Appendix A: Useful Previous Results

We first state the well-known Hoeffdeing's inequality, which establishes concentration bound for i.i.d. random variables.

LEMMA 7 **(Hoeffding's Inequality)**. *Let $X_1, \ldots, X_m$ be independent random variables such that $a_i \leq X_i \leq b_i$ almost surely for each $i \in [m]$. Then, denote by $S_n = \sum_{i=1}^m X_i$. It holds that*

$$P(S_n - \mathbb{E}[S_n] \geq t) \leq \exp(-\frac{2t^2}{\sum_{i=1}^m (b_i - a_i)^2})$$

We then state a general concentration bound for Markov chain with stationary distributions from Healy (2008).

LEMMA 8 **(Theorem 1.1 in Healy (2008))**. *Let $\boldsymbol{X} = (X_i, i \geq 1)$ be a Markov chain with a stationary distribution ϕ. Suppose that the distribution of X_1 is identical to the distribution of ϕ. Then, there exists a constant $\lambda > 0$ such that for any $\epsilon > 0$, it holds that*

$$P\left(\left|\sum_{i=1}^m X_i - \mathbb{E}\left[\sum_{i=1}^m X_i\right]\right| \geq \sqrt{m} \cdot \epsilon\right) \leq 2\exp\left(-\frac{\epsilon^2(1-\lambda)}{4}\right) \tag{21}$$

We also state the following lemma regarding the convexity of the pseudo-cost.

LEMMA 9 **(Proposition 1 in Bu et al. (2020))**. *We denote by $\hat{s}(\mu, Z) = s(q(\mu), z)$ where $q(\mu) = \min_q\{q : \mathbb{E}[s(q, Z)] \geq \mu\}$. We also denote the transformed cost function $TC(\mu) = \hat{C}_\infty^{\pi_{q(\mu)}}$. Suppose that the random supply function takes one of the four formulations specified in Section 3.2. Then, $TC(\mu)$ is a convex function over $[0, \bar{\mu}]$, where $\bar{\mu}$ satisfying $q(\bar{\mu}) = \bar{q}$.*

We finally state the following result, showing how the limiting inventory level can be bounded.

LEMMA 10 **(Lundberg's Inequality)**. *Denote by I_∞ as the limiting distribution of the stochastic process $I_{t+1} = (I_t + Q - D)^+$, where Q and D are two positive random variables. Then, there exists a constant ρ such that for any $a > 0$, we have*

$$P(I_\infty \geq a) \leq \exp(-\rho a).$$

Moreover, ρ is the adjustment coefficient of the random variable $Q - D$, which is defined as the solution to $\lambda(z) = 1$, where $\lambda(z) = \mathbb{E}[\exp(z \cdot (Q - D))]$.

Appendix B: Missing Proofs

Proof of Lemma 2. From Lemma 9, we know that the transformed cost function $\mathrm{TC}(\mu) = \hat{C}_\infty^{\pi_{q(\mu)}}$ is a convex function over $\mu \in [0, \bar{\mu}]$, which is a bounded region. Thus, we know that there exists a constant $\beta' > 0$ such that

$$|\mathrm{TC}(\mu_1) - \mathrm{TC}(\mu_2)| \leq \beta' \cdot |\mu_1 - \mu_2|, \ \forall \mu_1, \mu_2 \in [0, \bar{\mu}]. \tag{22}$$

For any $q_1, q_2 \in [0, \bar{q}]$, we now denote by $\mu_1 = \mathbb{E}[s(q_1, Z)]$ and $\mu_2 = \mathbb{E}[s(q_2, Z)]$. Moreover, for the random supply function taking one of the four formulations specified in Section 3.2, it is direct to check that there exists a constant $\alpha' > 0$ such that

$$|\mu_1 - \mu_2| \leq \alpha' \cdot |q_1 - q_2| \tag{23}$$

Plugging (23) into (22), we know that

$$|\hat{C}_\infty^{\pi_{q_1}} - \hat{C}_\infty^{\pi_{q_2}}| = |\mathrm{TC}(\mu_1) - \mathrm{TC}(\mu_2)| \leq \beta' \cdot |\mu_1 - \mu_2| \leq \alpha'\beta' \cdot |q_1 - q_2|$$

Therefore, we prove that $\hat{C}_\infty^{\pi_q}$ is Lipschitz continuous over q with a Lipschitz constant $\beta = \alpha'\beta'$. □

Proof of Lemma 3. For each epoch n, we denote by

$$\mathcal{B}_n = \{I_{\tau_{n'}} \le \kappa_1 \cdot \log T,\ \tilde{I}^{a^{n*}}_{\tau_{n'}} \le \kappa_1 \cdot \log T \text{ and } I_{\tau_{n'}+\kappa_2 \cdot \max\{\log T, 2L\}} = \tilde{I}^{a^{n'*}}_{\tau_{n'}+\kappa_2 \cdot \max\{\log T, 2L\}},\ \forall n' \le n\}.$$

In order to prove the lemma, it is sufficient to prove that

$$P(\mathcal{B}_n) \ge 1 - \frac{3n}{T^2}. \tag{24}$$

We prove (24) by using induction on the epoch n.

Clearly, when $n = 1$, we have that $P(I_{\tau_1} = 0) = 1$. From Lemma 10, there exists a constant $\kappa_1 > 0$ such that

$$P(\tilde{I}^{a^{1*}}_{\tau_1} \le \kappa_1 \cdot \log T) \ge 1 - \frac{1}{T^2}$$

by noting that the distribution of $\tilde{I}^{a^{1*}}_{\tau_1}$ is identical to the distribution of $I^{a^{1*}}_{\infty}$.

Now conditioning on the event $\{\tilde{I}^{a^{1*}}_{\tau_1} \le \kappa_1 \cdot \log T\}$, we proceed to bound the probability that event $\{I_{\tau_1+\kappa_2 \cdot \max\{\log T, 2L\}} = \tilde{I}^{a^{1*}}_{\tau_1+\kappa_2 \cdot \max\{\log T, 2L\}}\}$ happens. Note that the evolution of the stochastic process I_t in (10) is identical to the evolution of the stochastic process $\tilde{I}^{a^{1*}}_t$ in (13) for $t = \tau_1, \ldots, \tau_2 - 1$. Therefore, it is clear to see that the event $\{I_{\tau_1+\kappa_2 \cdot \max\{\log T, 2L\}} = \tilde{I}^{a^{1*}}_{\tau_1+\kappa_2 \cdot \max\{\log T, 2L\}}\}$ happens as long as

$$I_t = \tilde{I}^{a^{1*}}_t = 0, \quad \text{for some } t = \tau_1, \ldots, \tau_1 + \kappa_2 \cdot \max\{\log T, 2L\}. \tag{25}$$

We note that $I_{\tau_1} \le \tilde{I}^{a^{1*}}_{\tau_1}$ implies that $I_t \le \tilde{I}^{a^{1*}}_t$ for all $t = \tau_1, \ldots, \tau_1 + \kappa_2 \cdot \max\{\log T, 2L\}$. From the non-negativity of I_t and $\tilde{I}^{a^{1*}}_t$, we have that (25) holds as long as $\tilde{I}^{a^{1*}}_t = 0$ for some $t = \tau_1, \ldots, \tau_1 + \kappa_2 \cdot \max\{\log T, 2L\}$. As a result, a sufficient condition for (25) to hold is that

$$\sum_{t=\tau_1}^{\tau_1+\kappa_2 \cdot \max\{\log T, 2L\}} D_t - s(a^{1*}, Z_t) \ge \kappa_1 \cdot \log T$$

Note that $D_t - s(a^{1*}, Z_t)$ are i.i.d. random varibles for $t = \tau_1, \ldots, \tau_1 + \kappa_2 \cdot \max\{\log T, 2L\}$. We also denote by $\delta = \mathbb{E}[D] - \mathbb{E}[s(\bar{q}, Z)]$. Following Hoeffding's inequality (Lemma 7), we have that

$$P\left(\sum_{t=\tau_1}^{\tau_1+\kappa_2 \cdot \max\{\log T, 2L\}} D_t - s(a^{1*}, Z_t) \ge \kappa_1 \cdot \log T\right) \ge 1 - \exp\left(-\frac{2(\delta\kappa_2 \max\{\log T, 2L\} - \kappa_1 \log T)^2}{\kappa_2 \cdot \max\{\log T, 2L\} \cdot \bar{D}}\right) \ge 1 - \frac{1}{T^2}$$

where $\kappa_2 \ge \max\{\frac{2\kappa_1}{\delta}, \frac{4\bar{D}}{\delta^2}\} \ge \max\{\frac{2\kappa_1 \log T}{\delta \cdot \max\{\log T, 2L\}}, \frac{4\bar{D}\log T}{\delta^2 \cdot \max\{\log T, 2L\}}\}$.

The above derivation implies that

$$\begin{aligned}
P(\mathcal{B}_1) &= P\left(\mathcal{B}_1 \mid \tilde{I}^{a^{1*}}_{\tau_1} \le \kappa_1 \cdot \log T\right) \cdot P(\tilde{I}^{a^{1*}}_{\tau_1} \le \kappa_1 \cdot \log T) \\
&= P\left(\sum_{t=\tau_1}^{\tau_1+\kappa_2 \cdot \max\{\log T, 2L\}} D_t - s(a^{1*}, Z_t) \ge \kappa_1 \cdot \log T\right) \cdot P(\tilde{I}^{a^{1*}}_{\tau_1} \le \kappa_1 \cdot \log T) \\
&\ge (1 - \frac{1}{T^2}) \cdot (1 - \frac{1}{T^2}) \ge 1 - \frac{2}{T^2} \ge 1 - \frac{3}{T^2}
\end{aligned}$$

Therefore, we prove (24) for $n = 1$.

Now assume that (24) holds for epoch $n-1$. We consider epoch n. Clearly, from the definition of the stochastic process $\tilde{I}^{a^{n*}}_t$ in (13), $\tilde{I}^{a^{n*}}_t$ refreshes when $t = \tau_n$. As a result, the distribution of $\tilde{I}^{a^{n*}}_{\tau_n}$ is independent of the event $\mathcal{B}_{n-1}$ and is identical to the distribution of $I^{a^{n*}}_{\infty}$, which implies that

$$P(\tilde{I}^{a^{n*}}_{\tau_n} \le \kappa_1 \cdot \log T \mid \mathcal{B}_{n-1}) = P(\tilde{I}^{a^{n*}}_{\tau_n} \le \kappa_1 \cdot \log T) \ge 1 - \frac{1}{T^2} \tag{26}$$

where the second inequality follows from Lemma 10. Moreover, conditioning on $\mathcal{B}_{n-1}$, since I_{τ_n-1} couples with $\tilde{I}^{a^{(n-1)*}}_{\tau_n-1}$, we have that

$$P(I_{\tau_n-1} \le \kappa_1 \cdot \log T \mid \mathcal{B}_{n-1}) = P(\tilde{I}^{a^{(n-1)*}}_{\tau_n-1} \le \kappa_1 \cdot \log T \mid \mathcal{B}_{n-1}) = P(\tilde{I}^{a^{(n-1)*}}_{\tau_n-1} \le \kappa_1 \cdot \log T)/P(\mathcal{B}_{n-1}) \tag{27}$$

Note that the distribution of $\tilde{I}^{a^{(n-1)*}}_{\tau_n-1}$ is identical to the distribution of $I^{a^{(n-1)*}}_{\infty}$, which implies that

$$P(\tilde{I}^{a^{(n-1)*}}_{\tau_n-1} \le \kappa_1 \cdot \log T) = P(I^{a^{(n-1)*}}_{\infty} \le \kappa_1 \cdot \log T) \ge 1 - \frac{1}{T^2}$$

Therefore, by noting that $P(\mathcal{B}_{n-1}) \le 1$, from (27), we have that

$$P(I_{\tau_n-1} \le \kappa_1 \cdot \log T \mid \mathcal{B}_{n-1}) \ge 1 - \frac{1}{T^2} \tag{28}$$

From (26), (28) and the union bound, we have that

$$P(I_{\tau_n-1} \le \kappa_1 \cdot \log T \text{ and } \tilde{I}^{a^{n*}}_{\tau_n} \le \kappa_1 \cdot \log T \mid \mathcal{B}_{n-1}) \ge 1 - \frac{2}{T^2} \tag{29}$$

As a result, conditioning on $\mathcal{B}_{n-1}$, we know that

$$I_{\tau_n+L} \le L \cdot \bar{D} + \kappa_1 \cdot \log T \text{ and } \tilde{I}^{a^{n*}}_{\tau_n+L} \le L \cdot \bar{D} + \kappa_1 \cdot \log T \tag{30}$$

happens with a probability at least $1 - \frac{2}{T^2}$. It is clear to see that the event $\{I_{\tau_n+\max\{\log T, 2L\}} = \tilde{I}^{a^{n*}}_{\tau_n+\max\{\log T, 2L\}}\}$ happens as long as

$$I_t = \tilde{I}^{a^{n*}}_t = 0 \tag{31}$$

for some $t = \tau_n + L, \ldots, \tau_n + \max\{\log T, 2L\}$.

Suppose that $I_{\tau_n} \le \tilde{I}^{a^{n*}}_{\tau_n}$ (resp. $I_{\tau_n} \ge \tilde{I}^{a^{n*}}_{\tau_n}$), from the evolution of the stochastic process in (10) and (13), we have that $I_t \le \tilde{I}^{a^{n*}}_t$ (resp. $I_t \ge \tilde{I}^{a^{n*}}_t$) for any $t = \tau_n + L, \ldots, \tau_n + \max\{\log T, 2L\}$. Given that I_t and $\tilde{I}^{a^{n*}}_t$ must be non-negative (from definition), we conclude that if $I_{\tau_n} \le \tilde{I}^{a^{n*}}_{\tau_n}$ (resp. $I_{\tau_n} \ge \tilde{I}^{a^{n*}}_{\tau_n}$), then (31) happens as long as $\tilde{I}^{a^{n*}}_t = 0$ (resp. $I_t = 0$). Thus, a sufficient condition for (31) to happen is that

$$\sum_{t=\tau_n+L}^{\tau_n+\max\{\log T, 2L\}} D_t - s(a^{n*}, Z_t) \ge L \cdot \bar{D} + \kappa_1 \cdot \log T$$

Since $D_t - s(a^{n*}, Z_t)$ are i.i.d. random variable for $t = \tau_n + L, \ldots, \tau_n + \max\{\log T, 2L\}$, we denote by $\delta_n = \mathbb{E}_{D\sim F}[D] - \mathbb{E}_{Z\sim G}[s(a^{n*}, Z)] \ge \delta$. Following Hoeffding's inequality (Lemma 7), we have that

$$\begin{aligned} P\left(\sum_{t=\tau_n+L}^{\tau_n+\kappa_2\max\{\log T, 2L\}} D_t - s(a^{n*}, Z_t) \ge L \cdot \bar{D} + \kappa_1 \cdot \log T\right) &\ge 1 - \exp\left(-\frac{2(\kappa_2 \max\{\log T, 2L\} - L(\bar{D}+1) - \kappa_1 \log T)^2}{\kappa_2 \max\{\log T, 2L\} - L}\right) \\ &\ge 1 - \frac{1}{T^2} \end{aligned}$$

where $\kappa_2 \ge \max\{4, 2(\bar{D}+1+\kappa_1)\} \ge \max\{\frac{4\log T}{\max\{\log T, 2L\}}, 2(\bar{D}+1+\kappa_1)\}$. Therefore, we have that

$$P\left(I_{\tau_n+\kappa_2\max\{\log T, 2L\}} = \tilde{I}^{a^{n*}}_{\tau_n+\kappa_2\max\{\log T, 2L\}} \mid \mathcal{B}_{n-1} \text{ and (30) happens}\right) \ge 1 - \frac{1}{T^2}.$$

Combining (29) and the induction hypothesis that $P(\mathcal{B}_{n-1}) \ge 1 - \frac{3(n-1)}{T^2}$, we have that

$$\begin{aligned} P(\mathcal{B}_n) =& P(\mathcal{B}_{n-1}) \cdot P(I_{\tau_n-1} \le \kappa_1 \cdot \log T \text{ and } \tilde{I}^{a^{n*}}_{\tau_n} \le \kappa_1 \cdot \log T \mid \mathcal{B}_{n-1}) \\ &\cdot P\left(I_{\tau_n+\kappa_2\max\{\log T, 2L\}} = \tilde{I}^{a^{n*}}_{\tau_n+\kappa_2\max\{\log T, 2L\}} \mid \mathcal{B}_{n-1} \text{ and (30) happens}\right) \\ \ge& (1 - \frac{3(n-1)}{T^2}) \cdot (1 - \frac{2}{T^2}) \cdot (1 - \frac{1}{T^2}) \ge (1 - \frac{3(n-1)}{T^2}) \cdot (1 - \frac{3}{T^2}) \\ \ge& 1 - \frac{3n}{T^2} \end{aligned}$$

which completes our proof of the induction of (24) for each epoch n. Therefore, our proof of the lemma is completed. □

Proof of Lemma 4. Clearly, from (14), it is enough to compare the value of I_t and $\tilde{I}_t^{a^{n*}}$ for each epoch n and each period t in the epoch n. Note that we identify an event $\mathcal{B}$ in Lemma 3 that I_t and $\tilde{I}_t^{a^{n*}}$ couple with each other. We consider two situations where $\mathcal{B}$ happens or $\mathcal{B}$ not happens.

Case 1: We now assume that $\mathcal{B}$ happens. Then, we know that for each epoch $n \in [N]$ and each $t = \tau_n + \kappa_2 \cdot \max\{\log T, 2L\}, \ldots, \tau_{n+1} - 1$, the value of I_t and $\tilde{I}_t^{a^{n*}}$ are identical. Therefore, only when $t = \tau_n, \ldots, \tau_n + \kappa_2 \cdot \max\{\log T, 2L\}$, the value of I_t and $\tilde{I}_t^{a^{n*}}$ can be different. Moreover, note that the evolution of I_t in (10) is the same as the evolution of $\tilde{I}_t^{a^{n*}}$ in (13), except that the intial value I_{τ_n} is different from $\tilde{I}_{\tau_n}^{a^{n*}}$. We know that the gap between I_t and $\tilde{I}_t^{a^{n*}}$ can only become smaller. Therefore, we get that

$$|I_t - \tilde{I}_t^{a^{n*}}| \le |I_{\tau_n} - \tilde{I}_{\tau_n}^{a^{n*}}| \le \kappa_1 \cdot \log T \tag{32}$$

where the last inequality follows from the condition in the event $\mathcal{B}$. We have that

$$\begin{aligned} \left| \mathbb{E}\left[\sum_n \sum_{t=\tau_n}^{\tau_{n+1}-1} (I_t - \tilde{I}_t^{a^{n*}}) \mid \mathcal{B} \right] \right| &\le \sum_n \sum_{t=\tau_n}^{\tau_n + \kappa_2 \cdot \max\{\log T, 2L\}} \mathbb{E}[|I_t - \tilde{I}_t^{a^{n*}}| \mid \mathcal{B}] \\ &\le \sum_n \kappa_1 \cdot \log T \cdot \kappa_2 \cdot \max\{\log T, 2L\} \\ &= N \cdot \kappa_1 \kappa_2 \log T \cdot \max\{\log T, 2L\} \end{aligned} \tag{33}$$

Case 2: We now assume that $\mathcal{B}$ does not happen. Clearly, a direct upper bound on both I_t and $\tilde{I}_t^{a^{n*}}$ is that

$$I_t \le \bar{D} \cdot t \text{ and } \mathbb{E}[\tilde{I}_t^{a^{n*}} \mid \mathcal{B}^c] \le \bar{D} \cdot t$$

Therefore, we have that

$$\left| \mathbb{E}\left[\sum_n \sum_{t=\tau_n}^{\tau_{n+1}-1} (I_t - \tilde{I}_t^{a^{n*}}) \mid \mathcal{B}^c \right] \right| \le \bar{D} \cdot \sum_{t=1}^{T} t \le \bar{D} \cdot T^2 \tag{34}$$

However, from Lemma 3, we know that $P(\mathcal{B}^c) \le \frac{3N}{T^2}$. As a result, combining (33) and (34), we get that

$$\begin{aligned} \left| \mathbb{E}\left[\sum_n \sum_{t=\tau_n}^{\tau_{n+1}-1} (I_t - \tilde{I}_t^{a^{n*}}) \right] \right| &\le \left| \mathbb{E}\left[\sum_n \sum_{t=\tau_n}^{\tau_{n+1}-1} (I_t - \tilde{I}_t^{a^{n*}}) \mid \mathcal{B} \right] \right| \cdot P(\mathcal{B}) + \left| \mathbb{E}\left[\sum_n \sum_{t=\tau_n}^{\tau_{n+1}-1} (I_t - \tilde{I}_t^{a^{n*}}) \mid \mathcal{B}^c \right] \right| \cdot P(\mathcal{B}^c) \\ &\le \left| \mathbb{E}\left[\sum_n \sum_{t=\tau_n}^{\tau_{n+1}-1} (I_t - \tilde{I}_t^{a^{n*}}) \mid \mathcal{B} \right] \right| + \left| \mathbb{E}\left[\sum_n \sum_{t=\tau_n}^{\tau_{n+1}-1} (I_t - \tilde{I}_t^{a^{n*}}) \mid \mathcal{B}^c \right] \right| \cdot \frac{3N}{T^2} \\ &\le N \cdot \kappa_1 \kappa_2 \log T \cdot \max\{\log T, 2L\} + 3N\bar{D} \end{aligned}$$

which completes our proof. □

Proof of Lemma 5. The proof generalizes the proof of Lemma 3. For each epoch n, we denote by

$$\mathcal{C}_n = \{I^a_{\tau_{n'}} \le \kappa_1 \cdot \log T,\ \tilde{I}^a_{\tau_{n'}} \le \kappa_1 \cdot \log T \text{ and } I^a_{\tau_{n'} + \kappa_2 \cdot \max\{\log T, 2L\}} = \tilde{I}^{a^{n'*}}_{\tau_{n'} + \kappa_2 \cdot \max\{\log T, 2L\}},\ \forall a \in \mathcal{A}_{n'},\ \forall n' \le n\}.$$

In order to prove the lemma, it is sufficient to prove that

$$P(\mathcal{C}_n) \ge 1 - \frac{3(K+1)n}{T^2}. \tag{35}$$

We prove (35) by using induction on the epoch n.

Clearly, when $n = 1$, we have that $P(I^a_{\tau_1} = 0) = 1$ for all $a \in \mathcal{A}_1$. From Lemma 10, there exists a constant $\kappa_1 > 0$ such that for each $a \in \mathcal{A}_1$, it holds that

$$P(\tilde{I}^a_{\tau_1} \le \kappa_1 \cdot \log T) \ge 1 - \frac{1}{T^2}$$

by noting that the distribution of $\tilde{I}^a_{\tau_1}$ is identical to the distribution of I^a_∞ for each $a \in \mathcal{A}_1$.

Now conditioning on the event $\{\tilde{I}^a_{\tau_1} \le \kappa_1 \cdot \log T,\ \forall a \in \mathcal{A}_1\}$, we proceed to bound the probability that event $\{I^a_{\tau_1+\kappa_2\cdot\max\{\log T,2L\}} = \tilde{I}^a_{\tau_1+\kappa_2\cdot\max\{\log T,2L\}},\ \forall a \in \mathcal{A}_1\}$ happens. Note that the evolution of the stochastic process I^a_t in (9) is identical to the evolution of the stochastic process $\tilde{I}^a_t$ in (15) for $t = \tau_1, \ldots, \tau_2 - 1$. Therefore, for each $a \in \mathcal{A}_1$, it is clear to see that the event $\{I^a_{\tau_1+\kappa_2\cdot\max\{\log T,2L\}} = \tilde{I}^a_{\tau_1+\kappa_2\cdot\max\{\log T,2L\}}\}$ happens as long as

$$I^a_t = \tilde{I}^a_t = 0, \quad \text{for some } t = \tau_1, \ldots, \tau_1 + \kappa_2 \cdot \max\{\log T, 2L\}. \tag{36}$$

We note that $I^a_{\tau_1} \le \tilde{I}^a_{\tau_1}$ implies that $I^a_t \le \tilde{I}^a_t$ for all $t = \tau_1, \ldots, \tau_1 + \kappa_2 \cdot \max\{\log T, 2L\}$. From the non-negativity of I^a_t and $\tilde{I}^a_t$, we have that (36) holds as long as $\tilde{I}^a_t = 0$ for some $t = \tau_1, \ldots, \tau_1 + \kappa_2 \cdot \max\{\log T, 2L\}$. As a result, a sufficient condition for (36) to hold for a $a \in \mathcal{A}_1$ is that

$$\sum_{t=\tau_1}^{\tau_1+\kappa_2\cdot\max\{\log T,2L\}} D_t - s(a, Z_t) \ge \kappa_1 \cdot \log T$$

Note that $D_t - s(a, Z_t)$ are i.i.d. random varibles for $t = \tau_1, \ldots, \tau_1 + \kappa_2 \cdot \max\{\log T, 2L\}$. We also denote by $\delta = \mathbb{E}[D] - \mathbb{E}[s(\bar{q}, Z)]$. Following Hoeffding's inequality (Lemma 7), we have that

$$P\left(\sum_{t=\tau_1}^{\tau_1+\kappa_2\cdot\max\{\log T,2L\}} D_t - s(a, Z_t) \ge \kappa_1 \cdot \log T\right) \ge 1 - \exp\left(-\frac{2(\delta\kappa_2 \max\{\log T, 2L\} - \kappa_1 \log T)^2}{\kappa_2 \cdot \max\{\log T, 2L\} \cdot \bar{D}}\right) \ge 1 - \frac{1}{T^2}$$

where $\kappa_2 \ge \max\{\frac{2\kappa_1}{\delta}, \frac{4\bar{D}}{\delta^2}\} \ge \max\{\frac{2\kappa_1 \log T}{\delta \cdot \max\{\log T, 2L\}}, \frac{4\bar{D}\log T}{\delta^2 \cdot \max\{\log T, 2L\}}\}$.

The above derivation implies that

$$\begin{aligned}
P(\mathcal{B}_1) &= P\left(\mathcal{B}_1 \mid \tilde{I}^a_{\tau_1} \le \kappa_1 \cdot \log T,\ \forall a \in \mathcal{A}_1\right) \cdot P(\tilde{I}^a_{\tau_1} \le \kappa_1 \cdot \log T,\ \forall a \in \mathcal{A}_1) \\
&= P\left(\sum_{t=\tau_1}^{\tau_1+\kappa_2\cdot\max\{\log T,2L\}} D_t - s(a, Z_t) \ge \kappa_1 \cdot \log T,\ \forall a \in \mathcal{A}_1\right) \cdot P(\tilde{I}^{a^{1*}}_{\tau_1} \le \kappa_1 \cdot \log T,\ \forall a \in \mathcal{A}_1) \\
&\ge (1 - \frac{K+1}{T^2}) \cdot (1 - \frac{K+1}{T^2}) \ge 1 - \frac{2(K+1)}{T^2} \ge 1 - \frac{3(K+1)}{T^2}
\end{aligned}$$

where the first inequality follows from the union bound by noting that $|\mathcal{A}_1| \le K + 1$. Therefore, we prove (35) for $n = 1$.

Now assume that (35) holds for epoch $n-1$. We consider epoch n. Clearly, from the definition of the stochastic process $\tilde{I}^a_t$ in (15), $\tilde{I}^a_t$ refreshes when $t = \tau_n$. As a result, the distribution of $\tilde{I}^a_{\tau_n}$ is independent of the event $\mathcal{C}_{n-1}$ and is identical to the distribution of I^a_∞, which implies that

$$P(\tilde{I}^a_{\tau_n} \le \kappa_1 \cdot \log T \mid \mathcal{C}_{n-1}) = P(\tilde{I}^a_{\tau_n} \le \kappa_1 \cdot \log T) \ge 1 - \frac{1}{T^2}, \quad \forall a \in \mathcal{A}_n \tag{37}$$

where the second inequality follows from Lemma 10. Moreover, conditioning on $\mathcal{B}_{n-1}$, since $I^a_{\tau_n-1}$ couples with $\tilde{I}^a_{\tau_n-1}$ for each $a \in \mathcal{A}_n \subset \mathcal{A}_{n-1}$, we have that

$$P(I^a_{\tau_n-1} \le \kappa_1 \cdot \log T \mid \mathcal{C}_{n-1}) = P(\tilde{I}^a_{\tau_n-1} \le \kappa_1 \cdot \log T \mid \mathcal{C}_{n-1}) = P(\tilde{I}^a_{\tau_n-1} \le \kappa_1 \cdot \log T)/P(\mathcal{C}_{n-1}),\ \forall a \in \mathcal{A}_n \tag{38}$$

Note that the distribution of $\tilde{I}^a_{\tau_n-1}$ is identical to the distribution of I^a_∞, which implies that

$$P(\tilde{I}^a_{\tau_n-1} \le \kappa_1 \cdot \log T) = P(I^a_\infty \le \kappa_1 \cdot \log T) \ge 1 - \frac{1}{T^2},\ \forall a \in \mathcal{A}_n$$

Therefore, by noting that $P(\mathcal{C}_{n-1}) \le 1$, from (38), we have that

$$P(I^a_{\tau_n-1} \le \kappa_1 \cdot \log T \,|\, \mathcal{C}_{n-1}) \ge 1 - \frac{1}{T^2},\ \forall a \in \mathcal{A}_n. \tag{39}$$

From (37), (39) and the union bound, we have that

$$P(I_{\tau_n-1} \le \kappa_1 \cdot \log T \text{ and } \tilde{I}^{a^{n*}}_{\tau_n} \le \kappa_1 \cdot \log T,\ \forall a \in \mathcal{A}_n \,|\, \mathcal{C}_{n-1}) \ge 1 - \frac{2(K+1)}{T^2} \tag{40}$$

where we note that $|\mathcal{A}_n| \le K+1$. As a result, conditioning on $\mathcal{C}_{n-1}$, we know that

$$I^a_{\tau_n+L} \le L \cdot \bar{D} + \kappa_1 \cdot \log T \text{ and } \tilde{I}^a_{\tau_n+L} \le L \cdot \bar{D} + \kappa_1 \cdot \log T,\ \forall a \in \mathcal{A}_n \tag{41}$$

happens with a probability at least $1 - \frac{2(K+1)}{T^2}$. It is clear to see that for each $a \in \mathcal{A}_n$, the event $\{I^a_{\tau_n+\max\{\log T, 2L\}} = \tilde{I}^a_{\tau_n+\max\{\log T, 2L\}}\}$ happens as long as

$$I^a_t = \tilde{I}^a_t = 0 \tag{42}$$

for some $t = \tau_n + L, \ldots, \tau_n + \max\{\log T, 2L\}$.

For each $a \in \mathcal{A}_n$, suppose that $I^a_{\tau_n} \le \tilde{I}^a_{\tau_n}$ (resp. $I^a_{\tau_n} \ge \tilde{I}^a_{\tau_n}$), from the evolution of the stochastic process in (9) and (15), we have that $I^a_t \le \tilde{I}^a_t$ (resp. $I^a_t \ge \tilde{I}^a_t$) for any $t = \tau_n + L, \ldots, \tau_n + \max\{\log T, 2L\}$. Given that I^a_t and $\tilde{I}^a_t$ must be non-negative (from definition), we conclude that if $I^a_{\tau_n} \le \tilde{I}^a_{\tau_n}$ (resp. $I^a_{\tau_n} \ge \tilde{I}^a_{\tau_n}$), then (31) happens as long as $\tilde{I}^a_t = 0$ (resp. $I^a_t = 0$). Thus, a sufficient condition for (42) to happen is that

$$\sum_{t=\tau_n+L}^{\tau_n+\max\{\log T, 2L\}} D_t - s(a, Z_t) \ge L \cdot \bar{D} + \kappa_1 \cdot \log T$$

Since $D_t - s(a, Z_t)$ are i.i.d. random variable for $t = \tau_n + L, \ldots, \tau_n + \max\{\log T, 2L\}$, we denote by $\delta_{n,a} = \mathbb{E}_{D\sim F}[D] - \mathbb{E}_{Z\sim G}[s(a, Z)] \ge \delta$. Following Hoeffding's inequality (Lemma 7), for each $a \in \mathcal{A}_n$, we have that

$$\begin{aligned} P\left(\sum_{t=\tau_n+L}^{\tau_n+\kappa_2\max\{\log T, 2L\}} D_t - s(a, Z_t) \ge L \cdot \bar{D} + \kappa_1 \cdot \log T\right) &\ge 1 - \exp\left(-\frac{2(\kappa_2 \max\{\log T, 2L\} - L(\bar{D}+1) - \kappa_1 \log T)^2}{\kappa_2 \max\{\log T, 2L\} - L}\right) \\ &\ge 1 - \frac{1}{T^2} \end{aligned}$$

where $\kappa_2 \ge \max\{4, 2(\bar{D}+1+\kappa_1)\} \ge \max\{\frac{4\log T}{\max\{\log T, 2L\}}, 2(\bar{D}+1+\kappa_1)\}$. Therefore, from the union bound, we have that

$$P\left(I^a_{\tau_n+\kappa_2\max\{\log T, 2L\}} = \tilde{I}^a_{\tau_n+\kappa_2\max\{\log T, 2L\}},\ \forall a \in \mathcal{A}_n \,|\, \mathcal{C}_{n-1} \text{ and (41) happens}\right) \ge 1 - \frac{K+1}{T^2}.$$

Combining (40) and the induction hypothesis that $P(\mathcal{C}_{n-1}) \ge 1 - \frac{3(K+1)(n-1)}{T^2}$, we have that

$$\begin{aligned} P(\mathcal{C}_n) =& P(\mathcal{C}_{n-1}) \cdot P(I^a_{\tau_n-1} \le \kappa_1 \cdot \log T \text{ and } \tilde{I}^a_{\tau_n} \le \kappa_1 \cdot \log T,\ \forall a \in \mathcal{A}_n \,|\, \mathcal{C}_{n-1}) \\ &\cdot P\left(I^a_{\tau_n+\kappa_2\max\{\log T, 2L\}} = \tilde{I}^a_{\tau_n+\kappa_2\max\{\log T, 2L\}},\ \forall a \in \mathcal{A}_n \,|\, \mathcal{C}_{n-1} \text{ and (41) happens}\right) \\ \ge& (1 - \frac{3(K+1)(n-1)}{T^2}) \cdot (1 - \frac{2(K+1)}{T^2}) \cdot (1 - \frac{K+1}{T^2}) \\ \ge& (1 - \frac{3(K+1)(n-1)}{T^2}) \cdot (1 - \frac{3(K+1)}{T^2}) \\ \ge& 1 - \frac{3(K+1)n}{T^2} \end{aligned}$$

which completes our proof of the induction of (35) for each epoch n. Therefore, our proof of the lemma is completed. □

Proof of Lemma 6. We first show that for each epoch $n \in [N]$ and each action $a \in \mathcal{A}_n$, we can use the average value of $\tilde{I}_t^a$ for $t = \tau_n + \kappa_2 \cdot \max\{\log T, 2L\}$ to $\tau_{n+1} - 1$ to approximate the value of $\mathbb{E}[I_\infty^{\pi_a}]$, where the length of the confidence interval can be given by γ_n.

Clearly, $\{\tilde{I}_t^a\}_{t=\tau_n+\kappa_2\cdot\max\{\log T,2L\}}^{\tau_{n+1}-1}$ forms a Markov chain. We denote by $\boldsymbol{I}$ a vector such that

$$\boldsymbol{I} = (\tilde{I}_t^a, \forall t = \tau_n + \kappa_2 \cdot \max\{\log T, 2L\}, \ldots, \tau_{n+1} - 1)$$

We apply Lemma 8 to derive a concentration bound for $\boldsymbol{I}$. To be specific, for each epoch $n \le N-1$, we regard $\tilde{I}^a_{\tau_n+\kappa_2\cdot\max\{\log T,2L\}+i}$ as X_i for $i = 1, \ldots, \tau_{n+1} - 1 - \tau_n - \kappa_2 \cdot \max\{\log T, 2L\}$. Clearly, $\boldsymbol{I}$ is a Markov chain with stationary distributions and satisfies the conditions in Lemma 8.

We now denote $m = \tau_{n+1} - \tau_n - 1 - \kappa_2 \cdot \max\{\log T, 2L\}$. Then, from Lemma 8, there exists a constant λ such that for any $\epsilon > 0$, it holds

$$P\left(\left|\sum_{i=1}^m \tilde{I}^a_{\tau_n+\kappa_2\cdot\max\{\log T,2L\}+i} - \mathbb{E}\left[\sum_{i=1}^m \tilde{I}^a_{\tau_n+\kappa_2\cdot\max\{\log T,2L\}+i}\right]\right| \ge \sqrt{m}\cdot\epsilon\right) \le 2\exp\left(-\frac{\epsilon^2(1-\lambda)}{4}\right).$$

We now set $\epsilon = \frac{\gamma_n\cdot\sqrt{m}}{2}$. Then, we have

$$\exp\left(-\frac{\epsilon^2(1-\lambda)}{4}\right) \le \exp(-\frac{(1-\lambda)m\gamma_n^2}{16}) \tag{43}$$

We proceed to give a lower bound on $m\gamma_n^2$, which will imply an upper bound for (43). Note that

$$m = \tau_{n+1} - \tau_n - 1 - \kappa_2 \cdot \max\{\log T, 2L\} = \kappa_2 \cdot (\max\{\frac{1}{\gamma_n^2}\cdot\log T, 3L\} - \max\{\log T, 2L\})$$

If $\frac{1}{\gamma_2}\cdot\log T \ge 3L$, then we have

$$\max\{\frac{1}{\gamma_n^2}\cdot\log T, 3L\} - \max\{\log T, 2L\} = \frac{1}{\gamma_n^2}\cdot\log T - \max\{\log T, 2L\} \ge \frac{1}{3\gamma_n^2}\cdot\log T$$

If $\frac{1}{\gamma_2}\cdot\log T < 3L$, then we have

$$\max\{\frac{1}{\gamma_n^2}\cdot\log T, 3L\} - \max\{\log T, 2L\} = L \ge \frac{1}{3\gamma_n^2}\cdot\log T$$

Therefore, it holds that

$$m = \kappa_2 \cdot (\max\{\frac{1}{\gamma_n^2}\cdot\log T, 3L\} - \max\{\log T, 2L\}) \ge \frac{\kappa_2}{3\gamma_n^2}\cdot\log T \tag{44}$$

Plugging (44) into (43), we have

$$\exp\left(-\frac{\epsilon^2(1-\lambda)}{4}\right) \le \exp\left(-\frac{(1-\lambda)\kappa_2\log T}{12}\right) \le \frac{1}{T^2}$$

where $\kappa_2 \ge 24/(1-\lambda)$. We have

$$\begin{aligned}
&P\left(\left|\sum_{i=1}^m \tilde{I}^a_{\tau_n+\kappa_2\cdot\max\{\log T,2L\}+i} - \mathbb{E}\left[\sum_{i=1}^m \tilde{I}^a_{\tau_n+\kappa_2\cdot\max\{\log T,2L\}+i}\right]\right| \ge m\cdot\frac{\gamma_n}{2}\right) \\
=&P\left(\left|\sum_{i=1}^m \tilde{I}^a_{\tau_n+\kappa_2\cdot\max\{\log T,2L\}+i} - m\cdot\mathbb{E}\left[I_\infty^{\pi_a}\right]\right| \ge m\cdot\frac{\gamma_n}{2}\right) \\
\le&\frac{2}{T^2}
\end{aligned} \tag{45}$$

where the second inequality follows from the fact that the distribution of $\tilde{I}_t^a$ is identical to the distribution of $I_\infty^{\pi_a}$. Moreover, from Hoeffding's inequality (Lemma 7), it holds that

$$P\left(\left|\sum_{t=\tau_n+\kappa_2\cdot\max\{\log T,2L\}}^{\tau_{n+1}-1} s(a,Z_t)) - m\cdot\mathbb{E}[s(a,Z)]\right| \geq m\cdot\frac{\gamma_n}{2}\right) \leq 2\exp(-\frac{m\gamma_n^2}{2\bar{D}^2}) \leq 2\exp(-\frac{\kappa_2\log T}{6\bar{D}^2}) \leq \frac{2}{T^2} \tag{46}$$

where the second inequality follows from (44) and the third inequality follows from $\kappa_2 \geq 12\bar{D}^2$. Therefore, conditional on the event $\mathcal{C}$ happens, we have that

$$P\left(\left|\tilde{C}_n^a - \hat{C}_\infty^{\pi_a}\right| \leq (h+b)\cdot\frac{\gamma_n}{2} \mid \mathcal{C}\right) \geq 1-\frac{4}{T^2}$$

which implies that (from union bound over all $a\in\mathcal{A}$ and all $n\leq N-1$)

$$P(\mathcal{E}\mid\mathcal{C}) = P\left(\{|\tilde{C}_n^a - \hat{C}_\infty^{\pi_a}| \leq (h+b)\cdot\frac{\gamma_n}{2},\ \forall a\in\mathcal{A}_n, \forall 1\leq n\leq N-1\}\right) \geq 1-\frac{4(K+1)N}{T^2}$$

From Lemma 5, we know that $P(\mathcal{C})\geq 1-\frac{3(K+1)N}{T^2}$. Therefore, we have that

$$P(\mathcal{E}) = P(\mathcal{E}\mid\mathcal{C})\cdot P(\mathcal{C}) \geq 1-\frac{7(K+1)N}{T^2}$$

which completes our proof. □