MATH 589

Conditional expectation as a projection

Let us first recall some elementary facts about Hilbert spaces. If you are not familiar with these facts, please read the attached photocopies of several pages from Rudin's "Real and Complex analysis". These few pages are an easy reading and make this notes self-contained.

Let V be a real vector space with an inner product $(\cdot|\cdot)$. The norm of the vector $v \in V$ is $\|v\| = \sqrt{(v,v)}$ and the distance between vectors v_1, v_2 is $d(v_1, v_2) = \|v_1 - v_2\|$. d is a metric and V is called a real Hilbert space if (V, d) is a complete metric space (i.e. if every Cauchy sequence converges in V).

Throughout the rest of the notes, the letter V denotes a real Hilbert space.

Let $W \subset V$ be a closed subspace and

$$W^\perp = \{v : (v, w) = 0 \text{ for all } w \in W\}.$$

$W^\perp$ is a closed subspace and every vector $v \in V$ has a unique decomposition

$$v = w + w_\perp$$

where $v \in W$ and $v \in W^\perp$ (one writes $V = W \oplus W^\perp$). Obviously, $\|v\|^2 = \|w\|^2 + \|w_\perp\|^2$ (the Pythagorean theorem). The map

$$\mathcal{P}v = \mathcal{P}(w + w_\perp) = w$$

is linear and satisfies $\mathcal{P}^2 = \mathcal{P}$, $(\mathcal{P}v_1, v_2) = (v_1, \mathcal{P}v_2)$, $\mathrm{Ran}\mathcal{P} = W$, $\mathrm{Ker}\mathcal{P} = W^\perp$. $\mathcal{P}$ is called the orthogonal projection onto the subspace W. Note that $\mathcal{Q} := \mathbf{1} - \mathcal{P}$ is the orthogonal projection onto the subspace $W^\perp$. Obviously, $\mathcal{Q}v = w_\perp$ and $v = \mathcal{P}v + \mathcal{Q}v$.

The vector $\mathcal{P}v$ is called the projection of v onto W. $\mathcal{P}v$ is the nearest vector to v in W and the distance from v to W is $\|\mathcal{Q}v\|$. In other words,

$$\|\mathcal{Q}v\| = \inf_{x \in W} \|v - x\|.$$

If V is finite dimensional, i.e. $V = \mathbb{R}^n$, these facts are proven in introductory linear algebra courses (like MATH 133).

Let $(\Omega, \mathcal{F}, P)$ be a probability space and V a real vector space of random variables with finite variance (i.e. $\mathbb{E}(X^2) = \int_\Omega X^2 \mathrm{d}P < \infty$). The map

$$(X, Y) = \mathbb{E}(XY) = \int_\Omega XY \mathrm{d}P,$$

is an inner product and the convergence in the corresponding norm is equivalent to the m^2-convergence discussed last term (recall Section 9.4). In particular, V is a real Hilbert space, usually denoted by $L^2(\Omega, \mathrm{d}P)$.

Let $\mathcal{D} \subset \mathcal{F}$ be a σ-algebra and W a collection of all $\mathcal{D}$-measurable members of V. Obviously, W is a closed subspace of V. Let $\mathcal{P}$ be the orthogonal projection onto W.

Theorem. For all $X \in V$, $\mathcal{P}X = \mathbb{E}(X|\mathcal{D})$.

Proof. Let $D \in \mathcal{D}$. Since

$$(K_D, X) = (K_D, \mathcal{P}X) + (K_D, \mathcal{Q}X) = (K_D, \mathcal{P}X)$$

($(K_D, \mathcal{Q}X) = 0$ since $K_D \in W$ and $\mathcal{Q}X \in W^{\perp}$), we get

$$\int_D X \mathrm{d}P = \int_\Omega K_D X \mathrm{d}P = (K_D, X) = (K_D, \mathcal{P}X) = \int_D \mathcal{P}X \mathrm{d}P.$$

Hence, for all $D \in \mathcal{D}$,

$$\int_D \mathbb{E}(X|\mathcal{D}) \mathrm{d}P = \int_D \mathcal{P}X \mathrm{d}P,$$

and since both $\mathbb{E}(X|\mathcal{D})$ and $\mathcal{P}X$ are $\mathcal{D}$-measurable, we have $\mathbb{E}(X|\mathcal{D}) = \mathcal{P}X$ P-a.s. □

Let us consider the same elementary examples we have discussed the last class.

If $\mathcal{D} = \{\emptyset, \Omega\}$, then W consists of constant functions and

$$\mathbb{E}(X|\mathcal{D}) = (1, X) = \mathbb{E}(X).$$

If $\mathcal{D} = \mathcal{F}$, then $\mathcal{P} = \mathbf{1}$ (the identity map), and $\mathbb{E}(X|\mathcal{D}) = X$.

If $\mathcal{D} = \{\emptyset, A, A^c, \Omega\}$, where $0 < P(A) < 1$, then $\{K_A, K_{A^c}\}$ is an orthogonal basis for W and

$$\begin{aligned}\mathbb{E}(X|\mathcal{D}) = \mathcal{P}X &= \frac{(K_A, X)}{\|K_A\|^2} K_A + \frac{(K_{A^c}, X)}{\|K_{A^c}\|^2} K_{A^c} \\ &= \left(\frac{1}{P(A)} \int_A X \mathrm{d}P\right) K_A + \left(\frac{1}{P(A^c)} \int_{A^c} X \mathrm{d}P\right) K_{A^c}.\end{aligned}$$

It is an instructive exercise to reprove the results of Section 15.1 (for random variables with finite variance) by realizing conditional expectation as a projection in $L^2(\Omega, \mathrm{d}P)$.

As an illustration, we discuss Rao-Blackwell inequality and regression curves. According to the above, for any X with finite variance,

$$\begin{aligned}\mathbb{E}(X^2) = \|X\|^2 &= \|\mathcal{P}X\|^2 + \|\mathcal{Q}X\|^2 \\ &= \|\mathbb{E}(X|\mathcal{D})\|^2 + \|X - \mathbb{E}(X|\mathcal{D})\|^2 \\ &= \mathbb{E}(\mathbb{E}(X|\mathcal{D})^2) + \mathbb{E}((X - \mathbb{E}(X|\mathcal{D}))^2).\end{aligned}$$

Hence, $\mathbb{E}(\mathbb{E}(X|\mathcal{D})^2) \leq \mathbb{E}(X^2)$. By replacing X with $X - \mathbb{E}(X)$ we obtain

$$\mathrm{Var}(\mathbb{E}(X|\mathcal{D})) \leq \mathrm{Var}(X),$$

and this is the Rao-Blackwell inequality.

Similarly, $\mathbb{E}(X|\mathcal{D})$ is the random variable in W closest to X and

$$\mathbb{E}((X - \mathbb{E}(X|\mathcal{D}))^2) = \inf_{Z \in W} \mathbb{E}(X - Z)^2.$$

If $\mathcal{D} = \sigma(Y)$ for some random variable Y, then $\mathbb{E}(X|\mathcal{D}) = \mathbb{E}(X|Y) = \alpha(Y)$, where $\alpha(y) = \mathbb{E}(X|Y = y)$ is the regression curve of X on Y. Hence,

$$\mathbb{E}(X - \alpha(Y))^2 = \inf_g \mathbb{E}(X - g(Y))^2,$$

where the infimum is over all Borel functions g.

The above derivations are simple and transparent—the Rao-Blackwell inequality and the existence of regression curve are immediate consequences of Pythagorean theorem in the Hilbert space $L^2(\Omega, \mathrm{d}P)$ and have an obvious geometric meaning. The important point is that proper mathematical structures (in this case Hilbert spaces) may both clarify the meaning of a result and simplify its proof.

Finally, we mention that the projection method can be also used in the study of conditional expectation of general integrable random variables. The vector space of all integrable random variables with the norm $\mathbb{E}(|X|) = \int_\Omega |X| \mathrm{d}P$ is denoted $L^1(\Omega, \mathrm{d}P)$. $L^1(\Omega, \mathrm{d}P)$ is a Banach space and $L^2(\Omega, \mathrm{d}P)$ is dense in $L^1(\Omega, \mathrm{d}P)$. The properties of conditional expectation, established by the projection method for the elements of $L^2(\Omega, \mathrm{d}P)$, are easily extended by limiting arguments to the elements of $L^1(\Omega, \mathrm{d}P)$.