Convergence Rate of Sample Mean for φ-Mixing Random Variables with Heavy-Tailed Distributions*

TANG Fuquan      HAN Dong*
Department of Statistics, School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai, 200240, China

September 20, 2022

Abstract

This article studies the convergence rate of the sample mean for φ-mixing dependent random variables with finite means and infinite variances. Dividing the sample mean into sum of the average of the main parts and the average of the tailed parts, we not only obtain the convergence rate of the sample mean but also prove that the convergence rate of the average of the main parts is faster than that of the average of the tailed parts.

Keywords: convergence rate; sample mean; φ-mixing sequence; heavy-tailed distribution

2010 Mathematics Subject Classification: 62F12

1 Introduction

A kind of phenomenon which often happens in many random systems is that the mean is finite but the variance infinite. For example, consider the log return of Dow Jones Industrial Average $X_k = 100 \log(Y_k/Y_{k-1})$, where $Y_k$ denotes the closed price at $k$th day. By using the data from 2019/04/26 to 2020/02/11, we can check that they are not independent and the tailed index of the Hill estimator can be estimated as 2.1292. This means that the mean is finite but the variance infinite. One of questions which we are concerned here is that what is the convergence rate of the sample mean $n^{-1} \sum_{k=1}^{n} X_k$ in probability when $X_k, 1 \leq k \leq n$, are mutually dependent and all the variances are infinite?

Let us recall some known results. Let $X_n,n \geq 1$, be i.i.d. sequence with the finite $1+\alpha$ moment $E|X_k|^{1+\alpha} = \mu_\alpha$ for some $\alpha \in (0,1]$. Let $\delta \in (0,1)$. Bubeck et al.[2] proposed a truncated empirical mean $\hat{\mu}_T = n^{-1} \sum_{k=1}^{n} X_k I(|X_k| \leq b_k)$ to estimate the mean $\mu$ and proved that the convergence rate of $|\hat{\mu}_T - \mu|$ is $O(\log(1/\delta)/n)^{\alpha/(1+\alpha)}$ with the probability at least $1 - \delta$, where $b_k = (bk/\log(1/\delta))^{1/(1+\alpha)}$, $b \geq \mu_\alpha$ and $I(.)$ is the indicator function. Lee et al.[6] further presented a novel robust estimator with the same convergence rate, where information about

*Supported by National Natural Science Foundation of China (11531001)
* Corresponding author, E-mail: donghan@sjtu.edu.cn
μ₀ is not required in prior. Avella-medina et al.[1] considered Huber’s M-estimator[4] \( \hat{\mu}_H \) of \( \mu \) which is defined as the solution to \( \sum_{k=1}^n \psi_H(X_k - \mu) = 0 \), where \( \psi_H(x) = \min\{H, \max(-H, x)\} \) is the Huber function. They also obtained the same convergence rate of \( |\hat{\mu}_H - \mu| \) when \( H = (\mu_\alpha n / \log(2\delta^{-1}))^{1/(1+\alpha)} \).

We know that the tailed part of a random variable \( X \) with the heavy-tailed distribution can best reflect the characteristics of probability. However, the above authors did not consider the convergence rate of the average of the tailed parts \( n^{-1} \sum_{k=1}^n [X_k I(|X_k| > b_n)] \), where \( b_n \to \infty \) as \( n \to \infty \). This is a problem worth studying. Another problem said by Bubeck et al[2] is that "It is unclear whether similar results could be obtained for heavy-tailed bandits with dependent reward processes", that is, the known results above heavily rely on the independence of the processes.

In this paper, we will investigate the convergence rate of the sample mean for \( \varphi \)-mixing dependent random variables with heavy-tailed distributions and try to solve the above two problems by a different method.

2 Convergence rate of the sample mean

Let \( \{X_n, n \geq 1\} \) be a random sequence satisfying the following \( \varphi \)-mixing condition

\[
\varphi(m) \leq C(m + 1)^{-r} \text{ for some constant } r > 2 \text{ and a positive constant } C. \tag{1}
\]

\[
\varphi(m) = \max_{k \geq 1, A \in \mathcal{F}_k, \mathbb{P}(A) > 0, B \in \mathcal{F}_{k+m}} |\mathbb{P}(B|A) - \mathbb{P}(B)|
\]

for \( m \geq 1 \), where \( \varphi(m) \) is decreasing on \( m \), \( \mathbb{P}(\cdot) \) denotes the probability measure for the sequence \( \{X_k, k \geq 0\} \) and \( \mathcal{F}_k^m = \sigma(X_i, k \leq i \leq m) \) for \( 1 \leq k \leq m \leq +\infty \). From (1) it follows that

\[
\varphi := \sum_{m=1}^\infty \varphi^{1/2}(m) < \infty.
\]

Assume that there is a positive constant \( \alpha \) satisfying \( 0 < \alpha < 1 \) such that

\[
0 < \nu_\alpha = \max_{n \geq 1} \mathbb{E}(|X_n|^{1+\alpha}) < \infty. \tag{2}
\]

**Remark 1.** If \( \{X_n, n \geq 1\} \) has a common marginal heavy-tailed distribution with \( 0 < \mathbb{E}(|X_n|^{1+\alpha}) < \infty \), then the condition (2) holds.

In order to get the convergence rate of the average of the tailed parts, we first extend the result of the von Bahr-Esseen moment inequality for pairwise independent random variables in Theorem 2.1 in Chen et al.[3] to the case of \( \varphi \)-mixing random variables.

**Lemma 1.** Let the conditions (1)-(2) hold. There is a positive constant \( c_\alpha \) not depending on \( n \) such that

\[
\mathbb{E}\left|\sum_{k=1}^n Z_k(b)\right|^{1+\alpha} \leq 2^{1+\alpha} c_\alpha \nu_\alpha n. \tag{3}
\]
Proof. Let \( Z_k = Z_k(b) \), \( s = 1 + \alpha \) and \( M_n = \sum_{k=1}^{n} \mathbb{E}(|Z_k|^s) \). For any \( \varepsilon > 0 \), by the same method as proving (2.1)-(2.3) in Theorem 2.1 in Chen et al.\([3]\), we can get that

\[
\mathbb{E}\left| \sum_{k=1}^{n} Z_k \right|^s \leq (2 + \varepsilon) M_n + J_n
\]

and

\[
\sup_{x \geq (1 + \varepsilon) M_n} x^{-1/s} \left| \sum_{k=1}^{n} \mathbb{E}[Z_k I(|Z_k| \leq x^{1/s})] \right| = \sup_{x \geq (1 + \varepsilon) M_n} x^{-1/s} \left| \sum_{k=1}^{n} \mathbb{E}[Z_k I(|Z_k| > x^{1/s})] \right|
\]

\[
\leq \sup_{x \geq (1 + \varepsilon) M_n} x^{-1/s} x^{1/s-1} \sum_{k=1}^{n} \mathbb{E}(|Z_k|^s I(|Z_k| > x^{1/s}))
\]

\[
\leq (1 + \varepsilon)^{-1},
\]

where

\[
J_n = \int_{(1 + \varepsilon) M_n}^{\infty} \mathbb{P}\left( \left| \sum_{k=1}^{n} Z_k I(|Z_k| \leq x^{1/s}) \right| > x^{1/s} \right) dx.
\]

Let \( \bar{\varepsilon} = [1 - (1 + \varepsilon)^{-1}]^{-2} \). From Markov’s inequality, (1.11) in Wang et al.\([7]\) and the inequality of \( \varphi \)-mixing (Ibragimov\([5]\), \( \text{Cov}(\xi, \eta) \leq 2[\varphi(m)V\text{ar}(\xi)V\text{ar}(\eta)]^{1/2} \), it follows that

\[
J_n \leq \int_{(1 + \varepsilon) M_n}^{\infty} \mathbb{P}\left( \left| \sum_{k=1}^{n} Z_k I(|Z_k| \leq x^{1/s}) - \mathbb{E}(Z_k I(|Z_k| \leq x^{1/s})) \right| > [1 - (1 + \varepsilon)^{-1}] x^{1/s} \right) dx
\]

\[
\leq \bar{\varepsilon} \sum_{k=1}^{n} \int_{(1 + \varepsilon) M_n}^{\infty} x^{-2/s} \mathbb{E} \left( Z_k^2 I(|Z_k| \leq x^{1/s}) \right) dx
\]

\[
+ 2\bar{\varepsilon} \sum_{k=1}^{n} \int_{(1 + \varepsilon) M_n}^{\infty} x^{-2/s} \left( \sum_{j=k+1}^{n} \text{Cov}(Z_k I(|Z_k| \leq x^{1/s}), Z_j I(|Z_j| \leq x^{1/s})) \right) dx
\]

\[
\leq \bar{\varepsilon} (1 + 4\varphi) \sum_{k=1}^{n} \int_{(1 + \varepsilon) M_n}^{\infty} x^{-2/s} \mathbb{E} \left( Z_k^2 I(|Z_k| \leq x^{1/s}) \right) dx.
\]

Furthermore, like proving (2.5)-(2.7) in Theorem 2.1 in Chen et al.\([3]\), we have

\[
\int_{(1 + \varepsilon) M_n}^{\infty} x^{-2/s} \mathbb{E} \left( Z_k^2 I(|Z_k| \leq x^{1/s}) \right) dx \leq \left( \frac{2s}{2-s} + \frac{2}{2-s} \right) \mathbb{E}(|Z_k|^s).
\]

Hence, from (4) (5) and (6) it follows that

\[
\mathbb{E}\left| \sum_{k=1}^{n} Z_k \right|^{1+\alpha} \leq \left( 2 + \varepsilon + \frac{1 + 4\varphi}{[1 - (1 + \varepsilon)^{-1}]^2} \times \frac{4}{(1 - \alpha)^2} \right) M_n
\]

\[
= f(\varepsilon) M_n.
\]

Note that the function \( f(\varepsilon) \) in (7) is positive and continuous, and \( f(\varepsilon) \) goes to \( \infty \) as \( \varepsilon \downarrow 0 \) and \( \varepsilon \nearrow \infty \), respectively. This means that there is a positive constant \( c_\alpha \) such that \( c_\alpha = \min_{\varepsilon} \{ f(\varepsilon) \} \).
Thus, the inequality (3) follows from (7) since $\mathbb{E}(|Z_k(b)|^{1+\alpha}) \leq 2^{1+\alpha} \nu_\alpha$ for all $k \geq 1$. It completes the proof.

**Remark 2.** It is clear that when $X_n, n \geq 1$, are mutually independent and $\mathbb{E}(X_n^2) < \infty$, the inequality (3) holds for $\alpha = 1$, where $c_\alpha = 1/4$.

Let $Y_k(b) = X_kI(|X_k| \leq b) - \mathbb{E}(X_kI(|X_k| \leq b))$ for $b > 0$ and $k \geq 1$. Here $n^{-1}\sum_{k=1}^n Y_k(b)$ can be considered as the average of the main parts. Let

$$\beta = \frac{\alpha + 1}{(1+\alpha)(1+r)}, \quad \gamma = \frac{\alpha + \alpha'}{(1+\alpha)(1+r)}, \quad A = e^{1/4+\sqrt{\epsilon}C}\nu,$$

where $0 \leq \alpha' \leq \alpha$.

The following theorem shows the convergence rate of the sample mean in probability.

**Theorem 1.** Assume that the conditions (1)-(2) hold. Let $c$ be a positive constant, $\delta \in (0, 1)$, $p = 2\log(1/\delta)$, $a_1 = \sqrt{3c}c\nu_\alpha(1+4c)$ and $a_2 = 2(c_{\alpha}\nu_\alpha)^{1/(1+\alpha)}$. Then, there is a positive number $n_0$ depending on $\delta$ such that

$$P\left|\frac{1}{n}\sum_{k=1}^n |X_k - \mathbb{E}(X_k)| \geq a_1\left|\frac{A}{n}\right|^\beta + a_2\left|\frac{A}{n}\right|^\gamma\right| \leq 2A\delta^c + p^{-\alpha}\left|\frac{A}{n}\right|^{\alpha-\alpha'},$$

for $n \geq n_0$.

**Proof.** Let $b_n = (n/p)^\lambda$, $Y_k = Y_k(b_n) = X_kI(|X_k| \leq b_n) - \mathbb{E}(X_kI(|X_k| \leq b_n))$ and $Z_k = Z_k(b_n) = X_kI(|X_k| > b_n) - \mathbb{E}(X_kI(|X_k| > b_n))$, where $\lambda = (r - 1)/((1+r)(1+\alpha))^{-1}$. Dividing $X_k - \mathbb{E}(X_k)$ into two parts: the main part $Y_k$ and the tailed part $Z_k$, we have

$$P\left|\frac{1}{n}\sum_{k=1}^n |X_k - \mathbb{E}(X_k)| \geq a_1\left|\frac{A}{n}\right|^\beta + a_2\left|\frac{A}{n}\right|^\gamma\right| \leq P\left|\frac{1}{n}\sum_{k=1}^n Y_k \geq a_1\left|\frac{A}{n}\right|^\beta\right| + P\left|\frac{1}{n}\sum_{k=1}^n Z_k \geq a_2\left|\frac{A}{n}\right|^\gamma\right|$$

By Chebyshev’s inequality and Lemma 1 we have

$$P\left|\frac{1}{n}\sum_{k=1}^n Z_k \geq a_2\left|\frac{A}{n}\right|^\gamma\right| \leq \frac{\mathbb{E}\left|\sum_{k=1}^n Z_k\right|^{1+\alpha}}{(a_2n)^{1+\alpha}(p/n)^{(1+\alpha)\gamma}} \leq p^{-\alpha}\left|\frac{A}{n}\right|^{\alpha-\alpha'},$$

where $O((p/n)^\gamma)$ is the convergence rate of the average of the tailed part in probability.

Let $S_n = \sum_{k=1}^n Y_k$. As in Ibragimov [5] and Yang [8], the sum $S_n$ can be written as two parts

$$S_n = \sum_{i=1}^{l+1} \xi_i + \sum_{i=1}^l \eta_i, \quad \xi_i = \sum_{j=1}^m Y_{2(i-1)m+j}, \quad \eta_i = \sum_{j=1}^m Y_{2i-1)m+j}$$

for $1 \leq i \leq l$, $\xi_{l+1} = 0$ for $2lm \geq n$ and $\xi_{l+1} = Y_{2lm+1} + ... + Y_n$ for $2lm < n$, where

$$m = \lfloor 1/4 + \sqrt{\epsilon}C \rfloor, \quad l = \lfloor M^2 \log(1/\delta)(n/p)^{r/(1+r)} \rfloor,$$
where $\lceil x \rceil$ denotes the maximum integer of $x$ and $M$ is a large positive number satisfying that for given a small $0 < \epsilon \leq 1/2$, $1/M \leq \epsilon$. It is clear that $2m = 2[n/2] \approx n$ for large $n$.

Let $\delta_n = a_1(p/n)^\beta$. From Chebyshev’s inequality and Hölder inequality it follows that

$$P\left( \sum_{k=1}^{n} Y_k \geq n\delta_n \right) \leq \exp\{-n\delta_n\}E\exp\{\theta S_n\}$$

$$\leq \exp\{-n\delta_n\}\left( E\exp\{2\theta \sum_{i=1}^{l+1} \xi_i \} \right)^{1/2} \left( E\exp\{2\theta \sum_{i=1}^{l} \eta_i \} \right)^{1/2} \quad (13)$$

for $\theta > 0$. Furthermore, by the inequality of $\varphi$-mixing and (1.11) in Wang et al.\cite{7}, we have

$$E(\xi_i^2) = \sum_{j=1}^{m} E(Y_{2(i-1)m+j}^2) + 2 \sum_{j=1}^{m} \sum_{j'=j+1}^{m} E(Y_{(i-1)m+j}Y_{2(i-1)m+j'})$$

$$\leq (1 + 4\varphi) \sum_{j=1}^{m} E(X_{2(i-1)m+j}^2 I(|X_{2(i-1)m+j}| \leq b_n))$$

$$= (1 + 4\varphi) \sum_{j=1}^{m} E(|X_{2(i-1)m+j}|^{1+\alpha} |X_{2(i-1)m+j}|^{1-\alpha} I(|X_{2(i-1)m+j}| \leq b_n))$$

$$\leq m\nu_\alpha (1 + 4\varphi)(b_n)^{1-\alpha}.$$  

Note that $2\theta |\xi_i| \leq 2\theta mb_n \leq 1/M \leq \epsilon \leq 1/2$ for $1 \leq i \leq l + 1$ when $\theta \leq M(p/n)^\kappa$, where $\kappa = (\alpha + r)[(1 + r)(1 + \alpha)]^{-1}$. Let $\theta \leq M(p/n)^\kappa$. Hence,

$$\log E(\exp\{2\theta \xi_i\}) = \log \left( 1 + 2\theta^2 E(\xi_i^2)(1 + \epsilon) \right) \leq 2\theta^2 E(\xi_i^2)(1 + \epsilon)$$

and therefore,

$$E(\exp\{2\theta \xi_i\}) \leq \exp\{2\theta^2 E(\xi_i^2)(1 + \epsilon)\} \leq \exp\{2\theta^2 m\nu_\alpha (1 + 4\varphi)(b_n)^{1-\alpha}(1 + \epsilon)\}. \quad (14)$$

By using (14), the definition of $\{\xi_i\}$ in (11), the $\varphi$–mixing property and the similar method used by Yang \cite{8}, we can get that

$$E(\exp\{2\theta \sum_{i=1}^{l} \xi_i\}) = E\left( \exp\{2\theta \sum_{i=1}^{l-1} \xi_i\} E(\exp\{2\theta \xi_l\}|\xi_0^{l-1}) \right)$$

$$\leq (E(\exp\{2\theta \xi_l\}) + e^{\epsilon} \varphi(m)) E(\exp\{2\theta \sum_{i=1}^{l-1} \xi_i\})$$

$$\leq \exp\{e^{\epsilon} \varphi(m)l\} \exp\{2\theta^2 m\nu_\alpha (1 + 4\varphi)(b_n)^{1-\alpha}(1 + \epsilon)\}. \quad (15)$$

Similarly, we can obtain that

$$E(\exp\{2\theta \sum_{i=1}^{l} \eta_i\}) \leq \exp\{e^{\epsilon} \varphi(m)l\} \exp\{2\theta^2 m\nu_\alpha (1 + 4\varphi)(b_n)^{1-\alpha}(1 + \epsilon)\}. \quad (16)$$
Note that \(2\theta|\xi_{t+1}| \leq 1/M \leq \epsilon\) for \(\theta \leq M(p/n)^\kappa\), \(2lm \approx n\), \(\varphi(m)l \leq Cm^{-r}l \approx CM^{2r}\) and

\[e^{\frac{\epsilon}{2}}\exp\{e^\epsilon \varphi(m)l\} \leq A\]

for \(\epsilon = 1/2\) and \(M = 2\). It follows from (13), (15)-(16) that

\[P\left(\sum_{k=1}^{n} Y_k \geq n\delta_n\right) \leq e^{\epsilon/2} \exp\{e^\epsilon \varphi(m)l\} \exp\{-\theta n\delta_n\} \exp\{2\theta^2 ml\nu_\alpha(1 + 4\varphi)(b_n)^{1-\alpha}(1 + \epsilon)\}\]

\[\approx e^{\epsilon/2} \exp\{e^\epsilon \varphi(m)l\} \exp\{-n[\delta_n \theta - (1 + \epsilon)\nu_\alpha(1 + 4\varphi)(b_n)^{1-\alpha}\theta^2]\}\]

\[\leq A \exp\{-n[\delta_n \theta - (1 + 1/2)\nu_\alpha(1 + 4\varphi)(b_n)^{1-\alpha}\theta^2]\}\]

Note that the function \(g(\theta) = \delta_n \theta - (1 + 1/2)\nu_\alpha(1 + 4\varphi)(b_n)^{1-\alpha}\theta^2\) arrives its maximum value at \(\theta^*_n = \delta_n[3\nu_\alpha(1 + 4\varphi)(b_n)^{1-\alpha}]^{-1}, \delta^2_n(b_n)^{(1-\alpha)} = 6c\nu_\alpha(1 + 4\varphi)\log(1/\delta)/n\) and \(\kappa < 1\). Take the minimum number \(n_0\) such that \(\theta^*_n \leq 2(p/n)^\kappa\), we can get that

\[P\left(\sum_{k=1}^{n} Y_k \geq n\delta_n\right) \leq A \exp\{-n[\delta^2_n(6\nu_\alpha(1 + 4\varphi)(b_n)^{1-\alpha})]^{-1}\}\]

\[\leq A \exp\{-c\log(1/\delta) = A\delta^c\] (17)

for \(n \geq n_0\). Similarly, we have

\[P\left(\sum_{k=1}^{n} Y_k < -n\delta_n\right) \leq A\delta^c.\] (18)

This means that the convergence rate of the average of the main parts in probability is \(O((p/n)^\beta)\). By (9), (10), (17) and (18), we see that the inequality (8) holds. It completes the proof of Theorem 1.

**Remark 3.** It can be seen from (8), (9) and (10) that the convergence rate of the average of the main parts is faster than that of the average of the tailed parts since \(\beta > \alpha/(1 + \alpha) \geq \gamma\) for \(r < \infty\). Furthermore, taking \(\alpha' = \alpha\), we see that the convergence rate of the average of the tailed parts can arrive at \(O((p/n)^{\alpha/(1+\alpha)})\) with the probability at least \(1 - p^{-\alpha}\).

In the following three corollaries we assume that \(\mu = E(X_k)\) for all \(k \geq 1\). Let \(\alpha' < \alpha\). Take \(c_0\) and \(n_1\) such that \(2A\delta^{\alpha} = \delta/2\) and \(p^{-\alpha}(p/n_1)^{(\alpha-\alpha')/(1+r)} = \delta/2\). We have the following corollary.

**Corollary 1.** Let \(\alpha' < \alpha\) and \(c \geq c_0\). Then, with the probability at least \(1 - \delta\), one has

\[
\left|\frac{1}{n} \sum_{k=1}^{n} X_k - \mu\right| < a_1\left[\frac{2\log(1/\delta)}{n}\right]^{\beta} + a_2\left[\frac{2\log(1/\delta)}{n}\right]^\gamma
\]

for \(n \geq \max\{n_0, n_1\}\).

Note that \(\gamma = \alpha/(1 + \alpha)\) when \(\alpha' = \alpha\). Hence we have the following lemma.

**Corollary 2.** Let \(\alpha' = \alpha\) and \(c \geq c_0\). Then, with the probability at least \(1 - \delta/2 - (2\log(1/\delta))^{-\alpha}\), one has

\[
\left|\frac{1}{n} \sum_{k=1}^{n} X_k - \mu\right| < a_1\left[\frac{2\log(1/\delta)}{n}\right]^{\beta} + a_2\left[\frac{2\log(1/\delta)}{n}\right]^{\alpha/(1+\alpha)}
\]
for \( n \geq n_0 \).

It is known that \( X_n, n \geq 1 \), will become mutually independent when \( r = \infty \) in (1).

**Corollary 3.** If \( X_n, n \geq 1 \), are mutually independent, that is, \( r = \infty \) in (1) and therefore, \( a_1 = \sqrt{3c\nu_\alpha} \) and \( \beta = \gamma = \alpha/(1 + \alpha) \), then, with the probability at least \( 1 - \delta/2 - (2 \log(1/\delta))^{-\alpha} \), one has

\[
|\frac{1}{n} \sum_{k=1}^{n} X_k - \mu| < (a_1 + a_2) \left( \frac{2 \log(1/\delta)}{n} \right)^{\alpha/(1+\alpha)}
\]

for \( n \geq n_0 \) and \( c \geq c_0 \).

**Remark 4.** If \( X_n, n \geq 1 \), are i.i.d. sequence with \( E(|X_n|^{1+\alpha}) < \infty \) for \( 0 \leq \alpha \leq 1 \), then the conditions (1)-(2) hold.

### 3 Conclusion

By the different method we obtain the convergence rate of the sample mean for \( \varphi \)-mixing dependent random variables with finite means and infinite variances. Furthermore, we prove that the convergence rate \( O((2 \log(\delta^{-1})/n)^{\beta}) \) of the average of the main parts is faster than the convergence rate \( O((2 \log(\delta^{-1})/n)^{\alpha/(1+\alpha)}) \) at which the average of the tailed parts can arrive.

If \( X_n, n \geq 1 \), are i.i.d. sequence with \( E(|X_n|^{1+\alpha}) < \infty \) for \( 0 < \alpha \leq 1 \), it follows from the corollary 3 and Remark 3 that we not only get the same results obtained by Bubeck et al.\[2\] and Avella-medina et al.\[1\], but also show that the convergence rate of the average of the tailed parts is \( O((2 \log(\delta^{-1})/n)^{\alpha/(1+\alpha)}) \).

**Acknowledgements** We sincerely thank two reviewers for their precious comments on the manuscript.

### References

[1] AVELLA M M, BATTEY H S, FAN J Q, LI Q. Robust estimation of high-dimensional covariance and precision matrices [J]. *Biometrika*, 2018, 105(2), 271-284.

[2] BUBECK S, CESIA B N, LUGOSI G. Bandits with heavy tail [J]. *IEEE Trans. Inf. Theory*, 2013, 59(11), 7711-7717.

[3] CHEN P Y, BAI P, SUNG S H. The von Bahr-Esseen moment inequality for pairwise independent random variables and applications [J]. *J. Math. Anal. Appl.*, 2014, 419, 1290-1320.

[4] HUBER P. Robust estimation of a location parameter [J]. *Ann. Math. Statist.*, 1964, 35, 73-101.

[5] IBRAGIMOV I A. Some limit theorems for stationary processes [J]. *Theory of Probab. & Its Appl.*, 1962, 7(4), 349-382.

[6] LEE K J, YANG H J, LIM S B, OH S H. Optimal algorithms for stochastic multi-armed bandits with heavy tailed rewards. 34th Conference on Neural Information Processing Systems, NeurIPS 2020, Vancouver, Canada.

[7] Wang X J, Hu S H, Yang W Z, Shen Y. On complete convergence for weighted sums of \( \varphi \)-mixing random variables [J]. *J. Inequal. Appl.*, 2010, 2010, 1-13.

[8] YANG S. Almost sure convergence of weighted sums of mixing sequences(in Chinese) [J]. *J. Syst. Science and Math. Sciences*, 1995, 15(3), 254-265.