Equivalence of Conditional Expectation and Orthongonal Projection for jointly normal random variables

Phanindra Jampana

October 18, 2011

It can be shown that $f(Y_1, Y_2, \cdots, Y_n) = E[X|Y_1, Y_2, \cdots, Y_n]$ is the best MMSE estimator of the random variable X given the observations $Y_1, Y_2, \cdots, Y_n$.

Let $X, Y_1, Y_2, \cdots, Y_n$ be jointly normal such that $(X, Y_1, Y_2, \cdots, Y_n) \sim N(\mu, \Sigma)$ and denote the vector of random variables $(Y_1, Y_2, \cdots, Y_n)$ with Y. Similarly we use the notation y for the vector $(y_1, y_2, \cdots, y_n)$. Then, we know that

$$f_{X|Y_1,Y_2,\cdots,Y_n}(x, y_1, y_2, \cdots, y_n) \sim N(E[X]+\Sigma_{XY}\Sigma_{YY}^{-1}(y-E[Y]), \Sigma_{XX}-\Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX})$$

The conditional expectation given the values of $Y_1, Y_2, \cdots, Y_n$ is:

$$\begin{aligned} E[X|Y_1 = y_1, Y_2 = y_2, \cdots Y_n = y_n] &= \int_{-\infty}^{\infty} x f_{X|Y_1,Y_2,\cdots,Y_n}(x, y_1, y_2, \cdots, y_n) dx \\ &= E[X] + \Sigma_{XY}\Sigma_{YY}^{-1}(y - E[Y]) \end{aligned}$$

Hence, conditional expectation can be defined as :

$$E[X|Y] = E[X] + \Sigma_{XY}\Sigma_{YY}^{-1}(Y - E[Y])$$

We would like to show that the random variable $E[X|Y]$ is just the projection of X on the vector space spanned by the vectors in Y i.e. the vector space spanned by $Y_1, Y_2, \cdots, Y_n$. To be able to talk about projections we need an inner product on the space of random variables. This is defined for any random variables V, W as:

$$\langle V, W \rangle = E[VW]$$

It can be easily verified that this definition of the inner product satisfies the all the criterion for an inner product.

Let the projection of X on to the vector space spanned by $Y_1, Y_2, \cdots, Y_n$ be denoted by X_1. Then we know from definition that $X - X_1$ is orthogonal to every vector Y_j. In this note we will just be showing that if $X_1 = E[X|Y]$ then $X - X_1$ is indeed perpendicular to all the vectors Y_j. As the projection is unique, the final result immediately follows.

$$
\begin{aligned}
E[(X - X_1)Y_j] &= E[(X - E[X|Y])Y_j] \\
&= E[XY_j] - E[E[X|Y]Y_j] \\
&= E[XY_j] - E[(E[X] + \Sigma_{XY}\Sigma_{YY}^{-1}(Y - E[Y]))Y_j] \\
&= E[XY_j] - E[X]E[Y_j] + E[Y_j]\Sigma_{XY}\Sigma_{YY}^{-1}E[Y] - E(Y_j\Sigma_{XY}\Sigma_{YY}^{-1}Y) \\
&= E[(X - E[X])(Y_j - E[Y_j])] + \Sigma_{XY}\Sigma_{YY}^{-1}(E[Y_j]E[Y]) - \Sigma_{XY}\Sigma_{YY}^{-1}E[YY_j]) \\
&= E[(X - E[X])(Y_j - E[Y_j])] - \Sigma_{XY}\Sigma_{YY}^{-1}(E[(Y - E[Y])(Y_j - E[Y_j]))
\end{aligned}
$$

Note now that $E[(Y-E[Y])(Y_j-E[Y_j])$ is the jth column in the covariance matrix Σ_{YY}. Σ_{YY}^{-1} multiplied by that column is just the vector $\{0, 0, 0, \cdots, 1, 0, 0, \cdots, 0\}$ where the 1 is located at the jth co-ordinate. Σ_{XY} times this vector is just the element of Σ in the first row and $j + 1$th column. By definition the first term in the last expression is also the element of Σ in the first row and the $j + 1$th location. Therefore, the result is zero.