{\Gamma(n-k)}{n} \frac{e^{-q(n)}}{(q(n) - n \log \hat{\kappa})^{n-k}}.
\end{aligned}$$

因此，联合后验概率密度如下：

$$p(\kappa, \alpha \mid \boldsymbol{r}) = \frac{n[q(n) - n \log \hat{\kappa}(n)]^{n-k}}{\Gamma(n-k)} \alpha^{n-k} \kappa^{n\alpha-1} e^{-q(n)\alpha} \mathbb{1}[0 < \kappa \le \hat{\kappa}(n)],$$

由此得到 α 的边际后验分布为

$$p(\alpha \mid \boldsymbol{r}) = \frac{\alpha^{n-k-1}[q(n) - n \log \hat{\kappa}(n)]^{n-k}}{\Gamma(n-k)} e^{-\alpha(q(n) - n \log \hat{\kappa}(n))}. \tag{67}$$

因此，从边际后验分布中生成的样本 $\tilde{\alpha}$ 实际上服从形状参数为$n-k$、速率参数为 q的伽马分布（n）$-n$l og $\hat{\kappa}(n) = {}^{n}\hat{\alpha}$，即 $\tilde{\alpha} \sim$厄朗分布 $\left(n - k, \ {}^{n}\hat{\alpha}\right)$ 当 $n \in \mathbb{N}$时，以及当$n > k$时 $k \in \mathbb{Z}$。给定 $\tilde{\alpha}$时， κ 的条件后验分布为

$$\begin{aligned}
p(\kappa \mid \boldsymbol{r}, \alpha) &= \frac{p(\kappa, \alpha \mid \boldsymbol{r})}{p(\alpha \mid \boldsymbol{r})} \\
&= \frac{n\alpha}{\hat{\kappa}^{n\alpha}} \kappa^{n\alpha-1} \mathbb{1}[0 < \kappa \le \hat{\kappa}(n)].
\end{aligned} \tag{68}$$

因此，给定 α 时， κ 的累积分布函数（CDF）为

$$\mathbb{P}(\kappa \le x) = F(x \mid \boldsymbol{r}, \alpha = \tilde{\alpha}) = \left(\frac{x}{\hat{\kappa}(n)} \right)^{n\tilde{\alpha}}, \ 0 < x \le \hat{\kappa}(n). \tag{69}$$

的最大似然估计量 α 等价于最大后验（MAP）估计量，前提是采用杰弗里斯先验 [Li 等人，2022，Sun 等人， 2020]。

总之，在上述先验分布下，我们考虑 α上的边缘后验分布

$$p(\alpha \mid \boldsymbol{r}) = \mathrm{Erlang}\left(n-k, \frac{n}{\hat{\alpha}}\right)$$

以及 κ的条件后验分布的累积分布函数（CDF）

$$F(x \mid \boldsymbol{r}, \alpha = \tilde{\alpha}) = \left(\frac{x}{\hat{\alpha}(n)} \right)^{n\tilde{\alpha}}, \ 0 < x \le \hat{\kappa}(n).$$

请注意，我们要求最多进行 $\{2, k+1\}$ 次初始轮次，以避免出现不适当后验分布和非正规最大似然估计量。

F 技术引理

本节将介绍一些用于证明主要引理的技术引理。

引理 17. 设 Z 为服从自由度为$2n$的卡方分布的随机变量。那么，对于任意 $x \in (0,1)$

$$\mathbb{P}[Z \leq 2nx] \leq e^{-nh(x)},$$

其中 $h(x) = (x - 1 - \log x) \geq 0.$

证明。 设 X_i 为服从标准正态分布的随机变量，因此 $Z = \sum_{i=1}^{2n} X_i^2$成立。由引理20可得

$$\mathbb{P}[Z \leq 2nx] = \mathbb{P}\left[\frac{1}{2n}\sum_{i=1}^{2n} X_i^2 \leq x\right] \leq \exp\left\{\left(-2n \inf_{z \leq x} \Lambda^*(z)\right)\right\}.$$

根据矩生成函数的定义，可知

$$\Lambda^*(z) = \sup_{\lambda \in \mathbb{R}} \lambda z - \log \mathbb{E}\left[e^{\lambda X_1^2}\right] = \sup_{\lambda \in \mathbb{R}} \lambda z + \frac{1}{2}\log(1-2\lambda) = \frac{1}{2}(z - 1 - \log z),$$

由此得证。 □

G 已知结果

本节列出一些已证明的引理，后文将直接使用而不再证明。

引理 18 （伯恩斯坦不等式）。 设 X 为一个（σ^2, b）-次指数随机变量，满足 $\mathbb{E}[X] = \mu$ 和 $Var[X] = \sigma^2$，且

$$\mathbb{E}[e^{\lambda X}] \leq \exp\left\{\frac{\lambda^2\sigma^2}{2}\right\} \quad for\ |\lambda| \leq \frac{1}{b}.$$

设 X_i 相互独立且服从（σ^2, b）-次指数分布。则有

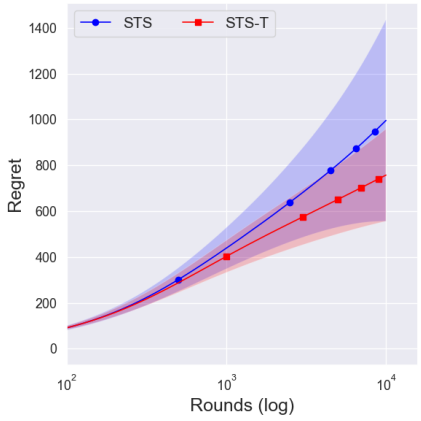
$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{s=1}^{n} X_i - \mu\right| \geq t\right) \leq 2\exp\left(-\frac{n}{2}\min\left\{\frac{t^2}{\sigma^2}, \frac{t}{b}\right\}\right).$$

更多细节请参阅 Vershynin [2018]。

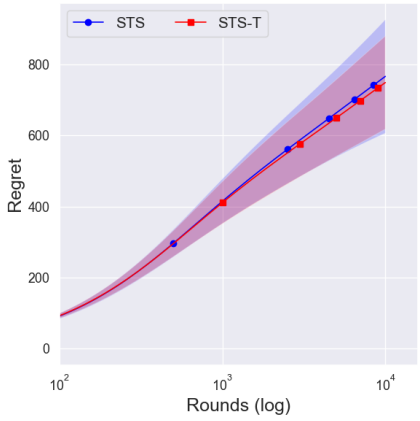
引理 19 （Wallace [1959]中的定理 4.1）。 设 F_n 为具有 n 个自由度的卡方分布的分布函数。对于所有 $t > n$、所有 $n > 0$，且当 $w(t) = \sqrt{t - n - n\log(t/n)}$，

$$1 - F_n(t) < \frac{d_n}{2}\mathrm{erfc}\left(\frac{w(t)}{\sqrt{2}}\right),$$

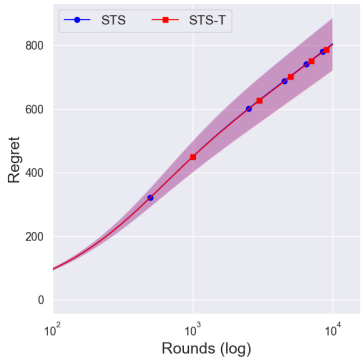
其中 $d_n = \frac{\left(\frac{n}{2}\right)^{\frac{n-1}{2}} e^{-\frac{n}{2}}\sqrt{2\pi}(n/2)}{\Gamma}$且 互补误差函数$(\cdot)$ 是互补误差函数。



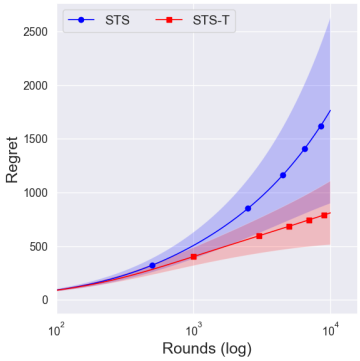
(a) 杰弗里斯先验 $k = 0$



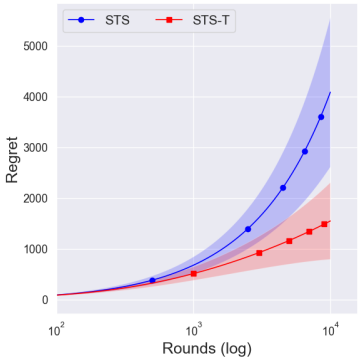
(b) 参考先验 $k = 1$



(c) 带有 $k = 3$ 的先验



(d) 带有 $k = -1$ 的先验



(e) 带有 $k = -3$ 的先验

图4：实线表示基于 100,000 次独立运行得到的平均遗憾。阴影区域表示四分之一标准差。

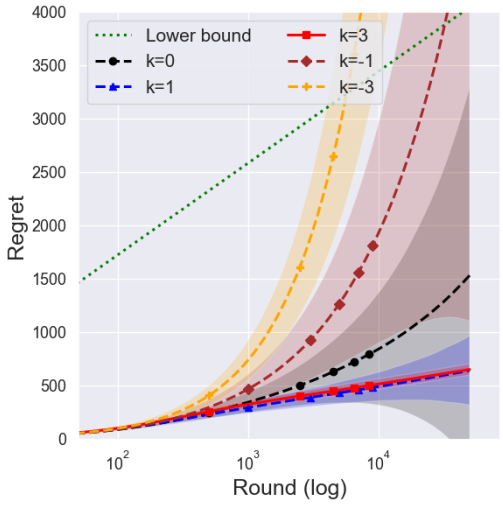
引理 20 (克拉默定理)。 设 $X_1, \ldots, X_n$ 是 $\mathbb{R}$ 上的独立同分布随机变量。那么，对于任意凸集 $C \in \mathbb{R}$,

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} X_i \in C\right] \leq \exp\left\{\left(-n \inf_{z \in C} \Lambda^*(z)\right)\right\},$$

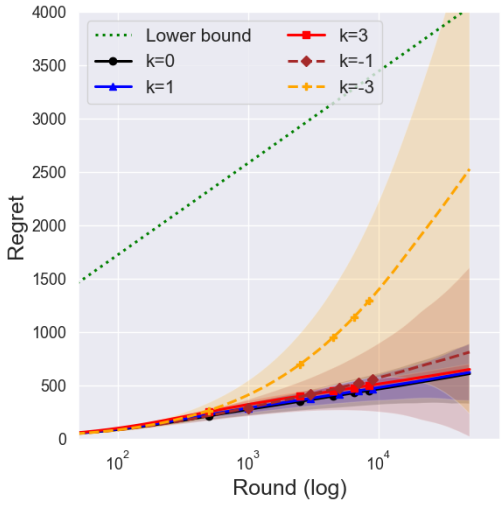
其中 $\Lambda^*(z) = \text{上确界}_{\lambda \in \mathbb{R}}\, \lambda z - \text{对数}\, \mathbb{E}[e^{\lambda X_1}]$。

引理 21 （Korda 等人 [2013]中的**A 项结果**）。当在尺度参数已知的帕累托分布下使用杰弗里斯先验作为先验分布时，对于足够小的 $\epsilon > 0$,

$$\sum_{t=1}^{T} \mathbb{E}\left[\mathbb{1}[j(t) \neq 1, \mathcal{M}_\epsilon^c(t)]\right] \leq \mathcal{O}\left(\epsilon^{-1}\right).$$



(a) 采用不同 k 时 STS 的累积遗憾（ $\boldsymbol{\theta}_4'$下）



(b) 采用不同 k 时 STS-T 的累积遗憾（在 下） $\boldsymbol{\theta}_4'$

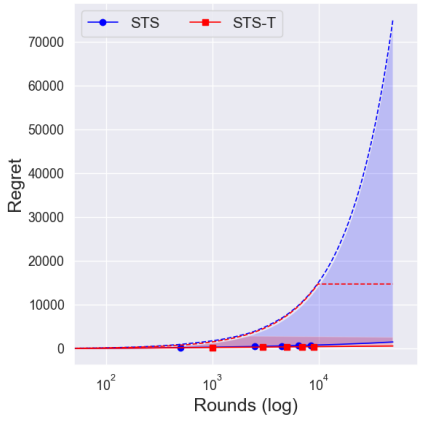
图5：实线表示在 100,000 次独立运行中，能够达到最优下界的先验分布的平均累积遗憾（3）。虚线表示无法达到最优下界的先验分布的平均累积遗憾（3）。绿色点线表示基于引理 1 的问题相关下界。

H 其他实验结果

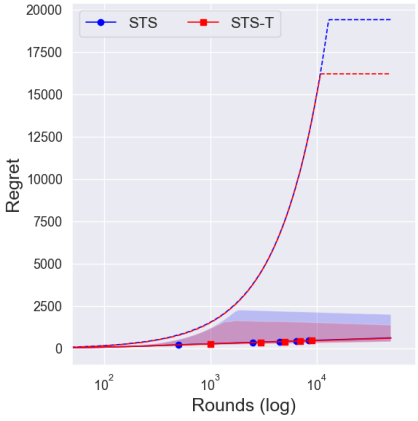
从图 4可以看出，STS 与 STS-T 之间的性能差异会随着 k减小而增大。由于截断过程旨在避免在 STS 使用先验分布$k \in \mathbb{Z}_{\leq 1}$时可能出现的极端情况，因此 STS 与 STS-T 使用先验分布 $k = 3$时没有差异，这是很自然的。我们还可以看到，STS-T 的改进在 k 减小时十分显著，此时极端情况很容易发生。

我们进一步考虑另一个四臂老虎机问题 $\boldsymbol{\theta}_4'$，其中 $\boldsymbol{\kappa} = (1.0, 1.5, 2.0, 2.0)$， $\boldsymbol{\alpha} = (1.2, 1.5, 1.8, 2.0)$，且 $\boldsymbol{\mu} = (5.0, 4.5, 4.5, 4.0)$。 $\boldsymbol{\theta}_4'$从以下意义上来说会比 $\boldsymbol{\theta}_4$更具挑战性： κ 决定了支撑集的左边界，其中较大的 κ 意味着臂的最小值更大。因此，如果次优臂的 κ 大于最优臂的相应值，那么在最初几次试验中，问题会变得更加困难。图5和图6展示了时间范围为 T =50,000、独立运行 10,000 次时的数值结果。尽管采用参考先验的 STS 与采用保守先验的 STS $k = 3$表现相似，但其性能变化很大。

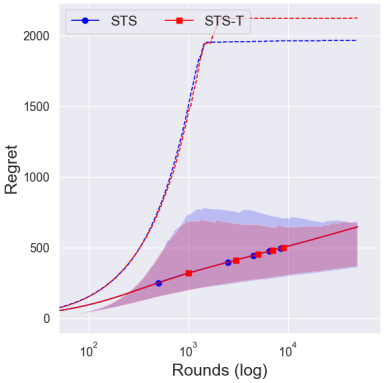
图 6(a) 和 6(b) 展示了截断过程的有效性，其中STS-T 的上尾 0.05% 遗憾远低于 STS。尽管 $k = -1$ 在图 99%的遗憾中心区间中也显示出巨大改进，如图6(d)所示，但 STS-T 使用 $k = -1$ 的表现较差，相比之下，图 $k \in \mathbb{Z}_{\geq 0}$ 中的先验分布在图 5(b)中表现更好。



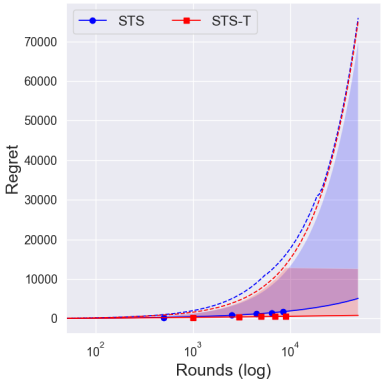
(a) 杰弗里斯先验 $k = 0$



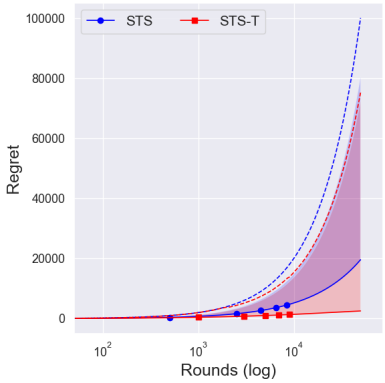
(b) 参考先验 $k = 1$



(c) 使用 $k = 3$ 的先验



(d) 使用 $k = -1$ 的先验



(e) 使用 $k = -3$ 的先验

图6：实线表示独立运行 10,000 次得到的平均遗憾。阴影区域和虚线分别表示中央99%区间和遗憾的上侧 0.05%。