A PROOF USING DERIVED DISTRIBUTIONS OF THE LINDEBERG CENTRAL LIMIT THEOREM

by

Walter L. Smith

Department of Statistics
University of North Carolina
Chapel Hill, North Carolina, 27514

Key Words: Central Limit Theorem, Characteristic Functions.

This research is supported by Office of Naval Research Grant No. N00014-83-K-0352.

SUMMARY

A short discussion of the ideas of "derived" distributions precedes a proof of the Lindeberg Central Limit Theorem. This proof is short, carries the necessity and sufficiency parts together, and renders their relationship clear. The derived distribution approach is then utilized to construct some illustrative examples of behavior differing from the familiar normal limit which occurs when the Lindeberg conditions are fully satisfied.

¶1. **Introduction.** The Lindeberg Central Limit Theorem is one of a few theorems that should, and typically do, constitute the major landmarks, as it were, of a serious first graduate course in Probability Theory. Most of the published proofs of this important theorem involve Fourier analysis in that they employ characteristic functions. Two noteworthy exceptions are the proofs of Trotter (1959) and (for the somewhat restricted case of lattice random variables, identically distributed, and therefore hardly meriting to be referred to as the "Lindeberg" theorem) due to Petrovsky and Kolmogorov (given in Khinchin (1948) and presented very clearly, in English, in Rosenblatt (1974)). However, the majority of the published proofs we have seen, including those in a number of much-used "standard" texts, essentially involve two distinct proofs, each somewhat intricate; one deals with the sufficiency part of the proof, the other with the necessity part. It is many years since I introduced into the first graduate course on probability theory at Chapel Hill a proof of the Lindeberg Theorem which makes use of so-called "derived" distributions; this proof is strikingly straightforward, the necessity and the sufficiency parts of the theorem are established simultaneously and their inter-relationship is rendered transparent. Professional colleagues at various academic institutions have long urged publication of this proof because it could, at the very least, be useful to others faced with the teaching of the Lindeberg Theorem, and more than that, it may provide, even to those not engaged in teaching, a new insight into that theorem. Having recently found a way to avoid one unattractive feature of the "Chapel Hill" proof as it has been taught for these many years (the use of the logarithm with all the ambiguities raised by that multi-valued function), I am finally persuaded to present this proof to a wider audience.

¶2. **Conditions for the theorem.** Let X_1, X_2, ... be an infinite sequnce of mutually independent random variables with finite means and variances. With no loss of generality and some gain in ease of discussion we may suppose that $\mathcal{E}X_n = 0$ for all n; we shall write σ_n^2 for the variance of X_n and set

$$s_n^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2 \tag{2.1}$$

for the variance of $S_n = X_1 + X_2 + \cdots + X_n$. Let us also write $F_n(x) = P\{X_n \leq x\}$ for the distribution function of X_n. Finally make the assumption of asymptotic negligibility:

$$\left\{ \max_{1 \leq j \leq n} \frac{\sigma_j^2}{s_n^2} \right\} \to 0, \text{ as } n \to \infty. \tag{2.2}$$

Let us refer to the all the assumptions described in the preceding paragraph as the Global Hypothesis. We may now introduce the famous Lindeberg condition:

Condition (LC): For every fixed $\epsilon > 0$, as $n \to \infty$,

$$\frac{1}{s_n^2} \sum_{j=1}^{n} \int_{|x| > \epsilon s_n} |x|^2 \, F_j(dx) \to 0 \tag{2.3}$$

Lindeberg (1922) established the following:

THEOREM I *If the global hypothesis be taken to hold, then, in order for S_n/s_n to be asymptotically $\mathcal{N}(0,1)$ as $n \to \infty$ it is necessary and sufficient that Condition (LC) hold.*

¶3. Derived Distributions. The author introduced the so-called "derived" distributions in his Ph. D. thesis (Smith (1953)) and has used them since in various ways; see (Smith (1959)) and, for more general results (Smith (1966)). For completeness of the present account we present briefly the necessary basic ideas, which are extremely simple, in spite of their availability in much greater generality elsewhere.

Let $F(\cdot)$ be the distribution function of a non-negative random variable whose first two moments μ_1 and μ_2 are both finite. It is well-known that $\mu_1 = \int_0^\infty \{1 - F(x)\}\,dx$; thus we can define a probability density function on $(0,\infty)$ by

$$f_{(1)}(x) = \frac{\{1 - F(x)\}}{\mu_1} \tag{3.1}$$

and we shall call this the first derived density function. It is a trivial matter to show that this new density has its first moment finite and equal to $\mu_2/2\mu_1 = \mu_1^{(1)}$, say. Thus, if we write $F_{(1)}(\cdot)$ for the df associated with $f_{(1)}(\cdot)$ we can introduce a second derived density by

$$f_{(2)}(x) = \frac{\{1 - F_{(1)}(x)\}}{\mu_1^{(1)}} \tag{3.2}$$

It is a comparatively easy matter to verify the following equation which relates the tail probabilities of $f_{(2)}(\cdot)$ to the original distribtuion function $F(\cdot)$:

$$\int_x^\infty f_{(2)}(u)\,du = \frac{1}{\mu_2}\int_x^\infty (x - z)^2 F(dz)\,. \tag{3.3}$$

Now let us introduce characteristic functions: write, for dummy real θ,

$$\phi(\theta) = \int_{0-}^\infty e^{i\theta x} F(dx). \tag{3.4}$$

Then, in an obvious notation, one can obtain by an integration by parts the result

$$(3.5) \qquad \phi_{(1)}(\theta) = \frac{1 - \phi(\theta)}{-\mu_1 i\theta}$$

and, by the same token

$$(3.6) \qquad \phi_{(2)}(\theta) = \frac{1 - \phi_{(1)}(\theta)}{-\mu_1^{(1)} i\theta} .$$

If we combine (3.5) and (3.6) we are led to the result

$$(3.7) \qquad \phi(\theta) = 1 + \mu_1 i\theta + \tfrac{1}{2}\mu_2 (i\theta)^2 \phi_{(2)}(\theta) .$$

This equation (3.7) is the basis of our proof; notice that it is valid for <u>all</u> θ and not, as is the case with the more familiar Taylor expansions of a characteristic function at the origin, for "small" θ only. All that remains for us to do, before we turn to the actual proof of the Central limit Theorem proper, is to extract from (3.7), which refers to a distribution on the non-negative half-line, a similar result for a distribution on the whole real line. However, before we do that let us notice that $\phi_{(2)}(\theta)$ is the characteristic function of the pdf $f_{(2)}(\cdot)$. Moreover, $f_{(2)}(\cdot)$ is a pdf with special properties; it is (as a glance at (3.1) and (3.2) will reveal) non-increasing, with a non-increasing derivative which tends to zero as $x \to \infty$. Thus (3.7) is remarkable in that it connects the arbitrary cf $\phi(\theta)$ to the cf of such a special convex pdf.

Let X be an unrestricted (to the non-negative reals) random variable with zero mean and finite variance σ^2. Let us set $p = \mathrm{P}\{X \geq 0\}$ and q = $\mathrm{P}\{X < 0\}$. Then we can regard X as arising from two non-negative random variables X^+ and X^-: with probability p, $X = X^+$; with probabilty q, $X = -X^-$. In an obvious notation we have

$$\begin{aligned}\mu_1 &= p\mu_1^+ - q\mu_1^- \\ \mu_2 &= p\mu_2^+ + q\mu_2^- \\ \phi(\theta) &= p\phi^+(\theta) + q\phi^-(-\theta)\end{aligned}$$

and from equations like (3.7) for both $\phi^+(\theta)$ and $\phi^-(\theta)$ and can then derive the result, remembering that $\mu_1 = 0$,

(3.8) $$\phi(\theta) = 1 - \tfrac{1}{2}\sigma^2\psi(\theta)$$

in which the function $\psi(\cdot)$ <u>is</u> <u>a</u> <u>characteristic</u> <u>function</u>

(3.9) $$\psi(\theta) = p\phi^+_{(2)}(\theta) + q\phi^-_{(2)}(-\theta)$$

associated with a density $g(\cdot)$, say, which is convex and monotonically decreasing to zero as x increases to ∞ on the positive half-line, convex and monotonically decreasing to zero as x decreases to $-\infty$ on the negative half-line. This result (3.8) has been the object of this section; it is a special case of a more general result [Smith (1999)] (which might well be presented at some earlier point in a probability course, when dealing with properties of characteristic functions). It is important to note that, like (3.7), (3.9) is true for <u>all</u> real θ.

¶4. **<u>Proof of Lindeberg's Theorem.</u>** We must first prove a crucial lemma; the one given here differs from the one long utilised in the same context in Chapel Hill; this newer lemma is what enables us to avoid difficulties caused by ambiguities of the logarithm function.

LEMMA 1 *For* j = 1, 2, ..., n; n = 1, 2, ..., ad inf., *let* $\{a_{jn}\}$ *be complex numbers such that*

$$(4.1)\qquad \sum_{j=1}^{n} |a_{jn}| \leq A < \infty$$

and suppose further that

$$(4.2)\qquad \alpha_n =_{\text{def}} \left\{ \max_{1 \leq j \leq n} |a_{jn}| \right\} \to 0, \quad \text{as } n \to \infty.$$

Then, as n → ∞,

$$(4.3)\qquad \prod_{j=1}^{n} (1 - a_{jn}) \sim \exp\left\{ -\sum_{j=1}^{j=n} a_{jn} \right\}.$$

PROOF. Let

$$\frac{e^{-a_{jn}}}{(1 - a_{jn})} = 1 + \xi_{jn}, \text{ say.}$$

Then

$$\xi_{jn} = \frac{\frac{1}{2!}a_{jn}^2 + \frac{1}{3!}a_{jn}^3 + \cdots}{1 - a_{jn}}$$

so

$$|\xi_{jn}| \leq \frac{\alpha_n |a_{jn}|}{2(1 - \alpha_n)^2}.$$

We may thus conclude that

$$(4.4)\qquad \sum_{j=1}^{n} |\xi_{jn}| \to 0, \text{ as } n \to \infty.$$

But

$$\left|\prod_{j=1}^{n}(1+\xi_{jn})-1\right| \leq \prod_{j=1}^{n}(1+|\xi_{jn}|)-1$$

$$\leq e^{\sum_{j=1}^{j=n}|\xi_{jn}|}-1$$

which, in view of (4.4), tends to zero and proves the lemma.

Let us now turn to the proof of Theorem I. In the notation already established, let us define the characteristic function of S_n/s_n to be

$$\Phi_n(\theta) = \mathcal{E}\, e^{i\theta S_n/s_n}.$$

Then, because of the mutual independence of the $\{X_n\}$,

$$\Phi_n(\theta) = \prod_{j=1}^{n}\phi_j(\theta/s_j)$$

$$= \prod_{j=1}^{n}(1-a_{jn}), \text{ say,}$$

where we have set

$$a_{jn} = \tfrac{1}{2}\sigma_{jn}^2\left(\frac{\theta^2}{s_n^2}\right)\psi_j(\theta/s_n).$$

The question is then whether the $\{a_{jn}\}$ so defined satisfy the conditions of the lemma. Because the ψ_j are characteristic functions it follows that for <u>all</u> θ we must have $|\psi_j(\theta)| \leq 1$ and so (4.2) is satisfied as because of the asymptotic negligibility assumed as part of the global hypothesis. For every fixed θ we have

(4.5) $$|a_{jn}| \leq \tfrac{1}{2}\sigma_{jn}^2\left(\frac{\theta^2}{s_n^2}\right),$$

from which it is plain that that (4.2) is satisfied. But from (4.5) it is equally plain that

(4.6) $$\sum_{j=1}^{n} |a_{jn}| \leq \tfrac{1}{2}\theta^2 < \infty,$$

and so condtion (4.1) of Lemma 1 is also satisfied.

Notice that from the global hypothesis alone we must have that, as $n \to \infty$, for every fixed θ,

(4.7) $$\Phi_n(\theta) \sim \exp\left\{ -\tfrac{1}{2}\theta^2\, \Psi_n(\theta)\right\}$$

where we have conveniently written

(4.8) $$\Psi_n(\theta) = \sum_{j=1}^{n} \frac{\sigma_j^2}{s_n^2}\, \psi_j(\theta/s_n).$$

From (4.8) it can be seen <u>that</u> $\Psi_n(\theta)$ <u>is</u> <u>also</u> <u>a</u> <u>characteristic</u> <u>function</u>! Thus the global hypothesis of Lindeberg's Theorem implies the asymptotic relationship between two sequences of characteristic functions. Moreover, we see that $\Psi_n(\theta)$ is associated with a pdf with special convexity properties since it is a convex linear combination of densities like the generic $g(\cdot)$ *of* ¶3 (especially after equation (3.9)).

Let us write G_n for the df associated with $\Psi_n(\theta)$ and U for the degenerate df (sometimes called the "HeaViside Unit Function") which places all the unit probability on the origin. Then it is plain from (4.7) that

(4.9) $$\Phi_n(\theta) \to e^{-\frac{1}{2}\theta^2}$$

if and only if

$$(4.10) \qquad \Psi_n(\theta) \;\rightarrow\; 1.$$

From the Continuity Theorem for characteristic functions it can be inferred that (4.9) holds if and only if the sequence of distribution functions $\{G_n\}$ converges weakly to U. This is essentially completes the proof of Lindeberg's Theorem although, of course, something must be said to relate the condition "$G_n \underset{w}{\Rightarrow} U$", which the present argument shows to arise in a natural way as the crucial condition for asymptotic normality, and the classical Condition (LC).

From (3.3) coupled with (4.8) we see that the condition $G_n \underset{w}{\Rightarrow} U$ is equivalent to the statement

$$(4.11) \qquad \frac{1}{s_n^2} \sum_{j=1}^{n} \int_{|x| > \epsilon s_n} (|x| - \epsilon s_n)^2 \, F_j(dx) \;\rightarrow\; 0, \qquad \text{as } n \rightarrow \infty.$$

Obviously Condition (LC) implies (4.11). On the other hand, in the integrals in (4.11), for the range $|x| > 2\epsilon s_n$ it is true that $|x| - \epsilon s_n > \frac{1}{2}|x|$ from which it should be apparent that (4.11) implies (LC) (with the unimportant change of ϵ into 2ϵ in the latter condition).

¶5. **Further remarks** Let us write K_n for the df of S/s_n, with cf Φ_n, and let us continue to write G_n for the df associated with Ψ_n. The argument we have given in the preceding section proves the following theorem, more general that the Lindeberg Theorem:

THEOREM II *Under the global hypothesis of* ¶1 $\{K_n\}$ *converges weakly to a limit df* K, *say, if and only if* $\{G_n\}$ *converges weakly to a (possibly defective) df* G, *say. If* $\Psi(\cdot)$ *be the cf associated with* G, *then* K *will have a finite variance given by* $\Psi(0)$.

PROOF If K_n converges weakly to K then $\Phi_n(\theta)$ tends to $\exp\left\{-\frac{1}{2}\theta^2\Psi(\theta)\right\}$ where we must also have $\Psi_n(\theta) \rightarrow \Psi(\theta)$. The usual weak-compactness argument will then show that we must have G_n converging weakly to some, possibly defective, df G of which function Ψ is the Fourier-Stieljes Transform.

Conversely, if the G_n converge weakly to some G then the $\{\Psi_n(\theta)\}$ tend to a limit $\Psi(\theta)$, the transform of G. Thus $\Psi(\theta)$ must be uniformly continuous and $0 \leq \Psi(0) \leq 1$. Thus $\Phi_n(\theta) \to \Phi(\theta)$ where the limit function must also be uniformly continuous and $\Phi(0) = 1$; the latter claim may easily be understood by letting $\theta \to 0$ in the equation $\Phi(\theta) = \exp\left\{-\frac{1}{2}\theta^2\Psi(\theta)\right\}$. From this result it follows that K must always be a proper distribution function,

It is well-known, and easily proved, that if

$$\liminf_{\theta=0} \frac{1 - \Re\Phi(\theta)}{\theta^2} < \infty$$

then K, the df associated with Φ, must have a finite variance; and hence, from a representation like (3.8), but possibly including (for argument's sake) a term corresponding to a first moment, it follows that K must in fact have a zero first moment and a finite variance given by:

$$\kappa^2 = \lim_{\theta=0} \frac{1 - \Phi(\theta)}{\frac{1}{2}\theta^2}.$$

Thus the claim $\kappa^2 = \Psi(0)$ also follows from equation $\Phi(\theta) = \exp\left\{-\frac{1}{2}\theta^2\Psi(\theta)\right\}$.

The approach to the Central Limit Theorem by means of the derived distributions makes it easy to construct examples of behavior differing from the familiar convergence predicted by the Lindeberg Theorem. We need two simple lemmas, however.

LEMMA 2 *Let* $g(x)$ *be an absolutely continuous function in* $\mathcal{L}_1(0,\infty)$ *such that both* $g(x)$ *and* $|g'(x)|$ *are non-increasing, and both tend to zero as* $x \to \infty$. *Also asume that* $|g'(0)| \leq 1$.

Then the distribution function $F(x;g) \equiv 1 - |g'(x)|$ *on* $[0,\infty)$ *has its first* *two moments finite and given by*

$$\mu_1(g) = g(0); \quad \mu_2(g) = 2\int_0^\infty g(x)\,dx$$

and a second derived pdf $f_{(2)}(x)$ *such that*

$$\tfrac{1}{2}\mu_2(g)\, f_{(2)}(x) = g(x).$$

The proof of Lemma 2 is straightforward, in view of what has gone earlier in this note, and so we do not give details. Even simpler to prove is the following:

LEMMA 3 *If* $g(x)$ *satisfies the conditions of* Lemma 2 *and* $a > 0$ *is any constant, then* $g_a(x) \equiv ag(ax)$ *also satisfies those conditions, and* $\mu_2(g_a) = \mu_2(g)$ *for all* $a > 0$.

Furthermore, suppose g_1 *and* g_2 *both satisfy the conditions of* Lemma 2. *Then, for any* $0 \le p,q \le 1$ *such that* $p+q=1$, $pg_1 + qg_2$ *also satisfies* Lemma 2, *and*

$$\mu_2(pg_1 + qg_2) = p\mu_2(g_1) + q\mu_2(g_2).$$

Plainly, if $g(x)$ on $(0,\infty)$ satisfies Lemma 2, we can use $g(|x|)$ to define $\sigma^2 f_{(2)}(x)$, where $f_{(2)}(x)$ is the second derived pdf of some symmetric pdf $f(x)$ on $(-\infty, +\infty)$. We have written σ^2 for the variance of this pdf $f(x)$. Let us also write $\sigma^2\psi(\theta)$ for the Fourier Transform of $g(|x|)$; thus $\psi(\theta)$ is the characteristic function of the second derived pdf. We have, by Lemma 3, the equation $\sigma^2 = 2\int_0^\infty g(x)\,dx$, and note that this variance is unchanged by scale changes made upon g.

<u>EXAMPLE 1</u> Let $g(|x|)$ and $\psi(\theta)$ be the functions introduced in the preceding paragraph, associated with a pdf $f(x)$ with zero mean and variance σ^2. Then, for $j = 1, 2, \ldots$ we can define the characteristic functions

$$\phi_j(\theta) = 1 - \tfrac{1}{2}\sigma^2\theta^2\psi(\theta\sqrt{j})$$

of random variables with zero means and variances all equal to σ^2. For this example we find

$$\Psi_n(\theta) = \frac{1}{n}\sum_{j=1}^{n} \psi(\theta\sqrt{j}/\sqrt{n}).$$

One can deduce from this that as $n \to \infty$

$$\Psi_n(\theta) \to \int_0^1 \psi(\theta\sqrt{u})\, du = \Psi(\theta), \text{ say.}$$

The limit cf Ψ is associated with the pdf

$$\frac{1}{\sigma^2}\int_0^1 \frac{g(|x|/\sqrt{u})}{\sqrt{u}}\, du$$

Thus we have a situation in which the global hypothesis is satisfied but the condition (LC) is not: yet there is still convergence, though not to a normal limit. The characteristic function of the limiting distribution is

$$\exp\left\{-\tfrac{1}{2}\theta^2\int_0^1 \psi(\theta\sqrt{u})\, du\right\}.$$

It may be verified from this last formula that the non-normal limit K still has unit variance.

EXAMPLE 2 We use Lemma 3. Pick $0 < \varpi < 1$ and set,

$$\psi_j(\theta) = \varpi\psi(\theta) + (1-\varpi)\psi(\theta j)$$

where ψ is the characteristic function introduced in the previous example. In this case $\Psi_n(\theta) \to \varpi$. To see this let us begin by noting that, since $\psi(0)=1$,

$$\frac{1}{n}\sum_{j=1}^{n} \psi(\theta/\sqrt{n}) \to 1 \qquad \text{as } n \to \infty.$$

Further, by the Riemann-Lebesgue lemma, $\psi(\theta) \rightarrow 0$ as $|\theta| \rightarrow \infty$. Therefore, given a small $\epsilon > 0$ we can find $\Delta(\epsilon)$ such that $|\psi(\theta)| < \epsilon$ for all $|\theta| > \Delta$. From this it is possible to show that for any fixed θ,

$$\left| \frac{1}{n} \sum_{j=1}^{n} \psi(\theta j/\sqrt{n}) \right| \leq \frac{\Delta\sqrt{n}}{n|\theta|} + \epsilon.$$

Since ϵ is arbitrary the claim is established. It follows that for this example the limit distribution is normal, but the variance is strictly less than one, depending on our choice of ϖ. There is an example of this phenomenon in Feller (1971).

I am indebted to Dr. Andrew Rosalsky for drawing my attention to the interesting and highly relevant paper of H. F. Trotter.

REFERENCES

W. Feller (1971), *An Introduction to Probability Theory and Its Applications*, 2d Edition, New York, John Wiley & Sons.

A. I. Khinchin (1948), *Asymptotische Gesetze des Wahrscheinlichkeitsrechnung*, New York: Springer.

M. Rosenblatt (1974), *Random Processes*, Second Edition, New York: Springer

W. L. Smith (1953), *Stochastic Sequences of Events*, Ph. D. Thesis, Cambridge University.

———————— (1959), On the cumulants of renewal processes, *Biometrika*, **46**, 1-29.

———————— (1966), A theorem on functions of characteristic functions and its application to some renewal theoretic random walk problems, *Proc. Fifth Berkeley Symposium*, Vol. II, Pt. 2, Berkeley, California: University of California Press.

H. F. Trotter (1959), An elementary proof of the Central Limit Theorem, *Archiv der Mathematik*, **10**, 226-234.