On the convergence of the MLE as an estimator of the learning rate in the Exp3 algorithm

Julien Aubert [1] **Luc Lehéricy** [1] **Patricia Reynaud-Bouret** [1]

Abstract

When fitting the learning data of an individual to algorithm-like learning models, the observations are so dependent and non-stationary that one may wonder what the classical Maximum Likelihood Estimator (MLE) could do, even if it is the usual tool applied to experimental cognition. Our objective in this work is to show that the estimation of the learning rate cannot be efficient if the learning rate is constant in the classical Exp3 (Exponential weights for Exploration and Exploitation) algorithm. Secondly, we show that if the learning rate decreases polynomially with the sample size, then the prediction error and in some cases the estimation error of the MLE satisfy bounds in probability that decrease at a polynomial rate.

1. Introduction

1.1. Context

Imagine that you observe a rat in a maze, learning progressively to find food. How would you guess the learning process (or the algorithm) it actually uses ? This question is of paramount importance in cognitive science where the problem is not to find the fastest or best learning algorithm to learn a specific task but to discover the most realistic learning model (always formulated as an algorithm) (Botvinick et al., 2009).

Many learning algorithms, and in particular all those associated with reinforcement learning, are often used to model real human or animal behavior (Sutton, 1988; Schultz et al., 1997). Realistic models may take into account attentional effects (Gluck & Bower, 1988), differences in reasoning (Mezzadri et al., 2022), or granularity of actions (Botvinick, 2008). Because all these realistic ingredients vary from one individual to another, it is clear that the fitting of a particular model should be done individual by individual and experience by experience (Estes, 1956).

The problem of proving that a certain model (or algorithm) is better suited to model reality than others is so crucial in cognitive science that the methodology for fitting any kind of learning algorithm to real learning data has been well established and emphasized (Wilson & Collins, 2019). Any scientist wishing to develop their own new learning model can follow the same numerical experiments to test whether their model is realistic or not.

The first step in the methodology is to apply the MLE to estimate the model parameters (Daw, 2011; Wilson & Collins, 2019). Recall that we are observing an individual learning a specific task. Therefore, not only the training data ($i.e.$ the observations) strongly depend on each other, but they are also often non stationary (otherwise the individual could not have learned). These data make the study of the MLE very complex from a theoretical point of view.

This is also why extensive simulations are required by (Wilson & Collins, 2019): depending on the set of chosen parameters, not only can a model learn or not learn, but there is also often a set of parameters for which the estimator behaves poorly. For instance in (Mezzadri, 2020)'s PhD thesis, large learning rates imply a too fast learning (if the model learns at all), which in turn prevents the MLE from performing well. Unfortunately, there is a lack of theoretical guarantees on whether it is possible to estimate the parameters of these models consistently. Our goal is to prove rigorously what can be said about the properties of the MLE when fitting a learning algorithm to real data.

1.2. Why Bandits Algorithms ?

This study strongly depends on the algorithm and the experiment. One typical example is the Skinner box experiment, introduced in the 30's, whose goal was to study rodents ability to undergo operant conditioning. Once inside the box, the rat could pull one or more levers and get a reward or a punishment for it. This simplified framework allowed scientists to work in a fully controlled environment. The

[1]Université Côte d'Azur, CNRS, LJAD, France. Correspondence to: Julien Aubert <julien.aubert@univ-cotedazur.fr>, Luc Lehéricy <luc.lehericy@univ-cotedazur.fr>, Patricia Reynaud-Bouret <Patricia.Reynaud-Bouret@univ-cotedazur.fr>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

Skinner box paradigm is still being discussed nowadays (Kang et al., 2021).

Instead of studying a particular cognitive model and in order to work within an established general theoretical framework, we focus on the adversarial multi armed bandit problem. The algorithm we specifically study (Exp3: Exponential weights for Exploration and Exploitation) is probably the simplest algorithm for adversarial bandits (Lattimore & Szepesvári, 2020). Even though it is not used in the cognition literature, it describes well the learning phenomena that occur in the Skinner box, and more generally all stimulus-action associations tasks (e.g. maze T-shaped, see introduction in (Lattimore & Szepesvári, 2020)). It also shares many features with famous cognitive algorithms such as component cue (Gluck & Bower, 1988): credit or loss updating, softmax transformation, etc. Finally, Exp3 has given rise to many variants with a variety of applications (e.g. Exp3-IX, Exp4 in (Lattimore & Szepesvári, 2020)) in more complex settings.

Our purpose here is not to prove any new performance result of Exp3 by minimizing the cumulative loss or regret. Instead, we observe an individual performing a learning task. We assume that it is following the Exp3 algorithm, and ask whether it is possible to infer the parameter of this specific Exp3 instance (that is, its learning rate) from the choices made by the individual. To our knowledge, the present work is the first to tackle this problem. It is why we cannot compare our results to the classical Reinforcement Learning literature: in ML, the classical purpose is to develop effective algorithms, whereas we propose to address the question: "how to estimate parameters of a ML algorithm based on the output of an individual using this algorithm?"

1.3. About Imitation Learning

Our framework could be seen as a particular imitation learning problem (Hussein et al., 2017), where the goal is to imitate the process by which a system learns a new task (and not only the final, calibrated, state of the system). However, this is not how imitation learning is typically devised: usually, the learner imitates a teacher who has already finished learning. In this sense, the input data of a classical imitation algorithm are not learning data.

Moreover, the motivation of the present work is different from imitation learning. Our aim is not to reproduce realistic learning curves, but to estimate the models and parameters that best fit an observed behavior.

1.4. Possible Applications of Learning Rate Estimation

As mentioned above, one of the motivations for this article is that MLE is the popular first step used to fit behavioral models on individual training data (Wilson & Collins, 2019). The aim of our article is to prove that this approach is justified from a mathematical point of view. This is the first step in giving theoretical credibility to a statistical method that has been used for years in behavioral neuroscience.

But another practical application is to use the bandit approach to model the responses of people with mental disorders, see (Bouneffouf et al., 2017) as an example. For example, smokers and non-smokers have been shown to behave differently when faced with a bandit problem (Addicott et al., 2013). Therefore, estimating the learning rates of individuals facing a bandit experience could aid in the early diagnosis of specific mental disorders such as apathy or Alzheimer's disease. Of course, such medical applications should be developed with psychiatrists to assess performance, acceptability, and ethical and societal impact.

1.5. Contributions

In Section 3, we prove in a particular case that trying to estimate constant learning rates leads to a bad estimation whatever the estimation procedure: the estimation error decreases more slowly than logarithmically with the number of observations.

In Section 4, in the framework where the learning rate decreases polynomially with the number of observations, we show a polynomial decrease in the prediction error and in a special case (Section 5) of the estimation error on a truncated MLE.

2. Model and Notation

2.1. Notation

The model of experiment on which the MLE will be fitted is as follows. For n successive iterations, the subject must choose an arm among K possible arms. Let $I_1^n = (I_t)_{1\leq t\leq n}$ be the sequence of observed arms and $(\mathcal{F}_t)_{t\leq n} := (\sigma(\{I_s,\ s\leq t\}))_{t\in n}$ the corresponding filtration.

As in classical cognitive experiments, the losses (or penalties) of the arms denoted by $\pi = (\pi_1, \dots, \pi_K)$ are constant over time. These losses are bounded between 0 and 1 and without loss of generality, we assume that

$$1 \geq \pi_1 \geq \pi_2 \geq \cdots \geq \pi_K \geq 0.$$

2.2. Learning Model: the Exp3 Algorithm

We assume that the subject picks an arm according to Exp3 (Algorithm 1), which in this sense becomes a learning model.

In the most general case, Exp3 is able to cope with losses that are depending on time and it has been proved that it

Algorithm 1 Exp3 (Exponential weights for Exploration and Exploitation for losses)

Let η be a positive real number, called the learning rate.
Let p_1^η be the uniform distribution over $1, \dots, K$.
For each $t = 1, \dots, n$,
• Draw an arm I_t from the probability distribution p_t^η.
• For each arm $i = 1, \dots, K$, compute the estimated loss $\tilde{\pi}_{i,t} = \frac{\pi_i}{p_{i,t}^\eta} \mathbb{1}_{I_t = i}$ and update the estimated cumulative loss $\tilde{L}_{i,t+1} = \tilde{L}_{i,t} + \tilde{\pi}_{i,t}$.
• Update the probability distribution of picking a given arm $p_{t+1}^\eta = (p_{1,t+1}^\eta, \dots, p_{K,t+1}^\eta)$, where for all $i \in \{1, \dots, K\}$

$$p_{i,t+1}^\eta = \frac{\exp(-\eta \tilde{L}_{i,t+1})}{\sum_{k=1}^K \exp(-\eta \tilde{L}_{k,t+1})}. \tag{1}$$

achieves sublinear bounds for the convergence of the pseudo-regret (Bubeck & Cesa-Bianchi, 2012).

However, for the purpose of the present work, we have simplified the set-up. Even if the algorithm is able to cope with time-varying losses, it is observed within a rigid framework planned by the cognitive experiment itself, the most classical framework of which corresponds to fixed losses. In cognitive science, robust and realistic learning processes are often assessed in toy situations where most of the variability is cancelled out.

The parameter η in Exp3 can originally be chosen as a time-varying parameter. Here, we have decided that this quantity is fixed during the experiment, as it is the parameter that the MLE estimates on the basis of n observations.

While the pseudo-regret bound obtained in (Bubeck & Cesa-Bianchi, 2012) is optimal for η of the order of $1/\sqrt{n}$, there are situations where the algorithm can learn even if η is constant (see Section 3) or decreases at a different rate. Therefore from a statistical point of view, it is not clear that the subject uses a constant η or a parameter η that tends to 0 with n. Let us go even further: whether the algorithm learns or not should not be an absolute criterion, since we might have subjects who are unable to learn even after many iterations and who might simply give up.

This is the main difference with other works on Exp3. Here we take Exp3 as a realistic model that can be fitted to data : the range of η that matters to us is the one for which we can guarantee a good quality estimate of the learning rate η.

In the sequel, we denote by $\mathbb{P}^\eta$ the probability when the learning parameter is η, and $\mathbb{E}_\eta[.]$ is the associated expectation. Finally, if f is a function, $f^{(i)}$ denotes its i-th iteration, and f^{-1} denotes its inverse when it exists. Also $O_n(1)$ is sequence that is bounded when n tends to infinity.

3. Tetration Behaviour

In this section, we illustrate the poor performance of the MLE when η is constant. We focus on a particular case where there are only two arms and $\pi_2 = 0$. By studying $KL(\mathbb{P}^\eta_{I_1^n} || \mathbb{P}^\delta_{I_1^n})$, the Kullback-Leibler (KL) divergence between $\mathbb{P}^\eta_{I_1^n}$ and $\mathbb{P}^\delta_{I_1^n}$ (the distributions of the vector of pulled arms I_1^n with the learning rates η and δ respectively), we show that some parameters do not separate well whatever the statistical procedure used.

3.1. A Particular Setup

It is sufficient with only two arms to look only at the time steps at which the worst arm (arm 1) is pulled, that is $T_0 = 0$ and for all $i \geq 0$, $T_{i+1} = \inf\{t \geq T_i + 1 \, , \, I_t = 1\}$.

Indeed, only at that times, is the probability changing: with $q_0^\eta = \frac{1}{2}$ and for all $i \geq 0$,

$$q_{i+1}^\eta = \frac{q_i^\eta e^{-\pi_1 \eta / q_i^\eta}}{(1 - q_i^\eta) + q_i^\eta e^{-\pi_1 \eta / q_i^\eta}}.$$

Note that $p_{1,t}^\eta$ is simply $p_{1,t}^\eta = \sum_{i \geq 0} q_i^\eta \mathbb{1}_{T_i + 1 \leq t \leq T_{i+1}}$.

We show in Lemma A.1 from the Appendix that the q_i^η's are decreasing and tend to 0, and that the increments $T_{i+1} - T_i$ are independent and geometrically distributed with parameter q_i^η.

Therefore when i increases, the distance between the T_i increases, making it more and more difficult to observe an error. This is what we quantify in the next paragraph.

3.2. Behavior of q_i^η

Define $I(\eta)$ and $J(n, \eta)$ as follows.

$$\begin{cases} I(\eta) & := \max\{i \in \mathbb{N} \, , \, q_i^\eta \geq \eta \pi_1\} \\ J(n, \eta) & := \max\{i \in \mathbb{N} \, , \, q_i^\eta \geq \frac{1}{n}\} \end{cases}$$

Note that when η is constant, so is $I(\eta)$.

The following proposition shows how fast q_i^η decreases for $i \geq I(\eta)$.

Proposition 3.1. *Let f be the function defined as $f(x) = e^{\frac{x}{2}}$. Then, for all $k \geq 0$,*

$$q_{I(\eta)+k+1}^\eta \leq \frac{1}{f^{(k)}(2)},$$

with the convention $f^{(0)} = Id$. In particular, $J(n, \eta) \leq I(\eta) + \log^*(n) + 1$, where

$$\log^*(n) = \max\{k \in \mathbb{N} \, , \quad (f^{-1})^{(k)}(n) \geq 2\}.$$

Proof. See section A.1.2 of the Appendix. □

This result shows that, as soon as $i \geq I(\eta)$, q_i^η decreases extremely rapidly down to 0, as fast as a tetration (that is an iterated exponentiation) (see (Knoebel, 1981)). This means reciprocally that the number of indices necessary to pass from $I(\eta)$ to $J(n, \eta)$ is in essence bounded. For instance, with $n = 10^{23}$, $\log^*(n) \leq 6$: even with an unrealistic number of observations, it means that after $I(\eta) + 7$ errors, which is a constant and quite small number, q_i^η becomes as small as 10^{-23}. This essentially means that this quantity becomes null, up to a reasonable computer precision.

Said differently (still with $n = 10^{23}$), the statistical problem is almost equivalent (up to computer precision) to the one consisting in observing a fixed number m of independent geometric variables (with $m \leq I(\eta) + 7$).

3.3. Bounded Kullback-Leibler Divergence

Let $\tau(n) = \max\{k \in \mathbb{N}\,, T_k \leq n\}$. Then,

$$KL(\mathbb{P}^\eta_{I_1^n}||\mathbb{P}^\delta_{I_1^n}) = \mathbb{E}_\eta\Bigg[\sum_{i=0}^{\tau(n)-1}((T_{i+1} - T_i - 1)\log\frac{1-q_i^\eta}{1-q_i^\delta} + \log\frac{q_i^\eta}{q_i^\delta})\Bigg] + \mathbb{E}_\eta\Bigg[(n - T_{\tau(n)})\log\frac{1-q^\eta_{\tau(n)}}{1-q^\delta_{\tau(n)}}\Bigg] \tag{2}$$

The previous proposition leads to the following result.

Theorem 3.2 (Existence of parameters that are hard to distinguish)**.** *Let R such that $0 < R\pi_1 < 1$ and let $\beta > 0$. There exists an integer n_0 depending on R, π_1 and β, a constant $\Delta > 0$ depending on π_1 and β and a constant $c > 0$ depending on π_1 such that for all $n \geq n_0$, there exists*

- $\delta \in [0, R]$ *such that* $\delta \geq \dfrac{c}{\log^*(n)}$,
- $\eta \in [0, R]$ *such that* $\eta = \delta + \dfrac{1}{(\log(n))^{1+\beta}}$

such that

$$KL(\mathbb{P}^\eta_{I_1^n}||\mathbb{P}^\delta_{I_1^n}) \leq \Delta. \tag{3}$$

Proof. See Section A.1.3 of the Appendix. □

This theorem becomes really interesting in the light of Theorem 2.2. of (Tsybakov, 2009) that tells us that if one finds two distributions such that $KL(\mathbb{P}^\eta_{I_1^n}||\mathbb{P}^\delta_{I_1^n}) \leq \Delta$, then any statistical procedure able to distinguish η from δ is doomed to make an error in probability of at least $1/4\exp(-\Delta)$.

This means that, whatever the method, these two parameters cannot be distinguished without a nearly fixed probability of error. This implies that any estimation procedure of η cannot converge at a faster rate than the distance between η and δ, that is a logarithmic rate in $O_n(\log(n)^{-(1+\beta)})$.

Note that due to technicalities, δ may have to tend to 0 in Theorem 3.2 but at a slower rate than $1/\log^*(n)$, which as said previously is almost constant. So in practice this corresponds to the case where the learning parameters are constants.

We think the lower logarithmic bound for $|\eta - \delta|$ is pessimistic and it is quite likely that the phenomenon appear even for $|\eta - \delta| = O_n((\log^*(n))^{-1})$, maybe to some power, but we have not been able to prove it. In any case, this rate is much slower than the polynomial rate we obtain in the next two sections for a decreasing learning rate and the truncated MLE. Moreover, simulations (see Section 6) at least confirm that for a fixed error (say 5%), the estimation error of the MLE is not decreasing as a function of n for a constant learning rate.

4. Decreasing Learning Rate

Let us now turn to the case where the parameter $\eta = \eta_n$ decreases with n, the number of observations. We consider the general case of K arms introduced in Section 2.1.

Allowing η_n to decrease with n is allowing $I(\eta_n)$ to grow with n. Proposition 3.1 intuitively shows that the observations that matter for the estimation are those obtained for $t \leq I(\eta_n)$. After $I(\eta_n)$, the probability of pulling the worst arm is so small that it is negligible from a numerical point of view and almost uninformative from an estimation point of view (see Section 3).

Truncating the observations has therefore a twofold interest: not only are we deleting uninformative data, but we are also removing the values in the log likelihood which could explode at the speed of a tetration and lead to numerical issues (see Section 6 for an illustration).

The question now is where to truncate the observations. From an estimation point of view, it is not possible to choose $I(\eta_n)$ as a truncating value since it depends on the unknown parameter η_n. Instead we introduce a parameter $\varepsilon > 0$, and given the number of observations n, we want to find a stopping term $\Upsilon_n \in \mathbb{N}$ guaranteeing that $p^{\delta_n}_{k,t}$, the probability of choosing an arm k at round t, remains greater than ε for all $t \leq \Upsilon_n$, whatever the choice of δ_n in the set of possible parameters.

4.1. The New Setup and the Truncated Log-likelihood

In the general setting, recall that K is the number of arms and π the sequence of losses. From now on, for some α fixed and known in $(0, 1)$, let $\eta_n = \dfrac{\eta_0}{n^\alpha \pi_1}$. We assume that the unknown parameter η_0 belongs to $\Theta = [r, R]$ with

$0 < r < R$ two known positive constants.

Let $0 < \varepsilon < \frac{1}{K}$ be a fixed known threshold, corresponding roughly in practice to the numerical precision.

The following result gives an absolute bound Υ_n that guarantees that the $p_{k,t}^{\delta_n}$'s are always larger than the threshold ε.

Proposition 4.1. *Let*

$$\Upsilon_n = \Big\lfloor (\frac{1}{K} - \varepsilon)\frac{n^\alpha}{R} \Big\rfloor. \tag{4}$$

Then

$$\forall \delta_0 \in \Theta,\ \forall t \leq \Upsilon_n,\ \forall 1 \leq k \leq K, \quad p_{k,t}^{\delta_n} \geq \varepsilon,$$

with $\delta_n = \dfrac{\delta_0}{n^\alpha \pi_1}$.

Proof. See Section A.2.1 of the Appendix. □

From now on, we define the truncated log likelihood as follows

$$\forall \delta_0 \in \Theta, \quad \ell_{n,\varepsilon}(\delta_0) := \sum_{t=1}^{\Upsilon_n} \sum_{k=1}^{K} \log(p_{k,t}^{\delta_n}) \mathbb{1}_{I_t = k}, \tag{5}$$

with $\delta_n = \dfrac{\delta_0}{n^\alpha \pi_1}$ and Υ_n given by (4).

4.2. Upperbound on the Prediction Error

Let $\eta_0 \in \Theta$ be the true parameter, meaning Exp3 is used by the subject with a learning parameter $\eta_n = \dfrac{\eta_0}{n^\alpha \pi_1}$. Let

$$\widehat{\eta}_0 = \mathrm{argmax}_{\delta_0 \in \Theta}\, \ell_{n,\varepsilon}(\delta_0).$$

The estimator of the learning rate of Exp3 is therefore $\widehat{\eta}_n = \dfrac{\widehat{\eta}_0}{n^\alpha \pi_1}$. The following result shows that the prediction error converges to zero with n.

Theorem 4.2. *For any $t \geq 0$, and any $\delta_0, \delta_0' \in \Theta$, let*

$$\|p_t^{\delta_n} - p_t^{\delta_n'}\|_2^2 = \sum_{k=1}^{K} |p_{k,t}^{\delta_n} - p_{k,t}^{\delta_n'}|^2,$$

with $\delta_n = \frac{\delta_0}{n^\alpha \pi_1}$ *(resp.* $\delta_n' = \frac{\delta_0'}{n^\alpha \pi_1}$*.)*

Then, for all $n \geq (R/\varepsilon^2)^{1/\alpha}$ and $x \geq 0$, with $\mathbb{P}_{\eta_n}$-probability at least $1 - e^{-x}$,

$$\frac{1}{\Upsilon_n} \sum_{t=1}^{\Upsilon_n} \|p_t^{\eta_n} - p_t^{\widehat{\eta}_n}\|_2^2 < \frac{9c}{\varepsilon} \left(\sqrt{\frac{x+1}{\Upsilon_n}} + \frac{x+1}{\Upsilon_n} \right),$$

where c is given by Lemma 4.3 below ($c = 11$ works).

Proof. See Section A.2.5 of the Appendix for a complete proof but a sketch of the proof with the main milestones is given in the next section. □

Note that if we take $\varepsilon = 10^{-15}$ the previous bound is rather pessimistic, but for $\varepsilon = 10^{-2}$ (and K much smaller than 100), we find almost identical Υ_n from a theoretical point of view (see Proposition 4.1). Even in practice (See Section 6), one can show that the first time, Υ_{max}, that the $p_{k,t}^{\eta_n}$'s pass below ε, does not vary much between both choices for ε. Since ε does not impact the estimation per se (only through Υ_n), one can opt for a rather reasonable choice of ε when using this bound. The fact that passing 10^{-2} or passing 10^{-15} happen at almost the same time t can be linked heuristically to the tetration phenomenon (see Proposition 3.1 that applies whatever η and in particular with $\eta = \eta_n$).

This bounds also means that if one stays away from ε, the behavior of the truncated MLE seems slower than the parametric convergence rate (that should be in Υ_n^{-1}), where Υ_n represents the number of observation used in the truncated log-likelihood. Nevertheless the rate is polynomial in $O_n(\Upsilon_n^{-1/2}) = O_n(n^{-\alpha/2})$. We do not know, if this bound is tight or not, but the simulations (see Section 6) seem to confirm this rate.

4.3. Sketch of Proof

1. By definition of $\widehat{\eta}_0$, we have $\ell_{n,\varepsilon}(\widehat{\eta}_0) \geq \ell_{n,\varepsilon}(\eta_0)$. By concentrating each quantity around its compensator in the martingale sense, we get a bound on a random version of a truncated Kullback-Leibler divergence, which is in turn lower bounded by the prediction error.

2. However, the concentration inequality is not straightforward because $\widehat{\eta}_0$ is random and depends on the same sample. Hence we need to control the deviation of a supremum of martingales.

3. To do so, we turn to the work of (Baraud, 2010), which provides a bound on supremum of the type $Z = \sup_{\delta \in \Theta} Y_\delta$ as long as $Y_\delta - Y_{\delta'}$ has a specific Laplace transform that shows a certain regularity w.r.t. a distance on Θ.

4. In particular, to prove this form of regularity of the Laplace transform, the adequate distance should control a distance on the $p_{k,t}^{\eta_n}$'s.

Let us describe now briefly how we do these steps, starting from step 4.

4.3.1. Step 4: Lipschitz character of p_t^δ

Lemma 4.3. *For any $t \geq 0$ and $\delta_0, \delta_0' \in \Theta$, let*

$$\|p_t^{\delta_n} - p_t^{\delta_n'}\|_\infty = \max_{1 \leq k \leq K} |p_{k,t}^{\delta_n} - p_{k,t}^{\delta_n'}|.$$

There exists a numerical constant c (for instance $c = 11$) such that for all $\delta_0, \delta_0' \in \Theta$, $n \geq (R/\varepsilon^2)^{1/\alpha}$ and $1 \leq t \leq \Upsilon_n = (\frac{1}{K} - \varepsilon)\frac{n^\alpha}{R}$,

$$\|p_t^{\delta_n'} - p_t^{\delta_n}\|_\infty \leq c\frac{|\delta_0 - \delta_0'|}{R}. \quad (6)$$

Proof. See Section A.2.2 of the Appendix. □

4.3.2. Step 3: Regularity of the Laplace transform

Let $\delta_0 \in \Theta$ and let

$$\begin{aligned} X_{\delta_n} &= \ell_{n,\varepsilon}(\delta_0) - \sum_{t=1}^{\Upsilon_n}\sum_{k=1}^{K} \log(p_{k,t}^{\delta_n}) p_{k,t}^{\eta_n} \\ &= \sum_{t=1}^{\Upsilon_n}\sum_{k=1}^{K} \log(p_{k,t}^{\delta_n}) \left[\mathbb{1}_{I_t = k} - p_{t,k}^{\eta_n}\right]. \end{aligned}$$

Note that, since the true learning rate parameter is η_n, the quantity X_{δ_n} can be seen as a martingale stopped at time Υ_n, whatever the value of δ_n.

The following proposition is a requirement to apply the work by (Baraud, 2010).

Proposition 4.4. *Let c be the constant defined in Lemma 4.3. For each $\delta_0, \delta_0' \in \Theta^2$, let*

$$d(\delta_0', \delta_0) = \frac{c}{R\varepsilon}|\delta_0 - \delta_0'|.$$

Then, for all $(\delta_0, \delta_0') \in \Theta^2$, $n \geq (R/\varepsilon^2)^{1/\alpha}$ and $|\lambda| < \frac{1}{d(\delta_0, \delta_0')}$,

$$\mathbb{E}_{\eta_n}[e^{\lambda(X_{\delta_n} - X_{\delta_n'})}] \leq \exp\left(\frac{(\lambda d(\delta_0', \delta_0))^2 \Upsilon_n}{2(1 - \lambda d(\delta_0', \delta_0))}\right).$$

Proof. See Section A.2.3 of the Appendix. □

4.3.3. Step 2: Resulting concentration inequality for the supremum

Next, it is sufficient to apply the work by (Baraud, 2010) to $Y_{\delta_0} = X_{\delta_n} - X_{\eta_n}$.

Theorem 4.5 (Bernstein-type inequality)**.** *Let c be the constant from Lemma 4.3. Let $Z_n = \sup_{\delta_0 \in \Theta}(X_{\delta_n} - X_{\eta_n})$. Then, for all $n \geq (R/\varepsilon^2)^{1/\alpha}$ and $x \geq 0$,*

$$\mathbb{P}_{\eta_n}\left[Z_n \geq \frac{18\,c}{\varepsilon}(\sqrt{\Upsilon_n}\sqrt{x+1} + x + 1)\right] \leq e^{-x}, \quad (7)$$

Proof. See Section A.2.4 of the Appendix. □

4.3.4. Some comments about Step 1 and the general methodology

Step 1 is very classic. This kind of methodology to derive non asymptotic results for Maximum Likelihood Estimators has been already used in much more general set-up, in particular by (Spokoiny, 2011). Note, however, that despite the very general nature of Spokoiny's work, it is not easy to apply it here, as the derivatives of the log-likelihood with respect to δ are not direct, due to the recursive nature of $p_{k,t}^{\delta_n}$. Therefore we have been forced to replace this kind of direct approach on the derivative with a less direct one. In particular, our general result stops at the prediction error and we are not able to give bounds directly on the estimation error, $|\widehat{\eta}_0 - \eta_0|$, because a general control on the derivatives of the log-likelihood is missing. We derive bounds on the estimation error only in one specific case which we detail in the next section, because in this case we can find recursive bounds on the derivatives.

5. Estimation Error in a Special Case

We prove an upper bound in large probability on the estimation error $|\eta_0 - \widehat{\eta}_0|$, only in the case when we have only two arms $K = 2$ and the loss of arm 2 is null, that is $\pi_2 = 0$.

5.1. Lower Bound on the Prediction Error

To pass from the control of the prediction error to the control of the estimation error, we need first to find an adequate lower bound.

Proposition 5.1. *Let $N_0 = 0$ and $N_t = \sum_{s=1}^{t} \mathbb{1}_{I_s = 1}$ be the number of times arm 1 is pulled until $t \geq 1$.*

There exists a constant $m_{\pi_1,\varepsilon} > 0$, depending only on π_1 and ε such that, for all $\delta_0, \eta_0 \in \Theta$

$$\sum_{t=1}^{\Upsilon_n} |p_{1,t}^{\delta_n} - p_{1,t}^{\eta_n}|^2 \geq m_{\pi_1,\varepsilon}|\delta_n - \eta_n|^2 \sum_{t=1}^{\Upsilon_n} N_{t-1}^2. \quad (8)$$

Proof. See Section A.3.1 of the Appendix. □

5.2. Upper Bound for the Estimation Error

Thanks to Theorem 4.2 and Proposition 5.1, the following bound on the estimation error holds with large probability.

Proposition 5.2. *For $x \geq 0$ and $n \in \mathbb{N}$, define by*

$$G_n(x) = \frac{2}{5}\sqrt{2(\log 2\Upsilon_n + x)}\Upsilon_n^{\frac{1}{2}} + \log 2\Upsilon_n + x$$

There exists a constant $M_{\pi_1,\varepsilon} > 0$, depending only on π_1 and ε and a positive sequence B_n verifying

$$B_n := \frac{1}{48}\Upsilon_n + O_n(1),$$

such that for all $n \geq (R/\varepsilon^2)^{\frac{1}{\alpha}}$ *and* $x \geq 0$ *such that* $B_n > G_n(x)$*, with* $\mathbb{P}_{\eta_n}$*-probability at least* $1 - 2e^{-x}$*,*

$$|\widehat{\eta}_0 - \eta_0| \leq M_{\pi_1,\varepsilon} \left(\frac{\sqrt{(x+1)\Upsilon_n} + x + 1}{B_n - G_n(x)} \right)^{\frac{1}{2}}.$$

Proof. See Section A.3.2 of the Appendix. □

Proposition 5.2 shows that for large n, $|\eta_0 - \widehat{\eta}_0| = O(\Upsilon_n^{-\frac{1}{4}}) = O(n^{-\frac{\alpha}{4}})$, which means that the truncated MLE converges at a polynomial rate, which is slower than the classic parametric rate, but clearly faster than the logarithmic rate obtained in Section 3 for a constant learning rate. As for the prediction error, we do not know if this rate is tight, but it seems to be confirmed by the simulations (see Section 6).

6. Numerical Illustrations

6.1. Numerical Setup

All the simulations have been conducted with `R`. The simulation of the n iterative subject's choices have been simulated according to an Exp3 algorithm with a given learning parameter η, K arms and losses given by $\pi = (\pi_1, ..., \pi_K)$. Note that if the $p^{\eta}_{k,t}$ are too small, the Exp3 simulation stops because "NA" can be returned when evaluating (1). In all the simulations that are proposed here, thanks to the choices of n and η this has never happened.

6.2. Constant Learning Rate

We choose $\eta = 0.3$. The log-likelihood is given by

$$\forall \delta \in (0.1, 0.8), \quad \ell_n(\delta) := \sum_{t=1}^{n} \sum_{k=1}^{K} \log(p^{\delta}_{k,t}) \mathbb{1}_{I_t = k}. \quad (9)$$

Figure 1 shows the shape of the log-likelihood when $K = 2$ and $\pi = (0.8, 0)$ on one simulation. For some parameters δ, the log-likelihood is computed as `-Inf` because some probabilities become null for the numerical precision (see the right part of the top sub-figure in Figure 1). Because of this repeated `-Inf` value, we have not been able to perform simulations for such large η with more than 2 arms.

For $K = 2$, $\pi = (0.8, 0)$ and $\eta = 0.3$, we maximised the log-likelihood thanks to the function `DEoptim` in `R` inside the interval $(0.1, 0.8)$, with the default parameters and a maxiter value equal to 50. For this set-up, `DEoptim` returns a correct estimator (red point in Figure 1), but the log-likelihood is very flat around the MLE, so an erroneous estimate is likely to occur.

In this setup, we performed 1000 simulations for various sample size between n =500, and n = 30000 (see Figure 2).

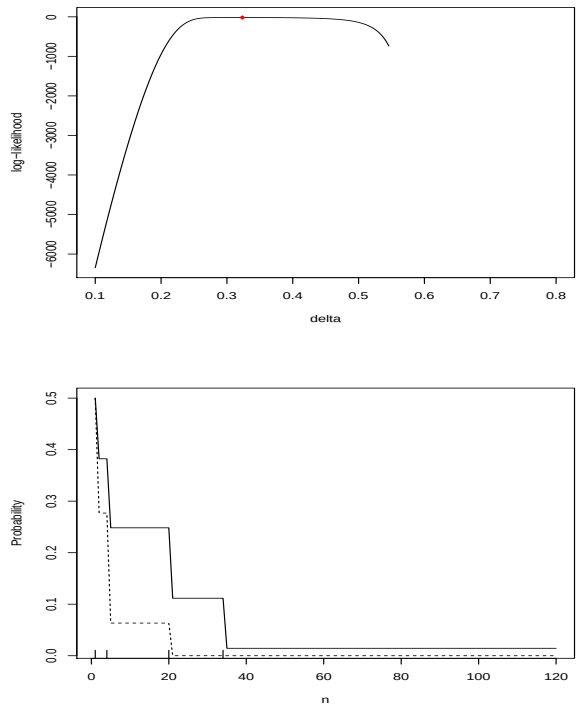

Figure 1. Log-likelihood. On top, the log-likelihood with the MLE in red. The simulation is performed via the Exp3 algorithm with $K = 2$, $\pi = (0.8, 0)$ and $\eta = 0.3$. The value of the log-likelihood at $\delta = 0.6$ is `-Inf`. Below, evolution of $p^{\eta}_{1,t}$ (plain black) and of $p^{\delta}_{1,t}$ (dotted black), with $\delta = 0.6$. After $n = 20$, $p^{\delta}_{1,t}$ is considered as null by the computer.

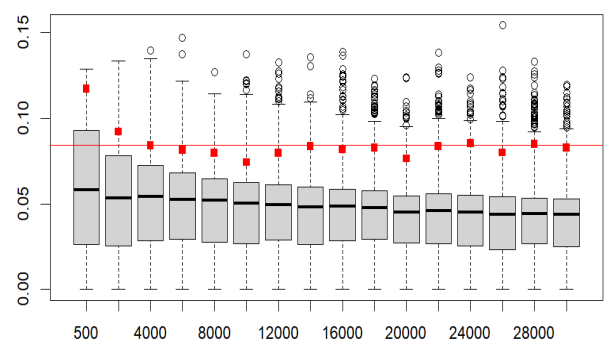

Figure 2. Estimation error when $\eta = 0.3$. In the set-up of Figure 1, boxplot over 1000 simulations of $|\hat{\eta} - \eta|/\eta$. The red squares are the 95% quantiles. The Spearman rank correlation test of decreasing character of these 95% quantiles has a p-value of 0.33. The red horizontal line is the mean of these quantiles.

While the variance of the error decreases with n, the overall trend in the estimation error does not. The Spearman rank correlation test does not detect any decreasing character

of the 95% quantile of the error as n increases. This is consistent with Theorem 3.2 which states there are some δ, η which are indistinguishable regardless of n. The simulations corroborate that the estimation cannot be done satisfactorily when η is constant and even suggest that the logarithmic lower bound of Theorem 3.2 is pessimistic.

6.3. Decreasing Learning Rate and the Stopping Criterion

Here we choose $\alpha = 1/2$, $\eta = \eta_n = \dfrac{0.3}{\pi_1 n^\alpha}$. The truncated log-likelihood is defined by

$$\forall \delta_0 \in (0.1, 0.8),\ \ \ell_n(\delta_n) := \sum_{t=1}^{\Upsilon_{max}} \sum_{k=1}^{K} \log(p_{k,t}^{\delta_n}) \mathbb{1}_{I_t = k}. \tag{10}$$

Following the configuration of Section 4 and to avoid values `-Inf` for the log-likelihood, we stop the log-likelihood, for a given ε, at

$$\Upsilon_{\max} = \sup\{t \in \mathbb{N} : p_{k,t}^{\delta_n} > \varepsilon,\ \ \forall \delta_n \in \texttt{grid}_n, \forall k\}, \tag{11}$$

where $\texttt{grid}_n$ is a regular grid of 50 points in $(\frac{0.1}{\pi_1\sqrt{n}}, \frac{0.8}{\pi_1\sqrt{n}})$ and $1 \le k \le K$.

Note that Υ_{max} is random and larger than the theoretical choice Υ_n of Section 4.

For a fixed n taking various values between $n = 500$ and $n = 30000$, we perform 1000 simulations with $K = 4$ arms and $\pi = (0.8, 0.6, 0.4, 0.2)$ and computed each time the corresponding Υ_{max}. The average is shown on Figure 3. In particular, we see that the bound Υ_n in $\sqrt{n}$ in Proposition 4.1 is a tight bound that reflects well the behavior of the first time to cross ε. We also see that the value of ε has almost no impact on the choice of the truncation inside the log-likelihood, since all curves are almost confounded.

6.4. Performance with Decreasing Learning Rate

We take again $\eta_n = \dfrac{0.3}{\pi_1\sqrt{n}}$. We perform 1000 simulations. The sample size n takes various values between 500 and 30000. We maximize the truncated log-likelihood given by (5) truncated at Υ_{max} given by (11) with $\varepsilon = 10^{-7}$, thanks to the `DEoptim` function of `R`. We consider two cases: $K = 2$ arms with $\pi = (0.8, 0)$ and $K = 4$ arms with $\pi = (0.8, 0.6, 0.4, 0.2)$.

As expected, in both configurations, Figure 4 shows that the prediction error seems to be decreasing in $n^{-1/4}$ (see Theorem 4.2).

The estimation error is illustrated in Figure 5. For $K = 2$, $\pi = (0.8, 0)$, we observe that the error approximately decreases in $n^{-1/8}$ (see Proposition 5.2). For the configuration

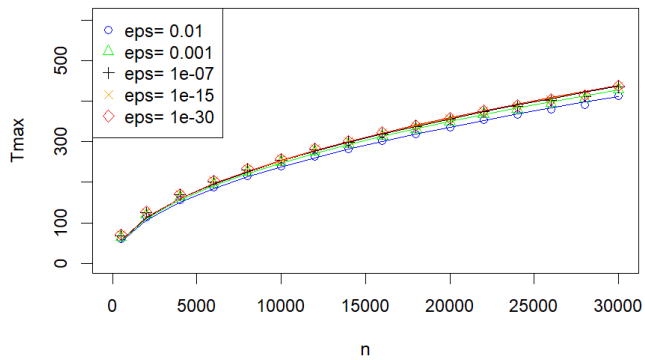

Figure 3. Evolution of $\Upsilon_{\max}$ as a function of n for different ε. The points are the average of the Υ_{max} over 1000 simulations with $K = 4$ arms and $\pi = (0.8, 0.6, 0.4, 0.2)$ and $\eta = \eta_n = 0.3/(\pi_1\sqrt{n})$. The curves for the different values of ε are found by regression of the points on the square root curve.

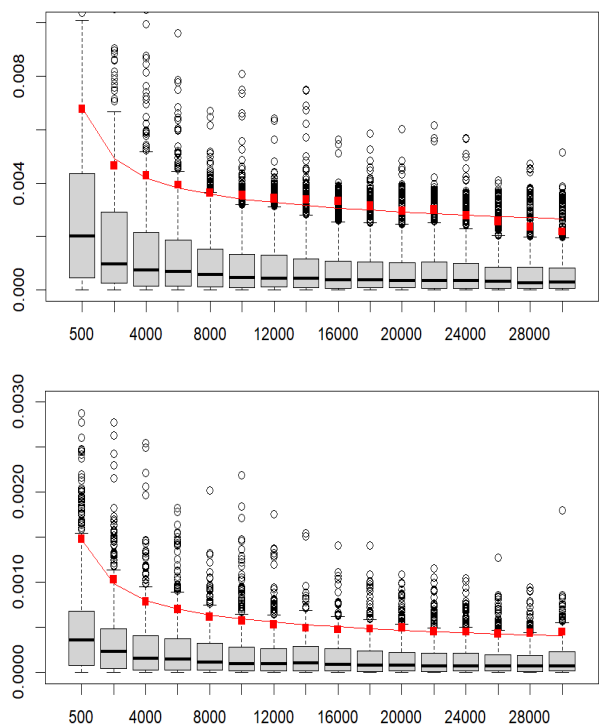

Figure 4. Prediction errors as defined in Theorem 4.2. On top, $K = 2$ and $\pi = (0.8, 0)$. Below, $K = 4$ and $\pi = (0.8, 0.6, 0.4, 0.2)$. 95% quantiles are in red. Spearman rank correlation test detects in both cases that the quantiles decrease with n (on top p-value $= 3.10^{-6}$, below p-value $< 2.10^{-16}$). The red line is obtained by a regression with respect to $n^{-1/4}$.

$K = 4$ and $\pi = (0.8, 0.6, 0.4, 0.2)$, which is not covered by Proposition 5.2, we still observe a decreasing charac-

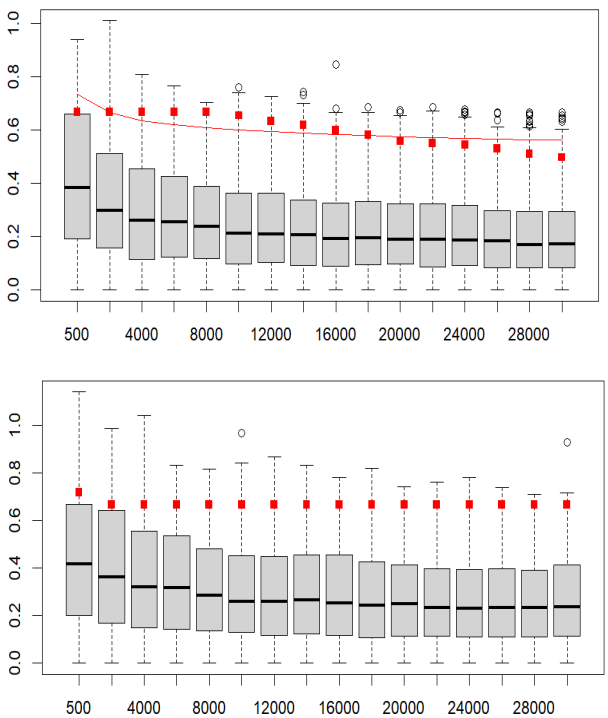

Figure 5. Estimation errors $|\hat{\eta} - \eta|/\eta$. On top, $K = 2$ and $\pi = (0.8, 0)$. Below, $K = 4$ and $\pi = (0.8, 0.6, 0.4, 0.2)$. 95% quantiles are in red. Spearman rank correlation test detects in both cases that the quantiles decrease with n (on top p-value $= 3.10^{-6}$, on the bottom p-value $< 2.10^{-16}$). On top, the red line is obtained by a regression with respect to $n^{-1/8}$.

ter of the 95% quantile as a function of n (p-value of the Spearman rank test is lower than 2.10^{-16}), but it seems to decrease much slower than $n^{-1/8}$.

7. Conclusion

In conclusion, we have shown that the estimation of the learning rate in Exp3 cannot be done correctly if the true learning rate parameter is constant, that is the estimation rate is at most logarithmic. But the MLE on truncated observations can estimate adequately learning rates that are decreasing at a polynomial rate with the number of observations. Note that the rates of convergence that we have shown either for the general prediction error or for the estimation error in particular cases are not the classic parametric rate.

Even if Exp3 is a toy learning model with respect to the vast literature in cognition, we believe that these phenomenons appear in a large variety of models well used in practice such as for instance (Kruschke, 1992) or (Gluck & Bower, 1988) and that our theoretical conclusions should be kept in mind even when working on more realistic models from a more practical point of view.

Acknowledgements

This work was supported by the French government, through the UCA[Jedi] and 3IA Côte d'Azur Investissements d'Avenir managed by the National Research Agency (ANR-15-IDEX-01 and ANR-19-P3IA-0002), by the interdisciplinary Institute for Modeling in Neuroscience and Cognition (NeuroMod) of the Université Côte d'Azur and directly by the National Research Agency (ANR-19-CE40-0024) with the ChaMaNe project. It is part of the explAIn project.

References

Addicott, M., Pearson, J., Wilson, J., Platt, M., and McClernon, F. Smoking and the bandit: a preliminary study of smoker and nonsmoker differences in exploratory behavior measured with a multiarmed bandit task. *Experimental and Clinical Psychopharmacology*, 21(1)(66-73), 2013.

Baraud, Y. A Bernstein-type inequality for suprema of random processes with applications to model selection in non-Gaussian regression. *Bernoulli*, 16(4):1064 – 1085, 2010.

Botvinick, M. Hierarchical models of behavior and prefrontal function. *Trends Cogn. Sci.*, 12(5):201–208, 2008.

Botvinick, M., Niv, Y., and Barto, A. Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition*, 113:262–280, 2009.

Bouneffouf, D., Rish, I., and Cecchi, G. A. Bandit models of human behavior: Reward processing in mental disorders. *CoRR*, abs/1706.02897, 2017. URL `http://arxiv.org/abs/1706.02897`.

Bubeck, S. and Cesa-Bianchi, N. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *CoRR*, abs/1204.5721, 2012.

Daw, N. D. Trial-by-trial data analysis using computational models. *Decision making, affect, and learning: Attention and performance XXIII*, 23(1), 2011.

Estes, W. K. The problem of inference from curves based on group data. *Psychological bulletin*, 53(2):134, 1956.

Gluck, M. and Bower, G. Evaluating an adaptive network model of human learning. *Journal of Memory and Language*, 27:166–195, 04 1988.

Houdré, C. and Reynaud-Bouret, P. Exponential inequalities for u-statistics of order two with constants. *Stochastic Inequalities and Applications. Progress in Probability*, 56, 01 2002.

Hussein, A., Gaber, M. M., Elyan, E., and Jayne, C. Imitation learning: A survey of learning methods. *ACM Computing Surveys (CSUR)*, 50(2):1–35, 2017.

Kang, X., Ri, H., Khalid, M., and Iida, H. Addictive games: Case study on multi-armed bandit game. *Information*, 12 (521), 2021.

Knoebel, R. Exponentials reiterated. *The American Mathematical Monthly*, 88(4):235–252, 1981.

Kruschke, J. K. Alcove: an exemplar-based connectionist model of category learning. *Psychological Review*, 99: 22–44, 1992.

Lattimore, T. and Szepesvári, C. *Bandit Algorithms*. Cambridge University Press, 2020.

Mezzadri, G. *Statistical inference for categorization models and presentation order*. Phd thesis, Université Côte d'Azur, LJAD, France, 2020.

Mezzadri, G., Laloë-Verdelhan, T., F., M., and Reynaud-Bouret, P. Hold-out strategy for selecting learning models: application to categorization subjected to presentation orders. *J. of Math. Psychology*, 2022.

Peel, T., Anthoine, S., and Ralaivola, L. Empirical Bernstein Inequalities for U-Statistics. In *Neural Information Processing Systems (NIPS)*, pp. 1903–1911, Vancouver, Canada, December 2010.

Schultz, W., Dayan, P., and Montague, P. A neural substrate of prediction and reward. *Science*, 275(5306):1593–1599, 1997.

Spokoiny, V. Parametric estimation. finite sample theory. *The Annals of Statistics*, 40, 11 2011.

Sutton, R. S. Learning to predict by the methods of temporal differences. *Machine learning*, 3(1):9–44, 1988.

Tsybakov, A. *Introduction to Nonparametric Estimation*. Springer New-York, 2009.

Wilson, R. C. and Collins, A. G. Ten simple rules for the computational modeling of behavioral data. *Elife*, 8: e49547, 2019.

A. MISSING PROOFS

The Appendix contains the omitted proofs, as well as some indications on the numerical illustrations.

A.1. For a constant learning rate

In this section, we include the proofs and results used for the study of the Kullback-Leibler divergence for a constant learning rate. Technical lemmas used throughout the proofs are gathered in Section A.1.4.

A.1.1. Properties of q_i^η

Lemma A.1. *The following properties hold:*

(i) Let $\eta > 0$. The sequence $(q_i^\eta)_{i\geq 0}$ is decreasing to 0 as $i \to +\infty$.

(ii) For all $i \in \mathbb{N}^*$, the function $\eta \longmapsto q_i^\eta$ is decreasing.

(iii) The sequence $(T_i)_{i\geq 0}$ is non-decreasing. Moreover, $T_i \to +\infty$ as $i \to +\infty$ almost surely.

(iv) Let $\eta > 0$. Under the distribution with parameter η, for all $i \in \mathbb{N}$, the $T_{i+1} - T_i$ are independent and geometrically distributed with parameter q_i^η.

(v) Let $S_n = \sum_{t=1}^n I_t$ be the number of times the worst arm is pulled before time n. Then, $S_n \to +\infty$ as $n \to +\infty$ almost surely.

(vi) Let $\eta > 0$. The sequence $(p_{1,t}^\eta)_{t\in\mathbb{N}^*}$ is decreasing to 0 as $t \to +\infty$ almost surely.

Proof of (i). Let $\eta > 0$. We show that $(q_i^\eta)_{i\geq 0}$ is a decreasing sequence by induction. Denote by g the function defined for all $\gamma \geq 0$ and $q \geq 0$ by

$$\begin{cases} g(\gamma, 0) = 0, \\ g(\gamma, q) = \dfrac{qe^{-\gamma\pi_1/q}}{(1-q) + qe^{-\gamma\pi_1/q}} = \dfrac{1}{(\frac{1}{q}-1)e^{\gamma\pi_1/q} + 1} \in (0,1] \quad \text{when } q > 0. \end{cases} \tag{12}$$

This function is continuous in both its parameters. The denominator is an increasing function of $1/q$, and thus $q \longmapsto g(\gamma, q)$ is an increasing function.

Moreover, $q_0^\eta = \frac{1}{2} > \frac{e^{-2\eta\pi_1}}{1+e^{-2\eta\pi_1}} = q_1^\eta$. Let $i \geq 0$, and assume $q_{i-1}^\eta \geq q_i^\eta$. Since g is an increasing function of q, by the induction hypothesis,

$$q_{i+1}^\eta = g(\eta, q_i^\eta) < g(\eta, q_{i-1}^\eta) = q_i^\eta.$$

Hence $(q_i^\eta)_{i\geq 0}$ is decreasing. Since it is lower bounded by 0, it converges to some $m \in [0, 1/2]$. Since g is continuous in its second parameter, $m = g(\gamma, m)$, that is

$$m = \frac{me^{-\pi_1\eta/m}}{(1-m) + me^{-\pi_1\eta/m}},$$

or equivalently,

$$m(1-m)(1 - e^{-\pi_1\eta/m}) = 0,$$

which admits only one solution in $[0, 1/2]$: $m = 0$. □

Proof of (ii). We show that $\eta \mapsto q_i^\eta$ is a decreasing function by induction on $i \in \mathbb{N}$. Let $\delta < \eta$. Firstly, for $i = 0$, $q_0^\eta = q_0^\delta = 1/2$. Let $i \in \mathbb{N}$, and assume that $q_i^\delta \geq q_i^\eta$. The function g from (12) is decreasing in its first parameter and increasing in its second, therefore

$$q_{i+1}^\eta = g(\eta, q_i^\eta) < g(\delta, q_i^\eta) \leq g(\delta, q_i^\delta) = q_{i+1}^\delta,$$

hence the result by induction. □

Proof of (iii). By definition of T_i, $T_{i+1} \geq T_i$ and $T_i \geq i$, hence the result. □

Proof of (iv). By definition, for all $t \in [T_i + 1, T_{i+1}]$, $p_t^\eta = q_i^\eta$, and therefore, for all $s \geq 1$,

$$\begin{aligned}
\mathbb{P}^\eta(T_{i+1} - T_i = s \,|\, (T_{j+1} - T_j)_{j \neq i}) &\\
&= \sum_{k \geq 1} \mathbb{P}^\eta(I_{k+1}^{k+s-1} = 0, I_{k+s} = 1 \,|\, T_i = k, (T_{j+1} - T_j)_{j \neq i}) \mathbb{P}^\eta(T_i = k \,|\, (T_{j+1} - T_j)_{j \neq i}) \\
&= (1 - q_i^\eta)^{s-1} q_i^\eta \sum_{k \geq 1} \mathbb{P}^\eta(T_i = k \,|\, (T_{j+1} - T_j)_{j \neq i}) \\
&= (1 - q_i^\eta)^{s-1} q_i^\eta,
\end{aligned}$$

which shows that $T_{i+1} - T_i$ is geometrically distributed with parameter q_i^η and independent from the other $T_{j+1} - T_j$. □

Proof of (v). The sequence $(S_n)_{n \geq 1}$ is non-decreasing, so, almost surely, it either tends to $+\infty$ or it converges to some limit $m \in \mathbb{N}$.

Consider the event where $\lim_n S_n = m \in \mathbb{N}$, then on this event, $T_m < +\infty$ and $T_{m+1} = +\infty$, in particular $T_{m+1} - T_m = +\infty$, which is an event with probability zero by (iv) since $q_m^\eta > 0$ for all $m \in \mathbb{N}$. Therefore, $\mathbb{P}^\eta(\lim_n S_n < +\infty) = 0$. □

Proof of (vi). First, note that $p_{1,t} = q_{S_{t-1}}^\eta$, where $(S_n)_{n \geq 0}$ is defined in (v), with the convention $S_0 = 0$. The sequence $(S_n)_{n \geq 1}$ is non-decreasing and tends to $+\infty$ almost surely by (v). Combining this with (i) shows that $(p_{1,t}^\eta)_{t \geq 1}$ is non-increasing and converges to 0 almost surely. □

A.1.2. Proof of Proposition 3.1: tetration behaviour of the updated probabilities.

Let us show by induction that for all $k \geq 1$:

$$q_{I(\eta)+k+1}^\eta \leq \frac{2 \min(\eta \pi_1, q_{I(\eta)+k}^\eta)}{f^{(k)}(2)} \leq \frac{1}{f^{(k)}(2)}.$$

For any $\eta > 0$ and $i \geq 0$, $1 - q_i^\eta + q_i^\eta e^{-\eta \pi_1 / q_i^\eta} \geq \frac{1}{2}$ because $q_i^\eta \leq \frac{1}{2}$. Therefore, $q_{i+1}^\eta \leq 2 q_i^\eta e^{-\eta \pi_1 / q_i^\eta}$. In particular,

$$\begin{aligned}
q_{I(\eta)+2}^\eta &\leq \frac{q_{I(\eta)+1}^\eta e^{-\eta \pi_1 / q_{I(\eta)+1}^\eta}}{1 - q_{I(\eta)+1}^\eta + q_{I(\eta)+1}^\eta e^{-\eta \pi_1 / q_{I(\eta)+1}^\eta}} \\
&\leq 2 q_{I(\eta)+1}^\eta e^{-\eta \pi_1 / q_{I(\eta)+1}^\eta} \\
&\leq 2 \min(\eta \pi_1, q_{I(\eta)+1}^\eta) \cdot e^{-1} = \frac{2 \min(\eta \pi_1, q_{I(\eta)+1}^\eta)}{f(2)} \leq \frac{1}{f(2)}.
\end{aligned}$$

Let $k \geq 2$. Assume the result is true for $k \geq 2$. Then,

$$\begin{aligned}
q_{I(\eta)+k+2}^\eta &\leq 2 q_{I(\eta)+k+1}^\eta e^{-\eta \pi_1 / q_{I(\eta)+k+1}^\eta} \\
&\leq 2 \min(\eta \pi_1, q_{I(\eta)+k+1}^\eta) \cdot e^{-\eta \pi_1 f^{(k)}(2) / (2 \eta \pi_1)} \\
&\leq 2 \min(\eta \pi_1, q_{I(\eta)+k+1}^\eta) \cdot \frac{1}{e^{f^{(k)}(2)/2}} \\
&\leq \frac{2 \min(\eta \pi_1, q_{I(\eta)+k+1}^\eta)}{f^{(k+1)}(2)} \leq \frac{1}{f^{(k+1)}(2)}.
\end{aligned}$$

Let us study the monotonicity of the sequence $(f^{(k)}(2))_{k \geq 0}$: $f^{(0)}(2) = 2 < e = f^{(1)}(2)$. Suppose the result is true for $k \in \mathbb{N}$. Then, since f is increasing,

$$f^{(k)}(2) < f^{(k+1)}(2) \Longrightarrow f^{(k+1)}(2) < f^{(k+2)}(2).$$

Hence, $(f^{(k)}(2))_{k\geq 0}$ is increasing.

For the second part of the Proposition, $\frac{1}{f^{(k)}(2)} < \frac{1}{n}$ implies $q^{\eta}_{I(\eta)+k+1} < \frac{1}{n}$, that is $k + I(\eta) + 1 \geq J(n,\eta) + 1$. Let $\log^*(n)$ be defined as in Proposition 3.1. For all $k \geq 1$,

$$\frac{1}{f^{(k)}(2)} < \frac{1}{n} \Longleftrightarrow n < f^{(k)}(2) \Longleftrightarrow k \geq \log^*(n) + 1,$$

because the sequence $\left(f^{(k)}(2)\right)_{k\geq 1}$ is increasing. In particular, taking $k = \log^*(n) + 1$,

$$J(n,\eta) \leq I(\eta) + \log^*(n) + 1.$$

A.1.3. Proof of Theorem 3.2

The proof is divided into two parts. We first show that the KL divergence can be bounded, up to an additive constant, by a sum whose number of terms is bounded by $J(n,\delta)$, that is:

$$KL(\mathbb{P}^{\eta}_{I_1^n}||\mathbb{P}^{\delta}_{I_1^n}) \leq \sum_{i=0}^{J(n,\delta)} 2\frac{q_i^{\delta} - q_i^{\eta}}{q_i^{\eta}} + n(J(n,\delta)+1)(1 - q^{\eta}_{J(n,\delta)})^{\lfloor \frac{n}{J(n,\delta)+1} \rfloor} + O_n(1).$$

We then show the existence of $\delta > 0$ and $\eta > \delta$, as in the theorem, that verifies that the KL divergence is bounded.

Proof of Theorem 3.2. Let us split the KL divergence into three parts. Assume in the following that $\eta > \delta$, so that $\log \frac{q_i^{\eta}}{q_i^{\delta}} \leq 0$. Recall from Equation (2) that

$$\begin{aligned}
KL(\mathbb{P}^{\eta}_{I_1^n}||\mathbb{P}^{\delta}_{I_1^n}) &= \sum_{i=0}^{n-1} \mathbb{E}_{\eta}[\mathbb{1}_{T_{i+1}\leq n}(T_{i+1} - T_i - 1)\log\frac{1-q_i^{\eta}}{1-q_i^{\delta}} + \mathbb{1}_{T_{i+1}\leq n}\log\frac{q_i^{\eta}}{q_i^{\delta}} \\
&\quad + \mathbb{E}_{\eta}\left[(n - T_{\tau(n)})\log\frac{1-q^{\eta}_{\tau(n)}}{1-q^{\delta}_{\tau(n)}}\right] \\
&= \underbrace{\sum_{i=0}^{J(n,\delta)} \mathbb{E}_{\eta}[\mathbb{1}_{T_{i+1}\leq n}(T_{i+1} - T_i - 1)\log\frac{1-q_i^{\eta}}{1-q_i^{\delta}} + \mathbb{1}_{T_{i+1}\leq n}\log\frac{q_i^{\eta}}{q_i^{\delta}}]}_{A} \\
&\quad + \underbrace{\sum_{i=J(n,\delta)+1}^{n-1} \mathbb{E}_{\eta}[\mathbb{1}_{T_{i+1}\leq n}(T_{i+1} - T_i - 1)\log\frac{1-q_i^{\eta}}{1-q_i^{\delta}} + \mathbb{1}_{T_{i+1}\leq n}\log\frac{q_i^{\eta}}{q_i^{\delta}}]}_{B} \\
&\quad + \underbrace{\mathbb{E}_{\eta}\left[(n - T_{\tau(n)})\log\frac{1-q^{\eta}_{\tau(n)}}{1-q^{\delta}_{\tau(n)}}\right]}_{C}.
\end{aligned}$$

Since $0 \leq q_i^\eta \leq q_i^\delta$ and $\log(1-\cdot)$ is non-increasing and 2-lipschitz on $[0, \frac{1}{2}]$, for all $i \geq 0$, if $\delta\pi_1 n \geq \log 2$ and $n \geq 3$,

$$\begin{aligned}
B &\leq \sum_{i=J(n,\delta)+1}^{n-1} \mathbb{E}_\eta\left[\mathbb{1}_{T_{i+1}\leq n}(T_{i+1}-T_i)\right] \log\frac{1-q_i^\eta}{1-q_i^\delta} \\
&\leq \sum_{i=J(n,\delta)+1}^{n-1} n\mathbb{P}_\eta(T_{i+1}\leq n) \times 2(q_i^\delta - q_i^\eta) \\
&\leq 2n \sum_{i=J(n,\delta)+1}^{n-1} q_i^\delta \\
&\leq 2n\frac{1}{n-2}\sum_{i\geq 0} e^{-\delta\pi_1 n i} \quad \text{by Lemma A.3} \\
&\leq \frac{6}{1-e^{-\delta\pi_1 n}} \leq 12.
\end{aligned}$$

For the same reasons,

$$\begin{aligned}
C &= \mathbb{E}_\eta[(n-T_{\tau(n)})\log\frac{1-q_{\tau(n)}^\eta}{1-q_{\tau(n)}^\delta}] \\
&\leq 2n\mathbb{E}_\eta[q_{\tau(n)}^\delta] \\
&= 2n\mathbb{E}_\eta[q_{\tau(n)}^\delta \mathbb{1}_{\tau(n)\leq J(n,\delta)}] + 2n\mathbb{E}_\eta[q_{\tau(n)}^\delta \mathbb{1}_{\tau(n)\geq J(n,\delta)+1}] \\
&\leq 2n \times \frac{1}{2}\mathbb{P}_\eta(\tau(n)\leq J(n,\delta)) + 2 \quad \text{by definition of } J(n,\delta) \\
&= n\mathbb{P}_\eta(T_{J(n,\delta)+1} > n) + 2 \\
&\leq n(J(n,\delta)+1)(1-q_{J(n,\delta)}^\eta)^{\lfloor \frac{n}{J(n,\delta)+1}\rfloor} + 2 \quad \text{by (18)}
\end{aligned}$$

Finally, provided that $\eta\pi_1 n \geq \log 2$ and $n \geq 3$,

$$\begin{aligned}
A &\leq \sum_{i=0}^{J(n,\delta)} \mathbb{E}_\eta[\mathbb{1}_{T_{i+1}\leq n}(T_{i+1}-T_i-1)]\log\frac{1-q_i^\eta}{1-q_i^\delta} \\
&\leq 2\sum_{i=0}^{J(n,\delta)} (q_i^\delta - q_i^\eta)\left\{\mathbb{E}_\eta[T_{i+1}-T_i] \wedge n\mathbb{P}_\eta(T_{i+1}\leq n)\right\} \\
&\leq 2\sum_{i=0}^{J(n,\delta)} (q_i^\delta - q_i^\eta)\left\{\frac{1}{q_i^\eta} \wedge n^2 q_i^\eta\right\} \quad \text{by (18)} \\
&\leq 2\sum_{i=0}^{(J(n,\eta)+1)\wedge J(n,\delta)} \frac{q_i^\delta - q_i^\eta}{q_i^\eta} + 2\frac{n^2}{n-2}\sum_{i=J(n,\eta)+2}^{+\infty} e^{-\eta\pi_1 n(i-J(n,\eta)-1)} \quad \text{by Lemma A.3} \\
&\leq 2\sum_{i=1}^{(J(n,\eta)+1)\wedge J(n,\delta)} \frac{q_i^\delta - q_i^\eta}{q_i^\eta} + 12ne^{-\eta\pi_1 n}.
\end{aligned}$$

Therefore,

$$KL(\mathbb{P}_{I_1^n}^\eta || \mathbb{P}_{I_1^n}^\delta) \leq 2\sum_{i=1}^{(J(n,\eta)+1)\wedge J(n,\delta)} \frac{q_i^\delta - q_i^\eta}{q_i^\eta} + n(J(n,\delta)+1)(1-q_{J(n,\delta)}^\eta)^{\lfloor \frac{n}{J(n,\delta)+1}\rfloor} + 14 + 12ne^{-\eta\pi_1 n}.$$

From Lemma A.5, by monotonicity of the sequence $(q_k^\delta/q_k^\eta)_{k\geq 0}$,

$$KL(\mathbb{P}^\eta_{I_1^n}||\mathbb{P}^\delta_{I_1^n}) \leq 2J(n,\delta)\frac{q^\delta_{J(n,\delta)} - q^\eta_{J(n,\delta)}}{q^\eta_{J(n,\delta)}} + n(J(n,\delta)+1)(1-q^\eta_{J(n,\delta)})^{\lfloor \frac{n}{J(n,\delta)+1}\rfloor} + 14 + 12ne^{-\eta\pi_1 n}. \quad (13)$$

By Proposition 3.1, for any $\delta > 0$ such that $\delta\pi_1 \geq \frac{1}{n}$,

$$I(\delta) \leq J(n,\delta) \leq I(\delta) + \log^*(n) + 1.$$

Let $R > 0$ be such that $\frac{1}{n} < R\pi_1 \leq 1/2$. By the inequalities above, for any δ such that $\frac{1}{n} < \delta\pi_1 < R\pi_1$,

$$J(n,\delta) - J(n,R) \geq I(\delta) - \Big(I(R) + \log^*(n) + 1\Big).$$

From Lemma A.6, $I(\delta) > I(R) + \log^*(n) + 1$ (and thus $J(n,\delta) > J(n,R)$) as soon as

$$\frac{1}{2\delta\pi_1} - 2 > \frac{2}{R\pi_1} - 4 + \log^*(n) + 1,$$

that is when

$$\delta < \frac{\frac{1}{2\pi_1}}{\frac{2}{R\pi_1} + \log^*(n) - 1}.$$

Therefore, by contraposition, choosing δ such that $J(n,\delta) = J(n,R)$ implies $\delta \geq \frac{\frac{1}{2\pi_1}}{\frac{2}{R\pi_1} + \log^*(n) - 1}$. By continuity of $\gamma \mapsto q^\gamma_{J(n,R)+1}$ and the fact that its limit is $1/2$ (resp. smaller than $1/n$) when $\gamma \to 0$ (resp. R), for all $n \geq 2$, there exists $\delta \in (0,R)$ such that $q^\delta_{J(n,R)+1} \in [\frac{1}{2n}, \frac{1}{n})$. Let δ_n be such a δ. In particular, $J(n,\delta_n) \leq J(n,R)$, thus is equal since $\delta \mapsto J(n,\delta)$ is non-increasing. Therefore,

$$\delta_n \geq \frac{\frac{1}{2\pi_1}}{\frac{2}{R\pi_1} + \log^*(n) - 1} \geq \frac{(4\pi_1)^{-1}}{\log^* n} =: \frac{c}{\log^* n}. \quad (14)$$

for n such that $\log^*(n) \geq \frac{2}{R\pi_1} - 1$. As a consequence, $\delta_n\pi_1 n \geq 1 \vee \log 2$ for all $n \geq n_0$, for some n_0 that depends only on R and π_1. From Lemma A.7, for any $\eta > \delta_n$,

$$q^{\delta_n}_{J(n,\delta_n)} - q^\eta_{J(n,\delta_n)} \leq G(n,\delta_n)(\eta - \delta_n),$$

where

$$G(n,\delta_n) = \frac{8}{\delta_n\pi_1}\frac{\left(\frac{8}{\delta_n\pi_1}\right)^{J(n,\delta_n)} - 1}{\frac{8}{\delta_n\pi_1} - 1} \leq 2\left(\frac{8}{\delta_n\pi_1}\right)^{J(n,\delta_n)}$$

In particular, for $\eta_n = \delta_n + \frac{1}{(\log n)^{1+\beta}}$,

$$q^{\delta_n}_{J(n,\delta_n)} - q^{\eta_n}_{J(n,\delta_n)} \leq \frac{G(n,\delta_n)}{(\log n)^{1+\beta}}. \quad (15)$$

Note that $\frac{G(n,\delta_n)}{(\log n)^\alpha} \to 0$ as $n \to +\infty$ for any $\alpha > 0$. Indeed, by Proposition 3.1 and Lemma A.6,

$$J(n,\delta_n) + 1 \leq I(\delta_n) + \log^*(n) + 2 \leq \frac{2}{\delta_n\pi_1} - 4 + \log^*(n) + 2 \leq 9\log^*(n) \quad (16)$$

by (14), as long as n is such that $\log^*(n) \geq \frac{2}{R\pi_1} - 1$, which also entails $\frac{1}{\delta_n\pi_1} \leq 4\log^*(n)$, so that $\log G(n,\delta_n) = O(\log^*(n)\log\log^*(n)) = o(\log\log n)$.

Therefore, since $q^{\delta_n}_{J(n,\delta_n)} \geq \frac{\delta\pi_1}{\log(2n)}$ by Lemma A.4 and the assumption $q^{\delta_n}_{J(n,\delta_n)+1} \geq 1/(2n)$,

$$q^{\eta_n}_{J(n,\delta_n)} \geq \frac{\delta_n \pi_1}{\log(2n)} - \frac{G(n,\delta_n)}{(\log n)^{1+\beta}} \geq \frac{1}{8\log^*(n)\log n} \tag{17}$$

for $n \geq n_0$, for some n_0 depending on R, π_1 and β. Therefore, by (16),

$$\begin{aligned}
n(J(n,\delta_n)+1)(1-q^{\eta_n}_{J(n,\delta_n)})^{\lfloor \frac{n}{J(n,\delta_n)+1} \rfloor} &\leq n(J(n,\delta_n)+1)\left(1 - \frac{1}{8\log^*(n)\log n}\right)^{\lfloor \frac{n}{J(n,\delta_n)+1} \rfloor} \\
&\leq n(J(n,\delta_n)+1)\exp\left(-\lfloor \frac{n}{J(n,\delta_n)+1} \rfloor \frac{1}{8\log^*(n)\log n}\right) \\
&\leq n \times 9\log^*(n)\exp\left(-\lfloor \frac{n}{9\log^*(n)} \rfloor \frac{1}{8\log^*(n)\log n}\right) \\
&= o_n(1),
\end{aligned}$$

where we used $\log(1-x) \leq -x$ for $x \in (0,1)$. Injecting this result in (13), and since $\eta_n \pi_1 n \geq \log n$ for $n \geq n_0'$ for some n_0' that depends only on π_1 and β,

$$KL(\mathbb{P}^{\eta_n}_{I_1^n} || \mathbb{P}^{\delta_n}_{I_1^n}) \leq 2J(n,\delta)\frac{q^{\delta_n}_{J(n,\delta_n)} - q^{\eta_n}_{J(n,\delta_n)}}{q^{\eta_n}_{J(n,\delta_n)}} + 14 + 12ne^{-\pi_1 n/(\log n)^{1+\beta}} + o_n(1).$$

Using (15) and (17), for n large enough,

$$\begin{aligned}
KL(\mathbb{P}^{\eta_n}_{I_1^n} || \mathbb{P}^{\delta_n}_{I_1^n}) &\leq 2J(n,\delta_n)\frac{\frac{G(n,\delta_n)}{(\log n)^{1+\beta}}}{\frac{1}{8\log^*(n)\log n}} + 14 + o_n(1). \\
&\leq 18\log^*(n)\frac{8\log^*(n)G(n,\delta_n)}{(\log n)^{\beta}} + 14 + o_n(1) \\
&= 14 + o_n(1)
\end{aligned}$$

since $G(n,\delta_n) \vee \log^*(n) = o_n((\log n)^\alpha)$ for any $\alpha > 0$. Note that this $o_n(1)$ only depends on π_1 and β. Therefore, there exists $\Delta_{\beta,\pi_1} > 0$ depending on π_1 and β and $n_0 \in \mathbb{N}$ depending on π_1, R and β, such that for all $n \geq n_0$,

$$KL(\mathbb{P}^{\eta_n}_{I_1^n} || \mathbb{P}^{\delta_n}_{I_1^n}) \leq \Delta_{\beta,\pi_1}.$$

□

A.1.4. Technical lemmas

Lemma A.2 (Bracketing the c.d.f. of T_i)**.** *Let $\eta > 0$. For all $i \in \mathbb{N}$,*

$$1 - i(1-q^\eta_{i-1})^{\lfloor \frac{n}{i} \rfloor} \leq \mathbb{P}^\eta(T_i \leq n) \leq 1 - (1-q^\eta_{i-1})^n \leq nq^\eta_{i-1}. \tag{18}$$

Proof. Since $T_i = T_i - T_0 = \sum_{k=0}^{i-1}(T_{k+1} - T_k) \geq T_i - T_{i-1}$,

$$\begin{aligned}
\mathbb{P}^\eta(T_i \leq n) &= \mathbb{P}^\eta(\sum_{k=0}^{i-1} T_{k+1} - T_k \leq n) \\
&\leq \mathbb{P}^\eta(T_i - T_{i-1} \leq n) \\
&= 1 - (1-q^\eta_{i-1})^n \quad \text{by Proposition A.1 (iv)} \\
&\leq 1 - (1 - nq^\eta_{i-1}) = nq^\eta_{i-1} \quad \text{by Bernoulli's inequality.}
\end{aligned}$$

On the other hand,

$$\mathbb{P}^\eta(T_i \le n) \ge \mathbb{P}^\eta(\forall k \in \{0,\dots,i-1\}, T_k - T_{k-1} \le \lfloor \frac{n}{i} \rfloor) \ge 1 - \sum_{k=0}^{i-1} (1-q_{k-1}^\eta)^{\lfloor \frac{n}{i} \rfloor} \quad \text{by union bound}$$
$$\ge 1 - i(1-q_{i-1}^\eta)^{\lfloor \frac{n}{i} \rfloor},$$

which concludes the proof. □

Lemma A.3. *For all $n \ge 3$ such that $\eta\pi_1 n \ge \log 2$, for all $i \ge J(n,\eta)+1$,*

$$q_i^\eta \le \frac{1}{n-2} e^{-\eta\pi_1 n(i-J(n,\eta)-1)}.$$

In particular, under the same conditions on n and η, for all $i \ge J(n,\eta)+2$,

$$q_i^\eta \le \frac{e^{-\eta\pi_1 n}}{n-2} \le e^{-\eta\pi_1 n}.$$

Proof. Write $h : (x,y) \in (0,1) \times \mathbb{R}_+ \longmapsto \frac{xe^{-y}}{1-x}$, so that $q_{i+1}^\eta \le h(q_i^\eta, \frac{\eta\pi_1}{q_i^\eta})$. This function h is increasing in x and decreasing in y.

Now, let $k \ge 1$, and assume that there exists $0 \le c_k < n$ such that $q_{J(n,\eta)+k}^\eta \le \frac{1}{n-c_k} e^{-\eta\pi_1 n(k-1)}$. By definition of $J(n,\eta)$, this holds for $k=1$ with $c_1 = 0$.

Since $i \mapsto q_i^\eta$ is decreasing and $J(n,\eta)+k+1 > J(n,\eta)$, $q_{J(n,\eta)+k+1}^\eta \le \frac{1}{n}$, so

$$q_{J(n,\eta)+k+1} \le h(q_{J(n,\eta)+k}, \frac{\eta\pi_1}{1/n}) = \frac{1}{n-c_k} e^{-\eta\pi_1 nk} \frac{1}{1 - \frac{1}{n-c_k} e^{-\eta\pi_1 n(k-1)}}$$
$$= \frac{e^{-\eta\pi_1 nk}}{n - c_k - e^{-\eta\pi_1 n(k-1)}}$$

thus the same inequality holds for $q_{J(n,\eta)+k+1}$ with $c_{k+1} = c_k + e^{-\eta\pi_1 n(k-1)}$, and for all $k \ge 1$, $c_k \le \frac{1}{1-e^{-\eta\pi_1 n}}$. Since $\eta\pi_1 n \ge \log 2$, $c_k \le 2 < n$ for all $k \ge 1$ (and $n \ge 3$), which concludes the proof. □

Lemma A.4. *Let $\eta > 0$. For all $i \ge 0$,*

$$q_i^\eta \ge \frac{\eta\pi_1}{\log(1/q_{i+1}^\eta)}.$$

In particular, for all $i \le J(n,\eta)-1$,

$$q_i^\eta \ge \frac{\eta\pi_1}{\log n}.$$

Proof. First, for any $i \ge 1$,

$$q_i^\eta = \frac{q_{i-1}^\eta e^{-\eta\pi_1/q_{i-1}^\eta}}{1 - q_{i-1}^\eta + q_{i-1}^\eta e^{-\eta\pi_1/q_{i-1}^\eta}} \le \frac{q_{i-1}^\eta}{1-q_{i-1}^\eta} e^{-\eta\pi_1/q_{i-1}^\eta} \le e^{-\eta\pi_1/q_{i-1}^\eta}$$

since $q_{i-1}^\eta \le 1/2$, which leads to the first inequality.

For the second, recall that by definition, for any $i \le J(n,\eta)$, $q_i^\eta \ge \frac{1}{n}$, and use the first inequality. □

Lemma A.5. *Let $\delta > \eta > 0$. Then, the sequence $\left(\frac{q_i^\delta}{q_i^\eta}\right)_{i\ge 0}$ is decreasing and tends to 0 when i tends to $+\infty$.*

Proof. Let $u_i := \frac{q_i^\delta}{q_i^\eta}$ for all $i \ge 0$. Then, by definition of q_i^η,

$$u_{i+1} = u_i \cdot \frac{e^{-\delta\pi_1/q_i^\delta}}{(1-q_i^\delta) + q_i^\delta e^{-\delta\pi_1/q_i^\delta}} \cdot \frac{(1-q_i^\eta) + q_i^\eta e^{-\eta\pi_1/q_i^\eta}}{e^{-\eta\pi_1/q_i^\eta}}.$$

Let $h(q,\gamma) := \dfrac{e^{-\gamma\pi_1/q}}{(1-q)+qe^{-\gamma\pi_1/q}}$. Then, for all $0<q<1$ and $\gamma>0$,

$$\begin{aligned}\frac{\partial h}{\partial \gamma}(q,\gamma) &= \frac{-\frac{\pi_1}{q}e^{-\gamma\pi_1/q}\Big(1-q+qe^{-\gamma\pi_1/q}\Big) - e^{-\gamma\pi_1/q}(-\pi_1 e^{-\gamma\pi_1/q})}{\Big(1-q+qe^{-\gamma\pi_1/q}\Big)^2}\\ &= \frac{-\frac{\pi_1}{q}e^{-\gamma\pi_1/q}\Big(1-q\Big)}{\Big(1-q+qe^{-\gamma\pi_1/q}\Big)^2} < 0.\end{aligned}$$

And, for all $0<q<1$ and $\gamma>0$,

$$\begin{aligned}\frac{\partial h}{\partial q}(q,\gamma) &= \frac{\frac{\gamma\pi_1}{q^2}e^{-\gamma\pi_1/q}\Big(1-q+qe^{-\gamma\pi_1/q}\Big) - e^{-\gamma\pi_1/q}\Big(-1+e^{-\gamma\pi_1/q}(1+\frac{\gamma\pi_1}{q})\Big)}{\Big(1-q+qe^{-\gamma\pi_1/q}\Big)^2}\\ &= \frac{\frac{\gamma\pi_1}{q^2}e^{-\gamma\pi_1/q}(1-q) - e^{-\gamma\pi_1/q}(-1+e^{-\gamma\pi_1/q})}{\Big(1-q+qe^{-\gamma\pi_1/q}\Big)^2}\\ &= e^{-\gamma\pi_1/q}\frac{\frac{\gamma\pi_1}{q^2}(1-q)+1-e^{-\gamma\pi_1/q}}{\Big(1-q+qe^{-\gamma\pi_1/q}\Big)^2} > 0.\end{aligned}$$

Hence h is decreasing in γ and increasing in q. Let $i\geq 0$. Since $\delta>\eta$, $q_i^\delta \leq q_i^\eta$ and $u_i>0$,

$$0 < \frac{u_{i+1}}{u_i} = \frac{h(q_i^\delta,\delta)}{h(q_i^\eta,\eta)} \leq \frac{h(q_i^\eta,\delta)}{h(q_i^\eta,\eta)} \begin{cases} < \frac{h(q_i^\eta,\eta)}{h(q_i^\eta,\eta)} = 1 \\ = \frac{e^{-\delta\pi_1/q_i^\eta}}{1-q_i^\eta+q_i^\eta e^{-\delta\pi_1/q_i^\eta}} \cdot \frac{1-q_i^\eta+q_i^\eta e^{-\eta\pi_1/q_i^\eta}}{e^{-\eta\pi_1/q_i^\eta}} \sim e^{-\frac{(\delta-\eta)}{q_i^\eta}\pi_1} \longrightarrow 0 \end{cases}$$

since $q_i^\eta \longrightarrow 0$ when $i\to+\infty$. Therefore, $(u_i)_{i\geq 1}$ is decreasing and tends to 0. □

Lemma A.6. *Let $\eta>0$ such that $\eta\pi_1<1$. For all $i\leq I(\eta)$,*

$$\frac{1}{2} - i\eta\pi_1 \leq q_i^\eta \leq \frac{1}{2} - i\frac{\eta\pi_1}{4}. \tag{19}$$

Therefore,

$$\frac{1}{2\eta\pi_1} - 2 \leq I(\eta)$$

and when $\eta\pi_1 \leq \frac{1}{2}$,

$$I(\eta) \leq \frac{2}{\eta\pi_1} - 4.$$

Proof. Let's prove (19) first, starting with the lower bound. For all i,

$$q_i^\eta - q_{i+1}^\eta = \frac{q_i^\eta(1-q_i^\eta)(1-e^{-\eta\pi_1/q_i^\eta})}{1-q_i^\eta+q_i^\eta e^{-\eta\pi_1/q_i^\eta}} \leq q_i^\eta(1-e^{-\eta\pi_1/q_i^\eta}) \leq q_i^\eta\frac{\eta\pi_1}{q_i^\eta} = \eta\pi_1$$

since $\frac{1-q_i^\eta}{1-q_i^\eta+q_i^\eta e^{-\eta\pi_1/q_i^\eta}} \leq 1$ and $1-e^{-x}\leq x$. Thus, by summation,

$$q_i^\eta \geq \frac{1}{2} - \eta\pi_1 i.$$

Conversely, for all $i \leq I(\eta)$,

$$\begin{aligned} q_i^\eta - q_{i+1}^\eta &= \frac{q_i^\eta(1-q_i^\eta)(1-e^{-\eta\pi_1/q_i^\eta})}{1-q_i^\eta+q_i^\eta e^{-\eta\pi_1/q_i^\eta})} \\ &\geq (1-q_i^\eta)q_i^\eta(1-e^{-\eta\pi_1/q_i^\eta}) \\ &\geq (1-q_i^\eta)(\eta\pi_1 - \frac{1}{2}\frac{(\eta\pi_1)^2}{q_i^\eta}) \quad \text{since } 1-e^{-x} \geq x - x^2/2 \\ &\geq (1-\frac{1}{2})(\eta\pi_1 - \frac{\eta\pi_1}{2}) \quad \text{since } \eta\pi_1 \leq q_i^\eta \leq 1/2 \text{ for } i \leq I(\eta) \\ &= \frac{\eta\pi_1}{4}. \end{aligned}$$

The upper bound on q_i^η follows by summation. For the bracketing $I(\eta)$, by definition of $I(\eta)$ and (19),

$$\eta\pi_1 \geq q_{I(\eta)+1}^\eta \geq \frac{1}{2} - (I(\eta)+1)\eta\pi_1$$

and when $\eta\pi_1 \leq 1/2$,

$$\eta\pi_1 \leq q_{I(\eta)}^\eta \leq \frac{1}{2} - I(\eta)\frac{\eta\pi_1}{4}.$$

Therefore,

$$\frac{1}{2\eta\pi_1} - 2 \leq I(\eta) \leq \frac{2}{\eta\pi_1} - 4.$$

□

Lemma A.7. *Let δ and η such that $\frac{1}{\pi_1} > \eta > \delta > 0$. Then for all $i \geq 0$,*

$$0 \leq q_i^\delta - q_i^\eta \leq (\eta-\delta)\frac{8}{\delta\pi_1}\frac{(\frac{8}{\delta\pi_1})^i - 1}{\frac{8}{\delta\pi_1} - 1}.$$

Proof. Let

$$g_1(\delta, q) = \frac{qe^{-\delta\pi_1/q}}{1-q+qe^{-\delta\pi_1/q}}.$$

From Equation (21), for $\frac{1}{\pi_1} > \eta > 0$ and $q \in (0, \frac{1}{2})$,

$$\begin{aligned} \left|\frac{\partial g_1(\eta,q)}{\partial\eta}\right| &= \frac{(1-q)\pi_1 e^{-\eta\pi_1/q}}{(1-q+qe^{-\eta\pi_1/q})^2} \\ &\leq \frac{\pi_1 e^{-\eta\pi_1/q}}{1-q+qe^{-\eta\pi_1/q}} \\ &\leq 2\pi_1 e^{-\eta\pi_1/q}, \end{aligned}$$

since $1-q+qe^{-\eta\pi_k/q} \geq 1-q \geq \frac{1}{2}$. And from Equation (22), for $\frac{1}{\pi_1} > \eta > 0$ and $q \in (0, \frac{1}{2})$,

$$\begin{aligned} \frac{\partial g_1(\eta,q)}{\partial q} &= \frac{e^{-\eta\pi_1/q}(1+\frac{(1-q)\eta\pi_1}{q})}{(1-q+qe^{-\eta\pi_1/q})^2} \\ &\leq \frac{e^{-\eta\pi_1/q}}{q}4(1+\eta\pi_1) \\ &\leq 8\frac{e^{-\eta\pi_1/q}}{q}. \end{aligned}$$

In particular, since $\pi_1 \leq 1 \leq \frac{1}{q}$,

$$\max\left(\frac{\partial g_1(\eta,q)}{\partial q}, \left|\frac{\partial g_1(\eta,q)}{\partial\eta}\right|\right) \leq 8\frac{e^{-\delta\pi_1/q}}{q} \leq \frac{8}{\delta\pi_1},$$

where we used $e^x \geq x$ applied to $x = \frac{\delta\pi_1}{q}$. Let $\eta > \delta$, then by the mean value theorem, for all $i \geq 0$,

$$\begin{aligned} q_{i+1}^{\delta} - q_{i+1}^{\eta} &= q_{i+1}^{\delta} - g_1(\eta, q_i^{\delta}) + g_1(\eta, q_i^{\delta}) - q_{i+1}^{\eta} \\ &\leq 8\frac{e^{-\delta\pi_1/q_i^{\delta}}}{q_i^{\delta}}\left(q_i^{\delta} - q_i^{\eta} + \eta - \delta\right) \\ &\leq \frac{8}{\delta\pi_1}(q_i^{\delta} - q_i^{\eta} + \eta - \delta). \end{aligned}$$

Therefore, since $q_0^{\delta} = q_0^{\eta} = \frac{1}{2}$, iterating this equation leads to: for all $i \geq 0$,

$$\begin{aligned} q_i^{\delta} - q_i^{\eta} &\leq \sum_{k=0}^{i-1}\left(\frac{8}{\delta\pi_1}\right)^{k+1}(\eta - \delta) \\ &= (\eta - \delta)\frac{8}{\delta\pi_1}\frac{(\frac{8}{\delta\pi_1})^i - 1}{\frac{8}{\delta\pi_1} - 1}. \end{aligned}$$

The inequality $q_i^{\delta} - q_i^{\eta} \geq 0$ follows from the fact that $\delta \mapsto q_i^{\eta}$ is non-increasing, see Lemma A.1 (ii). □

A.2. For a decreasing learning rate

In this section, we prove all the results concerning Section 4.

A.2.1. Proof of Proposition 4.1: choice of a stopping time.

Let $\eta_0 \in \Theta$ and $1 \leq k \leq K$.

$$p_{k,t+1}^{\eta_n} = \frac{p_{k,t}^{\eta_n} e^{-\eta_n\pi_k/p_{k,t}^{\eta_n}}}{(1 - p_{k,t}^{\eta_n}) + p_{k,t}^{\eta_n} e^{-\eta_n\pi_k/p_{k,t}^{\eta_n}}}\mathbb{1}_{I_t=k} + \sum_{\substack{j=1 \\ j\neq k}}^{K}\frac{p_{k,t}^{\eta_n}}{(1 - p_{j,t}^{\eta_n}) + p_{j,t}^{\eta_n} e^{-\eta_n\pi_j/p_{j,t}^{\eta_n}}}\mathbb{1}_{I_t=j}.$$

For any $q \in [0,1]$, $1 - q + qe^{-\eta\pi_k/q} \leq 1$. Therefore,

$$p_{k,t+1}^{\eta_n} \geq p_{k,t}^{\eta_n} e^{-\eta_n\pi_k/p_{k,t}^{\eta_n}}\mathbb{1}_{I_t=k} + \sum_{\substack{j=1 \\ j\neq k}}^{K} p_{k,t}^{\eta_n}\mathbb{1}_{I_t=j}.$$

Since $e^{-\eta_n\pi_k/p_{k,t}^{\eta_n}} \leq 1$,

$$p_{k,t+1}^{\eta_n} \geq p_{k,t}^{\eta_n} e^{-\eta_n\pi_k/p_{k,t}^{\eta_n}}\mathbb{1}_{I_t=k} + \sum_{\substack{j=1 \\ j\neq k}}^{K} p_{k,t}^{\eta_n} e^{-\eta_n\pi_k/p_{k,t}^{\eta_n}}\mathbb{1}_{I_t=j} = p_{k,t}^{\eta_n} e^{-\eta_n\pi_k/p_{k,t}^{\eta_n}} \geq p_{k,t}^{\eta_n}(1 - \frac{\eta_n\pi_k}{p_{k,t}^{\eta_n}}) = p_{k,t}^{\eta_n} - \eta_n\pi_k.$$

Summing from 1 to t, since $p_{k,1}^{\eta_n} = \frac{1}{K}$,

$$p_{k,t}^{\eta_n} \geq \frac{1}{K} - \eta_n\pi_k t.$$

Hence, choosing $\Upsilon_n = \left\lfloor (\frac{1}{K} - \varepsilon)\frac{n^{\alpha}}{R}\right\rfloor$ implies that for all $\eta_0 \in \Theta$, $t \leq \Upsilon_n$ and $1 \leq k \leq K$,

$$\varepsilon \leq \frac{1}{K} - \frac{R}{n^{\alpha}\pi_1}\pi_1\Upsilon_n = \frac{1}{K} - R_n\pi_1\Upsilon_n \leq \frac{1}{K} - \eta_n\pi_1\Upsilon_n \leq \frac{1}{K} - \eta_n\pi_k\Upsilon_n \leq \frac{1}{K} - \eta_n\pi_k t \leq p_{k,t}^{\eta_n}.$$

A.2.2. Proof of Lemma 4.3: lipschitz character of the updated probability.

Recall that, for $\eta_0 \in \Theta$, for all $1 \leq k \leq K$, for all $t \leq \Upsilon_n - 1$,

$$p_{k,t+1}^{\eta_n} = \frac{p_{k,t}^{\eta_n} e^{-\eta_n\pi_k/p_{k,t}^{\eta_n}}}{(1 - p_{k,t}^{\eta_n}) + p_{k,t}^{\eta_n} e^{-\eta_n\pi_k/p_{k,t}^{\eta_n}}}\mathbb{1}_{I_t=k} + \sum_{\substack{j=1 \\ j\neq k}}^{K}\frac{p_{k,t}^{\eta_n}}{(1 - p_{j,t}^{\eta_n}) + p_{j,t}^{\eta_n} e^{-\eta_n\pi_j/p_{j,t}^{\eta_n}}}\mathbb{1}_{I_t=j}.$$

For all $1 \leq k \leq K$, denote by h_k and g_k the following functions:

$$h_k(\eta, q) := \frac{1}{1 - q + qe^{-\eta\pi_k/q}} \quad \text{and} \quad g_k(\eta, q) = qe^{-\eta\pi_k/q} h_k(\eta, q).$$

Then,

$$p_{k,t+1}^{\eta_n} = g_k(\eta_n, p_{k,t}^{\eta_n})\mathbb{1}_{I_t=k} + \sum_{\substack{j=1 \\ j\neq k}}^{K} p_{k,t}^{\eta_n} h_j(\eta_n, p_{j,t}^{\eta_n})\mathbb{1}_{I_t=j}.$$

Let $\delta_0, \delta_0' \in \Theta$. Let $1 \leq k \leq K$. Then, for all $t \leq \Upsilon_n - 1$,

$$|p_{k,t+1}^{\delta_n} - p_{k,t+1}^{\delta_n'}| \leq |g_k(\delta_n, p_{k,t}^{\delta_n}) - g_k(\delta_n', p_{k,t}^{\delta_n'})|\mathbb{1}_{I_t=k} + \sum_{\substack{j=1 \\ j\neq k}}^{K} |p_{k,t}^{\delta_n} h_j(\delta_n, p_{j,t}^{\delta_n}) - p_{k,t}^{\delta_n'} h_j(\delta_n', p_{j,t}^{\delta_n'})|\mathbb{1}_{I_t=j}. \tag{20}$$

The following lemma controls the derivatives of h_k and g_k and is proved at the end of the current section. Recall that $\pi_1 \geq \pi_k \geq 0$ for all k.

Lemma A.8. *Let $k \in \{1, \dots, K\}$. For all $\eta > 0$ such that $\eta\pi_1 < 1$,*

$$\begin{cases} 0 \leq h_k(\eta, q) \leq \frac{1}{1-\eta\pi_1} \\ 0 \leq \frac{\partial h_k}{\partial q}(\eta, q) \leq \frac{1}{(1-\eta\pi_1)^2}(\frac{\eta\pi_1}{q} \wedge 1)^2 \\ 0 \leq \frac{\partial h_k}{\partial \eta}(\eta, q) \leq \frac{\pi_1}{1-\eta\pi_1} \\ -\pi_1 \leq \frac{\partial g_k}{\partial \eta}(\eta, q) \leq 0 \\ 0 \leq \frac{\partial g_k}{\partial q}(\eta, q) \leq \frac{1}{(1-\eta\pi_1)^2}. \end{cases}$$

Let us now consider the two cases. Let n_0 be such that $Rn_0^{-\alpha} \leq \varepsilon^2$, so that $\eta_n\pi_1 \leq \varepsilon^2 \leq \varepsilon$ for all $n \geq n_0$ and $\eta_0 \in \Theta$, and let $R_n = \frac{R}{n^\alpha \pi_1}$. Recall that $\Upsilon_n = (\frac{1}{K} - \varepsilon)\frac{n^\alpha}{R}$ is chosen such that $p_{k,t}^{\eta_n} \geq \varepsilon$ for all k, $t \leq \Upsilon_n$ and $\eta_0 \in \Theta$ with Proposition 4.1.

- If $I_t = j \neq k$, then, by the triangle inequality,

$$\begin{aligned} |p_{k,t+1}^{\delta_n} - p_{k,t+1}^{\delta_n'}| &\leq |p_{k,t}^{\delta_n} h_j(\delta_n, p_{j,t}^{\delta_n}) - p_{k,t}^{\delta_n} h_j(\delta_n', p_{j,t}^{\delta_n'})| + |p_{k,t}^{\delta_n} h_j(\delta_n', p_{j,t}^{\delta_n'}) - p_{k,t}^{\delta_n'} h_j(\delta_n', p_{j,t}^{\delta_n'})| \\ &\leq |h_j(\delta_n, p_{j,t}^{\delta_n}) - h_j(\delta_n', p_{j,t}^{\delta_n'})| + h_j(\delta_n', p_{j,t}^{\delta_n'})|p_{k,t}^{\delta_n} - p_{k,t}^{\delta_n'}| \\ &\leq |h_j(\delta_n, p_{j,t}^{\delta_n}) - h_j(\delta_n, p_{j,t}^{\delta_n'})| + |h_j(\delta_n, p_{j,t}^{\delta_n'}) - h_j(\delta_n', p_{j,t}^{\delta_n'})| + h_j(\delta_n', p_{j,t}^{\delta_n'})\|p_t^{\delta_n} - p_t^{\delta_n'}\|_\infty. \end{aligned}$$

  Then, by Lemma A.8 and the mean value theorem, for all $n \geq n_0$,

$$|p_{k,t+1}^{\delta_n} - p_{k,t+1}^{\delta_n'}| \leq \left(\frac{1}{1 - R_n\pi_1} + \frac{(R_n\pi_1)^2}{\varepsilon^2(1 - R_n\pi_1)^2}\right)\|p_t^{\delta_n} - p_t^{\delta_n'}\|_\infty + \frac{\pi_1}{1 - R_n\pi_1}|\delta_n - \delta_n'|.$$

- If $I_t = k$, then, by the triangle inequality, Lemma A.8 and the mean value theorem, for all $n \geq n_0$,

$$\begin{aligned} |p_{k,t+1}^{\delta_n} - p_{k,t+1}^{\delta_n'}| &\leq |g_k(\delta_n, p_{k,t}^{\delta_n}) - g_k(\delta_n', p_{k,t}^{\delta_n})| + |g_k(\delta_n', p_{k,t}^{\delta_n}) - g_k(\delta_n', p_{k,t}^{\delta_n'})| \\ &\leq \pi_1|\delta_n - \delta_n'| + \frac{1}{(1 - R_n\pi_1)^2}|p_{k,t}^{\delta_n'} - p_{k,t}^{\delta_n}| \\ &\leq \pi_1|\delta_n - \delta_n'| + \frac{1}{(1 - R_n\pi_1)^2}\|p_t^{\delta_n'} - p_t^{\delta_n}\|_\infty. \end{aligned}$$

Thus, almost surely, for all $n \geq n_0$ and $t \leq \Upsilon_n - 1$,

$$\|p_{t+1}^{\delta_n'} - p_{t+1}^{\delta_n}\|_\infty \leq \frac{1 - R_n\pi_1 + (\frac{R_n\pi_1}{\varepsilon})^2}{(1 - R_n\pi_1)^2}\|p_t^{\delta_n'} - p_t^{\delta_n}\|_\infty + \frac{\pi_1}{1 - R_n\pi_1}|\delta_n - \delta_n'|.$$

Now, $\|p_1^{\delta_n'} - p_1^{\delta_n}\|_\infty = 0$, so that

$$\|p_t^{\delta_n'} - p_t^{\delta_n}\|_\infty \leq \frac{\pi_1}{1 - R_n\pi_1}|\delta_n - \delta_n'| \sum_{s=1}^{t-1} \left(\frac{1 - R_n\pi_1 + (\frac{R_n\pi_1}{\varepsilon})^2}{(1 - R_n\pi_1)^2} \right)^{s-1}.$$

The latter quantity is an increasing function of t, thus for all $t \leq \Upsilon_n$,

$$\|p_t^{\delta_n'} - p_t^{\delta_n}\|_\infty \leq \frac{\pi_1}{1 - R_n\pi_1}|\delta_n - \delta_n'| \frac{\left(\frac{1 - R_n\pi_1 + (\frac{R_n\pi_1}{\varepsilon})^2}{(1 - R_n\pi_1)^2} \right)^{\Upsilon_n - 1} - 1}{\frac{1 - R_n\pi_1 + (\frac{R_n\pi_1}{\varepsilon})^2}{(1 - R_n\pi_1)^2} - 1}.$$

Now,

$$\frac{1 - R_n\pi_1 + (\frac{R_n\pi_1}{\varepsilon})^2}{(1 - R_n\pi_1)^2} = 1 + R_n\pi_1 + u.$$

where $0 \leq u \leq c(R_n\pi_1)^2$ for some numerical constant c, for instance $c = \frac{76}{9}$, since $R_n\pi_1 \leq \varepsilon^2 \leq \frac{1}{4}$. Therefore,

$$\left(\frac{1 - R_n\pi_1 + (\frac{R_n\pi_1}{\varepsilon})^2}{(1 - R_n\pi_1)^2} \right)^{\Upsilon_n - 1} = \exp((\Upsilon_n - 1)\log(1 + R_n\pi_1 + u)) = \exp(\Upsilon_n R_n\pi_1 + \tilde{u})$$

where $|\tilde{u}| \leq R_n\pi_1 + \tilde{c}\Upsilon_n(R_n\pi_1)^2$ for some numerical constant $\tilde{c}$, for instance $\tilde{c} = 22$. Recall that $\Upsilon_n = (\frac{1}{K} - \varepsilon)\frac{n^\alpha}{R} = \frac{\frac{1}{K} - \varepsilon}{R_n\pi_1}$, so that $|\tilde{u}| \leq 12R_n\pi_1 \leq 3$ and

$$\left(\frac{1 - R_n\pi_1 + (\frac{R_n\pi_1}{\varepsilon})^2}{(1 - R_n\pi_1)^2} \right)^{\Upsilon_n - 1} = \exp(\frac{1}{K} - \varepsilon)\exp(\tilde{u}).$$

All in all,

$$\begin{aligned} \|p_t^{\delta_n'} - p_t^{\delta_n}\|_\infty &\leq \frac{\pi_1|\delta_n - \delta_n'|}{R_n\pi_1 + u} \left(\exp(\frac{1}{K} - \varepsilon)\exp(\tilde{u}) - 1 \right) \\ &\leq c' \frac{|\delta_0 - \delta_0'|}{R} \end{aligned}$$

for some numerical constant c', for instance $c' = 11$.

Proof of Lemma A.8. First of all, using $e^{-x} \geq 1 - x$ and $1 - \eta\pi_k \geq 1 - \eta\pi_1 > 0$,

$$h_k(\eta, q) \leq \frac{1}{1 - \eta\pi_k} \leq \frac{1}{1 - \eta\pi_1}.$$

Let's study the derivatives of h_k:

$$\frac{\partial h_k}{\partial q}(\eta, q) = \frac{1 - e^{-\eta\pi_k/q}(1 + \frac{\eta\pi_k}{q})}{(1 - q + qe^{-\eta\pi_k/q})^2} \geq 0$$

since $e^{-x}(1 + x) \leq 1$, and

$$\frac{\partial h_k}{\partial q}(\eta, q) \leq \begin{cases} \frac{1 - (1 - \frac{\eta\pi_k}{q})(1 + \frac{\eta\pi_k}{q})}{(1 - q + qe^{-\eta\pi_k/q})^2} \leq \frac{(\frac{\eta\pi_1}{q})^2}{(1 - \eta\pi_1)^2} & \text{if } \eta\pi_k/q \leq 1 \\ \frac{1}{(1 - q + qe^{-\eta\pi_k/q})^2} \leq \frac{1}{(1 - \eta\pi_1)^2} & \text{if } \eta\pi_k/q > 1 \end{cases}$$

Similarly, since $1 - q + qe^{-x} \geq (1 - q)e^{-x} + qe^{-x} = e^{-x}$ for any $x \geq 0$,

$$\frac{\partial h_k}{\partial \eta}(\eta, q) = \frac{\pi_k e^{-\eta\pi_k/q}}{(1 - q + qe^{-\eta\pi_k/q})^2} \leq \frac{\pi_k}{1 - \eta\pi_k} \leq \frac{\pi_1}{1 - \eta\pi_1}.$$

Now let's study the derivatives of g_k. For all $q \in (\varepsilon, 1]$,

$$\frac{\partial g_k(\eta, q)}{\partial \eta} = \frac{-(1-q)\pi_k e^{-\eta\pi_k/q}}{(1-q+qe^{-\eta\pi_k/q})^2} \leq 0. \tag{21}$$

Thus,

$$-\frac{\partial g_k(\eta, q)}{\partial \eta} \leq \frac{(1-q)\pi_k}{1-q+qe^{-\eta\pi_k/q}} \leq \pi_k \leq \pi_1,$$

and

$$\begin{aligned}
\frac{\partial g_k(\eta, q)}{\partial q} &= e^{-\eta\pi_k/q} h_k(\eta, q) + q\frac{\eta\pi_k}{q^2} e^{-\eta\pi_k/q} h_k(\eta, q) + qe^{-\eta\pi_k/q}\frac{\partial h_k(\eta, q)}{\partial q} \\
&= \frac{e^{-\eta\pi_k/q}}{(1-q+qe^{-\eta\pi_k/q})^2}\left((1+\frac{\eta\pi_k}{q})(1-q+qe^{-\eta\pi_k/q}) + q(1-e^{-\eta\pi_k/q}(1+\frac{\eta\pi_k}{q}))\right) \\
&= \frac{e^{-\eta\pi_k/q}(1+\frac{(1-q)\eta\pi_k}{q})}{(1-q+qe^{-\eta\pi_k/q})^2} \geq 0.
\end{aligned} \tag{22}$$

Now, since $e^x \geq 1+x$, $1+\frac{(1-q)\eta\pi_k}{q} \leq 1+\frac{\eta\pi_k}{q} \leq e^{\eta\pi_k/q}$, so

$$\frac{\partial g_k(\eta, q)}{\partial q} \leq \frac{1}{(1-q+qe^{-\eta\pi_k/q})^2} \leq \frac{1}{(1-\eta\pi_k)^2} \leq \frac{1}{(1-\eta\pi_1)^2}.$$

□

A.2.3. Proof of Proposition 4.4

For $\delta_0 \in \Theta$, recall that $X_{\delta_n} = \ell_{n,\varepsilon}(\delta_0) - \sum_{t=1}^{\Upsilon_n}\sum_{k=1}^{K}\log(p_{k,t}^{\delta_n})p_{k,t}^{\eta_n}$. For $t \geq 1$, let

$$Y_t^{\delta_n} := \sum_{k=1}^{K}\log(p_{k,t}^{\delta_n})\mathbb{1}_{I_t=k}$$

and let

$$C_t^{\delta_n} = \mathbb{E}_{\eta_n}[Y_t^{\delta_n}|\mathcal{F}_t] = \sum_{k=1}^{K}\log(p_{k,t}^{\delta_n})p_{k,t}^{\eta_n}$$

be its compensator. Then, $X_{\delta_n} = \sum_{t=1}^{\Upsilon_n}(Y_t^{\delta_n} - C_t^{\delta_n})$. Since $\widehat{\eta}_0$ is the maximum likelihood estimator, the following inequalities hold:

$$\begin{aligned}
\sum_{t=1}^{\Upsilon_n}(C_t^{\eta_n} - C_t^{\widehat{\eta}_n}) &\leq \underbrace{\ell_{n,\varepsilon}(\widehat{\eta}_0) - \ell_{n,\varepsilon}(\eta_0)}_{\geq 0} - \sum_{t=1}^{\Upsilon_n}(C_t^{\widehat{\eta}_n} - C_t^{\eta_n}) \\
&= \sum_{t=1}^{\Upsilon_n}(Y_t^{\widehat{\eta}_n} - Y_t^{\eta_n} - (C_t^{\widehat{\eta}_n} - C_t^{\eta_n})) \\
&\leq \sup_{\delta_0\in\Theta}\sum_{t=1}^{\Upsilon_n}(Y_t^{\delta_n} - Y_t^{\eta_n} - (C_t^{\delta_n} - C_t^{\eta_n})).
\end{aligned}$$

Let $\delta_0, \delta_0' \in \Theta$. Let $Z_1 = 0$ and for $2 \leq t \leq \Upsilon_n$, let $Z_t = \sum_{s=1}^{t-1}(Y_s^{\delta_n} - Y_s^{\delta_n'} - (C_s^{\delta_n} - C_s^{\delta_n'}))$. Then, $(Z_t)_t$ is an $(\mathcal{F}_t)_t$-martingale. Indeed,

$$\mathbb{E}_{\eta_n}[Z_{t+1}|\mathcal{F}_t] = \underbrace{\mathbb{E}_{\eta_n}[Y_t^{\delta_n} - Y_t^{\delta_n'}|\mathcal{F}_t] - (C_t^{\delta_n} - C_t^{\delta_n'})}_{=0} + \sum_{s=1}^{t-1}(Y_s^{\delta_n} - Y_s^{\delta_n'} - (C_s^{\delta_n} - C_s^{\delta_n'})) = Z_t.$$

Let $A_t^k = \sum_{s=2}^t \mathbb{E}_{\eta_n}[(Z_s - Z_{s-1})^k|\mathcal{F}_{s-1}]$ if $t \geq 2$ and 0 if $t = 1$. Let $\lambda \geq 0$. From Lemma 3.3 in (Houdré & Reynaud-Bouret, 2002), the sequence $(\mathcal{E}_t)_{t\geq 1}$ defined for all integers $t \geq 1$ by

$$\mathcal{E}_t = \exp(\lambda Z_t - \sum_{k\geq 2} \frac{\lambda^k}{k!} A_t^k)$$

is a supermartingale with respect to the filtration $(\mathcal{F}_t)_t$. As a result, for any $t \geq 1$,

$$\mathbb{E}_{\eta_n}[\mathcal{E}_t] \leq \mathbb{E}_{\eta_n}[\mathcal{E}_1] = 1.$$

Therefore, for any $t \geq 1$,

$$\mathbb{E}_{\eta_n}[\exp(\lambda Z_t)] \leq \mathbb{E}_{\eta_n}[\exp(\sum_{k\geq 2} \frac{\lambda^k}{k!} A_t^k)]. \tag{23}$$

Let $t \geq 2$. When there exists a constant C_Z such that $|Z_s - Z_{s-1}| \leq C_Z$ for all $2 \leq s \leq t$ almost surely, it holds for all $k \geq 2$,

$$|A_t^k| \leq \sum_{s=2}^t \mathbb{E}_{\eta_n}[(Z_s - Z_{s-1})^2|\mathcal{F}_{s-1}] C_Z^{k-2}. \tag{24}$$

For all $u \geq 2$,

$$Z_u - Z_{u-1} = \sum_{k=1}^K \log(\frac{p_{k,u}^{\delta_n}}{p_{k,u}^{\delta'_n}})(\mathbb{1}_{I_u=k} - p_{k,u}^{\eta_n}).$$

Hence,

$$|Z_u - Z_{u-1}| \leq \sum_{k=1}^K |\log(\frac{p_{k,u}^{\delta_n}}{p_{k,u}^{\delta'_n}})| \cdot |\mathbb{1}_{I_u=k} - p_{k,u}^{\eta_n}| \leq \left(\sup_{1\leq k\leq K} |\log(\frac{p_{k,u}^{\delta_n}}{p_{k,u}^{\delta'_n}})|\right) \sum_{k=1}^K |\mathbb{1}_{I_u=k} - p_{k,u}^{\eta_n}|.$$

Then, using the lipschitzianity of the $\log$ function on $[\varepsilon, 1]$ and Lemma 4.3,

$$|Z_u - Z_{u-1}| \leq \frac{2c}{R\varepsilon}|\delta_0 - \delta'_0|.$$

Therefore, Equation (24) becomes

$$|A_t^k| \leq (\frac{2c}{R\varepsilon}|\delta_0 - \delta'_0|)^{k-2} \sum_{s=2}^t \mathbb{E}_{\eta_n}[(Z_s - Z_{s-1})^2|\mathcal{F}_{s-1}]. \tag{25}$$

Let us control the order 2 moment in the above equation:

$$\begin{aligned}
\sum_{s=2}^t \mathbb{E}_{\eta_n}[(Z_s - Z_{s-1})^2|\mathcal{F}_{s-1}] &= \sum_{s=2}^t \mathbb{E}_{\eta_n}[(Y_s^{\delta_n} - Y_s^{\eta_n} - (C_s^{\delta_n} - C_s^{\eta_n}))^2|\mathcal{F}_{s-1}] \\
&= \sum_{s=2}^t \mathbb{E}_{\eta_n}\left[\left(\sum_{k=1}^K \log(\frac{p_{k,s}^{\delta_n}}{p_{k,s}^{\delta'_n}})(\mathbb{1}_{I_t=k} - p_{k,s}^{\eta_n})\right)^2 |\mathcal{F}_{s-1}\right] \\
&= \sum_{s=2}^t \mathbb{E}_{\eta_n}\left[\sum_{k=1}^K \log(\frac{p_{k,s}^{\delta_n}}{p_{k,s}^{\delta'_n}})^2(\mathbb{1}_{I_s=k} - p_{k,s}^{\eta_n})^2|\mathcal{F}_{s-1}\right] \\
&\quad + \sum_{s=2}^t \mathbb{E}_{\eta_n}\left[\sum_{j\neq k} \log(\frac{p_{k,s}^{\delta_n}}{p_{k,s}^{\delta'_n}}) \log(\frac{p_{j,s}^{\delta_n}}{p_{j,s}^{\delta'_n}})(\mathbb{1}_{I_s=k} - p_{k,s}^{\eta_n})(\mathbb{1}_{I_s=j} - p_{k,s}^{\eta_n}))|\mathcal{F}_{s-1}\right] \\
&= \sum_{s=2}^t \sum_{k=1}^K \log(\frac{p_{k,s}^{\delta_n}}{p_{k,s}^{\delta'_n}})^2(1 - p_{k,s}^{\eta_n})p_{k,s}^{\eta_n} - \sum_{s=2}^t \sum_{j\neq k} \log(\frac{p_{k,s}^{\delta_n}}{p_{k,s}^{\delta'_n}}) \log(\frac{p_{j,s}^{\delta_n}}{p_{j,s}^{\delta'_n}})p_{k,s}^{\eta_n} p_{j,s}^{\eta_n},
\end{aligned}$$

where for the last line we used that

$$\mathbb{E}_{\eta_n}[(\mathbb{1}_{I_s=k} - p_{k,s}^{\eta_n})^2|\mathcal{F}_{s-1}] = \mathbb{V}_{\eta_n}[\mathbb{1}_{I_s=k}|\mathcal{F}_{s-1}] = (1 - p_{k,s}^{\eta_n})p_{k,s}^{\eta_n}$$

and

$$\mathbb{E}_{\eta_n}[(\mathbb{1}_{I_s=k} - p_{k,s}^{\eta_n})(\mathbb{1}_{I_s=\ell} - p_{\ell,s}^{\eta_n})|\mathcal{F}_{s-1}] = -p_{\ell,s}^{\eta_n}p_{k,s}^{\eta_n}.$$

Therefore,

$$\begin{aligned}
\sum_{s=2}^{t}\mathbb{E}_{\eta_n}[(Z_s - Z_{s-1})^2|\mathcal{F}_{s-1}] &= \sum_{s=2}^{t}\sum_{k=1}^{K}(\log(\frac{p_{k,s}^{\delta_n}}{p_{k,s}^{\delta'_n}}))^2 p_{k,s}^{\eta_n} - \sum_{s=2}^{t}\sum_{j,k}\log(\frac{p_{k,s}^{\delta_n}}{p_{k,s}^{\delta'_n}})\log(\frac{p_{j,s}^{\delta_n}}{p_{j,s}^{\delta'_n}})p_{k,s}^{\eta_n}p_{j,s}^{\eta_n} \\
&= \sum_{s=2}^{t}\left(\sum_{k=1}^{K}(\log(\frac{p_{k,s}^{\delta_n}}{p_{k,s}^{\delta'_n}}))^2 p_{k,s}^{\eta_n} - \left(\sum_{k=1}^{K}\log(\frac{p_{k,s}^{\delta_n}}{p_{k,s}^{\delta'_n}})p_{k,s}^{\eta_n}\right)^2\right) \\
&\leq \sum_{s=2}^{t}\sum_{k=1}^{K}(\log(\frac{p_{k,s}^{\delta_n}}{p_{k,s}^{\delta'_n}}))^2 p_{k,s}^{\eta_n}.
\end{aligned}$$

The $\log$ function is lipschitz on $[\varepsilon, 1]$, and its lipschitz constant is $\frac{1}{\varepsilon}$. Therefore,

$$\sum_{s=2}^{t}\mathbb{E}_{\eta_n}[(Z_s - Z_{s-1})^2|\mathcal{F}_{s-1}] \leq \frac{1}{\varepsilon^2}\sum_{s=2}^{t}\sum_{k=1}^{K}(p_{k,s}^{\delta_n} - p_{k,s}^{\delta'_n})^2 p_{k,s}^{\eta_n},$$

and by Lemma 4.3,

$$\begin{aligned}
\sum_{s=2}^{t}\mathbb{E}_{\eta_n}[(Z_s - Z_{s-1})^2|\mathcal{F}_{s-1}] &\leq \frac{c^2}{R^2\varepsilon^2}\sum_{s=2}^{t}\sum_{k=1}^{K}|\delta_0 - \delta'_0|^2 p_{k,s}^{\eta_n} \\
&= \frac{c^2}{R^2\varepsilon^2}|\delta_0 - \delta'_0|^2(t-1). \qquad (26)
\end{aligned}$$

Injecting (26) and (25) in (23),

$$\mathbb{E}_{\eta_n}[\exp(\lambda Z_{\Upsilon_n+1})] \leq \mathbb{E}_{\eta_n}\left[\exp\left(\lambda^2\frac{c^2}{R^2\varepsilon^2}|\delta_0 - \delta'_0|^2\Upsilon_n\sum_{k\geq 2}\frac{\lambda^{k-2}}{k!}(\frac{2c}{R\varepsilon}|\delta_0 - \delta'_0|)^{k-2}\right)\right].$$

For all $k \geq 0$, $(k+2)! \geq 2^{k+1}$, so

$$\forall|\lambda| < (\frac{c}{R\varepsilon}|\delta_0 - \delta'_0|)^{-1}, \quad \mathbb{E}_{\eta_n}[\exp(\lambda Z_{\Upsilon_n+1})] \leq \exp\left(\frac{(\frac{\lambda c}{R\varepsilon}|\delta_0 - \delta'_0|)^2\Upsilon_n}{2(1 - \frac{\lambda c}{R\varepsilon}|\delta_0 - \delta'_0|)}\right).$$

To conclude, note that $Z_{\Upsilon_n+1} = X_{\delta_n} - X_{\delta'_n}$.

A.2.4. Proof of Theorem 4.5

The goal is to apply Theorem 2.1 of (Baraud, 2010). The distance $d(\cdot,\cdot)$ derives from a norm, and for all $\delta_0 \in \Theta$,

$$d(\delta_0, \eta_0) = \frac{c}{R\varepsilon}|\delta_0 - \eta_0| \leq \frac{c(R-r)}{R\varepsilon} \leq \frac{c}{\varepsilon}.$$

Therefore, Assumption 2.2 in (Baraud, 2010) is verified for the constants $v = \frac{c}{\varepsilon}\sqrt{\Upsilon_n}$ and $b = \frac{c}{\varepsilon}$, so that

$$\forall x \geq 0, \quad \mathbb{P}_{\eta_n}\left[Z_n \geq \kappa(v\sqrt{x+1} + b(x+1))\right] \leq e^{-x}$$

with $\kappa = 18$.

A.2.5. PROOF OF THEOREM 4.2

Take the notations of Section A.2.3. Since $\widehat{\eta}_0$ is the maximum likelihood estimator, the following inequalities hold:

$$\begin{aligned}\sum_{t=1}^{\Upsilon_n}(C_t^{\eta_n}-C_t^{\widehat{\eta}_n}) &\leq \underbrace{\ell_{n,\varepsilon}(\widehat{\eta}_0)-\ell_{n,\varepsilon}(\eta_0)}_{\geq 0}-\sum_{t=1}^{\Upsilon_n}(C_t^{\widehat{\eta}_n}-C_t^{\eta_n})\\ &= \sum_{t=1}^{\Upsilon_n}(Y_t^{\widehat{\eta}_n}-Y_t^{\eta_n}-(C_t^{\widehat{\eta}_n}-C_t^{\eta_n}))\\ &\leq \sup_{\delta_0\in\Theta}\sum_{t=1}^{\Upsilon_n}(Y_t^{\delta_n}-Y_t^{\eta_n}-(C_t^{\delta_n}-C_t^{\eta_n})).\end{aligned}$$

Note that

$$\sum_{t=1}^{\Upsilon_n}(C_t^{\eta_n}-C_t^{\widehat{\eta}_n})=\sum_{t=1}^{\Upsilon_n}\sum_{k=1}^{K}\log(\frac{p_{k,t}^{\eta_n}}{p_{k,t}^{\widehat{\eta}_n}})p_{k,t}^{\eta_n},$$

so that using Pinsker's inequality,

$$\sum_{t=1}^{\Upsilon_n}(C_t^{\eta_n}-C_t^{\widehat{\eta}_n})\geq 2\sum_{t=1}^{\Upsilon_n}\left(\sum_{k=1}^{K}|p_{k,t}^{\eta_n}-p_{k,t}^{\widehat{\eta}_n}|\right)^2.$$

By comparison of the norms,

$$=\left(\sum_{k=1}^{K}|p_{k,t}^{\eta_n}-p_{k,t}^{\widehat{\eta}_n}|\right)^2\geq\sum_{k=1}^{K}|p_{k,t}^{\eta_n}-p_{k,t}^{\widehat{\eta}_n}|^2=\|p_t^{\eta_n}-p_t^{\widehat{\eta}_n}\|_2^2.$$

Thus,

$$\sum_{t=1}^{\Upsilon_n}(C_t^{\eta_n}-C_t^{\widehat{\eta}_n})\geq 2\sum_{t=1}^{\Upsilon_n}\|p_t^{\eta_n}-p_t^{\widehat{\eta}_n}\|_2^2.$$

All in all,

$$2\sum_{t=1}^{\Upsilon_n}\|p_t^{\eta_n}-p_t^{\widehat{\eta}_n}\|_2^2\leq\sup_{\delta_0\in\Theta}\sum_{t=1}^{\Upsilon_n}(Y_t^{\delta_n}-Y_t^{\eta_n}-(C_t^{\delta_n}-C_t^{\eta_n}))=Z_n.$$

Using Theorem 4.5 with the same constants,

$$\forall x\geq 0,\quad \mathbb{P}_{\eta_n}\left[2\sum_{t=1}^{\Upsilon_n}\|p_t^{\eta_n}-p_t^{\widehat{\eta}_n}\|_2^2\geq\frac{18c}{\varepsilon}\left(\sqrt{\Upsilon_n}\sqrt{x+1}+x+1)\right)\right]\leq e^{-x},$$

that is,

$$\forall x\geq 0,\quad \mathbb{P}_{\eta_n}\left[\frac{1}{\Upsilon_n}\sum_{t=1}^{\Upsilon_n}\|p_t^{\eta_n}-p_t^{\widehat{\eta}_n}\|_2^2\geq\frac{9c}{\varepsilon}\left(\sqrt{\frac{x+1}{\Upsilon_n}}+\frac{x+1}{\Upsilon_n}\right)\right]\leq e^{-x}.$$

A.3. Application to a special case

In this section, we prove all the results of Section 5. Some lemmas are used for the two propositions. They can be found at the end of each section.

A.3.1. PROOF OF PROPOSITION 5.1: A LOWER BOUND ON THE PREDICTION ERROR.

By Lemma A.9, there exists D_{π_1} such that

$$|p_{1,t}^{\widehat{\eta}_n}-p_{1,t}^{\eta_n}|\geq D_{\pi_1}|\widehat{\eta}_n-\eta_n|e^{-(t-2)R_n\pi_1/\varepsilon}N_{t-1}.$$

Taking the square and summing,

$$\sum_{t=1}^{\Upsilon_n} |p_{1,t}^{\delta_n} - p_{1,t}^{\eta_n}|^2 \geq D_{\pi_1}^2 |\widehat{\eta}_n - \eta_n|^2 \sum_{t=1}^{\Upsilon_n} e^{-2(t-2)R_n\pi_1/\varepsilon} N_{t-1}^2 \geq D_{\pi_1}^2 |\widehat{\eta}_n - \eta_n|^2 e^{-2\Upsilon_n R_n \pi_1/\varepsilon} \sum_{t=1}^{\Upsilon_n} N_{t-1}^2.$$

Since $\Upsilon_n = (\frac{1}{2} - \varepsilon)\frac{n^\alpha}{R}$,

$$\sum_{t=1}^{\Upsilon_n} |p_{1,t}^{\delta_n} - p_{1,t}^{\eta_n}|^2 \geq D_{\pi_1}^2 |\widehat{\eta}_n - \eta_n|^2 e^{-(1-2\varepsilon)/\varepsilon} \sum_{t=1}^{\Upsilon_n} N_{t-1}^2.$$

Let $m_{\pi_1,\varepsilon} := D_{\pi_1}^2 e^{-(1-2\varepsilon)/\varepsilon}$. Then,

$$\sum_{t=1}^{\Upsilon_n} |p_{1,t}^{\delta_n} - p_{1,t}^{\eta_n}|^2 \geq m_{\pi_1,\varepsilon} |\widehat{\eta}_n - \eta_n|^2 \sum_{t=1}^{\Upsilon_n} N_{t-1}^2,$$

which is the desired result.

Lemma A.9. *Recall that* $R = \max\{\delta_0, \delta_0 \in \Theta\} = \dfrac{(\frac{1}{K} - \varepsilon)n^\alpha}{\Upsilon_n}$. *Let* $N_t = \sum_{s=1}^t \mathbb{1}_{I_s=1}$. *Let* η_0, $\delta_0 \in \Theta$. *There exists* $D_{\pi_1} > 0$, *such that for any* $t \in \{1, \dots, \Upsilon_n\}$

$$|p_{1,t+1}^{\delta_n} - p_{1,t+1}^{\eta_n}| \geq D_{\pi_1} |\eta_n - \delta_n| \sum_{s=1}^t e^{-(t-s)R_n\pi_1/\varepsilon} \mathbb{1}_{I_s=1} \geq D_{\pi_1} |\eta_n - \delta_n| e^{-(t-1)R_n\pi_1/\varepsilon} N_t. \tag{27}$$

Proof. Suppose $\delta_0 \leq \eta_0$. Then, for all $t \geq 1$,

$$p_{1,t}^{\delta_n} \geq p_{1,t}^{\eta_n}.$$

Indeed, the result holds for $t = 1$. Assume it is true for $t \in \mathbb{N}$. If $I_t = 2$, then

$$p_{1,t+1}^{\delta_n} = p_{1,t}^{\delta_n} \geq p_{1,t}^{\eta_n} = p_{1,t+1}^{\eta_n}.$$

If $I_t = 1$, then

$$p_{1,t+1}^{\delta_n} = g(p_{1,t}^{\delta_n}, \delta_n) \geq g(p_{1,t}^{\delta_n}, \eta_n) \geq g(p_{1,t}^{\eta_n}, \eta_n) = p_{1,t+1}^{\eta_n}$$

where g is the function defined in the proof of Lemma A.1 and verifies that $q \mapsto g(q, \eta)$ is an increasing function and $\eta \mapsto g(q, \eta)$ is a decreasing function. Hence the result. Let's examine

$$p_{1,t+1}^{\delta_n} - p_{1,t+1}^{\eta_n} = (p_{1,t}^{\delta_n} - p_{1,t}^{\eta_n})\mathbb{1}_{I_t=2} + (g(p_{1,t}^{\delta_n}, \delta_n) - g(p_{1,t}^{\eta_n}, \eta_n))\mathbb{1}_{I_t=1}.$$

Recall that $R = \max \Theta$. Using Lemma A.10,

$$\begin{aligned} g(p_{1,t}^{\delta_n}, \delta_n) - g(p_{1,t}^{\eta_n}, \eta_n) &\geq g(p_{1,t}^{\delta_n}, \delta_n) - g(p_{1,t}^{\eta_n}, \delta_n) + g(p_{1,t}^{\eta_n}, \delta_n) - g(p_{1,t}^{\eta_n}, \eta_n) \\ &\geq e^{-R_n\pi_1/\varepsilon}(p_{1,t}^{\delta_n} - p_{1,t}^{\eta_n}) + D_{\pi_1}(\eta_n - \delta_n). \end{aligned}$$

Therefore,

$$\begin{aligned} p_{1,t+1}^{\delta_n} - p_{1,t+1}^{\eta_n} &\geq (p_{1,t}^{\delta_n} - p_{1,t}^{\eta_n})\mathbb{1}_{I_t=2} + \Big(\underbrace{e^{-R_n\pi_1/\varepsilon}}_{\leq 1}(p_{1,t}^{\delta_n} - p_{1,t}^{\eta_n}) + D_{\pi_1}(\eta_n - \delta_n)\Big)\mathbb{1}_{I_t=1} \\ &\geq e^{-R_n\pi_1/\varepsilon}(p_{1,t}^{\delta_n} - p_{1,t}^{\eta_n})\mathbb{1}_{I_t=2} + (e^{-R_n\pi_1/\varepsilon}(p_{1,t}^{\delta_n} - p_{1,t}^{\eta_n}) + D_{\pi_1}(\eta_n - \delta_n))\mathbb{1}_{I_t=1} \\ &\geq e^{-R_n\pi_1/\varepsilon}(p_{1,t}^{\delta_n} - p_{1,t}^{\eta_n}) + D_{\pi_1}(\eta_n - \delta_n)\mathbb{1}_{I_t=1}. \end{aligned}$$

Therefore, iterating,

$$p_{1,t+1}^{\delta_n} - p_{1,t+1}^{\eta_n} \geq D_{\pi_1}(\eta_n - \delta_n) \sum_{s=1}^t \mathbb{1}_{I_s=1} e^{-(t-s)R_n\pi_1/\varepsilon}.$$

The proof is the same for $\delta_0 > \eta_0$. □

Lemma A.10. *For all $\gamma\pi_1 \in (0,1]$ and $q \in [\max(\varepsilon, \gamma\pi_1), \frac{1}{2}]$,*

$$|\frac{\partial g_1(\gamma,q)}{\partial \gamma}| \geq \frac{\pi_1 e^{-1}}{2} =: D_{\pi_1} \quad \textit{and} \quad \frac{\partial g_1(\gamma,q)}{\partial q} \geq e^{-\gamma\pi_1/\varepsilon}.$$

Proof. In (21), for all $\gamma\pi_1 \in (0,1]$ and $q \in [\max(\varepsilon, \gamma\pi_1), \frac{1}{2}]$,

$$\begin{aligned}
|\frac{\partial g_1(\gamma,q)}{\partial \gamma}| &= \frac{(1-q)\pi_1 e^{-\gamma\pi_1/q}}{((1-q)+qe^{-\gamma\pi_1/q})^2} \\
&\geq \frac{\pi_1 e^{-1}}{2((1-q)+qe^{-\gamma\pi_1/q})^2} \quad \text{since } q \leq \frac{1}{2} \\
&\geq \frac{\pi_1 e^{-1}}{2} =: D_{\pi_1} \quad \text{using that } 1-q+qe^{-\gamma\pi_1/q} \leq 1.
\end{aligned}$$

Similarly, using (22), for all $q \in [\varepsilon, \frac{1}{2}]$,

$$\begin{aligned}
\frac{\partial g_1(\gamma,q)}{\partial q} &= \frac{e^{-\gamma\pi_1/q}(1+\frac{1-q}{q}\gamma\pi_1)}{((1-q)+qe^{-\gamma\pi_1/q})^2} \\
&\geq \frac{e^{-\gamma\pi_1/\varepsilon}}{((1-q)+qe^{-\gamma\pi_1/q})^2} \\
&\geq e^{-\gamma\pi_1/\varepsilon}.
\end{aligned}$$

□

Lemma A.11. *Let δ_1, δ_2, ε_1, ε_2 be positive numbers. Let X and Y be two random variables such that $\mathbb{P}(X \geq \varepsilon_1) \leq \delta_1$ and $\mathbb{P}(Y \geq \varepsilon_2) \leq \delta_2$. Then,*

$$\mathbb{P}(X+Y \geq \varepsilon_1+\varepsilon_2) \leq \delta_1+\delta_2.$$

Proof. It is a direct consequence from the inclusion $\{X+Y \geq \varepsilon_1+\varepsilon_2\} \subset \{X \geq \varepsilon_1\} \cup \{Y \geq \varepsilon_1\}$ and the union bound. □

A.3.2. PROOF OF PROPOSITION 5.2: AN UPPER BOUND ON THE ESTIMATION ERROR.

Let $A_n : \dfrac{\Upsilon_n(\Upsilon_n-1)(2\Upsilon_n-1)}{96}$. By Lemma A.12, for all $y > 0$,

$$\mathbb{P}_{\eta_n}\left(A_n - \Big(\frac{2}{5}\sqrt{2\log\frac{2\Upsilon_n}{y}}\Upsilon_n^{\frac{5}{2}} + \Upsilon_n^2 \log\frac{2\Upsilon_n}{y}\Big) \geq \sum_{t=1}^{\Upsilon_n} N_{t-1}^2\right) \leq y. \tag{28}$$

By Proposition 4.2, for all $n \geq (R/\varepsilon^2)^{1/\alpha}$ and $x \geq 0$,,

$$\mathbb{P}_{\eta_n}\left(\sum_{t=1}^{\Upsilon_n} \|p_t^{\eta_n} - p_t^{\widehat{\eta}_n}\|_2^2 < \frac{9c}{\varepsilon}\left(\sqrt{(x+1)\Upsilon_n} + x + 1\right)\right) \geq 1 - e^{-x}. \tag{29}$$

By Proposition 5.1, there exists a constant $m_{\pi_1,\varepsilon} > 0$, such that for all $\eta_0, \widehat{\eta}_0 \in \Theta$,

$$\sum_{t=1}^{\Upsilon_n} |p_{1,t}^{\widehat{\eta}_n} - p_{1,t}^{\eta_n}|^2 \geq m_{\pi_1,\varepsilon}|\widehat{\eta}_n - \eta_n|^2 \sum_{t=1}^{\Upsilon_n} N_{t-1}^2.$$

Since $\displaystyle\sum_{t=1}^{\Upsilon_n} |p_{1,t}^{\widehat{\eta}_n} - p_{1,t}^{\eta_n}|^2 = \frac{1}{2}\sum_{t=1}^{\Upsilon_n} \|p_t^{\eta_n} - p_t^{\widehat{\eta}_n}\|_2^2$, injecting this result in Equation (29),

$$\mathbb{P}_{\eta_n}\left(2m_{\pi_1,\varepsilon}|\widehat{\eta}_n - \eta_n|^2 \sum_{t=1}^{\Upsilon_n} N_{t-1}^2 > \frac{9c}{\varepsilon}\left(\sqrt{(x+1)\Upsilon_n} + x + 1\right)\right) \leq e^{-x},$$

that is,

$$\mathbb{P}_{\eta_n}\left(|\widehat{\eta}_n-\eta_n|^2\sum_{t=1}^{\Upsilon_n}N_{t-1}^2>\frac{9c}{2m_{\pi_1,\varepsilon}\varepsilon}\left(\sqrt{(x+1)\Upsilon_n}+x+1\right)\right)\leq e^{-x}. \tag{30}$$

Combining (28) and (30) with Lemma A.11,

$$\begin{aligned}
&\mathbb{P}_{\eta_n}\left(|\widehat{\eta}_n-\eta_n|^2A_n>\frac{9c}{2m_{\pi_1,\varepsilon}\varepsilon}\left(\sqrt{(x+1)\Upsilon_n}+x+1\right)+|\widehat{\eta}_n-\eta_n|^2\Big(\frac{2}{5}\sqrt{2\log\frac{2\Upsilon_n}{y}}\Upsilon_n^{\frac{5}{2}}+\Upsilon_n^2\log\frac{2\Upsilon_n}{y}\Big)\right)\\
&\leq\mathbb{P}_{\eta_n}\left(|\widehat{\eta}_n-\eta_n|^2\sum_{t=1}^{\Upsilon_n}N_{t-1}^2>\frac{9c}{2m_{\pi_1,\varepsilon}\varepsilon}\left(\sqrt{(x+1)\Upsilon_n}+x+1\right)\right)\\
&+\mathbb{P}_{\eta_n}\left(A_n-\sum_{t=1}^{\Upsilon_n}N_{t-1}^2\geq\Big(\frac{2}{5}\sqrt{2\log\frac{2\Upsilon_n}{y}}\Upsilon_n^{\frac{5}{2}}+\Upsilon_n^2\log\frac{2\Upsilon_n}{y}\Big)\right)\\
&\leq e^{-x}+y.
\end{aligned}$$

Therefore, with probability at least $1-(e^{-x}+y)$,

$$|\widehat{\eta}_n-\eta_n|^2\left(A_n-\Big(\frac{2}{5}\sqrt{2\log\frac{2\Upsilon_n}{y}}\Upsilon_n^{\frac{5}{2}}+\Upsilon_n^2\log\frac{2\Upsilon_n}{y}\Big)\right)\leq\frac{9c}{2m_{\pi_1,\varepsilon}\varepsilon}\left(\sqrt{(x+1)\Upsilon_n}+x+1\right).$$

Recall that $\Upsilon_n=\big(\frac{1}{2}-\varepsilon\big)\frac{n^\alpha}{R}$ and that

$$|\widehat{\eta}_n-\eta_n|=\frac{1}{n^\alpha\pi_1}|\widehat{\eta}_0-\eta_0|=\frac{\frac{1}{2}-\varepsilon}{R\pi_1}|\widehat{\eta}_0-\eta_0|\frac{1}{\Upsilon_n}.$$

Let $B_n:=\frac{A_n}{\Upsilon_n^2}$. Then, with probability at least $1-(e^{-x}+y)$,

$$|\widehat{\eta}_0-\eta_0|^2\left(B_n-\Big(\frac{2}{5}\sqrt{2\log\frac{2\Upsilon_n}{y}}\Upsilon_n^{\frac{1}{2}}+\log\frac{2\Upsilon_n}{y}\Big)\right)\leq\frac{9c(R\pi_1)^2}{2(\frac{1}{2}-\varepsilon)^2m_{\pi_1,\varepsilon}\varepsilon}\left(\sqrt{(x+1)\Upsilon_n}+x+1\right),$$

that is,

$$|\widehat{\eta}_0-\eta_0|\leq\frac{R\pi_1}{\frac{1}{2}-\varepsilon}\sqrt{\frac{9c}{2m_{\pi_1,\varepsilon}\varepsilon}}\left(\frac{\sqrt{(x+1)\Upsilon_n}+x+1}{B_n-\Big(\frac{2}{5}\sqrt{2\log\frac{2\Upsilon_n}{y}}\Upsilon_n^{\frac{1}{2}}+\log\frac{2\Upsilon_n}{y}\Big)}\right)^{\frac{1}{2}}.$$

Choose $y=e^{-x}$, then with probability at least $1-2e^{-x}$,

$$|\widehat{\eta}_0-\eta_0|\leq\frac{R\pi_1}{\frac{1}{2}-\varepsilon}\sqrt{\frac{9c}{2m_{\pi_1,\varepsilon}\varepsilon}}\left(\frac{\sqrt{(x+1)\Upsilon_n}+x+1}{B_n-\Big(\frac{2}{5}\sqrt{2(\log2\Upsilon_n+x)}\Upsilon_n^{\frac{1}{2}}+\log2\Upsilon_n+x\Big)}\right)^{\frac{1}{2}}.$$

Lemma A.12. *Let* $A_n:=\frac{\Upsilon_n(\Upsilon_n-1)(2\Upsilon_n-1)}{96}$. *For all* $y>0$,

$$\mathbb{P}_{\eta_n}\left(A_n-\sum_{t=1}^{\Upsilon_n}N_{t-1}^2\geq\frac{2}{5}\sqrt{2\log\frac{2\Upsilon_n}{y}}\Upsilon_n^{\frac{5}{2}}+\Upsilon_n^2\log\frac{2\Upsilon_n}{y}\right)\leq y.$$

Proof. For $t\geq0$, let $M_0=0$ and $M_t=\sum_{s=1}^t\left(\mathbb{1}_{I_s=1}-p_{1,s}^{\eta_n}\right)$ if $t\geq1$. Then (M_t) is an $(\mathcal{F}_t)$-martingale. Indeed,

$$\mathbb{E}_{\eta_n}\left[M_t|\mathcal{F}_{t-1}\right]=\sum_{s=1}^{t-1}\left(\mathbb{1}_{I_s=1}-p_{1,s}^{\eta_n}\right)+\underbrace{\mathbb{E}_{\eta_n}\left[\mathbb{1}_{I_s=1}|\mathcal{F}_{t-1}\right]-p_{1,t}^{\eta_n}}_{=0}=M_{t-1}.$$

Using Lemma 3.3 in (Houdré & Reynaud-Bouret, 2002), for all λ,

$$\mathcal{E}_t := \exp\left(\lambda M_t - \sum_{k\geq 2}\frac{\lambda^k}{k!}\sum_{s=1}^{t}\mathbb{E}_{\eta_n}\left[(M_s - M_{s-1})^k|\mathcal{F}_{s-1}\right]\right)$$

is an $(\mathcal{F}_t)$-supermartingale. Therefore,

$$\mathbb{E}_{\eta_n}[\mathcal{E}_t] \leq \mathbb{E}_{\eta_n}[\mathcal{E}_1] = 1,$$

that is,

$$\mathbb{E}_{\eta_n}\left[e^{\lambda M_t}\right] \leq \mathbb{E}_{\eta_n}\left[\exp\left(\sum_{k\geq 2}\frac{\lambda^k}{k!}\sum_{s=1}^{t}\mathbb{E}_{\eta_n}\left[(M_s - M_{s-1})^k|\mathcal{F}_{s-1}\right]\right)\right] \tag{31}$$

$$\leq \mathbb{E}_{\eta_n}\left[\exp\left(\sum_{k\geq 2}\frac{|\lambda|^k}{k!}\sum_{s=1}^{t}\mathbb{E}_{\eta_n}\left[|M_s - M_{s-1}|^k|\mathcal{F}_{s-1}\right]\right)\right] \tag{32}$$

using the monotonicity of $\exp$ and the triangle inequality. The difference $|M_s - M_{s-1}|$ is almost surely bounded:

$$|M_s - M_{s-1}| = |\mathbb{1}_{I_s=1} - p_{1,s}^{\eta_n}| \leq 1, \tag{33}$$

and it conditional variance can be controlled by

$$\sum_{s=1}^{t}\mathbb{E}_{\eta_n}\left[(M_s - M_{s-1})^2|\mathcal{F}_{s-1}\right] = \sum_{s=1}^{t}\mathbb{E}_{\eta_n}\left[(\mathbb{1}_{I_s=1} - p_{1,s}^{\eta_n})^2|\mathcal{F}_{s-1}\right] = \sum_{s=1}^{t}p_{1,s}^{\eta_n}(1 - p_{1,s}^{\eta_n}) \leq \frac{t}{4}, \tag{34}$$

since $p_{1,s}^{\eta_n} \leq \frac{1}{2}$. Therefore, (33) and (34) imply

$$\sum_{s=1}^{t}\mathbb{E}_{\eta_n}\left[|M_s - M_{s-1}|^k|\mathcal{F}_{s-1}\right] = \sum_{s=1}^{t}\mathbb{E}_{\eta_n}\left[|M_s - M_{s-1}|^{k-2}(M_s - M_{s-1})^2|\mathcal{F}_{s-1}\right] \tag{35}$$

$$\leq \sum_{s=1}^{t}\mathbb{E}_{\eta_n}\left[(M_s - M_{s-1})^2|\mathcal{F}_{s-1}\right] \leq \frac{t}{4}. \tag{36}$$

Injecting this equation in (31) shows that for all λ,

$$\mathbb{E}_{\eta_n}\left[e^{\lambda M_t}\right] \leq \mathbb{E}_{\eta_n}\left[\exp\left(\sum_{k\geq 2}\frac{|\lambda|^k}{k!}\frac{t}{4}\right)\right] = \exp\left(\lambda^2\sum_{k\geq 0}\frac{|\lambda|^k}{(k+2)!}\frac{t}{4}\right).$$

Given that $(k+2)! \geq 2^{k+1}$ for all $k \geq 0$, for all $\lambda \in (-2, 2)$,

$$\mathbb{E}_{\eta_n}\left[e^{\lambda M_t}\right] \leq \exp\left(\frac{\lambda^2}{2}\sum_{k\geq 0}\frac{|\lambda|^k}{2^k}\frac{t}{4}\right) \leq \exp\left(\frac{\lambda^2 t}{8(1 - \frac{|\lambda|}{2})}\right). \tag{37}$$

Let $x > 0$ and $\lambda \in (0, 2)$. Then,

$$\mathbb{P}_{\eta_n}\big(|M_t| \geq x\big) \leq \mathbb{P}_{\eta_n}\big(M_t \geq x\big) + \mathbb{P}_{\eta_n}\big(-M_t \geq x\big).$$

Using (37) and Chernoff's bound,

$$\mathbb{P}_{\eta_n}\big(|M_t| \geq x\big) \leq 2\exp\left(\frac{\lambda^2 t/2}{4(1 - \frac{\lambda}{2})} - \lambda x\right).$$

It can be shown that

$$\sup_{\lambda\in[0,2)}\left(\lambda x-\frac{\lambda^2/2t}{4(1-\lambda/2)}\right)=t\cdot h(2x/t)\geq\frac{2x^2}{t+2x}$$

where $h(u)=1+u-\sqrt{1+2u}$ for $u>0$, using that $h(u)\geq\frac{u^2}{2(1+u)}$ for $u>0$. Therefore,

$$\mathbb{P}_{\eta_n}\big(|M_t|\geq x\big)\leq 2\exp\left(-\frac{2x^2}{t+2x}\right).$$

That is

$$\mathbb{P}_{\eta_n}\Big(\big|N_t-\sum_{s=1}^t p_{1,s}^{\eta_n}\big|\geq x\Big)\leq 2\exp\left(-\frac{2x^2}{t+2x}\right) \tag{38}$$

For all $t\leq\Upsilon_n$, $0\leq N_t\leq t$ and $0\leq\sum_{s=1}^t p_{1,s}^{\eta_n}\leq t$, so

$$\big|N_t^2-\big(\sum_{s=1}^t p_{1,s}^{\eta_n}\big)^2\big|=\big|N_t-\sum_{s=1}^t p_{1,s}^{\eta_n}\big|\big(N_t+\sum_{s=1}^t p_{1,s}^{\eta_n}\big)\leq 2t\big|N_t-\sum_{s=1}^t p_{1,s}^{\eta_n}\big|.$$

Therefore,

$$\begin{aligned}\mathbb{P}_{\eta_n}\Big(\big(\sum_{s=1}^t p_{1,s}^{\eta_n}\big)^2-x\geq N_t^2\Big)&\leq\mathbb{P}_{\eta_n}\Big(\big|N_t^2-\big(\sum_{s=1}^t p_{1,s}^{\eta_n}\big)^2\big|\geq x\Big)\\&\leq\mathbb{P}_{\eta_n}\Big(2t\big|N_t-\sum_{s=1}^t p_{1,s}^{\eta_n}\big|\geq x\Big)\leq 2\exp\left(-\frac{x^2}{2(t^3+xt)}\right).\end{aligned}$$

Lemma A.13 (Inequality Reversal (Lemma 1 in (Peel et al., 2010))). *Let X be a random variable and a, $b>0$, c, $d\geq 0$ such that*

$$\forall x>0,\mathbb{P}(|X|\geq x)\leq a\exp\left(-\frac{bx^2}{c+dx}\right).$$

Then,

$$\forall y>0,\mathbb{P}\left(|X|\geq\sqrt{\frac{c}{b}\log\frac{a}{y}}+\frac{d}{b}\log\frac{a}{y}\right)\leq y.$$

Using this Lemma, the previous equation implies: for all $u>0$,

$$\mathbb{P}_{\eta_n}\left(\big(\sum_{s=1}^t p_{1,s}^{\eta_n}\big)^2-N_t^2\geq H(t,u)\right)\leq u,$$

where $H(t,u):=t^{\frac{3}{2}}\sqrt{2\log\frac{2}{u}}+2t\log\frac{2}{u}$. Using Lemma A.11 and summing the probabilities from $t=1$ to Υ_n,

$$\mathbb{P}_{\eta_n}\Big(\sum_{t=1}^{\Upsilon_n}\big(\sum_{s=1}^{t-1}p_{1,s}^{\eta_n}\big)^2-\sum_{t=1}^{\Upsilon_n}N_{t-1}^2\geq\sum_{t=1}^{\Upsilon_n}H(t-1,u)\Big)\leq\Upsilon_n u. \tag{39}$$

Now,

$$\begin{aligned}\sum_{t=1}^{\Upsilon_n}H(t-1,u)&\leq\sqrt{2\log\frac{2}{u}}\int_0^{\Upsilon_n}t^{\frac{3}{2}}dt+\Upsilon_n(\Upsilon_n-1)\log\frac{2}{u}\\&\leq\frac{2}{5}\sqrt{2\log\frac{2}{u}}\Upsilon_n^{\frac{5}{2}}+\Upsilon_n^2\log\frac{2}{u}.\end{aligned}$$

Choosing $y = \Upsilon_n u$, (39) becomes

$$\mathbb{P}_{\eta_n}\Big(\sum_{t=1}^{\Upsilon_n}\big(\sum_{s=1}^{t-1} p_{1,s}^{\eta_n}\big)^2 - \sum_{t=1}^{\Upsilon_n} N_{t-1}^2 \geq \frac{2}{5}\sqrt{2\log\frac{2\Upsilon_n}{y}}\Upsilon_n^{\frac{5}{2}} + \Upsilon_n^2 \log\frac{2\Upsilon_n}{y}\Big) \leq y. \tag{40}$$

From the proof of Proposition 4.1 (see Section A.2.1), for $s \leq \Upsilon_n$, $p_{1,s}^{\eta_n} \geq \frac{1}{2} - \eta_n \pi_1 s \geq \frac{1}{2} - \frac{R}{n^\alpha} s$. Therefore,

$$\sum_{s=1}^{t-1} p_{1,s}^{\eta_n} \geq \frac{1}{2}(t-1) - \frac{R}{n^\alpha}\frac{t(t-1)}{2} = \frac{t-1}{2}\left(1 - \frac{R}{n^\alpha} t\right),$$

and since $t \leq \Upsilon_n = \Big(\frac{1}{2} - \varepsilon\Big)\frac{n^\alpha}{R}$,

$$\sum_{s=1}^{t-1} p_{1,s}^{\eta_n} \geq \frac{t-1}{2}\left(1 - (\frac{1}{2} - \varepsilon)\frac{t}{\Upsilon_n}\right) \geq \frac{t-1}{4},$$

so that

$$\sum_{t=1}^{\Upsilon_n}\big(\sum_{s=1}^{t-1} p_{1,s}^{\eta_n}\big)^2 \geq \frac{1}{16}\sum_{t=1}^{\Upsilon_n}(t-1)^2 =: A_n = \frac{\Upsilon_n^3}{48} + O(\Upsilon_n^2).$$

Injecting this equation in Equation (40),

$$\begin{aligned}
&\mathbb{P}_{\eta_n}\Big(A_n - \sum_{t=1}^{\Upsilon_n} N_{t-1}^2 \geq \frac{2}{5}\sqrt{2\log\frac{2\Upsilon_n}{y}}\Upsilon_n^{\frac{5}{2}} + \Upsilon_n^2\log\frac{2\Upsilon_n}{y}\Big) \\
&\leq \mathbb{P}_{\eta_n}\Big(\sum_{t=1}^{\Upsilon_n}\big(\sum_{s=1}^{t-1} p_{1,s}^{\eta_n}\big)^2 - \sum_{t=1}^{\Upsilon_n} N_{t-1}^2 \geq \frac{2}{5}\sqrt{2\log\frac{2\Upsilon_n}{y}}\Upsilon_n^{\frac{5}{2}} + \Upsilon_n^2\log\frac{2\Upsilon_n}{y}\Big) \leq y.
\end{aligned}$$

□

B. NUMERICAL ILLUSTRATIONS

The numerical illustrations have been realized using `R`. All the material necessary for the reproduction of the simulations of the article can be found on ArXiv or on GitHub : https://github.com/JulienAubert3/Exp3R. Let's explain the link between the different data files and the available algorithms.

- All the necessary functions are in the file `myfunctions.R`. It contains in particular the procedure `Exp3`, but also the functions which calculate the MLE for a constant learning rate and the truncated MLE for a decreasing learning rate.
- The algorithm `est_etaconstant_nveclarge.R` generates the data file `Data_eta03_nveclarge1.Rdata` containing the estimators for a constant learning rate necessary for Figure 2.
- The algorithm `Tmax_comp.R` generates the data file `Tmax_eta03_dec_K41.Rdata` containing the $\Upsilon_{\max}$ necessary for Figure 3.
- The algorithm `est_etadecreasing.R` generates the data file `Data_eta03_dec1.Rdata` or `Data_eta03_dec_K41.Rdata` (depending on the number of arms) containing the estimators for a decreasing learning rate respectively for 2 arms and 4 arms. These data are used in Figures 4 and 5.
- The Markdown file `Test_Exp3.Rmd` reproduces all the figures from the article for these different data files. To spare the reader lengthy computations, the output of the functions `est_etaconstant_nveclarge.R`, `Tmax_comp.R` and `est_etadecreasing.R`, is made available in the previous enumerated data files.