PROBABILITY THEORY

MATH 154

Unit 23: Conditional Expectation

23.1. Conditional probability $\mathrm{P}[A|B]$ for events leads to **conditional expectation** for σ algebras: it is denoted $\mathrm{E}[X|\mathcal{B}]$ if $\mathcal{B} \subset \mathcal{A}$ is a sub-σ-algebra.

Theorem 1 (Kolmogorov)**.** *Given $X \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathrm{P})$ and a sub σ-algebra $\mathcal{B} \subset \mathcal{A}$. There exists a random variable $Y \in \mathcal{L}^1(\Omega, \mathcal{B}, \mathrm{P})$ denoted $\mathrm{E}[X|\mathcal{B}]$ satisfying $\int_A Y \, dP = \int_A X \, dP$ for all $A \in \mathcal{B}$. We call it $\mathrm{E}[X|\mathcal{B}]$*

Proof. For $X = \sum_i a_i 1_{A_i} \in \mathcal{S}$ define $\mathrm{E}[X; A] = \sum_i a_i 1_{A_i \cap A}/\mathrm{P}[A]$. For $X \in \mathcal{L}^1$ define $\mathrm{E}[X; A]$ as a limit. We also write $\int_A X \, d\mathrm{P}$ for this conditional integration. Define the two measures $\tilde{\mathrm{P}}[A] = \mathrm{P}[A]$ and $\mathrm{P}'[A] = \int_A X \; dP = \mathrm{E}[X; A]$ on the measure space $(\Omega, \mathcal{B})$. Given a set $B \in \mathcal{B}$ with $\tilde{P}[B] = 0$, then $\mathrm{P}'[B] = 0$ so that P' is absolutely continuous with respect to $\tilde{P}$. The **Radon-Nykodym** theorem from real analysis gives a random variable $Y \in \mathcal{L}^1(\mathcal{B})$ with $\mathrm{P}'[A] = \int_A X \; dP = \int_A Y \; dP$. □

23.2. Examples:
a) if $\mathcal{B} = \{\emptyset, \Omega\}$, then $\mathrm{E}[X|\mathcal{B}] = \mathrm{E}[X]$.
b) if $\mathcal{B} = \{\emptyset, \Omega, B, B^c\}$ then $E[X|\mathcal{B}]$ takes the value $\int_B X \; dP/\mathrm{P}[B]$ on B and the value $\int_{B^c} X \; dP/\mathrm{P}[B^c]$ on B^c.
c) if $\mathcal{B} = \mathcal{A}$, then $E[X|\mathcal{B}] = X$.
d) if $\mathcal{B} = \mathcal{A}_X$ is the σ algebra generated by X, then $E[X|\mathcal{B}] = X$.
e) if $X(x, y)$ is a continuous function on the unit square $\Omega = [0, 1]^2$ with $\mathrm{P} = dxdy$ as a probability measure and where $Y(x, y) = x$. In that case, $\mathrm{E}[X|Y]$ is a function of x alone, given by $\mathrm{E}[X|Y](x) = \int_0^1 f(x, y) \; dy$. It is called **a conditional integral**.

23.3. The map $X \in \mathcal{L}^1(\Omega, \mathcal{A}, \mathrm{P}) \to \mathrm{E}[X, \mathcal{B}] \in \mathcal{L}^1(\Omega, \mathcal{B}, \mathrm{P})$ is a **projection**. To see this geometrically, we work in the Hilbert space $\mathcal{L}^2$:

Theorem 2. *Conditional expectation $X \mapsto \mathrm{E}[X|\mathcal{B}]$ is the projection $\mathcal{L}^2(\mathcal{A}) \to \mathcal{L}^2(\mathcal{B})$.*

Proof. The space $\mathcal{L}^2(\mathcal{B})$ of square integrable $\mathcal{B}$-measurable functions is a linear subspace of $\mathcal{L}^2(\mathcal{A})$. When identifying functions which agree almost everywhere, then $L^2(\mathcal{B})$ is a Hilbert space which is a linear subspace of the Hilbert space $L^2(\mathcal{A})$. For any $X \in \mathcal{L}^2(\mathcal{A})$, there exists a unique projection $p(X) \in \mathcal{L}^2(\mathcal{B})$. The orthogonal complement $\mathcal{L}^2(\mathcal{B})^\perp$ is defined as

$$\mathcal{L}^2(\mathcal{B})^\perp = \{Z \in \mathcal{L}^2(\mathcal{A}) \mid (Z, Y) := \mathrm{E}[Z \cdot Y] = 0 \text{ for all } Y \in \mathcal{L}^2(\mathcal{B}) \} \; .$$

By the definition of the conditional expectation, we have for $A \in \mathcal{B}$

$$(X - \mathrm{E}[X|\mathcal{B}], 1_A) = \mathrm{E}[X - \mathrm{E}[X|\mathcal{B}]; A] = 0 \; .$$

Therefore $X - \mathrm{E}[X|\mathcal{B}] \in \mathcal{L}^2(\mathcal{B})^\perp$. Because the map $q(X) = \mathrm{E}[X|\mathcal{B}]$ satisfies $q^2 = q$, it is linear and has the property that $(1-q)(X)$ is perpendicular to $\mathcal{L}^2(\mathcal{B})$, the map q is a projection which must agree with p. □

Theorem 3 (Properties)**.** *For all $X, X_n, Y \in \mathcal{L}^1$:*
(1) Linearity: The map $X \mapsto \mathrm{E}[X|\mathcal{B}]$ is linear.
(2) Positivity: $X \geq 0 \Rightarrow \mathrm{E}[X|\mathcal{B}] \geq 0$.
(3) Tower property: $\mathcal{C} \subset \mathcal{B} \subset \mathcal{A} \Rightarrow \mathrm{E}[E[X|\mathcal{B}]|\mathcal{C}] = \mathrm{E}[X|\mathcal{C}]$.
(4) Cond. Fatou: $|X_n| \leq X$, $\mathrm{E}[\liminf_{n\to\infty} X_n|\mathcal{B}] \leq \liminf_{n\to\infty} \mathrm{E}[X_n|\mathcal{B}]$.
(5) Cond. dominated convergence: $|X_n| \leq X, X_n \to X$ a.e. $\Rightarrow \mathrm{E}[X_n|\mathcal{B}] \to \mathrm{E}[X|\mathcal{B}]$ a.e.
(6) Cond. Jensen: if h is convex, then $\mathrm{E}[h(X)|\mathcal{B}] \geq h(E[X|\mathcal{B}])$.
(7) Especially $||\mathrm{E}[X|\mathcal{B}]||_p \leq ||X||_p$.
(9) Extracting knowledge: For $Z \in \mathcal{L}^\infty(\mathcal{B})$, one has $\mathrm{E}[ZX|\mathcal{B}] = Z\mathrm{E}[X|\mathcal{B}]$.
(9) Independence: if X is independent of $\mathcal{C}$, then $\mathrm{E}[X|\mathcal{C}] = \mathrm{E}[X]$.

Proof. (1) The conditional expectation is a projection by the previous theorem so linear.
(2) If $Y = \mathrm{E}[X|\mathcal{B}]$ would be negative on a set of positive measure, then $A = Y^{-1}((-\infty, -1/n]) \in \mathcal{B}$ would have positive probability for some n. This would lead to the contradiction $0 \leq \mathrm{E}[1_A X] = \mathrm{E}[1_A Y] \leq -n^{-1} m(A) < 0$.
(3) Use that $P'' \ll P' \ll P$ implies $P'' = Y'P' = Y'YP$ and $P'' \ll P$ gives $P'' = ZP$ so that $Z = Y'Y$ almost everywhere.
This is especially useful when applied to the algebra $\mathcal{C}_Y = \{\emptyset, Y, Y^c, \Omega\}$. Because $X \leq Y$ almost everywhere if and only if $\mathrm{E}[X|\mathcal{C}_Y] \leq \mathrm{E}[Y|\mathcal{C}_Y]$ for all $Y \in \mathcal{B}$.
(4)-(5) The conditional versions of the Fatou lemma or the dominated convergence theorem are true, if they are true conditioned with $\mathcal{C}_Y$ for each $Y \in \mathcal{B}$. The tower property reduces these statements to versions with $\mathcal{B} = \mathcal{C}_Y$ which are then on each of the sets Y, Y^c the usual theorems.
(6) Chose a sequence $(a_n, b_n) \in \mathbb{R}^2$ such that $h(x) = \sup_n a_n x + b_n$ for all $x \in \mathbb{R}$. We get from $h(X) \geq a_n X + b_n$ that almost surely $\mathrm{E}[h(X)|\mathcal{G}] \geq a_n \mathrm{E}[X|\mathcal{G}] + b_n$. These inequalities hold therefore simultaneously for all n and we obtain almost surely

$$\mathrm{E}[h(X)|\mathcal{G}] \geq \sup_n (a_n \mathrm{E}[X|\mathcal{G}] + b_n) = h(E[X|\mathcal{G}]) \; .$$

(7) This is a special case of (6) using $h(x) = |x|^p$.
(8) It is enough to condition it to each algebra $\mathcal{C}_Y$ for $Y \in \mathcal{B}$. The tower property reduces these statements to linearity.
(9) By linearity, we can assume $X \geq 0$. For $B \in \mathcal{B}$ and $C \in \mathcal{C}$, the random variables $X1_B$ and 1_C are independent so that $\mathrm{E}[X1_{B\cap C}] = \mathrm{E}[X1_B 1_C] = \mathrm{E}[X1_B]\mathrm{P}[C]$. The random variable $Y = \mathrm{E}[X|\mathcal{B}]$ is $\mathcal{B}$ measurable and because $Y1_B$ is independent of $\mathcal{C}$ we get $\mathrm{E}[(Y1_B)1_C] = \mathrm{E}[Y1_B]\mathrm{P}[C]$ so that $\mathrm{E}[1_{B\cap C}X] = \mathrm{E}[1_{B\cap C}Y]$. The measures on $\sigma(\mathcal{B}, \mathcal{C})$

$$\mu : A \mapsto \mathrm{E}[1_A X], \nu : A \mapsto \mathrm{E}[1_A Y]$$

agree therefore on the π-system of the form $B \cap C$ with $B \in \mathcal{B}$ and $C \in \mathcal{C}$ and consequently everywhere on $\sigma(\mathcal{B}, \mathcal{C})$. □