A BRIEF PRIMER ON CONDITIONAL EXPECTATION

Conditional expectation is a topic that I found somewhat obscure as a student. However there is really nothing to be afraid of.

Let $(X, \mathcal{F}, \mu)$ be a probability space (I have emphasised the σ-algebra $\mathcal{F}$, as it plays an important rôle in the discussion). Let $\mathcal{F}' \subseteq \mathcal{F}$ be a sub σ-algebra of $\mathcal{F}$. One may define a *conditional expectation operator* $\mathbb{E}(\,\cdot\,|\mathcal{F}') : L^1(X, \mathcal{F}) \to L^1(X, \mathcal{F}')$; we denote the image of a function f under this map by $\mathbb{E}(f|\mathcal{F}')$.

We will give the definition and some basic properties of this map in a moment. But what does it mean? I find it easiest to visualise the construction in the case that X is a finite set and μ is the counting measure, normalised so that $\mu(X) = 1$. In this case the σ-algebras $\mathcal{F}$ and $\mathcal{F}'$ define partitions of X into basic cells which we may call atoms. Since $\mathcal{F}' \subseteq \mathcal{F}$, the partition into the atoms of $\mathcal{F}$ is a *refinement* of that into the atoms of $\mathcal{F}'$.

What is $L^1(X, \mathcal{F})$? In this finite setting it (and all the other spaces $L^p(X, \mathcal{F})$) are just the space of $\mathcal{F}$-measurable functions, that is to say functions which are constant on atoms of $\mathcal{F}$. Suppose that $f \in L^1(X, \mathcal{F})$. To get a function in $L^1(X, \mathcal{F}')$, we replace f by its average over each atom defined by $\mathcal{F}'$.

Think of f as a random variable (which is, after all, just another name for a real-valued function on a probability space). Suppose that $g \in L^1(X, \mathcal{F})$ is another random variable. We may use g to define a σ-algebra $\mathcal{F}'$, by dividing X into atoms on which g is constant. Then $\mathbb{E}(f|\mathcal{F}') : X \to \mathbb{R}$, more normally written in this context as $\mathbb{E}(f|g)$, may be thought of as follows: $\mathbb{E}(f|g)(x)$ is the expected value of f conditioned upon the event that $\{x' : g(x') = g(x)\}$, where conditioning here is in the sense of IA Probability.

Hopefully this gives some intuition. When the space X is infinite, as is generally the case in this course, the picture is more complicated than the finite case suggests and it does not make sense to think in terms of a partition into atoms. My favourite way to define the conditional expectation operator $\mathbb{E}(\cdot|\mathcal{F}') : L^1(X, \mathcal{F}) \to L^1(X, \mathcal{F}')$ is by first defining it on $L^2(X, \mathcal{F}) \subseteq L^1(X, \mathcal{F})$. In this setting we have Hilbert space theory available to us, and indeed it is very natural to define

$$\mathbb{E}(\,\cdot\,|\mathcal{F}') : L^2(X, \mathcal{F}) \to L^2(X, \mathcal{F}')$$

to be simply the projection onto the closed subspace $L^2(X, \mathcal{F}')$. Suppose that f also lies in $L^1(X, \mathcal{F})$. Then $\operatorname{sgn}(\mathbb{E}(f|\mathcal{F}'))$ lies in $L^2(X, \mathcal{F}')$ and so

$$\langle f - \mathbb{E}(f|\mathcal{F}'), \operatorname{sgn}(\mathbb{E}(f|\mathcal{F}'))\rangle = 0.$$

Thus we have the inequality

$$\|\mathbb{E}(f|\mathcal{F}')\|_1 = \int f \operatorname{sgn}(\mathbb{E}(f|\mathcal{F}'))\, d\mu \leqslant \|f\|_1. \tag{0.1}$$

for all $f \in L^1(X, \mathcal{F}) \cap L^2(X, \mathcal{F})$. Since L^1 is dense in L^2, it is not hard to see that we may define a unique extension of $\mathbb{E}(\,\cdot\,|\mathcal{F}')$ to a linear operator from $L^1(X, \mathcal{F})$ to $L^1(X, \mathcal{F}')$ such that (0.1) is still satisfied (that is, the operator norm is 1).

Another way to define the conditional expectation is to invoke the *Radon-Nikodym theorem.* (By this point students will be aware of my distaste for bringing out big theorems when this is unnecessary. This issue is slightly compounded here by the fact the proof of Radon-Nikodym in Rudin's book actually uses Hilbert space techniques similar to those we discussed above.) The Radon-Nikodym theorem is an important result in measure theory, so one should probably be aware of the statement. It involves two measures μ_1, μ_2 on a space X equipped with a σ-algebra $\mathcal{F}$. We say that μ_2 is *absolutely continuous* with respect to μ_1 if $\mu_2(E) > 0$ implies that $\mu_1(E) > 0$.

Theorem 0.1 (Radon-Nikodym)**.** *Suppose that μ_2 is absolutely continuous with respect to μ_1. Then there is a (unique) function $g \in L^1(X, \mathcal{F}, \mu_1)$ such that $\mu_2(E) = \int g 1_E \, d\mu_1$ for all $E \in \mathcal{F}$.*

The function f is called the Radon-Nikodym derivative of μ_2 with respect to μ_1 and is usually denoted $d\mu_2/d\mu_1$.

Now suppose that $f \in L^1(X, \mathcal{F})$. Then we may define a measure μ' on $(X, \mathcal{F}')$ by setting $\mu'(E) := \int f 1_E \, d\mu$ for all $E \in \mathcal{F}'$. The measure μ may be restricted to a measure on $(X, \mathcal{F}')$, which we also denote by μ'. As measures on $(X, \mathcal{F}')$, μ' is absolutely continuous with respect to μ. Indeed if $\mu(E) = 0$ then it is easy to see that $\int f 1_E \, d\mu = 0$, say by approximating f by an increasing sequence of step functions. By the Radon-Nikodym theorem there is some $g \in L^1(X, \mathcal{F}')$ such that

$$\int f 1_E \, d\mu = \mu'(E) = \int g 1_E \, d\mu \tag{0.2}$$

for all $E \in \mathcal{F}'$, and we define $g := \mathbb{E}(f|\mathcal{F}')$.

The two definitions we have given coincide: to see this, use (0.2) to show that if $f \in L^2(X, \mathcal{F})$ then $g \in L^2(X, \mathcal{F}')$, then use it again to see that $f - \mathbb{E}(f|\mathcal{F}')$ is orthogonal to all functions in $L^2(X, \mathcal{F}')$. Now check that with the second definition the map $f \mapsto \mathbb{E}(f|\mathcal{F}')$ is a bounded linear functional, and hence conclude that it must be the same object as before.

We conclude by listing some properties of the conditional expectation map. When thinking about these and other properties of the conditional expectation operators, I always keep the example of a finite measure space in mind.

Theorem 0.2 (Basic properties of conditional expectation)**.** *Suppose that $(X, \mu, \mathcal{F})$ is a probability space and that $\mathcal{F}'' \subseteq \mathcal{F}' \subseteq \mathcal{F}$ are σ-algebras. Let $f \in L^1(X, \mathcal{F})$. Then*

(i) *If $g \in L^\infty(X, \mathcal{F}')$ then multiplication by g commutes with $\mathbb{E}(\,\cdot\,|\mathcal{F}')$.*
(ii) *If $f \geqslant 0$ for a.e. x then $\mathbb{E}(f|\mathcal{F}') \geqslant 0$ for a.e. x.*
(iii) $\mathbb{E}(\,\cdot\,|\mathcal{F}'') \circ \mathbb{E}(\,\cdot\,|\mathcal{F}') = \mathbb{E}(\,\cdot\,|\mathcal{F}'')$.
(iv) *$\mathbb{E}(\,\cdot\,|\mathcal{F}')$ is a contraction in L^p for all $1 \leqslant p \leqslant \infty$.*
(v) *Suppose that $T : X \to X$ is measurable and that $\mathcal{F}'$ is the σ-algebra of T-invariant sets. Then $\mathbb{E}(f \circ T|\mathcal{F}') = \mathbb{E}(f|\mathcal{F}')$.*

Sketch Proof. I use the first (Hilbert space) definition. To check (i), suppose that $f \in L^2(X, \mathcal{F})$ and that $h \in L^2(X, \mathcal{F}')$. Then $gh \in L^2(X, \mathcal{F}')$, and so

$$\langle gf - g\mathbb{E}(f|\mathcal{F}'), h\rangle = \langle f - \mathbb{E}(f|\mathcal{F}'), gh\rangle = 0.$$

It follows that $g\mathbb{E}(f|\mathcal{F}') = \mathbb{E}(gf|\mathcal{F}')$ for all such f, and the same is true for all $f \in L^1(X, \mathcal{F})$ by a limiting argument.

(ii) If not then $\max(\mathbb{E}(f|\mathcal{F}'), 0)$ is a function in $L^2(X, \mathcal{F}')$ which is closer to f than $\mathbb{E}(f|\mathcal{F}')$, contrary to the definition.

(iii) is immediate for $f \in L^2(X, \mathcal{F})$, and again we can extend to $f \in L^1(X, \mathcal{F})$ by limiting arguments.

(iv) It follows easily from (ii) and the linearity of conditional expectation that

$$\|\mathbb{E}(f|\mathcal{F}')\|_\infty \leqslant \|f\|_\infty.$$

The fact that $\mathbb{E}(\,\cdot\,|\mathcal{F}')$ is a contraction in L^p follows from the fact that it is a contraction in L^1 and the Riesz-Thorin interpolation theorem. We may also argue directly. Suppose that $f \in L^\infty(X, \mathcal{F})$, so that $\mathbb{E}(f|\mathcal{F}') \in L^\infty(X, \mathcal{F}')$. Then, since every function in sight is bounded and hence in L^2, we have

$$\langle f - \mathbb{E}(f|\mathcal{F}'), |\mathbb{E}(f|\mathcal{F}')|^{p-1} \operatorname{sgn}(\mathbb{E}(f|\mathcal{F}'))\rangle = 0,$$

and therefore by Hölder's inequality

$$\begin{aligned}
\|\mathbb{E}(f|\mathcal{F}')\|_p^p &= \langle f, |\mathbb{E}(f|\mathcal{F}')|^{p-1} \operatorname{sgn}(\mathbb{E}(f|\mathcal{F}'))\rangle \\
&\leqslant \|f\|_p \||\mathbb{E}(f|\mathcal{F}')|^{p-1} \operatorname{sgn}(\mathbb{E}(f|\mathcal{F}'))\|_q \\
&= \|f\|_p \|\mathbb{E}(f|\mathcal{F}')\|_p^{p/q}
\end{aligned}$$

where, of course, $1/p + 1/q = 1$. It follows that conditional expectation is an L^p-contraction when restricted to functions in $L^\infty(X, \mathcal{F})$. However these functions are dense in $L^p(X, \mathcal{F})$ and so the result follows.

(v) Suppose that $f \in L^2(X, \mathcal{F})$ and that $g \in L^2(X, \mathcal{F}')$, that is to say g is a T-invariant function or in other words $U_T g = g$, where U_T is the Koopman operator associated to T. Recall from the proof of the von Neumann ergodic theorem that $U_T^* g = g$. It follows that

$$\begin{aligned}
\langle f \circ T - \mathbb{E}(f|\mathcal{F}'), g\rangle &= \langle U_T f, g\rangle - \langle \mathbb{E}(f|\mathcal{F}'), g\rangle \\
&= \langle f, U_T^* g\rangle - \langle \mathbb{E}(f|\mathcal{F}'), g\rangle \\
&= \langle f, g\rangle - \langle \mathbb{E}(f|\mathcal{F}'), g\rangle \\
&= \langle f - \mathbb{E}(f|\mathcal{F}'), g\rangle = 0.
\end{aligned}$$

It follows that $\mathbb{E}(f \circ T|\mathcal{F}') = \mathbb{E}(f|\mathcal{F}')$ for all $f \in L^2(X, \mathcal{F})$. As usual, the same is true for f in L^1 by a limiting argument. $\square$