

# EFFICIENT CONCENTRATION WITH GAUSSIAN APPROXIMATION

BY MORGANE AUSTERN <sup>1,a</sup> AND LESTER MACKEY <sup>2,b</sup>

<sup>1</sup>*Department of Statistics, Harvard University*, <sup>a</sup>[morgane.austern@gmail.com](mailto:morgane.austern@gmail.com)

<sup>2</sup>*Microsoft Research New England*, <sup>b</sup>[lmackey@microsoft.com](mailto:lmackey@microsoft.com)

Concentration inequalities for the sample mean, like those due to Bernstein, Hoeffding, and Bentkus, are valid for any sample size but overly conservative, yielding confidence intervals that are unnecessarily wide. The central limit theorem (CLT) provides asymptotic confidence intervals with optimal width, but these are invalid for all sample sizes. To resolve this tension, we develop new computable concentration inequalities for bounded variables with asymptotically optimal size, finite-sample validity, and sub-Gaussian decay. These bounds enable the construction of efficient confidence intervals with correct coverage for any sample size and efficient empirical Berry-Esseen bounds that require no prior knowledge of the population variance. We derive our inequalities by tightly bounding non-uniform Kolmogorov and Wasserstein distances to a Gaussian using zero-bias couplings and Stein's method of exchangeable pairs and demonstrate practical improvements over the Bernstein, Hoeffding, Bentkus, Berry-Esseen, Feller-Cramér, Romano-Wolf, empirical Bernstein, empirical Bentkus, and coin-betting inequalities.

**1. Introduction.** Concentration inequalities for the sample mean are ubiquitous in probability theory, statistics, and machine learning. Given  $n$  observations from an infinite sequence of independent and identically distributed (i.i.d.) random variables  $(W_i)_{i=1}^\infty$ , they allow us to give finite-sample and high-probability guarantees that the sample mean  $\bar{W}_n \triangleq \frac{1}{n} \sum_{i=1}^n W_i$  is close to the population mean  $\mathbb{E}[W_1]$ . More specifically, they provide upper bounds for the probability  $\mathbb{P}(\bar{W}_n - \mathbb{E}[W_1] > t/\sqrt{n})$  for each  $t \geq 0$ . Such inequalities lie at the heart of decision-making in reinforcement learning [3], generalization guarantees in high-dimensional statistics, machine learning, and deep learning [5, 68, 70], and the design [49] and selection [50] of efficient learning procedures.

However, standard concentration inequalities are overly conservative yielding confidence intervals that are unnecessarily wide and generalization guarantees that are weaker than needed. This is notably the case for the commonly used concentration inequalities of Hoeffding [38] and Bernstein [12]. For bounded random variables  $W_i \in [0, R]$  with variance  $\sigma^2 \triangleq \text{Var}(W_i) > 0$ , the Hoeffding and Bernstein inequalities respectively state that the scaled deviation  $S_n \triangleq \sqrt{n}(\bar{W}_n - \mathbb{E}[W_1])$  satisfies

$$(1) \quad \mathbb{P}(S_n > \sigma u) \leq \exp\left(-\frac{2u^2\sigma^2}{R^2}\right) \quad \text{and} \quad \mathbb{P}(S_n > \sigma u) \leq \exp\left(-\frac{u^2}{2(1+Ru/(3\sigma\sqrt{n}))}\right), \quad \forall u \geq 0.$$

Meanwhile, the central limit theorem (CLT) identifies the exact limit for each tail probability:

$$(2) \quad \mathbb{P}(S_n > \sigma u) \xrightarrow{n \rightarrow \infty} \Phi^c(u) \quad \text{for all } u \in \mathbb{R},$$

where  $\Phi$  is the cumulative distribution function (CDF) of a standard normal distribution and  $\Phi^c = 1 - \Phi$ . As a result, standard confidence intervals based on the CLT are asymptotically

---

*MSC2020 subject classifications:* Primary 60F10, 62G15.

*Keywords and phrases:* Efficient concentration inequality, Gaussian approximation, tail bound, quantile bound, confidence interval, empirical Berry-Eseen bound, zero-bias coupling, Stein's method, non-uniform Kolmogorov distance, Wasserstein distance.

exact and often much narrower than those obtained using concentration inequalities. However, these intervals are typically only asymptotically valid and provide incorrect coverage for every sample size  $n$ .

The choice between loose but valid concentration inequalities and tight but invalid CLT-based approximations is very unsatisfying. In this paper we derive new bounds that offer the best of both worlds: our new concentration inequalities are both finite-sample valid and *efficient*—that is, asymptotically of minimal width when scaled by  $\sqrt{n}$ . For example, our primary result, Thm. 1, implies that, for all  $u \geq \delta_n \triangleq \frac{R}{\sigma\sqrt{n}}$ ,

$$(3) \quad \mathbb{P}(S_n > \sigma u) \leq \Phi^c(u) + \frac{R}{\sigma\sqrt{n}}\varphi(u - \delta_n) + \frac{R(4+2u)}{\sigma\sqrt{n}}e^{-\frac{\sigma^2}{R^2}(u - \frac{3}{2}\delta_n)^2_+}$$

where  $\varphi(u) \triangleq \frac{1}{\sqrt{2\pi}}e^{-u^2/2}$  is the Lebesgue density of a standard normal. Underlying this implication is a new, computable concentration inequality formed by explicitly bounding a non-uniform Kolmogorov distance between the sample mean and a Gaussian.

It is informative to compare the result (3) with those obtained using classical CLT corrections. The Berry-Esseen bound [28] guarantees that

$$(4) \quad \mathbb{P}(S_n > \sigma u) \leq \Phi^c(u) + \frac{C_{R,\sigma}}{\sqrt{n}} \quad \text{for all } u \geq 0$$

and a constant  $C_{R,\sigma}$  depending only on  $R$  and  $\sigma$ . This yields an efficient concentration inequality, but the bound is overly conservative as the correction is independent of  $u$ . Non-uniform Berry-Esseen bounds [14, 51] ameliorate this behavior by identifying a constant  $\tilde{C}_{R,\sigma}$  satisfying

$$(5) \quad \mathbb{P}(S_n > \sigma u) \leq \Phi^c(u) + \frac{\tilde{C}_{R,\sigma}}{\sqrt{n}(1+u)^3} \quad \text{for all } u \geq 0.$$

Appealingly, this non-uniformity yields tighter bounds for larger  $u$ . However, the correction has only cubic, that is,  $O(u^{-3})$ , decay in  $u$  as the underlying argument only exploits the existence of a third moment of  $\bar{W}_n$ . Quantile coupling inequalities [see, e.g., 48] like the groundbreaking Komlós-Major-Tusnády approximations [43, 44, Thm. 1] and the strong embedding bounds of Bhattacharjee and Goldstein [13], Chatterjee [19] improve this  $u$  dependence for  $W_1$  with finite exponential moments but provide at best exponential decay in  $u$  and  $O(\log n/\sqrt{n})$  decay in  $n$ . By exploiting the boundedness of  $W_1$ , our new efficient correction term (3) guarantees faster, sub-Gaussian  $e^{-\Omega(u^2)}$  decay and eliminates the extraneous  $\log n$  factor present in prior quantile coupling inequalities. The aforementioned quantile coupling and strong embedding bounds are also unsuitable for practical deployment due to their unidentified constants. Crucially for our applications, our Thm. 1 is fully computable, allowing us to develop practical efficient confidence regions in Secs. 5 and 6.

In Sec. 4, we supplement our primary result with a computational refinement that yields tighter tail and quantile bounds for larger deviations by carefully controlling the  $p$ -Wasserstein distance to Gaussianity. For example, our efficient Wasserstein tail bound, Thm. 5, improves the dependence on  $u$  at the expense of a worse dependence on  $n$  and provides an explicit relative error bound in the spirit of classical Cramér-type inequalities [20, 23, 30, 31, 54]:

$$\begin{aligned} \mathbb{P}(S_n > \sigma u) &\leq \Phi^c(u - \delta_{u,n}) + \frac{\varphi(u)}{\sqrt{n}} \quad \text{for } \delta_{u,n} \triangleq \frac{eK_{R,\sigma}}{\sqrt{n}} \lceil \log \left( \frac{\sqrt{n}}{\varphi(u)} \right) \rceil \\ &\leq \Phi^c(u) \cdot \left( e^{(u+1)\delta_{u,n}} + \frac{1}{\sqrt{n}} \frac{\varphi(u)}{\Phi^c(u)} \right) \quad \text{for all } u > \delta_{u,n}. \end{aligned}$$

Here,  $K_{R,\sigma}$  is an explicit constant depending only on  $(R, \sigma)$  that we define in Sec. 4. Unlike Thm. 5, most Cramér-type relative error bounds, including those derived by Cramér [23], Petrov [54], Chen, Fang and Shao [20], and Fang and Koike [30], are unsuitable for practical

use due to unidentified constants. A notable exception is the inequality of Feller [31, Thm. 1] which delivers the Cramér-type bound

$$(6) \quad \mathbb{P}(S_n > \sigma u) \leq \Phi^c(u) \exp\left(\frac{6u^3 c_{R,\sigma}}{7\sqrt{n} - 84u c_{R,\sigma}}\right) \left(1 + \frac{9\sqrt{2\pi} c_{R,\sigma}}{\sqrt{n}} \frac{\varphi(u)}{\Phi^c(u)}\right) \quad \text{for all } u \in (0, \frac{\sqrt{n}}{12c_{R,\sigma}})$$

for a constant  $c_{R,\sigma}$  depending only on  $(R, \sigma)$ . We will see in Sec. 6 that Thm. 5 provides a substantially tighter bound than the Feller-Cramér inequality (6) in all of our experimental settings.

Another important comparison is with the “missing-factor” bounds of Eaton [25, 26], Talagrand [67], Pinelis [55, 56, 58], and Bentkus [8–11], which imply tail decay proportional to  $\Phi^c(u) \sim \frac{1}{u} \varphi(u)$ . A representative example is the computable bound of Bentkus, Kalosha and Van Zuijlen [11, Thm. 2.1],

$$(7) \quad \mathbb{P}(S_n > \sigma u) \leq \inf_{t \in [0, u]} \frac{\mathbb{E}[(G_n - t)_+^2]}{(u - t)^2} \quad \text{for all } u > 0$$

where  $G_n \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^n \epsilon_i$  for i.i.d.  $(\epsilon_i)_{i \geq 1}$  with  $\mathbb{P}(\epsilon_i = \frac{R_\sigma}{\sigma}) = \frac{\sigma^2}{R_\sigma^2 + \sigma^2} = 1 - \mathbb{P}(\epsilon_i = -\frac{\sigma}{R_\sigma})$ ,

$$(R_\sigma) \quad R_\sigma \triangleq \frac{1}{2}R + \frac{1}{2}\sqrt{R^2 - 4\sigma^2},$$

and  $a_+ \triangleq \max(a, 0)$  for  $a \in \mathbb{R}$ . Like our new inequalities in Thms. 1 and 5, the Bentkus bound (7) eschews a closed-form to obtain a significantly tighter tail bound that is still straightforwardly computable [45, App. C]. However, unlike our new inequalities, the Bentkus bound (7) is inefficient with a limit at least twice as large as the ideal width  $\Phi^c(u)$ .

**PROPOSITION 1** (Inefficiency of Bentkus bound). *For each  $u > 0$ ,*

$$\inf_{t \in [0, u]} \frac{\mathbb{E}[(G_n - t)_+^2]}{(u - t)^2} \xrightarrow{n \rightarrow \infty} \inf_{t \in [0, u]} \frac{\mathbb{E}[(Z - t)_+^2]}{(u - t)^2} \geq 2\Phi^c(u).$$

**PROOF.** Fix any  $u > 0$  and any positive  $\epsilon < \mathbb{E}[(Z - u)_+^2]$ , where  $Z$  is a standard Gaussian. By the 2-Wasserstein CLT [16, Thm. 1], there exists an integer  $N$  such that  $\sup_{t \leq u} |\mathbb{E}[(G_n - t)_+^2] - \mathbb{E}[(Z - t)_+^2]| \leq \epsilon$  for all  $n \geq N$ . Now, fix any  $n \geq N$  and suppose

$$f_n(t) \triangleq \mathbb{E}[(G_n - t)_+^2]/(u - t)^2 \leq f_n(0) = \mathbb{E}[(G_n)_+^2]/u^2 \leq 1/u^2$$

for some  $t \in [0, u)$ . Then we necessarily have

$$(\mathbb{E}[(Z - u)_+^2] - \epsilon)/(u - t)^2 \leq (\mathbb{E}[(Z - t)_+^2] - \epsilon)/(u - t)^2 \leq f_n(t) \leq 1/u^2.$$

Parallel logic ensures  $f(t) \triangleq \frac{\mathbb{E}[(Z - t)_+^2]}{(u - t)^2} \leq f(0)$  only if  $\frac{1}{(u - t)^2} \leq \frac{1}{u^2 \mathbb{E}[(Z - u)_+^2]}$ . Therefore,

$$|\inf_{t \in [0, u]} f_n(t) - \inf_{t \in [0, u]} f(t)| \leq \frac{\epsilon}{u^2 (\mathbb{E}[(Z - u)_+^2] - \epsilon)}.$$

The result now follows from the arbitrariness of  $\epsilon$  and the inequality of [11, Thm. 7.1].  $\square$

In Sec. 5, we apply our new tools to develop practical *empirical Berry-Esseen bounds* that are efficient and finite-sample valid even when the variance parameter  $\sigma^2$  is unknown. In Sec. 6, we confirm numerically that our new bounds yield improvements over the Hoeffding, Bernstein, and Bentkus inequalities, the uniform and non-uniform Berry-Esseen corrections, the generalized Cramér bound of Feller, the popular but inefficient empirical Bernstein [3, 49, 50] and empirical Bentkus [45] bounds, and the efficient  $I_{n,3}$  confidence interval of Romano and Wolf [61]. We conclude with an application to Monte Carlo integration (Sec. 6.3) in which our efficient empirical Berry-Esseen bounds yield narrower confidence intervals than the state-of-the-art predictable plug-in and betting intervals of Waudby-Smith and Ramdas [69] for larger sample sizes. Sec. 7 presents a discussion of these results and related work.

**2. Defining Efficient Concentration.** To match the setting of the classical Hoeffding and Bernstein inequalities (1), we will focus on random variables satisfying the following distributional assumptions.

**ASSUMPTION  $(R, \sigma)$ .** *The scaled deviations  $(S_n)_{n \geq 1}$  satisfy  $S_n = \sqrt{n}(\bar{W}_n - \mathbb{E}[W_1])$  for i.i.d. variables  $(W_i)_{i=1}^\infty$  with  $\bar{W}_n \triangleq \frac{1}{n} \sum_{i=1}^n W_i$ ,  $\text{Var}(W_1) = \sigma^2 > 0$ , and  $W_1 \in [0, R]$  almost surely.*

Our first inferential goal is to tightly upper bound the tail probability  $\mathbb{P}(S_n > \sigma u)$  for a given threshold  $u \geq 0$ . The CLT provides an asymptotic lower bound for this problem as  $\mathbb{P}(S_n > \sigma u)$  is known to converge precisely to  $\Phi^c(u)$  as  $n$  increases.

**PROPOSITION 2** (Asymptotic lower bound for valid tail bounds). *Fix any  $u \geq 0$  and any sequence of candidate tail bounds  $(\delta_n(u))_{n \geq 1}$ . Under Assump.  $(R, \sigma)$ , if  $\mathbb{P}(S_n > \sigma u) \leq \delta_n(u)$  for all  $n$ , then  $\Phi^c(u) \leq \liminf_{n \rightarrow \infty} \delta_n(u)$ .*

**PROOF.** Suppose that a sequence  $\delta_n(u)$  satisfies  $\liminf_{n \rightarrow \infty} \delta_n(u) < \Phi^c(u)$ . Then there exists an  $\epsilon > 0$  such that  $\delta_n(u) \leq \Phi^c(u) - \epsilon$  for infinitely many  $n$ . However, by the CLT [see, e.g., 24, Thm. 3.4.1], there exists an  $n_\epsilon$  such that, for all  $n > n_\epsilon$ ,  $\Phi^c(u) - \epsilon < \mathbb{P}(S_n > \sigma u)$ . Therefore,  $\delta_n(u) < \mathbb{P}(S_n > \sigma u)$  for infinitely many  $n$ , confirming the claim via its contrapositive.  $\square$

Unfortunately, the CLT limit does not provide a suitable tail bound for any finite  $n$ . However, by tightly bounding the distance between the distribution of  $S_n$  and the distribution of a Gaussian we can correct the asymptotic bound to obtain one that is both valid in finite samples and asymptotically exact. We will call such bounds *efficient concentration inequalities*.

Our second inferential goal is to tightly bound the quantiles of  $S_n$ . That is, given a tail probability  $\delta \in (0, 1)$  we wish to find  $q_n(R, \delta, \sigma)$  (a measurable function of  $(W_i)_{i=1}^n$ ,  $R$ ,  $\sigma$ , and  $\delta$ ) satisfying  $\mathbb{P}(S_n > q_n(R, \delta, \sigma)) \leq \delta$ . Such quantile bounds immediately deliver both one- and two-sided confidence intervals for the population mean  $\mathbb{E}[W_1]$  as

$$\mathbb{P}(\bar{W}_n - \frac{q_n(R, \delta, \sigma)}{\sqrt{n}} \leq \mathbb{E}[W_1]) \wedge \mathbb{P}(|\mathbb{E}[W_1] - \bar{W}_n| \leq \frac{q_n(R, \delta/2, \sigma)}{\sqrt{n}}) \geq 1 - \delta.$$

The interval efficiency theory of Romano and Wolf [61] implies that the CLT once again provides an asymptotic lower bound for any valid sequence of quantile bounds.

**PROPOSITION 3** (Asymptotic lower bound for valid quantile bounds). *Fix any  $R, \sigma > 0$ , any  $\delta \in (0, \frac{1}{2})$ , and any nonnegative candidate quantile bounds  $(q_n(R, \delta, \sigma))_{n \geq 1, \sigma > 0}$ . Under Assump.  $(R, \sigma)$ , if  $\mathbb{P}(S_n > q_n(R, \delta, \sigma)) \leq \delta$  for all  $n$ , then  $(q_n(R, \delta, \sigma))_{n \geq 1}$  is not asymptotically concentrated<sup>1</sup> on  $[0, a]$  for any  $a < \sigma\Phi^{-1}(1 - \delta)$ . Here,  $\Phi^{-1}$  is the quantile function of a standard normal distribution.*

**PROOF.** The result follows by applying Thm. 2.1 of Romano and Wolf [61] to the conservative confidence intervals  $I_n(\delta, \sigma) = \bar{W}_n + \frac{1}{\sqrt{n}}[-q_n(R, \delta, \sigma), q_n(R, \delta, \sigma)]$  for the unknown mean  $\mathbb{E}[W_1]$ .  $\square$

To construct *efficient quantile bounds*, we will once again tightly bound the distance between  $S_n$  and its Gaussian limit.

---

<sup>1</sup>A sequence of nonnegative random variables  $(X_n)_{n \geq 1}$  is asymptotically concentrated on  $[0, a]$  if  $(X_n - a)_+ \xrightarrow{p} 0$ .

**2.1. Notation.** Hereafter, we will write  $a_n \lesssim b_n$  to indicate that two sequences  $(a_n)_{n \geq 1}$  and  $(b_n)_{n \geq 1}$  satisfy  $a_n \leq b_n + o_n(b_n)$  and use the shorthand  $\|X\|_\infty = \sup \text{support}(|X|)$  and  $\|X\|_p \triangleq \mathbb{E}[|X|^p]^{1/p}$  when a random variable  $X$  is bounded or has a  $p$ -th absolute moment for some  $p \geq 1$ . We will also make regular use of the parameter  $R_\sigma$  which provides a priori bounds on the summand mean  $\mathbb{E}[W_1]$  and deviation  $|W_1 - \mathbb{E}[W_1]|$ :

LEMMA 1 (Summand mean and deviation bounds). *Under Assump.  $(R, \sigma)$ , we have*

$$R - R_\sigma \leq \mathbb{E}[W_1] \leq R_\sigma \text{ and } |W_1 - \mathbb{E}[W_1]| \leq R_\sigma \text{ almost surely.}$$

PROOF. Since  $W_1 \in [0, R]$ , we have  $\mathbb{E}[W_1]^2 = \mathbb{E}[W_1^2] - \sigma^2 \leq R\mathbb{E}[W_1] - \sigma^2$ . Hence  $\mathbb{E}[W_1]$  must lie between the roots  $\frac{1}{2}R \pm \frac{1}{2}\sqrt{R^2 - 4\sigma^2}$  of this quadratic inequality.  $\square$

**3. Efficient Concentration with Zero-Bias Couplings.** To derive our initial efficient concentration inequality, we tightly bound a non-uniform Kolmogorov distance between the scaled deviation  $S_n$  and a Gaussian using *zero-bias couplings*. As in Chen, Goldstein and Shao [21, Prop. 2.1], we say that  $S^*$  has the *zero-bias distribution* for a mean-zero random variable  $S$  with  $\sigma^2 := \text{Var}(S) < \infty$  if the distribution of  $S^*$  is absolutely continuous with Lebesgue density  $p^*(x) = \mathbb{E}[S \mathbb{I}(S > x)]/\text{Var}(S)$  or, equivalently, if  $\sigma^2 \mathbb{E}[f'(S^*)] = \mathbb{E}[Sf(S)]$  for all absolutely continuous  $f$  with  $\mathbb{E}[|Sf(S)|] < \infty$ . Our primary result, proved in Sec. 3.1, uses a close coupling of  $S_n$  and its zero-biased counterpart to establish efficient concentration.

THEOREM 1 (Efficient zero-bias tail bounds). *Under Assump.  $(R, \sigma)$ , for all  $u \geq 0$  and  $\lambda \in [0, 1]$ ,*

$$\mathbb{P}(S_n > \sigma u + \frac{R}{\sqrt{n}}) \leq \Phi^c(u) + \frac{R}{\sigma\sqrt{n+R}u} [h_u(\lambda u) - u\Phi^c(u) + (h_u(u) - h_u(\lambda u)) Q_n(\lambda\sigma u)],$$

where, for  $\Delta_n \triangleq \frac{R_\sigma}{4\sqrt{n}}$  and all  $w \leq u$ ,

$$(8) \quad h_u(w) \triangleq (w + \frac{(1+w^2)\Phi(w)}{\varphi(w)})\Phi^c(u) \quad \text{and}$$

$$(9) \quad Q_n(u) \triangleq \min\left(e^{-\frac{2(u-\Delta_n)_+^2}{R^2}}, e^{-\frac{(u-\Delta_n)_+^2}{2(v_n^2 + R_\sigma(u-\Delta_n)/(3\sqrt{n}))}}, \Phi^c\left(\frac{u-\Delta_n}{\tilde{v}_n}\right) + \frac{0.56}{\sqrt{n}} \frac{R_\sigma v_n^2 + \beta_n}{\tilde{v}_n^3}\right),$$

for  $v_n^2 \triangleq \sigma^2 + \frac{1}{9n}(R_\sigma^2 - 6\sigma^2)$ ,  $\tilde{v}_n^2 \triangleq \sigma^2(1 - \frac{89}{144n})$ , and  $\beta_n \triangleq \min(\frac{1}{4}R_\sigma, R_{\frac{\sqrt{55}\sigma}{12}} - R_\sigma)(\frac{\sigma^2}{3n} + \frac{R_\sigma^2}{9n})$ .

Moreover, under the same conditions,

$$(10) \quad \mathbb{P}(S_n > \sigma u \frac{\sqrt{n+1}}{\sqrt{n}} + \frac{R_\sigma}{\sqrt{n}}) \leq \Phi^c(u) + \frac{R}{\sigma\sqrt{n+1+R}u} [h_u(\lambda u) - u\Phi^c(u) + (h_u(u) - h_u(\lambda u)) Q_{n+1}(\lambda\sigma u \frac{\sqrt{n}}{\sqrt{n+1}})].$$

Like the popular concentration inequalities due to Hoeffding and Bernstein (1), Thm. 1 is valid and computable for any sample size  $n$ . However, unlike the Hoeffding and Bernstein inequalities, Thm. 1 is also efficient and converges to the asymptotically exact Gaussian tail bound at a  $O(1/\sqrt{n})$  rate. In fact, as our next corollary demonstrates, the suboptimality of Thm. 1 also decays at a sub-Gaussian  $e^{-\Omega(u^2)}$  rate in  $u$ , faster than the more conservative Berry-Esseen (4), non-uniform Berry-Esseen (5), and quantile coupling inequalities [13, 18, 19, 43, 44, 48].

COROLLARY 1 (Efficient sub-Gaussian tail bound). *Under Assump.  $(R, \sigma)$ ,*

$$\mathbb{P}(S_n > \sigma u) \leq \Phi^c(u) + \frac{R}{\sigma\sqrt{n}}\varphi(u - \delta_n) + \frac{R(4+2u)}{\sigma\sqrt{n}} e^{-\frac{\sigma^2}{R^2}(u - \frac{3}{2}\delta_n)_+^2}$$

for all  $u \geq \delta_n \triangleq \frac{R}{\sigma\sqrt{n}}$ .

PROOF. Fix any  $u \geq \delta_n$ , introduce the shifted value  $v = u - \delta_n$ , and instantiate the notation of Thm. 1. For each  $w \geq 0$ , we have  $Q_n(w) \leq e^{-2(w - \Delta_n)_+^2/R^2}$  by (9) and

$$a_v(w) \triangleq \frac{(1+w^2)\Phi(w)}{\varphi(w)} \Phi^c(v) \leq \frac{2(1+w^2)\Phi(w)}{v + \sqrt{v^2 + 8/\pi}} e^{-\frac{v^2 - w^2}{2}} \leq (2+w)\Phi(w)e^{-\frac{v^2 - w^2}{2}}$$

by Abramowitz and Stegun [1, 7.1.13]. Invoking Thm. 1 with  $\lambda = 1/\sqrt{2}$  therefore yields

$$\begin{aligned} & \mathbb{P}(S_n > \sigma u) - \Phi^c(v) \\ & \leq \frac{R}{\sigma\sqrt{n+R}v} [(a_v(\lambda v) - (1-\lambda)v\Phi^c(v))(1 - Q_n(\lambda\sigma v)) + a_v(v)Q_n(\lambda\sigma v)] \\ & \leq \frac{R(2+v)\Phi(v)}{\sigma\sqrt{n+R}v} \left( e^{-\frac{(1-\lambda^2)v^2}{2}} + e^{-\frac{2(\lambda\sigma v - \Delta_n)_+^2}{R^2}} \right) = \frac{R(2+v)\Phi(v)}{\sigma\sqrt{n+R}v} \left( e^{-\frac{v^2}{4}} + e^{-\frac{(\sigma v - \sqrt{2}\Delta_n)_+^2}{R^2}} \right) \\ & \leq \frac{2R(2+v)\Phi(v)}{\sigma\sqrt{n+R}v} e^{-\frac{(\sigma v - \sqrt{2}\Delta_n)_+^2}{R^2}} \leq \frac{2R(2+u)}{\sigma\sqrt{n}} e^{-\frac{\sigma^2}{R^2}(u - \frac{3}{2}\delta_n)_+^2}. \end{aligned}$$

The advertised result now follows from the relation

$$\Phi^c(v) = \Phi^c(u) + \int_v^u \varphi(x)dx \leq \Phi^c(u) + (u - v)\varphi(v) = \Phi^c(u) + \delta_n\varphi(u - \delta_n).$$

□

While the bound in Thm. 1 is more complex than that in Cor. 1, it is straightforward to compute<sup>2</sup> and significantly tighter in practice. In Secs. 5 and 6, we will use our efficient concentration inequalities to develop efficient empirical Berry-Esseen bounds and efficient confidence intervals for Monte Carlo integration.

3.1. *Proof of Thm. 1: Efficient zero-bias tail bounds.* The proof of Thm. 1 relies on four auxiliary results. The first, proved in Sec. 3.2, provides a tail bound for a sum,  $T_n$ , of independent random variables in terms of a sum,  $T'_n$ , of interpolated zero-biased variables.

**THEOREM 2** (Unidentical zero-bias tail bound). *Suppose independent  $(V_i)_{i \geq 1}$  satisfy  $\mathbb{E}[V_i] = 0$ ,  $\text{Var}(V_i) = \sigma_i^2$ , and  $\sup \text{support}(V_i) - \inf \text{support}(V_i) \leq R$  for all  $i \in \mathbb{N}$ . For each  $n \in \mathbb{N}$ , define the scaled sum  $T_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n V_i$  with variance parameter  $\bar{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n \sigma_i^2$  and auxiliary variable*

$$T'_n \triangleq T_n + \frac{1}{\sqrt{n}}(V'_{I_n} - V_{I_n}) \text{ for } V'_i \triangleq V_i^* + U(V_i - V_i^*) \text{ and } \mathbb{P}(I_n = i) = \frac{1}{n} \frac{\sigma_i^2}{\bar{\sigma}_n^2} \mathbb{I}(1 \leq i \leq n),$$

where each  $V_i^*$  has the zero-bias distribution of  $V_i$ , and  $U \sim \text{Uniform}([0, 1])$ ,  $(I_n)_{n \geq 1}$ , and  $(V_i^*)_{i \geq 1}$  are mutually independent and independent of  $(V_i)_{i \geq 1}$ . Then, for all  $u \geq 0$  and  $\lambda \in [0, 1]$ ,

$$\begin{aligned} \mathbb{P}(T_n > \bar{\sigma}_n u + \frac{R}{\sqrt{n}}) & \leq \frac{\bar{\sigma}_n \sqrt{n}}{\bar{\sigma}_n \sqrt{n} + R u} \Phi^c(u) + \frac{R}{\bar{\sigma}_n \sqrt{n} + R u} (h_u(u) - h_u(\lambda u)) \mathbb{P}(T'_n > \lambda \bar{\sigma}_n u) \\ & \quad + \frac{R}{\bar{\sigma}_n \sqrt{n} + R u} h_u(\lambda u). \end{aligned}$$

The second result, also proved in Sec. 3.2, provides an alternative tail bound when the summands underlying  $T_n$  are i.i.d.

---

<sup>2</sup>See [https://github.com/lmackey/gauss\\_conc](https://github.com/lmackey/gauss_conc) for our open-source Python implementation.

**THEOREM 3** (Identical zero-bias tail bound). *Instantiate the notation and assumptions of Thm. 2. If  $(V_i)_{i=1}^{n+1}$  are i.i.d. with  $\sigma \triangleq \sigma_1$  and  $V_1 \geq -R'$  almost surely, then*

$$\begin{aligned} & \mathbb{P}(T_n > \sigma u \frac{\sqrt{n+1}}{\sqrt{n}} + \frac{R'}{\sqrt{n}}) \\ & \leq \frac{\sigma \sqrt{n+1}}{\sigma \sqrt{n+1} + R u} \Phi^c(u) + \frac{R}{\sigma \sqrt{n+1} + R u} (h_u(u) - h_u(\lambda u)) \mathbb{P}(T'_{n+1} > \lambda \sigma u \frac{\sqrt{n}}{\sqrt{n+1}}) \\ & \quad + \frac{R}{\sigma \sqrt{n+1} + R u} h_u(\lambda u). \end{aligned}$$

The third result bounds the moments of interpolated zero-biased variables.

**LEMMA 2** (Properties of  $V'_i$ ). *Instantiate the notation and assumptions of Thm. 2. For any  $i \in \mathbb{N}$ , if  $|V_i| \leq R'$  almost surely, then  $\sup \text{support}(V'_i) - \inf \text{support}(V'_i) \leq R$ ,*

$$\begin{aligned} |\mathbb{E}[V'_i]| & \leq \frac{R'}{4}, \quad \mathbb{E}[V'^2_i] \leq \frac{\sigma_i^2}{3} + \frac{R'^2}{9}, \quad \text{Var}(V'_i) \geq \frac{55\sigma_i^2}{144}, \\ |V'_i| & \leq R', \quad \text{and} \quad |V'_i - \mathbb{E}[V'_i]| \leq R_i \triangleq \min\left(\frac{1}{2}R + \frac{1}{2}\sqrt{R^2 - \frac{220}{144}\sigma_i^2}, \frac{5}{4}R'\right) \quad \text{almost surely.} \end{aligned}$$

**PROOF.** Fix any  $i \in \{1, \dots, n\}$ , and suppose  $|V_i| \leq R'$  almost surely. We invoke the definition of  $V'_i$ , Lem. 2.1(iv) of Goldstein and Reinert [34], and our boundedness assumption in turn to find that

$$|\mathbb{E}[V'_i]| = \frac{1}{2}|\mathbb{E}[V_i^*]| = \frac{1}{4\sigma_i^2}|\mathbb{E}[V_i^3]| \leq \frac{R'}{4}.$$

The same invocations, coupled with the independence of  $V_i, U$ , and  $V_i^*$ , imply

$$\mathbb{E}[V'^2_i] = \mathbb{E}[U^2]\mathbb{E}[V_i^2] + \mathbb{E}[(1-U)^2]\mathbb{E}[V_i^{*2}] = \frac{1}{3}\sigma_i^2 + \frac{1}{9\sigma_i^2}\mathbb{E}[V_i^4] \leq \frac{1}{3}\sigma_i^2 + \frac{1}{9}R'^2.$$

The same upper bound holds for  $\text{Var}(V'_i) = \mathbb{E}[V'^2_i] - \mathbb{E}[V'_i]^2$ . Moreover, since  $\mathbb{E}[V_i^3]^2 \leq \mathbb{E}[V_i^2]\mathbb{E}[V_i^4]$  by Cauchy-Schwarz and  $\mathbb{E}[V_i^2]^2 \leq \mathbb{E}[V_i^4]$  by Jensen's inequality, we have

$$\text{Var}(V'_i) \geq \frac{1}{3}\sigma_i^2 + \frac{1}{9\sigma_i^2}\mathbb{E}[V_i^4] - \frac{1}{16\sigma_i^2}\mathbb{E}[V_i^4] = \frac{1}{3}\sigma_i^2 + \frac{7}{144\sigma_i^2}\mathbb{E}[V_i^4] \geq \frac{1}{3}\sigma_i^2 + \frac{7}{144}\sigma_i^2 = \frac{55}{144}\sigma_i^2.$$

Next, by Goldstein and Reinert [34, Lem. 2.1(ii)], the support of  $V_i^*$  is the closed convex hull of the support of  $V_i$ . Therefore,  $|V'_i| \leq R'$  almost surely by the triangle inequality, and

$$\sup \text{support}(V'_i) - \inf \text{support}(V'_i) = \sup \text{support}(V_i) - \inf \text{support}(V_i) \leq R.$$

The second almost sure bound on  $|V'_i - \mathbb{E}[V'_i]|$  now follows from the triangle inequality, while the first follows from Lem. 1 and our lower estimate for  $\text{Var}(V'_i)$ .  $\square$

The fourth result bounds the tail probabilities of the interpolated zero-biased sum  $T'_n$  using concentration inequalities for sums of potentially identically distributed random variables.

**LEMMA 3** ( $T'_n$  tail bound). *Under the notation and assumptions of Lem. 2, for each  $t \in \mathbb{R}$ ,*

$$\begin{aligned} \mathbb{P}(T'_n > t + \frac{R'}{4\sqrt{n}}) & \leq \frac{1}{n} \sum_{i=1}^n \frac{\sigma_i^2}{\bar{\sigma}_n^2} \mathbb{P}\left(\frac{1}{\sqrt{n}}(V'_i - \mathbb{E}[V'_i]) + \frac{1}{\sqrt{n}} \sum_{j \neq i} V_j > t\right) \\ & \leq \frac{1}{n} \sum_{i=1}^n \frac{\sigma_i^2}{\bar{\sigma}_n^2} \min\left(e^{-\frac{2t^2}{R'^2}}, e^{-\frac{t^2}{2(v_{n,i}^2 + \frac{R'^2}{3\sqrt{n}})}}, \Phi^c\left(\frac{t}{\tilde{v}_{n,i}}\right) + \frac{0.56}{\sqrt{n}} \frac{R'v_{n,i}^2 + \frac{R_i - R'}{n}(\frac{\sigma_i^2}{3} + \frac{R'^2}{9})}{\tilde{v}_{n,i}^3}\right) \end{aligned}$$

for  $v_{n,i}^2 \triangleq \bar{\sigma}_n^2 + (R'^2 - 6\sigma_i^2)/(9n)$  and  $\tilde{v}_{n,i}^2 \triangleq \bar{\sigma}_n^2 - 89\sigma_i^2/(144n)$ .

PROOF. The first claim follows from the definition  $T'_n = \frac{1}{\sqrt{n}}V'_{I_n} + \frac{1}{\sqrt{n}}\sum_{j \neq I_n} V_j$  and the Lem. 2 bound,  $\mathbb{E}[V'_i] \leq \frac{R'}{4}$  for each index  $i$ . Since  $\mathbb{E}[|V'_i - \mathbb{E}[V'_i]|^3] \leq R_i \text{Var}(V'_i)$  and  $\mathbb{E}[|V_i|^3] \leq R'\sigma_i^2$  for each  $i \in \mathbb{N}$ , the second follows by combining the upper and lower estimates of Lem. 2 with the Hoeffding [38, Thm. 1], Bernstein [17, Cor. 2.11], and Berry-Esseen [63, Thm. 1] tail bounds for unidentically distributed summands.  $\square$

Under Assump.  $(R, \sigma)$ , the first result of Thm. 1 now follows from a direct application of Thm. 2 and Lem. 3 with i.i.d.  $V_i = W_i - \mathbb{E}[W_i]$ ,  $T_n = S_n$ ,  $\sigma_i = \sigma$ , and, by Lem. 1,  $R' = R_\sigma$ . Analogously, under Assump.  $(R, \sigma)$ , the bound (10) follows from Thm. 3 and Lem. 3.

**3.2. Proofs of Thms. 2 and 3: Unidentical and identical zero-bias tail bounds.** The crux of our argument, inspired by Ross [62, Thm. 3.27] and proved in Sec. A, shows that a zero-biased version  $S^*$  of a random variable  $S$  has tails close to those of a Gaussian whenever  $S^*$  and  $S$  have a close coupling.

**THEOREM 4** (General zero-bias tail bounds). *Suppose that  $S^*$  has the zero-bias distribution for a random variable  $S$  with  $\mathbb{E}[S] = 0$  and  $\text{Var}(S) = 1$  and that  $S^* - S \leq \delta$  almost surely. For  $U \sim \text{Uniform}([0, 1])$  independent of  $(S, S^*)$ , define the intermediate variable  $S' \triangleq S^* + U(S - S^*)$ . Then, for all  $u \geq 0$  and  $\lambda \in [0, 1]$ ,*

$$\mathbb{P}(S^* > u) - \Phi^c(u) \leq \delta [h_u(\lambda u) + (h_u(u) - h_u(\lambda u))\mathbb{P}(S' > \lambda u) - u\mathbb{P}(S' > u)],$$

for  $h_u$  defined in (8). If, in addition,  $S - S^* \leq \delta$  almost surely,

$$\mathbb{P}(S > u + \delta) \leq \frac{1}{1 + \delta u}\Phi^c(u) + \frac{\delta}{1 + \delta u}(h_u(\lambda u) + (h_u(u) - h_u(\lambda u))\mathbb{P}(S' > \lambda u)).$$

Hence, to establish Thm. 2, it only remains to construct a suitable zero-bias coupling for  $T_n/\bar{\sigma}_n$ . By Chen, Goldstein and Shao [21, Lem. 2.8],  $T_n^* = T_n + \frac{1}{\sqrt{n}}(V_{I_n}^* - V_{I_n})$  has the zero-bias distribution for  $T_n$ . Therefore, by the zero-bias definition,  $T_n^*/\bar{\sigma}_n$  has the zero-bias distribution for  $T_n/\bar{\sigma}_n$ . Furthermore, by Chen, Goldstein and Shao [21, Sec. 2.3.3],  $\text{support}(V_i^*) = [\inf \text{support}(V_i), \sup \text{support}(V_i)]$  for each  $i$ , so  $|T_n^*/\bar{\sigma}_n - T_n/\bar{\sigma}_n| = |V_{I_n}^* - V_{I_n}|/(\bar{\sigma}_n\sqrt{n}) \leq R/(\bar{\sigma}_n\sqrt{n})$  almost surely. The advertised result thus follows from the second claim of Thm. 4 applied with  $S = T_n/\bar{\sigma}_n$ ,  $S^* = T_n^*/\bar{\sigma}_n$ , and  $\delta = R/(\bar{\sigma}_n\sqrt{n})$ .

Now suppose  $(V_i)_{i=1}^{n+1}$  are i.i.d. with  $\sigma \triangleq \sigma_1$  and  $V_1 \geq -R'$  almost surely. By Chen, Goldstein and Shao [21, Lem. 2.8],  $T_{n+1}^* = T_{n+1} + \frac{1}{\sqrt{n+1}}(V_{I_{n+1}}^* - V_{I_{n+1}})$  has the zero-bias distribution for  $T_{n+1}$ . By symmetry,  $\sqrt{n}T_n + V_{n+1}^* \stackrel{d}{=} \sqrt{n+1}T_{n+1}^*$  and  $\sqrt{n}T_n + V'_{n+1} \stackrel{d}{=} \sqrt{n+1}T'_{n+1}$ . Since  $V_{n+1}^* \wedge V'_{n+1} \geq -R'$  almost surely by Chen, Goldstein and Shao [21, Sec. 2.3.3],

$$\begin{aligned} \mathbb{P}(T_n > \sigma u \frac{\sqrt{n+1}}{\sqrt{n}} + \frac{R'}{\sqrt{n}}) &\leq \mathbb{P}(T_n + \frac{1}{\sqrt{n}}(V_{n+1}^* \wedge V'_{n+1}) > \sigma u \frac{\sqrt{n+1}}{\sqrt{n}}) \\ &\leq \mathbb{P}(T_{n+1}^* > \sigma u) \wedge \mathbb{P}(T'_{n+1} > \sigma u). \end{aligned}$$

Thm. 3 now follows from the first result of Thm. 4 with  $S = T_{n+1}/\sigma$ ,  $S^* = T_{n+1}^*/\sigma$ , and  $\delta = R/(\sigma\sqrt{n+1})$ .

**4. Computational Refinement with Wasserstein Approximation.** To obtain a practical refinement of our initial concentration inequality (Thm. 1), we next bound a second, complementary distance to Gaussianity, the  $p$ -Wasserstein distance,

$$\mathcal{W}_p(S_n, \mathcal{N}(0, \sigma^2)) \triangleq \inf_{\gamma \in \Gamma(S_n, \mathcal{N}(0, \sigma^2))} \mathbb{E}_{(X, Y) \sim \gamma} (|X - Y|^p)^{1/p} \quad \text{for } p \geq 1,$$

where  $\Gamma(S_n, \mathcal{N}(0, \sigma^2))$  is the set of all couplings between the law of  $S_n$  and the normal distribution  $\mathcal{N}(0, \sigma^2)$ . In Sec. 4.1, we explicitly bound each  $p$ -Wasserstein distance to Gaussianity to obtain the following efficient and computable concentration inequalities.

**THEOREM 5** (Efficient Wasserstein tail bound). *Under Assump.  $(R, \sigma)$ , for any  $u \geq 0$  and auxiliary upper bounds  $\hat{Q}_n(u)$  and  $\hat{Q}_n^d(u)$  on  $P(S_n > \sigma u)$  and  $P(|S_n| > \sigma u)$  respectively, we have*

$$(11) \quad \begin{aligned} \mathbb{P}(S_n > \sigma u) &\leq \min \left( \inf_{p \in \mathbb{N}, \rho \in (0,1)} \Phi^c(\rho u) + \frac{\omega_p^R(\sigma)^p}{(1-\rho)^p u^p}, \hat{Q}_n(u) \right) \quad \text{and} \\ \mathbb{P}(|S_n| > \sigma u) &\leq \min \left( \inf_{p \in \mathbb{N}, \rho \in (0,1)} 2\Phi^c(\rho u) + \frac{\omega_p^R(\sigma)^p}{(1-\rho)^p u^p}, \hat{Q}_n^d(u) \right) \end{aligned}$$

where  $\omega_p^R(\sigma)$  is a computable bound on  $\frac{1}{\sigma} \mathcal{W}_p(S_n, \mathcal{N}(0, \sigma^2))$  defined in Lem. 5 and satisfying  $\omega_p^R(\sigma) \leq K_{R,\sigma} \frac{p}{\sqrt{n}}$  for all  $p \in \{1\} \cup [2, \infty)$  and a constant  $K_{R,\sigma}$  depending only on  $(R, \sigma)$  defined in Lem. 6.

**REMARK 1** (Auxiliary tail bounds). *We include the auxiliary bounds  $\hat{Q}_n(u)$  and  $\hat{Q}_n^d(u)$  to emphasize that our efficient bounds can be paired with any valid tail bounds to simultaneously reap the large-sample benefits of the former and the small-sample benefits of the latter. A convenient default choice is to set  $\hat{Q}_n^d(u) = 2\hat{Q}_n(u)$  and  $\hat{Q}_n(u)$  equal to the minimum of the zero-bias (Thm. 1), Hoeffding (1), Bernstein (1), Berry-Esseen (4), and non-uniform Berry-Esseen (5) bounds.*

Thm. 5 immediately gives rise to the following looser but evidently efficient tail bounds.

**COROLLARY 2** (Efficiency and relative error). *Under the conditions of Thm. 5,*

$$(12) \quad \begin{aligned} \mathbb{P}(S_n > \sigma u) &\leq \Phi^c(u - \delta_{u,n}) + \frac{\varphi(u)}{\sqrt{n}} \quad \text{for} \quad \delta_{u,n} \triangleq \frac{eK_{R,\sigma}}{\sqrt{n}} \lceil \log \left( \frac{\sqrt{n}}{\varphi(u)} \right) \rceil \\ &\leq \Phi^c(u) \cdot \left( e^{(u+1)\delta_{u,n}} + \frac{1}{\sqrt{n}} \frac{\varphi(u)}{\Phi^c(u)} \right) \quad \text{for all} \quad u > \delta_{u,n}. \end{aligned}$$

**PROOF.** Fix any  $u > \delta_{u,n}$ . Since  $\omega_p^R(\sigma) \leq K_{R,\sigma} \frac{p}{\sqrt{n}}$  for all  $p \in \mathbb{N}$  by Thm. 5, we may choose  $p = \lceil \log(\frac{\sqrt{n}}{\varphi(u)}) \rceil \in \mathbb{N}$  and  $\rho = 1 - e \omega_p^R(\sigma)/u \in (0, 1)$  in the upper bound (11) to deduce the first inequality. The log-concavity of  $\Phi^c$  [4, Thm. 2] additionally implies that

$$\Phi^c(u - \delta_{u,n}) \leq \Phi^c(u) \exp(\delta_{u,n} \varphi(u)/\Phi^c(u)).$$

The second inequality therefore follows from the Mills ratio bound  $\frac{\varphi(u)}{\Phi^c(u)} \leq u + 1$  [15].  $\square$

While the  $n$  dependence of Cor. 2 is no better than that established for Thm. 1, the  $u$  dependence is improved, yielding tighter bounds for larger deviations  $u$  and a relative error bound (12) in the spirit of classical Cramér-type inequalities. For practical applications, the Wasserstein tail bound of Thm. 5 is substantially tighter than Cor. 2, while still admitting a straightforward implementation.<sup>2</sup> Hence, for practical use, we recommend pairing the zero-bias and Wasserstein inequalities as in Rem. 1 to simultaneously inherit the benefits of each efficient bound.

In addition, to obtain efficient, computable quantile bounds, we recommend simply inverting the tail bounds of Thm. 5 and Rem. 1. Our next result, also proved in Sec. 4.1, makes this precise.

**THEOREM 6** (Efficient Wasserstein quantile bound). *Under Assump.  $(R, \sigma)$ , for any  $\delta \in (0, 1)$  and auxiliary deterministic bounds  $\tilde{q}_n$  and  $\tilde{q}_n^d$  satisfying  $P(S_n > \tilde{q}_n(R, \delta, \sigma)) \leq \delta$  and  $P(|S_n| > \tilde{q}_n^d(R, \delta, \sigma)) \leq \delta$ , we have*

$$\mathbb{P}(S_n > q_n(R, \delta, \sigma)) \leq \delta \quad \text{and} \quad \mathbb{P}(|S_n| > q_n^d(R, \delta, \sigma)) \leq \delta,$$

for  $\omega_p^R(\sigma)$  as in Thm. 5 and

$$q_n(R, \delta, \sigma) \triangleq \min \left( \inf_{p \in \mathbb{N}, \rho \in (0,1)} \frac{\sigma \omega_p^R(\sigma)}{(\delta(1-\rho))^{1/p}} + \sigma \Phi^{-1}(1-\delta\rho), \tilde{q}_n(R, \delta, \sigma) \right),$$

$$q_n^d(R, \delta, \sigma) \triangleq \min \left( \inf_{p \in \mathbb{N}, \rho \in (0,1)} \frac{\sigma \omega_p^R(\sigma)}{(\delta(1-\rho))^{1/p}} + \sigma \Phi^{-1}(1-\frac{\delta\rho}{2}), \tilde{q}_n^d(R, \delta, \sigma) \right).$$

REMARK 2 (Auxiliary quantile bounds). *Our efficient bounds can be paired with any deterministic quantile bounds  $\tilde{q}_n(R, \delta, \sigma)$  and  $\tilde{q}_n^d(R, \delta, \sigma)$  to inherit the benefits of each. Convenient defaults are provided by numerically inverting the default auxiliary tail bounds  $\hat{Q}_n(u)$  and  $\hat{Q}_n^d(u)$  of Rem. 1.*

In the sequel, we build upon the quantile bounds of Thm. 6 and Rem. 2 to develop efficient empirical Berry-Esseen bounds that require no prior knowledge of the variance parameter  $\sigma$ .

4.1. *Proof of Thms. 5 and 6.* Thms. 5 and 6 will follow directly from three auxiliary results. The first, proved in Sec. 4.2, shows that any upper bound on the Wasserstein distance  $\frac{1}{\sigma} \mathcal{W}_p(S_n, \mathcal{N}(0, \sigma^2))$  also yields tail and quantile bounds for  $S_n$ .

LEMMA 4 (Tail and quantile bounds from Wasserstein bounds). *Under Assump. (R,  $\sigma$ ), suppose that  $\frac{1}{\sigma} \mathcal{W}_p(S_n, \mathcal{N}(0, \sigma^2)) \leq \omega_p$  for some  $p \geq 1$  and  $\omega_p \geq 0$ . Then, for any  $u \geq 0$ ,*

$$\mathbb{P}(S_n > \sigma u) \leq \inf_{\rho \in (0,1)} \left\{ \Phi^c(\rho u) + \frac{\omega_p^p}{(1-\rho)^p u^p} \right\} \quad \text{and}$$

$$\mathbb{P}(|S_n| > \sigma u) \leq \inf_{\rho \in (0,1)} \left\{ 2\Phi^c(\rho u) + \frac{\omega_p^p}{(1-\rho)^p u^p} \right\}.$$

Moreover, for all  $\delta \in (0, 1)$ ,

$$\mathbb{P}(S_n > \inf_{\rho \in (0,1)} \frac{\sigma \omega_p}{(\delta(1-\rho))^{1/p}} + \sigma \Phi^{-1}(1-\delta\rho)) \leq \delta \quad \text{and}$$

$$\mathbb{P}(|S_n| > \inf_{\rho \in (0,1)} \frac{\sigma \omega_p}{(\delta(1-\rho))^{1/p}} + \sigma \Phi^{-1}(1-\frac{1}{2}\delta\rho)) \leq \delta.$$

Our second result, proved in Sec. B, establishes that  $\omega_p^R(\sigma)$  is indeed an upper bound on  $\frac{1}{\sigma} \mathcal{W}_p(S_n, \mathcal{N}(0, \sigma^2))$ .

LEMMA 5 (Wasserstein upper bound). *Under Assump. (R,  $\sigma$ ), for  $p = 1$  or  $p \geq 2$ ,*

$$\frac{1}{\sigma} \mathcal{W}_p(S_n, \mathcal{N}(0, \sigma^2)) \leq \omega_p^R(\sigma) \quad \text{and} \quad \omega_p^{R,\kappa}(\sigma) \leq \omega_p^{R,\kappa,2}(\sigma) \quad \text{for all } \kappa > \frac{\tilde{R}^2}{n}$$

where

$$(\omega_p^R) \quad \omega_p^R(\sigma) \triangleq \begin{cases} \frac{R_\sigma}{\sigma \sqrt{n}} & \text{if } p = 1, \\ \inf_{\kappa > \frac{\tilde{R}^2}{n}} \omega_p^{R,\kappa,1}(\sigma) & \text{if } n+1 \geq p \geq 2, \\ \sqrt{p-1} \left(1 + \frac{R_\sigma}{\sigma}\right) & \text{if } p > n+1, \end{cases}$$

and, for each  $p \in [2, n+1]$ ,  $\kappa > 0$ , and  $j \in \{1, 2\}$ ,

$$\omega_p^{R,\kappa,j}(\sigma) \triangleq \left[ \|Z\|_p \left( \frac{\pi}{2} - \sin^{-1} \left( \sqrt{1 - \frac{\tilde{R}^2}{n\kappa}} \right) \right) + (b_{2,j}^{\kappa,p,\tilde{R}}) + (b_{3,j}^{\kappa,p,\tilde{R}}) \right] \begin{cases} \frac{1}{\sqrt{1 - \frac{\tilde{R}^2}{n\kappa}}} & \text{if } p > 2 \\ 1 & \text{if } p = 2, \end{cases}$$

for  $\tilde{R} \triangleq R/\sigma$ ,  $Z \sim \mathcal{N}(0, 1)$ , and  $(b_{2,1}^{\kappa,p,\tilde{R}})$ ,  $(b_{3,1}^{\kappa,p,\tilde{R}})$ ,  $(b_{2,2}^{\kappa,p,\tilde{R}})$ , and  $(b_{3,2}^{\kappa,p,\tilde{R}})$  defined in (22), (23), (25), and (26) respectively.

Our final lemma, proved in Sec. 4.3, verifies the growth rate of  $\omega_p^R(\sigma)$ .

**LEMMA 6 (Growth of  $\omega_p^R(\sigma)$ ).** *Instantiate the notation of Lem. 5 and define  $\widetilde{R}_\sigma \triangleq \frac{R_\sigma}{\sigma}$ . If  $p = 1$  or  $p \geq 2$ , then  $\omega_p^R(\sigma) \leq \frac{K_{R,\sigma} p}{\sqrt{n}}$  for*

$$K_{R,\sigma} \triangleq 1 + \max \left( \widetilde{R}_\sigma, \frac{3e}{4} \sqrt{R_\sigma^2 - \sigma^2} + \sqrt{2 \max(R_\sigma^2 - \sigma^2, \sigma^2)} \right) \\ + \left( \frac{3}{16} + \frac{\sqrt{2} \log(4e) \tilde{R}^{1-2/p} (8\pi)^{1/4}}{\sqrt{3e}} + \frac{\tilde{R}e}{2\sqrt{2}} + \frac{\tilde{R}^{2-2/p}}{\sqrt{2}} \frac{\sigma \pi^{1/4} e^{19/300} (e^{\frac{\tilde{R}^2}{2}} - 1)}{\sqrt{3}} \right).$$

**4.2. Proof of Lem. 4: Tail and quantile bounds from Wasserstein bounds.** Fix any  $u \geq 0$ ,  $\rho \in (0, 1)$ ,  $p \geq 1$ , and  $\epsilon > 0$ . By the definition of the Wasserstein distance, we can find  $G \sim \mathcal{N}(0, \sigma^2)$  that respects  $\|S_n - G\|_p \leq \epsilon + \mathcal{W}_p(S_n, \mathcal{N}(0, \sigma^2))$ . Hence, by applying the union bound and Markov's inequality in turn, we find that

$$\mathbb{P}(S_n > \sigma u) = \mathbb{P}(G + S_n - G > \sigma u) \leq \mathbb{P}(G > \rho \sigma u) + \mathbb{P}(S_n - G > (1 - \rho) \sigma u) \\ \leq \Phi^c(\rho u) + \frac{\mathbb{E}(|S_n - G|^p)}{(1 - \rho)^p \sigma^p u^p} \leq \Phi^c(\rho u) + \frac{(\epsilon + \mathcal{W}_p(S_n, \mathcal{N}(0, \sigma^2)))^p}{(1 - \rho)^p \sigma^p u^p}.$$

As this holds for arbitrary  $\epsilon > 0$  we obtain

$$\mathbb{P}(S_n > \sigma u) \leq \Phi^c(\rho u) + \frac{\mathcal{W}_p(S_n, \mathcal{N}(0, \sigma^2))^p}{(1 - \rho)^p \sigma^p u^p}.$$

The triangle inequality and identical reasoning yield

$$\mathbb{P}(|S_n| > \sigma u) \leq \mathbb{P}(|G| > \rho \sigma u) + \mathbb{P}(|S_n - G| > (1 - \rho) \sigma u) \leq 2\Phi^c\left(\frac{\rho \sigma u}{\sigma}\right) + \frac{(\epsilon + \mathcal{W}_p(S_n, \mathcal{N}(0, \sigma^2)))^p}{(1 - \rho)^p \sigma^p u^p}.$$

Since  $\epsilon > 0$  is arbitrary, we conclude that

$$\mathbb{P}(|S_n| > t) \leq 2\Phi^c\left(\frac{\rho t}{\sigma}\right) + \frac{\mathcal{W}_p(S_n, \mathcal{N}(0, \sigma^2))^p}{(1 - \rho)^p t^p}.$$

Now, fix any  $\delta \in (0, 1)$  and write  $t_{\delta, \rho} \triangleq \frac{\mathcal{W}_p(S_n, \mathcal{N}(0, \sigma^2))}{(\delta(1 - \rho))^{1/p}} + \sigma \Phi^{-1}(1 - \delta \rho)$ . We apply the union bound, Markov's inequality, and the Wasserstein condition in turn to find

$$\mathbb{P}(S_n > t_{\delta, \rho}) = \mathbb{P}(S_n - G + G > t_{\delta, \rho}) \\ \leq \mathbb{P}(G > \sigma \Phi^{-1}(1 - \delta \rho)) + \mathbb{P}\left(S_n - G > \frac{\mathcal{W}_p(S_n, \mathcal{N}(0, \sigma^2))}{(\delta(1 - \rho))^{1/p}}\right) \\ \leq \delta \rho + \frac{\delta(1 - \rho)(\mathcal{W}_p(S_n, \mathcal{N}(0, \sigma^2)) + \epsilon)^p}{\mathcal{W}_p(S_n, \mathcal{N}(0, \sigma^2))^p}.$$

Since this holds for any  $\epsilon > 0$ , we deduce that

$$\mathbb{P}(S_n > t_{\delta, \rho}) \leq \delta \rho + \delta(1 - \rho) = \delta.$$

Moreover, since  $\rho$  was arbitrary and each  $t_{\delta, \rho}$  is deterministic, we further have

$$\mathbb{P}(S_n > \inf_{\rho \in (0, 1)} t_{\delta, \rho}) \leq \delta.$$

Finally, the triangle inequality and identical reasoning yield

$$\mathbb{P}(|S_n| > \inf_{\rho \in (0, 1)} t_{\delta, \rho}^d) \leq \delta \quad \text{for} \quad t_{\delta, \rho}^d \triangleq \frac{\mathcal{W}_p(S_n, \mathcal{N}(0, \sigma^2))}{(\delta(1 - \rho))^{1/p}} + \sigma \Phi^{-1}(1 - \frac{1}{2} \delta \rho).$$

4.3. *Proof of Lem. 6: Growth of  $\omega_p^R(\sigma)$ .* If  $p = 1$ , then  $\omega_p^R(\sigma) = \frac{\tilde{R}_\sigma}{\sqrt{n}}$ . If  $p > n + 1$ ,

$$\omega_p^R(\sigma) = \sqrt{p-1}(1 + \tilde{R}_\sigma) \leq \frac{p}{\sqrt{n}}(1 + \tilde{R}_\sigma).$$

Now suppose  $p \in [2, n + 1]$ . By Lem. 5, we have, for all  $\kappa > \frac{R^2}{\sigma^2 n}$ ,

$$\begin{aligned} \omega_p^R(\sigma) & \frac{(M_{n,\kappa}\mathbb{I}(p>2)+\mathbb{I}(p=2))\sqrt{n}}{p} \\ & \leq \frac{\sqrt{(p-1)(p+2)}\sigma}{2p} \left[ \frac{\sqrt{p+2}n^{1/p}}{2\sqrt{n}} (\max(\tilde{R}_\sigma^2 - 1, 1))^{1-1/p} + \sqrt{\tilde{R}_\sigma^2 - 1} \frac{\sqrt{e}(2e)^{1/p}}{\sqrt{2}} \right] \\ & + \frac{C_{n,p}\sigma}{p} \frac{\tilde{R}^2}{3\sqrt{n}\kappa} e^{19/300} \pi^{1/4} [e^{\frac{1}{2}(p-1)(1-\frac{\tilde{R}^2}{n\kappa})\kappa} - 1] \log\left(\frac{1+M_{n,\kappa}}{1-M_{n,\kappa}}\right) \\ & + \frac{R\left(1+\frac{\tilde{R}\tilde{U}_{n,p}}{\sqrt{n}}\right)}{\sqrt{\kappa}} \frac{\pi^{1/4} e^{19/300} \sqrt{p-1}}{4\sqrt{3}} [e^{\frac{1}{2}(p-1)\kappa M_{n,\kappa}^2} - 1] + \frac{\sqrt{n}\sqrt{p-1}}{p} \left[ \frac{\pi}{2} - \sin^{-1}(M_{n,\kappa}) \right] \end{aligned}$$

where  $M_{n,\kappa} = \sqrt{1 - \frac{\tilde{R}^2}{n\kappa}}$  and  $C_{n,p}$  is a constant defined in (19) that satisfies  $C_{n,p} \leq \tilde{R}^{1-2/p} \sqrt{8\pi}^{1/p} \frac{\sqrt{p}}{\sqrt{e}}$  by Batir [7, Thm. 1.5]. Since we additionally have  $\frac{\pi}{2} - \sin^{-1}(M_{n,\kappa}) \leq \frac{\tilde{R}^2}{n\kappa M_{n,\kappa}}$  by the mean value theorem, we can write

$$\begin{aligned} \omega_p^R(\sigma) & \frac{\sqrt{n}\sqrt{1-\frac{\tilde{R}^2(p-1)}{nC}}\mathbb{I}(p>2)}{p} \\ & \leq \frac{\sqrt{(p-1)(p+2)}\sigma}{2p} \left( \sqrt{\tilde{R}_\sigma^2 - 1} \frac{\sqrt{e}(2e)^{1/p}}{\sqrt{2}} + \frac{\tilde{R}}{\sqrt{C}} \frac{\pi^{1/4} e^{19/300} (p-1)}{4p\sqrt{3}} [e^{\frac{C}{2}} - 1] \right) \\ & + \frac{\sqrt{p}\log(\frac{4nC}{(p-1)\tilde{R}^2})\sigma}{\sqrt{n}} \left\{ \frac{\tilde{R}^{3-2/p}\sqrt{8\pi}^{1/p}}{3\sqrt{e}C} \frac{p-1}{p} e^{19/300} \pi^{1/4} [e^{\frac{C}{2}} - 1] \right\} \\ & + \frac{\sqrt{p+2}\sigma}{\sqrt{n}^{1-2/p}} \left( \frac{(\max(\tilde{R}_\sigma^2 - 1, 1))^{1-1/p}}{2} + \frac{\sqrt{p+2}\sigma}{\sqrt{n}} \left[ \frac{\tilde{R}^2 \frac{\sqrt{e}(2e)^{1/p}}{\sqrt{2}}}{\sqrt{C}} \frac{\pi^{1/4} e^{19/300} (p-1)}{4p\sqrt{3}} [e^{\frac{C}{2}} - 1] \right] \right) \\ & + \frac{(p+2)\sigma\tilde{R}^{3-2/p}}{n^{1-1/p}} \frac{1}{2\sqrt{C}} \frac{\pi^{1/4} e^{19/300} (p-1)}{4p\sqrt{3}} [e^{\frac{C}{2}} - 1] + \frac{\tilde{R}^2 \sqrt{p-1}^3}{p\sqrt{n}C \sqrt{1 - \frac{\tilde{R}^2(p-1)}{nC}}} \end{aligned}$$

in terms of the reparameterization  $C \triangleq \kappa(p-1)$ . The advertised result now follows from the choosing  $C = \tilde{R}^2$ , since  $\kappa = \frac{\tilde{R}^2}{p-1} \geq \frac{\tilde{R}^2}{n}$  and  $(p+2)(\log(p)-1) \leq n^2/2$  for  $p \leq n+1$  and  $p \rightarrow \sqrt{p+2}n^{1/p}$  is decreasing for  $(p+2)(\log(p)-1) \leq n^2/2$ .

**5. Application: Efficient *Empirical Berry-Esseen Bounds*.** The preceding sections assumed that the variance parameter  $\sigma^2$  was a known quantity, but, in many applications, the variance is unknown and can only be estimated from data. In this section, we leverage our efficient known-variance bounds to develop efficient quantile bounds that are valid even when  $\sigma$  is unknown. We refer to these constructions as *empirical Berry-Esseen* bounds as they combine Gaussian approximation in the spirit of Berry-Esseen (4) with empirical variance estimation.

We begin by showing how to convert generic known-variance quantile bounds into valid empirical-variance quantile bounds. Our construction in Lem. 7 makes use of a generic confidence interval that contains the unknown variance with high probability.

**LEMMA 7 (Empirical quantile bounds).** *Consider any nonnegative interval  $[\hat{\sigma}_{\text{low},a}^2, \hat{\sigma}_{\text{up},a}^2]$  and any nonnegative quantile bounds  $\tilde{q}_n$  and  $\tilde{q}_n^d$  satisfying*

$$\mathbb{P}(\sigma \notin [\hat{\sigma}_{\text{low},a}, \hat{\sigma}_{\text{up},a}]) \leq a, \quad \mathbb{P}(S_n > \tilde{q}_n(R, \delta, \sigma)) \leq \delta, \quad \text{and} \quad \mathbb{P}(|S_n| > \tilde{q}_n^d(R, \delta, \sigma)) \leq \delta$$

whenever  $S_n$  satisfies Assump.  $(R, \sigma)$  and  $\delta, a \in (0, 1)$ . Then, under Assump.  $(R, \sigma)$ , we have, for each confidence level  $\delta \in (0, 1)$  and  $a_n \in (0, \delta)$ ,

$$\begin{aligned}\mathbb{P}(S_n > \hat{q}_n(R, \delta, a_n)) &\leq \delta \quad \text{for } \hat{q}_n(R, \delta, a) \triangleq \sup_{\tilde{\sigma} \in [\hat{\sigma}_{\text{low},a}, \hat{\sigma}_{\text{up},a}]} \tilde{q}_n(R, \delta - a, \tilde{\sigma}) \quad \text{and} \\ \mathbb{P}(|S_n| > \hat{q}_n^d(R, \delta, a_n)) &\leq \delta \quad \text{for } \hat{q}_n^d(R, \delta, a) \triangleq \sup_{\tilde{\sigma} \in [\hat{\sigma}_{\text{low},a}, \hat{\sigma}_{\text{up},a}]} \tilde{q}_n^d(R, \delta - a, \tilde{\sigma}).\end{aligned}$$

PROOF. Fix any  $\delta > 0$  and  $a \in (0, \delta)$ . Under Assump.  $(R, \sigma)$ , the union bound implies

$$\mathbb{P}(|S_n| > \hat{q}_n^d(R, \delta, a)) \leq \mathbb{P}(|S_n| > \tilde{q}_n^d(R, \delta - a, \sigma)) + \mathbb{P}(\sigma \notin [\hat{\sigma}_{\text{low},a}, \hat{\sigma}_{\text{up},a}]) \leq \delta.$$

The one-sided result is obtained identically using the event  $\{S_n > \tilde{q}_n(R, \delta - a, \sigma)\}$ .  $\square$

Finally, we show how to convert efficient known-variance bounds into efficient empirical Berry-Esseen bounds through appropriate choice of the variance confidence interval and the confidence parameter  $a_n$ . For this purpose we use the sharp (and valid) empirical Bernstein variance confidence intervals of Martinez-Taboada and Ramdas [47, Sec. 4.4] which depend on the auxiliary functions

$$\Psi_E(x) \triangleq -\log(1 - x) - x \quad \text{and} \quad \Psi_P(x) \triangleq e^x - x - 1$$

and the following readily-computed quantities for  $i \in \{1, \dots, n\}$  and  $a \in (0, 1)$ :

- the  $\mu \triangleq \mathbb{E}[W_1]$  estimates:  $\hat{\mu}_i \triangleq \frac{R}{2i} + \frac{1}{i} \sum_{j=1}^{i-1} W_j$ ,
- the  $\sigma^2$  estimates:  $\hat{\sigma}_i^2 \triangleq \frac{R^2}{4i} + \frac{1}{i} \sum_{j=1}^{i-1} (W_j - \hat{\mu}_j)^2$ ,
- the  $m_4^2 \triangleq \text{Var}((W_1 - \mu)^2)$  estimates:  $\hat{m}_{4,i}^2 \triangleq \frac{R^4}{8i} + \frac{1}{i} \sum_{j=1}^{i-1} ((W_j - \hat{\mu}_j)^2 - \hat{\sigma}_j^2)^2$ ,
- the upper estimate weights:  $\lambda_{i,a} \triangleq \sqrt{\frac{2R^4 \log(2/a)}{n\hat{m}_{4,i}^2}} \wedge \frac{1}{2}$ ,
- the auxiliary weights:  $\tilde{\lambda}_{i,a} \triangleq \sqrt{\frac{2R^2 \log((2+2\log(n))/a)}{\hat{\sigma}_i^2 i \log(1+i)}} \wedge 2$ ,
- the lower estimate weights:  $\lambda'_{i,a} \triangleq \begin{cases} \lambda_{i, \frac{a \log(n)}{1+\log(n)}} & \text{if } \frac{\log(\frac{2+2\log(n)}{a}) + \sum_{j=1}^{i-1} \Psi_P(\tilde{\lambda}_{j,a}) \hat{\sigma}_j^2 / R^2}{\sum_{j=1}^{i-1} \tilde{\lambda}_{j,a}} \leq 1 \\ 0 & \text{otherwise,} \end{cases}$
- the  $\sigma^2$  upper estimate:

$$(13) \quad \hat{\sigma}_{\text{up},a}^2 \triangleq \frac{R^2 \log(2/a) + \sum_{i=1}^n \lambda_{i,a} (W_i - \hat{\mu}_i)^2 + \Psi_E(\lambda_{i,a}) ((W_i - \hat{\mu}_i)^2 - \hat{\sigma}_i^2)^2 / R^2}{\sum_{i=1}^n \lambda_{i,a}},$$

- and the  $\sigma^2$  lower estimate:

$$(14) \quad \hat{\sigma}_{\text{low},a}^2 \triangleq \frac{-R^2 \log(2/a) + \sum_{i=1}^n \lambda'_{i,a} (W_i - \hat{\mu}_i)^2 - \Psi_E(\lambda'_{i,a}) ((W_i - \hat{\mu}_i)^2 - \hat{\sigma}_i^2)^2 / R^2}{\sum_{i=1}^n \lambda'_{i,a}}.$$

**THEOREM 7** (Efficient empirical Berry-Esseen bounds). *Instantiate Assump.  $(R, \sigma)$ , and consider any bounds  $(\tilde{q}_n, \tilde{q}_n^d)$  satisfying  $\tilde{q}_n(R, \delta_n, \sigma_n) \rightarrow \sigma \Phi^{-1}(1 - \delta)$  and  $\tilde{q}_n^d(R, \delta_n, \sigma_n) \rightarrow \sigma \Phi^{-1}(1 - \frac{\delta}{2})$  whenever  $(\sigma_n, \delta_n) \rightarrow (\sigma, \delta)$ . If  $a_n \rightarrow 0$  and  $\frac{\log(a_n)}{n} \rightarrow 0$ , then*

$$\hat{q}_n(R, \delta, a_n) \triangleq \sup_{\tilde{\sigma} \in [\hat{\sigma}_{\text{low},a_n}, \hat{\sigma}_{\text{up},a_n}]} \tilde{q}_n(R, \delta - a_n, \tilde{\sigma}) \xrightarrow{a.s.} \sigma \Phi^{-1}(1 - \delta) \quad \text{and}$$

$$\hat{q}_n^d(R, \delta, a_n) \triangleq \sup_{\tilde{\sigma} \in [\hat{\sigma}_{\text{low},a_n}, \hat{\sigma}_{\text{up},a_n}]} \tilde{q}_n^d(R, \delta - a_n, \tilde{\sigma}) \xrightarrow{a.s.} \sigma \Phi^{-1}(1 - \frac{\delta}{2})$$

when  $\hat{\sigma}_{\text{up},a_n}$  and  $\hat{\sigma}_{\text{low},a_n}$  are chosen as in (13) and (14).

PROOF. The strong law of large numbers [24, Thm. 2.4.1] and Prop. C.1 of [47] imply that  $\hat{\mu}_n \xrightarrow{a.s.} \mu$ ,  $\hat{\sigma}_n^2 \xrightarrow{a.s.} \sigma^2$ , and  $\hat{m}_{4,n}^2 \xrightarrow{a.s.} m_4^2$ . Thus, with probability 1, there exists a finite integer  $\tilde{N}_{\min}$  for which  $\hat{\sigma}_i^2 > \hat{\sigma}^2/2$  and  $\tilde{\lambda}_{i,a_n} < 2$  whenever  $\tilde{N}_n \triangleq \tilde{N}_{\min} \vee \log(\frac{2+2\log(n)}{a_n}) \leq i \leq n$ . Defining  $\tilde{d}_i \triangleq \sqrt{\frac{2R^2 i \log((2+2\log(n))/a_n)}{\log(i+1)}}$  for each  $i \leq n$ , we can therefore write

$$\begin{aligned} \frac{1}{\tilde{d}_i} \sum_{j=1}^{i-1} \tilde{\lambda}_{j,a_n}^2 &= \frac{1}{\tilde{d}_i} \sum_{j=1}^{(i-1) \wedge \tilde{N}_n} \tilde{\lambda}_{j,a_n}^2 + \sqrt{\frac{\log(i+1)}{i}} \sum_{j=\tilde{N}_n+1}^{i-1} \frac{\sqrt{2R^2 \log((2+2\log(n))/a_n)}}{\hat{\sigma}_j^2 j \log(j+1)} \\ &\leq \frac{4\tilde{N}_n \sqrt{\log(i+1)}}{\sqrt{2R^2 i \log((2+2\log(n))/a_n)}} + \frac{2(1+\log(i-1))}{\sigma^2} \sqrt{\frac{2R^2 \log((2+2\log(n))/a_n) \log(i+1)}{i \log(2)}} \quad \text{and} \\ \frac{1}{\tilde{d}_i} \sum_{j=1}^{i-1} \tilde{\lambda}_{j,a_n} &= \frac{1}{\tilde{d}_i} \sum_{j=1}^{(i-1) \wedge \tilde{N}_n} \tilde{\lambda}_{j,a_n} + \sqrt{\frac{\log(i+1)}{i}} \sum_{j=\tilde{N}_n+1}^{i-1} \frac{1}{\hat{\sigma}_j \sqrt{j \log(j+1)}} \geq \frac{2\sqrt{i-2}\sqrt{\tilde{N}_n+1}}{R\sqrt{i}}. \end{aligned}$$

Since  $\Psi_P(x) \leq x^2$  for all  $x \in [0, 2]$ , we have  $\sum_{j=1}^{i-1} \Psi_P(\tilde{\lambda}_{j,a_n}) \xrightarrow{a.s.} \sum_{j=1}^{i-1} \tilde{\lambda}_{j,a_n}^2$  for each  $i$  and therefore  $\lambda'_{i,a_n} = \lambda_{i,a'_n} < \frac{1}{2}$  whenever  $i \geq N'_{\min} \triangleq c_{R,\sigma} \tilde{N}_n \log(n+1)^3$  for some constant  $\tilde{c}_{R,\sigma}$  depending only on  $(R, \sigma)$  and  $a'_n \triangleq \frac{a_n \log(n)}{1+\log(n)}$ .

Similarly, Assump.  $(R, \sigma)$  and the inequalities  $0 \leq \Psi_E(x) \leq x^2$  for all  $x \in [0, \frac{1}{2}]$  imply

$$(15) \quad 0 \leq \sum_{i=1}^n \Psi_E(\lambda_{i,a_n}) \frac{((W_i - \hat{\mu}_i)^2 - \hat{\sigma}_i^2)^2}{R^2} \xrightarrow{a.s.} \sum_{i=1}^n \min(R^4 \lambda_{i,a_n}^2, \lambda_{i,a_n} \frac{((W_i - \hat{\mu}_i)^2 - \hat{\sigma}_i^2)^2}{2R^2}).$$

We will now show that  $\hat{\sigma}_{\text{up},a_n} \xrightarrow{a.s.} \sigma$  and  $\hat{\sigma}_{\text{low},a_n} \xrightarrow{a.s.} \sigma$  by considering two cases.

**Case 1:**  $m_4 > 0$ . Since  $\frac{\log(a_n)}{n} \rightarrow 0$ , with probability 1 a finite integer  $N_{\min}$  exists for which  $\lambda_{i,a_n} < \frac{1}{2}$  whenever  $N_{\min} < i \leq n$ . Therefore, since  $d_n \triangleq \sqrt{2R^4 n \log(2/a_n)} \rightarrow \infty$ ,  $\log(a_n)/n \rightarrow 0$ , and each  $\lambda_{i,a_n} \leq \frac{1}{2}$ , we have

$$\begin{aligned} \frac{1}{d_n} \sum_{i=1}^n \lambda_{i,a_n} &= \frac{1}{d_n} \sum_{i=1}^{n \wedge N_{\min}} \lambda_{i,a_n} + \frac{1}{n} \sum_{i=N_{\min}+1}^n \frac{1}{\hat{m}_{4,i}} \xrightarrow{a.s.} \frac{1}{m_4} \quad \text{and} \\ \frac{1}{d_n} \sum_{i=1}^n \lambda_{i,a_n}^2 &= \frac{1}{d_n} \sum_{i=1}^{n \wedge N_{\min}} \lambda_{i,a_n}^2 + \sqrt{\frac{2R^4 \log(2/a_n)}{n}} \frac{1}{n} \sum_{i=N_{\min}+1}^n \frac{1}{\hat{m}_{4,i}^2} \xrightarrow{a.s.} 0. \end{aligned}$$

The almost sure boundedness,  $(W_i - \hat{\mu}_i)^2 \in [0, R]$  for all  $i$ , further implies that

$$\begin{aligned} \frac{1}{d_n} \sum_{i=1}^n \lambda_{i,a_n} (W_i - \hat{\mu}_i)^2 \\ = \frac{1}{d_n} \sum_{i=1}^{n \wedge N_{\min}} \lambda_{i,a_n} (W_i - \hat{\mu}_i)^2 + \frac{1}{n} \sum_{i=N_{\min}+1}^n [(\frac{1}{\hat{m}_{4,i}} - \frac{1}{m_4})(W_i - \hat{\mu}_i)^2 + \frac{(W_i - \hat{\mu}_i)^2}{m_4}] \xrightarrow{a.s.} \frac{\sigma^2}{m_4}. \end{aligned}$$

These results, together with the  $\Psi_E$  estimates (15) and the assumption  $\frac{\log(a_n)}{n} \rightarrow 0$ , ensure

$$|\hat{\sigma}_{\text{up},a_n}^2 - \sigma^2| \xrightarrow{a.s.} \left| \frac{\frac{1}{d_n} \sum_{i=1}^n \lambda_{i,a_n} (W_i - \hat{\mu}_i)^2}{\frac{1}{d_n} \sum_{i=1}^n \lambda_{i,a_n}} - \sigma^2 \right| + \frac{\frac{1}{d_n} R^2 \log(2/a_n) + R^4 \frac{1}{d_n} \sum_{i=1}^n \lambda_{i,a_n}^2}{\frac{1}{d_n} \sum_{i=1}^n \lambda_{i,a_n}} \xrightarrow{a.s.} 0.$$

Identical reasoning with  $(N'_{\min}, a'_n)$  substituted for  $(N_{\min}, a_n)$  ensures that  $\hat{\sigma}_{\text{low},a_n} \xrightarrow{a.s.} \sigma$ .

**Case 2:**  $m_4 = 0$ . Since  $(W_1 - \mu)^2 \xrightarrow{a.s.} 0$ , we have  $\hat{\sigma}_n^2 - \sigma^2 \xrightarrow{a.s.} \frac{1}{n} (\frac{R^2}{4} - \sigma^2) \rightarrow 0$ ,

$$|(W_n - \hat{\mu}_n)^2 - \sigma^2| \xrightarrow{a.s.} |(\mu - \hat{\mu}_n)^2 + 2(W_n - \mu)(\mu - \hat{\mu}_n)| \xrightarrow{a.s.} (\mu - \hat{\mu}_n)^2 + 2R|\mu - \hat{\mu}_j| \xrightarrow{a.s.} 0,$$

and therefore  $((W_n - \hat{\mu}_n)^2 - \hat{\sigma}_n^2)^2 \xrightarrow{a.s.} 0$ . Moreover, since  $\hat{m}_{4,i}^2 \xrightarrow{a.s.} R^4$  for each  $i$ , we have

$$\sum_{i=1}^n \lambda_{i,a_n} \xrightarrow{a.s.} \sum_{i=1}^n (\sqrt{16 \log(2/a_n)/(R^4 n)} \wedge \frac{1}{2}) = \sqrt{16n \log(2/a_n)/R^4} \wedge \frac{n}{2} \rightarrow \infty.$$

Therefore, our assumption  $\frac{\log(a_n)}{n} \rightarrow 0$ , the  $\Psi_E$  estimates (15), and the Silverman-Toeplitz theorem [37, Thm. 2] imply that

$$|\hat{\sigma}_{\text{up},a_n}^2 - \sigma^2| \leq \left| \frac{\sum_{i=1}^n \lambda_{i,a_n} (W_i - \hat{\mu}_i)^2}{\sum_{i=1}^n \lambda_{i,a_n}} - \sigma^2 \right| + \frac{R^2 \log(2/a_n) + \sum_{i=1}^n \lambda_{i,a_n} ((W_i - \hat{\mu}_i)^2 - \hat{\sigma}_i^2)^2 / (2R^2)}{\sum_{i=1}^n \lambda_{i,a_n}} \xrightarrow{a.s.} 0.$$

Parallel reasoning implies that  $\hat{\sigma}_{\text{low}, a_n}^2 \xrightarrow{a.s.} \sigma^2$  as

$$\sum_{i=1}^n \lambda'_{i, a_n} \geq \sum_{i=N'_{\min}}^n \left( \sqrt{\frac{16 \log(2/a'_n)}{R^4 n}} \wedge \frac{1}{2} \right) = \sqrt{\frac{16(n-N'_{\min})^2 \log(2/a'_n)}{R^4 n}} \wedge \frac{n-N'_{\min}}{2} \rightarrow \infty.$$

Since  $\hat{\sigma}_{\text{up}, a_n}$  and  $\hat{\sigma}_{\text{low}, a_n}$  both converge almost surely to  $\sigma$ , any

$$\tilde{\sigma}_{n,1} \in \underset{\tilde{\sigma} \in [\hat{\sigma}_{\text{low}, a_n}, \hat{\sigma}_{\text{up}, a_n}]}{\operatorname{argmax}} \tilde{q}_n(R, \delta - a_n, \tilde{\sigma}) \quad \text{and} \quad \tilde{\sigma}_{n,2} \in \underset{\tilde{\sigma} \in [\hat{\sigma}_{\text{low}, a_n}, \hat{\sigma}_{\text{up}, a_n}]}{\operatorname{argmax}} \tilde{q}_n^d(R, \delta - a_n, \tilde{\sigma}).$$

must converge almost surely to  $\sigma$  as well. Finally, since  $a_n \rightarrow 0$ ,

$$\hat{q}_n^d(R, \delta, a_n) = \tilde{q}_n^d(R, \delta - a_n, \tilde{\sigma}_{n,2}) \xrightarrow{a.s.} -\sigma \Phi^{-1}\left(\frac{\delta}{2}\right) \quad \text{and} \quad \hat{q}_n(R, \delta, a_n) \xrightarrow{a.s.} -\sigma \Phi^{-1}(\delta).$$

□

**REMARK 3** (Default settings). *As convenient default settings, one can choose  $a_n = \frac{\delta}{\sqrt{n}} \Phi^{-1}(1 - \delta)$  for a one-sided quantile bound and  $a_n = \frac{\delta}{\sqrt{n}} \Phi^{-1}(1 - \frac{\delta}{2})$  for a two-sided quantile bound and take  $(\tilde{q}_n, \tilde{q}_n^d)$  to be the efficient quantile bounds  $(q_n, q_n^d)$  of Thm. 6 with the default auxiliary bounds of Rem. 2.*

**6. Numerical Evaluation.** We now turn to a numerical evaluation of our efficient bounds. Python code implementing our bounds and reproducing all plots can be found at

[https://github.com/lmackey/gauss\\_conc/](https://github.com/lmackey/gauss_conc/).

**6.1. Efficient quantile bounds.** Fig. 1 compares the efficient two-sided quantile bound of Thm. 6 and Rem. 2 with the Bernstein quantile bound  $P(|S_n| > \frac{R_\sigma}{3\sqrt{n}} \log(2/\delta) + \sigma \sqrt{2 \log(2/\delta)}) \leq \delta$  [17, Thm. 2.10] and the two-sided quantile bounds obtained by inverting the Hoeffding (1) and Bentkus (7) tail bounds. As anticipated, the efficient bound converges to the optimal asymptotic size as  $n$  increases, while the inefficient Bernstein, Hoeffding, and Bentkus bounds remain bounded away from the optimum for all  $n$ .

Fig. 2 compares the efficient quantile bound of Thm. 6 and Rem. 2 to the two-sided quantile bounds obtained by inverting the Feller-Cramér (6), Berry-Esseen (BE) (4), and non-uniform BE (5) tail bounds with  $C_{R,\sigma} = \min(.3328(\frac{R_\sigma}{\sigma} + .429), .33554(\frac{R_\sigma}{\sigma} + .415))$  [64] and  $\tilde{C}_{R,\sigma} = \min(17.36\frac{R_\sigma}{\sigma}, 15.70\frac{R_\sigma}{\sigma} + 0.646)$  [65, p. 54].<sup>3</sup> The new efficient bound provides tighter estimates for higher confidence levels  $1 - \delta$ , due to the improved underlying tail decay.

**6.2. Application: Efficient empirical Berry-Esseen bounds.** Fig. 3 compares the efficient empirical Berry-Esseen (EBE) bound of Thm. 7 and Rem. 3 with the Hoeffding quantile bound,  $\mathbb{P}(S_n > R\sqrt{\log(1/\delta)/2}) \leq \delta$ ; the empirical Bernstein quantile bound [49, Thm. 4] commonly deployed in reinforcement learning [3, 50],

$$\mathbb{P}\left(S_n > \hat{\sigma} \sqrt{\log(2/\delta) \frac{n}{n-1}} + \frac{7}{3} R \log(2/\delta) \frac{\sqrt{n}}{n-1}\right) \leq \delta \quad \text{for} \quad \hat{\sigma}^2 \triangleq \frac{1}{n} \sum_{i \leq n} (W_i - \bar{W}_n)^2;$$

an empirical Bentkus quantile bound based on Kuchibhotla and Zheng [45, (33)]; and the efficient  $I_{n,3}$  quantile bound of Romano and Wolf [61],<sup>4</sup> with its free parameter  $\beta_n$  set to match our default value,  $a_n = \frac{\delta}{\sqrt{n}} \Phi^{-1}(1 - \delta)$ .

As advertised, the efficient EBE bound converges to the optimal asymptotic size as  $n$  increases, while the empirical Bernstein, empirical Bentkus, and Hoeffding bounds remain bounded away from the optimum for all  $n$ . Moreover, the efficient EBE bound provides a consistently tighter estimate than the Romano-Wolf bound.

<sup>3</sup>The BE, non-uniform BE, and Feller-Cramér bounds can be infinite, but we constrain them to be no larger than the always-valid Hoeffding bound.

<sup>4</sup>The Romano-Wolf bound can be infinite, but we constrain it to be no larger than the Hoeffding  $1 - (\delta - \beta_n)$  quantile bound.

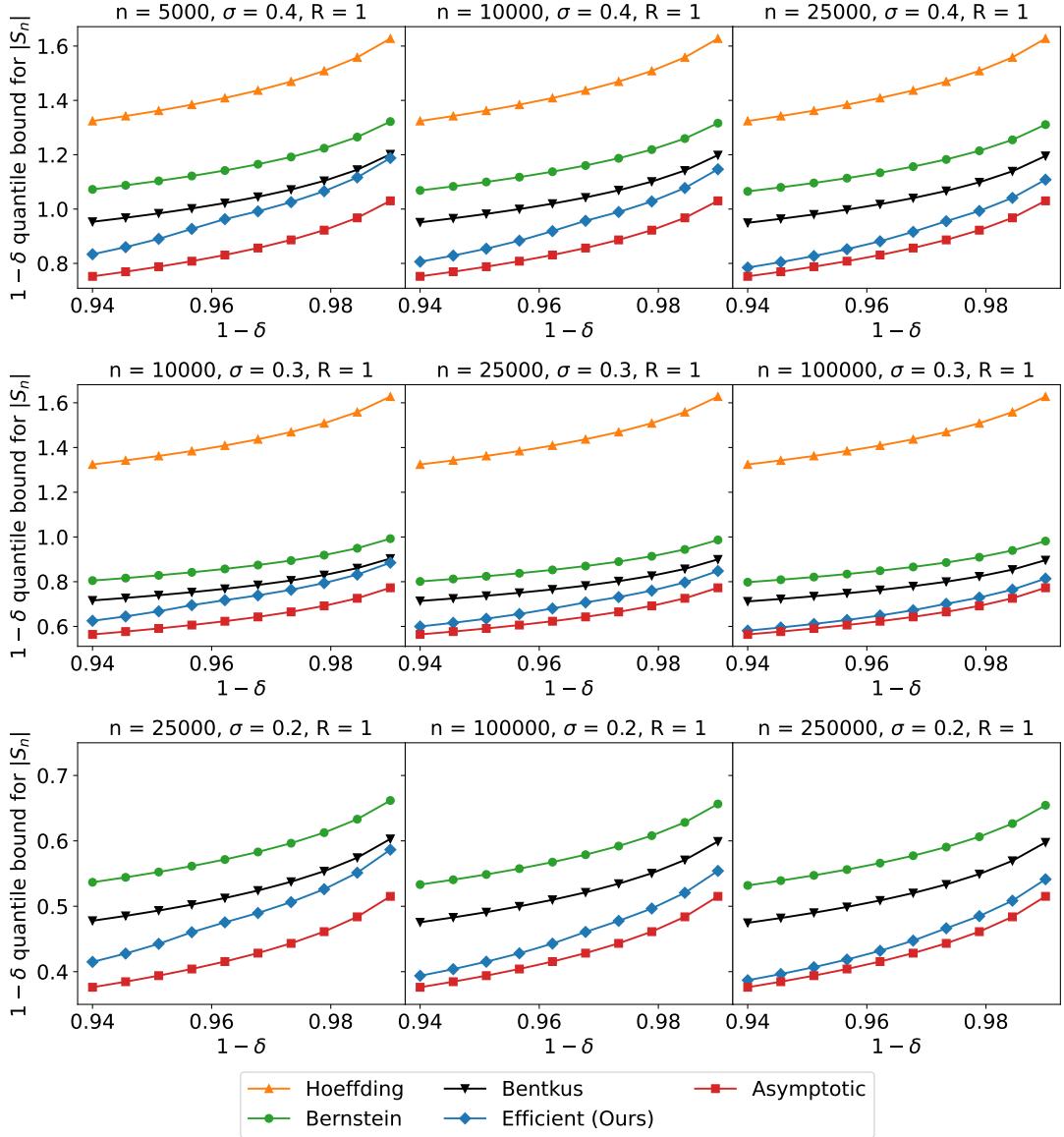


Fig 1: **Quantile bounds for  $|S_n|$**  with fixed boundedness parameter  $R = 1$  and varying sample size  $n$  and variance  $\sigma^2$ . Unlike the inefficient Hoeffding, Bernstein, and Bentkus bounds, the efficient quantile bound of Thm. 6 and Rem. 2 converges to the optimal asymptotic size as  $n$  increases.

6.3. *Application: Monte Carlo Confidence Intervals for Numerical Integration.* A common use of concentration inequalities is in numerical integration, to provide a confidence interval for the expectation  $\mathbb{E}(f(X_1))$  of a function  $f$  using an i.i.d. sample  $(f(X_i))_{i=1}^n$ . Recently, Jain et al. [40] employed the state-of-the-art predictable plug-in empirical Bernstein (PrPl-EB) and hedged capital betting (HCB) intervals of Waudby-Smith and Ramdas [69] for this purpose, as they provide some of the narrowest confidence intervals known for bounded observations. However, the PrPl-EB intervals are also provably inefficient, as they asymptotically match the suboptimal width of Bernstein's inequality rather than the optimal width of the CLT [69, Sec. E.3]. Fig. 4 compares the PrPl-EB and HCB intervals with the efficient empirical Berry-Esseen (EBE) interval of Thm. 7 and Rem. 3 for each of six benchmark test

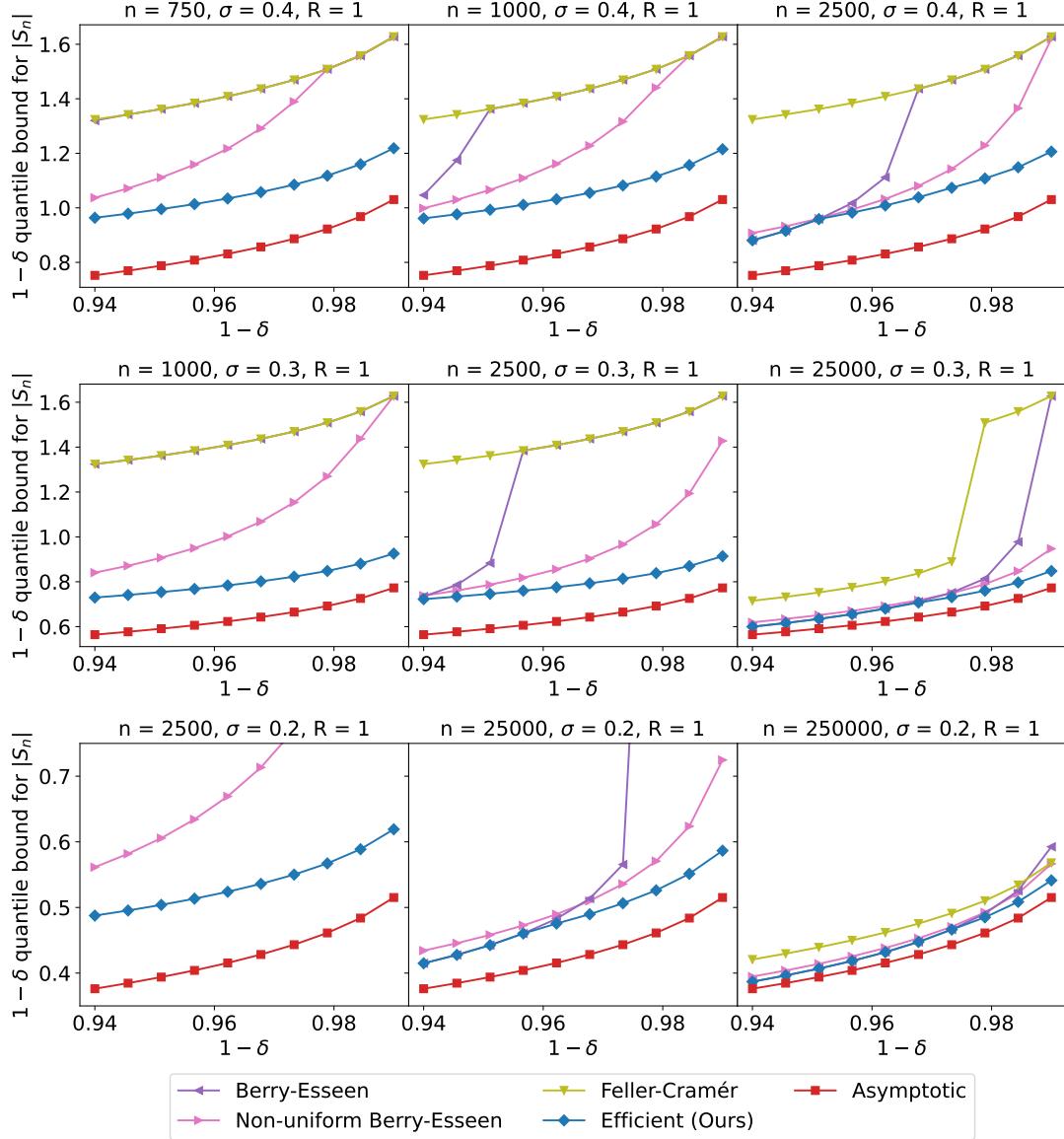


Fig 2: **Efficient quantile bounds for  $|S_n|$**  with fixed boundedness parameter  $R = 1$  and varying sample size  $n$  and variance  $\sigma^2$ . The efficient quantile bound of Thm. 6 and Rem. 2 provides tighter estimates than the Berry-Esseen, non-uniform Berry-Esseen, and Feller-Cramér bounds for higher confidence levels  $1 - \delta$ .

functions studied by Jain et al. [40]. In each case, we find that the PrPl-EB and HCB intervals remain bounded away from the optimal asymptotic width, while the EBE interval converges to the optimum, resulting in the tightest bounds for larger sample sizes.

**7. Discussion and Related Work.** In this work, we have derived new computable tail and quantile bounds for the scaled deviations  $S_n = \sqrt{n}(\bar{W}_n - \mathbb{E}(W_1))$  with asymptotically optimal size, finite-sample validity, and sub-Gaussian decay. These bounds enable the construction of efficient confidence intervals with correct coverage for any sample size. Our concentration inequalities arise from new computable bounds on a non-uniform Kolmogorov distance and the  $p$ -Wasserstein distances to a Gaussian.

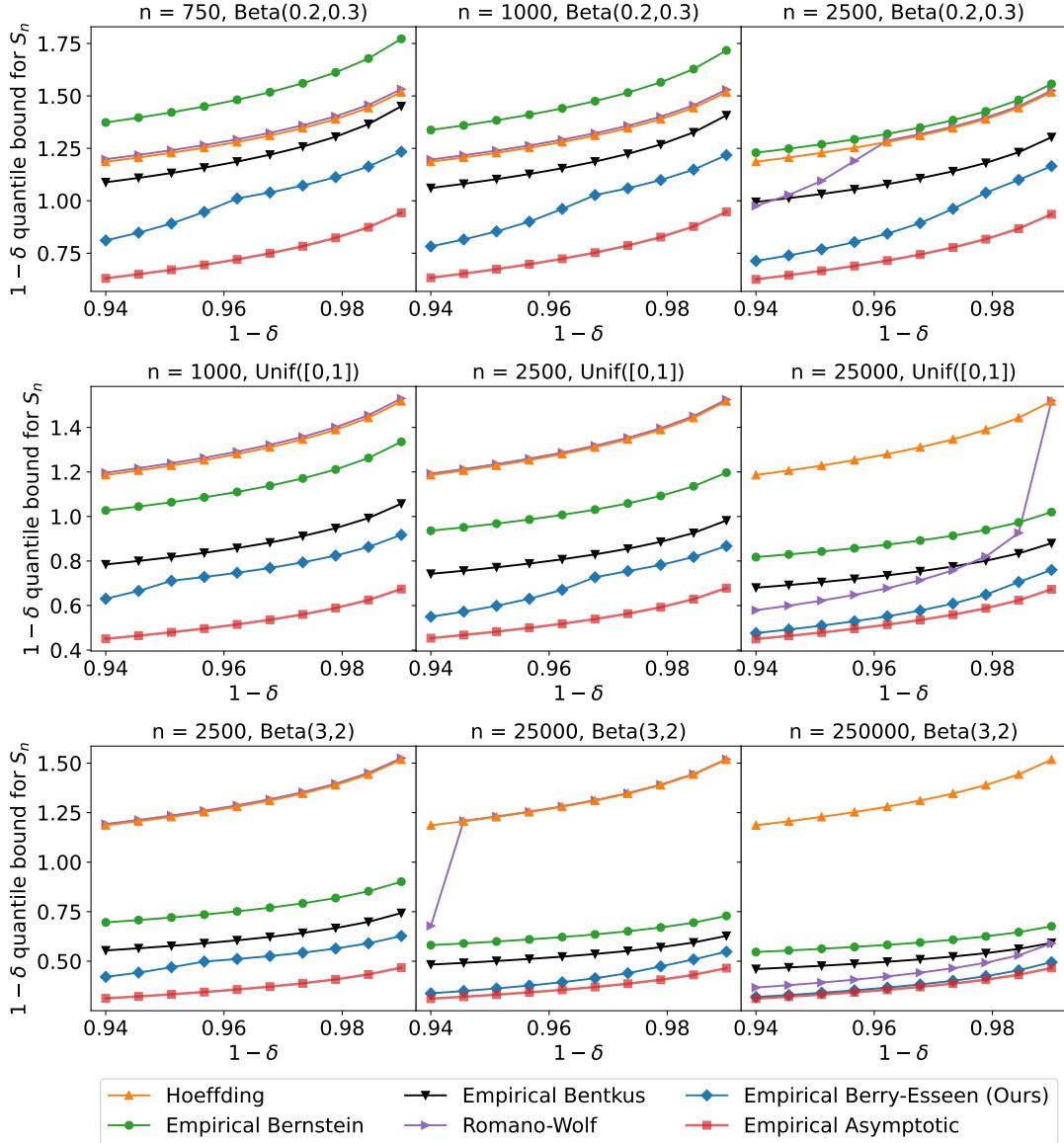


Fig 3: **Empirical quantile bounds for  $S_n$**  with fixed boundedness parameter  $R = 1$  and varying sample size  $n$  and empirical variance  $\hat{\sigma}^2$ . The efficient empirical Berry-Esseen (EBE) bound of Thm. 7 and Rem. 3 converges to the optimal asymptotic size as  $n$  increases, while the empirical Bernstein, empirical Bentkus, and Hoeffding bounds remain bounded away from the optimum for all  $n$ . In addition, the efficient EBE bound provides tighter estimates than the efficient  $I_{n,3}$  interval of Romano and Wolf [61].

The notion of efficient confidence intervals for the mean was introduced by Romano and Wolf [61]. In their Thm. 2.1, [Romano and Wolf](#) showed that efficient—that is, asymptotically minimal-length when scaled by  $\sqrt{n}$ —confidence intervals must converge to the width of the asymptotic Gaussian intervals implied by the CLT (2). Moreover, Sec. 3 of [Romano and Wolf](#) surveys a number of procedures for constructing confidence intervals that are either finite-sample invalid (including the bootstrap [27] and methods based on Edgeworth expansions [36]) or inefficient (including the methods of Anderson [2] and Gasko [33]). [Romano and Wolf](#) conclude by developing an efficient valid confidence interval for the mean of vari-

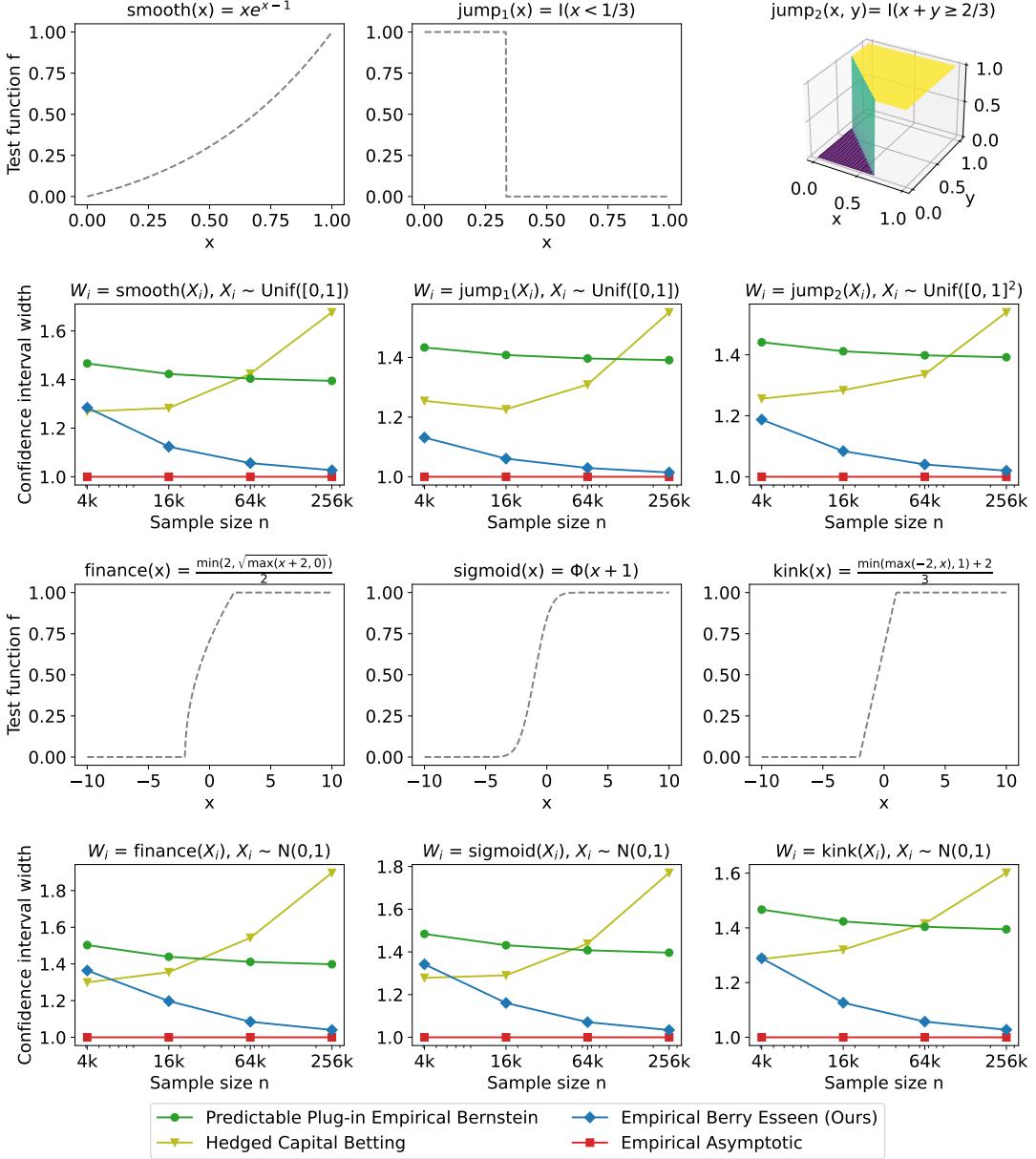


Fig 4: **Widths of Monte Carlo confidence intervals for  $\mathbb{E}(f(X_1))$**  relative to the empirical asymptotic width  $\hat{\sigma}\Phi^{-1}(1 - \frac{\delta}{2})\frac{2}{\sqrt{n}}$  with  $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (W_i - \bar{W}_n)^2$  and confidence level  $1 - \delta = 0.95$ . We display relative widths averaged across 10 independent replicates of the experiment. The efficient empirical Berry-Esseen intervals of Thm. 7 and Rem. 3 converge to the optimal asymptotic size as  $n$  increases, while the predictable plug-in empirical Bernstein and hedged capital betting bounds remain bounded away from the optimum for all  $n$ .

ables supported on  $[0, 1]$  but report that it is “unfortunately, much too wide for a reasonable sample.” Our new efficient bounds are developed in an entirely different manner, and, as we demonstrate in Sec. 6, improve upon the Romano-Wolf interval and the most commonly used empirical concentration inequalities.

Our zero-bias coupling arguments generalize the uniform,  $u$ -independent bounds of Chen, Goldstein and Shao [21] and Ross [62] to derive tighter non-uniform bounds with sub-

Gaussian decay. Our Wasserstein-bounding arguments build on the pioneering work of Bonis [16] who derived Wasserstein convergence rates for the CLT with inexplicit constants. Our new arguments lead to tighter estimates of the distance to Gaussianity and explicit, practical constants.

Our results also suggest simple strategies for making any concentration inequality or confidence region efficient. For efficient concentration, one can simply take the minimum of any existing tail bound and Thm. 1 or Thm. 5 to simultaneously reap the small-sample benefits of the former and the large-sample benefits of the latter. For efficient confidence, one can divide the total confidence budget between an existing region and the efficient region of Thm. 7 and then intersect the two regions. The result will remain efficient if the budget allocated to the auxiliary region vanishes as  $n$  grows. These strategies are particularly relevant given the recent renewed interest in deriving tighter concentration inequalities for bounded random variables [see, e.g., 42, 53, 69].

In Sec. 6.3, we demonstrated the usefulness of our bounds by constructing tighter confidence intervals for Monte Carlo integration. A second natural application is to the construction of *risk-controlling prediction sets*, that is, sets of predicted outcomes that are guaranteed to have high expected utility with high probability according to a given quality measure [6]. Bates et al. [6] reduce the problem of risk control to constructing confidence intervals for the unknown expected utility and employ the concentration inequalities of Hoeffding [38], Bentkus [10], Maurer and Pontil [49], and Waudby-Smith and Ramdas [69] to form their prediction sets. A tighter confidence interval based on efficient concentration would yield more informative and hence more actionable prediction sets.

Finally, while our work has focused on concentration of the sample mean assuming boundedness, non-zero variance, and a sampling distribution independent of the sample size, we conjecture that our analyses can be adapted to (1) derive efficient, computable concentration inequalities for more general classes of asymptotically normal statistics or for self-normalized statistics, in the spirit of Jing, Shao and Wang [41] and Pinelis [57]; (2) take advantage of other favorable properties of the  $W_1$  distribution like symmetry or a vanishing third moment; and (3) account for the non-Gaussian limits that arise when the distribution underlying  $(W_i)_{i=1}^n$  is allowed to vary with the sample size  $n$ . We leave these important challenges for future work.

**Acknowledgments.** We thank Diego Martinez-Taboada and Aaditya Ramdas for providing their empirical Bernstein variance bound code and Aadit Jain, Fred J. Hickernell, Art B. Owen, and Aleksei G. Sorokin for sharing their Monte Carlo confidence interval code.

## APPENDIX: TABLE OF CONTENTS

A Proof of Thm. 4: General zero-bias tail bounds . . . . .	20
A.1 Proof of Lem. A.1: Growth of $h_u$ . . . . .	22
B Proof of Lem. 5: Wasserstein upper bound . . . . .	23
C Additional Lemmas . . . . .	30
References . . . . .	32

## APPENDIX A: Proof of Thm. 4: General zero-bias tail bounds

Fix any  $u \in \mathbb{R}$ . Our proof structure, based on Stein's method, mimics that of Ross [62, Thm. 3.27] but employs  $u$ -dependent bounds in place of the  $u$ -independent bounds invoked by Ross.

*Solving the Stein equation.* First we define the *Stein equation* [66]

$$f'_u(w) - wf_u(w) = \mathbb{I}(w \leq u) - \Phi(u).$$

By Chen, Goldstein and Shao [21, Lem. 2.2], the absolutely continuous function

$$(16) \quad f_u(w) = \sqrt{2\pi} \exp\left(\frac{w^2}{2}\right) \Phi(w \wedge u) \Phi^c(w \vee u)$$

solves the Stein equation, and we can therefore write

$$\mathbb{P}(S^* \leq u) - \Phi(u) = \mathbb{E}[f'_u(S^*) - S^* f_u(S^*)] = \mathbb{E}[S f_u(S) - S^* f_u(S^*)],$$

where the final equality uses the definition of the zero-biased distribution. Negating both sides, we obtain

$$\mathbb{P}(S^* > u) - \Phi^c(u) = \mathbb{E}[S^* f_u(S^*) - S f_u(S)].$$

*Bounding the Stein solution.* Now let  $g_u(w) = wf_u(w)$ . By Chen, Goldstein and Shao [21, Eq. (2.81)], the function

$$(17) \quad \begin{aligned} h_u(w) &\triangleq \begin{cases} \sqrt{2\pi} \Phi^c(u)((1+w^2) \exp(w^2/2) \Phi(w) + w/\sqrt{2\pi}) & \text{if } w \leq u \\ \sqrt{2\pi} \Phi(u)((1+w^2) \exp(w^2/2) \Phi^c(w) - w/\sqrt{2\pi}) & \text{if } w > u \end{cases} \\ &= w(\Phi^c(u) - \mathbb{I}(w > u)) + (1+w^2)f_u(w) \\ &= w(\mathbb{I}(w \leq u)\Phi^c(u) - \mathbb{I}(w > u)\Phi(u)) + (1+w^2)f_u(w) \end{aligned}$$

matches the derivative of  $g_u(w)$  whenever  $w \neq u$ . Since the absolute continuity of  $f_u$  implies that  $g_u$  is absolutely continuous on compact intervals, the fundamental theorem of calculus yields

$$\mathbb{E}[g_u(S^*) - g_u(S)] = \mathbb{E}[\int_0^1 h_u(S^* + x(S - S^*))(S^* - S)dx] = \mathbb{E}[h_u(S')(S^* - S)].$$

Our next result, proved in Sec. A.1 provides a suitable  $u$ -dependent bound on  $h_u$ .

LEMMA A.1 (Growth of  $h_u$ ). *For any  $u \geq 0, \lambda \in [0, 1]$ , and  $w \in \mathbb{R}$ ,*

$$0 \leq h_u(w) \leq h_u(\lambda u) + (h_u(u) - h_u(\lambda u))\mathbb{I}(w > \lambda u) - u\mathbb{I}(w > u).$$

for  $h_u$  defined in (17).

*$S^*$  tail bound.* Fix any  $\lambda \in [0, 1]$ . Lem. A.1, our almost sure assumption  $S^* - S \leq \delta$ , and the definition of  $S'$  together imply that

$$(18) \quad \begin{aligned} \mathbb{P}(S^* > u) - \Phi^c(u) &\leq \delta \mathbb{E}[h_u(S')] \\ &\leq \delta [h_u(\lambda u) + (h_u(u) - h_u(\lambda u))\mathbb{P}(S' > \lambda u) - u\mathbb{P}(S' > u)]. \end{aligned}$$

*From  $S$  tails to  $S^*$  tails.* When  $S - S^* \leq \delta$  almost surely, we additionally have

$$\begin{aligned} S - S' &\leq (S - S^*)(1 - U) \leq \delta \quad \text{and therefore} \\ \mathbb{P}(S > u + \delta) &= \mathbb{P}(S^* \wedge S' > u + \delta + S^* \wedge S' - S) \\ &\leq \mathbb{P}(S^* \wedge S' > u) \leq \mathbb{P}(S^* > u) \wedge \mathbb{P}(S' > u). \end{aligned}$$

Combining this inequality with our  $S^*$  tail bound (18) yields

$$\mathbb{P}(S > u + \delta) \leq \Phi^c(u) + \delta [h_u(\lambda u) + (h_u(u) - h_u(\lambda u))\mathbb{P}(S' > \lambda u) - u\mathbb{P}(S > u + \delta)].$$

Rearranging the terms of this expression yields the final advertised result.

**A.1. Proof of Lem. A.1: Growth of  $h_u$ .** Our proof relies on the following  $\Phi^c$  property.

LEMMA A.2 (Growth of  $\Phi^c$ ). *The following function is decreasing for  $w \geq 0$ :*

$$b(w) \triangleq \sqrt{2\pi}(1+w^2) \exp(w^2/2) \Phi^c(w) - w.$$

PROOF. Fix any  $w \geq 0$ . The derivative of  $b$  takes the form

$$b'(w) = \Phi^c(w)(3w + w^3)/\varphi(w) - (2 + w^2).$$

Since the Mills ratio  $\frac{\Phi^c(w)}{\varphi(w)} < \frac{2+w^2}{3w+w^3}$  [46, Thm. 2.1],  $b'(w)$  is negative and  $b$  is decreasing.  $\square$

Fix any  $u \geq 0, \lambda \in [0, 1]$ , and  $w \in \mathbb{R}$ . We divide our proof into cases based on  $w$ .

*Lower bound,  $w \neq u$ :* Since  $g_u$  is differentiable for  $w \neq u$  and increasing by Chen, Goldstein and Shao [21, Lem. 2.3],  $h_u(w) = g'_u(w) \geq 0$ .

*Lower bound,  $w = u$ :* We have  $h_u(u) = \lim_{v \uparrow u} h_u(v) \geq 0$ .

*Upper bound,  $w > u$ :* Since  $b$  is decreasing (Lem. A.2), the definition of  $f_u$  (16) implies that

$$\begin{aligned} (1+w^2)f_u(w)\mathbb{I}(w > u) &= \sqrt{2\pi}(1+w^2) \exp(w^2/2) \Phi^c(w)\Phi(u)\mathbb{I}(w > u) \\ &\leq (b(u) + w)\Phi(u)\mathbb{I}(w > u). \end{aligned}$$

Hence,

$$h_u(w)\mathbb{I}(w > u) = (1+w^2)f_u(w)\mathbb{I}(w > u) - w\Phi(u)\mathbb{I}(w > u) \leq b(u)\Phi(u)\mathbb{I}(w > u).$$

*Upper bound,  $\lambda u \geq w \geq 0$ :* Since  $g_u$  is increasing [21, Lem. 2.3],  $f_u$  is increasing for  $w < u$ , and  $\lambda u f_u(\lambda u) + \Phi^c(u)$  is nonnegative, we have

$$\begin{aligned} h_u(w)\mathbb{I}(\lambda u \geq w \geq 0) &= (f_u(w) + w(w f_u(w) + \Phi^c(u)))\mathbb{I}(\lambda u \geq w \geq 0) \\ &\leq [f_u(\lambda u) + w(\lambda u f_u(\lambda u) + \Phi^c(u))]\mathbb{I}(\lambda u \geq w \geq 0) \\ &\leq [f_u(\lambda u) + \lambda u(\lambda u f_u(\lambda u) + \Phi^c(u))]\mathbb{I}(\lambda u \geq w \geq 0) \\ &= h_u(\lambda u)\mathbb{I}(\lambda u \geq w \geq 0) = h_u(\lambda u)(\mathbb{I}(w \geq 0) - \mathbb{I}(w > \lambda u)). \end{aligned}$$

*Upper bound,  $w < 0$ :* Since  $w < 0 \leq u$ , we use the definition of  $f_u$  (16), the fact that  $b$  is decreasing (Lem. A.2), and the nonnegativity of  $\Phi^c$  to derive

$$\begin{aligned} (1+w^2)f_u(w)\mathbb{I}(w < 0) &= \sqrt{2\pi} \exp(w^2/2) \Phi^c(|w|)\Phi^c(u)\mathbb{I}(w < 0) \\ &= (b(|w|) + |w|)\Phi^c(u)\mathbb{I}(w < 0) \\ &\leq (b(0) + |w|)\Phi^c(u)\mathbb{I}(w < 0). \end{aligned}$$

Since  $\lambda u \geq 0$  by assumption, our prior derivation implies  $h_u(\lambda u) \geq h_u(0)$  and hence,

$$\begin{aligned} h_u(w)\mathbb{I}(w < 0) &= ((b(0) + |w|) - |w|)\Phi^c(u)\mathbb{I}(w < 0) \\ &= b(0)\Phi^c(u)\mathbb{I}(w < 0) = \sqrt{\frac{\pi}{2}}\Phi^c(u)\mathbb{I}(w < 0) \\ &= h_u(0)\mathbb{I}(w < 0) \leq h_u(\lambda u)\mathbb{I}(w < 0). \end{aligned}$$

*Upper bound*,  $u \geq w > \lambda u$ . Since  $u, w, f_u(u)$ , and  $\Phi^c(u)$  are nonnegative,  $g_u$  is increasing [21, Lem. 2.3], and  $u \geq w$ , we have

$$\begin{aligned} h_u(w)\mathbb{I}(u \geq w > \lambda u) &= (f_u(w) + w(wf_u(w) + \Phi^c(u)))\mathbb{I}(u \geq w > \lambda u) \\ &\leq (f_u(u) + w(u f_u(u) + \Phi^c(u)))\mathbb{I}(u \geq w > \lambda u) \\ &\leq (f_u(u) + u(u f_u(u) + \Phi^c(u)))\mathbb{I}(u \geq w > \lambda u) \\ &= h_u(u)\mathbb{I}(u \geq w > \lambda u) = h_u(u)(\mathbb{I}(w > \lambda u) - \mathbb{I}(w > u)). \end{aligned}$$

*Complete upper bound*: Taken together, our upper bounds yield

$$\begin{aligned} h_u(w) &\leq b(u)\Phi(u)\mathbb{I}(w > u) + h_u(u)(\mathbb{I}(w > \lambda u) - \mathbb{I}(w > u)) \\ &\quad + h_u(\lambda u)(\mathbb{I}(w \geq 0) - \mathbb{I}(w > \lambda u)) + h_u(\lambda u)\mathbb{I}(w < 0) \\ &\leq (b(u)\Phi(u) - h_u(u))\mathbb{I}(w > u) + (h_u(u) - h_u(\lambda u))\mathbb{I}(w > \lambda u) + h_u(\lambda u) \\ &= (h_u(u) - h_u(\lambda u))\mathbb{I}(w > \lambda u) + h_u(\lambda u) - u\mathbb{I}(w > u). \end{aligned}$$

## APPENDIX B: Proof of Lem. 5: Wasserstein upper bound

We begin by defining some convenient shorthand notation. For each  $p \geq 1$  and  $q \in [0, 1]$ , we define  $\|\text{Binomial}(n, q)\|_p \triangleq \|V\|_p$  for  $V \sim \text{Binomial}(n, q)$  and make use of the constants

$$\begin{aligned} \widetilde{R}_\sigma &\triangleq R_\sigma/\sigma = \frac{1}{2}\tilde{R} + \frac{1}{2}\sqrt{\tilde{R}^2 - 4}, \\ A_p &\triangleq \frac{\sqrt{e}\sqrt{p+2}(2e)^{1/p}}{\sqrt{2}}, \quad A_{n,p}^* \triangleq \frac{(p+2)n^{1/p}}{2\sqrt{n}}, \quad \tilde{A}_{n,p} \triangleq A_{n,p}^*\tilde{R}^{-2/p}, \\ U_{n,p} &\triangleq A_p + \tilde{R}\tilde{A}_{n,p}, \quad \tilde{U}_{n,p} \triangleq \sqrt{2}A_p + 2^{1/p}\tilde{R}\tilde{A}_{n,p}, \quad \text{and} \end{aligned}$$

$$(19) \quad C_{n,p} \triangleq \min \begin{cases} \tilde{A}_{n,p}\tilde{R}2^{1/p} + \sqrt{2}A_p \\ \sqrt{2}\left(\frac{2\Gamma(\frac{p+1}{2})}{\sqrt{\pi}}\right)^{1/p} \begin{cases} \tilde{R}^{1-2/p} & \text{if } p < 4 \\ \tilde{R}^{1-2/p} \wedge \frac{\tilde{R}}{\sqrt{n}}\sqrt{\|\text{Binomial}(n, \frac{2}{\tilde{R}^2})\|_{p/2}} & \text{if } p \geq 4. \end{cases} \end{cases}$$

We will focus principally on establishing the bound  $\frac{1}{\sigma}\mathcal{W}_p(S_n, \mathcal{N}(0, \sigma^2)) \leq \omega_p^R(\sigma)$  and derive the secondary bound  $\omega_p^{R,\kappa,1}(\sigma) \leq \omega_p^{R,\kappa,2}(\sigma)$  through a series of asides demarcated by a vertical bar on the left-hand side. For the case of  $p = 1$ , we invoke Chen, Goldstein and Shao [21, Cor. 4.2] and Lem. 1 in turn to find that

$$\frac{1}{\sigma}\mathcal{W}_1(S_n, \mathcal{N}(0, \sigma^2)) \leq \frac{\mathbb{E}[|W_1 - \mathbb{E}(W_1)|^3]}{\sigma^3\sqrt{n}} \leq \frac{R_\sigma}{\sigma\sqrt{n}} = \omega_p^R(\sigma).$$

When  $p > n + 1$  the triangle inequality and Lems. C.1 and C.2 imply that

$$\sigma^{-1}\mathcal{W}_p(S_n, \mathcal{N}(0, \sigma^2)) \leq \|Z\|_p + \|S_n/\sigma\|_p \leq \sqrt{p-1} + \sqrt{p-1}R_\sigma/\sigma.$$

Now fix any  $p \in [2, n + 1]$ , define  $\tilde{S}_n \triangleq S_n/\sigma$ , and let  $Z \sim \mathcal{N}(0, 1)$  be independent of  $(W_i)_{i \geq 1}$ . For ease of notation, we will write  $\mathcal{W}_p(\tilde{S}_n, Z)$  in place of  $\mathcal{W}_p(\tilde{S}_n, \mathcal{N}(0, 1))$  and use  $(X_i)_{i \geq 1}$  to represent the centered and rescaled random variables

$$X_i \triangleq \frac{W_i - \mathbb{E}[W_i]}{\sigma\sqrt{n}},$$

which satisfy  $\tilde{S}_n = \sum_{i \leq n} X_i$ ,  $\text{Var}(X_1) = \frac{1}{n}$ , and  $\|X_1\|_\infty \leq \widetilde{R}_\sigma/\sqrt{n}$  by Lem. 1. Hence, for all  $k \geq 0$ , the moments of  $X_1^k$  can be upper bounded as

$$\|X_1^k\|_p = (\mathbb{E}[|X_1|^{kp}])^{1/p} \leq (\|X_1\|_\infty^{kp-2}\mathbb{E}[X_1^2])^{1/p} \leq \frac{\widetilde{R}_\sigma^k}{\widetilde{R}_\sigma^{2/p}\sqrt{n^k}}.$$

Consider a random index  $I \sim \text{Uniform}(\{1, \dots, n\})$  and a sequence  $(X'_i)_{i \geq 1} \stackrel{d}{=} (X_i)_{i \geq 1}$  with  $(I, (X'_i)_{i \geq 1}, (X_i)_{i \geq 1})$  mutually independent. Define an exchangeable copy of  $\tilde{S}_n$ ,

$$S'_n \triangleq \tilde{S}_n + (X'_I - X_I),$$

and the exchangeable pair difference,

$$Y \triangleq \tilde{S}_n - S'_n.$$

Let  $h_k(x) \triangleq e^{x^2/2} \frac{\partial^k}{\partial x^k} e^{-x^2/2}$  designate the  $k$ -th Hermite polynomial and define  $H_k \triangleq h_k(Z)$ . Finally, fix any  $\kappa > \frac{\tilde{R}^2}{n}$  and define

$$M_{n,\kappa} \triangleq \sqrt{1 - \frac{\tilde{R}^2}{n\kappa}}.$$

A slight modification of Bonis [16, Thm. 3] shows that

$$\begin{aligned} (20) \quad \mathcal{W}_p(\tilde{S}_n, Z) &\leq \int_0^{-\frac{1}{2} \log(1 - \frac{\tilde{R}^2}{n\kappa})} \|e^{-t} \tilde{S}_n - \frac{e^{-2t}}{\sqrt{1-e^{-2t}}} Z\|_p dt \\ &\quad + \int_{-\frac{1}{2} \log(1 - \frac{\tilde{R}^2}{n\kappa})}^{\infty} e^{-t} \|n\mathbb{E}[Y | \tilde{S}_n] - \tilde{S}_n\|_p dt \\ &\quad + \int_{-\frac{1}{2} \log(1 - \frac{\tilde{R}^2}{n\kappa})}^{\infty} \frac{e^{-2t} \|H_1\|_p}{\sqrt{1-e^{-2t}}} \left\| \frac{n}{2} \mathbb{E}[Y^2 | \tilde{S}_n] - 1 \right\|_p dt \\ &\quad + \sum_{k \geq 3} \int_{-\frac{1}{2} \log(1 - \frac{\tilde{R}^2}{n\kappa})}^{\infty} \frac{e^{-kt} \|H_{k-1}\|_p}{k! (\sqrt{1-e^{-2t}})^{k-1}} n \|\mathbb{E}[Y^k | \tilde{S}_n]\|_p dt \\ &\triangleq (a_0) + (a_1) + (a_2) + (a_3). \end{aligned}$$

**Bounding  $(a_0)$ .** We first bound  $(a_0)$ . To this end, fix  $\epsilon > 0$  and select  $G \sim \mathcal{N}(0, 1)$  independent from  $Z$  such that  $\|\tilde{S}_n - G\|_p \leq \mathcal{W}_p(\tilde{S}_n, Z) + \epsilon$ . By the triangle inequality,

$$\begin{aligned} (a_0) &\leq \int_0^{-\frac{1}{2} \log(1 - \frac{\tilde{R}^2}{n\kappa})} e^{-t} \|\tilde{S}_n - G\|_p + \|e^{-t} G - \frac{e^{-2t}}{\sqrt{1-e^{-2t}}} Z\|_p dt \\ &\stackrel{(a)}{=} (1 - M_{n,\kappa}) \|\tilde{S}_n - G\|_p + \|Z\|_p \int_0^{-\frac{1}{2} \log(1 - \frac{\tilde{R}^2}{n\kappa})} \frac{e^{-t}}{\sqrt{1-e^{-2t}}} dt \\ &\leq (1 - M_{n,\kappa}) (\mathcal{W}_p(\tilde{S}_n, Z) + \epsilon) + \|Z\|_p \left( \frac{\pi}{2} - \sin^{-1}(M_{n,\kappa}) \right) \end{aligned}$$

where (a) follows as  $e^{-t} G - \frac{e^{-2t}}{\sqrt{1-e^{-2t}}} Z \stackrel{d}{=} \frac{e^{-t}}{\sqrt{1-e^{-2t}}} Z$ . Since  $\epsilon > 0$  was arbitrary, we have

$$(a_0) \leq (1 - M_{n,\kappa}) \mathcal{W}_p(\tilde{S}_n, Z) + \|Z\|_p \left( \frac{\pi}{2} - \sin^{-1}(M_{n,\kappa}) \right).$$

In addition, if  $p = 2$ , by independence of  $Z$  and  $S_n$  we obtain that

$$\begin{aligned} (a_0) &\leq \int_0^{-\frac{1}{2} \log(1 - \frac{\tilde{R}^2}{n\kappa})} \sqrt{e^{-2t} \|\tilde{S}_n\|_2^2 + \frac{e^{-4t}}{1-e^{-2t}} \|Z\|_2^2} dt \\ &\leq \int_0^{-\frac{1}{2} \log(1 - \frac{\tilde{R}^2}{n\kappa})} \frac{e^{-t}}{\sqrt{1-e^{-2t}}} dt \leq \frac{\pi}{2} - \sin^{-1}(M_{n,\kappa}). \end{aligned}$$

These results together with the bound (20) yield the inequality

$$\mathcal{W}_p(\tilde{S}_n, Z) \leq \left[ \|Z\|_p \left( \frac{\pi}{2} - \sin^{-1} \left( \sqrt{1 - \frac{\tilde{R}^2}{n\kappa}} \right) \right) + (a_1) + (a_2) + (a_3) \right] \begin{cases} \frac{1}{M_{n,\kappa}} & \text{if } p > 2 \\ 1 & \text{if } p = 2. \end{cases}$$

**Bounding  $(a_1)$ .** We will next bound  $(a_1)$  of (20). Note that, since  $I \sim \text{Uniform}(\{1, \dots, n\})$ ,

$$\mathbb{E}[Y | \tilde{S}_n] = \frac{1}{n} \sum_{i \leq n} \mathbb{E}[X_i - X'_i | \tilde{S}_n] = \frac{1}{n} \sum_{i \leq n} X_i - \mathbb{E}[X_i] = \frac{1}{n} \tilde{S}_n,$$

and hence  $\|n\mathbb{E}[Y | \tilde{S}_n] - \tilde{S}_n\|_p = 0$ . Therefore  $(a_1) = 0$ .

**Bounding (a<sub>2</sub>).** We now turn to bounding (a<sub>2</sub>) of (20). By Jensen's inequality,

$$\|\frac{n}{2}\mathbb{E}[Y^2 | \tilde{S}_n] - 1\|_p \leq \|(\frac{1}{2}\sum_{i \leq n} \mathbb{E}[X_i^2] + X_i^2) - 1\|_p = \frac{1}{2}\|\sum_{i \leq n} X_i^2 - 1\|_p.$$

Moreover, by Lem. 1,  $|X_i^2 - 1/n| \stackrel{a.s.}{\leq} \max(\frac{1}{n}\tilde{R}_\sigma^2 - \frac{1}{n}, \frac{1}{n})$ . First suppose  $\frac{1}{n} \geq \frac{\tilde{R}_\sigma^2 - 1}{n}$ . Since  $\mathbb{E}[X_i^2] = 1/n$ , we know  $\|X_i^2 - 1/n\|_p$  is maximized when  $\frac{n}{\tilde{R}_\sigma^2} X_i^2 \sim \text{Bernoulli}(\frac{1}{\tilde{R}_\sigma^2})$ . Hence

$$\|X_i^2 - 1/n\|_p \leq \frac{1}{n} \left[ \frac{\tilde{R}_\sigma^2 - 1}{\tilde{R}_\sigma^2} \right]^{1/p} ((\tilde{R}_\sigma^2 - 1)^{p-1} + 1)^{1/p}.$$

Now suppose  $\tilde{R}_\sigma^2 - 1 \geq 1$ . We instead obtain that

$$\|X_i^2 - 1/n\|_p \leq (\|X_i^2 - 1/n\|_\infty^{p-2} \|X_i^2 - 1/n\|_2^2)^{1/p} \leq \frac{1}{n} (\tilde{R}_\sigma^2 - 1)^{1-1/p}.$$

Therefore using the Marcinkiewicz-Zygmund inequality (Lem. C.2) and the Rosenthal inequality (Lem. C.5) we find that

$$\|\frac{n}{2}\mathbb{E}[Y^2 | \tilde{S}_n] - 1\|_p \leq \frac{1}{2\sqrt{n}} \min \left\{ \frac{\sqrt{p-1}(\tilde{R}_\sigma^2 - 1)^{1-1/p}}{(\tilde{R}_\sigma^2 - 1)^{1-1/p} A_{n,p}^* + \sqrt{\tilde{R}_\sigma^2 - 1} A_p} \right.$$

for  $\tilde{R}_\sigma^2 \geq 2$  and

$$\|\frac{n}{2}\mathbb{E}[Y^2 | \tilde{S}_n] - 1\|_p \leq \frac{1}{2\sqrt{n}} \min \left\{ \frac{\sqrt{p-1} \left[ \frac{\tilde{R}_\sigma^2 - 1}{\tilde{R}_\sigma^2} \right]^{1/p} ((\tilde{R}_\sigma^2 - 1)^{p-1} + 1)^{1/p}}{\left[ \frac{\tilde{R}_\sigma^2 - 1}{\tilde{R}_\sigma^2} \right]^{1/p} ((\tilde{R}_\sigma^2 - 1)^{p-1} + 1)^{1/p} A_{n,p}^* + \sqrt{\tilde{R}_\sigma^2 - 1} A_p} \right\}$$

for  $\tilde{R}_\sigma^2 \leq 2$ . Alternatively, by Esseen [29, Eq. (2.8)] for  $p < 4$  and Cox and Kemperman [22, Thm. 2.6] for  $p \geq 4$ , we have the symmetrized estimate

$$\|\sum_{i \leq n} X_i^2 - 1\|_p \leq 2^{-1/p} \|\sum_{i \leq n} X_i^2 - (X_i')^2\|_p.$$

Since the random variables  $(X_i^2 - (X_i')^2)_{i \geq 1}$  are symmetric, with

$$\mathbb{E}[(X_i^2 - (X_i')^2)^2] \leq \frac{2}{n^2}(\tilde{R}_\sigma^2 - 1) \quad \text{and} \quad \mathbb{E}[(X_i^2 - (X_i')^2)^p] \leq \frac{2}{n^p}(\tilde{R}_\sigma^2 - 1)\tilde{R}_\sigma^{2(p-2)},$$

an improvement on the Marcinkiewicz-Zygmund inequality for symmetric random variables (Lem. C.3) implies

$$\|\sum_{i \leq n} X_i^2 - (X_i')^2\|_p \leq \begin{cases} \frac{\sqrt{2}2^{1/p}}{\sqrt{n}} \left( \frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}} \right)^{1/p} \tilde{R}_\sigma^{2(1-2/p)} (\tilde{R}_\sigma^2 - 1)^{1/p} & \text{if } p \geq 2 \\ \sqrt{2} \left( \frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}} \right)^{1/p} \frac{\tilde{R}_\sigma^2}{n} \|\text{Binomial}(n, \frac{2(\tilde{R}_\sigma^2 - 1)}{\tilde{R}_\sigma^4})\|_p & \text{if } p \geq 4. \end{cases}$$

Hence we finally obtain that

$$\begin{aligned} (21) \quad & \|\frac{n}{2}\mathbb{E}[Y^2 | \tilde{S}_n] - 1\|_p \\ & \leq D_{n,p} \triangleq \frac{1}{2\sqrt{n}} \min \left\{ \begin{array}{l} \sqrt{p-1} \max \left( (\tilde{R}_\sigma^2 - 1)^{1-1/p}, \left[ \frac{(\tilde{R}_\sigma^2 - 1)^p + \tilde{R}_\sigma^2 - 1}{\tilde{R}_\sigma^2} \right]^{1/p} \right) \\ \max \left( (\tilde{R}_\sigma^2 - 1)^{1-1/p}, \left[ \frac{(\tilde{R}_\sigma^2 - 1)^p + \tilde{R}_\sigma^2 - 1}{\tilde{R}_\sigma^2} \right]^{1/p} \right) A_{n,p}^* + \sqrt{\tilde{R}_\sigma^2 - 1} A_p \\ \sqrt{2} \left( \frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}} \right)^{1/p} \tilde{R}_\sigma^{2(1-2/p)} (\tilde{R}_\sigma^2 - 1)^{1/p} \\ \sqrt{2} 2^{-1/p} \left( \frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}} \right)^{1/p} \frac{\tilde{R}_\sigma^2}{\sqrt{n}} \sqrt{\|\text{Binomial}(n, \frac{2(\tilde{R}_\sigma^2 - 1)}{\tilde{R}_\sigma^4})\|_{p/2}} \quad \text{if } p \geq 4. \end{array} \right\} \end{aligned}$$

This implies that

$$\begin{aligned} & \int_{-\frac{1}{2} \log(1 - \frac{R^2}{n\kappa})}^{\infty} \left\| \frac{n}{2} \mathbb{E}[Y_t^2 \mid \tilde{S}_n] - 1 \right\|_p \frac{e^{-2t}}{\sqrt{1-e^{-2t}}} dt \\ & \leq D_{n,p} \int_{-\frac{1}{2} \log(1 - \frac{R^2}{n\kappa})}^{\infty} \frac{e^{-2t}}{\sqrt{1-e^{-2t}}} dt = D_{n,p} \int_0^{\sqrt{1 - \frac{R^2}{n\kappa}}} \frac{t}{\sqrt{1-t^2}} dt = M_{n,\kappa}^2 D_{n,p}. \end{aligned}$$

The definition (21) also implies the further upper bound

$$D_{n,p} \leq \frac{1}{2\sqrt{n}} (\max(\tilde{R}_\sigma^2 - 1, 1))^{1-1/p} \tilde{A}_{n,p}^* + \sqrt{\tilde{R}_\sigma^2 - 1} A_p.$$

Therefore since  $H_1(Z) = Z$  we obtain

$$(22) \quad (a_2) \leq (b_{2,1}^{\kappa,p,\tilde{R}}) \triangleq \|Z\|_p D_{n,p} M_{n,\kappa}^2.$$

Since  $\|Z\|_p \leq \sqrt{p-1}$  by Lem. C.1, we have the further upper bound

$$(23) \quad (b_{2,1}^{\kappa,p,\tilde{R}}) \leq (b_{2,2}^{\kappa,p,\tilde{R}}) \triangleq \frac{\sqrt{p-1}}{2\sqrt{n}} \left[ (\max(\tilde{R}_\sigma^2 - 1, 1))^{1-1/p} A_{n,p}^* + \sqrt{\tilde{R}_\sigma^2 - 1} A_p \right].$$

**Bounding (a<sub>3</sub>).** Finally, we turn to bounding (a<sub>3</sub>) of (20). Since

$$n \mathbb{E}[Y^k \mid X_1, \dots, X_n] = \sum_{i \leq n} \mathbb{E}[(X_i - X'_i)^k \mid X_1, \dots, X_n] \quad \text{for all } k \geq 1,$$

by Jensen's inequality, we have

$$(24) \quad n \left\| \mathbb{E}[Y^k \mid S_n] \right\|_p \leq \left\| \sum_{i \leq n} (X_i - X'_i)^k \right\|_p \quad \text{for all } k \geq 1.$$

We will derive different bounds for odd and even  $k$ , so we begin by writing

$$\begin{aligned} (a_3) & \leq \sum_{\substack{k \geq 3: \\ k \text{ is odd}}} \int_{-\frac{1}{2} \log(1 - \frac{R^2}{n\kappa})}^{\infty} \frac{e^{-kt} \|H_{k-1}\|_p}{k! (\sqrt{1-e^{-2t}})^{k-1}} n \left\| \mathbb{E}[Y^k \mid \tilde{S}_n] \right\|_p dt \\ & + \sum_{\substack{k \geq 4: \\ k \text{ is even}}} \int_{-\frac{1}{2} \log(1 - \frac{R^2}{n\kappa})}^{\infty} \frac{e^{-kt} \|H_{k-1}\|_p}{k! (\sqrt{1-e^{-2t}})^{k-1}} n \left\| \mathbb{E}[Y^k \mid \tilde{S}_n] \right\|_p dt \triangleq (a_{3,1}) + (a_{3,2}). \end{aligned}$$

Let  $k \geq 3$  be an *odd* integer. The random variables  $((X_i - X'_i)^k)_{i \geq 1}$  are symmetric and therefore have a mean of zero. Moreover, if we define  $W'_i \triangleq \sigma\sqrt{n}X'_i + \mathbb{E}[W_i]$ , then

$$\begin{aligned} \left\| (X_i - X'_i)^k \right\|_p &= \frac{1}{\sqrt{n^k} \sigma^k} \left\| |W_i - W'_i|^k \right\|_p \\ &\leq \frac{1}{\sqrt{n^k} \sigma^k} \left\| \max(W_i, W'_i)^{k-2/p} |W_i - W'_i|^{2/p} \right\|_p \\ &\leq \frac{1}{\sqrt{n^k} \sigma^{2/p}} \tilde{R}^{k-2/p} \left\| |W_i - W'_i|^{2/p} \right\|_p \leq \frac{2^{1/p} \tilde{R}^{k-2/p}}{\sqrt{n^k}}. \end{aligned}$$

Similarly, we also have

$$\left\| (X_i - X'_i)^k \right\|_2 \leq \frac{\sqrt{2} \tilde{R}^{k-1}}{\sqrt{n^k}}.$$

Therefore, Lems. C.2, C.3, and C.5 together imply  $\left\| \sum_{i \leq n} (X_i - X'_i)^k \right\|_p \leq \frac{\tilde{R}^{k-1}}{\sqrt{n^{k-1}}} C_{n,p}$ . Combining this with the inequality (24) we obtain

$$\begin{aligned} (a_{3,1}) &= \sum_{\substack{k \geq 3: \\ k \text{ is odd}}} \int_{-\frac{1}{2} \log(1 - \frac{R^2}{n\kappa})}^{\infty} \frac{e^{-tk} \|H_{k-1}\|_p}{k! \sqrt{1-e^{-2t}}^{k-1}} n \left\| \mathbb{E}[Y^k \mid S_n] \right\|_p dt \\ &\leq \sum_{\substack{k \geq 3: \\ k \text{ is odd}}} C_{n,p} \int_0^{\sqrt{1 - \frac{R^2}{n\kappa}}} \frac{x^{k-1} \tilde{R}^{k-1} \|H_{k-1}\|_p}{k! \sqrt{1-x^2}^{k-1} \sqrt{n^{k-1}}} dx \end{aligned}$$

$$\begin{aligned}
&\stackrel{(a)}{\leq} \sum_{k \geq 1} C_{n,p} \frac{\tilde{R}^{2k} \|H_{2k}\|_p}{n^k (2k+1)!} \int_0^{\sqrt{1-\frac{\tilde{R}^2}{n\kappa}}} \frac{x^{2k}}{(1-x^2)^k} dx \\
&\stackrel{(b)}{=} \frac{1}{2} \sum_{k \geq 1} C_{n,p} \frac{\tilde{R}^{2k} \|H_{2k}\|_p}{n(2k+1)!} \int_{\frac{\tilde{R}^2}{\kappa}}^{\frac{n}{\kappa}} \frac{1}{\sqrt{y}} \left(\frac{1}{y} - \frac{1}{n}\right)^{k-\frac{1}{2}} dy,
\end{aligned}$$

where (a) is obtained by noting that all odd numbers  $k$  can be written as  $2m+1$  for an  $m \in \mathbb{N}_+$  and (b) by the change of variables  $y = n(1-x^2)$ .

To further upper bound the right-hand side, we will invoke the Hermite polynomial moment bound (Lem. C.1)  $\|H_{k-1}\|_p \leq \sqrt{p-1}^{k-1} \sqrt{k-1!}$  and use two applications of Stirling's approximation [60] to conclude that, for all  $m \in \mathbb{N} \setminus \{0\}$ ,

$$\sqrt{(2m)!} \geq \sqrt{\sqrt{2\pi(2m)} \cdot (2m/e)^{2m} \cdot \exp\left(\frac{1}{12(2m)+1}\right)} \geq e^{-19/300} 2^m m! / (\pi m)^{1/4}.$$

These estimates imply that, for all  $K_p \in \mathbb{N}_+$ ,

$$\begin{aligned}
(a_{3,1}) &\leq \frac{C_{n,p}}{2} \left\{ \sum_{1 \leq k \leq K_p-1} \frac{\tilde{R}^{2k} \|H_{2k}\|_p}{n(2k+1)!} \int_{\frac{\tilde{R}^2}{\kappa}}^{\frac{n}{\kappa}} \frac{1}{\sqrt{y}} \left(\frac{1}{y} - \frac{1}{n}\right)^{k-\frac{1}{2}} dy \right. \\
&\quad \left. + e^{19/300} \pi^{1/4} \sum_{k \geq K_p} \frac{2^{-k} \tilde{R}^{2k} (p-1)^k k^{1/4}}{(2k+1) n k!} \int_{\frac{\tilde{R}^2}{\kappa}}^{\frac{n}{\kappa}} \frac{1}{\sqrt{y}} \left(\frac{1}{y} - \frac{1}{n}\right)^{k-\frac{1}{2}} dy \right\} \\
&\stackrel{(a)}{\leq} \frac{C_{n,p}}{2} \left\{ \sum_{1 \leq k \leq K_p-1} \frac{\tilde{R}^{2k} \|H_{2k}\|_p}{n(2k+1)!} \int_{\frac{\tilde{R}^2}{\kappa}}^{\frac{n}{\kappa}} \frac{1}{\sqrt{y}} \left(\frac{1}{y} - \frac{1}{n}\right)^{k-\frac{1}{2}} dy \right. \\
&\quad \left. + e^{19/300} \pi^{1/4} \frac{K_p^{1/4}}{(2K_p+1)} \sum_{k \geq K_p} \frac{2^{-k} \tilde{R}^{2k} (p-1)^k}{n k!} \int_{\frac{\tilde{R}^2}{\kappa}}^{\frac{n}{\kappa}} \frac{1}{\sqrt{y}} \left(\frac{1}{y} - \frac{1}{n}\right)^{k-\frac{1}{2}} dy \right\} \\
&\stackrel{(b)}{\leq} \frac{C_{n,p}}{2} \left\{ \sum_{1 \leq k \leq K_p-1} \frac{\tilde{R}^{2k}}{n} \left( \frac{\|H_{2k}\|_p}{(2k+1)!} - \frac{K_p^{1/4} 2^{-k} (p-1)^k e^{19/300} \pi^{1/4}}{(2K_p+1) k!} \right) \int_{\frac{\tilde{R}^2}{\kappa}}^{\frac{n}{\kappa}} \frac{1}{\sqrt{y}} \left(\frac{1}{y} - \frac{1}{n}\right)^{k-\frac{1}{2}} dy \right. \\
&\quad \left. + e^{19/300} \pi^{1/4} \frac{K_p^{1/4}}{(2K_p+1) n} \int_{\frac{\tilde{R}^2}{\kappa}}^{\frac{n}{\kappa}} \frac{1}{\sqrt{1-\frac{y}{n}}} \left[ e^{\frac{1}{2}(p-1)\tilde{R}^2 \left(\frac{1}{y} - \frac{1}{n}\right)} - 1 \right] dy \right\}
\end{aligned}$$

where (a) follows from the fact that  $x \rightarrow \frac{x^{1/4}}{(2x+1)}$  is decreasing and (b) from the fact that

$$e^{\frac{1}{2}(p-1)\tilde{R}^2 \left(\frac{1}{y} - \frac{1}{n}\right)} - 1 = \sum_{k=1}^{K_p-1} \frac{(\tilde{R}^2(p-1))^k 2^{-k}}{k!} \left(\frac{1}{y} - \frac{1}{n}\right)^k + \sum_{k=K_p}^{\infty} \frac{(\tilde{R}^2(p-1))^k 2^{-k}}{k!} \left(\frac{1}{y} - \frac{1}{n}\right)^k.$$

Since

$$\begin{aligned}
&\sum_{1 \leq k \leq K-1} \frac{\tilde{R}^{2k}}{n} \left( \frac{\|H_{2k}\|_p}{(2k+1)!} - \frac{K_p^{1/4} 2^{-k} (p-1)^k e^{19/300} \pi^{1/4}}{(2K+1) k!} \right) \int_{\frac{\tilde{R}^2}{\kappa}}^{\frac{n}{\kappa}} \frac{1}{\sqrt{y}} \left(\frac{1}{y} - \frac{1}{n}\right)^{k-\frac{1}{2}} dy \\
&\quad + e^{19/300} \pi^{1/4} \frac{K_p^{1/4}}{(2K+1) n} \int_{\frac{\tilde{R}^2}{\kappa}}^{\frac{n}{\kappa}} \frac{1}{\sqrt{1-\frac{y}{n}}} \left[ e^{\frac{1}{2}(p-1)\tilde{R}^2 \left(\frac{1}{y} - \frac{1}{n}\right)} - 1 \right] dy
\end{aligned}$$

is decreasing in  $K$ , its largest value is attained for  $K=1 \leq K_p$ . Moreover, for all  $x \leq \sqrt{1 - \frac{\tilde{R}^2}{n\kappa}}$ , we have  $\frac{1}{(1-x^2)^k} \leq \left(\frac{n\kappa}{\tilde{R}^2}\right)^{k-1} \frac{1}{1-x^2}$ . Hence a change of variables implies

$$\begin{aligned}
&\frac{C_{n,p}}{2} \left\{ \sum_{1 \leq k \leq K_p-1} \frac{\tilde{R}^{2k}}{n} \left( \frac{\|H_{2k}\|_p}{(2k+1)!} - \frac{K_p^{1/4} 2^{-k} (p-1)^k e^{19/300} \pi^{1/4}}{(2K_p+1) k!} \right) \int_{\frac{\tilde{R}^2}{\kappa}}^{\frac{n}{\kappa}} \frac{1}{\sqrt{y}} \left(\frac{1}{y} - \frac{1}{n}\right)^{k-\frac{1}{2}} dy \right. \\
&\quad \left. + e^{19/300} \pi^{1/4} \frac{K_p^{1/4}}{(2K_p+1) n} \int_{\frac{\tilde{R}^2}{\kappa}}^{\frac{n}{\kappa}} \frac{1}{\sqrt{1-\frac{y}{n}}} \left[ e^{\frac{1}{2}(p-1)\tilde{R}^2 \left(\frac{1}{y} - \frac{1}{n}\right)} - 1 \right] dy \right\} \\
&\leq C_{n,p} e^{19/300} \pi^{1/4} \frac{\tilde{R}^2}{3n\kappa} \sum_{k \geq 1} \int_0^{\sqrt{1-\frac{\tilde{R}^2}{n\kappa}}} \frac{x^{2k} (p-1)^k \kappa^k}{2^k k! (1-x^2)} dx \\
&\leq C_{n,p} \frac{\tilde{R}^2}{3n\kappa} e^{19/300} \pi^{1/4} \int_0^{\sqrt{1-\frac{\tilde{R}^2}{n\kappa}}} \frac{[e^{\frac{1}{2}(p-1)x^2 \kappa} - 1]}{(1-x^2)} dx
\end{aligned}$$

$$\left| \leq C_{n,p} \frac{\tilde{R}^2}{3n\kappa} e^{19/300} \pi^{1/4} [e^{\frac{1}{2}(p-1)(1-\frac{\tilde{R}^2}{n\kappa})\kappa} - 1] \log(\frac{1+M_{n,\kappa}}{1-M_{n,\kappa}}). \right.$$

Next suppose  $k \geq 4$  is even. Then  $((X_i - X'_i)^k)_{i \geq 1}$  are almost surely nonnegative. Moreover,

$$\mathbb{E}[(X_i - X'_i)^k] \leq 2 \frac{\tilde{R}^{k-2}}{\sqrt{n^k}} \quad \text{and} \quad \mathbb{E}[(X_i - X'_i)^{kp}] \leq 2 \frac{\tilde{R}^{kp-2}}{\sqrt{n^{kp}}}.$$

Therefore we can invoke a moment inequality for nonnegative random variables (Lem. C.4) to conclude that

$$n \|\mathbb{E}[Y^k | S_n]\|_p \leq \|\sum_{i \leq n} (X_i - X'_i)^k\|_p \leq \frac{\tilde{R}^k}{\sqrt{n^k}} \|\text{Binomial}(n, \frac{2}{\tilde{R}^2})\|_p.$$

Moreover by the triangle inequality and Lem. C.2 the following upper also holds

$$\begin{aligned} n \|\mathbb{E}[Y^k | \tilde{S}_n]\|_p &\leq \|\sum_{i \leq n} (X_i - X'_i)^k\|_p \\ &\leq n \mathbb{E}[(X_1 - X'_1)^k] + \|\sum_{i \leq n} (X_i - X'_i)^k - \mathbb{E}[(X_1 - X'_1)^k]\|_p \\ &\leq \frac{\tilde{R}^{k-2}}{\sqrt{n^{k-2}}} (1 + \frac{1}{\sqrt{n}} \tilde{U}_{n,p} \tilde{R}). \end{aligned}$$

Hence we obtain

$$n \|\mathbb{E}[Y^k | S_n]\|_p \leq \|\sum_{i \leq n} (X_i - X'_i)^k\|_p \leq \frac{B_{p,n} \tilde{R}^{k-2}}{\sqrt{n^{k-2}}}$$

for  $B_{p,n} \triangleq \min(\frac{\tilde{R}^2}{n} \|\text{Binomial}(n, \frac{2}{\tilde{R}^2})\|_p, 1 + \frac{1}{\sqrt{n}} \tilde{U}_{n,p} \tilde{R})$ . This gives us the upper estimate

$$\int_{-\frac{1}{2}}^{\infty} \log(1 - \frac{\tilde{R}^2}{n\kappa}) \frac{e^{-tk}}{\sqrt{1 - e^{-2t^{k-1}}}} n \|\mathbb{E}[Y^k | S_n]\|_p dt \leq B_{p,n} \int_{-\frac{1}{2}}^{\infty} \log(1 - \frac{\tilde{R}^2}{n\kappa}) \frac{e^{-tk} \tilde{R}^{k-2}}{\sqrt{1 - e^{-2t^{k-1}}} \sqrt{n^{k-2}}} dt.$$

To bound  $(a_{3,2})$ , it remains to bound

$$\begin{aligned} \sum_{\substack{k \geq 4: \\ k \text{ is even}}} \frac{\|H_{k-1}\|_p}{k!} \int_{-\frac{1}{2}}^{\infty} \log(1 - \frac{\tilde{R}^2}{n\kappa}) \frac{e^{-tk} \tilde{R}^{k-2}}{\sqrt{1 - e^{-2t^{k-1}}} \sqrt{n^{k-2}}} dt \\ \stackrel{(a)}{\leq} \sum_{\substack{k \geq 3: \\ k \text{ is odd}}} \frac{\tilde{R}^{k-1} \|H_k\|_p}{(k+1)! \sqrt{n^{k-1}}} \int_0^{\sqrt{1 - \frac{\tilde{R}^2}{n\kappa}}} \frac{x^k}{\sqrt{1 - x^{2k}}} dx \\ \stackrel{(b)}{\leq} \sum_{k \geq 1} \frac{\tilde{R}^{2k} \|H_{2k+1}\|_p}{(2k+2)!} \int_0^{\sqrt{1 - \frac{\tilde{R}^2}{n\kappa}}} \frac{x^{2k+1}}{\sqrt{1 - x^{2k+1}} n^k} dx \\ \stackrel{(c)}{\leq} \frac{1}{2} \sum_{k \geq 1} \frac{\tilde{R}^{2k} \|H_{2k+1}\|_p}{(2k+2)! \sqrt{n}} \int_{\frac{\tilde{R}^2}{\kappa}}^n \frac{1}{\sqrt{y}} \left(\frac{1}{y} - \frac{1}{n}\right)^k dy \end{aligned}$$

where (a) and (c) are obtained by a change of variable, and (b) is a consequence of the fact that every odd number can be written as  $2m+1$  for an  $m \geq 1$ . To upper bound this quantity we will again employ a Hermite polynomial moment bound (Lem. C.1),  $\|H_{k-1}\|_p \leq \sqrt{p-1}^{k-1} \sqrt{k-1!}$ , and use Stirling's approximation to deduce that, for all  $m \in \mathbb{N} \setminus \{0\}$ ,

$$\sqrt{(2m+1)!} = \sqrt{2m+1} \sqrt{2m!} \geq \sqrt{2m+1} e^{-19/300} 2^m m! / (\pi m)^{1/4}.$$

Hence for any  $K_p \in \mathbb{N}_+$  we obtain that

$$\begin{aligned} (a_{3,2}) &= \sum_{\substack{k \geq 4: \\ k \text{ is even}}} \frac{\|H_{k-1}\|_p}{k!} \int_{-\frac{1}{2}}^{\infty} \log(1 - \frac{\tilde{R}^2}{n\kappa}) \frac{e^{-tk}}{\sqrt{1 - e^{-2t^{k-1}}}} n \|\mathbb{E}((Y_t)^k | S_n)\|_p dt \\ &\leq \frac{B_{p,n}}{2} \sum_{K_p > k \geq 1} \frac{\tilde{R}^{2k} \|H_{2k+1}\|_p}{(2k+2)! \sqrt{n}} \int_{\frac{\tilde{R}^2}{\kappa}}^n \frac{1}{\sqrt{y}} \left(\frac{1}{y} - \frac{1}{n}\right)^k dy \\ &\quad + \frac{B_{p,n}}{4} \sum_{K_p \leq k} \frac{\tilde{R}^{2k} (p-1)^{k+1/2}}{(k+1) \sqrt{2k+1!} \sqrt{n}} \int_{\frac{\tilde{R}^2}{\kappa}}^n \frac{1}{\sqrt{y}} \left(\frac{1}{y} - \frac{1}{n}\right)^k dy \end{aligned}$$

$$\begin{aligned}
&\leq \frac{B_{p,n}}{2} \sum_{K_p > k \geq 1} \frac{\tilde{R}^{2k}}{\sqrt{n}} \left( \frac{\|H_{2k+1}\|_p}{(2k+2)!} - \frac{2^{-k} e^{19/300} \pi^{1/4} K_p^{1/4} \sqrt{p-1}^{2k+1}}{2(K_p+1) \sqrt{2K_p+1} k!} \right) \int_{\frac{\tilde{R}^2}{\kappa}}^n \frac{1}{\sqrt{y}} \left( \frac{1}{y} - \frac{1}{n} \right)^k dy \\
&\quad + \frac{B_{p,n}}{4} e^{19/300} \pi^{1/4} \frac{K_p^{1/4} \sqrt{p-1}}{(K_p+1) \sqrt{2K_p+1}} \frac{1}{\sqrt{n}} \int_{\frac{\tilde{R}^2}{\kappa}}^n \frac{1}{\sqrt{y}} \left( e^{\frac{1}{2}(p-1)\tilde{R}^2(\frac{1}{y}-\frac{1}{n})} - 1 \right) dy.
\end{aligned}$$

Since

$$\begin{aligned}
&\sum_{K > k \geq 1} \frac{\tilde{R}^{2k}}{\sqrt{n}} \left( \frac{\|H_{2k+1}\|_p}{(2k+2)!} - \frac{2^{-k} e^{19/300} \pi^{1/4} K_p^{1/4} \sqrt{p-1}^{2k+1}}{2(K_p+1) \sqrt{2K_p+1} k!} \right) \int_{\frac{\tilde{R}^2}{\kappa}}^n \frac{1}{\sqrt{y}} \left( \frac{1}{y} - \frac{1}{n} \right)^k dy \\
&\quad + e^{19/300} \pi^{1/4} \frac{K_p^{1/4} \sqrt{p-1}}{2(K_p+1) \sqrt{2K_p+1}} \frac{1}{\sqrt{n}} \int_{\frac{\tilde{R}^2}{\kappa}}^n \frac{1}{\sqrt{y}} \left( e^{\frac{1}{2}(p-1)\tilde{R}^2(\frac{1}{y}-\frac{1}{n})} - 1 \right) dy
\end{aligned}$$

is decreasing in  $K$ , its largest value is attained for  $K = 1 \leq K_p$ . Moreover for all  $x \leq \sqrt{1 - \frac{\tilde{R}^2}{n\kappa}}$  we have  $\frac{1}{\sqrt{1-x^2}^{2k+1}} \leq \left(\frac{n\kappa}{\tilde{R}^2}\right)^{k-1} \frac{1}{\sqrt{1-x^2}^3}$ . Hence a change of variables implies

$$\begin{aligned}
&\frac{B_{p,n}}{2} \sum_{K_p > k \geq 1} \frac{\tilde{R}^{2k}}{\sqrt{n}} \left( \frac{\|H_{2k+1}\|_p}{(2k+2)!} - \frac{2^{-k} e^{19/300} \pi^{1/4} K_p^{1/4} \sqrt{p-1}^{2k+1}}{2(K_p+1) \sqrt{2K_p+1} k!} \right) \int_{\frac{\tilde{R}^2}{\kappa}}^n \frac{1}{\sqrt{y}} \left( \frac{1}{y} - \frac{1}{n} \right)^k dy \\
&\quad + \frac{B_{p,n}}{4} e^{19/300} \pi^{1/4} \frac{K_p^{1/4} \sqrt{p-1}}{(K_p+1) \sqrt{2K_p+1}} \frac{1}{\sqrt{n}} \int_{\frac{\tilde{R}^2}{\kappa}}^n \frac{1}{\sqrt{y}} \left( e^{\frac{1}{2}(p-1)\tilde{R}^2(\frac{1}{y}-\frac{1}{n})} - 1 \right) dy \\
&\leq B_{p,n} \sum_{k \geq 1} \frac{\|H_{2k+1}\|_p}{(2k+2)!} \int_0^{\sqrt{1-\frac{\tilde{R}^2}{n\kappa}}} \frac{x^{2k+1} \tilde{R}^2 \sqrt{\kappa}^{2k-2}}{\sqrt{1-x^2}^3 n} dx \\
&\leq \frac{\tilde{R}^2 \left(1 + \frac{\tilde{R} \tilde{U}_{n,p}}{\sqrt{n}}\right) \pi^{1/4} e^{19/300} \sqrt{p-1}}{4\sqrt{3}} \left[ e^{\frac{1}{2}(p-1)\kappa \left(1 - \frac{\tilde{R}^2}{n\kappa}\right)} - 1 \right] \int_0^{\sqrt{1-\frac{\tilde{R}^2}{n\kappa}}} \frac{x}{\sqrt{1-x^2}^3} dx \\
&\leq \frac{\tilde{R} \left(1 + \frac{\tilde{R} \tilde{U}_{n,p}}{\sqrt{n}}\right) \pi^{1/4} e^{19/300} \sqrt{p-1}}{4\sqrt{3}} \left[ e^{\frac{1}{2}(p-1)\kappa \left(1 - \frac{\tilde{R}^2}{n\kappa}\right)} - 1 \right].
\end{aligned}$$

Therefore, to conclude, we obtain

$$\begin{aligned}
(a_3) &\leq (b_{3,1}^{\kappa,p,\tilde{R}}) \\
&\triangleq \frac{B_{p,n}}{2} \left\{ \sum_{K_p > k \geq 1} \frac{\tilde{R}^{2k}}{\sqrt{n}} \left( \frac{\|H_{2k+1}\|_p}{(2k+2)!} - \frac{2^{-k} e^{19/300} \pi^{1/4} K_p^{1/4} \sqrt{p-1}^{2k+1}}{2(K_p+1) \sqrt{2K_p+1} k!} \right) \int_{\frac{\tilde{R}^2}{\kappa}}^n \frac{1}{\sqrt{y}} \left( \frac{1}{y} - \frac{1}{n} \right)^k dy \right. \\
(25) \quad &\quad \left. + \frac{B_{p,n}}{4} e^{19/300} \pi^{1/4} \frac{K_p^{1/4} \sqrt{p-1}}{(K_p+1) \sqrt{2K_p+1}} \frac{1}{\sqrt{n}} \int_{\frac{\tilde{R}^2}{\kappa}}^n \frac{1}{\sqrt{y}} \left( e^{\frac{1}{2}(p-1)\tilde{R}^2(\frac{1}{y}-\frac{1}{n})} - 1 \right) dy \right\} \\
&\quad + \frac{C_{n,p}}{2} \left\{ \sum_{1 \leq k \leq K_p-1} \frac{\tilde{R}^{2k}}{n} \left( \frac{\|H_{2k}\|_p}{(2k+1)!} - \frac{K_p^{1/4} 2^{-k} (p-1)^k e^{19/300} \pi^{1/4}}{(2K_p+1) k!} \right) \int_{\frac{\tilde{R}^2}{\kappa}}^n \frac{1}{\sqrt{y}} \left( \frac{1}{y} - \frac{1}{n} \right)^{k-\frac{1}{2}} dy \right. \\
&\quad \left. + e^{19/300} \pi^{1/4} \frac{K_p^{1/4}}{(2K_p+1)n} \int_{\frac{1}{n}}^{\frac{\tilde{R}^2}{\kappa}} \frac{1}{y^{3/2} \sqrt{y-\frac{1}{n}}} \left[ e^{\frac{1}{2}(p-1)\tilde{R}^2(y-\frac{1}{n})} - 1 \right] dy \right\}.
\end{aligned}$$

Moreover,  $(b_{3,1}^{\kappa,p,\tilde{R}})$  can be further upper-bounded by

$$\begin{aligned}
(26) \quad (b_{3,1}^{\kappa,p,\tilde{R}}) &\leq (b_{3,2}^{\kappa,p,\tilde{R}}) \triangleq \frac{\tilde{R} \left(1 + \frac{\tilde{R} \tilde{U}_{n,p}}{\sqrt{n}}\right) \pi^{1/4} e^{19/300} \sqrt{p-1}}{4\sqrt{3}} \left[ e^{\frac{1}{2}(p-1)\kappa \left(1 - \frac{\tilde{R}^2}{n\kappa}\right)} - 1 \right] \\
&\quad + C_{n,p} \frac{\tilde{R}^2}{3n\kappa} e^{19/300} \pi^{1/4} [e^{\frac{1}{2}(p-1)(1-\frac{\tilde{R}^2}{n\kappa})\kappa} - 1] \log(\frac{1+M_{n,\kappa}}{1-M_{n,\kappa}}).
\end{aligned}$$

## APPENDIX C: ADDITIONAL LEMMAS

LEMMA C.1 (Hermite polynomial moment bound [16, Lem. 3]). *Let  $h_k(x) \triangleq e^{x^2/2} \frac{\partial^k}{\partial x^k} e^{-x^2/2}$  and  $H_k \triangleq h_k(Z)$  for  $Z \sim \mathcal{N}(0, 1)$ . Then the following holds for all  $k, p \in \mathbb{N}$ :*

$$\|H_k\|_p \leq \sqrt{k!} \sqrt{p-1}^k.$$

LEMMA C.2 (Marcinkiewicz-Zygmund inequality [59, Thm. 2.1]). *Suppose  $(\tilde{X}_i)_{i \geq 1}$  are centered i.i.d. observations admitting a finite  $p$ -th absolute moment for some  $p \geq 2$ . Then*

$$\left\| \frac{1}{\sqrt{n}} \sum_{i \leq n} \tilde{X}_i \right\|_p \leq \sqrt{p-1} \|\tilde{X}_1\|_p.$$

LEMMA C.3 (Improved Marcinkiewicz-Zygmund inequality for symmetric random variables). *Suppose  $(\tilde{X}_i)_{i \geq 1}$  are symmetric centered i.i.d. observations admitting a finite  $p$ -th absolute moment for some  $p \geq 2$ . Then*

$$\left\| \frac{1}{\sqrt{n}} \sum_{i \leq n} \tilde{X}_i \right\|_p \leq \sqrt{2} \left( \frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}} \right)^{1/p} \sqrt{n} \|\tilde{X}_1\|_p.$$

If  $p \geq 4$ ,  $\mathbb{E}[\tilde{X}_i^2] \leq \tilde{\sigma}^2$ , and  $\mathbb{E}[|\tilde{X}_i|^p] \leq b_p$ , we also have

$$\left\| \sum_{i \leq n} \tilde{X}_i \right\|_p \leq \sqrt{2} \left( \frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}} \right)^{1/p} \left( \frac{b_p}{\tilde{\sigma}^2} \right)^{1/(p-2)} \sqrt{\|\text{Binomial}(n, (\frac{\tilde{\sigma}^p}{b_p})^{2/(p-2)})\|_{p/2}}.$$

PROOF. As the random variables  $(\tilde{X}_i)_{i \geq 1}$  are symmetric, we know that

$$\sum_{i=1}^n \tilde{X}_i \stackrel{d}{=} \sum_{i=1}^n \epsilon_i |\tilde{X}_i|,$$

where  $(\epsilon_i)_{i \geq 1} \stackrel{\text{i.i.d.}}{\sim} \text{Uniform}(\{-1, 1\})$ . According to Haagerup [35, Thm. B] we have

$$\mathbb{E}[\left\| \sum_{i=1}^n \epsilon_i |\tilde{X}_i| \right\|^p \mid (\tilde{X}_i)_{i=1}^n] \leq 2^{p/2} \frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}} \left( \sum_{i=1}^n \tilde{X}_i^2 \right)^{p/2}.$$

The tower property and Jensen's inequality therefore imply that

$$(27) \quad \left\| \sum_{i=1}^n \tilde{X}_i \right\|_p \leq \sqrt{2} \left( \frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}} \right)^{1/p} (\mathbb{E}[\left( \sum_{i \leq n} \tilde{X}_i^2 \right)^{p/2}])^{1/p} \leq \sqrt{2} \left( \frac{\Gamma(\frac{p+1}{2})}{\sqrt{\pi}} \right)^{1/p} \sqrt{n} \|\tilde{X}_1\|_p.$$

Now suppose  $p \geq 4$ ,  $\mathbb{E}[\tilde{X}_i^2] \leq \tilde{\sigma}^2$ , and  $\mathbb{E}[\tilde{X}_i^p] \leq b_p$ . Then, by a moment inequality for non-negative random variables (Lem. C.4), we obtain

$$\left\| \sum_{i=1}^n \tilde{X}_i \right\|_{p/2} \leq \left( \frac{b_p}{\tilde{\sigma}^2} \right)^{2/(p-2)} \|\text{Binomial}(n, (\frac{\tilde{\sigma}^p}{b_p})^{2/(p-2)})\|_{p/2}.$$

Combining this with the inequality (27) yields the advertised conclusion.  $\square$

LEMMA C.4 (Moment inequality for nonnegative random variables). *Let  $(\tilde{X}_i)_{i \geq 1}$  be a sequence of i.i.d. random variables that are almost surely nonnegative. If, for some  $p \geq 2$ ,  $\mathbb{E}(\tilde{X}_1) \leq a$  and  $\mathbb{E}(\tilde{X}_i^p) \leq b$  for  $a, b > 0$ , then*

$$\left\| \sum_{i=1}^n \tilde{X}_i \right\|_p \leq \left( \frac{b}{a} \right)^{1/(p-1)} \|\text{Binomial}(n, (\frac{a^p}{b})^{\frac{1}{p-1}})\|_p$$

PROOF. Let  $(V_i)_{i \geq 1}$  be an i.i.d. sequence of Bernoulli random variables with

$$\mathbb{P}(V_i = 0) = 1 - \left( \frac{a^p}{b} \right)^{\frac{1}{p-1}} \quad \text{and} \quad \mathbb{P}(V_i = 1) = \left( \frac{a^p}{b} \right)^{\frac{1}{p-1}}.$$

Then, by Ibragimov and Sharakhmetov [39, Thm. 2],

$$\mathbb{E}[\left( \sum_{i=1}^n \tilde{X}_i \right)^p] \leq \mathbb{E}[\left( \sum_{i=1}^n \left( \frac{b}{a} \right)^{1/(p-1)} V_i \right)^p] \leq \left( \frac{b}{a} \right)^{p/(p-1)} \mathbb{E}[\left( \sum_{i=1}^n V_i \right)^p].$$

$\square$

LEMMA C.5 (Rosenthal's inequality with explicit constants). *Let  $(\tilde{X}_i)_{i \geq 1}$  be a sequence of centered i.i.d. observations. If  $\|\tilde{X}_1\|_p < \infty$  for some  $p \geq 2$ , then*

$$\left\| \frac{1}{\sqrt{n}} \sum_{i \leq n} \tilde{X}_i \right\|_p \leq \left( \frac{p}{2} + 1 \right) n^{1/p-1/2} \|\tilde{X}_1\|_p + 2^{1/p} \sqrt{p/2+1} e^{\frac{1}{2} + \frac{1}{p}} \|\tilde{X}_1\|_2.$$

PROOF. According to Nagaev and Pinelis [52, Thm. 2] we have

$$\left\| \frac{1}{\sqrt{n}} \sum_{i \leq n} \tilde{X}_i \right\|_p^p \leq \inf_{c > \frac{p}{2}} c^p n^{1-\frac{p}{2}} \|\tilde{X}_1\|_p^p + p c^{p/2} e^c B\left(\frac{p}{2}, c - \frac{p}{2}\right) \|\tilde{X}_1\|_2^p,$$

where  $B(\cdot, \cdot)$  is the Beta function. The choice  $c = \frac{p}{2} + 1$  yields

$$\begin{aligned} \left\| \frac{1}{\sqrt{n}} \sum_{i \leq n} \tilde{X}_i \right\|_p^p &\leq \left( \frac{p}{2} + 1 \right)^p n^{1-\frac{p}{2}} \|\tilde{X}_1\|_p^p + p \left( \frac{p}{2} + 1 \right)^{p/2} e^{\frac{p}{2}+1} B\left(\frac{p}{2}, 1\right) \|\tilde{X}_1\|_2^p \\ &= \left( \frac{p}{2} + 1 \right)^p n^{1-\frac{p}{2}} \|\tilde{X}_1\|_p^p + 2 \left( \frac{p}{2} + 1 \right)^{\frac{p}{2}} e^{\frac{p}{2}+1} \|\tilde{X}_1\|_2^p. \end{aligned}$$

The subadditivity of the  $p$ -th root now implies the result.  $\square$

LEMMA C.6 (Kolmogorov-Smirnov quantile bound). *Let  $(X_i)_{i \geq 1}$  be a sequence of i.i.d. random variables taking values in  $[0, 1]$  and, for a confidence level  $\alpha > 0$ , let  $q_{n,1}^{\text{KS}}(\alpha)$  and  $q_{n,2}^{\text{KS}}(\alpha)$  respectively be the  $1 - \alpha$  quantiles of the one-sided and two-sided Kolmogorov-Smirnov (KS) distribution with parameter  $n$ . Then,*

$$\mathbb{P}\left(\sup_{k \in \mathbb{N}} \frac{1}{n} \sum_{i \leq n} X_i^k - \mathbb{E}[X_1^k] \geq q_{n,1}^{\text{KS}}(\alpha)\right) \vee \mathbb{P}\left(\sup_{k \in \mathbb{N}} \left| \frac{1}{n} \sum_{i \leq n} X_i^k - \mathbb{E}[X_1^k] \right| \geq q_{n,2}^{\text{KS}}(\alpha)\right) \leq \alpha.$$

PROOF. We begin by proving the two-sided statement. Let  $\mu$  be the distribution of  $X_1$  and  $\hat{\mu}$  the empirical distribution of  $(X_i)_{i=1}^n$ . Romano and Wolf [61, Prop. 3.1] showed that

$$\sup_{k \in \mathbb{N}} \left| \frac{1}{n} \sum_{i \leq n} X_i^k - \mathbb{E}[X_1^k] \right| \leq d_K(\hat{\mu}, \mu) \triangleq \sup_{x \in [0,1]} |\mathbb{P}(X_1 \leq x) - \mathbb{P}_{X \sim \hat{\mu}}(X \leq x)|$$

for  $d_K(\hat{\mu}, \mu)$  the two-sided Kolmogorov distance between  $\hat{\mu}$  and  $\mu$ . Now let  $\text{KS}(n)$  denote the two-sided Kolmogorov-Smirnov distribution with parameter  $n$ . If  $X_1$  is continuous, then  $d_K(\hat{\mu}, \mu) \sim \text{KS}(n)$  [32, Thm. 1] and hence

$$(28) \quad \mathbb{P}(\sup_{k \in \mathbb{N}} \left| \frac{1}{n} \sum_{i \leq n} X_i^k - \mathbb{E}[X_1^k] \right| \geq q_n^{\text{KS}}(\alpha)) \leq \mathbb{P}(d_K(\hat{\mu}, \mu) \geq q_n^{\text{KS}}(\alpha)) \leq \alpha.$$

If  $X_1$  is not continuous, fix any  $\epsilon > 0$ , and define

$$X_i^\epsilon \triangleq \frac{1}{1+\epsilon} (X_i + \epsilon U_i)$$

for  $(U_i)_{i \geq 1}$  an i.i.d. sequence of uniform random variables in  $[0, 1]$ . Since the  $(X_i^\epsilon)_{i \geq 1}$  are continuous i.i.d. random variables on  $[0, 1]$ , we have

$$\mathbb{P}(\sup_{k \in \mathbb{N}} \left| \frac{1}{n} \sum_{i \leq n} (X_i^\epsilon)^k - \mathbb{E}((X_1^\epsilon)^k) \right| \geq q_n^{\text{KS}}(\alpha)) \leq \alpha.$$

As this holds for any arbitrary choice of  $\epsilon > 0$ , the result (28) holds.

A nearly identical proof establishes the one-sided result, since, using integration by parts,

$$\begin{aligned} \frac{1}{n} \sum_{i \leq n} X_i^k - \mathbb{E}[X_1^k] &= \int_0^1 x^k d(\hat{\mu} - \mu)(x) = \int_0^1 kx^{k-1} (\mathbb{P}_{X \sim \hat{\mu}}(X > x) - \mathbb{P}(X_1 > x)) dx \\ &\leq d_{K,1}(\mu, \hat{\mu}) \int_0^1 kx^{k-1} dx = d_{K,1}(\mu, \hat{\mu}) \end{aligned}$$

for  $d_{K,1}(\mu, \hat{\mu}) \triangleq \sup_{x \in [0,1]} |\mathbb{P}(X_1 \leq x) - \mathbb{P}_{X \sim \hat{\mu}}(X \leq x)|$ , the one-sided Kolmogorov distance.  $\square$

## REFERENCES

- [1] ABRAMOWITZ, M. and STEGUN, I. A. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables* **55**. US Government printing office, Washington, D.C.
- [2] ANDERSON, T. W. (1969). Confidence limits for the expected value of an arbitrary bounded random variable with a continuous distribution function Technical Report, STANFORD UNIV CA DEPT OF STATISTICS.
- [3] AUDIBERT, J.-Y., MUNOS, R. and SZEPESVÁRI, C. (2009). Exploration-exploitation tradeoff using variance estimates in multi-armed bandits. *Theoretical Computer Science* **410** 1876–1902.
- [4] BAGNOLI, M. and BERGSTROM, T. (2005). Log-concave probability and its applications. *Economic theory* **26** 445–469.
- [5] BARTLETT, P. L. and MENDELSON, S. (2002). Rademacher and Gaussian complexities: Risk bounds and structural results. *Journal of Machine Learning Research* **3** 463–482.
- [6] BATES, S., ANGELOPOULOS, A., LEI, L., MALIK, J. and JORDAN, M. (2021). Distribution-free, risk-controlling prediction sets. *Journal of the ACM (JACM)* **68** 1–34.
- [7] BATIR, N. (2008). Inequalities for the gamma function. *Archiv der Mathematik* **91** 554–563.
- [8] BENTKUS, V. (2001). An Inequality for Large Deviation Probabilities of Sums of Bounded i.i.d. Random Variables. *Lithuanian Mathematical Journal* **41** 112–119.
- [9] BENTKUS, V. (2002). An inequality for tail probabilities of martingales with bounded differences. *Lithuanian Mathematical Journal* **42** 255–261.
- [10] BENTKUS, V. (2004). On Hoeffding's inequalities. *Ann. Probab.* **32** 1650–1673.
- [11] BENTKUS, V., KALOSHA, N. and VAN ZUIJLEN, M. (2006). On domination of tail probabilities of (super) martingales: explicit bounds. *Lithuanian Mathematical Journal* **46** 1–43.
- [12] BERNSTEIN, S. (1924). On a modification of Chebyshev's inequality and of the error formula of Laplace. *Ann. Sci. Inst. Sav. Ukraine, Sect. Math* **1** 38–49.
- [13] BHATTACHARJEE, C. and GOLDSTEIN, L. (2016). On strong embeddings by Stein's method. *Electronic Journal of Probability* **21** 1–30.
- [14] BIKYALIS, A. (1966). Estimates of the remainder term in the central limit theorem, Litovsk. *Mat. Sb* **6** 323–346.
- [15] BIRNBAUM, Z. W. (1942). An Inequality for Mill's Ratio. *The Annals of Mathematical Statistics* **13** 245 – 246. <https://doi.org/10.1214/aoms/1177731611>
- [16] BONIS, T. (2020). Stein's method for normal approximation in Wasserstein distances with application to the multivariate central limit theorem. *Probability Theory and Related Fields* **178** 827–860.
- [17] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, Oxford. <https://doi.org/10.1093/acprof:oso/9780199535255.001.0001>
- [18] BRETAGNOLLE, J. and MASSART, P. (1989). Hungarian constructions from the nonasymptotic viewpoint. *The Annals of Probability* 239–256.
- [19] CHATTERJEE, S. (2012). A new approach to strong embeddings. *Probability Theory and Related Fields* **152** 231–264.
- [20] CHEN, L. H., FANG, X. and SHAO, Q.-M. (2013). From Stein identities to moderate deviations.
- [21] CHEN, L. H., GOLDSTEIN, L. and SHAO, Q.-M. (2011). *Normal approximation by Stein's method* **2**. Springer, Berlin.
- [22] COX, D. C. and KEMPERMAN, J. H. (1983). Sharp bounds on the absolute moments of a sum of two iid random variables. *The Annals of Probability* 765–771.
- [23] CRAMÉR, H. (1938). Sur un nouveau théorème-limite de la théorie des probabilités. *Actualités Scientifiques et Industrielles* **736** 5–23.
- [24] DURRETT, R. (2019). *Probability: theory and examples* **49**. Cambridge University Press, Cambridge.
- [25] EATON, M. L. (1970). A Note on Symmetric Bernoulli Random Variables. *The Annals of Mathematical Statistics* **41** 1223 – 1226. <https://doi.org/10.1214/aoms/1177696897>
- [26] EATON, M. L. (1974). A probability inequality for linear combinations of bounded random variables. *The Annals of Statistics* 609–614.
- [27] EFRON, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *7* 1 – 26. Institute of Mathematical Statistics. <https://doi.org/10.1214/aos/1176344552>
- [28] ESSEEN, C.-G. (1942). On the Liapunov limit error in the theory of probability. *Ark. Mat. Astr. Fys.* **28** 1–19.
- [29] ESSEEN, C.-G. (1975). Bounds for the absolute third moment. *Scandinavian Journal of Statistics* 149–152.
- [30] FANG, X. and KOIKE, Y. (2023). From p-Wasserstein bounds to moderate deviations. *Electronic Journal of Probability* **28** 1–52.

- [31] FELLER, W. (1943). Generalization of a probability limit theorem of Cramér. *Transactions of the American Mathematical Society* **54** 361–372.
- [32] FELLER, W. (1948). On the Kolmogorov-Smirnov Limit Theorems for Empirical Distributions. *The Annals of Mathematical Statistics* **19** 177 – 189. <https://doi.org/10.1214/aoms/1177730243>
- [33] GASKO, M. (1991). A new distribution-free lower confidence bound for the mean of a positive random variable Technical Report, Working Paper QS 92-006, San Jose State Univ. College of Business.
- [34] GOLDSTEIN, L. and REINERT, G. (1997). Stein's method and the zero bias transformation with application to simple random sampling. *The Annals of Applied Probability* **7** 935–952.
- [35] HAAGERUP, U. (1981). The best constants in the Khintchine inequality. *Studia Mathematica* **70** 231–283.
- [36] HALL, P. (2013). *The bootstrap and Edgeworth expansion*. Springer Science & Business Media, New York.
- [37] HARDY, G. H. (1956). *Divergent Series*. Clarendon Press, Oxford.
- [38] HOEFFDING, W. (1963). Probability Inequalities for Sums of Bounded Random Variables. *Journal of the American Statistical Association* **58** 13–30.
- [39] IBRAGIMOV, R. and SHARAKHMETOV, S. (2001). The best constant in the Rosenthal inequality for non-negative random variables. *Statistics & probability letters* **55** 367–376.
- [40] JAIN, A., HICKERNELL, F. J., OWEN, A. B. and SOROKIN, A. G. (2025). Empirical Bernstein and betting confidence intervals for randomized quasi-Monte Carlo. *arXiv preprint arXiv:2504.18677*.
- [41] JING, B.-Y., SHAO, Q.-M. and WANG, Q. (2003). Self-normalized Cramér-type large deviations for independent random variables. *The Annals of Probability* **31** 2167 – 2215. <https://doi.org/10.1214/aop/1068646382>
- [42] JUN, K.-S. and ORABONA, F. (2019). Parameter-free online convex optimization with sub-exponential noise. In *Conference on Learning Theory* 1802–1823. PMLR.
- [43] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's, and the sample DF. I. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **32** 111–131.
- [44] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **34** 33–58.
- [45] KUCHIBHOTLA, A. K. and ZHENG, Q. (2021). Near-Optimal Confidence Sequences for Bounded Random Variables. In *Proceedings of the 38th International Conference on Machine Learning* (M. MEILA and T. ZHANG, eds.). *Proceedings of Machine Learning Research* **139** 5827–5837. PMLR.
- [46] LEE, C.-I. C. (1992). On Laplace continued fraction for the normal integral. *Annals of the Institute of Statistical Mathematics* **44** 107–120.
- [47] MARTINEZ-TABOADA, D. and RAMDAS, A. (2025). Sharp empirical Bernstein bounds for the variance of bounded random variables. *arXiv preprint arXiv:2505.01987*.
- [48] MASON, D. M. and ZHOU, H. H. (2012). Quantile coupling inequalities and their applications. *Probability Surveys* **9** 439–479.
- [49] MAURER, A. and PONTIL, M. (2009). Empirical Bernstein Bounds and Sample-Variance Penalization. In *COLT*.
- [50] MNIH, V., SZEPESVÁRI, C. and AUDIBERT, J.-Y. (2008). Empirical Bernstein stopping. In *Proceedings of the 25th international conference on Machine learning* 672–679.
- [51] NAGAEV, S. V. (1965). Some limit theorems for large deviations. *Theory of Probability & Its Applications* **10** 214–235.
- [52] NAGAEV, S. and PINELIS, I. (1978). Some inequalities for the distribution of sums of independent random variables. *Theory of Probability & Its Applications* **22** 248–256.
- [53] ORABONA, F. and JUN, K.-S. (2023). Tight concentrations and confidence sequences from the regret of universal portfolio. *IEEE Transactions on Information Theory*.
- [54] PETROV, V. V. (1975). *Sums of Independent Random Variables*. De Gruyter, Berlin, Boston. <https://doi.org/doi:10.1515/9783112573006>
- [55] PINELIS, I. (1994). Extremal Probabilistic Problems and Hotelling's T2 Test Under a Symmetry Condition. *The Annals of Statistics* **22** 357–368.
- [56] PINELIS, I. (2006). Binomial upper bounds on generalized moments and tail probabilities of (super) martingales with differences bounded from above. *Lecture Notes-Monograph Series* 33–52.
- [57] PINELIS, I. (2007). Exact inequalities for sums of asymmetric random variables, with applications. *Probability Theory and Related Fields* **139** 605–635. <https://doi.org/10.1007/s00440-007-0055-4>
- [58] PINELIS, I. (2016). Optimal binomial, Poisson, and normal left-tail domination for sums of nonnegative random variables.
- [59] RIO, E. (2009). Moment inequalities for sums of dependent random variables under projective conditions. *Journal of Theoretical Probability* **22** 146–163.
- [60] ROBBINS, H. (1955). A remark on Stirling's formula. *The American Mathematical Monthly* **62** 26–29.
- [61] ROMANO, J. P. and WOLF, M. (2000). Finite sample nonparametric inference and large sample efficiency. *Annals of Statistics* 756–778.

- [62] ROSS, N. (2011). Fundamentals of Stein's method. *Probability Surveys* **8** 210–293.
- [63] SHEVTSOVA, I. G. (2010). An improvement of convergence rate estimates in the Lyapunov theorem. *Doklady Mathematics* **82** 862–864. <https://doi.org/10.1134/S1064562410060062>
- [64] SHEVTSOVA, I. (2011). On the absolute constants in the Berry-Esseen type inequalities for identically distributed summands. *arXiv preprint arXiv:1111.6554*.
- [65] SHEVTSOVA, I. (2017). On the absolute constants in Nagaev–Bikelis-type inequalities. In *Inequalities and Extremal Problems in Probability and Statistics* 47–102. Elsevier, London.
- [66] STEIN, C. (1986). Approximate computation of expectations. IMS.
- [67] TALAGRAND, M. (1995). The missing factor in Hoeffding's inequalities. In *Annales de l'IHP Probabilités et statistiques* **31** 689–702.
- [68] WAINWRIGHT, M. J. (2019). *High-Dimensional Statistics: A Non-Asymptotic Viewpoint. Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.
- [69] WAUDBY-SMITH, I. and RAMDAS, A. (2024). Estimating means of bounded random variables by betting. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **86** 1–27.
- [70] ZHOU, W., VEITCH, V., AUSTERN, M., ADAMS, R. P. and ORBANZ, P. (2019). Non-vacuous Generalization Bounds at the ImageNet Scale: a PAC-Bayesian Compression Approach. In *International Conference on Learning Representations*.