Supplementary of "Catoni-type Confidence Sequences under Infinite Variance"

Guanhua Fang[1], Sujay Bhatt[2], Ping Li[3], Gennady Samorodnitsky[4]

[1] *Fudan University* [2] *J.P. Morgan* [3] *VecML Inc.* [4] *Cornell University*

In this supplementary section, we provide more discussions and collect all the missing proofs. Connections and differences between Wang and Ramdas (2023) are made in Section A. Section B is for the proof of main results and Section C is for the supporting Lemma. Section D is dedicated for the proof of the lower bound of stitching methods.

Contents

A Connection and Difference between Wang and Ramdas (2023) 2

B Proofs of Main Results 3

C Proof of Supporting Lemma 26

D Proof of the Lower Bound of Stitching Method 28

A. Connection and Difference between Wang and Ramdas (2023)

A high-level comparison between our paper and Wang and Ramdas (2023) are summarized in Table 1, followed by detailed justifications.

	Wang & Ramdas 2023	Our paper
$1 < p < 2$	Simple modifications from $p = 2$ Sub-optimal constant	Complete Theory Nearly-optimal constant
Lower Bound	Only classical LIL result	New Catoni-style lower bound
Application	Hypothesis testing	Risk control & Confidence set

Table 1: Comparisons between our work and Wang and Ramdas (2023).

Wang and Ramdas (2023) provide a relatively complete theory on establishing confidence width under $p = 2$. However, when $1 < p < 2$, they only extend the result by straightforwardly using the result in Chen et al. (2021). By contrast, with refined calculations, our upper bound result (Theorem 2) is sharp in the sense that the constant in (4.9) reduces to $4\sigma^2$ and matches that in the case $p = 2$. Additionally, our constant $C_p := \left(\frac{p-1}{p}\right)^{p/2}\left(\frac{2-p}{p-1}\right)^{(2-p)/2}$ in $\psi(x)$ is tighter than $C_p = 1/p$. That is, we increase efficiency by $100(1/(C_p p) - 1)\%$ in terms of sample complexity. Moreover, Wang and Ramdas (2023) does not derive a Catoni-style lower bound, and instead they only cite the literature for the classical Law of Iterated Logarithm (LIL). By contrast, we provide lower bound results by giving insights that tuning $\{\lambda_i\}$'s only in Ville's inequality may still provide a sub-optimal width. Lastly, we provide two more applications, risk control and parameter confidence set construction, which bring more interest to the machine learning field.

B. Proofs of Main Results

In this section, we provide proofs of the theoretical results given in the main context.

Proof of Theorem 1: Let $S_n^+ = \sum_{i=1}^n \lambda_i(X_i-\mu)$ and $S_n^- = \sum_{i=1}^n -\lambda_i(X_i-\mu)$ denote two martingales. The confidence intervals and hence the sequences are obtained by applying Lemma 1 to each of these martingales. Let $a = \frac{1}{m_p b^{\frac{1}{p-1}}}\left(\left(\frac{2}{\alpha}\right)^{\frac{1}{p-1}} - 1\right)$. We have from Lemma 1,

$$\mathbb{P}\left(\forall\ n, \sum_{i=1}^n \lambda_i(X_i-\mu) \leq a + b\sum_{i=1}^n \lambda_i^2 \mathbb{E}[|X_i-\mu|^p|\mathcal{F}_{n-1}]\right)$$
$$\geq 1-\alpha/2,$$
$$\mathbb{P}\left(\forall\ n, -\sum_{i=1}^n \lambda_i(X_i-\mu) \leq a + b\sum_{i=1}^n \lambda_i^2 \mathbb{E}[|X_i-\mu|^p|\mathcal{F}_{n-1}]\right)$$
$$\geq 1-\alpha/2.$$

By using the fact that $\mathbb{E}[|X_i-\mu|^p|\mathcal{F}_{n-1}] \leq \upsilon_p$ and taking an union bound the result follows. The sequence that optimizes the width is calculated using (Waudby-Smith and Ramdas, 2024, Eq. (24-28)) and (Wang and Ramdas, 2023, Appendix A) as the minimizer of $b\upsilon_p\lambda^{p-1}+\frac{a}{t\lambda}$, solving which we obtain the desired sequence.

Proof of Theorem 2: For a fixed $x \in \mathbb{R}$ the following processes are also non-negative supermartingales:

$$M_n^+(x) = \prod_{i=1}^n \exp\big\{\phi\big(\lambda_i(X_i-x)\big)\big\} \tag{B.1}$$
$$\exp\left\{-(\mu-x)\sum_{i=1}^n \lambda_i - C_p\upsilon_p\sum_{i=1}^n \lambda_i^p t_i^{-(p-1)} - C_p|\mu-x|^p\sum_{i=1}^n \lambda_i^p(1-t_i)^{-(p-1)}\right\}$$

and

$$M_n^-(x) = \prod_{i=1}^n \exp\big\{-\phi\big(\lambda_i(X_i-x)\big)\big\} \tag{B.2}$$
$$\exp\left\{(\mu-x)\sum_{i=1}^n \lambda_i - C_p\upsilon_p\sum_{i=1}^n \lambda_i^p t_i^{-(p-1)} - C_p|\mu-x|^p\sum_{i=1}^n \lambda_i^p(1-t_i)^{-(p-1)}\right\}.$$

Note that for $x=\mu$ and $t_i \equiv 1$, these processes become M_n^+ and M_n^-, correspondingly. Denote $f_n(x)=\sum_{i=1}^n \phi\big(\lambda_i(X_i-x)\big)$. The maximal inequality for non-negative supermartingales, for every $x \in \mathbb{R}$ and $h>0$,

$$P\Big(\exp\left\{f_n(x)-(\mu-x)\sum_{i=1}^n \lambda_i - C_p\upsilon_p\sum_{i=1}^n \lambda_i^p t_i^{-(p-1)} - C_p|\mu-x|^p\sum_{i=1}^n \lambda_i^p(1-t_i)^{-(p-1)}\right\} \geq h\Big)$$
$$\leq 1/h,$$

which is the same as

$$P\biggl(f_n(x) \geq (\mu - x)\sum_{i=1}^n \lambda_i + C_p v_p \sum_{i=1}^n \lambda_i^p t_i^{-(p-1)} + C_p|\mu - x|^p \sum_{i=1}^n \lambda_i^p (1-t_i)^{-(p-1)} + \log h \biggr) \leq 1/h. \tag{B.3}$$

Choose $h = 2/\varepsilon_n$ for $0 < \varepsilon_n < 1$ and denote

$$B_n^+(x) = (\mu - x)\sum_{i=1}^n \lambda_i + C_p v_p \sum_{i=1}^n \lambda_i^p t_i^{-(p-1)} + C_p|m - x|^p \sum_{i=1}^n \lambda_i^p (1-t_i)^{-(p-1)} + \log 2/\varepsilon_n. \tag{B.4}$$

Then (B.3) translates into

$$P\Bigl(f_n(x) \geq B_n^+(x) \Bigr) \leq \varepsilon_n/2. \tag{B.5}$$

Consider now the equation

$$B_n^+(x) = -C_p v_p \sum_{i=1}^n \lambda_i^p - \log 2/\alpha. \tag{B.6}$$

We will establish conditions under which this equation has real roots. Assuming, for a moment, that such roots exist, let y_n denote the smallest such root. Using (B.6) with $x = y_n$ tells us that on an event of probability at least $1 - \varepsilon_n/2$, we have $f_n(y_n) < -C_p v_p \sum_{i=1}^n \lambda_i^p - \log 2/\alpha$. We conclude by the definition of $x_{-,n}$ in (4.6), that

$$\begin{aligned} & P\Bigl(x_{-,n} < y_n \ \text{ for all } n \text{ for which (B.6) has real roots} \Bigr) \\ \geq \ & 1 - \sum_{n=1}^{\infty} \varepsilon_n/2. \end{aligned} \tag{B.7}$$

We now establish conditions for the equation (B.6) to have real roots. The function B_n^+ is a strictly convex function of x, diverging to infinity at $\pm\infty$, so it has a unique minimum, achieved at the point

$$z_+ = \mu + \left(\frac{\sum_{i=1}^n \lambda_i}{pC_p \sum_{i=1}^n \lambda_i^p (1-t_i)^{-(p-1)}} \right)^{1/(p-1)},$$

and we have

$$\begin{aligned} B_n^+(z_n) &= -\frac{p}{p-1} \frac{(\sum_{i=1}^n \lambda_i)^{p/(p-1)}}{(pC_p \sum_{i=1}^n \lambda_i^p (1-t_i)^{-(p-1)})^{1/(p-1)}} \\ &\quad + C_p v_p \sum_{i=1}^n \lambda_i^p t_i^{-(p-1)} + \log 2/\varepsilon_n. \end{aligned} \tag{B.8}$$

If this minimal value satisfies

$$\begin{aligned} &-\frac{p}{p-1} \frac{(\sum_{i=1}^n \lambda_i)^{p/(p-1)}}{(pC_p \sum_{i=1}^n \lambda_i^p (1-t_i)^{-(p-1)})^{1/(p-1)}} + C_p v_p \sum_{i=1}^n \lambda_i^p t_i^{-(p-1)} + \log 2/\varepsilon_n \\ &\leq -C_p v_p \sum_{i=1}^n \lambda_i^p - \log 2/\alpha, \end{aligned} \tag{B.9}$$

then the equation (B.6) has real roots. Note that we can rewrite the condition (B.9) in the form

$$\begin{aligned} &C_p v_p \sum_{i=1}^n \lambda_i^p (1 + t_i^{-(p-1)}) + \log 2/\alpha + \log 2/\varepsilon_n \\ &\leq \frac{p}{p-1} \frac{(\sum_{i=1}^n \lambda_i)^{p/(p-1)}}{(pC_p \sum_{i=1}^n \lambda_i^p (1-t_i)^{-(p-1)})^{1/(p-1)}}. \end{aligned} \tag{B.10}$$

We claim that this condition holds for all large n, at least if (t_n) are bounded away from 0, and if ε_n is not too small. Indeed, in this case for some constant

C,
$$\sum_{i=1}^{n} \lambda_i^p (1 + t_i^{-(p-1)}) \le C \sum_{i=1}^{n} \lambda_i^p, \tag{B.11}$$
while
$$\frac{\left(\sum_{i=1}^n \lambda_i\right)^{p/(p-1)}}{\left(\sum_{i=1}^n \lambda_i^p (1-t_i)^{-(p-1)}\right)^{1/(p-1)}} \ge \frac{\left(\sum_{i=1}^n \lambda_i\right)^{p/(p-1)}}{\left(\sum_{i=1}^n \lambda_i^p\right)^{1/(p-1)}}. \tag{B.12}$$
Since the ratio of the expressions in the right-hand sides of (B.12) and (B.11) is
$$\frac{1}{C} \left(\frac{\left(\sum_{i=1}^n \lambda_i\right)}{\sum_{i=1}^n \lambda_i^p} \right)^{1/(p-1)} \to \infty$$
by (4.5), we conclude that (B.10) holds and, hence, the equation (B.6) has real roots, at least for all large n, as long ε_n does not go to zero too fast.

Notice that, if $\varepsilon_n \le 2$, then
$$\begin{aligned} B_n^+(\mu) &= C_p v_p \sum_{i=1}^n \lambda_i^p t_i^{-(p-1)} + \log 2/\varepsilon_n > 0 \\ &> -C_p v_p \sum_{i=1}^n \lambda_i^p - \log 2/\alpha. \end{aligned}$$
Furthermore, the minimum of B_n^+ is achieved to the right of μ. Therefore, under the condition (B.10), the equation (B.6) has one or two real roots to the right of μ, and y_n is the smallest of these roots.

For $x > \mu$ the equation (B.6) becomes
$$\begin{aligned} &C_p (x-\mu)^p \sum_{i=1}^n \lambda_i^p (1-t_i)^{-(p-1)} - (x-\mu) \sum_{i=1}^n \lambda_i \\ &+ C_p v_p \sum_{i=1}^n \lambda_i^p \big(1 + t_i^{-(p-1)}\big) + \log 2/\varepsilon_n + \log 2/\alpha = 0. \end{aligned} \tag{B.13}$$

We can rewrite (B.13) in the form

$$Kz^p - z + M = 0 \tag{B.14}$$

for $z = x - \mu > 0$ and

$$K = \frac{C_p \sum_{i=1}^n \lambda_i^p (1 - t_i)^{-(p-1)}}{\sum_{i=1}^n \lambda_i}$$

and

$$M = \frac{C_p v_p \sum_{i=1}^n \lambda_i^p \big(1 + t_i^{-(p-1)}\big) + \log 2/\varepsilon_n + \log 2/\alpha}{\sum_{i=1}^n \lambda_i}.$$

Setting $y = K^{1/(p-1)} z > 0$ and $D = K^{1/(p-1)} M$ transforms (B.14) into the equation

$$y^p - y + D = 0. \tag{B.15}$$

Let $\tau_n > 0$ and suppose that

$$D \leq \frac{\tau_n^{1/(p-1)}}{(1 + \tau_n)^{p/(p-1)}}. \tag{B.16}$$

Then the equation (B.15) has a positive solution $y(D)$ satisfying

$$y(D) \leq (1 + \tau_n) D,$$

which implies that

$$\begin{aligned} y_n &\leq \mu + (1 + \tau_n) M = \mu + (1 + \tau_n) \\ &\cdot \frac{C_p v_p \sum_{i=1}^n \lambda_i^p \big(1 + t_i^{-(p-1)}\big) + \log 2/\varepsilon_n + \log 2/\alpha}{\sum_{i=1}^n \lambda_i}. \end{aligned} \tag{B.17}$$

Note that the condition (B.16) can be rewritten in the form

$$
\begin{aligned}
&C_p v_p \sum_{i=1}^{n} \lambda_i^p (1 + t_i^{-(p-1)}) + \log 2/\alpha + \log 2/\varepsilon_n \\
&\le \frac{\tau_n^{1/(p-1)}}{(1+\tau_n)^{p/(p-1)}} \frac{(\sum_{i=1}^n \lambda_i)^{p/(p-1)}}{\left(C_p \sum_{i=1}^n \lambda_i^p (1-t_i)^{-(p-1)}\right)^{1/(p-1)}}.
\end{aligned} \tag{B.18}
$$

Similarly to the condition (B.10), this condition holds for all large n as long as ε_n and τ_n do not go to zero too fast. We conclude by (B.7) and (B.17)

$$
\begin{aligned}
&P\Bigg(x_{-,n} < \mu + (1+\tau_n) \frac{C_p v_p \sum_{i=1}^n \lambda_i^p \big(1 + t_i^{-(p-1)}\big) + \log 2/\varepsilon_n + \log 2/\alpha}{\sum_{i=1}^n \lambda_i} \\
&\qquad \text{for all } n \text{ for which (B.18) holds} \Bigg) \ge 1 - \sum_{i=1}^{\infty} \varepsilon_i/2.
\end{aligned} \tag{B.19}
$$

The same argument shows that

$$
\begin{aligned}
&P\Bigg(x_{+,n} > \mu - (1+\tau_n) \frac{C_p v_p \sum_{i=1}^n \lambda_i^p \big(1 + t_i^{-(p-1)}\big) + \log 2/\varepsilon_n + \log 2/\alpha}{\sum_{i=1}^n \lambda_i} \\
&\qquad \text{for all } n \text{ for which (B.18) holds} \Bigg) \ge 1 - \sum_{i=1}^{\infty} \varepsilon_i/2.
\end{aligned} \tag{B.20}
$$

We conclude by (B.19) and (B.20) that

$$
\begin{aligned}
&P\Bigg(\big|I_n(\alpha)\big| \le 2(1+\tau_n) \frac{C_p v_p \sum_{i=1}^n \lambda_i^p \big(1 + t_i^{-(p-1)}\big) + \log 2/\varepsilon_n + \log 2/\alpha}{\sum_{i=1}^n \lambda_i} \\
&\qquad \text{for all } n \text{ for which (B.18) holds} \Bigg) \ge 1 - \sum_{n=1}^{\infty} \varepsilon_n.
\end{aligned} \tag{B.21}
$$

With ε_n as in the statement, (B.21) is transformed into

$$P\Bigg(\big|I_n(\alpha)\big| \le 4(1+\tau_n)\frac{C_p v_p \sum_{i=1}^n \lambda_i^p\big(1+t_i^{-(p-1)}\big)+\log 2/\alpha}{\sum_{i=1}^n \lambda_i} \quad \text{(B.22)}$$

$$\text{for all } n \text{ for which (B.18) holds}\Bigg) \ge 1-\alpha\sum_{n=1}^{\infty}\exp\left\{-C_p v_p\sum_{i=1}^{n}\lambda_i^p\big(1+t_i^{-(p-1)}\big)\right\}.$$

It follows from (4.5) that the sum in the right hand side is finite, and can be made small if α is small.

Proof of Proposition 1. From (2.1), we have that

$$L_n(X_1, \cdots, X_n) := \text{ solution to } \sum_{i=1}^{n} \psi(\lambda_i(X_i - x)) = b_n,$$
$$U_n(X_1, \cdots, X_n) := \text{ solution to } \sum_{i=1}^{n} \psi(\lambda_i(X_i - x)) = a_n.$$

Written in another way, we have

$$\begin{aligned} b_n - a_n &= \sum_{i=1}^{n} \psi(\lambda_i(X_i - L_n)) - \sum_{i=1}^{n} \psi(\lambda_i(X_i - U_n)) \\ &= \sum_{i=1}^{n} \psi(\lambda_i(X_i - L_n)) - \psi(\lambda_i(X_i - U_n)) \end{aligned}$$

Using the fact that the Catoni [2012] influence function $\psi(\cdot)$ is $1-$Lipschitz, we have

$$\begin{aligned} b_n - a_n &\leq \sum_{i=1}^{n} |\psi(\lambda_i(X_i - L_n)) - \psi(\lambda_i(X_i - U_n))| \\ &\leq \sum_{i=1}^{n} |\lambda_i(X_i - L_n) - \lambda_i(X_i - U_n)| \\ &= \sum_{i=1}^{n} |\lambda_i(U_n - L_n)|. \end{aligned}$$

It follows that

$$W_n := |U_n - L_n| \geq \frac{b_n - a_n}{\sum_{i=1}^{n} \lambda_i}.$$

$\square$

Proof of Theorem 3:

We make use of the general law of iterated logarithm by Wittmann (1985) to support key arguments.

Let $s_n^2 = \sum_{i=1}^n \texttt{Var}(Y_i)$ for $n = 1, 2, \cdots$. We will first show that $\texttt{Var}(Y_i) \sim \lambda_i^2 \sigma^2$ as $i \to \infty$. Indeed,

$$\begin{aligned}
\mathbb{E}Y_i &= \mathbb{E}\psi(\lambda_i(X-\mu)) \\
&\leq \mathbb{E}[\log(1+\lambda_i(X-\mu)) + \lambda_i^2(X-\mu)^2/2] \\
&\leq \mathbb{E}[\lambda_i(X-\mu) + \lambda_i^2(X-\mu)^2/2] \\
&= \lambda_i^2\sigma^2/2.
\end{aligned}$$

We also have

$$\begin{aligned}
\mathbb{E}Y_i &\geq \mathbb{E}[-\log(1-\lambda_i(X-\mu) + \lambda_i^2(X-\mu)^2)/2] \\
&\geq \mathbb{E}[-(-\lambda_i(X-\mu) + \lambda_i^2(X-\mu)^2/2)] \\
&= -\lambda_i^2\sigma^2/2.
\end{aligned}$$

Therefore, $|\mathbb{E}Y_i| \leq \lambda_i^2\sigma^2/2$ for $i = 1, 2, \cdots$. Next,

$$\begin{aligned}
&\mathbb{E}Y_i^2 = \mathbb{E}\psi^2(\lambda_i(X-\mu)) \\
&\leq \mathbb{E}\Big\{[\log(1+\lambda_i(X-\mu) + \lambda_i^2(X-\mu)^2)/2]^2 \mathbb{1}(X \geq \mu)\Big\} \\
&+ \mathbb{E}\Big\{[\log(1-\lambda_i(X-\mu) + \lambda_i^2(X-\mu)^2)/2]^2 \mathbb{1}(X < \mu)\Big\}.
\end{aligned}$$

There is $x_0 > 0$ such that $\log(1+x) \le x^{1/2}$ for $x \ge x_0$. We have

$$\mathbb{E}\Big\{[\log(1+\lambda_i(X-\mu)+\lambda_i^2(X-\mu)^2)/2]^2 \mathbb{1}(X \ge \mu)\Big\}$$

$$\le \mathbb{E}\Big\{[\lambda_i(X-\mu)+\lambda_i^2(X-\mu)^2/2]^2 \mathbb{1}(\mu \le X \le \mu + x_0/\lambda_i\Big\}$$

$$+ \mathbb{E}\Big\{[\lambda_i(X-\mu)+\lambda_i^2(X-\mu)^2/2]^2 \mathbb{1}(X > \mu + x_0/\lambda_i)\Big\}$$

$$= \lambda_i^2 \mathbb{E}[(X-\mu)^2 \mathbb{1}(\mu \le X \le \mu + x_0/\lambda_i)] + o(\lambda_i^2)$$

$$= \lambda_i^2 \mathbb{E}[(X-\mu)^2 \mathbb{1}(X \ge \mu)] + o(\lambda_i^2). \tag{B.23}$$

Similarly,

$$\mathbb{E}\Big\{[\log(1-\lambda_i(X-\mu)+\lambda_i^2(X-\mu)^2)/2]^2 \mathbb{1}(X < \mu)\Big\}$$

$$= \lambda_i^2 \mathbb{E}[(X-\mu)^2 \mathbb{1}(X < \mu)] + o(\lambda_i^2). \tag{B.24}$$

From (B.23) and (B.24), we have that $\mathbb{E}Y_i^2 \le \lambda_i^2\sigma^2 + o(\lambda_i^2)$. On the other hand,

$$\mathbb{E}Y_i^2 \ge \mathbb{E}\Big\{[\log(1-\lambda_i(X-\mu)+\lambda_i^2(X-\mu)^2)/2]^2$$

$$\mathbb{1}(\mu \le X \le \mu + x_0/\lambda_i)\Big\}$$

$$+ \mathbb{E}\Big\{[\log(1+\lambda_i(X-\mu)+\lambda_i^2(X-\mu)^2)/2]^2$$

$$\mathbb{1}(\mu - x_0/\lambda_i \le X < \mu)\Big\}.$$

On the other hand, for every $\epsilon > 0$ there is $0 < x_1 < 1$ such that

$|\log(1+x)| \geq (1-\epsilon)|x|$ for all $|x| \leq x_1$. Therefore,

$$
\begin{aligned}
\mathbb{E}Y_i^2 \geq \mathbb{E}\Big\{&[\log(1-\lambda_i(X-\mu)+\lambda_i^2(X-\mu)^2)/2]^2 \\
&\mathbb{1}(\mu \leq X \leq \mu + x_1/\lambda_i)\Big\} \\
&+\mathbb{E}\Big\{[\log(1+\lambda_i(X-\mu)+\lambda_i^2(X-\mu)^2)/2]^2 \\
&\mathbb{1}(\mu - x_1/\lambda_i \leq X < \mu)\Big\} \\
&\geq (1-\epsilon)^2\lambda_i^2\mathbb{E}[(X-\mu)^2\mathbb{1}(|X-\mu| \leq x_1/\lambda_i)] \\
&= (1-\epsilon)^2\lambda_i^2\sigma_2 + o(\lambda_i^2).
\end{aligned}
$$

Since ϵ can be taken as small as we wish, then it holds $\mathbb{E}Y_i^2 \geq \lambda_i^2\sigma^2 + o(\lambda_i^2)$.

Therefore, we conclude that

$$\mathbb{E}Y_i^2 = \lambda_i^2\sigma^2 + o(\lambda_i^2).$$

It follows from the above arguments that

$$
\begin{aligned}
&s_n^2 \sim \sigma^2 \sum_{i=1}^n \lambda_i^2, \text{ and} \\
&\theta_n := (s_n 2 \log\log s_n^2)^{1/2} \sim \sigma\Big(2\sum_{i=1}^n \lambda_i^2 \log\log \sum_{i=1}^n \lambda_i^2\Big)^{1/2}.
\end{aligned}
$$

We verify that the condition $(2, \alpha)$ in Wittmann (1985) holds for sequence θ_n and ϑ. Denoting by c a generic positive constant that may change from time to time, we have for large n_0,

$$\begin{aligned}
&\sum_{n=n_0}^{\infty} \theta_n^{-(2+\vartheta)} \mathbb{E}|Y_i - \mathbb{E}Y_i|^{2+\vartheta} \\
\leq \;\; & c \sum_{n=n_0}^{\infty} \left(\sum_{i=1}^{n} \lambda_i^2 \log\log \sum_{i=1}^{n} \lambda_i^2 \right)^{-1-\vartheta/2} \mathbb{E}|Y_i|^{2+\vartheta}.
\end{aligned}$$

Using $\log(1+x+x^2/2) \leq 2\log(1+x/\sqrt{2})$ and $\log(1-x+x^2/2) \geq 2\log(1-x/\sqrt{2})$, we have

$$\begin{aligned}
\mathbb{E}|Y_i|^{2+\vartheta} &\leq \mathbb{E}[[2\log(1+\lambda_i(X-\mu)/\sqrt{2})]^{2+\vartheta} \mathbb{1}(X > \mu)] \\
&\quad + \mathbb{E}[[2\log(1-\lambda_i(X-\mu)/\sqrt{2})]^{2+\vartheta} \mathbb{1}(X < \mu)] \\
&\leq c\lambda_i^{2+\vartheta}. \qquad \text{(B.25)}
\end{aligned}$$

Therefore, as (λ_n) is non-increasing,

$$\begin{aligned}
&\sum_{n=n_0}^{\infty} \theta_n^{-(2+\vartheta)} \mathbb{E}|Y_i - \mathbb{E}Y_i|^{2+\vartheta} \\
\leq \;\; & c \sum_{n=n_0}^{\infty} \left(\sum_{i=1}^{n} \lambda_i^2 \right)^{-1-\vartheta/2} \lambda_i^{2+\vartheta} \leq c \sum_{n=n_0}^{\infty} \frac{1}{n^{1+\vartheta/2}} < \infty.
\end{aligned}$$

Hence the condition $(2, \alpha)$ in Wittmann (1985) holds. It follows that

$$\limsup_{n\to\infty} \frac{s_{n+1}}{s_n} = \limsup_{n\to\infty} \left(\frac{\sum_{i=1}^{n+1} \lambda_i^2}{\sum_{i=1}^{n} \lambda_i^2} \right)^{1/2} \leq \limsup_{n\to\infty} \left(\frac{n+1}{n} \right)^{1/2} = 1.$$

Therefore, by (Wittmann, 1985, Theorem 2.1)

$$\limsup_{n\to\infty} \theta_n^{-1} \sum_{i=1}^{n} (Y_i - \mathbb{E}Y_i) = 1, \text{ a.s.}$$

That is,

$$\limsup_{n\to\infty} \frac{\sum_{i=1}^{n}(Y_i - \mathbb{E}Y_i)}{\sigma\Big(2\sum_{i=1}^{n}\lambda_i^2 \log\log \sum_{i=1}^{n}\lambda_i^2\Big)^{1/2}} = 1, \text{ a.s.}$$

This means that the sequence $(\widetilde{a}_n)$ and $(\widetilde{b}_n)$ must satisfy

$$\widetilde{b}_n - \widetilde{a}_n > a\Big(2\sum_{i=1}^{n}\lambda_i^2 \log\log \sum_{i=1}^{n}\lambda_i^2\Big)^{1/2}$$

for any $a < 2\sigma\sqrt{2}$ and large n, which implies the same for $b_n - a_n$ and the result holds.

Here (B.25) relies on the existence of $\mathbb{E}[|X_i|^{2+\vartheta}]$. In fact, we establish an upper bound of $\mathbb{E}|Y_i|^{2+\vartheta}$ without assuming the $(2+\vartheta)$-moment of X_i.

To see this, we need to make use of the following lemma.

Lemma 1. *There exists a constant C_ϑ such that*

$$\log(1 + |x| + \frac{1}{2}x^2) \le C_\vartheta |x|^{\frac{2}{2+\vartheta}}. \tag{B.26}$$

Based on Lemma 1, we have

$$\mathbb{E}|Y_i|^{2+\vartheta} \le \mathbb{E}[\big(C_\vartheta |\lambda_i X_i|^{\frac{2}{2+\vartheta}}\big)^{2+\vartheta}] = C_\vartheta \lambda_i^2. \tag{B.27}$$

Hence, it holds

$$\begin{aligned} & \sum_{n=n_0}^{\infty} \theta_n^{-(2+\vartheta)} \mathbb{E}|Y_i - \mathbb{E}Y_i|^{2+\vartheta} \\ \le \;& c \sum_{n=n_0}^{\infty} \Big(\sum_{i=1}^{n}\lambda_i^2\Big)^{-1-\vartheta/2} C_\vartheta \lambda_n^2 < \infty. \end{aligned}$$

by assumption. Remark: the above inequality holds automatically when $\lambda_i = i^{-1/2}$ for any $i \geq 1$.

Proof of Theorem 4: The proof idea is very similar to that of Theorem 3. We first derive the upper and lower bounds of $\mathrm{Var}(Y_i)$ using the following lemma.

Lemma 2. *There exists constants C_1, C_2 such that*

$$\log(1 + |x| + C_p|x|^p) \leq C_1|x|^{p/2} \tag{B.28}$$

and

$$\log(1 + |x| + C_p|x|^p) \leq C_2|x|^{p/(2+\vartheta)} \tag{B.29}$$

Lemma 3. *There exist positive constants C_3, x_1, x_2 such that*

$$\log(1 + x + C_p|x|^p) \geq C_3|x|^{p/2}, \tag{B.30}$$

for any $x_1 \leq x \leq x_2$.

By Lemma 2, we have

$$\begin{aligned} \mathrm{Var}(Y_i) &\leq \mathbb{E}[Y_i^2] \\ &\leq \mathbb{E}[(C_1|\lambda_i X_i|^{p/2})^2] \\ &= C_1^2 \lambda_i^p v. \end{aligned} \tag{B.31}$$

By Lemma 3 and symmetry of X_i, we have

$$\begin{aligned}
\mathrm{Var}(Y_i) &= \mathbb{E}[Y_i^2] \\
&\geq \mathbb{E}[(C_3|\lambda_i X_i|^{p/2})^2 \mathbf{1}\{x_1 \leq |\lambda_i X_i| \leq x_2\}] \\
&= C_3^2 \lambda_i^p \mathbb{E}[(C_3|X_i|^{p/2})^2 \mathbf{1}\{x_1 \leq |\lambda_i X_i| \leq x_2\}] \\
&= C_3^2 \lambda_i^p \int_{\frac{x_1}{\lambda_i}}^{\frac{x_2}{\lambda_i}} x^p f(x) dx \\
&\geq C_3^2 C_{\vartheta'} \lambda_i^{p+\vartheta'} \qquad (B.32)
\end{aligned}$$

for any $\vartheta' > 0$ and $f(x) \propto x^{-(1+p+\vartheta')}$.

Again, by Lemma 2, we have

$$\begin{aligned}
\mathbb{E}[|Y_i|^{2+\vartheta}] &\leq C_2^{2+\vartheta} \mathbb{E}[(|\lambda_i X_i|^{p/(2+\vartheta)})^{2+\vartheta}] \\
&\leq C_2^{2+\vartheta} \lambda_i^p. \qquad (B.33)
\end{aligned}$$

Therefore,

$$\begin{aligned}
s_n &= \Theta(\sum_{i=1}^{n} \lambda_i^{p+\vartheta'}), \qquad (B.34) \\
\theta_n &= (s_n \log\log s_n^2)^{1/2}. \qquad (B.35)
\end{aligned}$$

Hence, it holds

$$\begin{aligned}
& \sum_{n=n_0}^{\infty} \theta_n^{-(2+\vartheta)} \mathbb{E}|Y_i - \mathbb{E}Y_i|^{2+\vartheta} \\
\leq\ & c \sum_{n=n_0}^{\infty} \Big(\sum_{i=1}^{n} \lambda_i^{p+\vartheta'} \Big)^{-1-\vartheta/2} \lambda_n^p < \infty.
\end{aligned}$$

Using arguments as in the proof of Theorem 3, we have

$$\tilde{b}_n - \tilde{a}_n \geq a'(\sum_{i=1}^{n} \lambda_i^{p+\vartheta'} \log\log \sum_{i=1}^{n} \lambda_i^{p+\vartheta'})^{1/2}$$

by adjusting constant a'.

Proof of Theorem 5.

For any t between a^{j-1} and a^j, we plug $\lambda_i^{(t)} \equiv \Lambda_j$ and α_j to inequality (6.15). Therefore, we get that

$$\begin{aligned}
& C_p v_p \sum_{i=1}^{t} \lambda_i^{(t)p} + \log(2/\alpha_j) \\
= \quad & C_p \sum_{i=1}^{t} \log(2/\alpha_j) a^{-j} + \log(2/\alpha_j) \\
= \quad & C_p \log(2/\alpha_j) t a^{-j} + \log(2/\alpha_j) \\
\leq \quad & (C_p + 1) \log(2/\alpha_j) \\
\leq \quad & (C_p + 1) \log(2 j^q \sum_{l=1}^{\infty} l^{-q}/\alpha) \\
\leq \quad & (C_p + 1) q \log(2 j \sum_{l=1}^{\infty} l^{-q}/\alpha) \\
\leq \quad & (C_p + 1) q \log(2(\log_a t + 1) \sum_{l=1}^{\infty} l^{-q}/\alpha).
\end{aligned} \tag{B.36}$$

Therefore, we use simplified notation as

$$\begin{aligned}
& L_t := \text{root of } \sum_{i=1}^{t} \psi(\lambda_i^{(t)}(X_i - x)) = b_t, \\
& \quad \text{with } \; b_t = (C_p + 1) q \log(2(\log_a t + 1) \sum_{l=1}^{\infty} l^{-q}/\alpha), \\
& U_t := \text{root of } \sum_{i=1}^{t} \psi(\lambda_i^{(t)}(X_i - x)) = a_t \\
& \quad \text{with } \; a_t = -(C_p + 1) q \log(2(\log_a t + 1) \sum_{l=1}^{\infty} l^{-q}/\alpha).
\end{aligned} \tag{B.37}$$

Next, we will show that there exists a sequence of constants c_t's such that

$$\begin{aligned} & c_t(\sum_{i=1}^{t} \lambda_i^{(t)})(U_t - L_t) \\ \leq \;& 2(C_p+1)q\log(2(\log_a t + 1)\sum_{l=1}^{\infty} l^{-q}/\alpha) \end{aligned} \tag{B.38}$$

holds with high probability and $c_t \to 1$. This then leads to the desired result since

$$\begin{aligned} |I_t| = U_t - L_t &\leq \frac{2(C_p+1)q\log(2(\log_a t+1)\sum_{l=1}^{\infty} l^{-q}/\alpha)}{c_t(\sum_{i=1}^{t}\lambda_i^{(t)})} \\ &\leq \frac{2v_p^{1/p}\big((C_p+1)q\log(2(\log_a t+1)\sum_{l=1}^{\infty} l^{-q}/\alpha)\big)^{1-1/p}}{c_t \cdot t(1/at)^{1/p}} \end{aligned} \tag{B.39}$$

By noticing (B.37), we know

$$\begin{aligned} & 2(C_p+1)q\log(2(\log_a t+1)\sum_{l=1}^{\infty} l^{-q}/\alpha) \\ = \;& \sum_{i=1}^{t}\psi(\lambda_i^{(t)}(X_i - L_t)) - \sum_{i=1}^{t}\psi(\lambda_i^{(t)}(X_i - U_t)) \\ = \;& \sum_{i=1}^{t}\int_{L_t}^{U_t}\lambda_i^{(t)}\psi'(\lambda_i^{(t)}(X_i - x))dx. \end{aligned} \tag{B.40}$$

Then it is equivalent to show that

$$\sum_{i=1}^{t}\lambda_i^{(t)}\psi'(\lambda_i^{(t)}(X_i - x)) \geq c_t\sum_{i=1}^{t}\lambda_i^{(t)} \tag{B.41}$$

holds with high probability for all t.

We choose $c_{1t} = \inf_{|x|\leq B_t}\psi'(x)$, where $0 < B_t < 1$ is a constant that may depend on time index t. Hence, we only need to show that

$$\sum_{i=1}^{t} \mathbf{1}\{|X_i| \leq B_t/\lambda_i^{(t)}\} \geq c_{2t}t$$

for some constant c_{2t}. We then can easily take $c_t = c_{1t}c_{2t}$ to conclude the proof.

To show this, we need to make use of the following concentration inequality for Bernoulli random variables.

Lemma 4.

$$\mathbb{P}(Z \leq \mathbb{E}[Z] - x) \leq \exp\{-x^2/2\mathbb{E}[Z]\}, \tag{B.42}$$

where $Z = \sum Z_i$ such that $Z_i \sim Bern(p)$.

In our case, for any fixed λ, we let $Z_i = \mathbf{1}\{|X_i| \leq B_t/\lambda_i^{(t)}\}$ and $p_t := \mathbb{E}[Z_i] \geq 1 - B_t^{-p} \cdot \log(\frac{2}{\alpha_j}) \cdot a^{-j}$ with $j = \lceil \log_a t \rceil$ by Markov inequality. Therefore, by Lemma 4 and taking $x = 2\sqrt{p_t \cdot t \cdot \log(\frac{t}{\alpha_0})}$ (where $0 < \alpha_0 < 1$ is a parameter that can be tuned), we have

$$\mathbb{P}(Z \leq \mathbb{E}[Z] - x) \leq \exp\{-x^2/2\mathbb{E}[Z]\} = \exp\{-4p_t t \cdot \log(\frac{t}{\alpha_0})/2p_t t\} = \frac{\alpha_0^2}{t^2}.$$

In other words, we have

$$\mathbb{P}(\exists t \geq 1; \sum_{i=1}^{t} \mathbf{1}\{|X_i| \leq B_t/\lambda_i^{(t)}\} \leq p_t \cdot t - 2\sqrt{t \cdot \log(\frac{t}{\alpha_0})}) \leq \alpha_0^2 \sum_{t=1} \frac{1}{t^2} \leq 2\alpha_0^2.$$

In a summary, with probability at least $1-\alpha_0^2$, we have

$$\sum_{i=1}^{t} \mathbf{1}\{|X_i| \le B_t/\lambda_i^{(t)}\} \ge c_{2t} t,$$

where $c_{2t} = p_t - 2\sqrt{\frac{\log(t/\alpha_0)}{t}}$. By straightforward calculations, we can find

$$\begin{aligned} c_{2t} &= p_t - 2\sqrt{\frac{\log(t/\alpha_0)}{t}} \\ &\ge 1 - B_t^{-p} \cdot \log(\frac{2}{\alpha_j}) \cdot a^{-j} - 2\sqrt{\frac{\log(t/\alpha_0)}{t}} \\ &\ge 1 - B_t^{-p} \cdot q\log(\frac{4C_q \log t}{\alpha}) \cdot \frac{a}{t} - 2\sqrt{\frac{\log(t/\alpha_0)}{t}}. \qquad \text{(B.43)} \\ &\quad (\text{since } \log(\frac{2}{\alpha_j}) \le q\log(\frac{4C_q \log t}{\alpha}) \text{and } a^{-j} \le a/t, \text{where } C_q = \sum_l l^{-q}) \end{aligned}$$

By taking $B_t = t^{-1/2p}$, we then have

$$c_{2t} \ge 1 - q\log(\frac{4C_q \log t}{\alpha}) \cdot \frac{a}{\sqrt{t}} - 2\sqrt{\frac{\log(t/\alpha_0)}{t}} \qquad \text{(B.44)}$$

and $c_{1t} = \inf_{|x|\le t^{-1/2p}} \psi'(x)$. Therefore, we have $c_t \to 1$ as both $c_{1t}, c_{2t} \to 1$ as $t \to \infty$. Then we can make α_0 arbitrary close to zero to conclude the proof.

Proof of Theorem 7. Note that $R(\beta)$ is a decreasing function of β so that the event $\{\exists\ n \text{ s.t. } R(\hat{\beta}_n) > r^*\}$ implies the event $\{\exists\ n \text{ s.t. } \hat{\beta}_n < \beta^*\}$ which further implies $\{\exists\ n \text{ s.t. } M_n(\beta^*) \ge 1/\alpha\}$ by definition (7.19). Using the fact that $M_n(\beta^*)$ is a non-negative supermartingale, we get

$$
\begin{aligned}
& \mathbb{P}(\exists\ n \text{ s.t. } R(\hat{\beta}_n) > r^*) \\
\leq\ & \mathbb{P}(\exists\ n \text{ s.t. } M_n(\beta^*) \geq 1/\alpha) \\
\leq\ & 1-\alpha
\end{aligned} \tag{B.45}
$$

by Ville's inequality. This concludes the proof.

Proof of Theorem 8. Recall that

$$
\delta_t := 2\frac{\left((C_p v_p + 1) q \log(2(\log_a t + 1)\sum_{l=1}^{\infty} l^{-q}/\alpha')\right)^{1-1/p}}{t(1/at)^{1/p}},
$$

with $\alpha' := \alpha/(d^2+d)$. By Theorem 5, we know that

$$
\mathbb{P}(|\widehat{XX^T}[j_1,j_2] - \mathbb{E}[XX^T][j_1,j_2]| \leq \delta_t,\ \forall\ t \geq t_0) \geq 1-\alpha' \tag{B.46}
$$

and

$$
\mathbb{P}(|\widehat{XY}[j] - \mathbb{E}[XY][j]| \leq \delta_t,\ \forall\ t \geq t_0) \geq 1-\alpha' \tag{B.47}
$$

for all $j \in [d]$ and $j_1, j_2 \in [d]$.

For any $\boldsymbol{\beta}$, we define

$$
\Delta(\boldsymbol{\beta}) := \hat{l}_n(\boldsymbol{\beta}) - \hat{l}_n(\boldsymbol{\beta}^*) - (l(\boldsymbol{\beta}) - l(\boldsymbol{\beta}^*)). \tag{B.48}
$$

By the optimality of $\hat{\boldsymbol{\beta}}_n$, we know

$$
\Delta(\hat{\boldsymbol{\beta}}_n) \leq -(l(\boldsymbol{\beta}) - l(\boldsymbol{\beta}^*)) \leq -\lambda_{min}\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*\|^2, \tag{B.49}
$$

where λ_{min} is the smallest eigenvalue of $\mathbb{E}[XX^T]$.

On the other hand, by the straightfoward calcuations, we have

$$\begin{aligned} & \Delta(\hat{\boldsymbol{\beta}}_n) \\ = \ & 2(\boldsymbol{\beta}^{*T}\widehat{XX^T} - \widehat{XY}^T)(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*) \\ & + (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*)^T(\widehat{XX^T} - \mathbb{E}[XX^T])(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*). \end{aligned} \tag{B.50}$$

By (B.46) - (B.47), it holds

$$\begin{aligned} & \|\boldsymbol{\beta}^{*T}\widehat{XX^T} - \widehat{XY}^T\| \\ \leq \ & \|\boldsymbol{\beta}^{*T}\widehat{XX^T} - \boldsymbol{\beta}^{*T}\mathbb{E}[XX^T]\| + \|\mathbb{E}[YX^T] - \widehat{XY}^T\| \\ \leq \ & d\|\boldsymbol{\beta}^*\|\delta_n + \sqrt{d}\delta_n \end{aligned} \tag{B.51}$$

Then we have

$$\begin{aligned} & \Delta(\hat{\boldsymbol{\beta}}_n) \\ \geq \ & -2(d\|\boldsymbol{\beta}^*\| + \sqrt{d})\delta_n\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*\| - d\delta_n\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*\|^2. \end{aligned} \tag{B.52}$$

Combining (B.49) and (B.49), we have

$$(\lambda_{min} - d\delta_n)\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*\|^2 \leq 2(d\|\boldsymbol{\beta}^*\| + \sqrt{d})\delta_n\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*\|, \tag{B.53}$$

which gives

$$\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*\| \leq 4(d\|\boldsymbol{\beta}^*\| + \sqrt{d})\delta_n/\lambda_{min},$$

for n satisfying $\delta_n \leq \lambda_{min}/2d$. This concludes the proof.

C. Proof of Supporting Lemma

Proof of Lemma 2 and 3.

In order to prove Lemma 2 and 3, we only need to show the following proposition.

Proposition 1. *For any $p \in (1,2]$ and $q \in (0,1]$, there exists a constant $C_{p,q}$ such that*

$$\log(1+|x|+C_p|x|^p) \leq C_{p,q}|x|^q. \tag{C.54}$$

If we can prove Proposition 1, then Lemma 2 holds with $p=2$ and $q=2/(2+\vartheta)$; Lemma 3 holds with $q=p/2$ or $q=p/(2+\vartheta)$.

Proof of Proposition 1. If suffices to show that

$$\sup_{x\in\mathbb{R}} \frac{\log(1+|x|+C_p|x|^p)}{|x|^q} \leq C_{p,q}, \tag{C.55}$$

which is equivalent to show

$$1+|x|+C_p|x|^p \leq \exp\{C_{p,q}|x|^q\} \tag{C.56}$$

for any x. Take $k_0 = \lceil p/q \rceil$, we only need to show that

$$1+|x|+C_p|x|^p \leq 1+\sum_{k=1}^{k_0} \frac{C_{p,q}^k}{k!}|x|^{qk}. \tag{C.57}$$

We take $C_{p,q}$ large enough such that

$$\frac{C_{p,q}^k}{k!} \geq 1+C_p, \text{ for } 1 \leq k \leq k_0. \tag{C.58}$$

Therefore, for any $0 \le |x| < 1$, we know

$$\begin{aligned}
1+|x|+C_p|x|^p &\le 1+(1+C_p)|x| \\
&\le 1+C_{p,q}|x|^q \\
&\le 1+\sum_{k=1}^{k_0}\frac{C_{p,q}^k}{k!}|x|^{qk} \\
&\le \exp\{C_{p,q}|x|^q\}. \qquad \text{(C.59)}
\end{aligned}$$

For $|x| > 1$, we have

$$\begin{aligned}
1+|x|+C_p|x|^p &\le 1+(1+C_p)|x|^p \\
&\le 1+\frac{C_{p,q}^{k_0}}{k_0!}|x|^{qk_0} \\
&\le 1+\sum_{k=1}^{k_0}\frac{C_{p,q}^k}{k!}|x|^{qk} \\
&\le \exp\{C_{p,q}|x|^q\}. \qquad \text{(C.60)}
\end{aligned}$$

Therefore, Proposition 1 holds and we conclude the proof.

Proof of Lemma 4. For fixed $x_1, x_2 > 0$, we write

$$F(x) = \frac{|x|^{p/2}}{\log(1+x+C_p|x|^p)}.$$

It is easy to check that $F(x)$ is a continuously differentiable function on interval $[x_1, x_2]$. Moreover, on compact set $[x_1, x_2]$, $\sup_{x\in[x_1,x_2]} F(x)$ exists and we denote it as $1/C_3$. Hence, it holds that $\log(1+x+C_p|x|^p) \ge C_3|x|^{p/2}$.

Proof of Lemma 5. See Theorem 4 in Chung and Lu (2006).

D. Proof of the Lower Bound of Stitching Method

Proof of Theorem 6. The idea of the proof is based on that of Theorem 1 in Tomkins (1974) with some modifications. We denote $Y_{n,i} := \psi(\lambda_i^{(n)} X_i) - \mathbb{E}[\psi(\lambda_i^{(n)} X_i)]$. Therefore, $Y_{n,i}$ is mean zero and $Y_{n,1}, Y_{n,2}, ..., Y_{n,n}$ are independent. We write $S_n = \sum_{i=1}^n Y_{n,i}$, $s_n = \mathbb{E}[S_n^2]$, and $t_n^2 = 2\log\log(s_n^2)$. We consider choosing a sequence of n_k's with n_k being the least integer such that $s_{n_k} \geq a^k$ where $a > 1$. This is possible when $n \cdot \lambda^{(n)} \to \infty$. We also define $V_k = \sum_{i=n_k+1}^{n_k} Y_{n,i}$ and $U_k = S_{n_k} - V_k$. Moreover, we denote $u_k^2 = \mathbb{E}[U_k^2]$, $v_k^2 = \mathbb{E}[V_k^2]$, $A_k = (2u_k^2 \log\log(u_k^2))^{1/2}$ and $B_k = (2v_k^2 \log\log(v_k^2))^{1/2}$. By straightforward calculation, we have

$$\frac{u_k^2}{s_{n_k}^2} \sim \frac{n_{k-1}}{n_k} \sim \frac{1}{a^{1-p/q}} \in (0,1)$$

and sequence $\{s_n\}$ is increasing and s_{n+1}/s_n is bounded by a. Therefore, conditions (c) and (d) of Theorem 1 in Tomkins (1974) are satisfied.

In the remaining, we only need to establish

$$\mathbb{P}(U_k \geq xA_k) \leq c_1(\log A_k)^{-x} \tag{D.61}$$

with some constants c_1 and $x > 1$, and

$$\mathbb{P}(V_k \geq xB_k) \geq c_2(\log B_k)^{-x} \tag{D.62}$$

with some constants c_2 and $x < 1$.

To prove (D.61) and (D.62), we make use of Berry–Esseen theorem. In our situation, when $1 < p < 2$, we know that $\mathbb{E}[|Y_{n,i}|^3] \le C'\lambda^{(n)p}$ by Lemma 2 by taking $v = 1$ and $\mathbb{E}[|Y_{n,i}|^2] \ge C(\lambda_i^{(n)})^{p+\vartheta'}$ by (B.32) and adjusting the constant C'. (For $p = 2$, ϑ' can be taken as 0.) Therefore, we have

$$|F_{U,k}(x) - \Phi(x)| \le C'' \frac{(\lambda^{(n)})^{-\frac{p+3\vartheta'}{2}}}{\sqrt{n_k}}, \tag{D.63}$$

where $F_{U,k}(x)$ is the cumulative distribution function (C.D.F.) of $\frac{U_k}{u_k}$ and $\Phi(x)$ is C.D.F. of the standard normal distribution. Similarly, we have

$$|F_{V,k}(x) - \Phi(x)| \le C'' \frac{(\lambda^{(n)})^{-\frac{p+3\vartheta'}{2}}}{\sqrt{n_k - n_{k-1}}}, \tag{D.64}$$

where $F_{V,k}(x)$ is the cumulative distribution function (C.D.F.) of $\frac{V_k}{v_k}$. Therefore, by (D.63), we have

$$\begin{aligned}
\mathbb{P}(U_k \ge xA_k) &\le 1 - \Phi((2x^2 \log\log(u_k))^{1/2}) + C'' \frac{(\lambda^{(n)})^{-\frac{p+3\vartheta'}{2}}}{\sqrt{n_k}} \\
&\le K\Big(\log(u_k^2)^{-x} + \frac{(\lambda^{(n)})^{-\frac{p+3\vartheta'}{2}}}{\sqrt{n_k}}\Big) \qquad \text{(D.65)} \\
&\le 2K\log(u_k^2)^{-x}, \qquad \text{(D.66)}
\end{aligned}$$

where the last inequality uses the fact that the term $\frac{(\lambda^{(n)})^{-\frac{p+3\vartheta'}{2}}}{\sqrt{n_k}}$ decreases the polynomially fast as $n_k \to \infty$ when $\vartheta' < (q-p)/3$, while the term $\log(u_k^2)^{-x}$ is of order $\log(n_k)^{-x}$ and it is the dominating term.

Similarly, we have

$$\begin{aligned}
\mathbb{P}(V_k \geq xB_k) &\geq 1 - \Phi((2x^2 \log\log(v_k))^{1/2}) - C'' \frac{(\lambda^{(n)})^{-\frac{p+3\vartheta'}{2}}}{\sqrt{n_k - n_{k-1}}} \\
&\geq K\Big(\log(v_k^2)^{-x} - \frac{(\lambda^{(n)})^{-\frac{p+3\vartheta'}{2}}}{\sqrt{n_k - n_{k-1}}}\Big) \qquad \text{(D.67)} \\
&\geq \frac{1}{2} K \log(v_k^2)^{-x}. \qquad \text{(D.68)}
\end{aligned}$$

Therefore, we have proved (D.63) and (D.64). Lastly, the proof is done by invoking Lemma 4 in Tomkins (1974). □

Remark 1. For $p = 2$, from Theorem 6, we have that the lower bound is of order $O(\sqrt{\log\log n/n})$ for the stitching method. When $1 < p < 2$, with the choice of $\lambda_i^{(n)} = n^{-1/q}$ and $q = p + \vartheta'$, the lower bound give the order of $\frac{(\log\log n)^{1/2}}{n^{(1-\frac{1+\vartheta'}{p+2\vartheta'})}}$. Therefore, a small gap exists between the lower bound and upper bound in the infinite variance case. It remains an open question whether we could further improve the lower bound or upper bound.

References

Chen, P., X. Jin, X. Li, and L. Xu (2021). A generalized catoni's m-estimator under finite α-th moment assumption with $\alpha \in (1, 2)$. *Electronic Journal of Statistics 15*(2), 5523–5544.

Chung, F. and L. Lu (2006). Concentration inequalities and martingale inequalities: a survey. *Internet mathematics 3*(1), 79–127.

Tomkins, R. (1974). On the law of the iterated logarithm for double sequences of random variables. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 30*(4), 303–314.

Wang, H. and A. Ramdas (2023). Catoni-style confidence sequences for heavy-tailed mean estimation. *Stochastic Processes and Their Applications 163*, 168–202.

Waudby-Smith, I. and A. Ramdas (2024). Estimating means of bounded random variables by betting. *Journal of the Royal Statistical Society Series B: Statistical Methodology 86*(1), 1–27.

Wittmann, R. (1985). A general law of iterated logarithm. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 68*(4), 521–543.