Indian J Pure Appl Math (2024) 55:966–973
https://doi.org/10.1007/s13226-024-00641-0

ORIGINAL RESEARCH

Tracking the mean of a piecewise stationary sequence

Ghurumuruhan Ganesan

Received: 30 April 2023 / Accepted: 4 March 2024 / Published online: 3 July 2024
© The Indian National Science Academy 2024

Abstract In this paper we study the problem of tracking the mean of a piecewise stationary sequence of independent random variables. First we consider the case where the transition times are known and show that a direct running average performs the tracking in short time and with high accuracy. We then use a single valued weighted running average with a tunable parameter for the case when transition times are unknown and establish deviation bounds for the tracking accuracy. Our result has applications in choosing the optimal rewards for the multiarmed bandit scenario.

Keywords Piecewise stationary sequence · mean tracking · weighted running average · multiarmed bandit

Mathematics Subject Classification Primary: 60C05 · 68T10

1 Introduction

The multiarmed bandit problem [2] is important from both theoretical and application perspectives and has implications in many fields including economics, machine and financial markets, to name a few. The basic setup is that there are multiple arms each equipped with a reward sequence and the average reward in each arm is different. The user (bandit) wants to latch on to the best arm in as short time as possible and therefore must balance exploration together with exploitation to minimize its overall regret. There are many known bounds regarding the exploration exploitation tradeoff [5,6] and for a recent survey of both theoretical and computational aspects, we refer to Slivkins [8].

Recently [4], there has been interest in the case when the reward sequence of an arm displays nonstationarity. In this case, a direct running average may not be feasible anymore and it is of interest to establish methods for detecting the best arm in a dynamic manner. Auer et al. [3] propose dynamic tracking for the two armed bandit problem and proceed by comparing the empirical means of both the arms (obtained from sliding window running averages) and declaring that a transition has occurred when the difference is significant. For more details and theoretical bounds on the regret we refer to Lattimore and Szepesvári [7].

In general, to compute dynamic running averages and invoke deviation estimates, we would have to store at least $O(\log t)$ samples for t rounds. In this paper, we describe and analyse a *single* valued recursive estimator that uses only knowledge of the minimum epoch durations and automatically performs tracking when the means change.

Communicated by Apoorva Khare, Rahul Roy.

G. Ganesan (✉)
IISER, Bhopal, India
E-mail: gganesan82@gmail.com

The paper is organized as follows: In Section 2, we describe the problem of tracking the mean of a piecewise stationary sequence and show that an unweighted running average performs well if the transition times are known. We also define a single valued weighted running average for the case when the transition times are known and describe its tracking performance in our main result Theorem 1. Next in Section 3, we prove Theorem 1.

2 Piecewise stationary sequence

For integer $t \geq 1$, let $\{X_j\}_{1\leq j\leq t}$ be positive independent random variables with $0 \leq X_j \leq 1$ and suppose there are deterministic integers $M = M(t)$ and $1 = s_1 < s_2 < \ldots < s_M = t$ such that

$$\mathbb{E}X_j = m(k) \text{ for each } s_k \leq j < s_{k+1} \text{ and } 1 \leq k \leq M-1. \tag{2.1}$$

In other words, the random variables $\{X_j\}$ are not necessarily stationary but enjoy "epochs" of stationarity in the sense of (2.1). Throughout this paper, we consider only first order stationarity, i.e., stationarity in the mean and for concreteness, define $s_k \leq j < s_{k+1}$ to be the k^{th} epoch and $\{s_k\}$ to be *transition times*.

It is of interest to "track" the mean $\{m(k)\}$ as it varies from epoch to epoch, using a *randomly* sampled subsequence of $\{X_j\}$ as follows. Let $\{Z_j\}_{1\leq j\leq t}$ be independent Bernoulli random variables (that are also independent of $\{X_j\}$) with

$$\mathbb{P}(Z_j = 1) = \epsilon_j = 1 - \mathbb{P}(Z_j = 0). \tag{2.2}$$

For $t \geq 1$, let $\mathcal{E}(t) = \{1 \leq j \leq t : Z_j = 1\}$ be the set of all sampling times and let $T(1) < T(2) < \cdots < T(w)$ be the random indices in $\mathcal{E}(t)$. Using $\{X_{T(l)}\}$ we are interested in tracking the mean sequence $\{m(k)\}$.

In the context of armed bandit sequential learning, $\{X_j\}$ denotes the reward sequence of an arm and $\{Z_j\}$ denotes the exploration sequence used to study and gain information regarding the reward sequence. There are usually multiple arms each with its own reward sequence and after every transition, the goal is *track* the arm with the largest reward, i.e., to latch on to the best arm as quickly as possible, using the rewards observed in all arms during the exploration sequence.

If the transition times $\{s_k\}$ are *known*, then we can track $\{m(k)\}$ simply using law of large numbers, provided each epoch in (2.1) is of large enough duration. We use the following deviation estimates to demonstrate our argument.

Lemma 1 (*a*) *Let* $\{W_j\}_{1\leq j\leq r}$ *be independent Bernoulli random variables with*

$$\mathbb{P}(W_j = 1) = 1 - \mathbb{P}(W_j = 0) > 0.$$

Setting $S_r := \sum_{j=1}^{r} W_j$, *we have for* $0 < \epsilon \leq \frac{1}{2}$ *that*

$$\mathbb{P}\left(|S_r - \mathbb{E}S_r| \geq \epsilon \mathbb{E}S_r\right) \leq 2\exp\left(-\frac{\epsilon^2}{4}\mathbb{E}S_r\right). \tag{2.3}$$

(*b*) *Let* $\{U_j\}_{1\leq j\leq r}$ *be positive independent random variables satisfying* $0 \leq U_j \leq 1$ *and set* $V_r := \sum_{j=1}^{r} \lambda_j U_j$ *where* $\lambda_j > 0$ *are positive numbers. For any* $\epsilon > 0$ *we have that*

$$\mathbb{P}\left(|V_r - \mathbb{E}V_r| \geq \epsilon \mathbb{E}V_r\right) \leq 2\exp\left(-\frac{\epsilon^2(\mathbb{E}V_r)^2}{\sum_{j=1}^{r}\lambda_j^2}\right). \tag{2.4}$$

For a proof of (2.3) and (2.4), we refer to Appendix *A* of Alon and Spencer [1].

Indeed, suppose that the transition times $\{s_k\}$ are known and satisfy

$$\min_k (s_{k+1} - s_k) \geq t^{\gamma_0} \text{ and } \min_k m(k) \geq \mu_0, \tag{2.5}$$

for some constants $0 < \gamma_0 \leq 1$ and $\mu_0 > 0$. Set $\epsilon_j = \frac{1}{j^\beta}$ where $0 < \beta < 1$ is a constant. For $0 < \gamma < \gamma_0$, let $\mathcal{E}_{k,\gamma}(t) \subset \mathcal{E}(t)$ be the set of all sampling indices lying between s_k and $s_k + t^\gamma$ and define

$$R_l := \frac{1}{\#\mathcal{E}_{k,\gamma}(t)} \sum_{T(j)\in\mathcal{E}_{k,\gamma}(t)} X_{T(j)} \tag{2.6}$$

for each l satisfying $s_k + t^\gamma \leq T(l) < s_{k+1}$. For $s_k \leq T(l) < s_k + t^\gamma$, we set $R_l = 0$. In words, we collect all observations lying between s_k and $s_k + t^\gamma$ and compute a single average as defined in (2.6), whose right hand side does not depend on l.

Using (2.3) and (2.4), we obtain the deviation estimate for R_l as follows. Since $s_k + t^\gamma < t$ for each k, the expected size of $\mathcal{E}_{k,\gamma}(t)$ is

$$\mathbb{E}\left(\#\mathcal{E}_{k,\gamma}(t)\right) = \sum_{j=s_k}^{s_k+t^\gamma} \frac{1}{j^\beta} \geq \sum_{j=t}^{t+t^\gamma} \frac{1}{j^\beta} \geq \frac{t^\gamma}{t^\beta}. \tag{2.7}$$

where $\#A$ denotes the cardinality of A. Therefore using (2.7) together with (2.3) we get that

$$\mathbb{P}\left(\#\mathcal{E}_{k,\gamma}(t) \geq C_1 t^{\gamma-\beta}\right) \geq 1 - \exp\left(-C_2 t^{\gamma-\beta}\right) \tag{2.8}$$

for some constants $C_1, C_2 > 0$, not depending on k. Similarly $\mathbb{E}(\#\mathcal{E}(t)) = \sum_{j=1}^{t} \frac{1}{j^\beta} \leq \frac{2t^{1-\beta}}{1-\beta}$ and so again using (2.3), we get that

$$\mathbb{P}\left(\#\mathcal{E}(t) \leq \frac{4t^{1-\beta}}{1-\beta}\right) \geq 1 - \exp\left(-C_3 t^{1-\beta}\right) \tag{2.9}$$

for some constant $C_3 > 0$. There are in total $M \leq t$ epochs and so defining

$$E_{nice} := \bigcap_{k=1}^{M-1} \left\{\#\mathcal{E}_{k,\gamma}(t) \geq C_1 t^{\gamma-\beta}\right\} \bigcap \left\{\#\mathcal{E}(t) \leq \frac{4t^{1-\beta}}{1-\beta}\right\}$$

we get from (2.8), (2.9) and the union bound that

$$\mathbb{P}(E_{nice}) \geq 1 - t\exp\left(-C_2 t^{\gamma-\beta}\right) - \exp\left(-C_3 t^{1-\beta}\right). \tag{2.10}$$

Given a realization $\omega \in E_{nice}$, the expected value $\mathbb{E}_\omega(R_l) = m(k)$ and so the Azuma-Hoeffding inequality (2.4) together with the second condition in (2.5) imply that

$$\begin{aligned}\mathbb{P}_\omega\left(|R_l - m(k)| \geq \epsilon m(k)\right) &\leq 2\exp\left(-C_4\epsilon^2 \#\mathcal{E}_{k,\gamma}(t)\right) \\ &\leq 2\exp\left(-C_4 C_1 \epsilon^2 t^{\gamma-\beta}\right)\end{aligned} \tag{2.11}$$

for each $s_k + t^\gamma \leq T(l) < s_{k+1}$ and some constant $C_4 = C_4(\mu_0) > 0$ not depending on k. There are $M \leq t$ epochs and so defining F_{good} to be the intersection of the event that

$$\#\mathcal{E}(t) \leq \frac{4t^{1-\beta}}{1-\beta} \tag{2.12}$$

and the event that

$$|R_l - m(k)| \leq \frac{m(k)}{t^b} \text{ for each } k \text{ and each } s_k + t^\gamma \leq T(l) < s_{k+1}, \tag{2.13}$$

we use (2.9) and (2.11) with $\epsilon = \frac{1}{t^b}$ together with the union bound, to get that F_{good} occurs with probability at least

$$1 - t\exp\left(-C_2 t^{\gamma-\beta}\right) - \exp\left(-C_3 t^{1-\beta}\right) - 2t\exp\left(-C_4 C_1 t^{\gamma(1-\beta)-2b}\right). \tag{2.14}$$

The expression in (2.14) converges to one as $t \to \infty$ provided we fix $b < \frac{\gamma(1-\beta)}{2}$.

The relations (2.12) and (2.13) capture the essential requirements we seek in tracking:
(1) From (2.12), we get that the number of samples used for tracking is much smaller than the overall duration t.
(2) The condition (2.13) states that once our tracking scheme comes close to the mean in an epoch, it stays that way for the rest of the epoch, even when new samples are encountered.

(3) The term t^γ could be interpreted as an upper bound for the time needed to track the mean with an accuracy of $1 - \frac{1}{t^b}$. Since we have chosen $\gamma < \gamma_0$ strictly, the tracking time in any epoch is much smaller than the total epoch duration, by (2.5).

The caveat in the above analysis is that we have crucially used the knowledge of the transition times $\{s_k\}$ for the tracking. This in fact allowed us to construct R_l by collecting only the samples observed between s_k and $s_k + t^\gamma$. In practice, $\{s_k\}$ are unknown and so we are interested in "universal" schemes that do not depend on $\{s_k\}$. In our main result below, we use the following recursive estimator to perform "automatic" tracking. For $l \geq 1$ let $\alpha_l := 1 - \frac{1}{l^\delta}$ where $0 < \delta < 1$ is a constant so that $\alpha_1 = 0$. Setting $Y_0 := 0$, we define the l^{th} weighted running average for $l \geq 1$ as

$$Y_l := \alpha_l Y_{l-1} + (1 - \alpha_l) X_{T(l)}. \tag{2.15}$$

The term δ could be interpreted as the importance or weight given to the past observations in deriving the current average.

We have the following result regarding the tracking performance of Y_l.

Theorem 1 *Suppose (2.5) holds and we set* $\epsilon_j = \frac{1}{j^\beta}$. *For every* $0 < \beta < \gamma < \gamma_0$ *and every* $0 < \delta < \frac{\gamma - \beta}{1 - \beta}$, *there are constants* $b, \theta > 0$ *such that*

$$\mathbb{P}\left(E_{good}\right) \geq 1 - \exp\left(-t^\theta\right) \tag{2.16}$$

where $E_{good} = E_{good}(b, \delta, \gamma)$ *is event that(2.12) and (2.13) hold, with* R_l *replaced by* Y_l.

In words, knowing only γ_0 and fixing the tracking tolerance γ, we can choose our parameters β and δ such that Y_l tracks the mean $\{m(k)\}$ with high probability.

3 Proof of Theorem 1

Let ω be a realization of the sequence $(Z_1, \ldots, Z_t)$ so that $\{T(j)\}_{1 \leq j \leq L}$ are fixed. We let $\mathbb{P}_\omega$ be the conditional distribution of $\mathbb{P}$, given the sequence ω. Also, for $1 \leq k \leq M - 1$ let

$$s_k + t^\gamma \leq T(L_k) < T(L_k + 1) < \cdots < T(U_k) \leq s_{k+1} - 1$$

be the set of random indices in $\mathcal{E}(t)$ lying between $s_k + t^\gamma$ and $s_{k+1} - 1$.

We prove Theorem 1 in three steps. In the first step, we obtain estimates for L_k, U_k and their difference. Using these estimates, we show in the second step that the expected value of the running average is close to the current expected value. Finally, we use the Azuma-Hoeffding inequality (2.4) in Lemma 1, to show that the true running averages are close to their respective expected values. Details follow.

Step 1: If $\mathcal{E}(r_1, r_2) = \{r_1 \leq j \leq r_2 : Z_j = 1\}$ then $\#\mathcal{E}(r_1, r_2) = \sum_{j=r_1}^{r_2} Z_j$ and so

$$\mathbb{E}\left(\#\mathcal{E}(r_1, r_2)\right) = \sum_{j=r_1}^{r_2} \frac{1}{j^\beta} := a(r_1, r_2)$$

where

$$\frac{r_2 - r_1}{r_2^\beta} \leq a(r_1, r_2) \leq 1 + \int_{r_1}^{r_2} \frac{dx}{x^\beta} \leq 1 + \frac{r_2^{1-\beta} - r_1^{1-\beta}}{1 - \beta} \tag{3.1}$$

for all r large. In terms of the notation following (2.2), we have $\mathcal{E}(r) = \mathcal{E}(1, r)$ and from (3.1), we see that $a(1, r) \geq \frac{r-1}{r^\beta} \geq \frac{r^{1-\beta}}{2}$ and also that $a(1, r) \leq 1 + \frac{r^{1-\beta}}{1-\beta} \leq \frac{2r^{1-\beta}}{1-\beta}$ for all r large. Consequently, setting

$$E_{samp}(r) := \left\{ a(1, r)\left(1 - \frac{1}{\log r}\right) \leq \#\mathcal{E}(r) \leq a(1, r)\left(1 + \frac{1}{\log r}\right) \right\}$$

and using the deviation estimate (2.3) with $\epsilon = \frac{1}{\log r}$, we have that

$$\mathbb{P}\left(E_{samp}(r)\right) \geq 1 - \exp\left(-2C_1 \frac{r^{1-\beta}}{(\log r)^2}\right),$$

for some constant $C_1 > 0$. We now define

$$E_{exp} := \bigcap_{r \geq (\log t)^{2/(1-\beta)}} E_{samp}(r)$$

and get from the union bound that

$$\mathbb{P}(E_{exp}) \geq 1 - \sum_{r \geq (\log t)^{2/(1-\beta)}} \exp\left(-2C_1 \frac{r^{1-\beta}}{(\log r)^2}\right) \tag{3.2}$$

for all t large.

For $0 < \epsilon < 1 - \beta$ strictly and all $r \geq (\log t)^{2/(1-\beta)} =: a$ with t large, we have

$$\exp\left(-2C_1 \frac{r^{1-\beta}}{(\log r)^2}\right) \leq \exp\left(-r^{\epsilon}\right)$$

and so from (3.2), we get

$$\mathbb{P}(E_{exp}) \geq 1 - \sum_{r \geq a} \exp\left(-r^{\epsilon}\right) = 1 - \sum_{l \geq 0} \exp\left(-(l+a)^{\epsilon}\right).$$

Using the mean value estimate $(l+a)^{\epsilon} - a^{\epsilon} \geq \epsilon l a^{\epsilon - 1} =: \lambda l$ we further get that

$$\begin{aligned}
\mathbb{P}(E_{exp}) &\geq 1 - \exp\left(-a^{\epsilon}\right) \sum_{l \geq 0} \exp\left(-l\lambda\right) \\
&\geq 1 - \exp\left(-a^{\epsilon}\right) \int_0^{\infty} \exp\left(-\lambda(x-1)\right) dx \\
&= 1 - \frac{1}{\lambda} \exp\left(-a^{\epsilon}\left(1 - \frac{\epsilon}{a}\right)\right) \\
&\geq 1 - \epsilon a^{1-\epsilon} \exp\left(-\frac{a^{\epsilon}}{2}\right)
\end{aligned}$$

since $\epsilon < 1$ and $a > 2$ for all t large. Setting ϵ to be slightly larger than $\frac{3(1-\beta)}{4}$ we then get that

$$\mathbb{P}(E_{exp}) \geq 1 - \exp\left(-(\log t)^{3/2}\right) \tag{3.3}$$

for all t large. We assume henceforth that E_{exp} occurs and this obtains upper and lower bounds for each L_k and U_k.

For a lower bound on the difference $L_{k+1} - U_k$, we argue as in (2.7) to get that

$$\mathbb{E}(L_{k+1} - U_k) \geq \sum_{j=s_{k+1}}^{s_{k+1}+t^{\gamma}} \frac{1}{j^{\beta}} \geq t^{\gamma-\beta}. \tag{3.4}$$

Choosing $\gamma > \beta$ strictly and setting

$$E_{diff} := \bigcap_{k=1}^{M-1} \left\{ L_{k+1} - U_k \geq \frac{t^{\gamma-\beta}}{2} \right\},$$

we get from the deviation estimate (2.3) and union bound that

$$\mathbb{P}(E_{diff}) \geq 1 - t \exp\left(-C_2 t^{\gamma-\beta}\right) \tag{3.5}$$

for some constant $C_2 > 0$, since the number of transitions $M \leq t$. Setting $E_{tot} := E_{exp} \cap E_{diff}$ we get from (3.3), (3.5) and the union bound that

$$\mathbb{P}(E_{tot}) \geq 1 - \exp\left(-(\log t)^{3/2}\right) - t \exp\left(-C_2 t^{\gamma-\beta}\right). \tag{3.6}$$

We assume henceforth that E_{tot} occurs so that for all $r \geq (\log t)^{2/(1-\beta)}$ and t large, we have

$$\frac{r^{1-\beta}}{4} \leq \#\mathcal{E}(r) \leq \frac{4r^{1-\beta}}{1-\beta}, \tag{3.7}$$

by (3.1).

Step 2: Let $m(k)$ be the value of the mean μ_j for $s_k \leq j < s_{k+1}$. In this step, we show that for each realization $\omega \in E_{tot}$, the expected value of the running average $\nu_l := \mathbb{E}_\omega Y_l$ is close to the current mean $m(k)$ for all l satisfying $L_k \leq l \leq U_k$.

Defining $W_j := X_{T(j)}$, we see that each Y_l is a linear sum of $\{W_j\}$ since by expanding the recursion (2.15) we get

$$Y_l = \sum_{j=1}^{l} W_j \theta_{j,l} \tag{3.8}$$

where

$$\theta_{j,l} = (1-\alpha_j) \prod_{i=j+1}^{l} \alpha_i$$

for $1 \leq j \leq l-1$ and $\theta_{l,l} = 1-\alpha_l$. By definition $\alpha_1 = 0$ and $\sum_{j=1}^{l} \theta_{j,l} = 1$ and so for each $L_1 \leq l \leq U_1$ we have that $\nu_l = m(1)$.

Next we consider the case $L_2 \leq l < U_2$ for which we have

$$\begin{aligned} \nu_l &= m(1) \sum_{j=L_0}^{U_1} \theta_{j,l} + m(2) \sum_{j=U_1+1}^{l} \theta_{j,l} \\ &= m(1) \prod_{j=U_1+1}^{l} \theta_{j,l} + m(2) \left(1 - \prod_{j=U_1+1}^{l} \theta_{j,l}\right). \end{aligned} \tag{3.9}$$

Since $l \geq L_2$, we have that

$$\begin{aligned} \prod_{j=U_1+1}^{l} \alpha_j &\leq \prod_{j=U_1+1}^{L_2} \alpha_j \\ &= \prod_{j=U_1+1}^{L_2} \left(1 - \frac{1}{j^\delta}\right) \\ &\leq \exp\left(-\sum_{j=U_1+1}^{L_2} \frac{1}{j^\delta}\right) \\ &\leq \exp\left(-\frac{(L_2 - U_1)}{L_2^\delta}\right). \end{aligned} \tag{3.10}$$

Since $\omega \in E_{exp}$, we get that $L_k \leq \frac{4t^{1-\beta}}{1-\beta}$ for any k and since $\omega \in E_{diff}$ as well, we also know that $L_{k+1} - U_k \geq \frac{t^{\gamma-\beta}}{2}$. Using these estimates in (3.10), we get that

$$\prod_{j=U_1+1}^{l} \alpha_j \leq \exp\left(-C_3 t^{\gamma-\beta-\delta(1-\beta)}\right) \tag{3.11}$$

for some constant $C_3 > 0$. Choosing $0 < \delta < \frac{\gamma-\beta}{1-\beta}$ strictly, we then get from (3.9) that

$$|\nu_l - m(2)| \leq (m(1) + m(2)) \exp\left(-C_3 t^{\gamma-\beta-\delta(1-\beta)}\right)$$

for all $L_2 \leq l \leq U_2$. Continuing this way iteratively, we get that for each $L_k \leq l \leq U_k$ that

$$|\nu_l - m(k)| \leq \exp\left(-C_3 t^{\gamma-\beta-\delta(1-\beta)}\right) \sum_{j=1}^{M} m(j)$$
$$\leq C_4 t \exp\left(-C_3 t^{\gamma-\beta-\delta(1-\beta)}\right) \tag{3.12}$$

for some constant $C_4 > 0$, since $0 \leq X_j \leq 1$ by Theorem statement and so $m(j) \leq 1$.

Step 3: In this step, we use the deviation estimate (2.4) to show that Y_l is close to ν_l. We therefore begin with an upper bound for $\sum_{j=1}^{l} \theta_{j,l}^2$.

By definition $\theta_{l,l}^2 = (1-\alpha_l)^2 = \frac{1}{l^{2\delta}}$ and for $1 \leq j \leq l-1$ we have that

$$\theta_{j,l}^2 = (1-\alpha_j)^2 \prod_{j+1 \leq i \leq l} \alpha_i^2$$
$$= \frac{1}{j^{2\delta}} \prod_{j+1 \leq i \leq l} \left(1 - \frac{1}{i^\delta}\right)^2$$
$$\leq \frac{1}{j^{2\delta}} \exp\left(-2 \sum_{j+1 \leq i \leq l} \frac{1}{i^\delta}\right)$$
$$\leq \frac{1}{j^{2\delta}} \exp\left(-\frac{2(l-j)}{l^\delta}\right).$$

If $j \leq l - l^\delta (\log l)^2$, then $\theta_{j,l}^2 \leq e^{-2(\log l)^2}$ and if $j \geq l - l^\delta (\log l)^2 > \frac{l}{2}$, then $\theta_{j,l}^2 \leq \frac{4}{l^{2\delta}}$. Combining we then get that

$$\sum_{j=1}^{l} \theta_{j,l}^2 \leq l e^{-2(\log l)^2} + l^\delta (\log l)^2 \cdot \frac{4}{l^{2\delta}} \leq \frac{5(\log l)^2}{l^\delta} \tag{3.13}$$

for all $l \geq L_0$, a large constant.

Plugging (3.13) into (2.4) in Lemma 1 and setting $\epsilon = \frac{1}{t^b}$ and using the fact that $m(k) \geq \mu_0$ a constant, we then get that

$$\mathbb{P}_\omega\left(|Y_l - m(k)| \geq \frac{m(k)}{t^b}\right) \leq 2\exp\left(-\frac{5\mu_0^2 l^\delta}{t^{2b}(\log l)^2}\right). \tag{3.14}$$

The estimate (3.14) is for a single epoch and there are $M \leq t$ epochs. Moreover, recalling that $\omega \in E_{tot}$, we get from (3.7) that

$$\frac{t^{\gamma(1-\beta)}}{4} \leq L_1 \leq l \leq \frac{4t^{1-\beta}}{1-\beta} \leq t.$$

Therefore from (3.14) and the union bound we get that

$$\mathbb{P}_\omega\left(E_{good}\right) \geq 1 - 2t\exp\left(-\frac{5\mu_0^2 t^{\delta\gamma(1-\beta)-2b}}{(\log t)^2}\right). \tag{3.15}$$

We choose b small so that $b < \frac{\delta\gamma(1-\beta)}{2}$. Averaging over $\omega \in E_{tot}$ and using (3.6), we then get (2.16). This completes the proof of the Theorem. □

Acknowledgements I thank Professors Rahul Roy, Thomas Mountford, C. R. Subramanian and the referee for crucial comments that led to an improvement of the paper. I also thank IMSc and IISER Bhopal for my fellowships.

Data availability Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article. No funding was received to assist with the preparation of this manuscript.

References

1. N. Alon and J. Spencer. (2008). *The Probabilistic Method*. Wiley Interscience.
2. P. Auer, N.C-Bianchi and P. Fischer. (2002). Finite Time Analysis of the Multiarmed Bandit Problem. *Machine Learning*, **47**, 235–256.
3. P. Auer, P. Gajane and R. Ortner. (2018). Adaptively Tracking the Best Arm with an Unknown Number of Distribution Changes. *European Workshop on Reinforcement Learning*, **14**, 1–8.
4. O. Besbes, Y. Gur and A. Zeevi. (2014). Stochastic multi-armed-bandit problem with nonstationary rewards. *Advances in Neural Information Processing Systems*, **27**, 199–207.
5. A. N. Burnetas and M. N. Katehakis. (1996). Optimal Adaptive Policies for Sequential Allocation Problems. *Advances in Applied Mathematics*, **17**, 122–142.
6. T. Lai and H. Robbins. (1985). Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, **6**, 4–22.
7. T. Lattimore and C. Szepesvári. (2020). *Bandit Algorithms*. Cambridge University Press.
8. A. Slivkins. (2019). Introduction to Multi-armed Bandits. *Foundations and Trends in Machine Learning*.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.