spl11.tex

Lecture 11. 1.11.2010.

Note on probability spaces.

The example of L10 showing that convergence in pr does not imply a.s. convergence depends on the existence of non-trivial (Lebesgue-)null sets. In a purely atomic measure space, there are no non-trivial null sets, examples such as this cannot be constructed, and convergence in pr and a.s. become the same. This is a rare case when we need to mention the probability space explicitly; usually we do not need to. In practice, we shall always have a probability space at least as rich as the Lebesgue probability space above, and so will not meet such examples.

4. Modes of convergence.

Recall (L9) that when all random variables are finite-valued, a.s. convergence is the same as almost uniform convergence.

Convergence in L_p is given by a norm, so also by a metric. By the Riesz-Fischer theorem (L6), L_p is complete: if $\|f_m - f_n\|_p \to 0$ as $m, n \to \infty$, then there is some $f \in L_p$ such that $f_n \to f$ in L_p.

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.

Given any sequence X_n converging in pr, there exists some subsequence converging a.s. (this also is due to F. Riesz in 1912). We quote this. Likewise, any sequence X_n converging in pth mean has an a.s. convergent subsequence.

Convergence in distribution.

We turn now to a weaker mode of convergence, which deals not with values of the random variables as above but with their distributions. If X_n, X are random variables with distribution functions F_n, F, we say that $X_n \to X$ *in distribution* (or *in law*),

$$X_n \to X \quad \text{in distribution,} \quad \text{or} \quad F_n \to F \quad \text{in distribution,}$$

if

$$Ef(X_n) \to Ef(X) \quad (n \to \infty) \quad \text{for all bounded continous functions } f,$$

equivalently, if

$$\int f(x)dF_n(x) \to \int f(x)dF(x) \qquad (n \to \infty)$$

for all such f. This mode of convergence is is also generated by a metric, the *Lévy metric*:

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

(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.)

We quote that the Lévy metric is complete.

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

There is no converse, but there is a partial converse. 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.

5. Characteristic functions.

If X has distribution function F, the *characteristic function* (CF) of X is

$$\phi(t) := Ee^{itX} = \int_{-\infty}^{\infty} e^{itx} dF(x) \qquad (t \in \mathbf{R}).$$

This is also the *Fourier-Stieltjes transform* of F ('Fourier transform, Stieltjes integral').

The CF has a number of important properties.

1. *Existence.* The CF always exists (the integral defining it always converges). Indeed,

$$|\phi(t)| = |\int e^{itx} dF(x)| \leq \int |e^{itx}| dF(x) = \int 1 dF(x) = 1.$$

2. *Continuity.* The CF is continuous, indeed uniformly continuous:

$$|\phi(t+u) - \phi(t)| = |\int e^{itx}(e^{itu} - 1) dF(x)| \leq \int |e^{itu} - 1|.1 dF(x) \to 0$$

as $t \to 0$, by dominated convergence.
3. *Uniqueness.* The CF determines the distribution function uniquely (so taking the CF loses no information). This is a general property of Fourier transforms; we quote this.
4. *Inversion formula.* There is an inversion formula (due to Lévy, 1937) giving the distribution function in terms of the CF. We omit this, as the formula is rarely useful.
5. *Continuity theorem* (Lévy, 1937). (i) If F_n, F have CFs ϕ_n, ϕ, and $F_n \to F$ in distribution, then

$$\phi_n(t) \to \phi(t) \qquad (n \to \infty) \qquad \text{uniformly in } t \text{ on compact sets.}$$

(ii) Conversely, if $\phi_n(t) \to \phi(t)$ pointwise, and the limit function $\phi(t)$ is continuous at $t = 0$, then ϕ is the CF of a distribution function, F say, and $F_n \to F$ in distribution.
6. *Moments.* For a random variable X, the kth *moment* of X is defined by

$$\mu_k := E[X^k].$$

The first moment is the *mean* or *expectation*, $\mu = E[X]$. (We use notation such as μ_X if there are other random variables present. Context will show whether μ denotes a mean or a measure.) If X has k moments (finite), we can expand the exponential e^{itX} in the definition of the CF and get $\sum_{j=0}^{k} (it)^j.E[X^j]/j!$ or $\sum_{j=0}^{k} (it)^j \mu_j/j!$, plus an error term. Analogy with Taylor's Theorem in Real Analysis suggests that this error term should be $o(t^k)$ at $t \to 0$. This is true; we quote it: if X has k moments finite, its CF satisfies

$$\phi(t) = \sum_{j=0}^{k} (it)^j \mu_j/j! + o(t^k) \qquad (t \to 0).$$

6. Independence.

Recall from your first course in Probability that events A, B are called independent if

$$P(A \cap B) = P(A).P(B)$$

(independence corresponds to product measures: see L9). Since $P(A) = EI_A$, this says

$$E[I_A.I_B] = E[I_A].E[I_B].$$

We generalize this. A family of events is *independent* if for any finite subfamily A_k $(k = 1, \ldots, n)$, the probability of the intersection of any subfamily is the product of the probabilities. A family of random variables is *independent* if, for any finite subfamily $\{X_k\}$ $(k = 1, \ldots, n)$ and any x_k, the events $\{X_k \leq x_k\}$ are independent; equivalently, the events $\{X_k \in A_k\}$ are independent for all measurable A_k.

Theorem (Multiplication Theorem). If $X_1, \ldots, X_n$ are independent and $g_1, \ldots, g_n$ are measurable,
(i) $g_1(X_1), \ldots, g_n(X_n)$ are independent;
(ii) If the g_k are bounded,

$$E[\prod_{k=1}^{n} g_k(X_k)] = \prod_{k=1}^{n} Eg_k(X_k).$$

Proof. (i)

$$P(g_k(X_k) \in A_k, k = 1, \ldots, n) = P(X_k \in g_k^{-1}(A_k), k = 1, \ldots, n)$$

$$= \prod_{k=1}^{n} P(X_k \in g_k^{-1}(A_k)) = \prod_{k=1}^{n} P(g_k(X_k) \in A_k),$$

proving independence of the $g_k(X_k)$.
(ii) For simple g_k, $g_k = \sum c_{k,i_k} I_{A_{k,i_k}}$,

$$E[\prod_{i=1}^{n} g_i(X_i)] = E[\prod_{k=1}^{n} \sum c_{k,i_k} I_{A_{k,i_k}}(X_k)].$$

By independence, on the RHS $E[\prod I] = E[I(\cap)] = P(\cap) = \prod P(.) = \prod E[I]$. The RHS thus factorizes, giving the result for simple g_k. The result extends to the general case by approximation. //