Bandit Algorithms Based on Thompson Sampling for Bounded Reward Distributions

Charles Riou CHARLES.RIOU@POLYTECHNIQUE.ORG
Ecole Polytechnique
RIKEN Center for Advanced Intelligence Project

Junya Honda HONDA@EDU.K.U-TOKYO.AC.JP
The University of Tokyo
RIKEN Center for Advanced Intelligence Project

Editors: Aryeh Kontorovich and Gergely Neu

Abstract

We focus on a classic reinforcement learning problem, called a multi-armed bandit, and more specifically in the stochastic setting with reward distributions bounded in $[0, 1]$. For this model, an optimal problem-dependent asymptotic regret lower bound has been derived. However, the existing algorithms achieving this regret lower bound all require to solve an optimization problem at each step, inducing a large complexity. In this paper, we propose two new algorithms, which we prove to achieve the problem-dependent asymptotic regret lower bound without requiring to solve any optimization problem. The first one, which we call Multinomial TS, is an adaptation of Thompson Sampling for Bernoulli rewards to multinomial reward distributions whose support is included in $\{0, \frac{1}{M}, \dots, 1\}$. This algorithm achieves the regret lower bound in the case of multinomial distributions with the aforementioned support, and it can be easily generalized to bounded reward distributions in $[0, 1]$ by randomly rounding the observed rewards. The second algorithm we introduce, which we call Non-parametric TS, is a randomized algorithm but it is not based on the posterior sampling in the strict sense. At each step, it computes an average of the observed rewards with random weight. Not only is it asymptotically optimal, but also it performs very well even for small horizons. Practically, it beats most state-of-the-art bandit algorithms, including some which require solving an optimization problem at each round.

Keywords: Thompson Sampling, Multi-armed Bandit Problem, Online Optimization

1. Introduction

The sequential decision-making problem, called a multi-armed bandit, consists of sequentially sampling one of unknown random variables called arms. At each round, an agent pulls an arm and gets a reward. The aim of the agent is to maximize the expectation of the sum of its rewards over a horizon T. Therefore, there arises a *regret* of not pulling the optimal arm (that is, the arm with the highest expected reward) every time the agent pulls a suboptimal arm. The aim of this problem can be expressed as the minimization of the total expected regret over the horizon T.

Various algorithms have been derived to the multi-armed bandit problem, which can be gathered in several categories. The following list is not exhaustive but includes some of the most important categories. The first category is derived from the algorithm UCB1 (Upper Confidence bound) (Auer et al., 2002), which relies on computing confidence intervals at each step, and includes many recent advances like the empirical KL-UCB (Cappé et al., 2013), which computes clever confidence

intervals solving a convex optimization problem at each step, or more recently Bootstrapped UCB (Hao et al., 2019), to name only a few. Bootstrapping is also a common technique for bandit problems, like the GIRO (Garbage In, Reward Out) (Kveton et al., 2019). The second category corresponds to a class of algorithms based on solving a convex optimization problem at each step, and includes, among others, DMED (Deterministic Minimum Empirical Divergence) (Honda and Takemura, 2010), IMED (Indexed Minimum Empirical Divergence) (Honda and Takemura, 2015) and the empirical KL-UCB (Cappé et al., 2013). The third category mentioned here relies on Thompson Sampling (TS) (see, for instance, Agrawal and Goyal, 2012; Kaufmann et al., 2012), a Bayesian randomized algorithm which selects an arm with its probability of being the best arm.

For this problem it is known that there exists an asymptotic lower bound on the regret (Lai and Robbins, 1985; Burnetas and Katehakis, 1996) and a policy with a regret matching this lower bound is called asymptotically optimal. However, the current asymptotically optimal policies all require to solve an optimization problem at each round (see, for instance, Cappé et al., 2013; Honda and Takemura, 2010) and are therefore quite slow and sometimes hard to apply online. In the specific case when the rewards are binary, some policies can determine the arm to pull in an analytic way without optimization. However, they are not expected to be optimal in more general cases.

A typical example of application of bandits with nonbinary rewards is the following. Let us assume, for instance, that you are a telecommunication company and that you are trying to provide the best communication possible to your clients. Given that you can transmit messages via two channels A or B and you want to attribute to your clients the channel with the best throughput, which channel should you assign to your clients? This problem can be modeled by a bandit problem with two arms, channel A and channel B. If the throughput of the communication can be measured by a certain number of packets (the higher the number of packets, the better the throughput), then both arms have reward distributions over integers. It can be normalized in order to become an element of a certain set $\{0, \frac{1}{M}, \dots, 1\}$ for a certain integer $M \geq 1$ which is the maximum number of packets. This is a bandit problem where all arms follow a multinomial distribution whose support is $\{0, \frac{1}{M}, \dots, 1\}$. If the throughput of the communication can take more general values, but is bounded, then it can be normalized in $[0, 1]$. This becomes a general bandit problem with reward distributions bounded in $[0, 1]$. In these two models, existing algorithms are either non-optimal, or require the computation of an optimization problem at each round, which is quite problematic, since such settings require quick online decision making.

In this paper, we propose two algorithms adapted respectively to the multinomial case and the case of general distributions over $[0, 1]$, which we prove to be asymptotically optimal. One of the major interests of both algorithms is that, despite their optimality, they do not require solving any optimization problem, which significantly enhances the applicability in online settings. However, they outperform some of the state-of-the-art bandit algorithms which require solving an optimization problem at each step of the algorithm, including the empirical KL-UCB (see Cappé et al., 2013).

Both of the proposed algorithms are based on Thompson Sampling (TS): the first one, which we call Multinomial TS, is an adaptation of TS to the multinomial case, and is asymptotically optimal for multinomial rewards. The second one, which we call Non-parametric TS, is an adaptation of TS in the sense that it is still a randomized algorithm, but it is not a Bayesian algorithm as we explain later. Non-parametric TS is asymptotically optimal for distributions over $[0, 1]$.

Although both algorithms are based on TS, their analysis is definitely non-trivial, for the following reasons. Multinomial TS is a natural adaptation of TS, but the analysis for the binary case (and that for one-parameter exponential families (see Korda et al., 2013)) cannot be used since it

heavily relies on the beta-binomial transform (see Agrawal and Goyal, 2012), which is based on the discussion of an order statistics and is not generalized to the multinomial case in a simple form. Non-parametric TS is not based on the posterior sampling in the strict sense, and is different from naive applications of non-parametric methods for posterior sampling that are computationally expensive. This non-parametric nature further makes the analysis difficult. In fact, the proof of the asymptotic optimality for distributions over $[0, 1]$ of an algorithm called the empirical KL-UCB (Cappé et al., 2013) is very recent (Garivier et al., 2018) because of such difficulty. On top of this difficulty comes the one implied by the randomization. Indeed, for this second reason, most TS-based algorithms do not have a theoretical guarantee. In the case of Non-parametric TS, this second difficulty adds to the first one.

2. Preliminaries

In this section, we formulate the multi-armed bandit problem and introduce the current state-of-the-art results and algorithms used to solve it.

For any distribution F, we will denote the support of F by $\mathrm{supp}(F)$ and $\mathrm{E}[F] := \mathrm{E}_{X\sim F}[X]$ the expectation of any random variable following distribution F. We consider a bandit with K arms, and an agent plays it T times (i.e. pulls an arm for T rounds). Each time the agent pulls arm $k \in [K] = \{1, \dots, K\}$, he/she will receive a reward $r \in [0, 1]$ drawn from distribution F_k. We assume that the received rewards are independent of each other. The aim of this problem is to maximize the expected reward, or in other words, to minimize the expected regret:

$$\mathrm{E}[R_T] := \mathrm{E}\left[\sum_{t=1}^{T}(\mu^* - \mu_{I(t)})\right],$$

where we denoted the expectation of arm k by $\mu_k := \mathrm{E}_{X\sim F_k}[X]$, $I(t)$ the arm selected by the agent at round t and $\mu^* := \max\{\mu_1, \dots, \mu_K\}$.

We consider a problem-dependent setting where K and μ_i are fixed, and T is sufficiently large. It was first proven in Lai and Robbins (1985) for single-parameter models and later in Burnetas and Katehakis (1996), that no strategy could beat systematically the asymptotic regret lower bound:

$$\mathrm{E}[R_T] \geq \sum_{i:\Delta_i>0} \frac{\Delta_i \log T}{\mathcal{K}_{\inf}(F_i, \mu^*)} + o(\log T), \tag{1}$$

where we denoted $\mathcal{K}_{\inf}(F_i, \mu^*) = \inf_{G:\ \mathrm{E}[G]>\mu^*} \mathrm{KL}(F_i\|G)$ and $\mathrm{KL}(F\|G)$ is the Kullback-Leibler divergence between the distributions F and G.

In the case of multinomial rewards of support included in $\{0, \frac{1}{M}, \dots, 1\}$, the optimal asymptotic regret lower bound is given by

$$\mathrm{E}[R_T] \geq \sum_{i:\Delta_i>0} \frac{\Delta_i \log T}{\mathcal{K}^{(M)}_{\inf}(F_i, \mu^*)} + o(\log T), \tag{2}$$

where we denoted $\mathcal{K}^{(M)}_{\inf}(F_i, \mu^*) = \inf\left\{\mathrm{KL}(F\|G) \mid \mathrm{supp}(G) \in \{0, \frac{1}{M}, \dots, 1\},\ \mathrm{E}(G) > \mu^*\right\}$.

When F_i is a Bernoulli distribution, $\mathcal{K}_{\inf}(F_i, \mu^*) = \mu_i \log \frac{\mu_i}{\mu^*} + (1 - \mu_i) \log \frac{1-\mu_i}{1-\mu^*}$ and some existing algorithms achieve the optimal lower bound like Thompson Sampling (see, for instance, Agrawal and Goyal, 2012; Kaufmann et al., 2012) or KL-UCB (Cappé et al., 2013).

Algorithm 1 Multinomial TS

Require: Horizon $T \geq 1$, number of arms $K \geq 1$, support size of the arm distributions $M \geq 1$.
Set $\alpha_m^k := 1$ for $k \in [K]$ and $m \in \{0, \dots, M\}$.
for $t = 1 \dots T$, **do**
  for $k = 1 \dots K$, **do**
    Sample $L_k \sim \mathrm{Dir}(\alpha_0^k, \alpha_1^k, \dots, \alpha_M^k)$.
  $I(t) := \arg\max_{k \in \{1,\dots,K\}} \left\{ \left(0, \frac{1}{M}, \frac{2}{M} \dots 1\right)^\top L_k \right\}$.
  Pull arm $I(t)$ and observe reward $r_t = \frac{m}{M}$ where $m \in \{0, 1, \dots, M\}$.
  Update $\alpha_m^{I(t)} := \alpha_m^{I(t)} + 1$.

In the more general case of reward distributions bounded in $[0, 1]$, some algorithms achieve the optimal regret bound, including DMED (Deterministic Minimum Empirical Divergence) (Honda and Takemura, 2010), IMED (Indexed Minimum Empirical Divergence) (Honda and Takemura, 2015) and the empirical KL-UCB (adaptation of KL-UCB, see Cappé et al., 2013). Still, the empirical KL-UCB algorithm for this model requires a nested optimization that maximizes the objective function represented by $\mathcal{K}_{\inf}$, which itself is expressed by a minimization problem. DMED and IMED algorithms do not require such a nested optimization, but they still need the computation of $\mathcal{K}_{\inf}$ at each round.

3. Proposed Algorithms

In this section, we propose two algorithms for the multi-armed bandit problem, based on Thompson sampling, that do not require solving an optimization problem, but instead generates samples from Dirichlet distributions, which are generalizations of beta distributions. We denote the Dirichlet distribution of parameters $(\alpha^1, \dots, \alpha^n)$ by $\mathrm{Dir}(\alpha^1, \dots, \alpha^n)$, whose density function is given by $\frac{\Gamma(\sum_{i=1}^n \alpha^i)}{\prod_{i=1}^n \Gamma(\alpha^i)} \prod_{i=1}^n x_i^{\alpha^i - 1}$ for $(x_1, \dots, x_n) \in [0, 1]^n$ such that $\sum_{i=1}^n x_i = 1$.

The first algorithm, Multinomial TS, is a simple adaptation of TS to multinomial distributions, which is given in Algorithm 1. TS for the binary case, which we will denote Binary TS, is a Bayesian algorithm which generates samples from a beta distribution, which is the conjugate of the Bernoulli distribution. In our case, we generate samples from a Dirichlet distribution, which is the conjugate of a multinomial distribution, and we expect it to be optimal in the case of multinomial arms.

At each round, instead of solving an optimization problem, Multinomial TS generates samples L_k for $k \in [K]$ from Dirichlet distributions of dimension $M + 1$, each L_k corresponding to the posterior sample on the parameter of the multinomial distribution of arm k. In fact, in the case where the reward distribution is not multinomial, we have to choose a parameter M and use the rounding technique explained below in the Remark 1. However, the performance of Multinomial TS depends on the choice of M. If M is very large, the asymptotic (in T) expected regret will be better, however, the algorithm will be slightly slower. On the other hand, if M is very large, the asymptotic regret will approach the optimal regret bound. In addition, Multinomial TS is a Bayesian algorithm, and more parameters are involved in the Dirichlet distribution as M becomes large. As a result, many rounds are needed for the posterior distributions to concentrate on the actual parameters of the arms for large M. Therefore, for small horizons T, M should also chosen to be small.

Algorithm 2 Non-parametric TS

Require: Horizon $T \geq 1$, number of arms $K \geq 1$.
for $k = 1 \dots K$, **do**
  Set $X_k := 1$ and $N_k := 1$.
for $t = 1 \dots T$, **do**
  for $k = 1 \dots K$, **do**
    Sample $L_k = \mathrm{Dir}(1_{N_k})$ where $1_{N_k} = \underbrace{(1, \dots, 1)}_{N_k \text{ elements}}$.
    $V_k := X_k^\top L_k$.
  $I(t) := \arg\max_{k \in \{1,\dots,K\}} \{V_k\}$.
  Pull arm $I(t)$ and observe reward r_t.
  Update $X_{I(t)} := (X_{I(t)}^\top, r_t^{I(t)})^\top$.
  Update $N_{I(t)} := N_{I(t)} + 1$.

Remark 1 The algorithm Multinomial TS can be used even if the arms do not follow multinomial distributions. A possible adaptation is the randomized rounding of the reward discussed in Agrawal and Goyal (2012) for the adaptation of Binary TS to more general bounded reward distributions: if $r \in \left[\frac{m}{M}, \frac{m+1}{M}\right)$ for some $m \in \{0, \dots, M\}$, then we generate a random variable $\tilde{r} = \frac{m+B}{M}$ for Bernoulli random variable $B \in \{0, 1\}$ with success probability $Mr - m$, and we regard $\tilde{r}$ as the observed reward instead of r. The generated (virtual) reward $\tilde{r} \in (r - \frac{1}{M}, r + \frac{1}{M})$ has the same expectation as r and follows a multinomial distribution over $\{0, \frac{1}{M}, \dots, 1\}$.

The second algorithm, Non-parametric TS, is given in Algorithm 2 which is a randomized algorithm like TS, but it is not a Bayesian algorithm in the strict sense. Whereas Multinomial TS samples a probability distribution over $\{0, \frac{1}{M}, \frac{2}{M}, \dots, 1\}$, Non-parametric TS samples, for any $k \in [K]$, a distribution L_k over $\{1, X_1^k, X_2^k, \dots, X_{N_k}^k\}$ where N_k is the number of times arm k has been pulled so far, X_i^k is the i-th reward observed from arm k and $L_k \sim \mathrm{Dir}(1, \dots, 1)$ is the uniform distribution on the probability simplex of dimension N_k. This means that the support of the sampled distribution of Non-parametric TS depends on the observed reward, which means that the sampled distribution is not a posterior sample with respect to a fixed prior distribution.

By investigating the property of Dirichlet distributions, we can see that Non-parametric TS coincides with Multinomial TS with support $\{0, \frac{1}{M}, \frac{2}{M}, \dots, 1\}$, when the reward follows a multinomial distribution over $\{0, \frac{1}{M}, \frac{2}{M}, \dots, 1\}$. However, in the case of Non-parametric TS, the prior is $\mathrm{Dir}(0, 0, ..., 0, 1)$. This is an improper prior and this impropriety also makes the analysis complicated but contributes to the asymptotic optimality.

It should be noted that if the observed rewards of an arm are $r_1, \dots, r_n$ we compute the random weighted average of $1, r_1, \dots, r_n$ with weight sampled from $\mathrm{Dir}(1, \dots, 1)$. Here, 1 in the support is important to create the exploration; without 1 in the support, the randomized average of each arm will be at most the maximum of the rewards observed so far. This is problematic because, for example, if the first reward X_1^1 of the arm 1 is unluckily smaller than any point on the support $\mathrm{supp}(F_2)$ of the arm 2 then the first arm will never be pulled again even if $\mu_1 > \mu_2$.

Remark 2 By the same discussion as the relation between Multinomial TS and Non-parametric TS, we see that $(1, X_1^k, X_2^k, \dots, X_{N_k}^k)^\top L_k$ for $L_k \sim \mathrm{Dir}(1_{N_k+1})$ has the same distribution as

Algorithm 3 Improved version of Non-parametric TS

Require: Horizon $T \geq 1$, number of arms $K \geq 1$.
for $k = 1 \dots K$, **do**
  Set $S_k := 1$ and $T_k := 1$.
for $t = 1 \dots T$, **do**
  for $k = 1 \dots K$, **do**
    Sample $L_k = \mathrm{Dir}(T_k)$.
    $V_k := S_k^\top L_k$.
  $I(t) := \arg\max_{k \in \{1,\dots,K\}} \{V_k\}$.
  Pull arm $I(t)$ and observe reward r_t.
  if $r_t = S_{I(t)}[m]$ *for some* m, **then**
    Update $T_{I(t)}[m] := T_{I(t)}[m] + 1$.
  else
    Update $S_{I(t)} := (S_{I(t)}^\top, r_t)^\top$ and $T_{I(t)} := (T_{I(t)}^\top, 1)^\top$.

$S_k^\top L_k$ for $L_k \sim \mathrm{Dir}(T_k)$, where $S_k = (S_k[1], S_k[2], \dots, S_k[s])$ is the set of non-identical elements of $(1, X_1^k, X_2^k, \dots, X_{N_k}^k)$ and $T_k = (T_k[1], T_k[2], \dots, T_k[s])$ is the set consisting of the number $T_k[i]$ of occurrence of element $S_k[i]$ in $(1, X_1^k, X_2^k, \dots, X_{N_k}^k)$. Therefore, sampling the latter one instead of the former one in Non-parametric TS does not affect the expected regret bound. Nevertheless, in some cases, the complexity of the algorithm is considerably reduced. For example, in the multinomial case, the complexity of the algorithm at round t is reduced from $O(t)$ to $O(Ks)$. The pseudo-code of the improved version of Non-parametric TS is given in Algorithm 3.

Remark 3 We can also see from the proof of the regret bound for Non-parametric TS that the theoretical analysis is largely the same (or simpler) even if we replace the posterior sample V_k with the empirical mean $\frac{1}{N_k}\sum_{i=1}^{N_k} X_i^k$ for k maximizing the empirical mean, though we do not give the formal analysis for this modification. This replacement reduces the complexity to from $O(t)$ to $O(K \log t)$ since each suboptimal arm is pulled at most $O(\log t)$ rounds.

4. Main Results

The two main results of this paper are that Multinomial TS achieves the optimal regret bound in (1) for the multinomial case and Non-parametric TS achieves the optimal regret bound in (2) for general distributions over $[0, 1]$.

Theorem 4 *Assume there are K multinomial arms of common support $\{0, \frac{1}{M}, \frac{2}{M}, \dots, 1\}$ where $M \geq 1$ is a natural number. Then, Multinomial TS achieves the optimal multinomial regret bound:*

$$\mathrm{E}[R_T] \leq \sum_{k:\Delta_k > 0} \frac{\Delta_k \log T}{\mathcal{K}_{\inf}^{(M)}(F_i, \mu^*)} + o(\log T).$$

Theorem 5 *Assume there are K arms whose distributions are supported on $[0, 1]$. Then, Non-parametric TS achieves the optimal regret bound:*

$$\mathrm{E}[R_T] \leq \sum_{k:\Delta_k > 0} \frac{\Delta_k \log T}{\mathcal{K}_{\inf}(F_i, \mu^*)} + o(\log T).$$

As discussed in Remark 1, we can use Multinomial TS by using the randomized rounding even if the reward is not supported on $\{0, \frac{1}{M}, \frac{2}{M}, \dots, 1\}$. The loss of the asymptotic regret by this rounding can be evaluated by the following lemma, the proof of which is given in Appendix I.

Lemma 6 *Let $X \in [0,1]$ be a random variable following distribution F and define $\tilde{F}$ as the distribution of $\tilde{X} := \frac{\lfloor MX \rfloor + B}{M}$ for an integer $M \geq 1$, where $B \in \{0,1\}$ is a Bernoulli random variable with success probability $MX - \lfloor MX \rfloor$ given X. Then, for $M > \frac{1}{1-\mu}$ we have*

$$\mathcal{K}_{\inf}(F,\mu) - \frac{1}{M(1-\mu)-1} \leq \mathcal{K}_{\inf}^{(M)}(\tilde{F},\mu) \leq \mathcal{K}_{\inf}(F,\mu). \tag{3}$$

By Lemma 6, we know that the main term in Multinomial TS gets closer to the optimal bound when M gets larger with rate $O(1/M)$. However, increasing M has a cost. Indeed, as for many Bayesian algorithms convergence, the regret of Multinomial TS can be decomposed into two phases: a pre-convergence phase and a post-convergence phase. The former is the phase during which the estimated parameters are converging towards the ones of the multinomial distribution. The latter is the phase after the estimated parameters have converged towards the true parameters of the distribution. If you increase M, then the exploration term in the post-convergence will be smaller, but it will take more time for the parameters to converge, so the pre-convergence phase will be longer. Therefore, as we can see from the proof of Theorem 4, the logarithmic term will be smaller but the constant term (hidden in the $o(\log T)$) will be larger and the algorithm will be worse for smaller values of T.

Non-parametric TS does not suffer from such a tradeoff, and its bound is optimal. In addition, as it is not a Bayesian algorithm in the strict sense, which estimate parameters and generate a distribution whose parameters are the ones estimated before. Instead, Non-parametric TS generates a distribution, relying directly on the observed rewards. This prevents typical errors of Bayesian algorithms due to early-stage estimation, and for this reason, Non-parametric TS performs also well for small values of T. However, this has a cost: the complexity of remembering all the rewards and sampling a growing dimension Dirichlet distribution at each round. Nevertheless, this problem can be avoided in certain cases, see Remark 2.

In the multinomial case, both algorithms coincide, except for the initialization of the weights. Multinomial TS initializes all the parameters to $p_m = 1$, while Non-parametric TS initializes $p^m = 0$ for $m \neq M$ and $p^M = 1$.

Both proofs have similar outlines, decomposed into a pre-convergence and a post-convergence, which we will present, before introducing the technical lemmas which lead to the proof.

General Outline of the Proof

We want to provide an upper bound on the expected regret

$$\mathrm{E}[R_T] = T\mu^* - \mathrm{E}\left[\sum_{t=1}^{T} \mu_{I(t)}\right].$$

Then, letting $N_i(T)$ be the number of times that arm i is pulled and $\Delta_i := \mu^* - \mu_i$ be the gap of arm i, we can rewrite the regret as

$$\mathrm{E}[R_T] = \sum_{i=1}^{K} (\mu^* - \mu_i)\mathrm{E}[N_i(T)] = \sum_{i=1}^{K} \Delta_i \mathrm{E}\left[\sum_{t=1}^{T} \mathbb{1}(I(t) = i)\right].$$

To derive the upper bound we want, it is enough to prove that, for any suboptimal arm i, the regret related to arm i satisfies

- $\mathrm{E}\left[\sum_{t=1}^{T} \mathbb{1}(I(t)=i)\right] = \frac{\log T}{\mathcal{K}_{\inf}^{(M)}(F_i,\mu^*)} + o(\log T)$ for Multinomial TS,
- $\mathrm{E}\left[\sum_{t=1}^{T} \mathbb{1}(I(t)=i)\right] = \frac{\log T}{\mathcal{K}_{\inf}(F_i,\mu^*)} + o(\log T)$ for Non-parametric TS.

To do that, we are going to introduce a distance d on distributions for each case. In the case of Multinomial TS, this distance is on the set of multinomial distributions whose support is included in $\{0, \frac{1}{M}, \dots, 1\}$. If D_1 and D_2 are two distributions of respective parameters $p_0, p_1, \dots, p_M$ and $q_0, q_1, \dots, q_M$, then we use the distance defined as

$$d(D_1, D_2) := \|p - q\|_\infty = \sup_{i \in \{0,\dots,M\}} |p_i - q_i|,$$

that is, the L^∞ distance between p and q in $\mathbb{R}^{M+1}$. In the case of Non-parametric TS, we use the Lévy distance for d. Recall that the Lévy distance between two cumulative distribution functions on $[0,1]$ F and G is defined by

$$D_L(F,G) = \inf\{\epsilon > 0 :\ \forall x \in [0,1],\ F(x-\epsilon) - \epsilon \leq G(x) \leq F(x+\epsilon) + \epsilon\}.$$

For more clarity, we use the following notations within the proof. We denote the true parameters of arm j by $p^j = (p_0^j, \dots, p_M^j)$, that is, $p_i^j = P_{X \sim F_j}[X = \frac{i}{M}]$. We denote the parameters of the posterior distribution of arm j by $\alpha^j(t) = (\alpha_0^j(t), \dots, \alpha_M^j(t))$. When there is no ambiguity on the arm due to the context, for the sake of clarity, we drop the superscript and simply denote $\alpha = (\alpha_0, \dots, \alpha_M)$ the parameters of the arm we are studying.

For $k \in [K]$, we denote by $V_k(t)$ the mean of the reward distribution of arm k sampled from the Dirichlet distribution at step t, that is, $V_k(t) = u^\top L_k$ for $u = \left(0, \frac{1}{M}, \dots, 1\right)$ in Multinomial TS and $V_k(t) = X_k^\top L_k$ in Non-parametric TS. We also denote $\hat{F}_k(t)$ the empirical cumulative distribution function of arm $k \in [K]$ at step t.

Let us decompose the regret related with suboptimal arm i into two terms:

$$\mathrm{E}\left[\sum_{t=1}^{T} \mathbb{1}(I(t)=i)\right] = \underbrace{\mathrm{E}\left[\sum_{t=1}^{T} \mathbb{1}\left(I(t)=i, V_i(t) \geq \mu^* - \epsilon_1, d(\hat{F}_{I(t)}(t), F_{I(t)}) \leq \epsilon_2\right)\right]}_{\text{(Post-CV)}}$$
$$+ \underbrace{\mathrm{E}\left[\sum_{t=1}^{T} \mathbb{1}\left(I(t)=i, \{V_i(t) < \mu^* - \epsilon_1 \cup d(\hat{F}_{I(t)}(t), F_{I(t)}) > \epsilon_2\}\right)\right]}_{\text{(Pre-CV)}}.$$

Using the notations previously introduced, rewriting both the terms (Post-CV) and (Pre-CV) in the case of Multinomial TS, we have

$$\begin{cases} \text{(Post-CV)} = \mathrm{E}\left[\sum_{t=1}^{T} \mathbb{1}\left(I(t)=i, u^\top L_{I(t)}(t) \geq \mu^* - \epsilon_1, \left\| \frac{\alpha^{I(t)}(t)}{N_{I(t)}(t)+M+1} - p^{I(t)} \right\|_\infty \leq \epsilon_2\right)\right], \\ \text{(Pre-CV)} = \mathrm{E}\Big[\sum_{t=1}^{T} \mathbb{1}\Big(I(t)=i, \\ \qquad\qquad \left\{u^\top L_{I(t)}(t) < \mu^* - \epsilon_1 \cup \left\| \frac{\alpha^{I(t)}(t)}{N_{I(t)}(t)+M+1} - p^{I(t)} \right\|_\infty > \epsilon_2\right\}\Big)\Big]. \end{cases}$$

In the case of Non-parametric TS, we have

$$\begin{cases} \text{(Post-CV)} = \mathrm{E}\left[\sum_{t=1}^{T} \mathbb{1}\left(I(t)=i, V_{I(t)}(t) \geq \mu^* - \epsilon_1, D_L(\hat{F}_{I(t)}(t), F_{I(t)}) \leq \epsilon_2\right)\right], \\ \text{(Pre-CV)} = \mathrm{E}\left[\sum_{t=1}^{T} \mathbb{1}\left(I(t)=i, \left\{V_{I(t)}(t) < \mu^* - \epsilon_1 \cup D_L(\hat{F}_{I(t)}(t), F_{I(t)}) > \epsilon_2\right\}\right)\right]. \end{cases}$$

The first term in the RHS is the post-convergence term, while the second one is the pre-convergence term. The post-convergence term is the exploration term, and thus the main term of the regret. The pre-convergence term is the regret implied before the algorithm seizes the true parameters of the arms. The proof of Theorem 4 relies on the following two propositions.

Proposition 7 *For Multinomial TS, we have, for any* $\epsilon_0 > 0$

$$\text{(Post-CV)} \leq \frac{(1+\epsilon_0)\log T}{\mathcal{K}_{\inf}^{(M)}(F_i, \mu^*)} + o(\log T).$$

Proposition 8 *For Multinomial TS, we have*

$$\text{(Pre-CV)} = O(1).$$

The key to those two propositions is to provide an upper and a lower bound on the probability $P_{L\sim\mathrm{Dir}(\alpha)}(L \in S)$ for a certain set $S = \{x \in P : \ u^\top x \geq \mu\}$ or $S = \{x \in P : \ u^\top x \leq \mu\}$ included in the probability simplex P, where $\mu \in [0,1]$. Those results, stated in Lemmas 13 and 14, are the following.

Lemma 9 *Assume that* $1^\top \alpha = N$ *and for any* $j \in \{0, 1, \ldots, M\}, \alpha_j \geq 1$. *We will denote* $P_\alpha = \frac{1}{N}\alpha$. *Let* $S \subset P$, *a closed convex set included in the probability simplex, and denote* $P^* := \arg\min_{x\in S} \mathrm{KL}(P_\alpha \| x)$. *Then, the following upper bound holds.*

$$P_{L\sim\mathrm{Dir}(\alpha)}(L \in S) \leq C_1 N^{M/2} \exp(-N\mathrm{KL}(P_\alpha \| P^*)).$$

In the particular case $S = \{x \in P : \ u^\top x \geq \mu\}$ *with* $\mu \geq u^\top \alpha$, *the following lower bound also holds:*

$$P_{L\sim\mathrm{Dir}(\alpha)}\left(L \in S\right) \geq C_2 N^{-\frac{M}{2}} \exp(-N\mathrm{KL}(P_\alpha \| P^*)) \frac{P_{\alpha_M}}{P^*_M},$$

where we denoted $C_1 := \frac{e^{1/12}}{\Gamma(M+1)} \left(\frac{1}{\sqrt{2\pi}}\right)^M$ *and* $C_2 := \left(\frac{1}{\sqrt{2\pi}}\right)^M e^{-(M+1)/12}$.

Those results provide exponential upper and lower bounds to an end-tail probability.

The proof of Theorem 5 relies on the following two propositions.

Proposition 10 *For Non-parametric TS, we have, for any* $\epsilon_0 > 0$

$$\text{(Post-CV)} \leq \frac{\log T}{\mathcal{K}_{\inf}(F_i, \mu^*) - \epsilon_0} + o(\log T).$$

Proposition 11 *For Non-parametric TS, we have*

$$\text{(Pre-CV)} = O(1).$$

Several lemmas using different proof techniques are used in the proof of those two propositions. To be more specific, the proofs are based on lower and upper bounds on the probabilities related to Dirichlet distributions in Lemmas 15 and 17, in addition to the previous lemma. These bounds are natural from the viewpoint of the large deviation theory but the derivation requires careful investigation of the properties of Dirichlet distributions.

The difficulty of the proof for Multinomial TS relies on the following factors. Unlike the regret analysis for Binary TS in Kaufmann et al. (2012) and Agrawal and Goyal (2012), the optimal asymptotic regret bound, which corresponds to the post-convergence term, requires the computation of the infimum of the Kullback-Leibler divergence. On the other hand in the case of Binary TS, where all arms follow Bernoulli distribution, this infimum can be trivially computed explicitly as $\mathcal{K}_{\inf}^{(1)}(F_i, \mu^*) = \mathrm{KL}(\mathrm{Ber}(\mathrm{E}[F_i]), \mathrm{Ber}(\mu^*))$. This difference adds a specific technicality in the proof of Proposition 7 for the post-convergence term of Multinomial TS. In addition to it, the techniques used for the binary case cannot be easily generalized, because it heavily relies on the beta-binomial transform. Instead of this transformation, our analysis uses the explicit form of the density function of the Dirichlet distribution in some places, and also uses the property that Dirichlet random variables can be generated by normalization of Gamma random variables in other places.

Regarding Non-parametric TS, the simple decomposition between a post-convergence and a pre-convergence phase is not trivial. Indeed, in the case of Bayesian algorithms like Binary TS or Multinomial TS, the pre-convergence phase corresponds to the phase of convergence of the parameters of the conjugate distribution. However, in the case of Non-parametric TS, the pre-convergence phase corresponds to the convergence in the algorithm of the empirical distribution of the reward, in the sense of the Lévy distance. We evaluate the convergence of the Lévy distance by reducing it to the evaluation of the L^∞ distance between cumulative distributions over the space of nondecreasing functions.

5. Simulation Results

In this section, we give results of two experiments to show the performance of the proposed two algorithms. Both experiments have been performed over a hundred trials each, that is, we have run each experiment a hundred times, and the curve is the average of these results.

We perform the first experiment on a horizon $T = 10^5$ with two multinomial arms of identical distribution support $\{0, \frac{1}{3}, \frac{2}{3}, 1\}$. The first arm has parameters $(0.1, 0.1, 0.4, 0.4)$ and $\mu_1 = 0.7$, and the second arm has parameters $(0.4, 0.4, 0.1, 0.1)$ and $\mu_2 = 0.3$. We will only compare Multinomial TS with $M = 3$ and Binary TS. Since Multinomial TS is designed to be optimal for multinomial distributions, comparing it to Binary TS will show how much it improves from Binary TS (which is optimal for Bernoulli arms), and how significant the difference in the regret is.

The results in Figure 1 show a clearcut difference between Multinomial TS and Binary TS. It appears quite clearly on the figure that the logarithmic coefficient is far better in the case of Multinomial TS. We notice, however, that for a small number of rounds, Binary TS seems to perform better than Multinomial TS. This is no surprise due to early-stage estimation. Indeed, since more parameters are estimated in Multinomial TS than in Binary TS, the pre-convergence phase is longer in Multinomial TS than in Binary TS, making it seemingly less performing on a short horizon.

The second experiment investigates the proposed algorithms in a more general setting where reward distributions are over $[0, 1]$ but not multinomial. In this experiment, we compare Non-parametric TS, Multinomial TS with parameter $M = 5$, the empirical KL-UCB from Cappé et al.

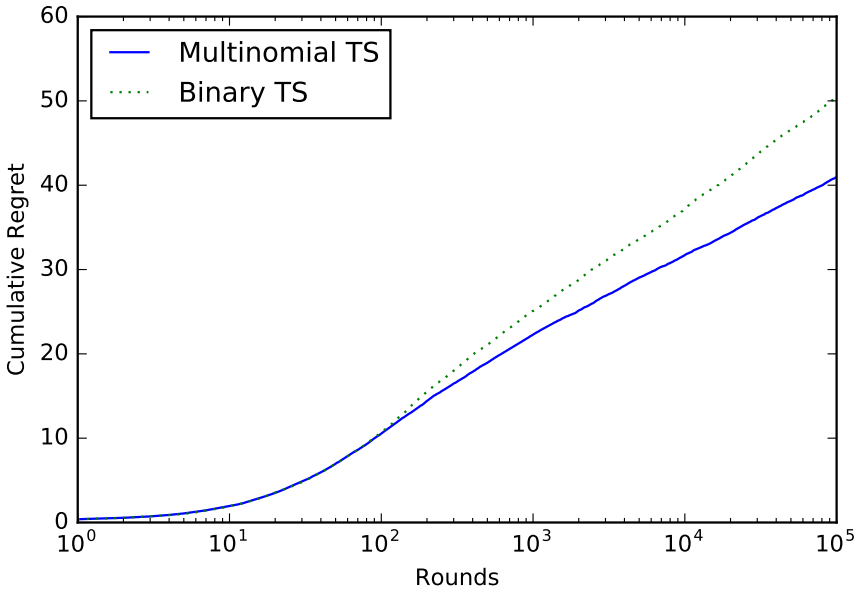

Figure 1: Comparison of Binary TS and Multinomial TS for multinomial rewards.

(2013), Binary TS and UCB1 from Auer et al. (2002). For this purpose, we conduct the experiment on the horizon $T = 10^4$, with two arms with exponential distributions truncated on $[0, 1]$. The first arm is an exponential distribution of rate parameter $l = 0.01$ which was then truncated on $[0, 1]$ with $\mu_1 \approx 0.499$ and the second arm is an exponential distribution of rate parameter $l = 10$ which was then truncated on $[0, 1]$ with $\mu_2 \approx 0.100$. The aim of this experiment is to compare our algorithms to the classic bandit algorithms UCB1 (see Auer et al., 2002) and Binary TS, and to the state-of-the-art algorithm called the empirical KL-UCB, which reaches the optimal regret lower bound.

From the result shown in Figure 2, we notice that Non-parametric TS outperforms or performs comparably with other algorithms, including the empirical KL-UCB. It is also interesting to see that it also performs very well for a small number of rounds. This is due to the fact that contrary to Binary TS and Multinomial TS, it does not estimate parameters but directly relies on the observed rewards. The algorithms Multinomial TS (for both values of M) and Binary TS seem to perform comparably. UCB1, however, with no surprise, performs not as well as all the other algorithms.

It should be noted that whereas the empirical KL-UCB performed comparably to Non-parametric TS, Non-parametric TS is still advantageous since, contrary to the empirical KL-UCB, it does not solve an optimization problem at each step, making it far quicker and easier to apply in online settings than the current state-of-the-art algorithm called the empirical KL-UCB.

6. Conclusion

In this paper, we proposed and analyzed two algorithms for the stochastic bandit with reward distributions over $[0, 1]$. The first one, Multinomial TS, is an adaptation of Thompson sampling for binary reward to the case of multinomial rewards of support included in $\{0, \frac{1}{M}, \ldots, 1\}$ and can also be used for general rewards bounded in $[0, 1]$ by the randomized rounding. The bound obtained in this case converges toward the optimal asymptotic regret bound for distributions bounded in $[0, 1]$

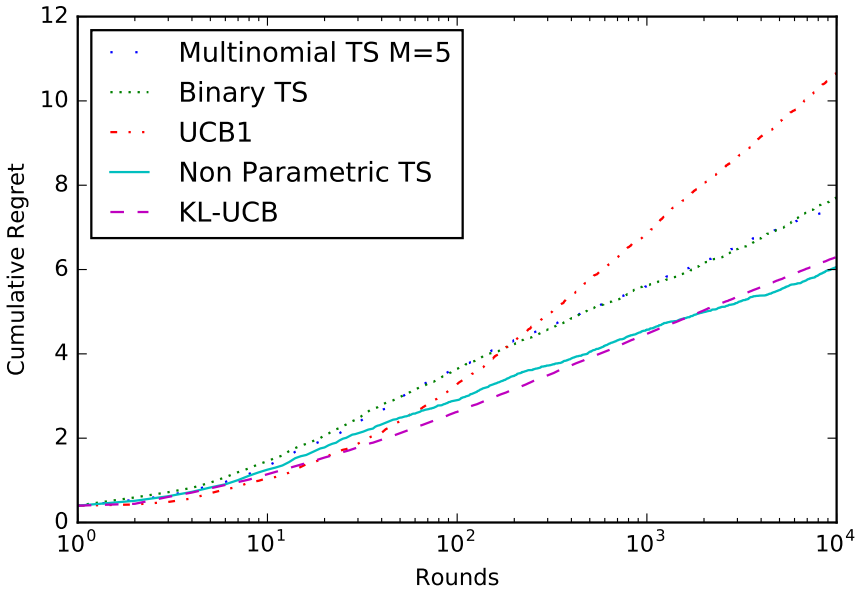

Figure 2: Comparison of UCB1, Binary TS, Multinomial TS and Non-parametric TS for truncated exponential rewards.

when M tends to infinity. The second one, Non-parametric TS, is a randomized algorithm in the more general case of reward distributions bounded in $[0, 1]$. It is not Bayesian in the strict sense, as it does not estimate parameters of a conjugate distribution before sampling. Thanks to this fact, it also performs well for a small number of rounds. For those reasons, it experimentally outperforms the classic bandit algorithms such as UCB1 and Binary TS, but also most state-of-the-art bandit algorithms, including some which require to solve an optimization problem at each step, like the empirical KL-UCB.

An important direction for future research is to give a finite-time regret bound to fully clarify the effect of M, which is currently hidden in the $O(1)$ term. A related direction is to clarify the effect of the prior for the Dirichlet distribution; although it is often reported that TS is not too sensitive to the choice of the prior, in our problem there are $M + 1$ or infinitely many parameters in the model and the choice may be more essential than models with few parameters.

Acknowledgments

We thank anonymous reviewers for helpful comments. JH was supported by KAKENHI 18K17998.

References

S. Agrawal and N. Goyal. Further optimal regret bounds for Thompson sampling. In *Artificial Intelligence and Statistics*, pages 99–107, 2012.

P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2):235–256, 2002.

A. N. Burnetas and M. N. Katehakis. Optimal adaptive policies for sequential allocation problems. *Advances in Applied Mathematics*, 17(2):122–142, 1996.

O. Cappé, A. Garivier, O. Maillard, R. Munos, and G. Stoltz. Kullback-Leibler upper confidence bounds for optimal sequential allocation. *The Annals of Statistics*, 41(3):1516–1541, 2013.

A. Garivier, H. Hadiji, P. Ménard, and G. Stoltz. Kl-ucb-switch: optimal regret bounds for stochastic bandits from both a distribution-dependent and a distribution-free viewpoints. *arXiv preprint*, 2018.

B. Hao, Y. Abbasi-Yadkori, Z. Wen, and G. Cheng. Bootstrapping upper confidence bound. *NeurIPS*, 2019.

J. Honda and A. Takemura. An asymptotically optimal bandit algorithm for bounded support models. In *COLT*, pages 67–79, 2010.

J. Honda and A. Takemura. Non-asymptotic analysis of a new bandit algorithm for semi-bounded rewards. *The Journal of Machine Learning Research*, 16(1):3721–3756, 2015.

E. Kaufmann, N. Korda, and R. Munos. Thompson sampling: an asymptotically optimal finite-time analysis. In *International Conference on Algorithmic Learning Theory*, pages 199–213, 2012.

N. Korda, E. Kaufmann, and R. Munos. Thompson sampling for 1-dimensional exponential family bandits. In *Advances in neural information processing systems*, pages 1448–1456, 2013.

B. Kveton, C. Szepesvari, S. Vaswani, Z. Wen, M. Ghazamvadeh, and T. Lattimore. Garbage in, reward out: Bootstrapping exploration in multi-armed bandits. *ICML*, 2019.

T. L. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6:4–22, 1985.

Appendix A. Proof of Proposition 7

In this section, we will look at the post-convergence term

$$\begin{aligned}
&\mathrm{E}\left[\sum_{t=1}^{T}\mathbb{1}\left(I(t)=i, u^\top L_{I(t)}(t)\geq \mu^*-\epsilon_1, \left\|\frac{\alpha^{I(t)}(t)}{N_{I(t)}(t)+M+1}-p^{I(t)}\right\|_\infty\leq\epsilon_2\right)\right]\\
&=\sum_{t=1}^{T}\sum_{n=1}^{T}\mathrm{E}\Bigg[\mathbb{1}\Bigg(I(t)=i, u^\top L_i(t)\geq\mu^*-\epsilon_1,\\
&\qquad\qquad\left\|\frac{\alpha^{i}(t)}{N_i(t)+M+1}-p^{i}\right\|_\infty\leq\epsilon_2, N_i(t)=n\Bigg)\Bigg].
\end{aligned}$$

Here note that if the event $\left\{I(t)=i, u^\top L_i(t)\geq\mu^*-\epsilon_1, \left\|\frac{\alpha^i(t)}{N_i(t)+M+1}-p^i\right\|_\infty\leq\epsilon_2, N_i(t)=n\right\}$ occurs at step t for a certain $n\in[T]$, then $N_i(t')>N_i(t)=n$ for any $t'>t$. Therefore, we deduce that, for any $n\in[T]$,

$$\sum_{t=1}^{T}\mathbb{1}\left(I(t)=i, u^\top L_i(t)\geq\mu^*-\epsilon_1, \left\|\frac{\alpha^{i}(t)}{N_i(t)+M+1}-p^{i}\right\|_\infty\leq\epsilon_2, N_i(t)=n\right)\leq 1.$$

We can then bound, for any $n_0\in[T]$:

$$\begin{aligned}
&\mathrm{E}\left[\sum_{t=1}^{T}\mathbb{1}\left(I(t)=i, u^\top L_{I(t)}(t)\geq \mu^*-\epsilon_1, \left\|\frac{\alpha^{I(t)}(t)}{N_{I(t)}(t)+M+1}-p^{I(t)}\right\|_\infty\leq\epsilon_2\right)\right]\\
&\leq n_0+\sum_{t=1}^{T}\sum_{n=n_0}^{T}\mathrm{E}\Bigg[\mathbb{1}\Bigg(I(t)=i, u^\top L_i(t)\geq\mu^*-\epsilon_1,\\
&\qquad\qquad\left\|\frac{\alpha^{i}(t)}{N_i(t)+M+1}-p^{i}\right\|_\infty\leq\epsilon_2, N_i(t)=n\Bigg)\Bigg]\\
&\leq n_0+\sum_{t=1}^{T}\sum_{n=n_0}^{T}P\left(u^\top L_i(t)\geq\mu^*-\epsilon_1, \left\|\frac{\alpha^{i}(t)}{N_i(t)+M+1}-p^{i}\right\|_\infty\leq\epsilon_2, N_i(t)=n\right)\\
&=n_0+\sum_{t=1}^{T}\sum_{n=n_0}^{T}P\left(u^\top L_i(t)\geq\mu^*-\epsilon_1\;\middle|\;\left\|\frac{\alpha^{i}(t)}{N_i(t)+M+1}-p^{i}\right\|_\infty\leq\epsilon_2, N_i(t)=n\right)\\
&\qquad\qquad\times P\left(\left\|\frac{\alpha^{i}(t)}{N_i(t)+M+1}-p^{i}\right\|_\infty\leq\epsilon_2, N_i(t)=n\right). \qquad (4)
\end{aligned}$$

Here note that by Lemma 13 in Appendix F.1 we have

$$\begin{aligned}
&P\left(u^\top L_i(t)\geq\mu^*-\epsilon_1\;\middle|\;\alpha^i(t), N_i(t)=n\right)\\
&\qquad\leq C_1(n+M+1)^{M/2}\exp(-(n+M+1)\mathrm{KL}(P_{\alpha^i(t)}\|P^*_{\mu^*-\epsilon_1})),
\end{aligned}$$

where $P^*_{\mu^*-\epsilon_1} = \arg\min_{x:\, u^T x \geq \mu^*-\epsilon_1} \mathrm{KL}(P_{\alpha^i(t)} \| x)$. But by definition, $\mathrm{KL}(P_{\alpha^i(t)} \| P^*_{\mu^*-\epsilon_1}) = \mathcal{K}_{\inf}(P_{\alpha^i(t)}, \mu^* - \epsilon_1)$, then we have

$$P\left(u^\top L_i(t) \geq \mu^* - \epsilon_1 \,\middle|\, \alpha^i(t), N_i(t) = n\right) \leq C_1 (n+M+1)^{M/2} \exp(-(n+M+1)\mathcal{K}_{\inf}(P_{\alpha^i(t)}, \mu^* - \epsilon_1)),$$

where $C_1 = \frac{e^{1/12}}{\Gamma(M+1)} \left(\frac{1}{\sqrt{2\pi}}\right)^M$. On the other hand, $\mathcal{K}_{\inf}(x, \mu^* - \epsilon_1)$ is continuous in $x \in [0,1]^{M+1}$ on the probability simplex with respect to the L^∞ distance from Honda and Takemura (2010, Theorem 7) and Lemma 18 in Appendix H. Therefore, for any $\epsilon_3 > 0$, there exist $\epsilon_2 > 0$ and constant $C_1' > 0$ such that

$$P\left(u^\top L_i(t) \geq \mu^* - \epsilon_1 \,\middle|\, \left\|\frac{\alpha^i(t)}{N_i(t)+M+1} - p^i\right\|_\infty \leq \epsilon_2, N_i(t) = n\right) \leq C_1' \exp(-(n+M+1)(\mathcal{K}_{\inf}(p^i, \mu^* - \epsilon_1) - \epsilon_3)).$$

Combining this with (4), we can bound

$$\begin{aligned}
&\mathrm{E}\left[\sum_{t=1}^T \mathbb{1}\left(I(t) = i, u^\top L_{I(t)}(t) \geq \mu^* - \epsilon_1, \left\|\frac{\alpha^{I(t)}(t)}{N_{I(t)}(t)+M+1} - p^{I(t)}\right\|_\infty \leq \epsilon_2\right)\right] \\
&\leq n_0 + C_1' \sum_{t=1}^T \exp(-(n_0+M+1)(\mathcal{K}_{\inf}(p^i, \mu^* - \epsilon_1) - \epsilon_3)) \\
&\qquad \times \sum_{n=n_0}^T P\left(\left\|\frac{\alpha^i(t)}{N_i(t)+M+1} - p^i\right\|_\infty \leq \epsilon_2, N_i(t) = n\right) \\
&\leq n_0 + C_1' \sum_{t=1}^T \exp(-(n_0+M+1)(\mathcal{K}_{\inf}(p^i, \mu^* - \epsilon_1) - \epsilon_3)) \\
&= n_0 + C_1' T \exp(-(n_0+M+1)(\mathcal{K}_{\inf}(p^i, \mu^* - \epsilon_1) - \epsilon_3)).
\end{aligned}$$

Choosing $n_0 = \frac{\log T}{\mathcal{K}_{\inf}(p^i, \mu^* - \epsilon_1) - \epsilon_3} - (M+1)$ provides the upper bound

$$\mathrm{E}\left[\sum_{t=1}^T \mathbb{1}\left(I(t) = i, u^\top L_{I(t)}(t) \geq \mu^* - \epsilon_1, \left\|\frac{\alpha^{I(t)}(t)}{N_{I(t)}(t)+M+1} - p^{I(t)}\right\|_\infty \leq \epsilon_2\right)\right] \leq \frac{\log T}{\mathcal{K}_{\inf}(p^i, \mu^* - \epsilon_1) - \epsilon_3} - M - 1 + C_1'.$$

In Honda and Takemura (2010, Theorem 7), it is proven that $\mu \longmapsto \mathcal{K}_{\inf}(F, \mu)$ is continuous for $\mu < 1$, and thus, we can deduce that for any $\epsilon_4 > 0$, there exists $\epsilon_1 > 0$ such that

$$\left|\mathcal{K}_{\inf}(p^i, \mu^* - \epsilon_1) - \mathcal{K}_{\inf}(p^i, \mu^*)\right| \leq \epsilon_4.$$

This implies that, for any $\epsilon_0 > 0$, there exist $\epsilon_1 > 0$ and $\epsilon_2 > 0$ such that

$$\mathrm{E}\left[\sum_{t=1}^{T} \mathbb{1}\left(I(t)=i, u^\top L_{I(t)}(t) \geq \mu^* - \epsilon_1, \left\|\frac{\alpha^{I(t)}(t)}{N_{I(t)}(t)+M+1} - p^{I(t)}\right\|_\infty \leq \epsilon_2\right)\right] \leq \frac{(1+\epsilon_0)\log T}{\mathcal{K}_{\inf}(p^i, \mu^*)} - M - 1 + C_1'.$$

■

Appendix B. Proof of Proposition 8

In this section, we evaluate the pre-convergence term, which is decomposed as

$$\mathrm{E}\left[\sum_{t=1}^{T} \mathbb{1}\left(I(t)=i, \left\{u^\top L_{I(t)}(t) < \mu^* - \epsilon_1 \cup \left\|\frac{\alpha^{I(t)}(t)}{N_{I(t)}(t)+M+1} - p^{I(t)}\right\|_\infty > \epsilon_2\right\}\right)\right]$$
$$\leq \underbrace{\mathrm{E}\left[\sum_{t=1}^{T} \mathbb{1}\left(\left\|\frac{\alpha^{I(t)}(t)}{N_{I(t)}(t)+M+1} - p^{I(t)}\right\|_\infty > \epsilon_2\right)\right]}_{\text{(A)}} + \underbrace{\mathrm{E}\left[\sum_{t=1}^{T} \mathbb{1}\left(u^\top L_{I(t)}(t) < \mu^* - \epsilon_1\right)\right]}_{\text{(B)}}.$$

We are going to bound each of the two pre-convergence terms, (A) and (B).

B.1. Bounding (A)

Bounding this term is quite easy actually and similar to the one in Agrawal and Goyal (2012). We bound the gap between the true parameters and the estimated parameters of the arm pulled at each step, so the gap will necessary vanish, and this is independent from the choice of the algorithm.

Letting $\tau^k(n)$ be the round of the n-th pull of arm $k \in [K]$, we can write

$$\begin{aligned}
&\mathrm{E}\left[\sum_{t=1}^{T} \mathbb{1}\left(\left\|\frac{\alpha^{I(t)}(t)}{N_{I(t)}(t)+M+1} - p^{I(t)}\right\|_\infty > \epsilon_2\right)\right] \\
&= \mathrm{E}\left[\sum_{k=1}^{K}\sum_{t=1}^{T} \mathbb{1}\left(\left\|\frac{\alpha^{k}(t)}{N_{k}(t)+M+1} - p^{k}\right\|_\infty > \epsilon_2, I(t)=k\right)\right] \\
&= \mathrm{E}\left[\sum_{k=1}^{K}\sum_{n=1}^{T}\sum_{t=\tau^k(n)}^{\tau^k(n+1)-1} \mathbb{1}\left(\left\|\frac{\alpha^{k}(t)}{n+M+1} - p^{k}\right\|_\infty > \epsilon_2, I(t)=k\right)\right] \\
&= \sum_{k=1}^{K}\sum_{n=1}^{T} \mathrm{E}\left[\mathbb{1}\left(\left\|\frac{\alpha^{k}(\tau^k(n))}{n+M+1} - p^{k}\right\|_\infty > \epsilon_2\right) \sum_{t=\tau^k(n)}^{\tau^k(n+1)-1} \mathbb{1}(I(t)=k)\right] \\
&= \sum_{k=1}^{K}\sum_{n=1}^{T} P\left(\left\|\frac{\alpha^{k}(\tau^k(n))}{n+M+1} - p^{k}\right\|_\infty > \epsilon_2\right).
\end{aligned}$$

Since $\alpha_i^k(\tau^k(n)) - 1$ follows the binomial distribution with n trials and success probability p_i^k, this term can be bounded using Hoeffding's inequality as

$$
\begin{aligned}
&P\left(\left\|\frac{\alpha^k(\tau^k(n))}{n+M+1} - p^k\right\|_\infty > \epsilon_2\right) \\
&= P\left(\max_{i\in\{0,..,M\}}\left|\frac{\alpha_i^k(\tau^k(n))}{n+M+1} - p_i^k\right| > \epsilon_2\right) \\
&\leq \sum_{i=0}^{M} P\left(\left|\frac{\alpha_i^k(\tau^k(n))}{n+M+1} - p_i^k\right| > \epsilon_2\right) \\
&\leq \sum_{i=0}^{M} P\left(\left|\frac{\alpha_i^k(\tau^k(n))-1}{n} - p_i^k\right| > \epsilon_2 - \left|\frac{(n+M+1)(\alpha_i^k(\tau^k(n))-1) - n\alpha_i^k(\tau^k(n))}{n(n+M+1)}\right|\right) \\
&= \sum_{i=0}^{M} P\left(\left|\frac{\alpha_i^k(\tau^k(n))-1}{n} - p_i^k\right| > \epsilon_2 - \left|\frac{(M+1)(\alpha_i^k(\tau^k(n))-1) - n}{n(n+M+1)}\right|\right) \\
&\leq \sum_{i=0}^{M} P\left(\left|\frac{\alpha_i^k(\tau^k(n))-1}{n} - p_i^k\right| > \epsilon_2 - \frac{M}{n}\right) \qquad (5) \\
&\leq (M+1)\min\left\{1, \exp\left(-2n\left(\epsilon_2 - \frac{M}{n}\right)^2\right)\right\},
\end{aligned}
$$

where (5) follows from $1 \leq \alpha_i^k(\tau^k(n)) \leq n+1$. Therefore,

$$
\begin{aligned}
\sum_{n=1}^{T} P\left(\left\|\frac{\alpha^k(\tau^k(n))}{n+M+1} - p^k\right\|_\infty > \epsilon_2\right) &\leq (M+1)\left(\frac{2M}{\epsilon_2} + \sum_{n=\left\lceil\frac{2M}{\epsilon_2}\right\rceil}^{T} \exp\left(-\frac{n\epsilon_2^2}{2}\right)\right) \\
&\leq (M+1)\left(\frac{2M}{\epsilon_2} + \sum_{n=1}^{T} \exp\left(-\frac{n\epsilon_2^2}{2}\right)\right) \\
&\leq (M+1)\left(\frac{2M}{\epsilon_2} + \frac{2}{\epsilon_2^2}\right).
\end{aligned}
$$

We can therefore obtain the bound of term (A) by

$$
\text{(A)} \leq K(M+1)\left(\frac{2M}{\epsilon_2} + \frac{2}{\epsilon_2^2}\right).
$$

B.2. Bounding (B)

The main difficulty of the regret analysis lies in bounding this term. To do that, we are going to decompose this term even more. Recall that we have assumed that the optimal arm is arm 1. We denote $\mathrm{Mult}(n,p)$ the multinomial distribution of parameters (n,p) where $p = (p_0,\ldots,p_M)$ for some $M \geq 1$ satisfies $\sum_{i=0}^{M} p_i = 1$ and $p_i \geq 0$ for any $i \in \{0,\ldots,M\}$. Then term (B) is expressed

as

$$\begin{aligned}
&\mathrm{E}\left[\sum_{t=1}^{T} \mathbb{1}(u^\top L_{I(t)}(t) < \mu^* - \epsilon_1)\right] \\
&= \mathrm{E}\left[\sum_{t=1}^{T}\sum_{n=1}^{T} \mathbb{1}(u^\top L_{I(t)}(t) < \mu^* - \epsilon_1, N_1(t) = n)\right] \\
&= \mathrm{E}\left[\sum_{n=1}^{T}\sum_{m=1}^{T} \mathbb{1}\left(\sum_{t=1}^{T} \mathbb{1}(u^\top L_{I(t)}(t) < \mu^* - \epsilon_1, N_1(t) = n) \geq m\right)\right],
\end{aligned}$$

where we used the property that, for any series of events (A_t),

$$\sum_{t=1}^{T} \mathbb{1}(A_t) = \sum_{m=1}^{T} \mathbb{1}\left(\sum_{t=1}^{T} \mathbb{1}(A_t) \geq m\right).$$

Then, if the event $\left\{u^\top L_1(t) > \mu^* - \epsilon_1,\ \max_{j\neq 1} u^\top L_j(t) \leq \mu^* - \epsilon_1,\ N_1(t) = n\right\}$ occurs at time t_0, then $N_1(t) = n$ will not hold for any $t > t_0$. Thus, denoting $\tau_1, \ldots, \tau_m$ the first m rounds at which the event $\left\{\max_{j\neq 1} u^\top L_j(t) \leq \mu^* - \epsilon_1,\ N_1(t) = n\right\}$ holds, it is necessary to have $u^\top L_1(t) \leq \mu^* - \epsilon_1$ at all $\tau_1, \ldots, \tau_m$ in order to have the event $\{\sum_{t=1}^{T} \mathbb{1}(u^\top L_{I(t)}(t) < \mu^* - \epsilon_1, N_1(t) = n) \geq m\}$. Thus, we can evaluate

$$\begin{aligned}
&\mathrm{E}\left[\sum_{t=1}^{T} \mathbb{1}(u^\top L_{I(t)}(t) < \mu^* - \epsilon_1)\right] \\
&\leq \sum_{n=1}^{T}\sum_{m=1}^{T} \mathrm{E}\left[\prod_{k=1}^{m} \mathbb{1}(u^\top L_1(\tau_k) \leq \mu^* - \epsilon_1)\right] \\
&\leq \sum_{n=1}^{T}\sum_{m=1}^{T} \mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\prod_{k=1}^{m} P\left(u^\top L(\tau_k) \leq \mu^* - \epsilon_1 \,\middle|\, \alpha(\tau_k)\right)\right] \\
&= \sum_{n=1}^{T} \mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\sum_{m=1}^{T} \left(P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L \leq \mu^* - \epsilon_1\right)\right)^m\right] \\
&\leq \sum_{n=1}^{T} \mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\frac{P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L \leq \mu^* - \epsilon_1\right)}{1 - P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L \leq \mu^* - \epsilon_1\right)}\right].
\end{aligned}$$

Then, we decompose the term within the sum as follows:

$$\begin{aligned}
&\mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\frac{P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L \leq \mu^* - \epsilon_1\right)}{1 - P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L \leq \mu^* - \epsilon_1\right)}\right] \\
&= \underbrace{\mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\mathbb{1}\left(\frac{u^\top \alpha}{n+M+1} > \mu^* - \frac{\epsilon_1}{2}\right)\frac{P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L \leq \mu^* - \epsilon_1\right)}{1 - P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L \leq \mu^* - \epsilon_1\right)}\right]}_{\text{(B1)}}
\end{aligned}$$

$$+\underbrace{\mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\mathbb{1}\left(\mu^*-\epsilon_1<\frac{u^\top\alpha}{n+M+1}\leq\mu^*-\frac{\epsilon_1}{2}\right)\frac{P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L\leq\mu^*-\epsilon_1\right)}{1-P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L\leq\mu^*-\epsilon_1\right)}\right]}_{\text{(B2)}}$$

$$+\underbrace{\mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\mathbb{1}\left(\frac{u^\top\alpha}{n+M+1}\leq\mu^*-\epsilon_1\right)\frac{P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L\leq\mu^*-\epsilon_1\right)}{1-P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L\leq\mu^*-\epsilon_1\right)}\right]}_{\text{(B3)}}.$$

Now, we are going to provide an upper bound for each of these terms.

B.2.1. BOUNDING (B1)

In the case where $\frac{u^\top\alpha}{n+M+1}>\mu^*-\frac{\epsilon_1}{2}$, we can use Lemma 13 in Appendix F.1 on tails of Dirichlet distributions to provide an upper bound to

$$P_{L\sim\mathrm{Dir}(\alpha)}(L\in S)\leq C_1(n+M+1)^{M/2}\exp(-(n+M+1)\mathrm{KL}(P_\alpha\|P^*)),$$

where we denoted $P^*:=\arg\min_{x\in S}\mathrm{KL}(P_\alpha\|x)$ and $S=\{x\in[0,1]^{M+1}:\ 1^\top x=1,\ u^\top x\leq\mu^*-\epsilon_1\}$. However, denoting $\delta:=\inf_{(a,b)\,:\,u^\top a\geq\mu^*-\epsilon_1/2,\ \mu^*-\epsilon_1\geq u^\top b}\mathrm{KL}(a\|b)>0$, we can also bound this term.

$$C_1(n+M+1)^{M/2}\exp(-(n+M+1)\mathrm{KL}(P_\alpha\|P^*))\leq C_1(n+M+1)^{M/2}\exp(-(n+M+1)\delta).$$

Then, there exists $n_1>0$ such that for any $n\geq n_1$, $C_1(n+M+1)^{M/2}\exp(-(n+M+1)\delta)<1$. Using this upper bound, we can then provide an upper bound to the term (B1) for any $n\geq n_1$,

$$\mathrm{E}\left[\mathbb{1}\left(\frac{u^\top\alpha}{n+M+1}>\mu^*-\frac{\epsilon_1}{2}\right)\frac{P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L\leq\mu^*-\epsilon_1\right)}{1-P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L\leq\mu^*-\epsilon_1\right)}\right]$$
$$\leq\frac{C_1(n+M+1)^{M/2}\exp(-(n+M+1)\delta)}{1-C_1(n+M+1)^{M/2}\exp(-(n+M+1)\delta)}.$$

Finally we have

$$\underbrace{\sum_{n=1}^{T}\mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\mathbb{1}\left(u^\top\alpha>(n+M+1)\left(\mu^*-\frac{\epsilon_1}{2}\right)\right)\frac{P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L\leq\mu^*-\epsilon_1\right)}{1-P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L\leq\mu^*-\epsilon_1\right)}\right]}_{(*)}$$
$$\leq\sum_{n=1}^{n_1-1}(*)+\sum_{n=n_1}^{T}\frac{C_1(n+M+1)^{M/2}}{\exp((n+M+1)\delta)-C_1(n+M+1)^{M/2}}$$
$$=O(1).$$

B.2.2. BOUNDING (B2)

We can then provide an upper bound to the expectation

$$\begin{aligned}
&\mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\mathbb{1}\left(\mu^*-\epsilon_1<\frac{u^\top\alpha}{n+M+1}\leq\mu^*-\frac{\epsilon_1}{2}\right)\frac{P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L\leq\mu^*-\epsilon_1\right)}{1-P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L\leq\mu^*-\epsilon_1\right)}\right]\\
&\leq\mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\mathbb{1}\left(\mu^*-\epsilon_1<\frac{u^\top\alpha}{n+M+1}\leq\mu^*-\frac{\epsilon_1}{2}\right)\frac{1}{1-P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L\leq\mu^*-\epsilon_1\right)}\right]\\
&=\mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\mathbb{1}\left(\mu^*-\epsilon_1<\frac{u^\top\alpha}{n+M+1}\leq\mu^*-\frac{\epsilon_1}{2}\right)\frac{1}{P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L\geq\mu^*-\epsilon_1\right)}\right].
\end{aligned}$$

Since $\alpha_i\neq 0$ for any $i\in\{0,\dots,M\}$, we can use Lemma 14 in Appendix F.2 on tails of Dirichlet distributions and we obtain

$$\begin{aligned}
P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L\geq\mu^*-\epsilon_1\right)&\geq P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L\geq\frac{1}{n+M+1}u^\top\alpha\right)\\
&\geq C_2(n+M+1)^{-\frac{M}{2}}\exp(-(n+M+1)\mathrm{KL}(P_\alpha\|P^*))\frac{P_{\alpha_M}}{P^*_M},
\end{aligned}$$

where we denoted $P^*:=\arg\min_{x:\ u^\top x\geq\mu^*-\epsilon_1}\mathrm{KL}(P_\alpha\|x)$ and $C_2:=\left(\frac{1}{\sqrt{2\pi}}\right)^M e^{-(M+1)/12}$. By definition of P^*, we deduce that $P^*=P_\alpha$, thus

$$P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L\geq\mu^*-\epsilon_1\right)\geq C_2(n+M+1)^{-\frac{M}{2}}.$$

Therefore, we deduce that

$$\begin{aligned}
&\mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\mathbb{1}\left(\mu^*-\epsilon_1<\frac{u^\top\alpha}{n+M+1}\leq\mu^*-\frac{\epsilon_1}{2}\right)\frac{P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L\leq\mu^*-\epsilon_1\right)}{1-P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L\leq\mu^*-\epsilon_1\right)}\right]\\
&\leq C_2^{-1}(n+M+1)^{\frac{M}{2}}\mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\mathbb{1}\left(\mu^*-\epsilon_1<\frac{u^\top\alpha}{n+M+1}\leq\mu^*-\frac{\epsilon_1}{2}\right)\right].
\end{aligned}$$

Then, using Hoeffding's inequality, we can provide the upper bound given by

$$\begin{aligned}
&\mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\mathbb{1}\left(\mu^*-\epsilon_1<\frac{u^\top\alpha}{n+M+1}\leq\mu^*-\frac{\epsilon_1}{2}\right)\right]\\
&\leq\mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\mathbb{1}\left(\frac{u^\top\alpha}{n+M+1}\leq\mu^*-\frac{\epsilon_1}{2}\right)\right]\\
&\leq\exp\left(-\frac{(n+M+1)\epsilon_1^2}{2}\right).
\end{aligned}$$

Thus, we conclude that

$$\begin{aligned}
&\mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\mathbb{1}\left(\mu^*-\epsilon_1<\frac{u^\top\alpha}{n+M+1}\leq\mu^*-\frac{\epsilon_1}{2}\right)\frac{P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L\leq\mu^*-\epsilon_1\right)}{1-P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L\leq\mu^*-\epsilon_1\right)}\right]\\
&\leq C_2^{-1}(n+M+1)^{\frac{M}{2}}\exp\left(-\frac{(n+M+1)\epsilon_1^2}{2}\right),
\end{aligned}$$

which proves that

$$\sum_{n=1}^{T} \mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\mathbb{1}\left(\mu^* - \epsilon_1 < \frac{u^\top \alpha}{n+M+1} \le \mu^* - \frac{\epsilon_1}{2}\right) \times \frac{P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L \le \mu^* - \epsilon_1\right)}{1 - P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L \le \mu^* - \epsilon_1\right)}\right] = O(1).$$

B.2.3. BOUNDING (B3)

We can eventually provide an upper bound to the expectation as follows:

$$\begin{aligned}&\mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\mathbb{1}\left(\frac{u^\top \alpha}{n+M+1} \le \mu^* - \epsilon_1\right) \frac{P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L \le \mu^* - \epsilon_1\right)}{1 - P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L \le \mu^* - \epsilon_1\right)}\right] \\ &\qquad \le \mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\mathbb{1}\left(\frac{u^\top \alpha}{n+M+1} \le \mu^* - \epsilon_1\right) \frac{1}{P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L \ge \mu^* - \epsilon_1\right)}\right].\end{aligned}$$

Let $S := \{x \in [0,1]^{M+1} : 1^\top x = 1,\ u^\top x \ge \mu^* - \epsilon_1\}$ and $P^* := \arg\min_{x\in S} \mathrm{KL}(P_\alpha \| x)$. Denoting $C_2 := \left(\frac{1}{\sqrt{2\pi}}\right)^M e^{-(M+1)/12}$, Lemma 14 from Appendix F.2 on tails of Dirichlet distributions provides the lower bound as follows

$$\begin{aligned}&P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L \ge \frac{u^\top \alpha}{n+M+1} + \Delta\right) \\ &\ge C_2 (n+M+1)^{-\frac{M}{2}} \exp(-(n+M+1)\mathrm{KL}(P_\alpha \| P^*)) \frac{P_{\alpha_M}}{P^*_M} \\ &\ge C_2 (n+M+1)^{-\frac{M}{2}} \exp(-(n+M+1)\mathrm{KL}(P_\alpha \| P^*)) P_{\alpha_M}.\end{aligned}$$

Therefore we have

$$\begin{aligned}&\mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\mathbb{1}\left(\frac{u^\top \alpha}{n+M+1} \le \mu^* - \epsilon_1\right) \frac{1}{P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L \ge \mu^* - \epsilon_1\right)}\right] \\ &\le C_2^{-1} (n+M+1)^{\frac{M}{2}} \\ &\quad \times \mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\mathbb{1}\left(\frac{u^\top \alpha}{n+M+1} \le \mu^* - \epsilon_1\right) \exp((n+M+1)\mathrm{KL}(P_\alpha \| P^*)) \frac{1}{P_{\alpha_M}}\right].\end{aligned}$$

Then, using the bound $P_{\alpha_M} \ge \frac{1}{n+M+1}$, we have

$$\begin{aligned}&\mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\mathbb{1}\left(\frac{u^\top \alpha}{n+M+1} \le \mu^* - \epsilon_1\right) \frac{1}{P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L \ge \mu^* - \epsilon_1\right)}\right] \\ &\le C_2^{-1} (n+M+1)^{\frac{M}{2}+1} \\ &\quad \times \mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\mathbb{1}\left(\frac{u^\top \alpha}{n+M+1} \le \mu^* - \epsilon_1\right) \exp((n+M+1)\mathrm{KL}(P_\alpha \| P^*))\right],\end{aligned}$$

where recall that $S = \{x \in [0,1]^{M+1} : 1^\top x = 1,\ u^\top x \ge \mu^* - \epsilon_1\}$ and we denoted $P^* := \arg\min_{x\in S} \mathrm{KL}(P_\alpha \| x)$.

Denoting by $\mathrm{H}(P)$ the entropy of the multinomial distribution of parameter P and $A := \left\{\alpha \in \{1,\dots,n+1\}^{M+1} : 1\right.$ we can directly bound the expectation by

$$\begin{aligned}
&\mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\mathbb{1}\left(\frac{u^\top\alpha}{n+M+1} \le \mu^* - \epsilon_1\right)\exp((n+M+1)\mathrm{KL}(P_\alpha\|P^*))\right] \\
&= \sum_{\alpha\in A} P_{X-1\sim\mathrm{Mult}(n,p)}(X=\alpha)\exp((n+M+1)\mathrm{KL}(P_\alpha\|P^*)) \\
&= \sum_{\alpha\in A} \exp(-(n+M+1)\mathrm{H}(P_\alpha))\exp((n+M+1)(\mathrm{KL}(P_\alpha\|p) - \mathrm{KL}(P_\alpha\|P^*))) \\
&\le \sum_{\alpha\in A} \exp(-(n+M+1)\mathrm{H}(P_\alpha))\exp(-(n+M+1)(\mathrm{KL}(P_\alpha\|p^*) - \mathrm{KL}(P_\alpha\|P^*))),
\end{aligned}$$

where we denoted $p^* := \arg\min_{x:\ u^\top x \ge u^\top p} \mathrm{KL}(P_\alpha\|x)$. Here, we can use a result from Honda and Takemura (2010, Lemma 13) which states that

$$\begin{aligned}
\mathrm{KL}(P_\alpha\|p^*) - \mathrm{KL}(P_\alpha\|P^*) &= \mathcal{K}_{\inf}(P_\alpha,\mu^*) - \mathcal{K}_{\inf}(P_\alpha,\mu^*-\epsilon_1) \\
&\ge \frac{(\mu^* - (\mu^*-\epsilon_1))^2}{2\mu^*(1-\mu^*+\epsilon_1)} \\
&\ge \frac{\epsilon_1^2}{2\mu^*(1-\mu^*+\epsilon_1)} \\
&> 0.
\end{aligned}$$

Denoting $C := \frac{\epsilon_1^2}{2\mu^*(1-\mu^*+\epsilon_1)} > 0$, we then have

$$\begin{aligned}
&\mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\mathbb{1}\left(\frac{u^\top\alpha}{n+M+1} \le \mu^* - \epsilon_1\right)\frac{1}{P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L \ge \mu^* - \epsilon_1\right)}\right] \\
&\qquad \le C_2^{-1}(n+M+1)^{\frac{M}{2}+1}\exp(-C(n+M+1))\sum_{\alpha\in A}\exp(-(n+M+1)\mathrm{H}(P_\alpha)).
\end{aligned}$$

Here, it is easy to bound the cardinal of A by a polynomial in n, considering $A \subset \{1,..,n+1\}^{M+1}$: $|A| \le (n+1)^{M+1}$, and to bound $\exp(-(n+M+1)\mathrm{H}(P_\alpha)) \le 1$. Therefore,

$$\begin{aligned}
&\mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\mathbb{1}\left(\frac{u^\top\alpha}{n+M+1} \le \mu^* - \epsilon_1\right)\frac{1}{P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L \ge \mu^* - \epsilon_1\right)}\right] \\
&\qquad\qquad \le C_2^{-1}(n+M+1)^{\frac{M}{2}+1}(n+1)^{M+1}\exp(-C(n+M+1)),
\end{aligned}$$

which implies that

$$\sum_{n=1}^{T}\mathrm{E}_{\alpha-1\sim\mathrm{Mult}(n,p)}\left[\mathbb{1}\left(\frac{u^\top\alpha}{n+M+1} \le \mu^* - \epsilon_1\right)\frac{1}{P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L \ge \mu^* - \epsilon_1\right)}\right] = O(1).$$

Appendix C. Proof of Proposition 10

In this section, we evaluate the post-convergence term given by

$$\mathrm{E}\left[\sum_{t=1}^{T} \mathbb{1}\left(I(t)=i, V_{I(t)}(t) \geq \mu^* - \epsilon_1, D_L(\hat{F}_{I(t)}(t), F_{I(t)}) \leq \epsilon_2\right)\right].$$

We can use the same discussion as the derivation of (4) in the proof of Proposition 7 for the Multinomial TS, and we can bound the above expectation by

$$\begin{aligned}
&\mathrm{E}\left[\sum_{t=1}^{T} \mathbb{1}\left(I(t)=i, V_{I(t)}(t) \geq \mu^* - \epsilon_1, D_L(\hat{F}_{I(t)}(t), F_{I(t)}) \leq \epsilon_2\right)\right] \\
&= n_0 + \sum_{t=1}^{T}\sum_{n=n_0}^{T} P\left(V_i(t) \geq \mu^* - \epsilon_1 \,\middle|\, D_L(\hat{F}_i(t), F_i) \leq \epsilon_2, N_i(t) = n\right) \\
&\qquad\qquad \times P\left(D_L(\hat{F}_i(t), F_i) \leq \epsilon_2, N_i(t) = n\right). \qquad (6)
\end{aligned}$$

By Lemma 15 in Appendix G.1 on conditional probabilities, for any $\eta \in (0,1)$ we have

$$\begin{aligned}
&P\left(V_i(t) \geq \mu^* - \epsilon_1 \,\middle|\, N_i(t) = n,\ D_L(\hat{F}_i(t), F_i) \leq \epsilon_2\right) \\
&\qquad\qquad \leq \frac{1}{\eta}\exp\left(-n\left(\mathcal{K}_{\inf}(\hat{F}_i(t), \mu^* - \epsilon_1) - \eta\frac{\mu^* - \epsilon_1}{1-(\mu^* - \epsilon_1)}\right)\right).
\end{aligned}$$

Since $\mathcal{K}_{\inf}(F, \mu)$ is continuous in F with respect to the Lévy distance for $\mu < 1$ from Honda and Takemura (2010, Theorem 7), for any $\epsilon_3 > 0$ there exists $\epsilon_2 > 0$ such that

$$D_L(\hat{F}, F_i) \leq \epsilon_2 \implies \left|\mathcal{K}_{\inf}(\hat{F}, \mu^* - \epsilon_1) - \mathcal{K}_{\inf}(F_i, \mu^* - \epsilon_1)\right| \leq \epsilon_3.$$

Therefore, for any $\eta \in (0,1)$ and for any $\epsilon_5 > 0$, there exist $\epsilon_1 > 0$ and $\epsilon_2 > 0$ such that

$$\begin{aligned}
&P\left(V_i(t) \geq \mu^* - \epsilon_1 \,\middle|\, N_i(t) = n,\ D_L(\hat{F}_i(t), F_i) \leq \epsilon_2\right) \\
&\leq \frac{1}{\eta}\exp\left(-n\left(\mathcal{K}_{\inf}(F_i, \mu^* - \epsilon_1) - \epsilon_3 - \eta\frac{\mu^* - \epsilon_1}{1-(\mu^* - \epsilon_1)}\right)\right) \\
&\leq \frac{1}{\eta}\exp\left(-n\left(\mathcal{K}_{\inf}(F_i, \mu^*) - \frac{\epsilon_1}{1-\mu^*} - \epsilon_3 - \eta\frac{\mu^* - \epsilon_1}{1-(\mu^* - \epsilon_1)}\right)\right),
\end{aligned}$$

where the last inequality follows from Honda and Takemura (2010, Theorem 6). This implies that, for any $\epsilon_0 > 0$, there exists $\eta \in (0,1)$, $\epsilon_1 > 0$ and $\epsilon_2 > 0$ such that

$$P\left(V_i(t) \geq \mu^* - \epsilon_1 \,\middle|\, N_i(t) = n,\ D_L(\hat{F}_i(t), F_i) \leq \epsilon_2\right) \ \leq\ \frac{1}{\eta}\exp\left(-n\left(\mathcal{K}_{\inf}(F_i, \mu^*) - \epsilon_0\right)\right).$$

Combining this with (6) we have

$$\begin{aligned}
&\mathrm{E}\left[\sum_{t=1}^{T}\mathbb{1}\left(I(t)=i, V_{I(t)}(t)\geq\mu^*-\epsilon_1, D_L(\hat{F}_{I(t)}(t),F_{I(t)})\leq\epsilon_2\right)\right]\\
&\leq n_0+\frac{1}{\eta}\sum_{t=1}^{T}\exp(-n_0(\mathcal{K}_{\inf}(F_i,\mu^*)-\epsilon_0))\sum_{n=n_0}^{T}P\left(D_L(\hat{F}_i(t),F_i)\leq\epsilon_2, N_i(t)=n\right)\\
&\leq n_0+\frac{1}{\eta}\sum_{t=1}^{T}\exp(-n_0(\mathcal{K}_{\inf}(F_i,\mu^*)-\epsilon_0))\\
&=n_0+\frac{1}{\eta}T\exp(-n_0(\mathcal{K}_{\inf}(F_i,\mu^*)-\epsilon_0)).
\end{aligned}$$

Choosing $n_0=\frac{\log T}{\mathcal{K}_{\inf}(F_i,\mu^*)-\epsilon_0}$ provides the upper bound

$$\mathrm{E}\left[\sum_{t=1}^{T}\mathbb{1}\left(I(t)=i, V_{I(t)}(t)\geq\mu^*-\epsilon_1, D_L(\hat{F}_{I(t)}(t),F_{I(t)})\leq\epsilon_2\right)\right]\leq\frac{\log T}{\mathcal{K}_{\inf}(F_i,\mu^*)-\epsilon_0}+\frac{1}{\eta}.$$

■

Appendix D. Proof of Proposition 11

In this section, we consider the pre-convergence term.

$$\begin{aligned}
&\mathrm{E}\left[\sum_{t=1}^{T}\mathbb{1}\left(I(t)=i, \left\{V_{I(t)}(t)<\mu^*-\epsilon_1\cup D_L(\hat{F}_{I(t)}(t),F_{I(t)})>\epsilon_2\right\}\right)\right]\\
&\qquad\leq\underbrace{\mathrm{E}\left[\sum_{t=1}^{T}\mathbb{1}(D_L(\hat{F}_{I(t)}(t),F_{I(t)})>\epsilon_2)\right]}_{\text{(A)}}\\
&\qquad\quad+\underbrace{\mathrm{E}\left[\sum_{t=1}^{T}\mathbb{1}\left(V_{I(t)}(t)<\mu^*-\epsilon_1,\ D_L(\hat{F}_{I(t)}(t),F_{I(t)})\leq\epsilon_2\right)\right]}_{\text{(B)}}.
\end{aligned}$$

We are going to bound each of the two pre-convergence terms, (A) and (B).

D.1. Bounding (A)

Using Lemma 18 in Appendix H, we know that

$$D_L(\hat{F}_{I(t)}(t),F_{I(t)})\leq\left\|\hat{F}_{I(t)}(t)-F_{I(t)}\right\|_\infty.$$

Therefore, denoting $\tau_n^{(k)}$ the n-th time at which arm k is pulled, we have

$$\begin{aligned}
(\mathrm{A}) &\leq \mathrm{E}\left[\sum_{t=1}^{T} \mathbb{1}\left(\left\|\hat{F}_{I(t)}(t) - F_{I(t)}\right\|_\infty > \epsilon_2\right)\right] \\
&= \sum_{k=1}^{K} \mathrm{E}\left[\sum_{t=1}^{T} \mathbb{1}\left(I(t)=k, \left\|\hat{F}_k(t) - F_k\right\|_\infty > \epsilon_2\right)\right] \\
&= \sum_{k=1}^{K} \mathrm{E}\left[\sum_{n=1}^{T} \mathbb{1}\left(\left\|\hat{F}_k(\tau_n^{(k)}) - F_k\right\|_\infty > \epsilon_2\right)\right] \\
&\leq \sum_{k=1}^{K} \left(1 + \mathrm{E}\left[\sum_{n=2}^{T} \mathbb{1}\left(\left\|\hat{F}_k(\tau_n^{(k)}) - F_k\right\|_\infty > \epsilon_2\right)\right]\right).
\end{aligned}$$

In this subsection, we use the notations:

- $X_1, \ldots, X_n$ the rewards obtained by arm k,
- $\hat{F}_n^{(k)}(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_i \leq x)$,
- $\check{F}_n^{(k)}(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_i < x)$,
- $X_{(1)} \leq \cdots \leq X_{(n)}$ the ordered rewards obtained from arm k.

With this in mind, notice that, for any $i \in \{1, .., n-1\}$:

- $\hat{F}_n^{(k)}$ is constant on $[X_{(i)}, X_{(i+1)})$,
- $\check{F}_n^{(k)}$ is constant on $(X_{(i)}, X_{(i+1)}]$.

Then, the increase of F_k and $\hat{F}_n^{(k)}$ implies that

$$\left\|\hat{F}_k(\tau_n^{(k)}) - F_k\right\|_\infty > \epsilon_2 \iff \begin{cases} & \exists i \in \{1,\ldots,n\} \left|\hat{F}_n^{(k)}(X_i) - F_k(X_i)\right| > \epsilon_2 \\ \text{or} & \exists i \in \{1,\ldots,n\} \left|\check{F}_n^{(k)}(X_i) - F_k(X_i)\right| > \epsilon_2 \\ \text{or} & \left|\hat{F}_n^{(k)}(0) - F_k(0)\right| > \epsilon_2 \\ \text{or} & \left|\check{F}_n^{(k)}(1) - F_k(1)\right| > \epsilon_2, \end{cases}$$

as it is visualized in Figure 3. Indeed, let us look at the plot where the red line represents $\hat{F}_n^{(k)}$ and the black line represents F_k. The distance $\left\|\hat{F}_n^{(k)} - F_k\right\|_\infty$ is equal to the longest double arrow.

Therefore,

$$\begin{aligned}
&\mathrm{E}\left[\mathbb{1}\left(\left\|\hat{F}_k(\tau_n^{(k)}) - F_k\right\|_\infty > \epsilon_2\right)\right] \\
&\leq \mathrm{E}\left[\sum_{i=1}^{n} \mathbb{1}\left(\left|\hat{F}_n^{(k)}(X_i) - F_k(X_i)\right| > \epsilon_2\right) + \sum_{i=1}^{n} \mathbb{1}\left(\left|\check{F}_n^{(k)}(X_i) - F_k(X_i)\right| > \epsilon_2\right)\right. \\
&\qquad \left. + \mathbb{1}\left(\left|\hat{F}_n^{(k)}(0) - F_k(0)\right| > \epsilon_2\right) + \mathbb{1}\left(\left|\check{F}_n^{(k)}(1) - F_k(1)\right| > \epsilon_2\right)\right].
\end{aligned}$$

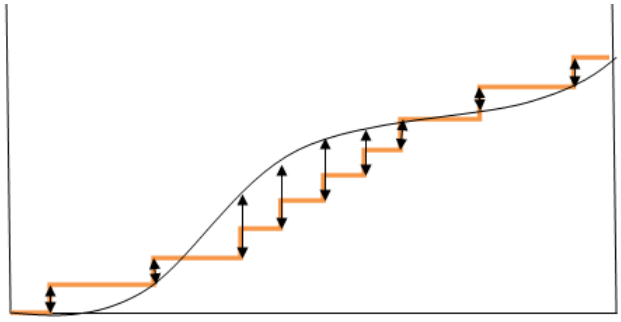

Figure 3: Visualizing the distance between $\hat{F}_n^{(k)}$ and F_k

But using Hoeffding's inequality, one can easily bound

$$P\left(\left|\hat{F}_n^{(k)}(0) - F_k(0)\right| > \epsilon_2\right) \leq \exp(-2n\epsilon_2^2),$$

and

$$P\left(\left|\check{F}_n^{(k)}(1) - F_k(1)\right| > \epsilon_2\right) \leq \exp(-2n\epsilon_2^2).$$

In addition to it, let us look carefully at $\left|\hat{F}_n^{(k)}(X_i) - F_k(X_i)\right|$, for $n \geq 2$. If we denote $U_i := F_k(X_i) \sim U([0,1])$, then for any $i \in \{1, \ldots, n\}$,

$$\begin{aligned}
\left|\hat{F}_n^{(k)}(X_i) - F_k(X_i)\right| &= \left|\frac{1}{n}\sum_{j=1}^{n} \mathbb{1}(X_j \leq X_i) - F_k(X_i)\right| \\
&= \left|\frac{1}{n}\sum_{j=1}^{n} \mathbb{1}\left(F_k(X_j) \leq F_k(X_i)\right) - F_k(X_i)\right| \\
&= \left|\frac{1}{n}\sum_{j=1}^{n} \mathbb{1}(U_j \leq U_i) - U_i\right| \\
&= \left|\frac{n-1}{n}\left(\frac{1}{n-1}\sum_{j\neq i} \mathbb{1}(U_j \leq U_i) - U_i\right) + \frac{1}{n} - \frac{1}{n}U_i\right| \\
&\leq \frac{n-1}{n}\left|\frac{1}{n-1}\sum_{j\neq i} \mathbb{1}(U_j \leq U_i) - U_i\right| + \frac{1}{n}(1 - U_i) \\
&\leq \frac{n-1}{n}\left|\frac{1}{n-1}\sum_{j\neq i} \mathbb{1}(U_j \leq U_i) - U_i\right| + \frac{1}{n}.
\end{aligned}$$

Therefore, using Hoeffding's inequality, for any $i \in \{1, \dots, n\}$,

$$\begin{aligned}
&\mathrm{E}\left[\mathbb{1}\left(\left|\hat{F}_n^{(k)}(X_i) - F_k(X_i)\right| > \epsilon_2\right)\right] \\
&\leq \mathrm{E}\left[\mathbb{1}\left(\frac{n-1}{n}\left|\frac{1}{n-1}\sum_{j\neq i}\mathbb{1}(U_j \leq U_i) - U_i\right| + \frac{1}{n} > \epsilon_2\right)\right] \\
&= \mathrm{E}\left[\mathrm{E}\left[\mathbb{1}\left(\left|\frac{1}{n-1}\sum_{j\neq i}\mathbb{1}(U_j \leq U_i) - U_i\right| > \epsilon_2 - \frac{1}{n-1}\right) \;\middle|\; U_i\right]\right] \\
&\leq \exp\left(-2(n-1)\left(\epsilon_2 - \frac{1}{n-1}\right)^2\right).
\end{aligned}$$

The same reasoning gives, for any $i \in \{1, \dots, n\}$,

$$\mathrm{E}\left[\mathbb{1}\left(\left|\check{F}_n^{(k)}(X_i) - F_k(X_i)\right| > \epsilon_2\right)\right] \leq \exp\left(-2(n-1)\left(\epsilon_2 - \frac{1}{n-1}\right)^2\right).$$

Therefore,

$$\begin{aligned}
\mathrm{E}\left[\mathbb{1}\left(\left\|\hat{F}_k(\tau_n^{(k)}) - F_k\right\|_\infty > \epsilon_2\right)\right] &\leq 2n\exp\left(-2(n-1)\left(\epsilon_2 - \frac{1}{n-1}\right)^2\right) + 2\exp(-2n\epsilon_2^2) \\
&\leq 2(n+1)\exp\left(-2(n-1)\left(\epsilon_2 - \frac{1}{n-1}\right)^2\right).
\end{aligned}$$

We can therefore bound term (A).

$$\text{(A)} \leq K\left(1 + \sum_{n=2}^{\infty} 2(n+1)\exp\left(-2(n-1)\left(\epsilon_2 - \frac{1}{n-1}\right)^2\right)\right).$$

D.2. Bounding (B)

In this subsection, we are going to bound the term (B) by decomposing the remaining term of the regret. Recall that we have assumed that the optimal arm is arm 1.

$$\begin{aligned}
&\mathrm{E}\left[\sum_{t=1}^{T}\mathbb{1}\left(V_{I(t)}(t) < \mu^* - \epsilon_1\right)\right] \\
&= \mathrm{E}\left[\sum_{t=1}^{T}\sum_{n=1}^{T}\mathbb{1}\left(V_{I(t)}(t) < \mu^* - \epsilon_1, N_1(t) = n\right)\right] \\
&= \mathrm{E}\left[\sum_{n=1}^{T}\sum_{m=1}^{T}\mathbb{1}\left(\sum_{t=1}^{T}\mathbb{1}\left(V_{I(t)}(t) < \mu^* - \epsilon_1, N_1(t) = n\right) \geq m\right)\right],
\end{aligned}$$

where we used the property that, for any series of events (A_t),

$$\sum_{t=1}^{T}\mathbb{1}(A_t) = \sum_{m=1}^{T}\mathbb{1}(\sum_{t=1}^{T}\mathbb{1}(A_t) \geq m).$$

Then, if the event $\{V_1(t) > \mu^* - \epsilon_1,\ \max_{j\neq 1} V_j(t) \leq \mu^* - \epsilon_1,\ N_1(t) = n\}$ occurs at time t_0, then $N_1(t) = n$ will not hold for any $t > t_0$. Thus, denoting $\tau_1, \ldots, \tau_m$ the first m rounds at which the event $\{\max_{j\neq 1} V_j(t) \leq \mu^* - \epsilon_1,\ N_1(t) = n\}$ holds, it is necessary to have $V_1(t) \leq \mu^* - \epsilon_1$ at all $\tau_1, \ldots, \tau_m$ so as to have the event $\left\{ \sum_{t=1}^{T} \mathbb{1}(V_{I(t)}(t) < \mu^* - \epsilon_1, N_1(t) = n) \geq m \right\}$. Thus, we can compute

$$
\begin{aligned}
&\mathrm{E}\left[\sum_{t=1}^{T} \mathbb{1}\left(V_{I(t)}(t) < \mu^* - \epsilon_1\right)\right] \\
&\leq \sum_{n=1}^{T}\sum_{m=1}^{T} \mathrm{E}\left[\prod_{k=1}^{m} \mathbb{1}\left(V_1(\tau_k) \leq \mu^* - \epsilon_1\right)\right] \\
&\leq \sum_{n=1}^{T}\sum_{m=1}^{T} \mathrm{E}_{X_1,\ldots,X_n\sim F_1}\left[\prod_{k=1}^{m} P\left(V(\tau_k) \leq \mu^* - \epsilon_1 \mid X_1, \ldots, X_n\right)\right] \\
&= \sum_{n=1}^{T} \mathrm{E}_{X_1,\ldots,X_n\sim F_1}\left[\sum_{m=1}^{T} \left(P_{L\sim \mathrm{Dir}(1)}\left(L^\top X \leq \mu^* - \epsilon_1 \,\middle|\, X_1, \ldots, X_n\right)\right)^m\right] \\
&\leq \sum_{n=1}^{T} \mathrm{E}_{X_1,\ldots,X_n\sim F_1}\left[\frac{P_{L\sim \mathrm{Dir}(1)}\left(L^\top X \leq \mu^* - \epsilon_1 \mid X_1, \ldots, X_n\right)}{1 - P_{L\sim \mathrm{Dir}(1)}\left(L^\top X \leq \mu^* - \epsilon_1 \mid X_1, \ldots, X_n\right)}\right].
\end{aligned}
$$

where in the last computation, all the variables refer to arm 1 (the optimal arm). But we dropped all 1 in the superscript for the sake of clarity (thus $X := X^{(1)}$ and $L := L_1$). We then perform the following decomposition of the events.

$$
\begin{aligned}
&\mathrm{E}_{X_1,\ldots,X_n\sim F_1}\left[\frac{P_{L\sim \mathrm{Dir}(1)}\left(X^\top L \leq \mu^* - \epsilon_1 \mid X_1, \ldots, X_n\right)}{1 - P_{L\sim \mathrm{Dir}(1)}\left(X^\top L \leq \mu^* - \epsilon_1 \mid X_1, \ldots, X_n\right)}\right] \\
&= \underbrace{\mathrm{E}_{X_1,\ldots,X_n\sim F_1}\left[\mathbb{1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \geq \mu^* - \frac{\epsilon_1}{2}\right)\frac{P_{L\sim \mathrm{Dir}(1)}\left(X^\top L \leq \mu^* - \epsilon_1 \mid X_1, \ldots, X_n\right)}{1 - P_{L\sim \mathrm{Dir}(1)}\left(X^\top L \leq \mu^* - \epsilon_1 \mid X_1, \ldots, X_n\right)}\right]}_{\text{(B1)}} \\
&\quad + \underbrace{\mathrm{E}_{X_1,\ldots,X_n\sim F_1}\left[\mathbb{1}\left(\mu^* - \frac{\epsilon_1}{2} > \frac{1}{n}\sum_{i=1}^{n} X_i \geq \mu^* - \epsilon_1\right) \times \frac{P_{L\sim \mathrm{Dir}(1)}\left(X^\top L \leq \mu^* - \epsilon_1 \mid X_1, \ldots, X_n\right)}{1 - P_{L\sim \mathrm{Dir}(1)}\left(X^\top L \leq \mu^* - \epsilon_1 \mid X_1, \ldots, X_n\right)}\right]}_{\text{(B2)}} \\
&\quad + \underbrace{\mathrm{E}_{X_1,\ldots,X_n\sim F_1}\left[\mathbb{1}\left(\mu^* - \epsilon_1 > \frac{1}{n}\sum_{i=1}^{n} X_i\right)\frac{P_{L\sim \mathrm{Dir}(1)}\left(X^\top L \leq \mu^* - \epsilon_1 \mid X_1, \ldots, X_n\right)}{1 - P_{L\sim \mathrm{Dir}(1)}\left(X^\top L \leq \mu^* - \epsilon_1 \mid X_1, \ldots, X_n\right)}\right]}_{\text{(B3)}}.
\end{aligned}
$$

We are going to provide an exponentially small bound to each of these terms. Recall that we denoted $X = (1, X_1, X_2, \ldots, X_n)$ with an additional 1 in the beginning.

D.2.1. BOUNDING (B1)

In this subsection, we provide an upper bound to term (B1), defined as

$$\mathrm{E}_{X_1,\dots,X_n\sim F_1}\left[\mathbb{1}\left(\frac{1}{n}\sum_{i=1}^{n}X_i \geq \mu^* - \frac{\epsilon_1}{2}\right)\frac{P_{L\sim\mathrm{Dir}(1)}\left(X^\top L \leq \mu^* - \epsilon_1 \mid X_1,\dots,X_n\right)}{1 - P_{L\sim\mathrm{Dir}(1)}\left(X^\top L \leq \mu^* - \epsilon_1 \mid X_1,\dots,X_n\right)}\right].$$

Applying the corollary of Lemma 15 in Appendix G.1 on conditional probabilities, for any $\eta \in (0,1)$,

$$\begin{aligned}&P_{L\sim\mathrm{Dir}(1)}\left(X^\top L \leq \mu^* - \epsilon_1 \mid X_1,\dots,X_n\right)\\ &\leq \frac{1}{\eta}\exp\left(-n\left(\mathcal{K}_{\inf}(\tilde{\hat{F}}, 1-\mu^*+\epsilon_1) - \eta\frac{1-\mu^*+\epsilon_1}{\mu^*-\epsilon_1}\right)\right).\end{aligned}$$

But under the assumption $\frac{1}{n}\sum_{i=1}^{n}X_i \geq \mu^* - \frac{\epsilon_1}{2}$, we know that

$$\mathrm{E}_{X\sim\tilde{\hat{F}}}[X] = \mathrm{E}_{X\sim\hat{F}}[1-X] \leq 1 - \mu^* + \frac{\epsilon_1}{2} < 1 - \mu^* + \epsilon_1.$$

Thus, denoting $\delta := \inf_{F,G:\ \mathrm{E}[F]\leq 1-\mu^*+\frac{\epsilon_1}{2}\ \mathrm{E}[G]\geq 1-\mu^*+\epsilon_1} \mathrm{KL}(F\|G) > 0$, we have that, for any $\eta > 0$,

$$P_{L\sim\mathrm{Dir}(1)}\left(X^\top L \leq \mu^* - \epsilon_1 \,\middle|\, X_1,\dots,X_n\right) \leq \frac{1}{\eta}\exp\left(-n\left(\delta - \eta\frac{1-\mu^*+\epsilon_1}{\mu^*-\epsilon_1}\right)\right).$$

In particular, let $\eta \in (0,1)$ such that $\eta\frac{1-\mu^*+\epsilon_1}{\mu^*-\epsilon_1} \leq \frac{\delta}{2}$. For such $\eta \in (0,1)$, we have

$$P_{L\sim\mathrm{Dir}(1)}\left(X^\top L \leq \mu^* - \epsilon_1 \,\middle|\, X_1,\dots,X_n\right) \leq \frac{1}{\eta}\exp\left(-n\frac{\delta}{2}\right).$$

Then, we can decompose

$$\begin{aligned}&\sum_{n=1}^{T}\mathrm{E}_{X_1,\dots,X_n\sim F_1}\left[\mathbb{1}\left(\frac{1}{n}\sum_{i=1}^{n}X_i \geq \mu^* - \frac{\epsilon_1}{2}\right)\frac{P_{L\sim\mathrm{Dir}(1)}\left(X^\top L \leq \mu^* - \epsilon_1 \mid X_1,\dots,X_n\right)}{1 - P_{L\sim\mathrm{Dir}(1)}\left(X^\top L \leq \mu^* - \epsilon_1 \mid X_1,\dots,X_n\right)}\right]\\ &\leq \sum_{n=1}^{T}\mathrm{E}_{X_1,\dots,X_n\sim F_1}\left[\mathbb{1}\left(\frac{1}{n}\sum_{i=1}^{n}X_i \geq \mu^* - \frac{\epsilon_1}{2}\right)\frac{1}{\eta\exp(n\frac{\delta}{2}) - 1}\right]\\ &\leq \sum_{n=1}^{T}\frac{1}{\eta\exp(n\frac{\delta}{2}) - 1},\end{aligned}$$

which proves that

$$\sum_{n=1}^{T}(\mathrm{B1}) = O(1).$$

D.2.2. BOUNDING (B2)

In this subsection, we provide an upper bound to

$$\begin{aligned}(\text{B2}) := \mathrm{E}_{X_1,\dots,X_n\sim F_1}\Bigg[\mathbb{1}\left(\mu^* - \frac{\epsilon_1}{2} > \frac{1}{n}\sum_{i=1}^n X_i \geq \mu^* - \epsilon_1\right) \\ \times \frac{P_{L\sim\mathrm{Dir}(1)}\left(X^\top L \leq \mu^* - \epsilon_1 \mid X_1,\dots,X_n\right)}{1 - P_{L\sim\mathrm{Dir}(1)}\left(X^\top L \leq \mu^* - \epsilon_1 \mid X_1,\dots,X_n\right)}\Bigg] \\ \leq \mathrm{E}_{X_1,\dots,X_n\sim F_1}\Bigg[\mathbb{1}\left(\mu^* - \frac{\epsilon_1}{2} > \frac{1}{n}\sum_{i=1}^n X_i \geq \mu^* - \epsilon_1\right) \\ \times \frac{1}{P_{L\sim\mathrm{Dir}(1)}\left(L^\top X \geq \mu^* - \epsilon_1 \mid X_1,\dots,X_n\right)}\Bigg].\end{aligned}$$

But in the case $\mu^* - \frac{\epsilon_1}{2} > \frac{1}{n}1^\top X \geq \mu^* - \epsilon_1$, we have

$$P_{L\sim\mathrm{Dir}(1)}\left(X^\top L \geq \mu^* - \epsilon_1 \;\middle|\; X_1,\dots,X_n\right) \geq P_{L\sim\mathrm{Dir}(1)}\left(X^\top L \geq \frac{1}{n}\sum_{i=1}^n X_i \;\middle|\; X_1,\dots,X_n\right).$$

But using Lemma 17 in Appendix G.2 on conditional probabilities, for $n \geq 2$, we have

$$\begin{aligned}P_{L\sim\mathrm{Dir}(1)}\left(X^\top L \geq \mu^* - \epsilon_1 \;\middle|\; X_1,\dots,X_n\right) &\geq P_{L\sim\mathrm{Dir}(1)}\left(X^\top L \geq \frac{1}{n}\sum_{i=1}^n X_i \mid X_1,\dots,X_n\right) \\ &\geq \left(1 - \frac{1}{n}\sum_{i=1}^n X_i\right)\frac{1}{25n^2} \\ &\geq \left(1 - \mu^* + \frac{\epsilon_1}{2}\right)\frac{1}{25n^2},\end{aligned}$$

since we are in the case $\mu^* - \frac{\epsilon_1}{2} > \frac{1}{n}\sum_{i=1}^n X_i \geq \mu^* - \epsilon_1$. Therefore, we can bound term (B2).

$$\begin{aligned}(\text{B2}) = \mathrm{E}_{X_1,\dots,X_n\sim F_1}\Bigg[\mathbb{1}\left(\mu^* - \frac{\epsilon_1}{2} > \frac{1}{n}\sum_{i=1}^n X_i \geq \mu^* - \epsilon_1\right) \\ \times \frac{P_{L\sim\mathrm{Dir}(1)}\left(X^\top L \leq \mu^* - \epsilon_1 \mid X_1,\dots,X_n\right)}{1 - P_{L\sim\mathrm{Dir}(1)}\left(X^\top L \leq \mu^* - \epsilon_1 \mid X_1,\dots,X_n\right)}\Bigg] \\ \leq \frac{25n^2}{1 - \mu^* + \frac{\epsilon_1}{2}}\mathrm{E}_{X_1,\dots,X_n\sim F_1}\left[\mathbb{1}\left(\mu^* - \frac{\epsilon_1}{2} > \frac{1}{n}\sum_{i=1}^n X_i \geq \mu^* - \epsilon_1\right)\right].\end{aligned}$$

But Hoeffding's inequality provides the bound

$$\mathrm{E}_{X_1,\dots,X_n\sim F_1}\left[\mathbb{1}\left(\mu^* - \frac{\epsilon_1}{2} > \frac{1}{n}\sum_{i=1}^n X_i \geq \mu^* - \epsilon_1\right)\right] \leq \exp(-n\frac{\epsilon_1^2}{2}).$$

Therefore, combining the results gives, for $n \geq 2$,

$$\text{(B2)} \leq \frac{1}{1-\mu^* + \frac{\epsilon_1}{2}} 25n^2 \exp(-n\frac{\epsilon_1^2}{2}),$$

which proves that:

$$\sum_{n=1}^{T} \text{(B2)} = O(1).$$

D.2.3. BOUNDING (B3)

In this subsection, we are going to provide an upper bound to term (B3), defined as

$$\mathrm{E}_{X_1,\ldots,X_n\sim F_1}\left[\mathbb{1}\left(\mu^* - \epsilon_1 > \frac{1}{n}\sum_{i=1}^{n} X_i\right) \frac{P_{L\sim\mathrm{Dir}(1)}\left(X^\top L \leq \mu^* - \epsilon_1 \mid X_1,\ldots,X_n\right)}{1 - P_{L\sim\mathrm{Dir}(1)}\left(X^\top L \leq \mu^* - \epsilon_1 \mid X_1,\ldots,X_n\right)}\right].$$

Let $M := \lceil \frac{3}{\epsilon_1} \rceil \geq 1$. For any $i \in \{1,\ldots,n\}$ we denote $\tilde{X}_i := \frac{\lfloor MX_i \rfloor}{M}$ and $\tilde{X} := (1, \tilde{X}_1, \ldots, \tilde{X}_n)$. For any $i \in \{0,\ldots,M\}$, let $\alpha_i := |\{j \in \{0,\ldots,n\} : \ \tilde{X}_j = \frac{i}{M}\}|$ be the number of samples that the discretized value is equal to $\frac{i}{M}$. The expectation we are interested in is bounded by

$$\begin{aligned}
\text{(B3)} &\leq \mathrm{E}_{X_1,\ldots,X_n\sim F_1}\left[\mathbb{1}\left(\mu^* - \epsilon_1 > \frac{1}{n}\sum_{i=1}^{n} X_i\right) \frac{1}{P_{L\sim\mathrm{Dir}(1)}\left(X^\top L \geq \mu^* - \epsilon_1 \mid X_1,\ldots,X_n\right)}\right] \\
&\leq \mathrm{E}_{X_1,\ldots,X_n\sim F_1}\left[\mathbb{1}\left(\mu^* - \epsilon_1 > \frac{1}{n}\sum_{i=1}^{n} \tilde{X}_i\right) \frac{1}{P_{L\sim\mathrm{Dir}(1)}\left(\tilde{X}^\top L \geq \mu^* - \epsilon_1 \,\middle|\, X_1,\ldots,X_n\right)}\right] \\
&\leq \mathrm{E}_{X_1,\ldots,X_n\sim F_1}\Bigg[\mathbb{1}\left(\mu^* - \epsilon_1 + \frac{1}{M} > \frac{1}{n}\sum_{i=1}^{n} \tilde{X}_i\right) \\
&\qquad\qquad \times \frac{1}{P_{L\sim\mathrm{Dir}(1)}\left(\tilde{X}^\top L \geq \mu^* - \epsilon_1 + \frac{1}{M} \,\middle|\, X_1,\ldots,X_n\right)}\Bigg] \\
&\leq \mathrm{E}_{X_1,\ldots,X_n\sim F_1}\Bigg[\mathbb{1}\left(\mu^* - \frac{2\epsilon_1}{3} > \frac{1}{n}\sum_{i=1}^{n} \tilde{X}_i\right) \\
&\qquad\qquad \times \frac{1}{P_{L\sim\mathrm{Dir}(1)}\left(\tilde{X}^\top L \geq \mu^* - \frac{2\epsilon_1}{3} \,\middle|\, X_1,\ldots,X_n\right)}\Bigg].
\end{aligned}$$

Recall that $P_\alpha = \frac{1}{n+1}(\alpha_0, \alpha_1, \ldots, \alpha_n)$ is the normalization of α. We denote $S := \{x \in [0,1]^{n+1} : 1^\top x = 1,\ u^\top x \geq \mu^* - \frac{2\epsilon_1}{3}\}$, and $P^* := \arg\min_{x\in S} \mathrm{KL}(P_\alpha \| x)$. Using Lemma 14 from Appendix F.2, we know that

$$P_{\tilde{L}\sim\mathrm{Dir}(\alpha)}\left(u^\top \tilde{L} \geq \mu\right) \geq C_2(n+1)^{-\frac{M}{2}} \exp(-(n+1)\mathrm{KL}(P_\alpha \| P^*)) \frac{P_{\alpha_M}}{P^*_M},$$

where $C_2 = \left(\frac{1}{\sqrt{2\pi}}\right)^M e^{-\frac{M+1}{12}}$, since $N = \sum_{i=0}^M \alpha_i = n+1$ in this setting. Then, we can clearly derive the lower bound

$$\begin{aligned} P_{\tilde{L}\sim\mathrm{Dir}(\alpha)}\left(u^\top \tilde{L} \geq \mu\right) &\geq C_2(n+1)^{-\frac{M}{2}} \exp(-(n+1)\mathrm{KL}(P_\alpha\|P^*))P_{\alpha_M} \\ &\geq C_2(n+1)^{-\frac{M}{2}-1} \exp(-(n+1)\mathrm{KL}(P_\alpha\|P^*)). \end{aligned}$$

Reinjecting in the computation and replacing C_2 by its value gives

$$\begin{aligned} \text{(B3)} &\leq \mathrm{E}_{X_1,\dots,X_n\sim F_1}\left[\mathbb{1}\left(\mu^* - \frac{2\epsilon_1}{3} > \frac{1}{n}\sum_{i=1}^n \tilde{X}_i\right) e^{\frac{M+1}{12}}(\sqrt{2\pi})^M (n+1)^{\frac{M}{2}+1}\right. \\ &\qquad\qquad \left. \times \exp((n+1)\mathrm{KL}(P_\alpha\|P^*))\right] \\ &= e^{\frac{M+1}{12}}(\sqrt{2\pi})^M (n+1)^{\frac{M}{2}+1} \\ &\qquad \times \mathrm{E}_{X_1,\dots,X_n\sim F_1}\left[\mathbb{1}\left(\mu^* - \frac{2\epsilon_1}{3} > \frac{1}{n}u^\top\alpha\right)\exp((n+1)\mathrm{KL}(P_\alpha\|P^*))\right] \\ &\leq e^{\frac{M+1}{12}}(\sqrt{2\pi})^M (n+1)^{\frac{M}{2}+1} \\ &\qquad \times \mathrm{E}_{X_1,\dots,X_n\sim F_1}\left[\mathbb{1}\left(\mu^* - \frac{2\epsilon_1}{3} > \frac{u^\top\alpha}{n+1}\right)\exp((n+1)\mathrm{KL}(P_\alpha\|P^*))\right]. \end{aligned}$$

We denote by p the distribution of $\tilde{X}_1$ and Y the random variable denoting the distribution of the α. Denoting $A := \{\alpha \in \{0,\dots,n+1\}^{M+1} :\ 1^\top\alpha = n+1,\ \frac{1}{n+1}u^\top\alpha \leq \mu\}$ and $\mathrm{H}(P)$ the entropy of the multinomial distribution of parameter P, we can directly compute the remaining expectation.

$$\begin{aligned} &\mathrm{E}_{X_1,\dots,X_n\sim F_1}\left[\mathbb{1}\left(\mu^* - \frac{2\epsilon_1}{3} > \frac{1}{n+1}u^\top\alpha\right)\exp((n+1)\mathrm{KL}(P_\alpha\|P^*))\right] \\ &= \sum_{\alpha\in A} P(Y=\alpha)\exp((n+1)\mathrm{KL}(P_\alpha\|P^*)) \\ &= \sum_{\alpha\in A} \exp(-(n+1)\mathrm{H}(P_\alpha))\exp(-(n+1)\mathrm{KL}(P_\alpha\|p))\exp((n+1)\mathrm{KL}(P_\alpha\|P^*)) \\ &= \sum_{\alpha\in A} \exp(-(n+1)\mathrm{H}(P_\alpha))\exp\left(-(n+1)(\mathrm{KL}(P_\alpha\|p) - \mathrm{KL}(P_\alpha\|P^*))\right). \end{aligned}$$

Now, recall that $u^\top P_\alpha \leq \mu^* - \frac{2\epsilon_1}{3}$ and $u^\top p = \mu^*$. Then, using a result from Honda and Takemura (2010, Lemma 13), we can bound

$$\begin{aligned} \mathrm{KL}(P_\alpha\|p) - \mathrm{KL}(P_\alpha\|P^*) &= \mathrm{KL}(P_\alpha\|p) - \mathcal{K}_{\inf}\left(P_\alpha, \mu^* - \frac{2\epsilon_1}{3}\right) \\ &\geq \mathcal{K}_{\inf}(P_\alpha,\mu^*) - \mathcal{K}_{\inf}\left(P_\alpha, \mu^* - \frac{2\epsilon_1}{3}\right) \\ &\geq \frac{(\mu^* - (\mu^* - \frac{2\epsilon_1}{3}))^2}{2\mu^*(1-\mu^*+\frac{2\epsilon_1}{3})} \end{aligned}$$

$$= \frac{2\epsilon_1^2}{9\mu^*(1-\mu^*+\frac{2\epsilon_1}{3})}$$
$$> 0.$$

Denoting $C := \frac{2\epsilon_1^2}{9\mu^*(1-\mu^*+\frac{2\epsilon_1}{3})}$, we can then bound

$$\begin{aligned}
&\sum_{\alpha\in A} \exp(-(n+1)\mathrm{H}(P_\alpha)) \exp(-(n+1)(\mathrm{KL}(P_\alpha\|p) - \mathrm{KL}(P_\alpha\|P^*))) \\
&\leq \sum_{\alpha\in A} \exp(-(n+1)\mathrm{H}(P_\alpha)) \exp(-C(n+1)) \\
&\leq \sum_{\alpha\in A} \exp(-C(n+1)) \\
&= |A| \exp(-C(n+1)) \\
&\leq (n+1)^{M+1} \exp(-C(n+1)).
\end{aligned}$$

Reinjecting in the computation, we can then bound

$$\begin{aligned}
\text{(B3)} &\leq e^{\frac{M+1}{12}} (\sqrt{2\pi})^M (n+1)^{\frac{M}{2}+1} \\
&\qquad \times \mathrm{E}_{X_1,\ldots,X_n\sim F_1}\left[\mathbb{1}\left(\mu^* - \frac{2\epsilon_1}{3} > \frac{u^\top \alpha}{n+1}\right) \exp((n+1)\mathrm{KL}(P_\alpha\|P^*))\right] \\
&\leq e^{\frac{M+1}{12}} (\sqrt{2\pi})^M (n+1)^{\frac{M}{2}+1} (n+1)^{M+1} \exp(-C(n+1)) \\
&= e^{\frac{M+1}{12}} (\sqrt{2\pi})^M (n+1)^{\frac{3M}{2}+2} \exp(-C(n+1)).
\end{aligned}$$

which proves that

$$\sum_{n=1}^{T} (B3) = O(1).$$

Appendix E. Application of Stirling Formula

In this small section, we are going to prove the following Lemma 12.

Lemma 12 *Let $\alpha := (\alpha_0, \ldots, \alpha_M)$ and $N := \sum_{j=0}^M \alpha_j$. Assume that for any $i \in \{0, \ldots, M\}, \alpha_i \geq 1$. We will denote $P_\alpha := \frac{1}{N}\alpha$, which implies that $1^\top P_\alpha = 1$. Then:*

$$\left(\frac{1}{\sqrt{2\pi}}\right)^M e^{-(M+1)/12} N^{-\frac{M}{2}} \prod_{i=0}^M \frac{1}{\sqrt{P_{\alpha_i}}} \leq \frac{\Gamma(N)}{N^{N-1}} \prod_{i=0}^M \frac{\alpha_i^{\alpha_i-1}}{\Gamma(\alpha_i)},$$

$$\frac{\Gamma(N)}{N^{N-1}} \prod_{i=0}^M \frac{\alpha_i^{\alpha_i-1}}{\Gamma(\alpha_i)} \leq \left(\frac{1}{\sqrt{2\pi}}\right)^M e^{1/12} N^{-\frac{M}{2}} \prod_{i=0}^M \frac{1}{\sqrt{P_{\alpha_i}}}.$$

Proof Applying Stirling's formula:

$$\Gamma(N) \geq \sqrt{2\pi} N^{N-1/2} e^{-N}.$$

Therefore,

$$\frac{\Gamma(N)}{N^{N-1}} \geq \sqrt{2\pi} e^{-N} \sqrt{N}.$$

We can apply Stirling's formula to each of the α_i, for $i \in \{0, \ldots, M\}$.

$$\Gamma(\alpha_i) \leq \sqrt{2\pi} e^{1/12} \alpha_i^{\alpha_i - 1/2} e^{-\alpha_i}.$$

Therefore,

$$\frac{\Gamma(\alpha_i)}{\alpha_i^{\alpha_i - 1}} \leq \sqrt{2\pi} e^{1/12} \sqrt{\alpha_i} e^{-\alpha_i}.$$

Using the fact that $\sum_{i=0}^{M} \alpha_i = N$, we deduce that

$$\begin{aligned}\prod_{i=0}^{M} \frac{\Gamma(\alpha_i)}{\alpha_i^{\alpha_i - 1}} &\leq \left(\sqrt{2\pi}\right)^{M+1} e^{(M+1)/12} e^{-N} \prod_{i=0}^{M} \sqrt{\alpha_i} \\ &= \left(\sqrt{2\pi}\right)^{M+1} e^{(M+1)/12} e^{-N} N^{\frac{M+1}{2}} \prod_{i=0}^{M} \sqrt{P_{\alpha_i}}.\end{aligned}$$

Then,

$$\prod_{i=0}^{M} \frac{\alpha_i^{\alpha_i - 1}}{\Gamma(\alpha_i)} \geq \left(\frac{1}{\sqrt{2\pi}}\right)^{M+1} e^{-(M+1)/12} e^{N} N^{-\frac{M+1}{2}} \prod_{i=0}^{M} \frac{1}{\sqrt{P_{\alpha_i}}}.$$

Therefore,

$$\frac{\Gamma(N)}{N^{N-1}} \prod_{i=0}^{M} \frac{\alpha_i^{\alpha_i - 1}}{\Gamma(\alpha_i)} \geq \left(\frac{1}{\sqrt{2\pi}}\right)^{M} e^{-(M+1)/12} N^{-\frac{M}{2}} \prod_{i=0}^{M} \frac{1}{\sqrt{P_{\alpha_i}}},$$

which is the desired lower bound.

Then, we try to derive the desired upper bound. Applying Stirling's formula,

$$\Gamma(N) \leq \sqrt{2\pi} e^{1/12} N^{N-1/2} e^{-N}.$$

Therefore,

$$\frac{\Gamma(N)}{N^{N-1}} \leq \sqrt{2\pi} e^{1/12} e^{-N} \sqrt{N}.$$

We can apply Stirling's formula to each of the α_i, for $i \in \{0, \ldots, M\}$.

$$\Gamma(\alpha_i) \geq \sqrt{2\pi} \alpha_i^{\alpha_i - 1/2} e^{-\alpha_i}$$

Therefore,

$$\frac{\Gamma(\alpha_i)}{\alpha_i^{\alpha_i - 1}} \geq \sqrt{2\pi} \sqrt{\alpha_i} e^{-\alpha_i}.$$

Using the fact that $\sum_{i=0}^{M} \alpha_i = N$, we deduce that

$$\begin{aligned}\prod_{i=0}^{M} \frac{\Gamma(\alpha_i)}{\alpha_i^{\alpha_i - 1}} &\geq \left(\sqrt{2\pi}\right)^{M+1} e^{-N} \prod_{i=0}^{M} \sqrt{\alpha_i} \\ &= \left(\sqrt{2\pi}\right)^{M+1} e^{-N} N^{\frac{M+1}{2}} \prod_{i=0}^{M} \sqrt{P_{\alpha_i}}.\end{aligned}$$

Then,

$$\prod_{i=0}^{M} \frac{\alpha_i^{\alpha_i - 1}}{\Gamma(\alpha_i)} \leq \left(\frac{1}{\sqrt{2\pi}}\right)^{M+1} e^N N^{-\frac{M+1}{2}} \prod_{i=0}^{M} \frac{1}{\sqrt{P_{\alpha_i}}}.$$

Therefore,

$$\frac{\Gamma(N)}{N^{N-1}} \prod_{i=0}^{M} \frac{\alpha_i^{\alpha_i - 1}}{\Gamma(\alpha_i)} \leq \left(\frac{1}{\sqrt{2\pi}}\right)^{M} e^{1/12} N^{-\frac{M}{2}} \prod_{i=0}^{M} \frac{1}{\sqrt{P_{\alpha_i}}},$$

which is the desired bound. ■

Appendix F. Bounds for Tails of Dirichlet Distributions

In this section, we prove lower and upper bounds of the probability of the end-tail of a Dirichlet distribution.

F.1. Upper Bounds for Tails of Dirichlet Distributions

In this section, we prove Lemma 13 below.

Lemma 13 *Assume $L \sim \mathrm{Dir}(\alpha_0, \alpha_1, \ldots, \alpha_M)$ a Dirichlet distribution over the probability simplex P. We assume that $1^\top \alpha = n + M + 1$ and for any $j \in \{0, 1, \ldots, M\}, \alpha_j \geq 1$. We will denote $P_\alpha = \frac{1}{n+M+1}\alpha$, the mean of the Dirichlet distribution. Let $S \subset P$, a closed convex set included in the probability simplex. Then, the following upper bound holds.*

$$P_{L \sim \mathrm{Dir}(\alpha)}(L \in S) \leq C_1 (n + M + 1)^{M/2} \exp(-(n + M + 1)\mathrm{KL}(P_\alpha \| P^*)),$$

where we denoted $P^* := \arg\min_{x \in S} \mathrm{KL}(P_\alpha \| x)$ and $C_1 := \frac{e^{1/12}}{\Gamma(M+1)} \left(\frac{1}{\sqrt{2\pi}}\right)^M$.

If P_α does not belong to S, this provides an exponential upper bound to an end-tail probability.

Proof We keep using the notation $N = n + M + 1$. Given that the Dirichlet distribution is the conjugate distribution of the multinomial distribution, we can write the following formula.

$$\begin{aligned} P_{L \sim \mathrm{Dir}(\alpha)}(L \in S) &= \frac{\int_{x \in S} \pi(x) P_{Z \sim \mathrm{Mult}(N,x)}(Z = \alpha) dx}{\int_{x \in P} \pi(x) P_{Z \sim \mathrm{Mult}(N,x)}(Z = \alpha) dx} \\ &= \frac{\int_{x \in S} P_{Z \sim \mathrm{Mult}(N,x)}(Z = \alpha) dx}{\int_{x \in P} P_{Z \sim \mathrm{Mult}(N,x)}(Z = \alpha) dx}, \end{aligned}$$

where π is the prior distribution, chosen as the uniform distribution over the simplex P. We are going to rewrite differently the term $\int_{x \in S} P_{Z \sim \mathrm{Mult}(N,x)}(Z = \alpha) dx$ and the term $\int_{x \in P} P_{Z \sim \mathrm{Mult}(N,x)}(Z = \alpha) dx$. Denoting $\mathrm{H}(P)$ the entropy of the multinomial distribution of parameter P, let us rewrite the numerator of the fraction:

$$\begin{aligned} &\int_{x \in S} P_{Z \sim \mathrm{Mult}(N,x)}(Z = \alpha) dx \\ &= \int_{x \in S} \exp(-N\mathrm{H}(P_\alpha)) \exp(-N\mathrm{KL}(P_\alpha \| x)) dx \\ &= \exp(-N\mathrm{H}(P_\alpha)) \exp(-N\mathrm{KL}(P_\alpha \| P^*)) \int_{x \in S} \exp(-N[\mathrm{KL}(P_\alpha \| x) - \mathrm{KL}(P_\alpha \| P^*)]) dx, \end{aligned}$$

where we denoted $P^* := \arg\min_{x \in S} \mathrm{KL}(P_\alpha \| x)$. Denoting $B(\alpha) := \frac{\prod_{i=0}^M \Gamma(\alpha_i)}{\Gamma(\sum_{i=0}^M \alpha_i)} = \frac{\prod_{i=0}^M \Gamma(\alpha_i)}{\Gamma(N)}$, recall that the Dirichlet distribution of parameters $\alpha_0, \dots, \alpha_M$, $\mathrm{Dir}(\alpha_0, \dots, \alpha_M)$ has density function $\frac{1}{B(\alpha)} \prod_{i=0}^M x_i^{\alpha_i - 1}$ over the probability simplex P. Rewriting the denominator of the fraction, we obtain

$$\begin{aligned}
\int_{x \in P} P_{Z \sim \mathrm{Mult}(N,x)}(Z = \alpha) dx &= \int_x \exp(-N\mathrm{H}(P_\alpha)) \exp(-N\mathrm{KL}(P_\alpha \| x)) dx \\
&= \exp(-N\mathrm{H}(P_\alpha)) \int_{x \in P} \prod_{i=0}^M \left(\frac{x_i}{P_{\alpha_i}} \right)^{\alpha_i} dx_i \\
&= \exp(-N\mathrm{H}(P_\alpha)) \frac{\int_{x \in P} \frac{1}{B(\alpha+1)} \prod_{i=0}^M (x_i)^{\alpha_i} dx_i}{\frac{1}{B(\alpha+1)} \prod_{i=0}^M (P_{\alpha_i})^{\alpha_i}} \\
&= \exp(-N\mathrm{H}(P_\alpha)) \frac{1}{\frac{1}{B(\alpha+1)} \prod_{i=0}^M (P_{\alpha_i})^{\alpha_i}}.
\end{aligned}$$

We want to divide the upper term by the lower term and see what we have. Then, dividing both the terms we previously obtained, we have

$$\begin{aligned}
\frac{\int_{x \in S} P(Z = \alpha) dx}{\int_{x \in P} P(Z = \alpha) dx} &= \exp(-N\mathrm{KL}(P_\alpha \| P^*)) \\
&\quad \times \int_{x \in S} \exp(-N[\mathrm{KL}(P_\alpha \| x) - \mathrm{KL}(P_\alpha \| P^*)]) dx \frac{1}{B(\alpha+1)} \prod_{i=0}^M (P_{\alpha_i})^{\alpha_i} \\
&= \exp(-N\mathrm{KL}(P_\alpha \| P^*)) \frac{B(\alpha)}{B(\alpha+1)} \prod_{i=0}^M P_{\alpha_i} \\
&\quad \times \int_{x \in S} \exp(-N[\mathrm{KL}(P_\alpha \| x) - \mathrm{KL}(P_\alpha \| P^*)]) dx \frac{1}{B(\alpha)} \prod_{i=0}^M (P_{\alpha_i})^{\alpha_i - 1}.
\end{aligned}$$

Then, it is easy to bound $\frac{B(\alpha)}{B(\alpha+1)} \leq 1$. Then,

$$\begin{aligned}
\frac{\int_{x \in S} P(Z = \alpha) dx}{\int_{x \in P} P(Z = \alpha) dx} &\leq \exp(-N\mathrm{KL}(P_\alpha \| P^*)) \prod_{i=0}^M P_{\alpha_i} \\
&\quad \times \int_{x \in S} \exp(-N[\mathrm{KL}(P_\alpha \| x) - \mathrm{KL}(P_\alpha \| P^*)]) dx \frac{1}{B(\alpha)} \prod_{i=0}^M (P_{\alpha_i})^{\alpha_i - 1}.
\end{aligned}$$

We will try to apply a simple bound, using Stirling's formula. Recall that we want to provide an upper bound to $\int_{x \in S} \exp(-N(\mathrm{KL}(P_\alpha \| x) - \mathrm{KL}(P_\alpha \| P^*))) dx$, but the integrand is bounded by 1 by definition of P^*. Therefore, this can be bounded by $\int_{x \in S} 1 dx \leq \int_{x \in P} 1 dx = \frac{1}{\Gamma(M+1)}$.

We also want to provide a lower bound to $\frac{1}{\frac{1}{B(\alpha)}\prod_{i=0}^{M}(P_{\alpha_i})^{\alpha_i-1}}$. Indeed, we would like to bound the quotient of the first one divided by the second one, by a polynomial of N. Let us compute

$$\begin{aligned}
&\int_{x\in S}\exp(-N(\mathrm{KL}(P_\alpha\|x)-\mathrm{KL}(P_\alpha\|P^*)))dx\frac{1}{B(\alpha)}\prod_{i=0}^{M}(P_{\alpha_i})^{\alpha_i-1}\\
&\leq \frac{1}{\Gamma(M+1)}\frac{1}{B(\alpha)}\prod_{i=0}^{M}(P_{\alpha_i})^{\alpha_i-1}\\
&= \frac{1}{\Gamma(M+1)}\frac{\Gamma(N)}{\prod_{i=0}^{M}\Gamma(\alpha_i)}\prod_{i=0}^{M}(P_{\alpha_i})^{\alpha_i-1}\\
&= \frac{1}{\Gamma(M+1)}\frac{\Gamma(N)}{\prod_{i=0}^{M}\Gamma(\alpha_i)}\prod_{i=0}^{M}\frac{\alpha_i^{\alpha_i-1}}{N^{\alpha_i-1}}\\
&= \frac{N^M}{\Gamma(M+1)}\frac{\Gamma(N)}{N^{N-1}}\prod_{i=0}^{M}\frac{\alpha_i^{\alpha_i-1}}{\Gamma(\alpha_i)}.
\end{aligned}$$

Replacing N by $n+M+1$, we obtain

$$\begin{aligned}
&\int_{x\in S}\exp(-(n+M+1)(\mathrm{KL}(P_\alpha\|x)-\mathrm{KL}(P_\alpha\|P^*)))dx\frac{1}{B(\alpha)}\prod_{i=0}^{M}(P_{\alpha_i})^{\alpha_i-1}\\
&\qquad\leq \frac{(n+M+1)^M}{\Gamma(M+1)}\frac{\Gamma(n+M+1)}{(n+M+1)^{n+M}}\prod_{i=0}^{M}\frac{\alpha_i^{\alpha_i-1}}{\Gamma(\alpha_i)}.
\end{aligned}$$

We can then use the upper bound of Lemma 12 in Appendix E (application of Stirling formula), which gives

$$\begin{aligned}
&\int_{x\in S}\exp(-(n+M+1)(\mathrm{KL}(P_\alpha\|x)-\mathrm{KL}(P_\alpha\|P^*)))dx\frac{1}{B(\alpha)}\prod_{i=0}^{M}(P_{\alpha_i})^{\alpha_i-1}\\
&\leq \frac{(n+M+1)^M}{\Gamma(M+1)}\left(\frac{1}{\sqrt{2\pi}}\right)^M e^{1/12}(n+M+1)^{-\frac{M}{2}}\prod_{i=0}^{M}\frac{1}{\sqrt{P_{\alpha_i}}}\\
&= C_1(n+M+1)^{\frac{M}{2}}\prod_{i=0}^{M}\frac{1}{\sqrt{P_{\alpha_i}}},
\end{aligned}$$

where we denoted $C_1=\frac{e^{1/12}}{\Gamma(M+1)}\left(\frac{1}{\sqrt{2\pi}}\right)^M$. Therefore, we have reached the desired upper bound.

$$\begin{aligned}
P_{L\sim\mathrm{Dir}(\alpha)}(L\in S) &\leq C_1\left(\prod_{i=0}^{M}\sqrt{P_{\alpha_i}}\right)(n+M+1)^{M/2}\exp(-(n+M+1)\mathrm{KL}(P_\alpha\|P^*))\\
&\leq C_1(n+M+1)^{M/2}\exp(-(n+M+1)\mathrm{KL}(P_\alpha\|P^*)).
\end{aligned}$$

■

F.2. Lower Bounds for Tails of Dirichlet Distributions

Let $L \sim \mathrm{Dir}(\alpha)$ be a random variable following Dirichlet distribution with parameter $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_M)$ such that $\alpha_M \geq 1$. We sometimes consider the case $\alpha_i = 0$, and in this case we redefine $L \sim \mathrm{Dir}(\alpha)$ as a random variable such that $L_i = 0$ and $(L_0, \ldots, L_{i-1}, L_{i+1}, \ldots, L_M) \sim \mathrm{Dir}(\alpha_0, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots, \alpha_M)$.

Let us denote $S := \{x \in [0,1]^{M+1} : \ 1^\top x = 1, \ u^\top x \geq \frac{1}{N} u^\top \alpha + \Delta\}$, where $\Delta \geq 0$ and $N = \sum_{i=0}^{M} \alpha_i$. We will also denote $P^* := \arg\min_{x \in S} \mathrm{KL}(P_\alpha \| x)$. In this section, we prove Lemma 14 given below.

Lemma 14 *For n sufficiently big:*

$$P_{L \sim \mathrm{Dir}(\alpha)} \left(u^\top L \geq \frac{u^\top \alpha}{N} + \Delta \right) \geq C_2 N^{-\frac{M}{2}} \exp(-N \mathrm{KL}(P_\alpha \| P^*)) \frac{P_{\alpha_M}}{P^*_M},$$

where we denoted $C_2 := \left(\frac{1}{\sqrt{2\pi}} \right)^M e^{-(M+1)/12}$.

This lemma provides a lower bound to the tails of Dirichlet distributions.

Proof Note that we always have $P^*_i > 0$ for any i such that $\alpha_i > 0$ from definition $P^* := \arg\min_{x \in S} \mathrm{KL}(P_\alpha \| x) = \sum_{i=0}^{M} P_{\alpha_i} \log \frac{P_{\alpha_i}}{x_i}$.

Let $S := \{x \in [0,1]^{M+1} : \ 1^\top x = 1, \ u^\top x \geq \frac{1}{N} u^\top \alpha + \Delta\}$. Let $S_2 := \{x \in [0,1]^{M+1} : 1^\top x = 1, \ \forall i \in \{0, \ldots, M-1\}, \ x_i \in [0, P^*_i]\}$. Then, we notice that $S_2 \subset S$, and therefore

$$P_{L \sim \mathrm{Dir}(\alpha)}(L \in S) \geq P_{L \sim \mathrm{Dir}(\alpha)}(L \in S_2).$$

Let us denote $\mathcal{I} = \{i \in \{0, 1, \ldots, M\} : \alpha_i > 0\}$, $\mathcal{I}^- = \{i \in \{0, 1, \ldots, M-1\} : \alpha_i > 0\}$, and $S_3 := \{x \in [0,1]^M : \ \forall i \in \mathcal{I}^-, \ x_i \in [0, P^*_i]\}$. Then we have

$$\begin{aligned}
P_{L \sim \mathrm{Dir}(\alpha)} \left(u^\top L \geq \frac{1}{N} u^\top \alpha + \Delta \right) &= P_{L \sim \mathrm{Dir}(\alpha)}(L \in S) \\
&\geq P_{L \sim \mathrm{Dir}(\alpha)}(L \in S_2) \\
&= \frac{\Gamma(N)}{\prod_{i \in \mathcal{I}} \Gamma(\alpha_i)} \int_{x \in S_3} \prod_{i \in \mathcal{I}^-} x_i^{\alpha_i - 1} \left(1 - \sum_{i=0}^{M-1} x_i \right)^{\alpha_M - 1} \prod_{i \in \mathcal{I}^-} dx_i \\
&\geq \frac{\Gamma(N)}{\prod_{i \in \mathcal{I}} \Gamma(\alpha_i)} (P^*_M)^{\alpha_M - 1} \int_{x \in S_3} \prod_{i \in \mathcal{I}^-} x_i^{\alpha_i - 1} dx_i \\
&= \frac{\Gamma(N)}{\prod_{i \in \mathcal{I}} \Gamma(\alpha_i)} (P^*_M)^{\alpha_M - 1} \prod_{i \in \mathcal{I}^-} \int_{x_i = 0}^{P^*_i} x_i^{\alpha_i - 1} dx_i \\
&= \frac{\Gamma(N)}{\prod_{i \in \mathcal{I}} \Gamma(\alpha_i)} (P^*_M)^{\alpha_M - 1} \prod_{i \in \mathcal{I}^-} \frac{(P^*_i)^{\alpha_i}}{\alpha_i}.
\end{aligned}$$

Then, recall that for any $i \in \{0, \dots, M\}, \alpha_i = NP_{\alpha_i}$, we can perform the computation

$$\begin{aligned}
\frac{\Gamma(N)}{\prod_{i\in\mathcal{I}}\Gamma(\alpha_i)}(P_M^*)^{\alpha_M-1}\prod_{i\in\mathcal{I}^-}\frac{(P_i^*)^{\alpha_i}}{\alpha_i} &= \frac{\Gamma(N)}{\prod_{i\in\mathcal{I}}\Gamma(\alpha_i)}\frac{1}{P_M^*}\prod_{i\in\mathcal{I}}(P_i^*)^{\alpha_i}\prod_{i\in\mathcal{I}^-}\frac{1}{\alpha_i} \\
&= \frac{\Gamma(N)}{\prod_{i\in\mathcal{I}}\Gamma(\alpha_i)}\frac{1}{P_M^*}\prod_{i\in\mathcal{I}}\left(\frac{P_i^*}{P_{\alpha_i}}\right)^{\alpha_i}\prod_{i\in\mathcal{I}}(P_{\alpha_i})^{\alpha_i}\prod_{i\in\mathcal{I}^-}\frac{1}{NP_{\alpha_i}} \\
&\geq \frac{\Gamma(N)}{\prod_{i\in\mathcal{I}}\Gamma(\alpha_i)}\frac{P_{\alpha_M}}{N^M P_M^*}\prod_{i\in\mathcal{I}}\left(\frac{P_i^*}{P_{\alpha_i}}\right)^{\alpha_i}\prod_{i\in\mathcal{I}}(P_{\alpha_i})^{\alpha_i-1}.
\end{aligned}$$

Here note that

$$\prod_{i\in\mathcal{I}}\left(\frac{P_i^*}{P_{\alpha_i}}\right)^{\alpha_i} = \exp(-N\mathrm{KL}(P_\alpha\|P^*)).$$

Then, reinjecting in the computation, we have

$$\begin{aligned}
&P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L \geq \frac{1}{N}u^\top\alpha + \Delta\right) \\
&\geq \frac{\Gamma(N)}{\prod_{i\in\mathcal{I}}\Gamma(\alpha_i)}\frac{P_{\alpha_M}}{N^M P_M^*}\exp(-N\mathrm{KL}(P_\alpha\|P^*))\prod_{i\in\mathcal{I}}(P_{\alpha_i})^{\alpha_i-1} \\
&= \frac{\Gamma(N)}{\prod_{i\in\mathcal{I}}\Gamma(\alpha_i)}\frac{P_{\alpha_M}}{N^M P_M^*}\exp(-N\mathrm{KL}(P_\alpha\|P^*))\prod_{i\in\mathcal{I}}\left(\frac{\alpha_i}{N}\right)^{\alpha_i-1} \\
&= \frac{\Gamma(N)}{N^{n+M}}\frac{P_{\alpha_M}}{P_M^*}\exp(-N\mathrm{KL}(P_\alpha\|P^*))\prod_{i\in\mathcal{I}}\frac{\alpha_i^{\alpha_i-1}}{\Gamma(\alpha_i)}.
\end{aligned}$$

Now, using the results of Lemma 12 in Appendix E (application of Stirling formula), we have

$$\begin{aligned}
&P_{L\sim\mathrm{Dir}(\alpha)}\left(u^\top L \geq \frac{1}{N}u^\top\alpha + \Delta\right) \\
&\geq \frac{\Gamma(N)}{N^{n+M}}\exp(-N\mathrm{KL}(P_\alpha\|P^*))\frac{P_{\alpha_M}}{P_M^*}\prod_{i\in\mathcal{I}}\frac{\alpha_i^{\alpha_i-1}}{\Gamma(\alpha_i)} \\
&\geq \left(\frac{1}{\sqrt{2\pi}}\right)^M e^{-(M+1)/12}N^{-\frac{M}{2}}\exp(-N\mathrm{KL}(P_\alpha\|P^*))\frac{P_{\alpha_M}}{P_M^*}\prod_{i\in\mathcal{I}}\frac{1}{\sqrt{P_{\alpha_i}}} \\
&= C_2 N^{-\frac{M}{2}}\exp(-N\mathrm{KL}(P_\alpha\|P^*))\frac{P_{\alpha_M}}{P_M^*}\prod_{i\in\mathcal{I}}\frac{1}{\sqrt{P_{\alpha_i}}} \\
&\geq C_2 N^{-\frac{M}{2}}\exp(-N\mathrm{KL}(P_\alpha\|P^*))\frac{P_{\alpha_M}}{P_M^*},
\end{aligned}$$

where we denoted $C_2 := \left(\frac{1}{\sqrt{2\pi}}\right)^M e^{-(M+1)/12}$, which is the result we wanted to prove. ■

Appendix G. Bounds for Conditional Random Average

In this section, we provide, for different values of μ, upper and lower bounds on the probability $P(L^\top X \geq \mu \mid X)$ where $L \sim \mathrm{Dir}(1)$.

G.1. Upper Bound for Conditional Random Average

Let $\mu \in [0,1]$. Assume $V = L^\top X$, where $L \sim \mathrm{Dir}(1^{n+1})$ a Dirichlet distribution, and denote $X = (\xi, X_1, \dots, X_n)$, where $X_1, \dots, X_n$ are iid random variables of distribution P, where ξ is a deterministic constant equal to 0 or 1. We want to prove the following upper bound on the following conditional probability.

Lemma 15 *For any $\eta \in (0,1)$,*

$$P_{L\sim\mathrm{Dir}(1^{n+1})}\left(L^\top X \geq \mu \;\middle|\; X\right) \leq \frac{1}{\eta}\exp\left(-n\left(\mathcal{K}_{\inf}(\hat{F},\mu) - \eta\frac{\mu}{1-\mu}\right)\right),$$

where we denoted $\mathcal{K}_{\inf}(F,\mu) := \inf_{G:\,\mathrm{E}[G]\geq\mu}\mathrm{KL}(F\|G)$ and $\hat{F}$ the empirical distribution of $X_1,\dots,X_n$.

This provides an exponential upper bound to an end-tail probability.

Corollary 16 *Applying this result to $1-X$ provides an upper bound for $P_{L\sim\mathrm{Dir}(1^{n+1})}(L^\top X \geq \mu|X)$. For any $\eta \in (0,1)$,*

$$\begin{aligned}
P_{L\sim\mathrm{Dir}(1^{n+1})}\left(L^\top X \leq \mu \;\middle|\; X\right) &= P_{L\sim\mathrm{Dir}(1^{n+1})}\left(1 - L^\top X \geq 1-\mu \;\middle|\; X\right)\\
&= P_{L\sim\mathrm{Dir}(1^{n+1})}\left(L^\top(1-X) \geq 1-\mu \;\middle|\; X\right)\\
&\leq \frac{1}{\eta}\exp\left(-n\left(\mathcal{K}_{\inf}(\tilde{\hat{F}},1-\mu) - \eta\frac{1-\mu}{\mu}\right)\right)
\end{aligned}$$

where we denoted by $\tilde{\hat{F}}$ the empirical distribution of $(1-X_1,\dots,1-X_n)$.

Proof Let $R_0,\dots,R_n$ iid exponential random variables of distribution $\mathcal{E}(1)$, and let us denote, for any $i \in \{0,\dots,n\}$, $R'_i := \frac{R_i}{\sum_{j=0}^n R_j}$. $L \sim \mathrm{Dir}(1^{n+1})$, and thus it has the same distribution as $R' = (R'_0,\dots,R'_n)$ and we can rewrite the probability

$$\begin{aligned}
P_{L\sim\mathrm{Dir}(1^{n+1})}\left(L^\top X \geq \mu \;\middle|\; X\right) &= P_{R'\sim\mathrm{Dir}(1^{n+1})}(R'^\top X \geq \mu \;\middle|\; X)\\
&= P_{R_0,\dots,R_n\sim\mathcal{E}(1)}\left(\frac{\sum_{i=0}^n R_i X_i}{\sum_{i=0}^n R_i} \geq \mu \;\middle|\; X\right)\\
&= P_{R_0,\dots,R_n\sim\mathcal{E}(1)}\left(\sum_{i=0}^n (X_i-\mu)R_i \geq 0 \;\middle|\; X\right).
\end{aligned}$$

Then, using Markov's inequality, for any $t \in [0, \frac{1}{1-\mu})$, we know that

$$\begin{aligned}
P_{L\sim\mathrm{Dir}(1^{n+1})}\left(L^\top X \geq \mu \;\middle|\; X\right) &\leq \mathrm{E}\left[\exp\left(t\sum_{i=0}^n (X_i-\mu)R_i\right) \;\middle|\; X_i\right]\\
&= \prod_{i=0}^n \mathrm{E}\left[\exp\left(t(X_i-\mu)R_i\right) \mid X\right]\\
&= \exp\left(\sum_{i=0}^n \Psi_{X_i}(t)\right),
\end{aligned}$$

where we denoted, for any $i \in \{0, \dots, n\}$, $\Psi_{X_i}(t) := \log \mathrm{E}[\exp(t(X_i - \mu_i)R_i) \mid X_i]$. Let us then compute $\Psi_{X_i}(t)$, for any $i \in \{0, \dots, n\}$.

$$\begin{aligned}\mathrm{E}[\exp(t(X_i - \mu)R_i) \mid X_i] &= \int_0^\infty \exp(t(X_i - \mu)x) \exp(-x)dx \\ &= \int_0^\infty \exp(-(1 - t(X_i - \mu))x)dx \\ &= \frac{1}{1 - t(X_i - \mu)}.\end{aligned}$$

Therefore,

$$\Psi_{X_i}(t) = -\log(1 - t(X_i - \mu)).$$

We deduce that

$$\exp\left(\sum_{i=0}^n \Psi_{X_i}(t)\right) = \exp\left(-\sum_{i=0}^n \log(1 - t(X_i - \mu))\right),$$

and thus, that for any $t \in [0, \frac{1}{1-\mu})$,

$$\begin{aligned}P_{L \sim \mathrm{Dir}(1^{n+1})}\left(L^\top X \geq \mu \;\middle|\; X\right) &\leq \exp\left(-\sum_{i=0}^n \log(1 - t(X_i - \mu))\right) \\ &= \exp\left(-\log(1 - t(\xi - \mu)) - n\frac{1}{n}\sum_{i=1}^n \log(1 - t(X_i - \mu))\right) \\ &= \frac{1}{1 - t(\xi - \mu)} \exp\left(-n\frac{1}{n}\sum_{i=1}^n \log(1 - t(X_i - \mu))\right) \\ &= \frac{1}{1 - t(\xi - \mu)} \exp(-n\phi(t)), \qquad (7)\end{aligned}$$

where we defined $\phi(t) := \frac{1}{n}\sum_{i=1}^n \log(1 - t(X_i - \mu))$.

Let $\eta \in (0, 1)$ be arbitrary. For this η, if $\xi \in \{0, 1\}$ and $t \in [0, \frac{1-\eta}{1-\mu}]$ then we have

$$\frac{1}{1 - t(\xi - \mu)} \leq \frac{1}{\eta}.$$

Therefore, since $t \in [0, \frac{1}{1-\mu})$ is arbitrary in (7), we have

$$P_{L \sim \mathrm{Dir}(1^{n+1})}\left(L^\top X \geq \mu \;\middle|\; X\right) \leq \frac{1}{\eta} \exp\left(-n \sup_{t \in [0, \frac{1-\eta}{1-\mu}]} \phi(t)\right). \qquad (8)$$

Here note that $\phi(t)$ is concave in t. Thus, for any $t \in [\frac{1-\eta}{1-\mu}, \frac{1}{1-\mu}]$ we have

$$\begin{aligned}
\phi(t) &\leq \phi\left(\frac{1-\eta}{1-\mu}\right) + \frac{\eta}{1-\mu}\phi'\left(\frac{1-\eta}{1-\mu}\right) \\
&= \phi\left(\frac{1-\eta}{1-\mu}\right) - \frac{\eta}{1-\mu}\mathrm{E}_{X\sim\hat{F}}\left[\frac{X-\mu}{1-\frac{1-\eta}{1-\mu}(X-\mu)}\right] \\
&\leq \phi\left(\frac{1-\eta}{1-\mu}\right) - \frac{\eta}{1-\mu}\frac{0-\mu}{1-\frac{1-\eta}{1-\mu}(0-\mu)} \\
&= \phi\left(\frac{1-\eta}{1-\mu}\right) + \frac{\eta\mu(1-\mu)}{(1-\mu\eta)(1-\mu)} \\
&\leq \phi\left(\frac{1-\eta}{1-\mu}\right) + \frac{\eta\mu}{1-\mu},
\end{aligned}$$

where the second inequality follows since $\frac{x-\mu}{1-t(x-\mu)}$ is increasing in $x \in [0,1]$. This implies that

$$\sup_{t\in[0,\frac{1}{1-\mu}]} \phi(t) \leq \sup_{t\in[0,\frac{1-\eta}{1-\mu}]} \phi(t) + \frac{\eta\mu}{1-\mu}. \tag{9}$$

From Honda and Takemura (2010, Theorem 8), we know that

$$\mathcal{K}_{\inf}(F,\mu) = \sup_{t\in[0,\frac{1}{1-\mu}]} \mathrm{E}_{X\sim F}[\log(1-t(X-\mu))]$$

and

$$\mathcal{K}_{\inf}(\hat{F},\mu) = \sup_{t\in[0,\frac{1}{1-\mu}]} \phi(t). \tag{10}$$

Putting (8)–(10) together, we obtain

$$\begin{aligned}
P_{L\sim\mathrm{Dir}(1^{n+1})}\left(L^\top X \geq \mu \,\middle|\, X\right) &\leq \frac{1}{\eta}\exp\left(-n\left(\sup_{t\in[0,\frac{1}{1-\mu}]}\phi(t) - \frac{\eta\mu}{1-\mu}\right)\right) \\
&\leq \frac{1}{\eta}\exp\left(-n\left(\mathcal{K}_{\inf}(\hat{F},\mu) - \frac{\eta\mu}{1-\mu}\right)\right)
\end{aligned}$$

for any $\eta \in [0,1)$. ■

G.2. Lower Bound for Conditional Random Average

Lemma 17 *Assume that $n \geq 2$ and let $L \sim \mathrm{Dir}(1^{n+1})$ and $X = (X_0, X_1, \ldots, X_n)$ where we know that $X_0 = 1$ is deterministic. Then, we have the lower bound*

$$P\left(X^\top L \geq \frac{1}{n}\sum_{i=1}^n X_i \,\middle|\, X\right) \geq \left(1 - \frac{1}{n}\sum_{i=1}^n X_i\right)\frac{1}{25n^2}.$$

Proof Since $L \sim \mathrm{Dir}(1^{n+1})$, then we know that $(L_0, \ldots, L_n) \sim \left(\frac{R_0}{\sum_{i=0}^n R_i}, \ldots, \frac{R_n}{\sum_{i=0}^n R_i}\right)$, where $R_0, \ldots, R_n$ independently follow the exponential distribution $\mathcal{E}(1)$ with rate parameter 1.

Then, we can compute

$$\begin{aligned} P\left(X^\top L \geq \frac{1}{n}\sum_{i=1}^n X_i \,\middle|\, X\right) &= P\left(\sum_{i=0}^n R_i X_i \geq \frac{1}{n}\sum_{i=1}^n X_i \sum_{i=0}^n R_i \,\middle|\, X\right) \\ &= P\left(\sum_{i=0}^n \left(X_i - \frac{1}{n}\sum_{i=1}^n X_i\right) R_i \geq 0 \,\middle|\, X\right) \\ &= 1 - P\left(\sum_{i=0}^n \left(X_i - \frac{1}{n}\sum_{i=1}^n X_i\right) R_i < 0 \,\middle|\, X\right). \end{aligned}$$

Then, we know by Markov's inequality that, for any random variable Y and for any $t > 0$, we have

$$P(Y < 0) \leq \mathrm{E}\left[e^{-tY}\right].$$

Thus, applying this small result, we have, for any $t > 0$,

$$\begin{aligned} P\left(X^\top L \geq \frac{1}{n}\sum_{i=1}^n X_i \,\middle|\, X\right) &\geq 1 - \mathrm{E}\left[\exp\left(-t\sum_{j=0}^n \left(X_j - \frac{1}{n}\sum_{i=1}^n X_i\right) R_j\right) \,\middle|\, X\right] \\ &= 1 - \prod_{j=0}^n \mathrm{E}\left[\exp\left(-t\left(X_j - \frac{1}{n}\sum_{i=1}^n X_i\right) R_j\right) \,\middle|\, X\right]. \end{aligned}$$

But we know that if $Y \sim \mathcal{E}(1)$, then for any $\lambda < 1$,

$$\begin{aligned} \mathrm{E}\left[e^{\lambda Y}\right] &= \int_0^\infty e^{\lambda y} e^{-y} dy \\ &= \frac{1}{1-\lambda}. \end{aligned}$$

Since, for any $t \in (0,1)$ and for any j, $\left|t\left(X_j - \frac{1}{n}\sum_{i=1}^n X_i\right)\right| < 1$, we can compute for any $t \in (0,1)$,

$$\mathrm{E}\left[\exp\left(-t\left(X_j - \frac{1}{n}\sum_{i=1}^n X_i\right) R_j\right) \,\middle|\, X\right] = \frac{1}{1 + t\left(X_j - \frac{1}{n}\sum_{i=1}^n X_i\right)}.$$

As a consequence, for any $t \in (0,1)$,

$$\begin{aligned} P\left(X^\top L \geq \frac{1}{n}\sum_{i=1}^n X_i \,\middle|\, X\right) &\geq 1 - \prod_{j=0}^n \frac{1}{1 + t\left(X_j - \frac{1}{n}\sum_{i=1}^n X_i\right)} \\ &= 1 - \frac{1}{1 + t\left(1 - \frac{1}{n}\sum_{i=1}^n X_i\right)} \prod_{j=1}^n \frac{1}{1 + t\left(X_j - \frac{1}{n}\sum_{i=1}^n X_i\right)}. \end{aligned}$$

Then, we are going to study carefully the polynomial in t, $\prod_{j=1}^{n}\left(1+t\left(X_j-\frac{1}{n}\sum_{i=1}^{n}X_i\right)\right)$ and provide a nice lower bound to it.

For any $j\in\{1,\dots,n\}$, let $a_j:=\left(X_j-\frac{1}{n}\sum_{i=1}^{n}X_i\right)$ and let $a_0:=\left(1-\frac{1}{n}\sum_{i=1}^{n}X_i\right)$. Recall that $\sum_{j=1}^{n}a_j=0$ and that for any $j\in\{1,\dots,n\}, |a_j|\leq 1$. We would like to prove that for any $|t|\leq\frac{1}{10n(n+1)}$,

$$\prod_{j=1}^{n}(1+ta_j)\geq 1-t^2\sum_{j=1}^{n}a_j^2.$$

First, note that:

$$\begin{aligned}2\sum_{i<j}a_ia_j+\sum_i a_i^2&=\left(\sum_{i=1}^{n}a_i\right)^2\\&=0.\end{aligned}$$

Then, we define the functions: $f(t):=\prod_{j=1}^{n}(1+ta_j)$ and $g(t):=f(t)-1+t^2\sum_i a_i^2$. We are going to prove that, for any $t\in\left[0,\frac{1}{10n(n+1)}\right]$, $g(t)\geq 0$. We notice that $g(0)=0$, so it is enough to prove that g is increasing on $\left[0,\frac{1}{10n(n+1)}\right]$. For any $t\in\left[0,\frac{1}{10n(n+1)}\right]$,

$$\begin{aligned}g(t)&=t^2\sum_{i=1}^{n}a_i^2+t^2\sum_{i<j}a_ia_j+\sum_{k=3}^{n}t^k\sum_{i_1<\dots<i_k}a_{i_1}\dots a_{i_k}\\&=\frac{1}{2}t^2\sum_{i=1}^{n}a_i^2+\sum_{k=3}^{n}t^k\sum_{i_1<\dots<i_k}a_{i_1}\dots a_{i_k}\end{aligned}$$

Let us compute the derivative of g and gather the terms by powers of t.

$$\begin{aligned}g'(t)&=t\sum_{i=1}^{n}a_i^2+\sum_{k=3}^{n}kt^{k-1}\sum_{i_1<\dots<i_k}a_{i_1}\dots a_{i_k}\\&=t\sum_{i=1}^{n}a_i^2+\sum_{k=2}^{n-1}(k+1)t^k\sum_{i_1<\dots<i_{k+1}}a_{i_1}\dots a_{i_{k+1}}.\end{aligned}$$

Now, we bound $\sum_{i_1<i_2<\dots<i_k<i_{k+1}}a_{i_1}\dots a_{i_k}a_{i_{k+1}}$. By symmetry, we can assume that: $|a_1|\leq|a_2|\leq\dots\leq|a_n|\leq 1$. Then, we can easily bound: $|a_{i_1}\dots a_{i_k}a_{i_{k+1}}|\leq|a_ia_j|$ where $i=\inf\{i_1,\dots,i_{k+1}\}$ and $j=\inf\{i_1,\dots,i_{k+1}\}-\{i\}$. As a consequence, if we bound the sum $\sum_{i_1<i_2<\dots<i_k<i_{k+1}}|a_{i_1}\dots a_{i_k}a_{i_{k+1}}|$ term by term, then we bound:

- $\binom{n-2}{k-1}$ terms by $|a_1a_2|$,
- $\binom{n-3}{k-1}$ terms by $|a_1a_3|$,
- $\binom{n-3}{k-1}$ terms by $|a_2a_3|$,

- $\binom{n-4}{k-1}$ terms by $|a_1 a_4|$,
- ...
- $\binom{n-j}{k-1}$ terms by $|a_i a_j|$ for any $i \in \{1, \dots, j-1\}$.

Indeed, there are $\binom{n-j}{k-1}$ subsets of $k+1$ elements of $\{1, \dots, n\}$ whose smallest elements are i and j.

Therefore, we can bound, for any $k \geq 2$,

$$\begin{aligned}
\left| \sum_{i_1 < i_2 < .. < i_k < i_{k+1}} a_{i_1} \dots a_{i_k} a_{i_{k+1}} \right| &\leq \sum_{i<j} \binom{n-j}{k-1} |a_i a_j| \\
&\leq \sum_{i<j} \binom{n-j}{k-1} \frac{a_i^2 + a_j^2}{2} \\
&= \sum_{j=1}^{n} \binom{n-j}{k-1} \sum_{i=1}^{j-1} \frac{a_i^2 + a_j^2}{2} \\
&= \sum_{j=1}^{n} \binom{n-j}{k-1} \left(\frac{j-1}{2} a_j^2 + \frac{1}{2} \sum_{i=1}^{j-1} a_i^2 \right) \\
&= \sum_{j=1}^{n} \binom{n-j}{k-1} \frac{j-1}{2} a_j^2 + \sum_{j=1}^{n} \binom{n-j}{k-1} \frac{1}{2} \sum_{i=1}^{j-1} a_i^2 \\
&=: \sum_{j=1}^{n} \beta_j a_j^2.
\end{aligned}$$

Let us now bound the β_j for $j \in \{1, \dots, n\}$.

$$\begin{aligned}
\beta_j &= \binom{n-j}{k-1} \frac{j-1}{2} + \sum_{l=j+1}^{n} \binom{n-l}{k-1} \frac{1}{2} \\
&= \frac{1}{2} \left((j-1) \binom{n-j}{k-1} + \sum_{l=j+1}^{n} \binom{n-l}{k-1} \right) \\
&\leq \frac{1}{2} \sum_{l=0}^{n} \binom{n-l}{k-1} \\
&= \frac{1}{2} \frac{1}{(k-1)!} \sum_{l=0}^{n} \frac{(n-l)!}{(n-l-k+1)!} \\
&\leq \frac{1}{2} \frac{1}{(k-1)!} (n+1) \frac{n!}{(n-k+1)!}.
\end{aligned}$$

Thus, re-injecting the result in the previous sum, we can bound

$$\left| \sum_{i_1<i_2<\cdots<i_k<i_{k+1}} a_{i_1} \dots a_{i_k} a_{i_{k+1}} \right| \leq \frac{1}{2} \sum_{i=1}^{n} a_i^2 \frac{1}{(k-1)!}(n+1)\frac{n!}{(n-k+1)!}.$$

There, we can eventually study the second sum in the derivative of g. For any $t \in \left[0, \frac{1}{10n(n+1)}\right]$,

$$\begin{aligned}
&\left| \sum_{k=2}^{n-1} (k+1)t^k \sum_{i_1<\cdots<i_{k+1}} a_{i_1} \dots a_{i_{k+1}} \right| \\
&\leq \sum_{k=2}^{n-1} (k+1)t^k \frac{1}{2} \sum_{i=1}^{n} a_i^2 \frac{1}{(k-1)!}(n+1)\frac{n!}{(n-k+1)!} \\
&= \frac{n+1}{2} \left(\sum_{i=1}^{n} a_i^2 \right) \sum_{k=2}^{n-1} \frac{k+1}{(k-1)!} \frac{n!}{(n-k+1)!} t^k \\
&= \frac{n+1}{2} \left(\sum_{i=1}^{n} a_i^2 \right) \sum_{k=2}^{n-1} \frac{k+1}{(k-1)!} \frac{n!}{(n-k+1)!n^k} (nt)^k \\
&= \frac{n+1}{2n} \left(\sum_{i=1}^{n} a_i^2 \right) \sum_{k=2}^{n-1} \frac{k+1}{(k-1)!} \frac{n!}{(n-k+1)!n^{k-1}} (nt)^k \\
&\leq \frac{n+1}{2n} \left(\sum_{i=1}^{n} a_i^2 \right) \sum_{k=2}^{n-1} \frac{k+1}{(k-1)!} (nt)^k \\
&= \frac{(n+1)t}{2} \left(\sum_{i=1}^{n} a_i^2 \right) \sum_{k=2}^{n-1} \frac{k+1}{(k-1)!} (nt)^{k-1} \\
&\leq \frac{(n+1)t}{2} \left(\sum_{i=1}^{n} a_i^2 \right) \sum_{k=2}^{\infty} \frac{k+1}{(k-1)!} (nt)^{k-1}.
\end{aligned}$$

But we know that, for any $x \geq 0$,

$$\begin{aligned}
\sum_{k=2}^{\infty} \frac{k+1}{(k-1)!} x^{k-1} &= \sum_{k=2}^{\infty} \frac{(k+1)k}{k!} x^{k-1} \\
&= \frac{d^2}{dx^2} \left(\sum_{k=2}^{\infty} \frac{1}{k!} x^{k+1} \right) \\
&= \frac{d^2}{dx^2} \{ x(\exp(x) - 1 - x) \} \\
&= \frac{d}{dx} \{ x \exp(x) + \exp(x) - 2x - 1 \} \\
&= x \exp(x) + 2 \exp(x) - 2.
\end{aligned}$$

As a consequence, for $n \geq 1$, we have that

$$\sum_{k=2}^{\infty} \frac{k+1}{(k-1)!} (nt)^{k-1} = e^{nt}(nt+2) - 2.$$

Thus, for $n \geq 1$,

$$\left|\sum_{k=2}^{n-1}(k+1)t^k \sum_{i_1<\cdots<i_{k+1}} a_{i_1}\dots a_{i_{k+1}}\right| \leq \frac{1}{2}t(n+1)\left(e^{nt}(nt+2)-2\right)\left(\sum_{i=1}^{n} a_i^2\right).$$

But since $t \in \left[0, \frac{1}{10n(n+1)}\right]$, then $nt \in (0,1)$ so $e^{nt} \leq 1 + ent$. Therefore,

$$\begin{aligned} e^{nt}(nt+2) &\leq (1+ent)(nt+2) \\ &= 2 + (2e+1)nt + en^2t^2. \end{aligned}$$

Thus,

$$e^{nt}(nt+2) - 2 \leq (2e+1)nt + en^2t^2.$$

Therefore, for any $t \in \left[0, \frac{1}{10n(n+1)}\right]$,

$$\left|\sum_{k=2}^{n-1}(k+1)t^k \sum_{i_1<\cdots<i_{k+1}} a_{i_1}\dots a_{i_{k+1}}\right| \leq \frac{1}{2}t(n+1)\left((2e+1)nt+en^2t^2\right)\left(\sum_{i=1}^{n} a_i^2\right).$$

We can now provide a lower bound to $g'(t)$ for $t \in \left[0, \frac{1}{10n(n+1)}\right]$ by

$$\begin{aligned} g'(t) &= t\sum_{i=1}^{n} a_i^2 + \sum_{k=3}^{n} kt^{k-1} \sum_{i_1<i_2<\cdots<i_k} a_{i_1}\dots a_{i_k} \\ &\geq t\left(\sum_{i=1}^{n} a_i^2\right) - \frac{1}{2}t(n+1)\left((2e+1)nt+en^2t^2\right)\left(\sum_{i=1}^{n} a_i^2\right) \\ &= t\left(\sum_{i=1}^{n} a_i^2\right)\left(1 - \frac{1}{2}(n+1)\left((2e+1)nt+en^2t^2\right)\right) \\ &\geq t\left(\sum_{i=1}^{n} a_i^2\right)\left(1 - \frac{1}{2}\left(\frac{2e+1}{10} + \frac{e}{100(n+1)}\right)\right) \\ &\geq t\left(\sum_{i=1}^{n} a_i^2\right)\left(1 - \frac{1}{2}\left(\frac{2e+1}{10} + \frac{e}{10}\right)\right) \\ &= t\left(\sum_{i=1}^{n} a_i^2\right)\left(1 - \frac{3e+1}{20}\right) \\ &\geq t\left(\sum_{i=1}^{n} a_i^2\right). \end{aligned}$$

We then deduce that g is increasing on $\left[0, \frac{1}{10n(n+1)}\right]$, but recall that

$$g(t) = f(t) - 1 - t^2\sum_i a_i^2,$$

where $f(t) := \prod_{j=1}^{n}(1+ta_j)$. Since $g(0) = 0$ and g is increasing on $\left[0, \frac{1}{10n(n+1)}\right]$, for any $t \in \left[0, \frac{1}{10n(n+1)}\right]$ we have $g(t) \geq 0$. It implies, for any $t \in \left[0, \frac{1}{10n(n+1)}\right]$,

$$f(t) \geq 1 - t^2 \sum_{i=1}^{n} a_i^2 > 0.$$

We deduce that, for any $t \in \left[0, \frac{1}{10n(n+1)}\right]$,

$$1 - \frac{1}{1+ta_0}\prod_{i=1}^{n}\frac{1}{1+ta_i} \geq 1 - \frac{1}{1+ta_0}\frac{1}{1-t^2\sum_{i=1}^{n}a_i^2}.$$

Therefore, for any $t \in \left[0, \frac{1}{10n(n+1)}\right]$,

$$P\left(X^\top L \geq \frac{1}{n}\sum_{i=1}^{n} X_i \middle| X\right) \geq 1 - \frac{1}{1+ta_0}\frac{1}{1-t^2\sum_{i=1}^{n}a_i^2}.$$

Since $a_0 > 0$, we can bound the RHS by using polynomial series as

$$\begin{aligned}\frac{1}{1+ta_0} &= \sum_{n=0}^{\infty}(-ta_0)^n \\ &\leq 1 - ta_0 + t^2a_0^2\end{aligned}$$

and

$$\begin{aligned}\frac{1}{1-t^2\sum_{i=1}^{n}a_i^2} &= \sum_{n=0}^{\infty}\left(t^2\sum_{i=1}^{n}a_i^2\right)^n \\ &\leq 1 + 2t^2\sum_{i=1}^{n}a_i^2.\end{aligned}$$

Therefore, for any $t \in \left[0, \frac{1}{10n(n+1)}\right]$,

$$\begin{aligned}P\left(X^\top L \geq \frac{1}{n}\sum_{i=1}^{n} X_i \middle| X\right) &\geq 1 - (1 - ta_0 + t^2a_0^2)\left(1 + 2t^2\sum_{i=1}^{n}a_i^2\right) \\ &= a_0 t - \left(2\sum_{i=1}^{n}a_i^2 + a_0^2\right)t^2 + 2a_0\sum_{i=1}^{n}a_i^2 t^3 - 2a_0^2\sum_{i=1}^{n}a_i^2 t^4.\end{aligned}$$

In particular, if $t = \frac{1}{20n^2}$, we have

$$\begin{aligned}P\left(X^\top L \geq \frac{1}{n}\sum_{i=1}^{n} X_i \middle| X\right) \geq \frac{a_0}{20}\frac{1}{n^2} - \left(2\sum_{i=1}^{n}a_i^2 + a_0^2\right)\frac{1}{400n^4} \\ + 2a_0\sum_{i=1}^{n}a_i^2\frac{1}{8000n^6} - 2a_0^2\sum_{i=1}^{n}a_i^2\frac{1}{160000n^8},\end{aligned}$$

where the last three terms can be bounded by

$$\begin{aligned}
&\left| -\left(2\sum_{i=1}^{n} a_i^2 + a_0^2\right)\frac{1}{400n^4} + 2a_0\sum_{i=1}^{n} a_i^2\frac{1}{8000n^6} - 2a_0^2\sum_{i=1}^{n} a_i^2\frac{1}{160000n^8}\right| \\
&\leq (2n+1)\frac{1}{400n^4} + 2n\frac{1}{8000n^6} + 2n\frac{1}{160000n^8} \\
&= \frac{1}{100n^3}.
\end{aligned}$$

Thus, since $a_0 > 0$,

$$\begin{aligned}
P\left(X^\top L \geq \frac{1}{n}\sum_{i=1}^{n} X_i \middle| X\right) &\geq a_0\frac{1}{20n^2} - \frac{1}{100n^3} \\
&\geq a_0\frac{1}{25n^2}.
\end{aligned}$$

Recall that $a_0 = \left(1 - \frac{1}{n}\sum_{i=1}^{n} X_i\right)$, we conclude

$$P\left(X^\top L \geq \frac{1}{n}\sum_{i=1}^{n} X_i \middle| X\right) \geq \left(1 - \frac{1}{n}\sum_{i=1}^{n} X_i\right)\frac{1}{25n^2},$$

which is the result we wanted to prove. ■

Appendix H. Domination of the Lévy Distance by the Infinite Distance

In this section we show the following lemma on the relation between the Lévy distance and the L^∞ distance.

Lemma 18 *Let F and G two cumulative distribution functions on $[0,1]$. Then,*

$$D_L(F,G) \leq \|F - G\|_\infty.$$

Proof Recall that $D_L(F,G) = \inf\{\epsilon > 0 :\ \forall x \in [0,1],\ F(x-\epsilon) - \epsilon \leq G(x) \leq F(x+\epsilon) + \epsilon\}$, where in this definition, we naturally extended the definition of F and G on $\mathbb{R}$ by $\forall x \leq 0, F(x) = G(x) = 0$ and $\forall x \geq 1, F(x) = G(x) = 1$. Let us denote $\epsilon := \|F - G\|_\infty$, then, for any $x \in [0,1]$,

$$|F(x) - G(x)| \leq \epsilon.$$

In other words, for any $x \in [0,1]$,

$$F(x) - \epsilon \leq G(x) \leq F(x) + \epsilon,$$

which implies that, for any $x \in [0,1]$, $F(x-\epsilon) - \epsilon \leq G(x) \leq F(x+\epsilon) + \epsilon$ because F is increasing. ■

Appendix I. Proof of Lemma 6

First note that

$$\mathcal{K}_{\inf}^{(M)}(\tilde{F},\mu) = \max_{t\in[0,(1-\mu)^{-1}]} \mathrm{E}[\log(1-t(\tilde{X}-\mu))],$$
$$\mathcal{K}_{\inf}(F,\mu) = \max_{t\in[0,(1-\mu)^{-1}]} \mathrm{E}[\log(1-t(X-\mu))].$$

Since

$$\frac{\mathrm{d}}{\mathrm{d}\mu}\mathcal{K}_{\inf}^{(M)}(\tilde{F},\mu) \le \frac{1}{1-\mu},$$

by Honda and Takemura (2010, Theorem 6), the first inequality of (6) is obtained by

$$\begin{aligned}
\mathcal{K}_{\inf}^{(M)}(\tilde{F},\mu) &\ge \mathcal{K}_{\inf}^{(M)}(\tilde{F},\mu+1/M) - \frac{1}{M(1-\mu-1/M)} \\
&= \max_{t\in[0,(1-\mu)^{-1}]} \mathrm{E}[\log(1-t(\tilde{X}-\mu-1/M))] - \frac{1}{M(1-\mu-1/M)} \\
&\ge \max_{t\in[0,(1-\mu)^{-1}]} \mathrm{E}[\log(1-t(X-\mu))] - \frac{1}{M(1-\mu-1/M)} \\
&= \mathcal{K}_{\inf}(F,\mu) - \frac{1}{M(1-\mu-1/M)}.
\end{aligned}$$

The second inequality of (6) is derived from

$$\begin{aligned}
\mathcal{K}_{\inf}^{(M)}(\tilde{F},\mu) &= \max_{t\in[0,(1-\mu)^{-1}]} \mathrm{E}[\log(1-t(\tilde{X}-\mu))] \\
&= \max_{t\in[0,(1-\mu)^{-1}]} \mathrm{E}[\mathrm{E}[\log(1-t(\tilde{X}-\mu))|X]] \\
&\le \max_{t\in[0,(1-\mu)^{-1}]} \mathrm{E}[\log(1-t(\mathrm{E}[\tilde{X}|X]-\mu))] \\
&= \max_{t\in[0,(1-\mu)^{-1}]} \mathrm{E}[\log(1-t(X-\mu))] \\
&= \mathcal{K}_{\inf}(F,\mu),
\end{aligned}$$

where the inequality follows from Jensen's inequality. ■