Nonparametric recursive quantile estimation

Michael Kohler, Adam Krzyżak, Harro Walk
Nonparametric recursive quantile estimation

Michael Kohler\textsuperscript{1}, Adam Krzyżak\textsuperscript{2,*}, and Harro Walk\textsuperscript{3}

\textsuperscript{1} Fachbereich Mathematik, Technische Universität Darmstadt, Schloßgartenstr. 7, 64289 Darmstadt, Germany, email: kohler@mathematik.tu-darmstadt.de
\textsuperscript{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
\textsuperscript{3} Fachbereich Mathematik, Universität Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany, email: walk@mathematik.uni-stuttgart.de.

May 31, 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) $G(y) = \mathbb{P}\{Y \leq y\}$. 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 $q_{\alpha}^{\text{lower}} = \inf\{y \in \mathbb{R} : G(y) \geq \alpha\}$ and $q_{\alpha}^{\text{upper}} = \sup\{y \in \mathbb{R} : G(y) \leq \alpha\}$.

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 $q_{\alpha}^{\text{upper}}$, which implies that there exists a uniquely determined quantile $q_{\alpha} = q_{\alpha}^{\text{upper}} = q_{\alpha}^{\text{lower}}$. 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 $q_{\alpha}$ 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 \( q_\alpha \) is given by order statistics. Let \( Y_1, \ldots, Y_n \) be the order statistics of \( Y_1, \ldots, Y_n \), i.e., \( Y_1, \ldots, Y_n \) is a permutation of \( Y_1, \ldots, Y_n \) such that \( Y_{1:n} \leq \ldots \leq Y_{n:n} \). Then we can estimate \( q_\alpha \) by

\[
\overline{q}_{\alpha,n} = Y_{\lceil n\alpha \rceil:n}.
\]

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 \( q_\alpha \), 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

\[
|\overline{q}_{\alpha,n} - q_\alpha| = O_p\left( \frac{1}{\sqrt{n}} \right) \tag{2}
\]

where \( X_n = O_p(Y_n) \) is defined as follows. For nonnegative random variables \( X_n \) and \( Y_n \) we say that \( X_n = O_p(Y_n) \) if

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

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

\[
Z_{n+1} = Z_n - \frac{D_n}{n} \cdot (I\{Y_n \leq Z_n\} - \alpha) \tag{3}
\]

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) and Holst (1987) by additional use of a recursive estimate of \( g(q_\alpha) \) or, for \( g \) Hölder continuous at \( q_\alpha \), 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, Méritier 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 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 \( q_\alpha \) 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
\[
\frac{\partial^k m}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}(x) - \frac{\partial^k m}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}(z)
\]
exists and satisfies
\[
\left| \frac{\partial^k m}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}(x) - \frac{\partial^k m}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}(z) \right| \leq C \cdot \| x - z \|^s
\]
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 known distribution \( \mu \). We assume that we have available a deterministic approximation \( \tilde{m}_n \) of \( m \) which satisfies
\[
\| \tilde{m}_n - m \|_{\infty, [-l_n, l_n]^d} \leq \log^{p+1}(n) \cdot n^{-p/d}
\]
for sufficiently large \( n \) for some \( 0 < p \leq d \), where \( l_n = \log(n) \). Set
\[
m_n = \tilde{m}_n - \log^{p+1}(n) \cdot n^{-p/d}.
\]
Then we have
\[ \|m_n - m\|_{\infty, [-l_n, l_n]^d} \leq 2 \cdot \log^{p+1}(n) \cdot n^{-p/d} \]  
(5)
and
\[ m_n(x) \leq m(x) \quad \text{for all } x \in [-l_n, l_n]^d \]  
(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 \( q_\alpha \). 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
\[ H_n(B) = \frac{\mu \left( \left\{ x \in [-l_n, l_n]^d : m_n(x) \leq Z_n \right\} \cup \left( [-l_n, l_n]^d \cap B \right) \right)}{G_n(Z_n)} \quad (B \in \mathcal{B}^d) \]
where \( \mathcal{B}^d \) is the set of all Borel sets in \( \mathbb{R}^d \) and where
\[ \bar{G}_n (z) = \mu \left( \left\{ x \in [-l_n, l_n]^d : m_n(x) \leq z \right\} \cup \left( [-l_n, l_n]^d \cap B \right) \right). \]  
(7)

By construction, the distribution \( H_n \) has the Radon-Nikodym derivative (conditional on \( Z_n \))
\[ \frac{dH_n}{d\mu}(x) = \frac{I_{\{m_n(x) \leq Z_n\}} \cdot I_{\{x \in [-l_n, l_n]^d\}} + I_{\{x \notin [-l_n, l_n]^d\}}}{G_n(Z_n)}. \]

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
\[ Z_{n+1} = Z_n - \frac{D_n}{n} \cdot \left( I_{\{m(X_n^{(IS)}) \leq Z_n\}} \cdot \bar{G}_n (Z_n) - \alpha \right), \]  
(8)
where \( D_n = \log^2(n) \) and
\[ \bar{G}_n (z) = \frac{1}{n} \sum_{i=1}^{n} (I_{\{m_n(X_{n,i}) \leq z, X_{n,i} \in [-l_n, l_n]^d\}} + I_{\{X_{n,i} \notin [-l_n, l_n]^d\}}) \quad (z \in \mathbb{R}). \]

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

**Theorem 1** Let \( X, X_{1,1}, X_{2,1}, X_{2,2}, X_{3,1}, X_{3,2}, X_{3,3}, \ldots \) 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 \( q_\alpha \) be the \( \alpha \)-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 \( q_\alpha \).
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

$$\mathbb{P}\{X \not\in [-\log(n), \log(n)]^d\} > 0 \quad (n \in \mathbb{N}) \quad \text{and} \quad \mathbb{E}\{e^{\|X\|}\} < \infty$$

imply

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

**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 $\mu$ 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

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

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. However, in this case the numerical computation of the above integral risks to lose the advantage of the recursive algorithm in terms of computational time.

**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 $\infty$ 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

\[
Z_{n+1} = Z_n - \frac{D_n}{n} \cdot (G(Z_n) - G(q_n)) + \frac{D_n}{n} \cdot V_n,
\]

where

\[
V_n = G(Z_n) - I_{\{m(X_n^{IS}) \leq Z_n\}} \cdot \hat{G}_n(Z_n)
\]

\[
= G(Z_n) - I_{\{m(X_n^{IS}) \leq Z_n\}} \cdot \hat{G}_n(Z_n) - I_{\{m(X_n^{IS}) \leq Z_n\}} \cdot \left( \hat{G}_n(Z_n) - \tilde{G}_n(Z_n) \right)
\]

and

\[
\tilde{G}_n(z) = P \{ m_n(X) \leq z, X \in [-l_n, l_n]^d \} + P \{ X \notin [-l_n, l_n]^d \}
\]

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

In the first step of the proof we show

\[
\left| E \left\{ I_{\{m(X_n^{IS}) \leq Z_n\}} \cdot \tilde{G}_n(Z_n) \mid Z_n = z \right\} - G(z) \right| \leq c_1/n \tag{11}
\]

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

\[
E \left\{ I_{\{m(X_n^{IS}) \leq Z_n\}} \mid Z_n = z \right\}
= \int I_{\{m(t) \leq z\}} dH_n(t)
= \frac{1}{\tilde{G}_n(z)} \int I_{\{m(t) \leq z\}} \cdot \left( I_{\{t \notin [-l_n, l_n]^d : m_n(t) \leq z\}} + I_{\{t \notin [-l_n, l_n]^d\}} \right) d\mu(t)
= \frac{P \{ m(X) \leq z, X \in [-l_n, l_n]^d \} + P \{ X \notin [-l_n, l_n]^d \}}{\tilde{G}_n(z)}, \tag{12}
\]

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

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

By the Markov inequality and assumption (9) we get

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

which implies (11).
In the second step of the proof we show

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

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

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

we get

\[
\begin{align*}
\text{Var} \left\{ I_{\{m(X_n^{(IS)} \leq Z_n)\}} \cdot \hat{G}_n(Z_n) \big| Z_n = z \right\} &= \hat{G}_n(z)^2 \cdot \left( \mathbb{E}\{ I_{\{m(X_n^{(IS)} \leq z)\}} | Z_n = z \} - \left( \mathbb{E}\{ I_{\{m(X_n^{(IS)} \leq z)\}} | Z_n = z \} \right)^2 \right) \\
&= \hat{G}_n(z) \cdot \left( P\{m(X) \leq z, X \in [-l_n, l_n]^d\} + P\{X \notin [-l_n, l_n]^d\} \right) \\
&\quad - \left( P\{m(X) \leq z, X \in [-l_n, l_n]^d\} + P\{X \notin [-l_n, l_n]^d\} \right)^2 \\
&= \left( P\{m_n(X) \leq z, X \in [-l_n, l_n]^d\} - P\{m(X) \leq z, X \in [-l_n, l_n]^d\} \right) \\
&\quad \cdot \left( P\{m(X) \leq z, X \in [-l_n, l_n]^d\} + P\{X \notin [-l_n, l_n]^d\} \right) \\
&\leq P\{m_n(X) \leq z, X \in [-l_n, l_n]^d\} - P\{m(X) \leq z, X \in [-l_n, l_n]^d\} \\
&\leq P\{m(X) \leq z + 2 \cdot \log^{p+1}(n) \cdot n^{-p/d}, X \in [-l_n, l_n]^d\} - P\{m(X) \leq z, X \in [-l_n, l_n]^d\} \\
&= \int_{[-l_n, l_n]^d} (I_{\{m(x) \leq z + 2 \cdot \log^{p+1}(n) \cdot n^{-p/d}\}} - I_{\{m(x) \leq z\}}) P_X(dx) \\
&\leq G(z + 2 \cdot \log^{p+1}(n) \cdot n^{-p/d}) - G(z) \\
&\leq c_2 \cdot \log^{p+1}(n) \cdot n^{-p/d},
\end{align*}
\]

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

\[
Z_n \rightarrow q_n \quad \text{a.s.} \quad \tag{14}
\]

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

\[
\mathbb{E} \left\{ I_{\{m(X_n^{(IS)} \leq Z_n)\}} \cdot \left( \hat{G}_n(Z_n) - \hat{G}_n(Z_n) \right) \big| F_{n-1} \right\} = 0
\]

and

\[
\text{Var} \left\{ I_{\{m(X_n^{(IS)} \leq Z_n)\}} \cdot \left( \hat{G}_n(Z_n) - \hat{G}_n(Z_n) \right) \big| F_{n-1} \right\} \leq \mathbb{E} \left\{ \left( \hat{G}_n(Z_n) - \hat{G}_n(Z_n) \right)^2 \big| F_{n-1} \right\} \leq \frac{1}{n}. \tag{15}
\]

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

\[
\mathbb{E}\{V_n|F_{n-1}\} \leq \frac{c_3}{n} \tag{16}
\]
and
\[
\mathbb{E} \{ V_n^2 | \mathcal{F}_{n-1} \} \leq \left( \mathbb{E} \{ V_n | \mathcal{F}_{n-1} \} \right)^2 + \text{Var} \{ V_n | \mathcal{F}_{n-1} \} \\
\leq \frac{c_3^2}{n^2} + 2 \cdot \text{Var} \left\{ I_{\{ m(X_n^{(i)}) \leq Z_n \}} \cdot \tilde{G}_n(Z_n) \right\} \mathcal{F}_{n-1} \\
+ 2 \cdot \text{Var} \left\{ I_{\{ m(X_n^{(i)}) \leq Z_n \}} \cdot (\tilde{G}_n(Z_n) - \tilde{G}_n(Z_n)) \right\} \mathcal{F}_{n-1} \\
\leq \frac{c_3^2}{n^2} + 2 \cdot c_2 \cdot \log^{p+1}(n) \cdot n^{-p/d} + \frac{2}{n} \\
\leq c_4 \cdot \log^{p+1}(n) \cdot n^{-p/d}, (17)
\]
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
\[
\log^{-6-p}(n) \cdot n^{1+p/d} \cdot \mathbb{E} \{ |Z_n - q_\alpha|^2 \cdot I_{B_N} \} \to 0 \quad (n \to \infty) \quad (18)
\]
for all sufficiently large \( N \in \mathbb{N} \). Because of (14) we have \( \mathbb{P}(B_N) \to 1 \quad (N \to \infty) \), consequently the assertion of Theorem 1 is implied by
\[
\mathbb{P} \left\{ |Z_n - q_\alpha| > c_5 \cdot \log^{3+p/2}(n) \cdot n^{-1/2-p/(2d)} \quad \text{and } B_N \text{ holds} \right\} \to 0 \quad (n \to \infty)
\]
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
\[
\mathbb{E} \{ |Z_{n+1} - q_\alpha|^2 | \mathcal{F}_{n-1} \} \leq \left( 1 - c_6 \cdot \frac{D_n}{n} \right) |Z_n - q_\alpha|^2 + c_7 \cdot \frac{D_n^2}{n^2} \cdot \log^{p+1}(n) \cdot n^{-p/d} \quad (19)
\]
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 \( V_n = G(Z_n) - I_{\{ m(X_n^{(i)}) \leq Z_n \}} \cdot \tilde{G}_n(Z_n) \). Here \( Z_n \) and \( A_n \) are \( \mathcal{F}_{n-1} \) measurable. Using \( 2 \cdot a \cdot b \leq
\[ a^2/\delta + b^2 \cdot \delta \text{ for } a, b \in \mathbb{R} \text{ and } \delta > 0 \text{ this implies in case } |Z_n - q_\alpha| < \epsilon \]

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

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

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

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

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 \( n \)

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

\[
\leq \mathbf{E} \left\{ |Z_n - q_\alpha|^2 \cdot I_{B_{N,n-1}} \right\}
\]

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

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

An iterative application of this argument yields for any sufficiently large \( N \in \mathbb{N} \) and \( \lfloor n/2 \rfloor > N \)

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

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

\[
\leq c_8 \cdot \log^{p+5}(n) \cdot n^{-1-p/d} + \epsilon^2 \cdot \exp \left( -c_9 \cdot \log^2(\lfloor n/2 \rfloor) \right).
\]

The proof is complete. \( \square \)

4 Acknowledgment

The first author would like to thank the German Research Foundation (DFG) for funding this project within the Collaborative Research Centre 805. The second author would like to acknowledge the support of Natural Sciences and Engineering Research Council of Canada.
References


Michael Kohler  
Fachbereich Mathematik  
Technische Universität Darmstadt  
Schloßgartenstr. 7  
64289 Darmstadt  
Germany  
E-Mail: kohler@mathematik.tu-darmstadt.de

Adam Krzyżak  
Department of Computer Science and Software Engineering  
Concordia University  
1455 De Maisonneuve Blvd. West  
Montreal  
Quebec  
Canada H3G 1M8  
E-Mail: krzyzak@cs.concordia.ca

Harro Walk  
Universität Stuttgart  
Fachbereich Mathematik  
Pfaffenwaldring 57  
70569 Stuttgart  
Germany  
E-Mail: walk@mathematik.uni-stuttgart.de
Erschienene Preprints ab Nummer 2007/2007-001
Komplette Liste:  http://www.mathematik.uni-stuttgart.de/preprints

2014-010 Köster, I.: Finite Groups with Sylow numbers \( \{q^a, a, b\} \)
2014-009 Kohnert, D.: Hausdorff Dimension of Rings
2014-008 Steinwart, I.: Measuring the Capacity of Sets of Functions in the Analysis of ERM
2014-007 Steinwart, I.: Convergence Types and Rates in Generic Karhunen-Loève Expansions with Applications to Sample Path Properties
2014-004 Markhasin, L.: \( L_2\)- and \( S_{p,d}^B\)-discrepancy of (order 2) digital nets
2014-003 Markhasin, L.: Discrepancy and integration in function spaces with dominating mixed smoothness
2014-002 Eberts, M.; Steinwart, I.: Optimal Learning Rates for Localized SVMs
2014-001 Giesselmann, J.: A relative entropy approach to convergence of a low order approximation to a nonlinear elasticity model with viscosity and capillarity
2013-016 Steinwart, I.: Fully Adaptive Density-Based Clustering
2013-015 Steinwart, I.: Some Remarks on the Statistical Analysis of SVMs and Related Methods
2013-014 Rohde, C.; Zeiler, C.: A Relaxation Riemann Solver for Compressible Two-Phase Flow with Phase Transition and Surface Tension
2013-012 Moroianu, A.; Semmelmann, U.: Generalized Killing Spinors on Spheres
2013-010 Corli, A.; Rohde, C.; Schleper, V.: Parabolic Approximations of Diffusive-Dispersive Equations
2013-009 Nava-Yazdani, E.; Polthier, K.: De Casteljau’s Algorithm on Manifolds
2013-008 Bächle, A.; Margolis, L.: Rational conjugacy of torsion units in integral group rings of non-solvable groups
2013-007 Knarr, N.; Stroppel, M.J.: Heisenberg groups over composition algebras
2013-006 Knarr, N.; Stroppel, M.J.: Heisenberg groups, semifields, and translation planes
2013-003 Kabil, B.; Rohde, C.: The Influence of Surface Tension and Configurational Forces on the Stability of Liquid-Vapor Interfaces


2012-018 Kimmerle, W.; Konovalov, A.: On the Prime Graph of the Unit Group of Integral Group Rings of Finite Groups II

2012-017 Stroppel, B.; Stroppel, M.: Desargues, Doily, Dualities, and Exceptional Isomorphisms


2012-015 Steinke, G.F.; Stroppel, M.J.: Simple groups acting two-transitively on the set of generators of a finite elation Laguerre plane

2012-014 Steinke, G.F.; Stroppel, M.J.: Finite elation Laguerre planes admitting a two-transitive group on their set of generators


2012-012 Moroianu, A.; Semmelmann, U.: Weakly complex homogeneous spaces

2012-011 Moroianu, A.; Semmelmann, U.: Invariant four-forms and symmetric pairs

2012-010 Hamilton, M.J.D.: The closure of the symplectic cone of elliptic surfaces

2012-009 Hamilton, M.J.D.: Iterated fibre sums of algebraic Lefschetz fibrations

2012-008 Hamilton, M.J.D.: The minimal genus problem for elliptic surfaces

2012-007 Ferrario, P.: Partitioning estimation of local variance based on nearest neighbors under censoring

2012-006 Stroppel, M.: Buttons, Holes and Loops of String: Lacing the Doily

2012-005 Hantsch, F.: Existence of Minimizers in Restricted Hartree-Fock Theory

2012-004 Grundhöfer, T.; Stroppel, M.; Van Maldeghem, H.: Unitals admitting all translations

2012-003 Hamilton, M.J.D.: Representing homology classes by symplectic surfaces

2012-002 Hamilton, M.J.D.: On certain exotic 4-manifolds of Akhmedov and Park

2012-001 Jentsch, T.: Parallel submanifolds of the real 2-Grassmannian

2011-028 Spreer, J.: Combinatorial 3-manifolds with cyclic automorphism group


2011-026 Müller, S.: Bootstrapping for Bandwidth Selection in Functional Data Regression


2011-023 Györfi, L.; Walk, H.: Strongly consistent nonparametric tests of conditional independence

2011-022 Ferrario, P.G.; Walk, H.: Nonparametric partitioning estimation of residual and local variance based on first and second nearest neighbors

2011-021 Eberts, M.; Steinwart, I.: Optimal regression rates for SVMs using Gaussian kernels

2011-018 Hänel, A.; Schulz, C.; Wirth, J.: Embedded eigenvalues for the elastic strip with cracks
2011-017 Wirth, J.: Thermo-elasticity for anisotropic media in higher dimensions
2011-015 Ferrario, P.: Nonparametric Local Averaging Estimation of the Local Variance Function
2011-012 Knarr, N.; Stroppel, M.: Baer involutions and polarities in Moufang planes of characteristic two
2011-011 Knarr, N.; Stroppel, M.: Polarities and planar collineations of Moufang planes
2011-010 Jentsch, T.; Moroianu, A.; Semmelmann, U.: Extrinsic hyperspheres in manifolds with special holonomy
2011-008 Stroppel, M.: Orthogonal polar spaces and unitals
2011-007 Nagl, M.: Charakterisierung der Symmetrischen Gruppen durch ihre komplexe Gruppenalgebra
2011-006 Solanes, G.; Teufel, E.: Horo-tightness and total (absolute) curvatures in hyperbolic spaces
2011-005 Ginoux, N.; Semmelmann, U.: Imaginary Kählerian Killing spinors I
2011-002 Alexandrov, B.; Semmelmann, U.: Deformations of nearly parallel \( G_2 \)-structures
2010-018 Kimmerle, W.; Konovalov, A.: On integral-like units of modular group rings
2010-017 Gauduchon, P.; Moroianu, A.; Semmelmann, U.: Almost complex structures on quaternion-Kähler manifolds and inner symmetric spaces
2010-016 Moroianu, A.; Semmelmann, U.: Clifford structures on Riemannian manifolds
2010-015 Grafarend, E.W.; Kühlner, W.: A minimal atlas for the rotation group \( SO(3) \)
2010-014 Weidl, T.: Semiclassical Spectral Bounds and Beyond
2010-013 Stroppel, M.: Early explicit examples of non-desarguesian plane geometries
2010-012 Effenberger, F.: Stacked polytopes and tight triangulations of manifolds
2010-010 Kohler, M.; Krzyżak, A.; Walk, H.: Estimation of the essential supremum of a regression function
2010-008  Poppitz, S.; Stroppel, M.:  Polarieties of Schellhammer Planes
2010-005  Kaltenbacher, B.; Walk, H.:  On convergence of local averaging regression function estimates for the regularization of inverse problems
2010-004  Kühnel, W.; Solanes, G.:  Tight surfaces with boundary
2010-003  Kohler, M; Walk, H.:  On optimal exercising of American options in discrete time for stationary and ergodic data
2010-002  Gulde, M.; Stroppel, M.:  Stabilizers of Subspaces under Similitudes of the Klein Quadric, and Automorphisms of Heisenberg Algebras
2010-001  Leitner, F.:  Examples of almost Einstein structures on products and in cohomogeneity one
2009-008  Griesemer, M.; Zenk, H.:  On the atomic photoeffect in non-relativistic QED
2009-007  Griesemer, M.; Moeller, J.S.:  Bounds on the minimal energy of translation invariant n-polaron systems
2009-005  Bächle, A, Kimmerle, W.:  Torsion subgroups in integral group rings of finite groups
2009-003  Walk, H.:  Strong laws of large numbers and nonparametric estimation
2009-002  Leitner, F.:  The collapsing sphere product of Poincaré-Einstein spaces
2009-001  Brehm, U.; Kühnel, W.:  Lattice triangulations of $E^3$ and of the 3-torus
2008-006  Kohler, M.; Krzyżak, A.; Walk, H.:  Upper bounds for Bermudan options on Markovian data using nonparametric regression and a reduced number of nested Monte Carlo steps
2008-004  Leitner, F.:  Conformally closed Poincaré-Einstein metrics with intersecting scale singularities
2008-003  Effenberger, F.; Kühnel, W.:  Hamiltonian submanifolds of regular polytope
2008-002  Hertweck, M.; Höft, C.R.; Kimmerle, W.:  Finite groups of units and their composition factors in the integral group rings of the groups $PSL(2,q)$
2007-006  Weidl, T.:  Improved Berezin-Li-Yau inequalities with a remainder term
2007-005  Frank, R.L.; Loss, M.; Weidl, T.:  Polya's conjecture in the presence of a constant magnetic field
2007-003  Lesky, R.H.; Racke, R.:  Elastic and electro-magnetic waves in infinite waveguides
2007-002  Teufel, E.:  Spherical transforms and Radon transforms in Moebius geometry
2007-001  Meister, A.:  Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions