ST908 tutorial: some remarks on conditional expectation

Suppose X is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E}|X| < \infty$ and $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$. The conditional expectation of X given $\mathcal{G}$ is the random variable Z (which we write $Z = \mathbb{E}(X|\mathcal{G})$) such that

1. $\mathbb{E}|Z| < \infty$;
2. Z is $\mathcal{G}$-measurable;
3. $\mathbb{E}(X1_G) = \mathbb{E}(Z1_G)$ for all $G \in \mathcal{G}$.

In this definition, the motivations of the first two properties are quite trivial - the first property excludes some irregular cases where the expectation blows up, and the second property is just saying that we are able to determine the value of $\mathbb{E}(X|\mathcal{G})$ given the information of $\mathcal{G}$. The meaning behind the last property is less obvious but we could still get some intuitions informally by looking at a simple example from geometry as an analogue.

1 A simple geometry problem

Consider this question: on a two-dimensional plane we are given a point $X = (1, 3)$ and a straight line $\mathcal{G} : y = x$. Let $Z = (z_1, z_2)$ be a point on the line $\mathcal{G}$. What should Z be if the distance between X and Z has to be minimised?

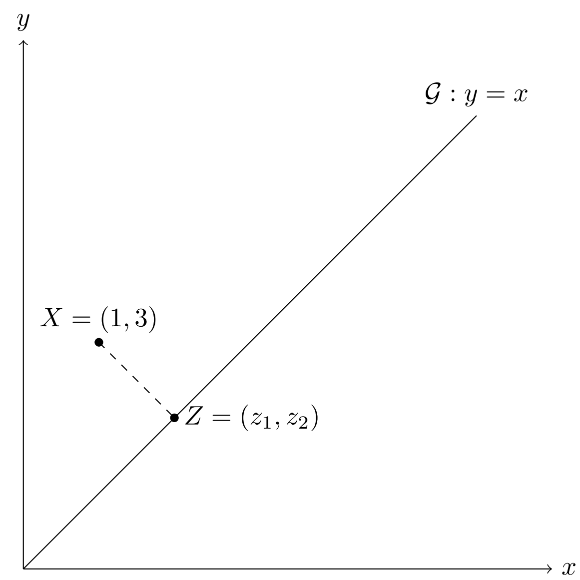

Figure 1: Z is the point on $\mathcal{G}$ which minimises the distance between X and Z.

Obviously, if the distance between X and Z is minimised, then the straight line joining X and Z should be perpendicular to $\mathcal{G}$. Recall that we can summarise the dependence between two vectors $\overrightarrow{u}$ and $\overrightarrow{v}$ by the angle θ between them which is given by

$$\frac{\overrightarrow{u} \cdot \overrightarrow{v}}{|\overrightarrow{u}||\overrightarrow{v}|} = \cos\theta.$$

If $\overrightarrow{u}$ and $\overrightarrow{v}$ are perpendicular to each other, we have $\theta = 90°$ and thus $\overrightarrow{u} \cdot \overrightarrow{v} = 0$.

Let $\overrightarrow{X} = (1, 3)$ and $\overrightarrow{Z} = (z_1, z_2)$ be the vector representations of X and Z. Then the straight line joining X and Z can be represented by the vector $\overrightarrow{Z} - \overrightarrow{X} = (z_1 - 1, z_2 - 3)$. This vector is perpendicular to $\mathcal{G}$. Thus if we take any vector $\overrightarrow{G}$ along $\mathcal{G}$, say choose $\overrightarrow{G} = (1, 1)$, we have

$$(\overrightarrow{Z} - \overrightarrow{X}) \cdot \overrightarrow{G} = 0$$

and hence

$$(z_1 - 1) + (z_2 - 3) = 0.$$

Together with the fact that Z lies on $\mathcal{G}$ which suggests $z_1 = z_2$, we obtain $z_1 = z_2 = 2$.

2 Conditional expectation as the "best predictor"

Now consider a different question: on a probability triple $(\Omega, \mathcal{F}, \mathbb{P})$, there is a random variable X. Suppose you want to know the value of X but you are only given a partial information set $\mathcal{G}$ which may not be sufficient to tell you the exact value of X. What is your "best predictor" of X based on the partial information $\mathcal{G}$?

Ideally your "best predictor", call it Z, should minimise some kind of "distance measure". One example criteria could be the average squared difference of

$$\mathbb{E}\left((Z - X)^2\right).$$

This problem is very similar to our previous geometry example: we are given a point (random variable) X, and we would like to find a point on $\mathcal{G}$ (a $\mathcal{G}$-measurable random variable) Z such that the geometric distance between X and Z (the average squared difference $\mathbb{E}\left((Z - X)^2\right)$) is minimised. Motivated by the geometry example, the best predictor Z should satisfy the property that "$(Z - X)$ is perpendicular to all the random variables on $\mathcal{G}$". But what does it mean?

Of course, we no longer have the nice geometric interpretation of " orthogonality" once we work with random variables. But recall how we measure the degree of dependence between two vectors in a two-dimensional setting by $\cos\theta$. For random variables U and V, a good analogue is the correlation measure which is

$$\frac{\mathbb{E}(UV)}{\sqrt{\mathbb{E}(U^2)\mathbb{E}(V^2)}} = corr(U, V).$$

Here $\mathbb{E}(UV)$ plays a similar role as $\overrightarrow{u} \cdot \overrightarrow{v}$ while $\sqrt{\mathbb{E}(U^2)}$ and $\sqrt{\mathbb{E}(V^2)}$ can be interpreted as "measures of length" just like $|\overrightarrow{u}|, |\overrightarrow{v}|$. The orthogonal feature between $Z - X$ and the elements on $\mathcal{G}$ can then be represented by the condition

$$\mathbb{E}((Z - X)\widetilde{G}) = 0$$

for any $\mathcal{G}$-measurable random variable $\widetilde{G}$. To simplify things, it is sufficient to consider $\widetilde{G}$ in form of indicator function 1_G for any $G \in \mathcal{G}$. Thus we have

$$\mathbb{E}((Z - X)1_G) = 0$$

or equivalently

$$\mathbb{E}(X1_G) = \mathbb{E}(Z1_G).$$

In summary, the third property in the definition of conditional expectation is some kind of "orthogonality criteria". With this condition in place, we could interpret $\mathbb{E}(X|\mathcal{G})$ as the "best predictor" of X given the information of $\mathcal{G}$ ("best" in the sense that its average squared difference with X is minimised).