I. Introduction

This is about the classical Lindeberg-Lévy-Feller Central limit theorem for independent random variables. One considers a double array \( \{X_{n,j}; j=1,2,\ldots, k_n; n=1,2,\ldots\} \) with the negligibility assumption

\[
(1) \quad \sup_j \mathbb{P}\{|X_{n,j}| > \varepsilon\} \to 0
\]

for each \( \varepsilon > 0 \) as \( n \to \infty \).

Lindeberg (1920) (1922) gave sufficient conditions for the convergence of the distribution of \( S_n = \sum X_{n,j} \) to Gaussian \( \mathcal{N}(0,1) \). His proof, relying only on a simple Taylor expansion argument, is very elementary.

The converse theorem was proved, in a particular case, by Lévy in 1935. Lévy's proof is also "elementary" in the sense that no Fourier transforms are used. For the particular case of normed sums, Feller (1935) gave a proof that relies heavily on Fourier transforms. According to Lévy (1937) Khintchin had an elementary proof.

The present paper gives a proof that does not use Fourier transforms in any manner, even where their use would be effective and simple.

2. A Basic Inequality

The present section contains a basic inequality on fluctuations of sums of independent variables in terms of second and fourth moments.

Let the \( X_{n,j} \) be as described and assume in addition to the negligibility hypothesis (1) that

\[
(2) \quad S_n = \sum_j X_{n,j} \quad \text{has a distribution that converges to } \mathcal{L}(\xi) \quad \text{where}
\]

\[
E\xi = 0, \quad E\xi^2 = 1, \quad \text{and} \quad E\xi^4 = 3.
\]
The problem is to show that:

\[ (3) \quad P \left\{ \sup_{j} |Y_{n_j}| > \varepsilon \right\} \to 0 \quad \text{as} \quad n \to \infty. \]

For the square integrable case with convergence of second moments, the usual Lindeberg condition follows easily from (3).

Let \( \{X'_{n_j}; j=1,2,\ldots; n, n=1,2,\ldots\} \) be an independent copy of the \( X_{n_j} \). Consider \n
\[ Y_{n_j} = \frac{1}{\sqrt{2}} (X_{n_j} - X'_{n_j}). \]

Then the \( Y_{n_j} \) ratio—by the negligibility condition (1) and the limit requirement (2)—

It is sufficient to prove (3) for the \( Y_{n_j} \). Indeed, according to Lévy's symmetrization inequalities, if \( m_{n_j} \) is a median of \( X_{n_j} \) one has

\[ P \left\{ \sup_{j} |X_{n_j} - m_{n_j}| > \varepsilon \right\} \leq 2 P \left\{ \sup_{j} |Y_{n_j}| > \varepsilon \right\} . \]

Hence, by (1)

\[ P \left\{ \sup_{j} |X_{n_j}| > 3 \varepsilon \right\} \leq 2 P \left\{ \sup_{j} |Y_{n_j}| > \varepsilon \right\} + 2 \sup_{j} P \{ |X_{n_j}| > \varepsilon \} \]

whenever the last term on the right is less than unity, hence always.

Now take a fixed integer \( m \) and partition the sum \( S_n = \sum Y_{n_j} \) into \( m \) parts \( S_n(1), S_n(2), \ldots, S_n(m) \) with

\[ S_n(1) = \sum_{j} Y_{n_j}; k_i \leq j < k_{i+1}, n \}

Taking subsequences, if necessary, assume that

\[ \mathcal{L}[S_n(1), i=1, \ldots, m] \to \mathcal{L}[Z_i, i=1, \ldots, m] \]
Then \( \sum_{i} E Z_{i} = 0 \), \( E(\sum_{i} Z_{i})^{2} = 1 \) and \( E(\sum_{i} Z_{i})^{4} = 3 \).

Note that \( \mathcal{L}(Z_{i}) = \mathcal{L}(-Z_{i}) \) by the symmetry of the \( Y_{nj} \). Hence \( E Z_{i} = 0 \) for all \( i \). Furthermore, \( E Z_{i}^{2} = v_{i} < \infty \) and \( E Z_{i}^{4} < \infty \).

**Lemma 1.** Under the conditions (1) (2) the variables \( Z_{i} \) satisfy the relations

\[
\sum_{i} E Z_{i}^{4} = 3 \sum v_{i}^{2},
\]

\[
\sum_{i} P(|Z_{i}| > \varepsilon) \leq \frac{3 \sum v_{i}^{2}}{\varepsilon^{4}}.
\]

Hence, for \( n \) sufficiently large,

\[
\sum_{i} P(|S_{n}(i)| > 2\varepsilon) \leq \varepsilon + \frac{3 \sum v_{i}^{2}}{\varepsilon^{4}}.
\]

**Proof.** Statement (5) follows from (4) by Markov's inequality. Statement (6) follows then from \( \mathcal{L}[S_{n}(1)] = \mathcal{L}(Z_{1}) \). Thus it is enough to prove (4).

Let \( W(k) = \sum_{i<k} Z_{i} \). Then

\[
E[W(k) + Z_{k}]^{4} = E[W(k)]^{4} + 6 E[W(k)]^{2} E Z_{k}^{2} + E Z_{k}^{4}.
\]

Thus, summing up and writing \( v_{i} = E Z_{i}^{2} \) one has

\[
3 = E(\sum_{i} Z_{i})^{4} = \sum_{i} E Z_{i}^{4} + 6 \sum_{i} v_{i+1}^{2} \sum_{j \leq i} v_{j}^{2} = \sum_{i} E Z_{i}^{4} + 3\{(E v_{i})^{2} - (E v_{i})^{2}\}.
\]

Now \( \sum v_{i} = E(\sum_{i} Z_{i})^{2} = 1 \). Thus \( \sum_{i} E Z_{i}^{4} = 3 \sum v_{i}^{2} \). Hence the result.
It is clear that statement (6) in Lemma 1 is a step towards (3). In fact to get (3) from it, it will be sufficient to (a) show that the \( E \nu_j^2 \) can be made small and (b) apply once more Paul Lévy symmetrization inequalities in the form

\[
(7) \quad P\{ \sup_s \{ E [Y_{nj}^2; k_{ij} < s < s < k_{i+1}^+] \} \} \leq 2 P\{ |S_n(1)| > \varepsilon \}.
\]

We shall now proceed to show that \( E \nu_j^2 \) can be made arbitrarily small, first in the square integrable case and then in the general case.

3. The Square Integrable Case

Suppose that the \( Y_{nj}^2 \) are as in Section 2, but assume further that \( \sigma_{nj}^2 = E Y_{nj}^2 \) exists and that

\[
(8) \quad \sup_n \Sigma \sigma_{nj}^2 = c < \infty \quad \text{and} \quad \beta_n = \sup_j \sigma_{nj}^2 > 0.
\]

In this case one can take for \( k_{in} \) the first integer such that

\[
E \sigma_{nj}^2; j < k_{in} \geq \frac{\sigma_n^2}{m} c.
\]

The variance of the sum \( S_n(1) \) is then at most \( \frac{c}{m} + \beta_n \). Thus the variance \( v_1 \) of the limit variables \( Z_i \) does not exceed \( \frac{c}{m} \) and \( \Sigma \nu_j^2 \leq \frac{k}{m} \). Therefore the assertion (3) follows for this case.

4. The General Case

Suppose again that the \( Y_{nj}^2 \) are as in Section 2 with \( \mathcal{D}(\xi_{nj}) \Rightarrow \mathcal{D}(\xi) \),

\[
E \xi = 0, \quad E \xi^2 = 1, \quad E \xi^4 = 3.
\]

We claim first that

\[
(9) \quad \text{For } \varepsilon > 0 \text{ there is an } n(\varepsilon) \text{ and a } b(\varepsilon) \text{ such that } n > n(\varepsilon) \text{ implies } \sum_j P\{ |Y_{nj}| > b(\varepsilon) \} < \varepsilon.
\]
For any fixed number $a$

$$\sup_n \Sigma E(\{Y_{nj}|^2 I|Y_{nj}|<a\}_{j}) < \infty.$$ 

To prove statement (9) note again by Lévy's inequalities

$$P(\sup_s [Y_{nj}; j<s]) > b \leq 2 P(\{|S_n| > b\})$$

with $S_n = \Sigma Y_{nj}; j \leq k_n$. Thus

$$P(\sup_j |Y_{nj}| > 2b \leq 2 P(\{|S_n| > b\}).$$

This gives (9).

To prove (10), assume the contrary. Then taking a subsequence if necessary one can assume that, for some fixed $a$,

$$\Sigma E Y_{nj}^2 I(|Y_{nj}| < a) \geq b_n^4$$

with $b_n \to \infty$. Thus we can assume $b_n > a$.

Consider then the variables

$$V_{nj} = \frac{1}{b_n} Y_{nj} I(|Y_{nj}| < b_n)$$

They are bounded by unity. The sum of their variances is larger than $b_n^2$. Thus, by the direct Lindeberg theorem $\mathcal{L}(\Sigma V_{nj}) \to \mathcal{N}(0,1)$. However, by (1) (2) and statement (9) $\mathcal{L}(b_n \Sigma V_{nj}) \to \mathcal{L}(\xi)$ with $E \xi^2 = 1$. This is a contradiction. Thus (10) must hold.

Now fix a value of $a$, say $a = 1$. Let

$$S_{nj}^2 = E Y_{nj}^2 I(|Y_{nj}| < a)$$

and

$$\alpha_{nj} = P(|Y_{nj}| > a).$$
Partition the total sum as before, considering integers \( k_{in} \) such that

\[
\sum_{ \{ s_{nj} + a_n \} } \leq j < k_{i+1,n} \leq \frac{c}{m}
\]

and let

\[
S_n(1) = \sum_{ \{ Y_{nj} ; k_{in} \leq j < k_{i+1,n} \} }
\]

Taking subsequence if necessary one can assume that

\[
\mathcal{L}[S_n(1)] \rightarrow \mathcal{L}(Z_1)
\]

Since \( \mathcal{L}(\Sigma Z_1) = \mathcal{L}(\zeta) \) with \( E\zeta I = 3 \), one must also have \( E Z_1 I \leq 3 \).

However, by Chebyshev

\[
P\{\sum_{ \{ Y_{nj} I[Y_{nj} \leq a] ; k_{in} \leq j < k_{i+1,n} \} } \geq \tau \} \leq \frac{c}{m} \frac{1}{\sigma^2 \tau^2}
\]

Also

\[
P\{ \sum_{ \{ Y_{nj} I[Y_{nj} > a]\neq 0 \} }\} \leq \sum_{ \{ a_j ; k_{in} \leq j < k_{i+1,n} \} }
\]

Hence

\[
P\{|S_n(1)| > \tau\} \leq \frac{2rc}{m} \frac{1}{\tau^2}
\]

Hence

\[
P\{|Z_1| > \tau\} \leq \frac{2rc}{m} \frac{1}{\tau^2}
\]

Also,

\[
E Z_1^2 \leq E Z_1^2 I[|Z_1| \leq \tau] + \frac{1}{\tau^2} E Z_1^4 I[|Z_1| > \tau]
\]

\[
\leq \tau^2 + \tau^2 P\{|Z_1| > \tau\} + \frac{1}{\tau^2} E Z_1^4
\]

If one takes

\[
t^2 = \frac{3}{\varepsilon} , \quad \tau^2 = \varepsilon , \quad \text{and} \quad \frac{6c}{m} \leq \varepsilon^2 \tau^2 \leq \varepsilon^2
\]

this becomes at most \( 3\varepsilon \).

Thus the partitioning of Section 2 is still possible, with \( \Sigma Z_1^2 \) as small as desired.
5. **Additional Comments on the General Case**

Once can give a different proof using an argument of Lévy as follows. The variables $y_{nj}$ have the same distribution as

$$y_{nj}^* = (1-y_{nj})U_{nj} + \xi_{nj}v_{nj} + (y_{nj}-\xi_{nj})U_{nj}$$

where the $\xi, y, U$ and $V$ are all independent and where

$$(12) \quad \xi_{nj} = \xi_{nj}^2, \quad y_{nj} = y_{nj}^2,$$

$$\alpha_{nj} = P[|y_{nj}|=1] = P[|\xi_{nj}|=1] = P[|y_{nj}|>a]$$

$$(13) \quad \mathcal{L}(U_{nj}) = \mathcal{L}(Y_{nj} \mid |Y_{nj}|<a)$$

$$(14) \quad \mathcal{L}(V_{nj}) = \mathcal{L}(Y_{nj} \mid |Y_{nj}|>a).$$

Here, $\alpha_n = \sup \alpha_{nj} + 0$. The variance of $\Sigma(y_{nj}-\xi_{nj})U_{nj}$ is at most $2\alpha_n(1-\alpha_n)\Sigma E U_{nj}^2$.

That of $\Sigma(1-y_{nj})U_{nj}$ is $\Sigma(1-\alpha_n)E U_{nj}^2 \geq (1-\alpha)\Sigma E U_{nj}^2$.

Thus, the term $\Sigma(y_{nj}-\xi_{nj})U_{nj}$ is always negligible compared to the rest. Indeed, if $\Sigma E U_{nj}^2$ stays bounded that term tends to zero in quadratic mean.

If $\text{var} \Sigma(1-y_{nj})U_{nj} = \sigma_n^2$ tends to infinity, the term $\Sigma(1-y_{nj})U_{nj}$ has a distribution close to $\mathcal{N}(0,\sigma_n^2)$ in Kolmogorov distance (by the CLT). Thus

$$P\{x < \Sigma[(1-y_{nj})U_{nj} + \xi_{nj}v_{nj}] \leq x + \tau \sigma_n\}$$

will eventually be smaller than $e + \frac{\tau}{\sqrt{\sigma_n}}$. By Chebyshev

$$P\{\Sigma(y_{nj}-\xi_{nj})U_{nj} \geq \tau \sigma_n\} \leq \frac{\alpha_n}{\tau^2}.$$

However, for any two variables $X$ and $Y$ (not necessarily independent)
\[ P[X+Y \leq t] - P[X \leq t] \leq P[|Y| > r] + \sup_x P[x \leq X \leq x+r] . \]

Thus, removing the cross term \( \Sigma(n_j - \xi_j) U n_j \) will not change the limiting distribution.

It follows that if \( \mathcal{L}(\Sigma n_j) \Rightarrow \mathcal{L}(\xi) \), the sum \( \sigma_n^2 \) must be bounded because the limiting distribution is that of a sum of two independent terms, one \( U \), from \( \Sigma(n_j - \xi_j) U n_j \), the other, \( V \) from \( \Sigma \xi_j V n_j \).

By the same calculation as in Section 2,

\[
E |\Sigma(n_j - \xi_j) U n_j|^4 \leq \sum_1 E |(n_j - \xi_j) U n_j|^4 + 3 \sigma_n^4
\]

\[
\leq a^2 \sigma_n^2 + 3 \sigma_n^4 .
\]

This is bounded. Hence the \( |\Sigma(n_j - \xi_j) U n_j|^2 \) are uniformly integrable and

\[
\lim \sigma_n^2 \leq E \xi^2 = 1 .
\]

This is true for every \( a \). Hence one can find a sequence \( \{a_n\} \), \( a_n \to \infty \) for which the corresponding \( \sigma_n^2 \) are still bounded.

For the corresponding \( \Sigma(n_j - \xi_j) U n_j \) Statement (9) says that

\[
\mathcal{L}(\Sigma(n_j - \xi_j) U n_j) \Rightarrow \mathcal{L}(\xi)
\]

and the argument of Section 3 is applicable.

Using the fact that

\[
(15) \quad \mathcal{L}(\Sigma \xi_j) \text{ behaves like } \mathcal{L}\{\Sigma(n_j - \xi_j) U n_j + \Sigma \xi_j V n_j\}
\]

one can also proceed differently.

The argument that gave Statement (9) in Section 4 shows also that for a value of \( a_0 \) that is large enough, but fixed, \( \sum P[|Y_n| > a_0] \) remains bounded.

Therefore, because of the boundedness of the variances (as above)

\[
\sum P[|Y_n| > a]
\]

remains bounded for every fixed \( a \), however small.
In such a case it is a very simple matter to prove that \( \mathcal{L}(\sum_{n,j} Y_{n,j}) \) can be approximated by a Poissonized sum

\[
T_n = \sum_{j} \sum_{k=1}^{v_j} Y_{n,j,k}
\]

where all the variables are independent, \( v_j \) is Poisson, \( E v_j = 1 \) and \( \mathcal{L}(Y_{n,j,k}) = \mathcal{L}(Y_{n,j}) \).

(Take a small, apply Lindeberg's technique to \( \sum (1-y_{n,j}) U_{n,j} \). For the part \( \sum E_{n,j} V_{n,j} \) use the fact that if \( p \) is small one can find a pair of variables \( M, N \) where \( P(M=1) = 1 - P(M=0) = p \), where \( N \) is Poisson, \( E N = p \) and \( P_n(N \neq N) \leq p^2 \).

Another simple fact is that if \( N \) is Poisson, \( E N = \lambda \) a sum \( \sum Y_{n,j,k} \) has a distribution

\[
\sum_{k} e^{-\lambda} \frac{\lambda^k}{k!} Q(k)
\]

where \( Q(k) \) is the distribution of a sum \( \sum_{i=1}^{k} Y_{n,j,i} \).

Multiplying series one sees that a sum such as

\[
\sum_{k} v_1 Y_{n1,k} + v_2 Y_{n2,k}
\]

has the same distribution as \( \sum_{k} X_k \) when \( N \) is Poisson \( E N = 2 \) and \( X_k \) has for distribution the average of the distributions:

\[
\mathcal{L}(X_k) = \frac{1}{2} [\mathcal{L}(Y_{n1,k}) + \mathcal{L}(Y_{n2,k})]
\]

(In other words \( e^{x+y} = e^x e^y \). Now note that the distribution of \( T_n \) is the same as that of

\[
\sum_{k} X_k, \quad \text{with} \quad P_n = \mathcal{L}(X_k) = \frac{1}{k_n} \sum_{j} \mathcal{L}(Y_{n,j})
\]

with \( k_n \) such that \( \sum_{k} X_k \).
N_n Poisson, E N_n = k_n (k_n is the range \( j = 1,2,\ldots,k_n \) for \( \sum_j Y_{nj} \)).

Note also that if \( \mathcal{L}(T_n) \rightarrow \mathcal{L}(\xi) \) with \( E \xi^4 = 3 \), one can truncate the \( P_n \) to a \( P'_n \) carried by \((-a_n, a_n)\) so that if \( \mathcal{L}(X'_k) = P'_n \) the Poissonized sum has fourth moments converging to that of \( \xi \).

However, the Poissonized sum has fourth moment

\[
k_n \nu_{n4} + 3 k_n^2 \sigma_n^4
\]

(easy, from Section 2 for instance), where \( \nu_{n4} = E(X'_k)^4 \), \( \sigma_n^2 = E(X'_k)^2 \). Here \( k_n \sigma_n^2 + 1 \). So for the fourth moments to tend to 3 it is necessary that

\[
k_n \nu_{n4} = \sum_j E \nu_{nj} I(|Y_{nj}| < a_n)
\]

tend to zero.

Now it is clear that the argument about multiplying exponential series is equivalent to that where one multiplies characteristic functions, but it is not at all necessary to use characteristic functions for it.

My feeling is that, all ready in 1934, Lévy knew all the relevant facts such as (9), (15), the possibility of Poissonizing etc. but somehow he did not use them.