Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 53-1 (2020) 634–639

A new index for information gain in the Bayesian framework ⋆

Prem Jagadeesan *,***,**** **Karthik Raman** **,***,**** **Arun K. Tangirala** *,***,****

* *Department of Chemical Engineering, Indian Institute of Technology (IIT) Madras*
** *Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, IIT Madras*
*** *Robert Bosch Centre for Data Science and Artificial Intelligence (RBC-DSAI), IIT Madras*
**** *Initiative for Biological Systems Engineering, IIT Madras, Chennai — 600 036, India.*

Abstract:
In data-driven dynamical modeling, precise estimation of the parameters of large models from limited data has been considered a challenging task. The precision of the parameter estimates is highly dependent upon the information contained in the data; Loss of practical identifiability and sloppiness in the model structure are major challenges in estimating parameters precisely and closely related to 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, Fisher Information computed via maximum likelihood estimation cannot accommodate any known prior knowledge about the parameters. Prior knowledge of the parameters along with informative experiments will improve the precision of the estimates. Bayesian estimation accommodates prior information in the form of a p.d.f. There has been very little work in the literature for quantifying information in the Bayesian framework. In this work, we introduce a new method for estimating information gain in the Bayesian framework using what is known as the Bhattacharyya coefficient. It is seen that the bounds of the coefficient have an insightful interpretation naturally in terms of information gain on the parameter of interest. We also demonstrate using case studies that the information gain of each parameter is an indication of loss of practical identifiability and sloppy parameters. It is also shown that the proposed information gain can be used as a model selection tool in black-box identification.

Keywords: Information Gain, Bhattacharyya Coefficient, Approximate Bayesian Computation (ABC), Model Sloppiness, Practical Identifiability and Model selection.

1. INTRODUCTION

Identifying complex models with a large number of parameters from data poses multiple challenges during an identification exercise. One of the crucial and commonly encountered challenges is identifiability, the ability to estimate a unique model. Model outputs of unidentifiable models are identical. Loss of identifiability can occur due to the nature of the model structure and/or uninformative experiments. Identifiability is a necessary condition for computing the error bounds of the parameter estimates. Practical unidentifiability is the inability to estimate parameters precisely with the given data set. Within the identifiable set of models, there exists a subset of models whose model predictions are nearly identical. Those models are called sloppy models (Gutenkunst et al., 2007). Model sloppiness can be characterized as large regions in the parameter space over which model predictions are nearly identical. Estimating parameters for sloppy models from noisy data can lead to huge parameter uncertainties (Raman et al., 2017).

Although both identifiability and sloppiness are properties of the model, they are also affected by information contained in the data and the estimation algorithm itself (Tangirala, 2014). Quantifying information contained in the data concerning a parameter becomes imperative to find a remedy for the same. Information contained in the data concerning a parameter though seems like an isolated quantity, in practice, it is coupled with the method of extraction of information. Fisher Information is one of the widely used information metrics that quantify information contained in the data concerning each parameter in the model structure, is dependent on the method of estimation (Tangirala, 2014). Fisher Information uses the likelihood to quantify the information which implicitly assumes the use of the maximum likelihood estimation algorithm. Hence use of Fisher Information as an information metric is greatly constrained by the estimation

⋆ Corresponding author, e-mail: arunkt@iitm.ac.in.

Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2020.06.106

method adopted. The fundamental limitation of the Fisher information estimated from maximum likelihood is the inability to incorporate the already known information. Fisher information will not be able to quantify any new information about a parameter apart from the already known information. Furthermore, Fisher Information is an unbounded measure.

Bayesian estimation overcomes this limitation and any known information about the parameters prior to the experiment can be incorporated in the form of a p.d.f (Tangirala, 2014). Though philosophically different, mathematically maximum likelihood is contained in the Bayesian estimation by assuming uniform prior distributions. For non-uniform priors, a more generalized version of Fisher Information called Kullback-Leibler (KL) divergence is used to find the information gain in the data from prior to the posterior distribution. KL divergence is also an unbounded measure. A new measure of information is proposed in Martin (1984) for the Bayesian framework which uses both squared Hellinger distance and Kullback-Leibler divergence to detect incorrect priors. The distance between the priors and posterior is viewed as the measure of consonance for the given model. The proposed information index in this work can also be viewed as the squared Hellinger distance.

In this work, we propose an information gain index that uses the Bhattacharyya coefficient(B_c), which has advantages that overcome the limitations of KL divergence. B_c is a measure of overlap between two statistical samples or population (Kailath, 1971) and being a bounded measure makes it a natural choice for information index in the Bayesian framework. Bhattacharyya Coefficient has been previously used in feature extraction (Coleman and Andrews, 1979) and optimal signal selection (Kailath, 1971).

In the case of complex and highly nonlinear systems, the difficulty in constructing posterior distributions paved the way to 'likelihood free' versions of Bayesian estimation known as Approximate Bayesian Computation (ABC) (Sunnåker et al., 2013). ABC finds widespread applications in systems biology to estimate complex non-linear models where sample prior and posterior distributions are obtained. The proposed new information gain index is estimated in the ABC framework. The two main contributions of the current work are

- An index for information gain from prior to posterior in a given data set assuming Gaussian priors.
- Application to the detection of practically unidentifiable parameters, sloppy parameters, and model selection.

The rest of this paper is organized as follows. Section 2 reviews the preliminary definitions pertaining to this paper. In section 3, we illustrate the problem statement and method for computing the proposed information index and its interpretation in the ABC framework. Section 4 contains simulation studies and the paper ends with some concluding remarks and future directions.

2. PRELIMINARIES

2.1 Practical/Numerical identifiability

Practical identifiability is the ability to estimate the parameters precisely with the given data set (DiStefano, 2013). It is quantified by the precision and confidence interval of the parameter estimates. Estimates of variances can be obtained using

$$\Sigma_{\hat{\theta}} = \hat{I}(\theta)^{-1} \tag{1}$$

In this work, we have used a relative confidence interval to quantify practical identifiability. 95% relative confidence interval for a Gaussian distributed random variable is computed by

$$RCI = \frac{\mu + 1.95\sigma}{\mu - 1.95\sigma} - 1 \tag{2}$$

Parameters with $RCI > 1$ indicates greater than 100% uncertainty and can be considered as practically unidentifiable. Practical identifiability is the precision of the parameter estimates and it is subjected to vary with the modeling exercise. Practical identifiability is different from the concept identifiability itself. Identifiability is binary conditionTangirala (2014). Loss of Practical identifiability can occur due to insufficient input excitation, low signal to noise ratio or due to optimization algorithms. Parameters with a wide confidence interval can also be a result of overfitting. In black-box identification, improper choice of the model structure may lead to loss of practical identifiability. In grey-box modeling, sloppy parameters under noisy measurements can lead to practical unidentifiability (Raman et al., 2017).

2.2 Model sloppiness

In certain models, there are regions in the parameter space over which the model predictions are nearly identical. It is quantified by the condition number of the Hessian of the cost function (Chis et al., 2016). The Hessian of the cost function can be approximated as

$$H_{ij} = \frac{1}{N} \sum_{n=1}^{N} \frac{\partial y}{\partial \log \theta_i} \frac{\partial y}{\partial \log \theta_j} \tag{3}$$

where y denotes model output.

While using the least square estimation algorithm assuming Gaussian data, the Hessian of the cost function is essentially an estimate of the Fisher information matrix:

$$\hat{\mathbf{I}}(\theta) = H \tag{4}$$

The sloppiness of a model is computed by the ratio of minimum to the maximum eigenvalues of the Fisher Information matrix/Hessian of the cost function. From Chis et al. (2016), a model can be considered sloppy if

$$\frac{\lambda_{\min}}{\lambda_{\max}} \leq 10^{-3} \tag{5}$$

Model sloppiness can be a result of both model structure and data (Gutenkunst et al., 2007). It is predominately observed in multi-parameter models where a large number of parameter combinations are insensitive to model output.

2.3 Bhattacharyya coefficient and Bhattacharyya distance

Bhattacharyya distance (B_d)is a measure of similarity between two statistical distributions (Bhattacharyya, 1943). 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

$$B_c = \int_{-\infty}^{\infty} \sqrt{f_1(\theta) f_2(\theta)} \tag{6}$$

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

$$B_d = -ln(B_c) \tag{7}$$

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 derived in Coleman and Andrews (1979) as:

$$B_d(f_1, f_2) = \frac{1}{4} ln\left(\frac{1}{4}\left(\frac{\sigma_{f_1}^2}{\sigma_{f_2}^2} + \frac{\sigma_{f_2}^2}{\sigma_{f_1}^2} + 2\right)\right) + \frac{1}{4}\left(\frac{(\mu_{f_1} - \mu_{f_2})^2}{\sigma_{f_1}^2 + \sigma_{f_2}^2}\right) \tag{8}$$

The Bhattacharya coefficient (B_c) can be computed by

$$B_c = e^{-B_d} \tag{9}$$

Bhattacharyya distance is a symmetric measure but does not obey triangle inequality. In this work we quantify the information gain as

$$\beta = 1 - B_c \tag{10}$$

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.

3. METHODOLOGY

In this section, we illustrate the methodology for computing the proposed information gain index in an Approximate Bayesian (ABC) computation framework.

3.1 Problem statement

Given a data set y_N, model structure $\mathcal{M}$, and prior knowledge of the parameters in the form of $f(\theta)$,

(1) Quantify the information gain from prior to posterior.
(2) Use the information gain to detect parameters that are practically unidentifiable and sloppy.
(3) Use the proposed index to select parsimonious black-box models.

3.2 Bayesian inference

At the heart of the Bayesian inference is the Bayes rule for conditional probability. The true model parameters are considered as random variables (Tangirala, 2014).

$$f(\theta|y_N) = \frac{f(y_N|\theta) f(\theta)}{f(y_N)} \tag{11}$$

$$f(\theta|y_N) = C f(y_N|\theta) f(\theta) \tag{12}$$

$f(\theta|y_N)$ is called the posterior distribution of the parameter θ as this quantity is computed after collecting data and $f(\theta)$ is the prior of the parameter θ. The constant C is adjusted to obtain a legitimate posterior p.d.f of the parameter θ. The quantity $f(y_N|\theta)$ is the likelihood function.

Priors capture the knowledge of the parameter θ *before* data; they are usually characterized by a tractable family of distributions. In the context of complex models, it is much more difficult to construct the likelihood functions. Therefore, it is important to turn towards likelihood-free methods such as the ABC rejection algorithm.

3.3 Approximate Bayesian Computation (ABC)

ABC approximates the likelihood function by numerical simulation (Sunnåker et al., 2013). The sampled prior is plugged into the model $\mathcal{M}$ to generate $\hat{y}_N$. The parameter is accepted if

$$d(\hat{y}_N, y_N) \leq \epsilon \tag{13}$$

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 (Sunnåker et al., 2013). The choice of conjugate priors helps to fix the posterior family of distributions the same as prior distributions. In this work, we focus on the normal distribution that falls in the domain of conjugate priors. In order to have a normalized ϵ we use $\epsilon = 1 - R^2$. Here R^2 is the measure of goodness of fit (Tangirala, 2014).

3.4 ABC rejection algorithm

Choose the simulation number N and acceptance threshold ϵ. Let observed data be y_N.
Step 1: Declare the Gaussian prior for each parameter $f_p(\theta_i) = \mathcal{N}(\mu_{\theta_i}, \sigma_{\theta_i}^2)$
Step 2: Sample θ_i from prior distribution
Step 3: Simulate model data set $\hat{y}_N$ from sampled θ_i
Step 4: if $d(\hat{y_N}, y_N) < \epsilon$ accept and store the parameter vector in an array A.
Step 5: Repeat Step 2 to Step 4 a total of N times.

3.5 Computing Bhattacharyya coefficient

Step 1: Compute the sample mean and variance of each parameter in the sampled prior and posterior distributions.
Step 2: Compute the Bhattacharyya distance B_d between prior and posterior distributions for each parameter θ_i using Equation 8
Step 3: Compute Bhattacharyya coefficient B_c for each parameter θ_i using Equation 9
Step 4: $(1 - B_c)$ gives the estimate of information gain.

3.6 Interpretation of the bounds of the Bhattacharyya coefficient in the ABC framework

The bounds of B_c have a meaningful interpretation in the ABC framework, which made it a natural choice of an index for information gain.

(1) If all the samples in the prior are present in posterior, then, from Equation 8 the B_d is zero and hence the information gain $\beta = 0$. This indicates that no new information is available in the data about the parameters θ_i apart from prior information.
(2) If only one of the samples from the prior is accepted in the posterior, then the variance of the posterior distribution is zero from Equation 8, the Bhattacharyya distance is infinite and hence the information gain $\beta = 1$. This indicates the information available in the data shrunk the uncertainty in the prior to the maximum.

In the ABC framework the proposed information index can be viewed as the information gain concerning a parameter with a known prior. Maximizing the information index is perceived as shrinking the uncertainty in the prior knowledge using the data.

4. SIMULATION STUDIES

Three different case studies of varying complexities are taken to demonstrate the application of the proposed information index in detecting practical identifiability, sloppy parameters and in model selection. We first illustrate a discrete-time FIR model, followed by a simple linear multiscale system, and ultimately a more realistic pharmacokinetic model used to understand the role of information gain in detecting practical identifiability, sloppiness and model selection.

4.1 A discrete-time FIR model

Consider a three-parameter FIR model given in Tangirala (2014), excited by a single frequency. This experiment is a classic example of the loss of identifiability due to data.

$$y[k] = \theta_1 u[k-1] + \theta_2 u[k-2] + \theta_3 u[k-3] \tag{14}$$

When the system is excited with the single frequency $u[k] = A\sin(2\pi f k)$, the output of the model becomes

$$y[k] = \theta_1 \sin(\omega_0 k - \phi) + \theta_2 \sin(\omega_0 k - 2\phi) + \theta_3 \sin(\omega_0 k - 3\phi) \tag{15a}$$

$$= \theta_1' \sin(\omega_0 k - \phi) + \theta_2' \sin(\omega_0 k - 3\phi) \tag{15b}$$

where $\theta_1' = \left(\theta_1 + \frac{\theta_2}{2\cos\phi}\right)$ and $\theta_2' = \left(\theta_3 + \frac{\theta_2}{2\cos\phi}\right)$

The three-parameter FIR model is manifested as a two-parameter model under single frequency input. This is due to insufficient input excitation. The model is simulated using the parameter values $\theta_1 = 1$, $\theta_2 = 0.6$, $\theta_3 = 0.3$ and with the input $u[k] = \sin(2\pi 0.1k)$. The posterior distributions are estimated with $N = 10,000$ number of simulations and threshold $\epsilon = 0.05$.

Table 1. Summary statistics of FIR(3)

θ	Prior μ	Prior σ^2	Post $\hat{\mu}$	Post $\hat{\sigma^2}$	β	RCI
θ_1	1	0.09	1.01	0.011	0.19	0.5
θ_2	0.6	0.02	0.59	0.017	0.03	1.6
θ_3	0.3	0.01	0.30	0.006	0.01	2.2

The relative confidence interval of $\theta_2 > 1$ and $\theta_3 > 1$ indicates practical unidentifiability of these parameters. The information gain column in the Table 1 indicates that the data does not contain information to estimate three parameters with good precision.

Now, an FIR(2) model is estimated with the same data fixing all the experimental conditions. The parameters are practically identifiable and information gain concerning both the parameters is quite high compared to FIR(3) model. The proposed information gain index thus can be a tool for model selection.

Table 2. Summary statistics of FIR(2)

θ	Prior μ	Prior σ^2	Post $\hat{\mu}$	Post $\hat{\sigma^2}$	β	RCI
θ_1	1	0.09	0.85	0.004	0.41	0.37
θ_2	0.6	0.02	0.95	0.003	0.67	0.27

In discrete-time black-box identification, the information gain index can also be used as a tool to detect overfitting. Parameters with low information gain can be eliminated and the model can be re-estimated.

4.2 A linear multiscale system

Consider the linear state space model. The model is multiscale with the parameter values $a = 1$, $b = 101$ and $c = 100$

$$\mathcal{M} : \begin{cases} \begin{bmatrix} \dot{x_1} \\ \dot{x_2} \end{bmatrix} = \begin{bmatrix} 0 & -a \\ b & -c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ y(t) = x_1(t) + x_2(t) + e(t), \\ e(t) \approx \mathcal{N}(0, \sigma_e^2) \end{cases} \tag{16}$$

The initial conditions of the system are $x_1(0) = x_2(0) = 1$. The output of the system y_N is corrupted by white Gaussian noise with zero mean and variance is adjusted to give SNR of the signal $y_N = 100$. The parameters have Gaussian priors, the parameters of the p.d.f are given in Table 3. The posterior distributions are estimated using the ABC rejection algorithm setting the number of simulations $N = 10000$, and the tolerance $\epsilon = 0.005$.

The Fisher information matrix is estimated by inverting the covariance matrix of the parameters. The ratio of eigenvalues quantifies the sloppiness of the model.

$$\hat{\mathbf{I}}(\theta) = \begin{bmatrix} 439.34 & 2.19 & -2.10 \\ 2.19 & 0.02 & -0.18 \\ -2.10 & -0.18 & 0.02 \end{bmatrix} \text{ and } \frac{\lambda_{min}}{\lambda_{max}} = 10^{-4}.$$

The ratio of eigenvalues indicates sloppiness. The eigenvectors corresponding to the sloppy and stiff directions are shown in the figure.

Table 3. Summary statistics of the multiscale system

θ	Prior μ	Prior σ^2	Post $\hat{\mu}$	Post $\hat{\sigma^2}$	β	RCI
a	1	1	1.00	0.004	0.61	0.29
b	100	100	100.45	71.85	0.007	0.39
c	101	100	101.27	69.30	0.007	0.38

Figures 1 and 2 indicate the projection of stiff and sloppy directions on the bare parameter axis. The parameters b

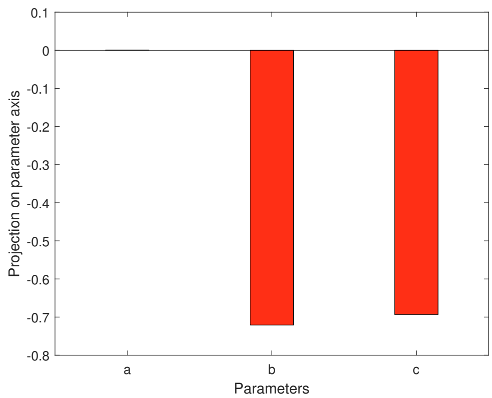

Fig. 1. Eigenvectors of the sloppy direction

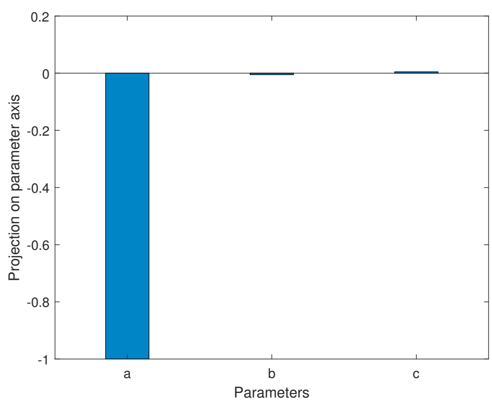

Fig. 2. Eigenvectors of the stiff direction

and c have maximum contribution to the sloppy direction and minimum contribution to the stiff direction. The information gain column (β_{θ_i}) in the Table 3 shows that parameters that contribute to the sloppy directions have extremely low information gain and the parameters that contribute to the stiff directions have high information gain.

The relative confidence interval of the parameters b and c falls in the sloppy direction is wider than the parameter a in stiff direction. But in this case, the parameters b and c are practically identifiable as the relative confidence interval is less than one. It can be seen that the presence of noise has widened the confidence intervals of the parameters that contribute to the sloppy direction. In ABC framework a parameter may become practically unidentifiable if it has a significant contribution to sloppy direction with weak (uninformative) prior.

4.3 A pharmacokinetic model

We now consider a two-compartment pharmacokinetic model with two states, one input, and four parameters, with only one of the states measured. The model equations are defined as follows:

$$\mathcal{M} : \begin{cases} \dot{x_1} &= -Kx_1 + k_{21}x_2 + bu(t) \\ \dot{x_2} &= k_{12}x_1 - k_{21}x_2 \\ y(t) &= x_1(t) + e(t) \\ e(t) &\approx \mathcal{N}(0, \sigma_e^2) \end{cases} \tag{17}$$

The model $\mathcal{M}$ is a slightly modified version of the model given in Villaverde et al. (2019). The parameter vector $\theta^* = [K\ k_{21}\ b\ k_{12}]^T = [10\ 5\ 10\ 2]^T$ is known with a prior knowledge of each parameter defined by a Gaussian p.d.f. The initial conditions of the states are assumed to be known $x_1(0) = x_2(0) = 0$.

The input $u(t)$ is a step input of magnitude 0.1 at time $t = 0$. The simulation time $t = 0$ to $t = 10$ with 101 equally sampled data points are generated. The measurement noise is added so that the signal to noise ratio (SNR) of the measurement signal (y_N) is 100. The sample posterior distribution for each parameter is constructed using ABC rejection algorithm with $N = 50000$ simulations, tolerance $\epsilon = 0.01$ and the distance function. $d(y_N, \hat{y_N}) = ||y_N - \hat{y_N}||_2^2$.

Table 4. Summary statistics of the pharmacokinetic model

θ	Prior μ	Prior σ^2	Post μ	Post σ^2	β	RCI
K	10	25	12.77	5.91	0.14	0.91
k_{21}	5	4	5.74	2.60	0.04	2.41
b	10	25	10.30	1.31	0.32	0.23
k_{12}	2	1	1.84	0.70	0.01	14.31

From the Table 4, it can be seen that the relative confidence interval of parameter $k_{21} > 1$ and $k_{12} > 10$ which indicates that these parameters have uncertainty greater than 100% and hence can be considered practically unidentifiable. From Table 4, the information gain of these respective parameters are 1% and 4% respectively, which is extremely low and indicates that there is not much additional information in this data, except the prior information.

However, the parameter K has nearly 14% and b has 32% information gain and their relative confidence intervals are less than 1 which makes them practically identifiable. It can be seen that practically identifiable parameters will have high information gain. A parameter can be considered practically unidentifiable if the information gain is extremely low for a prior with large uncertainty, otherwise called as weak prior.

Next, we use a different set of experimental conditions to study the sensitivity of the information gain to the changes in experimental conditions. we consider two cases a) Non-zero initial conditions , $x_1(0) = x_2(0) = 1$ and b) Reducing sample size to $n = 51$ data points by reducing the simulation end time $t = 5$ with zero initial conditions.

Table 5. Information gain for initial conditions $x_1(0) = x_2(0) = 1$

Parameters	RCI	β
K	1.6	0.30
k_{21}	3.0	0.11
b	1.4	0.30
k_{12}	10.8	0.0007

Table 6. Information gain for the sample size $n = 51$

Parameters	RCI	β
K	0.66	0.05
k_{21}	1.3	0.008
b	0.63	0.04
k_{12}	7.5	0.001

From Table 5 and Table 6 it is evident that the proposed information gain is affected by the quantity and quality of the data. Non-zero initial conditions have improved the information concerning the parameters K and $k21$ while the reduced sample size ended up in very low information gain concerning all the parameters in the model. The sensitivity of the information gain to experimental conditions indicates the possibility for optimal experiment design using the proposed information gain.

The case studies demonstrate the application of the proposed information index in detecting identifiable and sloppy parameters. Though there is no strict cut-off on the information gain to detect loss of practical identifiability and sloppy parameters, the proposed index gives insights into relative information content concerning each parameter which will help to design informative experiments focusing on the specific parameter of interest.

5. DISCUSSION AND CONCLUSION

Practical identifiability and sloppiness are frequently encountered challenges in any identification exercise that will result in poor parameter estimates, particularly such as those in dynamical systems-level modeling of biological systems. In this study, we propose a new index that has many advantages over conventional metrics like Fisher information and KL divergence. Firstly, this index is bounded and the bounds have a natural meaning when used in an ABC framework. The zero value of the information gain indicates that there is no new information in the data apart from the prior and a unity value indicates that the data has shrunk the uncertainty in the prior to the maximum. Second, the proposed index can be used to design experiments that would reduce sloppiness and increase the precision of the particular parameter of interest. When estimated in the Bayesian framework it can be seen that sloppy parameters become practically unidentifiable which is not necessary in case of least square estimation. Lastly, the proposed information gain can also be used to detect overfitting and as a model selection criterion in discrete-time black-box identification.

We perceived certain limitations on the proposed information gain. First, it is highly dependent on the prior distribution and the threshold ϵ. A highly informative prior and a very large ϵ may also lead to very low information gain. In such cases using the proposed information index to detect unidentifiable and sloppy parameters may be misleading. A sufficiently small ϵ and a weak prior(uninformative) will be appropriate in detecting practical identifiability and sloppiness. Although we have employed Gaussian priors in our illustrative examples, our proposed information index itself is not restricted to Gaussian priors. Indeed, the proposed method can be used for non-Gaussian priors using the Equation 6. However, in such cases, the densities have to be estimated, which may be computationally expensive, compared to merely estimating the moments of the p.d.f.

The proposed index also opens several avenues for future explorations. A natural extension of this is to use the proposed information gain index to find optimal input design that reduces sloppiness and improves identifiability. In pharmacokinetic and pharmacodynamic modeling the proposed index can be used to decide the number of compartments for a given experimental data set. In sum, we believe that this index provides another useful way to interrogate complex dynamic models, notably in the context of available data. Our approach also underlines the power of Bayesian approaches to characterize system sloppiness and identifiability.

REFERENCES

Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta math. Soc.*, 35, 99–109.

Chis, O.T., Villaverde, A.F., Banga, J.R., and Balsa-Canto, E. (2016). On the relationship between sloppiness and identifiability. *Mathematical Biosciences*, 282, 147–161.

Coleman, G.B. and Andrews, H.C. (1979). Image segmentation by clustering. *Proceedings of the IEEE*, 67, 773–785.

DiStefano, J.J. (ed.) (2013). *Dynamic Systems Biology Modeling and Simulation*. Academic Press.

Gutenkunst, R.N., Waterfall, J.J., Casey, F.P., Brown, K.S., Myers, C.R., and Sethna, J.P. (2007). Universally sloppy parameter sensitivities in systems biology models. *PLOS Computational Biology*, 3.

Kailath, T. (1971). The divergence and bhattacharyya distance measures in signal selection. *IEEE Transactions on Communication Technology*, 15, 52–60.

Martin, C.L. (1984). Applications of the distance measures between the prior and posterior distributions. *Iowa State University.*

Raman, D.V., Anderson, J., and Papachristodoulou, A. (2017). Delineating parameter unidentifiabilities in complex models. *Physical review.*

Sunnåker, M., Busetto, A.G., Numminen, E., Coranderand, J., Folland, M., and Dessimoz, C. (2013). Approximate Bayesian computation. *PLOS Computational Biology.*

Tangirala, A.K. (ed.) (2014). *Principles of System Identification: Theory and Practice*. CRC press.

Villaverde, A.F., Evans, N.D., Chappell, M.J., and Banga, J.R. (2019). Input-dependent structural identifiability of nonlinear systems. *IEEE Control Systems Letters*, 3, 272–279.