Constrained Best Arm Identification with Fixed Confidence

Dual Degree Project Stage II Report

submitted in partial fulfillment of the requirements

for the degree of

Bachelor of Technology in Electrical Engineering and
Master of Technology in Communications and Signal Processing

by

Kartik Nair Variar

(Roll No: 18D070051)

Supervisor:

Prof. Jayakrishnan Nair

Department of Electrical Engineering

Indian Institute of Technology Bombay

June 2023

Dissertation Approval

The dissertation entitled

Constrained Best Arm Identification with Fixed Confidence

by

Kartik Nair Variar

is approved for the degree of

Bachelor of Technology in Electrical Engineering and Master of Technology in Communications and Signal Processing

Prof. Jayakrishnan Nair
Department of Electrical Engineering
(Supervisor)

Prof. D. Manjunath
Department of Electrical Engineering
(Chairperson and Examiner)

Prof. Prasanna Chaporkar
Department of Electrical Engineering
(Examiner)

Date: June 2023
Place: Mumbai.

Declaration

I declare that this written submission represents my ideas in my own words and where other ideas or words or diagrams have been included from books/ papers/electronic media, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will result in disciplinary action by the institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken.

Kartik Nair Variar

(Roll No:18D070051)

June 2023

Abstract

In this thesis, we state the multi armed-bandit problem with constraints. Additionally, we define the best-arm identification problem in the fixed confidence setting for constrained multi-armed bandits. We derive a general lower bound on the expected stopping time for a *sound* algorithm in the constrained MAB setting and propose an Action Elimination based algorithm as well as a Track-and-Stop algorithm for this setting. We derive upper bounds on the expected stopping time of these algorithms and perform experiments to compare their performance.
Keywords: constrained multi-armed bandits, best arm identification, action elimination, confidence bounds, track-and-stop, optimal proportions.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

1.1 Multi-armed Bandit Problem

A multi-armed bandit problem is a sequential game consisting of K probability distributions $\nu_1, \ldots, \nu_K$ such that, at every time step $t = 1, 2, \ldots$, an arm $A_t \in \mathcal{A} = \{1, \ldots, K\}$ is chosen. The environment then samples an *independent* reward X_t from the distribution ν_{A_t}. The **best arm** is defined as the arm that optimizes a pre-specified attribute (e.g., the mean rewards, μ_k).

The above framework is not directly applicable to problems where more than one attribute is is taken into consideration. Examples of such problems can be in,

- **Clinical trials:** Maximize the efficacy while limiting the side effects.
- **Portfolio optimization:** Maximize mean returns subject to a risk appetite.

We propose a modification to the above framework so as to incorporate multiple attributes. The approach taken is to *optimize one attribute subject to constraints on others.* A general constrained MAB problem is stated in this chapter, however, the problem will be made more specific in subsequent chapters, based on the algorithm.

1.2 Best Arm Identification with Fixed Confidence

The best arm identification problem requires us to identify an arm that optimizes the pre-specified attribute as fast and accurately as possible, irrespec-

tive of the number of *bad* arm pulls. A policy (π) is defined by a *sampling rule*, a *stopping rule*, with T as the stopping time, and a *decision rule*, $\hat{a}_T$. The goal is to guarantee that $\hat{a}_T$ belongs to the set of *best arms*, with the highest possible probability while minimizing the number of pulls T.

In the fixed confidence setting, a *fixed confidence*, $\delta \in (0, 1)$ is provided and we are required to obtain a best arm/an optimal arm, with probability at least $1 - \delta$, using the least possible number of pulls, i.e., guarantee that $\mathbb{P}(\hat{a}_T \notin B) \leq \delta$, where B is the set of optimal arms (such a strategy is called δ-PAC) while minimizing $\mathbb{E}[T]$.

Definition. *A triple* $(\pi, T, \hat{a}_T)$ *is* ***δ-sound*** *for an environment class* $\mathcal{E}$ *if* $\forall \nu \in \mathcal{E}$,

$$\mathbb{P}_{\nu\pi}(T < \infty,\ \hat{a}_T \notin B) \leq \delta$$

We will now, extend the above definitions to the constraint bandit setting and define best arm identification with fixed confidence in this setting.

1.3 Constrained Multi-armed Bandit Problem

Consider an MAB problem with K *arms*, labelled $1, 2, \ldots K$. Let $\mathcal{A} = \{1, \ldots, K\}$. Each arm is associated with a, possibly multi-dimensional, probability distribution ν_k corresponding to arm $k \in \mathcal{A}$. Suppose $\nu_k \in \mathcal{C}$, the space of possible arm distributions, define m attributes $g_0, g_1, \ldots, g_m$ such that one of the attributes is an *objective* (WLOG, say g_0) and the rest are constraints. The user specifies a bound on each of the constraint attributes, let us denote this by $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_m)$. Thus, a constrained MAB instance is defined by the pair $(\nu, \boldsymbol{\tau})$

Definition. *The* ***feasibility criterion*** $(\boldsymbol{\phi_a})$ *of an arm a is a Boolean such that, $\phi_a = 0$ if the attributes corresponding to arm a satisfy the user specified constraints. Such an arm is called a feasible arm. Otherwise, $\phi_a = 1$ and such an arm is called an infeasible arm.*

Definition. *The* ***feasibility criterion*** $(\boldsymbol{\theta})$ *of an instance $(\nu, \boldsymbol{\tau})$ is a Boolean such that, $\theta = 0$ if there exists an arm satisfying the user specified constrains. Such an instance is called a feasible instance. Otherwise, $\theta = 1$ and such an instance is called an infeasible instance.*

Definition. *An optimal arm as is defined as follows,*

- ***Feasible Instance:*** *A feasible arm $a^* \in \mathcal{A}$ that optimizes the objective attribute g_0*

- ***Infeasible Instance:*** *It's definition depends on the problem statement.*

Given a constrained MAB instance $(\nu, \boldsymbol{\tau})$ the objective of an algorithm for best arm identification in such a setting is to identify (a^*, θ^*), where θ^* is the feasibility criterion if the instance, as fast and accurately as possible, irrespective of the number of sub-optimal pulls. A policy π is defined by a *sampling rule*, a *stopping rule* with stopping time T and a *decision rule* $(\hat{a}_T, \hat{\theta}_T)$.

In the fixed confidence setting, a *fixed confidence*, $\delta \in (0, 1)$ is provided and we are required to obtain an optimal arm and the instance feasibility criterion, with probability at least $1 - \delta$, using the least possible number of pulls, i.e., guarantee that $\mathbb{P}(\hat{a}_T \notin B, \hat{\theta}_T \neq \theta^*) \leq \delta$, where B is the set of optimal arms while minimizing $\mathbb{E}[T]$.

Definition. *A triple* $(\pi, T, (\hat{a}_T, \hat{\theta}_T))$ *is* ***δ-sound*** *for an environment class* $\mathcal{E}$ *if for all* $\nu \in \mathcal{E}$*,*

$$\mathbb{P}_{\nu\pi}(T < \infty, \hat{a}_T \notin B \text{ or } \hat{\theta}_T \neq \theta^*) \leq \delta$$

We will now look at **Constrained Action Elimination** (Con-AE) and a version of the **Track-and-stop** algorithm from multi-armed bandit problems with a single constraint.

Note: The problem formulations for the two algorithms are different, as they were worked on at different times. A more general problem formulation can be formulated to neatly encompass both the algorithms. This exercise has not been taken in this thesis.

Part I

Action Elimination

Chapter 2

Problem Formulation for Action Elimination

In this chapter, we shall define our constrained stochastic MAB problem for the action elimination algorithm. To keep the exposition simple, we consider a single constraint (i.e., m = 1). Formally,

- Consider an MAB problem with K *arms*, labelled $1, 2, \ldots$. Each arm is associated with a, possibly multi-dimensional, probability distribution $\nu(k)$ corresponding to arm $k \in [K]$. Suppose $\nu(k) \in \mathcal{C}$, the space of possible arm distributions, define

  Definition. *Objective attribute function* - $g_0 : \mathcal{C} \to \mathbb{R}$

  Definition. *Constraint attribute function* - $g_1 : \mathcal{C} \to \mathbb{R}$

- The user provides a threshold $\tau \in \mathbb{R}$, which specifies the upper bound on attribute g_1.

- An *instance* of the constrained MAB problem is defined by (ν, τ), where $\nu = \{\nu(k) : k \in [K]\}$.

- The arms that satisfy $g_1(\nu(.)) \leq \tau$ are called *feasible arms*. The rest are called *infeasible arms*. The set of feasible arms is denoted by $\mathcal{K}(\nu)$.

- An instance (ν, τ) is said to be feasible if $\mathcal{K}(\nu) \neq \phi$ and is said to be infeasible if $\mathcal{K}(\nu) = \phi$.

2.1 Feasible Instance

- An *optimal arm* is defined as the arm that minimizes $g_0(\nu(.))$, subject to the constraint that $g_1(\nu(.)) \leq \tau$.
- Let the optimal value be $g_0^* = \min_{k\in[K]} g_0(\nu(k))$. Arms with $g_0(\nu(.))$ larger than g_0^* are called sub-optimal arms.
- Infeasible arms with g_0 smaller that g_0^* are called *deceiver arms*. The set of all deceiver arms is denoted by $\mathcal{K}^d(\nu) = \{k \in [K] : g_1(\nu(k)) \geq \tau, g_0(\nu(k)) \leq g_0^*\}$.

2.2 Infeasible Instance

- The best among all infeasible arms is defined as the arm with the lowest constraint attribute. It is denoted by k_{min} and the lowest constraint attribute is denoted by $g_{1,min} = \min_{k\in[K]} g_1(\nu(k))$.

2.3 Concentration Inequalities on Attribute Estimators

Suppose that for $i \in \{0, 1\}$ and distribution $F \in \mathcal{C}$, $\exists$ an estimator $\hat{g}_i^t(F)$ of $g_i(F)$ using t iid samples of F, satisfying the following concentration inequality,

$$\exists a_i : \ \forall \epsilon > 0, \ \mathbb{P}(|\hat{g}_i^t(F) - g_i(F)| \geq \epsilon) \leq 2\exp(-a_i t\epsilon^2)$$

For simplicity, for each arm k, we denote $g_i(\nu(k))$ as $g_{i,k}$ and $\hat{g}_i^t(\nu(k))$ as $\hat{g}_{i,k}^t$.

Given a *fixed confidence* δ, we are required to find the optimal arm in the feasible instance, or the best among infeasible arms in the infeasible instance with a probability $\geq 1 - \delta$ in minimum number of trails/pulls. Formally,

- **Feasible Instance:** Obtain k^* with probability $\geq 1 - \delta$ in minimum possible number of trials.
- **Infeasible Instance:** Obtain k_{min} with probability $\geq 1 - \delta$ in minimum possible number of trials.

In **Part II**, we will look at a different problem formulation for the *Track-and-stop* algorithm

Chapter 3

Constrained Action Elimination Algorithm

3.1 Preliminaries

In an action elimination based algorithm, the arms are eliminated using the lower and upper confidence bounds of the attributes. Based on our problem formulation, we can define the confidence bounds as follows:

Definition. *The confidence bound for arm k and attribute i after t pulls is given by*

$$\alpha_{i,k}^t = \sqrt{\frac{1}{a_i t} log\left(\frac{cKt^2}{\delta}\right)}$$

Where $c > \frac{2\pi^2}{3}$ is a constant. Using the above, we define the UCB for the attribute i and arm k after t pulls as

$$U_{i,k}^t = \hat{g}_{i,k}^t + \alpha_{i,k}^t$$

and the LCB for the attribute i and arm k after t pulls as

$$L_{i,k}^t = \hat{g}_{i,k}^t - \alpha_{i,k}^t$$

We shall show that the true means $g_{i,k}$ lie between the UCB and LCB with a high probability governed by the fixed confidence δ. Let us define the events that the true means lie in between the UCB and LCB as follows:

Definition. *Define $X_{i,k}^t,\ \forall (i,k,t) \in \{0,1\} \times [K] \times \mathbb{Z}_+$ as*

$$X_{i,k}^t \equiv L_{i,k}^t \leq g_{i,k} \leq U_{i,k}^t \iff |\hat{g}_{i,k}^t - g_{i,k}| \leq \alpha_{i,k}^t.$$

Lemma 1. *The events $X_{i,k}^t$ occur $\forall (i,k,t) \in \{0,1\} \times [K] \times \mathbb{Z}_+$ with probability at least $1-\delta$. That is,*

$$\mathbb{P}\left(\bigcap_{i=1}^{2}\bigcap_{k=1}^{K}\bigcap_{t=1}^{\infty} X_{i,k}^t\right) \geq 1-\delta$$

Proof. We shall show that

$$\mathbb{P}\left(\bigcup_{i=1}^{2}\bigcup_{k=1}^{K}\bigcup_{t=1}^{\infty} \neg X_{i,k}^t\right) < \delta$$

Applying union bounds on the LHS,

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{2}\bigcup_{k=1}^{K}\bigcup_{t=1}^{\infty} \neg X_{i,k}^t\right) &\leq \sum_{i=1}^{2}\sum_{k=1}^{K}\sum_{t=1}^{\infty} \mathbb{P}\left(\neg X_{i,k}^t\right) \\
&= \sum_{i=1}^{2}\sum_{k=1}^{K}\sum_{t=1}^{\infty} \mathbb{P}\left(|\hat{g}_{i,k}^t - g_{i,k}| > \alpha_{i,k}^t\right) \\
&\leq \sum_{i=1}^{2}\sum_{k=1}^{K}\sum_{t=1}^{\infty} 2\exp\left(-a_i t(\alpha_{i,k}^t)^2\right) \\
&= \sum_{i=1}^{2}\sum_{k=1}^{K}\sum_{t=1}^{\infty} 2\exp\left(log(\frac{\delta}{cKt^2})\right) \\
&= \sum_{i=1}^{2}\sum_{k=1}^{K}\sum_{t=1}^{\infty} 2\frac{\delta}{cKt^2} = \frac{4\delta}{c}\frac{\pi^2}{6} < \delta
\end{aligned}$$

□

Therefore, from **lemma 1**, we observe that

$$|\hat{g}_{i,k}^t - g_{i,k}| \leq \alpha_{i,k}^t,\ \forall (i,k,t) \in \{0,1\} \times [K] \times \mathbb{Z}_+$$

with probability at least $1-\delta$ and $L_{i,k}^t$ and $U_{i,k}^t$ are the lower and upper confidence bounds of the attributes $g_{i,k},\ \forall (i,k,t) \in \{0,1\} \times [K] \times \mathbb{Z}_+$.

In the next section, we shall look at the algorithm description of the constrained Action Elimination algorithm.

3.2 Algorithm Description

In this section, we present an algorithm for constrained best arm identification, assuming that we have estimators for each attribute that satisfy the concentration inequality.

The algorithm, which we refer to as *constrained action elimination* (Con-AE) algorithm, is based on the successive elimination algorithm of Even-Dar et al.(2006). For a feasible instance, the algorithm identifies the optimal arm (feasible arm with lowest g_0). For an infeasible instance, the algorithm identifies the arm with the highest feasibility. Con-AE follows the following steps: (For simplicity, we shall consider the case of unique optimal arm)

1. Define Ω_n as the set of *active arms*, arms that are yet to be eliminated.
2. Sample all active arms r_n number of times. For simplicity, let $r_n = 1 \forall n \in \mathbb{Z}_+$
3. Define $\hat{K}_{F,n}$ as the set of active arms that are guaranteed to be feasible with probability $\geq 1 - \delta$.
4. Eliminate those arms whose LCB of optimizer (g_0) is greater than the minimum UCB of g_0 among all guaranteed feasible arms.

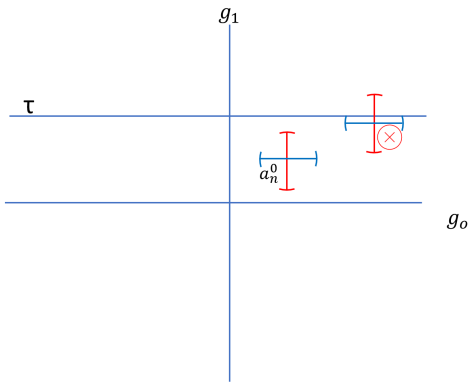
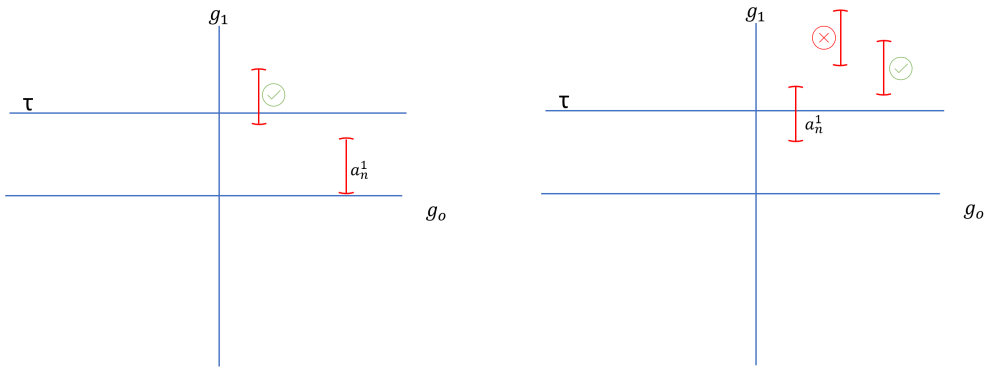


Figure 3.1: $L_{0,k}^t \geq \hat{U}_0^t$ case

5. Eliminate those arms whose LCB of constraint (g_1) is greater than the minimum UCB of g_1 among all active arms, given that their LCB is greater than the threshold τ.



(a) $L_{1,k}^t \geq \hat{U}_1^t$ but $L_{1,k}^t \leq \tau$ (b) $L_{1,k}^t \geq \hat{U}_1^t$ and $L_{1,k}^t \geq \tau$

Figure 3.2: $L_{1,k}^t \geq \hat{U}_1^t$ and $L_{1,k}^t \geq \tau$ case

6. If $|\Omega_n| > 1$ go back to step 2. The algorithm terminates when $|\Omega_n| = 1$.

The constrained action elimination algorithm has been written in pseudo code form below.

Algorithm 1 Con-AE

procedure CON-AE(K, τ, δ)
  $\Omega_n \leftarrow [K]$
  while $|\Omega_n| > 1$ **do**
    sample arms indexed by Ω_n, r_n times

    set $\hat{\mathcal{K}}_{F,n} \leftarrow \{k \in \Omega_n : U_{1,k}^t \leq \tau\}$

    set $E_1 \leftarrow \{k \in \Omega_n : L_{1,k}^t \geq \tau \ and \ L_{1,k}^t \geq \hat{U}_1^t\}$

    set $E_2 \leftarrow \{k \in \Omega_n : L_{0,k}^t \geq \hat{U}_0^t\}$

    $\Omega_n \leftarrow \Omega_n \setminus (E_1 \cup E_2)$

    where $\hat{U}_0^t = \min_{k \in \hat{K}_{F,n}}(U_{0,k}^t)$ and $\hat{U}_1^t = \min_{k \in \Omega_n}(U_{1,k}^t)$
  end while
end procedure

3.3 Algorithm Analysis

Definition. *Given $\alpha_{i,k}^t$ and $\Delta \in \mathbb{R}$. Define $f_{c,\delta}(\Delta, a_i)$ as the minimum time (t) required for $\Delta \geq \alpha_{i,k}^t$. (This definition works $\forall k \in [K]$ as the $\alpha_{i,k}^t$s are equal $\forall k \in [K]$)*

Even-Dar et al.(2006) showed that,

$$f_{c,\delta}(\Delta, a_i) = O\left(\frac{2}{a_i \Delta^2} log\left(\sqrt{\frac{2}{a_i}}\frac{K}{\Delta\delta}\right)\right)$$

. when $\Delta > 0$ and 0 otherwise.

We shall use the above definition to derive our upper bound on the stopping time for the constrained action elimination algorithm.

Definition. *Define constraint gap (infeasiblity gap) as*

$$\Delta_k^{Con} = g_{1,k} - \tau$$

Δ_k^{Con} can be positive or negative based on the value of $g_{1,k}$. Define $\Delta_{min}^{Con} = g_{1,min} - \tau$ as the minimum value of the constraint gap achieved by arm k_{min}. ($\Delta_{min}^{Con} < 0$ for a feasible instance)

Definition. *Let k^* be the optimal arm and let $g_0^* = g_{0,k^*}$. Define the sub-optimality gap as*

$$\Delta_k = g_{0,k} - g_0^*$$

$\Delta_k \geq 0$ for all feasible arms. There may exist deceiver arms for which $\Delta_k < 0$.

3.3.1 Feasible Instance

Theorem 1. *Given that the instance is feasible, with probability $\geq 1 - \delta$, the Con-AE algorithm terminates, outputting the optimal arm in at most T rounds given by,*

$$T = \sum_{k \in [K] \setminus k^*} \min\left(\max\left(f_{c,\delta}\left(\frac{\Delta_k^{Con}}{2}, a_1\right), f_{c,\delta}\left(\frac{\Delta_k^{Con} - \Delta_{min}^{Con}}{4}, a_1\right)\right), \max\left(f_{c,\delta}\left(\frac{-\Delta_{k^*}^{Con}}{2}, a_1\right), f_{c,\delta}\left(\frac{\Delta_k}{4}, a_0\right)\right)\right)$$

Assuming that $\nexists$ any $k \in [K]$ such that $g_{1,k} = \tau$

Proof. Consider the Con-AE algorithm as stated above. WLOG, let $r_n = 1 \forall n \in \mathbb{Z}_+$. Thus, at the end of the n^{th} epoch, each active arm ($k \in \Omega_n$) has been pulled $t = n$ number of times. From, Lemma 1, we know that,

$$|\hat{g}^t_{i,k} - g_{i,k}| \leq \alpha^t_{i,k},\ \forall (i,k,t) \in \{0,1\} \times [K] \times \mathbb{Z}_+$$

with probability at least $1 - \delta$.

Part 1 *Termination with optimal arm*
Let $k^* \in \Omega_n$ with a^0_n and a^1_n be the arms corresponding to $\hat{U}^t_0$ and $\hat{U}^t_1$ respectively,

1. $L^t_{1,k^*} \leq \tau,\ \forall t \in \mathbb{Z}_+$ by definition of optimal arm k^* in the feasible instance.

2. $L^t_{0,k^*} \leq U^t_{0,a^0_n} \implies \hat{g}^t_{0,k^*} - \hat{g}^t_{0,a^0_n} \leq 2\alpha^t_{0,k^*}$ *or* $2\alpha^t_{0,a^0_n}$, since the $\alpha^t_{i,k}$s are independent of k. Observe that,

$$\begin{aligned}\hat{g}^t_{0,k^*} - \hat{g}^t_{0,a^0_n} &= \hat{g}^t_{0,k^*} - g^*_0 - \hat{g}^t_{0,a^0_n} + g_{0,a^0_n} - \Delta_{a^0_n} \\ &\leq 2\alpha^t_{0,k^*}\end{aligned}$$

   Hence, k^* satisfies the above condition too.

Therefore, $k^* \in \Omega_{n+1}$. As $k^* \in [K]$, by induction $k^* \in \Omega_n\ \forall n \geq 1$. Thus we conclude that if the algorithm terminates, it outputs the optimal arm.

Part 2 *Bounding the time required for termination*
An arm k will be eliminated if

1. $L^t_{1,k} \geq \tau$ and $L^t_{1,k} \geq U^t_{1,a^1_n}$. Thus, the time required will be maximum of the time required for $L^t_{1,k} \geq \tau$ and $L^t_{1,k} \geq U^t_{1,a^1_n}$.

   - *Time required for* $L^t_{1,k} \geq \tau$
     Observe that $L^t_{1,k} \geq \tau \implies \hat{g}^t_{1,k} - \tau \geq \alpha^t_{1,k}$. Taking the LHS,

$$\hat{g}^t_{1,k} - \tau = \hat{g}^t_{1,k} - g_{1,k} + \Delta^{Con}_k \geq -\alpha^t_{1,k} + \Delta^{Con}_k$$

     With high probability, the time required above will be less than the time when $-\alpha^t_{1,k} + \Delta^{Con}_k \geq \alpha^t_{1,k} \implies \frac{\Delta^{Con}_k}{2} \geq \alpha^t_{1,k}$. This is only possible when $\Delta^{Con}_k \geq 0$, i.e., arm k is infeasible. Thus, from our definition, the minimum time required will be $f_{c,\delta}\left(\frac{\Delta^{Con}_k}{2}, a_1\right)$.

- *Time required for* $L^t_{1,k} \geq U^t_{1,a^1_n}$
  Observe that $L^t_{1,k} \geq U^t_{1,a^1_n} \implies \hat{g}^t_{1,k} - \hat{g}^t_{1,a^1_n} \geq 2\alpha^t_{1,k}$, since the $\alpha^t_{i,k}$ is independent of the arm k. Taking the LHS,

$$\begin{aligned}\hat{g}^t_{1,k} - \hat{g}^t_{1,a^1_n} \geq \hat{g}^t_{1,k} - \hat{g}^t_{1,min} &= \hat{g}^t_{1,k} - g_{1,k} - \hat{g}^t_{1,min} + g_{1,min} + \Delta^{Con}_k - \Delta^{Con}_{min} \\ &\geq -2\alpha^t_{1,k} + \Delta^{Con}_k - \Delta^{Con}_{min}\end{aligned}$$

  With high probability, the time required above will be less than the time when $-2\alpha^t_{1,k} + \Delta^{Con}_k - \Delta^{Con}_{min} \geq 2\alpha^t_{1,k} \implies \frac{\Delta^{Con}_k - \Delta^{Con}_{min}}{4} \geq \alpha^t_{1,k}$. Thus, from our definition, the minimum time required will be $f_{c,\delta}\left(\frac{\Delta^{Con}_k - \Delta^{Con}_{min}}{4}, a_1\right)$.

Combining the above two conditions, we get that the time required above is bounded by $\max\left(f_{c,\delta}\left(\frac{\Delta^{Con}_k}{2}, a_1\right), f_{c,\delta}\left(\frac{\Delta^{Con}_k - \Delta^{Con}_{min}}{4}, a_1\right)\right)$.

2. $\hat{\mathcal{K}}_{F,n} \neq \phi$ and $L^t_{1,k} \geq U^t_{1,a^0_n}$. Thus, the time required will be at most the maximum of the time required for $k^* \in \hat{\mathcal{K}}_{F,n}$ and $L^t_{0,k} \geq U^t_{0,a^0_n}$.

   - *Time required for* $k^* \in \hat{\mathcal{K}}_{F,n}$
     Observe that $k^* \in \hat{\mathcal{K}}_{F,n} \implies U^t_{1,k^*} \leq \tau \implies \hat{g}^t_{1,k^*} - \tau \leq -\alpha^t_{1,k^*}$. Taking the LHS,

$$\hat{g}^t_{1,k^*} - \tau = \hat{g}^t_{1,k^*} - g_{1,k^*} + \Delta^{Con}_{k^*} \leq \alpha^t_{1,k^*} + \Delta^{Con}_{k^*}$$

     With high probability, the time required above will be less than the time when $\alpha^t_{1,k^*} + \Delta^{Con}_{k^*} \leq -\alpha^t_{1,k^*} \implies \frac{-\Delta^{Con}_{k^*}}{2} \geq \alpha^t_{1,k^*}$. Thus, from our definition, the minimum time required will be $f_{c,\delta}\left(\frac{-\Delta^{Con}_{k^*}}{2}, a_1\right)$.

   - *Time required for* $L^t_{0,k} \geq U^t_{0,a^0_n}$
     Observe that $L^t_{0,k} \geq U^t_{0,a^0_n} \implies \hat{g}^t_{0,k} - \hat{g}^t_{0,a^0_n} \geq 2\alpha^t_{0,k}$, since the $\alpha^t_{i,k}$ is independent of the arm k. Taking the LHS,

$$\begin{aligned}\hat{g}^t_{0,k} - \hat{g}^t_{0,a^0_n} \geq \hat{g}^t_{0,k} - \hat{g}^t_{0,k^*} &= \hat{g}^t_{0,k} - g_{0,k} - \hat{g}^t_{0,k^*} + g_{0,k^*} + \Delta_k \\ &\geq -2\alpha^t_{0,k} + \Delta_k\end{aligned}$$

     With high probability, the time required above will be less than the time when $-2\alpha^t_{0,k} + \Delta_k \geq 2\alpha^t_{0,k} \implies \frac{\Delta_k}{4} \geq \alpha^t_{0,k}$. Thus, from our definition, the minimum time required will be $f_{c,\delta}\left(\frac{\Delta_k}{4}, a_0\right)$.

Combining the above two conditions, we get that the time required above is bounded by $\max\left(f_{c,\delta}\left(\frac{-\Delta^{Con}_{k^*}}{2}, a_1\right), f_{c,\delta}\left(\frac{\Delta_k}{4}, a_0\right)\right)$.

Combining the two cases, we get that the time required to eliminate sub-optimal arm k is bounded as follows:

$$\min\left(\max\left(f_{c,\delta}\left(\frac{\Delta_k^{Con}}{2}, a_1\right), f_{c,\delta}\left(\frac{\Delta_k^{Con} - \Delta_{min}^{Con}}{4}, a_1\right)\right), \max\left(f_{c,\delta}\left(\frac{-\Delta_{k^*}^{Con}}{2}, a_1\right), f_{c,\delta}\left(\frac{\Delta_k}{4}, a_0\right)\right)\right)$$

Thus time required to eliminate all sub-optimal arms is bounded by T given as

$$T = \sum_{k\in[K]\setminus k^*} \min\left(\max\left(f_{c,\delta}\left(\frac{\Delta_k^{Con}}{2}, a_1\right), f_{c,\delta}\left(\frac{\Delta_k^{Con} - \Delta_{min}^{Con}}{4}, a_1\right)\right), \max\left(f_{c,\delta}\left(\frac{-\Delta_{k^*}^{Con}}{2}, a_1\right), f_{c,\delta}\left(\frac{\Delta_k}{4}, a_0\right)\right)\right)$$

Hence proved. □

3.3.2 Infeasible Instance

Theorem 2. *Given that the instance is infeasible, with probability $\geq 1-\delta$, the Con-AE algorithm terminates, outputting the best arm among the infeasible arms (k_{min}) in at most T rounds given by,*

$$T = \sum_{k\in[K]\setminus k_{min}} \max\left(f_{c,\delta}\left(\frac{\Delta_k^{Con}}{2}, a_1\right), f_{c,\delta}\left(\frac{\Delta_k^{Con} - \Delta_{min}^{Con}}{4}, a_1\right)\right)$$

Assuming that $\nexists$ any $k \in [K]$ such that $g_{1,k} = \tau$

Proof. Consider the Con-AE algorithm as stated above. WLOG, let $r_n = 1 \forall n \in \mathbb{Z}_+$ and k_{min} be the unique best arm among infesible arm. Thus, at the end of the n^{th} epoch, each active arm ($k \in \Omega_n$) has been pulled $t = n$ number of times. From, Lemma 1, we know that,

$$|\hat{g}_{i,k}^t - g_{i,k}| \leq \alpha_{i,k}^t, \ \forall (i,k,t) \in \{0,1\} \times [K] \times \mathbb{Z}_+$$

with probability at least $1 - \delta$.

Part 1 *Termination with best among infeasible arms*
Let $k_{min} \in \Omega_n$ with a_n^0 and a_n^1 be the arms corresponding to $\hat{U}_0^t$ and $\hat{U}_1^t$ respectively,

1. $L^t_{1,k_{min}} \leq U^t_{1,a^1_n} \implies \hat{g}^t_{1,k_{min}} - \hat{g}^t_{1,a^1_n} \leq 2\alpha^t_{1,k_{min}}$ $or 2\alpha^t_{1,a^1_n}$, since the $\alpha^t_{i,k}$s are independent of k. Observe that,

$$\begin{aligned}\hat{g}^t_{1,k_{min}} - \hat{g}^t_{1,a^1_n} &= \hat{g}^t_{1,k_{min}} - g_{1,min} - \hat{g}^t_{1,a^1_n} + g_{1,a^1_n} - \Delta^{Con}_{a^1_n} + \Delta^{Con}_{min} \\ &\leq 2\alpha^t_{0,k_{min}}\end{aligned}$$

   Hence, k_{min} satisfies the above condition.

2. $L^t_{0,k_{min}} \leq U^t_{0,a^0_n}$ holds $\forall t \in \mathbb{Z}_+$ as $\hat{K}_{F,n} = \phi$. Hence, k_{min} satisfies the above condition too.

Therefore, $k_{min} \in \Omega_{n+1}$. As $k_{min} \in [K]$, by induction $k_{min} \in \Omega_n \ \forall n \geq 1$. Thus we conclude that if the algorithm terminates, it outputs the best among infeasible arm.

Part 2 *Bounding the time required for termination*
An arm k will be eliminated if $L^t_{1,k} \geq \tau$ and $L^t_{1,k} \geq U^t_{1,a^1_n}$. Thus, the time required will be maximum of the time required for $L^t_{1,k} \geq \tau$ and $L^t_{1,k} \geq U^t_{1,a^1_n}$.

- *Time required for* $L^t_{1,k} \geq \tau$
  Observe that $L^t_{1,k} \geq \tau \implies \hat{g}^t_{1,k} - \tau \geq \alpha^t_{1,k}$. Taking the LHS,

$$\hat{g}^t_{1,k} - \tau = \hat{g}^t_{1,k} - g_{1,k} + \Delta^{Con}_k \geq -\alpha^t_{1,k} + \Delta^{Con}_k$$

  With high probability, the time required above will be less than the time when $-\alpha^t_{1,k} + \Delta^{Con}_k \geq \alpha^t_{1,k} \implies \frac{\Delta^{Con}_k}{2} \geq \alpha^t_{1,k}$. This is only possible when $\Delta^{Con}_k \geq 0$, i.e., arm k is infeasible. Thus, from our definition, the minimum time required will be $f_{c,\delta}\left(\frac{\Delta^{Con}_k}{2}, a_1\right)$.

- *Time required for* $L^t_{1,k} \geq U^t_{1,a^1_n}$
  Observe that $L^t_{1,k} \geq U^t_{1,a^1_n} \implies \hat{g}^t_{1,k} - \hat{g}^t_{1,a^1_n} \geq 2\alpha^t_{1,k}$, since the $\alpha^t_{i,k}$ is independent of the arm k. Taking the LHS,

$$\begin{aligned}\hat{g}^t_{1,k} - \hat{g}^t_{1,a^1_n} \geq \hat{g}^t_{1,k} - \hat{g}^t_{1,min} &= \hat{g}^t_{1,k} - g_{1,k} - \hat{g}^t_{1,min} + g_{1,min} + \Delta^{Con}_k - \Delta^{Con}_{min} \\ &\geq -2\alpha^t_{1,k} + \Delta^{Con}_k - \Delta^{Con}_{min}\end{aligned}$$

  With high probability, the time required above will be less than the time when $-2\alpha^t_{1,k} + \Delta^{Con}_k - \Delta^{Con}_{min} \geq 2\alpha^t_{1,k} \implies \frac{\Delta^{Con}_k - \Delta^{Con}_{min}}{4} \geq \alpha^t_{1,k}$. Thus, from our definition, the minimum time required will be $f_{c,\delta}\left(\frac{\Delta^{Con}_k - \Delta^{Con}_{min}}{4}, a_1\right)$.

Combining the above two conditions, we get that the time required above is bounded by
$\max\left(f_{c,\delta}\left(\frac{\Delta_k^{Con}}{2}, a_1\right), f_{c,\delta}\left(\frac{\Delta_k^{Con}-\Delta_{min}^{Con}}{4}, a_1\right)\right)$.

Thus time required to eliminate all remaining arms is bounded by T given as

$$T = \sum_{k\in[K]\setminus k_{min}} \max\left(f_{c,\delta}\left(\frac{\Delta_k^{Con}}{2}, a_1\right), f_{c,\delta}\left(\frac{\Delta_k^{Con}-\Delta_{min}^{Con}}{4}, a_1\right)\right)$$

Hence proved. □

Part II

Track-and-stop

Chapter 4

General Lower Bound

Before stating the algorithm, we will look at a general lower bound on the expected stopping time for a δ-sound algorithm in the constrained multi-armed bandit setting.

Let $\mathcal{E}$ be an set of arbitrary K-armed constrained bandit environments, and $\nu \in \mathcal{E}$. We define $(a^*(\nu), \theta^*(\nu))$ as the set of optimal arms and the instance feasibility criterion, and $\mathcal{E}_{alt} = \{\nu' \in \mathcal{E} : (a^*(\nu'), \theta^*(\nu')) \cap (a^*(\nu), \theta^*(\nu)) = \phi\}$, which is the set of constrained bandits in $\mathcal{E}$ with different optimal arms/feasibility criteria than ν.

Theorem 3. *Assume that $(\pi, T, (\hat{a}_T, \hat{\theta}_T))$ is δ-sound and let $\nu \in \mathcal{E}$. Then $\mathbb{E}_{\nu\pi}[T] \geq c^*(\nu) \log\left(\frac{1}{4\delta}\right)$, where*

$$c^*(\nu)^{-1} = \sup_{\alpha \in \mathcal{P}_{K-1}} \left(\inf_{\nu' \in \mathcal{E}_{alt}(\nu)} \left(\sum_{a=1}^{K} \alpha_a D(\nu_a, \nu'_a) \right) \right)$$

with $c^*(\nu) = \infty$ if $c^*(\nu)^{-1} = 0$, where $D(\nu_a, \nu'_a)$ is KL-Divergence between distributions ν_a and ν'_a.

Proof. The result is trivial when $\mathbb{E}_{\nu\pi}[T] = \infty$. Now, assume that $\mathbb{E}_{\nu\pi}[T] < \infty \implies \mathbb{P}(T = \infty) = 0$. Let $\nu' \in \mathcal{E}_{alt}(\nu)$ and define event $E = \{T < \infty \text{ and } (\hat{a}_T, \hat{\theta}_T) \notin (a^*(\nu'), \theta^*(\nu'))\}$. Then,

$$\begin{aligned}
2\delta &\geq \mathbb{P}_{\nu\pi}(T < \infty \text{ and } (\hat{a}_T, \hat{\theta}_T) \notin (a^*(\nu), \theta^*(\nu))) + \mathbb{P}_{\nu'\pi}(T < \infty \text{ and } (\hat{a}_T, \hat{\theta}_T) \notin (a^*(\nu'), \theta^*(\nu'))) \\
&\geq \mathbb{P}_{\nu\pi}(E^c) + \mathbb{P}_{\nu'\pi}(E) \\
&\geq \frac{1}{2} \left(-\sum_{a=1}^{K} \mathbb{E}_{\nu\pi}[N_a(T)] D(\nu_a, \nu'_a) \right)
\end{aligned}$$

Where $N_a(T)$ is the number of pulls of arm a after time T. The first inequality comes from the definition of δ-soundness and the last follows from the

Bretagnolle-Huber inequality.The second inequality holds because,

$$\begin{aligned} E^c &= \{T = \infty\} \cup \{T < \infty \text{ and } (\hat{a}_T, \hat{\theta}_T) \in (a^*(\nu'), \theta^*(\nu'))\} \\ &\subseteq \{T = \infty\} \cup \{T < \infty \text{ and } (\hat{a}_T, \hat{\theta}_T) \notin (a^*(\nu), \theta^*(\nu))\} \end{aligned}$$

Therefore,

$$\sum_{a=1}^{K} \mathbb{E}_{\nu\pi}[N_a(T)] D(\nu_a, \nu'_a) \geq \log\left(\frac{1}{4\delta}\right)$$

Using the definition of $c^*(\nu)$,

$$\begin{aligned} \frac{\mathbb{E}_{\nu\pi}[T]}{c^*(\nu)} &= \mathbb{E}_{\nu\pi}[T] \sup_{\alpha \in \mathcal{P}_{K-1}} \left(\inf_{\nu' \in \mathcal{E}_{alt}(\nu)} \left(\sum_{a=1}^{K} \alpha_a D(\nu_a, \nu'_a) \right) \right) \\ &\geq \mathbb{E}_{\nu\pi}[T] \inf_{\nu' \in \mathcal{E}_{alt}(\nu)} \left(\sum_{a=1}^{K} \frac{\mathbb{E}_{\nu\pi}[N_a(T)]}{\mathbb{E}_{\nu\pi}[T]} D(\nu_a, \nu'_a) \right) \\ &= \inf_{\nu' \in \mathcal{E}_{alt}(\nu)} \left(\sum_{a=1}^{K} \frac{\mathbb{E}_{\nu\pi}[N_a(T)]}{D} (\nu_a, \nu'_a) \right) \geq \log\left(\frac{1}{4\delta}\right) \end{aligned}$$

In the special case that $c^*(\nu)^{-1} = 0$, the assumption that $\mathbb{E}_{\nu\pi}[T] < \infty$ would lead to a contradiction. □

Chapter 5

Problem Formulation for Track-and-stop

In this chapter, we shall define our constrained stochastic MAB problem for the Track-and-stop algorithm. To keep the exposition simple, we consider a single constraint. Formally,

- Let $\mathcal{C}$ be the set of two-dimensional Gaussian distributions with covariance matrix I_2, characterized by their means, $\boldsymbol{\mu} = (\mu(1), \mu(2))$.
- Consider an MAB problem with K *arms*, labelled $1, 2, \ldots K$. Each arm is associated with a, probability distribution $\nu_k \in \mathcal{C}$ corresponding to arm $k \in \mathcal{A}$. Thus, each arm can be characterized by two attributes, with the *objective* being $\mu_k(1)$ and the *constraint* being $\mu_k(2)$.
- The user provides a threshold $\tau \in \mathbb{R}$ which acts as an upper bound to the constraint attribute.
- An *instance* of the constrained MAB problem is defined by (ν, τ), where $\nu = \{\nu(k) : k \in \mathcal{A}\}$.
- The arms that satisfy $\mu_k(2) \leq \tau$ are called *feasible arms*. The rest are called *infeasible arms*. The set of feasible arms is denoted by $\mathcal{K}(\nu)$.
- An instance (ν, τ) is said to be feasible if $\mathcal{K}(\nu) \neq \phi$ and is said to be infeasible if $\mathcal{K}(\nu) = \phi$.

5.1 Feasible Instance

- An *optimal arm*(a^*) is defined as the arm with the largest value of $\mu(1)$, subject to the constraint that $\mu(2) \leq \tau$, i.e., $a^* \in \mathcal{K}(\nu)$ and

$\mu_{a^*}(1) > \max_{k \in \mathcal{K}(\nu), k \neq a^*} (\mu_k(1))$.

- Arms with $\mu_k(1) < \mu_{a^*}(1)$ are called sub-optimal arms.
- Infeasible arms with $\mu_k(1) \geq \mu_{a^*}(1)$ are called *deceiver arms*. The set of all deceiver arms is denoted by $\mathcal{K}^d(\nu) = \{k \in \mathcal{A} : \mu_k(2) > \tau, \mu_k(1) \geq \mu_{a^*}(1)\}$.

5.2 Infeasible Instance

- If the instance is infeasible, the best among all infeasible arms is defined as the arm with the lowest constraint attribute and denoted by a^*, i.e., $\mu_{a^*}(2) < \min_{k \neq a^*}(\mu_k(2))$

For simplicity we have assume that there is a unique optimal arm. The goal is to provide an algorithm that outputs a^* with probability $\geq 1 - \delta$ while minimizing the sample complexity, $\mathbb{E}[T]$. Without loss of generality, we will assume that arm 1 is the optimal arm/best among infeasible arms.

Chapter 6

Analysis of Optimal Proportions

For the Specific problem stated in the previous chapter, we will show that the supremum in the general lower bound is actually a maximum, we define

$$\alpha^*(\nu) = \arg\max_{\alpha \in \mathcal{P}_{K-1}} \left(\inf_{\nu' \in \mathcal{E}_{alt}(\nu)} \left(\sum_{a=1}^{K} \alpha_a D(\nu_a, \nu'_a) \right) \right)$$

Let $\mathcal{E}$ be the environmental class of Gaussian bandits with covariance matrix I_2, characterized by their meas $\boldsymbol{\mu} = (\mu(1), \mu(2))$. Clearly, $\mathcal{E}_{alt}(\nu) = \{\nu' \in \mathcal{E} : a^*(\nu') \neq 1 \text{ or } \theta^*_{\nu'} \neq \theta^*_{\nu}\}$. Thus, $\mathcal{E}_{alt}(\nu)$ can rewritten as,

$$\mathcal{E}_{alt}(\nu) = \{\nu' \in \mathcal{E} : a^*(\nu') \neq 1, \phi^*_1(\nu') = \phi^*_1(\nu)\} \cup \{\nu' \in \mathcal{E} : \phi^*_1(\nu') \neq \phi^*_1(\nu)\}$$

This is because, we can change the *outcome* by either making alternate arm optimal without changing the feasibility criterion of arm 1, or by just changing the feasibility criterion of arm 1. If instance (ν, τ) is a feasible instance, i.e, $\theta^*_{\nu} = 0$, then,

$$\mathcal{E}_{alt}(\nu) = \bigcup_{a \neq 1} \{\nu' \in \mathcal{E} : \mu'_a(1) > \mu'_1(1);\ \mu_1(2), \mu'_a(2) \leq \tau\} \cup \{\nu' \in \mathcal{E} : \mu'_1(2) > \tau\}$$

If instance (ν, τ) is a infeasible instance, i.e, $\theta^*_{\nu} = 1$, then,

$$\mathcal{E}_{alt}(\nu) = \bigcup_{a \neq 1} \{\nu' \in \mathcal{E} : \tau < \mu'_a(2) < \mu'_1(2)\} \cup \{\nu' \in \mathcal{E} : \mu'_1(2) > \tau\}$$

Let us define sets I and II as follows,

$$I = \begin{cases} \bigcup_{a\neq 1}\{\nu' \in \mathcal{E} : \mu'_a(1) > \mu'_1(1);\ \mu'_1(1), \mu'_a(2) \leq \tau\}, & \text{if } (\nu,\tau) \text{ is feasible} \\ \bigcup_{a\neq 1}\{\nu' \in \mathcal{E} : \tau < \mu'_a(2) < \mu'_1(2)\}, & \text{if } (\nu,\tau) \text{ is infeasible} \end{cases}$$

$$II = \begin{cases} \{\nu' \in \mathcal{E} : \mu'_1(2) > \tau\}, & \text{if } (\nu,\tau) \text{ is feasible} \\ \{\nu' \in \mathcal{E} : \mu'_1(2) \leq \tau\}, & \text{if } (\nu,\tau) \text{ is infeasible} \end{cases}$$

Using the above, we can rewrite $c^*(\nu)^{-1}$ as,

$$\begin{aligned} c^*(\nu)^{-1} &= \sup_{\alpha\in\mathcal{P}_{K-1}} \left(\min \left(\inf_{\nu'\in I} \left(\sum_{a=1}^{K} \alpha_a D(\nu_a, \nu'_a) \right), \inf_{\nu'\in II} \left(\sum_{a=1}^{K} \alpha_a D(\nu_a, \nu'_a) \right) \right) \right) \\ &= \sup_{\alpha\in\mathcal{P}_{K-1}} \left(\min \left(\min_{a\neq 1} \inf_{\nu'\in I_a} \left(\sum_{a=1}^{K} \alpha_a D(\nu_a, \nu'_a) \right), \inf_{\nu'\in II} \left(\sum_{a=1}^{K} \alpha_a D(\nu_a, \nu'_a) \right) \right) \right) \end{aligned}$$

where,

$$I_a = \begin{cases} \{\nu' \in \mathcal{E} : \mu'_a(1) > \mu'_1(1);\ \mu_1(2), \mu'_a(2) \leq \tau\}, & \text{if } (\nu,\tau) \text{ is feasible} \\ \{\nu' \in \mathcal{E} : \tau < \mu'_a(2) \geq \mu'_1(2)\}, & \text{if } (\nu,\tau) \text{ is infeasible} \end{cases}$$

6.1 Computing $\inf_{\nu'\in I_a} \left(\sum_{a=1}^{K} \alpha_a D(\nu_a, \nu'_a) \right)$

$\inf_{\nu'\in I_a} \left(\sum_{a=1}^{K} \alpha_a D(\nu_a, \nu'_a) \right)$ is an optimization problem, whose solution can be found through observation for various cases.

6.1.1 $\mu_a(1) < \mu_1(1)$ and $\mu_1(2), \mu_a(2) \leq \tau$

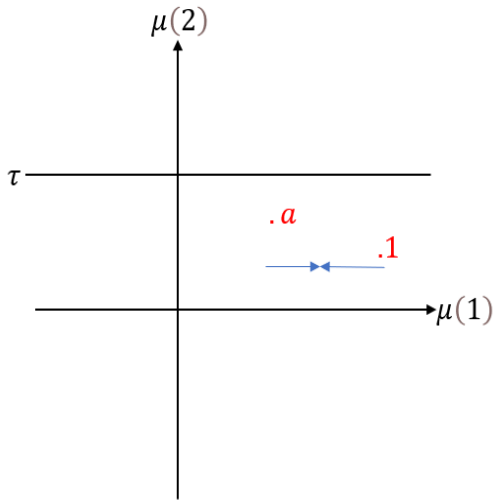


Figure 6.1: $mu_a(1) < \mu_1(1)$ and $\mu_1(2), \mu_a(2) \leq \tau$

Solution: $\mu'_a(1) = \mu'_a(1) = x$, $\mu'_1(2) = \mu_1(2), \mu'_a(2) = \mu_a(2)$ and $\mu'_b(i) = \mu_b(i)$, $\forall b \in \mathcal{A} \setminus \{1, a\}$ and $i \in \{1, 2\}$. Optimizing on x,

$$\frac{\partial}{\partial x}\left[\alpha_1\left\{\frac{1}{2}(\mu_1(1)-x)^2\right\} + \alpha_a\left\{\frac{1}{2}(\mu_a(1)-x)^2\right\}\right] = 0$$
$$\implies x = \frac{\alpha_1\mu_1(1) + \alpha_a\mu_a(1)}{\alpha_1 + \alpha_a}$$

This gives,

$$\inf_{\nu' \in I_a}\left(\sum_{a=1}^{K}\alpha_a D(\nu_a, \nu'_a)\right) = \frac{\alpha_1\alpha_a}{\alpha_1+\alpha_a}\left[\frac{1}{2}(\mu_1(1)-\mu_a(1))^2\right]$$

6.1.2 $\mu_a(1) < \mu_1(1)$ and $\mu_a(2) > \tau$

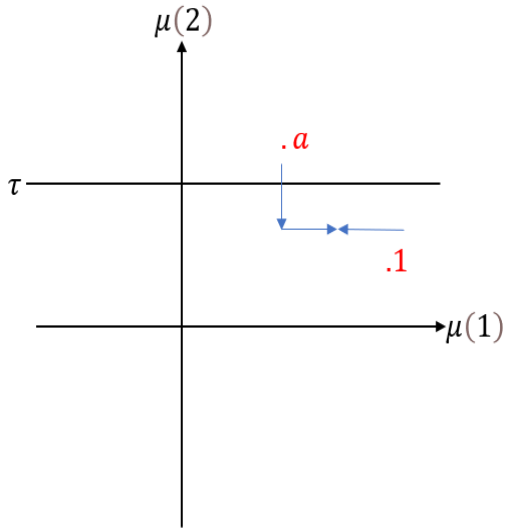


Figure 6.2: $\mu_a(1) < \mu_1(1)$ and $\mu_a(2) > \tau$

Solution: $\mu'_a(1) = \mu'_a(1) = x$, $\mu'_1(2) = \mu_1(2), \mu'_a(2) = \tau$ and $\mu'_b(i) = \mu_b(i)$, $\forall b \in \mathcal{A} \setminus \{1, a\}$ and $i \in \{1, 2\}$. Upon optimizing, we get,

$$\inf_{\nu' \in I_a} \left(\sum_{a=1}^{K} \alpha_a D(\nu_a, \nu'_a) \right) = \frac{\alpha_1 \alpha_a}{\alpha_1 + \alpha_a} \left[\frac{1}{2} (\mu_1(1) - \mu_a(1))^2 \right] + \alpha_a \left[\frac{1}{2} (\mu_a(2) - \tau)^2 \right]$$

6.1.3 $\mu_a(1) \geq \mu_1(1)$ and $\mu_a(2) > \tau$

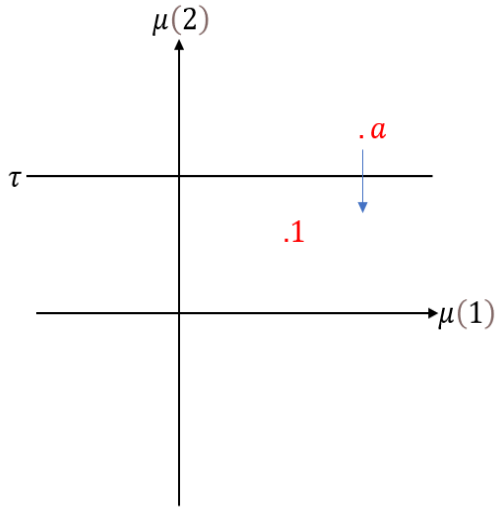


Figure 6.3: $\mu_a(1) \geq \mu_1(1)$ and $\mu_a(2) > \tau$

Solution: $\mu'_a(1) = \mu_a(1)$, $\mu'_a(2) = \tau$ and $\mu'_b(i) = \mu_b(i)$, $\forall b \in \mathcal{A} \setminus \{a\}$ and $i \in \{1, 2\}$. Upon optimizing, we get,

$$\inf_{\nu' \in I_a} \left(\sum_{a=1}^{K} \alpha_a D(\nu_a, \nu'_a) \right) = \alpha_a \left[\frac{1}{2} (\mu_a(2) - \tau)^2 \right]$$

6.1.4 $\tau < \mu_1(2) < \mu_a(2)$

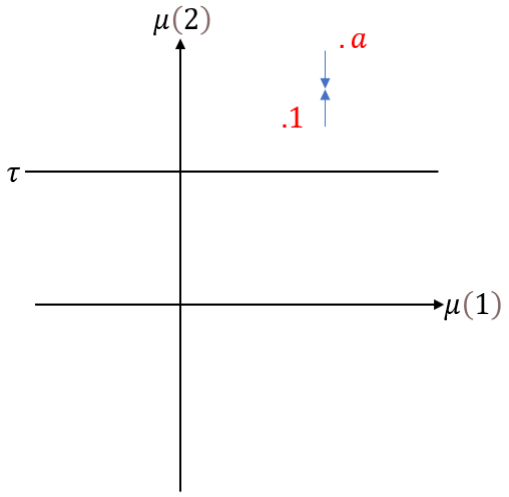


Figure 6.4: $\tau < \mu_1(2) < \mu_a(2)$

Solution: $\mu'_a(2) = \mu'_a(2) = x$, $\mu'_1(1) = \mu_1(), \mu'_a(1) = \mu_a(1)$ and $\mu'_b(i) = \mu_b(i)$, $\forall b \in \mathcal{A} \setminus \{1, a\}$ and $i \in \{1, 2\}$. Upon optimizing, we get,

$$\inf_{\nu' \in I_a} \left(\sum_{a=1}^{K} \alpha_a D(\nu_a, \nu'_a) \right) = \frac{\alpha_1 \alpha_a}{\alpha_1 + \alpha_a} \left[\frac{1}{2} (\mu_1(2) - \mu_a(2))^2 \right]$$

6.2 Computing $\inf_{\nu' \in II} \left(\sum_{a=1}^{K} \alpha_a D(\nu_a, \nu'_a) \right)$

The optimization problem can be solved by just changing the feasibility criterion of arm 1.

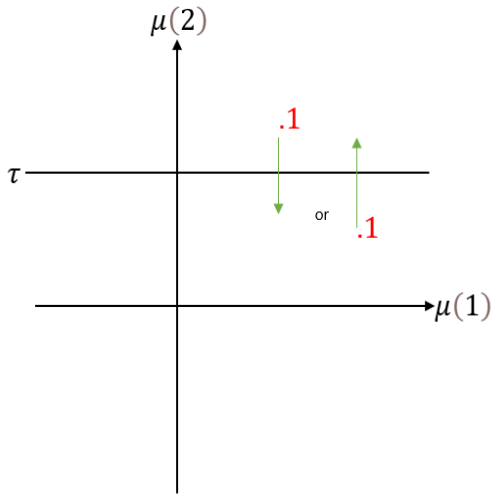


Figure 6.5: Changing the feasibility criterion of arm a

Solution: $\mu_1'(1) = \mu_1(1)$, $\mu_1'(2) = \tau$ and $\mu_b'(i) = \mu_b(i)$, $\forall b \in \mathcal{A} \setminus \{a\}$ and $i \in \{1, 2\}$. Upon optimizing, we get,

$$\inf_{\nu' \in II} \left(\sum_{a=1}^{K} \alpha_a D(\nu_a, \nu_a') \right) = \alpha_1 \left[\frac{1}{2} (\mu_1(2) - \tau)^2 \right]$$

Now, we define a function $G_a(x_a)$ as follows,

$$G_a(x_a) = \begin{cases} \frac{x_a}{1+x_a} \left[\frac{1}{2} (\mu_1(1) - \mu_a(1))^2 \right], \\ \text{if } \mu_a(1) < \mu_1(1) \text{ and } \mu_1(2), \mu_a(2) \leq \tau \\ \\ \frac{x_a}{1+x_a} \left[\frac{1}{2} (\mu_1(1) - \mu_a(1))^2 \right] + x_a \left[\frac{1}{2} (\mu_a(2) - \tau)^2 \right], \\ \text{if } \mu_a(1) < \mu_1(1) \text{ and } \mu_a(2) > \tau \text{ and } \mu_1(2) \leq \tau \\ \\ x_a \left[\frac{1}{2} (\mu_a(2) - \tau)^2 \right], \\ \text{if } \mu_a(1) > \mu_1(1) \text{ and } \mu_a(2) > \tau \text{ and } \mu_1(2) \leq \tau \\ \\ \frac{x_a}{1+x_a} \left[\frac{1}{2} (\mu_1(2) - \mu_a(2))^2 \right], \\ \text{if } \tau < \mu_a(1) < \mu_1(1) \end{cases}$$

We notice that $G_a(x_a)$ is a monotonically increasing function. Also, let $\alpha^*(\nu)$ be an element in

$$\begin{aligned}&\arg\max_{\alpha\in\mathcal{P}_{K-1}}\left(\min\left(\min_{a\neq 1}\inf_{\nu'\in I_a}\left(\sum_{a=1}^{K}\alpha_a D(\nu_a,\nu'_a)\right),\inf_{\nu'\in II}\left(\sum_{a=1}^{K}\alpha_a D(\nu_a,\nu'_a)\right)\right)\right)\\ =&\arg\max_{\alpha\in\mathcal{P}_{K-1}}\left(\min\left(\alpha_1\min_{a\neq 1}G_a\left(\frac{\alpha_a}{\alpha_1}\right),\ \alpha_1\left(\frac{1}{2}(\mu_1(2)-\tau)^2\right)\right)\right)\end{aligned}$$

The above holds because $\alpha_1^*(\nu)\neq 0$. Introducing $x_a^*=\frac{\alpha_a^*}{\alpha_1^*}$. Thus,

$$\alpha_1^*=\frac{1}{1+\sum_{k=2}^{K}x_k^*}\text{ and, for }a\geq 2,\ \alpha_a^*=\frac{x^*}{1+\sum_{k=2}^{K}x_k^*}$$

and $(x_2^*,\ldots,x_K^*)\in\mathbb{R}^{K-1}$ is such that,

$$(x_2^*,\ldots,x_K^*)\in\arg\max_{(x_2,\ldots,x_K)\in\mathbb{R}^{K-1}}\left(\min\left(\frac{\min_{a\neq 1}G_a(x_a)}{1+\sum_{k=2}^{K}x_k},\ \frac{\left(\frac{1}{2}(\mu_1(2)-\tau)^2\right)}{1+\sum_{k=2}^{K}x_k}\right)\right)$$

Additionally, we introduce $(\tilde{x}_2,\ldots,\tilde{x}_K)\in\mathbb{R}^{K-1}$ such that,

$$(\tilde{x}_2,\ldots,\tilde{x}_K)\in\arg\max_{(x_2,\ldots,x_K)\in\mathbb{R}^{K-1}}\left(\frac{\min_{a\neq 1}G_a(x_a)}{1+\sum_{k=2}^{K}x_k}\right)$$

$$\tilde{y}=\frac{1+\sum_{k=2}^{K}G_a^{-1}(\tilde{y})}{\sum_{k=2}^{K}(G_a^{-1})'(\tilde{y})}$$

Lemma 2. *$G_a(\tilde{x}_a)=\tilde{y}\in[0,y_{max}),\ \forall a\in\{2,\ldots,K\}$, where y_{max} comes from the definition of $G_a(x_a)$.*

Proof. Let $B=\{b\in\{2,\ldots,K\}:G_b(\tilde{x}_b)=\min_{a\neq 1}G_a(\tilde{x}_a)\}$ and $A=\{2,\ldots,K\}\backslash B$. Assume $A\neq\phi$. For all $a\in A$ and $b\in B$, $G_a(\tilde{x}_a)>G_b(\tilde{x}_b)$. As $G_a(x_a)$ is monotonically increasing, $\exists\ \epsilon>0$ such that, for all $a\in A$ and $b\in B$

$$G_a\left(\tilde{x}_a-\frac{\epsilon}{|A|}\right)>G_a\left(\tilde{x}_b+\frac{\epsilon}{|B|}\right)>G_b(\tilde{x}_b)$$

Introducing $\bar{x}_a=\tilde{x}_a-\frac{\epsilon}{|A|},\ \forall a\in A$ and $\bar{x}_b=\tilde{x}_b+\frac{\epsilon}{|B|},\ \forall b\in B$. There exists a $b\in B$ such that,

$$\frac{\min_{a\neq 1}G_a(\bar{x}_a)}{1+\sum_{k=2}^{K}\bar{x}_k}=\frac{\min_{a\neq 1}G_b\left(\tilde{x}_b+\frac{\epsilon}{|B|}\right)}{1+\sum_{k=2}^{K}\tilde{x}_k}>\frac{\min_{a\neq 1}G_b(\tilde{x}_b)}{1+\sum_{k=2}^{K}\tilde{x}_k}$$

This contradicts the fact that,

$$(\tilde{x}_2, \ldots, \tilde{x}_K) \in \underset{(x_2,\ldots,x_K)\in\mathbb{R}^{K-1}}{\arg\max} \left(\frac{\min_{a\neq 1} G_a(x_a)}{1 + \sum_{k=2}^{K} x_k} \right)$$

Hence, $A = \phi$ and $\exists\ \tilde{y} \in [0, y_{max})$ such that,

$$G_a(\tilde{x}_a) = \tilde{y},\ \forall a \in \{2, \ldots, K\}$$

where y_{max} comes from the definition of G_a. □

Lemma 3. $\min_{a\neq 1}(G_a(\tilde{x}_a)) = \tilde{y} \in [0, y_{max})$, *where* y_{max} *is determined by the arms, is the unique solution of*

$$\tilde{y} = \frac{1 + \sum_{a=2}^{K} G_a^{-1}(\tilde{y})}{\sum_{a=2}^{K} (G_a^{-1})'(\tilde{y})}$$

if it exists. Else, $\tilde{y} = y_{max}^{-}$.

proof. From lemma 2,

$$\tilde{y} \in \underset{y\in[0,y_{max})}{\arg\max} \frac{y}{1 + \sum_{a=2}^{K} G_a^{-1}(y)}$$

Differentiating w.r.t y and equating to 0 at $y = \tilde{y}$, we get,

$$\tilde{y} = \frac{1 + \sum_{a=2}^{K} G_a^{-1}(\tilde{y})}{\sum_{a=2}^{K} (G_a^{-1})'(\tilde{y})}$$

To show that the above has a unique solution, we observe that,

$$\frac{\partial}{\partial y} \frac{y}{1 + \sum_{a=2}^{K} G_a^{-1}(y)} = \frac{1 + \sum_{a=2}^{K} (G_a^{-1}(y) - y(G_a^{-1})'(y))}{(1 + 1 + \sum_{a=2}^{K} G_a^{-1}(y))^2} = 0$$

is at most a single root, and is maxima. For this, we derive $G_a^{-1}(y)$ and $(G_a^{-1})'(y)$ and observe that properties of $(G_a^{-1}(y) - y(G_a^{-1})'(y))$ for all types of arms. We observe that it is either 0 or decreasing for $y \in [0, y_{max})$. If no root exists then, $\tilde{y} = y_{max}^{-}$ □

Lemma 4. $\min_{a\neq 1}(G_a(x_a^*)) \leq \frac{1}{2}(\mu_1(2) - \tau)^2$

proof. Assume that $\min_{a\neq 1}(G_a(x_a^*)) > \frac{1}{2}(\mu_1(2)-\tau)^2$,

$$\implies \min\left(\frac{\min_{a\neq 1} G_a\left(x_a^*\right)}{1+\sum_{k=2}^K x_k^*},\ \frac{\frac{1}{2}(\mu_1(2)-\tau)^2}{1+\sum_{k=2}^K x_k^*}\right) = \frac{\frac{1}{2}(\mu_1(2)-\tau)^2}{1+\sum_{k=2}^K x_k^*}$$

As $\min_{a\neq 1}(G_a(x_a^*)) > \frac{1}{2}(\mu_1(2)-\tau)^2 \implies G_a(x_a^*) > \frac{1}{2}(\mu_1(2)-\tau)^2,\ \forall a \in \{2,\dots,K\}$ and $G_a(x_a)$ is a monotonically increasing function, for all $a \in \{2,\dots,K\}$, there exist, $\bar{x}_a < x_a^* : G_a(\bar{x}_a) = \frac{1}{2}(\mu_1(2)-\tau)^2$. This implies,

$$\min\left(\frac{\min_{a\neq 1} G_a\left(\bar{x}_a\right)}{1+\sum_{k=2}^K \bar{x}_k},\ \frac{\frac{1}{2}(\mu_1(2)-\tau)^2}{1+\sum_{k=2}^K \bar{x}_k}\right) = \frac{\frac{1}{2}(\mu_1(2)-\tau)^2}{1+\sum_{k=2}^K \bar{x}_k}$$
$$> \frac{\frac{1}{2}(\mu_1(2)-\tau)^2}{1+\sum_{k=2}^K x_k^*} = \min\left(\frac{\min_{a\neq 1} G_a\left(x_a^*\right)}{1+\sum_{k=2}^K x_k^*},\ \frac{\frac{1}{2}(\mu_1(2)-\tau)^2}{1+\sum_{k=2}^K x_k^*}\right)$$

This contradicts the fact that,

$$(x_2^*,\dots,x_K^*) \in \underset{(x_2,\dots,x_K)\in\mathbb{R}^{K-1}}{\arg\max}\left(\min\left(\frac{\min_{a\neq 1} G_a\left(x_a\right)}{1+\sum_{k=2}^K x_k},\ \frac{\left(\frac{1}{2}(\mu_1(2)-\tau)^2\right)}{1+\sum_{k=2}^K x_k}\right)\right)$$

Therefore, $\min_{a\neq 1}(G_a(x_a^*)) \leq \frac{1}{2}(\mu_1(2)-\tau)^2$ □

Lemma 5. $G_a(x_a^*) = y^* \in [0,\ y_{max}),\ \forall a \in \{2,\dots,K\}$

proof. From lemma 4, we know that $\min_{a\neq 1}(G_a(x_a^*)) \leq \frac{1}{2}(\mu_1(2)-\tau)^2$. This implies that,

$$\frac{\min_{a\neq 1} G_a\left(x_a^*\right)}{1+\sum_{k=2}^K x_k^*} \leq \frac{\frac{1}{2}(\mu_1(2)-\tau)^2}{1+\sum_{k=2}^K x_k^*}$$

If $\min_{a\neq 1}(G_a(x_a^*)) < \frac{1}{2}(\mu_1(2)-\tau)^2$, let $B = \{b \in \{2,\dots,K\} : G_b(x_b^*) = \min_{a\neq 1} G_a(x_a^*)\}$ and $A = \{2,\dots,K\} \setminus B$. Assume $A \neq \phi$. For all $a \in A$ and $b \in B$, $G_a(x_a^*) > G_b(x_b^*)$ and $\frac{1}{2}(\mu_1(2)-\tau)^2 > G_b(x_b^*)$. As $G_a(x_a)$ is monotonically increasing, $\exists\ \epsilon > 0$ such that, for all $a \in A$ and $b \in B$

$$G_a\left(x_a^* - \frac{\epsilon}{|A|}\right) > G_a\left(x_b^* + \frac{\epsilon}{|B|}\right) > G_b(x_b^*)$$

and

$$\frac{1}{2}(\mu_1(2)-\tau)^2 > G_a\left(x_b^* + \frac{\epsilon}{|B|}\right) > G_b(x_b^*)$$

Introducing $\bar{x}_a = x_a^* - \frac{\epsilon}{|A|},\ \forall a \in A$ and $\bar{x}_b = x_b^* + \frac{\epsilon}{|B|},\ \forall b \in B$. There exists a $b \in B$ such that,

$$\min\left(\frac{\min_{a\neq 1} G_a\left(\bar{x}_a\right)}{1+\sum_{k=2}^{K}\bar{x}_k},\ \frac{\left(\frac{1}{2}(\mu_1(2)-\tau)^2\right)}{1+\sum_{k=2}^{K}\bar{x}_k}\right) = \frac{\min_{a\neq 1} G_a\left(\bar{x}_a\right)}{1+\sum_{k=2}^{K}x_k^*}$$

$$=\frac{G_b\left(x_b^*+\frac{\epsilon}{|B|}\right)}{1+\sum_{k=2}^{K}x_k^*} > \frac{G_b\left(x_b^*\right)}{1+\sum_{k=2}^{K}x_k^*} = \min\left(\frac{\min_{a\neq 1} G_a\left(x_a^*\right)}{1+\sum_{k=2}^{K}x_k^*},\ \frac{\frac{1}{2}(\mu_1(2)-\tau)^2}{1+\sum_{k=2}^{K}x_k^*}\right)$$

This contradicts the fact that $A \neq \phi$. Therefore, $G_a(x_a^*) = y^*,\ \forall a \in \{2, ..., K\}$.

If $\min_{a\neq 1}(G_a(x_a^*)) = \frac{1}{2}(\mu_1(2)-\tau)^2$, let $B = \{b \in \{2,\ldots,K\} : G_b(x_b^*) = \min_{a\neq 1} G_a(x_a^*)\}$ and $A = \{2,\ldots,K\} \setminus B$. Assume $A \neq \phi$. For all $a \in A$ and $b \in B$, $G_a(x_a^*) > G_b(x_b^*)$. As $G_a(x_a)$ is monotonically increasing, $\forall a \in A,\ \exists\ \epsilon_a > 0$ such that, $G_a(x_A^* - \epsilon_a) = \frac{1}{2}(\mu_1(2)-\tau)^2$.Introducing $\bar{x}_a = x_a^* - \epsilon_a,\ \forall a \in A$ and $\bar{x}_b = x_b^*,\ \forall b \in B$.

$$\min\left(\frac{\min_{a\neq 1} G_a\left(\bar{x}_a\right)}{1+\sum_{k=2}^{K}\bar{x}_k},\ \frac{\left(\frac{1}{2}(\mu_1(2)-\tau)^2\right)}{1+\sum_{k=2}^{K}\bar{x}_k}\right) = \frac{\frac{1}{2}(\mu_1(2)-\tau)^2}{1+\sum_{k=2}^{K}\bar{x}_k}$$

$$> \frac{\frac{1}{2}(\mu_1(2)-\tau)^2}{1+\sum_{k=2}^{K}x_k^*} = \min\left(\frac{\min_{a\neq 1} G_a\left(x_a^*\right)}{1+\sum_{k=2}^{K}x_k^*},\ \frac{\frac{1}{2}(\mu_1(2)-\tau)^2}{1+\sum_{k=2}^{K}x_k^*}\right)$$

This contradicts the fact that $A \neq \phi$. Therefore, $G_a(x_a^*) = y^* = \frac{1}{2}(\mu_1(2) - \tau)^2\ , \forall a \in \{2, ..., K\}$. Additionally we note that,

$$y^* \in \underset{y\in[0,y_{max})}{\arg\max}\left(\min\left(\frac{y}{1+\sum_{k=2}^{K}G_a^{-1}(y)},\ \frac{\left(\frac{1}{2}(\mu_1(2)-\tau)^2\right)}{1+\sum_{k=2}^{K}G_a^{-1}(y)}\right)\right)$$

□

Theorem 4. *If* $\min_{a\neq 1}(G_a(\tilde{x}_a)) \leq \frac{1}{2}(\mu_1(2)-\tau)^2$ *then,*

$$(\tilde{x}_2,\ldots,\tilde{x}_K) \in \underset{(x_2,\ldots,x_K)\in\mathbb{R}^{K-1}}{\arg\max}\left(\min\left(\frac{\min_{a\neq 1} G_a\left(x_a\right)}{1+\sum_{k=2}^{K}x_k},\ \frac{\left(\frac{1}{2}(\mu_1(2)-\tau)^2\right)}{1+\sum_{k=2}^{K}x_k}\right)\right)$$

Thus $x_a^* = \tilde{x}_a = G_a^{-1}(\tilde{y})$. *Else,* $x_a^* = G_a^{-1}\left(\frac{1}{2}(\mu_1(2)-\tau)^2\right)$ *from which* α^* *can be easily computed.*

proof. If $\min_{a\neq 1}(G_a(\tilde{x}_a)) \leq \frac{1}{2}(\mu_1(2)-\tau)^2$, then $\tilde{y} \leq \frac{1}{2}(\mu_1(2)-\tau)^2$. Therefore for all $(x_2,\ldots,x_K) \in \mathbb{R}^{K-1}$,

$$\frac{\min_{a\neq 1} G_a(x_a)}{1+\sum_{k=2}^K x_k} \leq \frac{\min_{a\neq 1} G_a(\tilde{x}_a)}{1+\sum_{k=2}^K \tilde{x}_k} \leq \frac{\frac{1}{2}(\mu_1(2)-\tau)^2}{1+\sum_{k=2}^K \tilde{x}_k}$$

This implies,

$$\min\left(\frac{\min_{a\neq 1} G_a(x_a)}{1+\sum_{k=2}^K x_k},\ \frac{\frac{1}{2}(\mu_1(2)-\tau)^2}{1+\sum_{k=2}^K x_k}\right) \leq \min\left(\frac{\min_{a\neq 1} G_a(\tilde{x}_a)}{1+\sum_{k=2}^K \tilde{x}_k},\ \frac{\frac{1}{2}(\mu_1(2)-\tau)^2}{1+\sum_{k=2}^K \tilde{x}_k}\right)$$

Hence,

$$(\tilde{x}_2,\ldots,\tilde{x}_K) \in \underset{(x_2,\ldots,x_K)\in\mathbb{R}^{K-1}}{\arg\max}\left(\min\left(\frac{\min_{a\neq 1} G_a(x_a)}{1+\sum_{k=2}^K x_k},\ \frac{\left(\frac{1}{2}(\mu_1(2)-\tau)^2\right)}{1+\sum_{k=2}^K x_k}\right)\right)$$

and $x_a^* = \tilde{x}_a = G_a^{-1}(\tilde{y})$.

If $\min_{a\neq 1}(G_a(\tilde{x}_a)) > \frac{1}{2}(\mu_1(2)-\tau)^2$. Let $y^* \in [0, y_{max})$ be an element in,

$$\underset{y\in[0,y_{max})}{\arg\max}\left(\min\left(\frac{y}{1+\sum_{k=2}^K G_a^{-1}(y)},\ \frac{\left(\frac{1}{2}(\mu_1(2)-\tau)^2\right)}{1+\sum_{k=2}^K G_a^{-1}(y)}\right)\right)$$

From From the proof of lemma 3, we know that,

$$\frac{y}{1+\sum_{k=2}^K G_a^{-1}(y)}$$

has a unique maxima, if any. Therefore, in $y \in [0,\tilde{y}]$, the above function is monotonically increasing. Since, $G_a(x_a)$ is monotonically increasing, $G_a^{-1}(y)$ is also monotonically increasing in $y \in [0, y_{max})$ and hence $y \in [0,\tilde{y}]$. Therefore,

$$\frac{\left(\frac{1}{2}(\mu_1(2)-\tau)^2\right)}{1+\sum_{k=2}^K G_a^{-1}(y)}$$

is a monotonically decreasing function of y in $[0,\tilde{y}]$. Hence,

$$\min\left(\frac{y}{1+\sum_{k=2}^K G_a^{-1}(y)},\ \frac{\left(\frac{1}{2}(\mu_1(2)-\tau)^2\right)}{1+\sum_{k=2}^K G_a^{-1}(y)}\right)$$

is maximized when,

$$\frac{y}{1+\sum_{k=2}^K G_a^{-1}(y)} = \frac{\left(\frac{1}{2}(\mu_1(2)-\tau)^2\right)}{1+\sum_{k=2}^K G_a^{-1}(y)}$$

That is, $y^* = \frac{1}{2}(\mu_1(2) - \tau)^2$ if $\frac{1}{2}(\mu_1(2) - \tau)^2 \in [0, \tilde{y}]$. Now, if $\min_{a \neq 1}(G_a(\tilde{x}_a)) > \frac{1}{2}(\mu_1(2) - \tau)^2 \implies \tilde{y} > \frac{1}{2}(\mu_1(2) - \tau)^2$. This implies that $\frac{1}{2}(\mu_1(2) - \tau)^2 \in [0, \tilde{y}]$. Therefore, $y^* = \frac{1}{2}(\mu_1(2) - \tau)^2$. From lemma 5, $G_a(x_a^*) = y^*$, $\forall a \in \{2, ..., K\} \implies x_a^* = G_a^{-1}(y^*) = G_a^{-1}\left(\frac{1}{2}(\mu_1(2) - \tau)^2\right)$.

Therefore, $x_a^* = G_a^{-1}(y^*)$, where $y^* = \min\left(\tilde{y}, \frac{1}{2}(\mu_1(2) - \tau)^2\right)$. □

Chapter 7

Track-and-stop Algorithm

7.1 Sampling Rule: D-Tracking

This sampling strategy is exactly the same as the D-tracking strategy propounded in the paper by Garivier et al. (2016). Let $(\hat{\nu}(t), \tau)$ be the estimate of the bandit instance (ν, τ) characterized by the empirical means, $(\hat{\boldsymbol{\mu}}_1(t), \ldots, \hat{\boldsymbol{\mu}}_K(t))$. Let $N_a(t)$ be the number of pull of arm a until time t. Define, $U_t = \{a : N_a(t) < \sqrt{t} - K/2\}$. The D-tracking rule (A_t) is sequentially defined as,

$$A_t = \begin{cases} \arg\min_{a \in U_t}(U_t), \text{if } U_t \neq \phi & \text{(forced exploration)} \\ \arg\max_{a \in \mathcal{A}}(t\alpha_a^*(\hat{\nu}) - N_a(t)), \text{if } U_t = \phi & \text{(direct tracking)} \end{cases}$$

Proposition 1. *The D-Tracking sampling rules satisfy,*

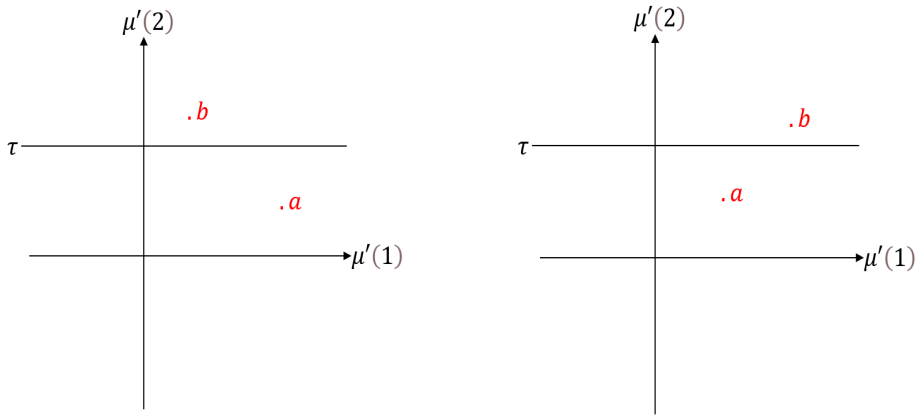
$$\mathbb{P}_\alpha \left(\lim_{t \to \infty} \frac{N_a(t)}{t} = \alpha^*(\nu) \right) = 1$$

Using the above proposition, we can show almost sure convergence of the stopping time to our lower bound.

7.2 Stopping Rule

Consider a two-armed bandit instance with arms a and b and feasibility criterion of the instance given by $\gamma \in \{0, 1\}$. Arm a is *best/optimal* arm given that the instance γ is feasible, i.e., $\gamma = 0$, if,

1. $\mu'_a(2) \leq \tau$ and $\mu'_b(2) \geq \tau$



(a) $\mu'_a(1) \geq \mu'_b(1)$ (b) $\mu'_a(1) \leq \mu'_b(1)$

Figure 7.1: $\mu'_a(2) \leq \tau$ and $\mu'_b(2) \geq \tau$

2. $\mu'_a(2), \mu'_b(2) \leq \tau$ and $\mu'_a(1) \geq \mu'_b(1)$

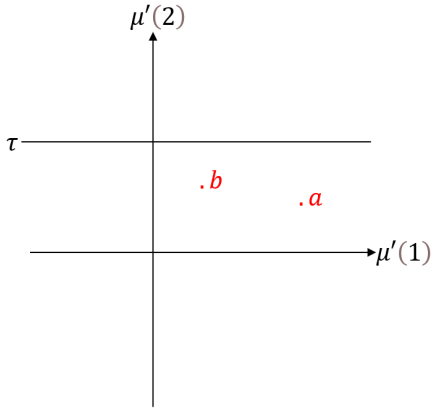


Figure 7.2: $\mu'_a(2), \mu'_b(2) \leq \tau$ and $\mu'_a(1) \geq \mu'_b(1)$

Similarly, arm a is the *best/optimal* arm given that the instance γ is infeasible, i.e., $\gamma = 1$, if,

1. $\tau \leq \mu'_a(2) \leq \mu'_b(2)$

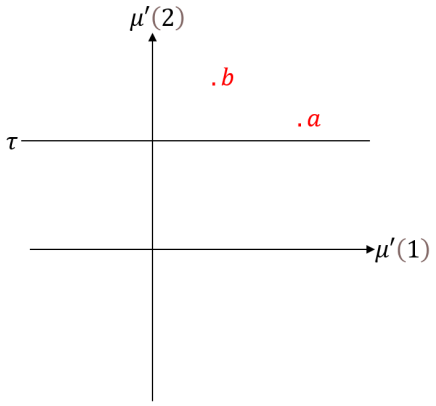


Figure 7.3: $\tau \leq \mu'_a(2) \leq \mu'_b(2)$

Let us denote arm a being *optimal* in the two-armed instance as $a \geq b$.

Now, we shall look at the a more general K-armed bandit instance. Let $\theta \in \{0, 1\}$ represent the feasibility criterion of the above instance such that for a feasible instance, $\theta = 0$ otherwise, $\theta = 1$. For all arms $a, b \in \mathcal{A}$ and instances $\theta \in \{0, 1\}$, let us define hypothesis $\mathcal{H}_0$ as follows,

- Consider the two-armed sub-instance formed by arms a and b. Let γ be the feasibility criterion of that sub-instance.

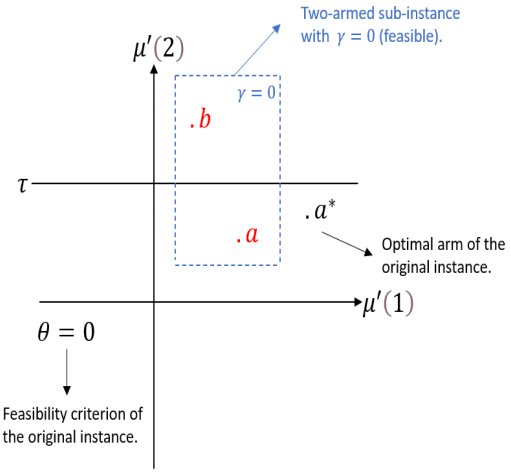
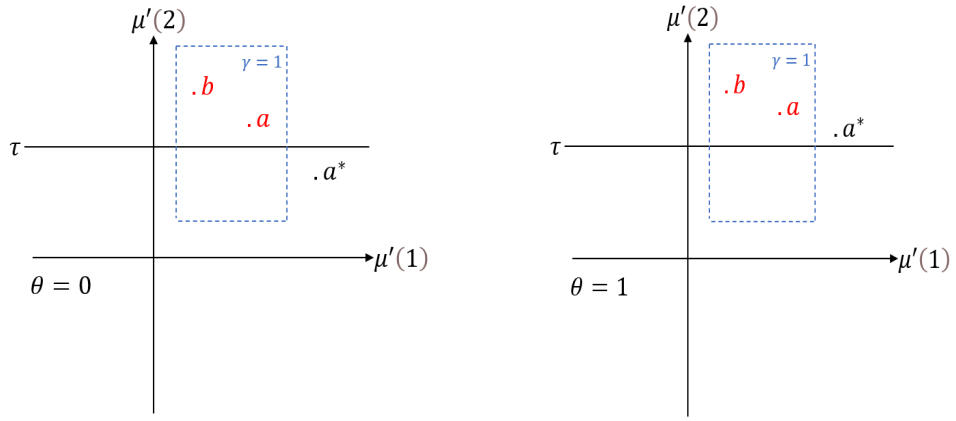


Figure 7.4: Two-armed sub-instance with $\gamma = 0$ and $\theta = 0$

- In the sub-instance so formed, arm a is *optimal* given feasibility criterion γ. In short, $\mathcal{H}_0 = (a \geq b, \gamma)$

The following figure shows the remaining possible combinations of instance and sub-instance feasibility criteria,



(a) $\gamma = 1, \theta = 0$ (b) $\gamma = 1, \theta = 1$

Figure 7.5: Remaining possible combinations of γ and θ

Using the previous definition, we can define the alternative hypothesis $\mathcal{H}_1$ as arm b is *optimal* in the sub-instance given feasibility criterion γ or the feasibility criterion for the sub-instance is γ^c. In short, $\mathcal{H}_1 = (b \geq a, \gamma)$ *or* (γ^c). We shall use these definition to define $Z_{a,b,\gamma}(t)$.

For any arms $a, b \in \mathcal{A}$ and their sub-instance feasibility criterion γ, we can use the Generalized Likelihood Ratio statistic to define $Z_{a,b,\gamma}(t)$ as follows,

$$Z_{a,b,\gamma}(t) := \log \frac{\max_{\mathcal{H}_0} \mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right) \mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)}{\max_{\mathcal{H}_1} \mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right) \mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)}$$

where $\bar{X}^a_{N_a(t)} = (X_s : A_s = a, s \leq t)$ is the vector of rewards obtained from arm a at time t, and $\mathcal{L}_\nu(Z_1, \ldots, Z_n)$ is the likelihood of n i.i.d. observations such that $Z_i \sim \mathcal{N}(\mu, I_2)$,

$$\mathcal{L}_\nu(Z_1,\ldots,Z_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-\left(Z_i(1)-\mu(1)\right)^2 - \left(Z_i(2)-\mu(2)\right)^2}{2}\right)$$
$$= \frac{1}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left\{\left(Z_i(1)-\mu(1)\right)^2 + \left(Z_i(2)-\mu(2)\right)^2\right\}\right]$$

7.2.1 Calculation of $Z_{a,b,\gamma}(t)$ for Various Cases

As log(.) is a monotonically increasing function, we can rewrite $Z_{a,b,\gamma}(t)$ as,

$$Z_{a,b,\gamma}(t) = \log \frac{\max_{\mathcal{H}_0} \mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)}{\max_{\mathcal{H}_1} \mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)}$$
$$= \max_{\mathcal{H}_0}\left[\log\left(\mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)\right)\right] - \max_{\mathcal{H}_1}\left[\log\left(\mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)\right)\right]$$

Cancelling out the constant $-\frac{N_a(t)+N_b(t)}{2}\log(2\pi)$, we get

$$Z_{a,b,\gamma}(t) = \max_{\mathcal{H}_0}\left[-\frac{1}{2}\sum_{s\in[t]:A_s=a}\left\{\left(X_s(1)-\mu'_a(1)\right)^2 + \left(X_s(2)-\mu'_a(2)\right)^2\right\}\right]$$
$$+ \max_{\mathcal{H}_0}\left[-\frac{1}{2}\sum_{s\in[t]:A_s=b}\left\{\left(X_s(1)-\mu'_b(1)\right)^2 + \left(X_s(2)-\mu'_b(2)\right)^2\right\}\right]$$
$$- \max_{\mathcal{H}_1}\left[-\frac{1}{2}\sum_{s\in[t]:A_s=a}\left\{\left(X_s(1)-\mu'_a(1)\right)^2 + \left(X_s(2)-\mu'_a(2)\right)^2\right\}\right]$$
$$- \max_{\mathcal{H}_1}\left[-\frac{1}{2}\sum_{s\in[t]:A_s=b}\left\{\left(X_s(1)-\mu'_b(1)\right)^2 + \left(X_s(2)-\mu'_b(2)\right)^2\right\}\right]$$

We assume that after t rounds, hypothesis $\mathcal{H}_0$ is *apparently* true, i.e., the empirical means satisfy $\mathcal{H}_0$ given the apparent sub-instance feasibility

criterion $\hat{\gamma}$. Following this assumption and maximizing w.r.t $\mu'_a(1)$, we get,

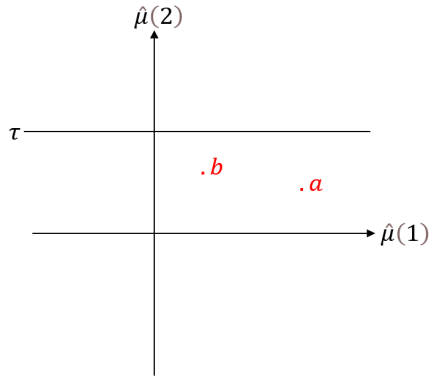
$$\frac{\partial\left[-\frac{1}{2}\sum_{s\in[t]:A_s=a}\left\{(X_s(1)-\mu'_a(1))^2+(X_s(2)-\mu'_a(2))^2\right\}\right]}{\partial\mu'_a(1)} = -\sum_{s\in[t]:A_s=a}(X_s(1)-\mu'_a(1)) = 0$$

$$\implies \mu'_a(1) = \frac{\sum_{s\in[t]:A_s=a}X_s(1)}{N_a(t)} = \hat{\mu}_a(t)$$

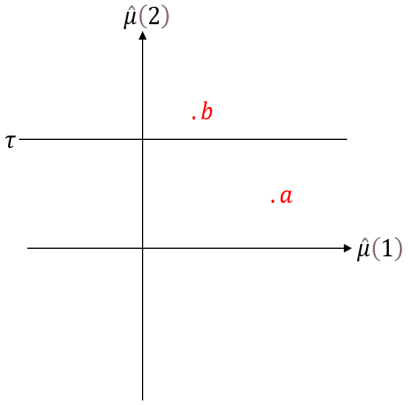
Similarly, maximizing w.r.t $\mu'_a(2), \mu'_b(1)$ and $\mu'_b(2)$, we get,

$$\begin{aligned}\max_{\mathcal{H}_0}\left[\log\left(\mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)\right)\right] = &-\frac{1}{2}\sum_{s\in[t]:A_s=a}\left\{(X_s(1)-\hat{\mu}_a(1))^2+(X_s(2)-\hat{\mu}_a(2))^2\right\}\\ &-\frac{1}{2}\sum_{s\in[t]:A_s=b}\left\{(X_s(1)-\hat{\mu}_b(1))^2+(X_s(2)-\hat{\mu}_b(2))^2\right\}\end{aligned}$$

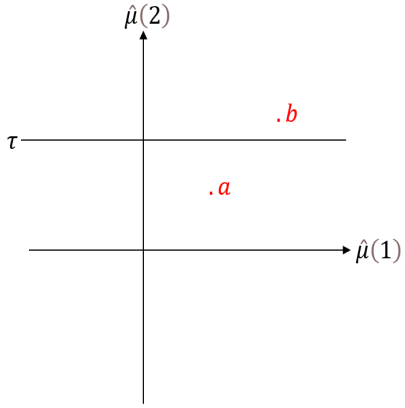
The next step is to obtain $\max_{\mathcal{H}_1}\left[\log\left(\mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)\right)\right]$ given that $\mathcal{H}_0$ is *apparently* true. The hypothesis $\mathcal{H}_0$ is *apparently* true if the empirical means of arms a, b and the apparent sub-instance feasibility criterion are represented by one of the following,



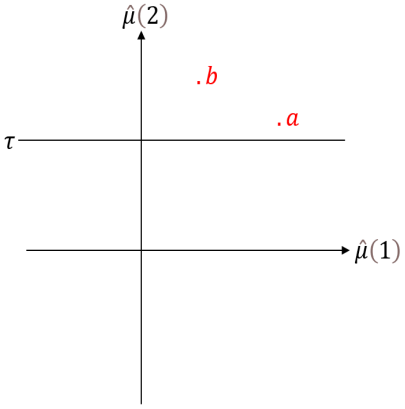
(a) $\hat{\mu}_a(2), \hat{\mu}_b(2) \leq \tau, \hat{\mu}_a(1) \geq \hat{\mu}_b(1)$ and $\hat{\gamma} = 0$



(b) $\hat{\mu}_a(2) \leq \tau, \hat{\mu}_b(2) \geq \tau, \hat{\mu}_a(1) \geq \hat{\mu}_b(1)$ and $\hat{\gamma} = 0$



(c) $\hat{\mu}_a(2) \leq \tau, \hat{\mu}_b(2) \geq \tau, \hat{\mu}_a(1) \leq \hat{\mu}_b(1)$ and $\hat{\gamma} = 0$



(d) $\hat{\mu}_a(2) \leq \hat{\mu}_b(2)$ and $\hat{\gamma} = 1$

Figure 7.6: Possible cases that satisfy hypothesis $\mathcal{H}_0$

We will now obtain $\max_{\mathcal{H}_1} \left[\log\left(\mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)\right)\right]$ for each of the above cases separately.

1. $\hat{\mu}_a(2), \hat{\mu}_b(2) \leq \tau, \hat{\mu}_a(1) \geq \hat{\mu}_b$ **and** $\hat{\gamma} = 0$

We know that,

$$\max_{\mathcal{H}_1}\left[\log\left(\mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\ \mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)\right)\right] =$$

$$\max_{\mathcal{H}_1}\left[-\frac{1}{2}\sum_{s\in[t]:A_s=a}\left\{(X_s(1)-\hat{\mu}_a(1))^2+(X_s(2)-\hat{\mu}_a(2))^2\right\}\right]$$

$$+\max_{\mathcal{H}_1}\left[-\frac{1}{2}\sum_{s\in[t]:A_s=b}\left\{(X_s(1)-\hat{\mu}_b(1))^2+(X_s(2)-\hat{\mu}_b(2))^2\right\}\right]$$

Since we have assumed the that the instance apparently satisfies $\hat{\mu}_a(2), \hat{\mu}_b(2) \leq \tau, \hat{\mu}_a(1) \geq \hat{\mu}_b(1)$ and $\hat{\gamma} = 0$, we will look at the *nearest* alternatives that will make the alternate hypothesis $\mathcal{H}_1$ *apparently* true. The idea being that, to maximize $\log\left(\mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)\right)$, we must deviate μ'_a, μ'_b by the least amount possible from $\hat{\mu}_a, \hat{\mu}_b$ so as to make the alternate hypothesis *apparently* true. This is shown in the following figure,

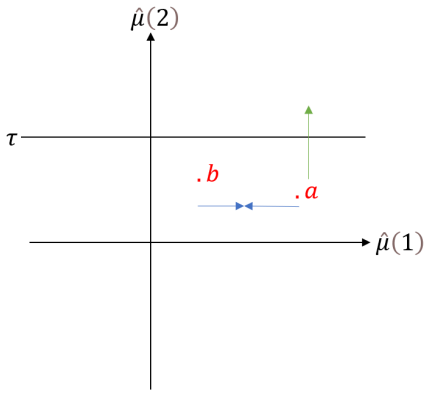


Figure 7.7: *Nearest* alternative possibilities that make $\mathcal{H}_1$ *apparently* true

Case 1. Keeping $\mu'_a(2), \mu'_b(2)$ at $\hat{\mu}_a(2)$ and $\hat{\mu}_b(2)$ respectively and bringing $\mu'_a(1), \mu'_b(1)$ to a common value, say, $\mu'_a(1) = \mu'_b(1) = x$. (This case is represented by the blue arrows in the figure.)

$$\max_{\mathcal{H}_1}\left[\log\left(\mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\ \mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)\right)\right] =$$
$$\max_x\left[-\frac{1}{2}\left\{\sum_{s\in[t]:A_s=a}(X_s(1)-x)^2+\sum_{s\in[t]:A_s=b}(X_s(1)-x)^2\right\}\right]$$
$$-\frac{1}{2}\left\{\sum_{s\in[t]:A_s=a}(X_s(2)-\hat{\mu}_a(2))^2+\sum_{s\in[t]:A_s=b}(X_s(2)-\hat{\mu}_b(2))^2\right\}$$

Maximizing w.r.t. x,

$$\frac{\partial\left[-\frac{1}{2}\left\{\sum_{s\in[t]:A_s=a}(X_s(1)-x)^2+\sum_{s\in[t]:A_s=b}(X_s(1)-x)^2\right\}\right]}{\partial x}=-\sum_{s\in[t]:A_s=a}(X_s(1)-x)$$
$$-\sum_{s\in[t]:A_s=b}(X_s(1)-x)=0\implies x=\frac{\sum_{s\in[t]:A_s=a}X_s(1)+\sum_{s\in[t]:A_s=b}X_s(1)}{N_a(t)+N_b(t)}=:\hat{\mu}_{a,b}(t)$$

Therefore,

$$\max_{\mathcal{H}_1}\left[\log\left(\mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\ \mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)\right)\right] =$$
$$-\frac{1}{2}\sum_{s\in[t]:A_s=a}\left\{(X_s(1)-\hat{\mu}_{a,b}(1))^2+(X_s(2)-\hat{\mu}_a(2))^2\right\}$$
$$-\frac{1}{2}\sum_{s\in[t]:A_s=b}\left\{(X_s(1)-\hat{\mu}_{a,b}(1))^2+(X_s(2)-\hat{\mu}_b(2))^2\right\}$$

Case 2. Keeping $\mu'_a(1),\mu'_b(1),\mu'_b(2)$ at $\hat{\mu}_a(1),\hat{\mu}_b(1)$ and $\hat{\mu}_b(2)$ respectively and taking $\mu'_a(2)$ just above the threshold, i.e., $\mu'_a(2)=\tau$. (This case is represented by the green arrow in the figure.)

$$\max_{\mathcal{H}_1}\left[\log\left(\mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\ \mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)\right)\right] =$$
$$-\frac{1}{2}\sum_{s\in[t]:A_s=a}\left\{(X_s(1)-\hat{\mu}_a(1))^2+(X_s(2)-\tau)^2\right\}$$
$$-\frac{1}{2}\sum_{s\in[t]:A_s=b}\left\{(X_s(1)-\hat{\mu}_b(1))^2+(X_s(2)-\hat{\mu}_b(2))^2\right\}$$

Combining Case 1 and Case 2, we get,

$$\max_{\mathcal{H}_1} \left[\log\left(\mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)\right)\right] = \max\left(Case\ 1,\ Case\ 2\right)$$

Therefore,

$$\begin{aligned} Z_{a,b,\gamma}(t) = &-\frac{1}{2}\sum_{s\in[t]:A_s=a}\left\{(X_s(1)-\hat{\mu}_a(1))^2 + (X_s(2)-\hat{\mu}_a(2))^2\right\} \\ &-\frac{1}{2}\sum_{s\in[t]:A_s=b}\left\{(X_s(1)-\hat{\mu}_b(1))^2 + (X_s(2)-\hat{\mu}_b(2))^2\right\} \\ &-\max\left(Case\ 1,\ Case\ 2\right) \end{aligned}$$

Observe that,

$$(X_s(i)-\mu'_k(i))^2 - (X_s(1)-\hat{\mu}_k(i))^2 = (2X_s(i)-\mu'_k(i)-\hat{\mu}_k(i))\,(\hat{\mu}_k(i)-\mu'_k(i))$$

for all $k \in \{a,b\}$ and $i \in \{1,2\}$, irrespective of the particular value of $\mu'_k(i)$. Hence,

$$\begin{aligned} \frac{1}{2}\sum_{s\in[t]:A_s=k}&\left\{(X_s(i)-\mu'_k(i))^2 - (X_s(1)-\hat{\mu}_k(i))^2\right\} \\ &= \frac{1}{2}\sum_{s\in[t]:A_s=k}(2X_s(i)-\mu'_k(i)-\hat{\mu}_k(i))\,(\hat{\mu}_k(i)-\mu'_k(i)) \\ &= \frac{N_k(t)}{2}\,(\hat{\mu}_k(i)-\mu'_k(i))^2 \end{aligned}$$

Therefore, we can simplify $Z_{a,b,\gamma}(t)$ as follows,

$$Z_{a,b,\gamma}(t) = \min\left\{\frac{N_a(t)}{2}(\hat{\mu}_a(1)-\hat{\mu}_{a,b}(1))^2 + \frac{N_b(t)}{2}(\hat{\mu}_b(1)-\hat{\mu}_{a,b}(1))^2,\ \frac{N_a(t)}{2}(\hat{\mu}_a(2)-\tau)^2\right\}$$

2. $\hat{\mu}_a(2) \le \tau, \hat{\mu}_b(2) \ge \tau, \hat{\mu}_a(1) \ge \hat{\mu}_b(1)$ **and** $\hat{\gamma} = 0$

Using similar ideas as the previous subsection, the *nearest* alternative possibilities that make $\mathcal{H}_1$ *apparently* true are shown in the following figure,

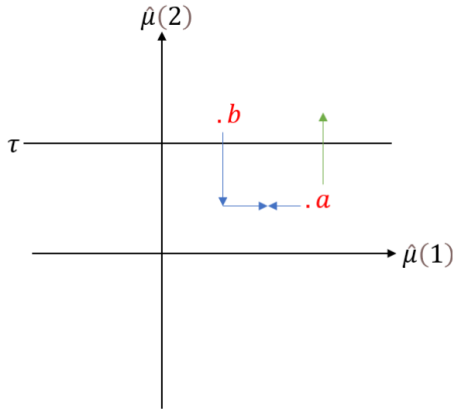


Figure 7.8: *Nearest* alternative possibilities that make $\mathcal{H}_1$ *apparently* true

Case 1. Keeping $\mu'_a(2)$ at $\hat{\mu}_a(2)$, bringing $\mu'_b(2)$ just below the threshold, i.e, $\mu'_b(2) = \tau$ and bringing $\mu'_a(1), \mu'_b(1)$ to a common value, say, $\mu'_a(1) = \mu'_b(1) = x$. (This case is represented by the blue arrows in the figure.)

$$\max_{\mathcal{H}_1} \left[\log\left(\mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\ \mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)\right)\right] =$$
$$-\frac{1}{2}\sum_{s\in[t]:A_s=a}\left\{(X_s(1)-\hat{\mu}_{a,b}(1))^2+(X_s(2)-\hat{\mu}_a(2))^2\right\}$$
$$-\frac{1}{2}\sum_{s\in[t]:A_s=b}\left\{(X_s(1)-\hat{\mu}_{a,b}(1))^2+(X_s(2)-\tau)^2\right\}$$

Case 2. Keeping $\mu'_a(1), \mu'_b(1), \mu'_b(2)$ at $\hat{\mu}_a(1), \hat{\mu}_b(1)$ and $\hat{\mu}_b(2)$ respectively and taking $\mu'_a(2)$ just above the threshold, i.e., $\mu'_a(2) = \tau$. (This case is represented by the green arrow in the figure.)

$$\max_{\mathcal{H}_1} \left[\log\left(\mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\ \mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)\right)\right] =$$
$$-\frac{1}{2}\sum_{s\in[t]:A_s=a}\left\{(X_s(1)-\hat{\mu}_a(1))^2+(X_s(2)-\tau)^2\right\}$$
$$-\frac{1}{2}\sum_{s\in[t]:A_s=b}\left\{(X_s(1)-\hat{\mu}_b(1))^2+(X_s(2)-\hat{\mu}_b(2))^2\right\}$$

Combining Case 1 and Case 2, we get,

$$\max_{\mathcal{H}_1}\left[\log\left(\mathcal{L}_{\nu_a'}\left(\bar{X}^a_{N_a(t)}\right)\mathcal{L}_{\nu_b'}\left(\bar{X}^b_{N_b(t)}\right)\right)\right] = \max\left(Case\ 1,\ Case\ 2\right)$$

Hence,

$$\begin{aligned} Z_{a,b,\gamma}(t) = &-\frac{1}{2}\sum_{s\in[t]:A_s=a}\left\{\left(X_s(1)-\hat{\mu}_a(1)\right)^2+\left(X_s(2)-\hat{\mu}_a(2)\right)^2\right\} \\ &-\frac{1}{2}\sum_{s\in[t]:A_s=b}\left\{\left(X_s(1)-\hat{\mu}_b(1)\right)^2+\left(X_s(2)-\hat{\mu}_b(2)\right)^2\right\} \\ &-\max\left(Case\ 1,\ Case\ 2\right) \end{aligned}$$

Therefore, upon simplification,

$$Z_{a,b,\gamma}(t) = \min\left\{\frac{N_a(t)}{2}(\hat{\mu}_a(1)-\hat{\mu}_{a,b}(1))^2+\frac{N_b(t)}{2}(\hat{\mu}_b(1)-\hat{\mu}_{a,b}(1))^2+\frac{N_b(t)}{2}(\hat{\mu}_b(2)-\tau)^2,\ \frac{N_a(t)}{2}(\hat{\mu}_a(2)-\tau)^2\right\}$$

3. $\hat{\mu}_a(2)\le\tau, \hat{\mu}_b(2)\ge\tau, \hat{\mu}_a(1)\le\hat{\mu}_b(1)$ **and** $\hat{\gamma}=0$

For the possibility considered in this subsection, the *nearest* alternative possibilities that make $\mathcal{H}_1$ *apparently* true are shown in the following figure,

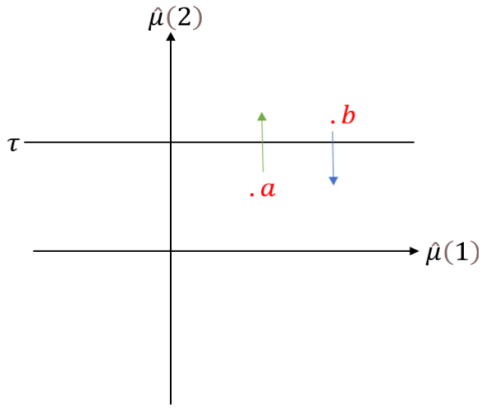


Figure 7.9: *Nearest* alternative possibilities that make $\mathcal{H}_1$ *apparently* true

Case 1. Keeping $\mu'_a(1), \mu'_a(2), \mu'_b(1)$ at $\hat{\mu}_a(1), \hat{\mu}_a(2)$ and $\hat{\mu}_b(1)$ respectively and bringing $\mu'_b(2)$ just below the threshold, i.e., $\mu'_b(2) = \tau$. (This case is represented by the green arrow in the figure.)

$$\begin{aligned}\max_{\mathcal{H}_1} \left[\log\left(\mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\ \mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)\right)\right] = \\ -\frac{1}{2}\sum_{s\in[t]:A_s=a}\left\{\left(X_s(1)-\hat{\mu}_a(1)\right)^2+\left(X_s(2)-\hat{\mu}_a(2)\right)^2\right\} \\ -\frac{1}{2}\sum_{s\in[t]:A_s=b}\left\{\left(X_s(1)-\hat{\mu}_b(1)\right)^2+\left(X_s(2)-\tau\right)^2\right\}\end{aligned}$$

Case 2. Keeping $\mu'_a(1), \mu'_b(1), \mu'_b(2)$ at $\hat{\mu}_a(1), \hat{\mu}_b(1)$ and $\hat{\mu}_b(2)$ respectively and taking $\mu'_a(2)$ just above the threshold, i.e., $\mu'_a(2) = \tau$. (This case is represented by the green arrow in the figure.)

$$\begin{aligned}\max_{\mathcal{H}_1} \left[\log\left(\mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\ \mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)\right)\right] = \\ -\frac{1}{2}\sum_{s\in[t]:A_s=a}\left\{\left(X_s(1)-\hat{\mu}_a(1)\right)^2+\left(X_s(2)-\tau\right)^2\right\} \\ -\frac{1}{2}\sum_{s\in[t]:A_s=b}\left\{\left(X_s(1)-\hat{\mu}_b(1)\right)^2+\left(X_s(2)-\hat{\mu}_b(2)\right)^2\right\}\end{aligned}$$

Combining Case 1 and Case 2, we get,

$$\max_{\mathcal{H}_1} \left[\log\left(\mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)\right)\right] = \max\left(Case\ 1,\ Case\ 2\right)$$

Hence,

$$\begin{aligned}Z_{a,b,\gamma}(t) = -\frac{1}{2}\sum_{s\in[t]:A_s=a}\left\{\left(X_s(1)-\hat{\mu}_a(1)\right)^2+\left(X_s(2)-\hat{\mu}_a(2)\right)^2\right\} \\ -\frac{1}{2}\sum_{s\in[t]:A_s=b}\left\{\left(X_s(1)-\hat{\mu}_b(1)\right)^2+\left(X_s(2)-\hat{\mu}_b(2)\right)^2\right\} \\ -\max\left(Case\ 1,\ Case\ 2\right)\end{aligned}$$

Therefore, upon simplification,

$$Z_{a,b,\gamma}(t) = \min\left\{\frac{N_b(t)}{2}(\hat{\mu}_b(2)-\tau)^2,\ \frac{N_a(t)}{2}(\hat{\mu}_a(2)-\tau)^2\right\}$$

4. $\hat{\mu}_a(2) \leq \hat{\mu}_b(2)$ **and** $\hat{\gamma} = 1$

For the possibility considered in this subsection, the *nearest* alternative possibilities that make $\mathcal{H}_1$ *apparently* true are shown in the following figure,

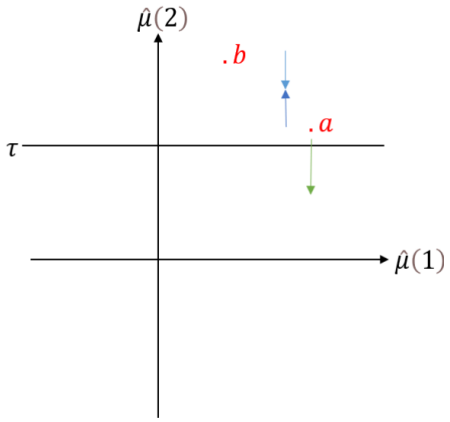


Figure 7.10: *Nearest* alternative possibilities that make $\mathcal{H}_1$ *apparently* true

Case 1. Keeping $\mu'_a(1), \mu'_b(1)$ at $\hat{\mu}_a(1)$ and $\hat{\mu}_b(1)$ respectively and bringing $\mu'_a(2), \mu'_b(2)$ to a common value, say, $\mu'_a(2) = \mu'_b(2) = x$. (This case is represented by the blue arrows in the figure.)

$$\max_{\mathcal{H}_1}\left[\log\left(\mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\ \mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)\right)\right] =$$
$$-\frac{1}{2}\sum_{s\in[t]:A_s=a}\left\{(X_s(1)-\hat{\mu}_a(1))^2+(X_s(2)-\hat{\mu}_{a,b}(2))^2\right\}$$
$$-\frac{1}{2}\sum_{s\in[t]:A_s=b}\left\{(X_s(1)-\hat{\mu}_b(1))^2+(X_s(2)-\hat{\mu}_{a,b}(2))^2\right\}$$

Case 2. Keeping $\mu'_a(1), \mu'_b(1), \mu'_b(2)$ at $\hat{\mu}_a(1), \hat{\mu}_b(1)$ and $\hat{\mu}_b(2)$ respectively and bringing $\mu'_a(2)$ just below the threshold, i.e., $\mu'_a(2) = \tau$. (This case is represented by the green arrow in the figure.)

$$\max_{\mathcal{H}_1}\left[\log\left(\mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\ \mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)\right)\right] = \\ -\frac{1}{2}\sum_{s\in[t]:A_s=a}\left\{(X_s(1)-\hat{\mu}_a(1))^2+(X_s(2)-\tau)^2\right\} \\ -\frac{1}{2}\sum_{s\in[t]:A_s=b}\left\{(X_s(1)-\hat{\mu}_b(1))^2+(X_s(2)-\hat{\mu}_b(2))^2\right\}$$

Combining Case 1 and Case 2, we get,

$$\max_{\mathcal{H}_1}\left[\log\left(\mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)\right)\right] = \max\left(Case\ 1,\ Case\ 2\right)$$

Hence,

$$Z_{a,b,\gamma}(t) = -\frac{1}{2}\sum_{s\in[t]:A_s=a}\left\{(X_s(1)-\hat{\mu}_a(1))^2+(X_s(2)-\hat{\mu}_a(2))^2\right\} \\ -\frac{1}{2}\sum_{s\in[t]:A_s=b}\left\{(X_s(1)-\hat{\mu}_b(1))^2+(X_s(2)-\hat{\mu}_b(2))^2\right\} \\ -\max\left(Case\ 1,\ Case\ 2\right)$$

Therefore, upon simplification,

$$Z_{a,b,\gamma}(t) = \min\left\{\frac{N_a(t)}{2}(\hat{\mu}_a(2)-\hat{\mu}_{a,b}(2))^2+\frac{N_b(t)}{2}(\hat{\mu}_b(2)-\hat{\mu}_{a,b}(2))^2,\ \frac{N_a(t)}{2}(\hat{\mu}_a(2)-\tau)^2\right\}$$

We shall now summarize the above results in form of a table

Table 7.1: $Z_{a,b,\gamma}(t)$ for various cases

Cases *apparently* satisfying $\mathcal{H}_0$	$Z_{a,b,\gamma}(t)$
$\hat{\mu}_a(2), \hat{\mu}_b(2) \leq \tau$, $\hat{\mu}_a(1) \geq \hat{\mu}_b(1)$ and $\hat{\gamma} = 0$	$Z_{a,b,\gamma}(t) = \min\left\{ \frac{N_a(t)}{2}(\hat{\mu}_a(1) - \hat{\mu}_{a,b}(1))^2 + \frac{N_b(t)}{2}(\hat{\mu}_b(1) - \hat{\mu}_{a,b}(1))^2, \frac{N_a(t)}{2}(\hat{\mu}_a(2) - \tau)^2 \right\}$
$\hat{\mu}_a(2) \leq \tau$, $\hat{\mu}_b(2) \geq \tau$, $\hat{\mu}_a(1) \geq \hat{\mu}_b(1)$ and $\hat{\gamma} = 0$	$Z_{a,b,\gamma}(t) = \min\left\{ \frac{N_a(t)}{2}(\hat{\mu}_a(1) - \hat{\mu}_{a,b}(1))^2 + \frac{N_b(t)}{2}(\hat{\mu}_b(1) - \hat{\mu}_{a,b}(1))^2 + \frac{N_b(t)}{2}(\hat{\mu}_b(2) - \tau)^2, \frac{N_a(t)}{2}(\hat{\mu}_a(2) - \tau)^2 \right\}$
$\hat{\mu}_a(2) \leq \tau$, $\hat{\mu}_b(2) \geq \tau$, $\hat{\mu}_a(1) \leq \hat{\mu}_b(1)$ and $\hat{\gamma} = 0$	$Z_{a,b,\gamma}(t) = \min\left\{ \frac{N_b(t)}{2}(\hat{\mu}_b(2) - \tau)^2, \frac{N_a(t)}{2}(\hat{\mu}_a(2) - \tau)^2 \right\}$
$\hat{\mu}_a(2) \leq \hat{\mu}_b(2)$ and $\hat{\gamma} = 1$	$Z_{a,b,\gamma}(t) = \min\left\{ \frac{N_a(t)}{2}(\hat{\mu}_a(2) - \hat{\mu}_{a,b}(2))^2 + \frac{N_b(t)}{2}(\hat{\mu}_b(2) - \hat{\mu}_{a,b}(2))^2, \frac{N_a(t)}{2}(\hat{\mu}_a(2) - \tau)^2 \right\}$

7.2.2 From $Z_{a,b,\gamma}(t)$ to the Stopping Rule

We will now look at what happens when the hypothesis $\mathcal{H}_0$ is *apparently* false.

1. Consider the case where the sub-instance formed by any two arms a and b has an *apparent* feasibility criterion, $\hat{\gamma}_t$, after t rounds. Let us calculate $Z_{a,b,\hat{\gamma}_t^c}(t)$. In this case, we observe that,

$$\max_{\mathcal{H}_1} \left[\log\left(\mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\ \mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)\right)\right] = \\ -\frac{1}{2}\sum_{s\in[t]:A_s=a}\left\{(X_s(1)-\hat{\mu}_a(1))^2+(X_s(2)-\hat{\mu}_a(2))^2\right\} \\ -\frac{1}{2}\sum_{s\in[t]:A_s=b}\left\{(X_s(1)-\hat{\mu}_b(1))^2+(X_s(2)-\hat{\mu}_b(2))^2\right\}$$

   Therefore,

$$\begin{aligned} Z_{a,b,\hat{\gamma}_t^c}(t) =& \max_{\mathcal{H}_0}\left[\log\left(\mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)\right)\right] - \max_{\mathcal{H}_1}\left[\log\left(\mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\mathcal{L}_{\nu'_b}\left(\bar{X}^b_{N_b(t)}\right)\right)\right] \\ =& -\frac{1}{2}\sum_{s\in[t]:A_s=a}\left\{(X_s(1)-\mu'_a(1))^2+(X_s(2)-\mu'_a(2))^2\right\} \\ & -\frac{1}{2}\sum_{s\in[t]:A_s=b}\left\{(X_s(1)-\mu'_b(1))^2+(X_s(2)-\mu'_b(2))^2\right\} \\ & +\frac{1}{2}\sum_{s\in[t]:A_s=a}\left\{(X_s(1)-\hat{\mu}_a(1))^2+(X_s(2)-\hat{\mu}_a(2))^2\right\} \\ & +\frac{1}{2}\sum_{s\in[t]:A_s=b}\left\{(X_s(1)-\hat{\mu}_b(1))^2+(X_s(2)-\hat{\mu}_b(2))^2\right\} \\ =& -\frac{N_a(t)}{2}\left\{(\hat{\mu}_a(1)-\mu'_a(1))^2+(\hat{\mu}_a(2)-\mu'_a(2))^2\right\} \\ & -\frac{N_b(t)}{2}\left\{(\hat{\mu}_b(1)-\mu'_b(1))^2+(\hat{\mu}_b(2)-\mu'_b(2))^2\right\} \end{aligned}$$

   which is clearly non-positive.

2. Now, let us consider the sub-instance formed by the *apparently* optimal arm after t rounds, $\hat{a}_t$, and any other arm $a \in \mathcal{A}\setminus\{\hat{a}_t\}$. Let the *apparent* criterion of this sub-instance be $\hat{\gamma}_t$. Clearly, $\hat{\gamma}_t$ is also the *apparent* feasibility criterion of the arm $\hat{a}_t$ and, thus, the original instance. Therefore, $\hat{\theta}_t = \hat{\gamma}_t$. Let us calculate $Z_{a,\hat{a}_t,\hat{\gamma}_t}$. Similar to the above, we

observe that,

$$\max_{\mathcal{H}_1}\left[\log\left(\mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\ \mathcal{L}_{\nu'_{\hat{a}_t}}\left(\bar{X}^{\hat{a}_t}_{N_{\hat{a}_t}(t)}\right)\right)\right] = -\frac{1}{2}\sum_{s\in[t]:A_s=a}\left\{(X_s(1)-\hat{\mu}_a(1))^2+(X_s(2)-\hat{\mu}_a(2))^2\right\} -\frac{1}{2}\sum_{s\in[t]:A_s=\hat{a}_t}\left\{(X_s(1)-\hat{\mu}_{\hat{a}_t}(1))^2+(X_s(2)-\hat{\mu}_{\hat{a}_t}(2))^2\right\}$$

Therefore,

$$Z_{a,\hat{a}_t,\hat{\gamma}_t} = \max_{\mathcal{H}_0}\left[\log\left(\mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\mathcal{L}_{\nu'_{\hat{a}_t}}\left(\bar{X}^{\hat{a}_t}_{N_{\hat{a}_t}(t)}\right)\right)\right] - \max_{\mathcal{H}_1}\left[\log\left(\mathcal{L}_{\nu'_a}\left(\bar{X}^a_{N_a(t)}\right)\mathcal{L}_{\nu'_{\hat{a}_t}}\left(\bar{X}^{\hat{a]}_t}_{N_{\hat{a}_t}(t)}\right)\right)\right]$$
$$= -\frac{N_a(t)}{2}\left\{(\hat{\mu}_a(1)-\mu'_a(1))^2+(\hat{\mu}_a(2)-\mu'_a(2))^2\right\} - \frac{N_{\hat{a}_t}(t)}{2}\left\{(\hat{\mu}_{\hat{a}_t}(1)-\mu'_{\hat{a}_t}(1))^2+(\hat{\mu}_{\hat{a}_t}(2)-\mu'_{\hat{a}_t}(2))^2\right\}$$

which, again, is non-positive.

Intuitively, the stopping rule can be written as follows:

$$T_\delta = \inf\left\{t\in\mathbb{N}:\exists(a,\theta)\in\mathcal{A}\times\{0,1\},\forall b\in\mathcal{A}\setminus\{a\}, Z_{a,b,\theta}>\beta(t,\delta)\right\}$$
$$= \inf\left\{t\in\mathbb{N}: Z(t) := \max_{a,\theta}\min_{b\in\mathcal{A}\setminus\{a\}} Z_{a,b,\theta} > \beta(t,\delta)\right\}$$

Observe that after t rounds, $\min_{b\in\mathcal{A}\setminus\{a\}} Z_{a,b,\theta}$ is non-negative if and only if $a=\hat{a}_t$ (since, for any arm $a\in\mathcal{A}\setminus\{\hat{a}_t\}$ and $\theta\in\{0,1\}$, $Z_{a,\hat{a}_t,\theta}$ is non-positive, as shown above) and $\theta=\hat{\theta}_t$ (since, for any sub-instance formed by arm $\hat{a}_t$ and arm $b\in\mathcal{A}\setminus\{\hat{a}_t\}$, $\hat{\gamma}_t=\hat{\theta}_t$ which is equal to the *apparent* feasibility criterion of the arm $\hat{a}_t$). Therefore, $Z(t)=\min_{b\in\mathcal{A}\setminus\{a\}} Z_{\hat{a}_t,b,\hat{\theta}_t}$ whenever there is a unique *apparently* optimal arm, $\hat{a}_t$ after t rounds. The *apparently* optimal ($\hat{a}_{T_\delta}$) arm after T_δ rounds and it's *apparent* feasibility criterion are the final decisions (*decision rule*).

7.3 Proof of δ-soundness

Theorem 5. *Let $\delta\in(0,1)$ and $\alpha>2$. Irrespective of the sampling strategy, there exists a constant $C=C(\alpha,K)$, such that using the above stopping rule with*

$$\beta(t,\delta)=\log\left(\frac{Ct^\alpha}{\delta^2}\right)$$

ensures that for all $\nu \in \mathcal{S}$, $\mathbb{P}_\nu(T_\delta < \infty, (\hat{a}_{T_\delta}, \hat{\theta}_{T_\delta}) \neq (a^*, \theta^*)) \leq \delta$

Proof. For some $a, b \in \mathcal{A}$, let the **apparent** sub-instance criterion be $\hat{\gamma}$. We know that,

$$\begin{aligned}
Z_{a,b,\hat{\gamma}}(t) = &-\frac{1}{2}\sum_{s\in[t]:A_s=a}\left\{(X_s(1)-\hat{\mu}_a(1))^2+(X_s(2)-\hat{\mu}_a(2))^2\right\}\\
&-\frac{1}{2}\sum_{s\in[t]:A_s=b}\left\{(X_s(1)-\hat{\mu}_b(1))^2+(X_s(2)-\hat{\mu}_b(2))^2\right\}\\
&-\max_{\mathcal{H}_1}\left[-\frac{1}{2}\sum_{s\in[t]:A_s=a}\left\{(X_s(1)-\mu'_a(1))^2+(X_s(2)-\mu'_a(2))^2\right\}\right]\\
&-\max_{\mathcal{H}_1}\left[-\frac{1}{2}\sum_{s\in[t]:A_s=b}\left\{(X_s(1)-\mu'_b(1))^2+(X_s(2)-\mu'_b(2))^2\right\}\right]
\end{aligned}$$

Upon simplification,

$$\begin{aligned}
Z_{a,b,\hat{\gamma}}(t) = &\min_{\mathcal{H}_1}\left[\frac{N_a(t)}{2}\left\{(\hat{\mu}_a(1)-\mu'_a(1))^2+(\hat{\mu}_a(2)-\mu'_a(2))^2\right\}\right]\\
&+\min_{\mathcal{H}_1}\left[\frac{N_b(t)}{2}\left\{(\hat{\mu}_b(1)-\mu'_b(1))^2+(\hat{\mu}_b(2)-\mu'_b(2))^2\right\}\right]
\end{aligned}$$

For every arm a, b that are such that the actual sub-instance formed by them satisfies the hypothesis $\mathcal{H}_1$, i.e., μ_a, μ_b and γ satisfy $\mathcal{H}_1$, but the *apparent* sub-instance after t rounds satisfies $\mathcal{H}_0$,

$$\begin{aligned}
Z_{a,b,\hat{\gamma}}(t) = &\min_{\mathcal{H}_1}\left[\frac{N_a(t)}{2}\left\{(\hat{\mu}_a(1)-\mu'_a(1))^2+(\hat{\mu}_a(2)-\mu'_a(2))^2\right\}\right]\\
&+\min_{\mathcal{H}_1}\left[\frac{N_b(t)}{2}\left\{(\hat{\mu}_b(1)-\mu'_b(1))^2+(\hat{\mu}_b(2)-\mu'_b(2))^2\right\}\right]\\
\leq &\left[\frac{N_a(t)}{2}\left\{(\hat{\mu}_a(1)-\mu_a(1))^2+(\hat{\mu}_a(2)-\mu_a(2))^2\right\}\right]\\
&+\left[\frac{N_b(t)}{2}\left\{(\hat{\mu}_b(1)-\mu_b(1))^2+(\hat{\mu}_b(2)-\mu_b(2))^2\right\}\right]
\end{aligned}$$

Now,

$$\begin{aligned}\mathbb{P}_\nu(T_\delta < \infty, (\hat{a}_{T_\delta}, \hat{\theta}_{T_\delta}) \neq (a^*, \theta^*)) &\leq \mathbb{P}_\nu(\exists (a,\theta) \in \mathcal{A} \setminus a^* \times \{0,1\}, \exists t \in \mathbb{N} : (\hat{\mu}_a, \hat{\mu}_a^*, \theta) \vdash \mathcal{H}_0, \\ &\quad Z_{a,a^*,\theta}(t) > \beta(t,\delta)) \\ &\leq \mathbb{P}_\nu(\exists t \in \mathbb{N} : \exists (a,\theta) \in \mathcal{A} \setminus a^* \times \{0,1\} : N_a(t) d(\hat{\mu}_a, \mu_a) \\ &\quad + N_{a^*}(t) d(\hat{\mu}_{a^*}, \mu_{a^*}) > \beta(t,\delta)) \\ &\leq \mathbb{P}_\nu(\exists t \in \mathbb{N} : \sum_{a=1}^{K} N_a(t) d(\hat{\mu}_a, \mu_a) > \beta(t,\delta))\end{aligned}$$

where $d(\mu_a, \mu_b)$ is the KL-divergence between two two-dimensional normal distributions with means μ_a and μ_b and same covariance matrix I_2. Let $d'(\mu_a, \mu_b)$ be the KL-divergence between two normal distributions (1-D) with means μ_a, μ_b and unit variance. We can write,

$$\sum_{a=1}^{K} N_a(t) d(\hat{\mu}_a, \mu_a) = \sum_{a=1}^{K} N_a(t) d'(\hat{\mu}_a(1), \mu_a(1)) + \sum_{a=1}^{K} N_a(t) d'(\hat{\mu}_a(2), \mu_a(2))$$

Therefore,

$$\begin{aligned}\mathbb{P}_\nu(\exists t \in \mathbb{N} : \sum_{a=1}^{K} N_a(t) d(\hat{\mu}_a, \mu_a) > \beta(t,\delta)) &\leq \mathbb{P}_\nu(\exists t \in \mathbb{N} : \sum_{a=1}^{K} N_a(t) d'(\hat{\mu}_a(1), \mu_a(1)) > \beta(t,\delta)/2) \\ &\quad + \mathbb{P}_\nu(\exists t \in \mathbb{N} : \sum_{a=1}^{K} N_a(t) d'(\hat{\mu}_a(2), \mu_a(2)) > \beta(t,\delta)/2) \\ &\leq 2 \sum_{t=1}^{\infty} e^{K+1} \left(\frac{\beta(t,\delta)^2 \log(t)}{4K} \right)^K e^{-\frac{\beta(t,\delta)}{2}}\end{aligned}$$

The above follows from union bounding arguments and a generalization of Theorem 2 of Magureanu et al. (2014) to single parameter Gaussian families. With exploration rate of the form $\beta(t,\delta) = \log(Ct^\alpha/\delta^2)$, for $\alpha > 2$, choosing C that satisfies,

$$\sum_{t=1}^{\infty} \frac{e^{K+1}}{(4K)^K} \frac{(\log^2(Ct^\alpha/\delta^2)\log(t))^K}{t^{\alpha/2}} \leq \frac{\sqrt{C}}{2}$$

gives a probability of error $\leq \delta$. □

Chapter 8

Numerical Experiments and Conclusions

8.1 Numerical Experiments

In this chapter we look at numerical experiments that were performed on four different bandit instances. We ran the Action elimination algorithm from part I and the track-and-stop algorithm from part II. The $\beta(t, \delta)$ (exploration rate) that was stated in the δ-soundness proof of the track-and-stop algorithm is highly over-conservative in practice. We instead used $\beta(t, \delta) = \log((\log(t)+1)/\delta)$ which has not yet been allowed in theory, but is still over-conservative in practice.

We performed a Monte-Carlo simulation of the two algorithms on a set four bandit instances over $N = 1000$ experiments. The following tables show the expected stopping times of the algorithms for three values of δ, i.e., $\delta = 0.1,\ 0.05,\ 0.01$.

Table 8.1: $\boldsymbol{\mu}(1) = [1.5, 1.4, 1.3, 1.2],\ \boldsymbol{\mu}(2) = [2.9, 2.7, 1.6, 2.54], \tau = 2.5$

Algorithm	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.01$
Action Elimination	21018	27786	29974
Track-and-stop	13198	14764	19545

Table 8.2: $\boldsymbol{\mu}(1) = [2.5, 2, 1.5, 1],\ \boldsymbol{\mu}(2) = [3, 2, 3, 1.5], \tau = 2.5$

Algorithm	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.01$
Action Elimination	125	133	149
Track-and-stop	103	119	159

Table 8.3: $\boldsymbol{\mu}(1) = [1.5, 1.4, 1.3, 1.2],\ \boldsymbol{\mu}(2) = [2.7, 1.4, 2.2, 2.6], \tau = 2.5$

Algorithm	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.01$
Action Elimination	8936	9457	10119
Track-and-stop	3974	4488	6085

Table 8.4: $\boldsymbol{\mu}(1) = [1.5, 1.42, 1.47, 1.38],\ \boldsymbol{\mu}(2) = [2.58, 2.41, 2.23, 2.64], \tau = 2.5$

Algorithm	$\delta = 0.1$	$\delta = 0.05$	$\delta = 0.01$
Action Elimination	39261	40236	43517
Track-and-stop	17873	17294	22451

8.2 Conclusions

- Both action elimination and track-and-stop are over conservative in practice.
- Upon running a single experiment on the track-and stop algorithm, we observed that the variance of the stopping times was very large. It is still unclear as to why this is the case.
- The track-and-stop algorithm generally performs better in terms of expected stopping time, except for the simplest case.
- As δ decreases to zero, the stopping time also increase.

Chapter 9

References

1. https://homes.cs.washington.edu/~jamieson/resources/bestArmSurvey.pdf
2. https://arxiv.org/pdf/2006.09649.pdf
3. https://link.springer.com/content/pdf/10.1007/3-540-45435-7_18.pdf
4. https://tor-lattimore.com/downloads/book/book.pdf
5. https://www.andrew.cmu.edu/course/10-703/textbook/BartoSutton.pdf
6. https://arxiv.org/abs/1602.04589
7. https://arxiv.org/abs/1405.4758