Mahesh Pandurang Mohite
IIT Kanpur

Self Project
Parameter and State estimation

Please note: New Problem starts from new page

Overview of project : (ABC Rejection Algorithm)

By passing the calculation of liklihood

ABC-Algorithm

1. Simulate the prior $\theta^* \sim \pi(\theta)$
2. Plug θ^* in the model and obtain y^* (It is equivalent to writing $y^* \sim p(y|\theta^*)$)
3. if $d(y^*, y) < \epsilon$, then store θ^* where $d(y^*, y)$ can be any distance function

Important result(s)

- When $\epsilon \to 0$ then $\pi_\epsilon(\theta|y) \to \pi(\epsilon|y)$
- When $\epsilon \to \infty$ then $\pi_\epsilon(\theta|y) \to \pi(\theta)$
- This means if we have $\epsilon \to 0$ then the information gain would be more i.e. β (as defined in the project paper)
- If we have $\epsilon \to \infty$ then we learn nothing which implies the information gain would be low

Bhattacharya distance , Bhattacharya coefficient and information gain

- B_d for two normal distributions is given by

$$B_d(f_1, f_2) = \frac{1}{4} ln\left(\frac{1}{4}\left(\frac{(\sigma_{f_1})^2}{(\sigma_{f_2})^2} + \frac{(\sigma_{f_2})^2}{(\sigma_{f_1})^2}\right)\right) + \frac{1}{4}\left(\frac{(\mu_{f_1} - \mu_{f_2})^2}{(\sigma_{f_1})^2 + (\sigma_{f_2})^2}\right)$$

- B_c is given by $B_c = e^{-B_d}$
- Information gain β is given by $\beta = 1 - B_c$, β_{θ_i} is bounded between 0 and 1 , lower and upper bounds represent the amount of new information contained in the data apart from the priors

> **Problem 1**
> Consider the model structure given as
>
> $$y[k] - a_1 y[k-1] = b_1 u[k-1] - b_2 u[k-2] + e[k]$$
>
> The priors of the parameters are given as (All the parameters have Gaussian priors a_1 : $\mathcal{N}(0.4, 0.2)$, $b_1 : \mathcal{N}(0.5, 0.2)$, $b_1 : \mathcal{N}(0.3, 0.3)$
> Using the information and data set , do the following
>
> 1. Estimate the posterior distribution of the parameters of the difference equation model using the ABC rejection method. Calculate the reduction in uncertainty from prior to posterior ?
> 2. Compute point estimates from the posterior distribution of parameters.
> 3. Compute the information gain with respect to each parameter using equation (4) [$B_c = e^{-B_d}$] where β_i is the information gain with respect to each parameter.

Solution:

1. Prior and Posterior distribution for various ϵ values are :

- A total of 3000 samples are taken , out of which for various values of ϵ different number of samples are taken(accepted) .
- Going from higher ϵ to lower ϵ , $\epsilon = [100, 10, 5, 1, 0.7]$
- The distance function utilized for simulation or in decision making step is 2-norm distance between `ystar` (generated from prior) and `y` (original data)
- Accepted values of paramters are stored in `a1_accepted` , `b1_accepted` and `b2_accepted` for a_1, b_1, b_2 parameters respectively
- By performing adtest , it is clear that posterior distribution follows gaussian distribution
- fit1 - corresponds to prior distribution , fit2 - corresponds to posterior distribution is all figures

$\epsilon = 100$

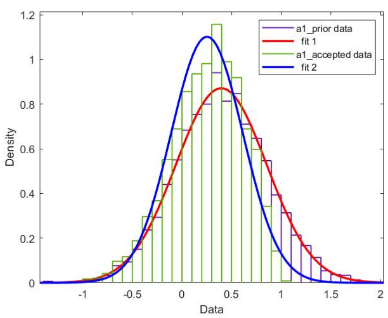

Figure 1: a_1

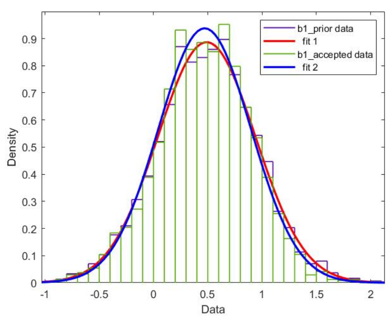

Figure 2: b_1

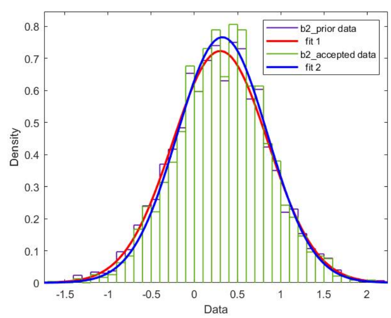

Figure 3: b_2

Figure 4: All three parameters prior (RED) (fit1) and posterior (BLUE) (fit2) distributions

$\epsilon = 10$

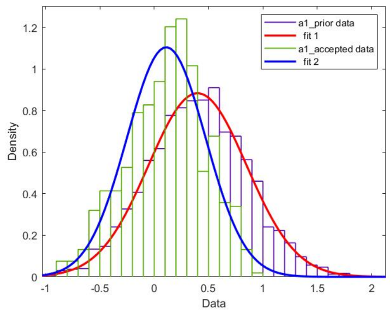

Figure 5: a_1

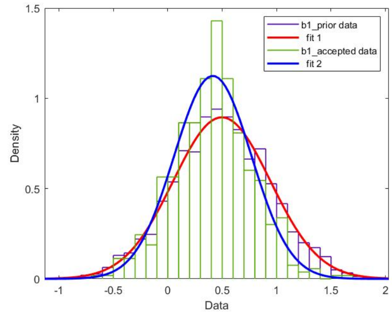

Figure 6: b_1

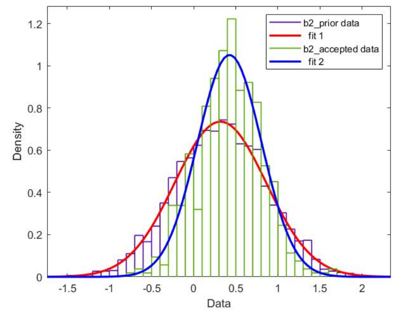

Figure 7: b_2

Figure 8: All three parameters prior (RED) and posterior (BLUE) distributions

$\epsilon = 5$

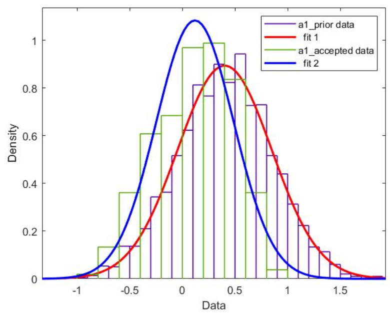

Figure 9: a_1

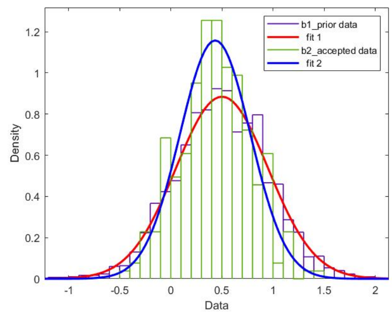

Figure 10: b_1

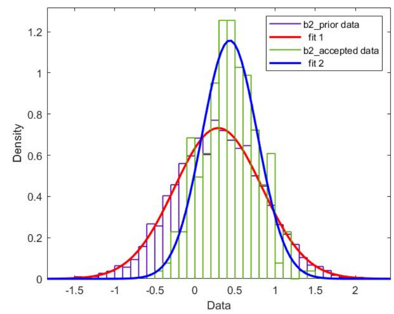

Figure 11: b_2

$\epsilon = 1$

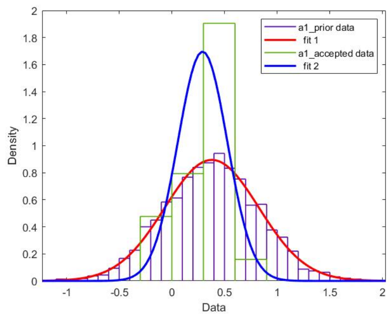

Figure 12: a_1

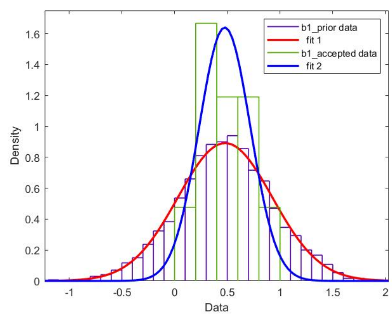

Figure 13: b_1

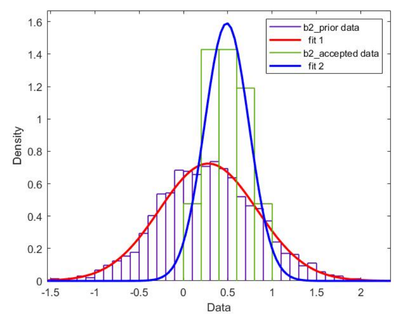

Figure 14: b_2

$\epsilon = 0.7$

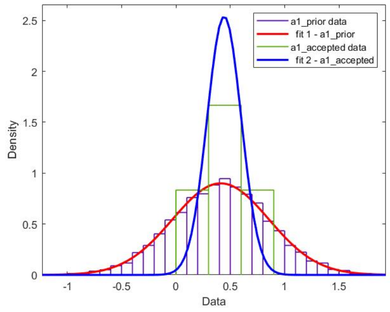

Figure 15: a_1

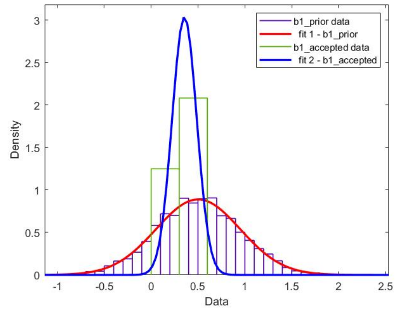

Figure 16: b_1

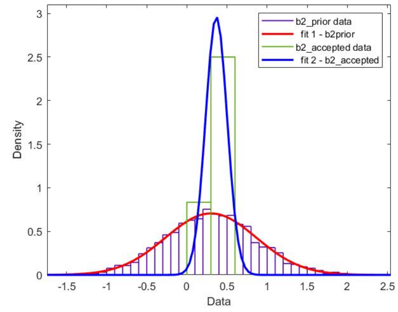

Figure 17: b_2

Figure 18:

Reduction in uncertainity from prior to posterior in all 5 cases is
(It calculated as the difference of variance of prior and variance of posterior)
Reduction is given as reduction [a_1 , b_1 , b_2] parameter respectively (also say as reduction is variability of a parameter)

- For $\epsilon = 100$ reduction is [0.0787 0.0215 0.0332]./[0.2 0.2 0.3] $\implies$ [39.35% 10.75% 11.07 %] reduction. Sample accepted : 2395
- For $\epsilon = 10$ reduction is [0.3675 0.3645 0.5013] $\implies$ [36.75% 36.45% 50.13%] reduction. Sample accepted : 532
- For $\epsilon = 5$ reduction is [0.3175 0.4750 0.5940] $\implies$ [31.75% 47.50% 59.40%] reduction. Sample accepted : 263
- For $\epsilon = 1$ reduction is [0.7165 0.7055 0.8020] $\implies$ [71.65% 70.55% 80.20%] reduction. Sample accepted : 21
- For $\epsilon = 0.7$ reduction is [0.8595 0.9195 1.0053] $\implies$ [85.95% 91.95% 100.53%] reduction. Sample accepted : 8

Here we can see if we tighten the value of epsilon , the reduction in uncertainity increases which is beneficial for us and also the number of sample accepted reduces . Here in our case we can take optimal value of epsilon to be 1 as for epsilon value 0.7 some-times the sample accepted are very less which will not give fair idea .

2. Compute point estimates from the posterior distribution of parameters :

Here in all cases posterior follows gaussian distribution , Estimate of true value would be mode of data (MAP: Maximum a posteriori probability estimate) but since it is gaussian distribution mean equal to mode and median , Hence point esimate (MAP estimate would be the mean of posterior distribution)

- For $\epsilon = 100$ `a1_est = 0.2519` , `b1_est =0.4714` , `b2_est = 0.3239`
- For $\epsilon = 10$ `a1_est = 0.1113` , `b1_est =0.4138` , `b2_est = 0.4254`
- For $\epsilon = 5$ `a1_est = 0.1194` , `b1_est =0.4177` , `b2_est = 0.4348`
- For $\epsilon = 1$ `a1_est = 0.2920` , `b1_est =0.4760` , `b2_est = 0.4882`
- For $\epsilon = 0.7$ `a1_est = 0.4439` , `b1_est =0.3555` , `b2_est = 0.3674`

3. Information Gain w.r.t. each parameter :

Information gain is stored in matrix `beta` (Matlab variable)
So , `beta(1)` gives information gain in parameter a_1
`beta(2)` gives information gain in parameter b_1
`beta(3)` gives information gain in parameter b_2

Information gain for various values of ϵ are given below : Note : Read $\begin{bmatrix} a_1 \\ b_1 \\ b_2 \end{bmatrix}$ as Information Gain in a_1 , Information Gain in b_1 ....

- For $\epsilon = 100$, $\begin{bmatrix} a_1 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 0.0285 \\ 0.0010 \\ 0.0011 \end{bmatrix} = \begin{bmatrix} 2.85\% \\ 0.10\% \\ 0.11\% \end{bmatrix}$
- For $\epsilon = 10$, $\begin{bmatrix} a_1 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 0.0715 \\ 0.0180 \\ 0.0372 \end{bmatrix} = \begin{bmatrix} 7.15\% \\ 1.80\% \\ 3.72\% \end{bmatrix}$
- For $\epsilon = 5$, $\begin{bmatrix} a_1 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 0.0663 \\ 0.0294 \\ 0.0611 \end{bmatrix} = \begin{bmatrix} 6.63\% \\ 2.94\% \\ 6.11\% \end{bmatrix}$
- For $\epsilon = 1$ $\begin{bmatrix} a_1 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 0.0983 \\ 0.0839 \\ 0.1572 \end{bmatrix} = \begin{bmatrix} 9.83\% \\ 8.39\% \\ 15.72\% \end{bmatrix}$
- For $\epsilon = 0.7$ $\begin{bmatrix} a_1 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 0.2076 \\ 0.2834 \\ 0.3310 \end{bmatrix} = \begin{bmatrix} 20.76\% \\ 28.34\% \\ 33.10\% \end{bmatrix}$

Problem 2

The information gain β is known to be used for model order estimation. Consider the structured state-space model in (1) . The given state-space model is of observable canonical form.

$$\mathbf{x}[k+1] = \begin{bmatrix} -a_1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -a_{n-1} & \vdots & \vdots & \vdots & 1 \\ -a_n & 0 & 0 & \dots & 0 \end{bmatrix} \mathbf{x}[k] + \begin{bmatrix} b_1 - a_1 * b_0 \\ \vdots \\ b_{n-1} - a_{n-1} * b_0 \\ b_n - a_n * b_0 \end{bmatrix} \mathbf{u}[k] \tag{1}$$

$$y = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix} \mathbf{x}[k] + b_0 \mathbf{u}[k] + e[k]$$

Using the given dataset, do the following

1. Using the data set given fit first-order and second-order state-space models.
2. Compare the improvement in the fit with the information loss in parameters when the order is increased
3. Can you suggest a method for the order estimation based on your observation in step 1 and step 2

Note: Take the prior for all parameters to follow , $\theta_i \sim \mathcal{N}(0.5, 1)$

Solution:

1. Using the data set given fit first-order and second-order models :

- Parameters required to fit first-order model is 3 parameters , 1 for y and 2 for u
  First order model

$$x_1[k+1] = -a_1 x_1[k] + (b_1 - a_1 b_0)\mathbf{u}[k] \tag{2}$$

$$x_1[k] = -a_1 x_1[k-1] + (b_1 - a_1 b_0)\mathbf{u}[k-1] \tag{3}$$

$$y[k] = x_1[k] + b_0 \mathbf{u}[k] \tag{4}$$

$$y[k-1] = x_1[k-1] + b_0 \mathbf{u}[k-1] \tag{5}$$

Using eq. (4) , (3) , Equation (4) can be re-written as

$$y[k] = -a_1 x_1[k-1] + (b_1 - a_1 b_0)\mathbf{u}[k-1] + b_0 \mathbf{u}[k] \tag{6}$$

From eq(5) , Putting the value of $x_1[k-1]$, we get

$$y[k] = -a_1(y[k-1] - b_0 \mathbf{u}[k-1]) + (b_1 - a_1 b_0)\mathbf{u}[k-1] + b_0 \mathbf{u}[k]$$
$$y[k] = -a_1 * y[k-1] + a_1 * b_0 \mathbf{u}[k-1] + b_1 * \mathbf{u}[k-1] - a_1 * b_0 * \mathbf{u}[k-1] + b_0 \mathbf{u}[k]$$
$$y[k] = -a_1 * y[k-1] + \cancel{a_1 * b_0 \mathbf{u}[k-1]} + b_1 * \mathbf{u}[k-1] - \cancel{a_1 * b_0 * \mathbf{u}[k-1]} + b_0 \mathbf{u}[k]$$
$$y[k] = -a_1 * y[k-1] + a_1 * b_0 \mathbf{u}[k-1] + b_1 * \mathbf{u}[k-1] - a_1 * b_0 * \mathbf{u}[k-1] + b_0 \mathbf{u}[k]$$
$$y[k] = -a_1 * y[k-1] + b_1 * \mathbf{u}[k-1] + b_0 \mathbf{u}[k]$$

Hence , the first model is

$$y[k] = -a_1 * y[k-1] + b_1 * \mathbf{u}[k-1] + b_0 * \mathbf{u}[k] + e[k] \tag{7}$$

with three parameters a_1, b_1, b_0

- Parameters required to fit second-order model is 5 parameters , 2 for y and 3 for u
  For second order model $x \in R^2, y \in R^1, u \in R^1$ and state space matrices are $A \in R^{2\times 2}, B \in R^{2\times 1}, C \in R^{1\times 2}, D \in R^{1 times 1}$
  State space model will look like (given below)

$$\begin{bmatrix} x_1[k+1] \\ x_2[k+1] \end{bmatrix} = \begin{bmatrix} -a_1 & 1 \\ -a_2 & 0 \end{bmatrix} \begin{bmatrix} x_1[k] \\ x_2[k] \end{bmatrix} + \begin{bmatrix} b_1 - a_1 * b_0 \\ b_2 - a_2 * b_0 \end{bmatrix} \mathbf{u}[k] \tag{8}$$

From state space model we can obtain below two equations

$$x_1[k+1] = -a_1 * x_1[k] + x_2[k] + (b_1 - a_1 * b_0)\mathbf{u}[k] \tag{9}$$
$$x_2[k+1] = -a_2 * x_1[k] + (b_2 - a_2 * b_0)\mathbf{u}[k] \tag{10}$$

and

$$y[k] = x_1[k] + b_0\mathbf{u}[k] + e[k] \tag{11}$$

Substituting $x_1[k]$ by using equation 9 , we get

$$y[k] = -a_1 * x_1[k-1] + x_2[k-1] + (b_1 - a_1 * b_0)\mathbf{u}[k-1] + b_0\mathbf{u}[k] \tag{12}$$

Now , Substituting $x_2[k-1]$ by using equation 10 , we get

$$y[k] = -a_1 * x_1[k-1] + -a_2 * x_1[k-2] + (b_2 - a_2 * b_0)\mathbf{u}[k-2] + (b_1 - a_1 * b_0)\mathbf{u}[k-1] + b_0\mathbf{u}[k] \tag{13}$$

Solving the above equation

$$\begin{aligned} y[k] = -a_1 * x_1[k-1] + -a_2 * x_1[k-2] + b_2 * \mathbf{u}[k-2] - a_2 * b_0 * \mathbf{u}[k-2] \\ + b_1 * \mathbf{u}[k-1] - a_1 * b_0 * \mathbf{u}[k-1] + b_0 * \mathbf{u}[k] \end{aligned}$$

Substituting $x_1[k-1], x_1[k-2]$ From equation 11 , we get

$$\begin{aligned} y[k] = -a_1 * (y[k-1] - b_0 * \mathbf{u[k-1]}) + -a_2 * (y[k-2] - b_0 * \mathbf{u[k-2]}) + b_2 * \mathbf{u}[k-2] - a_2 * b_0 * \mathbf{u}[k-2] \\ + b_1 * \mathbf{u}[k-1] - a_1 * b_0 * \mathbf{u}[k-1] + b_0 * \mathbf{u}[k] \end{aligned} \tag{14}$$

Expanding we get

$$\begin{aligned} y[k] = -a_1 * y[k-1] + a_1 * b_0 * u[k-1] + -a_2 * y[k-2] + a_2 * b_0 * \mathbf{u[k-2]} + b_2 * \mathbf{u}[k-2] - a_2 * b_0 * \mathbf{u}[k-2] \\ + b_1 * \mathbf{u}[k-1] - a_1 * b_0 * \mathbf{u}[k-1] + b_0 * \mathbf{u}[k] \end{aligned}$$

Cancelling terms

$$\begin{aligned} y[k] = -a_1 * y[k-1] + \cancel{a_1 * b_0 * u[k-1]} - a_2 * y[k-2] + \cancel{a_2 * b_0 * \mathbf{u[k-2]}} + b_2 * \mathbf{u}[k-2] - \cancel{a_2 * b_0 * \mathbf{u}[k-2]} \\ + b_1 * \mathbf{u}[k-1] - \cancel{a_1 * b_0 * \mathbf{u}[k-1]} + b_0 * \mathbf{u}[k] \end{aligned} \tag{15}$$

Hence , we get second order model as given below

$$y[k] = -a_1 * y[k-1] - a_2 * y[k-2] + b_2 * \mathbf{u}[k-2] + b_1 * \mathbf{u}[k-1] + b_0 * \mathbf{u}[k] + e[k] \tag{16}$$

Hence , the model consists of 5 parameters a_1, a_2, b_2, b_1, b_0 .
Now we have obtained the difference equation models for both order = 1 and order = 2

Order = 1

- A total of 3000 samples are taken , with the value of $\epsilon = 50$
- Model is $y[k] = -a_1 * y[k-1] + b_1 * \mathbf{u}[k-1] + b_0 * \mathbf{u}[k] + e[k]$
- Three parameters a_1, b_1, b_0 need to be estimated using ABC algorithm
- Samples accepted were 122
- prior and posterior distribution of a_1 and b_1 parameters (RED - prior , BLUE - posterior)

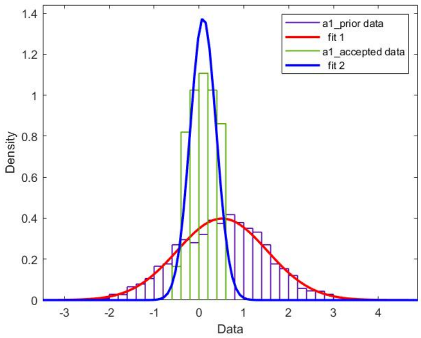

Figure 19: a_1

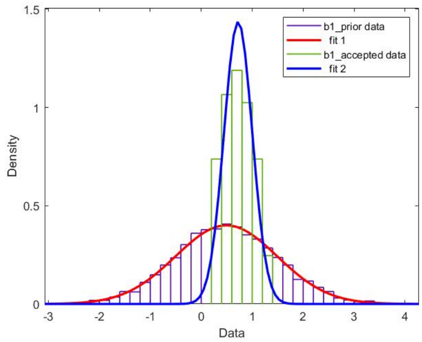

Figure 20: b_1

- Reduction in uncertainity is [0.9209 0.9174 0.9884]*100 % $\|(\mathrm{a}_1, b_1, b_0$) respectively
- All the posterior distribution of parameters follows the ad-test and follows gaussian (normal) distribution , so the point estimate is the mean of distribuion which is
  `est` =
  0.0952 0.7242 0.0284 $\|(\mathrm{a}_1, b_1, b_0$) respectively
- Information gain parameter is
  `beta` =
  0.2991 0.2899 0.3472 $\|(\mathrm{a}_1, b_1, b_0$) respectively

Order = 2

- A total of 3000 samples are taken , with the value of $\epsilon = 50$
- Model is $y[k] = -a_1 * y[k-1] - a_2 * y[k-2] + b_2 * \mathbf{u}[k-2] + b_1 * \mathbf{u}[k-1] + b_0 * \mathbf{u}[k] + e[k]$
- Three parameters a_1, a_2, b_2, b_1, b_0 need to be estimated using ABC algorithm
- Samples accepted were 12
- prior and posterior distribution of a_1 and b_1 parameters (RED - prior , BLUE - posterior)

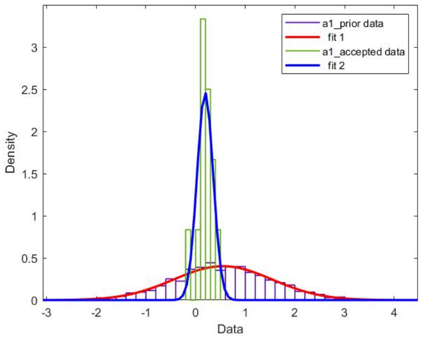

Figure 21: a_1

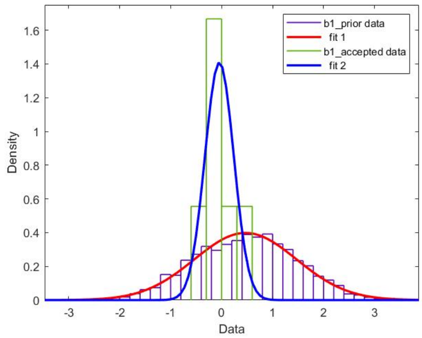

Figure 22: b_1

- Reduction in uncertainity is [0.9491 0.9590 0.9034 0.9775 0.9239]*100 % $\|(\mathrm{a}_1, b_1, b_0$) respectively
- All the posterior distribution of parameters follows the ad-test and follows gaussian (normal) distribution , so the point estimate is the mean of distribuion which is
  `est` =

  [0.1923 0.1923 -0.0443 0.2137 -0.0443] $\|(\mathrm{a}_1, a_2, b_1, b_2, b_0$) respectively
- Information gain parameter is
  `beta` =

  [0.2638 0.4541 0.2581 0.2979 0.3190] $\|(\mathrm{a}_1, a_2, b_1, b_2, b_0$) respectively

Observations

- Diagramitically fit seems to be better for order 2 than order 1 (more peakish in nature)
- Information gain parameter is more or less similar for both order for this particular case
- Variance of parameters for particular simulation is on the higher side for order-2 that is `var(a1)` (order2) $>$ `var(a1)` (order1) , for example .
- From above point we can say the point estimates for parameter(s) is better in order 1 than in order 2 due to less variability
- Other information criteria that we can make use for order estimation can be BIC index (Bayesian information Criteria) , For the two models if we have ΔBIC greater than 2 and less than 6 that is $\Delta\text{BIC} \in [2, 6]$ in favour of model which has higher BIC value .