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 $t \geqslant 1$ by R_t. Since R_t is unknown, we can consider the runlength as a random variable taking values $r_t \in \mathscr{R}_t = [0, t-1]$. Thereby, let $p\left(r_t|\mathbf{x}_{1:t}\right) = \mathbb{P}\Big\{R_t = r_t\big|\mathbf{X}_{1:t} = \mathbf{x}_{1:t}\Big\}$ denotes the distribution of R_t given the sequence of observations $\mathbf{x}_{1:t}$. ($p\left(r_t|\mathbf{x}_{1:t}\right)$ is a short hand notation).

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

$$\mathbb{P}\Big\{R_t = r_t\big|\mathbf{X}_{1:t} = \mathbf{x}_{1:t}\Big\} = \frac{\mathbb{P}\Big\{R_t = r_t, \mathbf{X}_{1:t} = \mathbf{x}_{1:t}\Big\}}{\mathbb{P}\Big\{\mathbf{X}_{1:t} = \mathbf{x}_{1:t}\Big\}}. \tag{11}$$

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

$$\begin{aligned}
&\mathbb{P}\Big\{R_t = r_t, \mathbf{X}_{1:t} = \mathbf{x}_{1:t}\Big\} \overset{(a)}{=} \sum_{r_{t-1}\in\mathscr{R}_{t-1}} \mathbb{P}\Big\{R_t = r_t, \mathbf{X}_{1:t} = \mathbf{x}_{1:t}, R_{t-1} = r_{t-1}\Big\} \\
&\overset{(b)}{=} \sum_{r_{t-1}\in\mathscr{R}_{t-1}} \mathbb{P}\Big\{R_t = r_t, X_t = x_t\big|R_{t-1} = r_{t-1}, \mathbf{X}_{1:t-1} = \mathbf{x}_{1:t-1}\Big\}\mathbb{P}\Big\{R_{t-1} = r_{t-1}, \mathbf{X}_{1:t-1} = \mathbf{x}_{1:t-1}\Big\} \\
&\overset{(c)}{=} \sum_{r_{t-1}\in\mathscr{R}_{t-1}} \mathbb{P}\Big\{X_t = x_t\big|\cancel{R_t = r_t}, R_{t-1} = r_{t-1}, \mathbf{X}_{1:t-1} = \mathbf{x}_{1:t-1}\Big\}\mathbb{P}\Big\{R_t = r_t\big|R_{t-1} = r_{t-1}, \cancel{\mathbf{X}_{1:t-1} = \mathbf{x}_{1:t-1}}\Big\} \\
&\qquad\qquad \times \mathbb{P}\Big\{R_{t-1} = r_{t-1}, \mathbf{X}_{1:t-1} = \mathbf{x}_{1:t-1}\Big\} \\
&\overset{(d)}{=} \sum_{r_{t-1}\in\mathscr{R}_{t-1}} \underbrace{\mathbb{P}\Big\{R_t = r_t\big|R_{t-1} = r_{t-1}\Big\}}_{p(r_t|r_{t-1})}\underbrace{\mathbb{P}\Big\{X_t = x_t\big|R_{t-1} = r_{t-1}, \mathbf{X}_{1:t-1} = \mathbf{x}_{1:t-1}\Big\}}_{p(x_t|r_{t-1}, \mathbf{x}_{1:t-1})}\mathbb{P}\Big\{R_{t-1} = r_{t-1}, \mathbf{X}_{1:t-1} = \mathbf{x}_{1:t-1}\Big\}.
\end{aligned} \tag{12}$$

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 $\mathbf{X}_{1:t-1}$ and X_t do not depend on R_t.

Thus, combining Equation (11) and Equation (12) we get (using the short-hand notations):

$$p\left(r_t|\mathbf{x}_{1:t}\right) \propto \sum_{r_{t-1}\in\mathscr{R}_{t-1}} p\left(r_t|r_{t-1}\right) p\left(x_t|r_{t-1}, \mathbf{x}_{1:t-1}\right) p\left(r_{t-1}|\mathbf{x}_{1:t-1}\right). \tag{13}$$

So, given the previous runlength distribution $p\left(r_{t-1}|\mathbf{x}_{1:t-1}\right)$, one can thus build a *message-passing algorithm* for the current run-length distribution $p(r_t\big|\mathbf{x}_{1:t})$ by calculating:

1. the underlying predictive model (UPM) $p\left(x_t\big|r_{t-1}, \mathbf{x}_{1:t-1}\right)$,
2. the hazard function $p\left(r_t|r_{t-1}\right)$.

It should be noted that at each time t, the runlength R_t either continues to grow (which corresponds to the event $\big\{R_t = R_{t-1} + 1\big\}$) or a change occurs which corresponds to $\big\{R_t = 0\big\}$. Thus, from equation (13), we get the following recursive runlength distribution estimation:

- Growth probability:

$$p\left(r_t = r_{t-1} + 1|\mathbf{x}_{1:t}\right) \propto p\left(r_t|r_{t-1}\right) p\left(x_t|r_{t-1}, \mathbf{x}_{1:t-1}\right) p\left(r_{t-1}|\mathbf{x}_{1:t-1}\right). \tag{14}$$

- Change-point probability:

$$p\left(r_t = 0|\mathbf{x}_{1:t}\right) \propto \sum_{r_{t-1}\in\mathscr{R}_{t-1}} p\left(r_t|r_{t-1}\right) p\left(x_t|r_{t-1},\mathbf{x}_{1:t-1}\right) p\left(r_{t-1}|\mathbf{x}_{1:t-1}\right). \tag{15}$$

Hazard function. According to Equation (14) and Equation (15), the runlength distribution estimation need to compute the change-point prior $\mathbb{P}(R_t \big| R_{t-1})$, which is done following the assumption that hazard function is a constant $h \in (0,1)$ in the sense that $\mathbb{P}(R_t \big| R_{t-1})$ is independent of r_{t-1} and is constant, giving rise, a priori, to geometric inter-arrival times for change points.

$$\mathbb{P}\Big\{R_t \big| R_{t-1}\Big\} = h\mathbb{I}\{R_t = 0\} + (1-h)\,\mathbb{I}\{R_t = R_{t-1}+1\}. \tag{16}$$

Then, injecting Equation (16) into Equation (14) and Equation (15) we get:

$$p\left(r_t = r_{t-1}+1\big|\mathbf{x}_{1:t}\right) \propto (1-h)\,p\left(x_t\big|r_{t-1},\mathbf{x}_{1:t-1}\right) \times p\left(r_{t-1}\big|\mathbf{x}_{1:t-1}\right), \tag{17}$$

$$p\left(r_t = 0\big|\mathbf{x}_{1:t}\right) \propto h \sum_{r_{t-1}\in\mathscr{R}_{t-1}} p\left(x_t\big|r_{t-1},\mathbf{x}_{1:t-1}\right) \times p\left(r_{t-1}\big|\mathbf{x}_{1:t-1}\right). \tag{18}$$

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:*

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

Proof of Lemma 1:

You only need to see that:

$$\begin{aligned}
V_t &= \sum_{s=1}^{t} v_{s,t} \\
&= \sum_{s=1}^{t-1} v_{s,t} + v_{t,t} \\
&= (1-h)\sum_{s=1}^{t-1} \exp\left(-l_{s,t}\right) v_{s,t-1} + h\sum_{s=1}^{t-1} \exp\left(-l_{s,t}\right) v_{s,t-1} \\
&= \sum_{s=1}^{t-1} \exp\left(-l_{s,t}\right) v_{s,t-1}.
\end{aligned}$$

□

Proof of Lemma 2:

First, for all $t \geqslant 2$, we have:

$$
\begin{aligned}
&V_t = \sum_{i=1}^{t} v_{i,t} \\
&V_t = v_{1,t} + \sum_{i=2}^{t-1} v_{i,t} + v_{t,t} \\
&V_t = (1-h)^{t-1} \exp\left(-\widehat{L}_{1:t}\right) V_1 + \sum_{i=2}^{t-1} (1-h)^{t-i} \exp\left(-\widehat{L}_{i:t}\right) hV_i + hV_t. \\
&\Leftrightarrow V_t = \sum_{i=1}^{t} \underbrace{(1-h)^{t-i} \exp\left(-\widehat{L}_{i:t}\right) h^{1(i\neq 1)}}_{\alpha_{t,i}} V_i \text{ with convention: } L_{i,j} = 0 \Leftrightarrow i > j. \\
&\Leftrightarrow V_t = \sum_{i=1}^{t} \alpha_{t,i} V_i. \\
&\Leftrightarrow (1-\alpha_{t,t}) V_t = \sum_{i=1}^{t-1} \alpha_{t,i} V_i.
\end{aligned}
$$

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 $V_1 = 1$):

$$
\begin{aligned}
&\forall t \geqslant 4, \\
&V_t = \left(\beta_{t,1} + \sum_{i_1=1}^{t-2} \beta_{t,t-i_1}\beta_{t-i_1,1} + \sum_{k=3}^{t-1}\sum_{i_1=1}^{t-k}\sum_{i_2=i_1+1}^{t-(k-1)} \cdots \sum_{i_{k-1}=i_{k-2}+1}^{t-2} \beta_{t,t-i_1}\beta_{t-i_1,t-i_2}\cdots\beta_{t-i_{k-1},1}\right) V_1 \\
&= \beta_{t,1} + \sum_{i_1=1}^{t-2} \beta_{t,t-i_1}\beta_{t-i_1,1} + \sum_{k=3}^{t-1}\sum_{i_1=1}^{t-k}\sum_{i_2=i_1+1}^{t-(k-1)} \cdots \sum_{i_{k-1}=i_{k-2}+1}^{t-2} \beta_{t,t-i_1}\beta_{t-i_1,t-i_2}\cdots\beta_{t-i_{k-1},1}. \\
&V_3 = \beta_{3,1} + \beta_{3,2}\beta_{2,1}. \\
&V_2 = \beta_{2,1}.
\end{aligned}
$$

which can be concatenated in the following form:

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

$$
\tilde{V}_{k:t} = \overbrace{\sum_{i_1=1}^{t-k}\sum_{i_2=i_1+1}^{t-(k-1)} \ldots \sum_{i_{k-1}=i_{k-2}+1}^{t-2}}^{\binom{t-2}{k-1}} \exp\left(-\widehat{L}_{1:i_1}\right) \times \prod_{j=1}^{k-2} \exp\left(-\widehat{L}_{i_j+1:i_{j+1}}\right) \exp\left(-\widehat{L}_{i_{k-1}+1:t-1}\right),
$$

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

□

Proof of Lemma 3:

First, notice that the cumulative loss $\widehat{L}_{s:t}$ can be written as follows:

$$\widehat{L}_{s,t} = -\log \prod_{s'=s}^{t} \mathtt{Lp}\left(x_s | \mathbf{x}_{s':s-1}\right)$$

Then, we only need to show by induction that:

$$\forall \mathbf{x}_{1:n} \in \{0,1\}^n \quad \prod_{s=1}^{n} \mathtt{Lp}\left(x_s | \mathbf{x}_{1:s-1}\right) = \frac{1}{(n+1)\binom{n}{\sum_{i=1}^{n} x_i}}.$$

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 $\mathtt{Lp}\left(\cdot|\cdot\right)$, we obtain:

$$\begin{cases} \mathtt{Lp}\left(1|\emptyset\right) = 1/2 = \frac{1}{(1+1)\binom{1}{1}}, \\ \mathtt{Lp}\left(0|\emptyset\right) = 1/2 = \frac{1}{(1+1)\binom{1}{0}}. \end{cases}$$

Step 2: Assume that for some $\mathbf{x}_{1:n} \in \{0,1\}^n$, we have:

$$\prod_{s=1}^{n} \mathtt{Lp}\left(x_s | \mathbf{x}_{1:s-1}\right) = \frac{1}{(n+1)\binom{n}{\sum_{i=1}^{n} x_i}}. \tag{19}$$

Then, let us verify that:

$$\forall \mathbf{x}_{n+1} \in \{0,1\} \quad \prod_{s=1}^{n+1} \mathtt{Lp}\left(x_s | \mathbf{x}_{1:s-1}\right) = \frac{1}{(n+2)\binom{n+1}{\sum_{i=1}^{n+1} x_i}}.$$

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:

$$\prod_{s=1}^{n+1} \mathtt{Lp}\left(x_s | \mathbf{x}_{1:s-1}\right) = \prod_{s=1}^{n} \mathtt{Lp}\left(x_s | \mathbf{x}_{1:s-1}\right) \mathtt{Lp}\left(1 | \mathbf{x}_{1:n}\right).$$

Using the definition of the predictor and the assumption (19), we obtain:

$$\begin{aligned} \prod_{s=1}^{n+1} \mathtt{Lp}\left(x_s | \mathbf{x}_{1:s-1}\right) &= \frac{1}{(n+1)\binom{n}{\sum_{i=1}^{n} x_i}} \times \frac{\sum_{i=1}^{n} x_i + 1}{n+2} \\ &\overset{(a)}{=} \frac{\left(\sum_{i=1}^{n} x_i + 1\right) \times \left(\sum_{i=1}^{n} x_i\right)! \times \left(\sum_{i=1}^{n} \bar{x}_i\right)!}{(n+2)(n+1)\, n!} \\ &= \frac{\left(\sum_{i=1}^{n} x_i + 1\right)! \times \left(\sum_{i=1}^{n} \bar{x}_i + 0\right)!}{(n+2)(n+1)!} \\ &= \frac{\left(\sum_{i=1}^{n+1} x_i\right)! \times \left(\sum_{i=1}^{n+1} \bar{x}_i\right)!}{(n+2)(n+1)!} \\ &= \frac{1}{(n+2)\binom{n+1}{\sum_{i=1}^{n+1} x_i}}. \end{aligned}$$

where (a) holds using the definition of the Binomial operator.

Case 2: $x_{n+1} = 0$ Observe that:

$$\prod_{s=1}^{n+1} \mathtt{Lp}\left(x_s|\mathbf{x}_{1:s-1}\right) = \prod_{s=1}^{n} \mathtt{Lp}\left(x_s|\mathbf{x}_{1:s-1}\right) \mathtt{Lp}\left(0|\mathbf{x}_{1:n}\right).$$

Using the definition of the predictor and the assumption (19), we obtain:

$$\begin{aligned}
\prod_{s=1}^{n+1} \mathtt{Lp}\left(x_s|\mathbf{x}_{1:s-1}\right) &= \frac{1}{(n+1)\binom{n}{\sum_{i=1}^n x_i}} \times \frac{\sum_{i=1}^n \bar{x}_i + 1}{n+2} \\
&\overset{(b)}{=} \frac{\left(\sum_{i=1}^n \bar{x}_i + 1\right) \times \left(\sum_{i=1}^n x_i\right)! \times \left(\sum_{i=1}^n \bar{x}_i\right)!}{(n+2)(n+1)\, n!} \\
&= \frac{\left(\sum_{i=1}^n x_i + 0\right)! \times \left(\sum_{i=1}^n \bar{x}_i + 1\right)!}{(n+2)(n+1)!} \\
&= \frac{\left(\sum_{i=1}^{n+1} x_i\right)! \times \left(\sum_{i=1}^{n+1} \bar{x}_i\right)!}{(n+2)(n+1)!} \\
&= \frac{1}{(n+2)\binom{n+1}{\sum_{i=1}^{n+1} x_i}}.
\end{aligned}$$

where (b) holds using the definition of the Binomial operator. □

Proof of Lemma 4:

The proof follows three main steps:

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

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

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

$$\forall\, n \geqslant 1, \forall k \in [0,n] \quad \frac{n^n}{k^k (n-k)^{n-k}} \frac{\exp(b_1)}{\sqrt{n}} \leqslant \binom{n}{k} \leqslant \frac{n^n}{k^k (n-k)^{n-k}} \text{ with } b_1 = -\frac{1}{6} - \frac{1}{2}\log(2\pi). \tag{20}$$

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

$$\widehat{L}_{s:t} = \log\left(n_{s:t} + 1\right) + \log\binom{n_{s:t}}{\Sigma_{s:t}}.$$

Then by letting $\Phi(x) = x \log x$ and by following Equation (20), we obtain the following two bounds:

$$\begin{cases} \widehat{L}_{s:t} & \leqslant \log\left(n_{s:t}+1\right) + \Phi\left(n_{s:t}\right) - \Phi\left(\Sigma_{s:t}\right) - \Phi\left(\bar{\Sigma}_{s:t}\right), \\ \widehat{L}_{s:t} & \geqslant \log\left(n_{s:t}+1\right) + \Phi\left(n_{s:t}\right) - \Phi\left(\Sigma_{s:t}\right) - \Phi\left(\bar{\Sigma}_{s:t}\right) - \frac{9}{8} - \frac{1}{2}\log n_{s:t}. \end{cases} \tag{21}$$

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

$$\Sigma_{s:t}\log\Sigma_{s:t} + \bar{\Sigma}_{s:t}\log\bar{\Sigma}_{s:t} = \Sigma_{s:t}\log\theta + \bar{\Sigma}_{s:t}\log\bar{\theta} + n_{s:t}\log n_{s:t} + n_{s:t}\boldsymbol{kl}\left(\frac{\Sigma_{s:t}}{n_{s:t}}, \theta\right). \tag{22}$$

Then, using Equations (21) with Equation (22), we obtain:

for the upper bound of the loss $\widehat{L}_{s:t}$

$$\begin{aligned}\widehat{L}_{s:t} &\leqslant \log\left(n_{s:t}+1\right) - \Sigma_{s:t}\log\frac{\Sigma_{s:t}}{n_{s:t}} - \bar{\Sigma}_{s:t}\log\frac{\bar{\Sigma}_{s:t}}{n_{s:t}} \\ &\leqslant \log\left(n_{s:t}+1\right) - \Sigma_{s:t}\log\Sigma_{s:t} - \bar{\Sigma}_{s:t}\log\bar{\Sigma}_{s:t} + n_{s:t}\log n_{s:t} \\ &\overset{(a)}{\leqslant} \log\left(n_{s:t}+1\right) - \Sigma_{s:t}\log\theta - \bar{\Sigma}_{s:t}\log\bar{\theta} - n_{s:t}\boldsymbol{kl}\left(\frac{\Sigma_{s:t}}{n_{s:t}},\theta\right) \\ &\overset{(b)}{\leqslant} \log\left(n_{s:t}+1\right) - \Sigma_{s:t}\log\theta - \bar{\Sigma}_{s:t}\log\bar{\theta},\end{aligned}$$

where (a) holds by using Equation (22) and (b) holds using the positiveness of the Kullback Leibler divergence ($\boldsymbol{kl}\left(\bullet,\bullet\right) \geqslant 0$),

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

$$\begin{aligned}\widehat{L}_{s:t} &\geqslant \log\left(n_{s:t}+1\right) - \frac{1}{2}\log n_{s:t} - \Sigma_{s:t}\log\theta - \bar{\Sigma}_{s:t}\log\bar{\theta} - n_{s:t}\boldsymbol{kl}\left(\frac{\Sigma_{s:t}}{n_{s:t}},\theta\right) + b_1. \\ &\geqslant \log\left(n_{s:t}+1\right) - \frac{1}{2}\log n_{s:t} - \Sigma_{s:t}\log\theta - \bar{\Sigma}_{s:t}\log\bar{\theta} - n_{s:t}\boldsymbol{kl}\left(\widehat{\mu}_{s:t},\theta\right) - \frac{9}{8}.\end{aligned}$$

□

Proof of Lemma 5 and Lemma 6:

The interested reader can refer for more details on the proofs of Lemma 5 and Lemma 6 to the manuscript untitled "Mathematics of Statistical Sequential Decision Making" `https://pdfs.semanticscholar.org/9099/c0f71185adce7705beb78d595abc817c33d6.pdf` □

Proof of Lemma 7:

Step 1 Without a loss of generality, we consider that $r = 1$ and we consider that the sequence $(x_t)_t$ has σ-sub Gaussian noise meaning that:

$$\forall t, \forall \lambda \in \mathbb{R}, \quad \log\mathbb{E}\left[\exp\left(\lambda\left(x_t - \mathbb{E}\left[x_t\right]\right)\right)\right] \leqslant \frac{\lambda^2\sigma^2}{2} \tag{23}$$

Note that the Bernoulli case is a σ sub-Gaussianity case where $\sigma = \frac{1}{2}$. Indeed:

$$\forall \lambda \in \mathbb{R}, \quad \log\mathbb{E}_{X\sim B(p)}\exp(\lambda(X-p)) \leqslant \frac{\lambda^2}{8}$$

Let $\bar{z}_{s+1:t} = \widehat{\mu}_{s+1:t} - \mathbb{E}\left[\widehat{\mu}_{s+1:t}\right]$ be the centered empirical mean using observations from $s+1$ to t. We first introduce for each $\lambda \in \mathbb{R}$ and each $s \leqslant t$ the following quantity:

$$B^{\lambda}_{s,t} = \exp\left(\lambda(t-s)\bar{z}_{s+1:t} - \frac{\lambda^2\sigma^2(t-s)}{2}\right)$$

Note that $\left(B^{\lambda}_{s,t}\right)_{t\in[s,\infty]\cap\mathrm{N}}$ is a non-negative supermartingale. Let us introduce $B_{s,t} = \mathbb{E}\left[B^{\Lambda}_{s,t}\right]$, where $\Lambda \sim \mathscr{N}\left(0, \frac{1}{\sigma^2(t-s)c}\right)$, for some $c > 0$. We note that by simple algebra,

$$|\bar{z}_{s+1:t}| = \sqrt{\frac{2\sigma^2(1+c)}{t-s}\ln\left(B_{s,t}\sqrt{1+1/c}\right)}$$

In particular, choosing $c = 1/(t-s)$, it comes for all deterministic $g(t) > 0$, that

$$\begin{aligned}\mathbb{P}\left(\exists t, \exists s < t, |\bar{z}_{s+1:t}| \geqslant \sqrt{\frac{2\sigma^2\left(\frac{t-s+1}{t-s}\right)}{t-s}\ln\left(\frac{g(t)\sqrt{1+t-s}}{\delta}\right)}\right) &= \mathbb{P}\Big\{\exists t, \exists s < t, B_{s,t} \geqslant \frac{g(t)}{\delta}\Big\}\\ &\leqslant \mathbb{P}\left(\exists t, \max_{s<t} B_{s,t} \geqslant \frac{g(t)}{\delta}\right)\\ &\leqslant \delta\mathbb{E}\left[\max_{t}\frac{\max_{s<t} B_{s,t}}{g(t)}\right]\end{aligned}$$

Step 2 This leads to study the quantity $\frac{\max_{s<t} B_{s,t}}{g(t)}$. To this end, it is convenient to introduce $\bar{B}_t = \frac{\sum_{s<t} B_{s,t}}{g(t)}$ for $t > 1$. Indeed, for every random stopping time $\tau > 1$,

$$\mathbb{E}\left[\frac{\max_{s<\tau} B_{s,\tau}}{g(\tau)}\right] \leqslant \mathbb{E}\left[\bar{B}_\tau\right] = \mathbb{E}\left[\bar{B}_2 + \sum_{t=2}^{\infty}\left(\bar{B}_{t+1} - \bar{B}_t\right)\mathbb{I}\{\tau > t\}\right]$$

Further, we note that, conveniently

$$\bar{B}_{t+1} - \bar{B}_t = \frac{B_{t,t+1}}{g(t+1)} + \sum_{s<t}\left(\frac{B_{s,t+1}}{g(t+1)} - \frac{B_{s,t}}{g(t)}\right)$$

Next, by construction, we note that

$$\mathbb{E}\left[B_{s,t+1} \mid \mathscr{F}_t\right] = \frac{\sigma}{\sqrt{2\pi}}\int_{\mathbb{R}}\mathbb{E}\left[B^{\lambda}_{s,t+1} \mid \mathscr{F}_t\right]e^{-\frac{\lambda^2\alpha^2}{2}}d\lambda \leqslant \frac{\sigma}{\sqrt{2\pi}}\int_{\mathbb{R}} B^{\lambda}_{s,t}e^{-\frac{\lambda^2 q^2}{2}}d\lambda = B_{s,t}$$

Thus, since $\mathbb{I}\{\tau > t\} \in \mathscr{F}_t$, we deduce that

$$\begin{aligned}\mathbb{E}\left[\frac{\max_{s<\tau} B_{s,\tau}}{g(\tau)}\right] &\leqslant \mathbb{E}\left[\bar{B}_2\right] + \sum_{t=2}^{\infty}\frac{\mathbb{E}\left[B_{t,t+1}\right]}{g(t+1)} + \sum_{t=1}^{\infty}\sum_{s<t}\mathbb{E}\left[\left(\frac{1}{g(t+1)} - \frac{1}{g(t)}\right)B_{s,t}\mathbb{I}\{\tau > t\}\right]\\ &= \mathbb{E}\left[\bar{B}_2\right] + \sum_{t=2}^{\infty}\frac{\mathbb{E}\left[B_{t,t+1}\right]}{g(t+1)} + \sum_{t=1}^{\infty}\sum_{s<t}\left(\frac{1}{g(t+1)} - \frac{1}{g(t)}\right)\underbrace{\mathbb{E}\left[B_{s,t}\mathbb{I}\{\tau > t\}\right]}_{\geqslant 0}\end{aligned}$$

Hence, choosing g as an increasing function of t ensures that the last sum is upper bounded by 0. since on the other hand $\mathbb{E}\left[B_{t,t+1}\right] \leqslant 1$ and $\mathbb{E}\left[\bar{B}_2\right] \leqslant 1/g(2)$, we deduce that

$$\mathbb{E}\left[\frac{\max_{s<\tau} B_{s,\tau}}{g(\tau)}\right] \leqslant \frac{1}{g(2)} + \sum_{t=2}^{\infty}\frac{1}{g(t+1)} = \sum_{t=2}^{\infty}\frac{1}{g(t)}$$

Choosing $g(t) = Ct\ln^2(t)$ for $t > 1$ yields

$$\mathbb{E}\left[\frac{\max_{s<\tau} B_{s,\tau}}{g(\tau)}\right] \leqslant \frac{1}{C\ln(2)}$$

Plugging-in this in the control of the deviation and choosing $C = 1/\ln(2)$ thus gives

$$P\left(\exists t, \exists s < t \quad |\bar{z}_{s+1:t}| \geqslant \sqrt{\frac{2\sigma^2\left(1+\frac{1}{t-s}\right)}{t-s}\ln\left(\frac{t\ln^2(t)\sqrt{t+1-s}}{\ln(2)\delta}\right)}\right) \leqslant \delta$$

since on the other hand, by the classical Laplace method (see Lemma 8),

$$\mathbb{P}\left(\exists s, \quad |\bar{z}_{1:s}| \geqslant \sqrt{\frac{2\sigma^2\left(1+\frac{1}{s}\right)}{s}\ln\left(\frac{\sqrt{s+1}}{\delta}\right)}\right) \leqslant \delta$$

we conclude by using the triangular inequality $\sqrt{z}_{1:s} - \bar{z}_{s+1:t}| \leqslant |\bar{z}_{1:s}| + |\bar{z}_{s+1:t}\ |$ together with a union bound argument.

□

Lemma 8 (Uniform confidence intervals)**.** *Let $Y_1, \ldots Y_t$ be a sequence of t i.i.d. real-valued random variables with mean μ, such that $Y_t - \mu$ is σ-sub-Gaussian. Let $\widehat{\mu}_t = \frac{1}{t}\sum_{s=1}^{t} Y_s$ be the empirical mean estimate. Then, for all $\delta \in (0,1)$, it holds*

$$\mathbb{P}\left(\exists t \in \mathbb{N}, \quad |\widehat{\mu}_t - \mu| \geqslant \sigma\sqrt{\left(1+\frac{1}{t}\right)\frac{2\ln(\sqrt{t+1}/\delta)}{t}}\right) \leqslant \delta$$

(The "Laplace" method refers to using the Laplace method of integration for optimization)

Proof of Lemma 8:

We introduce for a fixed $\delta \in [0,1]$ the random variable

$$\tau = \min\left\{t \in \mathbb{N} : \widehat{\mu}_t - \mu \geqslant \sigma\sqrt{\left(1+\frac{1}{t}\right)\frac{2\ln(\sqrt{1+t}/\delta)}{t}}\right\}$$

This quantity is a random stopping time for the filtration $\mathscr{F} = (\mathscr{F}_t)_t$, where $\mathscr{F}_t = \sigma(Y_1, \ldots, Y_t)$, since $\{\tau \leqslant m\}$ is $\mathscr{F}_m$ -measurable for all m. We want to show that $\mathbb{P}(\tau < \infty) \leqslant \delta$. To this end, for any λ, and t, we introduce the following quantity

$$M_t^\lambda = \exp\left(\sum_{s=1}^{t}\left(\lambda(Y_s - \mu) - \frac{\lambda^2\sigma^2}{2}\right)\right)$$

By assumption, the centered random variables are σ -sub-Gaussian and it is immediate to show that $\left\{M_t^\lambda\right\}_{t\in\mathbb{N}}$ is a non-negative super-martingale that satisfies $\ln\mathbb{E}\left[M_t^\lambda\right] \leqslant 0$ for all t. It then follows that $M_\infty^\lambda = \lim_{t\to\infty} M_t^\lambda$ is almost surely well-defined and so, M_τ^λ as well. Further, using the face that M_t^λ and $\{\tau > t\}$ are $\mathscr{F}_t$ measurable, it comes

$$\begin{aligned}
\mathbb{E}\left[M_\tau^\lambda\right] &= \mathbb{E}\left[M_1^\lambda\right] + \mathbb{E}\left[\sum_{t=1}^{\tau-1} M_{t+1}^\lambda - M_t^\lambda\right] \\
&= 1 + \sum_{t=1}^{\infty}\mathbb{E}\left[\left(M_{t+1}^\lambda - M_t^\lambda\right)\mathbb{I}\{\tau > t\}\right] \\
&= 1 + \sum_{t=1}^{\infty}\mathbb{E}\left[\left(\mathbb{E}\left[M_{t+1}^\lambda \mid \mathscr{F}_t\right] - M_t^\lambda\right)\mathbb{I}\{\tau > t\}\right] \\
&\leqslant 1
\end{aligned}$$

The next step is to introduce the auxiliary variable $\Lambda = \mathscr{N}\left(0, \sigma^{-2}\right)$, independent of all other variables, and study the quantity $M_t = \mathbb{E}\left[M_t^{\Lambda} \mid \mathscr{F}_\infty\right]$. Note that the standard deviation of Λ is σ^{-1} due to the fact we consider σ -sub-Gaussian random variables. We immediately get $\mathbb{E}\left[M_\tau\right] = \mathbb{E}\left[\mathbb{E}\left[M_\tau^{\Lambda} \mid \Lambda\right]\right] \leqslant 1$. For convenience, let $S_t = t\left(\mu_t - \mu\right)$. By construction of M_t, we have

$$\begin{aligned}
M_t &= \frac{1}{\sqrt{2\pi\sigma^{-2}}} \int_{\mathbb{R}} \exp\left(\lambda S_t - \frac{\lambda^2\sigma^2 t}{2} - \frac{\lambda^2\sigma^2}{2}\right) d\lambda \\
&= \frac{1}{\sqrt{2\pi\sigma^{-2}}} \int_{\mathbb{R}} \exp\left(-\left[\lambda\sigma\sqrt{\frac{t+1}{2}} - \frac{S_t}{\sigma\sqrt{2(t+1)}}\right]^2 + \frac{S_t^2}{2\sigma^2(t+1)}\right) d\lambda \\
&= \exp\left(\frac{S_t^2}{2\sigma^2(t+1)}\right) \frac{1}{\sqrt{2\pi\sigma^{-2}}} \int_{\mathbb{R}} \exp\left(-\lambda^2\sigma^2\frac{t+1}{2}\right) d\lambda \\
&= \exp\left(\frac{S_t^2}{2\sigma^2(t+1)}\right) \frac{\sqrt{2\pi\sigma^{-2}/(t+1)}}{\sqrt{2\pi\sigma^{-2}}}
\end{aligned}$$

Thus, we deduce that

$$|S_t| = \sigma\sqrt{2(t+1)\ln\left(\sqrt{t+1}M_t\right)}$$

We conclude by applying a simple Markov inequality:

$$\mathbb{P}\left(\tau\left|\widehat{\mu}_\tau - \mu\right| \geqslant \sigma\sqrt{2(\tau+1)\ln(\sqrt{\tau+1}/\delta)}\right) = \mathbb{P}\left(M_\tau \geqslant 1/\delta\right) \leqslant \mathbb{E}\left[M_\tau\right]\delta$$

□

C. Proofs of Theorems

Proof of Theorem 2:

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

Step 1: Rewriting Lemma 5 and Lemma 6

- Let: $\widehat{\mu}_t$ denotes the empirical mean over the sequence $x_1, ..., x_t \sim \mathcal{B}(\theta)$, then:

$$\forall\delta \in (0,1), \forall\alpha > 1 \quad \mathbb{P}_\theta\Big\{\underbrace{\forall t \in \mathbb{N}^\star : \boldsymbol{kl}\left(\widehat{\mu}_t, \theta\right) < \frac{\alpha}{t}\log\frac{\log(\alpha t)\log(t)}{\log^2(\alpha)\delta}}_{E^{(1)}_{\theta,\delta,\alpha}}\Big\} \geqslant 1-\delta \tag{24}$$

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

$$\forall\delta \in (0,1), \forall\alpha > 1 \quad \mathbb{P}_\theta\Big\{\underbrace{\forall t \in \mathbb{N}^\star, \forall s \in (r,t] : \boldsymbol{kl}\left(\widehat{\mu}_{s:t}, \theta\right) < \frac{\alpha}{n_{s:t}}\log\frac{n_{r:t}\log^2(n_{r:t})\log(\alpha n_{s:t})\log(n_{s:t})}{\log(2)\log^2(\alpha)\delta}}_{E^{(2)}_{\theta,\delta,\alpha}}\Big\} \geqslant 1-\delta \tag{25}$$

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

Step 2: Equivalent events. First, notice that:

$$\begin{aligned}\Big\{\exists t > r,\ s \in (r,t] : \vartheta_{r,s,t} \geqslant \vartheta_{r,r,t}\Big\} &\Leftrightarrow \Big\{\exists t > r,\ s \in (r,t] : \quad \log \vartheta_{r,s,t} \geqslant \log \vartheta_{r,r,t}\Big\}. \\ &\overset{(a)}{\Leftrightarrow} \Big\{\exists t > r,\ s \in (r,t] : -\log \eta_{r,s,t} \leqslant \widehat{L}_{r:t} - \widehat{L}_{s:t} - \widehat{L}_{r:s-1}\Big\},\end{aligned}$$

where (a) comes directly from Equation (10).

Step 3: Using the cumulative loss controls. Then, note that $\forall \delta \in (0,1), \forall \alpha > 1$ we get:

$$\begin{aligned}
&\mathbb{P}_\theta\Big\{\exists t > r,\ s \in (r,t] : \vartheta_{r,s,t} \geqslant \vartheta_{r,r,t}\Big\} = \mathbb{P}\Big\{\exists t > r,\ s \in (r,t] : \log \vartheta_{r,s,t} \geqslant \log \vartheta_{r,r,t}\Big\} \\
&= \mathbb{P}_\theta\Big\{\exists t > r,\ s \in (r,t] : \quad -\log \eta_{r,s,t} \leqslant \widehat{L}_{r:t} - \widehat{L}_{r:s-1} - \widehat{L}_{s:t}\Big\} \\
&\overset{(b)}{\leqslant} \mathbb{P}_\theta\Bigg\{\exists t > r,\ s \in (r,t] : -\log \eta_{r,s,t} \leqslant \log \frac{\sqrt{n_{r:s-1} \times n_{s:t}} \times (n_{r:t}+1)}{(n_{r:s-1}+1) \times (n_{s:t}+1)} + n_{r:s-1}\boldsymbol{kl}\left(\widehat{\mu}_{r:s-1}, \theta\right) + n_{s:t}\boldsymbol{kl}\left(\widehat{\mu}_{s:t}, \theta\right) + \frac{9}{4}\Bigg\} \\
&\overset{(c)}{\leqslant} \mathbb{P}_\theta\Bigg\{\exists t > r,\ s \in (r,t] : -\log \eta_{r,s,t} \leqslant \log \frac{n_{r:t}+1}{\sqrt{n_{r:s-1} \times n_{s:t}}} + n_{r:s-1}\boldsymbol{kl}\left(\widehat{\mu}_{r:s-1}, \theta\right) + n_{s:t}\boldsymbol{kl}\left(\widehat{\mu}_{s:t}, \theta\right) + \frac{9}{4}\Bigg\} \\
&\leqslant \frac{\delta}{2} + \mathbb{P}_\theta\Bigg\{\exists t > r,\ s \in (r,t] : -\log \eta_{r,s,t} \leqslant \log \frac{n_{r:t}+1}{\sqrt{n_{r:s-1} \times n_{s:t}}} + n_{r:s-1}\boldsymbol{kl}\left(\widehat{\mu}_{r:s-1}, \theta\right) + n_{s:t}\boldsymbol{kl}\left(\widehat{\mu}_{s:t}, \theta\right) + \frac{9}{4} \bigcap E^{(1)}_{\theta,\delta/2,\alpha}\Bigg\} \\
&\overset{(d)}{\leqslant} \frac{\delta}{2} + \mathbb{P}_\theta\Bigg\{\exists t > r, s \in (r,t] : \log \frac{1}{\eta_{r,s,t}} \leqslant \log \frac{n_{r:t}+1}{\sqrt{n_{r:s-1} \times n_{s:t}}} + \alpha \log \frac{2\log(\alpha n_{r:s-1})\log(n_{r:s-1})}{\log^2(\alpha)\delta} + n_{s:t}\boldsymbol{kl}\left(\widehat{\mu}_{s:t}, \theta\right) + \frac{9}{4}\Bigg\} \\
&\leqslant \delta + \mathbb{P}_\theta\Bigg\{\exists t > r, s \in (r,t] : \log \frac{1}{\eta_{r,s,t}} \leqslant \log \frac{n_{r:t}+1}{\sqrt{n_{r:s-1} n_{s:t}}} + \alpha \log \frac{2\log(\alpha n_{r:s-1})\log n_{r:s-1}}{\log^2(\alpha)\delta} + n_{s:t}\boldsymbol{kl}\left(\widehat{\mu}_{s:t}, \theta\right) + \frac{9}{4} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\bigcap E^{(2)}_{\theta,\delta/2,\alpha}\Bigg\} \\
&\overset{(e)}{\leqslant} \delta + \mathbb{P}_\theta\Bigg\{\exists t > r,\ s \in (r,t] : -\log \eta_{r,s,t} \leqslant \log \frac{n_{r:t}+1}{\sqrt{n_{r:s-1} \times n_{s:t}}} + \frac{9}{4} + \alpha \log \frac{2\log(\alpha n_{r:s-1})\log(n_{r:s-1})}{\log^2(\alpha)\delta} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad + \alpha \log \frac{2n_{r:t}\log^2(n_{r:t})\log(\alpha n_{s:t})\log(n_{s:t})}{\log(2)\log^2(\alpha)\delta}\Bigg\}
\end{aligned}$$

(b) holds by using Lemma 4, (c) holds thanks to $(n_{r:s-1}+1) \times (n_{s:t}+1) > n_{r:s-1} \times n_{s:t}$, (d) holds true thanks to Equation 24 and (e) holds true thanks to Equation 25.

Step 4: Building the sufficient condition on $\eta_{r,s,t}$ Thus, by using $\exp(-\frac{9}{4}) > \frac{1}{10}$, we get the following condition on $\eta_{r,s,t}$:

$$\begin{aligned}
\eta_{r,s,t} &< \frac{\sqrt{n_{r:s-1} \times n_{s:t}}}{10\left(n_{r:t}+1\right)} \times \left(\frac{\log^2(\alpha)\delta}{2\log(\alpha n_{r:s-1})\log(n_{r:s-1})} \times \frac{\log(2)\log^2(\alpha)\delta}{2n_{r:t}\log^2(n_{r:t})\log(\alpha n_{s:t})\log(n_{s:t})}\right)^\alpha \\
&= \frac{\sqrt{n_{r:s-1} \times n_{s:t}}}{10\left(n_{r:t}+1\right)} \times \left(\frac{\log(4\alpha)\log(2)\delta^2}{4n_{r:t}\log(\alpha n_{r:t})\log^2(n_{r:t})\log(n_{r:t})}\right)^\alpha \\
&= \frac{\sqrt{n_{r:s-1} \times n_{s:t}}}{10\left(n_{r:t}+1\right)} \times \left(\frac{\log(4\alpha+2)\delta^2}{4n_{r:t}\log((\alpha+3)\,n_{r:t})}\right)^\alpha,
\end{aligned}$$

which allows us to get the following control:

$$\boxed{\mathbb{P}_\theta\Big\{\exists t > r, s \in (r,t] : \vartheta_{r,s,t} \geqslant \vartheta_{r,r,t}\Big\} \leqslant \delta.}$$

□

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 $\Phi\left(\Sigma_{s:t}\right)$ to $\Phi\left(\widehat{\mu}_{s:t}\right)$ such that:

$$\forall\,(s,t): \quad \Phi\left(\Sigma_{s:t}\right) + \Phi\left(\bar{\Sigma}_{s:t}\right) - \Phi\left(n_{s:t}\right) = n_{s:t}\left(\Phi\left(\widehat{\mu}_{s:t}\right) + \Phi\left(1-\widehat{\mu}_{s:t}\right)\right).$$

Then, observe that :

$$\begin{aligned} &n_{r:s-1}\left(\Phi\left(\widehat{\mu}_{r:s-1}\right) + \Phi\left(1-\widehat{\mu}_{r:s-1}\right)\right) + n_{s:t}\left(\Phi\left(\widehat{\mu}_{s:t}\right) + \Phi\left(1-\widehat{\mu}_{s:t}\right)\right) \\ &- n_{r:t}\left(\Phi\left(\widehat{\mu}_{r:t}\right) + \Phi\left(1-\widehat{\mu}_{r:t}\right)\right) = n_{r:s-1}\boldsymbol{kl}\left(\widehat{\mu}_{r:s-1}, \widehat{\mu}_{r:t}\right) + n_{s:t}\boldsymbol{kl}\left(\widehat{\mu}_{s:t}, \widehat{\mu}_{r:t}\right). \end{aligned} \tag{26}$$

Finally, observe that:

$$n_{r:s-1}\left(\widehat{\mu}_{r:s-1} - \widehat{\mu}_{r:t}\right)^2 + n_{s:t}\left(\widehat{\mu}_{s:t} - \widehat{\mu}_{r:t}\right)^2 = \frac{n_{r:s-1}n_{s:t}}{n_{r:t}}\left(\widehat{\mu}_{r:s-1} - \widehat{\mu}_{s:t}\right)^2. \tag{27}$$

Then, we will also need a useful notation as $f_{r,s,t}$ (which comes directly from Lemma **??**):

$$f_{r,s,t} = \log\left(n_{r:s-1}+1\right) + \log\left(n_{s:t}+1\right) - \frac{1}{2}\log n_{r:t} + \frac{9}{8}.$$

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

$$\forall\; s \in [r:t) \quad |\widehat{\mu}_{r:s-1} - \widehat{\mu}_{s:t}| \geqslant \Delta_{r,s,t} - \mathcal{C}_{r,s,t,\delta}, \tag{28}$$

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

$$\Delta_{r,s,t} = \left|\mathbb{E}\left[\widehat{\mu}_{r:s-1} - \widehat{\mu}_{s:t}\right]\right| = \begin{cases} \frac{n_{\tau_c:t}}{n_{s:t}}\left|\theta_1 - \theta_2\right| = \frac{n_{\tau_c:t}}{n_{s:t}}\Delta & \text{if } s < \tau_c \leqslant t, \\ \frac{n_{r:\tau_c-1}}{n_{r:s-1}}\left|\theta_1 - \theta_2\right| = \frac{n_{r:\tau_c-1}}{n_{r:s-1}}\Delta & \text{if } \tau_c \leqslant s \leqslant t. \end{cases} \tag{29}$$

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

$$\mathbb{P}\Big\{\exists t > \tau_c : \texttt{Restart}_{r:t} = 1\Big\} > 1-\delta.$$

Notice that:

$$\left\{\forall\, t > \tau_c : \texttt{\textbf{Restart}}_{r:t} = 0\right\} \Leftrightarrow \left\{\forall\, t > \tau_c, \forall s \in (r,t] : \log\vartheta_{r,s,t} \leqslant \log\vartheta_{r,r,t}\right\}.$$
$$\Leftrightarrow \left\{\forall\, t > \tau_c, \forall s \in (r,t] : \log\eta_{r,s,t} \leqslant \widehat{L}_{r:s-1} + \widehat{L}_{s:t} - \widehat{L}_{r:t}\right\}.$$
$$\overset{(a)}{\Rightarrow} \Big\{\forall\, t > \tau_c, \forall s \in (r,t] : \log\eta_{r,s,t} \leqslant f_{r,s,t} + \Phi\left(\Sigma_{r:s-1}\right) + \Phi\left(\bar{\Sigma}_{r:s-1}\right) - \Phi\left(n_{r:s-1}\right) + \Phi\left(\Sigma_{s:t}\right) + \Phi\left(\bar{\Sigma}_{s:t}\right).$$
$$- \Phi\left(n_{s:t}\right) - \Phi\left(\Sigma_{r:t}\right) - \Phi\left(\bar{\Sigma}_{r:t}\right) + \Phi\left(n_{r:t}\right)\Big\}.$$
$$\overset{(b)}{\Rightarrow} \left\{\forall\, t > \tau_c, \forall s \in (r,t] : \log\eta_{r,s,t} \leqslant f_{r,s,t} - n_{r:s-1}\boldsymbol{kl}\left(\widehat{\mu}_{r:s-1}, \widehat{\mu}_{r:t}\right) - n_{s:t}\boldsymbol{kl}\left(\widehat{\mu}_{s:t}, \widehat{\mu}_{r:t}\right)\right\}.$$
$$\overset{(c)}{\Rightarrow} \left\{\forall\, t > \tau_c, \forall s \in (r,t] : \log\eta_{r,s,t} \leqslant f_{r,s,t} - 2n_{r:s-1}\left(\widehat{\mu}_{r:s-1} - \widehat{\mu}_{r:t}\right)^2 - 2n_{s:t}\left(\widehat{\mu}_{s:t} - \widehat{\mu}_{r:t}\right)^2\right\}.$$
$$\overset{(d)}{\Rightarrow} \left\{\forall\, t > \tau_c, \forall s \in (r,t] : \log\eta_{r,s,t} \leqslant f_{r,s,t} - 2 \times \frac{n_{r:s-1}n_{s:t}}{n_{r:t}}\left(\widehat{\mu}_{r:s-1} - \widehat{\mu}_{s:t}\right)^2\right\}.$$
$$\Rightarrow \left\{\forall\, t > \tau_c, \forall s \in (r,t] : 2 \times \frac{n_{r:s-1}n_{s:t}}{n_{r:t}}\left(\widehat{\mu}_{r:s-1} - \widehat{\mu}_{s:t}\right)^2 \leqslant f_{r,s,t} - \log\eta_{r,s,t}\right\}.$$
$$\overset{(e)}{\Rightarrow} \left\{\forall\, t > \tau_c, \forall s \in (r,t] : \sqrt{\frac{n_{r:s-1}n_{s:t}}{n_{r:t}}}\left|\widehat{\mu}_{r:s-1} - \widehat{\mu}_{s:t}\right| \leqslant \frac{\sqrt{f_{r,s,t} - \log\eta_{r,s,t}}}{\sqrt{2}}\right\}.$$

where (a), holds true thanks to Equation (21), (b) holds true thanks to Equation (26), (c) holds true thanks to the Pinsker Inequality taking the following form: $\forall\left(\theta_1, \theta_2\right) \in [0,1]^2\, \boldsymbol{kl}\left(\theta_1, \theta_2\right) \geqslant 2\left(\theta_1 - \theta_2\right)^2$. (d) holds true thanks to Equation (27) and (e) holds true under the condition that $\eta_{r,s,t} \leqslant \exp\left(f_{r,s,t}\right)$.

Therefore, we obtain:

$$\begin{aligned}
\mathbb{P}\Big\{\forall t > \tau_c : \texttt{\textbf{Restart}}_{r:t} = 0\Big\} &\leqslant \mathbb{P}\Big\{\forall\, t > \tau_c, \forall s \in (r,t] : \sqrt{\frac{n_{r:s-1}n_{s:t}}{n_{r:t}}}\left|\widehat{\mu}_{r:s-1} - \widehat{\mu}_{s:t}\right| \leqslant \frac{\sqrt{f_{r,s,t} - \log\eta_{r,s,t}}}{\sqrt{2}}\Big\} \\
&\overset{(f)}{\leqslant} \delta + \mathbb{P}\Big\{\forall\, t > \tau_c, \forall s \in (r,t] : \sqrt{\frac{n_{r:s-1}n_{s:t}}{n_{r:t}}}\left(\Delta_{r,s,t} - \mathcal{C}_{r,s,t,\delta}\right) \leqslant \frac{\sqrt{f_{r,s,t} - \log\eta_{r,s,t}}}{\sqrt{2}}\Big\} \\
&= \delta + \mathbb{P}\Big\{\forall\, t > \tau_c, \forall s \in (r,t] : \frac{n_{r:s-1}n_{s:t}}{n_{r:t}}\left(\Delta_{r,s,t} - \mathcal{C}_{r,s,t,\delta}\right)^2 \leqslant \frac{f_{r,s,t} - \log\eta_{r,s,t}}{2}\Big\} \\
&= \delta + \mathbb{P}\Big\{\forall\, t > \tau_c, \forall s \in (r,t] : \underbrace{1 - \frac{f_{r,s,t} - \log\eta_{r,s,t}}{2n_{r,s-1} \times \left(\Delta_{r,s,t} - \mathcal{C}_{r,s,t,\delta}\right)^2}}_{A} \leqslant \frac{n_{r:s-1}}{n_{r:t}}\Big\},
\end{aligned} \tag{30}$$

where (f) holds true thanks to Equation (28) (We recall that the relative gap $\Delta_{r,s,t}$ is defined in Equation (29)). 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:

$$\begin{cases} A > 0 & \Leftrightarrow \eta_{r,s,t} > \exp\left(-2n_{r,s-1}\left(\Delta_{r,s,t} - \mathcal{C}_{r,s,t,\delta}\right)^2\right)\exp\left(f_{r,s,t}\right), \\ A < 1 & \Leftrightarrow \eta_{r,s,t} < \exp\left(f_{r,s,t}\right) = \frac{(n_{r:s-1}+1)(n_{s:t}+1)}{\sqrt{n_{r:t}}}\exp\left(\frac{9}{8}\right). \end{cases} \tag{31}$$

The second condition in Equation (31) is always satisfied since, we have:

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

Therefore, from Equation (30) we get the following implication:

$$\Big\{\exists t > \tau_c, s \in (r,t] : 1 + \frac{\log\eta_{r,s,t} - f_{r,s,t}}{2n_{r,s-1}\left(\Delta_{r,s,t} - \mathcal{C}_{r,s,t,\delta}\right)^2} > \frac{n_{r:s-1}}{n_{r:t}}\Big\} \Rightarrow \mathbb{P}\Big\{\exists t > \tau_c : \texttt{\textbf{Restart}}_{r:t} = 1\Big\} > 1 - \delta.$$

In other words, the change-point τ_c is detected at time t (with probability at least $1-\delta$) if for some $s \in (r,t]$, we have:

$$1 + \frac{\log \eta_{r,s,t} - f_{r,s,t}}{2n_{r,s-1} \times \left(\Delta_{r,s,t} - \mathcal{C}_{r,s,t,\delta}\right)^2} > \frac{n_{r:s-1}}{n_{r:t}}. \tag{32}$$

Step 3: Non-asymptotic expression of the detection delay $\mathfrak{D}_{\Delta,r,\tau_c}$ To build the detection delay, we need to ensure the existence of $s \in (r,t]$ such that Equation (32) is satisfied. In particular, Equation (32) can be satisfied for $s = \tau_c$. By this way, a condition to detect the change-point τ_c is written as follows:

$$1 + \frac{\log \eta_{r,\tau_c,t} - f_{r,\tau_c,t}}{2n_{r,\tau_c-1} \times \left(\Delta - \mathcal{C}_{r,\tau_c,t,\delta}\right)^2} > \frac{n_{r:\tau_c-1}}{n_{r:t}}. \tag{33}$$

To build the delay, we should introduce the following variable: $d = t - \tau_c + 1 = n_{\tau_c:t} \in \mathbb{N}^\star$.

Thus from Equation (33), we obtain:

$$1 + \frac{\log \eta_{r,\tau_c,d+\tau_c-1} - f_{r,\tau_c,d+\tau_c-1}}{2n_{r,\tau_c-1}\left(\Delta - \mathcal{C}_{r,\tau_c,d+\tau_c-1,\delta}\right)^2} > \frac{n_{r:\tau_c-1}}{n_{r:\tau_c-1}+d}. \Leftrightarrow d > \frac{\left(1 - \frac{\mathcal{C}_{r,\tau_c,d+\tau_c-1,\delta}}{\Delta}\right)^{-2}}{2\Delta^2} \times \frac{-\log \eta_{r,\tau_c,d+\tau_c-1} + f_{r,\tau_c,d+\tau_c-1}}{1 + \frac{\log \eta_{r,\tau_c,d+\tau_c-1} - f_{r,\tau_c,d+\tau_c-1}}{2n_{r,\tau_c-1}\left(\Delta - \mathcal{C}_{r,\tau_c,d+\tau_c-1,\delta}\right)^2}}.$$

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

$$\mathfrak{D}_{\Delta,r,\tau_c} = \min\left\{ d \in \mathbb{N}^\star : d > \frac{\left(1 - \frac{\mathcal{C}_{r,\tau_c,d+\tau_c-1,\delta}}{\Delta}\right)^{-2}}{2\Delta^2} \times \frac{-\log \eta_{r,\tau_c,d+\tau_c-1} + f_{r,\tau_c,d+\tau_c-1}}{1 + \frac{\log \eta_{r,\tau_c,d+\tau_c-1} - f_{r,\tau_c,d+\tau_c-1}}{2n_{r,\tau_c-1}\left(\Delta - \mathcal{C}_{r,\tau_c,d+\tau_c-1,\delta}\right)^2}} \right\}.$$

□