Probability Theory II

Stefan Geiss

October 3, 2007

Contents

Introduction

This script is a continuation of [1] so that we shall assume that the reader is familiar with the basics from this previous course.
Let us motivate this course by some examples.

Example 1. Assume that we perform some experiment several times under identical conditions and get, each time, a measurement denoted by $f_1, f_2, f_3, ...$ To get the *true* quantity (whatever this means) we naturally consider

$$\frac{1}{n}(f_1 + \cdots + f_n)$$

for large n and hope that

$$\frac{1}{n}(f_1 + \cdots + f_n) \to_n \text{ true value.}$$

In probability we model $f_1, f_2, f_3, ...$ as *independent* random variables.

Now one has to ask: Does the sum $\frac{1}{n}(f_1 + \cdots + f_n)$ converge and, if yes, in what sense. This yields us to the *almost sure convergence* and the *Strong Law of Large Numbers.*

Example 2. We consider a random walk

$$S_n := \varepsilon_1 + \cdots + \varepsilon_n)$$

where $\varepsilon_1, \varepsilon_2, \varepsilon_3, ...$ are independent and

$$\text{probability}(\varepsilon_k = 1) = \text{ probability}(\varepsilon_k = -1) = \frac{1}{2}.$$

Now we make the time steps smaller and smaller and hope that the rescaled random walk converges to a mathematical object that can be identified.

What is the rescaling factor in the state variable that we get an interesting meaningful random object. Or ion other words: How do we have to choose the constants $0 < c_1 \leq c_2 \leq c_3 \leq \cdots$ such that

$$\frac{\varepsilon_1 + \cdots + \varepsilon_n}{c_n}$$

converges to a random variable? This will lead us to the *weak convergence* and to the *Central Limit Theorem.*

Example 3. We assume a source which sends out particles to one side with a uniformly distributed (random) angle which ranges from zero to 180 degree. There is wall in distance of 1 meter. The function $f(\theta)$ gives the position of the particle which hits the wall where $\theta = \pi/2$ gives position 0. The function f is random because it depends on the random angle θ. Consider a second wall parallel to the first one, but with a distance of 2 meters. Now we think in two different ways: firstly, the particle sent out hits the first wall and and will be resent out to the second wall. Secondly, we do the experiment in one step and wait until the particle hits the second wall. Is it possible that both experiments give the same distribution? The answer is *yes*. Knowing this we can analyze the distribution in an abstract manner: Namely, we obtain that the *distributions* of

$$f_1 + f_2 \quad \text{and} \quad 2f$$

are the same, where f_1 and f_2 are *independent copies* of f. The property of this distribution is to be 1-stable, the distribution is called CAUCHY distribution. [1] Stable distributions are used (for example) in stochastic modeling and in probabilistic methods in functional analysis. We shall prove the existence of stable distributions by the help of *characteristic functions*.

[1] Augustin Louis Cauchy, 21 Aug 1789 (Paris) - 23 May 1857 (Sceaux), real and complex analysis.

Chapter 1

Modes of convergence

1.1 Almost sure convergence

Definition 1.1.1. Let f_n, $f : \Omega \to \mathbb{R}$ be random variables where $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. We say that f_n converges almost surely to f ($f_n \xrightarrow[a.s.]{} f$) if

$$\mathbb{P}(\{\omega \in \Omega : |f_n(\omega) - f(\omega)| \xrightarrow[n]{} 0\}) = 1$$

Remark 1.1.2. (i) To formulate the above definition we need $\{\omega : |f_n(\omega) - f(\omega)| \xrightarrow[n]{} 0\}) \in \mathcal{F}$. This follows from

$$\begin{aligned} & \{\omega \in \Omega : |f_n(\omega) - f(\omega)| \xrightarrow[n]{} 0\} \\ = & \left\{\omega : \forall\, m \geq 1 \;\exists\, k \geq 1 \;s.t.\; \forall\, n \geq k \;\; |f_n(\omega) - f(\omega)| < \frac{1}{m}\right\} \\ = & \bigcap_{m=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} \left\{\omega : |f_n(\omega) - f(\omega)| < \frac{1}{m}\right\} \in \mathcal{F}. \end{aligned}$$

(ii) The above definition depends on the measure $\mathbb{P}$. In general one does not have that

$$\mathbb{P}(\{\omega : |f_n(\omega) - f(\omega)| \xrightarrow[n]{} 0\}) = 1$$

iff

$$Q(\{\omega : |f_n(\omega) - f(\omega)| \xrightarrow[n]{} 0\}) = 1$$

if Q is another measure on $\mathcal{F}$.

(iii) Only few properties of f_n are transfered to f by almost sure convergence. Take, for example, $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, and λ to be the Lebesgue measure ($\lambda([a,b]) = b - a$). Let f_n be

$$f_n(\omega) := \begin{cases} n^2 2^{n+1}\omega, & \omega \in \left[0, \frac{1}{2n}\right] \\ n2^{n+1} - n^2 2^{n+1}\omega, & \omega \in \left(\frac{1}{2n}, \frac{1}{n}\right] \\ 0, & \omega \in \left(\frac{1}{n}, 1\right]. \end{cases}$$

The function $f_n : [0,1] \to \mathbb{R}$ is continuous so that f_n is a random variable. Moreover $\lim_n f_n(\omega) = 0$ for all $\omega \in [0,1]$. On the other side

$$\int_0^1 f_n(\omega)\, d\lambda(\omega) = \int_0^1 f_n(t)\, dt = 2^{n-1} \xrightarrow[n]{} \infty.$$

A useful characterization of the almost sure convergence is given by

Proposition 1.1.3. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $f_n, f : \Omega \to \mathbb{R}$ be random variables. Then the following assertions are equivalent.*

(i) $f_n \xrightarrow[a.s.]{} f$.

(ii) $\lim_n \mathbb{P}(\{\omega : \sup_{k \geq n} |f_k(\omega) - f(\omega)| > \varepsilon\}) = 0$ *for all* $\varepsilon > 0$.

Proof. For $\varepsilon > 0$ and $n \geq 1$ define

$$A_n^\varepsilon = \{\omega : \sup_{k\geq n} |f_k(\omega) - f(\omega)| > \varepsilon\} = \bigcup_{k=n}^\infty \{\omega : |f_k(\omega) - f(\omega)| > \varepsilon\} \in \mathcal{F}$$

so that $A_1^\varepsilon \supseteq A_2^\varepsilon \supseteq \ldots$,

$$\lim_{n\to\infty} \mathbb{P}(A_n^\varepsilon) = \mathbb{P}\left(\bigcap_{k=1}^\infty A_n^\varepsilon\right)$$

and

$$\bigcup_{n=1}^\infty A_n^\varepsilon = \{\omega : \forall\ n = 1,2,\ldots \ \sup_{k\geq n} |f_k(\omega) - f(\omega)| > \varepsilon\}.$$

(i)$\Rightarrow$(ii) Let $\Omega_0 := \{\omega : \lim_n f_n(\omega) = f(\omega)\} \in \mathcal{F}$. Hence for all $\omega \in \Omega_0$ there exists $n(\omega) \geq 1$ with $\sup_{k\geq n(\omega)} |f_k(\omega) - f(\omega)| \leq \varepsilon$ so that

$$\Omega_0 \subseteq \left(\bigcup_{n=1}^\infty A_n^\varepsilon\right)^{\mathrm{C}} \quad \text{and} \quad \bigcup_{n=1}^\infty A_n^\varepsilon \subseteq \Omega_0^{\mathrm{C}}.$$

Hence

$$0 = \mathbb{P}\left(\bigcup_{n=1}^\infty A_n^\varepsilon\right) = \lim_n \mathbb{P}\left(A_n^\varepsilon\right).$$

(ii)$\Rightarrow$(i) We have

$$\mathbb{P}\left(\bigcup_{n=1}^\infty (A_n^\varepsilon)^{\mathrm{C}}\right) = 1 \quad \text{and} \quad \mathbb{P}\left(\bigcap_{N=1}^\infty \bigcup_{n=1}^\infty \left(A_n^{\frac{1}{N}}\right)^{\mathrm{C}}\right) = 1\,.$$

Finally,

$$\omega \in \bigcap_{N=1}^\infty \bigcup_{n=1}^\infty \left(A_n^{\frac{1}{N}}\right)^{\mathrm{C}}$$

if and only if for all $N = 1,2,\ldots$ there exists $n \geq 1,2,\ldots$ such that

$$\sup_{k\geq n} |f_k(\omega) - f(\omega)| \leq \frac{1}{N}.$$

$\square$

1.2 Convergence in probability

Although we saw in Remark 1.1.2(3) that a.s. convergence may be a weak notation, this notation is still sometimes too strong.

Example 1.2.1. $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, λ is the Lebesgue measure on $\mathcal{F}$. Define

$$\begin{aligned} f_1(\omega) &= \chi_{[0,\frac{1}{2})}(\omega)\,,\ f_2(\omega) = \chi_{[\frac{1}{2},1)}(\omega)\,, \\ f_3(\omega) &= \chi_{[0,\frac{1}{4})}(\omega)\,,\ f_4(\omega) = \chi_{[\frac{1}{4},\frac{1}{2})}(\omega)\,, \ldots,\ f_6(\omega) = \chi_{[\frac{3}{4},1)}(\omega)\,, \\ f_7(\omega) &= \chi_{[0,\frac{1}{8})}(\omega)\,,\ \ldots \end{aligned}$$

We have the feeling that $\lim_n f_n(\omega) = 0$, but in what sense? We do not have a.s. convergence since $\#\{n : f_n(\omega) = 1\} = \infty$ for all $\omega \in [0,1)$.

The way out is

Definition 1.2.2. Let $f_n\,, f : \Omega \to \mathbb{R}^d$ be random variables, where $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. Then f_n converges to f in probability ($f_n \xrightarrow[\mathbb{P}]{} f$) if

$$\lim_n \mathbb{P}(\{\omega : |f_n(\omega) - f(\omega)| > \varepsilon\}) = 0$$

for all $\varepsilon > 0$.

Example (Example 1.2.1 continued)**.** We have that $f_n \xrightarrow[\lambda]{} 0$. In fact,

$$\lim_n \lambda(\{\omega \in [0,1] : |f_n(\omega)| > \varepsilon\}) \leq \lim_n \lambda(\{\omega \in [0,1] : f_n(\omega) \neq 0\}) = 0$$

since

$$\lambda(\{\omega \in [0,1] : f_n(\omega) \neq 0\}) = \begin{cases} \frac{1}{2} & \text{for } n = 1,2 \\ \frac{1}{4} & \text{for } n = 3,4,5,6 \\ \frac{1}{8} & \text{for } n = 7,\ldots \\ \vdots & \end{cases} .$$

Proposition 1.2.3. *Let $f_n, f : \Omega \to \mathbb{R}^d$ be random variables and $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Then one has the following:*

(i) *If $f_n \xrightarrow[a.s.]{} f$, then $f_n \xrightarrow[\mathbb{P}]{} f$.*

(ii) *If $f_n \xrightarrow[\mathbb{P}]{} f$, then there exists $n_1 < n_2 < n_3 < \ldots$ such that $f_{n_k} \xrightarrow[a.s.]{} f$, as $k \to \infty$.*

Proof. (i) Follows from Proposition 1.1.3 and

$$\mathbb{P}(\{\omega : |f_n(\omega) - f(\omega)| > \varepsilon\}) \le \mathbb{P}(\{\omega : \sup_{k \ge n} |f_k(\omega) - f(\omega)| > \varepsilon\}).$$

(ii) Let n_1 be the smallest $n \ge 1$ such that for all $k \ge n$ one has

$$\mathbb{P}(\{\omega : |f_k(\omega) - f(\omega)| > \frac{1}{2}\}) < \frac{1}{2}.$$

Let n_2 be the smallest $n > n_1$ such that for all $k \ge n$ one has

$$\mathbb{P}(\{\omega : |f_k(\omega) - f(\omega)| > \frac{1}{2^2}\}) < \frac{1}{2^2}.$$

Continuing the same way we get

$$\mathbb{P}(\{\omega : |f_k(\omega) - f(\omega)| > \frac{1}{2^\ell}\}) < \frac{1}{2^\ell}.$$

for $k \ge n_\ell$ and $1 \le n_1 < n_2 < n_3 < \ldots$. Letting

$$A_\ell := \{\omega : |f_{n_\ell}(\omega) - f(\omega)| > \frac{1}{2^\ell}\})$$

we get

$$\mathbb{P}(A_\ell) < \frac{1}{2^\ell} \quad \text{and} \quad \sum_{\ell=1}^{\infty} \mathbb{P}(A_\ell) < \infty.$$

The Borel-Cantelli-lemma implies that

$$\mathbb{P}(\{\omega : \#\{n : \omega \in A_n\} = \infty\}) = 0.$$

Hence $\mathbb{P}(\{\omega : \#\{n : \omega \in A_n\} < \infty\}) = 1$. For those ω we have that

$$|f_{n_\ell}(\omega) - f(\omega)| \le \frac{1}{2^\ell}$$

for $\ell \ge \ell(\omega)$ which gives $f_{n_\ell} \xrightarrow[a.s.]{} f$. □

Example (Example 1.2.1 continued)**.** *What is a possible subsequence? One can take*

$$f_1 = \chi_{[0\frac{1}{2})},\ f_3 = \chi_{[0\frac{1}{4})},\ f_7 = \chi_{[0\frac{1}{8})}, \ldots.$$

Proposition 1.2.4. *For random variables $f, g : \Omega \to \mathbb{R}$ we let*

$$d(f,g) := \int_\Omega \frac{|f(\omega) - g(\omega)|}{1 + |f(\omega) - g(\omega)|} d\mathbb{P}(\omega),$$

Then the following assertions are equivalent:

(i) $d(f_n, f) \xrightarrow[n]{} 0$.

(ii) $f_n \underset{\mathbb{P}}{\longrightarrow} f$.

Proof. (i) $\Longrightarrow$ (ii) $d(f_n, f) \underset{n}{\longrightarrow} 0$ means

$$\int_\Omega \frac{|f(\omega) - g(\omega)|}{1 + |f(\omega) - g(\omega)|} d\mathbb{P}(\omega) \underset{n}{\longrightarrow} 0,$$

so that by Ĉebyšev's inequality, for $\lambda > 0$,

$$\lambda \mathbb{P}\left(\left\{\omega : \frac{|f_n(\omega) - f(\omega)|}{1 + |f_n(\omega) - f(\omega)|} \geq \lambda\right\}\right) \leq \int_\Omega \frac{|f_n(\omega) - f(\omega)|}{1 + |f_n(\omega) - f(\omega)|} d\mathbb{P}(\omega) \underset{n}{\longrightarrow} 0.$$

Given $\varepsilon > 0$ we find $\lambda(\varepsilon) > 0$ such that if $|x| > \varepsilon$, then $\frac{|x|}{1+|x|} > \lambda(\varepsilon)$. Hence

$$\begin{aligned}
\mathbb{P}(\{\omega : |f_n(\omega) - f(\omega)| > \varepsilon\}) &\leq \mathbb{P}\left(\left\{\omega : \frac{|f_n(\omega) - f(\omega)|}{1 + |f_n(\omega) - f(\omega)|} > \lambda(\varepsilon)\right\}\right) \\
&\leq \frac{1}{\lambda(\varepsilon)} \int_\Omega \frac{|f_n(\omega) - f(\omega)|}{1 + |f_n(\omega) - f(\omega)|} d\mathbb{P}(\omega)
\end{aligned}$$

and $\mathbb{P}(\{\omega : |f_n(\omega) - f(\omega)| > \varepsilon\}) \underset{n}{\longrightarrow} 0$.
(ii) $\Longrightarrow$ (i) For all $\varepsilon > 0$ we have that

$$\begin{aligned}
\int_\Omega \frac{|f_n(\omega) - f(\omega)|}{1 + |f_n(\omega) - f(\omega)|} d\mathbb{P}(\omega) &= \int_{\{|f_n - f| > \varepsilon\}} \frac{|f_n(\omega) - f(\omega)|}{1 + |f_n(\omega) - f(\omega)|} d\mathbb{P}(\omega) \\
&\quad + \int_{\{|f_n - f| \leq \varepsilon\}} \frac{|f_n(\omega) - f(\omega)|}{1 + |f_n(\omega) - f(\omega)|} d\mathbb{P}(\omega) \\
&\leq \mathbb{P}(\{\omega \in \Omega : |f_n(\omega) - f(\omega)| > \varepsilon\}) + \frac{\varepsilon}{1 + \varepsilon},
\end{aligned}$$

since the function $\frac{x}{1+x} = 1 - \frac{1}{1+x}$ is monotone for $x \geq 0$. Given $\theta > 0$, we take $\varepsilon > 0$ such that $\frac{\varepsilon}{1+\varepsilon} \leq \frac{\theta}{2}$ and then $n_0 \geq 1$ such that $\mathbb{P}(\{\omega : |f_n(\omega) - f(\omega)| > \varepsilon\}) \leq \frac{\theta}{2}$ for $n \geq n_0$. Hence $d(f_n, f) \leq \theta$ for $n \geq n_0$. □

Proposition 1.2.5. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be fixed and $f, g, h : \Omega \to \mathbb{R}$ be random variables. Then one has that*

(i) $d(f, g) = 0$ *iff* $\mathbb{P}(f = g) = 1$.

(ii) $d(f, g) = d(g, f)$.

(iii) $d(f, h) \leq d(f, g) + d(g, h)$.

(iv) *If $f_1, f_2, ... : \Omega \to \mathbb{R}$ are random variables such that for all $\varepsilon > 0$ there exists an $n(\varepsilon) \geq 1$ such that $d(f_n, f_m) \leq \varepsilon$ for all $m, n \geq n(\varepsilon)$, i.e. if $(f_n)_{n \geq 1}$ is a Cauchy sequence with respect to d, then there exists a random variable $f : \Omega \to \mathbb{R}$ such that*

$$\lim_n d(f_n, f) = 0.$$

Proof. (i) We have that

$$\begin{aligned} d(f,g)=0 &\Leftrightarrow \int_\Omega \frac{|f(\omega)-g(\omega)|}{1+|f(\omega)-g(\omega)|}d\mathbb{P}(\omega)=0 \\ &\Leftrightarrow \mathbb{P}\left(\left\{\omega : \frac{|f(\omega)-g(\omega)|}{1+|f(\omega)-g(\omega)|}=0\right\}\right)=1 \\ &\Leftrightarrow \mathbb{P}\left(\{\omega : f(\omega)=g(\omega)\}\right)=1 \\ &\Leftrightarrow f=g \text{ a.s.} \end{aligned}$$

(ii) follows by definition.
(iii) Here we get

$$\begin{aligned} d(f,h) &= \int_\Omega \frac{|f(\omega)-h(\omega)|}{1+|f(\omega)-h(\omega)|}d\mathbb{P}(\omega) \\ &\leq \int_\Omega \frac{|f(\omega)-g(\omega)|+|g(\omega)-h(\omega)|}{1+|f(\omega)-g(\omega)|+|g(\omega)-h(\omega)|}d\mathbb{P}(\omega) \\ &\leq \int_\Omega \frac{|f(\omega)-g(\omega)|}{1+|f(\omega)-g(\omega)|}d\mathbb{P}(\omega)+\int_\Omega \frac{|g(\omega)-h(\omega)|}{1+|g(\omega)-h(\omega)|}d\mathbb{P}(\omega) \\ &= d(f,g)+d(g,h). \end{aligned}$$

(iv) For the proof of this statement we need

Definition 1.2.6. A sequence $f_1, f_2, ... : \Omega \to \mathbb{R}$ of random variables is a Cauchy sequence in probability (or fundamental in probability) provided that for all $\varepsilon > 0$ there exists $n(\varepsilon) \geq 1$ such that for all $k, l \geq n(\varepsilon)$ one has that

$$\mathbb{P}(|f_k - f_l| > \varepsilon) \leq \varepsilon.$$

Proposition 1.2.7. *Let $(f_n)_{n=1}^\infty$ by a Cauchy sequence in probability. Then there exists a random variable $f : \Omega \to \mathbb{R}$ such that $f_n \to_{\mathbb{P}} f$.*

Proof. We proceed as in the proof of Proposition 1.2.3 and find $1 \leq n_1 < n_2 < \ldots$ such that

$$\mathbb{P}(\{\omega : |f_k(\omega) - f_l(\omega)| > \frac{1}{2^j}\}) < \frac{1}{2^j}$$

for $k, l \geq n_j$. Taking the sequence $(f_{n_j})_{j=1}^\infty$ we get that

$$\mathbb{P}(\{\omega : |f_{n_{j+1}}(\omega) - f_{n_j}(\omega)| > \frac{1}{2^j}\}) < \frac{1}{2^j}$$

and

$$\sum_{j=1}^\infty \mathbb{P}(\{\omega : |f_{n_{j+1}}(\omega) - f_{n_j}(\omega)| > \frac{1}{2^j}\}) < \infty.$$

Applying the Borel-Cantelli Lemma we get that

$$\mathbb{P}(\{\omega : |f_{n_{j+1}}(\omega) - f_{n_j}(\omega)| > \frac{1}{2^j} \text{ infinitely often}\}) = 0.$$

Hence

$$\mathbb{P}(\{\omega : \sum_{j=1}^{\infty} |f_{n_{j+1}}(\omega) - f_{n_j}(\omega)| < \infty\}) = 1.$$

We set

$$f(\omega) := \begin{cases} f_{n_1}(\omega) + \sum_{j=1}^{\infty}(f_{n_{j+1}}(\omega) - f_{n_j}(\omega)) & : \sum_{j=1}^{\infty}(f_{n_{j+1}} - f_{n_j}) < \infty \\ 0 & : \text{else} \end{cases}$$

and get that $f_{n_j} \xrightarrow[a.s.]{} f$. Finally, we have to check that $f_n \xrightarrow[\mathbb{P}]{} f$. For $\varepsilon > 0$ one gets

$$\begin{aligned} & \mathbb{P}(\{\omega : |f_n(\omega) - f(\omega)| > \varepsilon\}) \\ \leq\ & \mathbb{P}(\{\omega : |f_n(\omega) - f_{n_j}(\omega)| + |f_{n_j}(\omega) - f(\omega)| > \varepsilon\}) \\ \leq\ & \mathbb{P}(\{\omega : |f_n(\omega) - f_{n_j}(\omega)| > \frac{\varepsilon}{2}\}) + \mathbb{P}(\{\omega : |f_{n_j}(\omega) - f(\omega)| > \frac{\varepsilon}{2}\}) \end{aligned}$$

and

$$\lim_{n,j\to\infty} \mathbb{P}(\{\omega : |f_n(\omega) - f_{n_j}(\omega)| > \frac{\varepsilon}{2}\}) = 0,$$

$$\lim_{j\to\infty} \mathbb{P}(\{\omega : |f_{n_j}(\omega) - f(\omega)| > \frac{\varepsilon}{2}\}) = 0.$$

□

Now we can finish our proof by showing that $(f_n)_{n=1}^{\infty}$ is a Cauchy sequence with respect to d if and only if it is a Cauchy sequence in probability.

Assume that $(f_n)_{n=1}^{\infty}$ is a Cauchy sequence with respect to d. For $\lambda > 0$ we have that

$$\lambda \mathbb{P}\left(\left\{\omega : \frac{|f_k(\omega) - f_l(\omega)|}{1 + |f_k(\omega) - f_l(\omega)|} > \lambda\right\}\right) \leq d(f_k, f_l) \leq \eta$$

for $k, l \geq n(\eta) \geq 1$. For $\lambda := \frac{\varepsilon}{1+\varepsilon}$ with $\varepsilon > 0$ this gives that

$$\frac{\varepsilon}{1+\varepsilon} \mathbb{P}\left(\{\omega : |f_k(\omega) - f_l(\omega)| > \varepsilon\}\right) \leq \eta$$

for $k, l \geq n(\eta) \geq 1$. Choosing $\eta = \varepsilon^2$ we end up with

$$\mathbb{P}\left(\{\omega : |f_k(\omega) - f_l(\omega)| > \varepsilon\}\right) \leq \varepsilon(1 + \varepsilon)$$

for $k, l \geq n(\varepsilon^2) \geq 1$.

Now assume that $(f_n)_{n=1}^{\infty}$ is a Cauchy sequence in probability. Then for all $\varepsilon > 0$ there exists $n(\varepsilon) \geq 1$ such that for all $k, l \geq n(\varepsilon)$ $\mathbb{P}(\{\omega : |f_k(\omega) - f_l(\omega)| > \varepsilon\}) \leq \varepsilon$. We get that

$$\begin{aligned}\int_\Omega \frac{|f_k(\omega) - f_l(\omega)|}{1 + |f_k(\omega) - f_l(\omega)} d\mathbb{P}(\omega) &\leq \mathbb{P}(\{\omega : |f_k(\omega) - f_l(\omega)| > \varepsilon\}) + \frac{\varepsilon}{1+\varepsilon} \\ &\leq \varepsilon + \frac{\varepsilon}{1+\varepsilon} = \frac{2\varepsilon + \varepsilon^2}{1+\varepsilon}\end{aligned}$$

for $k, l \geq n(\varepsilon) \geq 1$. □

To formalize the above we recall the notion of a metric space:

Definition 1.2.8. Let $M \neq \emptyset$.

(i) The pair (M, d) is called *metric space,* if $d : M \times M \to [0, \infty)$ satisfies

  (a) $d(x, x) = 0$ (reflexivity),

  (b) $d(x, y) = d(y, x)$ (symmetry),

  (c) $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality).

(ii) A set $G \subseteq M$ is open, if and only if for all $x \in G$ there exists $\varepsilon > 0$ such that $y \in G$ for all y such that $d(x, y) < \varepsilon$.

Convergence in probability can be described by a suitable metric space those elements consists of equivalence classes.

Definition 1.2.9. Let M be an arbitrary set. We say that a relation $x \sim y$ is *an equivalence class relation* on M, provided that

(i) $x \sim x$ for all $x \in M$ (reflexivity),

(ii) if $x \sim y$, then $y \sim x$ (symmetry),

(iii) if $x \sim y$ and $y \sim z$, then $x \sim z$ (transitivity).

For two elements $x, y \in M$ we have either $x \sim y$ or x and y are not in the relation $x \sim y$. Based on this one introduces *equivalence classes:* We have

$$M = \bigcup_{i \in I} M_i,$$

with

(i) $M_i \neq \emptyset$,

(ii) $M_i \cap M_j = \emptyset$, if $i \neq j$,

(iii) if $x, y \in M_i$, then $x \sim y$, and

(iv) if $x \in M_i$ and $y \in M_j$ with $i \neq j$, then x and y are not in the relation $\sim$.

An element $x_i \in M_i$ is called *representative.* In probability theory, we define for random variables $f, g : \Omega \to \mathbb{R}$ the relation $f \sim g$, provided that $\mathbb{P}(\{\omega : f(\omega) = g(\omega)\}) = 1$. We denote by $L_0(\Omega, \mathcal{F}, \mathbb{P})$ the set of all equivalence classes with respect to $\sim$ and let

$$\hat{d}(\hat{f}, \hat{g}) := d(f, g)$$

for $f \in \hat{f}$ and $g \in \hat{g}$. Then we get

Proposition 1.2.10. *The space* $[L_0(\Omega, \mathcal{F}, \mathbb{P}), \hat{d}]$ *is a complete metric space.*

1.3 Convergence in mean

Definition 1.3.1. Let $p \in (0, \infty)$ and let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Given random variables $f_n, f : \Omega \to \mathbb{R}$ we say that f_n converges to f *with respect to the p-th mean* ($f_n \xrightarrow[L_p]{} f$) provided that

$$\lim_{n\to\infty} \int_\Omega |f_n(\omega) - f(\omega)|^p d\mathbb{P}(\omega) = 0.$$

Since f_n and f are random variables and $x \mapsto |x|^p$ is a continuous function, $\omega \mapsto |f_n(\omega) - f(\omega)|^p$ is a non-negative random variable and we may integrate.

Proposition 1.3.2. *If* $f_n \xrightarrow[L_p]{} f$*, then* $f_n \xrightarrow[\mathbb{P}]{} f$.

Proof. Let $\varepsilon > 0$. Then

$$\begin{aligned} \mathbb{P}(\{\omega : |f_n(\omega) - f(\omega)| > \varepsilon\}) &= \mathbb{P}(\{\omega : |f_n(\omega) - f(\omega)|^p > \varepsilon^p\}) \\ &\leq \frac{1}{\varepsilon^p} \int_\Omega |f_n(\omega) - f(\omega)|^p d\mathbb{P}(\omega) \xrightarrow[n]{} 0. \end{aligned}$$

□

Proposition 1.3.3. (Minkowski inequality) *Let* $0 < p < \infty$ *and* $f, g : \Omega \to \mathbb{R}$ *be random variables. Then one has*

$$\begin{aligned} &\left(\int_\Omega |f(\omega) + g(\omega)|^p d\mathbb{P}(\omega) \right)^{\frac{1}{p}} \\ \leq\ & c_p \left[\left(\int_\Omega |f(\omega)|^p d\mathbb{P}(\omega) \right)^{\frac{1}{p}} + \left(\int_\Omega |g(\omega)|^p d\mathbb{P}(\omega) \right)^{\frac{1}{p}} \right], \end{aligned}$$

where $c_p = 1$ *for* $1 \leq p < \infty$ *and* $c_p = 2^{\frac{1}{p}-1}$ *for* $0 < p < 1$.

Proof. We only prove the case $0 < p \le 1$. Here we get $|a+b|^p \le |a|^p + |b|^p$ for all $a, b \in \mathbb{R}$, so that

$$\begin{aligned} & \left(\int_\Omega |f(\omega) + g(\omega)|^p d\mathbb{P}(\omega) \right)^{\frac{1}{p}} \\ \le \;& \left(\int_\Omega |f(\omega)|^p d\mathbb{P}(\omega) + \int_\Omega |g(\omega)|^p d\mathbb{P}(\omega) \right)^{\frac{1}{p}} \\ \le \;& 2^{\frac{1}{p}-1} \left[\left(\int_\Omega |f(\omega)|^p d\mathbb{P}(\omega) \right)^{\frac{1}{p}} + \left(\int_\Omega |g(\omega)|^p d\mathbb{P}(\omega) \right)^{\frac{1}{p}} \right], \end{aligned}$$

where, for $1 \le q = \frac{1}{p} < \infty$, we have used $|a+b|^q \le 2^{q-1}(|a|^q + |b|^q)$. □

From the Minkowski inequality, we get immediately

Proposition 1.3.4. *If $f_n, g_n, f, g : \Omega \to \mathbb{R}$, $f_n \xrightarrow[L_p]{} f$ and $g_n \xrightarrow[L_p]{} g$, then $f_n + g_n \xrightarrow[L_p]{} f + g$.*

Now, we recall

Proposition 1.3.5. (Hölder's inequality) *Let $f, g : \Omega \to \mathbb{R}$ be random variables, $1 < p < \infty$, $1 < q < \infty$ and $1 = \frac{1}{p} + \frac{1}{q}$. Then*

$$\int_\Omega |f(\omega) g(\omega)| d\mathbb{P}(\omega) \le \left(\int_\Omega |f(\omega)|^p d\mathbb{P}(\omega) \right)^{\frac{1}{p}} \left(\int_\Omega |g(\omega)|^q d\mathbb{P}(\omega) \right)^{\frac{1}{q}}.$$

As an application we get

Proposition 1.3.6. *Let $0 < p < q < \infty$. Then from $f_n \xrightarrow[L_q]{} f$ it follows $f_n \xrightarrow[L_p]{} f$.*

Proof. Let $r := \frac{q}{p} \in (1, \infty)$ and let $1 = \frac{1}{r} + \frac{1}{r*}$. Then

$$\begin{aligned} & \int_\Omega |f_n(\omega) - f(\omega)|^p d\mathbb{P}(\omega) \\ = \;& \int_\Omega |f_n(\omega) - f(\omega)|^p \cdot 1 d\mathbb{P}(\omega) \\ \le \;& \left(\int_\Omega \left(|f_n(\omega) - f(\omega)|^p\right)^r d\mathbb{P}(\omega) \right)^{\frac{1}{r}} \left(\int_\Omega 1^{r*} d\mathbb{P}(\omega) \right)^{\frac{1}{r*}} \\ = \;& \left(\int_\Omega |f_n(\omega) - f(\omega)|^q d\mathbb{P}(\omega) \right)^{\frac{p}{q}}. \end{aligned}$$

□

We conclude with the Lorentz-spaces L_p. Before we do this we recall the notation of a Banach space.

Definition 1.3.7. Let X be a linear space and $\| \cdot \| : X \to [0, \infty)$. Then $[X, \| \cdot \|]$ is called Banach space, provided that

(i) $\|x\| = 0$, if and only if $x = 0$,

(ii) $\|\lambda x\| = |\lambda| \|x\|$ for all $\lambda \in \mathbb{R}$ ($\lambda \in \mathbb{C}$) and $x \in X$,

(iii) $\|x + y\| \le \|x\| + \|y\|$ and

(iv) for all $(x_n)_{n=1}^{\infty} \subseteq X$ such that for all $\varepsilon > 0$ there exists $n(\varepsilon) \ge 1$ such that for all $m, n \ge n(\varepsilon)$ $\|x_m - x_n\| \le \varepsilon$ there is an $x \in X$ such that $\lim_n \|x_n - x\|_{L_p} = 0$.

Remark 1.3.8. Properties (i), (ii), and (iii) say that $\| \cdot \|$ is a norm.

Example 1.3.9. For $X = \mathbb{R}^n$ and $\|x\| = \|(x_1, ..., x_n)\| := (x_1^2 + \cdots + x_n^2)^{\frac{1}{n}}$ we obtain a norm on $\mathbb{R}^n$.

In our setting there are some problems with property (i). For this reason we have to introduce equivalence classes.

Definition 1.3.10. For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $f : \Omega \to \mathbb{R}$ we let

$$\hat{f} := \{g : \Omega \to \mathbb{R} \text{ random variable}, \mathbb{P}(\{\omega \in \Omega : f(\omega) = g(\omega)\}) = 1\}.$$

The set $\hat{f}$ is called *equivalence class* (with respect to $\mathbb{P}$) and $f \in \hat{f}$ is called representative.

Proposition 1.3.11. *One has that*

(i) $\hat{f} = \hat{g}$ *or* $\hat{f} \cap \hat{g} = \emptyset$ *for all random variables* $f, g : \Omega \to \mathbb{R}$, *and that*

(ii) $\mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$ *is the disjoint union of the equivalence classes.*

For $f, g \in \mathcal{L}_0(\Omega, \mathcal{F}, \mathbb{P})$ and $\lambda \in \mathbb{R}$ we introduce the linear operations

$$\begin{aligned} \lambda \hat{f} &:= \widehat{\lambda f}, \\ \hat{f} + \hat{g} &:= \widehat{f + g}. \end{aligned}$$

Definition 1.3.12. For $p \in [1, \infty)$ and a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we let

(i) $\mathcal{L}_p := \left\{ f : \Omega \to \mathbb{R} \text{ random variable}, \|f\|_{L_p} := \left(\int_\Omega |f|^p d\mathbb{P} \right)^{\frac{1}{p}} < \infty \right\}$,

(ii) $L_p(\Omega, \mathcal{F}, \mathbb{P}) := \left\{ \hat{f} : f \in \mathcal{L}_p(\Omega, \mathcal{F}, \mathbb{P}) \right\}$ and $\|\hat{f}\|_{L_p} := \|f\|_{L_p}$.

Proposition 1.3.13. $\left[L_p(\Omega,\mathcal{F},\mathbb{P}),\|\cdot\|_{L_p}\right]$ *is a Banach space.*

Proof. (i) $\|\hat{f}\|_{L_p}=0$ iff $f=0$ a.s. iff $\hat{f}=0$.
(ii) is clear and (iii) is a consequence of the Minkowski inequality.
(iv) Assume a Cauchy sequence $(\hat{f}_n)_{n=1}^\infty\subseteq L_p$. Then $(f_n)_{n=1}^\infty$ is a Cauchy sequence in probability since

$$\mathbb{P}\left(\{\omega:|f_n(\omega)-f_m(\omega)|>\lambda\}\right)\leq\frac{1}{\lambda^p}\|\hat{f}_n(\omega)-\hat{f}_m(\omega)\|_{L_p}^p\leq\varepsilon$$

for $n,m\geq n(\lambda,\varepsilon)\geq 1$. Hence there is a limit $f:\Omega\to\mathbb{R}^d$ such that $f_n \xrightarrow[\mathbb{P}]{} f$. It remains to show that $f_n \xrightarrow[L_p]{} f$ as well. We pick a subsequence $(f_{n_k})_{k=1}^\infty$ such that $\lim_k f_{n_k}=f$ $\mathbb{P}-a.s.$. Applying the lemma of Fatou gives

$$\int_\Omega|f_m(\omega)-f(\omega)|^p d\mathbb{P}(\omega)\leq\liminf_k\int_\Omega|f_m(\omega)-f_{n_k}(\omega)|^p d\mathbb{P}(\omega)\leq\varepsilon$$

for $m\geq m(\varepsilon)\geq 1$. □

Remark 1.3.14. (i) One can define $L_p(\Omega,\mathcal{F},\mathbb{P})$ for $0<p<1$ in the same way. We get

(a) $\|x+y\|_{L_p}\leq c_p\left(\|x\|_{L_p}+\|y\|_{L_p}\right)$ (quasi-triangle inequality) and

(b) (L_p,d) is a complete metric space with $d(\hat{f},\hat{g}):=\|\hat{f}-\hat{g}\|_{L_p}^p$.

(ii) One can define $L_p(\Omega,\mathcal{F},\mathbb{P})$ for $p=\infty$ as well, where

$$\|\hat{f}\|_{L_\infty}:=\operatorname*{ess\,sup}_{\omega\in\Omega}|f(\omega)|=\inf\left\{\sup_{\omega\in\Omega\setminus N}|f(\omega)|:N\in\mathcal{F},\mathbb{P}(N)=0\right\}.$$

Again, $[L_\infty(\Omega,\mathcal{F},\mathbb{P}),\|\cdot\|_{L_\infty}]$ is a Banach space.

Proposition 1.3.15. *Assume that* $\lim_n f_n=f$ $\mathbb{P}-a.s.$, $p\in(0,\infty)$ *and* $\mathbb{E}\sup_n|f_n|^p<\infty$. *Then* $\hat{f}\in L_p$ *and* $\lim_n\mathbb{E}|f_n-f|^p=0$.

Proof. By Lebesgue's theorem of dominated convergence we get that

$$\mathbb{E}|f|^p=\lim_n\mathbb{E}|f_n|^p<\infty.$$

Hence

$$\mathbb{E}\sup_n|f_n-f|^p\leq c_p\mathbb{E}\left(\sup_n|f_n|^p+|f|^p\right)<\infty$$

and, again by dominated convergence,

$$0=\mathbb{E}\lim_n|f_n-f|^p=\lim_n\mathbb{E}|f_n-f|^p.$$

□

Summary:

a.s. convergence	$\Longrightarrow$	L_p-convergence if $\sup_n \lvert f_n\rvert^p < \infty$
L_p-convergence	$\Longrightarrow$	a.s. convergence for a subsequence
a.s. convergence	$\Longrightarrow$	convergence in probability
convergence in probability	$\Longrightarrow$	a.s. convergence for a subsequence
L_p-convergence	$\Longrightarrow$	convergence in probability
convergence in probability	$\Longrightarrow$	L_p-convergence if $\sup_n \lvert f_n\rvert^p < \infty$

Bibliography

[1] C. Geiss and S. Geiss. *An introduction to probability theory.*