ma414l1.tex
Lecture 1. 12.1.2012

I. ANALYSIS; PROBABILITY

1. Lebesgue Measure and Integral

We recall *Lebesgue measure* (MA411 Probability and Measure) λ: defined on intervals $(a,b]$ by $\lambda((a,b]) := b-a$ (so λ is translation-invariant), and then on the σ-field generated by the intervals (Borel σ-field) by the Carathéodory Extension Theorem, and then by completion to the σ-field of Lebesgue-measurable sets (σ-field generated by the Borel sets and the null sets – sets of measure 0). This gives the mathematics of *length.* In particular, λ is a measure – is σ- (countably) additive; also, non-measurable sets occur (in profusion) – though to construct one explicitly requires the Axiom of Choice, AC (recall Zermelo-Fraenkel set theory, ZF, the logical foundations of ordinary Mathematics, and ZFC, ZF + AC). Similarly for *area* in the plane, starting from $\lambda((a_1,b_1]\times(a_2,b_2]) := (b_1-a_1)(b_2-a_2)$, and *volume* in three-dimensional Euclidean space, starting from $\lambda((a_1,b_1]\times(a_2,b_2]\times(a_3,b_3]) := (b_1-a_1)(b_2-a_2)(b_3-a_3)$; similarly also for Lebesgue measure λ, or λ_k, in Euclidean k-space. Lebesgue measure is by construction invariant under translations; it is also invariant under rotations (start from plane polars coordinates, or spherical polars in k dimensions, and use uniqueness of the Carathéodory extension procedure); combining, Lebesgue measure is invariant under the Euclidean motion group.

A *measurable function* f from a measurable space $(\Omega,\mathcal{A})$ to the reals with the Borel σ-field is a function such that the inverse image $f^{-1}(B) := \{x : f(x)\in B\}$ of any Borel set B (equivalently, any interval I) is in $\mathcal{A}$. If $\mathcal{A}$ is the σ-field of Lebesgue-measurable sets, we call f (Lebesgue-) measurable; similarly for f Borel-measurable if $\mathcal{A}$ is the Borel σ-field $\mathcal{B}$.

Recall also the *Lebesgue integral.* This is defined for indicator functions of intervals $f = I_{(a,b]}$ by $\int f$, or $\int f(x)dx$, $:= b-a$. This extends to simple functions (linear combinations of indicators of intervals) by linearity. A non-negative measurable function f is the increasing limit of a sequence of simple functions f_n, and one can extend the integral to such f by $\int f := \lim \int f_n$. The limit, which exists as the integral is order-preserving and $f_n \uparrow f$, can be shown to be independent of the approximating sequence (or, one can define $\int f := \sup\{\int f_n : f_n \le f, f_n \text{ simple}\}$); one calls such f (Lebesgue-) *integrable* if $\int f < +\infty$. One can extend to measurable functions of either sign by linearity: if $f = f_+ - f_-$ with $f_\pm$ non-negative and measurable, $\int f := \int f_+ - \int f_-$. This gives the class of (Lebesgue-) *integrable* functions, L_1.

From the definition of the integral above, if two functions agree a.e. (almost everywhere – except on a null set), then their integrals agree. This means that the integral is really defined, not for functions but for *equivalence classes* of functions, under the equivalence relation of equality a.e. So L_1 above is really a space of equivalence classes of functions, not of functions (we use the same letter as before so as not to complicate the notation – there is no risk of ambiguity).

For $p > 0$, L_p (or L^p) is the class of (equivalence classes of) functions f with $f^p \in L_1$. These are the Lebesgue spaces of L_p spaces.

Properties of the integral. The (Lebesgue) integral is

1. linear: $\int af + bg = a \int f + b \int g$ (here a, b are constants and f, g are functons, understood);
2. order-preserving: $f \leq g$ implies $\int f \leq \int g$;
3. absolute: f is integrable iff $|f|$ is integrable (we have to restrict to integrals $< +\infty$ in the non-negative case, so as to avoid nonsensical expressions $\infty - \infty$ when we extend from the non-negative to the general case);
4. an extension of the (proper) Riemann integral: if $f : [a, b] \to \mathbf{R}$ is Riemann-integrable, it is Lebesgue-integrable to the same value.
5. A function f on $[a, b]$ is Riemann-integrable iff it is continuous a.e. This shows that many more functions are Lebesgue-integrable than are Riemann-integrable. For example, the indicator function f of the rational numbers in $[0, 1]$ is Lebesgue-integrable, to 0, as it is a.e. 0 (the set of rationals is countable, and any countable set has measure 0). But this f is discontinuous everywhere (as the rationals are dense in the reals) – so it is as far from being Riemann-integrable as it could be.

Convergence theorems.

We often need to interchange limit and integral, to conclude $\int \lim f_n = \lim \int f_n$. This can be done in Real Analysis using the Riemann integral, provided that $f_n \to f$ *uniformly.* The Riemann integral is also suitable in Complex Analysis, where if functions f_n are holomorphic (= analytic, or regular) and $f_n \to f$ uniformly on compact subsets of a domain D where they are holomorphic, then

(i) f is also holomorphic;
(ii) the derivatives $f_n^{(k)}$ are (holomorphic and) convergent: $f_n^{(k)} \to f$ on D (the proof uses Morera's Theorem and the Cauchy Integral Formulae from

Complex Analysis).

But for Real Analysis beyond a first course, and for Probability Theory (and hence also, Mathematical Finance), uniform convergence is much too restrictive. We need pointwise convergence, *plus* something. We quote the three classic results of this kind (M, F, D – they are in fact equivalent).

M (Lebesgue's Monotone Convergence Theorem – monotone convergence). If $f_n \uparrow f$ and f_n are integrable, $\int f_n \uparrow \int f$ (if f is integrable; $\int f_n \uparrow +\infty$ otherwise).

F (Fatou's Lemma). (i) If f_n are measurable and bounded below by an integrable function g, $f_n \to f$ a.e., and $\sup_n \int f_n \le K$, then f is integrable and $\int f \le K$;

(ii)

$$\int \liminf_{n\to\infty} f_n \le \liminf_{n\to\infty} \int f_n.$$

Of course, applying the result to $-f_n$ gives the alternative form: for functions bounded above by an integrable function, $\int \limsup \ge \limsup \int$.

D (Lebesgue's dominated convergence theorem – dominated convergence). If f_n are measurable, $f_n \to f$ and $|f_n| \le g$ with g μ-integrable, then

$$\int f_n \to \int f.$$

2. General measure and integral.

All this goes through much more generally, and is no harder in full generality. In Measure Theory, one has a measurable space $(\Omega \mathcal{A})$ as above (Ω is the base space on which we work, $\mathcal{A}$ a σ-field of subsets of it – a class of subsets closed under complements and countably many set-theoretic operations – unions $\cup$, intersections $\cap$ and set-theoretic differences $\backslash$. A *measure* μ is a non-negative set-function defined on $\mathcal{A}$ and countable additive:

$$\mu(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$$

for disjoint $A_n \in \mathcal{A}$. A *signed measure* is a σ-additive set-function not restricted to be non-negative; a signed measure can also be defined as the difference of two measures (see below). Then $(\Omega, \mathcal{A}, \mu)$ is called a *measure space.*

The *integral* $\int f d\mu$ is defined just as the Lebesgue integral is defined, but in place of $\lambda((a, b])$ one has $\mu(A)$, where A runs through a class of sets $A \in \mathcal{A}$ big enough to generate the whole σ-algebra $\mathcal{A}$. Then everything said above about the Lebesgue integral (w.r.t. Lebesgue measure λ) extends verbatim to this case, with $\int f d\mu$ in place of $\int f = \int f d\lambda$.

When μ is a measure of total mass 1, $\mu(\Omega) = 1$, μ is called a *probability measure* (= probability), usually written P, and then $(\Omega, \mathcal{A}, \mathcal{P})$ is called a *probability space*; the base space Ω is called the *sample space*. The points $\omega \in \Omega$ are called the *sample points*, and indicate the randomness present. There are two prototypes:

1. *Tossing a fair coin*. Write 1 for 'head', 0 for 'tail'. $\Omega = \{0, 1\}$; $\mathcal{A}$ is the power set of Ω (class of all subsets – there are $2^2 = 4$ here, as Ω has 2 points); $\mu(\{0\}) = \mu(\{1\}) = 1/2$.

2. *Drawing a random number from* $[0, 1]$. Here $\Omega = [0, 1]$, $\mathcal{A}$ is the σ-field of (Lebesgue-) measurable sets in $[0, 1]$, $P = \lambda$ is Lebesgue measure (so here *probability = length*). This gives us the *Lebesgue probability space*, $[0, 1]$ for short. Similarly for drawing two random numbers from $[0, 1]$ independently (here probability = area, in the unit square), or three (here probability = volume, in 3-space). Similarly also for drawing n such numbers independently, leading to Lebesgue measure in n-space.

It turns out that one can 'let $n \to \infty$' here, and model drawing infinitely many random numbers independently from $[0, 1]$. Remarkably, it also turns out that this gives us the Lebesgue probability space above (see below).

If μ is a measure and

$$\nu(A) := \int_A f d\mu$$

for some measurable function $f \geq 0$, then $Q(A) \geq 0$ for all $A \in \mathcal{A}$ (Q is non-negative), and if $A = \cup_{n=1}^{\infty} A_n$ with $A_n \in \mathcal{A}$ disjoint, then

$$\nu(A) = \int_A f d\mu = \int I_A f d\mu = \int \sum_n I_{A_n} d\mu = \sum \int_{Am} f d\mu = \sum \nu(A_n) :$$

ν is σ-additive. So ν is a measure. Also, if $\mu(A) = 0$, then also $\nu(A) = \int_A f d\mu = 0$, as any integral over a set of measure 0 is 0. So $\mu(A) = 0$ implies $\nu(A) = 0$. We write this as

$$\nu << \mu,$$

and say that ν is *absolutely continuous* w.r.t. μ.

It turns out that this is the only way in which absolute continuity can arise: $\nu << \mu$ iff $\nu = \int f d\mu$ for some measurable $f \geq 0$. This is the *Radon-Nikodym theorem* (RN), which we quote; then f is called the *Radon-Nikodym derivative* of ν w.r.t. μ, written

$$f = d\nu/d\mu.$$

Thus formally,
$$\nu(A) = \int_A d\nu = \int_A \frac{d\nu}{d\mu} d\mu = \int_A f d\mu.$$
If both $\nu << \mu$ and $\mu << \nu$, we call the measures μ and ν *equivalent*. Then both RN derivatives exist, and
$$d\nu/d\mu = 1/(d\mu/d\nu).$$
In this case, μ and ν have the same null sets.

3. Probability

In a probability space, we call the measurable functions *random variables*. Recall from your first course in Probability Theory that for a random variable X we define its *distribution function* F (or F_X) by
$$F(x) := P(\{\omega : X(\omega) \leq x\}), \text{ or } P(X \leq x) \qquad (x \in \mathbf{R}).$$
The measurability restriction is exactly the same as requiring that the distribution function F be defined.

In a measurable space, we also call the integral the *expectation*, E. Thus for a random variable (rv) X,
$$E[X], \text{ or } EX := \int_\Omega X(\omega) dP(\omega) \text{ or } \int X dP.$$
The expectation is a real number, and has the interpretation of a weighted average of the values of the rv X, weighted according to their probability. The two prototypical cases here are:
(i) *Discrete case.* Here X takes values x_n (finitely or countably many), with probabilities $f(x_n) > 0$. Then
$$P(A) = \sum_{n: x_n \in A} f(x_n), \qquad EX = \sum_n x_n f(x_n)$$
(the series must be *absolutely* convergent, or the expectation is not defined – this restriction is needed, to ensure linearity of expectation).
(ii) *Density case.* Here X takes values in some interval (or half-line, or the whole line),
$$P(A) = \int_A f(x)dx, \quad F(x) = P(X \leq x) = \int_{-\infty}^x f(u)du, \quad EX = \int_{-\infty}^\infty x f(x)dx$$

for some $f \geq 0$ which integrates to 1. Then f is called the (probability) *density* (function) of F, X.

We recall some basic examples. The first two are discrete; the last three have densities.

Binomial distribution, $B(n,p)$. For $p \in [0,1]$ (or $(0,1)$ to exclude trivialities), $q := 1-p$,

$$P(X = k) = \binom{n}{k} p_k q^{n-k} \qquad (k = 0, 1, \ldots, n).$$

Then (check) $EX = np$, $var\ X = npq$ (here the *variance*, or variability, is $var\ X := E[(X - EX)^2]$, and then $var\ X = E[X^2] - (EX)^2$). Then X models the number of heads in a n independent tosses of a biased coin that falls heads with probability p (so tails with probability q). The case $n = 1$ is called the *Bernoulli distribution*, $B(p)$.

Poisson distribution, $P(\lambda)$. For $k = 0, 1, 2, \ldots$,

$$P(X = k) = e^{-\lambda}\lambda^k/k!$$

Then (check) $EX = \lambda$, $var\ X = \lambda$.

Uniform distribution, $U(a,b)$. This has density $f(x) = 1/(b-a)$ on $[a,b]$, 0 elsewhere. Then (check) $EX = (a+b)/2$, $var\ X = (b-a)^2/12$. The special case $U(0,1)$ has $f(x) = I_{[0,1]}$, so for $0 \leq a \leq b \leq 1$,

$$P(a \leq X \leq b) = b - a :$$

probability = length. This gives the Lebesgue probability space above, and models drawing a random number uniformly from $[0,1]$.

Exponential distribution, $E(\lambda)$, for $\lambda > 0$. This has

$$f(x) = \lambda e^{-\lambda x}, \quad F(x) = P(X \leq x) = 1 - e^{-\lambda x} \quad (x \geq 0), \quad EX = 1/\lambda.$$

Normal distribution $N(\mu, \sigma^2)$. This has density

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\{-\frac{1}{2}(x-\mu)^2/\sigma^2\}.$$

Then (check) $EX = \mu$, $varX = \sigma^2$. The special case $\mu = 0$, $\sigma = 1$ is called the *standard normal*, $N(0,1) = \Phi$, with density ϕ:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad \Phi(x) = P(X \leq x) = \int_{-\infty}^{x} \phi(u)du.$$

4. Stieltjes integrals.

In both the Riemann integral and the Lebesgue integral, intervals $(a, b]$ played a major role, and we used that they have length $b - a$. The "dx" in $\int f(x)dx$ (Riemann or Lebesgue) comes from this. We now generalize this.

Suppose that F is a non-decreasing function. Then F can have at worst jump discontinuities; we take F to be *right*-continuous at any jump points. We replace the length $b-a$ of $(a, b]$ by $F(b) - F(a)$. This gives a *set-function* μ_F, defined by

$$\mu_F((a, b]) := F(b) - F(a).$$

If in the Riemann integral we replace lower R-sums $\sum m_i(x_{i+1} - x_i)$ by $\sum m_i(F(x_{i+1}) - F(x_i))$, and similarly for upper R-sums, we obtain an extension of the R-integral, called the *Riemann-Stieltjes integral* or RS-integral (Thomas STIELTJES (1856- 1894) in 1894/5). It is written $\int_a^b f(x)dF(x)$; here f is called the *integrand*, F the *integrator*. Care is needed if both integrand and integrator can have common points of discontinuity. We shall need to allow F to have jumps; we restrict to f continuous accordingly.

If in the definition of the measure-theoretic integral we take μ to be μ_F on half-open intervals $(a, b]$ as above, and then construct the integral as with the Lebesgue integral but with $\mu_F((a, b]) := F(b) - F(a)$ in place of $b - a$, we obtain the *Lebesgue-Stieltjes integral* or LS-integral.

Such Stieltjes integrals are important in Probability Theory. A random variable (rv) (X say) has a (probability) distribution function, F $(= F_X)$. Then the LS-integral $\int g(x)dF(x)$ has the interpretation of an *expectation*, $Eg(X)$ of the function $g(X)$ of the rv X.

Signed measures.

While length/area/volume, probability and (gravitational) mass are all non-negative, electrostatic charge can have either sign. A *signed measure* is a countably additive set function (not necessarily non-negative). The measure theory of signed measures is fairly simple: a signed measure μ can be written uniquely as

$$\mu = \mu^+ - \mu^-,$$

where $\mu^{\pm}$ are measures, with disjoint supports (the support of a measure is the largest set whose complement is null). This is the *Hahn-Jordan theorem* (Hans HAHN (1879-1934) in 1948, posth., Camille JORDAN (1838-1922) in 1881).

We can extend the LS integral from non-decreasing integrands F to

$$F = F_1 - F_2$$

that are the difference of two non-decreasing functions, in the obvious way:

$$\int f dF := \int f dF_1 - \int f dF_2$$

(both terms on the right must be finite – we must avoid '$\infty - \infty$'). This gives the LS integral with integrator the F, or the corresponding signed measure (cf. the Hahn-Jordan theorem).

So suitable integrators are differences of monotone functions. But how do we recognize them? For an interval $[a, b]$, let $\mathcal{P}$ be a partition:

$$a = x_0 < x_1 < \ldots < x_n = b.$$

For a function F, the *variation* of F over the partition $\mathcal{P}$ is

$$var(F, \mathcal{P}) := \sum |F(x_{i+1} - F(x_i)|.$$

Call F of *finite variation* (FV) on $[a, b]$ if

$$var_{[a,b]} F := \sup\{var(F, \mathcal{P})\} < \infty,$$

where $\mathcal{P}$ varies over all partitions. Of course, monotone functions are of FV: if F is monotone, $var(F, \mathcal{P}) = |F(b) - F(a)|$, so taking the sup over $\mathcal{P}$, $var_{[a,b]} F = |F(b) - F(a)|$. Of course also, we need to restrict to finite intervals (or compact sets): the case $F(x) \equiv x$ generating Lebesgue measure is the prototype, but x, though of FV on finite intervals, has infinite variation over the real line. We quote:

Theorem (Jordan, 1881). The following are equivalent:
(i) F is the difference of two monotone functions;
(ii) F is of finite variation (FV) on intervals $[a, b]$.

Later we will encounter *stochastic integrals* $\int h dX$, where both the integrand h and the integrator X are random (stochastic processes). These will be of two types: X of FV, when we can use LS-integrals as above, and X not FV (e.g.: Brownian motion, Ch. II) when we will need an entirely new kind of integral, the *Itô integral* (Ch. III).