m3f22l8.tex
Lecture 8. 21.10.2016

L_p spaces.

For $p \geq 1$, the L_p spaces $L_p(\mathbb{R}^k)$ on $\mathbb{R}^k$ are the spaces of measurable functions f with L_p-norm

$$\|f\|_p := \left(\int |f|^p\right)^{\frac{1}{p}} < \infty.$$

Riemann integrals.

Our first exposure to integration is the 'Sixth-Form integral', taught non-rigorously at school. Mathematics undergraduates are taught a rigorous integral (in their first or second years), the *Riemann integral* [G.B. RIEMANN (1826-1866)] – essentially this is just a rigourization of the school integral. It is much easier to set up than the Lebesgue integral, but much harder to manipulate.

For finite intervals $[a, b]$,we quote:
(i) for any function f Riemann-integrable on $[a, b]$, it is Lebesgue-integrable to the same value (but many more functions are Lebesgue integrable);
(ii) f is Riemann-integrable on $[a, b]$ iff it is continuous a.e. on $[a, b]$. Thus the question, "Which functions are Riemann-integrable?" cannot be answered without the language of measure theory – which then gives one the technically superior Lebesgue integral anyway.
Note. Integration is like summation (which is why Leibniz gave us the integral sign $\int$, as an elongated S). Lebesgue was a very practical man – his father was a tradesman – and used to think about integration in the following way. Think of a shopkeeper totalling up his day's takings. The Riemann integral is like adding up the takings – notes and coins – *in the order in which they arrived.* By contrast, the Lebesgue integral is like totalling up the takings *in order of size* - from the smallest coins up to the largest notes. This is obviously better! In mathematical effect, it exchanges 'integrating by x-values' (abscissae) with 'integrating by y-values' (ordinates).

Lebesgue-Stieltjes integral.

Suppose that $F(x)$ is a *non-decreasing* function on $\mathbb{R}$:

$$F(x) \leq F(x) \qquad \text{if } x \leq y$$

(prime example: F a probability distribution function). Such functions can have at most countably many discontinuities, which are at worst jumps. We may without loss re-define F at jumps so as to be *right-continuous*.

We now generalise the starting points above:
(i) *Measure.* We take $\mu((a, b]) := F(b) - F(a)$.
(ii) *Integral.* We take $\int_a^b 1 := F(b) - F(a)$.
We may now follow through the successive extension procedures used above. We obtain:
(i) *Lebesgue-Stieltjes measure* μ, or μ_F,
(ii) *Lebesgue-Stieltjes integral* $\int f \, d\mu$, or $\int f \, d\mu_F$, or even $\int f \, dF$.
Similarly in higher dimensions; we omit further details.
Finite variation (FV).

If instead of being monotone non-decreasing, F is the *difference* of two such functions, $F = F_1 - F_2$, we can define the integrals $\int f \, dF_1$, $\int f \, dF_2$ as above, and then define

$$\int f \, dF = \int f \, d(F_1 - F_2) := \int f \, dF_1 - \int f \, dF_2.$$

If $[a, b]$ is a finite interval and F is defined on $[a, b]$, a finite collection of points, $x_0, x_1, \ldots, x_n$ with $a = x_0 < x_1 < \cdots < x_n = b$, is called a *partition* of $[a, b]$, $\mathcal{P}$ say. The sum $\sum_{i=1}^n |F(x_i) - F(x_{i-1})|$ is called the *variation* of F over the partition. The least upper bound of this over all partitions $\mathcal{P}$ is called the *variation* of F over the interval $[a, b]$, $V_a^b(F)$:

$$V_a^b(F) := \sup_{\mathcal{P}} \sum |F(x_i) - F(x_{i-1})|.$$

This may be $+\infty$; but if $V_a^b(F) < \infty$, F is said to be of *finite variation (FV)* on $[a, b]$, $F \in FV_a^b$ (*bounded variation*, BV, is also used). If F is of finite variation on all finite intervals, F is said to be *locally of finite variation*, $F \in FV_{loc}$; if F is of finite variation on the real line, F is of *finite variation*, $F \in FV$.

We quote (*Jordan's theorem*) that the following are equivalent:
(i) F is locally of finite variation;
(ii) F can be written as the difference $F = F_1 - F_2$ of two monotone functions.
So the above procedure defines the integral $\int f \, dF$ when the *integrator* F is of *finite variation*.

3 Probability.

Probability spaces.

The mathematical theory of probability can be traced to 1654, to correspondence between PASCAL (1623-1662) and FERMAT (1601-1665). However, the theory remained both incomplete and non-rigorous till the 20th century. It turns out that the Lebesgue theory of measure and integral sketched above is exactly the machinery needed to construct a rigorous theory of probability adequate for modelling reality (option pricing, etc.) for us. This was realised by the great Russian mathematician and probabilist A.N.KOLMOGOROV (1903-1987), whose classic book of 1933, *Grundbegriffe der Wahrscheinlichkeitsrechnung* [Foundations of probability theory] inaugurated the modern era in probability.

Recall from your first course on probability that, to describe a random experiment mathematically, we begin with the *sample space* Ω, the set of all possible outcomes. Each point ω of Ω, or *sample point*, represents a possible – random – outcome of performing the random experiment. For a set $A \subseteq \Omega$ of points ω we want to know the probability $P(A)$ (or $\Pr(A), \mathrm{pr}(A)$). We clearly want

1. $P(\emptyset) = 0,\ P(\Omega) = 1$.

2. $P(A) \geq 0$ for all A.

3. If $A_1, A_2, \ldots, A_n$ are disjoint, $P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$
(finite additivity – fa), which, as above we will strengthen to

3*. If $A_1, A_2 \ldots$ (*ad inf.*) are disjoint,

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \quad \text{(countable additivity – ca).}$$

4. If $B \subseteq A$ and $P(A) = 0$, then $P(B) = 0$ (completeness).

Then by 1 and 3 (with $A = A_1,\ \Omega \setminus A = A_2$),

$$P(A^c) = P(\Omega \setminus A) = 1 - P(A).$$

So the class $\mathcal{F}$ of subsets of Ω whose probabilities $P(A)$ are defined should be closed under countable, disjoint unions and complements, and contain the empty set $\emptyset$ and the whole space Ω. Such a class is called a σ-*field* of subsets of Ω [or sometimes a σ-algebra, which one would write $\mathcal{A}$]. For each $A \in \mathcal{F}$, $P(A)$ should be defined (and satisfy $1, 2, 3*, 4$ above). So, $P : \mathcal{F} \to [0, 1]$ is a set-function,

$$P : A \mapsto P(A) \in [0, 1] \quad (A \in \mathcal{F}).$$

The sets $A \in \mathcal{F}$ are called *events*. Finally, 4 says that all subsets of null-sets (events) with probability zero (we will call the empty set $\emptyset$ empty, not null) should be null-sets (completeness). A *probability space*, or *Kolmogorov triple*, is a triple $(\Omega, \mathcal{F}, P)$ satisfying these *Kolmogorov axioms* 1,2,3*,4 above. A probability space is a mathematical model of a random experiment.
Random variables.

Next, recall random variables X from your first probability course. Given a random outcome ω, you can calculate the value $X(\omega)$ of X (a scalar – a real number, say; similarly for vector-valued random variables, or random vectors). So, X is a function from Ω to $\mathbb{R}$, $X \to \mathbb{R}$,

$$X : \omega \mapsto X(\omega) \quad (\omega \in \Omega).$$

Recall also that the *distribution function* of X is defined by

$$F(x), \quad \text{or} \quad F_X(x), \quad := P\Big(\{\omega : X(\omega) \le x\}\Big), \quad \text{or} \quad P(X \le x), \qquad (x \in \mathbb{R}).$$

We can only deal with functions X for which all these probabilities are defined. So, for each x, we need $\{\omega : \ X(\omega) \le x\} \in \mathcal{F}$. We summarize this by saying that X is *measurable* with respect to the σ-field $\mathcal{F}$ (of events), briefly, X is *$\mathcal{F}$-measurable*. Then, X is called a *random variable* [non-$\mathcal{F}$-measurable X cannot be handled, and so are left out]. So,
(i) a random variable X is an $\mathcal{F}$-measurable function on Ω;
(ii) a function on Ω is a random variable (is measurable) iff its distribution function is defined.
Generated σ-fields.

The smallest σ-field containing all the sets $\{\omega : X(\omega) \le x\}$ for all real x [equivalently, $\{X < x\}$, $\{X \ge x\}$, $\{X > x\}$][1] is called the σ-field *generated* by X, written $\sigma(X)$. Thus,

$$X \text{ is } \mathcal{F}\text{-measurable [is a random variable] iff } \sigma(X) \subseteq \mathcal{F}.$$

When the (random) value $X(\omega)$ is *known*, we know *which* of the events in the σ-field generated by X have happened: these are the events $\{\omega : X(\omega) \in B\}$, where B runs through the Borel σ-field [the σ-field generated by the intervals – it makes no difference whether open, closed etc.] on the line.

[1]Here, and in Measure Theory, whether intervals are open, closed or half-open doesn't matter. In Topology, such distinctions are crucial. One can combine Topology and Measure Theory, but we must leave this here.