# 10

# Limit Theorems

## 10.1 Central Limit Theorem, the Lindeberg Condition

Limit Theorems describe limiting distributions of appropriately scaled sums of a large number of random variables. It is usually assumed that the random variables are either independent, or almost independent, in some sense. In the case of the Central Limit Theorem that we prove in this section, the random variables are independent and the limiting distribution is Gaussian. We first introduce the definitions.

Let $\xi_1, \xi_2, ...$ be a sequence of independent random variables with finite variances, $m_i = \mathrm{E}\xi_i$, $\sigma_i^2 = \mathrm{Var}(\xi_i)$, $\zeta_n = \sum_{i=1}^n \xi_i$, $M_n = \mathrm{E}\zeta_n = \sum_{i=1}^n m_i$, $D_n^2 = \mathrm{Var}(\zeta_n) = \sum_{i=1}^n \sigma_i^2$. Let $F_i = F_{\xi_i}$ be the distribution function of the random variable $\xi_i$.

**Definition 10.1.** *The Lindeberg condition is said to be satisfied if*

$$\lim_{n\to\infty} \frac{1}{D_n^2} \sum_{i=1}^n \int_{\{x:|x-m_i|\geq \varepsilon D_n\}} (x-m_i)^2 dF_i(x) = 0$$

*for every* $\varepsilon > 0$.

*Remark 10.2.* The Lindeberg condition easily implies that $\lim_{n\to\infty} D_n = \infty$ (see formula (10.5) below).

**Theorem 10.3. (Central Limit Theorem, Lindeberg Condition)** *Let $\xi_1, \xi_2, ...$ be a sequence of independent random variables with finite variances. If the Lindeberg condition is satisfied, then the distributions of $(\zeta_n - M_n)/D_n$ converge weakly to $N(0,1)$ distribution as $n \to \infty$.*

*Proof.* We may assume that $m_i = 0$ for all $i$. Otherwise we can consider a new sequence of random variables $\widetilde{\xi}_i = \xi_i - m_i$, which have zero expectations, and for which the Lindeberg condition is also satisfied. Let $\varphi_i(\lambda)$ and $\varphi_{\tau_n}(\lambda)$ be the characteristic functions of the random variables $\xi_i$ and $\tau_n = \frac{\zeta_n}{D_n}$ respectively. By Theorem 9.7, it is sufficient to prove that for all $\lambda \in \mathbb{R}$

$$\varphi_{\tau_n}(\lambda) \to e^{-\frac{\lambda^2}{2}} \quad \text{as} \quad n \to \infty. \tag{10.1}$$

Fix $\lambda \in \mathbb{R}$ and note that the left-hand side of (10.1) can be written as follows:

$$\varphi_{\tau_n}(\lambda) = \mathrm{E}e^{i\lambda\tau_n} = \mathrm{E}e^{i(\frac{\lambda}{D_n})(\xi_1+\ldots+\xi_n)} = \prod_{i=1}^{n} \varphi_i(\frac{\lambda}{D_n}).$$

We shall prove that

$$\varphi_i(\frac{\lambda}{D_n}) = 1 - \frac{\lambda^2\sigma_i^2}{2D_n^2} + a_i^n \tag{10.2}$$

for some $a_i^n = a_i^n(\lambda)$ such that for any $\lambda$

$$\lim_{n\to\infty} \sum_{i=1}^{n} |a_i^n| = 0. \tag{10.3}$$

Assuming (10.2) for now, let us prove the theorem. By Taylor's formula, for any complex number $z$ with $|z| < \frac{1}{4}$

$$\ln(1+z) = z + \theta(z)|z|^2, \tag{10.4}$$

with $|\theta(z)| \le 1$, where ln denotes the principal value of the logarithm (the analytic continuation of the logarithm from the positive real semi-axis to the half-plane $\mathrm{Re}(z) > 0$).

We next show that

$$\lim_{n\to\infty} \max_{1\le i\le n} \frac{\sigma_i^2}{D_n^2} = 0. \tag{10.5}$$

Indeed, for any $\varepsilon > 0$,

$$\max_{1\le i\le n} \frac{\sigma_i^2}{D_n^2} \le \max_{1\le i\le n} \frac{\int_{\{x:|x|\ge\varepsilon D_n\}} x^2 dF_i(x)}{D_n^2} + \max_{1\le i\le n} \frac{\int_{\{x:|x|\le\varepsilon D_n\}} x^2 dF_i(x)}{D_n^2}.$$

The first term on the right-hand side of this inequality tends to zero by the Lindeberg condition. The second term does not exceed $\varepsilon^2$, since the integrand does not exceed $\varepsilon^2 D_n^2$ on the domain of integration. This proves (10.5), since $\varepsilon$ was arbitrary.

Therefore, when $n$ is large enough, we can put $z = -\frac{\lambda^2\sigma_i^2}{2D_n^2} + a_i^n$ in (10.4) and obtain

$$\sum_{i=1}^{n} \ln \varphi_i(\frac{\lambda}{D_n}) = \sum_{i=1}^{n} \frac{-\lambda^2\sigma_i^2}{2D_n^2} + \sum_{i=1}^{n} a_i^n + \sum_{i=1}^{n} \theta_i |\frac{-\lambda^2\sigma_i^2}{2D_n^2} + a_i^n|^2$$

with $|\theta_i| \le 1$. The first term on the right-hand side of this expression is equal to $-\frac{\lambda^2}{2}$. The second term tends to zero due to (10.3). The third term tends to zero since

$$\sum_{i=1}^{n} \theta_i \Big| \frac{-\lambda^2 \sigma_i^2}{2D_n^2} + a_i^n \Big|^2 \le \max_{1 \le i \le n} \{ \frac{\lambda^2 \sigma_i^2}{2D_n^2} + |a_i^n| \} \sum_{i=1}^{n} ( \frac{\lambda^2 \sigma_i^2}{2D_n^2} + |a_i^n| )$$

$$\le c(\lambda) \max_{1 \le i \le n} \{ \frac{\lambda^2 \sigma_i^2}{2D_n^2} + |a_i^n| \},$$

where $c(\lambda)$ is a constant, while the second factor converges to zero by (10.3) and (10.5). We have thus demonstrated that

$$\lim_{n \to \infty} \sum_{i=1}^{n} \ln \varphi_i ( \frac{\lambda}{D_n} ) = -\frac{\lambda^2}{2},$$

which clearly implies (10.1). It remains to prove (10.2). We use the following simple relations:

$$e^{ix} = 1 + ix + \frac{\theta_1(x) x^2}{2},$$

$$e^{ix} = 1 + ix - \frac{x^2}{2} + \frac{\theta_2(x) x^3}{6},$$

which are valid for all real $x$, with $|\theta_1(x)| \le 1$ and $|\theta_2(x)| \le 1$. Then

$$\varphi_i ( \frac{\lambda}{D_n} ) = \int_{-\infty}^{\infty} e^{\frac{i\lambda}{D_n} x} dF_i(x) = \int_{|x| \ge \varepsilon D_n} (1 + \frac{i\lambda}{D_n} x + \frac{\theta_1(x) (\lambda x)^2}{2D_n^2}) dF_i(x)$$

$$+ \int_{|x| < \varepsilon D_n} (1 + \frac{i\lambda x}{D_n} - \frac{\lambda^2 x^2}{2D_n^2} + \frac{\theta_2(x) |\lambda x|^3}{6D_n^3}) dF_i(x)$$

$$= 1 - \frac{\lambda^2 \sigma_i^2}{2D_n^2} + \frac{\lambda^2}{2D_n^2} \int_{|x| \ge \varepsilon D_n} (1 + \theta_1(x)) x^2 dF_i(x)$$

$$+ \frac{|\lambda|^3}{6D_n^3} \int_{|x| < \varepsilon D_n} \theta_2(x) |x|^3 dF_i(x).$$

Here we have used that

$$\int_{-\infty}^{\infty} x dF_i(x) = \mathrm{E}\xi_i = 0.$$

In order to prove (10.2), we need to show that

$$\sum_{i=1}^{n} \frac{\lambda^2}{2D_n^2} \int_{|x| \ge \varepsilon D_n} (1 + \theta_1(x)) x^2 dF_i(x) + \sum_{i=1}^{n} \frac{|\lambda|^3}{6D_n^3} \int_{|x| < \varepsilon D_n} \theta_2(x) |x|^3 dF_i(x) \to 0. \tag{10.6}$$

The second sum in (10.6) can be estimated as

$$\Big| \sum_{i=1}^{n} \frac{|\lambda|^3}{6D_n^3} \int_{|x| < \varepsilon D_n} \theta_2(x) |x|^3 dF_i(x) \Big|$$

$$\leq \Big|\sum_{i=1}^{n}\frac{|\lambda|^3\varepsilon}{6D_n^3}\int_{|x|<\varepsilon D_n}\theta_2(x)x^2 D_n dF_i(x)\Big| \leq \sum_{i=1}^{n}\frac{|\lambda|^3\varepsilon\sigma_i^2}{6D_n^2} = \frac{\varepsilon|\lambda|^3}{6},$$

which can be made arbitrarily small by selecting a sufficiently small $\varepsilon$. The first sum in (10.6) tends to zero by the Lindeberg condition. □

*Remark 10.4.* The proof can be easily modified to demonstrate that the convergence in (10.1) is uniform on any compact set of values of $\lambda$. We shall need this fact in the next section.

The Lindeberg condition is clearly satisfied for every sequence of independent identically distributed random variables with finite variances. We therefore have the following Central Limit Theorem for independent identically distributed random variables.

**Theorem 10.5.** *Let $\xi_1, \xi_2, \ldots$ be a sequence of independent identically distributed random variables with $m = \mathrm{E}\xi_1$ and $0 < \sigma^2 = \mathrm{Var}(\xi_1) < \infty$. Then the distributions of $(\zeta_n - nm)/\sqrt{n}\sigma$ converge weakly to $N(0,1)$ distribution as $n \to \infty$.*

Theorem 10.3 also implies the Central Limit Theorem under the following Lyapunov condition.

**Definition 10.6.** *The Lyapunov condition is said to be satisfied if there is a $\delta > 0$ such that*

$$\lim_{n\to\infty}\frac{1}{D_n^{2+\delta}}\sum_{i=1}^{n}\mathrm{E}(|\xi_i - m_i|^{2+\delta}) = 0.$$

**Theorem 10.7. (Central Limit Theorem, Lyapunov Condition)** *Let $\xi_1, \xi_2, \ldots$ be a sequence of independent random variables with finite variances. If the Lyapunov condition is satisfied, then the distributions of $(\zeta_n - M_n)/D_n$ converge weakly to $N(0,1)$ distribution as $n \to \infty$.*

*Proof.* Let $\varepsilon, \delta > 0$. Then,

$$\frac{\int_{\{x:|x-m_i|\geq\varepsilon D_n\}}(x-m_i)^2 dF_i(x)}{D_n^2}$$

$$\leq \frac{\int_{\{x:|x-m_i|\geq\varepsilon D_n\}}(x-m_i)^{2+\delta} dF_i(x)}{D_n^2(\varepsilon D_n)^\delta} \leq \varepsilon^{-\delta}\frac{\mathrm{E}(|\xi_i - m_i|^{2+\delta})}{D_n^{2+\delta}}.$$

Therefore, a sequence of random variables satisfying the Lyapunov condition also satisfies the Lindeberg condition. □

If condition (10.5) is satisfied, then the Lindeberg condition is not only sufficient, but also necessary for the Central Limit Theorem to hold. We state the following theorem without providing a proof.

**Theorem 10.8. (Lindeberg-Feller)** *Let $\xi_1, \xi_2, ...$ be a sequence of independent random variables with finite variances such that the condition (10.5) is satisfied. Then the Lindeberg condition is satisfied if and only if the Central Limit Theorem holds, that is the distributions of $(\zeta_n - M_n)/D_n$ converge weakly to $N(0,1)$ distribution as $n \to \infty$.*

There are various generalizations of the Central Limit Theorem, not presented here, where the condition of independence of random variables is replaced by conditions of weak dependence in some sense. Other important generalizations concern vector-valued random variables.

## 10.2 Local Limit Theorem

The Central Limit Theorem proved in the previous section states that the measures on $\mathbb{R}$ induced by normalized sums of independent random variables converge weakly to the Gaussian measure $N(0,1)$. Under certain additional conditions this statement can be strengthened to include the point-wise convergence of the densities. In the case of integer-valued random variables (where no densities exist) the corresponding statement is the following Local Central Limit Theorem, which is a generalization of the de Moivre-Laplace Theorem.

Let $\xi$ be an integer-valued random variable. Let $X = \{x_1, x_2, ...\}$ be the finite or countable set consisting of those values of $\xi$ for which $p_j = \mathrm{P}(\xi = x_j) \neq 0$. We shall say that $\xi$ spans the set of integers $\mathbb{Z}$ if the greatest common divisor of all the elements of $X$ equals 1.

**Lemma 10.9.** *If $\xi$ spans $\mathbb{Z}$, and $\varphi(\lambda) = \mathrm{E}e^{i\xi\lambda}$ is the characteristic function of the variable $\xi$, then for any $\delta > 0$*

$$\sup_{\delta \leq |\lambda| \leq \pi} |\varphi(\lambda)| < 1 \ . \tag{10.7}$$

*Proof.* Suppose that $x\lambda_0 \in \{2k\pi, k \in \mathbb{Z}\}$ for some $\lambda_0$ and all $x \in X$. Then $\lambda_0 \in \{2k\pi, k \in \mathbb{Z}\}$, since 1 is the largest common divisor of all the elements of $X$. Therefore, if $\delta \leq |\lambda| \leq \pi$, then $x\lambda \notin \{2k\pi, k \in \mathbb{Z}\}$ for some $x \in X$. This in turn implies that $e^{i\lambda x} \neq 1$. Recall that

$$\varphi(\lambda) = \sum_{x_j \in X} p_j e^{i\lambda x_j} .$$

Since $\sum_{x_j \in X} p_j = 1$ and $p_j > 0$, the relation $e^{i\lambda x} \neq 1$ for some $x \in X$ implies that $|\varphi(\lambda)| < 1$. Since $|\varphi(\lambda)|$ is continuous,

$$\sup_{\delta \leq |\lambda| \leq \pi} |\varphi(\lambda)| < 1 \ .$$

□

Let $\xi_1, \xi_2, ...$ be a sequence of integer-valued independent identically distributed random variables. Let $m = \mathrm{E}\xi_1$, $\sigma^2 = \mathrm{Var}(\xi_1) < \infty$, $\zeta_n = \sum_{i=1}^n \xi_i$, $M_n = \mathrm{E}\zeta_n = nm$, $D_n^2 = \mathrm{Var}(\zeta_n) = n\sigma^2$. We shall be interested in the probability of the event that $\zeta_n$ takes an integer value $k$. Let $\mathrm{P}_n(k) = \mathrm{P}(\zeta_n = k)$, $z = z(n,k) = \frac{k - M_n}{D_n}$.

**Theorem 10.10. (Local Limit Theorem)** *Let $\xi_1, \xi_2, ...$ be a sequence of independent identically distributed integer-valued random variables with finite variances such that $\xi_1$ spans $\mathbb{Z}$. Then*

$$\lim_{n\to\infty} (D_n \mathrm{P}_n(k) - \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}) = 0 \tag{10.8}$$

*uniformly in $k$.*

*Proof.* We shall prove the theorem for the case $m = 0$, since the general case requires only trivial modifications. Let $\varphi(\lambda)$ be the characteristic function of each of the variables $\xi_i$. Then the characteristic function of the random variable $\zeta_n$ is

$$\varphi_{\zeta_n}(\lambda) = \varphi^n(\lambda) = \sum_{k=-\infty}^{\infty} \mathrm{P}_n(k) e^{i\lambda k}.$$

Thus $\varphi^n(\lambda)$ is the Fourier series with coefficients $\mathrm{P}_n(k)$, and we can use the formula for Fourier coefficients to find $\mathrm{P}_n(k)$:

$$2\pi \mathrm{P}_n(k) = \int_{-\pi}^{\pi} \varphi^n(\lambda) e^{-i\lambda k} d\lambda = \int_{-\pi}^{\pi} \varphi^n(\lambda) e^{-i\lambda z D_n} d\lambda.$$

Therefore, after a change of variables we obtain

$$2\pi D_n \mathrm{P}_n(k) = \int_{-\pi D_n}^{\pi D_n} e^{-i\lambda z} \varphi^n(\frac{\lambda}{D_n}) d\lambda.$$

From the formula for the characteristic function of the Gaussian distribution

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\lambda z - \frac{\lambda^2}{2}} d\lambda = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\lambda z - \frac{\lambda^2}{2}} d\lambda.$$

We can write the difference in (10.8) multiplied by $2\pi$ as a sum of four integrals:

$$2\pi (D_n \mathrm{P}_n(k) - \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}) = I_1 + I_2 + I_3 + I_4,$$

where

$$I_1 = \int_{-T}^{T} e^{-i\lambda z} (\varphi^n(\frac{\lambda}{D_n}) - e^{-\frac{\lambda^2}{2}}) d\lambda,$$

$$I_2 = -\int_{|\lambda| > T} e^{-i\lambda z - \frac{\lambda^2}{2}} d\lambda,$$

$$I_3 = \int_{\delta D_n \leq |\lambda| \leq \pi D_n} e^{-i\lambda z} \varphi^n(\frac{\lambda}{D_n}) d\lambda,$$

$$I_4 = \int_{T \leq |\lambda| < \delta D_n} e^{-i\lambda z} \varphi^n(\frac{\lambda}{D_n}) d\lambda,$$

where the positive constants $T < \delta D_n$ and $\delta < \pi$ will be selected later. By Remark 10.4, the convergence $\lim_{n\to\infty} \varphi^n(\frac{\lambda}{D_n}) = e^{-\frac{\lambda^2}{2}}$ is uniform on the interval $[-T, T]$. Therefore $\lim_{n\to\infty} I_1 = 0$ for any $T$.

The second integral can be estimated as follows:

$$|I_2| \leq \int_{|\lambda|>T} |e^{-i\lambda z - \frac{\lambda^2}{2}}| d\lambda = \int_{|\lambda|>T} e^{-\frac{\lambda^2}{2}} d\lambda,$$

which can be made arbitrarily small by selecting $T$ large enough, since the improper integral $\int_{-\infty}^{\infty} e^{-\frac{\lambda^2}{2}} d\lambda$ converges.

The third integral is estimated as follows:

$$|I_3| \leq \int_{\delta D_n \leq |\lambda| \leq \pi D_n} |e^{-i\lambda z} \varphi^n(\frac{\lambda}{D_n})| d\lambda \leq 2\pi\sigma\sqrt{n}(\sup_{\delta \leq |\lambda| \leq \pi} |\varphi(\lambda)|)^n,$$

which tends to zero as $n \to \infty$ due to (10.7).

In order to estimate the fourth integral, we note that the existence of the variance implies that the characteristic function is a twice continuously differentiable complex-valued function with $\varphi'(0) = im = 0$ and $\varphi''(0) = -\sigma^2$. Therefore, applying the Taylor formula to the real and imaginary parts of $\varphi$, we obtain

$$\varphi(\lambda) = 1 - \frac{\sigma^2 \lambda^2}{2} + o(\lambda^2) \quad \text{as} \quad \lambda \to 0.$$

For $|\lambda| \leq \delta$ and $\delta$ sufficiently small, we obtain

$$|\varphi(\lambda)| \leq 1 - \frac{\sigma^2 \lambda^2}{4} \leq e^{-\frac{\sigma^2 \lambda^2}{4}}.$$

If $|\lambda| \leq \delta D_n$, then

$$|\varphi(\frac{\lambda}{D_n})|^n \leq e^{\frac{-n\sigma^2\lambda^2}{4D_n^2}} = e^{-\frac{\lambda^2}{4}}.$$

Therefore,

$$|I_4| \leq 2\int_T^{\delta D_n} e^{-\frac{\lambda^2}{4}} d\lambda \leq 2\int_T^{\infty} e^{-\frac{\lambda^2}{4}} d\lambda.$$

This can be made arbitrarily small by selecting sufficiently large $T$. This completes the proof of the theorem. □

When we studied the recurrence and transience of random walks on $\mathbb{Z}^d$ (Section 6) we needed to estimate the probability that a path returns to the origin after $2n$ steps:

$$u_{2n} = \mathrm{P}(\sum_{j=1}^{2n} \omega_j = 0).$$

Here $\omega_j$ are independent identically distributed random variables with values in $\mathbb{Z}^d$ with the distribution $p_y$, $y \in \mathbb{Z}^d$, where $p_y = \frac{1}{2d}$ if $y = \pm e_s$, $1 \le s \le d$, and 0 otherwise.

Let us use the characteristic functions to study the asymptotics of $u_{2n}$ as $n \to \infty$. The characteristic function of $\omega_j$ is equal to

$$\mathrm{E}e^{i(\lambda,\omega_i)} = \frac{1}{2d}(e^{i\lambda_1} + e^{-i\lambda_1} + ... + e^{i\lambda_d} + e^{-i\lambda_d}) = \frac{1}{d}(\cos(\lambda_1) + ... + \cos(\lambda_d)),$$

where $\lambda = (\lambda_1, ..., \lambda_d) \in \mathbb{R}^d$. Therefore, the characteristic function of the sum $\sum_{j=1}^{2n} \omega_j$ is equal to $\varphi_{2n}(\lambda) = \frac{1}{d^{2n}}(\cos(\lambda_1) + ... + \cos(\lambda_d))^{2n}$. On the other hand,

$$\varphi_{2n}(\lambda) = \sum_{k \in \mathbb{Z}^d} \mathrm{P}_n(k) e^{i(\lambda,k)},$$

where $\mathrm{P}_n(k) = \mathrm{P}(\sum_{j=1}^{2n} \omega_j = k)$. Integrating both sides of the equality

$$\sum_{k \in \mathbb{Z}^d} \mathrm{P}_n(k) e^{i(\lambda,k)} = \frac{1}{d^{2n}}(\cos(\lambda_1) + ... + \cos(\lambda_d))^{2n}$$

over $\lambda$, we obtain

$$(2\pi)^d u_{2n} = \frac{1}{d^{2n}} \int_{-\pi}^{\pi} ... \int_{-\pi}^{\pi} (\cos(\lambda_1) + ... + \cos(\lambda_d))^{2n} d\lambda_1 ... d\lambda_d.$$

The asymptotics of the latter integral can be treated with the help of the so-called Laplace asymptotic method. The Laplace method is used to describe the asymptotic behavior of integrals of the form

$$\int_D f(\lambda) e^{sg(\lambda)} d\lambda,$$

where $D$ is a domain in $\mathbb{R}^d$, $f$ and $g$ are smooth functions, and $s \to \infty$ is a large parameter. The idea is that if $f(\lambda) > 0$ for $\lambda \in D$, then the main contribution to the integral comes from arbitrarily small neighborhoods of the maxima of the function $g$. Then the Taylor formula can be used to approximate the function $g$ in small neighborhoods of its maxima. In our case the points of the maxima are $\lambda_1 = ... = \lambda_d = 0$ and $\lambda_1 = ... = \lambda_d = \pm\pi$. We state the result for the problem at hand without going into further detail:

$$\int_{-\pi}^{\pi} ... \int_{-\pi}^{\pi} (\cos(\lambda_1) + ... + \cos(\lambda_d))^{2n} d\lambda_1 ... d\lambda_d$$
$$= \int_{-\pi}^{\pi} ... \int_{-\pi}^{\pi} e^{2n \ln|\cos(\lambda_1)+...+\cos(\lambda_d)|} d\lambda_1 ... d\lambda_d$$
$$\sim c \sup(|\cos(\lambda_1) + ... + \cos(\lambda_d)|)^{2n} n^{-\frac{d}{2}} = c d^{2n} n^{-\frac{d}{2}},$$

which implies that $u_{2n} \sim c n^{-\frac{d}{2}}$ as $n \to \infty$ with another constant $c$.

## 10.3 Central Limit Theorem and Renormalization Group Theory

The Central Limit Theorem states that Gaussian distributions can be obtained as limits of distributions of properly normalized sums of independent random variables. If the random variables $\xi_1, \xi_2, ...$ forming the sum are independent and identically distributed, then it is enough to assume that they have a finite second moment.

In this section we shall take another look at the mechanism of convergence of normalized sums, which may help explain why the class of distributions of $\xi_i$, for which the central limit theorem holds, is so large. We shall view the densities (assuming that they exist) of the normalized sums as iterations of a certain non-linear transformation applied to the common density of $\xi_i$. The method presented below is called the renormalization group method. It can be generalized in several ways (for example, to allow the variables to be weakly dependent). We do not strive for maximal generality, however. Instead, we consider again the case of independent random variables.

Let $\xi_1, \xi_2, ...$ be a sequence of independent identically distributed random variables with zero expectation and finite variance. We define the random variables

$$\zeta_n = 2^{-\frac{n}{2}} \sum_{i=1}^{2^n} \xi_i, \quad n \geq 0.$$

Then

$$\zeta_{n+1} = \frac{1}{\sqrt{2}}(\zeta_n' + \zeta_n''),$$

where

$$\zeta_n' = 2^{-\frac{n}{2}} \sum_{i=1}^{2^n} \xi_i, \quad \zeta_n'' = 2^{-\frac{n}{2}} \sum_{i=2^n+1}^{2^{n+1}} \xi_i.$$

Clearly, $\zeta_n'$ and $\zeta_n''$ are independent identically distributed random variables. Let us assume that $\xi_i$ have a density, which will be denoted by $p_0$. Note that $\zeta_0 = \xi_1$, and thus the density of $\zeta_0$ is also $p_0$. Let us denote the density of $\zeta_n$ by $p_n$ and its distribution by $P_n$. Then

$$p_{n+1}(x) = \sqrt{2} \int_{-\infty}^{\infty} p_n(\sqrt{2}x - u) p_n(u) du.$$

Thus the sequence $p_n$ can be obtained from $p_0$ by iterating the non-linear operator $T$, which acts on the space of densities according to the formula

$$Tp(x) = \sqrt{2} \int_{-\infty}^{\infty} p(\sqrt{2}x - u) p(u) du, \tag{10.9}$$

that is $p_{n+1} = Tp_n$ and $p_n = T^n p_0$. Note that if $p$ is the density of a random variable with zero expectation, then so is $Tp$. In other words,

$$\int_{-\infty}^{\infty} x(Tp)(x)dx = 0 \quad \text{if} \quad \int_{-\infty}^{\infty} xp(x)dx = 0. \tag{10.10}$$

Indeed, if $\zeta'$ and $\zeta''$ are independent identically distributed random variables with zero mean and density $p$, then $\frac{1}{\sqrt{2}}(\zeta'+\zeta'')$ has zero mean and density $Tp$. Similarly, for a density $p$ such that $\int_{-\infty}^{\infty} xp(x)dx = 0$, the operator $T$ preserves the variance, that is

$$\int_{-\infty}^{\infty} x^2(Tp)(x)dx = \int_{-\infty}^{\infty} x^2 p(x)dx. \tag{10.11}$$

Let $p_G(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ be the density of the Gaussian distribution and $\mu_G$ the Gaussian measure on the real line (the measure with the density $p_G$). It is easy to check that $p_G$ is a fixed point of $T$, that is $p_G = Tp_G$. The fact that the convergence $P_n \Rightarrow \mu_G$ holds for a wide class of initial densities is related to the stability of this fixed point.

In the general theory of non-linear operators the investigation of the stability of a fixed point starts with an investigation of its stability with respect to the linear approximation. In our case it is convenient to linearize not the operator $T$ itself, but a related operator, as explained below.

Let $H = L^2(\mathbb{R}, \mathcal{B}, \mu_G)$ be the Hilbert space with the inner product

$$(f,g) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x)\overline{g}(x) \exp(-\frac{x^2}{2})dx.$$

Let $h$ be an element of $H$, that is a measurable function such that

$$||h||^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} h^2(x) \exp(-\frac{x^2}{2})dx < \infty.$$

Assume that $||h||$ is small. We perturb the Gaussian density as follows:

$$p_h(x) = p_G(x) + \frac{h(x)}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}) = \frac{1}{\sqrt{2\pi}}(1 + h(x)) \exp(-\frac{x^2}{2}).$$

In order for $p_h$ to be a density of a probability measure, we need to assume that

$$\int_{-\infty}^{\infty} h(x) \exp(-\frac{x^2}{2})dx = 0. \tag{10.12}$$

Moreover, in order for $p_h$ to correspond to a random variable with zero expectation, we assume that

$$\int_{-\infty}^{\infty} xh(x) \exp(-\frac{x^2}{2})dx = 0. \tag{10.13}$$

Let us define a non-linear operator $\widetilde{L}$ by the implicit relation

$$Tp_h(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})(1 + (\widetilde{L}h)(x)). \tag{10.14}$$

Thus,

$$T^n p_h(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})(1 + (\widetilde{L}^n h)(x)).$$

This formula shows that in order to study the behavior of $T^n p_h(x)$ for large $n$, it is sufficient to study the behavior of $\widetilde{L}^n h$ for large $n$. We can write

$$Tp_h(x) =$$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (1 + h(\sqrt{2}x - u)) \exp(-\frac{(\sqrt{2}x - u)^2}{2})(1 + h(u)) \exp(-\frac{u^2}{2}) du$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-\frac{(\sqrt{2}x - u)^2}{2} - \frac{u^2}{2}) du$$

$$+ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-\frac{(\sqrt{2}x - u)^2}{2} - \frac{u^2}{2})(h(\sqrt{2}x - u) + h(u)) du + O(||h||^2)$$

$$= \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}) + \frac{\sqrt{2}}{\pi} \int_{-\infty}^{\infty} \exp(-x^2 + \sqrt{2}xu - u^2) h(u) du + O(||h||^2)$$

$$= \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})(1 + (Lh)(x)) + O(||h||^2),$$

where the linear operator $L$ is given by the formula

$$(Lh)(x) = \frac{2}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp(-\frac{x^2}{2} + \sqrt{2}xu - u^2) h(u) du. \tag{10.15}$$

It is referred to as the Gaussian integral operator. Comparing two expressions for $Tp_h(x)$, the one above and the one given by (10.14), we see that

$$\widetilde{L}h = Lh + O(||h||^2),$$

that is $L$ is the linearization of $\widetilde{L}$ at zero.

It is not difficult to show that (10.15) defines a bounded self-adjoint operator on $H$. It has a complete set of eigenvectors, which are the Hermite polynomials

$$h_k(x) = \exp(\frac{x^2}{2})(\frac{d}{dx})^k \exp(-\frac{x^2}{2}), \quad k \geq 0.$$

The corresponding eigenvalues are $\lambda_k = 2^{1-\frac{k}{2}}$, $k \geq 0$. We see that $\lambda_0, \lambda_1 > 1$, $\lambda_2 = 1$, while $0 < \lambda_k \leq 1/\sqrt{2}$ for $k \geq 3$. Let $H_k$, $k \geq 0$, be one-dimensional subspaces of $H$ spanned by $h_k$. By (10.12) and (10.13) the initial vector $h$ is orthogonal to $H_0$ and $H_1$, and thus $h \in H \ominus (H_0 \oplus H_1)$.

If $h \perp H_0$, then $\widetilde{L}(h) \perp H_0$ follows from (10.14), since (10.12) holds and $p_h$ is a density. Similarly, if $h \perp H_0 \oplus H_1$, then $\widetilde{L}(h) \perp H_0 \oplus H_1$ follows from (10.10) and (10.14). Thus the subspace $H \ominus (H_0 \oplus H_1)$ is invariant not only for $L$, but also for $\widetilde{L}$. Therefore we can restrict both operators to this subspace, which can be further decomposed as follows:

$$H \ominus (H_0 \oplus H_1) = H_2 \oplus [H \ominus (H_0 \oplus H_1 \oplus H_2)].$$

Note that for an initial vector $h \in H \ominus (H_0 \oplus H_1)$, by (10.11) the operator $\widetilde{L}$ preserves its projection to $H_2$, that is

$$\int_{-\infty}^{\infty} (x^2 - 1)h(x) \exp(-\frac{x^2}{2}) = \int_{-\infty}^{\infty} (x^2 - 1)(\widetilde{L}h)(x) \exp(-\frac{x^2}{2}).$$

Let $U$ be a small neighborhood of zero in $H$, and $H^h$ the set of vectors whose projection to $H_2$ is equal to the projection of $h$ onto $H_2$. Let $U^h = U \cap H^h$. It is not difficult to show that one can choose $U$ such that $\widetilde{L}$ leaves $U^h$ invariant for all sufficiently small $h$. Note that $\widetilde{L}$ is contracting on $U^h$ for small $h$, since $L$ is contracting on $H \ominus (H_0 \oplus H_1 \oplus H_2)$. Therefore it has a unique fixed point. It is easy to verify that this fixed point is the function

$$f_h(x) = \frac{1}{\sigma(p_h)} \exp(\frac{x^2}{2} - \frac{x^2}{2\sigma^2(p_h)}) - 1,$$

where $\sigma^2(p_h)$ is the variance of a random variable with density $p_h$,

$$\sigma^2(p_h) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2(1 + h(x)) \exp(-\frac{x^2}{2})dx.$$

Therefore, by the contracting mapping principle,

$$\widetilde{L}^n h \to f_h \quad \text{as } n \to \infty,$$

and consequently

$$T^n p_h(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})(1 + (\widetilde{L}^n h)(x)) \to$$

$$\frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})(1 + f_h(x)) = \frac{1}{\sqrt{2\pi}\sigma(p_h)} \exp(-\frac{x^2}{2\sigma^2(p_h)}).$$

We see that $T^n p_h(x)$ converges in the space $H$ to the density of the Gaussian distribution with variance $\sigma^2(p_h)$. This easily implies the convergence of distributions.

It is worth stressing again that the arguments presented in this section were based on the assumption that $h$ is small, thus allowing us to state the convergence of the normalized sums $\zeta_n$ to the Gaussian distribution, provided the distribution of $\xi_i$ is a small perturbation of the Gaussian distribution. The proof of the Central Limit Theorem in Section 10.1 went through regardless of this assumption.

## 10.4 Probabilities of Large Deviations

In the previous chapters we considered the probabilities

$$\mathrm{P}(|\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n} m_i| \geq t)$$

with $m_i = \mathrm{E}\xi_i$ for sequences of independent random variables $\xi_1, \xi_2, ...,$ and we estimated these probabilities using the Chebyshev Inequality

$$\mathrm{P}(|\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n} m_i| \geq t) \leq \frac{\sum_{i=1}^{n} d_i}{t^2}, \quad d_i = \mathrm{Var}(\xi_i).$$

In particular, if the random variables $\xi_i$ are identically distributed, then for some constant $c$ which does not depend on $n$, and with $d = d_1$:

a) for $t = c\sqrt{n}$ we have $\frac{d}{c^2}$ on the right-hand side of the inequality;

b) for $t = cn$ we have $\frac{d}{c^2 n}$ on the right-hand side of the inequality.

We know from the Central Limit Theorem that in the case a) the corresponding probability converges to a positive limit as $n \to \infty$. This limit can be calculated using the Gaussian distribution. This means that in the case a) the order of magnitude of the estimate obtained from the Chebyshev Inequality is correct. On the other hand, in the case b) the estimate given by the Chebyshev Inequality is very crude. In this section we obtain more precise estimates in the case b).

Let us consider a sequence of independent identically distributed random variables. We denote their common distribution function by $F$. We make the following assumption about $F$

$$R(\lambda) = \int_{-\infty}^{\infty} e^{\lambda x} dF(x) < \infty \tag{10.16}$$

for all $\lambda$, $-\infty < \lambda < \infty$. This condition is automatically satisfied if all the $\xi_i$ are bounded. It is also satisfied if the probabilities of large values of $\xi_i$ decay faster than exponentially.

We now note several properties of the function $R(\lambda)$. From the finiteness of the integral in (10.16) for all $\lambda$, it follows that the derivatives

$$R'(\lambda) = \int_{-\infty}^{\infty} x e^{\lambda x} dF(x), \quad R''(\lambda) = \int_{-\infty}^{\infty} x^2 e^{\lambda x} dF(x)$$

exist for all $\lambda$. Let us consider $m(\lambda) = \frac{R'(\lambda)}{R(\lambda)}$. Then

$$m'(\lambda) = \frac{R''(\lambda)}{R(\lambda)} - (\frac{R'(\lambda)}{R(\lambda)})^2 = \int_{-\infty}^{\infty} \frac{x^2}{R(\lambda)} e^{\lambda x} dF(x) - (\int_{-\infty}^{\infty} \frac{x}{R(\lambda)} e^{\lambda x} dF(x))^2.$$

We define a new distribution function $F_\lambda(x) = \frac{1}{R(\lambda)} \int_{(-\infty, x]} e^{\lambda t} dF(t)$ for each $\lambda$. Then $m(\lambda) = \int_{-\infty}^{\infty} x dF_\lambda(x)$ is the expectation of a random variable with this distribution, and $m'(\lambda)$ is the variance. Therefore $m'(\lambda) > 0$ if $F$ is a non-trivial distribution, that is it is not concentrated at a point. We exclude the latter case from further consideration. Since $m'(\lambda) > 0$, $m(\lambda)$ is a monotonically increasing function.

We say that $M^+$ is an upper limit in probability for a random variable $\xi$ if $\mathrm{P}(\xi > M^+) = 0$, and $\mathrm{P}(M^+ - \varepsilon \leq \xi \leq M^+) > 0$ for every $\varepsilon > 0$. One can define the lower limit in probability in the same way. If $\mathrm{P}(\xi > M) > 0$ ($\mathrm{P}(\xi < M) > 0$) for any $M$, then $M^+ = \infty$ ($M^- = -\infty$). In all the remaining cases $M^+$ and $M^-$ are finite. The notion of the upper (lower) limit in probability can be recast in terms of the distribution function as follows:

$$M^+ = \sup\{x : F(x) < 1\}, \quad M^- = \inf\{x : F(x) > 0\}.$$

**Lemma 10.11.** *Under the assumption (10.16) on the distribution function, the limits for $m(\lambda)$ are as follows:*

$$\lim_{\lambda\to\infty} m(\lambda) = M^+, \qquad \lim_{\lambda\to-\infty} m(\lambda) = M^-.$$

*Proof.* We shall only prove the first statement since the second one is proved analogously. If $M^+ < \infty$, then from the definition of $F_\lambda$

$$\int_{(M^+,\infty)} dF_\lambda(x) = \frac{1}{R(\lambda)} \int_{(M^+,\infty)} e^{\lambda x} dF(x) = 0$$

for each $\lambda$. Note that $\int_{(M^+,\infty)} dF_\lambda(x) = 0$ implies that

$$m(\lambda) = \int_{(-\infty,M^+]} x dF_\lambda(x) \leq M^+,$$

and therefore $\lim_{\lambda\to\infty} m(\lambda) \leq M^+$. It remains to prove the opposite inequality.

Let $M^+ \leq \infty$. If $M^+ = 0$, then $m(\lambda) \leq 0$ for all $\lambda$. Therefore, we can assume that $M^+ \neq 0$. Take $M \in (0, M^+)$ if $M^+ > 0$ and $M \in (-\infty, M^+)$ if $M^+ < 0$. Choose a finite segment $[A, B]$ such that $M < A < B \leq M^+$ and $\int_{[A,B]} dF(x) > 0$. Then

$$\int_{(-\infty,M]} e^{\lambda x} dF(x) \leq e^{\lambda M},$$

while

$$\int_{(M,\infty)} e^{\lambda x} dF(x) \geq e^{\lambda A} \int_{[A,B]} dF(x),$$

which implies that

$$\int_{(-\infty,M]} e^{\lambda x} dF(x) = o(\int_{(M,\infty)} e^{\lambda x} dF(x)) \quad \text{as} \quad \lambda \to \infty.$$

Similarly,

$$\int_{(-\infty,M]} xe^{\lambda x} dF(x) = O(e^{\lambda M}),$$

while

$$|\int_{(M,\infty)} xe^{\lambda x} dF(x)| = |\int_{(M,M^+]} xe^{\lambda x} dF(x)| \geq \min(|A|,|B|)e^{\lambda A} \int_{[A,B]} dF(x),$$

which implies that

$$\int_{(-\infty,M]} xe^{\lambda x} dF(x) = o(\int_{(M,\infty)} xe^{\lambda x} dF(x)) \quad \text{as} \quad \lambda \to \infty.$$

Therefore,

$$\lim_{\lambda\to\infty} m(\lambda) = \lim_{\lambda\to\infty} \frac{\int_{(-\infty,\infty)} xe^{\lambda x} dF(x)}{\int_{(-\infty,\infty)} e^{\lambda x} dF(x)} = \lim_{\lambda\to\infty} \frac{\int_{(M,\infty)} xe^{\lambda x} dF(x)}{\int_{(M,\infty)} e^{\lambda x} dF(x)} \geq M.$$

Since $M$ can be taken to be arbitrary close to $M^+$, we conclude that $\lim_{\lambda\to\infty} m(\lambda) = M^+$. □

We now return to considering the probabilities of the deviations of sums of independent identically distributed random variables from the sums of their expectations. Consider $c$ such that $m = \mathrm{E}\xi_i < c < M^+$. We shall be interested in the probability $\mathrm{P}_{n,c} = \mathrm{P}(\xi_1 + ... + \xi_n > cn)$. Since $c > m$, this is the probability of the event that the sum of the random variables takes values which are far away from the mathematical expectation of the sum. Such values are called large deviations (from the expectation). We shall describe a method for calculating the asymptotics of these probabilities which is usually called Kramer's method.

Let $\lambda_0$ be such that $m(\lambda_0) = c$. Such $\lambda_0$ exists by Lemma 10.11 and is unique since $m(\lambda)$ is strictly monotonic. Note that $m = m(0) < c$. Therefore $\lambda_0 > 0$ by the monotonicity of $m(\lambda)$.

**Theorem 10.12.** $\mathrm{P}_{n,c} \leq B_n(R(\lambda_0)e^{-\lambda_0 c})^n$, *where* $\lim_{n\to\infty} B_n = \frac{1}{2}$.

*Proof.* We have

$$\mathrm{P}_{n,c} = \int ... \int_{x_1+...+x_n>cn} dF(x_1)...dF(x_n)$$

$$\leq (R(\lambda_0))^n e^{-\lambda_0 cn} \int ... \int_{x_1+...+x_n>cn} \frac{e^{\lambda_0(x_1+...+x_n)}}{(R(\lambda_0))^n} dF(x_1)...dF(x_n)$$

$$= (R(\lambda_0)e^{-\lambda_0 c})^n \int ... \int_{x_1+...+x_n>cn} dF_{\lambda_0}(x_1)...dF_{\lambda_0}(x_n).$$

To estimate the latter integral, we can consider independent identically distributed random variables $\widetilde{\xi}_1, ..., \widetilde{\xi}_n$ with distribution $F_{\lambda_0}$. The expectation of such random variables is equal to $\int_{\mathbb{R}} x dF_{\lambda_0}(x) = m(\lambda_0) = c$. Therefore

$$\int ... \int_{x_1+...+x_n>cn} dF_{\lambda_0}(x_1)...dF_{\lambda_0}(x_n) = \mathrm{P}(\widetilde{\xi}_1 + ... + \widetilde{\xi}_n > cn)$$
$$= \mathrm{P}(\widetilde{\xi}_1 + ... + \widetilde{\xi}_n - nm(\lambda_0) > 0)$$
$$= \mathrm{P}(\frac{\widetilde{\xi}_1 + ... + \widetilde{\xi}_n - nm(\lambda_0)}{\sqrt{nd(\lambda_0)}} > 0) \to \frac{1}{2}$$

as $n \to \infty$. Here $d(\lambda_0)$ is the variance of the random variables $\widetilde{\xi}_i$, and the convergence of the probability to $\frac{1}{2}$ follows from the Central Limit Theorem. □

The lower estimate turns out to be somewhat less elegant.

**Theorem 10.13.** *For any $b > 0$ there exists $p(b, \lambda_0) > 0$ such that*

$$\mathrm{P}_{n,c} \geq (R(\lambda_0)e^{-\lambda_0 c})^n e^{-\lambda_0 b\sqrt{n}} p_n,$$

*with* $\lim_{n\to\infty} p_n = p(b, \lambda_0) > 0$.

*Proof.* As in Theorem 10.12,

$$\mathrm{P}_{n,c} \geq \int ... \int_{cn<x_1+...+x_n<cn+b\sqrt{n}} dF(x_1)...dF(x_n)$$
$$\geq (R(\lambda_0))^n e^{-\lambda_0(cn+b\sqrt{n})} \int ... \int_{cn<x_1+...+x_n<cn+b\sqrt{n}} dF_{\lambda_0}(x_1)...dF_{\lambda_0}(x_n).$$

The latter integral, as in the case of Theorem 10.12, converges to a positive limit by the Central Limit Theorem. □

In Theorems 10.12 and 10.13 the number $R(\lambda_0)e^{-\lambda_0 c} = r(\lambda_0)$ is involved. It is clear that $r(0) = 1$. Let us show that $r(\lambda_0) < 1$. We have

$$\ln r(\lambda_0) = \ln R(\lambda_0) - \lambda_0 c = \ln R(\lambda_0) - \ln R(0) - \lambda_0 c.$$

By Taylor's formula,

$$\ln R(\lambda_0) - \ln R(0) = \lambda_0 (\ln R)'(\lambda_0) - \frac{\lambda_0^2}{2}(\ln R)''(\lambda_1),$$

where $\lambda_1$ is an intermediate point between 0 and $\lambda_0$. Furthermore,

$$(\ln R)'(\lambda_0) = \frac{R'(\lambda_0)}{R(\lambda_0)} = m(\lambda_0) = c, \quad \text{and} \quad (\ln R)''(\lambda_1) > 0,$$

since it is the variance of the distribution $F_{\lambda_1}$. Thus

$$\ln r(\lambda_0) = -\frac{\lambda_0^2}{2}(\ln R)''(\lambda_1) < 0.$$

From Theorems 10.12 and 10.13 we obtain the following corollary.

**Corollary 10.14.**

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathrm{P}_{n,c} = \ln r(\lambda_0) < 0.$$

*Proof.* Indeed, let $b = 1$ in Theorem 10.13. Then

$$\ln r(\lambda_0) - \frac{\lambda_0}{\sqrt{n}} - \frac{\ln p_n}{n} \leq \frac{\ln \mathrm{P}_{n,c}}{n} \leq \ln r(\lambda_0) + \frac{1}{n} \ln B_n.$$

We complete the proof by taking the limit as $n \to \infty$. □

This corollary shows that the probabilities $\mathrm{P}_{n,c}$ decay exponentially in $n$. In other words, they decay much faster than suggested by the Chebyshev Inequality.

## 10.5 Other Limit Theorems

The Central Limit Theorem applies to sums of independent identically distributed random variables when the variances of these variables are finite. When the variances are infinite, different Limit Theorems may apply, giving different limiting distributions.

As an example, we consider a sequence of independent identically distributed random variables $\xi_1, \xi_2, ...$, whose distribution is given by a symmetric density $p(x)$, $p(x) = p(-x)$, such that

$$p(x) \sim \frac{c}{|x|^{\alpha+1}} \quad \text{as} \quad |x| \to \infty, \tag{10.17}$$

where $0 < \alpha < 2$ and $c$ is a constant. The condition of symmetry is imposed for the sake of simplicity. Consider the normalized sum

$$\eta_n = \frac{\xi_1 + ... + \xi_n}{n^{\frac{1}{\alpha}}}.$$

**Theorem 10.15.** *As $n \to \infty$, the distributions of $\eta_n$ converge weakly to a limiting distribution whose characteristic function is $\psi(\lambda) = e^{-c_1|\lambda|^\alpha}$, where $c_1$ is a function of $c$.*

*Remark 10.16.* For $\alpha = 2$, the convergence to the Gaussian distribution is also true, but the normalization of the sum is different:

$$\eta_n = \frac{\xi_1 + ... + \xi_n}{n^{\frac{1}{2}} \ln n}.$$

*Remark 10.17.* For $\alpha = 1$ we have the convergence to the Cauchy distribution.

In order to prove Theorem 10.15, we shall need the following lemma.

**Lemma 10.18.** *Let $\varphi(\lambda)$ be the characteristic function of the random variables $\xi_1$, $\xi_2, \ldots$. Then,*

$$\varphi(\lambda) = 1 - c_1|\lambda|^\alpha + o(|\lambda|^\alpha) \quad \textit{as} \quad \lambda \to 0.$$

*Remark 10.19.* This is a particular case of the so-called Tauberian Theorems, which relate the behavior of a distribution at infinity to the behavior of the characteristic function near $\lambda = 0$.

*Proof.* Take a constant $M$ large enough, so that the density $p(x)$ can be represented as $p(x) = \frac{c(1+g(x))}{|x|^{\alpha+1}}$ for $|x| \geq M$, where $g(x)$ is a bounded function, $g(x) \to 0$ as $|x| \to \infty$. For simplicity of notation, assume that $\lambda \to 0+$. For $\lambda < 1/M$ we break the integral defining $\varphi(\lambda)$ into five parts:

$$\begin{aligned}
\varphi(\lambda) = \int_{-\infty}^{-\frac{1}{\lambda}} p(x)e^{i\lambda x}dx + \int_{-\frac{1}{\lambda}}^{-M} p(x)e^{i\lambda x}dx + \int_{-M}^{M} p(x)e^{i\lambda x}dx \\
+ \int_{M}^{\frac{1}{\lambda}} p(x)e^{i\lambda x}dx + \int_{\frac{1}{\lambda}}^{\infty} p(x)e^{i\lambda x}dx \\
= I_1(\lambda) + I_2(\lambda) + I_3(\lambda) + I_4(\lambda) + I_5(\lambda).
\end{aligned}$$

The integral $I_3(\lambda)$ is a holomorphic function of $\lambda$ equal to $\int_{-M}^{M} p(x)dx$ at $\lambda = 0$. The derivative $I_3'(0)$ is equal to $\int_{-M}^{M} p(x)ixdx = 0$, since $p(x)$ is an even function. Therefore, for any fixed $M$

$$I_3(\lambda) = \int_{-M}^{M} p(x)dx + O(\lambda^2) \quad \text{as} \quad \lambda \to 0.$$

Using a change of variables and the Dominated Convergence Theorem, we obtain

$$\begin{aligned}
I_1(\lambda) = \int_{-\infty}^{-\frac{1}{\lambda}} p(x)e^{i\lambda x}dx = \int_{-\infty}^{-\frac{1}{\lambda}} \frac{c(1+g(x))}{|x|^{\alpha+1}} e^{i\lambda x}dx \\
= c\lambda^\alpha \int_{-\infty}^{-1} \frac{(1+g(\frac{y}{\lambda}))}{|y|^{\alpha+1}} e^{iy}dy \sim c\lambda^\alpha \int_{-\infty}^{-1} \frac{e^{iy}}{|y|^{\alpha+1}}dy.
\end{aligned}$$

Similarly,

$$I_5(\lambda) \sim c\lambda^\alpha \int_{1}^{\infty} \frac{e^{iy}}{|y|^{\alpha+1}}dy.$$

Next, since $p(x)$ is an even function,

$$I_2(\lambda) + I_4(\lambda) = \int_{-\frac{1}{\lambda}}^{-M} p(x)(e^{i\lambda x} - 1 - i\lambda x)dx + \int_{M}^{\frac{1}{\lambda}} p(x)(e^{i\lambda x} - 1 - i\lambda x)dx$$

$$+\int_{-\frac{1}{\lambda}}^{-M} p(x)dx + \int_{M}^{\frac{1}{\lambda}} p(x)dx. \tag{10.18}$$

The third term on the right-hand side is equal to

$$\int_{-\frac{1}{\lambda}}^{-M} p(x)dx = \int_{-\infty}^{-M} p(x)dx - \int_{-\infty}^{-\frac{1}{\lambda}} \frac{c(1+g(x))}{|x|^{\alpha+1}}dx$$

$$= \int_{-\infty}^{-M} p(x)dx + c_0\lambda^{\alpha} + o(\lambda^{\alpha}) \ ,$$

where $c_0$ is some constant. Similarly,

$$\int_{M}^{\frac{1}{\lambda}} p(x)dx = \int_{M}^{\infty} p(x)dx + c_0\lambda^{\alpha} + o(\lambda^{\alpha}).$$

The first two terms on the right-hand side of (10.18) can be treated with the help of the same change of variables that was used to find the asymptotics of $I_1(\lambda)$. Therefore, taking into account the asymptotic behavior of each term, we obtain

$$I_1(\lambda) + I_2(\lambda) + I_3(\lambda) + I_4(\lambda) + I_5(\lambda)$$
$$= \int_{-\infty}^{\infty} p(x)dx - c_1\lambda^{\alpha} + o(\lambda^{\alpha}) = 1 - c_1\lambda^{\alpha} + o(\lambda^{\alpha}),$$

where $c_1$ is another constant. □

*Proof of Theorem 10.15.* The characteristic function of $\eta_n$ has the form

$$\varphi_{\eta_n}(\lambda) = \mathrm{E}e^{i\lambda\frac{\xi_1+\ldots+\xi_n}{n^{1/\alpha}}} = (\varphi(\frac{\lambda}{n^{1/\alpha}}))^n.$$

In our case, $\lambda$ is fixed and $n \to \infty$. Therefore we can use Lemma 10.18 to conclude

$$(\varphi(\frac{\lambda}{n^{1/\alpha}}))^n = (1 - \frac{c_1|\lambda|^{\alpha}}{n} + o(\frac{1}{n}))^n \to e^{-c_1|\lambda|^{\alpha}}.$$

By remark 9.11, the function $e^{-c_1|\lambda|^{\alpha}}$ is a characteristic function of some distribution. □

Consider a sequence of independent identically distributed random variables $\xi_1, \xi_2, \ldots$ with zero expectation. While both Theorem 10.15 and the Central Limit Theorem state that the normalized sums of the random variables converge weakly, there is a crucial difference in the mechanisms of convergence. Let us show that, in the case of the Central Limit Theorem, the contribution of each individual term to the sum is negligible. This is not so in the situation described by Theorem 10.15. For random variables with distributions of the form (10.17), the largest term of the sum is commensurate with the entire sum.

First consider the situation described by the Central Limit Theorem. Let $F(x)$ be the distribution function of each of the random variables $\xi_1, \xi_2, ...$, which have finite variance. Then, for each $a > 0$, we have

$$n\mathrm{P}(|\xi_1| \geq a\sqrt{n}) = n \int_{|x| \geq a\sqrt{n}} dF(x) \leq \frac{1}{a^2} \int_{|x| \geq a\sqrt{n}} x^2 dF(x).$$

The last integral converges to zero as $n \to \infty$ since $\int_{\mathbb{R}} x^2 dF(x)$ is finite.

The Central Limit Theorem states that the sum $\xi_1 + ... + \xi_n$ is of order $\sqrt{n}$ for large $n$. We can estimate the probability that the largest term in the sum is greater than $a\sqrt{n}$ for $a > 0$. Due to the independence of the random variables,

$$\mathrm{P}(\max_{1 \leq i \leq n} |\xi_i| \geq a\sqrt{n}) \leq n\mathrm{P}(|\xi_1| \geq a\sqrt{n}) \to 0 \quad \text{as} \quad n \to \infty.$$

Let us now assume that the distribution of each random variable is given by a symmetric density $p(x)$ for which (10.17) holds. Theorem 10.15 states that the sum $\xi_1 + ... + \xi_n$ is of order $n^{\frac{1}{\alpha}}$ for large $n$. For $a > 0$ we can estimate from below the probability that the largest term in the sum is greater than $an^{\frac{1}{\alpha}}$. Namely,

$$\mathrm{P}(\max_{1 \leq i \leq n} |\xi_i| \geq an^{\frac{1}{\alpha}}) = 1 - \mathrm{P}(\max_{1 \leq i \leq n} |\xi_i| < an^{\frac{1}{\alpha}}) = 1 - (\mathrm{P}(|\xi_1| < an^{\frac{1}{\alpha}}))^n$$

$$= 1 - (1 - \mathrm{P}(|\xi_1| \geq an^{\frac{1}{\alpha}}))^n.$$

By (10.17),

$$\mathrm{P}(|\xi_1| \geq an^{\frac{1}{\alpha}}) \sim \int_{|x| \geq an^{\frac{1}{\alpha}}} \frac{c}{|x|^{\alpha+1}} dx = \frac{2c}{\alpha a^\alpha n}.$$

Therefore,

$$\lim_{n \to \infty} \mathrm{P}(\max_{1 \leq i \leq n} |\xi_i| \geq an^{\frac{1}{\alpha}}) = \lim_{n \to \infty} (1 - (1 - \frac{2c}{\alpha a^\alpha n})^n) = 1 - \exp(-\frac{2c}{\alpha a^\alpha}) > 0.$$

This justifies our remarks on the mechanism of convergence of sums of random variables with densities satisfying (10.17).

Consider an arbitrary sequence of independent identically distributed random variables $\xi_1, \xi_2, ....$ Assume that for some $A_n$, $B_n$ the distributions of the normalized sums

$$\frac{\xi_1 + ... + \xi_n - A_n}{B_n} \tag{10.19}$$

converge weakly to a non-trivial limit.

**Definition 10.20.** *A distribution which can appear as a limit of normalized sums (10.19) for some sequence of independent identically distributed random variables $\xi_1, \xi_2, ...$ and some sequences $A_n$, $B_n$ is called a stable distribution.*

There is a general formula for characteristic functions of stable distributions. It is possible to show that the sequences $A_n$, $B_n$ cannot be arbitrary. They are always products of power functions and the so-called "slowly changing" functions, for which a typical example is any power of the logarithm.

Finally, we consider a Limit Theorem for a particular problem in one-dimensional random walks. It provides another example of a proof of a Limit Theorem with the help of characteristic functions. Let $\xi_1, \xi_2, ...$ be the consecutive moments of return of a simple symmetric one-dimensional random walk to the origin. In this case $\xi_1$, $\xi_2 - \xi_1$, $\xi_3 - \xi_2, ...$ are independent identically distributed random variables. We shall prove that the distributions of $\xi_n/n^2$ converge weakly to a non-trivial distribution.

Let us examine the characteristic function of the random variable $\xi_1$. Recall that in Section 6.2 we showed that the generating function of $\xi_1$ is equal to

$$F(z) = \mathrm{E}z^{\xi_1} = 1 - \sqrt{1 - z^2}.$$

This formula holds for $|z| < 1$, and can be extended by continuity to the unit circle $|z| = 1$. Here, the branch of the square root with the non-negative real part is selected. Now

$$\varphi(\lambda) = \mathrm{E}e^{i\lambda\xi_1} = \mathrm{E}(e^{i\lambda})^{\xi_1} = 1 - \sqrt{1 - e^{2i\lambda}}.$$

Since $\xi_n$ is a sum of independent identically distributed random variables, the characteristic function of $\frac{\xi_n}{n^2}$ is equal to

$$(\varphi(\frac{\lambda}{n^2}))^n = (1 - \sqrt{1 - e^{\frac{2i\lambda}{n^2}}})^n = (1 - \frac{\sqrt{-2i\lambda}}{n} + o(\frac{1}{n}))^n \sim e^{\sqrt{-2i\lambda}}.$$

By Remark 9.11, this implies that the distribution of $\frac{\xi_n}{n^2}$ converges weakly to the distribution with the characteristic function $e^{\sqrt{-2i\lambda}}$.

## 10.6 Problems

**1.** Prove the following Central Limit Theorem for independent identically distributed random vectors. Let $\xi_1 = (\xi_1^{(1)}, ..., \xi_1^{(k)}), \xi_2 = (\xi_2^{(1)}, ..., \xi_2^{(k)}), ...$ be a sequence of independent identically distributed random vectors in $\mathbb{R}^k$. Let $m$ and $D$ be the expectation and the covariance matrix, respectively, of the random vector $\xi_1$. That is,

$$m = (m^1, ..., m^k), \;\; m^i = \mathrm{E}\xi_1^{(i)}, \;\; \text{and} \;\; D = (d^{ij})_{1 \le i,j \le k}, \;\; d^{ij} = \mathrm{Cov}(\xi_1^{(i)}, \xi_1^{(j)}).$$

Assume that $|d^{ij}| < \infty$ for all $i, j$. Prove that the distributions of

$$(\xi_1 + .. + \xi_n - nm)/\sqrt{n}$$

converge weakly to $N(0, D)$ distribution as $n \to \infty$.

**2.** Two people are playing a series of games against each other. In each game each player either wins a certain amount of money or loses the same amount of money, both with probability 1/2. With each new game the stake increases by a dollar. Let $S_n$ denote the change of the fortune of the first player by the end of the first $n$ games.

(a) Find a function $f(n)$ such that the random variables $S_n/f(n)$ converge in distribution to some limit which is not a distribution concentrated at zero and identify the limiting distribution.

(b) If $R_n$ denotes the change of the fortune of the second player by the end of the first $n$ games, what is the limit, in distribution, of the random vectors $(S_n/f(n), R_n/f(n))$?

**3.** Let $\xi_1, \xi_2, ...$ be a sequence of independent identically distributed random variables with $\mathrm{E}\xi_1 = 0$ and $0 < \sigma^2 = \mathrm{Var}(\xi_1) < \infty$. Prove that the distributions of $(\sum_{i=1}^n \xi_i)/(\sum_{i=1}^n \xi_i^2)^{1/2}$ converge weakly to $N(0, 1)$ distribution as $n \to \infty$.

**4.** Let $\xi_1, \xi_2, ...$ be independent identically distributed random variables such that $\mathrm{P}(\xi_n = -1) = \mathrm{P}(\xi_n = 1) = 1/2$. Let $\zeta_n = \sum_{i=1}^n \xi_i$. Prove that

$$\lim_{n\to\infty} \mathrm{P}(\zeta_n = k^2 \ \text{ for some } k \in \mathbb{N}) = 0.$$

**5.** Let $\omega = (\omega_0, \omega_1, ...)$ be a trajectory of a simple symmetric random walk on $\mathbb{Z}^3$. Prove that for any $\varepsilon > 0$

$$\mathrm{P}(\lim_{n\to\infty} (n^{\varepsilon - \frac{1}{6}} ||\omega_n||) = \infty) = 1.$$

**6.** Let $\xi_1, \xi_2, ...$ be independent identically distributed random variables such that $\mathrm{P}(\xi_n = -1) = \mathrm{P}(\xi_n = 1) = 1/2$. Let $\zeta_n = \sum_{i=1}^n \xi_i$. Find the limit

$$\lim_{n\to\infty} \frac{\ln \mathrm{P}((\zeta_n/n) > \varepsilon)}{n}.$$

**7.** Let $\xi_1, \xi_2, ...$ be independent identically distributed random variables with the Cauchy distribution. Prove that

$$\liminf_{n\to\infty} \mathrm{P}(\max(\xi_1, ..., \xi_n) > xn) \geq \exp(-\pi x).$$

for any $x \geq 0$.

**8.** Let $\xi_1, \xi_2, ...$ be independent identically distributed random variables with the uniform distribution on the interval $[-1/2, 1/2]$. What is the limit (in distribution) of the sequence

$$\zeta_n = (\sum_{i=1}^{n} 1/\xi_i)/n.$$

**9.** Let $\xi_1, \xi_2, ...$ be independent random variables with uniform distribution on $[0, 1]$. Given $\alpha \in \mathbb{R}$, find $a_n$ and $b_n$ such that the sequence

$$(\sum_{i=1}^{n} i^\alpha \xi_i - a_n)/b_n$$

converges in distribution to a limit which is different from zero.

**10.** Let $\xi_1, \xi_2, ...$ be independent random variables with uniform distribution on [0,1]. Show that for any continuous function $f(x, y, z)$ on $[0, 1]^3$

$$\frac{1}{\sqrt{n}}(\sum_{j=1}^{n} f(\xi_j, \xi_{j+1}, \xi_{j+2}) - n \int_0^1 \int_0^1 \int_0^1 f(x, y, z) dx dy dz)$$

converges in distribution.

http://www.springer.com/978-3-540-25484-3

Theory of Probability and Random Processes
Koralov, L.; Sinai, Y.G.
2007, XI, 358 p., Softcover
ISBN: 978-3-540-25484-3