A Control-Treatment Proofs

A.1 IV Estimator Proof for Control-Treatment Setting

A.1.1 Model

We present our model as a system of linear equations expressed as such

$$\begin{cases} y_i = \theta x_i + g(u_i) + \varepsilon_i \\ x_i = f(u_i, z_i) \end{cases} \tag{19}$$

The recommendation z_i and the action choice x_i to be binary, i.e, either 0 or 1. This correspond to our control-treatment setting where the control action correspond to 0, and the treatment correspond to 1. The reward y_i is given by the action x_i, treatment effect θ, zero-mean noise ε and the unobserved confounding shift $g(u_i)$. Our goal is to consistently estimate the treatment effect θ. This is complicated by the unobserved confounding shift $g(u_i)$ and its correlation with the action choice x_i.

We model a linear relationship between the action x_i and the recommendation z_i by introducing two population probability that an agent choose the treatment given a recommendation γ_0 and γ_1, formally defined as

$$\gamma_0 = \mathbb{P}[x_i = 1 | z_i = 0] \qquad \text{and}$$
$$\gamma_1 = \mathbb{P}[x_i = 1 | z_i = 1]$$

Then, we can rewrite our equation modelling x_i as such:

$$\begin{aligned} x_i &= \gamma_1 z_i + \gamma_0(1 - z_i) + (f(u_i, z_i) - \gamma_1 z_i - \gamma_0(1 - z_i)) \\ &= \gamma_1 z_i + \gamma_0(1 - z_i) + \eta_i \\ &= \gamma z_i + \gamma_0 + \eta_i \end{aligned}$$

where $\eta_i = f(u_i, z_i) - \gamma_1 z_i - \gamma_0(1 - z_i)$ is a mean-zero random variable and $\gamma = \gamma_1 - \gamma_0$ is the population *compliance coefficient*. Now, we can rewrite the reward y_i as

$$\begin{aligned} y_i &= \theta\left(\gamma z_i + \gamma_0 + \eta_i\right) + g(u_i) + \varepsilon_i \\ &= \underbrace{\theta\,\gamma}_{\beta}\, z_i + \theta\gamma_0 + \theta\eta_t + g(u_i) + \varepsilon_i \end{aligned}$$

Let the operator $\bar{\cdot}$ denote the mean, such that $\bar{y} := \frac{1}{n}\sum_{i=1}^n y_i$ and $\bar{g} := \frac{1}{n}\sum_{i=1}^n g(u_i)$, etc. Then,

$$\bar{y} = \beta\bar{z} + \theta\gamma_0 + \theta\bar{\eta} + \bar{g} + \bar{\varepsilon}$$

Thus, the centered reward and treatment choice at round i are given as:

$$\begin{cases} y_i - \bar{y}_i = \beta(z_i - \bar{z}) + \theta(\eta_i - \bar{\eta}) + g(u_i) - \bar{g} + \varepsilon_i - \bar{\varepsilon} \\ x_i - \bar{x}_i = \gamma(z_i - \bar{z}) + \eta_i - \bar{\eta} \end{cases} \tag{20}$$

This formulation of the centered reward $y_i - \bar{y}$ allows us to express and bound the error between the treatment effect θ and its instrumental variable estimate $\hat{\theta}_n$, which we show in the following section.

A.1.2 Finite Sample Analysis

Given $n \in \mathbb{R}$ samples $(z, x, y)_n$, we would like to bound the difference between the predicted exogenous treatment effect, denoted $\hat{\theta}_n$, and the true exogenous treatment effect θ.

Theorem 2.1 (Treatment effect approximation bound)**.** *Given a sample set $(z_i, x_i, y_i)_n$, which contains n samples of instrument z, treatment x, and reward y, we bound the difference between the true treatment effect θ and the predicted treatment effect $\hat{\theta}_n$ derived via IV regression over $(z_i, x_i, y_i)_n$. Let Υ be a upper bound on the confounding term $g(u_i)$. For some confidence $\delta > 0$, with probability at least $1 - \delta$:*

$$\left|\hat{\theta}_n - \theta\right| \leq \frac{(2\sigma_\varepsilon + 4\Upsilon)\sqrt{2n\log(4/\delta)}}{\left|\sum_{i=1}^n (x_i - \bar{x})(z_i - \bar{z})\right|}.$$

Proof. We can estimate $\hat{\theta}_n$ via a Two-Stage Least Squares (2SLS). In the first stage, we regress $y_i - \bar{y}$ onto $z_i - \bar{z}$ to get the empirical estimate $\hat{\beta}_n$ and $x_i - \bar{x}$ onto $z_i - \bar{z}$ to get $\hat{\gamma}_n$ as such:

$$\hat{\beta}_n := \frac{\sum_{i=1}^{n}(y_i - \bar{y})(z_i - \bar{z})}{\sum_{i=1}^{n}(z_i - \bar{z})^2} \qquad \text{and} \qquad \hat{\gamma}_n := \frac{\sum_{i=1}^{n}(x_i - \bar{x})(z_i - \bar{z})}{\sum_{i=1}^{n}(z_i - \bar{z})^2} \tag{21}$$

Second, we take their quotient as the predicted treatment effect $\hat{\theta}_n$, i.e.

$$\begin{aligned} \hat{\theta}_n &= \frac{\hat{\beta}_n}{\hat{\gamma}_n} \qquad (22) \\ &= \frac{\sum_{i=1}^{n}(y_i - \bar{y})(z_i - \bar{z})}{\sum_{i=1}^{n}(z_i - \bar{z})^2} \frac{\sum_{i=1}^{n}(z_i - \bar{z})^2}{\sum_{i=1}^{n}(x_i - \bar{x})(z_i - \bar{z})} \\ &= \frac{\sum_{i=1}^{n}(y_i - \bar{y})(z_i - \bar{z})}{\sum_{i=1}^{n}(x_i - \bar{x})(z_i - \bar{z})} \qquad (23) \end{aligned}$$

Hence, the absolute value of the difference

$$\begin{aligned} \left|\hat{\theta}_n - \theta\right| &= \left|\frac{\sum_{i=1}^{n}(y_i - \bar{y})(z_i - \bar{z})}{\sum_{i=1}^{n}(x_i - \bar{x})(z_i - \bar{z})} - \theta\right| \\ &= \left|\frac{\sum_{i=1}^{n}\left(\theta\gamma(z_i - \bar{z}) + \theta(\eta_i - \bar{\eta}) + g(u_i) - \bar{g} + \varepsilon_i - \bar{\varepsilon}\right)(z_i - \bar{z})}{\sum_{i=1}^{n}(x_i - \bar{x})(z_i - \bar{z})} - \theta\right| \quad \text{(by eq. (20))} \\ &= \left|\theta + \frac{\sum_{i=1}^{n}\left(g(u_i) - \bar{g} + \varepsilon_i - \bar{\varepsilon}\right)(z_i - \bar{z})}{\sum_{i=1}^{n}(x_i - \bar{x})(z_i - \bar{z})} - \theta\right| \\ &= \frac{\left|\sum_{i=1}^{n}\left(g(u_i) - \bar{g} + \varepsilon_i - \bar{\varepsilon}\right)(z_i - \bar{z})\right|}{\left|\sum_{i=1}^{n}(x_i - \bar{x})(z_i - \bar{z})\right|} \\ &= \frac{\left|\sum_{i=1}^{n}\left(g(u_i) - \bar{g} + \varepsilon_i - \bar{\varepsilon}\right)(z_i - \bar{z})\right|}{\left|\sum_{i=1}^{n}(x_i - \bar{x})(z_i - \bar{z})\right|} \end{aligned}$$

Finally, we need to find an upper bound for the numerator $\left|\sum_{i=1}^{n}\left(g(u_i) - \bar{g} + \varepsilon_i - \bar{\varepsilon}\right)(z_i - \bar{z})\right|$. We do so in lemma A.1. □

Lemma A.1. *For some $\delta \in (0,1)$, with probability at least $1-\delta$, we have*

$$\left|\sum_{i=1}^{n}\left(g(u_i) - \bar{g} + \varepsilon_i - \bar{\varepsilon}\right)(z_i - \bar{z})\right| \le (2\sigma_\varepsilon + 4\Upsilon)\sqrt{2n\log(4/\delta)} \tag{24}$$

Proof.

$$\left|\sum_{i=1}^{n}\left(g(u_i)-\bar{g}+\varepsilon_i-\bar{\varepsilon}\right)(z_i-\bar{z})\right|$$

$$=\left|\sum_{i=1}^{n}\left(g(u_i)-\mathbb{E}[g(u)]+\mathbb{E}[g(u)]-\bar{g}+\varepsilon_i-\bar{\varepsilon}\right)(z_i-\bar{z})\right|$$

$$=\left|\sum_{i=1}^{n}\left(g(u_i)-\mathbb{E}[g(u)]+\varepsilon_i\right)z_i-\sum_{i=1}^{n}\left(g(u_i)-\mathbb{E}[g(u)]+\varepsilon_i\right)\bar{z}+\sum_{i=1}^{n}\left(\mathbb{E}[g(u)]-\bar{g}-\bar{\varepsilon}\right)(z_i-\bar{z})\right|$$

$$=\left|\sum_{i=1}^{n}\left(g(u_i)-\mathbb{E}[g(u)]+\varepsilon_i\right)z_i-\sum_{i=1}^{n}\left(g(u_i)-\mathbb{E}[g(u)]+\varepsilon_i\right)\bar{z}\right| \qquad \text{(since } \textstyle\sum_{i=1}^{n}(z_i-\bar{z})=0\text{)}$$

$$\leq\left|\sum_{i=1}^{n}\left(g(u_i)-\mathbb{E}[g(u)]+\varepsilon_i\right)z_i\right|+\left|\sum_{i=1}^{n}\left(g(u_i)-\mathbb{E}[g(u)]+\varepsilon_i\right)\right|$$

(by the triangle inequality and $|\bar{z}|\leq 1$)

$$\leq\left|\sigma_\varepsilon\sqrt{2n_1\log(1/\delta_1)}+2\Upsilon\sqrt{\frac{n_1\log(1/\delta_2)}{2}}\right|+\left|\sigma_\varepsilon\sqrt{2n\log(1/\delta_3)}+2\Upsilon\sqrt{\frac{n\log(1/\delta_4)}{2}}\right|$$

(by Chernoff Bound, where $n_1:=\sum_{i=1}^{n}z_i$)

$$\leq\left(2\sigma_\varepsilon+4\Upsilon\right)\sqrt{2n\log(4/\delta)}$$

(since $n_1\leq n$ and by Union Bound, where $\delta:=\delta_1+\delta_2+\delta_3+\delta_4$)

□

Corollary A.2 (Lower bound on $|\sum_{i=1}^{n}(x_i-\bar{x})(z_i-\bar{z})|$)**.** *Fix any phase. Let p_{BIC} be the proportion in the population of agents who are BIC in this phase. Then, the denominator of the approximation bound given by the samples of this phase*

$$\left|\sum_{i=1}^{n}(x_i-\bar{x})(z_i-\bar{z})\right|\geq\begin{cases}n\bar{z}(1-\bar{z}) & \textit{if the proportion } p_{BIC}=1;\\ \bar{z}\left(np_{BIC}(1-\bar{z})-\sqrt{\frac{n(1-\bar{z})\log(1/\delta)}{2}}\right) & \textit{with probability at least } 1-\delta \textit{ otherwise.}\end{cases}$$

Proof. Fix any phase. Recall that $\bar{z}$ is the probability of recommending arm 1 in this phase. Also let n_0 and n_1 be the number of samples with agents of type 0 and type 1, respectively. Let $n_{0,0}$ be the number of samples with agents of type 0 with recommendations of arm 0.

$$\begin{aligned}\left|\sum_{i=1}^{n}(x_i-\bar{x})(z_i-\bar{z})\right| &= \left|\sum_{i=1}^{n}(x_i-\bar{x})z_i-\bar{z}\sum_{i=1}^{n}(x_i-\bar{x})\right|\\ &=\left|\sum_{i=1}^{n}(1-\bar{x})\mathbb{1}[z_i=1]-\bar{z}\sum_{i=1}^{n}(x_i-\bar{x})\right|\end{aligned}$$

(since algorithms 1 & 2 are BIC for type 0, $z_i=1\Rightarrow x_i=1$; otherwise, $z_i=0$)

$$\begin{aligned}&=\left|\bar{z}n(1-\bar{x})-\bar{z}\sum_{i=1}^{n}x_i+\bar{z}\sum_{i=1}^{n}\bar{x}\right|\\ &=\left|\bar{z}n-\bar{z}n\bar{x}+\bar{z}n\bar{x}-\bar{z}\sum_{i=1}^{n}x_i\right|\\ &=\bar{z}\left|n-\sum_{i=1}^{n}x_i\right|\\ &=\bar{z}n_{0,0}\end{aligned}$$

Now, we need to lower bound $n_{0,0}$. Let p_{BIC} be the proportion in the population of agents who are BIC in this phase. Thus, in any round where the recommendation $z_i = 0$, the probability that an agent takes the recommendation is p_{BIC}. Thus, the expected value

$$\mathbb{E}[n_{0,0}] = np_{BIC}(1-\bar{z}).$$

Now, there are two cases:

1. If the proportion $p_{BIC} = 1$, i.e. in a phase when our recommendation is BIC for all agents, then $n_{0,0} = \mathbb{E}[n_{0,0}] = n(1-\bar{z})$. In this case, the denominator
$$\left|\sum_{i=1}^{n}(x_i - \bar{x})(z_i - \bar{z})\right| = n\bar{z}(1-\bar{z})$$
2. Otherwise, we may take a high probability bound on the number of agents $n_{0,0}$ that take our recommendation of arm 0. With probability at least $1-\delta$,
$$n_{0,0} \geq np_{BIC}(1-\bar{z}) - \sqrt{\frac{n(1-\bar{z})\log(1/\delta)}{2}}.$$
Therefore, with probability at least $1-\delta$,
$$\left|\sum_{i=1}^{n}(x_i - \bar{x})(z_i - \bar{z})\right| \geq \bar{z}\left(np_{BIC}(1-\bar{z}) - \sqrt{\frac{n(1-\bar{z})\log(1/\delta)}{2}}\right)$$
.

□

A.2 Sampling Stage Proofs

A.2.1 Sampling Stage Type 0 BIC Proof

Lemma 3.1 (Control-Treatment Sampling Stage BIC for Type 0)**.** *Algorithm 1 with parameters (ρ, ℓ_0, ℓ_1) completes in $\ell_1\rho + \ell_0 + \ell_1$ rounds. Algorithm 1 is BIC for all agents of type 0 if we hold the assumption above and the parameters satisfy:*

$$\rho \geq 1 + \frac{4(\mu_0^0 - \mu_0^1 + \nu\,\mathbb{P}_{\mathcal{P}_0}[\xi_0])}{\mathbb{P}_{\mathcal{P}_0}[\xi_0]}. \tag{11}$$

Proof. Recall that for type 0, the *fighting chance* event:

$$\xi_0 = \left\{\frac{1}{\ell_1}\sum_{t=1}^{\ell_1} y_t^1 - \frac{1}{\ell_0}\sum_{t=1}^{\ell_0} y_t^0 > 2\Upsilon + \sigma_\varepsilon\sqrt{\frac{2\log(1/\delta)}{\ell_0}} + \sigma_\varepsilon\sqrt{\frac{2\log(1/\delta)}{\ell_1}} + \frac{1}{2} + \nu\right\} \tag{25}$$

happens with probability at least $1-\delta$ for confidence $\delta > 0$.

Let $G := \mathop{\mathbb{E}}_{\mathcal{P}_0}[\theta | S_{1,\ell_0}, S_{2,\ell_1}]$. It is sufficient to show that $\mathop{\mathbb{E}}_{\mathcal{P}_0}[G|z_t = 0]\mathop{\mathbb{P}}_{\mathcal{P}_0}[z_t = 0] \geq 0$ (by claim A.3).

Claim A.3. *Assume that we have $\mathop{\mathbb{E}}_{\mathcal{P}_0}[\theta|z_t = 1]\mathop{\mathbb{P}}_{\mathcal{P}_0}[z_t = 1] \geq 0$. Then we also have $\mathop{\mathbb{E}}_{\mathcal{P}_0}[\theta|z_t = 0]\mathop{\mathbb{P}}_{\mathcal{P}_0}[z_t = 0] < 0$.*

Proof. Algorithm 1 is designed in a way such that we always recommend arm 1 with some positive possibility, i.e. $\mathbb{P}_{\mathcal{P}_0}[z_t = 1] > 0$.
Similarly, assume that we also recommend arm 0 with some positive probability: $\mathbb{P}_{\mathcal{P}_0}[z_t = 0] > 0$.
Assume that the expected difference in mean rewards $\mathop{\mathbb{E}}_{\mathcal{P}_0}[\theta] \leq 0$. Then, we have:

$$\begin{aligned} 0 &\geq \mathop{\mathbb{E}}_{\mathcal{P}_0}[\theta] \\ &= \mathop{\mathbb{E}}_{\mathcal{P}_0}[\theta|z_t = 0]\mathop{\mathbb{P}}_{\mathcal{P}_0}[z_t = 0] + \mathop{\mathbb{E}}_{\mathcal{P}_0}[\theta|z_t = 1]\mathop{\mathbb{P}}_{\mathcal{P}_0}[z_t = 1] \end{aligned}$$

Since, by assumption, the product $\mathbb{E}_{\mathcal{P}_0}[\theta|z_t = 1]\,\mathbb{P}[z_t = 1] \geq 0$, it must be that the product

$$\mathbb{E}_{\mathcal{P}_0}[\theta|z_t = 0] \mathbb{P}_{\mathcal{P}_0}[z_t = 0] \leq 0.$$

Hence, when given an recommendation for arm 0, the agent of type 0 will take our recommendation. □

Note that we have

$$\begin{aligned}
&\mathbb{E}_{\mathcal{P}_0}[G|z_t = 1] \mathbb{P}_{\mathcal{P}_0}[z_t = 1] \\
&= \mathbb{E}_{\mathcal{P}_0}[G|z_t = 1 \,\&\, p_t \in P - Q] \mathbb{P}_{\mathcal{P}_0}[z_t = 1 \,\&\, p_t \in P - Q] \\
&\quad + \mathbb{E}_{\mathcal{P}_0}[G|z_t = 1 \,\&\, p_t \in Q] \mathbb{P}_{\mathcal{P}_0}[z_t = 1 \,\&\, p_t \in Q] \\
&= \mathbb{E}_{\mathcal{P}_0}[G|\xi_0] \mathbb{P}_{\mathcal{P}_0}[\xi_0] \left(1 - \frac{1}{\rho}\right) + \mathbb{E}_{\mathcal{P}_0}[G|z_t = 1 \,\&\, p_t \in Q] \mathbb{P}_{\mathcal{P}_0}[z_t = 1 \,\&\, p_t \in Q] \\
&= \mathbb{E}_{\mathcal{P}_0}[G|\xi_0] \mathbb{P}_{\mathcal{P}_0}[\xi_0] \left(1 - \frac{1}{\rho}\right) + \mathbb{E}_{\mathcal{P}_0}[G|p_t \in Q] \mathbb{P}_{\mathcal{P}_0}[p_t \in Q], \quad \text{(by definition } p_t \in Q \Rightarrow z_t = 0\text{)} \\
&= \mathbb{E}_{\mathcal{P}_0}[G|\xi_0] \mathbb{P}_{\mathcal{P}_0}[\xi_0] \left(1 - \frac{1}{\rho}\right) + \mathbb{E}_{\mathcal{P}_0}[\theta|p_t \in Q] \mathbb{P}_{\mathcal{P}_0}[p_t \in Q] \quad \text{(by Law of Total Expectation)} \\
&= \mathbb{E}_{\mathcal{P}_0}[G|\xi_0] \mathbb{P}_{\mathcal{P}_0}[\xi_0] \left(1 - \frac{1}{\rho}\right) + \mathbb{E}_{\mathcal{P}_0}[\theta] \mathbb{P}_{\mathcal{P}_0}[p_t \in Q] \\
&= \mathbb{E}_{\mathcal{P}_0}[G|\xi_0] \mathbb{P}_{\mathcal{P}_0}[\xi_0] \left(1 - \frac{1}{\rho}\right) + (\mu_0^1 - \mu_0^0) \mathbb{P}_{\mathcal{P}_0}[p_t \in Q] \\
&= \mathbb{E}_{\mathcal{P}_0}[G|\xi_0] \mathbb{P}_{\mathcal{P}_0}[\xi_0] \left(1 - \frac{1}{\rho}\right) + \frac{1}{\rho}(\mu_0^1 - \mu_0^0)
\end{aligned}$$

For agents of type i to follow our recommendation for arm 2, we need to pick ρ:

$$\begin{aligned}
\mathbb{E}_{\mathcal{P}_0}[G|z_t = 1] \mathbb{P}_{\mathcal{P}_0}[z_t = 1] &\geq \phi(u_t) \\
\mathbb{E}_{\mathcal{P}_0}[G|\xi_0] \mathbb{P}_{\mathcal{P}_0}[\xi_0] \left(1 - \frac{1}{\rho}\right) + \frac{1}{\rho}(\mu_0^1 - \mu_0^0) &\geq \phi(u_t) \mathbb{P}_{\mathcal{P}_0}[\xi_0] \\
\mathbb{E}_{\mathcal{P}_0}[G|\xi_0] \mathbb{P}_{\mathcal{P}_0}[\xi_0] - \frac{1}{\rho} \mathbb{E}_{\mathcal{P}_0}[G|\xi_0] \mathbb{P}_{\mathcal{P}_0}[\xi_0] + \frac{1}{\rho}(\mu_0^1 - \mu_0^0) &\geq \phi(u_t) \mathbb{P}_{\mathcal{P}_0}[\xi_0] \\
\rho \mathbb{E}_{\mathcal{P}_0}[G|\xi_0] \mathbb{P}_{\mathcal{P}_0}[\xi_0] - \rho\phi(u_t) \mathbb{P}_{\mathcal{P}_0}[\xi_0] &\geq \mathbb{E}_{\mathcal{P}_0}[G|\xi_0] \mathbb{P}_{\mathcal{P}_0}[\xi_0] - (\mu_0^1 - \mu_0^0) \\
\rho &\geq 1 + \frac{\mu_0^0 - \mu_0^1 + \phi(u_t)\,\mathbb{P}_{\mathcal{P}_0}[\xi_0]}{\mathbb{E}_{\mathcal{P}_0}[G|\xi_0]\,\mathbb{P}_{\mathcal{P}_0}[\xi_0] - \phi(u_t)\,\mathbb{P}_{\mathcal{P}_0}[\xi_0]} \\
\rho &\geq 1 + \frac{\mu_0^0 - \mu_0^1 + \nu\,\mathbb{P}_{\mathcal{P}_0}[\xi_0]}{\mathbb{E}_{\mathcal{P}_0}[G|\xi_0]\,\mathbb{P}_{\mathcal{P}_0}[\xi_0] - \nu\,\mathbb{P}_{\mathcal{P}_0}[\xi_0]}
\end{aligned}$$

This comes down to finding a lower bound on the denominator of the expression above. First, we define clean events C_1 and C_2 where the average errors $\frac{1}{\ell_0}\sum_{i=1}^{\ell_0} \varepsilon_i$ and $\frac{1}{\ell_1}\sum_{i=1}^{\ell_1} \varepsilon_i$ are bounded (following Corollary C.5):

$$C_1 := \left\{ \left| \frac{1}{\ell_0} \sum_{i=1}^{\ell_0} \varepsilon_i \right| \leq \sigma_\varepsilon \sqrt{\frac{2\log(1/\delta)}{\ell_0}} \right\}. \tag{26}$$

$$C_2 := \left\{ \left| \frac{1}{\ell_1} \sum_{i=1}^{\ell_1} \varepsilon_i \right| \leq \sigma_\varepsilon \sqrt{\frac{2\log(1/\delta)}{\ell_1}} \right\}. \tag{27}$$

Define another clean event C' where both C_1 and C_2 happens simultaneously. The event C' occurs with probability at least $1 - \delta$ for $\delta < \delta' := \frac{\mathbb{P}_{\mathcal{P}_0}[\xi_0]}{8}$.

We have

$$
\begin{aligned}
\mathop{\mathbb{E}}_{\mathcal{P}_0}[G|\xi_0]\mathop{\mathbb{P}}_{\mathcal{P}_0}[\xi_0] &= \mathop{\mathbb{E}}_{\mathcal{P}_0}[G|\xi_0, C']\mathop{\mathbb{P}}_{\mathcal{P}_0}[\xi_0, C'] + \mathop{\mathbb{E}}_{\mathcal{P}_0}[G|\xi_0, \neg C']\mathop{\mathbb{P}}_{\mathcal{P}_0}[\xi_0, \neg C'] \\
&\geq \mathop{\mathbb{E}}_{\mathcal{P}_0}[G|\xi_0, C']\mathop{\mathbb{P}}_{\mathcal{P}_0}[\xi_0, C'] - \delta \qquad \text{(since } G \geq -1 \text{ and } \mathbb{P}_{\mathcal{P}_0}[\neg C'] < \delta) \\
&= \mathop{\mathbb{E}}_{\mathcal{P}_0}[G|\xi_0, C'](\mathop{\mathbb{P}}_{\mathcal{P}_0}[\xi_0] - \mathop{\mathbb{P}}_{\mathcal{P}_0}[\xi_0, \neg C']) - \delta \\
&\geq \mathop{\mathbb{E}}_{\mathcal{P}_0}[G|\xi_0, C'](\mathop{\mathbb{P}}_{\mathcal{P}_0}[\xi_0] - \mathop{\mathbb{P}}_{\mathcal{P}_0}[\neg C']) - \delta \\
&\geq \mathop{\mathbb{E}}_{\mathcal{P}_0}[G|\xi_0, C'](\mathop{\mathbb{P}}_{\mathcal{P}_0}[\xi_0] - \delta) - \delta \\
&= \mathop{\mathbb{E}}_{\mathcal{P}_0}[G|\xi_0, C']\mathop{\mathbb{P}}_{\mathcal{P}_0}[\xi_0] - \delta(1 + \mathop{\mathbb{E}}_{\mathcal{P}_0}[G|\xi_0, C']) \\
&\geq \mathop{\mathbb{E}}_{\mathcal{P}_0}[G|\xi_0, C']\mathop{\mathbb{P}}_{\mathcal{P}_0}[\xi_0] - 2\delta
\end{aligned} \tag{28}
$$

This comes down to finding a lower bound on the denominator of the expression above. We can reduce the dependency of the denominator to a single prior-dependent constant $\mathbb{P}_{\mathcal{P}_0}[\xi_0]$ if we lower bound the prior-dependent expected value $\mathbb{E}_{\mathcal{P}_0}[G|\xi_0]$. That way, assuming we know the prior and can calculate the probability of event ξ_0, we can pick an appropriate ρ to satisfy the BIC condition for all agents of type i. Remember that event $\xi_0 = \{\bar{y}^1_{\ell_0} + C < \bar{y}^2_{\ell_1}\}$ where $C = 2\Upsilon + \sigma_\varepsilon\sqrt{\frac{2\log(1/\delta)}{\ell_0}} + \sigma_\varepsilon\sqrt{\frac{2\log(1/\delta)}{\ell_0}} + \frac{1}{2} + \nu$ and $G := \mathop{\mathbb{E}}_{\mathcal{P}_0}[\theta|S_{1,\ell_0}, S_{2,\ell_1}]$. Then, the expected value

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{\mathcal{P}_0}[G|\xi_0, C'] \\
&= \mathop{\mathbb{E}}_{\mathcal{P}_0}[\mathop{\mathbb{E}}_{\mathcal{P}_0}[\theta|S_{1,\ell_0}, S_{2,\ell_1}]|\xi_0, C'] \\
&= \mathop{\mathbb{E}}_{\mathcal{P}_0}\left[\theta\middle|2\Upsilon + \sigma_\varepsilon\sqrt{\frac{2\log(1/\delta)}{\ell_0}} + \sigma_\varepsilon\sqrt{\frac{2\log(1/\delta)}{\ell_0}} + \frac{1}{2} + \nu + \bar{y}^0_{\ell_0} < \bar{y}^1_{\ell_1}, C'\right] \\
&= \mathop{\mathbb{E}}_{\mathcal{P}_0}\left[\theta\middle|2\Upsilon + \sigma_\varepsilon\sqrt{\frac{2\log(1/\delta)}{\ell_0}} + \sigma_\varepsilon\sqrt{\frac{2\log(1/\delta)}{\ell_0}} + \frac{1}{2} + \nu + \frac{1}{\ell_0}\sum_{t=1}^{\ell_0} y^0_t < \frac{1}{\ell_1}\sum_{t=1}^{\ell_1} y^1_t, C'\right] \\
&= \mathop{\mathbb{E}}_{\mathcal{P}_0}\Bigg[\theta\Bigg|2\Upsilon + \sigma_\varepsilon\sqrt{\frac{2\log(1/\delta)}{\ell_0}} + \sigma_\varepsilon\sqrt{\frac{2\log(1/\delta)}{\ell_0}} + \frac{1}{2} + \nu + \\
&\qquad\qquad \frac{1}{\ell_0}\sum_{t=1}^{\ell_0} g(u_t) + \varepsilon_t < \theta + \frac{1}{\ell_1}\sum_{t=1}^{\ell_1} g(u_t) + \varepsilon_t, C'\Bigg] \\
&> \mathop{\mathbb{E}}_{\mathcal{P}_0}\Bigg[\theta\Bigg|2\Upsilon + \sigma_\varepsilon\sqrt{\frac{2\log(1/\delta)}{\ell_0}} + \sigma_\varepsilon\sqrt{\frac{2\log(1/\delta)}{\ell_1}} + \frac{1}{2} + \nu - \\
&\qquad\qquad \Upsilon - \sigma_\varepsilon\sqrt{\frac{2\log(1/\delta)}{\ell_0}} < \theta + \Upsilon + \sigma_\varepsilon\sqrt{\frac{2\log(1/\delta)}{\ell_1}}\Bigg] \qquad (29) \\
&\qquad\qquad \text{(since } g(u_t) > -\Upsilon \text{ and } \tfrac{1}{\ell_1}\textstyle\sum_{t=1}^{\ell_1}\varepsilon_t > -\sigma_\varepsilon\sqrt{\frac{2\log(1/\delta)}{\ell_1}} \text{ by event } C'\text{)} \\
&> \mathop{\mathbb{E}}_{\mathcal{P}_0}\left[\theta\middle|\frac{1}{2} + \nu < \theta\right] \\
&> \frac{1}{2} + \nu
\end{aligned} \tag{30}
$$

Hence, the lower bound on the denominator is

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathcal{P}_0}[G|\xi_0] \mathop{\mathbb{P}}_{\mathcal{P}_0}[\xi_0] - \nu \mathop{\mathbb{P}}_{\mathcal{P}_0}[\xi_0] &\geq \mathop{\mathbb{E}}_{\mathcal{P}_0}[G|\xi_0, C'] \mathop{\mathbb{P}}_{\mathcal{P}_0}[\xi_0] - 2\delta - \nu \mathop{\mathbb{P}}_{\mathcal{P}_0}[\xi_0] && \text{(by Equation 28)} \\
&> \frac{1}{2} \mathop{\mathbb{P}}_{\mathcal{P}_0}[\xi_0] + \nu \mathop{\mathbb{P}}_{\mathcal{P}_0}[\xi_0] - 2\delta - \nu \mathop{\mathbb{P}}_{\mathcal{P}_0}[\xi_0] && \text{(by Equation 30)} \\
&= \frac{1}{4} \mathop{\mathbb{P}}_{\mathcal{P}_0}[\xi_0] + \frac{1}{4} \mathop{\mathbb{P}}_{\mathcal{P}_0}[\xi_0] - 2\delta && (31) \\
&= 2\delta' + \frac{\mathbb{P}_{\mathcal{P}_0}[\xi_0]}{4} - 2\delta && \text{(since } \delta' = \tfrac{1}{8} \mathbb{P}_{\mathcal{P}_0}[\xi_0]) \\
&\geq \frac{\mathbb{P}_{\mathcal{P}_0}[\xi_0]}{4} && \text{(since } \delta < \delta')
\end{aligned}$$

Hence, we can pick :

$$\rho \geq 1 + \frac{4(\mu_0^0 - \mu_0^1 + \nu \mathbb{P}_{\mathcal{P}_0}[\xi_0])}{\mathbb{P}_{\mathcal{P}_0}[\xi_0]}$$

to satisfy the BIC condition for all agents of type 0. □

A.2.2 Sampling Stage Estimation Bound

Theorem 3.2 (Sampling Stage Estimation Bound)**.** *With n total samples collected from algorithm 1 –run with exploration probability ρ large enough so that our recommendation BIC for type 0 agents (see lemma 3.1),– we form an estimate θ_n of the treatment effect θ. With probability at least $1 - \delta$,*

$$\left|\hat{\theta}_n - \theta\right| \leq \frac{2\rho\left(\sigma_\varepsilon + 2\Upsilon\right)\sqrt{2n\log(6/\delta)}}{\left(np_0(1-1/\rho) - \sqrt{\frac{n(1-1/\rho)\log(6/\delta)}{2}}\right)}$$

Proof. Recall that the mean recommendation $\bar{z} = 1/\rho$ in the sampling stage and the proportion of agents who are type 0 be p_0. By lemma 3.1, algorithm 1 is BIC for agents of type 0. Therefore, we may apply corollary A.2 where we let the proportion $p_{BIC} = p_0$. □

A.3 Racing Stage

A.3.1 Racing Stage Type 0 BIC proof

Lemma 4.1 (Control-Treatment Racing Stage BIC for Type 0)**.** *Fix some constant $\tau \in (0, 1)$. Remember that $\phi(u_t) \in [-\nu, \nu]$ for all u_t and $\mathbb{P}_{\mathcal{P}_0}[\theta > \tau]$ be the prior probability that $\theta > \tau$ for agents of type 0. Let the approximation bound s_L after the sampling stage be*

$$s_L := \frac{(2\sigma_\varepsilon + 4\Upsilon)\sqrt{2L\log(4/\delta)}}{\left|\sum_{i=1}^{L}(x_i - \bar{x})(z_i - \bar{z})\right|} \leq \frac{\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta > \tau]}{4} - \frac{\nu}{2}. \tag{13}$$

Then, during the first phase of the racing stage, the recommendations from algorithm 2 are BIC for agents of type 0.

Proof. Let the decision threshold s_q for a phase q be defined as in algorithm 2. Assume that after the elimination, at every iteration a sample of the eliminated arm is also drawn, but not revealed to the agent.

We define the event $\mathcal{C}_1$ as the accuracy guarantee of $\hat{\theta}$ such that:

$$\mathcal{C}_1 := \left\{\forall q \geq L : |\theta - \hat{\theta}_q| < s_q\right\} \tag{32}$$

where L is the number of samples after running the sampling stage.

Let X_q^a be the number of arm $a \in \{0, 1\}$ samples in phase q. The decision threshold s_q in phase q is unbounded if $X_q^a = 0$ for either arm a. We also define event $\mathcal{C}_2$ where there is at least one sample of each arm in each phase:

$$\mathcal{C}_2 := \left\{\forall q \geq L, a : X_q^a \geq 1\right\} \tag{33}$$

Recall that p_0 of the population is of type 0. Then, if algorithm 2 is BIC for all agents of type 0 and we recommend each arm at least h times in phase q, we have $X_q^0 > h$ and $X_q^0 \sim B(h, p_0)$. Hence, we have

$$\mathbb{P}[X_q^0 \geq 1] = 1 - p_0^h = \delta'. \tag{34}$$

Let $\tau \in (0,1)$. Fix phase $q \geq L$, and some agent t in this phase. In order to prove that algorithm 2 is BIC for all agents of type 0, we want to show that

$$\mathop{\mathbb{E}}_{\mathcal{P}_0}[\theta | z_t = 1] \mathop{\mathbb{P}}_{\mathcal{P}_0}[z_t = 1] \geq \phi(u_t). \tag{35}$$

Note that, by assumption, the parameter

$$\begin{aligned} h &\geq \frac{\log\left(\frac{3\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta\geq\tau]+4}{4\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta\geq\tau]+4}\right)}{\log(1-p_2)} \\ \Rightarrow \left(\frac{1}{2}\right)^h &\geq \frac{3\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \geq \tau] + 4}{4\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \geq \tau] + 4} \\ \Rightarrow \delta' = 1 - \left(\frac{1}{2}\right)^h &\leq \frac{\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \geq \tau]}{4\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \geq \tau] + 4}. \end{aligned}$$

From theorem 2.1, with probability $\delta > 0$ we have that

$$\mathbb{P}[\neg\mathcal{C}_1 | G] \leq \delta$$

Define event $\mathcal{C}$ such that both events $\mathcal{C}_1$ and $\mathcal{C}_2$ hold simultaneously in a round:

$$\mathcal{C} := \{\forall q \geq L_1 : \mathcal{C}_1 \;\&\; \mathcal{C}_2\} \tag{36}$$

Using union bound, we have

$$\begin{aligned} \mathbb{P}[\neg\mathcal{C}|G] &\leq \mathbb{P}[\neg\mathcal{C}_1|G] + \mathbb{P}[\neg\mathcal{C}_2|G] \qquad (37) \\ &\leq \delta + \delta' \\ &\leq \delta_\tau + \delta' \\ &\leq \frac{\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \geq \tau]}{4\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \geq \tau] + 4} + \frac{\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \geq \tau]}{4\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \geq \tau] + 4} \\ &= \frac{\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \geq \tau]}{2\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \geq \tau] + 2} \qquad (38) \end{aligned}$$

Therefore, since $\theta \geq -1$, we have:

$$\begin{aligned} \mathop{\mathbb{E}}_{\mathcal{P}_0}[\theta|z_t = 1] \mathop{\mathbb{P}}_{\mathcal{P}_0}[z_t = 1] &= \mathop{\mathbb{E}}_{\mathcal{P}_0}[\theta|z_t = 1, \mathcal{C}] \mathop{\mathbb{P}}_{\mathcal{P}_0}[z_t = 1, \mathcal{C}] + \mathop{\mathbb{E}}_{\mathcal{P}_0}[\theta|z_t = 1, \neg C] \mathop{\mathbb{P}}_{\mathcal{P}_0}[z_t = 1, \neg C] \\ &\geq \mathop{\mathbb{E}}_{\mathcal{P}_0}[\theta|z_t = 1, \mathcal{C}] \mathop{\mathbb{P}}_{\mathcal{P}_0}[z_t = 1, \mathcal{C}] - \frac{\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \geq \tau]}{2\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \geq \tau] + 2} \end{aligned}$$

We want to upper bound the first term. This can be done by splitting it into four cases based on the value of θ. We have:

$$\begin{aligned} \mathop{\mathbb{E}}_{\mathcal{P}_0}[\theta|z_t = 1, \mathcal{C}] \mathop{\mathbb{P}}_{\mathcal{P}_0}[z_t = 1, \mathcal{C}] = &\mathop{\mathbb{E}}_{\mathcal{P}_0}[\theta|z_t = 1, \mathcal{C}, \theta \geq \tau] \mathop{\mathbb{P}}_{\mathcal{P}_0}[z_t = 1, \mathcal{C}, \theta \geq \tau] \\ &+ \mathop{\mathbb{E}}_{\mathcal{P}_0}[\theta|z_t = 1, \mathcal{C}, 0 \leq \theta < \tau] \mathop{\mathbb{P}}_{\mathcal{P}_0}[z_t = 1, \mathcal{C}, 0 \leq \theta < \tau] \\ &+ \mathop{\mathbb{E}}_{\mathcal{P}_0}[\theta|z_t = 1, \mathcal{C}, -2s_q < \theta < 0] \mathop{\mathbb{P}}_{\mathcal{P}_0}[z_t = 1, \mathcal{C}, -2s_q < \theta < 0] \\ &+ \mathop{\mathbb{E}}_{\mathcal{P}_0}[\theta|z_t = 1, \mathcal{C}, \theta \leq -2s_q] \mathop{\mathbb{P}}_{\mathcal{P}_0}[z_t = 1, \mathcal{C}, \theta \leq -2s_q] \end{aligned} \tag{39}$$

By definition of s_q, we have:

$$\begin{aligned} 2s_q &\le 2s_L \\ &\le \frac{\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \ge \tau]}{2} - \nu \\ &\le \tau \end{aligned}$$

Conditional on $\mathcal{C}$, the empirical estimate $\hat{\theta}_q > \theta - s_q \ge 2s_q$. Hence, when we have $\theta \ge \tau \ge 2s_q$, then $\hat{\theta} \ge s_q$ and arm 0 must have already been eliminated at phase $q \ge L_1$. This implies that the probability $\mathbb{P}_{\mathcal{P}_0}[z_t = 1, \mathcal{C}, \theta \ge \tau] = \mathbb{P}_{\mathcal{P}_0}[\mathcal{C}, \theta \ge \tau]$. Similarly, when we have $\theta \le -2s_q$, then $\hat{\theta} \le -s_q$ and arm 1 must have already been eliminated at phase $q \ge L_1$. Hence, arm 1 could not be recommended in that case and the probability $\mathbb{P}_{\mathcal{P}_0}[z_t = 1, \mathcal{C}, \theta \le -2s_q] = 0$.
We can then rewrite equation (39) as

$$\begin{aligned} \mathop{\mathbb{E}}_{\mathcal{P}_0}[\theta|z_t = 1, \mathcal{C}] \mathop{\mathbb{P}}_{\mathcal{P}_0}[z_t = 1, \mathcal{C}] \ge& \mathop{\mathbb{E}}_{\mathcal{P}_0}[\theta|z_t = 1, \mathcal{C}, \theta \ge \tau] \mathop{\mathbb{P}}_{\mathcal{P}_0}[z_t = 1, \mathcal{C}, \theta \ge \tau] \\ &+ \mathop{\mathbb{E}}_{\mathcal{P}_0}[\theta|z_t = 1, \mathcal{C}, 0 \le \theta < \tau] \mathop{\mathbb{P}}_{\mathcal{P}_0}[z_t = 1, \mathcal{C}, 0 \le \theta < \tau] \\ &+ \mathop{\mathbb{E}}_{\mathcal{P}_0}[\theta|z_t = 1, \mathcal{C}, -2s_q < \theta < 0] \mathop{\mathbb{P}}_{\mathcal{P}_0}[z_t = 1, \mathcal{C}, -2s_q < \theta < 0] \\ \ge& \tau \mathop{\mathbb{P}}_{\mathcal{P}_0}[\mathcal{C}, \theta \ge \tau] + 0 \cdot \mathop{\mathbb{P}}_{\mathcal{P}_0}[z_t = 1, \mathcal{C}, 0 \le \theta < \tau] - 2s_q \mathop{\mathbb{P}}_{\mathcal{P}_0}[z_t = 1, \mathcal{C}, -2s_q < \theta < 0] \\ \ge& \tau \mathop{\mathbb{P}}_{\mathcal{P}_0}[\mathcal{C}, \theta \ge \tau] - 2s_q \\ \ge& \tau \mathop{\mathbb{P}}_{\mathcal{P}_0}[\mathcal{C}, \theta \ge \tau] - \frac{\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \ge \tau]}{2} + \nu \\ \ge& \tau \mathop{\mathbb{P}}_{\mathcal{P}_0}[\mathcal{C}|\theta \ge \tau] \mathop{\mathbb{P}}_{\mathcal{P}_0}[\theta \ge \tau] - \frac{\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \ge \tau]}{2} + \nu \\ \ge& \tau(1 - (\delta + \delta')) \mathop{\mathbb{P}}_{\mathcal{P}_0}[\theta \ge \tau] - \frac{\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \ge \tau]}{2} + \nu \\ =& \left(\frac{1}{2} - (\delta + \delta')\right) \tau \mathop{\mathbb{P}}_{\mathcal{P}_0}[\theta \ge \tau] + \nu \\ \ge& \left(\frac{1}{2} - \frac{\frac{1}{2}\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \ge \tau]}{\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \ge \tau] + 1}\right) \tau \mathop{\mathbb{P}}_{\mathcal{P}_0}[\theta \ge \tau] + \nu \qquad \text{(by Equation 38)} \\ =& \frac{1}{2}\left(\frac{\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \ge \tau] + 1 - \tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \ge \tau]}{\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \ge \tau] + 1}\right) \tau \mathop{\mathbb{P}}_{\mathcal{P}_0}[\theta \ge \tau] + \nu \\ =& \frac{1}{2}\left(\frac{\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \ge \tau]}{\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \ge \tau] + 1}\right) + \nu \end{aligned}$$

Hence, we have

$$\begin{aligned} \mathop{\mathbb{E}}_{\mathcal{P}_0}[\theta|z_t = 1] \mathop{\mathbb{P}}_{\mathcal{P}_0}[z_t = 1] &\ge \mathbb{E}[\theta|z_t = 1, \mathcal{C}] \mathop{\mathbb{P}}_{\mathcal{P}_0}[z_t = 1, \mathcal{C}] - \frac{\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \ge \tau]}{2\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \ge \tau] + 2} \\ &\ge \frac{\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \ge \tau]}{2\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \ge \tau] + 2} + \nu - \frac{\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \ge \tau]}{2\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta \ge \tau] + 2} \\ &= \nu \ge \phi(u_t) \end{aligned}$$

Therefore, Algorithm 2 fulfills equation 35 and is BIC for all agents of type 0. □

A.3.2 Racing Stage First Part Estimation Bound

Theorem 4.2 (First Part Racing Stage Estimation Bound)**.** *With n total samples collected from the first part of algorithm 2 where the type 0 BIC criterion on the sampling stage approximation bound is met (see lemma 4.1), form an estimate θ_n of the treatment effect θ. With probability at least $1 - \delta$,*

$$\left|\hat{\theta}_n - \theta\right| \le \frac{8\left(\sigma_\varepsilon + \Upsilon\right)\sqrt{2n\log(6/\delta)}}{np_0 - \sqrt{n\log(6/\delta)}}$$

Proof. Recall that the mean recommendation $\bar{z} = 1/2$ in the first part of the racing stage and the proportion of agents who are type 0 be p_0. By lemma 3.1, algorithm 1 is BIC for agents of type 0. Therefore, we may apply corollary A.2 where we let the proportion $p_{BIC} = p_0$. □

A.3.3 Racing Stage Type 1 BIC Proof

Lemma 4.3 (Control-Treatment Racing Stage BIC for Type 1)**.** *Fix some constant $\tau \in (0, 1)$. Remember that $\phi(u_t) \in [-\nu, \nu]$ for all u_t and $\mathbb{P}_{\mathcal{P}_1}[\theta < -\tau]$ be the prior probability that $\theta < -\tau$ for agents of type 1. Let the approximation bound s_{L_1} after the first racing stage be*

$$s_{L_1} := \frac{2\left(\sigma_\varepsilon + 2\Upsilon\right)\sqrt{2L_1 \log(4/\delta)}}{\left|\sum_{i=1}^{L_1}(x_i - \bar{x})(z_i - \bar{z})\right|} \leq \frac{\tau\,\mathbb{P}_{\mathcal{P}_1}[\theta < -\tau]}{4} - \frac{\nu}{2}. \tag{14}$$

Then, during the second phase of the racing stage, the recommendations from algorithm 2 are BIC for agents of type 1.

Proof. Let L_1 be the biggest phase length of the first racing stage. Let the decision threshold s_q for a phase q be defined as in algorithm 2. Assume that after the elimination, at every iteration a sample of the eliminated arm is also drawn, but not revealed to the agent.
We define the event $\mathcal{C}_1$ as the accuracy guarantee of $\hat{\theta}$ such that:

$$\mathcal{C}_1 := \left\{\forall q \geq L_1 : |\theta - \hat{\theta}_q| < s_q\right\} \tag{40}$$

where L is the number of samples after running the sampling stage.
Let X_q^a be the number of arm $a \in \{0, 1\}$ samples in phase q. The decision threshold s_q in phase q is unbounded if $X_q^a = 0$ for either arm a. We also define event $\mathcal{C}_2$ where there is at least one sample of each arm in each phase:

$$\mathcal{C}_2 := \left\{\forall q \geq L, a : X_q^a \geq 1\right\} \tag{41}$$

Recall that p_0 of the population is of type 0. Then, if algorithm 2 is BIC for all agents of type 0 and we recommend each arm at least h times in phase q, we have $X_q^0 > h$ and $X_q^0 \sim B(h, p_0)$. Hence, we have

$$\mathbb{P}[X_q^0 \geq 1] = 1 - p_0^h = \delta'. \tag{42}$$

Let $\tau \in (0, 1)$. Fix phase $q \geq L_1$, and some agent t in this phase. In order to prove that algorithm 2 is BIC for all agents of type 1, we want to show that

$$\mathop{\mathbb{E}}_{\mathcal{P}_1}[\theta | z_t = 0] \mathop{\mathbb{P}}_{\mathcal{P}_1}[z_t = 0] < \phi(u_t). \tag{43}$$

Note that, by assumption, the parameter

$$\begin{aligned}
h &\geq \frac{\log\left(\frac{3\tau\,\mathbb{P}_{\mathcal{P}_1}[\theta<-\tau]+4}{4\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta<-\tau]+4}\right)}{\log(1-p_2)} \\
\Rightarrow \left(\frac{1}{2}\right)^h &\geq \frac{3\tau\,\mathbb{P}_{\mathcal{P}_1}[\theta < -\tau] + 4}{4\tau\,\mathbb{P}_{\mathcal{P}_1}[\theta < -\tau] + 4} \\
\Rightarrow \delta' = 1 - \left(\frac{1}{2}\right)^h &\leq \frac{\tau\,\mathbb{P}_{\mathcal{P}_1}[\theta < -\tau]}{4\tau\,\mathbb{P}_{\mathcal{P}_1}[\theta < -\tau] + 4}.
\end{aligned}$$

From theorem 2.1, with probability $\delta > 0$ we have that

$$\mathbb{P}[\neg\mathcal{C}_1 | G] \leq \delta$$

Define event $\mathcal{C}$ such that both events $\mathcal{C}_1$ and $\mathcal{C}_2$ hold simultaneously in a round:

$$\mathcal{C} := \{\forall q \geq L_1 : \mathcal{C}_1 \;\&\; \mathcal{C}_2\} \tag{44}$$

Using union bound, we have

$$
\begin{aligned}
\mathbb{P}[\neg \mathcal{C}|G] &\leq \mathbb{P}[\neg \mathcal{C}_1|G] + \mathbb{P}[\neg \mathcal{C}_2|G] \quad (45)\\
&\leq \delta + \delta' \\
&\leq \delta_\tau + \delta' \\
&\leq \frac{\tau\, \mathbb{P}_{\mathcal{P}_1}[\theta < -\tau]}{4\tau\, \mathbb{P}_{\mathcal{P}_1}[\theta < -\tau] + 4} + \frac{\tau\, \mathbb{P}_{\mathcal{P}_1}[\theta < -\tau]}{4\tau\, \mathbb{P}_{\mathcal{P}_1}[\theta < -\tau] + 4} \\
&= \frac{\tau\, \mathbb{P}_{\mathcal{P}_1}[\theta < -\tau]}{2\tau\, \mathbb{P}_{\mathcal{P}_1}[\theta < -\tau] + 2} \quad (46)
\end{aligned}
$$

Therefore, since $\theta \leq 1$, we have:

$$
\begin{aligned}
\mathop{\mathbb{E}}_{\mathcal{P}_1}[\theta|z_t = 0] \mathop{\mathbb{P}}_{\mathcal{P}_1}[z_t = 0] &= \mathop{\mathbb{E}}_{\mathcal{P}_1}[\theta|z_t = 0, \mathcal{C}] \mathop{\mathbb{P}}_{\mathcal{P}_1}[z_t = 0, \mathcal{C}] + \mathop{\mathbb{E}}_{\mathcal{P}_1}[\theta|z_t = 0, \neg C] \mathop{\mathbb{P}}_{\mathcal{P}_1}[z_t = 0, \neg C] \\
&\leq \mathop{\mathbb{E}}_{\mathcal{P}_1}[\theta|z_t = 0, \mathcal{C}] \mathop{\mathbb{P}}_{\mathcal{P}_1}[z_t = 0, \mathcal{C}] + \frac{\tau\, \mathbb{P}_{\mathcal{P}_1}[\theta < -\tau]}{2\tau\, \mathbb{P}_{\mathcal{P}_1}[\theta < -\tau] + 2}
\end{aligned}
$$

We want to upper bound the first term. This can be done by splitting it into four cases based on the value of θ. We have:

$$
\begin{aligned}
\mathop{\mathbb{E}}_{\mathcal{P}_1}[\theta|z_t = 0, \mathcal{C}] \mathop{\mathbb{P}}_{\mathcal{P}_1}[z_t = 0, \mathcal{C}] &= \mathop{\mathbb{E}}_{\mathcal{P}_1}[\theta|z_t = 0, \mathcal{C}, \theta \geq 2s_q] \mathop{\mathbb{P}}_{\mathcal{P}_1}[z_t = 0, \mathcal{C}, \theta \geq 2s_q] \\
&+ \mathop{\mathbb{E}}_{\mathcal{P}_1}[\theta|z_t = 0, \mathcal{C}, 0 \leq \theta < 2s_q] \mathop{\mathbb{P}}_{\mathcal{P}_1}[z_t = 0, \mathcal{C}, 0 \leq \theta < 2s_q] \\
&+ \mathop{\mathbb{E}}_{\mathcal{P}_1}[\theta|z_t = 0, \mathcal{C}, -\tau \leq \theta < 0] \mathop{\mathbb{P}}_{\mathcal{P}_1}[z_t = 0, \mathcal{C}, -\tau \leq \theta < 0] \quad (47) \\
&+ \mathop{\mathbb{E}}_{\mathcal{P}_1}[\theta|z_t = 0, \mathcal{C}, \theta \leq -\tau] \mathop{\mathbb{P}}_{\mathcal{P}_1}[z_t = 0, \mathcal{C}, \theta < -\tau]
\end{aligned}
$$

By definition of s_q, we have:

$$
\begin{aligned}
2s_q &\leq 2s_{L_1} \\
&\leq \frac{\tau\, \mathbb{P}_{\mathcal{P}_1}[\theta < -\tau]}{2} - \nu \\
&\leq \tau
\end{aligned}
$$

Conditional on $\mathcal{C}$, the empirical estimate $\hat{\theta}_q > \theta - s_q \geq 2s_q$. Hence, when we have $\theta \geq 2s_q$, then $\hat{\theta} \geq s_q$ and arm 0 must have already been eliminated at phase $q \geq L_1$. This implies that the probability $\mathbb{P}_{\mathcal{P}_1}[z_t = 0, \mathcal{C}, \theta \geq 2s_q] = 0$. Similarly, when we have $\theta \leq -\tau \leq -2s_q$, then $\hat{\theta} \leq -s_q$ and arm 1 must have already been eliminated at phase $q \geq L_1$. Hence, arm 1 could not be recommended in that case and the probability $\mathbb{P}_{\mathcal{P}_1}[z_t = 0, \mathcal{C}, \theta < -\tau] = \mathbb{P}_{\mathcal{P}_1}[\mathcal{C}, \tau < -\tau]$.

We can then rewrite equation (47) as

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{\mathcal{P}_1}[\theta|z_t=0,\mathcal{C}]\mathop{\mathbb{P}}_{\mathcal{P}_1}[z_t=0,\mathcal{C}]\\
=&\mathop{\mathbb{E}}_{\mathcal{P}_1}[\theta|z_t=0,\mathcal{C},0\le\theta<2s_q]\mathop{\mathbb{P}}_{\mathcal{P}_1}[z_t=0,\mathcal{C},0\le\theta<2s_q]\\
&+\mathop{\mathbb{E}}_{\mathcal{P}_1}[\theta|z_t=0,\mathcal{C},-\tau\le\theta<0]\mathop{\mathbb{P}}_{\mathcal{P}_1}[z_t=0,\mathcal{C},-\tau\le\theta<0]\\
&+\mathop{\mathbb{E}}_{\mathcal{P}_1}[\theta|z_t=0,\mathcal{C},\theta<-\tau]\mathop{\mathbb{P}}_{\mathcal{P}_1}[z_t=0,\mathcal{C},\theta<-\tau]\\
\le&-\tau\mathop{\mathbb{P}}_{\mathcal{P}_1}[\mathcal{C},\theta<-\tau]+0\cdot\mathop{\mathbb{P}}_{\mathcal{P}_0}[z_t=1,\mathcal{C},-\tau\le\theta<0]+2s_q\mathop{\mathbb{P}}_{\mathcal{P}_1}[z_t=0,\mathcal{C},0\le\theta<2s_q]\\
\le&2s_q-\tau\mathop{\mathbb{P}}_{\mathcal{P}_1}[\mathcal{C},\theta\le-\tau]\\
\le&\frac{\tau\,\mathbb{P}_{\mathcal{P}_1}[\theta<-\tau]}{2}-\nu-\tau\mathop{\mathbb{P}}_{\mathcal{P}_1}[\mathcal{C},\theta<-\tau]\\
\le&\frac{\tau\,\mathbb{P}_{\mathcal{P}_1}[\theta<-\tau]}{2}-\nu-\tau\mathop{\mathbb{P}}_{\mathcal{P}_1}[\mathcal{C}|\theta<-\tau]\mathop{\mathbb{P}}_{\mathcal{P}_1}[\theta<-\tau]\\
\le&\frac{\tau\,\mathbb{P}_{\mathcal{P}_1}[\theta<-\tau]}{2}-\nu-\tau(1-\delta-\delta')\mathop{\mathbb{P}}_{\mathcal{P}_1}[\theta<-\tau]\\
=&\left(\frac{-1}{2}+\delta+\delta'\right)\tau\mathop{\mathbb{P}}_{\mathcal{P}_1}[\theta<-\tau]-\nu\\
\le&\left(\frac{-1}{2}+\frac{\tau\,\mathbb{P}_{\mathcal{P}_1}[\theta<-\tau]}{2\tau\,\mathbb{P}_{\mathcal{P}_1}[\theta<-\tau]+2}\right)\tau\mathop{\mathbb{P}}_{\mathcal{P}_1}[\theta<-\tau]-\nu && \text{(by Equation 46)}\\
=&\frac{-1}{2}\left(\frac{\tau\,\mathbb{P}_{\mathcal{P}_1}[\theta<-\tau]+1-\tau\,\mathbb{P}_{\mathcal{P}_1}[\theta<-\tau]}{\tau\,\mathbb{P}_{\mathcal{P}_1}[\theta<-\tau]+1}\right)\tau\mathop{\mathbb{P}}_{\mathcal{P}_1}[\theta<-\tau]-\nu\\
=&\frac{-1}{2}\left(\frac{\tau\,\mathbb{P}_{\mathcal{P}_1}[\theta<-\tau]}{\tau\,\mathbb{P}_{\mathcal{P}_1}[\theta<-\tau]+1}\right)+\nu
\end{aligned}
$$

Hence, we have

$$
\begin{aligned}
\mathop{\mathbb{E}}_{\mathcal{P}_1}[\theta|z_t=0]\mathop{\mathbb{P}}_{\mathcal{P}_1}[z_t=0]&\le\mathbb{E}[\theta|z_t=0,\mathcal{C}]\mathop{\mathbb{P}}_{\mathcal{P}_1}[z_t=0,\mathcal{C}]+\frac{\tau\,\mathbb{P}_{\mathcal{P}_1}[\theta<-\tau]}{2\tau\,\mathbb{P}_{\mathcal{P}_1}[\theta<-\tau]+2}\\
&\le\frac{-\tau\,\mathbb{P}_{\mathcal{P}_1}[\theta<-\tau]}{2\tau\,\mathbb{P}_{\mathcal{P}_1}[\theta<-\tau]+2}-\nu+\frac{\tau\,\mathbb{P}_{\mathcal{P}_1}[\theta<-\tau]}{2\tau\,\mathbb{P}_{\mathcal{P}_1}[\theta<-\tau]+2}\\
&=-\nu\le\phi(u_t)
\end{aligned}
$$

Therefore, Algorithm 2 fulfills equation 43 and is BIC for all agents of type 1. □

A.3.4 Racing Stage Second Part Estimation Bound

Theorem 4.4 (Second Part Racing Stage Estimation Bound)**.** *With n total samples collected from the second part of algorithm 2 where the BIC criterion for both types on the sampling stage approximation bound is met (see lemmas 4.1 and 4.3), form an estimate θ_n of the treatment effect θ. With probability at least $1-\delta$,*

$$\left|\hat{\theta}_n-\theta\right|\le\frac{8\left(\sigma_\varepsilon+\Upsilon\right)\sqrt{2n\log(6/\delta)}}{n}$$

Proof. Recall that the mean recommendation $\bar{z}=1/2$ in the racing stage. By lemmas 4.1 and 4.3, algorithm 2 in the second stage is BIC for all agents. Therefore, we may apply corollary A.2 where we let the proportion $p_{BIC}=1$. □

B Regret Proof

B.1 Ex-post Regret Proof

We recall Lemma 4.6:

Lemma 4.6 (Regret). *Algorithms 1 and 2 with parameters $\ell_1, \rho \in \mathbb{N}$ achieves ex-post regret*

$$R(T) \leq \ell_1 \rho + O(\sqrt{T \log T}) \tag{17}$$

where ℓ_1 is the sampling stage's phase length and $1/\rho$ is the exploration probability in the sampling stage.

Proof. From Theorem (2.1), the *clean event* $\mathcal{C}_1$ happens with probability at least $1 - |\theta|$ for some $|\theta| \in (0, 1)$. Hence, the probability of $\neg\mathcal{C}_1$ is at most $|\theta|$. The probability of not getting an arm 1 sample in each phase, $\neg\mathcal{C}_2$ is $|\theta|'$. Hence, the expected ex-post regret conditional on $\neg\mathcal{C}$ is at most $T(|\theta| + |\theta|)'$.

Assume that $\mathcal{C}$ holds. Observe that since $\mathcal{C}$ holds, we have $|\theta| \leq |\hat{\theta}_q| + s_q$ where s_q is the decision threshold of phase q. Before the stopping criteria is invoked, we also have $|\hat{\theta}_q| \leq s_q$. Hence, we have:

$$\begin{aligned}
|\theta| \leq 2s_q &= \frac{(16\sigma_\varepsilon + 16\Upsilon)\sqrt{2n\log(6T/\delta)}}{np_0 - \sqrt{n\log(6T/\delta)}} \\
\Rightarrow \frac{|\theta|}{16(\sigma_\varepsilon + \Upsilon)\sqrt{2\log(6T/\delta)}} &\leq \frac{1}{\sqrt{n} - \sqrt{\log(6T/\delta)}} \\
\Rightarrow n &\leq \frac{512(\sigma_\varepsilon + \Upsilon)^2 \log(6T/\delta)}{|\theta|^2 p_0^2} + \frac{\log(6T/\delta)}{p_0^2} + \frac{32\sqrt{2}(\sigma_\varepsilon + \Upsilon)\log(6T/\delta)}{|\theta| p_0^2}
\end{aligned}$$

Therefore, the main loop must end by when the phase length is

$$n = \frac{512(\sigma_\varepsilon + \Upsilon)^2 \log(6T/\delta)}{|\theta|^2 p_0^2} + \frac{\log(6T/\delta)}{p_0^2} + \frac{32\sqrt{2}(\sigma_\varepsilon + \Upsilon)\log(6T/\delta)}{|\theta| p_0^2}$$

During the racing stage, each phase we give out $2h$ recommendations for the two arms sequentially, so each arm gets recommended $n/2$ times. If arm 0 is indeed the best arm, then on average, the regret for each phase is $0.75n|\theta|$. If arm 1 is the best arm, then on average, the regret for each phase is $0.25n|\theta|$. Conditional on event $\mathcal{C}$, the arm a^* at the end of the racing stage is the best arm so no more regret is collected after the racing stage is finished. Sine we are doubling the phase length after each phase, the total length of the racing stage is upper bounded by $2n$, where n is the biggest phase length of the racing stage.

Then, if arm 0 is the best arm, the total accumulated regret for the racing stage is

$$\begin{aligned}
R_2(T) \leq & \frac{768(\sigma_\varepsilon + \Upsilon)^2 \log(6T/\delta)}{|\theta| p_0^2} + \frac{1.5\log(6T/\delta)|\theta|}{p_0^2} \\
& + \frac{48\sqrt{2}(\sigma_\varepsilon + \Upsilon)\log(6T/\delta)}{p_0^2}
\end{aligned} \tag{48}$$

On the other hand, if arm 1 is the best arm, the total accumulated regret for the racing stage is

$$\begin{aligned}
R_2(T) \leq & \frac{256(\sigma_\varepsilon + \Upsilon)^2 \log(6T/\delta)}{|\theta| p_0^2} + \frac{0.5\log(6T/\delta)|\theta|}{p_0^2} \\
& + \frac{16\sqrt{2}(\sigma_\varepsilon + \Upsilon)\log(6T/\delta)}{p_0^2}
\end{aligned} \tag{49}$$

Observe that the ex-post regret for each round t of the entire algorithm (both the sampling stage and the racing stage) is at most that of the racing stage plus $|\theta|$ per each round of the sampling stage. Alternatively, we can also upper bound the regret by $|\theta|$ per each round of the algorithm. Therefore, we can derive the ex-post regret for the entire algorithm: If arm 0 is the best arm, then the total regret of both Algorithm 1 and Algorithm 2 is:

$$R(T) \leq \min\left\{ L_1|\theta| + \frac{768(\sigma_\varepsilon + \Upsilon)^2 \log(6T/\delta)}{|\theta| p_0^2} + \frac{1.5\log(6T/\delta)|\theta|}{p_0^2} \right. \tag{50}$$

$$\left. + \frac{48\sqrt{2}(\sigma_\varepsilon + \Upsilon)\log(6T/\delta)}{p_0^2}, T|\theta| \right\} \tag{51}$$

$$\leq \ell_1 + O(\sqrt{T\log(T/\delta)}) \tag{52}$$

If arm 1 is the best arm, then the total regret of both Algorithm 1 and Algorithm 2 is:

$$R(T) \leq \min \left\{ L_1 \rho |\theta| + \frac{256(\sigma_\varepsilon + \Upsilon)^2 \log(6T/\delta)}{|\theta| p_0^2} + \frac{0.5 \log(6T/\delta)|\theta|}{p_0^2} \right. \tag{53}$$

$$\left. + \frac{16\sqrt{2}(\sigma_\varepsilon + \Upsilon)\log(6T/\delta)}{p_0^2}, T|\theta| \right\} \tag{54}$$

$$\leq \ell_1 \rho + O(\sqrt{T \log(T/\delta)}) \tag{55}$$

□

B.2 Expected Regret Proof

We recall Lemma 4.7:

Lemma 4.7 (Expected Regret)**.** *Algorithms 1 and 2 with parameters $\ell_1, \rho \in \mathbb{N}$ achieves expected regret*

$$\mathbb{E}[R(T)] = O(\sqrt{T \log T}) \tag{18}$$

Proof. From Theorem 2.1, we have the probability that the *clean event* $\mathcal{C}_1$ happens is at least $1 - |\theta|$ for some $|\theta| \in (0, 1)$. Hence, the probability of $\neg\mathcal{C}_1$ is at most $|\theta|$. The probability of not getting an arm 1 sample in each phase, $\neg\mathcal{C}_2$ is $|\theta|'$. Hence, The expected ex-post regret conditional on $\neg\mathcal{C}$ is at most $T(|\theta| + |\theta|)'$.
Now, we can set parameters $|\theta|, \ell_1$, and ρ in terms of the time horizon T, in order to obtain an expected regret bound strictly relative to T. First, we set $|\theta| = 1/T^2$. Thus, for arbitrary constants τ and $\mathbb{P}[G \geq \tau]$, we may set T sufficiently large such that

$$|\theta| + |\theta|' \leq |\theta|_\tau + |\theta|' = \frac{\tau\, \mathbb{P}[G \geq \tau]}{2\,(\tau\, \mathbb{P}[G \geq \tau] + 1)}.$$

Next, we can set $\ell_1 = \sqrt{T}$. Thus, for arbitrary constant ν and τ, we can always find T sufficiently large such that the estimation bound at the end of the sampling stage is smaller than

$$|\theta - \hat{\theta}| \leq \frac{\tau\, \mathbb{P}_{\mathcal{P}_0}[\theta > \tau]}{4} - \frac{\nu}{2}$$

Lastly, we set $\rho = \sqrt{\log(T)}$. Thus, for arbitrary constants μ_0^0, μ_0^1, Υ, σ_ε and $\mathbb{P}[\xi_i]$, we may set T sufficiently large such that

$$\rho \geq 1 + \frac{4(\mu_2^1 - \mu_2^2 + \nu\, \mathbb{P}_{\mathcal{P}_0}[\xi_0])}{\mathbb{P}[\xi_0]}$$

Therefore, the expected regret of our entire algorithm is:

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathcal{P}_2}[R(T)] &= \mathbb{E}[R(T)|\neg\mathcal{C}]\, \mathbb{P}[\neg\mathcal{C}] + \mathbb{E}[R(T)|\mathcal{C}]\, \mathbb{P}[\mathcal{C}] \\
&\leq T(|\theta| + |\theta|') + \Big((L_1 + L_1\rho) + O(\sqrt{T \log(T)}\Big) \\
&= \frac{1}{T} + \left(\sqrt{T} + \sqrt{\log(T)}\sqrt{T}\right) + O\left(\sqrt{T \log(T)}\right) \\
&= \frac{1}{T} + O\left(\sqrt{T \log(T)}\right) + O\left(\sqrt{T \log(T)}\right) \\
&= O\left(\sqrt{T \log(T)}\right)
\end{aligned}$$

□

Table 2: Type-Specific Regret for Two Arms & Two Types with partially BIC Algorithm

Best arm / Type	Type-specific regret
Arm 0 / Type 0	$O\left(\frac{\log T}{(\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta\geq\tau]-\nu)^2}\right)+O(\sqrt{T\log T})$
Arm 0 / Type 1	$O\left(\frac{\log T}{(\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta\geq\tau]-\nu)^2}\right)+O\left(\frac{\log T}{(\tau\,\mathbb{P}_{\mathcal{P}_1}[\theta<-\tau]-\nu)^2}\right)+O(\sqrt{T\log T})$
Arm 1 / Type 0	$O\left(\frac{\log T}{(\tau\,\mathbb{P}_{\mathcal{P}_0}[\theta\geq\tau]-\nu)^2}\right)+O(\sqrt{T\log T})$
Arm 1 / Type 1	$O(\sqrt{T\log T})$

B.3 Type-specific Regret Proof

We recall Lemma 4.8:

Lemma 4.8 (Type-specific regret)**.** *Algorithms 1 and 2 with parameters $\ell_1, \rho \in \mathbb{N}$ achieves type-specific regrets in table 1.*

Proof. The regret of our algorithm, as given by Lemma (4.6) is

$$R(T) \leq \ell_1\rho + O(\sqrt{T\log T}) \tag{56}$$

where ℓ_1 is the number of arm 1 samples collected in the sampling stage, ρ is the number of phases in the sampling stage.
Let L_1 be the number of samples needed to make agents of type 1 follow our recommendation for arm 0. Half of the population is agents of type 0 and the other half is agents of type 1. We can derive type-specific regret for each type of agents using our algorithm conditioned on which arm is the best arm overall.

1. If arm 0 is the best arm overall:
   Type 0 regret: Since our algorithms guarantee that agents of type 0 will always follow our recommendations, the regret for agents of type 0 is derived similar to the ex-post regret in Lemma 4.6.
   $$R_0 \leq 0.5\ell_1\rho + O(\sqrt{T\log T}) \tag{57}$$
   Type 1 regret: Since agents of type 1 only follow our recommendation and take arm 0 in the second part of the racing stage, the regret of type 1 is then the regret accumulated by all agents of type 1 in the sampling stage, the first part of the racing stage and the second part of the racing stage.
   $$\begin{aligned} R_1 &\leq 0.5(\ell_1\rho + L_1 - \ell_1)|\theta| + O(\sqrt{T\log T}) &(58)\\ &\leq 0.5\ell_1\rho + L_1 + O(\sqrt{T\log T}) &(59)\end{aligned}$$
2. If arm 1 is the best arm overall:
   Type 0 regret: Similar to the case where arm 0 is the best arm overall, the regret for agents of type 0 can be derived from the ex-post regret in Lemma 4.6
   $$R_0 \leq 0.5\ell_1|\theta| + O(\sqrt{T\log T}) \tag{60}$$
   Type 1 regret: Since agents of type 1 only follow our recommendation and take arm 0 in the second part of the racing stage, the regret for agents of type 1 is then the regret accumulated in the second part of the racing stage.
   $$R_1 \leq O(\sqrt{T\log T}) \tag{61}$$

□

C Theorems and Lemmas

Theorem C.1. *(Gaussian Tail Bound) Let $X_1, \ldots, X_n$ be independently and identically distributed Gaussian variables such that $X_i \sim \mathcal{N}(\mu, \sigma^2)$ for $i \in [1, n]$, mean μ, and variance σ. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}(\mu, \sigma^2/n)$. Then,*

$$\mathbb{P}\left[\left|\bar{X} - \mu\right| \geq \varepsilon\right] \leq 2\exp\left\{-\frac{n\varepsilon^2}{2\sigma^2}\right\}. \tag{62}$$

Corollary C.2. *(High-probability bound on the sum of n i.i.d. random Gaussian variables) Let $\sum_{i=1}^{n} X_i$ be the sum of Gaussian variables such that $X_i \sim \mathcal{N}(\mu, \sigma)$. Note that $\sum_{i=1}^{n} X_i = n\bar{X}$. Then, with some confidence level $\delta > 0$, with probability at least $1 - \delta$,*

$$\left|\sum_{i=1}^{n} X_i - \mu\right| \leq \sigma\sqrt{2n\log(2/\delta)}. \tag{63}$$

Corollary C.3. *(High-probability bound on the sum of n i.i.d. random Chi-squared variables) Let $Y := \sum_{i=1}^{n} X_i^2$ be a Chi-squared variable with n degrees of freedom. Then, with some confidence level $\delta > 0$, with probability at least $1 - \delta$,*

$$\sqrt{Y} < \sigma^2\left(\sqrt{n} + 2\sqrt{\log(1/\delta)}\right) \tag{64}$$

Theorem C.4. *(Chernoff Bound for unbounded sub-Gaussian random variables) Let $X_1, \ldots, X_n$ be independent sub-Gaussian random variables with parameter σ. Let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. For all $\varepsilon > 0$,*

$$\mathbb{P}\left[\left|\overline{X}\right| \geq \varepsilon\right] \leq \exp\left\{\frac{-n\varepsilon^2}{2\sigma^2}\right\}.$$

Corollary C.5. *(High probability bound on the sum of unbounded sub-Gaussian random variables) For any $\delta \in (0, 1)$, with probability at least $1 - \delta$,*

$$\left|\overline{X}\right| < \sigma\sqrt{\frac{2\log(1/\delta)}{n}}$$

Theorem C.6. *(Chernoff/Hoeffding's inequality) Let $X_1, \ldots, X_n$ be independent and bounded random variables such that $a \leq X_i \leq b$ for all i. Then*

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X_i] \geq \varepsilon\right] \leq \exp\left(\frac{-2n\varepsilon^2}{(b-a)^2}\right)$$

Corollary C.7. *(High probability upper bound on the sum of bounded random variables) For any $\delta \in (0, 1)$, with probability at least $1 - \delta$,*

$$\mathbb{E}[X] - \frac{1}{n}\sum_{i=1}^{n} X_i \leq (b-a)\sqrt{\frac{\log(1/\delta)}{2n}},$$

where $X_i \in [a, b]$ for all i from 1 to n.

Theorem C.8. *(Multiplicative Chernoff inequality) Let $X_1, \ldots, X_n$ be independent and identically distributed random variables such that $X_i \in [0, 1]$. Then,*

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X_i] > \tau\,\mathbb{E}[X_i]\right] \leq \exp(-\,\mathbb{E}[X_i]n\tau^2/3) \tag{65}$$

Corollary C.9. *(High probability upper bound on the sum of bounded random variables) For any $\delta \in (0, 1)$, with probability at least $1 - \delta$,*

$$\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X_i]\right| \leq \sqrt{\frac{3\ln(1/\delta)}{\mathbb{E}[X_i]n}}\,\mathbb{E}[X_i] \tag{66}$$

Lemma C.10. *(Cauchy-Schwarz Inequality) For any n-dimensional vectors $u, v \in \mathbb{R}^n$, the L^2-norm of the inner product of u and v is less than or equal to the L^2-norm of u times the L^2-norm of v, i.e.*

$$\|\langle u, v\rangle\|_2 \leq \|u\|_2 \cdot \|v\|_2 \,.$$

Theorem C.11. *(Union bound): For a countable set of events $A_1, A_2, \ldots$, we have*

$$\mathbb{P}\left[\bigcup_i A_i\right] \leq \sum_i \mathbb{P}(A_i) \tag{67}$$