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 $p(r_t|\mathbf{x}_{1:t}) = \mathbb{P}\{R_t = r_t|\mathbf{X}_{1:t} = \mathbf{x}_{1:t}\}$ denotes the distribution of $R_t$ given the sequence of observations $\mathbf{x}_{1:t}$. ($p(r_t|\mathbf{x}_{1:t})$ is a short hand notation).

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

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

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

\[
\mathbb{P}\{R_t = r_t, \mathbf{X}_{1:t} = \mathbf{x}_{1:t}\} \overset{(a)}{=} \sum_{r_{t-1} \in \mathcal{R}_{t-1}} \mathbb{P}\{R_t = r_t, \mathbf{X}_{1:t} = \mathbf{x}_{1:t}, R_{t-1} = r_{t-1}\}
\]

\[
\overset{(b)}{=} \sum_{r_{t-1} \in \mathcal{R}_{t-1}} \mathbb{P}\{R_t = r_t, \mathbf{X}_t = x_t|R_{t-1} = r_{t-1}, \mathbf{X}_{1:t-1} = \mathbf{x}_{1:t-1}\} \mathbb{P}\{R_{t-1} = r_{t-1}, \mathbf{X}_{1:t-1} = \mathbf{x}_{1:t-1}\}
\]

\[
\overset{(c)}{=} \sum_{r_{t-1} \in \mathcal{R}_{t-1}} \mathbb{P}\{X_t = x_t|R_{t-1} = r_{t-1}, \mathbf{X}_{1:t-1} = \mathbf{x}_{1:t-1}\} \mathbb{P}\{R_t = r_t|R_{t-1} = r_{t-1}, \mathbf{X}_{1:t-1} = \mathbf{x}_{1:t-1}\}
\]

\[
\times \mathbb{P}\{R_{t-1} = r_{t-1}, \mathbf{X}_{1:t-1} = \mathbf{x}_{1:t-1}\}
\]

\[
\overset{(d)}{=} \sum_{r_{t-1} \in \mathcal{R}_{t-1}} \left[\mathbb{P}\{R_t = r_t|R_{t-1} = r_{t-1}\} \mathbb{P}\{X_t = x_t|R_{t-1} = r_{t-1}, \mathbf{X}_{1:t-1} = \mathbf{x}_{1:t-1}\}\right] \frac{p(x_t|r_{t-1}, \mathbf{x}_{1:t-1})}{p(r_t|r_{t-1})} \mathbb{P}\{R_{t-1} = r_{t-1}, \mathbf{X}_{1:t-1} = \mathbf{x}_{1:t-1}\}. \tag{12}
\]

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):

\[
p(r_t|\mathbf{x}_{1:t}) \propto \sum_{r_{t-1} \in \mathcal{R}_{t-1}} p(r_t|r_{t-1}) p(x_t|r_{t-1}, \mathbf{x}_{1:t-1}) p(r_{t-1}|\mathbf{x}_{1:t-1}). \tag{13}
\]

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

1. the underlying predictive model (UPM) $p(x_t|r_{t-1}, \mathbf{x}_{1:t-1})$,
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 (13), we get the following recursive runlength distribution estimation:

- **Growth probability:**

\[
p(r_t = r_{t-1} + 1|\mathbf{x}_{1:t}) \propto p(r_t|r_{t-1}) p(x_t|r_{t-1}, \mathbf{x}_{1:t-1}) p(r_{t-1}|\mathbf{x}_{1:t-1}). \tag{14}
\]
Restarted Bayesian Online Change Point Detector

- Change-point probability:
  \[
  p(\tau = 0|x_{1:t}) \propto \sum_{\tau_{t-1} \in \mathbb{R}_{t-1}} p(\tau_{t-1}) p(x_t|\tau_{t-1}, x_{1:t-1}) p(\tau_{t-1}|x_{1:t-1}).
  \] (15)

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

\[
P\left\{ \tau_t|\tau_{t-1} \right\} = h I\{\tau_t = 0\} + (1 - h) I\{\tau_t = \tau_{t-1} + 1\}.
\] (16)

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

\[
p(\tau = \tau_{t-1} + 1|x_{1:t}) \propto (1 - h) p(x_t|\tau_{t-1}, x_{1:t-1}) \times p(\tau_{t-1}|x_{1:t-1}),
\] (17)

\[
p(\tau = 0|x_{1:t}) \propto h \sum_{\tau_{t-1} \in \mathbb{R}_{t-1}} p(x_t|\tau_{t-1}, x_{1:t-1}) \times p(\tau_{t-1}|x_{1:t-1}).
\] (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:

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

Proof of Lemma 1:

You only need to see that:

\[
V_t = \sum_{s=1}^{t} v_{s,t} \\
= \sum_{s=1}^{t-1} v_{s,t} + v_{t,t} \\
= (1 - h) \sum_{s=1}^{t-1} \exp(-l_{s,t}) v_{s,t-1} + h \sum_{s=1}^{t-1} \exp(-l_{s,t}) v_{s,t-1} \\
= \sum_{s=1}^{t-1} \exp(-l_{s,t}) v_{s,t-1}.
\] □

Proof of Lemma 2:
First, for all \( t \geq 2 \), we have:

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

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

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

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

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

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

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 \( V_1 = 1 \)):

\[
\forall t \geq 4,
V_t = \left( \beta_{t,1} + \sum_{i_1=1}^{t-2} \beta_{t,t-i_1} \beta_{t-i_1,1} + \sum_{i_1=1}^{t-2} \sum_{i_2=i_1+1}^{t-k} \sum_{i_3=i_2+1}^{t-(k-1)} \cdots \sum_{i_{k-1}+1}^{t-2} \beta_{t,t-i_1,t-i_2 \cdots t-i_{k-1},1} \right) V_1
\]

\[
= \beta_{t,1} + \sum_{i_1=1}^{t-2} \beta_{t,t-i_1} \beta_{t-i_1,1} + \sum_{i_1=1}^{t-2} \sum_{i_2=i_1+1}^{t-k} \sum_{i_3=i_2+1}^{t-(k-1)} \cdots \sum_{i_{k-1}+1}^{t-2} \beta_{t,t-i_1,t-i_2 \cdots t-i_{k-1},1}.
\]

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

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

which can be concatenated in the following form:

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

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

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

\[
\square
\]

---

**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:

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

**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 $\mathbb{P}(\cdot|\cdot)$, 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:

$$\prod_{s=1}^{n} \mathbb{P}(x_s|x_{1:s-1}) = \frac{1}{(n+1) \left( \sum_{i=1}^{n} x_i \right)}.$$  \hfill (19)

Then, let us verify that:

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

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:

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

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

$$\begin{align*}
\prod_{s=1}^{n+1} \mathbb{P}(x_s|x_{1:s-1}) &= \frac{1}{(n+1) \left( \sum_{i=1}^{n} x_i + 1 \right)} \times \frac{\sum_{i=1}^{n} x_i + 1}{n + 2} \\
&= \frac{\left( \sum_{i=1}^{n} x_i + 1 \right) \times (\sum_{i=1}^{n} x_i)! \times (\sum_{i=1}^{n} \bar{x}_i)!}{(n+2)(n+1)n!} \\
&= \frac{\left( \sum_{i=1}^{n+1} x_i \right)! \times (\sum_{i=1}^{n+1} \bar{x}_i)!}{(n+2)(n+1)!} \\
&= \frac{1}{(n+2) \left( \sum_{i=1}^{n+1} x_i \right)}.
\end{align*}$$

where (a) holds using the definition of the Binomial operator.
Case 2: $x_{n+1} = 0$. Observe that:

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

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

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

(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:

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

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

$$
\forall n \geq 1, \forall k \in [0, n] \quad \frac{n^n}{k^k (n-k)^{n-k}} \exp \left( b_1 \right) \leq \frac{n^n}{k^k (n-k)^{n-k}} \exp \left( b_1 \right) \leq \frac{n^n}{k^k (n-k)^{n-k}} \exp \left( b_1 \right)
$$

with $b_1 = -\frac{1}{6} - \frac{1}{2} \log (2\pi)$.  \hfill (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:

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

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

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

\[ \]  

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

$$
\Sigma_{s:t} \log \Sigma_{s:t} + \Sigma_{s:t} \log \Sigma_{s:t} = \Sigma_{s:t} \log \theta + \Sigma_{s:t} \log \theta + n_{s:t} \log n_{s:t} + n_{s:t} \theta + n_{s:t} \theta \left( \frac{\Sigma_{s:t}}{n_{s:t}} , \theta \right).
$$

Then, using Equations (21) with Equation (22), we obtain:
for the upper bound of the loss $\hat{L}_{s:t}$

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

where (a) holds by using Equation (22) and (b) holds using the positiveness of the Kullback-Leibler divergence ($\text{KL}(\bullet, \bullet) \geq 0$).

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

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

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" [link](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 $(x_t)$ has $\sigma$-sub Gaussian noise meaning that:

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

Note that the Bernoulli case is a $\sigma$ sub-Gaussianity case where $\sigma = \frac{1}{4}$. Indeed:

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

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

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

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

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

\[
P\left(\exists t, \exists s < t, |\bar{z}_{s+1:t}| \geq \sqrt{\frac{2\sigma^2(t-s+1)}{t-s}} \ln \left(\frac{g(t)\sqrt{1+t-s}}{\delta}\right)\right) = P\left\{\exists t, \exists s < t, B_{s,t} \geq \frac{g(t)}{\delta}\right\}
\]

\[
\leq P\left(\exists t, \max_{s < t} B_{s,t} \geq \frac{g(t)}{\delta}\right)
\]

\[
\leq \delta \mathbb{E}\left[\max_{t} \max_{s < t} B_{s,t} \frac{g(t)}{g(t)}\right]
\]

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

Indeed, for every random stopping time \(\tau > 1\),

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

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

\[
\mathbb{E}[B_{s,t+1} \mid \mathcal{F}_t] = \frac{\sigma}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathbb{E}[B_{s,t+1} \mid \mathcal{F}_t] e^{-\frac{x^2}{2}} d\lambda \leq \frac{\sigma}{\sqrt{2\pi}} \int_{\mathbb{R}} B_{s,t} e^{-\frac{x^2}{2}} d\lambda = B_{s,t}
\]

Thus, since \(\mathbb{I}\{\tau > t\} \in \mathcal{F}_t\), we deduce that

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

\[
= \mathbb{E}\left[\bar{B}_2\right] + \sum_{t=2}^\infty \mathbb{E}[B_{t,t+1} \mid \mathcal{F}_t] + \sum_{t=1}^\infty \sum_{s < t} \left(\frac{1}{g(t+1)} - \frac{1}{g(t)}\right) \mathbb{E}[B_{s,t} \mathbb{I}\{\tau > t\}]_{\geq 0}
\]

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

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

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

\[
\mathbb{E}\left[\max_{s < \tau} B_{s,\tau} \frac{g(\tau)}{g(\tau)}\right] \leq \frac{1}{C \ln(2)}
\]
Plugging-in this in the control of the deviation and choosing $C = 1/\ln(2)$ thus gives

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

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

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

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

\[\Box\]

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

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

(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

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

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

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

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

$$E[M^\lambda_\tau] = E[M^\lambda_\tau] + E\left[ \sum_{t=1}^{\tau-1} M^\lambda_{t+1} - M^\lambda_t \right]$$

$$= 1 + \sum_{t=1}^{\infty} E \left[ (M^\lambda_{t+1} - M^\lambda_t) \mathbb{I}_{\{ \tau > t \}} \right]$$

$$= 1 + \sum_{t=1}^{\infty} E \left[ (E[M^\lambda_{t+1} | \mathcal{F}_t] - M^\lambda_t) \mathbb{I}_{\{ \tau > t \}} \right]$$

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

$$M_t = \frac{1}{\sqrt{2\pi\sigma^{-2}}} \int_\mathbb{R} \exp \left( \lambda S_t - \frac{\lambda^2 \sigma^2 t}{2} - \frac{\lambda^2 \sigma^2}{2} \right) d\lambda$$

$$= \frac{1}{\sqrt{2\pi\sigma^{-2}}} \int_\mathbb{R} \exp \left( - \lambda \sigma \sqrt{\frac{t+1}{2}} \frac{S_t}{\sigma \sqrt{2(t+1)}} + \frac{S_t^2}{2\sigma^2(t+1)} \right) d\lambda$$

$$= \exp \left( \frac{S_t^2}{2\sigma^2(t+1)} \right) \frac{1}{\sqrt{2\pi\sigma^{-2}}} \int_\mathbb{R} \exp \left( - \lambda^2 \sigma^2 t + 1 \right) d\lambda$$

Thus, we deduce that

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

We conclude by applying a simple Markov inequality:

$$\mathbb{P} \left( \tau | \tilde{\mu}_t - \mu | \geq \sigma \sqrt{2(\tau + 1) \ln (\sqrt{\tau + 1} / \delta)} \right) = \mathbb{P} (M_\tau \geq 1/\delta) \leq \mathbb{E} [M_\tau] \delta$$

___

### 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: $\hat{\mu}_t$ denotes the empirical mean over the sequence $x_1, \ldots, x_t \sim \mathcal{B}(\theta)$, then:

$$\forall \delta \in (0, 1), \forall \alpha > 1 \quad \mathbb{P}_\theta \left\{ \forall t \in \mathbb{N}^*: kl(\hat{\mu}_t, \theta) < \frac{\alpha}{t} \log \frac{\log (\alpha t) \log(t)}{\log^2(\alpha) \delta} \right\} \geq 1 - \delta \quad (24)$$

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

$$\forall \delta \in (0, 1), \forall \alpha > 1 \quad \mathbb{P}_\theta \left\{ \forall t \in \mathbb{N}^*, \forall s \in (r, t): kl(\hat{\mu}_{s,t}, \theta) < \frac{\alpha}{n_{s,t}} \log \frac{n_{r,t} \log^2(n_{r,t}) \log(\alpha n_{s,t}) \log(n_{s,t})}{\log(2) \log^2(\alpha) \delta} \right\} \geq 1 - \delta \quad (25)$$

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: $\{\exists t > r, \textbf{Restart}_{r,t} = 1\}$ which is equivalent to the event $\{\exists t > r, \ s \in (r, t): \theta_{r,s,t} \geq \theta_{r,r,t}\}$. 


Step 2: Equivalent events. First, notice that:

\[
\{ \exists t > r, \ s \in (r, t] : \vartheta_{r,s,t} \geq \vartheta_{r,r,t} \} \iff \{ \exists t > r, \ s \in (r, t] : \log \vartheta_{r,s,t} \geq \log \vartheta_{r,r,t} \}.
\]

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

\((b)\) holds by using Lemma 4, \((c)\) holds thanks to Lemma 5.

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

\[
\mathbb{P}_\theta \{ \exists t > r, \ s \in (r, t] : \vartheta_{r,s,t} \geq \vartheta_{r,r,t} \} = \mathbb{P}_\theta \{ \exists t > r, \ s \in (r, t] : -\log \eta_{r,s,t} \leq \hat{L}_{r:t} - \hat{L}_{r:s-1} - \hat{L}_{s:t} \}
\]

\((b)\) holds by using Lemma 4, \((c)\) holds thanks to Equation (10), \((d)\) 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 \(\exp(-\frac{9}{4}) > \frac{1}{17}\), we get the following condition on \(\eta_{r,s,t}\):

\[
\eta_{r,s,t} < \sqrt{n_{r:s-1} \times n_{s:t}} \left( \frac{\log^2(\alpha) \delta}{2 \log(\alpha n_{r:s-1}) \log(n_{r:s-1})} \times \frac{\log(2) \log^2(\alpha) \delta}{2 n_{r:t} \log^2(n_{r:t}) \log(n_{s:t})} \right)^\alpha
\]

\[
= \sqrt{n_{r:s-1} \times n_{s:t}} \left( \frac{\log(4\alpha) \log(2) \delta^2}{4 n_{r:t} \log(\alpha n_{r:t}) \log^2(n_{r:t}) \log(n_{r:t})} \right)^\alpha
\]

\[
= \sqrt{n_{r:s-1} \times n_{s:t}} \left( \frac{\log(4\alpha + 2) \delta^2}{4 n_{r:t} \log((\alpha + 3) n_{r:t})} \right)^\alpha,
\]
which allows us to get the following control:

\[ \mathbb{P}_{\vartheta} \left\{ \exists t > r, s \in (r, t] : \vartheta_{r,s,t} \geq \vartheta_{r,r,t} \right\} \leq \delta. \]

\( \square \)

**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:

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

Then, observe that:

\[
n_{r,s,t} (\Phi (\hat{\mu}_{r:s-1}) + \Phi (1 - \hat{\mu}_{r:s-1})) + n_{s:t} (\Phi (\hat{\mu}_{s:t}) + \Phi (1 - \hat{\mu}_{s:t}))
- n_{r,t} (\Phi (\hat{\mu}_{r:t}) + \Phi (1 - \hat{\mu}_{r:t})) = n_{r,s} k_1 l (\hat{\mu}_{r:s-1}, \hat{\mu}_{r:t}) + n_{s:t} k_1 l (\hat{\mu}_{s:1}, \hat{\mu}_{r:t}).
\]

Finally, observe that:

\[
n_{r,s-1} (\hat{\mu}_{r:s-1} - \hat{\mu}_{r:t})^2 + n_{s:t} (\hat{\mu}_{s:t} - \hat{\mu}_{r:t})^2 = \frac{n_{r,s-1} n_{s:t}}{n_{r:t}} (\hat{\mu}_{r:s-1} - \hat{\mu}_{s:t})^2.
\]

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

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

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

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

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

\[
\Delta_{r,s,t} = |\mathbb{E} [\hat{\mu}_{r:s-1} - \hat{\mu}_{s:t}]| = \begin{cases} \frac{n_{r:t}}{n_{r:s-1}} |\theta_1 - \theta_2| = \frac{n_{r:s-1}}{n_{r:s-1}} \Delta & \text{if } s < \tau_c \leq t, \\ \frac{n_{r:s-1}}{n_{r:s-1}} |\theta_1 - \theta_2| = \frac{n_{r:s-1}}{n_{r:s-1}} \Delta & \text{if } \tau_c \leq s \leq t. \end{cases}
\]

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

\[
\mathbb{P} \{ \exists t > \tau_c : \text{Restart}_{r:t} = 1 \} > 1 - \delta.
\]

Notice that:
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: \( \forall (\theta_1, \theta_2) \in [0, 1]^2 \) \( k(\theta_1, \theta_2) \geq 2 (\theta_1 - \theta_2)^2 \). (d) holds true thanks to Equation (27) and (e) holds true under the condition that \( \eta_{r,s,t} \leq \exp (f_{r,s,t}) \).

Therefore, we obtain:

\[
\Pr \left\{ \forall t > \tau_c : \text{Restart}^{r,t} = 0 \right\} \leq \Pr \left\{ \forall t > \tau_c, \forall s \in (r,t] : \sqrt{\frac{n_{r,s-1} n_{s,t}}{n_{r,t}}} |\tilde{\mu}_{r,s-1} - \tilde{\mu}_{s,t}| \leq \sqrt{f_{r,s,t} - \log \eta_{r,s,t}} \left\} (f) \leq \delta + \Pr \left\{ \forall t > \tau_c, \forall s \in (r,t] : \sqrt{\frac{n_{r,s-1} n_{s,t}}{n_{r,t}}} (\Delta_{r,s,t} - C_{r,s,t,\delta}) \leq \sqrt{f_{r,s,t} - \log \eta_{r,s,t}} \right\} \\
= \delta + \Pr \left\{ \forall t > \tau_c, \forall s \in (r,t] : \frac{n_{r,s-1} n_{s,t}}{n_{r,t}} (\Delta_{r,s,t} - C_{r,s,t,\delta})^2 \leq \frac{f_{r,s,t} - \log \eta_{r,s,t}}{2} \right\} \\
= \delta + \Pr \left\{ \forall t > \tau_c, \forall s \in (r,t] : 1 - \frac{f_{r,s,t} - \log \eta_{r,s,t}}{2 n_{r,s-1} (\Delta_{r,s,t} - C_{r,s,t,\delta})^2} \leq \frac{n_{r,s-1} n_{s,t}}{n_{r,t}} \right\}, (30)
\]

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:

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

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

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

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

\[
\left\{ \exists t > \tau_c, s \in (r,t] : 1 + \frac{\log \eta_{r,s,t} - f_{r,s,t}}{2 n_{r,s-1} (\Delta_{r,s,t} - C_{r,s,t,\delta})^2} > \frac{n_{r,s-1} n_{s,t}}{n_{r,t}} \right\} \Rightarrow \Pr \left\{ \exists t > \tau_c : \text{Restart}^{r,t} = 1 \right\} > 1 - \delta.
\]
In other words, the change-point $\tau_c$ is detected at time $t$ (with probability at least $1 - \delta$) if for some $s \in (r, t]$, we have:

$$1 + \frac{\log \eta_{r,s,t} - f_{r,s,t}}{2n_{r,s-1} \times (\Delta_{r,s,t} - C_{r,s,t,\delta})^2} > \frac{n_{r,s-1}}{n_{r,t}}.$$  \hfill (32)

**Step 3: Non-asymptotic expression of the detection delay** $\mathcal{D}_{\Delta,r,\tau_c}$ 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 $\tau_c$ is written as follows:

$$1 + \frac{\log \eta_{r,\tau_c,d} - f_{r,\tau_c,d}}{2n_{r,\tau_c-1} \times (\Delta - C_{r,\tau_c,t,\delta})^2} > \frac{n_{r,\tau_c-1}}{n_{r,t}}.$$  \hfill (33)

To build the delay, we should introduce the following variable: $d = t - \tau_c + 1 = n_{r,t} \in \mathbb{N}^*$. Thus from Equation (33), we obtain:

$$1 + \frac{\log \eta_{r,\tau_c,d+\tau_c-1} - f_{r,\tau_c,d+\tau_c-1}}{2n_{r,\tau_c-1}(\Delta - C_{r,\tau_c,d+\tau_c-1,\delta})^2} > \frac{n_{r,\tau_c-1}}{n_{r,t}} \Rightarrow d > \left(1 - \frac{C_{r,\tau_c,d+\tau_c-1,\delta}}{\Delta}\right)^{-2} \times \frac{-\log \eta_{r,\tau_c,d+\tau_c-1} + f_{r,\tau_c,d+\tau_c-1}}{1 + \frac{\log \eta_{r,\tau_c,d+\tau_c-1} - f_{r,\tau_c,d+\tau_c-1}}{2n_{r,\tau_c-1}(\Delta - C_{r,\tau_c,d+\tau_c-1,\delta})^2}}.$$

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

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

$\square$