spl10.tex
Lecture 10. 1.11.2010.

II. PROBABILITY; CONDITIONAL EXPECTATION

1. 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 of Ch. I 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), which, as in Ch. I we will strengthen to
3*. If $A_1, A_2 \ldots$ (*ad inf.*) are disjoint,

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

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. That is, $\mathcal{F}$ should be a *σ-field* of subsets of Ω. 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.

2. 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 $\mathbf{R}$, $X \to \mathbf{R}$,

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

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

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

We can only deal with functions X for which all these probabilities are defined. So, for each x, we need $\{\omega : \; X(\omega) \le x\} \in \mathcal{F}$ – that is, 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, $\{X < x\}$, $\{X \ge x\}$, $\{X > X\}$] 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 $\{\omega : X(\omega) \in B\}$,

for $B \in \mathcal{B}$, the Borel σ-field [generated by the intervals] on the line.
Interpretation. Think of $\sigma(X)$ as representing *what we know when we know* X, or in other words *the information contained in* X (or in knowledge of X). This is reflected in Doob's lemma (L9):

$$\sigma(X) \subseteq \sigma(Y) \quad \text{iff} \quad X = g(Y)$$

for some measurable function g. For, knowing Y means we know $X := g(Y)$ – but not vice-versa, unless the function g is one-to-one [injective], when the inverse function g^{-1} exists, and we can go back via $Y = g^{-1}(X)$.

3. Expectation.

As in Ch. I, a measure determines an integral. A probability measure P, being a special kind of measure [a measure of total mass one] determines a special kind of integral, called an *expectation.*
Definition. The *expectation* E of a random variable X on $(\Omega, \mathcal{F}, P)$ is defined by

$$EX := \int_\Omega X \, dP, \text{ or } \int_\Omega X(\omega) \, dP(\omega).$$

If X is real-valued, say, with distribution function F, recall that EX is defined in your first course on probability by

$$EX := \int x f(x) \, dx \text{ if } X \text{ has a density } f$$

or if X is discrete, taking values X_n, $(n = 1, 2, \ldots)$ with probability function $f(x_n)(\geq 0)$, $(\sum x_n f(x_n) = 1)$,

$$EX := \sum x_n f(x_n).$$

These two formulae are the special cases (for the density and discrete cases) of the general formula

$$EX := \int_{-\infty}^{\infty} x \, dF(x)$$

where the integral on the right is a Lebesgue-Stieltjes integral. This in turn agrees with the definition above, since if F is the distribution function of X,

$$\int_\Omega X \, dP = \int_{-\infty}^{\infty} x \, dF(x)$$

follows by transformation of the integral (or *change of variable formula*), on applying the map $X : \Omega \to \mathbf{R}$ (L7). Then X maps to the identity, x; Ω maps to the real line; P maps to the image measure $X(P) = P \circ X^{-1} = P(X^{-1})$, which is called the (probability) *distribution*, or *law*, of X. It is a Lebesgue-Stieltjes measure on the line; the corresponding LS function is called the *distribution function* of X, F or F_X. We use the same letter for the LS measure and function, whence the F as the integrator in the equation above.

Note that the P-integral on the left, and the LS-integral on the right, are both *absolute* integrals (f is integrable iff $|f|$ is integrable), as these are measure-theoretic integrals. This explains why, when we meet expectation in our first course on Probability as a sum or integral, the sum or integral (if infinite) should be restricted to be *absolutely convergent* (correct your undergraduate notes if it wasn't!) Without the restriction to absolute convergence, we lose the vitally important property of linearity of expectation.

Glossary. We now have two parallel languages, measure-theoretic and probabilistic:

Measure	Probability
Integral	Expectation
Measurable set	Event
Measurable function	Random variable
almost-everywhere (a.e.)	almost-surely (a.s.).

We extend this by calling a.e. convergence in the measure case *a.s. convergence* in the probability case, and convergence in measure in the measure case *convergence in probability* (in prob., or in pr.) in the probability case. So (L9) neither of a.s. convergence or convergence in pth mean implies the other; each implies convergence in pr., but not conversely. (That a.s. convergence implies convergence in pr follows easily from Egorov's theorem.)
Example. We show by example that convergence in pr does not imply a.s. convergence (a fact known to F. Riesz in 1912). On the *Lebesgue measure space* $[0, 1]$ (i.e., $([0, 1], \mathcal{L}, \lambda)$, let

$$s_n := 1/2 + 1/3 + \ldots + 1/n \quad (mod\ 1), \quad A_n := [s_{n-1}, s_n], \quad X_n := I_{A_n}.$$

Since the harmonic series diverges, the X_n endlessly move rightwards through the interval $[0, 1]$, exiting right and reappearing left. So the X_n do not converge anywhere, in particular are not a.s. convergent. But since $X_n = 0$ except on a set of probability $1/n$, $X_n \to 0$ in probability.