A Proof for Theorems

We prove Theorem 2 before Theorem 1, since the former one includes more technical steps and main parts of the two proofs are similar.

A.1 Proof of Theorem 2 (C-TS)

Proof. By definition, $\mu_a := E\left[Y|a\right] = \sum_{i=1}^{k^n} E\left[Y|Pa_Y = Z_i\right] P\left(Pa_Y = Z_i|a\right)$, $a^* = \text{argmax}_a\, \mu_a$.

Define:

$$\begin{aligned} T_Z(t) &:= \sum_{s=1}^{t} \mathbb{1}_{\{Z_{(s)}=Z\}}, \\ \hat{\mu}_Z(t) &:= \frac{1}{T_Z(t)} \sum_{s=1}^{t} Y_s \mathbb{1}_{\{Z_{(s)}=Z\}}, \\ \mu_Z &:= E\left[Y|Pa_Y = Z\right], \end{aligned}$$

where $Z_{(s)}$ denotes the observed values of parent nodes for Y, in round s. Note that $\hat{\mu}_Z(t) = 0$ when $T_Z(t) = 0$.

Let E be the event that for all $t \in [T]$, $i \in [k^n]$ such that $\max_{a\in\mathcal{A}} P(Pa_Y = Z_i|a) > 0$, we have

$$|\hat{\mu}_{Z_i}(t-1) - \mu_{Z_i}| \leq \sqrt{\frac{2log(1/\delta)}{1 \vee T_{Z_i}(t-1)}}.$$

For fixed t and i, by Sub-Gaussian property, we can show

$$\begin{aligned} P\left(|\hat{\mu}_{Z_i}(t) - \mu_{Z_i}| \geq \sqrt{\frac{2\log(1/\delta)}{1 \vee T_{Z_i}(t)}}\right) &= \mathbb{E}\left[P\left(|\hat{\mu}_{Z_i}(t) - \mu_{Z_i}| \geq \sqrt{\frac{2\log(1/\delta)}{1 \vee T_{Z_i}(t)}}\middle| Z_{(1)}, \ldots, Z_{(t)}\right)\right] \\ &\leq \mathbb{E}\left[2\delta\right] = 2\delta. \end{aligned}$$

By union bound, we have $P\left(E^c\right) \leq 2\delta T k^n$.

The Bayesian regret can be written as

$$BR_T = \mathbb{E}\left[\sum_{t=1}^{T} \left(\mu_{a^*} - \mu_{a_t}\right)\right] = \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}\left[\mu_{a^*} - \mu_{a_t}|\mathcal{F}_{t-1}\right]\right],$$

where $\mathcal{F}_{t-1} = \sigma\left(a_1, Z_1, Y_1, \ldots, a_{t-1}, Z_{t-1}, Y_{t-1}\right)$.

The key insight is to notice that by definition of Thompson Sampling,

$$P\left(a^* = \cdot|\mathcal{F}_{t-1}\right) = P\left(a_t = \cdot|\mathcal{F}_{t-1}\right). \tag{1}$$

Further, define $\text{UCB}_a(t) := \sum_{j=1}^{k^n} \text{UCB}_{Z_j}(t) P(Pa_Y = Z_j|a)$, we can bound the conditional expected difference between optimal arm and the arm played at round t using equation 1 by

$$\begin{aligned} &\mathbb{E}\left[\mu_{a^*} - \mu_{a_t}|\mathcal{F}_{t-1}\right] \\ &= \mathbb{E}\left[\mu_{a^*} - \text{UCB}_{a_t}(t-1) + \text{UCB}_{a_t}(t-1) - \mu_{a_t}|\mathcal{F}_{t-1}\right] \\ &= \mathbb{E}\left[\mu_{a^*} - \text{UCB}_{a^*}(t-1) + \text{UCB}_{a_t}(t-1) - \mu_{a_t}|\mathcal{F}_{t-1}\right]. \end{aligned}$$

Next by tower rule, we have

$$BR_T = \mathbb{E}\left[\sum_{t=1}^{T} \left(\mu_{a^*} - \text{UCB}_{a^*}(t-1) + \text{UCB}_{a_t}(t-1) - \mu_{a_t}\right)\right].$$

On event E^c, by the original definition of BR_T we have $BR_T \leq 2T$. On event E, the first term is negative showing by the definition of $\text{UCB}_{\mathbf{Z}_j}, j = 1, \ldots, k^n$ and

$$\mu_{a^*} - \text{UCB}_{a^*}(t-1) = \sum_{j=1}^{k^n} \left(\mathbb{E}\left[Y|Pa_Y = Z_j\right] - \text{UCB}_{Z_j}(t-1) \right) P(Pa_Y = Z_j|a^*) \leq 0,$$

because $\mathbb{E}\left[Y|Pa_Y = Z_j\right] - \text{UCB}_{Z_j}(t-1) \leq 0$ on event E. Also on event E, the second term can be bounded by

$$\begin{aligned} \mathbb{1}_E \sum_{t=1}^{T} (\text{UCB}_{a_t}(t-1) - \mu_{a_t}) &= \mathbb{1}_E \sum_{t=1}^{T} \sum_{j=1}^{k^n} \left(\text{UCB}_{Z_j}(t-1) - \mathbb{E}\left[Y|Pa_Y = Z_j\right] \right) P(Pa_Y = Z_j|a_t) \\ &\leq \mathbb{1}_E \sum_{t=1}^{T} \sum_{j=1}^{k^n} \sqrt{\frac{8\log(1/\delta)}{1 \vee T_{Z_j}(t-1)}} P(Pa_Y = Z_j|a_t) \\ &\leq \mathbb{1}_E \sum_{t=1}^{T} \sum_{j=1}^{k^n} \sqrt{\frac{8\log(1/\delta)}{1 \vee T_{Z_j}(t-1)}} \left(P(Pa_Y = Z_j|a_t) - \mathbb{1}_{\{Z_{(t)}=Z_j\}} + \mathbb{1}_{\{Z_{(t)}=Z_j\}} \right). \end{aligned} \tag{2}$$

The second part of equation 2 can be bounded by

$$\begin{aligned} \mathbb{1}_E \sum_{t=1}^{T} \sum_{j=1}^{k^n} \sqrt{\frac{8\log(1/\delta)}{1 \vee T_{Z_j}(t-1)}} \mathbb{1}_{\{Z_{(t)}=Z_j\}} &\leq \mathbb{1}_E \sum_{j=1}^{k^n} \int_0^{T_{Z_j}(T)} \sqrt{\frac{8\log(1/\delta)}{s}} ds \\ &\leq \sum_{j=1}^{k^n} \sqrt{32 T_{Z_j}(T) \log(1/\delta)} \\ &\leq \sqrt{32 k^n T \log(1/\delta)}. \end{aligned}$$

For the first part of equation 2, we define $X_t := \sum_{s=1}^{t} \sum_{j=1}^{k^n} \sqrt{\frac{8\log(1/\delta)}{1 \vee T_{Z_j}(s-1)}} \left(P(Pa_Y = Z_j|a_s) - \mathbb{1}_{\{Z_{(s)}=Z_j\}} \right)$, $X_0 := 0$. Note that $\{X_t\}_{t=0}^T$ is a martingale sequence and we have

$$\begin{aligned} |X_t - X_{t-1}|^2 &= \left| \sum_{j=1}^{k^n} \sqrt{\frac{8\log(1/\delta)}{1 \vee T_{Z_j}(t-1)}} \left(P(Pa_Y = Z_j|a_t) - \mathbb{1}_{\{Z_{(t)}=Z_j\}} \right) \right|^2 \\ &\leq 32\log(1/\delta). \end{aligned}$$

By applying Azuma's inequality we have

$$P(|X_T| > \sqrt{k^n T \log(T)} \log(T)) \leq \exp\left(-\frac{k^n \log^3(T)}{32\log(1/\delta)} \right).$$

We take $\delta = 1/T^2$, combine the first and second part of equation 2, we show that with probability $1 - P(E^c) - \exp\left(-\frac{k^n \log^2(T)}{64}\right) = 1 - 2k^n/T - \exp\left(-\frac{k^n \log^2(T)}{64}\right)$,

$$R_T \leq 16\sqrt{k^n T \log(T)} \log(T).$$

Thus the Bayesian regret can be bounded by:

$$\begin{aligned} \mathbb{E}\left[R_T\right] &\leq P(E^c) \times 2T + \exp\left(-\frac{k^n \log^2(T)}{64} \right) \times 2T + \sqrt{64 k^n T \log(T)} \log(T) \\ &\leq C\sqrt{k^n T \log(T)} \log(T). \end{aligned}$$

where C is a constant and the above inequality holds for large T. Thus we have proved that $\mathbb{E}\left[R_T\right] = \tilde{O}\left(\sqrt{k^n T}\right)$. □

A.2 Proof of Theorem 1 (C-UCB)

Proof. Let E be the event that for all $t \in [T]$, $j \in [k^n]$, we have

$$\left|\hat{\mu}_{Z_j}(t-1) - \mathbb{E}\left[Y|Pa_Y = Z_j\right]\right| \leq \sqrt{\frac{2\log(1/\delta)}{1 \vee T_{Z_j}(t-1)}}.$$

Use same proof idea in Theorem 2, we have $P(E^c) \leq 2\delta T k^n$. Define $\text{UCB}_a(t) := \sum_{j=1}^{k^n} \text{UCB}_{Z_j}(t) P(Pa_Y = Z_j|a)$, the regret can be rewritten as

$$\begin{aligned} R_T &= \sum_{t=1}^{T} (\mu_{a^*} - \mu_{a_t}) \\ &= \sum_{t=1}^{T} \left(\mu_{a^*} - \text{UCB}_{a_t}(t-1) + \text{UCB}_{a_t}(t-1) - \mu_{a_t}\right). \end{aligned}$$

On event E^c, $R_T \leq 2T$. On event E we can show

$$\begin{aligned} \mu_{a^*} - \text{UCB}_{a_t}(t-1) &= \sum_{j=1}^{k^n} \mathbb{E}\left[Y|Pa_Y = Z_j\right] P(Pa_Y = Z_j|a^*) - \sum_{j=1}^{k^n} \text{UCB}_{Z_j}(t-1) P(Pa_Y = Z_j|a_t) \\ &\leq \sum_{j=1}^{k^n} \text{UCB}_{Z_j}(t-1) P(Pa_Y = Z_j|a^*) - \sum_{j=1}^{k^n} \text{UCB}_{Z_j}(t-1) P(Pa_Y = Z_j|a_t) \leq 0, \end{aligned}$$

where the last inequality follows by the way to choose a_t in Algorithm 1, the second last inequality follows by the definition of event E. Thus on event E we have

$$\begin{aligned} R_T &\leq \sum_{t=1}^{T} \left(\text{UCB}_{a_t}(t-1) - \mu_{a_t}\right) \\ &= \sum_{t=1}^{T} \sum_{j=1}^{k^n} \left(\text{UCB}_{Z_j}(t-1) - \mathbb{E}\left[Y|Pa_Y = Z_j\right]\right) P(Pa_Y = Z_j|a_t) \\ &\leq \sum_{t=1}^{T} \sum_{j=1}^{k^n} \sqrt{\frac{8\log(1/\delta)}{1 \vee T_{Z_j}(t-1)}} P(Pa_Y = Z_j|a_t) \\ &\leq \sum_{t=1}^{T} \sum_{j=1}^{k^n} \sqrt{\frac{8\log(1/\delta)}{1 \vee T_{Z_j}(t-1)}} \left(P(Pa_Y = Z_j|a_t) - \mathbb{1}_{\{Z_{(t)} = Z_j\}} + \mathbb{1}_{\{Z_{(t)} = Z_j\}}\right). \end{aligned} \tag{3}$$

The second part of Equation 3 can be bounded by

$$\begin{aligned} \sum_{t=1}^{T} \sum_{j=1}^{k^n} \sqrt{\frac{8\log(1/\delta)}{1 \vee T_{Z_j}(t-1)}} \mathbb{1}_{\{Z_{(t)} = Z_j\}} &\leq \sum_{j=1}^{k^n} \int_0^{T_{Z_j}(T)} \sqrt{\frac{8\log(1/\delta)}{s}} ds \\ &\leq \sum_{j=1}^{k^n} \sqrt{32 T_{Z_j}(T) \log(1/\delta)} \\ &\leq \sqrt{32 k^n T \log(1/\delta)}. \end{aligned}$$

For the first part of equation 3, we define $X_t := \sum_{s=1}^{t} \sum_{j=1}^{k^n} \sqrt{\frac{8\log(1/\delta)}{1 \vee T_{Z_j}(s-1)}} \left(P(Pa_Y = Z_j|a_s) - \mathbb{1}_{\{Z_{(s)} = Z_j\}}\right)$, $X_0 := 0$. Note that $\{X_t\}_{t=0}^{T}$ is a martingale sequence.

$$\begin{aligned} |X_t - X_{t-1}|^2 &= \left|\sum_{j=1}^{k^n} \sqrt{\frac{8\log(1/\delta)}{1 \vee T_{Z_j}(t-1)}} \left(P(Pa_Y = Z_j|a_t) - \mathbb{1}_{\{Z_{(t)} = Z_j\}}\right)\right|^2 \\ &\leq 32 \log(1/\delta). \end{aligned}$$

By applying Azuma's inequality we have

$$P(|X_T| > \sqrt{k^n T \log(T)} \log(T)) \leq \exp\left(-\frac{k^n \log^3(T)}{32\log(1/\delta)}\right).$$

We take $\delta = 1/T^2$, combine the first and second part of equation 3, with probability $1-P(E^c)-\exp\left(-\frac{k^n \log^2(T)}{64}\right) = 1 - 2k^n/T - \exp\left(-\frac{k^n \log^2(T)}{64}\right)$, the regret can be bounded by

$$R_T \leq 16\sqrt{k^n T \log(T)} \log(T).$$

Thus the expected regret can be bounded by:

$$\begin{aligned}\mathbb{E}\left[R_T\right] &\leq P(E^c) \times 2T + \exp\left(-\frac{k^n \log^2(T)}{64}\right) \times 2T + \sqrt{64 k^n T \log(T)} \log(T) \\ &\leq C\sqrt{k^n T \log(T)} \log(T)\end{aligned}$$

where C is a constant, above inequality holds for large T. Thus we prove $\mathbb{E}\left[R_T\right] = \tilde{O}\left(\sqrt{k^n T}\right)$ □

A.3 Proof of Theorem 3 (CL-TS)

Lemma 1. *(Lattimore and Szepesvári, 2020) Notations same as algorithm 4 and algorithm 5. Let $\delta \in (0,1)$. Then with probability at least $1-\delta$ it holds that for all $t \in \mathbb{N}$,*

$$\left\|\hat{\theta}_t - \theta\right\|_{V_t(\lambda)} \leq \sqrt{\lambda}\left\|\theta\right\|_2 + \sqrt{2\log\left(\frac{1}{\delta}\right) + \log\left(\frac{\det V_t(\lambda)}{\lambda^d}\right)}.$$

Furthermore, if $\|\theta^*\| \leq m_2$, then $P(\exists t \in \mathbb{N}^+ : \theta^* \notin \mathcal{C}_t) \leq \delta$ with

$$\mathcal{C}_t = \left\{\theta \in \mathbb{R}^d : \left\|\hat{\theta}_{t-1} - \theta\right\|_{V_{t-1}(\lambda)} \leq m_2\sqrt{\lambda} + \sqrt{2\log\left(\frac{1}{\delta}\right) + \log\left(\frac{\det V_{t-1}(\lambda)}{\lambda^d}\right)}\right\}.$$

Lemma 2. *(Lattimore and Szepesvári, 2020) Let $x_1, \ldots, x_T \in \mathbb{R}^d$ be a sequence of vectors with $\|x_t\|_2 \leq L < \infty$ for all $t \in [T]$, then*

$$\sum_{t=1}^{T}\left(1 \wedge \|x_t\|^2_{V_{t-1}^{-1}}\right) \leq 2\log\left(\det V_T\right) \leq 2d\log\left(1 + \frac{TL^2}{d}\right),$$

where $V_t = I_d + \sum_{s=1}^{t} x_s x_s^T$.

Proof. W define $\beta = 1 + \sqrt{2\log(T) + d\log\left(1+\frac{T}{d}\right)}$ and $V_t = I_d + \sum_{s=1}^{t} m_{a_s} m_{a_s}^T$ same as Algorithm 5, where $m_a := \sum_{i=1}^{k^n} f(Z_i)P(Pa_Y = Z_i|a)$. Define upper confidence bound $\text{UCB}_t : \mathcal{A} \to \mathbb{R}$ by

$$\text{UCB}_t(a) = \max_{\theta\in\mathcal{C}_t}\langle\theta, m_a\rangle =< \hat{\theta}_{t-1}, m_a > +\beta\left\|m_a\right\|_{V_{t-1}^{-1}},$$

where $\mathcal{C}_t = \left\{\theta \in \mathbb{R}^d : \left\|\theta - \hat{\theta}_{t-1}\right\|_{V_{t-1}} \leq \beta\right\}$. By Lemma 1, we have

$$P\left(\exists t \leq T : \left\|\hat{\theta}_{t-1} - \theta\right\|_{V_{t-1}} \geq 1 + \sqrt{2\log(T) + \log\left(\det V_t\right)}\right) \leq \frac{1}{T}.$$

And note $\|m_a\|_2 \leq 1$, thus by geometric means inequality we have

$$\det V_t \leq \left(trace(\frac{V_t}{d})\right)^d \leq \left(1 + \frac{T}{d}\right)^d.$$

Thus, by $\|\theta\|_2 \le 1$,

$$P\left(\exists t \le T : \left\|\hat{\theta}_{t-1} - \theta\right\|_{V_{t-1}} \ge 1 + \sqrt{2\log(T) + d\log\left(1 + \frac{T}{d}\right)}\right) \le \frac{1}{T}.$$

Let E_t be the event that $\left\|\hat{\theta}_{t-1} - \theta\right\|_{V_{t-1}} \le \beta$, $E := \cap_{t=1}^{T} E_t$, $a^* := \mathrm{argmax}_a \sum_{i=1}^{k^n} \langle f(Z_i), \theta \rangle P(Pa_Y = Z_i|a)$, which is a random variable in this setting because θ is random. Then

$$\begin{aligned}
BR_T &= \mathbb{E}\left[\sum_{t=1}^{T}\left\langle \sum_{i=1}^{k^n} f(Z_i)\left(P\left(Pa_Y = Z_i|a^*\right) - P\left(Pa_Y = Z_i|a_t\right)\right), \theta\right\rangle\right] \\
&= \mathbb{E}\left[\mathbb{1}_{E^c}\sum_{t=1}^{T}\left\langle \sum_{i=1}^{k^n} f(Z_i)\left(P(Pa_Y = Z_i|a^*) - P(Pa_Y = Z_i|a_t)\right), \theta\right\rangle\right] \\
&+ \mathbb{E}\left[\mathbb{1}_{E}\sum_{t=1}^{T}\left\langle \sum_{i=1}^{k^n} f(Z_i)\left(P(Pa_Y = Z_i|a^*) - P(Pa_Y = Z_i|a_t)\right), \theta\right\rangle\right] \\
&\le 2TP(E^c) + \mathbb{E}\left[\mathbb{1}_{E}\sum_{t=1}^{T}\left\langle \sum_{i=1}^{k^n} f(Z_i)\left(P(Pa_Y = Z_i|a^*) - P(Pa_Y = Z_i|a_t)\right), \theta\right\rangle\right] \\
&\le 2 + \mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}_{E_t}\left\langle \sum_{i=1}^{k^n} f(Z_i)\left(P(Pa_Y = Z_i|a^*) - P(Pa_Y = Z_i|a_t)\right), \theta\right\rangle\right]. \qquad (4)
\end{aligned}$$

Again, we know from equation 1 such that $P(a^* = \cdot|\mathcal{F}_{t-1}) = P(a_t = \cdot|\mathcal{F}_{t-1})$, where $\mathcal{F}_{t-1} = \sigma(Z_1, a_1, Y_1, \ldots, Z_{t-1}, a_{t-1}, Y_{t-1})$. Thus we have

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{1}_{E_t}\left\langle \sum_{i=1}^{k^n} f(Z_i)\left(P\left(Pa_Y = Z_i|a^*\right) - P\left(Pa_Y = Z_i|a_t\right)\right), \theta\right\rangle \middle| \mathcal{F}_{t-1}\right] \\
=&\mathbb{1}_{E_t}\mathbb{E}\left[\left\langle \sum_{i=1}^{k^n} f(Z_i)\left(P\left(Pa_Y = Z_i|a^*\right) - P\left(Pa_Y = Z_i|a_t\right)\right), \theta\right\rangle \middle| \mathcal{F}_{t-1}\right] \\
=&\mathbb{1}_{E_t}\mathbb{E}\left[\left\langle \sum_{i=1}^{k^n} f(Z_i)P\left(Pa_Y = Z_i|a^*\right), \theta\right\rangle - UCB_t(a^*) + UCB_t(a_t) - \left\langle \sum_{i=1}^{k^n} f(Z_i)P(Pa_Y = Z_i|a_t), \theta\right\rangle \middle| \mathcal{F}_{t-1}\right] \\
\le&\mathbb{1}_{E_t}\mathbb{E}\left[UCB_t(a_t) - \left\langle \sum_{i=1}^{k^n} f(Z_i)P(Pa_Y = Z_i|a_t), \theta\right\rangle \middle| \mathcal{F}_{t-1}\right] \\
\le&\mathbb{1}_{E_t}\mathbb{E}\left[\left\langle \sum_{i=1}^{k^n} f(Z_i)P(Pa_Y = Z_i|a_t), \hat{\theta}_{t-1} - \theta\right\rangle \middle| \mathcal{F}_{t-1}\right] + \beta\left\|\sum_{i=1}^{k^n} f(Z_i)P(Pa_Y = Z_i|a)\right\|_{V_{t-1}^{-1}} \\
\le&2\beta\left\|\sum_{i=1}^{k^n} f(Z_i)P(Pa_Y = Z_i|a)\right\|_{V_{t-1}^{-1}}.
\end{aligned}$$

Substituting into the second term of equation 4,

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}_{E_t}\left\langle\sum_{i=1}^{k^n} f(Z_i)\left(P(Pa_Y = Z_i|a^*) - P(Pa_Y = Z_i|a_t)\right), \theta\right\rangle\right] \\
\leq& 2\mathbb{E}\left[\beta\sum_{t=1}^{T}\left(1 \wedge \left\|\sum_{i=1}^{k^n} f(Z_i)P(Pa_Y = Z_i|a)\right\|_{V_{t-1}^{-1}}\right)\right] \\
\leq& 2\sqrt{T\mathbb{E}\left[\beta^2\sum_{t=1}^{T}\left(1 \wedge \left\|\sum_{i=1}^{k^n} f(Z_i)P(Pa_Y = Z_i|a)\right\|_{V_{t-1}^{-1}}^2\right)\right]} \text{ (By Cauchy-Schwartz)} \\
\leq& 2\sqrt{2dT\beta^2\log\left(1+\frac{T}{d}\right)} \text{ (By Lemma 2).}
\end{aligned}
$$

Putting together we prove

$$
BR_T \leq 2 + 2\sqrt{2dT\beta^2\log\left(1+\frac{T}{d}\right)} = \tilde{O}\left(d\sqrt{T}\right). \tag{5}
$$

□

A.4 Proof of Theorem 3 (CL-UCB)

Proof. Define $\beta = 1 + \sqrt{2\log(T) + d\log\left(1+\frac{T}{d}\right)}$, by Lemma 1 and above proof for CL-TS we have

$$
P(\exists t \leq T : \left\|\hat{\theta}_{t-1} - \theta^*\right\|_{V_{t-1}} \geq \beta) \leq \frac{1}{T},
$$
$$
P(\exists t \in \mathbb{N}^+ : \theta^* \notin \mathcal{C}_t) \leq \frac{1}{T},
$$

where $\mathcal{C}_t = \left\{\theta \in \mathbb{R}^d : \left\|\theta - \hat{\theta}_{t-1}\right\|_{V_{t-1}} \leq \beta\right\}$.

Let $\tilde{\theta}_t$ denote a θ that satisfies $\langle\tilde{\theta}_t, a_t\rangle = UCB_t(a_t)$. Again let E_t be the event that $\left\|\hat{\theta}_{t-1} - \theta^*\right\|_{V_{t-1}} \leq \beta$, let $E = \bigcap E_t$, $a^* = \operatorname{argmax}_a \sum_{j=1}^{k^n}\langle f(Z_j), \theta\rangle P(Pa_Y = Z_j|a)$. Then on event E_t, using the fact that $\theta^* \in \mathcal{C}_t$ we have

$$
\langle\theta^*, \sum_{j=1}^{k^n} f(Z_j)P(Pa_Y = Z_j|a^*)\rangle \leq UCB_t(a^*) \leq UCB_t(a_t) = \langle\tilde{\theta}_t, \sum_{j=1}^{k^n} f(Z_j)P(Pa_Y = Z_j|a_t)\rangle
$$

Thus we can bound the difference of expected reward between optimal arm and a_t by

$$
\begin{aligned}
\mu_{a^*} - \mu_{a_t} &= \langle\theta^*, \sum_{j=1}^{k^n} f(Z_j)P(Pa_Y = Z_j|a^*)\rangle - \langle\theta^*, \sum_{j=1}^{k^n} f(Z_j)P(Pa_Y = Z_j|a_t)\rangle \\
&\leq \langle\tilde{\theta}_t - \theta^*, \sum_{j=1}^{k^n} f(Z_j)P(Pa_Y = Z_j|a_t)\rangle \\
&\leq 2 \wedge 2\beta\left\|\sum_{j=1}^{k^n} f(Z_j)P(Pa_Y = Z_j|a_t)\right\|_{V_{t-1}^{-1}} \\
&\leq 2\beta\left(1 \wedge \left\|\sum_{j=1}^{k^n} f(Z_j)P(Pa_Y = Z_j|a_t)\right\|_{V_{t-1}^{-1}}\right).
\end{aligned}
$$

So the expected regret can be further bounded by:

$$
\begin{aligned}
\mathbb{E}\left[R_T\right] &= \mathbb{E}\left[\sum_{t=1}^{T}(\mu_{a^*}-\mu_{a_t})\right] = \mathbb{E}\left[\mathbb{1}_E\sum_{t=1}^{T}(\mu_{a^*}-\mu_{a_t})\right] + \mathbb{E}\left[\mathbb{1}_{E^c}\sum_{t=1}^{T}(\mu_{a^*}-\mu_{a_t})\right] \\
&\leq \mathbb{E}\left[\sum_{t=1}^{T}(\mu_{a^*}-\mu_{a_t})\mathbb{1}_{E_t}\right] + \mathbb{E}\left[\mathbb{1}_{E^c}\sum_{t=1}^{T}(\mu_{a^*}-\mu_{a_t})\right] \\
&\leq 2\beta\sum_{t=1}^{T}\left(1\wedge\left\|\sum_{j=1}^{k^n} f(Z_j)P(Pa_Y=Z_j|a_t)\right\|_{V_{t-1}^{-1}}\right) + 2TP(E^c) \\
&\leq 2+2\beta\sqrt{T\sum_{t=1}^{T}\left(1\wedge\left\|\sum_{j=1}^{k^n} f(Z_j)P(Pa_Y=Z_j|a_t)\right\|^2_{V_{t-1}^{-1}}\right)} \text{ (By Cauchy-Schwartz)} \\
&\leq 2+2\beta\sqrt{2dT\log\left(1+\frac{T}{d}\right)} \text{ (By Lemma 2)}
\end{aligned}
$$

□

A.5 Proof of Claim 1

Proof. Denote the reward variable for action a by $Y|_a$ and denote the reward variable given fixed parent values by $Y|_{\text{Pa}_Y=\mathbf{Z}}$. According to the causal information, $Y|_a$ can be represented as a weighted sum of $Y|_{\text{Pa}_Y=\mathbf{Z}}$:

$$Y|_a = \sum_{\mathbf{Z}} P(\text{Pa}_Y = \mathbf{Z}|a)Y|_{\text{Pa}_Y=\mathbf{Z}}. \tag{6}$$

In the statement of claim 1 we know that $Y|_{\text{Pa}_Y=\mathbf{Z}}$ are independent Gaussian distributions, therefore $Y|_a$, a weighted sum of Gaussian distributions still follows a Gaussian distribution. It remains to show the variance of $Y|_a$ is less than 1.

$$
\begin{aligned}
\text{Var}(Y|_a) &= \sum_{\mathbf{Z}} P(\text{Pa}_Y = \mathbf{Z}|a)^2\text{Var}(Y|_{\text{Pa}_Y=\mathbf{Z}}) & (7)\\
&\leq \sum_{\mathbf{Z}} P(\text{Pa}_Y = \mathbf{Z}|a)^2 \leq \sum_{\mathbf{Z}} P(\text{Pa}_Y = \mathbf{Z}|a) = 1, & (8)
\end{aligned}
$$

where the first inequality above uses the condition that $\text{Var}(Y|_{\text{Pa}_Y=\mathbf{Z}}) \leq 1$. We show that the reward for every arm $Y|_a$ is Gaussian distributed with variance less than 1, thus the bandit environment ν' described in the claim is an instance in Gaussian bandit environment class. □

A.6 Proof of Theorem 4

We first introduce an important concept.

Definition 2 (p-order Policy)**.** *For K-arm unstructured Gaussian bandit environments $\mathcal{E} := \mathcal{E}_K(\mathcal{N})$ and policy π, whose regret, on any $\nu \in \mathcal{E}$, is bounded by CT^p for some $C > 0$ and $p > 0$. We call this policy class $\Pi(\mathcal{E}, C, T, p)$, the class of p-order policies.*

Note that UCB and TS are in this class with $C = C'_\epsilon\sqrt{K}$ and $p = 1/2 + \epsilon$ with some $C'_\epsilon > 0$ for arbitrary small ϵ.

We use the following result to prove our theorem.

Theorem 5 (Finite-time, instance-dependent regret lower bound for p-order policies, Theorem 16.4 in Lattimore and Szepesvári (2020))**.** *Let $\nu \in \mathcal{E}_K(\mathcal{N})$ be a K-arm Gaussian bandit with mean vector $\mu \in \mathbb{R}^K$ and suboptimality gaps $\Delta \in [0, \infty)^K$. Let*

$$\mathcal{E}(\nu) = \{\nu' \in \mathcal{E}_K(\mathcal{N}) : \mu_i(\nu') \in [\mu_i, \mu_i + 2\Delta_i]\}.$$

Suppose π is a p-order policy such that $\exists C > 0$ and $p \in (0,1)$, $R_T(\pi, \nu') \leq CT^p$ for all T and $\nu' \in \mathcal{E}(\nu)$. Then for any $\epsilon \in (0,1]$,

$$\mathbb{E}R_T(\pi,\nu) \geq \frac{2}{(1+\epsilon)^2} \sum_{i:\Delta_i>0} \left(\frac{(1-p)\log(T) + \log(\frac{\epsilon\Delta_i}{8C})}{\Delta_i} \right)^+,$$

where $(x)^+ = \max(x,0)$ is the positive part of $x \in \mathbb{R}$.

Proof of Theorem 4. Consider the bandit environment ν described in section 4. By claim 1 we know ν is an instance in unstructured Gaussian bandit environment class, so we can further apply Theorem 5. The size of three types of actions are all $3^N/3$. For Type 1 actions, its gap compared to the optimal actions is Δ, for Type 0 actions, gap is $p_1\Delta$. Plugging into the results of Theorem 5, for every p-order policy over $\mathcal{E}(\nu)$, we have

$$\mathbb{E}R_T(\pi,\nu) \geq \frac{1}{2}\frac{3^N}{3} \left(\frac{(1-p)\log(T) + \log(\frac{\Delta}{8C})}{\Delta} \right)^+ + \frac{1}{2}\frac{3^N}{3} \left(\frac{(1-p)\log(T) + \log(\frac{p_1\Delta}{8C})}{p_1\Delta} \right)^+. \tag{9}$$

In particular, choose $\Delta = 8\rho CT^{p-1}$, we get

$$(1-p)\log(T) + \log(\frac{\Delta}{8C}) = \log(\rho),$$
$$(1-p)\log(T) + \log(\frac{p_1\Delta}{8C}) = \log(p_1\rho).$$

Note that $\sup_{\rho>0} \log(\rho)/\rho = \exp(-1) \approx 0.35$, and we next plug above two equations in Equation 9 to get

$$\mathbb{E}R_T(\pi,\nu) \geq \frac{3^N}{3}\frac{0.35}{8CT^{p-1}}.$$

Now consider π to be UCB, by plugging in $C = C'_\epsilon\sqrt{3^N}$ and $p = 1/2 + \epsilon$ we have

$$\mathbb{E}R_T(UCB,\nu) \geq \frac{0.35}{24C'_\epsilon}\sqrt{3^N}T^{1/2-\epsilon}.$$

□

B Probability Tables Used in Experiments

i	1	2	3
$P(X_1 = i)$	0.3	0.4	0.3
$P(X_2 = i)$	0.3	0.3	0.4
$P(X_3 = i)$	0.5	0.3	0.2
$P(X_4 = i)$	0.25	0.25	0.5
$P(W_1 = 1 \mid X_1 = i)$	0.2	0.5	0.8
$P(W_2 = 1 \mid X_2 = i)$	0.3	0.2	0.8
$P(W_3 = 1 \mid X_3 = i)$	0.4	0.6	0.5
$P(W_4 = 1 \mid X_4 = i)$	0.3	0.5	0.6

Table 1: Marginal and conditional probabilities for pure simulation experiment in section 5.1.1, numbers are randomly selected.

i	1	2	3	4
$P(X_1 = i)$	0.2	0.2	0.6	
$P(X_2 = i)$	0.05	0.6	0.3	0.05
$P(Z_3 = i)$	0.5	0.2	0.3	
$P(Z_1 = 1 \| X_2 = i)$	0.7	0.7	0.3	0.3
$P(Z_2 = 1 \| X_1 = 3, X_2 = i)$	0.6	0.7	0.6	0.5
$P(Z_2 = 1 \| X_1 \neq 3, X_2 = i)$	0.8	0.9	0.5	0.2

Table 2: Marginal and conditional probabilities for email campaign causal graph.