Parameter Estimation

- the problem of estimating a (time invariant) parameter x
- given the measurements $z(j) = h[j, x, w(j)]$ $j = 1, \ldots, k$
- made in the presence of the disturbances (noises) $w(j)$
- find a function of the k observations, $\hat{x}(k) \triangleq \hat{x}[k, Z^k]$
- observations are denoted as $Z^k \triangleq \{z(j)\}_{j=1}^{k}$
- that estimates the value of x in some sense
- this function is called the estimator and its value is the estimate
- the estimation error corresponding to the estimate $\hat{x}$ is $\tilde{x} \triangleq x - \hat{x}$
- alternate notation when k is fixed $\hat{x}(Z) \triangleq \hat{x}[k, Z^k]$

Models for Parameter Estimation

- two models used in estimation of a (time invariant) parameter:
- nonrandom - there is an unknown true value x_0 also called non-Bayesian or Fisher approach
- random - the parameter is a random variable with a prior pdf $p(x)$
- a realization of x according to $p(x)$ is assumed to have occurred,
- this value then stays constant during the measurement process
- called the Bayesian approach; described as below:
- starts with the prior pdf of the parameter from which one can obtain its posterior pdf (a posterior pdf) using Bayes' formula

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

  c is the normalization constant, which does not depend on x
- the posterior pdf can be used in several ways to estimate x

Models for Parameter Estimation

- Non-Bayesian (Likelihood Function) approach
- there is no prior pdf associated with the parameter
- one can not define a posterior pdf for it
- one has the pdf of the measurements conditioned on the parameter
- called the likelihood function (LF) of the parameter

$$\Lambda_Z(x) \triangleq p(Z|x) \quad \text{or} \quad \Lambda_k(x) \triangleq p(z^k|x)$$

Maximum Likelihood (ML) and Maximum A Posterior (MAP) Estimators

- ML Estimator: a common method of estimating nonrandom parameter
- maximizes the likelihood function
- x is an unknown constant
- $\hat{x}^{ML}(Z)$ being a function of the set of random observation Z is a random variable

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

- MLE is the solution of the likelihood function

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

Maximum A Posterior estimator (MAP)

- estimate for a random parameter
- maximization of the posterior pdf

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

normalization constant is irrelevant for the maximization

- MAP estimate depends on the observations Z and through them on the realization of x is obviously a random variable

MLE vs. MAP Estimator with Gaussian Prior

- consider the signle measurement
- $z = x + w$ of the unknown parameter x in the presence of the additive measurement noise w,
- assumed to be a normally (Gaussian) distributed random variable with mean zero and variance σ^2 $w \sim \mathcal{N}(0, \sigma^2)$
- first assume x is an unknown constant (no prior information about it is available)
- the likelihood function of x

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

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

- the peak or mode of the above equation occurs at $x = z$

MLE vs. MAP Estimator with Gaussian Prior

- Next assume that the prior information about the parameter is that x is Gaussian with mean $\bar{x}$ and variance σ_0^2

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

assume x is independent of w

- the posterior pdf of x conditioned on the observation z is

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

where $c = 2\pi\sigma\sigma_0 p(z)$ is the normatlization constant indepedent of x

- normalization constant which guarantees that the pdf integrates to unity

MLE vs. MAP Estimator with Gaussian Prior

- it can be shown that the posterior pdf of x is (i.e., Gaussian)

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

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

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

- maximization of $p(x|z)$ with respect to x yields

$$\hat{x}^{\mathrm{MAP}} = \xi(z)$$

- $\xi(z)$ is the MAP estimator for the random parameter x

MLE vs. MAP Estimator with Gaussian Prior

- MAP estimator for purely Gaussian problem is a weighted combination of
  1. z the MLE which is the peak (or mode) of the likelihood function
  2. $\bar{x}$ which is the peak of the prior pdf of the parameter to be estimated

$$
\begin{aligned}
\hat{x}^{\text{MAP}} &= (\sigma_0^{-2} + \sigma^{-2})^{-1}\sigma_0^{-2}\bar{x} + (\sigma_0^{-2} + \sigma^{-2})^{-1}\sigma^{-2}z \\
&= (\sigma_0^{-2} + \sigma^{-2})^{-1}\left[\frac{\bar{x}}{\sigma_0^2} + \frac{z}{\sigma^2}\right]
\end{aligned}
$$

- the weightings of the prior mean and the measurement are inversely proportional to their variances
- $\sigma_1^{-2} = \sigma_0^{-2} + \sigma^{-2}$
- the inverse variances (also called information) are additive
- this additive property of information holds in general when the information sources are independent

Bayesian vs. Non-Bayesian

- $\hat{x}^{ML}$ based on non-Bayesian approach and $\hat{x}^{MAP}$ based on Bayesian
- $\hat{x}^{MAP}$ coincides with $\hat{x}^{ML}$ for a certain prior pdf called a diffuse pdf

$$\lim_{\sigma_0 \to \infty} \xi(z) = z$$

- Sufficient statistic and likelihood equation
- if the likelihood function of a parameter can be decomposed as follows

$$\Lambda(x) \triangleq p(Z|x) = f_1[g(Z), x] f_2(Z)$$

- then the ML estimate of x depends only on the function $g(Z)$ called the sufficient statistic, rather than on the entire data set Z
- the sufficient statistic summarizes the information about x contained in the entire data

Example :Sufficient statistic and likelihood equation

- consider the scalar measurements $z(j) = x + w(j)$ $j = 1, \ldots, k$
- if the noise compoents $w(j)$ $j = 1, \ldots, k$ are independent and identically distributed zero-mean Gaussian random variables with variance σ^2

$$w(j) \sim \mathcal{N}(0, \sigma^2) \quad \text{then} \quad z(j) \sim \mathcal{N}(x, \sigma^2)$$

and conditioned on x, the observations $z(j)$ are mutually independent

- the likelihood function of x in terms of $Z^k \triangleq \{z(j), j = 1, \ldots, k\}$ is then

$$\begin{aligned}\Lambda_k(x) \triangleq p(Z^k|x) &\triangleq p[z(1), \ldots, z(k)|x] \\ &= \prod_{j=1}^{k} \mathcal{N}[z(j); x, \sigma^2] = ce^{-\frac{1}{2\sigma^2}\sum_{j=1}^{k}[z(j)-x]^2}\end{aligned}$$

Example :Sufficient statistic and likelihood equation

- it can be rewritten into the product of two functions

$$
\begin{aligned}
\Lambda_k(x) &= ce^{-\frac{1}{2\sigma^2}\sum_{j=1}^{k} z(j)^2 + \frac{1}{2\sigma^2} 2\sum_{j=1}^{k} z(j)x - \frac{1}{2\sigma^2}kx^2} \\
&= ce^{-\frac{1}{2\sigma^2}\sum_{j=1}^{k} z(j)^2} e^{-\frac{1}{2\sigma^2}kx[x-\frac{2}{k}\sum_{j=1}^{k} z(j)]} \\
&\triangleq f_z(Z) f_1[g(Z), x] \\
f_2(Z) &\triangleq ce^{-\frac{1}{2\sigma^2}\sum_{j=1}^{k} z(j)^2} \\
f_1[g(Z), x] &\triangleq e^{-\frac{1}{2\sigma^2}kx[x-2\bar{z}]} \\
g(Z) &\triangleq \frac{1}{k}\sum_{j=1}^{k} z(j) \triangleq \bar{z}
\end{aligned}
$$

- $\bar{z}$ is the sufficient statistic for estimating x

Least Squares and Minimum Mean Square Error Estimatio

- LS Estimator
- another common estimation procedure for nonrandom parameters
- given the scalar and nonlinear measurements $z(j) = h(j,x) + w(j)$ $j = 1, \ldots, k$
- the least squares estimator (LSE) of x is

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

- it is nonlinear LS problem
- if the function h is linear in x then it is linear LS problem
- does not make any assumptions about the "measurement errors" or "noises" $w(j)$

Least Squares and Minimum Mean Square Error Estimatio

- if these are independent and identically distributed zero-mean Gaussian random variables, that is
- $w(j) \sim \mathcal{N}(0, \sigma^2)$ then LSE coincides with the MLE under these assumptions
- $z(j) \sim \mathcal{N}[h(j,x), \sigma^2]$ $j = 1, \ldots, k$
- likelihood function of x is then

$$\begin{aligned} \Lambda_k(x) &\triangleq p(Z^k|x) \triangleq p[z(1), \ldots, z(k)|x] \\ &= \prod_{j=1}^{k} \mathcal{N}[z(j); h(j,k), \sigma^2] = ce^{-\frac{1}{2\sigma^2}\sum_{j=1}^{k}[z(j)-h(j,x)]^2} \end{aligned}$$

- the minimization of LSE problem is equivalent to the maximization of ML approach