Available online at www.sciencedirect.com
ScienceDirect
Expo. Math. 25 (2007) 215–233
EXPOSITIONES MATHEMATICAE
www.elsevier.de/exmath

# Lindeberg's central limit theorem à la Hausdorff

S.D. Chatterji

*Institut de mathématiques, Ecole Polytechnique Fédérale de Lausanne, Station 8, CH-1015 Lausanne, Switzerland*

Received 6 March 2006

## Abstract

We give a self-contained proof of Lindeberg's central limit theorem based on a presentation due to Hausdorff. Various historical remarks are also appended.


***MSC 2000:*** primary 60-03 60F05secondary 01A60

***Keywords:*** Central limit theorem; Lindeberg's condition

## 1. Introduction

Recently, while editing Hausdorff's Nachlass for the publication of vol. 5 of his collected works [9], we came across an attractive presentation of the central limit theorem of probability theory (under Liapounov's third moment conditions) using Lindeberg's novel and elementary proof of 1922 [10]. Part of the attractiveness lay in the fact that the theorem was presented in finitary terms and included a rate of convergence; although the latter was far from being optimal, it turned out that Hausdorff's arrangement of the proof could easily give the complete Lindeberg theorem with a rate of convergence included in its statement. Hausdorff's work was contained in his Lecture Notes from a course on probability theory given at the University of Bonn in 1923; Hausdorff probably avoided giving the full Lindeberg theorem in his lectures in order to concentrate on the essentials. It seemed to us to be of general interest to write out a complete proof of Lindeberg's theorem with a useful

*E-mail address:* srishti.chatterji@epfl.ch.

0723-0869/$ - see front matter © 2006 Elsevier GmbH. All rights reserved.
doi:10.1016/j.exmath.2006.10.003

remainder term included, using only the most elementary tools used by Lindeberg himself. The proofs we present here are essentially Lindeberg's as adjusted (mildly) by Hausdorff with some very minor modifications made by us; in fact, the remainder term is implicitly contained in Lindeberg's original paper. However, as some authors seemed to have felt that Lindeberg's original presentation was somewhat "intricate" (cf. [6, foot-note, p. 256] it seemed to us useful to work out the full details of Lindeberg's elementary proof in such a way that it can be presented easily in any introductory graduate course on probability or measure theory. The proof uses nothing more than Fubini's theorem for product probability measures in $\mathbb{R}^2$ and elementary calculus; use is made of the language of linear operators between normed spaces without using any non-trivial results of the latter theory; no use is made of characteristic functions (Fourier transforms).

Although Hausdorff's original 1923 Lecture Notes are available (in German, in [9, vol. 5]) accompanied by our commentaries and although various versions of Lindeberg's proof have already appeared in English (cf. [6, p. 256]) it appeared to us not without interest to give a full presentation in English, indicating at the same time where Lindeberg's original formulation was quite different from that given in present day text-books. Further historical comments are given at the end.

In Section 2 , we state two theorems formulated in purely measure-theoretical terms; in Sections 3–5 we prove them in detail. In Section 6, a brief complement is added to translate the theorems into the usual language of probability theory using random variables; this will allow the reader to compare them easily with the work which appears in standard probability texts. Section 7 concludes with various historical remarks.

The main body of the paper (Sections 2–5) could certainly be compressed to half its length by suppressing the elementary details provided there. We have preferred to spell out everything as explicitly as possible so that the material can be used easily for class-room teaching.

## 2. Statement of the theorems

We shall be concerned with probability measures in $\mathbb{R}$; if $\alpha$ is such a measure (defined on the Borel subsets of $\mathbb{R}$) we shall write

$$M(\alpha) = \int_{-\infty}^{\infty} x \, \mathrm{d}\alpha(x) = \text{mean value of } \alpha$$

only if $\int_{-\infty}^{\infty} |x| \, \mathrm{d}\alpha(x) < \infty$. If $M(\alpha) = 0$, we write

$$V(\alpha) = \int_{-\infty}^{\infty} x^2 \, \mathrm{d}\alpha(x), \quad C^3(\alpha) = \int_{-\infty}^{\infty} |x|^3 \, \mathrm{d}\alpha(x),$$

$V(\alpha)$ will be called the variance of $\alpha$ and $C^3(\alpha)$ the third absolute moment of $\alpha$. Further

$$F_\alpha(x) = \alpha(]-\infty, x]), \quad x \in \mathbb{R},$$

will stand for the distribution function associated with $\alpha$. The normal probability measure (in $\mathbb{R}$) $v = N(0; \sigma^2)$ with mean 0 and variance $\sigma^2$, $0 < \sigma < \infty$, is defined by

$$v(A) = \int_A \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-x^2/2\sigma^2}\,\mathrm{d}x, \quad A \subset \mathbb{R}, \quad A \text{ Borel},$$

If $\alpha$, $\beta$ are two probability measures in $\mathbb{R}$, their convolution $\alpha * \beta$ is the probability measure in $\mathbb{R}$ defined by

$$(\alpha * \beta)(A) = \int_A \alpha(A - x)\,\mathrm{d}\beta(x), \quad A \subset \mathbb{R}, \quad A \text{ Borel},$$

here $A - x$ is the set formed by $\{a - x;\, a \in A\}$.

**Theorem 1** (*Liapounov's central limit theorem*). *Let $\alpha_1, \ldots, \alpha_n$ be n probability measures in $\mathbb{R}$ with $M(\alpha_j) = 0$, $V(\alpha_j) = a_j^2$, $C(\alpha_j) < \infty$, $1 \leqslant j \leqslant n$; let*

$$b_n^2 = a_1^2 + \cdots + a_n^2, \quad d_n^3 = C^3(\alpha_1) + \cdots + C^3(\alpha_n), \quad \mu = \alpha_1 * \cdots * \alpha_n$$

*and $v = N(0; b_n^2)$. Then*

$$\sup_{x\in\mathbb{R}} |F_\mu(x) - F_v(x)| \leqslant \chi_0 \Lambda_n^{1/4}, \tag{1}$$

*where $\Lambda_n = (d_n/b_n)^3$ and $0 < \chi_0 < \infty$ is an absolute constant independent of $n, \alpha_1, \ldots, \alpha_n$.*

For the statement of Lindeberg's general theorem we define (with Lindeberg) the quantity $L(\alpha)$ for any probability measure $\alpha$ in $\mathbb{R}$; to do so, we use the function $s : \mathbb{R} \to \mathbb{R}$ given by

$$s(x) = \begin{cases} |x|^3 & \text{if } |x| < 1, \\ x^2 & \text{if } |x| \geqslant 1 \end{cases} \tag{2}$$

and put

$$L(\alpha) = \int_{-\infty}^{\infty} s(x)\,\mathrm{d}\alpha(x).$$

Note that $L(\alpha) < \infty$ iff $\int_{-\infty}^{\infty} |x|^2\,\mathrm{d}\alpha(x) < \infty$.

**Theorem 2** (*Lindeberg's central limit theorem*). *Let $\alpha_1, \ldots, \alpha_n$ be n probability measures in $\mathbb{R}$ with $M(\alpha_j) = 0$, $V(\alpha_j) = a_j^2$, $1 \leqslant j \leqslant n$, $a_1^2 + \cdots + a_n^2 = 1$. If $\mu = \alpha_1 * \cdots * \alpha_n$, $v = N(0; 1)$ then*

$$\sup_{x\in\mathbb{R}} |F_\mu(x) - F_v(x)| \leqslant \chi L_n^{1/4}, \tag{3}$$

*where $L_n = \sum_{j=1}^{n} L(\alpha_j)$ and $0 < \chi < \infty$ is an absolute constant independent of $n, \alpha_1, \ldots, \alpha_n$.*

## 3. Some lemmas

For ease of exposition, we have collected together here some notation and elementary lemmas for use in the proofs of the two main theorems stated above.

Let $B(\mathbb{R})$ be the (Banach) space of all bounded Borel functions $g : \mathbb{R} \to \mathbb{R}$ with $\|g\| = \sup_x |g(x)|$. For any probability measure $\alpha$ in $\mathbb{R}$ and for any $g$ in $B(\mathbb{R})$ we write

$$T_\alpha g(x) = \int_{-\infty}^{\infty} g(x-y)\,\mathrm{d}\alpha(y).$$

**Lemma 1.** *$T_\alpha : B(\mathbb{R}) \to B(\mathbb{R})$ is a positive linear operator with $\|T_\alpha\| \leqslant 1$.*

The proof is immediate; let us simply recall that

$$\|T_\alpha\| = \sup\{\|T_\alpha g\| : \|g\| \leqslant 1\}$$

and that

$$|T_\alpha g(x)| \leqslant \int_{-\infty}^{\infty} |g(x-y)|\,\mathrm{d}\alpha(y) \leqslant \|g\|.$$

Recall also that the positivity of $T_\alpha$ means simply that $T_\alpha g(x) \geqslant 0$ for all $x \in \mathbb{R}$ whenever $g(x) \geqslant 0$ for all $x \in \mathbb{R}$.

Note further that $T_\alpha 1_A(x) = \alpha(x-A)$ ($A \subset \mathbb{R}$, $A$ Borel where $1_A$ is the indicator function of the set $A$ and $x - A = \{x - a : a \in A\}$. Hence, if $\alpha, \beta$ are two probability measures in $\mathbb{R}$, then $T_\alpha g = T_\beta g$ for all $g \in B(\mathbb{R})$ iff $\alpha = \beta$.

**Lemma 2.** *If $\alpha, \beta$ are any two probability measures in $\mathbb{R}$, then $\alpha * \beta = \beta * \alpha$ and*

$$T_\alpha T_\beta = T_{\alpha*\beta} = T_\beta T_\alpha. \tag{4}$$

*Further, if $\beta_j = N(0; a_j^2)$, $1 \leqslant j \leqslant n$,* then

$$\beta_1 * \cdots * \beta_n = N(0; b_n^2),$$

*where $b_n^2 = a_1^2 + \cdots + a_n^2$.*

The proof of the statement concerning the convolution of normal probability measures will not be given here since it is quite elementary and standard. The relation (4) is very well-known and its proof can be found in many places; for completeness, we sketch here its proof very briefly, using Fubini's theorem. First note that for any $h \in B(\mathbb{R})$, we have

$$\int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} h(x+y)\,\mathrm{d}\alpha(x) \right) \mathrm{d}\beta(y) = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} h(x+y)\,\mathrm{d}\beta(y) \right) \mathrm{d}\alpha(x) \tag{5}$$

by Fubini's theorem. By taking $h = 1_A$ in (5), $A \subset \mathbb{R}$, $A$ Borel, we obtain

$$\int_{-\infty}^{\infty} \alpha(A-y)\,\mathrm{d}\beta(y) = \int_{-\infty}^{\infty} \beta(A-x)\,\mathrm{d}\alpha(x), \tag{6}$$

which gives $\alpha * \beta = \beta * \alpha$ by the very definition of convolutions given in Section 2; further, for all $x \in \mathbb{R}$, (see remark after Lemma 1)

$$T_\beta(T_\alpha 1_A)(x) = \int_{-\infty}^{\infty} T_\alpha 1_A(x-y)\,\mathrm{d}\beta(y) = \int_{-\infty}^{\infty} \alpha(x-y-A)\,\mathrm{d}\beta(y)$$
$$= (\alpha * \beta)(x - A)$$

and, similarly

$$T_\alpha(T_\beta 1_A)(x) = (\beta * \alpha)(x - A).$$

Thus, since $\alpha * \beta = \beta * \alpha$,

$$T_{\alpha*\beta}(1_A) = T_\alpha(T_\beta 1_A) = T_\beta(T_\alpha a_A),$$

whence follows

$$T_{\alpha*\beta}(g) = T_\alpha(T_\beta g) = T_\beta(T_\alpha g)$$

for all $g \in B(\mathbb{R})$ by linearity and passage to limit. This establishes (4) and Lemma 2.

**Lemma 3.** *Let $\{A_1, \ldots, A_n\}, \{B_1, \ldots, B_n\}$ be linear operators on any normed space $E$* with $\|A_i\| \leqslant 1, \|B_i\| \leqslant 1, 1 \leqslant i \leqslant n$ *and* $A_i A_j = A_j A_i, B_i B_j = B_j B_i$ for $1 \leqslant i, j \leqslant n$. *Then*

$$\|A_1 \ldots A_n - B_1 \ldots B_n\| \leqslant \sum_{i=1}^{n} \|A_i - B_i\|. \tag{7}$$

The proof is by induction; for $n=1$, (7) is trivial. Supposing (7) to hold for $n=m$, we prove it for $n = m+1$; write $P_j = A_1 \ldots A_j$, $Q_j = B_1 \ldots B_j$ and note that $\|P_j\| \leqslant 1$, $\|Q_j\| \leqslant 1$; using commutativity of the $A's$ and that of the $B's$, we have

$$\begin{aligned} P_{m+1} - Q_{m+1} &= P_m A_{m+1} - Q_m B_{m+1}, \\ &= A_{m+1} P_m - B_{m+1} P_m + B_{m+1} P_m - B_{m+1} Q_m, \\ &= (A_{m+1} - B_{m+1}) P_m + B_{m+1}(P_m - Q_m), \end{aligned}$$

which gives (using the inductive hypothesis)

$$\|P_{m+1} - Q_{m+1}\| \leqslant \|A_{m+1} - B_{m+1}\| + \|P_m - Q_m\| \leqslant \sum_{i=1}^{m+1} \|A_i - B_i\|$$

establishing (7) for $n = m+1$.

**Lemma 4.** *Let $\mathscr{C}_b^{(k)}(\mathbb{R})$ be the space of functions $g : \mathbb{R} \to \mathbb{R}$ which are k times continuously differentiable with $\|g^{(k)}\| < \infty$; then any $g \in \mathscr{C}_b^{(k)}(\mathbb{R})$ is in $\mathscr{C}_b^{(j)}(\mathbb{R})$ for $j = 0, 1, \ldots, k$; further if $g \in \mathscr{C}_b^{(k)}(\mathbb{R})$ then $T_\alpha g \in \mathscr{C}_b^{(k)}(\mathbb{R})$ and*

$$\|(T_\alpha g)^{(j)}\| \leqslant \|g^{(j)}\|, \quad 0 \leqslant j \leqslant k,$$

*where $\alpha$ is any probability measure in $\mathbb{R}$.*

The first part of Lemma 4 is elementary and follows from the mean-value theorem; the proof of the second part is contained in the fact that

$$(T_\alpha g)^{(j)}(x) = \int_{-\infty}^{\infty} g^{(j)}(x-y)\,\mathrm{d}\alpha(y)$$

whenever $g \in \mathscr{C}_b^{(j)}(\mathbb{R})$.

**Lemma 5.** *Let $\alpha$, $\beta$ be two probability measures in $\mathbb{R}$ with $M(\alpha)=0$, $M(\beta)=0$, $V(\alpha)=V(\beta)$ and $C(\alpha)$, $C(\beta)$ finite; then for any $g \in \mathscr{C}_b^{(3)}(\mathbb{R})$:*

$$\|T_\alpha g - T_\beta g\| \leqslant \tfrac{1}{6}\|g^{(3)}\|\{C^3(\alpha) + C^3(\beta)\}.$$

**Proof.** Using Taylor's expansion of order 3 we have, for $t$, $x$ in $\mathbb{R}$,

$$g(t-x) = g(t) - xg'(t) + \frac{x^2}{2}g''(t) - \frac{x^3}{6}\rho(t,x),$$

where $|\rho(t,x)| \leqslant \|g^{(3)}\|$. If $\sigma^2 = V(\alpha) = V(\beta)$ we obtain by integrating with respect to $\mathrm{d}\alpha$ that

$$T_\alpha g(t) = \int_{-\infty}^{\infty} g(t-x)\,\mathrm{d}\alpha(x) = g(t) + \frac{\sigma^2}{2}g''(t) - \frac{1}{6}\int_{-\infty}^{\infty} x^3\rho(t,x)\,\mathrm{d}\alpha(x),$$

whence

$$\left|T_\alpha g(t) - g(t) - \frac{\sigma^2}{2}g''(t)\right| \leqslant \tfrac{1}{6}\|g^{(3)}\|C^3(\alpha),$$

similarly,

$$\left|T_\beta g(t) - g(t) - \frac{\sigma^2}{2}g''(t)\right| \leqslant \tfrac{1}{6}\|g^{(3)}\|C^3(\beta),$$

whence

$$|T_\alpha g(t) - T_\beta g(t)| \leqslant \tfrac{1}{6}\|g^{(3)}\|\{C^3(\alpha) + C^3(\beta)\},$$

which proves Lemma 5.

**Lemma 6.** *Let $\alpha_j, \beta_j, 1 \leqslant j \leqslant n$, be probability measures in $\mathbb{R}$ such that $M(\alpha_j) = 0$, $M(\beta_j) = 0$, $V(\alpha_j) = V(\beta_j)$, $C(\alpha_j)$, $C(\beta_j)$ finite for $1 \leqslant j \leqslant n$; let $\mu = \alpha_1 * \cdots * \alpha_n$, $v = \beta_1 * \cdots * \beta_n$; then for any $g \in \mathscr{C}_b^{(3)}(\mathbb{R})$,*

$$\|T_\mu g - T_v g\| \leqslant \frac{1}{6}\|g^{(3)}\| \sum_{j=1}^{n}\{C^3(\alpha_j) + C^3(\beta_j)\}.$$

This follows immediately from Lemmas 1–5 since

$$T_\mu = T_{\alpha_1}T_{\alpha_2}\ldots T_{\alpha_n}, \quad T_v = T_{\beta_1}T_{\beta_2}\ldots T_{\beta_n}.$$

**Lemma 7.** *If* $\beta = N(0;\sigma^2)$, $M(\alpha)=0$, $V(\alpha)=\sigma^2$, $C^3(\alpha)<\infty$, *then*

$$C^3(\beta) < 2\,C^3(\alpha).$$

**Proof.** It is well-known that

$$C^3(\beta) = \int_{-\infty}^{\infty} |x|^3 \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-x^2/2\sigma^2}\,\mathrm{d}x = \frac{2\sqrt{2}}{\sqrt{\pi}}\sigma^3,$$

also

$$\sigma^3 = \left(\int_{-\infty}^{\infty} |x|^2\,\mathrm{d}\alpha(x)\right)^{3/2} \leqslant \int_{-\infty}^{\infty} |x|^3\,\mathrm{d}\alpha(x) = C^3(\alpha)$$

by Jensen's inequality. Hence

$$C^3(\beta) = \frac{2\sqrt{2}}{\sqrt{\pi}}\sigma^3 \leqslant \frac{2\sqrt{2}}{\sqrt{\pi}} C^3(\alpha) < 2\,C^3(\alpha). \qquad \square$$

**Lemma 8.** *Let* $\alpha_j, \beta_j, 1\leqslant j\leqslant n$, *be probability measures in* $\mathbb{R}$ *with* $\beta_j = N(0;a_j^2)$, $M(\alpha_j)=0$, $V(\alpha_j)=a_j^2$, $C^3(\alpha_j)<\infty$, $1\leqslant j\leqslant n$; *let* $\mu = \alpha_1 * \cdots * \alpha_n$, $\nu = \beta_1 * \cdots * \beta_n$; *then, for* $g \in \mathscr{C}_b^{(3)}(\mathbb{R})$,

$$\|T_\mu g - T_\nu g\| \leqslant \frac{1}{2}\|g^{(3)}\| \sum_{j=1}^{n} C^3(\alpha_j).$$

This follows immediately from Lemmas 6 and 7.

**Lemma 9.** *Let* $g \in B(\mathbb{R})$ *be such that* $0\leqslant g\leqslant 1$ *with (for some* $\ell > 0$*):*

$$g(x) = \begin{cases} 0 & \text{if } x\leqslant 0,\\ 1 & \text{if } x\geqslant \ell,\end{cases}$$

*then for any two probability measures* $\alpha$, $\beta$ *in* $\mathbb{R}$, *we have, for* $t\in\mathbb{R}$:

(i)

$$T_\alpha g(t) \leqslant F_\alpha(t) \leqslant T_\alpha g(t+\ell),$$
$$T_\beta g(t) \leqslant F_\beta(t) \leqslant T_\beta g(t+\ell),$$

(ii)

$$F_\alpha(t) - F_\beta(t) \leqslant \|T_\alpha g - T_\beta g\| + F_\beta(t+\ell) - F_\beta(t),$$
$$F_\beta(t) - F_\alpha(t) \leqslant \|T_\alpha g - T_\beta g\| + F_\beta(t) - F_\beta(t-\ell).$$

*Further, if* $\beta = N(0;\sigma^2)$ *then, for* $t\in\mathbb{R}$,

(iii)

$$|F_\alpha(t) - F_\beta(t)| \leqslant \|T_\alpha g - T_\beta g\| + \frac{\ell}{\sigma\sqrt{2\pi}}.$$

**Proof.** (i) $T_\alpha g(t) = \int_{-\infty}^{\infty} g(t-x)\,\mathrm{d}\alpha(x) = \int_{]-\infty,t[} g(t-x)\,\mathrm{d}\alpha(x) \leqslant \int_{]-\infty,t[} \mathrm{d}\alpha(x) \leqslant F_\alpha(t)$ since $g(t-x)=0$ if $x \geqslant t$ and $g \leqslant 1$; further, since $g \geqslant 0$,

$$\begin{aligned} T_\alpha g(t+\ell) &= \int_{-\infty}^{\infty} g(t+\ell-x)\,\mathrm{d}\alpha(x) \geqslant \int_{]-\infty,t]} g(t+\ell-x)\,\mathrm{d}\alpha(x) \\ &= \int_{]-\infty,t]} \mathrm{d}\alpha(x) = F_\alpha(t) \end{aligned}$$

because $g(t+\ell-x)=1$ if $x \leqslant t$. The same holds for $\beta$; this proves (i).

(ii) This follows from (i); indeed, from (i),

$$\begin{aligned} F_\alpha(t) - F_\beta(t) &\leqslant T_\alpha g(t+\ell) - F_\beta(t) \\ &= \{T_\alpha g(t+\ell) - T_\beta g(t+\ell)\} + T_\beta g(t+\ell) - F_\beta(t) \\ &\leqslant \|T_\alpha g - T_\beta g\| + F_\beta(t+\ell) - F_\beta(t), \end{aligned}$$

similarly, from (i), $F_\beta(t-\ell) \leqslant T_\beta g(t-\ell+\ell) = T_\beta g(t)$ which gives

$$\begin{aligned} F_\beta(t) - F_\alpha(t) &= \{F_\beta(t) - F_\beta(t-\ell)\} + F_\beta(t-\ell) - F_\alpha(t) \\ &\leqslant \{F_\beta(t) - F_\beta(t-\ell)\} + T_\beta g(t) - T_\alpha g(t) \\ &\leqslant \{F_\beta(t) - F_\beta(t-\ell)\} + \|T_\alpha g - T_\beta g\| \end{aligned}$$

proving (ii). For (iii), we simply note that

$$F_\beta(t+\ell) - F_\beta(t) = \int_t^{t+\ell} \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-x^2/2\sigma^2}\,\mathrm{d}x \leqslant \frac{\ell}{\sigma\sqrt{2\pi}}$$

and similarly that

$$F_\beta(t) - F_\beta(t-\ell) \leqslant \frac{\ell}{\sigma\sqrt{\pi}},$$

this along with (ii) proves (iii). □

## 4. Proof of Theorem 1

We use the same notations as in the statement of the theorem; a key element in the proof is to notice that $v = N(0; b_n^2)$ is equal to $\beta_1 * \cdots * \beta_n$ where $\beta_i = N(0; a_i^2)$, $1 \leqslant i \leqslant n$ (cf. Lemma 2).

Fix a function $h : \mathbb{R} \to \mathbb{R}$ in $\mathscr{C}_b^{(3)}(\mathbb{R})$ such that

$$h(x) = \begin{cases} 0 & \text{if } x \leqslant 0, \\ 1 & \text{if } x \geqslant 1, \end{cases}$$

with $0 < h(x) < 1$ if $0 < x < 1$; let $m = \|h^{(3)}\|$. For example, we may take

$$h(x) = \int_0^x t^3(1-t)^3\,\mathrm{d}t \Big/ \int_0^1 t^3(1-t)^3\,\mathrm{d}t \tag{8}$$

if $0 < x < 1$ and $h(x) = 0$ for $x \leqslant 0$, $h(x) = 1$ for $x \geqslant 1$. Let $\ell$ be some number, $0 < \ell < \infty$, which we shall fix later and put $g(x) = h(x/\ell)$; then $g \in \mathscr{C}_b^{(3)}(\mathbb{R})$, $0 \leqslant g \leqslant 1$, and

$$g^{(3)}(x) = \ell^{-3} h^{(3)}(x/\ell), \quad \|g^{(3)}\| = m/\ell^3.$$

From Lemma 8 we have

$$\|T_\mu g - T_\nu g\| \leqslant \frac{m}{2\ell^3} \sum_{j=1}^{n} C^3(\alpha_j) = \frac{m}{2\ell^3} d_n^3.$$

From Lemma 9(iii) (with $\alpha = \mu$, $\beta = \nu$) we have, for $t \in \mathbb{R}$,

$$|F_\mu(t) - F_\nu(t)| \leqslant \frac{m}{2\ell^3} d_n^3 + \frac{\ell}{b_n\sqrt{2\pi}}. \tag{9}$$

We now take $\ell = b_n^{1/4} d_n^{3/4}$; this essentially minimizes the right-hand side of (9); then (9) becomes

$$|F_\mu(t) - F_\nu(t)| \leqslant \left(\frac{m}{2} + \frac{1}{\sqrt{2\pi}}\right)\left(\frac{d_n}{b_n}\right)^{3/4},$$

which proves (1) with $\chi_0 = (m/2) + (1/\sqrt{2\pi})$.

N.B. If we choose $h$ as given by (8), then $m$ turns out to be less than 54 and so $\chi_0 < 28$. For reasons given later, we find it not very useful to determine the best value of $\chi_0$.

## 5. Proof of Theorem 2

This proof is very similar to that of Theorem 1; we just need a lemma in which the third moments of Lemmas 5, 6 and 8 are replaced by integrals involving the Lindeberg $s$-function (2). First, we give an elementary lemma relating the variance with the $s$-function integral.

**Lemma 10.** *Let $\alpha$ be a probability measure in $\mathbb{R}$ with $M(\alpha) = 0$, $V(\alpha) = a^2$ with $0 < a < 1$; then*

$$a^3 < 4\int_{-\infty}^{\infty} s(x)\,\mathrm{d}\alpha(x).$$

**Proof.** Suppose first that

$$\int_{|x| \geqslant 1} x^2\,\mathrm{d}\alpha(x) \geqslant \frac{a^2}{4},$$

in this case the inequality of the lemma is trivial since

$$\int_{-\infty}^{\infty} s\,\mathrm{d}\alpha \geqslant \int_{|x|\geqslant 1} x^2\,\mathrm{d}\alpha(x) \geqslant \frac{a^2}{4} > \frac{a^3}{4}$$

because $0 < a < 1$ and $a^2 > a^3$. Next, suppose that

$$\int_{|x|\geqslant 1} x^2\,\mathrm{d}\alpha(x) < \frac{a^2}{4},$$

then

$$\int_{|x|<1} x^2\,\mathrm{d}\alpha(x) = a^2 - \int_{|x|\geqslant 1} x^2\,\mathrm{d}\alpha(x) > \frac{3a^2}{4}$$

and

$$\int_{\frac{a}{2}<|x|<1} x^2\,\mathrm{d}\alpha(x) > \frac{3a^2}{4} - \int_{|x|\leqslant\frac{a}{2}} x^2\,\mathrm{d}\alpha(x) \geqslant \frac{3a^2}{4} - \frac{a^2}{4} = \frac{a^2}{2}$$

so that

$$\int_{|x|<1} |x|^3\,\mathrm{d}\alpha(x) \geqslant \int_{\frac{a}{2}<|x|<1} |x|^3\,\mathrm{d}\alpha(x) \geqslant \frac{a}{2}\int_{\frac{a}{2}<|x|<1} |x|^2\,\mathrm{d}\alpha(x) > \frac{a^3}{4},$$

hence,

$$\int_{-\infty}^{\infty} s(x)\,\mathrm{d}\alpha(x) \geqslant \int_{|x|<1} |x|^3\,\mathrm{d}\alpha(x) > \frac{a^3}{4},$$

which again gives the desired inequality, which is thus established in all cases where $0 < a < 1$. □

**Lemma 11.** *Let* $\alpha$ *be any probability measure in* $\mathbb{R}$ *with* $M(\alpha) = 0$, $V(\alpha) = a^2$ *and let* $\beta = N(0; a^2)$. *If* $g \in \mathscr{C}_b^{(3)}(\mathbb{R})$ *with* $\|g^{(3)}\| \leqslant m$, $\|g^{(2)}\| \leqslant m$, *then*

$$\|T_\alpha g - T_\beta g\| \leqslant m\left\{\int_{-\infty}^{\infty} s\,\mathrm{d}\alpha + \sqrt{\frac{2}{\pi}}\frac{1}{3}a^3\right\},$$

*further, if* $0 < a < 1$, *then*

$$\|T_\alpha g - T_\beta g\| \leqslant 3m\int_{-\infty}^{\infty} s\,\mathrm{d}\alpha.$$

**Proof.** Write, for $t, x$ in $\mathbb{R}$,

$$g(t-x) = g(t) - xg'(t) + \frac{x^2}{2}g''(t) + R(t,x) \tag{10}$$

and note that (for some $\theta$ between 0 and $x$)

$$R(t,x)=\begin{cases} -\dfrac{x^3}{6}g^{(3)}(t-\theta) & \text{if } |x|<1,\\[2mm] \dfrac{x^2}{2}\{g^{(2)}(t-\theta)-g^{(2)}(t)\} & \text{if } |x|\geqslant 1,\end{cases}$$

so that

$$|R(t,x)|\leqslant \begin{cases} \dfrac{m|x|^3}{6} & \text{if } |x|<1,\\[2mm] mx^2 & \text{if } |x|\geqslant 1.\end{cases} \tag{11}$$

Integrating (10) with respect to $\mathrm{d}\alpha(x)$ we obtain

$$T_\alpha g(t)=g(t)+\frac{a^2}{2}g''(t)+\int_{-\infty}^{\infty}R(t,x)\,\mathrm{d}\alpha(x)$$

whence, using (11), we have

$$\left|T_\alpha g(t)-g(t)-\frac{a^2}{2}g''(t)\right|\leqslant m\int_{-\infty}^{\infty}s\,\mathrm{d}\alpha. \tag{12}$$

Similarly, using $\beta$ in place of $\alpha$, we have

$$\begin{aligned}\left|T_\beta g(t)-g(t)-\frac{a^2}{2}g''(t)\right| &\leqslant \int_{-\infty}^{\infty}|R(t,x)|\,\mathrm{d}\beta(x)\\ &\leqslant \frac{m}{6}\int_{-\infty}^{\infty}|x|^3\,\mathrm{d}\beta(x)=\frac{m}{6}\cdot\frac{2\sqrt{2}}{\sqrt{\pi}}a^3=\frac{m}{3}a^3\sqrt{\frac{2}{\pi}}\end{aligned} \tag{13}$$

using the fact that, for all $t, x$, $|R(t,x)|=|(x^3/6)h^{(3)}(t-\theta)|\leqslant(|x|^3/6)m$. Combining (12) and (13) we obtain the first inequality of Lemma 11; if $0<a<1$, we use Lemma 10 to obtain the second inequality. □

We now pass to the proof of Theorem 2; note first that since $a_1^2+\cdots+a_n^2=1$, we may and do suppose that $0<a_j<1$, $1\leqslant j\leqslant n$; as in the proof of Lemma 1, we observe also that $\nu=N(0;1)$ equals $\beta_1*\cdots*\beta_n$ where $\beta_j=N(0;a_j^2)$, $1\leqslant j\leqslant n$.

Fix a function $h\in\mathscr{C}_b^{(3)}(\mathbb{R})$ exactly as in the proof of Theorem 1 in Section 4 with $0<h(x)<1$ if $0<x<1$ and $h(x)=0$ if $x\leqslant 0$, $h(x)=1$ if $x\geqslant 1$ and put $m=\|h^{(3)}\|$; it is easy to see that for such a function $h$, $\|h^{(2)}\|\leqslant m$ since $h^{(2)}(x)=0$ if $x\leqslant 0$ and if $x\geqslant 1$ and $|h^{(2)}(x)|\leqslant mx$ if $0<x<1$ because

$$h^{(2)}(x)=\int_0^x h^{(3)}(t)\,\mathrm{d}t,\quad 0<x<1.$$

Now take any number $\ell$, $0<\ell\leqslant 1$ (to be fixed later) and define $g$ by $g(t)=h(t/\ell)$; then $g(t)=0$ if $t\leqslant 0$, $g(t)=1$ if $t\geqslant\ell$, $0\leqslant g\leqslant 1$, $g^{(3)}(t)=\ell^{-3}h^{(3)}(t/\ell)$. To this $g$ apply

Lemma 11 with $\alpha=\alpha_j$, $\beta=\beta_j=N(0;\, a_j^2)$, $1\leqslant j\leqslant n$, where the $\alpha_j$'s are probability measures in $\mathbb{R}$ as in the statement of Theorem 2; since $0<a_j<1$, we have

$$\|T_{\alpha_j}g - T_{\beta_j}g\| \leqslant \frac{3m}{\ell^3}\int_{-\infty}^{\infty} s\,\mathrm{d}\alpha_j,\ \ 1\leqslant j\leqslant n.$$

We now use Lemmas 2 and 3 with $\mu=\alpha_1 * \cdots * \alpha_n$, $v=\beta_1 * \cdots * \beta_n = N(0;\,1)$ to obtain

$$\|T_\mu g - T_v g\| \leqslant \frac{3m}{\ell^3}\sum_{j=1}^{n}\int_{-\infty}^{\infty} s\,\mathrm{d}\alpha_j = \frac{3m}{\ell^3}L_n.$$

Finally, from Lemma 9(iii) (with $\alpha=\mu$, $\beta=v$) we have, for $t\in\mathbb{R}$,

$$|F_\mu(t)-F_v(t)| \leqslant \frac{3m}{\ell^3}L_n + \frac{1}{\sqrt{2\pi}}\ell. \tag{14}$$

Note that $L_n\leqslant 1$ since

$$L_n = \sum_{j=1}^{n}\int_{-\infty}^{\infty} s\,\mathrm{d}\alpha_j \leqslant \sum_{j=1}^{n}\int_{-\infty}^{\infty} x^2\,\mathrm{d}\alpha_j = \sum_{j=1}^{n} a_j^2 = 1.$$

Now take $\ell = L_n^{1/4}$ in (14); this gives, for $t\in\mathbb{R}$,

$$|F_\mu(t)-F_v(t)| \leqslant \left(3m + \frac{1}{\sqrt{2\pi}}\right)L_n^{1/4} = \chi L_n^{1/4},$$

where $\chi = 3m + (1/\sqrt{2\pi})$.

N.B. With the choice of $h$ as in (8) of Section 4, we have $m<54$ and $\chi<163$.

## 6. Translation into probabilistic language and complements

In terms of real-valued random variables $x_1,\ldots,x_n$ which are mutually independent with $\mathbb{E}x_j=0$, $\mathrm{Var}(x_j)=a_j^2<\infty$ and $b_n^2=a_1^2+\cdots+a_n^2$ Theorems 1 and 2 can be written as follows:

**Theorem 1′.** *If further $\mathbb{E}|x_j|^3<\infty$, $1\leqslant j\leqslant n$, with $d_n^3=\sum_{j=1}^n \mathbb{E}|x_j|^3$ and $\Lambda_n=(d_n/b_n)^3$ then*

$$\sup_{t\in\mathbb{R}}\left|\mathscr{P}\left(\frac{x_1+\cdots+x_n}{b_n}\leqslant t\right) - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t}\mathrm{d}^{-u^2/2}\,\mathrm{d}u\right| \leqslant \chi_0\Lambda_n^{1/4}, \tag{15}$$

*where $0<\chi_0<\infty$ is some absolute constant.*

**Theorem 2′.** *We have,*

$$\sup_{t\in\mathbb{R}}\left|\mathscr{P}\left(\frac{x_1+\cdots+x_n}{b_n}\leqslant t\right) - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t}\mathrm{e}^{-u^2/2}\,\mathrm{d}u\right| \leqslant \chi L_n^{1/4}, \tag{16}$$

*where*

$$L_n = \sum_{j=1}^{n} \mathbb{E}s(x_j/b_n)$$

*and* $0 < \chi < \infty$ *is some absolute constant.*

Thus, from Theorem 1′, if $\Lambda_n \to 0$ as $n \to \infty$ (Liapounov's condition), then,

$$\mathscr{L}\left(\frac{x_1 + \cdots + x_n}{b_n}\right) \to N(0, 1) \text{ as } n \to \infty, \tag{17}$$

which is Liapounov's central limit theorem; as usual, $\mathscr{L}(y)$ denotes the law of the random variable $y$. Similarly, from theorem 2′, if $L_n \to 0$ as $n \to \infty$ (Lindeberg's condition) then again (17) holds which is Lindeberg's central limit theorem.

The condition $L_n \to 0$ as $n \to \infty$ is the original form of the condition as given in Lindeberg's 1922 paper [10]; this condition is equivalent to the following: let

$$L_{n,\varepsilon} = \frac{1}{b_n^2} \sum_{j=1}^{n} \mathbb{E}(|x_j|^2; |x_j| \geqslant \varepsilon b_n), \quad \varepsilon > 0,$$

here $\mathbb{E}(y; A)$ stands for $\int_A y \, \mathrm{d}\mathscr{P} = \int_{y^{-1}(A)} t \, \mathrm{d}\alpha(t)$ where $\alpha = \mathscr{L}(y)$ and $A$ any measurable subset of the underlying probability space where the random variables and the basic probability measure $\mathscr{P}$ are defined; then $L_n \to 0$ as $n \to \infty$ iff $L_{n,\varepsilon} \to 0$ as $n \to \infty$ for any $\varepsilon > 0$. This equivalence is easy to establish by proving that if $0 < \varepsilon < 1$, then

$$L_{n,\varepsilon} \leqslant \frac{1}{\varepsilon} L_n, \quad L_n \leqslant L_{n,\varepsilon} + \varepsilon, \tag{18}$$

note that if $\varepsilon < \varepsilon'$ then $L_{n,\varepsilon} \geqslant L_{n,\varepsilon'}$.

To prove (18), we note that if $0 < \varepsilon < 1$, then

$$\begin{cases} \varepsilon x^2 \leqslant s(x) \leqslant x^2 & \text{if } |x| \geqslant \varepsilon, \\ s(x) = |x|^3 & \text{if } |x| < \varepsilon. \end{cases} \tag{19}$$

Hence,

$$\begin{aligned} L_n &= \sum_{j=1}^{n} \int_{-\infty}^{\infty} s(x/b_n) \, \mathrm{d}\alpha_j(x) \geqslant \sum_{j=1}^{n} \int_{|x| \geqslant \varepsilon b_n} s(x/b_n) \, \mathrm{d}\alpha_j(x) \\ &\geqslant \varepsilon \sum_{j=1}^{n} \int_{|x| \geqslant \varepsilon b_n} (x^2/b_n^2) \, \mathrm{d}\alpha_j(x) = \varepsilon L_{n,\varepsilon}, \end{aligned}$$

this proves the first inequality in (18); further use of (19) gives

$$\begin{aligned} L_n &= \sum_{j=1}^{n}\left\{\int_{|x|\geqslant \varepsilon b_n} s(x/b_n)\,\mathrm{d}\alpha_j(x) + \int_{|x|<\varepsilon b_n} s(x/b_n)\,\mathrm{d}\alpha_j(x)\right\} \\ &\leqslant \sum_{j=1}^{n}\left\{\int_{|x|\geqslant \varepsilon b_n} (x^2/b_n^2)\,\mathrm{d}\alpha_j(x) + \int_{|x|<\varepsilon b_n} (|x|^3/b_n^3)\,\mathrm{d}\alpha_j(x)\right\} \\ &\leqslant L_{n,\varepsilon} + \frac{\varepsilon}{b_n^2}\sum_{j=1}^{n}\int_{|x|<\varepsilon b_n} x^2\,\mathrm{d}\alpha_j(x) \\ &\leqslant L_{n,\varepsilon} + \frac{\varepsilon}{b_n^2}\sum_{j=1}^{n}\int_{-\infty}^{\infty} x^2\,\mathrm{d}\alpha_j(x) \\ &= L_{n,\varepsilon} + \varepsilon, \end{aligned}$$

which establishes (18) for $0<\varepsilon<1$.

It is in the form, $L_{n,\varepsilon}\to 0$ as $n\to\infty$ for all $\varepsilon>0$, that the Lindeberg condition is usually given in most standard probability texts like Feller [6], Chow and Teicher [3]; in the latter, there is an interesting alternative form (cf. [3, p. 295]), still different from Lindeberg's original condition.

It is well-known further that, provided that $\mathbb{E}|x_j|^3<\infty$, $j\geqslant 1$, then

$$\lim_{n\to\infty}\Lambda_n = 0 \Rightarrow \lim_{n\to\infty} L_{n,\varepsilon}=0 \quad \text{for all } \varepsilon>0,$$

where $\Lambda_n$ is defined in Theorem 1′ (or Theorem 1). This is proved, for example in [3, p. 298], under the more general condition of $\mathbb{E}|x_j|^{2+\delta}<\infty$ (for some $\delta>0$), i.e.

$$\lim_{n\to\infty}\frac{1}{b_n^{2+\delta}}\sum_{j=1}^{n}\mathbb{E}|x_j|^{2+\delta}=0 \quad \text{for some } \delta>0 \Rightarrow \lim_{n\to\infty} L_{n,\varepsilon}=0 \quad \text{for all } \varepsilon>0.$$

It is a famous theorem of Feller (1935) that Lindeberg's condition is an optimal one for the validity of the central limit theorem in the sense that if $\mathbb{E}x_j^2<\infty$, $j\geqslant 1$ and (17) holds and if $a_n/b_n\to 0$, $b_n\to\infty$ as $n\to\infty$ (where $a_n^2=\mathbb{E}x_n^2$ and $b_n^2=a_1^2+\cdots+a_n^2$) then Lindeberg's condition holds (see [3, p. 296; 6, p. 492]). For a deeper, intuitive understanding of this we must refer to appropriate texts (e.g. [6]).

Let us recall here the classical result of Berry-Esséen (1941–1945) which gives the optimal estimates in (1) or (15) by replacing $\chi_0\Lambda_n^{1/4}$ there by $c\Lambda_n$ where $0<c<\infty$ is an absolute constant; much work has gone into an exact determination of $c$; the best result seems to be that of Paul van Beek giving $c\leqslant 0.7975$ (cf. Journal of Probability Theory and Applications, vol. 23, 1972). A proof of various forms of the Berry-Esséen theorem appears in several books (cf. e.g. [3, p. 304; 6, Chapter XVI]).

As is clear from the Berry-Esséen theorem cited here, the rate of convergence contained in the statements of Theorem 1 or 1′ is far from being the optimal one. This becomes more vivid in the case of independent identically distributed random variables $x_1, x_2, \ldots$ in Theorem 1′ i.e. in the case where $\alpha_1=\alpha_2=\ldots$ in Theorem 1; here $d_n^3=n\gamma^3$, $b_n^2=n\sigma^2$

where $\gamma^3 = \mathbb{E}|x_1|^3 = \int_{-\infty}^{\infty} |x|^3 \, \mathrm{d}\alpha_1(x)$ and $\sigma^2 = \mathrm{Var}(x_1) = \int_{-\infty}^{\infty} |x|^2 \, \mathrm{d}\alpha_1(x)$ so that $\Lambda_n = (d_n/b_n)^3 = (\gamma/\sigma)^3 n^{-1/2}$. Thus the rate of convergence to 0 according to Theorems 1, 1′ is of the type const. $n^{-1/8}$ whereas the true rate according to the Berry-Eséen theorem is given by const. $n^{-1/2}$. This is the reason why we have found it uninteresting to obtain the best values for $\chi_0$, $\chi$ in our Theorems 1 and 2.

As regards the rate of convergence in Theorems 2 or 2′, standard monographs have not been informative. However, research reported in [12] indicates that the estimates of Theorem 2 or 2′ as given in (3) or (16) can be improved to $A \cdot L_n$, where $A$ is some absolute constant (instead of our $\chi L_n^{1/4}$); we do not know of any attempt at estimating the constant $A$; there also are more general results using functions other than the $s$-function of Lindeberg; cf. Petrov's article [11] in [12, p. 5]. The Lindeberg method has been extended to infinite-dimensional vector spaces in the article of Bentkus et al. [1] in [12] (cf. pp. 42–50).

## 7. Historical and other remarks

This is not the place to give a complete history of the complicated evolution of the central limit theorem; a clear and detailed account (starting with de Moivre, Laplace, Cauchy, Poisson, Bessel, Chebyshev, Markov, Liapounov and others) has been given in Hald [8]. We shall only make a few remarks concerning Lindeberg and Hausdorff.

Lindeberg (1876–1932) had first proved (in 1920) a limit theorem using third moments in a Finnish journal (cf. exact reference in his paper [10]); at the time of the writing of the 1920 paper, Lindeberg was not aware of Liapounov's work of 1900–1901 and he refers only to Chebyshev, Markov and von Mises. Already, his 1920 paper contains most of the ingredients of his basic method which avoids characteristic functions; he does not seem to have been aware of the latter which were already used efficiently by Liapounov. Lindeberg's 1922 paper [10] mentions Liapounov's work but does not make any use of it; his 1922 paper is a careful simplification of his previous 1920 paper. Contrary to what some have said (e.g. [13, foot-note p. 284]), Lindeberg did not borrow from Liapounov the "ingenious artifice" of comparing each $\alpha_i$ (in the notation of Theorem 1) with $N(0; a_i^2)$; this very important " artifice" seems to be due originally to Liapounov but it was tacitly used by Lindeberg already in his 1920 paper before he was aware of Liapounov's paper. An account of Liapounov's original work can be found in [13, p. 284]; Liapounov's work also contained a good estimation of the rate of convergence which was then subsequently improved by Cramér before ending up with the Berry-Esséen result mentioned before. See [7, p. 201].

The Finnish mathematician Lindeberg's early studies were entirely in the area of partial differential equations, calculus of variations and complex functions. It was only after 1918 that he began to work in probability and statistics; after his two remarkable papers on the central limit theorem, he worked for the rest of his life (until 1932) on the derivation of the properties of the distributions of various statistics which appear in sampling theory. More information on Lindeberg can be obtained from Elfving [5, pp. 153–161].

Lindeberg's 1922 paper [10] contains other equivalent formulations of his condition but not the one which appears in standard text-books these days; the latter seems to have been introduced explicitly by Feller in 1935 and has been used almost universally since then.

Lindeberg's 1922 paper contains five explicitly formulated theorems, each stated in finitary terms. To give a flavour of them, let us state his first theorem almost the way Lindeberg formulated it. Let $x_1, \dots, x_n$ be $n$ independent real-valued random variables with $\mathbb{E}x_i = 0$, $\mathbb{E}x_j^2 = a_j^2$, $a_1^2 + \cdots + a_n^2 = 1$, let $U(x) = \mathbb{P}\{x_1 + \cdots + x_n \leqslant x\}$, $\Phi(x) = \int_{-\infty}^{x} (1/\sqrt{2\pi}) \mathrm{e}^{-t^2/2}\, \mathrm{d}t$; then given $\varepsilon > 0$ there exists $\eta > 0$ such that for all $x$:

$$|U(x) - \Phi(x)| < \varepsilon$$

as soon as

$$\sum_{j=1}^{n} \mathbb{E}|x_j|^3 < \eta.$$

Lindeberg does not use the expectation symbol, replacing it instead by a Stieltjes integral; there are other similar formal differences between the above formulation and Lindeberg's. Lindeberg could have stated his first theorem as the statement

$$|U(x) - \Phi(x)| < C \left\{ \sum_{j=1}^{n} \mathbb{E}|x_j|^3 \right\}^{1/4},$$

with $C = 3$ since this is what he had actually proved just before stating his theorem. Note that this is equivalent to Theorems 1 and 1′ (with a different constant) and it immediately gives Liapounov's central limit theorem.

Lindeberg's second theorem is a simple deduction from his first theorem; here he supposes $|x_j| \leqslant d_n$, $1 \leqslant j \leqslant n$, (bounded random variables) with $\mathbb{E}x_j = 0$, $\mathbb{E}x_j^2 = a_j^2$, $b_n^2 = a_1^2 + \cdots + a_n^2$, $x_n$'s independent; then (with notation as above) for any $\varepsilon > 0$ there exists $\eta > 0$ such that

$$\left| U(x) - \int_{-\infty}^{\infty} \frac{1}{b_n \sqrt{2\pi}} \mathrm{e}^{-t^2/2b_n^2}\, \mathrm{d}t \right| < \varepsilon$$

as soon as $(d_n/b_n) < \eta$. Lindeberg seems to have considered this theorem as essential for mathematical statistics; he says so in the introduction to his paper. Now comes Lindeberg's third and most general theorem: given $\varepsilon > 0$ there exists $\eta > 0$ such that

$$|U(x) - \Phi(x)| < \varepsilon$$

as soon as $\sum_{j=1}^{n} \mathbb{E}s(x_j) < \eta$ under the hypothesis that $a_1^2 + \cdots + a_n^2 = 1$, other notations being held fixed as before. Here again, he could have stated his theorem in the form

$$|U(x) - \Phi(x)| < C \left\{ \sum_{j=1}^{n} \mathbb{E}s(x_j) \right\}^{1/4},$$

(where $0 < C < \infty$ is some absolute constant) since this is implicitly contained in his work. Of course, this is exactly what we have stated in Theorems 2 and 2′. Lindeberg's fourth theorem replaces his condition based on $\sum_{j=1}^{n} \mathbb{E}s(x_j)$ by an equivalent one (different from Feller's) and his fifth and last theorem is a mere renormalization of the fourth in case $\mathbb{E}x_j \neq 0$ and $\sum_{j=1}^{n} \mathbb{E}x_j^2 \neq 1$.

Curiously, contrary to what some authors have affirmed, Lindeberg does not state the most useful case of his general theorem where the random variables $x_j$ are independent and *identically* distributed with $\mathbb{E}x_1 = 0$, $\mathbb{E}x_1^2 = \sigma^2$, $0 < \sigma < \infty$; as is easy to show, in this case Lindeberg's condition is verified (most easily in the form $L_{n,\varepsilon} \to 0$ as $n \to \infty$, for all $\varepsilon > 0$) and

$$\mathscr{L}\left(\frac{x_1 + \cdots + x_n}{\sigma\sqrt{n}}\right) \to N(0, 1).$$

The observation that the central limit theorem holds in this very simple form in the independent, identically distributed case seems to be due to Lévy. Lévy's work is independent of Lindeberg's and almost contemporaneous; it uses characteristic functions systematically, a theory which Lévy himself had developed for the purpose of studying limit theorems concerning sums of independent random variables. Lévy goes on to study many deeper theorems concerning these sums; however, Lindeberg's condition and method remain Lindeberg's entirely original contribution. For references to Lévy's work and further developments one should consult the classic book of Gnedenko and Kolmogorov [7].

We have given a detailed summary of Lindeberg's work in order to clarify its exact nature; its contents are often represented in an incomplete and misleading manner. The Lindeberg method, presented by various text-books (basing themselves essentially on an article of Trotter from 1959, see e.g. [6]), remains an useful analytical approach to the central limit theorem, even when the presentation does not completely reflect its full power. As shown by the work reported in [12], it can be used fruitfully for further investigations. Dalang [4] uses Lindeberg's method in an elementary way in proving the central limit theorem in the special case of real-valued random variables which are independent and identically distributed.

The German mathematician Hausdorff (1868–1942) is best known for his work in set theory and topology although he himself has gone on record as having declared himself to be an analyst as well. His magnum opus is "Grundzüge der Mengenlehre" (1914; [9, vol. 2] is a recent annotated edition); besides introducing the general notion of "Hausdorff spaces", it contains an enormous amount of material on abstract set theory, topology of $\mathbb{R}^n$ (for example, a full proof of the Jordan curve theorem), much real analysis and Lebesgue integration theory. The book contains the first modern (and correct) proof of Borel's law of normal numbers. This major book of Hausdorff was never translated into English; what exists in English translation (under the title "Set theory") concerns a revised and enlarged edition (from 1927) of only parts of the material contained in his Grundzüge. Besides these books, he published 41 mathematical papers (all in German) containing such pearls as the Baker–Campbell–Hausdorff formula, Hausdorff paradox, Hausdorff dimension, Hausdorff summability, Hausdorff moment problem, Hausdorff-Toeplitz theorem and much else; the items named here can all be found in vol. 4 of his collected works [9].

Hausdorff also left behind some 26,000 pages of manuscript covering much of the mathematics of the first third of the 20th century. There are several manuscripts on probability theory of which some have been reproduced in vol. 5 of [9]; the Lecture Notes in which Theorem 1 was found will appear in this volume. Hausdorff actually formulates Theorem 1 rather in the form of our Theorem 1′, using the terminology of random variables. It is to be remembered that the definition of a random variable was never given explicitly until the appearance of Kolmogorov's 1933 axiomatization of probability theory; indeed, even eminent writers like Lévy, Fréchet and others continued to shy away from an exact definition of a random variable as late as 1950. For these authors (as for Hausdorff in 1923) a real random variable $x$ is something with which one associates a probability measure $\alpha$ in $\mathbb{R}$ and all calculations and reasonings concerning $x$ are then done with $\alpha$. For more remarks on this, see [7] and our comments in vol. 5 of [9]. Except for this reservation, we have found Hausdorff's writings (including all his manuscripts) whether in probability theory or elsewhere exemplary for their rigour and clarity; there may be occasional slips but the desire to present matters as succinctly and precisely as possible seem evident in all his writings.

Hausdorff's collected works have been planned to be published in 9 volumes (of which four have already appeared); these will include his considerable philosophical and literary production (2 volumes) and his far-flung and interesting correspondence (vol. 9). Vol. 1 should contain a detailed biography; of this we note only the following tragic element: persecuted by the Nazi regime, he was driven to commit suicide along with his wife and sister-in-law. His only daughter lived until 1991 and with the help of others, managed to preserve all the scientific Nachlass of Hausdorff. Awaiting the appearance of vol. 1 of the collected works [9], readers can find much of interest about Hausdorff in [2].

## Appendix. May 2006

A referee has reminded us of two useful historical references:

(i) S.L. Zabell, Alan turing and the central limit theorem, Amer. Math. Monthly 102 (1995) 483–94.

(ii) L. LeCam, The central limit theorem around 1935, Statist. Sci. 1 (1986) 78–91 (followed by comments of Trotter, Doob, Pollard).

In (i), Zabell points out that Alan Turing (in an unpublished 1934 manuscript, his fellowship dissertation, which I have not seen) gives a proof of a version of the Central Limit Theorem along lines similar to that of Lindeberg. In (ii), LeCam, amongst other things, mentions the possibility of extending Lindeberg's method to infinite-dimensional vector spaces. However, neither (i) nor (ii) states Lindeberg's original condition.

It is clear from Theorems 1 and 2 that the limit measure $v$ which appears there must be Normal. A trivial way to see this is to take each $\alpha_j$ appearing there to be Normal. A deeper observation is that since Theorem 2 implies its validity for the case of independent and identically distributed random variables, the limit measure $v$ must be a stable probability measure having finite variance, whence it must be Normal; for an explanation of this last statement, see Chapter 7 of [7]. I do not feel that Lindeberg's method, as presented here, can give any deeper characterization of the Normal distribution.

## References

[1] V. Bentkus, F. Götze, V. Paulaskas, A. Rackauskas, The accuracy of Gaussian approximation in Banach spaces, in: Yu.V. Prokhorov, V. Statulevicius (Eds.), Limit Theorems of Probability Theory, Springer, Berlin, 2000, pp. 25–111.
[2] E. Brieskorn (Hrsg.), Felix Hausdorff zum Gedächtnis, Band I, Vieweg, Wiesbaden, 1996.
[3] Y.S. Chow, H. Teicher, Probability Theory, second ed., Springer, Berlin, 1988.

[4] R. Dalang, Une démonstration élémentaire du théorème central limite, Elemente der Mathematik 61 (2006) 65–73.
[5] G. Elfving, The History of Mathematics in Finland 1828–1918, Societas Scientiarum Fennica, Helsinki, 1981.
[6] W. Feller, An Introduction to Probability Theory and its Applications, vol. II, Wiley, London, 1966.
[7] B.V. Gnedenko, A.N. Kolmogorov, Limit Distributions for Sums of Independent Random Variables (translated from Russian by K.L. Chung), Addison-Wesley, Reading, MA, 1954.
[8] A. Hald, The History of Mathematical Statistics from 1750 to 1930, Wiley, London, 1998.
[9] F. Hausdorff, Gesammelte Werke 9 vols. (vol. 2, 2002; vol. 4, 2000; vol. 7, 2004; vol. 5, 2006), Springer, Berlin, 2000.
[10] J.W. Lindeberg, Eine neue Herleitung des Exponentialgesetzes in der Wahscheinlichkeitsrechnung, Math. Zeit. 15 (1922) 211–225.
[11] V.V. Petrov, Classical-type limit theorems for sums of independent random variables, in: Yu.V. Prokhorov, V. Statulevicius (Eds.), Limit Theorems of Probability Theory, Springer, Berlin, 2000, pp. 1–24.
[12] Yu.V. Prokhorov, V. Statulevicius (Eds.), Limit Theorems of Probability Theory, Springer, Berlin, 2000.
[13] J.V. Uspensky, Introduction to Mathematical Probability, McGraw-Hill, London, 1937.