ML and CRLB

Sensor Fusion

Fredrik Gustafsson
fredrik.gustafsson@liu.se

Gustaf Hendeby
gustaf.hendeby@liu.se

Linköping University

BLUE and MVUE

Repetion of the linear model (which is linear in x):

$$y_k = H_k x + e_k, \quad \operatorname{Cov}(e_k) = R_k, \quad k = 1, \dots, N,$$
$$\mathbf{y} = \mathbf{H}x + \mathbf{e}, \quad \operatorname{Cov}(\mathbf{e}) = \mathbf{R}.$$

The Best Linear Unbiased Estimator (BLUE) $\hat{x}$ is defined as

- Linear: it has the form $\hat{x} = K\mathbf{y}$
- Unbiased: we require that $\mathrm{E}\left(\hat{x}\right) = x^0$
- Best: it minimizes the covariance $\operatorname{Cov}(\hat{x}) = \mathrm{E}\left(\hat{x} - x^0\right)\left(\hat{x} - x^0\right)^T$.

WLS is BLUE for the linear model.

In contrast, the Minimum Variance Unbiased Estimator (MVU) minimizes the covariance for any nonlinear estimator $\hat{x} = k(\mathbf{y})$, where the function $k(\mathbf{y})$ is constrained to give an unbiased estimator $\mathrm{E}(k(\mathbf{y})) = x$.

Maximum Likelihood

Definition of the maximum likelihood (ML) estimator

$$\hat{x} = \arg\max_{x} p(\mathbf{y}|x),$$

that is, the value of x that maximizes the likelihood function (distribution of measurements if x was given).
This is the same as minimizing the negative log likelihood

$$\hat{x} = \arg\min_{x} -\log p(\mathbf{y}|x) = \arg\min_{x} V^{ML}(x)$$

which looks similar to the LS and WLS solutions, but with a different cost function.
The ML is normally neither BLUE or MVU, but has other nice properties.

Maximum Likelihood for Gaussian Distribution

The likelihood function for the Gaussian distribution is

$$p(y_{1:N}|x) = \frac{1}{(2\pi)^{Nn_y/2}\prod_{k=1}^{N}\sqrt{\det(R_k)}} e^{-\frac{1}{2}\sum_{k=1}^{N}(y_k - H_k x)^T R_k^{-1}(y_k - H_k x)}$$
$$= \frac{1}{(2\pi)^{Nn_y/2}\prod_{k=1}^{N}\sqrt{\det(R_k)}} e^{-\frac{1}{2}V^{WLS}(x)}.$$

The (two times) negative log likelihood is thus

$$V^{ML}(x) = -2\log\big(p(y_{1:N}|x)\big) = Nn_y \log(2\pi) + \sum_{k=1}^{N}\log\big(\det(R_k)\big) + V^{WLS}(x).$$

Since the first two terms do not depend on x, ML and WLS are the same *for the Gaussian distribution*, so $\hat{x}^{ML} = \hat{x}^{WLS}$.

Maximum Likelihood for non-Gaussian Distributions

- For non-Gaussian distributions, ML is generally better then WLS.
- WLS is still BLUE for non-Gaussian distributions.
- That is, ML has smaller covariance, thus ML is *better* than WLS.
- Is ML also MVU, that is, the best?
- Yes, but only asymptotically!
- Asymptotically, the ML estimate reaches the *Cramér-Rao Lower Bound* (CRLB)

$$\hat{x}^{ML} \to \mathcal{N}(x^o, \mathcal{I}^{-1})$$

- No estimator can beat the CRLB, thus ML is the at least asymptotically (large amount of data) case.

Fisher Information Matrix

The CRLB is related to the *Fisher Information Matrix* (FIM) defined from the log likelihood. Key idea: every single data point provides a piece of information measured as

$$\mathcal{I}_k(x) = \mathrm{E}\left[\left(\frac{d \log p(y_k|x)}{dx} \right) \left(\frac{d \log p(y_k|x)}{dx} \right)^T \right].$$

and the total information is the sum of all information

$$\mathcal{I}_{1:N}(x) = \sum_{k=1}^{N} \mathcal{I}_k(x).$$

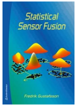
The CRLB theorem states that no unbiased estimator can beat the bound

$$\mathrm{Cov}(\hat{x}) \geq \mathcal{I}_{1:N}^{-1}(x^0).$$

Note 1: the FIM has to be evaluated at the true x^0.
Note 2: for the Gaussian case of the linear model $\mathcal{I}_k(x) = H_k^T R_k^{-1} H_k$, the FIM is independent of x and $\mathcal{I}_{1:N}^{-1} = P^{WLS}$, proving that WLS is the best estimator in this case.

Summary ML and CRLB

- ML maximizes the likelihood for the observed measurements, given the parameter x, with respect to x.
- For the linear model with Gaussian noise, WLS is also ML.
- For the linear model with non-Gaussian noise, ML is usually better than WLS.
- Asymptotically, ML is unbeatable in the MVU sense.
- The CRLB gives a lower bound on all estimators.
- ML attains the CRLB asymptotically.
- Most of the ML theory holds also for non-linear models.



Sections 2.4 and 2.5