pfsl9.tex
Lecture 9. 30.10.2013

We quote (see e.g. [GS], 7.2): if $1 \le p \le r$,
(a) $L_r \subset L_p$ [true in any finite measure space, but not in general];
(b) convergence in rth mean implies convergence in pth mean
[as expected: the higher the moment, the more restrictive the condition].

Neither of these two strong modes of convergence implies the other.

Definition. We say that $X_n \to X$ *in probability (in pr)* if for all $\epsilon > 0$,

$$P(|X_n - X| > \epsilon) \to 0 \qquad (n \to \infty).$$

This is a mode of convergence of intermediate strength. Each of the two strong modes above implies it, but not conversely.

There is a partial converse, due to F. Riesz: if $X_n \to X$ in probability, there is a subsequence along which $X_n \to X$ a.s.

Finally, we have a (very useful but) weak mode of convergence.

Definition. $X_n \to X$ *in distribution* if

$$E[f(X_n)] \to E[f(X)]$$

for all bounded continuous functions f.

The content of Lévy's convergence theorem for CFs (II.3 above) is that such behaviour on the particular functions

$$f(x) := e^{itx}$$

for t real suffices here.

As the names suggest, the intermediate mode convergence in probability implies the weak mode convergence in distribution, but not conversely.

Metrics and completeness.

Recall that a *metric* $d = d(.,.)$ is a distance function, generalising that in Euclidean space. Some but not all metrics are generated by *norms* $\|.\|$, again as in Euclidean space:

$$d(x,y) = \|x - y\|.$$

Recall a sequence $\{s_n\}$ is called *Cauchy* if it satisfies the *Cauchy condition*

$$\forall \epsilon > 0 \qquad \exists N \text{ s.t. } \forall m, n \ge N, \ \ |s_m - s_n| < \epsilon,$$

and *convergent to s* if

$$\forall \epsilon > 0 \qquad \exists N \ \text{ s.t. } \ \forall n \geq N, \ \ |s_n - s| < \epsilon.$$

Convergence implies the Cauchy condition (by the triangle inequality). The converse is *true* for the reals $\mathbb{R}$ (Cauchy's General Principle of Convergence), and similarly for the complex plane $\mathbb{C}$, but *false* for the rationals $\mathbb{Q}$. We call a metric space *complete* if every Cauchy sequence is convergent; thus $\mathbb{R}$, $\mathbb{C}$ are complete, but $\mathbb{Q}$ not.

Convergence in pth mean (or in L_p) is metric, and generated by the L_p-*norm*:

$$\|X\|_p := (E[|X|^p])^{1/p}.$$

By the *Riesz-Fischer theorem*, the L_p-spaces are complete.

Convergence in probability is also given by a metric:

$$d(X,Y) := E\Big(\frac{|X-Y|}{1+|X-Y|}\Big).$$

This metric is also complete.

Convergence in distribution is also generated by a metric, the *Lévy metric*:

$$d(F,G) := \inf\{\epsilon > 0 : F(x-\epsilon) - \epsilon \leq G(x) \leq F(x+\epsilon) + \epsilon \text{ for all } x\}$$

(the French probabilist Paul LÉVY (1886-1971) in 1937) (it is not obvious, but it is true, that d is a metric): if F_n, F are distribution functions,

$$F_n \to F \quad \text{in distribution} \quad \Leftrightarrow \quad d(F_n, F) \to 0.$$

This is also equivalent to

$$F_n(x) \to F(x) \qquad (n \to \infty) \qquad \text{at all continuity points } x \text{ of } F.$$

The restriction to continuity points x of F here is vital: take X_n, X as constants c_n, c with $c_n \to c$. We should clearly have $c_n \to c$ in distribution regarded as random variables; the distribution function F of c is 0 to the left of c and 1 at c and to the right; pointwise convergence takes place everywhere *except* c (the only interesting point here).

We quote that the Lévy metric is complete.

Egorov's theorem; almost uniform convergence. We quote (D. F. EGOROV (1869-1931) in 1911)

Egorov's theorem. If $X_n \to Z$ a.s., then for all $\epsilon > 0$ there exists a set of probability $< \epsilon$ off which $X_n \to X$ uniformly (in ω). This property is called *almost uniform convergence.* So Egorov's theorem states that almost sure and almost uniform convergence are equivalent.

It follows from this that almost sure convergence ('strong') implies convergence in probability ('weak'), as above.

Convergence in probability ('intermediate') implies convergence in distribution ('weak'). We quote this.

There is no converse, but there is a partial converse [which we shall use below]. If X_n converges in distribution to a *constant* c, then since the distribution function of the constant c is 0 to the left of c and 1 at c and to the right, it is easy to see that in fact $X_n \to c$ in probability.

Example. We show by example that convergence in pr does not imply a.s. convergence (a fact known to F. Riesz in 1912). On the *Lebesgue measure space* $[0,1]$ (i.e., $([0,1], \Lambda, \lambda)$, let

$$s_n := 1/2 + 1/3 + \ldots + 1/n \quad (mod\ 1), \quad A_n := [s_{n-1}, s_n], \quad X_n := I_{A_n}.$$

Since the harmonic series diverges, the X_n endlessly move rightwards through the interval $[0,1]$, exiting right and reappearing left. So the X_n do not converge anywhere, in particular are not a.s. convergent. But since $X_n = 0$ except on a set of probability $1/n$, $X_n \to 0$ in probability.

Three classical convergence theorems. We quote (see e.g. SP L6, 8):

M (Lebesgue's monotone convergence theorem). If $X_n \geq 0$, $X_n \uparrow X$, then $E[X_n] \uparrow E[X]$.

F (Fatou's lemma). If $X_n \geq 0$, then $E[\liminf X_n] \leq \liminf E[X_n]$.

D (Lebesgue's dominated convergence theorem). If $X_n \to X$ a.s., $|X_n| \leq Y$ with $E[Y] < \infty$, then $E[X_n] \to E[Y]$.

2. The Weak Law of Large Numbers (WLLN) and the Central Limit Theorem (CLT).

Recall that by Real Analysis,

$$(1 + \frac{x}{n})^n \to e^x \qquad (n \to \infty)$$

(this expresses compound interest, or exponential growth, as the limit of simple interest as the interest is compounded more and more often). This extends also to complex number z, and to $z_n \to z$:

$$(1 + \frac{z_n}{n})^n \to e^z \qquad (n \to \infty).$$

The next result is due to Lévy in 1925, and in more general form to the Russian probabilist A. Ya. KHINCHIN (1894-1956) in 1929 and to Kolmogorov in 1928/29.

Theorem (Weak Law of Large Numbers, WLLN). If X_i are iid with mean μ,

$$\frac{1}{n}\sum_1^n X_k \to \mu \qquad (n \to \infty) \qquad \text{in probability.}$$

Proof. If the X_k have CF $\phi(t)$, then as the mean μ exists $\phi(t) = 1 + i\mu t + o(t)$ as $t \to 0$. So $(X_1 + \ldots + X_n)/n$ has CF

$$E\exp\{it(X_1 + \ldots + X_n)/n\} = [\phi(t/n)]^n = [1 + \frac{i\mu t}{n} + o(1/n)]^n,$$

for fixed t and $n \to \infty$. By above, the RHS has limit $e^{i\mu t}$ as $n \to \infty$. But $e^{i\mu t}$ is the CF of the constant μ. So by Lévy's continuity theorem,

$$(X_1 + \ldots + X_n)/n \to \mu \qquad (n \to \infty) \qquad \text{in distribution.}$$

Since the limit μ is constant, by II.4 (L11), this gives

$$(X_1 + \ldots + X_n)/n \to \mu \qquad (n \to \infty) \qquad \text{in probability. //}$$

As the name implies, the Weak LLN can be strengthened, to the Strong LLN (with a.s. convergence in place of convergence in probability). We turn to this later, but proceed with a refinement of the method above, in which we retain one more term in the Taylor expansion of the CF. Recall first that the CF of the standard normal distribution $\Phi = N(0,1)$, with density $\phi(x)$ and distribution function $\Phi(x)$

$$\phi(x) := \frac{e^{-x^2/2}}{\sqrt{2\pi}}, \qquad \Phi(x) := \int_\infty^x \phi(u)du$$

is $e^{-t^2/2}$.

Theorem (Central Limit Theorem, CLT). If $X_1, \ldots X_n, \ldots$ are iid with mean μ and variance σ^2, and $S_n := X_1 + \ldots + X_n$, then

$$(S_n - n\mu)/(\sigma\sqrt{n}) \to \Phi = N(0,1) \qquad (n \to \infty) \qquad \text{in distribution.}$$