A smoothing stochastic algorithm for quantile estimation

Aboubacar Amiri *, Baba Thiam

Université Lille 3, Laboratoire EQUIPPE, EA 4018, Villeneuve d'Ascq, France

ARTICLE INFO

ABSTRACT

In this paper, we provide the almost-sure convergence and the asymptotic normality of a smooth version of the Robbins–Monro algorithm for the quantile estimation. A Monte Carlo simulation study shows that our proposed method works well within the framework of a data stream.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Quantile estimation is an active field in statistics and has been widely explored in various areas, including finance, economics, biology and hydrology. There are many circumstances under which it is essential to know the quantiles of a given distribution. An overview of the use of special quantiles such as quartiles, deciles and percentiles in statistical applications appears in the monograph of Fisz (1963), whereas recent works on applications of fractile estimation can be found in Jorion (1996), Ridder (1998) and Gourieroux et al. (2000) and the references therein.

In financial econometrics, technological innovations and trade globalization have provided the financial community the opportunity to develop sophisticated statistical techniques for risk management. Several measures of risk have been introduced, the most popular being the so-called Value at Risk (VaR). This measure has become the standard and most useful risk measure for some financial institutions. Practitioners define the VaR as the worst expected loss over a given horizon at a fixed confidence level $p \in (0, 1)$. More precisely, for some probability p, a bank might be able to know that the daily loss of its trading portfolio will exceed the value of the VaR with probability $1 - p$. According to this definition, the VaR can be mathematically defined as the p-quantile of the daily loss distribution. We refer the reader to Scaillet (2003) for a historical background on the VaR.

As another example, in hydrology, sometimes a researcher wants to obtain the value ℓ_p exceeded by the height of a river with a given confidence probability p. Such information commonly forms the basis for decision making and regulatory action.

---

* Corresponding author.
  E-mail address: aboubacar.amiri@univ-lille3.fr (A. Amiri).

In this paper, we address the estimation of the quantile of a random variable X defined in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with values on $\mathbb{R}$. Various estimators of quantiles have been introduced in the literature using parametric, semi-parametric or nonparametric approaches.

Many parametric developments have been investigated and are most often based on the assumption of normality of the underlying distributions. In this parametric context, quantile estimation is reduced to an estimation of a finite number of parameters. However, in many real situations, these parametric considerations can be rather stringent.

Another method is to consider a more general nonparametric approach that makes no explicit assumptions about underlying distributions. The traditional estimator of this type is the empirical quantile. However, as indicated by Daouia et al. (2011), this estimator suffers from a lack of efficiency due to the variation of individual order statistics.

The asymptotic properties of the empirical quantile were studied by David (1970), whereas Harrel and Davis (1982) introduced a nonparametric quantile estimator and compared it to the empirical quantile. Kalgh and Lachenbruch (1982) proposed an estimator that is a linear combination of order statistics. Nadaraya (1964) introduced a kernel estimator that requires inverting the kernel estimator of the distribution function. The mean square error of the kernel quantile estimator was established by Azzalini (1981), whereas the almost-sure convergence of the estimator was studied by Kulczycki and Dawidowicz (1999) under ergodic conditions. Gourieroux et al. (2000) demonstrated the best behavior of the kernel estimator for the VaR through an empirical study. Scaillet (2005) derived the asymptotic properties of kernel estimators of conditional expected shortfalls for stationary and strong mixing processes. Cai and Wang (2008) established an asymptotic normality and the consistency of the nonparametric estimator of the conditional VaR and obtained expected shortfall functions by inverting the weighted double kernel local linear estimate of the conditional distribution function.

Although, the kernel estimator can be preferable to the sample estimator, it suffers from computational problems. In fact, there is no explicit expression for the inverse of the kernel distribution function estimator. Thus, quantile estimation is obtained by numerical approximation methods such as the Newton–Raphson method. However, the convergence of this algorithm requires a large number of iterations and can thus be time-consuming for large sample sizes. This situation is more apparent in data streams, *i.e.*, in which sample data are obtained by an observational mechanism that allows for an increase in sample size over time. In recent years, data streams have become an increasingly important area of research. Common data streams include Twitter activity, the Facebook news stream, Internet packet data, stock market activity, credit card transactions and Internet and phone usage. We refer to Aggarwal (2007) for further information about and examples of data streams. In the data stream setting, kernel estimates must be completely recalculated when each additional item of data is received, after which a numerical approximation is provided. To bypass this problem, one may consider an alternative recursive estimator obtained by a stochastic approximation algorithm. The most famous use of the stochastic algorithm in the nonparametric framework is described in the work of Kiefer and Wolfowitz (1952), who constructed an algorithm that allows for the approximation of the point at which a regression function reaches its maximum. The simplest type of stochastic approximation quantile estimators are based on the work of Robbins and Monro (1951). These pioneering works have been widely discussed and extended in many directions (see, among many others, Blum, 1954, Fabian, 1967, Fabian, 1968, Kushner and Clark, 1978 and Ruppert, 1982). Blum (1954) demonstrated the almost-sure convergence of the Robbins–Monro procedure, and the asymptotic normality of the procedure was proved by Sacks (1958) under suitable conditions. Furthermore, Holst (1987) improved the theory and demonstrated the almost-sure convergence and asymptotic distribution of the quantile estimator based on the Robbins–Monro procedure within the contexts of independent and dependent observations.

In this paper, we study a smooth version of the Robbins–Monro algorithm, which bears a resemblance in its structure to the so-called adaptive Conditional Autoregressive Value-at-Risk (CAViaR) model, introduced by Engle and Manganelli (2004). We establish the almost-sure convergence of our quantile estimator and derive an asymptotic normality.

The main contributions of our work are (i) the generalization of the previous work of Holst (1987) to a smooth version, (ii) the illustration of the theoretical results developed in the above-cited references through a simulation study, (iii) the improvement of the smooth algorithm of Gourieroux et al. (2000) with regard to estimation errors and computational cost and (iv) the generalization of the adaptive CAViaR model in the one-dimensional stationary stream data setting.

The paper is organized as follows. In Section 2, we propose our algorithm, and in Section 3, the asymptotic results are established. Section 4 is devoted to a simulation study in which our estimator is compared with other quantile estimators. Finally, the proofs of the results are discussed in Section 5, and an Appendix is provided.

2. Presentation of the algorithm

Let X be a random variable valued on $\mathbb{R}$ with a probability density function f, and let F denote the cumulative distribution function of the random variable. Assume that F has a uniquely defined p-order quantile q_p for a given $p \in (0, 1)$.

We wish to estimate the solution $x = q_p$ of the equation $F(x) = p$ based on a sample of independent and identically distributed (i.i.d.) random variables $X_1, \ldots, X_n$, with the same distribution as X. This problem is a particular case of the general issue addressed by stochastic approximation methods that are intended to find the unique root of the function $M(x) = \mathbb{E}(N(x, X))$, for a given function N. To solve this problem, Robbins and Monro (1951) propose the following relationship:

$$x_n = x_{n-1} + \gamma_n \eta_n(x_{n-1}), \tag{1}$$

where η_n is a random function that depends on X_n such that $\mathbb{E}(\eta_n(x)) = M(x)$ and (γ_n) is a sequence of positive constants such that $\sum_{n \geq 1} \gamma_n^2 < \infty$ and $\sum_{n \geq 1} \gamma_n = \infty$. The initial value of x_n may be chosen arbitrarily or at random.

If we set $\psi_p(x) := p - \mathbb{1}_{\{x<0\}}$, then we have $p - F(x) = \mathbb{E}\left(\psi_p(X_n - x)\right)$. In addition, if (γ_n) is a sequence of the form $\gamma_n = (An)^{-1}$ with $A > 0$ and following (1), an estimator of q_p is given by

$$x_n = x_{n-1} + \frac{\psi_p(X_n - x_{n-1})}{An}. \tag{2}$$

The algorithm (2) was studied by Blum (1954) for $0 < A < 2f(q_p)$. It is shown in the cited reference that if $A = f(q_p)$, the asymptotic variance of x_n is minimal, while its asymptotic distribution is the same as that of the order statistic estimator. Thus, as suggested in Robinson (1975) or Holst (1987), the constant A can be replaced by the current estimate of $f(q_p)$.

The algorithm (2) involves an indicator function in the incremental estimation of the quantile. This discrete component increases abruptness because the step gives the same consideration to the observations that are close to the quantile estimate and the observations that are extremely larger or smaller than the quantile estimate. Thus, instead of choosing η_n as an unbiased estimator of $p - F(x)$, we propose a smoothed asymptotically unbiased replacement that might improve the performance.

The quantile estimator to be studied in the current paper is given by the following algorithm

$$\begin{cases} x_n = x_{n-1} + (na_n)^{-1}\left[p - H\left(\dfrac{x_{n-1} - X_n}{b_n}\right)\right] \\ a_n = \max\left\{\mu, \min\{f_n, \nu \ln(n+1)\}\right\} \\ f_n = \left(1 - n^{-1}\right)f_{n-1} + \dfrac{1}{nh_n}K\left(\dfrac{x_{n-1} - X_n}{h_n}\right), \end{cases} \tag{3}$$

where f_0, x_0, μ and ν are some positive constants, (h_n) and (b_n) are sequences of bandwidths, K is a positive bounded kernel function and $H(z) = \int_{-\infty}^{z} K(u)du$. Note that f_n is the current recursive kernel density estimator of $f(q_p)$ and can be rewritten as

$$f_n = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i}K\left(\frac{x_{i-1} - X_i}{h_i}\right),$$

such that when at time n the new observation X_n occurs, it is controlled when it is in the vicinity of the latest estimate x_{n-1}. The sequence x_n estimates q_p.

Note that the choice $H(z) = \frac{1}{1+e^{-z}}$ implies that $H\left(\frac{t}{b_n}\right) \to \mathbb{1}_{\{t\geq 0\}}$ as $b_n \to 0$. Thus, our estimator bears a resemblance in its structure to the adaptive CAViaR model, introduced by Engle and Manganelli (2004) to estimate a quantile that changes over time. CAViaR models are estimated by a quantile regression procedure. In addition, our algorithm might also be useful when we replace kernel estimation by sieve or wavelet estimation (see Cosma et al., 2007). Gourieroux et al. (2000) propose a smooth algorithm for quantile estimation based on the Newton–Raphson procedure via the following relationship:

$$x_{n,k} = x_{n-1,k} + \frac{p - \widehat{F}_n\left(x_{n-1,k}\right)}{\widehat{f}_n\left(x_{n-1,k}\right)}, \quad k = 1, 2, \ldots, k_{\max}, \tag{4}$$

where $k_{\max} \geq 1 \widehat{f}_n(\cdot)$ and $\widehat{f}_n(\cdot)$ are the density and distribution kernel estimators, respectively, based on the sample $X_1, \ldots, X_n$. For each sample size n, the total computational cost of (3) is $O(n)$, which is a tremendous gain for our proposed algorithm when compared to the computational cost of (4), which is $O(nk_{\max})$. Thus, the quantile estimation using (4) requires time that depends on the sample size and the number of iterations of the Newton–Raphson procedure, which tends to infinity. Clearly, in a stream data setting, it can be impossible to perform the quantile computation at every step, in addition to there being a lack of space to store all past values of the sample.

3. Assumptions and main results

Before presenting our results, let us consider the following assumptions:

(A1) K is a Lipschitz and continuous function such that

$$\int_{\mathbb{R}} K(u)du = 1, \qquad \int_{\mathbb{R}} uK(u)du = 0 \quad \text{and} \quad \int_{\mathbb{R}} u^2 K(u)du < \infty.$$

(A2) f is a bounded function and is twice continuously differentiable on a neighborhood of q_p.

(A3) (b_n) and (h_n) are sequences of bandwidths such that $b_n \to 0$, $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$. Furthermore, (h_n) satisfies $\sum_{k\geq 1}\frac{1}{k^2 h_k^2} < \infty$.

To obtain a rate of almost-sure consistency for the recursive quantile estimator, we will suppose that the following additional assumptions hold.

(A4) The bandwidth (b_n) is such that $\sum_{k\geq 1} k^{\delta-1}b_k^2 < \infty$ for any $\delta < \frac{1}{2}$.

Assumptions (A1)–(A3) are common in the nonparametric estimation field, whereas the last part of (A3) plays a crucial role in our calculation. (A1) is fulfilled by Gaussian and uniform kernels. The choice $h_n = cn^{-a}$, $0 < a < 1/2$, with $c > 0$, is a typical example of a bandwidth satisfying (A3). Assumption (A4) is satisfied when $b_n = cn^{-b}$, with $b > \delta/2$ and $c > 0$.

The following theorem establishes the pointwise almost-sure convergence of x_n and f_n.

Theorem 1. *Let* (x_n) *and* (f_n) *be generated by the algorithm* (3), *and define* $\Delta = \max \{\mu, f(q_p)\}$.

(a) *If Assumptions* (A1)–(A3) *hold, then*

$$\lim_{n \to \infty} |x_n - q_p| = 0 \quad a.s. \quad and \quad \lim_{n \to \infty} |a_n - \Delta| = 0 \quad a.s.$$

(b) *Under assumptions* (A1)–(A4), *if* $\mu \leq f(q_p)$ *and* $\delta < 1/2$, *then*

$$\lim_{n \to \infty} n^\delta (x_n - q_p) = 0 \quad a.s.$$

Similar results were established by Robinson (1975) for a non-smoothed stochastic algorithm. Based on the work of Robinson (1975), the same estimator was studied by Holst (1987) within the context of i.i.d. observations. In particular, if $H\left(\frac{x_{n-1} - X_n}{b_n}\right)$ is replaced by $\mathbb{1}_{\{x_{n-1} - X_n < 0\}}$ and $h_n = n^{-b}$ with $0 < b < 1/2$, we find the results obtained in the above mentioned references. Other results such as Theorem 1 have been demonstrated for various estimators. Indeed, following (2), we could replace the constant A by any estimator of $f(q_p)$ based on sample observations. Tierney (1983) studied the almost-sure convergence of a spaced-efficient recursive procedure based on the stochastic approximation, which corresponds to the case in which the constant A is estimated by $\min \left\{\frac{1}{n} \sum_{i=1}^{n} \frac{\mathbb{1}_{\{|x_{i-1} - X_i| < h_i\}}}{2h_i}, \ a_0 n^{-a_1}\right\}$, with $a_0 > 0$ and $0 < a_1 < 1/2$.

The following theorem allows for the asymptotic normality of x_n to be obtained.

Theorem 2. *Let* (x_n) *and* (f_n) *be generated by the algorithm* (3). *Assume that assumptions* (A1), (A3) *and* (A4) *hold.*

(a) *If* $\mu \leq f(q_p)$, *then*

$$\sqrt{n}(x_n - q_p) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \ p(1-p)/f^2(q_p)\right).$$

(b) *Moreover, if there exist* $c_1 \geq 0$, $c_2 > 0$ *and* $0 < \beta < 1$ *such that* $nh_n^5 \to c_1$ *and* $n^{1-\beta} h_n \to c_2$, *then*

$$\sqrt{nh_n}(f_n - f(q_p)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(m, \sigma^2\right),$$

where m and σ^2 are defined by

$$m = \sqrt{c_1} \frac{f''(q_p)}{2 - \beta} \int_{\mathbb{R}} z^2 K(z) dz \quad and \quad \sigma^2 = \frac{f(q_p)}{2 - \beta} \int_{\mathbb{R}} K^2(z) dz.$$

The choices of bandwidths previously provided are typical examples satisfying the conditions involved in Theorem 2. Because the results of Theorem 2 depend on unknown quantities, usable asymptotic distributions can be provided, replacing $f(q_p)$ by its current estimate with the choice of bandwidth $h_n \sim n^{-a}$, $1/5 < a < 1/2$. This choice allows for the bias term m to be canceled. Therefore, a confidence interval of q_p with level $1 - \alpha$ is defined by

$$\left[x_n - z_{1-\frac{\alpha}{2}} \sqrt{p(1-p)/nf_n^2}; \ x_n + z_{1-\frac{\alpha}{2}} \sqrt{p(1-p)/nf_n^2}\right],$$

where $z_{1-\frac{\alpha}{2}}$ is the $(1 - \frac{\alpha}{2})$-quantile of the standard Gaussian distribution.

4. Numerical studies

This section provides a comparison between our proposed algorithm (label (1)) and two other quantile estimators presented in the literature, namely a non-smooth stochastic algorithm (label (2)) proposed by Holst (1987) and the non-recursive kernel method (label (3)) proposed by Gourieroux et al. (2000). Our objective is to test these different estimation methods via a simulation study.

We examine the case of data streams in which the observations are continually captured over time and real-time updates of the models are required. In practice, it is often difficult to obtain real-time stream data, but this type of scenario can be generated from the so-called *stream* package available in the **R** software environment. This tool has support for reading and writing data streams through an **R** connection that allows for data generated outside of the **R** environment to be accessed. In our example, we use data streams with a dimensionality of 1 simulated and stored on disk. Once the simulated database is created and saved on disk, we can open a connection to the file where it was written and treat it as a stream of data. In this context, in which there is a time-varying sample size, the quantile estimations are initiated with $n_0 = \lfloor n_{\max}/4 \rfloor$ observations, and the models will be continuously updated until they reach a sample size of $n_{\max}$. In this simulation, $n_{\max}$ is the sample size at the end of the online estimation process. Clearly, the aim is therefore to compute the values of the quantile estimators based on the respective sequences $X_1, \ldots, X_{n_0+t}$, for $1 \leq t \leq n_{\max} - n_0$. However, to optimize the

process and avoid unnecessary calculations, we add convergence conditions in both algorithms; specifically, the algorithms are terminated if the absolute distance between two values of the quantile estimator obtained in two successive steps is less than 10^{-6}. The same stop condition is also used for the Newton–Raphson procedure at each step n. Throughout the simulation study, the maximum number of iterations of the Newton–Raphson algorithm is fixed to $k_{\max} = 100$.

Three distributions are selected: Standard Normal, Standard Lognormal and Weibull $(1, 1)$. The sample sizes generated from these distributions are 100, 200, 300 and 400.

4.1. Pointwise estimation

To select the smoothing parameters, we apply a heuristic method proposed by Kulczycki and Dawidowicz (1999) as follows.

- Compute a pilot kernel density estimator $\tilde{f}_n$ with the bandwidth $\tilde{h}_n = 1.06 \times \sigma_n n^{-1/5}$, σ_n being the standard deviation of the sample $X_1, \ldots, X_n$.
- Compute the modifying parameters $s_i = \left(\frac{\tilde{f}_n(X_i)}{\gamma} \right)^{-1/2}$, $i = 1, \ldots, n$, where γ is the geometric mean of $\tilde{f}_n(X_1), \tilde{f}_n(X_2), \ldots,$ $\tilde{f}_n(X_n)$.

Our proposed algorithm and the non-smooth stochastic algorithm of Holst (1987) are provided using the modified smooth parameters $h_n = \tilde{h}_n s_n$, $n \geq 1$. This choice of bandwidths fulfills assumption (A4) because it can be shown that the pilot estimator $\tilde{f}_n$ converges uniformly to f. To simplify, we set $b_n = h_n$, but our algorithm can be improved by other choices of b_n.

For the estimator developed by Nadaraya (1964), the quantile estimation is obtained by the Newton–Raphson procedure suggested by Gourieroux et al. (2000) and Kulczycki and Dawidowicz (1999), where the density function is estimated by the adaptive estimator $\widehat{f}_n(x) = \frac{1}{n\tilde{h}_n} \sum_{i=1}^{n} \frac{1}{s_i} K \left(\frac{x-X_i}{\tilde{h}_n s_i} \right)$, whereas the estimator of the distribution function is given by $\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} H \left(\frac{x-X_i}{\tilde{h}_n s_i} \right)$.

The Gaussian kernel is considered in the implementation of both algorithms. We consider 100 replications and compute the mean square errors (MSEs) for $p = 0.05, 0.5, 0.75$ and 0.95. For the initialization constants x_{n_0} and f_{n_0}, we choose the empirical quantile computed through the n_0 observations and the corresponding kernel density estimator. Concerning the other quantities involved in the stochastic algorithms, we set $\nu = 1$, whereas μ is selected by a cross-validation procedure over 100 equi-spaced points of the set $[0, 1]$. More precisely, we consider $\mu \in \{0, 1/100, 2/100, \ldots, 1\}$ and the cross-validation criterion

$$\text{CV}(\mu) = \frac{1}{100} \sum_{i=1}^{100} \left(x_{n,i}^{(\mu)} - q_p \right)^2 ,$$

where $x_{n,i}^{(\mu)}$ is the estimated quantile for the ith stage of the Monte Carlo procedure given the sample size n and $p \in (0, 1)$. Then, we select the value μ_{CV} of μ that minimizes $\text{CV}(\mu)$.

Table 1 summarizes the performances of the different methods. From these results, it can be observed that (1) and (2) are relatively more efficient (they performed better than (3) in 73% of the cases), whereas our proposed algorithm performed relatively better than the non-smooth algorithm in 52% of the cases, even if for the Weibull distribution, the estimator (2) was 56% more efficient than (1). Estimators (1) and (2) are superior to estimator (3) mainly because to perform (3) at each step, the researcher must first estimate the density and the distribution functions and then substitute these estimators into the Newton–Raphson approximation. Unfortunately, no method performs better than any other one for any given probability level and sample size. Globally, the estimation errors of the three methods are relatively close.

4.2. Computational times

Next, we compare the performance of both estimators in terms of computational time, which is another criterion that is used when evaluating the efficiency of algorithms in a data stream setting. Table 2 summarizes the different results. As expected, estimators (1) and (2) show a clear advantage with regard to computational time compared with estimator (3). Additionally, it can be observed that our estimator is efficient with regard to computational time. The estimator is efficient because the proposed algorithm can quickly reach the fixed threshold, unlike the two other methods.

In short, the advantage of our proposed approach involves the computational time gain that takes place without a substantial increase in the estimation errors.

5. Proofs

In the proofs, c will denote a constant whose value is unimportant and may vary from line to line. In addition, let $\mathcal{F}_n$ be the σ-algebra generated by $X_1, \ldots, X_n$ and set

$$H_{b_n}(x_{n-1}) = H \left(\frac{x_{n-1} - X_n}{b_n} \right) \quad \text{and} \quad K_{h_n}(x_{n-1}) = K \left(\frac{x_{n-1} - X_n}{h_n} \right) .$$

Table 1
Efficiency of quantile estimators.

	p	(1)	(2)	(3)	(1)	(2)	(3)
		$n_{\max} = 100$			$n_{\max} = 200$		
	0.05	0.0215	0.0434	0.0361	0.0052	0.0121	0.0196
	0.5	0.0037	0.0046	0.0087	0.0084	0.0065	0.0069
	0.75	0.0227	0.0255	0.0185	0.0056	0.0072	0.0084
Normal	0.95	0.0441	0.0691	0.0296	0.0064	0.0113	0.0159
		$n_{\max} = 300$			$n_{\max} = 400$		
	0.05	0.0080	0.0064	0.0295	0.0437	0.0182	0.0316
	0.5	0.0098	0.0097	0.0102	0.0008	0.0009	0.0009
	0.75	0.0059	0.0064	0.0111	0.0032	0.0037	0.0027
	0.95	0.0166	0.0093	0.0174	0.0021	0.0012	0.0109
		$n_{\max} = 100$			$n_{\max} = 200$		
	0.05	0.0361	0.0373	0.1611	0.0046	0.0020	0.2693
	0.5	0.0048	0.0053	0.0123	0.0035	0.0040	0.0177
	0.75	0.0254	0.0200	0.0286	0.0416	0.0415	0.0415
Lognormal	0.95	0.1888	0.1960	0.0662	0.6455	0.7153	0.6402
		$n_{\max} = 300$			$n_{\max} = 400$		
	0.05	0.0003	0.0006	0.2473	0.0037	0.0003	0.2294
	0.5	0.0045	0.0036	0.0088	0.0032	0.0031	0.0084
	0.75	0.0258	0.0274	0.0204	0.0187	0.0189	0.0192
	0.95	0.5930	0.5932	0.4223	0.3257	0.4179	0.3704
		$n_{\max} = 100$			$n_{\max} = 200$		
	0.05	0.0085	0.0005	0.0731	0.0040	0.0002	0.0452
	0.5	0.0086	0.0067	0.0088	0.0035	0.0039	0.0040
	0.75	0.0034	0.0056	0.0193	0.0128	0.0124	0.0109
Weibull(1, 1)	0.95	0.0114	0.0154	0.1865	0.0827	0.0660	0.0981
		$n_{\max} = 300$			$n_{\max} = 400$		
	0.05	0.0022	0.0002	0.0303	0.0019	0.0001	0.0286
	0.5	0.0066	0.0050	0.0035	0.0015	0.0013	0.0019
	0.75	0.0073	0.0071	0.0081	0.0057	0.0057	0.0064
	0.95	0.0515	0.0769	0.0461	0.0495	0.0423	0.0511

Table 2
Computational time in seconds of the three algorithms.

$n_{\max}$	100	200	300	400	500
(1)	5.965	39.257	136.382	222.272	539.415
(2)	7.315	53.579	181.802	411.728	806.535
(3)	12.751	77.226	226.176	488.098	906.395

5.1. Proof of Theorem 1

In the sequel, we assume that n is large enough. It follows from (3) that

$$\mathbb{E}\left[(x_n - q_p)^2 \,|\mathcal{F}_{n-1} \right] \leq (x_{n-1} - q_p)^2 + 2n^{-1}a_n^{-1} (x_{n-1} - q_p)(p - \mathbb{E}H_{b_n}(x_{n-1})) + cn^{-2}.$$

Integration by parts and a change of variables with the help of Taylor's formula allows for the following expression

$$p - \mathbb{E}H_{b_n}(x_{n-1}) = p - \frac{1}{b_n}\int_{\mathbb{R}} K\left(\frac{x_{n-1} - u}{b_n} \right) F(u)du = \int_{\mathbb{R}} K(v)\left[F(q_p) - F(x_{n-1} - b_n v) \right] dv$$

$$= F(q_p) - F(x_{n-1}) + \tau_n,$$

with

$$\tau_n = \frac{b_n^2}{2}\int_{\mathbb{R}} v^2 K(v) f'(x_{n-1} + t b_n v) dv \quad \text{where } 0 < t < 1.$$

Since $\tau_n \to 0$ as $n \to \infty$ and q_p is uniquely defined, then $|\tau_n| \leq \inf_{\{x:|x-q_p|>0\}} \left| F(q_p) - F(x) \right|$ which implies that $(x_{n-1} - q_p)$ $\left(F(x_{n-1}) - F(q_p) - \tau_n \right) \geq 0$, for n large enough. Consequently,

$$\mathbb{E}\left[(x_n - q_p)^2 \,|\mathcal{F}_{n-1} \right] \leq (x_{n-1} - q_p)^2 - 2n^{-1}a_n^{-1} \left| x_{n-1} - q_p \right| \left| F(x_{n-1}) - F(q_p) - \tau_n \right| + cn^{-2}.$$

Lemma 3 (see the Appendix) allows to conclude that $|x_{n-1} - q_p| \to \xi$ a.s., where ξ is a random variable and

$$\sum_{k=2}^{+\infty} k^{-1} a_k^{-1} \left|x_{k-1} - q_p\right| \left|F(x_{k-1}) - F(q_p) - \tau_k\right| < \infty \quad \text{a.s.} \tag{5}$$

It remains to be shown that $\xi = 0$ a.s. Suppose that this is not the case; then there exists a set S with $\mathbb{P}(S) > 0$ such that $\xi(\omega) > 0$ for every $\omega \in S$. Let $\omega \in S$; therefore, there exists $\epsilon > 0$ such that $|x_{n-1}(\omega) - q_p| > \epsilon$. Since q_p is uniquely defined, we obtain $\inf_{\{x:|x-q_p|>\epsilon\}} |F(q_p) - F(x)| > 0$. By the definition of a_n (see Eq. (3)) and because $\tau_n \to 0$ as $n \to 0$, there exists an integer N such that

$$\sum_{k=2}^{+\infty} k^{-1} a_k^{-1} \left|x_{k-1} - q_p\right| \left|F(x_{k-1}) - F(q_p) - \tau_k\right| \geq \epsilon/(2\nu) \inf_{\{x:|x-q_p|>\epsilon\}} \left|F(q_p) - F(x)\right| \sum_{n \geq N} (n \ln(1+n))^{-1} = \infty,$$

which contradicts (5), and the first part of Theorem 1(a) follows. To show that $\lim_{n \to \infty} |a_n - \Delta| = 0$ a.s., define

$$v_n = n^{-1} \sum_{k=1}^{n} \frac{1}{h_k} \mathbb{E}\left[K_{h_k}(x_{k-1}) | \mathcal{F}_{k-1}\right],$$

therefore, we have

$$f_n - v_n = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{h_k} \left\{K_{h_k}(x_{k-1}) - \mathbb{E}\left[K_{h_k}(x_{k-1}) | \mathcal{F}_{k-1}\right]\right\}$$

$$= \frac{1}{n} \sum_{k=1}^{n} \frac{U_k}{h_k}, \quad \text{with } U_k = K_{h_k}(x_{k-1}) - \mathbb{E}\left[K_{h_k}(x_{k-1}) | \mathcal{F}_{k-1}\right].$$

Because $\text{Var}(U_k) < \infty$, using (A3), we obtain $\sum_{k \geq 1} \text{Var}\left(\frac{U_k}{kh_k}\right) \leq c \sum_{k \geq 1} \frac{1}{k^2 h_k^2} < \infty$. Using the stability theorem (see Loève, 1978, p. 53), $f_n - v_n \to 0$ a.s. Next, we have

$$v_n = n^{-1} \sum_{k=1}^{n} \int_{\mathbb{R}} K(z) \left[f(x_{k-1} - h_k z) - f(x_{k-1})\right] dz + n^{-1} \sum_{k=1}^{n} f(x_{k-1}) \tag{6}$$

and because f is continuous, the first term on the right-hand side of (6) tends to zero almost surely, which ensures that $v_n \to f(q_p)$ a.s., thus completing the proof of (a). Concerning (b), from (3), we can write

$$x_n - q_p = x_{n-1} - q_p + n^{-1} a_n^{-1} \left[F(x_{n-1}) - H_{b_n}(x_{n-1})\right] + n^{-1} a_n^{-1} \left[p - F(x_{n-1})\right].$$

Taylor's expansion of F in the neighborhood of q_p yields $F(x_{n-1}) - p = (x_{n-1} - q_p) \left[f(q_p) + o(1)\right]$. It follows that

$$x_n - q_p = (x_{n-1} - q_p) \left[1 - n^{-1} a_n^{-1} \left(f(q_p) + o(1)\right)\right] + n^{-1} a_n^{-1} \left[F(x_{n-1}) - H_{b_n}(x_{n-1})\right]. \tag{7}$$

Obviously, $n^{\delta} = (n-1)^{\delta} \left[1 + \frac{1}{n-1}\right]^{\delta} = (n-1)^{\delta} \left[1 + n^{-1}[\delta + o(1)]\right]$, and thus, one may write

$$n^{\delta} \left(x_n - q_p\right) = (n-1)^{\delta} (x_{n-1} - q_p) \left[1 - \frac{f(q_p) + o(1)}{na_n}\right]\left[1 + \frac{\delta + o(1)}{n}\right] + \frac{F(x_{n-1}) - H_{b_n}(x_{n-1})}{n^{1-\delta} a_n}$$

$$= (n-1)^{\delta}(x_{n-1} - q_p) \left[1 - \frac{f(q_p) + o(1)}{na_n} + \frac{\delta + o(1)}{n}\right] + \left[\frac{F(x_{n-1}) - H_{b_n}(x_{n-1})}{n^{1-\delta} a_n}\right]$$

$$:= (n-1)^{\delta}(x_{n-1} - q_p)(1 - U_n) + \frac{V_n}{n^{1-\delta} a_n}.$$

From Lemma 4 (see the Appendix), the second part of Theorem 1 will be proven if we show that

$$U_n \geq 0, \quad \sum_{n \geq 1} U_n = \infty \quad \text{and} \quad \sum_{n \geq 1} n^{\delta-1} a_n^{-1} V_n < \infty.$$

To this end, note that $\lim_{n \to \infty} a_n = f(q_p)$; therefore, we have $a_n^{-1} f(q_p) \geq \delta$ for sufficiently large n, which implies that $\sum_{n \geq 1} U_n = \infty$ is obvious.

Now, let us show that $\sum_{n \geq 1} n^{\delta-1} a_n^{-1} V_n < \infty$. Set $S_n = \sum_{k=1}^{n} k^{\delta-1} a_k^{-1} V_k$. Clearly, $S_n = S_{n1} + S_{n2}$, where

$$S_{n1} = \sum_{k=1}^{n} k^{\delta-1} a_k^{-1} \left[F(x_{k-1}) - \mathbb{E}(H_{b_k}(x_{k-1}))\right] \quad \text{and}$$

$$S_{n2} = \sum_{k=1}^{n} k^{\delta-1} a_k^{-1} W_k \quad \text{where } W_k = \mathbb{E}(H_{b_k}(x_{k-1})) - H_{b_k}(x_{k-1}).$$

Because $\text{Var}(H_{b_k}(x_{k-1})) < \infty$, we obtain

$$\sum_{k=1}^{+\infty} k^{2\delta-2} a_k^{-2} \, \text{Var}(W_k) \leq c \sum_{k}^{+\infty} k^{2\delta-2} a_k^{-2} \leq c \sum_{k}^{+\infty} k^{2\delta-2} < \infty$$

as soon as $\delta < 1/2$. Because W_k is a martingale difference sequence, Theorem 2.8.7 in Stout (1974, p. 65), ensures that S_{n2} converges. On the other hand, using integration by parts, we can write

$$\mathbb{E}\left[H_{b_k}(x_{k-1})\right] - F(x_{k-1}) = \int_{\mathbb{R}} [F(x_{k-1} - b_k u) - F(x_{k-1})] \, K(u) du.$$

A second-order Taylor expansion of $F(x_{k-1} - b_k u)$ provides that

$$\mathbb{E}\left[H_{b_k}(x_{k-1})\right] - F(x_{k-1}) = \frac{F''(q_p)}{2} b_k^2 \int_{\mathbb{R}} u^2 K(u) du \left[1 + o(1)\right].$$

Therefore, Assumption (A4) ensures that S_{n1} converges, which concludes the proof of Theorem 1. ∎

5.2. Proof of Theorem 2

From (7), one may write $x_n - q_p = (x_{n-1} - q_p)\left[1 - n^{-1}\Gamma_n\right] + n^{-1}a_n^{-1}V_n$, with

$$\Gamma_n = a_n^{-1}\left(f(q_p) + o(1)\right) \quad \text{and} \quad V_n = \left[F(x_{n-1}) - H_{b_n}(x_{n-1})\right].$$

Because $a_n^{-1} \to (f(q_p))^{-1}$, using Lemma 5 (see the Appendix), we can obtain (a) if we show that

$$\text{(i)} \ \mathbb{E}\left[V_n^2 | \mathcal{F}_{n-1}\right] \to p(1-p) \quad \text{and} \quad \text{(ii)} \ \mathbb{E}\left[V_n^2 \mathbb{1}_{\{V_n^2 \geq rn\}}\right] \to 0 \quad \text{as } n \to \infty \text{ for every } r > 0.$$

For (i), let us write

$$\mathbb{E}\left[V_n^2 | \mathcal{F}_{n-1}\right] = \left[F(x_{n-1}) - \mathbb{E}(H_{b_n}(x_{n-1}))\right]^2 + \text{Var}\left(H_{b_n}(x_{n-1})\right).$$

From Theorem 1(a), it is obvious that $\mathbb{E}\left[H_{b_n}(x_{n-1})\right] = F(x_{n-1}) + o(1) = p + o(1)$. Moreover, by integration by parts and a change of variables, and with the help of the definition of H, we obtain

$$\mathbb{E}\left[H_{b_n}^2(x_{n-1})\right] = \frac{2}{b_n} \int_{\mathbb{R}} K\left(\frac{x_{n-1}-t}{b_n}\right) H\left(\frac{x_{n-1}-t}{b_n}\right) F(t) dt$$

$$= 2 \int_{\mathbb{R}} K(u) H(u) F(x_{n-1} - b_n u) du = p + o(1).$$

We deduce that $\mathbb{E}\left[V_n^2 \mid \mathcal{F}_{n-1}\right] = p(1-p)\left[1 + o(1)\right]$. Because V_n is bounded, (ii) is trivially satisfied, which proves (a). Now, for (b), consider the following decomposition

$$f_n - f(q_p) = (1 - n^{-1})\left(f_{n-1} - f(q_p)\right) + n^{-(1+\beta)/2}V_n^* + n^{-1-\beta/2}T_n,$$

where

$$V_n^* = \frac{1}{\sqrt{n^{1-\beta}}h_n}\left[K_{h_n}(x_{n-1}) - h_n f(x_{n-1}) - h_n B_n(x_{n-1})\right] \quad \text{and} \quad T_n = n^{\beta/2}\left[f(x_{n-1}) + B_n(x_{n-1}) - f(q_p)\right],$$

with $B_n(x) = \int_{\mathbb{R}} K(z)\left[f(x - h_n z) - f(x)\right] dz$. Following the same lines of the proof of (a), by applying Lemma 5 once more, it suffices to show that

$$\mathbb{E}\left[V_n^{*2} | \mathcal{F}_{n-1}\right] \to \Sigma := \frac{f(q_p)}{c_2} \int_{\mathbb{R}} K^2(z) dz, \tag{8}$$

$$\mathbb{E}\left[V_n^{*2} \mathbb{1}_{\{V_n^{*2} \geq rn\}}\right] \to 0 \quad \text{as } n \to \infty \text{ for every } r > 0, \tag{9}$$

$$T_n \to T := \sqrt{\frac{c_1}{c_2}} \frac{f''(q_p)}{2} \int_{\mathbb{R}} z^2 K(z) dz \quad \text{as } n \to \infty. \tag{10}$$

For (10), Taylor's formula and Assumptions (A1) and (A2) ensure that there exists $t \in]0, 1[$ such that

$$B_n(x_{n-1}) = \frac{h_n^2}{2} \int_{\mathbb{R}} z^2 K(z) f''(x_{n-1} - t h_n z) dz.$$

Therefore, because $nh_n^5 \to c_1 \geq 0$ and $n^{1-\beta}h_n \to c_2$, using Theorem 1(b), we obtain $\lim_{n\to\infty} T_n = T$, and (10) follows. Concerning (8), we have

$$
\begin{aligned}
\mathbb{E}\left[V_n^{*2}|\mathcal{F}_{n-1}\right] &= \frac{1}{n^{1-\beta}h_n^2}\int_\mathbb{R} K^2\left(\frac{x_{n-1}-u}{h_n}\right)f(u)du + o(1) \\
&= \frac{f(x_{n-1})}{n^{1-\beta}h_n}\int_\mathbb{R} K^2(z)dz + \frac{1}{n^{1-\beta}h_n}\int_\mathbb{R} K^2(z)\left[f(x_{n-1}-h_n z) - f(x_{n-1})\right]dz + o(1) \\
&= \frac{f(q_p)}{c_2}\int_\mathbb{R} K^2(z)dz\left[1+o(1)\right] = \Sigma\left[1+o(1)\right].
\end{aligned}
$$

Let us now prove the Lindeberg condition (9). Because $\mathbb{E}\left[V_n^{*2}\right] < \infty$ and $V_n^{*2} \leq ch_n^{-1}$, then for any $r > 0$,

$$
\mathbb{E}\left[V_n^{*2}\mathbb{1}_{\{V_n^{*2}\geq rn\}}\right] \leq ch_n^{-1}\mathbb{E}\left[\mathbb{1}_{\{V_n^{*2}\geq rn\}}\right] \leq ch_n^{-1}\mathbb{P}\left[V_n^{*2}\geq rn\right] \leq cr^{-1}(nh_n)^{-1}\mathbb{E}\left[V_n^{*2}\right],
$$

and we can deduce (9). With the notations of Lemma 5, we have $\Gamma_n = \Phi_n = 1, \alpha = 1$. Therefore,

$$
n^{\beta/2}\left(f_n - f(q_p)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left((1-\beta/2)^{-1}T, (2-\beta)^{-1}\Sigma\right),
$$

where Σ and T are respectively defined in (8) and (10). The decomposition

$$
\sqrt{nh_n}\left(f_n - f(q_p)\right) = \sqrt{n^{1-\beta}h_n}\,n^{\beta/2}\left(f_n - f(q_p)\right)
$$

and the condition $n^{1-\beta}h_n \to c_2$ with $c_2 > 0$ allow for the conclusion of the proof of Theorem 2. ∎

Acknowledgments

Appendix

In this section, we present technical lemmas used in our proofs. In their paper, Robbins and Siegmund (1971) proved the following lemma.

Lemma 3. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots$ be a sequence of sub σ-fields of $\mathcal{F}$. Let U_n, β_n, ξ_n and ζ_n, $n \in \mathbb{N}^*$ be nonnegative $\mathcal{F}_n$-measurable random variables such that*

$$
\mathbb{E}\left[U_{n+1}|\mathcal{F}_n\right] \leq (1+\beta_n)U_n + \xi_n - \zeta_n.
$$

Therefore, on the set $\{\sum_n \beta_n < \infty, \sum_n \xi_n < \infty\}$, U_n converges a.s. to a random variable and $\sum_n \zeta_n < \infty$ a.s.

Moreover, the following lemma is stated in Venter (1976).

Lemma 4. *If ξ_n is a real sequence satisfying $\xi_{n+1} = (1-a_n)\xi_n + b_n$ where $a_n \geq 0, a_n \to 0, \sum_n a_n = \infty$ and $\sum_n b_n$ converges, then $\xi_n \to 0$ as $n \to \infty$.*

Finally, we present the lemma of Fabian (1968) which is adapted to our context.

Lemma 5. *Let $\mathcal{F}_n \subset \mathcal{F}$. Suppose $U_n, V_n, T_n, T, \Gamma_n, \Phi_n, \Sigma, \Phi \in \mathbb{R}$ and $\Gamma > 0$. Suppose $\Gamma_n, \Phi_{n-1}, V_{n-1}$ are $\mathcal{F}_n$-measurable, $0 < \alpha \leq 1, \beta \geq 0, \beta_+ = \begin{cases} \beta \text{ if } \alpha = 1 \\ 0 \text{ if } \alpha \neq 1 \end{cases}$ and*

$$
\begin{aligned}
&\Gamma_n \to \Gamma, \qquad \Phi_n \to \Phi, \qquad T_n \to T \quad or \quad \mathbb{E}|T_n - T| \to 0 \\
&\mathbb{E}[V_n|\mathcal{F}_n] = 0, \qquad \mathbb{E}\left[V_n^2|\mathcal{F}_n\right] \to \Sigma \text{ and}
\end{aligned}
$$

for every $r > 0$, $\lim_{j\to\infty}\mathbb{E}\left[V_j^2\mathbb{1}_{\{V_j^2\geq rj^\alpha\}}\right] = 0$ or $\alpha = 1$, $\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^n\mathbb{E}\left[V_j^2\mathbb{1}_{\{V_j^2\geq rj^\alpha\}}\right] = 0$.

Suppose that $\beta_+ < 2\Gamma$ and $U_{n+1} = \left(1 - n^{-\alpha}\Gamma_n\right)U_n + n^{-(\alpha+\beta)/2}\Phi_n V_n + n^{-\alpha-\beta/2}T_n$. Then

$$
n^{\beta/2}U_n \xrightarrow{\mathcal{D}} \mathcal{N}\left((\Gamma - \beta_+/2)^{-1}T, (2\Gamma - \beta_+)^{-1}\Sigma\right).
$$

References

Aggarwal, C., 2007. Data Streams Models and Algorithms. Springer.
Azzalini, A., 1981. A note on the estimation of a distribution function and quantiles by a kernel method. Biometrika 68, 326–328.
Blum, J.R., 1954. Multidimensional stochastic approximation methods. Ann. Math. Statist. 25, 737–744.

Cai, Z., Wang, X., 2008. Nonparametric estimation of conditional VaR and expected shortfall. J. Econometrics 147, 120–130.

Cosma, A., Scaillet, O., von Sachs, R., 2007. Multivariate wavelet-based shape preserving estimation for dependent observations. Bernoulli 13, 301–329.

Daouia, A., Gardes, L., Girard, S., 2011. Nadaraya's estimates for large quantiles and free disposal support curves. In: Exploring Research Frontiers in Contemporary Statistics and Econometrics. pp. 1–22.

David, H.A., 1970. Order Statistics. Wiley.

Engle, R.F., Manganelli, S., 2004. CAViaR: conditional autoregressive value at risk by regression quantiles. J. Bus. Econom. Statist. 22 (4), 367–381.

Fabian, V., 1967. Stochastic approximation of minima with improved asymptotic speed. Ann. Math. Statist. 38, 191–200.

Fabian, V., 1968. On asymptotic normality in stochastic approximation. Ann. Math. Statist. 39 (4), 1327–1332.

Fisz, M., 1963. Probability Theory and Mathematical Statistics. Wiley, New-York.

Gourieroux, C., Laurent, J.P., Scaillet, O., 2000. Sensitivity analysis of values at risk. J. Empir. Finance 7, 225–245.

Harrel, F., Davis, C., 1982. A new distribution free quantile estimation. Biometrika 69, 635–640.

Holst, U., 1987. Recursive estimation of quantiles using recursive kernel density estimators. Sequential Anal. 6 (3), 217–237.

Jorion, P., 1996. Risk2: measuring the risk in value at risk. Financ. Anal. 47–56.

Kalgh, W.D., Lachenbruch, P.A., 1982. Generalized quantile estimator. Comm. Statist. Theory Methods 11, 2217–2238.

Kiefer, J., Wolfowitz, J., 1952. Stochastic approximation of the maximum of a regression function. Ann. Math. Statist. 23, 462–466.

Kulczycki, P., Dawidowicz, A.L., 1999. Kernel estimator of quantile. Univ. Jagiellonicue Acta Math. 37, 101–112.

Kushner, H.J., Clark, D.S., 1978. Stochastic Approximation Methods for Constrained and Unconstrained Systems. Springer, New York.

Loève, M., 1978. Probability Theory II, fourth ed. Springer Verlag, New York.

Nadaraya, E.A., 1964. Some new estimates for distribution functions. Theory Probab. Appl. 15, 497–500.

Ridder, T., 1998. Basics of statistical VaR-estimation. In: Risk Measurement, Econometrics and Neural Networks. pp. 161–187.

Robbins, H., Monro, S., 1951. A stochastic approximation method. Ann. Math. Statist. 22 (3), 400–407.

Robbins, R., Siegmund, D., 1971. A convergence theorem for non-negative almost supermartingales and some applications. In: Rustagi, J.S. (Ed.), Optimizing Methods in Statistics. Academic Press, New-York, pp. 233–257.

Robinson, D.W., 1975. Nonparametric Quantile Estimation Through Stochastic Approximation (Ph.D. thesis). Naval Postgraduate School, Monterey, California.

Ruppert, D., 1982. Almost sure approximations to the Robbins–Monro and Kiefer–Wolfowitz processes with dependent noise. Ann. Probab. 10, 178–187.

Sacks, J., 1958. Asymptotic distribution of stochastic approximation procedures. Ann. Math. Statist. 29, 373–4057.

Scaillet, O., 2003. The origin and development of VaR. In: Modern Risk Management: A History, 15th Anniversary of Risk Magazine. Risk Publications, pp. 151–158.

Scaillet, O., 2005. Nonparametric estimation of conditional expected shortfall. Insurance Risk Manage. J. 72, 639–660.

Stout, W., 1974. Almost Sure Convergence. Academic Press.

Tierney, L., 1983. A space-efficient recursive procedure for estimating a quantile of an unknown distribution. SIAM J. Sci. Stat. Comput. 4, 706–711.

Venter, J.H., 1976. An extension of the Robbins–Monro procedure. Ann. Math. Statist. 38, 181–190.