Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 10, n. 1, 2023.

Trabalho apresentado no XLII CNMAC, Universidade Federal de Mato Grosso do Sul - Bonito - MS, 2023

Medidas de distância aplicadas a Indetificabilidade de Modelos matemáticos

Douglas S. de Albuquerque[1] Renato S. Silva[2]
LNCC, Petrópolis, RJ

Modelos Matemáticos são fundamentais para entender algum fenômeno físico ou, até mesmo, realizar previsões que auxiliem tomadas de decisão. Desta maneira, determinar os parâmetros que compõem as equações do modelo é uma tarefa essencial, bem como, quantificar a incerteza associada a eles. Normalmente, os valores dos parâmetros são desconhecidos, sendo comum que sejam obtidos através de dados coletados. Este processo é chamado de Calibração.

A Inferência Bayesiana apresenta vantagem sobre os métodos tradicionais de calibração, pois permite caracterizar a distribuição de probabilidade dos parâmetros do modelo [1]. Tal fato é de extrema importância, uma vez que possibilita aplicar o efeito dessa variabilidade em previsões, além de ser uma ferramenta poderosa para avaliar o processo de calibração e a qualidade dos dados.

Um modelo é dito identificável se, dado um conjunto de parâmetros q que pertence ao espaço de parâmetros admissíveis Q, um par de parâmetros com valores diferentes aplicados na entrada do sistema gerar apenas valores diferentes na saída. Ou seja:

$$f(q) = f(q^*) \Longrightarrow q = q^* \text{ tal que } q, q^* \in Q \tag{1}$$

A identificabilidade de um modelo é uma propriedade fundamental em qualquer processo de calibração, pois permite determinar o erro associado aos parâmetros calibrados. Na Inferência Bayesiana, o problema da não-identificabilidade pode ser solucionado estabelecendo distribuições *a priori* próprias [2]. Entretanto, é possível que toda informação obtida pela distribuição *a posteriori* foi fornecida pela *a priori*, indicando que nenhuma nova informação foi aprendida no processo da inferência [2].

O conceito de identificabilidade está relacionado ao conceito de aprendizagem Bayesiana, onde o aprendizado obtido pelos parâmetros é verificado pela diferença entre as suas distribuições *a priori* e *a posteriori* [2]. Assim, neste trabalho utilizamos como critério de identificabilidade, ou como critério de aprendizagem, a distância entre tais distribuições [3, 4], pela sua simplicidade e baixo custo computacional. Como medidas foram utilizadas as distâncias: de Hellinger, de Bhattacharyya, a divergência de Kullback-Leibler e Kolmogorov-Smirnov [3, 5].

Para demonstrar e avaliar as medidas utilizadas, foi escolhido o modelo epidemiológico SIRS (**S**uscetível-**I**nfectado-**R**ecuperado-**S**uscetível) [6]. Trata-se de um modelo empregado para entender o comportamento de uma doença contagiosa a partir do contato realizado entre um indivíduo saudável com um indivíduo contaminado. De modo que, uma vez recuperado, o indivíduo não possui imunidade permanente, retornando à classe de suscetíveis após um período de tempo [6]. O modelo SIRS é definido como:

[1]dougalbu@posgrad.lncc.br
[2]rssr@lncc.br

$$\begin{aligned}\frac{dS}{dt} &= -\frac{\beta}{N}SI + wR\\ \frac{dI}{dt} &= \frac{\beta}{N}SI - \gamma I\\ \frac{dR}{dt} &= \gamma I - wR\end{aligned} \tag{2}$$

sendo β a taxa de contato, γ a taxa de recuperação, w a taxa de perda da imunidade, t o tempo e $N = S(t) + I(t) + R(t)$ a população total.

Utilizando-se a hipótese que o número N de indivíduos expostos a doença não é o número que representa a totalidade da população, o modelo se torna não-identificável, permitindo o seu uso nas medidas de avaliação da identificabilidade. Neste contexto, realizamos, computacionalmente, a sua calibração utilizando Inferência Bayesiana, através do método MCMC (*Markov Chain Monte Carlo*, Monte Carlo via Cadeias de *Markov*) [1], e checamos a identificabilidade para cada parâmetro. Os resultados apresentados na Tabela 1 evidenciam que todas as medidas foram capazes de apontar a não-identificabilidade do modelo.

Tabela 1: Valores obtidos a partir de uma Cadeia de Markov com 100000 iterações.

Medidas de Distância	Parâmetro N	Parâmetro w	Parâmetro β	Parâmetro γ
Hellinger	0.3992	0.9356	0.5008	0.8394
Bhattacharyya	0.1837	2.2211	0.3173	1.2754
Kullback-Leibler	987	15465	1386	16405
Kolmogorov-Smirnov	0.4433	0.9621	0.5201	0.6789

Agradecimentos

Esse trabalho foi financiado em parte pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – CÓDIGO DE FINANCIAMENTO 001.

Referências

[1] R. C. Smith. **Uncertainty Quantification: Theory, Implementation, and Applications**. USA: Society for Industrial e Applied Mathematics, 2013. ISBN: 161197321X.

[2] Y. Xie e B. P. Carlin. "Measures of Bayesian learning and identifiability in hierarchical models". Em: **Journal of Statistical Planning and Inference** 136.10 (2006), pp. 3458–3477. DOI: 10.1016/j.jspi.2005.04.003.

[3] C. L. Martin. "Applications of the distance measures between the prior and posterior distributions". Tese de doutorado. Ames, Iowa: Iowa State University, 1984.

[4] D. S. de Albuquerque. "Avaliação de Critérios de Convergência Aplicados ao Método MCMC". Dissertação de mestrado. LNCC, 2023.

[5] E. L. Boone, J. R. W. Merrick e M. J. Krachey. "A Hellinger distance approach to MCMC diagnostics". Em: **Journal of Statistical Computation and Simulation** 84.4 (2014), pp. 833–849.

[6] M. J. Keeling e P. Rohani. **Modeling infectious diseases in humans and animals**. New Jersey: Princeton University Press, 2011.

