

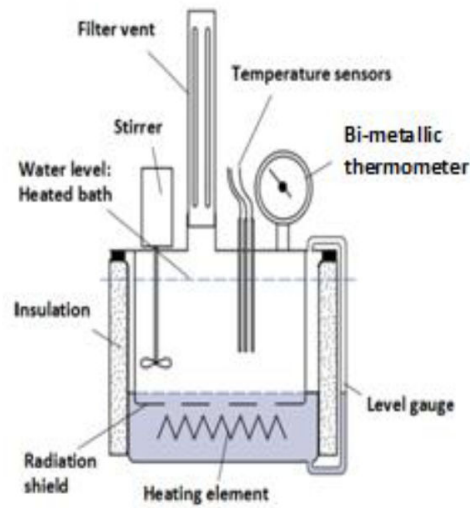
Bayes Estimation

Table of Contents

Motivating Example.....	1
The True Parametrization.....	3
The Sensor Noise Models.....	3
Dataset Size.....	3
A-Priori Knowledge on the Parameter	3
Bayesian Estimation of the Expected Value of a Random Variable - First Scenario: One Sensor Only.....	5
First Scenario: Using Only One Observation.....	5
Without Using Any A-Priori Knowledge on the Parameter	5
Using the A-Priori Knowledge: Bayes Estimation.....	6
Comparing the Results.....	6
Second Scenario: Using More than One Observation.....	7
Without Using Any A-Priori Knowledge on the Parameter	8
Using the A-Priori Knowledge: Bayes Estimation.....	8
Comparing the Results.....	9
Bayesian Estimation of the Expected Value of a Random Variable - Second Scenario: Using Both Sensors	10
First Scenario: Using Only Two Observation.....	11
Without Using Any A-Priori Knowledge on the Parameter	11
Using the A-Priori Knowledge: Bayes Estimation.....	12
Comparing the Results.....	13
Second Scenario: Using All the Observations.....	14
Without Using Any A-Priori Knowledge on the Parameter	14
Using the A-Priori Knowledge: Bayes Estimation.....	15
Comparing the Results.....	17

Motivating Example

Consider one more time the insulated stirrer experiment, introduced in the live scripts of the L7 part of the class material.



We want to measure the water temperature in an insulated stainless steel stirrer, as in Fig. 1. The water is heated with an immersion heater. We have both a thermocouple and an RTD sensor to perform the measurements. The experiment involves heating a predetermined amount of water and measuring temperature using both devices. The stirrer has good insulating capabilities, so it is assumed that the water maintains a constant temperature after the heating phase. The thermocouple and the RTD sensor provide measurements that are random variables, all with the same expected value but with different variances (and constant for each device).

This time consider the following scenarios:

1. Taking in consideration only one sensor (for example the thermocouple), collect $2N$ measurements and estimate the unknown constant water temperature, without any a priori knowledge on the effective temperature value.
2. Using the same N data, estimate the water temperature by means of a Bayes estimation approach, so taking into consideration not only the acquired data but also some a priori knowledge regarding the water temperature, i.e. the expected value, the variance and the probability density function of the temperature as random variable together with the expected value, the variance of the noise affecting the measurements and the cross-covariance of the noise and the unknown temperature. Compare the result with that obtained using the previous approach.
3. Consider now both the sensors and acquire at every time instant a sample using each of the sensors (so collect two observations at each time instant). Repeat such an acquisition N times, collecting $2N$ data. Estimate the unknown constant temperature by means of the weighted sampled average estimator, and compare the result with those obtained previously.
4. Using the dataset just collected, using both the sensors, apply a bayesian approach to estimate the constant temperature. Take into consideration as a-priori knowledge a the expected value, the variance (using both the sensors) and the probability density function of the temperature as random variable together with the expected value, the variance of the noise affecting the measurements and the cross-covariance of the noise (consider both the sensor) and the unknown temperature.

```
clear
close all
clc
```

The True Parametrization

```
theta_o = 48.5; % the effective value of the constant parameter to be estimated
```

The Sensor Noise Models

```
sigma2_v1 = 4; % the variance of the white noise stationary process describing the first
mu_v1 = 0; % the expected value for the white noise acting as additive disturbance for the first

sigma2_v2 = 10; % the variance of the white noise stationary process describing the
               % second sensor measurement disturbance
mu_v2 = 0; % the expected value for the white noise acting as additive disturbance for the second
```

Dataset Size

In every scenario we collect $2N$ samples of the random variable $d(t) = \theta + \epsilon(t)$ $\epsilon \sim \text{WN}(0, \lambda_\epsilon^2)$

```
N = 7; % we have to collect 2N data, using only one sensor (so 2N data) or both (N data)
```

A-Priori Knowledge on the Parameter ϑ

As a-priori knowledge regarding the unknown parameter ϑ to be estimated, what we need is

- information regarding the **a-priori distribution** $p(\vartheta)$ of the parameter ϑ (as random variable) to be estimated (i.e. the a-priori probability density function, for example a gaussian probability function, or an uniform one, or something else, if appropriate)
- information about the **covariance** between the parameter ϑ acting as random variable and the white noise affecting the observations $d(t)$.
- information concerning the **joint and marginal distributions** of the r.v. ϑ and the observations $d(\cdot)$

Let us assume that the **a-priori distribution** $p(\vartheta)$ is **gaussian**

$$p(\vartheta) = \frac{1}{\sqrt{2\pi}\sigma_\vartheta} e^{-\frac{(\vartheta-\mu_\vartheta)^2}{2\sigma_\vartheta^2}}$$

where we assume

$$\mu_\vartheta = 48.0 \quad \sigma_\vartheta^2 = 1$$

The assumptions we made allow us to state that the effective parameter ϑ^o belongs to the interval $[\mu_\vartheta - 3\sigma_\vartheta, \mu_\vartheta + 3\sigma_\vartheta] = [+45.0, +51.0]$ with confidence level 99%

```
% the parameters of the A-PRIORI distribution of theta_o
mu_theta = 47.50;
var_theta = 1;
```

Let us assume that the r.v. ϑ and the white noise $\epsilon(\cdot)$ are **independent random variables**, i.e.

$$\text{cov}(\vartheta, \epsilon) = 0$$

```
aPriori_cov_theta_noise = 0;
```

Finally, let us assume that the r.v. ϑ and the observations $d(t) = \theta^o + \epsilon(t)$, $\epsilon \sim \text{WN}(0, \lambda_\epsilon^2)$ are marginally and jointly gaussian distributed

$$\begin{bmatrix} d \\ \vartheta \end{bmatrix} \sim \mathcal{G}\left(\begin{bmatrix} d_m \\ \mu_\vartheta \end{bmatrix}, \begin{bmatrix} \lambda_{dd} & \lambda_{d\vartheta} \\ \lambda_{\vartheta d} & \lambda_{\vartheta\vartheta} \end{bmatrix}\right)$$

where $d_m = E[d(\cdot)]$, $\lambda_{dd} = \text{var}[d(\cdot)]$, $\lambda_{\vartheta d} = \text{cov}[\vartheta d]$.

Considering $d(t) = \theta^o + \epsilon(t)$, $\epsilon \sim \text{WN}(0, \lambda_\epsilon^2)$, we obtain

$$d_m = E[d(t)] = E[\vartheta] + E[\epsilon(t)] = \mu_\vartheta + 0 = \mu_\vartheta$$

and, since the two random variables ϑ and ϵ are independent

$$\lambda_{dd} = \text{var}[d(t)] = \text{var}[\vartheta] + \text{var}[\epsilon(t)] = \sigma_\vartheta^2 + \lambda_\epsilon^2$$

The effective value of λ_{dd} depends on the different scenarios (i.e. are we using only one sensor or both the available sensors?), so that value will be computed in every scenario. The a-priori expected value of the observations is always

```
d_m = mu_theta;
```

and the term $\lambda_{\vartheta d}$ is (owing that ϑ and ϵ are independent random variables)

$$\lambda_{\vartheta d} = \text{cov}[\vartheta \cdot d] = E[(\vartheta - \mu_{\vartheta}) \cdot (d - d_m)] = E[(\vartheta - \mu_{\vartheta}) \cdot (\vartheta + \epsilon - \mu_{\vartheta})] = E[(\vartheta - \mu_{\vartheta}) \cdot (\vartheta - \mu_{\vartheta})] = \sigma_{\vartheta}^2$$

Bayesian Estimation of the Expected Value of a Random Variable - First Scenario: One Sensor Only

Let's collect the data using only the first sensor

$$d_1(t) = \vartheta + \eta_1(t) \quad \vartheta \sim \mathcal{G}(\mu_{\vartheta}, \sigma_{\vartheta}^2) \quad \eta_1(\cdot) \sim \text{WN}(0, \lambda_1^2) \quad t = 1, 2, \dots, 2N$$

Assume that the white noise $\eta_1(\cdot)$ is

```
sigma1 = sqrt(sigma2_v1); % the standard deviation of the noise
D1data = theta_o + (mu_v1 + sigma1*randn(2*N,1)); % the data - what if the sensor is b
```

First Scenario: Using Only One Observation

$$d_1 = \vartheta + \eta_1 \quad \vartheta \sim \mathcal{G}(\mu_{\vartheta}, \sigma_{\vartheta}^2) \quad \eta_1(\cdot) \sim \text{WN}(0, \lambda_1^2)$$

Using only one measurement, how can we **estimate** the effective parameter value $\hat{\vartheta}$?

Without Using Any A-Priori Knowledge on the Parameter ϑ

The best (and only) estimate we can compute is indeed the observation we actually have:

```
d1 = D1data(1);
hat_theta_no_aPriorik = d1 % the estimate coincides with the only value available

hat_theta_no_aPriorik = 46.4711
```

Moreover, the variance of such a "raw" estimate is the a-priori variance of the observation d

$$\lambda_{dd} = \text{var}[d] = \text{var}[\vartheta] + \text{var}[\eta] = \sigma_{\vartheta}^2 + \lambda_1^2$$

```
var_hat_theta_no_aPriorik = sigma2_v1 + var_theta
```

```
var_hat_theta_no_aPriorik = 5
```

Using the A-Priori Knowledge: Bayes Estimation

Now let us apply the bayesian approach to estimate ϑ^o using the unique observation we collected so far, together with the a-priori knowledge we have regarding the parameter ϑ

```
lambda_dd = var_theta + sigma2_v1; % the avariance of the observations
```

```
lambda_theta_d = var_theta;
```

```
lambda_d_theta = lambda_theta_d';
```

```
Lambda_theta_d_MAT = [lambda_dd , lambda_d_theta; ...  
                      lambda_theta_d , var_theta ]; % the joint variance mat
```

The a-posteriori estimate $\hat{\vartheta}$ is

$$\begin{cases} \hat{\vartheta} = \vartheta_m + \lambda_{\vartheta d} \lambda_{dd}^{-1} (d - d_m) \\ \text{var}(\vartheta - \hat{\vartheta}) = \lambda_{\vartheta\vartheta} - \lambda_{\vartheta d} \lambda_{dd}^{-1} \lambda_{d\vartheta} \end{cases}$$

```
hat_thetaBayes = mu_theta + (lambda_theta_d/lambda_dd)*(d1-d_m)
```

```
hat_thetaBayes = 47.2942
```

```
var_hat_thetaBayes = var_theta - (lambda_theta_d/lambda_dd)*lambda_d_theta
```

```
var_hat_thetaBayes = 0.8000
```

Comparing the Results

```
a =3; % in order to build the confidence interval, we will use +/- a*std() as extre
```

```
figure('Units','normalized','Position',[0.1, 0.1,0.90,0.80]);
```

```
hat_theta = [hat_theta_no_aPriorik ; hat_thetaBayes]; % the estimated mean value
```

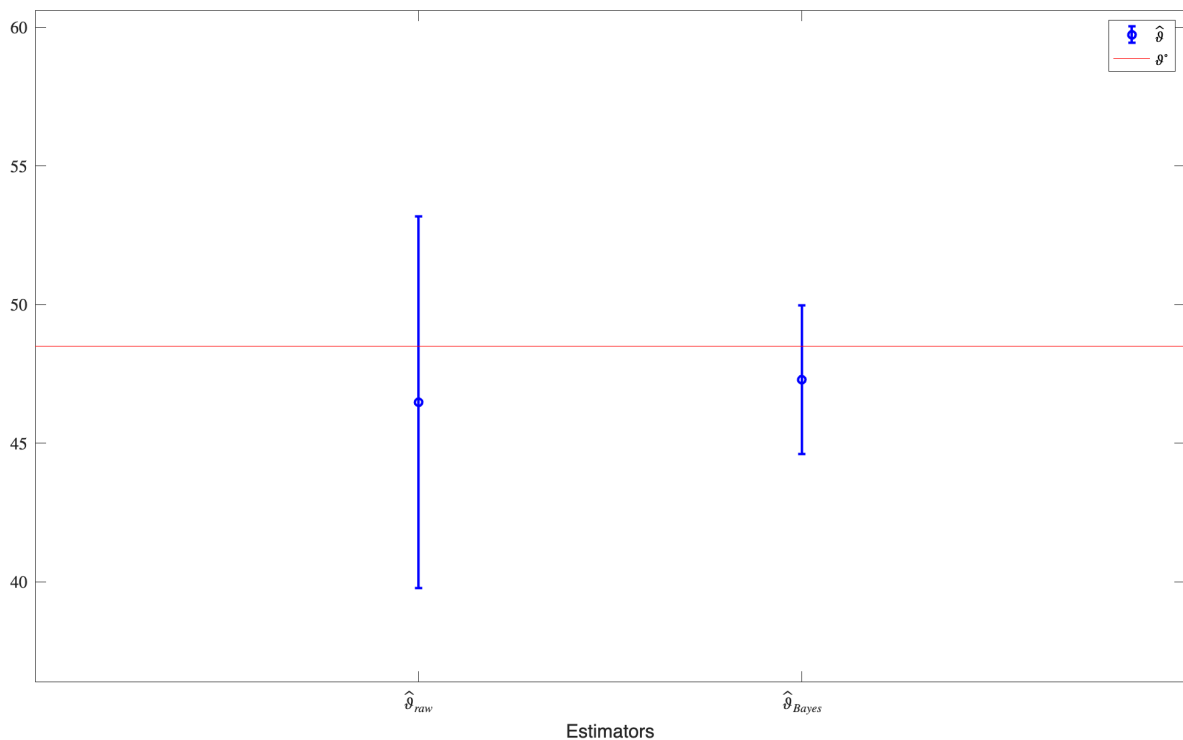
```

hat_sigma = [sqrt(var_hat_theta_no_aPrioriK); sqrt(var_hat_thetaBayes)];
% the std. dev. of the estimated mean value

R = numel(hat_theta);

errorbar(hat_theta,a*hat_sigma,'ob','linewidth',2)
hold on
plot((0:R+1),theta_o*ones(1,R+2),'r')
hold off
axis([0 R+1 0.75*theta_o 1.25*theta_o])
legend('$\hat{\vartheta}$','$\hat{\vartheta}^{\circ}$','Interpreter','latex')
xlabel('Estimators')
ax = gca;
ax.XTick = [1 2];
ax.XTickLabel = {'$\hat{\vartheta}_{raw}$','$\hat{\vartheta}_{Bayes}$'};
ax.TickLabelInterpreter = "latex";
%title(['Sample mean with ',num2str(a),' sigma interval'])
set(gca, 'FontSize', 14);

```



Second Scenario: Using More than One Observation

This time we collect $2N$ observations

$$d_1(t) = \vartheta^o + \eta_1(t) \quad \eta_1(\cdot) \sim \text{WN}(0, \lambda_1^2) \quad t = 1, 2, \dots, 2N$$

How can we **estimate** the effective parameter value $\hat{\vartheta}$?

Without Using Any A-Priori Knowledge on the Parameter ϑ

The best estimate we can compute is indeed the sampled average of the observations:

```
hat_theta_2N_no_aPrioriK = mean(D1data)
```

```
hat_theta_2N_no_aPrioriK = 47.8059
```

Moreover, the variance of such a nestimate is the sampled variance of the measurements:

```
var_hat_theta_2N_no_aPrioriK = var(D1data)
```

```
var_hat_theta_2N_no_aPrioriK = 3.7498
```

Using the A-Priori Knowledge: Bayes Estimation

Now let us apply the bayesian approach to estimate ϑ^o using the $2N$ observation we collected so far, together with the a-priori knowledge we have regarding the parameter ϑ

$$\Lambda_{dd} = \begin{bmatrix} \text{var}[d(1)] & \text{cov}[d(1) \cdot d(2)] & \text{cov}[d(1) \cdot d(3)] & \dots & \text{cov}[d(1) \cdot d(2N)] \\ \text{cov}[d(2) \cdot d(1)] & \text{var}[d(2)] & \text{cov}[d(2) \cdot d(3)] & \dots & \text{cov}[d(2) \cdot d(2N)] \\ \text{cov}[d(3) \cdot d(1)] & \text{cov}[d(3) \cdot d(2)] & \text{var}[d(3)] & \dots & \text{cov}[d(3) \cdot d(2N)] \\ & & \ddots & & \\ \text{cov}[d(2N) \cdot d(1)] & \text{cov}[d(2N) \cdot d(2)] & \dots & \text{cov}[d(2N) \cdot d(2N-1)] & \text{var}[d(2N)] \end{bmatrix}$$

where

$$\text{var}[d(i)] = \text{var}[\vartheta] + \text{var}[\eta(i)] = \sigma_{\vartheta}^2 + \lambda_{\eta}^2 \quad i = 1, 2, \dots, 2N$$

$$\text{cov}[d(i) \cdot d(j)] = \text{E}[(d(i) - d_m) \cdot (d(j) - d_m)] = \text{E}[(\vartheta + \eta(i) - \mu_{\vartheta}) \cdot (\vartheta + \eta(j) - \mu_{\vartheta})] = \text{var}[\vartheta] + \text{E}[\eta(i) \cdot \eta(j)] = \sigma_{\vartheta}^2 \quad \forall i \neq j$$

```
Lambda_dd = ones(2*N).*var_theta + diag(ones(2*N,1).*(sigma2_v1));  
% the variance of the observations is now a square matrix, of order 2N  
% the observations are independent of each other and each has the same variance
```


$$\Lambda_{\vartheta d} = [\text{cov}[\vartheta \cdot d(1)] \quad \text{cov}[\vartheta \cdot d(2)] \quad \text{cov}[\vartheta \cdot d(3)] \quad \cdots \quad \text{cov}[\vartheta \cdot d(2N)]] \quad \Lambda_{d\vartheta} = [\Lambda_{\vartheta d}]^T$$

where

$$\text{cov}[\vartheta \cdot d(i)] = E[(\vartheta - \mu_{\vartheta}) \cdot (d(i) - d_m)] = E[(\vartheta - \mu_{\vartheta}) \cdot (\vartheta + \eta(i) - \mu_{\vartheta})] = E[(\vartheta - \mu_{\vartheta}) \cdot (\vartheta - \mu_{\vartheta})] = \sigma_{\vartheta}^2 \quad \forall i$$

```
Lambda_theta_d = var_theta .* ones(1, 2*N); % a row matrix
Lambda_d_theta = Lambda_theta_d'; % a column matrix

Lambda_theta_d_MAT = [Lambda_dd, Lambda_d_theta; ...
                      Lambda_theta_d, var_theta]; % the joint variance matrix
```

The a-posteriori estimate $\hat{\vartheta}$ is

$$\begin{cases} \hat{\vartheta} = \vartheta_m + \Lambda_{\vartheta d} \Lambda_{dd}^{-1} (d - d_m) \\ \text{var}(\vartheta - \hat{\vartheta}) = \Lambda_{\vartheta\vartheta} - \Lambda_{\vartheta d} \Lambda_{dd}^{-1} \Lambda_{d\vartheta} \end{cases}$$

```
hat_theta2N_Bayes = mu_theta + Lambda_theta_d * inv(Lambda_dd)*(D1data-d_m)
```

```
hat_theta2N_Bayes = 47.7379
```

```
var_hat_theta2N_Bayes = var_theta - Lambda_theta_d * (Lambda_dd \ Lambda_d_theta)
```

```
var_hat_theta2N_Bayes = 0.2222
```

Comparing the Results

```
figure('Units','normalized','Position',[0.1, 0.1,0.90,0.80]);

hat_theta = [hat_theta_2N_no_aPrioriK ; hat_theta2N_Bayes]; % the estimated mean value
hat_sigma = [sqrt(var_hat_theta_2N_no_aPrioriK); sqrt(var_hat_theta2N_Bayes)];
% the std. dev. of the estimated mean value

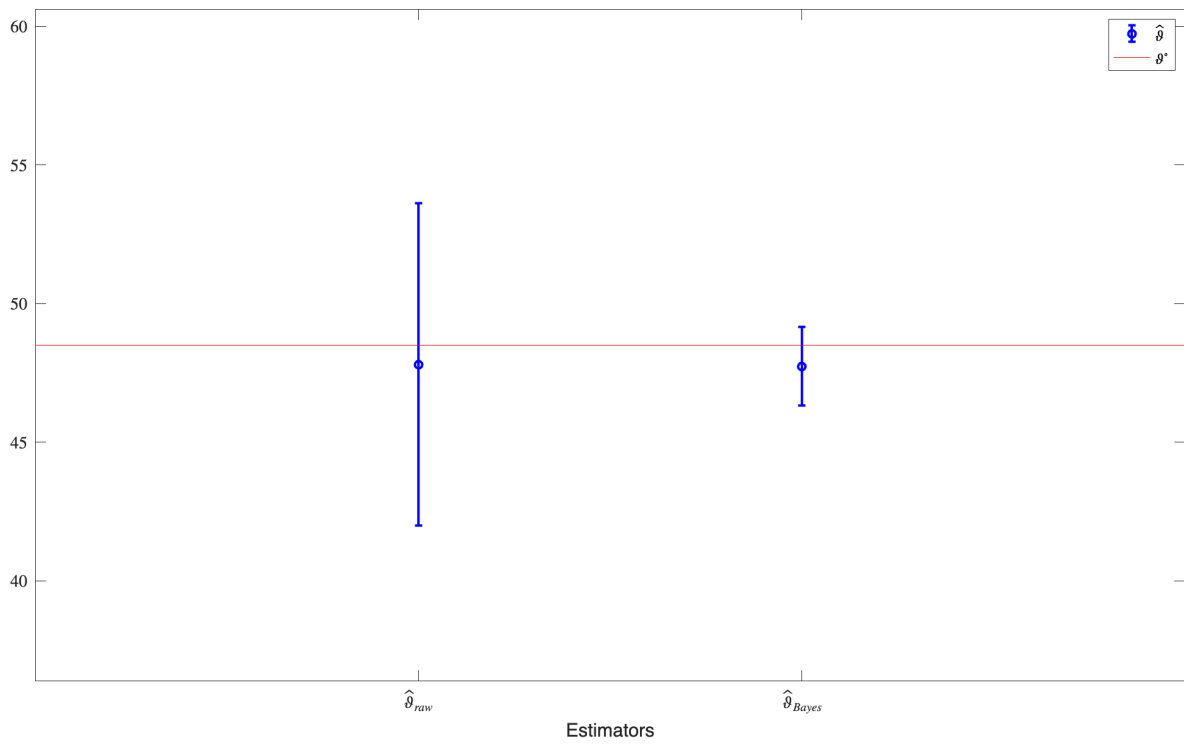
R = numel(hat_theta);

errorbar(hat_theta,a*hat_sigma,'ob','linewidth',2)
hold on
plot((0:R+1),theta_o*ones(1,R+2),'r')
hold off
axis([0 R+1 0.75*theta_o 1.25*theta_o])
legend('$\hat{\vartheta}$','$\vartheta^{\circ}$','Interpreter','latex')
xlabel('Estimators')
ax = gca;
ax.XTick = [1 2];
```

```

ax.XTickLabel = {'$\hat{\vartheta}_{raw}$', '$\hat{\vartheta}_{Bayes}$'};
ax.TickLabelInterpreter = "latex";
%title(['Sample mean with ', num2str(a), ' sigma interval'])
set(gca, 'FontSize', 14);

```



Bayesian Estimation of the Expected Value of a Random Variable - Second Scenario: Using Both Sensors

Suppose now to employ both the sensors and **to acquire the measurements using each sensor**, having in total two sets of observations:

$$\begin{cases} d_1(t) = \vartheta + \eta_1(t) & \eta_1(\cdot) \sim \text{WN}(0, \lambda_1^2) \\ d_2(t) = \vartheta + \eta_2(t) & \eta_2(\cdot) \sim \text{WN}(0, \lambda_2^2) \end{cases} \quad t = 1, 2, \dots, N$$

Collect the data using the first and the second sensor:

```

sigma1 = sqrt(sigma2_v1); % the standard deviation of the noise for the 1st device
D1data = theta_o + (mu_v1 + sigma1*randn(N,1)); % the data - what if the sensor is biased

sigma2 = sqrt(sigma2_v2); % the standard deviation of the noise for the 2nd sensor
D2data = theta_o + (mu_v2 + sigma2*randn(N,1)); % the data - what if the sensor is biased

```

First Scenario: Using Only Two Observation

Suppose we acquire a measurement for each sensor, using them simultaneously

$$d_1 = \vartheta + \eta_1 \quad \vartheta \sim \mathcal{G}(\mu_\vartheta, \sigma_\vartheta^2) \quad \eta_1(\cdot) \sim \text{WN}(0, \lambda_1^2)$$

$$d_2 = \vartheta + \eta_2 \quad \vartheta \sim \mathcal{G}(\mu_\vartheta, \sigma_\vartheta^2) \quad \eta_2(\cdot) \sim \text{WN}(0, \lambda_2^2)$$

Using only these measurements, how can we **estimate** the effective parameter value $\hat{\vartheta}$?

Without Using Any A-Priori Knowledge on the Parameter ϑ

The best estimate we can compute is indeed the weighted sampled average of the observations we actually have:

```
d1 = D1data(1);  
d2 = D2data(1);  
% the observations d1 and d2
```

The weighted sampled-average estimator is

$$\hat{\vartheta}_N = \sum_{i=1}^N \alpha(i) d(i) \quad \sum_{i=1}^N \alpha(i) = 1$$

where

$$\alpha(i) = \frac{1}{\text{var}[d(i)]} \cdot \frac{1}{\sum_{i=1}^N \frac{1}{\text{var}[d(i)]}}$$

Thus

```
alpha1 = (1/sigma2_v1); % the weights for d1  
alpha2 = (1/sigma2_v2); % the weights for d2  
  
hat_theta_W_no_aPriorik = ((alpha1 .* d1) + (alpha2 .* d2)) / ((1/sigma2_v1) + (1/sigma2_v2))  
  
hat_theta_W_no_aPriorik = 51.3866
```

The variance of the estimate takes the expression

$$\text{var}(\hat{\vartheta}_N) = \frac{1}{\sum_{i=1}^N \frac{1}{\text{var}[d(i)]}}$$

```
var_hat_theta_W_no_aPrioriK = 1/((1/sigma2_v1) + (1/sigma2_v2))
```

```
var_hat_theta_W_no_aPrioriK = 2.8571
```

Using the A-Priori Knowledge: Bayes Estimation

Now let us apply the bayesian approach to estimate ϑ^o using the observations d_1 and d_2 , together with the a-priori knowledge we have regarding the parameter ϑ

$$\Lambda_{dd} = \begin{bmatrix} \text{var}[d_1] & \text{cov}[d_1 \cdot d_2] \\ \text{cov}[d_2 \cdot d_1] & \text{var}[d_2] \end{bmatrix}$$

where

$$\text{var}[d_i] = \text{var}[\vartheta] + \text{var}[\eta_i] = \sigma_\vartheta^2 + \lambda_i^2 \quad i = 1, 2$$

$$\text{cov}[d_1 \cdot d_2] = E[(d_1 - d_m) \cdot (d_2 - d_m)] = E[(\vartheta + \eta_1 - \mu_\vartheta) \cdot (\vartheta + \eta_2 - \mu_\vartheta)] = \text{var}[\vartheta] + E[\eta_1 \cdot \eta_2] = \sigma_\vartheta^2$$

```
Lambda_dd = [var_theta + sigma2_v1, var_theta; ...
             var_theta, var_theta + sigma2_v2]; % the covariance matrix
```

$$\Lambda_{\vartheta d} = [\text{cov}[\vartheta \cdot d_1] \quad \text{cov}[\vartheta \cdot d_2]] \quad \Lambda_{d\vartheta} = [\Lambda_{\vartheta d}]^T$$

where

$$\text{cov}[\vartheta \cdot d_i] = E[(\vartheta - \mu_\vartheta) \cdot (d_i - d_m)] = E[(\vartheta - \mu_\vartheta) \cdot (\vartheta + \eta_i - \mu_\vartheta)] = E[(\vartheta - \mu_\vartheta) \cdot (\vartheta - \mu_\vartheta)] = \sigma_\vartheta^2 \quad \forall i$$

```
Lambda_theta_d = [var_theta, var_theta];
Lambda_d_theta = Lambda_theta_d';

Lambda_theta_d_MAT = [lambda_dd, lambda_d_theta; ...
                    lambda_theta_d, var_theta]; % the joint covariance matrix
```

The a-posteriori estimate $\hat{\vartheta}$ is

$$\begin{cases} \hat{\vartheta} = \vartheta_m + \Lambda_{\vartheta d} \Lambda_{dd}^{-1} (d - d_m) \\ \text{var}(\vartheta - \hat{\vartheta}) = \Lambda_{\vartheta\vartheta} - \Lambda_{\vartheta d} \Lambda_{dd}^{-1} \Lambda_{d\vartheta} \end{cases}$$

```
D = [d1; d2]; % the vector containing the observations
hat_thetaW_Bayes = mu_theta + Lambda_theta_d * (Lambda_dd \ (D-d_m*ones(size(D))))
```

```
hat_thetaW_Bayes = 48.5076
```

```
% the estimate
```

```
var_hat_thetaW_Bayes = var_theta - Lambda_theta_d * (Lambda_dd \ Lambda_d_theta)
```

```
var_hat_thetaW_Bayes = 0.7407
```

```
% the variance
```

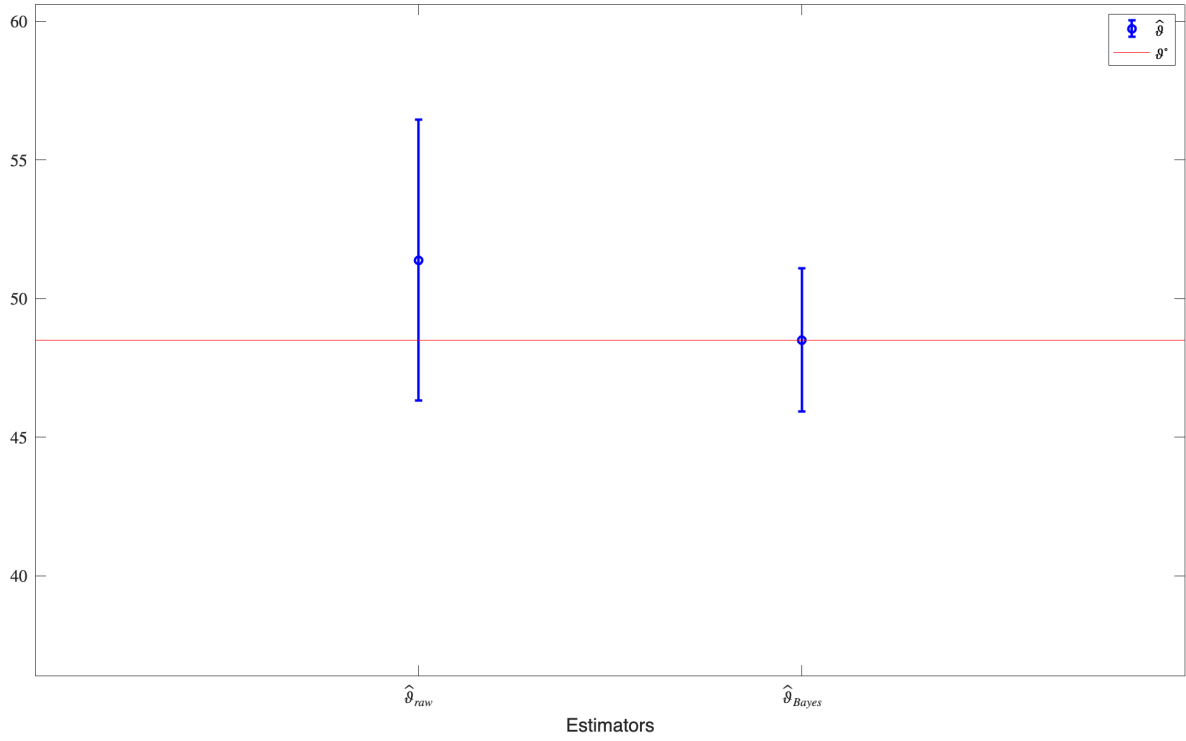
Comparing the Results

```
figure('Units','normalized','Position',[0.1, 0.1,0.90,0.80]);

hat_theta = [hat_theta_W_no_aPrioriK ; hat_thetaW_Bayes]; % the estimated mean value
hat_sigma = [sqrt(var_hat_theta_W_no_aPrioriK); sqrt(var_hat_thetaW_Bayes)];
% the std. dev. of the estimated mean value

R = numel(hat_theta);

errorbar(hat_theta,a*hat_sigma,'ob','linewidth',2)
hold on
plot((0:R+1),theta_o*ones(1,R+2),'r')
hold off
axis([0 R+1 0.75*theta_o 1.25*theta_o])
legend('$\hat{\vartheta}$','$\vartheta^{\circ}$','Interpreter','latex')
xlabel('Estimators')
ax = gca;
ax.XTick = [1 2];
ax.XTickLabel = {'$\hat{\vartheta}_{\text{raw}}$','$\hat{\vartheta}_{\text{Bayes}}$'};
ax.TickLabelInterpreter = "latex";
%title(['Sample mean with ',num2str(a),' sigma interval'])
set(gca, 'FontSize', 14);
```



Second Scenario: Using All the Observations

Suppose we acquire N measurement for each sensor, using them simultaneously

$$d_1(t) = \vartheta + \eta_1(t) \quad \vartheta \sim \mathcal{G}(\mu_\vartheta, \sigma_\vartheta^2) \quad \eta_1(\cdot) \sim \text{WN}(0, \lambda_1^2) \quad t = 1, 2, \dots, N$$

$$d_2(t) = \vartheta + \eta_2(t) \quad \vartheta \sim \mathcal{G}(\mu_\vartheta, \sigma_\vartheta^2) \quad \eta_2(\cdot) \sim \text{WN}(0, \lambda_2^2) \quad t = 1, 2, \dots, N$$

How can we **estimate** the effective parameter value $\hat{\vartheta}$?

Without Using Any A-Priori Knowledge on the Parameter ϑ

The best estimate we can compute is the weighted sampled average of the observations:

When $d(i) \in D_i$

$$\forall j : d(j) \in D_i, \quad \text{var}(d(j)) = \sigma_{v_i}^2 \quad \Rightarrow \quad \alpha(j) = \frac{1}{\text{var}[d(j)]} \cdot \frac{1}{\sum_{i=1}^N \frac{1}{\text{var}[d(i)]}} = \frac{1}{\sigma_{v_1}^2} \cdot \frac{1}{\frac{N_1}{\sigma_{v_1}^2} + \frac{N_2}{\sigma_{v_2}^2}}$$

Thus

$$\hat{\vartheta}_{WN} = \frac{\sum_{p=1}^{N_1} \frac{d_{1,p}}{\sigma_{v_1}^2} + \sum_{q=1}^{N_2} \frac{d_{2,q}}{\sigma_{v_2}^2}}{\frac{N_1}{\sigma_{v_1}^2} + \frac{N_2}{\sigma_{v_2}^2}}. \quad \sigma_{\hat{\vartheta}_{WN}}^2 = \left(\frac{N_1}{\sigma_{v_1}^2} + \frac{N_2}{\sigma_{v_2}^2} \right)^{-1}$$

whereas

$$\hat{\vartheta}_N = \frac{1}{N} \sum_{i=1}^N d(i) \quad \text{var}(\hat{\vartheta}_N) = \frac{1}{N^2} \sum_{i=1}^N \text{var}(d(i)) = \frac{N_1 \sigma_{v_1}^2 + N_2 \sigma_{v_2}^2}{N^2}$$

```
alpha1 = (1/sigma2_v1)*ones(N,1); % the weights for the D1 dataset
alpha2 = (1/sigma2_v2)*ones(N,1); % the weights for the D2 dataset
```

```
hat_theta_W_N__no_aPriorik = (sum(alpha1 .* D1data) + sum(alpha2 .* D2data)) / ((N/sig
```

```
hat_theta_W_N__no_aPriorik = 48.9320
```

Moreover, the variance of such a nestimate is the sampled variance of the measurements:

```
var_hat_theta_W_N__no_aPriorik = 1/((N/sigma2_v1) + (N/sigma2_v2))
```

```
var_hat_theta_W_N__no_aPriorik = 0.4082
```

Using the A-Priori Knowledge: Bayes Estimation

Now let us apply the bayesian approach to estimate ϑ^o using the $2N$ observation we collected so far, together with the a-priori knowledge we have regarding the parameter ϑ

$$\Lambda_{dd} = \begin{bmatrix} \Lambda_{D_1 D_1} & \Lambda_{D_1 D_2} \\ \Lambda_{D_2 D_1} & \Lambda_{D_2 D_2} \end{bmatrix}$$

where

$$\Lambda_{D_i D_i} = \begin{bmatrix} \text{var}[d_i(1)] & \text{cov}[d_i(1) \cdot d_i(2)] & \text{cov}[d_i(1) \cdot d_i(3)] & \dots & \text{cov}[d_i(1) \cdot d_i(N)] \\ \text{cov}[d_i(2) \cdot d_i(1)] & \text{var}[d_i(2)] & \text{cov}[d_i(2) \cdot d_i(3)] & \dots & \text{cov}[d_i(2) \cdot d_i(N)] \\ \text{cov}[d_i(3) \cdot d_i(1)] & \text{cov}[d_i(3) \cdot d_i(2)] & \text{var}[d_i(3)] & \dots & \text{cov}[d_i(3) \cdot d_i(N)] \\ & & \ddots & & \\ \text{cov}[d_i(N) \cdot d_i(1)] & \text{cov}[d_i(N) \cdot d_i(2)] & \dots & \text{cov}[d_i(N) \cdot d_i(N-1)] & \text{var}[d_i(N)] \end{bmatrix} \quad i = 1, 2$$

and

$$\text{var}[d_i(t)] = \text{var}[\vartheta] + \text{var}[\eta_i(t)] = \sigma_\vartheta^2 + \lambda_i^2 \quad i = 1, 2 \quad t = 1, 2, \dots, N$$

$$\text{cov}[d_k(i) \cdot d_k(j)] = \text{E}[(d_k(i) - d_m) \cdot (d_k(j) - d_m)] = \text{E}[(\vartheta + \eta_k(i) - \mu_\vartheta) \cdot (\vartheta + \eta_k(j) - \mu_\vartheta)] = \text{var}[\vartheta] + \text{E}[\eta_k(i) \cdot \eta_k(j)] = \sigma_\vartheta^2 \quad \forall i \neq j \quad k = 1, 2$$

Moreover

$$\Lambda_{D_1 D_2} = \text{E} \left\{ \begin{bmatrix} d_1(1) - d_m \\ d_1(2) - d_m \\ \vdots \\ d_1(N) - d_m \end{bmatrix} \cdot \begin{bmatrix} d_2(1) - d_m & d_2(2) - d_m & d_2(3) - d_m & \dots & d_2(N) - d_m \end{bmatrix}^\top \right\} \quad \Lambda_{D_2 D_1} = [\Lambda_{D_1 D_2}]^\top$$

```
Lambda_dd_1 = ones(N).*var_theta + diag(ones(N,1).*(sigma2_v1));
Lambda_dd_2 = ones(N).*var_theta + diag(ones(N,1).*(sigma2_v2));

Lambda_d1d2 = ones(N).*var_theta; Lambda_d2d1 = Lambda_d1d2';

Lambda_DD = [Lambda_dd_1, Lambda_d1d2; ...
             Lambda_d2d1, Lambda_dd_2];

% the variance of the observations is now a square matrix, of order 2N
% the observations are independent of each other and each has the same variance
```

$$\Lambda_{\vartheta D} = \begin{bmatrix} \text{cov}[\vartheta \cdot d_1(1)] & \text{cov}[\vartheta \cdot d_1(2)] & \dots & \text{cov}[\vartheta \cdot d_1(N)] & \text{cov}[\vartheta \cdot d_2(1)] & \text{cov}[\vartheta \cdot d_2(2)] & \dots & \text{cov}[\vartheta \cdot d_2(N)] \end{bmatrix} \quad \Lambda_{d\vartheta} = [\Lambda_{\vartheta d}]^\top$$

where

$$\text{cov}[\vartheta \cdot d_k(i)] = \text{E}[(\vartheta - \mu_\vartheta) \cdot (d_k(i) - d_m)] = \text{E}[(\vartheta - \mu_\vartheta) \cdot (\vartheta + \eta_k(i) - \mu_\vartheta)] = \text{E}[(\vartheta - \mu_\vartheta) \cdot (\vartheta - \mu_\vartheta)] = \sigma_\vartheta^2 \quad \forall i \quad k = 1, 2$$

```
Lambda_theta_D = [var_theta .* ones(1, N) var_theta .* ones(1, N)]; % a row matrix
Lambda_D_theta = Lambda_theta_D'; % a column matrix

Lambda_theta_D_MAT = [Lambda_DD, Lambda_D_theta; ...
```



```
Lambda_theta_D , var_theta ]; % the joint variance mat
```

The a-posteriori estimate $\hat{\vartheta}$ is

$$\begin{cases} \hat{\vartheta} = \vartheta_m + \Lambda_{\vartheta D} \Lambda_{DD}^{-1} (D - D_m) \\ \text{var}(\vartheta - \hat{\vartheta}) = \Lambda_{\vartheta\vartheta} - \Lambda_{\vartheta D} \Lambda_{DD}^{-1} \Lambda_{D\vartheta} \end{cases}$$

```
D_data = [D1data; D2data];
```

```
hat_theta_W_2N_Bayes = mu_theta + Lambda_theta_D * inv(Lambda_DD)*(D_data-d_m)
```

```
hat_theta_W_2N_Bayes = 48.5169
```

```
var_hat_theta_W_2N_Bayes = var_theta - Lambda_theta_D * (Lambda_DD \ Lambda_D_theta)
```

```
var_hat_theta_W_2N_Bayes = 0.2899
```

Comparing the Results

```
figure('Units','normalized','Position',[0.1, 0.1,0.90,0.80]);
```

```
hat_theta = [hat_theta_W_N__no_aPrioriK ; hat_theta_W_2N_Bayes]; % the estimated m
hat_sigma = [sqrt(var_hat_theta_W_N__no_aPrioriK); sqrt(var_hat_theta_W_2N_Bayes)];
% the std. dev. of the estimated mean value
```

```
R = numel(hat_theta);
```

```
errorbar(hat_theta,a*hat_sigma,'ob','linewidth',2)
```

```
hold on
```

```
plot((0:R+1),theta_o*ones(1,R+2),'r')
```

```
hold off
```

```
axis([0 R+1 0.75*theta_o 1.25*theta_o])
```

```
legend('$\hat{\vartheta}$','$\vartheta^{\circ}$','Interpreter','latex')
```

```
xlabel('Estimators')
```

```
ax = gca;
```

```
ax.XTick = [1 2];
```

```
ax.XTickLabel = {'$\hat{\vartheta}_{\text{raw}}$','$\hat{\vartheta}_{\text{Bayes}}$'};
```

```
ax.TickLabelInterpreter = "latex";
```

```
%title(['Sample mean with ',num2str(a),' sigma interval'])
```

```
set(gca, 'FontSize', 14);
```

