Nonparametric recursive quantile estimation

Michael Kohler[1], Adam Krzyżak[2,*], and Harro Walk[3]

[1] *Fachbereich Mathematik, Technische Universität Darmstadt, Schlossgartenstr. 7, 64289 Darmstadt, Germany, email: kohler@mathematik.tu-darmstadt.de*

[2] *Department of Computer Science and Software Engineering, Concordia University, 1455 De Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8, email: krzyzak@cs.concordia.ca*

[3] *Fachbereich Mathematik, Universität Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany, email: walk@mathematik.uni-stuttgart.de.*

April 19, 2014

Abstract

A simulation model with outcome $Y = m(X)$ is considered, where X is an $\mathbb{R}^d$-valued random variable and $m : \mathbb{R}^d \to \mathbb{R}$ is p-times continuously differentiable. It is shown that an importance sampling Robbins-Monro type quantile estimate achieves for $0 < p \leq d$ the rate of convergence $\log^{3+p/2}(n) \cdot n^{-1/2-p/(2d)}$.

AMS classification: Primary 62G05; secondary 62G20.

Key words and phrases: Nonparametric quantile estimation, importance sampling, rate of convergence, Robbins-Monro procedure.

1 Introduction

Let Y be a real-valued random variable with cumulative distribution function (cdf) $G(y) = \mathbf{P}\{Y \leq y\}$. In this article we are interested in estimating quantiles of Y of level $\alpha \in (0,1)$, which can be defined as any value between

$$q_\alpha^{lower} = \inf\{y \in \mathbb{R} \,:\, G(y) \geq \alpha\} \quad \text{and} \quad q_\alpha^{upper} = \sup\{y \in \mathbb{R} \,:\, G(y) \leq \alpha\}.$$

Throughout this paper we assume that Y has a bounded density g with respect to the Lebesgue-Borel-measure which is positive in a neighborhood of q_α^{upper}, which implies that there exists a uniquely determined quantile $q_\alpha = q_\alpha^{upper} = q_\alpha^{lower}$. Let $Y, Y_1, Y_2, \ldots$ be independent and identically distributed. Given $Y_1, \ldots, Y_n$, we are interested in estimates $\hat{q}_{n,\alpha} = \hat{q}_{n,\alpha}(Y_1, \ldots, Y_n)$ of q_α with the property that the error $\hat{q}_{n,\alpha} - q_\alpha$ converges quickly towards zero in probability as $n \to \infty$.

---

*Corresponding author. Tel: +1-514-848-2424 ext. 3007, Fax: +1-514-848-2830

Running title: *Recursive quantile estimation*

One of the simplest estimates of q_α is given by order statistics. Let $Y_{1:n}, \ldots, Y_{n:n}$ be the order statistics of $Y_1, \ldots, Y_n$, i.e., $Y_{1:n}, \ldots, Y_{n:n}$ is a permutation of $Y_1, \ldots, Y_n$ such that $Y_{1:n} \leq \ldots \leq Y_{n:n}$. Then we can estimate q_α by

$$\bar{q}_{\alpha,n} = Y_{\lceil n\alpha \rceil:n}.$$

The properties of this estimate can be studied using the results from order statistics. In particular Theorem 8.5.1 in Arnold, Balakrishnan and Nagaraja (1992) implies that in case that Y has a density g which is continuous and positive at q_α we have

$$\sqrt{n} \cdot g(q_\alpha) \cdot \frac{Y_{\lceil n\alpha \rceil:n} - q_\alpha}{\sqrt{\alpha \cdot (1-\alpha)}} \to N(0,1) \quad \text{in distribution.} \tag{1}$$

Consequently we have

$$|\bar{q}_{\alpha,n} - q_\alpha| = O_{\mathbf{P}}\left(\frac{1}{\sqrt{n}}\right) \tag{2}$$

where $X_n = O_{\mathbf{P}}(Y_n)$ is defined as follows. For nonnegative random variables X_n and Y_n we say that $X_n = O_{\mathbf{P}}(Y_n)$ if

$$\lim_{c \to \infty} \limsup_{n \to \infty} \mathbf{P}(X_n > c \cdot Y_n) = 0.$$

In order to compute the above estimate one needs to sort the given data $Y_1, \ldots, Y_n$ in increasing order, which requires an amount of time of order $n \cdot \log(n)$ and an amount of space of order n (the latter one in order to save all values of the data points simultaneously). In case that one wants to compute a quantile estimate for a very large sample size, a recursive estimate might be more appropriate. Such a recursive estimate can be computed by applying the Robbins-Monro procedure to estimate the root of $G(z) - \alpha$. In its most simple form one starts here with an arbitrary random variable Z_1, e.g., $Z_1 = 0$, and defines the quantile estimate Z_n recursively via

$$Z_{n+1} = Z_n - \frac{D_n}{n} \cdot \left(I_{\{Y_n \leq Z_n\}} - \alpha\right) \tag{3}$$

for some suitable sequence $D_n \geq 0$. Refined versions of the above simple Robbins-Monro estimate achieve the same rate of convergence as in (1) and (2), explicitly in Tierney (1983) by additional use of a recursive estimate of $g(q_\alpha)$ or, for g Hölder continuous at q_α, as a consequence of general results on averaged Robbins-Monro estimates due to Ruppert (1991) and Polyak and Juditsky (1992).

In this paper we consider a simulation model, e.g., of a technical system, where the random variable Y is given by $Y = m(X)$ for some known measurable function $m : \mathbb{R}^d \to \mathbb{R}$ and some $\mathbb{R}^d$-valued random variable X. In this framework we construct an importance sampling variant of the above recursive estimate, which is based on a suitably defined approximation m_n of m. In case that the function m is p-times continuously differentiable and that X satisfies a proper exponential moment condition we show that this importance sampling variant of the recursive estimate achieves up to some logarithmic factor a rate of convergence of order $n^{-1/2-p/(2d)}$ for $0 < p \leq d$.

The Robbins-Monro procedure was originally proposed by Robbins and Monro (1951) and further developed and investigated as well as applied in many different situations, cf., e.g., the monographs Benveniste, Métivier and Priouret (1990), Ljung, Pflug and Walk (1992), Chen (2002) and Kushner and Yin (2003), and the literature cited therein. Importance sampling is a technique to improve estimation of the expectation of a function by sample averages. Quantile estimation using importance sampling has been considered by Cannamela, Garnier and Iooss (2008), Egloff and Leippold (2010) and Morio (2012). In this paper we use ideas from Kohler et al. (2014) and use importance sampling combined with an approximation of the underlying function m in order to improve the rate of convergence of our recursive estimate of the quantile. Until step n of our recursive procedure we use evaluations of m at at most n nonrandom points in order to construct an approximation of m, and one evaluation of m at each of the n sequential estimates of q_α produced by importance sampling and the sequential algorithm.

Throughout this paper we use the following notations: $\mathbb{N}$, $\mathbb{N}_0$ and $\mathbb{R}$ are the sets of positive integers, nonnegative integers and real numbers, respectively. For a real number z we denote by $\lceil z \rceil$ the smallest integer larger than or equal to z. $\|x\|$ is the Euclidean norm of $x \in \mathbb{R}^d$. For $f : \mathbb{R}^d \to \mathbb{R}$ and $A \subseteq \mathbb{R}^d$ we set

$$\|f\|_{\infty,A} = \sup_{x \in A} |f(x)|.$$

Let $p = k + s$ for some $k \in \mathbb{N}_0$ and $0 < s \leq 1$, and let $C > 0$. A function $m : \mathbb{R}^d \to \mathbb{R}$ is called (p, C)-smooth, if for every $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$ with $\sum_{j=1}^d \alpha_j = k$ the partial derivative $\frac{\partial^k m}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}$ exists and satisfies

$$\left| \frac{\partial^k m}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}(x) - \frac{\partial^k m}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}(z) \right| \leq C \cdot \|x - z\|^s$$

for all $x, z \in \mathbb{R}^d$.

The main result is formulated in Section 2 and its proof is provided in Section 3.

2 Main result

We combine a Robbins-Monro estimate with importance sampling in order to improve the rate of convergence. Here we assume that our data is given by $Y = m(X)$ for some known measurable function $m : \mathbb{R}^d \to \mathbb{R}$ and some $\mathbb{R}^d$-valued random variable X with unknown distribution μ. We assume that we have available a deterministic approximation $\tilde{m}_n$ of m which satisfies

$$\|\tilde{m}_n - m\|_{\infty,[-l_n,l_n]^d} \leq \log^{p+1}(n) \cdot n^{-p/d} \tag{4}$$

for sufficiently large n for some $0 < p \leq d$, where $l_n = \log(n)$. Set

$$m_n = \tilde{m}_n - \log^{p+1}(n) \cdot n^{-p/d}.$$

3

Then we have

$$\|m_n - m\|_{\infty,[-l_n,l_n]^d} \leq 2 \cdot \log^{p+1}(n) \cdot n^{-p/d} \tag{5}$$

and

$$m_n(x) \leq m(x) \quad \text{for all } x \in [-l_n, l_n]^d \tag{6}$$

for sufficiently large n (more precisely, for $n \geq n_0$, where $n_0 \in \mathbb{N}$ is some unknown positive deterministic integer).

We recursively define a sequence of estimates Z_n of q_α. We start by choosing an arbitrary (w.l.o.g. deterministic) Z_1, e.g., $Z_1 = 0$. After having constructed already $Z_1, \ldots, Z_n$, we choose a random variable $X_n^{(IS)}$ such that $X_n^{(IS)}$ has the distribution

$$H_n(B) = \frac{\mu\left(\left(\{x \in [-l_n, l_n]^d \, : \, m_n(x) \leq Z_n\} \cup ([-l_n, l_n]^d)^c\right) \cap B\right)}{\bar{G}_n(Z_n)} \quad (B \in \mathcal{B}^d)$$

where

$$\bar{G}_n(z) = \mu\left(\{x \in [-l_n, l_n]^d \, : \, m_n(x) \leq z\} \cup ([-l_n, l_n]^d)^c\right). \tag{7}$$

By construction, the distribution H_n has the Radon-Nikodym derivative (conditional on Z_n)

$$\frac{dH_n}{d\mu}(x) = \frac{I_{\{m_n(x) \leq Z_n\}} \cdot I_{\{x \in [-l_n,l_n]^d\}} + I_{\{x \notin [-l_n,l_n]^d\}}}{\bar{G}_n(Z_n)}.$$

A realization of such a random variable can be constructed using a rejection method: We generate independent realizations of X until we observe a realization x which satisfies either $x \in [-l_n, l_n]^d$ and $m_n(x) \leq Z_n$ or $x \notin [-l_n, l_n]^d$, which we then use as the realization of our $X_n^{(IS)}$.

Furthermore we choose independent and identically distributed random variables $X_{n,1}, X_{n,2}, \ldots, X_{n,n}$ distributed as X, which are independent of all other random variables constructed or used until this point. Then we set

$$Z_{n+1} = Z_n - \frac{D_n}{n} \cdot \left(I_{\{m(X_n^{(IS)}) \leq Z_n\}} \cdot \tilde{G}_n(Z_n) - \alpha\right), \tag{8}$$

where $D_n = \log^2(n)$ and

$$\tilde{G}_n(z) = \frac{1}{n} \sum_{i=1}^{n} \left(I_{\{m_n(X_{n,i}) \leq z, X_{n,i} \in [-l_n,l_n]^d\}} + I_{\{X_{n,i} \notin [-l_n,l_n]^d\}}\right) \quad (z \in \mathbb{R}).$$

Our main result gives an upper bound on the error of this quantile estimate.

Theorem 1 *Let X, $X_{1,1}$, $X_{2,1}$, $X_{2,2}$, $X_{3,1}$, $X_{3,2}$, $X_{3,3}$, ... be independent and identically distributed $\mathbb{R}^d$-valued random variables and let $m : \mathbb{R}^d \to \mathbb{R}$ be a measurable function. Let $\alpha \in (0,1)$ and let q_α be the α-quantile of $Y = m(X)$. Assume that $Y = m(X)$ has a bounded density g with respect to the Lebesgue-Borel measure which is bounded away from zero in a neighborhood of q_α.*

Define $X_n^{(IS)}$ as above, where m_n satisfies (5) and (6) for some $0 < p \leq d$, and let $\hat{q}_{\alpha,n}^{(IS)} = Z_n$ be the Robbins-Monro importance sampling quantile estimate defined above with $D_n = \log^2(n)$. Then

$$\mathbf{P}\left\{ X \notin [-\log(n), \log(n)]^d \right\} > 0 \quad (n \in \mathbb{N}) \quad and \quad \mathbf{E}\{e^{\|X\|}\} < \infty \tag{9}$$

imply

$$\hat{q}_{\alpha,n}^{(IS)} \to q_\alpha \quad a.s. \quad and \quad \left| \hat{q}_{\alpha,n}^{(IS)} - q_\alpha \right| = O_{\mathbf{P}}\left(\log^{3+p/2}(n) \cdot n^{-1/2 - p/(2d)} \right).$$

Remark 1. The construction of an approximation m_n which satisfies (4) in case of a (p, C)–smooth function m can be obtained, e.g., by spline approximation of the function m using n points in $[-\log(n), \log(n)]^d$ (cf., e.g., Kohler (2013) or Kohler et al. (2014)), which can be either equidistantly chosen in $[-\log(n), \log(n)]^d$ or can be recursively defined such that for computation of m_{n+1} evaluations of m used for computation of m_n are used again. However, if we compute such a spline approximation we end up with an algorithm which needs for this again linear space in n, so the above advantage of the recursive algorithm disappears. But we can also use less data points for this spline approximator and require instead a higher degree of smoothness for m in order to achieve the same rate of convergence as above but with less requirement for space. E.g., if the spline approximator is based only on $\sqrt{n}$ evaluations of m on equidistant points but m is $(2 \cdot p, C)$–smooth, then (4) still holds, so our algorithm, which requires now only space of order $\sqrt{n}$, still achieves the rate of convergence in Theorem 1. Hence as the importance sampling algorithm in Kohler et al. (2014), our newly proposed algorithm achieves in this case a faster rate of convergence than the estimate based on order statistics, but it requires less space to be computed than the order statistics or the estimate in Kohler et al. (2014). It should also be noted that compared to Kohler et al. (2014) the newly proposed estimate needs a stronger smoothness assumption on m to achieve the better rate of convergence mentioned above.

Remark 2. If μ is known and has a known density f, then we can avoid the observation of $X_{i,j}$ by replacing $\tilde{G}_n(z)$ by its expectation

$$\bar{G}_n(z) = \int_{\mathbb{R}^d} f(t) \cdot \left(I_{\{t \in [-l_n, l_n]^d \,:\, m_n(t) \leq z\}} + I_{\{t \notin [-l_n, l_n]^d\}} \right) dt,$$

which can be computed, e.g., by numerical integration. The proof of Theorem 1 implies that the corresponding estimate achieves the same rate of convergence as the estimate in Theorem 1.

Remark 3. In the context of Theorem 1 one can modify the rejection method which yields $X_n^{(IS)}$, $n \in \mathbb{N}$, by generating at most n independent realizations of X, i.e., by stopping latest at the n-th trial. This leads to a modification of $X_n^{(IS)}$ replacing it by ∞ if none of the n trials yields the desired realization (and then setting $m(\infty) = \infty$). It is possible to show that in this case the assertion of Theorem 1 remains valid.

3 Proof of Theorem 1

Without loss of generality we assume that (5) and (6) hold for all $n \in \mathbb{N}$ (otherwise we start our Robbins-Monro procedure at step n_0 instead of at step 1). We have

$$Z_{n+1} = Z_n - \frac{D_n}{n} \cdot (G(Z_n) - G(q_\alpha)) + \frac{D_n}{n} \cdot V_n, \tag{10}$$

where

$$
\begin{aligned}
V_n &= G(Z_n) - I_{\{m(X_n^{(IS)}) \leq Z_n\}} \cdot \tilde{G}_n(Z_n) \\
&= G(Z_n) - I_{\{m(X_n^{(IS)}) \leq Z_n\}} \cdot \bar{G}_n(Z_n) - I_{\{m(X_n^{(IS)}) \leq Z_n\}} \cdot \left(\tilde{G}_n(Z_n) - \bar{G}_n(Z_n) \right)
\end{aligned}
$$

and

$$\bar{G}_n(z) = \mathbf{P}\left\{ m_n(X) \leq z, X \in [-l_n, l_n]^d \right\} + \mathbf{P}\left\{ X \notin [-l_n, l_n]^d \right\}$$

(cf., (7)). Let $\mathcal{F}_n$ be the σ-field generated by $X_1^{(IS)}$, ..., $X_n^{(IS)}$, $X_{1,1}$, $X_{2,1}$, $X_{2,2}$, ..., $X_{n,1}$, ..., $X_{n,n}$. Then Z_n is measurable with respect to $\mathcal{F}_{n-1}$.

In the first step of the proof we show

$$\left| \mathbf{E}\left\{ I_{\{m(X_n^{(IS)}) \leq Z_n\}} \cdot \bar{G}_n(Z_n) \big| Z_n = z \right\} - G(z) \right| \leq \frac{c_1}{n} \tag{11}$$

for all $z \in \mathbb{R}$. By definition of $X_n^{(IS)}$ and (6) we have for $z \in \mathbb{R}$

$$
\begin{aligned}
&\mathbf{E}\left\{ I_{\{m(X_n^{(IS)}) \leq Z_n\}} \bigg| Z_n = z \right\} \\
&= \int I_{\{m(t) \leq z\}} dH_n(t) \\
&= \frac{1}{\bar{G}_n(z)} \int I_{\{m(t) \leq z\}} \cdot \left(I_{\{t \in [-l_n, l_n]^d : m_n(t) \leq z\}} + I_{\{t \notin [-l_n, l_n]^d\}} \right) d\mu(t) \\
&= \frac{\mathbf{P}\left\{ m(X) \leq z, X \in [-l_n, l_n]^d \right\} + \mathbf{P}\left\{ X \notin [-l_n, l_n]^d \right\}}{\bar{G}_n(z)},
\end{aligned}
\tag{12}
$$

hence the left-hand side of (11) is equal to

$$\left| \mathbf{P}\left\{ m(X) \leq z, X \in [-l_n, l_n]^d \right\} + \mathbf{P}\left\{ X \notin [-l_n, l_n]^d \right\} - \mathbf{P}\{ m(X) \leq z \} \right| \leq \mathbf{P}\{ X \notin [-l_n, l_n]^d \}.$$

By the Markov inequality and assumption (9) we get

$$\mathbf{P}\{ X \notin [-l_n, l_n]^d \} \leq \mathbf{P}\{ \|X\| \geq \log(n) \} \leq \frac{\mathbf{E}\{ \exp(\|X\|) \}}{\exp(\log(n))} \leq \frac{c_1}{n},$$

which implies (11).

In the second step of the proof we show

$$\mathbf{Var}\left\{ I_{\{m(X_n^{(IS)}) \leq Z_n\}} \cdot \bar{G}_n(Z_n) \big| Z_n = z \right\} \leq c_2 \cdot \log^{p+1}(n) \cdot n^{-p/d} \tag{13}$$

for all $z \in \mathbb{R}$. By (12), (5) and (6), which implies

$$\mathbf{P}\{m_n(X) \leq z, X \in [-l_n, l_n]^d\} \geq \mathbf{P}\{m(X) \leq z, X \in [-l_n, l_n]^d\},$$

we get

$$
\begin{aligned}
&\mathbf{Var}\left\{I_{\{m(X_n^{(IS)}) \leq Z_n\}} \cdot \bar{G}_n(Z_n) \big| Z_n = z\right\} \\
&= \bar{G}_n(z)^2 \cdot \left(\mathbf{E}\{I_{\{m(X_n^{(IS)}) \leq z\}} | Z_n = z\} - \left(\mathbf{E}\{I_{\{m(X_n^{(IS)}) \leq z\}} | Z_n = z\}\right)^2\right) \\
&= \bar{G}_n(z) \cdot \left(\mathbf{P}\{m(X) \leq z, X \in [-l_n, l_n]^d\} + \mathbf{P}\left\{X \notin [-l_n, l_n]^d\right\}\right) \\
&\qquad\qquad - \left(\mathbf{P}\{m(X) \leq z, X \in [-l_n, l_n]^d\} + \mathbf{P}\left\{X \notin [-l_n, l_n]^d\right\}\right)^2 \\
&= \left(\mathbf{P}\{m_n(X) \leq z, X \in [-l_n, l_n]^d\} - \mathbf{P}\{m(X) \leq z, X \in [-l_n, l_n]^d\}\right) \\
&\qquad\qquad \cdot \left(\mathbf{P}\{m(X) \leq z, X \in [-l_n, l_n]^d\} + \mathbf{P}\left\{X \notin [-l_n, l_n]^d\right\}\right) \\
&\leq \mathbf{P}\{m_n(X) \leq z, X \in [-l_n, l_n]^d\} - \mathbf{P}\{m(X) \leq z, X \in [-l_n, l_n]^d\} \\
&\leq \mathbf{P}\{m(X) \leq z + 2 \cdot \log^{p+1}(n) \cdot n^{-p/d}, X \in [-l_n, l_n]^d\} - \mathbf{P}\{m(X) \leq z, X \in [-l_n, l_n]^d\} \\
&= \int_{[-l_n, l_n]^d} \left(I_{\{m(x) \leq z + 2 \cdot \log^{p+1}(n) \cdot n^{-p/d}\}} - I_{\{m(x) \leq z\}}\right) \mathbf{P}_X(dx) \\
&\leq G(z + 2 \cdot \log^{p+1}(n) \cdot n^{-p/d}) - G(z) \\
&\leq c_2 \cdot \log^{p+1}(n) \cdot n^{-p/d},
\end{aligned}
$$

where we have used in the last inequality that G has a bounded density. This implies (13).

In the third step of the proof we show

$$Z_n \to q_\alpha \quad a.s. \tag{14}$$

By construction $X_{n,1}, \dots, X_{n,n}$ are independent of $X_n^{(IS)}$ and Z_n which implies

$$\mathbf{E}\left\{I_{\{m(X_n^{(IS)}) \leq Z_n\}} \cdot \left(\tilde{G}_n(Z_n) - \bar{G}_n(Z_n)\right) \Big| \mathcal{F}_{n-1}\right\} = 0$$

and

$$\mathbf{Var}\left\{I_{\{m(X_n^{(IS)}) \leq Z_n\}} \cdot \left(\tilde{G}_n(Z_n) - \bar{G}_n(Z_n)\right) \Big| \mathcal{F}_{n-1}\right\} \leq \mathbf{E}\left\{\left(\tilde{G}_n(Z_n) - \bar{G}_n(Z_n)\right)^2 \Big| \mathcal{F}_{n-1}\right\} \leq \frac{1}{n}. \tag{15}$$

According to this and (11) and (13), the random variable V_n in (10) satisfies

$$|\mathbf{E}\{V_n | \mathcal{F}_{n-1}\}| \leq \frac{c_3}{n} \tag{16}$$

and

$$\begin{aligned}
\mathbf{E}\left\{V_n^2 | \mathcal{F}_{n-1}\right\} &\leq (\mathbf{E}\{V_n | \mathcal{F}_{n-1}\})^2 + \mathbf{Var}\{V_n | \mathcal{F}_{n-1}\} \\
&\leq \frac{c_3^2}{n^2} + 2 \cdot \mathbf{Var}\left\{I_{\{m(X_n^{(IS)}) \leq Z_n\}} \cdot \bar{G}_n(Z_n) \Big| \mathcal{F}_{n-1}\right\} \\
&\quad + 2 \cdot \mathbf{Var}\left\{I_{\{m(X_n^{(IS)}) \leq Z_n\}} \cdot \left(\tilde{G}_n(Z_n) - \bar{G}_n(Z_n)\right) \Big| \mathcal{F}_{n-1}\right\} \\
&\leq \frac{c_3^2}{n^2} + 2 \cdot c_2 \cdot \log^{p+1}(n) \cdot n^{-p/d} + \frac{2}{n} \\
&\leq c_4 \cdot \log^{p+1}(n) \cdot n^{-p/d}, \tag{17}
\end{aligned}$$

since $p \leq d$. By a theorem of Gladyshev (1965) on the Robbins-Monro algorithm (see, e.g., Ljung, Pflug and Walk (1982), p. 8, Theorem 1.9, applied with some random H_n satisfying $|H_n| \leq c/n$) one immediately obtains (14).

Choose $\epsilon > 0$ such that

$$\frac{G(z) - G(q_\alpha)}{z - q_\alpha} > g(q_\alpha)/2 > 0 \quad \text{whenever } |z - q_\alpha| < \epsilon$$

and let B_N be the event that $|Z_n - q_\alpha| < \epsilon$ for all $n \geq N$. *In the fourth step of the proof* we show that the assertion of Theorem 1 follows from step 3 and

$$\log^{-6-p}(n) \cdot n^{1+p/d} \cdot \mathbf{E}\left\{|Z_n - q_\alpha|^2 \cdot I_{B_N}\right\} \to 0 \quad (n \to \infty) \tag{18}$$

for all sufficiently large $N \in \mathbb{N}$. Because of (14) we have $\mathbf{P}(B_N) \to 1$ $(N \to \infty)$, consequently the assertion of Theorem 1 is implied by

$$\mathbf{P}\left\{|Z_n - q_\alpha| > c_5 \cdot \log^{3+p/2}(n) \cdot n^{-1/2-p/(2d)} \quad \text{and } B_N \text{ holds}\right\} \to 0 \quad (n \to \infty)$$

for all $N \in \mathbb{N}$. By the Markov inequality this in turn follows from (18).

In the fifth step of the proof we show that $|Z_n - q_\alpha| < \epsilon$ implies

$$\mathbf{E}\left\{|Z_{n+1} - q_\alpha|^2 \Big| \mathcal{F}_{n-1}\right\} \leq \left(1 - c_6 \cdot \frac{D_n}{n}\right)|Z_n - q_\alpha|^2 + c_7 \cdot \frac{D_n^2}{n^2} \cdot \log^{p+1}(n) \cdot n^{-p/d} \tag{19}$$

for some constants $c_6, c_7 > 0$ and n sufficiently large. From (10) we get

$$Z_{n+1} - q_\alpha = \left(1 - \frac{D_n}{n} \cdot A_n\right) \cdot (Z_n - q_\alpha) + \frac{D_n}{n} \cdot V_n$$

where

$$A_n = \frac{G(Z_n) - G(q_\alpha)}{Z_n - q_\alpha}$$

and $V_n = G(Z_n) - I_{\{m(X_n^{(IS)}) \leq Z_n\}} \cdot \tilde{G}_n(Z_n)$. Here Z_n and A_n are $\mathcal{F}_{n-1}$ measurable. Using $2 \cdot a \cdot b \leq$

$a^2/\delta + b^2 \cdot \delta$ for $a, b \in \mathbb{R}$ and $\delta > 0$ this implies in case $|Z_n - q_\alpha| < \epsilon$

$$\mathbf{E}\left\{|Z_{n+1} - q_\alpha|^2 \,\big|\, \mathcal{F}_{n-1}\right\}$$

$$= \left(1 - \frac{D_n}{n} \cdot A_n\right)^2 |Z_n - q_\alpha|^2 + \frac{D_n^2}{n^2} \cdot \mathbf{E}\left\{V_n^2 \big| \mathcal{F}_{n-1}\right\}$$

$$+ 2 \cdot \left(1 - \frac{D_n}{n} \cdot A_n\right) \cdot (Z_n - q_\alpha) \cdot \frac{D_n}{n} \cdot \mathbf{E}\{V_n | \mathcal{F}_{n-1}\}$$

$$\leq \left(1 - \frac{D_n}{n} \cdot A_n\right)^2 \left(1 + \frac{D_n \cdot A_n}{n}\right) |Z_n - q_\alpha|^2 + \frac{D_n^2}{n^2} \cdot \mathbf{E}\left\{V_n^2 \big| \mathcal{F}_{n-1}\right\}$$

$$+ \frac{n}{D_n \cdot A_n} \cdot \frac{D_n^2}{n^2} \cdot \left(\mathbf{E}\left\{V_n \big| \mathcal{F}_{n-1}\right\}\right)^2.$$

If $|Z_n - q_\alpha| < \epsilon$ then we have $A_n > g(q_\alpha)/2 = c_6 > 0$. This together with (16) and (17) and the uniform boundedness of A_n (which is a consequence of the boundedness of the density g of G) imply (19).

In the sixth (and final) step of the proof we finish the proof by showing (18). Let $B_{N,n}$ be the event that $|Z_k - q_\alpha| < \epsilon$ for all $N \leq k \leq n$. Because of $I_{B_N} \leq I_{B_{N,n-1}} \leq I_{B_{N,n-2}}$ and the $\mathcal{F}_{n-2}$-measurability of $I_{B_{N,n-1}}$ we can conclude from step 5 for sufficiently large n

$$\mathbf{E}\left\{|Z_n - q_\alpha|^2 \cdot I_{B_N}\right\} \leq \mathbf{E}\left\{\mathbf{E}\left\{|Z_n - q_\alpha|^2 | \mathcal{F}_{n-2}\right\} \cdot I_{B_{N,n-1}}\right\}$$

$$\leq \left(1 - c_6 \cdot \frac{\log^2(n)}{n}\right) \mathbf{E}\left\{|Z_{n-1} - q_\alpha|^2 \cdot I_{B_{N,n-2}}\right\} + c_7 \cdot \log^{p+5}(n) \cdot n^{-2-p/d}.$$

An iterative application of this argument yields for any sufficiently large $N \in \mathbb{N}$ and $\lceil n/2 \rceil > N$

$$\mathbf{E}\left\{|Z_n - q_\alpha|^2 \cdot I_{B_N}\right\}$$

$$\leq \sum_{k=\lceil n/2 \rceil + 1}^{n} c_7 \cdot \log^{p+5}(k) \cdot k^{-2-p/d} \prod_{l=k+1}^{n} \left(1 - c_6 \cdot \frac{\log^2(l)}{l}\right) + \epsilon^2 \cdot \prod_{l=\lceil n/2 \rceil + 1}^{n} \left(1 - c_6 \cdot \frac{\log^2(l)}{l}\right)$$

$$\leq c_8 \cdot \log^{p+5}(n) \cdot n^{-1-p/d} + \epsilon^2 \cdot \exp\left(-\frac{1}{2} \cdot c_1 \cdot \log^2(\lceil n/2 \rceil)\right).$$

The proof is complete. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

4   Acknowledgment

References

[1] Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (1992). *A First Course in Order Statistics.* John Wiley & Sons.

[2] Benveniste, A., Métivier, M., and Priouret, P. (1990). *Adaptive Algorithms and Stochastic Approximation.* Springer-Verlag, New York.

[3] Cannamela, C, Garnier, J., and Iooss, B. (2008). Controlled stratification for quantile estimation. *The Annals of Applied Statistics*, vol. 2, no. 4, pp. 1554-1580.

[4] Chen, H.-F. (2002). *Stochastic Approximation and Its Applications.* Kluwer Academic Publishers, Boston.

[5] Egloff, D., and Leippold, M. (2010). Quantile estimation with adaptive importance sampling. *The Annals of Statistics*, vol. 38, no. 2, pp. 1244-1278.

[6] Gladyshev, E. G. (1965). On stochastic approximation. *Theory of Probability and Applications*, vol. 10, no. 2, pp. 275-278.

[7] Kohler, M. (2013). Optimal global rates of convergence for noiseless regression estimation problems with adaptively chosen design. Submitted for publication.

[8] Kohler, M., Krzyżak, A., Tent, R., and Walk, H. (2014). Nonparametric quantile estimation using importance sampling. Submitted for publication.

[9] Kushner, H. J., and Yin, G. (2003). *Stochastic Approximation and Recursive Algorithms and Applications*, 2nd edition, Springer-Verlag, New York.

[10] Ljung, L., Pflug, G., and Walk, H. (1992). *Stochastic Approximation and Optimization of Random Systems.* Birkhäuser Verlag, Basel.

[11] Morio, J. (2012). Extreme quantile estimation with nonparametric adaptive importance sampling *Simulation Modelling Practice and Theory*, vol. 27, pp. 76-89.

[12] Polyak, B. T., and Juditsky, A. B. (2002). Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization*, vol. 30, no. 4, pp. 838-855.

[13] Robbins, H., and Monro, S. (1951). A stochastic apprximation method. *The Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400-407.

[14] Ruppert, D. (1991). Stochastic approximation. In. *Handbook of Sequential Analysis* (eds. B. K. Gosh and P. K. Sen), Ch. 22, pp. 503-529. Marcel Dekker, New York.

[15] Tierney, L. (1983). A space-efficient recursive procedure for estimating a quantile of an unknown distribution. *SIAM Journal on Scientific and Statistical Computation*, vol. 4, no. 4, pp. 706-711.