2 Estimation Theory
2.3: ML and MAP Estimators

Estimation Theory

1. **Introduction to Estimation Theory**
   - Assessing Estimator Performance
   - Minimum Variance Unbiased Estimator
   - Function Estimation
2. **Cramer-Rao Bound and Efficient Estimator**
   - Cramer-Rao Bound
   - Examples
3. **Maximum Likelihood & Maximum a Posteriori Estimator**
   - Classical estimation: Maximum Likelihood Estimator
   - The Bayesian framework: Maximum a Posteriori Estimator

ML and MAP Estimators

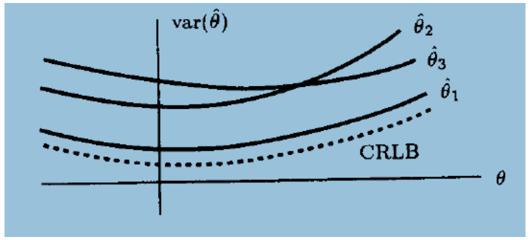
1. **Introduction**
   - Non efficient estimators
2. **Maximum Likelihood Estimator**
   - Properties of the ML estimators
   - Examples
3. **Maximum a Posteriori Estimator**
   - Bayesian framework
   - Examples

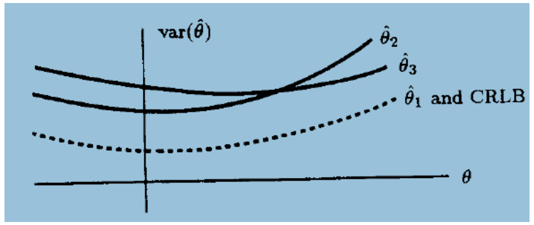
Cramer-Rao Lower Bound

2.3

The CRLB states that exists a **lower bound of the variance** of the whole set of unbiased estimators of a parameter θ.

It proposes a mechanism that, it some cases, allows obtaining this estimator, that is named **efficient**





Cramer-Rao Lower Bound

The **variance of any unbiased estimator** $\hat{\theta}$ must satisfy:

$$\mathrm{var}\left(\hat{\theta}\right) \geq \frac{1}{-E\left\{\dfrac{\partial^2 \ln f_{\underline{\mathbf{x}}}\left(\underline{\mathbf{x}}\,;\theta\right)}{\partial\theta^2}\right\}}$$

And the **equality is satisfied** when, for some function $k(\theta)$:

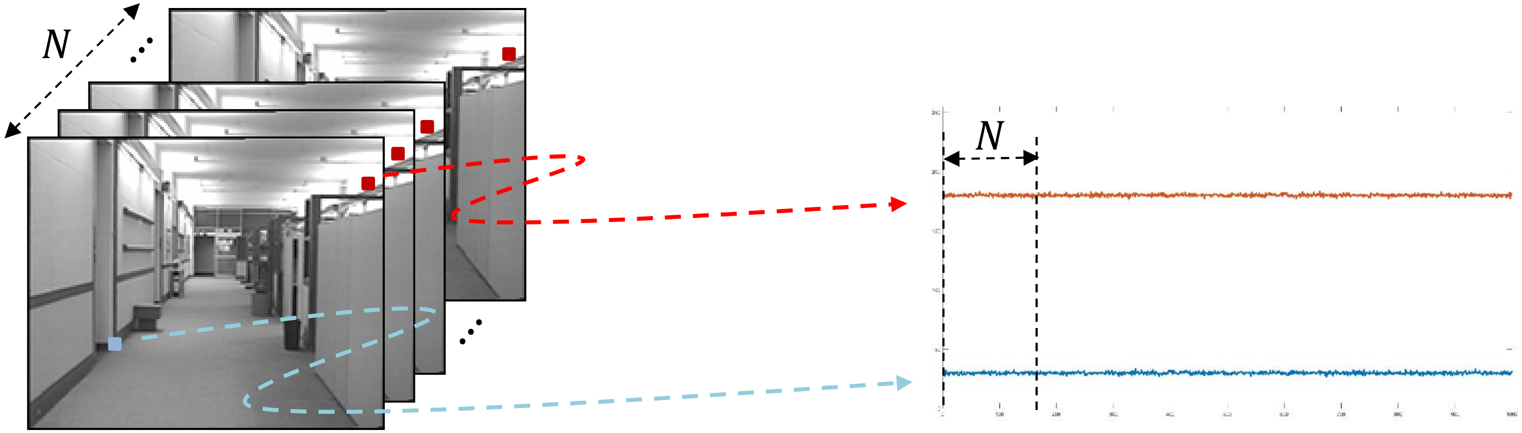
$$\frac{\partial \ln f_{\underline{\mathbf{x}}}\left(\underline{\mathbf{x}}\,;\theta\right)}{\partial\theta} = k(\theta)\left(\hat{\theta}_{opt}\left(\underline{\mathbf{x}}\right) - \theta\right)$$

Non efficient estimators

Nevertheless, in some cases, **there is no (feasible) estimator** that satisfies the Cramer-Rao Lower Bound:

- Given N samples of a process that can modeled as $\underline{\mathbf{x}} = A\underline{\mathbf{1}} + \underline{\mathbf{w}}$, compute and efficient estimator of its mean (A) and variance (σ^2) :

  Note: $W[n]$ is a **Gaussian, stationary, white noise**. The vector parameter is $\underline{\boldsymbol{\theta}} = (A, \sigma^2)^T$



Let us assume that the noise introduced by the camera sensors needs to be determined.
Joint estimation of the mean and variance values of every pixel based on N samples

Non efficient estimators

2.3

Nevertheless, in some cases, **there is no (feasible) estimator** that satisfies the Cramer-Rao Lower Bound:

- Given N samples of a process that can modeled as $\underline{\mathbf{x}} = A\underline{\mathbf{1}} + \underline{\mathbf{w}}$, compute and efficient estimator of its mean (A) and variance (σ^2) :

  Note: $W[n]$ is a **Gaussian, stationary, white noise**. The vector parameter is $\underline{\boldsymbol{\theta}} = (A, \sigma^2)^T$

$$\operatorname{var}(\hat{\theta}_i) \geq \left[\underline{\underline{\mathbf{I}}}^{-1}(\underline{\boldsymbol{\theta}})\right]_{ii}$$

$$\left[\underline{\underline{\mathbf{I}}}(\underline{\boldsymbol{\theta}})\right]_{ij} = -E\left\{\frac{\partial^2 \ln f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}; \underline{\boldsymbol{\theta}})}{\partial\theta_i \partial\theta_j}\right\} \quad \Rightarrow \quad \boxed{\underline{\underline{\mathbf{I}}}(\underline{\boldsymbol{\theta}}) = \begin{bmatrix} N/\sigma^2 & 0 \\ 0 & N/2\sigma^4 \end{bmatrix}}$$

$$\nabla_{\underline{\boldsymbol{\theta}}}\left(f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}; \underline{\boldsymbol{\theta}})\right) = \underline{\underline{\mathbf{I}}}^{-1}(\underline{\boldsymbol{\theta}})\left(\underline{\boldsymbol{\theta}}_{opt}(\underline{\mathbf{x}}) - \underline{\boldsymbol{\theta}}\right) \quad \Rightarrow \quad \boxed{\hat{A}_{opt} = \frac{1}{N}\sum_{n=1}^{N} x[n]} \quad \boxed{\operatorname{var}(\hat{A}_{opt}) = \frac{\sigma^2}{N}}$$

As $\underline{\underline{\mathbf{I}}}(\underline{\boldsymbol{\theta}})$ is diagonal, the p-dimensional problem becomes p 1D problems:

$$\boxed{\hat{\sigma}^2_{opt} = \frac{1}{N}\sum_{n=1}^{N}(x[n] - A)^2} \quad \boxed{\operatorname{var}(\hat{\sigma}^2_{opt}) = \frac{2\sigma^4}{N}}$$

Non efficient estimators

- Given N independent samples of a Laplacian process ($\underline{\mathbf{x}} = m\underline{\mathbf{1}} + \underline{\boldsymbol{w}}$), we want to estimate their mean (m):

  Note: $W[n]$ is a **Laplacian, stationary, white noise**

$$f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}\,;\theta) = [\text{INDEP.}] = \prod_{i=1}^{N} f_x(x_i\,;\theta)$$

$$\boxed{f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}\,;m) = \prod_{i=1}^{N} \tfrac{1}{2\lambda} \exp\left[-\frac{|x_i - m|}{\lambda}\right]}$$

Let us try to factorized: $\dfrac{\partial \ln f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}\,;\theta)}{\partial\theta} = k(\theta)\left(\hat{\theta}_{opt}(\underline{\mathbf{x}}) - \theta\right)$

$$f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}\,;m) = \frac{1}{(2\lambda)^N} \exp\left[-\frac{\sum_{i=1}^{N}|x_i - m|}{\lambda}\right] \quad \Rightarrow \quad L(\underline{\mathbf{x}}\,;m) = -N\ln 2\lambda - \frac{1}{\lambda}\sum_{i=1}^{N}|x_i - m|$$

$$\frac{\partial L(\underline{\mathbf{x}}\,;m)}{\partial m} = -\frac{1}{\lambda}\sum_{i=1}^{N}\frac{\partial}{\partial m}|x_i - m| \quad \Rightarrow \quad \boxed{\frac{\partial L(\underline{\mathbf{x}}\,;m)}{\partial m} = \frac{1}{\lambda}\sum_{i=1}^{N}\text{sign}(x_i - m)}$$

It cannot be factorized

ML and MAP Estimators

1. **Introduction**
   - Non efficient estimators
2. **Maximum Likelihood Estimator**
   - Properties of the ML estimators
   - Examples
3. **Maximum a Posteriori Estimator**
   - Bayesian framework
   - Examples

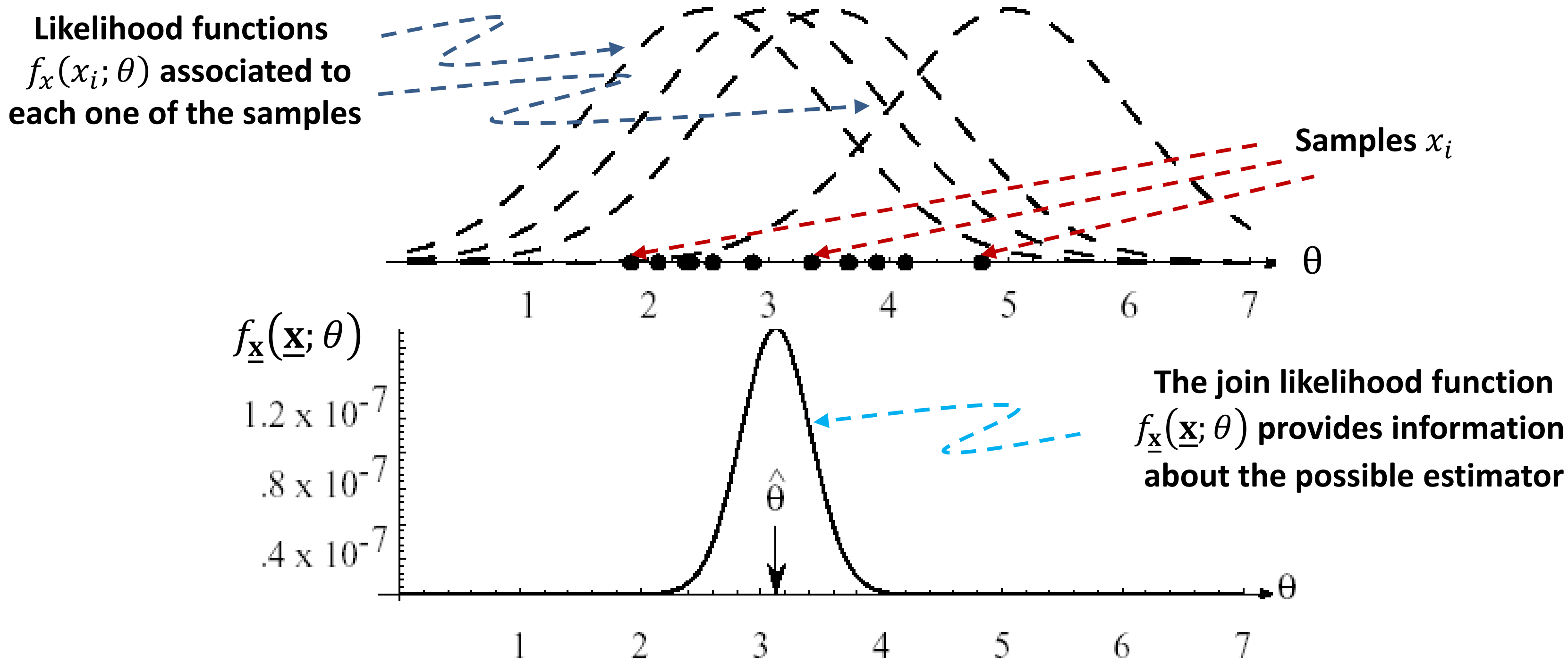
Likelihood function

2.3

Interpretation of the likelihood function:

Let us analyze the case of the likelihood function ($f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}};\theta)$) of a set of N Gaussian, independent samples:

$$f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}\,;\theta) = \prod_{i=1}^{N} f_x(x_i\,;\theta)$$

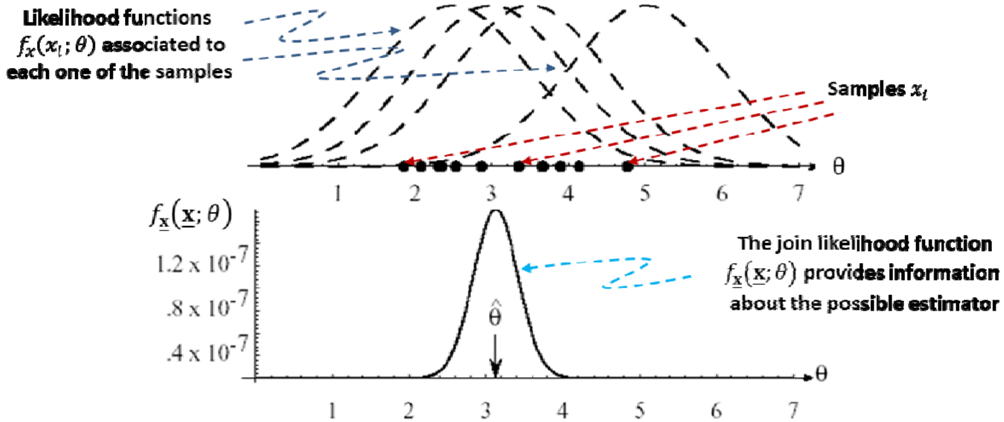


The Maximum Likelihood Estimator

2.3

Let us define the ML estimator ($\underline{\hat{\boldsymbol{\theta}}}_{ML}$) as:

$$\boxed{\underline{\hat{\boldsymbol{\theta}}}_{ML} = \max_{\underline{\boldsymbol{\theta}}} f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}\,;\underline{\boldsymbol{\theta}})}$$



Properties of the ML estimator:

1. **Asymptotically unbiased** (and in a large number of cases, **unbiased**)
2. **Asymptotically efficient** (when N increases, its variance attains CRLB)
3. **Efficiency**: When there exists an efficient estimator, it is the ML estimator
4. **Gaussian** for N large: it is characterized by its mean and variance
5. **Invariance**: The ML estimator of a function of a parameter $\alpha = g(\theta)$ can be obtained as

$$\boxed{\hat{\alpha}_{ML} = g(\hat{\theta}_{ML})}$$

Efficiency and the ML estimator

2.3

Properties of the ML estimator:

3. Efficiency: When there exists an efficient estimator, it is the ML estimator.

If there exists an **efficient estimator**, the following factorization has been possible:

$$\frac{\partial \ln f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}\,;\theta)}{\partial\theta} = k(\theta)\big(g(\underline{\mathbf{x}}) - \theta\big)$$

As $\ln(.)$ is a **monotonically increasing** function, the positions of the extrema do not change:

- **Note**: When computing the ML estimator, we will use the **log-likehood function**

$$\frac{\partial \ln f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}\,;\theta)}{\partial\theta} = 0 \quad \Leftrightarrow \quad \frac{\partial f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}\,;\theta)}{\partial\theta} = 0$$

Thus, if there is an efficient estimator, **the CR and the ML estimators** are the same:

$$\frac{\partial \ln f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}\,;\theta)}{\partial\theta} = 0 \quad \Leftrightarrow \quad \big(g(\underline{\mathbf{x}}) - \theta\big) = 0$$

$$\hat{\theta}_{ML} = g(\underline{\mathbf{x}}) = \hat{\theta}_{opt}(\underline{\mathbf{x}}) = \hat{\theta}_{CR}$$

ML and MAP Estimators

2.3

1. **Introduction**
   - Non efficient estimators
2. **Maximum Likelihood Estimator**
   - Properties of the ML estimators
   - Examples
3. **Maximum a Posteriori Estimator**
   - Bayesian framework
   - Examples

Example of ML estimator (I)

2.3

- Given N samples of a process that can be modeled as $\underline{\mathbf{x}} = A\underline{\mathbf{1}} + \underline{\mathbf{w}}$, compute the ML estimator of its mean (A) and variance (σ^2) .
  Note: $W[n]$ is a **Gaussian, stationary, white noise**.

Generic expression of a multivariate Gaussian ▶

$$f_{\underline{w}}(\underline{\mathbf{w}}) = \frac{1}{\sqrt{(2\pi)^N \left|\underline{\underline{\mathbf{C}}}_{\underline{w}}\right|}} \exp\left[-\frac{[\underline{\mathbf{w}} - \underline{\mathbf{m}}_{\underline{w}}]^T \underline{\underline{\mathbf{C}}}_{\underline{w}}^{-1} [\underline{\mathbf{w}} - \underline{\mathbf{m}}_{\underline{w}}]}{2}\right]$$

STATIONARY
WHITE NOISE $\Rightarrow E\{w[n]\} = \emptyset \quad r_w[\ell] = \sigma^2 \delta[\ell]$

$$\underline{\underline{C}}_w = \left[E\{w[n]\} = \emptyset\right] = \underline{\underline{R}}_w = \left[\text{WHITE, STAT.}\right] = \sigma^2 \underline{\underline{I}}$$

AS $\underline{x} = A\underline{1} + \underline{w} \Rightarrow \underline{w} = \underline{x} - A\underline{1} \leftarrow \begin{cases} \underline{w} : N(\underline{\emptyset}, \sigma^2 \underline{\underline{I}}) \\ \underline{x} : N(\underline{A}, \sigma^2 \underline{\underline{I}}) \end{cases}$

Example of ML estimator (I)

2.3

$$f_{\underline{w}}(\underline{x};\underline{\theta}) = \frac{1}{\left[(2\pi)^N \sigma^{2N}\right]^{1/2}} \cdot \exp\left[-\frac{[\underline{x}-\underline{A}]^T \frac{1}{\sigma^2}\underline{\underline{I}}\,[\underline{x}-\underline{A}]}{2}\right] =$$

$$f(\underline{x};\underline{\theta}) = \frac{1}{\left[2\pi\sigma^2\right]^{N/2}} \cdot \exp\left[-\frac{[\underline{x}-\underline{A}]^T[\underline{x}-\underline{A}]}{2\sigma^2}\right]$$

TWO COMMENTS:

* THE PARAMETERS ARE $\underline{\theta} = [A, \sigma^2]^T$ AND THE PDF IS NOT PARAMETERIZED WITH RESPECT TO $\underline{\theta}$ DIRECTLY
* WE ARE GOING TO DEVELOP THE STUDY USING SCALAR NOTATION. (VECTOR NOTATION IS POSSIBLE AS WELL)

Example of ML estimator (I)

2.3

$$f(\underline{x};\underline{\theta}) = \frac{1}{[2\pi\sigma^2]^{N/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x[n]-A)^2\right]$$

$$L(\underline{x};\underline{\theta}) = \ln f(\underline{x},\underline{\theta}) = -\frac{N}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x[n]-A)^2$$

LET US FIRST COMPUTE ALL PARTIAL DERIVATIVES:

$$\frac{\partial L(\underline{x};\underline{\theta})}{\partial A} = \frac{\partial}{\partial A}\left[-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x[n]-A)^2\right] =$$

$$= \frac{1}{2\sigma^2}\, 2\sum_{n=1}^{N}(x[n]-A) \qquad (1)$$

$$\frac{\partial L(\underline{x};\underline{\theta})}{\partial \sigma^2} = \frac{\partial}{\partial \sigma^2}\left[-\frac{N}{2}\ln\sigma^2\right] - \sum_{n=1}^{N}(x[n]-A)^2\,\frac{\partial}{\partial\sigma^2}\left[\frac{1}{2\sigma^2}\right] =$$

$$= -\frac{N}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{n=1}^{N}(x[n]-A)^2 \qquad (2)$$

Example of ML estimator (I)

2.3

$$\nabla_{\underline{\theta}} L(\underline{x};\underline{\theta}) = \left(\frac{\partial L(\underline{x};\underline{\theta})}{\partial A}, \frac{\partial L(\underline{x};\underline{\theta})}{\partial \sigma^2} \right)^T = \underline{\emptyset}$$

$$(1) \Rightarrow \frac{\partial L(\underline{x};\underline{\theta})}{\partial A} = \frac{1}{\sigma^2} \sum_{n=1}^{N} (x[n] - A) = \emptyset$$

$$\hat{A}_{ML} = \frac{1}{N} \sum_{n=1}^{N} x[n] \Rightarrow \hat{A}_{ML} = \hat{A}_{OPT} = \hat{A}_{CR} = \hat{A}_{EFF}$$

$$\text{with} \quad VAR(\hat{A}_{ML}) = \frac{\sigma^2}{N} = VAR(\hat{A}_{CR})$$

$$(2) \Rightarrow \frac{\partial L(\underline{x};\underline{\theta})}{\partial \sigma^2} = -\frac{N}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \sum_{n=1}^{N} (x[n] - A)^2 = \emptyset$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{n=1}^{N} (x[n] - A)^2 \Rightarrow \nabla_{\underline{\theta}} L(\underline{x};\underline{\theta}) = \underline{\emptyset}$$

$$\hat{\sigma}^2_{ML} = \frac{1}{N} \sum_{n=1}^{N} (x[n] - \hat{A}_{ML})^2 \Rightarrow VAR(\hat{\sigma}^2_{ML}) \geq VAR(\hat{\sigma}^2_{CR})$$

Example of ML estimator (II)

2.3

- Given N samples of a process that can modeled as $\underline{\mathbf{x}} = \theta\underline{\mathbf{1}} + \underline{\mathbf{w}}$, compute the ML estimator of its mean (θ).

  Note: $W[n]$ is a **Gaussian, stationary, colored noise**.

Generic expression of a multivariate Gaussian ▶

$$f_{\underline{w}}(\underline{\mathbf{w}}) = \frac{1}{\sqrt{(2\pi)^N \left|\underline{\underline{\mathbf{C}}}_{\underline{w}}\right|}} \exp\left[-\frac{\left[\underline{\mathbf{w}} - \underline{\mathbf{m}}_{\underline{w}}\right]^T \underline{\underline{\mathbf{C}}}_{\underline{w}}^{-1} \left[\underline{\mathbf{w}} - \underline{\mathbf{m}}_{\underline{w}}\right]}{2}\right]$$

STATIONARY
COLORED NOISE $\Rightarrow E\{w[n]\} = \emptyset \qquad r_w[l] \neq \sigma_w^2\, \delta[l$

$$\underline{\underline{C}}_w = \left[E\{w[n]\} = \emptyset\right] = \underline{\underline{R}}_w = \cancel{\left[\text{WHITE, STAT.}\right] = \sigma_w^2\, \underline{\underline{I}}}$$

$$\text{AS} \quad \underline{x} = \theta\, \underline{1} + \underline{w} \Rightarrow \underline{w} = \underline{x} - \theta\, \underline{1} \leftarrow \begin{cases} \underline{w} : N(\underline{\emptyset}, \underline{\underline{C}}_w) \\ \underline{x} : N(\underline{\theta}, \underline{\underline{C}}_w) \end{cases}$$

Example of ML estimator (II)

2.3

$$f_{\underline{w}}(\underline{x};\theta) = \frac{1}{\left[(2\pi)^N |\underline{\underline{C}}_w|\right]^{1/2}} \cdot \exp\left[-\frac{1}{2}\left[\underline{x}-\theta\underline{1}\right]^T \underline{\underline{C}}_w^{-1}\left[\underline{x}-\theta\underline{1}\right]\right]$$

$$L(\underline{x};\theta) = \ln f(\underline{x};\theta) = -\frac{1}{2}\ln (2\pi)^N |\underline{\underline{C}}_w| - \frac{1}{2}\left[\underline{x}-\theta\underline{1}\right]^T \underline{\underline{C}}_w^{-1}\left[\underline{x}-\theta\underline{1}\right]$$

$$\frac{\partial L(\underline{x};\theta)}{\partial\theta} = -\frac{1}{2}\frac{\partial}{\partial\theta}\left[\underline{x}^T\underline{\underline{C}}_w^{-1}\underline{x} - \underline{x}^T\underline{\underline{C}}_w^{-1}\theta\underline{1} - \theta\underline{1}^T\underline{\underline{C}}_w^{-1}\underline{x} + \theta^2\underline{1}^T\underline{\underline{C}}_w^{-1}\underline{1}\right] =$$

$$\frac{\partial L(\underline{x};\theta)}{\partial\theta} = \frac{1}{2}\left[\underline{x}^T\underline{\underline{C}}_w^{-1}\underline{1} + \underline{1}^T\underline{\underline{C}}_w^{-1}\underline{x} - 2\theta\,\underline{1}^T\underline{\underline{C}}_w^{-1}\underline{1}\right] =$$

$$= \left[\underline{x}^T\underline{\underline{C}}_w^{-1}\underline{1} = [\text{esc}] = \left[\underline{x}^T\underline{\underline{C}}_w^{-1}\underline{1}\right]^T = \underline{1}^T\left[\underline{x}^T\underline{\underline{C}}_w^{-1}\right]^T = \underline{1}^T\underline{\underline{C}}_w^{-1}\underline{x}\right] =$$

$$\frac{\partial L(\underline{x};\theta)}{\partial\theta} = \underline{1}^T\underline{\underline{C}}_w^{-1}\underline{x} - \theta\,\underline{1}^T\underline{\underline{C}}_w^{-1}\underline{1} = \underline{1}^T\underline{\underline{C}}_w^{-1}\underline{1}\left[\frac{\underline{1}^T\underline{\underline{C}}_w^{-1}\underline{x}}{\underline{1}^T\underline{\underline{C}}_w\underline{1}} - \theta\right]$$

Example of ML estimator (II)

$$\frac{\partial L(\underline{x};\theta)}{\partial \theta} = \underline{1}^T \underline{\underline{C}}_w^{-1} \underline{x} - \theta\, \underline{1}^T \underline{\underline{C}}_w^{-1} \underline{1} = \underline{1}^T \underline{\underline{C}}_w^{-1} \underline{1} \left[\frac{\underline{1}^T \underline{\underline{C}}_w^{-1} \underline{x}}{\underline{1}^T \underline{\underline{C}}_w \underline{1}} - \theta \right]$$

$$\frac{\partial L(\underline{x};\theta)}{\partial \theta} = \emptyset \Rightarrow \hat{\theta}_{ML}(\underline{x}) = \frac{\underline{1}^T \underline{\underline{C}}_w^{-1} \underline{x}}{\underline{1}^T \underline{\underline{C}}_w^{-1} \underline{1}}$$

$$\text{AS } \hat{\theta}_{ML}(\underline{x}) = \hat{\theta}_{CB}(x) \Rightarrow \text{VAR}(\hat{\theta}_{ML}) = \text{VAR}(\hat{\theta}_{CR}) = \frac{1}{\underline{1}^T \underline{\underline{C}}_w^{-1} \underline{1}}$$

Example of ML estimator (III)

2.3

- Given N independent samples of a Laplacian process ($\underline{\mathbf{x}} = m\underline{\mathbf{1}} + \underline{\boldsymbol{w}}$), we want to obtain the ML estimator of their mean (m) and diversity (λ):

  Note: $W[n]$ is a **Laplacian, stationary, white noise**. The vector parameter is $\underline{\boldsymbol{\theta}} = (m, \lambda)$

$$f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}\,;\underline{\boldsymbol{\theta}}) = [\text{INDEP.}] = \prod_{i=1}^{N} f_x(x_i\,;\underline{\boldsymbol{\theta}})$$

$$\boxed{f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}\,;\underline{\boldsymbol{\theta}}) = \prod_{i=1}^{N} \tfrac{1}{2\lambda} \exp\left[-\frac{|x_i - m|}{\lambda}\right]}$$

$$f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}\,;\underline{\boldsymbol{\theta}}) = \frac{1}{(2\lambda)^N} \exp\left[-\frac{\sum_{i=1}^{N} |x_i - m|}{\lambda}\right] \quad \Rightarrow L(\underline{\mathbf{x}}\,;\underline{\boldsymbol{\theta}}) = -N \ln 2\lambda - \frac{1}{\lambda}\sum_{i=1}^{N} |x_i - m|$$

The ML estimator implies:

$$\nabla_{\underline{\boldsymbol{\theta}}} L(\underline{\mathbf{x}}\,;\underline{\boldsymbol{\theta}}) = \left(\frac{\partial L(\underline{\mathbf{x}}\,;m)}{\partial m}, \frac{\partial L(\underline{\mathbf{x}}\,;\lambda)}{\partial \lambda}\right)^{\boldsymbol{T}} = \underline{\mathbf{0}}$$

Example of ML estimator (III)

2.3

$$\nabla_{\underline{\boldsymbol{\theta}}} L(\underline{\mathbf{x}}\,;\underline{\boldsymbol{\theta}}) = \left(\frac{\partial L(\underline{\mathbf{x}}\,;m)}{\partial m}, \frac{\partial L(\underline{\mathbf{x}}\,;\lambda)}{\partial \lambda}\right)^{\boldsymbol{T}} = \underline{\mathbf{0}} \qquad L(\underline{\mathbf{x}}\,;\underline{\boldsymbol{\theta}}) = -N\ln 2\lambda - \frac{1}{\lambda}\sum_{i=1}^{N}|x_i - m|$$

Computing $\widehat{m}_{ML}$:

$$\frac{\partial L(\underline{\mathbf{x}}\,;m)}{\partial m} = -\frac{1}{\lambda}\sum_{i=1}^{N}\frac{\partial}{\partial m}|x_i - m| = 0 \quad\Rightarrow\quad \frac{\partial L(\underline{\mathbf{x}}\,;m)}{\partial m} = \frac{1}{\lambda}\sum_{i=1}^{N}\operatorname{sign}(x_i - m) = 0$$

$$\sum_{i=1}^{N}\operatorname{sign}(x_i - m) = 0 \quad\Rightarrow\quad \boxed{\widehat{m}_{ML} = \text{Median}\,(x_1, x_2, \cdots, x_N)}$$

Computing $\hat{\lambda}_{ML}$:

$$\frac{\partial L(\underline{\mathbf{x}}\,;\lambda)}{\partial \lambda} = -\frac{N}{\lambda} + \frac{1}{\lambda^2}\sum_{i=1}^{N}|x_i - m| = 0 \quad\Rightarrow\quad \boxed{\hat{\lambda}_{ML} = \frac{1}{N}\sum_{i=1}^{N}|x_i - \widehat{m}_{ML}|}$$

ML and MAP Estimators

1. **Introduction**
   - Non efficient estimators
2. **Maximum Likelihood Estimator**
   - Properties of the ML estimators
   - Examples
3. **Maximum a Posteriori Estimator**
   - Bayesian framework
   - Examples

Bayesian framework

2.3

A Bayesian estimator models the parameter we are attempting to estimate as a **realization of a random variable**, instead of as a constant unknown parameter.

With this approach, we can include **the prior pdf of the parameter** ($f_{\boldsymbol{\theta}}(\theta)$) which summarizes our a priori knowledge about the parameter.

$$\hat{\theta}_{MAP} = \max_{\theta} f_{\underline{\mathbf{x}},\theta}\left(\underline{\mathbf{x}}\,,\theta\right) = \max_{\theta} f_{\underline{\mathbf{x}}}\left(\underline{\mathbf{x}}\,/\theta\right) f_{\boldsymbol{\theta}}(\theta)$$

Note: Conceptually, $f_{\underline{\mathbf{x}}}\left(\underline{\mathbf{x}}\,;\theta\right)$ is a family of pdf's and $f_{\underline{\mathbf{x}}}\left(\underline{\mathbf{x}}\,/\theta\right)$ is a conditional pfd

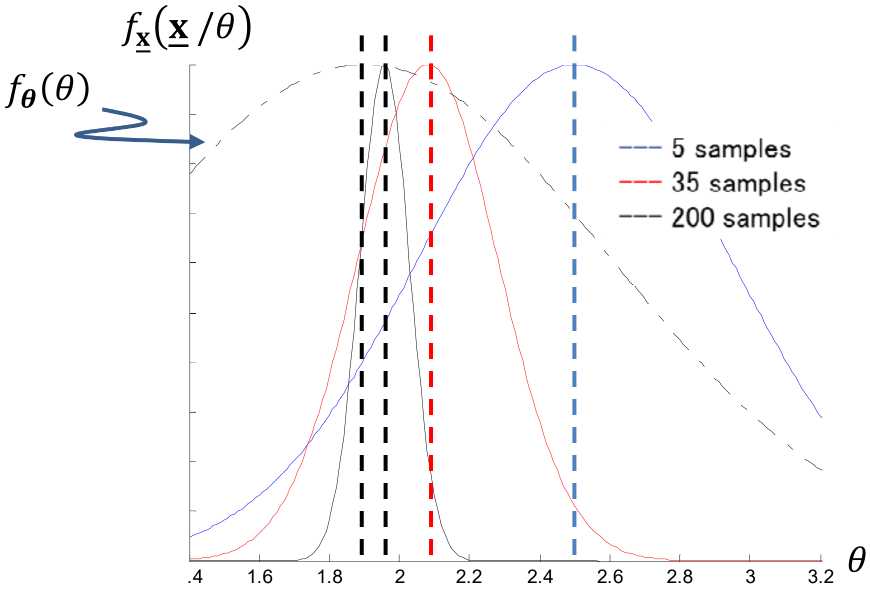
It is called **Maximum a Posterior** (**MAP**) estimator, since it can be formulated as:

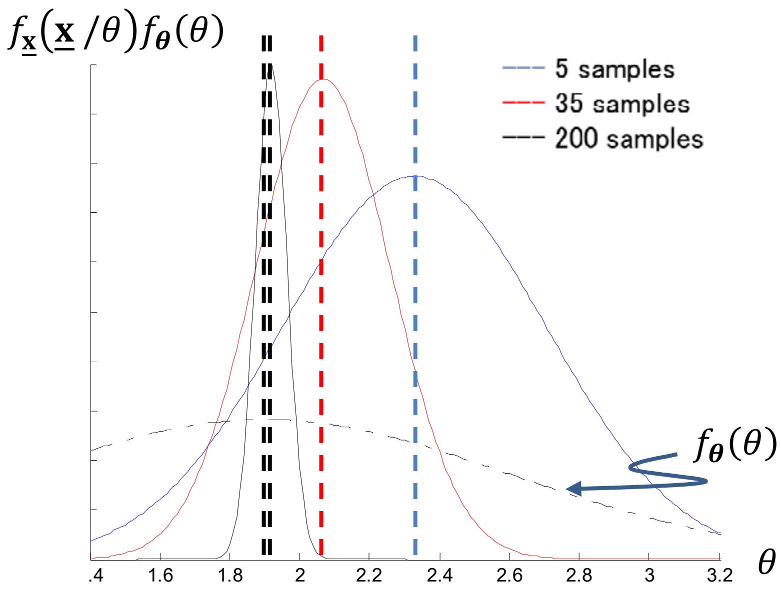
$$\hat{\theta}_{MAP} = \underline{\max_{\theta} f_{\theta}\left(\theta/\underline{\mathbf{x}}\right)} = \max_{\theta} \frac{f_{\underline{\mathbf{x}}}\left(\underline{\mathbf{x}}\,/\theta\right) f_{\boldsymbol{\theta}}(\theta)}{f_{\underline{\mathbf{x}}}\left(\underline{\mathbf{x}}\right)} = \max_{\theta} f_{\underline{\mathbf{x}}}\left(\underline{\mathbf{x}}\,/\theta\right) f_{\boldsymbol{\theta}}(\theta)$$

MAP versus ML

2.3

Increasing number of samples: The conditional probability ($f_{\underline{\mathbf{x}}}(\underline{\mathbf{x}}\,/\theta)$) will be sharper around θ_0 as the number of samples N increases. In this case, if the information provided by $f_{\boldsymbol{\theta}}(\theta)$ is correct, both estimators tend to be the same.

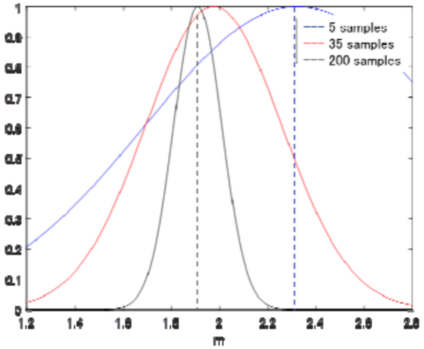


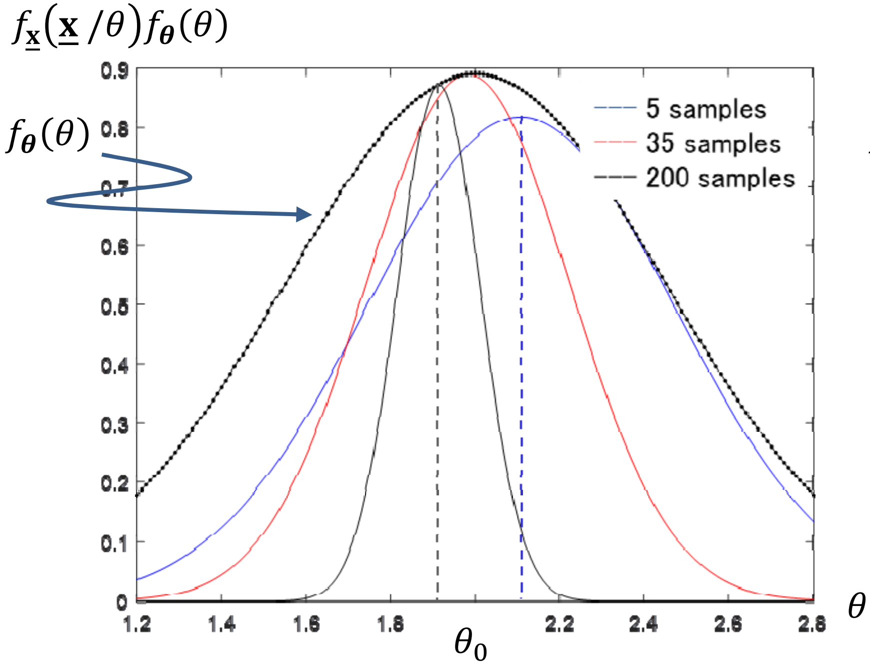


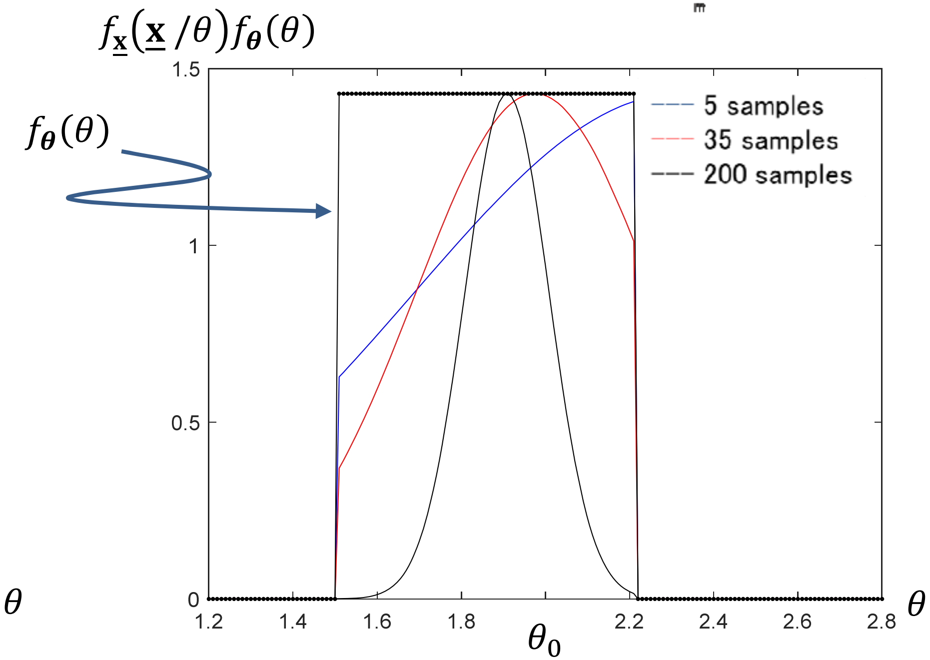
ML and MAP estimation of the mean (θ_0 = 2) of one realization of variable number N of Gaussian samples and a Gaussian prior. In the plot, likelihood functions have been normalized to better compare results

MAP with different priors

No prior: If we do not have any prior information about the parameter to be estimated, its pdf ($f_{\boldsymbol{\theta}}(\theta)$) is a constant and any possible value has the same likelihood. Then, the MAP estimator becomes the ML estimator.







MAP estimation of the mean (θ_0 = 2) of one realization of variable number N of Gaussian samples with a (1) Gaussian and (2) uniform prior. In the plot, likelihood functions have been normalized to better compare results

ML and MAP Estimators

1. **Introduction**
   - Non efficient estimators
2. **Maximum Likelihood Estimator**
   - Properties of the ML estimators
   - Examples
3. **Maximum a Posteriori Estimator**
   - Bayesian framework
   - Examples

Example of MAP estimator (I)

2.3

- Given N samples of a process that can modeled as $\underline{\mathbf{x}} = \mu \underline{\mathbf{1}} + \underline{\mathbf{w}}$, compute the MAP estimator of its mean (μ), knowing that it is a random variable with distribution $N(\mu_m, \sigma_m^2)$

  Note: $W[n]$ is a **Gaussian, stationary, colored noise**.

Generic expression of a multivariate Gaussian ▶

$$f_{\underline{w}}(\underline{\mathbf{w}}) = \frac{1}{\sqrt{(2\pi)^N \left|\underline{\underline{\mathbf{C}}}_{\underline{w}}\right|}} \exp\left[-\frac{\left[\underline{\mathbf{w}} - \underline{\mathbf{m}}_{\underline{w}}\right]^T \underline{\underline{\mathbf{C}}}_{\underline{w}}^{-1} \left[\underline{\mathbf{w}} - \underline{\mathbf{m}}_{\underline{w}}\right]}{2}\right]$$

$$f(\underline{x}, \theta) = f(\underline{x}/\theta) \cdot f(\theta) = [\theta = \mu] = f(\underline{x}/\mu) \cdot f(\mu) = f(x, \mu)$$

$$f(x, \mu) = \frac{1}{\sqrt{(2\pi)^N |\underline{\underline{C}}_x|}} \cdot \exp\left[-\frac{1}{2}[\underline{x} - \mu \underline{1}]^T \underline{\underline{C}}_x^{-1} [\underline{x} - \mu \underline{1}]\right] \cdot$$

$$\cdot \frac{1}{\sqrt{2\pi\sigma_\mu^2}} \cdot \exp\left[-\frac{[\mu - \mu_m]^2}{2\sigma_\mu^2}\right] =$$

Example of MAP estimator (I)

2.3

$$L(\underline{x},\mu) = -\frac{N}{2}\ln 2\pi\,|\underline{\underline{C}}_x| - \frac{1}{2}\left[\underline{x}-\mu\underline{1}\right]^T \underline{\underline{C}}_x^{-1}\left[\underline{x}-\mu\underline{1}\right] +$$

$$-\frac{1}{2}\ln 2\pi\sigma_\mu^2 - \frac{1}{2\sigma_\mu^2}\left[\mu-\mu_m\right]^2$$

$$\Rightarrow \nabla_\mu L(\underline{x},\mu) = \emptyset \quad \Rightarrow \frac{\partial}{\partial\mu} L(\underline{x},\mu) = \emptyset$$

$$\frac{\partial}{\partial\mu} L(\underline{x},\mu) = \frac{1}{2}\left(2\,\underline{1}^T\underline{\underline{C}}_x^{-1}\underline{x} - 2\mu\,\underline{1}^T\cdot\underline{\underline{C}}_x^{-1}\cdot\underline{1}\right) - \frac{1}{\sigma_\mu^2}(\mu-\mu_m) = \emptyset$$

$$\underline{1}^T\underline{\underline{C}}_x^{-1}\underline{x} - \mu\,\underline{1}^T\underline{\underline{C}}_x^{-1}\underline{1} - \frac{\mu}{\sigma_\mu^2} + \frac{\mu_m}{\sigma_\mu^2} = \emptyset$$

$$\mu\left[\underline{1}^T\underline{\underline{C}}_x^{-1}\underline{1} - \frac{1}{\sigma_\mu^2}\right] = \underline{1}^T\underline{\underline{C}}_x^{-1}\underline{x} + \frac{\mu_m}{\sigma_\mu^2}$$

Example of MAP estimator (I)

2.3

$$\mu \left[\underline{1}^T \underline{\underline{C}}_x^{-1} \underline{1} - \frac{1}{\sigma_\mu^2} \right] = \underline{1}^T \underline{\underline{C}}_x^{-1} \underline{x} + \frac{\mu_m}{\sigma_\mu^2}$$

$$\hat{\mu}_{MAP} = \frac{\underline{1}^T \underline{\underline{C}}_x^{-1} \underline{x} + \frac{\mu_m}{\sigma_\mu^2}}{\underline{1}^T \underline{\underline{C}}_x^{-1} \underline{1} + \frac{1}{\sigma_\mu^2}}$$

WHAT HAPPENS WHEN: $\sigma_w^2 \to \infty$; $\sigma_w^2 \to 0$?

Example of ML estimator (IV)

2.3

- We have 2 measures of a magnitude $z_i = x + v_i$, with different errors. The errors are Gaussian, zero mean, with variance σ_i^2 and independent. Compute the ML estimator of the magnitude to be measured.

$$f(\underline{z};\theta) = [\text{IND.}] = f(z_1;\theta)\cdot f(z_2;\theta) \Rightarrow x=\theta$$

$$f(\underline{z};x) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left[-\frac{[z_1-x]^2}{2\sigma_1^2}\right]\cdot \frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left[-\frac{(z_2-x)^2}{2\sigma_2^2}\right]$$

$$L(\underline{z};x) = -\frac{1}{2} \ln 2\pi\sigma_1^2 - \frac{(z_1-x)^2}{2\sigma_1^2} - \frac{1}{2}\ln 2\pi\sigma_2^2 - \frac{(z_2-x)^2}{2\sigma_2^2}$$

$$\frac{\partial}{\partial x} L(\underline{z};x) = \frac{1}{\sigma_1^2}(z_1-x) + \frac{1}{\sigma_2^2}(z_2-x) = 0$$

Example of ML estimator (IV)

$$x\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right) = \frac{z_1}{\sigma_1^2}+\frac{z_2}{\sigma_2^2} \Rightarrow \hat{x}_{ML} = \frac{1}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}}\left(\frac{z_1}{\sigma_1^2}+\frac{z_2}{\sigma_2^2}\right)$$

$$\hat{x}_{ML} = \frac{\sigma_1^2\cdot\sigma_2^2}{\sigma_1^2+\sigma_2^2}\left(\frac{z_1}{\sigma_1^2}+\frac{z_2}{\sigma_2}\right) \Rightarrow$$

ALL INFORMATION IS USEFUL ... WEIGHTED BY ITS QUALITY.

GENERALIZATION TO N SAMPLES:

$$\Rightarrow \hat{x}_{ML} = \frac{1}{\sum_{i=1}^{N}\frac{1}{\sigma_i^2}}\cdot\sum_{i=1}^{N}\frac{z_i}{\sigma_i^2}$$