A. Discussion of Assumptions

In this section, we prove that the combinatorial semi-bandit and the cascading bandit satisfy Assumptions 1 and 2 proposed in Section 5.

A.1. Combinatorial Semi-Bandits

Notice that in a combinatorial semi-bandit, the action $a = (a_1, \ldots, a_K)$, and

$$r(a, \theta) = \sum_{k=1}^{K} \theta^{(a_k)} = \sum_{l=1}^{L} \theta^{(l)} \mathbf{1}(l \in a).$$

Thus, for any l, $r(a, \theta)$ is weakly increasing in $\theta^{(l)}$. Hence Assumption 1 is satisfied. On the other hand, we have

$$\begin{aligned} |r(a, \theta_1) - r(a, \theta_2)| &= \left| \sum_{l=1}^{L} \left(\theta_1^{(l)} - \theta_2^{(l)} \right) \mathbf{1}(l \in a) \right| \\ &\leq \sum_{l=1}^{L} \left| \theta_1^{(l)} - \theta_2^{(l)} \right| \mathbf{1}(l \in a) = \sum_{l=1}^{L} P\left(E^{(l)} \middle| \theta_2, a \right) \left| \theta_1^{(l)} - \theta_2^{(l)} \right|, \end{aligned} \tag{7}$$

where the last quality follows from the fact that all nodes in a combinatorial semi-bandit is observed, and hence $P\left(E^{(l)} \middle| \theta, a\right) = \mathbf{1}(l \in a)$ for all θ. Thus, Assumption 2 is satisfied with $C = 1$.

A.2. Cascading Bandits

For a cascading bandit, the action is $a = (a_1, \ldots, a_K)$, and

$$r(a, \theta) = 1 - \prod_{k=1}^{K} (1 - \theta^{(a_k)}) = 1 - \prod_{l \in a} (1 - \theta^{(l)}).$$

Thus, for any l, $r(a, \theta)$ is weakly increasing in $\theta^{(l)}$. Hence Assumption 1 is satisfied. On the other hand, from Kveton et al. (2015a), we have

$$\begin{aligned} r(a, \theta_1) - r(a, \theta_2) &= \sum_{k=1}^{K} \prod_{k_1=1}^{k-1} \left(1 - \theta_2^{(a_{k_1})}\right) \left(\theta_1^{(a_k)} - \theta_2^{(a_k)}\right) \prod_{k_2=k+1}^{K} \left(1 - \theta_1^{(a_{k_2})}\right) \\ &= \sum_{k=1}^{K} P\left(E^{(a_k)} \middle| \theta_2, a\right) \left(\theta_1^{(a_k)} - \theta_2^{(a_k)}\right) \prod_{k_2=k+1}^{K} \left(1 - \theta_1^{(a_{k_2})}\right), \end{aligned}$$

where the second equality follows from $P\left(E^{(a_k)} \middle| \theta_2, a\right) = \prod_{k_1=1}^{k-1} \left(1 - \theta_2^{(a_{k_1})}\right)$. Thus, we have

$$\begin{aligned} |r(a, \theta_1) - r(a, \theta_2)| &= \left| \sum_{k=1}^{K} P\left(E^{(a_k)} \middle| \theta_2, a\right) \left(\theta_1^{(a_k)} - \theta_2^{(a_k)}\right) \prod_{k_2=k+1}^{K} \left(1 - \theta_1^{(a_{k_2})}\right) \right| \\ &\leq \sum_{k=1}^{K} P\left(E^{(a_k)} \middle| \theta_2, a\right) \left|\theta_1^{(a_k)} - \theta_2^{(a_k)}\right| \prod_{k_2=k+1}^{K} \left(1 - \theta_1^{(a_{k_2})}\right) \\ &\leq \sum_{k=1}^{K} P\left(E^{(a_k)} \middle| \theta_2, a\right) \left|\theta_1^{(a_k)} - \theta_2^{(a_k)}\right|, \end{aligned}$$

where the last inequality follows from $\prod_{k_2=k+1}^{K} \left(1 - \theta_1^{(a_{k_2})}\right) \in [0, 1]$. Thus, Assumption 2 is satisfied with $C = 1$.

B. Proof for Theorem 1

Proof:
Recall that the stochastic instantaneous reward is $r(x,z)$. Note that $r(x,z)$ is bounded since its domain is finite. Without loss of generality, we assume that $r(x,z) \in [0,B]$. Thus, for any action a and probability measure $\theta \in [0,1]^{d+L}$, we have $r(a,\theta) \in [0,B]$.

Define $R_t = r(a^*, \theta_*) - r(a_t, \theta_*)$, then by definition, we have

$$R_B(n) = \sum_{t=1}^{n} \mathbb{E}[R_t] = \sum_{t=1}^{n} \mathbb{E}\left[E[R_t | \mathcal{H}_{t-1}]\right],$$

where $\mathcal{H}_{t-1}$ is the "history" by the end of time $t-1$, which includes all the actions and observations by that time[5]. For any parameter index $i = 1, \ldots, d+L$ and any time t, we define $N_t^{(i)} = \sum_{\tau=1}^{t} \mathbf{1}\left[E_\tau^{(i)}\right]$ as the number of times that the samples corresponding to parameter $\theta_*^{(i)}$ have been observed by the end of time t, and $\hat{\theta}_t^{(i)}$ as the empirical mean for $\theta_*^{(i)}$ based on these $N_t^{(i)}$ observations. Then we define the upper confidence bound (UCB) $U_t^{(i)}$ and the lower confidence bound (LCB) $L_t^{(i)}$ as

$$U_t^{(i)} = \begin{cases} \min\left\{\hat{\theta}_t^{(i)} + c\left(t, N_t^{(i)}\right), 1\right\} & \text{if } N_t^{(i)} > 0 \\ 1 & \text{otherwise} \end{cases}$$

$$L_t^{(i)} = \begin{cases} \max\left\{\hat{\theta}_t^{(i)} - c\left(t, N_t^{(i)}\right), 0\right\} & \text{if } N_t^{(i)} > 0 \\ 0 & \text{otherwise} \end{cases}$$

where $c\left(t, N\right) = \sqrt{\frac{1.5\log(t)}{N}}$ for any positive integer t and N. Moreover, we define a probability measure $\tilde{\theta}_t \in [0,1]^{d+L}$ as

$$\vartheta_t^{(i)} = \begin{cases} U_t^{(i)} & \text{if } i \in \mathcal{I}^+ \\ L_t^{(i)} & \text{if } i \in \mathcal{I}^- \end{cases}$$

Since both $N_{t-1}^{(i)}$ and $\hat{\theta}_{t-1}^{(i)}$ are conditionally deterministic given $\mathcal{H}_{t-1}$, and $\mathcal{I}^+$ and $\mathcal{I}^-$ are deterministic, by the definitions above, U_{t-1}, L_{t-1} and ϑ_{t-1} are also conditionally deterministic given $\mathcal{H}_{t-1}$. Moreover, as is discussed in Russo & Van Roy (2014), since we apply exact Thompson sampling `idTS`, θ_* and θ_t are conditionally i.i.d. given $\mathcal{H}_{t-1}$, and $a^* = \arg\max_a r(a, \theta_*)$ and $a_t = \arg\max_a r(a, \theta_t)$. Thus, conditioning on $\mathcal{H}_{t-1}$, $r(a^*, \vartheta_{t-1})$ and $r(a_t, \vartheta_{t-1})$ are i.i.d., consequently, we have

$$\begin{aligned} \mathbb{E}[R_t | \mathcal{H}_{t-1}] &= \mathbb{E}[r(a^*, \theta_*) - r(a_t, \theta_*) | \mathcal{H}_{t-1}] \\ &= \mathbb{E}[r(a^*, \theta_*) - r(a^*, \vartheta_{t-1}) | \mathcal{H}_{t-1}] + \mathbb{E}[r(a_t, \vartheta_{t-1}) - r(a_t, \theta_*) | \mathcal{H}_{t-1}]. \end{aligned} \tag{8}$$

To simplify the exposition, for any time t and $i = 1, \ldots, d+L$, we define

$$G_t^{(i)} = \left\{\left|\theta_*^{(i)} - \hat{\theta}_t^{(i)}\right| > c\left(t, N_t^{(i)}\right),\ N_t^{(i)} > 0\right\} = \left\{\theta_*^{(i)} > U_t^{(i)} \text{ or } \theta_*^{(i)} < L_t^{(i)}\right\}. \tag{9}$$

Notice that $\overline{\bigcup_{i=1}^{d+L} G_t^{(i)}} = \bigcap_{i=1}^{d+L} \overline{G_t^{(i)}} = \{L_t \leq \theta_* \leq U_t\}$. Moreover, from Assumption 1, if $L_t \leq \theta_* \leq U_t$, based on the definition of ϑ_t, we have $r(a, \theta_*) \leq r(a, \vartheta_t)$ for all action a. Thus, we have

$$\begin{aligned} r(a^*, \theta_*) - r(a^*, \vartheta_{t-1}) &\overset{(a)}{=} [r(a^*, \theta_*) - r(a^*, \vartheta_{t-1})]\,\mathbf{1}\left(L_{t-1} \leq \theta_* \leq U_{t-1}\right) \\ &\quad + [r(a^*, \theta_*) - r(a^*, \vartheta_{t-1})]\,\mathbf{1}\left(\bigcup_{i=1}^{d+L} G_{t-1}^{(i)}\right) \\ &\overset{(b)}{\leq} [r(a^*, \theta_*) - r(a^*, \vartheta_{t-1})]\,\mathbf{1}\left(\bigcup_{i=1}^{d+L} G_{t-1}^{(i)}\right) \\ &\overset{(c)}{\leq} B\mathbf{1}\left(\bigcup_{i=1}^{d+L} G_{t-1}^{(i)}\right) \overset{(d)}{\leq} B\sum_{i=1}^{d+L} \mathbf{1}\left(G_{t-1}^{(i)}\right), \end{aligned} \tag{10}$$

[5] Rigorously speaking, $\{\mathcal{H}_t\}_{t=0}^{n-1}$ is a filtration and $\mathcal{H}_{t-1}$ is a σ-algebra.

where equality (a) is simply a decomposition based on indicators, inequality (b) follows from the fact that $r(a,\theta_*) \leq r(a,\vartheta_{t-1})$ if $L_{t-1} \leq \theta_* \leq U_{t-1}$, inequality (c) follows from the fact that $r(X,Z) \in [0,B]$ for all (X,Z) and hence $r(a,\theta) \in [0,B]$ for all a and θ, and inequality (d) trivially follows from the union bound of the indicators.

On the other hand, we have

$$\begin{aligned} r(a_t,\vartheta_{t-1}) - r(a_t,\theta_*) = &\left[r(a_t,\vartheta_{t-1}) - r(a_t,\theta_*)\right] \mathbf{1}\left(L_{t-1} \leq \theta_* \leq U_{t-1}\right) \\ &+ \left[r(a_t,\vartheta_{t-1}) - r(a_t,\theta_*)\right] \mathbf{1}\left(\bigcup_{i=1}^{d+L} G_{t-1}^{(i)}\right). \end{aligned}$$

Similarly as the above analysis, we have

$$\left[r(a_t,\vartheta_{t-1}) - r(a_t,\theta_*)\right] \mathbf{1}\left(\bigcup_{i=1}^{d+L} G_{t-1}^{(i)}\right) \leq B \sum_{i=1}^{d+L} \mathbf{1}\left(G_{t-1}^{(i)}\right). \tag{11}$$

On the other hand, we have

$$\begin{aligned} \left[r(a_t,\vartheta_{t-1}) - r(a_t,\theta_*)\right] \mathbf{1}\left(L_{t-1} \leq \theta_* \leq U_{t-1}\right) &\overset{(a)}{\leq} C \sum_{i=1}^{d+L} P\left(E_t^{(i)} \middle| \theta_*, a_t\right) \left|\vartheta_{t-1}^{(i)} - \theta_*^{(i)}\right| \mathbf{1}\left(L_{t-1} \leq \theta_* \leq U_{t-1}\right) \\ &\overset{(b)}{\leq} C \sum_{i=1}^{d+L} P\left(E_t^{(i)} \middle| \theta_*, a_t\right) \left[U_{t-1}^{(i)} - L_{t-1}^{(i)}\right] \mathbf{1}\left(L_{t-1} \leq \theta_* \leq U_{t-1}\right) \\ &\overset{(c)}{\leq} C \sum_{i=1}^{d+L} P\left(E_t^{(i)} \middle| \theta_*, a_t\right) \left[U_{t-1}^{(i)} - L_{t-1}^{(i)}\right], \end{aligned}$$

where inequality (a) follows from Assumption 2, inequality (b) follows trivially from $L_{t-1} \leq \theta_* \leq U_{t-1}$ and the definition of ϑ_{t-1}, and inequality (c) follows from the fact that $U_{t-1}^{(i)} > L_{t-1}^{(i)}$ always holds, no matter what θ_* is. Combining the above results, we have

$$\begin{aligned} \mathbb{E}[R_t|\mathcal{H}_{t-1}] &\leq C \sum_{i=1}^{d+L} \mathbb{E}\left[P\left(E_t^{(i)} \middle| \theta_*, a_t\right) \left[U_{t-1}^{(i)} - L_{t-1}^{(i)}\right] \middle| \mathcal{H}_{t-1}\right] + 2B \sum_{i=1}^{d+L} \mathbb{E}\left[\mathbf{1}\left(G_{t-1}^{(i)}\right) \middle| \mathcal{H}_{t-1}\right] \\ &\overset{(a)}{=} C \sum_{i=1}^{d+L} \mathbb{E}\left[P\left(E_t^{(i)} \middle| \theta_*, a_t\right) \middle| \mathcal{H}_{t-1}\right] \left[U_{t-1}^{(i)} - L_{t-1}^{(i)}\right] + 2B \sum_{i=1}^{d+L} \mathbb{E}\left[\mathbf{1}\left(G_{t-1}^{(i)}\right) \middle| \mathcal{H}_{t-1}\right] \\ &\overset{(b)}{=} C \sum_{i=1}^{d+L} \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\left(E_t^{(i)}\right) \middle| \theta_*, a_t\right] \middle| \mathcal{H}_{t-1}\right] \left[U_{t-1}^{(i)} - L_{t-1}^{(i)}\right] + 2B \sum_{i=1}^{d+L} \mathbb{E}\left[\mathbf{1}\left(G_{t-1}^{(i)}\right) \middle| \mathcal{H}_{t-1}\right] \\ &\overset{(c)}{=} C \sum_{i=1}^{d+L} \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\left(E_t^{(i)}\right) \middle| \theta_*, a_t, \mathcal{H}_{t-1}\right] \middle| \mathcal{H}_{t-1}\right] \left[U_{t-1}^{(i)} - L_{t-1}^{(i)}\right] + 2B \sum_{i=1}^{d+L} \mathbb{E}\left[\mathbf{1}\left(G_{t-1}^{(i)}\right) \middle| \mathcal{H}_{t-1}\right] \\ &\overset{(d)}{=} C \sum_{i=1}^{d+L} \mathbb{E}\left[\mathbf{1}\left(E_t^{(i)}\right) \left[U_{t-1}^{(i)} - L_{t-1}^{(i)}\right] \middle| \mathcal{H}_{t-1}\right] + 2B \sum_{i=1}^{d+L} \mathbb{E}\left[\mathbf{1}\left(G_{t-1}^{(i)}\right) \middle| \mathcal{H}_{t-1}\right], \end{aligned}$$

where (a) follows from the fact that U_{t-1} and L_{t-1} are deterministic conditioning on $\mathcal{H}_{t-1}$, (b) follows from the definition of $P\left(E_t^{(i)} \middle| \theta_*, a_t\right)$, (c) follows from that fact that conditioning on θ_* and a_t, $E_t^{(i)}$ is independent of $\mathcal{H}_{t-1}$, and (d) follows from the tower property. Thus we have

$$R_B(n) \leq C\mathbb{E}\left[\sum_{i=1}^{d+L} \sum_{t=1}^{n} \mathbf{1}\left(E_t^{(i)}\right) \left[U_{t-1}^{(i)} - L_{t-1}^{(i)}\right]\right] + 2B \sum_{i=1}^{d+L} \sum_{t=1}^{n} P\left(G_{t-1}^{(i)}\right). \tag{12}$$

We first bound the second term. Notice that we have $P\left(G_{t-1}^{(i)}\right) = \mathbb{E}\left[P\left(G_{t-1}^{(i)} \middle| \theta_*\right)\right]$. For any θ_*, we have

$$P\left(G_{t-1}^{(i)} \middle| \theta_*\right) = P\left(\left|\theta_*^{(i)} - \hat{\theta}_{N_{t-1}^{(i)}}^{(i)}\right| > c\left(t, N_{t-1}^{(i)}\right),\, N_{t-1}^{(i)} > 0 \middle| \theta_*\right),$$

where we use subscript $N_{t-1}^{(i)}$ for $\hat{\theta}$ to emphasize it is an empirical mean over $N_{t-1}^{(i)}$ samples. Following the union bound developed in Auer et al. (2002), we have

$$P\left(G_{t-1}^{(i)}\middle|\theta_*\right) = P\left(\left|\theta_*^{(i)} - \hat{\theta}_{N_{t-1}^{(i)}}^{(i)}\right| > c\left(t, N_{t-1}^{(i)}\right),\, N_{t-1}^{(i)} > 0\middle|\theta_*\right)$$
$$\overset{(a)}{\leq} \sum_{N=1}^{t-1} P\left(\left|\theta_*^{(i)} - \hat{\theta}_N^{(i)}\right| > c\left(t, N\right)\middle|\theta_*\right) \overset{(b)}{\leq} \sum_{t=1}^{N-1} \frac{2}{t^3} < \frac{2}{t^2},$$

where inequality (a) follows from the union bound over the realization of $N_{t-1}^{(i)}$, and inequality (b) follows from the Hoeffding's inequality. Since the above inequality holds for any θ_*, we have $P\left(G_{t-1}^{(i)}\right) < \frac{2}{t^2}$. Thus,

$$\sum_{i=1}^{d+L}\sum_{t=1}^{n} P\left(G_{t-1}^{(i)}\right) < \sum_{i=1}^{d+L}\sum_{t=1}^{n} \frac{2}{t^2} < (d+L)\sum_{t=1}^{\infty}\frac{2}{t^2} = \frac{(d+L)\pi^2}{3}.$$

We now try to bound the first term of equation 12. Notice that trivially, we have

$$U_{t-1}^{(i)} - L_{t-1}^{(i)} \leq 2c\left(t, N_{t-1}^{(i)}\right)\mathbf{1}\left(N_{t-1}^{(i)} > 0\right) + \mathbf{1}\left(N_{t-1}^{(i)} = 0\right)$$
$$= 2\sqrt{\frac{1.5\log(t)}{N_{t-1}^{(i)}}}\mathbf{1}\left(N_{t-1}^{(i)} > 0\right) + \mathbf{1}\left(N_{t-1}^{(i)} = 0\right)$$
$$\leq \sqrt{6\log(n)}\frac{1}{\sqrt{N_{t-1}^{(i)}}}\mathbf{1}\left(N_{t-1}^{(i)} > 0\right) + \mathbf{1}\left(N_{t-1}^{(i)} = 0\right).$$

Thus, we have

$$\sum_{i=1}^{d+L}\sum_{t=1}^{n}\mathbf{1}\left(E_t^{(i)}\right)\left[U_{t-1}^{(i)} - L_{t-1}^{(i)}\right] \leq \sqrt{6\log(n)}\sum_{i=1}^{d+L}\sum_{t=1}^{n}\frac{1}{\sqrt{N_{t-1}^{(i)}}}\mathbf{1}\left(E_t^{(i)},\, N_{t-1}^{(i)} > 0\right) + (d+L).$$

Notice that from the Cauchy–Schwarz inequality, we have

$$\sum_{i=1}^{d+L}\sum_{t=1}^{n}\frac{1}{\sqrt{N_{t-1}^{(i)}}}\mathbf{1}\left(E_t^{(i)},\, N_{t-1}^{(i)} > 0\right) \leq \sqrt{\sum_{t=1}^{n}\sum_{i=1}^{d+L}\mathbf{1}\left(E_t^{(i)}\right)}\sqrt{\sum_{i=1}^{d+L}\sum_{t=1}^{n}\frac{1}{N_{t-1}^{(i)}}\mathbf{1}\left(N_{t-1}^{(i)} > 0\right)}. \tag{13}$$

Moreover, we have

$$\sum_{i=1}^{d+L}\sum_{t=1}^{n}\frac{1}{N_{t-1}^{(i)}}\mathbf{1}\left(N_{t-1}^{(i)} > 0\right) < (d+L)\sum_{N=1}^{n}\frac{1}{N} < (d+L)\left(1 + \int_{z=1}^{n}\frac{1}{z}dz\right) = (d+L)(1+\log(n)).$$

Consequently, we have

$$\mathbb{E}\left[\sum_{i=1}^{d+L}\sum_{t=1}^{n}\mathbf{1}\left(E_t^{(i)}\right)\left[U_{t-1}^{(i)} - L_{t-1}^{(i)}\right]\right] \leq \sqrt{6(d+L)\log(n)\left(1+\log(n)\right)}\,\mathbb{E}\left[\sqrt{\sum_{t=1}^{n}\sum_{i=1}^{d+L}\mathbf{1}\left(E_t^{(i)}\right)}\right] + (d+L).$$

Moreover, we have

$$\mathbb{E}\left[\sqrt{\sum_{t=1}^{n}\sum_{i=1}^{d+L}\mathbf{1}\left(E_t^{(i)}\right)}\right] \leq \sqrt{\sum_{t=1}^{n}\mathbb{E}\left[\sum_{i=1}^{d+L}\mathbf{1}\left(E_t^{(i)}\right)\right]} \overset{(a)}{=} \sqrt{\sum_{t=1}^{n}\mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{d+L}\mathbf{1}\left(E_t^{(i)}\right)\middle|a_t\right]\right]}$$
$$\leq \sqrt{\sum_{t=1}^{n}\mathbb{E}\left[\max_a \mathbb{E}\left[\sum_{i=1}^{d+L}\mathbf{1}\left(E_t^{(i)}\right)\middle|a\right]\right]} \overset{(b)}{=} \sqrt{\sum_{t=1}^{n}\mathbb{E}\left[O_{\max}\right]} = \sqrt{nO_{\max}}, \tag{14}$$

where equality (a) follows from the tower property, and equality (b) follows from the definition of $O_{\max}$. Thus, we have

$$\sum_{i=1}^{d+L}\sum_{t=1}^{n}\mathbf{1}\left(E_t^{(i)}\right)\left[U_{t-1}^{(i)}-L_{t-1}^{(i)}\right]\leq\sqrt{6(d+L)O_{\max}n\log(n)\left(1+\log(n)\right)}+(d+L)$$

Putting everything together, we have

$$\begin{aligned}R_B(n)&\leq C\sqrt{6(d+L)O_{\max}n\log(n)\left(1+\log(n)\right)}+\left(C+\frac{2\pi^2}{3}B\right)(d+L)\\&=\mathcal{O}\left(C\sqrt{(d+L)O_{\max}n}\log(n)\right).\end{aligned}\tag{15}$$

q.e.d.

C. Pseudocode of `idTSinc`

The pseudocode of `idTSinc` is summarized in Algorithm 2.

Algorithm 2 `idTSinc`: A computationally efficient variant of `idTSvi`.

Input: $\epsilon > 0$
Randomly initialize q
for $t = 1, \ldots, n$ **do**
    Sample θ_t proportionally to $q(\theta_t)$
    Take action $a_t = \arg\max_{a\in\mathcal{A}^K} r(a, \theta_t)$
    Observes x_t and receive reward $r(x_t, z_t)$
    Randomly initialize q
    Calculate $\mathcal{L}(q)$ using (3) and set $\mathcal{L}'(q) = -\infty$
    while $\mathcal{L}(q) - \mathcal{L}'(q) \geq \epsilon$ **do**
        Set $\mathcal{L}'(q) = \mathcal{L}(q)$
        Update $q_t(z_t)$ using (4), for all z_t
        Update $q(\theta)$ using (5)
        Update $\mathcal{L}(q)$ using (3)
    end while
end for