Modes of Convergence in Probability Theory

David Mandel

November 5, 2015

Below, fix a probability space $(\Omega, \mathcal{F}, P)$ on which all random variables $\{X_n\}$ and $X$ are defined. All random variables are assumed to take values in $\mathbb{R}$. Propositions marked with “★” denote results that rely on our finite measure space. That is, those marked results may not hold on a non-finite measure space. Since we already know uniform convergence $\implies$ pointwise convergence this proof is omitted, but we include a proof that shows pointwise convergence $\implies$ almost sure convergence, and hence uniform convergence $\implies$ almost sure convergence.

The hierarchy we will show is diagrammed in Fig. 1 where some famous theorems that demonstrate the type of convergence are in parentheses: (SLLN) = strong long of large numbers, (WLLN) = weak law of large numbers, (CLT) = central limit theorem. In parameter estimation, $\hat{\theta}_n$ is said to be a consistent estimator or $\theta$ if $\hat{\theta}_n \to \theta$ in probability. For example, by the SLLN, $\bar{X}_n \to \mu$ a.s., and hence $\bar{X}_n \to \mu$ in probability. Therefore the sample mean is a consistent estimator of the population mean.

![Hierarchy of modes of convergence in probability.

Figure 1: Hierarchy of modes of convergence in probability.](image-url)
1 Definitions of Convergence

1.1 Modes from Calculus

Definition \( X_n \to X \) pointwise if \( \forall \omega \in \Omega, \forall \epsilon > 0, \exists N \in \mathbb{N} \) such that \( \forall n \geq N \),
\[
|X_n(\omega) - X(\omega)| < \epsilon.
\]

Definition \( X_n \to X \) uniformly if \( \forall \epsilon > 0, \exists N \in \mathbb{N} \) such that \( \forall \omega \in \Omega \) and \( \forall n \geq N \),
\[
|X_n(\omega) - X(\omega)| < \epsilon.
\]

1.2 Modes Unique to Measure Theory

Definition \( X_n \to X \) in probability if \( \forall \epsilon > 0, \forall \delta > 0, \exists N \in \mathbb{N} \) such that \( \forall n \geq N \),
\[
P(|X_n - X| \geq \epsilon) < \delta.
\]
Or, \( X_n \to X \) in probability if \( \forall \epsilon > 0 \),
\[
\lim_{n \to \infty} P(|X_n - X| \geq 0) = 0.
\]
The explicit epsilon-delta definition of convergence in probability is useful for proving a.s. \( \implies \) prob.

Definition \( X_n \to X \) in \( L^p \), \( p \geq 1 \), if \( \forall \epsilon > 0, \exists N \in \mathbb{N} \) such that \( \forall n \geq N \),
\[
\int_{\Omega} |X_n(\omega) - X(\omega)| dP(\omega) =: \|X_n - X\|_{L^p}^p < \epsilon.
\]
Or, \( X_n \to X \) in \( L^p \) if \( \lim_{n \to \infty} \|X_n - X\|_{L^p}^p = 0 \).

Definition \( X_n \to X \) almost surely (a.s.) if \( \exists E \in \mathcal{F} \) with \( P(E) = 0 \) such that \( \forall \omega \in E^c \) and \( \forall \epsilon > 0, \exists N \in \mathbb{N} \) such that \( \forall n \geq N \),
\[
|X_n(\omega) - X(\omega)| < \epsilon.
\]
Or, \( X_n \to X \) a.s. if \( \exists E \in \mathcal{F} \) with \( P(E) = 0 \) such that \( X_n \to X \) pointwise on \( E^c \).

1.3 Mode Unique to Probability Theory

Definition \( X_n \to X \) in distribution if the distribution functions of the \( X_n \) converge pointwise to the distribution function of \( X \) at all points \( x \) where \( F(x) \) is continuous. That is, \( X_n \to X \) in distribution if \( \forall x \in \mathbb{R} \) such that \( F(x) \) is continuous, \( \forall \epsilon > 0, \exists N \in \mathbb{N} \) such that \( \forall n \geq N \),
\[
|F_n(x) - F(x)| < \epsilon.
\]
Or, \( X_n \to X \) in distribution if \( \forall x \in \mathbb{R} \) such that \( F(x) \) is continuous,
\[
\lim_{n \to \infty} F_n(x) = F(x).
\]
2 Convergence Results

Proposition Pointwise convergence \(\implies\) almost sure convergence.

Proof Let \(\omega \in \Omega\), \(\epsilon > 0\) and assume \(X_n \to X\) pointwise. Then \(\exists N \in \mathbb{N}\) such that \(\forall n \geq N, |X_n(\omega) - X(\omega)| < \epsilon\). Hence \(X_n \to X\) almost surely since this convergence takes place on all sets \(E \in \mathcal{F}\).

Proposition Uniform convergence \(\implies\) convergence in probability.

Proof Let \(\epsilon > 0\) and assume \(X_n \to X\) uniformly. Then \(\exists N \in \mathbb{N}\) such that \(\forall \omega \in \Omega\) and \(\forall n \geq N, |X_n(\omega) - X(\omega)| < \epsilon\). Let \(n \geq N\). Then \(P(|X_n - X| \geq \epsilon) = P(\emptyset) = 0\), so of course \(P(|X_n - X| \geq \epsilon) \to 0\) as \(n \to \infty\).

Proposition \(L^p\) convergence \(\implies\) convergence in probability.

Proof Let \(\epsilon > 0\) and assume \(X_n \to X\) in \(L^p\), \(p \geq 1\). Then

\[
P(|X_n - X| \geq \epsilon) \leq \frac{||X_n - X||_{L^p}^p}{\epsilon^p} \to 0 \text{ as } n \to \infty,
\]

where the first inequality is Chebyshev’s inequality.

\section*{★ Proposition} Uniform convergence \(\implies\) \(L^p\) convergence.

Proof Let \(\epsilon > 0\) and assume \(X_n \to X\) uniformly. Then \(\exists N \in \mathbb{N}\) such that \(\forall \omega \in \Omega\) and \(\forall n \geq N, |X_n(\omega) - X(\omega)| < \epsilon\). Let \(n \geq N\). Then

\[
||X_n - X||_{L^p}^p = \int_{\Omega} |X_n(\omega) - X(\omega)|^p dP(\omega) < \epsilon P(\Omega) = \epsilon.
\]

\section*{★ Proposition} Almost sure convergence \(\implies\) convergence in probability. This result relies on Egoroff’s theorem, which we state now without proof.

Theorem 2.1 (Egoroff) Let \(P(\Omega) < \infty\) and assume \(X_n \to X\) a.s. Then \(\forall \delta > 0, \exists E \in \mathcal{F}\) with \(P(E) < \delta\) such that \(X_n \to X\) uniformly on \(E^c\).

Proof Let \(\epsilon > 0\). Also let \(\delta > 0\) and let \(E \in \mathcal{F}\) such a set as in Egoroff’s theorem. Since \(X_n \to X\) uniformly on \(E^c\), \(\exists N \in \mathbb{N}\) such that \(\forall \omega \in E^c\) and \(\forall n \geq N, |X_n(\omega) - X(\omega)| < \epsilon\). Let \(n \geq N\). Then

\[
P(|X_n - X| \geq \epsilon) = P(|X_n - X| \geq \epsilon \cap E) + P(|X_n - X| \geq \epsilon \cap E^c)
\]

\[
< \delta + P(\emptyset) = \delta.
\]
Proposition Convergence in probability $\implies$ convergence in distribution.

Proof First, a lemma.

Lemma 2.2 Let $X_n$ and $X$ be random variables. Then $\forall \epsilon > 0$ and $x \in \mathbb{R}$,

\[
P(X_n \leq x) \leq P(X \leq x + \epsilon) + P(|X_n - X| > \epsilon),
\]
\[
P(x \leq x - \epsilon) \leq P(X_n \leq x) + P(|X_n - X| > \epsilon),
\]

Proof (lemma) For the first line,

\[
P(X_n \leq x) = P(X_n \leq x, X \leq x + \epsilon) + P(X_n \leq x, X > x + \epsilon)
\]
\[
\leq P(X \leq x + \epsilon) + P(X_n - X \leq x - X, x - X < -\epsilon)
\]
\[
\leq P(X \leq x + \epsilon) + P(X_n - X < -\epsilon)
\]
\[
\leq P(X \leq x + \epsilon) + P(X_n - X < -\epsilon) + P(X_n - X > \epsilon)
\]
\[
= P(X \leq x + \epsilon) + P(|X_n - X| > \epsilon).
\]

For the second line,

\[
P(x \leq x - \epsilon) = P(X \leq x - \epsilon, X_n \leq x) + P(X \leq x - \epsilon, X_n > x)
\]
\[
\leq P(X_n \leq x) + P(x - X \geq \epsilon, X_n - X > x - X)
\]
\[
\leq P(X_n \leq x) + P(X_n - X > \epsilon)
\]
\[
\leq P(X_n \leq x) + P(X_n - X > \epsilon) + P(X_n - X < -\epsilon)
\]
\[
= P(X_n \leq x) + P(|X_n - X| > \epsilon).
\]

Let $x \in \mathbb{R}$ be a continuity point of $F$ and let $\epsilon > 0$. The lemma then tells us that

\[
F_n(x) \leq F(x + \epsilon) + P(|X_n - X| > \epsilon),
\]
\[
F(x - \epsilon) \leq F_n(x) + P(|X_n - X| > \epsilon).
\]

Hence

\[
F(x - \epsilon) - P(|X_n - X| > \epsilon) \leq F_n(x) \leq F(x + \epsilon) + P(|X_n - X| > \epsilon).
\]

Now let $n \to \infty$, in which case $P(|X_n - X| > \epsilon) \to 0$:

\[
F(x - \epsilon) \leq \liminf_{n \to \infty} F_n(x) \leq \limsup_{n \to \infty} F_n(x) \leq F(x + \epsilon).
\]

Finally let $\epsilon \to 0^+$ and use continuity of $F$ at $x$:

\[
F(x) \leq \liminf_{n \to \infty} F_n(x) \leq \limsup_{n \to \infty} F_n(x) \leq F(x),
\]

hence $\lim_{n \to \infty} F_n(x) = F(x)$. □
3 References

References

Proofs of Convergence of Random Variables

[2] University of Alabama, Birmingham
Convergence in Distribution (Lecture Notes)
http://www.math.uah.edu/stat/dist/Convergence.html

Real Analysis: Modern Techniques and Their Applications