1.9
Convergence: Almost Surely and in Probability

For nets of maps defined on a single, fixed probability space \((\Omega, \mathcal{A}, P)\), convergence almost surely and in probability are frequently used modes of stochastic convergence, stronger than weak convergence. In this section we consider their nonmeasurable extensions together with the concept of almost uniform convergence, which is equivalent to outer almost sure convergence for sequences, but stronger and more useful for general nets.

1.9.1 Definition. Let \(X_\alpha, X: \Omega \to \mathbb{D}\) be arbitrary maps.

(i) \(X_\alpha\) converges in outer probability to \(X\) if \(d(X_\alpha, X)^* \to 0\) in probability; this means that \(P(d(X_\alpha, X)^* > \epsilon) = P^*(d(X_\alpha, X) > \epsilon) \to 0\), for every \(\epsilon > 0\), and is denoted by \(X_\alpha \overset{P^*}{\to} X\).

(ii) \(X_\alpha\) converges almost uniformly to \(X\) if, for every \(\epsilon > 0\), there exists a measurable set \(A\) with \(P(A) \geq 1 - \epsilon\) and \(d(X_\alpha, X) \to 0\) uniformly on \(A\); this is denoted \(X_\alpha \overset{a. u.}{\to} X\).

(iii) \(X_\alpha\) converges outer almost surely to \(X\) if \(d(X_\alpha, X)^* \to 0\) almost surely for some versions of \(d(X_\alpha, X)^*\); this is denoted \(X_\alpha \overset{a. s.}{\to} X\).

(iv) \(X_\alpha\) converges almost surely to \(X\) if \(P_*(\lim d(X_\alpha, X) = 0) = 1\); this is denoted \(X_\alpha \overset{a. s.}{\to} X\).

The first three concepts in this list are the most important. The fourth is tricky — it does not behave as one might expect, in general. In fact, for nonmeasurable sequences, even convergence everywhere (which is stronger than (iv)) does not imply any of the other three forms of convergence. This is one of those phenomena that remind us that measurability, though often
present, should not be taken too lightly. The counterexample given below isn’t even very complicated, though contrived, perhaps.

For sequences the second and third modes of convergence are equivalent, but for general nets the third loses much of its value. First, there is the nuisance that for general nets outer almost sure convergence depends on the versions of the minimal measurable covers one uses. Second, even for measurable nets $X_\alpha$, outer almost sure convergence is rather weak. For instance, it does not imply convergence in probability. For these reasons we consider (iii) and (iv) for sequences only. The Problems and Complements section gives more details about the general implications between the four forms of convergence. In the following text, we use the notation $X_\alpha$ for a general net and $X_n$ for a sequence, without further mention.

For sequences things are partly as they should be. Outer almost sure convergence (iii) is stronger than convergence in outer probability. Equivalence of almost uniform and outer almost sure convergence, part (iii) of the following lemma, is known as Egorov’s theorem.

**1.9.2 Lemma.** Let $X$ be Borel measurable. Then

1. $X_n \xrightarrow{as} X$ implies $X_n \xrightarrow{P} X$;
2. $X_n \xrightarrow{P} X$ if and only if every subsequence $X_{n'}$ has a further subsequence $X_{n''}$ with $X_{n''} \xrightarrow{as} X$;
3. $X_n \xrightarrow{as} X$ if and only if $X_n \xrightarrow{au} X$.

**1.9.3 Lemma.** Let $X$ be Borel measurable. Then

1. $X_\alpha \xrightarrow{au} X$ if and only if $\sup_{\beta \geq \alpha} d(X_\beta, X)^* \xrightarrow{P} 0$ if and only if $\sup_{\beta \geq \alpha} d(X_\beta, X)^* \xrightarrow{P} 0$;
2. $X_\alpha \xrightarrow{au} X$ implies $X_n \xrightarrow{P} X$.

**Proofs.** (iii). Suppose $X_n$ converges outer almost surely to $X$. Fix $\varepsilon > 0$. Set $A_n^k = \{\sup_{m \geq n} d(X_m, X)^* > k\}$. Then, for every fixed $k$, it holds that $P(A_n^k) \downarrow 0$ as $n \to \infty$. Choose $n_k$ with $P(A_n^k) \leq \varepsilon/2^k$, and set $A = \Omega - \bigcup_{k=1}^\infty A_{n_k}^k$. Then $P(A) \geq 1 - \varepsilon$ and $d(X_n, X)^* \leq k$, for $n \geq n_k$ and $\omega \in A$. Thus $X_n$ converges to $X$ almost uniformly. Conversely, suppose $X_n$ converges almost uniformly to $X$. Fix $\varepsilon > 0$, and let $A$ be as in the definition of almost uniform convergence. Fix $\eta > 0$. Then $d(X_n, X)^* \xrightarrow{1_A} 0$ for sufficiently large $n$, since $\eta$ is a measurable function and $\eta \geq d(X_n, X)^* \xrightarrow{1_A}$ for sufficiently large $n$. Thus $d(X_n, X)^* \to 0$ for almost all $\omega \in A$.

Next consider the second lemma first.

(i). It is easy to see that the first statement implies the second, which implies the third. For the converse, fix $\varepsilon > 0$. Take $\alpha_k$ such that $P^*(\sup_{\beta \geq \alpha_k} d(X_\beta, X) > k) \leq \varepsilon/2^k$. Call the set within brackets $A_k$, and set $A = \Omega - \bigcup_{k=1}^\infty A_k^*$. Then $P(A) \geq 1 - \varepsilon$, and for every $\omega \in A$ and $\alpha \geq \alpha_k$, it holds that $d(X_\alpha, X)^* \leq k$.

(ii). This is an immediate consequence of (i).