

Basic concepts in estimation

Random and nonrandom parameters

Definitions of estimates

- ML Maximum Likelihood
- MAP Maximum *A Posteriori*
- LS Least Squares
- MMSE Minimum Mean square Error

Measures of quality of estimates

- Unbiased, variance, consistency, the Cramer-Rao lower bound, Fisher information, Efficiency

The problem of parameter estimation

Parameter is usually assumed to be time-invariant quantity,
vs. "time varying parameter"

Given measurements $z(j)$ $j=1,..k$ to estimate parameter x , in
presence of disturbances $w(j)$

$$z(j) = h[j, x, w(j)], \quad j = 1, \dots, k$$

find a function of k observation

$$Z^k = \{ z(j) \}_{j=1}^k$$

$$\hat{x}(k) \text{ or } \hat{x}(Z) \stackrel{\Delta}{=} \hat{x}[k, Z^k]$$

This function is called **estimator**, and the value is called
estimate

The **estimation error**

$$\tilde{x} \stackrel{\Delta}{=} x - \hat{x}$$

The problem of parameter estimation

Models for estimation of a parameter

1. Nonrandom, "unknown constant", unknown true value x_0 , **non-Bayesian** or **Fisher** approach
2. Random, the parameter is a random variable with a prior (*a priori*) pdf $p(x)$, the **Bayesian** approach

The Bayesian approach

from the prior pdf, one can obtain posterior (*a posteriori*) pdf using Bayes' formula

$$p(x|Z) = \frac{p(Z|x)p(x)}{p(Z)} = \frac{1}{c} p(Z|x)p(x)$$

c normalization constant

the posterior pdf can be used in several ways to estimate x

The problem of parameter estimation

The Non-Bayesian approach

- No the prior pdf associated with parameter, one cannot define posterior pdf for it
- One has the pdf of the measurements conditioned on the parameter, **Likelihood Function LF** of the parameter

$$\Lambda_Z(x) \stackrel{\Delta}{=} p(Z|x) \quad \text{or} \quad \Lambda_k(x) \stackrel{\Delta}{=} p(Z^k|x)$$

as a measure of how "likely" the parameter value is for the observations, LF serves as a measure of **evidence from the data**.

ML and MAP estimators

Maximization of the LF relative to x produces **Maximum Likelihood Estimator (MLE)**

$$\hat{x}^{ML}(Z) = \arg \max_x \Lambda_Z(x) = \arg \max_x p(Z|x)$$

MLE, being a function of random observations Z , is a random variable

MLE is the solution of likelihood equation

$$\frac{d\Lambda_Z(x)}{dx} = \frac{d p(Z|x)}{dx} = 0$$

ML and MAP estimators

Maximum A Posteriori Estimator (MAP), for a random parameter, follows from the maximization of the posterior pdf

$$\hat{x}^{MAP}(Z) = \arg \max_x p(x|Z) = \arg \max_x [p(Z|x)p(x)]$$

MAP estimate is a random variable

Normalization constant c (from Bayes formula) is irrelevant for the maximization

MLE versus MAP with Gaussian prior

Consider the single measurement

$$z = x + w \quad \text{where} \quad p(w) = N(0, \sigma^2)$$

Assume that x is **unknown constant**, no a prior information

$$\Lambda(x) = p(z|x) = N(z; x, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-x)^2}{2\sigma^2}}$$

then
$$\hat{x}^{ML}(Z) = \arg \max_x \Lambda(x) = z$$

since the peak or mode occurs at $x = z$

MLE versus MAP with Gaussian prior

Assume now that a prior information about x is **Gaussian**,

$$p(x) = N(x; \bar{x}, \sigma_0^2)$$

which is independent of w , then

the **posterior pdf** of x , on the observation z , is

$$p(x|z) = \frac{p(z|x)p(x)}{p(z)} = \frac{1}{c} e^{-\frac{(z-x)^2}{2\sigma^2} - \frac{(x-\bar{x})^2}{2\sigma_0^2}}$$

$$c = 2\pi\sigma\sigma_0 p(z)$$

where

c is the normalization constant independent of x , however it contain a Gaussian pdf $p(z)$

MLE versus MAP with Gaussian prior

rearrangement of the exponent, the 3rd term from $p(z)$

$$\begin{aligned}
 & -\frac{(z-x)^2}{2\sigma^2} - \frac{(x-\bar{x})^2}{2\sigma_0^2} + \frac{(z-\bar{x})^2}{(\sigma^2 + \sigma_0^2)} = -\frac{(z^2 - 2zx + x^2)\sigma_0^2 + (x^2 - 2x\bar{x} + \bar{x}^2)\sigma^2 - \frac{(z^2 - 2z\bar{x} + \bar{x}^2)\sigma_0^2\sigma^2}{(\sigma^2 + \sigma_0^2)}}{2\sigma^2\sigma_0^2} = \\
 & -\frac{x^2 - 2x\frac{(z\sigma_0^2 + \bar{x}\sigma^2)}{\sigma_0^2 + \sigma^2} + \frac{(z^2\sigma_0^2 + \bar{x}^2\sigma^2)}{\sigma_0^2 + \sigma^2} - \frac{(z^2 - 2z\bar{x} + \bar{x}^2)\sigma_0^2\sigma^2}{(\sigma_0^2 + \sigma^2)^2}}{\frac{2\sigma^2\sigma_0^2}{\sigma_0^2 + \sigma^2}} = \\
 & -\frac{x^2 - 2x\frac{(z\sigma_0^2 + \bar{x}\sigma^2)}{\sigma_0^2 + \sigma^2} + \frac{(z^2\sigma_0^4 + z^2\sigma_0^2\sigma^2 + \bar{x}^2\sigma^2\sigma_0^2 + \bar{x}^2\sigma^4) - (z^2 - 2z\bar{x} + \bar{x}^2)\sigma_0^2\sigma^2}{(\sigma_0^2 + \sigma^2)^2}}{\frac{2\sigma^2\sigma_0^2}{\sigma_0^2 + \sigma^2}} = \\
 & -\frac{x^2 - 2x\frac{(z\sigma_0^2 + \bar{x}\sigma^2)}{\sigma_0^2 + \sigma^2} + \frac{z^2\sigma_0^4 + 2z\bar{x}\sigma_0^2\sigma^2 + \bar{x}^2\sigma^4}{(\sigma_0^2 + \sigma^2)^2}}{\frac{2\sigma^2\sigma_0^2}{\sigma_0^2 + \sigma^2}} = \\
 & -\frac{\left(x - \frac{(z\sigma_0^2 + \bar{x}\sigma^2)}{\sigma_0^2 + \sigma^2}\right)^2}{\frac{2\sigma^2\sigma_0^2}{\sigma_0^2 + \sigma^2}}
 \end{aligned}$$

MLE versus MAP with Gaussian prior

the **posterior pdf** of x , on the observation z , becomes

$$p(x|z) = \frac{p(z|x)p(x)}{p(z)} = N(x; \xi(z), \sigma_1^2) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{[x-\xi(z)]^2}{2\sigma_1^2}}$$

where

$$\xi(z) \triangleq \frac{(z\sigma_0^2 + \bar{x}\sigma^2)}{\sigma_0^2 + \sigma^2} = \frac{\sigma^2}{\sigma_0^2 + \sigma^2} \bar{x} + \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2} z$$

$$\sigma_1^2 \triangleq \frac{\sigma^2 \sigma_0^2}{\sigma_0^2 + \sigma^2}$$

$$\hat{x}^{MAP}(z) = \arg \max_x p(x|z) = \xi(z)$$

MLE versus MAP with Gaussian prior

note that MAP estimator is a **weighted combination** of

1. z the ML estimator, peak of the likelihood function
2. \bar{x} the peak of the a prior pdf of the parameter

$$\hat{x}^{MAP}(z) = \arg \max_x p(x|z) = \xi(z) =$$

$$\frac{\sigma^2}{\sigma_0^2 + \sigma^2} \bar{x} + \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2} z = \left(\sigma_0^{-2} + \sigma^{-2} \right)^{-1} \left[\frac{\bar{x}}{\sigma_0^2} + \frac{z}{\sigma^2} \right]$$

the weighting is inversely proportional to their variance

directly proportional to information,

= inverse variance

Least Squares and Minimum Mean Square Error estimation

Another common estimation procedure for nonrandom parameters is **Least Squares method (LS)**

$$z(j) = h(j, x) + w(j), \quad j = 1, \dots, k$$

the LSE of x is

$$\hat{x}_{LS}(k) = \arg \min_x \left\{ \sum_{j=1}^k [z(j) - h(j, x)]^2 \right\}$$

Nonlinear LS-problem, linear LS-problem treated in detail
no assumptions about the "measurement errors"

Least Squares and Minimum Mean Square Error estimation

if the "measurement errors" are independent Gaussian random variables

$$p(w(j)) = N(0, \sigma^2)$$

then LSE coincides with the MLE, which is shown:

$$p(z(j)) = N(h(j, x), \sigma^2), \quad j = 1, \dots, k$$

Likelihood Function of x is

$$\Lambda_k(x) \stackrel{\Delta}{=} p(Z^k | x) = p[z(1), \dots, z(k) | x] = \prod_{j=1}^k N(z(j); h(j, x), \sigma^2) = c e^{-\frac{1}{2\sigma^2} \sum_{j=1}^k (z(j) - h(j, x))^2}$$

Least Squares and Minimum Mean Square Error estimation

$$\hat{x}_{ML} = \arg \max_x \ln \Lambda_k(x) =$$

$$\arg \max_x \left\{ - \sum_{j=1}^k [z(j) - h(j, x)]^2 \right\} =$$

$$\arg \min_x \left\{ \sum_{j=1}^k [z(j) - h(j, x)]^2 \right\} =$$

$$\hat{x}_{LS}(k)$$

under Gaussian assumption LS method is a disguised ML approach

Least Squares and Minimum Mean Square Error estimation

For **random parameter** the counterpart of LS is the **Minimum Mean Square Error (MMSE) Estimator**

$$\hat{x}^{MMSE}(Z) = \arg \min_{\hat{x}} E[(\hat{x} - x)^2 | Z]$$

$$\frac{dE[(\hat{x} - x)^2 | Z]}{d\hat{x}} = E[2(\hat{x} - x) | Z] = 2(\hat{x} - E[x | Z]) = 0 \Rightarrow$$

$$\hat{x}^{MMSE}(Z) = E[x | Z] \stackrel{\Delta}{=} \int_{-\infty}^{\infty} xp(x|Z)dx$$

the solution is **the conditional mean** of x

Least Squares and Minimum Mean Square Error estimation

Some LS-estimators

Single measurement of the unknown parameter x

$$z = x + w, \quad \hat{x}_{LS} = \arg \min_x \left[(z - x)^2 \right] = z = \hat{x}^{ML}$$

if the noise is *Gaussian*

Several measurement of the unknown parameter x ,

Gaussian independent noise terms

$$z(j) = x + w(j), \quad j = 1, \dots, k$$

sample mean

sample average

$$\hat{x}_{LS} = \arg \min_x \left\{ \sum_{j=1}^k [z(j) - x]^2 \right\} = \frac{1}{k} \sum_{j=1}^k z(j) = \bar{z}$$

$$= \arg \max_x \left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^k [z(j) - x]^2 \right\} = \arg \max_x \ln \Lambda_k(x) = \hat{x}^{ML}$$

Least Squares and Minimum Mean Square Error estimation

MMSE versus MAP Estimator

Single measurement of the unknown random parameter x with a Gaussian priori pdf, noise is also Gaussian

$$z = x + w, \quad p(x|z) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\xi(z))^2}{2\sigma_1^2}}$$

the **mean** of this Gaussian pdf is $\xi(z)$, which is also the **mode** of this pdf. Thus

$$\hat{x}^{MMSE} = E[x|z] = \xi(z) = \hat{x}^{MAP}$$

MMSE estimator coincides with MAP estimator due to the the fact that mean and mode of Gaussian coincide

$$p(x|z) = N\left(x; \hat{x}^{MMSE}, \sigma_1^2\right) = N\left(x; \hat{x}^{MAP}, \sigma_1^2\right)$$

Unbiased estimators

For a *nonrandom parameter*, an estimator is **unbiased** if

$$E[\hat{x}(k, Z^k)] = x_0, \quad \text{pdf } p(Z^k | x = x_0)$$

where x_0 is the true value of the parameter.

Bayesian case, x is random with a prior pdf $p(x)$

$$E[\hat{x}(k, Z^k)] = E[x]$$

General definition with estimation error

$$\tilde{x} \stackrel{\Delta}{=} x - \hat{x}; \quad E[\tilde{x}] = 0$$

Unbiasedness of ML and MAP estimator

Single measurement, **ML estimator**

$$z = x + w, \quad \hat{x}^{ML} = z$$

$$E[\hat{x}^{ML}] = E[z] = E[x_0 + w] = E[x_0] + E[w] = x_0$$

unbiased

Single measurement, **MAP estimator**, prior pdf $p(x)$

$$z = x + w, \quad \hat{x}^{MAP} = \xi(z) \triangleq \frac{\sigma^2}{\sigma_0^2 + \sigma^2} \bar{x} + \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2} z$$

$$E[\hat{x}^{MAP}] = E[\xi(z)] = \frac{\sigma^2}{\sigma_0^2 + \sigma^2} \bar{x} + \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2} E[z]$$

$$= \frac{\sigma^2}{\sigma_0^2 + \sigma^2} \bar{x} + \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2} [\bar{x} + E(w)] = \bar{x} = E(x)$$

unbiased

Bias in the ML estimation of two parameters

Estimating (ML) of unknown **mean** x and **variance** σ^2 of a set of k measurements

mean

$$\Lambda_k(x, \sigma) = p[z(1), \dots, z(k) | x, \sigma] = \frac{1}{(2\pi)^{\frac{k}{2}} \sigma^k} e^{-\frac{1}{2\sigma^2} \sum_{j=1}^k (z(j)-x)^2}$$

$$\hat{x}^{ML} = \bar{z} = \frac{1}{k} \sum_{j=1}^k z(j)$$

sample variance

$$\frac{\partial \ln \Lambda_k(x, \sigma)}{\partial \sigma} = -\frac{k}{\sigma} + \frac{1}{\sigma^3} \sum_{j=1}^k (z(j)-x)^2 = 0 \Rightarrow -\frac{k}{\sigma} + \frac{1}{\sigma^3} \sum_{j=1}^k (z(j)-\hat{x}^{ML})^2 = 0$$

$$\Rightarrow [\hat{\sigma}^{ML}(k)]^2 = \frac{1}{k} \sum_{j=1}^k (z(j)-\hat{x}^{ML})^2 = \frac{1}{k} \sum_{j=1}^k \left(z(j) - \frac{1}{k} \sum_{i=1}^k z(i) \right)^2$$

Bias in the ML estimation of two parameters

mean

$$E[\hat{x}^{ML}] = \bar{z} = E\left[\frac{1}{k} \sum_{j=1}^k z(j)\right] = x_0$$

sample mean estimator **unbiased**

sample variance

$$\Rightarrow E\left\{\left[\hat{\sigma}^{ML}(k)\right]^2\right\} = E\left\{\frac{1}{k} \sum_{j=1}^k \left(z(j) - \frac{1}{k} \sum_{i=1}^k z(i)\right)^2\right\} = \frac{k-1}{k} \sigma_0^2$$

sample variance estimator **biased**, (asymptotically unbiased)

In order to be unbiased

$$\left[\hat{\sigma}^{ML}(k)\right]^2 = \frac{1}{k-1} \sum_{j=1}^k \left(z(j) - \frac{1}{k} \sum_{j=1}^k z(j)\right)^2$$

The variance of an estimator

Non-Bayesian case, ML or LS

$$\text{var}[\hat{x}(Z)] \stackrel{\Delta}{=} E\{\hat{x}(Z) - E[\hat{x}(Z)]\}^2$$

if estimator unbiased

$$\text{var}[\hat{x}(Z)] = E\{\hat{x}(Z) - x_0\}^2$$

if estimator is biased then **Mean Square Error (MSE)**

$$MSE[\hat{x}(Z)] = E\{\hat{x}(Z) - x_0\}^2$$

Bayesian case, MAP

Unconditional (MSE) $MSE[\hat{x}(Z)] = E\{\hat{x}(Z) - x\}^2$

$$MSE[\hat{x}(Z)] = E\left[E\left\{[\hat{x}(Z) - x]^2 \mid Z\right\}\right] = E\left[MSE[\hat{x}(Z) \mid Z]\right]$$

inside brackets is **conditional MSE**

The variance of an estimator

For the **MMSE** estimator, the conditional MSE is

$$E\left[\left[\hat{x}^{MMSE}(Z) - x\right]^2 \middle| Z\right] = E\left[\left[x - E(x|Z)\right]^2 \middle| Z\right] = \text{var}(x|Z)$$

that is, the **conditional variance** of x given Z .

Averaging over Z yields

$$E\left[\text{var}(x|Z)\right] = E\left[\left[x - E(x|Z)\right]^2\right]$$

which is the **unconditional variance** of the estimate \hat{x}^{MMSE}

”average squared error over all possible observations”

The variance of an estimator

General definition

with **the estimation error** $\tilde{x} \stackrel{\Delta}{=} x - \hat{x}$

the expected value of the square of the estimation error is the estimator's variance or MSE

$$E[\tilde{x}^2] = \begin{cases} \text{var}(\hat{x}) & \text{if } \hat{x} \text{ unbiased and } x \text{ is nonrandom} \\ \text{MSE}(\hat{x}) & \text{if } \hat{x} \text{ biased} \end{cases}$$

The square root of the variance (or MSE)

$$\sigma_{\hat{x}} = \sqrt{\text{var}(\hat{x})}$$

is its **standard error**, also called **standard deviation of the estimation error**

The variance of an estimator

Comparison of variances of an ML and MAP estimator

Single observation

$$\text{var}[\hat{x}^{ML}] = E\{\hat{x}^{ML} - x_0\}^2 = E\{z - x_0\}^2 \stackrel{\Delta}{=} \sigma^2$$

$$\text{var}[\hat{x}^{MAP}] = E\{\hat{x}^{MAP} - x\}^2 =$$

$$E\left\{\left[\frac{\sigma^2}{\sigma_0^2 + \sigma^2} \bar{x} + \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2} (x + w) - x\right]^2\right\} =$$

$$E\left\{\left[\frac{\sigma^2}{\sigma_0^2 + \sigma^2} (\bar{x} - x) + \frac{\sigma_0^2}{\sigma_0^2 + \sigma^2} w\right]^2\right\} = \frac{\sigma_0^2 \sigma^2}{\sigma_0^2 + \sigma^2} < \sigma^2 = \text{var}[\hat{x}^{ML}]$$

This is due to the availability of of the *a priori* information

Consistency and efficiency of estimators

Consistency is here an asymptotic property

An estimator of a *nonrandom* parameter is **consistent estimator**, if the estimate converges to the true value in some sense

Using **convergence in the mean square** criterion then,

$$\lim_{k \rightarrow \infty} E \left[\left[\hat{x}(k, Z^k) - x_0 \right]^2 \right] = 0 \quad \text{Over } Z^k$$

is the condition for **consistency in the mean square sense**

for a *random* parameter

$$\lim_{k \rightarrow \infty} E \left[\left[\hat{x}(k, Z^k) - x \right]^2 \right] = 0 \quad \text{Over } Z^k \text{ and } x$$

$$\lim_{k \rightarrow \infty} E \left[\tilde{x}(k, Z^k) \right] = 0 \quad \text{For } \textit{unbiased} \text{ estimators}$$

Consistency and efficiency of estimators
The Cramer-Rao Lower Bound (CRLB) and
the Fisher Information Matrix (FIM)

CRLB: The *mean square error* corresponding to the estimator of a parameter cannot be smaller than a certain quantity

Scalar case

Scalar *nonrandom* parameter x with an *unbiased* estimator

$$E\left[\left[\hat{x}(k, Z^k) - x_0\right]^2\right] \geq J^{-1} \quad \text{where}$$

$$J \stackrel{\Delta}{=} - E\left[\frac{\partial^2 \Lambda(x)}{\partial x^2}\right]_{x=x_0} = E\left[\left[\frac{\partial \Lambda(x)}{\partial x}\right]^2\right]_{x=x_0}$$

J is the **Fisher Information**

$\Lambda(x)$ is the likelihood function

x_0 the *true value* of the parameter.

Consistency and efficiency of estimators
The Cramer-Rao Lower Bound (CRLB) and
the Fisher Information Matrix (FIM)

Scalar case

For scalar *random* parameter x with an *unbiased* estimator

$$E\left[\left[\hat{x}(k, Z^k) - x\right]^2\right] \geq J^{-1} \quad \text{where}$$
$$J \triangleq -E\left[\frac{\partial^2 p(Z, x)}{\partial x^2}\right] = E\left[\left[\frac{\partial p(Z, x)}{\partial x}\right]^2\right]$$

J is the **Fisher Information**

If the estimator variance is equal to CRLB, then such an estimator is called efficient.

Consistency and efficiency of estimators

The Cramer-Rao Lower Bound (CRLB) and the Fisher Information Matrix (FIM)

Vector case

For *nonrandom* vector parameter, the CRLB states that the covariance matrix of an *unbiased* estimator is bounded from below:

$$E\left[\left[\hat{x}(Z) - x_0\right]\left[\hat{x}(Z) - x_0\right]^T\right] \geq J^{-1} \quad \text{where}$$

$$J \stackrel{\Delta}{=} -E\left[\nabla_x \nabla_x^T \ln \Lambda(x)\right]_{x=x_0} = E\left[\left[\nabla_x \ln \Lambda(x)\right]\left[\nabla_x \ln \Lambda(x)\right]^T\right]_{x=x_0}$$

J is the **Fisher Information Matrix (FIM)**

$\Lambda(x)$ is the likelihood function

x_0 the *true value* of the parameter

$$A \geq B \Leftrightarrow C \stackrel{\Delta}{=} A - B \geq 0 \quad \text{pos. def.}$$