Parameter and State Estimation (CH5115)

Term Project

November 2021

Information gain in the Bayesian framework

The precision of parameter estimates is highly dependent upon the information contained in the data. Therefore, quantifying information is an important step in data-driven modeling. Quantifying information is a well-studied problem in the frequentist approach, where Fisher Information is one of the widely used metrics. However, in Bayesian framework information is viewed as some distance measure between prior and posterior distributions. A new method for estimating information gain in the Bayesian framework using what is known as the Bhattacharyya coefficient is recently proposed.

Bhattacharyya Coefficient and the Information gain

Bhattacharyya distance (B_d) is a measure of similarity between two statistical distributions. The B_c can be used to quantify the relative closeness of two samples. The B_c of two densities $f_1(\theta)$ and $f_2(\theta)$ is given as

$$B_c = \int_{-\infty}^{\infty} \sqrt{f_1(\theta) f_2(\theta)} \tag{1}$$

The B_c is bounded between $0 \leq B_c \leq 1$. A distance measure associated with this is the B_d.

$$B_d = -ln(B_c) \tag{2}$$

The Bhattacharyya distance is bounded between $0 \leq B_d \leq \infty$. The B_d for two normal distributions can be calculated by estimating the mean and variances. The simplified form of Bhattacharyya distance for two Gaussian distributed random variables is given below

$$B_d(f_1, f_2) = \frac{1}{4} ln\left(\frac{1}{4}\left(\frac{\sigma_{f_1}^2}{\sigma_{f_2}^2} + \frac{\sigma_{f_2}^2}{\sigma_{f_1}^2} + 2\right)\right) + \frac{1}{4}\left(\frac{(\mu_{f_1} - \mu_{f_2})^2}{\sigma_{f_1}^2 + \sigma_{f_2}^2}\right) \tag{3}$$

The Bhattacharya coefficient (B_c) can be computed by

$$B_c = e^{-B_d} \tag{4}$$

The information gain is given as

$$\boxed{\beta = 1 - B_c} \tag{5}$$

The β_{θ_i} is bounded between zero and one. The lower and upper bounds represent the amount of new information contained in the data apart from the priors.

Approximate Bayesian Computation

Approximate Bayesian Computation (ABC) is is used to estimate posterior distributions of the model parameters. ABC is a computational method rooted in Bayesian inference. In Bayesian inference likelihood is a crucial component. For simple models, an analytical formula for likelihood can be derived. But, in most cases, the likelihood functions of the observations are unknown or intractable analytically. In such cases, likelihood-free methods such as ABC become very useful in estimating the posterior distribution of model parameters. ABC bypasses the evaluation of likelihood function and computed posterior distributions. In this way ABC can be applied to wide variety of complex models to estimate parameters. A brief overview of the ABC rejection algorithm is presented below,

ABC Rejection Algorithm

ABC approximates the likelihood function by numerical simulation. Let the data be y_N. The sampled prior from $f(\theta)$ is plugged into the model $\mathcal{M}$ to generate the predictions $\hat{y}_N$. The parameter is accepted as a part of posterior distribution $f(\theta|y_N)$ if

$$d(\hat{y}_N, y_N) \leq \epsilon \tag{6}$$

where $d(\hat{y}_N, y_N)$ is some distance function. A sufficiently small ϵ and an appropriate distance function will approximate the true posterior distribution reasonably well. For a detailed explanation of ABC refer [1]

Problem 1

Consider the model structure given below,

$$y[k] - a_1 y[k-1] = b_1 u[k-1] - b_2 u[k-2] + e[k] \tag{7}$$

The priors for the both the data set are given below. All the parameters have Gaussian priors. $e[k]$ follows Gaussian distribution.

Parameter	data1
a_1	$\mathcal{N}(0.4, 0.2)$
b_1	$\mathcal{N}(0.5, 0.2)$
b_2	$\mathcal{N}(0.3, 0.3)$

Table 1: Priors for the parameters

Using the given 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)

   where β_i is the information gain with respect to each parameter.

Problem 2

The information gain β is known to be used for model order estimation. Consider the structured state-space model in (8). The given state-space model is of observable canonical form. Using the given data set, do the following,

$$\mathbf{x}[k+1] = \begin{bmatrix} -a_1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \\ -a_{n-1} & \vdots & \vdots & \dots & 1 \\ -a_n & 0 & 0 & \dots & 0 \end{bmatrix} \mathbf{x}[k] + \begin{bmatrix} b_1 - a_1 b_0 \\ \vdots \\ b_{n-1} - a_{n-1} b_0 \\ b_n - a_n b_0 \end{bmatrix} \mathbf{u}[k] \tag{8}$$

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

1. Using the given data set, 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:

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

References

[1] Busetto, A. G., Numminen, E., Corander, J., Foll, M., Sunna, M., & Dessimoz, C. (2013). Approximate Bayesian Computation. 9(136461). https://doi.org/10.1371/journal.pcbi.1002803