Optimum Nonlinear Estimation

Consider two random variables \( X(\lambda) \) and \( Y(\lambda) \) defined on the same sample space \( S \). We have previously looked at the optimal linear MMSE estimate of \( Y \) in terms of \( X \). In this exercise we will look at the form of the optimal nonlinear estimate of \( Y \) given the observation \( X \). Specifically let us look at the conditional mean estimate given by:

\[
\hat{Y} = g(X) = E\{Y|X\} = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy.
\]

Note that this estimate does not depend on \( Y \) and depends only on the conditioned value of \( X \), since the variable \( y \) has been integrated out. The mean-squared error incurred by this estimate is given by:

\[
\epsilon_g^2 = E\{(Y - g(X))^2\}
\]

Consider another estimate of \( Y \) given by \( \hat{Y}_2 = h(X) \). The corresponding error incurred by this estimate is given by:

\[
\epsilon_h^2 = E\{(Y - h(X))^2\}
\]

Relating this error to the error incurred by the conditional mean estimator we have:

\[
\epsilon_h^2 = E\{(Y - g(X) + g(X) - h(X))^2\}.
\]

This can further be written as the sum of three terms:

\[
\epsilon_h^2 = \epsilon_g^2 + 2E\{(Y - g(X))(g(X) - h(X))\} + E\{(g(X) - h(X))^2\}.
\]

The third term is always positive as is the first term. We will look at the second term in detail. Using iterated expectation we can rewrite this:

\[
E\{(Y - g(X))(g(X) - h(X))\} = E_X \left[ E_{Y|X} \{(Y - g(X))(g(X) - h(X))\}\right]
\]

Evaluating the inner expectation we have:

\[
E_{Y|X} \{(Y - g(X))(g(X) - h(X))\} = (g(X) - h(X))E_{Y|X} \{Y - g(X)\} = 0.
\]

We can therefore write the error incurred by the estimate \( h(X) \) as

\[
\epsilon_h^2 = \epsilon_g^2 + E\{(g(X) - h(X))^2\} \longleftarrow \epsilon_h^2 \geq \epsilon_g^2.
\]

The statement made above has two important implications:
1. The conditional mean estimator is the best non-linear estimate of $Y$ given the observations $X$.

2. Any other estimator be it linear or nonlinear will always have an estimation error larger than the conditional mean estimator.

3. In the special case where $X$ and $Y$ are jointly Gaussian then the optimal non-linear estimate becomes the optimal linear estimate.