3 Weak Law of Large Numbers

As previously: $0 < a_n \uparrow \infty$.

In the sequel: $(X_n)_{n\in\mathbb{N}}$ pairwise uncorrelated.

Particular case, $(X_n)_{n\in\mathbb{N}}$ pairwise independent and $X_n \in \mathfrak{L}^2$ for every $n \in \mathbb{N}$, see Theorem III.5.6.

Theorem 1 (Khintchine)**.** If

$$\lim_{n\to\infty} \frac{1}{a_n^2} \cdot \sum_{i=1}^n \mathrm{Var}(X_i) = 0,$$

then

$$\frac{1}{a_n} \cdot \sum_{i=1}^n (X_i - \mathrm{E}(X_i)) \xrightarrow{P} 0.$$

Proof. Without loss of generality $\mathrm{E}(X_n) = 0$ for every $n \in \mathbb{N}$. For $\varepsilon > 0$ the Chebyshev-Markov inequality and Bienaymé's Theorem yield

$$P\Big(\Big\{\Big|\frac{1}{a_n} \cdot S_n\Big| \geq \varepsilon\Big\}\Big) \leq \frac{1}{\varepsilon^2} \cdot \mathrm{Var}\Big(\frac{1}{a_n} \cdot S_n\Big) = \frac{1}{\varepsilon^2} \cdot \frac{1}{a_n^2} \cdot \sum_{i=1}^n \mathrm{Var}(X_i). \qquad (1)$$

□

Remark 1. Assume that $\sup_{n\in\mathbb{N}} \mathrm{Var}(X_n) < \infty$. Then Theorem 1 is applicable for any sequence $(a_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} a_n/\sqrt{n} = \infty$.

Example 1. Consider an independent sequence $(X_n)_{n\in\mathbb{N}}$ with

$$P(\{X_n = 0\}) = 1 - \frac{1}{n\log(n+1)}, \qquad P(\{X_n = \pm n\}) = \frac{1}{2n\log(n+1)}.$$

Hence

$$\mathrm{E}(X_n) = 0, \qquad \mathrm{Var}(X_n) = \frac{n}{\log(n+1)},$$

and

$$\frac{1}{n^2} \cdot \sum_{i=1}^n \mathrm{Var}(X_i) \leq \frac{1}{\log(n+1)}.$$

Thus $1/n \cdot S_n \xrightarrow{P} 0$ due to Theorem 1, but $1/n \cdot S_n \xrightarrow{P\text{-a.s.}} 0$ does not hold, see `Übung` 11.2.

4 Characteristic Functions

We use the notation $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ for the euclidean inner product and norm. Recall that $\mathfrak{M}(\mathbb{R}^k)$ denotes the class of all probability measures on $(\mathbb{R}^k, \mathfrak{B}_k)$.

Given: a probability measure $\mu \in \mathfrak{M}(\mathbb{R}^k)$.

Definition 1. $f : \mathbb{R}^k \to \mathbb{C}$ is *μ-integrable* if $\Re f$ and $\Im f$ are μ-integrable, in which case

$$\int f \, d\mu = \int \Re f \, d\mu + \imath \cdot \int \Im f \, d\mu.$$

Definition 2. The mapping $\widehat{\mu} : \mathbb{R}^k \to \mathbb{C}$ with

$$\widehat{\mu}(y) = \int \exp(\imath \langle x, y \rangle) \, \mu(dx), \qquad y \in \mathbb{R}^k,$$

is called the *Fourier transform* of μ.

Example 1.

(i) For a discrete probability measure

$$\mu = \sum_{j=1}^{\infty} \alpha_j \cdot \varepsilon_{x_j}$$

we have

$$\widehat{\mu}(y) = \sum_{j=1}^{\infty} \alpha_j \cdot \exp(\imath \langle x_j, y \rangle).$$

For instance, if $\mu = \pi(\lambda)$ is the Poisson distribution with parameter $\lambda > 0$, then

$$\begin{aligned}\widehat{\mu}(y) &= \exp(-\lambda) \cdot \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \cdot \exp(\imath j y) = \exp(-\lambda) \cdot \exp(\lambda \cdot \exp(\imath y)) \\ &= \exp(\lambda \cdot (\exp(\imath y) - 1)).\end{aligned}$$

(ii) If $\mu = f \cdot \lambda_k$ then

$$\widehat{\mu}(y) = \int \exp(\imath \langle x, y \rangle) \cdot f(x) \, \lambda_k(dx).$$

For any λ_k-integrable function f, the right-hand side defines its *Fourier transform*, see also Analysis or Funktionalanalysis. For instance, if μ is the k-dimensional standard normal distribution, i.e.,

$$f(x) = (2\pi)^{-k/2} \cdot \exp(-\|x\|^2/2),$$

then

$$\widehat{\mu}(y) = \exp(-\|y\|^2/2).$$

See Bauer (1996, p. 187) for the case $k = 1$. Use Fubini's Theorem for $k > 1$.

Theorem 1.

(i) $\widehat{\mu}$ is uniformly continuous on $\mathbb{R}^k$,

(ii) $|\widehat{\mu}(y)| \leq 1 = \widehat{\mu}(0)$ for $y \in \mathbb{R}^k$,

(iii) for $n \in \mathbb{N}$, $a_1, \dots, a_n \in \mathbb{C}$, and $y_1, \dots, y_n \in \mathbb{R}^k$,

$$\sum_{j,\ell=1}^{n} a_j \cdot \overline{a_\ell} \cdot \widehat{\mu}(y_j - y_\ell) \geq 0$$

(*positive semi-definite*).

Proof. Ad (i): Observe that

$$\left|\exp(\imath\langle x, y_1\rangle) - \exp(\imath\langle x, y_2\rangle)\right| \leq \|x\| \cdot \|y_1 - y_2\|.$$

For $\varepsilon > 0$ take $r > 0$ such that $\mu(B) \geq 1 - \varepsilon$, where $B = \{x \in \mathbb{R}^k : \|x\| \leq r\}$. Then

$$\begin{aligned} \left|\widehat{\mu}(y_1) - \widehat{\mu}(y_2)\right| &\leq \int_B \left|\exp(\imath\langle x, y_1\rangle) - \exp(\imath\langle x, y_2\rangle)\right| \mu(dx) + 2 \cdot \varepsilon \\ &\leq r \cdot \|y_1 - y_2\| + 2 \cdot \varepsilon. \end{aligned}$$

Properties (ii) and (iii) are easily verified. □

Remark 1. *Bochner's Theorem* states that every continuous, positive semi-definite function $\varphi : \mathbb{R}^k \to \mathbb{C}$ with $\varphi(0) = 1$ is the Fourier transform of a probability measure on $(\mathbb{R}^k, \mathfrak{B}_k)$. See Bauer (1996, p. 184) for references.

In the sequel: $X, Y, \dots$ are k-dimensional random vectors on a probability space $(\Omega, \mathfrak{A}, P)$.

Definition 3. The *characteristic function* of X is given by

$$\varphi_X = \widehat{P_X}.$$

Remark 2. Due to Theorem II.9.1

$$\varphi_X(y) = \int_{\mathbb{R}^k} \exp(\imath\langle x, y\rangle)\, P_X(dx) = \int_\Omega \exp(\imath\langle X(\omega), y\rangle)\, P(d\omega).$$

Theorem 2.

(i) For every linear mapping $T : \mathbb{R}^k \to \mathbb{R}^\ell$

$$\varphi_{T \circ X} = \varphi_X \circ T^t.$$

(ii) For independent random vectors X and Y

$$\varphi_{X+Y} = \varphi_X \cdot \varphi_Y.$$

In particular, for $a \in \mathbb{R}^k$,

$$\varphi_{X+a} = \exp(\imath\langle a, \cdot\rangle) \cdot \varphi_X.$$

Proof. Ad (i): Let $z \in \mathbb{R}^\ell$. Use $P_{T\circ X} = T(P_X)$ to obtain

$$\varphi_{T\circ X}(z) = \int_{\mathbb{R}^k} \exp(\imath\langle T(x), z\rangle)\, P_X(dx) = \varphi_X(T^t(z)).$$

Ad (ii): Let $z \in \mathbb{R}^k$. Fubini's Theorem and Theorem III.5.5 imply

$$\varphi_{X+Y}(z) = \int_{\mathbb{R}^{2k}} \exp(\imath\langle x+y, z\rangle)\, P_{(X,Y)}(d(x,y)) = \varphi_X(z)\cdot\varphi_Y(z).$$

□

Corollary 1 (Convolution Theorem)**.** For probability measures $\mu_j \in \mathfrak{M}(\mathbb{R})$,

$$\widehat{\mu_1 * \mu_2} = \widehat{\mu}_1 \cdot \widehat{\mu}_2.$$

Proof. Use Theorem 2.(ii) and Theorem III.5.8. □

Example 2. For $\mu = N(m, \sigma^2)$ with $\sigma \geq 0$ and $m \in \mathbb{R}$

$$\widehat{\mu}(y) = \exp(\imath m y)\cdot \exp(-\sigma^2 y^2/2).$$

See Example 1.(ii) and Theorem 2.

Lemma 1. For $z \in \mathbb{R}$ and $\sigma > 0$

$$\int \exp(-\imath y z)\cdot \widehat{\mu}(y)\, N(0, \sigma^{-2})(dy) = \int \exp\big(-(z-x)^2/(2\sigma^2)\big)\, \mu(dx).$$

Proof. See Gänssler, Stute (1977, p. 92). □

Lemma 2. For $\sigma_n > 0$ with $\lim_{n\to\infty} \sigma_n = 0$,

$$N(0, \sigma_n^2) * \mu \xrightarrow{\text{w}} \mu.$$

Proof. Consider independent random variables X_n and Y such that $X_n \sim N(0, \sigma_n^2)$ and $Y \sim \mu$. Then $X_n \xrightarrow{\mathfrak{L}^2} 0$, and therefore $X_n + Y \xrightarrow{\mathfrak{L}^2} Y$, which implies

$$X_n + Y \xrightarrow{\text{d}} Y.$$

□

Theorem 3 (Uniqueness Theorem)**.** For probability measures $\mu_j \in \mathfrak{M}(\mathbb{R}^k)$,

$$\mu_1 = \mu_2 \qquad \Leftrightarrow \qquad \widehat{\mu}_1 = \widehat{\mu}_2.$$

Proof. '⇒' holds by definition. '⇐': See Bauer (1996, Thm. 23.4) or Billingsley (1979, Sec. 29) for the case $k > 1$. Here: the case $k = 1$. For $\sigma > 0$ and $A \in \mathfrak{B}$

$$N(0,\sigma^2) * \mu_j(A) = \int\int 1_A(z+x)\, N(0,\sigma^2)(dz)\, \mu_j(dx),$$

and

$$\int 1_A(z+x)\, N(0,\sigma^2)(dz) = (2\pi\sigma^2)^{-1/2} \cdot \int 1_A(z+x) \cdot \exp(-z^2/(2\sigma^2))\, \lambda_1(dz)$$
$$= (2\pi\sigma^2)^{-1/2} \cdot \int_A \exp(-(z-x)^2/(2\sigma^2))\, \lambda_1(dz).$$

Therefore

$$N(0,\sigma^2) * \mu_j(A) = (2\pi\sigma^2)^{-1/2} \cdot \int_A \int \exp(-(z-x)^2/(2\sigma^2))\, \mu_j(dx)\, \lambda_1(dz).$$

Use Lemma 1 to conclude that

$$\forall\, \sigma > 0: \; N(0,\sigma^2) * \mu_1 = N(0,\sigma^2) * \mu_2.$$

Then, by Lemma 2 and Corollary III.3.1, $\mu_1 = \mu_2$. □

Example 3. For independent random variables X_1 and X_2 with $X_j \sim \pi(\lambda_j)$ we have $X_1 + X_2 \sim \pi(\lambda_1 + \lambda_2)$.
Proof: Theorem 2 and Example 1.(i) yield

$$\varphi_{X_1+X_2}(y) = \exp(\lambda_1 \cdot (\exp(\imath y) - 1)) \cdot \exp(\lambda_2 \cdot (\exp(\imath y) - 1))$$
$$= \exp((\lambda_1 + \lambda_2) \cdot (\exp(\imath y) - 1)).$$

Use Theorem 3.

Lemma 3. For every $\varepsilon > 0$ and every probability measure $\mu \in \mathfrak{M}(\mathbb{R})$,

$$\mu(\{x \in \mathbb{R} : |x| \geq 1/\varepsilon\}) \leq 7/\varepsilon \cdot \int_0^\varepsilon (1 - \Re\widehat{\mu}(y))\, dy.$$

Proof. Clearly

$$\Re\widehat{\mu}(y) = \int_{\mathbb{R}} \cos(xy)\, \mu(dx).$$

Hence, with the convention $\sin(0)/0 = 1$,

$$1/\varepsilon \cdot \int_0^\varepsilon (1 - \Re\widehat{\mu}(y))\, dy = 1/\varepsilon \cdot \int_{[0,\varepsilon]} \int_{\mathbb{R}} (1 - \cos(xy))\, \mu(dx)\, \lambda_1(dy)$$
$$= \int_{\mathbb{R}} \left(1/\varepsilon \cdot \int_0^\varepsilon (1 - \cos(xy))\, dy \right) \mu(dx)$$
$$= \int_{\mathbb{R}} (1 - \sin(\varepsilon x)/(\varepsilon x))\, \mu(dx)$$
$$\geq \inf_{|z| \geq 1} (1 - \sin(z)/z) \cdot \mu(\{x \in \mathbb{R} : |\varepsilon x| \geq 1\}).$$

Finally,

$$\inf_{|z| \geq 1} (1 - \sin(z)/z) \geq 1/7.$$

□

Theorem 4 (Continuity Theorem, Lévy)**.**

(i) Let $\mu, \mu_n \in \mathfrak{M}(\mathbb{R}^k)$ for $n \in \mathbb{N}$. Then

$$\mu_n \xrightarrow{\text{w}} \mu \qquad \Rightarrow \qquad \forall\, y \in \mathbb{R}^k : \lim_{n\to\infty} \widehat{\mu_n}(y) = \widehat{\mu}(y).$$

(ii) Let $\mu_n \in \mathfrak{M}(\mathbb{R}^k)$ for $n \in \mathbb{N}$, and let $\varphi : \mathbb{R}^k \to \mathbb{C}$ be continuous at 0. Then

$$\forall\, y \in \mathbb{R}^k : \lim_{n\to\infty} \widehat{\mu_n}(y) = \varphi(y) \qquad \Rightarrow \qquad \exists\, \mu \in \mathfrak{M}(\mathbb{R}^k) : \widehat{\mu} = \varphi \wedge \mu_n \xrightarrow{\text{w}} \mu.$$

Proof. Ad (i): Note that $x \mapsto \exp(\imath\langle x, y\rangle)$ is bounded and continuous on $\mathbb{R}^k$.
Ad (ii): See Bauer (1996, Thm. 23.8) or Billingsley (1979, Sec. 29) for the case $k > 1$. Here: the case $k = 1$.
We first show that

$$\{\mu_n : n \in \mathbb{N}\} \text{ is tight.} \tag{1}$$

By Lemma 3

$$\mu_n(\{x \in \mathbb{R} : |x| \geq 1/\varepsilon\}) \leq c_n(\varepsilon)$$

with

$$c_n(\varepsilon) = 7/\varepsilon \cdot \int_0^\varepsilon (1 - \Re\widehat{\mu_n}(y))\, dy.$$

The dominated convergence theorem and the continuity of φ at 0 yield

$$\lim_{n\to\infty} c_n(\varepsilon) = c(\varepsilon)$$

with

$$c(\varepsilon) = 7/\varepsilon \cdot \int_0^\varepsilon (1 - \Re\widehat{\varphi}(y))\, dy,$$

if ε is sufficiently small. Given $\delta > 0$ take $\varepsilon > 0$ such that

$$c(\varepsilon) \leq \delta/2.$$

Furthermore, take $n_0 \in \mathbb{N}$ such that, for every $n \geq n_0$,

$$|c_n(\varepsilon) - c(\varepsilon)| \leq \delta/2.$$

Hence, for $n \geq n_0$,

$$\mu_n(\{x \in \mathbb{R} : |x| \geq 1/\varepsilon\}) \leq \delta,$$

and hereby we get (1).
Thus, by Prohorov's Theorem,

$$\{\mu_n : n \in \mathbb{N}\} \text{ is relatively compact.} \tag{2}$$

We fix a probability measure $\mu \in \mathfrak{M}(\mathbb{R})$ such that $\mu_{n_k} \xrightarrow{\text{w}} \mu$ for a suitable subsequence of $(\mu_n)_{n\in\mathbb{N}}$. By assumption and (i), we get $\widehat{\mu} = \varphi$ as well as the following fact:

$$\text{if } \mu_{n_k} \xrightarrow{\text{w}} \nu \text{ for any subsequence } (\mu_{n_k})_{k\in\mathbb{N}}, \text{ then } \nu = \mu, \tag{3}$$

see Theorem 3.
We claim that $\mu_n \xrightarrow{\text{w}} \mu$. Due to Remarks III.2.3 and III.3.4.(ii) it suffices to show that every subsequence of $(\mu_n)_{n\in\mathbb{N}}$ contains a subsequence that converges weakly to μ. The latter property follows from (2) and (3). $\square$

Corollary 2. Weak convergence in $\mathfrak{M}(\mathbb{R}^k)$ is equivalent to pointwise convergence of Fourier transforms.

Example 4. Let $\mu_n = B(n, p_n)$ and assume that

$$\lim_{n\to\infty} n \cdot p_n = \lambda > 0.$$

Then

$$\mu_n \xrightarrow{\text{w}} \pi(\lambda).$$

Proof: Übung 11.4.

5 The Central Limit Theorem

Given: a triangular array of random variables X_{nk}, where $n \in \mathbb{N}$ and $k \in \{1, \dots, r_n\}$ with $r_n \in \mathbb{N}$.
Assumptions:

(i) $X_{nk} \in \mathfrak{L}^2$ for every $n \in \mathbb{N}$ and $k \in \{1, \dots, r_n\}$,

(ii) $(X_{n1}, \dots, X_{nr_n})$ independent for every $n \in \mathbb{N}$.

Put

$$S_n = \sum_{k=1}^{r_n} (X_{nk} - \mathrm{E}(X_{nk}))$$

and

$$\sigma_{nk}^2 = \mathrm{Var}(X_{nk}), \qquad s_n^2 = \mathrm{Var}(S_n) = \sum_{k=1}^{r_n} \sigma_{nk}^2.$$

Additional assumption:

(iii) $s_n^2 > 0$ for every $n \in \mathbb{N}$.

Normalization

$$S_n^* = \frac{1}{s_n} \cdot S_n = \sum_{k=1}^{r_n} \frac{X_{nk} - \mathrm{E}(X_{nk})}{s_n}$$

for $n \in \mathbb{N}$. Clearly

$$\mathrm{E}(S_n^*) = 0 \qquad \wedge \qquad \mathrm{Var}(S_n^*) = 1.$$

Question: convergence in distribution of $(S_n^*)_{n\in\mathbb{N}}$?
For notational convenience: all random variables X_{nk} are defined on a common probability space $(\Omega, \mathfrak{A}, P)$.

Example 1. $(X_n)_{n\in\mathbb{N}}$ i.i.d. with $X_1 \in \mathfrak{L}^2$ and $\mathrm{Var}(X_1) = \sigma^2 > 0$. Put $m = \mathrm{E}(X_1)$, take

$$r_n = n, \qquad X_{nk} = X_k.$$

Then

$$S_n^* = \frac{\sum_{k=1}^n X_k - n \cdot m}{\sqrt{n} \cdot \sigma}.$$

In the sequel we assume, without loss of generality,

$$\mathrm{E}(X_{nk}) = 0 \quad \wedge \quad s_n = 1$$

for $n \in \mathbb{N}$ and $k \in \{1, \dots, r_n\}$, hence

$$S_n^* = \sum_{k=1}^{r_n} X_{nk}$$

(Otherwise, consider the random variables $(X_{nk} - \mathrm{E}(X_{nk}))/s_n$.)

Definition 1.

(i) *Lyapunov condition*

$$\exists\, \delta > 0 : \lim_{n\to\infty} \sum_{k=1}^{r_n} \mathrm{E}(|X_{nk}|^{2+\delta}) = 0.$$

(ii) *Lindeberg condition*

$$\forall\, \varepsilon > 0 : \lim_{n\to\infty} \sum_{k=1}^{r_n} \int_{\{|X_{nk}|\geq\varepsilon\}} X_{nk}^2 \, dP = 0.$$

(iii) *Feller condition*

$$\lim_{n\to\infty} \max_{1\leq k\leq r_n} \mathrm{Var}(X_{nk}) = 0.$$

(iv) The random variables X_{nk} are *asymptotically negligible* if

$$\forall\, \varepsilon > 0 : \lim_{n\to\infty} \max_{1\leq k\leq r_n} P(\{|X_{nk}| > \varepsilon\}) = 0.$$

Lemma 1. The conditions from Definition 1 satisfy

$$\text{(i)} \Rightarrow \text{(ii)} \Rightarrow \text{(iii)} \Rightarrow \text{(iv)}.$$

Moreover, (iii) implies $\lim_{n\to\infty} r_n = \infty$.

Proof. From

$$\int_{\{|X_{nk}|\geq\varepsilon\}} X_{nk}^2 \, dP \leq \frac{1}{\varepsilon^\delta} \cdot \int_{\{|X_{nk}|\geq\varepsilon\}} |X_{nk}|^{2+\delta} \, dP \leq \frac{1}{\varepsilon^\delta} \cdot \mathrm{E}(|X_{nk}|^{2+\delta})$$

we get '(i) ⇒ (ii)'. From

$$\mathrm{Var}(X_{nk}) \leq \varepsilon^2 + \int_{\{|X_{nk}|\geq\varepsilon\}} X_{nk}^2 \, dP$$

we get '(ii) ⇒ (iii)'. The Chebyshev-Markov inequality yields '(iii) ⇒ (iv)'. Finally,

$$1 = \mathrm{Var}(S_n^*) \leq r_n \cdot \max_{1\leq k\leq r_n} \mathrm{Var}(X_{nk}),$$

so that (iii) implies $\lim_{n\to\infty} r_n = \infty$. ☐

Example 2. Example 1 continued in the case $m = 0$. We take $r_n = n$ and

$$X_{nk} = \frac{X_k}{\sqrt{n} \cdot \sigma}$$

to obtain

$$\sum_{k=1}^{n} \int_{\{|X_{nk}| \geq \varepsilon\}} X_{nk}^2 \, dP = \frac{1}{\sigma^2} \cdot \int_{\{|X_1| \geq \varepsilon \cdot \sqrt{n} \cdot \sigma\}} X_1^2 \, dP.$$

Hence the Lindeberg condition is satisfied.

In the sequel

$$\varphi_{nk} = \varphi_{X_{nk}}$$

denotes the characteristic function of X_{nk}.

Lemma 2. For $y \in \mathbb{R}$ and $\varepsilon > 0$

$$\left|\varphi_{nk}(y) - (1 - \sigma_{nk}^2/2 \cdot y^2)\right| \leq y^2 \cdot \left(\varepsilon \cdot |y| \cdot \sigma_{nk}^2 + \int_{\{|X_{nk}| \geq \varepsilon\}} X_{nk}^2 \, dP\right).$$

Proof. For $u \in \mathbb{R}$

$$\left|\exp(\imath u) - (1 + \imath u - u^2/2)\right| \leq \min(u^2,\ |u|^3/6),$$

see Billingsley (1979, Eqn. (26.4)). Hence

$$\begin{aligned}
&\left|\varphi_{nk}(y) - (1 - \sigma_{nk}^2/2 \cdot y^2)\right| \\
&\quad = \left|\mathrm{E}(\exp(\imath \cdot X_{nk} \cdot y)) - \mathrm{E}\Big(1 + \imath \cdot X_{nk} \cdot y - X_{nk}^2 \cdot y^2/2)\right| \\
&\quad \leq \mathrm{E}\big(\min(y^2 \cdot X_{nk}^2\,,\ |y|^3 \cdot |X_{nk}|^3)\big) \\
&\quad \leq |y|^3 \cdot \int_{\{|X_{nk}| < \varepsilon\}} \varepsilon \cdot X_{nk}^2 \, dP + y^2 \cdot \int_{\{|X_{nk}| \geq \varepsilon\}} X_{nk}^2 \, dP \\
&\quad \leq \varepsilon \cdot |y|^3 \cdot \sigma_{nk}^2 + y^2 \cdot \int_{\{|X_{nk}| \geq \varepsilon\}} X_{nk}^2 \, dP.
\end{aligned}$$

□

Lemma 3. Put

$$\Delta_n(y) = \prod_{k=1}^{r_n} \varphi_{nk}(y) - \exp(-y^2/2), \qquad y \in \mathbb{R}.$$

If the Lindeberg condition is satisfied, then

$$\forall\, y \in \mathbb{R} : \lim_{n \to \infty} \Delta_n(y) = 0.$$

Proof. Since $|\varphi_{nk}(y)| \leq 1$ and $|\exp(-\sigma_{nk}^2/2 \cdot y^2)| \leq 1$, we get

$$\begin{aligned}
|\Delta_n(y)| &= \Big|\prod_{k=1}^{r_n} \varphi_{nk}(y) - \prod_{k=1}^{r_n} \exp(-\sigma_{nk}^2/2 \cdot y^2)\Big| \\
&\leq \sum_{k=1}^{r_n} \left|\varphi_{nk}(y) - \exp(-\sigma_{nk}^2/2 \cdot y^2)\right|
\end{aligned}$$

by induction, see Billingsley (1979, Lemma 27.1).
We assume

$$\max_{1\le k\le r_n} \sigma_{nk}^2 \cdot y^2 \le 1\,,$$

which holds for fixed $y \in \mathbb{R}$ if n is sufficiently large, see Lemma 1. Using

$$0 \le u \le 1/2 \quad \Rightarrow \quad |\exp(-u) - (1-u)| \le u^2$$

and Lemma 2 we obtain

$$\begin{aligned} |\Delta_n(y)| &\le \sum_{k=1}^{r_n} |\varphi_{nk}(y) - (1 - \sigma_{nk}^2/2 \cdot y^2)| + \sum_{k=1}^{r_n} \sigma_{nk}^4/4 \cdot y^4 \\ &\le y^2 \cdot \Big(\varepsilon \cdot |y| + \sum_{k=1}^{r_n} \int_{\{|X_{nk}|\ge\varepsilon\}} X_{nk}^2 \, dP \Big) + y^4/4 \cdot \max_{1\le k\le r_n} \sigma_{nk}^2 \end{aligned}$$

for every $\varepsilon > 0$. Thus Lemma 1 yields

$$\limsup_{n\to\infty} |\Delta_n(y)| \le |y|^3 \cdot \varepsilon.$$

□

Theorem 1 (Central Limit Theorem)**.** The following properties are equivalent:

(i) $(X_{nk})_{n,k}$ satisfies the Lindeberg condition.

(ii) $P_{S_n^*} \xrightarrow{\text{w}} N(0,1)$ and $(X_{nk})_{n,k}$ satisfies the Feller condition.

(iii) $P_{S_n^*} \xrightarrow{\text{w}} N(0,1)$ and the random variables X_{nk} are asymptotically negligible.

Proof. '(i) $\Rightarrow$ (ii)': Due to Lemma 1 we only have to prove the weak convergence. Recall that $\widehat{\mu}(y) = \exp(-y^2/2)$ for the standard normal distribution μ. Consider the characteristic function $\varphi_n = \varphi_{S_n^*}$ of S_n^*. By Theorem 4.2.(ii)

$$\varphi_n = \prod_{k=1}^{r_n} \varphi_{nk},$$

and therefore Lemma 3 implies

$$\forall\, y \in \mathbb{R}: \lim_{n\to\infty} \varphi_n(y) = \widehat{\mu}(y).$$

It remains to apply Corollary 4.2.
Lemma 1 yields '(ii) $\Rightarrow$ (iii)'. See Billingsley (1979, p. 314–315) for the proof of '(iii) $\Rightarrow$ (i)'. □

Corollary 1. Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d. with $X_1 \in \mathfrak{L}^2$ and $\sigma^2 = \mathrm{Var}(X_1) > 0$. Then

$$\frac{\sum_{k=1}^n X_k - n \cdot \mathrm{E}(X_1)}{\sqrt{n} \cdot \sigma} \xrightarrow{\text{d}} Z$$

where $Z \sim N(0,1)$.

Proof. Theorem 1 and Example 2. □

Example 3. Example 2 continued, and Corollary 1 reformulated. Let

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{x} \exp(-u^2/2)\, du, \qquad x \in \mathbb{R},$$

denote the distribution function of the standard normal distribution, and let

$$\delta_n = \sup_{x\in\mathbb{R}} \left| P(\{S_n \le x \cdot \sqrt{n} \cdot \sigma\}) - \Phi(x) \right| = \sup_{x\in\mathbb{R}} \left| P(\{S_n \le x\}) - \Phi(x/(\sqrt{n}\cdot\sigma)) \right|. \quad (1)$$

Due to the Central Limit Theorem and Theorem III.3.2

$$\lim_{n\to\infty} \delta_n = 0.$$

Theorem 2 (Berry-Esséen)**.** Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d. with $X_1 \in \mathfrak{L}^3$, $\mathrm{E}(X_1) = 0$, and $\mathrm{Var}(X_1) = \sigma^2 > 0$. For δ_n given by (1)

$$\forall\, n \in \mathbb{N}: \ \delta_n \le \frac{6 \cdot \mathrm{E}(|X_1|^3)}{\sigma^3} \cdot \frac{1}{\sqrt{n}}.$$

Proof. See Gänssler, Stute (1977, Section 4.2). □

Example 4. Example 3 continued with

$$P_{X_1} = \frac{1}{2} \cdot (\varepsilon_1 + \varepsilon_{-1}). \quad (2)$$

Since $(-X_n)_{n\in\mathbb{N}}$ is i.i.d. as well, and since $P_{-X_1} = P_{X_1}$, we have

$$P(\{S_{2n} \le 0\}) = P(\{S_{2n} \ge 0\}),$$

which yields

$$P(\{S_{2n} \le 0\}) = \frac{1}{2} \cdot (1 + P(\{S_{2n} = 0\})).$$

From Example 1.3 we know that

$$P(\{S_{2n} = 0\}) \approx \frac{1}{\sqrt{\pi n}}\,,$$

and therefore

$$\delta_{2n} \ge P(\{S_{2n} \le 0\}) - \frac{1}{2} = \frac{1}{2} \cdot P(\{S_{2n} = 0\}) \approx \frac{1}{2\sqrt{\pi n}}.$$

Hence the upper bound from Theorem 2 cannot be improved in terms of powers of n.

Example 5. Example 3 continued, i.e., $(X_n)_{n\in\mathbb{N}}$ is i.i.d. with $X_1 \in \mathfrak{L}^2$, $\mathrm{E}(X_1) = 0$, and $\mathrm{Var}(X_1) = \sigma^2 > 0$. Recall that $S_n = \sum_{i=1}^n X_i$.
Let

$$B_c = \{\limsup_{n\to\infty} S_n/\sqrt{n} \ge c\}, \qquad c > 0.$$

Using Remark 1.2.(ii) we get

$$P(B_c) \geq P(\limsup_{n\to\infty}\{S_n/\sqrt{n} > c\}) \geq \limsup_{n\to\infty} P(\{S_n/\sqrt{n} > c\}) = 1 - \Phi(c/\sigma) > 0.$$

Kolmogorov's Zero-One Law yields

$$P(B_c) = 1,$$

and therefore

$$P(\{\limsup_{n\to\infty} S_n/\sqrt{n} = \infty\}) = P\left(\bigcap_{c\in\mathbb{N}} B_c\right) = 1.$$

By symmetry

$$P(\{\liminf_{n\to\infty} S_n/\sqrt{n} = -\infty\}) = 1.$$

In particular, for P_{X_1} given by (2),

$$P(\limsup_{n\to\infty}\{S_n = 0\}) = 1,$$

see also Example 1.3 and Übung 10.2.

6 Law of the Iterated Logarithm

Given: an i.i.d. sequence $(X_n)_{n\in\mathbb{N}}$ of random variables on $(\Omega, \mathfrak{A}, P)$.

Definition 1. $(S_n)_{n\in\mathbb{N}}$ with $S_n = \sum_{k=1}^n X_k$ is called the associated *random walk*.

In the sequel we assume

$$X_1 \in \mathfrak{L}^2 \quad \wedge \quad \mathrm{E}(X_1) = 0 \quad \wedge \quad \mathrm{Var}(X_1) = \sigma^2 > 0.$$

Remark 1. For every $\varepsilon > 0$, with probability one,

$$\lim_{n\to\infty} \frac{S_n}{\sqrt{n}\cdot(\log n)^{1/2+\varepsilon}} = 0,$$

see Remark 2.2. On the other hand, with probability one,

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{n}} = \infty \quad \wedge \quad \liminf_{n\to\infty} \frac{S_n}{\sqrt{n}} = -\infty,$$

see Example 5.5.

Question: precise description of the fluctuation of $(S_n(\omega))_{n\in\mathbb{N}}$ for P-almost every ω? In particular: existence of a deterministic sequence $(\gamma(n))_{n\in\mathbb{N}}$ of positive reals such that, with probability one,

$$\limsup_{n\to\infty} \frac{S_n}{\gamma(n)} = 1 \quad \wedge \quad \liminf_{n\to\infty} \frac{S_n}{\gamma(n)} = -1?$$

Notation: $L((u_n)_{n\in\mathbb{N}})$ is the set of all limit points in $\overline{\mathbb{R}}$ of a sequence $(u_n)_{n\in\mathbb{N}}$ in $\mathbb{R}$.

Let

$$\gamma(n) = \sqrt{2n\cdot\log(\log n)\cdot\sigma^2}, \qquad n \geq 3,$$

where log denotes the logarithm with base e.

Theorem 1 (Strassen's Law of the Iterated Logarithm).
With probability one,

$$L\left(\left(\frac{S_n}{\gamma(n)}\right)_{n\in\mathbb{N}}\right) = [-1,\, 1].$$

Proof. See Bauer (1996, §33). □

Corollary 1 (Hartman and Wintner's Law of the Iterated Logarithm).
With probability one,

$$\limsup_{n\to\infty} \frac{S_n}{\gamma(n)} = 1 \quad \wedge \quad \liminf_{n\to\infty} \frac{S_n}{\gamma(n)} = -1.$$

Literature

H. Bauer, *Probability Theory*, de Gruyter, Berlin, 1996.

P. Billingsley, *Probability and Measure*, Wiley, New York, first edition 1979, third edition 1995.

Y. S. Chow, H. Teicher, *Probability Theory*, Springer, New York, first editon 1978, third edition 1997.

R. M. Dudley, *Real Analysis and Probability*, Cambridge University Press, Cambridge, 2002.

J. Elstrodt, *Maß- und Integrationstheorie*, Springer, Berlin, first edition 1996, fifth edition, 2007.

K. Floret, *Maß- und Integrationstheorie*, Teubner, Stuttgart, 1981.

P. Gänssler, W. Stute, *Wahrscheinlichkeitstheorie*, Springer, Berlin, 1977.

E. Hewitt, K. Stromberg, *Real and Abstract Analysis*, Springer, Berlin, 1965.

A. Irle, *Finanzmathematik*, Teubner, Stuttgart, 1998.

A. Klenke, *Wahrscheinlichkeitstheorie*, Springer, Berlin, first edition 2006, second edition 2008.

K. R. Parthasarathy, *Probability Measures on Metric Spaces*, Academic Press, New York, 1967.

A. N. Širjaev, *Wahrscheinlichkeit*, Deutscher Verlag der Wissenschaften, Berlin, 1988.
A. N. Shiryayev, *Probability*, Springer, New York, 1984.

J. Yeh, *Martingales and Stochastic Analysis*, World Scientific, Singapore, 1995.

Index