Section 6

0 - 1 Laws. Convergence of random series.

Consider a sequence \((X_i)_{i \geq 1}\) of real valued independent random variables and let \(\sigma((X_i)_{i \geq 1})\) be a \(\sigma\)-algebra of events generated by this sequence, i.e. \(\{(X_i)_{i \geq 1} \in B\}\) for \(B\) in the cylindrical \(\sigma\)-algebra on \(\mathbb{R}^N\).

**Definition.** An event \(A \in \sigma((X_i)_{i \geq 1})\) is called a tail event if \(A \in \sigma((X_i)_{i \geq n})\) for all \(n \geq 1\).

For example, if \(A_i \in \sigma(X_i)\) then
\[
A_i \text{ i.o.} = \bigcap_{n \geq 1} \bigcup_{i \geq n} A_i
\]
is a tail event. It turns out that such events have probability 0 or 1.

**Theorem 9** *(Kolmogorov’s 0-1 law)* If \(A\) is a tail event then \(\mathbb{P}(A) = 0\) or \(1\).

**Proof.** For a finite subset \(F = \{i_1, \ldots, i_n\} \subset \mathbb{N}\), let us denote by \(X_F = (X_{i_1}, \ldots, X_{i_n})\). A \(\sigma\)-algebra \(\sigma((X_i)_{i \geq 1})\) is generated by algebra
\[
\{X_F \in B : F- \text{ finite } \subseteq \mathbb{N}, B \in \mathcal{B}(\mathbb{R}^{|F|})\}.
\]
By approximation lemma, we can approximate any event \(A \in \sigma((X_i)_{i \geq 1})\) by events in this generating algebra. Therefore, for any \(\varepsilon > 0\) there exists a set \(A'\) in this algebra such that \(\mathbb{P}(A \Delta A') \leq \varepsilon\) and by definition \(A' \in \sigma(X_1, \ldots, X_n)\) for large enough \(n\). This implies
\[
|\mathbb{P}(A) - \mathbb{P}(A')| \leq \varepsilon, \quad |\mathbb{P}(A) - \mathbb{P}(AA')| \leq \varepsilon.
\]
Since \(A\) is a tail event, \(A \in \sigma((X_i)_{i \geq n+1})\) which means that \(A, A'\) are independent, i.e. \(\mathbb{P}(AA') = \mathbb{P}(A)\mathbb{P}(A')\).

We get
\[
\mathbb{P}(A) \approx \mathbb{P}(AA') = \mathbb{P}(A)\mathbb{P}(A') \approx \mathbb{P}(A)^2
\]
and letting \(\varepsilon \to 0\) proves that \(\mathbb{P}(A) = \mathbb{P}(A)^2\).

**Examples.**
1. \(\{\sum_{i \geq 1} X_i \text{ converges}\}\) is a tail event, it has probability 0 or 1.
2. Consider series \(\sum_{i \geq 1} X_iz^i\) on a complex plane, \(z \in \mathbb{C}\). Its radius of convergence is
\[
r = \liminf_{i \to \infty} |X_i|^{-\frac{1}{4}}.
\]
For any \(x \geq 0\), event \(\{r \leq x\}\) is, obviously, a tail event. This implies that \(r = \text{const} \) with probability 1.

\(\blacksquare\)
The Savage-Hewitt 0 - 1 law.

Next we will prove a stronger result under more restrictive assumption that the r.v.s \( X_i, i \geq 1 \) are not only independent but also identically distributed with the law \( \mu \). Without loss of generality, we can assume that each \( X_i \) is given by the identity \( X_i(x) = x \) on its sample space \((\mathbb{R}, \mathcal{B}, \mu)\). By Kolmogorov’s consistency theorem the entire sequence \((X_i)_{i \geq 1}\) can be defined on the sample space \((\mathbb{R}^\mathbb{N}, \mathcal{B}^\mathbb{N}, \mathbb{P})\) where \( \mathcal{B}^\mathbb{N} \) is the cylindrical \( \sigma \)-algebra and \( \mathbb{P} \) is the measure guaranteed by the Caratheodory extension theorem. In our case \( X_i \)'s are i.i.d. and \( \mathbb{P} = \mu^\mathbb{N} \) is called the infinite product measure. It will be convenient to use the notation \( \sigma((X_i)_{i \geq 1}) \) for the cylindrical \( \sigma \)-algebra since similar notation can be used for the cylindrical \( \sigma \)-algebra on any subset of coordinates.

**Definition.** An event \( A \in \sigma((X_i)_{i \geq 1}) \) is called exchangeable/symmetric if for all \( n \geq 1 \),
\[
(x_1, x_2, \ldots, x_n, x_{n+1}, \ldots) \in A \implies (x_n, x_2, \ldots, x_{n-1}, x_1, x_{n+1}, \ldots) \in A.
\]

In other words, the set \( A \) is symmetric under permutations of a finite number of coordinates. Note that any tail event is symmetric.

**Theorem 10 (Savage-Hewitt 0-1 law)** If \( A \) is symmetric then \( \mathbb{P}(A) = 0 \) or 1.

**Proof.** Given a sequence \( x = (x_1, x_2, \ldots) \) let us define an operator
\[
\Gamma x = (x_{n+1}, x_{2n}, x_1, \ldots, x_n, x_{2n+1}, \ldots)
\]
that switches the first \( n \) coordinates with the second \( n \) coordinates. Since \( A \) is symmetric,
\[
\Gamma A = \{\Gamma x : x \in A\} = A.
\]

By the Approximation Lemma 2 for any \( \varepsilon > 0 \) for large enough \( n \), there exists \( A_n \in \sigma(X_1, \ldots, X_n) \) such that \( \mathbb{P}(A_n \Delta A) \leq \varepsilon \). Clearly,
\[
B_n = \Gamma A_n \in \sigma(X_{n+1}, \ldots, X_{2n})
\]
and
\[
\mathbb{P}(B_n \Delta A) = \mathbb{P}(\Gamma A_n \Delta \Gamma A) \quad \text{by i.i.d.}
\]
\[
\mathbb{P}(A_n \Delta A) \leq \varepsilon,
\]
which implies that \( \mathbb{P}(\{A_n \Delta B_n\} \Delta A) \leq 2\varepsilon \). Therefore, we can conclude that
\[
\mathbb{P}(A) \approx \mathbb{P}(A_n), \quad \mathbb{P}(A) \approx \mathbb{P}(A_n B_n) = \mathbb{P}(A_n) \mathbb{P}(B_n) = \mathbb{P}(A_n)^2
\]
where we used the fact that the events \( A_n, B_n \) are defined in terms of different sets of coordinates and, thus, are independent. Letting \( \varepsilon \to 0 \) implies that \( \mathbb{P}(A) = \mathbb{P}(A)^2 \).

**Example.** Let \( S_n = X_1 + \ldots + X_n \) and let
\[
r = \limsup_{n \to \infty} \frac{S_n - a_n}{b_n}.
\]
Event \( \{r \leq x\} \) is symmetric since changing the order of any finite set of coordinates does not affect \( S_n \) for large enough \( n \). As a result, \( \mathbb{P}(r \leq x) = 0 \) or 1, which implies that \( r = \text{const} \) with probability 1.

**Random series.** We already saw above that, by Kolmogorov’s 0-1 law, the series \( \sum_{i \geq 1} X_i \) for independent \( (X_i)_{i \geq 1} \) converges with probability 0 or 1. This means that either \( S_n = X_1 + \ldots + X_n \) converges to its limit \( S \) with probability one, or with probability one it does not converge. Two section back, before the proof of the strong law of large numbers, we saw the example of a sequence which with probability one does not converge yet converges to 0 in probability. In case when with probability one \( S_n \) does not converge, is it still possible that it converges to some random variable in probability? The answer is no because we will now prove that for random series convergence in probability implies a.s. convergence.
Theorem 11 (Kolmogorov’s inequality) Suppose that \((X_i)_{i \geq 1}\) are independent and \(S_n = X_1 + \ldots + X_n\). If for all \(j \leq n\),
\[ P(|S_n - S_j| \geq a) \leq P(<a) \quad (6.0.1) \]
then for \(x > a\),
\[ P\left( \max_{1 \leq j \leq n} |S_j| \geq x \right) \leq \frac{1}{1 - p} P(|S_n| > x - a). \]

Proof. First of all, let us notice that this inequality is obvious without the maximum because (6.0.1) is equivalent to \(1 - p \leq P(|S_n - S_j| < a)\) and we can write
\[ (1 - p)P(|S_j| \geq x) \leq P(|S_n - S_j| < a)P(|S_j| \geq x) \]
\[ = P(|S_n - S_j| < a, |S_j| \geq x) \leq P(|S_n| > x - a). \]
The equality is true because events \(\{|S_j| \geq x\}\) and \(\{|S_n - S_j| < a\}\) are independent since the first depends only on \(X_1, \ldots, X_j\) and the second only on \(X_{j+1}, \ldots, X_n\). The last inequality is true simply by triangle inequality. To deal with the maximum, instead of looking at an arbitrary partial sum \(S_j\) we will look at the first partial sum that crosses level \(x\). We define that first time by \(\tau = \min\{j \leq n : |S_j| \geq x\}\) and let \(\tau = n + 1\) if all \(|S_j| < x\). Notice that event \(\{\tau = j\}\) also depends only on \(X_1, \ldots, X_j\) so we can again write
\[ (1 - p)P(\tau = j) \leq P(|S_n - S_j| < a)P(\tau = j) \]
\[ = P(|S_n - S_j| < a, \tau = j) \leq P(|S_n| > x - a, \tau = j). \]
The last inequality is again true by triangle inequality because when \(\tau = j\) we have \(|S_j| \geq x\) and
\[ \left\{ |S_n - S_j| < a, \tau = j \right\} \subseteq \left\{ |S_n| > x - a, \tau = j \right\}. \]
It remains to add up over \(j \leq n\) to get
\[ (1 - p)P(\tau \leq n) \leq P(|S_n| > x - a, \tau \leq n) \leq P(|S_n| > x - a) \]
and notice that \(\tau \leq n\) is equivalent to \(\max_{j \leq n} |S_j| \geq x\).

Theorem 12 (Kolmogorov) If the series \(\sum_{i \geq 1} X_i\) converges in probability then it converges almost surely.

Proof. Suppose that partial sums \(S_n\) converge to some r.v. \(S\) in probability, i.e. for any \(\varepsilon > 0\), for large enough \(n \geq n_0(\varepsilon)\) we have \(P(|S_n - S| \geq \varepsilon) \leq \varepsilon\). If \(k \geq j \geq n \geq n_0(\varepsilon)\) then
\[ P(|S_k - S_j| \geq 2\varepsilon) \leq P(|S_k - S| \geq \varepsilon) + P(|S_j - S| \geq \varepsilon) \leq 2\varepsilon. \]
Next, we use Kolmogorov’s inequality for \(x = 4\varepsilon\) and \(a = 2\varepsilon\) (we let partial sums start at \(n\)):
\[ P\left( \max_{n \leq j \leq k} |S_j - S_n| \geq 4\varepsilon \right) \leq \frac{1}{1 - 2\varepsilon} P(|S_k - S_n| \geq 2\varepsilon) \leq \frac{2\varepsilon}{1 - 2\varepsilon} \leq 3\varepsilon, \]
for small \(\varepsilon\). The events \(\{\max_{n \leq j \leq k} |S_j - S_n| \geq 4\varepsilon\}\) are increasing as \(k \uparrow \infty\) and by continuity of measure
\[ P\left( \max_{n \leq j \leq k} |S_j - S_n| \geq 5\varepsilon \right) \leq 3\varepsilon. \]
Finally, since \(P(|S_n - S| \geq \varepsilon) \leq \varepsilon\) we get
\[ P\left( \max_{n \leq j} |S_j - S| \geq 5\varepsilon \right) \leq 4\varepsilon. \]
This kind of "maximal" statement about any sequence $S_j$ is actually equivalent to its a.s. convergence. To see this take $\varepsilon = \frac{1}{m^2}$, take $n(m) = n_0(\varepsilon)$ and consider an event

$$A_m = \left\{ \max_{n(m) \leq j} |S_j - S| \geq \frac{5}{m^2} \right\}.$$ 

We proved that

$$\sum \mathbb{P}(A_m) \leq \sum \frac{4}{m^2} < \infty$$

and by the Borel-Cantelli lemma, $\mathbb{P}(A_m \text{ i.o.}) = 0$. This means that with probability 1 for large enough (random) $m$,

$$\max_{j \geq n(m)} |S_j - S| \leq \frac{5}{m^2}$$

and, therefore, $S_n \to S$ a.s.

Let us give one easy-to-check criterion for convergence of random series, which is called Kolmogorov’s strong law of large numbers.

**Theorem 13** If $(X_i)_{i \geq 1}$ is a sequence of independent random variables such that $\mathbb{E}X_i = 0$ and $\sum_{i \geq 1} \mathbb{E}X_i^2 < \infty$ then $\sum_{i \geq 1} X_i$ converges a.s.

**Proof.** It is enough to prove convergence in probability. Let us first show the existence of a limit of partial sums $S_n$ over some subsequence. For $m < n$,

$$\mathbb{P}(|S_n - S_m| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \mathbb{E}(S_n - S_m)^2 = \frac{1}{\varepsilon^2} \sum_{m < i < n} \mathbb{E}X_i^2 \xrightarrow{m \to \infty} 0.$$ 

If we take $\varepsilon = \frac{1}{l^2}$ then for large enough $m(l)$ and for any $n \geq m(l)$,

$$\mathbb{P} \left( |S_n - S_m(l)| \geq \frac{1}{l^2} \right) \leq \frac{1}{l^2}. \quad (6.0.2)$$

W.l.o.g. we can assume that $m(l + 1) \geq m(l)$ so that

$$\mathbb{P} \left( |S_{m(l+1)} - S_m(l)| \geq \frac{1}{l^2} \right) \leq \frac{1}{l^2}.$$ 

Then,

$$\sum_{l \geq 1} \mathbb{P} \left( |S_{m(l+1)} - S_m(l)| \geq \frac{1}{l^2} \right) \leq \sum_{l \geq 1} \frac{1}{l^2} < \infty$$

and by Borel-Cantelli

$$\mathbb{P} \left( |S_{m(l+1)} - S_m(l)| \geq \frac{1}{l^2} \text{ i.o.} \right) = 0.$$ 

As a result, for large enough (random) $l$ and for $k > l$,

$$|S_{m(l)} - S_{m(l)}| \leq \sum_{l \geq 1} \frac{1}{l^2} < \frac{1}{l-1}.$$ 

This means that $(S_{m(l)})_{l \geq 1}$ is a Cauchy sequence and there exists the limit $S = \lim S_{m(l)}$. (6.0.2) implies that $S_n \to S$ in probability.
\textbf{Example.} Consider random series $\sum_{i \geq 1} \frac{\varepsilon_i}{i^{\alpha}}$ where $\mathbb{P}(\varepsilon_i = \pm 1) = \frac{1}{2}$. We have

$$\sum_{i \geq 1} \mathbb{E} \left( \left( \frac{\varepsilon_i}{i^{\alpha}} \right)^2 \right) = \sum_{i \geq 1} \frac{1}{i^{2\alpha}} < \infty \text{ if } \alpha > \frac{1}{2},$$

so the series converges a.s. for such $\alpha$. \hfill \qed