# Stat 761

## Properties of Conditional Expectation

The "regression function" $\phi(x) = \mathrm{E}(Y|x)$ is a measurable function of $x$. In particular, $\mathrm{E}(Y|X) = \phi(X)$ *is a random variable and is a function of the random variable X.*

In the following (and in general) the abbreviation "a.s." stands for "almost surely" and means that a given statement holds "with probability one."

1. **Linearity:** $\mathrm{E}(aY|X) = a\mathrm{E}(Y|X)$ a.s..
2. **Additivity:** $\mathrm{E}(Y + Z|X) = \mathrm{E}(Y|X) + \mathrm{E}(Z|X)$ a.s.
3. **Constant-preservation:** $\mathrm{E}(1|X) = 1$ a.s.
4. **Positivity:** If $Y \geq 0$ a.s., then $\mathrm{E}(Y|X) \geq 0$ a.s.
5. **Contraction:** $|\mathrm{E}(Y|X)| \leq \mathrm{E}(|Y|\,|X)$ a.s.
6. **Least-squares property:** See problem 4.13.
7. **Conditional probability:** $\mathrm{P}(A|X) = \mathrm{E}(I_A|X)$. ($I_A$ is the indicator of the event $A$.)
8. Analog of the Theorem of the Unconscious Statistician holds for $\mathrm{E}(g(Y)|X)$.
9. **Composition:** $\mathrm{E}(\mathrm{E}(Y|X)) = \mathrm{E}(Y)$ (Theorem 4.4.3, p.164).
10. **Averaging** (or **filtering**): If $h(x)$ is a (measurable) function, then

$$\mathrm{E}(h(X) \cdot Y|X) = h(X) \cdot \mathrm{E}(Y|X) \;\text{ a.s.}$$

    This is a generalization of **linearity**.

To define conditional expectation in full generality, one does this all backwards! One *first* defines $\mathrm{E}(Y|X)$ as a random variable which is a function of $X$ and which satisfies a certain property, namely:

$$\int_A \mathrm{E}(Y|X)dP = \int_A Y dP \;\text{ for every measurable set } A.$$

*Then* the regression function $\mathrm{E}(Y|x)$ is defined as the value of the random variable $\mathrm{E}(Y|X)$ on the event $X = x$. Conditional probability is defined *last of all* by the formula $\mathrm{P}(A|X) = \mathrm{E}(I_A|X)$, where $I_A$ is the indicator of the event $A$, and $\mathrm{P}(A|x)$ is defined accordingly.

*After* this is done, one has to show that under appropriate conditions on the distribution of $(X, Y)$, $\mathrm{P}(Y \in B|x)$ is computed by summing or integrating where the summand/integrand is precisely the conditional pmf/pdf. All this is properly the domain of a course in measure-theoretic probability (for example, Math 771). See also the texts by Tucker (my favorite) and Loève.