$$y(t) = a \exp(-bt) + \epsilon(t), \quad \epsilon(t) \sim \mathcal{N}(0, \sigma^2)$$

with $\sigma = 0.02$.

Original simulation values are $a = 0.4$ and $b = 3$. We denote $\theta := (a, b)^T$

1. Likelihood density.

   - Write the likelihood of θ given $y(t)$

     We have from the Bayes formula:

     $$p(\theta|y(t)) = \frac{p\big(y(t)|\theta\big)p(\theta)}{p\big(y(t)\big)} \propto p\big(y(t)|\theta\big)p(\theta)$$

     The likelihood is:

     $$\mathcal{L}(\theta) = p\big(y(t)|\theta\big) = \prod_{i=1}^{n} p\big(y(t_i)|\theta\big) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-1/2\left(\frac{y(t_i) - ae^{-bt_i}}{\sigma}\right)^2}$$

     So the likelihood is,

     $$\boxed{\mathcal{L}(\theta) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\big(y(t_i) - ae^{-bt_i}\big)^2\right)}$$

   - The negative log likelihood

     The negative log likelihood $\ell(\theta)$ of θ is given by:

     $$= -\log(\mathcal{L}(\theta))$$

     $$= -\log\left(\left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\big(y(t_i) - ae^{-bt_i}\big)^2\right)\right)$$

     $$= \frac{n}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{n}\big(y(t_i) - ae^{-bt_i}\big)^2$$

     So, the negative log likelihood is:

     $$\boxed{\ell(\theta) = \frac{n}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{n}\big(y(t_i) - ae^{-bt_i}\big)^2}$$

2. Optimization:

The maximum likelihood estimates $\hat{\theta}_{ML}$ is obtained by maximizing the likelihood or equivantly by minimizing the negative log likelihood. i.e:

$$\hat{\theta}_{ML} = \underset{\theta}{\operatorname{argmin}}\ \ell(\theta) = \underset{a,b}{\operatorname{argmin}}\ \frac{n}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y(t_i) - ae^{-bt_i}\right)^2$$

3. Priors and integrations

- Choose a Gaussian prior for θ and define a posterior distribution. Calculate the negative log posterior. Calculate the MAP and CM estimators.

  The conditionnal mean estimator is defined by:

  $$\hat{\theta}_{CM} = \mathbb{E}(\theta|y) = \int_{\mathbb{R}^2}\theta p(\theta|y)d\theta$$

  But we know that the posterior is define as:

  $$p(\theta|y) \propto p(\theta)\mathcal{L}(\theta)$$
  $$= p(\theta)\frac{1}{\sqrt{(2\pi)^n|\sigma^2 I|}}\exp\left(-\frac{1}{2\sigma^2}\|y - f(t,\theta)\|_2^2\right)$$

  where $\|y - f(t,\theta)\|_2^2 = \sum_{i=1}^{n}\left(y(t_i) - f(t_i,\theta)\right)^2$ with $f(t_i,\theta) = ae^{-bt_i}$
  By choosing a gaussian distributed prior with mean μ and variance matrix Σ, i.e.

  $$p(\theta) = \sqrt{\frac{1}{(2\pi)^2|\Sigma|}}\exp\left(-\frac{1}{2}(\theta - \mu)^T\Sigma^{-1}(\theta - \mu)\right)$$

  Putting everything back into the conditional mean estimator formula yield

  $$\hat{\theta}_{CM} = \sqrt{\frac{1}{(2\pi)^{n+2}|\Sigma||\sigma^2 I|}}\int_{\mathbb{R}^2}\theta\exp\left(-\frac{1}{2}(\theta - \mu)^T\Sigma^{-1}(\theta - \mu) - \frac{1}{2\sigma^2}\|y - f(t,\theta)\|_2^2\right)d\theta$$

  The MAP estimator is obtained by minimizing the negative log posterior.
  The negative log posterior is

  $$\ell_p(\theta) = -\log\left(p(\theta|y)\right)$$
  $$= -\log\left(p(\theta)\right) + \frac{n}{2}\log(2\pi) + \frac{1}{2}\log(|\sigma^2 I|) + \frac{1}{2\sigma^2}\|y - f(t,\theta)\|_2^2$$

  With the choosen prior, the MAP estimator is then given by

  $$\hat{\theta}_{MAP} = \underset{\theta}{argmin}\ \ C + \frac{1}{2}(\theta - \mu)^T\Sigma^{-1}(\theta - \mu) + \frac{1}{2\sigma^2}\|y - f(t,\theta)\|_2^2$$

  Where C is a constant.

- Do the same as above, but choose a prior distribution that is uniformly distributed.
  Let's consider a uniform distributed prior with parameters c_0 and c_1 i.e.

  $$p(\theta) = \begin{cases} 1 & \text{if } a,b \in [c_0, c_1] \\ 0 & \text{else} \end{cases}$$

2

ACcording to these parameters, the CM and MAP estimate becomes:

$$\hat{\theta}_{CM} = \frac{1}{\sqrt{(2\pi)^n |\sigma^2 I|}} \int_{c_0}^{c_1} \int_{c_0}^{c_1} \theta \exp\big(-\frac{1}{2\sigma^2}\|y - f(t,\theta)\|_2^2\big) d\theta$$

and,

$$\hat{\theta}_{MAP} = \underset{\theta}{argmin} - \log(p(\theta)) + \log(|\sigma^2 I|) + \big(\frac{n}{2}\big)\log(2\pi) + \frac{1}{2\sigma^2}\|y - f(t,\theta)\|_2^2 \qquad (1)$$
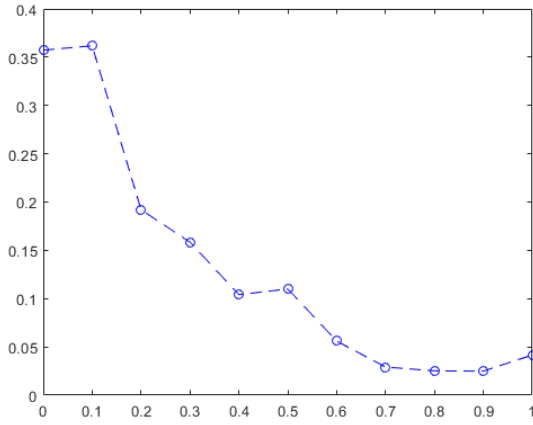
- Marginal densities of a and b:
  The marginal densities here are computed py using only the prior Gaussian distribution so $p(\theta) \sim \mathcal{N}(\mu, \Sigma)$. Therefore,

$$p(a) = \frac{1}{\sqrt{(2\pi)^2 |\Sigma|}} \int_{\mathbb{R}} \exp\big(-\frac{1}{2}(\theta - \mu)^T \Sigma^{-1}(\theta - \mu)\big) db$$

$$p(b) = \frac{1}{\sqrt{(2\pi)^2 |\Sigma|}} \int_{\mathbb{R}} \exp\big(-\frac{1}{2}(\theta - \mu)^T \Sigma^{-1}(\theta - \mu)\big) da$$

4. Visualizations

- Plot data and ground truth in the same plot.

(a) Data
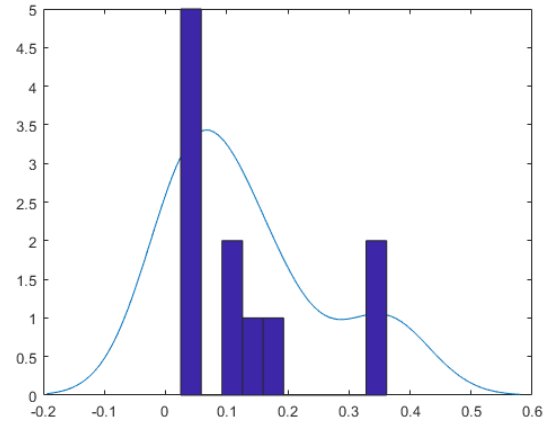
(b) Data distribution

Figure 1: Data and groundthruth

- Take all the parameter estimates calculated above and plot the corresponding curves.
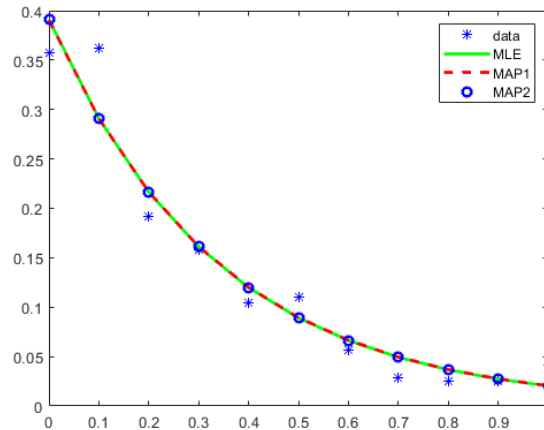

Figure 2: curves for each estimator

In the figure (2) above, with a gaussian prior of parameters $\mu = 0$ and variance $\Sigma = \sigma^2 I$ and a uniform prior of parameter $0, 1$ we can see how they all coincide with the MLE extimator.

- marginal density of a and b

  We choosed a gaussian prior distribution for θ to get the marginal density of a and b. The parameters are $\mu = [0 \quad 2]$ and $\Sigma = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}$. The implementation gives:
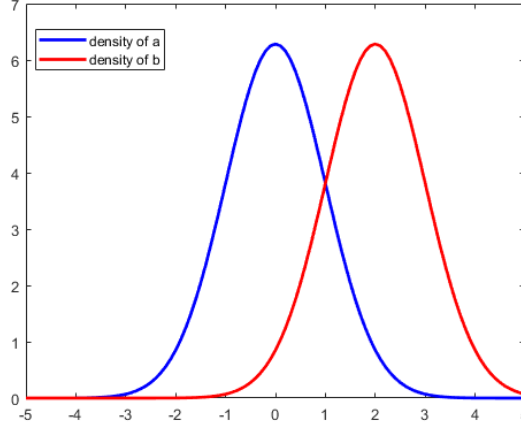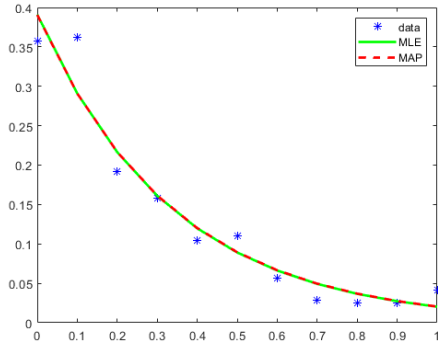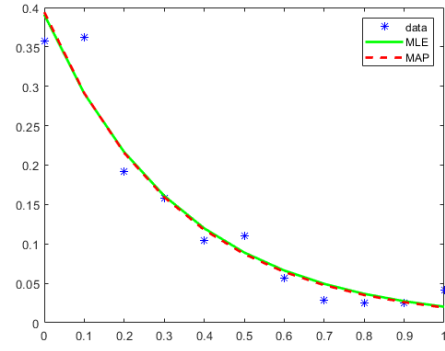

Figure 3: Marginal density of the parameter with a gaussian prior

- Tune the prior parameters and study how the parameter values affect the estimators and curves.
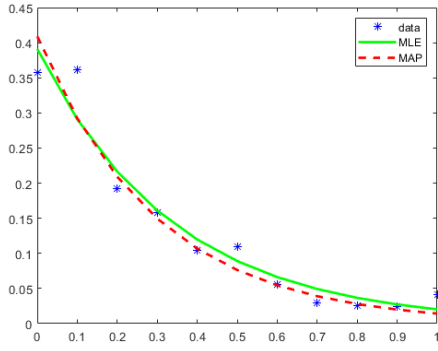
  We already found in figure (2) that the MAP estimator with the Uniform prior of parameter $0, 1$ and the gaussian prior with parameter $\mu = [0 \quad 0]$ and $\Sigma = \sigma^2 I$ coincide with the MLE estimator. During the whole experience, we found that the MLE estimator always coincide with the MAP estimator when the prior is uniformly distributed no matter what are the parameters. This can be explained by the fact in the additionnal the first term of the MAP estimator in equation (1) will be equal to zero if the prior is uniform and we only remain with the same function to minimize as in the case of the likelihood. So, we will only tuned the parameters of the Gaussian prior in this report. First we fixed the variance at $\Sigma = \sigma^2 I$ and we played with different values of the mean. Here are the result.
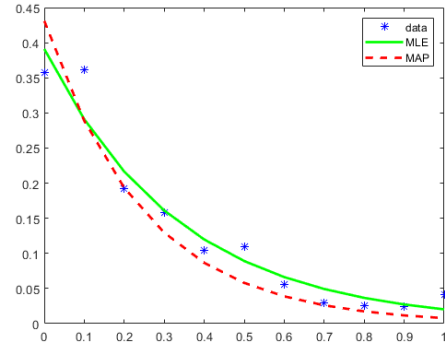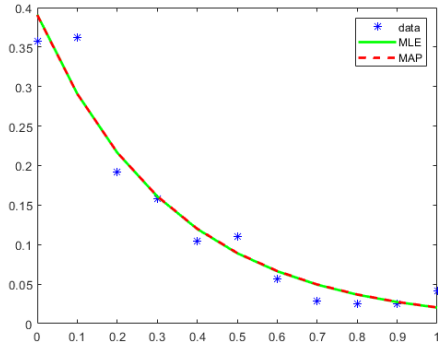
(a) $\mu = \begin{bmatrix} 0 & 0 \end{bmatrix}$

(b) $\mu = \begin{bmatrix} 10 & 20 \end{bmatrix}$

(c) $\mu = \begin{bmatrix} 50 & 100 \end{bmatrix}$

(d) $\mu = \begin{bmatrix} 100 & 200 \end{bmatrix}$

Secondly, we fixed the mean as $\mu = \begin{bmatrix} 0 & 0 \end{bmatrix}$ and we use the variance of the form $\Sigma = \sigma^2 I$ were we played with different values of σ to see what happened.
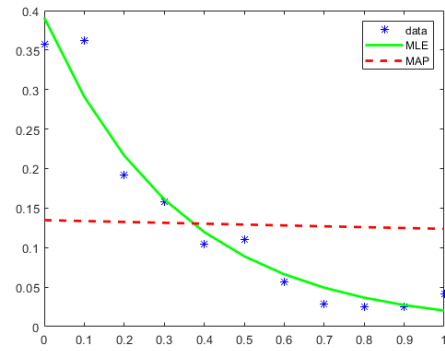

(a) $\sigma = 0.02$
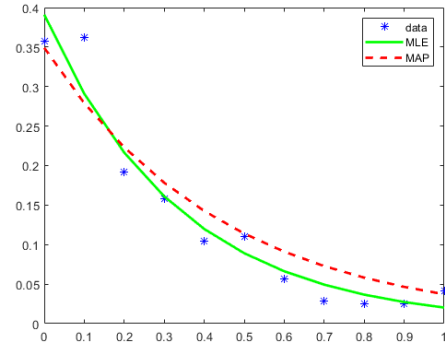
(b) $\sigma = 0.1$
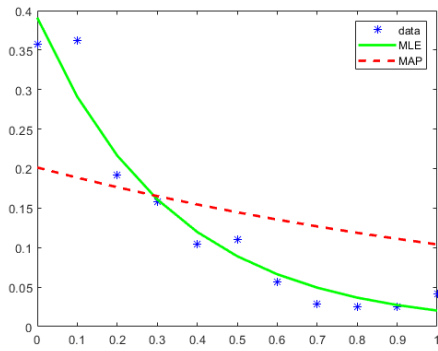
(c) $\sigma = 0.25$

(d) $\sigma = 0.5$

The graphs shows that the prior is more sensitive to the variance matrix than to the mean and

5

a small change in the variance matrix rise to a big change on the MAP estimator.