Steiner's theorem and least squares

Márton Balázs* and Bálint Tóth*

October 19, 2014

This little write-up is part of important foundations of probability that were left out of the unit Probability 1 due to lack of time and prerequisites. Here we prove Steiner's theorem and apply it to prove that the conditional expectation can be used to find the least square estimator of a random variable in a conditional setting.

The question we investigate is, first in general and later in a conditional setting, the best guess we can make for the value of a random variable X. We denote this guess by c, and this is a good guess if $X - c$ is small. But this latter being a random quantity we should specify this a bit more. The first idea that comes to mind is to look for c which makes $\mathbf{E}|X-c|$ minimal. This can be done, and the outcome is that the minimising c will be the *median* of the distribution of X. A non-negative function that is more handy than absolute values is the square function, we'll use this one below. We thus look for the number c that makes $\mathbf{E}(X-c)^2$ minimal. The following theorem is well known in physics in the context of moments of inertia.

Theorem 1 (Steiner's theorem) *Let X be a random variable, then*

$$\mathbf{E}(X-c)^2 = \mathbf{Var}X + (c - \mathbf{E}X)^2.$$

Proof.

$$\begin{aligned}\mathbf{E}(X-c)^2 &= \mathbf{E}\big((X-\mathbf{E}X)+(\mathbf{E}X-c)\big)^2 = \mathbf{E}(X-\mathbf{E}X)^2 + \mathbf{E}(\mathbf{E}X-c)^2 + \mathbf{E}\big(2(X-\mathbf{E}X)\cdot(\mathbf{E}X-c)\big) = \\ &= \mathbf{Var}X + \mathbf{E}(c-\mathbf{E}X)^2 + 2(\mathbf{E}X-c)\cdot\mathbf{E}(X-\mathbf{E}X) = \mathbf{Var}X + \mathbf{E}(c-\mathbf{E}X)^2 + 0.\end{aligned}$$

□

Corollary 2 *The smallest expected square deviation of X is from its mean, and agrees with the variance.*

Next we investigate the same problem in terms of conditional distributions. Let X and Y be random variables, and suppose that the value of Y is given. This can provide partial information on X if the variables are dependent on each other. The question is our best guess on the value of X, given Y. This guess $c(Y)$ is allowed to depend on Y this time. Therefore we seek the function $c(Y)$ that minimises the conditional square deviation

$$\mathbf{E}\big((X-c(Y))^2 \,|\, Y\big).$$

As the conditional expectation is a proper expectation, repeating the above arguments gives that the optimal guess is $c(Y) = \mathbf{E}(X\,|\,Y)$, and the conditionally expected square deviation from this is just the conditional variance $\mathbf{Var}(X\,|\,Y)$.

*University of Bristol / Budapest University of Technology and Economics