pfsl2.tex

Lecture 2. 10.10.2012

Definition. A *σ-field* (or σ-algebra) $\mathcal{A}$ is a class containing the whole set, closed under complements, and closed under countable disjoint unions (the "σ" here is from the German Summe = sum – the old-fashioned notation for a union is a sum).

The natural domain of definition of a measure is a σ-field:

Definition. A *measurable space* is a pair $(\Omega, \mathcal{A})$, where $\mathcal{A}$ is a σ-field of sets $A \subset \Omega$.

A *measure space* is a triple $(\Omega, \mathcal{A}, \mu)$, where μ is a measure defined on $\mathcal{A}$ (that is, $\mu(A)$ is defined on all the sets $A \subset \Omega$).

A *probability measure* is a measure P of mass 1, $P(\Omega) = 1$; then $(\Omega, \mathcal{A}, P)$ is a *probability space*.

Axiomatic Probability Theory as Measure Theory for measures of mass 1 is due to A. N. KOLMOGOROV (1903-87) in his 1933 book *Grundbegriffe der Wahrscheinlichkeitsrechnung*.

Examples. On the real line $\mathbf{R}$, the intervals $I; [a, b]$ are (Lebesgue) measurable, with (Lebesgue) measure

$$\mu([a, b]) := b - a. \qquad (L)$$

The σ-field generated by the intervals (= smallest σ-field containing the intervals, = intersection of all σ-fields containing the intervals – this is a σ-field) is called the *Borel* σ-field $\mathcal{B}$; its sets are called the *Borel sets* B (Emile BOREL (1871-1956, thesis of 1893). One can check that it does not matter whether we use closed intervals $[a, b]$, open ones (a, b), half-open ones $(a, b]$, $[a, b)$, semi-infinite intervals $(-\infty, a]$, etc. – they all generate the same σ-field.

A subset of a Borel set of measure 0 need not be a Borel set. Nevertheless, one feels that "a subset of a set of measure 0 should also have measure 0" – or, as we call sets of measure 0 *null sets*, "a subset of a nullset should also be a null set". It turns out that this is true for the σ-field generated by the *intervals and the null sets* together. These are called the *Lebesgue measurable sets*, $\mathcal{L}$. This process of including all subsets of null sets as null sets always works, and is called *completion.* Thus $\mathcal{L}$ is the completion of $\mathcal{B}$.

The measure μ obtained on $\mathcal{L}$ from (L) in this way is called *Lebesgue measure*; $\mathcal{L}$ is the natural domain of definition of μ.

Of course, the real line $\mathbf{R}$ has infinite Lebesgue measure (= length). But, it often suffices in Analysis, and even more in Probability, so work with the unit interval $[0,1]$. Then $([0,1],\mathcal{L},\lambda)$, where $\mathcal{L}$ here denotes the Lebesgue-measurable subsets of $[0,1]$ and μ Lebesgue measure on them, is called the *Lebesgue probability space* (see below).

Measurable functions; integrals. If f is a function from a measurable space $(\Omega,\mathcal{A})$ to the reals $(\mathbf{R},\mathcal{B})$, one calls f *measurable* if

$$f^{-1}(B) \in \mathcal{A} \text{ for all } B \in \mathcal{B}$$

– that is, inverse images of Borel sets are measurable.

These are the 'nice' functions, and we may restrict ourselves to them.

A (measurable) function of the form

$$f = \sum_{i=1}^{n} c_i I_{(a_i,b_i]}$$

is called a *simple function.* We can define the *integral* $\int f d\mu$ of a simple function with respect to the measure μ by

$$\int f d\mu := \sum_{i=1}^{n} c_i \mu((a_i,b_i])$$

when this is finite; we then say that f is μ-integrable, and write $f \in L_1(\mu)$ (L for Lebesgue; 1 for the first power, f). When it is $+\infty$, $\int f d\mu$ is undefined and f is not μ-integrable.

It turns out that a non-negative measurable function f is always the increasing limit of simple functions f_n, and that

$$\int f d\mu := \lim_{n\to\infty} \int f_n d\mu$$

defines $\int f d\mu$ uniquely (there are many such increasing sequences f_n, but they all give the same limit above).

Writing

$$x_+ := \max(x,0), \qquad x_- := -\min(x,0)$$

for the *positive part* and *negative part* of x, we may check that

$$|x| = x_+ + x_-, \qquad x = x_+ - x_-.$$

We can extend the definition above from non-negative measurable functions to general measurable functions by linearity:

$$\int f d\mu := \int f_+ d\mu - \int f_- d\mu.$$

Of course, this only holds when both integrals on the right are defined (are finite). So then

$$\int |f| d\mu = \int f_+ d\mu + \int f_- d\mu.$$

Thus f is μ-integrable iff $|f|$ is: the (measure-theoretic) integral here is an *absolute* integral, as we saw before. Also, the integral is easily seen to be *linear*: if $f, g \in L_1(\mu)$ and a, b are constants, then $af + bg \in L_1(\mu)$ and

$$\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu.$$

As one might suspect from the definition above, one can change the values of f on a μ-null set without changing the value of $\int f d\mu$. So: we are really dealing here with, not individual functions f themselves, but *equivalence classes*, under the equivalence relation

$$f \equiv g \qquad \text{iff} \qquad f = g \qquad \mu - a.e.,$$

where 'μ-a.e.' ('μ-almost everywhere' means 'except on a μ-null set').

For us, our (positive) measure (or integrator) μ, a set-function, will be obtained from a (non-decreasing) point function (which to save letters we also write μ), vanishing at some reference point x_0, by

$$\mu((a, b]) = \mu(b) - \mu(a). \qquad (LS)$$

The LS here is for *Lebesgue-Stieltjes* (the μ on the left is a LS *measure*, that on the right is a LS *measure function*). Thus for Lebesgue measure $\mu(x) \equiv x$ and $x_0 = 0$; for probability measures P, the point function is the distribution function (below), and $x_0 = -\infty$.

Random variables. When the measure space is a probability space $(\Omega, \mathcal{A}, P)$, we call the sets $A \in \mathcal{A}$ *events*. These are the sets A whose probabilities $P(A)$ are defined (this is consistent, both with ordinary speech and with usage in one's first exposure to Probability). We call a measurable function a *random variable*. In this case, we will use notation such as X, Y etc. rather than f, g etc. We call $\int_\Omega X dP$ the *expectation* of X, $E[X]$:

$$E[X] := \int_\Omega X dP.$$

By above, the expectation is *linear*:

$$E[aX + bY] = aE[X] + bE[Y].$$

Note. We need an absolute integral, as here, to get linearity of expectation. Without the restriction that $E[X]$ exists iff $E[|X|]$ exists, linearity of the expectation may fail. Recall from Analysis: *absolutely* convergent sums, integrals etc. may be rearranged at will. *Conditionally* convergent sums, integrals etc. are very dangerous: they result from 'cancelling infinities'. Note also that $a + b$ makes sense, not just for real numbers a and b, but for one or both of a or b $+\infty$ (then their sum is also $+\infty$); similarly for $-\infty$. But we must avoid the meaningless symbol "$\infty - \infty$". In much the same way, we must avoid the meaningless "$0/0$", as we know from Calculus.
Distribution functions. If X is a random variable (measurable function), the inverse image $X^{-1}(B) \in \mathcal{A}$ for all Borel sets B – equivalently, this holds for all B in some set that generates the Borel σ-field $\mathcal{B}$. The half-lines $(-\infty, x]$ $(x \in \mathbf{R})$ form such a set. So X is a random variable (rv) iff $X^{-1}((-\infty, x]) \in \mathcal{A}$ for each x, that is, $\{X \le x\} \in \mathcal{A}$ (is an event), that is, iff

$$F(x) := P(\{X \le x\})$$

is defined. Now the function F here (or F_X, if we need to distinguish between F_X and F_Y say) is called the (probability) *distribution function* (or just distribution, or d/n fn) is defined: *X is a random variable iff its distribution function is defined.*
Densities. If for some function $f \ge 0$ one has

$$F(x) := P(\{X \le x\}) = \int_{-\infty}^{x} f(u)du \qquad (x \in \mathbf{R}),$$

one calls f the (probability) *density* (function) of F, or X. Call this the *density case*, and such F *absolutely continuous* (SP L7). Then $f \ge 0$ corresponds to F non-decreasing. Then $F'(x) = f(x)$, but only a.e. (SP L7).
Example: the uniform distribution $U[0,1]$. On the Lebesgue probability space, U is a uniformly distributed random variable:

$$P(U \in (a, b]) = b - a \qquad (0 \le a \le b \le 1)$$

('probability = length'). This has distribution and density functions

$$F(x) = 0 \quad (x \le 0), \quad x \quad (0 \le x \le 1), \quad 1 \quad (x \ge 1); \quad f(x) = I_{[0,1]}(x).$$

Here F fails to be differentiable at the end-points 0 and 1 of the support interval $[0, 1]$ – but this exceptional set is of measure 0.