An elementary proof of the converse of the Lindeberg-Lévy Theorem

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

$$(1) \qquad \sup_j P\{|X_{nj}| > \varepsilon\} \to 0$$

for each $\varepsilon > 0$ as $n \to \infty$.

Lindeberg (1920) (1922) gave sufficient conditions for the convergence of the distribution of $S_n = \Sigma X_{nj}$ 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_{nj} be as described and assume in addition to the negligibility hypothesis (1) that

(2) $\quad S_n = \sum_j X_{nj}$ has a distribution that converges to $\mathcal{L}(\xi)$ where

$$E\xi = 0, \quad E\xi^2 = 1, \quad \text{and} \quad E\xi^4 = 3 .$$

The problem is to show that:

(3) $P\{\sup_j |X_{nj}| > \varepsilon\} \to 0$ as $n \to \infty$.

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

Let $\{X'_{nj}; j=1,2,\ldots k_n; n=1,2,\ldots\}$ be an independent copy of the X_{nj} . Consider

$$Y_{nj} = \frac{1}{\sqrt{2}}(X_{nj}-X'_{nj}) .$$

Then the Y_{nj} Satisfy ~~ratio~~ by the negligibility condition (1) and the limit requirement (2).

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

$$P\{\sup_j |X_{nj}-m_{nj}| > \varepsilon\} \leq 2\, P\{\sup |Y_{nj}| > \varepsilon\} .$$

Hence, by (1)

$$P\{\sup |X_{nj}| > 3\varepsilon\} \leq 2\, P\{\sup |Y_{nj}| > \varepsilon\} + 2 \sup_j P\{|X_{nj}| > \varepsilon\}$$

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

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

$$S_n(i) = \sum_j \{Y_{nj}; k_{in} \leq j < k_{i+1,n}\} .$$

Taking subsequences, if necessary, assume that

$$\mathscr{L}\{S_n(i); i=1,\ldots,m\} \to \mathscr{L}\{Z_i; i=1,\ldots,m\} .$$

Then $E\sum_i Z_i = 0$, $E(\sum_i Z_i)^2 = 1$ and $E(\sum Z_i)^4 = 3$.

Note that $\mathscr{L}(Z_i) = \mathscr{L}(-Z_i)$ 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 valuables* Z_i *satisfy the relations*

$$\sum_i E\, Z_i^4 = 3 \sum v_i^2 \ , \tag{4}$$

$$\sum P\{|Z_i| > \varepsilon\} \leq \frac{3 \sum v_i^2}{\varepsilon^4} \ . \tag{5}$$

Hence, for n *sufficiently large,*

$$\sum P[|S_n(i)| > 2\varepsilon] \leq \varepsilon + \frac{3 \sum v_i^2}{\varepsilon^4} \ . \tag{6}$$

Proof. Statement (5) follows from (4) by Markov's inequality. Statement (6) follows then from $\mathscr{L}[S_n(i)] \to \mathscr{L}(Z_i)$. 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

$$\begin{aligned} 3 = E(\sum Z_i) &= \sum E\, Z_i^4 \\ &\quad + 6 \sum_i v_{i+1}^2 \sum_{j \leq i} v_j^2 \\ &= \sum E\, Z_i^4 + 3\{(\sum v_i)^2 - (\sum v_i^2)\} \ . \end{aligned}$$

Now $\sum v_i = E(\sum Z_i)^2 = 1$. Thus $\sum E\, Z_i^4 = 3 \sum v_i^2$. 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 $\Sigma\, v_j^2$ can be made small and (b) apply once more Paul Lévy symmetrization inequalities in the form

$$(7) \quad P\{\sup_s\{\sum_j [Y_{nj}; k_i \leq j < s]; k_i \leq s < k_{i+1}\} > \} \leq 2\, P\{|S_n(i)| > \varepsilon\} \quad .$$

We shall now proceed to show that $\Sigma\, v_j^2$ 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} are as in Section 2, but assume further that $\sigma_{nj}^2 = E\, Y_{nj}^2$ exists and that

$$(8) \qquad \sup_n \Sigma\, \sigma_{nj}^2 \leq c < \infty \quad \text{and} \quad \beta_n = \sup_j \sigma_{nj}^2 \to 0 \ .$$

In this case one can take for value of k_{in} the first integer such that

$$\Sigma[\sigma_{nj}^2; j \leq k_{in}] \geq \frac{\sigma'}{m}\, c \ .$$

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

4. The General Case

Suppose again that the Y_{nj} are as in Section 2 with $\mathscr{L}(\Sigma Y_{nj}) \to \mathscr{L}(\xi)$, $E\,\xi = 0$, $E\xi^2 = 1$, $E\xi^4 = 3$.

We claim first that

(9) For $\varepsilon > 0$ there is an $n(\varepsilon)$ and a $b(\varepsilon)$ such that $n \geq n(\varepsilon)$ implies

$$\sum_j P\{|Y_{nj}| > b(\varepsilon)\} < \varepsilon \quad .$$

(10) For any fixed number a

$$\sup_n \sum_j E\{|Y_{nj}|^2 I|Y_{nj}|\leq a\} < \infty \quad .$$

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

$$P\{\sup_s |\Sigma[Y_{nj};j\leq s]|>b\} \quad \leq 2\ P\{|S_n|>b\}$$

with $S_n = \sum_j Y_{nj}\ ;\ j \leq k_n$. 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 ,

$$\Sigma\, E\ Y_{nj}^2\ I[|Y_{nj}|\leq a] \geq b_n^4$$

with $b_n^4 \to \infty$. Thus we can assume $b_n > a$.

Consider then the variables

$$V_{nj} = \frac{1}{b_n} Y_{nj} [I|Y_{nj}|\leq b_n]$$

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

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

$$s_{nj}^2 = E\ Y_{nj}^2\ I[|Y_{nj}|\leq a]$$

and

$$\alpha_{nj} = P[|Y_{nj}|>a] \quad .$$

Partition the total sum as before, considering integers k_{in} such that

$$\Sigma\{[s_{nj}^2+\alpha_{nj}];\ k_{in} \leq j < k_{i+1,n}\} < \frac{c}{m}$$

and let

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

Taking subsequence if necessary one can assume that

$$\mathscr{L}[S_n(i)] \to \mathscr{L}(Z_i) \ .$$

Since $\mathscr{L}(\Sigma Z_i) = \mathscr{L}(\xi)$ with $E\,\xi^4 = 3$, one must also have $E\,Z_i^4 \leq 3$.

However, by Chebyshev

$$P\{\Sigma[Y_{nj}I[|Y_{nj}|\leq a];k_{in}\leq j<k_{i+1,n}] \geq \tau\} \leq \frac{\cancel{\tau}c}{m}\frac{1}{\tau^2} \qquad \frac{c}{m}\frac{1}{\tau^2} \quad \frac{c}{m\tau^2}$$

Also

$$P\left\{\{\Sigma[Y_{nj}I|Y_{nj}|>a]\neq 0\} \leq \Sigma[\alpha_j;k_{in}\leq j<k_{i+1,n}]\right\} \ .$$

Hence $\quad P\{|S_n(i)|>\tau\} \leq \frac{2\cancel{\tau}c}{m}\frac{1}{\tau^2}$, if $\tau \leq 1$.

Hence $\quad P\{|Z_i|>\tau\} \leq \frac{2\cancel{\tau}c}{m}\frac{1}{\tau^2}$.

Also,

$$E\,Z_i^2 \leq E\,Z_i^2\,I[|Z_i|\leq t] + \frac{1}{t^2}\,E\,Z_i^4\,I[|Z_i|>t]$$

$$\leq \tau^2 + t^2\,P[|Z_i|>\tau] + \frac{1}{t^2}E\,Z_i^4 \ .$$

If one takes

$$t^2 = \frac{3}{\varepsilon}\ , \quad \tau^2 = \varepsilon\ , \quad \text{and} \quad \frac{6c}{m} \leq \cancel{\varepsilon^{3/2}} \quad \varepsilon^3$$

this becomes at most 3ε.

Thus the partitioning of Section 2 is still possible, with Σv_i^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

$$Y^*_{nj} = (1-y_{nj})\, U_{nj} + \xi_{nj}\, V_{nj} \quad + \quad (y_{nj}-\xi_{nj})U_{nj}$$

where the ξ, y, U and V are all independent and where

(12) $$\xi_{nj} = \xi^2_{nj}\ ,\ y_{nj} = y^2_{nj}\ ,$$

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

(13) $$\mathscr{L}(U_{nj}) = \mathscr{L}\{Y_{nj}\,|\,|Y_{nj}|\leq a\}$$

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

Here, $\alpha_n = \sup_j \alpha_{nj} \to 0$. The variance of $\sum_j (y_{nj}-\xi_{nj})U_{nj}$ is at most $2\alpha_n(1-\alpha_n)\,\Sigma\,E\,U^2_{nj}$. That of $\Sigma(1-y_{nj})U_{nj}$ is $\Sigma(1-\alpha_{nj})\,E\,U^2_{nj} \geq (1-\alpha)\,\Sigma\,E\,U^2_{nj}$.

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

If $\operatorname{var} \Sigma(1-y_{nj})U_{nj} = \sigma^2_n$ tends to infinity, the term $\Sigma(1-y_{nj})U_{nj}$ has a distribution close to $\mathcal{N}(0,\sigma^2_n)$ 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 $\varepsilon + \frac{\tau}{\sqrt{2\pi}}$. By Chebyshev

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

However, for any two variables X and Y (not necessarily independent)

$$|P[X+Y\leq t] - P[X\leq t]| \leq P[|Y|>\tau] + \sup_x P[x\leq X\leq x+\tau] \quad .$$

Thus, removing the cross term $\Sigma(y_{nj}-\xi_{nj})U_{nj}$ will not change the limiting distribution.

It follows that if $\mathscr{L}(\Sigma Y_{nj}) \to \mathscr{L}(\xi)$, the sum σ_n^2 must be bounded because the limiting distribution is that of a sum of two independent terms, one U , from $\Sigma(1-y_{nj})U_{nj}$, the other, V from $\Sigma\, \xi_{nj} V_{nj}$.

By the same calculation as in Section 2 ,

$$E\,|\Sigma(1-y_{nj})U_{nj}|^4 \leq \sum_i E|(1-y_{nj})U_{nj}|^4 + 3\,\sigma_n^4$$

$$\leq a^2\,\sigma_n^2 + 3\,\sigma_n^4 \quad .$$

This is bounded. Hence the $|\Sigma(1-y_{nj})U_{nj}|^2$ 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 σ_n^2 are still bounded.

For the corresponding $\Sigma(1-y_{nj})U_{nj}$ Statement (9) says that

$$\mathscr{L}[\Sigma(1-y_{nj})U_{nj}] \to \mathscr{L}(\xi)$$

and the argument of Section 3 is applicable.

Using the fact that

$$(15) \qquad \mathscr{L}(\Sigma Y_{nj}) \text{ behaves like } \mathscr{L}\{\Sigma(1-y_{nj})U_{nj} + \Sigma\, \xi_{nj}V_{nj}\}$$

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, $\sum_i P\{|Y_{nj}|>a_0\}$ remains bounded.

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

$$\sum_i P[|Y_{nj}|>a]$$

remains bounded for every fixed a , however small.

In such a case it is a very simple matter to prove that $\mathscr{L}(\Sigma Y_{nj})$ can be approximated by a Poissonized sum

$$T_n = \sum_j \sum_{k=1}^{v_j} Y_{nj,k}$$

where all the variables are independent, v_j is Poisson, $E\, v_j = 1$ and $\mathscr{L}(Y_{nj,k}) = \mathscr{L}(Y_{nj})$.

(Take a small, apply Lindeberg's technique to $\Sigma(1-y_{nj})U_{nj}$. For the part $\Sigma\, \xi_{nj}\, V_{nj}$ 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 $\underset{N}{P_r}(M \neq N) \leq p^2$).

Another simple fact is that if N is Poisson, $E\, N = \lambda$ a sum $\Sigma\, Y_{nj,k}$ has a distribution

$$\Sigma\, e^{-\lambda} \frac{\lambda^k}{k!} Q^{(k)}$$

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

Multiplying series one sees that a sum such as

$$\sum_k^{v_1} Y_{n1,k} + \sum_k^{v_2} Y_{n2,k}$$

has the same distribution as $\sum_k^{N} X_k$ when N is Poisson $E\, N = 2$ and X_k has for distribution the average of the distributions:

$$\mathscr{L}(X_k) = \frac{1}{2}\, [\mathscr{L}(Y_{n1,k}) + \mathscr{L}(Y_{n2,k})] \quad .$$

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

$$\sum_k^{N_n} X_k\,, \quad \text{with} \quad P_n = \mathscr{L}(X_k) = \frac{1}{k_n} \sum_j^{k_n} \mathscr{L}(Y_{nj}) \quad .$$

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

However, the Poissonized sum has fourth moment

$$k_n\,\mu_{n4} + 3\,k_n^2\,\sigma_n^4$$

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

$$k_n\,\mu_{n,4} = \Sigma\,E\,Y_{nj}^4\,I(|Y_{nj}|<a_n)$$

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.