PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 27, Fasc. 2 (2007), pp. 153–166

FINITE DIFFERENCE EQUATIONS AND CONVERGENCE RATES IN THE CENTRAL LIMIT THEOREM

BY

LARS LINDHAGEN* (UPPSALA)

Abstract. We apply the theory of finite difference equations to the central limit theorem, using interpolation of Banach spaces and Fourier multipliers. Let S_n^* be a normalized sum of i.i.d. random vectors, converging weakly to a standard normal vector $\mathcal{N}$. When does $\|Eg(x+S_n^*)-Eg(x+\mathcal{N})\|_{L_p(dx)}$ tend to zero at a specified rate? We show that, under moment conditions, membership of g in various Besov spaces is often sufficient and sometimes necessary. The results extend to signed probability.

2000 AMS Mathematics Subject Classification: Primary: 46B70, 60F05, 65M06; Secondary: 35K05, 42B15, 65M15.

Key words and phrases: Finite difference equations, central limit theorem, convergence rate, interpolation theory, Fourier multipliers, Besov spaces, signed probability.

1. INTRODUCTION

Let $X_1, X_2, \ldots$ be i.i.d. random vectors with zero mean and identity covariance matrix, and put $S_n^* = n^{-1/2}\sum_{j=1}^n X_j$, converging weakly to a standard normal vector $\mathcal{N}$. The purpose of this paper is to examine the convergence rate of functionals of the type

$$\|Eg(x+S_n^*)-Eg(x+\mathcal{N})\|_{L_p(dx)}, \qquad 1 \leqslant p \leqslant \infty.$$

In particular, we are interested in conditions on the function g guaranteeing that the above tends to zero at a specified rate. It turns out that membership in suitable Besov spaces is often sufficient and sometimes necessary.

This brings several old results under the same roof. If $p = \infty$, we get translation invariant bounds on the convergence of $Eg(S_n^*)$ to $Eg(\mathcal{N})$. With g equal to the one-dimensional Heaviside function, the results reduce to information on the L_p convergence of distribution functions. As is shown in Sec-

* This work was done while funded by Uppsala University.

tion 5.1, our general theorems recover, and in some cases improve, large portions of the older results from the literature.

The paper is based on convergence theorems for finite difference equations approximating the heat equation. These are extended, using interpolation theory and Fourier multipliers as main tools. We thus exploit the equivalence between a time discretization of a PDE, and an approximation of the normal distribution by a sum of random variables. The first one to use such connections for probabilistic applications seems to be Zukov [29]. We also mention the similarities to the operator method invented by Trotter [26]. Although well-known, such connections seem to have been subsequently underplayed.

The paper is organized as follows: After some preliminaries, we review a few convergence results for finite difference equations, and apply these to the central limit theorem in Sections 3 and 4. A final section contains comparisons with related results, some converse results, and an extension to signed probability. The main result is Theorem 4.9.

The author wishes to thank his former advisor Svante Janson, Uppsala University, for valuable suggestions. Thanks also to Henrik Brandén and Göran Hamrin for many rewarding discussions.

2. PRELIMINARIES

2.1. Notation. We shall work on $\boldsymbol{R}^d$ with the standard scalar product $x \cdot y$, norm $|x|$, convolution $f * g$, and Fourier transform $\mathscr{F}[f](\xi) = \hat{f}(\xi) = \int f(x) e^{-i\xi \cdot x} dx$. For $1 \leqslant p \leqslant \infty$, $L_p = L_p(\boldsymbol{R}^d, dx)$ is the usual Lebesgue space. Moreover, $\mathscr{S}$ is the Schwartz space of rapidly decreasing functions with dual $\mathscr{S}'$ (see [11]).

Fourier multipliers from L_q to L_p are defined as follows: $\varphi \in M_q^p$ if the mapping $g \mapsto \mathscr{F}^{-1}[\varphi \hat{g}]$ is bounded from L_q to L_p. If $U \subset \boldsymbol{R}^d$ is open, the space $M_q^p(U)$ of local Fourier multipliers essentially arises by disregarding ξ outside U, cf. [14].

For multiindices α we use the standard notation $x^\alpha = \prod x_i^{\alpha_i}$, $|\alpha| = \sum \alpha_i$ and $\alpha! = \prod \alpha_i!$, whereas $\partial^\alpha f$ denotes the corresponding partial derivative. Finally, c and C are small and large positive constants, not necessarily the same on each occurrence.

2.2. Two probability lemmas. We recall a few facts from probability theory. The first lemma is a simple exercise in conditioning.

LEMMA 2.1. *Let $X_1, X_2, \ldots$ be i.i.d. random vectors and put $S_n = \sum_{j=1}^n X_j$. Define $Tf(x) = Ef(x+X_1)$. Then the iterates of T are $T^n f(x) = Ef(x+S_n)$.*

Let $X = (X^{(1)}, \ldots, X^{(d)})$ be a random vector with characteristic function $\phi(\xi) = Ee^{i\xi \cdot X}$. Then next lemma follows e.g. from [28], Remark 3.2.

LEMMA 2.2. *Let X be non-singular. Then, for any $\delta > 0$,*

$$\sup_I \int_I |\phi(\xi)|^n \, d\xi = O(n^{-d/2}),$$

where the supremum is taken over all cubes $I \subset \mathbf{R}^d$ with edge length δ.

Finally, the cumulants $\kappa_\alpha = \kappa_\alpha(X)$ are defined by a formal expansion of $\log\phi$. Thus, if $E|X|^r < \infty$ for some $r > 0$, then

$$\log\phi(\xi) = \sum_{|\alpha| \leqslant r} \frac{\kappa_\alpha}{\alpha!}(i\xi)^\alpha + O(|\xi|^r) \tag{2.1}$$

as $\xi \to 0$, with O replaced by o if r is an integer. In particular, $\kappa_i = EX^{(i)}$ and $\kappa_{ij} = \mathrm{Cov}(X^{(i)}, X^{(j)})$.

2.3. Besov spaces. We give the basic definitions and properties of the Besov space $B_p^{s,q}$, $1 \leqslant p, q \leqslant \infty$. For proofs and more material, see [2] and [15].

Let $\varphi \in \mathscr{S}(\mathbf{R}^d)$ be such that $\hat{\varphi} \geqslant 0$ has support in $\{\frac{1}{2} \leqslant |\xi| \leqslant 2\}$, and $\sum_{j\in \mathbf{Z}} \hat{\varphi}_j(\xi) = 1$ for $\xi \neq 0$, where $\hat{\varphi}_j(\xi) = \hat{\varphi}(2^{-j}\xi)$. Moreover, put $\hat{\Phi} = 1 - \sum_{j=1}^{\infty} \hat{\varphi}_j$. Then $B_p^{s,q}$ consists of those $f \in \mathscr{S}'$ for which

$$\|f\|_{B_p^{s,q}} := \|\Phi * f\|_p + \Big\{\sum_{j=1}^{\infty} (2^{js}\|\varphi_j * f\|_p)^q\Big\}^{1/q} < \infty, \tag{2.2}$$

with the usual modification if $q = \infty$. The homogeneous Besov space $\dot{B}_p^{s,q}$ is defined analogously, by dropping the term $\|\Phi * f\|_p$ and summing over $j \in \mathbf{Z}$.

For $s > 0$ there is a more transparent definition that gives equivalent norms. Let $s = S + \sigma > 0$, S integral, $0 < \sigma \leqslant 1$. Then $f \in B_p^{s,q}$ iff

$$\|f\|_{B_p^{s,q}} := \|f\|_p + \sum_{|\alpha| = S} \left\{\int_0^\infty \left(\frac{\omega_p^k(t, \partial^\alpha f)}{t^\sigma}\right)^q \frac{dt}{t}\right\}^{1/q} < \infty, \tag{2.3}$$

where $k = 1$ if $\sigma < 1$ and $k = 2$ if $\sigma = 1$, and ω_p^k are the following smoothness moduli in L_p:

$$\omega_p^1(t, f) = \sup_{|h| \leqslant t} \|f(x+h) - f(x)\|_{L_p(dx)}$$

and

$$\omega_p^2(t, f) = \sup_{|h| \leqslant t} \|f(x+h) - 2f(x) + f(x-h)\|_{L_p(dx)}.$$

Homogeneous Besov spaces may be defined similarly, by dropping the term $\|f\|_p$ in (2.3). Then $B_p^{s,q} = \dot{B}_p^{s,q} \cap L_p$ and $\dot{B}_p^{s,q_1} \subset \dot{B}_p^{s,q_2}$ if $q_1 \leqslant q_2$.

We remark that functions in homogeneous spaces are only defined modulo polynomials, whence all formulae have to be interpreted with some care. For example, the smoothness bound presupposes that the representative f has been chosen properly modulo high order polynomials, cf. [15].

2.4. Interpolation theory. Let X_0 and X_1 be two Banach spaces embedded in some common larger topological vector space. Given $0<\theta<1$ and $1 \leqslant q \leqslant \infty$, there are several ways to define an interpolation space X, intermediate to X_0 and X_1, see [2] for definitions and constructions. We shall use the following instances of the real method $X=(X_0, X_1)_{\theta,q}$ and the complex method $X=[X_0, X_1]_\theta$:

$$(L_p, \dot{B}_p^{s,r})_{\theta,q} = \dot{B}_p^{\theta s,q}, \tag{2.4}$$

$$[L_{p_0}, L_{p_1}]_\theta = L_p, \tag{2.5}$$

where $1/p=(1-\theta)/p_0+\theta/p_1$ and $s>0$.

The main point is that one can interpolate operators as well:

THEOREM 2.3. *Suppose that T is a bounded linear operator from X_0 to Y_0 and from X_1 to Y_1. Let $X=(X_0, X_1)_{\theta,q}$, and similarly for Y. Then T is bounded from X to Y with norm*

$$\|T\|_{X,Y} \leqslant \|T\|_{X_0,Y_0}^{1-\theta} \|T\|_{X_1,Y_1}^{\theta}.$$

The same holds for complex interpolation: $X=[X_0, X_1]_\theta$ etc.

3. FINITE DIFFERENCE EQUATIONS

In this section we collect some results on finite difference approximations of certain partial differential equations. As general references, we mention [18] and [24]. We shall be only concerned with homogeneous parabolic equations with constant coefficients, although some of the results are valid in greater generality [16], [27]. Put $D=-i\partial$ and consider the initial value problem

$$\begin{aligned} &\partial u/\partial t + P(D)u = 0, \qquad && t>0,\ x \in \boldsymbol{R}^d, \\ &u = g, && t=0, \end{aligned} \tag{3.1}$$

where P is a homogeneous polynomial of degree m with constant coefficients. We assume that $P(\xi)>0$ for $\xi \neq 0$, implying that $P(D)$ is an elliptic differential operator. Then $-P(D)$ is the infinitesimal generator of a strongly continuous (in L_p) semigroup of solution operators $\{E(t);\ t \geqslant 0\}$. In other words, $u(t,\cdot)=E(t)g$, where $E(t)g=\mathscr{F}^{-1}[e(t,\xi)\hat{g}(\xi)]$ and $e(t,\xi)=e^{-tP(\xi)}$ (see [16]).

A classical finite difference approximation of (3.1) has the form

$$v(t+k, x) = \sum_{\tau \in \boldsymbol{Z}^d} b_\tau v(t, x+h\tau)$$

for some constants b_τ. We shall be a little more general, considering the scheme

$$\begin{aligned} &v(t+k, x) = \int_{\boldsymbol{R}^d} v(t, x+hy)\, d\nu(y), \qquad && t \in T_h,\ x \in \boldsymbol{R}^d, \\ &v = g, && t=0, \end{aligned} \tag{3.2}$$

where ν is a signed measure on $\boldsymbol{R}^d$ and $T_h = \{0, k, 2k, \ldots\}$. Let $k = h^m$, where m is the order of the PDE, and consider the solution operator of (3.2): $E_h(k)g(x) = \int g(x+hy)\,d\nu(y)$. Putting

$$e_h(k, \xi) = \int e^{iy\cdot h\xi}\,d\nu(y) \quad \text{and} \quad e_h(Nk, \xi) = e_h(k, \xi)^N,$$

it is clear that $E_h(t)g = \mathscr{F}^{-1}[e_h(t, \xi)\hat{g}(\xi)]$ for any $t \in T_h$. We assume that E_h is *stable* on L_p, i.e. that

$$\|E_h(t)g\|_p \leqslant C\|g\|_p \tag{3.3}$$

for some $C = C_p$ and all $t \in T_h$. Although this is in general a quite involved condition (see e.g. [18]), it will mostly hold trivially in our applications; see however Section 5.3 below.

We now say that $E_h(t)$ *approximates* $E(t)$ *in the strong sense with accuracy order* $\mu > 0$ if

$$e_h(k, \xi) = e(k, \xi) + |h\xi|^{m+\mu}Q(h\xi),$$

where $\xi^\alpha \partial^\alpha Q(\xi)$ is bounded on a punctured neighbourhood of the origin for $|\alpha| \leqslant L$, L being the smallest integer greater than $d/2$. The latter condition arises from the use of Michlin's multiplier theorem [10], and is precisely what is needed for the proof of Theorem 3.1 below. The less restrictive condition that Q is bounded close to the origin will be called *accuracy in the standard sense*. If, in addition, Q is bounded away from zero, we say that the *approximation order is exactly* μ.

The following result is due to Löfström [14]. It was formulated for lattice measures ν and with stronger regularity assumptions on Q, but holds just as well in the present situation; cf. [14], Remark 7.5.

THEOREM 3.1. *Suppose that* $E_h(t)$ *approximates* $E(t)$ *in the strong sense with order exactly* μ. *Let* $0 < s \leqslant \mu$ *and* $g \in L_p$. *Then the following are equivalent*:

(i) $\|E_h(t)g - E(t)g\|_p = O(h^s)$ *uniformly in* $t \in T_h$ *as* $h \to 0$.

(ii) $g \in B_p^{s,\infty}$.

If so, the bound in (i) *holds uniformly for* g *in the unit ball of* $B_p^{s,\infty}$. *Moreover, the approximation is saturated*: *If* $\|E_h(t)g - E(t)g\|_p = o(h^\mu)$ *uniformly in* t, *then* $g = 0$.

Remark 3.2. The assumptions of exact order and $g \in L_p$ are not needed for (ii) $\Rightarrow$ (i). Thus, (i) remains valid under the weaker assumption $g \in \dot{B}_p^{s,\infty}$. We also remark that the proof of the converse implication uses the bound for small t (say $t \leqslant 1$) only [14].

We close the section with an observation on more general bounds of the type

$$\|E_h(t)g - E(t)g\|_p \leqslant Ct^\gamma h^s \|g\|_{\dot{B}_q^{s,r}}.$$

Namely, the dependence on t is trivial in such cases, whence one needs only consider the case $t = 1$.

PROPOSITION 3.3. *Let $1 \leqslant p, q, r \leqslant \infty$ and put $\rho = 1/q - 1/p$. Then the following are equivalent:*

(i) $\|E_h(1)g - E(1)g\|_p \leqslant Ch^s \|g\|_{\dot{B}_q^{s,r}}$.

(ii) $\|E_h(t)g - E(t)g\|_p \leqslant Ct^{-\rho d/m} h^s \|g\|_{\dot{B}_q^{s,r}}$ *for all* $h > 0$ *and* $t \in T_h$.

Proof. Assume (i) and take $t = Nk$. Let $1 = N\tilde{k} = N\tilde{h}^m$, and put $\tilde{g}(x) = g(t^{1/m}x)$. The result follows from the identities

$$E_h(t)g(x) = (E_{\tilde{h}}(1)\tilde{g})(t^{-1/m}x) \quad \text{and} \quad \|g(ax)\|_{\dot{B}_q^{s,r}} = |a|^{s-d/q} \|g(x)\|_{\dot{B}_q^{s,r}};$$

we omit the details. ∎

4. APPLICATIONS TO THE CENTRAL LIMIT THEOREM

We now restrict ourselves to the situation relevant for probabilistic applications, see however Section 5.3 below. To this end, consider the heat equation $\partial u/\partial t = \frac{1}{2}\Delta u$, where Δ is the Laplacian in $\mathbf{R}^d$. Thus, in the notation of Section 3, $m = 2$, $P(\xi) = \frac{1}{2}|\xi|^2$, and $k = h^2$. Suppose, in addition, that ν is a probability measure. The stability condition (3.3) then holds with $C = 1$ by Jensen's inequality. Let $X, X_1, X_2, \ldots$ be i.i.d. random vectors with distribution ν, and put $S_n = \sum_{j=1}^n X_j$. Then $e_h(k, \xi) = \phi(h\xi)$, where ϕ is the characteristic function of X. If this difference scheme is to have positive accuracy, it is clear that X must have zero mean and identity covariance matrix (cf. (2.1)), so that $S_n^* := S_n/\sqrt{n}$ converges in distribution to a standard normal vector $\mathscr{N}$.

Now, let $n = k^{-1}$ be an integer, put $Z_t^{(n)} = S_{nt}/\sqrt{n}$ for $t \in T_h$, and extend this to all $t \geqslant 0$ by linear interpolation. Then $Z_t^{(n)}$ converges weakly to d-dimensional standard Brownian motion W as $n \to \infty$ (see [3]). Since $E_h(k)g(x) = Eg(x + hX)$, Lemma 2.1 and the expression for $e(t, \xi)$ above give

$$(4.1) \qquad \begin{aligned} E_h(t)g(x) &= Eg(x + Z_t^{(n)}), \qquad t \in T_h, \\ E(t)g(x) &= Eg(x + W_t), \qquad t \geqslant 0. \end{aligned}$$

For $t = 1$ we have $Z_1^{(n)} = S_n^*$ and $W_1 \stackrel{d}{=} \mathscr{N}$.

Let us call X (or ν) *approximately normal of the order* μ in any of the two senses discussed above if the difference scheme (3.2) has the same accuracy. This order of normality is nearly equivalent to the following moment conditions. Recall that L is the smallest integer exceeding $d/2$.

PROPOSITION 4.1. *Let $\mu > 0$. Suppose that X has zero mean, identity covariance matrix, that $E|X|^{2+\mu} < \infty$, and that all cumulants $\kappa_\alpha = \kappa_\alpha(X)$ of order $3 \leqslant |\alpha| < 2 + \mu$ vanish. Then X is approximately normal of the order μ in the standard sense. The extra condition $E|X|^L < \infty$ guarantees normality in the strong sense.*

If μ is an integer, the conditions above are satisfied, and, in addition,

$$\sum_{|\alpha|=2+\mu} \frac{\kappa_\alpha}{\alpha!} \xi^\alpha \neq 0 \quad \text{for } \xi \neq 0,$$

then the order of normality is exactly μ.

Proof. This is a standard approximation result for characteristic functions. The final statement follows from (2.1). ∎

Remark 4.2. In one dimension the last statement reduces to $\kappa_{2+\mu} \neq 0$. Moreover, we need not distinguish between the strong and the standard sense if the dimension $d \leqslant 3$.

Remark 4.3. The first moment condition of Proposition 4.1 can be slightly relaxed. If μ is non-integral, we can do with the weak condition $P(|X| > T) \leqslant CT^{-(2+\mu)}$. This can be seen in dimension one by modifying the proof of Boas [4], from which the general case follows by considering different directions separately and appealing to uniformity (the implied constant depends on C and μ only).

If μ is an integer and the dimension $d = 1$, an argument similar to that of Pitman [17] shows that this weak condition still suffices if combined with the assumption $E[X^{2+\mu} I\{|X| \leqslant T\}] = O(1)$. This is of course an improvement only if μ is odd.

The measurement of smoothness in L_p for convergence in L_p in Theorem 3.1 is natural, but one can also pass from smoothness in L_q to convergence in L_p if $q \leqslant p$. In a more restrictive setting, this was proved for $q = 1$ and $p = \infty$ by Thomée and Wahlbin [25]. Recall the notion of Fourier multipliers in Section 2.1. We let $U_j = \{2^{j-1} < |\xi| < 2^{j+1}\}$ and put $\|\varphi\|_j = \|\varphi\|_{M_q^p(U_j)}$ when there is no risk of confusion. The following result is implied by the inequalities of Hausdorff–Young and Hölder.

LEMMA 4.4. *Let $q \leqslant 2 \leqslant p$ and put $1/r = 1/q - 1/p$. Then $L_r \subset M_q^p$. More precisely, $\|\varphi\|_{M_q^p} \leqslant C_d \|\varphi\|_{L_r}$, where C_d is a constant depending on the dimension d only.*

In the sequel, we shall use the following notation: ε is a fixed, sufficiently small number and $\rho = 1/q - 1/p$. In line with Proposition 3.3, we take $t = Nk = 1$, so that $N = n = h^{-2}$. Moreover,

$$R_h(\xi) = e_h(1, \xi) - e(1, \xi) = \phi(h\xi)^n - \exp(-|\xi|^2/2).$$

Our main result will be deduced from the following two lemmas, which we prove for $r < \infty$ only.

LEMMA 4.5. *Let X be approximately normal in the strong sense of order μ, and let $q \leqslant p$. Then $\sum_{h2^j \leqslant \varepsilon} 2^{-js} \|R_h\|_j \leqslant Ch^s$ for $0 < s \leqslant \mu$.*

Proof. For $p = q$ this was proved by Löfström [14], Theorem 7.1. Suppose that $q \leqslant 2 \leqslant p$ and let $r = \rho^{-1}$ as in Lemma 4.4. Since $|\phi(\xi)| \leqslant \exp(-c|\xi|^2)$ for $|\xi| \leqslant 2\varepsilon$, say, and $|\phi(\xi) - \exp(-|\xi|^2/2)| \leqslant C|\xi|^{2+s}$, we have, for $\xi \in U_j$,

$$|R_h(\xi)| = \left|(\phi(h\xi) - \exp(-|h\xi|^2/2)) \sum_{l=0}^{n-1} \phi(h\xi)^{n-l-1} \exp(-l|h\xi|^2/2)\right|$$
$$\leqslant C|h\xi|^{2+s} \sum_{l=0}^{n-1} \exp(-c(n-l-1)|h\xi|^2) \exp(-l|h\xi|^2/2)$$
$$\leqslant Ch^s |\xi|^{2+s} \exp(-c|\xi|^2).$$

Hence

$$\|R_h\|_j \leqslant C\|R_h\|_{L_r(U_j)} \leqslant Ch^s \Big\{\int_{U_j} |\xi|^{r(2+s)} \exp(-c|\xi|^2)\, d\xi\Big\}^{1/r}$$
$$\leqslant Ch^s \min\{1, 2^{j(\rho d + 2 + s)}\},$$

and we need only sum over j. The general case follows by interpolating between the cases $p = q$ and $q \leqslant 2 \leqslant p$, using Theorem 2.3 and (2.5). ∎

LEMMA 4.6. *Assume that X has a non-singular covariance matrix and let $q \leqslant p$. Then $\|R_h\|_j \leqslant Ch^{\rho d} 2^{j\rho\alpha}$ for $h2^j > \varepsilon$.*

Proof. If $p = q$, then $\|R_h\|_j \leqslant 2 = C$ by stability. For $q \leqslant 2 \leqslant p$ and r as above, we have $\|\exp(-|\xi|^2/2)\|_{L_r(U_j)} \leqslant C$. Moreover, by Lemma 2.2,

$$\|\phi(h\xi)^n\|_{L_r(U_j)}^r = Ch^{-d} \int_{hU_j} |\phi(\xi)|^{rn}\, d\xi \leqslant C(h2^j)^d,$$

so that $\|R_h\|_j \leqslant C(1 + (h2^j)^{d/r}) \leqslant C(h2^j)^{\rho d}$. By interpolation, the same holds for any $q \leqslant p$. ∎

PROPOSITION 4.7. *Suppose that X is approximately normal in the strong sense of order μ. Let $q \leqslant p$ and put $\rho = 1/q - 1/p$. Then*

$$\|E_h(1)g - E(1)g\|_p \leqslant C\|g\|_{\dot{B}_q^{s,\infty}} h^s$$

for $\rho d < s \leqslant \mu$ and $g \in \dot{B}_q^{s,\infty}$. A similar bound holds if $s = \rho d$ and $g \in \dot{B}_q^{s,1}$.

Proof. We prove the case $s > \rho d$ only, the other one being similar. Let $\{\varphi_j\}$ be as in Section 2.3, and put $g_j = \varphi_j * g$. Then $\|E_h(1)g - E(1)g\|_p$ does not exceed

$$\sum_{j \in \mathbf{Z}} \|\mathscr{F}^{-1}[R_h \hat{g}_j]\|_p \leqslant \sum_{j \in \mathbf{Z}} \|R_h\|_j \|g_j\|_q \leqslant \|g\|_{\dot{B}_q^{s,\infty}} \sum_{j \in \mathbf{Z}} 2^{-js} \|R_h\|_j.$$

By Lemmas 4.5 and 4.6, the last sum above is, in turn, majorized by

$$Ch^s + Ch^{\rho d} \sum_{h2^j > \varepsilon} 2^{-j(s-\rho d)} \leqslant Ch^s. ∎$$

To remove the assumption $s \geqslant \rho d$ seems to require additional assumptions. A sufficient one is the following

DEFINITION 4.8. The random vector X is called *eventually smooth* if S_n^* is absolutely continuous with a bounded density for some n.

For example, this holds if ϕ belongs to L_p for some $p < \infty$ or, by the Hausdorff–Young inequality, if X has a bounded density in L_p for some $p > 1$. There exist singular distributions that are eventually smooth [19]. If X is eventually smooth, then S_n^* has uniformly bounded densities for large n, in fact they converge uniformly to the standard normal density [20]. By interpolation, the densities are bounded in any L_p.

If p, q, and ρ are as in Proposition 4.7 and X is eventually smooth, it thus follows from a standard convolution estimate ([11], Corollary 4.5.2) that $\|E_h(1)g\|_p \leqslant C\|g\|_q$, h small. Since the same holds for $E(1)$, we have

$$\|(E_h(1)-E(1))g\|_p \leqslant C\|g\|_{L_q}, \tag{4.2}$$

$$\|(E_h(1)-E(1))g\|_p \leqslant Ch^{\rho d}\|g\|_{\dot{B}_q^{\rho d,1}}, \tag{4.3}$$

$$\|(E_h(1)-E(1))g\|_p \leqslant Ch^{\mu}\|g\|_{\dot{B}_q^{\mu,\infty}}. \tag{4.4}$$

Interpolating between (4.2) and (4.3) or (4.4), using Theorem 2.3 and (2.4), we see that this proves our main result if combined with (4.1). The final statement follows from the proof of Lemma 4.5. For applications, the theorem should be combined with Proposition 4.1 and Remarks 4.2 and 4.3.

THEOREM 4.9. *Let* $1 \leqslant q \leqslant p \leqslant \infty$, *put* $\rho = 1/q - 1/p$, *and take* $s > 0$. *Suppose that* X *is approximately normal in the strong sense of order* $\mu \geqslant \max\{s, \rho d\}$. *If* $s \leqslant \rho d$, *then suppose that* X *is eventually smooth, and if* $s = \rho d$, *suppose, in addition, that* $\mu > \rho d$. *Then, for large* n,

$$\|Eg(x+S_n^*) - Eg(x+\mathcal{N})\|_{L_p(dx)} \leqslant C\|g\|_{\dot{B}_q^{s,\infty}} n^{-s/2}. \tag{4.5}$$

If $q \leqslant 2 \leqslant p$, *it suffices to have accuracy in the standard sense.*

Remark 4.10. By Proposition 4.7 and its proof, a few other cases can be handled without density assumptions. Thus, a bound similar to (4.5) holds for $s = \rho d \leqslant \mu$ and $g \in \dot{B}_q^{s,1}$. If $\mu \leqslant \rho d < s$ and $g \in \dot{B}_p^{s,\infty}$, we get the convergence rate $n^{-\mu/2}$.

Remark 4.11. Proposition 3.3 and (4.1) turn Theorem 4.9 into an invariance principle:

$$\|Eg(x+Z_t^{(n)}) - Eg(x+W_t)\|_{L_p(dx)} \leqslant Ct^{-\rho d/2}\|g\|_{\dot{B}_q^{s,\infty}} n^{-s/2}$$

uniformly for $t \in T_h$.

Remark 4.12. The result cannot be extended to $p < q$. This follows from Theorem 3.1, Remark 3.2 and Proposition 3.3; we omit the details. In dimension one we can also appeal to Proposition 5.3 below. As a comparison, note that $M_q^p = \{0\}$ if $p < q < \infty$ (see [10]).

Remark 4.13. If $p = \infty$, then Theorem 3.1 and Proposition 4.7 both give uniform bounds. Since this is not destroyed by interpolation, the bounds in Theorem 4.9 hold with true suprema for $p = \infty$, rather than essential ones.

Remark 4.14. The constant C in (4.5) really depends on s, since different, but equivalent, norms may occur in interpolations like (2.4). For example, the implied ordo constant in Theorem 3.1 grows (at most) like s^{-1} as $s \to 0$ if the norm (2.2) is used [14].

Assuming only finite second-order moment, a similar result, without information on the convergence rate, holds. For its statement, let C_0 be the space of continuous functions vanishing at infinity.

THEOREM 4.15. *Let* $1 \leqslant p \leqslant \infty$ *and the dimension* $d \leqslant 3$. *Suppose that* X *has zero mean and identity covariance matrix. Then*

$$\|Eg(x+Z_t^{(n)}) - Eg(x+W_t)\|_{L_p(dx)} \to 0$$

uniformly in $t \in T_h$ *as* $n \to \infty$ *provided that* $g \in L_p$ *(for* $p < \infty$*) or that* $g \in C_0$ *(for* $p = \infty$*).*

Proof. If $g \in \dot{B}_p^{0,1}$, then the conclusion is valid by an argument similar to the proof of Proposition 4.7. The result then follows by density (note that $\dot{B}_p^{0,1} \subset L_p$ by the "homogeneous" counterpart to (2.2)) or from the Lax–Richtmyer equivalence theorem [18]. ■

5. RELATED AND CONVERSE RESULTS

5.1. Comparison with related results. There are a few special cases of Theorem 4.9, that overlap existing results. Let us remark at once that our techniques give no asymptotic expansions; we always assume sufficiently many cumulants to vanish. This seems to be the usual approach when discussing L_p-convergence [8], [12], although we believe that such expansions are possible.

Firstly, if $p = \infty$, we get translation-invariant bounds on the convergence rate of $Eg(S_n^*)$ to $Eg(\mathcal{N})$. These may be compared to those of Barbour [1], treating the univariate case by Stein's method. (In the multivariate case, more complicated bounds have been given by Götze and Hipp [6].) For translation invariance, we take $p = 0$ in Barbour's main theorem ([1], p. 294). His assertion is then that if $s > 0$, $E|X|^{2+s} < \infty$, and the cumulants $\kappa_\alpha(X)$ vanish for $3 \leqslant \alpha < 2+s$, then

$$Eg(S_n^*) = Eg(\mathcal{N}) + O(n^{-s/2})$$

provided that $g \in \dot{B}_\infty^{s,\infty,\#}$. The latter is the same space as $\dot{B}_\infty^{s,\infty}$, except that a Lipschitz condition is used also for integral s, cf. Section 2.3. We recognize this as the case $d = 1$, $p = q = \infty$ of Theorem 4.9, but with a smaller function space for integral s. Thus, we can make some extensions and improvements of Barbour's result, including the assumption of only a weak moment condition, cf. Remark 4.3. Note however that [1] allows for non-identical distributions and non-translation invariant bounds, as well as asymptotic expansions.

As a second application, we consider distribution functions. This situation has been thoroughly studied, and definite results were given by Ibragimov [12]. We shall see that our general results contain large parts of these. To this end, take $d = 1$ and let g be the Heaviside function, for convenience defined as the indicator function of $(-\infty, 0]$. Then $g \in \dot{B}_p^{1/p,\infty}(\mathbf{R})$, and $Eg(x+X) = F(-x)$, where F is the distribution function of X. Taking $q = s^{-1}$, Theorem 4.9 yields

PROPOSITION 5.1. *Let X be one-dimensional with $E|X|^{2+s} < \infty$ for some $0 < s \leqslant 1$ and let $s^{-1} \leqslant p \leqslant \infty$. If $p = \infty$, then suppose, in addition, that X is eventually smooth and that $E|X|^r < \infty$ for some $r > 2+s$. Then $\|F_n - \Phi\|_p \leqslant Cn^{-s/2}$, where F_n is the distribution function of S_n^* and Φ is that of $\mathcal{N}$.*

As a comparison, Theorem 4.3 of Ibragimov [12] shows that no restrictions on p or extra assumptions are necessary. Note that Ibragimov assumes a weak moment condition, equivalent to our Remark 4.3.

Correcting a misprint, Theorem 1 of Heyde and Nakata [8] shows that if Cramér's condition (C) holds: $\limsup_{\xi \to \pm\infty} |\phi(\xi)| < 1$, and if $\kappa_3(X) = \ldots = \kappa_{k-1}(X) = 0$ but $\kappa_k(X) \neq 0$, then $\|F_n - \Phi\|_p$ decreases like $n^{-(k-2)/2}$ for any p. This is natural, since, in our language, X is approximately normal of order exactly $k-2$. The difference from our Theorem 4.9, which cannot go beyond approximation order one, lies in Cramér's continuity condition. The quoted result fails for lattice distributions, since F_n then has jumps of order $n^{-1/2}$; see also Remark 5.4 below. Quite generally, Cramér's condition is known to serve as a substitute for regularity of functions, cf. [5].

5.2. Converse results. We turn to a few converse results. Our first instance, showing that Besov spaces are essentially the "correct" spaces is a direct consequence of Theorem 3.1. For the definition of $Z_t^{(n)}$, see Section 4.

PROPOSITION 5.2. *Suppose that X is approximately normal in the strong sense of order exactly $\mu > 0$, and let $0 < s \leqslant \mu$ and $g \in L_p$. If*

$$\|Eg(x+Z_t^{(n)}) - Eg(x+W_t)\|_{L_p(dx)} \leqslant Cn^{-s/2}$$

uniformly in $t \in T_h$, then $g \in B_p^{s,\infty}$. If the above is $o(n^{-\mu/2})$, then $g = 0$.

Given convergence for a single t, a weaker conclusion is possible. The following is a slight generalization of Theorem 2.2 in [25], treating $p = \infty$, and may be proved similarly.

PROPOSITION 5.3. *Assume that X has a one-dimensional lattice distribution and is approximately normal of order $s > 1$ in the standard sense. If $g \in L_p$ and*

$$\|Eg(x+S_n^*)-Eg(x+\mathcal{N})\|_{L_p(dx)} \leqslant Cn^{-s/2},$$

then $g \in B_p^{s-1,\infty}$.

Remark 5.4. The assumption on lattice distribution cannot be omitted. The result of Heyde and Nakata quoted in Section 5.1 shows that the result is false under Cramér's condition (C).

Let us finally discuss the optimality of the convergence rate $n^{-s/2}$.

PROPOSITION 5.5. *Given $s > 0$, there exists a function $g \in \bigcap_{1 \leqslant p \leqslant \infty} B_p^{s,\infty}$ such that if the first component of X is integer valued, and X is approximately normal of order s in the standard sense, then, for all p,*

$$\limsup_{n\to\infty} n^{s/2} \|Eg(x+S_n^*)-Eg(x+\mathcal{N})\|_{L_p(dx)} > 0.$$

If $p = \infty$, we can substitute $|Eg(S_n^*)-Eg(\mathcal{N})|$ for the norm, possibly replacing g by a translate.

Proof. The case $p = \infty$ is, except for the last claim, essentially Theorem 10.2 of Hedstrom [7]. For general p, it may be proved similarly to Theorem 2.1 in [25]. The final statement follows by localizing the latter proof by a suitable test function, we omit the details. ∎

The instance $d = 1$, $p = \infty$ may be compared to Theorem 2 of Borisov et al. [5], also dealing with lattice variables. The latter assumes that $s > 1$ is non-integral, and that $E|X|^{\lfloor s \rfloor + 3} < \infty$, where $\lfloor s \rfloor$ is the integral part of s. If the cumulants of order between 3 and s vanish, their conclusion is that given $\varphi(n) \to +\infty$, there exists $g \in B_\infty^{s,\infty}$ such that

$$\limsup_{n\to\infty} \varphi(n)\, n^{s/2} |Eg(S_n^*)-Eg(\mathcal{N})| > 0.$$

Comparing, we see that we can treat more values of s, that we need between two and three moments less, and that the extra weight $\varphi(n)$ has been removed. Moreover, [7] and [25] give actual constructions of g, whereas [5] relies on the Banach–Steinhaus theorem for the existence of such a function.

5.3. A signed central limit theorem. The results in Section 3 hold without the assumption that v is a positive measure. The same is true for those in Section 4 as long as the conclusion of Lemma 2.2 is valid, for example if the difference method is parabolic in the sense of John [13], [25]. This observation leads to higher-order central limit theorems in signed probability theory; see [9] and the references therein for a survey of related results.

Let P be a signed measure on a measurable space Ω such that $P(\Omega) = 1$, and define the basic probabilistic concepts as usual. In particular, a random variable is in this subsection a measurable function on Ω. As for expectations, put $EX = \int XdP$ and $|E|X = \int X|dP|$. Let $m \geqslant 2$ be even and consider the one-dimensional PDE

$$\frac{\partial u}{\partial t}+\frac{(-1)^{m/2}}{m!}\frac{\partial^m u}{\partial x^m} = 0,$$

solved by a random variable $\mathcal{N}_m$ with characteristic function $\exp(-\xi^m/m!)$. Let $X, X_1, X_2, \ldots$ be i.i.d. random variables with $EX = EX^2 = \ldots = EX^{m-1} = 0$ and $EX^m = (-1)^{m/2+1}$, and put $S_n^* = n^{-1/m}\sum_{j=1}^n X_j$. If, in addition, $|E||X|^{m+s} < \infty$ and $EX^k = E\mathcal{N}_m^k$ for $k < m+s$, then

$$||Eg(x+S_n^*)-Eg(x+\mathcal{N}_m)||_{L_p(dx)} \leqslant C\,||g||_{\dot{B}_q^{s,\infty}}\, n^{-s/m},$$

provided that the stability condition (3.3) holds. Without extra moment assumptions, a counterpart to Theorem 4.15 is valid. The basic convergence rate is thus $n^{-1/m}$, as suggested by Studnev [22].

If $p = 2$, stability is equivalent to $|\phi(\xi)| \leqslant 1$, where ϕ is the characteristic function of X. Similar criteria in L_p have been given for lattice distributions by Strang [21] and Thomée [23].

REFERENCES

[1] A. D. Barbour, *Asymptotic expansions based on smooth functions in the central limit theorem*, Probab. Theory Related Fields 72 (1986), pp. 289–303.

[2] J. Bergh and J. Löfström, *Interpolation Spaces*, Springer, New York 1976.

[3] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York 1968.

[4] R. P. Boas, *Lipschitz behavior and integrability of characteristic functions*, Ann. Math. Statist. 38 (1967), pp. 32–36.

[5] I. S. Borisov, D. A. Panchenko, and G. I. Skilyagina, *On minimal smoothness conditions for asymptotic expansions of moments in the CLT*, Siberian Adv. Math. 8 (1998), pp. 80–95.

[6] F. Götze and C. Hipp, *Asymptotic expansions in the central limit theorem under moment conditions*, Z. Wahrsch. Verw. Gebiete 42 (1978), pp. 67–87.

[7] G. W. Hedstrom, *The rate of convergence of some difference schemes*, SIAM J. Numer. Anal. 5 (1968), pp. 363–406.

[8] C. C. Heyde and T. Nakata, *On the asymptotic equivalence of L_p metrics for convergence to normality*, Z. Wahrsch. Verw. Gebiete 68 (1984), pp. 97–106.

[9] K. J. Hochberg, *Central limit theorem for signed distributions*, Proc. Amer. Math. Soc. 79 (1980), pp. 298–302.

[10] L. Hörmander, *Estimates for translation invariant operators in L_p spaces*, Acta Math. 104 (1960), pp. 93–140.

[11] L. Hörmander, *The Analysis of Linear Partial Differential Operators I*, 2nd edition, Springer, New York 1990.

[12] I. A. Ibragimov, *On the accuracy of Gaussian approximation to the distribution functions of sums of independent variables*, Theory Probab. Appl. 11 (1966), pp. 559–579.

[13] F. John, *On integration of parabolic equations by difference methods. I. Linear and quasi-linear equations for the infinite interval*, Comm. Pure Appl. Math. 5 (1952), pp. 155–211.

[14] J. Löfström, *Besov spaces in theory of approximation*, Ann. Mat. Pura Appl. (4) 85 (1970), pp. 93–184.
[15] J. Peetre, *New Thoughts on Besov Spaces*, Duke Univ. Press, 1976.
[16] J. Peetre and V. Thomée, *On the rate of convergence for discrete initial-value problems*, Math. Scand. 21 (1967), pp. 159–176.
[17] E. J. G. Pitman, *On the derivatives of a characteristic function at the origin*, Ann. Math. Statist. 27 (1956), pp. 1156–1160.
[18] R. D. Richtmyer and K. W. Morton, *Difference Methods for Initial-value Problems*, 2nd edition, Interscience, New York 1967.
[19] S. Saeki, *On convolution squares of singular measures*, Illinois J. Math. 24 (1980), pp. 225–232.
[20] T. L. Shervashidze, *On the convergence of densities of sums of independent random vectors*, Lecture Notes in Math. No 1021, Springer, New York 1982, pp. 576–586.
[21] G. Strang, *Polynomial approximation of Bernstein type*, Trans. Amer. Math. Soc. 105 (1962), pp. 525–535.
[22] Yu. P. Studnev, *Some generalizations of limit theorems in probability theory*, Theory Probab. Appl. 12 (1967), pp. 668–672.
[23] V. Thomée, *Stability of difference schemes in the maximum-norm*, J. Differential Equations 1 (1965), pp. 273–292.
[24] V. Thomée, *Finite difference methods for linear parabolic equations*, in: *Handbook of Numerical Analysis I*, North-Holland, Amsterdam 1990, pp. 5–196.
[25] V. Thomée and L. Wahlbin, *Convergence rates of parabolic difference schemes for non-smooth data*, Math. Comp. 28 (1974), pp. 1–13.
[26] H. F. Trotter, *An elementary proof of the central limit theorem*, Arch. Math. 10 (1959), pp. 226–234.
[27] O. B. Widlund, *On the rate of convergence for parabolic difference schemes II*, Comm. Pure Appl. Math. 23 (1970), pp. 79–96.
[28] A. Yu. Zaitsev, *On the Gaussian approximation of convolutions under multidimensional analogues of S. N. Bernstein's inequality condition*, Probab. Theory Related Fields 74 (1987), pp. 535–566.
[29] A. I. Zukov, *A limit theorem for difference operators* (in Russian), Uspekhi Mat. Nauk 14 (1959), pp. 129–136.

Saab Systems
S:t Olofsgatan 9A
753 21 Uppsala, Sweden
E-mail: lars.lindhagen@saabsystems.se

Received on 22.4.2005