spl9.tex
Lecture 9. 28.10.2011.

7. Further results.

Doob's lemma

For real-valued measurable functions $f : (\Omega, \mathcal{A}) \to (\mathbf{R}, \mathcal{B})$, we write

$$\sigma(f) := f^{-1}(\mathcal{B}) := \{f^{-1}(B) : B \in \mathcal{B}\},$$

and call this the σ-field *generated* by f.

The next result is due to the American probabilist J. L. Doob (1910-2004) (in the Supplement on Measure Theory to his 1953 book *Stochastic Processes*).

Doob's lemma. For f, g real-valued measurable functions, the following are equivalent:
(i) $\sigma(f) \subset \sigma(g)$;
(ii) $f = h(g)$ for some measurable function h.

Proof. If (ii) holds and $f = h(g)$,

$$\sigma(f) = f^{-1}(\mathcal{B}) = (h(g))^{-1}(\mathcal{B}) = g^{-1}(h^{-1}(\mathcal{B})) \subset g^{-1}(\mathcal{B}) = \sigma(g),$$

as h is measurable, giving (i). Conversely, if (i) holds, consider first the case of f an indicator, $f = I_A$, $A \in \mathcal{A}$. As $\sigma(f) \subset \sigma(g)$, $A \in \sigma(g)$, so $A = g^{-1}(B)$ for some Borel set $B \in \mathcal{B}$. Then

$$f = I_A = I_{B(g)} = h(g)$$

with $h = I_B$. So the result holds for indicators f. It extends to simple f by linearity, and to general f by approximation, as before. //

Observe that when we take a function h of something, as here, we in general *lose information* (e.g. when $h(x) = x^2$, we lose the sign: from x^2 we can recover only $\pm x$). When h is one-to-one (surjective) and so the inverse function h^{-1} exists, we can go the other way, and so no information is lost, but not in general. This gives us the interpretation of Doob's lemma: $\sigma(f)$ represents the *information* contained in f. We shall make extensive use of this in later chapters.

Inequalities.

In addition to Minkowski's inequality (L6 – 'the L_p-norm is a norm'), there are other inequalities for integrals that we shall need. First, for an index $p \geq 1$, we define its *conjugate* index q by

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Thus also

$$p = \frac{q}{q-1}, \qquad q = \frac{p}{p-1};$$

$q \geq 1$. If $p = 1$, $q = \infty$; if $p = \infty$, $q = 1$; $p = q$ iff $p = q = 2$. We have Hölder's inequality (Otto HÖLDER (1859-1937) in 1884): for conjugate indices $p, q > 1$ and $f \in L_p$, $g \in L_q$, $fg \in L_1$ and

$$|\int fg| \leq (\int |f|^p)^{1/p} (\int |g|^q)^{1/q} : \quad \|fg\|_1 \leq \|f\|_p . \|g\|_q.$$

We note in particular the case $p = q = 2$, the *Cauchy-Schwarz inequality* (A. L. CAUCHY (1789 - 1857) in 1821, H. A. SCHWARZ (1843 - 1921) in 1885):

$$|\int fg| \leq (\int |f|^2)^{1/2} (\int |g|^2)^{1/2} : \quad \|fg\|_1 \leq \|f\|_2 . \|g\|_2.$$

Call a function f *convex* if for all $\lambda \in [0, 1]$,

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$

(geometric interpretation: 'graph lies above chord', or 'graph bends upwards'). Then (Jensen's inequality: J. L. W. V. JENSEN (1859-1925) in 1906): for ϕ convex and $f \geq$ measurable,

$$\int \phi(f) d\mu / \int d\mu \leq \phi(\int f d\mu / \int d\mu).$$

For proofs, see e.g. [S] Ch. 12, and Problems/Solutions 3.

Modes of convergence.

We need several modes in which a sequence of measurable functions f_n can converge. The most obvious is *pointwise convergence* ($f_n(x) \to f(x)$ for all points x), but bearing in mind that we deal with functions only up to sets of measure 0, this is too stringent, and needs to be qualified to *convergence almost everywhere* (a.e.):

$$f_n(x) \to f(x) \qquad (n \to \infty) \qquad a.e.$$

The second is convergence in pth mean ($p \geq 1$):

$$f_n \to f \quad \text{in } p\text{th mean, or in } L_p, \text{ means} \quad \|f_n - f\|_p \to 0 \quad (n \to \infty).$$

The third is it convergence in measure: $f_n \to f$ *in measure* (w.r.t. μ) means that for all $\epsilon > 0$,

$$\mu\{x : |f_n(x) - f(x)| > \epsilon\} \to 0 \quad (n \to \infty).$$

We think of convergence a.e. and in pth mean as *strong* modes of convergence. They are not comparable: neither implies the other. We think of convergence in measure as an *intermediate* mode of convergence: it is implied by each of the first two, but not conversely. Later (Ch. II) we will meet a *weak* mode of convergence (convergence *in distribution*), which is implied by convergence in measure, but not conversely.

Littlewood's three principles.

The British analyst J. E. LITTLEWOOD (1885-1977) formulated in 1944 three heuristic principles:
(i) every [measurable] set is nearly a finite union of rectangles;
(ii) every [measurable] function is nearly continuous;
(iii) every convergent sequence of [measurable] functions is nearly uniformly continuous.

Littlewood's first principle is made precise by saying that Lebesgue measure is *regular*: a measurable set A can be approximated from without by an $\mathcal{O}_\delta$ set of the same measure and from within by an $\mathcal{F}_\sigma$ set of the same measure (and for each $\epsilon > 0$, from without by an open set and from within by a closed set to within measure ϵ). Regular measures are important more generally, and connect Measure Theory and Topology.

Littlewood's second principle is expressed by *Luzin's theorem* (N. N. LUZIN (or Lusin) (1883-1950) in 1912): if $f : [a, b] \to \mathbf{R}$ is measurable, for each $\epsilon > 0$ there is a continuous function that coincides with f off a set of measure $< \epsilon$.

Littlewood's third principle is expressed by *Egorov's theorem* (D. F. EGOROV (1869-1931) in 1911): if measurable $f_n \to f$ on a set of finite measure, with f_n, f finite-valued, then for each $\epsilon > 0$ there is a set of measure $< \epsilon$ off which $f_n \to f$ uniformly. We then say that $f_n \to f$ *almost uniformly*.

Egorov's theorem extends to a.e. convergence, and then the converse is also true (and easy to prove). So for finite measure spaces and finite-valued functions converging to a finite-valued limit, a.e. convergence is the same as almost uniform convergence.

Product measures and Fubini's theorem.

Given two measure spaces $(\Omega_i, \mathcal{S}_i, \mu_i)$, one can form the cartesian product $\Omega := \Omega_1 \times \Omega_2$, the σ-algebra $\mathcal{A}$ generated by the sets $A_1 \times A_2$ for $A_i \in \mathcal{A}_i$, and the measure μ defined by $\mu(A_1 \times A_2) := \mu_1(A_1) \times \mu_2(A_2)$ extended to $\mathcal{A}$. This is called the *product* measure space, and μ is the *product measure*. We shall see in II.6 (L11) that this is relevant to *independence* of random variables. One can extend to finite products by induction, and also to infinite products (S. KAKUTANI (1911-2004) in 1943). Again, this is relevant to infinite replication of experiments (in the simplest case, coin-tossing).

Fubini's theorem (Guido FUBINI (1879-1943) in 1907) concerns double and repeated integrals. If $f(x, y)$ is a function of two variables, let f_x be the function of y with x held constant, and similarly for f_y. Then if $f \in L_1(\mu)$,

$$\int \int_{\Omega} f d\mu = \int_{\Omega_1} (\int_{\Omega_2} f_x d\mu_2) d\mu_1 = \int_{\Omega_2} (\int_{\Omega_1} f_y d\mu_1) d\mu_2$$

This results from the integral being *absolute*: for a non-absolute integral such as the Riemann integral, complications regarding cancellation can arise.

The Daniell integral.

We began with measure and turned to to integration – in particular, we went from Lebesgue measure to Lebesgue integration. We point out that one can instead go in the reverse direction. This was done by P. J. DANIELL (1889-1946) in 1918. Here, the integral is defined by its properties (linear, order-preserving, continuous – 'integral of limit = limit of integral', as in the Lebesgue convergence theorems). The integral is treated as the primary concept, and measure and its properties are deduced from there. Compare the reversal in Calculus: we learn Differential Calculus first and then Integral Calculus, but this is the reverse of the historical order, by two thousand years.