Chapter 3

Independent Sums

3.1 Independence and Convolution

One of the central ideas in probabilty is the notion of independence. In intuitive terms two events are independent if they have no influence on each other. The formal definition is

Definition 3.1. *Two events A and B are said to be independent if*

$$P[A \cap B] = P[A]P[B].$$

Exercise 3.1. If A and B are independent prove that so are A^c and B.

Definition 3.2. *Two random variables X and Y are independent if the events $X \in A$ and $Y \in B$ are independent for any two Borel sets A and B on the line i.e.*

$$P[X \in A, Y \in B] = P[X \in A]P[Y \in B].$$

for all Borel sets A and B.

There is a natural extension to a finite or even an infinite collection of random variables.

Definition 3.3. *A finite collection collection $\{X_j : 1 \le j \le n\}$ of random variables are said to be independent if for any n Borel sets $A_1, \ldots, A_n$ on the line*

$$P\Big[\cap_{1\le j\le n}[X_j \in A_j]\Big] = \Pi_{1\le j\le n} P[X_j \in A_j].$$

Definition 3.4. *An infinite collection of random variables is said to be independent if every finite subcollection is independent.*

Lemma 3.1. *Two random variables X, Y defined on (Ω, Σ, P) are independent if and only if the measure induced on $\mathbf{R}^2$ by (X, Y), is the product measure $\alpha \times \beta$ where α and β are the distributions on $\mathbf{R}$ induced by X and Y respectively.*

Proof. Left as an exercise. □

The important thing to note is that if X and Y are independent and one knows their distributions α and β, then their joint distribution is automatically determined as the product measure.

If X and Y are independent random variables having α and β for their distributions, the distribution of the sum $Z = X+Y$ is determined as follows. First we construct the product measure $\alpha\times\beta$ on $\mathbf{R}\times\mathbf{R}$ and then consider the induced distribution of the function $f(x, y) = x+y$. This distribution, called the convolution of α and β, is denoted by $\alpha * \beta$. An elementary calculation using Fubini's theorem provides the following identities.

$$(\alpha * \beta)(A) = \int \alpha(A - x)\, d\beta = \int \beta(A - x)\, d\alpha \tag{3.1}$$

In terms of characteristic function, we can express the characteristic function of the convolution as

$$\begin{aligned}\int \exp[\,i\,t\,x\,] d(\alpha * \beta) &= \int\int \exp[\,i\,t\,(x + y)\,]\, d\,\alpha\, d\,\beta \\ &= \int \exp[\,i\,t\,x\,]\, d\,\alpha \int \exp[\,i\,t\,x\,]\ d\,\beta\end{aligned}$$

or equivalently

$$\phi_{\alpha*\beta}(t) = \phi_\alpha(t)\phi_\beta(t) \tag{3.2}$$

which provides a direct way of calculating the distributions of sums of independent random variables by the use of characteristic functions.

Exercise 3.2. If X and Y are independent show that for any two measurable functions f and g, $f(X)$ and $g(Y)$ are independent.

Exercise 3.3. Use Fubini's theorem to show that if X and Y are independent and if f and g are measurable functions with both $E[|f(X)|]$ and $E[|g(Y)|]$ finite then

$$E[f(X)g(Y)] = E[f(X)]E[g(Y)].$$

Exercise 3.4. Show that if X and Y are any two random variables then $E(X+Y) = E(X) + E(Y)$. If X and Y are two independent random variables then show that

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

where

$$\mathrm{Var}(X) = E\big[[X - E[X]]^2\big] = E[X^2] - [E[X]]^2.$$

If $X_1, X_2, \cdots, X_n$ are n independent random variables, then the distribution of their sum $S_n = X_1 + X_2 + \cdots + X_n$ can be computed in terms of the distributions of the summands. If α_j is the distribution of X_j, then the distribution of μ_n of S_n is given by the convolution $\mu_n = \alpha_1 * \alpha_2 * \cdots * \alpha_n$ that can be calculated inductively by $\mu_{j+1} = \mu_j * \alpha_{j+1}$. In terms of their characteristic functions $\psi_n(t) = \phi_1(t)\phi_2(t)\cdots\phi_n(t)$. The first two moments of S_n are computed easily.

$$E(S_n) = E(X_1) + E(X_2) + \cdots E(X_n)$$

and

$$\begin{aligned}
\mathrm{Var}(S_n) &= E[S_n - E(S_n)]^2 \\
&= \sum_j E[X_j - E(X_j)]^2 \\
&\quad +2 \sum_{1\le i<j\le n} E[X_i - E(X_i)][X_j - E(X_j)].
\end{aligned}$$

For $i \neq j$, because of independence

$$E[X_i - E(X_i)][X_j - E(X_j)] = E[X_i - E(X_i)]E[X_j - E(X_j)] = 0$$

and we get the formula

$$\mathrm{Var}(S_n) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n). \tag{3.3}$$

3.2 Weak Law of Large Numbers

Let us look at the distribution of the number of succeses in n independent trials, with the probability of success in a single trial being equal to p.

$$P\{S_n = r\} = \binom{n}{r} p^r (1-p)^{n-r}$$

and

$$\begin{aligned}
P\{|S_n - np| \geq n\delta\} &= \sum_{|r-np|\geq n\delta} \binom{n}{r} p^r (1-p)^{n-r} \\
&\leq \frac{1}{n^2\delta^2} \sum_{|r-np|\geq n\delta} (r-np)^2 \binom{n}{r} p^r (1-p)^{n-r} \qquad (3.4) \\
&\leq \frac{1}{n^2\delta^2} \sum_{1\leq r\leq n} (r-np)^2 \binom{n}{r} p^r (1-p)^{n-r} \\
&= \frac{1}{n^2\delta^2} E[S_n - np]^2 = \frac{1}{n^2\delta^2}\mathrm{Var}(S_n) \qquad (3.5) \\
&= \frac{1}{n^2\delta^2} np(1-p) \qquad (3.6) \\
&= \frac{p(1-p)}{n\delta^2}.
\end{aligned}$$

In the step (3.4) we have used a discrete version of the simple inequality

$$\int_{x:g(x)\geq a} g(x) d\alpha \geq g(a)\alpha[x : g(x) \geq a]$$

with $g(x) = (x - np)^2$ and in (3.6) have used the fact that $S_n = X_1 + X_2 + \cdots + X_n$ where the X_i are independent and have the simple distribution

$P\{X_i = 1\} = p$ and $P\{X_i = 0\} = 1 - p$. Therefore $E(S_n) = np$ and $\mathrm{Var}(S_n) = n\mathrm{Var}(X_1) = np(1-p)$

It follows now that

$$\lim_{n\to\infty} P\{|S_n - np| \geq n\delta\} = \lim_{n\to\infty} P\{|\frac{S_n}{n} - p| \geq \delta\} = 0$$

or the average $S_n n$ converges to p in probability. This is seen easily to be equivalent to the statement that the distribution of $\frac{S_n}{n}$ converges to the distribution degenerate at p. See (2.12).

The above argument works for any sequence of independent and identically distributed random variables. If we assume that $E(X_i) = m$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$, then $E(\frac{S_n}{n}) = m$ and $\mathrm{Var}(\frac{S_n}{n}) = \frac{\sigma^2}{n}$. Chebychev's inequality states that for any random variable X

$$\begin{aligned} P\{|X - E[X]| \geq \delta\} &= \int_{|X-E[X]|\geq\delta} dP \\ &\leq \frac{1}{\delta^2}\int_{|X-E[X]|\geq\delta} [X - E[X]]^2 dP \\ &= \frac{1}{\delta^2}\int [X - E[X]]^2 dP \\ &= \frac{1}{\delta^2}\mathrm{Var}(X). \end{aligned} \tag{3.7}$$

This can be used to prove the weak law of large numbers for the general case of independent identically distributed random variables with finite second moments.

Theorem 3.2. *If $X_1, X_2 \ldots, X_n, \ldots$ is a sequence of independent identically distributed random variables with $E[X_j] \equiv m$ and $\mathrm{VarX_j} \equiv \sigma^2$ then for*

$$S_n = X_1 + X_2 + \cdots + X_n$$

we have

$$\lim_{n\to\infty} P\left[|\frac{S_n}{n} - m| \geq \delta\right] = 0$$

for any $\delta > 0$.

Proof. Use Chebychev's inequality to estimate

$$P\left[|\frac{S_n}{n} - m| \geq \delta\right] \leq \frac{1}{\delta^2}\text{Var}(\frac{S_n}{n}) = \frac{\sigma^2}{n\delta^2}.$$

□

Actually it is enough to assume that $E|X_i| < \infty$ and the existence of the second moment is not needed. We will provide two proofs of the statement

Theorem 3.3. *If $X_1, X_2, \cdots X_n$ are independent and identically distributed with a finite first moment and $E(X_i) = m$, then $\frac{X_1+X_2+\cdots+X_n}{n}$ converges to m in probability as $n \to \infty$.*

Proof. **1.** Let C be a large constant and let us define X_i^C as the *truncated* random variable $X_i^C = X_i$ if $|X_i| \leq C$ and $X_i^C = 0$ otherwise. Let $Y_i^C = X_i - X_i^C$ so that $X_i = X_i^C + Y_i^C$. Then

$$\begin{aligned}\frac{1}{n}\sum_{1\leq i\leq n} X_i &= \frac{1}{n}\sum_{1\leq i\leq n} X_i^C + \frac{1}{n}\sum_{1\leq i\leq n} Y_i^C \\ &= \xi_n^C + \eta_n^C.\end{aligned}$$

If we denote by $a_C = E(X_i^C)$ and $b_C = E(Y_i^C)$ we always have $m = a_C + b_C$. Consider the quantity

$$\begin{aligned}\delta_n &= E[|\frac{1}{n}\sum_{1\leq i\leq n} X_i - m|] \\ &= E[|\xi_n^C + \eta_n^C - m|] \\ &\leq E[|\xi_n^C - a_C|] + E[|\eta_n^C - b_C|] \\ &\leq \left[E[|\xi_n^C - a_C|^2]\right]^{\frac{1}{2}} + 2E[|Y_i^C|].\end{aligned} \tag{3.8}$$

As $n \to \infty$, the truncated random variables X_i^C are bounded and independent. Theorem 3.2 is applicable and the first of the two terms in (3.8) tends to 0. Therefore taking the limsup as $n \to \infty$, for any $0 < C < \infty$,

$$\limsup_{n\to\infty} \delta_n \leq 2E[|Y_i^C|].$$

If we now let the cutoff level C to go to infinity, by the integrability of X_i, $E\big[|Y_i^C|\big] \to 0$ as $C \to \infty$ and we are done. The final step of establishing that for any sequence Y_n of random variables, $E[|Y_n|] \to 0$ implies that $Y_n \to 0$ in probability, is left as an exercise and is not very different from Chebychev's inequality.

Proof 2. We can use characteristic functions. If we denote the characteristic function of X_i by $\phi(t)$, then the characteristic function of $\frac{1}{n}\sum_{1\le i\le n} X_i$ is given by $\psi_n(t) = [\phi(\frac{t}{n})]^n$. The existence of the first moment assures us that $\phi(t)$ is differentiable at $t = 0$ with a derivative equal to im where $m = E(X_i)$. Therefore by Taylor expansion

$$\phi(\frac{t}{n}) = 1 + \frac{i\,m\,t}{n} + o(\frac{1}{n}).$$

Whenever $na_n \to z$ it follows that $(1 + a_n)^n \to e^z$. Therefore,

$$\lim_{n\to\infty} \psi_n(t) = \exp[\,i\,m\,t\,]$$

which is the characteristic function of the distribution degenerate at m. Hence the distribution of $\frac{S_n}{n}$ tends to the degenerate distribution at the point m. The weak law of large numbers is thereby established. ☐

Exercise 3.5. If the underlying distribution is a Cauchy distribution with density $\frac{1}{\pi(1+x^2)}$ and characteristic function $\phi(t) = e^{-|t|}$, prove that the weak law does not hold.

Exercise 3.6. The weak law may hold sometimes even if the mean does not exist. If we dampen the tails of the Cauchy ever so slightly with a density $f(x) = \frac{c}{(1+x^2)\log(1+x^2)}$, show that the weak law of large numbers holds.

Exercise 3.7. In the case of the Binomial distribution with $p = \frac{1}{2}$, use Stirling's formula

$$n! \simeq \sqrt{2\pi}\; e^{-n}\; n^{n+12}$$

to estimate the probability

$$\sum_{r\ge nx} \binom{n}{r} \frac{1}{2^n}$$

and show that it decays geometrically in n. Can you calculate the geometric ratio

$$\rho(x) = \lim_{n\to\infty} \left[\sum_{r\ge nx} \binom{n}{r} \frac{1}{2^n}\right]^{\frac{1}{n}}$$

explicitly as a function of x for $x > \frac{1}{2}$?

3.3 Strong Limit Theorems

The weak law of large numbers is really a result concerning the behavior of

$$\frac{S_n}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

where $X_1, X_2, \cdots, X_n, \ldots$ is a sequence of independent and identically distributed random variables on some probability space $(\Omega, \mathcal{B}, P)$. Under the assumption that X_i are integrable with an integral equal to m, the weak law asserts that as $n \to \infty$, $\frac{S_n}{n} \to m$ in Probability. Since almost everywhere convergence is generally stronger than convergence in Probability one may ask if

$$P\left[\omega : \lim_{n\to\infty} \frac{S_n(\omega)}{n} = m\right] = 1$$

This is called the Strong Law of Large Numbers. Strong laws are statements that hold for almost all ω.

Let us look at functions of the form $f_n = \chi_{A_n}$. It is easy to verify that $f_n \to 0$ in probability if and only if $P(A_n) \to 0$. On the other hand

Lemma 3.4. (Borel-Cantelli lemma). *If*

$$\sum_n P(A_n) < \infty$$

then

$$P\big[\omega : \lim_{n\to\infty} \chi_{A_n}(\omega) = 0\big] = 1.$$

If the events A_n are mutually independent the converse is also true.

Remark 3.1. Note that the complementary event

$$\big[\omega : \limsup_{n\to\infty} \chi_{A_n}(\omega) = 1\big]$$

is the same as $\cap_{n=1}^{\infty} \cup_{j=n}^{\infty} A_j$, or the event that infinitely many of the events $\{A_j\}$ occcur.

The cnclusion of the next exercise will be used in the proof.

Exercise 3.8. Prove the following variant of the monotone convergence theorem. If $f_n(\omega) \geq 0$ are measurble functions the set $E = \{\omega : S(\omega) = \sum_n f_n(\omega) < \infty\}$ is measurable and $S(\omega)$ is a measurable function on E. If each f_n is integrable and $\sum_n E[f_n] < \infty$ then $P[E] = 1$, $S(\omega)$ is integrable and $E[S(\omega)] = \sum_n E[f_n(\omega)]$.

Proof. By the previous exercise if $\sum_n P(A_n) < \infty$, then $\sum_n \chi_{A_n}(\omega) = S(\omega)$ is finite almost everywhere and

$$E(S(\omega)) = \sum_n P(A_n) < \infty.$$

If an infinite series has a finite sum then the n-th term must go to 0, thereby proving the direct part. To prove the converse we need to show that if $\sum_n P(A_n) = \infty$, then $\lim_{m\to\infty} P(\cup_{n=m}^{\infty} A_n) > 0$. We can use independence and the continuity of probability under monotone limits, to calculate for every m,

$$\begin{aligned} P(\cup_{n=m}^{\infty} A_n) &= 1 - P(\cap_{n=m}^{\infty} A_n^c) \\ &= 1 - \prod_{n=m}^{\infty}(1 - P(A_n)) \quad \text{(by independence)} \\ &\geq 1 - e^{-\sum_m^{\infty} P(A_n)} \\ &= 1 \end{aligned}$$

and we are done. We have used the inequality $1 - x \leq e^{-x}$ familiar in the study of infinite products. □

Another digression that we want to make into measure theory at this point is to discuss Kolmogorov's consistency theorem. How do we know that there are probability spaces that admit a sequence of independent identically distributed random variables with specified distributions? By the construction of product measures that we outlined earlier we can construct a measure on $\mathbf{R}^n$ for every n which is the joint distribution of the first n random variables. Let us denote by P_n this probability measure on $\mathbf{R}^n$. They are consistent in the sense that if we project in the natural way from $\mathbf{R}^{n+1} \to \mathbf{R}^n$, P_{n+1} projects to P_n. Such a family is called a consistent family of finite dimensional distributions. We look at the space $\Omega = \mathbf{R}^\infty$ consisting of all real sequences $\omega = \{x_n : n \geq 1\}$ with a natural σ-field Σ generated by the field $\mathcal{F}$ of finite dimensional cylinder sets of the form $B = \{\omega : (x_1, \cdots, x_n) \in A\}$ where A varies over Borel sets in $\mathbf{R}^n$ and varies over positive integers.

Theorem 3.5. (Kolmogorov's Consistency Theorem). *Given a consistent family of finite dimensional distributions P_n, there exists a unique P on (Ω, Σ) such that for every n, under the natural projection $\pi_n(\omega) = (x_1, \cdots, x_n)$, the induced measure $P\pi_n^{-1} = P_n$ on $\mathbf{R}^n$.*

Proof. The consistency is just what is required to be able to define P on $\mathcal{F}$ by

$$P(B) = P_n(A).$$

Once we have P defined on the field $\mathcal{F}$, we have to prove the countable additivity of P on $\mathcal{F}$. The rest is then routine. Let $B_n \in \mathcal{F}$ and $B_n \downarrow \Phi$, the empty set. If possible let $P(B_n) \geq \delta$ for all n and for some $\delta > 0$. Then $B_n = \pi_{k_n}^{-1} A_{k_n}$ for some k_n and without loss of generality we assume that $k_n = n$, so that $B_n = \pi_n^{-1} A_n$ for some Borel set $A_n \subset \mathbf{R}^n$. According to Exercise 3.8 below, we can find a closed bounded subset $K_n \subset A_n$ such that

$$P_n(A_n - K_n) \leq \frac{\delta}{2^{n+1}}$$

and define $C_n = \pi_n^{-1} K_n$ and $D_n = \cap_{j=1}^n C_j = \pi_n^{-1} F_n$ for some closed bounded set $F_n \subset K_n \subset \mathbf{R}^n$. Then

$$P(D_n) \geq \delta - \sum_{j=1}^n \frac{\delta}{2^{j+1}} \geq \frac{\delta}{2}.$$

$D_n \subset B_n$, $D_n \downarrow \Phi$ and each D_n is nonempty. If we take $\omega^{(n)} = \{x_j^n : j \geq 1\}$ to be an arbitrary point from D_n, by our construction $(x_1^n, \cdots x_m^n) \in F_m$ for $n \geq m$. We can definitely choose a subsequence (diagonlization) such that $x_j^{n_k}$ converges for each j producing a limit $\omega = (x_1, \cdots, x_m, \cdots)$ and, for every m, we will have $(x_1, \cdots, x_m) \in F_m$. This implies that $\omega \in D_m$ for every m, contradicting $D_n \downarrow \Phi$. We are done. $\square$

Exercise 3.9. We have used the fact that given any borel set $A \subset \mathbf{R}^n$, and a probability measure α on $\mathbf{R}^n$, for any $\epsilon > 0$, there exists a closed bounded subset $K_\epsilon \subset A$ such that $\alpha(A - K_\epsilon) \leq \epsilon$. Prove it by showing that the class of sets $\mathcal{A}$ with the above property is a monotone class that contains finite disjoint unions of measurable rectangles and therefore contains the Borel σ-field. To prove the last fact, establish it first for $n = 1$. To handle $n = 1$, repeat the same argument starting from finite disjoint unions of right-closed left-open intevals. Use the countable additivity to verify this directly.

Remark 3.2. Kolmogorov's consistency theorem remains valid if we replace $\mathbf{R}$ by an arbitrary complete separable metric space X, with its Borel σ-field. However it is not valid in complete generality. See [8]. See Remark 4.7 in this context.

The following is a strong version of the Law of Large Numbers.

Theorem 3.6. *If $X_1, \cdots, X_n \cdots$ is a sequence of independent identically distributed random variables with $E|X_i|^4 = C < \infty$, then*

$$\lim_{n\to\infty} \frac{S_n}{n} = \lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = E(X_1)$$

with probability 1.

Proof. We can assume without loss of generality that $E[X_i] = 0$. Just take $Y_i = X_i - E[X_i]$. A simple calculation shows

$$E[(S_n)^4] = nE[(X_1)^4] + 3n(n-1)E[(X_1)^2]^2 \leq nC + 3n^2\sigma^4$$

and by applying a Chebychev type inequality using fourth moments,

$$P[|\frac{S_n}{n}| \geq \delta\,] = P[\,|S_n| \geq n\delta\,] \leq \frac{nC + 3n^2\sigma^4}{n^4\delta^4}.$$

We see that

$$\sum_{n=1}^{\infty} P[\,|\frac{S_n}{n}| \geq \delta\,] < \infty$$

and we can now apply the Borel-Cantelli Lemma. □

3.4 Series of Independent Random variables

We wish to investigate conditions under which an infinite series with independent summands

$$S = \sum_{j=1}^{\infty} X_j$$

converges with probability 1. The basic steps are the following inequalities due to Kolomogorov and Lévy that control the behaviour of sums of independent random variables. They both deal with the problem of estimating

$$T_n(\omega) = \sup_{1\leq k\leq n} |S_k(\omega)| = \sup_{1\leq k\leq n} |\sum_{j=1}^{k} X_j(\omega)|$$

where $X_1, \cdots, X_n$ are n independent random variables.

Lemma 3.7. (Kolmogorov's Inequality). *Assume that* $EX_i = 0$ *and* $\mathrm{Var}(X_i) = \sigma_i^2 < \infty$ *and let* $s_n^2 = \sum_{j=1}^n \sigma_j^2$. *Then*

$$P\{T_n(\omega) \geq \ell\} \leq \frac{s_n^2}{\ell^2}. \tag{3.9}$$

Proof. The important point here is that the estimate depends only on s_n^2 and not on the number of summands. In fact the Chebychev bound on S_n is

$$P\{|S_n| \geq \ell\} \leq \frac{s_n^2}{\ell^2}$$

and the supremum does not cost anything.

Let us define the events $E_k = \{|S_1| < \ell, \cdots, |S_{k-1}| < \ell, |S_k| \geq \ell\}$ and then $\{T_n \geq \ell\} = \cup_{k=1}^n E_k$ is a disjoint union of E_k. If we use the independence of $S_n - S_k$ and $S_k \chi_{E_k}$ that only depends on $X_1 \cdots, X_k$

$$\begin{aligned} P\{E_k\} &\leq \frac{1}{\ell^2}\int_{E_k} S_k^2\, dP \\ &\leq \frac{1}{\ell^2}\int_{E_k} \left[S_k^2 + (S_n - S_k)^2\right] dP \\ &= \frac{1}{\ell^2}\int_{E_k} \left[S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2\right] dP \\ &= \frac{1}{\ell^2}\int_{E_k} S_n^2\, dP. \end{aligned}$$

Summing over k from 1 to n

$$P\{T_n \geq \ell\} \leq \frac{1}{\ell^2}\int_{T_n \geq \ell} S_n^2\, dP \leq \frac{s_n^2}{\ell^2}.$$

eatblishing (3.9) ☐

Lemma 3.8. (Lévy's Inequality). *Assume that*

$$P\{|X_i + \cdots + X_n| \geq \frac{\ell}{2}\} \leq \delta$$

for all $1 \leq i \leq n$. *Then*

$$P\{T_n \geq \ell\} \leq \frac{\delta}{1 - \delta}. \tag{3.10}$$

Proof. Let E_k be as in the previous lemma.

$$\begin{aligned} P\left\{(T_n \geq \ell) \cap |S_n| \leq \frac{\ell}{2}\right\} &= \sum_{k=1}^{n} P\left\{E_k \cap |S_n| \leq \frac{\ell}{2}\right\} \\ &\leq \sum_{k=1}^{n} P\left\{E_k \cap |S_n - S_k| \geq \frac{\ell}{2}\right\} \\ &= \sum_{k=1}^{n} P\left\{|S_n - S_k| \geq \frac{\ell}{2}\right\} P(E_k) \\ &\leq \delta \sum_{k=1}^{n} P(E_k) \\ &= \delta P\{T_n \geq \ell\}. \end{aligned}$$

On the other hand,

$$P\left\{(T_n \geq \ell) \cap |S_n| > \frac{\ell}{2}\right\} \leq P\left\{|S_n| > \frac{\ell}{2}\right\} \leq \delta.$$

Adding the two,

$$P\big\{T_n \geq \ell\big\} \leq \delta P\big\{T_n \geq \ell\big\} + \delta$$

or

$$P\big\{T_n \geq \ell\big\} \leq \frac{\delta}{1-\delta}$$

proving (3.10) □

We are now ready to prove

Theorem 3.9. (Lévy's Theorem). *If $X_1, X_2, \ldots, X_n, \ldots$ is a seqence of independent random variables, then the following are equivalent.*

(i) The distribution α_n of $S_n = X_1 + \cdots + X_n$ converges weakly to a probability distribution α on $\mathbf{R}$.

(ii) The random variable $S_n = X_1 + \cdots + X_n$ converges in probability to a limit $S(\omega)$.

(iii) The random variable $S_n = X_1 + \cdots + X_n$ converges with probability 1 *to a limit $S(\omega)$.*

Proof. Clearly $(iii) \Rightarrow (ii) \Rightarrow (i)$ are trivial. We will establish $(i) \Rightarrow (ii) \Rightarrow (iii)$.

$(i) \Rightarrow (ii)$**.** The characteristic functions $\phi_j(t)$ of X_j are such that

$$\phi(t) = \prod_{i=1}^{\infty} \phi_j(t)$$

is a convergent infinite product. Since the limit $\phi(t)$ is continuous at $t = 0$ and $\phi(0) = 1$ it is nonzero in some interval $|t| \leq T$ around 0. Therefore for $|t| \leq T$,

$$\lim_{\substack{n\to\infty \\ m\to\infty}} \prod_{m+1}^{n} \phi_j(t) = 1.$$

By Exercise 3.10 below, this implies that for all t,

$$\lim_{\substack{n\to\infty \\ m\to\infty}} \prod_{m+1}^{n} \phi_j(t) = 1$$

and consequently, the distribution of $S_n - S_m$ converges to the distribution degenerate at 0. This implies the convergence in probability to 0 of $S_n - S_m$ as $m, n \to \infty$. Therefore for each $\delta > 0$,

$$\lim_{\substack{n\to\infty \\ m\to\infty}} P\{|S_n - S_m| \geq \delta\} = 0$$

establishing (ii).

$(ii) \Rightarrow (iii)$**.** To establish (iii), because of Exercise 3.11 below, we need only show that for every $\delta > 0$

$$\lim_{\substack{n\to\infty \\ m\to\infty}} P\Big\{\sup_{m<k\leq n} |S_k - S_m| \geq \delta\Big\} = 0$$

and this follows from (ii) and Lévy's inequality. □

Exercise 3.10. Prove the inequality $1 - \cos 2t \leq 4(1 - \cos t)$ for all real t. Deduce the inequality $1 - \text{Real}\,\phi(2t) \leq 4[1 - \text{Real}\,\phi(t)]$, valid for any characteristic function. Conclude that if a sequence of characteristic functions converges to 1 in an interval around 0, then it converges to 1 for all real t.

Exercise 3.11. Prove that if a sequence S_n of random variables is a Cauchy sequence in Probability, i.e. for each $\delta > 0$,

$$\lim_{\substack{n\to\infty\\ m\to\infty}} P\{|S_n - S_m| \geq \delta\} = 0$$

then there is a random variable S such that $S_n \to S$ in probability, i.e for each $\delta > 0$,

$$\lim_{n\to\infty} P\{|S_n - S| \geq \delta\} = 0.$$

Exercise 3.12. Prove that if a sequence S_n of random variables satisfies

$$\lim_{\substack{n\to\infty\\ m\to\infty}} P\Big\{ \sup_{m<k\leq n} |S_k - S_m| \geq \delta\Big\} = 0$$

for every $\delta > 0$ then there is a limiting random variable $S(\omega)$ such that

$$P\left\{ \lim_{n\to\infty} S_n(\omega) = S(\omega)\right\} = 1.$$

Exercise 3.13. Prove that whenever $X_n \to X$ in probability the distribution α_n of X_n converges weakly to the distribution α of X.

Now it is straightforward to find sufficient conditions for the convergence of an infinite series of independent random variables.

Theorem 3.10. (Kolmogorov's one series Theorem). *Let a sequence $\{X_i\}$ of independent random variables, each of which has finite mean and variance, satisfy $E(X_i) = 0$ and $\sum_{i=1}^{\infty} \mathrm{Var}(X_i) < \infty$, then*

$$S(\omega) = \sum_{i=1}^{\infty} X_i(\omega)$$

converges with probability 1.

Proof. By a direct application of Kolmogorov's inequality

$$\begin{aligned} \lim_{\substack{n\to\infty\\ m\to\infty}} P\Big\{ \sup_{m<k\leq n} |S_k - S_m| \geq \delta\Big\} &\leq \lim_{\substack{n\to\infty\\ m\to\infty}} \frac{1}{\delta^2} \sum_{j=m+1}^{n} E(X_i^2) \\ &= \lim_{\substack{n\to\infty\\ m\to\infty}} \frac{1}{\delta^2} \sum_{j=m+1}^{n} \mathrm{Var}(X_i) = 0. \end{aligned}$$

Therefore

$$\lim_{\substack{n\to\infty\\ m\to\infty}} P\{\sup_{m<k\le n} |S_k - S_m| \ge \delta\} \le 0.$$

We can also prove convergence in probability

$$\lim_{\substack{n\to\infty\\ m\to\infty}} P\{|S_n - S_m| \ge \delta\} = 0$$

by a simple application of Chebychev's inequality and then apply Lévy's Theorem to get almost sure convergence. □

Theorem 3.11. (Kolomogorov's two series theorem). *Let $a_i = E[X_i]$ be the means and $\sigma_i^2 = \mathrm{Var}(X_i)$ the variances of a sequence of independent random variables $\{X_i\}$. Assume that $\sum_i a_i$ and $\sum_i \sigma_i^2$ converge. Then the series $\sum_i X_i$ converges with probability 1.*

Proof. Define $Y_i = X_i - a_i$ and apply the previous (one series) theorem to Y_i. □

Of course in general random variables need not have finite expectations or variances. If $\{X_i\}$ is any sequence of random variables we can take a cut off value C and define $Y_i = X_i$ if $|X_i| \le C$, and $Y_i = 0$ otherwise. The Y_i are then independent and bounded in absolute value by C. The theorem can be applied to Y_i and if we impose the additional condition that

$$\sum_i P\{X_i \ne Y_i\} = \sum_i P\{|X_i| > C\} < \infty$$

by an application of Borel-Cantelli Lemma, with Probabilty 1, $X_i = Y_i$ for all sufficiently large i. The convergence of $\sum_i X_i$ and $\sum_i Y_i$ are therefore equivalent. We get then the sufficiency part of

Theorem 3.12. (Kolmogorov's three series theorem). *For the convergence of an infinite series of independent random variables $\sum_i X_i$ it is necessary and sufficient that all the three following infinite series converge.*

(i) For some cut off value $C > 0$, $\sum_i P\{|X_i| > C\}$ converges.

(ii) If Y_i is defined to equal X_i if $|X_i| \le C$, and 0 otherwise, $\sum_i E(Y_i)$ converges.

(iii) With Y_i as in (ii), $\sum_i \mathrm{Var}(Y_i)$ converges.

Proof. Let us now prove the converse. If $\sum_i X_i$ converges for a sequence of independent random variables, we must necessarily have $|X_n| \le C$ eventually with probability 1. By Borel-Cantelli Lemma the first series must converge. This means that in order to prove the necessity we can assume without loss of generality that $|X_i|$ are all bounded say by 1. we may also assume that $E(X_i) = 0$ for each i. Otherwise let us take independent random variables X_i' that have the same distribution as X_i. Then $\sum_i X_i$ as well as $\sum_i X_i'$ converge with probability 1 and therefore so does $\sum_i (X_i - X_i')$. The random variables $Z_i = X_i - X_i'$ are independent and bounded by 2. They have mean 0. If we can show $\sum \mathrm{Var}(Z_i)$ is convergent, since $\mathrm{Var}(Z_i) = 2\mathrm{Var}(X_i)$ we would have proved the convergence of the the third series. Now it is elementary to conclude that since both $\sum_i X_i$ as well as $\sum_i (X_i - E(X_i))$ converge, the series $\sum_i E(X_i)$ must be convergent as well. So all we need is the following lemma to complete the proof of necessity. $\square$

Lemma 3.13. *If $\sum_i X_i$ is convergent for a series of independent random variables with mean 0 that are individually bounded by C, then $\sum_i \mathrm{Var}(X_i)$ is convergent.*

Proof. Let $F_n = \{\omega : |S_1| \le \ell, |S_2| \le \ell, \cdots, |S_n| \le \ell\}$ where $S_k = X_1 + \cdots + X_k$. If the series converges with probablity 1, we must have, for some ℓ and $\delta > 0$, $P(F_n) \ge \delta$ for all n. We have

$$\begin{aligned}
\int_{F_{n-1}} S_n^2 \, dP &= \int_{F_{n-1}} [S_{n-1} + X_n]^2 \, dP \\
&= \int_{F_{n-1}} [S_{n-1}^2 + 2S_{n-1}X_n + X_n^2] \, dP \\
&= \int_{F_{n-1}} S_{n-1}^2 \, dP + \sigma_n^2 P(F_{n-1}) \\
&\ge \int_{F_{n-1}} S_{n-1}^2 \, dP + \delta\sigma_n^2
\end{aligned}$$

and on the other hand,

$$\begin{aligned}
\int_{F_{n-1}} S_n^2 \, dP &= \int_{F_n} S_n^2 \, dP + \int_{F_{n-1} \cap F_n^c} S_n^2 \, dP \\
&\le \int_{F_n} S_n^2 \, dP + P(F_{n-1} \cap F_n^c) \, (\ell + C)^2
\end{aligned}$$

providing us with the estimate

$$\delta\sigma_n^2 \leq \int_{F_n} S_n^2 \, dP - \int_{F_{n-1}} S_{n-1}^2 \, dP + P(F_{n-1} \cap F_n^c) \, (\ell + C)^2.$$

Since $F_{n-1} \cap F_n^c$ are disjoint and $|S_n| \leq \ell$ on F_n ,

$$\sum_{j=1}^{\infty} \sigma_j^2 \leq \frac{1}{\delta} [\ell^2 + (\ell + C)^2].$$

This concludes the proof. □

3.5 Strong Law of Large Numbers

We saw earlier that in Theorem 3.6 that if $\{X_i\}$ is sequence of i.i.d. (independent identically distributed) random variables with zero mean and a finite fourth moment then $\frac{X_1+\cdots+X_n}{n} \to 0$ with probability 1. We will now prove the same result assuming only that $E|X_i| < \infty$ and $E(X_i) = 0$.

Theorem 3.14. *If $\{X_i\}$ is a sequence of i.i.d random variables with mean* 0*,*

$$\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = 0$$

with probability 1*.*

Proof. We define

$$Y_n = \begin{cases} X_n & \text{if } |X_n| \leq n \\ 0 & \text{if } |X_n| > n \end{cases}$$

$$a_n = P[X_n \neq Y_n], \quad b_n = E[Y_n] \text{ and } c_n = \mathrm{Var}(Y_n).$$

First we note that (see exercise 3.14 below)

$$\sum_n a_n = \sum_n P[|X_1| > n] \leq E|X_1| < \infty$$

$$\lim_{n\to\infty} b_n = 0$$

and

$$\begin{aligned}\sum_n \frac{c_n}{n^2} \le \sum_n \frac{E[Y_n^2]}{n^2} &= \sum_n \int_{|x|\le n} \frac{x^2}{n^2}\, d\alpha \\ &= \int x^2 \Big(\sum_{n\ge x} \frac{1}{n^2}\Big)\, d\alpha \le C \int |x|\, d\alpha < \infty\end{aligned}$$

where α is the common distribution of X_i. From the three series theorem and the Borel-Cantelli Lemma, we conclude that $\sum_n \frac{Y_n - b_n}{n}$ as well as $\sum_n \frac{X_n - b_n}{n}$ converge almost surely. It is elementary to verify that for any series $\sum_n \frac{x_n}{n}$ that converges, $\frac{x_1+\cdots+x_n}{n} \to 0$ as $n \to \infty$. We therefore conclude that

$$P\Big\{\lim_{n\to\infty} \big[\frac{X_1+\cdots+X_n}{n} - \frac{b_1+\cdots+b_n}{n}\big] = 0\Big\} = 1$$

Since $b_n \to 0$ as $n \to \infty$, the theorem is proved. □

Exercise 3.14. Let X be a nonnegative random variable. Then

$$E[X] - 1 \le \sum_{n=1}^{\infty} P[X_n \ge n] \le E[X]$$

In particular $E[X] < \infty$ if and only if $\sum_n P[X \ge n] < \infty$.

Exercise 3.15. If for a sequence of i.i.d. random variables $X_1, \cdots, X_n, \cdots$, the strong law of large numbers holds with some limit, i.e.

$$P[\lim_{n\to\infty} \frac{S_n}{n} = \xi\,] = 1$$

for some random variable ξ, which may or may not be a constant with probability 1, then show that necessarily $E|X_i| < \infty$. Consequently $\xi = E(X_i)$ with probabilty 1.

One may ask why the limit cannot be a proper random variable. There is a general theorem that forbids it called Kolmogorov's Zero-One law. Let us look at the space Ω of real sequences $\{x_n : n \ge 1\}$. We have the σ-field $\mathcal{B}$, the product σ-field on Ω. In addition we have the sub σ-fields $\mathcal{B}^n$ generated by $\{x_j : j \ge n\}$. $\mathcal{B}^n$ are $\downarrow$ with n and $\mathcal{B}^\infty = \cap_n \mathcal{B}^n$ which is also a σ-field is called the tail σ-field. The typical set in $\mathcal{B}^\infty$ is a set depending only on the tail behavior of the sequence. For example the sets $\{\omega : x_n \text{ is bounded }\}$, $\{\omega : \limsup_n x_n = 1\}$ are in $\mathcal{B}^\infty$ whereas $\{\omega : \sup_n |x_n| = 1\}$ is not.

Theorem 3.15. (Kolmogorov's Zero-One Law). *If $A \in \mathcal{B}^\infty$ and P is any product measure (not necessarily with identical components) $P(A) = 0$ or 1.*

Proof. The proof depends on showing that A is independent of itself so that $P(A) = P(A \cap A) = P(A)P(A) = [P(A)]^2$ and therefore equals 0 or 1. The proof is elementary. Since $A \in \mathcal{B}^\infty \subset \mathcal{B}^{n+1}$ and P is a product measure, A is independent of $\mathcal{B}_n = \sigma$-field generated by $\{x_j : 1 \le j \le n\}$. It is therefore independent of sets in the field $\mathcal{F} = \cup_n \mathcal{B}_n$. The class of sets $\mathcal{A}$ that are independent of A is a monotone class. Since it contains the field $\mathcal{F}$ it contains the σ-field $\mathcal{B}$ generated by $\mathcal{F}$. In particular since $A \in \mathcal{B}$, A is independent of itself. □

Corollary 3.16. *Any random variable measurable with respect to the tail σ-field $\mathcal{B}^\infty$ is equal with probaility 1 to a constant relative to any given product measure.*

Proof. Left as an exercise. □

Warning. For different product measures the constants can be different.

Exercise 3.16. How can that happen?

3.6 Central Limit Theorem.

We saw before that for any sequence of independent identically distributed random variables $X_1, \cdots, X_n, \cdots$ the sum $S_n = X_1 + \cdots + X_n$ has the property that

$$\lim_{n\to\infty} \frac{S_n}{n} = 0$$

in probability provided the expectation exists and equals 0. If we assume that the Variance of the random variables is finite and equals $\sigma^2 > 0$, then we have

Theorem 3.17. *The distribution of $\frac{S_n}{\sqrt{n}}$ converges as $n \to \infty$ to the normal distribution with density*

$$p(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp[-\frac{x^2}{\sigma^2}]. \tag{3.11}$$

Proof. If we denote by $\phi(t)$ the characteristic function of any X_i then the characteristic function of $\frac{S_n}{\sqrt{n}}$ is given by

$$\psi_n(t) = [\phi(\frac{t}{\sqrt{n}})]^n$$

We can use the expansion

$$\phi(t) = 1 - \frac{\sigma^2 t^2}{2} + o\,(t^2)$$

to conclude that

$$\phi(\frac{t}{\sqrt{n}}) = 1 - \frac{\sigma^2 t^2}{2n} + o\,(\frac{1}{n})$$

and it then follows that

$$\lim_{n\to\infty} \psi_n(t) = \psi(t) = \exp[-\frac{\sigma^2 t^2}{2}].$$

Since $\psi(t)$ is the characteristic function of the normal distribution with density $p(x)$ given by equation (3.11), we are done. □

Exercise 3.17. A more direct proof is possible in some special cases. For instance if each $X_i = \pm 1$ with probability $\frac{1}{2}$, S_n can take the values $n - 2k$ with $0 \le k \le n$,

$$P[S_n = 2k - n] = \frac{1}{2^n}\binom{n}{k}$$

and

$$P[a \le \frac{S_n}{\sqrt{n}} \le b] = \frac{1}{2^n} \sum_{k: a\sqrt{n} \le 2k-n \le b\sqrt{n}} \binom{n}{k}.$$

Use Stirling's formula to prove directly that

$$\lim_{n\to\infty} P[a \le \frac{S_n}{\sqrt{n}} \le b] = \int_a^b \frac{1}{\sqrt{2\pi}} \exp[-\frac{x^2}{2}]\, dx.$$

Actually for the proof of the central limit theorem we do not need the random variables $\{X_j\}$ to have identical distributions. Let us suppose that they all have zero means and that the variance of X_j is σ_j^2. Define $s_n^2 =$

$\sigma_1^2 + \cdots + \sigma_n^2$. Assume $s_n^2 \to \infty$ as $n \to \infty$. Then $Y_n = \frac{S_n}{s_n}$ has zero mean and unit variance. It is not unreasonable to expect that

$$\lim_{n\to\infty} P[Y_n \le a] = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} \exp[-\frac{x^2}{2}]\, dx$$

under certain mild conditions.

Theorem 3.18. (Lindeberg's theorem). *If we denote by α_i the distribution of X_i, the condition (known as Lindeberg's condition)*

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=1}^{n} \int_{|x|\ge \epsilon s_n} x^2 d\alpha_i = 0$$

for each $\epsilon > 0$ is sufficient for the central limit theorem to hold.

Proof. The first step in proving this limit theorem as well as other limit theorems that we will prove is to rewrite

$$Y_n = X_{n,1} + X_{n,2} + \cdots + X_{n,k_n} + A_n$$

where $X_{n,j}$ are k_n mutually independent random variables and A_n is a constant. In our case $k_n = n$, $A_n = 0$, and $X_{n,j} = \frac{X_j}{s_n}$ for $1 \le j \le n$. We denote by

$$\phi_{n,j}(t) = E[e^{i\,t\,X_{n,j}}] = \int e^{i\,t\,x} d\alpha_{n,j} = \int e^{i\,t\,\frac{x}{s_n}} d\alpha_j = \phi_j(\frac{t}{s_n})$$

where $\alpha_{n,j}$ is the distribution of $X_{n,j}$. The functions ϕ_j and $\phi_{n,j}$ are the characteristic functions of α_j and $\alpha_{n,j}$ respectively. If we denote by μ_n the distribution of Y_n, its characteristic function $\hat{\mu}_n(t)$ is given by

$$\hat{\mu}_n(t) = \prod_{j=1}^{n} \phi_{n,j}(t)$$

and our goal is to show that

$$\lim_{n\to\infty} \hat{\mu}_n(t) = \exp[-\frac{t^2}{2}].$$

This will be carried out in several steps. First, we define

$$\psi_{n,j}(t) = \exp[\phi_{n,j}(t) - 1]$$

and

$$\psi_n(t) = \prod_{j=1}^{n} \psi_{n,j}(t).$$

We show that for each finite T,

$$\lim_{n\to\infty} \sup_{|t|\le T} \sup_{1\le j\le n} |\phi_{n,j}(t) - 1| = 0$$

and

$$\sup_n \sup_{|t|\le T} \sum_{j=1}^{n} |\phi_{n,j}(t) - 1| < \infty.$$

This would imply that

$$\begin{aligned}
&\lim_{n\to\infty} \sup_{|t|\le T} \Big| \log \hat{\mu}_n(t) - \log \psi_n(t) \Big| \\
&\quad \le \lim_{n\to\infty} \sup_{|t|\le T} \sum_{j=1}^{n} \Big| \log \phi_{n,j}(t) - [\phi_{n,j}(t) - 1] \Big| \\
&\quad \le \lim_{n\to\infty} \sup_{|t|\le T} C \sum_{j=1}^{n} |\phi_{n,j}(t) - 1|^2 \\
&\quad \le C \lim_{n\to\infty} \Big\{ \sup_{|t|\le T} \sup_{1\le j\le n} |\phi_{n,j}(t) - 1| \Big\} \Big\{ \sup_{|t|\le T} \sum_{j=1}^{n} |\phi_{n,j}(t) - 1| \Big\} \\
&\quad = 0
\end{aligned}$$

by the expansion

$$\log r = \log(1 + (r-1)) = r - 1 + O(r-1)^2.$$

The proof can then be completed by showing

$$\lim_{n\to\infty} \sup_{|t|\le T} \Big| \log \psi_n(t) + \frac{t^2}{2} \Big| = \lim_{n\to\infty} \sup_{|t|\le T} \left| \left[\sum_{j=1}^{n} (\phi_{n,j}(t) - 1) \right] + \frac{t^2}{2} \right| = 0.$$

We see that

$$
\begin{aligned}
\sup_{|t|\le T}\big|\phi_{n,j}(t)-1\big| &= \sup_{|t|\le T}\left|\int\big[\exp[i\,t\,x\,]-1\big]d\alpha_{n,j}\right| \\
&= \sup_{|t|\le T}\left|\int\big[\exp[i\,t\,\frac{x}{s_n}\,]-1\big]d\alpha_j\right| \\
&= \sup_{|t|\le T}\left|\int\big[\exp[i\,t\,\frac{x}{s_n}\,]-1-i\,t\,\frac{x}{s_n}\big]d\alpha_j\right| \qquad (3.12)\\
&\le C_T\int\frac{x^2}{s_n^2}d\alpha_j \qquad (3.13)\\
&= C_T\int_{|x|<\epsilon s_n}\frac{x^2}{s_n^2}\,d\alpha_j + C_T\int_{|x|\ge\epsilon s_n}\frac{x^2}{s_n^2}\,d\alpha_j \\
&\le C_T\epsilon^2 + C_T\frac{1}{s_n^2}\int_{|x|\ge\epsilon s_n}x^2 d\alpha_j. \qquad (3.14)
\end{aligned}
$$

We have used the mean zero condition in deriving equation 3.12 and the estimate $|e^{ix}-1-ix|\le cx^2$ to get to the equation 3.13. If we let $n\to\infty$, by Lindeberg's condition, the second term of equation (3.14) goes to 0. Therefore

$$\limsup_{n\to\infty}\sup_{1\le j\le k_n}\sup_{|t|\le T}\big|\phi_{n,j}(t)-1\big|\le\epsilon^2 C_T.$$

Since, $\epsilon>0$ is arbitrary, we have

$$\lim_{n\to\infty}\sup_{1\le j\le k_n}\sup_{|t|\le T}\big|\phi_{n,j}(t)-1\big|=0.$$

Next we observe that there is a bound,

$$\sup_{|t|\le T}\sum_{j=1}^{n}\big|\phi_{n,j}(t)-1\big|\le C_T\sum_{j=1}^{n}\int\frac{x^2}{s_n^2}d\alpha_j\le C_T\frac{1}{s_n^2}\sum_{j=1}^{n}\sigma_j^2=C_T$$

uniformly in n. Finally for each $\epsilon>0$,

$$
\begin{aligned}
&\lim_{n\to\infty}\sup_{|t|\le T}\left|\left[\sum_{j=1}^{n}(\phi_{n,j}(t)-1)\right]+\frac{t^2}{2}\right| \\
&\quad\le \lim_{n\to\infty}\sup_{|t|\le T}\sum_{j=1}^{n}\left|\phi_{n,j}(t)-1+\frac{\sigma_j^2 t^2}{2s_n^2}\right| \\
&\quad= \lim_{n\to\infty}\sup_{|t|\le T}\sum_{j=1}^{n}\left|\int\left[\exp[i\,t\,\frac{x}{s_n}]-1-\,i\,t\,\frac{x}{s_n}+\frac{t^2x^2}{2s_n^2}\right]d\alpha_j\right| \\
&\quad\le \lim_{n\to\infty}\sup_{|t|\le T}\sum_{j=1}^{n}\left|\int_{|x|<\epsilon s_n}\left[\exp[i\,t\,\frac{x}{s_n}]-1-\,i\,t\,\frac{x}{s_n}+\frac{t^2x^2}{2s_n^2}\right]d\alpha_j\right| \\
&\qquad+ \lim_{n\to\infty}\sup_{|t|\le T}\sum_{j=1}^{n}\left|\int_{|x|\ge\epsilon s_n}\left[\exp[i\,t\,\frac{x}{s_n}]-1-\,i\,t\,\frac{x}{s_n}+\frac{t^2x^2}{2s_n^2}\right]d\alpha_j\right| \\
&\quad\le \lim_{n\to\infty}C_T\sum_{j=1}^{n}\int_{|x|<\epsilon s_n}\frac{|x|^3}{s_n^3}\,d\alpha_j \\
&\qquad+ \lim_{n\to\infty}C_T\sum_{j=1}^{n}\int_{|x|\ge\epsilon s_n}\frac{x^2}{s_n^2}\,d\alpha_j \\
&\quad\le \epsilon C_T\limsup_{n\to\infty}\sum_{j=1}^{n}\int\frac{x^2}{s_n^2}\,d\alpha_j \\
&\qquad+ \lim_{n\to\infty}C_T\sum_{j=1}^{n}\int_{|x|\ge\epsilon s_n}\frac{x^2}{s_n^2}\,d\alpha_j \\
&\quad= \epsilon C_T
\end{aligned}
$$

by Lindeberg's condition. Since $\epsilon > 0$ is arbitrary the result is proven. □

Remark 3.3. The key step in the proof of the central limit theorem under Lindeberg's condition, as well as in other limit theorems for sums of independent random variables, is the analysis of products

$$\psi_n(t) = \Pi_{j=1}^{k_n}\phi_{n,j}(t).$$

The idea is to replace each $\phi_{n,j}(t)$ by $\exp[\phi_{n,j}(t)-1]$, changing the product to the exponential of a sum. Although each $\phi_{n,j}(t)$ is close to 1, making the idea reasonable, in order for the idea to work one has to show that the sum $\sum_{j=1}^{k_n}|\phi_{n,j}(t)-1|^2$ is negligible. This requires the boundedness of $\sum_{j=1}^{k_n}|\phi_{n,j}(t)-1|$. One has to use the mean 0 condition or some suitable centering condition to cancel the first term in the expansion of $\phi_{n,j}(t)-1$ and control the rest from sums of the variances.

Exercise 3.18. Lyapunov's condition is the following: for some $\delta>0$

$$\lim_{n\to\infty}\frac{1}{s_n^{2+\delta}}\sum_{j=1}^{n}\int|x|^{2+\delta}\,d\alpha_j=0.$$

Prove that Lyapunov's condition implies Lindeberg's condition.

Exercise 3.19. Consider the case of mutually independent random variables $\{X_j\}$, where $X_j=\pm a_j$ with probability $\frac{1}{2}$. What do Lyapunov's and Lindeberg's conditions demand of $\{a_j\}$? Can you find a sequence $\{a_j\}$ that does not satisfy Lyapunov's condition for any $\delta>0$ but satisfies Lindeberg's condition? Try to find a sequence $\{a_j\}$ such that the central limit theorem is not valid.

3.7 Accompanying Laws.

As we stated in the previous section, we want to study the behavior of the sum of a large number of independent random variables. We have k_n independent random variables $\{X_{n,j}:1\le j\le k_n\}$ with respective distributions $\{\alpha_{n,j}\}$. We are interested in the distribution μ_n of $Z_n=\sum_{j=1}^{k_n}X_{n,j}$. One important assumption that we will make on the random variables $\{X_{n,j}\}$ is that no single one is significant. More precisely for every $\delta>0$,

$$\lim_{n\to\infty}\sup_{1\le j\le k_n}P[\,|X_{n,j}|\ge\delta\,]=\lim_{n\to\infty}\sup_{1\le j\le k_n}\alpha_{n,j}[\,|x|\ge\delta\,]=0. \tag{3.15}$$

The condition is referred to as *uniform infinitesimality*. The following construction will play a major role. If α is a probability distribution on the line and $\phi(t)$ is its characteristic function, for any nonnegative real number $a>0$, $\psi_a(t)=\exp[a(\phi(t)-1)]$ is again a characteristic distribution. In fact, if we denote by α^k the k-fold convolution of α with itself, ψ_a is seen to be the characteristic function of the probability distribution

$$e^{-a}\sum_{j=0}^{\infty}\frac{a^j}{j!}\alpha^j$$

which is a convex combination α^j with weights $e^{-a}\,\frac{a^j}{j!}$. We use the construction mostly with $a = 1$. If we denote the probability distribution with characteristic function $\psi_a(t)$ by $e_a(\alpha)$ one checks easily that $e_{a+b}(\alpha) = e_a(\alpha) * e_b(\alpha)$. In particular $e_a(\alpha) = e_{\frac{a}{n}}(\alpha)^n$. Probability distributions β that can be written for each $n \geq 1$ as the n-fold convolution β_n^n of some probability distribution β_n are called infinitely divisible. In particular for every $a \geq 0$ and α, $e_a(\alpha)$ is an infinitely divisible probability distribution. These are called compound Poisson distributions. A special case when $\alpha = \delta_1$ the degenerate distribution at 1, we get for $e_a(\delta_1)$ the usual Poisson distribution with parameter a. We can interpret $e_a(\alpha)$ as the distribution of the sum of a random number of independent random variables with common distribution α. The random n has a distribution which is Poisson with parameter a and is independent of the random variables involved in the sum.

In order to study the distribution μ_n of Z_n it will be more convenient to replace $\alpha_{n,j}$ by an infinitely divisible distribution $\beta_{n,j}$. This is done as follows. We define

$$a_{n,j} = \int_{|x|\leq 1} x\, d\alpha_{n,j},$$

$\alpha'_{n,j}$ as the translate of $\alpha_{n,j}$ by $-a_{n,j}$, i.e.

$$\alpha'_{n,j} = \alpha_{n,j} * \delta_{-a_{n,j}},$$
$$\beta'_{n,j} = e_1(\alpha_{n,j}),$$
$$\beta_{n,j} = \beta'_{n,j} * a_{n,j}$$

and finally

$$\lambda_n = \prod_{j=1}^{k_n}\beta_{n,j}$$

A main tool in this subject is the following theorem. We assume always that the uniform infinitesimality condition (3.15) holds. In terms of notation, we will find it more convenient to denote by $\hat{\mu}$ the characteristic function of the probability distribution μ.

Theorem 3.19. (Accompanying Laws.) *In order that, for some constants A_n, the distribution $\mu_n * \delta_{A_n}$ of $Z_n + A_n$ may converge to the limit μ it is necessary and sufficient that, for the same constants A_n, the distribution $\lambda_n * \delta_{A_n}$ converges to the same limit μ.*

Proof. First we note that, for any $\delta > 0$,

$$\begin{aligned}
\limsup_{n\to\infty} \sup_{1\le j\le k_n} |a_{n,j}| &= \limsup_{n\to\infty} \sup_{1\le j\le k_n} \left| \int_{|x|\le 1} x \, d\alpha_{n,j} \right| \\
&\le \limsup_{n\to\infty} \sup_{1\le j\le k_n} \left| \int_{|x|\le \delta} x \, d\alpha_{n,j} \right| \\
&\quad + \limsup_{n\to\infty} \sup_{1\le j\le k_n} \left| \int_{\delta<|x|\le 1} x \, d\alpha_{n,j} \right| \\
&\le \delta + \limsup_{n\to\infty} \sup_{1\le j\le k_n} \alpha_{n,j}[\,|x| \ge \delta\,] \\
&= \delta.
\end{aligned}$$

Therefore

$$\lim_{n\to\infty} \sup_{1\le j\le k_n} |a_{n,j}| = 0.$$

This means that $\alpha'_{n,j}$ are uniformly infinitesimal just as $\alpha_{n,j}$ were. Let us suppose that n is so large that $\sup_{1\le j\le k_n} |a_{n,j}| \le \frac{1}{4}$. The advantage in going from $\alpha_{n,j}$ to $\alpha'_{n,j}$ is that the latter are better centered and we can calculate

$$\begin{aligned}
a'_{n,j} &= \int_{|x|\le 1} x \, d\alpha'_{n,j} \\
&= \int_{|x-a_{n,j}|\le 1} (x - a_{n,j}) \, d\alpha_{n,j} \\
&= \int_{|x-a_{n,j}|\le 1} x \, d\alpha_{n,j} - a_{n,j}\, \alpha_{n,j}[\,|x - a_{n,j}| \le 1\,] \\
&= \int_{|x-a_{n,j}|\le 1} x \, d\alpha_{n,j} - a_{n,j} + \alpha_{n,j}[\,|x - a_{n,j}| > 1\,]
\end{aligned}$$

and estimate $|a'_{n,j}|$ by

$$|a'_{n,j}| \le C\alpha_{n,j}[\,|x| \ge \frac{3}{4}\,] \le C\alpha'_{n,j}[\,|x| \ge \frac{1}{2}\,].$$

In other words we may assume without loss of generality that $\alpha_{n,j}$ satisfy the bound

$$|a_{n,j}| \le C\alpha_{n,j}[\,|x| \ge \frac{1}{2}\,] \tag{3.16}$$

and forget all about the change from $\alpha_{n,j}$ to $\alpha'_{n,j}$. We will drop the primes and stay with just $\alpha_{n,j}$. Then, just as in the proof of the Lindeberg theorem, we proceed to estimate

$$\begin{aligned}
\lim_{n\to\infty} \sup_{|t|\le T} \big|\log \hat{\lambda}_n(t) - \log \hat{\mu}_n(t)\big| & \\
&\le \lim_{n\to\infty} \sup_{|t|\le T} \Big|\sum_{j=1}^{k_n} \big[\log \hat{\alpha}_{n,j}(t) - (\hat{\alpha}_{n,j}(t)-1)\big]\Big| \\
&\le \lim_{n\to\infty} \sup_{|t|\le T} \sum_{j=1}^{k_n} \big|\log \hat{\alpha}_{n,j}(t) - (\hat{\alpha}_{n,j}(t)-1)\big| \\
&\le \lim_{n\to\infty} \sup_{|t|\le T} C\sum_{j=1}^{k_n} |\hat{\alpha}_{n,j}(t)-1|^2 \\
&= 0.
\end{aligned}$$

provided we prove that if either λ_n or μ_n has a limit after translation by some constants A_n, then

$$\sup_n \sup_{|t|\le T} \sum_{j=1}^{k_n} \big|\hat{\alpha}_{n,j}(t) - 1\big| \le C < \infty. \tag{3.17}$$

Let us first suppose that λ_n has a weak limit as $n \to \infty$ after translation by A_n. The characteristic functions

$$\exp\Big[\sum_{j=1}^{k_n} (\hat{\alpha}_{n,j}(t) - 1)) + itA_n\Big] = \exp[f_n(t)]$$

have a limit, which is again a characteristic function. Since the limiting characteristic function is continuous and equals 1 at $t = 0$, and the convergence is uniform near 0, on some small interval $|t| \le T_0$ we have the bound

$$\sup_n \sup_{|t|\le T_0} \big[1 - \operatorname{Re} f_n(t)\big] \le C$$

or equivalently

$$\sup_n \sup_{|t|\le T_0} \sum_{j=1}^{k_n} \int (1-\cos t\,x\,)\,d\alpha_{n,j} \le C$$

and from the subadditivity property $(1-\cos 2\,t\,x\,) \le 4(1-\cos\, t\,x)$ this bound extends to arbitrary interval $|t| \le T$,

$$\sup_n \sup_{|t|\le T} \sum_{j=1}^{k_n} \int (1-\cos t\,x\,)\,d\alpha_{n,j} \le C_T.$$

If we integrate the inequality with respect to t over the interval $[-T, T]$ and divide by $2T$, we get

$$\sup_n \sum_{j=1}^{k_n} \int (1-\frac{\sin T\,x}{Tx})\,d\alpha_{n,j} \le C_T$$

from which we can conclude that

$$\sup_n \sum_{j=1}^{k_n} \alpha_{n,j}[\,|x| \ge \delta\,] \le C_\delta < \infty$$

for every $\delta > 0$ by choosing $T = \frac{2}{\delta}$. Moreover using the inequality $(1-\cos x) \ge c\,x^2$ valid near 0 for a suitable choice of c we get the estimate

$$\sup_n \sum_{j=1}^{k_n} \int_{|x|\le 1} x^2\,d\alpha_{n,j} \le C < \infty.$$

Now it is straight forward to estimate, for $t \in [-T, T]$,

$$\begin{aligned}
|\hat{\alpha}_{n,j}(t) - 1| &= \left| \int [\exp(i\,t\,x\,) - 1]\, d\alpha_{n,j} \right| \\
&= \left| \int_{|x|\le 1} [\exp(i\,t\,x\,) - 1]\, d\alpha_{n,j} \right| \\
&\qquad + \left| \int_{|x|>1} [\exp(i\,t\,x\,) - 1]\, d\alpha_{n,j} \right| \\
&\le \left| \int_{|x|\le 1} [\exp(i\,t\,x\,) - 1 - i\,t\,x]\, d\alpha_{n,j} \right| \\
&\qquad + \left| \int_{|x|>1} [\exp(i\,t\,x\,) - 1]\, d\alpha_{n,j} \right| + T\,|a_{n,j}| \\
&\le C_1 \int_{|x|\le 1} x^2\, d\alpha_{n,j} + C_2\, \alpha_{n,j}\, [x : |x| \ge \frac{1}{2}\,]
\end{aligned}$$

which proves the bound of equation (3.17).

Now we need to establish the same bound under the assumption that μ_n has a limit after suitable translations. For any probability measure α we define $\bar{\alpha}$ by $\bar{\alpha}(A) = \alpha(-A)$ for all Borel sets. The distribution $\alpha * \bar{\alpha}$ is denoted by $|\alpha|^2$. The characteristic functions of $\bar{\alpha}$ and $|\alpha|^2$ are respectively $\bar{\hat{\alpha}}(t)$ and $|\hat{\alpha}(t)|^2$ where $\hat{\alpha}(t)$ is the characteristic function of α. An elementary but important fact is $|\alpha * A|^2 = |\alpha|^2$ for any translate A. If μ_n has a limit so does $|\mu_n|^2$. We conclude that the limit

$$\lim_{n\to\infty} |\hat{\mu}_n(t)|^2 = \lim_{n\to\infty} \prod_{j=1}^{k_n} |\hat{\alpha}_{n,j}(t)|^2$$

exists and defines a characteristic function which is continuous at 0 with a value of 1. Moreover because of uniform infinitesimality,

$$\lim_{n\to\infty} \inf_{|t|\le T} |\hat{\alpha}_{n,j}(t)| = 1.$$

It is easy to conclude that there is a $T_0 > 0$ such that, for $|t| \le T_0$,

$$\sup_n \sup_{|t|\le T_0} \sum_{j=1}^{k_n} [1 - |\hat{\alpha}_{n,j}(t)|^2] \le C_0 < \infty$$

and by subadditivity for any finite T,

$$\sup_n \sup_{|t|\le T} \sum_{j=1}^{k_n} [1 - |\hat{\alpha}_{n,j}(t)|^2] \le C_T < \infty$$

providing us with the estimates

$$\sup_n \sum_{j=1}^{k_n} |\alpha_{n,j}|^2 \, [\, |x| \ge \delta \,] \le C_\delta < \infty \tag{3.18}$$

for any $\delta > 0$, and

$$\sup_n \sum_{j=1}^{k_n} \int \int_{|x-y|\le 2} (x-y)^2 d\alpha_{n,j}(x) \, d\alpha_{n,j}(y) \le C < \infty. \tag{3.19}$$

We now show that estimates (3.18) and (3.19) imply (3.17)

$$\begin{aligned}
|\alpha_{n,j}|^2 [\, x : |x| \ge \frac{\delta}{2} \,] &\ge \int_{|y|\le\frac{\delta}{2}} \alpha_{n,j} \, [x : |x-y| \ge \frac{\delta}{2} \,] \, d\alpha_{n,j}(y) \\
&\ge \alpha_{n,j} \, [\, x : |x| \ge \delta \,] \, \alpha_{n,j} \, [\, x : |x| \le \frac{\delta}{2} \,] \\
&\ge \frac{1}{2} \, \alpha_{n,j} \, [\, x : |x| \ge \delta \,]
\end{aligned}$$

by uniform infinitesimality. Therfore 3.18 implies that for every $\delta > 0$,

$$\sup_n \sum_{j=1}^{k_n} \alpha_{n,j} \, [\, x : |x| \ge \delta \,] \le C_\delta < \infty. \tag{3.20}$$

We now turn to exploiting (3.19). We start with the inequality

$$\begin{aligned}
&\int \int_{|x-y|\le 2} (x-y)^2 d\alpha_{n,j}(x) \, d\alpha_{n,j}(y) \\
\ge \quad &\Big\{ \alpha_{n,j} \, [y : |y| \le 1\,] \Big\} \Big\{ \inf_{|y|\le 1} \int_{|x|\le 1} (x-y)^2 d\alpha_{n,j}(x) \Big\}.
\end{aligned}$$

The first term on the right can be assumed to be at least $\frac{1}{2}$ by uniform infinitesimality. The second term

$$\begin{aligned}\int_{|x|\le 1}(x-y)^2 d\alpha_{n,j}(x) &\ge \int_{|x|\le 1} x^2 d\alpha_{n,j}(x) - 2y\int_{|x|\le 1} x\, d\alpha_{n,j}(x)\\ &\ge \int_{|x|\le 1} x^2 d\alpha_{n,j}(x) - 2\Big|\int_{|x|\le 1} x\, d\alpha_{n,j}(x)\Big|\\ &\ge \int_{|x|\le 1} x^2 d\alpha_{n,j}(x) - C\alpha_{n,j}\,[x : |x| \ge \frac{1}{2}\,].\end{aligned}$$

The last step is a consequence of estimate (3.16) that we showed we could always assume.

$$\int_{|x|\le 1} x\, d\alpha_{n,j}(x) \le C\alpha_{n,j}\,[x : |x| \ge \frac{1}{2}\,]$$

Because of estimate (3.20) we can now assert

$$\sup_n \sum_{j=1}^{k_n} \int_{|x|\le 1} x^2\, d\alpha_{n,j} \le C < \infty. \tag{3.21}$$

One can now derive (3.17) from (3.20) and (3.21) as in the earlier part. □

Exercise 3.20. Let $k_n = n^2$ and $\alpha_{n,j} = \delta_{\frac{1}{n}}$ for $1 \le j \le n^2$. $\mu_n = \delta_n$ and show that without centering $\lambda_n * \delta_{-n}$ converges to a different limit.

3.8 Infinitely Divisible Distributions.

In the study of limit theorems for sums of independent random variables infinitely divisible distributions play a very important role.

Definition 3.5. *A distribution μ is said to be infinitely divisible if for every positive integer n, μ can be written as the n-fold convolution $(\lambda_n *)^n$ of some other probability distribution λ_n.*

Exercise 3.21. Show that the normal distribution with density

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp[-\frac{x^2}{2}]$$

is infinitely divisible.

Exercise 3.22. Show that for any $\lambda \geq 0$, the Poisson distribution with parameter λ

$$p_\lambda(n) = \frac{e^{-n}\lambda^n}{n!} \quad \text{for} \quad n \geq 0$$

is infinitely divisible.

Exercise 3.23. Show that any probabilty distribution supported on a finite set $\{x_1, \ldots, x_k\}$ with

$$\mu[\{x_j\}] = p_j$$

and $p_j \geq 0, \sum_{j=1}^k p_j = 1$ is infinitely divisible if and only if it is degenrate, i.e. $\mu[\{x_j\}] = 1$ for some j.

Exercise 3.24. Show that for any nonnegative finite measure α with total mass a, the distribution

$$e(F) = e^{-a} \sum_{j=0}^{\infty} \frac{(\alpha *)^j}{j!}$$

with characteristic function

$$\widehat{e(F)}(t) = \exp[\int (e^{itx} - 1) d\alpha]$$

is an infinitely divisible distribution.

Exercise 3.25. Show that the convolution of any two infinitely divisible distributions is again infinitely divisible. In particular if μ is infinitely divisible so is any translate $\mu * \delta_a$ for any real a.

We saw in the last section that the asymptotic behavior of $\mu_n * \delta_{A_n}$ can be investigated by means of the asymptotic behavior of $\lambda_n * \delta_{A_n}$ and the characteristic function $\hat{\lambda}_n$ of λ_n has a very special form

$$\begin{aligned}
\hat{\lambda}_n &= \prod_{j=1}^{k_n} \exp[\,\hat{\beta}_{n,j}(t) - 1 + i\,t\,a_{n,j}\,] \\
&= \exp\Big[\sum_{j=1}^{k_n} \int [\,e^{i\,t\,x} - 1\,]\,d\beta_{n,j} \,+\, i\,t \sum_{j=1}^{k_n} a_{n,j}\,\Big] \\
&= \exp\Big[\int [\,e^{i\,t\,x} - 1\,]\,dM_n \,+\, i\,t\,a_n\Big] \\
&= \exp\Big[\int [\,e^{i\,t\,x} - 1 \,-\, i\,t\,\theta(x)\,]\,dM_n \,+\, i\,t\,\Big[\int \theta(x)\,dM_n + a_n\,\Big]\Big] \\
&= \exp\Big[\int [\,e^{i\,t\,x} - 1 \,-\, i\,t\,\theta(x)\,]\,dM_n \,+\, i\,t\,b_n\,\Big]. \qquad (3.22)
\end{aligned}$$

We can make any reasonable choice for $\theta(\cdot)$ and we will need it to be a bounded continuous function with

$$|\theta(x) - x| \leq C|x|^3$$

near 0. Possible choices are $\theta(x) = \frac{x}{1+x^2}$, or $\theta(x) = x$ for $|x| \leq 1$ and sign (x) for $|x| \geq 1$. We now investigate when such things will have a weak limit. Convoluting with δ_{A_n} only changes b_n to $b_n + A_n$.

First we note that

$$\hat{\mu}(t) = \exp\big[\int [\,e^{i\,t\,x} - 1 - i\,t\,\theta(x)\,]\,dM + i\,t\,a\big]$$

is a characteristic function for any measure M with finite total mass. In fact it is the characteristic function of an infinitely divisible probability distribution. It is not necessary that M be a finite measure for μ to make sense. M could be infinite, but in such a way that it is finite on $\{x : |x| \geq \delta\}$ for every $\delta > 0$, and near 0 it integrates x^2 i.e.,

$$M[x : |x| \geq \delta] \quad < \infty \qquad \text{for all} \quad \delta > 0, \tag{3.23}$$

$$\int_{|x|\leq 1} x^2 dM \quad < \infty. \tag{3.24}$$

To see this we remark that

$$\hat{\mu}_\delta(t) = \exp\big[\int_{|x|\geq\delta} [\,e^{i\,t\,x} - 1 - i\,t\,\theta(x)\,]\,dM + i\,t\,a\big]$$

is a characteristic function for each $\delta > 0$ and because

$$|e^{i\,t\,x} - 1 - i\ t\,x\,| \leq C_T\,x^2$$

for $|t| \leq T$, $\hat{\mu}_\delta(t) \to \hat{\mu}(t)$ uniformly on bounded intervals where $\hat{\mu}(t)$ is given by the integral

$$\hat{\mu}(t) = \exp\big[\int [\,e^{i\,t\,x} - 1 - i\,t\,\theta(x)\,]\,dM + i\,t\,a\big]$$

which converges absolutely and defines a characteristic function. Let us call measures that satisfy (3.23) and (3.24), that can be expressed in the form

$$\int \frac{x^2}{1+x^2}\, dM < \infty \tag{3.25}$$

admissible *Lévy* measures. Since the same argument applies to $\frac{M}{n}$ and $\frac{a}{n}$ instead of M and a, for any admissible *Lévy* measure M and real number a, $\hat{\mu}(t)$ is in fact an infinitely divisible characteristic function. As the normal distribution is also an infinitely divisible probability distribution, we arrive at the following

Theorem 3.20. *For every admissible Lévy measure M, $\sigma^2 > 0$ and real a*

$$\hat{\mu}(t) = \exp\Big[\int [\, e^{i\,t\,x} - 1 - i\,t\,\theta(x)\,]\, dM + i\,t\,a - \frac{\sigma^2 t^2}{2}\Big]$$

is the characteristic function of an infinitely divisible distribution μ.

We will denote this distribution μ by $\mu = \mathbf{e}\,(M, \sigma^2, a)$. The main theorem of this section is

Theorem 3.21. *In order that $\mu_n = \mathbf{e}\,(M_n, \sigma_n^2, a_n)$ may converge to a limit μ it is necessary and sufficient that $\mu = \mathbf{e}\,(M, \sigma^2, a)$ and the following three conditions (3.26) (3.27) and (3.28) are satisfied.*

For every bounded continuous function f that vanishes in some neighborhood of 0,

$$\lim_{n\to\infty} \int f(x) dM_n = \int f(x) dM. \tag{3.26}$$

For some (and therefore for every) $\ell > 0$ such that $\pm\,\ell$ are continuity points for M, i.e., $M\{\pm\,\ell\} = 0$

$$\lim_{n\to\infty} \Big[\sigma_n^2 + \int_{-\ell}^{\ell} x^2 dM_n\Big] = \Big[\sigma^2 + \int_{-\ell}^{\ell} x^2 dM\Big]. \tag{3.27}$$

$$a_n \to a \text{ as } n \to \infty. \tag{3.28}$$

Proof. Let us prove the sufficiency first. Condition (3.26) implies that for every ℓ such that $\pm\,\ell$ are continuity points of M

$$\lim_{n\to\infty}\int_{|x|\geq\ell}[\,e^{i\,t\,x}-1-i\,t\,\theta(x)\,]\,dM_n=\int_{|x|\geq\ell}[\,e^{i\,t\,x}-1-i\,t\,\theta(x)\,]\,dM$$

and because of condition (3.27), it is enough to show that

$$\begin{aligned}\lim_{\ell\to 0}\limsup_{n\to\infty}\bigg|\int_{-\ell}^{\ell}[\,e^{i\,t\,x} &- 1-i\,t\,\theta(x)+\frac{t^2x^2}{2}\,]\,dM_n\\ &- \int_{\ell}^{\ell}[\,e^{i\,t\,x}-1-i\,t\,\theta(x)+\frac{t^2x^2}{2}\,]\,dM\bigg|\\ &= 0\end{aligned}$$

in order to conclude that

$$\begin{aligned}\lim_{n\to\infty}\bigg[&- \frac{\sigma_n^2t^2}{2}+\int[\,e^{i\,t\,x}-1-i\,t\,\theta(x)]\,dM_n\bigg]\\ &= \bigg[-\frac{\sigma^2t^2}{2}+\int[\,e^{i\,t\,x}-1-i\,t\,\theta(x)]\,dM\bigg].\end{aligned}$$

This follows from the estimates

$$\left|e^{i\,t\,x}-1-i\,t\,\theta(x)+\frac{t^2x^2}{2}\right|\leq C_T\,|x|^3$$

and

$$\int_{-\ell}^{\ell}|x|^3dM_n\leq\ell\int_{-\ell}^{\ell}|x|^2dM_n.$$

Condition (3.28) takes care of the terms involving a_n .

We now turn to proving the necessity. If μ_n has a weak limit μ then the absolute values of the characteristic functions $|\hat{\mu}_n(t)|$ are all uniformly close to 1 near 0. Since

$$|\hat{\mu}_n(t)|=\exp\left[-\int(1-\cos t\,x)\,dM_n-\frac{\sigma_n^2t^2}{2}\right]$$

taking logarithms we conclude that

$$\lim_{t\to 0}\sup_n\left[\frac{\sigma_n t^2}{2}+\int(1-\cos t\,x)\,dM_n\right]=0.$$

This implies (3.29), (3.30) and (3.31)below.

For each $\ell>0$,

$$\sup_n M_n\{x:|x|\geq \ell\}<\infty \tag{3.29}$$

$$\lim_{A\to\infty}\sup_n M_n\{x:|x|\geq A\}=0. \tag{3.30}$$

For every $0\leq \ell<\infty$,

$$\sup_n\left[\sigma_n^2+\int_{-\ell}^{\ell}|x|^2 dM_n\right]<\infty. \tag{3.31}$$

We can choose a subsequence of M_n (which we will denote by M_n as well) that 'converges' in the sense that it satisfies conditions (3.26) and (3.27) of the Theorem. Then $\mathbf{e}\,(M_n,\sigma_n^2,0)$ converges weakly to $\mathbf{e}\,(M,\sigma^2,0)$. It is not hard to see that for any sequence of probability distributions α_n if both α_n and $\alpha_n*\delta_{a_n}$ converge to limits α and β respectively, then necessarily $\beta=\alpha*\delta_a$ for some a and $a_n\to a$ as $n\to\infty$. In order complete the proof of necessity we need only establish the uniqueness of the representation, which is done in the next lemma. ☐

Lemma 3.22. (Uniqueness). *Suppose* $\mu=\mathbf{e}\,(M_1,\sigma_1^2,a_1)=\mathbf{e}\,(M_2,\sigma_2^2,a_2)$, *then* $M_1=M_2$, $\sigma_1^2=\sigma_2^2$ *and* $a_1=a_2$.

Proof. Since $\hat{\mu}(t)$ never vanishes by taking logarithms we have

$$\begin{aligned}\psi(t) &= \left[-\frac{\sigma_1^2 t^2}{2}+\int[\,e^{i\,t\,x}-1-i\,t\,\theta(x)\,]\,dM_1+i\,t\,a_1\right]\\ &= \left[-\frac{\sigma_2^2 t^2}{2}+\int[\,e^{i\,t\,x}-1-i\,t\,\theta(x)\,]\,dM_2+i\,t\,a_2\right].\end{aligned} \tag{3.32}$$

We can verify that for any admissible *Lévy* measure M

$$\lim_{t\to\infty}\frac{1}{t^2}\int[\,e^{i\,t\,x}\,-1\,-\,i\,t\,\theta(x)\,]\,dM=0.$$

Consequently

$$\lim_{t\to\infty}\frac{\psi(t)}{t^2}=\sigma_1^2=\sigma_2^2$$

leaving us with

$$\begin{aligned}\psi(t)\;&=\;\Big[\int[\,e^{i\,t\,x}\,-1\,-\,i\,t\,\theta(x)\,]\,dM_1\,+\,i\,t\,a_1\Big]\\&=\;\Big[\int[\,e^{i\,t\,x}\,-1\,-\,i\,t\,\theta(x)\,]\,dM_2\,+\,i\,t\,a_2\Big]\end{aligned}$$

for a different ψ. If we calculate

$$H(s,t)=\frac{\psi(t+s)+\psi(t-s)}{2}-\psi(t)$$

we get

$$\int e^{i\,t\,x}\,(1-\cos s\,x)dM_1=\int e^{i\,t\,x}\,(1-\cos s\,x)dM_2$$

for all t and s. Since we can and do assume that $M\{0\}=0$ for any admissible *Levy* measure M we have $M_1=M_2$. If we know that $\sigma_1^2=\sigma_2^2$ and $M_1=M_2$ it is easy to see that a_1 must equal a_2. $\square$

Finally

Corollary 3.23. (Lévy-Khintchine representation) *Any infinitely divisible distribution admits a representation* $\mu=\mathbf{e}\,(M,\sigma^2,a)$ *for some admissible Lévy measure* M*,* $\sigma^2>0$ *and real number* a*.*

Proof. We can write $\mu=\mu_n{*}^n=\mu_n*\mu_n*\cdots*\mu_n$ with n terms. If we show that $\mu_n\Rightarrow\delta_0$ then the sequence is uniformly infinitesimal and by the earlier theorem on accompanying laws μ will be the limit of some $\lambda_n=\mathbf{e}\,(M_n,0,a_n)$ and therefore has to be of the form $\mathbf{e}\,(M,\sigma^2,a)$ for some choice of admissible

Levy measure M, $\sigma^2 > 0$ and real a. In a neighborhood around 0, $\hat{\mu}(t)$ is close to 1 and it is easy to check that

$$\hat{\mu}_n(t) = [\hat{\mu}(t)]^{\frac{1}{n}} \to 1$$

as $n \to \infty$ in that neighborhood. As we saw before this implies that $\mu_n \Rightarrow \delta_0$.
$\square$

Applications.

1. Convergence to the Poisson Distribution. Let $\{X_{n,j} : 1 \le j \le k_n\}$ be k_n independent random variables taking the values 0 or 1 with probabilities $1 - p_{n,j}$ and $p_{n,j}$ respectively. We assume that

$$\lim_{n\to\infty} \sup_{1\le j\le k_n} p_{n,j} = 0$$

which is the uniform infinitesimality condition. We are interested in the limiting distribution of $S_n = \sum_{j=1}^{k_n} X_{n,j}$ as $n \to \infty$. Since we have to center by the mean we can pick any level say $\frac{1}{2}$ for truncation. Then the truncated means are all 0. The accompanying laws are given by $\mathbf{e}\,(M_n, 0, a_n)$ with $M_n = (\sum p_{n,j})\delta_1$ and $a_n = (\sum p_{n,j})\,\theta(1)$. It is clear that a limit exists if and only if $\lambda_n = \sum p_{n,j}$ has a limit λ as $n \to \infty$ and the limit in such a case is the Poisson distribution with parameter λ.

2. Convergence to the normal distribution. If the limit of $S_n = \sum_{j=1}^{k_n} X_{n,j}$ of k_n uniformly infinitesimal mutually independent random variables exists, then the limit is Normal if and only if $M \equiv 0$. If $a_{n,j}$ is the centering needed, this is equivalent to

$$\lim_{n\to\infty} \sum_j P[|X_{n,j} - a_{n,j}| \ge \epsilon] = 0$$

for all $\epsilon > 0$. Since $\lim_{n\to\infty} \sup_j |a_{n,j}| = 0$, this is equivalent to

$$\lim_{n\to\infty} \sum_j P[|X_{n,j}| \ge \epsilon] = 0$$

for each $\epsilon > 0$.

3. The limiting variance and the mean are given by

$$\sigma^2 = \lim_{n\to\infty} \sum_j E\Big\{[X_{n,j} - a_{n,j}]^2 : |X_{n,j} - a_{n,j}| \le 1\Big\}$$

and

$$a = \lim_{n\to\infty} \sum_j a_{n,j}$$

where

$$a_{n,j} = \int_{|x|\le 1} x \, d\alpha_{n,j}$$

Suppose that $E[X_{n,j}] = 0$ for all $1 \le j \le k_n$ and n. Assume that $\sigma_n^2 = \sum_j E\{[X_{n,j}]^2\}$ and $\sigma^2 = \lim_{n\to\infty} \sigma_n^2$ exists. What do we need in order to make sure that the limiting distribution is normal with mean 0 and variance σ^2? Let $\alpha_{n,j}$ be the distribution of $X_{n,j}$.

$$|a_{n,j}|^2 = \Big|\int_{|x|\le 1} x \, d\alpha_{n,j}\Big|^2 = \Big|\int_{|x|>1} x \, d\alpha_{n,j}\Big|^2 \le \alpha_{n,j}[\,|x|>1\,] \int |x|^2 \, d\alpha_{n,j}$$

and

$$\begin{aligned}\sum_{j=1}^{k_n} |a_{n,j}|^2 &\le \Big\{\sum_{1\le j\le k_n} \int |x|^2 \, d\alpha_{n,j}\Big\}\Big\{\sup_{1\le j\le k_n} \alpha_{n,j}[\,|x|>1\,]\Big\}\\ &\le \sigma_n^2\Big\{\sup_{1\le j\le k_n} \alpha_{n,j}[\,|x|>1\,]\Big\}\\ &\to 0.\end{aligned}$$

Because $\sum_{j=1}^{k_n} |a_{n,j}|^2 \to 0$ as $n \to \infty$ we must have

$$\sigma^2 = \lim_{n\to\infty} \sum \int_{|x|\le \ell} |x|^2 \, d\alpha_{n,j}$$

for every $\ell > 0$ or equivalently

$$\lim_{n\to\infty} \sum \int_{|x|>\ell} |x|^2 \, d\alpha_{n,j} = 0$$

for every ℓ establishing the necessity as well as sufficiency in Lindeberg's Theorem. A simple calculation shows that

$$\sum_j |a_{n,j}| \le \sum_j \int_{|x|>1} |x|\, d\alpha_{n,j} \le \sum_j \int_{|x|>1} |x|^2\, d\alpha_{n,j} = 0$$

establishing that the limiting Normal distribution has mean 0.

Exercise 3.26. What happens in the Poisson limit theorem (application 1) if $\lambda_n = \sum_j p_{n,j} \to \infty$ as $n \to \infty$? Can you show that the distribution of $\frac{S_n - \lambda_n}{\sqrt{\lambda_n}}$ converges to the standard Normal distribution?

3.9 Laws of the iterated logarithm.

When we are dealing with a sequence of independent identically distributed random variables $X_1, \cdots, X_n, \cdots$ with mean 0 and variance 1, we have a strong law of large numbers asserting that

$$P\left\{\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = 0\right\} = 1$$

and a central limit theorem asserting that

$$P\left\{\frac{X_1 + \cdots + X_n}{\sqrt{n}} \le a\right\} \to \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} \exp[-\frac{x^2}{2}]\, dx$$

It is a reasonable question to ask if the random variables $\frac{X_1 + \cdots + X_n}{\sqrt{n}}$ themselves converge to some limiting random variable Y that is distributed according to the the standard normal distribution. The answer is no and is not hard to show.

Lemma 3.24. *For any sequence n_j of numbers $\to \infty$,*

$$P\left\{\limsup_{j\to\infty} X_1 + \cdots + X_{n_j}\sqrt{n_j} = +\infty\right\} = 1$$

Proof. Let us define

$$Z = \limsup_{j\to\infty} X_1 + \cdots + X_{n_j}\sqrt{n_j}$$

which can be $+\infty$. Because the normal distribution has an infinitely long tail, i.e the probability of exceeding any given value is positive, we must have

$$P\big[Z \geq a\big] > 0$$

for any a. But Z is a random variable that does not depend on the particular values of $X_1, \cdots, X_n$ and is therefore a set in the tail σ-field. By Kolmogorov's zero-one law $P\big[Z \geq a\big]$ must be either 0 or 1. Since it cannot be 0 it must be 1. $\square$

Since we know that $\frac{X_1+\cdots+X_n}{n} \to 0$ with probability 1 as $n \to \infty$, the question arises as to the rate at which this happens. The law of the iterated logarithm provides an answer.

Theorem 3.25. *For any sequence $X_1, \cdots, X_n, \cdots$ of independent identically distributed random variables with mean 0 and Variance 1,*

$$P\bigg\{\limsup_{n\to\infty} \frac{X_1 + \cdots + X_n}{\sqrt{n \log\log n}} = \sqrt{2}\bigg\} = 1.$$

We will not prove this theorem in the most general case which assumes only the existence of two moments. We will assume instead that $E[|X|^{2+\alpha}] < \infty$ for some $\alpha > 0$. We shall first reduce the proof to an estimate on the tail behavior of the distributions of $\frac{S_n}{\sqrt{n}}$ by a careful application of the Borel-Cantelli Lemma. This estimate is obvious if $X_1, \cdots, X_n, \cdots$ are themselves normally distributed and we will show how to extend it to a large class of distributions that satisfy the additional moment condition. It is clear that we are interested in showing that for $\lambda > \sqrt{2}$,

$$P\bigg\{S_n \geq \lambda\sqrt{n\log\log n} \;\text{ infinitely often }\;\bigg\} = 0.$$

It would be sufficient because of Borel-Cantelli lemma to show that for any $\lambda > \sqrt{2}$,

$$\sum_n P\bigg\{S_n \geq \lambda\sqrt{n\log\log n}\bigg\} < \infty.$$

This however is too strong. The condition of the Borel-Cantelli lemma is not necessary in this context because of the strong dependence between the partial sums S_n. The function $\phi(n) = \sqrt{n\log\log n}$ is clearly well defined and non-decreasing for $n \geq 3$ and it is sufficient for our purposes to show that for any $\lambda > \sqrt{2}$ we can find some sequence $k_n \uparrow \infty$ of integers such that

$$\sum_n P\left\{\sup_{k_{n-1} \leq j \leq k_n} S_j \geq \lambda\,\phi(k_{n-1})\right\} < \infty. \tag{3.33}$$

This will establish that with probability 1,

$$\limsup_{n\to\infty} \frac{\sup_{k_{n-1}\leq j\leq k_n} S_j}{\phi(k_{n-1})} \leq \lambda$$

or by the monotonicity of ϕ,

$$\limsup_{n\to\infty} \frac{S_n}{\phi(n)} \leq \lambda$$

with probability 1. Since $\lambda > \sqrt{2}$ is arbitrary the upper bound in the law of the iterated logarithm will follow. Each term in the sum of 3.33 can be estimated as in Levy's inequality,

$$P\left\{\sup_{k_{n-1}\leq j\leq k_n} S_j \geq \lambda\,\phi(k_{n-1})\right\} \leq 2\,P\left\{S_{k_n} \geq (\lambda-\sigma)\,\phi(k_{n-1})\right\}$$

with $0 < \sigma < \lambda$, provided

$$\sup_{1\leq j\leq k_n - k_{n-1}} P\left\{|S_j| \geq \sigma\phi(k_{n-1})\right\} \leq \frac{1}{2}.$$

Our choice of k_n will be $k_n = [\rho^n]$ for some $\rho > 1$ and therefore

$$\lim_{n\to\infty} \frac{\phi(k_{n-1})}{\sqrt{k_n}} = \infty$$

and by Chebychev's inequality, for any fixed $\sigma > 0$,

$$\begin{aligned}
\sup_{1\leq j\leq k_n} P\left\{|S_j| \geq \sigma\phi(k_{n-1})\right\} &\leq \frac{E[S_n^2]}{[\sigma\phi(k_{n-1})]^2} \\
&= \frac{k_n}{[\sigma\phi(k_{n-1})]^2} \\
&= \frac{k_n}{\sigma^2 k_{n-1}\log\log k_{n-1}} \\
&= o(1) \quad \text{as} \quad n \to \infty.
\end{aligned} \tag{3.34}$$

By choosing σ small enough so that $\lambda - \sigma > \sqrt{2}$ it is sufficient to show that for any $\lambda' > \sqrt{2}$,

$$\sum_n P\Big\{S_{k_n} \geq \lambda' \,\phi(k_{n-1})\Big\} < \infty.$$

By picking ρ sufficiently close to 1, (so that $\lambda'\sqrt{\rho} > \sqrt{2}$), because $\frac{\phi(k_{n-1})}{\phi(k_n)} = \frac{1}{\sqrt{\rho}}$ we can reduce this to the convergence of

$$\sum_n P\Big\{S_{k_n} \geq \lambda \,\phi(k_n)\Big\} < \infty \tag{3.35}$$

for all $\lambda > \sqrt{2}$.

If we use the estimate $P[X \geq a] \leq \exp[-\frac{a^2}{2}]$ that is valid for the standard normal distribution, we can verify 3.35.

$$\sum_n \exp\big[-\frac{\lambda^2(\phi(k_n))^2}{2\,k_n}\big] < \infty$$

for any $\lambda > \sqrt{2}$.

To prove the lower bound we select again a subsequece, $k_n = [\rho^n]$ with some $\rho > 1$, and look at $Y_n = S_{k_{n+1}} - S_{k_n}$, which are now independent random variables. The tail probability of the Normal distribution has the lower bound

$$\begin{aligned}
P[X \geq a] &= \frac{1}{\sqrt{2\pi}}\int_a^\infty \exp[-\frac{x^2}{2}]dx \\
&\geq \frac{1}{\sqrt{2\pi}}\int_a^\infty \exp[-\frac{x^2}{2} - x](x+1)dx \\
&\geq \frac{1}{\sqrt{2\pi}}\exp[-\frac{(a+1)^2}{2}].
\end{aligned}$$

If we assume Normal like tail probabilities we can conclude that

$$\sum_n P\Big\{Y_n \geq \lambda\phi(k_{n+1})\Big\} \geq \sum_n \exp\big[-\frac{1}{2}[1 + \frac{\lambda\phi(k_{n+1})}{\sqrt{(\rho^{n+1}-\rho^n)}}]^2\big] = +\infty$$

provided $\frac{\lambda^2\rho}{2(\rho-1)} < 1$ and conclude by the Borel-Cantelli lemma, that $Y_n = S_{k_{n+1}} - S_{k_n}$ exceeds $\lambda\phi(k_{n+1})$ infinitely often for such λ. On the other hand from the upper bound we already have (replacing X_i by $-X_i$)

$$P\left\{\limsup_n \frac{-S_{k_n}}{\phi(k_{n+1})} \le \frac{\sqrt{2}}{\sqrt{\rho}}\right\} = 1.$$

Consequently

$$P\left\{\limsup_n \frac{S_{k_{n+1}}}{\phi(k_{n+1})} \ge \sqrt{\frac{2(\rho-1)}{\rho}} - \frac{\sqrt{2}}{\sqrt{\rho}}\right\} = 1$$

and therefore,

$$P\left\{\limsup_n \frac{S_n}{\phi(n)} \ge \sqrt{\frac{2(\rho-1)}{\rho}} - \frac{\sqrt{2}}{\sqrt{\rho}}\right\} = 1.$$

We now take ρ arbitrarily large and we are done.

We saw that the law of the iterated logarithm depends on two things.
(i). For any $a > 0$ and $p < \frac{a^2}{2}$ an upper bound for the probability

$$P[S_n \ge a\sqrt{n \log\log n}] \le C_p[\log n]^{-p}$$

with some constant C_p
(ii). For any $a > 0$ and $p > \frac{a^2}{2}$ a lower bound for the probability

$$P[S_n \ge a\sqrt{n \log\log n}] \ge C_p[\log n]^{-p}$$

with some, possibly different, constant C_p.

Both inequalities can be obtained from a uniform rate of convergence in the central limit theorem.

$$\sup_a \left| P\{\frac{S_n}{\sqrt{n}} \ge a\} - \int_a^\infty \frac{1}{\sqrt{2\pi}} \exp[-\frac{x^2}{2}]\, dx \right| \le Cn^{-\delta} \tag{3.36}$$

for some $\delta > 0$ in the central limit theorem. Such an error estimate is provided in the following theorem

Theorem 3.26. (Berry-Esseen theorem). *Assume that the i.i.d. sequence $\{X_j\}$ with mean zero and variance one satisfies an additional moment condition $E|X|^{2+\alpha} < \infty$ for some $\alpha > 0$. Then for some $\delta > 0$ the estimate (3.36) holds.*

Proof. The proof will be carried out after two lemmas.

Lemma 3.27. *Let $-\infty < a < b < \infty$ be given and $0 < h < \frac{b-a}{2}$ be a small positive number. Consider the function $f_{a,b,h}(x)$ defined as*

$$f_{a,b,h}(x) = \begin{cases} 0 & \text{for } -\infty < x \le a-h \\ \frac{x-a+h}{2h} & \text{for } a-h \le x \le a+h \\ 1 & \text{for } a+h \le x \le b-h \\ 1 - \frac{x-b+h}{2h} & \text{for } b-h \le x \le b+h \\ 0 & \text{for } b+h \le x < \infty. \end{cases}$$

For any probability distribution μ with characteristic function $\hat{\mu}(t)$

$$\int_{-\infty}^{\infty} f_{a,b,h}(x)\, d\mu(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{\mu}(y) \frac{e^{-i\,a\,y} - e^{-i\,b\,y}}{i\,y} \frac{\sin h\,y}{h\,y}\, dy.$$

Proof. This is essentially the Fourier inversion formula. Note that

$$f_{a,b,h}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\,x\,y} \frac{e^{-i\,a\,y} - e^{-i\,b\,y}}{i\,y} \frac{\sin h\,y}{h\,y}\, dy.$$

We can start with the double integral

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\,x\,y} \frac{e^{-i\,a\,y} - e^{-i\,b\,y}}{i\,y} \frac{\sin h\,y}{h\,y}\, dy\, d\mu(x)$$

and apply Fubini's theorem to obtain the lemma. ∎

Lemma 3.28. *If λ, μ are two probability measures with zero mean having $\hat{\lambda}(\cdot), \hat{\mu}(\cdot)$ for respective characteristic functions. Then*

$$\int_{-\infty}^{\infty} f_{a,h}(x)\, d(\lambda - \mu)(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} [\hat{\lambda}(y) - \hat{\mu}(y)] \frac{e^{-i\,a\,y}}{i\,y} \frac{\sin h\,y}{h\,y}\, dy$$

where $f_{a,h}(x) = f_{a,\infty,h}(x)$, is given by

$$f_{a,h}(x) = \begin{cases} 0 & \text{for } -\infty < x \le a-h \\ \frac{x-a+h}{2h} & \text{for } a-h \le x \le a+h \\ 1 & \text{for } a+h \le x < \infty. \end{cases}$$

Proof. We just let $b \to \infty$ in the previous lemma. Since $|\hat{\lambda}(y) - \hat{\mu}(y)| = o(|y|)$, there is no problem in applying the Riemann-Lebesgue Lemma. We now proceed with the proof of the theorem.

$$\lambda[[a,\infty)] \le \int f_{a-h,h}(x)\, d\lambda(x) \le \lambda[[a-2h,\infty)]$$

and

$$\mu[[a,\infty)] \le \int f_{a-h,h}(x)\, d\mu(x) \le \mu[[a-2h,\infty)].$$

Therefore if we assume that μ has a density bounded by C,

$$\lambda[[a,\infty)] - \mu[[a,\infty)] \le 2hC + \int f_{a-h,h}(x)\, d(\lambda-\mu)(x).$$

Since we get a similar bound in the other direction as well,

$$\begin{aligned}
\sup_a |\lambda[[a,\infty)] - \mu[[a,\infty)]| &\le \sup_a \Big| \int f_{a-h,h}(x)\, d(\lambda-\mu)(x) \Big| \\
&\quad + 2hC \\
&\le \frac{1}{2\pi} \int_{-\infty}^{\infty} |\hat{\lambda}(y) - \hat{\mu}(y)| \frac{|\sin h\, y\,|}{h\, y^2}\, dy \\
&\quad + 2hC. \qquad (3.37)
\end{aligned}$$

∎

Now we return to the proof of the theorem. We take λ to be the distribution of $\frac{S_n}{\sqrt{n}}$ having as its characteristic function $\hat{\lambda}_n(y) = [\phi(\frac{y}{\sqrt{n}})]^n$ where $\phi(y)$ is the characteristic function of the common distribution of the $\{X_i\}$ and has the expansion

$$\phi(y) = 1 - \frac{y^2}{2} + O(|y|^{2+\alpha})$$

for some $\alpha > 0$. We therefore get, for some choice of $\alpha > 0$,

$$|\hat{\lambda}_n(y) - \exp[-\frac{y^2}{2}]| \le C \frac{|y|^{2+\alpha}}{n^\alpha} \quad \text{if} \quad |y| \le n^{\frac{\alpha}{2+\alpha}}.$$

Therefore for $\theta = \frac{\alpha}{2+\alpha}$

$$
\begin{aligned}
\int_{-\infty}^{\infty} \big|\hat\lambda_n(y) &- \exp[-\frac{y^2}{2}]\big| \frac{|\sin h\, y\,|}{h\, y^2}\, dy \\
&= \int_{|y|\le n^\theta} \big|\hat\lambda_n(y) - \exp[-\frac{y^2}{2}]\big| \frac{|\sin h\, y\,|}{h\, y^2}\, dy \\
&\qquad + \int_{|y|\ge n^\theta} \big|\hat\lambda_n(y) - \exp[-\frac{y^2}{2}]\big| \frac{|\sin h\, y\,|}{h\, y^2}\, dy \\
&\le \frac{C}{h}\Big\{ \int_{|y|\le n^\theta} \frac{|y|^\alpha}{n^\alpha}\, dy + \int_{|y|\ge n^\theta} \frac{dy}{|y|^2} \Big\} \\
&\le C\frac{n^{(\alpha+1)\theta-\alpha} + n^{-\theta}}{h} \\
&= \frac{C}{hn^{\frac{\alpha}{\alpha+2}}}
\end{aligned}
$$

Substituting this bound in 3.37 we get

$$
\sup_a |\lambda_n[[a,\infty)] - \mu[[a,\infty)]| \le C_1\, h + \frac{C}{h\, n^{\frac{\alpha}{2+\alpha}}}\ .
$$

By picking $h = n^{-\frac{\alpha}{2(2+\alpha)}}$ we get

$$
\sup_a |\lambda_n[[a,\infty)] - \mu[[a,\infty)]| \le C\, n^{-\frac{\alpha}{2(2+\alpha)}}
$$

and we are done. $\square$