Appendix A. Proofs

A.1. Proof of theorem 1

Let a_t denotes the arm with the highest index at time t, i.e. $a_t = \arg\max_a \mathfrak{I}_{a,t}^{\texttt{Bayesian-CPD-TS}}$. First note that at each time t, if the arm a is played, then the Bayesian-CPD-TS algorithm is either sampling a random arm or playing the arm with the highest index. So the probability that arm a is chosen at time t when a is not the optimal arm is written as:

$$\mathbb{P}\left(A_t = a \neq a_t^\star\right) \leqslant \frac{\alpha}{A} + (1-\alpha)\mathbb{P}\left(a_t = a \neq a_t^\star\right)$$

Using the definition of $\overline{N}_{a,T}$, we have:

$$\begin{aligned}
\mathbb{E}\left[\overline{N}_{a,T}\right] &\leqslant \sum_{t=1}^{T} \mathbb{P}\left(A_t = a \neq a_t^\star\right) \\
&\leqslant \sum_{t=1}^{T} \left(\frac{\alpha}{A} + (1-\alpha)\mathbb{P}\left(a_t = a \neq a_t^\star\right)\right) \\
&\leqslant \frac{\alpha}{A}T + \underbrace{\sum_{t=1}^{T} \mathbb{P}\left(a_t = a \neq a_t^\star\right)}_{(a)}
\end{aligned}$$

Now, we need to upper bound the term (a). For this purpose, let us consider an experiment of the Bayesian-CPD-TS over T plays. Let F_a denote the number of false alarms up to time T and $D_{a,k}$ denote the detection delay of k-th change-point on arm a, where $a \leqslant \mathrm{NC}_{a,T}$. By the way, the total number of detection points, when the change detection algorithm RBOCPD signals an alarm on arm a is upper bounded by $\mathrm{NC}_{a,T} + F_a$. Recall that $\tau_a(t)$ is the latest detection time (which include also false alarms). For each arm a, we define $\mathcal{T}_a$ as the set of times slots that no change-point occurs i.e.

$$\mathcal{T}_a = \{t \in [1,T] : \mu_{a,s} = \mu_{a,t} \text{ and } \tau_a(t)+1 \leqslant s \leqslant t, t \geqslant \tau_a(t)+1\}$$

Following this, we have:

$$(a) \leqslant \mathbb{E}\left[\sum_{k=1}^{\mathrm{NC}_{a,T}} D_{a,k} + \sum_{t\in\mathcal{T}_a} \mathbb{I}\left\{a_t = a \neq a_t^\star\right\}\right]$$

Note that during a stationary period, we can easily use the regret upper control of Thompson Sampling to control the quantity $\mathbb{I}\left\{a_t = a \neq a_t^\star\right\}$. Thus, following analysis in Kaufmann et al. (2012b), we have (in the case where $\mathcal{T}_a$ is a deterministic set related to change-point k):

$$\forall \varepsilon \in (0,1), \exists C_{a,k} > 0 : \sum_{t \in \mathcal{T}_a} \mathbb{P}\left(a_t = a \neq a_t^\star\right) \leqslant (1+\varepsilon) \times \frac{\log |\mathcal{T}_a| + \log\log |\mathcal{T}_a|}{\boldsymbol{kl}\left(\theta_{a,[k]}, \theta^\star_{[k]}\right)} + C_{a,k}$$

where $|\mathcal{T}_a|$ denotes the length of the period $\mathcal{T}_a$.
Following this, since $|\mathcal{T}_a| \leqslant T$ we have naturally :

$$\forall \varepsilon \in (0,1), \exists C_{a,k} > 0 : \sum_{t \in \mathcal{T}_a} \mathbb{P}\left(a_t = a \neq a_t^\star\right) \leqslant (1+\varepsilon) \times \frac{\log T + \log\log T}{\boldsymbol{kl}\left(\theta_{a,[k]}, \theta^\star_{[k]}\right)} + C_{a,k}$$

And then,

$$\forall \varepsilon \in (0,1), \exists C_{a,k} > 0 : \sum_{t \in \mathcal{T}_a} \mathbb{P}\left(a_t = a \neq a_t^\star\right) \leqslant (1+\varepsilon) \times \frac{\log T + \log\log T}{\min\limits_{k \in [1, K_T], a \neq a_k^\star} \boldsymbol{kl}\left(\theta_{a,[k]}, \theta^\star_{[k]}\right)} + C_{a,k}$$

Finally, by applying the expectation operator, we get:

$$\begin{aligned}
\mathbb{E}\left[\overline{N}_{a,T}\right] &\leqslant \frac{\alpha}{A} T + \underbrace{\sum_{t=1}^{T} \mathbb{P}\left(a_t = a \neq a_t^\star\right)}_{(a)} \\
&\leqslant \frac{\alpha}{A} T + \sum_{k=1}^{\mathrm{NC}_{a,T}} \mathbb{E}\left[D_{a,k}\right] + \left(\mathrm{NC}_{a,T} + \mathbb{E}\left[F_T\right]\right) \times (1+\varepsilon) \times \frac{\log T + \log\log T}{\min\limits_{k \in [1, K_T], a \neq a_k^\star} \boldsymbol{kl}\left(\theta_{a,[k]}, \theta^\star_{[k]}\right)} + C
\end{aligned}$$

where $\mathbb{E}\left[F_T\right]$ denotes the expected number of false alarm raised up to horizon T and C a problem dependant constant depending on all $C_{a,k}$.

A.2. Proof of Theorem 2

Regarding the false alarm control, it comes directly from Theorem 1 in the analysis of the restarted Bayesian online changepoint detector in Alami et al. (2020).
Indeed, we have:

$$\begin{aligned}
\forall \delta' \in (0,1) : \mathbb{E}\left[F_T\right] &\leqslant \sum_{k=1}^{K_T} \mathbb{P}\left(\exists\ t \in [\tau_k + 1, \tau_{k+1} - 1) : \texttt{RBOCPD_Restart}(Y_{A_t,1}, ..., Y_{A_t, N_{A_t,t}}) = 1\right) \\
&\leqslant K_T \delta'.
\end{aligned}$$

Thus, by choosing $\delta' = \frac{\delta}{K_T}$, we upper bound $\mathbb{E}\left[F_T\right] \leqslant \delta$.
Then, the control of the detection delay comes also from theorem 2 in the analysis of the restarted Bayesian online change-point detector in Alami et al. (2020).

Indeed we upper bound the detection delay of change point $\tau_{a,k}$ related to arm a (with some $\delta' \in (0,1)$)

$$\mathbb{E}\left[D_{a,k}\right] = \min\left\{ d \in \mathbb{N}^{\star} : d > \frac{\left(1 - \frac{\mathcal{C}_{\tau_{a,k}, d+\tau_{a,k}-1,\delta}}{\Lambda_{a,[k]}}\right)^{-2}}{2\Lambda_{a,[k]}^2} \times \frac{-\log \eta_{\tau_{a,k}, d+\tau_{a,k}-1} + f_{\tau_{a,k}, d+\tau_{a,k}-1}}{1 + \frac{\log \eta_{\tau_{a,k}, d+\tau_{a,k}-1} - f_{1,\tau_{a,k}, d+\tau_{a,k}-1}}{2n_{1:\tau_{a,k}-1}\left(\Lambda_{a,[k]} - \mathcal{C}_{\tau_{a,k}, d+\tau_{a,k}-1,\delta}\right)^2}} \right\}, \tag{5}$$

where:

$$\mathcal{C}_{s,t,\delta} = \frac{\sqrt{2}}{2}\left(\sqrt{\frac{1 + \frac{1}{n_{1:s-1}}}{n_{1:s-1}} \log\left(\frac{2\sqrt{n_{1:s}}}{\delta'}\right)} + \sqrt{\frac{1 + \frac{1}{n_{s:t}}}{n_{s:t}} \log\left(\frac{2n_{1:t}\sqrt{n_{s:t}+1}\log^2\left(n_{1:t}\right)}{\log(2)\delta'}\right)}\right). \tag{6}$$

with $f_{s,t} = \log n_{1:s} + \log n_{s:t+1} - \frac{1}{2}\log n_{1:t} + \frac{9}{8}$ and the decreasing function $n_{i:j} = j - i + 1$ and $\eta \in (0,1)$.

Indeed assuming that we collect enough samples between two consecutive change-points, we upper bound the detection delay of change point $\tau_{a,k}$ related to arm a by its behavior in the asymptotic regime such that:

$$\mathbb{E}\left[D_{a,k}\right] = \mathcal{O}\left(\frac{o\left(\log \frac{1}{\delta'}\right)}{2\alpha \times \Lambda_{a,[k]}^2}\right) \leqslant \mathcal{O}\left(\frac{o\left(\log \frac{1}{\delta'}\right)}{2\alpha \times \min\limits_{a:\Lambda_{a,[k]}\neq 0)} \Lambda_{a,[k]}^2}\right)$$

Finally, by choosing $\delta' = \frac{\delta}{K_T}$ we get the result of Theorem 2.

A.3. Proof of Corollary 1

The result of corollary 1 comes directly by injecting the result of Theorem 2 into Theorem 1 after summing over all the arms.