MTH 664

Lectures 17, 18, & 19

Yevgeniy Kovchegov
Oregon State University
Topics:

- Statistical independence.
- Laws of large numbers (LLN).
- Borel-Cantelli Lemma.
- Kolmogorov’s Maximal Inequality.
Modes of convergence.

Let \((\Omega, \mathcal{F}, P)\) be a probability space, and \(X_1, X_2, \ldots, X\) are random variables over \((\Omega, \mathcal{F})\).

- We say that \(X_n\) converges to \(X\) \textbf{\(P\)-almost everywhere (P-a.e.)} if

\[
P \left\{ \omega \in \Omega : \limsup_{n \to \infty} |X_n(\omega) - X(\omega)| > 0 \right\} = 0
\]

Since \(P\) is a probability measure, we can also say that \(X_n\) converges to \(X\) \textbf{\(P\)-almost surely (P-a.s.)}.

- Given \(p > 0\). We say that \(X_n\) converges to \(X\) \textbf{in} \(L^p(\Omega, \mathcal{F}, P)\) if

\[
\lim_{n \to \infty} \|X_n - X\|_{L^p} = \lim_{n \to \infty} \left( E[|X_n - X|^p] \right)^{1/p} = 0
\]

- We say that \(X_n\) converges to \(X\) \textbf{in probability} (or in \(P\)-measure) if for all \(\epsilon > 0\),

\[
\lim_{n \to \infty} P(|X_n - X| \geq \epsilon) = 0
\]
Modes of convergence.

**Lemma.** $X_n \to X$ $P$-almost surely if and only if
$$P( \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} \{ \omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \epsilon \} ) = 0$$
for any $\epsilon > 0$.

**Theorem.** (a). Either almost sure convergence or $L^p$-convergence implies convergence in probability.

(b). Conversely, if $Y_n := \sup_{j: j \geq n} |X_j| \to 0$ in probability, then $X_n \to 0$ $P$-almost surely.

(c). If $X_n \to 0$ in probability and $|X_n| \leq Y$ ($P$-a.s.) for some $Y \in L^p(\Omega, \mathcal{F}, P)$, then $X_n \to 0$ in $L^p$. 
Statistical independence.

Consider a probability space \((\Omega, \mathcal{F}, P)\).

- Events \(A\) and \(B\) in \((\Omega, \mathcal{F})\) are independent if
  \[
  P(A \cap B) = P(A)P(B)
  \]
  Thus, if \(P(B) > 0\), \(P(A|B) = P(A)\).

- Two \(\sigma\)-algebras \(\mathcal{G}_1 \subseteq \mathcal{F}\) and \(\mathcal{G}_2 \subseteq \mathcal{F}\) are said to be independent if all pairs of events \(A \in \mathcal{G}_1\) and \(B \in \mathcal{G}_2\) are independent.

- Events \(A_1, \ldots, A_n\) are independent if any two nonoverlapping subcollections
  \[
  A_{i_1}, \ldots, A_{i_k} \quad \text{and} \quad A_{j_1}, \ldots, A_{j_r}
  \]
  generate two independent \(\sigma\)-algebras
  \[
  \sigma(A_{i_1}, \ldots, A_{i_k}) \quad \text{and} \quad \sigma(A_{j_1}, \ldots, A_{j_r}).
  \]
**Statistical independence.**

Consider a probability space \((\Omega, \mathcal{F}, P)\).

- Random variables \(X_1, \ldots, X_n\) are **independent** if
  \[ X_1^{-1}(B_1), \ldots, X_n^{-1}(B_n) \]
  are independent for all Borel \(B_1, \ldots, B_n \in \mathcal{B}\).

- Equivalently, \(X_1, \ldots, X_n\) are **independent random variables** if
  \[
P(X_1 \in B_1, \ldots, X_n \in B_n) = \prod_{j=1}^{n} P(X_j \in B_j) \quad \forall B_1, \ldots, B_n \in \mathcal{B}
  \]
  So, the distribution of \(X = (X_1, \ldots, X_n)\) is a product measure
  \[
  \mu_1 \times \ldots \times \mu_n
  \]

- \(X_1, \ldots, X_n\) are **independent random variables** if and only if
  \[
  E[\phi_1(X_1) \cdot \ldots \cdot \phi_n(X_n)] = E[\phi_1(X_1)] \cdot \ldots \cdot E[\phi_n(X_n)]
  \]
  for all Borel measurable \(\{\phi_j\}\).
**Statistical independence.**

- \( X_1, \ldots, X_n \) are independent random variables if and only if
  \[
  E[\phi_1(X_1) \cdot \ldots \cdot \phi_n(X_n)] = E[\phi_1(X_1)] \cdot \ldots \cdot E[\phi_n(X_n)]
  \]
  for all Borel measurable \( \{\phi_j\} \).

- \( X \) and \( Y \) in \( L^2(\Omega, P) \) are said to be uncorrelated if their covariance
  \[
  Cov(X, Y) = E\left[ (X - E[X])(Y - E[Y]) \right] = 0
  \]

- If \( X \) and \( Y \) in \( L^2(\Omega, P) \) are independent, they are uncorrelated.

- If \( X_1, \ldots, X_n \) are pairwise uncorrelated, then
  \[
  Var(X_1 + \ldots + X_n) = Var(X_1) + \ldots + Var(X_n)
  \]
Kolmogorov Extension Theorem.  

Kolmogorov Extension Theorem. For each \( n \in \mathbb{N} \) let \( \mu_n \) be a probability measure over \( (\mathbb{R}^n, \mathcal{B}^n) \). And let \( \{\mu_n\} \) be consistent, i.e. for any \( n \in \mathbb{N} \) and \( \forall A_1, \ldots, A_n \in \mathcal{B} \),

\[
\mu_{n+1}(A_1 \times \ldots \times A_n \times \mathbb{R}) = \mu_n(A_1 \times \ldots \times A_n)
\]

Then there is a unique probability measure \( \pi \) on \( (\mathbb{R}^n, \mathcal{B}^n) \) such that for any \( n \in \mathbb{N} \) and \( \forall A_1, \ldots, A_n \in \mathcal{B} \),

\[
\pi(A_1 \times \ldots \times A_n \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \ldots) = \mu_n(A_1 \times \ldots \times A_n)
\]

Independent identically distributed (i.i.d.) random variables.  
If \( X_1, X_2, \ldots, X_n \) are independent identically distributed random variables, each with probability distribution \( \mu \), then \( (X_1, \ldots, X_n) \) is distributed according probability measure

\[
\mu_n = \mu \times \ldots \times \mu
\]

over \( (\mathbb{R}^n, \mathcal{B}^n) \).

The Kolmogorov Extension Theorem implies the existence of \( \pi = \mu \times \mu \times \ldots \) on \( (\mathbb{R}^n, \mathcal{B}^n) \) in which case \( X_1, X_2, \ldots \) is a sequence of i.i.d. random variables.
Laws of Large Numbers (WLLN vs. SLLN).

Let $X_1, X_2, \ldots$ be independent identically distributed (i.i.d.) random variables on a probability space $(\Omega, \mathcal{F}, P)$ with finite mean

$$\rho = E[X_j] < \infty$$

Let $S_n = X_1 + \ldots + X_n$.

**The Weak Law of Large Numbers (Khinchin, 1929).**
If $X_1, X_2, \ldots$ are in $L^1(\Omega, P)$, then, as $n \to \infty$,

$$\frac{S_n}{n} \to \rho \quad \text{in} \quad L^1(\Omega, P),$$

and hence, in probability, i.e., $P \left( \left| \frac{S_n}{n} - \rho \right| \geq \epsilon \right) \to 0$.

**The Strong Law of Large Numbers (Kolmogorov, 1933).**
If $X_1, X_2, \ldots$ are in $L^1(\Omega, P)$, then, as $n \to \infty$,

$$\lim_{n \to \infty} \frac{S_n}{n} = \rho \quad P - a.s.$$
**Strong Law of Large Numbers.**

The proof of the Strong Law of Large Numbers (SLLN) utilizes the following two probabilistic results, important on their own.

**The Borel-Cantelli Lemma.** Consider a probability space \((\Omega, \mathcal{F}, P)\) and a collection of events \(\{A_n\}_{n=1,2,...} \in \mathcal{F} \).

- If \(\sum_{n=1}^{\infty} P(A_n) < \infty\) then \(\sum_{n=1}^{\infty} 1_{A_n} < \infty\) \(P - a.s.\).

- If \(A_1, A_2, \ldots\) are pairwise independent and \(\sum_{n=1}^{\infty} P(A_n) = \infty\) then \(\sum_{n=1}^{\infty} 1_{A_n} = \infty\) \(P - a.s.\).

**Kolmogorov’s Maximal Inequality.** Let \(S_n = X_1 + \ldots + X_n\). If \(X_1, X_2, \ldots\) are independent random variables in \(L^2(\Omega, P)\), then \(\forall \lambda > 0\) and any \(n \in \mathbb{N}\),

\[
P\left( \max_{1 \leq k \leq n} |S_k - E[S_k]| \geq \lambda \right) \leq \frac{Var(S_n)}{\lambda^2}
\]
The Borel-Cantelli Lemma.

Consider a probability space \((\Omega, \mathcal{F}, P)\) and a collection of events \(\{A_n\}_{n=1}^{\infty}\) in \(\mathcal{F}\).

(a) If \(\sum_{n=1}^{\infty} P(A_n) < \infty\) then \(\sum_{n=1}^{\infty} 1_{A_n} < \infty\) \(P - a.s.\)

Proof. By the Monotone Convergence Theorem,

\[
\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} E[1_{A_n}] = \lim_{N \to \infty} E \left[ \sum_{n=1}^{N} 1_{A_n} \right] = E \left[ \lim_{N \to \infty} \sum_{n=1}^{N} 1_{A_n} \right] = E \left[ \sum_{n=1}^{\infty} 1_{A_n} \right] < \infty
\]

Hence, \(\sum_{n=1}^{\infty} 1_{A_n} \in L^1(\Omega, P)\) and therefore

\[
\sum_{n=1}^{\infty} 1_{A_n} < \infty \quad P - a.s.
\]
The Borel-Cantelli Lemma.  (b) If $A_1, A_2, \ldots$ are pairwise independent and \[ \sum_{n=1}^{\infty} P(A_n) = \infty \] then \[ \sum_{n=1}^{\infty} 1_{A_n} = \infty \quad P - a.s. \]

Proof. If \[ \sum_{n=1}^{\infty} P(A_n) = \infty, \] then

\[
\text{Var} \left( \sum_{n=1}^{N} 1_{A_n} \right) = \sum_{n=1}^{N} \text{Var}(1_{A_n}) = \sum_{n=1}^{N} P(A_n) \cdot (1 - P(A_n)) \leq \sum_{n=1}^{N} P(A_n) = E \left[ \sum_{n=1}^{N} 1_{A_n} \right]
\]

and by Chebyshev's inequality, for any $\epsilon > 0$,

\[
P \left( \left| \sum_{n=1}^{N} 1_{A_n} - E \left[ \sum_{n=1}^{N} 1_{A_n} \right] \right| \geq \epsilon E \left[ \sum_{n=1}^{N} 1_{A_n} \right] \right) \leq \frac{1}{\epsilon^2 E \left[ \sum_{n=1}^{N} 1_{A_n} \right]} = \frac{1}{\epsilon^2 \cdot \sum_{n=1}^{N} P(A_n)}
\]

Thus \[ \frac{\sum_{n=1}^{N} 1_{A_n}}{\sum_{n=1}^{N} P(A_n)} \] converges to 1 in probability, and \[ P \left( \sum_{n=1}^{\infty} 1_{A_n} < \infty \right) = 0. \]
Kolmogorov’s Maximal Inequality.

Kolmogorov’s Maximal Inequality. Let $S_n = X_1 + \ldots + X_n$. If $X_1, X_2, \ldots$ are independent random variables in $L^2(\Omega, P)$, then $\forall \lambda > 0$ and any $n \in \mathbb{N}$,

$$P\left( \max_{1 \leq k \leq n} |S_k - E[S_k]| \geq \lambda \right) \leq \frac{\text{Var}(S_n)}{\lambda^2}$$

**Proof.** Assume $E[X_j] = 0$ for all $j$, as otherwise we can consider $\tilde{X}_j = X_j - E[X_j]$. Thus we need to prove

$$P\left( \max_{1 \leq k \leq n} |S_k| \geq \lambda \right) \leq \frac{E[S_n^2]}{\lambda^2}$$

Let $A_1 = \{|S_1| \geq \lambda\}$, and for all $k \geq 2$, let $A_k = \{|S_1| < \lambda, \ldots, |S_{k-1}| < \lambda, |S_k| \geq \lambda\}$. Since $A_1, A_2, \ldots$ are disjoint, and $S_n^2 \geq 2(S_n - S_k)S_k + S_k^2$,

$$E[S_n^2] \geq \sum_{k=1}^n E[S_n^2 \cdot 1_{A_k}] \geq 2 \sum_{k=1}^n E[(S_n - S_k)S_k \cdot 1_{A_k}] + \sum_{k=1}^n E[S_k^2 \cdot 1_{A_k}]$$

Next, since $S_n - S_k$ and $S_k \cdot 1_{A_k}$ are independent random variables,

$$E[(S_n - S_k)S_k \cdot 1_{A_k}] = E[(S_n - S_k)] \cdot E[S_k \cdot 1_{A_k}] = 0$$

and $E[S_n^2] \geq \sum_{k=1}^n E[S_k^2 \cdot 1_{A_k}] \geq \lambda^2 \sum_{k=1}^n P(A_k) = \lambda^2 P(\cup_{k=1}^n A_k)$ \qed
The Strong Law of Large Numbers (Kolmogorov, 1933).

Consider i.i.d. \( X_1, X_2, \ldots \), \( \rho = E[X_j] \), and let \( S_n = X_1 + \ldots + X_n \).

(a). If \( X_1, X_2, \ldots \) are in \( L^1(\Omega, P) \), then, as \( n \to \infty \), \( \lim_{n \to \infty} \frac{S_n}{n} = \rho \) \( P - a.s. \)

(b). If \( P\left( \limsup_{n \to \infty} \frac{|S_n|}{n} < \infty \right) > 0 \), then \( X_1, X_2, \ldots \) are in \( L^1(\Omega, P) \).

Proof of part (b). Suppose \( E[|X_j|] = \infty \). Now,

\[
\limsup_{n \to \infty} \frac{|X_n|}{n} = \limsup_{n \to \infty} \frac{|S_n - S_{n-1}|}{n} \leq \limsup_{n \to \infty} \frac{|S_n| + |S_{n-1}|}{n} \leq 2 \cdot \limsup_{n \to \infty} \frac{|S_n|}{n}
\]

Next, for any fixed \( m > 0 \),

\[
E[|X_1|] \leq mE\left[ \left\lfloor m^{-1} |X_1| \right\rfloor \right] \leq m \cdot \sum_{n=0}^{\infty} P(|X_1| \geq nm) = m \cdot \sum_{n=0}^{\infty} P(|X_n| \geq nm)
\]

So, \( \sum_{n=0}^{\infty} P\left( \frac{|X_n|}{n} \geq m \right) = \infty \), and the Borel-Cantelli Lemma (b) implies \( \limsup_{n \to \infty} \frac{|X_n|}{n} \geq m \) \( a.s. \). Thus, as \( m \to \infty \), \( \limsup_{n \to \infty} \frac{|X_n|}{n} = \infty \) \( a.s. \).

Hence, \( \limsup_{n \to \infty} \frac{|X_n|}{n} \leq 2 \cdot \limsup_{n \to \infty} \frac{|S_n|}{n} \) implies \( \limsup_{n \to \infty} \frac{|S_n|}{n} = \infty \) \( a.s. \). □
The Strong Law of Large Numbers (Kolmogorov, 1933).

Consider i.i.d. $X_1, X_2, \ldots$, $\rho = E[X_j]$, and let $S_n = X_1 + \ldots + X_n$.

(a). If $X_1, X_2, \ldots$ are in $L^1(\Omega, P)$, then, as $n \to \infty$, $\lim_{n \to \infty} \frac{S_n}{n} = \rho$ $P$-a.s.

Proof of part (a). Assume $E[X_j] = 0$ for all $j$, as otherwise we can consider $\tilde{X}_j = X_j - E[X_j]$. We consider two cases, $L^2$ and general.

$L^2$ case. If $X_1, X_2, \ldots$ are in $L^2(\Omega, P)$, then $\sigma^2 = Var(X_j) = E[X_j^2] < \infty$, and by Kolmogorov’s Maximal Inequality, for any $\epsilon > 0$,

$$P\left(\max_{1 \leq k \leq N} |S_k| \geq \epsilon N\right) \leq \frac{E[S_N^2]}{\epsilon^2 N^2} = \frac{\sigma^2}{\epsilon^2 N}$$

Next, let $N = 2^n$, then

$$\sum_{n=1}^{\infty} P\left(\max_{1 \leq k \leq 2^n} |S_k| \geq \epsilon \cdot 2^n\right) \leq \sum_{n=1}^{\infty} \frac{\sigma^2}{\epsilon^2 \cdot 2^n} < \infty$$

Borel-Cantelli Lemma $\implies \max_{1 \leq k \leq 2^n} |S_k| < \epsilon \cdot 2^n$ for all but finitely many $n$’s, $a.s.$
Proof of part (a) (continued). Assume $E[X_j] = 0$ for all $j$.

$L^2$ case. If $X_1, X_2, \ldots$ are in $L^2(\Omega, P)$.

We used Kolmogorov’s Maximal Inequality and Borel-Cantelli Lemma to show for any $\epsilon > 0$,

$$\max_{1 \leq k \leq 2^n} |S_k| < \epsilon \cdot 2^n$$

for all but finitely many $n$’s, a.s. Next we use a sandwich trick: For any $m$ there is $n$ s.t. $2^{n-1} < m \leq 2^n$ and therefore

$$|S_m| \leq \max_{1 \leq k \leq 2^n} |S_k| < \epsilon \cdot 2^n = 2\epsilon \cdot 2^{n-1} < 2\epsilon m$$

for $m$ large enough, a.s. Then

$$P \left( \bigcap_{M \in \mathbb{N}, \; \epsilon = 1/M} \left\{ \limsup_{m \to \infty} \frac{|S_m|}{m} \leq 2\epsilon \right\} \right) = 1$$

proving SLLN for the case when $X_1, X_2, \ldots$ are in $L^2(\Omega, P)$.

For the general case, we use the truncation argument.
Proof of part (a) (continued). Assume $E[X_j] = 0$ for all $j$.

We proved SLLN for the case when $X_1, X_2, \ldots$ are in $L^2(\Omega, P)$.

General case. Suppose $X_1, X_2, \ldots$ are in $L^1(\Omega, P)$. Let

$$\tilde{X}_j = X_j \cdot 1_{|X_j| \leq j} \quad \text{and} \quad \tilde{S}_n = \tilde{X}_1 + \ldots + \tilde{X}_n$$

Observe that $\infty > E[|X_j|] \geq \sum_{j=0}^{\infty} P(|X_j| > j)$, and by the Borel-Cantelli Lemma, $X_j = \tilde{X}_j$ for all but finitely many $j$, a.s.

Thus

$$\left| \frac{S_n}{n} - \frac{\tilde{S}_n}{n} \right| \to 0 \quad \text{a.s.}$$

Now,

$$|E[\tilde{S}_n]| = |E[\tilde{S}_n - S_n]| = \left| \sum_{j=1}^{n} E[X_j \cdot 1_{|X_j| > j}] \right| \leq \sum_{j=1}^{n} E[|X_1| \cdot 1_{|X_1| > j}] \leq E[|X_1| \cdot \min(|X_1|, n)]$$

Thus, by the Monotone Convergence Theorem, $\frac{E[\tilde{S}_n]}{n} \to 0$.

Hence, we only need to show that $\frac{\tilde{S}_n - E[\tilde{S}_n]}{n} \to 0 \; a.s.$
Proof of part (a) (continued). Assume $E[X_j] = 0$ for all $j$. We let $\tilde{X}_j = X_j \cdot 1_{|X_j| \leq j}$ and $\tilde{S}_n = \tilde{X}_1 + \ldots + \tilde{X}_n$.

Now, by the Kolmogorov’s Maximal Inequality (with independent but not necessarily i.i.d. $\tilde{X}_j$’s),

$$P\left( \max_{1 \leq k \leq N} |\tilde{S}_k - E[\tilde{S}_k]| \geq \epsilon N \right) \leq \frac{Var(\tilde{S}_N)}{\epsilon^2 N^2} = \frac{Var(\tilde{X}_1 + \ldots + \tilde{X}_N)}{\epsilon^2 N^2}$$

$$\leq \frac{E[\tilde{X}_1^2] + \ldots + E[\tilde{X}_N^2]}{\epsilon^2 N^2} = \sum_{j=1}^{N} \frac{E[X_1^2 \cdot 1_{X_1 \leq j}]}{\epsilon^2 N^2}$$

and, summing up over $N = 2^n$,

$$\sum_{n=1}^{\infty} P\left( \max_{1 \leq k \leq 2^n} |\tilde{S}_k - E[\tilde{S}_k]| \geq \epsilon \cdot 2^n \right) \leq \sum_{n=1}^{\infty} \sum_{j=1}^{2^n} \frac{E[X_1^2 \cdot 1_{X_1 \leq j}]}{\epsilon^2 2^{2n}}$$

$$= \frac{1}{\epsilon^2} \sum_{j=1}^{\infty} \left( E[X_1^2 \cdot 1_{X_1 \leq j}] \cdot \sum_{n: 2^n \geq j} 2^{-2n} \right) \leq \frac{16}{3\epsilon^2} \sum_{j=1}^{\infty} \frac{E[X_1^2 \cdot 1_{X_1 \leq j}]}{j^2}$$

$$= \frac{16}{3\epsilon^2} E \left[ X_1^2 \cdot \sum_{j: j \geq |X_1|} \frac{1}{j^2} \right] \quad \text{as} \quad \sum_{n=\lfloor \log_2 j \rfloor}^{\infty} 2^{-2n} \leq \frac{4^{2-\log_2 j}}{3}.$$
Proof of part (a) (continued). Assume $E[X_j] = 0$ for all $j$.

We let $\tilde{X}_j = X_j \cdot 1_{|X_j| \leq j}$ and $\tilde{S}_n = \tilde{X}_1 + \ldots + \tilde{X}_n$, and showed

$$\sum_{n=1}^{\infty} P\left( \max_{1 \leq k \leq 2^n} \left| \tilde{S}_k - E[\tilde{S}_k] \right| \geq \epsilon \cdot 2^n \right) \leq \frac{16}{3\epsilon^2} E \left[ X_1^2 \cdot \sum_{j: j \geq |X_1|} \frac{1}{j^2} \right] \leq \frac{32}{3\epsilon^2} E[|X_1|] < \infty$$

as $x \cdot \sum_{j: j \geq x} \frac{1}{j^2} \leq 2$ for all $x > 0$.

Thus, by the Borel-Cantelli Lemma, for any $\epsilon > 0$,

$$\max_{1 \leq k \leq 2^n} \left| \tilde{S}_k - E[\tilde{S}_k] \right| < \epsilon \cdot 2^n$$

for all but finitely many $n$’s, a.s. Next we use a sandwich trick: For any $m$ there is $n$ s.t. $2^{n-1} < m \leq 2^n$ and therefore

$$\left| \tilde{S}_m - E[\tilde{S}_m] \right| \leq \max_{1 \leq k \leq 2^n} \left| \tilde{S}_k - E[\tilde{S}_k] \right| < \epsilon \cdot 2^n = 2\epsilon \cdot 2^{n-1} < 2\epsilon m$$

for $m$ large enough, a.s. Then

$$P\left( \bigcap_{M \in \mathbb{N}, \; \epsilon=1/M} \left\{ \limsup_{m \to \infty} \frac{\left| \tilde{S}_m - E[\tilde{S}_m] \right|}{m} \leq 2\epsilon \right\} \right) = 1$$

proving SLLN for the case when $X_1, X_2, \ldots$ are in $L^1(\Omega, P)$. □