

Stochastic approximation algorithms for superquantiles estimation*

Bernard Bercu[†] Manon Costa[‡] Sébastien Gadat[§]

Abstract

This paper is devoted to two different two-time-scale stochastic approximation algorithms for superquantile, also known as conditional value-at-risk, estimation. We shall investigate the asymptotic behavior of a Robbins-Monro estimator and its convexified version. Our main contribution is to establish the almost sure convergence, the quadratic strong law and the law of iterated logarithm for our estimates via a martingale approach. A joint asymptotic normality is also provided. Our theoretical analysis is illustrated by numerical experiments on real datasets.

Keywords: stochastic approximation; quantile and superquantile; conditional value-at-risk; limit theorems.

MSC2020 subject classifications: Primary 62L20, Secondary 60F05; 62P05.

Submitted to EJP on July 28, 2020, final version accepted on May 22, 2021.

1 Introduction

Estimating quantiles has a longstanding history in statistics and probability. Except in parametric models where explicit formula are available, the estimation of quantiles is a real issue. The most common way to estimate quantiles is to make use of order statistics, see among other references [1, 13]. Another strategy is to make use of stochastic approximation algorithms and the pioneering work in this vein is the celebrated paper by Robbins and Monro [23].

Let X be an integrable continuous random variable with cumulative distribution function F and probability density function f . For some $\alpha \in]0, 1[$, the quantile θ_α of order α of F is the smallest solution of the equation

$$F(\theta_\alpha) = \mathbb{P}(X \leq \theta_\alpha) = \alpha, \tag{1.1}$$

*The authors gratefully acknowledge financial support from the Agence Nationale de la Recherche (MaSDOL grant ANR-19-CE23-0017).

[†]Université de Bordeaux, France. E-mail: bernard.bercu@math.u-bordeaux.fr

[‡]Université de Toulouse, France. E-mail: manon.costa@math.univ-toulouse.fr

[§]Université de Toulouse and Institut Universitaire de France, France. E-mail: sebastien.gadat@tse-fr.eu

whereas the superquantile ϑ_α of order α is defined by

$$\vartheta_\alpha = \mathbb{E}[X | X \geq \theta_\alpha] = \frac{\mathbb{E}[X I_{fX \geq \theta_\alpha}]}{\mathbb{P}(X \geq \theta_\alpha)} = \frac{\mathbb{E}[X I_{fX \geq \theta_\alpha}]}{1 - \alpha}. \tag{1.2}$$

One can observe that the superquantile provides more information on the tail of the distribution of the random variable X . Our goal in this paper is to simultaneously estimate quantiles and superquantiles, also respectively known as value-at-risk and conditional value-at-risk, which have become increasingly popular as measures of risk in finance [25, 26].

The paper is organized as follows. Section 2 is devoted to a brief overview of the previous literature on the recursive estimation of quantiles and superquantiles. The main results of the paper are given in Section 3. We propose the almost sure convergence of two-time-scale stochastic approximation algorithms for superquantile estimation. The quadratic strong law (QSL) as well as the law of iterated logarithm (LIL) of our stochastic algorithms are also provided. Moreover, we establish the joint asymptotic normality of our estimates. Numerical experiments on real data are given in Section 7. All technical proofs are postponed to Sections 5 and 6.

2 Overview of existing literature

A wide range of literature exists already on the recursive estimation of quantiles [23]. However, to the best of our knowledge, only a single paper is available on the recursive estimation of superquantiles [2]. In many practical situations where the data are recorded online with relatively high speed, or when the data are simply too numerous to be handled in batch systems, it is more suitable to implement a recursive strategy where quantiles and superquantiles are sequentially estimated with the help of stochastic approximation algorithms [9], [18]. We also refer the reader to [6, 7, 14, 15] for the online estimation of geometric medians and variances.

Bardou et al. [2] have previously studied the averaged version [22, 27] of a one-time-scale stochastic algorithm in order to estimate θ_α and ϑ_α . Here, we have chosen to investigate a two-time-scale stochastic algorithm [5, 11, 17, 20] which performs pretty well and offers more flexibility than the one-time-scale algorithm. Let (X_n) be a sequence of independent and identically distributed random variables sharing the same distribution as X . We shall extend the statistical analysis of [2] by studying the two-time-scale stochastic algorithm given, for all $n \geq 1$, by

$$\begin{aligned} \vartheta_{n+1} &= \vartheta_n - a_n I_{fX_{n+1} \leq \vartheta_n} - \alpha, \\ \vartheta_{n+1} &= \vartheta_n + b_n \frac{X_{n+1}}{1 - \alpha} I_{fX_{n+1} > \vartheta_n} - \vartheta_n, \end{aligned} \tag{2.1}$$

where the initial values ϑ_1 and ϑ_1 are deterministic real numbers that can be arbitrarily chosen and the steps (a_n) and (b_n) are two positive sequences of real numbers strictly smaller than one, decreasing towards zero such that

$$\sum_{n=1}^{\infty} a_n = +\infty, \quad \sum_{n=1}^{\infty} b_n = +\infty \quad \text{and} \quad \sum_{n=1}^{\infty} a_n^2 < +\infty, \quad \sum_{n=1}^{\infty} b_n^2 < +\infty. \tag{2.2}$$

We shall also investigate the asymptotic behavior of the convexified version of algorithm (2.1), based on the Rockafellar-Uryasev's identity [25] and given, for all $n \geq 1$, by

$$\begin{aligned} \vartheta_{n+1} &= \vartheta_n - a_n I_{fX_{n+1} \leq \vartheta_n} - \alpha \\ \vartheta_{n+1} &= \vartheta_n + b_n \left(\vartheta_n + \frac{(X_{n+1} - \vartheta_n)}{1 - \alpha} I_{fX_{n+1} > \vartheta_n} \right) - \vartheta_n, \end{aligned} \tag{2.3}$$

where as before the initial values x_1 and θ_1 are deterministic real numbers that can be arbitrarily chosen. We also refer the reader to the original contribution [3] where this convexification first appeared. The almost sure convergence

$$\lim_{n \rightarrow \infty} \bar{x}_n = \theta \quad \text{a.s.} \quad (2.4)$$

is a famous result that was established by Robbins and Monro [23], Robbins and Siegmund [24]. Moreover, the asymptotic normality is due to Sacks, see Theorem 1 in [28]. It requires the additional assumption that the probability density function f is differentiable with bounded derivative in every neighborhood of the quantile θ . More precisely, if the step $a_n = a_1/n$ where $a_1 > 0$ and $2a_1 f'(\theta) > 1$, we have the asymptotic normality

$$P\left(\frac{\bar{x}_n - \theta}{\sqrt{\frac{a_1^2 (1 - \theta)}{2a_1 f'(\theta) - 1}}}\right) \xrightarrow{D} N(0, 1) \quad (2.5)$$

One can observe that in the special case where the value $f'(\theta) > 0$ is known, it is possible to minimise the previous limiting variance by choosing $a_1 = 1/f'(\theta)$ and to obtain from (2.5) the asymptotic efficiency

$$P\left(\frac{\bar{x}_n - \theta}{\sqrt{\frac{1}{f'^2(\theta)}}}\right) \xrightarrow{D} N(0, 1)$$

Moreover, if the steps $a_n = a_1/n$ where $2a_1 f'(\theta) = 1$, we have from Theorem 2 in [19] the asymptotic normality

$$P\left(\frac{\bar{x}_n - \theta}{\sqrt{\frac{1}{\log n}}}\right) \xrightarrow{D} N(0, a_1^2 (1 - \theta)) \quad (2.6)$$

Some useful refinements on the asymptotic behavior of the sequence (\bar{x}_n) are also well-known. The LIL was first proved by Gaposkin and Krasulina, see Theorem 1 in [12] and Corollary 1 in [16]. More precisely, if the step $a_n = a_1/n$ where $2a_1 f'(\theta) > 1$, we have the LIL

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{n^{1/2} (\bar{x}_n - \theta)}{2 \log \log n} &= \liminf_{n \rightarrow \infty} \frac{n^{1/2} (\bar{x}_n - \theta)}{2 \log \log n} \\ &= \frac{a_1^2 (1 - \theta)}{2a_1 f'(\theta) - 1} \quad \text{a.s.} \end{aligned} \quad (2.7)$$

In particular, it follows from (2.7) that

$$\limsup_{n \rightarrow \infty} \frac{n (\bar{x}_n - \theta)^2}{2 \log \log n} = \frac{a_1^2 (1 - \theta)}{2a_1 f'(\theta) - 1} \quad \text{a.s.} \quad (2.8)$$

which is the limiting variance in (2.5). The QSL is due to Lai and Robbins, see Lemma 1 and Theorem 2 in [19] as well as Theorem 3 in [21]. More precisely, they proved that

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n X_k^2 = \frac{a_1^2 (1 - \theta)}{2a_1 f'(\theta) - 1} \quad \text{a.s.} \quad (2.9)$$

Besides the classical choice $a_n = a_1/n$ where $a_1 > 0$, slower step-size $a_n = a_1/n^a$ where $a_1 > 0$ and $1/2 < a < 1$ have been studied in depth. We refer the reader to pioneer work of Chung [8] and to Fabian [10] who obtained that the asymptotic normality still holds for the Robbins-Monro algorithm. More precisely, if $f'(\theta) > 0$, they showed that

$$P\left(\frac{\bar{x}_n - \theta}{\sqrt{\frac{a_1 (1 - \theta)}{2f'(\theta)}}}\right) \xrightarrow{D} N(0, 1) \quad (2.10)$$

In addition, it follows from Lai and Robbins [19] or Pelletier [21] that

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{n^a}{2(1-a)\log n} &= \liminf_{n \rightarrow \infty} \frac{n^a}{2(1-a)\log n} \\ &= \frac{a_1 (1 - \frac{1}{2})}{2f(\frac{1}{2})} \quad \text{a.s.} \end{aligned} \quad (2.11)$$

In particular,

$$\limsup_{n \rightarrow \infty} \frac{n^a}{2(1-a)\log n} = \frac{a_1 (1 - \frac{1}{2})}{2f(\frac{1}{2})} \quad \text{a.s.} \quad (2.12)$$

Moreover, we also have from [19], [21] that

$$\lim_{n \rightarrow \infty} \frac{1}{n^{1-a}} \sum_{k=1}^n X_k^a = \frac{a_1 (1 - \frac{1}{2})}{2(1-a)f(\frac{1}{2})} \quad \text{a.s.} \quad (2.13)$$

The restrictive assumption $2a_1f(\frac{1}{2}) > 1$, which involves the knowledge of $f(\frac{1}{2})$, is no longer needed. However, the convergence rate n^a is always slower than n , which means that the choice $a_n = a_1/n$ theoretically outperforms the one of $a_n = a_1/n^a$, at least asymptotically.

In the special case of the one-time-scale stochastic algorithm where $a_n = b_n$, Bardou et al. [2] proved the almost sure convergences

$$\lim_{n \rightarrow \infty} \bar{a}_n = \frac{1}{2} \quad \text{and} \quad \lim_{n \rightarrow \infty} \bar{\#}_n = \frac{1}{2} \quad \text{a.s.} \quad (2.14)$$

using an extended version of Robbins-Monro theorem together with Cesaro and Kronecker lemmas, see e.g. Theorem 1.4.26 in [9]. They also state without proof that

$$\lim_{n \rightarrow \infty} \bar{\#}_n = \frac{1}{2} \quad \text{a.s.} \quad (2.15)$$

Yet, other almost sure asymptotic properties for the sequences (\bar{a}_n) and $(\bar{\#}_n)$, such as the LIL and the QSL, are still missing. Bardou et al. also established in Theorem 2.4 of [2] the joint asymptotic normality of the averaged version [22, 27] of their one-time-scale stochastic algorithm

$$P \left(\frac{\bar{a}_n - \frac{1}{2}}{\sqrt{\frac{1}{n}}} \leq \frac{\bar{\#}_n - \frac{1}{2}}{\sqrt{\frac{1}{n}}} \leq \frac{1}{\sqrt{n}} \right) \rightarrow 0; \quad (2.16)$$

where the asymptotic covariance matrix is explicitly calculated,

$$\bar{\Sigma}_n = \frac{1}{n} \sum_{k=1}^n \Sigma_k \quad \text{and} \quad \bar{\#}_n = \frac{1}{n} \sum_{k=1}^n \#_k$$

We will show that our two-time-scale stochastic algorithms given by (2.1) and (2.3) allow us to avoid the Ruppert and Polyak-Juditsky averaging principle. Moreover, they perform pretty well both from a theoretical and a practical point of view and offer more flexibility than the one-time-scale stochastic algorithm.

3 Main results

In order to state our main results, it is necessary to introduce some assumptions.

- (A₁) The probability density function f is differentiable with bounded derivative in every neighborhood of $\frac{1}{2}$ and such that $f(\frac{1}{2}) > 0$.

(A₂) The function ψ defined, for all $x \in \mathbb{R}$, by $\psi(x) = f(x) + f'(x)$ is bounded in every neighborhood of g .

Our first result concerns the basic almost sure convergence of the two-time-scale stochastic algorithms (2.1) and (2.3) to the superquantile $\#$.

Theorem 3.1. Assume that (A₁) holds and that the random variable X is square integrable. Then, we have the almost sure convergences

$$\lim_{n \rightarrow \infty} \frac{\#_n}{n} = \# \quad \text{a.s.} \quad (3.1)$$

$$\lim_{n \rightarrow \infty} \frac{\#_n}{a_n} = \# \quad \text{a.s.} \quad (3.2)$$

Our proof is slightly different from that of Bardou et al. [12] established for the one-time-scale stochastic algorithm where $a_n = b_n$. It can be found in Section 5.1 for sake of completeness. We now focus our attention on the almost sure rates of convergence of the sequences $(\#_n)$ and $(\#_n)$. We divide our analysis into two parts depending on the step size (b_n) in the superquantile recursive procedure. First of all, we shall consider the optimal step $b_n = b_1 n$. Then, we shall study the case where $b_n = b_1 n^b$ with $1/2 < b < 1$. For all $x \in \mathbb{R}$, denote

$$\sigma^2(x) = \frac{1}{(1-x)^2} \text{Var}(X |_{f(x) > g}) \quad \text{and} \quad \tau^2(x) = \frac{1}{(1-x)^2} \text{Var}((X - g) |_{f(x) > g}): \quad (3.3)$$

It follows from straightforward calculation that

$$\sigma^2(\frac{\#_n}{n}) = \sigma^2(\frac{\#}{n}) \frac{1}{1 - \frac{\#}{n}} \quad (2\#):$$

Consequently, as soon as $\frac{\#}{n} > 0$, we always have $\sigma^2(\frac{\#_n}{n}) > \sigma^2(\frac{\#}{n})$ since $\frac{\#}{n} < \frac{\#}{n}$.

Theorem 3.2. Assume that (A₁) and (A₂) hold and that the random variable X has a moment of order strictly greater than 2. Moreover, suppose that the step sequences (a_n) and (b_n) are given by

$$a_n = \frac{a_1}{n^a} \quad \text{and} \quad b_n = \frac{b_1}{n}$$

where $a_1 > 0, b_1 > 1/2$ and $1/2 < a < 1$. Then, $(\#_n)$ and $(\#_n)$ share the same QSL

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n \frac{\#_k}{n} \#^2 = \frac{b_1^2}{2b_1 - 1} \sigma^2(\frac{\#}{n}) \quad \text{a.s.} \quad (3.4)$$

In addition, they also share the same LIL

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{n}{2 \log \log n} \frac{\#_n}{n} \#^{1/2} &= \liminf_{n \rightarrow \infty} \frac{n}{2 \log \log n} \frac{\#_n}{n} \#^{1/2} \\ &= \frac{b_1^2}{2b_1 - 1} \sigma(\frac{\#}{n}) \quad \text{a.s.} \end{aligned} \quad (3.5)$$

In particular,

$$\limsup_{n \rightarrow \infty} \frac{n}{2 \log \log n} \frac{\#_n}{n} \#^2 = \frac{b_1^2}{2b_1 - 1} \sigma^2(\frac{\#}{n}) \quad \text{a.s.}$$

Remark 3.1. In the special case where the step sequence (b_n) is given by

$$b_n = \frac{1}{n+1};$$

it is easy to see that $\hat{\mu}_n$ and $\hat{\mu}_n^*$ both reduce to

$$\hat{\mu}_n = \frac{1}{n} \sum_{k=1}^n \frac{X_k}{1} \mathbb{1}_{\{X_k > k^{-1}\}}$$

and

$$\hat{\mu}_n^* = \frac{1}{n} \sum_{k=1}^n \frac{X_k}{k+1} + \frac{1}{n} \sum_{k=1}^n \frac{X_k}{k-1} \mathbb{1}_{\{X_k > k^{-1}\}};$$

In this setting, we immediately obtain from Theorem 3.2 that

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n \hat{\mu}_k^* \#^2 = \frac{1}{2} \quad \text{a.s.}$$

and

$$\limsup_{n \rightarrow \infty} \frac{n}{2 \log \log n} \hat{\mu}_n \#^2 = \frac{1}{2} \quad \text{a.s.}$$

Theorem 3.3. Assume that (A_1) and (A_2) hold and that the random variable X has a moment of order strictly greater than 2. Moreover, suppose that the step sequences (a_n) and (b_n) are given by

$$a_n = \frac{a_1}{n^a} \quad \text{and} \quad b_n = \frac{b_1}{n^b}$$

where $a_1 > 0, b_1 > 0$ and $1=2 < a < b < 1$. Then, $(\hat{\mu}_n)$ and $(\hat{\mu}_n^*)$ share the same QSL

$$\lim_{n \rightarrow \infty} \frac{1}{n^{1-b}} \sum_{k=1}^n \hat{\mu}_k \#^2 = \frac{b_1}{2(1-b)} \quad \text{a.s.} \quad (3.6)$$

In addition, they also share the same LIL

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{n^b}{2(1-b) \log n} \hat{\mu}_n \#^{1=2} &= \liminf_{n \rightarrow \infty} \frac{n^b}{2(1-b) \log n} \hat{\mu}_n \#^{1=2} \\ &= \frac{b_1}{2} \quad \text{a.s.} \end{aligned} \quad (3.7)$$

In particular,

$$\limsup_{n \rightarrow \infty} \frac{n^b}{2(1-b) \log n} \hat{\mu}_n \#^2 = \frac{b_1}{2} \quad \text{a.s.}$$

Remark 3.2. Similar computations in the case where $1=2 < b < a < 1$ would lead to the same results for the convexified algorithm $(\hat{\mu}_n^*)$. However, for the standard algorithm $(\hat{\mu}_n)$, it is necessary to replace the asymptotic variance $\frac{1}{2}$ by $\frac{1}{2(1-b)}$. This emphasizes the interest of using the convexified algorithm.

We now focus our attention on the asymptotic normality of our two-time-scale stochastic algorithms (2.1) and (2.3).

Theorem 3.4. Assume that (A_1) and (A_2) hold and that the random variable X has a moment of order strictly greater than 2. Moreover, suppose that the step sequences (a_n) and (b_n) are given by

$$a_n = \frac{a_1}{n^a} \quad \text{and} \quad b_n = \frac{b_1}{n^b}$$

where $a_1 > 0, b_1 > 0$ and $1/2 < a < b < 1$ with $b_1 > 1/2$ if $b = 1$. Then, $(\hat{\mu}_n)$ and $(\hat{\sigma}_n)$ share the same joint asymptotic normality

$$\begin{pmatrix} \sqrt{n}(\hat{\mu}_n - \mu) \\ \sqrt{n}(\hat{\sigma}_n - \sigma) \end{pmatrix} \xrightarrow{L} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \right) \quad (3.8)$$

where the asymptotic variances are given by

$$\sigma^2 = \frac{a_1(1 - \mu)}{2f(\mu)}$$

and

$$\sigma^2 = \begin{cases} \frac{b_1^2(1 - \mu)}{2b_1 - 1} & \text{if } b = 1; \\ \frac{b_1(1 - \mu)}{2} & \text{if } b < 1; \end{cases}$$

Remark 3.3. One can observe that the asymptotic covariance matrix in (3.8) is diagonal. It means that, at the limit, the two algorithms for quantile and superquantile estimation are no longer correlated. This is due to the fact that we use two different time scales contrary to Bardou et al. [2]. Moreover, in the special case where $b = 1$, we also recover the same asymptotic variance as the one obtained in [2] for the averaged version of their one-time-scale stochastic algorithm.

Remark 3.4. The asymptotic variance $\sigma^2(\mu)$ can be estimated by

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{k=1}^n \frac{X_k - \hat{\mu}_n}{1 - \hat{\mu}_n} \frac{1}{1 - \hat{\mu}_n} \mathbb{1}_{\{X_k > \hat{\mu}_n\}} \frac{1}{n} \sum_{k=1}^n \frac{X_k - \hat{\mu}_n}{1 - \hat{\mu}_n} \mathbb{1}_{\{X_k > \hat{\mu}_n\}}^2$$

Via the same lines as in the proof of the almost sure convergences (3.1) and (3.2), one can verify that $\hat{\sigma}_n^2 \xrightarrow{a.s.} \sigma^2(\mu)$. Therefore, using Slutsky's Theorem, we deduce from (3.8) that $(\hat{\mu}_n)$ and $(\hat{\sigma}_n)$ share the same asymptotic normality

$$\begin{pmatrix} \sqrt{n}(\hat{\mu}_n - \mu) \\ \sqrt{n}(\hat{\sigma}_n - \sigma) \end{pmatrix} \xrightarrow{L} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \right) \quad (3.9)$$

where

$$\sigma^2 = \begin{cases} \frac{b_1^2(1 - \mu)}{2b_1 - 1} & \text{if } b = 1; \\ \frac{b_1(1 - \mu)}{2} & \text{if } b < 1; \end{cases}$$

Convergence (3.9) allows us to construct asymptotic confidence intervals for the superquantile $\hat{\mu}_n$.

4 Our martingale approach

All our analysis relies on a decomposition of our estimates as sum of a martingale increment and a drift term. More precisely, it follows from (2.1) and (2.3) that for all $n \geq 1$,

$$\begin{aligned} \hat{\mu}_{n+1} &= (1 - b_n)\hat{\mu}_n + b_n Y_{n+1} \\ \hat{\sigma}_{n+1} &= (1 - b_n)\hat{\sigma}_n + b_n Z_{n+1} \end{aligned} \quad (4.1)$$

where

$$Y_{n+1} = \frac{X_{n+1}}{1} I_{\{X_{n+1} > g\}}$$

and

$$Z_{n+1} = g + \frac{(X_{n+1} - g)}{1} I_{\{X_{n+1} > g\}}$$

Let $H(\cdot)$ and $L(\cdot)$ be the functions defined, for all $x \in \mathbb{R}$, by

$$H(x) = \frac{1}{1} E[X I_{\{X > g\}}] \quad \text{and} \quad L(x) = g + \frac{1}{1} E[(X - g) I_{\{X > g\}}] \quad (4.2)$$

Denote by \mathcal{F}_n the σ -algebra generated by the random variables $X_1; \dots; X_n$. It follows from the fact g is \mathcal{F}_n -measurable together with the fact that X_{n+1} is independent of \mathcal{F}_n that almost surely

$$E[Y_{n+1} | \mathcal{F}_n] = H(g) \quad \text{and} \quad E[Z_{n+1} | \mathcal{F}_n] = L(g):$$

It allows us to split Y_{n+1} and Z_{n+1} as sum of a martingale increment and a drift term, $Y_{n+1} = M_{n+1} + H(g)$ and $Z_{n+1} = N_{n+1} + L(g)$. One can also verify that $E[M_{n+1}^2 | \mathcal{F}_n] = \sigma^2(g)$ and $E[N_{n+1}^2 | \mathcal{F}_n] = \sigma^2(g)$ where the two variances are given by (3.3). Then, we immediately deduce from (4.1) that for all $n \geq 1$,

$$\begin{aligned} M_{n+1} &= (1 - b_n) M_n + b_n (Y_{n+1} - H(g)) \\ N_{n+1} &= (1 - b_n) N_n + b_n (Z_{n+1} - L(g)): \end{aligned} \quad (4.3)$$

Hereafter, assume for the sake of simplicity that for all $n \geq 1$, $b_n < 1$, since this is true for n large enough. Let (P_n) be the increasing sequence of positive real numbers defined by

$$P_n = \prod_{k=1}^n (1 - b_k)^{-1} \quad (4.4)$$

with the convention that $P_0 = 1$. Since $(1 - b_n) P_n = P_{n-1}$, we obtain from (4.3) that

$$\begin{aligned} M_{n+1} &= P_{n-1}^{-1} M_n + P_n b_n (Y_{n+1} - H(g)) \\ N_{n+1} &= P_{n-1}^{-1} N_n + P_n b_n (Z_{n+1} - L(g)) \end{aligned}$$

which implies the martingale decomposition

$$\begin{aligned} M_{n+1} &= \frac{1}{P_n} M_1 + M_{n+1} + H_{n+1} \\ N_{n+1} &= \frac{1}{P_n} N_1 + N_{n+1} + L_{n+1} \end{aligned} \quad (4.5)$$

where

$$M_{n+1} = \sum_{k=1}^n b_k P_k^{-1} M_{k+1}; \quad N_{n+1} = \sum_{k=1}^n b_k P_k^{-1} N_{k+1} \quad (4.6)$$

$$H_{n+1} = \sum_{k=1}^n b_k P_k^{-1} H(g); \quad L_{n+1} = \sum_{k=1}^n b_k P_k^{-1} L(g): \quad (4.7)$$

Our strategy is to establish the asymptotic behavior of the two martingales (M_n) and (N_n) as well as to determine the crucial role played by the two drift terms (H_n) and (L_n) . Several results in our analysis rely on the following keystone lemma which concerns the convexity properties of the functions H and L defined in (4.2).

Lemma 4.1. Assume that (A_1) and (A_2) hold. Then, L is a convex function such that $L(x) = \#$, $L^0(x) = 0$ and that for all $x \in \mathbb{R}$,

$$L^0(x) = L(x) - \frac{f(x)}{1} \tag{4.8}$$

In addition, we also have $H(x) = \#$ and that for all $x \in \mathbb{R}$,

$$H(x) = H(x) + \frac{f(x)}{1} \tag{4.9}$$

Proof. It follows from (4.2) that for all $x \in \mathbb{R}$,

$$L^0(x) = \frac{F(x)}{1} \quad \text{and} \quad L^{0q}(x) = \frac{f(x)}{1}$$

Consequently, L is a convex function such that $L^0(x) = 0$. Hence, we deduce from a Taylor expansion with integral remainder that for all $x \in \mathbb{R}$,

$$L(x) = L(x) + \int_0^x (1-t)^2 L^{0q}(t) dt$$

where $t = x + t(x)$, which immediately leads to (4.8) using (A_1) . Unfortunately, H is not a convex function. However, we obtain from (4.2) that for all $x \in \mathbb{R}$,

$$H^0(x) = \frac{f(x)}{1} \quad \text{and} \quad H^{0q}(x) = \frac{f(x)}{1}$$

where $(x) = f(x) + f^0(x)$. Finally, (4.9) follows once again from a Taylor expansion with integral remainder together with (A_2) . □

We have just seen that the function H is not convex. Consequently, in order to prove sharp asymptotic properties for the sequence $(\#_n)$, it is necessary to slightly modify the first martingale decomposition in (4.5). For all $x \in \mathbb{R}$, let

$$\begin{aligned} G(x) &= F(x) - f(x); \\ R(x) &= H(x) - \# + C(x) \end{aligned} \tag{4.10}$$

where

$$C(x) = H^0(x) = \frac{f(x)}{1}$$

We deduce from (2.1), (4.3) and (4.10) that for all $n \geq 1$,

$$\begin{aligned} \#_{n+1} &= (1 - a_n f(x)) (\#_n) - a_n V_{n+1} + G(x); \\ \#_{n+1} - \# &= (1 - b_n) (\#_n - \#) + b_n W_{n+1} + R(x) - C(x) \end{aligned} \tag{4.11}$$

with $V_{n+1} = \int_{\#_n}^{\#_{n+1}} f(x) dx$. Hereafter, we shall consider a tailor-made weighted sum of our estimates given by $\#_1 = 0$ and, for all $n \geq 2$,

$$\#_n = (\#_n - \#) + \# \tag{4.12}$$

where $(\#_n)$ is a deterministic sequence, depending on (a_n) and (b_n) , which will be explicitly given below. It follows from (4.11) together with straightforward calculation that for all $n \geq 2$,

$$\#_{n+1} = (1 - b_n) \#_n + b_n W_{n+1} + R(x) + \frac{a_n}{b_n} \#_{n+1} G(x) + \#_{n+1} (x) \tag{4.13}$$

where

$$W_{n+1} = W_n + \frac{a_n}{b_n} V_{n+1} \tag{4.14}$$

and

$$W_{n+1} = \frac{1}{b_n} (1 - b_n) W_n + C b_n W_{n+1} (1 - a_n f(\cdot)) : \tag{4.15}$$

We have several strategies in order to simplify the expression of W_{n+1} . A first possibility that cancels several terms in (4.15) is to choose

$$W_{n+1} = \frac{C b_n}{f(\cdot) a_n} :$$

It clearly reduces W_{n+1} to

$$W_{n+1} = \frac{(1 - b_n) W_n}{b_n} :$$

Another more sophisticated choice, which only works if $a_n f(\cdot) b_n \notin 0$, is to take

$$W_{n+1} = \frac{C b_n}{a_n f(\cdot) b_n} : \tag{4.16}$$

It implies that

$$W_{n+1} = \frac{(1 - b_n)(W_n)}{b_n} : \tag{4.17}$$

The two choices are quite similar in the special case where $b_n = b_1 = n$. However, in the case where the step $b_n = b_1 = n^b$ with $a < b < 1$, the second choice outperforms the first one as W_n goes faster towards zero as n grows to infinity. Throughout the sequel, we shall make use of the second choice given by (4.16). We deduce from (4.13) the new martingale decomposition

$$W_{n+1} = \frac{1}{P_n} M_{n+1} + H_{n+1} + R_{n+1} \tag{4.18}$$

where

$$M_{n+1} = \sum_{k=1}^X b_k P_k W_{k+1}; \quad H_{n+1} = \sum_{k=1}^X b_k P_k W_{k+1} - k \tag{4.19}$$

and

$$R_{n+1} = \sum_{k=1}^X b_k P_k R(k) + \frac{a_k}{b_k} W_{k+1} G(k) : \tag{4.20}$$

5 Proofs of the almost sure convergence results

5.1 The basic almost sure properties.

The starting point in our analysis of the almost sure convergence of our estimates is the following lemma.

Lemma 5.1. Assume that (A_1) holds and that the random variable X is square integrable. Then, we have the almost sure convergences

$$\lim_{n \rightarrow \infty} \frac{M_{n+1}}{P_n} = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{N_{n+1}}{P_n} = 0 \quad \text{a.s.} \tag{5.1}$$

Proof. Let (M_n) and (N_n) be the two locally square integrable martingales

$$M_n = \sum_{k=1}^X b_k W_{k+1}; \quad N_n = \sum_{k=1}^X b_k W_{k+1} - k :$$

Their predictable quadratic variations [9] are respectively given by

$$h_{i_n} = \sum_{k=1}^{X-1} b_k^2 \mathbb{1}_{(k)} \quad \text{and} \quad h_{i_n} = \sum_{k=1}^{X-1} b_k^2 \mathbb{1}_{(k)}:$$

It follows from convergence (2.4) and the continuity of the variances $\mathbb{1}_{(k)}$ and $\mathbb{1}_{(k)}$ given by (3.3) that $\mathbb{1}_{(k)}$ and $\mathbb{1}_{(k)}$ a.s. Consequently, we get from the right-hand side of (2.2) that

$$\lim_{n \rightarrow \infty} h_{i_n} < +\infty \quad \text{and} \quad \lim_{n \rightarrow \infty} h_{i_n} < +\infty \quad \text{a.s.} \quad (5.2)$$

Therefore, we obtain from the strong law of large numbers for martingales given e.g. by theorem 1.3.24 in [9] that (i_n) and (i_n) both converge almost surely. The rest of the proof proceeds in a standard way with the help of Kronecker's lemma. As a matter of fact, we can deduce from the left-hand side of (2.2) that the sequence (P_n) , defined in (4.4), is strictly increasing to infinity. In addition, we just showed the almost sure convergence of the series

$$\sum_{n=1}^{\infty} b_n \mathbb{1}_{n+1} \quad \text{and} \quad \sum_{n=1}^{\infty} b_n \mathbb{1}_{n+1}:$$

Consequently, we immediately deduce from Kronecker's lemma that

$$\lim_{n \rightarrow \infty} \frac{1}{P_n} \sum_{k=1}^{X-1} b_k P_{k+1} = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{1}{P_n} \sum_{k=1}^{X-1} b_k P_{k+1} = 0 \quad \text{a.s.}$$

which is exactly what we wanted to prove. □

Proof of Theorem 3.1. We recall from (4.5) that for all $n \geq 1$,

$$\begin{aligned} \mathbb{1}_{n+1} &= \frac{1}{P_n} \mathbb{1}_1 + M_{n+1} + H_{n+1} \\ \mathbb{1}_{n+1} &= \frac{1}{P_n} \mathbb{1}_1 + N_{n+1} + L_{n+1} \end{aligned}$$

We have from (2.4) together with the continuity of the functions H and L that

$$\lim_{n \rightarrow \infty} H(i_n) = H(\cdot) \quad \text{a.s.} \quad (5.3)$$

and

$$\lim_{n \rightarrow \infty} L(i_n) = L(\cdot) \quad \text{a.s.} \quad (5.4)$$

One can observe that $H(\cdot) = L(\cdot) = \#$. Moreover, it is easy to see that for all $n \geq 1$, $b_n P_n = P_n - P_{n-1}$. Hence, we obtain by a telescoping argument that

$$\sum_{k=1}^{X-1} b_k P_k = P_n - P_0; \quad (5.5)$$

which leads to

$$\lim_{n \rightarrow \infty} \frac{1}{P_n} \sum_{k=1}^{X-1} b_k P_k = 1:$$

Therefore, it follows from Toeplitz's lemma that

$$\lim_{n \rightarrow \infty} \frac{H_{n+1}}{P_n} = \# \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{L_{n+1}}{P_n} = \# \quad \text{a.s.} \quad (5.6)$$

Finally, we find from (4.5), (5.1) and (5.6) that

$$\lim_{n \rightarrow \infty} \mathbb{1}_n = \# \quad \text{and} \quad \lim_{n \rightarrow \infty} \mathbb{1}_n = \# \quad \text{a.s.} \quad (5.7)$$

which completes the proof of Theorem 3.1. □

5.2 A keystone lemma.

The QSL as well as the LIL for our estimates require the sharp asymptotic behavior of the sequence (P_n) defined in (4.4). Surprisingly, to the best of our knowledge, the following keystone lemma is new. It involves the famous Euler-Riemann zeta function.

Lemma 5.2. Assume that for some $0 < b_1 < 1$,

$$P_n = \sum_{k=1}^n \left(1 - \frac{b_1}{k}\right)^{b_1} \quad (5.8)$$

Then, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n^{b_1}} P_n = \Gamma(1 - b_1) \quad (5.9)$$

where Γ stands for the Euler gamma function. Moreover, suppose that $1/2 < b_1 < 1$ and that

$$P_n = \sum_{k=1}^n \left(1 - \frac{b_1}{k^b}\right)^{b_1} \quad (5.10)$$

where $1/2 < b < 1$. Then, we have

$$\lim_{n \rightarrow \infty} \frac{1}{\exp(cn^{1-b})} P_n = \exp(-c) \quad (5.11)$$

with $c = b_1(1 - b)$ and the limiting value

$$= \sum_{n=2}^{\infty} \frac{b_1^n}{n} \zeta(bn)$$

where ζ stands for the Riemann zeta function.

Remark 5.1. The link between the first case $b = 1$ and the second case $1/2 < b < 1$ is given by the following formula due to Euler. For all $|x| < 1$,

$$\log(1 - x) = -x - \sum_{n=2}^{\infty} \frac{x^n}{n} \zeta(n)$$

where ζ is the Euler-Mascheroni constant.

Remark 5.2. The case $b_1 = 1$ can be treated in the same way. For example, concerning the first part of Lemma 5.2, it is only necessary to replace P_n defined in (5.8) by

$$P_n = \sum_{k=1+\{bb_1c\}}^n \left(1 - \frac{b_1}{k}\right)^{b_1}$$

where $\{bb_1c\}$ is the integer part of bb_1c . Then, we obtain that

$$\lim_{n \rightarrow \infty} \frac{1}{n^{b_1}} P_n = \frac{\Gamma(1 - b_1)}{\Gamma(1 - bb_1c)}$$

where $\Gamma(1 - b_1) / \Gamma(1 - bb_1c)$ stands for the fractional part of b_1 .

Proof. In the first case $b = 1$, we clearly have

$$P_n = \sum_{k=1}^n \left(1 - \frac{b_1}{k}\right)^{b_1} = \frac{(n+1)(1 - b_1)}{(n+1 - b_1)} \quad (5.12)$$

It is well-known that for any $c > 0$,

$$\lim_{n \rightarrow \infty} \frac{(n+c)^n}{(n)n^c} = 1: \tag{5.13}$$

Hence, we obtain from (5.12) and (5.13) that

$$\lim_{n \rightarrow \infty} \frac{1}{n^{b_1}} P_n = (1 - b_1): \tag{5.14}$$

The second case $1=2 < b < 1$ is much more difficult to handle. It follows from the Taylor expansion of the natural logarithm given for all $|x| < 1$ by

$$\log(1 - x) = -\sum_{k=1}^{\infty} \frac{x^k}{k}$$

that

$$\begin{aligned} \log(P_n) &= \sum_{k=1}^n \log\left(1 - \frac{b_1}{k^b}\right) = \sum_{k=1}^n \sum_{j=1}^{\infty} \frac{1}{j} \left(\frac{b_1}{k^b}\right)^j = \sum_{j=1}^{\infty} \frac{1}{j} \sum_{k=1}^n \frac{b_1^j}{k^{bj}}; \\ &= b_1 \sum_{k=1}^n \frac{1}{k^b} + \sum_{j=2}^{\infty} \frac{b_1^j}{j} \sum_{k=1}^n \frac{1}{k^{bj}}. \end{aligned} \tag{5.15}$$

It is well-known that

$$\lim_{n \rightarrow \infty} \frac{1}{n^{1-b}} \sum_{k=1}^n \frac{1}{k^b} = \frac{1}{1-b}.$$

In addition, as $b > 1=2$, we always have for all $n \geq 2$, $b^j > 1$. Consequently,

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{k^{bj}} = \zeta(b^j)$$

where ζ is the Riemann zeta function. Therefore, we obtain from (5.15) that

$$\lim_{n \rightarrow \infty} \frac{1}{\exp(cn^{1-b})} P_n = \exp(-c) \tag{5.16}$$

where $c = b_1(1 - b)$ and the limiting value

$$= \sum_{j=2}^{\infty} \frac{b_1^j}{j} \zeta(b^j): \quad \square$$

5.3 The fast step size case.

The proof of Theorem 3.2 relies on the following lemma which provides the QSL and the LIL for the martingales (M_n) and (N_n) .

Lemma 5.3. Assume that the step sequences (a_n) and (b_n) are given by

$$a_n = \frac{a_1}{n^a} \quad \text{and} \quad b_n = \frac{b_1}{n}$$

where $a_1 > 0$, $b_1 > 1=2$ and $1=2 < a < 1$. Then, (M_n) and (N_n) share the same QSL

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n \frac{M_k^2}{P_{k-1}} = \frac{b_1^2}{2b_1 - 1} \zeta(2) \quad \text{a.s.} \tag{5.17}$$

In addition, they also share the same LIL

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{n}{2 \log \log n} \frac{M_n^{1=2}}{P_{n-1}} &= \liminf_{n \rightarrow \infty} \frac{n}{2 \log \log n} \frac{M_n^{1=2}}{P_{n-1}} \\ &= \frac{b_1^2}{2b_1 - 1} \quad () \quad \text{a.s.} \end{aligned} \quad (5.18)$$

Proof. We first focus our attention on the martingale (M_n) defined by

$$M_{n+1} = \prod_{k=1}^n b_k P_k W_{k+1}$$

where

$$W_{n+1} = V_{n+1} + \frac{a_n}{b_n} V_n$$

with $V_{n+1} = I_{F(X_{n+1}) \leq F(n)}$. We clearly have $E[V_{n+1} | F_n] = 0$, $E[W_{n+1} | F_n] = 0$, and $E[V_{n+1}^2 | F_n] = F(n)(1 - F(n))$, $E[W_{n+1}^2 | F_n] = \frac{2}{n} F(n)$ where for all $n \in \mathbb{N}$,

$$\frac{2}{n} F(n) = \frac{2}{n} F(n) + \frac{a_n}{b_n} \frac{2}{n} F(n)(1 - F(n)) - \frac{2a_n}{b_n} F(n)H(n): \quad (5.19)$$

We obtain from (4.16) that

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = \frac{1}{1}: \quad (5.20)$$

Consequently, we infer from (2.4), (5.19) and (5.20) that

$$\lim_{n \rightarrow \infty} \frac{2}{n} F(n) = \frac{2}{1} (2#) = \frac{2}{1} \quad \text{a.s.} \quad (5.21)$$

Hereafter, assume for the sake of simplicity that $1=2 < b_1 < 1$ inasmuch as the proof follows exactly the same lines for $b_1 = 1$. On the one hand, the predictable quadratic variation of (M_n) is given by

$$hMi_n = \sum_{k=1}^{n-1} b_k^2 P_k^2 \frac{2}{k} F(k):$$

On the other hand, as $b_1 > 1=2$, we obtain from convergence (5.9) in Lemma 5.2 that

$$\lim_{n \rightarrow \infty} \frac{1}{n^{2b_1 - 1}} \sum_{k=1}^{n-1} b_k^2 P_k^2 = \frac{b_1^2}{2b_1 - 1} \frac{2}{1} (1 - b_1):$$

Then, we deduce from (5.21) and Toeplitz's lemma that

$$\lim_{n \rightarrow \infty} \frac{1}{n^{2b_1 - 1}} hMi_n = \frac{b_1^2}{2b_1 - 1} \frac{2}{1} (1 - b_1) \frac{2}{1} \quad \text{a.s.} \quad (5.22)$$

Denote by f_n the explosion coefficient associated with the martingale (M_n) ,

$$f_n = \frac{hMi_n - hMi_{n-1}}{hMi_n}:$$

We obtain from (5.22) that

$$\lim_{n \rightarrow \infty} n f_n = 2b_1 - 1 \quad \text{a.s.} \quad (5.23)$$

It means that f_n converges to zero almost surely at rate n^{-1} . In addition, we already saw from (5.21) that

$$\lim_{n \rightarrow \infty} E[W_{n+1}^2 | F_n] = \frac{b_1^2}{2b_1 - 1} \quad \text{a.s.}$$

Furthermore, we assume that the random variable X has a moment of order p strictly greater than 2. It implies that

$$\sup_n E[|W_{n+1}|^p | F_n] < 1 \quad \text{a.s.}$$

Consequently, we deduce from the QSL for martingales given in theorem 3 of [4] that

$$\lim_{n \rightarrow \infty} \frac{1}{\log h M_n} \sum_{k=1}^n \frac{X_k^2}{h M_k^2} = 1 \quad \text{a.s.} \quad (5.24)$$

Hence, we obtain from the conjunction of (5.22) and (5.24) that

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n \frac{X_k^2}{P_{k-1}^2} = \frac{b_1^2}{2b_1 - 1} \quad \text{a.s.} \quad (5.25)$$

We shall now proceed to the proof of the LIL given by (5.18). We find from (5.23) that the explosion coefficient f_n satisfies

$$\sum_{n=1}^{\infty} f_n^{p-2} < +\infty \quad \text{a.s.}$$

Therefore, we deduce from the LIL for martingales [29], see also corollary 6.4.25 in [9] that

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{2h M_n \log \log h M_n} M_n^{1/2} &= \liminf_{n \rightarrow \infty} \frac{1}{2h M_n \log \log h M_n} M_n^{1/2} \\ &= 1 \quad \text{a.s.} \end{aligned} \quad (5.26)$$

Hence, it follows from the conjunction of (5.14), (5.22) and (5.26) that

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{n}{2 \log \log n} \frac{M_n^{1/2}}{P_{n-1}} &= \liminf_{n \rightarrow \infty} \frac{n}{2 \log \log n} \frac{M_n^{1/2}}{P_{n-1}} \\ &= \frac{b_1^2}{2b_1 - 1} \quad \text{a.s.} \end{aligned}$$

which is exactly what we wanted to prove. Finally, concerning the martingale (N_n) given by

$$N_{n+1} = \sum_{k=1}^n b_k P_{k+1};$$

the only minor change is that $E[\sum_{n+1}^2 | F_n] = \frac{b_1^2}{2b_1 - 1} (n)$. However, we already saw that $\frac{b_1^2}{2b_1 - 1} (n) \sim \frac{b_1^2}{2b_1 - 1} (n)$ a.s. Consequently, (M_n) and (N_n) share the same QSL and the same LIL, which completes the proof of Lemma 5.3. \square

Proof of Theorem 3.2. We shall only prove Theorem 3.2 in the special case where $b_n = b_1 = n$ with $1/2 < b_1 < 1$ inasmuch as the proof in the case $b_1 < 1$ follows essentially the same lines. First of all, we focus our attention on the standard estimator \hat{p}_n .

Our strategy is first to establish the QSL for the sequence (\hat{p}_n) given by (4.12) and then to come back to \hat{p}_n . We recall from (4.18) that for all $n \geq 2$,

$$\hat{p}_{n+1} = \frac{1}{P_n} M_{n+1} + H_{n+1} + R_{n+1} :$$

We claim that the weighted sequence (X_n) satisfies the QSL

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n X_k^2 = \frac{b_1^2}{2b_1 - 1} \quad \text{a.s.} \quad (5.27)$$

As a matter of fact, we already saw from (5.17) that

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n \frac{M_k}{P_{k-1}} = \frac{b_1^2}{2b_1 - 1} \quad \text{a.s.}$$

Hence, in order to prove (5.27), it suffices to show that

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n \frac{H_k}{P_{k-1}} = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n \frac{R_k}{P_{k-1}} = 0 \quad \text{a.s.} \quad (5.28)$$

On the one hand, it follows from (2.12) that for n large enough and for all $k \leq n$,

$$k^2 \leq 2D_a \frac{\log k}{k^a} \quad \text{a.s.} \quad (5.29)$$

where

$$D_a = \frac{2a_1(1-a)(1-\alpha)}{f(\alpha)}$$

Consequently, we obtain from (4.19) and (5.29) that

$$|H_{n+1}| = O \left(\sum_{k=1}^n \frac{b_k P_{k+1}^p \log k}{k^{a-2}} \right) \quad \text{a.s.} \quad (5.30)$$

Furthermore, one can easily check from (4.16) and (4.17) that

$$\lim_{n \rightarrow \infty} n^{1-a} P_{n+1} = \frac{(1-a)C}{a_1 f(\alpha)}$$

In addition, we also recall from convergence (5.9) in Lemma 5.2 that

$$\lim_{n \rightarrow \infty} \frac{1}{n^{b_1}} P_n = (1 - b_1)$$

Hence, we deduce from (5.30) that

$$|H_{n+1}| = O \left(\sum_{k=1}^n \frac{p \log k}{k^{2-b_1-a-2}} \right) \quad \text{a.s.} \quad (5.31)$$

It follows from (5.31) that

$$\sum_{n=1}^{\infty} \frac{H_n}{P_{n-1}} < +\infty \quad \text{a.s.} \quad (5.32)$$

As a matter of fact, let $d = 2 - b_1 - a = 2$. If $d > 1$ that is $b_1 < 1 - a = 2$, we obtain from (5.31) that $|H_n| = O(1)$ a.s. Consequently, as $b_1 > 1 - a = 2$, (5.32) holds true. In addition, if $d = 1$ that is $b_1 = 1 - a = 2$, we deduce from (5.31) that $|H_n| = O((\log n)^{3-2})$ a.s. which implies that

$$\sum_{n=1}^{\infty} \frac{H_n}{P_{n-1}} = O \left(\sum_{n=1}^{\infty} \frac{(\log n)^3}{n^{2b_1}} \right) = O(1) \quad \text{a.s.}$$

Moreover, if $d < 1$ that is $b_1 > 1$ $a=2$, we get from (5.31) that $|jH_n| = O(\log n)^{1-2}n^{1-d}$ a.s. leading to

$$\sum_{n=1}^{\infty} \frac{H_n^2}{P_{n-1}} = O \sum_{n=1}^{\infty} \frac{(\log n)^2}{n^{2b_1+2d-2}} = O \sum_{n=1}^{\infty} \frac{(\log n)^2}{n^{2-a}} = O(1) \quad \text{a.s.}$$

On the other hand, (4.20) together with (4.9) and (4.10) imply that

$$|jR_{n+1}| = O \sum_{k=1}^n b_k P_k^2 \quad \text{a.s.} \quad (5.33)$$

It follows from (5.29) and (5.33) that

$$|jR_{n+1}| = O \sum_{k=1}^n \frac{\log k}{k^{1+a-b_1}} \quad \text{a.s.} \quad (5.34)$$

which clearly leads to

$$\sum_{n=1}^{\infty} \frac{R_n^2}{P_{n-1}} < +\infty \quad \text{a.s.} \quad (5.35)$$

Therefore, we obtain from (5.32) and (5.35) that the two convergences in (5.28) hold true, which immediately implies (5.27). Hereafter, one can notice from (4.12) that

$$\sum_{k=1}^n \frac{1}{k^2} = \sum_{k=1}^n \frac{1}{k^2} + \sum_{k=1}^n \frac{1}{k^2} + 2 \sum_{k=1}^n \frac{1}{k^2} : \quad (5.36)$$

Hence, in order to prove (3.4), it suffices to show that

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n \frac{1}{k^2} = 0 \quad \text{a.s.}$$

and to make use of the Cauchy-Schwarz inequality. Denote

$$S_n = \sum_{k=1}^n \frac{1}{k^2}.$$

We have from (2.13) that as soon as $f(x) > 0$,

$$\lim_{n \rightarrow \infty} \frac{1}{n^{1-a}} S_n = \frac{a_1 (1-a)}{2(1-a)f(x)} \quad \text{a.s.} \quad (5.37)$$

Furthermore, we obtain from a simple Abel transform that

$$\sum_{k=1}^n \frac{1}{k^2} = \frac{1}{n} S_n + \sum_{k=1}^{n-1} \left(\frac{1}{k} - \frac{1}{k+1} \right) S_k : \quad (5.38)$$

We obtain from (4.16) that

$$\lim_{n \rightarrow \infty} n^{1-a} S_n = \frac{b_1 C}{a_1 f(x)} : \quad (5.39)$$

Then, we deduce from (5.37) that

$$\lim_{n \rightarrow \infty} n^{1-a} \frac{1}{n} S_n = \frac{b_1^2 C^2 (1-a)}{2a_1 (1-a) f^3(x)} \quad \text{a.s.}$$

which implies that

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n \frac{1}{k} = 0 \quad \text{a.s.} \quad (5.40)$$

In addition, we also have from (4.16) that

$$\lim_{n \rightarrow \infty} n^3 \sum_{k=1}^n \frac{1}{k^2} = 2(1-a) \frac{b_1 C}{a_1 f(\cdot)}^2 :$$

It clearly ensures via (5.37) that

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n \frac{X_k}{k} = 0 \quad \text{a.s.} \quad (5.41)$$

Then, it follows from (5.38) together with (5.40) and (5.41) that

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n \frac{X_k}{k} = 0 \quad \text{a.s.} \quad (5.42)$$

Consequently, we obtain from (5.27) together with (5.36), (5.42) and the Cauchy-Schwarz inequality that $(\hat{\#}_n)$ satisfies the QSL

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n \frac{X_k}{k} = \frac{b_1^2}{2b_1 - 1} \quad \text{a.s.}$$

The proof of the QSL for the convexified estimator $(\hat{\#}_n)$ is much more easier. We infer from (4.5), (4.7), (5.5) and the identity $L(\cdot) = \#$ that for all $n \geq 1$,

$$\begin{aligned} \hat{\#}_{n+1} - \# &= \frac{1}{P_n} (\hat{\#}_1 + N_{n+1} + L_{n+1} - P_n \#) ; \\ &= \frac{1}{P_n} (N_{n+1} + \mathbb{R}_{n+1}) \end{aligned} \quad (5.43)$$

where

$$\mathbb{R}_{n+1} = \hat{\#}_1 - \# + \sum_{k=1}^n \frac{X_k}{k} - L(\cdot) : \quad (5.44)$$

It follows from Lemma 5.3 that

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n \frac{N_k}{P_{k-1}} = \frac{b_1^2}{2b_1 - 1} \quad \text{a.s.} \quad (5.45)$$

Hence, in order to prove (3.4), it suffices to show that

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n \frac{\mathbb{R}_k}{P_{k-1}} = 0 \quad \text{a.s.} \quad (5.46)$$

We shall prove the stronger result

$$\sum_{n=1}^{\infty} \frac{\mathbb{R}_k}{P_{k-1}} < +\infty \quad \text{a.s.} \quad (5.47)$$

We obtain from (4.8) and (5.44) that for all $n \geq 1$,

$$\mathbb{R}_{n+1}^2 \leq 2 \hat{\#}_1^2 + M_f \sum_{k=1}^n \frac{X_k}{k} \quad (5.48)$$

where

$$M_f = \frac{\sum_{j=1}^n f_j^2}{2(1 - b_1)^2}.$$

As before, we obtain from a simple Abel transform that

$$\sum_{k=1}^n b_k P_k^2 = b_n P_n + \sum_{k=1}^{n-1} (b_k P_k - b_{k+1} P_{k+1}) k. \quad (5.49)$$

It is easy to see that

$$b_n P_n - b_{n+1} P_{n+1} = b_n P_n \frac{1 - b_1}{n + 1}.$$

It implies that $0 < n(b_n P_n - b_{n+1} P_{n+1}) < (1 - b_1)b_n P_n$. Hence, as $1/2 < a < 1$ and $1/2 < b_1 < 1$, we deduce from (5.4) and (5.37) that (5.47) holds true. Consequently, we deduce from (5.43) together with (5.45) and (5.46) that

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n \frac{X_k^2}{\#_k} = \frac{b_1^2}{2b_1 - 1} \quad \text{a.s.}$$

which is exactly the QSL given by (3.4).

It only remains to establish the LIL for our estimates $(\hat{\mu}_n)$ and $(\hat{\sigma}_n)$. We start by proving the LIL for the sequence (P_n) . We immediately obtain from (4.18) that

$$\frac{n^{1/2}}{2 \log \log n} P_n = \frac{n^{1/2}}{2 \log \log n} \frac{M_n + H_n + R_n}{P_{n-1}}. \quad (5.50)$$

We already saw in Lemma 5.3 that the martingale (M_n) satisfies the LIL given by (5.18). In addition, it is easy to see from (5.14), (5.31) and (5.34) that

$$\lim_{n \rightarrow \infty} n \frac{H_n^2}{P_{n-1}} = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} n \frac{R_n^2}{P_{n-1}} = 0 \quad \text{a.s.}$$

It clearly implies that

$$\lim_{n \rightarrow \infty} \frac{n^{1/2}}{2 \log \log n} \frac{H_n}{P_{n-1}} = 0 \quad \text{a.s.}$$

and

$$\lim_{n \rightarrow \infty} \frac{n^{1/2}}{2 \log \log n} \frac{R_n}{P_{n-1}} = 0 \quad \text{a.s.}$$

Therefore, we deduce from (5.18) and (5.50) that (P_n) satisfies the LIL

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{n^{1/2}}{2 \log \log n} P_n &= \liminf_{n \rightarrow \infty} \frac{n^{1/2}}{2 \log \log n} P_n \\ &= \frac{b_1^2}{2b_1 - 1} \quad \text{a.s.} \end{aligned} \quad (5.51)$$

Hereafter, one can observe from (4.12) that

$$\frac{n^{1/2}}{2 \log \log n} \hat{\mu}_n = \frac{n^{1/2}}{2 \log \log n} \mu_n + \frac{n^{1/2}}{2 \log \log n} \epsilon_n. \quad (5.52)$$

It follows from (2.11) and (5.39) that

$$\frac{n^{1/2}}{2 \log \log n} \epsilon_n = O\left(\frac{\log n}{n^{1-a} \log \log n}\right) \quad \text{a.s.}$$

which clearly leads to

$$\lim_{n \rightarrow \infty} \frac{n^{1=2}}{2 \log \log n} = 0 \quad \text{a.s.} \quad (5.53)$$

Consequently, we obtain (3.5) from (5.51), (5.52) and (5.53). The proof for the convexified estimator $(\#_n)$ is straightforward. We obtain from (5.43) that

$$\frac{n^{1=2}}{2 \log \log n} \#_n = \frac{n^{1=2}}{2 \log \log n} \frac{N_n + R_n}{P_{n-1}} \quad (5.54)$$

We already saw in Lemma 5.3 that the martingale (N_n) satisfies the LIL given by (5.18). In addition, it is easy to see from (5.14), (5.48) and (5.49) that

$$\lim_{n \rightarrow \infty} n \frac{R_n}{P_{n-1}} = 0 \quad \text{a.s.}$$

which clearly implies

$$\lim_{n \rightarrow \infty} \frac{n^{1=2}}{2 \log \log n} \frac{R_n}{P_{n-1}} = 0 \quad \text{a.s.} \quad (5.55)$$

Finally, we deduce (3.5) from (5.18), (5.54) and (5.55), which completes the proof of Theorem 3.2. \square

5.4 The slow step size case.

In order to prove Theorem 3.3, it is necessary to establish the following QSL and LIL for the martingales (M_n) and (N_n) .

Lemma 5.4. Assume that the step sequences (a_n) and (b_n) are given by

$$a_n = \frac{a_1}{n^a} \quad \text{and} \quad b_n = \frac{b_1}{n^b}$$

where $a_1 > 0, b_1 > 0$ and $1=2 < a < b < 1$. Then, (M_n) and (N_n) share the same QSL

$$\lim_{n \rightarrow \infty} \frac{1}{n^{1-b}} \sum_{k=1}^n \frac{M_k^2}{P_{k-1}} = \frac{b_1}{2(1-b)} \quad \text{a.s.} \quad (5.56)$$

In addition, they also share the same LIL

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{n^b}{2(1-b) \log n} \frac{M_n}{P_{n-1}} &= \liminf_{n \rightarrow \infty} \frac{n^b}{2(1-b) \log n} \frac{M_n}{P_{n-1}} \\ &= \frac{b_1}{2} \quad \text{a.s.} \end{aligned} \quad (5.57)$$

Proof. We recall that the martingale (M_n) and its predictable quadratic variation are given by

$$M_{n+1} = \sum_{k=1}^n b_k P_k W_{k+1} \quad \text{and} \quad \langle M \rangle_{n+1} = \sum_{k=1}^n b_k^2 P_k^2$$

where, thanks to (5.21),

$$\lim_{n \rightarrow \infty} \frac{\langle M \rangle_n}{n} = \frac{b_1}{2} \quad \text{a.s.}$$

It is not hard to see via a comparison series integral together with convergence in Lemma 5.2 that (5.11)

$$\lim_{n \rightarrow \infty} \frac{1}{b_n P_n^2} \sum_{k=1}^n X_k^2 P_k^2 = \frac{1}{2}. \quad (5.58)$$

Hence, we deduce from (5.58) and Toeplitz's lemma that

$$\lim_{n \rightarrow \infty} \frac{1}{b_n P_n^2} hM_{n+1} = \frac{2(\cdot)}{2} \quad \text{a.s.} \quad (5.59)$$

Denote by f_n the explosion coefficient associated with the martingale (M_n) ,

$$f_n = \frac{hM_n - hM_{n-1}}{hM_n}$$

It follows from the very definition of P_n given by (4.4) together with (5.59) that

$$\lim_{n \rightarrow \infty} n^b f_n = 2b_1 \quad \text{a.s.} \quad (5.60)$$

It means that f_n converges to zero almost surely at rate n^b where $1/2 < b < 1$. Furthermore, we assume that the random variable X has a moment of order p strictly greater than 2. It implies that

$$\sup_n E[|W_{n+1}|^p | F_n] < 1 \quad \text{a.s.}$$

Consequently, we deduce from the QSL for martingales given in theorem 3 of [4] that

$$\lim_{n \rightarrow \infty} \frac{1}{\log hM_n} \sum_{k=1}^n X_k^2 \frac{M_k^2}{hM_k} = 1 \quad \text{a.s.} \quad (5.61)$$

Therefore, we obtain from (5.16) and (5.59) together with (5.60) and (5.61) that

$$\lim_{n \rightarrow \infty} \frac{1}{n^{1-b}} \sum_{k=1}^n \frac{M_k^2}{P_k^{2-b}} = \frac{b_1}{2(1-b)} \quad \text{a.s.} \quad (5.62)$$

Hereafter, we focus our attention on the proof of the LIL given by (5.57). Since $b > 1/2$, we obtain from (5.60) that the explosion coefficient f_n satisfies

$$\sum_{n=1}^{\infty} f_n^{p-2} < +\infty \quad \text{a.s.}$$

Therefore, we deduce from the LIL for martingales [29], see also corollary 6.4.25 in [9] that

$$\limsup_{n \rightarrow \infty} \frac{1}{2hM_n \log \log hM_n} M_n^{1/2} = \liminf_{n \rightarrow \infty} \frac{1}{2hM_n \log \log hM_n} M_n^{1/2} = 1 \quad \text{a.s.} \quad (5.63)$$

Hence, we find from (5.16), (5.59) and (5.63) that

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{n^b}{2(1-b) \log n} \frac{M_n^{1/2}}{P_{n-1}} &= \liminf_{n \rightarrow \infty} \frac{n^b}{2(1-b) \log n} \frac{M_n^{1/2}}{P_{n-1}} \\ &= \frac{b_1}{2} \quad \text{a.s.} \end{aligned}$$

The proof for the martingale (N_n) is left to the reader inasmuch as it follows exactly the same lines than those for the martingale (M_n) . □

Proof of Theorem 3.3. We shall proceed as in the proof of Theorem 3.2. We already saw from (5.56) that

$$\lim_{n \uparrow} \frac{1}{n^{1-b}} \sum_{k=1}^n \frac{M_k^2}{P_{k-1}} = \frac{b_1}{2(1-b)} \quad \text{a.s.}$$

Our goal is to prove that the sequence (H_n) given by (4.12) satisfies the QSL

$$\lim_{n \uparrow} \frac{1}{n^{1-b}} \sum_{k=1}^n \frac{X_k^2}{k} = \frac{b_1}{2(1-b)} \quad \text{a.s.} \quad (5.64)$$

On the one hand, we have from (4.19) and (5.29) that

$$|H_{n+1}| = O \left(\sum_{k=1}^n \frac{P_k P_{k+1} \log k}{k^{a-2}} \right) \quad \text{a.s.}$$

In addition, one can easily check from (4.16) and (4.17) that

$$\lim_{n \uparrow} n^{1-a} H_{n+1} = \frac{(b-a)C}{a_1 f(\cdot)}:$$

Hence, we obtain from convergence (5.11) in Lemma 5.2 together with a comparison series integral as previously done in the proof of Theorem 3.2 that

$$|H_{n+1}| = O \left(\sum_{k=1}^n \frac{P_k \log k}{k^{1+b-a-2}} \right) = O \left(\frac{P_n \log n}{n^{1-a-2}} \right) \quad \text{a.s.} \quad (5.65)$$

Consequently, we deduce from (5.65) that

$$\sum_{n=1}^{\infty} \frac{|H_n|^2}{P_{n-1}} < +\infty \quad \text{a.s.} \quad (5.66)$$

On the other hand, we already saw from (5.33) that

$$|R_{n+1}| = O \left(\sum_{k=1}^n \frac{P_k P_{k-1} \log k}{k^2} \right) \quad \text{a.s.}$$

which implies that

$$|R_{n+1}| = O \left(\sum_{k=1}^n \frac{P_k \log k}{k^{a+b}} \right) = O \left(\frac{P_n \log n}{n^a} \right) \quad \text{a.s.} \quad (5.67)$$

Then, as $a > 1=2$, we find from (5.67)

$$\sum_{n=1}^{\infty} \frac{|R_n|^2}{P_{n-1}} < +\infty \quad \text{a.s.} \quad (5.68)$$

Therefore, we obtain from (5.66) and (5.68) that the QSL (5.64) holds true. In order to prove (3.6), it only remains to show via (5.36) that

$$\lim_{n \uparrow} \frac{1}{n^{1-b}} \sum_{k=1}^n \frac{X_k^2}{k} = 0 \quad \text{a.s.} \quad (5.69)$$

We recall from (5.38) that

$$\sum_{k=1}^n \frac{X_k^2}{k} = \sum_{k=1}^n \frac{X_k^2}{k} + \sum_{k=1}^n \left(\frac{X_k^2}{k} - \frac{X_{k+1}^2}{k+1} \right) k:$$

We obtain from (4.16) that

$$\lim_{n \rightarrow \infty} n^{b-a} = \frac{b_1 C}{a_1 f(\cdot)}; \tag{5.70}$$

Then, it follows from (5.37) and (5.70) that

$$\lim_{n \rightarrow \infty} \frac{1}{n^{1+a-2b}} = \frac{b_1^2 C^2 (1)}{2a_1(1-a)f^3(\cdot)} \quad \text{a.s.} \tag{5.71}$$

Consequently, as $a < b$, we deduce from (5.71) that

$$\lim_{n \rightarrow \infty} \frac{1}{n^{1-b}} = 0 \quad \text{a.s.} \tag{5.72}$$

By the same token, we also find from (5.37) and (5.70) that

$$\lim_{n \rightarrow \infty} \frac{1}{n^{1-b}} \sum_{k=1}^n \left(\frac{2}{k} - \frac{2}{k+1} \right) = 0 \quad \text{a.s.} \tag{5.73}$$

Then, we clearly obtain from (5.72) and (5.73) that convergence (5.69) holds true. As before, the proof of the QSL for the convexified estimator $(\hat{\#}_n)$ is much more easier and left to the reader. We now focus our attention on the LIL for our estimates $(\hat{\#}_n)$ and $(\hat{\#}_n)$. We start by proving the LIL for the sequence $(\hat{\#}_n)$ given by (4.12). We immediately obtain from (4.18) that

$$\frac{n^b}{2(1-b)\log n} = \frac{n^b}{2(1-b)\log n} \frac{M_n + H_n + R_n}{P_{n-1}}; \tag{5.74}$$

We already saw in Lemma 5.4 that the martingale (M_n) satisfies the LIL given by (5.57). In addition, as $b < 1 < 2a$, we get from (5.65) and (5.67) that

$$\lim_{n \rightarrow \infty} n^b \frac{H_n}{P_{n-1}} = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} n^b \frac{R_n}{P_{n-1}} = 0 \quad \text{a.s.}$$

which clearly ensures that

$$\lim_{n \rightarrow \infty} \frac{n^b}{2(1-b)\log n} \frac{H_n}{P_{n-1}} = 0 \quad \text{a.s.}$$

and

$$\lim_{n \rightarrow \infty} \frac{n^b}{2(1-b)\log n} \frac{R_n}{P_{n-1}} = 0 \quad \text{a.s.}$$

Consequently, we find from (5.57) and (5.74) that $(\hat{\#}_n)$ satisfies the LIL

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{n^b}{2(1-b)\log n} &= \liminf_{n \rightarrow \infty} \frac{n^b}{2(1-b)\log n} \\ &= \frac{b_1}{2} \quad \text{a.s.} \end{aligned} \tag{5.75}$$

Hereafter, we clearly have from (4.12) that

$$\frac{n^b}{2(1-b)\log n} \hat{\#}_n = \frac{n^b}{2(1-b)\log n} (n + \dots); \tag{5.76}$$

It follows from (2.11) and (5.70) that

$$\frac{n^b}{2(1-b)\log n} (n + \dots)^2 = O\left(\frac{1}{n^{b-a}}\right) \quad \text{a.s.}$$

Since $a < b$, it clearly implies that

$$\lim_{n \rightarrow \infty} \frac{n^b}{2(1-b)\log n} = 0 \quad \text{a.s.} \quad (5.77)$$

Therefore, we obtain (3.7) from (5.75), (5.76) and (5.77). The proof of the LIL for the convex estimator $(\hat{\theta}_n)$ is straightforward and left to the reader, which achieves the proof of Theorem 3.3. \square

6 Proofs of the asymptotic normality results

The proof of Theorem 3.4 relies on the central limit theorem for the two-time-scale stochastic algorithm given in Theorem 1 of Mokkadem and Pelletier [20]. To be more precise, the asymptotic normality for the couples $(n; \hat{\theta}_n)$ and $(n; \hat{\theta}_n)$ is a direct application of Theorem 1 in [20].

Proof of Theorem 3.4. We start with the proof for the standard estimator $\hat{\theta}_n$. As it was previously done in Section 5, our strategy is first to establish the joint asymptotic normality for the couple $(n; \hat{\theta}_n)$ where $\hat{\theta}_n$ is given by (4.12), and then to deduce the joint asymptotic normality for the couple $(n; \hat{\theta}_n)$. We have from (2.1) together with (4.13) that for all $n \geq 1$,

$$\begin{cases} X_{n+1} = X_n + a_n X_{n+1} \\ Y_{n+1} = Y_n + b_n Y_{n+1} \end{cases} \quad (6.1)$$

where

$$\begin{cases} X_{n+1} = f(n; X_n) + \epsilon_n + V_{n+1} \\ Y_{n+1} = g(n; Y_n) + \eta_n + W_{n+1} \end{cases}$$

with $f(\cdot; \cdot) = F(\cdot)$, $\epsilon_n = 0$, $V_{n+1} = F(X_{n+1}) - F(X_n)$ and $g(\cdot; \cdot) =$

$$\epsilon_n = R(X_n) + \frac{a_n}{b_n} G(X_n) + \eta_{n+1}(X_n);$$

$$W_{n+1} = \eta_{n+1} + \frac{a_n}{b_n} V_{n+1};$$

By denoting $\epsilon_n = 0$, we clearly have $f(\cdot; \cdot) = 0$ and $g(\cdot; \cdot) = 0$. To be more precise

$$\begin{cases} f(\cdot; \cdot) = f^{(0)}(\cdot) + o_p(\|\cdot\|) \\ g(\cdot; \cdot) = 0 + o_p(\|\cdot\|) \end{cases};$$

On the one hand, it follows from the conjunction of (4.9), (4.10) and (4.16) that

$$\epsilon_n = r_n^{(0)} + o_p(\|\cdot\|)$$

where $r_n^{(0)} = nu_{n+1}(X_n)$. On the other hand, we infer from (2.12) and (4.17) that

$$r_n^{(0)} = O\left(\frac{n^a \log n}{n}\right) = o_p\left(\frac{1}{b_n}\right) \quad \text{a.s.}$$

Furthermore, $E[V_{n+1} | F_n] = 0$, $E[W_{n+1} | F_n] = 0$, and we already saw in Sections 4 and 5 that $E[V_{n+1}^2 | F_n] = F(X_n)(1 - F(X_n))$ and $E[W_{n+1}^2 | F_n] = \sigma^2(X_n)$. One can also check that

$$E[V_{n+1} W_{n+1} | F_n] = F(X_n) H(X_n) \frac{a_n}{b_n} (1 - F(X_n));$$

It clearly implies that

$$\lim_{n \rightarrow \infty} \frac{E[V_{n+1}^2 | F_n]}{E[V_{n+1} W_{n+1} | F_n]} = \frac{E[V_{n+1}^2 | F_n]}{E[W_{n+1}^2 | F_n]} = \frac{(1 - F(X_n))}{\sigma^2(X_n)} \quad \text{a.s.}$$

Consequently, all the conditions of Theorem 1 in [20] are satisfied with

$$= \frac{(1)}{2f(\cdot)}$$

and

$$\# = \begin{cases} \frac{b_1}{2b_1 - 1} \frac{2(\cdot)}{1} & \text{if } b = 1; \\ \frac{2(\cdot)}{2} & \text{if } b < 1; \end{cases}$$

Therefore, as $\dots = 0$, we obtain from [20] the joint asymptotic normality

$$\begin{pmatrix} \frac{p}{n^a} \\ \frac{p}{n^b} \end{pmatrix} \xrightarrow{L} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \begin{pmatrix} 0 \\ \# \end{pmatrix} \quad (6.2)$$

where $\dots = a_1$ and $\# = b_1 \dots$. Hereafter, in order to prove the joint asymptotic normality for the couple $(\dots; \#_n)$, it is only necessary to show from the very definition of \dots_n given in (4.12) that

$$\lim_{n \rightarrow \infty} \frac{p}{n^b} \dots_n = 0 \quad \text{a.s.} \quad (6.3)$$

We already saw from (5.39) and (5.70) that

$$\lim_{n \rightarrow \infty} n^{b-a} \dots_n = \frac{b_1 C}{a_1 f(\cdot)} \quad (6.4)$$

Hence, we deduce from (2.12) and (6.4) that

$$\frac{p}{n^b} \dots_n(\dots_n) = O\left(\frac{p n^a \log n}{n^b}\right) \quad \text{a.s.}$$

which ensures that (6.3) holds true. Consequently, (3.8) clearly follows from (6.2) and (6.3). The proof for the convexified estimator $(\#_n)$ is much more easy to handle. We have from (2.3) that for all $n \geq 1$,

$$\begin{cases} \dots_{n+1} = \dots_n + a_n X_{n+1} \\ \#_{n+1} = \#_n + b_n Y_{n+1} \end{cases} \quad (6.5)$$

where

$$\begin{cases} X_{n+1} = f(\dots_n; \#_n) + \dots_n^{(\cdot)} + V_{n+1} \\ Y_{n+1} = g(\dots_n; \#_n) + \dots_n^{(\#)} + W_{n+1} \end{cases}$$

with $f(\cdot; \#) = F(\cdot)$, $\dots_n^{(\cdot)} = 0$, $V_{n+1} = F(\dots_n) - \mathbb{1}_{f X_{n+1} \leq \dots_n}$ and $g(\cdot; \#) = \# - \dots_n^{(\#)}$, $\dots_n^{(\#)} = L(\dots_n) - \#$, $W_{n+1} = Z_{n+1} - L(\dots_n)$, where we recall that $E[Z_{n+1} | F_n] = L(\dots_n)$ with $L(\cdot)$ given by (4.2). We clearly have $f(\cdot; \#) = 0$ and $g(\cdot; \#) = 0$. To be more precise,

$$\begin{cases} f(\cdot; \#) = f^{(0)}(\cdot) - 0 \\ g(\cdot; \#) = 0 - 1 \cdot \# - \# + O(jj) - jj^2 \end{cases} :$$

In addition, we deduce from (4.8) that $\dots_n^{(\#)} = L(\dots_n) - L(\cdot) = O(jj_n) - jj^2$. Furthermore, $E[V_{n+1} | F_n] = 0$, $E[W_{n+1} | F_n] = 0$, and we already saw in Sections 4 and 5 that $E[V_{n+1}^2 | F_n] = F(\dots_n)(1 - F(\dots_n))$ and $E[W_{n+1}^2 | F_n] = 2(\dots_n)$. One can also verify that $E[V_{n+1} W_{n+1} | F_n] = F(\dots_n) L(\dots_n) - \dots_n$. It clearly implies that

$$\lim_{n \rightarrow \infty} \frac{E[V_{n+1}^2 | F_n]}{E[V_{n+1} W_{n+1} | F_n]} = \frac{(1)}{(\#)} \frac{(\#)}{2(\cdot)} \quad \text{a.s.}$$

Consequently, our two-time-scale stochastic algorithm satisfies all the conditions of Theorem 1 in [20] where the asymptotic variances and $\#$ have been previously defined. Finally, we obtain the joint asymptotic normality (3.8) where a_1 and $\# = b_1 \#$, which completes the proof of Theorem 3.4. \square

7 Numerical experiments on real data

Simulated data. We briefly illustrate the asymptotic behavior of our two stochastic algorithms ($\hat{\#}_n$) and ($\hat{\#}_n$) with different tuning of parameters. Since we have several elements of variability in the parameters, we have chosen typical setups even if our presentation is not exhaustive. In our synthetic benchmark, we shall consider Exponential and Gamma distributions, even though explicit formula may be found for the pair ($\hat{\#}_n$; $\#$). First of all, we wish to point out that our recursive procedure is very fast for both algorithms since a set of 1000 observations is handled in less than 0.1 second with a standard laptop. Next, Figure 1 illustrates the good almost sure behavior of the standard and convexified algorithms both on Exponential and Gamma distributions. Here, we consider the $E(1=10)$ and $G(4; 3)$ distributions.

Figure 1: Almost sure convergence of our algorithms for $a = 0.5$ and $b_n = 1/n$.

Second, one can verify and compare the limiting variance of the asymptotic normality involved in Theorem 3.4 for several values of a and b . Figure 2 represents the histogram of the rescaled algorithms for several values of a and b . One can check that the convexified algorithm outperforms the standard algorithm as soon as $b < a$.

Figure 2: Distribution of the rescaled algorithms in different situations: left ($a = 2=3 < b = 4=5 < 1$), center ($b = 2=3 < a = 4=5 < 1$), right ($a = 2=3 < b = 1$).

One can also use our method to estimate online 95% confidence intervals for the superquantile $\#$ as explained in Remark 3.4. This is illustrated in Figure 3 with the Exponential and Gamma distributions with $a = 2=3$ and $b = 1$.

Real data. We finally illustrate, as a proof of concept, the use of our two algorithms on financial real-data that are freely available on the R-package tseries (EuStockMarkets

Figure 3: Online confidence interval with the convexified algorithm for the Exponential and Gamma distributions.

dataset). Some more recent resources may also be downloaded on the Yahoo! Finance website. We consider the four time series of the financial stock-markets DAX, CAC40, SMI, FTSE between 2014 and 2018 and compute the CVaR of the weekly log-returns, that are common indicators in the analysis of financial markets. It is commonly admitted as a reasonable approximation that in non-exceptional situations, the log-returns are not far from an independent and identically distributed set of observations. As a major interest in finance, we compute the negative CVaR at the level 10% and some 95% confidence intervals as well. Our results are presented in Figure 4 for the convexified algorithm tuned with the parameters $a = 2=3$, $a_1 = 5$ and $b = 1$, $b_1 = 3=4$.

Figure 4: Convexified algorithm on Yahoo! Finance datasets.

References

- [1] R. R. Bahadur. A note on quantiles in large samples. *Ann. Math. Statist.*, 37:577–580, 1966. MR-0189095
- [2] O. Bardou, N. Frikha, and G. Pages. Computing var and cvar using stochastic approximation and adaptive unconstrained importance sampling. *Monte Carlo Methods and Applications*, 15(3):173–210, 2009. MR-2573212
- [3] A. Ben-Tal and M. Teboulle. Expected utility, penalty functions, and duality in stochastic nonlinear programming. *Management Science*, 32(11):1445–1466, 1986. MR-0878887

- [4] B. Bercu. On the convergence of moments in the almost sure central limit theorem for martingales with statistical applications. *Stochastic Process. Appl.*, 11:157–173, 2004. MR-2049573
- [5] V. S. Borkar. Stochastic approximation with two time scales. *Systems Control Lett.*, 29:291–294, 1997. MR-1432654
- [6] H. Cardot, P. Cénac, and A. Godichon-Baggioni. Online estimation of the geometric median in Hilbert spaces: Nonasymptotic confidence balls. *Ann. Statist.*, 45(2):591–614, 2017. MR-3650394
- [7] H. Cardot, P. Cenac, and P. A. Zitt. Efficient and fast estimation of the geometric median in hilbert spaces with an averaged stochastic gradient algorithm. *Bernoulli*, 19:18–43, 2013. MR-3019484
- [8] K. L. Chung. On a stochastic approximation method. *Ann. Math. Statist*, 25:463–483, 1954. MR-0064365
- [9] M. Duflo. *Random iterative models*, volume 34 of *Applications of Mathematics*. Springer-Verlag, Berlin, 1997. MR-1485774
- [10] V. Fabian. On asymptotic normality in stochastic approximation. *Ann. Math. Statist*, 39:1327–1332, 1968. MR-0231429
- [11] S. Gadat, F. Panloup, and S. Saadane. Stochastic heavy ball. *Electronic Journal of Statistics*, pages 461–529, 2018. MR-3765604
- [12] V. Gaposkin and T. Krasulina. On the law of the iterated logarithm in stochastic approximation processes. *Theory of Probability and its Applications*, 19(4):844–850, 1975. MR-0365954
- [13] J. K. Ghosh. A new proof of the Bahadur representation of quantiles and an application. *Ann. Math. Statist.*, 42:1957–1961, 1971. MR-0297071
- [14] A. Godichon-Baggioni. Estimating the geometric median in hilbert spaces with stochastic gradient algorithms: L_p and almost sure rates of convergence. *Journal of Multivariate Analysis*, pages 209–222, 2015. MR-3477660
- [15] A. Godichon-Baggioni. Online estimation of the asymptotic variance for averaged stochastic gradient algorithms. *J. Statist. Plann. Inference*, pages 1–19, 2019. MR-3950590
- [16] G. Kersting. Almost sure approximation of the robbins-monro process by sums of independent random variables. *Ann. Probab.*, 5(6):954–965, 1977. MR-0494741
- [17] V. Konda and J. N. Tsitsiklis. Convergence rate of linear two-time-scale stochastic approximation. *Ann. Appl. Probab.*, 14:796–819, 2004. MR-2052903
- [18] H. J. Kushner and G. G. Yin. *Stochastic approximation and recursive algorithms and applications*, volume 35 of *Applications of Mathematics*. Springer-Verlag, New York, second edition, 2003. Stochastic Modelling and Applied Probability. MR-1993642
- [19] T. L. Lai and H. Robbins. Adaptive design and stochastic approximation. *Ann. Statist.*, 7(6):1196–1221, 1979. MR-0550144
- [20] A. Mokkadem and M. Pelletier. Convergence rate and averaging of nonlinear two-time-scale stochastic approximation algorithms. *Ann. Appl. Probab.*, 16:1671–1702, 2006. MR-2260078
- [21] M. Pelletier. On the almost sure asymptotic behaviour of stochastic algorithms. *Stochastic Process. Appl.* 78, 2:217–244, 1998. MR-1654569
- [22] B. T. Polyak and A. Juditsky. Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization*, 30:838–855, 1992. MR-1167814
- [23] H. Robbins and S. Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, 22:400–407, 1951. MR-0042668
- [24] H. Robbins and D. Siegmund. A convergence theorem for non negative almost supermartingales and some applications. *Optimizing methods in stat.*, pages 233–257, 1971. MR-0343355
- [25] R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *The Journal of Risk*, 2(3):21–41, 2000. MR-1835507
- [26] R. T. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26(7):1443–1471, 2002. MR-1835507

- [27] D. Ruppert. Efficient estimations from a slowly convergent robbins-monro process. *Technical Report, 781, Cornell university operations research and industrial engineering, Ithaca, NY*, 1988.
- [28] J. Sacks. Asymptotic distribution of stochastic sproximation procedures. *The Annals of Mathematical Statistics*, 29:373–405, 1958. MR-0098427
- [29] W. Stout. A martingale analogue of kolmogorov’s law of the iterated logarithm. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 15:279–290, 1970. MR-0293701

Acknowledgments. The authors warmly thank the two anonymous referees for their insightful comments and constructive suggestions which helped to improve the paper substantially.

Electronic Journal of Probability

Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS¹)
- Easy interface (EJMS²)

Economical model of EJP-ECP

- Non profit, sponsored by IMS³, BS⁴, ProjectEuclid⁵
- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund⁶ (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

¹LOCKSS: Lots of Copies Keep Stuff Safe <http://www.lockss.org/>

²EJMS: Electronic Journal Management System <http://www.vtex.lt/en/ejms.html>

³IMS: Institute of Mathematical Statistics <http://www.imstat.org/>

⁴BS: Bernoulli Society <http://www.bernoulli-society.org/>

⁵Project Euclid: <https://projecteuclid.org/>

⁶IMS Open Access Fund: <http://www.imstat.org/publications/open.htm>