TTT4275 Summary EstimationSpring 2019

Lecturer: Magne Hallstein Johnsen,

IES, NTNU

NTNU

Basic estimation 1

- General form for observation and estimator

  $x = f(\theta) + w$ and $\widehat{\theta} = g(x)$

- How to evaluate the estimator quality ?
  - Not so smart : by a lot of observations $x(n),\ n = 0, \ldots$
  - Our choice : by theory; i.e. no observations required!

- Defining two important properties :
  - Unbiased : $b(\widehat{\theta}) = E\{\widehat{\theta}\} - \theta = 0$
  - Variance : $= E\{(\widehat{\theta} - E\{\widehat{\theta}\})^2\}$

Basic estimation 2

- The overall best criterium for quality is $mse(\widehat{\theta}) = E\{(\hat{\theta} - \theta)^2\}$
- We showed that $mse(\widehat{\theta}) = var(\widehat{\theta}) + b^2(\widehat{\theta})$
- However minimizing $mse(\widehat{\theta})$ seldom gives feasible estimators
- We therefore choose a suboptimal strategy; i.e. restrict ourselves to unbiased estimators $b(\widehat{\theta}) = 0$ which results in $mse(\widehat{\theta}) = var(\widehat{\theta})$!
- Thus we want to find the unbiased estimator with minimum $mse = var$, and thereby shortened to MVU estimator.
- We also would like to find the smallest possible variance for any problem. This lower bound for the MVU is called the Cramer-Rao Lower Bound (CRLB).

The CRLB 1

a) Asume we know (or has estimated) $p(x, \theta)$

b) Further assume the 'regularity' condition is fulfilled :

$$E\{\frac{\delta log[p(x;\theta)])}{\delta\theta}\} = 0 \tag{1}$$

c) Then the CRLB is given by :

$$var(\widehat{\theta}) \geq E\{\frac{-1}{\frac{\delta^2 log[p(x;\theta)])}{\delta^2\theta}}\} = E\{\frac{1}{(\frac{\delta log[p(x;\theta)])}{\delta\theta})^2}\} \tag{2}$$

d) If equality is achieved we call the MVU estimator for **efficient** and the following reformulation applies

$$\delta log[p(x;\theta)] = I(\theta)[g(x) - \theta] \tag{3}$$

where $\widehat{\theta} = g(x)$ and $var(\widehat{\theta}) = \theta$

The vector CRLB

- In most problems we need to estimate more than one parameter, i.e. $\Theta = [\theta_1, \theta_2, \ldots, \theta_d]$ based on N observations $x = [x(0), x(1), .., x(N-1)]$
- We then define the Fisher Information matrix :

$$I(\Theta)_{ij} = E\{\frac{-1}{\frac{\delta^2 log[p(x;\Theta)])}{\delta\theta_i \delta\theta_j}}\} \tag{4}$$

- Any estimator must then fulfill the CRLB :

$$var(\hat{\theta})_{ii} \geq I^{-1}(\Theta)_{ii}$$

- Equality is achieved for an efficient (MVU) estimator which also will fulfill

$$\nabla_{\Theta} log[p(x;\Theta)] = I(\Theta)[\hat{\Theta} - \Theta] \tag{5}$$

where the covariance matrix of the estimator is given by

$$C(\hat{\Theta}) = I^{-1}(\Theta)$$

- Example 7 in the compendium gives a good introduction to the vector case

NTNU

The linear model for a problem and the resulting LSE estimator

- Many problems have a complex form such that good estimators are difficult to find.

- However, for some problems the observations are approximately linear in the unknown parameters. Thus we can write the following :

$$x = H\Theta + w \tag{6}$$

  where w is the model error and H is a known (observation) matrix

- We use the Least Square Error (LSE) criterium to find an estimator

$$LSE(\Theta) = (x - H\Theta)^T(x - H\Theta) \tag{7}$$

- Setting $\nabla LSE(\Theta) = 0$ we find

$$\hat{\Theta} = (H^T H)^{-1} H^T x \tag{8}$$

How good is the LSE estimator for the linear model problem

- The remaining question is how good this estimator is?

- The quality of the estimator can only be evaluated if we can calculate CRLB; i.e if we know $p(x, \Theta)$

- In the first case we assumed that the deviation can be approximated by independent (white) Gaussian noise; i.e. $p(w) = N(0, \sigma^2 I)$

- We then showed that the LSE-estimator fulfilled the requirement :

$$\nabla_{\Theta} log[p(x;\Theta)] = I(\Theta)[\hat{\Theta} - \Theta] \tag{9}$$

  with the corresponding CRLB equality :

  $Cov(\hat{\Theta}) = I^{-1}(\Theta) = \sigma^2(H^T H)^{-1}$

The linear model for a problem and the resulting LSE estimator

- The white noise assumption is often wrong. A more general approximation is to assume colored (correlated) Gaussian noise, i.e $\Sigma \neq \sigma^2 I$

- We solved the problem by filtering/whitening the noise and observation, i.e.

$$x' = Sx = SH\Theta + Sw = H'\Theta + w' \quad \text{where} \quad \Sigma^{-1} = S^T S \tag{10}$$

- We showed that $p(w') = N(0, I)$, i.e. white noise with unit power.

- Thus we ended up with the following efficient MVU estimator

$$\hat{\Theta} = (H^T \Sigma^{-1} H)^{-1} H^T \Sigma^{-1} x \tag{11}$$

$$Cov(\hat{\Theta}) = I^{-1}(\Theta) = (H^T \Sigma^{-1} H)^{-1} \tag{12}$$

The Best Linear Unbiased Estimator (BLUE) - 1

- The linear model gave an LSE-estimator which was linear in the observations $x = [x(0), \ldots, x(N-1)]$.

  This indicates that we should investigate linear estimators independent of the type of problem!

- Again we assume that we do not know the distribution $p(x;\theta)$

- It turns out that in order to find such an estimator we need to know the following (for a scalar case):

  a) The covariance matrix C_x
  b) The mean $E\{x(n)\} = s_n\theta \ n = 0, \ldots, N-1$ where all s_n are known

- The linear estimator is given by

$$\hat{\theta} = \sum_n a_n x(n) = a^T x \tag{13}$$

  where a must be found.

NTNU

BLUE 2

- Forcing the estimator to be unbiased results in the constraint $a^T s = 1$

- We also found the variance : $var(\hat{\theta}) = a^T C_x a$

- To find the best estimator we need to minimise $var(\hat{\theta})$, however while fulfilling the constraint. Thus we have to introduce the Lagrangian λ

$$L(\theta, \lambda) = a^T C_x a + \lambda(a^T s - 1) \tag{14}$$

- Minimizing $L(\theta, \lambda)$ results in

$$\hat{\theta} = \frac{s^T C_x^{-1} x}{s^T C_x^{-1} s} \tag{15}$$

$$var(\hat{\theta}) = \frac{1}{s^T C_x^{-1} s} \tag{16}$$

- Note that since we do not know $p(x; \theta)$ we do not know how close this variance is to the CRLB!

Maximum Likelihood Estimator (MLE)

- The LSE estimator (linear model approximation) and BLUE did not need knowledge of $p(x, \theta)$. Thus CRLB can not be found, and the estimator quality is generally unknown.

- The MLE requires knowledge of $p(x, \theta)$, thus CRLB can be found

- The term likelihood means $L(\theta/x) = p(x, \theta)$ where x is known and θ is unknown/variable

- MLE is generally not efficient, but is always **asymptotically** efficient, i.e.

$$\begin{aligned} lim_{N \to \infty} E\{\widehat{\theta}\} &= \theta \qquad (17) \\ lim_{N \to \infty} var(\widehat{\theta}) &= CRLB \end{aligned}$$

- As the name MLE indicates the estimator is found by

$$\widehat{\theta} = argmax_{\theta} L(\theta/x) \qquad (18)$$

Bayesian estimation -1

- Classical estimation (LSE, BLUE, MLE) : θ is unknown but deterministic, i.e.
$$p(x,\theta) = p(x/\theta)p(\theta) = p(x/\theta)$$

- Bayesian estimation (BMSE, MAP): θ is unknown and a stochastic variable with prior density $p(\theta)$

- Bayesian MSE is given by
$$BMSE(\acute{\theta}) = E\{(\theta - \acute{\theta})^2\} = \int\int (\theta - \acute{\theta})^2 p(x,\theta) d\theta dx \tag{19}$$

- Utilizing $p(x,\theta) = p(\theta/x)p(x)$
$$\begin{aligned} BMSE(\acute{\theta}) &= \int F(\acute{\theta}, x) p(x) dx \text{ where} \\ F(\acute{\theta}, x) &= \int (\theta - \acute{\theta})^2 p(\theta/x) d\theta \end{aligned}$$

Bayesian estimation -2

- Minimizing $BMSE(\acute{\theta})$ is eqivalent to minimizing the integrand $F(\acute{\theta}, x)$ as all variables inside the integrals are positive

- Minimizing $F(\acute{\theta}, x)$ by setting the derivative wrt. $\acute{\theta}$ equal to zero results in
$$\widehat{\theta} = E\{\theta/x\} = \int \theta p(\theta/x)dx \tag{20}$$

- The strategy then is first to find the posterior density $p(\theta/x)$ from
$$p(\theta/x) = \frac{p(x/\theta)p(\theta)}{p(x)}$$
and then calculate the integral in eq. 20.

- In practice it is seldom easy to calculate the integral. Thus a (suboptimal) strategy is to use the maximum value of the posterior (MAP)
$$\widehat{\theta} = argmax_\theta \, p(\theta/x) = argmax_\theta \frac{p(x/\theta)p(\theta)}{p(x)} = argmax_\theta \, p(x/\theta)p(\theta) \tag{21}$$

- If the posterior is symmetric MAP and minimum BMSE are identical