pfsl1.tex

Lecture 1. 9.10.2013

I. AXIOMATIC PROBABILITY THEORY

1. Length, area and volume. We shall mainly deal with area, as this is two-dimensional. We can draw pictures in two dimensions, and our senses respond to this; paper, whiteboards and computer screens are two-dimensional. By contrast, one-dimensional pictures are much less vivid, while three-dimensional ones are harder (they need the mathematics of perspective) – and dimensions higher than four are harder still.

Area.

1. *Rectangles*, base b, height h: area $A := bh$.
2. *Triangles*. $A = \frac{1}{2}bh$.

Proof: Drop a perpendicular from vertex to base; then extend each of the two triangles formed to a rectangle and use 1. above.

3. *Polygons*. Triangulate: choose a point in the interior and connect it to the vertices. This reduces the area A to the sum of areas of triangles; use 2. above.
4. *Circles*. We have a choice:

(a) Without calculus. Decompose the circle into a large number of equi-angular sectors. Each is approximately a triangle; use 2. above [the approximation boils down to $\sin\theta \sim \theta$ for θ small].

(b) With calculus and plane polar coordinates. Use $dA = dr.rd\theta = rdrd\theta$: $A = \int_0^r \int_0^{2\pi} rdrd\theta = \int_0^r rdr. \int_0^{2\pi} d\theta = \frac{1}{2}r^2.2\pi = \pi r^2$.

Note. The ancient Greeks essentially knew integral calculus – they could do this, and harder similar calculations [volume of a sphere $V = \frac{4}{3}\pi r^3$; surface area of a sphere $S = 4\pi r^2 dr$, etc.; note $dV = Sdr$].

What the ancient Greeks did not have is differential calculus [which we all learned first!] Had they had this, they would have had the idea of velocity, and differentiating again, acceleration. With this, they might well have got Newton's Law of Motion, Force = mass × acceleration. This triggered the Scientific Revolution. Had this happened in antiquity, the world would have been spared the Dark Ages and world history would have been completely different!

5. *Ellipses*, semi-axes a, b. Area $A = \pi ab$ (w.l.o.g., $a > b$).

Proof: cartesian coordinates: $dA = dx.dy$.

Reduce to the circle case: compress ['squash'] the x-axis in the ratio b/a [so $dx \mapsto dx.b/a$, $dA \mapsto dA.b/a$]. Now the area is $A = \pi b^2$, by 4. above. Now 'un-squash': dilate the x-axis in the ration a/b. So $A \mapsto A.a/b = \pi b^2.a/b = \pi ab$.

Fine – what next? We have already used *both* the coordinate systems to hand. There is no general way to continue this list.

The only general procedure is to superimpose finer and finer sheets of graph paper on our region, and count squares ('interior squares' and 'edge squares'). This yields numerical approximations – which is all we can hope for, and all we need, in general.

The question is whether this procedure always works. Where it is clearly most likely to fail is with highly irregular regions that are 'all edge and no middle'.

It turns out that this procedure does *not* always work; it works for *some but not all* sets – those whose structure is 'nice enough'. This goes back to the 1902 thesis of Henri LEBESGUE (1875-1941):
H. Lebesgue: Intégrale, longueur, aire. *Annali di Mat.* **7** (1902), 231-259.
Similarly in other dimensions. So: some but not all sets have a length/area/volume. Those which do are called *(Lebesgue) measurable*; length/area/volume is called *(Lebesgue) measure*; this subject is called Measure Theory. See e.g. SP Lecture 1 for details and references.

We first meet integration in just this context – finding areas under curves (say). The 'Sixth Form integral' proceeds by dividing up the range of integration on the x-axis into a large number of small subintervals, $[x, x + dx]$ say. This divides the required area up into a large number of thin strips, each of which is approximately rectangular; we sum the areas of these rectangles to approximate the area.

This informal procedure can be formalised, as the *Riemann integral* (G. F. B. RIEMANN (1826-66) in 1854). This (basically, the Sixth From integral formalised in the language of epsilons and deltas) is part of the undergraduate Mathematics curriculum. If you know it, fine; if not, don't worry.

The integration procedure that goes naturally with the Lebesgue approach above – the *Lebesgue integral* – basically replaces dividing up the region into thin strips with base $[x, x + dx]$ by using instead thin strips with base $[y, y + dy]$ (see the picture on the cover of [S]). The Lebesgue integral is much harder to set up than the Riemann integral, but much better: it is much more general, and much easier to handle.

One new aspect is that the Lebesgue integral is an *absolute* integral: we can integrate a function f iff we can integrate its modulus $|f|$. This means that formulae such as

$$\int_0^\infty \frac{\sin x}{x} dx = \frac{1}{2}\pi$$

(see e.g. M2P3 III.4, Lecture 28-29) – an *improper* Riemann integral, where $\int_0^\infty$ means $\lim_{X \to \infty} \int_0^X$ – have no counterpart in the Lebesgue theory.

It turns out that the mathematics of length, area and volume (*Lebesgue measure*) is actually no easier than that of a general measure. Correspondingly, the mathematics of the Lebesgue integral is actually no easier than that of the general (measure-theoretic) integral. For background, see e.g. [S].

2. Measure and integral; probability and expectation

A measure of total 'mass' 1 is called a *probability measure*; measure is then called *probability*; integral is then called *expectation*. So, by the above: a random variable X has an expectation, written $E[X]$, iff $|X|$ does – that is, iff

$$E[|X|] < \infty.$$

When you have met expectations before, they will have been [in general infinite] sums, or integrals. So the above says: the expectation is only defined if the relevant sum or integral is *absolutely convergent*. *Conditional convergence* is not good enough. We stress:

$$E[X] \text{ exists iff } E[|X|] \text{ exists, i.e. iff } E[|X| < \infty.$$

Measures and σ-fields

We write μ for a *measure* – length, area or volume, say ('m for measure'). Later on, we shall restrict to probability measures, which we write as P. This conveniently frees the letter μ for an expectation or mean ('m for mean').

Clearly a measure should add over disjoint sets on which it is defined: if A, B are disjoint, and measurable with measures $\mu(A)$, $\mu(B)$, then $A \cup B$ should be measurable, with measure

$$\mu(A \cup B) = \mu(A) + \mu(B).$$

That is, μ should be *additive* (over disjoint unions, understood). By induction, this should extend to finite disjoint unions (*finite additivity*): if $A_1, \ldots, A_n$ are measurable with measures $\mu(A_i)$, then their union should be measurable, with measure

$$\mu(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mu(A_i). \qquad (fa)$$

It turns out that for most purposes we get a better (and simpler!) theory by assuming additivity over *countable* disjoint unions (*countable additivity*):

if A_n are disjoint and measurable with measures $\mu(A_n)$, then their union is measurable, with measure

$$\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n). \qquad (ca)$$

Note. The contrast between (fa) and (ca) is interesting, but we have no time to discuss it here. For background, see e.g. "favca":
N. H. BINGHAM: Finite additivity versus countable additivity. *Electronic J. History of Probability and Statistics* **6**.1 (2010), 35p.

As we have seen, not all sets are measurable in general! As the above may suggest, *countability* is relevant here. Recall that an infinite set A is called *countable* if its elements can be put into one-to-one correspondence with the natural numbers. That is, we can *list* the elements of A:

$$A = \{a_1, \ldots, a_n, \ldots\}$$

(the list ends iff A is finite – recall fin = end, French). Otherwise, an infinite set is called *uncountable*. Think of the (finite or) countable sets as the 'little sets', uncountable sets as the 'big sets'. We quote:
the natural numbers, the integers and the rationals are countable;
the real numbers are uncountable; so are the numbers in any interval $I = [a, b]$ of positive length $b - a$; so are the complex numbers; so are the higher-dimensional analogues of these.

We will write our 'reference set' as Ω (see below). It turns out that, if Ω is countable, one can define a measure on *any* subset A of Ω in terms of the measures $\mu(\{a_n\})$ of singletons (one-point sets), by

$$\mu(A) := \sum_{n:a_n \in A} \mu(\{a_n\})$$

(note that both sides may be infinite: e.g., with *counting measure*, and all the singletons of mass 1 – take A the even integers).

This simple and obvious procedure does not suffice if Ω is uncountable. We write $\mathcal{A}$ for the class of measurable sets A. By (ca), $\mathcal{A}$ should be closed under countable disjoint unions. It should also be closed under *complements*: if $\mu(A)$ is defined, then $\mu(A^c)$ should be defined also (if A is nice enough for its measure to be defined, then A^c should be too, as we can specify it as 'not A', which is equivalent to specifying A). Also the empty set $\emptyset$ should be measurable with measure 0; hence its complement Ω should be measurable.