Probab. Theory Relat. Fields 126, 528–570 (2003)
Digital Object Identifier (DOI) 10.1007/s00440-003-0271-5

Yosef Rinott · Vladimir Rotar

On Edgeworth expansions for dependency-neighborhoods chain structures and Stein's method

Received: 28 October 2000 / Revised version: 26 February 2003 /
Published online: 12 May 2003 –

Abstract. Let W be the sum of dependent random variables, and $h(x)$ be a function. This paper provides an Edgeworth expansion of an arbitrary "length" for $E\{h(W)\}$ in terms of certain characteristics of dependency, and of the smoothness of h and/or the distribution of W. The core of the class of dependency structures for which these characteristics are meaningful is the local dependency, but in fact, the class is essentially wider. The remainder is estimated in terms of Lyapunov's ratios. The proof is based on a Stein's method.

1. Introduction and results

1.1. Background and motivations

This paper concerns the Edgeworth expansion for $E\{h(W)\}$, where h is a function, $W = X_1 + \cdots + X_n$, and X_i's are random variables (r.v.'s).

There are numerous papers on the subject for various types of dependency; see, e.g., papers (in the chronological order) by Statulevichius [41], Hipp [19], Jensen [22], and Malinovskii [31], where Markov chains are considered; Götze and Hipp [13], and Lahiri [25], concerning mixing summands; Rhee [37] on m-dependent r.v.'s; Heinrich [14], [15], [16] on m-dependent r.v.'s and random fields; Mykland [33] on expansions for martingales; Jensen [23] on expansions for random fields. There are many further references in these articles. There has been also a great deal of interest in the expansions for *U*-statistics; see, e.g., Korolyuk and Borovskikh [24], Bickel, Götze and van Zwet [6], Loh [27], Maesono [28], Bloznelis and Götze

Y. Rinott: Department of Mathematics, University of California, San Diego and Department of Statistics, Hebrew University. e-mail: rinott@pluto.mscc.huji.ac.il

Supported in part by NSF grant DMS-98-03623

V. Rotar: Department of Mathematics, San Diego State University and the Central Economics and Mathematics Institute of Russian Academy of Sciences
e-mail: vrotar@euclid.ucsd.edu

Supported in part by the Russian Foundation of Basic Research, grant # 00-01-00194, and by NSF grant DMS-98-03623

Mathematics Subject Classification (2000): Primary 62E20; Secondary 60E05

Key words or phrases: Edgeworth expansion – Local dependency – Stein's method – Non-complete U-statistics

[7], Bening [3]; and references therein. The paper by Bentkus, Götze and van Zwet [4] on symmetric statistics does not deal with sums but contains relevant ideas.

The results of Götze and Hipp [13] and Lahiri [25] are probably the closest to the present paper. They concern, in particular, an expansion for $E\{h(W)\}$ when h is sufficiently smooth, and X_i's are weakly dependent. The weakest condition on smoothness of h was established in [25]: for the Edgeworth expansion involving the first d moments in the main term (that is, the expansion has "length" $d-2$), the function h should have $d-1$ continuous derivatives. For m-dependent summands Lahiri requires only the existence of the derivative of order $d-2$.

For non-smooth functions [25] gives the expansion under a conditional Cramer condition and strong mixing with an exponential rate.

In the present paper we consider a different and rather broad class of dependency structures. For the remainder in the expansions below we provide not only asymptotic rates but give explicit bounds for the remainder in terms of Lyapunov's ratios and some characteristics of the dependence. The method of this paper is based on the Stein approach. Expansions for independent summands by Stein's method were studied by Barbour [1], and we use below several facts and ideas from this paper.

A typical example of the dependency structure under consideration is that specified in terms of *dependency neighborhoods* by indicating for every term in the sum, a set of other terms on which it "essentially" depends. More precisely, for each summand X we introduce a chain of collections of summands $\mathcal{N}_1,\ldots,\mathcal{N}_d$ such that $X \in \mathcal{N}_1 \subseteq \ldots \subseteq \mathcal{N}_d$. (Writing $\mathcal{N}' \subseteq \mathcal{N}''$, we mean that all r.v.'s from the collection $\mathcal{N}'$ belong to the collection $\mathcal{N}''$.) We assume that X may depend in a strong way only on terms from $\mathcal{N}_1$, and weakly in some sense on terms not in $\mathcal{N}_1$; the collection $\mathcal{N}_1$ essentially depends only on terms from $\mathcal{N}_2$, and weakly on terms outside of $\mathcal{N}_2$; and so on. Note that in the case of independent summands we would set all $\mathcal{N}_s$ to contain only X. The notion of weak dependence mentioned above will be quantified by conditions close to mixing, with an arbitrary rate. In particular, we may assume the rate to be less restrictive than the exponential rate which appears in most of the literature.

Note that we have ordered the summands by X_i, $i = 1, \ldots, n$, only for convenience: in fact, the above dependency structure need not be associated with an ordering of the summands, which may be compared with more "classical" CLT's where the dependence is specified in terms of an ordering (Markov chains, martingales, mixing, etc.).

The core of the class of dependencies which may be described in this way is the so called local dependency when X does not depend at all on the terms not involved in $\mathcal{N}_1$, the collection $\mathcal{N}_1$ does not depend on the terms not in $\mathcal{N}_2$, and so on. The simplest example is m-dependence. In introducing such structures we follow Stein [43].

However, the class of dependencies we deal with is essentially wider. A certain flexibility is due in particular to the fact that the sequence $\mathcal{N}_1,\ldots,\mathcal{N}_d$ considered in the theorems below is arbitrary, and for each particular dependency structure one can specify the most apt sequence of $\mathcal{N}$'s for which the error of approximation would be small. For related decompositions see also Barbour, Karoński and Ruciński [2], Rinott and Rotar [39], and references therein.

Concerning the remainder, we show that, as in the case of independence, if the main term in the approximation involves d first moments of the summands, and the moments of order $(d+1)$ are also finite, then under some natural conditions, the remainder has, roughly speaking, the order $O(n^{-(d-1)/2})$. (On the desirability of such a setup see, e.g., a discussion in Bentkus, Götze and van Zwet [4]). To this end we assume that h has a bounded derivative of the order $d-1$.

Comparing it with the paper of Lahiri [25] concerning mixing, it is worth noting that the goal of the mentioned paper is to obtain an error of $o(n^{-(d-2)/2})$, which requires just $E\{|X|^d(ln(1+|X|))^p\} < \infty$ for some p. If one wants to obtain the remainder of a higher order, say, $O(n^{-(d-1)/2})$, then a higher order, namely $d+1$, of the finite moment is needed. We, however, have managed to maintain in this situation the same order $(d-1)$ of the derivative of h.

In general, to obtain, by the method of this paper, a remainder of the order $O(n^{-(d-2+\alpha)/2})$ for an $\alpha \in (0,1]$, one should assume $h^{(d-2)}$, where $h^{(l)}$ denotes the l-th derivative of h, to satisfy the Lipschits condition of the order α, and the $(d+\alpha)$-th moments of X's to exist.

In the case of local dependency the smoothness condition on derivatives may be weakened if one requires the distribution of the sum itself to be smooth enough (in the usual statement of the problem for sufficiently large n). Say, if we assume that the distribution mentioned has a density whose m-th derivative exists and is absolutely integrable, then we need just $h^{(d-m)}$ to be bounded.

The condition on the smoothness of the distribution of W is certainly not mild, though for large n it is not so restrictive as it might seem. Say, in the independency case, as is well known, the distribution mentioned may be non-smooth for small n, and sufficiently smooth when n is large, due to the convolution effect. In the case of dependency one can expect a similar effect at least when the dependency is not too strong.

In this section we formulate general theorems; some particular schemes are considered in Section 2. The most typical example for the results of this paper is mixing on graphs, that is when the parameter indexing the summands, which is usually thought of as a "time" or "space" parameter, has values which may be identified with vertices of a graph. If the graph is a usual integer valued lattice in $\mathbb{Z}^k$, with edges connecting only nearest vertices, we deal with the usual mixing scheme for random fields, and for $k=1$ – with a process on a line. If the graph is arbitrary, the scheme is more complicated. This is especially true when the graph is random, and its structure may depend on the values of summands. In this case the above dependency neighborhoods may be random too. Stein's method works well in the case of graph related dependencies because it allows to reduce the whole dependency structure to the dependence of each separate variable on the others, which simplifies consideration.

However it is worth emphasizing that the results of this paper do not completely cover all known results for mixing. The bounds for the remainder below have a right order typically in the case of local dependency (though not only); for example, in the case of m-dependence on graphs (though m may be random, may depend on X's, etc.) However, in the classical mixing framework with exponentially decreasing mixing coefficients, the bounds below would have an order of $O([\ln n]^p n^{-(d-1)/2})$

for some p (see Section 2.4.1), which is a bit worse than the results for classical mixing for random processes or fields (see, e.g., [13], [25], [23]).

One can say that the method enables us to investigate more general situations, in particular more general dependency neighborhoods, but possibly at a cost of $\ln n$ in the numerator. The dependency neighborhoods should be viewed as a part of the parameters determining the dependency, which is in fact quite general. In the case of dependency expressed with graph-related neighborhoods, our approach allows in principle to write bound for arbitrary graphs, in terms of certain parameters, which include the neighborhoods.

1.2. Method

We study the following representation, the possibility of which was first pointed out and justified in a certain case in Barbour [1]. Let f be an r times differentiable function, and W be a r.v. with finite first $r+1$ moments. Then

$$E\{Wf(W)\} = \sum_{m=0}^{r} \frac{\gamma_{m+1}}{m!} E\{f^{(m)}(W)\} + R, \tag{1.1}$$

where γ_m is the m-th cumulant of W, and R is a remainder which may be small under suitable conditions; see [1] and below for details. If $E\{W\} = 0,\ \ E\{W^2\} = 1$, it follows from (1.1) that

$$E\{Wf(W)\} - E\{f'(W)\} = \sum_{m=2}^{r} \frac{\gamma_{m+1}}{m!} E\{f^{(m)}(W)\} + R. \tag{1.2}$$

For a given function h, denote by $\mathcal{S}(h)$ the Stein function f solving the differential equation $f'(w) - wf(w) = h(w) - \mathbf{\Phi}(h)$ (see, e.g., Stein [42], [43]), that is,

$$\mathcal{S}(h)(x) = \frac{1}{\varphi(x)} \int_{-\infty}^{x} [h(t) - \mathbf{\Phi}(h)]\varphi(t)dt,$$

where φ is the standard normal density, and $\mathbf{\Phi}(h) = \int_{-\infty}^{\infty} h(t)\varphi(t)dt$. For $f = \mathcal{S}(h)$, from (1.2) it follows that

$$E\{h(W)\} - \mathbf{\Phi}(h) = -\sum_{m=2}^{r} \frac{\gamma_{m+1}}{m!} E\{f^{(m)}(W)\} - R. \tag{1.3}$$

The main term in (1.3) specifies the proximity to normality in terms of cumulants which are small under very mild requirements, provided W is close to normal. Thus the main conditions to be imposed should concern the remainder. The essential difficulty, however, lies in the fact that it is hardly possible to estimate R efficiently in terms of cumulants or some other characteristics of W itself as a non-decomposable r.v.

Considering independent summands, Barbour [1] wrote down the representations (1.1) for each summand separately, and then combined them adroitly, proceeding in the spirit of Stein's approach, and using independence in a crucial way.

The main idea of this paper is to derive (1.1) or a representation close to it taking into account the structure of the r.v. W from the very beginning; see the key Proposition 12 in Section 3. The way in which we suggest to do this is not short, but leads to a proper remainder for certain dependency structures.

1.3. Results

We fix an integer $d \geq 2$, and assume that $EX_i = 0$, and $E|X_i|^{d+1} < \infty$ for all $i = 1, ..., n$. We fix also a function h, set $h_{\varepsilon+}(x) := \sup_{|y|\leq\varepsilon} h(x+y)$, and $h_{\varepsilon-}(x) := \inf_{|y|\leq\varepsilon} h(x+y)$, and suppose that

$$\omega(\varepsilon; h) := \sup_z \int (h_{\varepsilon+}(x+z) - h_{\varepsilon-}(x+z)) \left(1 + |x|^{3d}\right) \varphi(x)dx \to 0, \quad \text{as } \varepsilon \to 0. \tag{1.4}$$

If h is the indicator of an interval, or if $h^{(1)}$ is uniformly bounded, which will be one of the possible conditions below, (1.4) is clearly satisfied.

Consider a summand $X = X_i$. Henceforth we often suppress the index i in notations when it does not cause a misunderstanding. For each X we introduce d decompositions

$$W = W_k + \tilde{W}_k, \qquad k = 1, ..., d. \tag{1.5}$$

One can view W_k as a partial sum of summands defined by $W_k = \sum_{X_j \in \mathcal{N}_k} X_j$, with $X \in \mathcal{N}_1 \subseteq \mathcal{N}_2 \ldots$ as above, but it is not necessary. Formally the decompositions introduced are arbitrary. In particular, W_k may be the sum of a random number of summands, as in the example in Sections 2.2 and 2.4 below.

In order to provide a language in which the dependency will be described, we introduce the notations $W_0 = 0$, and $U_s = W_s - W_{s-1}$ for $1 \leq s \leq d$. Then $W_s = U_1 + ... + U_s$, and $\tilde{W}_s = U_{s+1} + \tilde{W}_{s+1}$ for $s = 1, ..., d-1$. So, in the situation of the previous example, $U_s = \sum_{X_j \in \mathcal{N}_s \setminus \mathcal{N}_{s-1}} X_j$. In the independence case we set all $W_k = X$.

Let

$$\mu_l = E\{(|X| + |U_1| + ... + |U_d|)^l\}. \tag{1.6}$$

We now characterize the dependency structure. Let $\mathcal{F}_0 = \sigma\{X\}$, $\mathcal{F}_s = \sigma\{X, U_1, U_2, ..., U_s\}$, the σ-algebra generated by the variables $\{X, U_1, U_2, ..., U_s\}$, and $\tilde{\mathcal{F}}_s = \sigma\{U_{s+1}, ..., U_d, \tilde{W}_d\}$, $s = 0, 1, ..., d-1$; $\tilde{\mathcal{F}}_d = \sigma\{\tilde{W}_d\}$. Thus, between $\mathcal{F}_s$ and $\tilde{\mathcal{F}}_{s+1}$ there is a "gap" U_{s+1}. Recalling that $X = X_i$, we may sometimes write $\mathcal{F}_{si}$ for $\mathcal{F}_s$, and similarly for other quantities in which the index i is usually suppressed.

The term *local dependency* is used if $\mathcal{F}_{si}$ and $\tilde{\mathcal{F}}_{s+1,i}$ are independent for all $s = 1, ..., d-1$ and $i = 1, \ldots, n$.

Next, using standard notations (see, e.g., Bradley [8], [9]), for two σ-algebras $\mathcal{A}$ and $\mathcal{B}$ we set

$$\phi(\mathcal{A}, \mathcal{B}) := \sup\{|P(B|A) - P(B)|;\ A \in \mathcal{A},\ B \in \mathcal{B},\ P(A) > 0\}. \tag{1.7}$$

Below it will be important for us that $\phi(\mathcal{A}, \mathcal{B})$ admits also the representation

$$\phi(\mathcal{A}, \mathcal{B}) = \frac{1}{2}\left\{ \sup_{Y_1 \in \mathcal{L}_1(\mathcal{A}),\, Y_2 \in \mathcal{L}_\infty(\mathcal{B})} \frac{|E\{Y_1Y_2\} - E\{Y_1\}E\{Y_2\}|}{\|Y_1\|_1 \|Y_2\|_\infty} \right\}. \tag{1.8}$$

(See, e.g., Bradley [9]. For the fact that the r.-h.s. of (1.8) does not exceed the r.-h.s. of (1.7) see, Bradley [9], also Iosifescu and Theodorescu[21, Lemma 1.1.9]; one can come arbitrarily close to achieving equality taking indicators as r.v.'s under consideration).

We define the *characteristic of dependence*

$$T = \sup_{s,i} \phi(\mathcal{F}_{si}, \tilde{\mathcal{F}}_{s+1,i}). \tag{1.9}$$

In the local dependency case, $T = 0$. In general, T is a counterpart of a well known quantity in the mixing framework; see, e.g., Bradley [8], [9], Doukhan [11], Ibragimov [20], Peligrad [34].

Now we turn to smoothness conditions. In the general case $T \neq 0$, we will assume that at least

$$||h^{(d-2)}||_\infty < \infty. \tag{1.10}$$

In the local dependency case, that is, when $T = 0$, conditions on smoothness of h may be weakened if we assume the distributions of $\tilde{W}$'s to be sufficiently smooth. More precisely, denote by $q_s(\cdot, \omega)$ the conditional density of $\tilde{W}_s$ with respect to $\mathcal{F}_s$ (provided that it exists for a.e. ω in the probability space. We omit obvious issues of non-uniqueness and regularity.) If for some m, $1 \le m \le d-1$, the density $q_s(x, \omega)$ is m times differentiable in x, set

$$\psi_m = \max_{s=0,\dots,d} Ess \sup_{\omega} \left\{ \max_{k=0,\dots,m-1} \|q_s^{(k)}(\cdot, \omega)\|_\infty + \max_{k=0,\dots,m} ||q_s^{(k)}(\cdot, \omega)||_1 \right\}.$$

Set $\psi_0 \equiv 1$. (In this case we do not need the density q to exist.) We will see that, if $\psi_m < \infty$ for an $m \ge 1$, then the order of the highest finite derivative of h needed, may be reduced.

Recall that in all quantities above, the index i of the summand $X = X_i$ was suppressed. So in fact $W_k = W_{ki}$, $U_k = U_{ki}$, $\mathcal{F}_s = \mathcal{F}_{si}$, $\psi_m = \psi_{mi}$, $\mu_l = \mu_{li}$. We define:

$$\Psi_m = \max_{1 \le i \le n} \psi_{mi}, \quad \bar{\mu}_l = \sum_{i=1}^n \mu_{li}, \quad \eta_k = \eta_{ki} = \sum_{l=1}^k \mu_{li},$$
$$\bar{\eta}_k = \sum_{i=1}^n \eta_{ki} \quad (= \sum_{l=1}^k \bar{\mu}_l), \quad \bar{z}_k = \sum_{l=2}^k \bar{\mu}_l. \tag{1.11}$$

To clarify the order of the quantities in (1.11), assume for a moment that W has already been normalized in some way, and suppose, as an example, that after a normalization $|X_i| \le Y/\sqrt{n}$ (in distribution), where Y is a positive r.v. with $(d+1)$

finite moments. Also suppose that the U_i's are sums of a bounded number of X's. Then it is easy to show that $\bar{\mu}_{l+2} = O(n^{-l/2})$, and $\bar{z}_k = O(1)$, or in other words, it is a bounded quantity. As to $\bar{\eta}_k$, it involves the first absolute moment $\bar{\mu}_1$, so this quantity is "large", having in the above situation the order $O(\sqrt{n})$. In the remainder it will be multiplied by T which is supposed to be "small".

Below, by convention, we set the quantities Ψ_m, $||h^{(k)}||$, and so on, to equal infinity, not only when these quantities are infinite but when they do not exist. It will not cause a misunderstanding. Also, by convention, $\infty \cdot 0 = 0$.

In order to formulate first results, define as usual (see, e.g., Petrov [35]), the signed measure

$$Q_\nu(dx) = \sum\nolimits_{(\nu)} p_{\mathbf{k}\nu} L_{\nu+2s}(dx), \tag{1.12}$$

where $L_m(dx) = (-1)^m \varphi^{(m)}(x)dx$,

$$p_{\mathbf{k}\nu} = \prod_{m=1}^{\nu} \frac{1}{k_m!} \left(\frac{\beta_{m+2}}{(m+2)!} \right)^{k_m}, \tag{1.13}$$

$s = s(\mathbf{k}) = k_1 + ... + k_\nu$, and the summation in $\sum_{(\nu)}$ is over all vectors of nonnegative integers $\mathbf{k} = (k_1, ..., k_\nu)$ such that $k_1 + 2k_2 + ... + \nu k_\nu = \nu$.

For the coefficients β_l in (1.13), we consider below two types of quantities. Either traditionally

(i) $\beta_l = \gamma_l$ where γ_l denotes the l-th cumulant of the sum W;

or

(ii) β_l equals another characteristic, $\bar{\alpha}_l$, which is defined in detail later in (3.3), and which coincides with the lth cumulant of W in the case of local dependency, that is, when $T = 0$. In particular, $\bar{\alpha}_2 = \sum_{i=1}^n \alpha_{2i}$, where $\alpha_{2i} = E\{X_i U_{1i}\}$. It is easy to see that in the case of local dependence, $\bar{\alpha}_2 = \mathbf{Var}\{W\}$. The other α's are certain combinations of the expectations of higher order products of X_i, and $U_{1i}, ..., U_{si}$, while $\bar{\alpha}$'s are obtained by summing over i. The formal representation for $\bar{\alpha}_l$ coincides with that in the case of the local dependency (see (3.3) for detail). Note also that under some mild conditions the characteristics γ and $\bar{\alpha}$ are asymptotically close which is reflected below in Proposition 4.

When dealing with the characteristics $\bar{\alpha}$ in the first theorem below, we suppose $\bar{\alpha}_2 > 0$, which is a rather mild condition in the case of normal convergence. If $\bar{\alpha}_2 > 0$, we can "normalize" W by dividing it by $\bar{\alpha}_2^{1/2}$, which amounts to assuming $\bar{\alpha}_2 = 1$. When choosing a normalization, we in fact choose the normal distribution by which we approximate the distribution of the original sum. So, setting $\bar{\alpha}_2 = 1$, we approximate the original sum by the normal distribution with the variance not equal to the variance of the sum but to $\bar{\alpha}_2$. We will see that such a choice of the approximating normal and $\bar{\alpha}_l$, $l > 2$, for coefficients in (1.13), leads not only to explicit coefficients in the expansion but to weaker conditions on dependency and a better remainder. Also $\bar{\alpha}$'s may work when the traditional setup does not. The following very simple

Example shows that the choice of the characteristics mentioned is not just an artifact of the above approach but can reflect the essence of the matter. Let $Z_1, Z_2, ...$ be independent r.v.'s with the same distribution as a r.v. Z for which $EZ = 0,\ EZ^2 = 1$, and for a fixed n the vector

$$(Y_{1n}, ..., Y_{nn}) = \begin{cases} (Z_1, ..., Z_n) & \text{with probability } 1 - 1/n \\ (Z, ..., Z) & \text{with probability } 1/n \end{cases}$$

(we skip formalities of defining all r.v.'s above in one space). Let $S_n = Y_{1n} + ... + Y_{nn}$. Obviously, $ES_n = 0$, $Var\{S_n\} = 2n - 1$, and the normalization by $\sqrt{Var\{S_n\}}$ is not proper since $(S_n/\sqrt{n})$ converges in distribution to a standard normal r.v. On the other hand, if we set $X_i = X_{in} = Y_{in}/\sqrt{n}$, and choose in (1.5) $W_k = W_{ki} = X_i$ for all $k = 1, ..., d;\ i = 1, ..., n$, then $U_{1i} = X_i$, and as is easy to see, $\bar{\alpha}_2 = n^{-1} \sum_{i=1}^{n} E\{Y_{in}^2\} = 1$, that is, gives a correct normalization. It is easy to verify that in this case $T = 1/n$, and Proposition 1 below will imply that $Eh(W) - \Phi(h) = O(1/\sqrt{n})$, that is, the right rate in the CLT. The derivation of examples for asymptotic expansions is a bit cumbersome, so we restrict ourselves to the above example.

Returning to $\bar{\alpha}_l$ in the general case note that it admits an explicit (though complicated) representation involving a smaller number of mixed moments of the r.v.'s X_i in comparison with γ's (see (3.3) in Section 3 for detail); and they have a desirable order: for $l = 2, ..., d$

$$|\bar{\alpha}_l| \le C(l)\bar{\mu}_l, \tag{1.14}$$

where here and below the symbol $C(l)$ denotes a constant, perhaps different in different formulas, depending only on l.

If again $|X_i| \le Y/\sqrt{n}$, as will follow from (3.3), we have $\bar{\alpha}_2 = O(1),\ \bar{\alpha}_{l+2} = O(n^{-l/2})$, for $l \ge 0$, and all coefficients in Q_ν in (1.12) have the order $O(n^{-\nu/2})$.

We turn to the first result which uses quantities $\bar{\alpha}_l$, relegating their formal definition through mixed moments of X's to (3.3) of Section 3.

Set

$$\hat{h}_p = \max_{k=0,...,p} ||h^{(k)}||_\infty$$

Before presenting the main theorem consider, for ease of reading, a special case of it.

Proposition 1. *Let in (1.13)* $\beta_k = \bar{\alpha}_k$ *defined in (3.3). Assume that* $EX_i = 0$ *and* $\bar{\alpha}_2 = 1$. *Then there exists a constant* $C(d)$ *such that*

$$Eh(W) - \Phi(h) = \sum_{\nu=1}^{d-2} \int h(x) Q_\nu(dx) + R_d, \tag{1.15}$$

where

$$|R_d| \le C(d)\hat{h}_{d-1}(1 + \bar{z}_d)^{d-2} \left[\bar{\mu}_{d+1} + \bar{\eta}_d T\right].$$

(For a definition of $\bar{z}_d$ see (1.11).) Note that the term $(1 + \bar{z}_d)^{d-2}$ is typically bounded and close to $(1 + \bar{\mu}_2)^{d-2}$ due to normalization, and since $\bar{\mu}_k$ is "small" for $k \geq 3$. The above proposition is a corollary of

Theorem 2. *Let (1.4) hold, and in (1.13) let $\beta_k = \bar{\alpha}_k$ defined in (3.3). Assume that $EX_i = 0$ and $\bar{\alpha}_2 = 1$. Then there exists a constant $C(d)$ such that the expansion (1.15) holds with a remainder R_d satisfying for any $m = 0, \dots, d-1$*

$$|R_d| \leq C(d)(1 + \bar{z}_d)^{d-2} \left\{ \hat{h}_{d-1-m} \Psi_m \bar{\mu}_{d+1} + \hat{h}_{d-2} \bar{\eta}_d T \right\}. \tag{1.16}$$

Clearly, one can take the minimum of the r.-h.s. of (1.16) in m.

Remarks. (1) In the case of local dependency $T = 0$, and we have for any m

$$|R_d| \leq C(d)(1 + \bar{z}_d)^{d-2} \hat{h}_{d-1-m} \Psi_m \bar{\mu}_{d+1}.$$

In particular, if $\Psi_{d-1} < \infty$, the expansion can be considered for any bounded h satisfying (1.4), and in this case

$$|R_d| \leq C(d)(1 + \bar{z}_d)^{d-2} \|h\|_\infty \Psi_{d-1} \bar{\mu}_{d+1}.$$

(2) Let again $T = 0$, each W_{ki} be the sum of some number of summands X, and let M be the maximum number of summands in all W_{ki}'s. Suppose $\|X_i\|_{d+1} \leq A$ for all i. Then $\mu_l \leq (MA)^l$, and

$$|R_d| \leq C(d)[1 + n(MA)^2]^{d-2} n \hat{h}_{d-1} (MA)^{d+1}. \tag{1.17}$$

If, as before, $|X_i| \leq Y/\sqrt{n}$ (in distribution), where Y is a positive r.v. with $(d+1)$ finite moments, then $A = O(1/\sqrt{n})$ and the r.-h.s. of (1.17) has the order $O(n^{-(d-1)/2})$.

Now we turn to expansions in terms of cumulants of W, and with the normalization $Var\{W\} = 1$. The resulting theorem is rather a tribute to tradition: from a certain point of view Theorem 2 is better.

Using again standard notations (see, e.g., Bradley [8], [9]), for two σ-algebras $\mathcal{A}$ and $\mathcal{B}$ we set

$$\begin{aligned} \psi(\mathcal{A}, \mathcal{B}) := \sup \{ &|P(AB) - P(A)P(B)| / |P(A)P(B)| ; \\ &A \in \mathcal{A}, \;\; B \in \mathcal{B}, \; P(A), P(B) > 0 \}. \end{aligned} \tag{1.18}$$

Below we will use the fact that, provided that $\psi(\mathcal{A}, \mathcal{B})$ is finite,

$$\psi(\mathcal{A}, \mathcal{B}) = \sup_{Y_1 \in \mathcal{L}_1(\mathcal{A}),\, Y_2 \in \mathcal{L}_1(\mathcal{B})} \frac{|E\{Y_1 Y_2\} - E\{Y_1\}E\{Y_2\}|}{\|Y_1\|_1 \|Y_2\|_1}. \tag{1.19}$$

(See, e.g., [9]. For the fact that the r.-h.s. of (1.19) is less or equal than that of (1.18) see, e.g., Philipp [36, Lemma 3, page 157]; one can come arbitrarily close to achieving equality taking indicators as r.v.'s under consideration).

Now we define the *characteristic of dependence*

$$T' = \sup_{s,i} \psi(\mathcal{F}_{si},\ \tilde{\mathcal{F}}_{s+1,i}). \tag{1.20}$$

Again in the case of local dependency one may set $T' = 0$, and T' also has its counterpart in the mixing framework; see, e.g., again Bradley [8], [9], Doukhan [11], Peligrad [34].

As we mentioned before, $\bar{\alpha}_m = \gamma_m$ when $T' = 0$, so it makes sense to consider Theorem 3 only when $T' \neq 0$. But in this case, as will be seen from the formulation of Theorem 3, we require $\hat{h}_{d-1}$ to be bounded, so the introduction of characteristics Ψ's in this situation is superfluous.

Also, to avoid a cumbersome formulation, we introduce the condition

$$\bar{\eta}_d T' \leq 1/2^d, \quad \text{and} \quad T' \leq 1. \tag{1.21}$$

Since the bound on the remainder below contains the term $\bar{\eta}_d T'$, and since $\bar{\eta}_k$ typically is "large" [see also a remark after (1.11)], condition (1.21) does not "make the theorem below worse": if (1.21) is not true, the bound mentioned is not small in any case.

Theorem 3. *Let (1.4) and (1.21) hold, and in (1.13) set $\beta_k = \gamma_k$, the k-th cumulant of W, $k = 2, ..., d$. Assume that $EX_i = 0$ and $Var\{W\} = 1$. Then there exists a constant $C(d)$ such that the expansion (1.15) holds with*

$$|R_d| \leq C(d)\hat{h}_{d-1}(1 + \bar{z}_d)^{2(d-1)} \left[\bar{\mu}_{d+1} + \bar{\eta}_d T'\right]. \tag{1.22}$$

Remarks similar to those following Theorem 2 apply here too.

Finally we indicate the relation between the two expansions by showing that the quantities $\bar{\alpha}_k$ and the cumulants γ_k are close under suitable conditions:

Proposition 4. *Let $\bar{\mu}_k < \infty$, $T' \leq 1$ and $\bar{\eta}_1 T' \leq 1/2^k$. Then*

$$|\gamma_k - \bar{\alpha}_k| \leq C(k)\bar{\eta}_k(1 + \bar{z}_k)^{(k-1)/2} T'. \tag{1.23}$$

2. Some particular schemes

The aim of this section is not to consider schemes below in full generality but rather to illustrate possibilities and restrictions of the above approach. In particular, in all examples below we do not calculate explicitly coefficients in expansions: formally it may be done with use of (1.13) and (3.3), but leads to cumbersome formulas as in any complex enough scheme with dependent summands. For the same reason we consider in this section a smooth h. Our goal is to justify, for particular examples below, the validity of the expansion and to give bounds for remainders. While nowadays the computer may calculate coefficients well enough, the remainder could be still a hard problem for it.

We first consider a corollary from Theorem 2 concerning dependency neighborhoods. We use this corollary first for two relatively simple particular examples of random directed graphs, and next - for a general scheme based on undirected graphs.

2.1. *A corollary for dependency neighborhoods*

As usual, C and $C(\cdot)$ below indicate constants, absolute or depending only on arguments in parentheses, and possibly varying between equations and even in the same equation.

We assume now that the decompositions (1.5) are specified in terms of "real" neighborhoods as was mentioned in Section 1.1. Identifying the sets $\mathcal{N}$'s from Section 1.1 with the corresponding sets of indices, we consider for each summand X_i, a chain of sets $\{\mathcal{N}_{1i}, ..., \mathcal{N}_{di}\}$ where all $\mathcal{N}$'s are subsets of $\{1, ..., n\}$, and $i \in \mathcal{N}_{1i} \subseteq \ldots \subseteq \mathcal{N}_{di}$. The sets $\mathcal{N}$'s may be random. Let

$$W_{si} = \sum\nolimits_{j\in\mathcal{N}_{si}} X_j, \text{ and } U_{si} = \sum\nolimits_{j\in\mathcal{N}_{si}\setminus\mathcal{N}_{(s-1)i}} X_j, \text{ for } s = 1, ..., d.$$

Denoting by $|\mathcal{N}|$ the cardinality of $\mathcal{N}$, and keeping in mind that $i \in \mathcal{N}_{1i}$, we get from (1.6) that

$$\mu_{li} = E\{(|X_i| + |U_{1i}| + ... + |U_{di}|)^l\} \le 2^l E\left\{\left(\sum\nolimits_{j\in\mathcal{N}_{di}} |X_j|\right)^l\right\}$$
$$\le 2^l E\left\{|\mathcal{N}_{di}|^{l-1}\sum\nolimits_{j\in\mathcal{N}_{di}} |X_j|^l\right\} \le 2^l \hat{\mu}_l E\left\{|\mathcal{N}_{di}|^l\right\},$$

where

$$\hat{\mu}_l := \max_{i=1,\dots,n} Ess \sup_{\omega} \max_{j\in\mathcal{N}_{di}} E\{|X_j|^l \mid \mathcal{N}_{di}\}. \tag{2.1}$$

Hence, for $\bar{\mu}_l$ from (1.11)

$$\bar{\mu}_l \le C(l)\hat{\mu}_l H_l(n), \qquad \text{where} \qquad H_l(n) = \sum_{i=1}^n E\{|\mathcal{N}_{di}|^l\}.$$

Set $\hat{\eta}_k := \max_{1\le l\le k} \hat{\mu}_l$, $\hat{z}_k := \max_{2\le l\le k} \hat{\mu}_l$. Since $H_l(n)$ is increasing in l, for $\bar{\eta}_k$, $\bar{z}_k$ from (1.11), we have

$$\bar{z}_k \le C(k)\hat{z}_k H_k(n), \quad \bar{\eta}_k \le C(k)\hat{\eta}_k H_k(n).$$

Now it is easy to derive from Proposition 1

Proposition 5. *For the scheme above and under the conditions of Proposition 1, the expansion (1.15) holds with*

$$|R_d| \le C(d)\hat{h}_{d-1}[1 + (\hat{z}_d H_d(n))^{d-2}]H_{d+1}(n)[\hat{\mu}_{d+1} + \hat{\eta}_d T], \tag{2.2}$$

where T is defined as in Section 1.

Note that, if after a normalization, $|X_i| \le Y/\sqrt{n}$ (in distribution), where Y is a positive r.v. with $(d+1)$ finite moments, then $\hat{\mu}_l = O(n^{-l/2})$, $\hat{z}_k = O(n^{-1})$, $\hat{\eta}_k = O(n^{-1/2})$. So, if $H_{d+1}(n) = O(n)$, then if $\hat{h}_{d-1} < \infty$, we have $R_d = O\left(n^{-(d-1)/2} + \sqrt{n}T\right)$.

Next we consider two particular relatively simple examples.

2.2. *The nearest neighbor test*

Consider a sample of n i.i.d. points from an absolutely continuous distribution F in R^k, and the nearest neighbor graph whose n vertices are these points. This is a directed graph such that from each vertex there is a directed edge pointing to its nearest neighbor (with respect to Euclidean distance, say). Let each vertex be independently assigned one of two colors, say, green or blue, with probabilities ν and $1-\nu$, respectively. For a vertex i, let $N(i)$ denote its nearest neighbor, and $Y_i = 1$ if the vertices i and $N(i)$ are both assigned the green, and 0 otherwise, $i = 1, \dots, n$. Then $S = \sum_{i=1}^n Y_i$ counts the number of vertices i for which both i and $N(i)$ are green (with mutual nearest neighbors counted twice, once for each vertex).

If the color allocation is not done at random, and vertices having the same color tend to cluster together, we should expect large values of S. This phenomenon was used by Henze [18] (see also Henze [17], and other references there), who proposed a non-parametric statistic similar to S for testing equality of two distributions, and proved its asymptotic normality. Rinott and Rotar [39] provided rates considering several distributions (more than two colors), which led to a multivariate setup.

Here we consider asymptotic expansions. Clearly, $ES = n\nu^2$; for $Var\{S\}$ which depends on F, see [39]. Let $X_i = (Y_i - \nu^2) \big/ \sqrt{Var\{S\}}$, $W = \sum_{i=1}^n X_i$.

From calculations in [39], it follows in particular that $Var\{S\} \geq n\nu^2(1-\nu)$, and hence

$$|X_i| \leq 1 \Big/ \left[\sqrt{n}\left(\nu\sqrt{1-\nu}\right)\right]. \tag{2.3}$$

Using the fact that the sets $\mathcal{N}_{ji}$ may be random, we choose them depending on the graph G. First, define $\mathcal{S}_i$ to consist of i and all vertices which are connected with i by an edge. Certainly $\mathcal{S}_i$ depends on the realization of the graph mentioned. As was proved in [39], X_i does not depend on $\{X_l;\ l \notin \mathcal{S}_i\}$.

We point out the reason why we cannot nevertheless choose $\mathcal{S}_i$ as $\mathcal{N}_{1i}$, in a

Remark. Consider, in the general framework, a random set A of indices, and $\mathbf{X}_A = \{X_j;\ j \in A\}$. Assume that we have managed to specify a set of vertices $\tilde{A} \supseteq A$, such that any combination of a non-random number of r.v.'s X_k with $k \notin \tilde{A}$ does not depend on $\mathbf{X}_A$. Nevertheless it does not mean that $\mathbf{X}_A$ and $\mathbf{X}_{\tilde{A}^c} = \{X_k;\ k \notin \tilde{A}\}$ are independent since $|\tilde{A}|$ may depend on $\mathbf{X}_A$ (say, through $\sum_{j\in A} X_j$), and hence the whole collection $\mathbf{X}_{\tilde{A}^c}$ might depend on $\mathbf{X}_A$ just because the number of r.v.'s in $\mathbf{X}_{\tilde{A}^c}$, that is, $n - |\tilde{A}|$, may depend on $\mathbf{X}_A$. It may cause problems.

[39, p. 341] contains a small gap connected with the above issue, which can be easily fixed in the way we follow now. As is well-known, the degrees in the nearest neighbor graph in R^k are bounded by some constant $K(k)$ which depends on the dimension k. In particular $K(1) = 2$, $K(2) = 6$, $K(3) = 12$, and estimates are known for all k, see, e.g., Leech and Sloane [26]. If $|\mathcal{S}_i| = K(k)$, we set $\mathcal{N}_{1i} = \mathcal{S}_i$. Otherwise we add to $\mathcal{S}_i$ nearest $(K(k) - |\mathcal{S}_i|)$ points from the complement of $\mathcal{S}_i$, and define the new set as $\mathcal{N}_{1i}$. Next consider $\mathcal{S}_{2i} = \bigcup_{j\in\mathcal{N}_{1i}} S_j$. Obviously, $|\mathcal{S}_{2i}| \leq K^2(k)$. If $|\mathcal{S}_{2i}| = K^2(k)$, we set $\mathcal{N}_{2i} = \mathcal{S}_{2i}$, otherwise we add to $\mathcal{S}_{2i}$

any $(K^2(k) - |\mathcal{S}_{2i}|)$ points from outside of $\mathcal{S}_{2i}$, and so on. For $\mathcal{N}_{1i}, ..., \mathcal{N}_{di}$ so constructed,

$$|\mathcal{N}_{si}| = K^s(k) \tag{2.4}$$

It remains to apply (2.2). From (2.1), (2.3), and (2.4) it follows that

$$\hat{\mu}_l \leq 1 \Big/ \left[n^{l/2}(\nu\sqrt{1-\nu})^l\right], \quad \hat{z}_k \leq 1 \Big/ \left[n(\nu\sqrt{1-\nu})^k\right],$$
$$H_l(n) = nK^{ld}(k) = C(k, l, d)n \ .$$

Since by construction $T = 0$, the above bounds and (2.2) imply that for the scheme under consideration the expansion (1.15) holds with

$$|R_d| \leq C(k, d)\left[(\nu\sqrt{1-\nu})^{-d^2-1+d}\right]\hat{h}_{d-1}n^{-(d-1)/2}.$$

2.3. A random non-complete U-statistic

Let $\xi_1, \xi_2, ...$ be i.i.d. r.v.'s, and for a fixed n the vector $\mathbf{K}_n = \{K(1), ..., K(n)\}$, where K's are independent r.v.'s which are also independent of ξ's, and which take values from $\{1, ..., n\}$. Let $\psi(x, y)$ be a symmetric function, and $Y_i = \psi(\xi_i, \xi_{K(i)})$, $i = 1, ..., n$. For simplicity of calculations, we assume

$$P(K(i) = i) = 0. \tag{2.5}$$

Suppose $E\{Y_i\} = 0$, $E\{|Y_i|^{d+1}\} = E\{|\psi^{d+1}(\xi_1, \xi_2)|\} < \infty$. Set $S = S_n = \sum_{i=1}^n Y_i$.

It is convenient to view this as a directed graph with n vertices where each vertex is connected with one and only one vertex (a neighbor), and $\mathbf{K}_n$ specifies these connections; ξ's are assigned to vertices, and each Y is a function of the corresponding ξ, and its "neighbor". In Section 2.4 we consider a more general scheme, though for undirected graphs, where the number of neighbors is random and may be greater than two, ψ may depend on the sample (i, K_i), etc.

Set $N_j = \sum_{i=1}^n I(K(i) = j)$, the number of vertices connected with j, and $N_{jD} = \sum_{i \in D} I(K(i) = j)$ for $D \subseteq \{1, ..n\}$. Let $A_D = \{K(i) \in D \ \ \forall i \in D\}$, that is, D is isolated if A_D occurs.

We define a number $\lambda > 0$ such that for any D, and any $j \in D$

$$E\{N_{jD} \mid A_D\} \leq \lambda \ \text{ provided } \ P(A_D) > 0. \tag{2.6}$$

In other words, $E\{N_j\} \leq \lambda$ for each j, and the same is true for any N_{jD} given that D is isolated. Formally, since the scheme is finite, such a λ always exists. For the results below to be meaningful λ should not depend on n.

Let $p_{ij} = P(K(i) = j)$. Then, since K's are independent, $P(A_D) = \prod_{j \in D}(\sum_{k \in D} p_{jk})$, and (2.6) is implied by

$$\sum_{i \in D}\left(p_{ij} \Big/ \sum\nolimits_{k \in D} p_{ik}\right) \leq \lambda \quad \text{for all } D, \text{and } \ j \in D, \tag{2.7}$$

provided $(0/0) = 0$.

For example, (2.7) holds if $0 < \alpha \le (p_{ij}/p_{ik}) \le \beta$ for all $i,\ j \ne i,\ k \ne i$, and some α, β (say, each vertex "chooses a neighbor" with equal probabilities). On the other hand, the "classical" case $S = \psi(\xi_1, \xi_2) + \psi(\xi_2, \xi_3) + ... + \psi(\xi_{n-1}, \xi_n) + \psi(\xi_n, \xi_1)$ is covered by (2.7) too with $\lambda = 1$ since, as is easy to see, in this case the only D for which $P(A_D) > 0$, is $D = \{1, ..., n\}$.

Let $I_i = I(K(K(i)) = i),\ \sigma^2 = E\{\psi^2(\xi_1, \xi_2)\},\ c = E\{\psi(\xi_1, \xi_2)\psi(\xi_1, \xi_3)\}$. We assume $\sigma > 0,\ |c| < \sigma^2$. We need a simple

Lemma 6. *(1) The conditional variance*

$$E\{S^2 \,|\, \mathbf{K}_n\} = n(\sigma^2 + c) + (\sigma^2 - 2c)\sum_{i=1}^{n} I_i + c\sum_{i=1}^{n} N_i^2, \ \textit{and} \tag{2.8}$$

$$ES^2 = n(\sigma^2 + c) + (\sigma^2 - 2c)m_{1n} + cm_{2n}, \tag{2.9}$$

where $m_{1n} = \sum_{i=1}^n E\{I_i\} = \sum_{i=1}^n \sum_{j=1}^n p_{ij}p_{ji},\ m_{2n} = \sum_{i=1}^n E\{N_i^2\} = \sum_{i=1}^n (\lambda_i + \lambda_i^2) - \sum_{i=1}^n \sum_{j=1}^n p_{ij}^2,$ *with* $\lambda_i = E\{N_i\} = \sum_{j=1}^n p_{ji}$.

(2) Let $\Theta_n^2 := \frac{1}{n} E\{S_n^2 \,|\, \mathbf{K}_n\}$. *Then*

$$\Theta_n^2 \ge \sigma^2 > 0, \tag{2.10}$$

and under condition (2.7) for any natural l

$$E\{\Theta_n^{2l}\} \le C(l)\sigma^{2l}e^{2\lambda}. \tag{2.11}$$

(3) If the above scheme is given for an infinite sequence of natural n, and (2.7) holds uniformly in n, that is, for the same λ for all n, then for $\theta_n^2 = E\{\Theta_n^2\}$

$$\Theta_n^2 - \theta_n^2 \xrightarrow{P} 0, \quad \textit{as} \ \ n \to \infty. \tag{2.12}$$

Proofs of this and the next lemma are given in Section 5.1.

Set $X_i = Y_i \big/ \left[\theta_n\sqrt{n}\right],\ W = \sum_{i=1}^n X_i$, and denote by Z a standard normal r.v. independent of all other r.v.'s. In our opinion, it is more convenient and in a certain sense more efficient, to approximate the distribution of W not by that of Z itself but by the distribution of the r.v. $Z_n = [\Theta_n \big/ \theta_n]Z$, that is, by a weighted normal. This leads to a better accuracy of approximation: W is closer to Z_n than to Z. The computation of the coefficients in the main terms of the approximation to Z_n may appear cumbersome, but in fact even when approximating by Z an involved numerical calculation would be required. Certainly, in view of (2.10) and (2.12) $Z_n \xrightarrow{\mathcal{L}} Z$.

Let $Q_\nu(dx \,|\, \mathbf{K}_n)$ be measures (1.12) with the coefficients (1.13) where β_l equals the lth cumulant of $S/\left[\Theta_n\sqrt{n}\right]$ given $\mathbf{K}_n$. The structure of $Q_\nu(dx \,|\, \mathbf{K}_n)$ is the same as of $Q_\nu(dx)$: the dependence on $\mathbf{K}_n$ is reflected only in coefficients. Note also that, since below we will define the dependency neighborhoods in a way that T will vanish, the cumulants mentioned will coincide with the quantities $\bar{\alpha}_l$ defined in Section 1.

Proposition 7. *Let* $\gamma = \max_{l=2,\ldots,d+1}(E\{|\psi(\xi_1,\xi_2)|^l / \sigma^l\})$. *Then*

$$Eh(W) - Eh(Z_n) = \sum_{\nu=1}^{d-2} E\left\{\int h([\Theta_n/\theta_n]\,x)Q_\nu(dx \mid \mathbf{K}_n)\right\} + R_d, \qquad (2.13)$$

where

$$|R_d| \le C(d)e^{(d+1)\lambda}\hat{h}_{d-1}(1+\gamma)^{d-1}\frac{1}{n^{(d-1)/2}}. \qquad (2.14)$$

Proof. It is straightforward. Clearly

$$\begin{aligned} Eh(W)-Eh(Z_n) &= E\left\{E\left\{h\left([\Theta_n/\theta_n]\left[S/[\Theta_n\sqrt{n}]\right]\right)-h\left([\Theta_n/\theta_n]Z\right)\right| \mathbf{K}_n\right\}\right\} \\ &= E\left\{E\left\{\tilde{h}\left(S/[\Theta_n\sqrt{n}]\right)-\tilde{h}(Z)\right| \mathbf{K}_n\right\}\right\}, \end{aligned}$$

where $\tilde{h}(x) = \tilde{h}_n(x) = h\left([\Theta_n/\theta_n]x\right)$.

First, we fix $\mathbf{K}_n$ and apply Proposition 5 to estimate the remainder $R_d(\mathbf{K}_n)$ in the expansion

$$E\left\{\tilde{h}\left(S/[\Theta_n\sqrt{n}]\right)-\tilde{h}(Z)\right| \mathbf{K}_n\right\} = \sum_{\nu=1}^{d-2}\int \tilde{h}(x)Q_\nu(dx \mid \mathbf{K}_n) + R_d(\mathbf{K}_n). \qquad (2.15)$$

To this end we should define neighborhoods $\mathcal{N}_{1i}, \ldots, \mathcal{N}_{di}$. We take into account (2.5), so all assertions below about $\mathcal{N}_{is}$ may be true only a.s.

Let $\mathcal{N}_{1i} = \{i\} \cup \{K(i)\} \cup \{j;\ K(j) = i \text{ or } K(i)\}$. Obviously, Y_i does not depend on $\{Y_j;\ j \notin \mathcal{N}_{1i}\}$. Note also that all edges emanating from points in $\mathcal{N}_{1i}$ end at points from $\mathcal{N}_{1i}$ with one possible exception: the edge from $K(i)$. In other words, $K(K(i))$ may be out of $\mathcal{N}_{1i}$. In view of this, we set $\mathcal{N}_{2i} = \mathcal{N}_{1i} \cup \{K(K(i))\} \cup \{j;\ K(j) \in \mathcal{N}_{1i}\} \cup \{j;\ K(j) = K(K(i))\}$. Again, all edges emanating from points in $\mathcal{N}_{2i}$ end at points from $\mathcal{N}_{2i}$ save perhaps the edge from $K(K(i))$, and so on. So, setting $K_s(i) = K(K(\ldots K(i)))$ where the operation K appears s times, define $\mathcal{N}_{(s+1)i} = \mathcal{N}_{si} \cup \{K(K_s(i))\} \cup \{j;\ K(j) \in \mathcal{N}_{si} \setminus \mathcal{N}_{(s-1)i}\} \cup \{j;\ K(j) = K(K_s(i))\}$. It is straightforward to verify that in this case, given $\mathbf{K}_n$, we deal with local dependency, and hence $T = 0$.

Lemma 8. *Under condition (2.7) for any* l

$$E\,|\mathcal{N}_{di}|^l \le C(l,d)e^{2d\lambda}. \qquad (2.16)$$

A combinatorial proof is given in Section 5.1. We turn to an expansion using notations from Section 2.1. To estimate the remainder we use (2.2). It is easy to see that in our case, given $\mathbf{K}_n$, and in view of (2.10) $\hat{\mu}_l = E\{|\psi(\xi_1,\xi_2)|^l\} / (\Theta_n\sqrt{n})^l \le E\{|\psi(\xi_1,\xi_2)|^l\} / n^{l/2}\sigma^l$. Then $\hat{\mu}_l \le \gamma/n^{l/2}$, $\quad \hat{z}_k \le \gamma/n$.

Thus, in view of (2.2), and since $T = 0$ given $\mathbf{K}_n$,

$$|R_d(\mathbf{K}_n)| \le C(d) \left(\max_{k=0,\dots,d-1} \left(\Theta_n / \theta_n\right)^k \|h^{(k)}\|_\infty \right) \frac{(1+\gamma)^{d-1}}{n^{(d+1)/2}} \left[\frac{1}{n} H_d(n) \right]^{d-2} \cdot H_{d+1}(n)$$

$$\le C(d) \hat{h}_{d-1} (1+\gamma)^{d-1} \frac{1}{n^{3(d-1)/2}} [H_{d+1}(n)]^{d-1} \sum_{k=1}^{d-1} \left(\Theta_n / \sigma\right)^k,$$

where in this case $H_l(n) = \sum_{i=1}^n |\mathcal{N}_{di}|^l$ for given $\mathbf{K}_n$.

Next we compute $\mathrm{E}|R_d(\mathbf{K}_n)|$. By (2.16) and (2.11)

$$E\left\{ [H_{d+1}(n)]^{d-1} \sum_{k=1}^{d-1} \left(\Theta_n / \sigma\right)^k \right\} \le \sum_{k=1}^{d-1} \sigma^{-k} n^{d-2} \sum_{i=1}^{n} E\left\{ |\mathcal{N}_{di}|^{d^2-1} \Theta_n^k \right\}$$

$$\le \sum_{k=1}^{d-1} \sigma^{-k} n^{d-2} \sum_{i=1}^{n} \left(E\left\{ |\mathcal{N}_{di}|^{2(d^2-1)} \right\} E\left\{ \Theta_n^{2k} \right\} \right)^{1/2} \le C(d) n^{d-1} e^{(d+1)\lambda}.$$

The above bounds lead to (2.13) and (2.14). □

2.4. A graph related scheme with dependency neighborhoods

In this section we write some corollaries for the case when the graph specifying the dependency structure is non-directed, and the dependency neighborhoods are "geographical" neighborhoods with respect to the graph.

2.4.1. Mixing on non-directed graphs

Consider an arbitrary n-vertex simple graph. Let $\partial(i, j)$ denote the distance between vertices; $\partial(A, B)$ – the distance between two sets, A and B, of vertices; A^c – the complement of A, and $|A|$ – its cardinality; $O(i; r)$ – the r-neighborhood of the vertex i w.r.t. the graph.

We assign to each vertex i a random variable X_i, and set $\mathcal{F}(A) = \sigma(X_i;\ i \in A)$. The dependency structure is characterized by

$$T(r) = \max_i \sup_x \phi(\mathcal{F}(O(i; x)),\ \mathcal{F}(O^c(i; x + r))), \tag{2.17}$$

where ϕ is defined in (1.7). As before we assume $EX_i = 0$.

If the graph is, say, a usual integer valued lattice in $\mathbb{R}^k$ with edges connecting only nearest vertices, we deal with the usual mixing scheme for random fields; if the graph is arbitrary the scheme is more complicated.

As dependency neighborhoods $\mathcal{N}_{ji}$, we take just the sequence $O(i; r)$, $O(i; 2r)$, ..., $O(i; rd)$ where r is a free parameter. We again apply Proposition 5 observing that in our case

$$\hat{\mu}_l = \max_{i=1,\dots,n} E\{|X_i|^l\}$$

(see Section 2.1), and

$$H_l(n) = \sum_{i=1}^{n} |O(i; rd)|^l \le \Gamma^l(rd)n,$$

where $\Gamma(r) = \max_i |O(i; r)|$, a characteristic of the graph.

We fix r, and assume that the r.v. $W = \sum_1^n X_i$ has been already normalized in such a way that the characteristic $\bar{\alpha}_2 = 1$.

Setting $\hat{\eta}_k := \max_{1 \le l \le k} \hat{\mu}_l$, $\hat{z}_k := \max_{2 \le l \le k} \hat{\mu}_l$, it is easy to derive from (2.2)

Proposition 9. *In the framework of this section the expansion (1.15) holds with*

$$|R_d| \le C(d)\hat{h}_{d-1}[1 + (\hat{z}_d n)^{d-2}](\Gamma(rd))^{d^2-d+1} \cdot n \cdot [\hat{\mu}_{d+1} + \hat{\eta}_d T(r)]. \quad (2.18)$$

Corollary 10. *Assume that for some γ and $l \le d + 1$*

$$\hat{\mu}_l \le C(d)\gamma n^{-l/2}, \ \hat{z}_l \le C(d)\gamma n^{-1}, \ \hat{\eta}_l \le C(d)\gamma n^{-1/2} \quad (2.19)$$

(which reflects a "natural" order). Then (2.18) implies that for any $r > 0$

$$|R_d| \le C(d)\hat{h}_{d-1}(1 + \gamma)^{d-1}(\Gamma(rd))^{d^2-d+1}[n^{-(d-1)/2} + n^{1/2}T(r)]. \quad (2.20)$$

If $T(r) = 0$ for $r >$ some m (a sort of m-dependency on graphs), the bound above has a "right" order, and $R_d = O(n^{-(d-1)/2})$ provided $\Gamma(md)$ is bounded uniformly in n. (Say, for a complete graph it is not true since in this case $\Gamma(r) \equiv n - 1$.) If $\Gamma(r)$ is not bounded, one can choose an optimal $r = r(n)$ which depends on the degree of growth of $\Gamma(r)$. In particular, if $\Gamma(r) \le Cr^t$ for some t (as for "usual" random fields), while $T(r)$ is decreasing exponentially, as is easy to see, $R_d = O\left([\ln^p n]n^{-(d-1)/2}\right)$for some $p = p(d, t)$ which can be easily computed. As has been already told, the method allows to consider more general situations but possibly at a cost of $\ln^p n$.

2.4.2. A random incomplete U-statistic on a undirected graph

Next we consider a particular example of the above scheme. The model below admits an economic application, see Majumdar and Rotar [29], [30] and below for details. Consider again a simple n-vertex graph, and assign to each vertex i a random variable ξ_i. The dependency below remains non-trivial if ξ's are independent, but we do not suppose that and introduce the characteristic $T_1(r)$ which is the counterpart of characteristic (2.17) for ξ's. We define for each i a point-to-set map $\mathbf{A}_i(x, y)$, where x, y are numbers, and $\mathbf{A}_i$ takes values in the set of all subsets of $\{1, ..., n\}$. Suppose that $i \in \mathbf{A}_i(x, y)$ for all x, y, and set $A_i = \mathbf{A}_i(\xi_i, \eta_i)$, where $\eta_1, ..., \eta_n$ are given r.v.'s independent between themselves and of all other r.v.'s under consideration. (As a matter of fact, η's may be random elements of an arbitrary nature, but it would not make the scheme more general in essence.)

For example, for an integer-valued function $g(x)$ and a fixed number k, one can take the $g(\xi_i)$-neighborhood (with respect to the graph) of the vertex i , and choose at random k vertices in the neighborhood chosen, provided that k is not greater than the cardinality of the neighborhood.

Let $\Xi_i = \{\xi_j;\ j \in A_i\} = \{\xi_j;\ j \in \mathbf{A}_i(\xi_i, \eta_i)\}$. For brevity we call ξ's from Ξ_i "partners" of ξ_i. Let $\{G_i(x_1, ..., x_{|A|};\ A)\}, i = 1, ..., n$, be a given collection of functions such that for a fixed $A \subseteq \{1, ..., n\}$ the function $G_i(\cdot\,;\ A) : \mathbf{R}^{|A|} \to \mathbf{R}$, and symmetric. Set $X_i = G_i(\Xi_i;\ A_i)$; in view of symmetry of G this definition is correct. Let $W = \sum_1^n X_i$.

One of essential differences of this scheme from that of Section 2.3 is that the choice of a partner *depends now on the value of the r.v.* ξ.

In the economic setup mentioned (see, [29], [30] for detail) vertices are identified with "economic agents"; the graph reflects connections between agents, and the r.v. ξ_i characterizes the "state" of the i-th agent. Each agent chooses "partners" from a neighborhood with respect to the graph, and the neighborhood mentioned depends on the state of the agent. Say, more an agent is "powerful", larger the possibility of choice of possible partners. X's characterize production.

One may generalize the above scheme making the set A_i depending not only on ξ_i but other $\acute{\xi}$'s; say a vertex-agent may refuse to be a partner, or an agent cannot be a partner of too may agents simultaneously, etc. Here we restrict ourselves to the framework defined.

Note also that formally W above may be reduced to a non-random non-complete U-statistic of (weakly) dependent r.v.'s, though in this case the order of the statistic is not fixed (formally it may be any, up to n); all kernels, say, $U(x_{j_1}, x_{j_2}, ...) = U_{\mathbf{j}}(x_{j_1}, x_{j_2}, ...)$, that is, depend on the sample $\mathbf{j} = (j_1, j_2, ...)$, and this dependence cannot be reflected only through weights as in weighted U-statistics; the kernels are not assumed to be non-degenerated (so the Hoefding representation might not work); and the structure of dependence for ξ's is specified by a rather general graph. We would not exclude nevertheless that a combination of a technique for weighted U-statistics and some mixing technique could, in principle, work here, but it may be hard. On the other hand, it is easy to see that Proposition 9 immediately implies at least the following.

Proposition 11. *Suppose that the dependency neighborhoods are chosen in the same way as in Section 2.4.1 with a fixed r, and assume that the functions G have been already normalized in a way that* $EX_i = 0$, *and* $\bar{\alpha}_2 = 1$. *Set* $M_i = \max\{\partial(j, i);\ j \in \mathbf{A}_i(\xi_i, \eta_i)\}$, *and suppose that for an* $r_0 > 0$

$$P(\max_{i=1,\ldots,n} M_i \le r_0) = 1. \tag{2.21}$$

Then (2.18) holds with $T(r) = T_1(r - 2r_0)$ *if* $r > 2r_0$, *and* $T(r) = 1$ *otherwise.*

Again, under condition (2.19) we have the explicit bound (2.20).

Next we discuss the case when r_0 satisfying (2.21) either does not exist, or too large. Anyhow, we should first define a dependency neighborhood for each X_i, or in other terms, the r.v.'s W_{1i} and $\tilde{W}_{1i}$ from (1.5). The first that comes to mind is to define $\tilde{W}_{1i}$ as the sum of all X_j's such that the vertices they correspond to, do not have common "partners" with i. However in general case, unlike in Section 2.3, it could lead to a problem: which vertices are partners of i depends now on the value of ξ_i, and hence on the value of X_i, while the fact that a vertex j does not have a particular vertex as a partner influences the value of ξ_j, and hence of X_j. Attempts to fix it lead to cumbersome constructions.

Another obstacle which one should be aware of, is the same as mentioned in Remark in Section 2.2: the randomness of the cardinality of a neighborhood may cause problems too.

For these reasons, one can choose a simpler way, namely a truncation, which could make bounds for the remainder somewhat less accurate, but allows to avoid problems mentioned. More precisely we can consider r.v.'s

$$\bar{X}_i = \begin{cases} X_i & \text{if } M_i \le r_0, \\ Z_i & \text{if } M_i > r_0, \end{cases}$$

where r_0 is now a free parameter, and the r.v.'s Z_i are at our choice. The simplest way is to set $Z_i \equiv 0$, however if we want to maintain the smoothness of X's (if it is the case), we can take as Z's independent normal variables with appropriately chosen variances. Let $\bar{W} = \sum_1^n \bar{X}_i$.

It is easy to see that for any h

$$|E\{h(W)\} - E\{h(\bar{W})\}| \le 2\hat{h}_0 \sum_{i=1}^{n} P(M_i > r_0). \tag{2.22}$$

For $\bar{W}$ one can apply Proposition 11 directly after a proper rescaling which makes $E\bar{W} = 0,\ \bar{\alpha}_2 = 1$. Note that the remainder in this case will be practically the same as for original X's since the moments $E|\bar{X}_i|^l \le E|X_i|^l + E|Z_i|^l$. Because of lack of room, we omit concrete calculations. In particular again, if probabilities $P(M_i > r)$ vanish exponentially, the same is true for $T_1(r)$, while $\Gamma(r)$ is a power function, then, as is easy to verify, $R_d = O\left([\ln^p n]n^{-(d-1)/2}\right)$ for some p.

3. Proof of Theorem 2

We now define most of our notations, and one should refer to this part for notations used later in the paper.

For the first few preliminary results it is convenient to use $r = d - 1$. Thus, we consider $(X, U_1, ..., U_r, U_{r+1})$, where $X = X_i$ is a particular summand in $W = X_1 + \cdots + X_n$. We now fix the index i and suppress it in the notation of all quantities defined below. For example, the terms $U_1, ..., U_r, U_{r+1}$ depend on i.

We have $W_k = U_1 + \cdots + U_k,\ W = W_k + \tilde{W}_k$. Set $\mathbf{V}_s = (W_s, U_{s+1}, ..., U_r, U_{r+1})$, $\mathbf{U} = \mathbf{V}_0 = (U_1, ..., U_r, U_{r+1})$.

For $s + l \le r + 1,\ p \le l$ we define

$$\chi_l(s, p) = \sum_{|\mathbf{m}|=l}{}^{(p)} \frac{1}{\mathbf{m}!} E\{\mathbf{V}_s^{\mathbf{m}}\}, \quad \text{and} \quad \vartheta_{l+1}(p) = \sum_{|\mathbf{m}|=l}{}^{(p)} \frac{1}{\mathbf{m}!} E\{X\mathbf{U}^{\mathbf{m}}\},$$

where $\sum_{|\mathbf{m}|=l}{}^{(p)}$ is the sum over all $\mathbf{m} = (m_1, \ldots, m_{r+1})$, such that $m_i = 1, ..., r+1$ for all $i \le p$, and $m_i = 0$ for $p < i \le r+1$, and $|\mathbf{m}| = m_1 + \cdots + m_{r+1} = l$; also,

$\mathbf{U^m} = \prod_{i=1}^{r+1} U_i^{m_i}$, and $\mathbf{V}_s^{\mathbf{m}} = W_s^{m_1} U_{s+1}^{m_2} \cdots U_{r+1}^{m_{r-s+2}}$, and $\mathbf{m}! = m_1! \cdot \ldots \cdot m_{r+1}!$. It is easy to see that

$$\begin{aligned} |\chi_l(s,p)| &\le \sum_{|\mathbf{m}|=l}^{(p)} \frac{1}{\mathbf{m}!} E\{|\mathbf{V}_s^m|\} \le \frac{1}{l!} E\{(|W_s| + |U_{s+1}| + \cdots + |U_{s+l-1}|)^l\} \\ &\le \frac{1}{l!} E\{(|U_1| + \cdots + |U_{s+l-1}|)^l\} \le \frac{1}{l!} E\{(|U_1| + \cdots + |U_r|)^l\} \quad (3.1) \\ &\le E\{(|X| + |U_1| + \cdots + |U_{r+1}|)^l\} = \mu_l. \end{aligned}$$

Likewise

$$|\vartheta_{l+1}(s)| \le \mu_{l+1}. \tag{3.2}$$

Set $\Upsilon_{00} = 1$, $\Upsilon_{l0} = 0$ for $l \ge 1$, and for $m \ge 1$

$$\begin{aligned} \Upsilon_{lm} = \Upsilon_{lm}(s) = \sum_{l_1+\cdots+l_m=l} \sum_{p_1=1}^{l_1} \ldots \sum_{p_m=1}^{l_m} \chi_{l_1}(s,p_1)\chi_{l_2}(s+p_1,p_2) \cdots \\ \times \chi_{l_m}(s+p_1+\cdots+p_{m-1},p_m). \end{aligned}$$

For $l \le r$ set

$$c(s,l) = \sum_{m=0}^{[l/2]} (-1)^m \Upsilon_{lm}(s), \quad \tilde{c}(s,l) = \sum_{m=0}^{[l/2]} |\Upsilon_{lm}(s)|,$$
$$(\, c(s,0) = 1, \; c(s,1) = 0, \; s \ge 0\,).$$

Recalling that the above quantities depend on the suppressed index i, and the notation in (1.11), define $\alpha_{m+1} (= \alpha_{m+1,i})$ and $\bar{\alpha}_{m+1}$ by

$$\frac{\alpha_{m+1}}{m!} = \sum_{t+l=m, l\ge 1} \sum_{s=1}^{l} \vartheta_{l+1}(s) c(s+1,t), \quad \bar{\alpha}_{m+1} = \sum_{i=1}^{n} \alpha_{m+1,i}, \quad m = 1, \ldots, r. \tag{3.3}$$

The following facts are seen easily:

$$\begin{aligned} &\mu_l \mu_k \le \mu_{l+k}, \quad \text{and hence} \quad \Upsilon_{lm}, \; c(s,l), \; \tilde{c}(s,l) \le C(l)\mu_l \; \text{ for all } \; l \le r, \\ &\alpha_{m+1} \le C(m)\mu_{m+1} \end{aligned} \tag{3.4}$$

for constants $C(l)$ depending only on l. Henceforth $C(l)$ will indicate any such constants, possibly varying between equations and even in the same equation.

Next let $\mathcal{F}_s (= \mathcal{F}_{si})$ be the σ-algebra generated by the r.v.'s $X = X_i$, $U_{1i}, \ldots, U_{si}$, and for $l \le r$ and a sufficiently smooth function f, define

$$\hat{f}_l = max_{k \le l} ||f^{(k)}||_\infty,$$
$$\Gamma_l(f) = \max_{1 \le i \le n} \max_{s \le r+1} \sup_x Ess \sup_\omega |E\{ f^{(l)}(\tilde{W}_{si} + x) \Big| \mathcal{F}_{si}\}(\omega)|.$$

For now we assume that all derivatives of f which appear here and below exist. We return to this issue later, in the proof of Theorem 2.

The next Proposition is a further development of (1.1).

Proposition 12. *(THE MAIN LEMMA) For any sufficiently smooth f,*

$$E\{Wf(W)\} = \sum_{m=1}^{r} \frac{\bar{\alpha}_{m+1}}{m!} E\{f^{(m)}(W)\} + B_r(f),$$

with

$$|B_r(f)| \le C(r)[\bar{\mu}_{r+2}\Gamma_{r+1}(f) + \bar{\eta}_{r+1}\hat{f}_r T].$$

For the proof we need some lemmas.

Lemma 13. *For any sufficiently smooth f, and $k \le r$,*

$$\begin{aligned} f(W) &= f(\tilde{W}_1) + \sum_{p=1}^{k} \sum_{|\mathbf{m}| \le k}{}^{(p)} \frac{1}{\mathbf{m}!} \mathbf{U}^{\mathbf{m}} f^{(|\mathbf{m}|)}(\tilde{W}_{p+1}) + R_k \\ &= f(\tilde{W}_1) + \sum_{l=1}^{k} \sum_{p=1}^{l} \sum_{|\mathbf{m}|=l}{}^{(p)} \frac{1}{\mathbf{m}!} \mathbf{U}^{\mathbf{m}} f^{(l)}(\tilde{W}_{p+1}) + R_k, \end{aligned} \tag{3.5}$$

where

$$R_k = \sum_{p=1}^{k+1} \sum_{|\mathbf{m}|=k+1}{}^{(p)} \frac{m_p}{\mathbf{m}!} \mathbf{U}^{\mathbf{m}} \int_0^1 (1-t)^{(m_p-1)} f^{(k+1)}(\tilde{W}_p + tU_p)dt.$$

Proof. This follows by repeated Taylor expansions. First expand $f(W)$ in a Taylor series of $k+1$ terms about $\tilde{W}_1$ and a remainder containing $f^{(k+1)}$. Next expand $f'(\tilde{W}_1)$ into a k term Taylor series about $\tilde{W}_2$ and a remainder, and $f''(\tilde{W}_1)$ into a $k-1$ term Taylor series about $\tilde{W}_2$ and a remainder, etc. We skip a formal proof. □

Lemma 14. *For any sufficiently smooth f, and $k \le r$,*

$$\begin{aligned} E\{f(\tilde{W}_s)\} &= \sum_{l=0}^{k} c(s,l) E\{f^{(l)}(W)\} + M_{sk}(f) \\ &= E\{f(W)\} + \sum_{l=2}^{k} c(s,l) E\{f^{(l)}(W)\} + M_{sk}(f), \end{aligned} \tag{3.6}$$

where

$$|M_{sk}(f)| \le C(k)[\mu_{k+1}\Gamma_{k+1}(f) + \eta_k \hat{f}_k T].$$

The proof which uses (3.5) with $\tilde{W}_s$ and $\mathbf{V}_s$ replacing $\tilde{W}_1$, and $\mathbf{U}$, is given in Section 5.

Proof of Proposition 12. Since $E\{X\} = 0$, it follows from (3.5), (1.9) and (1.8) that

$$E\{Xf(W)\} = \sum_{l=1}^{r}\sum_{s=1}^{l}\sum_{|\mathbf{m}|=l}{}^{(s)}\frac{1}{\mathbf{m}!}E\{X\mathbf{U}^{\mathbf{m}}\}E\{f^{(l)}(\tilde{W}_{s+1})\} + E\{XR_r\} + 2T_r, \tag{3.7}$$

where

$$|T_r| \le \hat{f}_r\left[\sum_{l=1}^{r+1}\mu_l\right]T = \hat{f}_r\eta_{r+1}T, \quad \text{and} \quad |E\{XR_r\}| \le \mu_{r+2}\Gamma_{r+1}(f).$$

Computing the main term in (3.7), that is, ignoring for a while all terms in which T or R appear, and applying Lemma 14 again ignoring the remainder term M_{sk} (the notation $\approx$ is used to indicate this approximation), we have for $X_i = X$

$$\begin{aligned}
E\{Xf(W)\} &\approx \sum_{l=1}^{r}\sum_{s=1}^{l}\sum_{|\mathbf{m}|=l}{}^{(s)}\frac{1}{\mathbf{m}!}E\{X\mathbf{U}^{\mathbf{m}}\}E\{f^{(l)}(\tilde{W}_{s+1})\}\\
&= \sum_{l=1}^{r}\sum_{s=1}^{l}\sum_{|\mathbf{m}|=l}{}^{(s)}\frac{1}{\mathbf{m}!}E\{X\mathbf{U}^{\mathbf{m}}\}\sum_{t=0}^{r-l}c(s+1,t)E\{f^{(t+l)}(W)\}\\
&= \sum_{m=1}^{r}\left(\sum_{t+l=m,l\ge 1}\sum_{s=1}^{l}\vartheta_{l+1}(s)c(s+1,t)\right)E\{f^{(m)}(W)\}.
\end{aligned}$$

Summing over i we obtain the main term in the desired result.

Denote the remainder in the latter approximation by B_r (which still depends on i). By Lemma 14 applied to the function $f^{(l)}$, the term $M_{s+1,r-l}(f^{(l)})$ appearing below satisfies $|M_{s+1,r-l}(f^{(l)})| \le C(r)[\mu_{r-l+1}\Gamma_{r+1}(f) + \eta_{r-l}\hat{f}_rT]$. Therefore, we have (again allowing $C(r)$ to vary between equations),

$$\begin{aligned}
|B_r| &:= |E\{Xf(W)\} - \sum_{m=1}^{r}\frac{\alpha_{m+1}}{m!}E\{f^{(m)}(W)\}|\\
&\le \mu_{r+2}\Gamma_{r+1}(f) + \hat{f}_r\eta_{r+1}T + \sum_{l=1}^{r}\sum_{s=1}^{l}|\vartheta_{l+1}(s)|\cdot|M_{s+1,r-l}(f^{(l)})|\\
&\le \mu_{r+2}\Gamma_{r+1}(f) + \hat{f}_r\eta_{r+1}T + \sum_{l=1}^{r}\sum_{s=1}^{l}|\vartheta_{l+1}(s)|C(r)\\
&\quad \times[\mu_{r-l+1}\Gamma_{r+1}(f) + \eta_{r-l}\hat{f}_rT]\\
&\le C(r)[\mu_{r+2}\Gamma_{r+1}(f) + \hat{f}_r\eta_{r+1}T].
\end{aligned}$$

This holds for $X = X_i$, and by summing over i, Proposition 12 follows. □

Lemma 15. *For* $k, l \geq 2$

$$\bar{\mu}_k \bar{\mu}_l \leq \bar{\mu}_{k+l-2}\bar{\mu}_2, \qquad \bar{\mu}_k \bar{\eta}_l \leq \bar{\mu}_k \bar{\eta}_1 + \bar{\mu}_2 \bar{z}_{k+l-2}. \tag{3.8}$$

Proof. Note that if a r.v. Y has the mixture distribution $\frac{1}{n}\sum_i F_i$ where F_i denotes the distribution of $|X|+|U_1|+|U_2|+\cdots|U_{r+1}|$ with $X = X_i$, then the first required inequality becomes $EY^k EY^l \leq EY^{k+l-2}EY^2$. For the latter inequality, see, e.g., [32, p.74], and references therein. The second inequality in (3.8) will follow from the first if we write $\bar{\eta}_l = \bar{\eta}_1 + \bar{z}_l$. □

Let $f = \mathcal{S}(h)$ denote the Stein function for h. The next lemma follows from equation (41) in Lemma 5 of Barbour [1]; for $k = 1, 2$ it can be found in Stein [43].

Lemma 16. *For any* $k \geq 1$, *we have* $||f^{(k)}||_\infty \leq C(k)||h^{(k-1)}||_\infty$ *where the constant* $C(k)$ *depends only on* k.

We need more sophisticated properties of f. To clarify what will follow, note that f admits the representation (see [1]):

$$f(x) = \left(-I(x \geq 0)\int_0^\infty + I(x < 0)\int_{-\infty}^0\right) e^{-xz-z^2/2}[h(x+z) - \mathbf{\Phi}(h)]dz. \tag{3.9}$$

Differentiating (3.9), for $k \geq 1$ one can get that

$$f^{(k)}(x) = (-1)^{k-1}\mathbf{\Phi}(h)G(x;1;k,0) + \sum_{m=0}^{k}(-1)^{k-m}\binom{k}{m}G(x;h;k-m,m), \tag{3.10}$$

where

$$G(x;h;l,m) = \left(-I(x \geq 0)\int_0^\infty + I(x < 0)\int_{-\infty}^0\right) z^l e^{-xz-z^2/2}h^{(m)}(x+z)dz.$$

Set also $\mathbf{\Phi}_k(h) = \int_{-\infty}^\infty z^k h(z)\varphi(z)dz$ (so $\mathbf{\Phi}(h) = \mathbf{\Phi}_0(h)$), and for a r.v. W define

$$\begin{aligned}\Lambda_r(h;W) = &\max_{1\leq t\leq r^3,\, l\leq r^3,\, \nu\leq r} |E\{W^t G(W;h;l+t-1,\nu)\}| \\ &+ \max_{l\leq r^3,\, \nu\leq r} |E\{G(W;h;l,\nu)\}| + |E\{h^{(r)}(W)\}| \\ &+ \max_{l\leq r^3}|\mathbf{\Phi}_l(h)|.\end{aligned} \tag{3.11}$$

The complex expression in (3.11), and, in particular the term r^3 above, is required for the induction in the proofs of Theorem 2, as reflected in the following two lemmas whose proofs are given in Section 5.3. To avoid some superfluous explanations in proofs we assume below that integrals

$$\mathbf{\Phi}_l(h^{(k)}) \quad \text{are finite for integers } k \leq r,\ l \leq r^3. \tag{3.12}$$

Lemma 17. *For any integer $r \geq 2$, a sufficiently smooth h, and a r.v. W,*

$$|E\{f^{(r+1)}(W)\}| \leq C(r)\Lambda_r(h; W), \tag{3.13}$$

and for $p = 2, ..., r$,

$$\Lambda_{r+1-p}(f^{(p)}; W) \leq C(r)\Lambda_r(h; W). \tag{3.14}$$

Lemma 18. *Let W be a r.v. with a density $p(x)$ such that for some $m \leq r,\ m \geq 1$,*

$$p_m := \max_{k \leq m-1} \|p^{(k)}\|_\infty + \max_{k \leq m} ||p^{(k)}||_1 < \infty.$$

Then

$$\Lambda_r(h; W) \leq C(r)p_m[\max_{k=0,\dots,r-m} \|h^{(k)}\|_\infty]. \tag{3.15}$$

For $m = 0$ the bound (3.15) is always true with $p_0 = 1$.

Proof of Theorem 2. Note that for a given $X = X_i$ we have from (3.3) that $\alpha_2 = E\{XU_1\}$. We assume that the variables X_i are normalized so that $EX = 0$ and that $\bar{\alpha_2} = 1$. Also, assume that h, and hence the Stein function $f = \mathcal{S}(h)$ are sufficiently smooth so that all derivatives appearing below exist. Later we will remove the assumption by a standard smoothing argument. By Proposition 12

$$Eh(W) - \mathbf{\Phi}(h) = E\{f'(W) - Wf(W)\} = -\sum_{m=2}^{r} \frac{\bar{\alpha}_{m+1}}{m!} E\{f^{(m)}(W)\} - B_r(f), \tag{3.16}$$

where $|B_r(f)| \leq C(r)[\bar{\mu}_{r+2}\Gamma_{r+1}(f) + \bar{\eta}_{r+1}\hat{f}_r T]$.

Let $\Lambda_r(h; W; i, s, \omega)$ be defined in a manner similar to that of the quantity $\Lambda_r(h; W)$ in (3.11) except that the expectation E is replaced by $E\{\cdot|\mathcal{F}_{si}\}$. Then, by (3.13)

$$\Gamma_{r+1}(f) \leq C(r) \max_{1 \leq i \leq n} \max_{s \leq r+1} \sup_x Ess \sup_\omega \Lambda_r(h; \tilde{W}_{si} + x; i, s, \omega) =: C(r)\Lambda_r(h). \tag{3.17}$$

Thus

$$|B_r(f)| \leq C(r)[\bar{\mu}_{r+2}\Lambda_r(h) + \bar{\eta}_{r+1}\hat{f}_r T] \leq C(r)[\bar{\mu}_{r+2}\Lambda_r(h) + \bar{\eta}_{r+1}\hat{h}_{r-1}T], \tag{3.18}$$

by Lemma 16.

Consider the signed measure $Q_\nu(dx)$ defined in (1.12) and (1.13) with $\beta_l = \bar{\alpha}_l$.

Before proving Theorem 2 we prove a more cumbersome (but stronger, as shown later) result:

For $1 \leq t \leq r$,

$$Eh(W) - \mathbf{\Phi}(h) = \sum_{\nu=1}^{t-1} \int h(x)Q_\nu(dx) + K_t(h), \tag{3.19}$$

or equivalently

$$Eh(W)=\sum_{\nu=0}^{t-1}\int h(x)Q_\nu(dx)+K_t(h),\quad Q_0(dx)=\varphi(x)dx,\quad (3.20)$$

where $\quad |K_t(h)|\le C(t)(1+\bar z_{t+1})^{t-1}[\bar\mu_{t+2}\Lambda_t(h)+\bar\eta_{t+1}\hat h_{t-1}T].$

Note that the case $t=1$ is already established by (3.16) and (3.18) (with $r=1$, using also the relation $h(W)-\mathbf{\Phi}(h)=f'(W)-Wf(W)$ and Lemma 16), and we proceed by induction to prove the case $t=r$ assuming (3.20) holds for $1\le t\le r-1$. By (3.16), then by the induction hypothesis applied to $f^{(p)}$ with $t-1=r-p$, and finally by (1.12),

$$\begin{aligned}
&Eh(W)-\mathbf{\Phi}(h)\\
&\quad=-\sum_{p=2}^{r}\frac{\bar\alpha_{p+1}}{p!}\left(\sum_{\nu=0}^{r-p}\int f^{(p)}(x)Q_\nu(dx)+K_{r-p+1}(f^{(p)})\right)-B_r(f)\\
&\quad=-\sum_{p=2}^{r}\frac{\bar\alpha_{p+1}}{p!}\sum_{\nu=0}^{r-p}\sum\nolimits_{(\nu)}p_{\mathbf{k}\nu}\int f^{(p)}(x)L_{\nu+2s}(dx)\\
&\qquad-\sum_{p=2}^{r}\frac{\bar\alpha_{p+1}}{p!}K_{r-p+1}(f^{(p)})-B_r(f).
\end{aligned}\quad (3.21)$$

It is easy to see that (3.14) implies

$$\Lambda_{r+1-p}(f^{(p)})\le C(r)\Lambda_r(h).\quad (3.22)$$

Applying (3.4), Lemma 15, Lemma 16, (3.22), and the induction hypothesis on $K_{r-p+1}(f^{(p)})$, and again allowing $C(r)$ to vary between equations and to depend only on r, we obtain

$$\begin{aligned}
&\sum_{p=2}^{r}|\bar\alpha_{p+1}K_{r-p+1}(f^{(p)})|\\
&\quad\le C(r)\sum_{p=2}^{r}|\bar\mu_{p+1}(1+\bar z_{r-p+2})^{r-p}[\bar\mu_{r-p+3}\Lambda_{r+1-p}(f^{(p)})+\bar\eta_{r-p+2}\hat f_rT]\\
&\quad\le C(r)(1+\bar z_r)^{r-2}[\bar\mu_2\bar\mu_{r+2}\Lambda_r(h)+(\bar\mu_{p+1}\bar\eta_1+\bar\mu_2\bar z_{r+1})\hat h_{r-1}T]\\
&\quad\le C(r)(1+\bar z_{r+1})^{r-1}[\bar\mu_{r+2}\Lambda_r(h)+\bar\eta_{r+1}\hat h_{r-1}T].
\end{aligned}$$

This and (3.18) complete the induction step for the remainder $|K_r(h)|\le C_r(1+\bar\mu_2)^{r-1}[\bar\mu_{r+2}\Lambda_r(h)+\bar\eta_{r+1}\hat h_{r-1}T]$. Set $t+1=r+1=d$ in the above and define $R_d=R_{r+1}=K_r(h)$. By (3.19) R_d is the remainder in (1.15), and applying Lemma 18 to bound $\Lambda_r(h)$, we easily obtain (1.16). (The role of densities $p(x)$ in (3.15) is played by $q_s(x,\omega)$.)

Turning to the main term from (3.21), that is, $\sum_{p=2}^{r} \frac{\bar{\alpha}_{p+1}}{p!} \sum_{\nu=0}^{r-p} \sum_{(\nu)} p_{\mathbf{k}\nu} \int f^{(p)}(x) L_{\nu+2s}(dx)$, and using the formula (see [1])

$$\int f^{(p)}(x) L_\nu(dx) = -\frac{1}{p+\nu+1} \int h(x) L_{\nu+p+1}(dx), \tag{3.23}$$

it is straightforward to verify that

$$\begin{aligned} &-\sum_{p=2}^{r} \frac{\bar{\alpha}_{p+1}}{p!} \sum_{\nu=0}^{r-p} \sum\nolimits_{(\nu)} p_{\mathbf{k}\nu} \int f^{(p)}(x) L_{\nu+2s}(dx) \\ &= \sum_{l=1}^{r-1} \sum_{\nu+p=l+1} \frac{\bar{\alpha}_{p+1}}{p!} \sum\nolimits_{(\nu)} p_{\mathbf{k}\nu} \frac{1}{l+2(s+1)} \int h(x) L_{l+2(s+1)}(dx). \end{aligned} \tag{3.24}$$

In order to complete the proof of Theorem 2 we have to show that the latter expression coincides with the three equal (because $r-1=d-2$ and by (1.12)) expressions below:

$$\sum_{\nu=1}^{d-2} \int h(x) Q_\nu(dx) = \sum_{l=1}^{r-1} \int h(x) Q_l(dx) = \sum_{l=1}^{r-1} \int h(x) \sum\nolimits_{(l)} p_{\mathbf{k}l} L_{l+2s}(dx).$$

Thus we have to show (setting $p=i+1$ in (3.24)) that for any $l=1,\ldots,r-1$

$$\sum_{\nu+i=l} \frac{\bar{\alpha}_{i+2}}{(i+1)!} \sum\nolimits_{(\nu)} p_{\mathbf{k}\nu} \frac{1}{l+2(s+1)} L_{l+2(s+1)} = \sum\nolimits_{(l)} p_{\mathbf{k}l} L_{l+2s},$$

where in the left-hand side sum $1 \le i \le l$ and $0 \le \nu \le l-1$. Fix now s_0. Then it suffices to show that

$$\begin{aligned} &\sum_{\nu+i=l} \frac{\bar{\alpha}_{i+2}}{(i+1)!} \sum\nolimits_{(\nu),\, s(\mathbf{k}')=s_0-1} p_{\mathbf{k}'\nu} \frac{1}{l+2(s(\mathbf{k}')+1)} L_{l+2(s(\mathbf{k}')+1)} \\ &\quad = \sum\nolimits_{(l),\, s(\mathbf{k})=s_0} p_{\mathbf{k}l} L_{l+2s_0}, \end{aligned}$$

or equivalently

$$\sum_{\nu+i=l} \frac{\bar{\alpha}_{i+2}}{(i+1)!} \sum\nolimits_{(\nu),\, s(\mathbf{k}')=s_0-1} p_{\mathbf{k}'\nu} \frac{1}{l+2s_0} = \sum\nolimits_{(l),\, s(\mathbf{k})=s_0} p_{\mathbf{k}l}.$$

Fixing $\mathbf{k}$ with $s(\mathbf{k})=s_0$ and setting $\mathbf{k}(i)=\mathbf{k}-\mathbf{e}_i$ where $\mathbf{e}_i$ is the vector with all components equal to zero except for the ith which equals 1, the latter relation can be seen to follow from

$$\sum_{i=1}^{l} \frac{\bar{\alpha}_{i+2}}{(i+1)!} p_{\mathbf{k}(i)\nu} \frac{1}{l+2s_0} = p_{\mathbf{k}l}. \tag{3.25}$$

We rewrite the left-hand side as

$$\sum_{i=1}^{l} \frac{k_i(i+2)}{l+2s_0} \frac{\bar{\alpha}_{i+2}}{k_i(i+2)!} p_{\mathbf{k}(i),(l-i)}.$$

Noting that

$$\frac{\bar{\alpha}_{i+2}}{k_i(i+2)!} p_{\mathbf{k}(i),(l-i)} = p_{\mathbf{k}l}, \quad \text{and} \quad \sum_{i=1}^{l} \frac{k_i(i+2)}{l+2s_0} = 1,$$

we obtain (3.25) and the induction is complete.

This proves Theorem 2 for a smooth function h. A standard smoothing argument leads to the result for any h satisfying (1.4); see, e.g., Lemma 11.4 in Bhattacharya and Ranga Rao [5]. We omit here standard details, and note only the following. First, to apply the lemma mentioned one should take into account that $\|h^{(k)}_{\varepsilon+}\|_\infty \leq \|h^{(k)}\|_\infty$, $\|h^{(k)}_{\varepsilon-}\|_\infty \leq \|h^{(k)}\|_\infty$ for any k. Second, $||h^{(k)}||_\infty$ appearing in the bound of (1.16) of Theorem 2 is interpreted as infinity if the kth derivative of h does not exist. On the other hand, if $||h^{(1)}||_\infty < \infty$, then (1.4) clearly holds. □

4. On cumulants and Theorem 3

In this section we assume $EW^2 = 1$. Let $D_l = \max_{i\leq n} \max_{s\leq d} E\{|\tilde{W}_{si}|^l\}$.

First,

$$D_l \leq 2^{l-1}\{\max_{i\leq n} \mu_{li} + E\{|W|^l\}\} \leq 2^{l-1}\{\bar{\mu}_l + E\{|W|^l\}\}. \tag{4.1}$$

In particular, $D_2 \leq 2(1+\bar{\mu}_2)$. We have also $D_1 \leq \sqrt{2(1+\bar{\mu}_2)}$, $D_0 = 1$.

Lemma 19. *Assume*

$$T' \leq 1, \quad \textit{and} \quad T' \sum_{j=1}^{n} E\{|X_j|\} \leq 1/2^{m+1} \tag{4.2}$$

for an even m. Then for $p = 1, ..., m/2$, and some constant $C(p)$

$$D_{2p} \leq C(p)(1+\bar{z}_{2p})^p, \tag{4.3}$$

where again $\bar{z}_k = \sum_{l=2}^{k} \bar{\mu}_l$.

Remark. In view of previous discussions, the quantity $\bar{z}_k$ is typically bounded. The bound (4.3) is quite rough, but it suffices for our purpose. Also, note that the quantity on the left-hand side of (4.2) is smaller than $\bar{\eta}_1 T'$.

Proof. For $p = 1$ (4.3) is true. We provide the induction step from p to $p+1$, making use of (3.5) and (3.2). We apply (3.5) to the function $f(w) = w^{2p+1}$; the remainder in this case vanishes, and we have, allowing the constant $C(p)$ to vary as usual,

$$
\begin{aligned}
E\{W^{2p+2}\} &= \sum_{j=1}^{n} E\{X_j W^{2p+1}\} = \sum_{j=1}^{n} E\{X_j \tilde{W}_{1j}^{2p+1}\} \\
&\quad + \sum_{j=1}^{n} \sum_{l=1}^{2p+1} \sum_{s=1}^{l} \sum_{|\mathbf{m}|=l} {}^{(s)} \frac{(2p+1)!}{(2p+1-l)!\,\mathbf{m}!} E\left\{X_j \mathbf{U}_j^{\mathbf{m}} (\tilde{W}_{s+1,j})^{2p+1-l}\right\} \\
&\le T'(\sum_{j=1}^{n} E\{|X_j|\}) D_{2p+1} + (1+T')C(p) \sum_{l=1}^{2p+1} (\sum_{j=1}^{n} \mu_{l+1,j}) D_{2p+1-l} \\
&\le T'(\sum_{j=1}^{n} E\{|X_j|\}) D_{2p+1} + (1+T')C(p) \sum_{l=1}^{2p+1} \bar{\mu}_{l+1} D_{2p+1-l}. \qquad (4.4)
\end{aligned}
$$

Assume that $D_{2p+2} \ge 1$, otherwise (4.3) is trivial. By the condition of the lemma, $T'(\sum_j E\{|X_j|\}) \le 1/2^{2p+2}$. Then from (4.1), (4.4) it follows that

$$
\begin{aligned}
D_{2p+2} &\le 2^{2p+1}(\bar{\mu}_{2p+2} + E\{W^{2p+2}\}) \\
&\le 2^{2p+1}\bar{\mu}_{2p+2} + (1/2)(D_{2p+2})^{(2p+1)/(2p+2)} + C(p) \sum_{l=1}^{2p+1} \bar{\mu}_{l+1} D_{2p+1-l} \\
&\le 2^{2p+1}\bar{\mu}_{2p+2} + (1/2)D_{2p+2} + C(p) \sum_{l=1}^{2p+1} \bar{\mu}_{l+1} D_{2p+1-l},
\end{aligned}
$$

and hence,

$$
D_{2p+2} \le 2^{2p+2}\bar{\mu}_{2p+2} + C(p) \sum_{l=1}^{2p+1} \bar{\mu}_{l+1} D_{2p+1-l}.
$$

If l is odd, then by induction,

$$
\begin{aligned}
\bar{\mu}_{l+1} D_{2p+1-l} &\le \bar{\mu}_{l+1} C(p)(1+\bar{z}_{2p+1-l})^{(2p+1-l)/2} \\
&\le C(p)(1+\bar{z}_{2p+2})^{p+1}.
\end{aligned}
$$

If l is even, and consequently $2p+1-l$ is odd, we write

$$
\begin{aligned}
\bar{\mu}_{l+1} D_{2p+1-l} &\le \bar{\mu}_{l+1} D_{2p+2-l}^{(2p+1-l)/(2p+2-l)} \\
&\le \bar{\mu}_{l+1} C(p)(1+\bar{z}_{2p+2-l})^{(2p+1-l)/2} \le C(p)(1+\bar{z}_{2p+2})^{p+1}.
\end{aligned}
$$

It remains to combine the above bounds. □

We turn to comparing γ and $\bar{\alpha}$. For $k \le r$ we have

$$E\{W^{k+1}\} = \sum_{m=0}^{k} \binom{k}{m} \gamma_{m+1} E\{W^{k-m})\}.$$

This known relation is a particular case of (1.1), see also [1].

We need also a version of Proposition 12 for $f(w) = w^k$ with T' replacing T. Going through the proof of the proposition mentioned for this particular function, and noting that in this case $\Gamma_{k+1} \equiv 0$, one can easily see that

$$E\{W^{k+1}\} = \sum_{m=0}^{k} \binom{k}{m} \bar{\alpha}_{m+1} E\{W^{k-m}\} + R(k),$$

with a remainder satisfying

$$|R(k)| \le C(k)\bar{\eta}_{k+1} \left(\max_{i,s} \max_{l \le k} E\{|\tilde{W}_{si}^{l}|\} \right) T' = C(k)\bar{\eta}_{k+1} \left(\max_{l \le k} D_l \right) T'. \quad (4.5)$$

It is straightforward to verify that if

$$T' \le 1, \quad \bar{\eta}_1 T' < 1/2^{k+1}, \quad (4.6)$$

then (4.5) and (4.3) imply

$$|R(k)| \le C(k)\bar{\eta}_{k+1}(1 + \bar{z}_{k+1})^{k/2} T'$$

for k odd and even as well. Thus

$$|\gamma_{k+1} - \bar{\alpha}_{k+1}| \le \sum_{m=0}^{k-1} \binom{k}{m} |(\gamma_{m+1} - \bar{\alpha}_{m+1}) E\{W^{k-m}\}| + C(k)\bar{\eta}_{k+1}(1 + \bar{z}_{k+1})^{k/2} T'. \quad (4.7)$$

The last step is to derive from this and (4.3) by induction that if (4.6) holds, then

$$|\gamma_{k+1} - \bar{\alpha}_{k+1}| \le C(k)\bar{\eta}_{k+1}(1 + \bar{z}_{k+1})^{k/2} T'. \quad (4.8)$$

To this end it suffices first to observe that the bound (4.3) is certainly true for $E\{W^{2p}\}$ [one can, for example, write $E\{W^{2p}\} \le 2^{2p-1}(EX_1{}^{2p} + W_{11}^{2p})$]; and second to realize that if $(k - m)$ is even, then by induction

$$\begin{aligned}
|(\gamma_{m+1} - \bar{\alpha}_{m+1}) E\{W^{k-m}\}| &\le C(k)\bar{\eta}_{m+1}(1 + \bar{z}_{m+1})^{m/2} T' (1 + \bar{z}_{k-m})^{(k-m)/2} \\
&\le C(k)\bar{\eta}_{k+1}(1 + \bar{z}_{k+1})^{m/2} T' (1 + \bar{z}_{k+1})^{(k-m)/2} \le C(k)\bar{\eta}_{k+1}(1 + \bar{z}_{k+1})^{k/2} T',
\end{aligned}$$

and if $(k - m)$ is odd, then by induction

$$\begin{aligned}
&|(\gamma_{m+1} - \bar{\alpha}_{m+1}) E\{W^{k-m}\}| \\
&\quad \le C(k)\bar{\eta}_{m+1}(1 + \bar{z}_{m+1})^{m/2} T' (E\{W^{k-m+1}\})^{(k-m)/(k-m+1)} \\
&\quad \le C(k)\bar{\eta}_{k+1}(1 + \bar{z}_{k+1})^{m/2} T' ((1 + \bar{z}_{k-m+1})^{(k-m+1)/2})^{(k-m)/(k-m+1)} \\
&\quad \le C(k)\bar{\eta}_{k+1}(1 + \bar{z}_{k+1})^{k/2} T'.
\end{aligned}$$

This proves Proposition 4. □

In general, if (4.6) holds for $k = r$, then

$$\begin{aligned}
&\left|\sum_{m=1}^{r} \frac{\gamma_{m+1}}{m!} E\{f^{(m)}(W)\} - \sum_{m=1}^{r} \frac{\bar{\alpha}_{m+1}}{m!} E\{f^{(m)}(W)\}\right| \\
&\quad \le \sum_{m=1}^{r} \frac{|\gamma_{m+1} - \bar{\alpha}_{m+1}|}{m!} |E\{f^{(m)}(W)\}| \\
&\quad \le C(r)\bar{\eta}_{r+1}(1+\bar{z}_{r+1})^{r/2} T' \max_{m\le r} |E\{f^{(m)}(W)\}| \\
&\quad \le C(r)\bar{\eta}_r(1+\bar{z}_{r+1})^{r/2}\, \hat{f}_r T'. \qquad (4.9)
\end{aligned}$$

Proof of Theorem 3. We repeat most of the proof of Theorem 2 and highlight the differences. We now assume $EW^2 = 1$ instead of $\bar{\alpha}_2 = 1$, the assumption that led to (3.16). Also, assume that h, and hence the Stein function $f = \mathcal{S}(h)$ are sufficiently smooth so that all derivatives appearing below exist.

Note also that, since Theorem 2 does not involve the characteristics Ψ's, the proof is simpler and does not use Lemma 17.

By (3.16), (3.18), (3.15), (4.9), (1.19), (1.20), and the facts that $T \le T'$ and $\bar{\eta}_r \le \bar{\eta}_{r+1}$, we have

$$Eh(W) - \mathbf{\Phi}(h) = E\{f'(W) - Wf(W)\} = -\sum_{m=2}^{r} \frac{\gamma_{m+1}}{m!} E\{f^{(m)}(W)\} - B_r(f), \qquad (4.10)$$

where

$$|B_r(f)| \le C(r)[\bar{\mu}_{r+2}\hat{h}_r + \bar{\eta}_{r+1}(1+\bar{z}_{r+1})^{r/2} \hat{f}_r T']. \qquad (4.11)$$

In analogy to the proof of Theorem 2, we now prove the following by induction: if

$$\bar{\eta}_{r+1} T' \le 2^{-r-1}, \quad \text{and} \quad T' \le 1, \qquad (4.12)$$

then for $1 \le t \le r$

$$Eh(W) - \mathbf{\Phi}(h) = \sum_{\nu=1}^{t-1} \int h(x) Q_\nu(dx) + K_t(h), \qquad (4.13)$$

or equivalently

$$Eh(W) = \sum_{\nu=0}^{t-1} \int h(x) Q_\nu(dx) + K_t(h), \quad Q_0(dx) = \varphi(x)dx, \qquad (4.14)$$

$$\text{where} \qquad |K_t(h)| \le C(t)(1+\bar{z}_{t+1})^{2t}\hat{h}_t[\bar{\mu}_{t+2} + \bar{\eta}_{t+1}T'].$$

Note that the case $t = 1$ is already established by (4.10) and (4.11) (with $r = 1$, using also the relation $h(W) - \mathbf{\Phi}(h) = f'(W) - Wf(W)$ and Lemma 16), and we proceed by induction to prove the case $t = r$ assuming (4.14) holds for $1 \le t \le r-1$.

It is obvious also that (4.12) implies $\bar{\eta}_{t+1}T' \leq 2^{-t-1}$ for $t < r$. Then by (4.10), by the induction hypothesis applied to $f^{(p)}$ with $t - 1 = r - p$, and finally by (1.12),

$$\begin{aligned}
&Eh(W) - \mathbf{\Phi}(h) \\
&\quad = -\sum_{p=2}^{r} \frac{\gamma_{p+1}}{p!} \left(\sum_{\nu=0}^{r-p} \int f^{(p)}(x) Q_\nu(dx) + K_{r-p+1}(f^{(p)}) \right) - B_r(f) \\
&\quad = -\sum_{p=2}^{r} \frac{\gamma_{p+1}}{p!} \sum_{\nu=0}^{r-p} \sum\nolimits_{(\nu)} p_{\mathbf{k}\nu} \int f^{(p)}(x) L_{\nu+2s}(dx) \\
&\qquad - \sum_{p=2}^{r} \frac{\gamma_{p+1}}{p!} K_{r-p+1}(f^{(p)}) - B_r(f).
\end{aligned}$$

From (4.8) and (3.4) it follows that

$$|\gamma_{p+1}| \leq C(p) \left\{ \bar{\mu}_{p+1} + \bar{\eta}_{p+1}(1 + \bar{z}_{p+1})^{p/2} T' \right\}. \tag{4.15}$$

Applying Lemma 15, Lemma 16, (4.15) and the induction hypothesis on $K_{r-p+1}(f^{(p)})$, and again allowing $C(r)$ to vary between equations and to depend only on r, we obtain that for $p = 2, ..., r$

$$\begin{aligned}
&|\gamma_{p+1} K_{r-p+1}(f^{(p)})| \\
&\quad \leq C(p) \left[\bar{\mu}_{p+1} + \bar{\eta}_{p+1}(1 + \bar{z}_{p+1})^{p/2} T' \right] \\
&\qquad \times \left[(1 + \bar{z}_{r-p+2})^{2(r-p+1)} \hat{f}_{r+1} [\bar{\mu}_{r-p+3} + \bar{\eta}_{r-p+2} T'] \right] \\
&\quad \leq C(r)(1 + \bar{z}_r)^{2(r-1)} \hat{h}_r \left[\bar{\mu}_2 \bar{\mu}_{r+2} + (\bar{\mu}_{p+1}\bar{\eta}_1 + \bar{\mu}_2 \bar{z}_{r+1}) T' \right] \\
&\qquad + C(r)(1 + \bar{z}_{r+1})^{2r+2-3p/2} \bar{\eta}_{p+1} \hat{h}_r T' \left[\bar{\mu}_{r-p+3} + \bar{\eta}_{r-p+2} T' \right] \\
&\quad \leq C(r)\hat{h}_r \left\{ (1 + \bar{z}_r)^{2(r-1)} \left[\bar{\mu}_2 \bar{\mu}_{r+2} + \bar{z}_{r+1}\bar{\eta}_{r+1} T' \right] \right. \\
&\qquad \left. + (1 + \bar{z}_{r+1})^{2r-1} T' \left[\bar{\eta}_{r+1}\bar{\mu}_{r+1} + \bar{\eta}_{r+1}\bar{\eta}_r T' \right] \right\} \\
&\quad \leq C(r)\hat{h}_r \left\{ (1 + \bar{z}_{r+1})^{2r-1} \left[\bar{\mu}_{r+2} + \bar{\eta}_{r+1} T' \right] \right. \\
&\qquad \left. + (1 + \bar{z}_{r+1})^{2r} \left[\bar{\eta}_{r+1} T' + (\bar{\eta}_{r+1} T')^2 \right] \right\} \\
&\quad \leq C(r)(1 + \bar{z}_{r+1})^{2r} \hat{h}_r \left[\bar{\mu}_{r+2} + \bar{\eta}_{r+1} T' \right],
\end{aligned}$$

if $\quad \bar{\eta}_{r+1} T' \leq 2^{-r-1}$.

This, (4.11) and Lemma 16 complete the induction step for the remainder part. Set $t + 1 = r + 1 = d$ in the above and define $R_d = R_{r+1} = K_r(h)$. By (4.13) R_d is the remainder in the l.-h.s.. of (1.22), so we easily obtain (1.22).

The main term is treated exactly as in the proof of Theorem 2, and the proof is complete. □

5. Proofs of lemmas

5.1. Proofs of Lemmas 6 and 8

Proof of Lemma 6. As in Section 2.3 we set $\mathcal{N}_{1i} = \{i\} \cup \{K(i)\} \cup \{j;\ K(j) = i$ or $K(i)\}$. Let $V_{1i} = \sum_{j \in \mathcal{N}_{1i}} Y_j$. Then $E\{S^2 \,|\, \mathbf{K}_n\} = \sum_{i=1}^n E\{Y_i V_{1i} \,|\, \mathbf{K}_n\}$. It is straight forward to verify that $E\{Y_i V_{1i} \,|\, \mathbf{K}_n\} = \sigma^2 + (\sigma^2 - 2c)I_i + cN_i + cN_{K(i)}$. Taking into account that $\sum_{i=1}^n N_{K(i)} = \sum_{i=1}^n N_i^2$, and certainly $\sum_{i=1}^n N_i = n$, we have (2.8). The representations for $m_{1n},\ m_{2n}$, and hence (2.9), are straightforward. Note now that for any l

$$E\left\{N_i^l\right\} \le C(l)E\left\{e^{N_i}\right\} \le C(l)\left\{\exp\{(e-1)\sum_{j=1}^n p_{ji}\}\right\} \le C(l)e^{2\lambda}. \tag{5.1}$$

It is easy to calculate that

$$Var\left\{\sum_{i=1}^n I_i\right\} \le 2n, \quad Var\left\{\sum_{i=1}^n N_i^2\right\} \le \sum_{i=1}^n Var\left\{N_i^2\right\} \le Ce^{2\lambda}n \tag{5.2}$$

in view of (5.1).

This implies (2.12). To prove (2.10), observe that $\sum_{i=1}^n N_i^2 \ge n$, and hence $\Theta_n^2 \ge \sigma^2 + c - 2c + c = \sigma^2 > 0$. Furthermore, by (2.8), (5.1), and since $c < \sigma^2$,

$$\begin{aligned} E\{\Theta_n^{2l}\} &\le E\left\{\left(3\sigma^2 + \frac{\sigma^2}{n}\sum_{i=1}^n N_i^2\right)^l\right\} \le C(l)\sigma^{2l}\left\{1 + \frac{1}{n}\sum_{i=1}^n E\{N_i^{2l}\}\right\} \\ &\le C(l)\sigma^{2l}e^{2\lambda}. \end{aligned}$$

□

Proof of Lemma 8. By construction,

$$\begin{aligned} E\left|\mathcal{N}_{(s+1)i}\right|^l &\le E\left(|\mathcal{N}_{si}| + N_{K_{s+1}(i)} + \sum_{j \in \mathcal{N}_{si} \setminus \mathcal{N}_{(s-1)i}} N_j\right)^l \\ &\le C(l)E\left\{|\mathcal{N}_{si}|^l + N^l_{K_{s+1}(i)} + |\mathcal{N}_{si} \setminus \mathcal{N}_{(s-1)i}|^{l-1} \sum_{j \in \mathcal{N}_{si} \setminus \mathcal{N}_{(s-1)i}} N_j^l\right\}. \end{aligned}$$

For $s \le d$

$$\begin{aligned} E\{N^l_{K_s(i)}\} &= E\{E\{N^l_{K_s(i)}\Big|\, K(i), K_2(i), ..., K_s(i)\}\} \\ &\le E\Bigg\{E\Bigg\{\left(s + \sum_m {}^{(K(i),...,K_s(i))} I(K(m) = K_s(i))\right)^l \Bigg| \\ &\qquad K(i), K_2(i), ..., K_s(i)\Bigg\}\Bigg\} \\ &\le 2^l\left[s^l + \sum_{j=1}^n P(K_d(i)\} = j)E\{N_j^l\}\right] \le C(d,l)e^{2\lambda} \end{aligned}$$

in view of (5.1). Next we write

$$E\left\{|\mathcal{N}_{si}\setminus\mathcal{N}_{(s-1)i}|^{l-1}\sum_{j\in\mathcal{N}_{si}\setminus\mathcal{N}_{(s-1)i}}N_j^l\right\}$$
$$\leq E\left\{|\mathcal{N}_{si}|^{l-1}\sum_{j\in\mathcal{N}_{si}\setminus\mathcal{N}_{(s-1)i}}E\left\{N_j^l\middle|\mathcal{N}_{1i},...,\mathcal{N}_{si}\right\}\right\}.$$

As in (5.1), taking into account the structure of the sets $\mathcal{N}$'s described above, and denoting by $\mathcal{N}^c$ the complement of $\mathcal{N}$, it is easy to verify that for $j\in\mathcal{N}_{si}\setminus\mathcal{N}_{(s-1)i}$

$$E\left\{N_j^l\middle|\mathcal{N}_{1i},...,\mathcal{N}_{si}\right\}\leq C(l)E\left\{\exp\{N_j\}\middle|\mathcal{N}_{1i},...,\mathcal{N}_{si}\right\}$$
$$\leq C(l)\exp\left\{(e-1)\sum_{k\in\mathcal{N}_{si}^c}\left(p_{kj}\Big/\sum_{m\in\mathcal{N}_{(s-1)i}^c}p_{km}\right)\right\}\leq C(l)e^{2\lambda}$$

in view of condition (2.7). Combining above bounds for $s\leq d-1$ we have finally that

$$E\left|\mathcal{N}_{(s+1)i}\right|^l\leq C(l,d)e^{2\lambda}E\left|\mathcal{N}_{si}\right|^l. \tag{5.3}$$

On the other hand, since $|\mathcal{N}_{1i}|=1+N_i+N_{K(i)}-I_i$, (5.1), the inequality

$$E\left|N_{K(i)}\right|^l=\sum_{j=1}^n p_{ij}E\{N_j^l\,|K(i)=j\}\leq\sum_{j=1}^n p_{ij}E\{(1+N_j)^l\}\leq C(l)e^{2\lambda},$$

and (5.3) imply (2.16). □

5.2. Proof of Lemma 14

Before proving the first equality in (3.6), note that the second one follows simply from the facts (see definitions) that $c(s,0)=1,\ c(s,1)=0$.

We now prove the first equality in (3.6). Replacing the vector $\mathbf{U}$ in (3.5) by $\mathbf{V}_s$, and $\tilde{W}_1$ by $\tilde{W}_s$, taking expectations, using (1.8) and $E\{W_s\}=0$ (which allows us to start the sum below from $l=2$), we have that for any $k\leq r+1-s$

$$\begin{aligned}E\{f(W)\}&=E\{f(\tilde{W}_s)\}+\sum_{l=2}^k\sum_{p=1}^l\sum_{|\mathbf{m}|=l}{}^{(p)}\frac{1}{\mathbf{m}!}E\{\mathbf{V}_\mathbf{s}^\mathbf{m}\}E\{f^{(l)}(\tilde{W}_{s+p})\}\\&\quad+E\{R_{ks}\}+T_{ks}\\&=E\{f(\tilde{W}_s)\}+\sum_{l=2}^k\sum_{p=1}^l\chi_l(s,p)E\{f^{(l)}(\tilde{W}_{s+p})\}\\&\quad+E\{R_{ks}\}+T_{ks},\end{aligned}\tag{5.4}$$

where

$$\begin{aligned}
&|E\{R_{ks}\}| \\
&\quad = |E\{\sum_{p=1}^{k+1} \sum_{|\mathbf{m}|=k+1}{}^{(p)} \frac{m_p}{\mathbf{m}!} \mathbf{V}_{\mathbf{s}}^{\mathbf{m}} \int_0^1 (1-t)^{(m_p-1)} f^{(k+1)}(\tilde{W}_{s+p-1} + tU_{s+p-1}) dt\}| \\
&\quad \leq \mu_{k+1} \Gamma_{k+1}(f),
\end{aligned}$$

$$T_{ks} = \sum_{l=1}^{k} \sum_{p=1}^{l} \sum_{|\mathbf{m}|=l}{}^{(p)} \frac{1}{\mathbf{m}!} \left[E\{\mathbf{V}_{\mathbf{s}}^{\mathbf{m}} f^{(l)}(\tilde{W}_{s+p})\} - E\{\mathbf{V}_{\mathbf{s}}^{\mathbf{m}}\} E\{f^{(l)}(\tilde{W}_{s+p})\} \right],$$

so that

$$|T_{ks}| \leq \hat{f}_k \left[\sum_{l=1}^{k} \mu_l \right] T = \hat{f}_k \eta_k T.$$

Denoting $H_{ks}(f) = E\{R_{ks}\} + T_{ks}$ and $H_{0s} = 0$, we have $|H_{ks}(f)| \leq \mu_{k+1} \Gamma_{k+1}(f) + \hat{f}_k \eta_k T$. Thus,

$$\begin{aligned}
&E\{f(\tilde{W}_s)\} \\
&\quad = E\{f(W)\} - H_{ks}(f) - \sum_{l=2}^{k} \sum_{p=1}^{l} \chi_l(s,p) E\{f^{(l)}(\tilde{W}_{s+p})\} \\
&\quad = E\{f(W)\} - H_{ks}(f) - \sum_{l=2}^{k} \sum_{p=1}^{l} \chi_l(s,p) [E\{f^{(l)}(W)\} - H_{k-l,s+p}(f^{(l)})] \\
&\qquad + \sum_{l_1=2}^{k} \sum_{p_1=1}^{l} \chi_{l_1}(s,p_1) \sum_{l_2=2}^{k-l_1} \sum_{p_2=1}^{l_2} \chi_{l_2}(s+p_1,p_2) E\{f^{(l_1+l_2)}(\tilde{W}_{s+p_1+p_2})\},
\end{aligned} \tag{5.5}$$

where the last expression was obtained by applying the first relation to $f^{(l)}$. Therefore,

$$\begin{aligned}
|H_{k-l,s+p}(f^{(l)})| &= |E\{R_{k-l,s+p}(f^{(l)})\} + T_{k-l,s+p}(f^{(l)})| \\
&\leq \mu_{k-l+1} \Gamma_{k+1}(f) + \hat{f}_k \eta_{k-l} T.
\end{aligned} \tag{5.6}$$

We first treat only the main part, neglecting the remainder containing all terms with H's, using $\approx$ to indicate this approximation. Repeating the above scheme we obtain

$$\begin{aligned}E\{f(\tilde{W}_s)\} &\approx E\{f(W)\} - \sum_{l=2}^{k} \Upsilon_{l1}(s)E\{f^{(l)}(W)\} + \sum_{l=4}^{k} \Upsilon_{l2}(s)E\{f^{(l)}(W)\} \\ &\qquad - \sum_{l=6}^{k} \Upsilon_{l3}(s)E\{f^{(l)}(W)\} + ... \\ &\qquad + (-1)^{[k/2]} \sum_{l=2[k/2]}^{k} \Upsilon_{l[k/2]}(s)E\{f^{(l)}(W)\} \\ &= \sum_{m=0}^{[k/2]}(-1)^m \sum_{l=2m}^{k} \Upsilon_{lm}(s)E\{f^{(l)}(W)\} = \sum_{l=0}^{k} c(s,l)E\{f^{(l)}(W)\} \\ &= E\{f(W)\} + \sum_{l=2}^{k} c(s,l)E\{f^{(l)}(W)\}.\end{aligned}$$

It remains to consider the remainder. Using the bounds in (5.6) it is easy to see that the remainder in the above expression for $E\{f(\tilde{W}_s)\}$, which we denote by $M_{sk}(f)$, satisfies:

$$\begin{aligned}|M_{sk}(f)| &\leq |\sum_{m=0}^{[k/2]} \sum_{l=2m}^{k} \Upsilon_{lm}(s)[\mu_{k+1-l}\Gamma_{k+1-l}(f^{(l)}) + \hat{f}_k\eta_{k-l}T]| \\ &\leq \sum_{l=0}^{k} \tilde{c}(s,l)[\mu_{k+1-l}\Gamma_{k+1}(f) + \hat{f}_k\eta_{k-l}T],\end{aligned}$$

and the result follows easily using (3.4).

5.3. *Proofs of Lemmas 17 and 18*

We give first a

Proof of Lemma 18. Note that

$$\begin{aligned}&E\{W^t G(W;h;l,\nu)\} \\ &\quad = -\int_0^\infty \left(\int_0^\infty z^l e^{-xz-z^2/2} h^{(\nu)}(x+z)dz\right) x^t p(x)dx \\ &\qquad + \int_{-\infty}^0 \left(\int_{-\infty}^0 z^l e^{-xz-z^2/2} h^{(\nu)}(x+z)dz\right) x^t p(x)dx. \qquad (5.7)\end{aligned}$$

Let first $\nu \geq r-m$, $t \geq 0$, and $g(x,z;\ t) = x^t e^{-xz} p(x)$. Throughout this proof, when integrating by parts, we use the boundedness of the derivatives of h and that $p^{(k)}$ vanishes at infinity for $k \leq m-1$. (This is true due to the fact that, if for a differentiable function $p(x)$ its derivative is Lebesgue integrable, then $p(x)$

is absolutely continuous and has limits at $\pm\infty$; see, e.g., [10]. Therefore, since p is a density, $p(\pm\infty) = 0$.)

Changing the order of integration in the above double integrals, and integrating by parts $s = \nu + m - r$ times in the interior integrals, one would easily get that

$$\begin{aligned}
&E\{W^t G(W; h; l, \nu)\} \\
&= \int_{-\infty}^{\infty} z^l e^{-z^2/2} \Big(h^{(\nu-1)}(z) g(0, z;\ t) - h^{(\nu-2)}(z) g'_x(0, z;\ t) + ... \\
&+ (-1)^{s-1} h^{(\nu-s)}(z) g_x^{(s-1)}(0, z;\ t) \Big) dz \\
&- (-1)^s \int_0^{\infty} z^l e^{-z^2/2} \left(\int_0^{\infty} h^{(\nu-s)}(x+z) g_x^{(s)}(x, z;\ t) dx \right) dz \\
&+ (-1)^s \int_{-\infty}^{0} z^l e^{-z^2/2} \left(\int_{-\infty}^{0} h^{(\nu-s)}(x+z) g_x^{(s)}(x, z;\ t) dx \right) dz,
\end{aligned}$$

where $g_x^{(k)} = \partial^k g / \partial x^k$. (The main point here is that some of the integrals $\int_0^\infty$ and $\int_{-\infty}^0$ can be united in one integral $\int_{-\infty}^\infty$.)

For the first type of the above integrals, we can write that for $k \le s$

$$\begin{aligned}
&\left| \int_{-\infty}^{\infty} z^l e^{-z^2/2} h^{(\nu-k)}(z) g_x^{(k-1)}(0, z;\ t) dz \right| \\
&= \left| \int_{-\infty}^{\infty} h^{(\nu-s)}(z) \mathcal{D}_z^{s-k} [z^l e^{-z^2/2} g_x^{(k-1)}(0, z;\ t)] dz \right| \\
&\le \|h^{(\nu-s)}\|_\infty \int_{-\infty}^{\infty} |\mathcal{D}_z^{s-k} [z^l e^{-z^2/2} g_x^{(k-1)}(0, z;\ t)]| dz \\
&\le C(t, l, \nu) \|h^{(\nu-s)}\|_\infty \max_{k \le s-1} |p^{(k)}(0)| \\
&\le C(t, l, \nu) \|h^{(r-m)}\|_\infty \max_{k \le m-1} |p^{(k)}(0)|, \qquad (5.8)
\end{aligned}$$

since $s = \nu + m - r$.

Regarding the second type of integrals, first note that, as a matter of fact we should estimate expressions of the type $E\{W^t G(W; h; l+t-1, \nu)\}$ (see the definition of Λ_r in (3.11)). So, when working with $E\{W^t G(W; h; l, \nu)\}$ we can assume $l \ge t - 1$.

Furthermore, $g_x^{(s)}(x, z;\ t)$ is a linear combination of terms $x^{t-k} z^j e^{-xz} p^{(i)}(x)$, where $k + j + i = s$. If $l \ge t - 1$, noticing again that $\nu - s = r - m$, we have for $\nu \ge r - m$

$$\begin{aligned}
&\left| \int_0^{\infty} z^l e^{-z^2/2} \left(\int_0^{\infty} h^{(\nu-s)}(x+z) x^{t-k} z^j e^{-xz} p^{(i)}(x) dx \right) dz \right| \\
&\le \|h^{(\nu-s)}\|_\infty \int_0^{\infty} \left(\int_0^{\infty} z^{l+j} e^{-xz} e^{-z^2/2} dz \right) x^{t-k} |p^{(i)}(x)| dx \\
&\le C(\nu, l) \|h^{(r-m)}\|_\infty \int_0^{\infty} (1+x)^{-l-j-1} x^{t-k} |p^{(i)}(x)| dx
\end{aligned}$$

$$\leq C(\nu, l)\|h^{(r-m)}\|_\infty \int_0^\infty (1+x)^{t-1-l} |p^{(i)}(x)| dx$$
$$\leq C(\nu, l)\|h^{(r-m)}\|_\infty \cdot \|p^{(i)}\|_1. \tag{5.9}$$

The integral $\int_{-\infty}^0 z^l e^{-z^2/2} \int_{-\infty}^0 h^{(\nu-m)}(x+z) g_x^{(m)}(x,z) dx dz$ may be treated similarly. Combining the above bounds we obtain for $\nu \geq r - m$

$$\left|E\{W^t G(W; h; l+t-1, \nu)\}\right|$$
$$\leq C(t, l, \nu)\|h^{(r-m)}\|_\infty \left\{ \max_{k \leq m-1} |p^{(k)}(0)| + \max_{i \leq m} \|p^{(i)}\|_1 \right\}.$$

Similarly one can consider $|E\{G(W; h; l, \nu)\}|$.

The case $\nu < r - m$ may be considered in the same way, but in this case one should not integrate by parts but estimate the expression (5.7) directly in the spirit of (5.9). It will lead, in particular, instead of the term $\|h^{(r-m)}\|_\infty$ to the term $\|h^{(\nu)}\|_\infty$.

We turn to the last two terms in the definition (3.11). Integration by parts implies that

$$|E\{h^{(r)}(W)\}| \leq \|h^{(r-m)}\|_\infty \int_{-\infty}^{\infty} |p^{(m)}(x)| dx.$$

Also, obviously $|\Phi_l(h)| \leq C(l)\|h\|_\infty$, which completes the proof. □

For proving Lemma 17 we need two more lemmas which we prove in the end of this section.

Lemma 20. *For any r.v. W and sufficiently smooth h*

$$|E\{W^t G(W; f; l, n)\}| \leq C(l, n)[\max_{l+1 \leq s \leq l+n,\ k \leq n-1} |E\{W^{t+1} G(W; h; s, k)\}|$$
$$+ \max_{l \leq s \leq l+n+1,\ k \leq n-1} |E\{W^t G(W; h; s, k)\}| + |\mathbf{\Phi}(h)| E\{|W|^t/[1+|W|]^{l+n}\}]. \tag{5.10}$$

Lemma 21. *For any natural l, n*

$$|\Phi_l(f^{(n)})| \leq C(l, n) \max_{k \leq 2n+l+1} |\Phi_k(h)|, \tag{5.11}$$

provided that all integrals $\Phi_k(h^{(p)})$ are finite for $p \leq n, k \leq 2n+l+1$.

Proof of Lemma 17. Since $f'(x) - xf(x) = h(x) - \mathbf{\Phi}(h)$, for $k = 1, 2, \ldots$

$$f^{(k)}(x) = xf^{(k-1)}(x) + (k-1)f^{(k-2)}(x) + [h(x) - \mathbf{\Phi}(h)]^{(k-1)}. \tag{5.12}$$

Note also that for $h \equiv 1$

$$|G(x; 1; k, 0)| \leq C(k). \tag{5.13}$$

For (3.13), we have from (5.12), (3.10), (5.13):

$$\begin{aligned}
&|E\{f^{(r+1)}(W)\}| \\
&\quad \le |E\{Wf^{(r)}(W)\}| + r|E\{f^{(r-1)}(W)\}| + |E\{h^{(r)}(W)\}| \\
&\quad \le C(r)\Big\{ \max_{l\le r, \nu\le r} \{|E\{WG(W;h;l,\nu)\}| \\
&\qquad + |E\{G(W;h;l,\nu)\}|\} + |E\{h^{(r)}(W)\}| + \mathbf{\Phi}(h)\Big\} \\
&\quad \le C(r)\Lambda_r(h;W).
\end{aligned}$$

We turn to (3.14), starting with the last term in the definition (3.11). Since $2 \le p \le r$, by (5.11)

$$\begin{aligned}
\max_{l\le (r+1-p)^3} |\Phi_l(f^{(p)})| &\le C(r) \max_{l\le (r-1)^3} \max_{k\le 2r+l+1} |\Phi_k(h)| \\
&\le C(r) \max_{k\le r^3} |\Phi_k(h)| \le C(r)\Lambda_r(h;W).
\end{aligned}$$

For the third term in (3.11), by (3.13) which has been already proved, we have $|E\{\mathcal{D}^{r+1-p} f^{(p)}(W)\}| = |E\{f^{(r+1)}(W)\}| \le C(r)\Lambda_r(h;W)$.

Turn to the first term in (3.11). Noticing that $G(W; f^{(p)}; l+t-1, \nu)\} = G(W; f; l+t-1, p+\nu)$, and using (5.10) we have

$$\begin{aligned}
&\max_{1\le t\le (r+1-p)^3,\, l\le (r+1-p)^3,\, \nu\le r+1-p} |E\{W^t G(W; f^{(p)}; l+t-1, \nu)\}| \\
&\quad \le C(r)\Big\{ \max_{1\le t\le (r+1-p)^3,\, l\le (r+1-p)^3,\, p+\nu-1\le r} \Big[\max_{l+t\le s\le l+t-1+p+\nu,\ k\le p+\nu-1} \\
&\qquad |E\{W^{t+1}G(W;h;s,k)\}| \\
&\qquad + \max_{l+t-1\le s\le l+t+p+\nu,\ k\le p+\nu-1} |E\{W^t G(W;h;s,k)\}| \\
&\qquad + \mathbf{\Phi}(h)E\{|W|^t/[1+|W|]^{l+t-1+p+\nu}\}\Big]\Big\}.
\end{aligned}$$

Taking into account that $p \ge 2$, and setting in the two consecutive "inside" maxima $s = j+t$, and $s = j+t-1$, respectively, we get that the last quantity is not less than

$$\begin{aligned}
&C(r)\Big\{ \max_{1\le t\le (r-1)^3,\, l\le (r-1)^3,\, p+\nu-1\le r} \\
&\quad \Big[\max_{l\le j\le l+r,\ k\le r} |E\{W^{t+1}G(W;h;j+(t+1)-1,k)\}| \\
&\quad + \max_{l\le j\le l+r+2,\ k\le r} |E\{W^t G(W;h;j+t-1,k)\}| \\
&\quad + \mathbf{\Phi}(h)E\{|W|^t/[1+|W|]^{t-1+p}\}\Big]\Big\}
\end{aligned}$$

$$\le C(r)\Bigg\{\max_{t+1\le(r-1)^3+1}\Bigg(\max_{j\le(r-1)^3+r,\ k\le r}|E\{W^{t+1}G(W;h;\,j+(t+1)-1,k)\}|\Bigg)$$
$$+\max_{t\le(r-1)^3}\Bigg(\max_{j\le(r-1)^3+r+2,\ k\le r}|E\{W^tG(W;h;t+j-1,k)\}|+\boldsymbol{\Phi}(h)\Bigg)\Bigg\}$$
$$\le C(r)\left\{\max_{t\le r^3}\max_{0\le j\le r^3,\ k\le r}|E\{W^tG(W;h;t+j-1,k)\}|+\boldsymbol{\Phi}(h)\right\}$$
$$\le C(r)\Lambda_r(h;W).$$

Similarly, one can get

$$\max_{l\le(r+1-p)^3,\,\nu\le r+1-p}|E\{G(W;f^{(p)},l,\nu)\}|\le C(r)\Lambda_r(h;W). \qquad \square$$

Proof of Lemma 20. It is straightforward to derive from (5.12) that

$$\begin{aligned}&G(x;f;l,n)\\&\quad=xG(x;f;l,n-1)+G(x;f;l+1,n-1)+(n-1)G(x;f;l,n-2)\\&\quad+G(x;h-\boldsymbol{\Phi}(h);l,n-1).\end{aligned}\tag{5.14}$$

Furthermore, it is easy to verify that for $x\ge0$, and a function $q(x)$

$$\begin{aligned}&\int_0^\infty\int_0^\infty t^kz^le^{-(t+z)^2/2}e^{-(t+z)x}q(x+t+z)dtdz\\&\qquad=C_1(l,k)\int_0^\infty u^{l+k+1}e^{-u^2/2}e^{-ux}q(x+u)du,\end{aligned}\tag{5.15}$$

where $C_1(l,k)=l!k!/(l+k+1)!$ Similarly for $x<0$

$$\begin{aligned}&\int_{-\infty}^0\int_{-\infty}^0 t^kz^le^{-(t+z)^2/2}e^{-(t+z)x}q(x+t+z)dtdz\\&\qquad=-C_1(l,k)\int_{-\infty}^0 u^{l+k+1}e^{-u^2/2}e^{-ux}q(x+u)du.\end{aligned}\tag{5.16}$$

Making use of (3.10) and (5.15), it is straightforward to derive that for $x\ge0$

$$\begin{aligned}&\int_0^\infty z^le^{-xz-z^2/2}f^{(n)}(x+z)dz\\&\quad=\int_0^\infty z^le^{-xz-z^2/2}\Bigg(-\sum_{k=0}^n(-1)^{n-k}\binom{n}{k}\\&\qquad\times\int_0^\infty t^{n-k}e^{-(x+z)t-t^2/2}\mathcal{D}^k\{h(x+z+t)-\boldsymbol{\Phi}(h)\}dt\Bigg)dz\\&\quad=-\sum_{k=0}^n(-1)^{n-k}\binom{n}{k}\int_0^\infty\int_0^\infty t^{n-k}z^le^{-x(t+z)}e^{-(t+z)^2/2}\\&\qquad\times\mathcal{D}^k\{h(x+t+z)-\boldsymbol{\Phi}(h)\}dtdz \qquad (5.17)\\&\quad=-\sum_{k=0}^nC_2(l,n,k)\int_0^\infty u^{1+l+n-k}e^{-u^2/2}e^{-ux}\\&\qquad\times\mathcal{D}^k\{h(x+u)-\boldsymbol{\Phi}(h)\}du, \qquad (5.18)\end{aligned}$$

where $C_2(l,n,k)=(-1)^{n-k}\binom{n}{k}\frac{l!(n-k)!}{(l+n-k+1)!}$.

Following (5.16) we get similarly "the same" for $x \le 0$:

$$\int_{-\infty}^{0} z^l e^{-xz-z^2/2} f^{(n)}(x+z)dz$$
$$= -\sum_{k=0}^{n} C_2(l,n,k) \int_{-\infty}^{0} u^{1+l+n-k} e^{-u^2/2} e^{-ux} \mathcal{D}^k \{h(x+u) - \mathbf{\Phi}(h)\} du. \tag{5.19}$$

Thus,

$$G(x; f; l, n) = -\sum_{k=0}^{n} C_2(l,n,k) G(x; h - \mathbf{\Phi}(h); 1+l+n-k, k). \tag{5.20}$$

Below and above, as usual, $C(k,l,n)$ or $C(l,n)$ denote some coefficients depending only on k, l, n, and perhaps different in different formulas.

From (5.20) and (5.14), we get that

$$\begin{aligned} G(x; f; l, n) &= xG(x; f; l, n-1) + G(x; f; l+1, n-1) \\ &\quad + (n-1)G(x; f; l, n-2) + G(x; h - \mathbf{\Phi}(h); l, n-1) \\ &= \sum_{k=0}^{n-1} C(k,l,n)[xG(x; h - \mathbf{\Phi}(h); l+n-k, k) \\ &\quad + G(x; h - \mathbf{\Phi}(h); 1+l+n-k, k)] \\ &\quad + \sum_{k=0}^{n-2} C(k,l,n) G(x; h - \mathbf{\Phi}(h); l+n-1-k, k) \\ &\quad + G(x; h - \mathbf{\Phi}(h); l, n-1). \end{aligned}$$

Note that in the expression appearing above, $G(x; h-\mathbf{\Phi}(h); l, k) = G(x; h; l, k)$ unless $k = 0$. Also

$$\begin{aligned} |G(x; 1, l, 0)| &= \left| \left(I(x \ge 0) \int_0^\infty - I(x < 0) \int_{-\infty}^0 \right) z^l e^{-xz-z^2/2} dz \right| \\ &\le C(l)(1+|x|)^{-l-1}. \end{aligned}$$

Thus,

$$\begin{aligned} G(x; f; l, n) = \sum_{k=0}^{n-1} & C(k,l,n)[xG(x; h; l+n-k, k) + G(x; h; 1+l+n-k, k)] \\ &+ \sum_{k=0}^{n-2} C(k,l,n) G(x; h; l+n-1-k, k) \\ &+ G(x; h; l, n-1) + C(l,n)\mathbf{\Phi}(h) O([1+|x|]^{-l-n}), \end{aligned}$$

where the last term above arises from terms of the form $G(x; h - \mathbf{\Phi}(h); l, k)$, with $k = 0$. A straightforward derivation of (5.10) from the above completes the proof. □

Proof of Lemma 21. Combining (5.18) and (5.19) for $x = 0$, we have

$$
\begin{aligned}
|\Phi_l(f^{(n)})| &= |\int_{-\infty}^{\infty} z^l f^{(n)}(z)\varphi(z)dz| \\
&\le C(l,n)\sum_{k=0}^{n} |\int_{-\infty}^{\infty} z^{1+l+n-k}\mathcal{D}^k\{h(z)-\mathbf{\Phi}(h)\}\varphi(z)dz| \\
&\le C(l,n)\left\{|\mathbf{\Phi}(h)| + \max_{k\le n,\, s\le l+n+1} |\int_{-\infty}^{\infty} z^s h^{(k)}(z)\varphi(z)dz|\right\} \\
&\le C(l,n)\left\{|\mathbf{\Phi}(h)| + \max_{k\le n,\, s\le l+n+1} |\int_{-\infty}^{\infty} h(z)\mathcal{D}^k[z^s\varphi(z)]dz|\right\} \\
&\le C(l,n)\left\{|\mathbf{\Phi}(h)| + \max_{k\le n,\, s\le l+n+1}\max_{j\le s+k} |\int_{-\infty}^{\infty} z^j h(z)\varphi(z)dz|\right\} \\
&\le C(l,n)\max_{j\le 2n+l+1} |\Phi_j(h)|.
\end{aligned}
$$

(It is not difficult to realize that if all integrals involved are finite, in integration by parts above all limits at $\pm\infty$ vanish.)

Acknowledgements. We are grateful to the anonymous referees for their very useful remarks, and to Richard Bradley for a very useful discussion on different representations for characteristics (1.7)–(1.8), and (1.18)–(1.19).

References

[1] Barbour, A.D.: Asymptotic expansions based on smooth functions in the central limit theorem. Probab. Th. Rel. Fields **72**, 289–303 (1986)
[2] Barbour, A.D., Karoński, M., Ruciński, A.: A central limit theorem for decomposable random variables with applications to random graphs. J. Combin. Theory Ser. B **47**, 125–145 (1989)
[3] Bening, V.E. : On the rate of convergence and asymptotic expansions for U-statistics under alternatives. Proceedings of the 19th Seminar on Stability Problems for Stochastic Models, Part II (Vologda, 1998). J. Math. Sci. (New York) **99**(4), 1403–1407 (2000)
[4] Bentkus, V., Götze, F., van Zwet, W.R.: An Edgeworth expansion for symmetric statistics. Ann. Statist. **25**, 851–896 (1997)
[5] Bhattacharya, R.N., Ranga Rao, R.: Normal approximation and asymptotic expansion. Krieger, Melbourne, Fla, 1986
[6] Bickel, P.J., Götze, F., van Zwet, W.R.: The Edgeworth expansion for U-statistics of degree two. Ann. Statist. **14**(4), 1463–1484 (1986)
[7] Bloznelis, M., Götze, F.: One-term Edgeworth expansion for finite population U-statistics of degree two. Limit theorems of probability theory (Vilnius, 1999), Acta Appl. Math. **58**(1–3), 75–90 (1999)
[8] Bradley, R.C.: Basic properties of strong mixing conditions. Dependence in probability and statistics. Progr. Probab. Statist., **11**, (Oberwolfach, 1985), 165–192, Birkhäuser, Boston, Mass, 1985
[9] Bradley, R.C.: Introduction to strong mixing conditions. Volume 1. Technical Report, Department of Mathematics, Custom Publishing of IU, Indiana University, Bloomington, 2002

[10] Bruckner, A.M., Bruckner, J.B., Thomson, B.S.: Real Analysis. Prentice Hall, NJ, 1997

[11] Doukhan, P.: Mixing: properties and examples. Lecture notes in Statistics **85**, Springer-Verlag, 1994

[12] Feller, W.: An Introduction to Probability Theory and Its Applications II. Wiley, New York, 1971

[13] Götze, F., Hipp, C.: Asymptotic expansions for sums of weakly dependent random vectors. Z.Wahrsch. Verw, Gebeite **64**, 211–239 (1983)

[14] Lothar, H.: Some remarks on asymptotic expansions in the central limit theorem for m-dependent random variables. Math. Nachr. **122**, 151–155 (1985)

[15] Lothar, H.: Asymptotic expansions in the central limit theorem for a special class of m-dependent random fields. I. Math. Nachr. **134**, 83–106 (1987)

[16] Lothar, H.: Asymptotic expansions in the central limit theorem for a special class of m-dependent random fields. II. Lattice case. Math. Nachr. **145**, 309–327 (1990)

[17] Henze, N.: On the fraction of random points with specified nearest-neighbor interrelations and degree of attraction. Adv. in Appl. Probab. **19**(4), 873–895 (1987)

[18] Henze, N.: A multivariate two-sample test based on the number of nearest neighbor type coincidences. Ann Statist. **16**, 772–783 (1988)

[19] Hipp C.: Asymptotic expansions in the central limit theorem for compound and Markov processes. Z. Wahrsch. Verw. Gebeite **69**, 361–385 (1985)

[20] Ibragimov, I.A.: Some limit theorems for stationary processes. Theory Prob. Appl. **7**, 349–382 (1962)

[21] Iosifescu, M.M., Theodorescu, R.: Random Processes and Learning. Springer-Verlag, 1969

[22] Jensen, J.L.: Asymptotic expansions for strongly mixing Harris recurrent Markov chains. J. Scand. Statist. **16**, 47–64 (1989)

[23] Jensen, J.L.: A note on asymptotic expansions for sums over a weakly dependent random field with application to the Poisson and Strauss processes. Ann. Inst. Statist. Math. **45**(2), 353–360 (1993)

[24] Korolyuk, V.S., Borovskikh, Yu.V.: Expansions for U-statistics and von Mises functionals. Ukrain. Mat. Zh. **37**(4), 450–456, 543 (1985)

[25] Lahiri, S.N.: Refinements in asymptotic expansions for sums of weakly dependent random vectors. Ann. Probability **21**(2), 791–799 (1993)

[26] Leech, J., Sloane, N.J.A.: Sphere packings and error-correcting codes. Canad. J. Math **23**, 718–745 (1971)

[27] Loh, W.-L.: An Edgeworth expansion for U-statistics with weakly dependent observations, Statist. Sinica **6**(1), 171–186 (1996)

[28] Yoshihiko, M.: Asymptotic mean square errors of variance estimators for U-statistics and their Edgeworth expansions. J. Japan Statist. Soc. **28**(1), 1–19 (1998)

[29] Majumdar, M., Rotar, V.: Equilibrium prices in a random exchange economy with dependent summands. Economic Theory **15**, 531–550 (2000)

[30] Majumdar, M., Rotar, V.: Some General Results on Equilibrium Prices in Large Random Exchange Economies. Annals of Operation Research **114**, 245–261 (2002)

[31] Malinovskii, V.K.: Limit theorems for Harris-Markov chains. I. Theory Probab. Appl. **31**, 269–288 (1987)

[32] Marshall, A.W., Olkin, I.: Inequalities: theory of majorization and its applications. Academic Press, Inc, 1979

[33] Mykland, P.A.: Asymptotic expansions for martingales. Ann.Probability **21**(2), 800–818 (1993)

[34] Peligrad, M.: Recent advances in the central limit theorem and its weak invariance principle for mixing sequences of random variables (a survey). Dependence in probability

and statistics. Progr. Probab. Statist., **11**, (Oberwolfach, 1985), 193–223, Birkhäuser Boston, Boston, Mass, 1986
[35] Petrov, V.V.: Limit Theorems of Probability Theory. Sequences of independent random variables. Clarendon Press, Oxford, 1995
[36] Philipp, W.: The central limit problem for mixing sequences of random variables. Z. Wahrsch. verw. Gebiete **12**, 155–171 (1969)
[37] Rhee, W.T.: An Edgeworth expansion for a sum of m-dependent random variables. Internat. J. Math. Math. Sci. **8**(3), 563–569 (1985)
[38] Rinott, Y.: On normal approximation rates for certain sums of dependent random variables. J. Comput. Appl. Math. **55**, 135–143 (1994)
[39] Rinott, Y., Rotar, V.I.: A multivariate CLT for local dependence with $n^{-1/2}\log n$ rate, and applications to multivariate graph related statistics. J. Multivariate analysis **56**, 333–350 (1996)
[40] Rotar, V.I.: Probability Theory. World Scientific Publishers, NY, 1997
[41] Statulevichius, V.: Limit theorems for sums of random variables that are connected in a Markov chain. I,II,III, Litovsk. Mat. Sb., **9**, 345–362, 635–672; **10**, 161–169 (1969–1970)
[42] Stein, C.: A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. Proc. Sixth Berkeley Symp. Math. Statist. Probab. **2**, 583–602, Univ. California Press, Berkeley, 1972
[43] Stein, C.: Approximate Computation of Expectations. IMS, Hayward, Calif., 1986. 2