Appendix

A MORE DETAILS IN THE INTRODUCTION PART

A.1 WRONG USAGE OF HOEFFDING'S INEQUALITY

This part mainly discusses some results when we use Hoeffding's inequality wrongly for the unbounded data $\{X_i\}_{i=1}^n$.

Lemma 3. *If the Gaussian data $\{X_i\}_{i=1}^n \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$ are misspecified as bounded variable (a bound as a function of n) with high probability, and Hoeffding's inequality is wrongly adopted for the unbounded Gaussian data, it gives*

$$\mathrm{P}\left(\mu \in [\bar{X}_n \pm 2\sigma\sqrt{n^{-1}\log(4/\alpha)}[\sqrt{\log(4/\alpha)} + \sqrt{\log n}]\right) \geq 1 - \alpha.$$

Lemma 3 gives a loose CI, since it contains a $\sqrt{\log n}$ factor. In the remark after the proof, we extend Lemma 3 to some unbounded random variables with strongly log-concave distributions and finite sub-exponential norms $\|X_1\|_{w_1} < \infty$.

Proof of Lemma 3. The Borell–TIS inequality Giné & Nickl (2016) gives the probability of a deviation of the maximum of a centered Gaussian random variables (or stochastic processes) above from its expected value. WLOG, we assume $\{X_i\}_{i=1}^n \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$ are misspecified as bounded variable, by Borell-TIS inequality

$$\mathrm{P}(\max_{i\in[n]} X_i - \mathrm{E}[\max_{i\in[n]} X_i] > t) \leq e^{-t^2/(2\sigma^2)},$$

we have with probability at least $1 - \alpha/4$

$$\max_{i\in[n]} X_i \leq \sqrt{2\sigma^2\log(4/\alpha)} + \mathrm{E}[\max_{i\in[n]} X_i] \leq \sigma\sqrt{2}[\sqrt{\log(4/\alpha)} + \sqrt{\log n}],\ i = 1, 2, \cdots, n,$$

where we use the maximal inequality $\mathrm{E}[\max_{1\leq i\leq n} X_i] \leq \sigma\sqrt{2\log n}$ in Rigollet & Hütter (2019). Then,

$$\mathrm{P}(\max_{i\in[n]} X_i \geq \sigma\sqrt{2}[\sqrt{\log(4/\alpha)} + \sqrt{\log n}]) \leq \alpha/4$$

and $\mathrm{P}(\max_{i\in[n]}(-X_i) \geq \sigma\sqrt{2}[\sqrt{\log(4/\alpha)} + \sqrt{\log n}]) \leq \alpha/4$ by symmetric property. Conditioning on event

$$\max_{i\in[n]} |X_i| \leq \sigma_{n,\alpha} := \sigma\sqrt{2}[\sqrt{\log(4/\alpha)} + \sqrt{\log n}],$$

For i.i.d. $\{X_i\}_{i=1}^n$ with $a \leq X_i \leq b$, Hoeffding's inequality gives $\mathrm{P}(|\bar{X}_n - \mu| \geq \frac{b-a}{\sqrt{2}}\sqrt{\frac{1}{n}\log(\frac{2}{\delta})}) \leq \delta)$. Let $\delta = \alpha/2$ and $a = \sigma_{n,\alpha}, b = -\sigma_{n,\alpha}$

$$\begin{aligned}
&\mathrm{P}\left(|\bar{X}_n - \mu| \geq \sqrt{2}\sigma_{n,\alpha}\sqrt{n^{-1}\log(4/\alpha)}\right) \\
&\leq \mathrm{P}\left(|\bar{X}_n - \mu| \geq \sqrt{2}\sigma_{n,\alpha}\sqrt{n^{-1}\log(4/\alpha)}, \max_{i\in[n]}|X_i| \leq \sigma_{n,\alpha}\right) + \mathrm{P}\left(\max_{i\in[n]}|X_i| > \sigma_{n,\alpha}\right) \leq \alpha/2 + \alpha/2 = \alpha
\end{aligned}$$

Then, $\mathrm{P}(\mu \in [\bar{X}_n \pm 2\sigma\sqrt{n^{-1}\log(4/\alpha)}[\sqrt{\log(4/\alpha)} + \sqrt{\log n}]) \geq 1 - \alpha$. □

Lemma 3 can be extend to unbounded non-Gaussian r.v.s. We need some geometric aspect of the concentration inequalities; see Section 3.2 in Wainwright (2019). A function $\psi(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ is γ-strongly concave if there is some $\gamma > 0$ s.t.

$$\lambda\psi(\mathbf{x}) + (1-\lambda)\psi(\mathbf{y}) - \psi(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \leq \tfrac{\gamma}{2}\lambda(1-\lambda)\|\mathbf{x} - \mathbf{y}\|_2^2,\ \forall\, \lambda \in [0,1] \text{ and } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

A continuous probability density $f(\mathbf{x})$ and the corresponding r.v. is strongly log-concave if $f(\mathbf{x})$ is a strongly log-concave function.

Lemma 4 (Theorem 3.16 in Wainwright (2019)). *Let* P *be any* $\gamma-$*strongly log-concave distribution on* $\mathbb{R}^n$ *with* $\gamma > 0$. *Then for any* L*-Lipschitz function* $f : \mathbb{R}^n \to \mathbb{R}$ *w.r.t. Euclidean norm, we have*

$$\mathrm{P}[f(X) - \mathrm{E}f(X) \geq t] \leq e^{-\frac{\gamma t^2}{4L^2}} \text{ for } X \sim \mathrm{P} \text{ and } t \geq 0.$$

From Lemma 4, assume $\{X_i\}_{i=1}^n$ are independent r.v.s which are $\gamma-$strongly log-concave distributed satisfying $\mathrm{P}[f(X) - \mathrm{E}f(X) \geq t] \leq e^{-\frac{\gamma t^2}{4L^2}}$ for any $f : \mathbb{R}^n \to \mathbb{R}$ that is L-Lipschitz w.r.t. the Euclidean norm. Note that

$$\max_{i\in[n]} x_i - \max_{i\in[n]} y_i \leq |x_l - y_l| \leq \|x - y\|_2 \text{ for some } l \in [n].$$

For $L = 1$, thus we obtain the following max-concentration inequality.

Corollary 1 (Borell-TIS inequality for $\gamma-$strongly log-concave distributions). *If* $\{X_i\}_{i=1}^n$ *are independent r.v.s which are* $\gamma-$*strongly log-concave distributed, then*

$$\mathrm{P}(\max_{i\in[n]} X_i - \mathrm{E}[\max_{i\in[n]} X_i] > t) \leq e^{-\frac{t^2}{4/\gamma}}.$$

From Corollary 7.4 in Zhang & Chen (2021), one has maximal inequality for sub-exponential r.v.s $\mathrm{E}(\max_{i\in[n]} |X_i|) \leq \log(1+n)\|X_1\|_{w_1}$ if $\|X_1\|_{w_1} < \infty$. Then with probability at least $1 - e^{-\frac{t^2}{4/\gamma}}$,

$$\max_{i\in[n]} |X_i| \leq \mathrm{E}(\max_{i\in[n]} |X_i|) + t \leq \log(1+n)\|X_1\|_{w_1} + t.$$

A.2 Historical notes for sub-Gaussian and its optimal parameter

The MGF-based variance proxy in Definition 1 for sub-Gaussian distribution dates back to Kahane (1960), and it is not unique which cannot be view as the parameter. So it motivates Chow (1966) to defined the optimal variance proxy $\sigma_{opt}^2(X)$ as the unique parameter of sub-Gaussian distribution. $\sigma_{opt}^2(X)$ is also called the sub-Gaussian norm in Wang (2020) or sub-Gaussian diameter in Kontorovich (2014). The monograph Buldygin & Kozachenko (2000) gave comprehensive studies concerning metric characterizations for various sub-Gaussian norms of certain random variables.

From Chernoff's inequality, the exponential decay of the sub-Gaussian tail is obtained

$$\mathrm{P}\left(X \geq t\right) \leq \inf_{s>0} \exp\{-st\}\mathrm{E}\exp\{sX\} \leq \inf_{s>0} \exp(-st + \tfrac{\sigma_{opt}^2 s^2}{2}) = \exp(-\tfrac{t^2}{2\sigma_{opt}^2})$$

by minimizing the upper bound via putting $s = t/\sigma^2$. Moreover, for independent $\{X_i\}_{i=1}^n$ with $X_i \sim \mathrm{subG}(\sigma_i^2)$, we have sub-Gaussian Hoeffding's inequality (Chow, 1966)

$$\mathrm{P}(|\sum_{i=1}^n X_i| \geq t) \leq 2\exp\left\{-\frac{t^2}{2\sum_{i=1}^n \sigma_{opt}^2(X_i)}\right\} \leq 2\exp\left\{-\frac{t^2}{2\sum_{i=1}^n \sigma_i^2}\right\},\ t \geq 0 \tag{12}$$

for any variance proxies $\{\sigma_i^2\}_{i=1}^n$ of $\{X_i\}_{i=1}^n$.

The $\sigma_{opt}^2(X)$ not only characterizes the speed of decay in (12) but also naturally bounds the variance of X as well. To appreciate this, observe that by the definition of sub-Gaussian:

$$\begin{aligned}\frac{s^2}{2}\sigma_{opt}^2(X) + o(s^2) = \exp(\frac{\sigma_{opt}^2(X)s^2}{2}) - 1 \geq \mathrm{E}\exp(sX) - 1 &= s\mathrm{E}X + \frac{s^2}{2}\mathrm{E}X^2 + \cdots \\ &= \frac{s^2}{2}\cdot \mathrm{Var}\,X + o\left(s^2\right) \end{aligned}\tag{13}$$

(by dividing s^2 on both sides and taking $s \to 0$) which implies

$$\sigma_{opt}^2(X) \geq \mathrm{Var}\,X. \tag{14}$$

Thus, $\sigma_{opt}^2(X)$ provides a conservative lower bound for optimal proxy variance.

Interestingly, some special distributions whose variance can attain the $\sigma_{opt}^2(X)$. For example, Bernoulli r.v. $X \in \{0,1\}$ with mean $\mu \in (0,1)$ [denote $X \sim \text{Bern}(\mu)$] is sub-Gaussian with

$$1/4 \geq \sigma_{opt}^2(X) = \frac{(1-2\mu)}{2\log\frac{1-\mu}{\mu}} \geq \mu(1-\mu) = \text{Var}(X)$$

in Kearns & Saul (1998), while, Hoeffding's inequality shows a crude bound $X - \mu \sim \text{subG}(1/4)$. The inequality holds if the Bernoulli distributions is symmetric with $\mu = 1/2$, i.e. $\sigma_{opt}^2(X) = \lim\limits_{\mu\to 1/2} \frac{(1-2\mu)}{2\log\frac{1-\mu}{\mu}} = 1/4 = \text{Var}\, X$, and the inequality $\sigma_{opt}^2(X) = \text{Var}\, X$ define the strict sub-Gaussianity:

Definition 4 (Buldygin & Kozachenko (2000))**.** *For zero-mean $X \sim \text{subG}(\sigma^2)$ is called strict sub-Gaussian if* $\text{Var}\, X = \sigma_{opt}^2(X)$ *[denote $X \sim \text{ssubG}(\sigma_{opt}^2(X))$].*

The strict sub-Gaussian r.v.s include Gaussian, symmetric Beta, symmetric Bernoulli and $U[-c,c]$; Marchal et al. (2017) showed that by a second order ODE (with a unique solution of the Cauchy problem) $\text{Beta}(\alpha,\beta)$ has

$$\sigma_{\text{opt}}^2(\alpha,\beta) = \frac{\alpha}{(\alpha+\beta)x_0}\left(\frac{{}_1F_1(\alpha+1;\alpha+\beta+1;x_0)}{{}_1F_1(\alpha;\alpha+\beta;x_0)} - 1\right) \geq \text{Var}[\text{Beta}(\alpha,\beta)],$$

where x_0 is a unique solution of $\log\left({}_1F_1\left(\alpha;\alpha+\beta;x_0\right)\right) = \frac{\alpha x_0}{2(\alpha+\beta)}\left(1 + \frac{{}_1F_1(\alpha+1;\alpha+\beta+1;x_0)}{{}_1F_1(\alpha;\alpha+\beta;x_0)}\right)$. Finding the explicit expression and giving the iff condition for general distributions (such as unbounded or asymmetrical distributions) are still an open questions Marchal et al. (2017).

Similar to Definition 1, if sub-G variable is unbounded, we define the *optimal lower variance proxy* that renders a sharp reverse Chernoff inequality and a sharp lower tails of sub-Gaussian maxima in below.

Definition 5. *The optimal lower variance proxy for a sub-G X is defined as*

$$l_{opt}^2(X) := \sup\left\{l^2 \geq 0 : \mathbb{E}\exp(tX) \geq \exp\{l^2t^2/2\},\, \forall\, t \in \mathbb{R}\right\} = 2\inf{}_{t\in\mathbb{R}} t^{-2}\log[\text{E}\exp(tX)]. \tag{15}$$

Lemma 5 (A sharp reverse Chernoff inequality)**.** *Suppose that $l_{opt}^2(X) > 0$ for a sub-G r.v. X. For $t > 0$, then*

$$\text{P}(X \geq t) \geq C_{\sigma,l}^2(X)\exp\left\{-4[2\sigma_{opt}^2(X)/l_{opt}^4(X) - l_{opt}^{-2}(X)]t^2\right\},$$

where $C_{\sigma,l}(X) := \left(\frac{l_{opt}^2(X)}{4\sigma_{opt}^2(X) - l_{opt}^2(X)}\right)\left(\frac{4\sigma_{opt}^2(X) - 2l_{opt}^2(X)}{4\sigma_{opt}^2(X) - l_{opt}^2(X)}\right)^{2[2\sigma_{opt}^2(X)/l_{opt}^2(X) - 1]} \in (0,1).$

Proposition 2. *(a). Suppose that $l_{opt}^2(X) > 0$ for i.i.d. sub-G r.v. $\{X_i\}_{i=1}^n \sim X$. With probability at least $1-\delta$,*

$$\frac{l_{opt}(X)/\sigma_{opt}(X)}{2\sqrt{2\sigma_{opt}^2(X)/l_{opt}^2(X) - 1}}\sqrt{\log n - \log C_{\sigma,l}^{-2}(X) - \log\log\left(\tfrac{2}{\delta}\right)} \leq \max\nolimits_{1\leq i\leq n}\frac{X_i}{\sigma_{opt}(X)} \leq \sqrt{2[\log n + \log\left(\tfrac{2}{\delta}\right)},$$

where $C_{\sigma,l}(X) < 1$ is constant defined in Lemma 1 below; (b) if X is bounded variable, then $l_{opt}^2(X) = 0$.

The proof of Lemma 5 and Proposition 2 is similar to Lemma 1 and Theorem 1.

A.2.1 Remarks for Orlicz norm and other commonly used norms

In Remarks 2 and 1 below, we will show $\text{P}(|X| \geq t) \leq 2\exp\{-\frac{t^2}{2}/(2e\|X\|_{\psi_2}^2)\}$ and $\text{P}(|X| > t) \leq 2\exp\left\{-t^2/[2(\|X\|_{w_2}/\sqrt{2})^2]\right\}$ for all $t \geq 0$. The variance can be upper bounded by both norms as $(2\|X\|_{w_2})^2 \geq \text{Var}X$ and $(\sqrt{2}\|X\|_{\psi_2})^2 \geq \text{Var}X$.

Remark 1. *If $\text{E}\exp(|X|^2/\|X\|_{w_2}^2) \leq 2$ for $\|X\|_{w_2} < \infty$, then Markov's inequality gives for all $t \geq 0$*

$$\text{P}(|X| > t) \leq \text{P}(e^{|X/\|X\|_{w_2}|^2} \geq e^{t^2/\|X\|_{w_2}^2}) \leq 2\exp\left\{-\frac{t^2}{2}/(\frac{\|X\|_{w_2}}{\sqrt{2}})^2\right\}. \tag{16}$$

Using Lemma 1.5 in Rigollet & Hütter (2019):

$$\textit{if}\ \text{P}(|X| > t) \leq 2\exp\left\{-t^2/[2\sigma^2]\right\}\ \textit{with}\ \text{E}X = 0,\ \textit{then}\ \text{E}\exp\{sX\} \leq \exp\{4\sigma^2 s^2\} \tag{17}$$

for any $s \geq 0$. Therefore, we have

$$\mathrm{E}\exp\{sX\} \leq \exp\left\{(2\|X\|_{w_2})^2\frac{s^2}{2}\right\}.$$

By the same argument in (13)*, one has* $(2\|X\|_{w_2})^2 \geq \mathrm{Var}X$.

Example 2. *(i) For $X \sim N(0,\sigma^2)$, one has $\|X\|_{w_2} = \sqrt{\frac{8}{3}}\sigma$ (see Example 3.4 in Zhang & Chen (2021)), and then we have a crude tail bound from* (16)*, $\mathrm{P}(|X| > t) \leq 2\exp\left\{-\frac{t^2}{2}/(\frac{4}{3}\sigma^2)\right\}$; (ii) For $X \sim \mathrm{Bern}(0.5)$, one has $\|X-0.5\|_{w_2} = \frac{1}{2\sqrt{\log 2}}$, which leads to the bound $\mathrm{P}(|X-0.5| > t) \leq 2\exp\left\{-\frac{t^2}{2}/\frac{1}{4\log 2}\right\}$; (iii) For $X \sim [a,b]$, we have $\|X-\frac{a+b}{2}\|_{w_2} \leq \frac{b-a}{2\sqrt{\log 2}}$ by and then $\mathrm{P}(|X-\frac{a+b}{2}| > t) \leq \exp\left\{-\frac{t^2}{2}/\frac{(b-a)^2}{2\log 2}\right\}$. The results in (ii) and (iii) comes from the conclusion about bounded variables in Example 1 in Zhang & Wei (2022), while Hoeffding's inequality gives sharper tail inequalities $\mathrm{P}(|X-\mu| > t) \leq 2\exp\{-2t^2\}$ and $\mathrm{P}(|X-\frac{a+b}{2}| > t) \leq 2\exp\{-\frac{2t^2}{(b-a)^2}\}$ in (ii) and (iii).*

Remark 2. *Recall Vershynin (2010)'s definition of sub-Gaussian norm*

$$\|X\|_{\psi_2} = \max_{p\geq 1} p^{-1}\big(\mathrm{E}|X|^p\big)^{1/p}.$$

By $p! \geq (p/e)^p$, a crude bounds also appears. Indeed, by $\big(\mathrm{E}|X|^p\big)^{1/p} \leq K\sqrt{p}$ for all integer $p \geq 1$,

$$\mathrm{E}e^{c^{-1}X^2} = 1 + \sum_{p=1}^{\infty}\frac{c^{-p}X^{2p}}{p!} \leq 1 + \sum_{p=1}^{\infty}\frac{c^{-p}(2K^2p)^p}{p!} \leq 1 + \sum_{p=1}^{\infty}(\frac{2eK^2}{c})^p = 1 + (\frac{2eK^2}{c})\sum_{p=0}^{\infty}(\frac{2eK^2}{c})^p$$
$$[\frac{2eK^2}{c} < 1] = 1 + (\frac{2eK^2}{c})\sum_{p=0}^{\infty}(\frac{2eK^2}{c})^p = 1 + (\frac{2eK^2}{c})\frac{1}{1-\frac{2eK^2}{c}}.$$

Setting $\frac{2eK^2}{c} \leq s < 1$ and assign s such that $\mathrm{E}\exp\big(c^{-1}X^2\big) \leq 1 + \frac{s}{1-s} =: 2$, the solution is $s = 1/2$. Let $K = \|X\|_{\psi_2}$. We thus have $c \geq 4e\|X\|_{\psi_2}^2$. The $\mathrm{E}e^{X^2/(4e\|X\|_{\psi_2}^2)} \leq 2$ implies $\mathrm{P}(|X| \geq t) \leq 2e^{-t^2/(4e\|X\|_{\psi_2}^2)}$. by using (17). Therefore, one has

$$\mathrm{E}\exp\{tX\} \leq \exp\left\{(4\sqrt{e}\|X\|_{\psi_2})^2 t^2/2\right\}. \tag{18}$$

Example 3. *For $X \sim N(0,1)$, observe that*

$$\|X\|_{\varphi_2} = \max_{p\geq 1} p^{-1/2}\big(\mathrm{E}|X|^p\big)^{1/p} = \lim_{p\to\infty} p^{-1/2}\big(\mathrm{E}|X|^p\big)^{1/p} = \lim_{p\to\infty}\sqrt{\tfrac{2}{p}}[\tfrac{\Gamma((1+p)/2)}{\Gamma(1/2)}]^{1/p} = 1/\sqrt{2} \approx 0.7071.$$

For uniform distributed $X \sim U[-1,1]$, one has $\|X\|_{\varphi_2} = 0.4082$ and $\mathrm{P}(|X| \geq t) \leq 2e^{-t^2/(4e\cdot 0.4082^2)} = 2e^{-t^2/1.8122}$ comparing to Hoeffding's inequality with a sharper bound $\mathrm{P}\left(|X| \geq t\right) \leq 2e^{-t^2}$. For $X \sim \mathrm{Bern}(\mu)$, $\frac{1}{\sqrt{p}}[\mathrm{E}|X-\mu|^p]^{1/p} = \frac{1}{\sqrt{p}}[(1-\mu)\mu^p + \mu(1-\mu)^p]^{1/p}$. Let $\mu = 0.3$, we have $\|X\|_{\varphi_2} = 0.3240$ and $\mathrm{P}(|X| \geq t) \leq 2e^{-t^2/(4e\cdot 0.3240^2)} = 2e^{-t^2/1.1417}$ comparing to Hoeffding's inequality with a sharper bound $\mathrm{P}\left(|X| \geq t\right) \leq 2e^{-2t^2}$.

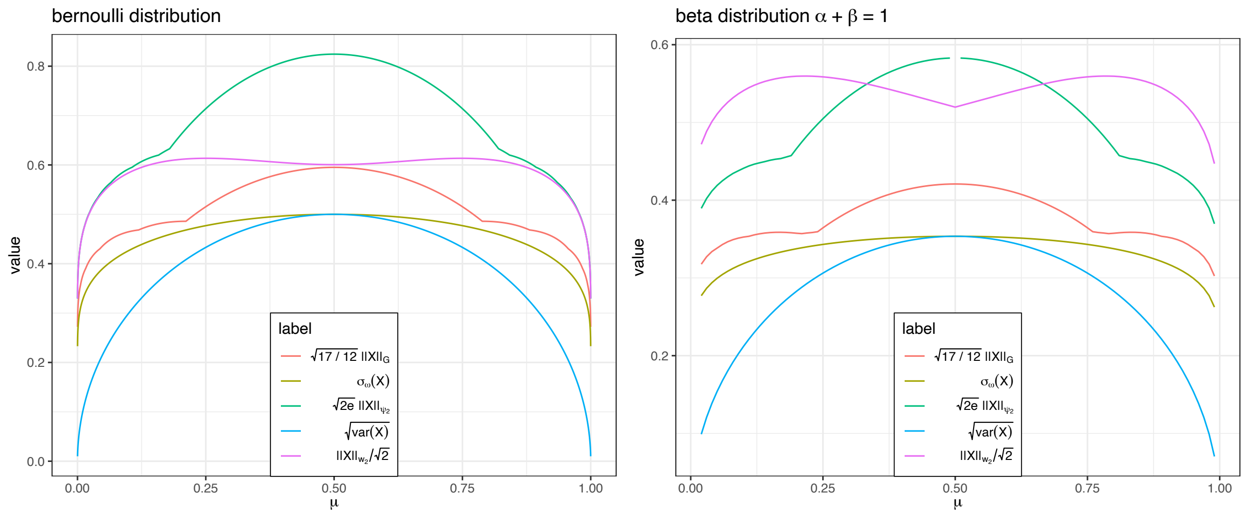
Example 4. *The* (18) *implies*

$$\mathrm{P}(|\sum_{i=1}^{n} X_i| \geq t) \leq 2\exp\left\{-\frac{t^2}{2\sum_{i=1}^{n}(4\sqrt{e}\|X_i\|_{w_2})^2}\right\},\ t \geq 0. \tag{19}$$

The $2\|X\|_{\psi_2}^2 \geq \mathrm{Var}\,X$ implies that for strictly sub-G independent variables $\{X_i\}_{i=1}^n$

$$\mathrm{P}(|\textstyle\sum_{i=1}^{n} X_i| \geq t) \leq 2\exp\left\{-\frac{t^2}{2\sum_{i=1}^{n}\mathrm{Var}(X_i)}\right\} \leq 2\exp\left\{-\frac{t^2}{4\sum_{i=1}^{n}\|X_i\|_{\psi_2}^2}\right\} < 2\exp\left\{-\frac{t^2}{2\sum_{i=1}^{n}(4\sqrt{e}\|X_i\|_{w_2})^2}\right\}.$$

Hence, by (19)*, $2\|\cdot\|_{w_2}$-norm leads to a looser concentration bound.*



Bernoulli $\quad\quad$ $\mathrm{Beta}(\alpha, \beta)$

Figure 5: The half length of $1-\delta$ confidence interval with different norms. The results are divided by $\sqrt{2\log(2/\delta)}$ to eliminate the affect of δ.

A.3 Details for Table 1

Let $\{X_i\}_{i=1}^n$ be i.i.d. r.v.s with $\mathrm{E}X_1 = 0, \mathrm{E}X_1^2 = \sigma^2 > 0$, and $\mathrm{E}\,|X_1|^3 = \rho < \infty$. Shevtsova (2013) gave a tighter estimate of the absolute constant in B-E bounds for $\bar{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$:

$$\Delta_n := \sup_{x\in\mathbb{R}} \left| \mathrm{P}\left(\frac{\sqrt{n}}{\sigma}\bar{X}_n \le x\right) - \Phi(x)\right| \le \frac{0.3328\left(\rho + 0.429\sigma^3\right)}{\sigma^3\sqrt{n}}, \ \forall\, n \ge 1, \tag{20}$$

where $\Phi(\cdot)$ is the cumulative distribution function of $N(0,1)$.

Consider Bernoulli samples $\{X_i\}_{i=1}^n \overset{\text{i.i.d.}}{\sim} \mathrm{Ber}\,(1/2)$ with $\sigma = 1/2$ and $\rho = 1/8$, and Zolotukhin et al. (2018) shown $\Delta_n \le 0.409954/\sqrt{n}$. Put $\delta = 0.05, 0.075, 0.1$. For $n \ge 1$, Hoeffding's inequality gives

$$\mathrm{P}\left(|\bar{X}_n - 1/2| \le \tfrac{1}{2\sqrt{n}} \cdot \sqrt{2\log(\tfrac{2}{\delta})}\right) \ge 1-\delta.$$

From B-E bounds (20), we have $\mathrm{P}(\frac{\sqrt{n}}{\sigma}(\bar{X}_n - 1/2) \le -x) - \Phi(-x) \le \Delta_n$ and $\mathrm{P}(\frac{\sqrt{n}}{\sigma}(1/2 - \bar{X}_n) \le -x) - \Phi(-x) \le \Delta_n$ by the symmetry of $\mathrm{Ber}\,(1/2)$. Set

$$\mathrm{P}\left(\tfrac{\sqrt{n}}{\sigma}(\bar{X}_n - 1/2) \le -x\right) \le \Delta_n + \Phi(-x) \le \tfrac{0.409954}{\sqrt{n}} + \Phi(-x) =: \tfrac{\delta}{2};$$
$$\mathrm{P}\left(\tfrac{\sqrt{n}}{\sigma}(\bar{X}_n - 1/2) \ge x\right) \le \Delta_n + \Phi(-x) \le \tfrac{0.409954}{\sqrt{n}} + \Phi(-x) =: \tfrac{\delta}{2},$$

where $x = -\Phi^{-1}(\frac{\delta}{2} - \frac{0.409954}{\sqrt{n}})$ with $\frac{\delta}{2} - \frac{0.409954}{\sqrt{n}} > 0$.

Then, it results in a $(1-\delta)100\%$'s non-asymptotic CI:

$$\mathrm{P}\left(|\bar{X}_n - 1/2| \le -\tfrac{1}{2\sqrt{n}} \cdot \Phi^{-1}(\tfrac{\delta}{2} - \tfrac{0.409954}{\sqrt{n}})\right) \ge 1-\delta$$

for $n \ge (0.8199/\delta)^2$, which require least sample sizes $n \ge 269, 120, 68$ for $\delta = 0.05, 0.075, 0.1$ respectively.

For symmetric data with zero mean and finite third moment, (20) gives a trivial bound if we put $\frac{0.3328(\rho+0.429\sigma^3)}{\sigma^3\sqrt{n}} \ge 1$, i.e. the B-E bound is useless when $n \le \frac{[0.3328(\rho+0.429\sigma^3)]^2}{\sigma^6}$.

B SMALL SAMPLE LEAVE-ONE-OUT AVERAGE IN MOMENT NORM ESTIMATIONS

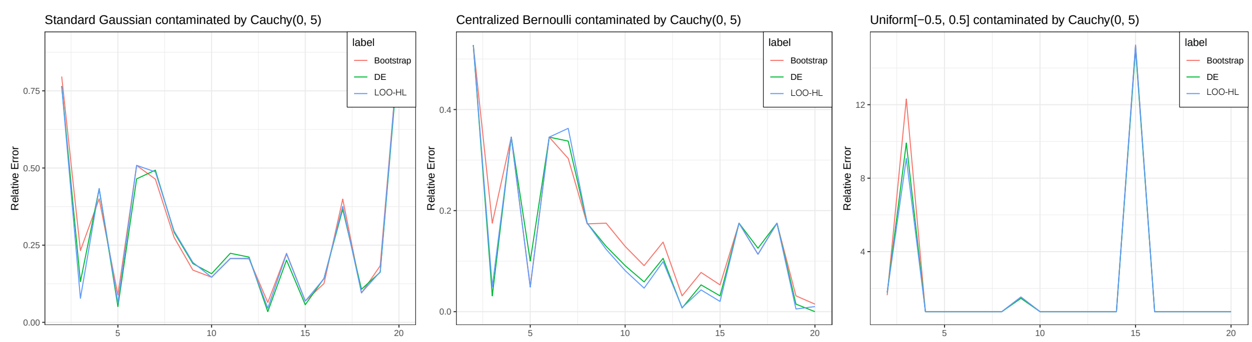
For the small sample zise ($n \le 20$), one has two other methods, except the direct empirical moment method (DE). The first one is the well-known Bootstrap. The non-parametric Bootstrap can reduce the estimator's variance and make it more robust (see p512 in Hesterberg (2011)). Especially, here we use $(n-1)$-out-of-n Bootstrap and construct $n-1$ Bootstrap estimators and then take the median of these estimators. The second robust method under small sample setting is called the leave-one-out Hodges-Lehmann method (LOO-HL) proposed by Rousseeuw & Verboven (2002). Specially, based on sample $X = (X_1, \ldots, X_n)^\top \in \mathbb{R}^n$, define LOO-HL empirical mean estimator as

$$\hat{\mu}_{\text{LOO}} := \text{med}\left\{ \frac{\text{DE}(X_{(-i)}) + \text{DE}(X_{(-j)})}{2} : 1 \le i < j \le n \right\},$$

where $\text{DE}(X_{(-i)}) := \sum_{k \ne i} X_k/(n-1)$ is the empirical mean estimator of $X_{(-i)} = \{X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\}$.

It is worthy to note that the leave-one-out method is different from classic Hodges–Lehmann empirical mean estimator $\hat{\mu}_{\text{HL}} := \text{med}_{i \le j} \frac{X_i + X_j}{2}$ which uses a single sample X_i instead of leave-one-out mean $\text{DE}(X_{(-i)})$ since in practice we find that classic Hodges-Lehmann method cannot render ideal performance.

To see the performance, we use the three methods above to calculate the relative estimators' errors based on small samples corresponding to the settings in the previous section, except we use 1% independent Cauchy$(0, 5)$ perturbation to contaminate the original distribution. The results is shown in Figure 6. It can be seen that the three methods can achieve relatively good performance, while the LOO-HL method gives less error overall and is obviously better than the other two methods when $2 \le n \le 4$.



Standard Gaussian Centralized bernoulli Uniform on $[-0.5, 0.5]$

Figure 6: The relative error of $\|\cdot\|_G$-norm estimation for three kind distributions by using direct empirical moment method (DE), Bootstrap, and Leave-one-out Hodges-Lehmann method (LOO-HL).

C PROOFS OF MAIN RESULTS

Proof of Lemma 2. Note that $\|X\|_G := \max_{m \in 2\mathbb{N}} \left[\frac{\text{E}X^m}{\text{E}Z^m}\right]^{1/m}$, where $Z \sim N(0,1)$. If the maximum take at $m = \infty$,

$$\max_{m=2k \in 2\mathbb{N}} \left[\frac{\text{E}X^m}{\text{E}Z^m}\right]^{1/m} = \max_{k \in \mathbb{N}} \left[\frac{2^k}{\sqrt{\pi}} \Gamma\left(k + \frac{1}{2}\right)\right]^{-1/(2k)} \|X\|_{2k}$$

is an increasing function for some sub-sequence $\{k_\ell\} \subseteq \{k\}$ such that $\lim_{\ell \to \infty} k_\ell = \infty$ when ℓ is large enough, where we use the formula of $2k$-th moment of standard normal distribution (see (18) in Winkelbauer (2012)).

Therefore,

$$\left[\frac{2^{k_\ell}}{\sqrt{\pi}} \Gamma\left(k_\ell + \frac{1}{2}\right)\right]^{-1/(2k_\ell)} \|X\|_{2k_\ell} \le \left[\frac{2^{k_\ell+1}}{\sqrt{\pi}} \Gamma\left(k_\ell + \frac{3}{2}\right)\right]^{-1/(2k_\ell+2)} \|X\|_{2k_\ell+2}$$

i.e. $\frac{2k_\ell+1}{2}\left[\frac{\sqrt{\pi}}{\Gamma(k_\ell+1/2)}\right]^{1/k_\ell} \le \|X\|_{2k_\ell+2}^{2k_\ell+2}/\|X\|_{2k_\ell}^{2k_\ell+2}$ for any k_ℓ is large enough. Let $\ell \to \infty$, we have

$$1 = \limsup_{\ell\to\infty} \frac{\|X\|_{2k_\ell+2}^{2k_\ell+2}}{\|X\|_{2k_\ell}^{2k_\ell+2}} \ge \lim_{\ell\to\infty} \frac{2k_\ell+1}{2}\left[\frac{\sqrt{\pi}}{\Gamma(k_\ell+1/2)}\right]^{1/k_\ell}$$
$$= \lim_{k_\ell\to\infty}(k_\ell+1/2)\left[\frac{\sqrt{\pi}}{\sqrt{2\pi}(k_\ell+1/2)^{k_\ell}e^{-(k_\ell+1/2)}}\right]^{1/k_\ell} = \lim_{\ell\to\infty}\frac{e^{1+1/(2k_\ell)}}{\sqrt{2}} = \frac{e}{\sqrt{2}} > 1,$$

which leads to a contradiction, where we use a fact that $\lim_{n\to\infty}\|X\|_n = \|X\|_\infty = \operatorname{esssup}|X|$, and hence $\limsup_{\ell\to\infty}\|X\|_{2k_\ell+2}^{2k_\ell+2}/\|X\|_{2k_\ell}^{2k_\ell+2} = 1$. As a result, one must have $\arg\max_{m\in 2\mathbb{N}}\left[\frac{\mathrm{E}X^m}{\mathrm{E}Z^m}\right]^{1/m} < \infty$. □

Proof of Theorem 2. If X_i is symmetric around zero, then we have by $\mathrm{E}X_i^{2k+1} = 0$ for $k \in \mathbb{N}_+$

$$\mathrm{E}e^{tX_i} = 1 + \sum_{k=1}^{\infty}\frac{t^{2k}\mathrm{E}X_i^{2k}}{(2k)!} \le 1 + \sum_{k=1}^{\infty}\frac{t^{2k}}{(2k)!}\frac{(2k)!\,\|X_i\|_G^{2k}}{2^k k!} = 1 + \sum_{k=1}^{\infty}\frac{\left(t^2\|X_i\|_G^2/2\right)^k}{k!} = \exp\{\frac{t^2\|X_i\|_G^2}{2}\}$$

for all $t \in \mathbb{R}$, where the last inequality is by the definition of $\|X_i\|_G < \infty$ such that $\mathrm{E}X_i^{2k} \le \frac{(2k)!}{2^k k!}\|X_i\|_G^{2k}$. Then it proves $X_i \sim \mathrm{subG}(\|X_i\|_G^2)$, which shows case (a).

For case (b), if X_i has zero mean, then we bound the odd moment by even moments. For $k = 1, 2, \cdots$ and $c_k > 0$, Cauchy's inequality and mean value inequality imply

$$\mathrm{E}|tX_i|^{2k+1} \le \left(c_k^{-1}\mathrm{E}|tX_i|^{2k}\cdot c_k\mathrm{E}|tX_i|^{2k+2}\right)^{1/2} \le \left(c_k^{-1}t^{2k}\mathrm{E}X_i^{2k} + c_k t^{2k+2}\mathrm{E}X_i^{2k+2}\right)/2.$$

So, $\frac{\mathrm{E}|tX_i|^3}{3!} \le \frac{c_1^{-1}t^2\mathrm{E}X_i^2 + c_1 t^4\mathrm{E}X_i^4}{2\cdot 3!}$, $\frac{\mathrm{E}|tX_i|^5}{5!} \le \frac{c_2^{-1}t^4\mathrm{E}X_i^4 + c_2 t^6\mathrm{E}X_i^6}{2\cdot 5!}$, and so on, which implies

$$\begin{aligned}
\mathrm{E}e^{tX_i} &\le 1 + \sum_{k=2}^{\infty}\frac{t^k\mathrm{E}|X_i|^k}{k!} \le 1 + \frac{t^2\mathrm{E}X_i^2}{2!} + \frac{c_1^{-1}t^2\mathrm{E}X_i^2 + c_1t^4\mathrm{E}X_i^4}{2\cdot 3!} + \frac{t^4\mathrm{E}X_i^4}{4!} \\
&+ \frac{c_2^{-1}t^4\mathrm{E}X_i^4 + c_2t^6\mathrm{E}X_i^6}{2\cdot 5!} + \frac{t^6\mathrm{E}X_i^6}{6!} + \frac{c_3^{-1}t^6\mathrm{E}X_i^6 + c_3t^8\mathrm{E}X_i^8}{2\cdot 7!} + \cdots \\
&\le 1 + \left(1 + \frac{c_1^{-1}}{3!}\right)\frac{t^2\mathrm{E}X_i^2}{2!} + \left(1 + \frac{4!c_1}{2\cdot 3!} + \frac{4!c_2^{-1}}{2\cdot 5!}\right)\frac{t^4\mathrm{E}X_i^4}{4!} + \left(1 + \frac{6!c_2}{2\cdot 5!} + \frac{6!c_3^{-1}}{2\cdot 7!}\right)\frac{t^6\mathrm{E}X_i^6}{6!} + \cdots \\
&\le 1 + \left(1 + \frac{1}{6c_1}\right)\frac{t^2\mathrm{E}X_i^2}{2!} + \left(1 + 2c_1 + \frac{c_2^{-1}}{10}\right)\frac{t^4\mathrm{E}X_i^4}{4!} + \left(1 + 3c_2 + \frac{c_3^{-1}}{14}\right)\frac{t^6\mathrm{E}X_i^6}{6!} + \cdots \\
&\le 1 + \left(1 + \frac{1}{6c_1}\right)\frac{t^2\mathrm{E}X_i^2}{2!} + \left(1 + 2c_1 + \frac{c_2^{-1}}{10}\right)\frac{t^4\mathrm{E}X_i^4}{4!} + \sum_{k=3}^{\infty}\left(1 + kc_{k-1} + \frac{c_k^{-1}}{4k+2}\right)\frac{t^{2k}\mathrm{E}X_i^{2k}}{(2k)!}. \qquad (21)
\end{aligned}$$

To bound (21), we assign $c_k = x^{-1}\cdot\frac{m^{k+1}}{2k+2}$ for $k \ge 2$ and $x, m > 0$. Consider the following system of equations:

$$\begin{cases} 1 + \frac{1}{6c_1} = m \\ 1 + 2c_1 + \frac{c_2^{-1}}{10} = m^2 \end{cases}$$

This system with $c_2 = \frac{1}{6\cdot x}m^3$ gives $1 + 2c_1 + 0.6x\left(1 + \frac{1}{6c_1}\right)^{-3} = \left(1 + \frac{1}{6c_1}\right)^2$ which could implies $c_1 = 0.4$ if we set $x = 0.9806308$. And then $m = 1 + \frac{1}{6c_1} = \frac{17}{12}$. Therefore, (21) has a further upper bound

$$\begin{aligned}
\mathrm{E}e^{tX_i} &\leq 1 + m\frac{t^2\mathrm{E}X_i^2}{2!} + m^2\frac{t^4\mathrm{E}X_i^4}{4!} + \sum_{k=3}^{\infty}\left(1 + kc_{k-1} + \frac{c_k^{-1}}{4k+2}\right)\frac{t^{2k}\mathrm{E}X_i^{2k}}{(2k)!} \\
&\leq \sum_{k=0}^{\infty}\frac{(\sqrt{m}t)^{2k}\mathrm{E}X_i^{2k}}{(2k)!} \leq 1 + \sum_{k=1}^{\infty}\frac{(\sqrt{m}t)^{2k}}{(2k)!}\frac{(2k)!\,\|X_i\|_G^{2k}}{2^k k!} \leq \exp\{\frac{t^2(\sqrt{17/12}\|X_i\|_G)^2}{2}\}, \qquad (22)
\end{aligned}$$

where the first inequality stems from $1 + \frac{m^k}{2x} + \frac{(k+1)x}{2k+1}\cdot m^{-k-1} \leq m^k$ with $m = 17/12, k = 3, 4, \cdots$, and the last inequality is by the definition of $\|X_i\|_G$. Thus we show $X_i \sim \mathrm{subG}(17\|X_i\|_G^2/12)$. □

Proof of Theorem 1. For (a), it remains to show the lower tail bound. For $t \geq 0$, by the independence of $\{X_i\}_{i=1}^n$,

$$\begin{aligned}
&\mathrm{P}\{\max_{1\leq i\leq n} X_i \leq t\} = \mathrm{P}\left(X_1 \leq t, \cdots, X_n \leq t\right) = \prod_{i=1}^n \mathrm{P}\left(X_i \leq t\right) \\
&[\text{Applying Lemma 1}] \ \leq \left(1 - C^2(X)\mathrm{e}^{-4[2\|X\|_G^2/\|X\|_{\tilde{G}}^2-1]t^2}\right)^n \leq \exp\left(-nC^2(X)\mathrm{e}^{-4[2\|X\|_G^2/\|X\|_{\tilde{G}}^2-1]t^2}\right),
\end{aligned}$$

where we use $1 - x \leq e^{-x}$ for all $x \in \mathbb{R}$ in the last inequality.

Let $\delta = \exp(-nC^2(X)\mathrm{e}^{-4[2\|X\|_G^2/\|X\|_{\tilde{G}}^2-1]})$ and we get $t = \frac{\|X\|_{\tilde{G}}/\|X\|_G}{2\sqrt{2\|X\|_G^2/\|X\|_{\tilde{G}}^2-1}}\sqrt{\log n - \log C^{-2}(X) - \log\log\left(\frac{2}{\delta}\right)}$.

For (b), if $X \leq M < \infty$, it shows

$$0 \leq \|X\|_{\tilde{G}} \leq \min_{k\geq 1}\left[\frac{2^k k!}{(2k)!}M^{2k}\right]^{1/(2k)} = M\min_{k\geq 1}\left[\frac{2^k k!}{(2k)!}\right]^{1/(2k)} = 0,$$

So we immediately get $\|X\|_{\tilde{G}} = 0$. □

Proof of Lemma 1. The proof is based on Paley–Zygmund inequality $\mathrm{P}(Z \geq \theta\mathrm{E}Z) \geq (1-\theta)^2\frac{\mathrm{E}[Z]^2}{\mathrm{E}[Z^2]}$ for a positive r.v. Z with finite variance, where $\theta \in (0, 1)$; see Page 47 in Boucheron et al. (2013).

Since X is symmetric around zero, one has $\mathrm{E}X^{2k+1} = 0$ for $k \in \mathbb{N}_+$, which gives for all $s \in \mathbb{R}$,

$$\mathrm{E}e^{sX} = 1 + \sum_{k=1}^{\infty}\frac{s^{2k}\mathrm{E}X^{2k}}{(2k)!} \geq 1 + \sum_{k=1}^{\infty}\frac{s^{2k}}{(2k)!}\frac{(2k)!\,\|X\|_{\tilde{G}}^{2k}}{2^k k!} = 1 + \sum_{k=1}^{\infty}\frac{\left(s^2\|X\|_{\tilde{G}}^2/2\right)^k}{k!} = \exp\{\frac{s^2\|X\|_{\tilde{G}}^2}{2}\}, \qquad (23)$$

where the last inequality stems from the definition of $\|X\|_{\tilde{G}} < \infty$ such that $\mathrm{E}X^{2k} \geq \frac{(2k)!}{2^k k!}\|X\|_{\tilde{G}}^{2k}$.

Let $Z = \exp\{sX\}$. The above Paley–Zygmund inequality and (23) imply

$$\begin{aligned}
\mathrm{P}(Z \geq t) &:= \mathrm{P}(\exp\{sX\} \geq \theta\exp\{\|X\|_{\tilde{G}}^2 s^2/2\}) \geq \mathrm{P}(\exp\{sX\} \geq \theta\mathrm{E}\exp\{sX\}) \geq (1-\theta)^2\frac{[\mathrm{E}\exp\{sX\}]^2}{\mathrm{E}\exp\{2sX\}} \\
&\geq (1-\theta)^2\frac{\exp\{\|X\|_{\tilde{G}}^2 s^2\}}{\exp\{2\|X\|_G^2 s^2\}} = (1-\theta)^2\exp\{-[2\|X\|_G^2 - \|X\|_{\tilde{G}}^2]s^2\} \qquad (24)
\end{aligned}$$

where $t := \|X\|_{\tilde{G}}^2 \frac{s}{2} + \frac{\log\theta}{s} > 0$, and the last inequality is from Theorem 2(a). Put $s = \frac{t+\sqrt{t^2+2\|X\|_{\tilde{G}}^2\log(1/\theta)}}{\|X\|_{\tilde{G}}^2}$, which is solved from equation $\frac{1}{2}\|X\|_{\tilde{G}}s^2 - ts - \log(1/\theta) = 0$, then we have

$$s^2 = \|X\|_{\tilde{G}}^{-4}\left[t + \sqrt{t^2 + 2\|X\|_{\tilde{G}}^2\log(1/\theta)}\right]^2 \leq \frac{4t^2 + 4\|X\|_{\tilde{G}}^2\log(1/\theta)}{\|X\|_{\tilde{G}}^4}.$$

Substitute this upper bound into (24), and it leads to

$$\begin{aligned}
\mathrm{P}(X \geq t) &\geq (1-\theta)^2 \exp\left\{-\frac{[2\|X\|_G^2 - \|X\|_{\tilde{G}}^2][4t^2 + 4\|X\|_{\tilde{G}}^2\log(1/\theta)]}{\|X\|_{\tilde{G}}^4}\right\} \\
&= (1-\theta)^2 \exp\{-[2\|X\|_G^2/\|X\|_{\tilde{G}}^4 - \|X\|_{\tilde{G}}^{-2}][4t^2 + 4\|X\|_{\tilde{G}}^2\log(1/\theta)]\} \\
&= (1-\theta)^2 \theta^{4[2\|X\|_G^2/\|X\|_{\tilde{G}}^4 - \|X\|_{\tilde{G}}^{-2}]} \exp\{-4[2\|X\|_G^2/\|X\|_{\tilde{G}}^4 - \|X\|_{\tilde{G}}^{-2}]t^2\}. \qquad (25)
\end{aligned}$$

Taking sup on $\theta \in (0,1)$ over the two sides of (25), we have

$$\begin{aligned}
&\mathrm{P}(X \geq t) \geq \exp\{-4[2\|X\|_G^2/\|X\|_{\tilde{G}}^4 - \|X\|_{\tilde{G}}^{-2}]t^2\} \sup_{\theta\in(0,1)}(1-\theta)^2\theta^{4[2\|X\|_G^2/\|X\|_{\tilde{G}}^2 - 1]} \\
&= \left(\frac{\|X\|_{\tilde{G}}^2}{4\|X\|_G^2 - \|X\|_{\tilde{G}}^2}\right)\left(\frac{4\|X\|_G^2 - 2\|X\|_{\tilde{G}}^2}{4\|X\|_G^2 - \|X\|_{\tilde{G}}^2}\right)^{2[2\|X\|_G^2/\|X\|_{\tilde{G}}^2 - 1]} \exp\{-4[2\|X\|_G^2/\|X\|_{\tilde{G}}^4 - \|X\|_{\tilde{G}}^{-2}]t^2\} \qquad (26)
\end{aligned}$$

where the supremum of $\sup_{\theta\in(0,1)}(1-\theta)^2\theta^{4[2\|X\|_G^2/\|X\|_{\tilde{G}}^2-1]}$ is attained at $\theta_0 = \frac{4\|X\|_G^2 - 2\|X\|_{\tilde{G}}^2}{4\|X\|_G^2 - \|X\|_{\tilde{G}}^2}$. □

Proof of Proposition 1. As in Hao et al. (2019), the sub-Weibull condition is that $X \sim \mathrm{subW}(\eta)$ is defined as a sub-Weibull r.v. with sub-Weibull index $\eta > 0$ if it has a finite sub-Weibull norm $\|X\|_{w_\eta} := \inf\{C \in (0,\infty) : \mathrm{E}[\exp(|X|^\eta/C^\eta)] \leq 2\}$. It is easy to see that, for sub-G X, we have $X \sim \mathrm{subW}(2)$. Write

$$\widetilde{\|X\|}_G^{2k_X} - \|X\|_G^{2k_X} = \frac{1}{(2k_X - 1)!!}\left[\frac{1}{n}\sum_{i=1}^n X_i^{2k_X} - \mathrm{E}X^{2k_X}\right].$$

Since $X \sim \mathrm{subW}(2)$, by Corollary 4 in Zhang & Wei (2022), we have $X^{2k_X} \sim \mathrm{subW}(1/k_X)$. Then apply Theorem 1 in Zhang & Wei (2022), we get

$$\mathrm{P}\left(\left|\widetilde{\|X\|}_G^{2k_X} - \|X\|_G^{2k_X}\right| \leq 2en^{-1/2}\|X\|_{\psi_{1/k_X}} C(k_X^{-1})\left\{\sqrt{t} + L_n\left(k_X^{-1}, 1_n^\top n^{-1}\|X\|_{\psi_{1/k_X}}\right)t^{k_X}\right\}\right) \geq 1 - 2e^{-t},$$

where constants $C(\cdot)$ and $L_n(\cdot,\cdot)$ is defined in Theorem 1 of Zhang & Wei (2022), and

$$\begin{aligned}
&L_n(k_X^{-1}, n^{-1}\|X\|_{\psi_{1/k_X}}1_n) \\
:=&\gamma^{2k_X}A(k_X^{-1})\frac{\|n^{-1}\|X\|_{\psi_{1/k_X}}1_n\|_\infty}{\|n^{-1}\|X\|_{\psi_{1/k_X}}1_n\|_2}1\{0 < k_X^{-1} \leq 1\} + \gamma^{2k_X}B(k_X^{-1})\frac{\|n^{-1}\|X\|_{\psi_{1/k_X}}1_n\|_\beta}{\|n^{-1}\|X\|_{\psi_{1/k_X}}1_n\|_2}1\{k_X^{-1} > 1\} \\
=&\gamma^{2k_X}A(k_X^{-1})\frac{\|n^{-1}\|X\|_{\psi_{1/k_X}}1_n\|_\infty}{\|n^{-1}\|X\|_{\psi_{1/k_X}}1_n\|_2} = \gamma^{2k_X}A(k_X^{-1})/\sqrt{n}, \quad (1/\theta + 1/\beta = 1).
\end{aligned}$$

□

Proof of Theorem 3. Let $\mathrm{MOM}_b[Y] := \mathrm{med}_{s\in[b]}\{\mathrm{P}_m^{B_s}Y\}$ be the MOM estimator for data $\{Y_i\}_{i=1}^n$.

Since b_S represents the number of sane block containing no outliers, and $\eta(\varepsilon)$ is a possitive fraction function for sane block such that $b_S \geq \eta(\varepsilon)b$. For $\epsilon > 0$, in fact, if

$$\textstyle\sum_{k\in[b_S]} 1_{\{|\mathrm{P}_m^{B_k}Y - \mathrm{E}Y| > \epsilon\}} \leq b_S - b/2, \text{ then } |\mathrm{MOM}_b[Y] - \mathrm{E}Y| \leq \epsilon.$$

The reason is that if at least $b/2$ sane block $\{B_k\}$ s.t. $|\mathrm{P}_m^{B_k} Y - \mathrm{E}Y| \le \epsilon$, then $|\operatorname{MOM}_b[Y] - \mathrm{E}Y| \le \epsilon$. We have

$$\{|\operatorname{MOM}_b[Y] - \mathrm{E}Y| \le \epsilon\} \supset \left\{ \left|\left\{k \in [b_S] : |\mathrm{P}_m^{B_k} Y - \mathrm{E}Y| \le \epsilon\right\}\right| \ge \frac{b}{2} \right\} = \left\{ \sum_{k\in[b_S]} 1_{\left\{|\mathrm{P}_m^{B_k} Y - \mathrm{E}Y| > \epsilon\right\}} \le b_S - \frac{b}{2} \right\}.$$

Then,

$$\begin{aligned}
&\mathrm{P}\{|\operatorname{MOM}_b[Y] - \mathrm{E}Y| \le \epsilon\} \ge \mathrm{P}\left\{ \sum_{k\in[b_S]} 1_{\left\{|\mathrm{P}_m^{B_k} Y - \mathrm{E}Y| > \epsilon\right\}} \le b_S - \frac{b}{2} \right\} \\
&= \mathrm{P}\left\{ \sum_{s\in[b_S]} [1_{\left\{|\mathrm{P}_m^{B_s} Y - \mathrm{E}Y| > \epsilon\right\}} - \mathrm{P}\left\{|\mathrm{P}_m^{B_s} Y - \mathrm{E}Y| > \epsilon\right\}] < b_S - \frac{b}{2} - b_S \mathrm{P}\left\{|\mathrm{P}_m^{B_s} Y - \mathrm{E}Y| > \epsilon\right\} \right\} \\
&\ge \mathrm{P}\left\{ \sum_{s\in[b_S]} [1_{\left\{|\mathrm{P}_m^{B_s} Y - \mathrm{E}Y| > \epsilon\right\}} - \mathrm{P}\left\{|\mathrm{P}_m^{B_s} Y - \mathrm{E}Y| > \epsilon\right\}] < b_S[1 - \frac{1}{2\eta(\varepsilon)} - \mathrm{P}\left\{|\mathrm{P}_m^{B_s} Y - \mathrm{E}Y| > \epsilon\right\}] \right\}, \qquad (27)
\end{aligned}$$

where the last inequality is by (M.2): $-\frac{b}{2} \ge -\frac{b_S}{2\eta(\varepsilon)}$. In (27), Chebyshev's inequality implies

$$\mathrm{P}\{|\mathrm{P}_m^{B_s} Y - \mathrm{E}Y| \ge 2\sqrt{\frac{\eta(\varepsilon)\mathrm{Var}Y}{m}}\} \le \mathrm{P}\{|\mathrm{P}_m^{B_s} Y - \mathrm{E}Y| \ge 2\sqrt{\frac{\eta(\varepsilon)\mathrm{Var}Y}{m}}\} \le \frac{1}{4\eta(\varepsilon)}.$$

Let $\epsilon = 2\sqrt{\frac{\eta(\varepsilon)\mathrm{Var}Y}{m}}$, and the last inequality shows

$$\mathrm{P}\left\{|\operatorname{MOM}_b[Y] - \mathrm{E}Y| \le \epsilon\right\} \ge \mathrm{P}\left\{ \sum_{s\in[b_S]} [1_{\left\{|\mathrm{P}_m^{B_s} Y - \mathrm{E}Y| > \epsilon\right\}} - \mathrm{P}\left\{|\mathrm{P}_m^{B_s} Y - \mathrm{E}Y| > \epsilon\right\}] < b_S[1 - \frac{3}{4\eta(\varepsilon)}] \right\}.$$

Since $\{1_{\{|\mathrm{P}_m^{B_s} Y - \mathrm{E}Y| > \epsilon\}}\}_{s\in[b_S]}$ are independent r.v. which is bounded by 1, Hoeffding's inequality shows

$$\mathrm{P}\left\{ \sum_{s\in[b_S]} [1_{\left\{|\mathrm{P}_m^{B_s} Y - \mathrm{E}Y| > \epsilon\right\}} - \mathrm{P}\left\{|\mathrm{P}_m^{B_s} Y - \mathrm{E}Y| > \epsilon\right\}] < b_S[1 - \frac{3}{4\eta(\varepsilon)}] \right\} \ge 1 - e^{-2\frac{[b_S(1-\frac{3}{4\eta(\varepsilon)})]^2}{\sum_{s=1}^{b_S}(1-0)^2}} = 1 - e^{-2b_S(1-\frac{3}{4\eta(\varepsilon)})^2}.$$

Therefore, we have

$$\mathrm{P}\left\{|\operatorname{MOM}_b[Y] - \mathrm{E}Y| \ge 2\sqrt{\frac{\eta(\varepsilon)\mathrm{Var}Y}{m}}\right\} \le e^{-2b_S(1-\frac{3}{4\eta(\varepsilon)})^2} \le e^{-2\eta(\varepsilon)b(1-\frac{3}{4\eta(\varepsilon)})^2}, \qquad (28)$$

where the last inequality is from $b_S \ge \eta(\varepsilon)b$.

Next, recall that

$$\|X\|_G = \max_{1\le k\le k_X}\left[\frac{\mathrm{E}X^{2k}}{(2k-1)!!}\right]^{1/(2k)} = \max_{1\le k\le \kappa_n}\left[\frac{\mathrm{E}X^{2k}}{(2k-1)!!}\right]^{1/(2k)} \text{ for any } \kappa_n \ge k_X \qquad (29)$$

and $\widehat{\|X\|}_{b,G} = \max_{1\le k\le k_X} \operatorname{med}_{s\in[b]}\{[\frac{1}{(2k-1)!!}\cdot \mathrm{P}_m^{B_s} X^{2k}]^{1/(2k)}\}$. Recall that $\underline{g}_{k,m}(\sigma_k)$ and $\bar{g}_{k,m}(\sigma_k)$ are the sequences s.t.

$$\left[\mathrm{E}X^{2k}/(2k-1)!!\right]^{1/(2k)}(1-\bar{g}_{k,m}(\sigma_k)) = \max_{1\le k\le\kappa_n}\left[-2[m/\eta(\varepsilon)]^{-1/2}\sigma_k^k/(\mathrm{E}X^{2k}) + \mathrm{E}X^{2k}/(2k-1)!!\right]^{1/(2k)}; \qquad (30)$$

$$[2[m/\eta(\varepsilon)]^{-1/2}\sigma_k^k/(\mathrm{E}X^{2k}) + 1]^{1/(2k)} = 1 + \underline{g}_{k,m}(\sigma_k) \text{ for any } m \in \mathbb{N} \text{ and } 1 \le k \le \kappa_n \text{ respectively.} \qquad (31)$$

For the first inequality, we have by (29)

$$\begin{aligned}
&\mathrm{P}\left\{\widehat{\|X\|}_{b,G} \leq [1 - \max_{1\leq k\leq \kappa_n} \bar{g}_{k,m}(\sigma_k)]\|X\|_G\right\} = \mathrm{P}\left\{\widehat{\|X\|}_{b,G} \leq \max_{1\leq k\leq \kappa_n}\left[\frac{\mathrm{E}X^{2k}}{(2k-1)!!}\right]^{1/(2k)}(1 - \max_{1\leq k\leq \kappa_n} \bar{g}_{k,m}(\sigma_k))\right\} \\
&\leq \mathrm{P}\left\{\widehat{\|X\|}_{b,G} \leq \max_{1\leq k\leq \kappa_n}\left[\frac{\mathrm{E}X^{2k}}{(2k-1)!!}\right]^{1/(2k)}(1 - \bar{g}_{k,m}(\sigma_k))\right\} \\
\text{[By (30)]}\ &= \mathrm{P}\left\{\widehat{\|X\|}_{b,G} \leq \left[-\frac{\sigma_k^k}{(2k-1)!!}\cdot\frac{2}{[m/\eta(\varepsilon)]^{1/2}} + \frac{\mathrm{E}X^{2k}}{(2k-1)!!}\right]^{1/(2k)}\right\} \\
&\leq \sum_{k=1}^{\kappa_n}\mathrm{P}\left\{\operatorname*{med}_{s\in[b]}\{[\frac{1}{(2k-1)!!}\cdot \mathrm{P}_m^{B_s}X^{2k}]^{1/(2k)}\} \leq \left[\frac{\sigma_k^k}{(2k-1)!!}\cdot\frac{2}{[m/\eta(\varepsilon)]^{1/2}} + \frac{\mathrm{E}X^{2k}}{!k}\right]^{1/(2k)}\right\} \\
&= \sum_{k=1}^{\kappa_n}\mathrm{P}\left\{\operatorname*{med}_{s\in[b]}\{\frac{1}{(2k-1)!!}\cdot \mathrm{P}_m^{B_s}X^{2k}\} \leq \frac{\mathrm{E}X^{2k}}{(2k-1)!!} - \frac{\sigma_k^k}{(2k-1)!!}\cdot\frac{2}{[m/\eta(\varepsilon)]^{1/2}}\right\} \\
&= \sum_{k=1}^{\kappa_n}\mathrm{P}\left\{\operatorname*{med}_{s\in[b]}\{\frac{1}{(2k-1)!!}\cdot [\mathrm{P}_m^{B_s}X^{2k} - \mathrm{E}X^{2k}]\} \leq -\frac{\sigma_k^k}{(2k-1)!!}\cdot\frac{2}{[m/\eta(\varepsilon)]^{1/2}}\right\} \\
&< \sum_{k=1}^{\kappa_n}\mathrm{P}\left\{|\operatorname*{med}_{s\in[b]}\{\mathrm{P}_m^{B_s}[X^{2k} - \mathrm{E}X^{2k}]\}| \geq \sigma_k^k\cdot\frac{2}{[m/\eta(\varepsilon)]^{1/2}}\right\} \leq \kappa_n e^{-2\eta(\varepsilon)b(1-\frac{3}{4\eta(\varepsilon)})^2},
\end{aligned}$$

where the last inequality is by (28) with $Y_i = X_i^{2k}$; and the assumption that $\sqrt{\mathrm{Var}X^{2k}} \leq \sigma_k^k,\ 1\leq k\leq \kappa_n$.

Let $\underline{g}_m(\sigma) := \max_{1\leq k\leq \kappa_n}\underline{g}_{k,m}(\sigma_k)$. For the second inequality, the definition of $g_{k,m}(\sigma_k)$ implies

$$\begin{aligned}
&\mathrm{P}\left\{\|X\|_G < \frac{\widehat{\|X\|}_{b,G}}{1+\underline{g}_m(\sigma)}\right\} = \mathrm{P}\left\{\widehat{\|X\|}_{b,G} - \max_{1\leq k\leq \kappa_n}\left[\frac{\mathrm{E}X^{2k}}{(2k-1)!!}\right]^{1/(2k)} > \underline{g}_m(\sigma)\max_{1\leq k\leq \kappa_n}\left[\frac{\mathrm{E}X^{2k}}{(2k-1)!!}\right]^{1/(2k)}\right\} \\
&\leq \mathrm{P}\left\{\widehat{\|X\|}_{b,G} - \max_{1\leq k\leq \kappa_n}\left[\frac{\mathrm{E}X^{2k}}{(2k-1)!!}\right]^{1/(2k)} > \underline{g}_{k,m}(\sigma_k)\max_{1\leq k\leq \kappa_n}\left[\frac{\mathrm{E}X^{2k}}{(2k-1)!!}\right]^{1/(2k)}\right\} \\
&= \mathrm{P}\left\{\widehat{\|X\|}_{b,G} > \max_{1\leq k\leq \kappa_n}\left[\frac{\mathrm{E}X^{2k}}{(2k-1)!!}\right]^{1/(2k)}(1+\underline{g}_{k,m}(\sigma_k))\right\} \\
&= \mathrm{P}\left\{\widehat{\|X\|}_{b,G} > \max_{1\leq k\leq \kappa_n}\left[\frac{\mathrm{E}X^{2k}}{(2k-1)!!}\right]^{1/(2k)}\left[\frac{\sigma_k^k}{\mathrm{E}X^{2k}}\cdot\frac{2}{[m/\eta(\varepsilon)]^{1/2}}+1\right]^{1/(2k)}\right\} \\
&= \mathrm{P}\left\{\widehat{\|X\|}_{b,G} > \max_{1\leq k\leq \kappa_n}\left[\frac{\sigma_k^k}{(2k-1)!!}\cdot\frac{2}{[m/\eta(\varepsilon)]^{1/2}} + \frac{\mathrm{E}X^{2k}}{(2k-1)!!}\right]^{1/(2k)}\right\} \\
&\leq \mathrm{P}\left\{\max_{1\leq k\leq \kappa_n}\operatorname*{med}_{s\in[b]}\{[\frac{1}{(2k-1)!!}\cdot \mathrm{P}_m^{B_s}X^{2k}]^{1/(2k)}\} > \left[\frac{\sigma_k^k}{(2k-1)!!}\cdot\frac{2}{[m/\eta(\varepsilon)]^{1/2}} + \frac{\mathrm{E}X^{2k}}{(2k-1)!!}\right]^{1/(2k)}\right\} \\
&\leq \sum_{k=1}^{\kappa_n}\mathrm{P}\left\{\operatorname*{med}_{s\in[b]}\left\{\left[\frac{1}{(2k-1)!!}\cdot \mathrm{P}_m^{B_s}X^{2k}\right]^{1/2k}\right\} > \left[\frac{\sigma_k^k}{(2k-1)!!}\cdot\frac{2}{[m/\eta(\varepsilon)]^{1/2}} + \frac{\mathrm{E}X^{2k}}{(2k-1)!!}\right]^{1/(2k)}\right\} \\
&= \sum_{k=1}^{\kappa_n}\mathrm{P}\left\{\operatorname*{med}_{s\in[b]}\left\{\frac{1}{(2k-1)!!}\cdot \mathrm{P}_m^{B_s}X^{2k}\right\} > \frac{\sigma_k^k}{(2k-1)!!}\cdot\frac{2}{[m/\eta(\varepsilon)]^{1/2}} + \frac{\mathrm{E}X^{2k}}{(2k-1)!!}\right\} \\
&= \sum_{k=1}^{\kappa_n}\mathrm{P}\left\{\operatorname*{med}_{s\in[b]}\left\{\mathrm{P}_m^{B_s}X^{2k}\right\} > \frac{2\sigma_k^k}{[m/\eta(\varepsilon)]^{1/2}} + \mathrm{E}X^{2k}\right\} = \sum_{k=1}^{\kappa_n}\mathrm{P}\left\{\operatorname*{med}_{s\in[b]}\left\{\mathrm{P}_m^{B_s}\left[X^{2k} - \mathrm{E}X^{2k}\right]\right\} > \frac{2\sigma_k^k}{[m/\eta(\varepsilon)]^{1/2}}\right\} \\
\text{[By (28)]}\ &< \sum_{k=1}^{\kappa_n}\mathrm{P}\left\{|\operatorname*{med}_{s\in[b]}\{\mathrm{P}_m^{B_s}[X^{2k} - \mathrm{E}X^{2k}]\}| \geq \sigma_k^k\cdot\frac{2}{[m/\eta(\varepsilon)]^{1/2}}\right\} \leq \kappa_n e^{-2\eta(\varepsilon)b(1-\frac{3}{4\eta(\varepsilon)})^2},
\end{aligned}$$

where the last inequality stems from $\sqrt{\mathrm{Var}X^{2k}} \leq \sigma_k^k,\ 1 \leq k \leq \kappa_n$ for a sequence $\{\sigma_k\}_{k\geq 1}$. $\square$

Proof of Theorem 4: Denote $Y_{k,i}$ represents the i-th value of reward of arm k in its history, where $i \in [T_k(t)]$ for any round t. It is needed to bound $\mathrm{E}T_k(t)$ as one has

$$\mathrm{E}\,\mathrm{Reg}_T \leq \sum_{k=2}^{K} \Delta_k \mathrm{E}T_k(t),$$

from (10), so we can only focus on some fixed arm. Hence, we can just drop the subscript k in $\{Y_{k,i}\}_{i=1}^{T_k(t)}$ as $\{Y_i\}_{i=1}^{T_k(t)}$.

We first give a lemma, which is crucial in the following proof.

Lemma 6. *Let $\{Y_i\}_{i=1}^n$ be independent r.v.s with $\mu_i = \mathrm{E}Y_i$, and assume that $Y_i - \mu_i$ is symmetric around zero with $\|Y_i - \mu_i\|_G \leq C$ and $\{w_i\}_{i=1}^n$ are i.i.d. Rademacher r.v.s independent of $\{Y_i\}_{i=1}^n$. Let $\overline{w} := \frac{1}{n}\sum_{i=1}^n w_i$. Then,*

$$\mathrm{P}\left(\frac{1}{n}\sum_{i=1}^{n}(w_i - \overline{w})(Y_i - \mu_i) \leq C\sqrt{\frac{2\log(1/\alpha)}{n}}\right) \geq 1 - \alpha.$$

Proof of Lemma 6. From Theorem 2 , we know that under the conditions in the lemma,

$$\mathrm{P}\left(\frac{1}{n}\sum_{i=1}^{n} a_i(Y_i - \mu_i) \leq C\|a\|_2 \frac{\sqrt{2\log(1/\alpha)}}{n}\right) \geq 1 - \alpha \tag{32}$$

for any vector $a := (a_1, \ldots, a_n)^\top \in \mathbb{R}^n$.

On the other hand, we have following inequalities,

$$\|w - \overline{w}\|_2^2 := \sum_{i=1}^{n}(w_i - \overline{w})^2 = \sum_{i=1}^{n} w_i^2 - n\overline{w}^2 = n(1 - \overline{w}^2) \leq n, \text{ and } \|w - \overline{w}\|_\infty \leq 2. \tag{33}$$

Therefore, one has

$$\mathrm{P}\left(\frac{1}{n}\sum_{i=1}^{n}(w_i - \overline{w})(Y_i - \mu_i) \leq C\sqrt{\frac{2\log(1/\alpha)}{n}}\right) = \mathrm{P}_w\mathrm{P}_y\left(\frac{1}{n}\sum_{i=1}^{n}(w_i - \overline{w})(Y_i - \mu_i) \leq C\sqrt{\frac{2\log(1/\alpha)}{n}}\right)$$
$$\geq \mathrm{P}_w\mathrm{P}_y\left(\frac{1}{n}\sum_{i=1}^{n}(w_i - \overline{w})(Y_i - \mu_i) \leq C\|w - \overline{w}\|_2 \frac{\sqrt{2\log(1/\alpha)}}{n}\right) \geq 1 - \alpha,$$

where the second inequality is by (33) and the last inequality applies (32). $\square$

Based on Lemma 6, next we can prove Theorem 4. We first state the assumptions for Theorem 4 in detail.

(UCB1) The rewards of k-the arm in round t, Y_k. And $\sqrt{\mathrm{Var}Y_k^{2\kappa}} \leq \sigma_{k,\kappa}^{\kappa},\ 1 \leq \kappa \leq k_{Y_k}$ for a sequence $\{\sigma_{k,\kappa}\}_{\kappa\geq 1}$;

(UCB2) For any $k \in [K]$

$$\overline{g}_{\kappa,m}(\sigma_{k,\kappa}) := 1 - \left[\mathrm{E}Y_k^{2\kappa}/(2\kappa-1)!!\right]^{-1/(2\kappa)} \max_{1\leq\kappa\leq\kappa_{Y_k}} \left[-2m^{-1/2}\sigma_{k,\kappa}^{\kappa}/(\mathrm{E}Y_k^{2\kappa}) + \mathrm{E}Y_k^{2\kappa}/(2\kappa-1)!!\right]^{1/(2\kappa)}$$

and

$$\underline{g}_{\kappa,m}(\sigma_{k,\kappa}) := [2m^{-1/2}\sigma_{k,\kappa}^{\kappa}/(\mathrm{E}Y_k^{2\kappa}) + 1]^{1/(2\kappa)} - 1$$

are both less than $n^{-1/2}$ for sufficient large m.

Proof of Theorem 4. Denote the population version of

$$\widehat{\varphi}_G(\mathbf{Y}_n) = \sqrt{\tfrac{2\log(4/\alpha)}{n}}\tfrac{\|\widehat{Y-\mu}\|_{b_k,G}}{1-n^{-1/2}} \quad \text{for} \quad \varphi_G(\mathbf{Y}_n) := \|Y-\mu\|_G\sqrt{\tfrac{2\log(4/\alpha)}{n}} \quad \text{with} \quad \mathrm{E}Y = \mu.$$

Theorem 2 gives

$$\mathrm{P}\big(|\overline{Y}-\mu| \geq \varphi_G(\mathbf{Y}_n)\big) \leq 2\exp\left\{-\frac{n^2\varphi_G^2(\mathbf{Y}_n)}{2n\|Y-\mu\|_G^2}\right\} = \alpha/2.$$

by the fact that $Y-\mu$ is symmetric around zero.

From Theorem 3, we take $b \geq 8\log(k_Y/\alpha)$ and define the event $\mathcal{E}_Y$ bellow associated with the MOM estimation of the intrinsic moment norm

$$\mathcal{E}_Y := \{\|\widehat{Y-\mu}\|_{b,G} \geq [1-\max_{1\leq k\leq k_Y/2}\overline{g}_{k,m}(\sigma_k)]\|Y-\mu\|_G\}$$

with probability at least $1-\alpha/2$. Note that

$$\begin{aligned}
&\mathrm{P}\big(|\overline{Y}-\mu| \geq \widehat{\varphi}_G(\mathbf{Y}_n)\big) \leq \mathrm{P}\big(|\overline{Y}-\mu| \geq \widehat{\varphi}_G(\mathbf{Y}_n),\, \mathcal{E}_Y\big) + \mathrm{P}\big(\mathcal{E}_Y^c\big)\\
&\leq \mathrm{P}\left(|\overline{Y}-\mu| \geq \sqrt{\frac{2\log(4/\alpha)}{n}}(1-n^{-1/2})^{-1}\|\widehat{Y-\mu}\|_{b,G},\, \mathcal{E}_Y\right) + \alpha/2\\
&\leq \mathrm{P}\left(|\overline{Y}-\mu| \geq (1-n^{-1/2})^{-1}[1-\max_{1\leq k\leq k_Y/2}\overline{g}_{k,m}(\sigma_k)]\sqrt{\frac{2\log(4/\alpha)}{n}}\|Y-\mu\|_G\right) + \alpha/2\\
&\leq \mathrm{P}\big(|\overline{Y}-\mu| \geq \varphi_G(\mathbf{Y}_n)) + \alpha/2 \leq \alpha.
\end{aligned}$$

where the last inequality is by taking m big enough such that $\max_{1\leq k\leq k_Y/2}\overline{g}_{k,m}(\sigma_k) \leq 1/\sqrt{n}$. Then, for $s \in \mathbb{N}_+$,

$$\mathrm{P}\big(|\overline{Y}_s-\mu_k| \leq \widehat{\varphi}_G(\mathbf{Y}_s)\big) \geq 1-\alpha.$$

Now for any $k \in [K]$ and fixed $T_k(t) = s$, we know that

$$\mathrm{P}\big(\overline{Y}_s-\mu_k \geq \widehat{\varphi}_G(\mathbf{Y}_s)\big) \leq \mathrm{P}\big(|\overline{Y}_k-\mu_k| \geq \widehat{\varphi}_G(\mathbf{Y}_n)\big) \leq \alpha.$$

By the non-asymptotic second-order correction (see Theorem 2.2 in Hao et al. (2019)) and the assumption that $Y-\mu$ is symmetric around zero, one has

$$\mathrm{P}\left\{\mu_k-\overline{Y}_s \geq q_{\alpha/2}(\mathbf{Y}_s-\overline{Y}_s) + \sqrt{\frac{2\log(4/\alpha)}{s}}\widehat{\varphi}_G(\mathbf{Y}_s)\right\} \leq 2\alpha,$$

where $q_{\alpha/2}(\mathbf{Y}_{B_k}-\overline{Y}_{B_k}) := q_{\alpha/2}\left(\mathbf{Y}_s-\overline{Y}_s, \frac{1}{s}\mathbf{1}_s\right)$.

Denote the UCB index $\mathrm{UCB}_k(t) = \overline{Y}_{T_k(t)} + h_\alpha(\mathbf{Y}_{T_k(t)})$, and the good event

$$\mathcal{E}_k := \{\mu_1 < \min_{t\in[T]}\mathrm{UCB}_1(t)\} \cap \left\{\overline{Y}_{B_k} + q_{\alpha/2}(\mathbf{Y}_s-\overline{Y}_s) + \sqrt{2\log(4/\alpha)/s}\cdot\widehat{\varphi}_G(\mathbf{Y}_s) < \mu_1\right\}, \qquad k \in [K],$$

where $B_k \in [T]$ is a constant to be chosen later. Following from the proof in (B.16)-(B.18) of Hao et al. (2019), we can gives that $T_k(t) \leq B_k$ and

$$\mathrm{E}T_k(t) \leq B_k + T\Big[2\alpha T + \mathrm{P}\big(\overline{Y}_{B_k} + q_{\alpha/2}(\mathbf{Y}_{B_k}-\overline{Y}_{B_k}) + \sqrt{2\log(4/\alpha)/B_k}\cdot\widehat{\varphi}_G(\mathbf{Y}_{B_k}) \geq \mu_1\big)\Big]. \tag{34}$$

On the other hand, from Lemma 6, then

$$\mathrm{P}\left(q_{\alpha/2}(\mathbf{Y}_{B_k}-\overline{Y}_{B_k}) \geq C\sqrt{\frac{2\log(4/\alpha)}{B_k}}\right) \leq \alpha/2.$$

Next, by applying MOM estimator, we need to the following assumptions for block $\{b_k\}_{k\in[K]}$ corresponding to Theorem 3. Here we include the subscript k to avoid confusion.

Under $\eta(\varepsilon) = 1$ with $\varepsilon = 0$, Theorem 3 ensures for $b_k \geq 8\log(k_Y/\alpha)$ and take m large enough such that $\max_{1\leq\kappa\leq k_Y} \underline{g}_{\kappa,m}(\sigma_\kappa) \leq n^{-1/2}$, then

$$
\begin{aligned}
&\mathrm{P}\big((1+n^{-1/2})C \leq \|\widehat{Y-\mu}\|_{b,G}\big) \\
\leq{}& \mathrm{P}\big((1+\max_{1\leq\kappa\leq k_Y/2} \underline{g}_{\kappa,m}(\sigma_{k,\kappa}))\,\|Y-\mu\|_G \leq \|\widehat{Y-\mu}\|_{b_k,G}\big) \leq k_Y e^{-b_k/8} \leq \alpha/2
\end{aligned}
$$

Hence, we have

$$
\begin{aligned}
&\mathrm{P}\Bigg(q_{\alpha/2}(\mathbf{Y}_{B_k}-\overline{Y}_{B_k}) + \sqrt{\frac{2\log(4/\alpha)}{B_k}}\widehat{\varphi}_G(\mathbf{Y}_{B_k}) \geq C\left[1+\sqrt{\frac{2\log(4/\alpha)}{B_k}}\frac{1+{B_k}^{-1/2}}{1-{B_k}^{-1/2}}\right]\sqrt{\frac{2\log(4/\alpha)}{B_k}}\Bigg) \\
\leq{}&\mathrm{P}\Bigg(q_{\alpha/2}(\mathbf{Y}_{B_k}-\overline{Y}_{B_k}) + \frac{2\log(4/\alpha)}{B_k}\frac{\|\widehat{Y-\mu}\|_{b_k,G}}{1-{B_k}^{-1/2}} \geq C\sqrt{\frac{2\log(4/\alpha)}{B_k}} + \frac{2\log(4/\alpha)}{B_k}\frac{(1+{B_k}^{-1/2})C}{1-B_k^{-1/2}}\Bigg) \\
\leq{}&\mathrm{P}\Bigg(q_{\alpha/2}(\mathbf{Y}_{B_k}-\overline{Y}_{B_k}) \geq C\sqrt{\frac{2\log(4/\alpha)}{B_k}}\Bigg) + \mathrm{P}\big(\|\widehat{Y-\mu}\|_{b_k,G} \geq (1+{B_k}^{-1/2})C\big)
\end{aligned}
$$

which implies with probability at least $1-\alpha$,

$$
\begin{aligned}
q_{\alpha/2}(\mathbf{Y}_{B_k}-\overline{Y}_{B_k}) + \sqrt{\frac{2\log(4/\alpha)}{B_k}}\widehat{\varphi}_G(\mathbf{Y}_{B_k}) &\leq C\left[1+\sqrt{\frac{2\log(4/\alpha)}{B_k}}\frac{1+{B_k}^{-1/2}}{1-B_k^{-1/2}}\right]\sqrt{\frac{2\log(4/\alpha)}{B_k}} \\
&\leq 2(2+\sqrt{2})C\sqrt{\frac{2\log(4/\alpha)}{B_k}}
\end{aligned}
$$

where $B_k \geq 8\log(k_{Y_k}/\alpha) \vee 2\log(4/\alpha)$ and $\max_{1\leq\kappa\leq k_{Y_k}/2} \overline{g}_{\kappa,m}(\sigma_{k,\kappa}) \leq 1/\sqrt{B_k}$ for each arm k.

Now, define the event $\mathcal{B}_k := \{q_{\alpha/2}(\mathbf{Y}_{B_k}-\overline{Y}_{B_k}) + \sqrt{2\log(4/\alpha)/B_k}\cdot\widehat{\varphi}_G(\mathbf{Y}_{B_k}) \leq \Delta_k/2\}$ with $\Delta_k := \mu_1-\mu_k$. Choose B_k as

$$
B_k = \frac{4^2(2+\sqrt{2})^2C^2}{\Delta_k^2}\log(4/\alpha) \geq 2, \tag{35}
$$

we have

$$
\begin{aligned}
\mathrm{P}\big(\mathcal{B}_k^c\big) &= \mathrm{P}\big(q_{\alpha/2}(\mathbf{Y}_{B_k}-\overline{Y}_{B_k}) + \sqrt{2\log(4/\alpha)/B_k}\cdot\widehat{\varphi}_G(\mathbf{Y}_{B_k}) > \Delta_k/2\big) \\
&\leq \mathrm{P}\Bigg(2(2+\sqrt{2})C\sqrt{\frac{\log(4/\alpha)}{B_k}} > \Delta_k/2\Bigg) + \alpha = 0 + \alpha = \alpha.
\end{aligned}
$$

Applying Theorem 2 for concentration of $\overline{Y}_{B_k}-\mu_k$ when B_k is chosen as in (35),

$$
\begin{aligned}
\mathrm{P}\big(\overline{Y}_{B_k} + q_{\alpha/2}(\mathbf{Y}_{B_k}-\overline{Y}_{B_k}) + \sqrt{2\log(4/\alpha)/B_k}\cdot\widehat{\varphi}_G(\mathbf{Y}_{B_k}) \geq \mu_1\big) &\leq \mathrm{P}\big(\overline{Y}_{B_k}-\mu_k \geq \Delta_k/2\big) + \mathrm{P}\big(\mathcal{B}_k^c\big) \\
&\leq 2\exp\Big\{-\frac{(B_k\Delta_k/2)^2}{2\cdot B_kC^2}\Big\} + \alpha \\
&= 2\exp\Big\{-\frac{B_k\Delta_k^2}{8C^2}\Big\} + \alpha.
\end{aligned}
$$

Taking account these results into (34), we get that

$$
\begin{aligned}
\mathrm{E}T_k(t) &\leq B_k + 2\alpha T^2 + \alpha T + 2T \exp\Big\{ - \frac{B_k \Delta_k^2}{8C^2} \Big\} \\
&= \frac{4^2(2+\sqrt{2})^2 C^2}{\Delta_k^2} \log(4/\alpha) + 2\alpha T^2 + \alpha T + 2T \exp\Big\{ -2(2+\sqrt{2})^2 \log(4/\alpha) \Big\} \\
&= \frac{16(2+\sqrt{2})^2 C^2}{\Delta_k^2} \log T + \frac{4}{T} + \frac{2}{T^{25+16\sqrt{2}}} + 8
\end{aligned}
$$

by taking $\alpha = 4/T^2$. Under the problem-dependent case, the regret is bounded by

$$
\mathrm{Reg}_T = \sum_{k=2}^{K} \Delta_k \mathrm{E}T_k(t) \leq 16(2+\sqrt{2})^2 C^2 \log T \sum_{k=2}^{K} \Delta_k^{-1} + \left(\frac{4}{T} + \frac{2}{T^{25+16\sqrt{2}}} + 8 \right) \sum_{k=2}^{K} \Delta_k.
$$

To get the problem-independent bound, we let $\Delta > 0$ as an arbitrary threshold, then decompose Reg_T, we get

$$
\begin{aligned}
\mathrm{Reg}_T &= \sum_{\Delta_k : \Delta_k < \Delta} \Delta_k \mathrm{E}T_k(t) + \sum_{\Delta_k : \Delta_k \geq \Delta} \Delta_k \mathrm{E}T_k(t) \\
&\leq T\Delta + 16(2+\sqrt{2})^2 C^2 \log T \sum_{\Delta_k : \Delta_k \geq \Delta} \Delta_k^{-1} + \left(\frac{4}{T} + \frac{2}{T^{25+16\sqrt{2}}} + 8 \right) \sum_{\Delta_k : \Delta_k \geq \Delta} \Delta_k \\
&\leq T\Delta + \frac{16(2+\sqrt{2})^2 C^2 K \log T}{\Delta} + \left(\frac{4}{T} + \frac{2}{T^{25+16\sqrt{2}}} + 8 \right) K \mu_1^* \\
&= 8(\sqrt{2}+2) C \sqrt{TK \log T} + \left(\frac{4}{T} + \frac{2}{T^{25+16\sqrt{2}}} + 8 \right) K \mu_1^*,
\end{aligned}
$$

by taking $\Delta = 8(2+\sqrt{2}) C \sqrt{(K \log T)/T}$. And finally, we take $C = \max_{k \in [K]} \|Y_k - \mu_k\|_G$. $\square$