NEW BERRY-ESSEEN AND WASSERSTEIN BOUNDS IN THE
CLT FOR NON-RANDOMLY CENTERED RANDOM SUMS BY
PROBABILISTIC METHODS

CHRISTIAN DÖBLER

Abstract. We prove abstract bounds on the Wasserstein and Kolmogorov
distances between non-randomly centered random sums of real i.i.d. random variables
with a finite third moment and the standard normal distribution. Except for the
case of mean zero summands, these bounds involve a coupling of the summation in-
dex with its size biased distribution as was previously considered in [GR96] for the
normal approximation of nonnegative random variables. When being specialized to
concrete distributions of the index like the Binomial, Poisson and Hypergeometric
distribution, our bounds turn out to be of the correct order of magnitude.

1. Introduction

Let \( N, X_1, X_2, \ldots \) be random variables on a common probability space such that
the \( X_j, j \geq 1 \), are real-valued and \( N \) assumes values in the nonnegative integers
\( \mathbb{Z}_+ = \{0, 1, \ldots \} \). Then, the random variable

\[
S := \sum_{j=1}^{N} X_j
\]

is called a random sum. Such random variables appear frequently in modern proba-
bility theory, as many models for example from physics, finance, reliability and risk
theory naturally lead to the consideration of such sums. Furthermore, sometimes a
model, which looks quite different from (1) at the outset, may be transformed into
a random sum and then general theory of such sums may be invoked to study the
original model [GK96]. There already exists a huge body of literature about the
asymptotic distributions of random sums. Their investigation evidently began with
the work [Rob48] of Robbins, who assumes that the random variables \( X_1, X_2, \ldots \) are
i.i.d. with a finite second moment and that \( N \) also has a finite second moment. One
of the results of [Rob48] is that under these assumptions asymptotic normality of the
index \( N \) automatically implies asymptotic normality of the corresponding random
sum. The book [GK96] gives a comprehensive description of the limiting behaviour
of such random sums under the assumption that the random variables \( N, X_1, X_2, \ldots \)
are independent. In particular, one may ask under what conditions the sum \( S \) in (1)
is asymptotically normal, where asymptotically refers to the fact that the random
index \( N \) in fact usually depends on a parameter, which is send either to infinity or to
zero. Once a CLT is known to hold, one might ask about the accuracy of the normal
approximation to the distribution of the given random sum. It turns out that it
is generally much easier to derive rates of convergence for random sums of centered random variables, or, which amounts to the same thing, for random sums centered by random variables than for random sums of not necessarily centered random variables. In the centered case one might, for instance, first condition on the value of the index $N$, then use known error bounds for sums of a fixed number of independent random variables like the classical Berry-Esseen theorem and, finally, take expectation with respect to $N$. This technique is illustrated e.g. in the manuscript [Dob12] and also works for non-normal limiting distributions like the Laplace distribution. For this reason we will mainly be interested in deriving sharp rates of convergence for the case of non-centered summands, but will also consider the mean-zero case and hint at the relevant differences. Also, we will not assume from the outset that the index $N$ has a certain fixed distribution like the Binomial or the Poisson, but will be interested in the general situation.

For non-centered summands and general index $N$, the relevant literature on rates of convergence in the random sums CLT seems quite easy to survey. Under the same assumptions as in [Rob48] the paper [Eng83] gives an upper bound on the Kolmogorov distance between the distribution of the random sum and a suitable normal distribution, which is proved to be sharp in some sense. However, this bound is not very explicit as it contains the Kolmogorov distance of $N$ to the normal distribution with the same mean and variance as $N$ as one of the terms appearing in the bound, for instance. This might make the task of applying this result difficult for a concrete distribution of $N$. Furthermore, the method of proof cannot be easily adapted to probability metrics different from the Kolmogorov distance like e.g. the Wasserstein distance. In [Kor87] a bound on the Kolmogorov distance is given which improves upon the result of [Eng83] with respect to the constants appearing in the bound. However, the bound given in [Kor87] is no longer strong enough to assure the well-known asymptotic normality of Binomial and Poisson random sums, unless the summands are centered.

To the best of our knowledge, the article [Sun14] is the only one, which gives bounds on the Wasserstein distance between random sums for general indices $N$ and the standard normal distribution. However, as mentioned by the same author in [Sun14], the results of [Sun13] generally do not yield accurate bounds, unless the summands are centered. Indeed, the results from [Sun13] do not even yield convergence in distribution for Binomial random sums of non-centered summands.

The main purpose of the present article is to combine Stein’s method of normal approximation with several modern probabilistic concepts like certain coupling constructions and conditional independence, to prove accurate abstract upper bounds on the distance between suitably standardized random sums of i.i.d. summands measured by two popular probability metrics, the Kolmogorov and Wasserstein distances. These upper bounds, in their most abstract forms (see Theorem 2.3 and Corollary 2.8 below), involve moments of the difference of a coupling of $N$ with its size-biased distribution but reduce to very explicit expressions if either $N$ has a concrete distribution like the Binomial, Poisson or dirac delta distribution, the summands $X_j$ are centered, or, if the distribution of $N$ is infinitely divisible. These special cases are extensively presented in order to illustrate the wide applicability and strength of our results. As indicated above, this seems to be the first work which gives Wasserstein bounds in the random sums CLT for general indices $N$, which reduce to bounds of optimal order, when specializing to concrete distributions like the Binomial and the
Poisson distributions. Using our abstract approach via size-bias couplings, we are also able to prove rates for Hypergeometric random sums. These do not seem to have been treated in the literature, yet. This is not a surprise, because the Hypergeometric distribution is conceptually more complicated than the Binomial or Poisson distribution, as it is neither a natural convolution of i.i.d. random variables nor infinitely divisible.

It should be mentioned that Stein’s method and coupling techniques have previously been used to bound the error of exponential approximation [PR11] and approximation by the Laplace distribution [PR12] of certain random sums.

The remainder of the article is structured as follows: In Section 2 we review the relevant probability distances, the size biased distribution and state our quantitative results on the normal approximation of random sums. Furthermore, we prove new identities for the distance of a nonnegative random variable to its size-biased distribution in three prominent distances and show that for some concrete distributions, natural couplings are $L^1$-optimal and, hence, yield the Wasserstein distance. In Section 3 we collect necessary facts from Stein’s method of normal approximation and introduce a variant of the zero-bias transformation, which we need for the proofs of our results. Then, in Section 4, the proof of our main theorems, Theorem 2.5 and Theorem 2.6 is given. Finally, Section 5 contains the proofs of some auxiliary results, needed for the proof in Section 4.

**2. Main results**

Recall that for probability measures $\mu$ and $\nu$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, their Kolmogorov distance is defined by

$$d_K(\mu, \nu) := \sup_{z \in \mathbb{R}} |\mu((-\infty, z]) - \mu((-\infty, z])| = \|F - G\|_\infty,$$

where $F$ and $G$ are the distribution functions corresponding to $\mu$ and $\mu$, respectively. Also, if both $\mu$ and $\nu$ have finite first absolute moment, then one defines the Wasserstein distance between them via

$$d_W(\mu, \nu) := \sup_{h \in \text{Lip}(1)} \left| \int h \mu - \int h \nu \right|,$$

where $\text{Lip}(1)$ denotes the class of all Lipschitz-continuous functions $g$ on $\mathbb{R}$ with Lipschitz constant not greater than 1. In view of Lemma 2.1 below, we also introduce the total variation distance between $\mu$ and $\nu$ by

$$d_{TV}(\mu, \nu) := \sup_{B \in \mathcal{B}(\mathbb{R})} |\mu(B) - \nu(B)|.$$

If the real-valued random variables $X$ and $Y$ have distributions $\mu$ and $\nu$, respectively, then we simply write $d_K(X, Y)$ for $d_K(\mathcal{L}(X), \mathcal{L}(Y))$ and similarly for the Wasserstein and total variation distances and also speak of the respective distance between the random variables $X$ and $Y$. Before stating our results, we have to review the concept of the size-biased distribution corresponding to a distribution supported on $[0, \infty)$. Thus, if $X$ is a nonnegative random variable with $0 < E[X] < \infty$, then a random variable $X^*$ is said to have the $X$-size biased distribution, if for all bounded and measurable functions $h$ on $[0, \infty)$

$$E[Xh(X)] = E[X]E[h(X^*)],$$

where $X^*$ is the random variable with the size-biased distribution of $X$. This is possible because

$$E[X^*] = E[X]$$

and

$$E[Xh(X^*)] = E[X]E[h(X^*)].$$

This is a direct consequence of the fact that $X^*$ is a random variable with the size-biased distribution of $X$ and that $E[X] = E[X^*]$.
see, e.g. [GR96], [AG10] or [AGK13]. Equivalently, the distribution of $X^s$ has Radon-Nikodym derivative with respect to the distribution of $X$ given by
\[
\frac{P(X^s \in dx)}{P(X \in dx)} = \frac{x}{E[X]} \, .
\]
which immediately implies both existence and uniqueness of the $X$-size biased distribution. Also note that (2) holds true for all measurable functions $h$ for which $E|Xh(X)| < \infty$. In consequence, if $X \in L^p(P)$ for some $1 \leq p < \infty$, then $X^s \in L^{p-1}(P)$ and
\[
E[(X^s)^{p-1}] = \frac{E[X^p]}{E[X]} \, .
\]

The following lemma, which seems to be new and might be of independent interest, gives identities for the distance of $X$ to $X^s$ in the three metrics mentioned above. The proof is deferred to the end of this section.

**Lemma 2.1.** Let $X$ be a nonnegative random variable such that $0 < E[X] < \infty$. Then, the following identities hold true:

(a) $d_K(X, X^s) = \frac{E[(X - E[X])1_{\{X > E[X]\}}]}{E[X]}$

(b) $d_{TV}(X, X^s) = \frac{E|X - E[X]|}{2E[X]}$

(c) If additionally $E[X^2] < \infty$, then $d_W(X, X^s) = \frac{\text{Var}(X)}{E[X]}$.

**Remark 2.2.** (a) It is well known (see e.g. [Dud02]) that the Wasserstein distance $d_W(X, Y)$ between the real random variables $X$ and $Y$ has the dual representation
\[
d_W(X, Y) = \inf_{(\hat{X}, \hat{Y}) \in \pi(X, Y)} E|\hat{X} - \hat{Y}| \, ,
\]
where $\pi(X, Y)$ is the collection of all couplings of $X$ and $Y$, i.e. of all pairs $(\hat{X}, \hat{Y})$ of random variables on a joint probability space such that $\hat{X} \overset{D}{=} X$ and $\hat{Y} \overset{D}{=} Y$. Also, the infimum in (3) is always attained, e.g. by the quantile transformation: If $U$ is uniformly distributed on $(0, 1)$ and if, for a distribution function $F$ on $\mathbb{R}$, we let
\[
F^{-1}(p) := \inf\{x \in \mathbb{R} : F(x) \geq p\}, \quad p \in (0, 1) \, ,
\]
denote the corresponding generalized inverse of $F$, then $F^{-1}(U)$ is a random variable with distribution function $F$. Thus, letting $F_X$ and $F_Y$ denote the distribution functions of $X$ and $Y$, respectively, it was proved e.g. in [Maj78] that
\[
\inf_{(\hat{X}, \hat{Y}) \in \pi(X, Y)} E|\hat{X} - \hat{Y}| = E|F_X^{-1}(U) - F_Y^{-1}(U)| = \int_0^1 |F_X^{-1}(t) - F_Y^{-1}(t)| dt \, .
\]
Furthermore, it is not difficult to see that $X^s$ is always stochastically larger than $X$, implying that there is a coupling $(\hat{X}, \hat{X}^s)$ of $X$ and $X^s$ such that $\hat{X}^s \geq \hat{X}$ (see [AG10] for details). In fact, this property is already achieved by the coupling via the quantile transformation. By the dual representation (3) and the fact that the coupling via the quantile transformation yields the minimum $L^1$ distance in
we can conclude that every coupling \((\hat{X}, \hat{X}^s)\) such that \(\hat{X}^s \geq \hat{X}\) is optimal in this sense, since
\[
E[\hat{X}^s - \hat{X}] = E[\hat{X}^s] - E[\hat{X}] = E[F_{X^s}^{-1}(U)] - E[F_X^{-1}(U)] \\
= E[F_{X^s}^{-1}(U) - F_X^{-1}(U)] = d_W(X, X^s).
\]
(b) Due to a result by Steutel [Ste73], the distribution of \(X\) is infinitely divisible, if and only if there exists a coupling \((X, X^s)\) of \(X\) and \(X^s\) such that \(X^s - X\) is nonnegative and independent of \(X\) (see e.g. [AG10] for a nice exposition and a proof of this result). According to (a) such a coupling always achieves the minimum \(L^1\)-distance.

**Example 2.3.**
(a) Let \(X \sim \text{Poisson}(\lambda)\) have the Poisson distribution with parameter \(\lambda > 0\). From the Stein characterization of Poisson(\(\lambda\)) (see [Che75]) it is known that
\[
E[Xf(X)] = \lambda E[f(X + 1)] = E[X]E[f(X + 1)]
\]
for all bounded and measurable \(f\). Hence, \(X + 1\) has the \(X\)-size biased distribution. As \(X + 1 \geq X\), by Remark 2.2 this coupling yields the minimum \(L^1\)-distance between \(X\) and \(X^s\), which is equal to 1 in this case.

(b) Let \(n\) be a positive integer, \(p \in (0, 1]\) and let \(X_1, \ldots, X_n\) be i.i.d. random variables such that \(X_1 \sim \text{Bernoulli}(p)\). Then,
\[
X := \sum_{j=1}^{n} X_j \sim \text{Bin}(n, p)
\]
has the Binomial distribution with parameters \(n\) and \(p\). From the construction in [GR96] one easily sees that
\[
X^s := \sum_{j=2}^{n} X_j + 1
\]
has the \(X\)-size biased distribution. As \(X^s \geq X\), by Remark 2.2 this coupling yields the minimum \(L^1\)-distance between \(X\) and \(X^s\), which is equal to
\[
d_W(X, X^s) = E[1 - X] = 1 - p = \frac{\text{Var}(X)}{E[X]}
\]
in accordance with Lemma 2.1.

(c) Let \(n, r, s\) be positive integers such that \(n \leq r + s\) and let \(X \sim \text{Hyp}(n; r, s)\) have the Hypergeometric distribution with parameters \(n, r\) and \(s\), i.e.
\[
P(X = k) = \binom{r}{k} \binom{s}{n-k} \binom{n}{r+s}, \quad k = 0, 1, \ldots, n.
\]
Then, \(E[X] = \frac{nr}{r+s}\) and, hence,
\[
P(X^s = k) = \frac{kP(X = k)}{E[X]} = \frac{k \binom{r}{k} \binom{s}{n-k} \binom{n}{r+s}}{\binom{r}{k-1} \binom{s}{n-1} \binom{r-1}{r+s}} = \binom{r-1}{r+s} \binom{s}{n-1}, \quad k = 1, 2, \ldots, n.
\]
Thus,
\[
X^s \overset{D}{=} Y + 1, \quad \text{where } Y \sim \text{Hyp}(n-1; r-1, s).
\]
Imagine an urn with \(r\) red and \(s\) silver balls. If we draw \(n\) times without replacement from this urn and denote by \(X\) the total number of drawn red balls,
then \( X \sim \text{Hyp}(n; r, s) \). For \( j = 1, \ldots, n \) denote by \( X_j \) the indicator of the event that a red ball is drawn at the \( j \)-th draw. Then, \( X = \sum_{j=1}^{n} X_j \). Also, fix one of the red balls in the urn and, for \( j = 2, \ldots, n \), denote by \( Y_j \) the indicator of the event that at the \( j \)-th draw this fixed red ball is drawn. Then, it is not difficult to see that

\[
Y := 1_{\{X_1=1\}} \sum_{j=2}^{n} X_j + 1_{\{X_1=0\}} \sum_{j=2}^{n} (X_j - Y_j) \sim \text{Hyp}(n-1; r-1, s)
\]

and, hence,

\[
X^s := Y + 1 = 1_{\{X_1=1\}} \sum_{j=2}^{n} X_j + 1_{\{X_1=0\}} \sum_{j=2}^{n} (X_j - Y_j) + 1
\]

\[
= 1_{\{X_1=1\}} X + 1_{\{X_1=0\}} (X + 1 - \sum_{j=2}^{n} Y_j)
\]

has the \( X \)-size biased distribution. Note that since \( \sum_{j=2}^{n} Y_j \leq 1 \) we have

\[
X^s - X = 1_{\{X_1=0\}} (1 - \sum_{j=2}^{n} Y_j) \geq 0,
\]

and consequently, by Remark 2.2 (a), the coupling \((X, X^s)\) is optimal in the \( L^1 \)-sense and yields the Wasserstein distance between \( X \) and \( X^s \):

\[
d_W(X, X^s) = E|X^s - X| = \frac{\text{Var}(X)}{E[X]} = \frac{n \frac{r}{r+s} \frac{s}{r+s} \frac{r+s-n}{r+s-1}}{\frac{n}{r+s}} = \frac{s(r+s-n)}{(r+s)(r+s-1)}.
\]

We now turn back to the asymptotic behaviour of random sums. We will rely on the following general assumptions and notation, which we adopt and extend from [Rob48].

**Assumption 2.4.** The random variables \( N, X_1, X_2, \ldots \) are independent, \( X_1, X_2, \ldots \) being i.i.d. and such that \( E|X_1|^3 < \infty \) and \( E[X^3] < \infty \). Furthermore, we let

\[
\alpha := E[N], \quad \beta^2 := E[N^2], \quad \gamma^2 := \text{Var}(N) = \beta^2 - \alpha^2, \quad \delta^3 := E[N^3],
\]

\[
a := E[X_1], \quad b^2 := E[X_1^2], \quad c^2 := \text{Var}(X_1) = b^2 - a^2, \quad \xi := E[X_1 - E[X_1]]^3.
\]

By Wald’s equation and the Blackwell-Girshick formula, from Assumption 2.4 we have

\[
(4) \quad \mu := E[S] = \alpha a \quad \text{and} \quad \sigma^2 := \text{Var}(S) = \alpha c^2 + a^2 \gamma^2.
\]

The main purpose of this paper is to assess the accuracy of the standard normal approximation to the normalized version

\[
(5) \quad W := \frac{S - \mu}{\sigma} = \frac{S - \alpha a}{\sqrt{\alpha c^2 + a^2 \gamma^2}}
\]

of \( S \) measured by the Kolmogorov and the Wasserstein distance, respectively. As can be seen from the paper [Rob48], under the general assumption that

\[
\sigma^2 = \alpha c^2 + a^2 \gamma^2 \rightarrow \infty,
\]

there are four typical situations in which \( W \) is asymptotically normal, which we will now briefly review.
1) \( a = 0 \neq c \) and \( \gamma = o(\alpha) \)

2) \( c \neq 0 \neq a \) and \( \gamma^2 = o(\alpha) \)

3) \( c \neq 0 \) and \( N \) itself is asymptotically normal

4) \( c = 0 \neq a \) and \( N \) itself is asymptotically normal

We remark that 1) and 2) roughly mean that \( N \) tends to infinity in a certain sense, but such that it only fluctuates slightly around its mean \( \alpha \). For instance, this is the case, whenever \( N \) is roughly equal to \( \alpha \). Of course, 4) is quite uninteresting, here, since in this case

\[
S = aN \quad \text{a.s.}
\]

Although we do not exclude such cases explicitly, we will taciturnly assume that \( c \neq 0 \) in what follows.

**Theorem 2.5.** Let Assumption 2.4 hold, let \( W \) be given by (5) and let \( Z \) have the standard normal distribution Also, let \( (N, N^*) \) be a coupling of \( N \) and \( N^* \) having the \( N \)-size biased distribution and define \( D := N^* - N \). Then,

\[
d_W(W, Z) \leq \frac{2c^2b\gamma^2}{\sigma^3} + \frac{3\alpha \xi}{\sigma^3} + \frac{\alpha a^2}{\sigma^2} \sqrt{\frac{2}{\pi}} \sqrt{\text{Var}(E[D | N])}
\]

\[
+ \frac{2\alpha a^2 b}{\sigma^3} E[1_{\{D<0\}} D^2] + \frac{\alpha |a| b^2}{\sigma^3} E[D^2]
\]

and

\[
d_K(W, Z) \leq \sum_{j=1}^{7} B_j,
\]

where \( B_1, B_2, B_4, B_5, B_6 \) and \( B_7 \) are defined in (70), (75), (82), (88), (97) and (103), respectively, and

\[
B_3 := \frac{\alpha a^2}{\sigma^2} \sqrt{\text{Var}(E[D | N])}.
\]

**Theorem 2.6.** Assume that Assumption 2.4 holds with \( a = E[X_1] = 0 \), let \( W \) be given by (5) and let \( Z \) have the standard normal distribution. Then,

\[
d_W(W, Z) \leq \frac{2\gamma}{\alpha} + \frac{3\xi}{c^3 \sqrt{\alpha}} \quad \text{and}
\]

\[
d_K(W, Z) \leq \left( \frac{\sqrt{2\pi} + 4}{4\alpha} \right) \gamma + \left( \frac{\xi (3\sqrt{2\pi} + 4)}{8c^3} + 1 \right) \frac{1}{\sqrt{\alpha}} + \left( \frac{7}{2} \sqrt{2} + 2 \right) \frac{\xi}{c^3 \alpha}
\]

\[
+ P(N = 0) + \left( \frac{2C_K \xi}{c^3} + \frac{\gamma}{\sqrt{\alpha} \sqrt{2\pi}} \right) \sqrt{E[1_{\{N\geq1\}} N^{-1}]}.
\]

**Remark 2.7.** (a) The proof will show that Theorem 2.6 holds as long as \( E[N^2] < \infty \). Thus, Assumption 2.4 could be slightly relaxed in this case. Also, for \( a \neq 0 \) it would be possible to extend the result to non-identically distributed summands.

(b) Theorem 2.6 is not a direct consequence of Theorem 2.5 as it is stated above. Actually, instead of Theorem 2.5 we could state a result, which would reduce to Theorem 2.6 if \( a = 0 \), but the resulting bounds would look even more cumbersome in the general case. Also, they would be of the same order as the bounds presented in Theorem 2.5 if \( a \neq 0 \). This is why we have refrained from presenting these bounds in the general case. Instead, we have chosen to prove Theorem 2.5 and Theorem 2.6 in parallel.
Since $N^*$ is stochastically larger than $N$, by Remark 2.2 (a) it is always possible to construct a coupling $(N, N^*)$ such that $D = N^* - N \geq 0$. For such a coupling we obtain

\begin{equation}
E[D] = E[D] = E[N^*] - E[N] = \frac{E[N^2]}{E[N]} - E[N] = \frac{\text{Var}(N)}{E[N]} = \frac{\gamma^2}{\alpha},
\end{equation}

and the bounds in Theorem 2.5 can be further simplified.

**Corollary 2.8.** Assume that Assumption 2.4 holds, let $W$ be given by (5) and let $Z$ have the standard normal distribution. Also, let $(N, N^*)$ be a coupling of $N$ and $N^*$ having the $N$-size biased distribution such that $D = N^* - N \geq 0$. Then, we have:

\begin{align*}
d_W(W, Z) &\leq \frac{2c^2b\gamma^2}{\sigma^3} + \frac{3\alpha\xi}{\sigma^3} + \frac{\alpha a^2}{\sigma^2} \sqrt{\frac{2}{\pi}} \sqrt{\text{Var}(E[D | N])} + \frac{\alpha |a| b^2}{\sigma^3} E[D^2] \\
d_K(W, Z) &\leq \frac{(\sqrt{2\pi} + 4)bc^2\alpha}{4\sigma^3} \sqrt{E[D^2]} + \frac{\xi \alpha (3\sqrt{2\pi} + 4)}{8\sigma^3} + \frac{c^3\alpha}{\sigma^3} \\
 &+ \frac{\alpha a^2}{\sigma^2} \sqrt{\text{Var}(E[D | N])} + \frac{\alpha |a| b^2}{2\sigma^3} \sqrt{E[(E[D^2 | N])^2]} \\
 &+ \frac{\alpha |a| b^2}{\sigma^3} \sqrt{2\pi} E[D^2] + \frac{\alpha |a| b^2}{\sigma^3} \sqrt{E[(E[D^2 | N])^2]} \\
 &+ \frac{\alpha |a| b^2}{\sigma^2} \sqrt{2\pi} E[D^2 1_{\{N\geq 1\}} N^{-1/2}] + \left(\frac{\xi \alpha |a| b^2}{\sigma^2} + \frac{\alpha b^2c}{\sigma^2 \sqrt{2\pi}}\right) E[D1_{\{N\geq 1\}} N^{-1/2}]
\end{align*}

**Remark 2.9.** (a) One may check that our bounds prove the CLT for random sums, if one of the situations 1) - 4) holds. Also note that if $a \neq 0$, then a necessary condition for the CLT to hold is that

\[
\frac{\alpha}{\sigma^2} E[D^2] = o(1) \quad \text{and} \quad \frac{\alpha}{\sigma \pi} \sqrt{\text{Var}(E[D | N])} = o(1).
\]

This should be compared to conditions which imply asymptotic normality for $N$ by size-bias couplings in [GR96].

(b) In many practical applications, we will have that a natural coupling of $N$ and $N^*$ yields $D \geq 0$ and, hence, Corollary 2.8 can be applied. For instance, by Remark 2.2 (b), this is the case, if the distribution of $N$ is infinitely divisible. In this case, the random variables $D$ and $N$ can be chosen to be independent and, thus, the bounds from Corollary 2.8 further simplify (see Corollary 2.10 below).

(c) However, though always possible, a coupling of $N$ and $N^*$ such that $D = N^* - N \geq 0$ is not always the most feasible, see Corollary 2.12 below, for example. This is why we have not restricted ourselves to the case $D \geq 0$.

(d) For distribution functions $F$ and $G$ on $\mathbb{R}$ and $1 \leq p < \infty$, one defines their $L^p$-distance by

\[
\|F - G\|_p := \left(\int_{\mathbb{R}} |F(x) - G(x)|^p dx\right)^{1/p}.
\]

It is known (see [Dud02]) that $\|F - G\|_1$ coincides with the Wasserstein distance of the corresponding distributions $\mu$ and $\nu$, say. By Hölder’s inequality, for
1 ≤ p < ∞, we have
\[ \|F - G\|_p \leq d_K(\mu, \nu)^{\frac{p-1}{p}} \cdot d_W(\mu, \nu)^{\frac{1}{p}}. \]

Thus, our results immediately yield bounds on the \(L^p\)-distances of \(\mathcal{L}(W)\) and \(N(0,1)\).

**Corollary 2.10.** In addition to the assumptions from Theorem 2.5, suppose that the distribution of the index \(N\) is infinitely divisible. Then, we have:
\[
d_W(W, Z) \leq \frac{2c^2b\gamma^2 + 3\alpha\xi}{\sigma^3} + \frac{(\alpha\delta^3 - \alpha^2\gamma^2 + \gamma^4 - \beta^4)|a|b^2}{c\sigma^3}
\]
\[
d_K(W, Z) \leq \frac{\xi\alpha(3\sqrt{2\pi} + 4)}{8\sigma^3} + \frac{c^3\alpha}{\sigma^3} + \left(\frac{7\sqrt{2} + 2}{\sqrt{\alpha}\xi}\right)\frac{\sqrt{\alpha}}{c\sigma^2} + \frac{c^2\alpha}{\sigma^2}P(N = 0)
\]
\[
+ \frac{|a|b^2(\delta^3\alpha + \gamma^4 - \beta^4 - \gamma^2\alpha^2)}{\alpha\sigma^3}\left(\frac{\sqrt{2\pi}}{8} + \frac{1}{2}\right)
\]
\[
+ \frac{\sqrt{\alpha\delta^3\gamma^4 - \beta^4 - \gamma^2\alpha^2}}{\alpha\sigma^3}\left(\frac{(\sqrt{2\pi} + 4)bc}{4\sigma^3} + \sqrt{P(N = 0)}\frac{|a|b}{\sigma^2}\right)
\]
\[
+ E[1_{\{N \geq 1\}}N^{-1/2}]
\]
\[
\left(\frac{|a|b^2(\delta^3\alpha + \gamma^4 - \beta^4 - \gamma^2\alpha^2)}{c\alpha\sigma^2\sqrt{2\pi}}\frac{\gamma^2\xi|a|b}{\sigma^2} + \frac{\alpha\xi}{c\sigma^2} + \frac{\gamma^2bc}{\sigma^2\sqrt{2\pi}}\right)
\]

**Proof.** By Remark 2.2 (b) we can choose \(D \geq 0\) independent of \(N\) such that \(N^* = N + D\) has the \(N\)-size biased distribution. Thus, by independence we obtain
\[
\text{Var}(D) = \text{Var}(N^*) - \text{Var}(N) = E\left[(N^*)^2\right] - E[N^*]^2 - \gamma^2
\]
\[
= \frac{E[N^3]}{E[N]} - \left(\frac{E[N^2]}{E[N]}\right)^2 - \gamma^2
\]
\[
= \frac{\delta^3}{\alpha} - \frac{\beta^4}{\alpha^2} - \gamma^2.
\]
This gives
\[
E[D^2] = \text{Var}(D) + E[D]^2 = \frac{\delta^3}{\alpha} + \frac{\gamma^4 - \beta^4}{\alpha^2} - \gamma^2.
\]
Also,
\[
\text{Var}(E[D|N]) = \text{Var}(E[D]) = 0 \quad \text{and} \quad \sqrt{E\left[\left(E[D^2|N]\right)^2\right]} = E[D^2]
\]
in this case. Now, the claim follows from Corollary 2.8.

\[ \square \]

In the case that \(N\) is constant, the results from Theorem 2.5 reduce to the known optimal convergence rates for sums of i.i.d. random variables with finite third moment, albeit with non-optimal constants (see e.g. [She11] and [Gol10] for comparison).

**Corollary 2.11.** In addition to the assumptions from Theorem 2.5, suppose that the index \(N\) is a positive constant. Then,
\[
d_W(W, Z) \leq \frac{3\xi}{c^3\sqrt{N}} \quad \text{and}
\]
\[
d_K(W, Z) \leq \frac{1}{\sqrt{N}} \left(1 + \left(\frac{7}{2}(1 + \sqrt{2}) + \frac{3\sqrt{2\pi}}{8}\right)\frac{\xi}{c^3}\right).
\]
Another typical situation when the distribution of \( W \) may be well approximated by the normal is if the index \( N \) is itself a sum of many i.i.d. variables. Our results yield very explicit convergence rates in this special case. This will be exemplified for the Wasserstein distance by the next corollary. A different way to prove bounds for the CLT by Stein’s method in this situation is presented in Theorem 10.6 of [CGS11].

**Corollary 2.12.** In addition to the assumptions from Theorem 2.5 suppose that the distribution of the index \( N \) is such that \( N = N_1 + \ldots + N_n \), where \( n \in \mathbb{N} \) and \( N_1, \ldots, N_n \) are i.i.d. nonnegative random variables such that \( E[N_1^3] < \infty \). Then, using the notation

\[
\begin{align*}
\alpha_1 & := E[N_1], \quad \beta_1^2 := E[N_1^2], \quad \gamma_1^2 := \text{Var}(N_1), \quad \delta_1^3 := E[N_1^3] \\
\sigma_1^2 & := c^2 \alpha_1 + a^2 \gamma_1^2
\end{align*}
\]

we have

\[
d_{W}(W, Z) \leq \frac{1}{\sqrt{n}} \left( \frac{2c^2 b \gamma_1^2}{\sigma_1^3} + \frac{3 \alpha_1 \xi}{\sigma_1^3} + \sqrt{\frac{2}{\pi}} \frac{\alpha_1 a^2 \gamma_1^2}{\sigma_1^3} + \frac{2 \alpha_1 (a^2 b + |a| b^2)}{\sigma_1^3} \left( \frac{\delta_3^3}{\alpha_1} - \beta_1^2 \right) \right).
\]

**Proof.** From [GR96] (see also [CGS11]) it is known that letting \( N_1^s \) be independent of \( N_1, \ldots, N_n \) and have the \( N_1 \)-size biased distribution, a random variable with the \( N \)-size biased distribution is given by

\[
N^s := N_1^s + \sum_{j=2}^{n} N_j, \quad \text{yielding} \quad D = N_1^s - N_1.
\]

Thus, by independence and since \( N_1, \ldots, N_n \) are i.i.d., we have

\[
E[D|N] = E[N_1^s] - \frac{1}{n} N
\]

and, hence,

\[
\text{Var}(E[D|N]) = \frac{\text{Var}(N)}{n^2} = \frac{\gamma_1^2}{n}.
\]

Clearly, we have

\[
\alpha = n \alpha_1, \quad \gamma^2 = n \gamma_1^2 \quad \text{and} \quad \sigma^2 = n \sigma_1^2.
\]

Also, using independence and (2),

\[
E[D^2] = E[N_1^2 - 2 N_1 N_1^s + (N_1^s)^2] = \beta_1^2 - 2 \alpha_1 E[N_1^s] + E[(N_1^s)^2]
\]

\[
= \beta_1^2 - 2 \alpha_1 \beta_1^2 + \frac{\delta_1^3}{\alpha_1} = \frac{\delta_1^3}{\alpha_1} - \beta_1^2.
\]

Thus, the bound follows from Theorem 2.5. \( \square \)

Very prominent examples of random sums, which are known to be asymptotically normal, are Poisson and Binomial random sums. The respective bounds, which follow from our abstract findings, are presented in the next two corollaries.
**Corollary 2.13.** In addition to the assumptions from Theorem 2.5 assume that $N \sim \text{Poisson} \left( \lambda \right)$ has the Poisson distribution with parameter $\lambda > 0$. Then,

$$d_W(W, Z) \leq \frac{1}{\sqrt{\lambda}} \left( \frac{2c^2}{b^2} + \frac{3\xi}{b^3} + \frac{|a|}{b} \right) \quad \text{and} \quad d_K(W, Z) \leq \frac{1}{\sqrt{\lambda}} \left( \frac{\sqrt{2\pi}}{4} + 1 + \frac{(3\sqrt{2\pi} + 4)\xi}{8b^3} + \frac{c^3}{b^3} + \left( \frac{7}{2} \sqrt{2} + 3 \right) \frac{\xi}{cb^2} \right.$$

$$\left. + \frac{|a|}{8b} \left( \sqrt{2\pi} + 4 + 3\xi \right) + \frac{|a|}{c\sqrt{2\pi}} + \frac{c}{b\sqrt{2\pi}} \right) + \frac{c^2}{b^2} e^{-\lambda} + \frac{|a|}{b} e^{-\lambda/2}.$$

**Proof.** We apply the result of Corollary 2.8. In this case, by Example 2.3 (a), we can choose $D = 1$, yielding that $E[D^2] = 1$ and $\text{Var}(E[D|N]) = 0$.

Note that

$$E[1_{(N \geq 1)} N^{-1/2}] \leq \sqrt{E[1_{(N \geq 1)} N^{-1}]}$$

by Jensen’s inequality. Also, using $k + 1 \leq 2k$ for all $k \in \mathbb{N}$, we can bound

$$E[1_{(N \geq 1)} N^{-1}] = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{kk!} \leq 2e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k + 1)k!} = \frac{2}{\lambda} e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k+1}}{(k + 1)!}$$

$$= \frac{2}{\lambda} e^{-\lambda} \sum_{l=2}^{\infty} \frac{\lambda^l}{l!} \leq \frac{2}{\lambda}.$$

Hence,

$$E[1_{(N \geq 1)} N^{-1/2}] \leq \frac{\sqrt{2}}{\sqrt{\lambda}}.$$

Noting that

$$\alpha = \gamma^2 = \lambda \quad \text{and} \quad \sigma^2 = \lambda(a^2 + c^2) = \lambda b^2,$$

the result follows from Corollary 2.8.

**Remark 2.14.** The Berry-Esseen bound presented in Corollary 2.13 is of the same order of $\lambda$ as the bound given in [KS12], which seems to be the best currently available, but has a worst constant. However, it should be mentioned that the bound in [KS12] was obtained using special properties of the Poisson distribution and does not seem likely to be easily transferable to other distributions of $N$.

**Corollary 2.15.** In addition to the assumptions from Theorem 2.5 suppose that $N \sim \text{Bin}(n, p)$ has the Binomial distribution with parameters $n \in \mathbb{N}$ and $p \in (0, 1)$.
Then,

\[ d_W(W, Z) \leq \frac{1}{\sqrt{np}(b^2 - pa^2)^{3/2}} \left( (2c^2 b + |a|b^2)(1 - p) + 3\xi \right. \]

\[ + \sqrt{\frac{2}{\pi}} a^2 p b^2 - pa^2 \sqrt{1 - p} \right) \quad \text{and} \]

\[ d_K(W, Z) \leq \frac{1}{\sqrt{np}(b^2 - pa^2)^{3/2}} \left( c^3 + \frac{(\sqrt{2\pi} + 4)b^2c\sqrt{1 - p}}{4} \right. \]

\[ + \frac{|a|b^2\sqrt{1 - p}}{2} + \left. \frac{|a|b^2\sqrt{2\pi}(1 - p)}{8} \right) \]

\[ + \sqrt{2\pi}(b(2b^2 - a^2)) \]

\[ + \frac{c^2}{b^2 - pa^2}(1 - p)^n + \frac{|a|b}{b^2 - pa^2}(1 - p)^{n+1}. \]

**Remark 2.16.** Bounds for binomial random sums have also been derived in [Sun14] using a technique developed in [Tih80]. Our bounds are of the same order \((np)^{-1/2}\) of magnitude.

**Proof of Corollary 2.15.** Here, we clearly have

\[ \alpha = np, \quad \gamma^2 = np(1 - p) \quad \text{and} \quad \sigma^2 = np(a^2(1 - p) + c^2). \]

Also, using the same coupling as in Example 2.3 (b) we have \(D \sim \text{Bernoulli}(1 - p)\),

\[ E[D^2] = E[D] = 1 - p \quad \text{and} \quad E[D|N] = 1 - \frac{N}{n}. \]

This yields

\[ \text{Var}(E[D|N]) = \frac{1}{n^2} \text{Var}(N) = \frac{p(1 - p)}{n}. \]

We have \(D^2 = D\) and, by Cauchy-Schwartz,

\[ E[D1_{\{N \geq 1\}}N^{-1/2}] \leq \sqrt{E[D^2]}\sqrt{E[1_{\{N \geq 1\}}N^{-1}]} = \sqrt{1 - p} \sqrt{E[1_{\{N \geq 1\}}N^{-1}]} . \]

Using

\[ \frac{1}{k} \binom{n}{k} \leq \frac{2}{n + 1} \binom{n + 1}{k + 1} \leq \frac{2}{n} \binom{n + 1}{k + 1} , \quad 1 \leq k \leq n , \]

we have

\[ E[1_{\{N \geq 1\}}N^{-1}] = \sum_{k=1}^{n} \frac{1}{k} \binom{n}{k} p^k(1 - p)^{n-k} \leq \frac{2}{n} \sum_{k=1}^{n} \binom{n + 1}{k + 1} p^k(1 - p)^{n-k} \]

\[ = \frac{2}{np} \sum_{l=2}^{n+1} \binom{n + 1}{l} p^l(1 - p)^{n+1-l} \leq \frac{2}{np}. \]
Thus,

\[ E[D1_{\{N \geq 1\}}N^{-1/2}] \leq \frac{\sqrt{2(p-1)}}{\sqrt{np}} \quad \text{and} \quad E[1_{\{N \geq 1\}}N^{-1/2}] \leq \frac{\sqrt{2}}{\sqrt{np}}. \]

Also, we can bound

\[ E \left[ (E[D^2 \mid N])^2 \right] \leq E[D^4] = E[D] = 1 - p. \]

Now, using \( a^2 + c^2 = b^2 \), the claim follows from Corollary 2.8.

\[ \square \]

**Corollary 2.17.** In addition to the assumptions from Theorem 2.9, suppose that \( N \sim \text{Hyp}(n; r, s) \) has the Hypergeometric distribution with parameters \( n, r, s \in \mathbb{N} \) such that \( n \leq \min\{r, s\} \). Then,

\[
d_W(W, Z) \leq \left( \frac{nr}{r+s} \right)^{-1/2} \left( 2b \frac{s}{c} \frac{r+s-n}{r+s-1} + \frac{3\xi}{c^3} a \frac{s}{r+s} \frac{r+s-n}{r+s-1} \right) \\
+ \frac{a^2}{c^2} \sqrt{\frac{2}{\pi}} \sqrt{\varepsilon(n, r, s)} \quad \text{and} \quad \\
d_K(W, Z) \leq \left( \frac{nr}{r+s} \right)^{-1/2} \left[ 1 + \left( \frac{\sqrt{2}\pi + 4}{4c} \frac{s}{r+s-1} \right)^{1/2} \\
+ \left( \frac{3\sqrt{2}\pi \sqrt{8}}{8} + \frac{r+s-n}{2} \right) \frac{\xi}{c^3} \frac{s}{r+s-1} \right)^{1/2} \\
+ \left( \frac{3\sqrt{2}\pi \sqrt{8}}{8} + \frac{r+s-n}{2} \right) \frac{\xi}{c^3} \frac{s}{r+s-1} \right)^{1/2} \\
+ \frac{a^2}{c^2} \sqrt{\varepsilon(n, r, s)} + \frac{a b}{c^2} \left( \frac{(s)_n}{(r+s)_n} \frac{s(r+s-n)}{(r+s)(r+s-1)} \right)^{1/2}.
\]

where \( \varepsilon(n, r, s) \) is defined in \( \square \) below and \((m)_n = m(m-1) \ldots (m-n+1)\) denotes the lower factorial.

**Proof.** In this case, we clearly have

\[
\alpha = \frac{nr}{r+s}, \quad \gamma = \frac{nr}{r+s} \frac{s}{r+s-1} \quad \text{and} \quad \\
\sigma^2 = \frac{nr}{r+s} \left( c^2 + a^2 \frac{s}{r+s-1} \right).
\]

Hence,

\[
c^2 \frac{nr}{r+s} \leq \sigma^2 \leq \frac{nr}{r+s} \left( c^2 + a^2 \frac{s}{r+s} \right) = \frac{nr}{r+s} \left( b^2 - a^2 \frac{s}{r+s} \right).
\]

We use the coupling constructed in Example 2.3 (c) but write \( N \) for \( X \) and \( N^s \) for \( X^s \), here. Recall that we have

\[ D = N^s - N = 1_{\{X_1 = 0\}} \left( 1 - \sum_{j=2}^{n} Y_j \right) \geq 0 \quad \text{and} \quad D = D^2. \]

Furthermore, we know that

\[ E[D] = E[D^2] = d_W(N, N^s) = \frac{\text{Var}(N)}{E[N]} = \frac{s(r+s-n)}{(r+s)(r+s-1)}. \]
Elementary combinatorics yield

\[ E[Y_j | X_1, \ldots, X_n] = r^{-1}1_{\{X_j=1\}}. \]

Thus,

\[ E[D | X_1, \ldots, X_n] = 1_{\{X_1=0\}} - \frac{1}{r}1_{\{X_1=0\}} \sum_{j=2}^{n} 1_{\{X_j=1\}} = 1_{\{X_1=0\}} \left(1 - \frac{N}{r}\right) \quad \text{and} \]

\[ E[D | N] = \left(1 - \frac{N}{r}\right)P(X_1 = 0 | N) = \left(1 - \frac{N}{r}\right) \frac{n-N}{nr} = \left(1 - \frac{N}{r}\right) \left(1 - \frac{N}{n}\right). \]

Using a computer algebra system, one may check that

\[ \text{Var}(E[D | N]) = \left(nrs - n^3rs - r^2s + 5n^2r^2s + 2n^3r^2s - 8n^2r^3s - 8n^2r^3s + 2nrs^5 \right. \]
\[ - n^3r^3s + 4r^4s + 10nr^4s + 3n^2r^4s - 4r^5s - 3nr^5s + r^6s + ns^2 \]
\[ - n^3s^2 - 2rs^2 + 4n^2rs^2 - 2n^3rs^2 - 14nr^2s^2 - 4n^2r^2s^2 + n^3r^2s^2 \]
\[ + 12r^3s^2 + 20nr^3s^2 + 2n^2r^3s^2 - 14r^4s^2 - 7nr^4s^2 + 4r^5s^2 - s^3 \]
\[ - n^2s^3 + 2n^3s^3 - 5nrs^3 + 4n^2rs^3 + n^3rs^3 + 13r^2s^3 + 8nr^2s^3 \]
\[ - 4n^2r^2s^3 - 18r^3s^3 - 3nr^3s^3 + 6r^4s^3 + ns^4 - n^3s^4 + 6rs^4 - 4nrs^4 \]
\[ - 2n^2rs^4 - 10r^2s^4 + 3mr^2s^4 + 4r^3s^4 + s^5 - 2ns^5 + n^2s^5 - 2rs^5 \]
\[ + r^2s^5 \right) \left(nr(r+s)^{2}(r+s-1)^{2}(r+s-2)(r+s-3)\right)^{-1} \]
\[ \varepsilon(n, r, s). \]

(7)

Also, by the conditional version of Jensen’s inequality

\[ E\left[ (E[D^2 | N])^2 \right] \leq E[D^4] = E[D] = \frac{s(r+s-n)}{(r+s)(r+s-1)}. \]

Using

\[ E[N^{-1}1_{\{N\geq 1\}}] = \binom{r+s}{n}^{-1} \sum_{k=1}^{n} \frac{1}{k} \binom{r}{k} \binom{s}{n-k} \]
\[ \leq \frac{2}{r+1} \binom{r+s}{n}^{-1} \sum_{l=2}^{n+1} \binom{r+1}{l} \binom{s}{n+1-l} \]
\[ \leq \frac{2}{r+1} \binom{r+s}{n}^{-1} \binom{r+1+s}{n+1} = \frac{2(r+s+1)}{(n+1)(r+1)} \leq 2 \frac{r+s}{nr} , \]

we get

\[ E[D1_{\{N\geq 1\}}N^{-1/2}] \leq \sqrt{E[D^2]} \sqrt{E[1_{\{N\geq 1\}}N^{-1}]} \leq \left(2 \frac{s(r+s-n)}{(r+s)(r+s-1)} \frac{r+s}{nr}\right)^{1/2} \]
\[ \leq \sqrt{2} \left( \frac{s}{nr} \frac{r+s-n}{r+s-1} \right)^{1/2} \quad \text{and} \]

\[ E[1_{\{N\geq 1\}}N^{-1/2}] \leq \sqrt{E[1_{\{N\geq 1\}}N^{-1}]} \leq \sqrt{2} \left( \frac{r+s}{nr} \right)^{1/2}. \]
Finally, we have
\[
P(N = 0) = \binom{s}{n} = \frac{s(s - 1) \cdots (s - n + 1)}{(r + s)(r + s - 1) \cdots (r + s - n + 1)} = \frac{(s)_n}{(r + s)_n}.
\]

Thus, the result follows from Corollary 2.8.

\[\square\]

Remark 2.18. (1) One can check that under the assumption \(n \leq \min\{r, s\}\) always
\[
\varepsilon(n, r, s) = O\left(\frac{\min\{r, s\}}{n(r + s)}\right).
\]

As always
\[
\frac{\min\{r, s\}}{n(r + s)} \leq \frac{r + s}{nr} = \frac{1}{E[N]},
\]

we conclude that the bounds in Theorem 2.17 are of order \(E[N]^{-1/2}\).

(2) One typical situation, in which a CLT for Hypergeometric random sums holds, is when \(N\), itself, is asymptotically normal. Using the the same coupling \((N, N^*)\) as in the above proof and the results from [GR96], one obtains that under the condition
\[
\frac{\max\{r, s\}}{n \min\{r, s\}} \longrightarrow 0
\]

the index \(N\) is asymptotically normal. This condition is stricter than that
\[
E[N]^{-1} = \frac{r + s}{nr} \longrightarrow 0,
\]

which implies the random sums CLT. For instance, choosing
\[
r \propto n^{1+\varepsilon}, \quad \text{and} \quad s \propto n^{1+\kappa}
\]

with \(\varepsilon, \kappa \geq 0\), then (8) holds, if and only if \(|\varepsilon - \kappa| < 1\), whereas (9) is equivalent to \(\kappa - \varepsilon < 1\) in this case.

Proof of Lemma 2.1. Let \(h\) be a measurable function such that all the expected values in (2) exist. By (2) we have
\[
\left|E[h(X^s)] - E[h(X)]\right| = \frac{1}{E[X]} \left|E[(X - E[X])h(X)]\right|.
\]

It is well known that
\[
d_{TV}(X, Y) = \sup_{h \in \mathcal{H}} \left|E[h(X)] - E[h(Y)]\right|,
\]

where \(\mathcal{H}\) is the class of all measurable functions on \(\mathbb{R}\) such that \(\|h\|_{\infty} \leq 1/2\). If \(\|h\|_{\infty} \leq 1/2\), then
\[
\frac{1}{E[X]} \left|E[(X - E[X])h(X)]\right| \leq \frac{E|X - E[X]|}{2E[X]}
\]

Hence, from (11) and (10) we conclude that
\[
d_{TV}(X, X^s) \leq \frac{E|X - E[X]|}{2E[X]}.
\]
On the other hand, letting
\[ h(x) := \frac{1}{2}(1_{\{x>E[X]\}} - 1_{\{x\leq E[X]\}}) \]
in (10) we have \( h \in \mathcal{H} \) and obtain
\[ |E[h(X^*)] - E[h(X)]| = \frac{E[X - E[X]]}{2E[X]} \]  
proving (b). Note that we have
\[ d_K(X, X^*) = \sup_{t \geq 0} |P(X^* > t) - P(X > t)| = \sup_{t \geq 0} \left( P(X^* > t) - P(X > t) \right) \]
(14)
where \( g_t := 1_{(t, \infty)} \). If \( 0 \leq t < E[X] \) we obtain
\[ E[(X - E[X])1_{\{X>0\}}] = E[(X - E[X])1_{\{t<X\leq E[X]\}}] + E[(X - E[X])1_{\{X\leq E[X]\}}] \]
(15)
\[ \leq E[(X - E[X])1_{\{X>E[X]\}}] \].
Thus, by (10) from (14), (15) and (16) we conclude the claim of (a). Finally, if \( h \) is 1-Lipschitz continuous, then
\[ E[(X - E[X])h(X)] = E[(X - E[X])(h(X) - h(E[X]))] \]
(16)
\[ \leq \|h'\|_\infty E[|X - E[X]|^2] = \text{Var}(X). \]
On the other hand, the function \( h(x) := x - E[X] \) is 1-Lipschitz and
\[ E[(X - E[X])h(X)] = \text{Var}(X). \]
Thus, also (c) is proved.

3. Elements of Stein’s method

In this section we review some well-known and also some recent results about Stein’s method of normal approximation. Our general reference for this topic is the book [CGS11]. Throughout, \( Z \) will denote a standard normal random variable. Stein’s method originated from Stein’s seminal observation (see [Ste72]) that a real-valued random variable \( X \) has the standard normal distribution, if and only if for all, say, Lipschitz-continuous functions \( f \), the identity
\[ E[f'(X)] = E[Xf(X)] \]
(17)
holds. For a given random variable \( W \), which is supposed to be asymptotically normal, and a Borel-measurable test function \( h \) on \( \mathbb{R} \) with \( E|h(Z)| < \infty \) it was then Stein’s idea to solve the Stein equation
\[ f'(x) - xf(x) = h(x) - E[h(Z)] \]
(18)
and to use properties of the solution \( f \) and of \( W \) in order to bound the right hand side of
\[ |E[h(W')] - E[h(Z)]| = |E[f'(W) - Wf(W)]| \]
rather than bounding the left hand side directly. For $h$ as above, by $f_h$ we denote the standard solution to the Stein equation (18) which is given by

$$f_h(x) = e^{x^2/2} \int_{-\infty}^{x} (h(t) - E[h(Z)]) e^{-t^2/2} dt$$

$$= -e^{x^2/2} \int_{x}^{\infty} (h(t) - E[h(Z)]) e^{-t^2/2} dt.$$  (19)

Note that generally $f_h$ is only differentiable and satisfies (18) at the continuity points of $h$. In order to be able to deal with distributions which might have point masses, if $x \in \mathbb{R}$ is a point at which $f_h$ is not differentiable, one defines

$$f_h'(x) := xf_h(x) + h(x) - E[h(Z)]$$

such that, by definition, $f_h$ satisfies (18) at each point $x \in \mathbb{R}$. This gives a Borel-measurable version of the derivative of $f_h$ in the Lebesgue sense. Properties of the solutions $f_h$ for various classes of test functions $h$ have been studied. Since we are only interested in the Kolmogorov and Wasserstein distances $s$, we either suppose that $h$ is 1-Lipschitz or that $h = h_z = 1_{(-\infty,z]}$ for some $z \in \mathbb{R}$. In the latter case we write $f_z$ for $f_{h_z}$.

We need the following properties of the solutions $f_h$. If $h$ is 1-Lipschitz, then it is well known (see e.g. [CGS11]) that $f_h$ is continuously differentiable and that both $f_h$ and $f'_h$ are Lipschitz-continuous with

$$\|f_h\|_\infty \leq 1, \quad \|f'_h\|_\infty \leq \sqrt{\frac{2}{\pi}} \quad \text{and} \quad \|f''_h\|_\infty \leq 2.$$  (21)

Here, for a function $g$ on $\mathbb{R}$, we denote by

$$\|g'\|_\infty := \sup_{x \neq y} \frac{|g(x) - g(y)|}{|x - y|}$$

its minimum Lipschitz constant. Note that if $g$ is absolutely continuous, then $\|g'\|_\infty$ coincides with the essential supremum norm of the derivative of $g$ in the Lebesgue sense. Hence, the double use of the symbol $\|\cdot\|_\infty$ does not cause any problems. For an absolutely continuous function $g$ on $\mathbb{R}$, a fixed choice of its derivative $g'$ and for $x, y \in \mathbb{R}$ we let

$$R_g(x, y) := g(x + y) - g(x) - g'(x)y$$

(22)

denote the remainder term of its first order Taylor expansion around $x$ at the point $x + y$. If $h$ is 1-Lipschitz, then we obtain for all $x, y \in \mathbb{R}$ that

$$|R_{f_h}(x, y)| = |f_h(x + y) - f_h(x) - f'_h(x)y| \leq y^2.$$  (23)

This follows from (21) via

$$|f_h(x + y) - f_h(x) - f'_h(x)y| = \left| \int_{x}^{x+y} \left( f'_h(t) - f'_h(x) \right) dt \right|$$

$$\leq \|f''_h\|_\infty \int_{x}^{x+y} |t - x| dt = \frac{y^2 \|f''_h\|_\infty}{2} \leq y^2.$$
For \( h = h_z \) we list the following properties of \( f_z \): The function \( f_z \) has the representation
\[
    f_z(x) = \begin{cases} 
        \frac{(1-\Phi(x))\Phi(z)}{\varphi(x)}, & x \leq z \\
        \frac{\Phi(x)(1-\Phi(z))}{\varphi(x)}, & x > z.
    \end{cases}
\]

Here, \( \Phi \) denotes the standard normal distribution function and \( \varphi := \Phi' \) the corresponding continuous density. It is easy to see from (24) that \( f_z \) is infinitely often differentiable on \( \mathbb{R}\setminus\{z\} \). Furthermore, it is well-known that \( f_z \) is Lipschitz-continuous with Lipschitz constant 1 and that it satisfies
\[
    0 < f_z(x) \leq f_0(0) = \frac{\sqrt{2\pi}}{4}, \quad x, z \in \mathbb{R}.
\]

These properties already easily yield that for all \( x, u, v, z \in \mathbb{R} \)
\[
    |(x+u)f_z(x+u) - (x+v)f_z(x+v)| \leq \left( |x| + \frac{\sqrt{2\pi}π}{4} \right) (|u| + |v|).
\]

Proofs of the above mentioned classic facts about the functions \( f_z \) can again be found in [CCST1], for instance. As \( f_z \) is not differentiable at \( z \) (the right and left derivatives do exist but are not equal) by the above convention we define
\[
    f'_z(z) := z f_z(z) + 1 - \Phi(z)
\]
such that \( f = f_z \) satisfies (18) with \( h = h_z \) for all \( x \in \mathbb{R} \). Furthermore, with this definition, for all \( x, z \in \mathbb{R} \) we have
\[
    |f'_z(x)| \leq 1.
\]

The following quantitative version of the first order Taylor approximation of \( f_z \) has recently been proved by Lachièze-Rey and Peccati [LRP15] and had already been used implicitly in [ET14]. Applying Convention (27), for all \( x, u, z \in \mathbb{R} \) we have
\[
    |R_{f_z}(x,u)| = |f_z(x+u) - f_z(x) - f'_z(x)u|
\]
\[
\leq \frac{u^2}{2} \left( |x| + \frac{\sqrt{2\pi}π}{4} \right) + |u| \left( 1_{\{x<z\leq x+u\}} + 1_{\{x+u\leq z\leq x\}} \right)
\]
\[
\leq \frac{u^2}{2} \left( |x| + \frac{\sqrt{2\pi}π}{4} \right) + |u| 1_{\{z-(u\wedge 0)\leq x\leq z-(u\vee 0)\}},
\]
where, here and elsewhere, we write \( x \vee y := \max(x,y) \) and \( x \wedge y := \min(x,y) \).

For the proof of Theorems 2.5 and 2.6 we need to recall a certain coupling construction, which has been efficiently used in Stein’s method of normal approximation: Let \( X \) be a real-valued random variable such that \( E[X] = 0 \) and \( 0 < E[X^2] < \infty \). In [GR97] it was proved that there exists a unique distribution for a random variable \( X^* \) such that for all Lipschitz continuous functions \( f \) the identity
\[
    E[Xf(X)] = \text{Var}(X)E[f'(X^*)]
\]
holds true. The distribution of \( X^* \) is called the \emph{X-zero biased distribution} and the distributional transformation which maps \( \mathcal{L}(X) \) to \( \mathcal{L}(X^*) \) is called the \emph{zero bias transformation}. It can be shown that (30) holds for all absolutely continuous functions \( f \) on \( \mathbb{R} \) such that \( E|Xf(X)| < \infty \). From the Stein characterization of the
family of normal distributions it is immediate that the fixed points of the zero bias transformation are exactly the centered normal distributions. Thus, if, for a given $X$, the distribution of $X^*$ is close to that of $X$, the distribution of $X$ is approximately a fixed point of this transformation and, hence, should be close to the normal distribution with the same variance as $X$. In [Gol04] this heuristic was made precise by showing the inequality

$$d_{W}(W, \sigma Z) \leq 2d_{W}(W, W^*),$$

where $W$ is a mean zero random variable with $0 < \sigma^2 = E[W^2] = \text{Var}(W) < \infty$, $W^*$ having the $W$-zero biased distribution is defined on the same probability space as $W$ and $Z$ is standard normally distributed. For mere technical reasons we introduce a variant of the zero bias transformation for not necessarily centered random variables. Thus, if $X$ is a real random variable with $0 < E[X^2] < \infty$, we say that a random variable $X_{nz}$ has the $X$-non-zero biased distribution, if for all Lipschitz-continuous functions $f$ it holds that

$$E[(X - E[X])f(X)] = \text{Var}(X)E[f'(X_{nz})].$$

Existence and uniqueness of the $X$-non-zero biased distribution immediately follow from Theorem 2.1 of [GR05] (or Theorem 2.1 of [Döb13] by letting $B(x) = x - E[X]$, there). Alternatively, let $Y := X - E[X]$ and $Y^*$ have the $Y$-zero biased distribution, it is easy to see that $X_{nz} := Y^* + E[X]$ fulfills the requirements for the $X$-non-zero biased distribution. Most of the properties of the zero bias transformation have natural analogs for the non-zero bias transformation, so we do not list them all, here. Since an important part of the proof of our main result relies on the so-called single summand property, however, we state the result for the sake of reference.

**Lemma 3.1 (single summand property).** Let $X_1, \ldots, X_n$ be independent random variables such that $0 < E[X_j^2] < \infty$, $j = 1, \ldots, n$. Define $\sigma_j^2 := \text{Var}(X_j)$, $j = 1, \ldots, n$, $S := \sum_{j=1}^n X_j$ and $\sigma^2 := \text{Var}(S) = \sum_{j=1}^n \sigma_j^2$. For each $j = 1, \ldots, n$ let $X_j^{nz}$ have the $X_j$-non-zero biased distribution and be independent of $X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_n$ and let $I \in \{1, \ldots, n\}$ be a random index, independent of all the rest and such that

$$P(I = j) = \frac{\sigma_j^2}{\sigma^2}, \quad j = 1, \ldots, n.$$

Then, the random variable

$$S_{nz} := S - X_I + X_I^{nz} = \sum_{i=1}^n 1_{\{i = I\}} \left( \sum_{j \neq i} X_j + X_i^{nz} \right)$$

has the $S$-non-zero biased distribution.

**Proof.** The proof is either analogous to the proof of Lemma 2.1 in [GR97] or else, the statement could be deduced from this result in the following way: Using the fact that $X_{nz} = Y^* + E[X]$ has the $X$-non-zero biased distribution and only if $Y^*$ has the $(X - E[X])$-zero biased distribution, we let $Y_j := X_j - E[X_j]$, $Y^*_j := X_j^{nz} - E[X_j]$, $j = 1, \ldots, n$ and $W := \sum_{j=1}^n Y_j = S - E[S]$. Then, from Lemma 2.1 in [GR97] we
know that
\[
W^* := W - Y_I + Y^*_I = S - E[S] + \sum_{j=1}^{n} 1_{\{I=j\}} \left( E[X_j] - X_j + X_j^{nz} - E[X_j] \right)
\]
\[
= S - X_I + X_j^{nz} - E[S] = S^{nz} - E[S]
\]
has the $W$-zero biased distribution, implying that $S^{nz}$ has the $S$-non-zero biased distribution.

\[\Box\]

4. Proof of Theorems 2.5 and 2.6

From now on we let $h$ be either 1-Lipschitz or $h = h_z$ for some $z \in \mathbb{R}$ and write $f = f_h$ given by (24). Since $f$ is a solution to (18), plugging in $W$ and taking expectations yields
\[
E[h(W)] - E[h(Z)] = E[f'(W) - W f(W)]
\]
As usual in Stein’s method of normal approximation, the main task is to rewrite the term $E[W f(W)]$ into a more tractable expression by exploiting the structure of $W$ and using properties of $f$. From (5) we have
\[
E[W f(W)] = \frac{1}{\sigma} E[(S - aN) f(W)] + \frac{a}{\sigma} E[(N - \alpha) f(W)] =: T_1 + T_2.
\]
For ease of notation, for $n \in \mathbb{Z}_+$ and $M$ any $\mathbb{Z}_+$-valued random variable we let
\[
S_n := \sum_{j=1}^{n} X_j, \quad W_n := \frac{S_n - \alpha a}{\sigma}, \quad S_M := \sum_{j=1}^{M} X_j \quad \text{and} \quad W_M := \frac{S_M - \alpha a}{\sigma},
\]
such that, in particular, $S = S_N$ and $W = W_N$. Using the decomposition
\[
E[f'(W)] = \frac{a \gamma^2}{\sigma^2} E[f'(W)] + \frac{a^2 \gamma^2}{\sigma^2} E[f'(W)]
\]
which is true by virtue of (1), from (34) we have
\[
E[h(W)] - E[h(Z)] = E[f'(W)] - T_1 - T_2
\]
\[
= E\left[ \frac{c^2 \alpha}{\sigma^2} f'(W) - \frac{1}{\sigma}(S - aN) f(W) \right]
\]
\[
+ E\left[ \frac{a^2 \gamma^2}{\sigma^2} f'(W) - \frac{a}{\sigma}(N - \alpha) f(W) \right] =: E_1 + E_2.
\]
We will bound the terms $E_1$ and $E_2$ separately. Using independence of $N$ and $X_1, X_2, \ldots$ for $T_1$ we obtain:
\[
T_1 = \frac{1}{\sigma} \sum_{n=0}^{\infty} P(N = n) E[(S_n - na) f(W_n)]
\]
\[
= \frac{1}{\sigma} \sum_{n=0}^{\infty} P(N = n) E[(S_n - na) g(S_n)],
\]
where
\[
g(x) := f\left( \frac{x - \alpha a}{\sigma} \right).
\]
Thus, if, for each $n \geq 0$, $S_n^{nz}$ has the $S_n$-non-zero biased distribution, from (37) and (32) we obtain that

$$T_1 = \frac{1}{\sigma} \sum_{n=0}^{\infty} P(N = n) \text{Var}(S_n)E \left[ g'(S_n^{nz}) \right]$$

$$= \frac{c^2}{\sigma^2} \sum_{n=0}^{\infty} nP(N = n)E \left[ f'(\frac{S_n^{nz} - \alpha a}{\sigma}) \right].$$

(38)

Note that if we let $M$ be independent of $S_1^{nz}, S_2^{nz}, \ldots$ and have the $N$-size biased distribution, then, from (38) we obtain

$$T_1 = \frac{c^2\alpha}{\sigma^2} E \left[ f'(\frac{S_M - \alpha a}{\sigma}) \right],$$

where

$$S_M^{nz} = \sum_{n=1}^{\infty} 1_{(M=n)} S_n^{nz}.$$  

(39)

We use Lemma 3.1 for the construction of the variables $S_n^{nz}, n \in \mathbb{N}$. Note, however, that by the i.i.d. property of the $X_j$ we actually do not need the mixing index $I$, here. Hence, we construct independent random variables

$$(N, M), X_1, X_2, \ldots$$

and $Y$
such that $M$ has the $N$-size biased distribution and such that $Y$ has the $X_1$-non-zero biased distribution. Then, for all $n \in \mathbb{N}$

$$S_n^{nz} := S_n - X_1 + Y$$

has the $S_n$-non-zero biased distribution and we have

$$\frac{S_M^{nz} - \alpha a}{\sigma} = \frac{S_M - \alpha a}{\sigma} + \frac{Y - X_1}{\sigma} = W_M + \frac{Y - X_1}{\sigma} =: W^*.$$  

(40)

Thus, from (40) and (39) we conclude that

$$T_1 = \frac{c^2\alpha}{\sigma^2} E \left[ f'(W^*) \right]$$

and

$$E_1 = \frac{c^2\alpha}{\sigma^2} E \left[ f'(W) - f'(W^*) \right].$$

(41)

We would like to mention that if $a = 0$, then $W^*$ has the $W$-zero biased distribution as $T_2 = 0$ and $\sigma^2 = c^2\alpha$ in this case. Before addressing $T_2$, we remark that the random variables appearing in $E_1$ and $E_2$, respectively, could possibly be defined on different probability spaces, if convenient, since they do not appear under the same expectation sign. Indeed, for $E_2$ we use the coupling $(N, N^s)$, which is given in the statements of Theorems 2.5 and 2.6 and which appears in the bounds via the difference $D = N^s - N$. In order to manipulate $E_2$ we thus assume that the random variables

$$(N, N^s), X_1, X_2, \ldots$$

are independent and that $N^s$ has the $N$-size biased distribution. Note that we do not assume here that $D = N^s - N \geq 0$, since sometimes a natural coupling yielding
a small value of $|D|$ does not satisfy this nonnegativity condition. In what follows we will use the notation

\begin{equation}
V := W_{N^*} - W_N = \frac{1}{\sigma} (S_{N^*} - S_N) \quad \text{and} \quad J := 1_{\{D \geq 0\}} = 1_{\{N^* \geq N\}}.
\end{equation}

Now we turn to rewriting $T_2$. Using independence of $N$ and $X_1, X_2, \ldots$, and of $N^*$ and $X_1, X_2, \ldots$, respectively, $E[N] = \alpha$ and the defining equation (2) of the $N$-size biased distribution, we obtain from (34) that

\begin{align*}
T_2 &= \frac{a}{\sigma} E[(N - \alpha) f(W_N)] = \frac{\alpha a}{\sigma} E[f(W_{N^*}) - f(W_N)] \\
&= \frac{\alpha a}{\sigma} E[1_{\{N^* \geq N\}} (f(W_{N^*}) - f(W_N))] + \frac{\alpha a}{\sigma} E[1_{\{N^* < N\}} (f(W_{N^*}) - f(W_N))] \\
&= \frac{\alpha a}{\sigma} E[J (f(W_N + V) - f(W_N))] - \frac{\alpha a}{\sigma} E[(1 - J) (f(W_{N^*} - V) - f(W_{N^*}))] \\
&= \frac{\alpha a}{\sigma} E[J V f'(W_N)] + \frac{\alpha a}{\sigma} E[J R_f(W_N, V)] \\
&\quad + \frac{\alpha a}{\sigma} E[(1 - J) V f'(W_{N^*})] - \frac{\alpha a}{\sigma} E[(1 - J) R_f(W_{N^*}, -V)],
\end{align*}

where $R_f$ was defined in (22). Note that we have

\begin{align*}
JV = 1_{\{N^* \geq N\}} \frac{1}{\sigma} \sum_{j=N+1}^{N^*} X_j \quad \text{and} \quad W_N = \sum_{j=1}^{N} X_j
\end{align*}

and, hence, the random variables $JV$ and $W_N$ are conditionally independent given $N$. Noting also that

\begin{align*}
E[JV \mid N] &= \frac{1}{\sigma} E[J \sum_{j=N+1}^{N^*} X_j \mid N] \\
&= \frac{1}{\sigma} E[J E\left[ \sum_{j=N+1}^{N^*} X_j \mid N, N^* \right] \mid N] \\
&= \frac{a}{\sigma} E[JD \mid N] = \frac{a}{\sigma} E[JD \mid N]
\end{align*}

we obtain that

\begin{align*}
\frac{\alpha a}{\sigma} E[JV f'(W_N)] &= \frac{\alpha a}{\sigma} E[E[JV \mid N] E[f'(W_N) \mid N]] \\
&= \frac{\alpha a^2}{\sigma^2} E[E[JD \mid N] E[f'(W_N) \mid N]] \\
&= \frac{\alpha a^2}{\sigma^2} E[E[J f'(W_N) \mid N]] \\
&= \frac{\alpha a^2}{\sigma^2} E[J f'(W_N)],
\end{align*}

where we have used for the next to last equality that also $D$ and $W_N$ are conditionally independent given $N$. In a similar fashion, using that $W_{N^*}$ and $1_{\{D < 0\}} V$ and also $W_{N^*}$ and $D$ are conditionally independent given $N^*$, one can show

\begin{align*}
\frac{\alpha a}{\sigma} E[(1 - J)V f'(W_{N^*})] &= \frac{\alpha a^2}{\sigma^2} E[(1 - J) D f'(W_{N^*})] .
\end{align*}
Hence, using
\[
\frac{\alpha a^2}{\sigma^2} E[D] = \frac{\alpha a^2 \gamma^2}{\sigma^2} = \frac{\alpha^2 \gamma^2}{\sigma^2}
\]
which follows from (6), from (36), (44), (45) and (46) we obtain
\[
E_2 = \frac{\alpha a^2}{\sigma^2} E \left[ (E[D] - D) f'(W_N) \right] + \frac{\alpha a^2}{\sigma^2} E \left[ (1 - J) D \left( f'(W_N) - f'(W_{N^s}) \right) \right]
\]
\[
- \frac{\alpha a}{\sigma} E \left[ J R_f(W_N, V) \right] + \frac{\alpha a}{\sigma} E \left[ (1 - J) R_f(W_{N^s}, -V) \right]
\]
(47) =: E_{2,1} + E_{2,2} + E_{2,3} + E_{2,4}.

Using conditional independence of $D$ and $W_N$ given $N$ as well as the Cauchy-Schwartz inequality, we can estimate
\[
|E_{2,1}| = \frac{\alpha a^2}{\sigma^2} E \left[ E[D - E[D] \mid N] E[f'(W_N) \mid N] \right]
\]
\[
\leq \frac{\alpha a^2}{\sigma^2} \sqrt{\text{Var}(E[D \mid N])} \sqrt{E \left[ (E[f'(W_N) \mid N])^2 \right]}
\]
\[
\leq \frac{\alpha a^2}{\sigma^2} \|f'\|_\infty \sqrt{\text{Var}(E[D \mid N])}.
\]
(48)

Now we will proceed by first assuming that $h$ is a 1-Lipschitz function. In this case, we choose the coupling $(M, N)$ used for $E_1$ in such a way that $M \geq N$. By Remark 2.2(a) such a construction of $(M, N)$ is always possible e.g. via the quantile transformation and that it achieves the Wasserstein distance, i.e.
\[
E|M - N| = E[M - N] = \frac{E[N^2]}{E[N]} - E[N] = \frac{\text{Var}(N)}{E[N]} = \frac{\gamma^2}{\alpha} = d_W(N, N^*).
\]

In order to bound $E_1$, we first derive an estimate for $E|W_M - W_N|$. We have
\[
E[|W_M - W_N| \mid N, M] = \frac{1}{\sigma} E[|S_M - S_N| \mid N, M] \leq \frac{|M - N|}{\sigma} E|X_1|
\]
(49)
and, hence,
\[
E|W_M - W_N| = E \left[ E[|W_M - W_N| \mid N, M] \right] = \frac{1}{\sigma} E[|S_M - S_N| \mid N, M]
\]
\[
\leq \frac{b}{\sigma} E[M - N] = \frac{b \gamma^2}{\sigma \alpha}.
\]
(50)

Then, using (21), (50) as well as the fact that the $X_j$ are i.i.d., for $E_1$ we obtain that
\[
|E_1| = \frac{c^2 \alpha}{\sigma^2} E \left[ f'(W_N) - f'\left( W_M + \frac{Y - X_1}{\sigma} \right) \right]
\]
\[
\leq \frac{2c^2 \alpha}{\sigma^2} \left( E|W_N - W_M| + \sigma^{-1} E|Y - X_1| \right)
\]
\[
\leq \frac{2c^2 \alpha}{\sigma^3} \left( \frac{b \gamma^2}{\alpha} + \frac{3}{2c^2} E|X_1 - E[X_1]|^3 \right)
\]
\[
= \frac{2c^2 \gamma^2}{\sigma^3} + \frac{3 \alpha \xi}{\sigma^3}.
\]
(52)
Similarly to (49) we obtain yielding
\[ Y \] the fact that which follows from an analogous one in the zero-bias framework (see [CGS11]) via the fact that \( Y - E[X_1] \) has the \( (X - E[X_1]) \)-zero biased distribution. Similarly to (49) we obtain
\[ E\left[ \left| V \right| \mid N, N^* \right] = E\left[ \left| W_{N^*} - W_N \right| \mid N, N^* \right] \leq \frac{2|b^2 - N|}{\sigma} = \frac{b|D|}{\sigma} \]
which, together with (21) yields that
\[ |E_{2,2}| = \frac{\alpha a^2}{\sigma^2} E\left[ (1 - J)D(f'(W_N) - f'(W_{N^*})) \right] \leq \frac{2\alpha a^2}{\sigma^2} E\left[ (1 - J)D(W_N - W_{N^*}) \right] \]
\[ = \frac{2\alpha a^2}{\sigma^2} E\left[ (1 - J)|D| E\left[ |V| \mid N, N^* \right] \right] \leq \frac{2\alpha a^2 b}{\sigma^3} E\left[ (1 - J)D^2 \right]. \]
We conclude the proof of the Wasserstein bounds by estimating \( E_{2,3} \) and \( E_{2,4} \). Note that by (23) we have
\[ |R_f(W_N, V)| \leq V^2 \quad \text{and} \quad |R_f(W_{N^*}, -V)| \leq V^2 \]
yielding
\[ |E_{2,3}| + |E_{2,4}| \leq \frac{\alpha|a|}{\sigma} E\left[ (1_{D \geq 0} + 1_{D < 0})V^2 \right] = \frac{\alpha|a|}{\sigma} E[V^2]. \]
Observe that
\[ E[V^2] = \frac{1}{\sigma^2} E\left[ (S_{N^*} - S_N)^2 \right] \]
\[ = \frac{1}{\sigma^2} \left( \text{Var}(S_{N^*} - S_N) + E[S_{N^*} - S_N]^2 \right) \]
and
\[ E[S_{N^*} - S_N] = E\left[ E[S_{N^*} - S_N \mid N, N^*] \right] = aE[D] = \frac{a \gamma^2}{\alpha}. \]
Further, from the variance decomposition formula we obtain
\[ \text{Var}(S_{N^*} - S_N) = E[\text{Var}(S_{N^*} - S_N \mid N, N^*)] + \text{Var}(E[S_{N^*} - S_N \mid N, N^*]) \]
\[ = E[c^2 |D|] + \text{Var}(aD) = c^2 E[D] + a^2 \text{Var}(D). \]
This together with (57) and (58) yields the bounds
\[ E[V^2] = E\left[ (W_{N^*} - W_N)^2 \right] = \frac{1}{\sigma^2} \left( c^2 E|D| + a^2 E[D^2] \right) \]
\[ \leq \frac{b^2}{\sigma^2} E[D^2], \]
where we have used the fact that \( D^2 \geq |D| \) and \( a^2 + c^2 = b^2 \) to obtain
\[ c^2 E|D| + a^2 E[D^2] \leq b^2 E[D^2]. \]
The asserted bound on the Wasserstein distance between \( W \) and \( Z \) from Theorem 2.5 now follows from (21), (36), (52), (55), (56) and (61).
If $a = 0$, then $E_1$ can be bounded more accurately than we did before. Indeed, using (60) with $N^s = M$ and applying the Cauchy-Schwartz inequality give

$$E |W_M - W_N| \leq \sqrt{E[(W_M - W_N)^2]} = \frac{c}{\sigma} \sqrt{E[M - N]} = \frac{c\gamma}{\sqrt{\alpha \sigma}},$$

as $c = b$ in this case. Plugging this into (51), we obtain

$$|E_1| \leq \frac{2c^2 \alpha}{\sigma^2} \left( E|W_M - W_N| + \sigma^{-1} E|Y - X_1| \right)$$
$$\leq \frac{2c^3 \gamma \sqrt{\alpha}}{\sigma^3} + \frac{2c^2 \alpha}{\sigma^3} E|Y - X_1|$$
$$\leq \frac{2c^3 \gamma \sqrt{\alpha}}{c^3 \alpha^{3/2}} + \frac{3\alpha \xi}{c^3 \alpha^{3/2}}$$
$$= \frac{2\gamma}{\alpha} + \frac{3\xi}{c^3 \sqrt{\alpha}},$$

which is the Wasserstein bound claimed in Theorem 2.6.

Next, we proceed to the proof of the Berry-Esseen bounds in Theorems 2.5 and 2.6. Bounding the quantities $E_1, E_{2,2}, E_{2,3}$ and $E_{2,4}$ in the case that $h = h_z$ is much more technically involved. Also, in this case we do not in general profit from choosing $M$ appearing in $T_1$ in such a way that $M \geq N$. This is why we let $M = N^s$ for the proof of the Kolmogorov bound in Theorem 2.5. Only for the proof of Theorem 2.6 we will later assume that $M \geq N$. We will write $f = f_z$ and introduce the notation

$$\tilde{V} := W^* - W = W_{N^s} + \sigma^{-1}(Y - X_1) - W_N = V + \sigma^{-1}(Y - X_1).$$

From (12) and the fact that $f$ solves the Stein equation (18) for $h = h_z$ we have

$$E_1 = \frac{c^2 \alpha}{\sigma^2} E[f'(W) - f'(W^*)]$$
$$= \frac{c^2 \alpha}{\sigma^2} E[W f(W) - W^* f(W^*)] + \frac{c^2 \alpha}{\sigma^2} (P(W \leq z) - P(W^* \leq z))$$
$$=: E_{1,1} + E_{1,2}. \tag{62}$$

In order to bound $E_{1,1}$ we apply (26) to obtain

$$|E_{1,1}| \leq \frac{c^2 \alpha}{\sigma^2} E \left[ \tilde{V} \left( \frac{\sqrt{2\pi}}{4} + |W| \right) \right]. \tag{63}$$

Using (60), (61) and (63) we have

$$E|\tilde{V}| \leq E|V| + \sigma^{-1} E|Y - X_1| \leq \sqrt{E[V^2]} + \frac{3\xi}{2\sigma c^2}$$
$$= \frac{1}{\sigma} \sqrt{c^2 E|D| + a^2 E[D^2]} + \frac{3\xi}{2\sigma c^2} \tag{64}$$
$$\leq \frac{b}{\sigma} \sqrt{E[D^2]} + \frac{3\xi}{2\sigma c^2}. \tag{65}$$
Furthermore, using independence of $W$ and $Y$, we have

$$E|(Y - X_1)W| \leq E|(Y - E[X_1])W| + E|(X_1 - E[X_1])W|$$

$$= E|Y - E[X_1]|E|W| + E|(X_1 - E[X_1])W|$$

$$\leq \frac{\xi}{2c^2} \sqrt{E[W^2]} + \sqrt{\text{Var}(X_1)E[W^2]} = \frac{\xi}{2c^2} + c.$$  
(66)

Finally, we have

$$E|VW| \leq \sqrt{E[V^2]} \sqrt{E[W^2]} = \frac{1}{\sigma} \sqrt{c^2 E[D] + a^2 E[D^2]}$$  
(67)

$$\leq \frac{b}{\sigma} \sqrt{E[D^2]}.$$  
(68)

From (63), (64), (65), (66), (67) and (68) we conclude that

$$|E_{1,1}| \leq \frac{c^2 \alpha}{\sigma^2} \left( \frac{\sqrt{2\pi}}{4\sigma} \sqrt{c^2 E[D] + a^2 E[D^2]} + \frac{3\xi \sqrt{2\pi}}{8c^2 \sigma} + \frac{\xi}{2c^2 \sigma} + \frac{c}{\sigma} \right)$$

$$+ \frac{1}{\sigma} \sqrt{c^2 E[D] + a^2 E[D^2]}$$

$$= \frac{c^2 \alpha (\sqrt{2\pi} + 4)}{4\sigma^3} \sqrt{c^2 E[D] + a^2 E[D^2]} + \frac{\xi \alpha (3\sqrt{2\pi} + 4)}{8\sigma^3} + \frac{c^3 \alpha}{\sigma^3}$$

$$\leq \left( \frac{\sqrt{2\pi} + 4}{4\sigma^3} \right) b c^2 \alpha \sqrt{E[D^2]} + \frac{\xi \alpha (3\sqrt{2\pi} + 4)}{8\sigma^3} + \frac{c^3 \alpha}{\sigma^3} =: B_1.$$  
(69)

In order to bound $E_{1,2}$ we need the following lemma, which will be proved in Section 5. In the following we denote by $C_K$ the Berry-Esseen constant for sums of i.i.d. random variables with finite third moment. It is known from [She11] that

$$C_K \leq 0.4748.$$  

In particular, $2C_K \leq 1$, which is substituted for $2C_K$ in the statements of Theorems 2.5 and 2.6. However, we prefer keeping the dependence of the bounds on $C_K$ explicit within the proof.

**Lemma 4.1.** With the above assumptions and notation we have for all $z \in \mathbb{R}$

$$|P(W^* \leq z) - P(W_{N^*} \leq z)| \leq \frac{1}{\sqrt{\alpha}} \left( \frac{7}{2} \sqrt{2} + 2 \right) \frac{\xi}{c^3} \text{ and }$$

$$|P(W_{N^*} \leq z) - P(W \leq z)| \leq P(N = 0) + \frac{b}{c \sqrt{2\pi}} E[D1_{(D \geq 0)}N^{-1/2}1_{(N \geq 1)}]$$

$$+ \frac{2C_K \xi}{c^3} E[1_{(D \geq 0)}N^{-1/2}1_{(N \geq 1)}]$$

$$+ \frac{1}{\sqrt{\alpha}} \left( \frac{b}{c \sqrt{2\pi}} \sqrt{E[D^21_{(D < 0)}]} + \frac{2C_K \xi}{c^3} \sqrt{P(D < 0)} \right).$$  
(72)
If $a = 0$ and $D \geq 0$, then for all $z \in \mathbb{R}$

$$|P(W_{N^*} \leq z) - P(W \leq z)| \leq P(N = 0) + \frac{2C_K \xi}{c^3} E[N^{-1/2}1_{\{N \geq 1\}}] + \frac{1}{\sqrt{2\pi}} E[\sqrt{D}N^{-1/2}1_{\{N \geq 1\}}]$$

(73)

$$\leq P(N = 0) + \left(\frac{2C_K \xi}{c^3} + \frac{\gamma}{\sqrt{\alpha\sqrt{2\pi}}}\right) \sqrt{E[1_{\{N \geq 1\}}]} N^{-1}\] .$$

(74)

Applying the triangle inequality to Lemma 4.1 yields the following bounds on $E_{1,2}$:

In the general situation of Theorem 2.3, we have

$$|E_{1,2}| \leq \left(\frac{7}{2}\sqrt{2} + 2\right) \frac{\sqrt{\alpha\xi}}{c^3\sigma^2} + \frac{2C_K \xi}{c^3} P(N = 0) + \frac{\alpha bc}{\sigma^2\sqrt{2\pi}} E[D1_{\{D \geq 0\}}N^{-1/2}1_{\{N \geq 1\}}] + \frac{2C_K \xi}{c^3} E[1_{\{D \geq 0\}}N^{-1/2}1_{\{N \geq 1\}}]$$

(75)

$$+ \frac{\alpha bc}{\sigma^2\sqrt{2\pi}} \sqrt{P(D < 0) =: B_2} .$$

If $a = 0$ and $D \geq 0$, then, keeping in mind that $\sigma^2 = \alpha c^2$ in this case,

$$|E_{1,2}| \leq \left(\frac{7}{2}\sqrt{2} + 2\right) \frac{\xi}{c^3\sqrt{\alpha}} + P(N = 0) + \frac{2C_K \xi}{c^3} E[N^{-1/2}1_{\{N \geq 1\}}] + \frac{1}{\sqrt{2\pi}} E[\sqrt{D}N^{-1/2}1_{\{N \geq 1\}}]$$

(76)

$$\leq \left(\frac{7}{2}\sqrt{2} + 2\right) \frac{\xi}{c^3\sqrt{\alpha}} + P(N = 0) + \left(\frac{2C_K \xi}{c^3} + \frac{\gamma}{\sqrt{\alpha\sqrt{2\pi}}}\right) \sqrt{E[1_{\{N \geq 1\}}]} N^{-1}\] .$$

(77)

The following lemma, which is also proved in Section 5, will be needed to bound the quantities $E_{2,2}$, $E_{2,3}$ and $E_{2,4}$ from (17).

**Lemma 4.2.** With the above assumptions and notation we have

$$E[J|V|1_{\{z-(\forall \leq 0)W_{N^*} \leq z-(\forall \leq 0)\}}] \leq \frac{b}{\sigma} \sqrt{P(N = 0)} \sqrt{E[J D^2]}$$

(78)

$$+ \frac{b^2}{c \sigma \sqrt{2\pi}} E[J D^2 1_{\{N \geq 1\}} N^{-1/2}] + \frac{2C_K \xi b}{c^3 \sigma} E[J D 1_{\{N \geq 1\}} N^{-1/2}] ,$$

$$E[(1 - J)|V|1_{\{z+(\forall \leq 0)W_{N^*} \leq z+(\forall \leq 0)\}}] \leq \frac{b^2}{c \sigma \sqrt{2\pi}} E[(1 - J) D^2 (N^s)^{-1/2}]$$

(79)

$$+ \frac{2C_K b \xi}{c^3 \sigma \sqrt{\alpha}} \sqrt{E[(1 - J) D^2]}$$

and

$$E[(1 - J)|D|1_{\{z+(\forall \leq 0)W_{N^*} \leq z+(\forall \leq 0)\}}] \leq \frac{b}{c \sqrt{2\pi}} E[(1 - J) D^2 (N^s)^{-1/2}]$$

(80)

$$+ \frac{2C_K \xi}{c^3 \sqrt{\alpha}} \sqrt{E[(1 - J) D^2]} .$$
Next, we derive a bound on $E_{2,2}$. Since $f$ solves the Stein equation \((18)\) for $h = h_z$ we have

\[
E_{2,2} = \frac{\alpha a^2}{\sigma^2} \mathbb{E}[(1 - J)D(W_N f(W_N) - W_{N^*} f(W_{N^*}))]
\]

\[
+ \frac{\alpha a^2}{\sigma^2} \mathbb{E}[(1 - J)D(1_{\{W_N \leq z\}} - 1_{\{W_{N^*} \leq z\}})]
\]

\[=: E_{2,2,1} + E_{2,2,2}. \tag{81} \]

Using

\[
28 \, \text{CHRISTIAN DÖBLER}
\]

\[W_N = W_{N^*} - V \]

and Lemma 5.1 we obtain from \((80)\) that

\[
|E_{2,2,2}| \leq \frac{\alpha a^2}{\sigma^2} \mathbb{E}[1_{\{D < 0\}}|D|1_{\{z^* < y \leq z^* + (V \vee 0)\}}]
\]

\[
\leq \frac{\alpha a^2 b}{\sigma^2 c \sqrt{2\pi}} \mathbb{E}[1_{\{D < 0\}}D^2(N^*)^{-1/2}] + \frac{2C_k \xi a^2 \sqrt{\alpha}}{c^3 \sigma^2} \sqrt{\mathbb{E}[1_{\{D < 0\}} D^2]} \tag{82} \]

\[=: B_4. \]

As to $E_{2,2,1}$, from \((26)\) we have

\[
|E_{2,2,1}| \leq \frac{\alpha a^2}{\sigma^2} \mathbb{E}[(1 - J)|DV|(|W_{N^*}| + \frac{\sqrt{2\pi}}{4})] \tag{83} \]

As

\[
\mathbb{E}[|V| \mid N, N^*] \leq \sqrt{\mathbb{E}[V^2 \mid N, N^*]} = \frac{1}{\sigma} \sqrt{c^2|D| + a^2 D^2} \leq \frac{b}{\sigma}|D|, \tag{84} \]

by conditioning, we see

\[
\mathbb{E}[(1 - J)|DV|] = \mathbb{E}[(1 - J)|D|\mathbb{E}[|V| \mid N, N^*]] \leq \frac{b}{\sigma} \mathbb{E}[(1 - J)D^2]. \tag{85} \]

Now, using the fact that conditionally on $N^*$, the random variables $W_{N^*}$ and $(1 - J)|DV|$ are independent, as well as the Cauchy-Schwartz inequality, we conclude that

\[
\mathbb{E}[(1 - J)DVW_{N^*}] = \mathbb{E}[\mathbb{E}[(1 - J)DVW_{N^*} \mid N^*]]
\]

\[
= \mathbb{E}[\mathbb{E}[(1 - J)DV \mid N^*]\mathbb{E}[|W_{N^*}| \mid N^*]]
\]

\[
\leq \sqrt{\mathbb{E}[(\mathbb{E}[(1 - J)DV \mid N^*])^2]} \sqrt{\mathbb{E}[(\mathbb{E}[|W_{N^*}| \mid N^*])^2]}
\]

\[\leq \frac{b}{\sigma} \sqrt{\mathbb{E}[(\mathbb{E}[(1 - J)D^2 \mid N^*])^2]} \sqrt{\mathbb{E}[W_{N^*}^2]}, \tag{86} \]

where we have used the conditional Jensen inequality, \((84)\) and

\[
\mathbb{E}[(1 - J)|DV| \mid N^*] = \mathbb{E}[(1 - J)|D|\mathbb{E}[|V| \mid N, N^*] \mid N^*]
\]

to obtain the last inequality. Using the defining relation \((2)\) of the size-biased distribution one can easily show that

\[
\mathbb{E}[W_{N^*}^2] = \frac{1}{\sigma^2} \mathbb{E}[c^2 N^* + a^2 (N^* - \alpha)^2] = \frac{c^2 \beta^2 + a^2 (\delta^3 - 2\alpha \beta^2 + \alpha^3)}{\alpha \sigma^2}, \tag{87} \]
Thus, from (92), (93) and (94) we see that
\[ |E_{2,2,1}| \leq \frac{\alpha^2 b \sqrt{2\pi}}{4\sigma^3} E[1_{\{D < 0\}} D^2] + a^2 b \frac{c^2 \beta^2 + a^2 (\delta^2 - 2\alpha \beta^2 + \alpha^3)}{\sigma^5} \sqrt{E[(E[1_{\{D < 0\}} D^2 \mid N^s])^2]} \]
\[ =: B_5. \]

It remains to bound the quantities \( E_{2,3} \) and \( E_{2,4} \) from (47) for \( f = f_z \). From (23) we have
\[ |E_{2,3}| = \frac{\alpha|\alpha|}{\sigma} |E[1_{\{D \geq 0\}} R_f(W, V)]| \leq \frac{\alpha|\alpha|}{2\sigma} E[J V^2 \left(|W| + \frac{\sqrt{2\pi}}{4}\right)] + \frac{\alpha|\alpha|}{\sigma} E[J |V| 1_{\{z-(V \lor 0) < W \leq z-(V \land 0)\}}] =: R_{1,1} + R_{1,2}. \]
Similarly to (60) we obtain
\[ E[J V^2] = \frac{1}{\sigma^2} (c^2 E[J D] + a^2 E[J D^2]) \leq \frac{b^2}{\sigma} E[J D^2] \]
from
\[ E[J V^2 \mid N, N^s] = J E[V^2 \mid N, N^s] = \frac{J}{\sigma^2} (c^2 |D| + a^2 D^2) \]
\[ = \frac{1}{\sigma^2} (c^2 JD + a^2 JD^2). \]

Also, observe that the random variables
\[ JV^2 = \sigma^{-1} 1_{\{N^s \geq N\}} \left( \sum_{j=N+1}^{N^s} X_j \right)^2 \quad \text{and} \quad W_N = \sigma^{-1} \left( \sum_{j=1}^{N} X_j - \alpha a \right) \]
are conditionally independent given \( N \). Hence, using the Cauchy-Schwartz inequality
\[ E[J V^2 \mid W_N] = E\left[E[J V^2 \mid W_N \mid N]\right] = E\left[E[J V^2 \mid N] E[|W_N| \mid N]\right] \leq \sqrt{E\left[(E[J V^2 \mid N])^2\right]} \sqrt{E\left[(E[|W_N| \mid N])^2\right]} . \]
From (91) and \( D^2 \geq |D| \) we conclude that
\[ E[J V^2 \mid N] = \frac{1}{\sigma^2} (c^2 E[J D \mid N] + a^2 E[J D^2 \mid N]) \leq \frac{b^2}{\sigma^2} E[J D^2 \mid N]. \]
Furthermore, by the conditional version of Jensen’s inequality we have
\[ E\left(E[|W_N| \mid N]\right)^2 \leq E\left[E[|W_N| \mid N]\right] = E[W_N^2] = 1. \]

Thus, from (92), (93) and (94) we see that
\[ E[J V^2 \mid W_N] \leq \frac{b^2}{\sigma^2} \sqrt{E\left[(E[J D^2 \mid N])^2\right]} . \]
Hence, (89), (90) and (95) yield

\[ R_{1,1} \leq \frac{\alpha |a| b^2}{2 \sigma^3} \sqrt{E \left[ \left( E [JD^2 \mid N] \right)^2 \right]} + \frac{\alpha |a| b^2 \sqrt{2\pi}}{8\sigma^3} E [JD^2]. \]  

Finally, from (89), (96) and (78) we get

\[ |E_{2,3}| \leq \frac{\alpha |a| b^2}{2\sigma^3} \sqrt{E \left[ \left( E [1_{\{D \geq 0\}} D^2 \mid N] \right)^2 \right]} + \frac{\alpha |a| b^2 \sqrt{2\pi}}{8\sigma^3} E [1_{\{D \geq 0\}} D^2] \]
\[ \quad + \frac{2\sigma \xi a |b|}{\sigma^2 \sqrt{2\pi}} E \left[ 1_{\{D \geq 0\}} D 1_{\{N \geq 1\}} N^{-1/2} \right] \]
\[ \quad + \frac{\alpha |a| b^2}{\sigma^2 \sqrt{2\pi}} E [1_{\{D \geq 0\}} D 1_{\{N \geq 1\}} N^{-1/2}] \]
\[ =: B_6. \]

Similarly, we have

\[ |E_{2,4}| = \frac{\alpha |a|}{\sigma} \left| E \left[ 1_{\{D < 0\}} R_f(W_{N^*}, -V) \right] \right| \]
\[ \leq \frac{\alpha |a|}{2\sigma} E \left[ (1 - J)V^2 \left( W_{N^*} + \frac{\sqrt{2\pi}}{4} \right) \right] \]
\[ + \frac{\alpha |a|}{\sigma} E \left[ (1 - J)|V| 1_{\{z+(V \land 0) < W_{N^*} \leq z+(V \lor 0)\}} \right] =: R_{2,1} + R_{2,2}. \]

Analogously to the above we obtain

\[ E \left[ (1 - J)V^2 \right] = \frac{1}{\sigma^2} \left( c^2 E \left[ (1 - J)|D| \right] + a^2 E \left[ (1 - J)D^2 \right] \right) \]
\[ \leq \frac{b^2}{\sigma^2} E \left[ (1 - J)D^2 \right] \quad \text{and} \]
\[ E \left[ (1 - J)V^2 \mid N^* \right] = \frac{1}{\sigma^2} \left( c^2 E \left[ (1 - J)|D| \mid N^* \right] + a^2 E \left[ (1 - J)D^2 \mid N^* \right] \right) \]
\[ \leq \frac{b^2}{\sigma^2} E \left[ (1 - J)D^2 \mid N^* \right]. \]

Additionally, using conditional independence of \((1 - J)V^2\) and \(W_{N^*}\) given \(N^*\), one has

\[ E \left[ (1 - J)V^2 \mid W_{N^*} \right] \leq \frac{b^2}{\sigma^2} \sqrt{E \left[ \left( E \left[ (1 - J)D^2 \mid N^* \right] \right)^2 \right]} \sqrt{E \left[ W_{N^*}^2 \right]} \]

Combining (87) and (101) we obtain

\[ E \left[ (1 - J)V^2 \mid W_{N^*} \right] \leq \frac{b^2}{\sigma^2} \sqrt{E \left[ \left( E \left[ (1 - J)D^2 \mid N^* \right] \right)^2 \right]} \left( \frac{c^2 \beta^2 + a^2 (\delta^3 - 2\alpha \beta^2 + \alpha^3)}{\alpha \sigma^2} \right)^{1/2}. \]
Thus, from (98), (99), (101) and (79) we conclude

\[
|E_2| \leq \frac{\alpha|a|^2\sqrt{2\pi}}{8\sigma^3}E[1_{\{D<0\}}D^2] + \frac{\alpha|a|^2}{2\sigma^3}\sqrt{E\left(\left(E[1_{\{D<0\}}D^2 | N^*]\right)^2\right)}
\]

\[
\cdot \left(\frac{c^2\beta^2 + a^2(\delta^3 - 2\alpha\beta^2 + \alpha^2)}{\alpha\sigma^2}\right)^{1/2}
\]

\[
+ \frac{\alpha|a|^2}{\sigma^2\sqrt{2\pi}}E[1_{\{D<0\}}D^2(N^*)^{-1/2}] + \frac{\sqrt{\alpha|a|2C_Kb\xi}}{\sigma^2}\sqrt{E[1_{\{D<0\}}D^2]}
\]

(103)

\[=: B_7.\]

The Berry-Esseen bound stated in Theorem 2.5 follows from (36), (62), (70), (75), (47), (48), (28), (81), (82), (88), (97) and (103). In order to obtain the Kolmogorov bound in Theorem 2.6 again, we choose \( M \) such that \( M \geq N \) and use the bounds (69) and (77) instead. The result then follows from (36) and (62).

5. PROOFS OF AUXILIARY RESULTS

Here, we give several rather technical proofs. We start with the following easy lemma, whose proof is omitted.

**Lemma 5.1.** For all \( x, u, v, z \in \mathbb{R} \) we have

\[
1\{x+u \leq z\} - 1\{x+v \leq z\} = 1\{z-u \leq x \leq z-v\} \quad \text{and}
\]

\[
|1\{x+u \leq z\} - 1\{x+v \leq z\}| = 1\{z-u \leq x \leq z-v\}.
\]

**Lemma 5.2 (Concentration inequality).** For all real \( t < u \) and for all \( n \geq 1 \) we have

\[
P(t < W_n \leq u) \leq \frac{\sigma(u-t)}{c\sqrt{2\pi}\sqrt{n}} + \frac{2C_K\xi}{c^3\sqrt{n}}.
\]

**Proof.** The proof uses the Berry-Esseen Theorem for sums of i.i.d. random variables with finite third moment as well as the following fact, whose proof is straightforward: For each real-valued random variable \( X \) and for all real \( r < s \) we have the bound

\[
P(r < X \leq s) \leq \frac{s-r}{\sqrt{2\pi}} + 2d_K(X, Z).
\]

A similar result was used in [PR11] in the framework of exponential approximation. Now, for given \( t < u \) and \( n \geq 1 \) by (104) and the Berry-Esseen Theorem we have

\[
P(t < W_n \leq u) = P\left(\frac{\sigma t + a(\alpha - n)}{c\sqrt{n}} < \frac{S_n - na}{c\sqrt{n}} \leq \frac{\sigma u + a(\alpha - n)}{c\sqrt{n}}\right)
\]

\[
\leq \frac{\sigma(u-t)}{c\sqrt{2\pi}\sqrt{n}} + \frac{2C_K\xi}{c^3\sqrt{n}}.
\]

\(\square\)

**Remark 5.3.** It is actually not strictly necessary to apply the Berry-Esseen Theorem in order to prove Lemma 5.2. Using known concentration results for sums of independent random variables like Proposition 3.1 from [CGS11], for instance, would yield a comparable result, albeit with worse constants.

In order to prove Lemma 4.1 we cite the following concentration inequality from [CGS11]:
Lemma 5.4. Let $Y_1, \ldots, Y_n$ be independent mean zero random variables such that
\[ \sum_{j=1}^{n} E[Y_j^2] = 1 \quad \text{and} \quad \zeta := \sum_{j=1}^{n} E|Y_j|^3 < \infty, \]
then with $S^{(i)} := \sum_{j \neq i} Y_j$ one has for all real $r < s$ and all $i = 1, \ldots, n$ that
\[ P(r \leq S^{(i)} \leq s) \leq \sqrt{2}(s - r) + 2(\sqrt{2} + 1)\zeta. \]

Proof of Lemma 5.4. We first prove (71). Define
\[ W_{N^s}^{(1)} := W_{N^s} - \sigma^{-1} X_1 = \frac{1}{\sigma} \left( \sum_{j=2}^{N^s} X_j - \alpha \right) \]
such that
\[ W_{N^s} = W_{N^s}^{(1)} + \sigma^{-1} X_1 \quad \text{and} \quad W^* = W_{N^s}^{(1)} + \sigma^{-1} Y. \]
Then, using Lemma 5.1 we have
\[
|P(W^* \leq z) - P(W_{N^s} \leq z)| = |P(W_{N^s}^{(1)} + \sigma^{-1} Y \leq z) - P(W_{N^s}^{(1)} + \sigma^{-1} X_1 \leq z)| \\
\leq P(z - \sigma^{-1}(X_1 \lor Y) < W_{N^s}^{(1)} \leq z - \sigma^{-1}(X_1 \land Y)) \\
= E \left[ P \left( \frac{\sigma z - (X_1 \lor Y) + a(\alpha - N^s + 1)}{c \sqrt{N^s}} < \sum_{j=2}^{N^s} \left( \frac{X_j - a}{c \sqrt{N^s}} \right) \leq \frac{\sigma z - (X_1 \land Y) + a(\alpha - N^s + 1)}{c \sqrt{N^s}} \right) \right] \\
\leq \sigma z - (X_1 \land Y) + a(\alpha - N^s + 1) \left| N^s \right|.
\]
(105)

Now note that conditionally on $N^s$ the random variables $W_{N^s}^{(1)}$ and $(X_1, Y)$ are independent and that the statement of Lemma 5.4 may be applied to the random variable in the middle term of the above conditional probability giving the bound
\[ |P(W^* \leq z) - P(W_{N^s} \leq z)| \leq E \left[ \frac{\sqrt{2}Y - X_1}{c \sqrt{N^s}} + \frac{2(\sqrt{2} + 1)\xi}{c^3 \sqrt{N^s}} \right]. \]
(106)

Noting that $(X_1, Y)$ and $N^s$ are independent and using (53) again, we obtain
\[ E\left[ \frac{Y - X_1}{\sqrt{N^s}} \right] \leq \frac{3}{2c^2} \xi E[(N^s)^{-1/2}] \leq \frac{3\xi}{2c^2 \sqrt{\alpha}}, \]
as
\[ E[(N^s)^{-1/2}] = \frac{E[\sqrt{N}]}{E[N]} \leq \frac{\sqrt{E[N]}}{E[N]} = \frac{1}{\sqrt{\alpha}} \]
by (2) and Jensen’s inequality. From (106), (107) and (108) the bound (71) follows. Next we prove (72). Using Lemma 5.1 we obtain
\[ |P(W_{N^s} \leq z) - P(W \leq z)| = |E[J(1_{\{W_N^s \leq z\}} - 1_{\{W \leq z\}})] \\
- E[(1 - J)(1_{\{W_{N^s} - V \leq z\}} - 1_{\{W_{N^s} \leq z\}})]| \\
\leq E[J1_{\{z - (V \lor 0) < W \leq z - (V \land 0)\}}] \\
+ E[(1 - J)1_{\{z + (V \land 0) < W_{N^s} \leq z + (V \lor 0)\}}]
\]
(109)
\[ =: A_1 + A_2. \]
To bound $A_1$ we write

$$A_1 = \sum_{n=0}^{\infty} E\left[J_1\{N=n\}1\{z-(V\lor 0)<W\leq z-(V\land 0)\}\right]$$

(110)

$$= \sum_{n=0}^{\infty} P\left(z-(V\lor 0)<W\leq z-(V\land 0) \mid D \geq 0, N = n\right) \cdot P\left(D \geq 0, N = n\right).$$

Now note that conditionally on the event that $D \geq 0$ and $N = n$ the random variables $W$ and $V$ are independent and

$$\mathcal{L}(W \mid D \geq 0, N = n) = \mathcal{L}(W_n).$$

Thus, using Lemma 5.2 we have for all $n \geq 1$:

$$P\left(z-(V\lor 0)<W\leq z-(V\land 0) \mid D \geq 0, N = n\right)$$

$$= P\left(z-(V\lor 0)<W_n\leq z-(V\land 0) \mid D \geq 0, N = n\right)$$

(111)

$$\leq E\left[\frac{\sigma[V]}{c\sqrt{2\pi}\sqrt{n}} + \frac{2C_k \xi}{c^3 \sqrt{n}} \mid D \geq 0, N = n\right].$$

From (110) and (111) we thus have

$$A_1 \leq P(N = 0) + \sum_{n=1}^{\infty} E\left[\frac{\sigma[V]}{c\sqrt{2\pi}\sqrt{n}} + \frac{2C_k \xi}{c^3 \sqrt{n}} \mid D \geq 0, N = n\right] P\left(D \geq 0, N = n\right)$$

$$= P(N = 0) + \sum_{n=1}^{\infty} E\left[1\{D \geq 0, N = n\}\left(\frac{\sigma[V]}{c\sqrt{2\pi}\sqrt{n}} + \frac{2C_k \xi}{c^3 \sqrt{n}}\right)\right]$$

(112)

$$= P(N = 0) + E\left[J_1\{N \geq 1\}\left(\frac{\sigma[V]}{c\sqrt{2\pi}\sqrt{N}} + \frac{2C_k \xi}{c^3 \sqrt{N}}\right)\right].$$

Now note that

$$E\left[J_1\{N \geq 1\}\mid V\mid N^{-1/2}\right] = E\left[J_1\{N \geq 1\}N^{-1/2}E\left[\mid V\mid \mid N, N^s\right]\right]$$

$$\leq E\left[J_1\{N \geq 1\}N^{-1/2}\sqrt{E[V^2 \mid N, N^s]}\right]$$

(113)

$$= \frac{1}{\sigma} E\left[J_1\{N \geq 1\}N^{-1/2}\sqrt{c^2 D + a^2 D^2}\right]$$

(114)

$$\leq \frac{b}{\sigma} E\left[J D 1\{N \geq 1\}N^{-1/2}\right].$$

It remains to bound $A_2$. We may assume that $P(D < 0) > 0$ since otherwise $A_2 = 0$. Noting that $N^s \geq 1$ almost surely, similarly to (110) we obtain

$$A_2 = \sum_{m=1}^{\infty} P\left(z+(V\land 0)<W_{N^s} \leq z+(V\lor 0) \mid D < 0, N^s = m\right)$$

(115)

$$\cdot P\left(D < 0, N^s = m\right).$$

Now, using the fact that conditionally on the event $\{N^s = m\} \cap \{D < 0\}$ the random variables $W_{N^s}$ and $V$ are independent and

$$\mathcal{L}(W_{N^s} \mid N^s = m, D < 0) = \mathcal{L}(W_m)$$
in the same manner as \((112)\) we find
\[
A_2 \leq E \left[ (1 - J) \left( \frac{\sigma |V|}{c \sqrt{2 \pi \sqrt{N^*}}} + \frac{2C_K \xi}{c^3 \sqrt{N^*}} \right) \right].
\]

Using \((2)\) we have
\[
E \left[ (N^*)^{-1} \right] = \frac{1}{E[N]} = \frac{1}{\alpha}.
\]

Thus, from the Cauchy-Schwartz inequality and \((117)\) we obtain
\[
E \left[ (1 - J) \frac{|V|}{\sqrt{N^*}} \right] \leq \sqrt{E \left[ (N^*)^{-1} \right]} \sqrt{E[(1 - J)V^2]}
\]
\[
= \frac{1}{\sigma \sqrt{\alpha}} \sqrt{c^2 E[|D|(1 - J)] + a^2 E[D^2(1 - J)]}
\]
\[
\leq \frac{b}{\sigma \sqrt{\alpha}} \sqrt{E[D^2(1 - J)]}.
\]

Similarly, we have
\[
E \left[ \frac{1 - J}{\sqrt{N^*}} \right] \leq \sqrt{P(D < 0)} \sqrt{E \left[ (N^*)^{-1} \right]} = \frac{\sqrt{P(D < 0)}}{\sqrt{\alpha}}.
\]

Thus, from \((112)\), \((114)\), \((116)\) and \((119)\) we see that \(E_1 + E_2\) is bounded from above by the right hand side of \((72)\). Using \((112)\) and \((113)\) instead gives the bounds \((73)\) and \((74)\).

\[\square\]

**Proof of Lemma 4.2.** We only prove \((78)\), the proofs of \((79)\) and \((80)\) being similar and easier. By the definition of conditional expectation given an event, we have
\[
E \left[ J | V | 1_{\{z-(V \lor 0) < W_N \leq z-(V \land 0)\}} \right]
= \sum_{n=0}^{\infty} E \left[ 1_{\{N=n,D \geq 0\}} | V | 1_{\{z-(V \lor 0) < W_n \leq z-(V \land 0)\}} \right] = E \left[ 1_{\{N=0\}} J | V | \right]
+ \sum_{n=1}^{\infty} E \left[ |V| 1_{\{z-(V \lor 0) < W_n \leq z-(V \land 0)\}} \right] | N = n, D \geq 0 \right] \cdot P(N = n, D \geq 0).
\]

Now, for \(n \geq 1\), using the fact that the random variables \(W_N\) and \(V\) are conditionally independent given the event \(\{D \geq 0\} \cap \{N = n\}\), from Lemma 5.2 we infer that
\[
E \left[ |V| 1_{\{z-(V \lor 0) < W_n \leq z-(V \land 0)\}} \right] | N = n, D \geq 0 \right]
= E \left[ |V| \left( \frac{\sigma |V|}{c \sqrt{2 \pi \sqrt{N}}} + \frac{2C_K \xi}{c^3 \sqrt{N}} \right) \right] | N = n, D \geq 0 \right]
\]

Combining \((121)\) and \((122)\) we get
\[
E \left[ J | V | 1_{\{z-(V \lor 0) < W_N \leq z-(V \land 0)\}} \right] \leq E \left[ 1_{\{N=0\}} J | V | \right]
+ \sum_{n=1}^{\infty} E \left[ |V| \left( \frac{\sigma |V|}{c \sqrt{2 \pi \sqrt{N}}} + \frac{2C_K \xi}{c^3 \sqrt{N}} \right) \right] | N = n, D \geq 0 \right] \cdot P(N = n, D \geq 0)
\]
\[
= E \left[ 1_{\{N=0\}} J | V | \right] + E \left[ 1_{\{N \geq 1\}} J | V | \left( \frac{\sigma |V|}{c \sqrt{2 \pi \sqrt{N}}} + \frac{2C_K \xi}{c^3 \sqrt{N}} \right) \right].
\]

Using Cauchy-Schwartz as well as

\[ E[JV^2] = E[J \mathbb{E}[V^2 \mid N, N^*]] \leq \frac{b^2}{\sigma^2} E[JD^2] \]

we obtain

\[ E\left[1_{\{N=0\}} J \mathbb{E}[V] \right] \leq \frac{b}{\sigma} \sqrt{P(N=0)} \sqrt{E[JD^2]} . \]

Analogously to (114) one can show that

\[ E\left[1_{\{N\geq 1\}} N^{-1/2} JV^2 \right] \leq \frac{b^2}{\sigma^2} E\left[1_{\{N\geq 1\}} N^{-1/2} JD^2 \right] . \]

Hence, bound (78) follows from (123), (124), (114) and (125).

\[ \square \]

References

[AG10] R. Arratia and L. Goldstein. Size bias, sampling, the waiting time paradox, and infinite divisibility: when is the increment independent? arXiv:1007.3910, 2010.

[AGK13] R. Arratia, L. Goldstein, and F. Kochman. Size bias for one and all. arXiv:1308.2729, 2013.

[CGS11] L. H. Y. Chen, L. Goldstein, and Q.-M. Shao. Normal approximation by Stein’s method. Probability and its Applications (New York). Springer, Heidelberg, 2011.

[Che75] L. H. Y. Chen. Poisson approximation for dependent trials. Ann. Probability, 3(3):534–545, 1975.

[Döb12] C. Döbler. On rates of convergence and Berry-Esseen bounds for random sums of centered random variables with finite third moments. arXiv:1212.5401, 2012.

[Döb13] C. Döbler. Distributional transformations without orthogonality relations. arXiv:1312.6093v2, 2013.

[Dud02] R. M. Dudley. Real analysis and probability, volume 74 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.

[Eng83] G. Englund. A remainder term estimate in a random-sum central limit theorem. Teor. Veroyatnost. i Primenen., 28(1):143–149, 1983.

[ET14] P. Eichelsbacher and C. Thäle. New Berry-Esseen bounds for non-linear functionals of Poisson random measures. Electron. J. Probab., 19:no. 102, 25, 2014.

[GK96] B. V. Gnedenko and V. Yu. Korolev. Random summation. CRC Press, Boca Raton, FL, 1996. Limit theorems and applications.

[GR96] L. Goldstein and Y. Rinott. Multivariate normal approximations by Stein’s method and size bias couplings. J. Appl. Probab., 33(1):1–17, 1996.

[GR97] L. Goldstein and G. Reinert. Stein’s method and the zero bias transformation with application to simple random sampling. Ann. Appl. Probab., 7(4):935–952, 1997.

[GR05] L. Goldstein and G. Reinert. Distributional transformations, orthogonal polynomials, and Stein characterizations. J. Theoret. Probab., 18(1):237–260, 2005.

[Kor87] V. Yu. Korolev. The accuracy of the normal approximation to the distribution of the sum of a random number of independent random variables. In Stability problems for stochastic models (Varna, 1985), volume 1233 of Lecture Notes in Math., pages 36–40. Springer, Berlin, 1987.

[KS12] V. Yu. Korolev and I. Shevtsova. An improvement of the Berry-Esseen inequality with applications to Poisson and mixed Poisson random sums. Scand. Actuar. J., (2):81–105, 2012.
[LRP15] R. Lachièze-Rey and G. Peccati. New Kolmogorov bounds for functionals of Binomial point processes. preprint, available at http://w3.mi.parisdescartes.fr/ rlachiez/recherche/articles/LRPKol.pdf, 2015.

[Maj78] P. Major. On the invariance principle for sums of independent identically distributed random variables. *J. Multivariate Anal.*, 8(4):487–517, 1978.

[PR11] E. A. Peköz and A. Röllin. New rates for exponential approximation and the theorems of Rényi and Yaglom. *Ann. Probab.*, 39(2):587–608, 2011.

[PR12] J. Pike and H. Ren. Stein’s method and the Laplace distribution. arXiv:1210.5775, 2012.

[Rob48] H. Robbins. The asymptotic distribution of the sum of a random number of random variables. *Bull. Amer. Math. Soc.*, 54:1151–1161, 1948.

[She11] I. Shevtsova. On the absolute constants in the Berry-Esseen type inequalities for identically distributed summands. arXiv:1111.6554, 2011.

[Ste72] C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, pages 583–602, Berkeley, Calif., 1972. Univ. California Press.

[Ste73] F. W. Steutel. Some recent results in infinite divisibility. *Stochastic Processes Appl.*, 1:125–143, 1973.

[Sun13] J. K. Sunklodas. $L_1$ bounds for asymptotic normality of random sums of independent random variables. *Lith. Math. J.*, 53(4):438–447, 2013.

[Sun14] J. K. Sunklodas. On the normal approximation of a binomial random sum. *Lith. Math. J.*, 54(3):356–365, 2014.

[Tih80] A. N. Tihomirov. Convergence rate in the central limit theorem for weakly dependent random variables. *Teor. Veroyatnost. i Primenen.*, 25(4):800–818, 1980.