Quantiles on Stream: An Application to Monte Carlo Simulation

Wei WANG

Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
E-mail: wangwei611@amss.ac.cn

Wai-Ki CHING

Advanced Modeling and Applied Computing Laboratory, Department of Mathematics, The University of Hong Kong, Hong Kong, China
E-mail: wching@hku.hk

Shouyang WANG[†]

Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
E-mail: sywang@amss.ac.cn

Lean YU

School of Economics and Management, Beijing University of Chemical Technology, Beijing 100029, China; Centre for Big Data Science, Beijing University of Chemical Technology, Beijing 100029, China
E-mail: yulean@amss.ac.cn

Abstract Monte Carlo simulation is an efficient method to estimate quantile. However, it becomes a serious problem when a huge sample size is required but the memory is insufficient. In this paper, we apply the stream quantile algorithm to Monte Carlo simulation in order to estimate quantile with limited memory. A rigorous theoretical analysis on the properties of the ϵ_n -approximate quantile is proposed in this paper. We prove that if $\epsilon_n = o(n^{-1/2})$, then the ϵ_n -approximate α -quantile computed by any deterministic stream quantile algorithm is a consistent and asymptotically normal estimator of the true quantile q_α . We suggest setting $\epsilon_n = 1/(n^{1/2} \log_{10} n)$ in practice. Two deterministic stream quantile algorithms, including of GK algorithm and ZW algorithm, are employed to illustrate the performance of the ϵ_n -approximate quantile. The numerical example shows that the deterministic stream quantile algorithm can provide desired estimator of the true quantile with less memory.

Keywords quantile; stream; Monte Carlo; simulation; estimation

1 Introduction

Quantiles are commonly used statistics to describe the data distribution. For example, the median (i.e., the 50% quantile) provides a robust way to reflect the average level of the population. Indeed, quantile computation is a fundamental task in data analysis. In practice, quantiles serve as important performance measures in many fields. For example, quantiles,

Received March 14, 2016, accepted April 23, 2016

[†] The corresponding author

also known as Value-at-Risks (VaRs), are widely accepted risk measures in finance industry^[1]. Quantiles can also be used to measure the quality of service (QoS) in service industry^[2].

Monte Carlo methods are powerful tools for solving many problems in science, engineering, finance and statistics, which rely on repeated random sampling to obtain numerical results. They are often the only accessible approaches when dealing with very complicated problems. For a comprehensive review of Monte Carlo methods, readers can refer to Rubinstein and Kroese^[3] and Kroese, et al.^[4]. Besides, Glasserman^[5] and Korn, et al.^[6] presented detailed discussions on Monte Carlo methods from the perspective of financial engineering.

Let X be a univariate random variable, and $F(\cdot)$ be the distribution function of X . For any $0 < \alpha < 1$, the α -quantile of X is defined as $q_\alpha = \inf\{t | F(t) \geq \alpha\}$ ^[7]. The usual approach of estimating the α -quantile by Monte Carlo simulation is as follows. First, we draw n iid (independent and identically distributed) samples, $X_1, X_2, \dots, X_n$, from the distribution $F(\cdot)$. The n samples are stored in memory immediately. Then the samples are sorted, and the sorted result can be given by $X_{1:n} \leq \dots \leq X_{r:n} \leq \dots \leq X_{n:n}$. Note that r is called the rank of $X_{r:n}$. Finally, the α -quantile of X (i.e., q_α) can be estimated by $X_{[\alpha n]:n}$ (called sample quantile). It is well known that $X_{[\alpha n]:n}$ is a strongly consistent and asymptotically normal estimator of q_α under some mild conditions^[7].

However, when the sample size is rather large and there is no sufficient memory to store the whole data, the usual approach for estimating the α -quantile becomes unavailable. For example, if we need to accurately estimate the 99% VaR of a credit portfolio that includes millions of obligors, or if we need to provide a benchmark value of the true quantile that is as accurate as possible, a huge sample size (e.g., $n = 10^{10}$) may be required, and it may cause insufficient memory. In this paper, we focus on how to estimate quantiles accurately with limited memory. Indeed, this problem is a special case of approximating quantiles on stream, since the samples can be generated by Monte Carlo simulation in a streaming fashion. A concise introduction about quantiles on stream can be found in Buragohain and Suri^[8]. Moreover, Munro and Paterson^[9] have shown that any algorithm that exactly computes the sample quantile of a sequence of n data elements in only p passes requires $\Omega(n^{1/p})$ space. Thus, approximation of quantiles is necessary for any stream quantile algorithm with sublinear space. A commonly used notion is ϵ -approximate quantile. For any $0 < \alpha < 1$ and an error parameter $0 < \epsilon < 1$, the ϵ -approximate α -quantile of the samples $X_1, X_2, \dots, X_n$ is any element $X_{r:n}$ satisfying $(\alpha - \epsilon)n \leq r \leq (\alpha + \epsilon)n$ ^[8].

There are a lot of studies about approximating quantiles on stream, which are mainly concentrated in the database literature. In these studies, the error parameter ϵ is generally set as a small positive constant that does not depend on the sample size n , since the length of the stream is often unknown or is treated as infinite. Then it brings tolerable errors for the estimates of the true quantiles. Consequently, there exist some obstacles when applying the algorithms of approximating quantiles on stream to Monte Carlo simulation. In Monte Carlo simulation, the sample size n is known in advance. Furthermore, an estimator of the quantile is required to possess good statistical properties (such as consistency and asymptotic normality). Thus, we need to propose a good estimator of the true quantile based on the ϵ -approximate quantile. By treating the error parameter ϵ as a function of the sample size n , i.e., $\epsilon = \epsilon_n$, we prove that

the ϵ_n -approximate α -quantile is a consistent and asymptotically normal estimator of the true quantile q_α if $n^{1/2}\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$. Moreover, we make a study on the error bound of the ϵ_n -approximate quantile. A note is that the focus of this paper is to make a theoretical analysis on the existing stream quantile algorithms, and we do not intend to propose a new algorithm. Therefore, the main contribution of this paper is that we fill the gaps between approximating quantiles on stream and Monte Carlo simulation by a rigorous theoretical analysis. The results we develop here can provide meaningful guidelines for estimating quantiles through Monte Carlo simulation and a stream quantile algorithm with limited memory.

The rest of the paper is organized as follows. In Section 2, we review the related work on approximating quantiles on stream. The theoretical analysis about the asymptotic properties and error bound of the ϵ -approximate quantile is presented in Section 3. A numerical example is given in Section 4 to show the results when estimating quantiles through Monte Carlo simulation and stream quantiles algorithms. The conclusions are drawn in Section 5.

2 Quantiles on Stream

In this section, we present a short review of the related studies on approximating quantiles on stream. Some notations should be introduced here. Let $\{A_n\}_{n=1}^\infty$ and $\{B_n\}_{n=1}^\infty$ be two sequences of real numbers (or random variables). The notation $A_n = O(B_n)$ means that there exists a positive constant c and a positive integer n_0 , such that $|A_n| \leq c|B_n|$ for all $n \geq n_0$. The notation $A_n = \Omega(B_n)$ means $B_n = O(A_n)$. The notation $A_n = o(B_n)$ means that $A_n/B_n \rightarrow 0$ as $n \rightarrow \infty$.

A data stream indicates that data elements can arrive one by one, or one block by one block. The length of the stream may be unknown. The total elements in the stream could not be stored because of the limited memory. Many effective algorithms aimed at computing the approximated quantiles on stream have been proposed in the related literature. These algorithms can be classified as randomized algorithms and deterministic algorithms^[8].

A random algorithm can find the ϵ -approximate quantile with a guaranteed probability. For example, Manku, et al.^[10] proposed a random algorithm to compute the ϵ -approximate quantile with probability at least $1 - \delta$ without a prior knowledge of the sample size n , using $O(\epsilon^{-1}(\log^2 \epsilon^{-1} + \log^2 \log \delta^{-1}))$ space. Other random algorithms can be referred to Manku, et al.^[11], Gilbert, et al.^[12], Cormode and Muthukrishnan^[13], Guha and McGregor^[14], Wang, et al.^[15], and Agarwal, et al.^[16]. However, there is a small but positive probability that a random algorithm fails to find the ϵ -approximate quantile. Thus, random algorithms are unsuited to our problem which requires a high accuracy on the estimation of the quantile.

A deterministic algorithm can guarantee that the computed quantile is the ϵ -approximate quantile with probability 1. For example, Manku, et al.^[11] presented a single pass deterministic algorithm that constructs an ϵ -approximate quantile summary with the prior knowledge of the sample size n , using $O(\epsilon^{-1} \log^2(\epsilon n))$ space. Greenwald and Khanna^[17] improved the algorithm by Manku, et al.^[11] to achieve a storage bound of $O(\epsilon^{-1} \log(\epsilon n))$ without a prior knowledge of the sample size n . Zhang and Wang^[18] presented a fast deterministic algorithm for computing ϵ -approximate quantile in high speed data streams, and the storage requirement is bounded by $O(\epsilon^{-1} \log^2(\epsilon n))$. Other deterministic algorithms can be referred to Cormode and

Muthukrishnan^[13], Guha and McGregor^[14], Wang, et al.^[15], Raatikainen^[19], and Cormode, et al.^[20].

We mainly consider the deterministic stream quantile algorithms in the context of Monte Carlo simulation in this paper. GK algorithm^[17] and ZW algorithm^[18] are two popular deterministic stream quantile algorithms, which are employed in the numerical example in Section 4. The two algorithms are designed with kinds of very fine operations. Since the main consideration and contribution in this paper are the theoretical analysis of the stream quantile, any deterministic stream quantile algorithm can be treated as a black box in our analysis, so we skip the discussions about the details of GK algorithm and ZW algorithm. Readers can refer to Greenwald and Khanna^[17] and Zhang and Wang^[18] for more details.

3 Theoretical Analysis

Suppose that we want to estimate the α -quantile of a random variable X , i.e., q_α , for a certain $\alpha \in (0, 1)$. Note that the sample size n is known in advance when implementing a Monte Carlo simulation, so we can treat ϵ as a function of n , i.e., $\epsilon = \epsilon_n$. The sample distribution function of $\{X_1, X_2, \dots, X_n\}$ is denoted by $F_n(\cdot)$. Moreover, let $X_{r_n:n}$ be the ϵ_n -approximate α -quantile computed by any deterministic stream quantile algorithm, satisfying $(\alpha - \epsilon_n)n \leq r_n \leq (\alpha + \epsilon_n)n$. To facilitate our discussion, we make a typical assumption on the density function of the random variable X .

Assumption 1 The random variable X has a density function $f(\cdot)$, satisfying $f(t)$ is continuous in the neighborhood of $t = q_\alpha$, and $f(q_\alpha) > 0$.

Assumption 1 implies that $\mathbb{P}(X = q_\alpha) = 0$, then q_α is unique and $F(q_\alpha) = \mathbb{P}(X \leq q_\alpha) = \alpha$. Consequently, we present a rigorous theoretical analysis on the properties and error bound of $X_{r_n:n}$ in this section.

3.1 Asymptotic Properties

The analysis on the asymptotic properties (e.g., consistency and asymptotic normality) of $X_{r_n:n}$ depends on Bahadur's asymptotic representation for order statistics^[7], which is summarized in the following lemma.

Lemma 1 (Bahadur) *Let $0 < \alpha < 1$. Suppose that F is twice differentiable at q_α , with $F'(q_\alpha) = f(q_\alpha) > 0$. Let $\{k_n\}_{n=1}^\infty$ be a sequence of positive integers ($1 \leq k_n \leq n$) such that*

$$\frac{k_n}{n} = \alpha + o\left(\frac{(\log n)^\Delta}{n^{1/2}}\right), \quad n \rightarrow \infty,$$

for some $\Delta \geq \frac{1}{2}$. Then

$$X_{k_n:n} = q_\alpha + \frac{k_n/n - F_n(q_\alpha)}{f(q_\alpha)} + R_n,$$

where with probability 1

$$R_n = O\left(n^{-3/4}(\log n)^{(\Delta+1)/2}\right), \quad n \rightarrow \infty.$$

The consistency and asymptotic normality of $X_{r_n:n}$ can be easily derived from Lemma 1, if we assume a sufficient condition on ϵ_n . The result is summarized in the following proposition.

Proposition 1 Suppose the conditions in Assumption 1 and Lemma 1 are satisfied. If $\epsilon_n = o(n^{-1/2})$, then the ϵ_n -approximate α -quantile, $X_{r_n:n}$, is a strongly consistent and asymptotically normal estimator of q_α , i.e., as $n \rightarrow \infty$,

$$X_{r_n:n} \xrightarrow{wp1} q_\alpha \quad \text{and} \quad n^{1/2}(X_{r_n:n} - q_\alpha) \xrightarrow{d} N\left(0, \frac{\alpha(1-\alpha)}{f^2(q_\alpha)}\right). \quad (1)$$

Proof According to $(\alpha - \epsilon_n)n \leq r_n \leq (\alpha + \epsilon_n)n$ and $\epsilon_n = o(n^{-1/2})$, we know that

$$\frac{r_n}{n} = \alpha + o\left(\frac{1}{n^{1/2}}\right), \quad n \rightarrow \infty.$$

Then applying with Lemma 1, with probability 1, we have

$$X_{r_n:n} = q_\alpha + \frac{\alpha - F_n(q_\alpha)}{f(q_\alpha)} + o\left(\frac{1}{n^{1/2}}\right) + O\left(\frac{(\log n)^{3/4}}{n^{3/4}}\right). \quad (2)$$

Note that the sample distribution function $F_n(\cdot)$ satisfies $F_n(q_\alpha) \xrightarrow{wp1} \alpha$ and $n^{1/2}(F_n(q_\alpha) - \alpha) \xrightarrow{d} N(0, \alpha(1-\alpha))^{[7]}$. Thus, Equation (1) can be easily derived from Equation (2). This completes the proof of Proposition 1. ■

Proposition 1 indicates that if ϵ_n converges to zero faster than $n^{-1/2}$, then the ϵ_n -approximate α -quantile computed by any deterministic stream quantile algorithm is a consistent and asymptotically normal estimator of the true quantile q_α . In practice, we suggest setting $\epsilon_n = 1/(n^{1/2} \log_{10} n)$. With such configuration, the storage requirements of GK algorithm and ZW algorithm are $O(\sqrt{n}(\log n)^2)$ and $O(\sqrt{n}(\log n)^3)$, respectively.

Based on Proposition 1, we can construct the confidence interval of q_α . First, we need to make an estimation about the density $f(q_\alpha)$. Kernel method provides a good solution. Kernel method has been studied extensively in the field of nonparametric statistics, especially in density estimation and nonparametric regression analysis. For a comprehensive review of kernel method, interested readers can refer to Bosq^[21], Li and Racine^[22], Pagan and Ullah^[23]. The kernel estimator of $f(q_\alpha)$ can be given by

$$\hat{f}_n(X_{r_n:n}) = \frac{1}{n\delta_n} \sum_{i=1}^n K\left(\frac{X_{r_n:n} - X_i}{\delta_n}\right). \quad (3)$$

In Equation (3), $K(\cdot)$ is a kernel function on $\mathbb{R}$ which satisfies some regularity conditions. For instance, the standard normal density function is a common used kernel function in practice. Besides, δ_n is called bandwidth which satisfies $\delta_n \rightarrow 0$ and $n\delta_n \rightarrow \infty$ as $n \rightarrow \infty$. We use a rule of thumb and suggest setting $\delta_n = 0.79 \cdot \text{IQR}(X) \cdot n^{-1/5}$, where IQR denotes the interquartile range^[23, 24].

We can prove that $\hat{f}_n(X_{r_n:n})$ in Equation (3) is a consistent estimator of $f(q_\alpha)$. The proof can be referred to Liu and Hong^[25]. Hence, for any $0 < \beta < 1$, an asymptotically valid 100 $\beta\%$ confidence interval of q_α can be given by

$$\left(X_{r_n:n} - \Phi^{-1}\left(\frac{1+\beta}{2}\right)\sqrt{\frac{\alpha(1-\alpha)}{n}}\frac{1}{\hat{f}_n(X_{r_n:n})}, \right.$$

$$X_{r_n:n} + \Phi^{-1}\left(\frac{1+\beta}{2}\right) \sqrt{\frac{\alpha(1-\alpha)}{n}} \frac{1}{\widehat{f}_n(X_{r_n:n})}, \quad (4)$$

where Φ is the distribution function of the standard normal random variable.

A simulation with two passes is required if we want to construct the confidence interval of q_α . In the first pass, the ϵ_n -approximate α -quantile $X_{r_n:n}$ is computed. Then the bandwidth δ_n can be calculated immediately. Given $X_{r_n:n}$ and δ_n computed in the first pass, $\widehat{f}(X_{r_n:n})$ can be computed with an online way by Equation (3) in the second pass. The same seed can be configured to generate the pseudo random numbers in the two passes.

3.2 Error Bound

In this subsection, we explore the error bound of the ϵ_n -approximate α -quantile $X_{r_n:n}$, i.e., the difference between $X_{r_n:n}$ and q_α . We require a lemma in Serfling^[7], which is shown as follows.

Lemma 2 *Let $0 < \alpha < 1$. Suppose the conditions in Assumption 1 are satisfied, and $\{k_n\}_{n=1}^\infty$ is a sequence of positive integers defined in Lemma 1. Then with probability 1*

$$|X_{k_n:n} - q_\alpha| \leq \frac{2(\log n)^\Delta}{f(q_\alpha)n^{1/2}}, \quad \text{for all } n \text{ sufficiently large.}$$

Thus, the error bound of $X_{r_n:n}$ is a straightforward result of Lemma 2, and can be summarized in the following proposition.

Proposition 2 *Suppose the conditions in Assumption 1 are satisfied. If $\epsilon_n = o(n^{-1/2})$, then with probability 1, the ϵ_n -approximate α -quantile, $X_{r_n:n}$, satisfies*

$$|X_{r_n:n} - q_\alpha| \leq \frac{2(\log n)^{1/2}}{f(q_\alpha)n^{1/2}}, \quad \text{for all } n \text{ sufficiently large.}$$

Proposition 2 claims that the error bound of $X_{r_n:n}$ is not greater than $2(\log n/n)^{1/2}/f(q_\alpha)$ when n is sufficiently large. Note that $2(\log n/n)^{1/2}/f(q_\alpha)$ tends to zero as $n \rightarrow \infty$. A direct application of Proposition 2 is that we can construct an asymptotically valid 100% confidence interval of q_α , which is given by

$$\left(X_{r_n:n} - \sqrt{\frac{\log n}{n}} \frac{2}{\widehat{f}_n(X_{r_n:n})}, X_{r_n:n} + \sqrt{\frac{\log n}{n}} \frac{2}{\widehat{f}_n(X_{r_n:n})} \right). \quad (5)$$

Comparing with the asymptotically valid $100\beta\%$ confidence interval of q_α for any $0 < \beta < 1$ in Equation (4), there is an extra infinitely large quantity $(\log n)^{1/2}$ in the half width of the asymptotically valid 100% confidence interval of q_α presented in Equation (5). So the 100% confidence interval in Equation (5) provides a loose estimate of q_α .

4 Numerical Examples

In this section, we propose a toy example to illustrate the performance of the ϵ_n -approximate quantile computed by a deterministic stream quantile algorithm. The example has closed forms for the quantiles and density function, so that we can validate the asymptotic properties and error bound of the ϵ_n -approximate quantile. We test the example under Windows 7 on a

machine with a 2.5GHz CPU and 4GB memory. The GK algorithm implemented by Wang et al.^[15] and the ZW algorithm implemented by Intel Math Kernel Library are used here.

Let X be a random variable that follows a normal distribution $N(\mu, \sigma^2)$ with $\mu = 0$ and $\sigma = 100$. We consider to estimate the 95%-quantile of X by GK algorithm and ZW algorithm. The usual approach is provided as a benchmark. We increase the sample size from 10^4 to 10^6 , and compare the performances of the three estimators, i.e., GK quantile, ZW quantile and the usual quantile. All the numerical results reported here are based on 1000 independent replications. Note that the true value of the 95%-quantile of X has a closed form that $q_{0.95} = \sigma \Phi^{-1}(0.95) = 164.4854$.

We first study the consistencies of the three estimators. Figure 1 shows the comparison of relative RMSEs (defined as the ratio of RMSE and the true value) amongst GK quantile, ZW quantile and the usual quantile. All the relative RMSEs become decreasing as the sample size increases, and are less than 0.002 when the sample size is larger than 5×10^5 . So the three estimators appear to be consistent. Besides, the relative RMSE of GK quantile is larger than the relative RMSEs of ZW quantile and the usual quantile, and the relative RMSEs of ZW quantile and the usual quantile are nearly same.

Figure 1 Performances of the relative RMSEs for GK quantile, ZW quantile and the usual quantile ($\alpha = 0.95$)

Figure 2 Performances of the 90% confidence intervals for GK quantile, ZW quantile and the usual quantile ($\alpha = 0.95$)

Consequently, we study the asymptotic normalities of the three estimators. Figure 2 shows the coverage probabilities of the 90% confidence intervals for GK quantile, ZW quantile and the usual quantile. We can see that when the sample size varies between 10^4 and 10^6 , the coverage probabilities associated with ZW quantile and the usual quantile are nearly 90%, yet the coverage probability associated with GK quantile is close but less than 90%. So the confidence intervals of the three estimators appears to be asymptotically valid.

Finally, we study the error bounds for the three estimators. Since the results are almost same for GK quantile, ZW quantile and the usual quantile, we only show the result for ZW quantile in Figure 3. The horizontal line in Figure 3 corresponds to the true quantile. The scatters around the horizontal line correspond to the estimators under different sample sizes and different runs. It is shown that the gap between the upper bound and the lower bound for the 100% confidence interval decreases as the sample size increases. Moreover, all the scatters

are in the area between the upper bound and the lower bound, and are far away from the two bounds. So it verifies that the 100% confidence interval proposed in Section 3 provides a loose estimate of the true quantile.

Figure 3 Error bound for ZW quantile ($\alpha = 0.95$)

In summary, the three estimators, GK quantile, ZW quantile and the usual quantile, appear to be consistent and asymptotically normal. GK quantile performs a little worse than ZW quantile and the usual quantile, which may be caused by the less memory used by GK algorithm. The performances of ZW quantile and the usual quantile are nearly same. So ZW algorithm provides a desired estimator with less memory than the usual approach.

5 Conclusions

In this paper, we apply the stream quantile algorithms to Monte Carlo simulation in order to estimate quantile with limited memory. We propose a rigorous theoretical analysis on the properties of the ϵ_n -approximate quantile computed by any deterministic stream quantile algorithm. We show that if the condition $\epsilon_n = o(n^{-1/2})$ is satisfied, then the ϵ_n -approximate α -quantile is a consistent and asymptotically normal estimator of the true quantile q_α . Two popular stream quantile algorithms, i.e., GK algorithm and ZW algorithm, are employed to illustrate the performance of the ϵ_n -approximate quantile. The storage requirements of GK algorithm and ZW algorithm are $O(\sqrt{n}(\log n)^2)$ and $O(\sqrt{n}(\log n)^3)$ respectively, with the configuration that setting $\epsilon_n = 1/(n^{1/2} \log_{10} n)$ in practice. We find that ZW algorithm performs nearly as same as the usual approach with less memory, yet GK algorithm produces a little loss on the estimated accuracy. Thus we suggest using ZW algorithm when handling quantile estimation or related problems through Monte Carlo simulation with limited memory. Note that GK algorithm and ZW algorithm are aimed at estimating quantiles on stream in general cases, yet quantile estimation through Monte Carlo simulation with a deterministic stream quantile algorithm is a special case. Hence, we may develop a more efficient deterministic stream quantile algorithm to estimate quantiles through Monte Carlo simulation with limited memory. It is a direction for future research.

References

- [1] Duffie D, Pan J. An overview of value at risk. *The Journal of Derivatives*, 1997, 4(3): 7–49.

- [2] Volk-Makarewicz W, Heidergott B. Sensitivity analysis of ranked data: From order statistics to quantiles. *Discrete Event Dynamic Systems*, 2015, 25(4): 453–495.
- [3] Rubinstein R Y, Kroese D P. *Simulation and the Monte Carlo method*. John Wiley & Sons, New York, 2011.
- [4] Kroese D P, Taimre T, Botev Z I. *Handbook of Monte Carlo methods*. John Wiley & Sons, New York, 2011.
- [5] Glasserman P. *Monte Carlo methods in financial engineering*. Springer-Verlag, New York, 2004.
- [6] Korn R, Korn E, Kroisandt G. *Monte Carlo methods and models in finance and insurance*. CRC Press, Boca Raton, 2010.
- [7] Serfling R J. *Approximation theorems of mathematical statistics*. John Wiley & Sons, New York, 1980.
- [8] Buragohain C, Suri S. Quantiles on streams. *Encyclopedia of Database Systems*. Springer US, 2009: 2235–2240.
- [9] Munro J I, Paterson M S. Selection and sorting with limited storage. *Theoretical Computer Science*, 1980, 12(3): 315–323.
- [10] Manku G S, Rajagopalan S, Lindsay B G. Random sampling techniques for space efficient online computation of order statistics of large datasets. *ACM SIGMOD Record*. ACM, 1999, 28(2): 251–262.
- [11] Manku G S, Rajagopalan S, Lindsay B G. Approximate medians and other quantiles in one pass and with limited memory. *ACM SIGMOD Record*. ACM, 1998, 27(2): 426–435.
- [12] Gilbert A C, Kotidis Y, Muthukrishnan S, et al. How to summarize the universe: Dynamic maintenance of quantiles. *Proceedings of the 28th International Conference on Very Large Data Bases. VLDB Endowment*, 2002: 454–465.
- [13] Cormode G, Muthukrishnan S. An improved data stream summary: The count-min sketch and its applications. *Journal of Algorithms*, 2005, 55(1): 58–75.
- [14] Guha S, McGregor A. Stream order and order statistics: Quantile estimation in random-order streams. *SIAM Journal on Computing*, 2009, 38(5): 2044–2059.
- [15] Wang L, Luo G, Yi K, et al. Quantiles over data streams: An experimental study. *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*. ACM, 2013: 737–748.
- [16] Agarwal P K, Cormode G, Huang Z, et al. Mergeable summaries. *ACM Transactions on Database Systems*, 2013, 38(4): 26.
- [17] Greenwald M, Khanna S. Space-efficient online computation of quantile summaries. *ACM SIGMOD Record*. ACM, 2001, 30(2): 58–66.
- [18] Zhang Q, Wang W. A fast algorithm for approximate quantiles in high speed data streams. *19th International Conference on Scientific and Statistical Database Management, IEEE*, 2007: 29.
- [19] Raatikainen K E E. Simultaneous estimation of several percentiles. *Simulation*, 1987, 49(4): 159–163.
- [20] Cormode G, Korn F, Muthukrishnan S, et al. Space- and time-efficient deterministic algorithms for biased quantiles over data streams. *Proceedings of the 25th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, ACM, 2006: 263–272.
- [21] Bosq D. *Nonparametric statistics for stochastic processes: Estimation and prediction*. Springer, New York, 1996.
- [22] Li Q, Racine J S. *Nonparametric econometrics: Theory and practice*. Princeton University Press, Princeton, 2007.
- [23] Pagan A, Ullah A. *Nonparametric econometrics*. Cambridge University Press, Cambridge, 1999.
- [24] Silverman B W. *Density estimation for statistics and data analysis*. CRC Press, Boca Raton, 1986.
- [25] Liu G, Hong L J. Kernel estimation of quantile sensitivities. *Naval Research Logistics*, 2009, 56(6): 511–525.