Model Order Selection Using F-Test

Alper Yazar

The selection of a suitable model and its parameters is a fundamental problem arising in many signal processing applications such as pole-zero modeling of deterministic signals, inverse filtering, data analysis etc. The main problem of modeling is the selection of a simple enough mathematical expression explaining the phenomena of interest.

In this work, we focus on a sequence of nested models where the more sophisticated model (the model with a large number of parameters) encapsulates the less sophisticated ones as a special case and examine the appropriate model order. We consider the linear models with unknown parameters for this purpose. A typical example is the polynomial fitting to a set of measurements. The selection of the polynomial order can be considered as the model order selected with the nested models.

Different from the classical model order selection problem, we examine the suitability of a particular model for a given application. As an illustrative application, we may focus on the determination of zero-crossing point of the sampled data. Figure 1 presents an example for zero-crossing point estimation for a sampled sinusoid. In this figure, the noiseless and noisy samples (at SNR of 20 dB) are shown. The goal is to estimate the zero-crossing point indicated in the figure. Typically, the zero-crossing point is located by fitting a straight line to the two samples with different signs. This approach assumes that the signal is sampled "fast" enough that the line fit is sufficient to determine the zero-crossing point. In the presence of noise, the zero-crossing point is detected with some error due to errors in line parameters. A reasonable idea is to use more than 2 points in line-fitting by a least squares type method. By using more points, the zero-crossing estimation error is expected to decrease provided that the samples utilized for line-fitting are indeed on a line. As in Figure 1, using large number of points may bring an estimation bias, since the imposed line structure does not match the signal structure. The goal is to use "right" number of points to reduce the zero-crossing estimation error.

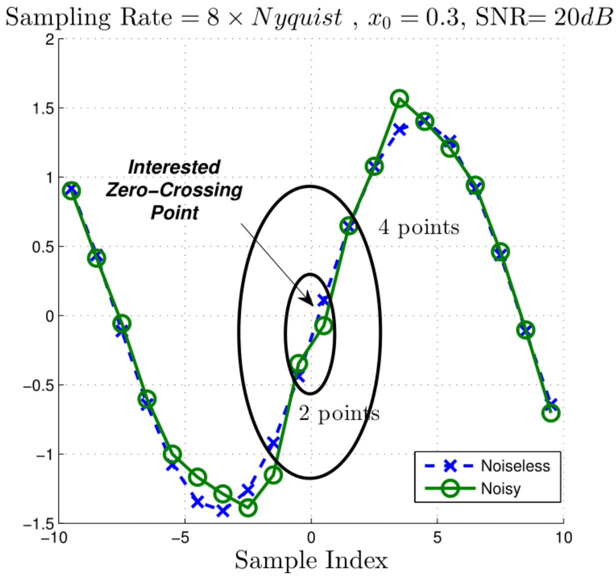

Figure 1. Noisy and Noiseless Sinusoidal Signal

In this study, we use a method known as F-Test method to check whether higher order model decreases RMSE in a significant way. In F-Test, two models are compared to each other. This comparison gives a number and that number is compared with a threshold value. Threshold is calculated based on criteria that how much RMSE difference between two models due to model order difference is acceptable. F-Test doesn't require actual expression of observed data and noise variance to determine threshold value. This is a benefit for cases where there is no a priori information about data and noise.