m3f22l10tex

Lecture 10 27.11.2016

Kolmogorov's approach: conditional expectations via σ-fields

The problem with the approach of L9 (discrete and density cases) is that joint densities need not exist – do not exist, in general. One of the great contributions of Kolmogorov's classic book of 1933 was the realization that measure theory – specifically, the *Radon-Nikodym theorem* –provides a way to treat conditioning in general, without assuming that we are in the discrete case or density case above.

Recall that the probability triple is $(\Omega, \mathcal{F}, \mathcal{P})$. Take $\mathcal{B}$ a sub-σ-field of $\mathcal{F}$, $\mathcal{B} \subset \mathcal{F}$ (recall: a σ-field represents information; the big σ-field $\mathcal{F}$ represents 'knowing everything', the small σ-field $\mathcal{B}$ represents 'knowing something').

Suppose that Y is a non-negative random variable whose expectation exists: $E[Y] < \infty$. The set-function

$$Q(B) := \int_B Y dP \qquad (B \in \mathcal{B})$$

is non-negative (because Y is), σ-additive – because

$$\int_B Y dP = \sum_n \int_{B_n} Y dP$$

if $B = \cup_n B_n$, B_n disjoint – and defined on the σ-algebra $\mathcal{B}$, so is a *measure* on $\mathcal{B}$. If $P(B) = 0$, then $Q(B) = 0$ also (the integral of anything over a null set is zero), so $Q << P$. By the Radon-Nikodym theorem (II.4), there exists a Radon-Nikodym derivative of Q with respect to P on $\mathcal{B}$, which is $\mathcal{B}$-measurable [in the Radon-Nikodym theorem as stated in II.4, we had $\mathcal{F}$ in place of $\mathcal{B}$, and got a random variable, i.e. an $\mathcal{F}$-measurable function. Here, we just replace $\mathcal{F}$ by $\mathcal{B}$.] Following Kolmogorov (1933), we call this Radon-Nikodym derivative the *conditional expectation* of Y *given* (or *conditional on*) $\mathcal{B}$, $E[Y|\mathcal{B}]$: this is $\mathcal{B}$-measurable, integrable, and satisfies

$$\int_B Y dP = \int_B E[Y|\mathcal{B}] dP \qquad \forall B \in \mathcal{B}. \qquad (*)$$

In the general case, where Y is a random variable whose expectation exists ($E[|Y|] < \infty$) but which can take values of both signs, decompose Y as

$$Y = Y_+ - Y_-$$

and define $E[Y|\mathcal{B}]$ by linearity as

$$E[Y|\mathcal{B}] := E[Y_+|\mathcal{B}] - E[Y_-|\mathcal{B}].$$

Suppose now that $\mathcal{B}$ is the σ-field generated by a random variable X: $\mathcal{B} = \sigma(X)$ (so $\mathcal{B}$ represents the information contained in X, or what we know when we know X). Then $E[Y|\mathcal{B}] = E[Y|\sigma(X)]$, which is written more simply as $E[Y|X]$. Its defining property is

$$\int_B Y dP = \int_B E[Y|X] dP \qquad \forall B \in \sigma(X).$$

Similarly, if $\mathcal{B} = \sigma(X_1, \cdots, X_n)$ ($\mathcal{B}$ is the information in $(X_1, \cdots, X_n)$) we write $E[Y|\sigma(X_1, \cdots, X_n]$ as $E[Y|X_1, \cdots, X_n]$:

$$\int_B Y dP = \int_B E[Y|X_1, \cdots, X_n] dP \qquad \forall B \in \sigma(X_1, \cdots, X_n).$$

Note. 1. To check that something is a conditional expectation: we have to check that it integrates the right way over the right sets [i.e., as in (*)].
2. From (*): if two things integrate the same way over all sets $B \in \mathcal{B}$, they have the same conditional expectation given $\mathcal{B}$.
3. For notational convenience, we use $E[Y|\mathcal{B}]$ and $E_{\mathcal{B}}Y$ interchangeably.
4. The conditional expectation thus defined coincides with any we may have already encountered - in regression or multivariate analysis, for example. However, this may not be immediately obvious. The conditional expectation defined above – via σ-fields and the Radon-Nikodym theorem – is rightly called by Williams ([W], p.84) 'the central definition of modern probability'. It may take a little getting used to. As with all important but non-obvious definitions, it proves its worth in action: see II.6 below for properties of conditional expectations, and Chapter III for stochastic processes, particularly martingales [defined in terms of conditional expectations].

§6. Properties of Conditional Expectations.

1. $\mathcal{B} = \{\emptyset, \Omega\}$. Here $\mathcal{B}$ is the *smallest* possible σ-field (*any* σ-field of subsets of Ω contains $\emptyset$ and Ω), and represents 'knowing nothing'.

$$E[Y|\{\emptyset, \Omega\}] = EY.$$

Proof. We have to check (*) of §5 for $B = \emptyset$ and $B = \Omega$. For $B = \emptyset$ both sides are zero; for $B = \Omega$ both sides are EY. //

2. $\mathcal{B} = \mathcal{F}$. Here $\mathcal{B}$ is the *largest* possible σ-field: 'knowing everything'.

$$E[Y|\mathcal{F}] = Y \qquad P-a.s.$$

Proof. We have to check (*) for *all* sets $B \in \mathcal{F}$. The only integrand that integrates like Y over *all* sets is Y itself, or a function agreeing with Y except on a set of measure zero.
Note. When we condition on $\mathcal{F}$ ('knowing everything'), we *know* Y (because we know everything). There is thus no uncertainty left in Y to average out, so taking the conditional expectation (averaging out remaining randomness) has no effect, and leaves Y unaltered.

3. If Y is $\mathcal{B}$-measurable, $E[Y|\mathcal{B}] = Y \qquad P-a.s.$
Proof. Recall that Y is *always* $\mathcal{F}$-measurable (this is the definition of Y being a random variable). For $\mathcal{B} \subset \mathcal{F}$, Y may not be $\mathcal{B}$-measurable, but if it is, the proof above applies with $\mathcal{B}$ in place of $\mathcal{F}$.
Note. If Y is $\mathcal{B}$-measurable, when we are given $\mathcal{B}$ (that is, when we condition on it), we *know* Y. That makes Y effectively a constant, and when we take the expectation of a constant, we get the same constant.

4. If Y is $\mathcal{B}$-measurable, $E[YZ|\mathcal{B}] = YE[Z|\mathcal{B}] \qquad P-a.s.$
We refer for the proof of this to [W], p.90, proof of (j).
Note. Williams calls this property 'taking out what is known'. To remember it: if Y is $\mathcal{B}$-measurable, then given $\mathcal{B}$ we know Y, so Y is effectively a constant, so can be taken out through the integration signs in (*), which is what we have to check (with YZ in place of Y).

5. If $\mathcal{C} \subset \mathcal{B}$, $E[E[Y|\mathcal{B}]|\mathcal{C}] = E[Y|\mathcal{C}] \qquad a.s.$
Proof. $E_{\mathcal{C}}E_{\mathcal{B}}Y$ is $\mathcal{C}$-measurable, and for $C \in \mathcal{C} \subset \mathcal{B}$,

$$\int_C E_{\mathcal{C}}[E_{\mathcal{B}}Y]dP = \int_C E_{\mathcal{B}}Y\,dP \qquad (\text{definition of } E_{\mathcal{C}} \text{ as } C \in \mathcal{C})$$

$$= \int_C Y\,dP \qquad (\text{definition of } E_{\mathcal{B}} \text{ as } C \in \mathcal{B}).$$

So $E_{\mathcal{C}}[E_{\mathcal{B}}Y]$ satisfies the defining relation for $E_{\mathcal{C}}Y$. Being also $\mathcal{C}$-measurable, it *is* $E_{\mathcal{C}}Y$ (a.s.). //

5'. If $\mathcal{C} \subset \mathcal{B}$, $E[E[Y|\mathcal{C}]|\mathcal{B}] = E[Y|\mathcal{C}] \quad a.s.$
Proof. $E[Y|\mathcal{C}]$ is $\mathcal{C}$-measurable, so $\mathcal{B}$-measurable as $\mathcal{C} \subset \mathcal{B}$, so $E[.|\mathcal{B}]$ has no effect on it, by 3.
Note. 5, 5' are the two forms of the *iterated conditional expectations property.* When conditioning on two σ-fields, one larger (finer), one smaller (coarser), the coarser rubs out the effect of the finer, either way round. This is also called the *coarse-averaging property*, or (Williams [W]) the *tower property.*

6. *Conditional Mean Formula.* $E[E[Y|\mathcal{B}]] = EY \quad P - a.s.$
Proof. Take $\mathcal{C} = \{\emptyset, \Omega\}$ in 5 and use 1. //
Example. Check this for the bivariate normal distribution considered above.
Note. Compare this with the *Conditional Variance Formula* of Statistics: see e.g. SMF, IV.6, Day 9.

7. *Role of independence.* If Y is independent of $\mathcal{B}$,

$$E[Y|\mathcal{B}] = E[Y] \qquad a.s.$$

Proof. See [**W**], p.88, 90, property (k).
Note. In the elementary definition $P(A|B) := P(A \cap B)/P(B)$ (if $P(B) > 0$), if A and B are independent (that is, if $P(A \cap B) = P(A).P(B)$), then $P(A|B) = P(A)$: conditioning on something independent has no effect. One would expect this familiar and elementary fact to hold in this more general situation also. It does – and the proof of this rests on the proof above.

Projections. In Property 5 (tower property), take $\mathcal{B} = \mathcal{C}$:

$$E[E[X|\mathcal{C}]|\mathcal{C}] = E[X|\mathcal{C}].$$

This says that the operation of taking conditional expectation given a sub-σ-field $\mathcal{C}$ is *idempotent* – doing it twice is the same as doing it once. Also, taking conditional expectation is a *linear* operation (it is defined via an integral, and integration is linear). Recall from Linear Algebra that we have met such idempotent linear operations before. They are the *projections.* (Example: $(x, y, z) \mapsto (x, y, 0)$ projects from 3-dimensional space onto the (x, y)-plane.) This view of conditional expectation as projection is useful and powerful; see e.g. [BK], [BF] or J. Neveu, *Discrete-parameter martingales* (North-Holland, 1975), I.2. It is particularly useful when one has not yet got used to conditional expectation defined measure-theoretically as above, as it gives us an alternative (and perhaps more familiar) way to think.