New Berry-Esseen and Wasserstein bounds in the CLT for non-randomly centered random sums by probabilistic methods

Christian Döbler

Université du Luxembourg
Unité de Recherche en Mathématiques
Campus Kirchberg
6, rue Richard Coudenhove-Kalergi
L-1359 Luxembourg,
Luxembourg
E-mail address: `christian.doebler@uni.lu`
URL: `wwwen.uni.lu/recherche/fstc/mathematics_research_unit/people/christian_doebler`

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 index with its size biased distribution as was previously considered in Goldstein and Rinott (1996) for the normal approximation of nonnegative random variables. When being specialized to concrete distributions of the summation index like the Binomial, Poisson and Hypergeometric distribution, our bounds turn out to be of the correct order of magnitude.

1. Introduction

Let $N, X_1, X_2, \dots$ 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 set of nonnegative integers $\mathbb{Z}_+ = \{0, 1, \dots\}$. Then, the random variable

$$S := \sum_{j=1}^{N} X_j \qquad (1.1)$$

is called a *random sum*. Such random variables appear frequently in modern probabiliy 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 (1.1) 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 Gnedenko and Korolev (1996). For example, by the recent so-called *master Steiner formula* from McCoy and Tropp (2014) the distribution of the metric projection of a standard Gaussian vector onto a closed convex cone in Euclidean space can be represented as a random sum of i.i.d. centered chi-squared random variables with the distribution of N given by the *conic intrinsic volumes* of the cone. Hence, this distribution belongs to the class of the so-called chi-bar-square distributions, which is ubiquitous in the theory of hypotheses testing with inequality constraints (see e.g. Dykstra, 1991 and Shapiro, 1988). This representation was used in Goldstein et al. (2014) to prove quantitative CLTs for both the distribution of the metric projection and the conic intrinsic volume distribution. These results are of interest e.g. in the field of compressed sensing.

There already exists a huge body of literature about the asymptotic distributions of random sums. Their investigation evidently began with the work Robbins (1948) 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 Robbins (1948) is that under these assumptions asymptotic normality of the index N automatically implies asymptotic normality of the corresponding random sum. The book Gnedenko and Korolev (1996) 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 (1.1) is asymptotically normal, where asymptotically refers to the fact that the *random index N* in fact usually depends on a parameter, which is sent 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 Döbler (2012) 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 Robbins (1948) the paper Englund (1983) 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 it difficult to apply this

result to 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 Korolev (1987) a bound on the Kolmogorov distance is given which improves upon the result of Englund (1983) with respect to the constants appearing in the bound. However, the bound given in Korolev (1987) is no longer strong enough to assure the well-known asymptotic normality of Binomial and Poisson random sums, unless the summands are centered. The paper Korolev (1988) generalizes the results from Englund (1983) to the case of not necessarily identically distributed summands and to situations, where the summands might not have finite absolute third moments. However, at least for non-centered summands, the bounds in Korolev (1988) still lack some explicitness.

To the best of our knowledge, the article Sunklodas (2013) 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 Sunklodas (2014), the results of Sunklodas (2013) generally do not yield accurate bounds, unless the summands are centered. Indeed, the results from Sunklodas (2013) do not even yield convergence in distribution for Binomial or Poisson 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. Using a simple inequality, this gives bounds for the whole classe of L^p distances of distributions, $1 \leq p \leq \infty$. These upper bounds, in their most abstract forms (see Theorem 2.5 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 being specialized 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. Indeed, every distribution of the summation index which allows for a close size-bias coupling should be amenable to our approach.

It should be mentioned that Stein's method and coupling techniques have previously been used to bound the error of exponential approximation Peköz and Röllin (2011) and approximation by the Laplace distribution Pike and Ren (2014) of certain random sums. In these papers, the authors make use of the fact that the exponential distribution and the Laplace distribution are the unique fixed points of certain distributional transformations and are able to succesfully couple the given

random sum with a random variable having the respective transformed distribution. In the case of the standard normal distribution, which is a fixed point of the zero-bias transformation from Goldstein and Reinert (1997), it appears natural to try to construct a close coupling with the zero biased distribution of the random sum under consideration. However, interestingly it turns out that we are only able to do so in the case of centered summands whereas for the general case an intermediate step involving a coupling of the index N with its size biased distribution is required for the proof. Nevertheless, the zero-bias transformation or rather an extension of it to non-centered random variables, plays an important role for our argument. This combination of two coupling constructions which belong to the classical tools of Stein's method for normal approximation is a new feature lying at the heart of our approach.

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 metrics 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 2.7 is given. Finally, Section 5 contains the proofs of some auxiliary results, needed for the proof of the Berry-Esseen bounds in Section 4.

2. Main results

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

$$d_{\mathcal{K}}(\mu, \nu) := \sup_{z \in \mathbb{R}} \left| \mu\big((-\infty, z]\big) - \mu\big((-\infty, z]\big) \right| = \|F - G\|_{\infty},$$

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

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

where $\text{Lip}(1)$ denotes the class of all Lipschitz-continous 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* bewtween μ and ν by

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

If the real-valued random variables X and Y have distributions μ and ν, respectively, then we simply write $d_{\mathcal{K}}(X, Y)$ for $d_{\mathcal{K}}\big(\mathcal{L}(X), \mathcal{L}(Y)\big)$ 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^s is said to have the X-*size biased distribution*, if for all bounded and measurable functions h on $[0, \infty)$

$$E[Xh(X)] = E[X]E[h(X^s)]\,, \tag{2.1}$$

see, e.g. Goldstein and Rinott (1996); Arratia and Goldstein (2010) or Arratia et al. (2013). 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 (2.1) 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

$$E\left[\left(X^s\right)^{p-1}\right] = \frac{E\left[X^p\right]}{E[X]}\,.$$

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) $d_{\mathcal{K}}(X, X^s) = d_{TV}(X, X^s) = \dfrac{E|X - E[X]|}{2E[X]}.$

(b) *If additionally $E[X^2] < \infty$, then $d_{\mathcal{W}}(X, X^s) = \dfrac{\mathrm{Var}(X)}{E[X]}.$*

Remark 2.2. (a) It is well known (see e.g. Dudley, 2002) that the Wasserstein distance $d_{\mathcal{W}}(X, Y)$ between the real random variables X and Y has the dual representation

$$d_{\mathcal{W}}(X, Y) = \inf_{(\hat{X}, \hat{Y}) \in \pi(X, Y)} E|\hat{X} - \hat{Y}|\,, \tag{2.2}$$

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 $\hat{X} \overset{\mathcal{D}}{=} X$ and $\hat{Y} \overset{\mathcal{D}}{=} Y$. Also, the infimum in (2.2) 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 Major (1978) 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 Arratia and Goldstein, 2010 for details). In fact, this property is already achieved by the coupling via the quantile transformation. By the

dual representation (2.2) and the fact that the coupling via the quantile trans-
formation yields the minimum L^1 distance in (2.2) 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

$$E\big|\hat{X}^s - \hat{X}\big| = E\big[\hat{X}^s\big] - E\big[\hat{X}\big] = E\big[F_{X^s}^{-1}(U)\big] - E\big[F_X^{-1}(U)\big]$$
$$= E\big|F_{X^s}^{-1}(U) - F_X^{-1}(U)\big| = d_{\mathcal{W}}(X, X^s).$$

Note also that, by the last computation and part by (b) of Lemma 2.1, we have

$$E\big[X^s\big] - E\big[X\big] = E\big[\hat{X}^s\big] - E\big[\hat{X}\big] = d_{\mathcal{W}}(X, X^s) = \frac{\mathrm{Var}(X)}{E[X]}.$$

(b) Due to a result by Steutel (1973), 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. Arratia and Goldstein, 2010 for
a nice exposition and a proof of this result). According to (a) such a coupling
always achieves the minimum L^1-distance.

(c) It might seem curious that according to part (a) of Lemma 2.1, the Kolmogorov
distance and the total variation distance between a nonnegative random vari-
able and one with its size biased distribution always coincide. Indeed, this holds
true since for each Borel-measurable set $B \subseteq \mathbb{R}$ we have the inequality

$$\big|P(X^s \in B) - P(X \in B)\big| \leq \big|P(X^s > m) - P(X > m)\big|$$
$$\leq d_{\mathcal{K}}(X, X^s),$$

where $m := E[X]$. Thus, the supremum in the definition

$$d_{TV}(X, X^s) = \sup_{B \in \mathcal{B}(\mathbb{R})} \big|P(X^s \in B) - P(X \in B)\big|$$

of the total variation distance is assumed for the set $B = (m, \infty)$. This can
be shortly proved and explained in the following way: For $t \in \mathbb{R}$, using the
defining property (2.1) of the size biased distribution, we can write

$$H(t) := P(X^s \leq t) - P(X \leq t) = m^{-1}E\big[(X - m))1_{\{X \leq t\}}\big].$$

Thus, for $s < t$ we have

$$H(t) - H(s) = m^{-1}E\big[(X - m))1_{\{s < X \leq t\}}\big],$$

and, hence, H is decreasing on $(-\infty, m)$ and increasing on (m, ∞). Thus, for
every Borel set $B \subseteq \mathbb{R}$ we conclude that

$$P(X^s \in B) - P(X \in B) = \int_{\mathbb{R}} 1_B(t)dH(t) \leq \int_{\mathbb{R}} 1_{B \cap (m, \infty)}(t)dH(t)$$

$$\leq \int_{\mathbb{R}} 1_{(m, \infty)}(t)dH(t) = P(X^s > m) - P(X > m).$$

Note that for this argumentation we heavily relied on the defining property (2.1)
of the size biased distribution which guaranteed the monotonicity property of
the difference H of the distribution functions of X^s and X, respectively. Since
X^s is stochastically larger than X, one might suspect that the coincidence of
the total variation and the Kolmogorov distance holds true in this more general
situation. However, observe that the fact that X^s dominates X stochastically
only implies that $H \leq 0$ but that it is the monotonicity of H on $(-\infty, m)$ and
on (m, ∞) that was crucial for the derivation.

Example 2.3. (a) Let $X \sim \text{Poisson}(\lambda)$ have the Poisson distribution with paramter $\lambda > 0$. From the Stein characterization of $\text{Poisson}(\lambda)$ (see Chen, 1975) 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 Goldstein and Rinott (1996) 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

$$d_{\mathcal{W}}(X, X^s) = E[1 - X_1] = 1 - p = \frac{\text{Var}(X)}{E[X]}$$

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.

$$P(X = k) = \frac{\binom{r}{k}\binom{s}{n-k}}{\binom{r+s}{n}}, \quad k = 0, 1, \ldots, n$$

with $E[X] = \frac{nr}{r+s}$. 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, $X = \sum_{j=1}^{n} X_j$ and since the X_j are exchangeable, the well-known construction of a random variable X^s wth the X-size biased distribution from Goldstein and Rinott (1996) gives that $X^s = 1 + \sum_{j=2}^{n} X_j'$, where

$$\mathcal{L}\big((X_2', \ldots, X_n')\big) = \mathcal{L}\big((X_2, \ldots, X_n) \,\big|\, X_1 = 1\big).$$

But given $X_1 = 1$ the sum $\sum_{j=2}^{n} X_j$ has the Hypergeometric distribution with parameters $n - 1, r - 1$ and s and, hence,

$$X^s \overset{\mathcal{D}}{=} Y + 1, \quad \text{where } Y \sim \text{Hyp}(n-1; r-1, s).$$

In order to construct an L^1-optimal coupling of X and X^s, 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,

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

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

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

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

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:

$$d_{\mathcal{W}}(X, X^s) = E\big|X^s - X\big| = \frac{\mathrm{Var}(X)}{E[X]} = \frac{n\frac{r}{r+s}\frac{s}{r+s}\frac{r+s-n}{r+s-1}}{\frac{nr}{r+s}} = \frac{s(r+s-n)}{(r+s)(r+s-1)}\,.$$

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 Robbins (1948).

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 $E[N^3] < \infty$. Furthermore, we let

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

$$a := E[X_1], \quad b^2 := E[X_1^2], \quad c^2 := \mathrm{Var}(X_1) = b^2 - a^2 \text{ and } d^3 := E\big|X_1 - E[X_1]\big|^3.$$

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

$$\mu := E[S] = \alpha a \quad \text{and} \quad \sigma^2 := \mathrm{Var}(S) = \alpha c^2 + a^2 \gamma^2\,. \qquad (2.3)$$

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

$$W := \frac{S - \mu}{\sigma} = \frac{S - \alpha a}{\sqrt{\alpha c^2 + a\gamma^2}} \qquad (2.4)$$

of S measured by the Kolmogorov and the Wasserstein distance, respectively. As can be seen from the paper Robbins (1948), under the general assumption that

$$\sigma^2 = \alpha c^2 + a^2 \gamma^2 \to \infty\,,$$

there are three typical situations in which W is asymptotically normal, which we will now briefly review.
1) $c \neq 0 \neq a$ and $\gamma^2 = o(\alpha)$
2) $a = 0 \neq c$ and $\gamma = o(\alpha)$
3) N itself is asymptotically normal and at least one of a and c is different from zero.

We remark that 1) roughly means that N tends to infinity in a certain sense, but such that it only fluctuates slightly around its mean α and, thus, behaves more or less as the constant α (tending to infinity). If $c = 0$ and $a \neq 0$, then we have

$$S = aN \quad \text{a.s.}$$

and asymptotic normality of S is equivalent to that of N. For this reason, unless specifically stated otherwise, we will from now on assume that $c \neq 0$. However, we would like to remark that all bounds in which c does not appear in the denominator also hold true in the case $c = 0$.

Theorem 2.5. *Suppose that Assumption 2.4 holds, let W be given by (2.4) and let Z have the standard normal distribution. Also, let (N, N^s) be a coupling of N and N^s having the N-size biased distribution such that N^s is also independent of $X_1, X_2, \ldots$ and define $D := N^s - N$. Then, we have the following bound:*

(a) $d_{\mathcal{W}}(W, Z) \leq \dfrac{2c^2 b \gamma^2}{\sigma^3} + \dfrac{3\alpha d^3}{\sigma^3} + \dfrac{\alpha a^2}{\sigma^2} \sqrt{\dfrac{2}{\pi}} \sqrt{\operatorname{Var}\big(E[D \,|\, N]\big)}$

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

(b) *If, additionally, $D \geq 0$, then we also have*

$$d_{\mathcal{K}}(W, Z) \leq \frac{(\sqrt{2\pi} + 4)bc^2\alpha}{4\sigma^3} \sqrt{E[D^2]} + \frac{d^3\alpha(3\sqrt{2\pi} + 4)}{8\sigma^3} + \frac{c^3\alpha}{\sigma^3}$$

$$+ \left(\frac{7}{2}\sqrt{2} + 2\right)\frac{\sqrt{\alpha}d^3}{c\sigma^2} + \frac{c^2\alpha}{\sigma^2}P(N = 0) + \frac{d^3\alpha}{c\sigma^2}E\big[N^{-1/2}1_{\{N \geq 1\}}\big]$$

$$+ \frac{\alpha a^2}{\sigma^2}\sqrt{\operatorname{Var}\big(E[D \,|\, N]\big)} + \frac{\alpha|a|b^2}{2\sigma^3}\sqrt{E\Big[\big(E[D^2 \,|\, N]\big)^2\Big]}$$

$$+ \frac{\alpha|a|b^2\sqrt{2\pi}}{8\sigma^3}E[D^2] + \frac{\alpha|a|b}{\sigma^2}\sqrt{P(N = 0)}\sqrt{E[D^2]}$$

$$+ \frac{\alpha|a|b^2}{c\sigma^2\sqrt{2\pi}}E\big[D^2 1_{\{N \geq 1\}}N^{-1/2}\big]$$

$$+ \left(\frac{d^3\alpha|a|b}{\sigma^2} + \frac{\alpha bc}{\sigma^2\sqrt{2\pi}}\right)E\big[D 1_{\{N \geq 1\}}N^{-1/2}\big].$$

Remark 2.6. (a) In many concrete situations, one has that a natural coupling of N and N^s yields $D \geq 0$ and, hence, Theorem 2.5 gives bounds on both the Wasserstein and Kolmogorov distances (note that the fourth summand in the bound on $d_{\mathcal{W}}(W, Z)$ vanishes if $D \geq 0$). 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, our bounds can further be simplified (see Corollary 2.9 below). Indeed, since N^s is always stochastically larger than N, by Remark 2.2 (a) it is always possible to construct a coupling (N, N^s) such that $D = N^s - N \geq 0$.

(b) However, although we know that a coupling of N and N^s such that $D = N^s - N \geq 0$ is always possible in principle, sometimes one would prefer working with a feasible and natural coupling which does not have this property. For instance, this is the case in the situation of Corollary 2.11 below. This is why we have not restricted ourselves to the case $D \geq 0$ but allow for arbitrary couplings (N, N^s). We mention that we also have a bound on the Kolmogorov distance between W and a standard normally distributed Z in this more general situation, which is given by

$$d_{\mathcal{K}}(W, Z) \leq \sum_{j=1}^{7} B_j,$$

where B_1, B_2, B_4, B_5, B_6 and B_7 are defined in (4.33), (4.38), (4.44), (4.50), (4.59) and (4.64), respectively, and

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

It is this bound what is actually proved in Section 4. Since it is given by a rather long expression in the most general case, we have decided, however, not to present it within Theorem 2.5.

(c) We mention that the our proof of the Wasserstein bounds given in Theorem 2.5 is only roughly five pages long and is not at all technical but rather makes use of probabilistic ideas and concepts. The extended length of our derivation is simply due to our ambition to present Kolmogorov bounds as well which, as usual within Stein's method, demand much more technicality.

The next theorem treats the special case of centered summands.

Theorem 2.7. *Suppose that Assumption 2.4 holds with $a = E[X_1] = 0$, let W be given by (2.4) and let Z have the standard normal distribution. Then,*

$$d_{\mathcal{W}}(W, Z) \leq \frac{2\gamma}{\alpha} + \frac{3d^3}{c^3 \sqrt{\alpha}} \quad and$$

$$d_{\mathcal{K}}(W, Z) \leq \frac{(\sqrt{2\pi} + 4)\gamma}{4\alpha} + \left(\frac{d^3(3\sqrt{2\pi} + 4)}{8c^3} + 1 \right) \frac{1}{\sqrt{\alpha}} + \left(\frac{7}{2}\sqrt{2} + 2 \right) \frac{d^3}{c^3 \alpha}$$

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

Remark 2.8. (a) The proof will show that Theorem 2.7 holds as long as $E[N^2] < \infty$. Thus, Assumption 2.4 could be slightly relaxed in this case.

(b) Theorem 2.7 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 2.7 if $a = 0$, but the resulting bounds would look more cumbersome in the general case. Also, they would be of the same order as the bounds presented in Theorem 2.5 in the case that $a \neq 0$. This is why we have refrained from presenting these bounds in the general case but have chosen to prove Theorem 2.5 and Theorem 2.7 in parallel. Note that, if $a \neq 0$, then a necessary condition for our bounds to imply the CLT is that

$$\frac{\alpha}{\sigma^3} E[D^2] = o(1) \quad \text{and} \quad \frac{\alpha}{\sigma^2} \sqrt{\mathrm{Var}\big(E[D|N]\big)} = o(1) \,. \tag{2.5}$$

This should be compared to the conditions which imply asymptotic normality for N by size-bias couplings given in Goldstein and Rinott (1996), namely

$$\frac{\alpha}{\gamma^3} E[D^2] = o(1) \quad \text{and} \quad \frac{\alpha}{\gamma^2} \sqrt{\mathrm{Var}\big(E[D|N]\big)} = o(1) \,. \tag{2.6}$$

If (2.6) holds, then from Goldstein and Rinott (1996) we know that N is asymptotically normal and, as was shown within the proof of Lemma 1 in Robbins (1948), this implies that $\gamma = o(\alpha)$. Since, if $a \neq 0$, (2.6) implies (2.5), we can conclude from Theorems 2.5 and 2.7 that W is asymptotically normal. In a nutshell, if the bounds from Goldstein and Rinott (1996) on the distance to normality of N tend to zero, then so do our bounds and, hence, yield the CLT for W. However, the validity of (2.6) is neither necessary for (2.5) to hold

nor for our bounds to imply asymptotic normality of W (see Remark 2.17 (b) below).

(c) 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 Dudley, 2002) that $\|F - G\|_1$ coincides with the Wasserstein distance of the corresponding distributions μ and ν, say. By Hölder's inequality, for $1 \leq p < \infty$, we have

$$\|F - G\|_p \leq d_{\mathcal{K}}(\mu, \nu)^{\frac{p-1}{p}} \cdot d_{\mathcal{W}}(\mu, \nu)^{\frac{1}{p}} .$$

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

(d) It would be possible to drop the assumption that the summands be identically distributed. For reasons of clarity of the presentation, we have, however, decided to stick to the i.i.d. setting. See also the discussion of possible generalizations before the proof of Lemma 2.1 at the end of this section.

Corollary 2.9. *Suppose that Assumption 2.4 holds, let W be given by (2.4) and let Z have the standard normal distribution. Furthermore, assume that the distribution of the index N is infinitely divisible. Then, we have*

$$d_{\mathcal{W}}(W, Z) \leq \frac{2c^2 b\gamma^2 + 3\alpha d^3}{\sigma^3} + \frac{(\alpha\delta^3 - \alpha^2\gamma^2 + \gamma^4 - \beta^4)|a|b^2}{\alpha\sigma^3} \quad and$$

$$d_{\mathcal{K}}(W, Z) \leq \frac{d^3\alpha(3\sqrt{2\pi} + 4)}{8\sigma^3} + \frac{c^3\alpha}{\sigma^3} + \left(\frac{7}{2}\sqrt{2} + 2 \right) \frac{\sqrt{\alpha}d^3}{c\sigma^2} + \frac{c^2\alpha}{\sigma^2} P(N = 0)$$

$$+ \frac{|a|b^2(\delta^3\alpha + \gamma^4 - \beta^4 - \gamma^2\alpha^2)}{\alpha\sigma^3} \left(\frac{\sqrt{2\pi}}{8} + \frac{1}{2} \right)$$

$$+ \sqrt{\delta^3\alpha + \gamma^4 - \beta^4 - \gamma^2\alpha^2} \left(\frac{(\sqrt{2\pi} + 4)bc^2}{4\sigma^3} + \sqrt{P(N = 0)} \frac{|a|b}{\sigma^2} \right)$$

$$+ E\left[1_{\{N \geq 1\}} N^{-1/2} \right] \left(\frac{|a|b^2(\delta^3\alpha + \gamma^4 - \beta^4 - \gamma^2\alpha^2)}{c\alpha\sigma^2\sqrt{2\pi}} + \frac{\gamma^2 d^3|a|b}{\sigma^2} + \frac{d^3\alpha}{c\sigma^2} + \frac{\gamma^2 bc}{\sigma^2\sqrt{2\pi}} \right) .$$

Proof: By Remark 2.2 (b) we can choose $D \geq 0$ independent of N such that $N^s = N + D$ has the N-size biased distribution. Thus, by independence we obtain

$$\text{Var}(D) = \text{Var}(N^s) - \text{Var}(N) = E\left[(N^s)^2 \right] - E[N^s]^2 - \gamma^2$$

$$= \frac{E[N^3]}{E[N]} - \left(\frac{E[N^2]}{E[N]} \right)^2 - \gamma^2$$

$$= \frac{\delta^3}{\alpha} - \frac{\beta^4}{\alpha^2} - \gamma^2 .$$

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,

$$\text{Var}\left(E[D|N] \right) = \text{Var}\left(E[D] \right) = 0 \quad \text{and} \quad \sqrt{E\left[\left(E[D^2 \mid N] \right)^2 \right]} = E[D^2]$$

in this case. Now, the claim follows from Theorem 2.5.

□

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. Shevtsova, 2011 and Goldstein, 2010 for comparison).

Corollary 2.10. *Suppose that Assumption 2.4 holds, let W be given by (2.4) and let Z have the standard normal distribution. Also, assume that the index N is a positive constant. Then,*

$$d_{\mathcal{W}}(W, Z) \leq \frac{3d^3}{c^3\sqrt{N}} \quad and$$

$$d_{\mathcal{K}}(W, Z) \leq \frac{1}{\sqrt{N}}\left(1 + \left(\frac{7}{2}\left(1 + \sqrt{2}\right) + \frac{3\sqrt{2\pi}}{8}\right)\frac{d^3}{c^3}\right).$$

Proof: In this case, we can choose $N^s = N$ yielding $D = 0$ and the result follows from Theorem 2.5.

□

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. Using the bound presented in Remark 2.6 (b) one would get a bound on the Kolmogorov distance, which is more complicated but of the same order of magnitude. A different way to prove bounds for the CLT by Stein's method in this special situation is presented in Theorem 10.6 of Chen et al. (2011). Their method relies on a general bound for the error of normal approximation to the distribution of a non-linear statistic of independent random variables which can be written as a linear statistic plus a small remainder term as well as on truncation and conditioning on N in order to apply the classical Berry-Esseen theorem. Though our method also makes use of conditioning on N, it is more directly tied to random sums and also relies on (variations of) classical couplings in Stein's method (see the proof in Section 4 for details).

Corollary 2.11. *Suppose that Assumption 2.4 holds, let W be given by (2.4) and let Z have the standard normal distribution. Additionally, assume that the distribution of the index N is such that $N \overset{\mathcal{D}}{=} N_1 + \ldots + N_n$, 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*

$$\alpha_1 := E[N_1], \quad \beta_1^2 := E[N_1^2], \quad \gamma_1^2 := \mathrm{Var}(N_1), \quad \delta_1^3 := E[N_1^3] \quad and$$
$$\sigma_1^2 := c^2\alpha_1 + a^2\gamma_1^2$$

we have

$$d_{\mathcal{W}}(W, Z) \leq \frac{1}{\sqrt{n}}\left(\frac{2c^2 b\gamma_1^2}{\sigma_1^3} + \frac{3\alpha_1 d^3}{\sigma_1^3} + \sqrt{\frac{2}{\pi}}\frac{\alpha_1 a^2\gamma_1^2}{\sigma_1^2} + \frac{2\alpha_1(a^2 b + |a|b^2)}{\sigma_1^3}\left(\frac{\delta_1^3}{\alpha_1} - \beta_1^2\right)\right).$$

Proof: From Goldstein and Rinott (1996) (see also Chen et al., 2011) 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,

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

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 (2.1),

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

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

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 2.12. *Suppose that Assumption 2.4 holds, let W be given by (2.4) and let Z have the standard normal distribution. Assume further that $N \sim \text{Poisson}(\lambda)$ has the Poisson distribution with parameter $\lambda > 0$. Then,*

$$d_{\mathcal{W}}(W, Z) \leq \frac{1}{\sqrt{\lambda}} \left(\frac{2c^2}{b^2} + \frac{3d^3}{b^3} + \frac{|a|}{b} \right) \quad \text{and}$$

$$d_{\mathcal{K}}(W, Z) \leq \frac{1}{\sqrt{\lambda}} \left(\frac{\sqrt{2\pi}}{4} + 1 + \frac{(3\sqrt{2\pi}+4)d^3}{8b^3} + \frac{c^3}{b^3} + \left(\frac{7}{2}\sqrt{2} + 3 \right) \frac{d^3}{cb^2} \right.$$

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

Proof: In this case, by Example 2.3 (a), we can choose $D = 1$, yielding that

$$E[D^2] = 1 \quad \text{and} \quad \operatorname{Var}\big(E[D|N]\big) = 0.$$

Note that

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

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

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

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

Hence,

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

Noting that

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

the result follows from Theorem 2.5.

$\square$

Remark 2.13. The Berry-Esseen bound presented in Corollary 2.12 is of the same order of λ as the bound given in Korolev and Shevtsova (2012), which seems to be the best currently available, but has a worst constant. However, it should be mentioned that the bound in Korolev and Shevtsova (2012) was obtained using special properties of the Poisson distribution and does not seem likely to be easily transferable to other distributions of N.

Corollary 2.14. *Suppose that Assumption 2.4 holds, let W be given by (2.4) and let Z have the standard normal distribution. Furthermore, assume that $N \sim \text{Bin}(n,p)$ has the Binomial distribution with parameters $n \in \mathbb{N}$ and $p \in (0,1]$. Then,*

$$d_{\mathcal{W}}(W,Z) \leq \frac{1}{\sqrt{np}\big(b^2 - pa^2\big)^{3/2}} \bigg(\big(2c^2b + |a|b^2\big)(1-p) + 3d^3$$

$$+ \sqrt{\frac{2}{\pi}}a^2p\sqrt{b^2 - pa^2}\sqrt{1-p} \bigg) \quad and$$

$$d_{\mathcal{K}}(W,Z) \leq \frac{1}{\sqrt{np}\big(b^2 - pa^2\big)^{3/2}} \bigg(c^3 + \frac{(\sqrt{2\pi}+4)bc^2\sqrt{1-p}}{4} + \frac{(3\sqrt{2\pi}+4)d^3}{8}$$

$$+ \frac{|a|b^2\sqrt{1-p}}{2} + \frac{|a|b^2\sqrt{2\pi}(1-p)}{8} \bigg)$$

$$+ \frac{1}{\sqrt{np}\big(b^2 - pa^2\big)} \bigg(\Big(\frac{9}{2}\sqrt{2}+2\Big)\frac{d^3}{c} + \sqrt{1-p}\big(a^2p + \sqrt{2}|a|bd^3\big)$$

$$+ \frac{\sqrt{2(1-p)}b\big(2b^2 - a^2\big)}{c\sqrt{2\pi}} \bigg)$$

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

Remark 2.15. Bounds for binomial random sums have also been derived in Sunklodas (2014) using a technique developed in Tihomirov (1980). Our bounds are of the same order $(np)^{-1/2}$ of magnitude.

Proof of Corollary 2.14: 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

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

We have $D^2 = D$ and, by Cauchy-Schwarz,

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

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

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

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

Thus,

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

Also, we can bound

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

Now, using $a^2 + c^2 = b^2$, the claim follows from Theorem 2.5.

$\square$

Corollary 2.16. *Suppose that Assumption 2.4 holds, let W be given by (2.4) and let Z have the standard normal distribution. Assume further that $N \sim \mathrm{Hyp}(n;r,s)$ has the Hypergeometric distribution with parameters $n, r, s \in \mathbb{N}$ such that $n \leq \min\{r,s\}$. Then,*

$$d_{\mathcal{W}}(W,Z) \leq \left(\frac{nr}{r+s}\right)^{-1/2}\left(\frac{2b}{c}\frac{s(r+s-n)}{(r+s)_2} + \frac{3d^3}{c^3} + \frac{|a|b^2}{c^2}\frac{s(r+s-n)}{(r+s)_2}\right)$$

$$+ K\frac{a^2}{c^2}\sqrt{\frac{2}{\pi}}\left(\frac{\min\{r,s\}}{n(r+s)}\right)^{1/2} \quad and$$

$$d_{\mathcal{K}}(W,Z) \leq \left(\frac{nr}{r+s}\right)^{-1/2}\left[1 + \frac{(\sqrt{2\pi}+4)b}{4c}\left(\frac{s(r+s-n)}{(r+s)_2}\right)^{1/2}\right.$$

$$+ \left(\frac{3\sqrt{2\pi}}{8} + \frac{9}{2}\sqrt{2} + \frac{5}{2}\right)\frac{d^3}{c^3} + \left(\frac{\sqrt{2\pi}}{8}+1\right)\frac{|a|b^2}{c^3}\frac{s(r+s-n)}{(r+s)_2}$$

$$+ \left(|a|b^2c^3\sqrt{2\pi} + \frac{|a|bd^3}{c^2} + \frac{b}{c\sqrt{2\pi}}\right)\left(\frac{2s(r+s-n)}{(r+s)_2}\right)^{1/2}\right]$$

$$+ \frac{(s)_n}{(r+s)_n} + K\frac{a^2}{c^2}\left(\frac{\min\{r,s\}}{n(r+s)}\right)^{1/2} + \frac{|a|b}{c^2}\left(\frac{(s)_n}{(r+s)_n}\frac{s(r+s-n)}{(r+s)_2}\right)^{1/2},$$

where K is a numerical constant and $(m)_n = m(m-1)\cdot\ldots\cdot(m-n+1)$ denotes the lower factorial.

Proof: In this case, we clearly have

$$\alpha = \frac{nr}{r+s}, \quad \gamma^2 = \frac{nr}{r+s} \frac{s}{r+s} \frac{r+s-n}{r+s-1} = \frac{nr}{r+s} \frac{s(r+s-n)}{(r+s)_2} \quad \text{and}$$

$$\sigma^2 = \frac{nr}{r+s} \left(c^2 + a^2 \frac{s}{r+s} \frac{r+s-n}{r+s-1} \right) = \frac{nr}{r+s} \left(c^2 + a^2 \frac{s(r+s-n)}{(r+s)_2} \right).$$

Hence,

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

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

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

Furthermore, we know that

$$E[D] = E[D^2] = d_{\mathcal{W}}(N, N^s) = \frac{\mathrm{Var}(N)}{E[N]} = \frac{s(r+s-n)}{(r+s)_2}.$$

Elementary combinatorics yield

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

Thus,

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

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

Using a computer algebra system, one may check that

$$\mathrm{Var}\big(E[D \,|\, N]\big) = \Big(nrs - n^3 rs - r^2 s + 5n^2 r^2 s + 2n^3 r^2 s - 8nr^3 s - 8n^2 r^3 s + 2nrs^5$$
$$- n^3 r^3 s + 4r^4 s + 10nr^4 s + 3n^2 r^4 s - 4r^5 s - 3nr^5 s + r^6 s + ns^2$$
$$- n^3 s^2 - 2rs^2 + 4n^2 rs^2 - 2n^3 rs^2 - 14nr^2 s^2 - 4n^2 r^2 s^2 + n^3 r^2 s^2$$
$$+ 12r^3 s^2 + 20nr^3 s^2 + 2n^2 r^3 s^2 - 14r^4 s^2 - 7nr^4 s^2 + 4r^5 s^2 - s^3$$
$$- n^2 s^3 + 2n^3 s^3 - 5nrs^3 + 4n^2 rs^3 + n^3 rs^3 + 13r^2 s^3 + 8nr^2 s^3$$
$$- 4n^2 r^2 s^3 - 18r^3 s^3 - 3nr^3 s^3 + 6r^4 s^3 + ns^4 - n^3 s^4 + 6rs^4 - 4nrs^4$$
$$- 2n^2 rs^4 - 10r^2 s^4 + 3nr^2 s^4 + 4r^3 s^4 + s^5 - 2ns^5 + n^2 s^5 - 2rs^5$$
$$+ r^2 s^5 \Big) \Big(nr(r+s)^2 (r+s-1)^2 (r+s-2)(r+s-3) \Big)^{-1}$$
$$=: \varepsilon(n, r, s). \tag{2.7}$$

One can check that under the assumption $n \le \min\{r, s\}$ always

$$\varepsilon(n, r, s) = O\left(\frac{\min\{r, s\}}{n(r+s)} \right).$$

Hence, there is a numerical constant K such that

$$\sqrt{\varepsilon(n,r,s)} \le K\left(\frac{\min\{r,s\}}{n(r+s)}\right)^{1/2}.$$

Also, by the conditional version of Jensen's inequality

$$E\left[\left(E\left[D^2 \mid N\right]\right)^2\right] \le E\left[D^4\right] = E[D] = \frac{s(r+s-n)}{(r+s)_2}.$$

Using

$$E\left[N^{-1}1_{\{N\ge 1\}}\right] = \binom{r+s}{n}^{-1}\sum_{k=1}^{n}\frac{1}{k}\binom{r}{k}\binom{s}{n-k}$$

$$\le \frac{2}{r+1}\binom{r+s}{n}^{-1}\sum_{l=2}^{n+1}\binom{r+1}{l}\binom{s}{n+1-l}$$

$$\le \frac{2}{r+1}\binom{r+s}{n}^{-1}\binom{r+1+s}{n+1} = \frac{2(r+s+1)}{(n+1)(r+1)} \le 2\frac{r+s}{nr},$$

we get

$$E\left[D1_{\{N\ge 1\}}N^{-1/2}\right] \le \sqrt{E[D^2]}\sqrt{E\left[1_{\{N\ge 1\}}N^{-1}\right]} \le \left(2\frac{s(r+s-n)}{(r+s)(r+s-1)}\frac{r+s}{nr}\right)^{1/2}$$

$$= \sqrt{2}\left(\frac{s}{nr}\frac{r+s-n}{r+s-1}\right)^{1/2} \quad\text{and}$$

$$E\left[1_{\{N\ge 1\}}N^{-1/2}\right] \le \sqrt{E\left[1_{\{N\ge 1\}}N^{-1}\right]} \le \sqrt{2}\left(\frac{r+s}{nr}\right)^{1/2}.$$

Finally, we have

$$P(N=0) = \frac{\binom{s}{n}}{\binom{r+s}{n}} = \frac{s(s-1)\cdot\ldots\cdot(s-n+1)}{(r+s)(r+s-1)\cdot\ldots\cdot(r+s-n+1)} = \frac{(s)_n}{(r+s)_n}.$$

Thus, the result follows from Theorem 2.5.

$\square$

Remark 2.17. (a) From the above proof we see that the numerical constant K appearing in the bounds of Corollary 2.16 could in principle be computed explicitly. Also, as always

$$\frac{\min\{r,s\}}{n(r+s)} \le \frac{r+s}{nr} = \frac{1}{E[N]},$$

we conclude that the bounds are of order $E[N]^{-1/2}$.

(b) One typical situation, in which a CLT for Hypergeometric random sums holds, is when N, itself, is asymptotically normal. Using the same coupling (N, N^s) as in the above proof and the results from Goldstein and Rinott (1996), one obtains that under the condition

$$\frac{\max\{r,s\}}{n\min\{r,s\}} \longrightarrow 0 \tag{2.8}$$

the index N is asymptotically normal. This condition is stricter than that

$$E[N]^{-1} = \frac{r+s}{nr} \longrightarrow 0, \tag{2.9}$$

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 (2.8) holds, if and only if $|\varepsilon - \kappa| < 1$, whereas (2.9) is equivalent to $\kappa - \varepsilon < 1$ in this case.

Before we end this section by giving the proof of Lemma 2.1, we would like to mention in what respects the results in this article could be generalized. Firstly, it would be possible do dispense with the assumption of independence among the summands $X_1, X_2, \ldots$. Of course, the terms appearing in the bounds would look more complicated, but the only essential change would be the emergence of the additional error term

$$E_3 := \frac{C}{\alpha\sigma} E|\alpha A_N - \mu N|,$$

where

$$A_N := \sum_{j=1}^{N} E[X_j] \quad \text{and} \quad \mu = E[A_N]$$

and where C is an explicit constant depending on the probabilistic distance chosen. Note that $E_3 = 0$ if the summands are either i.i.d. or centered.

Secondly, it would be possible in principle to allow for some dependence among the summands $X_1, X_2, \ldots$. Indeed, an inspection of the proof in Section 4 reveals that this dependence should be such that for the non-random partial sums bounds on the normal approximation exist and such that suitable couplings with the non-zero biased distribution (see Section 3) of those partial sums are available. The latter, however, have not been constructed yet in great generality, although Goldstein and Reinert (1997) gives a construction for summands forming a simple random sampling in the zero bias case.

It would be much more difficult to abandon the assumption about the independence of the summation index and the summands. This can be seen from Equation (4.9) below, in which the second identity would no longer hold, in general, if this independence was no longer valid. Also, one would no longer be able to freely choose the coupling (N, N^s) when specializing to concrete distributions of N.

Proof of Lemma 2.1: Let h be a measurable function such that all the expected values in (2.1) exist. By (2.1) we have

$$\left|E[h(X^s)] - E[h(X)]\right| = \frac{1}{E[X]}\left|E\left[(X - E[X])h(X)\right]\right|. \qquad (2.10)$$

It is well known that

$$d_{TV}(X, Y) = \sup_{h \in \mathcal{H}}\left|E[h(X)] - E[h(Y)]\right|, \qquad (2.11)$$

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

$$\frac{1}{E[X]}\left|E\left[(X - E[X])h(X)\right]\right| \leq \frac{E|X - E[X]|}{2E[X]}.$$

Hence, from (2.11) and (2.10) we conclude that

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

On the other hand, letting

$$h(x) := \frac{1}{2}\left(1_{\{x>E[X]\}} - 1_{\{x\leq E[X]\}}\right)$$

in (2.10) we have $h \in \mathcal{H}$ and obtain

$$\left|E[h(X^s)] - E[h(X)]\right| = \frac{E|X - E[X]|}{2E[X]}$$

proving the second equality of (a). Note that, since X^s is stochastically larger than X, we have

$$d_{\mathcal{K}}(X, X^s) = \sup_{t\geq 0}\left|P(X^s > t) - P(X > t)\right| = \sup_{t\geq 0}\left(P(X^s > t) - P(X > t)\right)$$

$$= \sup_{t\geq 0}\left(E[g_t(X^s)] - E[g_t(X)]\right), \tag{2.12}$$

where $g_t := 1_{(t,\infty)}$.
By (2.10), choosing $t = E[X]$ yields

$$d_{\mathcal{K}}(X, X^s) \geq E\left[(X - E[X])1_{\{X>E[X]\}}\right]. \tag{2.13}$$

If $0 \leq t < E[X]$ we obtain

$$E\left[(X - E[X])1_{\{X>t\}}\right] = E\left[(X - E[X])1_{\{t<X\leq E[X]\}}\right]$$
$$+ E\left[(X - E[X])1_{\{X>E[X]\}}\right]$$
$$\leq E\left[(X - E[X])1_{\{X>E[X]\}}\right]. \tag{2.14}$$

Also, if $t \geq E[X]$, then

$$E\left[(X - E[X])1_{\{X>t\}}\right] \leq E\left[(X - E[X])1_{\{X>E[X]\}}\right]. \tag{2.15}$$

Thus, by (2.10), from (2.12), (2.13), (2.14) and (2.15) we conclude that

$$d_{\mathcal{K}}(X, X^s) = E\left[(X - E[X])1_{\{X>E[X]\}}\right]. \tag{2.16}$$

Now, the remaining claim of (a) can be easily inferred from (2.16) and from the following two identities:

$$0 = E[X - E[X]] = E\left[(X - E[X])1_{\{X>E[X]\}}\right] - E\left[(X - E[X])1_{\{X\leq E[X]\}}\right]$$
$$= E\left[|X - E[X]|1_{\{X>E[X]\}}\right] - E\left[|X - E[X]|1_{\{X\leq E[X]\}}\right]$$

and

$$E|X - E[X]| = E\left[|X - E[X]|1_{\{X>E[X]\}}\right] + E\left[|X - E[X]|1_{\{X\leq E[X]\}}\right]$$
$$= 2E\left[(X - E[X])1_{\{X>E[X]\}}\right].$$

Finally, if h is 1-Lipschitz continuous, then

$$\left|E\left[(X - E[X])h(X)\right]\right| = \left|E\left[(X - E[X])(h(X) - h(E[X]))\right]\right|$$
$$\leq \|h'\|_{\infty}E\left[|X - E[X]|^2\right] = \text{Var}(X).$$

On the other hand, the function $h(x) := x - E[X]$ is 1-Lipschitz and

$$E\left[(X - E[X])h(X)\right] = \text{Var}(X).$$

Thus, also (b) is proved.

$\square$

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 Chen et al. (2011). Throughout, Z will denote a standard normal random variable. Stein's method originated from Stein's seminal observation (see Stein, 1972) that a real-valued random variable X has the standard normal distribution, if and only if the identity

$$E\big[f'(X)\big] = E\big[Xf(X)\big]$$

holds for each, say, continuously differentiable function f with bounded derivative. 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*

$$f'(x) - xf(x) = h(x) - E[h(Z)] \tag{3.1}$$

and to use properties of the solution f and of W in order to bound the right hand side of

$$\Big|E\big[h(W)\big] - E\big[h(Z)\big]\Big| = \Big|E\big[f'(W) - Wf(W)\big]\Big|$$

rather than bounding the left hand side directly. For h as above, by f_h we denote the standard solution to the Stein equation (3.1) which is given by

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

$$= -e^{x^2/2} \int_{x}^{\infty} \big(h(t) - E[h(Z)]\big) e^{-t^2/2} dt. \tag{3.2}$$

Note that, generally, f_h is only differentiable and satisfies (3.1) 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

$$f_h'(x) := xf_h(x) + h(x) - E[h(Z)] \tag{3.3}$$

such that, by definition, f_h satisfies (3.1) 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, 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. Chen et al., 2011) that f_h is continuously differentiable and that both f_h and f_h' are Lipschitz-continuous with

$$\|f_h\|_\infty \le 1, \quad \|f_h'\|_\infty \le \sqrt{\frac{2}{\pi}} \quad \text{and} \quad \|f_h''\|_\infty \le 2. \tag{3.4}$$

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

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

its minimum Lipschitz constant. Note that if g is absolutely continuous, then $\|g_h'\|_\infty$ 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

$$R_g(x, y) := g(x + y) - g(x) - g'(x)y \tag{3.5}$$

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

$$\left| R_{f_h}(x, y) \right| = \left| f_h(x + y) - f_h(x) - f_h'(x)y \right| \leq y^2. \tag{3.6}$$

This follows from (3.4) via

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

$$\leq \|f_h''\|_\infty \left| \int_x^{x+y} |t - x| dt \right| = \frac{y^2 \|f_h''\|_\infty}{2} \leq y^2.$$

For $h = h_z$ we list the following properties of f_z: The function f_z has the representation

$$f_z(x) = \begin{cases} \frac{(1-\Phi(z))\Phi(x)}{\varphi(x)}, & x \leq z \\ \frac{\Phi(z)(1-\Phi(x))}{\varphi(x)}, & x > z. \end{cases} \tag{3.7}$$

Here, Φ denotes the standard normal distribution function and $\varphi := \Phi'$ the corresponding continuous density. It is easy to see from (3.7) 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 $x, u, v, z \in \mathbb{R}$

$$\left| (x + u)f_z(x + u) - (x + v)f_z(x + v) \right| \leq \left(|x| + \frac{\sqrt{2\pi}}{4} \right) (|u| + |v|). \tag{3.8}$$

Proofs of the above mentioned classic facts about the functions f_z can again be found in Chen et al. (2011), 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 (3.3) we define

$$f_z'(z) := z f_z(z) + 1 - \Phi(z) \tag{3.9}$$

such that $f = f_z$ satisfies (3.1) with $h = h_z$ for all $x \in \mathbb{R}$. Furthermore, with this definition, for all $x, z \in \mathbb{R}$ we have

$$|f_z'(x)| \leq 1. \tag{3.10}$$

The following quantitative version of the first order Taylor approximation of f_z has recently been proved by Lachi and Peccati (2015) and had already been used implicitly in Eichelsbacher and Thäle (2014). Using (3.9), for all $x, u, z \in \mathbb{R}$ we have

$$\left| R_{f_z}(x, u) \right| = \left| f_z(x + u) - f_z(x) - f_z'(x)u \right|$$

$$\leq \frac{u^2}{2} \left(|x| + \frac{\sqrt{2\pi}}{4} \right) + |u| \left(1_{\{x < z \leq x+u\}} + 1_{\{x+u \leq z < x\}} \right)$$

$$= \frac{u^2}{2} \left(|x| + \frac{\sqrt{2\pi}}{4} \right) + |u| 1_{\left\{ z - (u \vee 0) < x \leq z - (u \wedge 0) \right\}}, \tag{3.11}$$

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 2.7 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 Goldstein and Reinert (1997) it was proved that there exists a unique distribution for a random variable X^* such that for all Lipschitz continuous functions f the identity

$$E[Xf(X)] = \operatorname{Var}(X)E[f'(X^*)] \tag{3.12}$$

holds true. The distribution of X^* is called the *X-zero biased distribution* and the distributional transformation which maps $\mathcal{L}(X)$ to $\mathcal{L}(X^*)$ is called the *zero bias transformation*. It can be shown that (3.12) 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 Goldstein (2004) this heuristic was made precise by showing the inequality

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

where X is a mean zero random variable with $0 < \sigma^2 = E[X^2] = \operatorname{Var}(X) < \infty$, X^* having the X-zero biased distribution is defined on the same probability space as X and Z is standard normally distributed. For merely 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

$$E\big[(X - E[X])f(X)\big] = \operatorname{Var}(X)E\big[f'(X^{nz})\big].$$

Existence and uniqueness of the X-non-zero biased distribution immediately follow from Theorem 2.1 of Goldstein and Reinert (2005) (or Theorem 2.1 of Döbler, 2015 by letting $B(x) = x - E[X]$, there). Alternatively, letting $Y := X - E[X]$ 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 $\sigma_j^2 := \operatorname{Var}(X_j)$, $j = 1, \ldots, n$, $S := \sum_{j=1}^n X_j$ and $\sigma^2 := \operatorname{Var}(S) = \sum_{j=1}^n \sigma_j^2$. 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 Goldstein and Reinert (1997) 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 if 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 Goldstein and Reinert (1997) we know that

$$W^* := W - Y_I + Y_I^* = S - E[S] + \sum_{j=1}^{n} 1_{\{I=j\}} \left(E[X_j] - X_j + X_j^{nz} - E[X_j] \right)$$

$$= S - X_I + X_I^{nz} - E[S] = S^{nz} - E[S]$$

has the W-zero biased distribution, implying that S^{nz} has the S-non-zero biased distribution.

$\square$

4. Proof of Theorems 2.5 and 2.7

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 (3.2). Since f is a solution to (3.1), plugging in W and taking expectations yields

$$E[h(W)] - E[h(Z)] = E[f'(W) - Wf(W)] . \qquad (4.1)$$

As usual in Stein's method of normal approximation, the main task is to rewrite the term $E[Wf(W)]$ into a more tractable expression be exploiting the structure of W and using properties of f. From (2.4) we have

$$E[Wf(W)] = \frac{1}{\sigma} E[(S - aN)f(W)] + \frac{a}{\sigma} E[(N - \alpha)f(W)] =: T_1 + T_2 . \qquad (4.2)$$

For ease of notation, for $n \in \mathbb{Z}_+$ and M any $\mathbb{Z}_+$-valued random variable we let

$$S_n := \sum_{j=1}^{n} X_j, \quad W_n := \frac{S_n - \alpha a}{\sigma}, \quad S_M := \sum_{j=1}^{M} X_j \quad \text{and } W_M := \frac{S_M - \alpha a}{\sigma} ,$$

such that, in particular, $S = S_N$ and $W = W_N$. Using the decomposition

$$E[f'(W)] = \frac{\alpha c^2}{\sigma^2} E[f'(W)] + \frac{a^2 \gamma^2}{\sigma^2} E[f'(W)]$$

which is true by virtue of (2.3), from (4.1) and (4.2) we have

$$E[h(W)] - E[h(Z)] = E[f'(W)] - T_1 - T_2$$

$$= E\left[\frac{c^2 \alpha}{\sigma^2} f'(W) - \frac{1}{\sigma}(S - aN)f(W) \right]$$

$$+ E\left[\frac{a^2 \gamma^2}{\sigma^2} f'(W) - \frac{a}{\sigma}(N - \alpha)f(W) \right] =: E_1 + E_2 . \qquad (4.3)$$

We will bound the terms E_1 and E_2 seperately. Using the independence of N and $X_1, X_2, \ldots$ for T_1 we obtain:

$$T_1 = \frac{1}{\sigma} \sum_{n=0}^{\infty} P(N=n) E\big[(S_n - na) f(W_n)\big]$$

$$= \frac{1}{\sigma} \sum_{n=0}^{\infty} P(N=n) E\big[(S_n - na) g(S_n)\big], \qquad (4.4)$$

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 (4.4) and (3) we obtain that

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

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

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

$$T_1 = \frac{c^2 \alpha}{\sigma^2} E\Big[f'\Big(\frac{S_M^{nz} - \alpha a}{\sigma}\Big)\Big], \qquad (4.5)$$

where

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

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

$$(N, M), X_1, X_2, \ldots \text{ and } Y$$

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

$$\frac{S_M^{nz} - \alpha a}{\sigma} = \frac{S_M - \alpha a}{\sigma} + \frac{Y - X_1}{\sigma} = W_M + \frac{Y - X_1}{\sigma} =: W^*. \qquad (4.6)$$

Thus, from (4.6) and (4.5) we conclude that

$$T_1 = \frac{c^2 \alpha}{\sigma^2} E\big[f'(W^*)\big] \qquad (4.7)$$

and

$$E_1 = \frac{c^2 \alpha}{\sigma^2} E\big[f'(W) - f'(W^*)\big]. \qquad (4.8)$$

We would like to mention that if $a = 0$, then, by (4.7), 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 2.7 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

$$V := W_{N^s} - W_N = \frac{1}{\sigma}\left(S_{N^s} - S_N\right) \quad \text{and} \quad J := 1_{\{D \geq 0\}} = 1_{\{N^s \geq N\}}.$$

Now we turn to rewriting T_2. Using the independence of N and $X_1, X_2, \ldots$, and that of N^s and $X_1, X_2, \ldots$, respectively, $E[N] = \alpha$ and the defining equation (2.1) of the N-size biased distribution, we obtain from (4.2) that

$$
\begin{aligned}
T_2 &= \frac{a}{\sigma}E[(N-\alpha)f(W_N)] = \frac{\alpha a}{\sigma}E\big[f(W_{N^s}) - f(W_N)\big] \\
&= \frac{\alpha a}{\sigma}E\Big[1_{\{N^s \geq N\}}\big(f(W_{N^s}) - f(W_N)\big)\Big] + \frac{\alpha a}{\sigma}E\Big[1_{\{N^s < N\}}\big(f(W_{N^s}) - f(W_N)\big)\Big] \\
&= \frac{\alpha a}{\sigma}E\big[J\big(f(W_N + V) - f(W_N)\big)\big] - \frac{\alpha a}{\sigma}E\big[(1-J)\big(f(W_{N^s} - V) - f(W_{N^s})\big)\big] \\
&= \frac{\alpha a}{\sigma}E\big[JV f'(W_N)\big] + \frac{\alpha a}{\sigma}E\big[JR_f(W_N, V)\big] \\
&\quad + \frac{\alpha a}{\sigma}E\big[(1-J)V f'(W_{N^s})\big] - \frac{\alpha a}{\sigma}E\big[(1-J)R_f(W_{N^s}, -V)\big],
\end{aligned}
\tag{4.9}
$$

where R_f was defined in (3.5). Note that we have

$$JV = 1_{\{N^s \geq N\}}\frac{1}{\sigma}\sum_{j=N+1}^{N^s} X_j \quad \text{and} \quad W_N = \frac{\sum_{j=1}^{N} X_j - \alpha a}{\sigma}$$

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

$$
\begin{aligned}
E\big[JV \,\big|\, N\big] &= \frac{1}{\sigma}E\bigg[J\sum_{j=N+1}^{N^s} X_j \,\bigg|\, N\bigg] \\
&= \frac{1}{\sigma}E\bigg[JE\Big[\sum_{j=N+1}^{N^s} X_j \,\Big|\, N, N^s\Big] \,\bigg|\, N\bigg] \\
&= \frac{a}{\sigma}E\big[JD \,\big|\, N\big] = \frac{a}{\sigma}E\big[JD \,\big|\, N\big]
\end{aligned}
$$

we obtain that

$$
\begin{aligned}
\frac{\alpha a}{\sigma}E\big[JV f'(W_N)\big] &= \frac{\alpha a}{\sigma}E\Big[E\big[JV \,\big|\, N\big]E\big[f'(W_N) \,\big|\, N\big]\Big] \\
&= \frac{\alpha a^2}{\sigma^2}E\Big[E\big[JD \,\big|\, N\big]E\big[f'(W_N) \,\big|\, N\big]\Big] \\
&= \frac{\alpha a^2}{\sigma^2}E\Big[E\big[JD f'(W_N) \,\big|\, N\big]\Big] \\
&= \frac{\alpha a^2}{\sigma^2}E\Big[JD f'(W_N)\Big],
\end{aligned}
\tag{4.10}
$$

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^s} and $1_{\{D<0\}}V$ and also W_{N^s} and D are conditionally independent given N^s, one can show

$$\frac{\alpha a}{\sigma}E\big[(1-J)Vf'(W_{N^s})\big] = \frac{\alpha a^2}{\sigma^2}E\big[(1-J)Df'(W_{N^s})\big]. \qquad (4.11)$$

Hence, using that

$$\frac{\alpha a^2}{\sigma^2}E[D] = \frac{\alpha a^2}{\sigma^2}E[N^s - N] = \frac{\alpha a^2}{\sigma^2}\frac{\gamma^2}{\alpha} = \frac{a^2\gamma^2}{\sigma^2}$$

from (4.3), (4.9), (4.10) and (4.11) we obtain

$$
\begin{aligned}
E_2 &= \frac{\alpha a^2}{\sigma^2}E\Big[\big(E[D]-D\big)f'(W_N)\Big] + \frac{\alpha a^2}{\sigma^2}E\Big[(1-J)D\big(f'(W_N)-f'(W_{N^s})\big)\Big] \\
&\quad - \frac{\alpha a}{\sigma}E\big[JR_f(W_N,V)\big] + \frac{\alpha a}{\sigma}E\big[(1-J)R_f(W_{N^s},-V)\big] \\
&=: E_{2,1} + E_{2,2} + E_{2,3} + E_{2,4}\,.
\end{aligned}
\qquad (4.12)
$$

Using the conditional independence of D and W_N given N as well as the Cauchy-Schwarz inequality, we can estimate

$$
\begin{aligned}
|E_{2,1}| &= \frac{\alpha a^2}{\sigma^2}\Big|E\Big[E\big[D-E[D]\,|\,N\big]E\big[f'(W_N)\,|\,N\big]\Big]\Big| \\
&\leq \frac{\alpha a^2}{\sigma^2}\sqrt{\mathrm{Var}\big(E[D\,|\,N]\big)}\sqrt{E\Big[\big(E\big[f'(W_N)\,|\,N\big]\big)^2\Big]} \\
&\leq \frac{\alpha a^2}{\sigma^2}\|f'\|_\infty\sqrt{\mathrm{Var}\big(E[D\,|\,N]\big)}\,.
\end{aligned}
\qquad (4.13)
$$

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 und that it achieves the Wasserstein distance, i.e.

$$E|M-N| = E[M-N] = \frac{E[N^2]}{E[N]} - E[N] = \frac{\mathrm{Var}(N)}{E[N]} = \frac{\gamma^2}{\alpha} = d_{\mathcal{W}}(N,N^s)\,.$$

In order to bound E_1, we first derive an estimate for $E|W_M - W_N|$. We have

$$
\begin{aligned}
E\big[|W_M-W_N|\,\big|\,N,M\big] &= \frac{1}{\sigma}E\big[|S_M-S_N|\,\big|\,N,M\big] \leq \frac{|M-N|}{\sigma}E|X_1| \\
&\leq \frac{b(M-N)}{\sigma}
\end{aligned}
\qquad (4.14)
$$

and, hence,

$$
\begin{aligned}
E|W_M-W_N| &= E\Big[E\big[|W_M-W_N|\,\big|\,N,M\big]\Big] = \frac{1}{\sigma}E\big[|S_M-S_N|\,\big|\,N,M\big] \\
&\leq \frac{b}{\sigma}E[M-N] = \frac{b\gamma^2}{\sigma\alpha}\,.
\end{aligned}
\qquad (4.15)
$$

Then, using (3.4), (4.15) as well as the fact that the X_j are i.i.d., for E_1 we obtain that

$$
\begin{aligned}
|E_1| &= \frac{c^2\alpha}{\sigma^2}\left|E\left[f'(W_N) - f'\left(W_M + \frac{Y - X_1}{\sigma}\right)\right]\right| \\
&\leq \frac{2c^2\alpha}{\sigma^2}\left(E|W_N - W_M| + \sigma^{-1}E|Y - X_1|\right) \qquad\qquad (4.16) \\
&\leq \frac{2c^2\alpha}{\sigma^3}\left(\frac{b\gamma^2}{\alpha} + \frac{3}{2c^2}E|X_1 - E[X_1]|^3\right) \\
&= \frac{2c^2b\gamma^2}{\sigma^3} + \frac{3\alpha d^3}{\sigma^3}\,. \qquad\qquad\qquad\qquad\qquad (4.17)
\end{aligned}
$$

Here, we have used the inequality

$$
E|Y - X_1| = E\big|Y - E[X_1] - (X_1 - E[X_1])\big| \leq \frac{3}{2\,\mathrm{Var}(X_1)}E\big|X_1 - E[X_1]\big|^3, \quad (4.18)
$$

which follows from an analogous one in the zero-bias framework (see Chen et al., 2011) via the fact that $Y - E[X_1]$ has the $(X - E[X_1])$ - zero biased distribution. Similarly to (4.14) we obtain

$$
E\big[|V| \,\big|\, N, N^s\big] = E\big[|W_{N^s} - W_N| \,\big|\, N, N^s\big] \leq \frac{b|N^s - N|}{\sigma} = \frac{b|D|}{\sigma}
$$

which, together with (3.4) yields that

$$
\begin{aligned}
|E_{2,2}| &= \frac{\alpha a^2}{\sigma^2}\left|E\left[(1 - J)D\big(f'(W_N) - f'(W_{N^s})\big)\right]\right| \\
&\leq \frac{2\alpha a^2}{\sigma^2}E\big|(1 - J)D(W_N - W_{N^s})\big| \\
&= \frac{2\alpha a^2}{\sigma^2}E\Big[(1 - J)|D|E\big[|V| \,\big|\, N, N^s\big]\Big] \leq \frac{2\alpha a^2 b}{\sigma^3}E\big[(1 - J)D^2\big]. \qquad (4.19)
\end{aligned}
$$

We conclude the proof of the Wasserstein bounds by estimating $E_{2,3}$ and $E_{2,4}$. Note that by (3.6) we have

$$
\big|R_f(W_N, V)\big| \leq V^2 \quad \text{and} \quad \big|R_f(W_{N^s}, -V)\big| \leq V^2
$$

yielding

$$
|E_{2,3}| + |E_{2,4}| \leq \frac{\alpha|a|}{\sigma}E\Big[\big(1_{\{D\geq 0\}} + 1_{\{D<0\}}\big)V^2\Big] = \frac{\alpha|a|}{\sigma}E[V^2]. \qquad (4.20)
$$

Observe that

$$
\begin{aligned}
E[V^2] &= \frac{1}{\sigma^2}E\Big[\big(S_{N^s} - S_N\big)^2\Big] \\
&= \frac{1}{\sigma^2}\Big(\mathrm{Var}\big(S_{N^s} - S_N\big) + \big(E[S_{N^s} - S_N]\big)^2\Big) \qquad (4.21)
\end{aligned}
$$

and

$$
E\big[S_{N^s} - S_N\big] = E\Big[E\big[S_{N^s} - S_N \,\big|\, N, N^s\big]\Big] = aE[D] = \frac{a\gamma^2}{\alpha}\,. \qquad (4.22)
$$

Further, from the variance decomposition formula we obtain

$$
\begin{aligned}
\mathrm{Var}\big(S_{N^s} - S_N\big) &= E\big[\mathrm{Var}\big(S_{N^s} - S_N \,\big|\, N, N^s\big)\big] + \mathrm{Var}\big(E\big[S_{N^s} - S_N, \big|\, N, N^s\big]\big) \\
&= E\big[c^2|D|\big] + \mathrm{Var}(aD) = c^2E|D| + a^2\,\mathrm{Var}(D)\,.
\end{aligned}
$$

This together with (4.21) and (4.22) yields the bounds

$$E[V^2] = E\big[(W_{N^s} - W_N)^2\big] = \frac{1}{\sigma^2}\Big(c^2 E|D| + a^2 E[D^2]\Big) \tag{4.23}$$

$$\leq \frac{b^2}{\sigma^2} E[D^2]\,, \tag{4.24}$$

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 (3.4), (4.3), (4.12), (4.17), (4.19), (4.20) and (4.24). If $a = 0$, then E_1 can be bounded more accurately than we did before. Indeed, using (4.23) with $N^s = M$ and applying the Cauchy-Schwarz inequality give

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

as $c = b$ in this case. Plugging this into (4.16), we obtain

$$\begin{aligned}
|E_1| &\leq \frac{2c^2\alpha}{\sigma^2}\Big(E|W_M - W_N| + \sigma^{-1}E|Y - X_1|\Big)\\
&\leq \frac{2c^3\gamma\sqrt{\alpha}}{\sigma^3} + \frac{2c^2\alpha}{\sigma^3}E|Y - X_1|\\
&\leq \frac{2c^3\gamma\sqrt{\alpha}}{c^3\alpha^{3/2}} + \frac{3\alpha d^3}{c^3\alpha^{3/2}}\\
&= \frac{2\gamma}{\alpha} + \frac{3d^3}{c^3\sqrt{\alpha}}\,,
\end{aligned}$$

which is the Wasserstein bound claimed in Theorem 2.7.

Next, we proceed to the proof of the Berry-Esseen bounds in Theorems 2.5 and 2.7. 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 2.7 we will later assume that $M \geq N$. We 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 (4.8) and the fact that f solves the Stein equation (3.1) for $h = h_z$ we have

$$\begin{aligned}
E_1 &= \frac{c^2\alpha}{\sigma^2} E\big[f'(W) - f'(W^*)\big]\\
&= \frac{c^2\alpha}{\sigma^2} E\big[Wf(W) - W^*f(W^*)\big] + \frac{c^2\alpha}{\sigma^2}\Big(P(W \leq z) - P(W^* \leq z)\Big)\\
&=: E_{1,1} + E_{1,2}\,. \tag{4.25}
\end{aligned}$$

In order to bound $E_{1,1}$ we apply (3.8) to obtain

$$|E_{1,1}| \leq \frac{c^2\alpha}{\sigma^2} E\Big[|\tilde{V}|\Big(\frac{\sqrt{2\pi}}{4} + |W|\Big)\Big]\,. \tag{4.26}$$

Using (4.23), (4.24) and (4.18) we have

$$E|\tilde{V}| \le E|V| + \sigma^{-1}E|Y - X_1| \le \sqrt{E[V^2]} + \frac{3d^3}{2\sigma c^2}$$

$$= \frac{1}{\sigma}\sqrt{c^2 E|D| + a^2 E[D^2]} + \frac{3d^3}{2\sigma c^2} \tag{4.27}$$

$$\le \frac{b}{\sigma}\sqrt{E[D^2]} + \frac{3d^3}{2\sigma c^2}\,. \tag{4.28}$$

Furthermore, using independence of W and Y, we have

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

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

$$\le \frac{d^3}{2c^2}\sqrt{E[W^2]} + \sqrt{\mathrm{Var}(X_1)E[W^2]} = \frac{d^3}{2c^2} + c\,. \tag{4.29}$$

Finally, we have

$$E|VW| \le \sqrt{E[V^2]}\sqrt{E[W^2]} = \frac{1}{\sigma}\sqrt{c^2 E|D| + a^2 E[D^2]} \tag{4.30}$$

$$\le \frac{b}{\sigma}\sqrt{E[D^2]}\,. \tag{4.31}$$

From (4.26), (4.27), (4.28), (4.29), (4.30) and (4.31) we conclude that

$$|E_{1,1}| \le \frac{c^2\alpha}{\sigma^2}\Big(\frac{\sqrt{2\pi}}{4\sigma}\sqrt{c^2 E|D| + a^2 E[D^2]} + \frac{3d^3\sqrt{2\pi}}{8c^2\sigma} + \frac{d^3}{2c^2\sigma} + \frac{c}{\sigma}$$

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

$$= \frac{c^2\alpha(\sqrt{2\pi} + 4)}{4\sigma^3}\sqrt{c^2 E|D| + a^2 E[D^2]} + \frac{d^3\alpha(3\sqrt{2\pi} + 4)}{8\sigma^3} + \frac{c^3\alpha}{\sigma^3} \tag{4.32}$$

$$\le \frac{(\sqrt{2\pi} + 4)bc^2\alpha}{4\sigma^3}\sqrt{E[D^2]} + \frac{d^3\alpha(3\sqrt{2\pi} + 4)}{8\sigma^3} + \frac{c^3\alpha}{\sigma^3} =: B_1\,. \tag{4.33}$$

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 $C_{\mathcal{K}}$ the Berry-Esseen constant for sums of i.i.d. random variables with finite third moment. It is known from Shevtsova (2011) that

$$C_{\mathcal{K}} \le 0.4748\,.$$

In particular, $2C_{\mathcal{K}} \le 1$, which is substituted for $2C_{\mathcal{K}}$ in the statements of Theorems 2.5 and 2.7. However, we prefer keeping the dependence of the bounds on $C_{\mathcal{K}}$ explicit within the proof.

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

$$\big|P(W^* \le z) - P(W_{N^s} \le z)\big| \le \frac{1}{\sqrt{\alpha}}\Big(\frac{7}{2}\sqrt{2} + 2\Big)\frac{d^3}{c^3} \quad and \tag{4.34}$$

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

$$+ \frac{2C_{\mathcal{K}}d^3}{c^3}E\big[1_{\{D\ge 0\}}N^{-1/2}1_{\{N\ge 1\}}\big] \tag{4.35}$$

$$+ \frac{1}{\sqrt{\alpha}}\Big(\frac{b}{c\sqrt{2\pi}}\sqrt{E\big[D^2 1_{\{D<0\}}\big]} + \frac{2C_{\mathcal{K}}d^3}{c^3}\sqrt{P(D < 0)}\Big).$$

If $a = 0$ and $D \geq 0$, then for all $z \in \mathbb{R}$

$$\left|P(W_{N^s} \leq z) - P(W \leq z)\right| \leq P(N = 0) + \frac{2C_{\mathcal{K}}d^3}{c^3}E\left[N^{-1/2}1_{\{N \geq 1\}}\right]$$

$$+ \frac{1}{\sqrt{2\pi}}E\left[\sqrt{D}N^{-1/2}1_{\{N \geq 1\}}\right] \tag{4.36}$$

$$\leq P(N = 0) + \left(\frac{2C_{\mathcal{K}}d^3}{c^3} + \frac{\gamma}{\sqrt{\alpha}\sqrt{2\pi}}\right)\sqrt{E\left[1_{\{N \geq 1\}}N^{-1}\right]}. \tag{4.37}$$

Applying the triangle inequality to Lemma 4.1 yields the following bounds on $E_{1,2}$: In the most general situation (Theorem 2.5 and Remark 2.6 (b)) we have

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

$$+ \frac{2C_{\mathcal{K}}d^3\alpha}{c\sigma^2}E\left[1_{\{D \geq 0\}}N^{-1/2}1_{\{N \geq 1\}}\right] + \frac{cb\sqrt{\alpha}}{\sigma^2\sqrt{2\pi}}\sqrt{E\left[D^21_{\{D < 0\}}\right]}$$

$$+ \frac{\alpha C_{\mathcal{K}}d^3}{c\sigma^2}\sqrt{P(D < 0)} =: B_2. \tag{4.38}$$

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

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

$$+ \frac{1}{\sqrt{2\pi}}E\left[\sqrt{D}N^{-1/2}1_{\{N \geq 1\}}\right]$$

$$\leq \left(\frac{7}{2}\sqrt{2} + 2\right)\frac{d^3}{c^3\sqrt{\alpha}} + P(N = 0) + \left(\frac{2C_{\mathcal{K}}d^3}{c^3} + \frac{\gamma}{\sqrt{\alpha}\sqrt{2\pi}}\right)\sqrt{E\left[1_{\{N \geq 1\}}N^{-1}\right]}. \tag{4.39}$$

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 (4.12).

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

$$E\left[J|V|1_{\{z-(V \vee 0) < W \leq z-(V \wedge 0)\}}\right] \leq \frac{b}{\sigma}\sqrt{P(N = 0)}\sqrt{E[JD^2]}$$

$$+ \frac{b^2}{c\sigma\sqrt{2\pi}}E\left[JD^21_{\{N \geq 1\}}N^{-1/2}\right] + \frac{2C_{\mathcal{K}}d^3b}{c^3\sigma}E\left[JD1_{\{N \geq 1\}}N^{-1/2}\right], \tag{4.40}$$

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

$$+ \frac{2C_{\mathcal{K}}bd^3}{c^3\sigma\sqrt{\alpha}}\sqrt{E\left[(1-J)D^2\right]} \quad and \tag{4.41}$$

$$E\left[(1-J)|D|1_{\{z+(V \wedge 0) < W_{N^s} \leq z+(V \vee 0)\}}\right] \leq \frac{b}{c\sqrt{2\pi}}E\left[(1-J)D^2(N^s)^{-1/2}\right]$$

$$+ \frac{2C_{\mathcal{K}}d^3}{c^3\sqrt{\alpha}}\sqrt{E\left[(1-J)D^2\right]}. \tag{4.42}$$

Next, we derive a bound on $E_{2,2}$. Since f solves the Stein equation (3.1) for $h = h_z$ we have

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

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

$$=: E_{2,2,1} + E_{2,2,2}. \tag{4.43}$$

Using

$$W_N = W_{N^s} - V$$

and Lemma 5.1, we obtain from (4.42) that

$$|E_{2,2,2}| \leq \frac{\alpha a^2}{\sigma^2} E\big[1_{\{D<0\}}|D|1_{\{z+(V\wedge 0)<W_{N^s}\leq z+(V\vee 0)\}}\big]$$

$$\leq \frac{\alpha a^2 b}{\sigma^2 c\sqrt{2\pi}} E\big[1_{\{D<0\}} D^2 (N^s)^{-1/2}\big] + \frac{2C_K d^3 a^2 \sqrt{\alpha}}{c^3 \sigma^2} \sqrt{E\big[1_{\{D<0\}} D^2\big]}$$

$$=: B_4. \tag{4.44}$$

As to $E_{2,2,1}$, from (3.8) we have

$$|E_{2,2,1}| \leq \frac{\alpha a^2}{\sigma^2} E\Big[(1-J)|DV|\Big(|W_{N^s}| + \frac{\sqrt{2\pi}}{4}\Big)\Big] \tag{4.45}$$

As

$$E\big[|V| \,\big|\, N, N^s\big] \leq \sqrt{E\big[V^2 \,\big|\, N, N^s\big]} = \frac{1}{\sigma}\sqrt{c^2|D| + a^2 D^2} \leq \frac{b}{\sigma}|D|, \tag{4.46}$$

by conditioning, we see

$$E\big[(1-J)|DV|\big] = E\Big[(1-J)|D|E\big[|V| \,\big|\, N, N^s\big]\Big] \leq \frac{b}{\sigma} E\big[(1-J)D^2\big]. \tag{4.47}$$

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

$$E\big|(1-J)DVW_{N^s}\big| = E\Big[E\big[|(1-J)DVW_{N^s}| \,\big|\, N^s\big]\Big]$$

$$= E\Big[E\big[(1-J)|DV| \,\big|\, N^s\big] E\big[|W_{N^s}| \,\big|\, N^s\big]\Big]$$

$$\leq \sqrt{E\Big[\big(E[(1-J)|DV| \,|\, N^s]\big)^2\Big]} \sqrt{E\Big[\big(E[|W_{N^s}| \,|\, N^s]\big)^2\Big]}$$

$$\leq \frac{b}{\sigma} \sqrt{E\Big[\big(E[(1-J)D^2 \,|\, N^s]\big)^2\Big]} \sqrt{E\big[W_{N^s}^2\big]}, \tag{4.48}$$

where we have used the conditional Jensen inequality, (4.46) and

$$E\big[(1-J)|DV| \,\big|\, N^s\big] = E\Big[(1-J)|D|E\big[|V| \,\big|\, N, N^s\big] \,\Big|\, N^s\Big]$$

to obtain the last inequality. Using the defining relation (2.1) of the size-biased distribution one can easily show that

$$E\big[W_{N^s}^2\big] = \frac{1}{\sigma^2} E\big[c^2 N^s + a^2(N^s - \alpha)^2\big] = \frac{c^2\beta^2 + a^2\big(\delta^3 - 2\alpha\beta^2 + \alpha^3\big)}{\alpha\sigma^2}, \tag{4.49}$$

which, together with (4.45), (4.47) and (4.48) yields that

$$
\begin{aligned}
|E_{2,2,1}| \leq {} & \frac{\alpha a^2 b\sqrt{2\pi}}{4\sigma^3} E\big[1_{\{D<0\}}D^2\big] \\
& + a^2 b \frac{c^2\beta^2 + a^2\left(\delta^3 - 2\alpha\beta^2 + \alpha^3\right)}{\sigma^5} \sqrt{E\Big[\big(E\big[1_{\{D<0\}}D^2 \,\big|\, N^s\big]\big)^2\Big]} \\
=: {} & B_5 \,.
\end{aligned}
\tag{4.50}
$$

It remains to bound the quantities $E_{2,3}$ and $E_{2,4}$ from (4.12) for $f = f_z$. From (3.11) we have

$$
\begin{aligned}
|E_{2,3}| = {} & \frac{\alpha|a|}{\sigma}\Big|E\big[1_{\{D\geq 0\}}R_f(W,V)\big]\Big| \\
\leq {} & \frac{\alpha|a|}{2\sigma} E\Big[JV^2\Big(|W| + \frac{\sqrt{2\pi}}{4}\Big)\Big] \\
& + \frac{\alpha|a|}{\sigma} E\Big[J|V|1_{\{z-(V\vee 0)<W\leq z-(V\wedge 0)\}}\Big] =: R_{1,1} + R_{1,2} \,.
\end{aligned}
\tag{4.51}
$$

Similarly to (4.23) we obtain

$$
\begin{aligned}
E\big[JV^2\big] = {} & \frac{1}{\sigma^2}\Big(c^2 E[JD] + a^2 E\big[JD^2\big]\Big) \\
\leq {} & \frac{b^2}{\sigma^2} E\big[JD^2\big]
\end{aligned}
\tag{4.52}
$$

from

$$
\begin{aligned}
E\big[JV^2 \,\big|\, N, N^s\big] = JE\big[V^2 \,\big|\, N, N^s\big] = {} & \frac{J}{\sigma^2}\Big(c^2|D| + a^2 D^2\Big) \\
= {} & \frac{1}{\sigma^2}\Big(c^2 JD + a^2 JD^2\Big) \,.
\end{aligned}
\tag{4.53}
$$

Also, recall that the random variables

$$
JV^2 = \sigma^{-1}1_{\{N^s\geq N\}}\Big(\sum_{j=N+1}^{N^s} X_j\Big)^2 \quad \text{and} \quad W_N = \sigma^{-1}\Big(\sum_{j=1}^{N} X_j - \alpha a\Big)
$$

are conditionally independent given N. Hence, using the Cauchy-Schwarz inequality

$$
\begin{aligned}
E\big[JV^2|W_N|\big] = E\Big[E\big[JV^2|W_N|\,\big|\,N\big]\Big] = {} & E\Big[E\big[JV^2 \,\big|\, N\big]E\big[|W_N|\,\big|\,N\big]\Big] \\
\leq {} & \sqrt{E\Big[\big(E\big[JV^2 \,\big|\, N\big]\big)^2\Big]}\sqrt{E\Big[\big(E\big[|W_N|\,\big|\,N\big]\big)^2\Big]} \,.
\end{aligned}
\tag{4.54}
$$

From (4.53) and $D^2 \geq |D|$ we conclude that

$$
E\big[JV^2 \,\big|\, N\big] = \frac{1}{\sigma^2}\Big(c^2 E[JD \,|\, N] + a^2 E\big[JD^2 \,\big|\, N\big]\Big) \leq \frac{b^2}{\sigma^2} E\big[JD^2 \,\big|\, N\big] \,.
\tag{4.55}
$$

Furthermore, by the conditional version of Jensen's inequality we have

$$
E\Big[\big(E\big[|W_N|\,\big|\,N\big]\big)^2\Big] \leq E\Big[E\big[W_N^2 \,\big|\, N\big]\Big] = E\big[W_N^2\big] = 1 \,.
\tag{4.56}
$$

Thus, from (4.54), (4.55) and (4.56) we see that

$$
E\big[JV^2|W_N|\big] \leq \frac{b^2}{\sigma^2}\sqrt{E\Big[\big(E\big[JD^2 \,\big|\, N\big]\big)^2\Big]} \,.
\tag{4.57}
$$

Hence, (4.51), (4.52) and (4.57) yield

$$R_{1,1} \le \frac{\alpha|a|b^2}{2\sigma^3}\sqrt{E\left[\left(E\left[JD^2 \mid N\right]\right)^2\right]} + \frac{\alpha|a|b^2\sqrt{2\pi}}{8\sigma^3}E[JD^2]. \qquad (4.58)$$

Finally, from (4.51), (4.58) and (4.40) we get

$$\begin{aligned}
|E_{2,3}| &\le \frac{\alpha|a|b^2}{2\sigma^3}\sqrt{E\left[\left(E\left[1_{\{D\ge0\}}D^2 \mid N\right]\right)^2\right]} + \frac{\alpha|a|b^2\sqrt{2\pi}}{8\sigma^3}E[1_{\{D\ge0\}}D^2] \\
&\quad + \frac{\alpha|a|b}{\sigma^2}\sqrt{P(N=0)}\sqrt{E[1_{\{D\ge0\}}D^2]} \\
&\quad + \frac{\alpha|a|b^2}{c\sigma^2\sqrt{2\pi}}E\left[1_{\{D\ge0\}}D^2 1_{\{N\ge1\}}N^{-1/2}\right] \\
&\quad + \frac{2C_{\mathcal{K}}d^3\alpha|a|b}{\sigma^2}E\left[1_{\{D\ge0\}}D 1_{\{N\ge1\}}N^{-1/2}\right] =: B_6. \qquad (4.59)
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
|E_{2,4}| &= \frac{\alpha|a|}{\sigma}\left|E\left[1_{\{D<0\}}R_f(W_{N^s}, -V)\right]\right| \\
&\le \frac{\alpha|a|}{2\sigma}E\left[(1-J)V^2\left(W_{N^s} + \frac{\sqrt{2\pi}}{4}\right)\right] \\
&\quad + \frac{\alpha|a|}{\sigma}E\left[(1-J)|V|1_{\{z+(V\wedge0)<W_{N^s}\le z+(V\vee0)\}}\right] =: R_{2,1} + R_{2,2}. \qquad (4.60)
\end{aligned}$$

Analogously to the above we obtain

$$\begin{aligned}
E\left[(1-J)V^2\right] &= \frac{1}{\sigma^2}\left(c^2 E\left[(1-J)|D|\right] + a^2 E\left[(1-J)D^2\right]\right) \\
&\le \frac{b^2}{\sigma^2}E\left[(1-J)D^2\right] \quad \text{and} \qquad\qquad (4.61) \\
E\left[(1-J)V^2 \mid N^s\right] &= \frac{1}{\sigma^2}\left(c^2 E\left[(1-J)|D| \mid N^s\right] + a^2 E\left[(1-J)D^2 \mid N^s\right]\right) \\
&\le \frac{b^2}{\sigma^2}E\left[(1-J)D^2 \mid N^s\right].
\end{aligned}$$

Using these as well as the conditional independence of $(1-J)V^2$ and W_{N^s} given N^s, one has

$$E\left[(1-J)V^2|W_{N^s}|\right] \le \frac{b^2}{\sigma^2}\sqrt{E\left[\left(E\left[(1-J)D^2 \mid N^s\right]\right)^2\right]}\sqrt{E\left[W_{N^s}^2\right]}. \qquad (4.62)$$

Combining (4.49) and (4.62) we obtain

$$\begin{aligned}
E\left[(1-J)V^2|W_{N^s}|\right] &\le \frac{b^2}{\sigma^2}\sqrt{E\left[\left(E\left[(1-J)D^2 \mid N^s\right]\right)^2\right]} \\
&\quad \left(\frac{c^2\beta^2 + a^2\left(\delta^3 - 2\alpha\beta^2 + \alpha^3\right)}{\alpha\sigma^2}\right)^{1/2}. \qquad (4.63)
\end{aligned}$$

Thus, from (4.60), (4.61), (4.63) and (4.41) we conclude

$$
\begin{aligned}
|E_{2,4}| \leq \; & \frac{\alpha|a|b^2\sqrt{2\pi}}{8\sigma^3} E\big[1_{\{D<0\}}D^2\big] + \frac{\alpha|a|b^2}{2\sigma^3}\sqrt{E\Big[\big(E\big[1_{\{D<0\}}D^2 \mid N^s\big]\big)^2\Big]} \\
& \cdot \left(\frac{c^2\beta^2 + a^2\big(\delta^3 - 2\alpha\beta^2 + \alpha^3\big)}{\alpha\sigma^2}\right)^{1/2} \\
& + \frac{\alpha|a|b^2}{\sigma^2 c\sqrt{2\pi}} E\big[1_{\{D<0\}}D^2(N^s)^{-1/2}\big] + \frac{\sqrt{\alpha}|a|2C_{\mathcal{K}}bd^3}{\sigma^2}\sqrt{E\big[1_{\{D<0\}}D^2\big]} \\
=: \; & B_7\,.
\end{aligned}
\tag{4.64}
$$

The Berry-Esseen bound stated in Remark 2.6 (b) follows from (4.3), (4.25), (4.33), (4.38), (4.12), (4.13), (3.10), (4.43), (4.44), (4.50), (4.59) and (4.64). This immediately yields the Berry-Esseen bound presented in Theorem 2.5 (b) because

$$
B_2 = B_4 = B_5 = B_7 = 0
$$

in this case. In order to obtain the Kolmogorov bound in Theorem 2.7, again, we choose M such that $M \geq N$ and use the bounds (4.32) and (4.39) instead. The result then follows from (4.3) and (4.25).

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*

$$
1_{\{x+u\leq z\}} - 1_{\{x+v\leq z\}} = 1_{\{z-v<x\leq z-u\}} - 1_{\{z-u<x\leq z-v\}} \quad \text{and}
$$

$$
\big|1_{\{x+u\leq z\}} - 1_{\{x+v\leq z\}}\big| = 1_{\{z-u\vee v<x\leq z-u\wedge v\}}\,.
$$

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

$$
P(t < W_n \leq u) \leq \frac{\sigma(u-t)}{c\sqrt{2\pi}\sqrt{n}} + \frac{2C_{\mathcal{K}}d^3}{c^3\sqrt{n}}\,.
$$

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

$$
P(r < X \leq s) \leq \frac{s-r}{\sqrt{2\pi}} + 2d_{\mathcal{K}}(X, Z)\,.
\tag{5.1}
$$

A similar result was used in Peköz and Röllin (2011) in the framework of exponential approximation. Now, for given $t < u$ and $n \geq 1$ by (5.1) and the Berry-Esseen Theorem we have

$$
\begin{aligned}
P(t < W_n \leq u) = \; & P\left(\frac{\sigma t + a(\alpha - n)}{c\sqrt{n}} < \frac{S_n - na}{c\sqrt{n}} \leq \frac{\sigma u + a(\alpha - n)}{c\sqrt{n}}\right) \\
\leq \; & \frac{\sigma(u-t)}{c\sqrt{2\pi}\sqrt{n}} + \frac{2C_{\mathcal{K}}d^3}{c^3\sqrt{n}}\,.
\end{aligned}
$$

$\square$

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 Chen et al. (2011), 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 Chen et al. (2011):

Lemma 5.4. *Let $Y_1, \ldots, Y_n$ be independent mean zero random variables such that*

$$\sum_{j=1}^{n} E[Y_j^2] = 1 \quad and \quad \zeta := \sum_{j=1}^{n} E|Y_j|^3 < \infty,$$

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 4.1: We first prove (4.34). Define

$$W_{N^s}^{(1)} := W_{N^s} - \sigma^{-1}X_1 = \frac{1}{\sigma}\left(\sum_{j=2}^{N^s} X_j - \alpha a\right)$$

such that

$$W_{N^s} = W_{N^s}^{(1)} + \sigma^{-1}X_1 \quad and \quad W^* = W_{N^s}^{(1)} + \sigma^{-1}Y.$$

Then, using Lemma 5.1 we have

$$\begin{aligned}
&\left| P(W^* \leq z) - P(W_{N^s} \leq z) \right| \\
&= \left| P(W_{N^s}^{(1)} + \sigma^{-1}Y \leq z) - P(W_{N^s}^{(1)} + \sigma^{-1}X_1 \leq z) \right| \\
&\leq P\left(z - \sigma^{-1}(X_1 \vee Y) < W_{N^s}^{(1)} \leq z - \sigma^{-1}(X_1 \wedge Y) \right) \\
&= E\left[P\left(\frac{\sigma z - (X_1 \vee Y) + a(\alpha - N^s + 1)}{c\sqrt{N^s}} < \sum_{j=2}^{N^s} \left(\frac{X_j - a}{c\sqrt{N^s}} \right) \right. \right. \\
&\qquad\qquad \left. \left. \leq \frac{\sigma z - (X_1 \wedge Y) + a(\alpha - N^s + 1)}{c\sqrt{N^s}} \,\middle|\, N^s \right) \right].
\end{aligned}$$

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 probabilty giving the bound

$$\left| P(W^* \leq z) - P(W_{N^s} \leq z) \right| \leq E\left[\frac{\sqrt{2}|Y - X_1|}{c\sqrt{N^s}} + \frac{2(\sqrt{2}+1)d^3}{c^3\sqrt{N^s}} \right]. \tag{5.2}$$

Noting that (X_1, Y) and N^s are independent and using (4.18) again, we obtain

$$E\left[\frac{|Y - X_1|}{\sqrt{N^s}} \right] \leq \frac{3}{2c^2}d^3 E\left[(N^s)^{-1/2}\right] \leq \frac{3d^3}{2c^2\sqrt{\alpha}}, \tag{5.3}$$

as

$$E\left[(N^s)^{-1/2}\right] = \frac{E[\sqrt{N}]}{E[N]} \leq \frac{\sqrt{E[N]}}{E[N]} = \frac{1}{\sqrt{\alpha}} \tag{5.4}$$

by (2.1) and Jensen's inequality. From (5.2), (5.3) and (5.4) the bound (4.34) follows.

Next we prove (4.35). Using Lemma 5.1 we obtain

$$
\begin{aligned}
\left|P(W_{N^s} \leq z) - P(W \leq z)\right| &= \left|E\left[J(1_{\{W+V \leq z\}} - 1_{\{W \leq z\}})\right]\right. \\
&\quad \left. - E\left[(1-J)(1_{\{W_{N^s}-V \leq z\}} - 1_{\{W_{N^s} \leq z\}})\right]\right| \\
&\leq E\left[J1_{\{z-(V \vee 0) < W \leq z-(V \wedge 0)\}}\right] \\
&\quad + E\left[(1-J)1_{\{z+(V \wedge 0) < W_{N^s} \leq z+(V \vee 0)\}}\right] \\
&=: A_1 + A_2 .
\end{aligned}
\tag{5.5}
$$

To bound A_1 we write

$$
\begin{aligned}
A_1 &= \sum_{n=0}^{\infty} E\left[J1_{\{N=n\}}1_{\{z-(V \vee 0) < W \leq z-(V \wedge 0)\}}\right] \\
&= \sum_{n=0}^{\infty} P\left(z-(V \vee 0) < W \leq z-(V \wedge 0) \,\middle|\, D \geq 0, N = n\right) \cdot P\left(D \geq 0, N = n\right) .
\end{aligned}
\tag{5.6}
$$

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 $n \geq 1$:

$$
\begin{aligned}
&P\left(z-(V \vee 0) < W \leq z-(V \wedge 0) \,\middle|\, D \geq 0, N = n\right) \\
&= P\left(z-(V \vee 0) < W_n \leq z-(V \wedge 0) \,\middle|\, D \geq 0, N = n\right) \\
&\leq E\left[\frac{\sigma|V|}{c\sqrt{2\pi}\sqrt{n}} + \frac{2C_{\mathcal{K}}d^3}{c^3\sqrt{n}} \,\middle|\, D \geq 0, N = n\right]
\end{aligned}
\tag{5.7}
$$

From (5.6) and (5.7) we thus have

$$
\begin{aligned}
A_1 &\leq P(N = 0) + \sum_{n=1}^{\infty} E\left[\frac{\sigma|V|}{c\sqrt{2\pi}\sqrt{n}} + \frac{2C_{\mathcal{K}}d^3}{c^3\sqrt{n}} \,\middle|\, D \geq 0, N = n\right]P\left(D \geq 0, N = n\right) \\
&= P(N = 0) + \sum_{n=1}^{\infty} E\left[1_{\{D \geq 0, N=n\}}\left(\frac{\sigma|V|}{c\sqrt{2\pi}\sqrt{n}} + \frac{2C_{\mathcal{K}}d^3}{c^3\sqrt{n}}\right)\right] \\
&= P(N = 0) + E\left[J1_{\{N \geq 1\}}\left(\frac{\sigma|V|}{c\sqrt{2\pi}\sqrt{N}} + \frac{2C_{\mathcal{K}}d^3}{c^3\sqrt{N}}\right)\right]
\end{aligned}
\tag{5.8}
$$

Now note that

$$
\begin{aligned}
E\left[J1_{\{N \geq 1\}}|V|N^{-1/2}\right] &= E\left[J1_{\{N \geq 1\}}N^{-1/2}E\left[|V| \,\middle|\, N, N^s\right]\right] \\
&\leq E\left[J1_{\{N \geq 1\}}N^{-1/2}\sqrt{E\left[V^2 \mid N, N^s\right]}\right] \\
&= \frac{1}{\sigma}E\left[J1_{\{N \geq 1\}}N^{-1/2}\sqrt{c^2D + a^2D^2}\right] \\
&\leq \frac{b}{\sigma}E\left[JD1_{\{N \geq 1\}}N^{-1/2}\right] .
\end{aligned}
\tag{5.9}
\tag{5.10}
$$

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 (5.6) we obtain

$$A_2 = \sum_{m=1}^{\infty} P\big(z + (V \wedge 0) < W_{N^s} \leq z + (V \vee 0) \,\big|\, D < 0, N^s = m\big)$$
$$\cdot P\big(D < 0, N^s = m\big).$$

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

$$\mathcal{L}(W_{N^s} | N^s = m, D < 0) = \mathcal{L}(W_m)$$

in the same manner as (5.8) we find

$$A_2 \leq E\left[(1 - J)\left(\frac{\sigma|V|}{c\sqrt{2\pi}\sqrt{N^s}} + \frac{2C_{\mathcal{K}}d^3}{c^3\sqrt{N^s}}\right)\right]. \tag{5.11}$$

Using (2.1) we have

$$E\big[(N^s)^{-1}\big] = \frac{1}{E[N]} = \frac{1}{\alpha}. \tag{5.12}$$

Thus, from the Cauchy-Schwarz inequality and (5.12) we obtain

$$E\left[(1 - J)\frac{|V|}{\sqrt{N^s}}\right] \leq \sqrt{E\big[(N^s)^{-1}\big]}\sqrt{E\big[(1 - J)V^2\big]}$$
$$= \frac{1}{\sigma\sqrt{\alpha}}\sqrt{c^2 E\big[|D|(1 - J)\big] + a^2 E\big[D^2(1 - J)\big]}$$
$$\leq \frac{b}{\sigma\sqrt{\alpha}}\sqrt{E\big[D^2(1 - J)\big]}. \tag{5.13}$$

Similarly, we have

$$E\left[\frac{1 - J}{\sqrt{N^s}}\right] \leq \sqrt{P(D < 0)}\sqrt{E\big[(N^s)^{-1}\big]} = \frac{\sqrt{P(D < 0)}}{\sqrt{\alpha}}. \tag{5.14}$$

Thus, from (5.8), (5.10), (5.11), (5.13) and (5.14) we see that $A_1 + A_2$ is bounded from above by the right hand side of (4.35). Using (5.8) and (5.9) instead gives the bounds (4.36) and (4.37).

<div style="text-align:right">□</div>

Proof of Lemma 4.2: We only prove (4.40), the proofs of (4.41) and (4.42) being similar and easier. By the definition of conditional expectation given an event, we have

$$E\big[J|V|1_{\{z-(V\vee0)<W_N\leq z-(V\wedge0)\}}\big]$$
$$= \sum_{n=0}^{\infty} E\big[1_{\{N=n,D\geq0\}}|V|1_{\{z-(V\vee0)<W_n\leq z-(V\wedge0)\}}\big] = E\big[1_{\{N=0\}}J|V|\big]$$
$$+ \sum_{n=1}^{\infty} E\big[|V|1_{\{z-(V\vee0)<W_n\leq z-(V\wedge0)\}} \,\big|\, N = n, D \geq 0\big] \cdot P(N = n, D \geq 0). \tag{5.15}$$

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

$$
E\big[|V|1_{\{z-(V\vee 0)<W_n\leq z-(V\wedge 0)\}}\,\big|\,N=n,D\geq 0\big]
$$
$$
= E\Big[|V|\Big(\frac{\sigma|V|}{c\sqrt{2\pi}\sqrt{n}}+\frac{2C_\mathcal{K}d^3}{c^3\sqrt{n}}\Big)\,\Big|\,N=n,D\geq 0\Big] \tag{5.16}
$$

Combining (5.15) and (5.16) we get

$$
E\big[J|V|1_{\{z-(V\vee 0)<W_N\leq z-(V\wedge 0)\}}\big]\leq E\big[1_{\{N=0\}}J|V|\big]
$$
$$
+\sum_{n=1}^{\infty}E\Big[|V|\Big(\frac{\sigma|V|}{c\sqrt{2\pi}\sqrt{n}}+\frac{2C_\mathcal{K}d^3}{c^3\sqrt{n}}\Big)\,\Big|\,N=n,D\geq 0\Big]\cdot P(N=n,D\geq 0)
$$
$$
= E\big[1_{\{N=0\}}J|V|\big]+E\Big[1_{\{N\geq 1\}}J|V|\Big(\frac{\sigma|V|}{c\sqrt{2\pi}\sqrt{N}}+\frac{2C_\mathcal{K}d^3}{c^3\sqrt{N}}\Big)\Big]. \tag{5.17}
$$

Using Cauchy-Schwarz as well as

$$
E\big[JV^2\big] = E\Big[JE\big[V^2\,\big|\,N,N^s\big]\Big]\leq \frac{b^2}{\sigma^2}E\big[JD^2\big]
$$

we obtain

$$
E\big[1_{\{N=0\}}J|V|\big]\leq \frac{b}{\sigma}\sqrt{P(N=0)}\sqrt{E\big[JD^2\big]}. \tag{5.18}
$$

Analogously to (5.10) one can show that

$$
E\big[1_{\{N\geq 1\}}N^{-1/2}JV^2\big]\leq \frac{b^2}{\sigma^2}E\big[1_{\{N\geq 1\}}N^{-1/2}JD^2\big]. \tag{5.19}
$$

Hence, bound (4.40) follows from (5.17), (5.18), (5.10) and (5.19).

<div align="right">□</div>

Acknowledgements

References

R. Arratia and L. Goldstein. Size bias, sampling, the waiting time paradox, and inifinite divisibility: when is the increment independent? *ArXiv Mathematics e-prints* (2010). arXiv:1007.3910.

R. Arratia, L. Goldstein and F. Kochman. Size bias for one and all. *ArXiv Mathematics e-prints* (2013). arXiv:1308.2729.

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). ISBN 978-3-642-15006-7. MR2732624.

Louis H.Y. Chen. Poisson approximation for dependent trials. *Ann. Probability* **3** (3), 534–545 (1975). MR0428387.

C. Döbler. On rates of convergence and berry-esseen bounds for random sums of centered random variables with finite third moments. *ArXiv Mathematics e-prints* (2012). arXiv:1212.5401.

C. Döbler. Distributional transformations without orthogonality relations. *J. Theoret. Probab.* page to appear (2015). DOI: 10.1214/EJP.v18-2104.

R.M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge (2002). ISBN 0-521-00754-2. Revised reprint of the 1989 original. MR1932358.

R. Dykstra. Asymptotic normality for chi-bar-square distributions. *Canad. J. Statist.* **19** (3), 297–306 (1991). MR1144147.

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). MR3275854.

G. Englund. A remainder term estimate in a random-sum central limit theorem. *Teor. Veroyatnost. i Primenen.* **28** (1), 143–149 (1983). MR691474.

B.V. Gnedenko and V.Y. Korolev. *Random summation*. CRC Press, Boca Raton, FL (1996). ISBN 0-8493-2875-6. Limit theorems and applications. MR1387113.

L. Goldstein. Normal approximation for hierarchical structures. *Ann. Appl. Probab.* **14** (4), 1950–1969 (2004). MR2099658.

L. Goldstein. Bounds on the constant in the mean central limit theorem. *Ann. Probab.* **38** (4), 1672–1689 (2010). MR2663641.

L. Goldstein, I. Nourdin and G. Peccati. Gaussian phase transitions and conic intrinsic volumes: Steining the steiner formula. *ArXiv Mathematics e-prints* (2014). arXiv:1411.6265.

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). MR1484792.

L. Goldstein and G. Reinert. Distributional transformations, orthogonal polynomials, and Stein characterizations. *J. Theoret. Probab.* **18** (1), 237–260 (2005). MR2132278.

L. Goldstein and Y. Rinott. Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.* **33** (1), 1–17 (1996). MR1371949.

V. 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). MR2929524.

V.Y. 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). MR886279.

V.Y. Korolev. Accuracy of the normal approximation for distributions of sums of a random number of independent random variables. *Teor. Veroyatnost. i Primenen.* **33** (3), 577–581 (1988). MR968403.

R. Lachi and G. Peccati. New kolmogorov bounds for functionals of binomial point processes. *ArXiv Mathematics e-prints* (2015). arXiv:1505.04640.

P. Major. On the invariance principle for sums of independent identically distributed random variables. *J. Multivariate Anal.* **8** (4), 487–517 (1978). MR520959.

M. B. McCoy and J.A. Tropp. From Steiner formulas for cones to concentration of intrinsic volumes. *Discrete Comput. Geom.* **51** (4), 926–963 (2014). MR3216671.

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). MR2789507.

J. Pike and H. Ren. Stein's method and the Laplace distribution. *ALEA Lat. Am. J. Probab. Math. Stat.* **11** (1), 571–587 (2014). MR3283586.

H. Robbins. The asymptotic distribution of the sum of a random number of random variables. *Bull. Amer. Math. Soc.* **54**, 1151–1161 (1948). MR0027974.

A. Shapiro. Towards a unified theory of inequality constrained testing in multivariate analysis. *Internat. Statist. Rev.* **56** (1), 49–62 (1988). MR963140.

I. Shevtsova. On the absolute constants in the berry-esseen type inequalities for identically distributed summands. *ArXiv Mathematics e-prints* (2011). arXiv:1111.6554.

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. Univ. California Press, Berkeley, Calif. (1972). MR0402873.

F.W. Steutel. Some recent results in infinite divisibility. *Stochastic Processes Appl.* **1**, 125–143 (1973). MR0372948.

J.K. Sunklodas. L_1 bounds for asymptotic normality of random sums of independent random variables. *Lith. Math. J.* **53** (4), 438–447 (2013). MR3132110.

J.K. Sunklodas. On the normal approximation of a binomial random sum. *Lith. Math. J.* **54** (3), 356–365 (2014). MR3240976.

A.N. Tihomirov. Convergence rate in the central limit theorem for weakly dependent random variables. *Teor. Veroyatnost. i Primenen.* **25** (4), 800–818 (1980). MR595140.