Proceedings of the 6th International Conference on Statistics: Theory and Applications (ICSTA'24)
Barcelona, Spain – August 19 - 21, 2024
Paper No. 161
DOI: 10.11159/icsta24.161

Bahadur Type Representation and Berry-Esseen bound for Sample Quantiles Under Weak Dependence

Bahaeddine Taoufik[1], Lahcen Douge[2]
[1]Department of Mathematics/Saint Joseph's University,
Philadelphia, USA
btaoufik@sju.edu; lahcen.douge@uca.ac.ma
[2]Laboratoire de Modélisation des Systèmes Complexes /Cadi Ayyad University
Marrakesh, Morocco

Abstract **-** In this paper, we investigate the Bahadur representation of the empirical quantile estimator based on ψ weakly dependent sequences. We obtain the optimal rate under the assumption of exponential decay of the dependence coefficient. As an application, we establish a Berry-Esseen bound and deduce the asymptotic normality. Additionally, we conduct Monte Carlo simulations to evaluate the finite sample performance of the proposed estimators for a weakly dependent and non-mixing model.

Keywords**:** Bahadur representation; sample quantiles; empirical process, weak dependence.

1. Introduction

The Bahadur representation proves valuable in establishing consistency and asymptotic normality for sample quantiles. Bahadur [1] initially established the asymptotic representation for sample quantiles using the empirical distribution function based on independent and identically distributed (i.i.d.) random variables. Kiefer [2] further provided exact rates in the Bahadur representation for i.i.d. sequences. Extensions of Bahadur-type representations, relaxing the independence assumption, have been explored by various authors. Notably, Sen [3] obtained results similar to Bahadur's for stationary ϕ-mixing processes, while Yoshihara [4] generalized Sen's findings for ϕ-mixing and α-mixing sequences. Recent research on Bahadur representation for sample quantiles includes contributions from [5], [6], and [7] for mixing sequences, as well as [8] for negatively associated sequences. Additionally, extensions of Bahadur's representations under weak dependence have been investigated by [9], [10], [11], [12], and [7]. Regarding Berry–Esseen bounds of sample quantiles, references such as [13] [Theorem 2.3.3 C], [14], [15], [16], and related literature provide further insights.

In this study, we establish the Bahadur representation for ψ-weakly dependent random variables under an exponential decay of the dependence coefficient and derive a corresponding Berry-Esseen bound.

A sequence of random variables $(X_i)_{i\in Z}$ is said to be ψ_w-dependent if for each integer $u, v \geq 1$ one has, for any sequences $(i_1, \dots, i_u)$ and $(j_1, \dots, j_v)$ such that $i_1 \leq \dots \leq i_u \leq j_1 \leq \dots \leq j_v$ and $j_1 - i_u := r$, for any bounded Lipchitz functions $G(\cdot)\ and\ H(\cdot)$, defined on $\mathbb{R}^u$ and $\mathbb{R}^v$ respectively,

$$\left|\mathrm{Cov}\left(G\left(X_{i_1}, \dots, X_{i_u}\right), H\left(X_{j_1}, \dots, X_{j_v}\right)\right)\right| \leq \psi_\omega(G, H, u, v)\, \theta_r, \quad \omega \in \{0,1\} \tag{1}$$

where $\psi_1(G, H, u, v) = u\mathrm{Lip}(G) + v\mathrm{Lip}(H)$ and $\psi_2(G, H, u, v) = uv\mathrm{Lip}(G)\mathrm{Lip}(H)$ with $\mathrm{Lip}(G)$ is the Lipchitz modulus of G. As an example of models that satisfy this weak dependence condition, we point out that associated and Gaussian sequences are ψ_2-weakly dependent, called also κ dependent. The Bernoulli shifts and bilinear processes are ψ_1-weakly dependent, called also η dependent. We refer the reader to [17] and [18] for more examples and details about these particular processes.

Let $(X_n)_{n\geq 1}$ has continuous distribution function F and Q denote the associated density and quantile function, respectively. For $0 < p < 1$, denote by $Q(p)$ the p th quantile of F. Given a sample $X_1, \dots, X_n$, define the empirical distribution function

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}},\ x \in R, \tag{2}$$

where $\mathbb{1}_A$ denotes the indicator function of a set A and let $Q_n(p) = \inf\{ x: F_n(x) \geq p\}$ the p th sample quantile.

2. Main Results

Throughout this paper $(Z_n)_{n\geq 1}$ is said to be $O_{a.s.}(r_n)$ if $\frac{Z_n}{r_n}$ is almost surely bounded, as $n \to \infty$.

2.1. Assumptions

(A1) f possesses a bounded derivative f' in a neighbourhood of $Q(p)$ and $f\big(Q(p)\big) > 0$.

(A2) $\theta_{w,r} \coloneqq \theta_r^{\frac{1}{w+1}} \leq C\, e^{-br}$, C and b are some positive constants.

2.2. Bahadur Representation

Theorem 2.1. *Under assumptions (A1) and (A2), we have*

$$|Q_n(p) - Q(p)| = O_{a.s.}\left(\frac{\log n}{n}\right)^{1/2}. \tag{3}$$

Theorem 2.2. *Under assumptions (A1) and (A2), we have*

$$Q_n(p) - Q(p) = \frac{p - F_n\big(Q(p)\big)}{f\big(Q(p)\big)} + O_{a.s.}\left(n^{-\frac{3}{4}} \log^{\gamma}(n)\right), \quad \gamma > 5/4. \tag{4}$$

2.3. Rates in the Central Limit Theorem

By using the result above, we can establish the rates in the uniformly asymptotic normality of the sample quantiles for ψ_w weakly dependent random variables. Denote

$$\sigma_p^2 \coloneqq \mathrm{Var}\big(\mathbb{1}_{\{X_1 \leq Q(p)\}}\big) + 2\sum_{j=2}^{\infty} \mathrm{Cov}\left(\mathbb{1}_{\{X_1 \leq Q(p)\}}, \mathbb{1}_{\{X_j \leq Q(p)\}}\right). \tag{5}$$

Theorem 2.3. Under assumptions (A1) and (A2), we have

$$\sup_{t\in R}\left|\mathbf{P}\left(\frac{\sqrt{n}\big(Q_n(p) - Q(p)\big)}{a_p} \leq t\right) - \Phi(t)\right| = O_{a.s.}\left(n^{-\frac{1}{3}}\right), \tag{6}$$

where $a_p \coloneqq \sigma_p / f\big(Q(p)\big)$ and Φ is the distribution function of a standard normal variable.

Under the conditions of Theorem 2.3, we have a.s. as $n \to \infty$,

$$\frac{\sqrt{n}\big(Q_n(p) - Q(p)\big)}{a_p} \xrightarrow{\mathscr{D}} \mathcal{N}(0,1)$$

where $\xrightarrow{\mathscr{D}}$ stand for the convergence in distribution.

3. Simulations

In this section, a simulation will be conducted to examine the numerical effectiveness of the empirical estimators in approximating the quantile function. The computing program codes are implemented using the programming language R. We investigate the asymptotic normality for this estimator in the case of a weakly dependent and non-mixing model. We simulate the following process

$$X_t = \frac{1}{2}(X_{t-1} + \epsilon_t), \quad t = 1, \dots, n, \tag{7}$$

with Bernoulli innovations $\mathbf{P}(\epsilon_t = 0) = \mathbf{P}(\epsilon_t = 1) = \frac{1}{2}$ and X_0 is distributed uniformly on the interval [0,1] and is independent of the sequence $(\epsilon_t)_{t\geq 1}$. For each $t \geq 1$, X_i presented in Figure 1, can be expressed in the following form:

$$X_t = \sum_{k\geq 0} 2^{-(k+1)}\, \epsilon_{t-k}. \tag{8}$$

This model is ψ_1-weakly dependent, with $\theta_r \leq C\exp(-ar)$, a and $C > 0$, but it satisfies no mixing conditions. We compute $S_{n,1}(p) = |Q_n(p) - Q(p)|$. We calculate the values of $S_{n,1}$ for 1000 times by taking $n = 600{,}4000$, with $p = 0.1, \dots, 0.9$. We depict the Q-Q plots of $\sqrt{n}S_{n,1}(p)$ versus the quantiles of standard normal distribution for $p = 0.1, \dots, 0.9$. According to Theorem 2.3, the distributions of $\sqrt{n}S_{n,1}(p)$ should be asymptotically normal. Figures 2 and 3 represent the Q-Q plots with $n = 600{,}4000$ respectively. The Q-Q plots show good fits of the $\sqrt{n}S_{n,1}(p)$ to normal distribution for all values of p. The simulation results are consistent with the theoretical results obtained in this paper. In addition, we investigate the behavior of the absolute mean bias, standard deviation and the p-value of the Shapiro-Wilk and Jarque-Bera normality tests. We conducted 1000 replications with varying values of $n = 200, 400, 600, 800$. The results are presented in Table 1. Simple inspection of the results reported in Table 1 allows us to deduce that, for different values of p, large sample sizes n result in smaller absolute mean bias, standard deviation and better closeness to the normality.

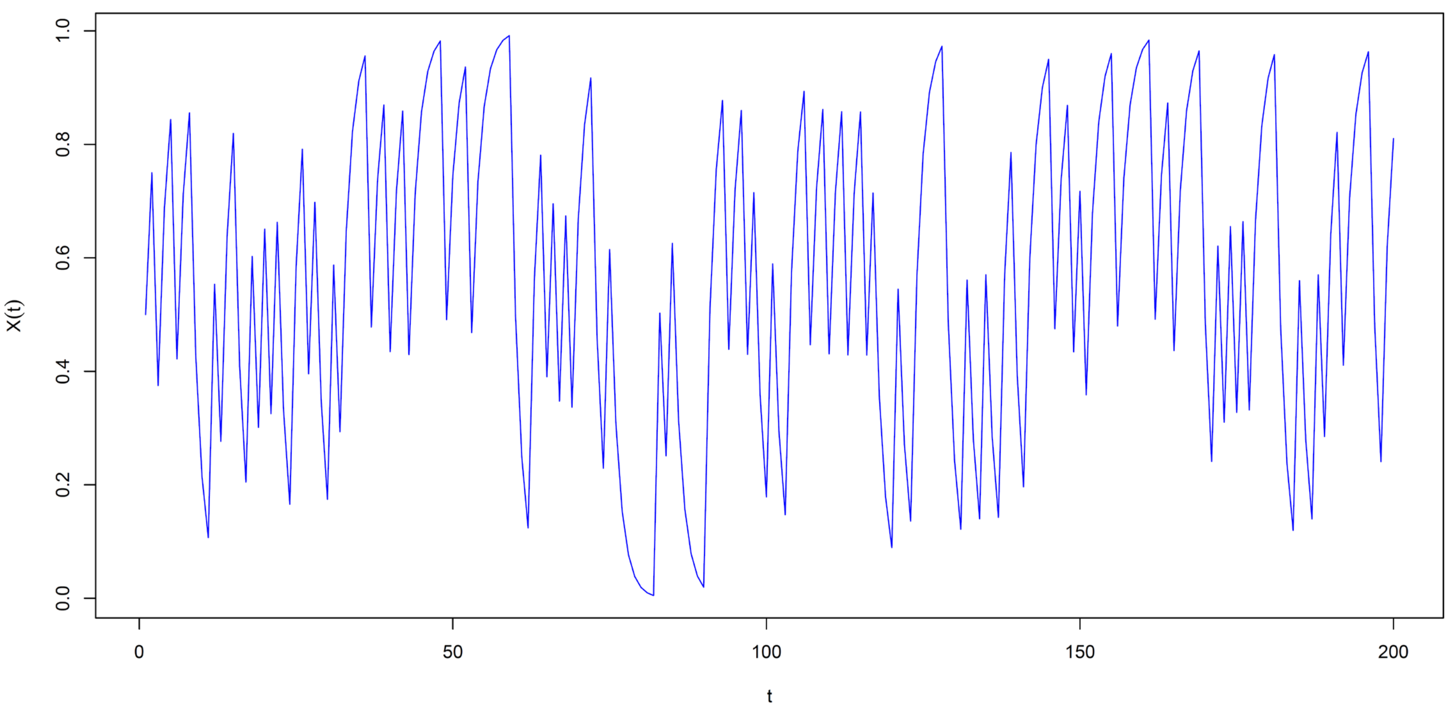

Fig. 1: Plot of Bernoulli shift for $n = 200$.

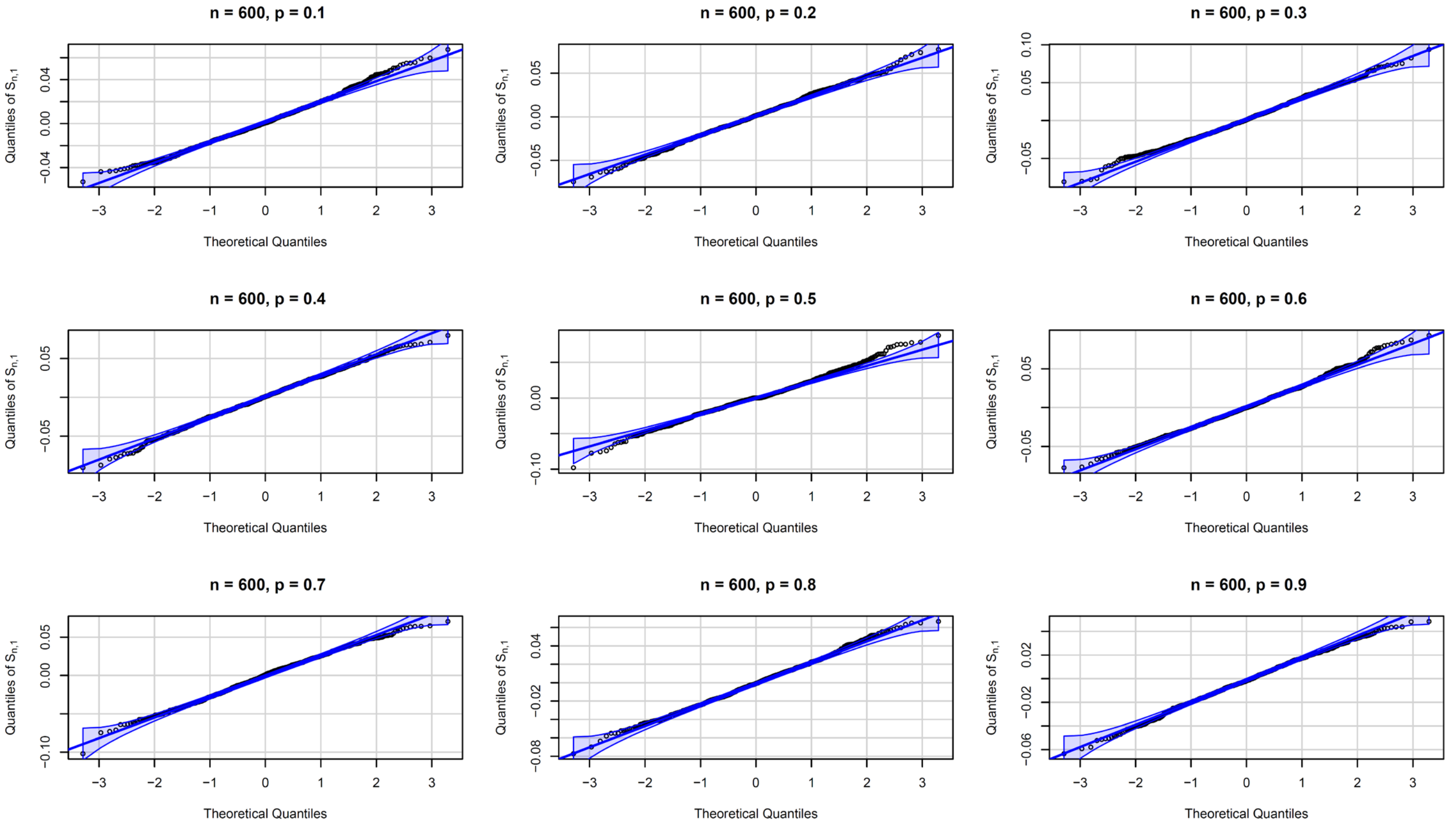

Fig. 2: Normal Q-Q plot of $S_{n,1}$ with different values of p for $n = 600$.

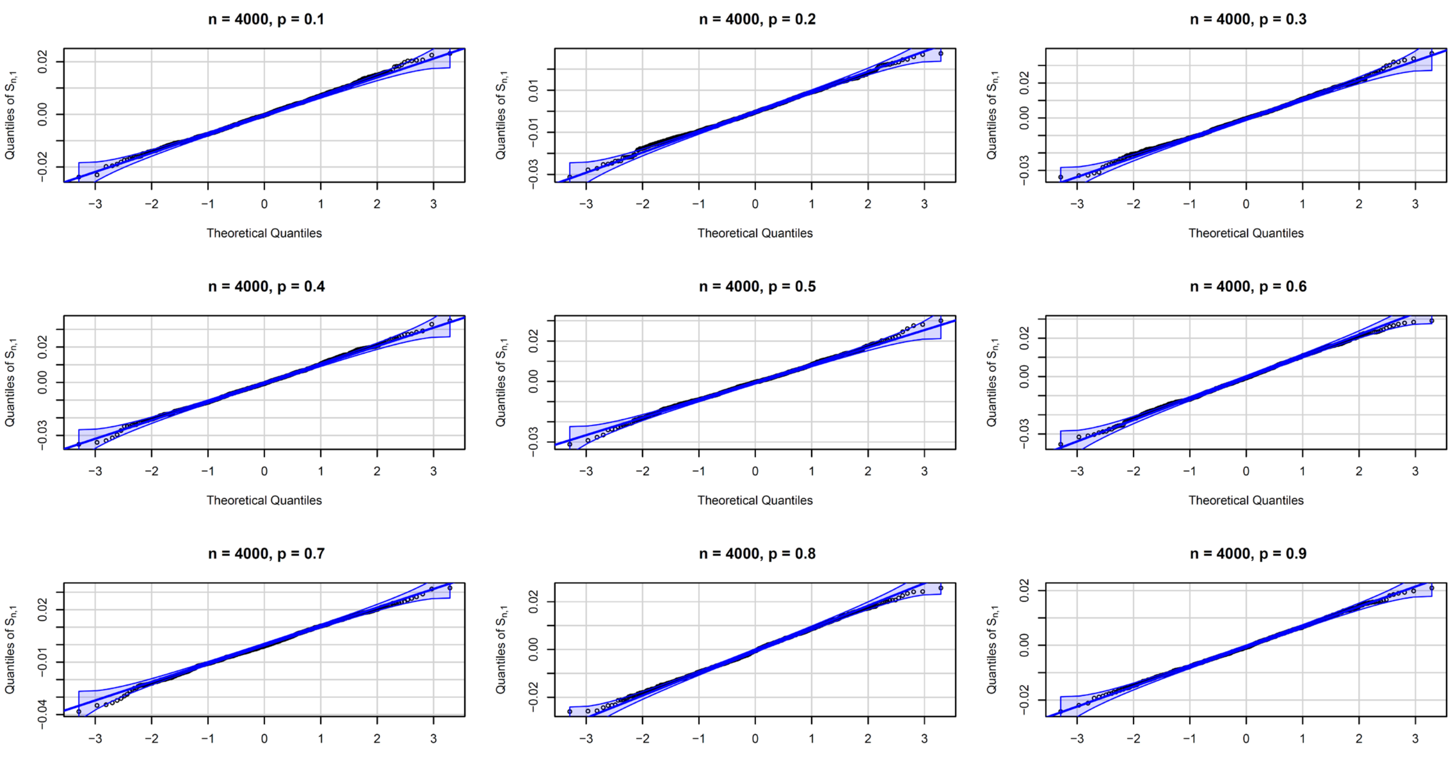

Fig. 3: Normal Q-Q plot of $S_{n,1}$ with different values of p for $n = 4000$.

Table 1: Results for the absolute means bias, standard deviation, and the p-values of Jarque-Bera and Shapiro-Wilk normality tests of $S_{n,1}$ for $n = 200, 400, 600, 800$.

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$n = 200$									
Mean bias of $S_{n,1}$	0.025881	0.032423	0.036332	0.037033	0.034126	0.037992	0.036759	0.033951	0.027693
sd of $S_{n,1}$	0.032222	0.040592	0.045339	0.04618	0.045394	0.047303	0.045572	0.042533	0.034250
Jarque-Bera Test of $S_{n,1}$	0.00076882	0.48127	0.52660	0.0078335	0.0013485	0.0032121	0.40469	0.0032844	7.4838e−06
Shapiro-Wilk Test of $S_{n,1}$	4.6834e−05	0.011088	0.039408	0.0012430	8.6110e−05	5.5703e−05	0.10700	0.0022121	1.7926e−06
$n = 400$									
Mean bias of $S_{n,1}$	0.019553	0.024051	0.027319	0.027350	0.024258	0.027365	0.026663	0.023394	0.019236
sd of $S_{n,1}$	0.024294	0.029871	0.03362	0.034442	0.032307	0.034590	0.033406	0.029738	0.024069
Jarque-Bera Test of $S_{n,1}$	0.001006	0.49224	0.033163	0.26469	6.3317e−07	0.0070512	0.43548	0.051708	7.7172e−05
Shapiro-Wilk Test of $S_{n,1}$	6.3733e−05	0.036555	0.0044032	0.069593	5.9607e−06	0.0056282	0.042307	0.036567	0.00016745
$n = 600$									
Mean bias of $S_{n,1}$	0.015034	0.018528	0.021704	0.021591	0.018771	0.0218	0.021387	0.018523	0.014935
sd of $S_{n,1}$	0.018957	0.023278	0.026911	0.027012	0.024767	0.027493	0.026308	0.023128	0.01863
Jarque-Bera Test of $S_{n,1}$	0.0039835	0.92111	0.12504	0.25498	8.07015e−05	0.0056627	0.054440	0.42713	0.10357
Shapiro-Wilk Test of $S_{n,1}$	0.0024283	0.57999	0.043693	0.30454	0.00022188	0.0058209	0.02052	0.29180	0.092670
$n = 800$									
Mean bias of $S_{n,1}$	0.01316	0.016329	0.019128	0.01915	0.016354	0.018908	0.018947	0.016351	0.013253
sd of $S_{n,1}$	0.01643	0.020328	0.023649	0.024029	0.021401	0.023818	0.023539	0.020493	0.016403
Jarque-Bera Test of $S_{n,1}$	0.032245	0.28179	0.022583	0.16401	0.012556	0.10743	0.010702	0.18946	0.11482
Shapiro-Wilk Test of $S_{n,1}$	0.009991	0.17416	0.012101	0.17771	0.00079938	0.031714	0.0034098	0.061744	0.037042

4. Conclusion

This work examines the Bahadur representation of the empirical estimator for the quantile function based on ψ_w weakly dependent sequences. Under the exponential decay of the dependence coefficient, the rates at which approximation occurs are optimal. Additionally, we derive a Berry-Esseen bound with a rate $O\left(n^{-\frac{1}{3}}\right)$ and employ Monte Carlo simulations to evaluate the asymptotic normality of the estimator proposed in this study.

Multiple avenues exist for further developing our method. The first is to investigate the behavior of the smooth quantile estimators. The second is to establish the asymptotic properties of the conditional quantile, or what is known as the expected shortfall, under weak dependence.

References

[1] R. R. Bahadur, "A Note on Quantiles in Large Samples," Ann. Math. Statist., pp. 577 - 580, 1966.

[2] J. Kiefer, "On Bahadur's representation of sample quantiles," Ann. Math. Statist, vol. 38, p. 1323–1342, 1967.

[3] P. K. Sen, "On the Bahadur representation of sample quantiles for sequences of ϕ-mixing random variables," J. Multivariate Anal., vol. 2, p. 77–95, 1972.

[4] K.-i. Yoshihara, "The Bahadur representation of sample quantiles for sequences of strongly mixing random variables," Statist. Probab. Lett., vol. 24, no. 4, p. 299–304, 1995.

[5] Y. Wang, W. Yang and S. Hu, "The Bahadur representation of sample quantiles for weakly dependent sequences," Stochastics, vol. 88, no. 3, p. 428–436, 2016.

[6] G. Xing and S. Yang, "On the Bahadur representation of sample quantiles for ψ-mixing sequences and its application," Comm. Statist. Theory Methods, vol. 48, no. 5, p. 1060–1072, 2019.

[7] Y. Wu, W. Yu and X. Wang, "The Bahadur representation of sample quantiles for φ-mixing random variables and its application," Statistics, vol. 55, no. 2, p. 426–444, 2021.

[8] S. Xu, L. Ge. and Y. Miao, "On the Bahadur representation of sample quantiles and order statistics for NA sequences," J. Korean Statist. Soc., vol. 42, no. 1, p. 1–7, 2013.

[9] W. B. Wu, "On the Bahadur representation of sample quantiles for dependent sequences," Ann. Statist., vol. 33, no. 4, p. 1934–1963, 2005.

[10] S. Sun, "The Bahadur representation for sample quantiles under weak dependence," Statist. Probab. Lett., vol. 76, no. 12, p. 1238–1244, 2006.

[11] W.-z. Yang, S.-h. Hu and X.-j. Wang, "The Bahadur representation for sample quantiles under dependent," Acta Math. Appl. Sin. Engl. Ser., vol. 35, no. 3, p. 521–531, 2019.

[12] E. Kong and Y. Xia, "Uniform Bahadur representation for nonparametric censored quantile regression: a redistribution-of-mass approach," Econometric Theory, vol. 33, no. 1, p. 42–261, 2017.

[13] R. J. Serfling,, Approximation theorems of mathematical statistics. Wiley Series in Probability and Mathematical Statistics, New York: John Wiley & Sons, Inc, 1980.

[14] S. N. Lahiri and S. Sun, "A Berry-Esseen theorem for sample quantiles under weak dependence," Ann. Appl. Probab., vol. 19, no. 1, p. 108–126, 2009.

[15] W. Yang, S. Hu, X. Wang and N. Ling, "The Berry-Esséen type bound of sample quantiles for strong mixing sequence," J. Statist. Plann. Inference, vol. 142, no. 3, p. 660–672, 2012.

[16] L. Douge, "A Berry-Esseen theorem for sample quantiles under association," Comm. Statist. Theory, vol. 51, no. 18, p. 6515–6528, 2022.

[17] J. Dedecker, P. Doukhan, G. Lang, R. L. José Rafael, S. Louhichi and C. Prieur, Weak Dependence: With Examples and Applications, Springer, 2007.

[18] P. Doukhan, H. Madre and M. Rosenbaum, "Weak dependence for infinite arch-type bilinear models," Statistics, vol. 41, no. 1, p. 31–45., 2007.