Classical Inference

- Sufficient statistics, factorization theorem,
- Fisher information, Cramér-Rao bound, efficient estimators, information inequality theorem,
- Minimum variance unbiased (MVU) estimator for linear models [the same as the best linear unbiased estimator (BLUE), for white noise reduces to the least-squares (LS) estimator],
- Maximum likelihood (ML) estimators, ML invariance principle, asymptotic properties of ML estimators.

Example 1: We observe two samples of a DC level in *correlated* Gaussian noise:

$$\begin{aligned} X[0] &= a + W[0] \\ X[1] &= a + W[1] \end{aligned}$$

where

$$\boldsymbol{W} = [W[0], W[1]]^T$$

is a zero-mean Gaussian vector with covariance

$$C = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

Here,

- a is the unknown parameter,
- ρ is the known correlation coefficient between $W[1]$ and $W[2]$, and
- σ^2 is the known noise variance.

(a) Compute the Cramér-Rao bound (CRB) for a and compare it to the white-noise CRB (i.e. $\rho = 0$).

(b) Compute MVU estimator of a and its variance given a.

Example 2: Consider the observed data set $x[n]\, n = 0, 1, \ldots, N-1$. modeled as

$$X[n] = a + W[n] \quad n = 0, 1, \ldots, N-1$$

where

- a is the unknown DC level, which is assumed to be positive $(a > 0)$, and
- $W[n]$ is zero-mean white Gaussian noise with *unknown* variance a. Hence, the unknown parameter a is reflected in *both* the mean and the variance of the observations.

(a) Find CRB for a.

(b) Find the equation that needs to be solved to compute the ML estimate of a. Obtain the closed-form expression for the ML estimate of a.

(c) Consider the sample-mean estimator of a:

$$\overline{X} = (1/N) \cdot \sum_{n=0}^{N-1} X[n].$$

Does $\overline{X}$ attain CRB as $N \nearrow \infty$?

Example 3: In many applications involving fitting a linear model

$$\boldsymbol{X} = H\,\boldsymbol{\theta} + \boldsymbol{W}$$

to sets of input and output measurements

$$H = \begin{bmatrix} \boldsymbol{h}_1^T \\ \boldsymbol{h}_2^T \\ \vdots \\ \boldsymbol{h}_N^T \end{bmatrix} \text{ (of size } N \times p) \quad \text{and} \quad \boldsymbol{x} = \begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix} \text{ (of size } N \times 1)$$

not only are the output measurements $\boldsymbol{X}$ noisy but the input 'measurements' $\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_N$ may also be noisy. Denote by "T" a transpose and by I_p the identity matrix of size p.

To account for both noise effects, we formulate the following measurement model:

$$X[n] = (\underbrace{\boldsymbol{\beta}_n + \boldsymbol{\epsilon}_n}_{\boldsymbol{h}_n})^T\,\boldsymbol{\theta} + W[n] \quad n = 0, 1, \ldots, N-1$$

where

- $\boldsymbol{\theta}$ is an unknown parameter vector of size $p \times 1$,
- $W[n]$ is a sequence of independent, identically distributed (i.i.d.) zero-mean Gaussian random variables having known variance σ^2, and

- $\boldsymbol{\epsilon}_n$ is an i.i.d. sequence of zero-mean Gaussian $p \times 1$ random vectors having known covariance matrix $\sigma_\epsilon^2 \, I_p$, and
- $\boldsymbol{\beta}_n$ are known vectors.

Find the likelihood function of the unknown parameter vector $\boldsymbol{\theta}$. What does this function reduce to when σ^2 is much larger than σ_ϵ^2?

Find the likelihood equation that must be satisfied by the ML estimate of $\boldsymbol{\theta}$ when σ^2 and σ_ϵ^2 are known.

Hint: Recall that the likelihood function of $\boldsymbol{\theta}$ is the probability density function (pdf) of the measurement vector $\boldsymbol{X}$ given the parameter vector $\boldsymbol{\theta}$.

Example 4: Problem 8.3 in Kay-I.