Estimation Notes 3.3 - 3.4 pp.28 - 35

Key Points : Estimator Accuracy Considerations; Cramer-Rao Lower Bound

1. Estimator Accuracy Considerations

1. In general, **the more the PDF is influenced by the unknown parameter, the better we should be able to estimate it**.

2. **Example 3.1 - PDF Dependence on Unknown Parameter**

If a single sample is observed as

$$x[0] = A + w[0]$$

where $w[0] \sim \mathcal{N}(0, \sigma^2)$, and it is desired to estimate A, then we expect a better estimate if σ^2 is small. Of course the variance is σ^2 and the estimator accuracy improves as it decreases.

Figure 3.1 shows an alternative way of viewing this, the PDF has been plotted versus the **unknown parameter** A for a **given value** of $x[0]$. The PDFs for two different σ^2 are

$$p_i(x[0]; A) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{1}{2\sigma_i^2}(x[0] - A)^2 \right] \tag{3.1}$$

for $i = 1, 2$. If $x[0] = 3$, $\sigma_1^2 < \sigma_2^2$, then we should be able to estimate A **more accurately** based on $p_1(x[0]; A)$

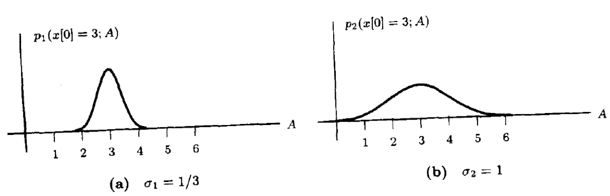

Figure 3.1 PDF dependence on unknown parameter.

3. When the PDF is viewed as a function of the unknown parameter (with $\mathbf{x}$ fixed), it is termed the **Likelihood Function**. For example, in (3.1), A is unknown parameter and $p_i(x[0]; A)$ is likelihood function.

4. Intuitively, the `\textbf{sharpness}` of the likelihood functions determines how accurately we can estimate the unknown parameter. We use **the negative of the second derivative of the logarithm of the likelihood function** at its peak. This is the **Curvature** of the log-likelihood function.

5. If we consider

$$\ln p(x[0]; A) = -\ln\sqrt{2\pi\sigma^2} - \frac{1}{2\sigma^2}(x[0] - A)^2$$

then the first derivative is

$$\frac{\partial \ln p(x[0]; A)}{\partial A} = -0 - \frac{1}{2\sigma^2}(-2x[0] + 2A) = \frac{1}{\sigma^2}(x[0] - A) \tag{3.2}$$

and the second derivative is

$$\underbrace{-\frac{\partial^2 \ln p(x[0]; A)}{\partial A^2}}_{\text{the Curvature of the log-likelihood function.}} = \frac{1}{\sigma^2} \tag{3.3}$$

Therefore, we have

$$var(\hat{A}) = \frac{1}{-\frac{\partial^2 \ln p(x[0];A)}{\partial A^2}} \tag{3.4}$$

and the variances decreases as the curvature increases.

6. A more appropriate measure of curvature is

$$-E\left[\frac{\partial^2 \ln p(x[0]; A)}{\partial A^2}\right] \qquad 3.5$$

7. The likelihood function, which depends on $x[0]$, is itself a **random variable**.

2. Cramer-Rao Lower Bound

8. **Theorem 3.1 (Cramer-Rao Lower Bound - Scalar Parameter)**

Assume the PDF $p(\mathbf{x};\theta)$ satisfies the **regularity** condition

$$E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] = 0 \quad \text{for all } \theta$$

where the expectation is taken with respect to $p(\mathbf{x};\theta)$. Then, the variance of any unbiased estimator $\hat{\theta}$ must satisfy

$$\begin{aligned} var(\hat{\theta}) &\geq \frac{1}{\text{Fisher Information } I(\theta) \text{ for the data } \mathbf{x}} \\ &= \frac{1}{-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right]} = \frac{1}{-\int \frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2} p(\mathbf{x};\theta) d\mathbf{x}} \end{aligned} \tag{3.6}$$

where **the derivative is evaluated at the true value of** θ and **the expectation is taken with respect to** $p(\mathbf{x};\theta)$. Furthermore, an unbiased estimator may be found that **attains the bound for all** θ **if and only if**

$$\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} = I(\theta)(g(\mathbf{x} - \theta)) \tag{3.7}$$

for some functions g and I. The MVU estimator is $\hat{\theta} = g(\mathbf{x})$, and the minimum variance is $1/I(\theta)$.

9. When the CRLB is attained (but it is NOT always attained), we have

$$var(\hat{\theta}) = \frac{1}{I(\theta)} \tag{3.10}$$

where

$$I(\theta) = -E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right]$$

10. An estimator which is unbiased and attains the CRLB, is said to be **efficient** that it efficiently uses the data. An MVU estimator **may or may not** be efficient.

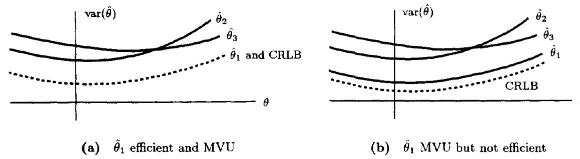

Figure 3.2 Efficiency vs. minimum variance.

11. For example, in Figure 3.2a, $\hat{\theta}_1$ is efficient in that it attains the CRLB. So it is also the MVU estimator. On the other hand, in Figure 3.2b, $\hat{\theta}_1$ does not attain the CRLB, and hence it is not efficient, but since its variance is uniformly less than that of all other unbiased estimators, it is the MVU estimator.

12. the **Fisher Information** $I(\theta)$ for the data $\mathbf{x}$

$$I(\theta) = -E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right] = -\int \frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2} p(\mathbf{x};\theta) d\mathbf{x} \tag{3.13}$$

when attains CRLB, the variance (i.e., $var(\hat{\theta})$) is the **reciprocal** of the Fisher information (i.e., $1/I(\theta)$). That means, **the more information, the lower the bound**.

13. the CRLB for N i.i.d observations is $1/N$ times that for one observation. For completely dependent samples such as $x[0] = \cdots = x[N-1]$, additional observations carry **NO** information, and the CRLB will **NOT** decrease with increasing data record length.