m3a22l9.tex

Lecture 9. 30.10.2015

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 from the following result, due to J. L. DOOB (1910-2004), which we quote:

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

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 $Y = g^{-1}(X)$.

Expectation.

A measure (II.1) determines an integral (II.2). 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 the *change of variable formula* for the measure-theoretic integral, on applying the map $X : \Omega \to \mathbb{R}$ (we quote this: see any book on Measure Theory).

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.)

§4. Equivalent Measures and Radon-Nikodym derivatives.

Given two measures P and Q defined on the same σ-field $\mathcal{F}$, we say that P is *absolutely continuous* with respect to Q, written

$$P << Q,$$

if $P(A) = 0$ whenever $Q(A) = 0$, $A \in \mathcal{F}$. We quote from measure theory the vitally important *Radon-Nikodym theorem:* $P << Q$ iff there exists a ($\mathcal{F}$-) measurable function f such that

$$P(A) = \int_A f \, dQ \quad \forall A \in \mathcal{F}$$

(note that since the integral of anything over a null set is zero, any P so representable is certainly absolutely continuous with respect to Q – the point is that the converse holds). Since $P(A) = \int_A dP$, this says that $\int_A dP = \int_A f \, dQ$ for all $A \in \mathcal{F}$. By analogy with the chain rule of ordinary calculus, we write dP/dQ for f; then

$$\int_A dP = \int_A \frac{dP}{dQ} dQ \quad \forall A \in \mathcal{F}.$$

Symbolically,

$$\text{if } P << Q, \quad dP = \frac{dP}{dQ} dQ.$$

The measurable function (= random variable) dP/dQ is called the *Radon-Nikodym derivative* (RN-derivative) of P with respect to Q.

If $P << Q$ and also $Q << P$, we call P and Q *equivalent* measures, written $P \sim Q$. Then dP/dQ and dQ/dP both exist, and

$$\frac{dP}{dQ} = 1 \Big/ \frac{dQ}{dP}.$$

For $P \sim Q$, $P(A) = 0$ iff $Q(A) = 0$: P and Q *have the same null sets.* Taking negations: $P \sim Q$ iff P, Q have the same sets of positive measure. Taking complements: $P \sim Q$ iff P, Q have the same sets of probability one [the same a.s. sets]. Thus the following are equivalent: $P \sim Q$ iff P, Q have the same null sets/the same a.s. sets/the same sets of positive measure.
Note. Far from being an abstract theoretical result, the Radon-Nikodym theorem is of key practical importance, in two ways:
(a) It is the key to the concept of conditioning ("using what we know" – §5, §6 below), which is of central importance throughout,
(b) The concept of equivalent measures is central to the key idea of mathematical finance, *risk-neutrality*, and hence to its main results, the *Black-Scholes formula, the Fundamental Theorem of Asset Pricing (FTAP)*, etc. The key to all this is that prices should be the *discounted expected values under the equivalent martingale measure.* Thus equivalent measures, and the operation of *change of measure*, are of central economic and financial importance. We shall return to this later in connection with the main mathematical result on change of measure, *Girsanov's theorem* (VI.4).

Recall that we first met the phrase 'equivalent martingale measure' in I.5 above. We now know what a measure is, and what equivalent measures are; we will learn about martingales in III.3 below.

§5.Conditional Expectations.

Suppose that X is a random variable, whose expectation exists (i.e. $E[|X|] < \infty$, or $X \in L_1$). Then $E[X]$, the expectation of X, is a scalar (a number) – non-random. The expectation operator E averages out all the randomness in X, to give its mean (a weighted average of the possible value of X, weighted according to their probability, in the discrete case).

It often happens that we have *partial information* about X – for instance, we may know the value of a random variable Y which is associated with X, i.e. carries information about X. We may want to average out over the remaining randomness. This is an expectation conditional on our partial information, or more briefly a conditional expectation.

This idea will be familiar already from elementary courses, in two cases (see e.g. [BF]):
1. *Discrete case*, based on the formula

$$P(A|B) := P(A \cap B)/P(B) \text{ if } P(B) > 0.$$

If X takes values $x_1, \cdots, x_m$ with probabilities $f_1(x_i) > 0$, Y takes values $y_1, \cdots, y_n$ with probabilities $f_2(y_j) > 0$, (X, Y) takes values (x_i, y_j) with

probabilities $f(x_i, y_j) > 0$, then
(i) $f_1(x_i) = \sum_j f(x_i, y_j)$, $\qquad f_2(y_j) = \sum_i f(x_i, y_j)$,
(ii) $P(Y = y_j | X = x_i) = P(X = x_i, Y = y_j)/P(X = x_i) = f(x_i, y_j)/f_1(x_i)$

$$= f(x_i, y_j)/\sum_j f(x_i, y_j).$$

This is the *conditional distribution* of Y *given* $X = x_i$, written

$$f_{Y|X}(y_j|x_i) = f(x_i, y_j)/f_1(x_i) = f(x_i, y_j)/\sum_j f(x_i, y_j).$$

Its expectation is

$$E[Y|X = x_i] = \sum_j y_j f_{Y|X}(y_j|x_i)$$

$$= \sum_j y_j f(x_i, y_j)/\sum_j f(x_i, y_j).$$

But this approach only works when the events on which we condition have *positive* probability, which only happens in the *discrete* case.
2. *Density case.* If (X, Y) has density $f(x, y)$,

$$X \text{ has density } f_1(x) := \int_{-\infty}^{\infty} f(x, y)dy, \qquad Y \text{ has density } f_2(y) := \int_{-\infty}^{\infty} f(x, y)dx.$$

We *define* the *conditional density* of Y *given* $X = x$ by the continuous analogue of the discrete formula above:

$$f_{Y|X}(y|x) := f(x, y)/f_1(x) = f(x, y)/\int_{-\infty}^{\infty} f(x, y)dy.$$

Its expectation is

$$E[Y|X = x] = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy = \int_{-\infty}^{\infty} y f(x, y)dy/\int_{-\infty}^{\infty} f(x, y)dy.$$

Example: Bivariate normal distribution, $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$.

$$E[Y|X = x] = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1),$$

the familiar *regression line* of statistics (linear model: [BF, Ch. 1]).