Fixed-confidence guarantees for Bayesian best-arm identification

Xuedong Shang, Rianne de Heide, Emilie Kaufmann, Pierre Ménard, Michal
Valko

Fixed-Confidence Guarantees for Bayesian Best-Arm Identification

Xuedong Shang[1,2]   **Rianne de Heide**[3,4]   **Emilie Kaufmann**[1,2,5]   **Pierre Ménard**[1]   **Michal Valko**[1,6]

[1]Inria Lille Nord Europe [2]Université de Lille [3]Leiden University [4]CWI [5]CNRS [6]DeepMind Paris

Abstract

We investigate and provide new insights on the sampling rule called Top-Two Thompson Sampling (`TTTS`). In particular, we justify its use for *fixed-confidence best-arm identification*. We further propose a variant of `TTTS` called Top-Two Transportation Cost (`T3C`), which disposes of the computational burden of `TTTS`. As our main contribution, we provide the first sample complexity analysis of `TTTS` and `T3C` when coupled with a very natural Bayesian stopping rule, for bandits with Gaussian rewards, solving one of the open questions raised by Russo (2016). We also provide new posterior convergence results for `TTTS` under two models that are commonly used in practice: bandits with Gaussian and Bernoulli rewards and conjugate priors.

1 Introduction

In multi-armed bandits, a learner repeatedly chooses an *arm* to play, and receives a reward from the associated unknown probability distribution. When the task is *best-arm* identification (BAI), the learner is not only asked to sample an arm at each stage, but is also asked to output a recommendation (i.e., a guess for the arm with the largest mean reward) after a certain period. Unlike in another well-studied bandit setting, the learner is not interested in maximizing the sum of rewards gathered during the exploration (or minimizing *regret*), but only cares about the quality of her recommendation. As such, BAI is a particular *pure exploration* setting (Bubeck et al., 2009).

Formally, we consider a finite-arm bandit model, which is a collection of K probability distributions, called arms $\mathcal{A} \triangleq \{1, \ldots, K\}$, parametrized by their means $\mu_1, \ldots, \mu_K$. We assume the (unknown) best arm is unique and we denote it by $I^\star \triangleq \arg\max_i \mu_i$. A best-arm identification strategy (I_n, J_n, τ) consists of three components. The first is a *sampling rule*, which selects an arm I_n at round n. At each round n, a vector of re-

wards $\mathbf{Y}_n = (Y_{n,1}, \cdots, Y_{n,K})$ is generated for all arms independently from past observations, but only Y_{n,I_n} is revealed to the learner. Let $\mathcal{F}_n$ be the σ-algebra generated by $(U_1, I_1, Y_{1,I_1}, \cdots, U_n, I_n, Y_{n,I_n})$, then I_n is $\mathcal{F}_{n-1}$-measurable, i.e., it can only depend on the past $n-1$ observations, and some exogeneous randomness, materialized into $U_{n-1} \sim \mathcal{U}([0,1])$. The second component is a $\mathcal{F}_n$-measurable *recommendation rule* J_n, which returns a guess for the best arm, and thirdly, the *stopping rule* τ, a stopping time with respect to $(\mathcal{F}_n)_{n \in \mathbb{N}}$, decides when the exploration is over.

BAI has already been studied within several theoretical frameworks. In this paper we consider the fixed-confidence setting, introduced by Even-dar et al. (2003), in which given a risk parameter δ, the goal is to ensure that probability to stop and recommend a wrong arm, $\mathbb{P}[J_\tau \neq I^\star]$, is smaller than δ, while minimizing the expected total number of samples to make this accurate recommendation, $\mathbb{E}[\tau]$. The most studied alternative is the fixed-budget setting for which the stopping rule τ is fixed to some (known) maximal budget n, and the goal is to minimize the error probability $\mathbb{P}[J_n \neq I^\star]$ (Audibert and Bubeck, 2010). Note that these two frameworks are very different in general and do not share transferable regret bounds (see Carpentier and Locatelli 2016 for an additional discussion).

Most of the existing sampling rules for the fixed-confidence setting depend on the risk parameter δ. Some of them rely on confidence intervals such as `LUCB` (Kalyanakrishnan et al., 2012), `UGapE` (Gabillon et al., 2012), or `lil'UCB` (Jamieson et al., 2014); others are based on eliminations such as `SuccessiveElimination` (Even-dar et al., 2003) and `ExponentialGapElimination` (Karnin et al., 2013). The first known sampling rule for BAI that does not depend on δ is the *tracking* rule proposed by Garivier and Kaufmann (2016), which is proved to achieve the minimal sample complexity when combined with the Chernoff stopping rule when δ goes to zero. Such an *anytime* sampling rule (neither depending on a risk δ or a budget n) is very appealing for applications, as advocated by Jun and Nowak (2016), who introduce the anytime best-arm identification framework. In this

paper, we investigate another anytime sampling rule for BAI: T̲op-T̲wo T̲hompson S̲ampling (TTTS), and propose a second anytime sampling rule: T̲op-T̲wo T̲ransportation C̲ost (T3C).

Thompson Sampling (Thompson, 1933) is a Bayesian algorithm well known for regret minimization, for which it is now seen as a major competitor to UCB-typed approaches (Burnetas and Katehakis, 1996; Auer et al., 2002; Cappé et al., 2013). However, it is also well known that regret minimizing algorithms cannot yield optimal performance for BAI (Bubeck et al., 2011; Kaufmann and Garivier, 2017) and as we opt Thompson Sampling for BAI, then its adaptation is necessary. Such an adaptation, TTTS, was given by Russo (2016) along with the other top-two sampling rules TTPS and TTVS. By choosing between two different candidate arms in each round, these sampling rules enforce the exploration of sub-optimal arms, that would be under-sampled by vanilla Thompson sampling due to its objective of maximizing rewards.

While TTTS appears to be a good anytime sampling rule for the fixed-confidence BAI when coupled with an appropriate stopping rule, so far there is no theoretical support for this employment. Indeed, the (Bayesian-flavored) asymptotic analysis of Russo (2016) shows that under TTTS, the posterior probability that $I^\star$ is the best arm converges almost surely to 1 at the best possible rate. However, this property does not by itself translate into sample complexity guarantees. Since the result of Russo (2016), Qin et al. (2017) proposed and analyzed TTEI, another Bayesian sampling rule, both in the fixed-confidence setting and in terms of posterior convergence rate. Nonetheless, similar guarantees for TTTS have been left as an open question by Russo (2016). In the present paper, we answer this open question. In addition, we propose T3C, a computationally more favorable variant of TTTS and extend the fixed-confidence guarantees to T3C as well.

Contributions (1) We propose a new Bayesian sampling rule, T3C, which is inspired by TTTS but easier to implement and computationally advantageous (2) We investigate two Bayesian stopping and recommendation rules and establish their δ-correctness for a bandit model with Gaussian rewards.[1] (3) We provide the first sample complexity analysis of TTTS and T3C for a Gaussian model and our proposed stopping rule. (4) Russo's posterior convergence results for TTTS were obtained under restrictive assumptions on the models and priors, which exclude the two mostly used in practice: Gaussian bandits with Gaussian priors and bandits with Bernoulli rewards[2] with Beta priors. We

---

[1]hereafter 'Gaussian bandits' or 'Gaussian model'
[2]hereafter 'Bernoulli bandits'

prove that optimal posterior convergence rates can be obtained for those two as well.

Outline In Section 2, we give a reminder of TTTS and introduce T3C along with our proposed recommendation and stopping rules. Then, in Section 3, we describe in detail two important notions of optimality that are invoked in this paper. The main fixed-confidence analysis follows in Section 4, and further Bayesian optimality results are given in Section 5. Numerical illustrations are given in Section 6.

2 Bayesian BAI Strategies

In this section, we give an overview of the sampling rule TTTS and introduce T3C. We provide details for Bayesian updating for Gaussian and Bernoulli models respectively, and introduce associated Bayesian stopping and recommendation rules.

2.1 Sampling rules

Both TTTS and T3C employ a Bayesian machinery and make use of a prior distribution Π_1 over a set of parameters Θ, that contains the unknown true parameter vector $\boldsymbol{\mu}$. Upon acquiring observations $(Y_{1,I_1}, \cdots, Y_{n-1,I_{n-1}})$, we update our beliefs according to Bayes' rule and obtain a posterior distribution Π_n which we assume to have density π_n w.r.t. the Lebesgue measure. Russo's analysis is restricted to strong regularity properties on the models and priors that exclude two important useful cases we consider in this paper: (1) the observations of each arm i follow a Gaussian distribution $\mathcal{N}(\mu_i, \sigma^2)$ with common known variance σ^2, with imposed Gaussian prior $\mathcal{N}(\mu_{1,i}, \sigma^2_{1,i})$, (2) all arms receive Bernoulli rewards with unknown means, with a uniform prior on each arm.

Gaussian model For Gaussian bandits with a $\mathcal{N}(0, \kappa^2)$ prior on each mean, the posterior distribution of μ_i at round n is Gaussian with mean and variance that are respectively given by

$$\frac{\sum_{\ell=1}^{n-1} \mathbb{1}\{I_\ell = i\}Y_{\ell,I_\ell}}{T_{n,i} + \sigma^2/\kappa^2} \quad \text{and} \quad \frac{\sigma^2}{T_{n,i} + \sigma^2/\kappa^2},$$

where $T_{n,i} \triangleq \sum_{\ell=1}^{n-1} \mathbb{1}\{I_\ell = i\}$ is the number of selections of arm i before round n. For the sake of simplicity, we consider improper Gaussian priors with $\mu_{1,i} = 0$ and $\sigma_{1,i} = +\infty$ for all $i \in \mathcal{A}$, for which

$$\mu_{n,i} = \frac{1}{T_{n,i}} \sum_{\ell=1}^{n-1} \mathbb{1}\{I_\ell = i\}Y_{\ell,I_\ell} \quad \text{and} \quad \sigma^2_{n,i} = \frac{\sigma^2}{T_{n,i}}.$$

Observe that in that case the posterior mean $\mu_{n,i}$ coincides with the empirical mean.

Beta-Bernoulli model For Bernoulli bandits with a uniform ($\mathcal{B}eta(1,1)$) prior on each mean, the posterior distribution of μ_i at round n is a Beta distribution with shape parameters $\alpha_{n,i} = \sum_{\ell=1}^{n-1} \mathbb{1}\{I_\ell = i\}Y_{\ell,I_\ell} + 1$ and $\beta_{n,i} = T_{n,i} - \sum_{\ell=1}^{n-1} \mathbb{1}\{I_\ell = i\}Y_{\ell,I_\ell} + 1$.

Now we briefly recall `TTTS` and introduce `T3C`.

Description of `TTTS` At each time step n, `TTTS` has two potential actions: (1) with probability β, a parameter vector $\boldsymbol{\theta}$ is sampled from Π_n, and `TTTS` chooses to play $I_n^{(1)} \triangleq \arg\max_{i \in \mathcal{A}} \theta_i$, (2) and with probability $1 - \beta$, the algorithm continues sampling new $\boldsymbol{\theta}'$ until we obtain a *challenger* $I_n^{(2)} \triangleq \arg\max_{i \in \mathcal{A}} \theta_i'$ that is different from $I_n^{(1)}$, and `TTTS` then selects the challenger.

Description of `T3C` One drawback of `TTTS` is that, in practice, when the posteriors become concentrated, it takes many Thompson samples before the challenger $I_n^{(2)}$ is obtained. We thus propose a variant of `TTTS`, called `T3C`, which alleviates this computational burden. Instead of re-sampling from the posterior until a different candidate appears, we define the challenger as the arm that has the lowest *transportation cost* $W_n(I_n^{(1)}, i)$ with respect to the first candidate (with ties broken uniformly at random).

Let $\mu_{n,i}$ be the empirical mean of arm i and $\mu_{n,i,j} \triangleq (T_{n,i}\mu_{n,i} + T_{n,j}\mu_{n,j})/(T_{n,i} + T_{n,j})$, then we define

$$W_n(i,j) \triangleq \begin{cases} 0 & \text{if } \mu_{n,j} \geq \mu_{n,i}, \\ W_{n,i,j} + W_{n,j,i} & \text{otherwise,} \end{cases} \quad (1)$$

where $W_{n,i,j} \triangleq T_{n,i}d(\mu_{n,i}, \mu_{n,i,j})$ for any i,j and $d(\mu; \mu')$ denotes the Kullback-Leibler between the distribution with mean μ and that of mean μ'. In the Gaussian case, $d(\mu; \mu') = (\mu - \mu')^2/(2\sigma^2)$ while in the Bernoulli case $d(\mu; \mu') = \mu\ln(\mu/\mu') + (1 - \mu)\ln(1 - \mu)/(1 - \mu')$. In particular, for Gaussian bandits

$$W_n(i,j) = \frac{(\mu_{n,i} - \mu_{n,j})^2}{2\sigma^2(1/T_{n,i} + 1/T_{n,j})}\mathbb{1}\{\mu_{n,j} < \mu_{n,i}\}.$$

The pseudo-code of `TTTS` and `T3C` are shown in Algorithm 1. Note that under the Gaussian model with improper priors, one should pull each arm once at the beginning for the sake of obtaining proper posteriors.

W_n in Line 12 of Algorithm 1 is the transportation cost defined in (1).

2.2 Rationale for `T3C`

In order to explain how `T3C` can be seen as an approximation of the re-sampling performed by `TTTS`, we first need to define the *optimal action probabilities*.

Algorithm 1 Sampling rule (`TTTS`/`T3C`)

```
1: Input: β
2: for n ← 1, 2, · · · do
3:     sample θ ∼ Π_n
4:     I^(1) ← arg max_{i∈A} θ_i
5:     sample b ∼ Bern(β)
6:     if b = 1 then
7:         evaluate arm I^(1)
8:     else
9:         repeat sample θ' ∼ Π_n
10:        I^(2) ← arg max_{i∈A} θ'_i          ⎫  TTTS
11:        until I^(2) ≠ I^(1)                 ⎭
12:        I^(2) ← arg min_{i≠I^(1)} W_n(I^(1), i), cf. (1)    T3C
13:        evaluate arm I^(2)
14:     end if
15:     update mean and variance
16:     t = t + 1
17: end for
```

Optimal action probability The optimal action probability $a_{n,i}$ is defined as the posterior probability that arm i is optimal. Formally, letting Θ_i be the subset of Θ such that arm i is the optimal arm,

$$\Theta_i \triangleq \left\{ \boldsymbol{\theta} \in \Theta \,\middle|\, \theta_i > \max_{j\neq i} \theta_j \right\},$$

then we define

$$a_{n,i} \triangleq \Pi_n(\Theta_i) = \int_{\Theta_i} \pi_n(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}.$$

With this notation, one can show that under `TTTS`,

$$\Pi_n\left(I_n^{(2)} = j | I_n^{(1)} = i\right) = \frac{a_{n,j}}{\sum_{k\neq i} a_{n,k}}. \quad (2)$$

Furthermore, when i coincides with the empirical best mean (and this will often be the case for $I_n^{(1)}$ when n is large due to posterior convergence) one can write

$$a_{n,j} \simeq \Pi_n(\theta_j \geq \theta_i) \simeq \exp\left(-W_n(i,j)\right),$$

where the last step is justified in Lemma 2 in the Gaussian case (and Lemma 26 in Appendix H.3 in the Bernoulli case). Hence, `T3C` replaces sampling from the distribution (2) by an approximation of its mode which is *easy to compute*. Note that directly computing the mode would require to compute $a_{n,j}$, which is much more costly than the computation of $W_n(i,j)$[3].

2.3 Stopping and recommendation rules

In order to use `TTTS` or `T3C` as sampling rule for fixed-confidence BAI, we need to additionally define stopping and recommendation rules. While Qin et al.

---

[3]the `TTPS` sampling rule (Russo, 2016) also requires the computation of $a_{n,i}$, thus we do not report simulations for this Bayesian sampling rule in Section 6

(2017) suggest to couple `TTEI` with the "frequentist" Chernoff stopping rule (Garivier and Kaufmann, 2016), we propose in this section natural Bayesian stopping and recommendation rule. They both rely on the optimal action probabilities defined above.

Bayesian recommendation rule   At time step n, a natural candidate for the best arm is the arm with largest optimal action probability, hence we define

$$J_n \triangleq \arg\max_{i \in \mathcal{A}} a_{n,i}.$$

Bayesian stopping rule   In view of the recommendation rule, it is natural to stop when the posterior probability that the recommended action is optimal is large, and exceeds some threshold $c_{n,\delta}$ which gets close to 1. Hence our Bayesian stopping rule is

$$\tau_\delta \triangleq \inf\left\{ n \in \mathbb{N} : \max_{i \in \mathcal{A}} a_{n,i} \geq c_{n,\delta} \right\}. \tag{3}$$

Links with frequentist counterparts   Using the transportation cost $W_n(i,j)$ defined in (1), the Chernoff stopping rule of Garivier and Kaufmann (2016) can actually be rewritten as

$$\tau_\delta^{\text{Ch.}} \triangleq \inf\left\{ n \in \mathbb{N} : \max_{i \in \mathcal{A}} \min_{j \in \mathcal{A} \setminus \{i\}} W_n(i,j) > d_{n,\delta} \right\}. \tag{4}$$

This stopping rule coupled with the recommendation rule $J_n = \arg\max_i \mu_{n,i}$.

As explained in that paper, $W_n(i,j)$ can be interpreted as a (log) Generalized Likelihood Ratio statistic for rejecting the hypothesis $\mathcal{H}_0 : (\mu_i < \mu_j)$. Through our Bayesian lens, we rather have in mind the approximation $\Pi_n(\theta_j > \theta_i) \simeq \exp\{-W_n(i,j)\}$, valid when $\mu_{n,i} > \mu_{n,j}$, which permits to analyze the two stopping rules using similar tools, as will be seen in the proof of Theorem 2.

As shown later in Section 4, τ_δ and $\tau_\delta^{\text{Ch.}}$ prove to be fairly similar for some corresponding choices of the thresholds $c_{n,\delta}$ and $d_{n,\delta}$. This endorses the use of the Chernoff stopping rule in practice, which does not require the (heavy) computation of optimal action probabilities. Still, our sample complexity analysis applies to the two stopping rules, and we believe that a frequentist sample complexity analysis of a fully Bayesian BAI strategy is a nice theoretical contribution.

Useful notation   We follow the notation of Russo (2016) and define the following measures of effort allocated to arm i up to time n,

$$\psi_{n,i} \triangleq \mathbb{P}\left[I_n = i | \mathcal{F}_{n-1}\right] \quad \text{and} \quad \Psi_{n,i} \triangleq \sum_{l=1}^{n} \psi_{l,i}.$$

In particular, for `TTTS` we have

$$\psi_{n,i} = \beta a_{n,i} + (1-\beta) a_{n,i} \sum_{j \neq i} \frac{a_{n,j}}{1 - a_{n,j}},$$

while for `T3C`

$$\psi_{n,i} = \beta a_{n,i} + (1-\beta) \sum_{j \neq i} a_{n,j} \frac{\mathbb{1}\{W_n(j,i) = \min_{k \neq j} W_n(j,k)\}}{\# \left| \arg\min_{k \neq j} W_n(j,k) \right|}.$$

3   Two Related Optimality Notions

In the fixed-confidence setting, we aim for building δ-correct strategies, i.e. strategies that identify the best arm with high confidence on any problem instance.

Definition 1. *A strategy (I_n, J_n, τ) is δ-correct if for all bandit models $\boldsymbol{\mu}$ with a unique optimal arm, it holds that $\mathbb{P}_{\boldsymbol{\mu}}\left[J_\tau \neq I^\star\right] \leq \delta$.*

Among δ-correct strategies, seek the one with the smallest sample complexity $\mathbb{E}\left[\tau_\delta\right]$. So far, `TTTS` has not been analyzed in terms of sample complexity; Russo (2016) focusses on posterior consistency and optimal convergence rates. Interestingly, both the smallest possible sample complexity and the fastest rate of posterior convergence can be expressed in terms of the following quantities.

Definition 2. *Let $\Sigma_K = \{\boldsymbol{\omega} : \sum_{k=1}^{K} \omega_k = 1\}$ and define for all $i \neq I^\star$*

$$C_i(\omega, \omega') \triangleq \min_{x \in \mathcal{I}} \omega d(\mu_{I^\star}; x) + \omega' d(\mu_i; x),$$

where $d(\mu, \mu')$ is the KL-divergence defined above and $\mathcal{I} = \mathbb{R}$ in the Gaussian case and $\mathcal{I} = [0,1]$ in the Bernoulli case. We define

$$\begin{aligned}
\Gamma^\star &\triangleq \max_{\boldsymbol{\omega} \in \Sigma_K} \min_{i \neq I^\star} C_i(\omega_{I^\star}, \omega_i), \\
\Gamma_\beta^\star &\triangleq \max_{\substack{\boldsymbol{\omega} \in \Sigma_K \\ \omega_{I^\star} = \beta}} \min_{i \neq I^\star} C_i(\omega_{I^\star}, \omega_i). \tag{5}
\end{aligned}$$

The quantity $C_i(\omega_{I^\star}, \omega_i)$ can be interpreted as a "transportation cost"[4] from the original bandit instance $\boldsymbol{\mu}$ to an alternative instance in which the mean of arm i is larger than that of $I^\star$, when the proportion of samples allocated to each arm is given by the vector $\boldsymbol{\omega} \in \Sigma_K$. As shown by Russo (2016), the $\boldsymbol{\omega}$ that maximizes (5) is unique, which allows us to define the β-optimal allocation $\boldsymbol{\omega}^\beta$ in the following proposition.

Proposition 1. *There is a unique solution $\boldsymbol{\omega}^\beta$ to the optimization problem (5) satisfying $\omega_{I^\star}^\beta = \beta$, and for all $i, j \neq I^\star$, $C_i(\beta, \omega_i^\beta) = C_j(\beta, \omega_j^\beta)$.*

---

[4]for which $W_n(I^\star, i)$ is an empirical counterpart

For models with more than two arms, there is no closed form expression for $\Gamma_\beta^\star$ or $\Gamma^\star$, even for Gaussian bandits with variance σ^2 for which we have

$$\Gamma_\beta^\star = \max_{\boldsymbol{\omega}:\omega_{I^\star}=\beta} \min_{i \neq I^\star} \frac{(\mu_{I^\star} - \mu_i)^2}{2\sigma^2(1/\omega_i + 1/\beta)}.$$

Bayesian β-optimality Russo (2016) proves that any sampling rule allocating a fraction β to the optimal arm ($\Psi_{n,I^\star}/n \to \beta$) satisfies $1 - a_{n,I^\star} \geq e^{-n(\Gamma_\beta^\star + o(1))}$ (a.s.) for large values of n. We define a *Bayesian β-optimal* sampling rule as a sampling rule matching this lower bound, i.e. satisfying $\Psi_{n,I^\star}/n \to \beta$ and $1 - a_{n,I^\star} \leq e^{-n(\Gamma_\beta^\star + o(1))}$.

Russo (2016) proves that TTTS with parameter β is Bayesian β-optimal. However, the result is valid only under strong regularity assumptions, excluding the two practically important cases of Gaussian and Bernoulli bandits. In this paper, we complete the picture by establishing Bayesian β-optimality for those models in Section 5. For the Gaussian bandit, Bayesian β-optimality was established for TTEI by Qin et al. (2017) with Gaussian priors, but this remained an open problem for TTTS.

A fundamental ingredient of these proofs is to establish the convergence of the allocation of measurement effort to the β-optimal allocation: $\Psi_{n,i}/n \to \omega_i^\beta$ for all i, which is equivalent to $T_{n,i}/n \to \omega_i^\beta$ (cf. Lemma 4).

β-optimality in the fixed-confidence setting In the fixed confidence setting, the performance of an algorithm is evaluated in terms of sample complexity. A lower bound given by Garivier and Kaufmann (2016) states that any δ-correct strategy satisfies $\mathbb{E}[\tau_\delta] \geq (\Gamma^\star)^{-1} \ln(1/(3\delta))$.

Observe that $\Gamma^\star = \max_{\beta \in [0,1]} \Gamma_\beta^\star$. Using the same lower bound techniques, one can also prove that under any δ-correct strategy satisfying $T_{n,I^\star}/n \to \beta$,

$$\liminf_{\delta \to 0} \frac{\mathbb{E}[\tau_\delta]}{\ln(1/\delta)} \geq \frac{1}{\Gamma_\beta^\star}.$$

This motivates the relaxed optimality notion that we introduce in this paper: A BAI strategy is called *asymptotically β-optimal* if it satisfies

$$\frac{T_{n,I^\star}}{n} \to \beta \quad \text{and} \quad \limsup_{\delta \to 0} \frac{\mathbb{E}[\tau_\delta]}{\ln(1/\delta)} \leq \frac{1}{\Gamma_\beta^\star}.$$

In the paper, we provide the first sample complexity analysis of a BAI algorithm based on TTTS (with the stopping and recommendation rules described in Section 2), establishing its asymptotic β-optimality.

As already observed by Qin et al. (2017), any sampling rule converging to the β-optimal allocation (i.e. satisfying $T_{n,i}/n \to w_i^\beta$ for all i) can be shown to satisfy

$\limsup_{\delta \to 0} \tau_\delta / \ln(1/\delta) \leq (\Gamma_\beta^\star)^{-1}$ almost surely, when coupled with the Chernoff stopping rule. The fixed confidence optimality that we define above is stronger as it provides guarantees on $\mathbb{E}[\tau_\delta]$.

4 Fixed-Confidence Analysis

In this section, we consider Gaussian bandits and the Bayesian rules using an improper prior on the means. We state our main result below, showing that TTTS and T3C are asymptotically β-optimal in the fixed confidence setting, when coupled with appropriate stopping and recommendation rules.

Theorem 1. *With $\mathcal{C}^{g_G}$ the function defined by Kaufmann and Koolen (2018), which satisfies $\mathcal{C}^{g_G}(x) \simeq x + \ln(x)$, we introduce the threshold*

$$d_{n,\delta} = 4\ln(4 + \ln(n)) + 2\mathcal{C}^{g_G}\left(\frac{\ln((K-1)/\delta)}{2}\right). \quad (6)$$

The TTTS and T3C sampling rules coupled with either

- *the Bayesian stopping rule (3) with threshold*

$$c_{n,\delta} = 1 - \frac{1}{\sqrt{2\pi}} e^{-\left(\sqrt{d_{n,\delta}} + \frac{1}{\sqrt{2}}\right)^2}$$

  and the recommendation rule $J_t = \arg\max_i a_{n,i}$
- *or the Chernoff stopping rule (4) with threshold $d_{n,\delta}$ and recommendation rule $J_t = \arg\max_i \mu_{n,i}$,*

form a δ-correct BAI strategy. Moreover, if all the arms means are distinct, it satisfies

$$\limsup_{\delta \to 0} \frac{\mathbb{E}[\tau_\delta]}{\log(1/\delta)} \leq \frac{1}{\Gamma_\beta^\star}.$$

We now give the proof of Theorem 1, which is divided into three parts. The **first step** of the analysis is to prove the δ-correctness of the studied BAI strategies.

Theorem 2. *Regardless of the sampling rule, the stopping rule (3) with the threshold $c_{n,\delta}$ and the Chernoff stopping rule with threshold $d_{n,\delta}$ defined in Theorem 1 satisfy $\mathbb{P}[\tau_\delta < \infty \wedge J_{\tau_\delta} \neq I^\star] \leq \delta$.*

To prove that TTTS and T3C allow to reach a β-optimal sample complexity, one needs to quantify how fast the measurement effort for each arm is concentrating to its corresponding optimal weight. For this purpose, we introduce the random variable

$$T_\beta^\varepsilon \triangleq \inf\left\{ N \in \mathbb{N} : \max_{i \in \mathcal{A}} |T_{n,i}/n - \omega_i^\beta| \leq \varepsilon, \forall n \geq N \right\}.$$

The **second step** of our analysis is a sufficient condition for β-optimality, stated in Lemma 1. Its proof is given in Appendix E. The same result was proven for the Chernoff stopping rule by Qin et al. (2017).

Lemma 1. *Let $\delta, \beta \in (0,1)$. For any sampling rule which satisfies $\mathbb{E}\left[T_\beta^\varepsilon\right] < \infty$ for all $\varepsilon > 0$, we have*

$$\limsup_{\delta \to 0} \frac{\mathbb{E}[\tau_\delta]}{\log(1/\delta)} \leq \frac{1}{\Gamma_\beta^\star},$$

if the sampling rule is coupled with stopping rule (3),

Finally, it remains to show that TTTS and T3C meet the sufficient condition, and therefore the **last step**, which is the core component and the most technical part our analysis, consists of showing the following.

Theorem 3. *Under TTTS or T3C, $\mathbb{E}\left[T_\beta^\varepsilon\right] < +\infty$.*

In the rest of this section, we prove Theorem 2 and sketch the proof of Theorem 3. But we first highlight some important ingredients for these proofs.

4.1 Core ingredients

Our analysis hinges on properties of the Gaussian posteriors, in particular on the following tails bounds, which follow from Lemma 1 of Qin et al. (2017).

Lemma 2. *For any $i, j \in \mathcal{A}$, if $\mu_{n,i} \leq \mu_{n,j}$*

$$\Pi_n\left[\theta_i \geq \theta_j\right] \leq \frac{1}{2} \exp\left\{-\frac{(\mu_{n,j} - \mu_{n,i})^2}{2\sigma_{n,i,j}^2}\right\}, \qquad (7)$$

$$\Pi_n\left[\theta_i \geq \theta_j\right] \geq \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(\mu_{n,j} - \mu_{n,i} + \sigma_{n,i,j})^2}{2\sigma_{n,i,j}^2}\right\}, \quad (8)$$

where $\sigma_{n,i,j}^2 \triangleq \sigma^2/T_{n,i} + \sigma^2/T_{n,j}$.

This lemma is crucial to control $a_{n,i}$ and $\psi_{n,i}$, the optimal action and selection probabilities.

4.2 Proof of Theorem 2

We upper bound the desired probability as follows

$$\mathbb{P}\left[\tau_\delta < \infty \wedge J_{\tau_\delta} \neq I^\star\right] \leq \sum_{i \neq I^\star} \mathbb{P}\left[\exists n \in \mathbb{N} : \alpha_{i,n} > c_{n,\delta}\right]$$

$$\leq \sum_{i \neq I^\star} \mathbb{P}\left[\exists n \in \mathbb{N} : \Pi_n(\theta_i \geq \theta_{I_\star}) > c_{n,\delta}, \mu_{n,I^\star} \leq \mu_{n,i}\right]$$

$$\leq \sum_{i \neq I^\star} \mathbb{P}\left[\exists n \in \mathbb{N} : 1 - c_{n,\delta} > \Pi_n(\theta_{I^\star} > \theta_i), \mu_{n,I^\star} \leq \mu_{n,i}\right].$$

The second step uses the fact that as $c_{n,\delta} \geq 1/2$, a necessary condition for $\Pi_n(\theta_i \geq \theta_{I_\star}) \geq c_{n,\delta}$ is that $\mu_{n,i} \geq \mu_{n,I_\star}$. Now using the lower bound (8), if $\mu_{n,I^\star} \leq \mu_{n,i}$, the inequality $1 - c_{n,\delta} > \Pi_n(\theta_{I^\star} > \theta_i)$ implies

$$\frac{(\mu_{n,i} - \mu_{n,I^\star})^2}{2\sigma_{n,i,I^\star}^2} \geq \left(\sqrt{\ln\frac{1}{\sqrt{2\pi}(1 - c_{n,\delta})}} - \frac{1}{\sqrt{2}}\right)^2 = d_{n,\delta},$$

where the equality follows from the expression of $c_{n,\delta}$ as function of $d_{n,\delta}$. Hence to conclude the proof it

remains to check that

$$\mathbb{P}\left[\exists n \in \mathbb{N} : \mu_{n,i} \geq \mu_{n,I^\star}, \frac{(\mu_{n,i} - \mu_{n,I^\star})^2}{2\sigma_{n,i,I^\star}^2} \geq d_{n,\delta}\right] \leq \frac{\delta}{K-1}. \quad (9)$$

To prove this, we observe that for $\mu_{n,i} \geq \mu_{n,I^\star}$,

$$\frac{(\mu_{n,i} - \mu_{n,I^\star})^2}{2\sigma_{n,i,I^\star}^2} = \inf_{\theta_i < \theta_{I^\star}} T_{n,i}d(\mu_{n,i}; \theta_i) + T_{n,I^\star}d(\mu_{n,I^\star}; \theta_{I^\star})$$

$$\leq T_{n,i}d(\mu_{n,i}; \mu_i) + T_{n,I^\star}d(\mu_{n,I^\star}; \mu_{I^\star}).$$

Corollary 10 of Kaufmann and Koolen (2018) then allows us to upper bound the probability

$$\mathbb{P}\left[\exists n \in \mathbb{N} : T_{n,i}d(\mu_{n,i}; \mu_i) + T_{n,I^\star}d(\mu_{n,I^\star}, \mu_{I^\star}) \geq d_{n,\delta}\right]$$

by $\delta/(K-1)$ for the choice of threshold given in (6), which completes the proof that the stopping rule (3) is δ-correct. The fact that the Chernoff stopping rule with the above threshold $d_{n,\delta}$ given above is δ-correct straightforwardly follows from (9).

4.3 Sketch of the proof of Theorem 3

We present a unified proof sketch of Theorem 3 for TTTS and T3C. While the two analyses follow the same steps, some of the lemmas given below have different proofs for TTTS and T3C, which can be found in Appendix C and Appendix D respectively.

We first state two important concentration results, that hold under any sampling rule.

Lemma 3. *[Lemma 5 of Qin et al. 2017] There exists a random variable W_1, such that for all $i \in \mathcal{A}$,*

$$\forall n \in \mathbb{N}, \quad |\mu_{n,i} - \mu_i| \leq \sigma W_1 \sqrt{\frac{\log(e + T_{n,i})}{1 + T_{n,i}}} \quad a.s.,$$

and $\mathbb{E}\left[e^{\lambda W_1}\right] < \infty$ for all $\lambda > 0$.

Lemma 4. *There exists a random variable W_2, such that for all $i \in \mathcal{A}$,*

$$\forall n \in \mathbb{N}, |T_{n,i} - \Psi_{n,i}| \leq W_2\sqrt{(n+1)\log(e^2 + n)} \quad a.s.,$$

and $\mathbb{E}\left[e^{\lambda W_2}\right] < \infty$ for any $\lambda > 0$.

Lemma 3 controls the concentration of the posterior means towards the true means and Lemma 4 establishes that $T_{n,i}$ and $\Psi_{n,i}$ are close. Both results rely on uniform deviation inequalities for martingales.

Our analysis uses the same principle as that of TTEI: We establish that T_β^ε is upper bounded by some random variable N which is a polynomial of the random variables W_1 and W_2 introduced in the above lemmas, denoted by $\text{Poly}(W_1, W_2) \triangleq \mathcal{O}(W_1^{c_1}W_2^{c_2})$, where c_1 and c_2 are two constants (that may depend on arms'

means and the constant hidden in the $\mathcal{O}$). As all exponential moments of W_1 and W_2 are finite, N has a finite expectation as well, which concludes the proof.

The first step to exhibit such an upper bound N is to establish that every arm is pulled sufficiently often.

Lemma 5. *Under TTTS or T3C, there exists* $N_1 = Poly(W_1, W_2)$ *s.t.* $\forall n \geq N_1$, *for all* i, $T_{n,i} \geq \sqrt{n/K}$, *almost surely.*

Due to the randomized nature of TTTS and T3C, the proof of Lemma 5 is significantly more involved than for a deterministic rule like TTEI. Intuitively, the posterior of each arm would be well concentrated once the arm is sufficiently pulled. If the optimal arm is undersampled, then it would be chosen as the first candidate with large probability. If a sub-optimal arm is undersampled, then its posterior distribution would possess a relatively wide tail that overlaps with or cover the somehow narrow tails of other overly-sampled arms. The probability of that sub-optimal arm being chosen as the challenger would be large enough then.

Combining Lemma 5 with Lemma 3 straightforwardly leads to the following result.

Lemma 6. *Under TTTS or T3C, fix a constant* $\varepsilon > 0$, *there exists* $N_2 = Poly(1/\varepsilon, W_1, W_2)$ *s.t.* $\forall n \geq N_2$,

$$\forall i \in \mathcal{A}, \quad |\mu_{n,i} - \mu_i| \leq \varepsilon.$$

We can then deduce a very nice property about the optimal action probability for sub-optimal arms from the previous two lemmas. Indeed, we can show that

$$\forall i \neq I^\star, \quad a_{n,i} \leq \exp\left\{ -\frac{\Delta_{\min}^2}{16\sigma^2} \sqrt{\frac{n}{K}} \right\}$$

for n larger than some $Poly(W_1, W_2)$. In the previous inequality, $\Delta_{\min}$ is the smallest mean difference among all the arms.

Plugging this in the expression of $\psi_{n,i}$, one can easily quantify how fast $\psi_{n,I^\star}$ converges to β, which eventually yields the following result.

Lemma 7. *Under TTTS or T3C, fix* $\varepsilon > 0$, *then there exists* $N_3 = Poly(1/\varepsilon, W_1, W_2)$ *s.t.* $\forall n \geq N_3$,

$$\left| \frac{T_{n,I^\star}}{n} - \beta \right| \leq \varepsilon.$$

The last, more involved, step is to establish that the fraction of measurement allocation to every sub-optimal arm i is indeed similarly close to its optimal proportion ω_i^β.

Lemma 8. *Under TTTS or T3C, fix a constant* $\varepsilon > 0$, *there exists* $N_4 = Poly(1/\varepsilon, W_1, W_2)$ *s.t.* $\forall n \geq N_4$,

$$\forall i \neq I^\star, \quad \left| \frac{T_{n,i}}{n} - \omega_i^\beta \right| \leq \varepsilon.$$

The major step in the proof of Lemma 8 for each sampling rule, is to establish that if some arm is oversampled, then its probability to be selected is exponentially small. Formally, we show that for n larger than some $Poly(1/\varepsilon, W_1, W_2)$,

$$\frac{\Psi_{n,i}}{n} \geq \omega_i^\beta + \xi \quad \Rightarrow \quad \psi_{n,i} \leq \exp\left\{ -f(n, \xi) \right\},$$

for some function $f(n, \xi)$ to be specified for each sampling rule, satisfying $f(n) \geq C_\xi \sqrt{n}$ (a.s.). This result leads to the concentration of $\Psi_{n,i}/n$, thus can be easily converted to the concentration of $T_{n,i}/n$ by Lemma 4.

Finally, Lemma 7 and Lemma 8 show that T_β^ε is upper bounded by $N \triangleq \max(N_3, N_4)$, which yields $\mathbb{E}[T_\beta^\varepsilon] \leq \max(\mathbb{E}[N_3], \mathbb{E}[N_4]) < \infty$.

5 Optimal Posterior Convergence

Recall that $a_{n,I^\star}$ denotes the posterior mass assigned to the event that action $I^\star$ (i.e. the true optimal arm) is optimal at time n. As the number of observations tends to infinity, we desire that the posterior distribution converges to the truth. In this section we show equivalently that the posterior mass on the complementory event, $1 - a_{n,I^\star}$, the event that arm $I^\star$ is not optimal, converges to zero at an exponential rate, and that it does so at optimal rate $\Gamma_\beta^\star$.

Russo (2016) proves a similar theorem under three confining boundedness assumptions (cf. Russo 2016, Asssumption 1) on the parameter space, the prior density and the (first derivative of the) log-normalizer of the exponential family. Hence, the theorems in Russo (2016) do not apply to the two bandit models most used in practise, which we consider in this paper: the Gaussian and Bernoulli model.

In the first case, the parameter space is unbounded, in the latter model, the derivative of the log-normalizer (which is $e^\eta/(1 + e^\eta)$) is unbounded. Here we provide two theorems, proving that under TTTS, the optimal, exponential posterior convergence rates are obtained for the Gaussian model with uninformative (improper) Gaussian priors (proof given in Appendix G), and the Bernoulli model with $\mathcal{B}eta(1,1)$ priors (proof given in Appendix H).

Theorem 4. *Under TTTS, for Gaussian bandits with improper Gaussian priors, it holds almost surely that*

$$\lim_{n \to \infty} -\frac{1}{n} \log(1 - a_{n,I^\star}) = \Gamma_\beta^\star.$$

Theorem 5. *Under TTTS, for Bernoulli bandits and uniform priors, it holds almost surely that*

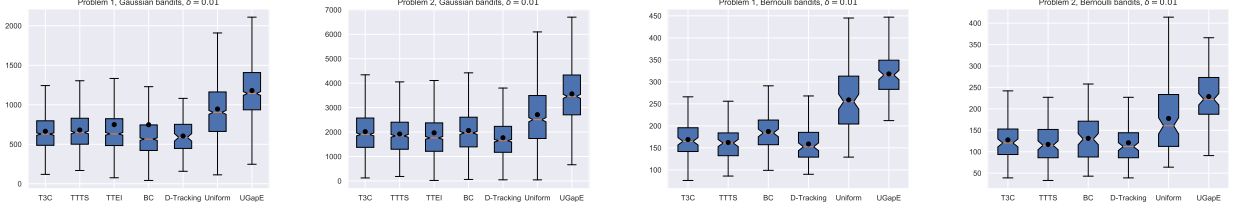$$\lim_{n \to \infty} -\frac{1}{n} \log(1 - a_{n,I^\star}) = \Gamma_\beta^\star.$$

Figure 1: black dots represent means and oranges lines represent medians.

Sampling rule	T3C	TTTS	TTEI	BC	D-Tracking	Uniform	UGapE
Execution time (s)	1.6×10^{-5}	2.3×10^{-4}	1×10^{-5}	1.4×10^{-5}	1.3×10^{-3}	6×10^{-6}	5×10^{-6}

Table 1: average execution time in seconds for different sampling rules.

6   Numerical Illustrations

This section is aimed at illustrating our theoretical results and supporting the practical use of Bayesian sampling rules for fixed-confidence BAI.

We experiment with three different Bayesian sampling rules: `T3C`, `TTTS` and `TTEI`, and we also include the Direct Tracking (`D-Tracking`) rule of Garivier and Kaufmann (2016) (which is adaptive to β), the `UGapE` (Gabillon et al., 2012) algorithm, and a uniform baseline. In order to make a fair comparison, we use the Chernoff stopping rule (4) and associated recommendation rule for all of the sampling rules, including the uniform one, except for `UGapE` which has its own stopping rule. Furthermore, we include a top-two variant of the Best Challenger (`BC`) heuristic (see, e.g., Ménard, 2019).

`BC` selects the empirical best arm $\widehat{I}_n$ with probability β and the maximizer of $W_n(\widehat{I}_n, j)$ with probability $1 - \beta$, but also performs forced exploration (selecting any arm sampled less than $\sqrt{n}$ times at round n). `T3C` can thus be viewed as a variant of `BC` in which no forced exploration is needed to converge to $\boldsymbol{\omega}^\beta$, due to the noise added by replacing $\widehat{I}_n$ with $I_n^{(1)}$.

We consider two simple instances with arms means given by $\boldsymbol{\mu}_1 = [0.5\ 0.9\ 0.4\ 0.45\ 0.44999]$, and $\boldsymbol{\mu}_2 = [1\ 0.8\ 0.75\ 0.7]$ respectively. We run simulations for both Gaussian (with $\sigma = 1$) and Bernoulli bandits with a risk parameter $\delta = 0.01$. Figure 1 reports the empirical distribution of τ_δ under the different sampling rules, estimated over 1000 independent runs.

These figures provide several insights: (1) `T3C` is competitive with, and sometimes slightly better than `TTTS` and `TTEI` in terms of sample complexity. (2) The `UGapE` algorithm has a larger sample complexity than the uniform sampling rule, which highlights the importance of the stopping rule in the fixed-confidence setting. (3) The fact that `D-Tracking` performs best is not surprising, since it converges to $\boldsymbol{\omega}^{\beta^\star}$ and achieves minimal sample complexity. However, in terms of computation time, `D-Tracking` is much worse than other sampling rules, as can be seen in Table 1, which reports the average execution time of one step of each sampling rule for μ_1 in the Gaussian case. (4) `TTTS` also suffers from computational costs, whose origins are explained in Section 2, unlike `T3C` and `TTEI`. Although `TTEI` is already computationally more attractive than `TTTS`, its practical benefits are limited to the Gaussian case, since the *Expected Improvement* (EI) does not have a closed form beyond this case and its approximation would be costly. In contrast, `T3C` can be applied for other distributions.

7   Conclusion

We have advocated the use of a Bayesian sampling rule for BAI. In particular, we proved that `TTTS` and a computationally advantageous approach `T3C`, are both β-optimal in the fixed-confidence setting, for Gaussian bandits. We further extended the Bayesian optimality properties established by Russo (2016) to more practical choices of models and prior distributions.

In order to be optimal, the sampling rules studied in this paper would need to use the oracle tuning $\beta^\star = \arg\max_{\beta \in [0,1]} \Gamma_\beta^\star$, which is not feasible. In future work, we will investigate an efficient online tuning of β to circumvent this issue. We also plan to investigate the extension of `T3C` to more general pure exploration problems, as an alternative to approaches recently proposed by Ménard (2019); Degenne et al. (2019).

Finally, it is also important to study Bayesian sampling rules in the fixed-budget setting which is more plausible in many application scenarios such as applying BAI for automated machine learning (Hoffman et al., 2014; Li et al., 2017; Shang et al., 2019).

References

Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. (2012). Online-to-confidence-set conversions and application to sparse stochastic bandits. In *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics (AIStats)*.

Audibert, J.-Y. and Bubeck, S. (2010). Best arm identification in multi-armed bandits. In *Proceedings of the 23rd Conference on Learning Theory (CoLT)*.

Auer, P., Cesa-Bianchi, N., and Fischer, P. (2002). Finite-time analysis of the multi-armed bandit problem. *Machine Learning Journal*, 47(2-3):235–256.

Bubeck, S., Munos, R., and Stoltz, G. (2009). Pure exploration in multi-armed bandits problems. In *Proceedings of the 20th International Conference on Algorithmic Learning Theory (ALT)*.

Bubeck, S., Munos, R., and Stoltz, G. (2011). Pure exploration in finitely-armed and continuous-armed bandits. *Theoretical Computer Science*, 412(19):1832–1852.

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

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

Carpentier, A. and Locatelli, A. (2016). Tight (lower) bounds for the fixed budget best arm identification bandit problem. In *Proceedings of the 29th Conference on Learning Theory (CoLT)*.

Degenne, R., Koolen, W., and Ménard, P. (2019). Non-asymptotic pure exploration by solving games. In *Advances in Neural Information Processing Systems 33 (NeurIPS)*.

Even-dar, E., Mannor, S., and Mansour, Y. (2003). Action elimination and stopping conditions for reinforcement learning. In *Proceedings of the 20th International Conference on Machine Learning (ICML)*.

Gabillon, V., Ghavamzadeh, M., and Lazaric, A. (2012). Best arm identification: A unified approach to fixed budget and fixed confidence. In *Advances in Neural Information Processing Systems 25 (NIPS)*.

Garivier, A. and Kaufmann, E. (2016). Optimal best arm identification with fixed confidence. In *Proceedings of the 29th Conference on Learning Theory (CoLT)*.

Hoffman, M. W., Shahriari, B., and de Freitas, N. (2014). On correlation and budget constraints in model-based bandit optimization with application to automatic machine learning. In *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics (AIStats)*.

Jamieson, K., Malloy, M., Nowak, R., and Bubeck, S. (2014). lil'UCB: An optimal exploration algorithm for multi-armed bandits. In *Proceedings of the 27th Conference on Learning Theory (CoLT)*.

Jun, K.-S. and Nowak, R. (2016). Anytime exploration for multi-armed bandits using confidence information. In *Proceedings of the 33rd International Conference on Machine Learning (ICML)*.

Kalyanakrishnan, S., Tewari, A., Auer, P., and Stone, P. (2012). PAC subset selection in stochastic multi-armed bandits. In *Proceedings of the 29th International Conference on Machine Learning (ICML)*.

Karnin, Z., Koren, T., and Somekh, O. (2013). Almost optimal exploration in multi-armed bandits. In *Proceedings of the 30th International Conference on Machine Learning (ICML)*.

Kaufmann, E. and Garivier, A. (2017). Learning the distribution with largest mean: two bandit frameworks. *ESAIM: Proceedings and Surveys*, 60:114–131.

Kaufmann, E. and Koolen, W. (2018). Mixture martingales revisited with applications to sequential tests and confidence intervals. *arXiv preprint arXiv:1811.11419*.

Li, L., Jamieson, K., DeSalvo, G., Talwalkar, A., and Rostamizadeh, A. (2017). Hyperband: Bandit-based configuration evaluation for hyperparameter optimization. In *Proceedings of the 5th International Conference on Learning Representations (ICLR)*.

Ménard, P. (2019). Gradient ascent for active exploration in bandit problems. *arXiv preprint arXiv:1905.08165*.

Qin, C., Klabjan, D., and Russo, D. (2017). Improving the expected improvement algorithm. In *Advances in Neural Information Processing Systems 30 (NIPS)*.

Russo, D. (2016). Simple Bayesian algorithms for best arm identification. In *Proceedings of the 29th Conference on Learning Theory (CoLT)*.

Shang, X., Kaufmann, E., and Valko, M. (2019). A simple dynamic bandit algorithm for hyperparameter tuning. In *6th Workshop on Automated*

Machine Learning at International Conference on Machine Learning (ICML-AutoML).

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285.

A   Outline

The appendix of this paper is organized as follows:

B   Useful Notation for the Appendices

In this section, we provide a list of useful notation that is applied in appendices (including reminders of previous notation in the main text and some new ones).

- Recall that $d(\mu_1; \mu_2)$ denotes the `KL`-divergence between two distributions parametrized by their means μ_1 and μ_2. For Gaussian distributions, we know that

$$d(\mu_1; \mu_2) = \frac{(\mu_1 - \mu_2)^2}{2\sigma^2}.$$

  When it comes to Bernoulli distributions, we denote this with kl, i.e.

$$kl(\mu_1; \mu_2) = \mu_1 \ln\left(\frac{\mu_1}{\mu_2}\right) + (1 - \mu_1) \ln\left(\frac{1 - \mu_1}{1 - \mu_2}\right).$$

- $\mathcal{B}eta(\cdot, \cdot)$ denotes a Beta distribution.
- $\mathcal{B}ern(\cdot)$ denotes a Bernoulli distribution.
- $\mathcal{B}(\cdot)$ denotes a Binomial distribution.
- $\mathcal{N}(\cdot, \cdot)$ denotes a normal distribution.
- $Y_{n,i}$ is the reward of arm i at time n.
- Y_{n,I_n} is the observation of the sampling rule at time n.
- $\mathcal{F}_n \triangleq \sigma(I_1, Y_{1,I_1}, I_2, Y_{2,I_2}, \cdots, I_n, Y_{n,I_n})$ is the filtration generated by the first n observations.
- $\psi_{n,i} \triangleq \mathbb{P}[I_n = i | \mathcal{F}_{n-1}]$.
- $\Psi_{n,i} \triangleq \sum_{l=1}^{n} \psi_{l,i}$.
- For the sake of simplicity, we further define $\overline{\psi}_{n,i} \triangleq \frac{\Psi_{n,i}}{n}$.
- $T_{n,i}$ is the number of pulls of arm i before round n.
- $\mathbf{T}_n$ denotes the vector of the number of arm selections.
- $I_n^\star \triangleq \arg\max_{i \in \mathcal{A}} \mu_{n,i}$ denotes the empirical best arm at time n.
- For any $a, b > 0$, define a function $C_{a,b}$ s.t. $\forall y$,

$$C_{a,b}(y) \triangleq (a + b - 1)kl(\frac{a-1}{a+b-1}; y).$$

- We define the minimum and the maximum means gap as

$$\Delta_{\min} \triangleq \min_{i \neq j} |\mu_i - \mu_j|; \Delta_{\max} \triangleq \max_{i \neq j} |\mu_i - \mu_j|.$$

- We introduce two indices

$$J_n^{(1)} \triangleq \arg\max_j a_{n,j}, J_n^{(2)} \triangleq \arg\max_{j \neq J_n^{(1)}} a_{n,j}.$$

  Note that $J_n^{(1)}$ coincides with the Bayesian recommendation index J_n.
- Two real-valued sequences (a_n) and (b_n) are are said to be logarithmically equivalent if

$$\lim_{n \to \infty} \frac{1}{n} \log\left(\frac{a_n}{b_n}\right) = 0,$$

  and we denote this by $a_n \doteq b_n$.

C Fixed-Confidence Analysis for TTTS

This section is entirely dedicated to TTTS.

C.1 Sufficient exploration of all arms, proof of Lemma 5 under TTTS

To prove this lemma, we introduce the two following sets of indices for a given $L > 0$: $\forall n \in \mathbb{N}$ we define

$$U_n^L \triangleq \{i : T_{n,i} < \sqrt{L}\},$$

$$V_n^L \triangleq \{i : T_{n,i} < L^{3/4}\}.$$

It is seemingly non trivial to manipulate directly TTTS's candidate arms, we thus start by connecting TTTS with TTPS (top two probability sampling). TTPS is another sampling rule presented by Russo (2016) for which the two candidate samples are defined as in Appendix B, we recall them in the following.

$$J_n^{(1)} \triangleq \arg\max_j a_{n,j}, J_n^{(2)} \triangleq \arg\max_{j \neq J_n^{(1)}} a_{n,j}.$$

Lemma 5 is proved via the following sequence of lemmas.

Lemma 9. *There exists $L_1 = Poly(W_1)$ s.t. if $L > L_1$, for all n, $U_n^L \neq \emptyset$ implies $J_n^{(1)} \in V_n^L$ or $J_n^{(2)} \in V_n^L$.*

Proof. If $J_n^{(1)} \in V_n^L$, then the proof is finished. Now we assume that $J_n^{(1)} \in \overline{V_n^L}$, and we prove that $J_n^{(2)} \in V_n^L$.

Step 1 According to Lemma 3, there exists $L_2 = \text{Poly}(W_1)$ s.t. $\forall L > L_2, \forall i \in \overline{U_n^L}$,

$$|\mu_{n,i} - \mu_i| \leq \sigma W_1 \sqrt{\frac{\log(e + T_{n,i})}{1 + T_{n,i}}}$$

$$\leq \sigma W_1 \sqrt{\frac{\log(e + \sqrt{L})}{1 + \sqrt{L}}}$$

$$\leq \sigma W_1 \frac{\Delta_{\min}}{4\sigma W_1} = \frac{\Delta_{\min}}{4}.$$

The second inequality holds since $x \mapsto \frac{\log(e+x)}{1+x}$ is a decreasing function. The third inequality holds for a large $L > L_2$ with $L_2 = \ldots$.

Step 2 We now assume that $L > L_2$, and we define

$$\overline{J_n^\star} \triangleq \arg\max_{j \in \overline{U_n^L}} \mu_{n,j} = \arg\max_{j \in \overline{U_n^L}} \mu_j.$$

The last equality holds since $\forall j \in \overline{U_n^L}$, $|\mu_{n,i} - \mu_i| \leq \Delta_{\min}/4$. We show that there exists $L_3 = \text{Poly}(W_1)$ s.t. $\forall L > L_3$,

$$\overline{J_n^\star} = J_n^{(1)}.$$

We proceed by contradiction, and suppose that $\overline{J_n^\star} \neq J_n^{(1)}$, then $\mu_{n,J_n^{(1)}} < \mu_{n,\overline{J_n^\star}}$, since $J_n^{(1)} \in \overline{V_n^L} \subset \overline{U_n^L}$. However, we have

$$a_{n,J_n^{(1)}} = \Pi_n \left[\theta_{J_n^{(1)}} > \max_{j \neq J_n^{(1)}} \theta_j \right]$$

$$\leq \Pi_n \left[\theta_{J_n^{(1)}} > \theta_{\overline{J_n^\star}} \right]$$

$$\leq \frac{1}{2} \exp\left\{ -\frac{(\mu_{n,J_n^{(1)}} - \mu_{n,\overline{J_n^\star}})^2}{2\sigma^2 (1/T_{n,J_n^{(1)}} + 1/T_{n,\overline{J_n^\star}})} \right\}.$$

The last inequality uses the Gaussian tail inequality (7) of Lemma 2. On the other hand,

$$
\begin{aligned}
|\mu_{n,J_n^{(1)}} - \mu_{n,\overline{J_n^\star}}| &= |\mu_{n,J_n^{(1)}} - \mu_{J_n^{(1)}} + \mu_{J_n^{(1)}} - \mu_{\overline{J_n^\star}} + \mu_{\overline{J_n^\star}} - \mu_{n,\overline{J_n^\star}}| \\
&\geq |\mu_{J_n^{(1)}} - \mu_{\overline{J_n^\star}}| - |\mu_{n,J_n^{(1)}} - \mu_{J_n^{(1)}} + \mu_{\overline{J_n^\star}} - \mu_{n,\overline{J_n^\star}}| \\
&\geq \Delta_{\min} - \left(\frac{\Delta_{\min}}{4} + \frac{\Delta_{\min}}{4}\right) \\
&= \frac{\Delta_{\min}}{2},
\end{aligned}
$$

and

$$
\frac{1}{T_{n,J_n^{(1)}}} + \frac{1}{T_{n,\overline{J_n^\star}}} \leq \frac{2}{\sqrt{L}}.
$$

Thus, if we take L_3 s.t.

$$
\exp\left\{-\frac{\sqrt{L_3}\Delta_{\min}^2}{16\sigma^2}\right\} \leq \frac{1}{2K},
$$

then for any $L > L_3$, we have

$$
a_{n,J_n^{(1)}} \leq \frac{1}{2K} < \frac{1}{K},
$$

which contradicts the definition of $J_n^{(1)}$. We now assume that $L > L_3$, thus $J_n^{(1)} = \overline{J_n^\star}$.

Step 3 We finally show that for L large enough, $J_n^{(2)} \in V_n^L$. First note that $\forall j \in \overline{V_n^L}$, we have

$$
a_{n,j} \leq \Pi_n\left[\theta_j \geq \theta_{\overline{J_n^\star}}\right] \leq \exp\left\{-\frac{L^{3/4}\Delta_{\min}^2}{16\sigma^2}\right\}. \tag{10}
$$

This last inequality can be proved using the same argument as Step 2. Now we define another index $J_n^\star \triangleq \arg\max_{j \in U_n^L} \mu_{n,j}$ and the quantity $c_n \triangleq \max(\mu_{n,J_n^\star}, \mu_{n,\overline{J_n^\star}})$. We can lower bound $a_{n,J_n^\star}$ as follows:

$$
\begin{aligned}
a_{n,J_n^\star} &\geq \Pi_n\left[\theta_{J_n^\star} \geq c_n\right] \prod_{j \neq J_n^\star} \Pi_n\left[\theta_j \leq c_n\right] \\
&= \Pi_n\left[\theta_{J_n^\star} \geq c_n\right] \prod_{j \neq J_n^\star; j \in U_n^L} \Pi_n\left[\theta_j \leq c_n\right] \prod_{j \in \overline{U_n^L}} \Pi_n\left[\theta_j \leq c_n\right] \\
&\geq \Pi_n\left[\theta_{J_n^\star} \geq c_n\right] \frac{1}{2^{K-1}}.
\end{aligned}
$$

Now there are two cases:

- If $\mu_{n,J_n^\star} > \mu_{n,\overline{J_n^\star}}$, then we have

$$
\Pi_n\left[\theta_{J_n^\star} \geq c_n\right] = \Pi_n\left[\theta_{J_n^\star} \geq \mu_{n,J_n^\star}\right] \geq \frac{1}{2}.
$$

- If $\mu_{n,J_n^\star} < \mu_{n,\overline{J_n^\star}}$, then we can apply the Gaussian tail bound (8) of Lemma 2, and we obtain

$$
\begin{aligned}
\Pi_n\left[\theta_{J_n^\star} \geq c_n\right] &= \Pi_n\left[\theta_{J_n^\star} \geq \mu_{n,\overline{J_n^\star}}\right] = \Pi_n\left[\theta_{J_n^\star} \geq \mu_{n,J_n^\star} + (\mu_{n,\overline{J_n^\star}} - \mu_{n,J_n^\star})\right] \\
&\geq \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(1 - \frac{\sqrt{T_{n,J_n^\star}}}{\sigma}(\mu_{n,J_n^\star} - \mu_{n,\overline{J_n^\star}})\right)^2\right\} \\
&= \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(1 + \frac{\sqrt{T_{n,J_n^\star}}}{\sigma}(\mu_{n,\overline{J_n^\star}} - \mu_{n,J_n^\star})\right)^2\right\}.
\end{aligned}
$$

On the other hand, by Lemma 3, we know that

$$|\mu_{n,J_n^\star} - \mu_{n,\overline{J_n^\star}}| = |\mu_{n,J_n^\star} - \mu_{J_n^\star} + \mu_{J_n^\star} - \mu_{\overline{J_n^\star}} + \mu_{\overline{J_n^\star}} - \mu_{n,\overline{J_n^\star}}|$$

$$\leq |\mu_{J_n^\star} - \mu_{\overline{J_n^\star}}| + \sigma W_1 \sqrt{\frac{\log(e + T_{n,J_n^\star})}{1 + T_{n,J_n^\star}}} + \sigma W_1 \sqrt{\frac{\log(e + T_{n,\overline{J_n^\star}})}{1 + T_{n,\overline{J_n^\star}}}}$$

$$\leq |\mu_{J_n^\star} - \mu_{\overline{J_n^\star}}| + 2\sigma W_1 \sqrt{\frac{\log(e + T_{n,J_n^\star})}{1 + T_{n,J_n^\star}}}$$

$$\leq \Delta_{\max} + 2\sigma W_1 \sqrt{\frac{\log(e + T_{n,J_n^\star})}{1 + T_{n,J_n^\star}}}.$$

Therefore,

$$\Pi_n\left[\theta_{J_n^\star} \geq c_n\right] \geq \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(1 + \frac{\sqrt{T_{n,J_n^\star}}}{\sigma}\left(\Delta_{\max} + 2\sigma W_1\sqrt{\frac{\log(e + T_{n,J_n^\star})}{1 + T_{n,J_n^\star}}}\right)\right)^2\right\}$$

$$\geq \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(1 + \frac{\sqrt{\sqrt{L}}}{\sigma}\left(\Delta_{\max} + 2\sigma W_1\sqrt{\frac{\log(e + \sqrt{L})}{1 + \sqrt{L}}}\right)\right)^2\right\}$$

$$\geq \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(1 + \frac{L^{1/4}\Delta_{\max}}{\sigma} + 2W_1\sqrt{\log(e + \sqrt{L})}\right)^2\right\}.$$

Now we have

$$a_{n,J_n^\star} \geq \max\left(\left(\frac{1}{2}\right)^K, \left(\frac{1}{2}\right)^{K-1}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(1 + \frac{L^{1/4}\Delta_{\max}}{\sigma} + 2W_1\sqrt{\log(e + \sqrt{L})}\right)^2\right\}\right),$$

and we have $\forall j \in \overline{V_n^L}$, $a_{n,j} \leq \exp\left\{-L^{3/4}\Delta_{\min}^2/(16\sigma^2)\right\}$, thus there exists $L_4 = \text{Poly}(W_1)$ s.t. $\forall L > L_4$, $\forall j \in \overline{V_n^L}$,

$$a_{n,j} \leq \frac{a_{n,J_n^\star}}{2},$$

and by consequence, $J_n^{(2)} \in V_n^L$.

Finally, taking $L_1 = \max(L_2, L_3, L_4)$, we have $\forall L > L_1$, either $J_n^{(1)} \in V_n^L$ or $J_n^{(2)} \in V_n^L$. $\qquad\square$

Next we show that there exists at least one arm in V_n^L for whom the probability of being pulled is large enough. More precisely, we prove the following lemma.

Lemma 10. *There exists $L_1 = \text{Poly}(W_1)$ s.t. for $L > L_1$ and for all n s.t. $U_n^L \neq \emptyset$, then there exists $J_n \in V_n^L$ s.t.*

$$\psi_{n,J_n} \geq \frac{\min(\beta, 1 - \beta)}{K^2} \triangleq \psi_{\min}.$$

Proof. Using Lemma 9, we know that $J_n^{(1)}$ or $J_n^{(2)} \in V_n^L$. On the other hand, we know that

$$\forall i \in \mathcal{A}, \psi_{n,i} = a_{n,i}\left(\beta + (1 - \beta)\sum_{j \neq i}\frac{a_{n,j}}{1 - a_{n,j}}\right).$$

Therefore we have

$$\psi_{n,J_n^{(1)}} \geq \beta a_{n,J_n^{(1)}} \geq \frac{\beta}{K},$$

since $\sum_{i \in \mathcal{A}} a_{n,i} = 1$, and

$$\psi_{n,J_n^{(2)}} \geq (1 - \beta) a_{n,J_n^{(2)}} \frac{a_{n,J_n^{(1)}}}{1 - a_{n,J_n^{(1)}}}$$

$$= (1 - \beta) a_{n,J_n^{(1)}} \frac{a_{n,J_n^{(2)}}}{1 - a_{n,J_n^{(1)}}}$$

$$\geq \frac{1 - \beta}{K^2},$$

since $a_{n,J_n^{(1)}} \geq 1/K$ and $\sum_{i \neq J_n^{(1)}} a_{n,i}/(1 - a_{n,J_n^{(1)}}) = 1$, thus $a_{n,J_n^{(2)}}/(1 - a_{n,J_n^{(1)}}) \geq 1/K$. $\qquad\square$

The rest of this subsection is quite similar to that of Qin et al. (2017). Indeed, with the above lemma, we can show that the set of poorly explored arms U_n^L is empty when n is large enough.

Lemma 11. *Under* TTTS, *there exists* $L_0 = Poly(W_1, W_2)$ *s.t.* $\forall L > L_0$, $U_{\lfloor KL \rfloor}^L = \emptyset$.

Proof. We proceed by contradiction, and we assume that $U_{\lfloor KL \rfloor}^L$ is not empty. Then for any $1 \leq \ell \leq \lfloor KL \rfloor$, U_ℓ^L and V_ℓ^L are non empty as well.

There exists a deterministic L_5 s.t. $\forall L > L_5$,

$$\lfloor L \rfloor \geq KL^{3/4}.$$

Using the pigeonhole principle, there exists some $i \in \mathcal{A}$ s.t. $T_{\lfloor L \rfloor, i} \geq L^{3/4}$. Thus, we have $|V_{\lfloor L \rfloor}^L| \leq K - 1$.

Next, we prove $|V_{\lfloor 2L \rfloor}^L| \leq K - 2$. Otherwise, since U_ℓ^L is non-empty for any $\lfloor L \rfloor + 1 \leq \ell \leq \lfloor 2L \rfloor$, thus by Lemma 10, there exists $J_\ell \in V_\ell^L$ s.t. $\psi_{\ell, J_\ell} \geq \psi_{\min}$. Therefore,

$$\sum_{i \in V_\ell^L} \psi_{\ell, i} \geq \psi_{\min},$$

and

$$\sum_{i \in V_{\lfloor L \rfloor}^L} \psi_{\ell, i} \geq \psi_{\min}$$

since $V_\ell^L \subset V_{\lfloor L \rfloor}^L$. Hence, we have

$$\sum_{i \in V_{\lfloor L \rfloor}^L} (\Psi_{\lfloor 2L \rfloor, i} - \Psi_{\lfloor L \rfloor, i}) = \sum_{\ell = \lfloor L \rfloor + 1}^{\lfloor 2L \rfloor} \sum_{i \in V_{\lfloor L \rfloor}^L} \psi_{\ell, i} \geq \psi_{\min} \lfloor L \rfloor.$$

Then, using Lemma 4, there exists $L_6 = \text{Poly}(W_2)$ s.t. $\forall L > L_6$, we have

$$\sum_{i \in V_{\lfloor L \rfloor}^L} (T_{\lfloor 2L \rfloor, i} - T_{\lfloor L \rfloor, i}) \geq \sum_{i \in V_{\lfloor L \rfloor}^L} (\Psi_{\lfloor 2L \rfloor, i} - \Psi_{\lfloor L \rfloor, i} - 2W_2\sqrt{\lfloor 2L \rfloor \log(e^2 + \lfloor 2L \rfloor)})$$

$$\geq \sum_{i \in V_{\lfloor L \rfloor}^L} (\Psi_{\lfloor 2L \rfloor, i} - \Psi_{\lfloor L \rfloor, i}) - 2KW_2\sqrt{\lfloor 2L \rfloor \log(e^2 + \lfloor 2L \rfloor)}$$

$$\geq \psi_{\min} \lfloor L \rfloor - 2KW_2 C_2 \lfloor L \rfloor^{3/4}$$

$$\geq KL^{3/4},$$

where C_2 is some absolute constant. Thus, we have one arm in $V_{\lfloor L \rfloor}^L$ that is pulled at least $L^{3/4}$ times between $\lfloor L \rfloor + 1$ and $\lfloor 2L \rfloor$, thus $|V_{\lfloor 2L \rfloor}^L| \leq K - 2$.

By induction, for any $1 \leq k \leq K$, we have $|V_{\lfloor kL \rfloor}^L| \leq K - k$, and finally if we take $L_0 = \max(L_1, L_5, L_6)$, then $\forall L > L_0$, $U_{\lfloor KL \rfloor}^L = \emptyset$. $\qquad\square$

We can finally conclude the proof of Lemma 5 for TTTS.

Proof of Lemma 5 Let $N_1 = KL_0$ where $L_0 = \text{Poly}(W_1, W_2)$ is chosen according to Lemma 11. For all $n > N_1$, we let $L = n/K$, then by Lemma 11, we have $U^L_{\lfloor KL \rfloor} = U^{n/K}_n$ is empty, which concludes the proof.

∎

C.2 Concentration of the empirical means, proof of Lemma 6 under TTTS

As a corollary of the previous section, we can show the concentration of $\mu_{n,i}$ to μ_i for TTTS[5].

By Lemma 3, we know that $\forall i \in \mathcal{A}$ and $n \in \mathbb{N}$,

$$|\mu_{n,i} - \mu_i| \le \sigma W_1 \sqrt{\frac{\log(e + T_{n,i})}{T_{n,i} + 1}}.$$

According to the previous section, there exists $N_1 = \text{Poly}(W_1, W_2)$ s.t. $\forall n \ge N_1$ and $\forall i \in \mathcal{A}$, $T_{n,i} \ge \sqrt{n/K}$. Therefore,

$$|\mu_{n,i} - \mu_i| \le \sqrt{\frac{\log(e + \sqrt{n/K})}{\sqrt{n/K} + 1}},$$

since $x \mapsto \log(e + x)/(x + 1)$ is a decreasing function. There exists $N'_2 = \text{Poly}(\varepsilon, W_1)$ s.t. $\forall n \ge N'_2$,

$$\sqrt{\frac{\log(e + \sqrt{n/K})}{\sqrt{n/K} + 1}} \le \sqrt{\frac{2(n/K)^{1/4}}{\sqrt{n/K} + 1}} \le \frac{\varepsilon}{\sigma W_1}.$$

Therefore, $\forall n \ge N_2 \triangleq \max\{N_1, N'_2\}$, we have

$$|\mu_{n,i} - \mu_i| \le \sigma W_1 \frac{\varepsilon}{\sigma W_1}.$$

C.3 Measurement effort concentration of the optimal arm, proof of Lemma 7 under TTTS

In this section we show that the empirical arm draws proportion of the true best arm for TTTS concentrates to β when the total number of arm draws is sufficiently large.

The proof is established upon the following lemmas. First, we prove that the empirical best arm coincides with the true best arm when the total number of arm draws goes sufficiently large.

Lemma 12. *Under TTTS, there exists $M_1 = \text{Poly}(W_1, W_2)$ s.t. $\forall n > M_1$, we have $I^\star_n = I^\star = J^{(1)}_n$ and $\forall i \ne I^\star$,*

$$a_{n,i} \le \exp\left\{-\frac{\Delta^2_{min}}{16\sigma^2}\sqrt{\frac{n}{K}}\right\}.$$

Proof. Using Lemma 6 with $\varepsilon = \Delta_{\min}/4$, there exists $N'_1 = \text{Poly}(4/\Delta_{\min}, W_1, W_2)$ s.t. $\forall n > N'_1$,

$$\forall i \in \mathcal{A}, |\mu_{n,i} - \mu_i| \le \frac{\Delta_{\min}}{4},$$

which implies that starting from a known moment, $\mu_{n,I^\star} > \mu_{n,i}$ for all $i \ne I^\star$, hence $I^\star_n = I^\star$. Thus, $\forall i \ne I^\star$,

$$\begin{aligned}
a_{n,i} &= \Pi_n\left[\theta_i > \max_{j \ne i}\theta_j\right]\\
&\le \Pi_n[\theta_i > \theta_{I^\star}]\\
&\le \frac{1}{2}\exp\left\{-\frac{(\mu_{n,i} - \mu_{n,I^\star})^2}{2\sigma^2(1/T_{n,i} + 1/T_{n,I^\star})}\right\}.
\end{aligned}$$

---

[5]this proof is the same as Proposition 3 of Qin et al. (2017)

The last inequality uses the Gaussian tail inequality of (7) Lemma 2. Furthermore,

$$
\begin{aligned}
(\mu_{n,i} - \mu_{n,I^\star})^2 &= (|\mu_{n,i} - \mu_{n,I^\star}|)^2 \\
&= (|\mu_{n,i} - \mu_i + \mu_i - \mu_{I^\star} + \mu_{I^\star} - \mu_{n,I^\star}|)^2 \\
&\geq (|\mu_i - \mu_{I^\star}| - |\mu_{n,i} - \mu_i + \mu_{I^\star} - \mu_{n,I^\star}|)^2 \\
&\geq \left(\Delta_{\min} - \left(\frac{\Delta_{\min}}{4} + \frac{\Delta_{\min}}{4} \right) \right)^2 = \frac{\Delta_{\min}^2}{4},
\end{aligned}
$$

and according to Lemma 5, we know that there exists $M_2 = \mathrm{Poly}(W_1, W_2)$ s.t. $\forall n > M_2$,

$$
\frac{1}{T_{n,i}} + \frac{1}{T_{n,I^\star}} \leq \frac{2}{\sqrt{n/K}}.
$$

Thus, $\forall n > \max\{N_1', M_2\}$, we have

$$
\forall i \neq I^\star, a_{n,i} \leq \exp\left\{ -\frac{\Delta_{\min}^2}{16\sigma^2} \sqrt{\frac{n}{K}} \right\}.
$$

Then, we have

$$
a_{n,I^\star} = 1 - \sum_{i \neq I^\star} a_{n,i} \geq 1 - (K-1) \exp\left\{ -\frac{\Delta_{\min}^2}{16\sigma^2} \sqrt{\frac{n}{K}} \right\}.
$$

There exists M_2' s.t. $\forall n > M_2'$, $a_{n,I^\star} > 1/2$, and by consequence $I^\star = J_n^{(1)}$. Finally taking $M_1 \triangleq \max\{N_1', M_2, M_2'\}$ concludes the proof. $\qquad\square$

Before we prove Lemma 7, we first show that $\Psi_{n,I^\star}/n$ concentrates to β.

Lemma 13. *Under TTTS, fix a constant $\varepsilon > 0$, there exists $M_3 = Poly(\varepsilon, W_1, W_2)$ s.t. $\forall n > M_3$, we have*

$$
\left| \frac{\Psi_{n,I^\star}}{n} - \beta \right| \leq \varepsilon.
$$

Proof. By Lemma 12, we know that there exists $M_1' = \mathrm{Poly}(W_1, W_2)$ s.t. $\forall n > M_1'$, we have $I_n^\star = I^\star = J_n^{(1)}$ and $\forall i \neq I^\star$,

$$
a_{n,i} \leq \exp\left\{ -\frac{\Delta_{\min}^2}{16\sigma^2} \sqrt{\frac{n}{K}} \right\}.
$$

Note also that $\forall n \in \mathbb{N}$, we have

$$
\psi_{n,I^\star} = a_{n,I^\star} \left(\beta + (1-\beta) \sum_{j \neq I^\star} \frac{a_{n,j}}{1 - a_{n,j}} \right).
$$

We proceed the proof with the following two steps.

Step 1 We first lower bound $\Psi_{n,I^\star}$ for a given ε. Take $M_4 > M_1'$ that we decide later, we have $\forall n > M_4$,

$$
\begin{aligned}
\frac{\Psi_{n,I^\star}}{n} &= \frac{1}{n}\sum_{l=1}^{n}\psi_{l,I^\star} = \frac{1}{n}\sum_{l=I^\star}^{M_4}\psi_{l,I^\star} + \frac{1}{n}\sum_{l=M_4+1}^{n}\psi_{l,I^\star} \\
&\geq \frac{1}{n}\sum_{l=M_4+1}^{n}\psi_{l,I^\star} \geq \frac{1}{n}\sum_{l=M_4+1}^{n}a_{l,I^\star}\beta \\
&= \frac{\beta}{n}\sum_{l=M_4+1}^{n}\left(1 - \sum_{j\neq I^\star}a_{l,j}\right) \\
&\geq \frac{\beta}{n}\sum_{l=M_4+1}^{n}\left(1 - (K-1)\exp\left\{-\frac{\Delta_{\min}^2}{16\sigma^2}\sqrt{\frac{l}{K}}\right\}\right) \\
&= \beta - \frac{M_4}{n}\beta - \frac{\beta}{n}\sum_{l=M_4+1}^{n}(K-1)\exp\left\{-\frac{\Delta_{\min}^2}{16\sigma^2}\sqrt{\frac{l}{K}}\right\} \\
&\geq \beta - \frac{M_4}{n}\beta - \frac{(n-M_4)}{n}\beta(K-1)\exp\left\{-\frac{\Delta_{\min}^2}{16\sigma^2}\sqrt{\frac{M_4}{K}}\right\} \\
&\geq \beta - \frac{M_4}{n}\beta - \beta(K-1)\exp\left\{-\frac{\Delta_{\min}^2}{16\sigma^2}\sqrt{\frac{M_4}{K}}\right\}.
\end{aligned}
$$

For a given constant $\varepsilon > 0$, there exists M_5 s.t. $\forall n > M_5$,

$$
\beta(K-1)\exp\left\{-\frac{\Delta_{\min}^2}{16\sigma^2}\sqrt{\frac{n}{K}}\right\} < \frac{\varepsilon}{2}.
$$

Furthermore, there exists $M_6 = \text{Poly}(\varepsilon/2, M_5)$ s.t. $\forall n > M_6$,

$$
\frac{M_5}{n}\beta < \frac{\varepsilon}{2}.
$$

Therefore, if we take $M_4 \triangleq \max\{M_1', M_5, M_6\}$, we have $\forall n > M_4$,

$$
\frac{\Psi_{n,I^\star}}{n} \geq \beta - \varepsilon.
$$

Step 2 On the other hand, we can also upper bound $\Psi_{n,I^\star}$. We have $\forall n > M_3$,

$$
\begin{aligned}
\frac{\Psi_{n,I^\star}}{n} &= \frac{1}{n}\sum_{l=1}^{n}\psi_{l,I^\star} \\
&= \frac{1}{n}\sum_{l=1}^{n}a_{l,I^\star}\left(\beta + (1-\beta)\sum_{j\neq I^\star}\frac{a_{l,j}}{1-a_{l,j}}\right) \\
&\leq \frac{1}{n}\sum_{l=1}^{n}a_{l,I^\star}\beta + \frac{1}{n}\sum_{l=1}^{n}a_{l,I^\star}(1-\beta)\sum_{j\neq I^\star}\frac{a_{l,j}}{1-a_{l,j}} \\
&\leq \beta + \frac{1}{n}\sum_{l=1}^{n}(1-\beta)\sum_{j\neq I^\star}\frac{a_{l,j}}{1-a_{l,j}} \\
&\leq \beta + \frac{1}{n}\sum_{l=1}^{n}(1-\beta)\sum_{j\neq I^\star}\frac{\exp\left\{-\frac{\Delta_{\min}^2}{16\sigma^2}\sqrt{\frac{l}{K}}\right\}}{1-\exp\left\{-\frac{\Delta_{\min}^2}{16\sigma^2}\sqrt{\frac{l}{K}}\right\}}.
\end{aligned}
$$

Since, for a given $\varepsilon > 0$, there exists M_8 s.t. $\forall n > M_8$,

$$\exp\left\{-\frac{\Delta_{\min}^2}{16\sigma^2}\sqrt{\frac{n}{K}}\right\} < \frac{1}{2},$$

and there exists M_9 s.t. $\forall n > M_9$,

$$(1-\beta)(K-1)\exp\left\{-\frac{\Delta_{\min}^2}{16\sigma^2}\sqrt{\frac{n}{K}}\right\} < \frac{\varepsilon}{4}.$$

Thus, $\forall n > M_{10} \triangleq \max\{M_8, M_9\}$,

$$\frac{\Psi_{n,I^\star}}{n} \leq \beta + \frac{1-\beta}{n}\left(\sum_{l=1}^{M_{10}}\sum_{j\neq I^\star}\frac{\exp\left\{-\frac{\Delta_{\min}^2}{16\sigma^2}\sqrt{\frac{l}{K}}\right\}}{1-\exp\left\{-\frac{\Delta_{\min}^2}{16\sigma^2}\sqrt{\frac{l}{K}}\right\}} + \sum_{l=M_{10}+1}^{n}\sum_{j\neq I^\star}\frac{\exp\left\{-\frac{\Delta_{\min}^2}{16\sigma^2}\sqrt{\frac{l}{K}}\right\}}{1-\exp\left\{-\frac{\Delta_{\min}^2}{16\sigma^2}\sqrt{\frac{l}{K}}\right\}}\right)$$

$$\leq \beta + \frac{1-\beta}{n}\sum_{l=1}^{M_{10}}\sum_{j\neq I^\star}\frac{\exp\left\{-\frac{\Delta_{\min}^2}{16\sigma^2}\sqrt{\frac{l}{K}}\right\}}{1-\exp\left\{-\frac{\Delta_{\min}^2}{16\sigma^2}\sqrt{\frac{l}{K}}\right\}} + 2(1-\beta)(K-1)\exp\left\{-\frac{\Delta_{\min}^2}{16\sigma^2}\sqrt{\frac{M_{10}}{K}}\right\}$$

$$\leq \beta + \frac{1-\beta}{n}\sum_{l=1}^{M_{10}}\sum_{j\neq I^\star}\frac{\exp\left\{-\frac{\Delta_{\min}^2}{16\sigma^2}\sqrt{\frac{l}{K}}\right\}}{1-\exp\left\{-\frac{\Delta_{\min}^2}{16\sigma^2}\sqrt{\frac{l}{K}}\right\}} + \frac{\varepsilon}{2}.$$

There exists $M_{11} = \text{Poly}(\varepsilon/2, M_{10})$ s.t. $\forall n > M_{11}$,

$$\frac{1-\beta}{n}\sum_{l=1}^{M_{10}}\sum_{j\neq I^\star}\frac{\exp\left\{-\frac{\Delta_{\min}^2}{16\sigma^2}\sqrt{\frac{l}{K}}\right\}}{1-\exp\left\{-\frac{\Delta_{\min}^2}{16\sigma^2}\sqrt{\frac{l}{K}}\right\}} < \frac{\varepsilon}{2}.$$

Therefore, $\forall n > M_7 \triangleq \max\{M_3, M_{11}\}$, we have

$$\frac{\Psi_{n,I^\star}}{n} \leq \beta + \varepsilon.$$

Conclusion Finally, combining the two steps and define $M_3 \triangleq \max\{M_4, M_7\}$, we have $\forall n > M_3$,

$$\left|\frac{\Psi_{n,I^\star}}{n} - \beta\right| \leq \varepsilon.$$

$\square$

With the help of the previous lemma and Lemma 4, we can finally prove Lemma 7.

Proof of Lemma 7 Fix an $\varepsilon > 0$. Using Lemma 4, we have $\forall n \in \mathbb{N}$,

$$\left|\frac{T_{n,I^\star}}{n} - \frac{\Psi_{n,I^\star}}{n}\right| \leq \frac{W_2\sqrt{(n+1)\log(e^2+n)}}{n}.$$

Thus there exists M_{12} s.t. $\forall n > M_{12}$,

$$\left|\frac{T_{n,I^\star}}{n} - \frac{\Psi_{n,I^\star}}{n}\right| \leq \frac{\varepsilon}{2}.$$

And using Lemma 13, there exists $M_3' = \text{Poly}(\varepsilon/2, W_1, W_2)$ s.t. $\forall n > M_3'$,

$$\left|\frac{\Psi_{n,I^\star}}{n} - \beta\right| \leq \frac{\varepsilon}{2}.$$

Again, according to Lemma 10, there exists M_3' s.t. $\forall n > M_3'$,

$$\frac{\Psi_{n,I^\star}}{n} \leq \beta + \frac{\varepsilon}{2}.$$

Thus, if we take $N_3 \triangleq \max\{M_3', M_{12}\}$, then $\forall n > N_3$, we have

$$\left| \frac{T_{n,I^\star}}{n} - \beta \right| \leq \varepsilon.$$

■

C.4 Measurement effort concentration of other arms, proof of Lemma 8 under TTTS

In this section, we show that, for TTTS, the empirical measurement effort concentration also holds for other arms than the true best arm. We first show that if some arm is overly sampled at time n, then its probability of being picked is reduced exponentially.

Lemma 14. *Under TTTS, for every $\xi \in (0,1)$, there exists $S_1 = Poly(1/\xi, W_1, W_2)$ such that for all $n > S_1$, for all $i \neq I^\star$,*

$$\frac{\Psi_{n,i}}{n} \geq \omega_i^\beta + \xi \quad \Rightarrow \quad \psi_{n,i} \leq \exp\{-\varepsilon_0(\xi)n\},$$

where ε_0 is defined in (11) below.

Proof. First, by Lemma 12, there exists $M_1'' = \text{Poly}(W_1, W_2)$ s.t. $\forall n > M_1''$,

$$I^\star = I_n^\star = J_n^{(1)}.$$

Then, following the similar argument as in Lemma 25, one can show that for all $i \neq I^\star$ and for all $n > M_1''$,

$$
\begin{aligned}
\psi_{n,i} &= a_{n,i}\left(\beta + (1-\beta)\sum_{j \neq i}\frac{a_{n,j}}{1 - a_{n,j}}\right) \\
&\leq a_{n,i}\beta + a_{n,i}(1-\beta)\frac{\sum_{j \neq i} a_{n,j}}{1 - a_{n,J_n^{(1)}}} \\
&= a_{n,i}\beta + a_{n,i}(1-\beta)\frac{\sum_{j \neq i} a_{n,j}}{1 - a_{n,I^\star}} \\
&\leq a_{n,i}\beta + a_{n,i}(1-\beta)\frac{1}{1 - a_{n,I^\star}} \\
&\leq \frac{a_{n,i}}{1 - a_{n,I^\star}} \\
&\leq \frac{\Pi_n\left[\theta_i \geq \theta_{I^\star}\right]}{\Pi_n\left[\cup_{j \neq I^\star}\theta_j \geq \theta_{I^\star}\right]} \\
&\leq \frac{\Pi_n\left[\theta_i \geq \theta_{I^\star}\right]}{\max_{j \neq I^\star}\Pi_n\left[\theta_j \geq \theta_{I^\star}\right]}.
\end{aligned}
$$

Using the upper and lower Gaussian tail bounds from Lemma 2, we have

$$
\psi_{n,i} \leq \frac{\exp\left\{-\dfrac{(\mu_{n,I^\star} - \mu_{n,i})^2}{2\sigma^2\left(1/T_{n,I^\star} + 1/T_{n,i}\right)}\right\}}{\exp\left\{-\min_{j \neq I^\star} \dfrac{1}{2}\left(\dfrac{(\mu_{n,I^\star} - \mu_{n,j})}{\sigma\sqrt{(1/T_{n,I^\star} + 1/T_{n,j})}} - 1\right)^2\right\}}
$$

$$
= \frac{\exp\left\{-n\dfrac{(\mu_{n,I^\star} - \mu_{n,i})^2}{2\sigma^2\left(n/T_{n,I^\star} + n/T_{n,i}\right)}\right\}}{\exp\left\{-n\left(\min_{j \neq I^\star} \dfrac{(\mu_{n,I^\star} - \mu_{n,j})}{\sqrt{2\sigma^2\left(n/T_{n,I^\star} + n/T_{n,j}\right)}} - \dfrac{1}{\sqrt{2n}}\right)^2\right\}},
$$

where we assume that $n > S_2 = \mathrm{Poly}(W_1, W_2)$ for which

$$
\frac{(\mu_{n,I^\star} - \mu_{n,i})^2}{\sigma^2\left(1/T_{n,I^\star} + 1/T_{n,i}\right)} \geq 1
$$

according to Lemma 5. From there we take a supremum over the possible allocations to lower bound the denominator and write

$$
\psi_{n,i} \leq \frac{\exp\left\{-n\dfrac{(\mu_{n,I^\star} - \mu_{n,i})^2}{2\sigma^2\left(n/T_{n,I^\star} + n/T_{n,i}\right)}\right\}}{\exp\left\{-n\left(\sup_{\boldsymbol{\omega}:\omega_{I^\star} = T_{n,I^\star}/n}\min_{j \neq I^\star} \dfrac{(\mu_{n,I^\star} - \mu_{n,i})}{\sqrt{2\sigma^2\left(1/\omega_{I^\star} + 1/\omega_j\right)}} - \dfrac{1}{\sqrt{2n}}\right)^2\right\}}
$$

$$
= \frac{\exp\left\{-n\dfrac{(\mu_{n,I^\star} - \mu_{n,i})^2}{2\sigma^2\left(n/T_{n,I^\star} + n/T_{n,i}\right)}\right\}}{\exp\left\{-n\left(\sqrt{\Gamma^\star_{T_{n,I^\star}/n}(\boldsymbol{\mu}_n)} - \dfrac{1}{\sqrt{2n}}\right)^2\right\}},
$$

where $\boldsymbol{\mu}_n \triangleq (\mu_{n,1}, \cdots, \mu_{n,K})$, and $(\beta, \boldsymbol{\mu}) \mapsto \Gamma^\star_\beta(\boldsymbol{\mu})$ represents a function that maps β and $\boldsymbol{\mu}$ to the parameterized optimal error decay that any allocation rule can reach given parameter β and a set of arms with means $\boldsymbol{\mu}$. Note that this function is continuous with respect to β and $\boldsymbol{\mu}$ respectively.

Now, assuming $\Psi_{n,i}/n \geq \omega_i^\beta + \xi$ yields that there exists $S_2' \triangleq \mathrm{Poly}(2/\xi, W_2)$ s.t. for all $n > S_2'$, $T_{n,i}/n \geq \omega_i^\beta + \xi/2$, and by consequence,

$$
\psi_{n,i} \leq \exp\left\{-n\left(\underbrace{\frac{(\mu_{n,I^\star} - \mu_{n,i})^2}{2\sigma^2\left(n/T_{n,I^\star} + 1/(\omega_i^\beta + \xi/2)\right)} - \Gamma^\star_{T_{n,I^\star}/n}(\boldsymbol{\mu}_n) - \frac{1}{2n} + \sqrt{\frac{2\Gamma^\star_{T_{n,I^\star}/n}(\boldsymbol{\mu}_n)}{n}}}_{\varepsilon_n(\xi)}\right)\right\}.
$$

Using Lemma 7, we know that for any ε, there exists $S_3 = \mathrm{Poly}(1/\varepsilon, W_1, W_2)$ s.t. $\forall n > S_3$, $|T_{n,I^\star}/n - \beta| \leq \varepsilon$, and $\forall j \in \mathcal{A}, |\mu_{n,j} - \mu_j| \leq \varepsilon$. Furthermore, $(\beta, \boldsymbol{\mu}) \mapsto \Gamma^\star_\beta(\boldsymbol{\mu})$ is continuous with respect to β and $\boldsymbol{\mu}$, thus for a given ε_0, there exists $S_3' = \mathrm{Poly}(1/\varepsilon_0, W_1, W_2)$ s.t. $\forall n > S_3'$, we have

$$
\left|\varepsilon_n(\xi) - \left(\frac{(\mu_{I^\star} - \mu_i)^2}{2\sigma^2\left(1/\beta + 1/(\omega_i^\beta + \xi/2)\right)} - \Gamma^\star_\beta\right)\right| \leq \varepsilon_0.
$$

Finally, define $S_1 \triangleq \max\{S_2, S_2', S_3'\}$, we have $\forall n > S_1$,

$$
\psi_{n,i} \leq \exp\left\{-\varepsilon_0(\xi)n\right\},
$$

where

$$\varepsilon_0(\xi) = \frac{(\mu_{I^\star} - \mu_i)^2}{2\sigma^2 \left(1/\beta + 1/(\omega_i^\beta + \xi/2)\right)} - \Gamma_\beta^\star + \varepsilon_0 \,. \tag{11}$$

$\square$

Next, starting from some known moment, no arm is overly allocated. More precisely, we show the following lemma.

Lemma 15. *Under* TTTS, *for every* ξ, *there exists* $S_4 = Poly(1/\xi, W_1, W_2)$ *s.t.* $\forall n > S_4$,

$$\forall i \in \mathcal{A}, \quad \frac{\Psi_{n,i}}{n} \le \omega_i^\beta + \xi.$$

Proof. From Lemma 14, there exists $S_1' = \mathrm{Poly}(2/\xi, W_1, W_2)$ such that for all $n > S_1'$ and for all $i \ne I^\star$,

$$\frac{\Psi_{n,i}}{n} \ge \omega_i^\beta + \frac{\xi}{2} \quad \Rightarrow \quad \psi_{n,i} \le \exp\left\{-\varepsilon_0(\xi/2)n\right\}.$$

Thus, for all $i \ne I^\star$,

$$\frac{\Psi_{n,i}}{n} \le \frac{S_1'}{n} + \frac{\displaystyle\sum_{\ell=S_1'+1}^{n} \psi_{\ell,i}\mathbb{1}\left(\frac{\Psi_{\ell,i}}{n} \ge \omega_i^\beta + \frac{\xi}{2}\right)}{n} + \frac{\displaystyle\sum_{\ell=S_1'+1}^{n} \psi_{\ell,i}\mathbb{1}\left(\frac{\Psi_{\ell,i}}{n} \le \omega_i^\beta + \frac{\xi}{2}\right)}{n}$$

$$\le \frac{S_1'}{n} + \frac{\displaystyle\sum_{\ell=1}^{n} \exp\left\{-\varepsilon_0(\xi/2)n\right\}}{n} + \frac{\displaystyle\sum_{\ell=S_1'+1}^{\ell_n(\xi)} \psi_{\ell,i}\mathbb{1}\left(\frac{\Psi_{\ell,i}}{n} \le \omega_i^\beta + \frac{\xi}{2}\right)}{n},$$

where we let $\ell_n(\xi) = \max\left\{\ell \le n : \Psi_{\ell,i}/n \le \omega_i^\beta + \xi/2\right\}$. Then

$$\frac{\Psi_{n,i}}{n} \le \frac{S_1'}{n} + \frac{\displaystyle\sum_{\ell=1}^{n} \exp\left\{-\varepsilon_0(\xi/2)n\right\}}{n} + \Psi_{\ell_n(\xi),i}$$

$$\le \frac{S_1' + (1 - \exp(-\varepsilon_0(\xi/2)))^{-1}}{n} + \omega_i^\beta + \frac{\xi}{2}$$

Then, there exists S_5 such that for all $n \ge S_5$,

$$\frac{S_1' + (1 - \exp(-\varepsilon_0(\xi/2)))^{-1}}{n} \le \frac{\xi}{2}.$$

Therefore, for any $n > S_4 \triangleq \max\{S_1', S_5\}$, $\Psi_{n,i} \le \omega_i^\beta + \xi$ holds for all $i \ne I^\star$. For $i = I^\star$, it is already proved for the optimal arm. $\square$

We now prove Lemma 8 under TTTS.

Proof of Lemma 8 From Lemma 15, there exists $S_4' = \mathrm{Poly}((K-1)/\xi, W_1, W_2)$ such that for all $n > S_4'$,

$$\forall i \in \mathcal{A}, \frac{\Psi_{n,i}}{n} \le \omega_i^\beta + \frac{\xi}{K-1}.$$

Using the fact that $\Psi_{n,i}/n$ and ω_i^β all sum to 1, we have $\forall i \in \mathcal{A}$,

$$\frac{\Psi_{n,i}}{n} = 1 - \sum_{j \ne i} \frac{\Psi_{n,j}}{n}$$

$$\ge 1 - \sum_{j \ne i} \left(\omega_j^\beta + \frac{\xi}{K-1}\right)$$

$$= \omega_i^\beta - \xi.$$

Thus, for all $n > S_4'$, we have

$$\forall i \in \mathcal{A}, \left| \frac{\Psi_{n,i}}{n} - \omega_i^\beta \right| \leq \xi.$$

And finally we use the same reasoning as the proof of Lemma 7 to link $T_{n,i}$ and $\Psi_{n,i}$. Fix an $\varepsilon > 0$. Using Lemma 4, we have $\forall n \in \mathbb{N}$,

$$\forall i \in \mathcal{A}, \left| \frac{T_{n,i}}{n} - \frac{\Psi_{n,i}}{n} \right| \leq \frac{W_2 \sqrt{(n+1) \log(e^2 + n)}}{n}.$$

Thus there exists S_5 s.t. $\forall n > S_5$,

$$\left| \frac{T_{n,I^\star}}{n} - \frac{\Psi_{n,I^\star}}{n} \right| \leq \frac{\varepsilon}{2}.$$

And using the above result, there exists $S_4'' = \mathrm{Poly}(2/\varepsilon, W_1, W_2)$ s.t. $\forall n > S_4''$,

$$\left| \frac{\Psi_{n,i}}{n} - \omega_i^\beta \right| \leq \frac{\varepsilon}{2}.$$

Thus, if we take $N_4 \triangleq \max\{S_4'', S_5\}$, then $\forall n > N_4$, we have

$$\forall i \in \mathcal{A}, \left| \frac{T_{n,i}}{n} - \omega_i^\beta \right| \leq \varepsilon.$$

$\blacksquare$

D   Fixed-Confidence Analysis for T3C

This section is entirely dedicated to T3C. Note that the analysis to follow share the same proof line with that of TTTS, and some parts even completely coincide with those of TTTS. For the sake of simplicity and clearness, we shall only focus on the parts that differ and skip some redundant proofs.

D.1   Sufficient exploration of all arms, proof of Lemma 5 under T3C

To prove this lemma, we still need the two sets of indices for under-sampled arms like in Appendix C.1. We recall that for a given $L > 0$: $\forall n \in \mathbb{N}$ we define

$$U_n^L \triangleq \{i : T_{n,i} < \sqrt{L}\},$$

$$V_n^L \triangleq \{i : T_{n,i} < L^{3/4}\}.$$

For T3C however, we investigate the following two indices,

$$J_n^{(1)} \triangleq \arg\max_j a_{n,j}, \widetilde{J_n^{(2)}} \triangleq \arg\min_{j \neq J_n^{(1)}} W_n(J_n^{(1)}, j).$$

Lemma 5 is proved via the following sequence of lemmas.

Lemma 16. *There exists $L_1 = Poly(W_1)$ s.t. if $L > L_1$, for all n, $U_n^L \neq \emptyset$ implies $J_n^{(1)} \in V_n^L$ or $\widetilde{J_n^{(2)}} \in V_n^L$.*

Proof. If $J_n^{(1)} \in V_n^L$, then the proof is finished. Now we assume that $J_n^{(1)} \in \overline{V_n^L} \subset \overline{U_n^L}$, and we prove that $J_n^{(2)} \in V_n^L$.

Step 1   Following the same reasoning as Step 1 and Step 2 of the proof of Lemma 9, we know that there exists $L_2 = \mathrm{Poly}(W_1)$ s.t. if $L > L_2$, then

$$\overline{J_n^\star} \triangleq \arg\max_{j \in \overline{U_n^L}} \mu_{n,j} = \arg\max_{j \in \overline{U_n^L}} \mu_j = J_n^{(1)}.$$

Step 2 Now assuming that $L > L_2$, and we show that for L large enough, $\widetilde{J_n^{(2)}} \in V_n^L$. In the same way that we proved (10) one can show that for all $\forall j \in \overline{V_n^L}$,

$$W_n(J_n^{(1)}, j) = \frac{(\mu_{n,I^\star} - \mu_{n,j})^2}{2\sigma^2\left(\dfrac{1}{T_{n,I^\star}} + \dfrac{1}{T_{n,j}}\right)} \geq \frac{L^{3/4}\Delta_{\min}^2}{16\sigma^2}.$$

Again, denote $J_n^\star \triangleq \arg\max_{j \in U_n^L} \mu_{n,j}$, we obtain

$$W_n(J_n^{(1)}, J_n^\star) = \begin{cases} 0 & \text{if } \mu_{n,J_n^\star} \geq \mu_{n,J_n^{(1)}}, \\ \dfrac{(\mu_{n,J_n^{(1)}} - \mu_{n,J_n^\star})^2}{2\sigma^2\left(\dfrac{1}{T_{n,J_n^{(1)}}} + \dfrac{1}{T_{n,J_n^\star}}\right)} & \text{else.} \end{cases}$$

In the second case, as already shown in Step 3 of Lemma 9 we have that

$$|\mu_{n,J_n^\star} - \mu_{n,\overline{J_n^\star}}| \leq \Delta_{\max} + 2\sigma W_1 \sqrt{\frac{\log(e + T_{n,J_n^\star})}{1 + T_{n,J_n^\star}}}$$

$$\leq \Delta_{\max} + 2\sigma W_1 \sqrt{\frac{\log(e + \sqrt{L})}{1 + \sqrt{L}}},$$

since $J_n^\star \in U_n^L$. We also know that

$$2\sigma^2\left(\frac{1}{T_{n,J_n^{(1)}}} + \frac{1}{T_{n,J_n^\star}}\right) \geq \frac{2\sigma^2}{T_{n,J_n^\star}} \geq \frac{2\sigma^2}{\sqrt{L}}.$$

Therefore, we get

$$W_n(J_n^{(1)}, J_n^\star) \leq \frac{\sqrt{L}}{2\sigma^2}\left(\Delta_{\max} + 2\sigma W_1 \sqrt{\frac{\log(e + \sqrt{L})}{1 + \sqrt{L}}}\right)^2.$$

On the other hand, we know that for all $j \in \overline{V_n^L}$,

$$W_n(J_n^{(1)}, j) \geq \frac{L^{3/4}\Delta_{\min}^2}{16\sigma^2}.$$

Thus, there exists L_3 s.t. if $L > L_3$, then

$$\forall j \in \overline{V_n^L}, \ W_n(J_n^{(1)}, j) \geq 2W_n(J_n^{(1)}, J_n^\star).$$

That means $\widetilde{J_n^{(2)}} \notin \overline{V_n^L}$ and by consequence, $\widetilde{J_n^{(2)}} \in V_n^L$.

Finally, taking $L_1 = \max(L_2, L_3)$, we have $\forall L > L_1$, either $J_n^{(1)} \in V_n^L$ or $\widetilde{J_n^{(2)}} \in V_n^L$. $\qquad\square$

Next we show that there exists at least one arm in V_n^L for whom the probability of being pulled is large enough. More precisely, we prove the following lemma.

Lemma 17. *There exists $L_1 = Poly(W_1)$ s.t. for $L > L_1$ and for all n s.t. $U_n^L \neq \emptyset$, then there exists $J_n \in V_n^L$ s.t.*

$$\psi_{n,J_n} \geq \frac{\min(\beta, 1 - \beta)}{K^2} \triangleq \psi_{\min}.$$

Proof. Using Lemma 16, we know that $J_n^{(1)}$ or $\widetilde{J_n^{(2)}} \in V_n^L$. We also know that under T3C, for any arm i, $\psi_{n,i}$ can be written as

$$\psi_{n,i} = \beta a_{n,i} + (1 - \beta)\sum_{j \neq i} a_{n,j} \frac{\mathbb{1}\{W_n(j,i) = \min_{k \neq j} W_n(j,k)\}}{|\arg\min_{k \neq j} W_n(j,k)|}.$$

Note that $(\psi_{n,i})_i$ sums to 1,

$$
\sum_i \psi_{n,i} = \beta + (1-\beta) \sum_j a_{n,j} \sum_{i \neq j} \frac{\mathbb{1}\{W_n(j,i) = \min_{k \neq j} W_n(j,k)\}}{\left| \arg\min_{k \neq j} W_n(j,k) \right|}
$$
$$
= \beta + (1-\beta) \sum_j a_{n,j} = 1 \,.
$$

Therefore, we have

$$
\psi_{n, J_n^{(1)}} \geq \beta a_{n, J_n^{(1)}} \geq \frac{\beta}{K}
$$

on one hand, since $\sum_{i \in \mathcal{A}} a_{n,i} = 1$. On the other hand, we have

$$
\psi_{n, \widetilde{J_n^{(2)}}} \geq (1-\beta) \frac{a_{n, J_n^{(1)}}}{K}
$$
$$
\geq \frac{1-\beta}{K^2},
$$

which concludes the proof. $\qquad\square$

The rest of this subsection is exactly the same to that of T3C. Indeed, with the above lemma, we can show that the set of poorly explored arms U_n^L is empty when n is large enough.

Lemma 18. *Under T3C, there exists $L_0 = Poly(W_1, W_2)$ s.t. $\forall L > L_0$, $U_{\lfloor KL \rfloor}^L = \emptyset$.*

Proof. See proof of Lemma 11 in Appendix C.1. $\qquad\square$

We can finally conclude the proof of Lemma 5 for T3C in the same way as for TTTS in Appendix C.1. $\qquad\blacksquare$

D.2 Concentration of the empirical means, proof of Lemma 6 under T3C

As a corollary of the previous section, we can show the concentration of $\mu_{n,i}$ to μ_i, and the proof remains the same as that of TTTS in Appendix C.2.

D.3 Measurement effort concentration of the optimal arm, proof of Lemma 7 under T3C

Next, we show that the empirical arm draws proportion of the true best arm for T3C concentrates to β when the total number of arm draws is sufficiently large. This proof also remains the same as that of TTTS in Appendix C.3.

D.4 Measurement effort concentration of other arms, proof of Lemma 8 under T3C

In this section, we show that, for T3C, the empirical measurement effort concentration also holds for other arms than the true best arm. Note that this part differs from that of TTTS.

We again establish first an over-allocation implies negligible probability result as follow.

Lemma 19. *Under T3C, for every $\xi \leq \varepsilon_0$ with ε_0 problem dependent, there exists $S_1 = Poly(1/\xi, W_1, W_2)$ such that for all $n > S_1$, for all $i \neq I^\star$,*

$$
\frac{\Psi_{n,i}}{n} \geq \omega_i^\beta + 2\xi \quad \Rightarrow \quad \psi_{n,i} \leq (K-1)\exp\left\{ -\frac{\Delta_{min}^2}{16\sigma^2} \sqrt{\frac{n}{K}} \right\} \,.
$$

Proof. Fix $i \neq I^\star$ s.t. $\Psi_{n,i}/n \geq \omega_i^\beta + 2\xi$, then using Lemma 4, there exists $S_2 = \text{Poly}(1/\xi, W_2)$ such that for any $n > S_2$, we have

$$
\frac{T_{n,i}}{n} \geq \omega_i^\beta + \xi.
$$

Then,

$$\psi_{n,i} \leq \beta a_{n,i} + (1-\beta) \sum_{j \neq i} a_{n,j} \mathbb{1}\{W_n(j,i) = \min_{k \neq j} W_n(j,k)\}$$

$$\leq \beta a_{n,i} + (1-\beta) \left(\sum_{j \neq i, I^\star} a_{n,j} + a_{n,I^\star} \mathbb{1}\{W_n(I^\star, i) = \min_{k \neq I^\star} W_n(I^\star, k)\} \right)$$

$$\leq \sum_{j \neq I^\star} a_{n,j} + \mathbb{1}\{W_n(I^\star, i) = \min_{k \neq I^\star} W_n(I^\star, k)\}.$$

Next we show that the indicator function term in the previous inequality equals to 0.

Using Lemma 3 and Lemma 7 for T3C, there exists $S_3 = \text{Poly}(1/\xi, W_1, W_2)$ such that for any $n > S_3$,

$$\left| \frac{T_{n,I^\star}}{n} - \beta \right| \leq \xi^2 \text{ and } \forall j \in \mathcal{A}, |\mu_{n,j} - \mu_j| \leq \xi^2.$$

Now if $\forall j \neq I^\star, i$, we have $T_{n,j}/n > \omega_j^\beta$, then

$$\frac{n-1}{n} = \sum_{j \in \mathcal{A}} \frac{T_{n,j}}{n}$$

$$= \frac{T_{n,I^\star}}{n} + \frac{T_{n,i}}{n} + \sum_{j \neq I^\star, i} \frac{T_{n,j}}{n}$$

$$> \beta - \varepsilon^2 + \omega_i^\beta + \varepsilon + \sum_{j \neq I^\star, i} \omega_j^\beta \geq 1,$$

which is a contradiction.

Thus there exists at least one $j_0 \neq I^\star, i$, such that $T_{n,j_0}/n \leq \omega_j^\beta$. Assuming $n > \max(S_2, S_3)$, we have

$$W_n(I^\star, i) - W_n(I^\star, j_0) = \frac{(\mu_{n,I^\star} - \mu_{n,i})^2}{2\sigma^2 \left(\frac{1}{T_{n,I^\star}} + \frac{1}{T_{n,i}} \right)} - \frac{(\mu_{n,I^\star} - \mu_{n,j_0})^2}{2\sigma^2 \left(\frac{1}{T_{n,I^\star}} + \frac{1}{T_{n,j_0}} \right)}$$

$$\geq \underbrace{\frac{(\mu_{I^\star} - \mu_i - 2\xi^2)^2}{2\sigma^2 \left(\frac{1}{\beta - \xi^2} + \frac{1}{\omega_i^\beta + \xi} \right)} - \frac{(\mu_{I^\star} - \mu_{j_0} + 2\xi^2)^2}{2\sigma^2 \left(\frac{1}{\beta + \xi^2} + \frac{1}{\omega_{j_0}^\beta} \right)}}_{W_{i,j_0}^\xi}.$$

According to Proposition 1, W_{i,j_0}^ξ converges to 0 when ξ goes to 0, more precisely we have

$$W_{i,j_0}^\xi = \frac{(\mu_{I^\star} - \mu_i)^2}{2\sigma^2} \left(\frac{\beta}{\beta + \omega_i^\beta} \right)^2 \xi + O(\xi^2),$$

thus there exists a ε_0 such that for all $\xi < \varepsilon_0$ it holds for all $i, j_0 \neq I^\star$, $W_{i,j_0}^\xi > 0$. It follows then

$$W_n(I^\star, i) - \min_{k \neq I^\star} W_n(I^\star, k) \geq W_n(I^\star, i) - W_n(I^\star, j_0) > 0,$$

and $\mathbb{1}\{W_n(I^\star, i) = \min_{k \neq I^\star} W_n(I^\star, k)\} = 0$.

Knowing that Lemma 12 is also valid for T3C, thus there exists $M_1 = \text{Poly}(4/\Delta_{\min}, W_1, W_2)$ such that for all $n > M_1$,

$$\forall j \neq I^\star, a_{n,j} \leq \exp\left\{ -\frac{\Delta_{\min}^2}{16\sigma^2} \sqrt{\frac{n}{K}} \right\},$$

which then concludes the proof by taking $S_1 \triangleq \max(M_1, S_2, S_3)$. □

The rest of this subsection almost coincides with that of TTTS. We first show that, starting from some known moment, no arm is overly allocated. More precisely, we show the following lemma.

Lemma 20. *Under T3C, for every ξ, there exists $S_4 = Poly(1/\xi, W_1, W_2)$ s.t. $\forall n > S_4$,*

$$\forall i \in \mathcal{A}, \quad \frac{\Psi_{n,i}}{n} \leq \omega_i^\beta + 2\xi.$$

Proof. See proof of Lemma 15 in Appendix C.4. Note that the previous step does not match exactly that of TTTS, so the proof would be slightly different. However, the difference is only a matter of constant, we thus still choose to skip this proof. □

It remains to prove Lemma 8 for T3C, which stays the same as that of TTTS.

Proof of Lemma 8 for T3C See proof of Lemma 8 for TTTS in Appendix C.4.

■

E Proof of Lemma 1

Finally, it remains to prove Lemma 1 under the Gaussian case before we can conclude for Theorem 1 for TTTS or T3C.

Lemma 1. *Let $\delta, \beta \in (0,1)$. For any sampling rule which satisfies $\mathbb{E}\left[T_\beta^\varepsilon\right] < \infty$ for all $\varepsilon > 0$, we have*

$$\limsup_{\delta \to 0} \frac{\mathbb{E}\left[\tau_\delta\right]}{\log(1/\delta)} \leq \frac{1}{\Gamma_\beta^\star},$$

if the sampling rule is coupled with stopping rule (3),

For the clarity, we recall the definition of generalized likelihood ratio. For any pair of arms i, j, We first define a weighted average of their empirical means,

$$\widehat{\mu}_{n,i,j} \triangleq \frac{T_{n,i}}{T_{n,i} + T_{n,j}}\widehat{\mu}_{n,i} + \frac{T_{n,j}}{T_{n,i} + T_{n,j}}\widehat{\mu}_{n,j}.$$

And if $\widehat{\mu}_{n,i} \geq \widehat{\mu}_{n,j}$, then the generalized likelihood ratio $Z_{n,i,j}$ for Gaussian noise distributions has the following analytic expression,

$$Z_{n,i,j} \triangleq T_{n,i}d(\widehat{\mu}_{n,i}; \widehat{\mu}_{n,i,j}) + T_{n,j}d(\widehat{\mu}_{n,j}; \widehat{\mu}_{n,i,j}).$$

We further define a statistic Z_n as

$$Z_n \triangleq \max_{i \in \mathcal{A}} \min_{j \in \mathcal{A}\backslash\{i\}} Z_{n,i,j}.$$

The following lemma stated by Qin et al. (2017) is needed in our proof.

Lemma 21. *For any $\zeta > 0$, there exists ε s.t. $\forall n \geq T_\beta^\varepsilon$, $Z_n \geq (\Gamma_\beta^\star - \zeta)n$.*

To prove Lemma 1, we need the Gaussian tail inequality (7) of Lemma 2.

Proof. We know that

$$1 - a_{n,I^\star} = \sum_{i \neq I^\star} a_{n,i}$$
$$\leq \sum_{i \neq I^\star} \Pi_n \left[\theta_i > \theta_{I^\star}\right]$$
$$= \sum_{i \neq I^\star} \Pi_n \left[\theta_i - \theta_{I^\star} > 0\right]$$
$$\leq (K-1)\max_{i \neq I^\star} \Pi_n \left[\theta_i - \theta_{I^\star} > 0\right].$$

We can further rewrite $\Pi_n \left[\theta_i - \theta_{I^\star} > 0 \right]$ as

$$\Pi_n \left[\theta_i - \theta_{I^\star} > \mu_{n,i} - \mu_{n,I^\star} + \mu_{n,I^\star} - \mu_{n,i} \right].$$

We choose ε sufficiently small such that the empirical best arm $I_n^\star = I^\star$. Then, for all $n \geq T_\beta^n$ and for any $i \neq I^\star$, $\mu_{n,I^\star} \geq \mu_{n,i}$. Thus, fix any $\zeta \in (0, \Gamma_\beta^\star/2)$ and apply inequality (7) of Lemma 2 with $\mu_{n,I^\star}$ and $\mu_{n,i}$, we have for any $n \geq T_\beta^\varepsilon$,

$$
\begin{aligned}
1 - a_{n,I^\star} &\leq (K-1) \max_{i \neq I^\star} \frac{1}{2} \exp\left\{ -\frac{(\mu_{n,I^\star} - \mu_{n,i})^2}{2\sigma_{n,i,I^\star}^2} \right\} \\
&= \frac{(K-1)\exp\{-Z_n\}}{2} \\
&\leq \frac{(K-1)\exp\left\{-(\Gamma_\beta^\star - \zeta)n\right\}}{2}.
\end{aligned}
$$

The last inequality is deduced from Lemma 21. By consequence,

$$\forall n \geq T_\beta^\varepsilon, \ln\left(1 - a_{n,I^\star}\right) \leq \ln \frac{K-1}{2} - (\Gamma_\beta^\star - \zeta)n.$$

On the other hand, we have for any n,

$$1 - c_{n,\delta} = \frac{\delta}{2n(K-1)\sqrt{2\pi e}\exp\left\{\sqrt{2\ln \frac{2n(K-1)}{\delta}}\right\}}.$$

Thus, there exists a deterministic time N s.t. $\forall n \geq N$,

$$
\begin{aligned}
\ln\left(1 - c_{n,\delta}\right) &= \ln \frac{\delta}{(K-1)\sqrt{8\pi e}} - \ln n - \sqrt{2\ln\frac{2n(K-1)}{\delta}} \\
&\geq \ln \frac{\delta}{2(K-1)\sqrt{2\pi e}} - \zeta n.
\end{aligned}
$$

Let $C_3 \triangleq (K-1)^2\sqrt{2\pi e}$, we have for any $n \geq N_0 \triangleq T_\beta^\varepsilon + N$,

$$\ln\left(1 - a_{n,I^\star}\right) - \ln\left(1 - c_{n,\delta}\right) \leq \ln \frac{C_3}{\delta} - (\Gamma_\beta^\star - 2\zeta)n, \tag{12}$$

and it is clear that $\mathbb{E}\left[N_0\right] < \infty$.

Let us consider the following two cases:

Case 1 There exists $n \in [1, N_0]$ s.t. $a_{n,I^\star} \geq c_{n,\delta}$, then by definition,

$$\tau_\delta \leq n \leq N_1.$$

Case 2 For any $n \in [1, N_0]$, we have $a_{n,I^\star} < c_{n,\delta}$, then $\tau_\delta \geq N_0 + 1$, thus by Equation 12,

$$
\begin{aligned}
0 &\leq \ln\left(1 - a_{\tau_\delta-1,I^\star}\right) - \ln\left(1 - c_{\tau_\delta-1,\delta}\right) \\
&\leq \ln \frac{C_3}{\delta} - (\Gamma_\beta^\star - 2\zeta)(\tau_\delta - 1),
\end{aligned}
$$

and we obtain

$$\tau_\delta \leq \frac{\ln(C_3/\delta)}{\Gamma_\beta^\star - 2\zeta} + 1.$$

Combining the two cases, and we have for any $\zeta \in (0, \Gamma_\beta^\star/2)$,

$$\tau_\delta \leq \max \left\{ N_0, \frac{\ln(C_3/\delta)}{\Gamma_\beta^\star - 2\zeta} + 1 \right\}$$

$$\leq N_0 + 1 + \frac{\ln(C_3)}{\Gamma_\beta^\star - 2\zeta} + \frac{\ln(1/\delta)}{\Gamma_\beta^\star - 2\zeta}.$$

Since $\mathbb{E}[N_1] < \infty$, therefore

$$\limsup_\delta \frac{\mathbb{E}[\tau_\delta]}{\log(1/\delta)} \leq \frac{1}{\Gamma_\beta^\star - 2\zeta}, \forall \zeta \in (0, \Gamma_\beta^\star/2),$$

which concludes the proof. $\qquad\square$

F Technical Lemmas

The whole fixed-confidence analysis for the two sampling rules are both substantially based on two lemmas: Lemma 5 of Qin et al. (2017) and Lemma 4. We prove Lemma 4 in this section.

Lemma 4. *There exists a random variable W_2, such that for all $i \in \mathcal{A}$,*

$$\forall n \in \mathbb{N}, |T_{n,i} - \Psi_{n,i}| \leq W_2 \sqrt{(n+1)\log(e^2 + n)} \ a.s.,$$

and $\mathbb{E}\left[e^{\lambda W_2}\right] < \infty$ for any $\lambda > 0$.

Proof. The proof shares some similarities with that of Lemma 6 of Qin et al. (2017). For any arm $i \in \mathcal{A}$, define $\forall n \in \mathbb{N}$,

$$D_n \triangleq T_{n,i} - \Psi_{n,i},$$

$$d_n \triangleq \mathbb{1}\{I_n = i\} - \psi_{n,i}.$$

It is clear that $D_n = \sum_{l=1}^{n-1} d_l$ and $\mathbb{E}[d_n|\mathcal{F}_{n-1}] = 0$. Indeed,

$$\begin{aligned}
\mathbb{E}[d_n|\mathcal{F}_{n-1}] &= \mathbb{E}[\mathbb{1}\{I_n = i\} - \psi_{n,i}|\mathcal{F}_{n-1}] \\
&= \mathbb{P}[I_n = i|\mathcal{F}_{n-1}] - \mathbb{E}[\mathbb{P}[I_n = i|\mathcal{F}_{n-1}]|\mathcal{F}_{n-1}] \\
&= \mathbb{P}[I_n = i|\mathcal{F}_{n-1}] - \mathbb{P}[I_n = i|\mathcal{F}_{n-1}] = 0.
\end{aligned}$$

The second last equality holds since $\mathbb{P}[I_n = i|\mathcal{F}_{n-1}]$ is $\mathcal{F}_{n-1}$-measurable. Thus D_n is a martingale, whose increment are 1 sub-Gaussian as $d_n \in [-1, 1]$ for all n.

Applying Corollary 8 of Abbasi-Yadkori et al. (2012)[6], it holds that, with probability larger than $1 - \delta$, for all n,

$$|D_n| \leq \sqrt{2(1+n)\ln\left(\frac{\sqrt{1+n}}{\delta}\right)}$$

which yields the first statement of Lemma 4.

We now introduce the random variable

$$W_2 \triangleq \max_{n \in \mathbb{N}} \max_{i \in \mathcal{A}} \frac{|T_{n,i} - \Psi_{n,i}|}{\sqrt{(n+1)\ln(e^2 + n)}}.$$

Applying the previous inequality with $\delta = e^{-x^2/2}$ yields

$$\mathbb{P}\left[\exists n \in \mathbb{N}^\star : |D_n| > \sqrt{(1+n)(\ln(1+n) + x^2)}\right] \leq e^{-x^2/2},$$

$$\mathbb{P}\left[\exists n \in \mathbb{N}^\star : |D_n| > \sqrt{(1+n)\ln(e^2 + n)x^2}\right] \leq e^{-x^2/2},$$

---

[6]but we could actually use several deviation inequalities that hold uniformly over time for martingales with sub-Gaussian increments

where the last inequality uses that for all $a, b \geq 2$, we have $ab \geq a + b$.

Consequently $\forall x \geq 2$, for all $i \in \mathcal{A}$

$$\mathbb{P}\left[\max_{n \in \mathbb{N}} \frac{|T_{n,i} - \Psi_{n,i}|}{\sqrt{(n+1)\log(e^2 + n)}} \geq x\right] \leq e^{-x^2/2}.$$

Now taking a union bound over $i \in \mathcal{A}$, we have $\forall x \geq 2$,

$$
\begin{aligned}
\mathbb{P}\left[W_2 \geq x\right] &\leq \mathbb{P}\left[\max_{i \in \mathcal{A}} \max_{n \in \mathbb{N}} \frac{|T_{n,i} - \Psi_{n,i}|}{(n+1)\log\left(\sqrt{e^2 + n}\right)} \geq x\right] \\
&\leq \mathbb{P}\left[\bigcup_{i \in \mathcal{A}} \max_{n \in \mathbb{N}} \frac{|T_{n,i} - \Psi_{n,i}|}{(n+1)\log\left(\sqrt{e^2 + n}\right)} \geq x\right] \\
&\leq \sum_{i \in \mathcal{A}} \mathbb{P}\left[\max_{n \in \mathbb{N}} \frac{|T_{n,i} - \Psi_{n,i}|}{(n+1)\log\left(\sqrt{e^2 + n}\right)} \geq x\right] \\
&\leq K e^{-x^2/2}.
\end{aligned}
$$

The previous inequalities imply that $\forall i \in \mathcal{A}$ and $\forall n \in \mathbb{N}$, we have $|T_{n,i} - \Psi_{n,i}| \leq W_2 \sqrt{(n+1)\log(e^2 + n)}$ almost surely. Now it remains to show that $\forall \lambda > 0, \mathbb{E}\left[e^{\lambda W_2}\right] < \infty$. Fix some $\lambda > 0$.

$$
\begin{aligned}
\mathbb{E}\left[e^{\lambda W}\right] &= \int_{x=1}^{\infty} \mathbb{P}\left[e^{\lambda W} \geq x\right] \mathrm{d}x = \int_{y=0}^{\infty} \mathbb{P}\left[e^{\lambda W} \geq e^{2\lambda y}\right] 2\lambda e^{2\lambda y} \, \mathrm{d}y \\
&= 2\lambda \int_{y=0}^{2} \mathbb{P}\left[W \geq 2y\right] e^{2\lambda y} \, \mathrm{d}y + 2\lambda \int_{y=2}^{\infty} \mathbb{P}\left[W \geq 2y\right] e^{2\lambda y} \, \mathrm{d}y \\
&\leq \underbrace{2\lambda \int_{y=0}^{2} \mathbb{P}\left[W \geq 2y\right] e^{2\lambda y} \, \mathrm{d}y}_{=e^{4\lambda - 1}} + \underbrace{2\lambda C_1 \int_{y=2}^{\infty} e^{-y^2/2} e^{2\lambda y} \, \mathrm{d}y}_{<\infty} < \infty,
\end{aligned}
$$

where C_1 is some constant.

$\square$

G Proof of Posterior Convergence for the Gaussian Bandit

G.1 Proof of Theorem 4

Theorem 4. *Under TTTS, for Gaussian bandits with improper Gaussian priors, it holds almost surely that*

$$\lim_{n \to \infty} -\frac{1}{n}\log(1 - a_{n,I^\star}) = \Gamma_\beta^\star.$$

From Theorem 2 in Qin et al. (2017), any allocation rule satisfying $T_{n,i}/n \to \omega_i^\beta$ for each $i \in \mathcal{A}$, satisfies

$$\lim_{n \to \infty} -\frac{1}{n}\log(1 - a_{n,I^\star}) = \Gamma_\beta^\star.$$

Therefore, to prove Theorem 4, it is sufficient to prove that under TTTS,

$$\forall i \in \{1, \ldots, K\}, \quad \lim_{n \to \infty} \frac{T_{n,i}}{n} \overset{a.s}{=} \omega_i^\beta. \tag{13}$$

Due to the concentration result in Lemma 4 that we restate below (and proved in Appendix C), which will be useful at several places in the proof, observe that

$$\lim_{n \to \infty} \frac{T_{n,i}}{n} \overset{a.s}{=} \omega_i^\beta \quad \Leftrightarrow \quad \lim_{n \to \infty} \frac{\Psi_{n,i}}{n} \overset{a.s}{=} \omega_i^\beta,$$

therefore it suffices to establish the convergence of $\overline{\psi}_{n,i} = \Psi_{n,i}/n$ to ω_i^β, which we do next. For that purpose, we need again the following maximality inequality lemma.

Lemma 4. *There exists a random variable W_2, such that for all $i \in \mathcal{A}$,*

$$\forall n \in \mathbb{N}, |T_{n,i} - \Psi_{n,i}| \leq W_2 \sqrt{(n+1)\log(e^2 + n)} \ a.s.,$$

and $\mathbb{E}\left[e^{\lambda W_2}\right] < \infty$ for any $\lambda > 0$.

Step 1: TTTS draws all arms infinitely often and satisfies $T_{n,I^\star}/n \to \beta$. More precisely, we prove the following lemma.

Lemma 22. *Under TTTS, it holds almost surely that*

1. *for all $i \in \mathcal{A}$, $\lim_{n \to \infty} T_{n,i} = \infty$.*

2. *$a_{n,I^\star} \to 1$.*

3. *$T_{n,I^\star}/n \to \beta$.*

Proof. Our first ingredient is a lemma showing the implications of finite measurement, and consistency when all arms are sampled infinitely often. Its proof follows standard posterior concentration arguments and is given in Appendix G.2.

Lemma 23 (Consistency and implications of finite measurement)**.**
Denote with $\overline{\mathcal{I}}$ the arms that are sampled only a finite amount of times:

$$\overline{\mathcal{I}} = \{i \in \{1, \ldots, k\} : \forall n, T_{n,i} < \infty\}.$$

If $\overline{\mathcal{I}}$ is empty, $a_{n,i}$ converges almost surely to 1 when $i = I^\star$ and to 0 when $i \neq I^\star$. If $\overline{\mathcal{I}}$ is non-empty, then for every $i \in \overline{\mathcal{I}}$, we have $\liminf_{n \to \infty} a_{n,i} > 0$ a.s.

First we show that $\sum_{n \in \mathbb{N}} T_{n,j} = \infty$ for each arm j. Suppose otherwise. Let $\overline{\mathcal{I}}$ again be the set of arms to which only finite measurement effort is allocated. Under TTTS, we have

$$\psi_{n,i} = a_{n,i} \left(\beta + (1-\beta) \sum_{j \neq i} \frac{a_{n,j}}{1 - a_{n,j}} \right),$$

so $\psi_{n,i} \geq \beta a_{n,i}$. Therefore, by Lemma 23, if $i \in \overline{\mathcal{I}}$, then $\liminf a_{n,i} > 0$ implies that $\sum_n \psi_{n,i} = \infty$. By Lemma 4, we then must have that $\lim_{n \to \infty} T_{n,i} = \infty$ as well: contradiction. Thus, $\lim_{n \to \infty} T_{n,i} = \infty$ for all i, and we conclude that $a_{n,I^\star} \to 1$, by Lemma 23.

For TTTS with parameter β this implies that $\overline{\psi}_{n,I^\star} \to \beta$, and since we have a bound on $|T_{n,i}/n - \overline{\psi}_{n,i}|$ in Lemma 4, we have $T_{n,I^\star}/n \to \beta$ as well. □

Step 2: Controlling the over-allocation of sub-optimal arms. The convergence of $T_{n,I^\star}/n$ to β leads to following interesting consequence, expressed in Lemma 24: if an arm is sampled more often than its optimal proportion, the posterior probability of this arm to be optimal is reduced compared to that of other sub-optimal arms.

Lemma 24 (Over-allocation implies negligible probability)**.** [7] *Fix any $\xi > 0$ and $j \neq I^\star$. With probability 1, under any allocation rule, if $T_{n,I^\star}/n \to \beta$, there exist $\xi' > 0$ and a sequence ε_n with $\varepsilon_n \to 0$ such that for any $n \in \mathbb{N}$,*

$$\frac{T_{n,j}}{n} \geq \omega_j^\beta + \xi \Rightarrow \frac{a_{n,j}}{\max_{i \neq I^\star} a_{n,i}} \leq e^{-n(\xi' + \varepsilon_n)}.$$

Proof. We have $\Pi_n(\Theta_{\cup i \neq I^\star}) = \sum_{i \neq I^\star} a_{n,i} = 1 - a_{n,I^\star}$, therefore $\max_{i \neq I^\star} a_{n,i} \leq 1 - a_{n,I^\star}$. By Theorem 2 of Qin et al. (2017) we have, as $T_{n,I^\star}/n \to \beta$,

$$\limsup_{n \to \infty} -\frac{1}{n} \log \left(\max_{i \neq I^\star} a_{n,i} \right) \leq \Gamma_\beta^\star.$$

---

[7]analogue of Lemma 13 of Russo (2016)

We also have the following from the standard Gaussian tail inequality, for $n \geq \tau$ after which $\mu_{n,I^\star} \geq \mu_{n,i}$, using that $\theta_i - \theta_{I^\star} \sim \mathcal{N}(\mu_{n,i} - \mu_{n,I^\star}, \sigma_{n,i}^2 + \sigma_{n,I^\star}^2)$ and $\sigma_{n,i}^2 + \sigma_{n,I^\star}^2 = \sigma^2(1/T_{n,i} + 1/T_{n,I^\star})$,

$$a_{n,i} \leq \Pi_n(\theta_i \geq \theta_{I^\star}) \leq \exp\left(\frac{-(\mu_{n,i} - \mu_{n,I^\star})^2}{2\sigma^2(1/T_{n,I^\star} + 1/T_{n,i})}\right) = \exp\left(-n\frac{(\mu_{n,i} - \mu_{n,1})^2}{2\sigma^2(n/T_{n,I^\star} + n/T_{n,i})}\right).$$

Thus, there exists a sequence $\varepsilon_n \to 0$, for which

$$\frac{a_{n,j}}{\max_{i \neq I^\star} a_{n,i}} \leq \frac{\exp\left\{-n\left(\frac{(\mu_{n,j} - \mu_{n,I^\star})^2}{2\sigma^2(n/T_{n,I^\star} + n/T_{n,j})} - \varepsilon_n/2\right)\right\}}{\exp\left\{-n\left(\Gamma_\beta^\star + \varepsilon_n/2\right)\right\})} = \exp\left\{-n\left(\frac{(\mu_{n,j} - \mu_{n,I^\star})^2}{2\sigma^2(n/T_{n,I^\star} + n/T_{n,j})} - \Gamma_\beta^\star - \varepsilon_n\right)\right\}.$$

Now we take a look at the two terms in the middle:

$$\frac{(\mu_{n,j} - \mu_{n,I^\star})^2}{2\sigma^2(n/T_{n,I^\star} + n/T_{n,j})} - \Gamma_\beta^\star.$$

Note that the first term is increasing in $T_{n,j}/n$. We have the definition from Qin et al. (2017), for any $j \neq I^\star$,

$$\Gamma_\beta^\star = \frac{(\mu_j - \mu_{I^\star})^2}{2\sigma^2\left(1/\omega_{I^\star}^\beta + 1/\omega_j^\beta\right)},$$

and we have the premise

$$\frac{T_{n,j}}{n} \geq \omega_j^\beta + \xi.$$

Combining these with the convergence of the empirical means to the true means (consistency, see Lemma 23), we can conclude that for all $\varepsilon > 0$, there exists a time n_0 such that for all later times $n \geq n_0$, we have

$$\frac{(\mu_{n,j} - \mu_{n,I^\star})^2}{2\sigma^2(n/T_{n,I^\star} + n/T_{n,j})} \geq \frac{(\mu_j - \mu_{I^\star})^2}{2\sigma^2\left(1/\beta + n/T_{n,j}\right)} - \varepsilon \geq \frac{(\mu_j - \mu_{I^\star})^2}{2\sigma^2\left(1/\beta + 1/(\omega_j^\beta + \xi)\right)} - \varepsilon > \Gamma_\beta^\star,$$

where the first inequality follows from consistency, the second from monotonicity in $T_{n,j}/n$. That means that there exist a $\xi' > 0$ such that

$$\frac{(\mu_{n,j} - \mu_{n,I^\star})^2}{2\sigma^2(n/T_{n,I^\star} + n/T_{n,j})} - \Gamma_\beta^\star > \xi',$$

and thus the claim follows that when $\frac{T_{n,j}}{n} \geq \omega_j^\beta + \xi$, we have

$$\frac{a_{n,j}}{\max_{i \neq I^\star} a_{n,i}} \leq \exp\left\{-n\left(\frac{(\mu_{n,j} - \mu_{n,I^\star})^2}{2\sigma^2(n/T_{n,I^\star} + n/T_{n,j})} - \Gamma_\beta^\star - \varepsilon_n\right)\right\} \leq e^{-n(\xi' + \varepsilon_n)}.$$

$\square$

Step 3: $\overline{\psi}_{n,i}$ converges to ω_i^β for all arms. To establish the convergence of the allocation effort of all arms, we rely on the same sufficient condition used in the analysis of Russo (2016), that we recall below.

Lemma 25 (Sufficient condition for optimality). [8] *Consider any adaptive allocation rule. If we have*

$$\overline{\psi}_{n,I^\star} \to \beta, \quad and \quad \sum_{n \in \mathbb{N}} \psi_{n,j} \mathbf{1}\left\{\overline{\psi}_{n,j} \geq \omega_j^\beta + \xi\right\} < \infty, \quad \forall j \neq I^\star, \xi > 0, \tag{14}$$

then $\overline{\psi}_n \to \psi^\beta$.

---

[8]Lemma 12 of Russo (2016)

First, note that from Lemma 22 we know that $T_{n,I^\star}/n \to \beta$, an by Lemma 4 this implies $\overline{\psi}_{n,I^\star} \to \beta$, hence we can use Lemma 25 to prove convergence to the optimal proportions. Thus, we now show that (14) holds under `TTTS`. Recall that $J_n^{(1)} = \arg\max_j a_{n,j}$ and $J_n^{(2)} = \arg\max_{j \neq J_n^{(1)}} a_{n,j}$. Since $a_{n,I^\star} \to 1$ by Lemma 22, there is some finite time τ after which for all $n > \tau$, $J_n^{(1)} = I^\star$. Under `TTTS`,

$$\psi_{n,i} = a_{n,i} \left(\beta + (1-\beta) \sum_{j \neq i} \frac{a_{n,j}}{1 - a_{n,j}} \right) \leq a_{n,i}\beta + a_{n,i}(1-\beta) \frac{\sum_{j \neq i} a_{n,j}}{1 - a_{n,J_n^{(1)}}} \leq a_{n,i}\beta + a_{n,i}(1-\beta) \frac{\sum_{j \neq i} a_{n,j}}{a_{n,J_n^{(2)}}}$$

$$\leq a_{n,i}\beta + a_{n,i}(1-\beta) \frac{1}{a_{n,J_n^{(2)}}} \leq \frac{a_{n,i}}{a_{n,J_n^{(2)}}},$$

where we use the fact that for $j \neq J_n^{(1)}$, we have $a_{n,J_n^{(1)}} \geq a_{n,j}$ and $a_{n,J_n^{(2)}} \leq 1 - a_{n,J_n^{(1)}}$. For $n \geq \tau$ this means that $\psi_{n,i} \leq a_{n,i}/\max_{j \neq I^\star} a_{n,i}$ for any $i \neq I^\star$.

By Lemma 24, there is a constant $\xi' > 0$ such and a sequence $\varepsilon_n \to 0$ such that

$$T_{n,i}/n \geq w_i^\beta + \xi \Rightarrow \frac{a_{n,i}}{\max_{j \neq I^\star} a_{n,j}} \leq e^{-n(\xi' - \varepsilon_n)}.$$

Now take a time τ large enough, such that for $n \geq \tau$ we have $|T_{n,j}/n - \overline{\psi}_{n,j}| \leq \xi$ (which can be found by Lemma 4). Then we have

$$\mathbb{1}\left\{ \overline{\psi}_{n,j} \geq \psi_j^\beta + \xi \right\} \leq \mathbb{1}\left\{ \frac{T_{n,j}}{n} \geq \omega_j^\beta + 2\xi \right\}$$

Therefore, for all $i \neq I^\star$, we have

$$\sum_{n \geq \tau} \psi_{n,i} \mathbb{1}\left\{ \overline{\psi}_{n,j} \geq \psi_j^\beta + \xi \right\} \leq \sum_{n \geq \tau} \psi_{n,i} \mathbb{1}\left\{ \frac{T_{n,j}}{n} \geq \omega_j^\beta + 2\xi \right\} \leq \sum_{n \geq \tau} e^{-n(\xi' - \varepsilon_n)} < \infty.$$

Thus (14) holds and the convergence to the optimal proportions follows by Lemma 25.

G.2 Proof of auxiliary lemmas

Proof of Lemma 23 Let $\overline{\mathcal{I}}$ be nonempty. Define

$$\mu_{\infty,n} \triangleq \lim_{n \to \infty} \mu_{n,i}, \text{ and } \sigma_{\infty,i}^2 \triangleq \lim_{n \to \infty} \sigma_{n,i}^2,$$

and recall that for $i \in \mathcal{A}$ for which $T_{n,i} = 0$, we have $\mu_{n_i} = \mu_{1,i} = 0$ and $\sigma_{n,i}^2 = \sigma_{1,i}^2 = \infty$, and if $T_{n,i} > 0$, we have

$$\mu_{n,i} = \frac{1}{T_{n,i}} \sum_{\ell=1}^{n-1} \mathbb{1}\{I_\ell = i\} Y_{\ell,I_\ell}, \text{ and } \sigma_{n,i}^2 = \frac{\sigma^2}{T_{n,i}}.$$

For all arms that are sampled infinitely often, we therefore have $\mu_{\infty,i} = \mu_i$ and $\sigma_{\infty,i}^2 = 0$. For all arms that are sampled only a finite number of times, i.e. $i \in \overline{\mathcal{I}}$, we have $\sigma_{\infty,i}^2 > 0$, and there exists a time n_0 after which for all $n \geq n_0$ and $i \in \overline{\mathcal{I}}$, we have $T_{n,i} = T_{n_0,i}$. Define

$$\Pi_\infty \triangleq \mathcal{N}(\mu_{\infty,1}, \sigma_{\infty,1}^2) \otimes \mathcal{N}(\mu_{\infty,2}, \sigma_{\infty,2}^2) \otimes \ldots \otimes \mathcal{N}(\mu_{\infty,k}, \sigma_{\infty,k}^2) = \bigotimes_{i \notin \overline{\mathcal{I}}} \delta_{\mu_i} \otimes \bigotimes_{i \in \overline{\mathcal{I}}} \Pi_{n_0}.$$

Then for each $i \in \mathcal{A}$ we define

$$a_{\infty,i} \triangleq \Pi_\infty \left(\theta_i > \max_{j \neq i} \theta_j \right).$$

Then we have for all $i \in \overline{\mathcal{I}}$, $a_{\infty,i} \in (0,1)$, since $\sigma_{\infty,i}^2 > 0$, and thus $a_{\infty,I^\star} < 1$.

When $\overline{\mathcal{I}}$ is empty, we have $a_{n,I^\star} = \Pi_n(\theta_{I^\star} > \max_{i \neq I^\star} \theta_i)$, but since $\Pi_\infty = \bigotimes_{i \in \mathcal{A}} \delta_{\mu_i}$, we have $a_{\infty,I^\star} = 1$ and $a_{\infty,i} = 0$ for all $i \neq I^\star$.

∎

H   Proof of Posterior Convergence for the Bernoulli Bandit

H.1   Preliminaries

We first introduce a crucial Beta tail bound inequality. Let $F_{a,b}^{\text{Beta}}$ denote the cdf of a Beta distribution with parameters a and b, and $F_{c,d}^{\text{B}}$ the cdf of a Binomial distribution with parameters c and d, then we have the following relationship, often called the 'Beta-Binomial trick',

$$F_{a,b}^{\text{Beta}}(y) = 1 - F_{a+b-1,y}^{\text{B}}(a-1),$$

so that we have

$$\mathbb{P}[X \geq x] = \mathbb{P}[B_{a+b-1,x} \leq a-1] = \mathbb{P}[B_{a+b-1,1-x} \geq b].$$

We can bound Binomial tails with Sanov's inequality:

$$\frac{e^{-nd(k/n,x)}}{n+1} \leq \mathbb{P}[B_{n,x} \geq k] \leq e^{-nd(k/n,x)},$$

where the last inequalities hold when $k \geq nx$.

Lemma 26. *Let $X \sim \mathcal{B}eta(a,b)$ and $Y \sim \mathcal{B}eta(c,d)$ with $0 < \frac{a-1}{a+b-1} < \frac{c-1}{c+d-1}$. Then we have $\mathbb{P}[X > Y] \leq De^{-C}$ where*

$$C = \inf_{\frac{a-1}{a+b-1} \leq y \leq \frac{c-1}{c+d-1}} C_{a,b}(y) + C_{c,d}(y),$$

and

$$D = 3 + \min\left(C_{a,b}\left(\frac{c-1}{c+d-1}\right), C_{c,d}\left(\frac{a-1}{a+b-1}\right)\right).$$

Note that this lemma is the Bernoulli version of Lemma 2.

Theorem 6. *Consider the Beta-Bernoulli setting. For $\beta \in (0,1)$, under any allocation rule satisfying $T_{n,I^\star}/n \to \omega_{I^\star}^\beta$,*

$$\lim_{n\to\infty} -\frac{1}{n}\log(1 - a_{n,I^\star}) \leq \Gamma_\beta^\star,$$

and under any allocation rule satisfying $T_{n,i}/n \to \omega_i^\beta$ for each $i \in \mathcal{A}$,

$$\lim_{n\to\infty} -\frac{1}{n}\log(1 - a_{n,I^\star}) = \Gamma_\beta^\star.$$

Proof. Denote again with $\overline{\mathcal{I}}$ again the set of arms sampled only finitely many times. For $\overline{\mathcal{I}}$ empty, we thus have $\mu_{\infty,i} \triangleq \lim_{n\to\infty} \mu_{n,i} = \mu_i$. The posterior variance is

$$\sigma_{n,i}^2 = \frac{\alpha_{n,i}\beta_{n,i}}{(\alpha_{n,i}+\beta_{n,i})^2(\alpha_{n,i}+\beta_{n,i}+1)} = \frac{(1 + \sum_{\ell=1}^{n-1}\mathbb{1}\{I_\ell = i\}Y_{\ell,I_\ell})(1 + T_{n,i} - \sum_{\ell=1}^{n-1}\mathbb{1}\{I_\ell = i\}Y_{\ell,I_\ell})}{(2+T_{n,i})^2(2+T_{n,i}+1)}.$$

We see that when $\overline{\mathcal{I}}$ is empty, we have $\sigma_{\infty,i}^2 \triangleq \lim_{n\to\infty} \sigma_{n,i}^2 = 0$, i.e., the posterior is concentrated.

Step 1: A lower bound when some arms are sampled only finitely often.   First, note that when $T_{n,i} = 0$ for some $i \in \mathcal{A}$, the empirical mean for that arm equals the prior mean $\mu_{n,i} = \alpha_{0,i}/(\alpha_{0,i}+\beta_{0,i})$, and the variance is strictly positive: $\sigma_{n,i}^2 = (\alpha_{0,i}\beta_{0,i})/\left((\alpha_{0,i}+\beta_{0,i})^2(\alpha_{0,i}+\beta_{0,i}+1)\right) > 0$. When $\overline{\mathcal{I}}$ is not empty, then for every $i \in \overline{\mathcal{I}}$ we have $\sigma_{\infty,i}^2 > 0$, and $a_{\infty,i} \in (0,1)$, implying $a_{\infty,I^\star} < 1$, and thus

$$\lim_{n\to\infty} -\frac{1}{n}\log(1 - a_{n,I^\star}) = -\frac{1}{n}\log(1 - a_{\infty,I^\star}) = 0.$$

Step 2: A lower bound when every arm is sampled infinitely often. Suppose now that $\overline{\mathcal{I}}$ is empty, then we have

$$\max_{i \neq I^\star} \Pi_n(\theta_i \geq \theta_{I^\star}) \leq 1 - a_{n,I^\star} \leq \sum_{i \neq I^\star} \Pi_n(\theta_i \geq \theta_{I^\star}) \leq (k-1) \max_{i \neq I^\star} \Pi_n(\theta_i \geq \theta_{I^\star}).$$

Thus, we have $1 - a_{n,I^\star} \leq (k-1) \max_{i \neq I^\star} \Pi_n(\theta_i \geq \theta_{I^\star})$ and also $1 - a_{n,I^\star} \doteq \max_{i \neq I^\star} \Pi_n(\theta_i \geq \theta_{I^\star})$. We have

$$\Gamma^\star = \max_{w \in W} \min_{i \neq I^\star} C_i(\omega_{I^\star}, \omega_i),$$

$$\Gamma^\star_\beta = \max_{w \in W; \omega_{I^\star} = \beta} \min_{i \neq I^\star} C_i(\beta, \omega_i), \text{ with}$$

$$C_i(\omega_{I^\star}, \omega_i) = \min_{x \in \mathbb{R}} \omega_{I^\star} d(\theta_{I^\star}; x) + \omega_i d(\theta_i; x) = \omega_{I^\star} d(\theta_{I^\star}; \overline{\theta}) + \omega_i d(\theta_i; \overline{\theta}),$$

where $\overline{\theta} \in [\theta_i, \theta_{I^\star}]$ is the solution to

$$A'(\overline{\theta}) = \frac{\omega_{I^\star} A'(\theta_{I^\star}) + \omega_i A'(\theta_i)}{\omega_{I^\star} + \omega_i}.$$

Since every arm is sampled infinitely often, when n is large, we have $\mu_{n,I^\star} > \mu_{n,i}$. Define $S_{n,i} \triangleq \sum_{\ell=1}^{n-1} \mathbb{1}\{I_\ell = i\} Y_{\ell,I_\ell}$. Recall that the posterior is a Beta distribution with parameters $a_{n,i} = S_{n,i} + 1$ and $\beta_{n,i} = T_{n,i} - S_{n,i} + 1$. Let $\tau \in \mathbb{N}$ be such that for every $n \geq \tau$, we have $S_{n,i}/(T_{n,i} + 1) < S_{n,I^\star}/(T_{n,I^\star} + 1)$. For the sake of simplicity, we define for any $i \in \mathcal{A}$ the interval

$$I_{i,I^\star} \triangleq \left[\frac{S_{n,i}}{T_{n,i} + 1}, \frac{S_{n,I^\star}}{T_{n,I^\star} + 1} \right].$$

Then using Lemma 26 with $a = S_{n,i} + 1, b = T_{n,i} - S_{n,i} + 1, c = S_{n,I^\star} + 1, d = T_{n,I^\star} - S_{n,I^\star} + 1$, we have

$$\Pi_n(\theta_i - \theta_{I^\star} \geq 0) \leq D \exp\left\{ - \inf_{y \in I_{i,I^\star}} C_{S_{n,i}+1, T_{n,i}-S_{n,i}+1}(y) + C_{S_{n,I^\star}+1, T_{n,I^\star}-S_{n,I^\star}+1}(y) \right\}.$$

This implies

$$\frac{1}{n} \log \left(\frac{\Pi_n(\theta_i \geq \theta_{I^\star})}{\exp\left\{ - \inf_{y \in I_{i,I^\star}} C_{S_{n,i}+1, T_{n,i}-S_{n,i}+1}(y) + C_{S_{n,I^\star}+1, T_{n,I^\star}-S_{n,I^\star}+1}(y) \right\}} \right) \leq \frac{1}{n} \log(D),$$

which goes to zero as n goes to infinity. Indeed replacing a, b, c, d by their values in the definition of D we get

$$D \leq 3 + (T_{n,i} - 1) kl \left(\frac{S_{n,i}}{T_{n,i} + 1}; \frac{S_{n,I^\star}}{T_{n,I^\star} + 1} \right)$$

$$\leq 3 + (n+1) kl \left(0; \frac{n}{n+1} \right) = (n+1) \log(n+1).$$

Hence,

$$\Pi_n(\theta_i \geq \theta_{I^\star}) \doteq \exp\left\{ - \inf_{y \in I_{i,I^\star}} C_{S_{n,i}+1, T_{n,i}-S_{n,i}+1}(y) + C_{S_{n,I^\star}+1, T_{n,I^\star}-S_{n,I^\star}+1}(y) \right\}.$$

We thus have for any i,

$$1 - a_{n,i} \doteq \max_{j \neq I^\star} \Pi_n \left[\theta_j \geq \theta_{I^\star} \right]$$

$$\doteq \max_{j \neq I^\star} \exp\left\{ - \inf_{y \in I_{j,I^\star}} C_{S_{n,j}+1, T_{n,j}-S_{n,j}+1}(y) + C_{S_{n,I^\star}+1, T_{n,I^\star}-S_{n,I^\star}+1}(y) \right\}$$

$$\doteq \exp\left\{ -n \min_{j \neq I^\star} \inf_{y \in I_{j,I^\star}} \frac{T_{n,j} + 1}{n} kl \left(\frac{S_{n,j}}{T_{n,j} + 1}; y \right) + \frac{T_{n,I^\star} + 1}{n} kl \left(\frac{S_{n,I^\star}}{T_{n,I^\star} + 1}; y \right) \right\}$$

$$\geq \exp\left\{ -n \max_{\boldsymbol{\omega}} \min_{j \neq I^\star} \inf_{y \in I_{j,I^\star}} \omega_i kl \left(\frac{S_{n,j}}{T_{n,j} + 1}; y \right) + \omega_{I^\star} kl \left(\frac{S_{n,I^\star}}{T_{n,j} + 1}; y \right) \right\}.$$

Fix some $\varepsilon > 0$, then there exists some $n_0(\varepsilon)$ such that for all $n \geq n_0(\varepsilon)$, we have for any j,

$$I_{j,I^\star} = \left[\frac{S_{n,j}}{T_{n,j}+1}, \frac{S_{n,I^\star}}{T_{n,I^\star}+1}, \right] \subset [\mu_j + \varepsilon, \mu_{I^\star} - \varepsilon] \triangleq I^\star_{j,\varepsilon},$$

and because KL-divergence is uniformly continuous on the compact interval $I^\star_{j,\varepsilon}$, there exists an n_1 such that for every $n \geq n_1$ we have

$$kl\left(\frac{S_{n,j}}{T_{n,j}+1}; y \right) \geq (1-\varepsilon)kl\left(\mu_j; y \right),$$

for any y and for all $j \in \mathcal{A}$. Therefore, we have

$$1 - a_{n,i} \doteq \exp\left\{ -n \max_{\boldsymbol{\omega}} \min_{j \neq I^\star} \inf_{y \in I_{j,I^\star}} \omega_j kl\left(\frac{S_{n,j}}{T_{n,j}+1}; y \right) + \omega_{I^\star} kl\left(\frac{S_{n,I^\star}}{T_{n,I^\star}+1}; y \right) \right\}$$

$$\geq \exp\left\{ -n \max_{\boldsymbol{\omega}} \min_{i \neq I^\star} \inf_{y \in I^\star_{j,\varepsilon}} \omega_i kl(\mu_j; y) + \omega_{I^\star} kl(\mu_{I^\star}; y) \right\}.$$

Therefore, we have

$$\limsup_{n \to \infty} -\frac{1}{n} \log(1 - a_{n,i}) \leq \Gamma^\star.$$

If $T_{n,i}/n \to \omega_i^\star$ for each $i \in \mathcal{A}$, we have

$$\lim_{n \to \infty} \inf_{y \in I_{i,I^\star}} \frac{T_{n,i}+1}{n} kl\left(\frac{S_{n,i}}{T_{n,i}+1}; y \right) + \frac{T_{n,I^\star}+1}{n} kl\left(\frac{S_{n,I^\star}}{T_{n,i}+1}; y \right)$$

$$= \inf_{y \in [\mu_i, \mu_{I^\star}]} \omega_i^\star kl(\mu_i; y) + \omega_{I^\star}^\star kl(\mu_{I^\star}; y)$$

$$= \Gamma^\star,$$

and thus

$$1 - a_{n,i} \doteq \exp\left\{ -n \max_{\boldsymbol{\omega}} \min_{j \neq I^\star} \inf_{y \in I_\varepsilon^\star} \omega_i kl(\mu_j; y) + \omega_{I^\star} kl(\mu_{I^\star}; y) \right\}$$

$$\doteq \exp\left\{ -n\Gamma^\star \right\},$$

implying

$$\lim_{n \to \infty} -\frac{1}{n} \log\left(1 - a_{n,i} \right) = \Gamma^\star.$$

Everything goes similarly when $\omega_{I^\star} = \beta \in (0,1)$, so under any sampling rule satisfying $T_{n,I^\star}/n \to \beta$ we have

$$\limsup_{n \to \infty} -\frac{1}{n} \log(1 - a_{n,i}) \leq \Gamma_\beta^\star$$

and under any sampling rule satisfying $T_{n,i}/n \to \omega_i^\beta$ for each $i \in \mathcal{A}$, we have

$$\lim_{n \to \infty} -\frac{1}{n} \log(1 - a_{n,i}) = \Gamma_\beta^\star.$$

$\square$

H.2   Proof of Theorem 5

Theorem 5. *Under TTTS, for Bernoulli bandits and uniform priors, it holds almost surely that*

$$\lim_{n \to \infty} -\frac{1}{n} \log(1 - a_{n,I^\star}) = \Gamma_\beta^\star.$$

From Theorem 6 we know that under any allocation rule satisfying $T_{n,i}/n \to \omega_i^\beta$ for every $i \in \mathcal{A}$, we have

$$\lim_{n \to \infty} -\frac{1}{n} \log\left(1 - a_{n,I^\star}\right) = \Gamma_\beta^\star.$$

Thus, we only need to prove that under TTTS, for all $i \in \mathcal{A}$, we have

$$\lim_{n \to \infty} \frac{T_{n,i}}{n} \overset{a.s}{=} \omega_i^\beta.$$

Just as for the proof of the Gaussian case, we can use Lemma 4 (proof in Appendix G.2), which implies

$$\lim_{n \to \infty} \frac{T_{n,i}}{n} \overset{a.s}{=} \omega_i^\beta \quad \Leftrightarrow \quad \lim_{n \to \infty} \frac{\Psi_{n,i}}{n} \overset{a.s}{=} \omega_i^\beta.$$

Therefore, it suffices to show convergence for $\overline{\psi}_{n,i} = \Psi_{n,i}/n$ to ω_i^β, which we will do next, following the same steps as in the proof for the Gaussian case.

Step 1: TTTS draws all arms infinitely often and satisfies $T_{n,I^\star}/n \to \beta$. We prove the following lemma.

Lemma 27. *Under TTTS, it holds almost surely that*

1. *for all $i \in \mathcal{A}$, $\lim_{n \to \infty} T_{n,i} = \infty$.*

2. *$a_{n,I^\star} \to 1$.*

3. *$\frac{T_{n,I^\star}}{n} \to \beta$.*

Proof. First, we give a lemma showing the implications of finite measurement, and consistency when all arms are sampled infinitely often, which provides a proof for 2. The proof of this lemma follows from the proof of Theorem 6, and is given in Appendix H.3.

Lemma 28 (Consistency and implications of finite measurement)**.**
Denote with $\overline{\mathcal{I}}$ the arms that are sampled only a finite amount of times:

$$\overline{\mathcal{I}} = \{i \in \{1, \ldots, k\} : \forall n, T_{n,i} < \infty\}.$$

If $\overline{\mathcal{I}}$ is empty, $a_{n,i}$ converges almost surely to 1 when $i = I^\star$ and to 0 when $i \neq I^\star$. If $\overline{\mathcal{I}}$ is non-empty, then for every $i \in \overline{\mathcal{I}}$, we have $\liminf_{n \to \infty} a_{n,i} > 0$ a.s.

Now we can show 1. of Lemma 27: we show that under TTTS, for each $j \in A$, we have $\sum_{n \in \mathbb{N}} T_{n,j} = \infty$. The proof is exactly equal to the proof for Gaussian arms.

Under TTTS, we have

$$\psi_{n,i} = a_{n,i} \left(\beta + (1-\beta) \sum_{j \neq i} \frac{a_{n,j}}{1 - a_{n,j}} \right),$$

so $\psi_{n,i} \geq \beta a_{n,i}$, therefore, by Lemma 23, if $i \in \overline{\mathcal{I}}$, then $\liminf a_{n,i} > 0$ implies that $\sum_n \psi_{n,i} = \infty$. By Lemma 4, we then must have that $\lim_{n \to \infty} T_{n,i} = \infty$ as well: contradiction. Thus, $\lim_{n \to \infty} T_{n,i} = \infty$ for all i, and we conclude that $a_{n,I^\star} \to 1$, by Lemma 23.

Lastly we prove point 3. of Lemma 27. For TTTS with parameter β, the above implies that $\overline{\psi}_{n,I^\star} \to \beta$, and since we have a bound on $|T_{n,i}/n - \overline{\psi}_{n,i}|$ in Lemma 4, we have $T_{n,I^\star}/n \to \beta$ as well.

$\square$

Step 2: Controlling the over-allocation of sub-optimal arms. Following the proof for the Gaussian case again, we can establish a consequence of the convergence of $T_{n,I^\star}/n$ to β : if an arm is sampled more often than its optimal proportion, the posterior probability of this arm to be optimal is reduced compared to that of other sub-optimal arms. We can prove this by using ingredients from the proof of the lower bound in Theorem 6.

Lemma 29 (Over-allocation implies negligible probability). [9]
Fix any $\xi > 0$ and $j \neq I^\star$. With probability 1, under any allocation rule, if $T_{n,I^\star}/n \to \beta$, there exist $\xi' > 0$ and a sequence ε_n with $\varepsilon_n \to 0$ such that for any $n \in \mathbb{N}$,

$$\frac{T_{n,j}}{n} \geq \omega_j^\beta + \xi \implies \frac{a_{n,j}}{\max_{i \neq I^\star} a_{n,i}} \leq e^{-n(\xi' + \varepsilon_n)}.$$

Proof. By Theorem 6, we have, as $T_{n,I^\star}/n \to \beta$,

$$\limsup_{n \to \infty} -\frac{1}{n} \log\left(\max_{i \neq I^\star} a_{n,i}\right) \leq \Gamma_\beta^\star,$$

since $\max_{i \neq I^\star} a_{n,i} \leq 1 - a_{n,I^\star}$. We also have from Lemma 26 a deviation inequality, so that we can establish the following logarithmic equivalence:

$$a_{n,j} \leq \Pi_n(\theta_j \geq \theta_{I^\star}) \doteq \exp\left\{-nC_j\left(w_{n,I^\star}, \omega_{n,j}\right)\right\} \doteq \exp\left\{-nC_j\left(\beta, \omega_{n,j}\right)\right\},$$

where we denote $\omega_{n,j} \triangleq \frac{T_{n,j}}{n}$. We can combine these results, which implies that there exists a non-negative sequence $\varepsilon_n \to 0$ such that

$$\frac{a_{n,j}}{\max_{i \neq I^\star} a_{n,i}} \leq \frac{\exp\left\{-nC_j\left(\beta, \omega_{n,j}\right) - \varepsilon_n/2\right\}}{\exp\left\{-n(\Gamma_\beta^\star + \varepsilon/2)\right\}} = \exp\left\{-n\left(C_j\left(\beta, \omega_{n,j}\right) - \Gamma_\beta^\star\right) - \varepsilon_n\right\}.$$

We know that $C_j\left(\beta, \omega_j^\beta\right)$ is strictly increasing in ω_j^β, and $C_j\left(\beta, \omega_j^\beta\right) = \Gamma_\beta^\star$, thus, there exists some $\xi' > 0$ such that

$$\omega_{n,j} \geq \omega_j^\beta + \xi \implies C_j\left(\beta, \omega_{n,j}\right) - \Gamma_\beta^\star > \xi'.$$

$\square$

Step 3: $\overline{\psi}_{n,i}$ converges to ω_i^β for all arms. To establish the convergence of the allocation effort of all arms, we rely on the same sufficient condition used in the analysis of Russo (2016), restated above in Lemma 25, and we will restate it here again for convenience.

Lemma 30 (Sufficient condition for optimality).
Consider any adaptive allocation rule. If

$$\overline{\psi}_{n,I^\star} \to \beta, \quad and \quad \sum_{n \in \mathbb{N}} \psi_{n,j} \mathbf{1}\left\{\overline{\psi}_{n,j} \geq \omega_j^\beta + \xi\right\} < \infty, \; \forall j \neq I^\star, \xi > 0, \tag{15}$$

then $\overline{\psi}_n \to \psi^\beta$.

First, note that from Lemma 27 we know that $\frac{T_{n,I^\star}}{n} \to \beta$, and by Lemma 4 this implies $\overline{\psi}_{n,I^\star} \to \beta$, hence we can use the lemma above to prove convergence to the optimal proportions. This proof is already given in Step 3 of the proof for the Gaussian case, and since it does not depend on the specifics of the Gaussian case, except for invoking Lemma 23 (consistency), which for the Bernoulli case we replace by Lemma 28, it gives a proof for the Bernoulli case as well. We conclude that (14) holds, and the convergence to the optimal proportions follows by Lemma 25.

---

[9]analogue of Lemma 13 of Russo (2016)

H.3   Proof of auxiliary lemmas

Proof of Lemma 26

$$\mathbb{P}\left[X > Y\right] = \mathbb{E}\left[\mathbb{P}\left[X > Y|Y\right]\right] \le \mathbb{E}\left[\mathbb{1}\{Y < \frac{a-1}{a+b-1}\} + \mathbb{1}\{Y \ge \frac{a-1}{a+b-1}\}\mathbb{P}\left[X > Y|Y\right]\right]$$

$$\le \exp\left\{-(c+d-1)kl\left(\frac{c-1}{c+d-1};\frac{a-1}{a+b-1}\right)\right\}$$

$$+ \mathbb{E}\left[\exp\left\{-(a+b-1)kl\left(\frac{a-1}{a+b-1};Y\right)\right\}\mathbb{1}\{Y \ge \frac{a-1}{a+b-1}\}\right],$$

Using the Beta-Binomial trick in the second inequality. Then we have (call the second half A)

$$A \le \mathbb{E}\left[\mathbb{1}\{\frac{a-1}{a+b-1} \le Y \le \frac{c-1}{c+d-1}\}\right]\exp\left\{-(a+b-1)kl\left(\frac{a-1}{a+b-1};Y\right)\right\}$$

$$+ \exp\left\{-(a+b-1)kl\left(\frac{a-1}{a+b-1};\frac{c-1}{c+d-1}\right)\right\}$$

(call the first half B). Denote with f the density of Y, then

$$B = \int_{\frac{a-1}{a+b-1}}^{\frac{c-1}{c+d-1}} \exp\left\{-(a+b-1)kl\left(\frac{a-1}{a+b-1};y\right)\right\}f(y)\,\mathrm{d}y.$$

Via integration by parts we obtain

$$B = \left[\exp\left\{-(a+b-1)kl\left(\frac{a-1}{a+b-1};y\right)\right\}\mathbb{P}\left[Y \le y\right]\right]_{\frac{a-1}{a+b-1}}^{\frac{c-1}{c+d-1}}$$

$$+ \int_{\frac{a-1}{a+b-1}}^{\frac{c-1}{c+d-1}} (a+b-1)\frac{\mathrm{d}}{\mathrm{d}y}kl\left(\frac{a-1}{a+b-1};y\right)\exp\left\{-C_{a,b}(y)\right\}P(Y \le y)\,\mathrm{d}y$$

$$\le \int_{\frac{a-1}{a+b-1}}^{\frac{c-1}{c+d-1}} (a+b-1)\frac{\mathrm{d}}{\mathrm{d}y}kl\left(\frac{a-1}{a+b-1};y\right)\exp\left\{-(C_{a,b}(y) + C_{c,d}(y))\right\}\,\mathrm{d}y$$

$$+ \exp\left\{-(a+b-1)kl\left(\frac{a-1}{a+b-1};\frac{c-1}{c+d-1}\right)\right\},$$

where the first inequality uses the Binomial trick again. Let

$$C = \inf_{\frac{a-1}{a+b-1} \le y \le \frac{c-1}{c+d-1}} (a+b-1)kl\left(\frac{a-1}{a+b-1};y\right) + (c+d-1)kl\left(\frac{c-1}{c+d-1};y\right) = \inf_{\frac{a-1}{a+b-1} \le y \le \frac{c-1}{c+d-1}} C_{a,b}(y) + C_{c,d}(y),$$

then note that in particular we have

$$C \le \min\left((a+b-1)kl\left(\frac{a-1}{a+b-1};\frac{c-1}{c+d-1}\right), (c+d-1)kl\left(\frac{c-1}{c+d-1};\frac{a-1}{a+b-1}\right)\right)$$

$$= \min\left(C_{a,b}\left(\frac{c-1}{c+d-1}\right), C_{c,d}\left(\frac{a-1}{a+b-1}\right)\right).$$

Then

$$B \le e^{-C}\int_{\frac{a-1}{a+b-1}}^{\frac{c-1}{c+d-1}} (a+b-1)\frac{\mathrm{d}}{\mathrm{d}y}kl(\frac{a-1}{a+b-1};y)\,\mathrm{d}y + e^{-C} = \left[(a+b-1)kl\left(\frac{a-1}{a+b-1};\frac{c-1}{c+d-1}\right) + 1\right]e^{-C}.$$

Thus we have

$$\mathbb{P}\left[X > Y\right] \le \left(3 + (a+b-1)kl\left(\frac{a-1}{a+b-1};\frac{c-1}{c+d-1}\right)\right)e^{-C}.$$

By symmetry, we have

$$\mathbb{P}\left[X > Y\right] \leq \left(3 + \min\left(C_{a,b}\left(\frac{c-1}{c+d-1}\right), C_{c,d}\left(\frac{a-1}{a+b-1}\right)\right)\right)e^{-C},$$

where

$$C = \inf_{\frac{a-1}{a+b-1} \leq y \leq \frac{c-1}{c+d-1}} (a+b-1)kl\left(\frac{a-1}{a+b-1}; y\right) + (c+d-1)kl\left(\frac{c-1}{c+d-1}; y\right).$$

∎

Proof of Lemma 28 Let $\overline{\mathcal{I}}$ be empty, then we have $\mu_{\infty,i} \triangleq \lim_{n\to\infty} \mu_{n,i} = \mu_i$. The posterior variance is

$$\sigma_{n,i}^2 = \frac{\alpha_{n,i}\beta_{n,i}}{(\alpha_{n,i}+\beta_{n,i})^2(\alpha_{n,i}+\beta_{n,i}+1)} = \frac{(1+\sum_{\ell=1}^{n-1}\mathbb{1}\{I_\ell=i\}Y_{\ell,I_\ell})(1+T_{n,i}-\sum_{\ell=1}^{n-1}\mathbb{1}\{I_\ell=i\}Y_{\ell,I_\ell})}{(2+T_{n,i})^2(2+T_{n,i}+1)},$$

We see that when $\overline{\mathcal{I}}$ is empty, we have $\sigma_{\infty,i}^2 \triangleq \lim_{n\to\infty} \sigma_{n,i}^2 = 0$, i.e., the posterior is concentrated.

When $T_{n,i} = 0$ for some $i \in \mathcal{A}$, the empirical mean for that arm equals to the prior mean $\mu_{n,i} = \alpha_{1,i}/(\alpha_{1,i}+\beta_{1,i})$, and the variance is strictly positive: $\sigma_{n,i}^2 = (\alpha_{n,i}\beta_{n,i})/\left((\alpha_{1,i}+\beta_{1,i})^2(\alpha_{1,i}+\beta_{1,i}+1)\right) > 0$. When $\overline{\mathcal{I}}$ is not empty, then for every $i \in \overline{\mathcal{I}}$ we have $\sigma_{\infty,i}^2 > 0$, and $\alpha_{\infty,i} \in (0,1)$, implying $\alpha_{\infty,I^\star} < 1$, hence the posterior is not concentrated.

∎