A. Additional details about the original formulation of the Bayesian Online Change Point Detection (Adams & MacKay, 2007)

**Notion of runlength.** In order to deal with the non-stationary behavior of the environment, the notion of runlength has been introduced by (Adams & MacKay, 2007). It represents the overall number of time steps since the last change-point. We denote the length of the current run at time $t \geq 1$ by $R_t$. Since $R_t$ is unknown, we can consider the runlength as a random variable taking values $r_t \in \mathcal{R}_t = [0, t - 1]$. Thereby, let $p(r_t | x_{1:t}) = \mathbb{P}\{R_t = r_t | X_{1:t} = x_{1:t}\}$ denotes the distribution of $R_t$ given the sequence of observations $x_{1:t}$. ($p(r_t | x_{1:t})$ is a short hand notation).

**Computation of $p(r_t | x_{1:t})$ based on a message passing algorithm.** (Adams & MacKay, 2007) have proposed an online recursive runlength estimation in order to calculate the runlength distribution $p(r_t | x_{1:t})$. More specifically to find:

$$
\mathbb{P}\{R_t = r_t | X_{1:t} = x_{1:t}\} = \frac{\mathbb{P}\{R_t = r_t, X_{1:t} = x_{1:t}\}}{\mathbb{P}\{X_{1:t} = x_{1:t}\}}. \tag{10}
$$

We seek the joint distribution over the past estimated runlengths $R_{t-1}$ as follows:

$$
\mathbb{P}\{R_t = r_t, X_{1:t} = x_{1:t}\} = \sum_{r_{t-1} \in \mathcal{R}_{t-1}} \mathbb{P}\{R_t = r_t, X_{1:t} = x_{1:t}, R_{t-1} = r_{t-1}\} \tag{a}
$$

$$
= \sum_{r_{t-1} \in \mathcal{R}_{t-1}} \mathbb{P}\{R_t = r_t, X_{t} = x_t | R_{t-1} = r_{t-1}, X_{1:t-1} = x_{1:t-1}\} \mathbb{P}\{R_{t-1} = r_{t-1}, X_{1:t-1} = x_{1:t-1}\} \tag{b}
$$

$$
= \sum_{r_{t-1} \in \mathcal{R}_{t-1}} \mathbb{P}\{X_t = x_t | R_{t-1} = r_{t-1}, X_{1:t-1} = x_{1:t-1}\} \mathbb{P}\{R_{t-1} = r_{t-1}, R_{t-1} = r_{t-1}, X_{1:t-1} = x_{1:t-1}\} \tag{c}
$$

$$
= \sum_{r_{t-1} \in \mathcal{R}_{t-1}} \mathbb{P}\{R_t = r_t | R_{t-1} = r_{t-1}\} \mathbb{P}\{X_t = x_t | R_{t-1} = r_{t-1}, X_{1:t-1} = x_{1:t-1}\} \mathbb{P}\{R_{t-1} = r_{t-1}, X_{1:t-1} = x_{1:t-1}\} \tag{d}
$$

where (a) holds true using a marginalization, (b) and (c) hold true using two chain rules, (d) holds true thanks to the fact that $R_t$ do not depend on $X_{1:t-1}$ and $X_t$ do not depend on $R_{t-1}$.

Thus, combining Equation (10) and Equation (11) we get (using the short-hand notations):

$$
p(r_t | x_{1:t}) \propto \sum_{r_{t-1} \in \mathcal{R}_{t-1}} p(r_t | r_{t-1}) p(x_t | r_{t-1}, x_{1:t-1}) p(r_{t-1} | x_{1:t-1}). \tag{12}
$$

So, given the previous runlength distribution $p(r_{t-1} | x_{1:t-1})$, one can thus build a message-passing algorithm for the current run-length distribution $p(r_t | x_{1:t})$ by calculating:

1. the underlying predictive model (UPM) $p(x_t | r_{t-1}, x_{1:t-1})$,
2. the hazard function $p(r_t | r_{t-1})$.

It should be noted that at each time $t$, the runlength $R_t$ either continues to grow (which corresponds to the event $\{R_t = R_{t-1} + 1\}$) or a change occurs which corresponds to $\{R_t = 0\}$. Thus, from equation (12), we get the following recursive runlength distribution estimation:

- **Growth probability:**

$$
p(r_t = r_{t-1} + 1 | x_{1:t}) \propto p(r_t | r_{t-1}) p(x_t | r_{t-1}, x_{1:t-1}) p(r_{t-1} | x_{1:t-1}). \tag{13}
$$
• Change-point probability:

\[ p(r_t = 0 | x_{1:t}) \propto \sum_{r_{t-1} \in \mathcal{R}_{t-1}} p(r_t | r_{t-1}) \, p(x_t | r_{t-1}, x_{1:t-1}) \, p(r_{t-1} | x_{1:t-1}). \] (14)

**Hazard function.** According to Equation (13) and Equation (14), the runlength distribution estimation need to compute the change-point prior \( P(R_t | R_{t-1}) \), which is done following the Assumption that hazard function is a constant \( h \in (0, 1) \) in the sense that \( P(R_t | R_{t-1}) \) is independent of \( r_{t-1} \) and is constant, giving rise, a priori, to geometric inter-arrival times for change points.

\[
P\{ R_t | R_{t-1} \} = h \mathbb{I}\{ R_t = 0 \} + (1 - h) \mathbb{I}\{ R_t = R_{t-1} + 1 \}.
\] (15)

Then, injecting Equation (15) into Equation (13) and Equation (14) we get:

\[
p(r_t = r_{t-1} + 1 | x_{1:t}) \propto (1 - h) \, p(x_t | r_{t-1}, x_{1:t-1}) \times p(r_{t-1} | x_{1:t-1}),
\] (16)

\[
p(R_t = 0 | x_{1:t}) \propto h \sum_{r_{t-1} \in \mathcal{R}_{t-1}} p(x_t | r_{t-1}, x_{1:t-1}) \times p(r_{t-1} | x_{1:t-1}).
\] (17)

**B. Proofs of Lemmas**

**Notation 2 (Useful short-hand notations).** In the following, for some element \( x \in [0, 1] \), we denote by \( \bar{x} \) its complementary such that: \( \bar{x} = 1 - x \). Then, we denote by \( \Sigma_{s:t} \) and \( \bar{\Sigma}_{s:t} \) the two following cumulative sums:

\[
\Sigma_{s:t} = \sum_{s=s}^{t} x_s \quad \text{and} \quad \bar{\Sigma}_{s:t} = \sum_{s=s}^{t} \bar{x}_s.
\]

---

**Proof of Lemma 1:**

You only need to see that:

\[
V_t = \sum_{s=1}^{t} \nu_{s,t} = \sum_{s=1}^{t-1} \nu_{s,t} + \nu_{t,t} = (1 - h) \sum_{s=1}^{t-1} \exp(-l_{s,t}) \, \nu_{s,t-1} + h \sum_{s=1}^{t-1} \exp(-l_{s,t}) \, \nu_{s,t-1} = \sum_{s=1}^{t-1} \exp(-l_{s,t}) \, \nu_{s,t-1}.
\]
First, for all \( t \geq 2 \), we have:

\[
V_t = \sum_{i=1}^{t} v_{i,t}
\]

\[
V_t = v_{1,t} + \sum_{i=2}^{t-1} v_{i,t} + v_{t,t}
\]

\[
V_t = (1 - h)^{t-1} \exp \left( -\tilde{L}_{1:t} \right) V_1 + \sum_{i=2}^{t-1} (1 - h)^{t-i} \exp \left( -\tilde{L}_{i:t} \right) hV_i + hV_t.
\]

\[
\Leftrightarrow V_t = \sum_{i=1}^{t} (1 - h)^{t-i} \exp \left( -\tilde{L}_{i:t} \right) hL(i\neq 1) V_i \text{ with convention: } L_{i,j} = 0 \Leftrightarrow i > j.
\]

\[
\Leftrightarrow V_t = \sum_{i=1}^{t} \alpha_{t,i} V_i.
\]

\[
\Rightarrow (1 - \alpha_{t,t}) V_t = \sum_{i=1}^{t-1} \alpha_{t,i} V_i.
\]

Finally, by letting:

\[
\beta_{t,i} = \frac{\alpha_{t,i}}{1 - h},
\]

we obtain the following expression of \( V_t \) (using the classical induction procedure and using \( V_1 = 1 \)):

\[
\forall t \geq 4,
\]

\[
V_t = \left( \beta_{t,1} + \sum_{i_1=1}^{t-2} \beta_{t,t-i_1,1} \beta_{t-i_1,1} + \sum_{k=3}^{t-1} \sum_{i_1=1}^{t-k} \sum_{i_2=i_1+1}^{t-k} \cdots \sum_{i_{k-1}=i_k-2+1}^{t-k} \beta_{t,t-i_1,1} \beta_{t-i_1,1} \beta_{t-i_1,1} \cdots \beta_{t-i_{k-1},1,1} \right) V_1
\]

\[
= \beta_{t,1} + \sum_{i_1=1}^{t-2} \beta_{t,t-i_1,1} \beta_{t-i_1,1} + \sum_{k=3}^{t-1} \sum_{i_1=1}^{t-k} \sum_{i_2=i_1+1}^{t-k} \cdots \sum_{i_{k-1}=i_k-2+1}^{t-k} \beta_{t,t-i_1,1} \beta_{t-i_1,1} \beta_{t-i_1,1} \cdots \beta_{t-i_{k-1},1,1}.
\]

\[
V_3 = \beta_{3,1} + \beta_{3,2} \beta_{2,1}.
\]

\[
V_2 = \beta_{2,1}.
\]

which can be concatenated in the following form:

\[
V_t = (1 - h)^{t-2} \sum_{k=1}^{(t-k)} \left( \frac{h}{1 - h} \right)^{k-1} \tilde{V}_{t,k}, \text{ where:}
\]

\[
\tilde{V}_{t,k} = \sum_{i_1=1}^{t-k} \sum_{i_2=i_1+1}^{t-k} \cdots \sum_{i_{k-1}=i_k-2+1}^{t-k} \exp \left( -\tilde{L}_{1:i_1} \right) \times \prod_{j=1}^{k-2} \exp \left( -\tilde{L}_{i_j+1:i_{j+1}} \right) \exp \left( -\tilde{L}_{i_{k-1}+1:t-1} \right),
\]

and \((1 - h)^{t-2} \sum_{k=1}^{(t-2)} \left( \frac{h}{1 - h} \right)^{k-1} \binom{t-2}{k-1} = 1\).

\[
\square
\]

**Proof of Lemma 3:**
First, notice that the cumulative loss $\hat{L}_{s,t}$ can be written as follows:

$$\hat{L}_{s,t} = - \log \prod_{s' = s}^{t} \mathbb{P}(x_{s'}|x_{s:s-1})$$

Then, we only need to show by induction that:

$$\forall x_{1:n} \in \{0, 1\}^n \quad \prod_{s=1}^{n} \mathbb{P}(x_s|x_{1:s-1}) = \frac{1}{(n + 1) \left(\sum_{i=1}^{n} x_i\right)}.$$  (18)

**Step 1:** For $n = 1$, we have to deal with two cases, $x_1 = 1$ and $x_1 = 0$. Using the definition of the predictor $\mathbb{P}(\cdot)$, we obtain:

$$\begin{cases}
\mathbb{P}(1|\emptyset) = 1/2 = \frac{1}{(1+1)(1)} , \\
\mathbb{P}(0|\emptyset) = 1/2 = \frac{1}{(1+1)(0)} .
\end{cases}$$

**Step 2:** Assume that for some $x_{1:n} \in \{0, 1\}^n$, we have:

$$\prod_{s=1}^{n} \mathbb{P}(x_s|x_{1:s-1}) = \frac{1}{(n + 1) \left(\sum_{i=1}^{n} x_i\right)}.$$  (18)

Then, let us verify that:

$$\forall x_{n+1} \in \{0, 1\} \quad \prod_{s=1}^{n+1} \mathbb{P}(x_s|x_{1:s-1}) = \frac{1}{(n + 2) \left(\sum_{i=1}^{n+1} x_i\right)} .$$

To this end, we need to deal with two cases, depending on the values taken by $x_{n+1}$.

**Case 1:** $x_{n+1} = 1$  Observe that:

$$\prod_{s=1}^{n+1} \mathbb{P}(x_s|x_{1:s-1}) = \prod_{s=1}^{n} \mathbb{P}(x_s|x_{1:s-1}) \hat{p}(1|x_{1:n}) .$$

Using the definition of the predictor and the assumption (18), we obtain:

$$\prod_{s=1}^{n+1} \mathbb{P}(x_s|x_{1:s-1}) = \frac{1}{(n + 1) \left(\sum_{i=1}^{n} x_i\right)} \times \frac{\sum_{i=1}^{n} x_i + 1}{n + 2}$$

$$= \frac{\left(\sum_{i=1}^{n} x_i + 1\right) \times \left(\sum_{i=1}^{n} x_i\right)! \times \left(\sum_{i=1}^{n} \bar{x}_i\right)!}{(n + 2) (n + 1) n!}$$

$$= \frac{\left(\sum_{i=1}^{n} x_i + 1\right)! \times \left(\sum_{i=1}^{n} \bar{x}_i + 0\right)!}{(n + 2) (n + 1)!}$$

$$= \frac{\left(\sum_{i=1}^{n+1} x_i\right)! \times \left(\sum_{i=1}^{n+1} \bar{x}_i\right)!}{(n + 2) (n + 1)!}$$

$$= \frac{1}{(n + 2) \left(\sum_{i=1}^{n+1} x_i\right)} .$$
Case 2: $x_{n+1} = 0$ Observe that:

$$\prod_{s=1}^{n+1} L_p(x_s | x_{1:s-1}) = \prod_{s=1}^{n} L_p(x_s | x_{1:s-1}) \tilde{p}(0 | x_{1:n}).$$

Using the definition of the predictor and the assumption (18), we obtain:

$$\prod_{s=1}^{n+1} L_p(x_s | x_{1:s-1}) = \frac{1}{(n+1) \left( \sum_{i=1}^{n} x_i \right)} \times \frac{\sum_{i=1}^{n} \bar{x}_i + 1}{n + 2} \times \frac{(\sum_{i=1}^{n} x_i)! \times (\sum_{i=1}^{n} \bar{x}_i)!}{(n + 2)(n + 1)!} \times \frac{1}{(n + 2) \left( \sum_{i=1}^{n+1} x_i \right)}.$$

Proof of Lemma 4:

The proof follows three main steps:

Step 1: Controlling the binomial $\binom{n}{k}$ Using the Stirling formula:

$$\forall n \geq 1, \quad \sqrt{2\pi n} \left( \frac{n}{e} \right)^n \leq n! \leq \sqrt{2\pi n} \left( \frac{n}{e} \right)^n \exp \left( \frac{1}{12} \right),$$

the control of the binomial $\binom{n}{k}$ takes the following form:

$$\forall n \geq 1, \forall k \in [0, n], \quad \frac{n^n}{k^k (n-k)^{n-k}} \frac{\exp(b_1)}{\sqrt{n}} \leq \binom{n}{k} \leq \frac{n^n}{k^k (n-k)^{n-k}} \text{ with } b_1 = -\frac{1}{6} - \frac{1}{2} \log(2\pi). \tag{19}$$

Step 2: First bounds for the cumulative loss $\hat{L}_{s:t}$ Following Lemma 3, we can rewrite the cumulative loss $\hat{L}_{s:t}$ as follows:

$$\hat{L}_{s:t} = \log(n_{s:t} + 1) + \log \left( \sum_{s:t} \right).$$

Then by letting $\Phi(x) = x \log x$ and by following Equation (19), we obtain the following two bounds:

$$\begin{align*}
\hat{L}_{s:t} &\leq \log(n_{s:t} + 1) + \Phi(n_{s:t}) - \Phi(\sum_{s:t}) - \Phi(\sum_{s:t}) - \frac{9}{8} - \frac{1}{2} \log n_{s:t}, \\
\hat{L}_{s:t} &\geq \log(n_{s:t} + 1) + \Phi(n_{s:t}) - \Phi(\sum_{s:t}) - \frac{9}{8} - \frac{1}{2} \log n_{s:t}. \tag{20}
\end{align*}$$

Step 3: Controlling the cumulative loss First, notice that:

$$\sum_{s:t} \log \sum_{s:t} + \sum_{s:t} \log \sum_{s:t} = \sum_{s:t} \log \theta + \sum_{s:t} \log \theta + n_{s:t} \log n_{s:t} + n_{s:t} k l \left( \frac{\sum_{s:t}}{n_{s:t}}, \theta \right). \tag{21}$$

Then, using Equations (20) with Equation (21), we obtain:
for the upper bound of the loss $\widehat{L}_{s,t}$

$$
\widehat{L}_{s,t} \leq \log (n_{s,t} + 1) - \sum_{s,t} \log \frac{\sum_{s,t}}{n_{s,t}} - \sum_{s,t} \log \frac{\sum_{s,t}}{n_{s,t}} \\
\leq \log (n_{s,t} + 1) - \sum_{s,t} \log \frac{\sum_{s,t}}{n_{s,t}} - \sum_{s,t} \log \left( \frac{\sum_{s,t} + n_{s,t}}{n_{s,t}} \right) \\
\leq \log (n_{s,t} + 1) - \sum_{s,t} \log \theta - \sum_{s,t} \log \bar{\theta} - n_{s,t} \text{KL} \left( \frac{\sum_{s,t}}{n_{s,t}}, \theta \right) \\
\leq \log (n_{s,t} + 1) - \sum_{s,t} \log \theta - \sum_{s,t} \log \bar{\theta} ,
$$

where (a) holds by Equation (21) and (b) holds using the positiveness of the Kullback-Leibler divergence ($\text{KL}(\cdot, \cdot) \geq 0$),

for the lower bound of the loss $\widehat{L}_{s,t}$

$$
\widehat{L}_{s,t} \geq \log (n_{s,t} + 1) - \frac{1}{2} \log n_{s,t} - \sum_{s,t} \log \theta - \sum_{s,t} \log \bar{\theta} - n_{s,t} \text{KL} \left( \frac{\sum_{s,t}}{n_{s,t}}, \theta \right) + b_1. \\
\geq \log (n_{s,t} + 1) - \frac{1}{2} \log n_{s,t} - \sum_{s,t} \log \theta - \sum_{s,t} \log \bar{\theta} - n_{s,t} \text{KL} (\mu_{s,t}, \theta) - \frac{9}{8}.
$$

Proof of Lemma 5, Lemma 6 and Lemma 7:

The interested reader can refer for more details on the proofs of Lemma 5, Lemma 6 and Lemma 7 to the manuscript titled “Mathematics of Statistical Sequential Decision Making” https://pdfs.semanticscholar.org/9099/c0ff71185adce7705beb78d595abc817c33d6.pdf

C. Proofs of Theorems

Proof of Theorem 2:

Assume that: $\forall t \in [t_0, \tau_c) \ x_{t_0:t} \sim B(\theta)$. The proof follows four main steps:

Step 1: Rewriting Lemma 5 and Lemma 6

- Let: $\mu_t$ denotes the empirical mean over the sequence $x_1, ..., x_t \sim B(\theta)$, then:

$$
\forall \delta \in (0, 1), \forall \alpha > 1 \ P_{\theta}\left\{ \forall t \in \mathbb{N}^*: \text{KL}(\mu_t, \theta) < \frac{\alpha}{t} \log \frac{\log(\alpha t) \log(t)}{\log^2(\alpha)\delta} \right\} \geq 1 - \delta \quad (22)
$$

- Let: $\mu_{s,t}$ denotes the empirical mean over the sequence $x_s, ..., x_t \sim B(\theta)$, then:

$$
\forall \delta \in (0, 1), \forall \alpha > 1 \ P_{\theta}\left\{ \forall t \in \mathbb{N}^*, \forall s \in (t_0, t]: \text{KL}(\mu_{s,t}, \theta) < \frac{\alpha}{n_{s,t}} \log \frac{n_{t_0,t} \log^2(n_{t_0,t}) \log(\alpha n_{s,t}) \log(n_{s,t})}{\log(2) \log^2(\alpha)\delta} \right\} \geq 1 - \delta \quad (23)
$$

Let us build a suitable value of $\eta_{t_0:s,t}$ in order to ensure the control of the false alarm on the period $[t_0, \tau_c)$. To this end, let us control the event: $\{ \exists t > t_0, \text{Restart}_{t_0:t} = 1 \}$ which is equivalent to the event $\{ \exists t > t_0, \ s \in (t_0, t]: \vartheta_{t_0:s,t} \geq \vartheta_{t_0,t_0:t} \}$. 

Step 2: Equivalent events. First, notice that:
\[
\{ \exists t > t_0, s \in (t_0, t] : \vartheta_{t_0,s,t} \geq \vartheta_{t_0,t_0,t} \} \iff \{ \exists t > t_0, s \in (t_0, t] : \log \vartheta_{t_0,s,t} \geq \log \vartheta_{t_0,t_0,t} \}.
\]
\[(b) \{ \exists t > t_0, s \in (t_0, t] : -\log \eta_{t_0,s,t} \leq \hat{L}_{t_0:t} - \hat{L}_{t_0:s-1} \}, \]
where (a) comes directly from Equation (6).

Step 3: Using the cumulative loss controls. Then, note that \( \forall \delta \in (0, 1), \forall \alpha > 1 \) we get:
\[
P_\theta \left\{ \exists t > t_0, s \in (t_0, t] : \vartheta_{t_0,s,t} \geq \vartheta_{t_0,t_0,t} \right\} = P_\theta \left\{ \exists t > t_0, s \in (t_0, t] : \log \vartheta_{t_0,s,t} \geq \log \vartheta_{t_0,t_0,t} \right\}
\]
\[
\leq P_\theta \left\{ \exists t > t_0, s \in (t_0, t] : -\log \eta_{t_0,s,t} \leq \hat{L}_{t_0:t} - \hat{L}_{t_0:s-1} - \hat{L}_{s:t} \right\}
\]
\[(b) \leq P_\theta \left\{ \exists t > t_0, s \in (t_0, t] : -\log \eta_{t_0,s,t} \leq \log \frac{n_{t_0:s-1} \times n_{s:t}}{(n_{t_0:s-1} + 1) \times (n_{s:t} + 1)} + n_{t_0:s-1} \log (\hat{\mu}_{t_0:s-1}) + n_{s:t} \log (\hat{\mu}_{s:t}) + \frac{9}{4} \right\}
\]
\[(c) \leq P_\theta \left\{ \exists t > t_0, s \in (t_0, t] : -\log \eta_{t_0,s,t} \leq \log \frac{n_{t_0:s-1} + 1}{n_{t_0:s-1} \times n_{s:t}} + n_{t_0:s-1} \log (\hat{\mu}_{t_0:s-1}) + n_{s:t} \log (\hat{\mu}_{s:t}) + \frac{9}{4} \right\}
\]
\[(d) \leq \delta + \frac{1}{2} P_\theta \left\{ \exists t > t_0, s \in (t_0, t] : \log \frac{1}{\eta_{t_0,s,t}} \leq \log \frac{n_{t_0:s-1} + 1}{n_{t_0:s-1} \times n_{s:t}} + n_{t_0:s-1} \log (\hat{\mu}_{t_0:s-1}) + n_{s:t} \log (\hat{\mu}_{s:t}) + \frac{9}{4} \right\}
\]
\[
\leq \delta + \frac{1}{2} P_\theta \left\{ \exists t > t_0, s \in (t_0, t] : \log \frac{1}{\eta_{t_0,s,t}} \leq \log \frac{n_{t_0:s-1} + 1}{n_{t_0:s-1} \times n_{s:t}} + n_{t_0:s-1} \log (\hat{\mu}_{t_0:s-1}) + n_{s:t} \log (\hat{\mu}_{s:t}) + \frac{9}{4} \right\}
\]
\[(d) \leq \delta + P_\theta \left\{ \exists t > t_0, s \in (t_0, t] : \log \frac{1}{\eta_{t_0,s,t}} \leq \log \frac{n_{t_0:s-1} + 1}{n_{t_0:s-1} \times n_{s:t}} + n_{t_0:s-1} \log (\hat{\mu}_{t_0:s-1}) + n_{s:t} \log (\hat{\mu}_{s:t}) + \frac{9}{4} \right\}
\]
\[(e) \leq \delta + P_\theta \left\{ \exists t > t_0, s \in (t_0, t] : \log \frac{1}{\eta_{t_0,s,t}} \leq \log \frac{n_{t_0:s-1} + 1}{n_{t_0:s-1} \times n_{s:t}} + \frac{9}{4} \right\}
\]
\[(b) \text{ holds by using Lemma 4}, \text{ (c) holds thanks to } (n_{t_0:s-1} + 1) \times (n_{s:t} + 1) > n_{t_0:s-1} \times n_{s:t}, \text{ (d) holds true thanks to Equation 22 and (e) holds true thanks to Equation 23.}
\]

Step 4: Building the sufficient condition on \( \eta_{t_0,s,t} \) Thus, by using \( \exp(-\frac{9}{4}) > \frac{1}{M} \), we get the following condition on \( \eta_{t_0,s,t} \):
\[
\eta_{t_0,s,t} < \frac{\sqrt{n_{t_0:s-1} \times n_{s:t}}}{10 (n_{t_0:t} + 1)} \times \left( \frac{\log^2(\alpha)\delta}{2 \log(\alpha) \log(n_{t_0:s-1}) \log(n_{s:t})} + \frac{\log(2) \log^2(\alpha)\delta}{2 n_{t_0:t} \log^2(n_{t_0:t}) \log(n_{s:t})} \right) ^\alpha
\]
\[
= \frac{\sqrt{n_{t_0:s-1} \times n_{s:t}}}{10 (n_{t_0:t} + 1)} \times \left( \frac{\log(4) \log(2)\delta^2}{4 n_{t_0:t} \log(\alpha) \log(n_{t_0:t})} \right) ^\alpha
\]
\[
= \frac{\sqrt{n_{t_0:s-1} \times n_{s:t}}}{10 (n_{t_0:t} + 1)} \times \left( \frac{\log(4) \log(2)\delta^2}{4 n_{t_0:t} \log((\alpha + 3) n_{t_0:t})} \right) ^\alpha,
\]
which allows us to get the following control:
\[
P_\theta \left\{ \exists t > t_0, s \in (t_0, t] : \vartheta_{t_0,s,t} \geq \vartheta_{t_0,t_0,t} \right\} \leq \delta.
Proof of Theorem 3:

The proof follows three main steps:

**Step 1: Some preliminaries**  Before building the detection delay, we need to introduce three intermediate results. The first result is to link the quantity $\Phi (\Sigma_{s:t})$ to $\Phi (\hat{\mu}_{s:t})$ such that:

$$\forall (s, t) : \Phi (\Sigma_{s:t}) + \Phi (\hat{\Sigma}_{s:t}) - \Phi (n_{s:t}) = n_{s:t} (\Phi (\hat{\mu}_{s:t}) + \Phi (1 - \hat{\mu}_{s:t})).$$

Then, observe that:

$$n_{t_0:s-1} (\Phi (\hat{\mu}_{t_0:s-1}) + \Phi (1 - \hat{\mu}_{t_0:s-1})) + n_{s:t} (\Phi (\hat{\mu}_{s:t}) + \Phi (1 - \hat{\mu}_{s:t}))$$

$$- n_{t_0:t} (\Phi (\hat{\mu}_{t_0:t}) + \Phi (1 - \hat{\mu}_{t_0:t})) = n_{t_0:s-1} k l (\hat{\mu}_{t_0:s-1}, \hat{\mu}_{t_0:t}) + n_{s:t} k l (\hat{\mu}_{s:t}, \hat{\mu}_{t_0:t}).$$

Finally, observe that:

$$n_{t_0:s-1} (\hat{\mu}_{t_0:s-1} - \hat{\mu}_{t_0:t})^2 + n_{s:t} (\hat{\mu}_{s:t} - \hat{\mu}_{t_0:t})^2 = \frac{n_{t_0:s-1} n_{s:t}}{n_{t_0:t}} (\hat{\mu}_{t_0:s-1} - \hat{\mu}_{s:t})^2.$$  

Finally, following Lemma 7, the control of the quantity $|\hat{\mu}_{t_0:s-1} - \hat{\mu}_{s:t}|$ takes the following form: (with a probability at least $1 - \delta$)

$$\forall s \in [t_0 : t] \quad |\hat{\mu}_{t_0:s-1} - \hat{\mu}_{s:t}| \geq \Delta_{t_0,s,t} - C_{t_0,s,t,\delta},$$

where $\Delta_{t_0,s,t}$ represents the relative gap and it takes the following form:

$$\Delta_{t_0,s,t} = |E [\hat{\mu}_{t_0:s-1} - \hat{\mu}_{s:t}]| = \left\{ \begin{array}{ll}
\frac{n_{s:t}}{n_{t_0:t}} |\theta_1 - \theta_2| & \text{if } s < \tau_c \leq t, \\
\frac{n_{t_0:s-1}}{n_{t_0:t}} |\theta_1 - \theta_2| & \text{if } \tau_c \leq s \leq t.
\end{array} \right.$$  

**Step 2: Building the sufficient conditions for detecting the change-point $\tau_c$**  First, assume that: $x_{t_0:\tau_c-1} \sim B (\theta_1)$, $x_{\tau_c:t} \sim B (\theta_2)$. Then, to build the detection delay, we need to prove that at some instant after $\tau_c$ the restart criterion $\text{Restart}_{t_0:t}$ is activated. In other words, we need to build the following guarantee:

$$\mathbb{P} \left\{ \exists t > \tau_c : \text{Restart}_{t_0:t} = 1 \right\} > 1 - \delta.$$  

Notice that:
where (a), holds true thanks to Equation (20), (b) holds true thanks to Equation (24), (c) holds true thanks to the Pinsker Inequality taking the following form: $\forall (\theta_1, \theta_2) \in [0, 1]^2 2k_l (\theta_1, \theta_2) \geq 2 (\theta_1 - \theta_2)^2$. (d) holds true thanks to Equation (25) and (e) holds true under the condition that $\eta_{t_0,s,t} \leq \exp (f_{t_0,s,t})$.

Therefore, we obtain:

\[
\begin{align*}
\mathbb{P}\{\forall t > \tau_c : \text{Restart}_{t_0,t} = 0\} & \leq \mathbb{P}\{\forall t > \tau_c, \forall s \in (t_0,t] : \sqrt{\frac{\eta_{t_0,s,t}}{n_{t_0,t}}} |\hat{\mu}_{t_0,s,t} - \hat{\mu}_{s,t} | \leq \frac{\sqrt{f_{t_0,s,t} - \log \eta_{t_0,s,t}}}{\sqrt{2}}\} \\
& \leq \delta + \mathbb{P}\{\forall t > \tau_c, \forall s \in (t_0,t] : \sqrt{\frac{\eta_{t_0,s,t}}{n_{t_0,t}}} (\Delta_{t_0,s,t} - C_{t_0,s,t})/2 \leq \frac{f_{t_0,s,t} - \log \eta_{t_0,s,t}}{\sqrt{2}}\} \\
& \leq \delta + \mathbb{P}\{\forall t > \tau_c, \forall s \in (t_0,t] : 1 - \frac{f_{t_0,s,t} - \log \eta_{t_0,s,t}}{2n_{t_0,s,t} \times (\Delta_{t_0,s,t} - C_{t_0,s,t})^2} \leq \frac{n_{t_0,s,t} - 1}{n_{t_0,t}}\};
\end{align*}
\]

where (f) holds true thanks to Equation (26) (We recall that the relative gap $\Delta_{t_0,s,t}$ is defined in Equation (27)). Before continuing the analysis, one need to verify that term $A$ is valid (i.e. $A \in [0, 1]$, otherwise the associated event cannot be controlled). So, notice that:

\[
\begin{align*}
A > 0 & \iff \eta_{t_0,s,t} > \exp \left( -2n_{t_0,s,t} (\Delta_{t_0,s,t} - C_{t_0,s,t})^2 \right) \exp (f_{t_0,s,t}), \\
A < 1 & \iff \eta_{t_0,s,t} < \exp (f_{t_0,s,t}) = \frac{(n_{t_0,s,t} + 1)}{\sqrt{n_{t_0,t}}} \exp \left( \frac{9}{8}\right).
\end{align*}
\]

The second condition in Equation (29) is always satisfied since, we have:

\[
\forall (t_0, s, t) : \frac{(n_{t_0,s,t} + 1)}{\sqrt{n_{t_0,t}}} \exp \left( \frac{9}{8}\right) > 1 \text{ and by definition, we have: } \eta_{t_0,s,t} < 1.
\]

Therefore, from Equation (28) we get the following implication:
In other words, the change-point $\tau_c$ is detected at time $t$ (with probability at least $1 - \delta$) if for some $s \in (t_0, t]$ , we have:

$$1 + \frac{\log \eta_{t_0, s, t} - f_{t_0, s, t}}{2n_{t_0, s-1} \times (\Delta_{t_0, s, t} - C_{t_0, s, t, \delta})^2} > \frac{n_{t_0; s-1}}{n_{t_0; t}}.$$

(30)

**Step 3: Non-asymptotic expression of the detection delay** $\mathcal{D}_{\Delta, t_0, \tau_c}$ To build the detection delay, we need to ensure the existence of $s \in (t_0, t]$ such that Equation (30) is satisfied. In particular, Equation (30) can be satisfied for $s = \tau_c$. By this way, a condition to detect the change-point $\tau_c$ is written as follows:

$$1 + \frac{\log \eta_{t_0, \tau_c, t} - f_{t_0, \tau_c, t}}{2n_{t_0, \tau_c-1} \times (\Delta - C_{t_0, \tau_c, t, \delta})^2} > \frac{n_{t_0; \tau_c-1}}{n_{t_0; t}}.$$

(31)

To build the delay, we should introduce the following variable: $d = t - \tau_c + 1 = n_{r_c; t} \in \mathbb{N}^*$. Thus from Equation (31), we obtain:

$$1 + \frac{\log \eta_{t_0, \tau_c, d + \tau_c - 1} - f_{t_0, \tau_c, d + \tau_c - 1}}{2n_{t_0, \tau_c-1} \times (\Delta - C_{t_0, \tau_c, d + \tau_c - 1, \delta})^2} > \frac{n_{t_0; \tau_c-1}}{n_{t_0; t}} \Leftrightarrow d > \frac{1 - \frac{C_{t_0, \tau_c, d + \tau_c - 1, \delta}}{\Delta}}{2\Delta^2} \times \frac{-\log \eta_{t_0, \tau_c, d + \tau_c - 1} + f_{t_0, \tau_c, d + \tau_c - 1}}{1 + \frac{\log \eta_{t_0, \tau_c, d + \tau_c - 1} - f_{t_0, \tau_c, d + \tau_c - 1}}{2n_{t_0, \tau_c-1} \times (\Delta - C_{t_0, \tau_c, d + \tau_c - 1, \delta})^2}}.$$

Finally, the change-point $\tau_c$ is detected (with a probability at least $1 - \delta$) with a delay not exceeding $\mathcal{D}_{\Delta, t_0, \tau_c}$, such that:

$$\mathcal{D}_{\Delta, t_0, \tau_c} = \min \{d \in \mathbb{N}^* : d > \frac{1 - \frac{C_{t_0, \tau_c, d + \tau_c - 1, \delta}}{\Delta}}{2\Delta^2} \times \frac{-\log \eta_{t_0, \tau_c, d + \tau_c - 1} + f_{t_0, \tau_c, d + \tau_c - 1}}{1 + \frac{\log \eta_{t_0, \tau_c, d + \tau_c - 1} - f_{t_0, \tau_c, d + \tau_c - 1}}{2n_{t_0, \tau_c-1} \times (\Delta - C_{t_0, \tau_c, d + \tau_c - 1, \delta})^2}}\}. \square$$