Module 2

Random Processes

Lesson 9

Introduction to Statistical Signal Processing

After reading this lesson, you will learn about

- ***Hypotheses testing***
- ***Unbiased estimation based on minimum variance***
- ***Mean Square Error (MSE)***
- ***Crammer - Rao Lower Bound (CRLB)***

As mentioned in Module #1, demodulation of received signal and taking best possible decisions about the transmitted symbols are key operations in a digital communication receiver. We will discuss various modulation and demodulation schemes in subsequent modules (Modules # 4 and #5). However, a considerable amount of generalization is possible at times and hence, it is very useful to have an overview of the underlying principles of statistical signal processing. For example, the process of signal observation and decision making is better appreciated if the reader has an overview of what is known as 'hypothesis testing' in detection theory. In the following, we present a brief treatise on some fundamental concepts and principles of statistical signal processing.

Hypothesis Testing

A hypothesis is a statement of a possible source of an observation. The observation in a statistical hypothesis is a random variable. The process of making a decision from an observation on 'which hypothesis is true (or correct)', is known as hypothesis testing.

Ex.#2.9.1. Suppose it is known that a modulator generates either a pulse 'P_1' or a pulse 'P2' over a time interval of 'T' second and 'r' is received in the corresponding observation interval of 'T' sec. Then, the two hypotheses of interest may be,

H_1: 'P_1' is transmitted.
H_2: 'P_2' is transmitted.

Note that 'P_1 is not transmitted' is equivalent to 'P_2 is transmitted' and vice versa, as it is definitely known that either 'P_1'or'P_2' has been transmitted.

Let us define a parameter (random variable) 'u' which is generated in the demodulator as a response to the received signal 'r', over the observation interval. The parameter 'u' being a function of an underlying random process, is defined as a random variable (single measurement over T) and is called an 'observation' in the context of hypothesis testing. An 'observation' is also known as a 'decision variable' in the jargon of digital communications. The domain (range of values) of 'u' is known as 'observation space'. The relevant hypothesis is, 'making a decision on whether H_1 or H_2 is correct' (upon observing 'u'). The observation space in the above example is the one dimensional real number axis. Next, to make decisions, the whole observation space is divided in appropriate 'decision regions' so as to associate the possible decisions from an observation 'r' with these regions.

Note that the observation space need not be the same as the range of pulses P_1 or P_2. However, decision regions can always be identified for the purpose of decision-making. If a decision doesn't match with the corresponding true hypothesis, an error is said to have occurred. An important aim of a communication receiver is to clearly identify the decision regions such that the (decision) errors are minimized.

Estimation

There are occasions in the design of digital communications systems when one or more parameters of interest are to be estimated from a set of signal samples. For example, it is very necessary to 'estimate' the frequency (or, instantaneous phase) of a received carrier-modulated signal as closely as possible so that the principle of 'coherent demodulation' may be used and the 'decision' errors can be minimized.

Let us consider a set of N random, discrete-time samples or data points, {x[0], x[1], x[2], …,x[N-1]} which depends (in some way) on a parameter θ which is unknown. The unknown parameter θ is of interest to us. Towards this, we express an 'estimator' $\hat{\theta}$ as a function of the data points:

$$\hat{\theta} = f\,(x[0], x[1], x[2], \ldots, x[N-1]) \qquad 2.9.1$$

Obviously, the issue is to find a suitable function f(.) such that we can obtain θ from our knowledge of $\hat{\theta}$ as precisely as we should expect. Note that, we stopped short of saying that at the end of our estimation procedure, '$\hat{\theta}$ will be equal to our parameter of interest θ'.

Next, it is important to analyze and mathematically model the available data points which will usually be of finite size. As the sample points are inherently random, it makes sense to model the data set in terms of some family of probability density function [pdf] , parameterized by the parameter θ: p(x[0], x[1], x[2], …,x[N-1] ; θ). It means that the family of pdf is dependent on 'θ' and that different values of θ may give rise to different pdf-s. The distribution should be so chosen that the mathematical analysis for estimation is easier. A Gaussian pdf is a good choice on several occasions. In general, if the pdf of the data depends strongly on θ, the estimation process is likely to be more fruitful.

A classical estimation procedure assumes that the unknown parameter θ is deterministic while the Bayesian approach allows the unknown parameter to be a random variable. In this case, we say that the parameter, being estimated, is a 'realization' of the random variable. If the set of sample values {x[0], x[1], x[2], …,x[N-1]} is represented by a data vector **x,** a joint pdf of the following form is considered before formulating the procedure of estimation:

$$p(\mathbf{x}, \theta) = p(\mathbf{x} \mid \theta).\, p(\theta) \qquad 2.9.2$$

Here $p(\theta)$ is the prior pdf of θ, the knowledge of which we should have before observing the data and $p(\mathbf{x} \mid \theta)$ is the conditional pdf of **x**, conditioned on our knowledge of θ. With these interpretations, an estimator may now be considered as a rule that assigns a value to θ for each realization of data vector **x** and the 'estimate' of θ is the value of θ for a specific realization of the data vector **x**. An important issue for any estimation is to assess the performance of an estimator. All such assessment is done statistically. As the process of estimation involves multiple computations, sometimes the efficacy of an estimator is decided in a practical application by its associated computational complexity. An 'optimal' estimator may need excessive computation while the performance of a suboptimal estimator may be reasonably acceptable.

An Unbiased Estimation based on Minimum Variance:

Let us try to follow the principle of a classical estimation approach, known as unbiased estimation. We expect the estimator to result in the true value of the unknown parameter on an average. The best estimator in this class will have minimum variance in terms of estimation error $(\hat{\theta} - \theta)$. Suppose we know that the unknown parameter θ is to lie within the interval $\theta_{min} < \theta < \theta_{max}$. The estimate is said to be unbiased if $E[\hat{\theta}] = \theta$ in the interval $\theta_{min} < \theta < \theta_{max}$. The criterion for optimal estimation that is intuitive and easy to conceive is the 'mean square error (MSE)', defined as below:

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] \qquad 2.9.3$$

However, from practical considerations, this criterion is not all good as it may not be easy to formulate a good estimator (such as 'minimum MSE') that can be directly expressed in terms of the data points. Usually, the MSE results in a desired variance term and an undesired 'bias' term, which makes the estimator dependent on the unknown parameter (θ). So, summarily, the intuitive 'mean square error (MSE)' approach does not naturally lead to an optimum estimator unless the bias is removed to formulate an 'unbiased' estimator.

A good approach, wherever applicable, is to constrain the bias term to zero and determine an estimator that will minimize the variance. Such an estimator is known as a 'minimum variance unbiased estimator'.

The next relevant issue in classical parameter estimation is to find the estimator. Unfortunately, there is no single prescription for achieving such an optimal estimator for all occasions. However, several powerful approaches are available and one has to select an appropriate method based on the situation at hand. One such approach is to determine the Crammer-Rao lower bound (CRLB) and to check whether an estimator at hand satisfied this bound. The CRLB results in a bound over the permissible range of the unknown parameter θ such that the variance of any unbiased estimator will be equal or greater than this bound.

Further, if an estimator exists whose variance equals the CRLB over the complete range of the unknown parameter θ, then that estimator is definitely a minimum variance unbiased estimator. The theory of CRLB can be used with reasonable ease to see for an application whether an estimator exists which satisfies the bound. This is important in modeling and performance evaluation of several functional modules in a digital receiver.

The CRLB theorem can be stated in multiple ways and we choose the simple form which is applicable when the unknown parameter is a scalar.

Theorem on Crammer- Rao Lower Bound

The variance of any unbiased estimator $\hat{\theta}$ will satisfy the following inequality provided the assumptions stated below are valid:

$$\operatorname{var}(\hat{\theta}) \geq 1 / \{ -E [(\partial^2 \ln p(\mathbf{x};\theta))/\partial\theta^2] \} = 1/ \{E[(\partial \ln p(\mathbf{x};\theta)/\partial\theta)^2]\} \qquad 2.9.4$$

Associated assumptions:

a) all statistical expectation operations are with respect to the pdf $p(\mathbf{x};\theta)$,

b) the pdf $p(\mathbf{x};\theta)$ also satisfies the following 'regularity' condition for all θ:

$$E [(\partial \ln p(\mathbf{x};\theta))/\partial\theta] = \int \{ (\partial \ln p(\mathbf{x};\theta))/\partial\theta \}.\, p(\mathbf{x};\theta)d\mathbf{x} = 0 \qquad 2.9.5$$

Further, an unbiased estimator, achieving the above bound may be found only if the following equality holds for some functions g(.) and I(.):

$$\partial \ln p(\mathbf{x};\theta))/\partial\theta = I(\theta).\{g(\mathbf{x}) - \theta \} \qquad 2.9.6$$

The unbiased estimator is given by $\hat{\theta} = g(\mathbf{x})$ and the resulting minimum variance of $\hat{\theta}$ is the reciprocal of $I(\theta)$, i.e. $1/I(\theta)$.

It is interesting to note that, $I(\theta) = -E [(\partial^2 \ln p(\mathbf{x};\theta))/\partial\theta^2] = E[(\partial \ln p(\mathbf{x};\theta)/\partial\theta)^2]$ and is known as Fisher information, associated with the data vector '**x**'.

CRLB for signals in White Gaussian Noise

Let us denote time samples of a deterministic signal, which is some function of the unknown parameterθ, as $y[n;\theta]$. Let us assume that we have received the samples after corruption by white Gaussian noise, denoted by $\omega[n]$, as:

$$x[n] = y[n;\theta] + \omega[n], \qquad n = 1, 2, \ldots, N$$

We wish to determine an expression for the Crammer-Rao Lower Bound with the above knowledge. Note that we can express $\omega[n]$ as $\omega[n] = x[n] - y[n;\theta]$.

As ω[n] is known to follow Gaussian probability distribution, we may write,

$$p(\mathbf{x};\theta) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[\frac{-1}{2\sigma^2}\sum_{n=1}^{N}(x[n] - y[n;\theta])^2\right] \qquad 2.9.7$$

Partial derivative of p(**x**; θ) with respect to θ gives:

$$\frac{\partial \ln p(x;\theta)}{\partial\theta} = \frac{1}{\sigma^2}\cdot \sum_{n=1}^{N}(x[n] - y[n;\theta])\frac{\partial y[n;\theta]}{\partial\theta} \qquad 2.9.8$$

Similarly, the second derivative leads to the following expression:

$$\frac{\partial^2 \ln p(x;\theta)}{\partial\theta^2} = \frac{1}{\sigma^2}\cdot \sum_{n=1}^{N}\left\{ (x[n] - y[n;\theta])\frac{\partial^2 y[n;\theta]}{\partial\theta^2} - \left(\frac{\partial y[n;\theta]}{\partial\theta}\right)^2\right\} \qquad 2.9.9$$

Now, carrying out the expectation operation gives,

$$E\left[\frac{\partial^2 \ln p(x;\theta)}{\partial\theta^2}\right] = -\frac{1}{\sigma^2}\cdot \sum_{n=1}^{N}\left(\frac{\partial y[n;\theta]}{\partial\theta}\right)^2 \qquad 2.9.10$$

Now, following the CRLB theorem, we see that the bound is given by the following inequality:

$$\mathrm{var}(\hat{\theta}) \geq \frac{\sigma^2}{\sum_{n=1}^{N}\left(\frac{\partial y[n;\theta]}{\partial\theta}\right)^2} \qquad 2.9.11$$

Problems

Q2.9.1) Justify why "Hypotheses testing" may be useful in the study of digital communications.

Q2.9.2) What is MSE? Explain its significance.