The Gaussian Data Assumption Leads to the Largest Cramér-Rao Bound

Arindam Bose

Estimation and Detection Theory

04/19/2016

Motivation

- **Criticism:** The Gaussian data assumption is sometimes criticized as being unrealistic in certain applications. But The Gaussian assumption is a natural choice when nothing is known about the exact data distribution.

- **The Fact:** The fact that the Gaussian distribution leads to the largest Cramér-Rao bound (CRB) in quite a general class of data distributions and for a significant set of parameter estimation problems. Consequently, the Gaussian CRB is the worst-case one.

Proof:

$$y(t) = \mathcal{Z}(t, \boldsymbol{\theta}) + \epsilon(t) \qquad t = 1, \dots, N$$

$$\epsilon(t) \sim \mathcal{N}(0, \sigma^2)$$

$\mathcal{Z}(s, \boldsymbol{\theta})$ is independent of $\epsilon(t)$ for $s \leq t$

Proof (contd.):

Transfer function description of an LRS sampled data system:

$$y(t) = \frac{B(z^{-1})}{A(z^{-1})} u(t) + \frac{C(z^{-1})}{D(z^{-1})} \epsilon(t)$$

Initial conditions: $\frac{B(0)}{A(0)} = 0, \frac{C(0)}{D(0)} = 1$

Proof (contd.):

$$\mathcal{Z}(t,\boldsymbol{\theta}) = \left[1 - \frac{D(z^{-1})}{C(z^{-1})}\right] \mathrm{y(t)} + \frac{B(z^{-1})}{A(z^{-1})}\frac{D(z^{-1})}{C(z^{-1})}u(t)$$

We can write,

$$y(t) = \mathcal{Z}(t,\boldsymbol{\theta}) + \epsilon(t) \qquad t = 1, \dots, N$$

$$\mathcal{Z}(s,\boldsymbol{\theta}) \text{ is independent of } \epsilon(t) \text{ for } s \leq t$$

Proof (contd.):

$p(t)$ depends on the past of $y(t)$ and $u(t)$:

$$E[p(t) - y(t)]^2 = E[p(t) - \mathcal{Z}(t, \boldsymbol{\theta})]^2 + E[\epsilon(t)]^2$$

$\Rightarrow \mathcal{Z}(t, \boldsymbol{\theta})$ is the min-variance one-step predictor of $y(t)$ denoted as $\hat{y}(t|t-1)$.

Proof (contd.):

Lets $\mathcal{p}(\epsilon)$ be the pdf of $\epsilon(t)$ such that $\mathcal{p}(\epsilon) > 0$

and $\mathcal{p}(\epsilon)$ satisfies the regularity conditions required in the standard derivation of the CRB.

Log-likelihood function:

$$g(\boldsymbol{\theta}) = \ln \prod_{t=1}^{N} \mathcal{p}(y(t) - \mathcal{Z}(t, \boldsymbol{\theta}))$$

$$= \sum_{t=1}^{N} \ln \mathcal{p}(y(t) - \mathcal{Z}(t, \boldsymbol{\theta}))$$

Proof (contd.):

The CRB matrix for $\{\boldsymbol{\theta}, \sigma^2\}$ is block diagonal with the block corresponding to $\boldsymbol{\theta}$ being equal to $\boldsymbol{F}^{-1}$.

$$\boldsymbol{F} = E\left[\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\boldsymbol{T}}}\right]$$

$$\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\sum_{t=1}^{N} \frac{\mathscr{p}'\left(\epsilon(t)\right)}{\mathscr{p}\left(\epsilon(t)\right)} \mathcal{Z}'(t, \boldsymbol{\theta})$$

$$= -\sum_{t=1}^{N} \rho(t)\, \mathcal{Z}'(t, \boldsymbol{\theta})$$

Proof (contd.):

$$\boldsymbol{F} = E\left[\sum_{t=1}^{N}\sum_{s=1}^{N}\rho(t)\,\rho(s)\mathcal{Z}'(t)\mathcal{Z}'^{T}(s)\right]$$

$$= \sum_{t=1}^{N} E[\rho^{2}(t)]\,E[\mathcal{Z}'(t)\mathcal{Z}'^{T}(t)]$$

$$+ \sum_{t=1}^{N}\sum_{s=1}^{t-1} E[\rho(t)]\,E[\rho(s)\mathcal{Z}'(t)\mathcal{Z}'^{T}(s)]$$

$$+ \sum_{t=1}^{N}\sum_{s=t+1}^{N} E[\rho(s)]\,E[\rho(t)\mathcal{Z}'(t)\mathcal{Z}'^{T}(s)]$$

$$= \gamma\boldsymbol{R}$$

$$\gamma = E[\rho^{2}(t)]\,,\ \boldsymbol{R} = \sum_{t=1}^{N} E[\rho^{2}(t)]\,E[\mathcal{Z}'(t)\mathcal{Z}'^{T}(t)]$$

Proof (contd.):

We see that the CRB depends on the data distribution only via the scalar γ.

$$\wp(\epsilon) = \text{const}\, e^{-\frac{1}{2\sigma^2}\epsilon^2}$$

$$\begin{aligned}
\gamma &= E[\rho^2(t)] \\
&= \int_{-\infty}^{\infty} \left[\frac{\wp'\big(\epsilon(t)\big)}{\wp\big(\epsilon(t)\big)}\right]^2 \wp(\epsilon)d\epsilon \\
&= \int_{-\infty}^{\infty} \frac{\epsilon^2}{\sigma^4}\wp(\epsilon)d\epsilon \\
&= \frac{1}{\sigma^2}
\end{aligned}$$

Proof (contd.):

For an arbitrary distribution:

$$E(\epsilon) = \int_{-\infty}^{\infty} \epsilon \wp(\epsilon) d\epsilon = 0 = \int_{-\infty}^{\infty} \wp(\epsilon) d\epsilon + \int_{-\infty}^{\infty} \epsilon \wp'(\epsilon) d\epsilon = 0$$

$$\Rightarrow \int_{-\infty}^{\infty} \epsilon \wp'(\epsilon) d\epsilon = -1$$

$$\Rightarrow 1 = \left[\int_{-\infty}^{\infty} \epsilon \wp'(\epsilon) d\epsilon\right]^2$$

$$= \left[\int_{-\infty}^{\infty} \epsilon \wp^{\frac{1}{2}}(\epsilon) \frac{\wp'(\epsilon)}{\wp^{\frac{1}{2}}(\epsilon)} d\epsilon\right]^2$$

$$\leq \left[\int_{-\infty}^{\infty} \epsilon^2 \wp(\epsilon) d\epsilon\right] \times \left[\int_{-\infty}^{\infty} \frac{\wp'^2(\epsilon)}{\wp(\epsilon)} d\epsilon\right] = \sigma^2 \gamma$$

$$\wp(\epsilon) = \text{const}\, \epsilon \wp(\epsilon)$$

Proof (contd.):

$$\wp'(\epsilon) = \text{const}\,\epsilon\wp(\epsilon)$$

$$\Rightarrow \wp(\epsilon) = \text{const}\, e^{-\frac{1}{2\sigma^2}\epsilon^2}$$

(Gaussian Probability Distribution Function)

Conclusion

$$\gamma \geq \frac{1}{\sigma^2}$$

- The CRB matrix associated with the general parameter estimation problem under discussion takes on its largest (unique) value under the Gaussian assumption.
- The parameter estimation methods or the experiment designs that are optimized based on the Gaussian CRB are min-max optimal in the sense that they yield the best CRB-related performance in the worst case.

Thank You

Questions?