Estimation Notes 1 pp.1 - 12

Key Points : Estimation in Signal Processing; The Mathematical Estimation Problem; Assessing Estimator Performance

1. Estimation in Signal Processing

1. **Modern Estimation Theory** can be found at the heart of many electronic signal processing systems designed to **extract information**.

2. Three examples: **Radar, Sonar, Speech**.

3. In all cases we are faced with the problem of extracting values of parameters based on continuous-time waveforms. Due to the use of digital computers to sample and store the continuous-time waveform, we have the equivalent problem of extracting parameter values from a **discrete-time** waveform or a **data set**.

4. We have the N-point data set $\{x[0], x[1], \ldots, x[N-1]\}$ which depends on an **unknown parameter** θ. We wish to determine θ based on the data or to define an estimator

$$\hat{\theta} = g(x[0], x[1], \ldots, x[N-1]) \tag{1.2}$$

where g is some function. This is what we called **Parameter Estimation**.

2. The Mathematical Estimation Problem

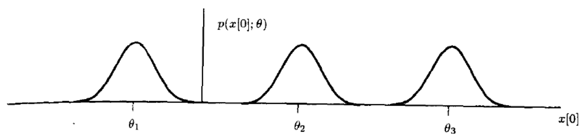

Figure 1.5 Dependence of PDF on unknown parameter

5. Since the data are **Inherently Random**, we describe it by its **Probability Density Function (PDF)** or $p(x[0], x[1], \ldots, x[N-1]; \theta)$

6. The PDF is parameterized by the unknown parameter θ, i.e., we have a class of PDFs where each one is different due to a different value θ. We will use a semicolon(;) to denote this **dependence**. As an example, if N = 1 and θ denotes the mean, then the PDF of data might be

$$p(x[0];\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x[0]-\theta)^2\right]$$

where is shown in Figure 1.5 for various of θ.

7. It should be clear that because the value of θ affects the probability of $x[0]$, we should be able to **infer the value of θ from the observed value of $x[0]$**.

8. Consider Figure 1.5, if the value of $x[0]$ is negative, it is doubtful that $\theta = \theta_2$. The value $\theta = \theta_1$ might be more reasonable. This specification of the PDF is critical in determining a good estimator.

9. In an actual problem we are not given a PDF but must choose one that is **not only consistent with the problem constraints and any prior knowledge, but one that is also mathematically tractable**.

10. Consider Figure 1.6, we might be conjectured that this data, although appearing to fluctuate wildly, actually is `\textbf{on the average}` increasing.

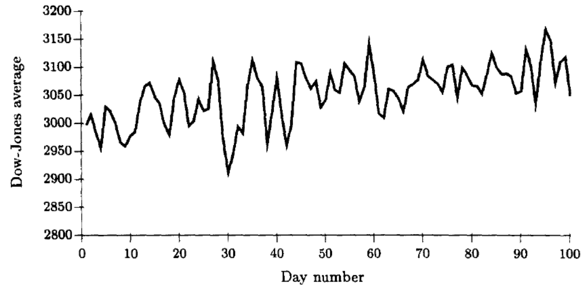

Figure 1.6 Hypothetical Dow-Jones average

11. To determine if this is true **we could assume that the data actually consist of a straight line embedded in random noise** or

$$x[n] = A + Bn + w[n] \quad n = 0, 1, \dots, N-1$$

where $w[n] \sim \mathcal{N}(0, \sigma^2)$(i.e., $w[n]$ is **white Gaussian Noise (WGN)**).

12. Then, the unknown parameters are A and B, which arranged as a vector become the vector parameter $\theta = [AB]^T$. Letting $\mathbf{x} = [x[0], x[1], \dots, x[N-1]]^T$, the PDF is

$$p(\mathbf{x}; \theta) = \frac{1}{(\sqrt{2\pi\sigma^2})^N} \exp\left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A - Bn)^2 \right] \qquad 1.3$$

13. A models the Dow-Jones average which is hovering around 3000 and the conjecture that it is increasing ($B > 0$ models this).

14. The assumption of WGN is making the mathematically model more tractable. It is reasonable unless there is strong evidence to the contrary, such as highly correlated noise.

15. The performance of any estimator obtained will be **Critically Dependent** on the PDF assumptions.

16. Estimation based on PDFs such as (1.3) is termed **Classical Estimation** in that the parameters of interest are assumed to be **Deterministic** but **Unknown**.

17. If we know a priori that the mean is somewhere around 3000, we might be more willing to constrain the estimator to produce values of A in the range $[2800, 3200]$. To **incorporate** this prior knowledge we can assume that A **is no longer deterministic but a random variable and assign it a PDF**, possibly uniform over the [2800, 3200] interval.

18. Then, any subsequent estimator will yield values in this range. Such an approach is termed **Bayesian Estimation**. The parameter θ we are attempting to estimate is then viewed as a **Realization** of the random variable θ. As such, the data are described by the **Joint PDF**

$$p(\mathbf{x}, \theta) = p(\mathbf{x}|\theta)p(\theta)$$

where $p(\theta)$ refers to the **prior PDF**, summarizing our knowledge about θ before any data are observed, and $p(\mathbf{x}|\theta)$ is a **conditional PDF**, summarizing our knowledge provided by the data $\mathbf{x}$ conditioned on knowing θ.

19. We should clear that $p(\mathbf{x}; \theta)$ means a family of PDFs; $p(\mathbf{x}|\theta)$ is a conditional PDF; and $p(\mathbf{x}, \theta)$ denotes a joint PDF.

20. Once the PDF has been specified, the problem becomes one of **determining an optimal estimator or function of the data**, as in (1.2)

$$\hat{\theta} = g(x[0], x[1], \ldots, x[N-1]) \tag{1.2}$$

21. Note that an estimator may depend on other parameters, but only if they are **known**.

22. An estimator may be thought of as **a rule that assigns a value to θ for each realization of $\mathbf{x}$**.

23. The estimate of θ is the **value of θ obtained for a given realization of $\mathbf{x}$**.

3. Assessing Estimator Performance

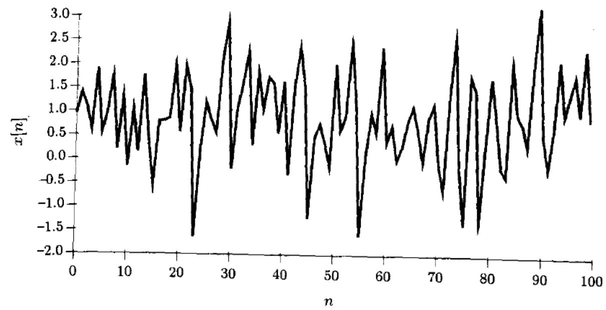

Figure 1.7 Realization of DC (Direct Current) level in noise.

24. We could model the Direct Current (DC) as

$$x[n] = A + w[n]$$

where $w[n]$ denotes some zero mean noise process. Based on the data set $\{x[0], x[1], \ldots, x[N-1]\}$, we would like to estimate A. It would be reasonable to estimate A as

$$\hat{A} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$$

25. From the data set in Figure 1.7, we can know that $\hat{A} = 0.9$, which is close to the true value of $A = 1$. Consider another estimator might be

$$\check{A} = x[0]$$

Intuitively, we would not expect this estimator to perform as well since it does not make use of all the data. In other words, there is **no averaging** to reduce the noise effects. However, $\check{A} = x[0] = 0.95$, which is closer to the true value of $A = 1$ than the sample mean estimate $\hat{A} = 0.9$. Can we say that $\check{A}$ is a better estimator than $\hat{A}$?

26. The answer is NO. The reason is **an estimator is a function of the data, which are random variables. The estimator is also a random variable, subject to many possible outcomes**.

27. The fact that $\check{A}$ is closer to the true value only means that **for the given realization of data**, as shown in Figure 1.7, the estimate $\check{A} = 0.95$ is closer to the true value than the estimate $\hat{A} = 0.9$. **To assess performance we must do so statistically**.

28. Let's prove that $\hat{A}$ is better than $\check{A}$. We assume 1) $w[n]$ have zero mean and are uncorrelated; 2) $w[n]$ have equal variance σ^2. Then we have

$$E(\hat{A}) = E\left(\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right) = \frac{1}{N}\sum_{n=0}^{N-1} E(x[n]) = A$$

$$E(\check{A}) = E(x[0]) = A$$

29. Since $var(Ax) = A^2 var(x)$, where x is scalar. Then the variances are

$$var(\hat{A}) = var\left(\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right) = \frac{1}{N^2}\sum_{n=0}^{N-1} var(x[n]) = \frac{1}{N^2}N\sigma^2 = \frac{\sigma^2}{N}$$

and

$$var(\check{A}) = var(x[0]) = \sigma^2 > var(\hat{A})$$

4. Keep in Mind

30. **An estimator is a random variable**.

31.The use of computer simulations for assessing estimation performance, although quite valuable for gaining insight and motivating conjectures, **is never conclusive**.

32. We will see that optimal estimators can sometimes be difficult to implement, requiring a multidimensional optimization or integration. In these situations, **alternative estimators that are suboptimal, but which can be implemented on a digital computer, may be preferred**. (END)