2 Strong Law of Large Numbers

Definition 1. $(X_n)_{n\in\mathbb{N}}$ *independent and identically distributed (i.i.d.)* iff $(X_n)_{n\in\mathbb{N}}$ is independent and $\forall\, n,k\ :\ X_n \overset{d}{=} X_k$.

Throughout this section: $(X_n)_{n\in\mathbb{N}}$ independent (but only i.i.d. if explicitly noted). Consider

$$C = \big\{(S_n)_{n\in\mathbb{N}} \text{ converges in } \mathbb{R}\big\}.$$

By Remark 1.1, $P(C) \in \{0,1\}$.
First we provide sufficient conditions for $P(C) = 1$ to hold.

Theorem 1 (Kolmogorov's inequality). Assume that $X_i \in \mathfrak{L}^2$ and $\operatorname{E} X_i = 0$ for all i. Then

$$P\Big(\Big\{\sup_{1\le k\le n} |S_k| \ge \varepsilon\Big\}\Big) \le \frac{1}{\varepsilon^2}\cdot \operatorname{Var}(S_n).$$

Proof. Let $1 \le k \le n$. We show that

$$\forall\, B \in \sigma(\{X_1,\dots,X_k\}) : \int_B S_k^2\, dP \le \int_B S_n^2\, dP. \tag{1}$$

Let $B \in \sigma(\{X_1,\dots,X_k\})$. We start with $S_n^2 = (S_k + S_n - S_k)^2$, which implies

$$\begin{aligned} \operatorname{E} 1_B S_n^2 &= \operatorname{E} 1_B S_k^2 + 2\operatorname{E}[(1_B S_k)\cdot(S_n - S_k)] + \operatorname{E} 1_B (S_n - S_k)^2 \\ &\ge \operatorname{E} 1_B S_k^2 + 2\operatorname{E}[(1_B S_k)\cdot(S_n - S_k)]\,. \end{aligned}$$

Moreover, it follows easily from Theorem III.5.4 that $1_B S_k$ and $S_n - S_k$ are independent. Hence Theorem III.5.6 yields

$$\operatorname{E}[(1_B S_k)\cdot(S_n - S_k)] = \operatorname{E}(1_B \cdot S_k)\cdot \operatorname{E}(S_n - S_k) = 0,$$

and thereby

$$\operatorname{E}(1_B \cdot S_n^2) \ge \operatorname{E}(1_B \cdot S_k^2).$$

This completes the proof of (1). For $k \le n$, define

$$A_k = \Big\{|S_\ell| < \varepsilon\,, \forall\, l < k \wedge |S_k| \ge \varepsilon\Big\}.$$

Then $A_k \in \sigma(\{X_1,\dots,X_k\})$, the A_k are disjoint and $\sup_{k\le n}|S_k| > \varepsilon$ iff one A_k happens; hence with the help of (1) we have

$$\begin{aligned} \varepsilon^2 \cdot P\Big(\Big\{\sup_{1\le k\le n}|S_k| \ge \varepsilon\Big\}\Big) &= \varepsilon^2 \cdot \sum_{k=1}^n P(A_k) \le \sum_{k=1}^n \int_{A_k} S_k^2\, dP \\ &\le \sum_{k=1}^n \int_{A_k} S_n^2\, dP \ \le\ \int_\Omega S_n^2\, dP \\ &= \operatorname{Var}(S_n). \end{aligned}$$

∎

Theorem 2. If $X_n \in \mathfrak{L}^2$ and $\mathrm{E}(X_n) = 0$ for all n, and

$$\sum_{i=1}^{\infty} \mathrm{Var}(X_i) < \infty,$$

then S_n converges a.s..

Proof. S_n converges iff it is Cauchy; hence, for

$$M := \inf_{n\in\mathbb{N}} \sup_{k\in\mathbb{N}} |S_{n+k} - S_n| \ ,$$

S_n converges iff $M = 0$. Fix $n \in \mathbb{N}$. Then $M > \varepsilon$ implies that for one $r \in \mathbb{N}$ we have $\sup_{1\le k\le r} |S_{n+k} - S_n| > \varepsilon$. Hence,

$$P(\{M > \varepsilon\}) \le \sup_r P\Big(\Big\{ \sup_{1\le k\le r} |S_{n+k} - S_n| > \varepsilon\Big\}\Big),$$

and Kolmogorov's inequality yields

$$P\Big(\Big\{ \sup_{1\le k\le r} |S_{n+k} - S_n| > \varepsilon\Big\}\Big) \le \frac{1}{\varepsilon^2} \cdot \sum_{i=n+1}^{n+r} \mathrm{Var}(X_i) \le \frac{1}{\varepsilon^2} \cdot \sum_{i=n+1}^{\infty} \mathrm{Var}(X_i).$$

Since n was arbitrary, we get $P(\{M > \varepsilon\}) = 0$ for every $\varepsilon > 0$, i.e., $M = 0$ a.s.. □

Example 1. Let $(Y_n)_{n\in\mathbb{N}}$ be i.i.d. with $\mathrm{E}\, Y_n = 0$, $\mathrm{E}\, Y_n^2 < \infty$, and let b_n such that $1/b_n^2$ is summable. Then

$$\sum_n \mathrm{Var}(Y_n/b_n) < \infty \ ,$$

hence $\sum_n Y_n/b_n$ converges.

In the sequel, $0 < a_n \uparrow \infty$. We now study convergence almost surely of $(S_n/a_n)_{n\in\mathbb{N}}$.

Lemma 1 (Kronecker's Lemma). Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$. Then if $\sum_{i=1}^{\infty} \frac{x_i}{a_i}$ converges, $\frac{1}{a_n} \cdot \sum_{i=1}^{n} x_i \to 0$.

Proof. Consider $\mathbb{N}$ with the counting measure γ, and define

$$f_n(i) := \frac{x_i}{a_i} \cdot \frac{a_i}{a_n} \cdot \mathbf{1}_{i\le n} \ .$$

Then $f_n \to 0$ pointwise, and since a_n is monotone, $|f_n(i)| \le \frac{x_i}{a_i}$, which is γ–integrable by assumption. Hence, by Lebesgue's theorem,

$$\frac{1}{a_n} \cdot \sum_{i\le n} x_i \ = \ \int_{\mathbb{N}} f_n \, d\gamma \ \to \ 0 \ .$$

□

Theorem 3 (Strong Law of Large Numbers, $\mathfrak{L}^2$ Case). If $X_n \in \mathfrak{L}^2$ for all n, and

$$\forall\, n \in \mathbb{N}: \; X_n \in \mathfrak{L}^2 \quad \wedge \quad \sum_{i=1}^{\infty} \frac{1}{a_i^2} \cdot \mathrm{Var}(X_i) < \infty \tag{2}$$

then

$$\frac{1}{a_n} \cdot \sum_{i=1}^{n} (X_i - \mathrm{E}(X_i)) \stackrel{P\text{-a.s.}}{\longrightarrow} 0.$$

Proof. Put $Y_n = 1/a_n \cdot (X_n - \mathrm{E}(X_n))$. Then $\mathrm{E}(Y_n) = 0$ and $(Y_n)_{n\in\mathbb{N}}$ is independent. Moreover,

$$\sum_{i=1}^{\infty} \mathrm{Var}(Y_i) = \sum_{i=1}^{\infty} \frac{1}{a_i^2} \cdot \mathrm{Var}(X_i) < \infty.$$

Thus $\sum_{i=1}^{\infty} Y_i$ converges P-a.s. due to Theorem 2. Apply Lemma 1. □

Remark 1.

1. Assume that the variances $\mathrm{Var}(X_n)$ are bounded and that $\varepsilon > 0$. Then it follows (with $a_n = n^{1/2}(\log n)^{1/2+\varepsilon}$) in particular that

$$n^{-1/2}(\log n)^{-1/2-\varepsilon} \cdot \Big[\sum_{i\le n} X_i - \mathrm{E}\sum_{i\le n} X_i\Big] \stackrel{P\text{-a.s.}}{\longrightarrow} 0\ .$$

   This means that for the 'cumulative effect' $\sum_{i\le n} X_i$ the deviation from mean 'typically' grows slower than $n^{1/2}(\log n)^{1/2+\varepsilon}$. (This will be refined by the CLT.) The independence of the X_n is of course crucial for this; if $X_1 = X_2 = \cdots$, we have a growth rate of n.

2. If additionally X_n is an i.i.d. sequence with $X_1 \in \mathfrak{L}^2$, we may choose $a_n = n$ and derive that

$$\frac{1}{n} \cdot \sum_{i=1}^{n} X_i \stackrel{P\text{-a.s.}}{\longrightarrow} \mathrm{E}(X_1).$$

   In fact, this conclusion already holds if $X_1 \in \mathfrak{L}^1$, see Theorem 4 below.

Example 2. Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d. with $P_{X_1} = p \cdot \delta_1 + (1-p) \cdot \delta_{-1}$. Due to the Strong Law of Large Numbers

$$\frac{1}{n} \cdot S_n \stackrel{P\text{-a.s.}}{\longrightarrow} 2p - 1.$$

Moreover, if $p = 1/2$, for every $\varepsilon > 0$

$$\frac{1}{\sqrt{n} \cdot (\log n)^{1/2+\varepsilon}} \cdot S_n \stackrel{P\text{-a.s.}}{\longrightarrow} 0.$$

Precise description of the fluctuation of $S_n(\omega)$ for P-a.e. $\omega \in \Omega$: Law of the Iterated Logarithm.

Lemma 2. Let $U_i,\ V_i,\ W \in \mathfrak{Z}(\Omega, \mathfrak{A})$ such that

$$\sum_{i=1}^{\infty} P(\{U_i \neq V_i\}) < \infty.$$

Then

$$\frac{1}{n} \cdot \sum_{i=1}^{n} U_i \xrightarrow{P\text{-a.s.}} W \quad \Leftrightarrow \quad \frac{1}{n} \cdot \sum_{i=1}^{n} V_i \xrightarrow{P\text{-a.s.}} W.$$

Proof. The Borel-Cantelli Lemma implies $P(\overline{\lim}_{i\to\infty}\{U_i \neq V_i\}) = 0$. □

Lemma 3. For $X \in \mathfrak{Z}_+(\Omega, \mathfrak{A})$

$$\mathrm{E}(X) \leq \sum_{k=0}^{\infty} P(\{X > k\}) \leq \mathrm{E}(X) + 1.$$

(Cf. Corollary II.8.2.)

Proof. We have

$$\mathrm{E}(X) = \sum_{k=1}^{\infty} \int_{\{k-1<X\leq k\}} X \, dP,$$

and therefore

$$\mathrm{E}(X) \leq \sum_{k=1}^{\infty} k \cdot P(\{k-1 < X \leq k\}) = \sum_{k=0}^{\infty} P(\{X > k\})$$

as well as

$$\mathrm{E}(X) \geq \sum_{k=1}^{\infty} (k-1) \cdot P(\{k-1 < X \leq k\}) \geq \sum_{k=0}^{\infty} P(\{X > k\}) - 1.$$

□

Theorem 4 (Strong Law of Large Numbers, i.i.d. Case). Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d. Then

$$\exists\, Z \in \mathfrak{Z}(\Omega, \mathfrak{A}) : \; \frac{1}{n} \cdot S_n \xrightarrow{P\text{-a.s.}} Z \quad \Leftrightarrow \quad X_1 \in \mathfrak{L}^1,$$

in which case $Z = \mathrm{E}(X_1)$ P-a.s.

Proof. '$\Rightarrow$': From the assumption we derive

$$\frac{1}{n} \cdot X_n = \frac{1}{n} \cdot S_n - \frac{n-1}{n} \cdot \frac{1}{n-1} \cdot S_{n-1} \xrightarrow{P\text{-a.s.}} 0.$$

Hence, for the independent events $A_n = \{|X_n| > n\}$ we have

$$P(\overline{\lim_{n\to\infty}} A_n) = 0\ .$$

The Borel-Cantelli Lemma implies

$$\sum_{n=1}^{\infty} \underbrace{P(A_n)}_{=P(|X_1|>n)} < \infty.$$

Use Lemma 3 to obtain $\mathrm{E}(|X_1|) < \infty$.

'⇐': Consider the truncated random variables

$$Y_n = \begin{cases} X_n & \text{if } |X_n| < n \\ 0 & \text{otherwise.} \end{cases}$$

We will first show that

$$\sum_{i=1}^{\infty} \frac{1}{i^2} \cdot \mathrm{Var}(Y_i) < \infty. \tag{3}$$

To this end, observe that

$$\begin{aligned} \mathrm{Var}(Y_i) \le \mathrm{E}(Y_i^2) &= \sum_{k=1}^{i} \mathrm{E}[Y_i^2 \cdot 1_{[k-1,k[}(|Y_i|)] \\ &= \sum_{k=1}^{i} \mathrm{E}[X_i^2 \cdot 1_{[k-1,k[}(|X_i|)] \\ &\le \sum_{k=1}^{i} k^2 \cdot P(\{k-1 \le |X_1| < k\}). \end{aligned}$$

Thus

$$\begin{aligned} \sum_{i=1}^{\infty} \frac{1}{i^2} \cdot \mathrm{Var}(Y_i) &\le \sum_{k=1}^{\infty} k^2 \cdot P(\{k-1 \le |X_1| < k\}) \cdot \sum_{i=k}^{\infty} \frac{1}{i^2} \\ &\le 2 \cdot \sum_{k=1}^{\infty} k \cdot P(\{k-1 \le |X_1| < k\}) \\ &\le 2 \cdot (\mathrm{E}(|X_1|) + 1) < \infty, \end{aligned}$$

cf. the proof of Lemma 3. (3) follows. Theorem 3 now asserts that

$$\frac{1}{n} \cdot \sum_{i=1}^{n} (Y_i - \mathrm{E}(Y_i)) \xrightarrow{P\text{-a.s.}} 0 \, .$$

Furthermore, Y_n is easily seen to be uniformly integrable, and thus

$$\lim_{n \to \infty} \mathrm{E}(Y_n) = \mathrm{E}(X_1) \, . \tag{4}$$

Due to (4),

$$\frac{1}{n} \cdot \sum_{i=1}^{n} Y_i \xrightarrow{P\text{-a.s.}} \mathrm{E}(X_1) \, .$$

Moreover,

$$\sum_{i=1}^{\infty} P(\{X_i \ne Y_i\}) < \infty, \tag{5}$$

since, by Lemma 3,

$$\sum_{i=1}^{\infty} P(\{X_i \ne Y_i\}) = \sum_{i=1}^{\infty} P(\{|X_i| \ge i\}) \le \sum_{i=0}^{\infty} P(\{|X_1| > i\}) \le \mathrm{E}(|X_1|) + 1 < \infty.$$

Finally, by Lemma 2 and (5)

$$\frac{1}{n} \cdot \sum_{i=1}^{n} X_i \overset{P\text{-a.s.}}{\longrightarrow} \mathrm{E}(X_1) \ .$$

$\square$

What happens if X_n is not integrable?

Theorem 5. Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d..

(i) If $\mathrm{E}(X_1^-) < \infty \wedge \mathrm{E}(X_1^+) = \infty$ then

$$\frac{1}{n} \cdot S_n \overset{P\text{-a.s.}}{\longrightarrow} \infty.$$

(ii) If $\mathrm{E}(|X_1|) = \infty$ then

$$\overline{\lim_{n\to\infty}} \left| \frac{1}{n} \cdot S_n \right| = \infty \ P\text{-a.s.}$$

Proof. (i) follows from Theorem 4, and (ii) is an application of the Borel-Cantelli Lemma, see Gänssler, Stute (1977, p. 131). $\square$

Remark 2. Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d. with $\mu = P_{X_1}$ and corresponding distribution function $F = F_{X_1}$. Suppose that μ is unknown, but observations $X_1(\omega), \ldots, X_n(\omega)$ are available for 'estimation of μ'.
Fix $x \in \mathfrak{R}$. Due to Theorem 4, we have

$$F_n(x, \omega) := \frac{\#\{i \le n \ : \ X_i(\omega) \le x\}}{n} \overset{P\text{-a.s.}}{\longrightarrow} F(x).$$

$F_n(x, \omega)$ is called the *empirical distribution function* $F_n(\cdot, \omega)$; analogously, one can define the *empirical distribution*

$$\mu_n(A, \omega) := \frac{\#\{i \le n \ : \ X_i(\omega) \in A\}}{n} \ .$$

To be precise, we know about the empirical distribution function that

$$\forall\, x \in \mathbb{R} \ \exists\, A \in \mathfrak{A} : \ P(A) = 1 \wedge \Big(\forall\, \omega \in A : \ \lim_{n\to\infty} F_n(x, \omega) = F(x)\Big).$$

Therefore

$$\exists\, A \in \mathfrak{A} : \ P(A) = 1 \wedge \Big(\forall\, q \in \mathbb{Q} \ \forall\, \omega \in A : \ \lim_{n\to\infty} F_n(q, \omega) = F(q)\Big),$$

which easily implies

$$\exists\, A \in \mathfrak{A} : \ P(A) = 1 \wedge \Big(\forall\, \omega \in A : \ \mu_n(\cdot, \omega) \overset{\mathrm{w}}{\longrightarrow} \mu\Big),$$

see p. 63, and Theorem III.3.2. This result can be strengthened to the *Glivenko-Cantelli Theorem*

$$\exists\, A \in \mathfrak{A} : \ P(A) = 1 \wedge \Big(\forall\, \omega \in A : \ \lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x, \omega) - F(x)| = 0\Big),$$

see Billingsley (1979, Theorem 20.6). (From Übung9.2, this result immediately follows for continuous F.)