Lecture 6: Convergence and related concepts

1. Borel-Cantelli Lemmas

Let $E_1, E_2, E_3, \ldots$ be a sequence of events on some probability space $(\Omega, \mathcal{F}, P)$. In many situations, an important question is whether infinitely many or only finitely many of these events occur. The Borel-Cantelli lemmas provide the answer to this question, but first we need some preliminary definitions and notations.

Define

$$\limsup_{n\to\infty} E_n = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} E_n = \{E_n \text{ infinitely often}\}$$

$$\liminf_{n\to\infty} E_n = \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} E_n = \{E_n \text{ eventually}\}.$$

One advantage of the above formal notation is that it makes clear that $\{E_n \text{ infinitely often}\}$ and $\{E_n \text{ eventually}\}$ are themselves events (since σ-algebras are closed under countable unions and intersections). In addition, note that

$$\begin{aligned}
\{E_n \text{ eventually}\}^c &= \{E_n^c \text{ infinitely often}\} \\
\{E_n \text{ infinitely often}\}^c &= \{E_n^c \text{ eventually}\} \\
I_{\limsup E_n} &= \limsup I_{E_n} \\
I_{\liminf E_n} &= \liminf I_{E_n}.
\end{aligned}$$

Applying Fatou's Lemma to the random variables I_{E_n}, we immediately obtain

$$P(\liminf E_n) \le \liminf P(E_n) \tag{1.1}$$

and applying Fatou to $1 - I_{E_n}$, we also obtain the reverse Fatou Lemma

$$P(\limsup E_n) \ge \limsup P(E_n). \tag{1.2}$$

Lemma 1.1. *(First Borel-Cantelli Lemma) Let $E_1, E_2, E_3, \ldots$ be a sequence of events such that $\sum_{n=1}^{\infty} P(E_n) < \infty$. Then $P(\limsup E_n) = P(E_n,\ i.o.) = 0$.*

Proof. Let $F_m = \bigcup_{n=m}^{\infty} E_n$ and $F = \limsup E_n$. Then since for each m, $F \subset F_m$ we have

$$P(F) \le P(F_m) \le \sum_{n=m}^{\infty} P(E_n).$$

The result follows by letting $m \to \infty$. $\square$

Lemma 1.2. *(Second Borel-Cantelli Lemma) Let $E_1, E_2, E_3, \ldots$ be a sequence of* independent *events such that $\sum_{n=1}^{\infty} P(E_n) = \infty$. Then $P(\limsup E_n) = P(E_n,\ i.o.) = 1$.*

Proof. We show that $P((\limsup E_n)^c) = 0$. Recall that

$$(\limsup E_n)^c = \liminf E_n^c = \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} E_n^c.$$

Let $p_n = P(E_n)$. Then by independence, we have

$$P(\bigcap_{n=m}^{\infty} E_n^c) = \prod_{n=m}^{\infty}(1-p_n).$$

Next, note that for $x \geq 0$, $1 - x \leq e^{-x}$, so since $\sum p_n = \infty$, we have

$$\prod_{n=m}^{\infty}(1-p_n) \leq \exp\left\{-\sum_{n\geq m} p_n\right\} = 0.$$

Finally, since the countable union of events of probability 0 has probability 0, we have $P((\limsup E_n)^c) = 0$. □

The subsequent sections contain many applications of these 2 results.

2. Modes of convergence

Fix a probability space $(\Omega, \mathcal{F}, P)$. In what follows, all random variables and events live on this same probability space, unless otherwise stated. For $p \geq 1$, let $L^p = L^p(\Omega, \mathcal{F}, P) = \{X : E(|X|^p) < \infty\}$ and $L^\infty = \{X : |X| \leq K \text{ a.s. for some } K\}$. Recall that $L^\infty \subset L^p \subset L^q \subset L^1$ for all $p \geq q$. Some definitions:

Convergence in probability A sequence of random variables $(X_n : n = 1, 2, 3, \ldots)$ is said to converge to random variable X in probability (as $n \to \infty$) if for every $\epsilon > 0$, $P(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$.

Convergence in L^p A sequence of random variables $(X_n : n = 1, 2, 3, \ldots)$ is said to converge to random variable X in L^p $(1 \leq p < \infty)$ if $E(|X_n - X|^p) \to 0$ as $n \to \infty$.

Note: Since for $p \geq q$ $L^p \subset L^q$, it is clear that $X_n \to X$ in L^q implies $X_n \to X$ in L^p.

Convergence almost surely A sequence of random variables $(X_n : n = 1, 2, 3, \ldots)$ is said to converge to random variable X almost surely if $P(X_n \to X) = 1$.

The definition of almost sure convergence raises an important technical question: implicit in the statement that $P(X_n \to X) = 1$ is that $\{X_n \to X\}$ is an event, but this is not obvious at first sight!

For a sequence of random variables $(X_n : n = 1, 2, 3, \ldots)$, let $\mathcal{T}_n = \sigma(X_{n+1}, X_{n+2}, \ldots)$ and $\mathcal{T} = \bigcap_n \mathcal{T}_n$. The σ-algebra $\mathcal{T}$ is called the *tail σ-algebra* for the sequence $(X_n : n = 1, 2, 3, \ldots)$. Clearly, $\mathcal{T} \subset \mathcal{F}$. The tail σ-algebra contains many important events related to convergence, for example, the sets $\{\lim X_n \text{ exists}\}$, $\{\sum_n X_n < \infty\}$ and $\{\lim \frac{X_1 + \cdots + X_n}{n} \text{ exists}\}$ all belong to $\mathcal{T}$ and hence are events. We will discuss the tail σ-algebra in greater detail later (Kolomogorov 0-1 law, strong law of large numbers etc.)

Of the 3 different modes of convergence introduced above, convergence in probability is the weakest, as the next two results demonstrate.

Proposition 2.1. *$X_n \to X$ a.s. implies $X_n \to X$ in probability.*

Proof. By definition, if $X_n \to X$ a.s. then for every $\epsilon > 0$, the set

$$\{\omega : \exists N(\omega, \epsilon) \text{ s.t. } |X_n(\omega) - X(\omega)| < \epsilon \ \forall n \geq N\}$$

has probability 1, or in other words, $P(|X_n - X| > \epsilon \text{ i.o}) = 0$. By the reverse Fatou lemma (1.2),

$$0 = P(\limsup\{|X_{n_k} - X| > \epsilon\}) \geq \limsup P(|X_{n_k} - X| > \epsilon).$$

Since probabilities are non-negative, this shows that $\lim P(|X_{n_k} - X| > \epsilon) = 0$. □

Proposition 2.2. *For any $p \geq 1$, $X_n \to X$ in L^p implies $X_n \to X$ in probability.*

Proof. Suppose that $P(|X_n - X| > \epsilon) \not\to 0$, then for some $\delta > 0$, there is an infinite sequence (n_k) such that $P(|X_{n_k} - X| > \epsilon) > \delta$ for all k. Then

$$E(|X_{n_k} - X|^p) \geq E(|X_{n_k} - X|^p;\ |X_{n_k} - X| > \epsilon) > \epsilon^p P(|X_{n_k} - X| > \epsilon) > \epsilon^p \delta$$

for all k, so that $E(|X_{n_k} - X|^p) \not\to 0$. □

It is not hard to construct examples to show that the converses to the above 2 results are false.

Example 2.1. Let $E_1, E_2, \dots$ be a sequence of independent events with $P(E_n) = 1/n$ and let $X_n = nI_{E_n}$. Then since for any $0 < \epsilon < 1$, $P(|X_n| > \epsilon) = P(E_n) = 1/n \to 0$, $X_n \to 0$ in probability. But $\sum P(|X_n| > \epsilon) = \sum 1/n = \infty$ so by Borel-Cantelli II, almost surely, $|X_n| > \epsilon$ for infinitely many n so $X_n \not\to 0$. Also $E(|X_n|^p) = n^p P(E_n) = n^{p-1}$, so $X_n \not\to 0$ in L^p for $p \geq 1$.

Convergence in distribution (weak convergence)

This idea of convergence differs in nature from the others above, in that the modes of convergence above are all in one way or another to do with sample path behaviour, in particular related to sets of the form $\{|X_n - X| > \epsilon\}$, whereas convergence in distribution is purely a property of the distributions of the random variables concerned.

Theorem 2.3. *Let $(F_n,\ n \geq 0)$ be a sequence of distribution functions and let F be a distribution function. Let ϕ_n and ϕ denote the associated characteristic functions:*

$$\phi_n(\theta) = \int_{\mathbb{R}} e^{i\theta x}\, dF_n(x), \quad \phi(\theta) = \int_{\mathbb{R}} e^{i\theta x}\, dF(x).$$

Then the following statements are equivalent:

1. *$F_n(x) \to F(x)$ for every x where F is* continuous*;*
2. *for every* continuous and bounded *function h,*
$$\int_{\mathbb{R}} h(x)\, dF_n(x) \to \int_{\mathbb{R}} h(x)\, dF(x);$$
3. *$\phi_n(\theta) \to \phi(\theta)$ for all θ.*

The proof of this theorem is beyond the scope of these lectures; we will simply treat the above equivalent statements as the definition of convergence in distribution. As can be seen from the statement of Theorem 2.3, there is no need for any mention of random variables at all although it is often helpful to think in terms of random variables. Thus, if X_n is a random variable whose distribution function is F_n (or equivalently whose characteristic function is ϕ_n) and X is a random variable whose distribution function is F, then we say that $X_n \to X$ in distribution if any of the statements in Theorem 2.3 holds; in particular, note that Theorem 2.3(2) says $E(h(X_n)) \to E(h(X))$.

The continuity requirements in Theorem 2.3 are crucial, as the following examples show.

Example 2.2. Let X_n have $N(0, \sigma_n^2)$ distribution where $\sigma_n^2 \to 0$. Then $X_n \to 0$ in distribution, but the limiting distribution function (of the deterministic random variable 0) is

$$F(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases}$$

Thus $F_n(0) = 1/2 \not\to F(0) = 1$.

Example 2.3. Let $X_n = 1/n$ be a deterministic random variable. Then $X_n \to 0$ in distribution but unless h is continuous, we cannot conclude that $h(1/n) \to h(0)$.

As pointed out earlier, convergence in distribution is purely a property of the distributions of the respective random variables – it makes sense to say that $X_n \to X$ even if the random variables are defined on different probability spaces. However, if X_n and X all live on the same probability space $(\Omega, \mathcal{F}, P)$, then we have the following result:

Proposition 2.4. *$X_n \to X$ in probability implies $X_n \to X$ in distribution.*

Proof. Suppose $P(|X_n - X| > \epsilon) \to 0$ for all $\epsilon > 0$. Let h be bounded and continuous and let $M = \max_y |h(y)|$. Fix an arbitrary $\delta > 0$. Since h is continuous, we can choose ϵ so that $|h(x) - h(y)| < \delta$ whenever $|x - y| \leq \epsilon$. Next, we can choose n so large that $P(|X_n - X| > \epsilon) < \delta$. Then

$$\begin{aligned} |E(h(X_n) - h(X))| &\leq E(|h(X_n) - h(X)|) \\ &= E(|h(X_n) - h(X)|;\ |X_n - X| \leq \epsilon) \\ &\qquad + E(|h(X_n) - h(X)|;\ |X_n - X| > \epsilon) \\ &< \delta + M\delta. \end{aligned}$$

Since δ is arbitrary, this shows that $|E(h(X_n) - h(X))| \to 0$. ∎

The following corollary immediately follows from Proposition 2.1:

Corollary. *$X_n \to X$ a.s. implies $X_n \to X$ in distribution.*

3. Uniform integrability

Definition 3.1. Let $\mathcal{R}$ be a family of random variables. Then $\mathcal{R}$ is said to be a *uniformly integrable* family if for any given $\epsilon > 0$, there exists K such that

$$E(|X|;\ |X| > K) < \epsilon \quad \forall X \in \mathcal{R}.$$

Example 3.1. *An important example of a uniformly integrable family:* Let $X \in L^1$. Then the family of conditional expectations $\{E(X|\mathcal{G}) : \ \mathcal{G} \subset \mathcal{F}\}$ is uniformly integrable. Fix an arbitrary ϵ and choose δ so small that for any event $E \in \mathcal{F}$, $P(E) < \delta$ implies that $E(|X|;\ E) < \epsilon$. Next choose K so large that $E(|X|)/K < \delta$. For any σ-algebra $\mathcal{G} \subset \mathcal{F}$, let $Y = E(X|\mathcal{G})$. By Jensen's inequality

$$|Y| \leq E(|X| \big| \mathcal{G}). \tag{3.1}$$

Taking expectations above shows that $E(|Y|) \leq E(|X|)$. Moreover,

$$E(|Y|) \geq E(|Y|;\ |Y| > K) \geq KP(|Y| > K)$$

so we have $KP(|Y| > K) \leq E(|Y|) \leq E(|X|)$ and thus $P(|Y| > K) \leq E(|X|)/K < \delta$ which implies $E(|X|;\ |Y| > K) < \epsilon$. But since $\{|Y| > K\} \in \mathcal{G}$, by (3.1) and the definition of conditional expectation

$$E(|Y|;\ |Y| > K) \leq E(E(|X| \big| \mathcal{G});\ |Y| > K) = E(|X|;\ |Y| > K) < \epsilon.$$

Example 3.2. *An example of a family which is not uniformly integrable:* Let X_n take values 0 of n such that $P(X_n = n) = 1/n$ and $P(X_n = 0) = 1 - 1/n$. But for any $K > 0$, if $n > K$ we have

$$E(|X_n|;\ |X_n| > K) = nP(X_n = n) = 1,$$

so the family $(X_n :\ n \geq 0)$ is not uniformly integrable.

Below are two sufficient conditions for uniform integrability which are applicable in a wide variety of situations.

Proposition 3.1.

1. *If there exists a random variable* $0 \leq Y \in L^1$ *such that* $|X| \leq Y$ *(a.s.) for all* $X \in \mathcal{R}$*, then* $\mathcal{R}$ *is uniformly integrable;*
2. *If* $\mathcal{R}$ *is bounded in* L^p *for some* $p > 1$*, i.e. there is some* A *such that* $E(|X|^p) \leq A$ *for all* $X \in \mathcal{R}$*, then* $\mathcal{R}$ *is uniformly integrable.*

Proof.

Proof of (1): Since $E(Y) < \infty$, for any ϵ, we can choose K so large that $E(Y;\ Y > K) < \epsilon$. But for all $X \in \mathcal{R}$,

$$E(|X|;\ |X| > K) \leq E(Y;\ Y > K) < \epsilon.$$

Proof of (2): Recall that $p > 1$ so $1 - p < 0$.

$$E(|X|;\ |X| > K) = E(|X|^p |X|^{1-p};\ |X| > K) \leq K^{1-p} E(|X|^p; |X| > K) \leq K^{1-p} A.$$

To finish the proof, note that for given ϵ, we can choose K so large that $K^{1-p} < \epsilon/A$. □

Note that in Proposition 3.1(2), it is crucial that $p > 1$; as Example 3.2 shows, L^1-boundedness is not sufficient for uniform integrability.

Finally, note that the condition in Proposition 3.1(1) is identical to that in the dominated convergence theorem and in fact uniform integrability gives an improvement on dominated convergence. The proof of the following can be found in Williams 1991:

Theorem 3.2. *Let X_n and X be random variables in L^1. Then $E(|X_n - X|) \to 0$ if and only the following 2 conditions are satisfied;*

(i) $X_n \to X$ in probability and

(ii) (X_n) is a uniformly inregrable family.

In particular, since $|E(X_n) - E(X)| \leq E(|X_n - X|)$, if conditions (i) and (ii) hold, then $E(X_n) \to E(X)$.