Journal of Machine Learning Research 24 (2023) 1-25 Submitted 1/22; Revised 11/22; Published 2/23

Recursive Quantile Estimation: Non-Asymptotic Confidence Bounds

Likai Chen LIKAI.CHEN@WUSTL.EDU
Department of Mathematics and Statistics
Washington University in St.Louis
St.Louis, MO, USA

Georg Keilbar GEORG.KEILBAR@UNIVIE.AC.AT
Department of Statistics and Operations Research
University of Vienna
Vienna, Austria

Wei Biao Wu WBWU@GALTON.UCHICAGO.EDU
Department of Statistics
University of Chicago
Chicago, IL, USA

Editor: Aurelien Garivier

Abstract

This paper considers the recursive estimation of quantiles using the stochastic gradient descent (SGD) algorithm with Polyak-Ruppert averaging. The algorithm offers a computationally and memory efficient alternative to the usual empirical estimator. Our focus is on studying the non-asymptotic behavior by providing exponentially decreasing tail probability bounds under mild assumptions on the smoothness of the density functions. This novel non-asymptotic result is based on a bound of the moment generating function of the SGD estimate. We apply our result to the problem of best arm identification in a multi-armed stochastic bandit setting under quantile preferences.

Keywords: Finite sample bounds, quantiles, stochastic gradient descent, Polyak-Ruppert averaging, recursive estimation

1. Introduction

The emergence of big data has brought serious challenges to traditional deterministic optimization methods. In many applications, the data arrives sequentially and the sample size is so large that a storage of the entire dataset is infeasible. In these situations, the stochastic gradient descent (SGD) algorithm (Robbins and Monro, 1951; Kiefer and Wolfowitz, 1952) provides a scalable alternative for estimation. The algorithm updates estimates recursively according to the gradient of the objective function. This recursive nature of the SGD algorithm makes it computationally and memory efficient. Thus, SGD is naturally suited for online learning problems. The convergence properties of the algorithm and its asymptotics have been analyzed thoroughly (Robbins and Siegmund, 1971; Ljung, 1977; Lai, 2003). Notable applications include anomaly detection (Ahmad et al., 2017) and matrix factorization (Mairal et al., 2010). The large sample properties of SGD are well established. For an

©2023 Likai Chen, Georg Keilbar and Wei Biao Wu.

License: CC-BY 4.0, see https://creativecommons.org/licenses/by/4.0/. Attribution requirements are provided at http://jmlr.org/papers/v24/22-0021.html.

averaged version of the algorithm (Ruppert, 1988; Polyak and Juditsky, 1992), it can be shown that the estimator converges with the optimal parametric rate to a Gaussian limit. To conduct inference, Chen et al. (2020) and Zhu et al. (2021) proposed methods to estimate the covariance matrix of the parameter estimates. Fang et al. (2018) and Fang (2019) proposed online bootstrap procedures to measure the uncertainty of SGD estimates.

In this paper, we consider the recursive estimation of quantiles. This classical problem is of great importance in a variety of applications ranging from finance (Engle and Manganelli, 2004), health care (Wang et al., 2018) and survival studies (Peng and Huang, 2008). The large sample properties of the traditional estimator for the quantile, which is based on order statistics, were studied in Bahadur (1966) and Kiefer (1967). The downside of the empirical estimator is that it is not memory-efficient in the presence of large and sequentially arriving datasets. Volgushev et al. (2019) and Chen et al. (2019) proposed new algorithms for the estimation of conditional quantiles taking into account these computational issues and memory requirements. Moreover, asymptotic normality provides limited insights on the performance of the estimator in finite samples. The study of the finite-sample behavior is an important task since in practical problems the sample size is always finite. Usually, obtaining such results requires more mathematical effort than merely obtaining asymptotic results and typically, this involves more restrictive assumptions on the tail behavior and the existence of moments.

The aim of this paper is to study the tail probability of the averaged version of the SGD algorithm for estimating quantiles in finite samples. As our main result, we derive an exponential bound on the tail probability, while only imposing weak assumptions on the smoothness of the distribution function. The proof relies on the decomposition of the gradient in the SGD algorithm into a martingale difference part, a shift part and a remainder part. Another key component of the proof is a novel bound on the moment generating function of the SGD estimate, which exploits the recursive behavior of the algorithm and the boundedness of the quantile score function.

The non-asymptotic behavior of the SGD estimate of quantiles with Polyak-Ruppert averaging was studied in Costa and Gadat (2021). They derived finite sample bounds for the L^p loss. Another closely related paper is Cardot et al. (2013), who proposed a SGD estimation procedure for the geometric median (Haldane, 1948), which is a multi-dimensional generalization of the median. The SGD solution has the same asymptotic behavior as the empirical estimator of the geometric median. The result can be easily generalized to the geometric quantiles proposed by Chaudhuri (1996). In a subsequent paper, Cardot et al. (2017) studied the finite sample performance of the SGD algorithm. In particular, they derived non-asymptotic confidence balls for the averaged version of the algorithm. While the geometric median is a generalization of the classic median, their main result does not directly apply to this univariate case. One contribution of our paper is the extension to the result of Cardot et al. (2017) to the univariate median with exponential tail bound. It should be noted that their bound is only valid for a sample size exceeding a certain rank. In contrast, our new non-asymptotic bound is valid for each finite sample size. The reason is that the bound derived in this paper is based on a bound of the moment generating function, while previous results only relied on a finite-sample bound for the L^2 risk.

We apply our novel finite sample bound to the problem of best arm identification in the context of stochastic multi-armed bandit models. We refer to the monograph of Lattimore and Szepesvári (2020) for an extensive overview on bandit algorithms. While the majority of the research on multi-armed bandits focused one the mean case, there are important arguments in favor of looking at the quantiles. First, quantiles are more robust location parameters compared to the mean and second, depending on the context of the application, the focus can be on different parts of the distribution. Previously, the problem of best arm identification in a quantile bandit settings were studied in Nikolakakis et al. (2021), Zhang and Ong (2021) and in Howard and Ramdas (2022). We consider a quantile version of the successive reject algorithm of Audibert et al. (2010) and the sequential elimination algorithm of Karnin et al. (2013).

The remainder of the paper is organized as follows. Section 2 provides an overview of the problem and introduces the SGD algorithm and its averaged version. The main theoretical results are presented in Section 3. In Section 4 we apply our probability bound to the problem of best arm identification. Section 5 concludes and Section 6 provides the proofs of the main results.

We now introduce some notation. For a function g, define $|g|_\infty := \sup_x |g(x)|$. For two sequences of positive numbers, (a_n) and (b_n), write $a_n \lesssim b_n$ if there there exists a positive constant C such that $a_n/b_n \leq C$ for all n. Alternatively, we write $a_n = O(b_n)$. We write $b_n = \Omega(a_n)$ if $a_n = O(b_n)$.

2. Overview of the Problem

In this paper, we are interested in estimating quantiles for high dimensional data. For a random vector $X = (X_1, X_2, \ldots, X_p)^\top \in \mathbb{R}^p$, the τ-th quantile of coordinate X_i is defined as the minimizer of the quantile loss function (see Figure 1),

$$Q_i(\tau) := \operatorname{argmin}_{x\in\mathbb{R}} \mathbb{E}\left\{(X_i - x)\left(\tau - \mathbf{1}_{x\geq X_i}\right)\right\}. \tag{1}$$

Denote the distribution function of X_i as $F_i(x)$, under the assumption of F_i being continuous, we have $F_i(Q_i(\tau)) = \tau$.

In financial applications, X_i could be the stock return of firm i and $Q_i(\tau)$ the corresponding value-at-risk (VaR) at confidence level τ. In survival studies, the variable of interest is the survival time of an individual.

Let $X_{i,1}, \ldots, X_{i,n}$ denote i.i.d. copies of the coordinates X_i, $1 \leq i \leq p$. Then a natural empirical estimator of $Q_i(\tau)$ takes the form

$$\widehat{Q}_i(\tau) := \operatorname{argmin}_{x\in\mathbb{R}} \sum_{k=1}^{n} \left\{(X_{i,k} - x)\left(\tau - \mathbf{1}_{X_{i,k}\geq x}\right)\right\}. \tag{2}$$

Asymptotic properties of the empirical estimator are well studied (Bahadur, 1966; Kiefer, 1967). It is well known that the estimator is strongly consistent and has an asymptotic normal distribution.

One of the problems of the classical estimator is that it is not memory efficient in the case of streaming data. Munro and Paterson (1980) showed that any algorithm exactly calculating quantiles in q passes requires $\Omega(n^{1/q})$ memory. Recent developments on estimating

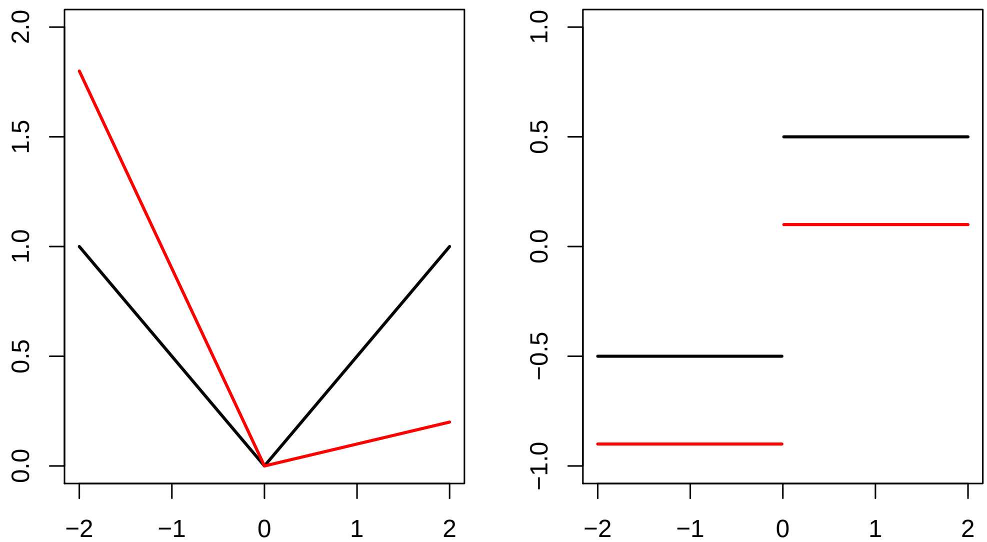

Figure 1: Quantile loss function (left panel) and score function (right panel) for quantile level $\tau = 0.5$ (black line) and for $\tau = 0.1$ (red line).)

quantiles in the case of streaming data are discussed in Luo et al. (2016). We therefore follow a different approach based on the SGD algorithm of Robbins and Monro (1951). Starting from a constant initial value $Y_{i,1}(\tau) = y_i$, with $\max_{1\le i\le p} |y_i| \le c_y$ and some constant $c_y > 0$, we have

$$Y_{i,k+1}(\tau) = Y_{i,k}(\tau) + \gamma_k\Big(\tau \mathbf{1}_{X_{i,k+1} > Y_{i,k}(\tau)} - (1-\tau)\mathbf{1}_{X_{i,k+1} \le Y_{i,k}(\tau)}\Big), \tag{3}$$

where the sequence of learning rates (γ_k) determines the convergence of the algorithm. In particular, the following assumptions need to be fulfilled,

$$\sum_{k=1}^{\infty} \gamma_k^2 < \infty \quad \text{and} \quad \sum_{k=1}^{\infty} \gamma_k = \infty.$$

The first condition ensures the convergence to some point in $\mathbb{R}$, while the second condition ensures the convergence to a unique minimizer $Q_i(\tau)$. We consider sequences of step sizes in the form of $\gamma_k = c_\gamma k^{-\beta}$, with $1/2 < \beta < 1$, and some constant $c_\gamma > 0$.

Due to favorable asymptotic properties, we consider an averaged version of the algorithm, which takes the form

$$\bar{Y}_{i,n}(\tau) = \sum_{k=1}^{n} Y_{i,k}(\tau)/n,$$

where $\bar{Y}_{i,0} = 0$. Note that the averaged estimator also can be updated recursively by $\bar{Y}_{i,n}(\tau) = (n-1)\bar{Y}_{i,n-1}(\tau)/n + Y_{i,n}/n$. Such an averaging step is known as Polyak-Ruppert averaging (Ruppert, 1988; Polyak and Juditsky, 1992). Estimators based on the averaged SGD algorithm converge almost surely to the true parameter and have the same Gaussian limit distribution as the empirical estimator defined in (2). For the estimation of quantiles, asymptotic normality of the solution of the averaged algorithm was shown by Bardou et al. (2009). In particular, it holds

$$\sqrt{n}\left\{\bar{Y}_{i,n}(\tau)-Q_i(\tau)\right\} \xrightarrow{d} N\left(0, \frac{\tau(1-\tau)}{f_i^2\left\{Q_i(\tau)\right\}}\right).$$

However, these asymptotic results do not provide any insights on how well the estimator will perform in finite samples. Recently, Gadat and Panloup (2023) derived non-asymptotic bounds on the L^2-loss for the recursive quantile estimator based on Polyak-Ruppert averaging. It is shown that for each $n \geq 1$ it holds that, given the optimal choice of the learning rate parameter β,

$$\mathbb{E}\big\{\bar{Y}_{i,n}(\tau)-Q_i(\tau)\big\}^2 = \frac{\tau(1-\tau)}{f_i^2\left\{Q_i(\tau)\right\}n} + O\left(n^{-5/4}\right).$$

Recently, Cardot et al. (2017) analyzed the finite sample tail behavior of the Polyak-Ruppert algorithm for estimating the geometric median. The geometric median is a multivariate generalization of the univariate median (Haldane, 1948; Minsker, 2015), which can easily be generalized to geometric quantiles (Chaudhuri, 1996). The geometric median is defined by

$$m := \operatorname{argmin}_{x\in H}\mathbb{E}\left(\|X-x\|-\|X\|\right),$$

where X is a random variable taking values in a separable Hilbert space H with corresponding norm $\|\cdot\|$. The algorithm and its asymptotic properties are studied in Cardot et al. (2013). In particular, it was shown that the algorithm is strongly consistent and asymptotically normal. Cardot et al. (2017) further studied the non-asymptotic properties. In the main theorem of the paper, the authors derived non-asymptotic confidence balls for the averaged solutions. Although the geometric median generalizes the univariate median, the asymptotic normality as well as the result on non-asymptotic confidence bounds only hold for the case of the dimensionality of the data being larger than 2, thus excluding the univariate median. The reason for this is condition (A3) in Cardot et al. (2017), which requires the existence of a constant C such that for all $x \in H$,

$$\mathbb{E}\left(\|X-x\|^{-2}\right) \leq C. \tag{4}$$

This condition does not hold for $H=\mathbb{R}^d$ with $d<3$. As a second drawback of the tail bound the sample size is required to exceed a certain rank n_δ, which might be prohibitively large. The order of the rank is $O((\frac{1}{\delta \log \delta})^6)$, which increases with decreasing confidence level δ. It is therefore crucial to obtain finite sample results which hold in dimension one and therefore include the important special case of the univariate median.

3. Theoretical Results

In this section, we shall present the main theoretical results of this paper. All proofs are deferred to Section 6. We are interested in the tail probability,

$$\mathbb{P}\Big(\max_{1\le i\le p}|\bar{Y}_{i,n}(\tau)-Q_i(\tau)|\ge x\Big),\ x>0.$$

At first, we derive a non-asymptotic probability bound for the averaged SGD solution for a single coordinate X_i. A simple tail bound can be obtained from the L^2 bound in Gadat and Panloup (2023) and Markov's inequality

$$\mathbb{P}\big(|\bar{Y}_{i,n}(\tau)-Q_i(\tau)|>x\big)\le\frac{\mathbb{E}\big\{\bar{Y}_{i,n}(\tau)-Q_i(\tau)\big\}^2}{x^2}=O\left(\frac{1}{nx^2}\right),\ \text{where } x>0.$$

However, the resulting probability bound is only algebraically decreasing in n and x. In the following we will present a much sharper bound which decreases exponentially fast.

3.1 A Bound on the Moment Generating Function

We need to impose the following smoothness condition on the density of X_i and the boundedness of the true quantiles within the interested region. Theses conditions are standard in the quantile literature.

Assumption 3.1 *Let $\tau_0<\tau_1$ be some finite constants in $(0,1)$. Assume the random variable X_i has a differentiable density function $f_i(x)$, with $c_\tau:=\min_{1\le i\le p}\inf_{\tau\in[\tau_0,\tau_1]}f_i(Q_i(\tau))>0$ and $\max_{1\le i\le p}|f_i'|_\infty\le c_f<\infty$. Moreover for some constant $M>0$, $|Q_i(\tau)|\le M$, for any $1\le i\le p$ and $\tau_0\le\tau\le\tau_1$.*

This assumption ensures the existence of a unique theoretical quantile. We require no assumptions on the tail behavior nor on the existence of any moments of X_i. In order to derive the tail probability bound, we directly bound the moment generating function of the SGD solution without averaging, $Y_{i,n}(\tau)$, in the following theorem.

Theorem 3.1 *Under Assumption 3.1, for $0<t\le cn^{(1-\beta)\beta}$ and $\tau_0\le\tau\le\tau_1$, we have*

$$\mathbb{E}(e^{t|Y_{i,n}(\tau)-Q_i(\tau)|})\le c'n^\beta, \tag{5}$$

where $c,c'>0$ are some constants independent of i,n,τ. Consequently, for any $x>0$,

$$\mathbb{P}\big(|Y_{i,n}(\tau)-Q_i(\tau)|>x\big)\le c'n^\beta\exp\big\{-cn^{\beta(1-\beta)}x\big\}. \tag{6}$$

The specific form of constants c and c' are presented in Subsection 6.1 in equations (20) and (21) respectively. The key idea of the proof is to exploit the recursive nature of the SGD algorithm in order to obtain a recursive bound on the moment generating function.

We now compare our bound (6) with the one in Cardot et al. (2017), namely the inequality on Page 10 in the Appendix of the latter paper which concerns medians with $\tau=1/2$:

$$\mathbb{P}\big(|Y_{i,n}(1/2)-Q_i(1/2)|>x\big)\le 2\exp\Big\{-\frac{C_0x^2n^\beta}{1+C_3x}\Big\}+\frac{C_1e^{-C_4n^{1-\beta}}}{x^2}+\frac{C_2}{n^\beta x}, \tag{7}$$

where $C_0, \ldots, C_4 > 0$ are constants. Asymptotically, the upper bound in (7) decays *polynomially* in x and n, while in comparison, our bound (6) is decreasing *exponentially fast*. We shall remark that their paper is concerned with the estimation of the geometric median, which is a multi-dimensional extension of the median, their theorems are not directly applicable in dimension one since condition (4) is violated in the latter case. Nonetheless, we show in Subsection 6.4 that their method can be extended to the univariate case so that (7) can be obtained.

3.2 Confidence Bounds for the Averaged SGD Algorithm

We now consider the averaged version of the SGD algorithm, which can have a better convergence property. The following theorem gives the non-asymptotic confidence bounds for the SGD algorithm with the Polyak-Ruppert averaging.

Theorem 3.2 *Under Assumption 3.1, for $\tau \in [\tau_0, \tau_1]$, we have for $n \geq 1$ and $x > 0$,*

$$\begin{aligned} &\mathbb{P}\big(|\bar{Y}_{i,n}(\tau) - Q_i(\tau)| > x\big) \\ \lesssim &n^{1+\beta}\exp\big\{ - cn^{(1-\beta^2)}x\big\} + n^{1+\beta}\exp\Big\{ - c'n^{\beta(1-\beta)}x^{1/2}\Big\} + \exp\big\{ - c''nx^2\big\}, \end{aligned} \tag{8}$$

where the constants in $\lesssim$, c, c', c'' are all independent of i, n, p. Consequently, we have the uniform upper bound,

$$\max_{1 \leq i \leq p} |\bar{Y}_{i,n}(\tau) - Q_i(\tau)| = O_{\mathbb{P}}(\log^2(np)n^{-2\beta(1-\beta)}). \tag{9}$$

This implies that if $\log(p)/n^{\beta(1-\beta)} \to 0$, then we have the uniform consistency.

The specific form of these constants can be found in equation (25) in Subsection 6.2. The uniform consistency in Theorem 3.2 allows extra high-dimensional settings with $p = \exp(o(n^{\beta(1-\beta)}))$. While our tail probability bound decreases with an exponential rate, we cannot make a statement whether the bound is optimal.

The proof of Theorem 3.2 relies on a decomposition of $Y_{i,k}(\tau)$ into a martingale difference part, a shift part and the remainder. This decomposition scheme is also used in Cardot et al. (2017). Note that (3) can be rewritten as

$$Y_{i,k+1}(\tau) = Y_{i,k}(\tau) + \gamma_k Z_{i,k+1}(\tau), \tag{10}$$

where

$$Z_{i,k+1}(\tau) := \tau\mathbf{1}_{X_{i,k+1} > Y_{i,k}(\tau)} - (1-\tau)\mathbf{1}_{X_{i,k+1} \leq Y_{i,k}(\tau)} = \tau - \mathbf{1}_{X_{i,k+1} \leq Y_{i,k}(\tau)}. \tag{11}$$

To obtain the martingale difference part, we further decompose the term $Z_{i,k+1}(\tau)$ into $\xi_{i,k+1} = Z_{i,k+1}(\tau) - \mathbb{E}(Z_{i,k+1}(\tau)|\mathcal{F}_{i,k})$ and $\mathbb{E}(Z_{i,k+1}(\tau)|\mathcal{F}_{i,k})$, where $\mathcal{F}_{i,k} = (X_{i,k}, X_{i,k-1}, \ldots)$. Hence (10) can be rewritten into

$$\gamma_k^{-1}(Y_{i,k+1}(\tau) - Y_{i,k}(\tau)) = \xi_{i,k+1} + \mathbb{E}(Z_{i,k+1}(\tau)|\mathcal{F}_{i,k}).$$

Furthermore, we have $\mathbb{E}(Z_{i,k+1}(\tau)|\mathcal{F}_k) = c(Y_{i,k}(\tau) - Q_i(\tau)) + \rho_{i,k}$, where $c = -f_i(Q_i(\tau))$, $c(Y_{i,k}(\tau) - Q_i(\tau))$ is the linear part and $\rho_{i,k}$ represents the remainder term. Hence we have the decomposition

$$c(Y_{i,k}(\tau) - Q_i(\tau)) = \gamma_k^{-1}(Y_{i,k+1}(\tau) - Y_{i,k}(\tau)) - \xi_{i,k+1} - \rho_{i,k}.$$

Let $\bar{\xi}_{i,n} = n^{-1}\sum_{k=1}^n \xi_{i,k+1}$ and $\bar{\rho}_{i,n} = n^{-1}\sum_{k=1}^n \rho_{i,k}$. Then

$$c(\bar{Y}_{i,n}(\tau) - Q_i(\tau)) = \sum_{k=1}^n \gamma_k^{-1}(Y_{i,k+1}(\tau) - Y_{i,k}(\tau))/n - \bar{\xi}_{i,n} - \bar{\rho}_{i,n}.$$

To bound the first and the third terms, we can adopt our bound on the moment generating function obtained in Theorem 3.1. For the second term, the martingale difference part, we shall apply Azuma's concentration inequality. Details are provided in Subsection 6.2.

In the following, we take a closer look at the three terms on the right hand side of the probability bound in (8). We want to determine which terms are the leading ones in different scenarios.

Remark 3.3 *When $x \geq c_1 n^{-1/2}$ for some constant $c_1 > 0$, then the last term in (8) will be dominated by the second one, that is*

$$\mathbb{P}\big(|\bar{Y}_{i,n}(\tau) - Q_i(\tau)| > x\big) \lesssim n^{1+\beta}\exp\big\{-cn^{(1-\beta^2)}x\big\} + n^{1+\beta}\exp\Big\{-c'n^{\beta(1-\beta)}x^{1/2}\Big\}. \tag{12}$$

In addition, if $x \gg n^{2\beta-2}$, then the second term $n^{1+\beta}\exp\{-c'n^{\beta(1-\beta)}x^{1/2}\}$ would dominate, otherwise, the first term $n^{1+\beta}\exp\{-cn^{(1-\beta^2)}x\}$ would be the leading one.

Figure 2 visualizes the simulation results for the tail probabilities for the averaged algorithm. The simulation confirms the statements of exponential decay bounds, and the asymptotic algebraic bounds (7) and (13) can be too crude.

Theorem 3.4 provides an asymptotic algebraic bound for the average SGD solution, $\bar{Y}_{i,n}(\tau)$, which is a one-dimensional analogue of the geometric median in Cardot et al. (2017). The conclusion of the comparison between (6) and (7) also applies here: our bound (8) decreases exponentially fast.

Theorem 3.4 *Under Assumption 3.1, for $\tau \in [\tau_0, \tau_1]$, we have for $n \geq 1$ and $x > 0$,*

$$\mathbb{P}\big(|\bar{Y}_{i,n}(\tau) - Q_i(\tau)| > x\big) \leq 2\exp\big\{-nx^2/2\big\} + \frac{c}{xn^{1-\beta/2}} + \frac{c'}{xn} + \frac{c''}{xn^{\beta}}, \tag{13}$$

where c, c', c'' are independent of i, τ, n.

At the technical level, the proof of Theorem 3.4, similar to the proof of Theorem 3.2, relies on the decomposition of the SGD solutions into three parts. However, instead of using a bound on the moment generating function of $Y_{i,n}(\tau)$, the proof is based on a finite-sample bound on the L^2 error derived by Costa and Gadat (2021). Consequently, one can only obtain asymptotically algebraic bounds. In comparison, our approach based on the moment generating function can lead to exponential bounds.

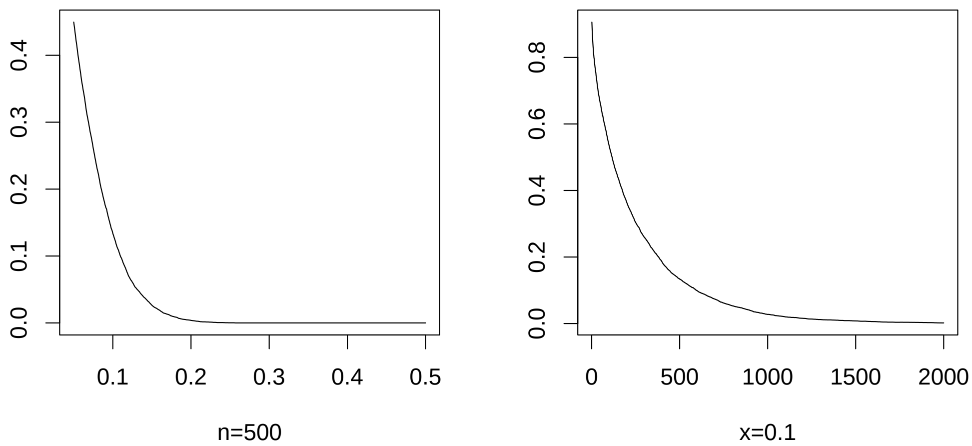

Figure 2: Tail probability as a function of x for $n = 500$ (left side), and as a function of n for $x = 0.1$ (right side), averaged over 10000 Monte Carlo iterations. The data is drawn i.i.d. from a t-distribution with 10 degrees of freedom, $\tau = 0.5$, while the SGD algorithm is initialized with a random draw from a uniform distribution on $[-1, 1]$ and $\beta = 0.7$.

4. Application to Best Arm Identification

4.1 Stochastic Quantile Bandits

As an application of our novel tail probability bound, we consider the problem of best arm identification in a multi-armed bandit setting. A p-armed stochastic bandit is a collection of probability distributions, $\nu = (F_i : i \in [p])$. For each round $t = 1, \ldots, n$, the agent chooses an action $A_t \in [p]$ and observes the reward X_{tA_t} drawn independently from the distribution of the chosen arm, F_{A_t}. The rewards of the other arms at time t are not observed. We refer to the monograph of Lattimore and Szepesvári (2020) for a comprehensive overview on bandit algorithms. We consider the setting of pure exploration (Bubeck et al., 2009). The learner is endowed with a fixed budget n and has to commit to one arm after the exploration phase in period $n+1$, according to a policy π. The goal is to select with action A_{n+1} the unique arm with the highest τ-quantile,

$$i^* = \mathrm{argmax}_{i \in [p]} Q_i(\tau).$$

In the following, we write $Q_{i^*}(\tau) = Q^*(\tau)$. Most of the literature on best arm identification is concerned with the selection of arms with the highest expected value. Focusing instead on quantiles has at least two advantages. First, quantiles are more robust location parameters than the expected value. Many existing results on best arm identification in the mean case rely on the restrictive assumption of sub-Gaussian distributions. And second, the agent might be interested in different regions of the distribution of rewards, depending on her risk attitudes and preferences. Quantile preferences were first studied by Manski (1988) and formalized as a choice-theoretic model in Rostek (2010). A recent extension of quantile preferences to the dynamic setting was proposed by de Castro and Galvao (2019). We define the suboptimality gap of a given arm i relative to the optimal arm, for a fixed quantile level τ,

$$\Delta_i^\tau = Q^*(\tau) - Q_i(\tau).$$

The agent's goal is to minimize the regret R_n^τ, which is defined as the expected difference in quantiles of her policy in comparison to playing the optimal arm i^*,

$$R_n^\tau(\pi, \nu) = \mathbb{E}\left\{Q^*(\tau) - Q_{A_{n+1}}(\tau)\right\} = \mathbb{E}\left(\Delta_{A_{n+1}}^\tau\right). \tag{14}$$

The expectation is taken with respect to the interaction of the bandit environment ν and the policy of the learner π. In addition, we define the probability of selecting a suboptimal arm after the exploration period,

$$e_n = \mathbb{P}\left(A_{n+1} \neq i^*\right). \tag{15}$$

Let $\Delta_{max}^\tau = \max_{i\in[p]} \Delta_i^\tau$ be the maximal suboptimality gap. Then we can obtain the following bound for the regret,

$$R_n^\tau(\pi, \nu) \leq \Delta_{max}^\tau e_n.$$

Our goal is to find a policy π which minimizes e_n and thus minimizes the regret R_n^τ. Other accounts on best arm identification in a quantile bandit settings were discussed in Nikolakakis et al. (2021), Szorenyi et al. (2015) and in Howard and Ramdas (2022). Best arm identification with a fixed budget is discussed in Zhang and Ong (2021). However, none of the above mentioned papers consider the recursive estimation of quantiles.

4.2 Algorithms and Instance-Dependent Regret Bounds

A naive policy would be to play each arm uniformly during the exploration phase and then commit to the arm with the largest estimated quantile. The pseudo code for this uniform exploration algorithm is provided in Algorithm 1.

Algorithm 1: Uniform Exploration Algorithm for Quantiles

```
for t = 1 to n do
 |  Choose A_t = 1 + (t mod (p))
end
Choose A_{n+1} = argmax_{i∈[p]} Ȳ_{i,⌊n/p⌋}(τ)
```

Although there is no trade-off between exploration and exploitation in our best arm identification setting, this strategy might be improved by allocating more actions to promising arms. The rationale is to avoid playing arms with a low estimated quantile by dividing the exploration period into different rounds in which one or several arms are eliminated and not played anymore in future rounds. For this purpose we adapt both the successive elimination algorithm of Audibert et al. (2010) and the sequential halving algorithm of Karnin et al. (2013) to the quantile case.

The successive elimination algorithm divides the exploration period into $p-1$ phases. After each phase, the arm with the lowest estimated quantile is eliminated. Within the phase, the arms are played uniformly. The recommended arm A_{n+1} is the single remaining arm after $p-1$ rounds. Algorithm 2 provides the pseudo code for the successive reject algorithm.

Algorithm 2: Successive Reject Algorithm for Quantiles

The budget n is allocated to the rounds r and arms i. For $t = 1, \dots, n$, only the reward of the arm which is played, X_{tA_t}, is observed and used in the SGD update. Let $S_1 = \{1, \dots, p\}$, $\overline{\log}(p) = \frac{1}{2} + \sum_{i=2}^{p} \frac{1}{i}$, and

$$n_r = \lceil \frac{1}{\overline{\log}(p)} \frac{n-p}{p+1-r} \rceil, \quad r = 1, \dots, p-1$$

for $r = 1$ **to** $p-1$ **do**

  For each $i \in S_r$, select i for $n_r - n_{r-1}$ rounds

  Set $S_{r+1} = S_r \backslash \text{argmin}_{i \in S_r} \bar{Y}_{i,n_r}(\tau)$

end

Choose $A_{n+1} = \text{argmax}_{i \in S_{p-1}} \bar{Y}_{i,n_{p-1}}(\tau)$

Note that the arm eliminated in the first round is played $n_1 = \lceil \frac{1}{\overline{\log}(p)} \frac{n-p}{p} \rceil$ times, the one eliminated in the second round is played $n_2 = \lceil \frac{1}{\overline{\log}(p)} \frac{n-p}{p-1} \rceil$ times and the two remaining arms in round $p-1$ are played $n_{p-1} = \lceil \frac{1}{\overline{\log}(p)} \frac{n-p}{2} \rceil$ times. It can be easily verified that the budget constraint is satisfied. In the following, we denote with $(i) \in \{1, \dots, p\}$ the i-th best arm, which implies $\Delta^{\tau}_{(1)} \leq \Delta^{\tau}_{(2)} \leq \dots \leq \Delta^{\tau}_{(p)}$. Further, we write the tail probability bound from Theorem 3.2 as a function of the sample size and the suboptimality gap,

$$B(n, \Delta) = C \left(n^{1+\beta} \exp\big\{ -cn^{(1-\beta^2)} \Delta \big\} + n^{1+\beta} \exp\Big\{ -c' n^{\beta(1-\beta)} \Delta^{1/2} \Big\} + \exp\big\{ -c'' n \Delta^2 \big\} \right),$$

where C, c, c' and c'' are positive constants independent of τ and n.

We bound the regret using the policy of the successive reject algorithm in the following theorem.

Theorem 4.1 *Let ν denote a p-armed stochastic bandit with F_i satisfying Assumption 3.1 and let π be the policy of the successive reject algorithm. Then,*

$$R^{\tau}_n(\nu, \pi) \leq p(p-1) \Delta^{\tau}_{max} 2 \max_{r \in \{1, \dots, p-1\}} B\left(\lceil \frac{1}{\overline{\log}(p)} \frac{n-p}{p+1-r} \rceil, \Delta^{\tau}_{(p+1-r)}/2 \right)$$

Proof Recall that $R_n^\tau = \mathbb{E}\{Q^*(\tau) - Q_{A_{t+1}}(\tau)\} \leq \Delta_{max}^\tau e_n$. In the following, we bound e_n.

$$\begin{aligned}
e_n &\leq \sum_{r=1}^{p-1} \sum_{i=p+1-r}^{p} \mathbb{P}\left\{\bar{Y}_{i^*,n_r}(\tau) \leq \bar{Y}_{(i),n_r}(\tau)\right\} \\
&= \sum_{r=1}^{p-1} \sum_{i=p+1-r}^{p} \mathbb{P}\left\{\bar{Y}_{(i),n_r}(\tau) - Q_{(i)}(\tau) + Q^*(\tau) - \bar{Y}_{i^*,n_r}(\tau) \geq \Delta_{(i)}^\tau\right\} \\
&\leq \sum_{r=1}^{p-1} \sum_{i=p+1-r}^{p} 2B(n_r, \Delta_{(i)}^\tau/2) \\
&\leq \sum_{r=1}^{p-1} r2B(n_r, \Delta_{(p+1-r)}^\tau/2) \\
&\leq p(p-1)2 \max_{r\in\{1,\dots,p-1\}} B(n_r, \Delta_{(p+1-r)}^\tau/2).
\end{aligned}$$

Plugging in the definition of n_r gives the desired result. ■

An alternative to the successive reject algorithm is the sequential halving algorithm proposed by Karnin et al. (2013). Instead of eliminating only one arm per phase, the algorithm dismisses half of the arms with the lowest estimated τ-quantile. The number of rounds is reduced to $\lfloor \log_2 p \rfloor - 1$.

Algorithm 3: Sequential Halving Algorithm for Quantiles

The budget n is allocated to the rounds r and arms i. For $t = 1, \dots, n$, only the reward of the arm which is played, X_{tA_t}, is observed and used in the SGD update.
Let $S_1 = \{1, \dots, p\}$,
for $r = 1$ **to** $\lceil \log_2 p \rceil$ **do**
  For each $i \in S_r$, select i for $n_r = \lfloor \frac{n}{|S_r|\lfloor \log_2 p \rceil} \rfloor$ times
  Let S_{r+1} denote the set of $\lfloor |S_r|/2 \rfloor$ arms in S_r with the highest value of $\bar{Y}_{i,n_r}(\tau)$.
end
Choose A_{n+1} by selecting the remaining arm in $S_{\lceil \log_2 p \rceil}$.

The following theorem bounds the regret of the policy following the sequential elimination algorithm.

Theorem 4.2 *Let ν denote a p-armed stochastic bandit with F_i satisfying Assumption 3.1 and let π be the policy of the sequential halving algorithm. Then,*

$$R_n^\tau(\nu, \pi) \leq 4\log_2 p \Delta_{max}^\tau \max_{r\in\{1,\dots,\log_2 p\}} B\left(\frac{2^r n}{p\log_2 p}, \Delta_{(p/(2^r))}^\tau/2\right).$$

Proof WLOG, assume that p is a power of 2. Note that we can bound the probability that an arbitrary arm i has an higher estimated quantile than the optimal arm in a given round r,

$$\mathbb{P}\left\{\bar{Y}_{i^*,n_r}(\tau) < \bar{Y}_{i,n_r}(\tau)\right\} \leq 2B(n_r, \Delta_i^\tau/2).$$

If the best arm i^* is eliminated in round r, at least $1/2|S_r|$ arms need to have a larger estimated quantile. Denote by N_r the number of arms with a larger estimated quantile than the optimal arm. Then we have,

$$\begin{aligned}\mathbb{E}\left(N_r\right) &= \sum_{i \in S_r} \mathbb{P}\left\{\bar{Y}_{i^*,n_r}(\tau) < \bar{Y}_{i,n_r}(\tau)\right\} \\ &\leq |S_r| 2B(n_r, \Delta^{\tau}_{(p/(2^r))}/2).\end{aligned}$$

Using Markov's inequality, we can bound the probability of eliminating the optimal arm in round r,

$$\begin{aligned}\mathbb{P}\left(i^* \notin S_{r+1}\right) &= \mathbb{P}\left(N_r > \frac{1}{2}|S_r|\right) \\ &\leq \frac{2\mathbb{E}\left(N_r\right)}{|S_r|} \\ &\leq 4B\left(n_r, \Delta^{\tau}_{(p/(2^r))}/2\right).\end{aligned}$$

Then we can use the union bound to bound e_n,

$$\begin{aligned}e_n &\leq 4 \sum_{r=1}^{\log_2 p} B(n_r, \Delta^{\tau}_{(p/(2^r))}/2) \\ &\leq 4\log_2 p \max_{r \in \{1,\ldots,\log_2 p\}} B\left(\frac{2^r n}{p\log_2 p}, \Delta^{\tau}_{(p/(2^r))}/2\right)\end{aligned}$$

■

We want to comment on the regret bounds. For a given bandit instance, ν, the probability of selecting a suboptimal arm (15) is decreasing exponentially fast in the budget n for both the successive reject and the sequential halving algorithm. This due to our tail probability result for the averaged SGD algorithm in Theorem 3.2 and since the number of times a given arm is played increases linearly with n for both algorithms. As a consequence, the simple regret in (14) is also decreasing exponentially fast in n. While most existing results on the mean bandit case rely on the assumption of sub-Gaussian distributions, we only need to impose a smoothness on the distribution. The number of arms, p, impacts the regret bounds in two ways. On the one hand, it increases the number of rounds, which is $O(p^2)$ for the successive reject algorithm and $O(\log_2 p)$ for the sequential halving algorithm. On the other hand, it decreases the fraction of the budget allocated to a certain arm in a given round.

Setting	n	UE	SR	SH
Setting 1	1000	0.792	0.190	0.377
	2000	0.562	0.058	0.213
	4000	0.196	0.005	0.110
	8000	0.133	0.000	0.040
Setting 2	1000	0.617	0.134	0.196
	2000	0.381	0.047	0.099
	4000	0.153	0.002	0.016
	8000	0.024	0.000	0.000
Setting 3	1000	0.411	0.162	0.220
	2000	0.355	0.101	0.176
	4000	0.251	0.024	0.088
	8000	0.076	0.000	0.028

Table 1: This table shows the regret (14) for the uniform exploration algorithm (UE), the successive reject algorithm (SR) and the sequential halving algorithm (SH) for different settings and choices for the budget n, averaged over 1000 Monte-Carlo iterations.

We analyze the performance of the algorithms in a short simulation study. We set $p = 16$ and $\tau = 0.5$, the data is drawn i.i.d. from a t-distribution with 10 degrees of freedom. For the SGD estimation, we set $\beta = 0.7$ and $c_\gamma = 1$, and the algorithm is initialized with a random draw from a uniform distribution on $[-1, 1]$. We consider three settings.

- Setting 1: One type of suboptimal arms, $Q_1(\tau) = 5$ and $Q_i(\tau) = 4$, $i = 2, \ldots, 16$.

- Setting 2: Arithmetic progression of suboptimal arms, $Q_1(\tau) = 5$ and $Q_i(\tau) = 5 - 0.25i$, $i = 2, \ldots, 16$.

- Setting 3: Two types of suboptimal arms, $Q_1(\tau) = 5$, $Q_i(\tau) = 4.5$, $i = 2, \ldots, 8$, and $Q_i(\tau) = 3$, $i = 9, \ldots, 16$.

Table 1 shows the performance of the three algorithms in terms of the regret for different choices for the budget n, based on 1000 Monte-Carlo iterations. In all settings, the successive reject algorithm (SR) performs best, followed by the sequential halving algorithm (SH) and the uniform exploration algorithm (UE) performs worst. It becomes clear that the regret can be effectively reduced with a growing budget, n. It should be noted that the dominance of the successive reject algorithm cannot be solely explained by Theorems 4.1 and 4.2. However, a plausible explanation arises by looking at the budget allocated to each arm in the first round of the sequential halving algorithm, which is $n_1 = \lfloor n/64 \rfloor$ in our simulation setting. The decision to reject a given arm thus hinges on a small sample size. The main advantage of the successive reject algorithm is that the optimal arm only needs to avoid being the arm with the lowest estimated quantile, while the sequential halving algorithm requires the optimal arm to be in the upper half of arms in order to survive the round.

5. Conclusion

This paper studies the non-asymptotic performance of the SGD algorithm with Polyak-Ruppert averaging for the recursive estimation of quantiles. The algorithm is a computationally and memory efficient alternative to traditional estimators based on order statistics. We derive an exponentially fast decreasing tail probability bound while only imposing assumption on the smoothness of the distribution. Instead of only bounding a few moments of the distribution, our proof relies on a bound on the moment generating function of the SGD solution.

6. Proofs for Results in Section 3

6.1 Proof of Theorem 3.1

Since our goal is to establish an exponential tail probability for $Y_{i,k}(\tau)$, we will need to work with its moment generating function. Woodroofe (1972) used characteristic and moment generating functions to derive normal approximations and large deviations of SGD solutions. Here we shall carry out a meticulous analysis and show that

$$\mathbb{E}(e^{t(Y_{i,k+1}(\tau)-Q_i(\tau))}) \le a_{k,t}\mathbb{E}(e^{t(Y_{i,k}(\tau)-Q_i(\tau))}) + c_0, \text{ where } a_{k,t} = 1 - c_1\gamma_k + c_2 t^2\gamma_k^2,$$

by considering two cases $X_{i,k+1} > Y_{i,k}(\tau)$ and $X_{i,k+1} \le Y_{i,k}(\tau)$ separately. Here c_0, c_1, c_2 are positive constants. Recursively applying above inequality, we have

$$\mathbb{E}(e^{t(Y_{i,n}(\tau)-Q_i(\tau))}) \le c_0\Big(1 + \sum_{k=k_t+1}^{n} \phi_k\Big) + \phi_{k_t}\mathbb{E}(e^{t(Y_{i,k_t}(\tau)-Q_i(\tau))}), \text{ where } \phi_k = \prod_{l=k}^{n} a_{l,t}$$

and $k_t > 0$ is a selected starting point. We choose $k_t = \lceil (c_\gamma t)^{1/\beta} \rceil$, so that $\sum_{k=k_t+1}^{n} \phi_k$ and the starting term $\phi_{k_t}\mathbb{E}(e^{t(Y_{i,k_t}(\tau)-Q_i(\tau))})$ are both of order n^β when $t \le cn^{\beta(1-\beta)}$.

Proof Recall that F_i is the distribution function of X_i and $t > 0$. We shall firstly bound the term $\mathbb{E}\{e^{t(Y_{i,k+1}(\tau)-Q_i(\tau))}\}$, the other one $\mathbb{E}\{e^{-t(Y_{i,k+1}(\tau)-Q_i(\tau))}\}$ will be handled at the end of the proof. To start with, by the iteration mechanism of SGD in (3) we have

$$\begin{aligned}
&\mathbb{E}\Big\{e^{t(Y_{i,k+1}(\tau)-Q_i(\tau))}\Big\} \\
=&\mathbb{E}\Big\{\mathbb{E}\Big(e^{t(Y_{i,k}(\tau)-Q_i(\tau)+\gamma_k\tau)}\mathbf{1}_{X_{i,k+1}>Y_{i,k}(\tau)} + e^{t(Y_{i,k}(\tau)-Q_i(\tau)-\gamma_k(1-\tau))}\mathbf{1}_{X_{i,k+1}\le Y_{i,k}(\tau)}\Big|Y_{i,k}(\tau)\Big)\Big\} \\
=&\mathbb{E}\Big\{e^{t(Y_{i,k}(\tau)-Q_i(\tau))}\Big[e^{t\gamma_k\tau}\big(1 - F_i(Y_{i,k}(\tau))\big) + e^{-t\gamma_k(1-\tau)}F_i(Y_{i,k}(\tau))\Big]\Big\} \\
=&\mathbb{E}\Big\{e^{t(Y_{i,k}(\tau)-Q_i(\tau))}\Big[e^{t\gamma_k\tau} - (e^{t\gamma_k\tau} - e^{-t\gamma_k(1-\tau)})F_i(Y_{i,k}(\tau))\Big]\Big\},
\end{aligned}$$

where the second equation is due to the independence between $X_{i,k+1}$ and $Y_{i,k}(\tau)$. Since $e^{t\gamma_k\tau} \ge 1 \ge e^{-t\gamma_k(1-\tau)}$, the term $(e^{t\gamma_k\tau} - e^{-t\gamma_k(1-\tau)})F_i(x)$ monotonically increases in x. Take

$L = c_\tau/(2c_f t)$, for any $k \geq (c_\gamma t)^{1/\beta}$ we have that $t\gamma_k \leq 1$ and

$$\begin{aligned}
&\mathbb{E}\big\{e^{t(Y_{i,k+1}(\tau)-Q_i(\tau))}\big\} \\
\leq&\mathbb{E}\Big\{e^{t(Y_{i,k}(\tau)-Q_i(\tau))}\big[e^{t\gamma_k\tau} - (e^{t\gamma_k\tau} - e^{-t\gamma_k(1-\tau)})F_i(L+Q_i(\tau))\big]\mathbf{1}_{Y_{i,k}(\tau)-Q_i(\tau)\geq L} \\
&\quad + e^{tL+t\gamma_k\tau}\mathbf{1}_{Y_{i,k}(\tau)-Q_i(\tau)<L}\Big\} \\
\leq&\mathbb{E}\Big\{e^{t(Y_{i,k}(\tau)-Q_i(\tau))}\big[e^{t\gamma_k\tau}(1-F_i(L+Q_i(\tau))) + e^{-t\gamma_k(1-\tau)}F_i(L+Q_i(\tau))\big]\Big\} + c_0,
\end{aligned} \tag{16}$$

where $c_0 = e^{c_\tau/(2c_f)+\tau}$. Since $t\gamma_k\tau \leq \tau < 1$, by Taylor's expansion, $e^{t\gamma_k\tau} \leq 1+t\gamma_k\tau+t^2\gamma_k^2\tau^2$. Then we have

$$\begin{aligned}
&e^{t\gamma_k\tau}(1-F_i(L+Q_i(\tau))) + e^{-t\gamma_k(1-\tau)}F_i(L+Q_i(\tau)) \\
\leq&(1+t\gamma_k\tau+t^2\gamma_k^2)(1-F_i(L+Q_i(\tau))) + (1-t\gamma_k(1-\tau)+t^2\gamma_k^2)F_i(L+Q_i(\tau)) \\
\leq&1-t\gamma_k\big[F_i(L+Q_i(\tau))-\tau\big] + t^2\gamma_k^2 \\
\leq&1-t\gamma_k\big(f_i(Q_i(\tau))-c_f L\big)L + t^2\gamma_k^2,
\end{aligned} \tag{17}$$

where the last inequality is due to the fact that $\tau = F_i(Q_i(\tau))$ and

$$F_i(L+Q_i(\tau)) - F_i(Q_i(\tau)) - f_i(Q_i(\tau))L \geq -|f_i|_\infty L^2.$$

Recall that $L = c_\tau/(2c_f t)$. Inserting above into (16), we have

$$\mathbb{E}(e^{t(Y_{i,k+1}(\tau)-Q_i(\tau))}) \leq \mathbb{E}(e^{t(Y_{i,k}(\tau)-Q_i(\tau))})(1-c_1\gamma_k+t^2\gamma_k^2) + c_0,$$

where $c_1 = c_\tau^2/(4c_f)$. Recursively applying above inequality, for $k_t = \lceil (c_\gamma t)^{1/\beta}\rceil$, we have

$$\mathbb{E}(e^{t(Y_{i,n+1}(\tau)-Q_i(\tau))}) \leq c_0\Big(1+\sum_{k=k_t+1}^{n}\phi_k\Big) + \phi_{k_t}\mathbb{E}(e^{t(Y_{i,k_t}(\tau)-Q_i(\tau))}) \tag{18}$$

where

$$\phi_k = \prod_{l=k}^{n}(1-c_1\gamma_l+t^2\gamma_l^2). \tag{19}$$

In the following, we shall bound the terms $\sum_{k=k_t+1}^{n}\phi_k$ and $\phi_{k_t}\mathbb{E}(e^{t(Y_{i,k_t}(\tau)-Q_i(\tau))})$ separately. Firstly, for the $\sum_{k=k_t+1}^{n}\phi_k$ part, note that $1+x \leq e^x$, hence

$$\begin{aligned}
\phi_k &\leq \exp\Big\{-\sum_{l=k}^{n}(c_1\gamma_l - t^2\gamma_l^2)\Big\} \\
&\leq \exp\Big\{-c_1c_\gamma(n^{1-\beta}-k^{1-\beta})/(1-\beta) + t^2c_\gamma^2(k^{-2\beta+1}-n^{-2\beta+1})/(2\beta-1)\Big\}.
\end{aligned}$$

To calculate $\sum_{k=k_t}^{n}\phi_k$, we shall deal with k close to n and far away from n separately. Denote $c_2 = c_1c_\gamma(1-(1/2)^{1-\beta})/(1-\beta)$. If $k_t \leq k \leq n/2$ and $t \leq cn^{\beta(1-\beta)}$, where

$$c = \min\big\{((2\beta-1)c_2/2)^\beta/c_\gamma, (c_1/c_\gamma)^{1/2}2^{-\beta-1/2}, ((1-\beta)c_2/4)^\beta/c_\gamma, c_2/(4(M+c_y))\big\}, \tag{20}$$

we have

$$\begin{aligned}\phi_k &\le \exp\big\{-c_2 n^{1-\beta} + c_\gamma^2 t^2 k_t^{-2\beta+1}/(2\beta-1)\big\}\\ &\le \exp\big\{-c_2 n^{1-\beta} + c_\gamma^{1/\beta} t^{1/\beta}/(2\beta-1)\big\}\\ &\le \exp\big\{-c_2 n^{1-\beta}/2\big\}.\end{aligned}$$

When $k \ge n/2$, by the mean value theorem,

$$\begin{aligned}\phi_k &\le \exp\Big\{-c_1 c_\gamma n^{-\beta}(n-k) + c_\gamma^2 t^2 (n/2)^{-2\beta}(n-k)\Big\}\\ &\le \exp\Big\{-c_\gamma n^{-\beta}(n-k)(c_1 - c_\gamma c^2 n^{-(2\beta-1)\beta} 2^{2\beta})\Big\}\\ &\le \exp\big\{-c_1 c_\gamma n^{-\beta}(n-k)/2\big\}.\end{aligned}$$

Combining the above two cases, we obtain that

$$\begin{aligned}\sum_{k=k_t+1}^{n} \phi_k &= \sum_{k=k_t+1}^{n/2} \phi_k + \sum_{k=n/2+1}^{n} \phi_k\\ &\le (n/2)\exp\big\{-c_2 n^{1-\beta}/2\big\} + \int_0^{n/2} \exp\big\{-c_1 c_\gamma n^{-\beta} x/2\big\} dx\\ &\le (n/2)\exp\big\{-c_2 n^{1-\beta}/2\big\} + (2/c_1 c_\gamma) n^\beta \big(1 - \exp\big\{-c_1 c_\gamma n^{1-\beta}/4\big\}\big)\\ &\le c_3 n^\beta,\end{aligned}$$

where $c_3 = 1/c_2 + 2/(c_1 c_\gamma)$.

Secondly, for the $\phi_{k_t}\mathbb{E}(e^{t(Y_{i,k_t}(\tau)-Q_i(\tau))})$ part, note that

$$|Y_{i,k}(\tau) - Q_i(\tau)| \le |y_i - Q_i(\tau)| + \sum_{j=1}^{k} |\gamma_j| \le |y_i - Q_i(\tau)| + (1-\beta)^{-1} c_\gamma k^{1-\beta}.$$

Hence

$$\phi_{k_t}\mathbb{E}(e^{t(Y_{i,k_t}(\tau))-Q_i(\tau))}) \le \exp\big\{-c_2 n^{1-\beta}/2 + (1-\beta)^{-1} c_\gamma t k_t^{1-\beta} + t|y_i - Q_i(\tau)|\big\} \le 1.$$

Therefore by (18), we have

$$\mathbb{E}(e^{t(Y_{i,n}(\tau)-Q_i(\tau))}) \le c_0\Big(1 + c_3 n^\beta\Big) + 1 \le c' n^\beta,$$

where

$$c' = c_0 c_3 + (c_0+1)/(c_0 c_3). \tag{21}$$

On the other hand, by the argument in (16) with truncated value replaced by $Q_i(\tau) - L$, we have

$$\begin{aligned}&\mathbb{E}\big\{e^{-t(Y_{i,k+1}(\tau)-Q_i(\tau))}\big\}\\ \le&\mathbb{E}\Big\{e^{-t(Y_{i,k}(\tau)-Q_i(\tau))}\big[e^{-t\gamma_k\tau}(1 - F_i(Q_i(\tau)-L)) + e^{t\gamma_k(1-\tau)}F_i(Q_i(\tau)-L)\big]\Big\} + c_0.\end{aligned}$$

Similar to (17), by Taylor's expansion we have

$$\begin{aligned}&e^{-t\gamma_k\tau}(1-F_i(Q_i(\tau)-L))+e^{t\gamma_k(1-\tau)}F_i(Q_i(\tau)-L)\\ \leq&1-t\gamma_k\big(f_i(Q_i(\tau))-c_fL)L+t^2\gamma_k^2.\end{aligned}$$

Then by the same argument as for $\mathbb{E}\big\{e^{t(Y_{i,k+1}(\tau)-Q_i(\tau))}\big\}$ case, we have

$$\mathbb{E}\big\{e^{-t(Y_{i,k+1}(\tau)-Q_i(\tau))}\big\}\leq c'n^{\beta},$$

for $0<t\leq cn^{\beta(1-\beta)}$. Combining the two cases we complete the proof of (5). To obtain the tail probability (6), by Markov's inequality with $t=cn^{\beta(1-\beta)}$, we have

$$\mathbb{P}(|Y_{i,n}(\tau)-Q_i(\tau)|\geq x)\leq e^{-tx}\mathbb{E}\Big[\exp\big\{t|Y_{i,n}(\tau)-Q_i(\tau)|\big\}\Big]=c'n^{\beta}e^{-cn^{\beta(1-\beta)}x}.$$

■

6.2 Proof of Theorem 3.2

Proof Let $G_{i,\tau}(x)=\tau-F_i(x)$. Recall the definition of $Z_{i,k+1}(\tau)$ in (11). Notice that $G_{i,\tau}(Y_{i,k}(\tau))=\mathbb{E}(Z_{i,k+1}(\tau)|\mathcal{F}_{i,k})$ and $G_{i,\tau}(Q_i(\tau))=0$. Hence $Z_{i,k+1}(\tau)$ can be written as

$$Z_{i,k+1}(\tau)=\xi_{i,k+1}+G'_{i,\tau}(Q_i(\tau))(Y_{i,k}(\tau)-Q_i(\tau))+\rho_{i,k}, \tag{22}$$

where $\xi_{i,k+1}=Z_{i,k+1}(\tau)-G_{i,\tau}(Y_{i,k}(\tau))$ is the martingale difference part with respect to the filtration $\mathcal{F}_{i,k}$ and

$$\rho_{i,k}=G_{i,\tau}(Y_{i,k}(\tau))-G'_{i,\tau}(Q_i(\tau))(Y_{i,k}(\tau)-Q_i(\tau)).$$

Then by (22), the SGD equation (10) can be written as

$$Y_{i,k+1}(\tau)=Y_{i,k}(\tau)+\gamma_k\big[\xi_{i,k+1}+G'_{i,\tau}(Q_i(\tau))(Y_{i,k}(\tau)-Q_i(\tau))+\rho_{i,k}\big]. \tag{23}$$

Averaging (23) for k from 1 to n leads to

$$n^{-1}\sum_{k=1}^{n}\gamma_k^{-1}(Y_{i,k+1}(\tau)-Y_{i,k}(\tau))=\bar{\xi}_{i,n}+G'_{i,\tau}(Q_i(\tau))(\bar{Y}_{i,n}(\tau)-Q_i(\tau))+\bar{\rho}_{i,n}.$$

By the summation by parts formula on the left hand side of the above identity, we obtain

$$\begin{aligned}&G'_{i,\tau}(Q_i(\tau))(\bar{Y}_{i,n}-Q_i(\tau))\\ =&n^{-1}\gamma_n^{-1}(Y_{i,n+1}(\tau)-Y_{i,1}(\tau))-n^{-1}\sum_{k=1}^{n-1}(Y_{i,k+1}(\tau)-Y_{i,1}(\tau))(\gamma_{k+1}^{-1}-\gamma_k^{-1})-\bar{\xi}_{i,n}-\bar{\rho}_{i,n}\\ =&\Big[n^{-1}\gamma_n^{-1}(Y_{i,n+1}(\tau)-Q_i(\tau))-n^{-1}\gamma_1^{-1}(Y_{i,1}(\tau)-Q_i(\tau))\Big]\\ &-n^{-1}\sum_{k=1}^{n-1}(Y_{i,k+1}(\tau)-Q_i(\tau))(\gamma_{k+1}^{-1}-\gamma_k^{-1})-\bar{\xi}_{i,n}-\bar{\rho}_{i,n}\\ =&(\mathrm{I}_{11}-\mathrm{I}_{12})-\mathrm{I}_2-\mathrm{I}_3-\mathrm{I}_4.\end{aligned} \tag{24}$$

For I_{11}, by Theorem 3.1, we have

$$\mathbb{P}(|\mathrm{I}_{11}| > x) \le c' n^{\beta} \exp\big\{ - c_{\gamma} c n^{(1+\beta)(1-\beta)} x \big\}.$$

For I_{12}, by the boundedness of $Y_{i,1}(\tau)$ and $Q_i(\tau)$ we have

$$|\mathrm{I}_{12}| \le (c_y + M)/(n c_{\gamma}).$$

For I_2, take $a_k = \beta^2 k^{-(1+\beta)(1-\beta)} n^{-\beta^2}$, such that $\sum_{k=1}^{n-1} a_k \le 1$. By Theorem 3.1 we have

$$\begin{aligned}
&\mathbb{P}\Big(\big|(Y_{i,k+1}(\tau) - Q_i(\tau))(\gamma_k^{-1} - \gamma_{k+1}^{-1})\big| > n a_k x \Big) \\
\le& c' k^{\beta} \exp\Big\{ - c n a_k k^{\beta(1-\beta)} x / |\gamma_k^{-1} - \gamma_{k+1}^{-1}| \Big\} \\
\le& c' n^{\beta} \exp\big\{ - c c_{\gamma} \beta^{-1} n a_k k^{(1+\beta)(1-\beta)} x \big\} \\
\le& c' n^{\beta} \exp\big\{ - c_1 n^{1-\beta^2} x \big\},
\end{aligned}$$

where the second inequality is due to $\gamma_{k+1}^{-1} - \gamma_k^{-1} \le c_{\gamma}^{-1} \beta k^{-(1-\beta)}$, $c_1 = c c_{\gamma} \beta$ and c, c' are constants in Theorem 3.1. Therefore, by above inequality we have

$$\begin{aligned}
\mathbb{P}(|\mathrm{I}_2| > x) &\le \sum_{k=1}^{n-1} \mathbb{P}\Big(|Y_{i,k+1}(\tau)(\gamma_k^{-1} - \gamma_{k+1}^{-1})| > n a_k x \Big) \\
&\le \sum_{k=1}^{n-1} c' n^{\beta} \exp\Big\{ - c_1 n^{1-\beta^2} x \Big\} \\
&\le c' n^{1+\beta} \exp\big\{ - c_1 n^{1-\beta^2} x \big\}.
\end{aligned}$$

To deal with I_3, since $(\xi_{i,k})_k$ are martingale differences with respect to filtration $\mathcal{F}_{i,k}$, and $|\xi_{i,k}| \le 1$, thus by Azuma's concentration inequality, we have

$$\mathbb{P}(|\mathrm{I}_3| > x) \le 2\exp\big\{ - n x^2/2 \big\}.$$

For I_4, by Assumption 3.1, $|\rho_{i,k}| \le c_f (Y_{i,k}(\tau) - Q_i(\tau))^2$, take $b_k = c_b k^{-2\beta(1-\beta)} n^{-1+2\beta(1-\beta)}$, where $c_b = (1 - 2\beta(1-\beta))^{-1}$, such that $\sum_{k=1}^{n} b_k \le 1$. Then by Theorem 3.1,

$$\begin{aligned}
\mathbb{P}(|\mathrm{I}_4| > x) &\le \sum_{k=1}^{n} \mathbb{P}\big(c_f (Y_{i,k}(\tau) - Q_i(\tau))^2 > n b_k x \big) \\
&\le c' \sum_{k=1}^{n} k^{\beta} \exp\big\{ - c k^{\beta(1-\beta)} (n b_k x / c_f)^{1/2} \big\} \\
&\le c' n^{1+\beta} \exp\big\{ - c_2 n^{\beta(1-\beta)} x^{1/2} \big\},
\end{aligned}$$

where $c_2 = c(c_b/c_f)^{1/2}$. Combining I_1-I_4, we obtain

$$\begin{aligned}
\mathbb{P}(|\bar{Y}_{i,n} - Q_i(\tau)| > x) \le\, & c' n^{\beta} \exp\big\{ - c_{\gamma} c n^{(1+\beta)(1-\beta)} (x/4 - (c_y + M)/(n c_{\gamma})) \big\} \\
& + c' n^{1+\beta} \exp\big\{ - c_1 n^{1-\beta^2} x/4 \big\} + 2\exp\big\{ - n (x/4)^2/2 \big\} \\
& + c' n^{1+\beta} \exp\big\{ - c_2 n^{\beta(1-\beta)} (x/4)^{1/2} \big\}, \qquad (25)
\end{aligned}$$

which completes the proof of (8). Let $x = C\log^2(np)n^{-2\beta(1-\beta)}$, where $C > 0$ is a sufficiently large constant, by (8) and the union bound, we have

$$\begin{aligned}
&\mathbb{P}\big(\max_{1\le i\le p} |\bar{Y}_{i,n}(\tau) - Q_i(\tau)| > x\big) \\
\lesssim& pn^{1+\beta}\exp\big\{-cn^{(1-\beta^2)}x\big\} + pn^{1+\beta}\exp\Big\{-c'n^{\beta(1-\beta)}x^{1/2}\Big\} + p\exp\big\{-c''nx^2\big\} \\
\lesssim& pn^{1+\beta}\exp\big\{-cC\log^2(np)n^{(1-\beta)^2}\big\} + pn^{1+\beta}\exp\Big\{-c'C^{1/2}\log(np)\Big\} \\
&+ p\exp\big\{-c''C^2n^{1-4\beta(1-\beta)}\log^4(np)\big\},
\end{aligned} \tag{26}$$

then (9) follows. ■

6.3 Proof of Theorem 3.4

Proof Recall the following decomposition,

$$\begin{aligned}
G'_{i,\tau}(Q_i(\tau))\big(\bar{Y}_{i,n}(\tau) - Q_i(\tau)\big) =& \big[n^{-1}\gamma_n^{-1}\left(Y_{i,n+1}(\tau) - Q_i(\tau)\right) - n^{-1}\gamma_1^{-1}\left(Y_{i,1}(\tau) - Q_i(\tau)\right)\big] \\
&- n^{-1}\sum_{k=1}^{n-1}\left(Y_{i,k+1}(\tau) - Q_i(\tau)\right)\left(\gamma_{k+1}^{-1} - \gamma_k^{-1}\right) - \bar{\xi}_{i,n} - \bar{\rho}_{i,n} \\
=:& \mathrm{I}_1 - \mathrm{I}_2 - \mathrm{I}_3 - \mathrm{I}_4.
\end{aligned}$$

Following the proof of Theorem 4.2 of Cardot et al. (2017), we again bound each term on the right-hand side separately. By Condition 3.1, we can apply Theorem 2.2 of Costa and Gadat (2021). For each $n \ge 1$, there is a constant $C > 0$ such that $\mathbb{E}(|Y_{i,n+1} - Q_i(\tau)|^2) \le Cn^{-\beta}$. For the I_1 part, we have

$$\mathbb{E}\left\{\left|\frac{Y_{i,n+1}(\tau) - Q_i(\tau)}{n\gamma_n}\right|^2\right\} \le n^{2\beta-2}\mathbb{E}\left\{|Y_{i,n+1}(\tau) - Q_i(\tau)|^2\right\} \le C\frac{1}{n^{2-\beta}}.$$

Applying Cauchy-Schwarz, we have for a constant $C_1 > 0$,

$$\mathbb{E}\left\{\left|\frac{Y_{i,n+1}(\tau) - Q_i(\tau)}{n\gamma_n}\right|\right\} \le C_1\frac{1}{n^{1-\beta/2}}.$$

Further, there is a constant $C_2 > 0$ such that

$$\mathbb{E}\left\{\left|\frac{Y_{i,1}(\tau) - Q_i(\tau)}{n\gamma_1}\right|\right\} \le \frac{C_2}{n}.$$

For the I_2 part, since $\gamma_{k+1}^{-1} - \gamma_k^{-1} \le 2\beta k^{\beta-1}$, there exists a constant $C_3 > 0$ such that

$$\begin{aligned}
\mathbb{E}\left\{\left|\frac{1}{n}\sum_{k=1}^{n-1}\left(Y_{i,k+1}(\tau) - Q_i(\tau)\right)\left(\gamma_{k+1}^{-1} - \gamma_k^{-1}\right)\right|\right\} &\le \frac{2\beta}{n}\sum_{k=1}^{n-1}\mathbb{E}\left(|Y_{i,k+1}(\tau) - Q_i(\tau)|\right)k^{\beta-1} \\
&\le \frac{C_3}{n^{1-\beta/2}}.
\end{aligned}$$

For the I_4 part, by condition 3.1, we have that $|\rho_{i,n}| \leq c_f(Y_{i,n}(\tau) - Q_i(\tau))^2$. Consequently, we have for a constant $C_4 > 0$,

$$\mathbb{E}\left(\left|\frac{1}{n}\sum_{k=1}^{n}\rho_{i,k}\right|\right) \leq \frac{c_f}{n}\sum_{k=1}^{n}\mathbb{E}\left\{|Y_{i,k}(\tau) - Q_i(\tau)|^2\right\} \leq \frac{c_f C}{n}\sum_{k=1}^{n}k^{-\beta} \leq C_4\frac{1}{n^{\beta}}.$$

Finally, we bound the martingale term by using Azuma's inequality,

$$\mathbb{P}(|\mathrm{I}_3| > x) \leq 2\exp\big\{-nx^2/2\big\}.$$

By Markov's inequality, we have

$$\begin{aligned}&\mathbb{P}\left\{\left|G'_{i,\tau}(Q_i(\tau))\left(\bar{Y}_{i,n}(\tau) - Q_i(\tau)\right)\right| > x\right\}\\ &\leq 2\exp\big\{-nx^2/2\big\} + \frac{C_1 + C_3}{xn^{1-\beta/2}} + \frac{C_2}{xn} + \frac{C_4}{xn^{\beta}}.\end{aligned}$$

■

6.4 Proof of (7)

Following the notation of Cardot et al. (2017), let Z_n be the SGD estimate of the univariate median, m, updated recursively via

$$Z_{k+1} = Z_k + \gamma_k\Big(\tau\mathbf{1}_{X_{k+1}>Z_k} - (1-\tau)\mathbf{1}_{X_{k+1}\leq Z_k}\Big), \tag{27}$$

with constant initial value $Z_0 = z$ and learning rate $\gamma_k = c_\gamma k^{-\alpha}$, with $1/2 < \alpha < 1$, and some constant $c_\gamma > 0$. Define $U_{n+1} := \tau\mathbf{1}_{X_{k+1}>Z_k} - (1-\tau)\mathbf{1}_{X_{k+1}\leq Z_k}$ and consider a sequence of σ-algebras $\mathcal{F}_n := \sigma(X_1, \ldots, X_n)$. Further, let $\Phi(Z_n) := \mathbb{E}\left(U_{n+1}|\mathcal{F}_n\right)$ and $\xi_{n+1} := \Phi(Z_n) - U_{n+1}$. Let $\delta_n := \Phi(Z_n) - f(m)(Z_n - m)$, where f is the density of X_k.

Consider the following decomposition of the SGD estimate of the median in (4.2) of Cardot et al. (2017) with $d = 1$,

$$\begin{aligned}Z_n - m &= Z_{n-1} - m - \gamma_n f(m)\left(Z_{n-1} - m\right) + \gamma_n\xi_n - \gamma_n\delta_{n-1}\\ &= \beta_{n-1}\left(Z_1 - m\right) + \beta_{n-1}M_n - \beta_{n-1}R_n\\ &=: \mathrm{I}_1 + \mathrm{I}_2 + \mathrm{I}_3,\end{aligned} \tag{28}$$

where

$$R_n := \sum_{k=1}^{n-1}\gamma_k\beta_k^{-1}\delta_k,$$

$$M_n := \sum_{k=1}^{n-1}\gamma_k\beta_k^{-1}\xi_{k+1},$$

and $\beta_n = \prod_{k=1}^{n}\alpha_k$ with $\alpha_l = 1 - \gamma_k f(m)$ and $\beta_0 = 1$.

We want to show that the following upper bound in the Proof of Theorem 4.1 on Page 10 in the Appendix of Cardot et al. (2017) is also valid in the univariate case, which is not covered due to the violation of Assumption (A3),

$$\mathbb{P}\left(|Z_n - m| > t\right) \leq 2\text{exp}\left\{-\frac{-C_0 t^2 n^{\alpha}}{1 + C_3 t}\right\} + \frac{C_1 e^{-C_4 n^{1-\alpha}}}{t^2} + \frac{C_2}{n^{\alpha} t},$$

for constants $C_0, C_1, C_2, C_3 > 0$. The bounds for parts I_1 and I_2 follow from the proof of Theorem 4.1. For part I_3, we have to assume that the random variable X has a differentiable density function $f(x)$, with $f(m) > 0$ and $|f'|_{\infty} \leq c_f < \infty$. Then we have, since $|\delta_k| \leq c_f (Z_k - m)^2$,

$$\begin{aligned}\mathbb{E}\left(|\beta_{n-1} R_n|\right) &\leq \sum_{k=1}^{n-1} \left|\beta_{n-1}\beta_k^{-1}\right| \mathbb{E}\left(|\delta_k|\right) \\ &\leq \sum_{k=1}^{n-1} \left|\beta_{n-1}\beta_k^{-1}\right| c_f \mathbb{E}\left(|Z_k - m|^2\right).\end{aligned}$$

By Theorem 2.2 of Costa and Gadat (2021) there exists a constant $C > 0$ such that $\mathbb{E}\left(|Z_k - m|^2\right) \leq C k^{-\alpha}$. Finally, we have

$$\begin{aligned}\mathbb{E}\left(|\beta_{n-1} R_n|\right) &\leq C_f C c_{\gamma} \sum_{k=1}^{n-1} \frac{1}{k^{2\alpha}} \left|\beta_{n-1}\beta_k^{-1}\right| \\ &= O(n^{-\alpha}).\end{aligned}$$

Acknowledgments

Georg Keilbar gratefully acknowledges financial support from the Deutsche Forschungsgemeinschaft via the IRTG 1792 "High Dimensional Nonstationary Time Series". Wei Biao Wu's research is partially supported by NSF grants. We are grateful to the editor and referees for their helpful comments and suggestions which render a much improved version.

References

Subutai Ahmad, Alexander Lavin, Scott Purdy, and Zuha Agha. Unsupervised real-time anomaly detection for streaming data. *Neurocomputing*, 262:134–147, 2017.

Jean-Yves Audibert, Sébastien Bubeck, and Rémi Munos. Best arm identification in multi-armed bandits. In *COLT*, pages 41–53, 2010.

Raghu Raj Bahadur. A note on quantiles in large samples. *The Annals of Mathematical Statistics*, 37(3):577–580, 1966.

Olivier Bardou, Noufel Frikha, and Gilles Pages. Computing var and cvar using stochastic approximation and adaptive unconstrained importance sampling. *Monte Carlo Methods and Applications*, 15(3):173–210, 2009.

Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in multi-armed bandits problems. In *International conference on Algorithmic learning theory*, pages 23–37. Springer, 2009.

Hervé Cardot, Peggy Cénac, and Pierre-André Zitt. Efficient and fast estimation of the geometric median in Hilbert spaces with an averaged stochastic gradient algorithm. *Bernoulli*, 19(1):18 – 43, 2013. doi: 10.3150/11-BEJ390. URL `https://doi.org/10.3150/11-BEJ390`.

Hervé Cardot, Peggy Cénac, Antoine Godichon-Baggioni, et al. Online estimation of the geometric median in hilbert spaces: Nonasymptotic confidence balls. *The Annals of Statistics*, 45(2):591–614, 2017.

Probal Chaudhuri. On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association*, 91(434):862–872, 1996.

Xi Chen, Weidong Liu, and Yichen Zhang. Quantile regression under memory constraint. *The Annals of Statistics*, 47(6):3244–3273, 2019.

Xi Chen, Jason D Lee, Xin T Tong, and Yichen Zhang. Statistical inference for model parameters in stochastic gradient descent. *The Annals of Statistics*, 48(1):251–273, 2020.

Manon Costa and Sébastien Gadat. Non asymptotic controls on a recursive superquantile approximation. *Electronic Journal of Statistics*, 15(2):4718–4769, 2021.

Luciano de Castro and Antonio F Galvao. Dynamic quantile models of rational behavior. *Econometrica*, 87(6):1893–1939, 2019.

Robert F Engle and Simone Manganelli. Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4):367–381, 2004.

Yixin Fang. Scalable statistical inference for averaged implicit stochastic gradient descent. *Scandinavian Journal of Statistics*, 46(4):987–1002, 2019.

Yixin Fang, Jinfeng Xu, and Lei Yang. Online bootstrap confidence intervals for the stochastic gradient descent estimator. *The Journal of Machine Learning Research*, 19(1):3053–3073, 2018.

Sébastien Gadat and Fabien Panloup. Optimal non-asymptotic analysis of the ruppert–polyak averaging stochastic algorithm. *Stochastic Processes and their Applications*, 156: 312–348, 2023.

JBS Haldane. Note on the median of a multivariate distribution. *Biometrika*, 35(3-4): 414–417, 1948.

Steven R Howard and Aaditya Ramdas. Sequential estimation of quantiles with applications to a/b testing and best-arm identification. *Bernoulli*, 28(3):1704–1728, 2022.

Zohar Karnin, Tomer Koren, and Oren Somekh. Almost optimal exploration in multi-armed bandits. In *International Conference on Machine Learning*, pages 1238–1246. PMLR, 2013.

Jack Kiefer. On bahadur's representation of sample quantiles. *The Annals of Mathematical Statistics*, 38(5):1323–1342, 1967.

Jack Kiefer and Jacob Wolfowitz. Stochastic estimation of the maximum of a regression function. *The Annals of Mathematical Statistics*, 23(3):462–466, 1952.

Tze Leung Lai. Stochastic approximation. *The Annals of Statistics*, 31(2):391–406, 2003.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Lennart Ljung. Analysis of recursive stochastic algorithms. *IEEE transactions on automatic control*, 22(4):551–575, 1977.

Ge Luo, Lu Wang, Ke Yi, and Graham Cormode. Quantiles over data streams: experimental comparisons, new analyses, and further improvements. *The VLDB Journal*, 25(4):449–472, 2016.

Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research*, 11(1), 2010.

Charles F Manski. Ordinal utility models of decision making under uncertainty. *Theory and Decision*, 25(1):79–104, 1988.

Stanislav Minsker. Geometric median and robust estimation in banach spaces. *Bernoulli*, 21(4):2308–2335, 2015.

Ian Munro and Mike S Paterson. Selection and sorting with limited storage. *Theoretical computer science*, 12(3):315–323, 1980.

Konstantinos E Nikolakakis, Dionysios S Kalogerias, Or Sheffet, and Anand D Sarwate. Quantile multi-armed bandits: Optimal best-arm identification and a differentially private scheme. *IEEE Journal on Selected Areas in Information Theory*, 2021.

Limin Peng and Yijian Huang. Survival analysis with quantile regression models. *Journal of the American Statistical Association*, 103(482):637–649, 2008.

Boris T Polyak and Anatoli B Juditsky. Acceleration of stochastic approximation by averaging. *SIAM journal on control and optimization*, 30(4):838–855, 1992.

Herbert Robbins and Sutton Monro. A stochastic approximation method. *Annals of Mathematical Statistics*, 22:400–407, 1951.

Herbert Robbins and David Siegmund. A convergence theorem for non negative almost supermartingales and some applications. In *Optimizing methods in statistics*, pages 233–257. Elsevier, 1971.

Marzena Rostek. Quantile maximization in decision theory. *The Review of Economic Studies*, 77(1):339–371, 2010.

David Ruppert. Efficient estimations from a slowly convergent robbins-monro process. Technical report, Cornell University Operations Research and Industrial Engineering, 1988.

Balazs Szorenyi, Róbert Busa-Fekete, Paul Weng, and Eyke Hüllermeier. Qualitative multi-armed bandits: A quantile-based approach. In *International Conference on Machine Learning*, pages 1660–1668. PMLR, 2015.

Stanislav Volgushev, Shih-Kang Chao, and Guang Cheng. Distributed inference for quantile regression processes. *The Annals of Statistics*, 47(3):1634–1662, 2019.

Lan Wang, Yu Zhou, Rui Song, and Ben Sherwood. Quantile-optimal treatment regimes. *Journal of the American Statistical Association*, 113(523):1243–1254, 2018.

Michael Woodroofe. Normal approximation and large deviations for the Robbins-Monroe process. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 21:329–338, 1972.

Mengyan Zhang and Cheng Soon Ong. Quantile bandits for best arms identification. In *International Conference on Machine Learning*, pages 12513–12523. PMLR, 2021.

Wanrong Zhu, Xi Chen, and Wei Biao Wu. Online covariance matrix estimation in stochastic gradient descent. *Journal of the American Statistical Association*, (just-accepted):1–30, 2021.