Supplementary Material for *Efficient Monte Carlo Counterfactual Regret Minimization in Games with Many Player Actions*

Richard Gibson, Neil Burch, Marc Lanctot, and Duane Szafron
Department of Computing Science, University of Alberta
Edmonton, Alberta, T6G 2E8, Canada
{rggibson | nburch | lanctot | dszafron}@ualberta.ca

1 Introduction

This supplementary material proves Theorems 4, 5, and 6 from the paper *Efficient Monte Carlo Counterfactual Regret Minimization in Games with Many Player Actions* and proves that Average Strategy Sampling (AS) exhibits the same regret bound given by Theorem 6.

2 Preliminaries

We begin by providing additional notation and definitions. For a history $h' \in H$, we say that the history h is a **prefix** of h', written $h \sqsubseteq h'$, if h' begins with the sequence h. For a history $h \in H_i$ and a strategy profile $\sigma \in \Sigma$, let $I(h)$ be the information set containing h and denote $\sigma(h, \cdot) = \sigma(I(h), \cdot)$. Similar to the definition of $\pi^\sigma(h, h')$, let $\pi_i^\sigma(h, h')$ and $\pi_{-i}^\sigma(h, h')$ be the probability contributed from player i and from all players/chance other than i respectively of history h' occurring after history h, given that history h has occurred. Furthermore, for $I \in \mathcal{I}_i$, define $\pi_{-i}^\sigma(I) = \sum_{h \in I} \pi_{-i}^\sigma(h)$.

Define the **counterfactual value for player i at h under σ** to be

$$v_i(h, \sigma) = \sum_{\substack{z \in Z \\ h \sqsubseteq z}} \pi_{-i}^\sigma(h) \pi^\sigma(h, z) u_i(z).$$

Notice that for $I \in \mathcal{I}_i$, perfect recall implies that

$$v_i(I, \sigma) = \sum_{h \in I} v_i(h, \sigma). \tag{1}$$

In addition, for $h \in H_i$ and a strategy $\sigma_i' \in \Sigma_i$, define

$$R_i^T(h, \sigma_i') = \sum_{t=1}^{T} (v_i(h, \sigma^t_{(I(h) \to \sigma_i')}) - v_i(h, \sigma_i^t))$$

to be the **counterfactual regret at h for σ_i'**, where $\sigma_{(I \to \sigma_i')}$ is the strategy profile σ except at I, we follow σ_i'. Note that by (1),

$$R_i^T(I, \sigma_i') = \sum_{h \in I} R_i^T(h, \sigma_i'). \tag{2}$$

Furthermore, define the **full counterfactual regret at h for σ_i'** to be

$$R_{i,\text{full}}^T(h, \sigma_i') = \sum_{t=1}^{T} (v_i(h, (\sigma_i', \sigma_{-i}^t)) - v_i(h, \sigma^t)).$$

The full counterfactual regret measures how much we wish we had played σ'_i at every history from h on, rather than playing σ^t at every time step. Notice that the regret $R_i^T = \max_{\sigma'_i \in \Sigma_i} R^T_{i,\text{full}}(\emptyset, \sigma'_i)$, where $\emptyset$ is the root of the game.

We now need some notation regarding reachable histories. Firstly, define $H_i = \{h \in H \mid P(h) = i\}$ to be the set of all histories belonging to player i. Next, for $h \in H_i$, define

$$Succ^1(h) = \{h' \in H_i \mid h \sqsubset h' \text{ and } \nexists h'' \in H_i \text{ such that } h \sqsubset h'' \sqsubset h'\}$$

to be the set of all possible next histories for player i before taking another action. For an integer $\ell > 1$, we recursively define

$$Succ^\ell(h) = \bigcup_{h' \in Succ^{\ell-1}(h)} Succ^1(h')$$

to be the set of all possible histories of player i reachable after exactly ℓ more actions by player i. Similarly, let

$$Z^1(h) = \{z \in Z \mid h \sqsubset z \text{ and } \nexists h' \in H_i \text{ such that } h \sqsubset h' \sqsubset z\}$$

be the set of all terminal histories where player i's last action was at h. Finally, define

$$D(h) = \{h\} \cup \bigcup_{\ell \geq 1} Succ^\ell(h)$$

to be the set of all nonterminal histories for player i descending from h.

3 Proof of Theorems 4, 5, and 6

Lemma A. *For $h \in H_i$ and $\sigma'_i \in \Sigma_i$,*

$$R^T_{i,\text{full}}(h, \sigma'_i) = \sum_{h' \in D(h)} \pi_i^{\sigma'}(h, h') R_i^T(h', \sigma'_i).$$

Proof. The proof is by strong induction on $|D(h)|$. Note that the base case $D(h) = \{h\}$ is trivial since $R^T_{i,\text{full}}(h, \sigma'_i) = R_i^T(h, \sigma'_i)$. For the induction step, assume that the lemma holds for all $h' \in H_i$ with $|D(h')| < |D(h)|$. To complete the proof, we must show that the lemma holds for h. To start,

$$\begin{aligned}
R^T_{i,\text{full}}(h, \sigma'_i) &= \sum_{t=1}^T v_i(h, (\sigma'_i, \sigma^t_{-i})) - \sum_{t=1}^T v_i(h, \sigma^t) \\
&= \sum_{t=1}^T \sum_{a \in A(h)} \sigma'_i(h, a) v_i(h, (\sigma'_{i(I(h)\to a)}, \sigma^t_{-i})) - \sum_{t=1}^T v_i(h, \sigma^t) \\
&= \sum_{a \in A(h)} \sigma'_i(h, a) \sum_{t=1}^T \left(\sum_{\substack{z \in Z^1(h) \\ ha \sqsubseteq z}} \pi_{-i}^{\sigma^t}(z) u_i(z) \right. \\
&\qquad \left. + \sum_{\substack{h' \in Succ^1(h) \\ ha \sqsubseteq h'}} v_i(h', (\sigma'_i, \sigma^t_{-i})) \right) - \sum_{t=1}^T v_i(h, \sigma^t).
\end{aligned} \tag{3}$$

Now, notice that for all $h' \in Succ^1(h)$, $D(h') \subset D(h)$ and $h \notin D(h')$, and so $|D(h')| < |D(h)|$ for all $h' \in Succ^1(h)$. Therefore, we may apply the induction hypothesis to each $h' \in Succ^1(h)$, giving us

$$\sum_{t=1}^T v_i(h', (\sigma'_i, \sigma^t_{-i})) = R^T_{i,\text{full}}(h', \sigma'_i) + \sum_{t=1}^T v_i(h', \sigma^t)$$

$$= \sum_{h'' \in D(h')} \pi_i^{\sigma'}(h', h'') R_i^T(h'', \sigma_i') + \sum_{t=1}^{T} v_i(h', \sigma^t)$$

for all $h' \in Succ^1(h)$. Substituting this into (3), after changing the order of summation, gives

$$\begin{aligned}
R_{i,\text{full}}^T(h, \sigma_i') = \sum_{a \in A(h)} \sigma_i'(h,a) \Bigg[& \sum_{t=1}^{T} \sum_{\substack{z \in Z^1(h) \\ ha \sqsubseteq z}} \pi_{-i}^{\sigma^t}(z) u_i(z) \\
& + \sum_{\substack{h' \in Succ^1(h) \\ ha \sqsubseteq h'}} \left(\sum_{h'' \in D(h')} \pi_i^{\sigma'}(h', h'') R_i^T(h'', \sigma_i') + \sum_{t=1}^{T} v_i(h', \sigma^t) \right) \Bigg] \\
& - \sum_{t=1}^{T} v_i(h, \sigma^t) \\
= & \sum_{a \in A(h)} \sigma_i'(h,a) \sum_{t=1}^{T} v_i(h, \sigma^t_{(I(h) \to a)}) - \sum_{t=1}^{T} v_i(h, \sigma^t) \\
& + \sum_{a \in A(h)} \sigma_i'(h,a) \sum_{\substack{h' \in Succ^1(h) \\ ha \sqsubseteq h'}} \sum_{h'' \in D(h')} \pi_i^{\sigma'}(h', h'') R_i^T(h'', \sigma_i') \\
= & \sum_{a \in A(h)} \sigma_i'(h,a) R_i^T(h,a) + \sum_{\substack{h' \in D(h) \\ h' \neq h}} \pi_i^{\sigma'}(h, h') R_i^T(h', \sigma_i') \\
= & \sum_{h' \in D(h)} \pi_i^{\sigma'}(h, h') R_i^T(h', \sigma_i'),
\end{aligned}$$

completing the proof. ■

Theorem 4.

$$R_i^T = \sum_{I \in \mathcal{I}_i} \pi_i^{\sigma^*}(I) R_i^T(I, \sigma_i^*).$$

Proof. We may assume that player i acts at the root of the game, $\emptyset$; otherwise, we may append a new root to the game that belongs to player i, is contained in a new, singleton information set, and has one action leading to the old root. Then,

$$\begin{aligned}
R_i^T &= \max_{\sigma_i' \in \Sigma_i} \sum_{t=1}^{T} (u_i(\sigma_i', \sigma_{-i}^t) - u_i(\sigma_i^t, \sigma_{-i}^t)) \\
&= R_{i,\text{full}}^T(\emptyset, \sigma_i^*) \\
&= \sum_{h \in H_i \setminus Z} \pi_i^{\sigma^*}(h) R_i^T(h, \sigma_i^*) \text{ by Lemma A} \\
&= \sum_{I \in \mathcal{I}_i} \sum_{h \in I} \pi_i^{\sigma^*}(h) R_i^T(h, \sigma_i^*) \\
&= \sum_{I \in \mathcal{I}_i} \pi_i^{\sigma^*}(I) \sum_{h \in I} R_i^T(h, \sigma_i^*) \text{ due to perfect recall} \\
&= \sum_{I \in \mathcal{I}_i} \pi_i^{\sigma^*}(I) R_i^T(I, \sigma_i^*),
\end{aligned}$$

where the last line follows by equation (2). ■

Theorem 5. *When using vanilla CFR, average regret is bounded by*

$$\frac{R_i^T}{T} \leq \frac{\Delta_i M_i(\sigma_i^*) \sqrt{|A_i|}}{\sqrt{T}}.$$

Proof. Following the proof of Theorem 2 [10],

$$\begin{aligned}
R_i^T &= \sum_{I \in \mathcal{I}_i} \pi_i^{\sigma^*}(I) R_i^T(I, \sigma_i^*) \text{ by Theorem 4} \\
&= \sum_{I \in \mathcal{I}_i} \pi_i^{\sigma^*}(I) \sum_{a \in A(I)} \sigma_i^*(I,a) R_i^T(I,a) \\
&= \sum_{I \in \mathcal{I}_i} \pi_i^{\sigma^*}(I) \max_{a \in A(I)} R_i^T(I,a) \\
&\le \sum_{I \in \mathcal{I}_i} \pi_i^{\sigma^*}(I) \sqrt{\sum_{a \in A(I)} T^2 (R_i^{T,+}(I,a)/T)^2} \\
&\le \sum_{I \in \mathcal{I}_i} \pi_i^{\sigma^*}(I) \Delta_i \sqrt{|A(I)|} \sqrt{\sum_{t=1}^{T} (\pi_{-i}^{\sigma^t}(I))^2} \\
&\text{by Theorem 6 of [10] with } \Delta_t = \Delta_i \pi_{-i}^{\sigma^t}(I) \\
&\le \Delta_i \sqrt{|A_i|} \sum_{B \in \mathcal{B}_i} \pi_i^{\sigma^*}(B) \sum_{I \in B} \sqrt{\sum_{t=1}^{T} (\pi_{-i}^{\sigma^t}(I))^2} \\
&\le \Delta_i \sqrt{|A_i|} \sum_{B \in \mathcal{B}_i} \pi_i^{\sigma^*}(B) \sqrt{|B| \sum_{t=1}^{T} \sum_{I \in B} \pi_{-i}^{\sigma^t}(I)} \\
&\text{by Lemma 6 of [10]} \\
&\le \Delta_i \sqrt{|A_i|} \sum_{B \in \mathcal{B}_i} \pi_i^{\sigma^*}(B) \sqrt{|B| T} \text{ by Lemma 16 of [10]} \\
&= \Delta_i \sqrt{|A_i| T} M_i(\sigma_i^*).
\end{aligned}$$

Dividing both sides by T gives the result. ■

We now prove a general, probabilistic bound that can be applied to any MCCFR sampling algorithm. We then use this bound to prove Theorem 6 and a similar bound for AS.

Lemma B. *Let $p, \delta \in (0,1]$. When using any MCCFR algorithm, if*

$$\sum_{I \in B} \left(\sum_{z \in Q \cap Z_I} \frac{\pi^{\sigma^t}(z[I], z) \pi_{-i}^{\sigma^t}(z[I])}{q(z)} \right)^2 \le \frac{1}{\delta^2}$$

for all $Q \in \mathcal{Q}$, $B \in \mathcal{B}_i$, and $t \le T$, then with probability at least $1-p$, average regret is bounded by

$$\frac{R_i^T}{T} \le \left(M_i(\sigma_i^*) + \frac{\sqrt{2 |\mathcal{I}_i| |\mathcal{B}_i|}}{\sqrt{p}} \right) \left(\frac{1}{\delta} \right) \frac{\Delta_i \sqrt{|A_i|}}{\sqrt{T}}.$$

Proof. Our proof follows that of Theorem 7 in [10]. To start, define

$$\Delta_i^t(I) = \Delta_i \sum_{z \in Q \cap Z_I} \frac{\pi^{\sigma^t}(z[I], z) \pi_{-i}^{\sigma^t}(z[I])}{q(z)}$$

so that the difference between two sampled counterfactual values at information set I is bounded by

$$\tilde{v}_i(I, \sigma_{(I \to a)}^t) - \tilde{v}_i(I, \sigma_{(I \to b)}^t) \le \Delta_i^t(I)$$

for all $a, b \in A(I)$. By our assumption, we then have

$$\sum_{I \in B} (\Delta_i^t(I))^2 \le \frac{(\Delta_i)^2}{\delta^2} \tag{4}$$

for all $B \in \mathcal{B}_i$.

Define $R_i^T(I) = \max_{a \in A(I)} R_i^T(I,a)$ and $\tilde{R}_i^T(I) = \max_{a \in A(I)} \tilde{R}_i^T(I,a)$. The proof will proceed as follows. First, we prove a bound on the weighted sum of the cumulative sampled counterfactual regrets $\sum_{I \in \mathcal{I}} \pi_i^{\sigma^*}(I)\tilde{R}_i^T(I)$. Secondly, we prove a probabilistic bound on the expected squared difference between $\sum_{I \in \mathcal{I}_i} \pi_i^{\sigma^*}(I) R_i^T(I)$ and $\sum_{I \in \mathcal{I}} \pi_i^{\sigma^*}(I)\tilde{R}_i^T(I)$, showing that the true counterfactual regrets are not too far from the sampled counterfactual regrets. Finally, we apply Theorem 4 to obtain the bound on the average regret.

For the first step,

$$\begin{aligned}
\sum_{I \in \mathcal{I}_i} \pi_i^{\sigma^*}(I)\tilde{R}_i^T(I) &\le \sum_{I \in \mathcal{I}_i} \pi_i^{\sigma^*}(I) \sqrt{T^2 \sum_{a \in A(I)} \left(\frac{\tilde{R}_i^{T,+}(I,a)}{T} \right)^2} \\
&\le \sum_{I \in \mathcal{I}_i} \pi_i^{\sigma^*}(I) \sqrt{|A(I)| \sum_{t=1}^{T} (\Delta_i^t(I))^2} \\
&\quad \text{by Theorem 6 of [10]} \\
&\le \sqrt{|A_i|} \sum_{B \in \mathcal{B}_i} \pi_i^{\sigma^*}(B) \sum_{I \in B} \sqrt{\sum_{t=1}^{T} (\Delta_i^t(I))^2} \\
&\le \sqrt{|A_i|} \sum_{B \in \mathcal{B}_i} \pi_i^{\sigma^*}(B) \sqrt{|B| \sum_{t=1}^{T} \sum_{I \in B} (\Delta_i^t(I))^2} \\
&\quad \text{by Lemma 5 of [10]} \\
&\le \sqrt{|A_i|} \sum_{B \in \mathcal{B}_i} \pi_i^{\sigma^*}(B) \sqrt{|B| T \frac{(\Delta_i)^2}{\delta^2}} \text{ by equation (4)} \\
&= \frac{\Delta_i M_i(\sigma_i^*) \sqrt{|A_i| T}}{\delta}.
\end{aligned} \tag{5}$$

Secondly, for $I \in \mathcal{I}_i$,

$$\begin{aligned}
\left(R_i^T(I) - \tilde{R}_i^T(I) \right)^2 &= \left(\max_{a \in A(I)} \sum_{t=1}^{T} r_i^t(I,a) - \max_{a \in A(I)} \sum_{t=1}^{T} \tilde{r}_i^t(I,a) \right)^2 \\
&\le \left(\max_{a \in A(I)} \sum_{t=1}^{T} \left(r_i^t(I,a) - \tilde{r}_i^t(I,a) \right) \right)^2 \\
&\le \max_{a \in A(I)} \left(\sum_{t=1}^{T} \left(r_i^t(I,a) - \tilde{r}_i^t(I,a) \right) \right)^2 \\
&\le \sum_{a \in A(I)} \left[\sum_{t=1}^{T} \left(r_i^t(I,a) - \tilde{r}_i^t(I,a) \right)^2 \right. \\
&\quad \left. + 2 \sum_{t=1}^{T} \sum_{t'=t+1}^{T} \left(r_i^t(I,a) - \tilde{r}_i^t(I,a) \right) \left(r_i^{t'}(I,a) - \tilde{r}_i^{t'}(I,a) \right) \right].
\end{aligned} \tag{6}$$

We now multiply both sides by $(\pi_i^{\sigma^*}(I))^2$ and take the expectation of both sides. Note that

$$\begin{aligned}
&\mathbf{E} \left[\left(r_i^t(I,a) - \tilde{r}_i^t(I,a) \right) \left(r_i^{t'}(I,a) - \tilde{r}_i^{t'}(I,a) \right) \right] \\
&= \mathbf{E} \left[\mathbf{E} \left[\left(r_i^{t'}(I,a) - \tilde{r}_i^{t'}(I,a) \right) \mid r_i^t(I,a), \tilde{r}_i^t(I,a) \right] \left(r_i^t(I,a) - \tilde{r}_i^t(I,a) \right) \right]
\end{aligned}$$

and that $\mathbf{E}\left[(r_i^{t'}(I,a) - \tilde{r}_i^{t'}(I,a)) \mid r_i^t(I,a), \tilde{r}_i^t(I,a)\right] = 0$ since for $t' > t$, $\tilde{r}_i^{t'}$ is an unbiased estimate of $r_i^{t'}$ given $\sigma^{t'}$. Thus from equation (6), we have

$$\begin{aligned}
\mathbf{E}\left[(\pi_i^{\sigma^*}(I))^2 \left(R_i^T(I) - \tilde{R}_i^T(I)\right)^2\right] &\le \sum_{a \in A(I)} \sum_{t=1}^{T} \mathbf{E}\left[(\pi_i^{\sigma^*}(I))^2 \left(r_i^t(I,a) - \tilde{r}_i^t(I,a)\right)^2\right] \\
&\le \sum_{a \in A(I)} \sum_{t=1}^{T} \mathbf{E}\left[\left(r_i^t(I,a)\right)^2 + \left(\tilde{r}_i^t(I,a)\right)^2\right] \\
&\le \sum_{a \in A(I)} \sum_{t=1}^{T} \left[\left(\pi_{-i}^{\sigma^t}(I)\right)^2 \Delta_i^2 + \left(\Delta_i^t(I)\right)^2\right]. \qquad (7)
\end{aligned}$$

We can now bound the expected squared difference between $\sum_{I \in \mathcal{I}_i} \pi_i^{\sigma^*}(I) R_i^T(I)$ and $\sum_{I \in \mathcal{I}_i} \pi_i^{\sigma^*}(I) \tilde{R}_i^T(I)$ by

$$\begin{aligned}
&\mathbf{E}\left[\left(\sum_{I \in \mathcal{I}_i} \pi_i^{\sigma^*}(I) \left(R_i^T(I) - \tilde{R}_i^T(I)\right)\right)^2\right] \\
&\quad \le \mathbf{E}\left[\left(\sum_{I \in \mathcal{I}_i} \left|\pi_i^{\sigma^*}(I) \left(R_i^T(I) - \tilde{R}_i^T(I)\right)\right|\right)^2\right] \\
&\quad \le \mathbf{E}\left[\left(\sqrt{|\mathcal{I}_i| \sum_{I \in \mathcal{I}_i} \left|\pi_i^{\sigma^*}(I) \left(R_i^T(I) - \tilde{R}_i^T(I)\right)\right|^2}\right)^2\right] \\
&\qquad \text{by Lemma 5 of [10]} \\
&\quad = |\mathcal{I}_i| \sum_{I \in \mathcal{I}_i} \mathbf{E}\left[\left(\pi_i^{\sigma^*}(I)\right)^2 \left(R_i^T(I) - \tilde{R}_i^T(I)\right)^2\right] \\
&\quad \le |\mathcal{I}_i| \sum_{I \in \mathcal{I}_i} \sum_{a \in A(I)} \sum_{t=1}^{T} \left[\left(\pi_{-i}^{\sigma^t}(I)\right)^2 \Delta_i^2 + \left(\Delta_i^t(I)\right)^2\right] \\
&\qquad \text{by equation (7)} \\
&\quad \le |\mathcal{I}_i||A_i| \sum_{B \in \mathcal{B}_i} \sum_{t=1}^{T} \left[\sum_{I \in B} \left(\pi_{-i}^{\sigma^t}(I)\right)^2 \Delta_i^2 + \sum_{I \in B} \left(\Delta_i^t(I)\right)^2\right] \\
&\quad \le |\mathcal{I}_i||A_i| \sum_{B \in \mathcal{B}_i} \sum_{t=1}^{T} \left[\Delta_i^2 + \frac{\Delta_i^2}{\delta^2}\right] \\
&\qquad \text{by Lemma 16 of [10] and equation (4)} \\
&\quad \le \frac{2|\mathcal{I}_i||A_i||\mathcal{B}_i| T \Delta_i^2}{\delta^2} \qquad (8)
\end{aligned}$$

Finally, with probability $1 - p$, we can bound the regret by

$$\begin{aligned}
R_i^T &= \sum_{I \in \mathcal{I}_i} \pi_i^{\sigma^*}(I) R_i^T(I) \text{ by Theorem 4} \\
&= \sum_{I \in \mathcal{I}_i} \pi_i^{\sigma^*}(I) \left(R_i^T(I) - \tilde{R}_i^T(I) + \tilde{R}_i^T(I)\right) \\
&\le \left|\sum_{I \in \mathcal{I}_i} \pi_i^{\sigma^*}(I) \left(R_i^T(I) - \tilde{R}_i^T(I)\right)\right| + \sum_{I \in \mathcal{I}_i} \pi_i^{\sigma^*}(I) \tilde{R}_i^T(I)
\end{aligned}$$

$$\leq \frac{1}{\sqrt{p}}\sqrt{\mathbf{E}\left[\left(\sum_{I\in\mathcal{I}_i}\pi_i^{\sigma^*}(I)\left(R_i^T(I)-\tilde{R}_i^T(I)\right)\right)^2\right]}+\frac{\Delta_i M_i(\sigma_i^*)\sqrt{|A_i|T}}{\delta}$$

by Lemma 2 of [10] and equation (5)

$$\leq \left(\frac{\sqrt{2|\mathcal{I}_i||\mathcal{B}_i|}}{\sqrt{p}}+M_i(\sigma_i^*)\right)\left(\frac{1}{\delta}\right)\Delta_i\sqrt{|A_i|T}$$

by equation (8). Dividing both sides by T gives the result. ■

Theorem 6′. *Let X be one of CS, ES, OS (assuming OS samples opponent actions according to σ_{-i}), or AS, let $p \in (0,1]$, and let $\delta = \min_{z\in Z} q_i(z) > 0$ over all $1 \leq t \leq T$. When using X, with probability $1-p$, average regret is bounded by*

$$\frac{R_i^T}{T} \leq \left(M_i(\sigma_i^*)+\frac{\sqrt{2|\mathcal{I}_i||\mathcal{B}_i|}}{\sqrt{p}}\right)\left(\frac{1}{\delta}\right)\frac{\Delta_i\sqrt{|A_i|}}{\sqrt{T}}.$$

Proof. By Lemma B, it suffices to show that

$$Y=\sum_{I\in B}\left(\sum_{z\in Q\cap Z_I}\frac{\pi^{\sigma^t}(z[I],z)\pi_{-i}^{\sigma^t}(z[I])}{q(z)}\right)^2 \leq \frac{1}{\delta^2}$$

for all $B \in \mathcal{B}_i$, $Q \in \mathcal{Q}$, and $t \leq T$. To that end, fix $B \in \mathcal{B}_i$, $Q \in \mathcal{Q}$, and $t \leq T$. Since X samples a single action at each $h \in H_c$ according to σ_c, there exists a unique $a_h^* \in A(h)$ such that if $z \in Q$ and $h \sqsubseteq z$, then $ha_h^* \sqsubseteq z$. Consider the new chance probability distribution $\hat{\sigma}_c$ defined according to

$$\hat{\sigma}_c(h,a)=\begin{cases}1 & \text{if } a=a_h^*\\ 0 & \text{if } a\neq a_h^*\end{cases}$$

for all $h \in H_c$, $a \in A(h)$. When $X \neq CS$, we also have a unique such action a_I^* for each $I \in \mathcal{I}_{-i}$ sampled according to σ_{-i}^t, so we can similarly define the new opponent profile $\hat{\sigma}_{-i}$ according to

$$\hat{\sigma}_{-i}(I,a)=\begin{cases}\sigma_{-i}^t(I,a) & \text{if } X = \text{ CS}\\ 1 & \text{if } X\neq \text{ CS and } a=a_I^*\\ 0 & \text{if } X\neq \text{ CS and } a\neq a_I^*\end{cases}$$

for all $I \in \mathcal{I}_{-i}$, $a \in A(I)$. Then

$$\begin{aligned}Y&=\sum_{I\in B}\left(\sum_{z\in Q\cap Z_I}\frac{\pi_i^{\sigma^t}(z[I],z)\pi_{-i}^{\sigma^t}(z)}{q(z)}\right)^2\\&=\sum_{I\in B}\left(\sum_{z\in Z_I}\frac{\pi_i^{\sigma^t}(z[I],z)\pi_{-i}^{\hat{\sigma}}(z)}{q_i(z)}\right)^2\\&\leq\frac{1}{\delta^2}\sum_{I\in B}\left(\sum_{z\in Z_I}\pi_{-i}^{\hat{\sigma}}(z)\right)^2\\&=\frac{1}{\delta^2}\sum_{I\in B}\left(\pi_{-i}^{\hat{\sigma}}(I)\right)^2\\&\leq\frac{1}{\delta^2},\end{aligned}$$

where the last line follows by Lemma 16 of [10]. ■